CONVERGENCE ANALYSIS OF INEXACT DESCENT ALGORITHM FOR MULTIOBJECTIVE OPTIMIZATIONS ON RIEMANNIAN MANIFOLDS WITHOUT CURVATURE CONSTRAINTS

XIANGMEI WANG, JINHUA WANG, AND CHONG LI

Abstract. We study the convergence issue for inexact descent algorithm (employing general step sizes) for multiobjective optimizations on general Riemannian manifolds (without curvature constraints). Under the assumption of the local convexity/quasi-convexity, local/global convergence results are established. On the other hand, without the assumption of the local convexity/quasi-convexity, but under a Kurdyka-Lojasiewicz-like condition, local/global linear convergence results are presented, which seem new even in Euclidean spaces setting and improve sharply the corresponding results in [24] in the case when the multiobjective optimization is reduced to the scalar case. Finally, for the special case when the inexact descent algorithm employing Armijo rule, our results improve sharply/extend the corresponding ones in [3, 2, 38].

1. Introduction

Let $F: \mathbb{R}^m \to \mathbb{R}^n$ be a vector function defined on $\mathbb{R}^m$. The multicriteria optimization problem consists of minimizing several objective functions simultaneously, which is formulated as follows:

(1.1) $\min_{x \in \mathbb{R}^m} F(x)$.

Since there is usually no single point which will minimize all given objective functions simultaneously, the concept of Pareto-optimality or efficiency is considered in stead of the concept of optimality. Recall from [15, 35] that a point $p \in \mathbb{R}^m$ is called a Pareto point of (1.1) (or an efficient point), if there does not exist a different point $q \in \mathbb{R}^m$ such that $F(q) \preceq F(p)$ and $F(q) \neq F(p)$ (where sign “$\preceq$” means the classical partial order on Euclidean space $\mathbb{R}^n$; see (2.2) in Section 2 for the definition.)

Problem (1.1) arises in many applications such as engineering disciplines, location science, statistics, management science; see, e.g., [5,6,13,24] and the references therein. One of the standard techniques for finding the Pareto points of (1.1) is the scalarization approach, which in fact tries to compute a discrete approximation to the whole set of the Pareto points. Since it was
proposed by Geoffrion in [19] for solving the multicriteria optimization problems in Euclidean spaces, the scalarization technique has been extensively studied in the literature; see, e.g., [7, 10, 20, 25, 16, 27] for more details. In general, the scalarization approach requires some parameters to be specified in advance, leaving the modeler and the decision-maker with the burden of choosing them. Another important approach for finding the Pareto points is the descent-type method. This type of method usually does not require any parameter information, which includes such as the (steepest) descent algorithm, Newton method, proximal point method and trust-region method; see, e.g., [15, 14, 9, 11, 17, 18, 5, 6, 30]. We are particularly interested in the (steepest) descent algorithm proposed by Fliege and Svaiter in [15] for solving the multicriteria optimization problem in Euclidean spaces, which was well-studied and has been extended to the multiobjective optimization (equipped with the partial order induced by a general closed convex pointed cone); see, e.g., [9, 11, 17, 18] and the references therein.

Recently, some important notions, techniques and approaches in Euclidean spaces have been extended to Riemannian manifold settings; see, e.g., [12, 21, 22, 23, 28, 40] and the references therein. As pointed out in [3], such extensions are natural and, in general, nontrivial; and enjoy some important advantages; see, e.g., [1, 33, 34, 41, 24] for more details. In particular, in [24], the gradient algorithm (employing general step sizes) was extended for scalar optimization problems on general Riemannian manifolds (without curvature constraints). Under the assumption of the local convexity/quasi-convexity (resp. weak sharp minima), local/global convergence (resp. linear convergence) results are established (see [24]).

One the other hand, the exact/inexact descent algorithm employing Armijo rule was recently extended to solve the multicritera optimization problem on Riemannian manifolds in [2, 3], where it was shown that the partial convergence property (i.e., each cluster point of the generated sequence by the inexact descent algorithm is a Pareto critical point) holds on general Riemannian manifolds, while the full convergence does for the (vector) objective function being quasi-convex on the whole manifold of nonnegative sectional curvatures; see [3, Theorems 5.1 and 5.2]. The further development of this full convergence results of the exact/inexact descent algorithm employing Armijo rule have been given in [38] where they were established under the following weaker assumption

(A) the objective function is quasi-convex only on a sub-level set which is of curvatures bounded from below.

The main purpose of the present paper is to study the local/global convergence issue for the inexact descent algorithm (employing general step sizes) for multiobjective optimizations on general Riemannian manifolds (without curvature constraints). The present paper contains two topics of convergence results for the descent algorithm employing more general step sizes (which includes the Armijo step sizes as a special case).
One is the local/global convergence for locally quasi-convex function $F$ which includes local convergence, that is, any sequence generated with initial point close enough to a critical point converges to a critical point (see Theorem 3.5, which seem new in the linear space setting), and the global convergence which means that any sequence generated with arbitrary initial point from the domain of the function $F$ does (see Theorem 5.1(i) and Corollary 5.3). In particular, the global convergence result is established for the descent algorithm employing the Armijo step sizes under the following weaker assumption than (A) (see Lemma 5.2):

\((H)\) The generated sequence \(\{p_k\}\) has a cluster point \(\bar{p}\) and $F$ is quasi-convex around \(\bar{p}\).

The other is the locally/globally linear convergence without locally quasi-convex assumption for $F$ which includes local convergence, that is, any sequence generated with initial point close enough to a weak Pareto optimum converges to a weak Pareto optimum(see Theorem 4.1 which seems new in the linear space setting in the case when the Kurdyka-Lojasiewicz-like property holds at the weak Pareto optimum), and the global convergence which means that any sequence generated with arbitrary initial point from the domain of the function $F$ does (see Theorem 5.1(ii) and Corollary 5.3), that is, if the following assumption is assumed, we show that the sequence \(\{p_k\}\) converges linearly:

- The generated sequence \(\{p_k\}\) has a cluster point \(\bar{p}\) which is a locally weak Pareto optimum, the Kurdyka-Lojasiewicz-like property holds at \(\bar{p}\) and the step sizes \(\{t_k\}\) has a positive lower bound.

(Note by Lemma 4.3 that the Armijo step sizes has a positive lower bound if Jacobian $JF$ is Lipschitz continuous around \(\bar{p}\)). To the best of our knowledge, this global linear convergence result also seems new even in the linear space setting.

Note that our results in the present paper extend/improve the corresponding results in [24] for scalar optimization problems on Riemannian manifolds to multiobjective optimizations on Riemannian manifolds. In particular, it should be remarked that for the linear convergence of the gradient method, our result improves sharply the corresponding result in [24] in the sense that we remove the local quasi-convexity assumption; see Remark 4.2.

The remaining of the paper is organized as follows. Some basic notions and notation on Riemannian manifolds and the inexact descent algorithm employing general step sizes for solving the multicriteria problem on Riemannian manifolds are presented in the next section. In Section 3, some related properties about the convexity properties of vector functions and some useful lemmas are presented, and local convergence results are established, while locally linear convergence result is presented in Section 4. Global convergence (resp. linear convergence) results are presented in the last section.
2. Preliminaries and inexact descent algorithm

2.1. Notation and notions on Riemannian manifolds.

The notation and notions on Riemannian manifolds used in the present paper are standard, and the readers are referred to some textbooks for more details; see, e.g., [8, 32, 34].

Let $M$ be a connected and complete $m$-dimensional Riemannian manifold. We use $\nabla$ to denote the Levi-Civita connection on $M$. Let $p \in M$, and let $T_pM$ stand the tangent space at $p$ to $M$. We denote by $\langle \cdot, \cdot \rangle_p$ the scalar product on $T_pM$ with the associated norm $\| \cdot \|_p$, where the subscript $p$ is sometimes omitted. For $q \in M$, let $\gamma : [0, 1] \to M$ be a piecewise smooth curve joining $p$ to $q$. Then, the arc-length of $\gamma$ is defined by $l(\gamma) := \int_0^1 \| \gamma'(t) \| dt$; and the Riemannian distance from $p$ to $q$ is defined by $d(p, q) := \inf_{\gamma} l(\gamma)$, where the infimum is taken over all piecewise smooth curves $\gamma : [0, 1] \to M$ joining $p$ to $q$. A smooth curve $\gamma$ is called a geodesic if and only if $\nabla_{\gamma'} \gamma' = 0$. A geodesic joining $p$ to $q$ is said to be minimal if its arc-length equals the Riemannian distance between $p$ and $q$. By the Hopf-Rinow theorem [8], $(M, d)$ is a complete metric space, and there is at least one minimal geodesic joining $p$ to $q$. The closed metric ball in $M$ centered at the point $p$ with radius $r > 0$ is denoted by $B(p, r)$, i.e.,

$$B(p, r) := \{ q \in M : d(p, q) \leq r \}.$$

Let $Q \subseteq M$ be a subset and $p, q \in Q$. The set of all geodesics $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\gamma(1) = q$ satisfying $\gamma([0, 1]) \subseteq Q$ is denoted by $\Gamma^Q_{pq}$, that is,

$$\Gamma^Q_{pq} := \{ \gamma : [0, 1] \to Q : \gamma(0) = p, \gamma(1) = q \text{ and } \nabla_{\gamma'} \gamma' = 0 \}.$$

Recall the convexity radius $r_{cvx}(p)$ of $p \in M$ which is defined by (2.1)

$$r_{cvx}(p) := \sup \left\{ r > 0 : \text{ each ball in } B(p, r) \text{ is strongly convex and each geodesic in } B(p, r) \text{ is minimal } \right\}.$$

Then, $r_{cvx}(p) > 0$ for any $p \in M$; see, e.g., [32, Theorem 5.3].

