Chapter 6

Quasi-Newton Method for Set Optimization Problems with Set-Valued Mapping Given by Finitely Many Vector-Valued Functions

6.1 Introduction

Optimization problems with set-valued objective functions or set-valued constraints have a number of applications in various areas of mathematical economics [13], finance [11], game theory, optimal control, and many others (see [70, 74]). The general expression of a set-valued optimization problem is given by

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

where F is a set-valued map from a nonempty subset X of $\mathbb{R}^n$ to $\mathbb{R}^m$ ordered by a convex cone $K \subset \mathbb{R}^m$. The solution set for the problem (A) provides an important generalization and unification for scalar as well as vector optimization problems using different approaches (see [21–23]).

6.2 Motivation and Contribution

Kuroiwa [23, 57, 78] was the first to define solutions of set optimization problems by preorder relations of sets. In [23], solution concepts have been defined based on the approach of comparing the sets that are values of the objective function. A detailed discussion on this is given in Chapter 5.

As the steepest descent method given by Bouza et al. in [1] is known to have linear convergence rate, we aim to derive a quasi-Newton method for set optimization problems. The set-valued objective that we consider here is defined by a finite number of twice continuously differentiable vector-valued functions. We use the ideas from [97, 111,124] for vector optimization problems. We have approximated the Hessian matrices with the help of BFGS approximation techniques. These problems have significant applications in uncertain optimization problems, as discussed in [81,125]. The proposed method in our work exhibits a superlinear convergence rate near the optimal solution and works well for highly nonlinear objective functions.

In this chapter, we discuss the interrelation of stationary points with weakly minimal solutions of (SOP). Next, we propose the quasi-Newton method for the considered set optimization problem. We define the well-definedness of the proposed algorithm with the existence of Armijo's step length condition and the boundedness of the norm of descent direction. After that, we analyze the convergence of the proposed quasi-Newton method. Further, we show the numerical implementation of our method with the help of suitable examples. Finally, we compare the results of the proposed algorithm with the results of the steepest descent method presented in [1].

6.3 Optimality Conditions for Set-Valued Mappings

In this chapter, we aim to derive a quasi-Newton method to identify weakly minimal solutions to 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 defined by (SOP).

First, we discuss results on optimality conditions for weakly minimal solutions of (SOP) under Assumption 1. These notions are the foundation for constructing the proposed quasi-Newton method to capture weakly minimal solutions of (SOP). The contribution of this section is divided into two subsections.

(i) To identify the sequence of iterates in the proposed quasi-Newton method, we figure out a family of vector optimization problems using the concept of partition set at a point. Thereafter, we discuss the concept of stationary point for (SOP) and interrelate stationary points with weakly minimal solutions using the defined family of vector optimization problems.

(ii) We derive a necessary optimality condition for weakly minimal points of (SOP). In the process of evaluating these points, we approximate the Hessian corresponding to each objective function $f^1, f^2, \ldots, f^p$ with the help of the BFGS method for vector optimization problems (see [124, 126, 127]).

6.3.1 Hessian Approximation and Necessary Condition for Weakly Minimal Solutions of (SOP)

To find a quadratic approximation of the functions $f^1, f^2, \ldots, f^p$, we use a positive definite approximation of their Hessian by the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) [128–131] methods.

Corresponding to a given initial point x_0, for the function f^i, $i \in [p]$, we generate a sequence of symmetric positive definite matrices $\{B^i(x_k)\}$ starting with an initial symmetric positive definite matrix $B^i(x_0)$. For each $i \in [p]$, the quadratic approximation of each f^i about x_k is given by

$$Q^i(x) = f^i(x_k) + \nabla f^i(x_k)^\top (x - x_k) + \tfrac{1}{2}(x - x_k)^\top B^i(x_k)(x - x_k).$$

We choose $\{B^i(x_k)\}$ that satisfies the quasi-Newton equation

$$B^i(x_{k+1})\ s_k = y_k^i,\ \ k = 0, 1, 2, \dots,$$

where $s_k = x_{k+1} - x_k$ and $y_k^i = \nabla f^i(x_{k+1}) - \nabla f^i(x_k)$. To ensure symmetric and positive definiteness of all the terms in the sequence $\{B^i(x_k)\}$, for a given symmetric and positive definite $B^i(x_0)$, we use the BFGS update formula at any point x_k by

$$B^i(x_{k+1}) = B^i(x_k) - \frac{B^i(x_k) s_k s_k^\top B^i(x_k)}{s_k^\top B^i(x_k) s_k} + \frac{y_k^i {y_k^i}^\top}{s_k^\top y_k^i}. \tag{6.1}$$

It can be observed from [132, Section 6.1] that for every $k \in \mathbb{N} \cup \{0\}$, if $B^i(x_k)$ is positive definite, then $B^i(x_{k+1})$ remains positive definite.

Remark 6.1 (See [132]). *The BFGS method satisfies the curvature condition $s_k^\top y_k^i > 0$. If each f^i, $i = 1, 2, \dots, p$, is strongly convex, then the curvature condition is satisfied by any two points x_k and x_{k+1}.*

Next, we derive a necessary condition for weakly minimal points of (SOP). We start with the following lemma.

Lemma 6.1 *If a point $\bar{x}$ is a local weakly minimal point of* (SOP)*, then $\bar{x}$ is a stationary point of* (SOP).

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

$$\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} \in -\text{int}(K) \text{ for all } j \in [\bar{w}]. \tag{6.2}$$

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

$$f^{\bar{a}_j}(x) = f^{\bar{a}_j}(\bar{x}) + \nabla f^{\bar{a}_j}(\bar{x})^\top (x - \bar{x}) + o(\|x - \bar{x}\|), \text{ where } \lim_{x \to \bar{x}} \frac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0. \tag{6.3}$$

Note that $\bar{x}$ is a local weakly minimal solution of (SOP). Therefore, by Theorem 1.5, $\bar{x}$ 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

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

From (6.3), there exists a neighborhood $B \subseteq U$ of $\bar{x}$ such that for all $j \in [\bar{w}]$,

$$\nabla f^{\bar{a}_j}(\bar{x})^\top (x - \bar{x}) \notin -\text{int}(K) \text{ for all } x \in B. \tag{6.4}$$

As B is a neighborhood of $\bar{x}$, there exists $\bar{t} > 0$ such that $x' = \bar{x} + \bar{t}\bar{u} \in B$. The relation (6.4) with $x = x'$ yields

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

which is contradictory to (6.2). Therefore, $\bar{x}$ is a stationary point for (SOP). $\square$

Lemma 6.2 *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*

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

where $B^{a_j}(x) \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, is strongly convex on $\mathbb{R}^n$.

Proof: For any $u \in \mathbb{R}^n$, we define a function $h_j : \mathbb{R}^n \to \mathbb{R}^m$ by

$$h_j(u) = \nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top B^{a_j}(x)u.$$

Since B^{a_j} is a symmetric positive definite matrix, there exists positive constant ρ_j such that

$$u^\top \nabla^2 h_j(x)u \succeq \rho_j \|u\|^2 e. \tag{6.5}$$

So, by Corollary 2.2 in [111], the function h_j is strongly convex, for all $j \in [w(x)]$. Hence, there exists $\mu > 0$ such that for any $u_1, u_2 \in \mathbb{R}^n$ and $\lambda \in [0,1]$,

$$h_j(\lambda u_1 + (1-\lambda)u_2) \preceq \lambda h_j(u_1) + (1-\lambda)h_j(u_2) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1 - u_2\|^2 e. \tag{6.6}$$

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

$$\begin{aligned} & g(\lambda u_1 + (1-\lambda)u_2) \\ = \ & \Psi_e(h_j(\lambda u_1 + (1-\lambda)u_2)) \\ \overset{(6.6)}{\leq} \ & \lambda \Psi_e(h_j(u_1)) + (1-\lambda)\Psi_e(h_j(u_2)) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1 - u_2\|^2 \\ = \ & \lambda g(u_1) + (1-\lambda)g(u_2) - \tfrac{\mu}{2}\lambda(1-\lambda)\|u_1 - u_2\|^2. \end{aligned}$$

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

Remark 6.2 *In view of Lemma 6.1, we note that every weakly minimal solution of* (SOP) *is a stationary point of* (SOP)*. Also, from Definition 1.37, we see that*

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

Thus, by Proposition 1.2(v)&(vi) and Remark 6.1, for a weakly minimal point $\bar{x}$, for any $a \in P_{\bar{x}}$ and $u \in \mathbb{R}^n$, there exists a_j with

$$\Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u + u^\top B^{a_j}(\bar{x})u\right) \geq \Psi_e\left(\nabla f^{a_j}(\bar{x})^\top u\right) + \tfrac{1}{2}\rho\|u\|^2 \geq 0, \tag{6.7}$$

where $\rho = \min\{\rho_1, \rho_2, \ldots, \rho_p\}$.

Next, we discuss a necessary condition for weakly minimal solutions of (SOP). For this, for any $x \in \mathbb{R}^n$, we define a function $\xi_x : P_x \times \mathbb{R}^n \to \mathbb{R}$ by

$$\xi_x(a, u) = \max_{j \in [w(x)]} \left\{\Psi_e(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2}u^\top\ B^{a_j}(x)u)\right\}, a \in P_x,\ u \in \mathbb{R}^n. \tag{6.8}$$

Then, by (6.7), at a stationary point $\bar{x}$, we have

$$\begin{aligned}
& \xi_{\bar{x}}(a, u) \geq 0 \ \ \forall\ a \in P_{\bar{x}} \text{ and } u \in \mathbb{R}^n \\
\implies\ & \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) \geq 0 \ \ \forall\ a \in P_{\bar{x}} \\
\implies\ & \forall a \in P_{\bar{x}} : 0 \leq \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) \leq \xi_{\bar{x}}(a, 0) = 0 \\
\implies\ & \forall a \in P_{\bar{x}} : \min_{u \in \mathbb{R}^n} \xi_{\bar{x}}(a, u) = 0.
\end{aligned} \tag{6.9}$$

Moreover, as for any $x \in \mathbb{R}^n$, P_x is finite, we note from Lemma 6.2 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 (6.9), we have

$$\xi_{\bar{x}}(a, \bar{u}_{a,\bar{x}}) = 0 \text{ if and only if } \bar{u}_{a,\bar{x}} = 0. \tag{6.10}$$

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

$$\Phi(x) = \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a, u). \tag{6.11}$$

Then, in view of (6.10) and (6.9), 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

$$\Phi(\bar{x}) = 0 \text{ and } \bar{u} = 0. \tag{6.12}$$

Accumulating all, we obtain the following result.

Proposition 6.1 (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 (6.8) and (6.11), respectively. Then, $\bar{u} = 0$.*

6.4 Quasi-Newton Method for (SOP)

The whole section is described in the following two subsections.

(i) At first, we discuss a few properties of Φ.

(ii) We propose the quasi-Newton method (Algorithm 2) and discuss the well-definedness of the proposed Algorithm 2. After that, we characterize the boundedness of the norm of descent direction for (SOP). Further, the existence of step length that satisfies the Armijo condition is derived.

6.4.1 Properties of Φ

In this subsection, we discuss a few properties of Φ, which play an important role in the convergence analysis of the proposed quasi-Newton method for (SOP).

