Chapter 6

A New Trust Region Method for Nonsmooth Multiobjective Optimization

6.1 Introduction

Generally, for moving from the current point to a new iterate, there are two fundamental strategies namely line search and trust region methods. The line search is based on searching a new iterate along a descent direction at each iteration, while trust-region methods are based on finding a new iterate within a ball centred at a current iteration. The trust-region methods are one of the most popular methods for solving smooth unconstrained optimization problems due to its convergence properties and robustness [120, 124]. The purpose of this chapter is to extend the trust-region method (from smooth to nonsmooth) to solve nonsmooth MOPs.

6.2 Motivation

Qu, Goh and Liang [26] introduced the following local quadratic model for locally Lipschitz functions:

$$Q_i(x_k, d) = \phi_i(x_k, d) + \frac{1}{2} d^\top B_i(x_k) d, \quad \text{for all } i \in \mathcal{I} = \{1, \}, \tag{6.1}$$

where at the k-th iteration $x_k \in \mathbb{R}^n$, $B_i(x_k)$ are given. In [26], the classical trust-region algorithm is extended for smooth MOPs to the nonsmooth case. Based on some assumptions on the functions ϕ_i, they established the global convergence property for the new method. However, they did not provide any specific choice for each function ϕ_i. In this article, we provide a choice for the functions ϕ_i in order to construct a new quadratic model for nonsmooth MOP, which is theoretically tractable and practically efficient. To be exact, for a given $\varepsilon > 0$, we substitute each function ϕ_i in (6.1) with a suitable element of Goldstein ε-subdifferential. We also present a practical way for (approximately) solving the newly proposed quadratic model.

6.3 Contributions

In this chapter, we introduce a new trust-region method to obtain the Pareto critical points for nonsmooth multiobjective optimization problems. The gradients and Hessians of the objective functions that are used when solving trust-region subproblem are approximated by using ϵ-subgradients and quasi-Newton methods, respectively. Using the BFGS updating formula for the Hessian approximation of the model, we show that the proposed algorithm is convergent under some mild and standard conditions on the objective functions.

The results of the proposed algorithm in this chapter are as follows.

(i) New local quadratic models corresponding to each objective functions are constructed.

(ii) The convergence property of the new proposed trust-region method is established under some suitable assumptions on the objective functions.

(iii) The implementaion of the proposed algorithm is tested on some examples.

In this paper, we deal with the following unconstrained nonsomooth MOP

$$\min_{x\in\mathbb{R}^n} f(x) = \left(f_1(x), f_2(x), \ldots, f_{\mathcal{P}}(x)\right)^\top, \tag{6.2}$$

where each objective functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz function for all $i \in \mathcal{I}$.

The directional derivative of f_i at x along the direction $d \in \mathbb{R}^n$ is defined as:

$$f_i'(x, d) = \lim_{\alpha\to 0} \frac{f_i(x + \alpha d) - f_i(x)}{\alpha}. \tag{6.3}$$

In real-life problems, it is common for objective functions to have conflicting nature, i.e., it is not possible to find a single solution that can minimize every objective functions simultaneously. For an nonsmooth MOP (6.2), the notion of optimality is Pareto optimality, as explained below:

Definition 6.1 *(*Pareto optimality *(e.g. [5])). A point $\hat{x} \in \mathbb{R}^n$ is called Pareto optimal for the problem* (6.2)*, if there is no $x \in \mathbb{R}^n$ such that $f(x) \leq f(\hat{x})$. The set of all Pareto optimal points of an nonsmooth MOP (6.2) is Pareto set.*

In practice, the optimality conditions are required to determine whether a point is Pareto optimal. For smooth MOPs, the optimality conditions have been derived via gradients of the objective functions. When objective functions are locally Lipschitz, the optimality conditions will depend on the concept of subdifferentials. Now, we provide some essential definitions and results that are generally used in nonsmooth analysis.

Definition 6.2 *(*Clarke subdifferential *(e.g. see [125])). Let $\Theta_i \in \mathbb{R}^n$ be the set of all nondifferentiable points of f_i, $i \in \mathcal{I}$. Then,*

$$\partial f_i(x) = conv(\{\mathcal{A} \in \mathbb{R}^n : \exists \langle x_l \rangle \in \mathbb{R}^n \setminus \Theta_i \text{ with } x_l \to x \text{ and } \nabla f_i(x_l) \to \mathcal{A} \text{ for } l \to \infty\})$$

is the Clarke-subdifferential of f_i at x.