Definition 2.1 below presents the notions of different kinds of convexities about subsets in $M$; see e.g., [23, 36].

Definition 2.1. A nonempty subset $Q$ of the Riemannian manifold $M$ is said to be

(a) weakly convex if and only if, for any $p, q \in Q$, there is a minimal geodesic of $M$ joining $p$ to $q$ and it is in $Q$;

(b) totally convex if and only if, for any $p, q \in Q$, all geodesics of $M$ joining $p$ to $q$ lie in $Q$.

Note by definition that the strong/total convexity implies the weakly convexity for any subset $Q$. 

2.2. Convexity.

Below, we recall the notion of convexity of a real-valued scalar function $f : M \to \mathbb{R}$. Item (b) in the following definition was known in [21, Definition 6.1 (b)] (for the convexity) and [31, Definition 2.2] (for the quasi-convexity).

**Definition 2.2.** Let $f : M \to \mathbb{R}$ and let $Q \subseteq M$ be weakly convex. Then, $f$ is said to be

(a) convex (resp. quasi-convex) on $Q$ if, for any $x, y \in Q$ and any geodesic $\gamma \in \Gamma_{xy}^{Q}$, the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex (resp. quasi-convex) on $[0, 1]$;

(b) pseudo-convex on $Q$ if $f$ is differentiable and for any $p, q \in Q$, any geodesic $\gamma \in \Gamma_{pq}^{Q}$, there holds:

$$\langle \nabla f(p), \gamma'(0) \rangle \geq 0 \implies f(q) \geq f(p).$$

(c) convex (resp. quasi-convex, pseudo-convex) if $f$ is convex (resp. quasi-convex, pseudo-convex) on $M$.

(d) convex (resp. quasi-convex, pseudo-convex) around $x \in M$ if $f$ is convex (resp. quasi-convex, pseudo-convex) on $B(x, r)$ for some $r > 0$.

It is clear that the convexity implies the quasi-convexity and pseudo-convexity (assuming $f$ is differentiable). The assertions in the following lemma can be proved directly by definition and are known for some special cases; see, e.g., [31, Theorems 5.1, 6.2] for assertion (i) and [29, Proposition 3.1] for assertion (ii).

**Lemma 2.3.** Let $f : M \to \mathbb{R}$ be differentiable. Let $Q$ be weakly convex and let $x \in Q$. Then, the following assertions hold.

(i) If $f$ is convex on $Q$, then it holds for any $y \in Q$ that

$$f(y) \geq f(x) + \langle \nabla f(x), \gamma_{xy}'(0) \rangle \quad \text{for all } \gamma_{xy} \in \Gamma_{xy}^{Q}.$$ 

(ii) If $f$ is quasi-convex on $Q$, then it holds for any $y \in Q$ with $f(y) \leq f(x)$ that

$$\langle \nabla f(x), \gamma_{xy}'(0) \rangle \leq 0 \quad \text{for all } \gamma_{xy} \in \Gamma_{xy}^{Q}.$$ 

Below, we extend the notions of different kinds of convexities to vector functions on $M$, which are known for the case when $Q = M$; see, items (a), (b) in [21, definition 5.1] and item (c) in [3, definition 5.1]. To proceed, as usual, we use “≤” and “<” to denote the classical partial order and the strictly partial order defined by

(2.2) $x \leq y$ (or $y \geq x$) $\iff y - x \in \mathbb{R}_+^n$ \quad \text{for } x, y \in \mathbb{R}^n$

and

$x < y$ (or $y > x$) $\iff y - x \in \mathbb{R}_{++}^n$ \quad \text{for } x, y \in \mathbb{R}^n,$

respectively, where

$$\mathbb{R}_+^n := \{ x = (x_i) \in \mathbb{R}^n : x_i \geq 0, i \in I \}$$

and

$$\mathbb{R}_{++}^n := \{ x = (x_i) \in \mathbb{R}^n : x_i > 0, i \in I \}.$$
Definition 2.4. Let \( Q \subseteq M \) be weakly convex. The vector function \( F : M \to \mathbb{R}^n \) is said to be

(a) convex on \( Q \) if for any \( p, q \in Q \) and any geodesic \( \gamma \in \Gamma_{pq}^Q \), there holds:
\[
F(\gamma(t)) \preceq (1-t)F(p) + tF(q) \quad \text{for any} \; t \in [0,1].
\]

(b) quasi-convex on \( Q \) if for any \( p, q \in Q \) and any geodesic \( \gamma \in \Gamma_{pq}^Q \), there holds:
\[
F(\gamma(t)) \preceq \max \{ F(p), F(q) \} \quad \text{for any} \; t \in [0,1].
\]

(c) pseudo-convex on \( Q \) if \( F \) is differentiable and for any \( p, q \in Q \), any geodesic \( \gamma \in \Gamma_{pq}^Q \), there holds:
\[
JF(p)(\gamma'(0)) \preceq 0 \implies F(q) \preceq F(p).
\]

Clearly for a vector function, the convexity implies both the pseudo-convexity (assuming that \( F \) is differentiable) and the quasi-convexity.

Proposition 2.5 below shows the equivalence between the convexity of \( F \) and its scalarization. Its proof is easy and so is omitted here.

Proposition 2.5. Let \( Q \subseteq M \) be weakly convex. \( F := (f_i)_{i \in I} : M \to \mathbb{R}^n \) is convex (resp. quasi-convex, pseudo-convex) on \( Q \) if and only if for each \( \{\alpha_i : i \in I\} \subset [0,1] \) with \( \sum_{i \in I} \alpha_i = 1 \), \( \sum_{i \in I} \alpha_i f_i \) is convex (resp. quasi-convex, pseudo-convex) on \( Q \).

Furthermore, using the same arguments for proving [3, Proposition 5.1] (for the case when \( Q := M \)), one can check the following lemma.

Lemma 2.6. Let \( F \) be a differentiable vector function. Then, \( F \) is quasi-convex on \( Q \) if and only if, for any \( p, q \in Q \) and any geodesic \( \gamma_{pq} \in \Gamma_{pq}^Q \),
\[
F(q) \preceq F(p) \implies JF(p)(\gamma_{pq}'(0)) \preceq 0.
\]

Consequently, \( F \) is pseudo-convex implies that it is quasi-convex.

The following lemma is useful; see [3, Proposition 5.2].

Lemma 2.7. If \( F \) is pseudo-convex (e.g., convex) (on \( M \)), then a point \( p \in M \) is a Pareto critical point of \( F \) if and only if it is a weak Pareto optimum of \( (2.4) \).

2.3. Multiobjective optimizations on Riemannian manifold.

Below, we consider a vector function \( F : M \to \mathbb{R}^n \) given by
\[
F(p) := (f_i(p))_{i \in I} = (f_1(p), f_2(p), \ldots, f_n(p)) \quad \text{for any} \; p \in M,
\]
where \( I := \{1, 2, \cdots, n\} \) and for each \( i \in I, f_i : M \to \mathbb{R} \) is a function defined on \( M \). The vector function \( F \) is said to be (continuously) differentiable if each \( f_i \) is (continuously) differentiable \( (i \in I) \). For a continuously differentiable vector function \( F \), the Riemannian Jacobian \( JF \) and its image at \( p \in M \) are respectively denoted by
\[
JF(p) := (\nabla f_i(p))_{i \in I} \quad \text{and} \quad \text{Im}JF(p) := \{JF(p)(v) : v \in T_p M\},
\]
where

\[(2.3) \quad JF(p)(v) := (\langle \nabla f_i(p), v \rangle)_{i \in I}.\]

In the remainder of this paper, we always assume that \( F := (f_i)_{i \in I} : M \to \mathbb{R}^n \) is continuously differentiable. The vector optimization problem considered in the present paper is denoted by

\[(2.4) \quad \min_{p \in M} F(p).\]

Recall that a point \( p \in M \) is called a (globally) Pareto (resp. weak Pareto) optimum of \( (2.4) \) if there does not exist other point \( q \in M \) such that

\[(2.5) \quad F(q) \preceq F(p) \quad \text{(resp.} \quad F(q) \prec F(p) \text{)} \quad \text{and} \quad F(q) \neq F(p)\]

(see, e.g., [15, 35] in Euclidean space settings). Furthermore, a point \( p \in M \) is called a locally Pareto (resp. weak Pareto) optimum of \( (2.4) \) if there exists a neighborhood \( U \subset M \) of \( p \) such that there does not exist other point \( q \in U \setminus \{p\} \) satisfying \( (2.5) \).

