EXTENDED NEWTON METHODS FOR MULTIOBJECTIVE OPTIMIZATION: MAJORIZING FUNCTION TECHNIQUE AND CONVERGENCE ANALYSIS

YAOHUA HU*, CARISA KWOK WAI YU†, JINHUA WANG‡, CHONG LI§, AND XIAOQI YANG¶

Abstract. We consider the extended Newton method for approaching a Pareto optimum of multiobjective optimization problems and establish its quadratic convergence criteria and estimation of radius of convergence ball under the assumption that the Hessians of objective functions satisfy an L-average Lipschitz condition. These convergence theorems significantly improve the corresponding ones in [SIAM J. Optim 20 (2009), pp. 602-626]. As applications of the obtained results, convergence theorems under the classical Lipschitz condition or the γ-condition are presented for multiobjective optimization, and the global quadratic convergence results of the extended Newton method with Armijo/Goldstein/Wolf line-search schemes are also provided.

Key words. Multiobjective optimization; Pareto optimum; Newton method; convergence criteria; L-average Lipschitz condition

AMS subject classifications. Primary, 90C29, 90C30; Secondary, 65K05

1. Introduction.

Let $U \subseteq \mathbb{R}^l$ be an open set, and let $F : U \to \mathbb{R}^m$ be a twice continuously differentiable function. In the present paper, we consider the following multiobjective optimization problem:

$$\min_{x \in U} F(x). \tag{1.1}$$

This type of problems has been widely studied by [3, 6, 21, 23] and extensively applied in various areas such as engineering [7, 11, 25], management science [2, 22, 33, 39] and environmental analysis [5, 17, 27].

Motivated by its extensive applications, a great amount of attention has been attracted to

*College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China (mayhhu@szu.edu.cn).
This author's work was supported in part by the National Natural Science Foundation of China (grant 11601343), Natural Science Foundation of Guangdong (grant 2016A030310038) and Interdisciplinary Innovation Team of Shenzhen University.

†Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong (carisayu@hsmc.edu.hk).
This author's work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China (grant UGC/FDS14/P03/14 and UGC/FDS14/P02/15).

‡Corresponding author. Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, P. R. China (wjh@zjut.edu.cn).
This author's work was supported in part by the National Natural Science Foundation of China (grant 11771397) and Zhejiang Provincial Natural Science Foundation of China (grant LY17A010021, LY17A010006).

§Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China (cli@zju.edu.cn).
This author's work was supported in part by the National Natural Science Foundation of China (grant 11571308) and Zhejiang Provincial Natural Science Foundation of China (grant LY18A010004).

¶Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (mayangxq@polyu.edu.hk).
This author's work was supported in part by the Research Grants Council of Hong Kong (PolyU 152128/17E).

the development of optimization algorithms, and many iterative methods have been proposed to approach a Pareto optimum of multiobjective optimization; see [3, 4, 8, 9, 10, 18, 19, 20, 35, 41] and references therein. Among them, one of the most important methods is the extended Newton method (with Armijo line-search scheme) introduced by Fliege et al. [19], which is an extension of the classical Newton method for solving nonlinear equations (see [40]). Comparing with other iterative methods for multiobjective optimization, it was pointed out in [19] that the extended Newton method enjoys several advantages: (a) it has a fast convergence rate under some mild conditions; (b) its subproblems can be solved effectively; and (c) it does not use a priori weighting factor or any other priori information for the objective functions. Due to these benefits, there is a great demand for further investigating the convergence theory of the extended Newton method, which is formally stated as follows (for undefined notations in the sequel, one can refer to section 2).

Algorithm 1.1.

Step 1. (Initialization) *Choose* $x_0 \in U$ *and* $\sigma \in (0,1)$, *and set* $n := 0$.
Step 2. (Direction search) *Solve the direction search problem* (2.3) *at* x_n *to obtain the search direction* $s(x_n)$ *and the associated value* $\theta(x_n)$.
Step 3. (Stopping criterion) *If* $\theta(x_n) = 0$, *then stop; otherwise, proceed to Step 4.*
Step 4. (Armijo line-search) *Choose* α_n *as the maximal value of* $\{2^{-s} : s \in \mathbb{N}\}$ *such that*

$$x_k + \alpha_k s(x_n) \in U \text{ and } F_j(x_n + \alpha_k s(x_n)) \le F_j(x_n) + \sigma\alpha_n\theta(x_n) \text{ for all } j = 1, \ldots, m.$$

Step 5. (Update) *Define* $x_{n+1} = x_n + \alpha_n s(x_n)$ *and set* $n := n+1$. *Go back to Step 2.*

Under the assumption that each $\nabla^2 F_j(\cdot)$ is positive definite and Lipschitz continuous on a convex subset of U (with a nonempty interior), the authors studied in [19] the convergence issue of Algorithm 1.1 for problem (1.1)and established the quadratic convergence results, which are in particular concerned with three types. The first one is the semi-local convergence theorem, in which the quadratic convergence to a local Pareto optimum is established under the assumptions, depending on a lot of parameters, at the initial point; see [19, Theorem 6.1] for details. The second one is the local convergence theorem (i.e., [19, Corollary 6.2]) that, for each local Pareto optimum x^*, there exists $r > 0$ such that the generated sequence converges to a local Pareto optimum at a quadratic rate whenever the initial point falls in $\mathbf{B}(x^*, r)$. The last one is the global convergence theorem (i.e., [19, Corollary 6.3]), in which the sequence starting from any initial point is shown to converge to a local Pareto optimum at a quadratic rate.

The purpose of the present paper is to continue the theoretical study of the extended Newton method for multiobjective optimization problems. We focus on the case when each $\nabla^2 F_j(\cdot)$ is Lipschitz continuous and develop a new approach to provide the quantitative convergence analysis for the extended Newton methods, not only Algorithm 1.1 but also the one without the line-search scheme (see Algorithm 3.1). Under the classical Lipschitz continuity assumption for the second derivatives $\nabla^2 F_j(\cdot)$, our main results, concerning also the three types of convergence properties mentioned above, are described as follows:

- Our theorem (i.e., Theorem 4.1) regarding the semi-local convergence property provides some explicit convergence criteria, which are only based on the data at an initial point and the Lipschitz constants of the second derivatives $\nabla^2 F_j(\cdot)$ around the initial point, for ensuring the convergence (to a local Pareto optimum) of Algorithms 3.1 and 1.1.

- Our theorem (i.e., Theorem 4.2) regarding the local convergence property provides some explicit estimates, which only depend on the data of a given local Pareto optimum and the Lipschitz constants of the second derivatives $\nabla^2 F_j(\cdot)$ around the local Pareto optimum, for the radii of the convergence balls of Algorithms 3.1 and 1.1.
- Our theorem (i.e., Theorem 4.5) regarding the global convergence property provides some sufficient conditions made on the cluster point for ensuring the global convergence of the extended Newton method not only with the Armijo line-search scheme (i.e., Algorithm 1.1) but also with Goldstein/Wolf line-search schemes (i.e., Algorithm 3.2).
- The results obtained in the present paper, containing the local, semi-local and the global types, provide explicit error estimates for any sequence generated by Algorithm 3.1 or 3.2 (and so Algorithm 1.1) in terms of the corresponding parameters/modulus, which improve the corresponding ones in [19]; see Theorem 6.1 and Corollaries 6.2, 6.3 therein.

Most of results (such as Theorems 3.4, 3.5, 3.8, 3.9 and so on) in the paper are new, and some of them (i.e., Theorems 4.1, 4.2 and 4.5), where less data is required, extend/improve partially the corresponding ones in [19, Theorem 6.1 and Corollaries 6.2, 6.3] as explained in Remark 4.1; in particular, an example is provided to show the case where the convergence result in the present paper (Theorem 4.1) is available but not the one in [19, Theorem 6.1]; see Example 4.1 for details.

Another important extension of the present paper is that the L-average Lipschitz condition is involved to the consideration of the convergence analysis of the extended Newton method. The L-average Lipschitz condition, which includes the classical Lipschitz condition and the γ-condition as special cases, was introduced by Wang [36] to unify and develop the convergence theory of the Newton method for solving an equation in a Banach space; this idea has been used extensively in numerical analysis and optimization problems; see [12, 28, 29, 30] and references therein, but not been found to be applied to study the multiobjective optimization problems. Note that the L-average Lipschitz condition implies the classical Lipschitz condition, but as shown in the theorems (see Theorems 4.1 and 4.2) in the present paper, the convergence criteria and/or the radius of the convergence ball of the extended Newton method depend heavily on the value of the Lipschitz constant on the involved balls. Indeed, as we will see in Example 4.2, one of the main advantages of adopting the L-average Lipschitz condition is that, in the case when the theorem under the classical Lipschitz condition is not available, it provides the possibility to choose a suitable non-negative and monotonically increasing function L such that the convergence theorem, which we established for the general L-average Lipschitz condition, is applicable to guarantee the convergence of the extended Newton method.

It should be remarked that the analysis tool used in the present paper is the majorizing function technique, which deviates significantly from that of [19]. The majorizing function technique has been widely used in the convergence analysis of Newton method for nonlinear equations [12, 13, 14, 16, 36, 37] and scalar optimization [15, 28], which enables us to establish an explicit convergence criterion and provides a precise estimation of the convergence radius.

To the best of our knowledge, this is the first work to develop the majorizing function technique for the convergence analysis of the extended Newton method for multiobjective optimization.

The paper is organized as follows. In section 2, we present the notations and preliminary results to be used in the present paper. The quadratic convergence criterion and the estimation of radius of convergence ball of the extended Newton method for multiobjective optimization problems are provided in section 3, under the L-average Lipschitz condition. In section 4, theorems under the classical Lipschitz condition, the global quadratic convergence results of the extended Newton method and theorems under the γ-condition are presented for multiobjective optimization problems.

2. Notation and preliminary results. The notations used in the present paper are standard in Euclidean spaces. As usual, for $x \in \mathbb{R}^l$ and $r > 0$, let $\mathbf{B}(x, r)$ and $\mathbf{B}[x, r]$ respectively denote the open and closed balls in $\mathbb{R}^l$, and let $\mathbb{R}^m_+$ and $\mathbb{R}^m_{++}$ denote the non-negative orthant and positive orthant of $\mathbb{R}^m$, respectively. The standard simplex in $\mathbb{R}^m$ is denoted by $\mathbf{\Delta}_m$, i.e.,

$$\mathbf{\Delta}_m := \{\lambda \in \mathbb{R}^m_+ : \sum_{i=1}^m \lambda_i = 1\}.$$

Let $\mathbb{R}^{m\times l}$ denote the space of all $m \times l$ matrices, and let I denote the identical matrix in $\mathbb{R}^{l\times l}$. For $M \in \mathbb{R}^{m\times l}$, the range of M is denoted by $\mathbf{R}(M)$. The following lemma regarding the inverses of the perturbations of nonsingular matrix is well-known; see for example [32, p.45].

Lemma 2.1. *Let $A, B \in \mathbb{R}^{l\times l}$ be such that A is invertible and $\|A^{-1}\|\|A - B\| < 1$. Then B is invertible and*

$$\|B^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|A - B\|}.$$

If $A, B \in \mathbb{R}^{n\times n}$ are additionally symmetric, then B is positive definite.

2.1. Preliminary results about multiobjective optimization. In the present paper, we consider the multiobjective optimization problem (1.1) with $U \subseteq \mathbb{R}^l$ being an open (not necessarily convex) set and $F : U \to \mathbb{R}^m$ being a vector-valued function, denoted by

$$F := (F_1, \ldots, F_m)^T, \tag{2.1}$$

where each $F_i : U \to \mathbb{R}$ is a twice continuously differentiable and real-valued function. For a convex subset $V \subseteq U$, F is said to be $\mathbb{R}^m$-convex on V if F_i is convex on V for each $i = 1, \ldots, m$. The following notions are about the Pareto optimum (also named efficient point).

Definition 2.2. *A point $x^* \in U$ is said to be*

(a) *a (global) Pareto optimum of F on U if there does not exist $y \in U$ such that $F(x^*) - F(y) \in \mathbb{R}^m_+$ and $F(y) \neq F(x^*)$;*
(b) *a weak Pareto optimum of F on U if there does not exist $y \in U$ such that $F(x^*) - F(y) \in \mathbb{R}^m_{++}$;*
(c) *a local Pareto optimum (resp. local weak Pareto optimum) if there exists a neighborhood $V \subseteq U$ of x^* such that x^* is a Pareto optimum (resp. weak Pareto optimum) of F on V.*

Obviously, every Pareto optimum is also a weak Pareto optimum, and each local Pareto optimum is a (global) Pareto optimum if U is convex and F is $\mathbb{R}^m$-convex on U.

For each $i \in \mathbb{N} := \{1, 2, \dots\}$, $C^i(U, \mathbb{R}^m)$ denotes the set of i-th continuously differentiable functions from U to $\mathbb{R}^m$. Let $x \in U$, $f \in C^2(U, \mathbb{R})$ and $F \in C^2(U, \mathbb{R}^m)$ given by (2.1). We use $\nabla f(x) \in \mathbb{R}^n$ and $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ to denote the gradient of f and the Hessian of f at x, respectively; while, the Jacobian of F and the second derivative of F at x are denoted by $\mathrm{D}F(x)$ and $\mathrm{D}^2F(x)$, respectively, that is,

$$\mathrm{D}F(x) = (\nabla F_1(x), \dots, \nabla F_m(x))^T \quad \text{and} \quad \mathrm{D}^2F(x) = (\nabla^2 F_1(x), \dots, \nabla^2 F_m(x))^T.$$

We say that $\mathrm{D}^2F(x)$ is positive definite if so is each $\nabla^2 F_i(x)$.