It is clear from Definition 6.2 that when the objective functions are continuously differentiable then $\partial f_i(x) = \{\nabla f_i(x)\}$. Furthermore, the set $\partial f_i(x)$ is nonempty, convex and compact for all $i \in \mathcal{I}$ (see [125]). We denote the set

$$\Lambda(x) = \text{conv}\left(\bigcup_{i=1}^{\mathcal{P}} \partial f_i(x)\right). \tag{6.4}$$

A vector $d \in \mathbb{R}^n$ is a descent direction for f at $x \in \mathbb{R}^n$ if the following condition satisfies:

$$\mathcal{A}^\top d < 0 \text{ for all } \mathcal{A} \in \Lambda(x). \tag{6.5}$$

Definition 6.3 *(Goldstein subdifferential (e.g. see [126])). The (Goldstein) ε-subdifferential of f_i at x with some $\varepsilon \geq 0$ is defined as*

$$\partial_\varepsilon f_i(x) = conv\left(\bigcup_{y \in B_\varepsilon(x)} \partial f_i(y)\right), \quad i \in \mathcal{I}.$$

Note that $\partial f_i(x) \subseteq \partial_\varepsilon f_i(x)\ \forall\ \varepsilon \geq 0$ and $\partial_\varepsilon f_i(x)$ is a nonempty, convex and compact set [127]. We denote the set $\Lambda_\varepsilon(x) = \text{conv}\left(\bigcup_{i=1}^{\mathcal{P}} \partial_\varepsilon f_i(x)\right)$.

The following definition provides the necessary optimality condition when MOP (6.2) is locally Lipschitz.

Definition 6.4 *A point $x \in \mathbb{R}^n$ is said to be Pareto critical of the nonsmooth MOP (6.2) if $0 \in \Lambda(x)$. When working with ε-subdifferential rather than Clarke subdifferential, the Pareto critical condition will be $0 \in \Lambda_\varepsilon(x)$.*

The following theorem provides a way to compute descent direction for each objective function f_i.

Theorem 6.1 *Let* $\varepsilon \geq 0$ *and* $\bar{v} = argmin_{\mathcal{A}\in\Lambda_\varepsilon(x)}\|\mathcal{A}\|^2$. *Then, for all* $i \in \mathcal{I}$

$$f_i(x - \alpha\bar{v}) \leq f_i(x) - \alpha\|\bar{v}\|^2 \quad \forall \quad \alpha \leq \frac{\varepsilon}{\|\bar{v}\|}.$$

Proof: See [128]. □

6.3.1 Motivation

Qu, Goh and Liang [26] introduced the following local quadratic model for locally Lipschitz functions:

$$Q_i(x, d) = \phi_i(x, d) + \frac{1}{2}d^\top B_i(x)d, \quad \text{for all } i \in \mathcal{I} \tag{6.6}$$

and extended the classical trust-region algorithm for smooth MOPs to the nonsmooth case. Based on some assumptions on the functions ϕ_i, they established the global convergence property for the new method. However, they did not provide any specific choice for each function ϕ_i. In this article, we provide a choice for the functions ϕ_i in order to construct a new quadratic model for nonsmooth MOP (6.2), which is theoretically tractable and practically efficient. To be exact, for a given $\varepsilon > 0$, we substitute each function ϕ_i in (6.6) with a suitable element of $\Lambda_\varepsilon(x)$. We also present a practical way for (approximately) solving the newly proposed quadratic model.

Let $\varepsilon > 0$ be given. One can compute the descent direction by using $\Lambda_\varepsilon(x)$. For this purpose, let

$$\mathsf{v} = \arg\min_{\mathcal{A}\in\Lambda_\varepsilon(x)} \|\mathcal{A}\|^2, \tag{6.7}$$

and $d = -\frac{\mathsf{v}}{\|\mathsf{v}\|}$. Motivated by Theorem 6.1, we regard v as an acceptable descent direction, if

$$f_i(x + \varepsilon d) \leq f_i(x) - \varepsilon \|\mathsf{v}\| \quad \forall \quad i \in \mathcal{I}. \tag{6.8}$$

Therefore, d is descent direction for each objective function. Note that the structure of computation of $\Lambda_\varepsilon(x)$ makes the computation of v impractical. Consequently, set $\Lambda_\varepsilon(x)$ can be approximated by its certain finite subsets. Indeed, as it has mentioned in [128], if $\mathrm{W} \subset \Lambda_\varepsilon(x)$, then $\mathrm{conv}\,(\mathrm{W})$ can be considered as an approximation of $\Lambda_\varepsilon(x)$. Precisely, assume that

$$\tilde{\mathsf{v}} = \arg \min_{\mathcal{A} \in \mathrm{conv}(\mathrm{W})} \|\mathcal{A}\|^2 \tag{6.9}$$

and $d = -\frac{\tilde{\mathsf{v}}}{\|\tilde{\mathsf{v}}\|}$. If for all $i \in \mathcal{I}$,

$$f_i(x + \varepsilon d) - f_i(x) \leq -c\varepsilon\|\tilde{\mathsf{v}}\|, \text{ for some } c \in (0, 1), \tag{6.10}$$

then $\tilde{\mathsf{v}}$ can be considered as an approximation of the descent direction. Otherwise, a new ε-subgradient $\mathcal{A}' \in \Lambda_\varepsilon(x)$ is calculated such that $\mathrm{W} \cup \{\mathcal{A}'\}$ yields a better descent direction. The following theorem guarantees that for each $i \in \mathcal{I}$, there exists some $\ell \in \left(0, \frac{\varepsilon}{\|\tilde{\mathsf{v}}\|}\right]$ such that $\mathcal{A}' \in \partial f_i(x + \ell\tilde{\mathsf{v}})$ improves the approximation of $\Lambda_\varepsilon(x)$.

