manuscript No.
(will be inserted by the editor)

Complexity bound of trust-region methods for convex smooth unconstrained multiobjective optimization

R. Garmanjani

Received: date / Accepted: date

Abstract In this paper, we analyze the worst-case complexity of trust-region methods for solving unconstrained smooth multiobjective optimization problems. We particularly focus on the method proposed by Villacorta et al. [J. Optim. Theory Appl., 160:865–889]. When the component functions are convex, we will derive a complexity bound of $\mathcal{O}(\epsilon^{-1})$ for driving some criticality measure below some given positive ϵ. The derived complexity bound recovers the complexity bound of classical trust-region methods for solving convex smooth unconstrained single-objective problems.

Keywords trust-region methods · multiobjective optimization · worst-case complexity · convex smooth unconstrained

Mathematics Subject Classification (2020) 90C25 · 90C29 · 90C30 · 90C60

1 Introduction

In multiobjective optimization, in contrast to single-objective optimization, one needs to minimize more than one possibly conflicting objectives at the same time. These kind of optimization problems appear in many applications in science, engineering, finance, and economic, among others. There are various approaches and a vast number of methods in literature which have been devoted to solving multiobjective optimization problems. An extensively studied approach for solving a multiobjective optimization problem is the so-called

R. Garmanjani
Department of Mathematics, FCT-UNL-CMA, Campus de Caparica, 2829-516 Caparica, Portugal.
Support for this author was provided by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) under the projects PTDC/MAT-APL/28400/2017 and UIDB/00297/2020. E-mail: r.garmanjani@fct.unl.pt

scalarization approach in which a single (and sometimes more than one) objective, made by, for instance, a weighted sum of all the objectives, is minimized. The weighting parameters are set by the optimizer based on his/her preference in advance or are chosen adaptively as the algorithm proceeds. Further details on such approach can been found in [1–3]. Another approach is based on the extension of single-objective optimization algorithms to multiobjective setting, and tries to find a descent direction for all the component functions at the same time, by solving a subproblem at each iteration. These methods do not use a set of parameters in advance for converting the problem to a single-objective optimization. They include steepest descent [4], Newton [5], trust-region methods [6–9], among others. Further information on such methods can be found in the survey [10] and the references therein.

In this paper, we consider unconstrained multiobjective optimization problems of the form

$$\min_{x \in \mathbb{R}^n} F(x) = (f_1(x), \ldots, f_m(x)), \tag{1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I := \{1, \ldots, m\}$, is a smooth function. For this class of optimization problems, under the convexity assumption of the component functions, we are interested in analyzing the worst-case complexity (WCC) of a trust-region method for driving the criticality measure below some given positive threshold. The WCC of an algorithm measures the cost of running the algorithm, which is expressed in terms of number of iterations or function evaluations, in the worst-case scenario when starting from any initial point and arriving at a point for which the stationary measure is small enough.

Hereafter, when we refer to nonconvex and convex case, we mean all the component functions in (1) are nonconvex and convex, respectively.

By setting $m = 1$ in (1), we will have an unconstrained single-objective optimization problem. The WCC analysis of optimization algorithms, when the goal is to find a point at which the norm of the gradient becomes less than the given positive threshold, has been the subject of many research studies in single-objective optimization of which we will review the ones more related to the current study. For nonconvex smooth unconstrained optimization problems, a WCC bound of $\mathcal{O}(\epsilon^{-2})$ has been derived for steepest descent methods [11] [see Example 1.2.3] and trust-region methods [12]. The same complexity bound has been shown to hold for adaptive regularization with cubic (ARC) [13]. A complexity bound of $\mathcal{O}(n^2\epsilon^{-2})$ has been obtained in [14] for derivative-free trust-region methods. Assuming the convexity of the objective function, a complexity bound of $\mathcal{O}(\epsilon^{-1})$ has been derived in [15] for nonlinear stepsize control (NSC) framework [16], which is a generalized framework for ARC and trust-region methods.

