ON CONVERGENCE RATES OF LINEARIZED PROXIMAL ALGORITHMS FOR CONVEX COMPOSITE OPTIMIZATION WITH APPLICATIONS

YAOHUA HU*, CHONG LI†, AND XIAOQI YANG‡

Abstract. In the present paper, we investigate a linearized proximal algorithm (LPA) for solving a convex composite optimization problem. Each iteration of the LPA is a proximal minimization of the convex composite function with the inner function being linearized at the current iterate. The LPA has the attractive computational advantage that the solution of each subproblem is a singleton, which avoids the difficulty as in the Gauss-Newton method (GNM) of finding a solution with minimum norm among the set of minima of its subproblem, while it still maintains the same local convergence rate as that of the GNM. Under the assumptions of local weak sharp minima of order p ($p \in [1, 2]$) and a quasi-regularity condition, we establish a local superlinear convergence rate for the LPA. We also propose a globalization strategy for the LPA based on a backtracking line-search and an inexact version of the LPA. We further apply the LPA to solve a (possibly nonconvex) feasibility problem, as well as a sensor network localization problem. Our numerical results illustrate that the LPA meets the demand for an efficient and robust algorithm for the sensor network localization problem.

Key words. Convex composite optimization, linearized proximal algorithm, weak sharp minima, quasi-regularity condition, feasibility problem, sensor network localization

AMS subject classifications. Primary, 65K05, 49M37; Secondary, 90C26

1. Introduction.

We consider the following convex composite optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) := h(F(x)), \tag{1.1}$$

where the outer function $h : \mathbb{R}^m \to \mathbb{R}$ is convex, and the inner function $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable. We denote by $h_{\min}$ and C the minimum value and the set of minima for the function h respectively, that is,

$$h_{\min} := \min_{y\in\mathbb{R}^m} h(y) \quad \text{and} \quad C := \arg\min_{y\in\mathbb{R}^m} h(y). \tag{1.2}$$

The convex composite optimization framework (1.1) provides a unified framework of a wide variety of important optimization problems, such as convex inclusions, nonsmooth and nonconvex optimization, penalty methods for nonlinear programming and regularized minimization problems; see [8, 13, 18, 31, 36] and references therein. Moreover, this model provides a unified framework for the development and analysis of optimization algorithms.

The development of optimization algorithms for solving problem (1.1) has attracted a great amount of

*Corresponding author. College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China (mayhhu@szu.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (grant 11526144) and The HD Video R & D Platform for Intelligent Analysis and Processing in Guangdong Engineering Technology Research Centre of Colleges and Universities (grant GCZX-A1409).

†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).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (mayangxq@polyu.edu.hk). This author was supported by the Research Grants Council of Hong Kong (PolyU 5292/13E) and the National Natural Science Foundation of China (grant 11431004).

attention. The famous Gauss-Newton method (GNM) has been extensively applied to solve problem (1.1) and is stated as follows.

ALGORITHM 1.1. Given $\rho \geq 1$, $\Delta \in (0, +\infty]$ and $x_0 \in \mathbb{R}^n$. Having x_k, the next iterate x_{k+1} is generated as follows. If $h(F(x_k)) = \min\{h(F(x_k) + F'(x_k)d) : \|d\| \leq \Delta\}$, then stop; otherwise, choose $d_k \in D_\Delta(x_k) := \arg\min_{\|d\|\leq\Delta}\{h(F(x_k) + F'(x_k)d)\}$ such that $\|d_k\| \leq \rho \operatorname{dist}(0, D_\Delta(x_k))$ is satisfied, and set $x_{k+1} = x_k + d_k$.

Many articles have been devoted to establish a local quadratic convergence rate of the GNM; see [9, 10, 21, 39] and references therein. In particular, Burke and Ferris [10] made a great contribution in the development of the GNM, whose work extended that of Womersley [39] without the assumption of the set of minima for h being a singleton, and also proposed a globalization strategy based on a backtracking line-search. Their work is based on the following two assumptions:

(A1) C is the set of weak sharp minima for h, and
(A2) a regularity condition of the inclusion $F(x) \in C$ holds

(see Definitions 2.2 and 2.5 for the details). Under assumptions (A1) and (A2), Burke and Ferris [10] proved the local quadratic convergence rate, as well as a global quadratic convergence rate of Algorithm 1.1 when a globalization strategy is included. Without assumption (A1), Li and Wang [21] established the same local quadratic convergence rate as that of Burke and Ferris [10] and they also proposed an inexact version of Algorithm 1.1 and established its local superlinear convergence.

However, from the practical point of view, it is inefficient to implement Algorithm 1.1, because the search direction d_k is found among the set $D_\Delta(x_k)$ possibly with minimum norm, and it is difficult to find d_k for many applications, especially for large scale problems. Hence, numerical algorithms with low cost and high efficiency are required for solving the convex composite optimization problem. The proximal point algorithm was originally presented by Martinet [23] and developed by Rockafellar [30] for finding a zero of a maximal monotone operator. Nowadays, the idea of proximal point algorithm is very popular and extensively applied in designing algorithms for structured optimization problems, and several variants of proximal point algorithms were proposed, such as the accelerated proximal point algorithm [25, 37], the proximal gradient algorithm [2, 40] and the alternating direction method of multipliers [7, 14]. In 2008, Lewis and Wright [18] used this idea to propose a linearized proximal algorithm, named ProxDescent, for solving the convex composite optimization problem (1.1) (or a more general problem where the outer function h is assumed to be extended-value and prox-regular, not necessarily convex). Each subproblem of the ProxDescent is a proximal minimization of the composite function with the inner function being linearized at the current iterate and the stepsize being updated to maintain a descent property. Thus their algorithm is a descent one. Their work is of high theoretical significance in investigating the properties of local solutions of the subproblem. They also proved a global convergence result of the ProxDescent, that is the cluster points of the sequence produced by ProxDescent are stationary points of (1.1). Recently, Sagastizábal [32] proposed a composite proximal bundle algorithm for solving (1.1) with a positively homogeneous convex function h, by employing a bundle of subgradient information of the outer function and gradient of the inner function at the current iterate, and established that the sequence produced by the algorithm either stops finitely or has a cluster point being a stationary point of (1.1).

In the present paper, we study the linearized proximal algorithm, named LPA, proposed in [18] but using general stepsizes for solving (1.1) and investigate its local convergence rates. As general stepsizes are used, the resulting algorithm is generally not a descent one. Hence our algorithm is significantly different from the ProxDescent. In fact, the introduction of the LPA was motivated by both the GNM and the proximal point algorithm. The LPA shares many of their advantages, as well as overcomes their disadvantages. The subproblem of the LPA is an unconstrained strongly convex optimization problem, which is easier to solve than that of the GNM. Consequently, the LPA has the attractive computational advantage that the solution of each subproblem is a singleton, which avoids the difficulty of finding a solution with minimum norm among the set of minima of its subproblem as in the GNM, while it still maintains the same local convergence rate as that of the GNM. Under the assumptions of local weak sharp minima of order p ($p \in [1, 2]$) and a quasi-regularity condition, we establish the local superlinear convergence rate for the LPA. This is the main contribution of the present paper. Based on a backtracking line-search, we also propose a globalization strategy for the LPA and obtain the global superlinear convergence result. Furthermore, we extend the LPA to the inexact setting and provide the superlinear convergence results of the inexact LPA similar to that of (exact) LPA. In particular, as a consequence of our main result, [18, Theorem 7.4] can be partially improved in the sense that any sequence generated by the ProxDescent for solving the convex composite optimization problem (1.1) is shown to converge to a global solution of (1.1) at a superlinear rate under the weak sharp minima and the regular condition; while [18, Theorem 7.4] only presented the convergence to a stationary point; see Remark 3.4 for details. Moreover, to the best of our knowledge, our results of the convergence rate on the LPA type algorithms (e.g., Theorems 3.2, 3.4 and 3.5) seem new in the literature.

The motivation of our work also stems from various applications. In particular, we consider the (possibly nonconvex) feasibility problem as an application of the convex composite optimization, which is at the core of the modeling of many problems in various areas of mathematics and physical sciences. For example, there has been an increasing use of ad hoc wireless sensor networks for monitoring the environmental information across an entire physical space. Typical networks of this type consist of a large number of inexpensive wireless sensors deployed in a geographical area with the ability to communicate with their neighbors within a limited radio range. The sensor network localization problem is to determine the positions of the sensors in a network by using the given incomplete pairwise distance measurements. However, the use of the GPS system is very an expensive solution to this requirement as a huge number of sensors are required. Therefore, there is a great demand for developing efficient and robust algorithms that can estimate or localize sensor positions in a network by using only the mutual distance measures that the wireless sensors receive from their neighbors. The sensor network localization problem can be cast into a nonconvex feasibility problem. We further reformulate the feasibility problem as framework (1.1) and then apply the LPA to solve the feasibility problem, as well as the sensor network localization problem. In particular, when applied to the sensor network localization problem, the numerical results illustrate that the LPA achieves the more precise solution, costs less CPUtime and requires less information (the small radio range and the few anchors) than that of the semidefinite relaxation technique; see the explanations on page 24 for details.

The paper is organized as follows. In section 2, we present the notation and preliminary results used in the present paper. In section 3, we investigate the local convergence property of the LPA under the assumptions of local weak sharp minima of order p and the quasi-regularity condition, and propose the globalized LPA and inexact LPA, as well as their convergence behavior. Applications to the feasibility problem and numerical experiments on the sensor network localization problem are demonstrated in section 4.

2. Notation and Preliminary Results. We consider the n-dimensional Euclidean space $\mathbb{R}^n$. We view a vector as a column one, and denote by $\langle x, y\rangle$ the inner product of two vectors $x, y \in \mathbb{R}^n$. We use $\|x\|$ to denote the standard Euclidean norm of x, that is, $\|x\| = \sqrt{\langle x, x\rangle}$. For $x \in \mathbb{R}^n$ and $\delta \in \mathbb{R}_+$, $\mathbf{B}(x, \delta)$

denotes the open ball of radius δ centered at x. For a closed convex subset $Z \subseteq \mathbb{R}^n$, the negative polar of Z, denoted by $Z^{\ominus}$, is defined by

$$Z^{\ominus} := \{y : \langle y, z\rangle \le 0, \text{ for each } z \in Z\}. \tag{2.1}$$

For a point x and a set Z, the Euclidean distance of x from Z, denoted by $\operatorname{dist}(x, Z)$, is defined by

$$\operatorname{dist}(x, Z) := \inf_{z\in Z} \|x - z\|.$$

We adopt the convention that $\operatorname{dist}(x, \emptyset) = +\infty$ for the whole paper. For $f : \mathbb{R}^n \to \mathbb{R}$, $\epsilon \ge 0$ and $X \subseteq \mathbb{R}^n$, the ϵ-optimal solution set of f over X is defined by

$$\epsilon\text{-}\arg\min_{x\in X} f(x) := \{x \in X : f(x) \le \inf_{y\in X} f(y) + \epsilon\}.$$

For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) := \{g : f(y) \ge f(x) + \langle g, y - x\rangle, \text{ for each } y \in \mathbb{R}^n\}.$$

For $F : \mathbb{R}^n \to \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$, we say that F is a $C^{1,1}$ function on X, denoted by $F \in C^{1,1}_L(X)$, if F is continuously differentiable with a Lipschitz continuous gradient F' on X, i.e., there exists $L > 0$ such that

$$\|F'(x) - F'(y)\| \le L\|x - y\| \quad \text{for each } x, y \in X.$$

A well-known property of the $C^{1,1}$ function is presented as follows; see [3, Proposition A.24].

Lemma 2.1. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $X \subseteq \mathbb{R}^n$. If $F \in C^{1,1}_L(X)$, then for all $x, y \in X$, it holds that*

$$\|F(y) - F(x) - F'(x)(y - x)\| \le \frac{L}{2}\|y - x\|^2.$$

The concepts of weak sharp minima were introduced by Burke and Ferris [12], and have been extensively studied and widely used to analyze the convergence properties of many algorithms; see [10, 21, 43, 44] and references therein. We recall in the following definition the concepts of weak sharp minima: items (b) and (c) were taken from Burke and Ferris [12] and Burke and Deng [11], respectively. Let $h : \mathbb{R}^m \to \mathbb{R}$, and let $h_{\min}$ and C be given in (1.2).

Definition 2.2. *Let $S \subseteq \mathbb{R}^m$ and $\eta > 0$. C is said to be*

(a) *the set of weak sharp minima for h on S with modulus η if*

$$h(y) - h_{\min} \ge \eta \operatorname{dist}(y, C) \quad \text{for each } y \in S;$$

(b) *the set of (global) weak sharp minima for h with modulus η if C is the set of weak sharp minima for h on $\mathbb{R}^n$ with modulus η;*
(c) *the set of local weak sharp minima for h at $\bar{y} \in C$ if there exist $\epsilon > 0$ and $\eta_\epsilon > 0$ such that C is the set of weak sharp minima for h on $\mathbf{B}(\bar{y}, \epsilon)$ with modulus η_ϵ.*

One natural extension of these concepts is that of (global and local) weak sharp minima of order p ($p \ge 1$); see [6, 16, 27, 35] and references therein. Item (b) in the following definition was introduced by Studniarski and Ward [35].

Definition 2.3. *Let $S \subseteq \mathbb{R}^m$, $\eta > 0$ and $p \ge 1$. C is said to be*

(a) *the set of weak sharp minima of order* p *for* h *on* S *with modulus* η *if*

$$h(y) - h_{\min} \geq \eta \, \mathrm{dist}^p(y, C) \quad \text{for each } y \in S; \tag{2.2}$$

(b) *the set of local weak sharp minima of order* p *for* h *at* $\bar{y} \in C$ *if there exist* $\epsilon > 0$ *and* $\eta_\epsilon > 0$ *such that* C *is the set of weak sharp minima of order* p *for* h *on* $\mathbf{B}(\bar{y}, \epsilon)$ *with modulus* η_ϵ.

