AN ACTIVE SET METHOD FOR MATHEMATICAL PROGRAMS WITH LINEAR COMPLEMENTARITY CONSTRAINTS

LIFENG CHEN* AND DONALD GOLDFARB†

April, 2007

Abstract. We study mathematical programs with linear complementarity constraints (MPLCC) for which the objective function is smooth. Current nonlinear programming (NLP) based algorithms including regularization methods and decomposition methods generate only weak (e.g., C- or M-) stationary points that may not be first-order solutions to the MPLCC. Piecewise sequential quadratic programming methods enjoy stronger convergence properties, but need to solve expensive subproblems. Here we propose a primal-dual active set projected Newton method for MPLCCs, that maintains the feasibility of all iterates. At every iteration the method generates a working set for predicting the active set. The projected step direction on the subspace associated with this working set is determined by the current dual iterate, while other elements in the step direction are computed by a Newton system. The major cost of a subproblem involves one matrix factorization and is comparable to that of NLP based algorithms. Our method has strong convergence properties. In particular, under the MPLCC-linear independence constraint qualification, any accumulation point of the generated iterates is a B-stationary solution (i.e., a first-order solution) to the MPLCC. The asymptotic rate of convergence is quadratic under additional MPLCC-second-order sufficient conditions and strict complementarity.

Key words. mathematical programs with linear complementarity constraints, equilibrium constraints, active set method, B-stationary point, global convergence, local convergence

1. Introduction. A mathematical program with equilibrium constraints (MPEC) is an optimization problem, whose constraint set includes variational inequalities. Specifically, an MPEC can be expressed by a bilevel programming problem in the form

$$\begin{array}{ll} \min_{z,y} & F(z,y) \\ \text{s.t.} & (z,y) \in \mathcal{F}, \\ & y \in \mathcal{Q}(z), \end{array} \tag{1.1}$$

where $F(\cdot)$ is the "first-level" objective function in the first-level variables $z \in \Re^n$ and "second-level" variables $y \in \Re^m$ and $\mathcal{F} \subseteq \Re^{n+m}$ is the joint feasible region. The set $\mathcal{Q}(z)$ is the solution set of a variational inequality parameterized by z with the feasible set $\bar{\mathcal{F}}(z)$, i.e.,

$$y \in \mathcal{Q}(z) \Leftrightarrow \{y \in \bar{\mathcal{F}}(z) | G(z,y)^\top (w-y) \geq 0,\ \forall w \in \bar{\mathcal{F}}(z)\}, \tag{1.2}$$

where $G : \Re^{n+m} \to \Re^m$ is a mapping in y and z and $\top$ denotes the transpose of a vector. In particular, if $G(z,y)$ is the gradient of a differentiable function $g(z,y)$ in the vector y, i.e., $G(z,y) = \nabla_y g(z,y)$, then the solution set $\mathcal{Q}(z)$ reduces to the first-order solution set of the optimization problem

$$\min_y \ g(z,y) \ \text{ s.t. } \ y \in \bar{\mathcal{F}}(z). \tag{1.3}$$

MPECs have received increasing interest in recent years because of their numerous applications in engineering and economics, see [7, 24, 26]. For example, in game theory, e.g., the static Stackelberg game can be formulated as an MPEC in which the variational inequality constraints arise from a competitive equilibrium [3].

Problem (1.1), being typically non-convex and non-differentiable, is intrinsically hard to solve. For example, even in the linear case, i.e., $F(\cdot)$ and $G(\cdot)$ are linear mappings and $\mathcal{F}$ and $\bar{\mathcal{F}}(z)$ for

*IEOR Department, Columbia University, New York, NY10027 (lifeng.chen@columbia.edu). Research supported by the Presidential Fellowship of Columbia University.

†IEOR Department, Columbia University, New York, NY10027 (gold@ieor.columbia.edu). Research supported in part by NSF Grant DMS 01-04282, DOE Grant DE-FG02-92EQ25126 and DNR Grant N00014-03-0514.

all z are polyhedra, problem (1.1) is strongly NP-hard. Indeed, in this case, problem (1.1) can be formulated as a mixed integer linear program and determining whether a solution is optimal to (1.1) is strongly NP-hard.

An attractive way to handle the underlying non-differentiability in the lower-level constraints of (1.1) is to replace the constraint $y \in \mathcal{Q}(z)$ by its first-order necessary conditions. Suppose the feasible sets $\mathcal{F}$ and every $\bar{\mathcal{F}}(z)$ are polyhedra, i.e., $\mathcal{F} = \{(z,y)|D_1 z + D_2 y \leq d\}$ and $\bar{\mathcal{F}}(z) = \{w|\bar{D}_1 z + \bar{D}_2 w \leq \bar{d}\}$ for some matrices $D_1,\ D_2,\ \bar{D}_1$ and $\bar{D}_2$ and some vectors d and $\bar{d}$ with corresponding dimensions. The necessary conditions for $y \in \mathcal{Q}(z)$ give rise to a system of equations and inequalities in the primal variables y and the dual variables u

$$G(z,y) + \bar{D}_2^\top u = 0,\ \bar{D}_1 z + \bar{D}_2 y \leq \bar{d},\ u \geq 0,\ (\bar{D}_1 z + \bar{D}_2 y - \bar{d})^\top u = 0. \tag{1.4}$$

Note that when $G(z,y) = \nabla_y g(z,y)$, (1.4) are the Karush-Kuhn-Tucker (KKT) optimality conditions for problem (1.3). These conditions become sufficient conditions for $y \in \mathcal{Q}(z)$ if the mapping $G(z,w)$ is monotone in w. Replacing the constraint $y \in \mathcal{Q}(z)$ in (1.1) by (1.4), problem (1.1) becomes a mathematical program with complementarity constraints (MPCC) with variables (z,y,u)

$$\begin{array}{rl} \min_{z,y} & F(z,y) \\ \text{s.t.} & D_1 z + D_2 y \leq d, \\ & \bar{D}_1 z + \bar{D}_2 y \leq \bar{d},\ u \geq 0 \\ & G(z,y) + \bar{D}_2^\top u = 0, \\ & (\bar{D}_1 z + \bar{D}_2 y - \bar{d})^\top u = 0. \end{array} \tag{1.5}$$

Our primary interest in this paper is the development of a solution method for problem (1.5) that possesses strong convergence properties. In particular, we will restrict ourselves to the situation in which $G(z,y)$ is a linear mapping in both y and z. In such a case problem (1.5) reduces to a so-called mathematical program with linear complementarity constraints (MPLCC). To simply notation, we will focus on MPLCCs in the standard form:

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & Ax = b, \\ & x_0 \geq 0, \\ & 0 \leq x_1 \perp x_2 \geq 0. \end{array} \tag{1.6}$$

The vector of variables x has three parts, i.e., $x = (x_0^\top, x_1^\top, x_2^\top)^\top$ with $x_0 \in \Re^n,\ x_1, x_2 \in \Re^m$. The complementarity operator $\perp$ requires that for each pair of nonnegative complementarity variables $x_{1,i}, x_{2,i} \geq 0$, either $x_{1,i} = 0$ or $x_{2,i} = 0$ for all $i = 1, ..., m$. Throughout the paper we assume that the feasible region of problem (1.6) is nonempty, that the objective function $f : \Re^{n+2m} \rightarrow \Re$ is twice continuously differentiable, and that the matrix $A \in \Re^{p\times(n+2m)}$ and the vector $b \in \Re^p$.

From a nonlinear programming (NLP) point of view, MPCCs are among the most highly degenerate problems since they violate at every feasible point the Mangasarian-Fromovitz constraint qualification (MFCQ), a key ingredient for stability. This lack of regularity may impose difficulties for existing nonlinear programming solvers. Interior-point algorithms do not directly apply since there is no strictly feasible solution to (1.6). Sequential quadratic programming (SQP) methods can get stuck due to infeasible QP subproblems. Solution methods that overcome these difficulties are discussed in the next subsection.

1.1. Related work. Research on solution methods for the continuous variant (1.6) of MEPCs has followed two main approaches: regularization and decomposition.

Although any smooth reformulation of the complementarity constraints in (1.6) violates the MFCQ, slightly relaxing these reformulations resolves this difficulty and results in problems satisfying the MFCQ. State-of-the-art NLP solvers can then be applied. The basic idea of regularization methods is to compute solutions for a sequence of the relaxed problems controlled by some parameter, obtaining a stationary solution to problem (1.6) when the parameter goes beyond a certain threshold or goes to the limit.

For instance, the regularization framework [31] introduces a positive parameter ϕ and relaxes the complementarity constraints to $\{x_1, x_2 \geq 0,\ x_1^\top x_2 \leq \phi\}$. Interior-point algorithms for NLPs are shown to be able to solve the resulting relaxations [23, 27]. Global and fast local convergence to stationary solutions of problem (1.6) is analyzed by letting ϕ converge to zero. In [5] fast local convergence is also established allowing ϕ to be bounded away from zero. An alternative to this relaxation is to introduce an ℓ_1-penalty for the complementarity constraints and add $rx_1^\top x_2$ to the objective function, where r is the penalty parameter. The resulting penalized problem is well behaved. Together with a suitable technique for updating r, the penalty scheme identifies a stationary solution of (1.6) as long as r is large enough or r goes to infinity [20]. Other regularization approaches that explore different relaxations include an elastic model [1, 2], smoothing schemes [6, 12, 17] as well as the approaches proposed in [16, 21, 32], etc.

Much work has been done on understanding the limiting behavior of regularization methods for MEPCs with increasingly tighter relaxations. It turns out that the convergence properties of all the proposed regularization schemes are surprisingly similar. In particular, any cluster point of the sequence of the first-order solutions for these relaxations is at least a C-stationary point of the MPEC. Moreover, if these first-order solutions are also second-order solutions to the relaxed problems, their cluster points are at least M-stationary points. The assumption made to achieve these results is, in general, the MPEC-linear independence constraint qualification (MPEC-LICQ), also known as the primal non-degeneracy assumption. An important question that remains open is whether regularization methods can be strengthened to provided convergence to first-order (i.e., B-stationary) solutions of the MPEC, since there are examples showing that regularization methods may converge to weak C- or M-stationary points at which trivial first-order descent directions exist. Such points are really spurious stationary points that are not locally optimal.

Another approach for attacking problem (1.6) is to use decomposition methods. These methods explore the disjunctive structure of the complementarity constraints and are essentially an extension of pivoting algorithms for complementarity problems. At each iteration, these methods specify two subsets B_1 and B_2 with $B_1 \cup B_2 = \{1, 2, ..., m\}$. By decomposing the complementarity constraints according to B_1 and B_2, they compute a stationary solution to the NLP subproblem

$$
\begin{array}{ll}
\min & f(x) \\
& Ax = b,\ x_0 \geq 0, \\
& x_{1,i} = 0,\ i \in B_1, \quad x_{1,i} \geq 0,\ i \in B_2 \backslash B_1, \\
& x_{2,i} = 0,\ i \in B_2, \quad x_{2,i} \geq 0,\ i \in B_1 \backslash B_2.
\end{array}
\tag{1.7}
$$

Obviously, if the sets B_1 and B_2 are able to identify respectively the correct active sets for x_1 and x_2 at an optimal solution, solutions to problems (1.7) also solve problem (1.6). Therefore, the key point for decomposition methods lies in how to choose B_1 and B_2 effectively, especially when the iterates are far from the solution. One way is to determine these sets according to the current values of the dual variables. Recent developments in decomposition methods show that they have guaranteed convergence to M-stationary points of problem (1.6) assuming primal non-degeneracy [13, 15]. In [14] this result is strengthened to show convergence to B-stationary points under an additional assumption that every M-stationary point is isolated.

An interesting attempt to solve general MPECs is to directly apply state-of-the-art NLP active set solvers, which do not rely on the existence of a strict interior. Based on a filter-SQP solver, Fletcher and Leyffer [9] solve MPECs as NLPs by replacing the complementarity constraints by the inequalities: $x_1 \geq 0,\ x_2 \geq 0$ and $x_1 x_2 \leq 0$. Surprisingly, quadratic convergence is observed for almost all tested MPEC problems. The reason for this is studied in [10].

Piecewise SQP (PSQP) methods are the natural extension of NLP SQP methods for solving MEPCs. The subproblem solved by PSQP methods at each iteration is a quadratic program with linear complementarity constraints (QPLCC). As a variant of SQP methods, it is not surprising that PSQP methods enjoy reasonably strong convergence properties. In particular, it has been shown that under rather weak conditions, PSQP methods converge quadratically in a small neighborhood of a second-order local solution [25, 28]. Unfortunately, since solving the QPLCC subproblem is much more expensive than solving a regular quadratic program, current PSQP methods are only conceptual. How to reduce this cost and still guarantee the strong convergence properties of PSQP methods is the subject of current research [22].

1.2. Contribution. Given the state of current solution methods for problem (1.6), this paper is aimed at an efficient method that not only is implementable with reasonable subproblem cost, but also enjoys strong global and local convergence to meaningful stationary solutions. In particular, we propose a primal-dual active set Newton method for problem (1.6). The method starts from a feasible point and maintains feasibility at every iterate. In many important applications, a feasible starting point to problem (1.6) is readily available through state-of-the-art codes for linear complementarity problems (LCP) or linear variational inequalities. Moreover, in many interesting cases such a feasible solution can be obtained in polynomial time, e.g., if the matrix defining the LCP is positive semidefinite, or the variational inequalities are strictly monotone.

At every iteration our method generates a working set for predicting the active set. The complementarity and inequality constraints in the working set are treated approximately as equality constraints depending on the current estimate for the dual variables, while those not in the set are temporarily ignored. In terms of projected Newton methods, the working set defines the projection space, in which elements of the step direction are chosen according to the multipliers. The subproblem solved at each step is as inexpensive to solve as the subproblem solved in NLP based algorithms. Specifically, the major cost involves only one matrix factorization besides the cost needed to guarantee the correct condition of the Newton system. The convergence properties of our algorithm are as strong as those of the PSQP methods and stronger than any that have been shown for NLP-based regularization or decomposition methods. In particular, we show that under the primal non-degeneracy condition, any accumulation point of the generated iterates is a B-stationary point (i.e., a first-order solution) to problem (1.6). Under additional MPLCC-second order sufficient conditions (SOSC) and strict complementarity, we show that the generated working set eventually identifies the final active set at the B-stationary solution. As a consequence, the convergence rate of our algorithm is locally quadratic.

As a first step for this research, in this paper we will focus on the global and local convergence theory of the proposed method. In particular, the paper is organized as follows. Section 2 presents notation and some definitions and preliminary results relevant to MPECs. In section 3, we describe our active set algorithm for solving problem (1.6). Global convergence to B-stationary solutions and fast local convergence are established in Sections 4 and 5, respectively. We conclude the paper in Section 6.

2. Preliminaries. We use standard notation in the literature of MPECs and NLPs. A letter with superscript or subscript k is related to the kth iteration. The subscript i means the ith element of a vector or the ith column of a matrix. $\emptyset$ stands for empty set. $\|\cdot\|$ denotes the Euclidean norm unless otherwise specified. For a vector d of dimension $n+2m$, we write $d = (d_0^\top, d_1^\top, d_2^\top)^\top$ with $d_0 \in \Re^n$, $d_1 \in \Re^m$ and $d_2 \in \Re^m$, respectively, as $(d_0; d_1; d_2)$. Moreover, for an index subset $\mathcal{D}$, $d_{0,\mathcal{D}}$ ($d_{1,\mathcal{D}}$, $d_{2,\mathcal{D}}$) denote the sub-vector of d_0 (d_1, d_2) with components $d_{0,i}$ ($d_{1,i}$, $d_{2,i}$) for $i \in \mathcal{D}$. For an index i, vectors $e_{0,i}$, $e_{1,i}$ and $e_{2,i}$ are respectively the $n+2m$ dimensional gradients of $x_{0,i}$, $x_{1,i}$ and $x_{2,i}$, i.e., all elements of $e_{0,i}$ ($e_{1,i}$, $e_{2,i}$) are zero except the one corresponding to variable $x_{0,i}$ ($x_{1,i}$, $x_{2,i}$), which is one. Matrices $e_0^\top$, $e_1^\top$ and $e_2^\top$ denote the Jacobian of x_0, x_1 and x_2, respectively. Given two vectors x and y of the same dimension q, we say $x \geq (>) y$ if and only if $x_i \geq (>) y_i$, $\forall i = 1, ..., q$, and $\min(x, y)$ is a vector whose ith element is $\min(x_i, y_i)$. We define the following index sets for the constraints in problem (1.6)

$$\mathcal{E} = \{1, ..., p\},\ \mathcal{I} = \{1, ..., n\},\ \mathcal{A} = \{1, ..., m\}.$$

The active sets for inequality and complementarity constraints at a feasible point x are defined by

$$\mathcal{I}(x) = \{i \in \mathcal{I} | x_{0,i} = 0\},\ \mathcal{A}_1(x) = \{i \in \mathcal{A} | x_{1,i} = 0\},\ \mathcal{A}_2(x) = \{i \in \mathcal{A} | x_{2,i} = 0\}.$$

In particular, the bi-active set $\mathcal{A}_{12}(x)$ for the complementarity constraints is defined as the intersection of $\mathcal{A}_1(x)$ and $\mathcal{A}_2(x)$, i.e., $\mathcal{A}_{12}(x) = \mathcal{A}_1(x) \cap \mathcal{A}_2(x)$.