Despite ubiquity of the WCC analysis of algorithms for single-objective optimization, there are rather few studies addressing the complexity of multiobjective optimization algorithms. The complexity of gradient descent for multiobjective optimization has been analyzed in [17]. A complexity bound of $\mathcal{O}(\epsilon^{-2})$ has been established for the nonconvex case for bringing the Pareto criticality measure below ϵ. The authors have further derived a complexity bound of $\mathcal{O}(\epsilon^{-1})$ and $\mathcal{O}(\log(\epsilon^{-1}))$ for driving a gap with regard to the weighted sum of the component functions. In [18], complexity bounds have been derived for gradient descent multiobjective optimization on Riemannian manifolds. Complexity bounds of $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(\epsilon^{-1})$ have been derived for the nonconvex and convex case, respectively. In [19], under the assumption of β–Hölder continuity of the p–th order derivative of component functions, the authors have developed a high-order regularization mutliobjective method along with a complexity bound of $\mathcal{O}(\epsilon^{-\frac{p+\beta}{p+\beta-1}})$, which when $p = \beta = 1$ (i.e. when f_is are continuously differentiable with Lipschitz continuous gradient) leads to the complexity bound of $\mathcal{O}(\epsilon^{-2})$. The complexity of the Direct Multisearch (DMS) algorithm and a particular version of it has been analyzed in [20], and a complexity bound of $\mathcal{O}(\epsilon^{-2})$ has been derived for the latter case. In [21], it has been shown that a stochastic multi-gradient method enjoys complexity bounds of $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(\epsilon^{-1})$ under convexity and strong convexity assumption of the component functions, respectively.

In [9], a trust-region multiobjective method, which is called TRMP, has been proposed for solving (1). The WCC of TRMP algorithm for the nonconvex case (i.e. when the component functions in (1) are nonconvex) has been analyzed in [22]. The authors have derived a complexity bound of $\mathcal{O}(\epsilon^{-2})$ for the method. To the best of our knowledge, when the component functions in (1) are convex, no complexity bound has been derived for trust-region methods and more specifically for TRMP algorithm. In this paper, we aim to address this gap. Under convexity assumption of the component functions in (1), we derive a complexity bound of $\mathcal{O}(\epsilon^{-1})$ for finding an ϵ–Pareto critical point (see Definition 2). This matches the bound derived in [18] for steepest descent multiobjective optimization on Riemannian manifolds. Furthermore, when $m = 1$, this recovers the complexity bound of driving the norm of gradient below ϵ, which has been derived in [15] for the trust-region methods under the NSC framework. We should further add that, although we use a minimax approach for converting the original problem to a single-objective optimization problem, the resulting function is no longer smooth. Therefore, the WCC result derived here is not resulted in a trivial way by applying the techniques used in [15] for establishing WCC bound of single-objective optimization problems, which rely on the smoothness of the objective function.

The rest of the paper is organized as follows. In Section 2 we present the TRMP algorithm in its original form as given in [9], and introduce some preliminary results. The WCC analysis of the method for the convex case is established in Section 3. Finally, we will draw some conclusions and mention some future lines of research in Section 4.

Notation. In this paper, we mostly follow the related notations used in [9]. For simplicity, we assume all the norms are Euclidean norm. Notation $|\cdot|$ represents the cardinality of a set. For set $X \subset \mathbb{R}^n$, $\text{conv}(X)$ denotes its convex hull. The subgradient of function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ with ∂f. We define the nonnegative orthant (which is also called *Pareto cone*) $\mathbb{R}^m_+ =: \{(u_1, \ldots, u_m) \in$

$\mathbb{R}^m | u_i \geq 0\}$, and the positive orthant $\mathbb{R}^m_{++} =: \{(u_1, \ldots, u_m) \in \mathbb{R}^m | u_i > 0\}$. Given $u, v \in \mathbb{R}^m$, we say $u \preceq v$ (i.e. v is dominated by u) iff $v - u \in \mathbb{R}^m_+$.

2 TRMP algorithm and some preliminary results

Throughout the paper, we will use the following quantity as the *criticality measure*:

$$\omega(x) \;:=\; -\min_{\|d\|\leq 1} \left(\max_{i \in I} \nabla f_i(x)^\top d \right). \tag{2}$$

We notice that for the single-objective case (i.e. $m = 1$ in (1)), we have $\omega(x) = \|\nabla F(x)\|$.