Theorem 6.2 *Let $c \in (0, 1)$, $W = \{\xi_1, \xi_2, \ldots, \xi_m\} \subset \Lambda_\varepsilon(x)$ and $\tilde{\mathsf{v}}$ be the solution of (6.9). If $f_i\left(x + \frac{\varepsilon}{\|\tilde{\mathsf{v}}\|}\tilde{\mathsf{v}}\right) > f_i(x) - c\varepsilon\|\tilde{\mathsf{v}}\|$ for some $i \in \mathcal{I}$, then there is some $\ell' \left(0, \frac{\varepsilon}{\|\tilde{\mathsf{v}}\|}\right]$ and $\mathcal{A}' \in \partial f_i(x + \ell'\tilde{\mathsf{v}})$ such that $\tilde{\mathsf{v}}^\top \mathcal{A}' > -c\|\tilde{\mathsf{v}}\|^2$.*

Proof: See [128]. □

The following algorithm describes a procedure for finding a descent direction. It has been demonstrated that Algorithm 8 terminates after a finite number of iterations under some standard assumptions(see [128]).

Algorithm 8 A method to compute descent direction

Aim: To compute a direction such that (6.8) is satisfied for each $f_i,\ i \in \mathcal{I}$

Input: Current point $x \in \mathbb{R}^n$, tolerances $\varepsilon,\ \hat{\delta} > 0$ and $c \in (0,1)$

1: Compute $\mathcal{A}_1^i \in \partial_\varepsilon f_i(x)$ for all $i \in \mathcal{I}$

2: Set $\mathrm{W}_1 = \{\mathcal{A}_1^1, \mathcal{A}_1^2, \ldots, \mathcal{A}_1^{\mathcal{P}}\}$ and $\jmath = 1$

3: Compute

$$v_\jmath = \arg \min_{v \in \mathrm{conv}(\mathrm{W}_\jmath)} \|v\|^2$$

4: If $\|v_\jmath\| < \hat{\delta}$ then stop otherwise set $\hat{d} = -\frac{v_\jmath}{\|v_\jmath\|}$ go to Step 5

5: Find the objective functions for which the condition (6.8) is not satisfied:

$$\gimel_\jmath = \left\{ i \in \mathcal{I} : f_i\left(x + \varepsilon \hat{d}\right) > f_i(x) - c\varepsilon\|v_\jmath\| \right\}$$

6: If $\gimel_\jmath$ is empty then stop otherwise set go to Step 7

7: For each $i \in \gimel_\jmath$, compute $\ell \in (0, \frac{\varepsilon}{\|v_\jmath\|})$ and $\mathcal{A}_\jmath^i \in \partial_\varepsilon f_i(x - \ell v_\jmath)$ such that

$$\langle v_\jmath, \mathcal{A}_\jmath^i \rangle < c\|v_\jmath\|^2$$

8: Set $\mathrm{W}_{\jmath+1} = \mathrm{W}_\jmath \cup \{\mathcal{A}_\jmath^i : i \in \gimel_\jmath\}$

9: Set $\jmath \leftarrow \jmath + 1$ and go to Step 1

Algorithm 8 has the following properties:

1. it terminates after a finite number of iterations (see [128]),

2. if $\tilde{\mathsf{v}}$ is the descent direction obtained by Algorithm 8 and $\bar{d} = -\frac{\tilde{\mathsf{v}}}{\|\tilde{\mathsf{v}}\|}$ then either $\|\tilde{\mathsf{v}}\| < \hat{\delta}$ or it satisfies $f_i\left(x + \varepsilon \bar{d}\right) \leq f_i(x) - c\varepsilon\|\tilde{\mathsf{v}}\| \ \ \forall\, i \in \mathcal{I}$.