A local optimal solution to problem (1.6) must be a first-order optimal solution, also known as a B-stationary point, at which no feasible descent directions exist. The primal stationarity condition for B-stationary points is given by

Definition 2.1. *A feasible point x is a B-stationary point of problem (1.6) if the following linear program with linear complementarity constraints is solved by $\triangle x = (\triangle x_0; \triangle x_1; \triangle x_2) = 0$,*

$$\begin{array}{ll} \min_{\triangle x} & \nabla f(x)^\top \triangle x \\ \text{s.t.} & A\triangle x = 0, \\ & x_0 + \triangle x_0 \geq 0, \\ & 0 \leq (x_1 + \triangle x_1) \perp (x_2 + \triangle x_2) \geq 0. \end{array} \tag{2.1}$$

Checking B-stationarity from Definition 2.1 is usually very difficult. Thus, it is more convenient to consider primal-dual stationarity conditions. For instance, under certain constraint qualifications, one can check whether a feasible point x is a weak stationary solution of problem (1.6); i.e., there exist dual variables $y \in \Re^p$ and $\lambda = (\lambda_0; \lambda_1; , \lambda_2) \in \Re^{n+2m}$ that along with x satisfy

$$\nabla_x \mathcal{L}(x, y, \lambda) = 0, \tag{2.2}$$

$$\lambda_0 \geq 0,\ \lambda_{0,i} x_{0,i} = 0,\ \forall i \in \mathcal{I}, \tag{2.3}$$

$$\lambda_{1,i} x_{1,i} = 0,\ \lambda_{2,i} x_{2,i} = 0,\ \forall i \in \mathcal{A}, \tag{2.4}$$

where $\mathcal{L}(x, y, \lambda)$ is the Lagrangian function

$$\mathcal{L}(x, y, \lambda) = f(x) + (Ax - b)^\top y - x_0^\top \lambda_0 - x_1^\top \lambda_1 - x_2^\top \lambda_2.$$

Unlike the simple multiplier rule for NLPs, there are several levels of stationarity conditions for MPLCCs depending on the sign to the complementarity constraint multipliers. The stationarity conditions of primary interest in this paper include C-stationarity, M-stationarity and S-stationarity.

Definition 2.2. *A feasible point x is called a C-stationary point of problem (1.6), if there exist y and λ such that (2.2), (2.3) and (2.4) hold and $\lambda_{1,i}\lambda_{2,i} \geq 0$ for all $i \in \mathcal{A}$.*

Definition 2.3. *A feasible point x is called a M-stationary point of problem (1.6), if there exist y and λ such that (2.2), (2.3) and (2.4) hold and either $\lambda_{1,i} \geq 0,\ \lambda_{2,i} \geq 0$ or $\lambda_{1,i}\lambda_{2,i} = 0$ for all $i \in \mathcal{A}_{12}(x)$.*

Definition 2.4. *A feasible point x is called a S-stationary point of problem (1.6), if there exist y and λ such that (2.2), (2.3) and (2.4) hold and $\lambda_{1,i} \geq 0,\ \lambda_{2,i} \geq 0$ for all $i \in \mathcal{A}_{12}(x)$.*

Note that S-stationarity is the strongest stationarity condition. A S-stationary point must be a M-stationary point and a M-stationary point must be a C-stationary point. Moreover, a S-stationary point must be a B-stationary point. B-stationarity does not necessarily imply S-stationarity and S-stationary points may not exist for problem (1.6). However, as is well known (see, e.g., [30]), B-stationarity is equivalent to S-stationarity under the following primal non-degeneracy condition:

Definition 2.5. *(MPLCC-LICQ) A feasible point x is said to satisfy the MPLCC-LICQ, if the following constraint gradients are linearly independent:*

$$\{A_i, i \in \mathcal{E}\} \cup \{e_{0,i},\ i \in \mathcal{I}(x)\} \cup \{e_{1,i},\ i \in \mathcal{A}_1(x)\} \cup \{e_{2,i},\ i \in \mathcal{A}_2(x)\}.$$

MPLCC-LICQ is important in both practice and theory. It allows the identification of a B-stationary solution as a S-stationary point, which is much easier than solving the LPLCC (2.1). In particular, given the validity of the MPLCC-LICQ at x, the unique multiplier vector (y, λ) that satisfies (2.2) and the complementarity conditions can be obtained by solving the linear system

$$\nabla f(x) + A^\top y - \sum_{i \in \mathcal{I}(x)} \lambda_{0,i} e_{0,i} - \sum_{i \in \mathcal{A}_1(x)} \lambda_{1,i} e_{1,i} - \sum_{i \in \mathcal{A}_2(x)} \lambda_{2,i} e_{2,i} = 0.$$

For other multipliers in λ set $\lambda_{0,i} = 0,\ \forall i \in \mathcal{I}\backslash\mathcal{I}(x)$, $\lambda_{1,i} = 0,\ \forall i \in \mathcal{A}\backslash\mathcal{A}_1(x)$ and $\lambda_{2,i} = 0,\ \forall i \in \mathcal{A}\backslash\mathcal{A}_2(x)$. If furthermore the multipliers corresponding to the bi-active set $\mathcal{A}_{12}(x)$ are all nonnegative, x is a S-stationary point and thus also a B-stationary point. Otherwise, x is not a B-stationary point.

We now define second-order sufficient conditions (SOSC) for a feasible point x to be a solution of problem (1.6).

Definition 2.6. *The MPLCC-SOSC hold at a feasible point x if x is a S-stationary point of problem (1.6) and there is a $\varrho > 0$ such that $d^\top \nabla^2 f(x) d \geq \varrho \|d\|^2$ for all*

$$d \in \mathcal{S}(x) = \{s \in \Re^{n+2m} | As = 0;\quad s_{0,i} = 0,\ \forall i \in \mathcal{I}(x); s_{1,i} = 0,\ \forall i \in \mathcal{A}_1(x);\ s_{2,i} = 0,\ \forall i \in \mathcal{A}_2(x)\}.$$

Strict complementarity for problem (1.6) is defined as follows.

Definition 2.7. *Let x be a S-stationary point of problem (1.6) and (y, λ) be a corresponding multiplier vector. Strict complementarity holds at (x, y, λ) if $\lambda_{0,i} > 0,\ \forall i \in \mathcal{I}(x)$, $\lambda_{1,i} \neq 0,\ \forall i \in \mathcal{A}_1(x)$ and $\lambda_{2,i} \neq 0,\ \forall i \in \mathcal{A}_2(x)$.*

So far we have described stationarity conditions for problem (1.6) from the perspective of MPECs. Now let us look at problem (1.6) from the view point of nonlinear optimization by writing it as:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax = b, \\ & x_0 \geq 0,\ x_1 \geq 0,\ x_2 \geq 0, \\ & x_1^\top x_2 \leq 0. \end{array} \tag{2.5}$$

A feasible point x is a first-order optimal solution to problem (2.5), if there exist multipliers $y \in \Re^p$, $z_0 \in \Re^n$, $z_1 \in \Re^m$, $z_2 \in \Re^m$ and $z_3 \in \Re$ such that the KKT conditions hold; i.e.,

$$\mathcal{R}(x, y, z) = \begin{bmatrix} \nabla_x \bar{\mathcal{L}}(x, y, z) \\ \min\{x_0, z_0\}, \\ \min\{x_1, z_1\}, \\ \min\{x_2, z_2\}, \\ \min\{x_1^\top x_2, z_3\} \end{bmatrix} = 0, \tag{2.6}$$

where $z = (z_0; z_1; z_2; z_3)$ and $\bar{\mathcal{L}}(x, y, z)$ is the Lagrangian function associated with the NLP (2.5),

$$\bar{\mathcal{L}}(x, y, z) = f(x) + (Ax - b)^\top y - x_0^\top z_0 - x_1^\top z_1 - x_2^\top z_2 + z_3 x_1^\top x_2. \tag{2.7}$$

Proposition 4.1 in [10] shows the strong relation between a KKT point of problem (2.5) and a S-stationary solution to problem (1.6), which we describe as follows.

PROPOSITION 2.8. *Suppose x is a feasible point of (1.6) and (2.5). If (x, y, z) satisfies the KKT conditions (2.6), then x is a S-stationary point of (1.6) with multipliers y and λ, where λ is defined by*

$$\lambda_0 = z_0, \quad \lambda_1 = z_1 - z_3 x_2, \quad \lambda_2 = z_2 - z_3 x_1. \tag{2.8}$$

If x is a S-stationary point of problem (1.6) with a corresponding multiplier vector (y, λ), then (x, y, z) satisfies the KKT conditions (2.6), where z is defined by

$$\begin{cases} z_0 = \lambda_0, \\ z_3 \in \{\theta \geq 0 | \lambda_1 + \theta x_2 \geq 0,\ \lambda_2 + \theta x_1 \geq 0\}, \\ z_1 = \lambda_1 + z_3 x_2, \\ z_2 = \lambda_2 + z_3 x_1. \end{cases} \tag{2.9}$$

The SOSC for problem (2.5) follows from the standard definition of the SOSC for general NLPs. In particular, it is easy to check that under the MPLCC-LICQ (implying the uniqueness of the multipliers) and the MPLCC-SOSC, the SOSC for problem (2.5) holds for any multipliers that satisfy (2.9).

3. Algorithm. Our algorithm is an active set Newton method, which maintains the feasibility of all iterates. At every iteration, the algorithm computes a working set that approximates the final active set at a solution. Given an iterate x and a multiplier estimate (y, z), the indices of the working set that correspond, respectively to x_0, x_1 and x_2, are those indices in the sets

$$\begin{cases} \mathcal{I}(x, y, z; \varepsilon) &= \{i \in \mathcal{I}\, | x_{0,i} \leq \min\{\varepsilon, \|\mathcal{R}(x, y, z)\|^{\eta}\}\}, \\ \mathcal{A}_1(x, y, z; \varepsilon) &= \{i \in \mathcal{A} | x_{1,i} \leq \min\{\varepsilon, \|\mathcal{R}(x, y, z)\|^{\eta}\}\}, \\ \mathcal{A}_2(x, y, z; \varepsilon) &= \{i \in \mathcal{A} | x_{2,i} \leq \min\{\varepsilon, \|\mathcal{R}(x, y, z)\|^{\eta}\}\}, \end{cases} \tag{3.1}$$

where $\varepsilon > 0$ and $\eta \in (0, 1)$ are parameters, $\mathcal{R}(x, y, z)$ is defined by (2.6). The set

$$\mathcal{A}_{12}(x, y, z; \varepsilon) = \mathcal{A}_1(x, y, z; \varepsilon) \cap \mathcal{A}_2(x, y, z; \varepsilon) \tag{3.2}$$

is called the bi-working set.

Let $\mathcal{W}$ denote the current working set; i.e., $\mathcal{W} \equiv (\mathcal{I}(x, y, z; \varepsilon), \mathcal{A}_1(x, y, z; \varepsilon), \mathcal{A}_2(x, y, z; \varepsilon))$. Also let the Jacobian of the inequality and complementarity constraints in this working set be denoted by

$$E(x, y, z; \varepsilon) = [E_0(x, y, z; \varepsilon), E_1(x, y, z; \varepsilon), E_2(x, y, z; \varepsilon)], \tag{3.3}$$

where

$$\begin{cases} E_0(x, y, z; \varepsilon) &= [e_{0,i},\ i \in \mathcal{I}(x, y, z; \varepsilon)], \\ E_1(x, y, z; \varepsilon) &= [e_{1,i},\ i \in \mathcal{A}_1(x, y, z; \varepsilon)], \\ E_2(x, y, z; \varepsilon) &= [e_{2,i},\ i \in \mathcal{A}_2(x, y, z; \varepsilon)]. \end{cases} \tag{3.4}$$

Using the notation $d_{\mathcal{W}} = (d_{\mathcal{I}(x,y,z;\varepsilon)}; d_{\mathcal{A}_1(x,y,z;\varepsilon)}; d_{\mathcal{A}_2(x,y,z;\varepsilon)})$, several Newton systems are solved at each iteration, where each system is a linearization of the nonlinear system of equations

$$\begin{aligned} \nabla f(x) + A^\top y - E(x, y, z; \varepsilon)\lambda_{\mathcal{W}} &= 0, \\ Ax - b &= 0, \\ x_{\mathcal{W}} - u_{\mathcal{W}} &= 0 \end{aligned}$$

for some appropriate choice of $u_{\mathcal{W}}$. In particular, letting $\varpi = u_{\mathcal{W}} - x_{\mathcal{W}}$, the Newton systems are

$$\mathcal{M}(\mathcal{H}, x, y, z; \varepsilon)(\triangle x; y^+; \lambda_{\mathcal{W}}^+) = (-\nabla f(x); 0; \varpi), \tag{3.5}$$

where the Newton matrix is given by

$$\mathcal{M}(\mathcal{H}, x, y, z; \varepsilon) = \begin{bmatrix} \mathcal{H} & A^\top & -E(x, y, z; \varepsilon) \\ A & 0 & 0 \\ E(x, y, z; \varepsilon)^\top & 0 & 0 \end{bmatrix}, \tag{3.6}$$

and $\mathcal{H}$ is the exact Hessian $\nabla^2 f(x)$ or an approximation to it.

At the start of each iteration, the system (3.5) is solved with $\varpi = 0$, which is equivalent to keeping the components of $x_{\mathcal{W}}$ fixed at their current values (i.e., setting $u_{\mathcal{W}} = x_{\mathcal{W}}$), to obtain an updated estimate $\lambda_{\mathcal{W}}^+$ of the dual variables. The values of the components of $\lambda_{\mathcal{W}}^+$ and $x_{\mathcal{W}}$ are then compared to determine which indices in the working set $\mathcal{W}$ are likely to be in an optimal active set, and ϖ is chosen so that a full Newton step will make the corresponding constraints active. If this step, which is modified by a step length parameter, keeps all variables nonnegative and satisfies an Armijo sufficient decrease condition, the modified step is taken, the multipliers are updated and the iteration is complete. Otherwise, the value of the tolerance ε used to define the working set in (3.1) is reduced to further refine the working set, and the above step computation procedure is repeated.

Now we are ready to present our active set algorithm for solving problem (1.6). The motivations underlying the computations performed by the algorithm are more fully explained in the remarks that follow it.

ALGORITHM 3.1.
Step 0. Initialization.
Parameters: $\eta \in (0,1)$, $z_3^{\max} > 0$, $\kappa \in (0,1)$, $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$, $\gamma > 1$, $\nu \in (0,1)$.
Initial data: a feasible point $x^0 \in \Re^{n+2m}$, $y^0 \in \Re^p$, $z^0 \in \Re^{n+2m+1}$ with $z_3^0 \geq 0$, $\mathcal{H}^0 = \nabla^2 f(x^0)$, $\varepsilon_0 > 0$.
Set $k \leftarrow 0$.

Step 1. Factorization of the Newton matrix.
Step 1.1. If the Jacobian $[A^\top, E(x^k, y^k, z^k; \varepsilon_k)]$ has full column rank, set $\varepsilon_{k,0} = \varepsilon_k$; otherwise, set $\varepsilon_{k,0} = 0$ and $\varepsilon_{k+1} = \kappa\varepsilon_k$.
Step 1.2. Modify $\mathcal{H}^k$ if necessary, so that

$$d^\top \mathcal{H}^k d > 0, \ \forall d \in \mathcal{S}^k, d \neq 0, \tag{3.7}$$

where

$$\mathcal{S}^k = \{s \in \Re^{n+2m} | As = 0; \quad s_{0,i} = 0, \ \forall i \in \mathcal{I}^k; s_{1,i} = 0, \ \forall i \in \mathcal{A}_1^k; \ s_{2,i} = 0, \ \forall i \in \mathcal{A}_2^k\}, \tag{3.8}$$

and

$$\mathcal{I}^k = \mathcal{I}(x^k, y^k, z^k; \varepsilon_{k,0}), \ \mathcal{A}_1^k = \mathcal{A}_1(x^k, y^k, z^k; \varepsilon_{k,0}), \ \mathcal{A}_2^k = \mathcal{A}_2(x^k, y^k, z^k; \varepsilon_{k,0}). \tag{3.9}$$

Step 2. Computation of a Newton direction.
Step 2.1. Solve

$$\mathcal{M}^k \left(\triangle x^k; y^{k+1}; \bar{\lambda}^{k+1}\right) = \left(-\nabla f(x^k); 0; 0\right) \tag{3.10}$$

where $\mathcal{M}^k = \mathcal{M}(\mathcal{H}^k, x^k, y^k, z^k; \varepsilon_{k,0})$ and $\bar{\lambda}^{k+1} = \left(\lambda^{k+1}_{0,\mathcal{I}^k}; \lambda^{k+1}_{1,\mathcal{A}_1^k}; \lambda^{k+1}_{2,\mathcal{A}_2^k}\right)$. Set $\lambda^{k+1}_{0,i} = 0,\ \forall i \in \mathcal{I}\backslash\mathcal{I}^k$, $\lambda^{k+1}_{1,i} = 0,\ \forall i \in \mathcal{A}\backslash\mathcal{A}_1^k$ and $\lambda^{k+1}_{2,i} = 0,\ \forall i \in \mathcal{A}\backslash\mathcal{A}_2^k$.
Step 2.2. If x^k and the multipliers (y^{k+1}, λ^{k+1}) satisfy the conditions in Definition 2.4, stop; x^k is a S-stationary point of problem (1.6). Otherwise, continue to Step 3.