Definition 1 [4,5] Let $x_* \in \mathbb{R}^n$. We call x_*

- a *local Pareto optimal* if there is no neighborhood $\mathcal{N}$ of x_* such that

$$F(x) \;\preceq\; F(x_*) \text{ and } F(x) \;\neq\; F(x_*), \quad \forall x \in \mathcal{N};$$

- a *Pareto critical* for F if

$$\text{range}(JF(x)) \cap (-\mathbb{R}^m_{++}) \;=\; \emptyset, \tag{3}$$

where range$(JF(x))$ denotes the linear subspace generated by the columns of Jacobian of F at x, and $-\mathbb{R}^m_{++} = \{-u | u \in \mathbb{R}^m_{++}\}$.

Let

$$\mathcal{D}(x) \;=\; \{d(x) : d(x) \text{ is a solution of (2)}\}.$$

The following lemma lists some properties of ω and its relation with Pareto criticality.

Lemma 1 *[4]*

(i) $\omega(x) \geq 0$, *for all* $x \in \mathbb{R}^n$;
(ii) *if* x *is Pareto critical of* (1), *then* $0 \in \mathcal{D}(x)$ *and* $\omega(x) = 0$;
(iii) *if* x *is not Pareto critical of* (1), *then* $\omega(x) > 0$ *and for any* $d \in \mathcal{D}(x)$ *we have*

$$\nabla f_j(x)^\top d \;\leq\; \max_{i \in I} \left\{ \nabla f_i(x)^\top d \right\} \;<\; 0, \quad \forall j \in I,$$

i.e., d *is a descent direction of* (1);
(iv) *the application* $x \mapsto \omega(x)$ *is continuous.*

Definition 2 [23] Let $x \in \mathbb{R}^n$. We call x an ϵ–Pareto critical point if

$$\omega(x) \;\leq\; \epsilon. \tag{4}$$

The following lemma states that the subgradient of a *pointwise maximum* of a some closed and convex functions at a point is equal to the convex hull of the subgradients of those functions that are active at that point.

Lemma 2 *[11, Lemma 3.1.10] Let functions $f_i(x)$, $i = 1, \dots, m$, be closed and convex. Then function*

$$\phi(x) := \max_{1 \le i \le m} f_i(x) \tag{5}$$

is also closed and convex. For any $x \in \operatorname{int}(\operatorname{dom}\phi) = \cap_{i=1}^{m} \operatorname{int}(\operatorname{dom} f_i)$ we have

$$\partial\phi(x) = \operatorname{conv}\{\partial f_i(x) : i \in I(x)\} \tag{6}$$

where $I(x) = \{i : f_i(x) = \phi(x)\}$.

In trust-region methods for minimizing single-objective optimization problems, to find a new point, typically a quadratic model within a region is minimized. If the magnitude of a quotient, which is resulted from dividing the *actual reduction* (i.e. reduction in function value) by the *predicted reduction* (i.e. reduction in model value), is close enough to 1, the new point would then be accepted and the trust-region radius is either retained or increased; otherwise the new point is rejected and the trust-region radius is reduced and the quadratic model is minimized within the new region. Further details on trust-region methods for single-objective optimization can be found in the comprehensive monograph [24].

In the following we briefly describe TRMP algorithm, as originally proposed in [9], which is a trust-region algorithm for solving (1). We refer the interested reader to [9] for further details on the method and its numerical performance.

In TRMP algorithm, which is an extension of the classical trust-region method [24], the idea of solving (1) revolves around substituting it with the following *strict scalarization*:

$$\min_{x \in \mathbb{R}^n} \phi(x) = \max_{i \in I} f_i(x). \tag{7}$$

(We notice that, in view of (6) and as the component functions f_i are assumed to be continuously differentiable, we have

$$\partial\phi(x) = \operatorname{conv}\{\nabla f_i(x) : i \in I(x)\}). \tag{8}$$

Remark 1 The aforementioned scalarization approach is different from the so-called *standard scalarization* approach, which is based on transforming the original problem to a single-objective function by a choosing a set of parameters *a priori*. In the case of the TRMP algorithm, similarly to the methods proposed in [4,5], these parameters are determined by solving a subproblem (see subproblem (9) below) at each iteration of the algorithm.

