Chapter 5

Newton Method for Set Optimization Problems with Set-Valued Mapping Being Finitely Many Vector-Valued Functions

5.1 Introduction

A general statement of a set-valued optimization problem is given by

$$\text{Minimize } F(x) \text{ subject to } x \in X,$$

where F is a set-valued map from a nonempty subset X of $\mathbb{R}^n$ to $\mathbb{R}^m$. For the solution concepts of set optimization problems, there are two main approaches, namely the vector approach [21, 22] and the set approach [21, 23]. In the vector approach, the decision maker's preference is based on comparing the vectors in the image set $F(X)$. An alternative definition of solutions of set optimization problems was studied [21, 23] via comparing the sets $F(x)$'s for all $x \in X$.

Optimization problems with set-valued maps have extensive applications in various areas, for instance, optimal control, game theory, mathematical economics [13], finance [11], differential inclusions [73], and many others [70, 74]. A list of detailed references of mathematical and practical applications of set optimization problems can be found in the introduction section of [1].

5.2 Motivation and Contribution

The research in the direction of set approach to defining solutions of set optimization problem was started with the works in [23, 57, 78], which considers preorder relations for comparing sets. A comprehensive discussion on this field has been discussed in [79]. The existing methods in the literature to solve set optimization problems fall into one of the following groups:

1. Algorithms *based on scalarization* have been discussed in [80, 81]. The methods proposed in these papers address a particular class of set-valued mappings characterized by robust counterparts of vector optimization problems. In [80, 81], a linear scalarization technique was utilized to analyse the optimistic solution of set optimization problem and extended the ϵ-constraint method for the ordering cones with nonnegative orthant to deal with the set optimization problem.

2. Algorithms of *sorting type* have been discussed in [82, 84, 85]. These methods deal with the set optimization problems having a finite feasible set. In [84, 85], Köbis incorporate the forward and backward reduction to the algorithms of [15, 86] and proposed extended version of the algorithms. After that, Günther and Popovici [87] evaluated the images of set-valued mappings whose values are in an increasing manner by scalarizing the mappings via a strongly monotone functional and employed a forward iteration procedure.

3. Eichfelder et al. [88] proposed a *branch and bound* technique for solving mul-

tiobjective optimization problems with decision uncertainty based on set-order relation to handling the uncertainties.

4. Jahn [90, 91] reported *derivative free* descent methods for set optimization.

5. Bouza et al. [1] introduced the study of conventional *gradient-based classical approach* (started with the steepest descent method) to solve set optimization problems with finite cardinality.

The drawbacks of the above-mentioned methods are in the first four types. These drawbacks are neatly declared in [1]. Motivated by the work of Bouza et al. [1] on generalizing gradient-based conventional methods, in this chapter, we derive a Newton method for the set optimization problems studied in [1] with a strong convexity assumption. The proposed method in this work exhibits a quadratic convergence near the optimal solution and works well for highly nonlinear objective functions (see Example 5.4). We derive optimality conditions for weakly minimal solutions. Next, we propose the Newton method for the considered set optimization problems and analyze its superlinear and quadratic convergence. Further, we provide the numerical performance of the proposed method in some test examples. Lastly, we compare the results of the proposed algorithm with the results of the steepest descent method presented in [1].

5.3 Optimality Conditions for Set-Valued Mappings

Below, we discuss the required fundamental results, basic definitions, and optimality conditions for set-valued mappings.

In this chapter, we aim to derive a Newton method to identify weakly minimal solutions of the following unconstrained set optimization problem. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a nonempty set-valued mapping. The unconstrained set optimization problem that we study is given by (SOP).

Throughout the analysis, we assume that each f^i's, $i = 1, 2, \ldots, p$ are K-strongly convex functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ with respect to a common $e \in \text{int}(K)$, i.e., there exist positive constants $\rho_1, \rho_2, \ldots, \rho_p$ and $e \in \text{int}(K)$ such that for each $i \in [p]$,

$$f^i(\lambda x_1 + (1-\lambda)x_2) \preceq \lambda f^i(x_1) + (1-\lambda) f^i(x_2) - \tfrac{1}{2}\rho_i \lambda(1-\lambda)\|x_1 - x_2\|^2 e \tag{5.1}$$

for all $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$.

Remark 5.1 *It view of Assumption 1, observe that the strong convexity condition (5.1) is equivalent to the existence of positive constants $\rho_1, \rho_2, \ldots, \rho_p$ such that for any $x, u \in \mathbb{R}^n$ and $i \in [p]$,*

$$u^\top \nabla^2 f^i(x) u \succeq \rho_i \|u\|^2 e. \tag{5.2}$$

Lemma 5.1 *A point $\bar{x}$ is a stationary point of (SOP) if and only if $\bar{x}$ is a local weakly minimal solution of (SOP).*

Proof: Let $\bar{x}$ be a local weakly minimal solution of (SOP). On the contrary, suppose $\bar{x}$ is not a stationary point of (SOP). Then, by Lemma 1.10, $\bar{x}$ is not a stationary point of (VOP) for at least one $\bar{a} = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_{\bar{w}})^\top \in P_{\bar{x}}$. Therefore, there exists $\bar{u} \in \mathbb{R}^n$ such that

$$\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} \in -\text{int}(K)\ \forall\ j \in [\bar{w}]. \tag{5.3}$$

Since $f^{a_j} : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable (Assumption 1) for all $j \in [\bar{w}]$, we have

$$f^{\bar{a}_j}(x) = f^{\bar{a}_j}(\bar{x}) + \nabla f^{\bar{a}_j}(\bar{x})^\top (x - \bar{x}) + o(\|x - \bar{x}\|), \tag{5.4}$$

where $\lim\limits_{x \to \bar{x}} \frac{o(\|x-\bar{x}\|)}{\|x-\bar{x}\|} = 0$. As $\bar{x}$ is a local weakly minimal solution of (SOP), by Lemma 1.10, $\bar{x}$ is a local weakly minimal point of (VOP) for all $a \in P_{\bar{x}}$. So, $\bar{x}$ is a local weakly minimal point of (VOP) for $\bar{a} \in P_{\bar{x}}$. Thus, there exists a neighborhood U of $\bar{x}$ such that

$$\nexists\ x \in U \text{ with } f^{\bar{a}_j}(x) - f^{\bar{a}_j}(\bar{x}) \notin -\text{int}(K) \text{ for all } j \in [\bar{w}].$$

From (5.4), there exists a neighborhood $B \subseteq U$ of $\bar{x}$ such that for all $j \in [\bar{w}]$,

$$\nabla f^{\bar{a}_j}(\bar{x})^\top (x - \bar{x}) \notin -\text{int}(K) \quad \forall\ x \in B \tag{5.5}$$

As B is a neighborhood of $\bar{x}$, there exists $\bar{t} > 0$ such that $x' = \bar{x} + \bar{t}u \in B$. The relation (5.5) with $x = x'$ yields

$$\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} \notin -\text{int}(K),$$

which is contradictory to (5.3). Therefore, $\bar{x}$ is a stationary point for (SOP).

Conversely, suppose that $\bar{x}$ is a stationary point of (SOP). As F is strongly K-convex (Assumption 1), we have for all $j \in [\bar{w}], x \in \mathbb{R}^n, \alpha \in (0, 1]$ that

$$\begin{aligned}
& \alpha f^{a_j}(x) + (1-\alpha) f^{a_j}(\bar{x}) - f^{a_j}(\alpha x + (1-\alpha)\bar{x}) \in K \\
\overset{\text{Def. 1.28}}{\Longrightarrow} \quad & (f^{a_j}(x) - f^{a_j}(\bar{x})) \preceq \tfrac{1}{\alpha}(f^{a_j}(\bar{x} + \alpha(x - \bar{x})) - f^{a_j}(\bar{x})) \\
\overset{\text{Def. 1.28}}{\Longrightarrow} \quad & (f^{a_j}(x) - f^{a_j}(\bar{x})) - \tfrac{1}{\alpha}(f^{a_j}(\bar{x} + \alpha(x - \bar{x})) - f^{a_j}(\bar{x})) \in K \\
\overset{\alpha \to 0+}{\Longrightarrow} \quad & (f^{a_j}(x) - f^{a_j}(\bar{x})) - \nabla f^{a_j}(\bar{x})(x - \bar{x}) \in K \\
\Longrightarrow \quad & \nabla f^{a_j}(\bar{x})(x - \bar{x}) \preceq (f^{a_j}(x) - f^{a_j}(\bar{x})) \\
\Longrightarrow \quad & \Psi_e(\nabla f^{a_j}(\bar{x})(x - \bar{x})) \leq \Psi_e(f^{a_j}(x) - f^{a_j}(\bar{x})).
\end{aligned} \tag{5.6}$$

Since $\bar{x}$ is a stationary point for (SOP), by Lemma 1.10, $\bar{x}$ is stationary point for (VOP) for all $a \in P_{\bar{x}}$. Thus, for any $u = x - \bar{x} \in \mathbb{R}^n$, there exists $j_u \in [\bar{w}]$ such that $\Psi_e(\nabla f^{a_{j_u}}(\bar{x})^\top u) \geq 0$. So, from (5.6) we get for any $x \in \mathbb{R}^n$ that

$$\begin{aligned}
& 0 \leq \Psi_e(\nabla f^{a_{j_u}}(\bar{x})(x - \bar{x})) \leq \Psi_e(f^{a_{j_u}}(x) - f^{a_{j_u}}(\bar{x})) \\
\overset{\text{Lemma 1.2}}{\Longrightarrow} \quad & f^{a_{j_u}}(x) - f^{a_{j_u}}(\bar{x}) \notin -\text{int } K
\end{aligned}$$

$$\implies f^{a_{j_0}}(x) \nprec f^{a_{j_0}}(\bar{x}).$$

Therefore, there exists $j_u \in [\bar{w}]$ for which there does not exist any $x \in \mathbb{R}^n$ such that $f^{a_{j_u}}(x) \prec f^{a_{j_0}}(\bar{x})$. This implies $\bar{x}$ is a weakly minimal element of (VOP) for all $a \in P_{\bar{x}}$. Therefore, by Lemma 1.10, $\bar{x}$ is a weakly minimal solution of (SOP). □

Lemma 5.2 *For any given $x \in \mathbb{R}^n$, $a \in P_x$ and $j \in [w(x)]$, the function $g : \mathbb{R}^n \to \mathbb{R}$ given by*

$$g(u) = \Psi_e\left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2}u^\top \nabla^2 f^{a_j}(x)u\right)$$

is strongly convex on $\mathbb{R}^n$.

Proof: Consider the function

$$h_j(u) = \nabla f^{a_j}(x)^\top u + \tfrac{1}{2}u^\top \nabla^2 f^{a_j}(x)u, \ u \in \mathbb{R}^n.$$

According to Remark 5.1, we have $u^\top \nabla^2 h_j(x)u \succeq \rho_j\|u\|^2 e$. Therefore, by Corollary 2.2 in [111], the function $h_j : \mathbb{R}^n \to \mathbb{R}^m$ is strongly convex. Hence, there exists $\mu > 0$ such that for any $u_1, u_2 \in \mathbb{R}^n$ and $\lambda \in [0,1]$,

$$h_j(\lambda u_1 + (1-\lambda)u_2) \preceq \lambda h_j(u_1) + (1-\lambda)h_j(u_2) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1-u_2\|^2 e. \tag{5.7}$$

Therefore, by Proposition 1.2, for any $u_1, u_2 \in \mathbb{R}^n$ and $\lambda \in [0,1]$, we have

$$\begin{aligned} & g(\lambda u_1 + (1-\lambda)u_2) \\ = \ & \Psi_e(h_j(\lambda u_1 + (1-\lambda)u_2)) \\ \overset{(5.7)}{\leq} \ & \lambda \Psi_e(h_j(u_1)) + (1-\lambda)\Psi_e(h_j(u_2)) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1-u_2\|^2 \\ = \ & \lambda g(u_1) + (1-\lambda)g(u_2) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1-u_2\|^2. \end{aligned}$$

Hence, g is strongly convex on $\mathbb{R}^n$. □

Next, we discuss a necessary condition for weakly minimal solutions of (SOP). From Lemma 5.1, we note that a weakly minimal solution of (SOP) is a stationary point of (SOP) and vice-versa. From Definition 1.37,

$$\begin{aligned} &\text{A point } \bar{x} \text{ is a stationary point of (SOP)} \\ \iff & \forall a \in P_{\bar{x}} \text{ such that for any } u \in \mathbb{R}^n\ \exists a_j \text{ with } \Psi_e(\nabla f^{a_j}(\bar{x})^\top u) \geq 0. \end{aligned}$$

So, by Proposition 1.2 (v) & (vi) and Remark 5.1, for a stationary point $\bar{x}$, for any $a \in P_{\bar{x}}$ and $u \in \mathbb{R}^n$ there exists a_j with

$$\begin{aligned} & u^\top \nabla^2 f^{a_j}(\bar{x}) u \succeq \rho_{a_j} \|u\|^2 e \\ \text{or, } & \nabla f^{a_j}(\bar{x})^\top u + u^\top \nabla^2 f^{a_j}(\bar{x}) u \succeq \nabla f^{a_j}(\bar{x})^\top u + \rho_{a_j} \|u\|^2 e \\ \text{or, } & \Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u + u^\top \nabla^2 f^{a_j}(\bar{x}) u\right) \geq \Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u + \rho_{a_j} \|u\|^2 e\right) \text{ by 1.2(iv)} \\ \text{or, } & \Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u + u^\top \nabla^2 f^{a_j}(\bar{x}) u\right) \geq \Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u\right) + \tfrac{1}{2}\rho_{a_j} \|u\|^2 \geq 0 \text{ by 1.2(vi),} \end{aligned} \tag{5.8}$$

where $\rho = \min\{\rho_1, \rho_2, \ldots, \rho_p\}$. So, for any $x \in \mathbb{R}^n$, if we define a function $\xi_x : P_x \times \mathbb{R}^n \to \mathbb{R}$ by

$$\xi_x(a, u) = \max_{j \in [w(x)]} \left\{ \Psi_e(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u) \right\}, a \in P_x,\ u \in \mathbb{R}^n, \tag{5.9}$$

then by (5.8), at a stationary point $\bar{x}$, we have

$$\begin{aligned} & \xi_{\bar{x}}(a, u) \geq 0\ \ \forall\ a \in P_{\bar{x}} \text{ and } u \in \mathbb{R}^n \\ \implies & \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) \geq 0\ \ \forall\ a \in P_{\bar{x}} \\ \implies & \forall a \in P_{\bar{x}} : 0 \leq \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) \leq \xi_{\bar{x}}(a, 0) = 0 \\ \implies & \forall a \in P_{\bar{x}} : \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) = 0. \end{aligned} \tag{5.10}$$

Moreover, as for any $x \in \mathbb{R}^n$, P_x is finite, we note from Lemma 5.2 that for any $a \in P_x$, the function $\xi_x(a, \cdot)$ is strongly convex in $\mathbb{R}^n$. Hence, the function $\xi_{\bar{x}}(a, \cdot)$ has a unique minimum over $\mathbb{R}^n$. If for $a \in P_{\bar{x}}$, $\bar{u}_{a,\bar{x}} \in \mathbb{R}^n$ be such that $\xi_{\bar{x}}(a, \bar{u}_{a,\bar{x}}) = \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u)$, then from (5.10), we have

$$\xi_{\bar{x}}(a, \bar{u}_{a,\bar{x}}) = 0 \text{ if and only if } \bar{u}_{a,\bar{x}} = 0. \tag{5.11}$$

As for any $x \in \mathbb{R}^n$, the partition set P_x is finite, ξ_x attains its minimum over the set $P_x \times \mathbb{R}^n$. Let us define a function $\Phi : \mathbb{R}^n \to \mathbb{R}$ by

$$\Phi(x) = \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a, u). \tag{5.12}$$

Then, for any $x \in \mathbb{R}^n$,

$$\Phi(x) = \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a, u) \leq \xi_x(a, 0) = 0. \tag{5.13}$$

Also, in view of (5.11) and (5.10), if for $(a, \bar{u}) \in P_{\bar{x}} \times \mathbb{R}^n$ we have $\Phi(\bar{x}) = \xi_{\bar{x}}(\bar{a}, \bar{u})$, then

$$\Phi(\bar{x}) = 0 \text{ and } \bar{u} = 0. \tag{5.14}$$

Accumulating all, we obtain the following result.