Recall from [2, 3], that a point \( p \in M \) is called a Pareto critical point of \( F \) if the image of \( JF(p) \) satisfies

\[\text{Im}(JF(p)) \cap (-\mathbb{R}^n_{++}) = \emptyset.\]

By definition, each (locally) Pareto optimum of \( F \) is a Pareto critical point of \( F \).

Let \( p \in M \) and assume that it is not a Pareto critical point of \( F \). By definition, there exists a direction \( v \in T_p M \) satisfying \( JF(p)(v) \in -\mathbb{R}^n_{++}, \) that is, \( v \) is a descent direction at \( p \). We shall give some notation related to the descent directions of \( F \) at \( p \). As done in [3], we consider the following unconstrained optimization problem on \( T_p M \):

\[(2.6) \quad \min_{v \in T_p M} \alpha_p(v) := \max_{i \in I} \langle \nabla f_i(p), v \rangle + \frac{1}{2} \|v\|^2.\]

Noting that \( \alpha_p \) is strongly convex on \( T_p M \), problem \( (2.6) \) has a unique solution. The solution of problem \( (2.6) \) and the associated value are denoted by \( v(p) \) and \( \alpha_p^* \) respectively, that is,

\[(2.7) \quad v(p) := \text{argmin}_{v \in T_p M} \alpha_p(v), \quad \alpha_p^* := \alpha_p(v(p)).\]

As pointed out in [3], the vector \( v(p) \) is in fact a descent direction at \( p \) and always called the steepest descent direction at \( p \). Furthermore, we need the concept of the \( \sigma \)-approximate steepest descent direction, which can be found in [3, Definition 4.2] (see also [11, Definition 3.4] for the Euclidean space version).

**Definition 2.8.** Let \( \sigma \in [0, 1) \). A vector \( v_p \in T_p M \) is said to be a \( \sigma \)-approximate steepest descent direction at \( p \) if it satisfies \( \alpha_p(v_p) \leq (1 - \sigma)\alpha_p^* \).

For convenience, for any \( p \in M \) and \( \sigma \in [0, 1) \), we use \( D_\sigma(p) \) to denote the set of all \( \sigma \)-approximate steepest descent direction at \( p \). It is clear that for
any \( \sigma \in [0,1) \), \( v(p) \in D_\sigma(p) \). The following lemma shows some properties related to the (approximate) steepest descent directions.

**Proposition 2.9.** Let \( p \in M \). The following assertions hold:

(i) \( v(p) = 0 \) (or \( \alpha_p^* = 0 \)) if and only if \( p \) is a Pareto critical point.

(ii) There exist \( \{\lambda_i : i \in I(p)\} \subseteq [0,1) \) with \( \sum_{i \in I(p)} \lambda_i = 1 \), such that

\[
(2.8) \quad v(p) = - \sum_{i \in I(p)} \lambda_i \nabla f_i(p),
\]

where \( I(p) := \{i \in I : \langle \nabla f_i(p), v(p) \rangle = \max_{j \in I} \langle \nabla f_j(p), v(p) \rangle\} \); and the function: \( M \ni p \mapsto v(p) \in T_p M \) is continuous on \( M \).

(iii) If \( p \in M \) is not a Pareto critical point and \( v_p \in D_\sigma(p) \), then there holds

\[
(2.9) \quad \alpha_p(v_p) := \max_{i \in I} \langle \nabla f_i(p), v_p \rangle + \frac{1}{2} \|v_p\|^2 < 0,
\]

which particularly implies that \( v_p \) is a descent direction. Furthermore, the following relation holds:

\[
(2.10) \quad \|v_p\| \geq (1 - \sqrt{\sigma})\|v(p)\|.
\]

(iv) Let \( p \in M \) be a Pareto critical point. Then, for any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that

\[
(2.11) \quad \|v_q\| \leq \varepsilon \quad \text{for any} \quad v_q \in D_\sigma(q), q \in B(p, \delta).
\]

**Proof.** Assertions (i)-(ii) are known in [3, Lemmas 4.1, 4.2]. To show assertion (iii), suppose that \( p \) is not a Pareto critical point. Then, we see from assertion (i) that \( v(p) \neq 0 \). First, we show that

\[
(2.12) \quad \alpha_p^* = -\frac{1}{2} \|v(p)\|^2.
\]

Granting this, we get that \( \alpha_p^* < 0 \), and so (2.10) is valid by recalling \( \alpha(v_p) \leq (1 - \sigma)\alpha_p^* \). To show (2.12), by definition of the subindex \( I(p) \), there holds

\[
(2.13) \quad \langle \nabla f_i(p), v(p) \rangle = \max_{i \in I} \langle \nabla f_i(p), v(p) \rangle \quad \text{for each} \quad i \in I(p).
\]

Note by (2.8) that there exist \( \{\lambda_i : i \in I(p)\} \subseteq [0,1) \) with \( \sum_{i \in I(p)} \lambda_i = 1 \) such that \( v(p) = - \sum_{i \in I(p)} \lambda_i \nabla f_i(p) \). Then, there holds:

\[
-\|v(p)\|^2 = \bigl( \sum_{i \in I(p)} \lambda_i \nabla f_i(p), v(p) \bigr) = \sum_{i \in I(p)} \lambda_i \langle \nabla f_i(p), v(p) \rangle.
\]

In view of (2.13), we get that \( \max_{i \in I} \langle \nabla f_i(p), v(p) \rangle = -\|v(p)\|^2 \), and so (2.12) holds by definition. Letting \( v_p \in D_\sigma(p) \), we estimate that

\[
-\langle v(p), v_p \rangle + \frac{1}{2} \|v_p\|^2 = \bigl( \sum_{i \in I(p)} \lambda_i \nabla f_i(p), v_p \bigr) + \frac{1}{2} \|v_p\|^2 \\
\leq \max_{i \in I} \langle \nabla f_i(p), v_p \rangle + \frac{1}{2} \|v_p\|^2 \\
\leq (1 - \sigma)\alpha_p^* = -\frac{1}{2}(1 - \sigma)\|v(p)\|^2,
\]

where the last equality is by (2.12). Then, we have that \( \|v(p) - v_p\|^2 \leq \sigma\|v(p)\|^2 \). This implies (2.10), and so assertion (iii) is shown.
To show assertion (iv), let \( p \) be a Pareto critical point. Then, by definition, for any \( v \in T_p M \), there exists an index \( i_v \in I \) such that \( \langle \nabla f_{i_v}(p), v \rangle \geq 0 \). Since \( F \) is continuously differentiable, for any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that

\[
\tag{2.14}
\langle \nabla f_{i_v}(q), P_{q,p}v \rangle \geq -\frac{\varepsilon^2}{2} \quad \text{for any } \ q \in \mathbb{B}(p, \delta).
\]

Fix \( q \in \mathbb{B}(p, \delta) \) and \( v_q \in D_\sigma(q) \). Then \( P_{p,q}v_q \in T_p M \) and it follows from (2.14) that there exists \( i_{v_q} \in I \) satisfying \( \langle \nabla f_{i_{v_q}}(q), P_{q,p}P_{p,q}v_q \rangle \geq -\frac{\varepsilon^2}{2} \). Thus, in view of \( v_q \in D_\sigma(q) \), we get by definition that

\[
0 \geq (1 - \sigma)\alpha_{q}^{*} \geq \alpha_{q}(v_q) \geq \langle \nabla f_{i_{v_q}}(q), v_q \rangle + \frac{1}{2}\|v_q\|^2 \geq -\frac{\varepsilon^2}{2} + \frac{1}{2}\|v_q\|^2,
\]
which shows (2.11), completing the proof.

We end this subsection by recalling the concept of \( s \)-compatible decent direction at \( p \). Recall from [11, Definition 3.4] (see also [3], where the authors used the notion of the compatible scalarization) that a vector \( v \in T_p M \) is said to be \( s \)-\emph{compatible} at \( p \) if there exist \( \{\alpha_i^p : i \in I\} \subset [0,1] \) with \( \sum_{i \in I} \alpha_i^p = 1 \) such that

\[
\tag{2.15}
v = -\sum_{i \in I} \alpha_i^p \nabla f_i(p).
\]

2.4. Inexact descent algorithm with general step sizes for multiobjective optimizations.

Below, we propose an inexact descent algorithm employing general step sizes for solving problem (2.4).