Suppose that Algorithm 8 returns $\tilde{\mathsf{v}}$ with $\|\tilde{\mathsf{v}}\| \geq \delta$. Then, we define the local quadratic models associated to each f_i as follows:

$$Q_i(x,d) = \tilde{\mathsf{v}}^\top d + \frac{1}{2} d^\top B_i(x) d, \quad i \in \mathcal{I},$$

where $B_i(x)$'s are the positive definite matrices. Now, for $x \in \mathbb{R}^n$, we define the trust region subproblem as follows:

$$\left.\begin{array}{ll}\text{minimize} & \max_{i\in\mathcal{I}} Q_i(x,d) = \tilde{\mathsf{v}}^\top d + \frac{1}{2} d^\top B_i(x) d \\ \text{subject to} & \|d\| \le \Delta.\end{array}\right\} \tag{6.11}$$

As the problem (6.11) is nonsmooth, It can be equivalently framed as the following convex quadratic optimization problem:

$$\left.\begin{array}{ll}\text{minimize} & t \\ \text{subject to} & \tilde{\mathsf{v}}^\top d + \frac{1}{2} d^\top B_i(x) d \le t, \quad i \in \mathcal{I} \\ & \|d\| \le \Delta,\end{array}\right\} \tag{6.12}$$

where $\tilde{\mathsf{v}}$ is the output of Algorithm 8 at the point x with $\varepsilon = \Delta$ and the matrices $B_i(x)$ are positive definite for all $i \in \mathcal{I}$.

In the next section, we provide the structure of the new proposed trust-region method.

6.4 Trust-region algorithm and its convergence analysis

In this section, we introduce a trust-region algorithm for obtaining critical points of MOP (6.2). Suppose that, at the point x_k, the set $\mathrm{W}_k \subset \Lambda_\varepsilon(x_k)$ is an approximation of $\Lambda_\varepsilon(x_k)$. Let

$$\tilde{\mathsf{v}}_k = \arg \min_{\mathcal{A}\in \mathrm{conv}(\mathrm{W}_k)} \|\mathcal{A}\|^2,$$

and assume that for all $i \in \mathcal{I}$, $f_i\left(x_k + \varepsilon \frac{\tilde{\mathsf{v}}_k}{\|\tilde{\mathsf{v}}_k\|}\right) - f_i(x) \le -c\varepsilon\|\tilde{\mathsf{v}}\|$, for some $c \in (0,1)$. Now, the local quadratic model corresponding to each objective function is defined by

$$Q_i(x_k, d) = \tilde{\mathsf{v}}_k^\top d + \frac{1}{2} d^\top B_i(x_k) d, \quad i \in \mathcal{I},$$

where each $B_i(x_k)$ is positive definite matrices. Based on this model, we outline the structure of the new proposed trust-region method for minimizing MOP (6.2).

Algorithm 9 Trust-region algorithm to solve an MOP (6.2)

Step 0: Given $x_0 \in \mathbb{R}^n$, Δ_0, $tol > 0$, $\tau_\Delta, \delta_1, \tau_\delta \in (0,1)$, $c_1 \in (0,1)$, $0 < c_2 < c_3 < 1$, $0 < \varsigma_1 < \varsigma_2 < 1$, $\varsigma_3 > 1$.
Step 1: Apply Algorithm 8 at x_k with $\varepsilon = \Delta_k$, $\delta = \delta_k$ and $c = c_1$. This algorithm terminates after a finite number of iterations (see [128]) and as a result, we get $\tilde{\mathrm{v}}_k$, either $\|\tilde{\mathrm{v}}_k\| < \delta_k$ or for the search direction $\bar{d}_k = -\frac{\tilde{\mathrm{v}}_k}{\|\tilde{\mathrm{v}}_k\|}$,

$$f_i\left(x + \Delta_k \bar{d}_k\right) - f_i(x_k) \leq -c\Delta_k \|\tilde{\mathrm{v}}_k\| \quad \forall\ i \in \mathcal{I}.$$

Step 2: If $\|\tilde{\mathrm{v}}_k\| < tol$, then stop. If $\|\tilde{\mathrm{v}}_k\| \leq \delta_k$, then set $\Delta_{k+1} = \tau_\Delta \Delta_k$, $x_{k+1} = x_k$, $\delta_{k+1} = \tau_\delta \delta_k$, $k \leftarrow k+1$ and go to Step 1. Else, set $\delta_{k+1} = \delta_k$ and go to Step 3.
Step 3: Solve the following subproblem:

$$\begin{aligned} &\text{minimize} && t \\ &\text{subject to} && \tilde{\mathrm{v}}_k^\top d + \frac{1}{2} d^\top B_i(x_k) d \leq t, \quad i \in \mathcal{I} \\ &&& \|d\| \leq \Delta_k, \end{aligned}$$

and let (d_k, t_k) be its approximate solution.
Step 4: If for all $i \in \mathcal{I}$, $f_i(x_k + d_k) - f_i(x_k) \leq c_1 \tilde{\mathrm{v}}_k^\top d_k$, then go to Step 5. Else set $\Delta_{k+1} = \tau_\Delta \Delta_k$, $x_{k+1} = x_k$, $k \leftarrow k+1$ and go to Step 1.
Step 5: Set