REMARK 2.1. We define the weak sharp minima constant of order p for h on S by

$$\eta_p(h; S) := \inf_{y \in S \setminus C} \frac{h(y) - h_{\min}}{\mathrm{dist}^p(y, C)},$$

and the local weak sharp minima constant of order p for h at $\bar{y} \in C$ by

$$\eta_p(h; \bar{y}) := \sup_{\epsilon > 0} \inf_{y \in \mathbf{B}(\bar{y}, \epsilon) \setminus C} \frac{h(y) - h_{\min}}{\mathrm{dist}^p(y, C)}. \tag{2.3}$$

Clearly, C is the set of weak sharp minima of order p for h on S (resp. the set of local weak sharp minima of order p for h at $\bar{y}$) if and only if $\eta_p(h; S) > 0$ (resp. $\eta_p(h; \bar{y}) > 0$).

The following lemma provides a useful property of the composition of a function, satisfying the weak sharp minima of order p, and a $C^{1,1}$ function, which will repeatedly be used in the study of the convergence behavior of the LPA.

LEMMA 2.4. *Let* $S \subseteq \mathbb{R}^m$, $\eta > 0$ *and* $p \geq 1$. *Let* C *be the set of weak sharp minima of order* p *for* h *on* S *with modulus* η. *Suppose that* $F \in C_L^{1,1}(X)$. *Then, for all* $x, y \in X$ *satisfying* $F(x) + F'(x)(y - x) \in S$, *it holds that*

$$\mathrm{dist}(F(y), C) \leq \frac{1}{2} L \|y - x\|^2 + \eta^{-\frac{1}{p}} \left(h(F(x) + F'(x)(y - x)) - h_{\min} \right)^{\frac{1}{p}}. \tag{2.4}$$

Proof. By Lemma 2.1 and (2.2), it follows that

$$\begin{aligned} \mathrm{dist}(F(y), C) &\leq \|F(y) - F(x) - F'(x)(y - x)\| + \mathrm{dist}(F(x) + F'(x)(y - x), C) \\ &\leq \tfrac{1}{2} L \|y - x\|^2 + \eta^{-\frac{1}{p}} \left(h(F(x) + F'(x)(y - x)) - h_{\min} \right)^{\frac{1}{p}}. \end{aligned}$$

The proof is complete. □

Associated to problem (1.1), we consider the inclusion

$$F(x) \in C, \tag{2.5}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable, and $C \subseteq \mathbb{R}^m$ is defined by (1.2). For $x \in \mathbb{R}^n$, let $D(x)$ be defined by

$$D(x) := \{d \in \mathbb{R}^n : F(x) + F'(x)d \in C\}. \tag{2.6}$$

The quasi-regularity condition in the following definition provides a local bound on the set $D(x)$.

DEFINITION 2.5. *Let* $S \subseteq \mathbb{R}^m$ *and* $\beta > 0$. *We say that*

(a) *a point* $\bar{x} \in \mathbb{R}^n$ *is a regular point of inclusion* (2.5) *if*

$$\ker(F'(\bar{x})^T) \cap (C - F(\bar{x}))^{\ominus} = \{0\};$$

(b) *inclusion* (2.5) *is said to satisfy the quasi-regularity condition on* S *with constant* β *if*

$$\operatorname{dist}(0, D(x)) \leq \beta \operatorname{dist}(F(x), C) \quad \text{for each } x \in S \tag{2.7}$$

(and so $D(x) \neq \emptyset$ *for each* $x \in S$*);*

(c) *a point* $\bar{x} \in \mathbb{R}^n$ *is a quasi-regular point of inclusion* (2.5) *if there exist* $r > 0$ *and* $\beta_r > 0$ *such that inclusion* (2.5) *satisfies the quasi-regularity condition on* $\mathbf{B}(\bar{x}, r)$ *with constant* β_r*.*

REMARK 2.2.

(a) The notion of a regular point was introduced and applied to establish the local convergence rate of the GNM for problem (1.1) in Burke and Ferris [10]. By [10, Proposition 3.3], one sees that any regular point of inclusion (2.5) is a quasi-regular point.

(b) The notion of the quasi-regular point was originally introduced by Li and Ng [20]. Recall from [20] that a point $\bar{x} \in \mathbb{R}^n$ is a quasi-regular point of inclusion (2.5) if there exist $r > 0$ and an increasing positive-valued function $\kappa(\cdot)$ on $[0, r)$ such that

$$\operatorname{dist}(0, D(x)) \leq \kappa(\|x - \bar{x}\|) \operatorname{dist}(F(x), C) \quad \text{for each } x \in \mathbf{B}(\bar{x}, r). \tag{2.8}$$

One can check directly by definition that this is equivalent to the concept of the quasi-regular point given in Definition 2.5.

(c) We define the quasi-regularity constant $\beta(\bar{x})$ as the infimum over all positive constants β_r for which inclusion (2.5) satisfies the quasi-regularity condition on $\mathbf{B}(\bar{x}, r)$ for some positive radius r, that is,

$$\beta(\bar{x}) := \inf_{r>0} \{\beta : (2.7) \text{ holds on } \mathbf{B}(\bar{x}, r)\}. \tag{2.9}$$

Then $\bar{x} \in \mathbb{R}^n$ is a quasi-regular point of inclusion (2.5) if and only if $\beta(\bar{x}) < +\infty$.

3. Linearized Proximal Algorithms and Convergence Analysis. Throughout the whole section, we always assume that $p \in [1, 2]$, unless otherwise specified. In this section, we shall investigate a linearized proximal algorithm (LPA) to solve problem (1.1), and establish the local convergence behavior of the LPA under the assumptions of the local weak sharp minima of order p and the quasi-regularity condition. We also provide a globalization strategy for the LPA by virtue of the backtracking line-search, and an inexact version of the LPA, together with their convergence analysis.

We proceed with the (inexact) linearized proximal mapping and some basic properties. Let $v > 0$ and $\epsilon \geq 0$. The linearized proximal mapping $\mathcal{LP}_{v,\epsilon} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined as, for each $x \in \mathbb{R}^n$, the ϵ-optimal solution set of the following optimization problem:

$$\min_{d \in \mathbb{R}^n} f(x; d) := h(F(x) + F'(x)d) + \frac{1}{2v}\|d\|^2, \tag{3.1}$$

that is,

$$\mathcal{LP}_{v,\epsilon}(x) := \epsilon\text{-arg}\min_{d \in \mathbb{R}^n} \left\{ h(F(x) + F'(x)d) + \frac{1}{2v}\|d\|^2 \right\}. \tag{3.2}$$

In the special case when $\epsilon = 0$, we write $\mathcal{LP}_v(x)$ for $\mathcal{LP}_{v,0}(x)$ for simplicity; note that, $\mathcal{LP}_v(x)$ is a singleton for each $x \in \mathbb{R}^n$. The following lemma presents some useful properties of the linearized proximal mapping.

LEMMA 3.1. *Let* $v > 0$ *and* $\epsilon > 0$*, and let* $x \in \mathbb{R}^n$ *satisfying* $D(x) \neq \emptyset$ *and* $d \in \mathcal{LP}_{v,\epsilon}(x)$*. Then the following statements hold:*

(i) $\|d\|^2 \le \mathrm{dist}^2(0, D(x)) + 2v\epsilon$,
(ii) $h(F(x) + F'(x)d) \le h_{\min} + \frac{1}{2v}\mathrm{dist}^2(0, D(x)) + \epsilon$.

Proof. Note by (2.6) that $h(F(x) + F'(x)\tilde{d}) = h_{\min}$ for each $\tilde{d} \in D(x)$. Then one has by definition (cf. (3.2)) that

$$h(F(x) + F'(x)d) + \frac{1}{2v}\|d\|^2 \le h(F(x) + F'(x)\tilde{d}) + \frac{1}{2v}\|\tilde{d}\|^2 + \epsilon = h_{\min} + \frac{1}{2v}\|\tilde{d}\|^2 + \epsilon.$$

Taking the infimum over $D(x)$ on the right-hand side of the above inequality, we obtain

$$h(F(x) + F'(x)d) + \frac{1}{2v}\|d\|^2 \le h_{\min} + \frac{1}{2v}\mathrm{dist}^2(0, D(x)) + \epsilon. \tag{3.3}$$

Thus, (i) and (ii) follow. □

3.1. Linearized Proximal Algorithm. This subsection is devoted to the study of the LPA. Note that the outer function h in convex composite optimization problem (1.1) is convex. The ProxDescent [18] for solving (1.1) is a special case of the following LPA (as the stepsize in ProxDescent is selected such that a descent property is satisfied: $h(F(x_k)) - h(F(x_k + d_k)) \ge \sigma\left(h(F(x_k)) - h(F(x_k) + F'(x_k)d_k)\right)$ for some $\sigma \in (0, 1)$).

Algorithm 3.1. Given an initial point $x_0 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k, we calculate the search direction $d_k := \mathcal{LP}_{v_k}(x_k)$ by solving the optimization problem (3.1) (with x_k in place of x). If $d_k = 0$, then it stops; otherwise, we set $x_{k+1} = x_k + d_k$.

Remark 3.1. In the special case when $h := \frac{1}{2}\|\cdot\|^2$, Algorithm 3.1 is reduced to the well-known Levenberg-Marquardt method [24] for solving the following nonlinear least squares problem:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|F(x)\|^2.$$

Indeed, applying Algorithm 3.1 to this problem, the first order optimality condition of (3.1) (with x_k in place of x) implies that

$$0 = F'(x_k)^\top(F(x_k) + F'(x_k)d_k) + \frac{d_k}{v}.$$

Thus, the closed formula of the iteration of Algorithm 3.1 is given by

$$x_{k+1} = x_k + d_k = x_k - v\left(I + vF'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k) \text{ for each } k = 0, 1, \ldots,$$

which is the Levenberg-Marquardt method (also the trust region method for the nonlinear least squares problem [42]).

The main theorem of this subsection is as follows. It provides some sufficient conditions around initial points ensuring the convergence of Algorithm 3.1. For the convergence results in the remainder of the present paper (i.e., Theorems 3.2, 3.4 and 3.5 and Corollaries 3.3 and 3.6), our analysis is, without loss of generality, only focused on the special case when the stepsizes are chosen to be a constant, that is, $v_k \equiv v$, unless otherwise specified, as the corresponding convergence results for the general case can be established similarly; see the explanation in Remark 3.2(a) for more details.

Theorem 3.2. *Let $\eta > 0$, $\beta > 0$ and $\bar{\delta} > 0$. Let $\bar{x} \in \mathbb{R}^n$, and let C be the set of weak sharp minima of order p for h on $\mathbf{B}(F(\bar{x}), \bar{\delta})$ with modulus η. Suppose that $F \in C_L^{1,1}(\mathbf{B}(\bar{x}, \bar{\delta}))$, and that inclusion* (2.5) *satisfies the quasi-regularity condition on* $\mathbf{B}(\bar{x}, \bar{\delta})$ *with constant* β*. Suppose further that there exists* $\delta > 0$ *such that*

(a) $\delta \le \min\left\{\frac{\bar{\delta}}{2}, \frac{2\bar{\delta}}{5L_0}\right\}$,
(b) $\mathrm{dist}(F(\bar{x}), C) < \frac{\delta}{2\beta}$,
(c) $\beta\left(L\delta + 2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right) \le 1$,

where L_0 *is the Lipschitz constant for* F *on* $\mathbf{B}(\bar{x}, \bar{\delta})$*. Then, there exists a neighborhood* $N(\bar{x})$ *of* $\bar{x}$ *such that, for any* $x_0 \in N(\bar{x})$*, the sequence* $\{x_k\}$ *generated by Algorithm 3.1 with initial point* x_0 *converges at a rate of* $\frac{2}{p}$ *to a solution* x^* *satisfying* $F(x^*) \in C$*.*

Proof. Set

$$\bar{\beta} := \frac{\delta - 2\beta \mathrm{dist}(F(\bar{x}), C)}{2\beta L_0} \quad \text{and} \quad r_0 := \min\{\delta, \bar{\beta}\}. \tag{3.4}$$

Then $r_0 > 0$ due to assumption (b). Let $x_0 \in N(\bar{x}) := \mathbf{B}(\bar{x}, r_0)$. Then one has that $\|x_0 - \bar{x}\| \le r_0 \le \delta < \bar{\delta}$ (by assumption (a)). Thus, by the choice of L_0, we have that

$$\|F(x_0) - F(\bar{x})\| \le L_0 r_0 \le L_0 \bar{\beta},$$

and it follows that

$$\mathrm{dist}(F(x_0), C) \le \|F(x_0) - F(\bar{x})\| + \mathrm{dist}(F(\bar{x}), C) \le L_0\bar{\beta} + \mathrm{dist}(F(\bar{x}), C) = \frac{\delta}{2\beta}, \tag{3.5}$$

where the last inequality follows from the definition of $\bar{\beta}$ in (3.4). We shall show by induction that the following estimates hold for all $i = 0, 1, 2, \ldots$:

$$\|x_i - \bar{x}\| < 2\delta (\le \bar{\delta}) \quad \text{and} \quad \mathrm{dist}(F(x_i), C) \le \frac{\delta}{\beta}\left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^i + i}. \tag{3.6}$$

Note that (3.6) holds for $i = 0$ (thanks to the choice of x_0 and (3.5)). Now assume that (3.6) holds for each $i \le k-1$. Then, by the assumed quasi-regularity condition, $D(x_i) \ne \emptyset$. Thus, Lemma 3.1 is applicable (with x_i, d_i, 0 in place of x, d, ϵ), and we conclude that

$$\|d_i\| \le \mathrm{dist}(0, D(x_i)) \le \beta \mathrm{dist}(F(x_i), C) \le \delta \left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^i + i} \quad \text{for each } i = 0, \ldots, k-1. \tag{3.7}$$

Hence

$$\|x_k - \bar{x}\| \le \sum_{i=0}^{k-1} \|d_i\| + \|x_0 - \bar{x}\| < \delta \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^i + i} + r_0.$$