Step 3. Line search.
Step 3.1. Set $j \leftarrow 0$ and

$$\gamma_k^{\max} = \gamma \max\{1;\ x^k_{1,i},\ i \in \mathcal{A}_1^k;\ x^k_{2,i},\ i \in \mathcal{A}_2^k\}, \tag{3.11}$$

$$\mathcal{I}^k_+ = \left\{i \in \mathcal{I}^k | \lambda^{k+1}_{0,i} \geq x^k_{0,i}\right\},\ \mathcal{I}^k_- = \left\{i \in \mathcal{I}^k | \lambda^{k+1}_{0,i} < x^k_{0,i}\right\}. \tag{3.12}$$

Step 3.2. Set $\mathcal{A}_1^{k,j} = \mathcal{A}_1(x^k, y^k, z^k; \varepsilon_{k,j})$ and $\mathcal{A}_2^{k,j} = \mathcal{A}_2(x^k, y^k, z^k; \varepsilon_{k,j})$. If $j = 0$, set $\alpha_{k,j} = 1$; otherwise, set

$$\alpha_{k,j} = \min\left\{\theta \,\middle|\, \beta\alpha_{k,j-1} \leq \theta \leq \alpha_{k,j-1};\ \theta \geq \frac{x^k_{1,i}}{\gamma_k^{\max}},\ \forall i \in \mathcal{A}_1^{k,j};\ \theta \geq \frac{x^k_{2,i}}{\gamma_k^{\max}},\ \forall i \in \mathcal{A}_2^{k,j}\right\}. \tag{3.13}$$

Let $\mathcal{A}_{12}^{k,j} = \mathcal{A}_1^{k,j} \cap \mathcal{A}_2^{k,j}$ and

$$\left\{\begin{array}{lcl}
\mathcal{B}_1^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | \lambda^{k+1}_{1,i} \geq x^k_{1,i},\ \lambda^{k+1}_{2,i} \geq x^k_{2,i}\right\}, \\
\mathcal{B}_2^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{1,i} = 0,\ \lambda^{k+1}_{2,i} < x^k_{2,i},\ x^k_{1,i} \leq \lambda^{k+1}_{1,i}\right\}, \\
\mathcal{B}_3^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{1,i} = 0,\ \lambda^{k+1}_{2,i} < x^k_{2,i},\ \lambda^{k+1}_{1,i} < x^k_{1,i},\ \lambda^{k+1}_{2,i} \leq \lambda^{k+1}_{1,i}\right\}, \\
\mathcal{B}_4^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{1,i} = 0,\ \lambda^{k+1}_{2,i} < x^k_{2,i},\ \lambda^{k+1}_{1,i} < x^k_{1,i},\ \lambda^{k+1}_{1,i} < \lambda^{k+1}_{2,i}\right\}, \\
\mathcal{B}_5^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{1,i} = 0,\ x^k_{2,i} \leq \lambda^{k+1}_{2,i},\ \lambda^{k+1}_{1,i} < x^k_{1,i}\right\}, \\
\mathcal{B}_6^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{2,i} = 0,\ \lambda^{k+1}_{1,i} < x^k_{1,i},\ x^k_{2,i} \leq \lambda^{k+1}_{2,i}\right\}, \\
\mathcal{B}_7^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{2,i} = 0,\ \lambda^{k+1}_{1,i} < x^k_{1,i},\ \lambda^{k+1}_{2,i} < x^k_{2,i},\ \lambda^{k+1}_{1,i} \leq \lambda^{k+1}_{2,i}\right\}, \\
\mathcal{B}_8^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{2,i} = 0,\ \lambda^{k+1}_{1,i} < x^k_{1,i},\ \lambda^{k+1}_{2,i} < x^k_{2,i},\ \lambda^{k+1}_{2,i} < \lambda^{k+1}_{1,i}\right\}, \\
\mathcal{B}_9^{k,j} & = & \left\{i \in \mathcal{A}_{12}^{k,j}\, | x^k_{2,i} = 0,\ x^k_{1,i} \leq \lambda^{k+1}_{1,i},\ \lambda^{k+1}_{2,i} < x^k_{2,i}\right\}, \\
\mathcal{B}_{10}^{k,j} & = & \left\{i \in \mathcal{A}_1^k\backslash\mathcal{A}_1^{k,j}\, | x^k_{1,i} \leq \lambda^{k+1}_{1,i}\right\}, \\
\mathcal{B}_{11}^{k,j} & = & \left\{i \in \mathcal{A}_1^k\backslash\mathcal{A}_1^{k,j}\, | \lambda^{k+1}_{1,i} < x^k_{1,i}\right\}, \\
\mathcal{B}_{12}^{k,j} & = & \left\{i \in \mathcal{A}_2^k\backslash\mathcal{A}_2^{k,j}\, | x^k_{2,i} \leq \lambda^{k+1}_{2,i}\right\}, \\
\mathcal{B}_{13}^{k,j} & = & \left\{i \in \mathcal{A}_2^k\backslash\mathcal{A}_2^{k,j}\, | \lambda^{k+1}_{2,i} < x^k_{2,i}\right\}.
\end{array}\right. \tag{3.14}$$

Set the components of the vector $\varpi^{k,j} = (\varpi_0^{k,j}; \varpi_1^{k,j}; \varpi_2^{k,j})$, where $\varpi_0^{k,j} \in \Re^{|\mathcal{I}^k|}$, $\varpi_1^{k,j} \in \Re^{|\mathcal{A}_1^k|}$ and

$\varpi_2^{k,j} \in \Re^{|\mathcal{A}_2^k|}$, as follows:

$$
(3.15)\qquad \begin{cases}
\varpi_{0,i}^{k,j} = -x_{0,i}^k, & & i \in \mathcal{I}_+^k,\\
\varpi_{0,i}^{k,j} = -\lambda_{0,i}^{k+1}, & & i \in \mathcal{I}_-^k,\\
\varpi_{1,i}^{k,j} = -x_{1,i}^k (= 0), & & i \in \mathcal{A}_1^{k,j} \backslash \mathcal{A}_{12}^{k,j},\\
& \varpi_{2,i}^{k,j} = -x_{2,i}^k (= 0), & i \in \mathcal{A}_2^{k,j} \backslash \mathcal{A}_{12}^{k,j},\\
\varpi_{1,i}^{k,j} = -x_{1,i}^k, & \varpi_{2,i}^{k,j} = -x_{2,i}^k, & i \in \mathcal{B}_1^{k,j},\\
\varpi_{1,i}^{k,j} = -x_{1,i}^k (= 0), & \varpi_{2,i}^{k,j} = -\lambda_{2,i}^{k+1}, & i \in \mathcal{B}_2^{k,j} \cup \mathcal{B}_3^{k,j},\\
\varpi_{1,i}^{k,j} = -\lambda_{1,i}^{k+1} + x_{2,i}^k/\alpha_{k,j}, & \varpi_{2,i}^{k,j} = -x_{2,i}^k/\alpha_{k,j}, & i \in \mathcal{B}_4^{k,j} \cup \mathcal{B}_5^{k,j},\\
\varpi_{1,i}^{k,j} = -\lambda_{1,i}^{k+1}, & \varpi_{2,i}^{k,j} = -x_{2,i}^k (= 0), & i \in \mathcal{B}_6^{k,j} \cup \mathcal{B}_7^{k,j},\\
\varpi_{1,i}^{k,j} = -x_{1,i}^k/\alpha_{k,j}, & \varpi_{2,i}^{k,j} = -\lambda_{2,i}^{k+1} + x_{1,i}^k/\alpha_{k,j}, & i \in \mathcal{B}_8^{k,j} \cup \mathcal{B}_9^{k,j},\\
\varpi_{1,i}^{k,j} = -x_{1,i}^k, & & i \in \mathcal{B}_{10}^{k,j},\\
\varpi_{1,i}^{k,j} = -\lambda_{1,i}^{k+1}, & & i \in \mathcal{B}_{11}^{k,j},\\
& \varpi_{2,i}^{k,j} = -x_{2,i}^k, & i \in \mathcal{B}_{12}^{k,j},\\
& \varpi_{2,i}^{k,j} = -\lambda_{2,i}^{k+1}, & i \in \mathcal{B}_{13}^{k,j}.
\end{cases}
$$

Step 3.3. Compute a trial step; i.e., using the factorization of $\mathcal{M}^k$ computed in Step 2.1, solve

$$
(3.16)\qquad \mathcal{M}^k \left(\triangle x^{k,j}; y^{k+1,j}; \bar{\lambda}^{k+1,j}\right) = \left(-\nabla f(x^k); 0; \varpi^{k,j}\right),
$$

where $\bar{\lambda}^{k+1,j} = \left(\lambda_{0,\mathcal{I}^k}^{k+1,j}; \lambda_{1,\mathcal{A}_1^k}^{k+1,j}; \lambda_{2,\mathcal{A}_2^k}^{k+1,j}\right)$.

Step 3.4. Check the following conditions:

$$
(3.17)\qquad x^k + \alpha_{k,j} \triangle x^{k,j} \geq 0,
$$

$$
(3.18)\qquad f(x^k + \alpha_{k,j} \triangle x^{k,j}) \leq f(x^k) + \sigma \alpha_{k,j} \nabla f(x^k)^\top \triangle x^{k,j}.
$$

If both (3.17) and (3.18) hold, set $j_k \leftarrow j$, $x^{k+1} = x^k + \alpha_{k,j} \triangle x^{k,j}$ and go to Step 4. Otherwise, set $\varepsilon_{k,j+1} = \nu \varepsilon_{k,j}$, $j \leftarrow j+1$ and go back to Step 3.2.

Step 4. Update.
Set $\mathcal{H}^{k+1} = \nabla^2 f(x^{k+1})$. Compute z^{k+1} by

$$
(3.19)\qquad \begin{cases}
z_0^{k+1} = \lambda_0^{k+1},\\
z_3^{k+1} = \min\{z_3^{\max}, \min\{\theta \geq 0 | \lambda_1^{k+1} + \theta x_2^{k+1} \geq 0,\ \lambda_2^{k+1} + \theta x_1^{k+1} \geq 0\}\},\\
z_1^{k+1} = \lambda_1^{k+1} + z_3^{k+1} x_2^{k+1},\\
z_2^{k+1} = \lambda_2^{k+1} + z_3^{k+1} x_1^{k+1}.
\end{cases}
$$

Set $k \leftarrow k+1$ and go back to Step 1.

Remark 3.1. As we have mentioned earlier, for many interesting applications a feasible starting point to problem (1.6) is readily available in polynomial time through state-of-the-art solvers for complementarity problems.

Remark 3.2. In Step 1.1 of Algorithm 3.1, we check whether the gradients of the equality constraints and the inequality constraints in the working set are linearly independent. Although this procedure may incur additional cost, in practice it can be accomplished as a side product of factorizing the Newton matrix. If these gradients are dependent, the algorithm sets $\varepsilon_{k,0} = 0$, i.e.,

$\mathcal{I}^k = \mathcal{I}(x^k)$, $\mathcal{A}_1^k = \mathcal{A}_1(x^k)$ and $\mathcal{A}_2^k = \mathcal{A}_2(x^k)$. Hence, if the MPLCC-LICQ holds at all feasible points, the gradients corresponding to the equality constraints and the inequality constraints in the working set are now independent. Moreover, if the Hessian estimate $\mathcal{H}$ is positive definite on the null space of the working set Jacobian, i.e., condition (3.7) holds, the Newton matrix $\mathcal{M}^k$ is nonsingular and as will be shown later, the directions computed in Step 2 and Step 3 are descent directions for the objective function. Given that the Jacobian in $\mathcal{M}^k$ has full column rank, condition (3.7) holds if and only if the inertia of the matrix obtained by symmetrizing $\mathcal{M}^k$ is $(n+2m, p+|\mathcal{I}^k|+|\mathcal{A}_1^k|+|\mathcal{A}_2^k|, 0)$, i.e ., the matrix has $n+2m$ positive eigenvalues, $p+|\mathcal{I}^k|+|\mathcal{A}_1^k|+|\mathcal{A}_2^k|$ negative eigenvalues and no zero eigenvalues. The inertia of a symmetric indefinite matrix is readily available from several linear system solvers that perform an $LBL^\top$ factorization. Whenever condition (3.7) fails with the exact Hessian $\mathcal{H} = \nabla^2 f(x)$, we can employ a common approach used in interior-point methods for non-convex NLPs of adding to $\mathcal{H}$ a multiple of the identity with increasing magnitude until (3.7) holds (see, e.g., [33, 34]). Note that (3.7) eventually holds with the exact Hessian as long as the iterates converge to a S-stationary solution of problem (1.6) at which the MPLCC-SOSC holds. Also note that since the linear systems (3.10) and (3.16) have the same coefficient matrix $\mathcal{M}^k$, only one matrix factorization is needed to compute the step directions at each iteration.

Remark 3.3. The Newton step $\triangle x^k$ computed in Step 2.1 is not directly used to update the current iterate x^k. Rather, it is used to compute new dual variable estimates y^{k+1} and λ^{k+1}, which are then used to actually compute trial primal steps $\triangle x^{k,j}$.

Remark 3.4. Our goal for using the working set is to predict the final active set. The optimality error $\|\mathcal{R}(x,y,z)\|^\eta$ used to define the working set in (3.1) is a superior error bound in the sense that if (x,y,z) is close enough to a second-order solution of problem (2.5), this optimality error is an upper bound on the distance of (x,y,z) to the solution. Therefore, the working set defined by (3.1) identifies the final active set once the iterates (x^k, y^k, z^k) are sufficiently close to a solution of problem (2.5) satisfying the SOSC. In this case we can deal with the constraints in the working set exactly as equality constraints and ignore the constraints outside the working set. Hence, the Newton equation yields $\triangle x_{l,i}^k = -x_{l,i}^k$ $(l = 0,1,2)$ for every constraint indexed by i in the working set.

Unfortunately, the error bound $\|\mathcal{R}(x,y,z)\|^\eta$ may not be valid when the iterates are far from a solution. In this case the working set may not correspond to the final active set. Hence, we need to determine which constraints in the working set are likely to be in an optimal active set and which are not. In Algorithm 3.1, this is done by comparing the values of $\lambda_{0,i}^{k+1}$ and $x_{0,i}^k$ for $i \in \mathcal{I}^k$ and the values of $\lambda_{l,i}^{k+1}$ and $x_{l,i}^k$ for $l = 1,2$ and $i \in \mathcal{A}_{12}^k$, where λ^{k+1} is the multiplier estimate computed from (3.10). The motivation is based on a simple observation that, for example, if $x_{0,i}$ eventually becomes active, the corresponding dual variable $\lambda_{0,i}$ will eventually become greater than or equal to $x_{0,i}$; otherwise, $\lambda_{0,i}$ will eventually become smaller than $x_{0,i}$ if strict complementarity holds. In Algorithm 3.1, the working set $\mathcal{I}^k$ is partitioned into two sets: $\mathcal{I}_-^k$ for those indices i such that $\lambda_{0,i}^{k+1} < x_{0,i}^k$ and $\mathcal{I}_+^k$ for those indices i such that $\lambda_{0,i}^{k+1} \geq x_{0,i}^k$. If $i \in \mathcal{I}_+^k$, the algorithm for the moment recognizes $x_{0,i}$ as likely to be an active variable and sets $\triangle x_{0,i}^{k,j} = -x_{0,i}^k$ (since $\triangle x_{0,i}^{k,j} = \varpi_{0,i}^{k,j}$ by (3.16) and (3.6) and $\varpi_{0,i}^{k,j} = -x_{0,i}^k$ by (3.15)); otherwise, $x_{0,i}$ is considered likely to be a basic variable and $\triangle x_{0,i}^{k,j} = -\lambda_{0,i}^{k+1}$. Note that setting $\triangle x_{0,i}^{k,j} = -\lambda_{0,i}^{k+1}$ is not only important to our proof in Lemma 3.2 that $\triangle x^{k,j}$ is a descent direction, but it allows a full step to be taken, since $x_{0,i}^k + \triangle x_{0,i}^{k,j} = x_{0,i}^k - \lambda_{0,i}^{k+1} \geq 0$. Also by setting $x_{0,i}$ in this manner, it may leave the working set in the next iteration if $\lambda_{0,i}^{k+1}$ is very negative.

The situation becomes more complicated when handling the working set for the complementarity constraints. Those variables in the working set that are outside of the bi-working set are always considered as likely to be active variables; hence, Newton equations yield $\triangle x_{l,i}^{k,j} = -x_{l,i}^k$ $(l = 1,2)$ (see the third and fourth equations in (3.15)). For the bi-working set $\mathcal{A}_{12}^{k,j}$, the underlying idea is the same as that described above for the nonnegativity constraints except that several cases are needed when both variables $x_{1,i}$ and $x_{2,i}$ in the bi-working set are considered likely to be basic variables (i.e., $\lambda^k_{1,i} < x^k_{1,i}$ and $\lambda^k_{2,i} < x^k_{2,i}$). Specifically, we cannot allow both $x_{1,i}$ and $x_{2,i}$ to become basic variables since Algorithm 3.1 maintains the feasibility of every iterate. To ensure this, Algorithm 3.1 further compares the values of $\lambda^{k+1}_{1,i}$ and $\lambda^{k+1}_{2,i}$ and defines the sets $\mathcal{B}^{k,j}_3$, $\mathcal{B}^{k,j}_4$, $\mathcal{B}^{k,j}_7$ and $\mathcal{B}^{k,j}_8$ in (3.14). If $\lambda^{k+1}_{1,i}$ is less than $\lambda^{k+1}_{2,i}$, $x_{2,i}$ is considered more likely to be an active variable than $x_{1,i}$; otherwise, $x_{1,i}$ is chosen. The values for $\varpi^{k,j}_{1,i}$ or $\varpi^{k,j}_{2,i}$ are set in the same manner as that for $\varpi^{k,j}_{0,i}$ except for two cases: $i \in \mathcal{B}^{k,j}_4 \cup \mathcal{B}^{k,j}_5$ or $i \in \mathcal{B}^{k,j}_8 \cup \mathcal{B}^{k,j}_9$. In these cases, the variables that are considered likely to be active are greater than zero. To guarantee the feasibility of the next iterate, we need to ensure that these variables are driven to zero. To this end, given a step size $\alpha_{k,j}$, Algorithm 3.1 computes a step $\triangle x^{k,j}$ for which $\triangle x^{k,j}_{2,i} = -x^k_{2,i}/\alpha_{k,j}$ for $i \in \mathcal{B}^{k,j}_4 \cup \mathcal{B}^{k,j}_5$ and $\triangle x^{k,j}_{1,i} = -x^k_{1,i}/\alpha_{k,j}$ for $i \in \mathcal{B}^{k,j}_8 \cup \mathcal{B}^{k,j}_9$ (see (3.15)). Hence, if $x^{k+1} = x^k + \alpha_{k,j}\triangle x^{k,j}$ is chosen as the next iterate, $x^{k+1}_{2,i}$ ($i \in \mathcal{B}^{k,j}_4 \cup \mathcal{B}^{k,j}_5$) and $x^{k+1}_{1,i}$ ($i \in \mathcal{B}^{k,j}_8 \cup \mathcal{B}^{k,j}_9$) are all zero. The other (complementarity) variables in these sets are considered likely to be basic variables, and we set $\triangle x^{k,j}_{1,i} = -\lambda^{k+1}_{1,i} + x^k_{2,i}/\alpha_{k,j}$ for $i \in \mathcal{B}^{k,j}_4 \cup \mathcal{B}^{k,j}_5$ and $\triangle x^{k,j}_{2,i} = -\lambda^{k+1}_{2,i} + x^k_{1,i}/\alpha_{k,j}$ for $i \in \mathcal{B}^{k,j}_8 \cup \mathcal{B}^{k,j}_9$. This, as will be shown later, guarantees that $\triangle x^{k,j}$ is a descent direction for the objective function. Since Algorithm 3.1 maintains $0 \le x^k_1 \perp x^k_2 \ge 0$ for every k, it follows that $x^k_{1,i} = 0$ (respectively, $x^k_{2,i} = 0$) if for some j, $i \in \mathcal{A}^k_1 \backslash \mathcal{A}^{k,j}_{12}$ (respectively, $i \in \mathcal{A}^k_2 \backslash \mathcal{A}^{k,j}_{12}$). Finally, we note that if $x^k_{1,i} = 0$ and $x^k_{2,i} = 0$, the index i could belong to two of the sets $\mathcal{B}^{k,j}_l$, $l = 2, ..., 9$. In this case we can simply choose to put i into either of the two sets. This does not interfere with our analysis.