$$\rho_k^i = \frac{f_i(x_k) - f_i(x_k + d_k)}{-Q_i(x_k, d_k)}.$$

Step 6: If $\rho_k^i \geq c_2 \ \ \forall\ i$, then set $x_{k+1} = x_k + d_k$; otherwise set $x_{k+1} = x_k$ and update the trust region radius as follows:

$$\Delta_{k+1} = \begin{cases} \varsigma_3 \Delta_k & \text{if } c_3 \leq \rho_k^i \ \forall\ i \\ \varsigma_2 \Delta_k & \text{if } c_2 \leq \rho_k^i \ \text{ and } \ \exists\ l \text{ such that } \rho_k^l < c_3 \\ \varsigma_1 \Delta_k & \text{if } \exists\ l \text{ such that } \rho_k^l \leq c_2 \end{cases}$$

Step 8: Update matrices $B_i(x_k)$'s by BFGS formula. Set $k \leftarrow k+1$ and go to Step 1.

In Algorithm 9, k-th iteration is considered *very successful* if $\rho_k^i \geq c_3$ for all $i \in \mathcal{I}$ and in this case, we increase the trust region radius, i.e., $\Delta_{k+1} = \varsigma_3 \Delta_k$. Also, the k-th iteration is considered *successful* if $\rho_k^i \geq c_2$ for all $i \in \mathcal{I}$ and there are some indexes l such that $\rho_k^l \leq c_3$. In this case, we update the trust region radius by taking $\Delta_{k+1} = \varsigma_2 \Delta_k$. In another case, the k-th iteration is said to be *unsuccessful* and we decrease the trust region radius, i.e., $\Delta_{k+1} = \varsigma_1 \Delta_k$. For the purpose of further analysis of the proposed algorithm, we consider the set of indices of the successful iterations defined by

$$\mathcal{S} = \{k \in \mathbb{N} : \rho_k^i \geq c_2 \text{ for all } i \in \mathcal{I}\}.$$

Note that if Algorithm 8 returns $\tilde{\mathrm{v}}_k$ with $\varepsilon = \Delta_k$ and $\|\tilde{\mathrm{v}}_k\| > \delta$, then for each $i \in \mathcal{I}$, we have $f_i\left(x_k + \Delta_k \bar{d}_k\right) - f_i(x_k) \leq -c\Delta_k \|\tilde{\mathrm{v}}_k\|$, for some constant $c \in (0,1)$ and $\bar{d}_k = -\frac{\tilde{\mathrm{v}}_k}{\|\tilde{\mathrm{v}}_k\|}$. Hence, in this case, a sufficient reduction is achieved in each objective function.

The bound for the reduction predicted by a quadratical model Q_i at x_k in a descent direction for f is given by in the following lemma.

Lemma 6.1 *Let $\tilde{\mathsf{v}} > 0$ is obtained by Algorithm 8 at x_k, i.e., x_k is not Pareto critical point. Then, for each $i \in \mathcal{I}$, there exists $\bar{\ell}_i > 0$ such that $\|\bar{\ell}_i \tilde{\mathsf{v}}\| \leq \Delta_k$ and*

$$Q_i(x_k, -\bar{\ell}_i \tilde{\mathsf{v}}) \leq Q_i(x_k, -\ell \tilde{\mathsf{v}}), \quad \forall\ \ell > 0 \ \text{ such that } \|\ell \tilde{\mathsf{v}}\| \leq \Delta_k. \tag{6.13}$$

Furthermore,

$$Q_i(x_k, 0) - Q_i(x_k, -\bar{\ell}_i \tilde{\mathsf{v}}) \geq \frac{1}{2}\|\tilde{\mathsf{v}}\| \min\left\{\frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k\right\},$$

where M is such that $\|\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}\| \leq M$ with $B_i(x_k)$ positive definite matrix for all $i \in \mathcal{I}$.

Proof: Consider the function $\varphi_i : \left[0, \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\right] \to \mathbb{R}$ defined as:

$$\varphi_i(\ell) = Q_i(x_k, -\ell \tilde{\mathsf{v}}).$$

The function ϕ_i is continuous in a compact set $\left[0, \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\right]$, then there exists $\bar{\ell}_i \in \left[0, \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\right]$ such that

$$Q_i(x_k, -\bar{\ell}_i \tilde{\mathsf{v}}) \leq Q_i(x_k, -\ell \tilde{\mathsf{v}}), \quad \forall\ \ell > 0 \ \text{ such that } \|\ell \tilde{\mathsf{v}}\| \leq \Delta_k.$$