Since $p \le 2$ and $r_0 \le \delta$ (see (3.4)), it follows that

$$\|x_k - \bar{x}\| < \delta + r_0 \le 2\delta. \tag{3.8}$$

Since $x_{k-1} \in \mathbf{B}(\bar{x}, 2\delta)$ and by the choice of L_0 ($\bar{\delta} \ge 2\delta$), one has that $\|F(x_{k-1}) - F(\bar{x})\| \le 2L_0\delta$ and $\|F'(x_{k-1})\| \le L_0$. Thus, by (3.7), it follows that

$$\|F(x_{k-1}) + F'(x_{k-1})d_{k-1} - F(\bar{x})\| \le \|F(x_{k-1}) - F(\bar{x})\| + \|F'(x_{k-1})\|\|d_{k-1}\| \le \frac{5}{2}L_0\delta \le \bar{\delta}$$

(due to assumption (a)). Hence Lemma 2.4 is applicable (with x_{k-1}, x_k, $\mathbf{B}(F(\bar{x}),\bar{\delta})$, $\mathbf{B}(\bar{x},\bar{\delta})$ in place of x, y, S, X), and we obtain that

$$\mathrm{dist}(F(x_k),C)\leq\frac{L}{2}\|d_{k-1}\|^2+\left(\frac{1}{\eta}\right)^{\frac{1}{p}}(h(F(x_{k-1})+F'(x_{k-1})d_{k-1})-h_{\min})^{\frac{1}{p}}.$$

By Lemma 3.1, it follows that

$$\begin{aligned}\mathrm{dist}(F(x_k),C)&\leq\tfrac{L}{2}\mathrm{dist}^2(0,D(x_{k-1}))+\left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}}\mathrm{dist}^{\frac{2}{p}}(0,D(x_{k-1}))\\&\leq\tfrac{L}{2}\delta^2\left(\tfrac{1}{2}\right)^{2\left(\left(\frac{2}{p}\right)^{k-1}+k-1\right)}+\left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2}{p}}\left(\tfrac{1}{2}\right)^{\frac{2}{p}\left(\left(\frac{2}{p}\right)^{k-1}+k-1\right)}\end{aligned}\tag{3.9}$$

(due to (3.7)). Since

$$2\left(\left(\frac{2}{p}\right)^{k-1}+k-1\right)\geq\frac{2}{p}\left(\left(\frac{2}{p}\right)^{k-1}+k-1\right)\geq\left(\frac{2}{p}\right)^{k}+k-1$$

(noting that $p\in[1,2]$ and $k\geq1$), it follows from (3.9) that

$$\mathrm{dist}(F(x_k),C)\leq\delta\left(L\delta+2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right)\left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^k+k}\leq\frac{\delta}{\beta}\left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^k+k},\tag{3.10}$$

where the last inequality holds by assumption (c). Combining (3.8) and (3.10), one sees that (3.6) holds for $i=k$ and so for each $i=0,1,2,\ldots$. This, together with Lemma 3.1(i) and (2.7), implies that

$$\|d_i\|\leq\mathrm{dist}(0,D(x_i))\leq\beta\mathrm{dist}(F(x_i),C)\leq\delta\left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^i+i}\quad\text{for each } i=0,1,2,\ldots.$$

Thus, $\{x_k\}$ is a Cauchy sequence, and then converges to a point x^*. Clearly, $F(x^*)\in C$ by (3.10), and

$$\|x_k-x^*\|\leq\sum_{i=k}^{+\infty}\|d_i\|\leq2\delta\left(\frac{1}{2}\right)^{\left(\frac{2}{p}\right)^k+k}.$$

This means that $\{x_k\}$ converges to x^* at a rate of $\frac{2}{p}$, and the proof is complete. □

In Theorem 3.2, we have established the local convergence theorem of Algorithm 3.1 under the assumptions of local weak sharp minima of order p and the quasi-regularity condition. For different order p, we make a remark on the assumptions and the local convergence rate as follows.

Remark 3.2.

(a) Theorem 3.2 is also true for Algorithm 3.1 when using the general stepsize sequence $\{v_k\}$, if the assumption (c) of Theorem 3.2 is changed as
  (c) $\beta\left(L\delta+2\left(\frac{1}{2\eta\inf v_k}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right)\leq1.$
  This remark is also valid for Theorems 3.4 and 3.5.
(b) When $p\in[1,2)$, Theorem 3.2 indicates the local superlinear convergence rate of Algorithm 3.1. In the special case when $p=1$, Theorem 3.2 shows the local quadratic convergence rate of Algorithm 3.1, which shares the same convergence rate as that of the GNM; see [10, 21]. The main difference between

convergence analysis of Algorithm 3.1 and that of the GNM stems from their different subproblems. In particular, Li and Wang [21] directly used the minimal property of GNM subproblem to derive the quadratic convergence property of $\|d_k\|$; while our convergence analysis of Algorithm 3.1 utilized Lemma 3.1 and the assumption of weak sharp minima to estimate the convergence rate of $\mathrm{dist}(F(x_k), C)$.

(c) When $p = 2$, Theorem 3.2 exhibits the local linear convergence rate of Algorithm 3.1. Furthermore, the assumption (c) of Theorem 3.2 is reduced to

(c) $\beta\left(L\delta + \left(\frac{2}{\eta v}\right)^{1/2}\right) \leq 1.$

This assumption (c) not only requires δ to be small, but also needs v to be large, which coincides with the property given by Rockafellar [30] that the proximal point algorithm reaches the linear convergence rate if the stepsize stays large enough.

Note that in Theorem 3.2, we do not assume $F(\bar{x}) \in C$; actually, we even do not need to assume the feasibility of inclusion (2.5). In the case when $F(\bar{x}) \in C$, the assumption (b) of Theorem 3.2 automatically holds. Thus, we present the local convergence property of Algorithm 3.1 as follows.

Corollary 3.3. *Let $\bar{x} \in \mathbb{R}^n$ satisfying the inclusion (2.5), and let C be the set of local weak sharp minima of order p for h at $F(\bar{x})$ with the local weak sharp minima constant $\eta_p(h; F(\bar{x}))$. Suppose that $F \in C_L^{1,1}(\mathbf{B}(\bar{x}, r))$ for some $r > 0$, and that $\bar{x}$ is a quasi-regular point of inclusion (2.5) with the quasi-regularity constant $\beta(\bar{x})$. Suppose further that $p \in [1, 2)$ or the stepsize $v > \frac{2\beta(\bar{x})^2}{\eta_p(h;F(\bar{x}))}$ (if $p = 2$). Then, there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that, for any $x_0 \in N(\bar{x})$, the sequence $\{x_k\}$ generated by Algorithm 3.1 with initial point x_0 converges at a rate of $\frac{2}{p}$ to a solution x^* satisfying $F(x^*) \in C$.*

Proof. Let $\bar{\epsilon} \in (0, \eta_p(h; F(\bar{x})))$ be such that

$$v > \frac{2(\beta(\bar{x}) + \bar{\epsilon})^2}{\eta_p(h; F(\bar{x})) - \bar{\epsilon}} \quad \text{if } p = 2. \tag{3.11}$$

Recall from the definition of $\eta_p(h; F(\bar{x}))$ in (2.3) and the definition of $\beta(\bar{x})$ in (2.9), there exists $\bar{\delta} \in (0, r)$, such that C is the set of weak sharp minima of order p for h on $\mathbf{B}(F(\bar{x}), \bar{\delta})$ with modulus $\eta := \eta_p(h; F(\bar{x})) - \bar{\epsilon}$ and that inclusion (2.5) satisfies the quasi-regularity condition on $\mathbf{B}(\bar{x}, \bar{\delta})$ with constant $\beta := \beta(\bar{x}) + \bar{\epsilon}$. We denote by L_0 the Lipschitz constant for F on $\mathbf{B}(\bar{x}, \bar{\delta})$. Set

$$\delta := \begin{cases} \min\left\{\frac{\bar{\delta}}{2}, \frac{2\bar{\delta}}{5L_0}, \frac{1}{2L\beta}, \left(\frac{2\eta v}{(4\beta)^p}\right)^{\frac{1}{2-p}}\right\}, & p \in [1, 2), \\ \min\left\{\frac{\bar{\delta}}{2}, \frac{2\bar{\delta}}{5L_0}, \frac{1}{L\beta}\left(1 - \left(\frac{2\beta^2}{\eta v}\right)^{\frac{1}{2}}\right)\right\}, & p = 2. \end{cases} \tag{3.12}$$

Then, one can directly check that $\delta > 0$ and satisfies the assumptions (a), (b) and (c) of Theorem 3.2. Thus, Theorem 3.2 is applicable and the conclusion follows. □

By the proof of Corollary 3.3 (and that of Theorem 3.2), we further have the following remark, which will be useful in the proof of Theorem 3.4.

Remark 3.3. Suppose that the assumptions of Corollary 3.3 are satisfied. Then, for any $\delta > 0$, there exists $r_\delta \in (0, \delta)$ such that any sequence $\{\tilde{x}_k\}$ generated by Algorithm 3.1 with initial point $\tilde{x}_0 \in \mathbf{B}(\bar{x}, r_\delta)$ satisfies the following property:

$$\|\tilde{x}_k - \bar{x}\| < \delta \text{ for any } k = 0, 1, 2, \ldots. \tag{3.13}$$

REMARK 3.4. As a consequence of Corollary 3.3, we can prove that any sequence $\{x_k\}$ generated by the ProxDescent [18] for solving the convex composite optimization problem (1.1) converges to a global solution of (1.1) at a rate of $\frac{2}{p}$, if there exists a cluster point $\bar{x}$ of $\{x_k\}$ such that C is the set of local weak sharp minima of order p ($1 \leq p < 2$) for h at $F(\bar{x})$ and $\bar{x}$ is a regular point of inclusion (2.5). Indeed, by [18, Theorem 7.4], one sees that $\bar{x}$ is a stationary point of problem (1.1), that is, $0 \in F'(\bar{x})^\top \circ \partial h(F(\bar{x}))$. This implies that $\partial h(F(\bar{x})) \cap \ker F'(\bar{x})^\top \neq \emptyset$. Note by definition that $\partial h(F(\bar{x})) \subseteq (C - F(\bar{x}))^\ominus$. Thus, $\emptyset \neq \partial h(F(\bar{x})) \cap \ker F'(\bar{x})^\top \subseteq (C - F(\bar{x}))^\ominus \cap \ker F'(\bar{x})^\top = \{0\}$; hence, $0 \in \partial h(F(\bar{x}))$ and $F(\bar{x}) \in C$. Then Corollary 3.3 is applicable to concluding that $\{x_k\}$ converges to a global solution of (1.1) at a rate of $\frac{2}{p}$.

3.2. Globalized LPA. By virtue of the backtracking line-search, this subsection is to propose a globalization strategy for the LPA and establish its global convergence theorem. The globalized LPA presented in the following paragraph is in the spirit of the ideas used in [8, 10].

ALGORITHM 3.2. Given constants $c \in (0,1)$ and $\gamma \in (0,1)$, an initial point $x_0 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k, we calculate the search direction $d_k := \mathcal{LP}_{v_k}(x_k)$ by solving the optimization problem (3.1). If $d_k = 0$, then it stops; otherwise, we set $x_{k+1} = x_k + t_k d_k$, where t_k is the maximum value of γ^s for $s = 0, 1, \ldots$, such that

$$h(F(x_k + \gamma^s d_k)) - h(F(x_k)) \leq c\gamma^s \left(h(F(x_k) + F'(x_k)d_k) + \frac{1}{2v_k}\|d_k\|^2 - h(F(x_k)) \right). \tag{3.14}$$

The idea of adopting the backtracking line-search strategy for solving the convex composite optimization problem originated from the works of Burke [8] and Burke and Ferris [10]. The backtracking line-search strategy preserves the descent property of the objective function (cf. (3.14)), which is critical in establishing the global convergence property of Algorithm 3.2 (cf. [8]). Li and Wang [21] also provided the similar globalization strategy for the GNM and proved the global quadratic convergence rate of the globalized GNM.

We now establish in the following theorem a global superlinear convergence result for Algorithm 3.2 under the assumptions of local weak sharp minima of order p and the regularity condition. In particular, if the local weak sharp minima is satisfied, then it indicates the global quadratic convergence rate of Algorithm 3.2, which shares the same convergence rate as that of the globalized GNM, under the same assumptions as in [10].

THEOREM 3.4. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.2 and assume that $\{x_k\}$ has a cluster point $\bar{x}$. Suppose that $1 \leq p < 2$ and that C is the set of local weak sharp minima of order p for h at $F(\bar{x})$. Suppose further that F is of class $C^{1,1}$ near $\bar{x}$, and that $\bar{x}$ is a regular point of inclusion* (2.5). *Then $F(\bar{x}) \in C$, and $\{x_k\}$ converges to $\bar{x}$ at a rate of $\frac{2}{p}$.*

Proof. We first claim that $F(\bar{x}) \in C$. Indeed, the sequence $\{x_k\}$ is also one generated by the descent methods studied in [8] (see (2.1) in [8], with $\{d_k\}$, $h(F(x_k) + F'(x_k)d_k) + \frac{1}{2v}\|d_k\|^2 - h(F(x_k))$ in place of D_k, Δ_k, which satisfy conditions (2.2) in [8]). Thus, [8, Theorems 2.4 and 5.3] can be applied to conclude that $\bar{x}$ is a stationary point of problem (1.1): $0 \in F'(\bar{x})^\top \circ \partial h(F(\bar{x}))$. Similar to the idea in Remark 3.4, we obtain $F(\bar{x}) \in C$, as desired to show.