Proposition 5.1 (Necessary condition for weakly minimal points). *Let $\bar{x}$ be a weakly minimal point of (SOP) and $\bar{a} \in P_{\bar{x}}$ and $\bar{u} \in \mathbb{R}^n$ be such that $\Phi(\bar{x}) = \xi_{\bar{x}}(\bar{a}, \bar{u})$, where $\xi_{\bar{x}}$ and Φ are as defined in (5.9) and (5.12), respectively. Then, $\bar{u} = 0$.*

Next, we derive a few properties of Φ, which play an important role in the convergence analysis of the proposed Newton method for (SOP).

Proposition 5.2 *The function Φ as given in (5.12) is continuous at any $\bar{x} \in \mathbb{R}^n$.*

Proof: Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ that converges to $\bar{x} \in \mathbb{R}^n$. We show that

$$\lim_{k\to\infty} \Phi(x_k) = \Phi(\bar{x}).$$

Since the set $P_{\bar{x}}$ is finite and $\xi_{\bar{x}}$ attains its minimum over the set $P_{\bar{x}} \times \mathbb{R}^n$, there exists $(\bar{a}, \bar{u}) \in P_{\bar{x}} \times \mathbb{R}^n$ such that $\Phi(\bar{x}) = \xi_{\bar{x}}(\bar{a}, \bar{u})$.

Let (a^k, u_k) be an element in $P_{x_k} \times \mathbb{R}^n$ such that $\Phi(x_k) = \xi_{x_k}(a^k, u_k)$. Such an element (a^k, u_k) exists since the set P_{x_k} is finite and ξ_{x_k} attains its minimum over the set $P_{x_k} \times \mathbb{R}^n$. Since Ψ_e is Lipschitz continuous on $\mathbb{R}^n$ (Proposition 1.2 (iii)) and $f^{a_j^k}$ is twice continuously differentiable for each $j \in [w(x_k)]$, therefore the function ξ_{x_k} is continuous on $P_{x_k} \times \mathbb{R}^n$. Thus, we get

$$\limsup_{k\to\infty} \Phi(x_k) = \limsup_{k\to\infty} \xi_{x_k}(a^k, u_k) \le \limsup_{k\to\infty} \xi_{x_k}(\bar{a}, \bar{u}) = \xi_{\bar{x}}(\bar{a}, \bar{u}) = \Phi(\bar{x}). \quad (5.15)$$

Let $L > 0$ be the Lipschitz constant of Ψ_e. Then, from the Definition (5.12) of Φ at $\bar{x}$, we observe that

$$\begin{aligned}
&\Phi(\bar{x}) \\
&= \min_{(a,u)\in P_{\bar{x}}\times\mathbb{R}^n} \xi_{\bar{x}}(a, u) \\
&\le \xi_{\bar{x}}(a^k, u_k) \\
&= \liminf_{k\to\infty} \xi_{\bar{x}}(a^k, u_k) \text{ since } \xi_{\bar{x}} \text{ is continuous} \\
&= \liminf_{k\to\infty} \left\{ \max_{j\in[w(\bar{x})]} \left(\Psi_e(\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(\bar{x}) u_k) \right) \right\} \\
&= \liminf_{k\to\infty} \left\{ \max_{j\in[w(\bar{x})]} \left(\Psi_e(\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(\bar{x}) u_k + \nabla f^{a_j^k}(x_k)^\top u_k \right.\right. \\
&\qquad \left.\left. + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k - \nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k) \right) \right\} \\
&\overset{1.2\text{(i)}}{\le} \liminf_{k\to\infty} \left\{ \max_{j\in[w(\bar{x})]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right.\right.
\end{aligned}$$

$$
\begin{aligned}
&+\Psi_e\left(\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(\bar{x})u_k - \nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k\right)\Big\}\Big\}\\
&= \liminf_{k\to\infty}\left\{\xi_{x_k}(a^k,u_k) + \max_{j\in[w(\bar{x})]}\left\{\Psi_e\left(\nabla f^{a_j^k}(\bar{x})^\top u_k\right.\right.\right.\\
&\qquad\left.\left.\left. +\tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(\bar{x})u_k - \nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k\right)\right\}\right\}\\
&\overset{1.2\text{(iii)}}{\leq} \liminf_{k\to\infty}\left\{\xi_{x_k}(a^k,u_k) + L\max_{j\in[w(\bar{x})]}\left\{\|\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(\bar{x})u_k\right.\right.\\
&\qquad\left.\left. -\nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k\|\right\}\right\}\\
&\leq \liminf_{k\to\infty}\left\{ \xi_{x_k}(a^k,u_k) + L\max_{j\in[w(\bar{x})]}\{\|\nabla f^{a_j^k}(\bar{x}) - \nabla f^{a_j^k}(x_k)\|\|u_k\|\}\right.\\
&\quad\left. +\tfrac{L}{2}\max_{j\in[w(\bar{x})]}\left\{\|u_k^\top\left(\nabla^2 f^{a_j^k}(\bar{x}) - \nabla^2 f^{a_j^k}(x_k)\right)u_k\|\right\}\right\}. \qquad (5.16)
\end{aligned}
$$

Note that for all $j \in [w]$, each $f^{a_j^k}$ is a twice continuously differentiable and the sequence $\{x_k\}$ converges to $\bar{x}$. Also, note that there is no loss of generality if $\{u_k\}$ is assumed to be in $\{u \in \mathbb{R}^n : \|u\| \leq 1\}$. Thus, we obtain from (5.16) that

$$
\Phi(\bar{x}) \leq \liminf_{k\to\infty}\ \xi_{x_k}(a^k,u_k) = \liminf_{k\to\infty}\ \Phi(x_k). \qquad (5.17)
$$

Finally, in view of (5.15) and (5.17), we conclude that

$$
\lim_{k\to\infty}\Phi(x_k) = \Phi(\bar{x}).
$$

Thus, the function Φ is continuous at $\bar{x}$. □

Proposition 5.3 *Let U be a nonempty subset of $\mathbb{R}^n$. Suppose there exists $p, q \in \mathbb{R}_{++}$ such that for any $x \in U$ and $a \in P_x$, $\nabla^2 f^{a_j}(x) \leq qI$ for all $j \in [w(x)]$, where I is $n \times n$ identity matrix. Then, for any $a \in P_x$, there exist $\lambda_j \geq 0$, $j \in [w(x)]$, with*

$\sum_{j=1}^{[w(x)]} \lambda_j = 1$ *such that*

$$|\Phi(x)| \leq \tfrac{3L}{2q} \left\| \sum_{j=1}^{[w(x)]} \lambda_j \nabla f^{a_j}(x) \right\|^2,$$

where L *is the Lipschitz constant of* Ψ_e.

Proof: Let $x \in U$ and P_x be the partition set of (SOP) at x. Note that for any $b_1, b_2, \ldots, b_{w(x)} \in \mathbb{R}$, the identity $\max\{b_1, b_2, \ldots, b_{w(x)}\} = \max\limits_{\lambda \in \Delta_{w(x)}} \sum_{j=1}^{w(x)} \lambda_j b_j$ holds, where $\Delta_{w(x)} = \{(\lambda_1, \lambda_2, \ldots, \lambda_{w(x)}) \in \mathbb{R}_+^{w(x)} : \sum_{j=1}^{w(x)} \lambda_j = 1\}$. Thus, in view of definition (5.12) of Φ, we have for any $a \in P_x$ that

$$\begin{aligned}
|\Phi(x)| &= \left| \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a,u) \right| \\
&= \left| \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \max_{j \in [w(x)]} \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u \right) \right\} \right| \\
&= \min_{(a,u) \in P_x \times \mathbb{R}^n} \left| \max_{\lambda \in \Delta_{w(x)}} \sum_{j=1}^{[w(x)]} \lambda_j \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u \right) \right| \\
&\leq \min_{(a,u) \in P_x \times \mathbb{R}^n} \sum_{j=1}^{w(x)} \left| \lambda_j \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u \right) \right| \\
&\quad \text{for some } \lambda \in \Delta_{w(x)} \\
&\overset{1.2\text{(iii)}}{\leq} \min_{(a,u) \in P_x \times \mathbb{R}^n} \sum_{j=1}^{w(x)} \lambda_j L \left\| \nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u \right\| \\
&\leq L \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \sum_{j=1}^{w(x)} \lambda_j \left\| \nabla f^{a_j}(x)^\top u \right\| + \sum_{j=1}^{w(x)} \lambda_j \left\| \tfrac{1}{2} u^\top \nabla^2 f^{a_j}(x) u \right\| \right\} \\
&\leq L \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \sum_{j=1}^{w(x)} \lambda_j \|\nabla f^{a_j}(x)^\top u\| + \tfrac{q}{2} \|u\|^2 \right\} \text{ as } \nabla^2 f^{a_j}(x) \leq qI (5.18)
\end{aligned}$$

As the function $u \mapsto \sum_{j=1}^{w(x)} \lambda_j \|\nabla f^{a_j}(x)^\top u\| + \frac{q}{2}\|u\|^2$ is a strongly convex function on $\mathbb{R}^n$, the first-order optimality condition implies that its minimum is obtained at $u = -\frac{1}{\gamma} \sum_{j=1}^{w(x)} \lambda_j \nabla f^{a_j}(x)$. Thus, (5.18) gives for any $a \in P_x$ that there exist $\lambda_j \geq 0$, $j \in w(x)$

with $\sum_{j=1}^{w(x)} \lambda_j = 1$ such that

$$|\Phi(x)| \leq \tfrac{3L}{2q} \left\| \sum_{j=1}^{[w(x)]} \lambda_j \nabla f^{a_j}(x) \right\|^2 .$$

□

5.4 Newton Method and Its Convergence Analysis

In this section, we propose a Newton method (Algorithm 1) for set optimization problems (SOP) with an F as given in Assumption 1. We start the algorithm by selecting an arbitrary initial point. If this point does not satisfy the necessary condition for a weakly minimal point as stated in Proposition 5.1, then we proceed to update this point as discussed in Algorithm 1. At each iteration, we select an element from the partition set of the current point, and then we evaluate a descent direction for (VOP) by following the ideas of [111, 112]. Once a descent direction is found, we employ a backtracking procedure similar to the classical Armijo-type method to find an appropriate step size and then update the iterate. We keep updating the iterate until the necessary condition in Proposition 5.1 for a weakly minimal point is met. The entire method is given in Algorithm 1.

Remark 5.2 *It is to be noted that for $p = 1$ in Algorithm 1, that is, for $F(x) = \{f^1(x)\}$, the Step Step 4 of Algorithm 1 reduces to finding u_k such that*

$$u_k = \underset{u \in \mathbb{R}^n}{argmin} \;\; \Psi_e \left(\nabla f^1(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^1(x_k) u_k \right).$$

In this case, the proposed Algorithm 1 reduces to the method given in [111] and [112]. In [111], Drummond-Svaiter functional and the support of a generator of the dual cone K^* have been used, and we have used the Gerstewitz function Ψ_e.

Algorithm 1 Newton Method for Set Optimization Problem (SOP)

Step 1 **Inputs**
Provide the objective function F with $f^1, f^2, \ldots, f^p$ being twice continuously differentiable and strongly convex vector-valued functions satisfying Assumption 1.

Step 2 **Initialization**
Choose an initial point $x_0 \in \mathbb{R}^n$, a trial step length $\beta \in (0,1)$, and a positive $\nu \in (0,1)$.
Set the iteration number $k = 0$.
Provide a value of the precision level $\epsilon > 0$ for termination.

Step 3 **Calculate the minimal set and the partition set at the k-th iteration**
Compute $M_k = \text{Min }(F(x_k), K) = \{r_1, r_2, \ldots, r_{w_k}\}$ and $w_k = |\text{ Min } F(x_k), K|$.
Find $P_k = P_{x_k} = I_{r_1} \times I_{r_2} \times \cdots \times I_{r_{w_k}}$, $p_k = |P_{x_k}|$, and $P_{x_k} = \{a_1, a_2, \ldots, a_{p_k}\}$, and for each $i \in [p_k]$, $a_i = (a_i^1, a_i^2, \ldots, a_i^{w_k}) \in P_{x_k}$, $a_i^j \in I_{r_j^{x_k}}$, $j \in [w_k]$.