The TRMP method uses the following quadratic approximation of function ϕ at point x:

$$m(x, H, d) \ := \ \max_{i \in I} \left\{ f_i(x) + \nabla f_i(x)^\top d \right\} + \frac{1}{2} d^\top H d, \qquad \text{with } H \in \mathbb{S}^{n \times n}.$$

Then, for step calculation, at each iteration of the TRMP algorithm the following subproblem is minimized:

$$
\begin{aligned}
m(x_k, H_k, d_k) \;&:=\; \max_{i \in I} \left\{ f_i(x_k) + \nabla f_i(x_k)^\top d_k \right\} + \frac{1}{2} d_k^\top H_k d_k, \\
\text{s.t. } \|d_k\| \;&\leq\; \Delta_k,
\end{aligned}
\tag{9}
$$

where Δ_k is the trust-region radius. Function ϕ and model m are then used to compute the quotient for acceptance of trial point. Finally, the strategy for updating the trust-region is similar to the classical trust-region methods in single-objection optimization. We notice, by setting $m = 1$, we recover the basic trust-region algorithms proposed in [24] for minimization of single-objective problems.

In the continuation, we will use the following notations:

$$
m_k(\cdot) \;:=\; m\,(x_k, H_k, \cdot)
$$

and

$$
\mathcal{B}_k \;:=\; \{d \in \mathbb{R}^n : \|d\| \leq \Delta_k\}\,.
$$

Algorithm 1 (TRMP Algorithm [9])

Step 0: Initialization. Choose x_0, H_0, Δ_0, $0 < \eta_1 \leq \eta_2 < 1$, and $0 < \gamma_1 \leq \gamma_2 < 1 < \gamma_3$. Set $k = 0$.

Step 1: Step calculation. Compute a step $d_k \in \mathcal{B}_k$ that "sufficiently reduces" the function m_k.

Step 3: Acceptance of trial point. Compute $\phi(x_k + d_k)$ and

$$
\rho_k \;=\; \frac{\phi(x_k) - \phi(x_k + d_k)}{m_k(0) - m_k(d_k)}. \tag{10}
$$

If $\rho_k \geq \eta_1$, then define $x_{k+1} = x_k + d_k$; otherwise define $x_{k+1} = x_k$.

Step 4: Trust-region radius update. Set

$$
\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{if } \rho_k \geq \eta_2, \quad\; [k \text{ very successful}]; \\ [\gamma_2 \Delta_k, \Delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \; [k \text{ successful}]; \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{if } \rho_k < \eta_1, \quad\; [k \text{ unsuccessful}]. \end{cases} \tag{11}
$$

Update H_k to H_{k+1}. Set $k = k + 1$ and go to Step 1.

We denote $\mathcal{S}$ the set of all successful or very successful iterations, i.e.,

$$
\mathcal{S} \;=\; \{k \geq 0 : \text{iteration } k \text{ is successful or very successful}\}, \tag{12}
$$

and $\mathcal{S}_\ell$ be the set of all such iterations up to iteration ℓ, i.e.,

$$
\mathcal{S}_\ell \;=\; \{k \leq \ell : k \in \mathcal{S}\}. \tag{13}
$$

We also denote the set of all unsuccessful iterations with $\mathcal{U}$ and the ones up to iteration ℓ with $\mathcal{U}_\ell$. Henceforth, for the sake of simplicity, we will call all successful and very successful iterations as successful iterations.

In the following, we will list several assumptions in a more formal way, which will be required in the continuation.

Assumption 1 *For each $i \in I$, the component function f_i is twice continuously differentiable on $\mathbb{R}^n$ and lower bounded.*

Remark 2 As it has been discussed in [9], the assumption on the lower boundedness of all the component functions f_i is not restrictive. As, if unbounded, we can replace (1) with the following problem:

$$\min_{x \in \mathbb{R}^n} F(x) = \left(\exp\left(f_1(x)\right), \ldots, \exp\left(f_m(x)\right)\right). \tag{14}$$

Then, one can easily see that all the component functions in (14) are convex and lower bounded and both problems (1) and (14) have the same set of Pareto critical points.

Assumption 2 *[9, Assumption 5.1] There exists a positive constant κ_{uFh} such that, for all $x \in \mathbb{R}^n$,*

$$\|\nabla^2 f_i(x)\| \;\leq\; \kappa_{uFh}, \quad \text{for all } i \in I.$$

When the gradient of the function f_i is Lipschitz continuous with constant $L_{\nabla f_i}$, then Assumption 2 is satisfied with $\kappa_{uFh} = L_{\nabla f_i}$

Assumption 3 *[9, Assumption 5.2] The matrix H_k is uniformly bounded, that is, there exists a constant $\kappa_H \geq 1$ such that*

$$\|H_k\| \;\leq\; \kappa_{umh}, \quad \text{for all } k.$$

We will use the same assumptions as [9], which states that the model is *sufficiently reduced* at every iteration. We refer the interested reader to [9] for justification on having such assumption.