Proposition 6.2 *The function Φ as given in (6.11) 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} is twice continuously differentiable for each $j \in [w(x)]$, the function ξ_x is continuous on $P_x \times \mathbb{R}^n$. Thus, we get

$$\limsup_{k\to\infty} \Phi(x_k) = \limsup_{k\to\infty} \xi_{x_k}(a^k, u_k) \le \limsup_{k\to\infty} \xi_{x_k}(\bar{a}, \bar{u}) = \xi_{\bar{x}}(\bar{a}, \bar{u}) = \Phi(\bar{x}). \quad (6.13)$$

Let $L > 0$ be a Lipschitz constant of Ψ_e. Then, from the definition (6.11) of Φ at $\bar{x}$, we observe that

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

$$
\begin{aligned}
&\quad +\Psi_e\left(\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2}u_k^\top B^{a_j^k}(\bar{x})u_k - \nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top B^{a_j^k}(x_k)u_k\right)\Bigg\}\Bigg\}\\
&= \liminf_{k\to\infty}\ \Bigg\{\xi_{x_k}(a^k,u_k) + \max_{j\in[w(\bar{x})]}\Bigg\{\Psi_e\left(\nabla f^{a_j^k}(\bar{x})^\top u_k\right.\\
&\qquad \left. +\tfrac{1}{2}u_k^\top B^{a_j^k}(\bar{x})u_k - \nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top B^{a_j^k}(x_k)u_k\right)\Bigg\}\Bigg\}\\
&\overset{1.2(\text{iii})}{\leq} \liminf_{k\to\infty}\ \Bigg\{\xi_{x_k}(a^k,u_k) + L\max_{j\in[w(\bar{x})]}\Big\{\|\nabla f^{a_j^k}(\bar{x})^\top u_k + \tfrac{1}{2}u_k^\top B^{a_j^k}(\bar{x})u_k\\
&\qquad -\nabla f^{a_j^k}(x_k)^\top u_k - \tfrac{1}{2}u_k^\top B^{a_j^k}(x_k)u_k\|\Big\}\Bigg\}\\
&\leq \liminf_{k\to\infty}\Bigg\{\ \xi_{x_k}(a^k,u_k) + L\max_{j\in[w(\bar{x})]}\{\|\nabla f^{a_j^k}(\bar{x}) - \nabla f^{a_j^k}(x_k)\|\|u_k\|\}\\
&\quad +\tfrac{L}{2}\max_{j\in[w(\bar{x})]}\left\{\|u_k^\top\left(B^{a_j^k}(\bar{x}) - B^{a_j^k}(x_k)\right)u_k\|\right\}\Bigg\}. \qquad (6.14)
\end{aligned}
$$

Note that for $j \in [w]$, 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 (6.14) that

$$\Phi(\bar{x}) \leq \liminf_{k\to\infty}\ \xi_{x_k}(a^k,u_k) = \liminf_{k\to\infty}\ \Phi(x_k). \qquad (6.15)$$

Finally, in view of (6.13) and (6.15), we conclude that

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

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

Proposition 6.3 *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 $a \in P_x$, $B^{a_j}(x) \leq qI$ 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 $\max\{b_1, b_2, \ldots, b_{w(x)}\} = \max_{\lambda \in \Delta_{w(x)}} \sum_{i=1}^{w(x)} \lambda_i b_i$ holds, where $\Delta_{w(x)} = \{(\lambda_1, \lambda_2, \ldots, \lambda_{w(x)}) \in \mathbb{R}_+^{w(x)} : \sum_{i=1}^{w(x)} \lambda_i = 1\}$. Thus, in view of the definition (6.11) of Φ, we have for any $a \in P_x$ that

$$
\begin{aligned}
|\Phi(x)| &= \left| \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a,u) \right| \\
&= \left| \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \max_{j \in [w(x)]} \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top B^{a_j}(x) u \right) \right\} \right| \\
&= \min_{(a,u) \in P_x \times \mathbb{R}^n} \left| \max_{\lambda \in \Delta_{w(x)}} \sum_{j=1}^{[w(x)]} \lambda_j \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top B^{a_j}(x) u \right) \right| \\
&\le \min_{(a,u) \in P_x \times \mathbb{R}^n} \sum_{j=1}^{w(x)} \left| \lambda_j \Psi_e \left(\nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top B^{a_j}(x) u \right) \right| \\
&\qquad \text{for any } \lambda \in \Delta_{w(x)} \\
&\overset{1.2\text{(iii)}}{\le} \min_{(a,u) \in P_x \times \mathbb{R}^n} \sum_{j=1}^{w(x)} \lambda_j L \left\| \nabla f^{a_j}(x)^\top u + \tfrac{1}{2} u^\top B^{a_j}(x) u \right\| \\
&\le L \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \sum_{j=1}^{w(x)} \lambda_j \left\| \nabla f^{a_j}(x)^\top u \right\| + \sum_{j=1}^{w(x)} \lambda_j \left\| \tfrac{1}{2} u^\top B^{a_j}(x) u \right\| \right\} \\
&\le L \min_{(a,u) \in P_x \times \mathbb{R}^n} \left\{ \sum_{j=1}^{w(x)} \lambda_j \| \nabla f^{a_j}(x)^\top u \| + \tfrac{\gamma}{2} \|u\|^2 \right\} \text{ as } B^{a_j}(x) \le qI. \qquad (6.16)
\end{aligned}
$$

Note that 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 $\mathbb{R}^n$. Therefore, 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, (6.16) gives that for all $\lambda_j \ge 0$ and $j \in w(x)$ with $\sum_{j=1}^{w(x)} \lambda_j = 1$, we have

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

□

6.4.2 Quasi-Newton Method for (SOP)

In this subsection, we propose a quasi-Newton method (Algorithm 2) 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 6.1, then we proceed to update this point as discussed in Algorithm 2. At each iteration, we select an element a from the partition set, and then a descent direction for (VOP) is evaluated by following the ideas of [86, 133]. 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 6.1 for a weakly minimal point is met. The entire method is given in Algorithm 2.

Remark 6.3 *It is to be noted that for $p = 1$ in Algorithm 2, that is, for $F(x) = \{f^1(x)\}$, the Step Step 4 of Algorithm 2 reduces to finding u_k such that*

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

In this case, the proposed Algorithm 2 reduces to the method given in [124] and [127, 134]. The difference is that we have used the Gerstewitz function instead of the support of a generator of the dual cone K^*.

Next, we show that Algorithm 2 is well-defined. The well-definedness of Algorithm 2 is based 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.

Algorithm 2 Quasi-Newton Method Algorithm 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.

Step 2 **Initialization**
Choose an initial point $x_0 \in \mathbb{R}^n$, a trial step length $\beta \in (0,1)$, and a positive $\nu \in (0,1)$.
Provide an initial symmetric positive definite matrix $B^i(x_0) \in \mathbb{R}^{n\times n}$, for each $i \in [p]$.
Set the iteration number $k = 0$.
Provide a value of the precision level $\varepsilon > 0$ 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 $(a^k, u_k) \in \underset{(a,u)\in P_k\times\mathbb{R}^n}{\text{argmin}}\ \xi_{x_k}(a,u)$, where $\xi_{x_k} : P_{x_k} \times \mathbb{R}^n \to \mathbb{R}$ is given by $\xi_{x_k}(a,u) = \max_{j\in[w_k]} \{\Psi_e(\nabla f^{a_j}(x_k)^\top u + \frac{1}{2} u^\top B^{a_j}(x_k) u)\}$.

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

Step 6 **Compute step length**
Evaluate the smallest value of q such that ν^q estimates the step length t_k by

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

Step 7 **Update the iterate and approximation**
Update $x_{k+1} \leftarrow x_k + t_k u_k$ and $k \leftarrow k+1$.
Compute $s_k = x_{k+1} - x_k$ and $y_k^i = \nabla f^i(x_{k+1}) - \nabla f^i(x_k)$, $i \in [p]$.
Evaluate the next symmetric positive definite matrix using the formula (6.1):

$$B^i(x_{k+1}) = B^i(x_k) - \frac{B^i(x_k) s_k s_k^\top B^i(x_k)}{s_k^\top B^i(x_k) s_k} + \frac{y_k^i {y_k^i}^\top}{s_k^\top y_k^i}.$$

Go to Step *Step* 3.

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

Therefore, Algorithm 2 is well-defined.

Next, we characterize the stationary points of (SOP) in terms of the functions ξ_x and Φ as defined in (6.8) and (6.11), respectively.

Theorem 6.1 *Let us consider the functions ξ_x and Φ given in (6.8) and (6.11), 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 $\widetilde{u} \in \mathbb{R}^n$ for which

$$\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top \widetilde{u}) < 0 \text{ for all } j \in [\bar{w}].$$

Thus, in view of the above relation, we conclude that

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

$$\leq \quad t\left\{\max_{j\in[w(\bar{x})]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\}+\tfrac{t}{2}\max_{j\in[w(\bar{x})]}\{\Psi_e(\widetilde{u}^\top B^{\widetilde{a}_j}(\bar{x})\widetilde{u})\}\right\}. \tag{6.17}$$

Choosing any t such that $0<t<\left(\frac{-2}{\max_{j\in[w(x)]}\{\Psi_e(\widetilde{u}^\top B^{\widetilde{a}_j}(\bar{x})\widetilde{u})\}}\right)\left(\max_{j\in[w(x)]}\{\Psi_e(\nabla f^{\widetilde{a}_j}(\bar{x})^\top\widetilde{u})\}\right)$, we obtain from (6.17) 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) $\implies$ (iii). It trivially follows from (6.12).

(iii) $\implies$ (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

$$\Psi_e(\nabla f^{\bar{a}_{\widetilde{j}}}(\bar{x})^\top\bar{u})\geq 0. \tag{6.18}$$

Note from Assumption 1 that for any $a\in P_x$ and $x\in\mathbb{R}^n$, we have $\bar{u}^\top B^{\bar{a}_{\widetilde{j}}}(\bar{x})\bar{u}>0$. Therefore, from (6.18) with the help of Proposition 1.2 (iv), we get

$$\begin{aligned}
&\Psi_e\left(\nabla f^{\bar{a}_{\widetilde{j}}}(\bar{x})^\top\bar{u}+\tfrac{1}{2}\bar{u}^\top B^{\bar{a}_{\widetilde{j}}}(\bar{x})\bar{u}\right)\geq 0\\
\text{or,}\quad &\max_{j\in[w(\bar{x})]}\left\{\Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top\bar{u}+\tfrac{1}{2}\bar{u}^\top B^{\bar{a}_j}(\bar{x})\bar{u})\right\}\geq 0\\
\text{or,}\quad &\xi_{\bar{x}}(\bar{a},\bar{u})\geq 0\\
\text{or,}\quad &\Phi(\bar{x})=0 \text{ from (6.9)}\\
\text{or,}\quad &\bar{u}=0 \text{ from (6.12)},
\end{aligned}$$

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

Remark 6.4 *In view of (6.12) and statements (i)–(iii) of Theorem 6.1, we obtain that $\bar{x}$ is a stationary point of (SOP) if and only if $\Phi(\bar{x})=0$ or $\bar{u}=0$.*

Next, we characterize an upper bound for the norm of quasi-Newton's direction u_k

generated by Algorithm 2 for (SOP). After that, we provide convergence analysis of Algorithm 2.