Step 4 **Computation of a descent direction**
Find $(a^k, u_k) \in \underset{(a,u)\in P_k \times \mathbb{R}^n}{\operatorname{argmin}} \xi_{x_k}(a,u)$, where $\xi_{x_k} : P_{x_k} \times \mathbb{R}^n \to \mathbb{R}$ is given by $\xi_{x_k}(a,u) = \max_{j\in[w_k]} \Psi_e\left(\nabla f^{a_j}(x_k)^\top u + \frac{1}{2} u^\top \nabla^2 f^{a_j}(x_k) u\right)$.

Step 5 **Stopping criterion**
If $\|u_k\| < \epsilon$, stop. Otherwise, go to Step *Step* 6.

Step 6 **Find the step length t_k by the following relation:**

$$t_k = \max_{q\in\mathbb{N}\cup\{0\}} \left\{\nu^q \ : f^{a_j^k}(x_k + \nu^q u_k) \preceq f^{a_j^k}(x_k) + \beta \nu^q \nabla f^{a_j^k}(x_k)^\top u_k \ \forall\, j \in [w_k]\right\}.$$

Step 7 **Update the iterate**
Update $x_{k+1} \leftarrow x_k + t_k u_k$ and $k \leftarrow k+1$, and go to Step *Step* 3.

The proposed Algorithm 1 extends the approach proposed in [111, 112] to the (SOP)*, and are found to be equivalent in case of vector optimization problems. It has been proved that the class of Gerstewitz functional Ψ_e is a specific instance of other methods discussed in [113].*

5.4.1 Convergence Analysis

In this section, first, we show that Algorithm 1 is well-defined. After that, we prove the convergence of Algorithm 1.

The well-definedness of Algorithm 1 essentially depends on the following two points:

(i) Existence of (a^k, u_k) in Step *Step* 4, which is assured by the discussion in the paragraph after Definition 1.37.

(ii) Existence of step length t_k in Step *Step* 6, which is assured by the result in Proposition 5.4.

Therefore, Algorithm 1 is well-defined.

Next, we characterize the stationary points of (SOP) in terms of the functions ξ_x and Φ as defined in (5.9) and (5.12), respectively.

Theorem 5.1 *Let us consider the functions ξ_x and Φ as given in (5.9) and (5.12), respectively. Let $(\bar{a}, \bar{u}) \in P_x \times \mathbb{R}^n$ be such that $\Phi(\bar{x}) = \xi_{\bar{x}}(\bar{a}, \bar{u})$. Then, the following conditions are equivalent:*

(i) The point $\bar{x}$ is a nonstationary point of (SOP).

(ii) $\Phi(\bar{x}) < 0$.

(iii) $\bar{u} \neq 0$.

Proof: (i) $\Longrightarrow$ (ii). Let us assume that the point $\bar{x}$ is a nonstationary point of (SOP). Then, in view of (1.2), there exists an $\widetilde{a} = (\widetilde{a}_1, \widetilde{a}_2, \ldots, \widetilde{a}_{\bar{w}}) \in P_{\bar{x}}$ and $u \in \mathbb{R}^n$ such that

$$\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top \widetilde{u}) < 0.$$

Thus, in view of the above relation, we have

$$\begin{aligned}
\Phi(\bar{x}) &= \min_{(a,u)\in P_{\bar{x}}\times\mathbb{R}^n} \xi_{\bar{x}}(a,u)\\
&\leq \xi_{\bar{x}}(\widetilde{a}, t\widetilde{u}) \text{ for any } t>0, \widetilde{u}\in\mathbb{R}^n, \text{ and } a\in P_{\bar{x}}\\
&= \max_{j\in[w(\bar{x})]} \Psi_e\left(\nabla f^{\widetilde{a}_j}(\bar{x})^\top t\widetilde{u} + \tfrac{1}{2}t\widetilde{u}^\top\nabla^2 f^{\widetilde{a}_j}(\bar{x})t\widetilde{u}\right)\\
&= t\max_{j\in[w(\bar{x})]} \Psi_e\left(\nabla f^{\widetilde{a}_j}(\bar{x})^\top \widetilde{u} + \tfrac{t}{2}\widetilde{u}^\top\nabla^2 f^{\widetilde{a}_j}(\bar{x})\widetilde{u}\right) \text{ by Proposition 1.2(ii)}\\
&\leq t\max_{j\in[w(\bar{x})]} \left\{\Psi_e\left(\nabla f^{\widetilde{a}_j}(\bar{x})^\top \widetilde{u}\right) + \tfrac{t}{2}\Psi_e\left(\widetilde{u}^\top\nabla^2 f^{\widetilde{a}_j}(\bar{x})\widetilde{u}\right)\right\} \text{ by Proposition 1.2(i)\&(ii)}\\
&\leq t\left\{\max_{j\in[w(\bar{x})]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\} + \tfrac{t}{2}\max_{j\in[w(\bar{x})]}\{\Psi_e(\widetilde{u}^\top\nabla^2 f^{\widetilde{a}_j}(\bar{x})\widetilde{u})\}\right\}. \qquad (5.19)
\end{aligned}$$

Choosing any t such that $0 < t < \left(\frac{-2}{\max_{j\in[w(x)]}\{\Psi_e(\widetilde{u}^\top\nabla^2 f^{\widetilde{a}_j}(\bar{x})\widetilde{u})\}}\right)\left(\max_{j\in[w(x)]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\}\right)$, we obtain from (5.19) that

$$\Phi(\bar{x}) < t\left\{\max_{j\in[w(\bar{x})]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\} - \max_{j\in[w(\bar{x})]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\}\right\} = 0.$$

(ii) $\Longrightarrow$ (iii). It trivially follows from (5.14).

(iii) $\Longrightarrow$ (i). Let us assume contrarily that $\bar{x}$ is a stationary point of (SOP) and $\bar{u} \neq 0$. Then, in view of (1.2), for $\bar{a}\in P_{\bar{x}}$, there exists $\widetilde{j}\in[\bar{w}]$ such that

$$\Psi_e(\nabla f^{\bar{a}_{\widetilde{j}}}(\bar{x})^\top\bar{u}) \geq 0. \qquad (5.20)$$

Note from Assumption 1 that for any $a\in P_x$ and $x\in\mathbb{R}^n$, we have $\bar{u}^\top\nabla^2 f^{\bar{a}_{\widetilde{j}}}(\bar{x})\bar{u} > 0$. Therefore, from (5.20) with the help of Proposition 1.2(iv), we get

$$\begin{aligned}
&\nabla f^{\bar{a}_{\widetilde{j}}}(\bar{x})^\top\bar{u} + \tfrac{1}{2}\bar{u}^\top\nabla^2 f^{\bar{a}_{\widetilde{j}}}(\bar{x})\bar{u} \geq 0\\
\text{or,}\quad &\Psi_e\left(\nabla f^{\bar{a}_{\widetilde{j}}}(\bar{x})^\top\bar{u} + \tfrac{1}{2}\bar{u}^\top\nabla^2 f^{\bar{a}_{\widetilde{j}}}(\bar{x})\bar{u}\right) \geq 0
\end{aligned}$$

$$\begin{aligned}
&\text{or,} \quad \max_{j\in[w(\bar{x})]} \left\{ \Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} + \tfrac{1}{2}\bar{u}^\top \nabla^2 f^{\bar{a}_j}(\bar{x})\bar{u}) \right\} \geq 0 \\
&\text{or,} \quad \xi_{\bar{x}}(\bar{a}, \bar{u}) \geq 0 \\
&\text{or,} \quad \Phi(\bar{x}) = 0 \text{ from (5.13)} \\
&\text{or,} \quad \bar{u} = 0 \text{ from (5.14)},
\end{aligned}$$

which is a contradiction to the considered assumption. Thus, $\bar{x}$ is a nonstationary point of (SOP). □

Remark 5.3 *In view of (5.13), (5.14), and statements (i)–(iii) of Theorem 5.1, we obtain that $\bar{x}$ is a stationary point of (SOP) if and only if $\Phi(\bar{x}) = 0$ or $\bar{u} = 0$.*

In the next theorem, we characterize an upper bound for the norm of Newton direction u_k, generated by Algorithm 1, for (SOP).

Theorem 5.2 *Let $\{x_k\}$ be the sequence of nonstationary points, $\{u_k\}$ be a sequence of directions generated by Algorithm 1, and $\{x_k\}$ be convergent. Then, the sequence $\{u_k\}$ is bounded.*

Proof: Let P_{x_k} be the partition set at x_k. Then, by Theorem 5.1, there exists $a^k \in P_{x_k}$ such that

$$\begin{aligned}
& \max_{j\in[w(x_k)]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k) \right\} < 0 \\
\Longrightarrow \quad & \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2}u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k) < 0 \ \forall j \in [w(x_k)] \\
\overset{(5.2)}{\Longrightarrow} \quad & \tfrac{1}{2}\rho_{a_j^k}\|u_k\|^2 < -\Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k) \ \forall j \in [w(x_k)] \\
\Longrightarrow \quad & \tfrac{1}{2}\rho_{a_j^k}\|u_k\|^2 < \max_{j\in[w(x_k)]} \{|\Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k)|\} \leq L \max_{j\in[w(x_k)]} \|\nabla f^{a_j^k}(x_k)^\top u_k\| \\
& \text{from Proposition 1.2(iii) and } L \text{ is a Lipschitz constant of } \Psi_e. \qquad (5.21)
\end{aligned}$$

Note that for every $k \in \mathbb{N}$, $f^{a_j^k}$'s are twice continuously differentiable and the sequence

$\{x_k\}$ is convergent. Therefore, there exists a positive constants C such that

$$C = \max_{j\in[w(x_k)]} \|\nabla f^{a_j^k}(x_k)\|. \tag{5.22}$$

Let $\rho = \min\{\rho_1, \rho_2, \ldots, \rho_p\}$, where $\rho_1, \rho_2, \ldots, \rho_p$ are as given in (5.2). Then, $\rho > 0$ and in view of (5.21) and (5.22), we observe that

$$\rho\|u_k\|^2 \leq 2CL\|u_k\| \implies \|u_k\| \leq \frac{2CL}{\rho}.$$

Thus, the sequence $\{u_k\}$ is bounded. □

Next, to prove the convergence of the proposed Algorithm 1, we first give a proposition on the existence of a step size along the chosen (descent) direction of F for the set optimization problem (SOP) by Algorithm 1 and discuss the notion of the regularity of a point for a set-valued mapping.

Proposition 5.4 *Let $\beta \in (0,1)$ and $(\bar{a}, \bar{u}) \in P_{\bar{x}} \times \mathbb{R}^n$ be such that $\Phi(\bar{x}) = \xi_{\bar{x}}(\bar{a}, \bar{u})$ and assume that the point $\bar{x}$ is not a stationary point of (SOP). Then, there exists $\widetilde{t} > 0$ such that for all $t \in (0, \widetilde{t}]$ and $j \in [\bar{w}]$,*

$$f^{\bar{a}_j}(\bar{x} + t\bar{u}) \preceq f^{\bar{a}_j}(\bar{x}) + \beta t \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}. \tag{5.23}$$

Additionally, for all $t \in (0, \widetilde{t}]$ and $j \in [\bar{w}]$, we have

$$F(\bar{x} + t\bar{u}) \preceq^l \{f^{\bar{a}_j}(\bar{x}) + \beta t \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}\}_{j\in[\bar{w}]} \prec^l F(\bar{x}). \tag{5.24}$$

Proof: Suppose (5.23) do not hold. Therefore, there exists a sequence $\{t_k\} \searrow 0$ and $j' \in [\bar{w}]$ such that

$$f^{\bar{a}_{j'}}(\bar{x} + t_k\bar{u}) - f^{\bar{a}_{j'}}(\bar{x}) - \beta t_k \nabla f^{\bar{a}_{j'}}(\bar{x})^\top \bar{u} \notin -K$$

$$\begin{aligned}
&\implies \lim_{k\to 0}\frac{f^{\bar{a}_{j'}}(\bar{x}+t_k\bar{u})-f^{\bar{a}_{j'}}(\bar{x})}{t_k}-\beta\nabla f^{\bar{a}_{j'}}(\bar{x})^\top\bar{u}\notin -K\\
&\implies (1-\beta)\nabla f^{\bar{a}_{j'}}(\bar{x})^\top\bar{u}\notin -\text{int}(K)\\
&\implies \nabla f^{\bar{a}_{j'}}(\bar{x})^\top\bar{u}\notin -\text{int}(K) \text{ since } \beta\in(0,1).
\end{aligned}\tag{5.25}$$

Note that $\bar{x}$ is not a stationary point of (SOP) and $(\bar{a},\bar{u})\in P_{\bar{x}}\times\mathbb{R}^n$ is such that $\Phi(\bar{x})=\xi_{\bar{x}}(\bar{a},\bar{u})$. Therefore, in view of Theorem 5.1, $\bar{u}\neq 0$ and

$$\begin{aligned}
&\xi_{\bar{x}}(\bar{a},\bar{u})=\Phi(\bar{x})<0\\
\implies &\max_{j\in[\bar{w}]}\Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u}+\tfrac{1}{2}\bar{u}^\top\nabla^2 f^{\bar{a}_j}(\bar{x})\bar{u})<0 \text{ since } \bar{u}\neq 0\\
\implies &\Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u}+\tfrac{1}{2}\bar{u}^\top\nabla^2 f^{\bar{a}_j}(\bar{x})\bar{u})<0 \text{ for all } j\in[\bar{w}]\\
\overset{(5.2)}{\implies} &\Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u})+\tfrac{1}{2}\rho_{\bar{a}_j}\|\bar{u}\|^2<0 \text{ for all } j\in[\bar{w}]\\
\implies &\Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u})<0\\
\implies &\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u}\in-\text{int}(K),
\end{aligned}\tag{5.26}$$

which is contradictory to (5.25). Therefore, for every $j\in[\bar{w}]$, relation (5.23) holds.

Now, from (5.23) and Proposition 1.1, we observe that for every $t\in(0,\bar{t}]$,

$$\begin{aligned}
F(\bar{x}) &\subseteq \{f^{\bar{a}_j}(x)\}_{j\in[\bar{w}]}+K\\
&\subseteq \{f^{\bar{a}_j}(\bar{x})+\beta t\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u}\}_{j\in[\bar{w}]}+K+\text{ int}(K)\\
&\subseteq \{f^{\bar{a}_j}(\bar{x}+t\bar{u})\}_{j\in[\bar{w}]}+K+K+\text{ int}(K) \text{ from Definition 1.28}\\
&\subseteq F(\bar{x}+t\bar{u})+\text{ int}(K),
\end{aligned}$$

which implies that for every $t\in(0,\bar{t}]$, we have $F(\bar{x}+t\bar{u})\prec^l F(\bar{x})$. □

Below, we define the notion of the regularity of a point with an essential property for a set-valued mapping, which has a significant role in the convergence of the proposed algorithm.