**Algorithm 2.1.** (Inexact descent algorithm with general step sizes)

**Step 0.** Select \( p_0 \in M, \sigma, \beta \in (0,1), R \in [1, +\infty) \) and set \( k := 0 \).

**Step 1.** If \( p_k \) is a Pareto critical point, then stop; otherwise select \( v_k \in D_\sigma(p_k) \) and construct the geodesic \( \gamma_k \) such that

\[
\tag{2.16}
\gamma_k(0) = p_k \quad \text{and} \quad \gamma'_k(0) = v_k.
\]

**Step 2.** Select the step size \( t_k \in (0, R] \) which satisfies the following inequality:

\[
\tag{2.17}
F(\gamma_k(t_k)) \leq F(p_k) + \beta t_k JF(p_k)(v_k).
\]

**Step 3.** Set \( p_{k+1} := \gamma_k(t_k) \), replace \( k \) by \( k + 1 \) and go to step 1.

Recall that Algorithm 2.1 is said to be well defined if for each \( k \in \mathbb{N} \), there always exists \( t_k \in (0, 1] \) satisfying (2.17) in Step 2. Let \( \nu \in (0, 1) \). Algorithm 2.1 is said to employ the (generalized) Armijo step sizes (cf. [3]) if each step size \( t_k \) in Step 2 is chosen by

\[
\tag{2.18}
t_k := \max\{\nu^{-i} : i \in \mathbb{N}, \ F(\gamma_k(\nu^{-i})) \leq F(p_k) + \beta \nu^{-i} JF(p_k)(v_k)\}.
\]
Define a mapping \( \varphi : M \to \mathbb{R} \) by
\[
\varphi(p) := \sup_{q \in M} \min_{i \in I} (f_i(p) - f_i(q)) \quad \text{for each } p \in M.
\]

Clearly, \( \varphi(p) \geq 0 \) for each \( p \in M \) and \( \varphi(p) = 0 \) if and only if \( p \) is a weak Pareto optimum of (2.4). Moreover, the following lemma quantifies some properties of the function \( \varphi \).

**Proposition 2.10.** (i) \( \varphi \) is locally Lipschitz continuous on \( M \), that is, for each \( \bar{p} \in M \), there exist \( \delta > 0 \) and \( L > 0 \) such that the function \( \varphi \) is Lipschitz continuous on \( B(\bar{p}, \delta) \) with modulus \( L \):
\[
|\varphi(p) - \varphi(p')| \leq Ld(p, p') \quad \text{for each } p, p' \in B(\bar{p}, \delta).
\]

(ii) Let \( p, q \in M \). If \( F(q) \preceq F(p) \), then \( \varphi(q) \leq \varphi(p) \).

(iii) Let \( \{p_k\} \) (together with associated sequences \( \{t_k\}, \{v_k\} \)) be a sequence generated by Algorithm 2.1. Then, we have the following estimate
\[
\frac{\beta t_k}{2} \Vert v_k \Vert^2 \leq \varphi(p_k) - \varphi(p_{k+1}) \quad \forall k \in \mathbb{N}.
\]

**Proof.** (i). Let \( \bar{p} \in M \). Noting that \( F \) is continuously differentiable, there exist \( \delta > 0 \) and \( L > 0 \) such that
\[
|f_i(p) - f_i(p')| \leq Ld(p, p') \quad \text{for each } p, p' \in B(\bar{p}, \delta) \quad \text{and for each } i \in I.
\]

Fix \( p, p' \in B(\bar{p}, \delta) \). Then, it follows that
\[
f_i(p) - f_i(q) \leq f_i(p') - f_i(q) + Ld(p, p') \quad \text{for each } i \in I \quad \text{and for each } q \in M,
\]
which implies that
\[
\sup_{q \in M} \min_{i \in I} (f_i(p) - f_i(q)) \leq \sup_{q \in M} \min_{i \in I} (f_i(p') - f_i(q)) + Ld(p, p'),
\]
that is,
\[
\varphi(p) - \varphi(p') \leq Ld(p, p').
\]
With similar technique, we can also check that
\[
\varphi(p') - \varphi(p) \leq Ld(p, p').
\]

Hence, it follows that
\[
\|\varphi(p) - \varphi(p')\| \leq Ld(p, p'),
\]
showing assertion (i).

(ii). It’s clearly by definition.

(iii). By the definition of function \( \varphi \), one has that
\[
\varphi(p_{k+1}) = \sup_{q \in M} \min_{i \in I} (f_i(p_{k+1}) - f_i(q))
\leq \sup_{q \in M} \min_{i \in I} (f_i(p_k) + \beta t_k \nabla f_i(p_k)^T v_k - f_i(q))
\leq \sup_{q \in M} \min_{i \in I} (f_i(p_k) - \frac{\beta t_k}{2} \|v_k\|^2 - f_i(q))
= -\frac{\beta t_k}{2} \|v_k\|^2 + \varphi(p_k),
\]

(2.21)
where the first inequality is by (2.17) and the second inequality thanks to $\nabla f_i(p_k)^T v_k \leq -\frac{1}{2} \|v_k\|^2$ by (2.9). Hence, (2.20) is seen to hold, completing the proof.

The following proposition is about some useful properties of sequence $\{p_k\}$ (together with $\{t_k\}$ and $\{v_k\}$) generated by Algorithm 2.1 which includes the partial convergence result for Algorithm 2.1 (see assertion (iii) below), while assertion (ii) improves the corresponding results in [3, Theorem 5.1(i)] where (2.23) holds under the assumption that $\{p_k\}$ has a cluster point.

**Proposition 2.11.** Algorithm 2.1 is well defined and each sequence $\{p_k\}$ generated by Algorithm 2.1 has the following properties:

(i) $\{F(p_k)\}$ is non-increasing monotonically and for any $k \in \mathbb{N}$:

\[
d(\gamma_k(t), p_k) \leq t \|v_k\| \text{ for any } t \in [0, t_k].
\]

(ii) \[
\sum_{k \in \mathbb{N}} t_k^2 \|v_k\|^2 < +\infty.
\]

(iii) If $\{t_k\}$ has a positive lower bound or that $\{t_k\}$ satisfies the Armijo step sizes, then each cluster point of the sequence $\{p_k\}$ is a Pareto critical point of $F$.

**Proof.** The well definedness of Algorithm 2.1 follows from [3, Proposition 4.1]. By Steps 2 and 3 of Algorithm 2.1 assertion (i) is clear.

By (2.20), one has

\[
\sum_{j=0}^{k} \frac{\beta t_j^2}{2} \|v_j\|^2 \leq \varphi(p_0) - \varphi(p_{k+1}) < \varphi(p_0) < +\infty,
\]

showing assertion (ii).

To show assertion (iii), suppose that $\{t_k\}$ has a positive lower bound. Then, it follows from (2.23) that $\|v_k\| \to 0$. Note by (2.10) $\|v_k\| \geq (1 - \sqrt{\sigma}) \|v(p_k)\|$. Hence, one has that $v(p_k) \to 0$ which, together with Proposition 2.9(i) and (ii), implies that each cluster point of the sequence $\{p_k\}$ is a Pareto critical point of $F$. In the case when $\{t_k\}$ satisfies the Armijo step sizes, the conclusion follows from [3, Theorem 5.1(ii)]. The proof is complete.

3. **Local convergence under locally quasi-convex assumption**

This section is devoted to establishing local convergence of Algorithm 2.1 under locally quasi-convex assumption. Firstly, we need some useful lemmas.

The inequality in the following lemma plays an important role in our study.
Lemma 3.1. Let $Q \subseteq M$ be weakly convex with nonempty interior, let $p \in \text{int}Q$ and $v \in T_pM$ be a s-compatible vector at $p$. Let $t \geq 0$ and $\gamma : [0, +\infty) \rightarrow M$ be the geodesic satisfying
\begin{equation}
\gamma(0) = p, \quad \gamma'(0) = v \neq 0 \quad \text{and} \quad \gamma([0, t]) \subset \text{int}Q.
\end{equation}
Suppose further that the sectional curvatures on $Q$ are bounded from below by some $\kappa < 0$, and that $F$ is quasi-convex on $Q$. Then the following inequality holds for any $q \in \text{int}Q$ satisfying $F(q) \preceq F(p)$:
\begin{equation}
\text{d}^2(\gamma(t), q) < \text{d}^2(p, q) + \frac{3t^2\|v\|^2}{2h(\sqrt{|\kappa|d(p, q)})} \quad \text{if} \quad \sqrt{|\kappa|\|v\|} \leq 1.
\end{equation}

Proof. By assumption, there exist $\{\alpha^p_i : i \in I\} \subset [0, 1]$ with $\sum_{i \in I} \alpha^p_i = 1$ such that $v = -\sum_{i \in I} \alpha^p_i \nabla f_i(p)$. Define a function $f : M \rightarrow \mathbb{R}$ by
\begin{equation}
f(\cdot) := \sum_{i \in I} \alpha^p_i f_i(\cdot).
\end{equation}
Then $f$ is quasi-convex (due to Proposition 2.5) and differentiable on $Q$, $\nabla f(p) = -v$, $f(q) \leq f(p)$. Hence, [24, Lemma 2.5] is applicable with $q, p$ in place of $z, x$ to concluding that (3.2) holds, completing the proof. \square

The following lemmas is known in [37, Lemma 2.3].

Lemma 3.2. Let $\{a_k\}, \{b_k\} \subset (0, +\infty)$ be two sequences satisfying
\begin{equation}
a_{k+1} \leq a_k(1 + b_k) \quad \text{for all} \quad k \in \mathbb{N},
\end{equation}
and $\sum_{k=0}^{\infty} b_k < \infty$. Then, $\{a_k\}$ is convergent and so it is bounded.