Assumption 4 *[9, Assumption 4.1] For all k,*

$$m_k(0) - m_k(d_k) \;>\; \kappa_{dlam}\,\omega(x_k) \min\left\{\frac{\omega(x_k)}{1 + \|H_k\|}, \Delta_k\right\},$$

for some $\kappa_{dlam} \in (0, 1)$.

The following lemma, which is a slightly modified version of Proposition 6.2 in [9], gives a lower bound on trust-region radius when the corresponding gradient is lower bounded.

Lemma 3 *Let Assumptions 3 and 4 hold. Suppose furthermore that there exists a constant $\kappa > 0$ such that $\omega(x_k) \geq \kappa$ for all k. Then*

$$\Delta_k \;\geq\; \kappa_1 \kappa, \quad \text{for all } k. \tag{15}$$

where

$$\kappa_1 \;=\; \frac{\gamma_1 \kappa_{dlam}(1 - \eta_2)}{\kappa_H}, \quad \kappa_H \;=\; \max\{\kappa_{uFh}, \kappa_{umh}\}, \tag{16}$$

and $\kappa_{dlam} \in (0, 1)$.

Proof See the proof of Proposition 6.2 in [9].

The following results states that the number of unsuccessful iterations is bounded by the number of successful iterations. As a result, in order to count the total number of iterations, it suffices to count the number of successful iterations. The proof follows along the lines of the proof of corresponding results given, for instance, in [15].

Theorem 1 *Given any* $\epsilon \in (0,1)$. *Let* $\omega(x_k) > \epsilon$ *for* $k = 0, 1, \cdots, \ell$. *Then Algorithm 1 takes at most* $|\mathcal{U}_\ell|$ *unsuccessful iterations, where*

$$|\mathcal{U}_\ell| \;\leq\; -\log_{\gamma_2}(\gamma_3)|\mathcal{S}_\ell| + \log_{\gamma_2}\frac{\kappa_1\epsilon}{\Delta_0}$$

for having $\|\omega_{\ell+1}\| \leq \epsilon$.

3 WCC of TRMP algorithm under convexity

In this section, we analyze the WCC of TRMP algorithm for minimization of (1). In the following assumption, we formally assume the convexity of the component functions along with the boundedness of the level set of their pointwise maximum function.

Assumption 5 *For each* $i \in I$, *the component function* f_i *is convex and there exists* $D \geq 1$ *such that* $L(x_0) \subset B(x_*; D)$, *where* x_* *is any global minimizer of* ϕ *defined in* (5).

We notice, in view of Lemma 2, that function ϕ is convex.

We will need the following lemma which shows that the difference between the function value at any point and optimal function value is of the same order as the value of ω at that point.

Lemma 4 *Let Assumption 5 hold. Then for any iteration* x_k *produced by Algorithm 1, we have*

$$\Theta_k \;:=\; \phi(x_k) - \phi(x_*) \;\leq\; D\omega(x_k). \tag{17}$$

Proof Since ϕ is convex, for all $x, y \in \mathbb{R}^n$, we have

$$\phi(y) \;\geq\; \phi(x) + g^\top(y - x),$$

for all $g \in \partial\phi(x)$. Since $\omega(x_k) = \|g_0\|$ for some $g_0 \in \partial\phi(x_k)$, by setting $x = x_k$, $y = x_*$, and using the Cauchy–Schwarz inequality, the desired inequality is resulted .

Now, as it was mentioned before, for counting the total number of iterations, we need to count only the number of successful iterations. In the continuation, without loss of generality, we shall assume that the set of successful iterations is not empty. The following theorem shows that the difference between the function value at any point and optimal function value is of the order of the inverse of the number of successful iterations.