Next, we show that there exists $\delta > 0$ such that the following implication holds for any k:

$$\|x_k - \bar{x}\| < \delta \Longrightarrow t_k = 1. \tag{3.15}$$

Suppose on the contrary that, there exist a sequence $\{\delta_i\} \subseteq (0,1)$ with $\delta_i \downarrow 0$ and a subsequence $\{k_i\} \subseteq \mathbb{N}$ such that $x_{k_i} \in \mathbf{B}(\bar{x}, \delta_i)$ and $t_{k_i} \neq 1$. Then, $x_{k_i} \to \bar{x}$ and, for each k_i,

$$h(F(x_{k_i}+d_{k_i})) - h(F(x_{k_i})) > c\left(h\left(F(x_{k_i}) + F'(x_{k_i})d_{k_i}\right) + \frac{1}{2v}\|d_{k_i}\|^2 - h(F(x_{k_i}))\right). \tag{3.16}$$

Hence, by the continuity of F and the assumption that $x_{k_i} \to \bar{x}$, it follows that

$$F(x_{k_i}) \to F(\bar{x}) \quad \text{and} \quad \operatorname{dist}(F(x_{k_i}), C) \to 0 \tag{3.17}$$

(as $F(\bar{x}) \in C$ as we showed before). By the assumptions, there exist $\bar{\delta} > 0$, $\eta > 0$ and $\beta > 0$ such that

$$h(z) - h_{\min} \geq \eta \operatorname{dist}^p(z, C) \quad \text{for each } z \in \mathbf{B}(F(\bar{x}), \bar{\delta}), \tag{3.18}$$

$$\operatorname{dist}(0, D(x)) \leq \beta \operatorname{dist}(F(x), C) \quad \text{for each } x \in \mathbf{B}(\bar{x}, \bar{\delta}), \tag{3.19}$$

and

$$\|F'(x) - F'(y)\| \leq L\|x-y\| \quad \text{for each } x, y \in \mathbf{B}(\bar{x}, \bar{\delta}).$$

Combining (3.17) and (3.19), we apply Lemma 3.1(i) to obtain that

$$\operatorname{dist}(0, D(x_{k_i})) \to 0 \quad \text{and} \quad \|d_{k_i}\| \to 0. \tag{3.20}$$

Thus, there exists an integer i_0 such that, for all $i \geq i_0$, the following inequalities hold:

$$\|x_{k_i} - \bar{x}\| < \frac{\bar{\delta}}{2}, \quad \|d_{k_i}\| < \frac{\bar{\delta}}{2}, \tag{3.21}$$

and

$$\|F(x_{k_i}+d_{k_i}) - F(\bar{x})\| < \bar{\delta}, \quad \|F(x_{k_i}) + F'(x_{k_i})d_{k_i} - F(\bar{x})\| < \bar{\delta}. \tag{3.22}$$

Then, it follows from Lemmas 3.1 that

$$h(F(x_{k_i}+d_{k_i})) - h_{\min} \leq h(F(x_{k_i}+d_{k_i})) - h\left(F(x_{k_i}) + F'(x_{k_i})d_{k_i}\right) + \frac{1}{2v}\operatorname{dist}^2(0, D(x_{k_i})). \tag{3.23}$$

Without loss of generality, we assume that h is Lipschitz continuous on $\mathbf{B}(F(\bar{x}), \bar{\delta})$ with Lipschitz constant K (using a smaller $\bar{\delta}$ if necessary). Now let $i \geq i_0$. Then, by (3.22) and (3.21), we conclude from Lemma 2.1 that

$$h(F(x_{k_i}+d_{k_i})) - h\left(F(x_{k_i}) + F'(x_{k_i})d_{k_i}\right) \leq K\|F(x_{k_i}+d_{k_i}) - F(x_{k_i}) - F'(x_{k_i})d_{k_i}\| \leq \frac{KL}{2}\operatorname{dist}^2(0, D(x_{k_i})),$$

and it follows from (3.23) that

$$h(F(x_{k_i}+d_{k_i})) - h_{\min} \leq \tau \operatorname{dist}^2(0, D(x_{k_i})), \tag{3.24}$$

where $\tau := \frac{KL}{2} + \frac{1}{2v} < +\infty$. This, together with (3.16), implies that

$$\begin{aligned} h_{\min} - h(F(x_{k_i})) + \tau \operatorname{dist}^2(0, D(x_{k_i})) \quad &\geq h(F(x_{k_i}+d_{k_i})) - h(F(x_{k_i})) \\ &> c\left(h(F(x_{k_i}) + F'(x_{k_i})d_{k_i}) + \tfrac{1}{2v}\|d_{k_i}\|^2 - h(F(x_{k_i}))\right) \\ &\geq c\left(h_{\min} + \tfrac{1}{2v}\|d_{k_i}\|^2 - h(F(x_{k_i}))\right). \end{aligned}$$

Hence

$$(1-c)\left(h_{\min}-h(F(x_{k_i}))\right)+\tau\mathrm{dist}^2(0,D(x_{k_i}))\geq\frac{c}{2v}\|d_{k_i}\|^2>0,\tag{3.25}$$

(noting that $d_{k_i}\neq 0$ by (3.16)). On the other hand, applying (3.18) and (3.19), we conclude that

$$(1-c)\left(h_{\min}-h(F(x_{k_i}))\right)\leq(c-1)\eta\beta^{-p}\mathrm{dist}^p(0,D(x_{k_i})).$$

Hence it follows from (3.25) that

$$0<(c-1)\eta\beta^{-p}+\tau\mathrm{dist}^{2-p}(0,D(x_{k_i})).$$

Since $\mathrm{dist}(0,D(x_{k_i}))\to 0$ (see (3.20)) and that $p<2$, we arrive by taking the limit at $0<(c-1)\eta\beta^{-p}$, which is clearly a contradiction. Thus, we establish the implication (3.15) for some $\delta>0$.

Finally, we show that $\{x_k\}$ converges to $\bar{x}$ at a rate of $\frac{2}{p}$. Let $\delta>0$ be such that the implication (3.15) holds for any k. Then, by Remark 3.3, there exists $r_\delta\in(0,\delta)$ such that any sequence $\{\tilde{x}_k\}$ generated by Algorithm 3.1 with initial point $\tilde{x}_0\in\mathbf{B}(\bar{x},r_\delta)$ satisfies (3.13). Since $\bar{x}$ is a cluster point of $\{x_k\}$, there exists integer j_0 such that $\|x_{j_0}-\bar{x}\|<r_\delta$. Let $\tilde{x}_0:=x_{j_0}\in\mathbf{B}(\bar{x},r_\delta)$, and let $\{\tilde{x}_k\}$ be generated by Algorithm 3.1 with $\tilde{x}_0$ being the initial point. Then we have that $\|\tilde{x}_k-\bar{x}\|<\delta$ for any $k=0,1,2,\ldots$. By Corollary 3.3, we may assume that $\{\tilde{x}_k\}$ is convergent (using a smaller positive number r_δ if necessary). Moreover, since $\|x_{j_0}-\bar{x}\|<r_\delta\leq\delta$, it follows from (3.15) that $t_{j_0}=1$. This means that $\tilde{x}_1$ and x_{j_0+1} are the same. Hence $\|x_{j_0+1}-\bar{x}\|<\delta$, and we further have that $t_{j_0+1}=1$. Inductively, we conclude that $t_k=1$ for all $k\geq j_0$. Thus $\{x_k\}_{k\geq j_0}$ coincides with $\{\tilde{x}_k\}$ and so is convergent (to $\bar{x}$) at a rate of $\frac{2}{p}$ (as so is $\{\tilde{x}_k\}$ as noted earlier). Therefore the proof is complete. □

3.3. Inexact LPA. In practical terms, it could be computationally very expensive to exactly solve the subproblem (3.1) in each iteration. In this section, we propose an inexact version of the LPA, which is to solve (3.1) only approximately in each iteration (with progressively better accuracy), and investigate its local convergence behavior. Specifically, we present the inexact version of the LPA as follows.

ALGORITHM 3.3. Given constants $M>0$ and $\alpha>2$, initial points $x_0\in\mathbb{R}^n$ and $d_{-1}\in\mathbb{R}^n$, and a sequence of stepsizes $\{v_k\}\subseteq(0,+\infty)$. Having x_k and d_{k-1}, we update

we set $\epsilon_k=M\|d_{k-1}\|^\alpha$ and determine x_{k+1} and d_k as follows. If $\mathcal{LP}_{v_k}(x_k)=0$, then it stops; else if $0\in\mathcal{LP}_{v_k,\epsilon_k}(x_k)$, then we set $d_k=\|d_{k-1}\|^{\alpha-1}d_{k-1}$ and $x_{k+1}=x_k+d_k$; otherwise, we calculate $d_k\in\mathcal{LP}_{v_k,\epsilon_k}(x_k)$ and set $x_{k+1}=x_k+d_k$.

The following theorem provides some sufficient conditions around initial points ensuring the convergence of Algorithm 3.3.

THEOREM 3.5. *Let $\eta>0$, $\beta>0$ and $\bar{\delta}>0$. Let $\bar{x}\in\mathbb{R}^n$ and C be the set of weak sharp minima of order p for h on $\mathbf{B}(F(\bar{x}),\bar{\delta})$ with modulus η. Suppose that $F\in C_L^{1,1}(\mathbf{B}(\bar{x},\bar{\delta}))$, and that inclusion (2.5) satisfies the quasi-regularity condition on $\mathbf{B}(\bar{x},\bar{\delta})$ with constant β. Suppose further that there exists $\delta>0$ such that*

(a) $\delta\leq\min\left\{\frac{\bar{\delta}}{3},\frac{2\bar{\delta}}{7L_0},\frac{1}{2}\left(\frac{1}{32vM}\right)^{\frac{1}{\alpha-2}}\right\}$,
(b) $\mathrm{dist}(F(\bar{x}),C)<\frac{\delta}{2\beta}$,
(c) $\beta\left(L\delta+2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right)\leq\frac{1}{2\sqrt{2}}$,

where L_0 is the Lipschitz constant for F on $\mathbf{B}(\bar{x}, \bar{\delta})$. Then, there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that, for any $x_0 \in N(\bar{x})$, any sequence $\{x_k\}$ generated by Algorithm 3.3 with initial points x_0 and $\|d_{-1}\| \le \left(\frac{\delta^2}{8vM}\right)^{\frac{1}{\alpha}}$, converges at a rate of $q := \min\left\{\frac{\alpha}{2}, \frac{2}{p}\right\}$ to a solution x^* satisfying $F(x^*) \in C$.

Proof. Let $\bar{\beta}$, r_0, and $N(\bar{x})$ be defined respectively as in the beginning of the proof for Theorem 3.2, and let $x_0 \in N(\bar{x})$. Then, as discussed there, we have that

$$\mathrm{dist}(F(x_0), C) \le \frac{\delta}{2\bar{\beta}} \quad \text{and} \quad \mathrm{dist}(0, D(x_0)) \le \frac{\delta}{2}. \tag{3.26}$$

By the assumed quasi-regularity condition, Lemma 3.1 is applicable, and it follows that

$$\|d_0\| \le \left(\mathrm{dist}^2(0, D(x_0)) + 2vM\|d_{-1}\|^{\alpha}\right)^{\frac{1}{2}} \le \frac{\sqrt{2}}{2}\delta. \tag{3.27}$$

We shall show by induction that the following estimates hold for each $i = 0, 1, 2, \ldots$:

$$\|x_i - \bar{x}\| < 3\delta, \quad \mathrm{dist}(F(x_i), C) \le \frac{\delta}{\bar{\beta}} \left(\frac{1}{2}\right)^{q^i + i} \quad \text{and} \quad \|d_i\| \le 2\delta \left(\frac{1}{2}\right)^{q^i + i}. \tag{3.28}$$

Note first that (3.28) holds for $i = 0$ (thanks to the choice of x_0, (3.26) and (3.27)). Next, assume that (3.28) holds for each $i \le k-1$. Then it follows that

$$\|x_k - \bar{x}\| \le \sum_{i=0}^{k-1} \|d_i\| + \|x_0 - \bar{x}\| \le 2\delta \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{q^i + i} + \delta < 3\delta. \tag{3.29}$$

Since $x_{k-1} \in \mathbf{B}(\bar{x}, 3\delta)$ and by the choice of L_0, one has that $\|F(x_{k-1}) - F(\bar{x})\| \le 3L_0\delta$ and $\|F'(x_{k-1})\| \le L_0$ (as $\bar{\delta} \ge 3\delta$). Thus, we have that

$$\|F(x_{k-1}) + F'(x_{k-1})d_{k-1} - F(\bar{x})\| \le \|F(x_{k-1}) - F(\bar{x})\| + L_0\|d_{k-1}\| \le \frac{7}{2}L_0\delta < \bar{\delta}$$

(due to assumption (a)). Hence Lemma 2.4 and Lemma 3.1(ii) are applicable to conclude that

$$\begin{aligned} \mathrm{dist}(F(x_k), C) &\le \tfrac{L}{2}\|d_{k-1}\|^2 + \left(\tfrac{1}{\eta}\right)^{\frac{1}{p}} \left(h(F(x_{k-1}) + F'(x_{k-1})d_{k-1}) - h_{\min}\right)^{\frac{1}{p}} \\ &\le \tfrac{L}{2}\|d_{k-1}\|^2 + \left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}} \left(\mathrm{dist}^2(0, D(x_{k-1})) + 2vM\|d_{k-2}\|^{\alpha}\right)^{\frac{1}{p}}. \end{aligned} \tag{3.30}$$

We now claim that

$$\mathrm{dist}(F(x_k), C) \le \frac{\delta}{\bar{\beta}} \left(\frac{1}{2}\right)^{q^k + k}. \tag{3.31}$$