Theorem 6.2 *Let $\{x_k\}$ be the sequence of nonstationary points, $\{u_k\}$ be a sequence of descent directions generated by Algorithm 2, and $\{x_k\}$ be convergent. Then, the sequence $\{u_k\}$ is bounded.*

Proof: Let P_{x_k} be the partition set at $\{x_k\}$ and $\{x_k\}$ be a sequence of nonstationary points that converges to $\bar{x}$ (say). Then, in view of Theorem 5.1, there exists $a^k \in P_{x_k}$ such that $\Phi(x_k) < 0$, i.e.,

$$\begin{aligned} & \max_{j \in [w(x_k)]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k) \right\} < 0 \\ \Longrightarrow \quad & \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k) < 0 \text{ for all } \ j \in [w(x_k)] \\ \overset{1.2(\text{vi})}{\Longrightarrow} \quad & \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k) + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k < 0 \text{ for all } \ j \in [w(x_k)]. \end{aligned} \tag{6.19}$$

Then, note that $W = \{x_k\} \cup \{\bar{x}\}$ is compact. Moreover, $B^1, B^2, \ldots, B^p$ are continuous functions, for each $i \in [p]$. Then, we have

$$\sigma_i = \min_{\|u\|=1, \ x \in W} u^\top B^i(x) u > 0. \tag{6.20}$$

On taking $\rho = \min\{\sigma_1, \sigma_2, \ldots, \sigma_p\}$, we have $u^\top B^i(x) u \geq \rho \|u\|^2$. Thus, in view of (6.19) and (6.20), we get

$$\begin{aligned} & \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k) + \tfrac{1}{2} \rho \|u_k\|^2 < 0 \text{ for all } \ j \in [w(x_k)] \\ \Longrightarrow \ & \tfrac{1}{2} \rho \|u_k\|^2 \leq -\Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k) \text{ for all } \ j \in [w(x_k)] \\ \Longrightarrow \ & \tfrac{1}{2} \rho \|u_k\|^2 \leq \max_{j \in [w(x_k)]} \{ |\Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k)| \} \\ \overset{1.2(\text{iii})}{\Longrightarrow} \ & \tfrac{1}{2} \rho \|u_k\|^2 \leq L \max_{j \in [w(x_k)]} \|\nabla f^{a_j^k}(x_k)^\top u_k\|, \ L \text{ is a Lipschitz constant of } \Psi_e \\ \Longrightarrow \ & \tfrac{1}{2} \rho \|u_k\|^2 \leq L \|u_k\| \max_{j \in [w(x_k)]} \|\nabla f^{a_j^k}(x_k)\| \end{aligned}$$

$$\implies \tfrac{1}{2}\rho\|u_k\|^2 \leq L\|u_k\| \max\{\|\nabla f^1(x_k)\|, \|\nabla f^2(x_k)\|, \ldots, \|\nabla f^p(x_k)\|\}. \tag{6.21}$$

Note that each f^i, $i \in [p]$, is twice continuously differentiable and the sequence $\{x_k\}$ is convergent. Therefore, $\max\{\|\nabla f^1(x_k)\|, \|\nabla f^2(x_k)\|, \ldots, \|\nabla f^p(x_k)\|\}$ is a convergent sequence, and hence bounded. Now, let us assume that C be an upper bound of $\max\{\|\nabla f^1(x_k)\|, \|\nabla f^2(x_k)\|, \ldots, \|\nabla f^p(x_k)\|\}$. Then, from (6.21), we observe that

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

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

Next, we 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 2.

Proposition 6.4 *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 $j \in [\bar{w}]$,*

$$f^{\bar{a}_j}(\bar{x} + t\bar{u}) \preceq f^{\bar{a}_j}(\bar{x}) + \beta t \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}. \tag{6.22}$$

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

$$F(\bar{x} + t\bar{u}) \preceq^l \{f^{\bar{a}_j}(\bar{x}) + \beta t \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}\}_{j \in [\bar{w}]} \prec^l F(\bar{x}). \tag{6.23}$$

Proof: Let us assume that (6.22) does not hold. Therefore, there exists a sequence $\{t_k\} \searrow 0$ and $j' \in [\bar{w}]$ such that

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

$$\implies \quad \nabla f^{\bar{a}_{j'}}(\bar{x})^\top \bar{u} \notin -\text{int}(K) \text{ since } \beta \in (0,1). \tag{6.24}$$

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 6.1, we have

$$\begin{aligned}
& \xi_{\bar{x}}(\bar{a}, \bar{u}) = \Phi(\bar{x}) < 0 \\
\implies \quad & \max_{j \in [\bar{w}]} \Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} + \tfrac{1}{2}\bar{u}^\top B^{\bar{a}_j}(\bar{x})\bar{u}) < 0 \text{ since } \bar{u} \neq 0 \\
\implies \quad & \Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} + \tfrac{1}{2}\bar{u}^\top B^{\bar{a}_j}(\bar{x})\bar{u}) < 0 \text{ for all } j \in [\bar{w}] \\
\overset{(6.5)}{\implies} \quad & \Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}) + \tfrac{1}{2}\rho_{\bar{a}_j}\|\bar{u}\|^2 < 0 \text{ for all } j \in [\bar{w}] \\
\implies \quad & \Psi_e(\nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u}) < 0 \\
\implies \quad & \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} \in -\text{int}(K),
\end{aligned} \tag{6.25}$$

which is a contradiction to (6.24). Therefore, we conclude that for every $j \in [\bar{w}]$, the relation (6.22) holds.

Now, for a nonstationary point $\bar{x}$, from (6.22), we observe that

$$f^{\bar{a}_j}(\bar{x}) + \beta t \nabla f^{\bar{a}_j}(\bar{x})^\top \bar{u} \prec f^{\bar{a}_j}(\bar{x}) \text{ for all } j \in [\bar{w}] \text{ and } t \in (0, \widetilde{t}]. \tag{6.26}$$

From Proposition 1.1, we observe that for every $t \in (0, \bar{t}]$

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

which implies that for every $t \in (0, \bar{t}]$, (6.23) holds. □

6.5 Convergence Analysis of Quasi-Newton Method

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 6.1 (Regular point [1]). *A point $\bar{x}$ 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 6.3 (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 method shown in Algorithm 2.

Theorem 6.3 *Let $\{x_k\}$ is an infinite sequence generated by Algorithm 2 and $\bar{x}$ is an accumulation point for the sequence $\{x_k\}$. Additionally, assume that $\bar{x}$ is a regular point of F. Then, $\bar{x}$ is a stationary point of (SOP).*

Proof: Without loss of generality, let $\{x_k\}$ be a subsequence of $\{x_k\}$ which converge to an accumulation point $\bar{x}$. We prove that $\bar{x}$ is stationary. Towards this, define the functional $\varsigma : \mathcal{P}(\mathbb{R}^m) \to \mathbb{R} \cup \{-\infty\}$ given by

$$\varsigma(A) = \inf_{z \in A} \Psi_e(z) \text{ for all } A \in \mathcal{P}(\mathbb{R}^m).$$

From Proposition 1.2(iv), the function Ψ_e is monotonic. Therefore, the functional ς is monotone with respect to the preorder $\preceq^l$, that is, for all $A, B \in \mathcal{P}(\mathbb{R}^m)$, we have

$$A \preceq^l B \implies \varsigma(A) \leq \varsigma(B).$$

Now in view of (6.23) of Proposition 5.4, for every $k = 0, 1, \ldots$, we obtain

$$\begin{aligned}
&\varsigma(F(x_{k+1}))\\
= \quad &\varsigma(F(x_k + t_k u_k))\\
\leq \quad &\min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) + \beta t_k \nabla f^{a_j^k}(x_k)^\top u_k \right) \right\}\\
\leq \quad &\min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) + \beta t_k (\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k) \right) \right\}\\
&\text{from Proposition 1.2(iv) and } u_k^\top B^{a_j^k}(x_k) u_k > 0\\
\leq \quad &\min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right) \right\}\\
&\text{from Proposition 1.2(i)}\\
\leq \quad &\min_{j\in[w_k]} \left\{ \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \max_{j'\in[w_k]} \Psi_e \left(\nabla f^{a_{j'}^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_{j'}^k}(x_k) u_k \right) \right\}\\
\leq \quad &\min_{j\in[w_k]} \Psi_e \left(f^{a_j^k}(x_k) \right) + \beta t_k \max_{j\in[w_k]} \left\{ \Psi_e \left(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right) \right\}\\
= \quad &\varsigma(F(x_k)) + \beta t_k \Phi(x_k). \qquad (6.27)
\end{aligned}$$

Therefore, for a fixed $k \in \mathbb{N} \cup \{0\}$, we get

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

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

$$-\beta \sum_{k=0}^{\kappa} t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right\} \leq \varsigma(F(x_0)) - \varsigma(F(x_{\kappa+1})). \quad (6.28)$$

Since $\{x_k\}$ is a convergent sequence and Ψ_e is monotonic, from (6.28), we have

$$-\beta \lim_{k\to\infty} \sum_{k=0}^{\kappa} t_k \max_{j\in[w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right\} \leq +\infty. \quad (6.29)$$

Given that $\{x_k\}$ is a sequence of nonstationary points, therefore 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 $k \in \mathbb{N} \cup \{0\}$, we get $\nabla f^{a_j^k}(x_k)^\top u_k \in -\operatorname{int}(K)$.

On proceeding in similar manner to (6.17), we can find $t_k > 0$ such that

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

In view of the above chosen t_k, we conclude that

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

Therefore, we get

$$-t_k \max_{j \in [w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right\} \geq 0. \tag{6.30}$$

On combining (6.29) and (6.30), and taking limit $k \to \infty$, we have

$$\lim_{k \to \infty} t_k \max_{j \in [w_k]} \left\{ \Psi_e(\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{a_j^k}(x_k) u_k \right\} = 0. \tag{6.31}$$

Since $\bar{x}$ is an accumulation point of the sequence $\{x_k\}$ and the sequences $\{t_k\}$ and $\{u_k\}$ are bounded (Theorem 6.3), therefore we can find $\bar{t} \in \mathbb{R}_+$, $\bar{u} \in \mathbb{R}^n$ and a subsequence $\mathcal{K} \in \mathbb{N}$ such that

$$t_k \xrightarrow{k \in \mathcal{K}} \bar{t} \text{ and } u_k \xrightarrow{k \in \mathcal{K}} \bar{u}.$$

Note that the number of points in $[p]$ is finite, and $\bar{x}$ is a regular point of F. Thus, in view of Lemma 6.3, for all $k \in \mathcal{K}, u \in \mathbb{R}^n$, we have $w_k = \bar{w}, P_{x_k} = \bar{P}, a^k = \bar{a}$ and

$$\begin{aligned} & \Phi(x_k) = \xi_{x_k}(\bar{a}, u_k) \leq \xi_{x_k}(a, u) \\ \text{and} \quad & \xi_{\bar{x}}(\bar{a}, \bar{u}) \leq \xi_{\bar{x}}(a, u) \text{ on taking limit } k \xrightarrow{\mathcal{K}} \infty. \end{aligned} \tag{6.32}$$

Now, we analyze the following two cases:

(i) Let $\bar{t} > 0$. In view of (6.31) and for all $k \in \mathcal{K}$ such that $w_k = \bar{w}, P_{x_k} = \bar{P}, a^k = \bar{a}$, we have

$$\begin{aligned} & \lim_{k \xrightarrow{\mathcal{K}} \infty} \max_{j \in [\bar{w}]} \left\{ \Psi_e(\nabla f^{\bar{a}_j}(x_k)^\top u_k + \tfrac{1}{2} u_k^\top B^{\bar{a}_j}(x_k) u_k \right\} = 0 \\ \implies & \lim_{k \xrightarrow{\mathcal{K}} \infty} \Phi(x_k) = 0. \end{aligned} \tag{6.33}$$

Thus, by Theorem 6.1, we conclude that $\bar{u} = 0$. Hence, $\bar{x}$ is a stationary point of (SOP).