The notion of a critical point is recalled in the following definition, which characterizes a necessary (but in general not sufficient) condition for Pareto optimality and was used in [20] and [19] to investigate a steepest descent algorithm and an extended Newton method for multiobjective optimization, respectively.

Definition 2.3. *A point $\bar{x} \in U$ is said to be a critical point of F if* $\mathbf{R}(\mathrm{D}F(\bar{x})) \cap (-\mathbb{R}^m_{++}) = \emptyset$.

Note that, in the case when $m = 1$, $\mathbf{R}(\mathrm{D}F(\bar{x})) \cap (-\mathbb{R}^m_{++}) = \emptyset$ is reduced to the classical optimality condition of scalar optimization. It follows from [19, Theorem 3.1] that, if $F \in C^2(U, \mathbb{R}^m)$ and $x^* \in U$ is such that $\mathrm{D}^2F(x^*)$ is positive definite, then

$$x^* \text{ is a critical point of } F \quad \Leftrightarrow \quad x^* \text{ is a local Pareto optimum of } F. \tag{2.2}$$

Following [19], associated to (1.1), we consider, for a point $x \in U$ such that $\mathrm{D}^2F(x)$ is positive definite, the following optimization problem:

$$\min_{s\in\mathbb{R}^n} \max_{j=1,\dots,m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s, \tag{2.3}$$

the solution of which is the Newton direction of the extended Newton method. Clearly, the function $s \mapsto \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s$ is strongly convex for each $j = 1, \dots, m$, and so, problem (2.3) has a unique minimizer. Since problem (2.3) can be framed as a convex quadratic optimization problem, it can be solved effectively. Let $V \subseteq U$ be convex such that $\mathrm{D}^2F(x)$ is positive definite for each $x \in V$. We use the functions $s : V \to \mathbb{R}^n$ and $\theta : V \to \mathbb{R}$ to denote the unique minimizer and the minimal value of problem (2.3), respectively, that is, for each $x \in V$,

$$s(x) := \arg\min_{s\in\mathbb{R}^n} \max_{j=1,\dots,m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s, \tag{2.4}$$

$$\theta(x) := \min_{s\in\mathbb{R}^n} \max_{j=1,\dots,m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s. \tag{2.5}$$

By the KKT optimality condition for problem (2.3), for each $x \in V$, there exist parameters $\lambda (:= \lambda(x)) \in \mathbf{\Delta}_m$ such that (see [19] for details)

$$s(x) = -\left[\sum_{j=1}^m \lambda_j(x) \nabla^2 F_j(x)\right]^{-1} \sum_{j=1}^m \lambda_j(x) \nabla F_j(x). \tag{2.6}$$

We end this subsection by recalling in the following lemmas some useful properties of the functions $s(x)$ and $\theta(x)$. Lemma 2.4 is taken from [19, Lemma 3.2].

LEMMA 2.4. *Let $V \subseteq U$ be convex and let $\bar{x} \in V$. Suppose that $\mathrm{D}^2 F(\bar{x})$ is positive definite. Then the following statements are true.*

(i) $\theta(\bar{x}) \leq 0$.
(ii) *$\bar{x}$ is not a critical point* $\Leftrightarrow$ $[\theta(\bar{x}) < 0]$ $\Leftrightarrow$ $[s(\bar{x}) \neq 0]$.
(iii) *If $\mathrm{D}^2 F(x)$ is positive definite for each $x \in V$, then s is bounded on any compact subset of V and θ is continuous on V.*

Let $F := (F_1, \ldots, F_m)^T \in C^2(U, \mathbb{R}^m)$. Throughout the whole paper, we define

$$F_\lambda(\cdot) := \sum_{j=1}^{m} \lambda_j F_j(\cdot) \quad \text{for each } \lambda := (\lambda_1, \ldots, \lambda_m)^T \in \mathbf{\Delta}_m. \tag{2.7}$$

Let $\lambda \in \mathbf{\Delta}_m$ and $x \in U$, and let $\rho_{\min}(\lambda, x)$ and $\rho_{\max}(\lambda, x)$ denote the minimum and maximum eigenvalues of the matrix $\nabla^2 F_\lambda(x)$, respectively, that is,

$$\rho_{\min}(\lambda, x) := \min\{z^T \nabla^2 F_\lambda(x) z| \ \ \|z\| = 1\} = \|\nabla^2 F_\lambda(x)^{-1}\|^{-1}$$

and

$$\rho_{\max}(\lambda, x) := \max\{z^T \nabla^2 F_\lambda(x) z| \ \ \|z\| = 1\} = \|\nabla^2 F_\lambda(x)\|. \tag{2.8}$$

Relation (2.9) and the first inequality of (2.10) in the following lemma are known in [19, Lemmas 4.2 and 4.3]; while the second inequality of (2.10) is a direct consequence of the first inequality of (2.9) and the first inequality of (2.10).

LEMMA 2.5. *Let $x \in U$ and let $\lambda \in \mathbf{\Delta}_m$ be such that $\nabla^2 F_\lambda(x)$ is positive definite. Then the following relations hold:*

$$\frac{\rho_{\min}(\lambda, x)}{2} \|s(x)\|^2 \leq |\theta(x)| \leq \frac{\rho_{\max}(\lambda, x)}{2} \|s(x)\|^2, \tag{2.9}$$

$$|\theta(x)| \leq \frac{1}{2} \|\nabla^2 F_\lambda(x)^{-1}\| \|\nabla F_\lambda(x)\|^2 \quad \textit{and} \quad \|s(x)\| \leq \|\nabla^2 F_\lambda(x)^{-1}\| \|\nabla F_\lambda(x)\|. \tag{2.10}$$

2.2. Preliminary results about majorizing function. To study the convergence properties of the extended Newton method for multiobjective optimization, we first recall some auxiliary results of a majorizing function. The majorizing function, originally introduced by Wang [36], is a powerful tool for the study of convergence criteria of the Newton method. Let $R > 0$ and let $L : [0, R) \to \mathbb{R}_+$ be a nondecreasing and integrable function. Let $a > 0$ satisfy

$$\frac{1}{R} \int_0^R L(u)(R - u) \mathrm{d}u > \frac{1}{a}. \tag{2.11}$$

Associated to the triple $(a, \beta; L)$, we define the pair of positive constants (r_a, b_a) and the majorizing function $h_a : [0, R) \to \mathbb{R}$ by

$$a \int_0^{r_a} L(u)\mathrm{d}u = 1, \qquad b_a = a \int_0^{r_a} L(u)u\mathrm{d}u. \tag{2.12}$$

and

$$h_a(t) := \beta - t + a \int_0^t L(u)(t-u)\mathrm{d}u \quad \text{for each } t \in [0, R], \tag{2.13}$$

respectively. Clearly, $b_a < r_a < R$ and h_a is twice differentiable on $[0, R)$ with its derivatives being given by

$$h_a'(t) = a \int_0^t L(u)\mathrm{d}u - 1 \quad \text{and} \quad h_a''(t) = aL(t) \quad \text{for each } t \in [0, R), \tag{2.14}$$

where and throughout the whole paper, $h_a'(0)$ means the right derivative of h_a at 0.

Let $\{t_{a,n}\}$ denote a sequence generated by the classical Newton method for approaching the zeros of the majorizing function h_a with the initial value $t_{a,0} = 0$. That is,

$$t_{a,n+1} := t_{a,n} - h_a'(t_{a,n})^{-1} h_a(t_{a,n}) \quad \text{for each } n \in \mathbb{N}. \tag{2.15}$$

Some useful properties of the majorizing function h_a and the sequence $\{t_{a,n}\}$ are presented in the following proposition, in which (i) is taken from [36, Lemma 1.2], while (ii) is well-known in the literature of the Newton method (cf. [36]).

PROPOSITION 2.6. *Suppose that $0 \le \beta \le b_a$. Then, the following assertions are true.*

(i) *h_a is strictly decreasing on $[0, r_a]$ and strictly increasing on $[r_a, R]$ with*

$$h_a(\beta) > 0, \quad h_a(r_a) = \beta - b_a \le 0 \quad \text{and} \quad h_a(R) > \beta > 0.$$

Moreover, if $\beta < b_a$, then h_a has two zeros r_a^* and r_a^{**} such that

$$\beta < r_a^* < \frac{r_a}{b_a}\beta < r_a < r_a^{**}; \tag{2.16}$$

if $\beta = b_a$, then h_a has a unique zero $r_a^* \in (\beta, R)$ (in fact, $r_a^* = r_a$).

(ii) *$\{t_{a,n}\}$ is monotonically increasing and converges to r_a^*.*

(iii) *If $\beta < b_\alpha$, then*

$$\lim_{n\to\infty} \frac{2t_{a,n+1} - t_{a,n} - r_a^*}{t_{a,n+1} - t_{a,n}} = 1 \ \text{ and } \ \overline{\lim}_{n\to\infty} \frac{r_a^* - t_{a,n+1}}{(2t_{a,n+1} - t_{a,n} - r_a^*)^2} \le -\frac{aL(r_a^*)}{2h'(r_a^*)}. \tag{2.17}$$

Proof. To complete the proof, we only need to show assertion (iii). For simplicity, we omit the first subscript a in the sequence $\{t_{a,n}\}$, namely, write $\{t_n\}$ for $\{t_{a,n}\}$. Then, one has by (2.15) and assertion (ii) of this proposition that

$$\lim_{n\to\infty} \frac{2t_{n+1} - t_n - r_a^*}{t_{n+1} - t_n} = 2 + \lim_{n\to\infty} \frac{1}{-h_a'(t_n)^{-1} \frac{h_a(t_n) - h_a(r_a^*)}{t_n - r_a^*}} = 1,$$

that is, the equality of (2.17) holds. On the other hand, note again by (2.15) that

$$\begin{aligned} r_a^* - t_{n+1} &= r_a^* - t_n + h_a'(t_n)^{-1} h_a(t_n) \\ &= -h_a'(t_n)^{-1} \int_0^1 [h_a'(t_n + t(r_a^* - t_n)) - h_a'(t_n)](r_a^* - t_n)\mathrm{d}t \\ &= -h_a'(t_n)^{-1} \int_0^1 \int_0^1 h_a''(t_n + \tau t(r_a^* - t_n)) t(r_a^* - t_n)\mathrm{d}\tau(r_a^* - t_n)\mathrm{d}t \\ &\le -h_a'(t_n)^{-1} \tfrac{aL(r_a^*)}{2}(r_a^* - t_n)^2, \end{aligned}$$

where the inequality holds because $h_a'(t_n) < 0$ (cf. (2.12) and (2.14)), $h_a''(\cdot) = aL(\cdot)$ (cf. (2.14)) and $L(\cdot)$ is nondecreasing. Then, we obtain

$$\frac{r_a^* - t_{n+1}}{(2t_{n+1} - t_n - r_a^*)^2} \le \frac{-h_a'(t_n)^{-1} \frac{aL(r_a^*)}{2}(r_a^* - t_n)^2}{(-2h_a'(t_n)^{-1} h_a(t_n) + t_n - r_a^*)^2} = \frac{-h_a'(t_n)^{-1} \frac{aL(r_a^*)}{2}}{(-2h_a'(t_n)^{-1} \frac{h_a(t_n) - h_a(r_a^*)}{t_n - r_a^*} + 1)^2},$$

and thus, the inequality of (2.17) is seen to hold. The proof is complete. □

The following lemma is useful for the convergence analysis of Newton method and is taken from [36, pp. 175]. Recall that $R > 0$ and $L : [0, R) \to \mathbb{R}_+$ is a nondecreasing and integrable function.

LEMMA 2.7. *Let $0 \le \zeta < R$, and let $\varphi : (0, R - \zeta) \to \mathbb{R}_+$ be defined by*

$$\varphi(t) := \frac{1}{t^2} \int_0^t L(\zeta + u)(t - u)\mathrm{d}u \quad \textit{for each } 0 < t < R - \zeta.$$

Then, φ is increasing on $(0, R - \zeta)$.

3. Convergence analysis of the extended Newton method. This section aims to establish the quadratic convergence criterion of the extended Newton method (without or with line-search scheme) for multiobjective optimization under an L-average Lipschitz condition. The extended Newton method without line-search scheme for solving the multiobjective optimization problem (1.1) is formally stated as follows.

ALGORITHM 3.1.

Step 1. *Choose $x_0 \in U$ and set $n := 0$.*
Step 2. *Solve problem (2.3) at x_n to obtain $s(x_n)$ as in (2.4).*
Step 3. *Update $x_{n+1} := x_n + s(x_n)$ and set $n := n + 1$. Go back to Step 2.*

Below, we propose the extended Newton method with line-search scheme. For this purpose, three kinds of the typical line-search rule for selecting the stepsize sequence $\{\alpha_n\}$ for Algorithm 3.2 are stated below, which have been widely used for optimization algorithms in the literature; see, e.g., [1, 26, 31].

DEFINITION 3.1. *Let $\sigma \in (0, 1)$ and let $\beta_1, \beta_2 \in (\sigma, 1)$. Given $n \ge 0$ and x_n. Let $s(x_n)$ and $\theta(x_n)$ be given by (2.4) and (2.5), respectively. A stepsize $\alpha_n \in (0, +\infty)$ such that $x_n + \alpha_n s(x_n) \in U$ is said to satisfy*

(i) *the Armijo rule if*

$$F_j(x_n + \alpha_n s(x_n)) \le F_j(x_n) + \sigma\alpha_n\theta(x_n) \quad \textit{for all } j = 1, \ldots, m, \tag{3.1}$$

and

$$\alpha_n := \max\{2^{-i} : i \in \mathbb{N},\ (3.1) \text{ holds with } 2^{-i} \text{ in place of } \alpha_n\};$$

(ii) *the Goldstein rule if* (3.1) *holds and*

$$F_j(x_n + \alpha_n s(x_n)) \geq F_j(x_n) + \beta_1 \alpha_n \theta(x_n) \quad \text{for all } j = 1, \ldots, m;$$

(iii) *the Wolf rule if* (3.1) *holds and*

$$\nabla F_j(x_n + \alpha_n s(x_n))^T s(x_n) \geq \beta_2 \theta(x_n) \quad \text{for all } j = 1, \ldots, m.$$

The extended Newton method with line-search scheme for solving the multiobjective optimization problem (1.1) is formally stated as follows.