Remark 3.5. The line search procedure described in Step 3 is a variant of the traditional backtracking line search using the Armijo condition (3.18). In particular, as we have seen from Remark 3.4, the step direction $\triangle x^{k,j}$ varies in a specified way as the step size $\alpha_{k,j}$ changes. This guarantees the feasibility of the generated trial iterate with respect to the complementarity constraints. If the trial point is feasible, i.e., (3.17) holds, and satisfies the Armijo condition (3.18), it is accepted as the next iterate. Otherwise, we decrease the working set parameter $\varepsilon_{k,j}$ by a factor and update the step size using (3.13). Decreasing $\varepsilon_{k,j}$ avoids the situation that the Armijo condition is rejected due to an overestimation of the active set. Consequently, $\mathcal{A}^{k,j}_1$ and $\mathcal{A}^{k,j}_2$ are, respectively, subsets of $\mathcal{A}^k_1$ and $\mathcal{A}^k_2$ for all j. Those variables outside $\mathcal{A}^{k,j}_1$ or $\mathcal{A}^{k,j}_2$ but in $\mathcal{A}^k_1$ or $\mathcal{A}^k_2$ are handled in the same way as are the inequality variables $x_{0,i}$ that belong to $\mathcal{I}^k$ (see the treatment related to $\mathcal{B}^{k,j}_{10}$-$\mathcal{B}^{k,j}_{13}$ in (3.15) and (3.14)). In Algorithm 3.1, the step size always starts from one and the update formula (3.13) tries to shorten a previous step size by a factor β unless the resulting step size is bigger than certain thresholds. These thresholds ensure that the vector $\varpi^{k,j}$ defined by (3.15) is bounded, and hence that the step direction is bounded. Finally, the feasibility of the iterate with respect to the equality constraints follows trivially since x^0 is feasible and $A\triangle x^{k,j}$ always equals zero for any k and j.

To simply notation, let

$$
\begin{array}{rcl}
E^k_0 & = & E_0(x^k, y^k, z^k; \varepsilon_{k,0}), \\
E^k_1 & = & E_1(x^k, y^k, z^k; \varepsilon_{k,0}), \\
E^k_2 & = & E_2(x^k, y^k, z^k; \varepsilon_{k,0}), \\
E^k & = & [E^k_0, E^k_1, E^k_2].
\end{array} \tag{3.20}
$$

In the remainder of this section, we show that Algorithm 3.1 is well-defined. To this end, we need the following assumptions.

A1. *Problem (1.6) is feasible.*

A2. *The MPLCC-LICQ holds on the feasible region of problem (1.6).*

As we have discussed in Remarks 3.4 and 3.5, our line search procedure always maintains the feasibility of every iterate as long as the starting point x^0 is feasible. Moreover, from Remark 3.2, under the MPLCC-LICQ, the constraint gradients involved in the Newton matrix $\mathcal{M}^k$ are linearly independent. Since condition (3.7) holds, the next lemma is easy to prove using Sylvester's law of inertia and is an immediate consequence of Proposition 2 in [11].

LEMMA 3.1. *Under Assumptions A1 and A2, every iterate generated by Algorithm 3.1 is feasible and the Newton matrix $\mathcal{M}^k$ in Step 2.1 of Algorithm 3.1 is nonsingular on each iteration.*

This shows that the proposed step directions for each iteration are all computable. The next lemma shows that these directions are non-ascent directions for the objective function $f(x)$.

LEMMA 3.2. *Under Assumptions A1 and A2, the directions $\triangle x^k$ and $\triangle x^{k,j}$ $(j = 0, 1, ...)$ satisfy*

$$\nabla f(x^k)^\top \triangle x^k = -(\triangle x^k)^\top \mathcal{H}^k \triangle x^k \le 0, \tag{3.21}$$

$$\nabla f(x^k)^\top \triangle x^{k,j} = \nabla f(x^k)^\top \triangle x^k + \left(\bar\lambda^{k+1}\right)^\top \varpi^{k,j} \le \nabla f(x^k)^\top \triangle x^k,\ j = 0, 1, ... \tag{3.22}$$

Proof. Let $\xi^{k+1} = \left(\left(y^{k+1}\right)^\top, -\left(\bar\lambda^{k+1}\right)^\top\right)^\top$, $\xi^{k+1,j} = \left(\left(y^{k+1,j}\right)^\top, -\left(\bar\lambda^{k+1,j}\right)^\top\right)^\top$, $\bar A^k = [A^\top, E^k]$, and $v^{k,j} = \left(0^\top, \left(\varpi^{k,j}\right)^\top\right)^\top$ with $0 \in \Re^p$. The linear systems (3.10) and (3.16) can be written as

$$\begin{bmatrix} \mathcal{H}^k & \bar A^k \\ \left(\bar A^k\right)^\top & 0 \end{bmatrix} \begin{bmatrix} \triangle x^k \\ \xi^{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \mathcal{H}^k & \bar A^k \\ \left(\bar A^k\right)^\top & 0 \end{bmatrix} \begin{bmatrix} \triangle x^{k,j} \\ \xi^{k+1,j} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ v^{k,j} \end{bmatrix} \tag{3.23}$$

Premultiplying the first block of equations in the first system in (3.23) by $\left(\triangle x^k\right)^\top$ and noting the second block of equations and condition (3.7) yields (3.21). Since $\mathcal{M}^k$ is nonsingular, we can write

$$\left(\mathcal{M}^k\right)^{-1} = \begin{bmatrix} \mathcal{H}^k & \bar A^k \\ \left(\bar A^k\right)^\top & 0 \end{bmatrix}^{-1} \equiv \begin{bmatrix} B & C \\ C^\top & D \end{bmatrix}, \quad \begin{bmatrix} \triangle x^k \\ \xi^{k+1} \end{bmatrix} = \begin{bmatrix} -B\nabla f(x^k) \\ -C\nabla f(x^k) \end{bmatrix}$$

and

$$\triangle x^{k,j} = -B\nabla f(x^k) + Cv^{k,j} = \triangle x^k + Cv^{k,j}.$$

Hence,

$$\begin{aligned} \nabla f(x^k)^\top \triangle x^{k,j} &= \nabla f(x^k)^\top \triangle x^k + \nabla f(x^k)^\top C v^{k,j} \\ &= \nabla f(x^k)^\top \triangle x^k - (\xi^{k+1})^\top v^{k,j}, \end{aligned} \tag{3.24}$$

which implies the equality in (3.22).

Moreover, according to the definition (3.15) of $\varpi^{k,j}$, we have from (3.24) that

$$\begin{aligned} -(\xi^{k+1})^\top v^{k,j} = (\bar\lambda^{k+1})^\top \varpi^{k,j} \quad = \quad & -\textstyle\sum_{i\in\mathcal{I}_+^k} \lambda_{0,i}^{k+1} x_{0,i}^k - \sum_{i\in\mathcal{I}_-^k} (\lambda_{0,i}^{k+1})^2 \\ & -\textstyle\sum_{i\in\mathcal{B}_1^{k,j}} \left(\lambda_{1,i}^{k+1} x_{1,i}^k + \lambda_{2,i}^{k+1} x_{2,i}^k\right) \\ & -\textstyle\sum_{i\in\mathcal{B}_2^{k,j}\cup\mathcal{B}_3^{k,j}} \left(\lambda_{2,i}^{k+1}\right)^2 \\ & -\textstyle\sum_{i\in\mathcal{B}_4^{k,j}\cup\mathcal{B}_5^{k,j}} \left(\left(\lambda_{1,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{2,i}^{k+1} - \lambda_{1,i}^{k+1}\right) x_{2,i}^k\right) \\ & -\textstyle\sum_{i\in\mathcal{B}_6^{k,j}\cup\mathcal{B}_7^{k,j}} \left(\lambda_{1,i}^{k+1}\right)^2 \\ & -\textstyle\sum_{i\in\mathcal{B}_8^{k,j}\cup\mathcal{B}_9^{k,j}} \left(\left(\lambda_{2,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{1,i}^{k+1} - \lambda_{2,i}^{k+1}\right) x_{1,i}^k\right) \\ & -\textstyle\sum_{i\in\mathcal{B}_{10}^{k,j}} \lambda_{1,i}^{k+1} x_{1,i}^k - \sum_{i\in\mathcal{B}_{11}^{k,j}} \left(\lambda_{1,i}^{k+1}\right)^2 \\ & -\textstyle\sum_{i\in\mathcal{B}_{12}^{k,j}} \lambda_{2,i}^{k+1} x_{2,i}^k - \sum_{i\in\mathcal{B}_{13}^{k,j}} \left(\lambda_{2,i}^{k+1}\right)^2. \end{aligned} \tag{3.25}$$

From the definitions of $\mathcal{I}_+^k$, $\mathcal{I}_-^k$ and $\mathcal{B}_l^{k,j}$ $(l = 1, ..., 13)$ given by (3.12) and (3.14), it follows that

$$
\left\{\begin{array}{ll}
\lambda_{0,i}^{k+1} x_{0,i}^k \geq (x_{0,i}^k)^2, & \forall i \in \mathcal{I}_+^k,\\
\lambda_{1,i}^{k+1} x_{1,i}^k \geq (x_{1,i}^k)^2 \text{ and } \lambda_{2,i}^{k+1} x_{2,i}^k \geq (x_{2,i}^k)^2, & \forall i \in \mathcal{B}_1^{k,j},\\
\lambda_{2,i}^{k+1} > \lambda_{1,i}^{k+1}, & \forall i \in \mathcal{B}_4^{k,j},\\
\lambda_{2,i}^{k+1} \geq x_{2,i}^k \geq x_{1,i}^k (= 0) > \lambda_{1,i}^{k+1}, & \forall i \in \mathcal{B}_5^{k,j},\\
\lambda_{1,i}^{k+1} > \lambda_{2,i}^{k+1}, & \forall i \in \mathcal{B}_8^{k,j},\\
\lambda_{1,i}^{k+1} \geq x_{1,i}^k \geq x_{2,i}^k (= 0) > \lambda_{2,i}^{k+1}, & \forall i \in \mathcal{B}_9^{k,j},\\
\lambda_{1,i}^{k+1} x_{1,i}^k \geq (x_{1,i}^k)^2, & \forall i \in \mathcal{B}_{10}^{k,j},\\
\lambda_{2,i}^{k+1} x_{2,i}^k \geq (x_{2,i}^k)^2, & \forall i \in \mathcal{B}_{12}^{k,j}.
\end{array}\right. \tag{3.26}
$$

Hence, every summation term in the right hand side of (3.25) is nonnegative and thus, it follows that $\nabla f(x^k)^\top \triangle x^{k,j} \leq \nabla f(x^k)^\top \triangle x^k \leq 0$. ∎

LEMMA 3.3. *Under Assumptions A1 and A2, $\triangle x^{k,j}$ is a descent direction for $f(x)$ at x^k; i.e., if Algorithm 3.1 does not terminate at iteration k, $\nabla f(x^k)^\top \triangle x^k < 0$ and $\nabla f(x^k)^\top \triangle x^{k,j} < 0$ for every j and k. Moreover, the line search step in Algorithm 3.1 is well-defined, i.e., a step size satisfying (3.17) and (3.18) can always be found in Step 3.*

Proof. Suppose $\nabla f(x^k)^\top \triangle x^{k,j} = 0$ for some iteration k and j. By (3.26) every summation term in the right hand side of (3.25) is nonnegative. Since $\triangle x^k \in \mathcal{S}^k$ by (3.10), $\triangle x^k$ satisfies condition (3.7). Consequently, we obtain from (3.21), (3.25) and (3.14) that:

$$(a1): \nabla f(x^k)^\top \triangle x^k = 0 \Rightarrow (\triangle x^k)^\top \mathcal{H}^k \triangle x^k = 0 \Rightarrow \triangle x^k = 0;$$

$$(a2): \sum_{i\in\mathcal{I}_+^k} \lambda_{0,i}^{k+1} x_{0,i}^k = 0 \text{ and } \lambda_{0,i}^{k+1} \geq x_{0,i}^k,\ \forall i \in \mathcal{I}_+^k \Rightarrow \lambda_{0,i}^{k+1} \geq 0,\ x_{0,i}^k = 0,\ \forall i \in \mathcal{I}_+^k;$$

$$(a3): \sum_{i\in\mathcal{I}_-^k} (\lambda_{0,i}^{k+1})^2 = 0 \Rightarrow \lambda_{0,i}^{k+1} = 0,\ \forall i \in \mathcal{I}_-^k;$$

$$(a4): \sum_{i\in\mathcal{B}_1^{k,j}} \left(\lambda_{1,i}^{k+1} x_{1,i}^k + \lambda_{2,i}^{k+1} x_{2,i}^k\right) = 0 \text{ and } \lambda_{1,i}^{k+1} \geq x_{1,i}^k,\ \lambda_{2,i}^{k+1} \geq x_{2,i}^k,\ \forall i \in \mathcal{B}_1^{k,j}$$
$$\Rightarrow \lambda_{1,i}^{k+1} \geq 0,\ x_{1,i}^k = 0,\ \lambda_{2,i}^{k+1} \geq 0,\ x_{2,i}^k = 0,\ \forall i \in \mathcal{B}_1^{k,j};$$

$$(a5): \sum_{i\in\mathcal{B}_2^{k,j}} \left(\lambda_{2,i}^{k+1}\right)^2 = 0 \text{ and } \lambda_{1,i}^{k+1} \geq x_{1,i}^k,\ \forall i \in \mathcal{B}_2^{k,j}$$
$$\Rightarrow \lambda_{2,i}^{k+1} = 0,\ \lambda_{1,i}^{k+1} \geq 0,\ x_{1,i}^k = 0,\ \forall i \in \mathcal{B}_2^{k,j};$$

$$(a6): \sum_{i\in\mathcal{B}_3^{k,j}} \left(\lambda_{2,i}^{k+1}\right)^2 = 0 \text{ and } \lambda_{2,i}^{k+1} \leq \lambda_{1,i}^{k+1} < x_{1,i}^k = 0,\ \forall i \in \mathcal{B}_3^{k,j} \Rightarrow \mathcal{B}_3^{k,j} = \emptyset;$$

$$(a7): \sum_{i\in\mathcal{B}_4^{k,j}} \left(\left(\lambda_{1,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{2,i}^{k+1} - \lambda_{1,i}^{k+1}\right) x_{2,i}^k\right) = 0 \text{ and } \lambda_{1,i}^{k+1} < \lambda_{2,i}^{k+1} < x_{2,i}^k,\ \forall i \in \mathcal{B}_4^{k,j}$$
$$\Rightarrow \lambda_{1,i}^{k+1} = 0,\ \lambda_{2,i}^{k+1} = 0,\ \forall i \in \mathcal{B}_4^{k,j};$$

$$(a8): \sum_{i\in\mathcal{B}_5^{k,j}} \left(\left(\lambda_{1,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{2,i}^{k+1} - \lambda_{1,i}^{k+1}\right) x_{2,i}^k\right) = 0 \text{ and}$$
$$\lambda_{1,i}^{k+1} < x_{1,i}^k = 0 \leq x_{2,i}^k \leq \lambda_{2,i}^{k+1},\ \forall i \in \mathcal{B}_5^{k,j} \Rightarrow \mathcal{B}_5^{k,j} = \emptyset;$$

$$(a9): \sum_{i\in\mathcal{B}_6^{k,j}} \left(\lambda_{1,i}^{k+1}\right)^2 = 0 \text{ and } \lambda_{2,i}^{k+1} \geq x_{2,i}^k,\ \forall i \in \mathcal{B}_6^{k,j}$$

$$\Rightarrow \lambda_{1,i}^{k+1} = 0,\ \lambda_{2,i}^{k+1} \geq 0,\ x_{2,i}^k = 0,\ \forall i \in \mathcal{B}_6^{k,j};$$

$$(a10): \sum_{i\in\mathcal{B}_7^{k,j}} \left(\lambda_{1,i}^{k+1}\right)^2 = 0 \text{ and } \lambda_{1,i}^{k+1} \leq \lambda_{2,i}^{k+1} < x_{2,i}^k = 0,\ \forall i \in \mathcal{B}_7^{k,j} \Rightarrow \mathcal{B}_7^{k,j} = \emptyset;$$

$$(a11): \sum_{i\in\mathcal{B}_8^{k,j}} \left(\left(\lambda_{2,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{1,i}^{k+1} - \lambda_{2,i}^{k+1}\right) x_{1,i}^k\right) = 0 \text{ and } \lambda_{2,i}^{k+1} < \lambda_{1,i}^{k+1} < x_{1,i}^k,\ \forall i \in \mathcal{B}_8^{k,j}$$

$$\Rightarrow \lambda_{2,i}^{k+1} = 0,\ \lambda_{1,i}^{k+1} = 0,\ \forall i \in \mathcal{B}_8^{k,j};$$

$$(a12): \sum_{i\in\mathcal{B}_9^{k,j}} \left(\left(\lambda_{2,i}^{k+1}\right)^2 + \frac{1}{\alpha_{k,j}}\left(\lambda_{1,i}^{k+1} - \lambda_{2,i}^{k+1}\right) x_{1,i}^k\right) = 0 \text{ and}$$

$$\lambda_{2,i}^{k+1} < x_{2,i}^k = 0 \leq x_{1,i}^k \leq \lambda_{1,i}^{k+1},\ \forall i \in \mathcal{B}_9^{k,j} \Rightarrow \mathcal{B}_9^{k,j} = \emptyset;$$

$$(a13): \sum_{i\in\mathcal{B}_{10}^{k,j}} \lambda_{1,i}^{k+1} x_{1,i}^k = 0 \text{ and } x_{1,i}^k \leq \lambda_{1,i}^{k+1},\ \forall i \in \mathcal{B}_{10}^{k,j} \Rightarrow \lambda_{1,i}^{k+1} \geq 0,\ x_{1,i}^k = 0,\ \forall i \in \mathcal{B}_{10}^{k,j};$$

$$(a14): \sum_{i\in\mathcal{B}_{11}^{k,j}} \left(\lambda_{1,i}^{k+1}\right)^2 = 0 \Rightarrow \lambda_{1,i}^{k+1} = 0,\ \forall i \in \mathcal{B}_{11}^{k,j};$$

$$(a15): \sum_{i\in\mathcal{B}_{12}^{k,j}} \lambda_{2,i}^{k+1} x_{2,i}^k = 0 \text{ and } x_{2,i}^k \leq \lambda_{2,i}^{k+1},\ \forall i \in \mathcal{B}_{12}^{k,j} \Rightarrow \lambda_{2,i}^{k+1} \geq 0,\ x_{2,i}^k = 0,\ \forall i \in \mathcal{B}_{12}^{k,j};$$

$$(a16): \sum_{i\in\mathcal{B}_{13}^{k,j}} \left(\lambda_{2,i}^{k+1}\right)^2 = 0 \Rightarrow \lambda_{2,i}^{k+1} = 0,\ \forall i \in \mathcal{B}_{13}^{k,j}.$$