(ii) Let $\bar{t} = 0$. Fix any $\kappa \in \mathbb{N}$. Since $t_k \xrightarrow{\mathcal{K}} \bar{t} = 0$, large enough ν^κ does not satisfy Armijo condition in Step *Step* 6 of Algorithm 2. 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

$$\begin{aligned} & f^{\bar{a}_{\bar{j}}}(x_k + \nu^\kappa u_k) \npreceq f^{\bar{a}_{\bar{j}}}(x_k) + \beta \nu^\kappa \nabla f^{\bar{a}_{\bar{j}}}(x_k)^\top u_k \\ \implies & \frac{f^{\bar{a}_{\bar{j}}}(x_k + \nu^\kappa u_k) - f^{\bar{a}_{\bar{j}}}(x_k)}{\nu^\kappa} - \beta \nabla f^{\bar{a}_{\bar{j}}}(x_k)^\top u_k \notin -K \\ \implies & \frac{f^{\bar{a}_{\bar{j}}}(\bar{x} + \nu^\kappa \bar{u}) - f^{\bar{a}_{\bar{j}}}(\bar{x})}{\nu^\kappa} - \beta \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ taking } k \xrightarrow{\mathcal{K}} +\infty \\ \implies & (1-\beta) \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ taking limit } k \to +\infty \\ \implies & \nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} \notin -\operatorname{int}(K) \text{ since } (1-\beta) \in (0,1). \end{aligned}$$

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

$$\begin{aligned} & \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u}) \geq 0 \\ \text{or,} \quad & \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} + \tfrac{1}{2} \bar{u}^\top B^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u}) \geq 0 \\ & \quad \text{by Proposition 1.2(iv) and } \bar{u}^\top B^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u} \succ 0 \\ \text{or,} \quad & 0 \leq \Psi_e(\nabla f^{\bar{a}_{\bar{j}}}(\bar{x})^\top \bar{u} + \tfrac{1}{2} \bar{u}^\top B^{\bar{a}_{\bar{j}}}(\bar{x}) \bar{u}) \\ \text{or,} \quad & 0 \leq \xi_{\bar{x}}(\bar{a}, \bar{u}) = \min_{(a,u) \in P_x \times \mathbb{R}^n} \xi_x(a, u) = \Phi(\bar{x}) \leq 0 \text{ from (6.11).} \end{aligned}$$

Thus, from Theorem 6.1, we conclude that $\bar{x}$ is a stationary point of (SOP).

□

Now, we recall some results that help in analyzing the convergence properties of the proposed quasi-Newton Algorithm 1. The first result follows from [135, Theorem 3.1], which says the following. Let $\{x_k\}$ be a sequence of nonstationary points converging to $\bar{x}$. Then, we have the relation given by

$$\lim_{k\to\infty} \frac{\|(B(x_k) - \nabla^2 f(\bar{x}))s_k\|}{\|s_k\|} = 0,$$

where $s_k = x_{k+1} - x_k = t_k u_k$, $B(x_k)$ is the BFGS approximation of the Hessian $\nabla^2 f(x_k)$, and f is twice continuously differentiable such that $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite. Now, from [124], and in view of above result, for any $\varepsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, the following holds

$$\lim_{k\to\infty} \frac{\|(B^{a_j^k}(x_k) - \nabla^2 f^{a_j^k}(\bar{x}))u_k\|}{\|u_k\|} < \varepsilon \text{ for every } j \in [w(x)].$$

Next, we recall the assumptions from [136], where for each $j \in [w(x)]$, the authors have estimated the error of approximating $\nabla f^{a_j^k}$ and $f^{a_j^k}$ by its linear and quadratic models, respectively.

Lemma 6.4 (See [136]). *Let U be a nonempty subset of $\mathbb{R}^n$ and $\varepsilon, \delta > 0$ be such that for any $x, y \in U$ with $\|y - x\| < \delta$, the following conditions hold:*

(i) For every $j \in [w(x)]$, we have

$$\|\nabla^2 f^{a_j}(y) - \nabla^2 f^{a_j}(x)\| < \tfrac{\varepsilon}{2}. \tag{6.34}$$

(ii) For every $j \in [w(x)]$ and $x, y \in U$ such that $\|y - x\| < \delta$, we have

$$\|\nabla f^{a_j}(y) - [\nabla f^{a_j}(x) + \nabla^2 f^{a_j}(x)(y - x)]\| < \tfrac{\varepsilon}{2}\|y - x\|. \tag{6.35}$$

(iii) Also, for every $j \in [w(x)]$ and $x, y \in U$ such that $\|y - x\| < \delta$, we have

$$|f^{a_j}(y) - \Big(f^{a_j}(x) + \nabla f^{a_j}(x)(y-x) + \tfrac{1}{2}(y-x)^\top \nabla^2 f^{a_j}(x)(y-x)\Big)| < \tfrac{\varepsilon}{4}\|y-x\|^2. \tag{6.36}$$

Now, we modify Lemma 6.4 to estimate the error of approximations, where we use the BFGS approximation of the second-order derivative term Hessian.

Lemma 6.5 *Let U be a nonempty subset of $\mathbb{R}^n$. Let $V \subset U$ be a convex subset and $\delta \in \mathbb{R}_+$ be constants such that $x, y \in V$ with $\|y - x\| < \delta$. Let $\{x_k\}$ be a sequence generated by Algorithm 1. Assume that for every $\varepsilon > 0$ and for every $j \in [w(x_k)]$, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, we have*

$$\frac{\big\|(\nabla^2 f^{a_j}(x_k) - B^{a_j}(x_k))(y - x_k)\big\|}{\|y - x_k\|} < \tfrac{\varepsilon}{2}. \tag{6.37}$$

Then, for any x_k and $k \geq k_0$, and any $x, y \in V$ such that $\|y - x_k\| < \delta$, we have

$$\|\nabla f^{a_j}(y) - (\nabla f^{a_j}(x_k) + B^{a_j}(x_k)(y - x_k))\| < \varepsilon\|y - x_k\| \tag{6.38}$$

$$\text{and } |f^{a_j}(y) - (f^{a_j}(x_k) + \nabla f^{a_j}(x_k)(y-x_k) + \tfrac{1}{2}(y-x_k)^\top B^{a_j}(x_k)(y-x_k))| < \tfrac{\varepsilon}{2}\|y-x_k\|^2, \tag{6.39}$$

for every $j \in [w(x)]$.

Proof: In view of (6.35) of Lemma 6.4, for every $j \in [w(x)]$ and $x, y \in U$ such that $\|y - x\| < \delta$, we have

$$\begin{aligned}
& \|\nabla f^{a_j}(y) - [\nabla f^{a_j}(x_k) + B^{a_j}(x_k)(y - x_k)]\| \\
\leq\ & \|\nabla f^{a_j}(y) - \nabla f^{a_j}(x_k) - \nabla^2 f^{a_j}(x_k)(y - x_k)]\| + \|(\nabla^2 f^{a_j}(x_k) - B^{a_j}(x_k))(y - x_k)]\| \\
<\ & \tfrac{\varepsilon}{2}\|y - x_k\| + \tfrac{\varepsilon}{2}\|y - x_k\| \text{ from (6.37)}
\end{aligned}$$

$$< \ \varepsilon\|y - x_k\|.$$

In the similar manner in (6.36) of Lemma 6.4, for every $j \in [w(x)]$ and $x, y \in U$ such that $\|y - x\| < \delta$, we have

$$\begin{aligned}
& \left\| f^{a_j}(y) - \left(f^{a_j}(x_k) + \nabla f^{a_j}(x_k)^\top (y - x_k) + \tfrac{1}{2}(y - x_k)^\top B^{a_j}(x_k)(y - x_k) \right) \right\| \\
\leq \ & \left\| f^{a_j}(y) - f^{a_j}(x_k) - \nabla f^{a_j}(x_k)^\top (y - x_k) - \tfrac{1}{2}(y - x_k)^\top \nabla^2 f^{a_j}(x_k)(y - x_k) \right\| \\
& + \left\| \tfrac{1}{2}(y - x_k)^\top (\nabla^2 f^{a_j}(x_k) - B^{a_j}(x_k))(y - x_k) \right\| \\
< \ & \tfrac{\varepsilon}{4}\|y - x_k\|^2 + \tfrac{\varepsilon}{4}\|y - x_k\|^2 \text{ from (6.37)} \\
< \ & \tfrac{\varepsilon}{2}\|y - x_k\|^2,
\end{aligned}$$

which is the required relation. □

Theorem 6.4 (Superlinear convergence). *Let $\{x_k\}$ be a sequence of nonstationary points generated by Algorithm 2 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) $pI \leq B^{a_j}(x) \leq qI$ *for all* $j \in [w(x)]$*, where* I *is* $n \times n$ *identity matrix,*

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

(iii) $\|(\nabla^2 f^{a_j}(x_k) - B^{a_j}(x_k))(y - x_k)]\| < \frac{\varepsilon}{2}$ *for all* $x, y \in V$ *with* $\|x - y\| < \delta$*, and*

(iv) $\frac{\varepsilon}{q} \leq 1 - 2\beta$*.*

Then, for sufficiently large $k \in \mathbb{N}$, we have $t_k = 1$ and the sequence $\{x_k\}$ converges superlinearly to $\bar{x} \in \mathbb{R}^n$.

Proof: From Theorem 6.3, it can be observed that the sequence $\{x_k\}$ converges to a stationary point $\bar{x}$ of (SOP). Moreover, each f^{a_j} is twice continuously differentiable.