Algorithm 3.2.

Step 1. *Choose* $x_0 \in U$, $\sigma \in (0,1)$, $\beta_1, \beta_2 \in (\sigma, 1)$ *and set* $n := 0$.
Step 2. *Solve problem* (2.3) *at* x_n *to obtain* $s(x_n)$ *and* $\theta(x_n)$ *as in* (2.4) *and* (2.5), *respectively.*
Step 3. *If* $\theta(x_n) = 0$, *then stop. Otherwise, proceed to Step 4.*
Step 4. *If* $x_n + s(x_n) \in U$ *and*

$$F_j(x_n + s(x_n)) \leq F_j(x_n) + \sigma\theta(x_n) \quad \text{for all } j = 1, \ldots, m,$$

then set $x_{n+1} := x_n + s(x_n)$, *and go to Step 6. Otherwise, go to Step 5.*
Step 5. *(Line search) Choose a stepsize* $\alpha_n \in (0, +\infty)$ *satisfying the Armijo rule, or the Goldstein rule, or the Wolf rule. Set* $x_{n+1} := x_n + \alpha_n s(x_n)$.
Step 6. *Set* $n := n + 1$. *Go back to Step 2.*

Obviously, a sequence generated by Algorithm 1.1 can be regarded as the one generated by Algorithm 3.2 with Step 5 using just the Armijo rule.

The notion of the L-average Lipschitz condition was introduced by Wang in [36] (but using the terminology "the center Lipschitz condition in the inscribed sphere with L-average") and has been widely used to analyze the convergence properties of the Newton method; see [28, 30] and references therein. We extend in the following definition the notion of the L-average Lipschitz condition to the setting of vector valued functions. Recall that $F := (F_1, \ldots, F_m)^T \in C^2(U, \mathbb{R}^m)$, and that $L : [0, R) \to \mathbb{R}_+$ is nondecreasing and integrable.

Definition 3.2. *Let* $x_0 \in U$ *and* $r \in (0, R)$ *be such that* $\mathbf{B}(x_0, r) \subseteq U$. $\mathrm{D}^2 F$ *is said to satisfy the* L*-average Lipschitz condition on* $\mathbf{B}(x_0, r)$ *if, for each* $i = 1, \ldots, m$ *and any* $x, y \in \mathbf{B}(x_0, r)$ *with* $\|x - x_0\| + \|y - x\| < r$, *the following inequality holds:*

$$\|\nabla^2 F_i(y) - \nabla^2 F_i(x)\| \leq \int_{\|x-x_0\|}^{\|x-x_0\|+\|y-x\|} L(u)\mathrm{d}u.$$

By definition, we can check that on $\mathbf{B}(x_0, r)$ with $r \in (0, R)$, the L-average Lipschitz condition implies the classical Lipschitz condition with Lipschitz constant being $L(r)$. The introduction of the L-average Lipschitz condition is beneficial to provide the more precise convergence criterion and estimation of convergence radius for the Newton method.

Fixing the triple $(x;a,r)$ with $x\in U$ and $(a,r)\in\mathbb{R}^2_+$, we consider the following assumption for $F\in C^2(U,\mathbb{R}^m)$ associated to the triple $(x;a,r)$ and L:

- $L:[0,R)\to\mathbb{R}_+$ is nondecreasing and integrable;
- a satifies (2.11), and $\mathrm{D}^2F(x)$ is positive definite with each $\|\nabla^2F_i(x)^{-1}\|\le a$; (3.2)
- $\mathrm{D}^2F(\cdot)$ satisfies the L-average Lipschitz condition on $\mathbf{B}(x,r)\subseteq U$.

LEMMA 3.3. *Suppose that F satisfies assumption* (3.2) *associated to* $(x_0;a,r)$ *and L, and that $r\le r_a$. Let $x\in\mathbf{B}(x_0,r)$, $\lambda\in\boldsymbol{\Delta}_m$ and F_λ be defined by* (2.7). *Then $\nabla^2F_\lambda(x)$ is positive definite, and*

$$\|\nabla^2F_\lambda(x)^{-1}\|\le\frac{\|\nabla^2F_\lambda(x_0)^{-1}\|}{1-a\int_0^{\|x_0-x\|}L(u)\mathrm{d}u}\le\frac{a}{1-a\int_0^{\|x_0-x\|}L(u)\mathrm{d}u}.$$

Proof. By assumption, one has that

$$\|\nabla^2F_\lambda(x_0)^{-1}\|\|\nabla^2F_\lambda(x)-\nabla^2F_\lambda(x_0)\|\le a\int_0^{\|x_0-x\|}L(u)\mathrm{d}u<a\int_0^{r_a}L(u)\mathrm{d}u=1$$

(by (2.12)). Hence, Lemma 2.1 is applicable and the conclusions hold. □

3.1. Convergence criterion. One of the main results of this subsection is presented in the following theorem, in which we provide a quadratic convergence criterion of the extended Newton method for multiobjective optimization under the assumption that the Hessians of objective functions satisfy the L-average Lipschitz condition. Theorem 3.4 not only extends [19, Theorem 6.1] under a weaker condition, but also improves it in the sense that the quantitative convergence result is provided here (see (3.6) below).

THEOREM 3.4. *Suppose that F satisfies assumption* (3.2) *associated to* $(x_0;a,r_a^*)$ *and L, and*

$$\|s(x_0)\|\le\beta\le b_a. \tag{3.3}$$

Then, the sequence $\{x_n\}$ generated by Algorithm 3.1 with initial point x_0 is well-defined, stays in $\mathbf{B}(x_0,r_a^*)$, and converges to a local Pareto optimum $\bar{x}\in\mathbf{B}[x_0,r_a^*]$. Moreover, the following error estimations hold for each $n\ge0$:

$$\|x_{n+1}-x_n\|=\|s(x_n)\|\le t_{a,n+1}-t_{a,n}, \tag{3.4}$$

and

$$\|x_n-\bar{x}\|\le r_a^*-t_{a,n}. \tag{3.5}$$

Moreover, if $\beta<b_a$, then there exists $N\in\mathbb{N}$ such that

$$\|x_{n+1}-\bar{x}\|\le\frac{r_a^*-t_{a,n+1}}{(2t_{a,n+1}-t_{a,n}-r_a^*)^2}\|x_n-\bar{x}\|^2\quad \textit{for each } n\ge N, \tag{3.6}$$

and so $\{x_n\}$ converges quadratically to $\bar{x}$.

Proof. Since $r_a^* \le r_a$ (cf. Proposition 2.6(i)), Lemma 3.3 is applicable to concluding that

$$\nabla^2 F_\lambda(x) \text{ is positive definite for any } x \in \mathbf{B}(x_0, r_a^*) \text{ and } \lambda \in \mathbf{\Delta}_m. \tag{3.7}$$

Furthermore, by assumption (3.2), it is easy to see that there exists a constant $c > 0$ such that

$$\sup_{\lambda \in \mathbf{\Delta}_m, x \in \mathbf{B}(x_0, r_a^*)} \rho_{\max}(\lambda, x) \le c, \tag{3.8}$$

where $\rho_{\max}(\lambda, x)$ is given by (2.8). We first show that $\{x_n\}$ is well-defined and (3.4). For simplicity, we, as before, omit the first subscript a in the sequence $\{t_{a,k}\}$, write $\{t_k\}$ for $\{t_{a,k}\}$. Thus, in view of Algorithm 3.1, (3.7) and (3.3), one has that x_1 is well-defined and $\|x_1 - x_0\| = \|s(x_0)\| \le \beta = t_1 - t_0$ (due to (2.15)), namely (3.4) holds for $n = 0$. Fix $k \in \mathbb{N}$. Below, we show the following implication:

$$\begin{array}{l} [x_n \text{ is well-defined for all } n = 0, 1, \ldots, k+1 \text{ and (3.4) holds for all } n = 0, \ldots, k] \\ \Rightarrow\ x_{k+2} \text{ is well-defined and } \|s(x_{k+1})\| \le (t_{k+2} - t_{k+1}) \left(\frac{\|s(x_k)\|}{t_{k+1} - t_k} \right)^2. \end{array} \tag{3.9}$$

Granting this, $\{x_n\}$ is well-defined and (3.4) is shown by mathematical induction. To proceed, suppose that x_n is well-defined for all $n = 0, 1, \ldots, k+1$ and (3.4) holds for all $n = 0, \ldots, k$. Recall from (2.6) that there exists $\lambda = (\lambda_1, \ldots, \lambda_m)^T \in \mathbf{\Delta}_m$ such that

$$s(x_k) = -\left[\sum_{j=1}^m \lambda_j \nabla^2 F_j(x_k) \right]^{-1} \sum_{j=1}^m \lambda_j \nabla F_j(x_k) = -\nabla^2 F_\lambda(x_k)^{-1} \nabla F_\lambda(x_k). \tag{3.10}$$

Note by the induction assumption that

$$\|x_{k+1} - x_0\| \le \sum_{i=0}^k \|x_{i+1} - x_i\| \le \sum_{i=0}^k (t_{i+1} - t_i) = t_{k+1} < r_a^* \tag{3.11}$$

(by Proposition 2.6(ii)). Consequently, $x_{k+1} \in \mathbf{B}(x_0, r_a^*)$. Thus, in view of Algorithm 3.1, (3.7) and (3.3), one has that x_{k+2} is well-defined. Furthermore, Lemma 3.3 is applicable to concluding that

$$\|\nabla^2 F_\lambda(x_{k+1})^{-1}\| \le \frac{a}{1 - a \int_0^{\|x_{k+1} - x_0\|} L(u) \mathrm{d}u} \le -a\, {h_a}'(t_{k+1})^{-1}, \tag{3.12}$$

because, by (2.14),

$$-{h_a}'(t_{k+1})^{-1} = \frac{1}{1 - a \int_0^{t_{k+1}} L(u) \mathrm{d}u}.$$

Observe further from (3.10) that

$$\nabla^2 F_\lambda(x_k) s(x_k) + \nabla F_\lambda(x_k) = 0.$$

Thus, by L-average Lipschitz condition assumption, we obtain

$$\begin{aligned}\|\nabla F_\lambda(x_{k+1})\| &= \|\nabla F_\lambda(x_k+s(x_k))-(\nabla^2F_\lambda(x_k)s(x_k)+\nabla F_\lambda(x_k))\|\\ &\le \int_0^1\|\nabla^2F_\lambda(x_k+ts(x_k))-\nabla^2F_\lambda(x_k)\|\|s(x_k)\|\mathrm{d}t\\ &\le \int_0^1\int_{\|x_k-x_0\|}^{\|x_k-x_0\|+t\|s(x_k)\|}L(u)\mathrm{d}u\|s(x_k)\|\mathrm{d}t\\ &= \int_0^{\|s(x_k)\|}L(\|x_k-x_0\|+u)(\|s(x_k)\|-u)\mathrm{d}u.\end{aligned} \tag{3.13}$$

Since by inductive assumption that $\|s(x_k)\|\le t_{k+1}-t_k$, it follows from Lemma 2.7 and (3.11) (with k in place of $k+1$) that

$$\int_0^{\|s(x_k)\|}L(\|x_k-x_0\|+u)(\|s(x_k)\|-u)\mathrm{d}u\le\frac{\|s(x_k)\|^2}{(t_{k+1}-t_k)^2}\int_0^{t_{k+1}-t_k}L(t_k+u)(t_{k+1}-t_k-u)\mathrm{d}u.$$

Note by (2.13)-(2.15) that

$$a\int_0^{t_{k+1}-t_k}L(t_k+u)(t_{k+1}-t_k-u)\mathrm{d}u=h_a(t_{k+1})-h_a(t_k)-h_a{}'(t_k)(t_{k+1}-t_k)=h_a(t_{k+1}). \tag{3.14}$$

Hence, we have from (3.13)-(3.14) that

$$a\|\nabla F_\lambda(x_{k+1})\|\le\frac{\|s(x_k)\|^2}{(t_{k+1}-t_k)^2}h_a(t_{k+1}).$$

Note by (2.10) that $\|s(x_{k+1})\|\le\|\nabla^2F_\lambda(x_{k+1})^{-1}\|\|\nabla F_\lambda(x_{k+1})\|$. It follows from (3.12) that

$$\|s(x_{k+1})\|\le -h_a{}'(t_{k+1})^{-1}h_a(t_{k+1})\frac{\|s(x_k)\|^2}{(t_{k+1}-t_k)^2}=(t_{k+2}-t_{k+1})\left(\frac{\|s(x_k)\|}{t_{k+1}-t_k}\right)^2.$$

Thus, implication (3.9) is proved.

Now, we show the convergence of $\{x_n\}$ to a local Pareto optimum. Since $\{t_n\}$ is monotonically increasing and converges to r_a^* (by Proposition 2.6(ii)), (3.4) shows that $\{x_n\}$ is a Cauchy sequence, and so, there exists $\bar{x}\in\mathbf{B}[x_0,r_a^*]$ such that $\lim_{n\to\infty}x_n=\bar{x}$. Furthermore, (3.4) says that $\lim_{n\to\infty}\|s(x_n)\|=0$. Observe further from (2.9) and (3.8) that $|\theta(x_n)|\le\frac{c}{2}\|s(x_n)\|^2$ for each $n\in\mathbb{N}$, and then, passing to the limits, we get that $\lim_{n\to\infty}|\theta(x_n)|=0$. Note by Lemma 2.4(iii) that θ is continuous and so $\theta(\bar{x})=0$. Then, by Lemma 2.4(ii), one has that $\bar{x}$ is a critical point, and thus, it is a local Pareto optimum (by (2.2)). Fix $n\in\mathbb{N}$. One has by (3.4) that $\|x_{n+l}-x_n\|\le t_{n+l}-t_n$ for each $l\in\mathbb{N}$, and so (3.5) is seen to hold by passing to the limits (as $l\to\infty$).