Note that from (a13) and (a15) we know $\mathcal{B}_{10}^{k,j} = \mathcal{B}_{12}^{k,j} = \emptyset$ since otherwise $\mathcal{B}_{10}^{k,j} \subseteq \mathcal{A}_1(x^k) \subseteq \mathcal{A}_1^{k,j}$ and $\mathcal{B}_{12}^{k,j} \subseteq \mathcal{A}_2(x^k) \subseteq \mathcal{A}_2^{k,j}$ contradicting the definitions of $\mathcal{B}_{10}^{k,j}$ and $\mathcal{B}_{12}^{k,j}$, respectively. Moreover, since $\mathcal{A}_1(x^k) \subseteq \mathcal{A}_1^{k,j}$ and $\mathcal{A}_2(x^k) \subseteq \mathcal{A}_2^{k,j}$, we have $x_{2,i}^k > x_{1,i}^k = 0$ for $i \in \mathcal{A}_1^{k,j} \backslash \mathcal{A}_{12}^{k,j}$ and $x_{1,i}^k > x_{2,i}^k = 0$ for $i \in \mathcal{A}_2^{k,j} \backslash \mathcal{A}_{12}^{k,j}$. On the other hand, all the multipliers in λ^{k+1} for which the corresponding primal iterates are not in the working set are set to zero according to Step 2.1 of Algorithm 3.1.

Summarizing all the results above, we obtain that

$$\lambda_0^{k+1} \geq 0;\ \left(\lambda_0^{k+1}\right)^\top x_0^k = 0;\ \lambda_{1,i}^{k+1}, \lambda_{2,i}^{k+1} \geq 0,\ \forall i \in \mathcal{A}_{12}^{k,j};\ \lambda_{1,i}^{k+1} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_1^{k,j};$$
$$\lambda_{2,i}^{k+1} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_2^{k,j};\ \lambda_{1,i}^{k+1} x_{1,i}^k = 0,\ \forall i \in \mathcal{A}_1^{k,j};\ \lambda_{2,i}^{k+1} x_{2,i}^k = 0,\ \forall i \in \mathcal{A}_2^{k,j}.$$

In view of the fact that $\mathcal{A}_1(x^k) \subseteq \mathcal{A}_1^{k,j}$ and $\mathcal{A}_2(x^k) \subseteq \mathcal{A}_2^{k,j}$, we further obtain that

$$\lambda_{1,i}^{k+1}, \lambda_{2,i}^{k+1} \geq 0,\ \forall i \in \mathcal{A}_{12}(x^k);\ \lambda_{1,i}^{k+1} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_1(x^k);\ \lambda_{2,i}^{k+1} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_2(x^k).$$

Hence, substituting $\triangle x^k = 0$ into the first equation of (3.10), we conclude from Definition 2.4 that x^k is a S-stationary point of problem (1.6) with the unique multiplier vector (y^{k+1}, λ^{k+1}). This means that Algorithm 3.1 should have been terminated in Step 2.2. Therefore, we always have $\nabla f(x^k)^\top \triangle x^{k,j} < 0$ for every j if the algorithm does not stop at Step 2.2 on iteration k.

Now we consider the second part of the lemma. Suppose at some iteration k, a step size satisfying (3.17) and (3.18) cannot be found. Thus, by Step 3.4 of Algorithm 3.1, parameters $\varepsilon_{k,j}$ are eventually driven to zero as j increases. As a consequence, the working sets $\mathcal{A}_1^{k,j}$ and $\mathcal{A}_2^{k,j}$ eventually become the active sets $\mathcal{A}_1(x^k)$ and $\mathcal{A}_2(x^k)$, respectively. Therefore, the sets defined in (3.14) and the vectors $\varpi^{k,j}$ given by (3.15) are eventually fixed. This means that the step directions $\triangle x^{k,j}$ are fixed for all large enough j. Let $\triangle x^{k,j} = \triangle \bar{x}^k$ for such j. Clearly, it follows from (3.15) and (3.14) that $\triangle \bar{x}_{0,i}^k \geq 0,\ \forall i \in \mathcal{I}(x^k)$, $\triangle \bar{x}_{1,i}^k \geq 0,\ \forall i \in \mathcal{A}_1(x^k)$ and $\triangle \bar{x}_{2,i}^k \geq 0,\ \forall i \in \mathcal{A}_2(x^k)$. Hence, there is a $\bar{\alpha} > 0$

such that $x^k + \alpha \triangle \bar{x}^k \geq 0$ for all $\alpha \in (0, \bar{\alpha}]$. Moreover, since $\nabla f(x^k)^\top \triangle \bar{x}^k < 0$ as shown above and $f(x)$ is continuously differentiable, there is a $\hat{\alpha} > 0$ with $\hat{\alpha} \leq \bar{\alpha}$ such that for any $\alpha \in (0, \hat{\alpha}]$,

$$f(x^k + \alpha \triangle \bar{x}^k) \leq f(x^k) + \sigma \alpha \nabla f(x^k)^\top \triangle \bar{x}^k.$$

Hence, violation of (3.17) and (3.18) implies that $\alpha_{k,j} > \hat{\alpha}$ for all j large enough. On the other hand, since $\mathcal{A}_1^{k,j} = \mathcal{A}_1(x^k)$ and $\mathcal{A}_2^{k,j} = \mathcal{A}_2(x^k)$ eventually, we know from (3.13) and the definition (3.11) of $\gamma_k^{\max}$ that $\alpha_{k,j} = \beta \alpha_{k,j-1}$ for all large enough j. This implies that the step sizes $\alpha_{k,j}$ tend to zero in the limit. Hence, we get a contradiction. ☐

So far we have established that under Assumptions A1 and A2, at every iteration Algorithm 3.1 either terminates at Step 2.2 with a S-stationary point to problem (1.6), or readily generates the next iterate and continues on to the next iteration. In particular, we have shown that the Newton matrix $\mathcal{M}^k$ is always non-singular, all step directions are computable and the line search is well-defined. To conclude, we have

PROPOSITION 3.4. *Under Assumptions A1 and A2, Algorithm 3.1 is well defined.*

4. Global convergence to B-stationarity. In this section we show that the iterates generated by Algorithm 3.1 converge to a S-stationary solution, and hence, under the MPLCC-LICQ, to a B-stationary solution of problem (1.6). In particular, we only consider the case that Algorithm 3.1 generates an infinite sequence of iterates as otherwise a B-stationary point is found in a finite number of steps. The next two assumptions are used in our analysis.

A3. *The level set $\{x \in \Re^{n+2m} | f(x) \leq f(x^0)\}$ is bounded on the feasible region of problem (1.6).*

A4. *The sequence $\{\mathcal{H}^k\}$ of Hessian estimates is bounded and there exists a $\mu > 0$ such that for all k, $d^\top \mathcal{H}^k d \geq \mu \|d\|^2$, $\forall d \in \mathcal{S}^k$.*

Since Algorithm 3.1 is a descent algorithm and $f(x)$ is continuously differentiable, Assumption A3 guarantees that the generated iterates are bounded and hence an accumulation point exists. Assumption A4, which can be ensured by careful modifications to the exact Hessian, ensures that the sequence $\{\mathcal{H}^k\}$ is uniformly positive definite on the null space defined by $\mathcal{S}^k$. This will be used to show that the Newton system is nonsingular in the limit.

LEMMA 4.1. *Under Assumptions A1, A2 and A3, there is a $\bar{\varepsilon} > 0$ such that $\varepsilon_{k,0} = \varepsilon_k = \bar{\varepsilon}$ for all large enough k.*

Proof. Suppose $\{\varepsilon_k\}$ tends to zero as k increases. Then by Step 1.1 of Algorithm 3.1, there is an infinite index set $\mathcal{K}$ on which the matrix $[A^\top, E(x^k, y^k, z^k; \varepsilon_k)]$ does not have full column rank. Since the iterates x^k are bounded by Assumption A3 and $\mathcal{I}$ and A are finite sets, there is an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{x^k\}_{\bar{\mathcal{K}}} \to \bar{x}$ and the sets $\mathcal{I}^k$, $\mathcal{A}_1^k$ and $\mathcal{A}_2^k$ defined by (3.9) are fixed for all $k \in \bar{\mathcal{K}}$. Suppose they are $\hat{\mathcal{I}}$, $\hat{\mathcal{A}}_1$ and $\hat{\mathcal{A}}_2$, respectively. Since x^k are all feasible, $\bar{x}$ is also feasible. Since $\{\varepsilon_k\} \to 0$, it follows from (3.1) that $\hat{\mathcal{I}} \subseteq \mathcal{I}(\bar{x})$, $\hat{\mathcal{A}}_1 \subseteq \mathcal{A}_1(\bar{x})$ and $\hat{\mathcal{A}}_2 \subseteq \mathcal{A}_2(\bar{x})$. Hence, we have found a feasible point $\bar{x}$ for which the gradients of equality constraints and a set of active inequality and complementarity constraints respectively in $\hat{\mathcal{I}}$ and $\hat{\mathcal{A}}_1$ and $\hat{\mathcal{A}}_2$ are linearly dependent. This contradicts Assumption A2. Thus, $\varepsilon_k = \bar{\varepsilon}$ for some $\bar{\varepsilon} > 0$ and all k large enough. This implies that matrix $[A^\top, E^k]$ eventually has full column rank by Step 1.1 of Algorithm 3.1, where E^k is defined in (3.20). Hence, $\varepsilon_{k,0} = \varepsilon_k = \bar{\varepsilon}$ for all k large enough. ☐

LEMMA 4.2. *Under Assumptions A1-A4, the sequences $\{x^k\}$, $\{\|(\mathcal{M}^k)^{-1}\|\}$, $\{\triangle x^k, y^{k+1}, \lambda^{k+1}\}$, $\{z^k\}$ and $\{\triangle x^{k,j}, y^{k+1,j}, \lambda^{k+1,j}\}$ for all j are bounded.*

Proof. Suppose there exists an infinite index set $\mathcal{K}$ such that the sequence $\{\|(\mathcal{M}^k)^{-1}\|\}$ tends to infinity as $k \in \mathcal{K}$ tends to infinity. Since the iterates x^k are bounded by Assumption A3 and the sets $\mathcal{I}$ and $\mathcal{A}$ are finite, there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{x^k\}_{\bar{\mathcal{K}}} \to \bar{x}$ and the working sets $\mathcal{I}^k$, $\mathcal{A}_1^k$ and $\mathcal{A}_2^k$ defined by (3.9) are fixed for all $k \in \bar{\mathcal{K}}$. Suppose they are $\hat{\mathcal{I}}$, $\hat{\mathcal{A}}_1$ and $\hat{\mathcal{A}}_2$, respectively. Moreover, the Jacobian matrix E^k defined in (3.20) and the set $\mathcal{S}^k$ are also fixed for all $k \in \bar{\mathcal{K}}$. Denote them by $\bar{E}$ and $\bar{\mathcal{S}}$, respectively. By Assumption A4, there is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\mathcal{H}^k \to \bar{\mathcal{H}}$ as $k \in \hat{\mathcal{K}} \to \infty$. Hence,

$$\{\mathcal{M}^k\}_{\hat{\mathcal{K}}} \to \bar{\mathcal{M}} = \begin{bmatrix} \bar{\mathcal{H}} & A^\top & -\bar{E} \\ A & 0 & 0 \\ \bar{E} & 0 & 0 \end{bmatrix}. \tag{4.1}$$

By Assumption A4, we have $d^\top \bar{\mathcal{H}} d \geq \mu \|d\|^2$ for all $d \in \bar{S}$, i.e., for all $d \in \Re^{n+2m}$ such that $Ad = 0$ and $\bar{E}^\top d = 0$. Since the columns of $[A^\top, \bar{E}]$ are independent by arguments similar to those used in the proof of Lemma 3.1, we know that $\bar{\mathcal{M}}$ is nonsingular. Hence, it follows that $\{(\mathcal{M}^k)^{-1}\}_{\hat{\mathcal{K}}} \to \bar{\mathcal{M}}^{-1}$. This contradicts the assumption that $\{\|(\mathcal{M}^k)^{-1}\|\}_{\mathcal{K}} \to \infty$. Hence, the sequence $\{\mathcal{M}^k\}$ is uniformly nonsingular.

The boundedness of $\{x^k\}$ implies the right hand side of (3.10) is bounded. Hence, the solution sequence $\{\triangle x^k, y^{k+1}, \lambda^{k+1}\}$ is bounded. Moreover, we have from (3.13) that for any j, $x^k_{1,i}/\alpha_{k,j} \leq \gamma_k^{\max}$, $\forall i \in \mathcal{A}_1^{k,j}$ and $x^k_{2,i}/\alpha_{k,j} \leq \gamma_k^{\max}$, $\forall i \in \mathcal{A}_2^{k,j}$, where $\{\gamma_k^{\max}\}$ is bounded because of the boundedness of $\{x^k\}$ and (3.11). Hence, we know from (3.15) that $\{\varpi^{k,j}\}$ is bounded and thus the right hand side of (3.16) is bounded for all j. This together with the uniform non-singularity of $\{\mathcal{M}^k\}$ imply that the sequence $\{\triangle x^{k,j}, y^{k+1,j}, \lambda^{k+1,j}\}$ is bounded for all j.

By the second equation in (3.19), $\{z_3^k\}$ is bounded above by $z_3^{\max} > 0$. Hence, we have from the boundedness of $\{\lambda^k\}$ and $\{x^k\}$ and (3.19) that $\{z^k\}$ is bounded as well. ☐

Lemma 4.3. *Suppose Assumptions A1-A4 hold. If there is an infinite index set $\mathcal{K}$ associated with a sequence of indexes $\{\bar{j}_k\}_{\mathcal{K}}$ such that $\min\{\|\mathcal{R}(x^k, y^k, z^k)\|^\eta, \varepsilon_{k,\bar{j}_k}\} \geq \tilde{\varepsilon}$ for some $\tilde{\varepsilon} > 0$ and all $k \in \mathcal{K}$ and $\{\nabla f(x^k)^\top \triangle x^{k,\bar{j}_k}\} \to 0$ as $k \in \mathcal{K} \to \infty$, then any limit point of the sequence $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\mathcal{K}}$ satisfies the S-stationarity conditions for problem (1.6).*

Proof. First note that $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\mathcal{K}}$ is bounded by Assumption A3 and Lemma 4.2. Assume to the contrary that for some infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$, $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\bar{\mathcal{K}}} \to (\bar{x}, \bar{y}, \bar{\lambda})$, but $(\bar{x}, \bar{y}, \bar{\lambda})$ does not satisfy the S-stationarity conditions for problem (1.6). Due to the boundedness of $\{\triangle x^k\}$ by Lemma 4.2 and the boundedness $\{\mathcal{H}^k\}$ by Assumption A4, there exists an infinite index set $\tilde{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\mathcal{H}^k\} \to \bar{\mathcal{H}}$ and $\{\triangle x^k\} \to \triangle \bar{x}$ as $k \in \tilde{\mathcal{K}} \to \infty$. Since the sets $\mathcal{I}$ and $\mathcal{A}$ are finite, there exists an infinite index set $\hat{\mathcal{K}} \subseteq \tilde{\mathcal{K}}$ such that $\mathcal{I}^k$, $\mathcal{I}^k_-$, $\mathcal{I}^k_+$, $\mathcal{A}^k_1$, $\mathcal{A}^k_2$, $\mathcal{A}_1^{k,\bar{j}_k}$, $\mathcal{A}_2^{k,\bar{j}_k}$, $\mathcal{A}_{12}^{k,\bar{j}_k}$ and $\mathcal{B}_l^{k,\bar{j}_k}$ $(l = 1, ..., 13)$ are all fixed for all $k \in \hat{\mathcal{K}}$. Suppose they are $\bar{\mathcal{I}}$, $\bar{\mathcal{I}}_-$, $\bar{\mathcal{I}}_+$, $\bar{\mathcal{A}}_1$, $\bar{\mathcal{A}}_2$, $\hat{\mathcal{A}}_1$, $\hat{\mathcal{A}}_2$, $\hat{\mathcal{A}}_{12}$ and $\bar{\mathcal{B}}_l$ $(l = 1, ..., 13)$, respectively.