Let $S \subset M$ be a subset. Recall that a sequence $\{p_k\} \subset M$ is said to be quasi-Fejér convergent to $S$ if, for any $s \in S$, there exists a sequence $\{\varepsilon_k\} \subset (0, +\infty)$ satisfying $\sum_{k=1}^{\infty} \varepsilon_k < \infty$ such that
\begin{equation}
\text{d}^2(p_{k+1}, s) \leq \text{d}^2(p_k, s) + \varepsilon_k \quad \text{for each} \quad k \in \mathbb{N}.
\end{equation}
We end this section with the following lemma, which provides some properties for quasi-Fejér convergent sequences (see e.g., [13, Theorem 4.3]).

Lemma 3.3. Let $\{p_k\} \subset M$ be a sequence quasi-Fejér convergent to $S$. Then, $\{p_k\}$ is bounded. If, furthermore, a cluster point $p$ of $\{p_k\}$ belongs to $S$, then $\lim_{k \rightarrow \infty} p_k = p$.

For the remainder of the paper, we make the following assumption:

(A_{sc}): Each vector $v_k$ in $\{v_k\}$ is s-compatible at $p_k$.

To study the local convergence of Algorithm 2.1, we further need the following assumption:
\begin{equation}
\bar{p} \text{ is a Pareto critical point and } F \text{ is quasi-convex around } \bar{p}.
\end{equation}

For the following key lemma, recall that $R$ is the constant given at the beginning of Algorithm 2.1.
Lemma 3.4. Suppose that assumption (3.5) holds. Then, for any \( \delta > 0 \), there exist \( \bar{\delta}, \hat{\delta}, c > 0 \) satisfying \( \bar{\delta} < \hat{\delta} < \frac{\delta}{2} \) such that, for any \( k \in \mathbb{N} \), if \( \{p_j : 0 \leq j \leq k+1\} \) generated by Algorithm 2.1 satisfies that

\[(3.6) \quad p_0 \in \mathbb{B}(\bar{p}, \bar{\delta}) \quad \text{and} \quad \{p_j : 1 \leq j \leq k\} \subset \mathbb{B}(\bar{p}, \hat{\delta}), \]

then one has that

\[(3.7) \quad d^2(p_{k+1}, q) \leq d^2(p_k, q) + 2R_k \|v_k\|^2 \leq d^2(p_0, q) + cd(p_0, q) \]

if \( q \in \mathbb{B}(\bar{p}, \hat{\delta}) \) satisfies \( F(q) \leq F(p_{k+1}) \), and that

\[(3.8) \quad p_{k+1} \in \mathbb{B}(\bar{p}, \hat{\delta}) \quad \text{if} \quad F(\bar{p}) \leq F(p_{k+1}). \]

Proof. Noting that any closed ball is compact, we have by [4, p. 166] that the curvatures of the ball \( \mathbb{B}(\bar{p}, r_{\text{cvx}}(\bar{p})) \) are bounded, where \( r_{\text{cvx}}(\bar{p}) \) is the convexity radius at \( \bar{p} \) defined in (2.1). Let \( \kappa < 0 \) be a lower bound of the curvatures of \( \mathbb{B}(\bar{p}, r_{\text{cvx}}(\bar{p})) \). Thanks to assumption (3.5), there exists \( \delta > 0 \) (using a smaller \( \delta \) if necessarily) such that \( F \) is quasi-convex on \( \mathbb{B}(\bar{p}, \delta) \) and that

\[(3.9) \quad \delta < \min \left\{ 1, \frac{r_{\text{cvx}}(\bar{p})}{\sqrt{|\kappa|}} \right\}. \]

Furthermore, let \( L > 0 \) be such that \( \varphi \) is Lipschitz continuous on \( \mathbb{B}(\bar{p}, \delta) \) with constant \( L \) (recalling Proposition 2.10(i)):

\[(3.10) \quad \varphi(p) - \varphi(q) \leq Ld(p, q) \quad \text{for any} \quad p, q \in \mathbb{B}(\bar{p}, \delta). \]

Now set \( c := \frac{2RL}{\delta} \) and choose \( \hat{\delta}, \bar{\delta}, \delta > 0 \) be such that

\[(3.11) \quad \hat{\delta} < \bar{\delta} < \frac{\delta}{2}, \quad (\bar{\delta} + c)\bar{\delta} \leq \bar{\delta}^2 \]

and

\[(3.12) \quad \|v_p\| \leq \frac{\beta \delta}{2R} \quad \text{for any} \quad p \in \mathbb{B}(\bar{p}, \hat{\delta}) \quad \text{and} \quad v_p \in D_{\sigma}(p), \]

(where existence of \( \hat{\delta} \) of the second item of (3.12) is because of Proposition 2.9(iv)). To proceed, we verify that the implication (3.6) \( \implies \) (3.7) holds for any \( k \in \mathbb{N} \), any \( \{p_j : 0 \leq j \leq k + 1\} \) generated by Algorithm 2.1 and any \( q \in \mathbb{B}(\bar{p}, \hat{\delta}) \) satisfying \( F(q) \leq F(p_{k+1}) \). Granting this and assuming that \( F(\bar{p}) \leq F(p_{k+1}) \). Then we estimate by (3.7) (applied to \( \bar{p} \) in place of \( q \) and noting \( d(p_0, \bar{p}) \leq \bar{\delta} \) that

\[(3.13) \quad d^2(p_{k+1}, \bar{p}) \leq d^2(p_0, \bar{p}) + cd(p_0, \bar{p}) \leq (\bar{\delta} + c)\bar{\delta} < \hat{\delta}^2, \]

which implies \( p_{k+1} \in \mathbb{B}(\bar{p}, \hat{\delta}) \). Then, the triple \( (c, \bar{\delta}, \delta) \) is as desired.

Thus to complete the proof, let \( k \in \mathbb{N} \), and let \( \{p_j : 0 \leq j \leq k + 1\} \) be generated by Algorithm 2.1 to satisfy (3.6). Fix \( j \in \{0, 1, \ldots, k\} \), and let
\( \gamma_j \) be the geodesic determined by (2.16). Then, \( d(p_j, \bar{\rho}) \leq \bar{\delta} \) by (3.6) and so \( \|v_j\| \leq \frac{\bar{\delta}}{2} \) by (3.12). Therefore it follows from (2.22) that, for any \( t \in [0, t_j] \),

\[
d(\gamma_j(t), \bar{\rho}) \leq d(\gamma_j(t), p_j) + d(p_j, \bar{\rho}) < t \|v_j\| + \bar{\delta} \leq \frac{\bar{\delta}}{2} + \frac{1}{2} \bar{\delta} < \delta,
\]

(noting that \( t \leq t_j \leq R \)) and then one has that

\[
\gamma_j([0, t_j]) \subseteq \text{int}\mathbb{B}(\bar{\rho}, \bar{\delta}) \subseteq \mathbb{B}(\bar{\rho}, \text{cvx}(\bar{\rho})�.
\]

Now let \( q \in \mathbb{B}(\bar{\rho}, \bar{\delta}) \) be such that \( F(q) \leq F(p_{k+1}) \). Then, we have that

\[
d(p_j, q) \leq d(p_j, \bar{\rho}) + d(q, \bar{\rho}) \leq 2\bar{\delta} < \delta.
\]

Noting that \( \sqrt{|\kappa|}\delta < 1 \) by the choice of \( \delta \) in (3.9), one has that

\[
h\left(\sqrt{|\kappa|}d(p_j, q)\right) \geq h(\sqrt{|\kappa|}\delta) \geq h(1) > \frac{3}{4}.
\]

Recalling that \( F(q) \leq F(p_j) \) and \( \sqrt{|\kappa|}t_j\|v_j\| \leq \sqrt{|\kappa|}R\|v_j\| \leq 1 \) by (3.9) and (3.12), it follows from (3.2) (with \( t_j, v_j \) and \( p_j \) in place of \( t, v \) and \( p \) that

\[
d^2(\gamma_j(t_j), q) \leq d^2(p_j, q) + \frac{3t_j^2\|v_j\|^2}{2h(\sqrt{|\kappa|}d(p_j, q))} \leq d^2(p_j, q) + 2Rt_j\|v_j\|^2,
\]

where the last inequality holds by (3.14) and \( t_j \leq R \). Since \( p_{k+1} = \gamma_k(t_k) \), it follows that

\[
d^2(p_{k+1}, q) \leq d^2(p_k, q) + 2Rt_k\|v_k\|^2.
\]

Moreover, we first estimate by (2.20) that

\[
\sum_{l=0}^{k} t_l\|v_j\|^2 \leq \sum_{l=0}^{k} \frac{2(\varphi(p_l) - \varphi(p_{l+1}))}{\beta} = \frac{2(\varphi(p_0) - \varphi(p_{k+1}))}{\beta} \leq \frac{2(\varphi(p_0) - \varphi(q))}{\beta},
\]

where the last inequality holds because \( \varphi(q) \leq \varphi(p_{k+1}) \) by \( F(q) \leq F(p_{k+1}) \) and Proposition 2.10(ii). Summing up the inequalities in (3.15) over \( 0 \leq j \leq k - 1 \), one concludes that

\[
d^2(p_k, q) + 2Rt_k\|v_k\|^2 \leq d^2(p_0, q) + \frac{2R}{\beta} (\varphi(p_0) - \varphi(q))\).
\]

This, together with (3.10), implies that

\[
d^2(p_k, q) + 2Rt_k\|v_k\|^2 \leq d^2(p_0, q) + cd(p_0, q)
\]

Thus (3.7) is seen to hold by (3.16), showing the implication. The proof is complete.