Theorem 2 *Let Assumptions 1, 2, 3, 4, and 5 hold. Then, by applying Algorithm 1, we have*

$$\Theta_\ell \;\leq\; \frac{1}{|\mathcal{S}_\ell|\eta_1\kappa_2}, \quad \forall \ell \geq 0,$$

where Θ_ℓ is given in (17) *and*

$$\kappa_2 \;=\; \frac{\kappa_{dlam}}{D^2}\min\left\{\frac{1}{1+\kappa_H}, \kappa_1\right\}.$$

Proof In view of Lemma 4, we have $\omega(x_k) \geq \Theta_k/D$. By applying Lemma 3 with $\kappa = \Theta_k/D$, we obtain $\Delta_k \geq \kappa_1\Theta_k/D$. Thus, in view of Assumption 4, we have

$$\begin{aligned} m_k(0) - m(d_k) \;&\geq\; \kappa_{dlam}\omega(x_k)\min\left\{\frac{\omega(x_k)}{1+\kappa_H}, \Delta_k\right\} \\ &\geq\; \frac{\kappa_{dlam}\Theta_k}{D}\min\left\{\frac{\Theta_k}{(1+\kappa_H)D}, \frac{\kappa_1\Theta_k}{D}\right\} \\ &\geq\; \frac{\kappa_{dlam}\Theta_k^2}{D^2}\min\left\{\frac{1}{1+\kappa_H}, \kappa_1\right\}. \end{aligned}$$

Now, for any $k \in \mathcal{S}$, we have

$$\phi(x_k) - \phi(x_{k+1}) \;\geq\; \eta_1\kappa_2\Theta_k^2, \tag{18}$$

Hence,

$$\Theta_k - \Theta_{k+1} \;\geq\; \eta_1\kappa_2\Theta_k^2, \quad k \in \mathcal{S},$$

which, by dividing both sides by $\Theta_k\Theta_{k+1}$ and considering that $\Theta_{k+1} \geq \Theta_k$ in view of Assumption 4, leads to

$$\frac{1}{\Theta_{k+1}} - \frac{1}{\Theta_k} \;\geq\; \eta_1\kappa_2\frac{\Theta_k}{\Theta_{k+1}} \geq \eta_1\kappa_2, \quad k \in \mathcal{S}.$$

By summing up the above inequalities up to ℓ we have,

$$\frac{1}{\Theta_\ell} \;\geq\; \frac{1}{\Theta_0} + |\mathcal{S}_\ell|\eta_1\kappa_2 \geq |\mathcal{S}_\ell|\eta_1\kappa_2 \quad \ell \geq 0.$$

Therefore,

$$\Theta_\ell \;\leq\; \frac{1}{|\mathcal{S}_\ell|\eta_1\kappa_2},$$

and the proof is completed.

In view of Theorem 1 and Theorem 2, the following corollary is readily resulted.

Corollary 1 *Let Assumptions 1, 2, 3, 4, and 5 hold. Then, for any $\epsilon \in (0, 1)$, Algorithm 1 needs at most $\mathcal{O}(\epsilon^{-1})$ iterations for finding a point $\bar{x} \in \mathbb{R}^n$ such that $\phi(\bar{x}) - \phi(x_*) = \Theta_\ell \leq \epsilon$.*

Now, we obtain an upper bound on the number of successful iterations that Algorithm 1 takes for finding an ϵ–Pareto critical point.

Theorem 3 *Let Assumptions 1, 2, 3, 4, and 5 hold. Then, for any $\epsilon \in (0,1)$, Algorithm 1 takes at most $\mathcal{O}(\epsilon^{-1})$ successful iterations to find a point $\bar{x}$ such that $\omega(\bar{x}) \leq \epsilon$.*

Proof Without loss of generality, we can choose $N \geq 1$ such that $|\mathcal{S}_N| \geq 2$. Let $0 \leq \ell < N$ be chosen in a way that $|\mathcal{S}_\ell| \geq 1$, and $|\mathcal{S}_N| \geq 2|\mathcal{S}_\ell|$. In view of Theorem 2, we have

$$\phi(x_\ell) - \phi_* \;=\; \Theta_\ell \;\leq\; \frac{1}{|\mathcal{S}_\ell|\eta_1\kappa_2}, \quad \forall \ell \geq 0.$$