Finally, we prove the quadratic convergence rate of $\{x_n\}$ to $\bar{x}$. Fix $n\in\mathbb{N}$, and note from (3.4) and implication (3.9) that

$$\|s(x_{n+j})\|\le(t_{n+j+1}-t_{n+j})\left(\frac{\|s(x_n)\|}{t_{n+1}-t_n}\right)^2\quad\text{for each } j\in\mathbb{N}. \tag{3.15}$$

In view of Algorithm 3.1, one sees that $\|x_i-x_{n+1}\|\le\sum_{j=n+1}^{i-1}\|s(x_j)\|$ for each $i>n+1$. Letting $i\to\infty$, one has by the convergence of $\{x_n\}$ to $\bar{x}$ and by (3.15) that

$$\|\bar{x}-x_{n+1}\|\le\sum_{j=n+1}^{\infty}\|s(x_j)\|\le(r_a^*-t_{n+1})\left(\frac{\|s(x_n)\|}{t_{n+1}-t_n}\right)^2\le\frac{r_a^*-t_{n+1}}{t_{n+1}-t_n}\|s(x_n)\| \tag{3.16}$$

(by (3.4)). Then, it follows that

$$\|\bar{x}-x_n\| \geq \|x_{n+1}-x_n\| - \|\bar{x}-x_{n+1}\| \geq \frac{2t_{n+1}-t_n-r_a^*}{t_{n+1}-t_n}\|s(x_n)\|. \tag{3.17}$$

By assumption that $\beta < b_a$, Proposition 2.6(iii) is applicable, and then we have by the equality of (2.17) that there exists $N \in \mathbb{N}$ such that

$$\frac{2t_{n+1}-t_n-r_a^*}{t_{n+1}-t_n} > 0 \quad \text{for each } n \geq N.$$

Therefore, combining (3.16) and (3.17), we obtain (3.6). This, together with the inequality of (2.17), ensures the quadratic convergence rate of $\{x_n\}$ to $\bar{x}$. The proof is complete. □

Theorem 3.5 below shows that under almost the same conditions as in Theorem 3.4, a sequence $\{x_n\}$ generated by Algorithm 1.1 or 3.2 with initial point x_0 is the one generated by Algorithm 3.1 with the same initial point x_0. Hence, all the conclusions of Theorem 3.4 hold for Algorithm 1.1 or 3.2.

THEOREM 3.5. *Suppose that F satisfies assumption* (3.2) *associated to* $(x_0; a, r_a)$ *and* L, *and*

$$\|s(x_0)\| \leq \beta \leq \frac{-3(1-\sigma)h_a'(r_a^*)}{aL(r_a^*)}. \tag{3.18}$$

Then, with initial point x_0, any sequence $\{x_n\}$ generated by Algorithm 3.2 coincides with the one generated by Algorithm 3.1; consequently, the conclusions of Theorem 3.4 hold.

Proof. Below, we only show the case when $\{x_n\}$ is a sequence generated by Algorithm 1.1 with initial point x_0 because the proof is similar for Algorithm 3.2. To furniture the proof of this theorem, fix $i \in \mathbb{N}$. First, we show the following implication:

$$[\|s(x_i)\| \leq t_{i+1}-t_i,\ \|x_i-x_0\| + \|s(x_i)\| \leq r_a^*] \Rightarrow [x_{i+1} = x_i + s(x_i)]. \tag{3.19}$$

For this purpose, we assume that

$$\|s(x_i)\| \leq t_{i+1}-t_i \quad \text{and} \quad \|x_i-x_0\| + \|s(x_i)\| \leq r_a^*. \tag{3.20}$$

Noting by Proposition 2.6 that $r_a^* \leq r_a$, we have $x_i \in \mathbf{B}(x_0, r_a)$, and then obtain from Lemma 3.3 that for each $\lambda \in \mathbf{\Delta}_m$, $\nabla^2 F_\lambda(x_i)$ is positive definite and

$$\|\nabla^2 F_\lambda(x_i)^{-1}\| \leq -ah_a'(\|x_i-x_0\|)^{-1}. \tag{3.21}$$

By assumption (3.20), one has $x_i + s(x_i) \in \mathbf{B}(x_0, r_a)$. Fix $j \in \{1, \ldots, m\}$. By the Taylor formula, one has that

$$\begin{aligned}
&F_j(x_i+s(x_i)) \\
&= F_j(x_i) + \nabla F_j(x_i)^T s(x_i) + \tfrac{1}{2}s(x_i)^T\nabla^2 F_j(x_i)s(x_i) \\
&\quad + \textstyle\int_0^1 s(x_i)^T(\nabla^2 F_j(x_i+ts(x_i)) - \nabla^2 F_j(x_i))s(x_i)(1-t)\mathrm{d}t \\
&\leq F_j(x_i) + \nabla F_j(x_i)^T s(x_i) + \tfrac{1}{2}s(x_i)^T\nabla^2 F_j(x_i)s(x_i) + \tfrac{L(r_a^*)}{6}\|s(x_i)\|^3,
\end{aligned}$$

where the inequality holds because

$$\|\nabla^2 F_j(x_i + ts(x_i)) - \nabla^2 F_j(x_i)\| \leq \int_{\|x_i - x_0\|}^{\|x_i - x_0\| + t\|s(x_i)\|} L(u)\mathrm{d}u \leq L(r_a^*)\|s(x_i)\|t$$

(due to assumption (3.2) and the fact that $L(\cdot)$ is nondecreasing and positive). By the definition of θ (cf. (2.5)), this implies that

$$\begin{aligned} F_j(x_i + s(x_i)) &\leq F_j(x_i) + \theta(x_i) + \tfrac{L(r_a^*)}{6}\|s(x_i)\|^3 \\ &= F_j(x_i) + \sigma\theta(x_i) + (1-\sigma)\theta(x_i) + \tfrac{L(r_a^*)}{6}\|s(x_i)\|^3, \end{aligned} \tag{3.22}$$

where $\sigma \in (0,1)$ is the parameter in Algorithm 1.1. Recall from (2.9) and Lemma 2.4(i) that

$$\theta(x_i) \leq -\frac{\rho_{\min}(\lambda, x_i)}{2}\|s(x_i)\|^2. \tag{3.23}$$

Recalling by (2.8) that $\rho_{\min}(\lambda, x_i) = \|\nabla^2 F_\lambda(x_i)^{-1}\|^{-1}$, it follows from (3.21) and (3.23) that

$$\theta(x_i) \leq \frac{1}{2a}h_a'(\|x_i - x_0\|)\|s(x_i)\|^2 \leq \frac{1}{2a}h_a'(r_a^*)\|s(x_i)\|^2, \tag{3.24}$$

where the last inequality holds because that $h'(\cdot)$ is monotonically increasing on $[0, r_a^*]$. Note that $\{t_{i+1} - t_i\}$ is monotonically decreasing (cf [28, Lemma 2.4]), and so, for each $i \in \mathbb{N}$, $t_{i+1} - t_i \leq t_1 - t_0 = \beta$ (by (2.15)). This, together with (3.20), implies that

$$\|s(x_i)\| \leq t_{i+1} - t_i \leq t_1 - t_0 = \beta \leq \frac{-3(1-\sigma)h_a'(r_a^*)}{aL(r_a^*)}$$

(due to (3.18)). Combining this with (3.24) yields that

$$(1-\sigma)\theta(x_i) + \frac{L(r_a^*)\|s(x_i)\|}{6}\|s(x_i)\|^2 \leq \left(\frac{L(r_a^*)\|s(x_i)\|}{3} + \frac{(1-\sigma)h_a'(r_a^*)}{a}\right)\frac{\|s(x_i)\|^2}{2} \leq 0;$$

then, (3.22) implies that

$$F_j(x_i + s(x_i)) \leq F_j(x_i) + \sigma\theta(x_i) \quad \text{for all } j = 1, \ldots, m.$$

Thus, in view of Algorithm 1.1, we have $x_{i+1} = x_i + s(x_i)$ and so (3.19) is seen to hold.

Below, we show by induction that $\{x_n\}$ coincides with the sequence generated by Algorithm 3.1 with the same initial point x_0, namely the following assertion holds for each $n \in \{0\} \cup \mathbb{N}$:

$$x_{n+1} = x_n + s(x_n) \tag{3.25}$$

Since $\|s(x_0)\| \leq \beta = t_1 - t_0 \leq r_a^*$ by (3.18) and Proposition 2.6(i), it follows from (3.19) that (3.25) holds for $n = 0$. Suppose that $x_1, \ldots, x_k$ are the same points as generated by Algorithm 3.1. Then, by Theorem 3.4, we have that $x_i \in \mathbf{B}(x_0, r_a^*)$ and $\|s(x_i)\| \leq t_{i+1} - t_i$ for $i = 1, \ldots, k$, and

$$\|x_k - x_0\| + \|s(x_k)\| \leq \|x_k - x_{k-1}\| + \cdots + \|x_1 - x_0\| + \|s(x_k)\| \leq t_{k+1} < r_a^*.$$

This implies that the assumptions of implication (3.19) hold when $i = k$. Then, it follows from implication (3.19) that $\alpha_k = 1$, and so, (3.25) holds for $n = k$. Thus, x_{k+1} is the same point as generated by Algorithm 3.1. Then, we obtain inductively that $\{x_n\}$ is same as the sequence generated by Algorithm 3.1 with same initial point x_0. Therefore, the conclusions of Theorem 3.4 hold and the proof is complete. □

3.2. Estimation of convergence radius. This subsection is devoted to providing an estimate of the radius of the convergence ball of the extended Newton method (without or with line-search scheme) for multiobjective optimization under an L-average Lipschitz condition. For this purpose, let $a^* > 0$ be such that (2.11) is reduced to

$$\frac{1}{R}\int_0^R L(u)(R-u)\mathrm{d}u > \frac{1}{a^*}. \tag{3.26}$$

Let (r_{a^*}, b_{a^*}) be the pair of positive constants given by (2.12) with a^* in place of a. Let $x^* \in U$ be a local Pareto optimum of F, and assume that F satisfies assumption (3.2) associated to $(x^*; a^*, r_{a^*})$ and L. Throughout this subsection, we always assume that $L(\cdot)$ is left-hand continuous. Write

$$\xi_{a^*} := a^* \max\{\|\nabla^2 F_i(x^*)\| : i = 1, \ldots, m\}. \tag{3.27}$$

Let $t \in (0, r_{a^*})$, and set

$$R_t := R - t, \quad a_t := \frac{a^*}{1 - a^* \int_0^t L(u)\mathrm{d}u}$$

and

$$\beta_t := \frac{a^* \int_0^t L(u)(t-u)\mathrm{d}u + \xi_{a^*} t}{1 - a^* \int_0^t L(u)\mathrm{d}u}.$$

Define the function $L_t : [0, R_t) \to \mathbb{R}$ by

$$L_t(u) := L(u+t) \quad \text{for each } u \in [0, R_t). \tag{3.28}$$

Let $\bar{r}_{a_t}, \bar{b}_{a_t}, \bar{r}^*_{a_t}$ denote the corresponding positive constants given by (2.12) and (2.16) with β_t, a_t, L_t in place of β, a, L.

LEMMA 3.6. *There holds that $\bar{r}_{a_t} + t = r_{a^*}$ and $\beta_t < \bar{b}_{a_t}$ for each $0 < t < \frac{b_{a^*}}{1+\xi_{a^*}}$, and there exists $0 < r < \frac{b_{a^*}}{1+\xi_{a^*}}$ such that*

$$\beta_t \le \frac{-3(1-\sigma)h'_{a_t}(\bar{r}^*_{a_t})}{a_t L_t(\bar{r}^*_{a_t})} \quad \textit{for each } t \in (0, r). \tag{3.29}$$

Proof. Let $0 < t < \frac{b_{a^*}}{1+\xi_{a^*}}$. By definition, it is easy to verify that $0 < \bar{r}_{a_t} \le r_{a^*}$ because $a_t \ge a^*$ and $L_t(u) \ge L(u)$ for each $u \in \mathbb{R}^+$. Observing further from the definition of r_{a^*} and $\bar{r}_{a_t}$, we have that

$$a^* \int_0^{\bar{r}_{a_t}} L_t(u)\mathrm{d}u = 1 - a^* \int_0^t L(u)\mathrm{d}u = a^* \int_0^{r_{a^*}} L(u)\mathrm{d}u - a^* \int_0^t L(u)\mathrm{d}u = a^* \int_t^{r_{a^*}} L(u)\mathrm{d}u.$$

This gives that

$$\int_t^{\bar{r}_{a_t}+t} L(u)\mathrm{d}u = \int_t^{r_{a^*}} L(u)\mathrm{d}u. \tag{3.30}$$

As $L(\cdot)$ is a nondecreasing and positive integrable function, one has from (3.30) that $\bar{r}_{a_t}+t = r_{a^*}$. This, together with the definitions of $\bar{b}_{a_t}, a_t, L_t$, implies that

$$\bar{b}_{a_t} = a_t \int_0^{\bar{r}_{a_t}} L_t(u)u\mathrm{d}u = a_t \int_0^{r_{a^*}-t} L_t(u)u\mathrm{d}u = \frac{a^* \int_t^{r_{a^*}} L(u)(u-t)\mathrm{d}u}{1 - a^* \int_0^t L(u)\mathrm{d}u}.$$

Then, by elementary calculus and noting that $a^* \int_0^{r_{a^*}} L(u)\mathrm{d}u = 1$, $b_{a^*} = a^* \int_0^{r_{a^*}} L(u)u\mathrm{d}u$ and $0 < t < \frac{b_{a^*}}{1+\xi_{a^*}}$, one checks that

$$a^* \int_0^t L(u)(t-u)\mathrm{d}u + \xi_{a^*} t < a^* \int_t^{r_{a^*}} L(u)(u-t)\mathrm{d}u,$$

or equivalently, $\beta_t < \bar{b}_{a_t}$. Thus the first assertion is shown.