Definition 5.1 (Regular point [1]). *Let U be a nonempty subset of $\mathbb{R}^n$. A point $\bar{x} \in U$ is said to be a regular point of F if it satisfies the following conditions:*

(i) Min $(F(\bar{x}), K) =$ WMin $(F(\bar{x}), K)$, and

(ii) the cardinality function w in Definition 1.35 is constant in the neighbourhood of $\bar{x}$.

Lemma 5.3 (See [1]). *Let us assume that $\bar{x} \in \mathbb{R}^n$ is a regular point of F. Then, there exists a neighbourhood U of $\bar{x}$ such that for every $x \in U$, $w(x) = \bar{w}$, and $P_x \subseteq P_{\bar{x}}$.*

Now, we present the main theorem of the chapter that proves the convergence of the proposed Algorithm 1.

Theorem 5.3 *Let $\{x_k\}$ be a sequence of nonstationary points generated by Algorithm 1 and $\bar{x}$ be one of its accumulation points. Additionally, assume that $\bar{x}$ is a regular point of F. Then, $\bar{x}$ is a stationary point of (SOP).*

Proof: Let $\{x_k\}$ be a sequence of nonstationary points and $\bar{x}$ be an accumulation point of $\{x_k\}$. We prove that $\bar{x}$ is stationary. Towards this, define a function $\varsigma : \mathcal{P}(\mathbb{R}^m) \to \mathbb{R} \cup \{-\infty\}$ by

$$\varsigma(A) = \inf_{z \in A} \Psi_e(z).$$

By Proposition 1.2(iv), the function ς is monotonic with respect to the preorder $\preceq^l$, i.e., for all $A, B \in \mathcal{P}(\mathbb{R}^m)$, we have

$$A \preceq^l B \implies \varsigma(A) \leq \varsigma(B). \tag{5.27}$$

Now in view of (5.24) of Proposition 5.4, for every $k = 0, 1, 2, \ldots$, we obtain

$$\varsigma(F(x_{k+1}))$$

$$
\begin{aligned}
&= \varsigma(F(x_k + t_k u_k)) \\
&\leq \min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) + \beta t_k \nabla f^{a_j^k}(x_k)^\top u_k \right) \right\} \\
&\leq \min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) + \beta t_k \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_{k,j}}(x_k) u_k \right) \right) \right\} \\
&\quad \text{from Proposition 1.2(iv) and } u_k^\top \nabla^2 f^{a_{k,j}}(x_k) u_k \succ 0 \\
&\leq \min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right\} \\
&\quad \text{from Proposition 1.2(i)} \\
&\leq \min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \max_{j'\in[w_k]} \Psi_e \left(\nabla f^{a_{j'}^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_{j'}^k}(x_k) u_k \right) \right\} \\
&\leq \min_{j\in[w_k]} \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \max_{j\in[w_k]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right\} \\
&= \varsigma(F(x_k)) + \beta t_k \Phi(x_k). \qquad (5.28)
\end{aligned}
$$

Therefore,

$$
-\beta t_k \max_{j\in[w_k]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right\} \leq \varsigma(F(x_k)) - \varsigma(F(x_{k+1})). \quad (5.29)
$$

On adding the above relation for $k = 0, 1, \ldots, \kappa$, we obtain

$$
-\beta \sum_{k=0}^{\kappa} t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right\} \leq \varsigma(F(x_0)) - \varsigma(F(x_{\kappa+1})). \quad (5.30)
$$

Since $\bar{x}$ is an accumulation point of the sequence $\{x_k\}$, we can find a subsequence $\mathcal{K} \in \mathbb{N}$ such that

$$
\{x_k\}_{k\in\mathcal{K}} \to \bar{x}, \ \{t_k\}_{k\in\mathcal{K}} \to \bar{t}, \quad \text{and} \quad \{u_k\}_{k\in\mathcal{K}} \to \bar{u}.
$$

In view of (5.13), the function ς in (5.28) is monotonic. Therefore, from (5.30), taking $\kappa \to \infty$ we deduce that

$$
-\beta \lim_{k\to\infty} \sum_{k=0}^{\kappa} t_k \max_{j\in[w_k]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right\} \leq +\infty. \quad (5.31)
$$

Now, note that if x_k is not a stationary point, then in view of (1.2), for every $a^k \in P_{x_k}, u_k \in \mathbb{R}^n, j \in [w(x_k)]$, and $k \in \mathcal{K}$, we get $\nabla f^{a_j^k}(x_k)^\top u_k \in -\operatorname{int}(K)$. On proceeding in a similar manner to (5.19), we can find $t_k > 0$ such that

$$0 < t_k < \left(\frac{-2}{\max\limits_{j\in[w_k]}\{\Psi_e(u_k^\top \nabla^2 f^{a_j^k}(x_k)u_k)} \right) \left(\max_{j\in[w_k]}\{\Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k)\} \right).$$

In view of the above chosen $t_k > 0, k \in \mathcal{K}$, we conclude that

$$\begin{aligned} & t_k \max_{j\in[w_k]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right) \right\} \\ \leq \; & t_k \left\{ \max_{j\in[w(\bar{x})]} \{\Psi_e(\nabla f^{a_j^k}(\bar{x})^\top \bar{u})\} - \max_{j\in[w(\bar{x})]} \{\Psi_e(\nabla f^{a_j^k}(\bar{x})^\top \bar{u})\} \right\} \\ = \; & 0. \end{aligned}$$

Therefore, we get

$$-t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right\} \geq 0. \tag{5.32}$$

On combining (5.31) and (5.32), and taking limit $k \to \infty$, we have

$$0 \leq -\sum_{k=0}^{\infty} t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \right\} \leq +\infty. \tag{5.33}$$

Hence, we obtain

$$\lim_{k\to\infty,\; k\in\mathcal{K}} t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k) \right\} = 0. \tag{5.34}$$

As the number of points in $[p]$ is finite, and $\bar{x}$ is a regular point of F, therefore in view of Lemma 5.3, for all $k \in \mathcal{K}, u \in \mathbb{R}^n$, we have $w_k = \bar{w}, P_{x_k} = \bar{P}, a^k = \bar{a} \in \bar{P}$ and

$$\Phi(x_k) = \xi_{x_k}(\bar{a}, u_k) \leq \xi_{x_k}(a, u)$$

$$\text{and} \quad \xi_{\bar{x}}(\bar{a}, \bar{u}) \leq \xi_{\bar{x}}(a, u) \text{ on taking limit } k \to \infty, k \in \mathcal{K}. \tag{5.35}$$

Now, we analyze the following two cases:

(i) Let $\bar{t} > 0$. In view of (5.34) and for all $k \in \mathcal{K}$ such that $w_k = \bar{w}, P_{x_k} = \bar{P}, a^k = \bar{a}$, we have

$$\begin{aligned} & \lim_{k \overset{\mathcal{K}}{\to} +\infty} \max_{j \in [\bar{w}]} \left\{ \Psi_e(\nabla f^{\bar{a}_j}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{\bar{a}_j}(x_k) u_k \right\} = 0 \\ \implies \quad & \lim_{k \overset{\mathcal{K}}{\to} +\infty} \Phi(x_k) = 0, \text{ i.e., } \Phi(\bar{x}) = 0. \end{aligned} \tag{5.36}$$

Thus, by Theorem 5.1, $\bar{u} = 0$. Hence, $\bar{x}$ is a stationary point of (SOP).

(ii) Let $\bar{t} = 0$. Fix any $\kappa \in \mathbb{N}$. Since $t_k \overset{\mathcal{K}}{\to} \bar{t} = 0$, large enough ν^κ does not satisfy Armijo condition in Step *Step* 6 of Algorithm 1. Therefore for all $k \in \mathcal{K}$ such that $w_k = \bar{w}, \; P_{x_k} = \bar{P}$, and $a^k = \bar{a}$, there exists $\bar{j} \in [\bar{w}]$ such that

$$\begin{aligned} & f^{\bar{a}_{\bar{j}}}(x_k + \nu^\kappa u_k) \not\preceq f^{\bar{a}_{\bar{j}}}(x_k) + \beta \nu^\kappa \nabla f^{\bar{a}_{\bar{j}}}(x_k)^\top u_k \\ \implies \quad & \frac{f^{\bar{a}_{\bar{j}}}(x_k + \nu^\kappa u_k) - f^{\bar{a}_{\bar{j}}}(x_k)}{\nu^\kappa} - \beta \nabla f^{\bar{a}_{\bar{j}}}(x_k)^\top u_k \notin -K \\ \implies \quad & \frac{f^{\bar{a}_{\bar{j}}}(\bar{x} + \nu^\kappa \bar{u}) - f^{\bar{a}_{\bar{j}}}(\bar{x})}{\nu^\kappa} - \beta \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ taking } k \overset{\mathcal{K}}{\to} +\infty \\ \implies \quad & (1 - \beta) \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ taking limit } k \to +\infty \\ \implies \quad & \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ since } (1 - \beta) \in (0, 1). \end{aligned}$$

Therefore, from (v) of Proposition 1.2, we have

$$\begin{aligned} & \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u}) \geq 0 \\ \text{or,} \quad & \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} + \tfrac{1}{2} \bar{u}^\top \nabla^2 f^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u}) \geq 0 \\ & \quad \text{from (iv) of Proposition 1.2 and } \bar{u}^\top \nabla^2 f^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u} \succ 0 \\ \text{or,} \quad & 0 \leq \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} + \tfrac{1}{2} \bar{u}^\top \nabla^2 f^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u}) \end{aligned}$$

$$\text{or,} \quad 0 \leq \xi_{\bar{x}}(\bar{a}, \bar{u}) = \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a, u) = \Phi(\bar{x}) \leq 0 \text{ from (5.13).}$$

Thus, from Theorems 5.1, we conclude that $\bar{x}$ is a stationary point of (SOP).

□

Next, we analyze the rate of convergence of the proposed Algorithm 1. It is found in the following Theorem 5.4 that under suitable assumptions, the step length t_k is eventually 1, and a subsequence of the generated sequence $\{x_k\}$ by Algorithm 1 converges superlinearly to a locally efficient solution. Towards this, at first, we recall the following result.

Lemma 5.4 (See [112]). *Let V be a nonempty convex subset of $\mathbb{R}^n$, and $\epsilon > 0$ and $\delta > 0$ be such that for any $x, y \in V$ with $\|y - x\| < \delta$,*

$$\|\nabla^2 f^{a_j}(y) - \nabla^2 f^{a_j}(x)\| < \epsilon \text{ for all } j \in [w(x)]. \tag{5.37}$$

Then, for every $j \in [w(x)]$, we have

$$\|\nabla f^{a_j}(y) - [\nabla f^{a_j}(x) + \nabla^2 f^{a_j}(x)^\top (y - x)]\| < \epsilon \|y - x\|. \tag{5.38}$$

If $\nabla^2 f^{a_j}$ is Lipschitz continuous on V with constant $\widetilde{L}$ for all $j \in [w(x)]$, then

$$\|\nabla f^{a_j}(y) - [\nabla f^{a_j}(x) + \nabla^2 f^{a_j}(x)^\top (y - x)]\| < \tfrac{\widetilde{L}}{2} \|y - x\|^2 \text{ for all } j \in w[(x)]. \tag{5.39}$$

Theorem 5.4 (Superlinear convergence). *Let $\{x_k\}$ be a sequence of nonstationary points generated by Algorithm 1 and $\bar{x}$ be one of its accumulation points. Additionally, assume that $\bar{x}$ is a regular point of F, and there exists a nonempty convex set $V \subseteq \mathbb{R}^n$ and $p > 0, q > 0, \delta > 0, \epsilon > 0$ for which the following conditions hold:*

(i) $\nabla^2 f^{a_j}(x) \leq qI$ *for all $j \in [w(x)]$, where I is $n \times n$ identity matrix,*

(ii) $\|\nabla^2 f^{a_j}(x) - \nabla^2 f^{a_j}(y)\| < \epsilon$ *for all* $x, y \in V$ *with* $\|x - y\| < \delta$*, and*

(iii) $\frac{\epsilon}{q} \leq 1 - 2\beta$.

Then, for sufficiently large $k \in \mathbb{N}$*, we have* $t_k = 1$ *and there exists a subsequence of* $\{x_k\}$ *that converges superlinearly to* $\bar{x}$.

Proof: From Theorem 5.3, we obtain that the sequence $\{x_k\}$ converges to $\bar{x}$ and $\bar{x}$ is a stationary point of (SOP).