In fact, if $k = 1$, then, (3.30), together with (3.26), (3.27) and the choice of d_{-1}, implies that

$$\begin{aligned} \mathrm{dist}(F(x_1), C) &\le \tfrac{L}{2}\|d_0\|^2 + \left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}} \left(\mathrm{dist}^2(0, D(x_0)) + 2vM\|d_{-1}\|^{\alpha}\right)^{\frac{1}{p}} \\ &\le \tfrac{L}{2}\left(\tfrac{\sqrt{2}}{2}\delta\right)^2 + \left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}} \left(\left(\tfrac{\delta}{2}\right)^2 + 2vM\tfrac{\delta^2}{8vM}\right)^{\frac{1}{p}} \\ &= \tfrac{1}{4}L\delta^2 + \left(\tfrac{1}{2\eta v}\right)^{\frac{1}{p}} \left(\tfrac{1}{2}\delta^2\right)^{\frac{1}{p}}, \end{aligned}$$

and so (3.31) is established because

$$\frac{1}{4}L\delta^2+\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\left(\frac{1}{2}\delta^2\right)^{\frac{1}{p}}=\frac{\delta}{4}\left(\frac{1}{2}\right)^{\frac{1}{p}-1}\left(\left(\frac{1}{2}\right)^{1-\frac{1}{p}}L\delta+2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right)\le\frac{\delta}{8\beta}\le\frac{\delta}{\beta}\left(\frac{1}{2}\right)^{q+1},$$

where the first inequality is true by assumption (c) and the facts that $\left(\frac{1}{2}\right)^{\frac{1}{p}-1}\in[1,\sqrt{2}]$ (noting $p\in[1,2]$). Now we consider the case when $k\ge 2$. Then, noting the following elementary inequality:

$$(a+b)^r\le a^r+b^r\quad\text{for any } a\ge 0,\ b\ge 0 \text{ and } r\in(0,1], \tag{3.32}$$

one has, from (3.30) and the induction assumption that (3.28) holds for each $i\le k-1$, that

$$\begin{aligned}\operatorname{dist}(F(x_k),C)&\le\frac{L}{2}\|d_{k-1}\|^2+\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\left(\operatorname{dist}^{\frac{2}{p}}(0,D(x_{k-1}))+(2vM)^{\frac{1}{p}}\|d_{k-2}\|^{\frac{\alpha}{p}}\right)\\&\le\frac{L}{2}(2\delta)^2\left(\tfrac{1}{2}\right)^{2(q^{k-1}+k-1)}\\&\quad+\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\left(\delta^{\frac{2}{p}}\left(\tfrac{1}{2}\right)^{\frac{2}{p}(q^{k-1}+k-1)}+(2vM)^{\frac{1}{p}}(2\delta)^{\frac{\alpha}{p}}\left(\tfrac{1}{2}\right)^{\frac{\alpha}{p}(q^{k-2}+k-2)}\right).\end{aligned}\tag{3.33}$$

Noting by $2vM\le\frac{1}{16}(2\delta)^{2-\alpha}$(that is, $\delta\le\frac{1}{2}\left(\frac{1}{32vM}\right)^{\frac{1}{\alpha-2}}$ by assumption (a)), we have that

$$(2vM)^{\frac{1}{p}}(2\delta)^{\frac{\alpha}{p}}\le\left(\frac{1}{16}(2\delta)^{2-\alpha}\right)^{\frac{1}{p}}(2\delta)^{\frac{\alpha}{p}}=\left(\frac{1}{2}\delta\right)^{\frac{2}{p}}, \tag{3.34}$$

and also note that

$$2(q^{k-1}+k-1)\ge q^k+k,\quad \frac{2}{p}(q^{k-1}+k-1)\ge q^k+k-1$$

and

$$\frac{\alpha}{p}(q^{k-2}+k-2)\ge q^k+k-2$$

(as $q=\min\{\frac{\alpha}{2},\frac{2}{p}\}$, $\alpha>2$, $p\in[1,2]$ and $k\ge 2$). It follows from (3.33) that

$$\begin{aligned}\operatorname{dist}(F(x_k),C)&\le\frac{L}{2}(2\delta)^2\left(\tfrac{1}{2}\right)^{q^k+k}+\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\left(\delta^{\frac{2}{p}}\left(\tfrac{1}{2}\right)^{q^k+k-1}+\left(\tfrac{1}{2}\delta\right)^{\frac{2}{p}}\left(\tfrac{1}{2}\right)^{q^k+k-2}\right)\\&=2\delta\left(L\delta+\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\left(\delta^{\frac{2-p}{p}}+\left(\tfrac{1}{2}\right)^{\frac{2}{p}-1}\delta^{\frac{2-p}{p}}\right)\right)\left(\tfrac{1}{2}\right)^{q^k+k}\\&\le 2\delta\left(L\delta+2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\right)\left(\tfrac{1}{2}\right)^{q^k+k}\\&<\frac{\delta}{\beta}\left(\tfrac{1}{2}\right)^{q^k+k},\end{aligned}$$

where the last inequality holds because, by assumption (c), $L\delta+2\left(\frac{1}{2\eta v}\right)^{\frac{1}{p}}\delta^{\frac{2-p}{p}}\le\frac{1}{2\sqrt{2}\beta}<\frac{1}{2\beta}$. Hence (3.31) is established. Thus, by (2.7), we have that

$$\operatorname{dist}(0,D(x_k))\le\beta\operatorname{dist}(F(x_k),C)<\delta\left(\frac{1}{2}\right)^{q^k+k}. \tag{3.35}$$

In view of Algorithm 3.3, if $0 \in \mathcal{LP}_{v,\epsilon_k}(x_k)$, then $d_k = \|d_{k-1}\|^{\alpha-1} d_{k-1}$. This, together with the induction assumption that (3.28) holds for $i = k-1$, implies that

$$\|d_k\| = \|d_{k-1}\|^\alpha \le (2\delta)^\alpha \left(\frac{1}{2}\right)^{\alpha(q^{k-1}+k-1)} < 2\delta \left(\frac{1}{2}\right)^{q^k+k} \tag{3.36}$$

(noting that $\alpha > 2 \ge q$;) otherwise, $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$, and it follows from Lemma 3.1(i) that

$$\|d_k\| \le \left(\mathrm{dist}^2(0, D(x_k)) + 2vM\|d_{k-1}\|^\alpha\right)^{\frac{1}{2}} \le \mathrm{dist}(0, D(x_k)) + (2vM)^{\frac{1}{2}} \|d_{k-1}\|^{\frac{\alpha}{2}}$$

(thanks to (3.32)). Then, by (3.35) and the induction assumption that (3.28) holds for $i = k-1$, it follows that

$$\|d_k\| \le \delta \left(\frac{1}{2}\right)^{q^k+k} + (2vM)^{\frac{1}{2}} (2\delta)^{\frac{\alpha}{2}} \left(\frac{1}{2}\right)^{\frac{\alpha}{2}(q^{k-1}+k-1)}.$$

Since $\frac{\alpha}{2}(q^{k-1}+k-1) \ge q^k + k - 1$ (as $\frac{\alpha}{2} \ge q \ge 1$) and since $(2vM)^{\frac{1}{2}}(2\delta)^{\frac{\alpha}{2}} \le \frac{1}{2}\delta$ by (3.34) (with 2 in place of p), it follows that

$$\|d_k\| \le \delta \left(\frac{1}{2}\right)^{q^k+k} + \frac{\delta}{2}\left(\frac{1}{2}\right)^{q^k+k-1} = 2\delta \left(\frac{1}{2}\right)^{q^k+k}. \tag{3.37}$$

Hence, combining (3.29), (3.31), (3.36) and (3.37), one checks that (3.28) holds for $i = k$ and so for each $i = 0, 1, 2, \ldots$. Consequently, $\{x_k\}$ is a Cauchy sequence, and converges to a point x^*, which, by (3.28), satisfies that $F(x^*) \in C$, and

$$\|x_k - x^*\| \le \sum_{i=k}^{+\infty} \|d_i\| \le 4\delta \left(\frac{1}{2}\right)^{q^k+k}.$$

Therefore, $\{x_k\}$ converges to x^* at a rate of q $\left(= \min\left\{\frac{\alpha}{2}, \frac{2}{p}\right\}\right)$, and the proof is complete. □

Remark 3.5.

(a) Algorithm 3.3 not only has the attractive computational advantage that the subproblems need to be solved only approximately, but also inherits the same convergence rate as that of Algorithm 3.1 if $\alpha \ge 4/p$.
(b) When $p \in [1, 2)$, Theorem 3.5 indicates the local superlinear convergence of Algorithm 3.3. In particular, if $p = 1$ and $\alpha \ge 4$, then it shows the local quadratic convergence rate of Algorithm 3.3, which shares the same convergence rate as that of the inexact GNM [21] under the weaker conditions. While, if $p = 2$, it exhibits the local linear convergence rate of Algorithm 3.3, where the assumption (c) not only requires δ to be small, but also needs v to stay large. This coincides with the property given by Rockafellar [30] that the proximal point algorithm reaches the linear convergence if the stepsize remains large enough.

Similar to the case of Algorithm 3.1, we have the local convergence property for Algorithm 3.3 as follows.

Corollary 3.6. *Let $\bar{x} \in \mathbb{R}^n$ satisfying inclusion* (2.5)*, and let C be the set of local weak sharp minima of order p for h at $F(\bar{x})$ with the local weak sharp minima constant $\eta_p(h; F(\bar{x}))$. Suppose that $F \in C_L^{1,1}(\bar{x}, r)$ for some $r > 0$, and that $\bar{x}$ is a quasi-regular point of inclusion* (2.5) *with the quasi-regularity constant $\beta(\bar{x})$. Suppose further that $p \in [1, 2)$ or the stepsize $v > \frac{16\beta(\bar{x})^2}{\eta_p(h;F(\bar{x}))}$ (if $p = 2$). Then, there exists a neighborhood*

$N(\bar{x})$ of $\bar{x}$ such that, for any $x_0 \in N(\bar{x})$, the sequence $\{x_k\}$ generated by Algorithm 3.3 with initial points x_0 and d_{-1} near 0, converges at a rate of $q := \min\left\{\frac{\alpha}{2}, \frac{2}{p}\right\}$ to a solution x^* satisfying $F(x^*) \in C$.

Proof. Let $\bar{\epsilon} \in (0, \eta_p(h; F(\bar{x})))$ such that

$$v > \frac{16(\beta(\bar{x}) + \bar{\epsilon})^2}{(\eta_p(h; F(\bar{x})) - \bar{\epsilon})} \quad \text{if } p = 2. \tag{3.38}$$

Recall from the definition of $\eta_p(h; F(\bar{x}))$ in (2.3) and the definition of $\beta(\bar{x})$ in (2.9), there exists $\bar{\delta} \in (0, r)$, such that C is the set of weak sharp minima of order p for h on $\mathbf{B}(F(\bar{x}), \bar{\delta})$ with modulus $\eta := \eta_p(h; F(\bar{x})) - \bar{\epsilon}$ and that inclusion (2.5) satisfies the quasi-regularity condition on $\mathbf{B}(\bar{x}, \bar{\delta})$ with constant $\beta := \beta(\bar{x}) + \bar{\epsilon}$. We denote by L_0 the Lipschitz constant for F on $\mathbf{B}(\bar{x}, \bar{\delta})$. Set

$$\delta := \begin{cases} \min\left\{\frac{\bar{\delta}}{2}, \frac{2\bar{\delta}}{7L_0}, \frac{1}{2}\left(\frac{1}{32vM}\right)^{\frac{1}{\alpha-2}}, \frac{1}{4\sqrt{2}L\beta}, \left(\frac{2\eta v}{(8\sqrt{2}\beta)^p}\right)^{\frac{1}{2-p}}\right\}, & p \in [1, 2), \\ \min\left\{\frac{\bar{\delta}}{2}, \frac{2\bar{\delta}}{7L_0}, \frac{1}{2}\left(\frac{1}{32vM}\right)^{\frac{1}{\alpha-2}}, \frac{1}{L\beta}\left(\frac{1}{2\sqrt{2}} - \left(\frac{2\beta^2}{\eta v}\right)^{\frac{1}{2}}\right)\right\}, & p = 2. \end{cases} \tag{3.39}$$

Then $\delta > 0$ and satisfies the assumptions (a), (b), (c) of Theorem 3.5. Thus, Theorem 3.5 is applicable and the conclusion follows. ☐

The proof of Corollary 3.6 (and that of Theorem 3.5) shows actually the following remark, which will be useful in the proof of Theorem 4.3.

REMARK 3.6.

(a) Theorem 3.5 and Corollary 3.6 remain true if Algorithm 3.3 is modified by choosing $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$ in any case (even when $0 \in \mathcal{LP}_{v,\epsilon_k}(x_k)$). Note that adopting $d_k = \|d_{k-1}\|^{\alpha-1} d_{k-1}$ in the case when $0 \in \mathcal{LP}_{v,\epsilon_k}(x_k)$ in Algorithm 3.3 is to avoid solving exactly subproblem (3.1) in the next iteration (otherwise, $d_k = 0$ could be chosen).

(b) Suppose that the assumptions of Corollary 3.6 are satisfied, and let $\{\tilde{x}_k\}$ be the sequence generated by Algorithm 3.3 (or with the modification that $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$ in any cases) with initial points $\tilde{x}_0$ and $\tilde{d}_{-1}$. Then, for any $\delta > 0$ and $M > 0$, there exists $r_\delta \in (0, \delta)$ such that the following property holds:

$$\text{If } \tilde{x}_0 \in \mathbf{B}(\bar{x}, r_\delta) \text{ and } \tilde{d}_{-1} \in \mathbf{B}(0, r_\delta), \text{ then } \|\tilde{x}_k - \bar{x}\| < \delta \text{ for any } k = 0, 1, 2, \ldots. \tag{3.40}$$

4. Application to Feasibility Problem. The feasibility problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences. It consists of finding a point in the intersection of a collection of closed sets (possibly nonconvex); see [1, 17] and references therein. The feasibility problem we consider here is to find a solution of the following system of inequalities:

$$g_i(x) \le 0 \quad \text{for each } i = 1, \ldots, m, \tag{4.1}$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ are all continuously differentiable for $i = 1, \ldots, m$. The solution set of (4.1) is denoted by X^*. The feasibility problem described above can be cast into framework (1.1) as the following two models:

$$\min_{x \in \mathbb{R}^n} h(F(x)), \quad \text{where } F := (g_1, \ldots, g_m)^\top \text{ and } h(\cdot) := \frac{1}{2}\mathrm{dist}^2(\cdot, \mathbb{R}^m_-), \tag{4.2}$$

and

$$\min_{x \in \mathbb{R}^n} h(F(x)), \quad \text{where } F := (g_1, \ldots, g_m)^\top \text{ and } h(\cdot) := \mathrm{dist}(\cdot, \mathbb{R}^m_-), \tag{4.3}$$

where $\mathbb{R}^m_- := \{x = (x_1, \ldots, x_m)^\top : x_i \le 0, i = 1, \ldots, m\}$.