To verify the second assertion, we note first by the first assertion that $\bar{r}^*_{a_t}$ is well-defined for each $t \in (0, \frac{b_{a^*}}{1+\xi_{a^*}})$. Furthermore, by definition, one can check that $\lim_{t\to 0^+} \beta_t = 0$, $\lim_{t\to 0^+} a_t = a^*$, $\lim_{t\to 0^+} \bar{r}_{a_t} = r_{a^*}$ and $\lim_{t\to 0^+} \bar{b}_{a_t} = b_{a^*}$. Hence, $\lim_{t\to 0^+} \bar{r}^*_{a_t} = 0$ thanks to (2.16) (applied to β_t, a_t, L_t in place of β, a, L) and so $\lim_{t\to 0^+} h'_{a_t}(\bar{r}^*_{a_t}) = -1$. Thus it follows from the left-hand continuity assumption for L that

$$\lim_{t\to 0^+} \frac{-3(1-\sigma)h'_{a_t}(\bar{r}^*_{a_t})}{a_t L_t(\bar{r}^*_{a_t})} \geq \frac{3(1-\sigma)}{a^* L(r^*_a)} > 0.$$

Since $\lim_{t\to 0^+} \beta_t = 0$ and the function $t \mapsto \beta_t$ is monotonically increasing on $[0, r_{a^*})$, it follows that there exists $0 < r \leq \frac{b_{a^*}}{1+\xi_{a^*}}$ to satisfy (3.29), and the proof is complete. □

Another useful proposition is as follows.

Proposition 3.7. *Suppose that F satisfies assumption* (3.2) *associated to* $(x^*; a^*, r_{a^*})$ *and* L. *Let* $x_0 \in \mathbf{B}\left(x^*, \frac{b_{a^*}}{1+\xi_{a^*}}\right)$ *and* $t := \|x_0 - x^*\|$. *Then, the following assertions hold:*

(i) *F satisfies assumption* (3.2) *associated to* $(x_0; a_t, \bar{r}_{a_t})$ *and* L_t.

(ii) $\|s(x_0)\| \leq \beta_t < \bar{b}_{a_t}$.

Proof. (i) We first show (2.11) holds with a_t, R_t, L_t in place of a, R, L, or equivalently,

$$\int_t^R L(u)(R-u)\mathrm{d}u \geq (R-t) \int_t^{r_{a^*}} L(u)\mathrm{d}u, \tag{3.31}$$

thanks to the first equality in (2.12) (applied to a^* in place of a) and the definitions of a_t, R_t, L_t. By (3.26) and the first equality in (2.12) (applied to a^* in place of a), one checks that

$$\int_t^R L(u)(R-u)\mathrm{d}u \geq \tfrac{R}{a^*} - \int_0^t L(u)(R-u)\mathrm{d}u = R\int_t^{r_{a^*}} L(u)\mathrm{d}u + \int_0^t L(u)u\mathrm{d}u.$$

This implies trivially (3.31), showing (2.11) (with a_t, R_t, L_t in place of a, R, L), namely the first assumption in (3.2) (associated to $(x_0; a_t, \bar{r}_{a_t})$ and L_t). To show the second assumption in (3.2), note first that $\|x_0 - x^*\| < \frac{b_{a^*}}{1+\xi_{a^*}} < b_{a^*} < r_{a^*}$, Lemma 3.3 is applicable to concluding that, for each $j = 1, \ldots, m$, $\nabla^2 F_j(x_0)$ is positive definite, and

$$\|\nabla^2 F_j(x_0)^{-1}\| \leq \frac{\|\nabla^2 F_j(x^*)^{-1}\|}{1 - a^* \int_0^{\|x_0 - x^*\|} L(u)\mathrm{d}u} \leq \frac{a^*}{1 - a^* \int_0^t L(u)\mathrm{d}u} = a_t;$$

consequently, the second assumption in (3.2) (associated to $(x_0; a_t, \bar{r}_{a_t})$ and L_t) is checked. Now let us verify the last assumption. To do this, let $x, y \in \mathbf{B}(x_0, \bar{r}_{a_t})$ be such that $\|x - x_0\| + \|y - x\| < \bar{r}_{a_t}$, and fix j. Then, as $t = \|x_0 - x^*\| \in \left(0, \frac{b_{a^*}}{1+\xi_{a^*}}\right)$, the first assertion of Lemma 3.6 is applicable to concluding that

$$\|x - x^*\| + \|y - x\| \le \|x_0 - x^*\| + \|x - x_0\| + \|y - x\| \le t + \bar{r}_{a_t} = r_{a^*}.$$

Thus it follows from the last assumption in (3.2) (associated to $(x^*; a^*, r_{a^*})$ and L) that

$$\|\nabla^2 F_j(y) - \nabla^2 F_j(x)\| \le \int_{t+\|x_0-x\|}^{t+\|x_0-x\|+\|x-y\|} L(u)\mathrm{d}u = \int_{\|x_0-x\|}^{\|x_0-x\|+\|x-y\|} L_t(u)\mathrm{d}u.$$

This shows the third assumption in (3.2) (associated to $(x_0; a_t, \bar{r}_{a_t})$ and L_t) and the proof for assertion (i) is complete.

(ii) By Lemma 3.6, we only need to show $\|s(x_0)\| \le \beta_t$. Noting that x^* is a local Pareto optimum of F, we obtain from (2.2) that x^* is a critical point of F. Therefore, it follows from Lemma 2.4 that $s(x^*) = 0$. Note by definition that there exists $\lambda(:= \lambda(x^*)) \in \mathbf{\Delta}_m$ (the KKT multipliers of problem (2.3)) such that

$$s(x^*) = -\left[\sum_{j=1}^m \lambda_j(x^*)\nabla^2 F_j(x^*)\right]^{-1} \sum_{j=1}^m \lambda_j(x^*)\nabla F_j(x^*) = -\nabla^2 F_\lambda(x^*)^{-1}\nabla F_\lambda(x^*).$$

Hence $\nabla F_\lambda(x^*) = 0$, and

$$\begin{aligned}\|\nabla F_\lambda(x_0) - \nabla^2 F_\lambda(x^*)(x_0 - x^*)\| &= \|\textstyle\int_0^1 (\nabla^2 F_\lambda(x^* + \tau(x_0 - x^*)) - \nabla^2 F_\lambda(x^*))(x_0 - x^*)\mathrm{d}\tau\| \\ &\le \int_0^1 \int_0^{t\tau} L(u)t\mathrm{d}u\mathrm{d}\tau = \int_0^t L(u)(t-u)\mathrm{d}u,\end{aligned}$$

thanks to the third assumption in (3.2) (associated to $(x^*; a^*, r_{a^*})$). Therefore,

$$\begin{aligned}a^*\|\nabla F_\lambda(x_0)\| &\le a^*\|\nabla F_\lambda(x_0) - \nabla^2 F_\lambda(x^*)(x_0 - x^*)\| + a^*\|\nabla^2 F_\lambda(x^*)\|t \\ &\le a^* \int_0^t L(u)(t-u)\mathrm{d}u + \xi_{a^*}t.\end{aligned} \tag{3.32}$$

Furthermore, by Lemma 3.3 one has that

$$\|\nabla^2 F_\lambda(x_0)^{-1}\| \le \frac{\|\nabla^2 F_\lambda(x^*)^{-1}\|}{1 - a^* \int_0^{\|x_0-x^*\|} L(u)\mathrm{d}u} \le \frac{a^*}{1 - a^* \int_0^t L(u)\mathrm{d}u}.$$

This, together with (3.32) and (2.10), implies that that

$$\|s(x_0)\| \le \|\nabla^2 F_\lambda(x_0)^{-1}\|\|\nabla F_\lambda(x_0)\| \le \frac{a^* \int_0^t L(u)(t-u)\mathrm{d}u + \xi_{a^*}t}{1 - a^* \int_0^t L(u)\mathrm{d}u} = \beta_t.$$

The proof is complete. □

Theorem 3.8. *Suppose that F satisfies assumption* (3.2) *associated to* $(x^*; a^*, r_{a^*})$ *and L. Let $x_0 \in \mathbf{B}(x^*, \frac{b_{a^*}}{1+\xi_{a^*}})$. Then, the sequence $\{x_n\}$ generated by Algorithm 3.1 with initial point x_0 is well-defined and converges quadratically to a local Pareto optimum of F.*

Proof. Let $x_0 \in \mathbf{B}(x^*, \frac{b_{a^*}}{1+\xi_{a^*}})$ and $t = \|x_0 - x^*\|$. By Proposition 3.7, F satisfies assumption (3.2) associated to $(x_0; a_t, \bar{r}_{a_t})$ and L_t, and (3.3) holds (with $\beta_t, \bar{b}_{a_t}$ in place of β, b_a). Hence, Theorem 3.4 is applicable (with $\beta_t, a_t, L_t, \bar{b}_{a_t}$ in place of β, a, L, b_a) and the conclusion follows. This completes the proof. ∎

Theorem 3.9 below shows that if F satisfies assumption (3.2) associated to $(x^*; a^*, r_{a^*})$ and L, then there exists $r > 0$ such that any sequence $\{x_n\}$ generated by Algorithm 1.1 or 3.2 with initial point $x_0 \in \mathbf{B}(x^*, r)$ converges quadratically to a local Pareto optimum of F. In the next section, we provide an explicitly estimate for the radius r because there $L(\cdot) \equiv L$.

THEOREM 3.9. *Suppose that F satisfies assumption* (3.2) *associated to* $(x^*; a^*, r_{a^*})$ *and L. Let $r \in \left(0, \frac{b_{a^*}}{1+\xi_{a^*}}\right)$ satisfy* (3.29)*. Then, for any $x_0 \in \mathbf{B}(x^*, r)$, any sequence $\{x_n\}$ generated by Algorithms 3.2 with initial point x_0 converges quadratically to a local Pareto optimum of F.*

Proof. Let $x_0 \in \mathbf{B}(x^*, r)$ and $t = \|x_0 - x^*\|$. Then $x_0 \in \mathbf{B}\left(x^*, \frac{b_{a^*}}{1+\xi_{a^*}}\right)$, and Proposition 3.7 is applicable to concluding that assertions (i) and (ii) there hold, namely F satisfies assumption (3.2) associated to $(x_0; a_t, \bar{r}_{a_t})$ and L_t, and $\|s(x_0)\| \le \beta_t$. Furthermore, by (3.29), (3.18) holds with $\beta_t, a_t, L_t, \bar{r}^*_{a_t}$ in place of β, a, L, r^*_a. Thus, the conclusion follows from Theorem 3.5 (applied to $\beta_t, a_t, L_t, \bar{r}^*_{a_t}$ in place of β, a, L, r^*_a), and the proof is complete. ∎

4. Applications.

By virtue of the results established in the preceding section, this section is devoted to establishing convergence analysis theorems under the classical Lipschitz condition or the γ-condition to multiobjective optimization. In particular, a global version of the extended Newton method is proposed and its global convergence is established.

4.1. Theorems under the classical Lipschitz condition and global version of the extended Newton method with its convergence.

4.1.1. Theorems under the classical Lipschitz condition.

Kantorovich's theorem [24] is one of the famous results on the Newton method, which provides a criterion for ensuring its quadratic convergence under the classical Lipschitz condition. The main point of Kantorovich's type premise is to let L mentioned in the preceding section be a constant function. In this case, the L-average Lipschitz condition of $\nabla^2 F_j$ is reduced to the classical Lipschitz condition of $\nabla^2 F_j$ for each $j = 1, \ldots, m$. That is, there are $L > 0$ and $r > 0$ such that

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(y)\| \le L\|x - y\| \quad \text{for each } x, y \in \mathbf{B}(x_0, r).$$

Then the function L_t defined in (3.28) is independent of the choice of t and coincides with L, that is, $L(\cdot) = L_t(\cdot) = L$ on $\mathbb{R}^+$. Thus, for any $a > 0$, one has that

$$b_a = \frac{1}{2aL}, \quad r_a = \frac{1}{aL},$$

and the majorizing functions h_a defined by (2.13) is reduced to

$$h_a(t) = \beta - t + \frac{aL}{2}t^2 \quad \text{for each } t \in \mathbb{R}.$$

Therefore, if $\beta \le \frac{1}{2aL}$, one has by (2.12), (2.15) and (2.16) (see also [36]) that

$$\bar{r}_a^* = r_a^* = \frac{1-\sqrt{1-2aL\beta}}{aL}, \tag{4.1}$$

$$t_{a,n} = \frac{1-q_a^{2^n-1}}{1-q_a^{2^n}} r_a^* \quad \text{and} \quad t_{a,n+1} - t_{a,n} = \frac{1-q_a}{1-q_a^{2^{n+1}}} q_a^{2^n-1} r_a^* \quad \text{for each } n \in \mathbb{N},$$

where

$$q_a := \frac{1-\sqrt{1-2aL\beta}}{1+\sqrt{1-2aL\beta}}, \tag{4.2}$$

and we adopt the convention that $\frac{1-q_a^{2^n-1}}{1-q_a^{2^n}} := 1 - (\frac{1}{2})^n$ and $\frac{1-q_a}{1-q_a^{2^{n+1}}} := \left(\frac{1}{2}\right)^{n+1}$ if $q_a = 1$.