Therefore, for any $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that for all $x, y \in \mathcal{B}(\bar{x}, \delta_\varepsilon)$, we have

$$\mathcal{B}(\bar{x}, \delta_\varepsilon) \subset V,\ \|B^{a_j}(x) - B^{a_j}(y)\| < \varepsilon \text{ with } \|x - y\| < \delta_\varepsilon.$$

Now, for $x \in U$, and $\Delta_{w(x)} = \{(\lambda_1, \lambda_2, \ldots, \lambda_{w(x)}) \in \mathbb{R}_+^{w(x)} \text{ with } \sum_{i=1}^{w(x)} \lambda_i = 1\}$, we define a function $\Theta_\lambda : V \times \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\Theta_\lambda(x, u) = \sum_{j=1}^{[w(x)]} \lambda_j \nabla f^{a_j}(x)^\top u + \tfrac{1}{2} \sum_{j=1}^{[w(x)]} \lambda_j u^\top B^{a_j}(x)^\top u,\ \lambda_{w(x)} \in \Delta_{w(x)}.$$

Note that for any $a^k \in P_{x_k}$ and $j \in [w(x_k)]$, each $f^{a_j^k}$ is twice continuously differentiable and strongly convex function. Moreover, for any $x_k \in \mathbb{R}^n$, and $a^k \in P_{x_k}$, 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 the existence of a minimizer, we conclude that

$$\sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_k) + \sum_{j=1}^{[w_k]} \lambda_j B^{a_j^k}(x_k)^\top u_k = 0 \tag{6.40}$$

$$\implies \quad u_k = -\left[\sum_{j=1}^{[w_k]} \lambda_j B^{a_j^k}(x_k)\right]^{-1} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_k)$$

$$\implies \quad u_k \leq -\tfrac{1}{q} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_k) \text{ since } B^{a_j^k}(x_k) \leq qI$$

$$\implies \quad u_k \leq -\tfrac{1}{q} \max_{\lambda \in \Delta_k} \sum_{j=1}^{[w_k]} \lambda_j \nabla f^{a_j^k}(x_k). \tag{6.41}$$

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

$f^{a_j^k}$, we have

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

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

Therefore, observing this identity in the above relation, we get

$$\begin{aligned}
& f^{a_j^k}(x_k+u_k) - f^{a_j^k}(x_k) \\
\leq\; & \nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{1}{2}u_k^\top B^{a_j^k}(x_k)u_k + \tfrac{\varepsilon}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\nabla f^{a_j^k}(x_k)^\top u_k + \tfrac{(q+\varepsilon)}{2}\|u_k\|^2 \text{ since } B^{a_j^k}(x) \leq qI \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\max_{j \in [w(x_k)]}\{\nabla f^{a_j}(x_k)^\top u_k\} + \tfrac{(q+\varepsilon)}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + (1-\beta)\max_{\lambda \in \Delta_k}\left\{\sum_{j=1}^{[w(x_k)]} \lambda_j \nabla f^{a_j^k}(x_k)^\top u_k\right\} + \tfrac{(q+\varepsilon)}{2}\|u_k\|^2 \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k - q(1-\beta)\|u_k\|^2 + \tfrac{(q+\varepsilon)}{2}\|u_k\|^2 \text{ from (6.41)} \\
\leq\; & \beta \nabla f^{a_j^k}(x_k)^\top u_k + \frac{\varepsilon - q(1-2\beta)}{2}\|u_k\|^2,
\end{aligned}$$

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

$$\begin{aligned}
& \left\|\sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_{k+1})\right\| \\
=\; & \left\|\sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k+u_k)\right\| \\
=\; & \left\|\sum_{j=1}^{[w(x_{k+1})]} \lambda_j \nabla f^{a_j^k}(x_k+u_k) - \left[\sum_{j=1}^{[w(x_k)]} \lambda_j \nabla f^{a_j^k}(x_k) + \sum_{j=1}^{[w(x_k)]} \lambda_j B^{a_j^k}(x_k)^\top u_k\right]\right\| \text{ from (6.40)} \\
\leq\; & \varepsilon\|u_k\| \text{ from (6.38) of Lemma 6.5.}
\end{aligned} \tag{6.42}$$

Now, combining assumption (i) and boundedness of $\{u_{k+1}\}$ (Theorem 6.2), we observe that

$$\tfrac{1}{2}u_k^\top B^{a_j^k}(x_k)u_k \leq \tfrac{q}{2}\|u_k\|^2 \text{ for any } j \in [w(x_k)]. \tag{6.43}$$

Therefore, incorporating the above relation in (6.19), we get

$$\begin{aligned}
&\|u_{k+1}\| \\
&\leq \tfrac{2L}{q} \max_{j\in[w(x_k)]} \left\|\nabla f^{a_j^k}(x_{k+1})\right\| \\
&\leq \tfrac{2L}{q} \left\{\max_{\lambda\in\Delta_k} \left\| \sum_{j=1}^{[w(x_k)]} \nabla f^{a_j^k}(x_{k+1}) \right\| \right\} \text{using } \max\{b_1, b_2, \ldots, b_{w(x)}\} = \max_{\lambda\in\Delta_{w(x)}} \sum_{i=1}^{w(x)} \lambda_i b_i \\
&\leq \tfrac{2L\varepsilon}{q}\|u_k\| \text{ from (6.42).}
\end{aligned}$$

In view of the above relation, we have

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

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

$$\begin{aligned}
\|x^{k+m} - x^{k+m+1}\| &\leq \left(\tfrac{2L\varepsilon}{q}\right)\|x^{k+m-1} - x^{k+m}\| \\
&\leq \left(\tfrac{2L\varepsilon}{q}\right)^2\|x^{k+m-1} - x^{k+m}\| \\
&\leq \cdots \leq \left(\tfrac{2L\varepsilon}{q}\right)^m\|x^k - x^{k+1}\|.
\end{aligned} \tag{6.44}$$

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

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

If we take $\varepsilon < \bar{\varepsilon}$ and $k \geq k_\varepsilon$, then by convergence of sequence $\{x_k\}$ and relation (6.44), 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 $\varepsilon < \bar{\varepsilon}$ and $k \geq k_\varepsilon$, then $\frac{\|\bar{x}-x^{k+1}\|}{\|\bar{x}-x^k\|} \leq \tau$. $\square$

6.6 Numerical Demonstrations and Execution of Results

In this section, we execute the proposed quasi-Newton Algorithm 2 on some numerical examples. We analyze the experimentation of Algorithm 2 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 parameter values:

- We considered the cone K to be a standard ordering cone, that is, $K = \mathbb{R}^2_+$ for all test instances except Example 6.5 and Example 6.6, and the parameter $e = (1, 1)^\top \in \text{int}(K)$ for the scalarizing function Ψ_e.

- The parameters β and ν in Step *Step* 6 for the line search of the Algorithm 2 was chosen as $\beta = 0.5$ and $\nu = 0.6$.

- The employed stopping criterion is $\|u_k\| < 0.001$, or a maximum number of 100 iterations was reached.

- To figure out the set Min $(F(x_k), K)$ at the k-th iteration in Step *Step* 3 of Algorithm 2, we adopted the common method of comparing the elements in $F(x_k)$.

- At the k-th iteration in Step *Step* 3 of Algorithm 2, for every $a^k \in P_k$, we compute the unique solution u_a of the function given by

$$\min_{u \in \mathbb{R}^n} \xi_{x_k}(a, u).$$

In the conclusion, we find $(a^k, u_k) = \underset{(a,u) \in P_k \times \mathbb{R}^n}{\operatorname{argmin}} \xi_{x_k}(a, u)$ 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 freshly introduced problems. For each test considered, we generated 100 initial points randomly and ran the algorithm. In the context of each experiment, we have presented a table with three columns. The resulting error is the value of $\|u_k\|$ at the end of the final iteration. The subsequent values are gathered for each test instance:

  - **Initial Points:** The value represents the first column of the table, which counts the number of initial points taken to solve the proposed Algorithm 2.
  - **Iterations:** This value presents the second 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 iterations in which the stopping condition is reached.
  - **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 2.

Furthermore, we compare the results of the proposed quasi-Newton's method (abbreviated as QNM) algorithm (Algorithm 2) with the existing steepest descent method (abbreviated as SD) for set optimization presented in [1].

Now, we discuss different test problems on which our algorithm was tested. The first test problem is freshly introduced.

Example 6.1 *Consider the set-valued function $F : \mathbb{R} \rightrightarrows \mathbb{R}^2$ defined as*

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

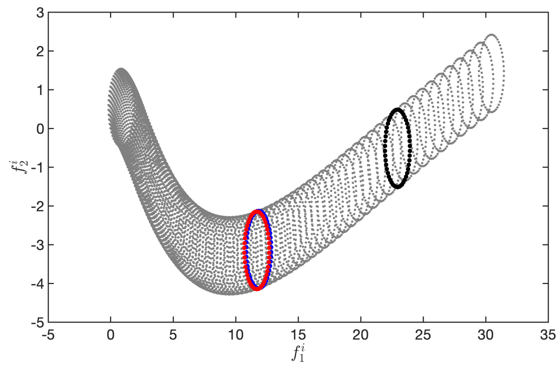
where for each $i \in [50]$, $f^i : \mathbb{R} \to \mathbb{R}^2$ is given by

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

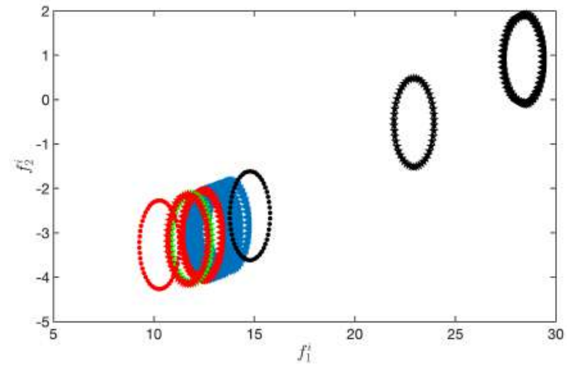
Output of Algorithm 2 for different initial points of Example 6.1 are depicted in Figure 6.1. Figure 6.1(a) depicts the sequence $\{F(x_k)\}$ generated by Algorithm 2 for the starting point as $x_0 = 2.3000$. In Figure 6.1(b), we exhibit the output of Algorithm 2 for three initial points.

The performance of Algorithm 2 for Example 6.1 is shown in Table 6.1. The values in

Table 6.1 show that the proposed method performs better than the existing SD method.



(a) The value of F at each iteration generated by Algorithm 2 for Example 6.1 for the initial point $x_0 = 2.3000$



(b) The value of F for three different initial points at each iteration generated by Algorithm 2 on Example 6.1

Figure 6.1: Obtained output of Algorithm 2 for Example 6.1

Table 6.1: Performance of Algorithm 2 on Example 6.1

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 22, 19.1700, 19, 21, 2.4621)	(3.2973, 44.5769, 37.5000, 37.6309, 3, 4.7245)
	SD	(1, 33, 32.4600, 32, 32, 0.5009)	(2.1389, 32.6268, 31.5719, 31.3467, 30, 0.5165)

Further, for the initial point $x_0 = 2.3000$*, the decreasing behavior in the values of vector-valued functions at each iteration has been exhibited in Table 6.2.*

Table 6.2: Output of Algorithm 2 on Example 6.1 with $x_0 = 2.3000$

k	$x_k^\top$	$f^{10}(x_k)$	$f^{25}(x_k)$	$f^{50}(x_k)$
0	2.3000	$(23.8454, -0.0901)$	$(23.0660, -1.5080)$	$(22.8153, 0.4762)$
1	1.8541	$(12.7453, -2.7029)$	$(11.9658, -4.1208)$	$(11.7151, -2.1366)$
2	1.8458	$(12.5946, -2.7214)$	$(11.8151, -4.1393)$	$(11.5644, -2.1551)$
3	1.8397	$(12.4845, -2.7343)$	$(11.7050, -4.1522)$	$(11.4543, -2.1679)$

Example 6.2 *Consider the set-valued function* $F : \mathbb{R} \rightrightarrows \mathbb{R}^3$ *defined as*

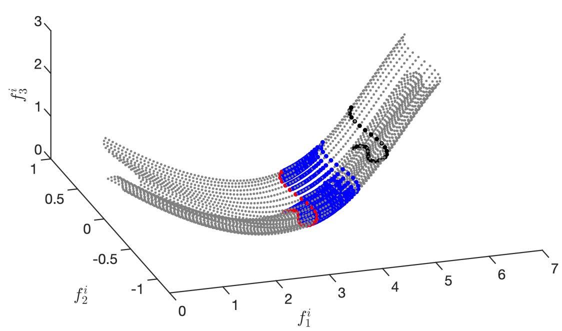
$$F(x) = \{f^1(x), f^2(x), \ldots, f^{30}(x)\},$$

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

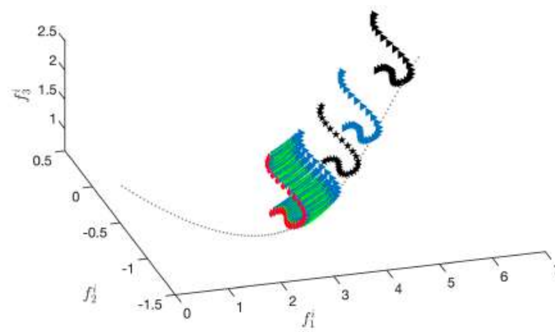
$$f^i(x) = \begin{pmatrix} 0.27\sin\left(\frac{2\pi(i-1))}{30}\right)\cos\left(\frac{2\pi(i-1))}{30}\right) + x^2 \\ \cos(2x) + \frac{1}{(1+e^{2x})} + 0.27\cos\left(\frac{2\pi(i-1))}{30}\right) \\ 0.27x^2 + \left(\frac{i-1}{30}\right) \end{pmatrix}.$$

Output of Algorithm 2 for different initial points of Example 6.2 are depicted in Figure 6.2. Figure 6.2(a) depicts the sequence $\{F(x_k)\}$ *generated by Algorithm 2 for the starting point as* $x_0 = 2.1300$. *In Figure 6.2(b), we exhibit the output of Algorithm 2 for three initial points and depict their corresponding* F*-values.*

The performance of Algorithm 2 for Example 6.2 is shown in Table 6.3. Moreover, we have compared the results of the QNM with the SD method for set optimization as presented in Table 6.3. The values in Table 6.3 show that the proposed method performs better than the existing SD method.