To prove superlinear convergence, note that each f^{a_j} is twice continuously differentiable. Therefore, for any $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that

$$\|\nabla^2 f^{a_j}(x) - \nabla^2 f^{a_j}(\bar{x})\| < \epsilon \text{ for all } x \in \mathcal{B}(\bar{x}, \delta_\epsilon) \subseteq V.$$

For $x_k \in V$, let $w_k = w(x_k)$. For any $\lambda \in \Delta_k = \{(\lambda_1, \lambda_2, \ldots, \lambda_{w_k}) \in \mathbb{R}^{w_k}_+ : \sum_{i=1}^{w_k} \lambda_i = 1\}$, we define a function $\Theta_\lambda : U \times \mathbb{R}^n \to \mathbb{R}^m$ by

$$\Theta_\lambda(x_k, u) = \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a^k_j}(x_k)^\top u + \tfrac{1}{2} \sum_{j=1}^{[w_k]} \lambda_j u^\top \nabla^2 f^{a^k_j}(x_k) u.$$

Note that for any $a^k \in P_{x_k}$ and $j \in [w(x_k)]$, each $f^{a^k_j}$ is twice continuously differentiable and strongly convex function. Moreover, the set P_{x_k} is finite. Therefore, the function $\Theta_\lambda(a, \cdot)$ is strongly convex in $\mathbb{R}^n$ and hence the function $\Theta_\lambda(a, \cdot)$ attains its minimum. Then, using Danskin's theorem (see Proposition 4.5.1, pp. 245–247 in [114]) and the first order necessary condition for a minimizer u_k of $\Theta_\lambda(x_k, \cdot)$, we have

$$\begin{aligned}
& \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a^k_j}(x_k) + \sum_{j=1}^{[w_k]} \lambda_j \nabla^2 f^{a^k_j}(x_k) u_k = 0 && (5.40)\\
\implies \quad & u_k = -\left[\sum_{j=1}^{[w_k]} \lambda_j \nabla^2 f^{a^k_j}(x_k)\right]^{-1} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a^k_j}(x_k) \\
\implies \quad & u_k \leq -\tfrac{1}{q} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a^k_j}(x_k) \text{ since } \nabla^2 f^{a^k_j}(x_k) \leq qI
\end{aligned}$$

$$\implies \qquad u_k \leq -\tfrac{1}{q}\max_{\lambda \in \Delta_k} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_k). \tag{5.41}$$

As the sequence $\{x_k\}$ converges to $\bar{x}$, there exists $k_\epsilon \in \mathbb{N}$ such that for all $k \geq k_\epsilon$, we have $x_k, x_k + u_k \in \mathcal{B}(\bar{x}, \delta_\epsilon)$. Now, using the second-order Taylor expansion at x_k of $f^{a_j^k}$, we have

$$f^{a_j^k}(x_k + u_k) \;\leq\; f^{a_j^k}(x_k) + \nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k + \tfrac{\epsilon}{2}\|u_k\|^2.$$

As $\max\{b_1, b_2, \ldots, b_{w_k}\} = \max\limits_{\lambda \in \Delta_{w_k}} \sum_{i=1}^{w_k} \lambda_i b_i$ holds, we get

$$\begin{aligned}
& f^{a_j^k}(x_k + u_k) - f^{a_j^k}(x_k) \\
\leq\; & \nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k + \tfrac{\epsilon}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{(q+\epsilon)}{2}\|u_k\|^2 \text{ since } \nabla^2 f^{a_j^k}(x) \leq qI \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\max_{j \in [w(x_k)]}\{\nabla f^{a_j}(x_k)^\top u_k\} + \tfrac{(q+\epsilon)}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\max_{\lambda \in \Delta_k}\left\{\sum_{j=1}^{[w(x_k)]} \lambda_j \nabla f^{a_j^k}(x_k)^\top u_k\right\} + \tfrac{(q+\epsilon)}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k - q(1-\beta)\|u_k\|^2 + \tfrac{(q+\epsilon)}{2}\|u_k\|^2 \text{ by (5.41)} \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + \frac{\epsilon - q(1-2\beta)}{2}\|u_k\|^2,
\end{aligned}$$

where from assumption (iii), we conclude that $\epsilon - q(1-2\beta) \leq 0$, and hence $t_k = 1$ holds. Now, for $k \geq k_\epsilon$, $\lambda \in \Delta_k$, and $j \in [w(x_k)]$, we have

$$\begin{aligned}
& \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_{k+1}) \right\| \\
=\; & \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k + u_k) \right\|
\end{aligned}$$

$$\overset{(5.40)}{=} \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k + u_k) - \left[\sum_{j=1}^{[w(x_k)]} \lambda_j \nabla f^{a_j^k}(x_k) + \sum_{j=1}^{[w(x_k)]} \lambda_j \nabla^2 f^{a_j^k}(x_k)^\top u_k \right] \right\|$$
$$\leq \epsilon \|u_k\| \text{ by (5.38) of Lemma 5.4.} \tag{5.42}$$

Combining assumption (i) and boundedness of $\{u_{k+1}\}$ (Theorem 5.2), we observe that

$$\tfrac{1}{2} u_k^\top \nabla^2 f^{a_j^k}(x_k) u_k \leq \tfrac{q}{2} \|u_k\|^2 \text{ for any } j \in [w(x_k)]. \tag{5.43}$$

Therefore, incorporating the above relation in (5.21), we get

$$\|u_{k+1}\| \leq \tfrac{2L}{q} \max_{j \in [w(x_k)]} \left\| \nabla f^{a_j^k}(x_{k+1}) \right\| \leq \tfrac{2L}{q} \max_{\lambda \in \Delta_k} \left\| \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_{k+1}) \right\| \overset{(5.42)}{\leq} \tfrac{2L\epsilon}{q} \|u_k\|.$$

In view of the above relation, we have

$$\|x^{k+1} - x^{k+2}\| = \|u^{k+1}\| \leq \tfrac{2L\epsilon}{q} \|u_k\| = \tfrac{2L\epsilon}{q} \|x^k - x^{k+1}\|,$$

and for any $k \geq 1$ and $m \geq 1$, we obtain

$$\begin{aligned} \|x^{k+m} - x^{k+m+1}\| &\leq \left(\tfrac{2L\epsilon}{q}\right) \|x^{k+m-1} - x^{k+m}\| \\ &\leq \left(\tfrac{2L\epsilon}{q}\right)^2 \|x^{k+m-1} - x^{k+m}\| \\ &\leq \cdots \leq \left(\tfrac{2L\epsilon}{q}\right)^m \|x^k - x^{k+1}\|. \end{aligned} \tag{5.44}$$

Now, we assume $0 < \tau < 1$ and define

$$\bar{\epsilon} = \min \left\{ q(1 - 2\beta), \tfrac{\tau}{1+2\tau} \left(\tfrac{q}{2L\epsilon}\right) \right\}.$$

If we take $\epsilon < \bar{\epsilon}$ and $k \geq k_\epsilon$, then by convergence of sequence $\{x_k\}$ and relation (5.44), we have

$$\|\bar{x} - x^{k+1}\| \leq \sum_{m=1}^{\infty} \|x^{k+m} - x^{k+m+1}\| \leq \sum_{m=1}^{\infty} \left(\tfrac{\tau}{1+2\tau}\right)^m \|x^k - x^{k+1}\|$$
$$= \tfrac{\tau}{1+\tau}\|x^k - x^{k+1}\|.$$

Therefore, we obtain

$$\|\bar{x} - x^k\| \geq \|x^k - x^{k+1}\| - \|x^{k+1} - \bar{x}\| \geq \tfrac{1}{1+\tau}\|x^k - x^{k+1}\|.$$

Hence, we can conclude that if $\epsilon < \bar{\epsilon}$ and $k \geq k_\epsilon$, then $\frac{\|\bar{x}-x^{k+1}\|}{\|\bar{x}-x^k\|} \leq \tau$. □

In the conventional Newton method, it is well known that once the initial point is chosen from the vicinity of the optimal solution, then the entire sequence of iterates resides in the same vicinity of the optimal solution and converges to it. The next result shows a similar observation in the proposed Newton method for (SOP).

Proposition 5.5 *Let $\bar{x}$ be a stationary point of F. Then, there exists $\rho > 0$ and a nonempty convex set $V \subseteq \mathbb{R}^n$ such that all the assumptions in Theorem 5.4 are satisfied for any chosen initial point $x_0 \in \mathcal{B}(\bar{x}, \rho)$. Moreover, if the initial point x_0 belongs to a compact level set of F and $\{x_k\}$ is the sequence generated from x_0, then the sequence $\{x_k\}$ converges to some stationary point $\widetilde{x}$ of (SOP).*

Proof: Given that $\bar{x}$ is a stationary point of F. Take $R > 0$ such that $\mathcal{B}(\bar{x}, R) \subseteq V$. Then, there exist $p > 0$ and $q > 0$ such that $pI \leq \nabla^2 f^{a_j}(x) \leq qI$ for all $x \in V$ and $j \in [w(x)]$, which is the assumption (i) of Theorem 5.4.
Now, take $\epsilon > 0$ such that $\epsilon \leq q(1 - 2\beta)$. Then, there exists $\delta > 0$ such that for all $x, y \in V$ and $j \in [w(\bar{x})]$,

$$\|\nabla^2 f^{a_j}(x) - \nabla^2 f^{a_j}(y)\| < \epsilon \text{ with } \|x - y\| < \delta,$$

which is the assumption (ii) of Theorem 5.4. Note that $\bar{x}$ is a stationary point of F. Then, in view of Theorem 5.1, we have $\Phi(\bar{x}) = 0$.

From Theorem 5.2, the function Φ is continuous. Thus, there exists $\rho \in (0, \frac{R}{2})$ such that for any $x \in \mathcal{B}(\bar{x}, \rho)$, we have

$$|\Phi(x) - \Phi(\bar{x})| \leq \tfrac{p}{2}\left[\min\left\{\delta, \left(\tfrac{R}{2}\right)\left(1 - \tfrac{\epsilon}{p}\right)\right\}\right]^2.$$

Given that $\{x_k\}$ is the sequence generated from x_0. Then, in view of Proposition 5.4, the sequence $\{F(x_k)\}$ is set-less decreasing. Moreover, the sequence $\{x_k\}$ is bounded and f^{a_j}'s are twice continuously differentiable. Therefore, the sequence $\{f^{a_j}(x_k)\}$ is also bounded. Thus, in view of Proposition 5.4 and Step *Step* 6 of Algorithm 1, we get

$$\lim_{k\to\infty} t_k \Phi(x_k) = 0.$$

Hence, we can observe that there exists a subsequence $\{x_{k_j}\}$ of $\{x_k\}$ such that $x_{k_j} \to \widetilde{x}$, for some stationary point $\widetilde{x} \in V$. If this claim is not true, i.e., $\widetilde{x}$ is not stationary, then in view of Proposition 5.4 and Step *Step* 6 of Algorithm 1, we have

$$\liminf_{j\to\infty} t_{k_j} > 0 \quad \text{or} \quad \lim_{j\to\infty} \Phi(x_{k_j}) = 0.$$

Therefore, by Proposition 5.2, we have $\Phi(\widetilde{x}) = 0$, which is contradictory to the result in Theorem 5.1. Thus, the point $\widetilde{x}$ is stationary. Additionally, in view of Proposition 5.5, we conclude that for j large enough, x_{k_j} belongs to the close neighborhood of and converges to stationary point $\widetilde{x}$. □

Next, we analyze the quadratic convergence of the proposed Algorithm 1.

Theorem 5.5 (Quadratic convergence). *Let the initial point x_0 belongs to a compact level set of F, $\{x_k\}$ be the corresponding sequence generated by Algorithm 1 and there exist $p > 0, q > 0$ for which $pI \leq \nabla^2 f^{a_j}(x) \leq qI$ for all $j \in [w(x)]$. Moreover, assume* *that for all $j \in [w(x)]$, $\nabla^2 f^{a_j}$ is Lipschitz continuous with a common Lipschitz constant $\widetilde{L}$. Then, the sequence $\{x_k\}$ converges quadratically to a stationary point of* (SOP).

Proof: From Theorem 5.3 and 5.4, we observe that $\bar{x}$ is a stationary point of (SOP) and $t_k = 1$ for large enough k. Since f^{a_j}'s are twice continuously differentiable, for any $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that for all $x, y \in \mathcal{B}(\bar{x}, \delta_\epsilon)$, we have

$$\|\nabla^2 f^{a_j}(x) - \nabla^2 f^{a_j}(y)\| < \epsilon. \tag{5.45}$$

Now proceeding in similar steps as in Theorem 5.4 and in view of (iii) of Lemma 5.4, we conclude that for $k \geq k_\epsilon$, we have

$$\begin{aligned} & \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_{k+1}) \right\| \\ = \quad & \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k + u_k) \right\| \\ \overset{(5.40)}{=} \quad & \left\| \sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k + u_k) - \left[\sum_{j=1}^{[w(x_k)]} \lambda_j \nabla f^{a_j^k}(x_k) + \sum_{j=1}^{[w(x_k)]} \lambda_j \nabla^2 f^{a_j^k}(x_k)^\top u_k \right] \right\| \\ \leq \quad & \tfrac{\widetilde{L}}{2}\|u_k\|^2 \text{ from (iii) of Lemma 5.4,} \end{aligned} \tag{5.46}$$

Note that $\max\{b_1, b_2, \ldots, b_{w(x)}\} = \max\limits_{\lambda \in \Delta_{w(x)}} \sum_{i=1}^{w(x)} \lambda_i b_i$ holds, where

$$\Delta_{w(x)} = \{(\lambda_1, \lambda_2, \ldots, \lambda_{w(x)}) \in \mathbb{R}_+^{w(x)} : \sum_{i=1}^{w(x)} \lambda_i = 1\}.$$

On combining the relation (5.46) with assumption (i) and boundedness of $\{u_{k+1}\}$ (Theorem 5.2), we get

$$\|u_{k+1}\| \leq \tfrac{2L}{q} \max_{j \in [w(x_{k+1})]} \left\| \nabla f^{a_{k,j}}(x_{k+1}) \right\|$$

$$\text{or, } \|u_{k+1}\| \leq \tfrac{2L}{q} \left\{ \max_{\lambda \in \Delta_k} \left\| \sum_{j=1}^{[w(x_k)]} \nabla f^{a_j^k}(x_{k+1}) \right\| \right\}$$

$$\text{or, } \|u_{k+1}\| \leq \tfrac{2L\widetilde{L}}{q} \|u_k\|^2 \text{ from (5.46).} \tag{5.47}$$

Since from Theorem 5.4 the sequence $\{x_k\}$ converges superlinearly to $\bar{x}$ of (SOP), then in view of (5.46) and (5.47) there exist N such that for $k \geq N$, we have

$$\|\bar{x} - x^{k+1}\| \leq \tau \|\bar{x} - x^k\| \text{ for some } \tau \in (0,1). \tag{5.48}$$

Further, triangle inequality, for $k = l \geq N$, we obtain

$$\|x^l - x^{l+1}\| \leq \|x^l - \bar{x}\| + \|\bar{x} - x^{l+1}\| \overset{(5.48)}{\leq} (1+\tau)\|\bar{x} - x^l\| \tag{5.49}$$

$$\text{and} \quad \|x^l - x^{l+1}\| \geq \|x^l - \bar{x}\| - \|\bar{x} - x^{l+1}\| \overset{(5.48)}{\geq} (1-\tau)\|\bar{x} - x^l\|. \tag{5.50}$$

Finally, from the first inequality of relation (5.49), we conclude that

$$\begin{aligned} (1-\tau)\|\bar{x} - x^{k+1}\| \leq \|x^{k+1} - x^{k+2}\| &= \|u_{k+1}\| \\ &\overset{(5.47)}{\leq} \tfrac{2L\widetilde{L}}{q}\|u_k\|^2 \\ &= \tfrac{2L\widetilde{L}}{q}\|x^k - x^{k+1}\|^2 \\ &\overset{(5.49)}{=} \tfrac{2L\widetilde{L}}{q}(1+\tau)^2\|x^k - \bar{x}\|^2, \end{aligned}$$

which proves the quadratic convergence of $\{x_k\}$ to $\bar{x}$. □

5.5 Numerical Demonstrations and Execution of Results

In this section, we implement the proposed Algorithm 1 on some numerical experiments. Algorithm 1 and its experimentation were performed in MATLAB R2023b software. This MATLAB software is installed in an IOS machine equipped with a 12-core CPU and 8 GB RAM. In the numerical implementation of the algorithm, we have considered the following:

- We take the cone K to be a standard ordering cone, that is, $K = \mathbb{R}^2_+$ for all test instances except for Examples 5.6 and 5.7. For the scalarizing function Ψ_e, we take $e = (1, 1, \ldots, 1)^\top \in \text{int}(K)$.
- The parameters β and ν in Step *Step* 6 for the line search of the Algorithm 1 is chosen as $\beta = 0.5$ and $\nu = 0.54$.
- The employed stopping criterion is $\|u_k\| < 0.001$ or a maximum number of 100 iterations is reached.
- To find the set $\text{Min}(F(x_k), K)$ at the k-th iteration in Step *Step* 3 of Algorithm 1, we adopted the crude method (pair-wise comparing) of comparing the elements in $F(x_k)$.
- At the k-th iteration in Step *Step* 4 of Algorithm 1, we find an

$$(a^k, u_k) \in \underset{(a,u)\in P_k\times\mathbb{R}^n}{\text{argmin}} \xi_{x_k}(a, u)$$

  with the help of an inbuilt function *fminsearch* in MATLAB.
- We have considered some test problems from the literature subjected to slight modifications, and some are freshly introduced. For each test considered, we generated 100 initial points randomly and ran the algorithm for each of the opted initial points. In the context of each experiment, we have presented a table with four columns:
  - **Number of initial points:** This value in the first column is the number of initial points taken to execute the proposed Algorithm 1.