From (3.10) we have $\triangle x^k \in \mathcal{S}^k$. Hence, by Assumption A4 and Lemma 3.2, it follows that $\nabla f(x^k)^\top \triangle x^{k,\bar{j}_k} \leq -\mu \|\triangle x^k\|^2$ for all k. By (3.26) the summation terms in the right hand side of (3.25) are all nonnegative. Since $\{\nabla f(x^k)^\top \triangle x^{k,\bar{j}_k}\}_{\hat{\mathcal{K}}} \to 0$, similarly to the analysis in Lemma 3.3, letting $k \in \hat{\mathcal{K}} \to \infty$ in (3.25) and using the definitions of $\bar{\mathcal{B}}_l$ $(l = 1, ..., 13)$ given in (3.14) yields that:

$$\begin{aligned}
&(b1): \triangle \bar{x} = 0;\\
&(b2): \bar{\lambda}_{0,i} \geq \bar{x}_{0,i} = 0,\ \forall i \in \bar{\mathcal{I}}_+;\\
&(b3): \bar{\lambda}_{0,i} = 0,\ \forall i \in \bar{\mathcal{I}}_-;\\
&(b4): \bar{\lambda}_{1,i} \geq \bar{x}_{1,i} = 0,\ \text{and}\ \bar{\lambda}_{2,i} \geq \bar{x}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_1;\\
&(b5): \bar{\lambda}_{2,i} = 0 \text{ and } \bar{\lambda}_{1,i} \geq \bar{x}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_2;\\
&(b6): \bar{x}_{1,i} = 0 \text{ and } \bar{\lambda}_{1,i} \geq \bar{\lambda}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_3;\\
&(b7): \bar{\lambda}_{1,i} = 0 \text{ and } \bar{x}_{2,i} \geq \bar{\lambda}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_4;\\
&(b8): \bar{\lambda}_{1,i} = 0 \text{ and } \bar{\lambda}_{2,i} \geq \bar{x}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_5;\\
&(b9): \bar{\lambda}_{1,i} = 0 \text{ and } \bar{\lambda}_{2,i} \geq \bar{x}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_6;\\
&(b10): \bar{x}_{2,i} = 0 \text{ and } \bar{\lambda}_{2,i} \geq \bar{\lambda}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_7;\\
&(b11): \bar{\lambda}_{2,i} = 0 \text{ and } \bar{x}_{1,i} \geq \bar{\lambda}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_8;
\end{aligned}$$

$$(b12): \bar{\lambda}_{2,i} = 0 \text{ and } \bar{\lambda}_{1,i} \geq \bar{x}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_9;$$
$$(b13): \bar{\lambda}_{1,i} \geq \bar{x}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_{10};$$
$$(b14): \bar{x}_{1,i} \geq \bar{\lambda}_{1,i} = 0,\ \forall i \in \bar{\mathcal{B}}_{11};$$
$$(b15): \bar{\lambda}_{2,i} \geq \bar{x}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_{12};$$
$$(b16): \bar{x}_{2,i} \geq \bar{\lambda}_{2,i} = 0,\ \forall i \in \bar{\mathcal{B}}_{13}.$$

By Step 2.1 of Algorithm 3.1, $\bar{\lambda}_{0,i} = 0$ for $i \in \mathcal{I} \backslash \bar{\mathcal{I}}$. This together with $(b2)$ and $(b3)$ imply

$$\bar{\lambda}_0 \geq 0,\ \left(\bar{\lambda}_0\right)^\top \bar{x}_0 = 0. \tag{4.2}$$

From $(b4) - (b12)$ we have

$$\bar{\lambda}_{1,i}, \bar{\lambda}_{2,i} \geq 0,\ \bar{\lambda}_{1,i}\bar{x}_{1,i} = 0,\ \bar{\lambda}_{2,i}\bar{x}_{2,i} = 0,\ \forall i \in \hat{\mathcal{A}}_{12}. \tag{4.3}$$

Since $\min\{\|\mathcal{R}(x^k, y^k, z^k)\|^\eta, \varepsilon_{k,\bar{j}_k}\} \geq \tilde{\varepsilon}$ for all $k \in \hat{\mathcal{K}}$, it follows that $\bar{x}_{1,i} \geq \tilde{\varepsilon}$ for all $i \in \mathcal{A} \backslash \hat{\mathcal{A}}_1$ and hence for all $i \in \bar{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_1$ and $\bar{x}_{2,i} \geq \tilde{\varepsilon}$ for all $i \in \mathcal{A} \backslash \hat{\mathcal{A}}_2$ and hence for all $i \in \bar{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_2$. Since $\bar{\mathcal{B}}_{10} \subseteq \bar{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_1$ and $\bar{\mathcal{B}}_{12} \subseteq \bar{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_2$, we have from $(b13)$ and $(b15)$ that $\bar{\mathcal{B}}_{10} = \bar{\mathcal{B}}_{12} = \emptyset$. Hence, $(b14)$ and $(b16)$ imply that $\bar{\lambda}_{1,i} = 0$ for $i \in \bar{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_1$ and $\bar{\lambda}_{2,i} = 0$ for $i \in \bar{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_2$. Moreover, by Step 2.1 of Algorithm 3.1, $\bar{\lambda}_{1,i} = 0$ for $i \in \mathcal{A} \backslash \bar{\mathcal{A}}_1$ and $\bar{\lambda}_{2,i} = 0$ for $i \in \mathcal{A} \backslash \bar{\mathcal{A}}_2$. Hence,

$$\bar{\lambda}_{1,i} = 0,\ \forall i \in \mathcal{A} \backslash \hat{\mathcal{A}}_1;\ \bar{\lambda}_{2,i} = 0,\ \forall i \in \mathcal{A} \backslash \hat{\mathcal{A}}_2. \tag{4.4}$$

Since $\min\{\|\mathcal{R}(x^k, y^k, z^k)\|^\eta, \varepsilon_{k,\bar{j}_k}\} \geq \tilde{\varepsilon}$, we have from (3.1) that $\bar{x}_{2,i} > \tilde{\varepsilon}$ for $i \in \hat{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_{12}$ and $\bar{x}_{1,i} > \tilde{\varepsilon}$ for $i \in \hat{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_{12}$. Moreover, since Algorithm 3.1 maintains the feasibility of every iterate, it follows that $\bar{x}_{1,i}\bar{x}_{2,i} = 0$ for every $i \in \mathcal{A}$ and hence that $\bar{x}_{1,i} = 0$ for $i \in \hat{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_{12}$ and $\bar{x}_{2,i} = 0$ for $i \in \hat{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_{12}$. This together with (4.3) imply

$$\bar{\lambda}_{1,i}\bar{x}_{1,i} = 0,\ \forall i \in \hat{\mathcal{A}}_1;\ \bar{\lambda}_{2,i}\bar{x}_{2,i} = 0,\ \forall i \in \hat{\mathcal{A}}_2. \tag{4.5}$$

Since $\mathcal{A}_1(\bar{x}) \subseteq \hat{\mathcal{A}}_1$ and $\mathcal{A}_2(\bar{x}) \subseteq \hat{\mathcal{A}}_2$, we obtain from (4.2)-(4.5) that

$$\bar{\lambda}_{1,i}, \bar{\lambda}_{2,i} \geq 0,\ \forall i \in \mathcal{A}_{12}(\bar{x});\ \bar{\lambda}_{1,i} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_1(\bar{x});\ \bar{\lambda}_{2,i} = 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_2(\bar{x}). \tag{4.6}$$

Finally, since $\triangle\bar{x} = 0$, letting $k \in \hat{\mathcal{K}} \to \infty$ in the first equation of (3.10) yields $\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\lambda}) = 0$. Combining this with (4.2) and (4.6) shows that $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies S-stationarity. This is a contradiction. □

LEMMA 4.4. *For any infinite index set $\mathcal{K}$ and any associated index set $\{\bar{j}_k\}_\mathcal{K}$, $\{\varepsilon_{k,\bar{j}_k}\}_\mathcal{K} \to 0$ if and only if $\{\alpha_{k,\bar{j}_k}\}_\mathcal{K} \to 0$.*

Proof. We have from (3.13) that $\alpha_{k,j} \geq \beta^j$ for any j. Hence, if $\{\alpha_{k,\bar{j}_k}\}_\mathcal{K} \to 0$, it follows that $\{\bar{j}_k\}_\mathcal{K} \to \infty$. Since $\varepsilon_{k,j} = \nu^j \varepsilon_{k,0}$ for any j, we have $\{\varepsilon_{k,\bar{j}_k}\}_\mathcal{K} \to 0$.

Suppose $\{\varepsilon_{k,\bar{j}_k}\}_\mathcal{K} \to 0$. Since $\varepsilon_{k,0} = \bar{\varepsilon} > 0$ for all k large enough by Lemma 4.1 and $\varepsilon_{k,j} = \nu^j \varepsilon_{k,0}$ for all j, it follows that $\{\bar{j}_k\}_\mathcal{K} \to \infty$. By (3.11) we have $\gamma_k^{\max} \geq \gamma > 1$ for every k. By the definitions of $\mathcal{A}_1^{k,j}$ and $\mathcal{A}_2^{k,j}$, it follows that for any j, $x_{1,i}^k \leq \varepsilon_{k,j},\ \forall i \in \mathcal{A}_1^{k,j}$ and $x_{2,i}^k \leq \varepsilon_{k,j},\ \forall i \in \mathcal{A}_2^{k,j}$. Hence, $\varepsilon_{k,j} \geq x_{1,i}^k/\gamma_k^{\max},\ \forall i \in \mathcal{A}_1^{k,j}$ and $\varepsilon_{k,j} \geq x_{2,i}^k/\gamma_k^{\max},\ \forall i \in \mathcal{A}_2^{k,j}$. This together with (3.13) imply that $\alpha_{k,j} \leq \max\{\beta\alpha_{k,j-1}, \varepsilon_{k,j}\}$. Hence, it follows that

$$\begin{aligned}\max\{\varepsilon_{k,j}, \alpha_{k,j}\} &\leq \max\{\beta\alpha_{k,j-1}, \varepsilon_{k,j}\}\\ &= \max\{\beta\alpha_{k,j-1}, \nu\varepsilon_{k,j-1}\}\\ &\leq \max\{\beta, \nu\}\max\{\alpha_{k,j-1}, \varepsilon_{k,j-1}\}\\ &\leq \max\{\beta^j, \nu^j\}\max\{\alpha_{k,0}, \varepsilon_{k,0}\}.\end{aligned}$$

This implies $\{\alpha_{k,\bar{j}_k}\}_{\mathcal{K}} \to 0$ as $\nu, \beta \in (0,1)$ and $\{\bar{j}_k\}_{\mathcal{K}} \to \infty$. □

LEMMA 4.5. *Suppose Assumptions A1-A4 hold. If there is an infinite index set $\mathcal{K}$ such that $\|\mathcal{R}(x^k, y^k, z^k)\|^\eta \geq \hat{\varepsilon}$ for some $\hat{\varepsilon} > 0$ and all $k \in \mathcal{K}$, then there exists a $\bar{\alpha} > 0$ such that for all $k \in \mathcal{K}$ large enough, all j and all $\alpha \in (0, \bar{\alpha}]$, (i) $x_0^k + \alpha \triangle x_0^{k,j} \geq 0$, (ii) $x_{1,i}^k + \alpha \triangle x_{1,i}^{k,j} \geq 0,\ \forall i \in \mathcal{A} \backslash (\mathcal{B}_8^{k,j} \cup \mathcal{B}_9^{k,j})$, and (iii) $x_{2,i}^k + \alpha \triangle x_{2,i}^{k,j} \geq 0,\ \forall i \in \mathcal{A} \backslash (\mathcal{B}_4^{k,j} \cup \mathcal{B}_5^{k,j})$.*

Proof. Let $\bar{\varepsilon}$ be defined as in Lemma 4.1. From the definitions of $\mathcal{I}^k$, $\mathcal{A}_1^k$ and $\mathcal{A}_2^k$, we have that for all $k \in \mathcal{K}$ large enough,

$$x_{0,i}^k > \min\{\hat{\varepsilon}, \bar{\varepsilon}\},\ \forall i \in \mathcal{I} \backslash \mathcal{I}^k;\ x_{1,i}^k > \min\{\hat{\varepsilon}, \bar{\varepsilon}\},\ \forall i \in \mathcal{A} \backslash \mathcal{A}_1^k;\ x_{2,i}^k > \min\{\hat{\varepsilon}, \bar{\varepsilon}\},\ \forall i \in \mathcal{A} \backslash \mathcal{A}_2^k.$$

Since $\{\triangle x^{k,j}\}_{\mathcal{K}}$ are bounded for all j by Lemma 4.2, there is a $\bar{\alpha} > 0$ such that for all $k \in \mathcal{K}$ large enough, all j and all $\alpha \in (0, \bar{\alpha}]$,

$$x_{0,i}^k + \alpha \triangle x_{0,i}^{k,j} \geq 0,\ \forall i \in \mathcal{I} \backslash \mathcal{I}^k;\ x_{1,i}^k + \alpha \triangle x_{1,i}^{k,j} \geq 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_1^k;\ x_{2,i}^k + \alpha \triangle x_{1,i}^{k,j} \geq 0,\ \forall i \in \mathcal{A} \backslash \mathcal{A}_2^k.$$

On the other hand, it follows from (3.16) that for any j, $\triangle x_{0,i}^{k,j} = \varpi_{0,i}^{k,j},\ \forall i \in \mathcal{I}^k$, $\triangle x_{1,i}^{k,j} = \varpi_{1,i}^{k,j},\ \forall i \in \mathcal{A}_1^k$ and $\triangle x_{2,i}^{k,j} = \varpi_{2,i}^{k,j},\ \forall i \in \mathcal{A}_2^k$. Hence, we have from (3.15) that for any j:

$$\begin{aligned}
&(c1): \triangle x_{0,i}^{k,j} = -x_{0,i}^k,\ \forall i \in \mathcal{I}_+^k; \quad (c2): \quad \triangle x_{0,i}^{k,j} = -\lambda_{0,i}^{k+1} \geq -x_{0,i}^k,\ \forall i \in \mathcal{I}_+^k;\\
&(c3): \triangle x_{1,i}^{k,j} = -x_{1,i}^k,\ \forall i \in (\mathcal{A}_1^{k,j} \backslash \mathcal{A}_{12}^{k,j}) \cup \mathcal{B}_1^{k,j} \cup \mathcal{B}_2^{k,j} \cup \mathcal{B}_3^{k,j} \cup \mathcal{B}_{10}^{k,j};\\
&(c4): \triangle x_{2,i}^{k,j} = -x_{2,i}^k,\ \forall i \in (\mathcal{A}_2^{k,j} \backslash \mathcal{A}_{12}^{k,j}) \cup \mathcal{B}_1^{k,j} \cup \mathcal{B}_6^{k,j} \cup \mathcal{B}_7^{k,j} \cup \mathcal{B}_{12}^{k,j};\\
&(c5): \triangle x_{1,i}^{k,j} = -\lambda_{1,i}^{k+1} > -x_{1,i}^k,\ \forall i \in \mathcal{B}_6^{k,j} \cup \mathcal{B}_7^{k,j} \cup \mathcal{B}_{11}^{k,j};\\
&(c6): \triangle x_{2,i}^{k,j} = -\lambda_{2,i}^{k+1} > -x_{2,i}^k,\ \forall i \in \mathcal{B}_2^{k,j} \cup \mathcal{B}_3^{k,j} \cup \mathcal{B}_{13}^{k,j};\\
&(c7): \triangle x_{1,i}^{k,j} = -\lambda_{1,i}^{k+1} + x_{2,i}^k/\alpha_{k,j} \geq -\lambda_{1,i}^{k+1} > -x_{1,i}^k,\ \forall i \in \mathcal{B}_4^{k,j} \cup \mathcal{B}_5^{k,j};\\
&(c8): \triangle x_{2,i}^{k,j} = -\lambda_{2,i}^{k+1} + x_{1,i}^k/\alpha_{k,j} \geq -\lambda_{2,i}^{k+1} > -x_{2,i}^k,\ \forall i \in \mathcal{B}_8^{k,j} \cup \mathcal{B}_9^{k,j}.
\end{aligned}$$

Now the result follows immediately. □

LEMMA 4.6. *Suppose Assumptions A1-A4 hold. If there is an infinite index set $\mathcal{K}$ such that (i) $\|\mathcal{R}(x^k, y^k, z^k)\|^\eta \geq \hat{\varepsilon}$ for some $\hat{\varepsilon} > 0$ and all $k \in \mathcal{K}$, and (ii) $\{\varepsilon_{k,j_k}\} \to 0$ as $k \in \mathcal{K} \to \infty$, where j_k is the index for which conditions (3.17) and (3.18) in Step 3.4 of Algorithm 3.1 are satisfied at iteration k, then any limit point of the sequence $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\mathcal{K}}$ satisfies the S-stationarity conditions for problem (1.6).*