Theorem 4.1 follows directly from Theorems 3.4 and 3.5, and establishes a quantitative convergence criterion of the extended Newton method for multiobjective optimization under the classical Lipschitz condition.

THEOREM 4.1. *Suppose that F satisfies assumption* (3.2) *associated to* $(x_0; a, r_a^*)$ *and* $L(\cdot) \equiv L$, *and* $\|s(x_0)\| \le \beta$. *Let q_a be given by* (4.2). *Then, with initial point x_0, we have the following assertions:*

(i) *If $\beta \le \frac{1}{2aL}$, then the sequence $\{x_n\}$ generated by Algorithm 3.1 is well-defined, stays in $\mathbf{B}(x_0, r_a^*)$, and converges to a local Pareto optimum $\bar{x} \in \mathbf{B}[x_0, r_a^*]$ with the following error estimations:*

$$\|x_{n+1} - x_n\| \le \frac{1-q_a}{1-q_a^{2^{n+1}}} q_a^{2^n-1} r_a^* \quad \text{and} \quad \|x_n - \bar{x}\| \le \frac{1-q_a}{1-q_a^{2^n}} q_a^{2^n-1} r_a^* \quad \text{for each } n \in \mathbb{N}. \tag{4.3}$$

(ii) *If $\beta < \frac{1}{2aL}$, then $\{x_n\}$ in* (i) *converges quadratically to $\bar{x}$ with the following error estimation for some $N \in \mathbb{N}$:*

$$\|x_{n+1} - \bar{x}\| \le \frac{q_a(1-q_a^{2^{n+1}})}{(1-q_a)(1-q_a^{2^n})^2 r_a^*} \|x_n - \bar{x}\|^2 \quad \text{for each } n \ge N. \tag{4.4}$$

(iii) *If $\beta \le \frac{-9(1-\sigma)^2+3(1-\sigma)\sqrt{1+9(1-\sigma)^2}}{aL}$, then $\beta < \frac{1}{2aL}$, and any sequence $\{x_n\}$ generated by Algorithm 3.2 coincides with the one generated by Algorithm 3.1, and satisfies* (4.3) *and* (4.4).

Proof. Assertions (i) and (ii) follow directly from Theorem 3.4. Then, it remains to show assertion (iii). In fact, assume that $\beta \le \frac{-9(1-\sigma)^2+3(1-\sigma)\sqrt{1+9(1-\sigma)^2}}{aL}$. Then $\beta < \frac{1}{2aL}$ because $-9(1-\sigma)^2 + 3(1-\sigma)\sqrt{1+9(1-\sigma)^2} < \frac{1}{2}$. Since $L(\cdot) \equiv L$, it follows from (4.1) that $\frac{-3(1-\sigma)h_a'(r_a^*)}{aL(r_a^*)} = \frac{3(1-\sigma)\sqrt{1-2aL\beta}}{aL}$. Thus, (3.18) holds because it is equivalent that $aL\beta \le 3(1-\sigma)\sqrt{1-2aL\beta}$, which is true by assumption. Hence, the conclusion follows from Theorem 3.5. □

REMARK 4.1. *Under the assumption made in Theorem 4.1, we see that there exist $V \subseteq \mathbf{B}(x_0, r_a^*)$, $\bar{a} := \frac{1}{a}$ and $\bar{b} > 0$ such that $\bar{a}\mathrm{I} \le \nabla^2 F_j(x) \le \bar{b}\mathrm{I}$ for all $x \in V$ and all $j = 1, \ldots, m$,* *where, for A, $B \in \mathbb{R}^{n\times n}$, $A \geq B$ means that $A - B$ is positive semi-definite. Thus [19, Theorem 6.1] could apply. However, Theorem 4.1 cannot be derived via a direct application of [19, Theorem 6.1]. In fact, Example 4.1 below illustrates the case where Theorem 4.1 is applicable but not [19, Theorem 6.1].*

EXAMPLE 4.1. *Let $\sigma \in (\frac{1}{2}, 1)$ and let τ satisfy*

$$(1-\sigma)\sigma < \tau \leq -9(1-\sigma)^2 + 3(1-\sigma)\sqrt{1+9(1-\sigma)^2}. \tag{4.5}$$

Consider problem (1.1) *with $m = l = 1$ and $F : \mathbb{R} \to \mathbb{R}$ defined by*

$$F(x) := -\tau x + \frac{1}{2}x^2 - \frac{1}{6}x^3 \quad \text{for each } x \in \mathbb{R}.$$

Then

$$F''(x) = 1 - x \quad \text{for each } x \in \mathbb{R}. \tag{4.6}$$

Let $x_0 = 0$. Then, one checks that

$$a := \|F''(x_0)^{-1}\| = 1, \quad \|s(x_0)\| = \|-(F''(x_0))^{-1}F'(x_0)\| = \tau, \tag{4.7}$$

and F'' satisfies the Lipschitz condition with modulus $L = 1$ on $[-1, 1]$. By (4.5)*, we see that Theorem 4.1(iii) is applicable to concluding that any sequence $\{x_n\}$ generated by Algorithm 3.2 (and so Algorithm 1.1) with initial point x_0 converges to a local Pareto optimum. We below show that [19, Theorem 6.1] is not applicable. To do this, suppose on the contrary that [19, Theorem 6.1] is applicable. Then, there exist $0 < r < 1$ and positive numbers $a_r, b_r, \delta, \varepsilon$ such that*

$$\frac{\varepsilon}{a_r} \leq 1-\sigma, \quad \|s(x_0)\| \leq \min\left\{\delta, r(1-\frac{\varepsilon}{a_r})\right\}, \quad a_r \leq F''(x) \leq b_r \text{ for all } x \in (-r, r), \tag{4.8}$$

and $\|F''(x) - F''(y)\| \leq \varepsilon$ for all $x, y \in (-r, r)$ with $\|x - y\| \leq \delta$. Then, by (4.6)*, without loss of generality, we take $a_r = 1 - r$ and $\delta = \varepsilon \leq (1-r)(1-\sigma)$. Thus, if $r \geq 1-\sigma$, one has that $\|s(x_0)\| \leq \delta \leq \sigma(1-\sigma)$. Below we shows that this is also true if $r \leq 1-\sigma$. Granting this, one has from* (4.7) *that $\tau \leq \sigma(1-\sigma)$, which is a contradiction to* (4.5)*. To proceed, assume $r \leq 1-\sigma$, and note that the function $t \mapsto \min\{t, r(1-\frac{t}{1-r})\}$ attains its maximum t_0 on $[0, (1-r)(1-\sigma)]$ at t_0 satisfying $t_0 = r(1-\frac{t_0}{1-r})$, i.e., $t_0 = r(1-r)$. Since $\sigma \in (\frac{1}{2}, 1)$ by assumption, it follows that $r \leq 1-\sigma \leq \frac{1}{2}$ and so $\min\{\delta, r(1-\frac{\delta}{1-r})\} \leq t_0 = r(1-r) \leq \sigma(1-\sigma)$. Thus we have by* (4.8) *that $\|s(x_0)\| \leq \min\{\delta, r(1-\frac{\delta}{1-r})\} \leq \sigma(1-\sigma)$, as desired to show.*

Theorem 4.2 below follows directly from Theorems 3.8 and 3.9, and provides explicit estimates of the convergence radius of the extended Newton method for multiobjective optimization under the classical Lipschitz condition. In particular, assertions (ii) improves the corresponding result in [19, Corollary 6.21], which only asserts the existence of such convergence radius under the stronger assumption than that for assertions (ii). Recall that x^* is a local Pareto optimum of F and ξ_{a^*} is defined by (3.27).

THEOREM 4.2. *Suppose that F satisfies assumption* (3.2) *associated to $(x^*; a^*, \frac{1}{a^*L})$ with $L(\cdot) \equiv L$. Let $x_0 \in \mathbf{B}(x^*, \frac{1}{2(1+\xi_{a^*})a^*L})$. Then, with initial point x_0, we have the following assertions:*

(i) *The sequence* $\{x_n\}$ *generated by Algorithm 3.1 is well-defined and converges quadratically to a local Pareto optimum of* F.

(ii) *If* $\|x_0 - x^*\| \le \frac{-9(1-\sigma)^2+3(1-\sigma)\sqrt{1+9(1-\sigma)^2}}{(1+4\xi_{a^*})a^*L}$, *then any sequence* $\{x_n\}$ *generated by Algorithm 3.2 with initial point* x_0 *is well-defined and converges quadratically to a local Pareto optimum of* F.

Proof. Assertion (i) follows directly from Theorem 3.8. Then, it remains to verify assertion (ii). To do this, write $r := \frac{-9(1-\sigma)^2+3(1-\sigma)\sqrt{1+9(1-\sigma)^2}}{(1+4\xi_{a^*})a^*L}$. Then $r < \frac{1}{2(1+\xi_{a^*})a^*L}$ (due to the fact $-9(1-\sigma)^2 + 3(1-\sigma)\sqrt{1+9(1-\sigma)^2} < \frac{1}{2}$), and, $La^*t < \frac{1}{2(1+\xi_{a^*})} < \frac{1}{2}$ for each $t \in (0, r)$. As $L(\cdot) \equiv L$, one checks that, for each $t \in (0, r)$,

$$\beta_t = \frac{a^* \int_0^t L(u)(t-u)\mathrm{d}u + \xi_{a^*}t}{1 - a^* \int_0^t L(u)\mathrm{d}u} = \frac{\frac{L}{2}a^*t^2 + \xi_{a^*}t}{1 - La^*t} < (\frac{1}{2} + 2\xi_{a^*})t \le (\frac{1}{2} + 2\xi_{a^*})r.$$

Moreover, since $a_tL = \frac{a^*L}{1-a^*Lt} < 2a^*L$, it follows that, for each $t \in (0, r)$,

$$\frac{-3(1-\sigma)h'_{a_t}(\bar{r}^*_{a_t})}{a_tL(\bar{r}^*_{a_t}))} = \frac{3(1-\sigma)\sqrt{1-2a_tL\beta_t}}{a_tL} \ge \frac{3(1-\sigma)\sqrt{1-2(1+4\xi_{a^*})a^*Lr}}{2a^*L} = (\frac{1}{2} + 2\xi_{a^*})r,$$

where the last equality holds by the definition of r. Thus, one checks that $r \in (0, \frac{b_{a^*}}{1+\xi_{a^*}})$ satisfies (3.29), and the conclusion follows from Theorem 3.9. □

4.1.2. Global convergence of Algorithm 3.2. This subsection aims to establish global convergence of Algorithm 3.2 under the classical Lipschitz condition.

The following proposition shows that any accumulation point of a sequence $\{x_n\}$ generated by Algorithm 3.2, where the stepsize $\{\alpha_n\}$ satisfies the Armijo rule, or the Goldstein rule, or the Wolf rule, is a critical point of F.

Proposition 4.3. *Let* $\{x_n\}$ *be a sequence generated by Algorithm 3.2. Then, any accumulation point* x^* *of* $\{x_n\}$ *such that* $\mathrm{D}^2F(x^*)$ *is positive definite and* D^2F *is Lipschitz continuous around* x^*, *is a local Pareto optimum of* F.

Proof. Let x^* be an accumulation point of $\{x_n\}$ such that $\mathrm{D}^2F(x^*)$ is positive definite and D^2F is Lipschitz continuous around x^*. Then, it's easy to show that $\mathrm{D}^2F(\cdot)$ is positive definite around x^*. By (2.2), we only need to verify that x^* is a critical point of F. As x^* is an accumulation point of $\{x_n\}$, there exists a subsequence $\{x_{n_i}\}$ such that $\lim_{i\to\infty} x_{n_i} = x^*$. Let $j \in \{1, \ldots, m\}$. Noting that $\{F_j(x_n)\}$ is monotonically nonincreasing (by Algorithm 3.2) and F_j is continuous, it follows that

$$\lim_{n\to\infty} F_j(x_n) = \lim_{i\to\infty} F_j(x_{n_i}) = F_j(x^*). \tag{4.9}$$

By (i) and (ii) of Lemma 2.4, to complete the proof, it suffices to verify that $\theta(x^*) \ge 0$. To do this, let

$$K_1 := \{i : F_j(x_{n_i} + s(x_{n_i})) \le F_j(x_{n_i}) + \sigma\theta(x_{n_i}) \text{ for all } j = 1, \ldots, m\}.$$

Then, we divide the proof into two cases.

Case 1. K_1 is infinite. Then, there exists a subsequence of $\{x_{n_i}\}$, denoted it by itself, such that

$$F_j(x_{n_i} + s(x_{n_i})) \le F_j(x_{n_i}) + \sigma\theta(x_{n_i}) \quad \text{for all } i \in \mathbb{N} \text{ and } j = 1, \ldots, m. \tag{4.10}$$

In view of Step 4 of Algorithm 3.2, one has that $x_{n_i+1} = x_{n_i} + s(x_{n_i})$. Passing to the limit as $i \to \infty$ in (4.10), we get from (4.9) that $\theta(x^*) \ge 0$ and the proof is complete in this case.