Also, $\bar{\ell}_i > 0$ as $\varphi_i'(0) = -\|\tilde{\mathsf{v}}\|^2 < 0$ (because $\tilde{\mathsf{v}}$ is descent direction for f at x_k).
In the second part, for each quadratic model the direction $\tilde{\mathsf{v}}$ is such that

$$\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}} > 0 \ \text{ for all } i \in \mathcal{I}.$$

The minimizer of the function $\varphi_i(\ell) = Q_i(x_k, -\ell \tilde{\mathsf{v}})$ is $\hat{\ell}_i = \frac{\|\tilde{\mathsf{v}}_k\|^2}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}} > 0$. Now, if the minimizer is inside the trust region, i.e., $\|\hat{\ell}_i \tilde{\mathsf{v}}\| \leq \Delta_k$, we take $\bar{\ell}_i = \hat{\ell}_i$, then

$$\begin{aligned}Q_i(x_k, -\bar{\ell}_i\tilde{\mathsf{v}}) =& -\frac{\|\tilde{\mathsf{v}}\|^2}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}}\|\tilde{\mathsf{v}}\|^2 + \frac{1}{2}\left(\frac{\|\tilde{\mathsf{v}}\|^2}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}}\right)^2 \tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}\\ =& -\frac{1}{2}\frac{\|\tilde{\mathsf{v}}_k\|^4}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}}.\end{aligned}$$

Now, using the assumption $\|d^\top B_i(x_k)d\| \leq M \quad \forall\, i \in \mathcal{I}$ and the fact $Q_i(x_k, 0) = 0$, we obtain

$$Q_i(x_k, 0) - Q_i(x_k, -\bar{\ell}_i\tilde{\mathsf{v}}) = \frac{1}{2}\frac{\|\tilde{\mathsf{v}}\|^4}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}} \geq \frac{1}{2}\frac{\|\tilde{\mathsf{v}}\|^4}{\|\tilde{\mathsf{v}}\|^2 M} = \frac{1}{2}\frac{\|\tilde{\mathsf{v}}\|^2}{M}. \tag{6.14}$$

On the other hand, if $\|\hat{\ell}_i\tilde{\mathsf{v}}\| > \Delta_k$, then $\hat{\ell}_i$ is taken as $\hat{\ell}_i = \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}$, then

$$Q_i(x_k, -\hat{\ell}_i\tilde{\mathsf{v}}) = -\frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\|\tilde{\mathsf{v}}_k\|^2 + \frac{1}{2}\left(\frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\right)^2 \tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}.$$

From $\|\hat{\ell}_i\tilde{\mathsf{v}}\| > \Delta_k$, we have $\frac{\|\tilde{\mathsf{v}}\|^2}{\tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}}\|\tilde{\mathsf{v}}\| = \|\hat{\ell}_i\tilde{\mathsf{v}}\| > \Delta_k = \bar{\ell}_i\|\tilde{\mathsf{v}}\|$. Hence,

$$\frac{\|\tilde{\mathsf{v}}\|^2}{\bar{\ell}_i} > \tilde{\mathsf{v}}^\top B_i(x_k)\tilde{\mathsf{v}}.$$

Therefore, using the last inequality, we have

$$\begin{aligned}Q_i(x_k, 0) - Q_i(x_k, -\bar{\ell}_i\tilde{\mathsf{v}}) >& \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\|\tilde{\mathsf{v}}\|^2 - \frac{1}{2}\left(\frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\right)^2 \frac{\|\tilde{\mathsf{v}}_k\|^2}{\bar{\ell}_i}\\ =& \frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\|\tilde{\mathsf{v}}\|^2 - \frac{1}{2}\frac{\Delta_k}{\|\tilde{\mathsf{v}}\|}\|\tilde{\mathsf{v}}\|^2\\ =& \frac{1}{2}\Delta_k\|\tilde{\mathsf{v}}\|.\end{aligned} \tag{6.15}$$

Summarizing (6.14) and (6.15), we have

$$Q_i(x_k, 0) - Q_i(x_k, -\bar{\ell}_i\tilde{\mathsf{v}}) \geq \frac{1}{2}\|\tilde{\mathsf{v}}\| \min\left\{\frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k\right\}.$$

□

Note 6.1 *Lemma 6.1 is valid for each $f_i,\ i \in \mathcal{I}$. Moreover, if d_k be the solution of the trust-region subproblem (6.12) at the kth iteration, then the reduction obtained in the quadratic models can be bounded below using the bound given in Lemma 6.1 (see [25]):*

$$Q_i(x_k, 0) - Q_i(x_k, d_k) \geq \frac{1}{2}\|\tilde{\mathsf{v}}\| \min\left\{\frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k\right\} > 0 \quad \forall\ i \in \mathcal{I}. \tag{6.16}$$

It is important to observe that the bound (6.16) is greater than zero then the ratios ρ_k^i in Step 5 of Algorithm 9 is well defined.

The next result establishes the conditions to accept a step.