(a) The value of F at each iteration generated by Algorithm 2 for Example 6.2 for the initial point $x_0 = 2.1300$



(b) The value of F for three different initial points at each iteration generated by Algorithm 2 on Example 6.2

Figure 6.2: Obtained output of Algorithm 2 for Example 6.2

Table 6.3: Performance of Algorithm 2 on Example 6.2

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 10, 5, 4.5800, 5, 1, 1.9719)	(2.5750, 23.7773, 14.9322, 16.2000, 2, 10.2129)
	SD	(1, 11, 10.220, 9, 2, 2.4887)	(1.2113, 28.1483, 26.1314, 26.1002, 25, 0.2662)

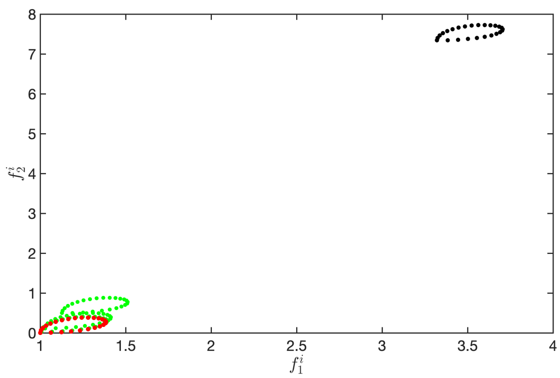
Example 6.3 *Consider the function $F : \mathbb{R} \rightrightarrows \mathbb{R}^2$ defined as*

$$F(x) = \{f^1(x), f^2(x), \ldots, f^{25}(x)\},$$

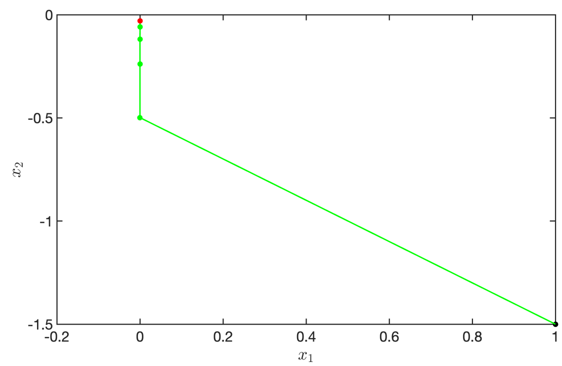
where for each $i \in [25]$, $f^i : \mathbb{R} \to \mathbb{R}^2$ is given by

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

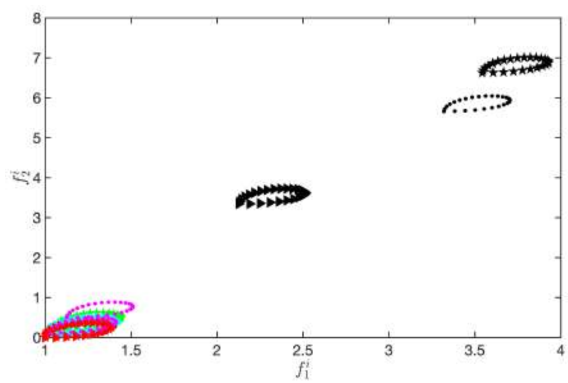
In Figure 6.3, the iterates generated by Algorithm 2 for different initial points taken



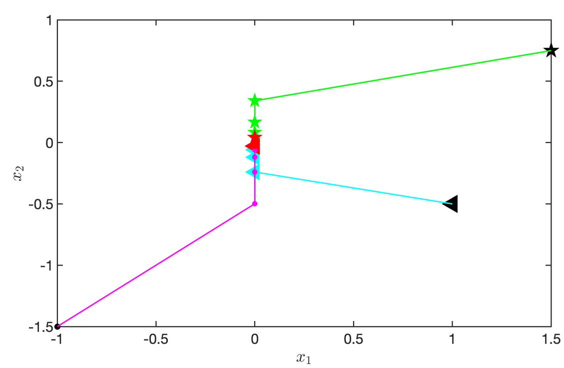
(a) The value of F at each iteration generated by Algorithm 2 for initial point $x_0 = (1.0000, -1.5000)^\top$ of Example 6.3



(b) TThe movement of subsequent x_k generated by Algorithm 2 of initial point $x_0 = (1.0000, -1.5000)^\top$ of Example 6.3



(c) The value of F at each iteration generated by Algorithm 2 for three different randomly chosen initial points of Example 6.3



(d) The movement of subsequent x_k generated by Algorithm 2 for three different randomly chosen initial points of Example 6.3

Figure 6.3: Obtained output of Algorithm 2 for Example 6.3

from the set $[-5,5] \times [-5,5]$ *are given. The sequence of iterates* $\{x_k\}$ *and the corresponding* $\{F(x_k)\}$ *generated by Algorithm 2 for an initial point* $x_0 = (1.0000, -1.5000)^\top$ *are given in Figure 6.3(b) and Figure 6.3(a), respectively. Moreover, for the other three randomly selected initial points, the sequence of iterates* $\{x_k\}$ *and the corresponding* $\{F(x_k)\}$ *generated by Algorithm 2 are shown in Figure 6.3(d) and Figure 6.3(c), respectively.*

The performance of Algorithm 2 for Example 6.3 is shown in Table 6.4. Moreover, we have compared the results of the QNM with the SD method for set optimization as presented in Table 6.4. The values in Table 6.4 show that the proposed method performs better than the existing SD method.

Table 6.4: Performance of Algorithm 2 on Example 6.3

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 6, 5.7400, 6, 6, 0.4845)	(1.2551, 42.8398, 23.8484, 29.5973, 11, 11.9814)
	SD	(1, 14, 13.9900, 14, 14, 0.1000)	(0.112, 4.2931, 3.2962, 3.2843, 3, 0.1097)

For the initial point $x_0 = (1.0000, -1.5000)^\top$*, the decreasing behavior in the values of vector-valued functions at each iteration has been exhibited in Table 6.5.*

Table 6.5: Output of Algorithm 2 on Example 6.3 with $x_0 = (1.0000, -1.5000)^\top$

k	$x_k^\top$	$f^5(x_k)$	$f^{15}(x_k)$	$f^{25}(x_k)$
0	$(1.0000, -1.5000)$	$(3.3806, 7.5748)$	$(3.6992, 7.6545)$	$(3.3833, 7.3454)$
1	$(-0.0000, -0.4987)$	$(1.1868, 0.7308)$	$(1.5054, 0.8105)$	$(1.1895, 0.5014)$
2	$(0.0000, -0.2392)$	$(1.0886, 0.3477)$	$(1.4072, 0.4275)$	$(1.0913, 0.1184)$
3	$(0.0000, -0.1185)$	$(1.0669, 0.2614)$	$(1.3855, 0.3412)$	$(1.0696, 0.0320)$
4	$(0.0001, -0.0591)$	$(1.0617, 0.2404)$	$(1.3802, 0.3201)$	$(1.0643, 0.0110)$
5	$(0.0001, -0.0295)$	$(1.0603, 0.2352)$	$(1.3789, 0.3149)$	$(1.0630, 0.0058)$

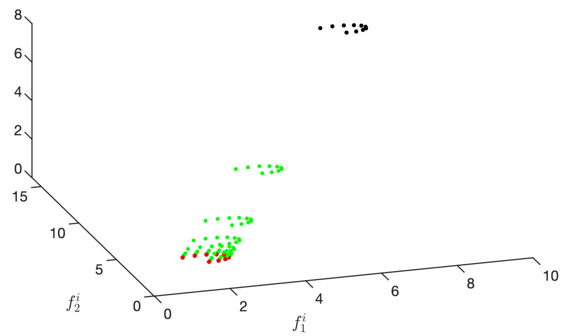
Example 6.4 *Consider the function* $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^3$ *defined as*

$$F(x) = \{f^1(x), f^2(x), \ldots, f^{10}(x)\},$$

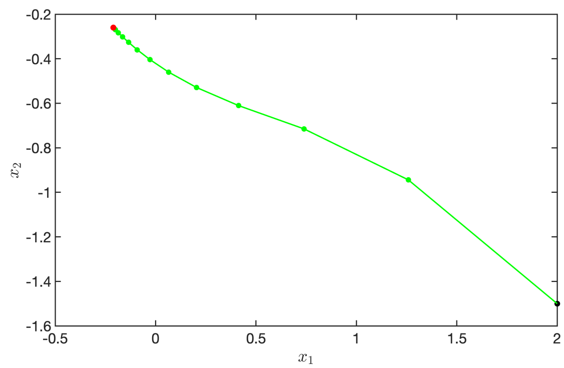
where for each $i \in [10]$, $f^i : \mathbb{R} \to \mathbb{R}^2$ *is given by*

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

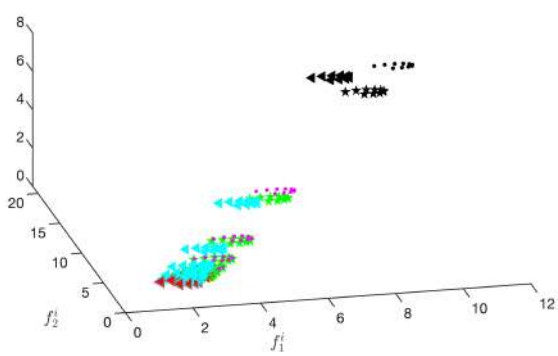
In Figure 6.4, the iterates generated by Algorithm 2 for different initial points taken



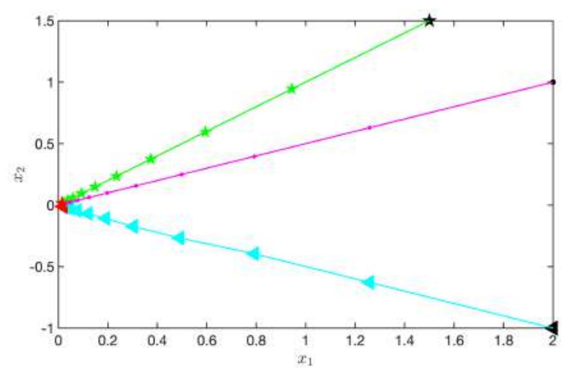
(a) The value of F at each iteration generated by Algorithm 2 for initial point $x_0 = (1.0000, -1.5000)^\top$ of Example 6.4



(b) The movement of subsequent x_k generated by Algorithm 2 for the initial point $x_0 = (1.0000, -1.5000)^\top$ of Example 6.4



(c) The value of F at each iteration generated by Algorithm 2 for three different randomly chosen initial points of Example 6.4



(d) TThe movement of subsequent x_k generated by Algorithm 2 for three different randomly chosen initial points of Example 6.4

Figure 6.4: Obtained output of Algorithm 2 for Example 6.4

from the set $[-4, 3] \times [-4, 3]$ *are given. The sequence of iterates* $\{x_k\}$ *and the corresponding* $\{F(x_k)\}$ *generated by Algorithm 2 for a selected initial point* $x_0 = (1.0000, -1.5000)^\top$ *are given in Figure 6.4(b) and Figure 6.4(a), respectively. Moreover, for three randomly* *selected initial points, the sequence of iterates* $\{x_k\}$ *and the corresponding* $\{F(x_k)\}$ *generated by Algorithm 2 are shown in Figure 6.4(d) and Figure 6.4(c), respectively.*

The performance of Algorithm 2 for Example 6.4 is shown in Table 6.6. Moreover, we have compared the results of the QNM with the SD method for set optimization as presented in Table 6.6. The values in Table 6.6 show that the proposed method performs better than the existing SD method.