Case 2. K_1 is finite. Then, there exist $j_0 \in \{1, \ldots, m\}$ and a subsequence of $\{x_{n_i}\}$, denoted it by itself, such that

$$F_{j_0}(x_{n_i} + s(x_{n_i})) > F_{j_0}(x_{n_i}) + \sigma\theta(x_{n_i}) \quad \text{for all } i \in \mathbb{N}.$$

Thus, in view of Step 5 in Algorithm 3.2 (cf. (3.1)) and Lemma 2.4(i), we have

$$F_{j_0}(x_{n_i}) - F_{j_0}(x_{n_i+1}) \ge -\sigma\alpha_{n_i}\theta(x_{n_i}) \ge 0,$$

where each $\alpha_{n_i} \in (0, +\infty)$ satisfies the Armijo rule, or the Goldstein rule, or the Wolf rule. This, together with (4.9), implies that $\lim_{i\to\infty} \alpha_{n_i}\theta(x_{n_i}) = 0$. Recall that θ is continuous around x^* (due to Lemma 2.4) and that $\lim_{i\to\infty} x_{n_i} = x^*$. We only need to consider the case when $\lim_{i\to\infty} \alpha_{n_i} = 0$ because, otherwise, one has that $\overline{\lim}_{i\to\infty}\alpha_{n_i} > 0$ and thus

$$\theta(x^*)\overline{\lim}_{i\to\infty}\alpha_{n_i} \ge \lim_{i\to\infty} \alpha_{n_i}\theta(x_{n_i}) = 0;$$

this implies $\theta(x^*) \ge 0$. To proceed, let $\zeta := \max\{\sigma, \beta_1, \beta_2\}$, and define for each n_i

$$\Theta(x_{n_i}) := \max_{k=1,2} \left\{ \frac{F_{j_0}(x_{n_i} + k\alpha_{n_i}s(x_{n_i})) - F_{j_0}(x_{n_i})}{k\alpha_{n_i}}, \nabla F_{j_0}(x_{n_i} + \alpha_{n_i}s(x_{n_i}))^T s(x_{n_i}) \right\}.$$

Then, $\zeta \in (0, 1)$. Below we show that

$$\limsup_{i\to\infty} \Theta(x_{n_i}) \le \theta(x^*) \quad \text{and} \quad \zeta\theta(x_{n_i}) \le \Theta(x_{n_i}) \ \text{ for each } n_i. \tag{4.11}$$

Granting this and noting $\lim_{i\to\infty} \theta(x_{n_i}) = \theta(x^*)$, one checks that $\theta(x^*) \ge \zeta\theta(x^*)$ and so $\theta(x^*) \ge 0$ (as $\zeta \in (0, 1)$), completing the proof.

The second relation in (4.11) holds by the choice of the stepsize α_{n_i} in Step 5 of Algorithm 3.2. To show the first one in (4.11), we first note θ is continuous around x^* and $\{s(x_{n_i})\}$ is bounded (due to Lemma 2.4(iii)). Note further that ∇F_{j_0} is continuous. It follows from $\lim_{i\to\infty} \alpha_{n_i} = 0$ and the inequality $\nabla F_{j_0}(x_{n_i})^T s(x_{n_i}) \le \theta(x_{n_i})$ (due to the definition of θ) that

$$\begin{aligned}
&\limsup\nolimits_{i\to\infty} \nabla F_{j_0}(x_{n_i} + \alpha_{n_i}s(x_{n_i}))^T s(x_{n_i}) \\
&\le \limsup\nolimits_{i\to\infty}((\nabla F_{j_0}(x_{n_i} + \alpha_{n_i}s(x_{n_i})) - \nabla F_{j_0}(x_{n_i}))^T s(x_{n_i}) + \theta(x_{n_i})) \\
&= \limsup\nolimits_{i\to\infty} \theta(x_{n_i}) = \theta(x^*).
\end{aligned}$$

Thus it remains to verify that

$$\limsup_{i\to\infty} \frac{F_{j_0}(x_{n_i} + k\alpha_{n_i}s(x_{n_i})) - F_{j_0}(x_{n_i})}{k\alpha_{n_i}} \le \theta(x^*) \quad \text{for } k = 1, 2. \tag{4.12}$$

To do this, consider a sequence $\{t_{n_i}\} \subseteq (0, +\infty)$ converging to zero. Then we have that

$$\lim_{i\to\infty} \int_0^1 (\nabla F_{j_0}(x_{n_i} + \tau t_{n_i} s(x_{n_i})) - \nabla F_{j_0}(x_{n_i}))^T s(x_{n_i}) \mathrm{d}\tau = 0 \tag{4.13}$$

as ∇F_{j_0} is continuous and $\{s(x_{n_i})\}$ is bounded. Note for each $i \in \mathbb{N}$ that

$$\begin{aligned} \frac{F_{j_0}(x_{n_i} + t_{n_i} s(x_{n_i})) - F_{j_0}(x_{n_i})}{t_{n_i}} = & \int_0^1 (\nabla F_{j_0}(x_{n_i} + \tau t_{n_i} s(x_{n_i})) - \nabla F_{j_0}(x_{n_i}))^T s(x_{n_i}) \mathrm{d}\tau \\ & + \nabla F_{j_0}(x_{n_i})^T s(x_{n_i}). \end{aligned}$$

Hence, thanks again to the inequality $\nabla F_{j_0}(x_{n_i})^T s(x_{n_i}) \le \theta(x_{n_i})$ (due to the definition of θ) and using again the continuity of θ, we conclude from (4.13) that

$$\limsup_{i\to\infty} \frac{F_{j_0}(x_{n_i} + t_{n_i} s(x_{n_i})) - F_{j_0}(x_{n_i})}{t_{n_i}} \le \limsup_{i\to\infty} \theta(x_{n_i}) = \theta(x^*).$$

Applying this fact to $\{\alpha_{n_i}\}$ and $\{2\alpha_{n_i}\}$ in place of $\{t_{n_i}\}$, one sees that (4.12) holds, and the proof is complete. □

Corollary 4.4. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.2. Suppose that the set $\bigcap_{j=1,\ldots,m}\{x \in U : F_j(x) \le F_j(x_0)\}$ is bounded. Then, there exists an accumulation point x^* of $\{x_n\}$. Furthermore, if x^* satisfies that $\mathrm{D}^2F(x^*)$ is positive definite and D^2F is Lipschitz continuous around x^*, then x^* is a local Pareto optimum of F.*

Proof. Note by Algorithm 3.2 that $\{F_j(x_n)\}$ is monotonically nonincreasing for each $j = 1, \ldots, m$. Hence, by assumption, we have that $\{x_n\} \subseteq \bigcap_{j=1,\ldots,m}\{x \in U : F_j(x) \le F_j(x_0)\}$ and so $\{x_n\}$ is bounded. Thus, there exists an accumulation point of $\{x_n\}$. Then, the conclusion follows from Proposition 4.3. □

Now we are ready to establish the global quadratic convergence of a sequence generated by Algorithm 3.2.

Theorem 4.5. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.2. Suppose that $\{x_n\}$ has an accumulation point x^* such that $\mathrm{D}^2F(x^*)$ is positive definite and D^2F is Lipschitz continuous around x^*. Then, x^* is a local Pareto optimum of F and $\{x_n\}$ converges quadratically to x^*.*

Proof. In view of Proposition 4.3, it suffices to show that $\{x_n\}$ converges quadratically to x^*. For this purpose, note by the Lipschitz continuity assumption that there exists a pair of positive numbers (r, L) such that each D^2F satisfies the Lipschitz condition with modulus L on $\mathbf{B}(x^*, r)$. Since each $\nabla^2 F_j(x^*)$ is positive definite by assumption, we can take

$$a^* > \max_{j=1,\ldots,m} \left\{ \frac{1}{rL}, \|\nabla^2 F_j(x^*)^{-1}\| \right\}.$$

Then, F satisfies assumption (3.2) associated to $(x^*; a^*, \frac{1}{a^*L})$ and $L(\cdot) \equiv L$. Let

$$\hat{r} = \frac{-9(1-\sigma)^2 + 3(1-\sigma)\sqrt{1 + 9(1-\sigma)^2}}{(1 + 4\xi_{a^*})a^*L},$$

and let $\{x_{n_i}\} \subseteq \{x_n\}$ be a subsequence such that $\lim_{i\to\infty} x_{n_i} = x^*$. Then there exists $i_0 \in \mathbb{N}$ such that $\|x_{n_{i_0}} - x^*\| \le \hat{r}$. Thus, Theorem 4.2(ii) is applicable to concluding that the sequence

$\{x_n\}_{n=n_{i_0}}^{\infty}$ converges quadratically to a local Pareto optimum of F. This completes the proof. □

4.2. Theorems under the γ-condition. Wang [36] introduced the γ-condition and completely improved Smale's results (cf. [34]) by using the technique of a majorizing function. Below, we present an analogue of γ-condition (with a slight difference) inspired by the one introduced in [36]. Let $r > 0$ and $\gamma > 0$ be such that $r\gamma \leq 1$.

DEFINITION 4.6. *Let $x_0 \in U$ and $r > 0$ be such that $\mathbf{B}(x_0, r) \subseteq U$. $\mathrm{D}F$ is said to satisfy the γ-condition on $\mathbf{B}(x_0, r)$ if,*

$$\|\nabla^3 F_i(x)\| \leq \frac{2\gamma}{(1-\gamma\|x-x_0\|)^3} \qquad \text{for each } i \in \{1, \ldots, m\} \text{ and } x \in \mathbf{B}(x_0, r).$$

REMARK 4.2. *As in [38, Lemma 3], one checks by definition that if F is analytic at x_0, then $\mathrm{D}F$ satisfies the γ-condition on $\mathbf{B}(x_0, \frac{1}{\gamma})$, where $\gamma := \max\limits_{i=1,\ldots,m}\{\sup_{k\geq 2}\|\frac{1}{k!}F_i^{(k+1)}(x_0)\|^{\frac{1}{k-1}}\}$.*

The following proposition shows that the γ-condition of $\mathrm{D}F$ implies the L-average Lipschitz condition of D^2F, the proof of which is easy and so is omitted here.

PROPOSITION 4.7. *Suppose that $\mathrm{D}F$ satisfies the γ-condition on $\mathbf{B}(x_0, r)$. Then, D^2F satisfies the L-average Lipschitz condition on $\mathbf{B}(x_0, \frac{1}{\gamma})$ with the function $L : [0, \frac{1}{\gamma}) \to \mathbb{R}_+$ defined by*

$$L(u) := \frac{2\gamma}{(1-\gamma u)^3} \quad \text{for each } u \in [0, \frac{1}{\gamma}). \tag{4.14}$$

Let $a > 0$ and $\beta \geq 0$. For $L(\cdot)$ given by (4.14), the majoring function h_a defined in (2.13) is reduced to

$$h_a(t) = \beta - t + \frac{a\gamma t^2}{1-\gamma t} \quad \text{for each } 0 \leq t < \frac{1}{\gamma}.$$

Then, it follows from (2.12) that

$$r_a = \left(1 - \sqrt{\frac{a}{1+a}}\right)\frac{1}{\gamma} \quad \text{and} \quad b_a = \left(1 + 2a - 2\sqrt{a(1+a)}\right)\frac{1}{\gamma}.$$

Let $\{t_{a,n}\}$ denote a sequence generated by the classical Newton method for approaching the zeros of h_a with the initial value $t_0 = 0$, and assume

$$\gamma\beta \leq 1 + 2a - 2\sqrt{a(1+a)}.$$

Then, by [36, P.180], the smaller zero r_a^* of h_a and the Newton sequence $\{t_{a,n}\}$ have the following closed forms:

$$r_a^* = \frac{1+\gamma\beta-\sqrt{\tau}}{2(1+a)\gamma}, \quad \text{and} \quad t_{a,n} = \frac{1-\mu^{2^n-1}}{1-\mu^{2^n-1}\eta}r_a^* \quad \text{for each } n \in \mathbb{N}, \tag{4.15}$$

where $\tau := (1+\gamma\beta)^2 - 4(1+a)\gamma\beta \geq 0$,

$$\mu := \frac{1-\gamma\beta-\sqrt{\tau}}{1-\gamma\beta+\sqrt{\tau}} \quad \text{and} \quad \eta := \frac{1+\gamma\beta-\sqrt{\tau}}{1+\gamma\beta+\sqrt{\tau}}. \tag{4.16}$$

Fixing the triple $(x; a, r)$ with $x \in U$ and $(a, r) \in \mathbb{R}^2_+$, we consider the following assumption for $F \in C^3(U, \mathbb{R}^m)$ associated to the triple $(x; a, r)$:

$$\begin{array}{ll} \bullet & \mathrm{D}^2F(x) \text{ is positive definite with each } \|\nabla^2 F_i(x)^{-1}\| \leq a; \\ \bullet & \mathrm{D}F \text{ satisfies the } \gamma\text{-condition on } \mathbf{B}(x,r) \subseteq U. \end{array} \tag{4.17}$$

Then, we have the following theorem about the quadratic convergence criterion of the extended Newton method under the γ-condition.