Lemma 6.2 *If d_k is the solution of (6.12) such that $\|\tilde{\mathsf{v}}_k\| \neq 0$ and*

$$\Delta_k \leq \frac{\|\tilde{\mathsf{v}}_k\|}{M} \quad \textit{for all } i \in \mathcal{I}, \tag{6.17}$$

then the k-th iteration is successful and $\Delta_{k+1} \geq \Delta_k$.

Proof: By using (6.17), we can write the bounds of quadratic models as follows:

$$Q_i(x_k, 0) - Q_i(x_k, d_k) \geq \frac{1}{2}\|\tilde{\mathsf{v}}\|\Delta_k. \tag{6.18}$$

On the other hand

$$\begin{aligned}
1 - \rho_k^i &= 1 - \frac{f_i(x_k) - f_i(x_k + d_k)}{-Q_i(x_k, d_k)} \\
&= \frac{-f_i(x_k) + f_i(x_k + d_k) - Q_i(x_k, d_k)}{-Q_i(x_k, d_k)} \\
&\leq \frac{-f_i(x_k) + c_1 \tilde{\mathsf{v}}_k^\top d_k + f_i(x_k) - \tilde{\mathsf{v}}_k^\top d_k - \frac{1}{2} d_k^\top B_i(x_k) d_k}{\frac{1}{2}\|\tilde{\mathsf{v}}\| \min\left\{\frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k\right\}} \\
&\leq \frac{(1 - c_1)\|\tilde{\mathsf{v}}\|\|d_k\|}{\frac{1}{2}\|\tilde{\mathsf{v}}\|\Delta_k} \\
&\leq 2(1 - c_1).
\end{aligned}$$

If $c_1 \geq c_3$, then we obtain $\rho_k^i \geq c_3 \ \ \forall\, i \in \mathcal{I}$. Thus, from Step 6 of Algorithm 9, the step is accepted and trust-region radius is increased, i.e., $\Delta_{k+1} \geq \Delta_k$. □

Note 6.2 *If x_k is not Pareto critical, i.e., $\tilde{\mathsf{v}}_k = \min\{\|\mathcal{A}\| : \mathcal{A} \in W_{\Delta_k}(x_k)\} > 0$. Therefore, after a finite number of inner iterations, in which the step is rejected and trust radius reduced, the inequality (6.17) will be satisfied and the step accepted.*

Lemma 6.3 *Assuming that there exists $\delta > 0$ such that $\|\tilde{\mathsf{v}}_k\| > \delta \ \ \forall\, k$, then there exists $\Delta_{\min}$ so that $\Delta_k > \Delta_{\min}$ for all k.*

Proof: Consider a subsequence of trust-region radius $\langle \Delta_{k_l} \rangle$ such that $\Delta_{k_l} \to 0$, where kth iteration is the first one which satisfies the following condition:

$$\Delta_{k+1} \leq \frac{\varsigma_1 \delta}{M}. \tag{6.19}$$

Since k is the first index satisfying the inequality $\varsigma_1 \Delta_k \leq \Delta_{k+1}$, we have

$$\Delta_k \leq \frac{\delta}{M} \leq \frac{\|\tilde{\mathsf{v}}_k\|}{M}.$$

Then, from Lemma 6.2, the previous iteration was very successful and radius must increase, which is a contradiction with the assumption that k is the first index that satisfies (6.19). Therefore, there does not exist k which satisfies (6.19) so that

$$\Delta_{k+1} > \frac{\varsigma_1 \delta}{M} = \Delta_{\min} \ \forall\, k.$$

□

Theorem 6.3 *If Algorithm 9 does not terminate after a finite number of iterations, then $\lim_{k\to\infty, k\in\mathcal{S}} \inf \|\tilde{\mathsf{v}}_k\| = 0$, where $\mathcal{S}$ be the index set of successful iterations.*

Proof: At the contrary, let us assume that there exists $\epsilon > 0$, so that

$$\|\tilde{\mathsf{v}}_k\| \geq \epsilon, \quad \text{for all } k \in \mathcal{S}. \tag{6.20}$$

For $k \in \mathcal{S}$, we have

$$\begin{aligned} f_i(x_k) - f_i(x_{k+1}) &\geq c_2 \left(Q_i(x_k, 0) - Q_i(x_k, d_k) \right) \\ &\geq c_2 \frac{1}{2} \|\tilde{\mathsf{v}}\| \min \left\{ \frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k \right\} \\ &\geq c_2 \frac{1}{2} \epsilon \min \left\{ \frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k \right\}. \end{aligned}$$