Table 6.6: Performance of Algorithm 2 on Example 6.4

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 9, 8.0300, 8, 8, 0.2642)	(1.1032, 49.9087, 27.6729, 22.9076, 17, 4.4759)
	SD	(1, 10, 9.6300, 10, 10, 0.4852)	(33.9218, 31.0397, 32.0548, 28.6446, 28, 1.6261)

In the next two examples (Example 6.5 and Example 6.6), we consider a cone different from $\mathbb{R}^m_+$ and observe the performance of Algorithm 2. The Example 6.5 is a slight modification of Test instance 5.1 discussed in [1] with respect to a cone $\mathbb{R}^m_+$.

Example 6.5 *Consider the function* $F : \mathbb{R} \rightrightarrows \mathbb{R}^2$ *defined as*

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

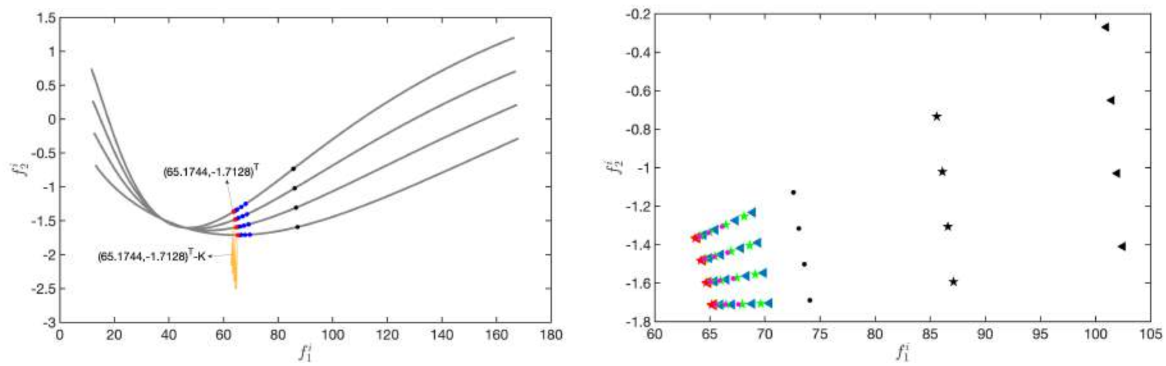
where for each $i \in [4], f^i : \mathbb{R} \to \mathbb{R}^2$ *is given by*

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

The cone is K *given by* $K = \{(z_1, z_2)^\top \in \mathbb{R}^2 : 6z_1 - 2z_2 \geq 0, -7z_1 + 10z_2 \geq 0\}$.

The output of Algorithm 2 for different initial points of Example 6.5 are depicted in Figure 6.5. The discretized segments represent the objective values that transverse a curve within the interval $[2.3350, 4.4010]$*. In Figure 6.5(a), we test our algorithm for an initial point* $x_0 = 4.0000$*. It can seen that the points depicted with red color are optimal points of* F *as the set* $(65.1744, -1.7128)^\top - K$ *does not contain any element* *of $F(x)$ other than $(65.1744, -1.7128)$ for all $x \in [2.3350, 4.4010]$. In Figure 6.5(b), we test our algorithm for three initial points and depict the output.*

The performance of Algorithm 2 for Example 6.5 is shown in Table 6.7. Moreover, we have compared the results of the QNM with the SD method for set optimization as presented in Table 6.7. The values in Table 6.7 show that the proposed method performs better than the existing SD method.



(a) The value of F at each iteration generated by Algorithm 2 for the initial point $x_0 = 4.3000$ of Example 6.5

(b) The value of F for three different initial points at each iteration generated by Algorithm 2 on Example 6.5

Figure 6.5: Obtained output of Algorithm 2 for Example 6.5

Table 6.7: Performance of Algorithm 2 on Example 6.5

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 6, 5.8900, 6, 6, 0.8275)	$(0.1013,\ 1.2640e+03,\ 1.0633e+03,\ 1.0639e+03,\ 24,\ 17.6970)$
	SD	(1, 8, 8.8900, 8, 8, 0.8275)	$(1.2113,\ 8.7830e+03,\ 8.6875e+03,\ 8.6873e+03,\ 36,\ 55.8729)$

Further, for the initial point $x_0 = 4.0000$, the decreasing behavior in the values of vector-valued functions at each iteration has been exhibited in Table 6.8.

Example 6.6 *Consider the set-valued function $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ defined as*

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

Table 6.8: Output of Algorithm 2 on Example 6.5 with $x_0 = 4.0000$

k	x_k	$f^1(x_k)$	$f^2(x_k)$	$f^3(x_k)$	$f^4(x_k)$
0	4.0000	(85.5982, −0.7345)	(86.0982, −1.0209)	(86.5982, −1.3073)	(87.0982, −1.5937)
1	3.7232	(68.1191, −1.2538)	(68.6191, −1.4047)	(69.1191, −1.5555)	(69.6191, −1.7064)
2	3.6932	(66.4532, −1.2981)	(66.9532, −1.4354)	(67.4532, −1.5727)	(67.9532, −1.7100)
3	3.6660	(64.9752, −1.3360)	(65.4752, −1.4613)	(65.9752, −1.5867)	(66.4752, −1.7120)
4	3.6503	(64.1331, −1.3569)	(64.6331, −1.4755)	(65.1331, −1.5941)	(65.6331, −1.7126)
5	3.6416	(63.6744, −1.3681)	(64.1744, −1.4830)	(64.6744, −1.5979)	(65.1744, −1.7128)
6	3.6371	(63.4378, −1.3738)	(63.9378, −1.4868)	(64.4378, −1.5998)	(64.9378, −1.7129)

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

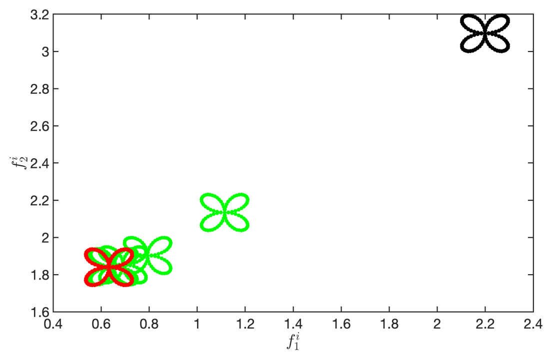
The output of Algorithm 2 for different initial points of Example 6.6 are depicted in Figure 6.5. The discretized segments represent the objective values that transverse a curve within the interval $[-\pi, \pi] \times [-\pi, \pi]$. Figure 6.6(a) exhibits the sequence $\{F(x_k)\}$ generated by Algorithm 2 for a chosen starting point, and Figure 6.6(b) tests Algorithm 2 for three different randomly chosen starting points.

The performance of Algorithm 2 for Example 6.6 is shown in Table 6.9. Moreover, we have compared the results of the QNM with the SD method for set optimization as presented in Table 6.9. The values in Table 6.9 show that the proposed method performs better than the existing SD method.

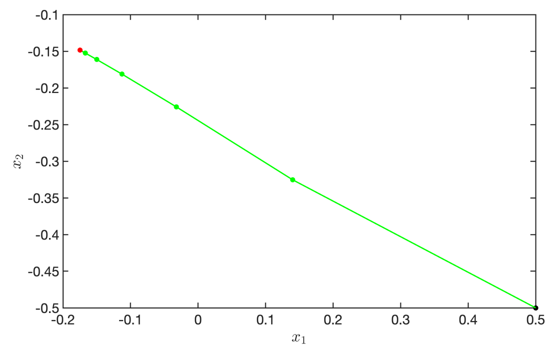
Table 6.9: Performance of Algorithm 2 on Example 6.6

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 33, 22.7200, 23, 29, 6.4418)	(3.2551, 67.4360, 45.8573, 46.2144, 3, 12.9461)
	SD	(1, 33, 22.4500, 22.5000, 25, 6.3776)	(1.2551, 63.9492, 43.9482, 44.1527, 3, 12.1495)

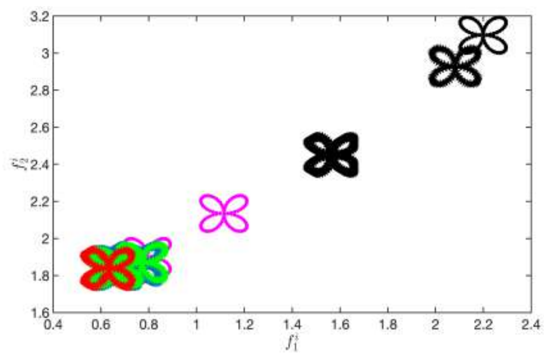
In the next example, we discuss the robust counterpart of a vector-valued facility location problem under uncertainty [115]. A complete discussion on this problem is given in [1].