Proof. First note that $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\mathcal{K}}$ is bounded by Lemma 4.2. Since $\{\varepsilon_{k,j_k}\}_{\mathcal{K}} \to 0$, we have $\{j_k\}_{\mathcal{K}} \to \infty$ and from Lemma 4.4 that $\{\alpha_{k,j_k}\}_{\mathcal{K}} \to 0$. Suppose there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{(x^k, y^{k+1}, \lambda^{k+1})\}_{\bar{\mathcal{K}}} \to (\bar{x}, \bar{y}, \bar{\lambda})$, but $(\bar{x}, \bar{y}, \bar{\lambda})$ does not satisfy the S-stationarity conditions for problem (1.6). Let

$$\tau = \frac{1}{2} \min\left\{\bar{\varepsilon}, \hat{\varepsilon}, \min\left\{\bar{x}_{1,i},\ i \in \mathcal{A} \backslash \mathcal{A}_1(\bar{x})\right\}, \min\left\{\bar{x}_{2,i},\ i \in \mathcal{A} \backslash \mathcal{A}_2(\bar{x})\right\}\right\},$$

where $\bar{\varepsilon}$ is given by Lemma 4.1. It is obvious that $\tau > 0$. Let $\hat{j}_k$ be the first index of the sequence $j = 0, 1, ...$ for which $\varepsilon_{k,j} \leq \tau$. Clearly, $\hat{j}_k < j_k$ for all $k \in \bar{\mathcal{K}}$ large enough as $\{j_k\}_{\bar{\mathcal{K}}} \to \infty$ and $\varepsilon_{k,\hat{j}_k} = \nu \varepsilon_{k,\hat{j}_{k-1}} > \nu\tau$, since $\varepsilon_{k,\hat{j}_{k-1}} > \tau$. Therefore, we have $\tau \geq \varepsilon_{k,\hat{j}_k} = \nu^{\hat{j}_k} \bar{\varepsilon} > \nu\tau$ as $\varepsilon_{k,0} = \bar{\varepsilon} > \tau$ by Lemma 4.1. Clearly, by (3.13) there is an index $\hat{j}$ such that $\hat{j}_k \leq \hat{j}$ and $\alpha_{k,\hat{j}_k} \geq \beta^{\hat{j}}$ for all $k \in \bar{\mathcal{K}}$. Also, since $\bar{\varepsilon} > \tau \geq \varepsilon_{k,\hat{j}_k}$, it follows that $\mathcal{A}_1^{k,\hat{j}_k} = \mathcal{A}_1(\bar{x})$, $\mathcal{A}_2^{k,\hat{j}_k} = \mathcal{A}_2(\bar{x})$ and $\mathcal{A}_{12}^{k,\hat{j}_k} = \mathcal{A}_{12}(\bar{x})$ for all $k \in \bar{\mathcal{K}}$ large enough. Since $(\bar{x}, \bar{y}, \bar{\lambda})$ does not satisfy the S-stationarity conditions and $\min\{\|\mathcal{R}(x^k, y^k, z^k)\|^\eta, \varepsilon_{k,\hat{j}_k}\} \geq \nu\tau$, we conclude from Lemma 4.3 that there is a $\rho > 0$ such that $\nabla f(x^k)^\top \triangle x^{k,\hat{j}_k} \leq -\rho$ for all $k \in \bar{\mathcal{K}}$.

Since $\{\triangle x^{k,\hat{j}_k}\}$ is bounded by Lemma 4.2 and $\mathcal{I}$ and $\mathcal{A}$ are finite sets, there exists an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\triangle x^{k,\hat{j}_k}\} \to \triangle\bar{x}$ and $\mathcal{I}_-^k$, $\mathcal{I}_+^k$, $\mathcal{A}_1^k$, $\mathcal{A}_2^k$, $\mathcal{A}_1^{k,\hat{j}_k}$, $\mathcal{A}_2^{k,\hat{j}_k}$, $\mathcal{A}_{12}^{k,\hat{j}_k}$ and $\mathcal{B}_l^{k,\hat{j}_k}$ $(l = 1, ..., 13)$ are all fixed on $\hat{\mathcal{K}}$. Suppose they are $\bar{\mathcal{I}}_-$, $\bar{\mathcal{I}}_+$, $\bar{\mathcal{A}}_1$, $\bar{\mathcal{A}}_2$, $\hat{\mathcal{A}}_1$, $\hat{\mathcal{A}}_2$, $\hat{\mathcal{A}}_{12}$ and $\bar{\mathcal{B}}_l$ $(l = 1, ..., 13)$, respectively. Clearly, we have $\nabla f(\bar{x})^\top \triangle\bar{x} \leq -\rho$. Let $\bar{\sigma} \in (\sigma, 1)$. Due to the continuous differentiability of $f(x)$, by standard arguments used for the Armijo condition (3.18), one can show that for all small enough $\alpha > 0$,

$$f(\bar{x} + \alpha\triangle\bar{x}) \leq f(\bar{x}) + \bar{\sigma}\alpha\nabla f(\bar{x})^\top \triangle\bar{x} < f(\bar{x}) + \sigma\alpha\nabla f(\bar{x})^\top \triangle\bar{x} - \alpha(\bar{\sigma} - \sigma)\rho. \tag{4.7}$$

Let $\hat{\alpha} \in (0, \min\{\bar{\alpha}, \beta^{\hat{j}}\})$, where $\bar{\alpha}$ is given by Lemma 4.5, such that for all $\alpha \in [\beta\hat{\alpha}, \hat{\alpha}]$, inequality (4.7) holds. Since $\{\alpha_{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$, it follows that $\alpha_{k,j_k} < \beta\hat{\alpha}$ for all $k \in \hat{\mathcal{K}}$ large enough. By (3.13), there is an index $\bar{j}_k$ for every large enough $k \in \hat{\mathcal{K}}$ such that $\alpha_{k,\bar{j}_k}$ belongs to the region $[\beta\hat{\alpha}, \hat{\alpha}]$ and the line search conditions (3.17) and (3.18) fail to hold for these $\bar{j}_k$. Since $\alpha_{k,\hat{j}_k} \geq \beta^{\hat{j}} \geq \hat{\alpha} \geq \alpha_{k,\bar{j}_k}$, it follows that $\hat{j}_k \leq \hat{j} \leq \bar{j}_k$. Since $\{\alpha_{k,\bar{j}_k}\}_{\hat{\mathcal{K}}}$ is bounded below by $\beta\hat{\alpha}$, we obtain by Lemma 4.4 that $\{\varepsilon_{k,\bar{j}_k}\}_{\hat{\mathcal{K}}}$ is bounded below by a strictly positive scalar. This implies that there is an index $\bar{j}$ such that $\bar{j}_k \leq \bar{j}$ and thus, $\varepsilon_{k,\bar{j}_k} \geq \nu^{\bar{j}}\bar{\varepsilon}$ for all $k \in \hat{\mathcal{K}}$ large enough. Moreover, since $\varepsilon_{k,\bar{j}_k} \leq \nu^{\hat{j}}\bar{\varepsilon} \leq \nu^{\hat{j}_k}\bar{\varepsilon} = \varepsilon_{k,\hat{j}_k} \leq \tau$, we have from the definition of τ that $\mathcal{A}_1^{k,\bar{j}_k} = \hat{\mathcal{A}}_1 = \mathcal{A}_1(\bar{x})$, $\mathcal{A}_2^{k,\bar{j}_k} = \hat{\mathcal{A}}_2 = \mathcal{A}_2(\bar{x})$ and $\mathcal{A}_{12}^{k,\bar{j}_k} = \hat{\mathcal{A}}_{12} = \mathcal{A}_{12}(\bar{x})$ for all $k \in \hat{\mathcal{K}}$ large enough. Hence, in view of (3.14), we know $\mathcal{B}_l^{k,\bar{j}_k} = \mathcal{B}_l^{k,\hat{j}_k} = \bar{\mathcal{B}}_l$ $(l = 1, ..., 13)$ for all $k \in \hat{\mathcal{K}}$ large enough. On the other hand, according to (3.15), the elements in $\varpi^{k,\bar{j}_k}$ and $\varpi^{k,\hat{j}_k}$ are identical except for those belonging to $\bar{\mathcal{B}}_l$, $l = 4, 5, 8, 9$. Since $\bar{\mathcal{B}}_l \subseteq \hat{\mathcal{A}}_{12} = \mathcal{A}_{12}(\bar{x})$ $(l = 4, 5, 8, 9)$ and the sequences $\{\alpha_{k,\bar{j}_k}\}$ and $\{\alpha_{k,\hat{j}_k}\}$ are bounded below by $\beta\hat{\alpha}$ on $\hat{\mathcal{K}}$, we obtain from (3.15) that as $k \in \hat{\mathcal{K}} \to \infty$,

$$\begin{aligned}
&\lim \varpi_{1,i}^{k,\bar{j}_k} = \lim \varpi_{1,i}^{k,\hat{j}_k} = -\bar{\lambda}_{1,i},\ \lim \varpi_{2,i}^{k,\bar{j}_k} = \lim \varpi_{2,i}^{k,\hat{j}_k} = 0,\ \forall i \in \bar{\mathcal{B}}_4 \cup \bar{\mathcal{B}}_5;\\
&\lim \varpi_{2,i}^{k,\bar{j}_k} = \lim \varpi_{2,i}^{k,\hat{j}_k} = -\bar{\lambda}_{2,i},\ \lim \varpi_{1,i}^{k,\bar{j}_k} = \lim \varpi_{1,i}^{k,\hat{j}_k} = 0,\ \forall i \in \bar{\mathcal{B}}_8 \cup \bar{\mathcal{B}}_9.
\end{aligned}$$

Therefore, we have proved $\lim \varpi^{k,\bar{j}_k} = \lim \varpi^{k,\hat{j}_k}$ as $k \in \hat{\mathcal{K}} \to \infty$. This together with the linear system (3.16) confirms $\{\triangle x^{k,\bar{j}_k}\}_{\hat{\mathcal{K}}} \to \triangle\bar{x}$.

From (3.16) and (3.15) we have

$$\begin{aligned}
\triangle x_{2,i}^{k,\bar{j}_k} &= \varpi_{2,i}^{k,\bar{j}_k} = -x_{2,i}^k/\alpha_{k,\bar{j}_k}, \quad \forall i \in \mathcal{B}_4^{k,\bar{j}_k} \cup \mathcal{B}_5^{k,\bar{j}_k},\\
\triangle x_{1,i}^{k,\bar{j}_k} &= \varpi_{1,i}^{k,\bar{j}_k} = -x_{1,i}^k/\alpha_{k,\bar{j}_k}, \quad \forall i \in \mathcal{B}_8^{k,\bar{j}_k} \cup \mathcal{B}_9^{k,\bar{j}_k}.
\end{aligned}$$

Therefore, it follows that

$$x_{2,i}^k + \alpha_{k,\bar{j}_k}\triangle x_{2,i}^{k,\bar{j}_k} = 0,\ \forall i \in \mathcal{B}_4^{k,\bar{j}_k} \cup \mathcal{B}_5^{k,\bar{j}_k};\ x_{1,i}^k + \alpha_{k,\bar{j}_k}\triangle x_{1,i}^{k,\bar{j}_k} = 0,\ \forall i \in \mathcal{B}_8^{k,\bar{j}_k} \cup \mathcal{B}_9^{k,\bar{j}_k}.$$

Since $\alpha_{k,\bar{j}_k} \in [\beta\hat{\alpha}, \hat{\alpha}]$ for all $k \in \hat{\mathcal{K}}$ large enough, there exists an infinite index set $\tilde{\mathcal{K}} \subseteq \hat{\mathcal{K}}$ such that $\{\alpha_{k,\bar{j}_k}\}_{\tilde{\mathcal{K}}} \to \tilde{\alpha}$ for some $\tilde{\alpha} \in [\beta\hat{\alpha}, \hat{\alpha}]$. Then since $\alpha_{k,\bar{j}_k} \leq \hat{\alpha} \leq \bar{\alpha}$ for $k \in \tilde{\mathcal{K}}$, we obtain from Lemma 4.5 that $x^k + \alpha_{k,\bar{j}_k}\triangle x^{k,\bar{j}_k} \geq 0$ for all $k \in \tilde{\mathcal{K}}$ large enough. On the other hand, by the continuity of $f(x)$ and $\nabla f(x)$ and the fact that $\{\alpha_{k,\bar{j}_k}\}_{\tilde{\mathcal{K}}} \to \tilde{\alpha} \in [\beta\hat{\alpha}, \hat{\alpha}]$ and $\{\triangle x^{k,\bar{j}_k}\}_{\tilde{\mathcal{K}}} \to \triangle\bar{x}$, we have from

(4.7) that for all $k \in \tilde{\mathcal{K}}$ large enough,

$$
\begin{aligned}
& f(x^k + \alpha_{k,\bar{j}_k} \triangle x^{k,\bar{j}_k}) - f(x^k) - \sigma \alpha_{k,\bar{j}_k} \nabla f(x^k)^\top \triangle x^{k,\bar{j}_k} \\
= \ & f(x^k + \alpha_{k,\bar{j}_k} \triangle x^{k,\bar{j}_k}) - f(\bar{x} + \tilde{\alpha} \triangle \bar{x}) - f(x^k) + f(\bar{x}) \\
& -\sigma \left(\alpha_{k,\bar{j}_k} \nabla f(x^k)^\top \triangle x^{k,\bar{j}_k} - \tilde{\alpha} \nabla f(\bar{x})^\top \triangle \bar{x} \right) \\
& + f(\bar{x} + \tilde{\alpha} \triangle \bar{x}) - f(\bar{x}) - \sigma \tilde{\alpha} \nabla f(\bar{x})^\top \triangle \bar{x} \\
= \ & f(\bar{x} + \tilde{\alpha} \triangle \bar{x}) - f(\bar{x}) - \sigma \tilde{\alpha} \nabla f(\bar{x})^\top \triangle \bar{x} \\
& + \mathcal{O}\left(\left\| \left(x^k - \bar{x}, \triangle x^{k,\bar{j}_k} - \triangle \bar{x}, \alpha_{k,\bar{j}_k} - \tilde{\alpha} \right) \right\| \right) \\
< \ & -\tilde{\alpha}(\bar{\sigma} - \sigma)\rho + \mathcal{O}\left(\left\| \left(x^k - \bar{x}, \triangle x^{k,\bar{j}_k} - \triangle \bar{x}, \alpha_{k,\bar{j}_k} - \tilde{\alpha} \right) \right\| \right) \\
\leq \ & 0.
\end{aligned} \tag{4.8}
$$

Consequently, we have proved that conditions (3.17) and (3.18) hold with index $\bar{j}_k$ for all $k \in \tilde{\mathcal{K}}$ large enough. Therefore, from Step 3.4 of Algorithm 3.1, we know that $j_k = \bar{j}_k$ for all $k \in \tilde{\mathcal{K}}$ large enough. Thus, the sequence $\{\varepsilon_{k,j_k}\}_{\tilde{\mathcal{K}}}$ is bounded away from zero as $\varepsilon_{k,\bar{j}_k} \geq \nu^{\bar{j}} \bar{\varepsilon}$ for all $k \in \tilde{\mathcal{K}}$ large enough. This contradicts the assumption that $\{\varepsilon_{k,j_k}\}_{\mathcal{K}} \to 0$, where $\mathcal{K} \supseteq \tilde{\mathcal{K}}$. ☐

LEMMA 4.7. *Suppose Assumptions A1-A4 hold. If there is an infinite index set* $\mathcal{K}$ *such that* $\{\|\mathcal{R}(x^k, y^k, z^k)\|\} \to 0$ *as* $k \in \mathcal{K} \to \infty$*, any limit points of the sequences* $\{(x^k, y^k, z^k)\}_{\mathcal{K}}$ *and* $\{(x^k, y^k, \lambda^k)\}_{\mathcal{K}}$ *satisfy the KKT conditions for problem (2.5) and the S-stationarity conditions for problem (1.6), respectively.*

Proof. First, by Lemma 4.2, the sequence $\{x^k, y^k, z^k, \lambda^k\}$ is uniformly bounded. Suppose for some infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$, $\{x^k, y^k, z^k, \lambda^k\} \to (\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ as $k \in \bar{\mathcal{K}} \to \infty$. Clearly, $(\bar{x}, \bar{y}, \bar{z})$ satisfies the KKT conditions (2.6) since $\|\mathcal{R}(\cdot)\|$ is the KKT error for problem (2.5). On the other hand, it follows from (3.19) that $\bar{\lambda}_0 = \bar{z}_0$, $\bar{z}_1 = \bar{\lambda}_1 + \bar{z}_3 \bar{x}_2$ and $\bar{z}_2 = \bar{\lambda}_2 + \bar{z}_3 \bar{x}_1$. Hence, we obtain from Proposition 2.8 that x is a S-stationary point of problem (1.6) with multipliers $\bar{y}$ and $\bar{\lambda}$. ☐

Now we are ready to state the global convergence properties of Algorithm 3.1. We ignore the case that Algorithm 3.1 stops with a S-stationary point in a finite number of iterations.

THEOREM 4.8. *Suppose Assumptions A1-A4 hold and Algorithm 3.1 generates an infinite sequence of iterates. Then any limit point of either the sequence* $\{x^k, y^k, \lambda^k\}$ *or the sequence* $\{x^k, y^{k+1}, \lambda^{k+1}\}$ *satisfies the S-stationarity conditions for problem (1.6).*

Proof. First note that the sequence $\{x^k, y^k, z^k, \lambda^k\}$ is uniformly bounded by Lemma 4.2. Consider any infinite index set $\mathcal{K}$ on which $\{x^k\} \to \bar{x}$, $\{(y^k, \lambda^k)\} \to (\hat{y}, \hat{\lambda})$ and $\{(y^{k+1}, \lambda^{k+1})\} \to (\bar{y}, \bar{\lambda})$ as $k \in \mathcal{K} \to \infty$. There are two cases to consider.

Case 1. There is an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\|\mathcal{R}(x^k, y^k, z^k)\|^\eta \geq \hat{\varepsilon}$ for some $\hat{\varepsilon} > 0$ and all $k \in \bar{\mathcal{K}}$ large enough. There are two subcases.