Now, we are ready to establish local convergence of Algorithm 2.1 under locally quasi-convex assumption.
Theorem 3.5. Let \( \bar{p} \in M \) be such that assumption (3.5) holds. Then, for any \( \delta > 0 \), there exist \( \hat{\delta}, \delta > 0 \) satisfying \( \hat{\delta} < \delta < \frac{\delta}{2} \) such that, for any sequence \( \{p_k\} \) generated by Algorithm 2.1 with initial point \( p_0 \in \mathbb{B}(\bar{p}, \hat{\delta}) \), if it satisfies

\[
(3.19) \quad \lim_{k \to +\infty} F(p_k) \preceq F(\bar{p}),
\]

then one has the following assertions:

(i) The sequence \( \{p_k\} \) stays in \( \mathbb{B}(\bar{p}, \hat{\delta}) \) and converges to a point \( p^* \).

(ii) If it is additionally assumed that \( \{t_k\} \) has a positive lower bound or that \( \{t_k\} \) satisfies the Armijo step sizes, then \( p^* \) is a critical point of \( F \).

Proof. By the assumed (3.5), Lemma 3.4 is applicable. Thus, for any \( \delta > 0 \), there exist \( \bar{\delta} < \hat{\delta} < \delta \) such that, for any sequence \( \{p_k\} \) generated by Algorithm 2.1 if it satisfies (3.6) then (3.8) holds (for any \( k \)); hence the following implication holds for each \( k \in \mathbb{N} \):

\[
(3.20) \quad [(3.6) \text{ and } (3.19) \text{ hold}] \implies p_{k+1} \in \mathbb{B}(\bar{p}, \hat{\delta}).
\]

Now, let \( \{p_k\} \) be a sequence generated by Algorithm 2.1 with initial point \( p_0 \in \mathbb{B}(\bar{p}, \hat{\delta}) \) such that (3.19) holds. Then one checks by (3.20) (applied to \( k = 0 \)) that \( p_1 \in \mathbb{B}(\bar{p}, \hat{\delta}) \), and concludes by mathematical induction that \( \{p_k\} \subset \mathbb{B}(\bar{p}, \hat{\delta}) \), showing the first conclusion of assertion (i). Consequently, the sequence \( \{p_k\} \) has at least one cluster point, say \( p^* \). Letting \( L_{\bar{\delta}} := \{p \in \mathbb{B}(\bar{p}, \hat{\delta}) : F(p) \preceq \inf_{k \in \mathbb{N}} F(p_k)\} \), one sees that \( p^* \in L_{\bar{\delta}} \) since \( \{F(p_k)\} \) is decreasing and \( F \) is continuous on \( \mathbb{B}(\bar{p}, \hat{\delta}) \) (using a smaller \( \delta \) if necessary). Then, (3.7) holds for each \( q \in L_{\bar{\delta}} \). Thanks to \( \sum_{k=1}^{\infty} t_k \|v_k\|^2 < +\infty \) by (2.23), we get that \( \{p_k\} \) is quasi-Fejér convergent to \( L_{\bar{\delta}} \). Hence, we conclude by Lemma 3.3 that \( \lim_{k \to \infty} p_k = p^* \) (recalling \( p^* \in L_{\bar{\delta}} \)). Thus, the second conclusion of assertion (i) is seen to hold.

Assertion (ii) is a direct consequence of assertion (i) and Proposition 2.11(iii). This completes the proof. \( \square \)

4. Local linear convergence without locally quasi-convex assumption

To study the linear convergence property, we need the following Kurdyka-Lojasiewicz-like property. Let \( \bar{p} \) be a locally weak Pareto optimum of \( F \). Consider the following condition on some ball \( \mathbb{B}(\bar{p}, r) \) with some constant \( \alpha > 0 \):

\[
(4.1) \quad \|v(p)\|^2 \geq \alpha \varphi(p) \quad \text{for each } p \in \mathbb{B}(\bar{p}, r),
\]

where \( v(p) \) is the steepest descent direction at \( p \) given by (2.7) and \( \varphi \) is defined by (2.19). Our second main result in this subsection is on the linear convergence property of Algorithm 2.1 without locally quasi-convex assumption. Note that, to guarantee the linear convergence, it is required in Theorem 4.1 that the corresponding step sizes \( \{t_k\} \) have a positive lower
bound, which is satisfied by the Armijo step sizes in the case when $JF(\cdot)$ is Lipschitz continuous around $\bar{p}$; see Lemma 4.3 below.

**Theorem 4.1.** Let $\bar{p} \in M$ be a weak Pareto optimum of (2.4) such that

$$d^2(p_k, p^*) \leq \mu \varphi(p_k) \leq \mu \rho^{2k} \varphi(p_0) \quad \text{for each } k \in \mathbb{N},$$

where $\mu := \frac{2R}{(1-\rho)^2}$ and $\rho := \sqrt{1 - \frac{\alpha \beta (1-\sqrt{\sigma})^2}{2}}$.

**Proof.** Note by definition that $\varphi(\bar{p}) = 0$. Recalling that $\varphi$ is continuous on $M$, one can choose $\delta > 0$ small enough such that $\delta < r$ and

$$\frac{1}{1-\rho} \sqrt{\frac{2R \varphi(p)}{\beta}} \leq r - \delta \quad \text{for all } p \in B(\bar{p}, \delta).$$

Below we show that $\delta$ is as desired. To this end, let $\{p_k\}$ be a sequence generated by Algorithm 2.1 with initial point $p_0 \in B(\bar{p}, \delta)$. Then by step 3 of Algorithm 2.1 and (2.20), the following relation holds for each $k, l \in \mathbb{N}$,

$$d^2(p_{k+l+1}, p_{k+l}) \leq R t_{k+l} \|v_{k+l}\|^2 \leq \frac{2R(\varphi(p_{k+l}) - \varphi(p_{k+l+1}))}{\beta} \leq \frac{2R \varphi(p_{k+l})}{\beta}$$

(notating $t_k \in (0, R]$ and $\varphi(p) \geq 0$ for all $p \in M$). We first show inductively that

$$\{p_k\} \subseteq B(\bar{p}, r).$$

Clearly, (4.6) holds for $k = 0$. Now assume that

$$\{p_j : j = 0, 1, \ldots, k\} \subseteq B(\bar{p}, r).$$

Then, it follows from (2.10) and (4.2) that

$$\|v_j\|^2 \geq (1 - \sqrt{\sigma})^2 \|v(p_j)\| \geq \alpha(1 - \sqrt{\sigma})^2 \varphi(p_j) \quad \text{for all } j = 0, 1, \ldots, k.$$

Hence, for all $j = 0, 1, \ldots, k$, one checks from (2.20) that

$$\varphi(p_{j+1}) \leq \varphi(p_j) - \frac{\beta t_j}{2} \|v_j\|^2 \leq \varphi(p_j) - \frac{\alpha(1 - \sqrt{\sigma})^2 \beta t_j}{2} \varphi(p_j) \leq \rho^2 \varphi(p_j).$$

Thus, we get that

$$\varphi(p_j) \leq \rho^{2j} \varphi(p_0) \quad \text{for all } j = 0, 1, \ldots, k.$$