Hence, by Assumption 4, we have

$$\begin{aligned}
\frac{1}{|\mathcal{S}_\ell|\eta_1\kappa_2} \;&\geq\; \phi(x_\ell) - \phi_* \\
&\geq\; \phi(x_{N+1}) - \phi_* + \phi(x_{\ell+1}) - \phi(x_{N+1}) \\
&=\; \phi(x_{N+1}) - \phi_* + \sum_{\substack{k=\ell+1 \\ k\in\mathcal{S}}}^{N} \phi(x_k) - \phi(x_{k+1}) \\
&\geq\; \phi(x_{N+1}) - \phi_* + \eta_1\kappa_{dlam} \sum_{\substack{k=\ell+1 \\ k\in\mathcal{S}}}^{N} \omega(x_k) \min\left\{\frac{\omega(x_k)}{1+\kappa_H}, \Delta_k\right\} \\
&\geq\; \eta_1\kappa_{dlam}|\mathcal{S}_\ell| \min_{\substack{0\leq k\leq N \\ k\in\mathcal{S}}} \left\{\omega(x_k)\min\left\{\frac{\omega(x_k)}{1+\kappa_H}, \Delta_k\right\}\right\}.
\end{aligned}$$

Thus, there exists $1 \leq k_0 \leq N$ such that

$$\frac{1}{|\mathcal{S}_\ell|^2\eta_1^2\kappa_{dlam}\kappa_2} \;\geq\; \omega(x_{k_0})\min\left\{\frac{\omega(x_{k_0})}{1+\kappa_H}, \Delta_{k_0}\right\}. \tag{19}$$

Now, we consider the two cases of $\frac{\omega(x_{k_0})}{1+\kappa_H} \leq \Delta_{k_0}$ and $\frac{\omega(x_{k_0})}{1+\kappa_H} \geq \Delta_{k_0}$ separately. When $\frac{\omega(x_{k_0})}{1+\kappa_H} \leq \Delta_{k_0}$, in view of (19), we have

$$\frac{1+\kappa_H}{|\mathcal{S}_\ell|^2\eta_1^2\kappa_{dlam}\kappa_2} \;\geq\; \omega(x_{k_0})^2.$$

Thus, for having $\omega(x_{k_0}) \leq \epsilon$, for some $0 \leq k_0 \leq N$, Algorithm 1 needs at most

$$|\mathcal{S}_N| = 2|\mathcal{S}_\ell| = 2\left\lceil \frac{\sqrt{1+\kappa_H}}{\kappa_1\eta_1\sqrt{\kappa_{dlam}\kappa_2}}\epsilon^{-1}\right\rceil$$

successful iterations.

On the other hand, when $\frac{\omega(x_{k_0})}{1+\kappa_H} \geq \Delta_{k_0}$, in view of (19), we have

$$\frac{1}{|\mathcal{S}_\ell|^2 \eta_1^2 \kappa_{dlam} \kappa_2 (1+\kappa_H)} \;\geq\; \Delta_{k_0}^2.$$

Thus, in order to obtain $\Delta_{k_0} \leq \kappa_1 \epsilon$, for some $0 \leq k_0 \leq N$, Algorithm 1 takes at most

$$|\mathcal{S}_N| = 2|\mathcal{S}_\ell| = 2\left\lceil \frac{1}{\kappa_1 \eta_1 \sqrt{\kappa_{dlam}\kappa_2(1+\kappa_H)}} \epsilon^{-1} \right\rceil$$

successful iterations. Now, in view of Lemma 3 with $\kappa = \epsilon$, if $\Delta_{k_0} < \kappa_1 \epsilon$ then we have $\omega(x_{k_1}) < \epsilon$ for some $0 \leq k_1 \leq N$.

Therefore, in both cases the number of successful iterations for driving ω below ϵ is of $\mathcal{O}(\epsilon^{-1})$.

Remark 3 We should mention that, although the technique for the proof of Theorem 3 has been inspired by the one used in [15], the function ϕ defined at (5) is not continuously differentiable unlike the objective function considered in [15].

Now, by combining Theorem 1 and Theorem 3, we will have a bound on the total number of iterations the TRMP algorithm requires for deriving an $\epsilon-$ Pareto critical point.

Corollary 2 *Let Assumptions 1, 2, 3, 4, and 5 hold. Then, for any $\epsilon \in (0,1)$, Algorithm 1 needs at most $\mathcal{O}(\epsilon^{-1})$ iterations for driving ω below ϵ.*

We notice that the bound obtained here matches the one derived for gradient descent multiobjective optimization on Riemannian manifolds [18]. In addition, by setting $m = 1$, we recover the bound derived in [15] for classical trust-region methods for minimization of convex smooth unconstrained single-objective optimization problems. Furthermore, as a byproduct our WCC analysis, we have derived a WCC bound of $\mathcal{O}(\epsilon^{-1})$ also for trust-region methods for a class of *structured nonsmooth* convex single-objective problems (i.e. pointwise maximum of convex functions).