- **Algorithm:** For the proposed algorithm, we use the abbreviation NM (Newton method), and for the existing steepest descent method [1], we use the abbreviation SD.
- **Iterations:** This value presents the third column with a 6-tuple (Min, Max, Mean, Median, Mode, SD) whose components are the minimum, maximum, mean, median, mode, and standard deviation of the number of iterations until the stopping condition is met.
- **CPU time:** This value indicates the third column, which is again a 6-tuple (Min, Max, Mean, Median, $\lceil$Mode$\rceil$, SD) that shows the minimum, maximum, mean, median, least integer greater or equal to mode, and standard deviation of the CPU time (in seconds) taken by the initial point in reaching the stopping condition.

Additionally, the numerical values are presented with precision up to four decimal places to ensure clarity. For every examined problem, the values of F at each iteration for the initial and final points are marked with black and red colors, respectively. We use shapes $\bullet$, $\star$, and $\blacktriangle$ to depict the values F for different initial points. Cyan, magenta, and green colors are used to represent the intermediate iterates for different initial points. Initial points are depicted in black, and the termination point is in red. If the initial point is depicted by black *bullet* $\bullet$, then the terminating is depicted by the red *bullet* $\bullet$, and the intermediate iterates by cyan *bullets* $\bullet$ or magenta *bullets* $\bullet$ or green *bullets* $\bullet$. That is, we use the same shape for depicting a complete sequence of iterates generated by Algorithm 1.

Furthermore, we compare the results of the proposed Newton's method (abbreviated as NM) with the existing steepest descent method (abbreviated as SD) for set optimization presented in [1].

Now, we discuss the different test problems on which the proposed algorithm is tested. The first is taken from [84] with some modifications.

Example 5.1 *Consider the function $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ defined as*

$$F(x) = \{f^1(x), f^2(x) \ldots, f^{20}(x)\},$$

where for each $i \in [20]$, $f^i : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$f^i(x) = \begin{pmatrix} x_1^2 + 0.5 \sin\left(\frac{2\pi(i-1)}{20}\right) + x_2^2 \\ 2x_1^2 + 0.5 \cos\left(\frac{2\pi(i-1)}{20}\right) + 2x_2^2 \end{pmatrix}.$$

Figure 5.1 shows the behaviour of Algorithm 1 for different initial points within the set $[-4, 4] \times [-4, 4]$. Firstly, Figure 5.1(a) depicts the sequence of iterates $\{F(x_k)\}$ generated by Algorithm 1 for the chosen initial point $x_0 = (2.5102, 0.0000)^\top$. Subsequently, the sequence of iterates $\{x_k\}$ corresponding to $\{F(x_k)\}$ is illustrated in Figure 5.1(b).

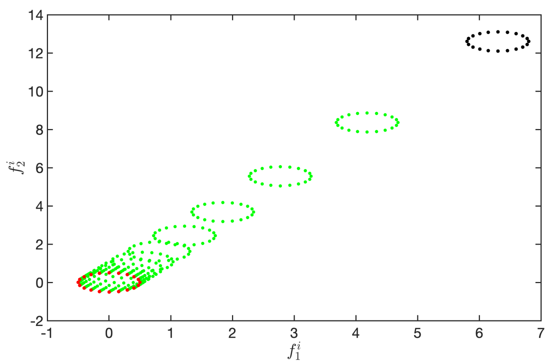
Further, in Figure 5.1(c), the sequence of iterates $\{F(x_k)\}$ generated by Algorithm 1 for three randomly selected initial points are depicted with cyan, magenta, and green colors. Additionally, the sequence of iterates $\{x_k\}$ corresponding to $\{F(x_k)\}$ generated by Algorithm 1 are shown in Figure 5.1(d).

Next, we have discussed the performance of Algorithm 1 for Example 5.1 in Table 5.1. Further, in Table 5.1, we have compared the results of NM with SD for set optimization. The values in Table 5.1 show that the proposed method performs better than the existing SD method.

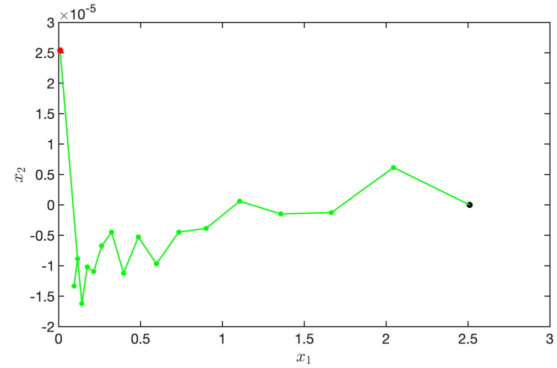
Table 5.1: Performance of Algorithm 1 on Example 5.1

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈ Mode ⌉, SD)
100	NM ($t_k = 1$)	(2, 2, 2, 2, 2, 0)	(2.0521, 3.1308, 2.1637, 2.1031, 2, 0.1601)
	NM	(1, 14, 10.6800, 11, 12, 2.6963)	(1.0345, 25.6842, 18.9496, 20.4787, 1, 4.9776)
	SD	(1, 23, 11.4800, 11, 10, 5.7375)	(1.0182, 41.4550, 19.9548, ,19.5869, 1, 10.6772)

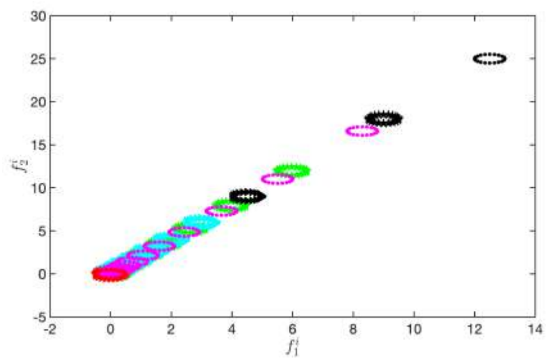
The next example, discussed below, is freshly introduced.



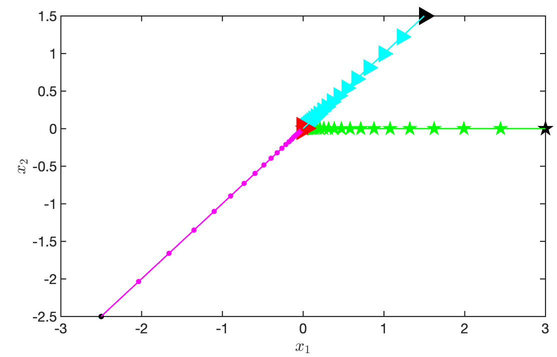
(a) The value of F at each iteration generated by Algorithm 1 for initial point $x_0 = (2.5102, 0.0000)^\top$ for Example 5.1



(b) The value of x_k at each iteration generated by Algorithm 1 for initial point $x_0 = (2.5102, 0.0000)^\top$ for Example 5.1



(c) The value of F at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.1



(d) The value of x_k at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.1

Figure 5.1: Obtained output of Algorithm 1 for Example 5.1

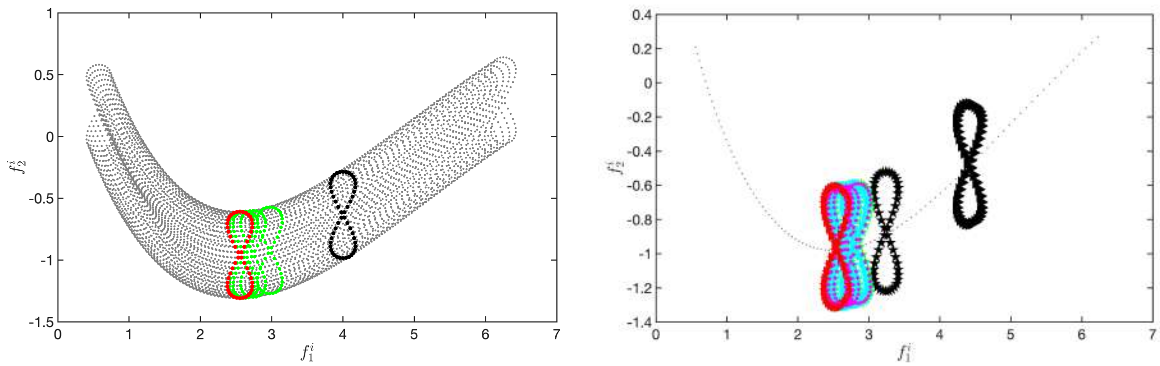
Example 5.2 *Consider the set-valued function $F : \mathbb{R} \rightrightarrows \mathbb{R}^2$ defined by*

$$F(x) = \{f^1(x), f^2(x), \dots, f^{50}(x)\},$$

where for each $i \in [50]$, the function $f^i : \mathbb{R} \to \mathbb{R}^2$ is given by

$$f^i(x) = \begin{pmatrix} 0.35 \sin\left(\frac{2\pi(i-1))}{50}\right) \cos\left(\frac{2\pi(i-1))}{50}\right) + x^2 \\ 0.35 \cos\left(\frac{2\pi(i-1))}{50}\right) + \frac{1}{(1+e^{2x})} + \cos(2x) \end{pmatrix}.$$

The output of Algorithm 1 for different initial points of Example 5.2 are depicted in



(a) The value of F at each iteration generated by Algorithm 1 for Example 5.2 for the initial point $x_0 = 2.0000$

(b) The value of F for three different initial points at each iteration generated by Algorithm 1 of Example 5.2

Figure 5.2: Obtained output of Algorithm 1 for Example 5.2

Figure 5.2. The discretized ∞-shaped segments represent the objective values that transverse a curve within the interval $[0.7700, 6.3000]$. *Figure 5.2(a) depicts the sequence* $\{F(x_k)\}$ *generated by Algorithm 1 for the starting point* $x_0 = 2.0000$. *In Figure 5.2(b), we exhibit the output sequence* $\{F(x_k)\}$ *generated by Algorithm 1 for three randomly chosen initial points.*

The performance of Algorithm 1 for Example 5.2 is shown in Table 5.2. Moreover, we have compared the results of the NM with the SD method for set optimization as presented in Table 5.2. The values in Table 5.2 show that the proposed method performs better than the existing SD method.

Table 5.2: Performance of Algorithm 1 on Example 5.2

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM ($t_k = 1$)	(1, 3, 1.9900, 2, 2, 0.3013)	(0.2215, 1.6828, 0.8634, 0.8416, 0, 0.1625)
	NM	(1, 5, 2.3600, 1, 1, 1.8175)	(0.3823, 3.9151, 1.6241, 0.9135, 0, 0.9606)
	SD	(1, 8, 6.0900, 7, 8, 2.4622)	(0.4170, 5.4045, 3.7906, 4.2701, 0, 1.3692)

Example 5.3 *Consider the function* $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^3$ *defined as*

$$F(x) = \{f^1(x), f^2(x), \dots, f^{14}(x)\},$$

where for each $i \in [14]$, $f^i : \mathbb{R}^2 \to \mathbb{R}^3$ is given by

$$f^i(x) = \begin{pmatrix} x_1^2 + x_2^2 + 0.25 \sin\left(\frac{2\pi(i-1)}{14}\right) \\ 4x_1^2 + 4x_2^2 + 0.25 \cos\left(\frac{2\pi(i-1)}{14}\right) \\ x_1^2 + x_2^2 + i \end{pmatrix}.$$

Figure 3.3 shows the behaviour of Algorithm 1 for different initial points within the set $[-3, 4]\times[-3, 4]$. Firstly, Figure 5.3(a) depicts the sequence of iterates $\{F(x_k)\}$ generated by Algorithm 1 for the chosen initial point $x_0 = (3.2302, -0.5102)^\top$. Subsequently, the sequence of iterates $\{x_k\}$ corresponding to $\{f(x_k)\}$ is illustrated in Figure 5.3(b).

Further, in Figure 5.3(c), the sequence of iterates $\{F(x_k)\}$ generated by Algorithm 1 for three randomly selected initial points are depicted with cyan, magenta, and green colors. Additionally, the sequence of iterates $\{x_k\}$ corresponding to $\{F(x_k)\}$ generated by Algorithm 1 are shown in Figure 5.3(d).

The performance of Algorithm 1 on Example 5.3 is shown in Table 5.3. The comparison of the results of NM with the SD method for set optimization is given in Table 5.3. The values in Table 5.3 show that the proposed method performs better than the existing SD method.