Thus, one can solve the feasibility problem (4.1) naturally by applying the Algorithms 3.1 or 3.3 to the reformulated models (4.2) and/or (4.3). In particular, when applied to the model (4.2), it follows from the first order optimality condition that, for any fixed x, solving the subproblem (3.1) (with h defined in (4.2)) is equivalent to solve the following nonlinear equations

$$F'(x)^\top (F(x) + F'(x)d)_+ + \frac{d}{v} = 0, \tag{4.4}$$

where x_+ denotes the componentwise nonnegative part of x. This motivates us to propose an algorithm for solving the feasibility problem (4.1), which is given in the following Algorithm 4.1. For the sake of simplicity, we introduce, for any $x \in \mathbb{R}^n$, an auxiliary function $H_x : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$H_x(d) := F'(x)^\top (F(x) + F'(x)d)_+ + \frac{d}{v} \quad \text{for each } d \in \mathbb{R}^n. \tag{4.5}$$

Algorithm 4.1. Given constants $M > 0$, $\alpha > 1$, initial points $x_0 \in \mathbb{R}^n$ and $d_{-1} \in \mathbb{R}^n$, and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k and d_{k-1}, we determine x_{k+1} and d_k as follows.

If $H_{x_k}(0) = 0$, then it stops; else if $\|H_{x_k}(0)\| \le M\|d_{k-1}\|^\alpha$, then we set $d_k = \|d_{k-1}\|^{\alpha-1} d_{k-1}$ and $x_{k+1} = x_k + d_k$; otherwise, we solve the nonlinear equations $H_{x_k}(d) = 0$ to obtain d_k such that

$$\|H_{x_k}(d_k)\| \le M\|d_{k-1}\|^\alpha, \tag{4.6}$$

and set $x_{k+1} = x_k + d_k$.

Similarly, applying directly Algorithm 3.3 to (4.3), we present the following algorithm for solving the feasibility problem (4.1).

Algorithm 4.2. Given constants $M > 0$ and $\alpha > 2$, initial points $x_0 \in \mathbb{R}^n$ and $d_{-1} \in \mathbb{R}^n$, and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k and d_{k-1}, we set $\epsilon_k = M\|d_{k-1}\|^\alpha$ and determine x_{k+1} and d_k as follows.

Let $f_k^* := \min_{d \in \mathbb{R}^n} \left\{ \mathrm{dist}(F(x_k) + F'(x_k)d, \mathbb{R}^m_-) + \frac{1}{2v_k}\|d\|^2 \right\}$. If $\mathrm{dist}(F(x_k), \mathbb{R}^m_-) = f_k^*$, then it stops; else if $\mathrm{dist}(F(x_k), \mathbb{R}^m_-) \le f_k^* + \epsilon_k$, then we set $d_k = \|d_{k-1}\|^{\alpha-1} d_{k-1}$ and $x_{k+1} = x_k + d_k$; otherwise, we set d_k to be an ϵ_k-optimal solution of

$$\min_{d \in \mathbb{R}^n} f(x_k; d) := \mathrm{dist}(F(x_k) + F'(x_k)d, \mathbb{R}^m_-) + \frac{1}{2v_k}\|d\|^2,$$

and $x_{k+1} = x_k + d_k$.

To obtain the convergence properties of Algorithms 4.1 and 4.2 by virtue of Corollary 3.6, we provide the following two propositions to show the weak sharp minima property and quasi-regularity condition for models (4.2) and (4.3). The first proposition is trivial by definition, and the second one is a consequence of [19, Proposition 4.1].

Proposition 4.1. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, and let $\bar{x} \in \mathbb{R}^n$ be such that $F(\bar{x}) \in \mathbb{R}^m_-$.*

(i) *Let $h := \frac{1}{2}\mathrm{dist}^2(\cdot, \mathbb{R}^m_-)$. Then $\mathbb{R}^m_-$ is the set of weak sharp minima of order 2 for h at $F(\bar{x})$ with $\eta_2(h; F(\bar{x})) = \frac{1}{2}$.*

(ii) *Let* $h := \mathrm{dist}(\cdot, \mathbb{R}^m_-)$*. Then* $\mathbb{R}^m_-$ *is the set of weak sharp minima for* h *at* $F(\bar{x})$ *with* $\eta_1(h; F(\bar{x})) = 1$*.*

To ensure the quasi-regularity condition of the inclusion $F(x) \in \mathbb{R}^m_-$, we introduce the Robinson constraint qualification at a point $\bar{x}$ satisfying $F(\bar{x}) \in \mathbb{R}^m_-$ (see [29, Definition 2]), that is, it holds that

$$0 \in \mathrm{int}\left\{F(\bar{x}) + \mathrm{im}F'(\bar{x}) + \mathbb{R}^m_+\right\}. \tag{4.7}$$

By [29, Theorem 3] (with $\mathbb{R}^n$ in place of C), one sees that the Robinson constraint qualification (4.7) is equivalent to the following condition

$$\mathrm{im}F'(\bar{x}) + \mathbb{R}^m_+ = \mathbb{R}^m. \tag{4.8}$$

Proposition 4.2. *Let* $F : \mathbb{R}^n \to \mathbb{R}^m$*, and let* $\bar{x} \in \mathbb{R}^n$ *be such that* $F(\bar{x}) \in \mathbb{R}^m_-$*. Suppose that* $F \in C^{1,1}_L(\mathbf{B}(\bar{x}, r))$ *for some* $r > 0$ *and that the Robinson constraint qualification* (4.7) *is satisfied. Let*

$$\bar{\beta} := \sup_{\|y\|\le 1} \inf_{F'(\bar{x})d \in y + \mathbb{R}^m_-} \|d\|. \tag{4.9}$$

Then, $\bar{x}$ *is a quasi-regular point of the inclusion* $F(x) \in \mathbb{R}^m_-$ *with the quasi-regularity constant* $\beta(\bar{x}) \le \bar{\beta}$*.*

Proof. Recall from [19, (2.10)] (with $\mathbb{R}^m_-$ in place of C) that the map $T^{-1}_{\bar{x}}$ and its norm $\|T^{-1}_{\bar{x}}\|$ are defined by

$$T^{-1}_{\bar{x}} y := \{d \in \mathbb{R}^n : F'(\bar{x})d \in y + \mathbb{R}^m_-\} \quad \text{for each } y \in \mathbb{R}^m,$$

and

$$\|T^{-1}_{\bar{x}}\| := \sup\{\inf\{\|d\| : d \in T^{-1}_{\bar{x}}(y)\} : \|y\| \le 1\},$$

respectively. Hence $\|T^{-1}_{\bar{x}}\| = \bar{\beta}$ by definition. Without loss of generality, we may assume that $r \le \frac{1}{L}$. Note that the Robinson constraint qualification (4.7) is equivalent to (4.8); hence $T_{\bar{x}}$ is surjective (and so $\|T^{-1}_{\bar{x}}\| < +\infty$). Then, one concludes from [19, Proposition 4.1(ii)] that the inclusion $F(x) \in \mathbb{R}^m_-$ satisfies the quasi-regularity condition on $\mathbf{B}(\bar{x}, t)$ with constant $\frac{\|T^{-1}_{\bar{x}}\|}{1-Lt}$ for any $t \in (0, r)$. Hence it follows from (2.9) that

$$\beta(\bar{x}) \le \inf_{t>0}\left\{\frac{\|T^{-1}_{\bar{x}}\|}{1 - Lt}\right\} = \bar{\beta},$$

and the proof is complete. □

In the following paragraph we establish the local linear convergence result for Algorithm 4.1 by showing that any generated sequence $\{x_k\}$ of this algorithm is also a sequence generated by Algorithm 3.3, but with the modification that $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$ in any case, with the same initial points and some suitable error controls $\{\epsilon_k\}$ for problem (4.2). Recall that $\bar{\beta}$ is defined by (4.9).

Theorem 4.3. *Let* $\bar{x} \in X^*$*. Suppose that* $F \in C^{1,1}_L(\mathbf{B}(\bar{x}, r))$ *for some* $r > 0$*,* $\mathrm{im}F'(\bar{x}) + \mathbb{R}^m_+ = \mathbb{R}^m$*, and that the stepsize* $v > 32\bar{\beta}^2$*. Then, there exists a neighborhood* $N(\bar{x})$ *of* $\bar{x}$ *such that, for any* $x_0 \in N(\bar{x})$*, any sequence* $\{x_k\}$ *generated by Algorithm 4.1 with initial points* x_0 *and* d_{-1} *near 0, linearly converges to a solution* $x^* \in X^*$*.*

Proof. Let $\delta := r$, $\tilde{\alpha} := 2\alpha$ and $\tilde{M} := M + L^2 + \frac{1}{v}$. By Remark 3.6(b), there exists $r_\delta \in (0, r)$ such that (3.40) holds with $\{r, \tilde{\alpha}, \tilde{M}\}$ in place of $\{\delta, \alpha, M\}$. Let $x_0 \in \mathbf{B}(\bar{x}, r_\delta)$ and $d_{-1} \in \mathbf{B}(0, r_\delta)$ be initial points, and let $\{x_k\}$ and $\{d_k\}$ be the sequences generated by Algorithm 4.1. Set $\epsilon_k := v\tilde{M}^2\|d_{k-1}\|^{2\alpha}$ for each $k = 0, 1, \ldots$, and let h be defined by (4.2). Fix $k \in \mathbb{N}$. We first show the following implication:

$$x_k \in \mathbf{B}(\bar{x}, r) \Longrightarrow d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k). \tag{4.10}$$

To do this, we assume that $x_k \in \mathbf{B}(\bar{x}, r)$. Without loss of generality, we may assume that $H_{x_k}(0) \neq 0$ (otherwise, $\mathcal{LP}_v(x_k) = 0$ and that $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$ is clear). We now claim that

$$\|H_{x_k}(d_k)\| \leq \tilde{M}\|d_{k-1}\|^{\alpha}. \tag{4.11}$$

In view of Algorithm 4.1, we only need to consider that case when $\|H_{x_k}(0)\| \leq M\|d_{k-1}\|^{\alpha}$ (since (4.11) automatically holds otherwise). By (4.5), one has that

$$\|H_{x_k}(d_k) - H_{x_k}(0)\| \leq \|F'(x_k)^\top\|\|(F(x_k) + F'(x_k)d_k)_+ - (F(x_k))_+\| + \frac{1}{v}\|d_k\|. \tag{4.12}$$

Note that y_+ is the projection of y onto $\mathbb{R}^m_+$. Then it follows from [15, Chapter A, (3.1.6)] that $\|(F(x_k) + F'(x_k)d_k)_+ - (F(x_k))_+\| \leq \|F'(x_k)d_k\|$. As $F \in C^{1,1}_L(\mathbf{B}(\bar{x}, r))$ and $x_k \in \mathbf{B}(\bar{x}, r)$, it follows from (4.12) that

$$\|H_{x_k}(d_k) - H_{x_k}(0)\| \leq \|F'(x_k)\|^2\|d_k\| + \frac{1}{v}\|d_k\| \leq \left(L^2 + \frac{1}{v}\right)\|d_{k-1}\|^{\alpha}$$

(due to $d_k = \|d_{k-1}\|^{\alpha-1}d_{k-1}$), and thus,

$$\|H_{x_k}(d_k)\| \leq \|H_{x_k}(0)\| + \|H_{x_k}(d_k) - H_{x_k}(0)\| \leq \tilde{M}\|d_{k-1}\|^{\alpha}.$$

Hence (4.11) is verified. To proceed, we define, for any $x \in \mathbb{R}^n$, the function $\phi_x : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\varphi_x(d) := \tfrac{1}{2}\mathrm{dist}^2(F(x) + F'(x)d, \mathbb{R}^m_-) + \tfrac{1}{2v}\|d\|^2 \quad \text{for each } d \in \mathbb{R}^n.$$

Note that φ_x (for fixed $x \in \mathbb{R}^n$) is a convex function and that $H_x(\cdot)$ defined in (4.5) is its gradient. Hence one has that

$$\varphi_{x_k}(d) \geq \varphi_{x_k}(d_k) + \langle H_{x_k}(d_k), d - d_k\rangle \quad \text{for any } d \in \mathbb{R}^n.$$

In particular, letting $d_k^* := \mathcal{LP}_v(x_k)$ (and so $\varphi_{x_k}(d_k^*) = \min_{d\in\mathbb{R}^n}\varphi_{x_k}(d)$), one concludes that

$$\varphi_{x_k}(d_k) \leq \varphi_{x_k}(d_k^*) - \langle H_{x_k}(d_k), d_k^* - d_k\rangle \leq \varphi_{x_k}(d_k^*) + \|H_{x_k}(d_k)\|\|d_k^* - d_k\|. \tag{4.13}$$

Moreover, by [30, Proposition 3], we have that $\|d_k^* - d_k\| \leq v\|H_{x_k}(d_k)\|$. This, together with (4.13) and (4.11), implies that

$$\varphi_{x_k}(d_k) \leq \varphi_{x_k}(d_k^*) + v\|H_{x_k}(d_k)\|^2 \leq \varphi_{x_k}(d_k^*) + v\tilde{M}^2\|d_{k-1}\|^{2\alpha} = \varphi_{x_k}(d_k^*) + \epsilon_k.$$

This means that $d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k)$. Thus, implication (4.10) is checked. Next we further verify that

$$d_k \in \mathcal{LP}_{v,\epsilon_k}(x_k) \quad \text{for each } k \in \mathbb{N}. \tag{4.14}$$

Granting this and noting by Propositions 4.1(i) and 4.2 that both the weak sharp minima assumption ($p = 2$) for h and the quasi-regularity assumption for the inclusion $F(x) \in \mathbb{R}^m_-$ are satisfied, one sees that Corollary 3.6 (and Remark 3.6(a)) is applicable; hence the conclusion follows.