THEOREM 4.8. *Suppose that F satisfies assumption (4.17) associated to $(x_0; a, r_a^*)$, and $\|s(x_0)\| \leq \beta$. Let μ and η be given by (4.16). Then, with initial point x_0, we have the following assertions:*

(i) *If $\beta \leq \left(1+2a-2\sqrt{a(1+a)}\right)\frac{1}{\gamma}$, then the sequence $\{x_n\}$ generated by Algorithm 3.1 is well-defined, stays in $\mathbf{B}(x_0, r_a^*)$, and converges to a local Pareto optimum $\bar{x} \in \mathbf{B}[x_0, r_a^*]$ with the following error estimation for each $n \in \mathbb{N}$:*

$$\|x_n - \bar{x}\| \leq \frac{(1-\eta)\mu^{2^n-1}}{1-\mu^{2^n-1}\eta} r_a^*. \tag{4.18}$$

(ii) *If $\beta < \left(1+2a-2\sqrt{a(1+a)}\right)\frac{1}{\gamma}$, then $\{x_n\}$ converges quadratically to $\bar{x}$ with the following error estimation for some $N \in \mathbb{N}$:*

$$\|x_{n+1} - \bar{x}\| \leq \frac{\mu(1-\mu^{2^{n+1}-1}\eta)(1-\mu^{2^n-1}\eta)^2}{(1-\eta)(1-\mu^{2^n}(2-\mu^{2^n-1}\eta))^2 r_a^*}\|x_n - \bar{x}\|^2 \quad \text{for each } n \geq N. \tag{4.19}$$

(iii) *If $\beta \leq \frac{3(1-\sigma)(1-\gamma\beta)(1-2\gamma\beta(1+2a)+\gamma^2\beta^2)}{2a\gamma(1+\gamma\beta)^3}$, then any sequence $\{x_n\}$ generated by Algorithm 3.2 coincides with the one generated by Algorithm 3.1, and satisfies (4.18) and (4.19).*

Proof. Assertions (i) and (ii) follow directly from Theorem 3.4 in combination with Proposition 4.7. Then, it remains to show assertion (iii). In fact, as $L(\cdot)$ is given by (4.14), it follows that

$$\frac{-3(1-\sigma)h_a'(r_a^*)}{aL(r_a^*)} = \frac{3(1-\sigma)(1-r_a^*\gamma)((1+a)(1-r_a^*\gamma)^2-a)}{2a\gamma}. \tag{4.20}$$

Note further by (4.15) that

$$r_a^*\gamma = \frac{1+\gamma\beta-\sqrt{\tau}}{2(1+a)} = \frac{(1+\gamma\beta)^2-\tau}{2(1+a)(1+\gamma\beta+\sqrt{\tau})} \leq \frac{2\gamma\beta}{1+\gamma\beta}.$$

Combing this with (4.20) gives that

$$\frac{3(1-\sigma)(1-\gamma\beta)(1-2\gamma\beta(1+2a)+\gamma^2\beta^2)}{2a\gamma(1+\gamma\beta)^3} \leq \frac{-3(1-\sigma)h_a'(r_a^*)}{aL(r_a^*)}.$$

Thus, if $\beta \leq \frac{3(1-\sigma)(1-\gamma\beta)(1-2\gamma\beta(1+2a)+\gamma^2\beta^2)}{2a\gamma(1+\gamma\beta)^3}$, then (3.18) holds. Hence, the conclusion follows from Proposition 4.7 and Theorem 3.5. □

Similarly, we have the following results by using Theorem 3.8 in combination with Proposition 4.7, regarding estimation of the radius of the convergence ball of the extended Newton method for multiobjective optimization under the γ-condition. Recall that x^* is a local Pareto optimum of F and ξ_{a^*} is defined by (3.27).

THEOREM 4.9. *Suppose that F satisfies assumption (4.17) associated to $(x^*; a^*, r_{a^*})$. Let $x_0 \in \mathbf{B}\left(x^*, \frac{1+2a^*-2\sqrt{a^*(1+a^*)}}{(1+\xi_{a^*})\gamma}\right)$. Then, with initial point x_0, we have the following assertions:*

(i) *The sequence $\{x_n\}$ generated by Algorithm 3.1 is well-defined and converges quadratically to a local Pareto optimum of F.*

(ii) *Let $0 < r < \frac{1+2a^*-2\sqrt{a^*(1+a^*)}}{(1+\xi_{a^*})\gamma}$ satisfy (3.29). Then for any $x_0 \in \mathbf{B}(x^*, r)$, any sequence $\{x_n\}$ generated by Algorithms 3.2 with initial point x_0 converges quadratically to a local Pareto optimum of F.*

The advantage of considering the L-average Lipschitz condition rather than the classical Lipschitz condition is shown in the following example, for which Theorem 4.8 is applicable but not Theorem 4.1.

EXAMPLE 4.2. *Consider problem (1.1) with $m = l = 1$ and $F : \mathbb{R} \to \mathbb{R}$ defined by*

$$F(x) := \begin{cases} (\tau - 1)x - \ln(1-x), & x \leq \frac{1}{2}, \\ (\tau+1)x - 2x^2 + \frac{8}{3}x^3 - \frac{5}{6} + \ln 2, & x \geq \frac{1}{2}. \end{cases}$$

where $\tau \in (10\sqrt{2} - 14, 3 - 2\sqrt{2})$. Then one checks that

$$F''(x) = \begin{cases} \frac{1}{(1-x)^2}, & x \leq \frac{1}{2}, \\ -4 + 16x, & x \geq \frac{1}{2}, \end{cases} \quad \text{and} \quad F'''(x) = \begin{cases} \frac{2}{(1-x)^3}, & x \leq \frac{1}{2}, \\ 16, & x \geq \frac{1}{2}. \end{cases}$$

Let $x_0 := 0$ and $\gamma := 1$. It follows that $a := \|F''(x_0)^{-1}\| = 1$, and that F' satisfies the γ-condition on $\mathbf{B}(x_0, 1)$. Note that

$$\beta := \|s(x_0)\| = \| - (F''(x_0))^{-1}F'(x_0)\| = \tau < 3 - 2\sqrt{2}.$$

Therefore, Theorem 4.8 is applicable to concluding that the sequence $\{x_n\}$ generated by Algorithm 3.1 with initial point x_0 converges to a local Pareto optimum of F. We below show that Theorem 4.1 is not applicable. To do this, we first note that F'' is also Lipschitz continuous on $\mathbf{B}(x_0, r)$ with the (least) Lipschitz constant K_r given by

$$K_r := \begin{cases} \frac{2}{(1-r)^3}, & r \leq \frac{1}{2}, \\ 16, & r \geq \frac{1}{2}. \end{cases} \tag{4.21}$$

Now suppose on the contrary that Theorem 4.1 is applicable. Then there exists a positive constant L such that

$$L \geq K_r, \quad r \geq \frac{1 - \sqrt{1 - 2L\tau}}{L} \quad \text{and} \quad \tau \leq \frac{1}{2L} \leq \frac{1}{2K_r}, \tag{4.22}$$

as $a = 1$ and $\beta = \tau$. Recalling $\tau > 10\sqrt{2} - 14 > \frac{1}{32}$, we have that $K_r < 16$, and then it follows from (4.21) that $r < \frac{1}{2}$. Hence $L \geq K_r = \frac{2}{(1-r)^3} \geq 2$. Consequently, by the second inequality in (4.22), we have that $\tau \leq r - \frac{Lr^2}{2}$ and so $\tau \leq r - r^2$. Combining this and the last inequaliity in (4.22), and (4.21), we have that $\tau \leq \min\{\frac{(1-r)^3}{4}, r - r^2\}$. Since the function $r \mapsto \frac{(1-r)^3}{4}$ is decreasing and $r \mapsto r - r^2$ increasing on $[0, \frac{1}{2}]$, it follows that, for each $r \in (0, \frac{1}{2})$,

$$\min\left\{\frac{(1-r)^3}{4}, r - r^2\right\} \leq s_0 - s_0^2 = 10\sqrt{2} - 14,$$

where $s_0 := 3 - 2\sqrt{2}$ is the least positive root of equation $\frac{(1-s)^3}{4} = s - s^2$. Therefore, $\tau \leq 10\sqrt{2} - 14$, which contradicts the choice of τ, and thus Theorem 4.1 is not applicable.

REFERENCES

[1] L. Armijo, *Minimization of functions having continuous partial derivatives*, Pac. J. Math., 16 (1966), pp. 1–3.

[2] U. Bagchi, *Simultaneous minimization of mean and variation of flow time and waiting time in single machine systems*, Oper. Res., 37 (1989), pp. 118–125.

[3] J. Branke, K. Deb, K. Miettinen, and R. Slowinski, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, Springer, Berlin, 2008.

[4] R. S. Burachik, C. Y. Kaya, and M. M. Rizvi, *A new scalarization technique and new algorithms to generate Pareto fronts*, SIAM J. Optim., 27 (2017), pp. 1010–1034.

[5] E. Carrizosa and J. B. G. Frenk, *Dominating sets for convex functions with some applications*, J. Optim. Theory Appl., 96 (1998), pp. 281–295.

[6] G. Y. Chen, X. X. Huang, and X. Q. Yang, *Vector Optimization: Set-Valued and Variational Analysis*, Springer, Berlin, 2005.

[7] I. Das and J. E. Dennis, *Normal-boundary intersection: A new method for generating Pareto optimal points in nonlinear multicriteria optimization problems*, SIAM J. Optim., 8 (1998), pp. 631–657.

[8] L. M. G. Drummond and A. N. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl., 28 (2004), pp. 5–29.

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

[10] G. Eichfelder, *Adaptive Scalarization Methods in Multiobjective Optimization*, Springer Verlag, Berlin, Heidelberg, 2008.

[11] H. Eschenauer, J. Koski, and A. Osyczka, *Multicriteria Design Optimization*, Springer, Berlin, 1990.

[12] O. P. Ferreira, *Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle*, IMA J. Numer.Anal., 29 (2009), pp. 746-759.

[13] O. P. Ferreira, *A robust semi-local convergence analysis of Newton's method for cone inclusion problems in Banach spaces under affine invariant majorant condition,* J. Comput. Appl. Math. 279 (2015), pp. 318-335.

[14] O. P. Ferreira, M. L. N. Goncalves, and P. R. Oliveira, *Local convergence analysis of the Gauss-Newton method under a majorant condition*, J. Complexity 27(1) (2011), pp. 111-125.

[15] O. P. Ferreira, M. L. N. Goncalves, and P. R. Oliveira, *Convergence of the Gauss–Newton method for convex composite optimization under a majorant condition*, SIAM J. Optim., 23 (2013), pp. 1757–1783.

[16] O. P. Ferreira and G. N. Silva, *Kantorovich's theorem on Newton's method for solving strongly regular generalized equation,* SIAM J. Optim. 27(2) (2017), pp. 910-926.

[17] J. Fliege, *OLAF – A general modeling system to evaluate and optimize the location of an air polluting facility*, OR Spektrum., 23 (2001), pp. 117–136.

[18] J. Fliege, *An efficient interior-point method for convex multicriteria optimization problems*, Math. Oper. Res., 31 (2006), pp. 825–845.

[19] J. Fliege, L. M. G. Drummond, and B. F. Svaiter, *Newton's method for multiobjective optimization*, SIAM J. Optim., 20 (2009), pp. 602–626.

[20] J. Fliege and B. F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res., 51 (2000), pp. 479–494.

[21] F. Giannessi, *Vector Variational Inequalities and Vector Equilibria: Mathematical Theories*, Kluwer Academic Publishers, Dordrecht, 2000.

[22] M. Gravel, I. M. Martel, R. Madeau, W. Price, and R. Tremblay, *A multicriterion view of optimal ressource allocation in job-shop production*, European J. Oper. Res., 61 (1992), pp. 230–244.

[23] J. Jahn, *Vector Optimization: Theory, Applications, and Extensions*, Springer, New York, 2011.

[24] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Pergamon, Oxford, 1982.

[25] I. Y. Kim and O. L. de Weck, *Adaptive weighted sum method for bi-objective optimization: Pareto fron generation*, Struct. Multidiscip. O., 29 (2005), pp. 149–158.

[26] K C. Kiwiel and K. Murty,*Convergence of the steepest descent method for minimization quasiconvex functions*, J. Optim. Theory Appl., 89(1996), pp. 221–226.

[27] T. M. Leschine, H. Wallenius, and W. A. Verdini, *Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions*, European J. Oper. Res., 56 (1992), pp. 278–289.

[28] C. Li and K. F. Ng, *Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization*, SIAM J. Optim., 18 (2007), pp. 613–642.

[29] C. Li and J. H. Wang, *Newton's method on Riemannian manifolds: Smale's point estimate theory under the modified γ-condition*, IMA J. Numer.Anal., 26 (2006), pp. 228–251.

[30] C. Li and J. H. Wang, *Newton's method for sections on Riemannian manifolds: Generalized covariant α-Theory*, J. Complexity, 24 (2008), pp. 423–451.

[31] I. Necedal and S. J. Wright, *Numerical Optimization. Springer Series in Operations Research*, Springer, New York ,1999.

[32] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, SIAM, Philadelphia, 2000.

[33] D. Prabuddha, J. B. Gosh, and C. E. Wells, *On the minimization of completion time variance with a bicriteria extension*, Oper. Res., 40 (1992), pp. 1148–1155.

[34] S. Smale, *Complexity theory and numerical analysis*, ACTA Numer., 6 (1997), pp. 523–551.

[35] F. Sourd and O. Spanjaard, *A multiobjective branch-and-bound framework: Application to the biobjective spanning tree problem*, Informs. J. Comput., 20 (2008), pp. 472–484.

[36] X. H. Wang, *Convergence of Newton's method and inverse function theorem in Banach spaces*, Math. Comput., 68 (1999), pp. 169–186.

[37] X. H. Wang, *Convergence of Newton's method and uniqueness of the solution of equations in Banach space*, IMA J. Numer. Anal., 20 (2000), pp. 123–134.

[38] X. H. Wang, D. F. Han, and F. Y. Sun, *Point estimations on deformated Newton's iteration,* Math. Num. Sin., 12 (1990), pp. 145-156; Chinese J. Num. Math. Appl., 12(1990), pp. 1-13.

[39] D. J. White, *Epsilon-dominating solutions in mean-variance portfolio analysis*, European J. Oper. Res., 105 (1998), pp. 457–466.

[40] T. J. Ypma, *Historical development of the Newton-Raphson method*, SIAM Rev., 37 (1995), pp. 531–551.

[41] P. J. Zufiria and J. A. Álvarez-Cubero, *Generalized lexicographic multiObjective combinatorial optimization. Application to cryptography*, SIAM J. Optim., 27 (2017), pp. 2182–2201.