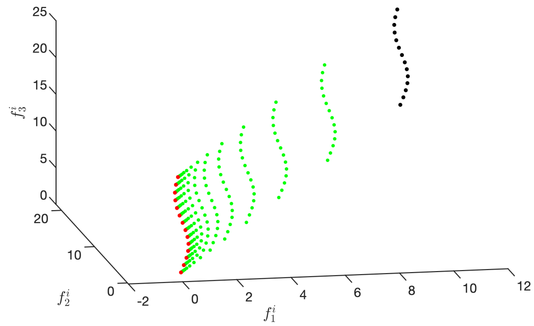
Table 5.3: Performance of Algorithm 1 on Example 5.3

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM	(1, 14, 10.8800, 12, 13, 2.8508)	(0.1190, 69.3281, 51.3830, 56.2429, 3, 13.9629)
	SD	(1, 14, 11.0500, 12, 13, 2.2128)	(0.1202, 65.2457, 50.2756, 54.3065, 21, 10.3982)

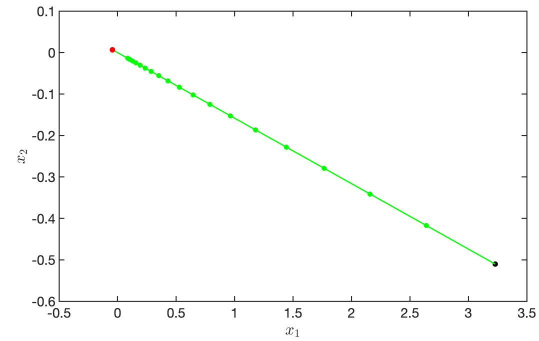
Below, we consider Examples 5.3 and 5.4, again motivated from [84].

Example 5.4 *Consider the function $F : \mathbb{R} \rightrightarrows \mathbb{R}^3$ defined by*

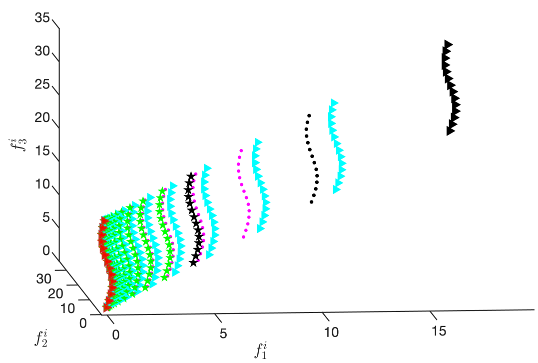
$$F(x) = \{f^1(x), f^2(x), \ldots, f^{30}(x)\},$$



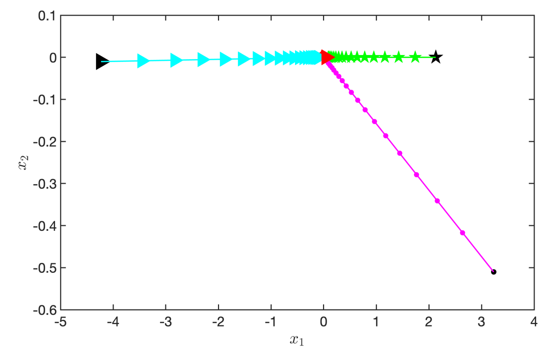
(a) The value of F at each iteration generated by Algorithm 1 for initial point $x_0 = (3.2302, -0.5102)^\top$ for Example 5.3



(b) The value of x_k at each iteration generated by Algorithm 1 for initial point $x_0 = (3.2302, -0.5102)^\top$ for Example 5.3



(c) The value of F at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.3



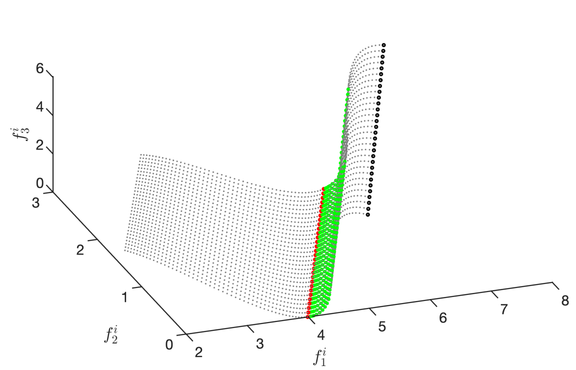
(d) The value of x_k at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.3

Figure 5.3: Obtained output of Algorithm 1 for Example 5.3

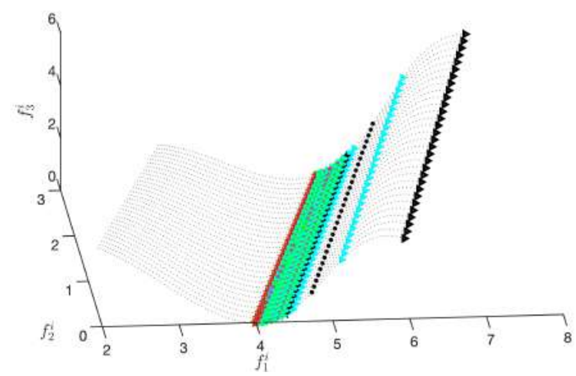
where for each $i \in [30]$, $f^i : \mathbb{R} \to \mathbb{R}^3$ *is given by*

$$f^i(x) = \begin{pmatrix} x^2 + \frac{(i-1)}{30} \\ (x^2-4)(\sin(x^2-4)) + \frac{(i-1)}{30} \\ \frac{(i-1)}{30}\ x^2 \end{pmatrix}.$$

The output of Algorithm 1 for Example 5.4 is depicted in Figure 5.4. The collection of the objective values for all $x \in [1.54, 2.16]$ *transverses a surface depicted by the dot-shaded region. Figure 5.4(a) depicts the sequence* $\{F(x_k)\}$ *generated by Algorithm 1 for the starting point* $x_0 = 2.1300$*. We test Algorithm 1 for three different randomly chosen initial points and observe their corresponding output* $\{F(x_k)\}$ *(marked with cyan,*



(a) The value of F at each iteration generated by Algorithm 1 for initial point $x_0 = 2.1300$ for Example 5.4



(b) The value of F at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.4

Figure 5.4: Obtained output of Algorithm 1 for Example 5.4

magenta, and green colors) as shown in Figure 5.4(b). The performance of Algorithm 1 for Example 5.4 is shown in Table 5.4.

Table 5.4: Performance of Algorithm 1 on Example 5.4

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM	(0, 19, 7.4200, 9, 1, 5.1231)	(0.0018, 267.5772, 125.0167, 129.7198, 0, 81.4232)

For the initial point $x_0 = 2.1300$*, the values of* F *across the generated iterates for NM and SD methods are exhibited in Table 5.5 and 5.6, respectively. We have observed that the SD method presented in [1] is not properly solving this problem due to the objective function's highly nonlinear nature and converging to the given initial point* x_0 *in zero number of iterations. On the other hand, the proposed NM method gives the minimal solution* $\bar{x} = 1.9957$ *for the chosen initial point* $x_0 = 2.1300$*. Therefore, we can conclude that the proposed NM Algorithm 1 is more efficient than the SD method for set optimization given in [1].*

In the next example, we consider the robust counterpart of a vector-valued facility location problem under uncertainty [115]. A detailed discussion on this problem is given in [1].

Table 5.5: Performance of Algorithm 1 on Example 5.4 with $t_k \in (0, 1)$

Iteration number (k)	x_k	$f^{10}(x_k)$	$f^{20}(x_k)$	$f^{30}(x_k)$
0	2.1300	(4.8369, 0.5746, 1.3611)	(5.1702, 0.9079, 2.8734)	(5.5036, 1.2413, 4.3857)
1	2.1115	(4.7584, 0.5029, 1.3375)	(5.0918, 0.8362, 2.8237)	(5.4251, 1.1695, 4.3098)
2	2.0929	(4.6802, 0.4411, 1.3141)	(5.0136, 0.7744, 2.7741)	(5.3469, 1.1078, 4.2342)
3	2.0744	(4.6031, 0.3905, 1.2909)	(4.9365, 0.7238, 2.7253)	(5.2698, 1.0572, 4.1597)
4	2.0559	(4.5267, 0.3510, 1.2680)	(4.8601, 0.6843, 2.6769)	(5.1934, 1.0176, 4.0858)
5	2.0373	(4.4506, 0.3226, 1.2452)	(4.7839, 0.6559, 2.6287)	(5.1173, 0.9893, 4.0122)
6	2.0244	(4.3982, 0.3096, 1.2295)	(4.7315, 0.6430, 2.5955)	(5.0649, 0.9763, 3.9616)
7	2.0179	(4.3719, 0.3052, 1.2216)	(4.7053, 0.6385, 2.5789)	(5.0386, 0.9718, 3.9362)
8	2.0114	(4.3457, 0.3021, 1.2137)	(4.6791, 0.6354, 2.5623)	(5.0124, 0.9688, 3.9109)
9	1.9957	(4.2828, 0.3003, 1.1948)	(4.6162, 0.6336, 2.5225)	(4.9495, 0.9670, 3.8501)

Table 5.6: Performance of SD method (see [1]) on Example 5.4 with $t_k \in (0, 1)$

Iteration number (k)	x_k	$f^{10}(x_k)$	$f^{20}(x_k)$	$f^{30}(x_k)$
0	2.1300	(4.8369, 0.5746, 1.3611)	(5.1702, 0.9079, 2.8734)	(5.5036, 1.2413, 4.3857)
1	2.1300	(4.8369, 0.5746, 1.3611)	(5.1702, 0.9079, 2.8734)	(5.5036, 1.2413, 4.3857)

Example 5.5 *Consider the function $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^3$ defined as*

$$F(x) = \{f^1(x), f^2(x), \ldots, f^{100}(x)\},$$

where for each $i \in [100]$, $f^i : \mathbb{R}^2 \to \mathbb{R}^3$ is given as

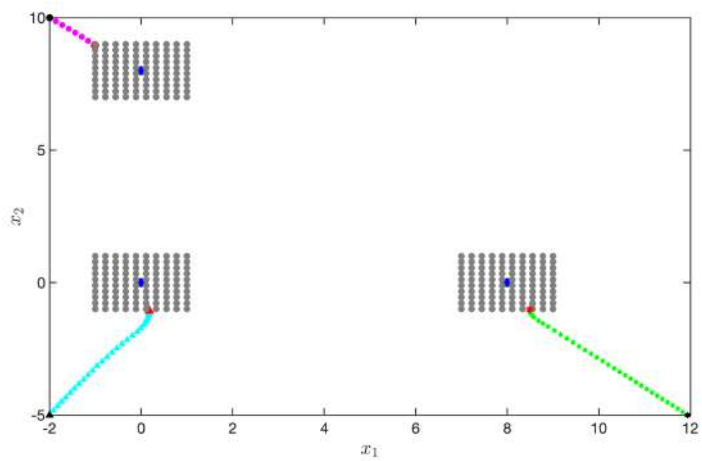
$$f^i(x) = \tfrac{1}{2}\begin{pmatrix} \|x - l_1 - u_i\|^2 \\ \|x - l_2 - u_i\|^2 \\ \|x - l_3 - u_i\|^2 \end{pmatrix},$$

where $l_1 = \begin{pmatrix} 0 \\ 8 \end{pmatrix}$, $l_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $l_3 = \begin{pmatrix} 8 \\ 0 \end{pmatrix}$. We consider a uniform partition set of 10 *points of the interval* $[-1, 1]$ *given by*

$$\mathcal{U} = \left\{-1, -1 + \tfrac{1}{s}, -1 + \tfrac{2}{s}, \ldots, -1 + \tfrac{2(s-1)}{s}, 1\right\} \text{ with } s = 4.5.$$

The set $\{u_i = (u_{1i}, u_{2i})^\top : i \in [100]\}$ *is an enumeration of the set* $\mathcal{U} \times \mathcal{U}$.

In Figure 5.5, the total of 100 *initial points were generated in the square* $[-50, 50] \times [-50, 50]$. *The grey points represent the set* $(l_1+u_i)\cup(l_2+u_i)\cup(l_3+u_i)$ *and the locations of* l_1, l_2, l_3 *are depicted in blue color. The values of* $F(x_k)$ *generated by Algorithm 1 for three different randomly chosen initial points are given with cyan, magenta, and green colors as shown in Fig. 5.5.*



(a) The value of F in argument space at each iteration generated by Algorithm 1 for three randomly chosen initial points for Example 5.6

Figure 5.5: Obtained output of Algorithm 1 for Example 5.5

Next, the performance of the proposed Algorithm 1 for the initial point $x_0 = (-2.0000, 10.0000)^\top$ *with* $t \in (0, 1)$ *and* $t_k = 1$ *is shown in Table 5.8 and 5.9. The decreasing behavior in the values of vector-valued functions at each iteration has been exhibited in Table 5.8.*

Table 5.7: Performance of Algorithm 1 on Example 5.5

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM ($t_k = 1$)	(3, 2.1000, 2 ,2, 0.3892)	(0.3677, 3.8689, 0.6442, 0.3840, 0, 0.8040)
	NM	(1, 14, 6.4700, 7, 1, 3.9503)	(0.2631, 17.114, 5.0776, 4.2074, 0, 3.9056)
	SD	(1, 65, 14.4200, 5.5000, 1, 16.6243)	(0.3871, 23.1876, 6.8118, 4.5792, 0, 5.8667)

In the next two examples (Example 5.6 and Example 5.7), we consider the ordering cone K different from $\mathbb{R}^m_+$ and observe the performance of the proposed Algorithm

Table 5.8: Performance of Algorithm 1 on Example 5.5 with $t \in (0, 1)$

Iteration number (k)	$x_k^\top$	$f^{25}(x_k)$	$f^{50}(x_k)$	$f^{75}(x_k)$	$f^{100}(x_k)$
0	$(-2, 10)$	$(0.6033, 0.7958, 0.8456)$	$(0.5820, 0.7793, 0.8399)$	$(0.6354, 0.7991, 0.8550)$	$(0.6295, 0.7841, 0.8502)$
1	$(-1.8614, 9.8614)$	$(0.5940, 0.7934, 0.8432)$	$(0.5715, 0.7765, 0.8374)$	$(0.6278, 0.7965, 0.8526)$	$(0.6224, 0.7812, 0.8478)$
2	$(-1.7210, 9.7210)$	$(0.5839, 0.7909, 0.8407)$	$(0.5601, 0.7736, 0.8348)$	$(0.6196, 0.7939, 0.8503)$	$(0.6150, 0.7782, 0.8454)$
3	$(-1.5794, 9.5782)$	$(0.5729, 0.7883, 0.8382)$	$(0.5478, 0.7707, 0.8322)$	$(0.6109, 0.7912, 0.8479)$	$(0.6073, 0.7752, 0.8430)$
4	$(-1.4381, 9.4317)$	$(0.5610, 0.7857, 0.8356)$	$(0.5345, 0.7677, 0.8296)$	$(0.6016, 0.7884, 0.8454)$	$(0.5994, 0.7720, 0.8404)$
5	$(-1.2971, 9.2813)$	$(0.5477, 0.7831, 0.8330)$	$(0.5202, 0.7647, 0.8269)$	$(0.5915, 0.7856, 0.8429)$	$(0.5914, 0.7688, 0.8379)$
6	$(-1.1556, 9.1258)$	$(0.5329, 0.7803, 0.8302)$	$(0.5051, 0.7615, 0.8240)$	$(0.5806, 0.7826, 0.8402)$	$(0.5832, 0.7654, 0.8352)$
7	$(-1.0149, 8.9666)$	$(0.5161, 0.7775, 0.8274)$	$(0.4901, 0.7582, 0.8212)$	$(0.5688, 0.7796, 0.8376)$	$(0.5752, 0.7619, 0.8325)$
8	$(-1.0029, 8.9639)$	$(0.5154, 0.7774, 0.8272)$	$(0.4885, 0.7581, 0.8210)$	$(0.5680, 0.7795, 0.8374)$	$(0.5744, 0.7618, 0.8323)$

Table 5.9: Performance of Algorithm 1 on Example 5.5 with $t_k = 1$

Iteration number (k)	$x_k^\top$	$f^{25}(x_k)$	$f^{50}(x_k)$	$f^{75}(x_k)$	$f^{100}(x_k)$
0	$(-5,\ -5)$	$(92.9382,\ 21.8270,\ 89.3822)$	$(109.9507,\ 29.9507,\ 101.0619)$	$(98.4942,\ 27.3830,\ 103.8278)$	$(116,\ 36,\ 116)$
1	$(-1,\ -1)$	$(39.6046,\ 0.4938,\ 36.0490)$	$(50.3946,\ 2.3950,\ 41.5062)$	$(40.7158,\ 1.6050,\ 46.0498)$	$(51.9996,\ 3.9999,\ 51.9999)$

1. The Example 5.6 is a slight modification of Test instance 5.1 discussed in [1] with respect to a cone other than $\mathbb{R}^m$.