To show (4.14), note first that $x_0 \in \mathbf{B}(\bar{x}, r_\delta) \subseteq \mathbf{B}(\bar{x}, r)$ and so $d_0 \in \mathcal{LP}_{v,\epsilon_0}(x_0)$ by implication (4.10). We next assume that $d_i \in \mathcal{LP}_{v,\epsilon_i}(x_i)$ for any $i \leq k$. Then, by (3.40) (with $\{r, \tilde{\alpha}, \tilde{M}\}$ in place of $\{\delta, \alpha, M\}$), we have that $x_{k+1} \in \mathbf{B}(\bar{x}, r)$ ans $d_{k+1} \in \mathcal{LP}_{v,\epsilon_{k+1}}(x_{k+1})$ by implication (4.10). Thus (4.14) is seen to hold by mathematical induction; hence the proof is complete. ∎

For Algorithm 4.2, we have the following local quadratic convergence result.

THEOREM 4.4. *Let $\bar{x} \in X^*$. Suppose that $F \in C_L^{1,1}(\mathbf{B}(\bar{x}, r))$ for some $r > 0$ and that $\mathrm{im}F'(\bar{x}) + \mathbb{R}^m_+ = \mathbb{R}^m$. Then, there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that, for any $x_0 \in N(\bar{x})$, any sequence $\{x_k\}$ generated by Algorithm 4.2 with initial points x_0 and d_{-1} near 0, quadratically converges to some $x^* \in X^*$.*

Proof. Note that the Algorithm 4.2 is a direction applying of Algorithm 3.3 to problem (4.3). Propositions 4.1(i) and 4.2 say that both the weak sharp minima assumption ($p = 1$) for h and the quasi-regularity assumption for the inclusion $F(x) \in \mathbb{R}^m_-$ are satisfied. Hence Corollary 3.6 is applicable, and the conclusion follows. ∎

For the subproblem of solving each nonlinear equations $H_{x_k}(d) = 0$ in Algorithm 4.1, there are many efficient methods such as the Newton-type methods and the trust region methods; see the monograph [26] for more details. Note that the function H_x (for fixed $x \in \mathbb{R}^n$) is p-order semismooth* everywhere for any $p > 0$ (which could be verified by definition). Recall from [28] that the semismooth Newton method for p-order semismooth functions converges locally at a rate of $1 + p$. This means that the semismooth Newton method is highly efficient in solving each nonlinear equations $H_{x_k}(d) = 0$ (indeed, one iteration is enough in most cases for our application in the sensor network localization problem below). This motivates us to present the following algorithm based on one semismooth Newton iteration for solving each nonlinear equations $H_{x_k}(d) = 0$.

ALGORITHM 4.3. Given initial points $x_0 \in \mathbb{R}^n$, $d_{-1} \in \mathbb{R}^n$, and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k and d_{k-1}, we calculate the search direction d_k by

$$V = F'(x_k)^\top \operatorname{diag}(\operatorname{sgn}(F'(x_k)d + F(x_k))_+)F'(x_k) + \frac{1}{v_k} I_n \quad \text{and} \quad d_k = d_{k-1} - V^{-1} H_{x_k}(d_{k-1}),$$

where $\operatorname{sgn}(\cdot)$ denotes the sign function, and set $x_{k+1} = x_k + d_k$.

Before conducting the numerical experiments, we make a remark on the comparison of the proposed algorithms for the feasibility problem.

REMARK 4.1.

(a) As showed by Theorems 4.4 and 4.3, Algorithm 4.2 achieves a local quadratic convergence rate while Algorithm 4.1 concludes a local linear convergence rate. However, their numerical efficiency depends on the costs of solving the corresponding subproblems. As illustrated in numerical experiments below, Algorithm 4.1 is more efficient and costs less CPUtime than Algorithm 4.2. This is because that the semismooth Newton method is highly efficient in solving each subproblem in Algorithm 4.1; while each

*A locally Lipschitzian function $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is said to be p-order semismooth ($p > 0$) at x if

$$\limsup_{h \to 0} \sup_{V \in \partial\phi(x+h)} \frac{Vh - \phi'(x; h)}{\|h\|^{1+p}} \text{ is bounded,}$$

where $\phi'(x; h)$ denotes the directional derivative of ϕ at x along h, and $\partial\phi(x)$ is the generalized Jacobian of ϕ at x (see [28]). An important family of the p-order semi-smooth functions are the semi-algebraic Lipschitz functions, which covers most of applications; see, e.g., [5].

subproblem in Algorithm 4.2 is a (large scale) nonsmooth convex optimization problem, and it usually takes much more time to solve this subproblem by using any popular algorithm such as the primal-dual interior point method [41] or the alternating direction method [14].

(b) Although we cannot provide the proof of the linear convergence rate of Algorithm 4.3, our numerical experiments below illustrate that it shares the same stability and linear convergence rate as Algorithm 4.1, and costs less CPUtime.

The rest of this section is devoted to demonstrate the performance of the LPA type algorithms on the sensor network localization problem, arising from the area of wireless sensor networks.

Typical wireless sensor networks consist of a large number of inexpensive wireless sensors deployed in a geographical area with the ability to communicate with their neighbors within a limited radio range. The sensor network localization problem is to estimate the positions of the sensors in a network by using the given incomplete pairwise distance measurements; see [4, 22, 38] and references therein. Formally, let $V_s = \{s_1, \ldots, s_n\} \subset \mathbb{R}^2$ and $V_a = \{a_{n+1}, \ldots, a_{n+m}\} \subset \mathbb{R}^2$ be the sets of sensors and anchors (a small quantity of sensors whose positions are known), respectively. For each pair of (sensor, sensor) or (sensor, anchor), if their distance is within the radio range (denoted by R), then they can detect each other and this edge is recorded in the set E_{ss} (the set of sensor-sensor edges) or E_{sa} (the set of sensor-anchor edges). That is, E_{ss} and E_{sa} denote the sets of sensor-sensor edges and sensor-anchor edges, whose length d_{ij} is less or equal to the radio range R, respectively. Thus, the sensor network localization problem can be cast into the feasibility problem of finding n locations $x_i \in \mathbb{R}^2$ $(i = 1, \ldots, n)$ such that

$$\begin{array}{ll} \|x_i - x_j\|^2 = d_{ij}^2, & (i,j) \in E_{ss}, \\ \|x_i - x_j\|^2 > R^2, & (i,j) \notin E_{ss}, \\ \|x_i - a_j\|^2 = \bar{d}_{ij}^2, & (i,j) \in E_{sa}, \\ \|x_i - a_j\|^2 > R^2, & (i,j) \notin E_{sa}, \end{array} \tag{4.15}$$

see [34]. In general, the problem (4.15) is difficult to solve (indeed, it is NP-hard; see [33]), as the quadratic constraints in it are nonconvex. As shown in (4.2) and (4.3), we can reformulate the feasibility problem (4.15) as a convex composite optimization problem (a small perturbation on the strict inequalities of (4.15) is required to maintain the closeness of the feasibility system), where the inner functions in (4.2) and (4.3) are given by

$$\begin{array}{llll} g_{i,j,1}(x) = \|x_i - x_j\|^2 - d_{ij}^2 & \text{and} & g_{i,j,2}(x) = d_{ij}^2 - \|x_i - x_j\|^2, & (i,j) \in E_{ss}, \\ g_{i,j,0}(x) = R^2 - \|x_i - x_j\|^2, & & & (i,j) \notin E_{ss}, \\ \bar{g}_{i,j,1}(x) = \|x_i - a_j\|^2 - \bar{d}_{ij}^2 & \text{and} & \bar{g}_{i,j,2}(x) = \bar{d}_{ij}^2 - \|x_i - a_j\|^2, & (i,j) \in E_{sa}, \\ \bar{g}_{i,j,0}(x) = R^2 - |x_i - a_j\|^2, & & & (i,j) \notin E_{sa}. \end{array}$$

Here, we will apply Algorithms 4.1, 4.2 and 4.3 to solve it in the numerical experiments.

Many works concentrate on the following relaxation model, neglecting all inequality constraints in (4.15),

$$\begin{array}{ll} \|x_i - x_j\|^2 = d_{ij}^2, & (i,j) \in E_{ss}, \\ \|x_i - a_j\|^2 = \bar{d}_{ij}^2, & (i,j) \in E_{sa}, \end{array} \tag{4.16}$$

see [4, 22] and references therein. We also apply Algorithms 4.1 and 4.2 to solve the relaxed problem (4.16).

One of the most popular and practical tools for solving the sensor network localization problem is the semidefinite relaxation (SDR) technique, which further relaxes (4.16) into a semidefinite programming; see, e.g., [4, 22]. We choose the MATLAB software† in [4] as the representative of the semidefinite relaxation technique. Furthermore, the proximal bundle method [32] is an implementable algorithm for solving the convex composite optimization problems, which is also compared in the numerical experiments. Note that a quadratic subproblem is solved to find the search direction in each iteration of the proximal bundle method, here we employ the CVX‡ to solve such subproblems.

All numerical experiments are implemented in MATLAB R2013b and executed on a personal desktop (Intel Core Duo E8500, 3.16 GHz, 4.00 GB of RAM). In the numerical experiments, the sensors and anchors are randomly placed in the unit square $[-0.5, 0.5]^2$:

$$V_s = -0.5 + 0.5 * rand(2, n) \quad \text{and} \quad V_a = -0.5 + 0.5 * rand(2, m).$$

The key criterion to characterize the performance of executed algorithms is the accuracy of the estimation $\{x_1, \ldots, x_n\}$, measured by the root mean square distance (RMSD):

$$\text{RMSD} = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \|s_i - x_i\|^2 \right)^{\frac{1}{2}}.$$

In order to facilitate the reading of the numerical results, we list the abbreviations of the algorithms for the sensor network localization problem in Table 4.1.

TABLE 4.1
List of the algorithms for solving the sensor network localization problem.

Abbreviations	Algorithms
SDR	**S**emi**D**efinite **R**elaxation method in [4], which is to solve the relaxed problem (4.16).
LPA-I	Algorithm 4.1 for solving the problem (4.15).
LPA-II	Algorithm 4.2 for solving the problem (4.15).
LPA-SN	Algorithm 4.3 for solving the problem (4.15).
LPA-I-R	Algorithm 4.1 for solving the **R**elaxed problem (4.16).
LPA-II-R	Algorithm 4.2 for solving the **R**elaxed problem (4.16).
CPB	The **C**omposite **P**roximal **B**undle method for solving the problem (4.15).
CPB-R	The **C**omposite **P**roximal **B**undle method for solving the **R**elaxed problem (4.16).

When implementing the LPA type algorithms, we set $M = 1$, $\alpha = 2$, $d_{-1} = rand(2, n)$, the constant stepsize $v = 100$ (unless otherwise specified), the stopping criterion of inner iteration (except the LPA-SN) as $H_{x_k}(d) < \max\{\|d_{k-1}\|^3, 1e-6\}$ or the number of iterations is greater than 50, and the stopping criterion of the LPA type algorithms as RMSD $< 1e - 10$ or the number of outer iterations is greater than 100. The initial starting point for The LPA-I and LPA-SN is chosen randomly, that for the LPA-II and LPA-II-R, CPB and CPB-R is set as $sensor + 0.2 * randn(2, n)$ and that for the LPA-I-R is set as $sensor + 0.5 * randn(2, n)$. Observing in the extensive simulations of the LPA type algorithms, we find that the number of semismooth Newton iterations is frequently 1 and occasionally 2 or 3 in the first 10 outer iterations, and always 1 in the rest of iterations. Hence, it is indicative that the semismooth Newton method is highly efficient in solving the subproblem (4.4) of Algorithm 4.1. On the other hand, solving the subproblem of Algorithm 4.2 seems

†The code and description are available in http://www.math.nus.edu.sg/∼mattohkc/SNLSDP.html.

‡CVX, designed by Michael Grant and Stephen Boyd, is a MATLAB-based modeling system for convex optimization. Detailed information is available at the website http://cvxr.com/cvx/.

much harder than the semismooth Newton method for the subproblem of Algorithm 4.1, which is consistent with Remark 4.1(a).

We first demonstrate the performance of the SDR, the LPA type and the CPB type algorithms on a randomly generated network of 100 sensors, 10 anchors and the radio range being 0.3. All the algorithms listed in Table 4.1 are tested in this experiment. The realization of the LPA-I is presented in Figure 4.1, where the anchors are denoted by diamonds, the true sensors are denoted by circles and their estimates by asterisks. The performance of all the algorithms are listed in Table 4.2. Three observations are indicated by Table 4.2: (i) The SDR, LPA-I, LPA-SN and LPA-I-R (based on model (4.2)) can achieve the estimation in a few seconds, while the LPA-II and LPA-II-R (based on model (4.3)), CPB and CPB-R are not suitable for the large scale sensor network localization problem, since they take too much time in solving the subproblems. (ii) We find that the performance of the LPA-I and LPA-SN do not depend on the choice of initial starting points. Thus we believe that the LPA-I and LPA-SN can converge globally, even though this property cannot be proved for the moment. The LPA-II, LPA-I-R and LPA-II-R converge locally, as shown in Theorems 4.3 and 4.4. The choices of the initial starting points also indicate that the LPA-I-R allows a larger region of the initial points than the LPA-II and LPA-II-R. (iii) The LPA-I, LPA-SN and LPA-I-R achieve a more precise solution and take less CPUtime than the SDR and CPB type algorithms do, since the LPA-I, LPA-SN and LPA-I-R converge fast and do not need any software package. Moreover, the LPA-I and LPA-SN consume more CPUtime than the LPA-I-R, because the LPA-I and LPA-SN are designed to solve the full version of feasibility problem (4.15), whose the number of constraints is about the triple of that of the relaxation problem (4.16), solved by LPA-I-R.