This, together with (4.3), implise that for all $j = 0, 1, \ldots, k$,

$$d^2(p_{j+1}, p_j) \leq \rho^{2j} \frac{2R \varphi(p_0)}{\beta}.$$
and so
\[
\begin{align*}
    d(p_{k+1}, \bar{p}) & \leq \sum_{j=0}^{k} d(p_{j+1}, p_{j}) + d(p_{0}, \bar{p}) \\
    & \leq \sum_{j=0}^{k} \sqrt{\rho_{j}^2 \frac{2R\varphi(p_{j})}{\beta}} + d(p_{0}, \bar{p}) \\
    & \leq \frac{1}{1-\rho} \sqrt{\frac{2R\varphi(p_{0})}{\beta}} + d(p_{0}, \bar{p}) \\
    & \leq r - \delta + \delta = r,
\end{align*}
\]
where the last inequality is by the choice of \(\delta\) (see (4.4)). Thus, (4.6) is valid by mathematical induction. Furthermore, by the arguments for proving (4.6), we see that (4.8) and (4.9) hold for all \(j \in \mathbb{N}\). Hence the following relations hold for each \(k, l \in \mathbb{N}\):
\[
\begin{align*}
    & \varphi(p_{k+l}) \leq \rho^{2l} \varphi(p_{k}) \quad \text{and} \quad \varphi(p_{k}) \leq \rho^{2k} \varphi(p_{0}).
\end{align*}
\]
Recalling \(\varphi(p_{k}) \geq 0\) for each \(k\), there holds
\[
\lim_{k \to \infty} \varphi(p_{k}) = 0.
\]
Combing (4.5) and (4.10) yields that
\[
\begin{align*}
    d(p_{k+l+1}, p_{k+l}) & \leq \rho^{l} \sqrt{\frac{2R\varphi(p_{k})}{\beta}} \quad \text{for any} \ k, l \in \mathbb{N},
\end{align*}
\]
and then
\[
\begin{align*}
    d(p_{k+l}, p_{k}) & \leq \sum_{j=1}^{l} d(p_{k+j}, p_{k+j-1}) \leq \frac{1}{1-\rho} \sqrt{\frac{2R\varphi(p_{k})}{\beta}}.
\end{align*}
\]
Thus, in view of (4.11), the sequence \(\{p_{k}\}\) is a Cauchy sequence, and then \(\{p_{k}\}\) converges to some point \(p^{*}\) satisfying \(\varphi(p^{*}) = 0\) (noting that \(\varphi\) is continuous), and so \(p^{*}\) is a weak Pareto optimum of (2.4). Letting \(l\) goes to infinite in (4.12) and noting the second item of (4.10), we have that
\[
\begin{align*}
    d(p_{k}, p^{*}) & \leq \frac{1}{1-\rho} \sqrt{\frac{2R\varphi(p_{k})}{\beta}} \leq \frac{1}{1-\rho} \sqrt{\frac{2R\varphi(p_{0})}{\beta}} \rho^{k}.
\end{align*}
\]
Hence, (4.3) is seen to hold, completing the proof.

Remark 4.2. Theorem 4.1 establishes the linear convergence property of Algorithm 2.1 without locally quasi-convex assumption, which seems new even in linear spaces setting. Furthermore, in the case when the multiobjective optimization is reduced to scalar optimization (i.e., \(I = \{1\}\)), our result improves sharply the corresponding result in [24] in the sense that we remove the local quasi-convexity assumption.

The following lemma provides a sufficient condition for the step size sequence \(\{t_{k}\}\) generated by the Armijo step sizes to have a positive lower bound.

Lemma 4.3. Let \(\bar{p} \in M\) be such that assumption (3.5) holds, and suppose that \(JF(\cdot)\) is Lipschitz continuous around \(\bar{p}\). Then, there exist \(\delta > 0\) and \(t > 0\) such that, for any \(p_{0} \in B(\bar{p}, \delta)\), if Algorithm 2.1 employs the Armijo step sizes and the generated sequence \(\{p_{k}\}\) satisfies (3.19), then the generated step sizes \(\{t_{k}\}\) satisfies that \(\inf_{k \in \mathbb{N}} t_{k} \geq t\).
Proof. By assumption, Theorem 3.5 is applicable to getting that, for any $\delta > 0$, there exist $\delta, \hat{\delta} > 0$ satisfying $\delta < \hat{\delta} < \frac{\delta}{2}$ with the property stated there. Without loss of generality, we may assume further that $3\nu^{-1}\hat{\delta} < r_{cvx}(\bar{p})$, and there exists $L > 0$ such that for each $i \in I$,

\begin{equation}
\|\nabla f_i(p) - P_{p,q}\nabla f_i(q)\| \leq Ld(p,q) \quad \text{for any } p, q \in B(\bar{p}, 3\nu^{-1}\hat{\delta})
\end{equation}

(where $\nu$ is chosen by the Armijo step size rule (2.18)).

Let $t := \min \left\{ \nu, \frac{\nu(1-\beta)}{2L} \right\}$. Below, we show that $t, \hat{\delta}$ are as desired. To do this, let $p_0 \in B(\bar{p}, \delta)$, and let $\{t_k\}$ and $\{p_k\}$ be the generated Armijo step sizes and the generated sequence by Algorithm 2.1 with initial point $p_0$, respectively. Now fix $k$ and assume that $t_k \leq \nu$. Then, by (2.18), we see that there exists $i \in I$ such that

\begin{equation}
f_i(\gamma_k(\nu^{-1}t_k)) - f_i(p_k) \geq \nu^{-1}\beta t_k \langle \nabla f_i(p_k), v_k \rangle.
\end{equation}

Noting that $B(\bar{p}, \delta)$ is strongly convex, one sees that $\gamma_k([0, t_k])$ is the unique minimal geodesic joining $p_k$ to $p_{k+1}$. Therefore $t_k \|v_k\| = d(p_k, p_{k+1})$, and it follows that

$$d(p_k, \gamma_k(\nu^{-1}t_k)) \leq \nu^{-1}t_k \|v_k\| = \nu^{-1}d(p_k, p_{k+1}) \leq 2\nu^{-1}\hat{\delta},$$

(see Theorem 3.5(i) for the last inequality). Thus, using the triangle inequality and noting that $1 < \nu^{-1}$, one checks that $\gamma_k(\nu^{-1}t_k) \in B(\bar{p}, 3\nu^{-1}\hat{\delta})$ because

$$d(\bar{p}, \gamma_k(\nu^{-1}t_k)) \leq d(\bar{p}, p_k) + \nu^{-1}d(p_k, p_{k+1}) \leq 3\nu^{-1}\hat{\delta}.$$ 

Using the mean value theorem, we can choose $\bar{t}_k \in (0, t_k)$ to satisfy that

\begin{equation}
f_i(\gamma_k(\nu^{-1}t_k)) - f_i(p_k) = \langle \nabla f_i(\gamma_k(\nu^{-1}\bar{t}_k)), \nu^{-1}t_k P_{\gamma_k, \gamma_k(\nu^{-1}\bar{t}_k), p_k} v_k \rangle.
\end{equation}

Since

$$\langle \nabla f_i(\gamma_k(\nu^{-1}\bar{t}_k)), P_{\gamma_k, \gamma_k(\nu^{-1}\bar{t}_k), p_k} v_k \rangle = \langle P_{\gamma_k, \gamma_k(\nu^{-1}\bar{t}_k)} \nabla f_i(\gamma_k(\nu^{-1}\bar{t}_k)) - \nabla f_i(p_k), v_k \rangle + \langle \nabla f_i(p_k), v_k \rangle \leq \nu^{-1}t_k L \|v_k\|^2 + \|\nabla f_i(p_k)\| \cdot \|v_k\| + \langle \nabla f_i(p_k), v_k \rangle,$$

where the last inequality holds by (1.13) (as $\gamma_k(\nu^{-1}\bar{t}_k) \in B(\bar{p}, 3\nu^{-1}\hat{\delta})$), it follows from (1.15) that

$$f_i(\gamma_k(\nu^{-1}t_k)) - f_i(p_k) \leq \nu^{-1}t_k(\nu^{-1}t_k L \|v_k\|^2 + \langle \nabla f_i(p_k), v_k \rangle).$$

Combining this and (4.14), we conclude that

$$0 \leq \nu^{-1}t_k L \|v_k\|^2 + (1 - \beta) \langle \nabla f_i(p_k), v_k \rangle.$$

Hence, it follows from (2.9) that

$$0 \leq \left( \nu^{-1}t_k L + (1 - \beta) \left( -\frac{1}{2} \right) \right) \|v_k\|^2.$$
This implies that $t_k \geq \frac{\nu(1-\beta)}{2L}$ (in the case when $t_k \leq \nu$), and so $\inf_{k\in\mathbb{N}} t_k \geq \min\left\{\nu, \frac{\nu(1-\beta)}{2L}\right\}$ as desired to show.

□

5. Global convergence

The following theorem regards the global convergence and the linear convergence of Algorithm 2.1. We emphasize that the convergence result as well as the linear convergence rate of Algorithm 2.1 is independent of the curvatures of $M$.

**Theorem 5.1.** Suppose that the sequence $\{p_k\}$ generated by Algorithm 2.1 has a cluster point $\bar{p}$. Then, the following assertions hold:

(i) If (3.5) holds, then $\{p_k\}$ converges to $\bar{p}$.

(ii) If $\bar{p}$ is a weak Pareto optimum of (2.4), $\inf_{k\geq0} \{t_k\} > 0$ and assumption (4.2) holds, then $\{p_k\}$ converges linearly to $\bar{p}$.

**Proof.** Noting that (3.19) is naturally satisfied as $\{F(p_k)\}$ is non-increasing monotone and $\bar{p}$ is a cluster point, we get from Theorem 3.5(i) that there exists $\delta > 0$ such that any sequence generated by Algorithm 2.1 with initial point in $B(\bar{p}, \delta)$ is convergent. Now $\bar{p}$ is a cluster point, so there exists some $k_0 \in \mathbb{N}$ such that $p_{k_0} \in B(\bar{p}, \delta)$. Thus, $\{p_k\}$ converges to some point, which in fact equals to $\bar{p}$ and assertion (i) holds.

With a similar argument that we did for assertion (i), but using Theorem 4.1 instead of Theorem 3.5(i), one sees that assertions (ii) holds. The proof is complete.