Example 5.6 *Consider the function* $F : \mathbb{R} \rightrightarrows \mathbb{R}^2$ *defined by*

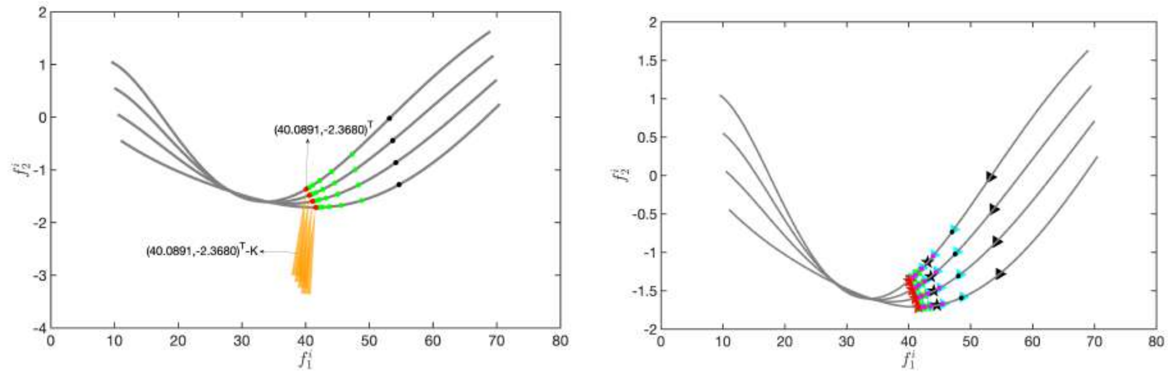
$$F(x) = \{f^1(x), f^2(x), f^3(x), f^4(x)\},$$

where for each $i \in [4], f^i : \mathbb{R} \to \mathbb{R}^2$ *is given as*

$$f^i(x) = \begin{pmatrix} 2x^2 + 4x + \frac{(i-3)}{2} \\ \frac{x}{2}\cos(x) - \frac{(i-3)}{2}\sin^2 x \end{pmatrix}.$$

The cone is K *given by* $K = \{(y_1, y_2)^\top \in \mathbb{R}^2 : 5y_1 - y_2 \geq 0, -9y_1 + 10y_2 \geq 0\}$.

Figure 2.3 shows the behaviour of Algorithm 1 for different initial points within the set $[2.3350, 4.4010]$. *The collection of objective function values for all* $x \in [2.3350, 4.4010]$ *transverse a surface as shown in Figure 2.3. Firstly, Figure 5.6(a) depicts the sequence of iterates* $\{F(x_k)\}$ *generated by Algorithm 1 for the selected initial point* $x_0 = 4.300$. *In Figure 5.6(a), the points depicted with red color collectively constitute a local weakly minimal point of* F*-values as the set* $(40.0891, -2.3680)^\top - K$ *does not contain any el-*



(a) The value of F at each iteration generated by Algorithm 1 for the initial point $x_0 = 4.3000$ for Example 5.6

(b) The value of F for three different initial points at each iteration generated by Algorithm 1 for Example 5.6

Figure 5.6: Obtained output of Algorithm 1 for Example 5.6

ement of the image set of F other than $(40.0891. -2.3680)$ for all $x \in [2.3350, 4.4010]$.

Subsequently, in Figure 5.6(b), we exhibit the output sequence $\{F(x_k)\}$ generated by Algorithm 1 for three randomly chosen initial points.

The performance of Algorithm 1 on Example 5.6 is shown in Table 5.10. Additionally, a comparison of the results of NM with the SD method for set optimization is presented in Table 5.10. The values in Table 5.10 show that the proposed method performs better than the existing SD method.

Table 5.10: Performance of Algorithm 1 on Example 5.6

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM	(1, 8, 7.0400, 7, 8, 0.8980)	(0.1778, 3.6665, 2.3651, 2.3388, 1, 0.3064)
	SD	(3, 10, 8.7800, 9, 9, 0.9813)	(1.0957, 3.8509, 2.6506, 2.6638, 2, 0.3428)

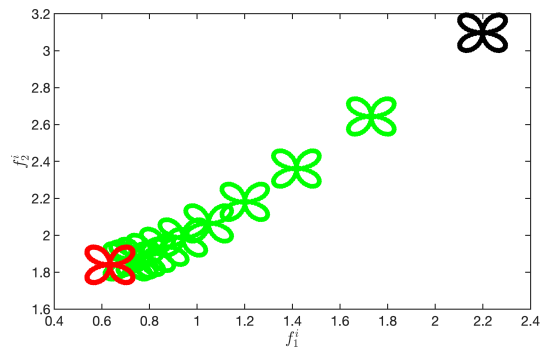
Example 5.7 *Consider the set-valued function $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ defined as*

$$F(x) = \{f^1(x), f^2(x) \dots, f^{100}(x)\},$$

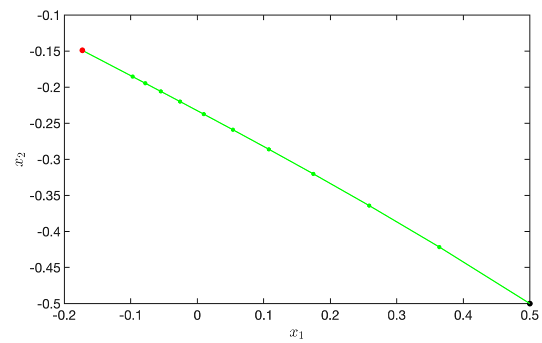
where for each $i \in [100]$, *the function* $f^i : \mathbb{R}^2 \to \mathbb{R}^2$ *is given as*

$$f^i(x) = \begin{pmatrix} x_1^2 + \sin(x_1) + x_1^2\cos(x_2) + 0.25\cos\left(\frac{2\pi(i-1)}{100}\right)\sin^2\left(\frac{2\pi(i-1)}{100}\right) + e^{x_1+x_2} + x_2^2 \\ 2x_1^2 + x_2^2\cos(x_1) + 0.25\cos\left(\frac{2\pi(i-1)}{100}\right)\sin^2\left(\frac{2\pi(i-1)}{100}\right) + \cos(x_2) + e^{x_1+x_2} + 2x_2^2 \end{pmatrix}.$$

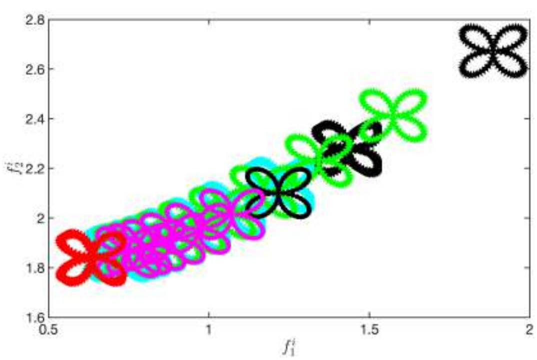
The cone K *is given by* $K = \{(z_1, z_2)^\top \in \mathbb{R}^2 : 2y_1 - 6y_2 \geq 0, -6y_1 + 7y_2 \geq 0\}$.



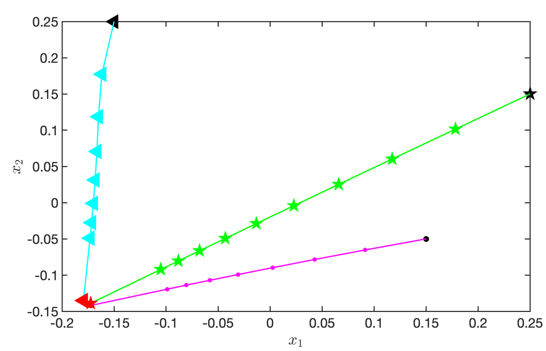
(a) The value of F at each iteration generated by Algorithm 1 for the initial point $x_0 = (0.5000, -0.5000)^\top$ for Example 5.7



(b) The value of x_k at each iteration generated by Algorithm 1 for for the initial point $x_0 = (0.5000, -0.5000)^\top$ for Example 5.7



(c) The value of F at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.7



(d) The value of x_k at each iteration generated by Algorithm 1 for three different randomly chosen initial points for Example 5.7

Figure 5.7: Obtained output of Algorithm 1 for Example 5.7

The output of Algorithm 1 on Example 5.7 is depicted in Figure 5.7. Figure 5.7(a) depicts the sequence $\{F(x_k)\}$ *generated by Algorithm 1 for the starting point* $x_0 = (0.5000, -0.5000)^\top$. *Figure 5.7(b) exhibits the sequence of iterates generated by Algorithm 1 for three different randomly chosen starting points.*

The performance of Algorithm 1 for Example 5.7 is shown in Table 5.11. A com-

parison of the results of NM with the SD method for set optimization is presented in Table 5.11. The values in Table 5.11 show that the proposed method performs better than the existing SD method.

Table 5.11: Performance of Algorithm 1 on Example 5.7

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	NM	(5, 12, 10.6600, 10, 10, 1.4718)	(47.8361, 109.4561, 80.7287, 71.5817, 38, 12.6247)
	SD	(7, 14, 10.5900, 10, 10, 1.8592)	(61.5775, 113.1363, 83.6014, 79.4974, 46, 15.1152)

5.6 Conclusion

In this chapter, we studied set optimization problems with respect to the lower set less relation, where the set-valued objective mapping is given by finitely many twice continuously differentiable strongly convex functions. We have proposed a Newton method (Algorithm 1) to generate a sequence of iterates that converges to a weakly minimal solution of the problem. To generate the sequence for a tactfully chosen initial point, at each iteration k, we choose an element a_k from the partition set $P_{x_k}$of the current iterate x_k, and then with the help of the concepts in [111, 112] we figured out the Newton direction u_k (*Step* 4) of the vector optimization problem (VOP) corresponding to a_k; this decent direction has been used to find the next iterate x_{k+1}. Algorithm 1 kept generating iterates until the stopping condition (*Step* 5) was met. The employed stopping condition is a necessary optimality condition (Proposition 5.1) of weakly minimal or stationary points of the considered problem (SOP).

The well-definedness and convergence analysis (Subsection 5.4.1) of the proposed Algorithm 1 for the derived Newton method has been discussed in detail. Towards certifying the well-definedness, we have ensured the existence of (a^k, u_k) in *Step* 4 and the existence of a step length t_k in *Step* 6 (Proposition 5.4). In the convergence analysis of Algorithm 1, we derived

(i) an equivalent condition of nonstationarity of a point (Proposition 5.1),

(ii) boundedness of the sequence of generated Newton direction (Proposition 5.3),

(iii) convergence of the generated sequence of iterates (Theorem 5.3) under a regularity condition (Definition 5.1),

(iv) superlinear convergence to a stationary point (Theorem 5.4) of the generated sequence under a regularity condition and uniform continuity of the Hessian matrices, and

(v) local quadratic convergence of the generated sequence under a regularity condition and Lipschitz continuity of the Hessian matrices (Theorem 5.5).

Finally, we tested the performance of the proposed Newton method on some existing and freshly introduced numerical test problems in section 4.5. It is found that the proposed Newton method outperforms the existing steepest descent method for strongly convex cases.

In this chapter, we have used the lower set less relation to compare images of the set-valued objective function of (SOP). Future research can be carried out by testing the proposed Algorithm 1 with other set relations such as certainly less order relation, possibly less order relation, min-max less order relation, and min-max certainly less order relation (see [15]). A comparison between these order relations is also given in [15]. While working with min-max less order relation and min-max certainly less order relation, one may note that usual derivative concepts, like epiderivatives or coderivatives, may not be suitable for optimality concepts. Thus, new derivative concepts need to be developed that align with these order relations.

As a future work, one can try to find the set of Minimal elements for the proposed work using the approaches discussed in [87, 116–119] and references therein. A comparison of their performance with different sorting functions can be observed. Moreover, in this chapter, we have used Gerstewitz's scalarizing function for ordering vectors. Further research can be performed on various other scalarizing functionals, such as separating functionals [117] with uniform level sets, Hiriart-Urruty functional [120], and Drummond-Svaiter functional [97]. Moreover, a comparison study can be made in the future, showing the performance profile of Newton's method with different scalarizing functionals.

Furthermore, one can try to extend the proposed Newton method to different step-size conditions, such as strong Wolfe or Armijo-Wolfe conditions, and a comparison between the performance of the methods with different step-size conditions can be observed. Global convergence of the proposed Newton's method can be observed by analyzing the work done in the articles [121–123].

****