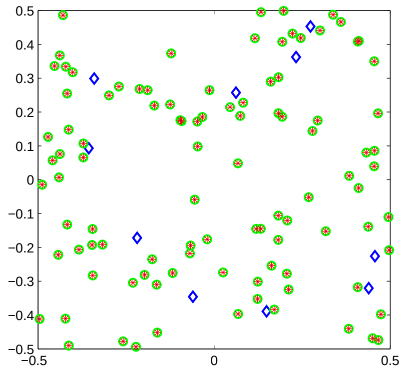


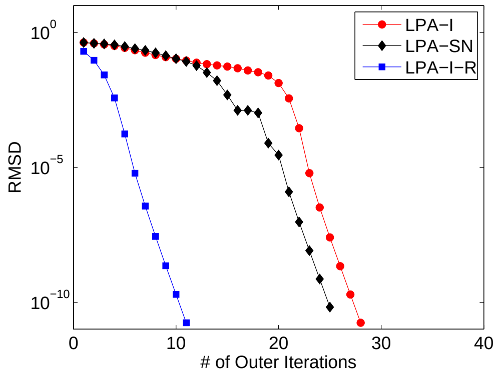
Fig. 4.1. *The LPA-I can successfully localize the positions of sensors in a wireless sensor network (100 sensors, 10 anchors, and radio range = 0.3), where the RMSD is* 5.3e-11 *and the CPUtime is* 5.8 *seconds.*

Table 4.2
The performance of SDR, the LPA type and the CPB type algorithms for a sensor network localization problem (100 sensors, 10 anchors, and radio range = 0.3).

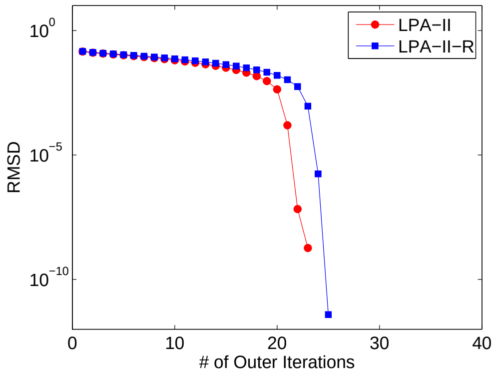
	SDR	LPA-I	LPA-SN	LPA-II	LPA-I-R	LPA-II-R	CPB	CPB-R
RMSD	1.9e-5	5.3e-11	4.5e-11	1.8e-10	6.1e-11	3.8e-15	1.3e-4	3.4e-4
CPUtime (seconds)	7.9	5.8	4.1	765	0.9	22	98	39

We also verify the local convergence rate of the LPA type algorithms by conducting extensive simulations. Figure 4.2 plots the RMSD of the estimation along the number of the outer iterations in a random trial.

Figure 4.2(a) illustrates the local linear convergence rate of the LPA-I, LPA-SN and LPA-I-R (based on model (4.2)), which is consistent with the theoretical analysis in Theorem 4.3. Figure 4.2(b) demonstrates the local quadratic convergence rate of the LPA-II and LPA-II-R (based on model (4.3)), which is consistent with the theoretical analysis in Theorem 4.4.



(a) The linear convergence rate.

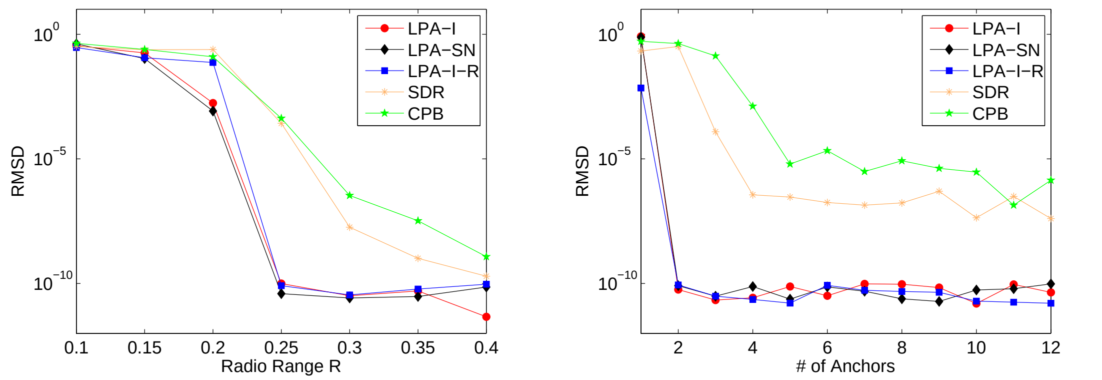


(b) The quadratic convergence rate.

FIG. 4.2. *The local convergence rate of the LPA type algorithms.*

The third experiment is performed to study the variation of RMSD when varying the circumstances (the radio range and the number of anchors) of the wireless sensor network of 100 of them. Figure 4.3(a) shows the variation of RMSD by increasing the radio range from 0.1 to 0.4 for the LPA-I, LPA-SN, LPA-I-R, SDR and CPB. When the radio range R is too low, there is no enough information between the sensors or anchors for the estimation to be effective. The accuracy is improved (the RMSD decreases) consistently for all algorithms as the radio range is increased. It is also illustrated that the LPA-I and LPA-SN can obtain more accurate estimation by using less information between the sensors or anchors (only need $R \geq 0.2$). Figure 4.3(b) illustrates the variation of RMSD by varying the number of anchors from 1 to 12. When the number of anchors is too small, the estimation fails since the information revealed in the network is not enough. The accuracy is enhanced consistently for all algorithms as the number of anchors increases. The perfect estimation is realized by SDR and CPB when the number of anchors is greater than 4 and 5, respectively, while the LPA-I, LPA-SN and LPA-I-R only require 2 anchors. This experiment indicates that the LPA type algorithms can achieve the perfect estimation (in higher precision) by using less information (the small radio range and the few anchors) than that of SDR.

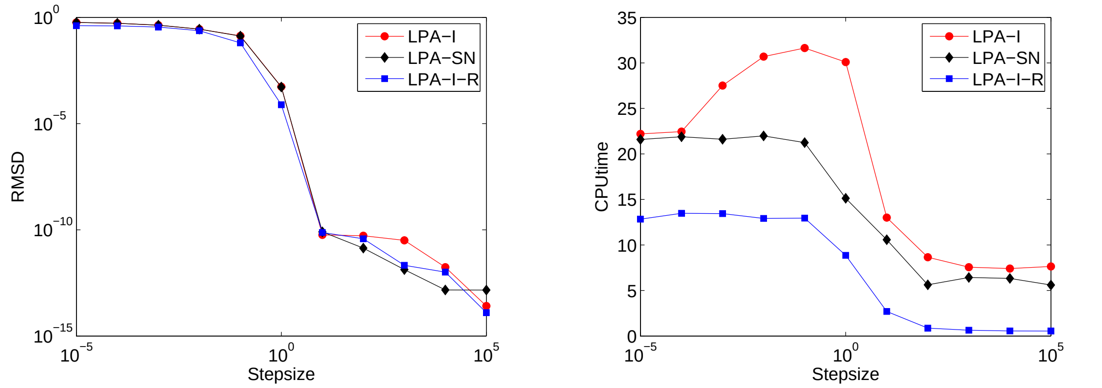
We finally demonstrate the effect of the stepsize on the LPA type algorithms to localize a wireless sensor network of 100 sensors and 10 anchors and the radio range being 0.3. Figure 4.4 shows the variation of RMSD and CPUtime when varying the stepsize from 10^{-5} to 10^5. As shown in Figure 4.4(a), the accuracy of the estimation is improved consistently for the LPA-I, LPA-SN and LPA-I-R as the stepsize increases. This is because the stepsize is indeed a proximal parameter and this numerical result is consistent with the theory on the proximal point algorithm in [30] (also Remarks 3.2(ii) and 3.5(ii)). Thus we conclude that the larger the proximal parameter, the better the performance. We further find that the perfect estimation requires the stepsize to be as large as $v \geq 10$. It is illustrated in Figure 4.4(b) that the CPUtime drops when $v = 10$ or 100 and decreases little when $v > 100$. Thus, in all experiments, we set the stepsize $v = 100$ as default; see the paragraph below Table 4.1.



(a) Varying the radio range of a network of 100 sensors and 10 anchors.

(b) Varying the number of anchors of a network of 100 sensors and $R = 0.3$.

Fig. 4.3. *Variation of RMSD when varying the circumstances of network.*



(a) RMSD

(b) CPUtime

Fig. 4.4. *Variation of RMSD and CPUtime when varying the stepsize.*

The conclusions of the numerical experiments can be summarized as follows. (i) The LPA-I, LPA-SN and LPA-I-R (based on model (4.2)) can achieve a more precise solution, take less CPUtime and require less information (the small radio range and the few anchors) than the SDR does. The LPA-II and LPA-II-R (based on model (4.3)) are not suitable for the large scale sensor network localization problem. (ii) The LPA-I and LPA-SN globally converge to the true sensors, while the LPA-II, LPA-I-R and LPA-II-R only locally converge. (iii) For the LPA-I, LPA-SN and LPA-I-R, the larger the stepsize, the more precise the estimation and the less the CPUtime. Further from the extensive simulations, we find that the LPA type algorithms are a little less robust than the SDR. In particular, the estimation is regarded as "successful" if the estimated RMSD is less than $1e-3$. Thus, the successful estimation rate of the SDR is about 96%, while the LPA-I and LPA-SN can only successfully localize 93% wireless sensor networks.

Acknowledgment. The authors are grateful to two anonymous reviewers for their valuable suggestions and remarks which helped to improve the quality of the paper.

REFERENCES

[1] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[3] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Cambridge, 1999.

[4] P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, and T.-C. Wang, *Semidefinite programming approaches for sensor network localization with noisy distance measurements*, IEEE Trans. Autom. Sci. Eng., 3 (2006), pp. 360–371.

[5] J. Bolte, A. Daniilidis, and A. S. Lewis, *Tame functions are semismooth*, Math. Program., 1-2 (2009), pp. 5–19.

[6] J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Math. Oper. Res., 20 (1995), pp. 801–817.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–124.

[8] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Math. Program., 33 (1985), pp. 260–279.

[9] J. V. Burke, *Second order necessary and sufficient conditions for convex composite NDO*, Math. Program., 38 (1987), pp. 287–302.

[10] J. V. Burke and M. C. Ferris, *A Gauss-Newton method for convex composite optimization*, Math. Program., 71 (1995), pp. 179–194.

[11] J. V. Burke and S. Deng, *Weak sharp minima revisited, part I: Basic theory*, Control Cybern., 31 (2002), pp. 439–469.

[12] J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359.

[13] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.

[14] B. He and X. Yuan, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[15] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer-Verlag, Berlin, 2001.

[16] Y. H. Hu, X. Q. Yang, and C.-K. Sim, *Inexact subgradient methods for quasi-convex optimization problems*, Eur. J. Oper. Res., 240 (2015), pp. 315–327.

[17] A. S. Lewis and J. Malick, *Alternating projections on manifolds*, Math. Oper. Res., 33 (2008), pp. 216–234.

[18] A. S. Lewis and S. J. Wright, *A proximal method for composite minimization*, tech. report, 2008.

[19] C. Li and K. F. Ng, *Convergence analysis of the Gauss-Newton method for convex inclusion and convex-composite optimization problems*, J. Math. Anal. Appl., 389 (2012), pp. 469–485.

[20] 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.

[21] C. Li and X. Wang, *On convergence of the Gauss-Newton method for convex composite optimization*, Math. Program., 91 (2002), pp. 349–356.

[22] Z.-Q. Luo, W.-K. Ma, A.M.-C. So, Y. Ye, and S. Zhang, *Semidefinite relaxation of quadratic optimization problems*, IEEE Signal Proc. Mag., 27 (2010), pp. 20–34.

[23] B. Martinet, *Regularisation d'indquations variationelles par approximations successives*, Rev. Francaise Inf. Rech. Oper., (1970), pp. 154–159.

[24] J. J. Moré, *The Levenberg-Marquardt algorithm: Implementation and theory*, in Numerical Analysis, G.A. Watson, ed., vol. 630 of Lecture Notes in Mathematics, Springer, Berlin, 1978, pp. 105–116.

[25] Yu. Nesterov, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161.

[26] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Science+Business Media, New York, 2006.

[27] J.-S. Pang, *Error bounds in mathematical programming*, Math. Program., 79 (1997), pp. 299–332.

[28] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[29] S. M. Robinson, *Stability theory for systems of inequalities, part II. Differentiable nonlinear systems.*, SIAM J. Numer. Anal., 13(4) (1976), pp. 497–513.

[30] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[31] R. T. Rockafellar, *First- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307 (1988), pp. 75–108.

[32] C. Sagastizábal, *Composite proximal bundle method*, Math. Program., 140 (2013), pp. 189–233.

[33] J. B. Saxe, *Embeddability of weighted graphs in k-space is strongly NP-hard*, in Proc. 17th Annual Allerton Conf. Communication, Control, and Computing, 1979, pp. 480–489.

[34] Q. Shi, C. He, and L. Jiang, *Normalized incremental subgradient algorithm and its application*, IEEE Trans. Signal Proces., 57 (2009), pp. 3759–3774.

[35] M. Studniarski and D. E. Ward, *Weak sharp minima: Characterizations and sufficient conditions*, SIAM J. Control Optim., 38 (1999), pp. 219–236.

[36] W. Sun and Y.-X. Yuan, *Optimization Theory and Methods*, Springer Science+Business Media, New York, 2006.

[37] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, tech. report, 2008.

[38] Z. Wang, S. Zheng, Y. Ye, and S. Boyd, *Further relaxations of the semidefinite programming approach to sensor network localization*, SIAM J. Optim., 19 (2008), pp. 655–673.

[39] R. S. Womersley, *Local properties of algorithms for minimizing nonsmooth composite functions*, Math. Program., 32 (1985), pp. 69–89.

[40] L. Xiao and T. Zhang, *A proximal-gradient homotopy method for the sparse least-squares problem*, SIAM J. Optim., 23 (2013), pp. 1062–1091.

[41] Y. Ye, *Interior Point Algorithms: Theory and Analysis*, John Wiley & Sons, New York, 1987.

[42] Y.-X. Yuan, *Conditions for convergence of trust region algorithms for nonsmooth optimization*, Math. Program., 31 (1985), pp. 220–228.

[43] X. Y. Zheng and K. F. Ng, *Strong KKT conditions and weak sharp solutions in convex-composite optimization*, Math. Program., 126 (2011), pp. 259–279.

[44] X. Y. Zheng and X. Q. Yang, *Weak sharp minima for semi-infinite optimization problems with applications*, SIAM J. Optim., 18 (2007), pp. 573–588.