The following lemma provides some sufficient conditions ensuring the boundedness of the sequence $\{p_k\}$ generated by Algorithm 2.1 (and so the existence of a cluster point). Set

\[ \mathcal{L}_0 := \{p \in M : F(p) \preceq F(p_0)\}. \]

**Lemma 5.2.** Let $\{p_k\}$ be a sequence generated by Algorithm 2.1 with initial point $p_0$. Then, $\{p_k\}$ is bounded provided one of the assumptions (a) and (b) holds:

(a) $\mathcal{L}_0$ is bounded.

(b) $\mathcal{L}_0$ is totally convex with its curvatures being bounded from below and $F$ is quasi-convex on $\mathcal{L}_0$ (e.g., $F$ is quasi-convex on $M$ and $M$ is of lower bounded curvatures).

**Proof.** Note that $\{p_k\} \subseteq \mathcal{L}_0$ as $\{F(p_k)\}$ is non-increasing monotone. Then, $\{p_k\}$ is clearly bounded under assumption (a). Under assumption (b), with a similar argument as in the proof for [38, Theorem 3.7], one can check that $\{p_k\}$ is bounded.

□

The following corollary is immediate from Theorem 5.1 and Lemma 5.2. Particularly, the global convergence result (assertion (i)) under assumption (b) in Lemma 5.2 extends the corresponding one in [38, Theorem 3.7] which was established for the case when Algorithm 2.1 employs the Armijo step.
sizes (noting that in this case any cluster point $\bar{p}$ of a generated sequence satisfies (3.5) by Proposition 2.11(iii)). As for assertion (ii), as far as we know, it is new even in the linear space setting.

**Corollary 5.3.** Suppose that one of assumptions (a) and (b) in Lemma 5.2 holds. Then, any sequence $\{p_k\}$ generated by Algorithm 2.1 has at least a cluster point $\bar{p}$; furthermore, if $\bar{p}$ satisfies (3.5), then assertions (i) and (ii) in Theorem 5.1 hold.

**Acknowledgments 1.** Research of the author was supported in part by the National Natural Science Foundation of China (grant numbers 11661019, 11771397).

**References**

[1] P. A. Absil, R. Mahony and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, 2008.

[2] G. C. Bento, O. P. Ferreira and P. R. Oliveira, *Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds*, J. Optim. Theory Appl., 154 (2012), pp. 88–107.

[3] G. C. Bento, J. X. D. C. Neto and P. S. M. Santo, *An Inexact Steepest Descent Method for Multicriteria Optimization on Riemannian Manifolds*, J. Optim. Theory Appl., 159 (2013), pp. 108–124.

[4] R. L. Bishop and R. J. Crittenden, *Geometry of Manifold*, Academic Press, New York and London, 1964.

[5] H. Bonnel, A. N. Iusem and B. F. Svaiter, *Proximal Methods in Vector Optimization*, Society for Industrial and Applied Mathematics, 2005.

[6] G. Y. Chen, X. Huang and X. Yang, *Vector Optimization*, Springer, Berlin Heidelberg, 2005.

[7] I. Das and J. E. Dennis, *Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems*, SIAM J. Optim., 8 (1998), pp. 631–657.

[8] M. P. DoCarmo, *Riemannian Geometry*, Birkhäuser Boston, Boston MA, 1992.

[9] L. M. G. Drummond and A. N. Iusem, *A projected gradient method for vector optimization problems*, Computational Optimization and Applications, 28 (2004), pp. 5–29.

[10] L. M. G. Drummond, N. Maculan and B. F. Svaiter, *On the choice of parameters for the weighting method in vector optimization*, Math. Program., 111 (2008), pp. 201–216.

[11] L. M. G. Drummond and B. F. Svaiter, *A steepest descent method for vector optimization*, J. Comput. Appl. Math., 175 (2005), pp. 395–414.

[12] O. P. Ferreira, L. R. Lucambo Pérez and S. Z. Németh, *Singularities of monotone vector fields and an extragradient-type algorithm*, J. Global Optim., 31 (2005), pp. 133–151.

[13] O. P. Ferreira and P. R. Oliveira, *Subgradient Algorithm on Riemannian Manifolds*, J. Optim. Theory Appl., 97 (1998), pp. 93–104.

[14] J. Fliege, L. M. Gra na Drummond and B. F. Svaiter, *Newton’s Method for Multiobjective Optimization*, SIAM J. Optim., 20 (2009), pp. 602–626.

[15] J. Fliege and B. F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res., 51 (2000), p. 479–C494.

[16] C. M. Fonseca and P. J. Fleming, *An overview of evolutionary algorithms in multiobjective optimization*, Evolutionary Computation, 3 (1995), pp. 1–16.
[17] E. H. Fukuda and L. M. G. Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization, 60 (2011), pp. 1009–1021.

[18] E. H. Fukuda and L. M. G. Drummond, *Inexact projected gradient method for vector optimization*, Computational Optimization and Applications, 54 (2013), pp. 493–493.

[19] A. M. Geoffrion, *Proper efficiency and the theory of vector maximization*, Journal of Mathematical Analysis and Applications, 22 (1968), pp. 618–630.

[20] J. Jahn, *Scalarization in vector optimization*, Math. Program., 29 (1984), pp. 203–218.

[21] C. Li, B. S. Mordukhovich, J. Wang and J. C. Yao, *Weak sharp minima on Riemannian manifolds*, SIAM J. Optim., 21 (2011), pp. 1523–1560.

[22] C. Li and J. C. Yao, *Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm*, SIAM J. Control Optim., 50 (2012), pp. 2486–2514.

[23] S. L. Li, C. Li, Y.C. Liou and J.C. Yao, *Existence of solutions for variational inequalities on Riemannian manifolds*, Nonlinear Anal., 71 (2009), pp. 5695–5706.

[24] J. H. Wang, X. M. Wang, C. Li and J. C. Yao, *Convergence Analysis of Gradient Algorithms on Riemannian Manifolds Without Curvature Constraints and Application to Riemannian Mass*, SIAM J. Optim., 2021, 31(1), 172-199.

[25] D. T. Luc, *Theory of Vector Optimization. Lecture Notes in Economy and Mathematical Systems*, Springer vol 319, Berlin Heidelberg New York, 1989.

[26] R. E. Mahony, *The constrained Newton method on Lie group and the symmetric eigenvalue problem*, linear Algebra Appl., 248 (1996), pp. 67–89.

[27] K. M. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer, Norwel, 1999.

[28] S. A. Miller and J. Malick, *Newton methods for nonsmooth convex minimization: Connections among U-Lagrangian, Riemannian Newton and SQP methods*, Math. Program., 104 (2005), pp. 609–633.

[29] S. Z. Nemeth, *Five kinds of monotone vector fields*, Pure Math. Appl., 9 (1998), pp. 417–428.

[30] J. H. Ryu and S. Kim, *A derivative-free Trust-Region method for biobjective optimization*, SIAM J. Optim., 24 (2014), pp. 334–362.

[31] E. A. Papa Quiroz, E. M. Quispe, and P. R. Oliveira, *Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds*, J. Math. Anal. Appl., 341 (2008), pp. 467–477.

[32] T. Sakai, *Riemannian Geometry. Translations of Mathematical Monographs*, Am. Math. Soc., Providence, 1996.

[33] S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, phd thesis, Harvard University Cambridge Massachusetts, 1994.

[34] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*. In: Mathematics and Its Applications, Kluwer Academic, Dordrecht, 1994.

[35] J. H. Wang, Y. Hu, C. K. W. Yu, C. Li and X. Yang, *Extended Newton methods for multiobjective optimization: majorizing function technique and convergence analysis*, SIAM J. Optim., 29 (2019), pp. 2388-2421.

[36] J. H. Wang, G. López, V. Martín-Márquez and C. Li, *Monotone and accretive vector fields on Riemannian manifolds*, J. Optim. Theory Appl., 146 (2010), pp. 691–708.

[37] X. M. Wang, *Subgradient algorithms on Riemannian manifolds of lower bounded curvatures*, Optimization, 67 (2018), pp. 179–194.

[38] X. M. Wang, *An Inexact Descent Algorithm for Multicriteria Optimizations on General Riemannian Manifolds*, ?.

[39] X. M. Wang, C. Li, and J. C. Yao, *Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures*, J. Optim. Theory Appl., 164 (2015), pp. 201–217.
[40] X. M. Wang, C. Li, J. H. Wang and J. C. Yao, Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds, SIAM J. Optim., 25 (2015), pp. 2334–2358.

[41] Y. Yang, Globally convergent optimization algorithms on Riemannian manifolds: uniform framework for unconstrained and constrained optimization, J. Optim. Theory Appl., 132 (2007), pp. 245–265.

[42] S. T. Yau, Non-existence of continuous convex functions on certain Riemannian manifolds, Math. Ann., 207 (1974), pp. 269–270.

(X. M. Wang) College of Science, Guizhou University, Guiyang 550025, P. R. China
Email address: xmwang2@gzu.edu.cn

(J. H. Wang) Department of Mathematics, Hangzhou Normal University, Hangzhou 311121, P. R. China
Email address: wangjh@hznu.edu.cn

(C. Li) Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China
Email address: cli@zju.edu.cn