4 Conclusions

In this paper, under the presence of convexity assumption of the component functions, we analyzed the WCC of a multiobjective trust-region method called TRMP, which has originally been presented in [9]. We show that, in order to drive some criticality measure below some predefined positive ϵ, the required number of iterations is at most of $\mathcal{O}(\epsilon^{-1})$. This bound matches the one derived for steepest descent methods for multiobjective optimization methods, and recovers the complexity bound of trust-region in single-objective convex smooth unconstrained optimization. The techniques developed in this work could be easily adapted to derive WCC results for other versions of multiobjective trust-region methods such as the ones developed in [6, 7].

References

1. G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer-Verlag Berlin Heidelberg, 2008.
2. J. Jahn. *Vector optimization*. Springer, 2009.
3. K. Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 2012.
4. J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.*, 51:479–494, 2000.
5. J. Fliege, L. M. Graña Drummond, and B. F. Svaiter. Newton's method for multiobjective optimization. *SIAM J. Optim.*, 20(2):602–626, 2009.
6. G. A. Carrizo, P. A. Lotito, and M. C. Maciel. Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Math. Program.*, 159:339–369, 2016.
7. S. Qu, M. Goh, and B. Liang. Trust region methods for solving multiobjective optimisation. *Optim. Methods Softw.*, 28(4):796–811, 2013.
8. J. Thomann and G. Eichfelder. A trust-region algorithm for heteregeneous multiobjective optimization. *SIAM J. Optim.*, 29:1017–1047, 2019.
9. K. D. V. Villacorta, P. R. Oliveira, and A. Soubeyran. A trust-region method for unconstrained multiobjective problems with applications in satisficing processes. *J. Optim. Theory Appl.*, 160(3):865–889, 2014.
10. E. H. Fukuda and L. M. Graña Drummond. A survey on multiobjective descent methods. *Pesquisa Operacional*, 34(3):585–620, 2014.
11. Y. Nesterov. *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, Dordrecht, 2004.
12. S. Gratton, A. Sartenaer, and Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM J. Optim.*, 19:414–444, 2008.
13. C. Cartis, N. I. M. Gould, and Ph. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. part II: worst-case function-evaluation complexity. *Math. Program.*, 130:295–319, 2011.
14. R. Garmanjani, D. Júdice, and L. N. Vicente. Trust-region methods without using derivatives: Worst case complexity and the non-smooth case. *SIAM J. Optim.*, 26:1987–2011, 2016.
15. R. Garmanjani. A note on the worst-case complexity of nonlinear stepsize control methods for convex smooth unconstrained optimization. *Optimization*, 2020, to appear.
16. Ph. L. Toint. Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. *Optim. Methods Softw.*, 28:82–95, 2013.
17. J. Fliege, A. I. F. Vaz, and L. N. Vicente. Complexity of gradient descent for multiobjective optimization. *Optim. Methods Softw.*, 34(5):949–959, 2019.
18. O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Iteration-complexity and asymptotic analysis of steepest decent method for multiobjective optimization on riemannian manifolds. *J. Optim. Theory Appl.*, 184:507–533, 2020.
19. L. Calderón, M. A. Diniz-Ehrhardt, and J. M. Martínez. On high-order model regularization for multiobjective optimization. *Optim. Methods Softw.*, 2020, to appear.
20. A. L. Custódio, Y. Diouane, R. Garmanjani, and E. Riccietti. Worst-case complexity bounds of directional direct-search methods for multiobjective derivative-free optimization. *J. Optim. Theory Appl.*, 2020, to appear.
21. S. Liu and L. N. Vicente. The stochastic multi-gradient algorithm for multi-objective optimization and its application to supervised machine learning. *arXiv preprint arXiv:1907.04472*, 2019.
22. G.N. Grapiglia, J. Yuan, and Y. Yuan. On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization. *Math. Program.*, 152:491–520, 2015.
23. G. Cocchi and M. Lapucci. An augmented lagrangian algorithm for multi-objective optimization. *Comput. Optim. Appl.*, 77:29–56, 2020.
24. A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.