Summing over all the successful iterations up to the k-th index,

$$\begin{aligned} f_i(x_0) - f_i(x_{k+1}) &= \sum_{l=0, l \in \mathcal{S}}^{k} \left(f_i(x_l) - f_i(x_{l+1}) \right) \\ &\geq \sum_{l=0, l \in \mathcal{S}}^{k} c_2 \frac{1}{2} \epsilon \min \left\{ \frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k \right\} \\ &\geq \sigma_k c_2 \frac{1}{2} \epsilon \min \left\{ \frac{\|\tilde{\mathsf{v}}\|}{M}, \Delta_k \right\}, \end{aligned}$$

where σ_k is the number of successful iterations up to the k-th. Then if $\lim_{k \to \infty, k \in \mathcal{S}} \sigma_k = \infty$, the difference $f_i(x_0) - f_i(x_{k+1})$ goes to infinite for (6.20) and f_i will not be bounded. Therefore, there does not exist $\varepsilon > 0$ that satisfies (6.20) for some $i \in \mathcal{I}$. □

Theorem 6.4 *Let x_k be generated by Algorithm 9 and assume that the level sets of the functions f_i, $i \in \mathcal{I}$ are bounded, then x_k converges to a Pareto critical point.*

Proof: Set $A = \{x \in \mathbb{R}^n : f_i(x) \leq f_i(x_0),\ i \in \mathcal{I}\}$. Due to the fact that the level sets of f_i, $i \in \mathcal{I}$ are bounded. Furthermore, the set A is closed because f_i, $i \in \mathcal{I}$ are continuous functions. Then A is compact. Therefore, the sequence x_k generated by Algorithm 9 has an accumulation point x^*. □

6.5 Numerical tests

In this section, we focus on providing some computational results of Algorithm 9 on some examples. Each example has been considered with different initial points generated at random. The algorithm has been coded in MATLAB 2020a. The code has been executed on a personal computer with a processor INTEL Core i5 with 8GB RAM.

For Algorithm 9, we set the parameters as follows: $\Delta_0 = 0.5,\ \tau_\Delta = 0.8.\ \delta_1 = 0.7,\ \tau_\delta = 0.6,\ c_1 = 0.1,\ c_2 = 0.2,\ c_3 = 0.95.$

Example 1 *Consider the following locally Lipschitz biobjective problem*

$$\min_{x\in\mathbb{R}^2} \left((x_1-1)^2 + (x_2-1)^2, x_1^2 + |x_2|\right)^\top . \tag{6.21}$$

The objective region (in magenta color) and obtained Pareto critical points (in blue color) by Algorithm 9 of the problem (6.21) is shown in Figure 6.1.

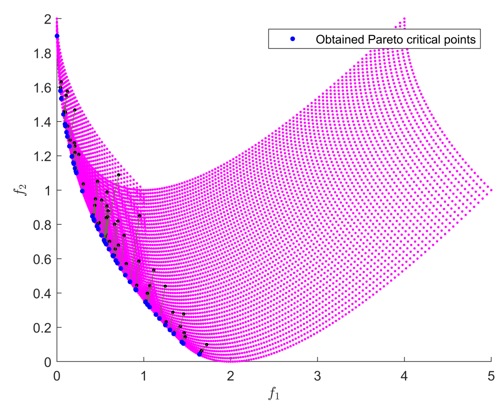

Figure 6.1: Obtained Pareto critical points by Algorithm 9

Example 2 *Consider the following biobjective optimization problem*

$$\min_{x\in\mathbb{R}^2} \left((x_1-1)^2 + (x_2-1)^2, |x_2 - 10|x_1|| + 0.5x_2\right)^\top . \tag{6.22}$$

The objective region (in magenta color) and obtained Pareto critical points (in blue color) by Algorithm 9 of the problem (6.22) is shown in Figure 6.2.

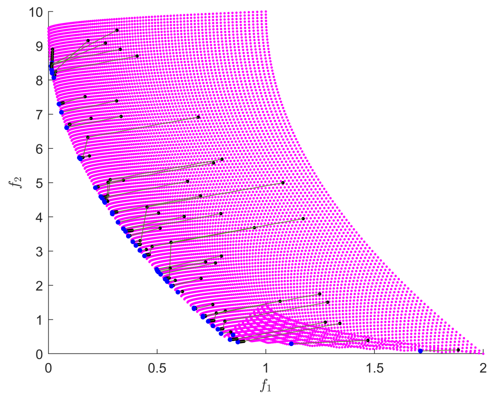

Figure 6.2: Obtained Pareto critical points by Algorithm 9

6.6 Conclusion

In this chapter, we have developed a nonsmooth trust-region method for unconstrained MOPs. Inspired by the work presented in [128], which computes efficiently an approximation of the descent direction after fewer number of subgradients, we construct practical local quadratic models in the proposed trust-region method for locally Lipschitz functions. The main result is the global convergence property of the new proposed trust-region method, which is established under some suitable assumptions on the objective functions. The new proposed trust-region method is tested on some nonsmooth examples.

****