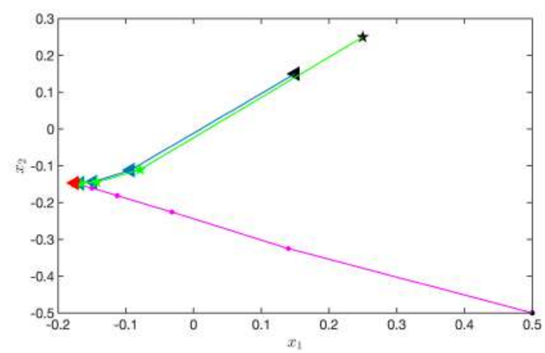
(a) The value of F at each iteration generated by Algorithm 2 for the initial point $x_0 = (0.5000, -0.5000)^\top$ of Example 6.6



(b) The movement of subsequent x_k generated by Algorithm 2 for the initial point $x_0 = (0.5000, -0.5000)^\top$ of Example 6.6



(c) The value of F for three different initial points at each iteration generated by Algorithm 2 for Example 6.6



(d) The movement of subsequent x_k generated by Algorithm 2 for the three different randomly chosen initial points of Example 6.6

Figure 6.6: Obtained output of Algorithm 2 for Example 6.6

Example 6.7 *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

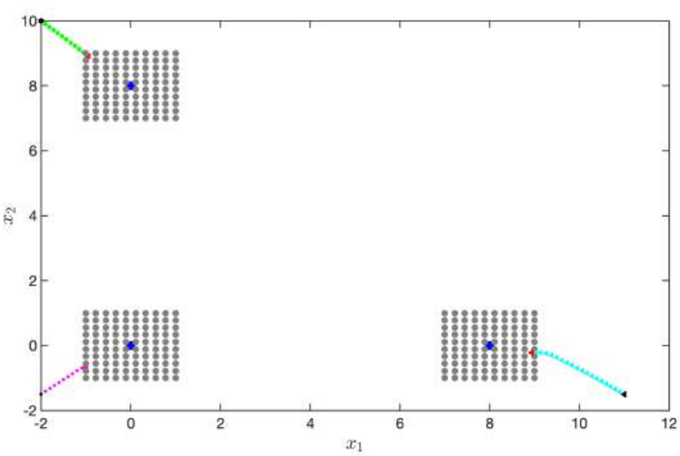
$$f^i(x) = \frac{1}{2}\begin{pmatrix} \|x - l_1 - u_i\|^2 \\ \|x - l_2 - u_i\|^2 \\ \|x - l_3 - u_i\|^2 \end{pmatrix},$$

where $l_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $l_2 = \begin{pmatrix} 8 \\ 0 \end{pmatrix}$ *and* $l_3 = \begin{pmatrix} 0 \\ 8 \end{pmatrix}$. *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* $\mathcal{U} \times \mathcal{U}$.

We generate the total of 100 *initial points in the square* $[-50, 50] \times [-50, 50]$ *as shown in Figure 6.7. The set* $(l_1 + u_i) \cup (l_2 + u_i) \cup (l_3 + u_i)$ *is represented by grey points and the locations of* l_1, l_2, l_3 *are depicted in blue color. The values of* $F(x_k)$ *generated by Algorithm 2 for three different randomly chosen initial points are given with cyan, magenta, and green colors as shown in Fig. 6.7.*



(a) The value of F in argument space at each iteration generated by Algorithm 2 for three randomly chosen initial points of Example 6.7

Figure 6.7: Obtained output of Algorithm 2 for Example 6.7

The performance of Algorithm 2 on Example 6.7 is shown in Table 6.10. A comparison of the results of QNM with the existing SD method is presented in Table 6.10. The *values in Table 6.10 show that the proposed method performs better than the existing SD method.*

Table 6.10: Performance of Algorithm 2 on Example 6.7

Number of initial points	Algorithm	Iterations (Min, Max, Mean, Median, Mode, SD)	CPU time (Min, Max, Mean, Median, ⌈Mode⌉, SD)
100	QNM	(1, 24, 11.4503, 6, 5123, 1.1421, 6.9534)	(1.2813, 19.2874, 4.002, , 1, 4.0954)
	SD	(1, 42, 12.1002, 4.2000, 3.2873, 13.2432)	(2.3412, 21.1771, 5.1217, 3.9272, 1, 4.2031)

For the initial point $x_0 = (11, -1.5)^\top$, the decreasing behavior in the values of vector-valued functions at each iteration has been exhibited in Table 6.11.

Table 6.11: Performance of Algorithm 2 on Example 6.7 with $x_0 = (-11, -1.5)^\top$

Iteration number (k)	$x_k^\top$	$f^{25}(x_k)$	$f^{50}(x_k)$	$f^{75}(x_k)$	$f^{100}(x_k)$
0	$(-11, -1.5)$	$(0.8169, 0.6536, 0.8581)$	$(0.8133, 0.6595, 0.8627)$	$(0.8008, 0.6148, 0.8481)$	$(0.7973, 0.6310, 0.8536)$
1	$(10.9017, -1.4198)$	$(0.8154, 0.6494, 0.8566)$	$(0.8117, 0.6552, 0.8612)$	$(0.7991, 0.6093, 0.8466)$	$(0.7955, 0.6261, 0.8521)$
2	$(10.8018, -1.3395)$	$(0.8138, 0.6451, 0.8552)$	$(0.8100, 0.6508, 0.8598)$	$(0.7974, 0.6034, 0.8450)$	$(0.7937, 0.6209, 0.8506)$
3	$(10.7001, -1.2589)$	$(0.8123, 0.6406, 0.8536)$	$(0.8084, 0.6462, 0.8583)$	$(0.7957, 0.5973, 0.8435)$	$(0.7918, 0.6155, 0.8490)$
4	$(10.5967, -1.1782)$	$(0.8107, 0.6359, 0.8521)$	$(0.8066, 0.6415, 0.8568)$	$(0.7940, 0.5908, 0.8418)$	$(0.7899, 0.6099, 0.8475)$
5	$(10.4915, -1.0973)$	$(0.8090, 0.6311, 0.8505)$	$(0.8049, 0.6365, 0.8552)$	$(0.7922, 0.5839, 0.8402)$	$(0.7880, 0.6041, 0.8459)$
6	$(10.3845, -1.0162)$	$(0.8074, 0.6260, 0.8489)$	$(0.8031, 0.6313, 0.8537)$	$(0.7904, 0.5767, 0.8385)$	$(0.7860, 0.5979, 0.8443)$
7	$(10.2756, -0.9349)$	$(0.8057, 0.6207, 0.8473)$	$(0.8012, 0.6258, 0.8521)$	$(0.7885, 0.5689, 0.8368)$	$(0.7840, 0.5915, 0.8426)$
8	$(10.1649, -0.8534)$	$(0.8039, 0.6152, 0.8456)$	$(0.7993, 0.6201, 0.8505)$	$(0.7866, 0.5606, 0.8350)$	$(0.7819, 0.5848, 0.8409)$
9	$(10.0565 - 0.7725)$	$(0.8022, 0.6096, 0.8439)$	$(0.7975, 0.6143, 0.8488)$	$(0.7848, 0.5520, 0.8333)$	$(0.7798, 0.5780, 0.8393)$
10	$(9.9454, -0.6912)$	$(0.8005, 0.6038, 0.8422)$	$(0.7956, 0.6082, 0.8472)$	$(0.7828, 0.5427, 0.8315)$	$(0.7777, 0.5707, 0.8375)$
11	$(9.8330, -0.6101)$	$(0.7987, 0.5977, 0.8405)$	$(0.7936, 0.6017, 0.8455)$	$(0.7809, 0.5326, 0.8297)$	$(0.7756, 0.5632, 0.8358)$
12	$(9.7218, -0.5290)$	$(0.7969, 0.5914, 0.8388)$	$(0.7917, 0.5950, 0.8439)$	$(0.7789, 0.5219, 0.8279)$	$(0.7734, 0.5554, 0.8341)$
13	$(9.6084, -0.4481)$	$(0.7951, 0.5849, 0.8370)$	$(0.7897, 0.5879, 0.8422)$	$(0.7769, 0.5101, 0.8260)$	$(0.7712, 0.5473, 0.8323)$
14	$(9.4937, -0.3672)$	$(0.7933, 0.5780, 0.8352)$	$(0.7877, 0.5804, 0.8404)$	$(0.7749, 0.4972, 0.8242)$	$(0.7690, 0.5388, 0.8305)$
15	$(9.3774 - 0.2963)$	$(0.7914, 0.5710, 0.8335)$	$(0.7856, 0.5728, 0.8388)$	$(0.7729, 0.4831, 0.8223)$	$(0.7667, 0.5309, 0.8288)$
16	$(9.2560, 0.2475)$	$(0.7895, 0.5634, 0.8319)$	$(0.7835, 0.5655, 0.8372)$	$(0.7707, 0.4674, 0.8207)$	$(0.7644, 0.5248, 0.8272)$
17	$(9.1363, -0.2200)$	$(0.7875, 0.5557, 0.8304)$	$(0.7815, 0.5589, 0.8359)$	$(0.7686, 0.4500, 0.8192)$	$(0.7621, 0.5209, 0.8259)$
18	$(9.0278, -0.2110)$	$(0.7857, 0.5484, 0.8292)$	$(0.7797, 0.5535, 0.8348)$	$(0.7666, 0.4322, 0.8180)$	$(0.7601, 0.5195, 0.8248)$
19	$(8.9339, -0.2151)$	$(0.7842, 0.5417, 0.8283)$	$(0.7781, 0.5495, 0.8340)$	$(0.7649, 0.4147, 0.8171)$	$(0.7584, 0.5200, 0.8241)$

6.7 Conclusion

In this chapter, we studied set optimization problems with respect to the lower set less relation. The objective mapping is given by a finite number of twice continuously differentiable functions. We have proposed a quasi-Newton method (Algorithm 2 to generate a sequence of iterates that converges to a weakly minimal solution of the problem. In the process of finding the weakly minimal solution for (SOP), we have approximated the Hessian matrices corresponding to given functions considered in Assumption 1 with the help of BFGS methods [128–131]. To generate the sequence, we have used a family of vector optimization problems (VOP). Then, for a suitably chosen element a_k from the partition set P_{x_k} of the current iterate x_k, we evaluate quasi-Newton direction u_k (*Step* 4) with the help of concepts in [133,137]. The process of generating iterates by Algorithm 2 continued until the stopping condition (*Step* 5) was met. We have discussed and ensured the well-definedness (Theorem 6.1) of Algorithm 2 with 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). For deriving the convergence analysis Algorithm 2, we have derived the following results.

(i) We have proved a condition of nonstationarity of a point (Proposition 6.1).

(ii) We analysed the boundedness of the sequence of generated Newton direction (Proposition 6.3).

(iii) We have derived the convergence of the generated sequence of iterates (Theorem 6.3) under a regularity condition (Definition 6.1).

(iv) We proved the superlinear convergence to a stationary point (Theorem 6.4) of the generated sequence under a regularity condition and uniform continuity of BFGS approximations and with the help of Lemma 6.5.

Finally, we tested the performance of the proposed quasi-Newton method on some existing and freshly introduced numerical test problems in Section 6.6. It is found that the proposed quasi-Newton method performed well when compared with the existing steepest descent method for strongly convex cases.

As a future reference, the proposed work can be tested for more practical problems similar to those discussed in [115]. In this chapter, we have used the lower set less ordering to compare the given sets. Research can be performed on other ordering relations also (given in [15]). To study the proposed quasi-Newton method on these relations, the usual derivative concepts like epiderivatives or coderivatives need separate attention. Moreover, in this chapter, we have used Armijo's step size condition to find the weakly minimal solution of (SOP). This work can be extended to different step size conditions, such as strong Wolfe or Armijo-Wolfe conditions, and a comparison of the performance of the method can be observed. Further, we have used the Gerstewitz scalarizing function for treating the set optimization problems (SOP). Future research can be performed on Hiriat-Urruty functional [120]. This involves the parameter set as the set of all compact generators of dual cone K^*. There are several conventional optimization methods in the literature that can be generated for set-valued optimization problems. A comparison between the performance of these methods can be analysed (see [123, 126, 138, 139] and references therein). Future research can focus on devising a quasi-Newton method whose convergent analysis may not require the used assumption of regular point; Use of a strong Wolfe line search instead of Armijo condition may be of great help in this direction, as observed in [140].

****