Case 1.1. There is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\varepsilon_{k,j_k} \geq \tilde{\varepsilon}$ for some $\tilde{\varepsilon} > 0$ and all $k \in \hat{\mathcal{K}}$ large enough. Then it follows from Lemma 4.4 that $\alpha_{k,j_k} \geq \tilde{\alpha}$ for some $\tilde{\alpha} > 0$ and all $k \in \hat{\mathcal{K}}$ large enough. Assumption A3 and the continuous differentiability of $f(x)$ imply that $f(x)$ is bounded below on the feasible region of problem (1.6). Therefore, we know from the Armijo condition (3.18) that $\{\nabla f(x^k)^\top \triangle x^{k,j_k}\} \to 0$ as $k \in \hat{\mathcal{K}} \to \infty$. Hence, it follows from Lemma 4.3 that $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies the S-stationarity conditions for problem (1.6).

Case 1.2. There is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\varepsilon_{k,j_k}\} \to 0$ as $k \in \hat{\mathcal{K}} \to \infty$. Then it follows from Lemma 4.6 that $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies the S-stationarity conditions for problem (1.6).

Case 2. There is an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\|\mathcal{R}(x^k, y^k, z^k)\|\} \to 0$ as $k \in \bar{\mathcal{K}} \to \infty$. We know from Lemma 4.7 that $(\bar{x}, \hat{y}, \hat{\lambda})$ satisfies the S-stationarity conditions for problem (1.6). ☐

5. Quadratic local convergence. In this section, we assume that Algorithm 3.1 generates an infinite sequence of iterates and for some infinite index set $\mathcal{K}$, $\{x^k\} \to x^*$, $\{(y^k, \lambda^k)\} \to (\bar{y}, \bar{\lambda})$ and $\{(y^{k+1}, \lambda^{k+1})\} \to (\hat{y}, \hat{\lambda})$ as $k \in \mathcal{K} \to \infty$. By Theorem 4.8, x^* is a S-stationary point of problem

(1.6). Under Assumption A2, there is a unique multiplier vector (y^*, λ^*) associated with x^*. Again we have from Theorem 4.8 that either $(\bar{y}, \bar{\lambda}) = (y^*, \lambda^*)$ or $(\hat{y}, \hat{\lambda}) = (y^*, \lambda^*)$. To establish the fast local convergence of Algorithm 3.1, we need the following assumptions in addition to A1-A4.

A5. *The MPLCC-SOSC hold at x^* – see Definition 2.6.*

A6. *Strict complementarity holds at (x^*, y^*, λ^*) – see Definition 2.7.*

Assumptions A5 and A6 are usually required to prove fast local convergence of NLP based algorithms. In our case Assumption A5 has two purposes. First, it allows the KKT error $\|\mathcal{R}(x,y,z)\|$ to properly measure the distance from a nearby point to the solution. On the other hand, it guarantees that the exact Hessian can be eventually used without any modification by Algorithm 3.1. Moreover, in order for the final active set to be correctly identified using the error bound $\|\mathcal{R}(x,y,z)\|^{\eta}$, we need the parameter $z_3^{\max}$ in Algorithm 3.1 to be chosen large enough so that

$$z_3^{\max} \geq z_3^* = \min\left\{0,\ -\frac{\lambda_{1,i}^*}{x_{2,i}^*}, i \in \mathcal{A}_1(x^*)\backslash\mathcal{A}_{12}(x^*);\ -\frac{\lambda_{2,i}^*}{x_{1,i}^*}, i \in \mathcal{A}_2(x^*)\backslash\mathcal{A}_{12}(x^*)\right\}. \tag{5.1}$$

Similarly to NLPs, a first-order solution of problem (1.6) is a strict local optimal solution under the MPLCC-SOSC and strict complementarity. This result can be proved in a straightforward way using the stability results in [29]. The next result is an easy consequence of the strict local optimality of x^*.

LEMMA 5.1. *Under Assumptions A1-A6, x^* is an isolated accumulation point of the sequence $\{x^k\}$ generated by Algorithm 3.1.*

Proof. First it follows from Theorem 7 in [30] that under Assumptions A5 and A6, x^* is a strict local minimizer of problem (1.6). This implies that there is a small neighborhood of x^* in which there are no other first-order solutions to problem (1.6). Since by Theorem 4.8 any accumulation point of x^k is a first-order solution to problem (1.6), the result follows immediately. □

The original version of the next result is from [19]. Here we cite a slightly different version given by Proposition 5.4 in [18].

PROPOSITION 5.2. *Assume that $\omega^* \in \Re^t$ is an isolated accumulation point of a sequence $\{\omega^k\} \subset \Re^t$ such that for every subsequence $\{\omega^k\}_{\bar{\mathcal{K}}}$ converging to ω^*, there is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\|\omega^{k+1} - \omega^k\|\}_{\hat{\mathcal{K}}} \to 0$. Then the whole sequence $\{\omega^k\}$ converges to ω^*.*

Let $z_1^* = \lambda_1^* + z_3^* x_2^*$ and $z_2^* = \lambda_2^* + z_3^* x_1^*$. Clearly, x^* is a KKT point of problem (2.5) and (y^*, z^*) is an associated multiplier vector. The next result shows that the sequence $\{(x^k, y^k, z^k, \lambda^k)\}$ converges to $(x^k, y^*, z^*, \lambda^*)$.

LEMMA 5.3. *Suppose Assumptions A1-A6 and (5.1) hold. If there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\mathcal{I}(x^*) \subseteq \mathcal{I}^k$, $\mathcal{A}_1(x^*) \subseteq \mathcal{A}_1^k$ and $\mathcal{A}_2(x^*) \subseteq \mathcal{A}_2^k$ for all $k \in \bar{\mathcal{K}}$, then the entire sequence $\{(x^k, y^k, z^k, \lambda^k)\}$ converges to $(x^k, y^*, z^*, \lambda^*)$.*

Proof. Since $\{\mathcal{H}^k\}$ is bounded by Assumption A4 and $\mathcal{I}$ and $\mathcal{A}$ are finite sets, there exists an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\mathcal{H}^k\}_{\hat{\mathcal{K}}} \to \bar{\mathcal{H}}$ and $\mathcal{I}^k$, $\mathcal{I}_-^k$, $\mathcal{I}_+^k$, $\mathcal{A}_1^k$, $\mathcal{A}_2^k$, $\mathcal{A}_1^{k,\hat{\jmath}_k}$, $\mathcal{A}_2^{k,\hat{\jmath}_k}$, $\mathcal{A}_{12}^{k,\hat{\jmath}_k}$ and $\mathcal{B}_l^{k,\hat{\jmath}_k}$ $(l = 1, ..., 13)$ are all fixed on $\hat{\mathcal{K}}$. Suppose they are $\bar{\mathcal{I}}$, $\bar{\mathcal{I}}_-$, $\bar{\mathcal{I}}_+$, $\bar{\mathcal{A}}_1$, $\bar{\mathcal{A}}_2$, $\hat{\mathcal{A}}_1$, $\hat{\mathcal{A}}_2$, $\hat{\mathcal{A}}_{12}$ and $\bar{\mathcal{B}}_l$ $(l = 1, ..., 13)$, respectively. Let $\bar{E} = [e_{0,i}, i \in \bar{\mathcal{I}};\ e_{1,i}, i \in \bar{\mathcal{A}}_1;\ e_{2,i}, i \in \bar{\mathcal{A}}_2]$. Consider the following linear system in (d_x, d_y, d_λ)

$$\left\{\begin{array}{lllcl} \bar{\mathcal{H}} d_x & +A^\top d_y & -\bar{E} d_\lambda & = & -\nabla f(x^*), \\ A d_x & & & = & 0, \\ \bar{E}^\top d_x & & & = & 0. \end{array}\right. \tag{5.2}$$

Since $\mathcal{I}(x^*) \subseteq \mathcal{I}^k$, $\mathcal{A}_1(x^*) \subseteq \mathcal{A}_1^k$ and $\mathcal{A}_2(x^*) \subseteq \mathcal{A}_2^k$ for all $k \in \hat{\mathcal{K}}$, it follows from (2.2), (2.3) and (2.4) that $(0, y^*, \bar{E}^\top \lambda^*)$ is a solution of (5.2). On the other hand, by Step 1.1 of Algorithm 3.1 and Assumption A2, $[A^\top, \bar{E}]$ has full column rank. Assumption A4 implies that $\bar{\mathcal{H}}$ is positive definite on the null space of $[A^\top, \bar{E}]^\top$. Therefore, similarly to Lemma 3.1, we conclude that the matrix in (5.2) is nonsingular and thus (5.2) has the unique solution $(0, y^*, \bar{E}^\top\lambda^*)$. By continuity, taking limit in (3.10) as $k \in \hat{\mathcal{K}} \to \infty$ yields (5.2). In particular, $\{(\triangle x^k, y^{k+1}, \bar{E}^\top\lambda^{k+1})\}_{\hat{\mathcal{K}}} \to (0, y^*, \bar{E}^\top\lambda^*)$. Notice that by Step 2.1 of Algorithm 3.1, the multipliers in λ^{k+1} for which the corresponding primal variables are not in the working set are set to zero. Hence, it follows that $\{(\triangle x^k, y^{k+1}, \lambda^{k+1})\}_{\hat{\mathcal{K}}} \to (0, y^*, \lambda^*)$.

Now we show $\{\varpi^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. First it follows from (3.15) that $\{\varpi_{0,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -x_{0,i}^*,\ \forall i \in \bar{\mathcal{I}}_+$ and $\{\varpi_{0,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -\lambda_{0,i}^*,\ \forall i \in \bar{\mathcal{I}}_-$. From the definition of $\mathcal{I}_+^k$ and $\mathcal{I}_-^k$, we know $x_{0,i}^* \le \lambda_{0,i}^*$ for $i \in \bar{\mathcal{I}}_+$ and $x_{0,i}^* \ge \lambda_{0,i}^*$ for $i \in \bar{\mathcal{I}}_-$. This together with the complementarity condition imply that $x_{0,i}^* = 0$ for $i \in \bar{\mathcal{I}}_+$ and $\lambda_{0,i}^* = 0$ for $i \in \bar{\mathcal{I}}_-$. Hence, it follows that $\{\varpi_0^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. Since every x^k is feasible to problem (1.6), we have from (3.15) that for all $k \in \hat{\mathcal{K}}$, $\varpi_{1,i}^{k,j_k} = -x_{1,i}^k = 0,\ \forall i \in \hat{\mathcal{A}}_1 \backslash \hat{\mathcal{A}}_{12}$ and $\varpi_{2,i}^{k,j_k} = -x_{2,i}^k = 0,\ \forall i \in \hat{\mathcal{A}}_2 \backslash \hat{\mathcal{A}}_{12}$. Since $x_{1,i}^* \le \lambda_{1,i}^*,\ \forall i \in \bar{\mathcal{B}}_{10}$ and $x_{2,i}^* \le \lambda_{2,i}^*,\ \forall i \in \bar{\mathcal{B}}_{12}$ in view of (3.14), we obtain from (3.15) and the complementarity condition that $\{\varpi_{1,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -x_{1,i}^* = 0,\ \forall i \in \bar{\mathcal{B}}_{10}$ and $\{\varpi_{2,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -x_{2,i}^* = 0,\ \forall i \in \bar{\mathcal{B}}_{12}$. Also from (3.15), it follows that $\{\varpi_{1,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -\lambda_{1,i}^*,\ \forall i \in \bar{\mathcal{B}}_{11}$ and $\{\varpi_{2,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to -\lambda_{2,i}^*,\ \forall i \in \bar{\mathcal{B}}_{13}$. Now we show $\lambda_{1,i}^* = 0,\ \forall i \in \bar{\mathcal{B}}_{11}$. Notice that by the definition of $\mathcal{B}_{11}^{k,j_k}$ we have $x_{1,i}^k > \varepsilon_{k,j_k} \ge 0,\ \forall i \in \bar{\mathcal{B}}_{11}$. Since every x^k is feasible, we have $x_{2,i}^k = 0,\ \forall i \in \bar{\mathcal{B}}_{11}$. Hence, if $x_{1,\bar{i}}^* = 0$ for some $\bar{i} \in \bar{\mathcal{B}}_{11}$, it follows that $\bar{i} \in \mathcal{A}_{12}(x^*)$. This together with Definition 2.4 of S-stationarity implies that $\lambda_{1,\bar{i}}^* \ge 0$. However, the definition of $\mathcal{B}_{11}^{k,j_k}$ implies $\lambda_{1,i}^* \le x_{1,i}^*,\ \forall i \in \bar{\mathcal{B}}_{11}$. Hence, it follows that $\lambda_{1,\bar{i}}^* = 0$. On the other hand, if $x_{1,\bar{i}}^* > 0$ for some $\bar{i} \in \bar{\mathcal{B}}_{11}$, we have from the complementarity condition that $\lambda_{1,\bar{i}}^* = 0$. Therefore, we obtain $\{\varpi_{1,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to \lambda_{1,i}^* = 0,\ \forall i \in \bar{\mathcal{B}}_{11}$. In the same way we can derive $\{\varpi_{2,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to \lambda_{2,i}^* = 0,\ \forall i \in \bar{\mathcal{B}}_{13}$ as well. There are two cases to analyze the behavior of the remaining elements of ϖ^{k,j_k}.

Case 1. $\{\alpha_{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. Then it follows from Lemma 4.4 that $\{\varepsilon_{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. This implies $\hat{\mathcal{A}}_1 \subseteq \mathcal{A}_1(x^*)$ and $\hat{\mathcal{A}}_2 \subseteq \mathcal{A}_2(x^*)$. Strict complementarity (Assumption A6) gives that $\lambda_{1,i}^* > x_{1,i}^* = 0$ and $\lambda_{2,i}^* > x_{2,i}^* = 0$ for all $i \in \hat{\mathcal{A}}_{12}$. Therefore, if follows from (3.14) that $\bar{\mathcal{B}}_1 = \hat{\mathcal{A}}_{12}$ and $\bar{\mathcal{B}}_l = \emptyset,\ l = 2, ..., 9$. Hence, we have from (3.15) that $\{\varpi_{1,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to x_{1,i}^* = 0$ and $\{\varpi_{2,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to x_{2,i}^* = 0$ for all $i \in \cup_{l=1,...,9}\bar{\mathcal{B}}_l$.

Case 2. $\alpha_{k,j_k} \ge \hat{\alpha}$ for some $\hat{\alpha} > 0$ and all $k \in \hat{\mathcal{K}}$. Since $f(x)$ is bounded below on the feasible region of problem (1.6) by Assumption A3, it follows from (3.18) that $\{\nabla f(x^k)^\top \triangle x^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. By following the same analysis as in the proof of Lemma 4.3, we obtain relations (b5)-(b12) with $\bar{x}$ and $\bar{\lambda}$ in them replaced by x^* and λ^*, respectively. Hence, it follows from (3.15) that $\{\varpi_{1,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$ and $\{\varpi_{2,i}^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$ for all $i \in \bar{\mathcal{B}}_1 \cup \bar{\mathcal{B}}_2 \cup \bar{\mathcal{B}}_3 \cup \bar{\mathcal{B}}_6 \cup \bar{\mathcal{B}}_7$. If $\bar{\mathcal{B}}_4 \cup \bar{\mathcal{B}}_5 \neq \emptyset$, for any $i \in \bar{\mathcal{B}}_4 \cup \bar{\mathcal{B}}_5$, it follows from (3.14) and (b7) and (b8) in the proof of Lemma 4.3 that $x_{1,i}^* = 0$ and $\lambda_{1,i}^* = 0$. This contradicts the strict complementarity assumption. Hence, $\bar{\mathcal{B}}_4 \cup \bar{\mathcal{B}}_5 = \emptyset$. Similarly, we can obtain $\bar{\mathcal{B}}_8 \cup \bar{\mathcal{B}}_9 = \emptyset$.

Summarizing the results above, we conclude that $\{\varpi^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. Therefore, letting $k \in \hat{\mathcal{K}} \to \infty$, (3.10) and (3.16) attain the same limit (5.2). Thus, $\{\triangle x^{k,j_k}\}_{\hat{\mathcal{K}}} \to 0$. By Lemma 5.1, x^* is an isolated accumulation point of $\{x^k\}$. Hence, we have from Proposition 5.2 that the whole sequence $\{x^k\}$ converges to x^*. Under the MPLCC-LICQ, (y^*, λ^*) is the unique multiplier vector associated with x^*. By Theorem 4.8, either $\{(y^k, \lambda^k)\}$ or $\{(y^{k+1}, \lambda^{k+1})\}$ converges to (y^*, λ^*). Hence, the entire sequence $\{(y^k, \lambda^k)\}$ converges to (y^*, λ^*) as $\{x^k\} \to x^*$. Moreover, since (5.1) holds, letting $k \to \infty$ in (3.19) gives the convergence of $\{z^k\}$ to z^*. □

We now show that the working set eventually identifies the active set at x^*.

Lemma 5.4. *Suppose Assumptions A1-A6 and (5.1) hold. The entire sequence* $\{(x^k, y^k, z^k, \lambda^k)\}$ *converges to* $(x^*, y^*, z^*, \lambda^*)$ *and for all* k *large enough,* $\mathcal{I}^k = \mathcal{I}(x^*)$, $\mathcal{A}_1^k = \mathcal{A}_1(x^*)$ *and* $\mathcal{A}_2^k = \mathcal{A}_2(x^*)$.

Proof. Suppose there is an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\|\mathcal{R}(x^k, y^k, z^k)\|^\eta \ge \hat{\varepsilon}$ for some $\hat{\varepsilon} > 0$ and all $k \in \bar{\mathcal{K}}$. Then it follows from Lemma 4.1 that $\min\{\|\mathcal{R}(x^k, y^k, z^k)\|^\eta, \varepsilon_{k,0}\} \ge \min\{\hat{\varepsilon}, \bar{\varepsilon}\} > 0$, where $\bar{\varepsilon}$ is defined in Lemma 4.1. Hence, we have from (3.1) that $\mathcal{I}(x^*) \subseteq \mathcal{I}^k$, $\mathcal{A}_1(x^*) \subseteq \mathcal{A}_1^k$ and $\mathcal{A}_2(x^*) \subseteq \mathcal{A}_2^k$ for all $k \in \bar{\mathcal{K}}$ large enough. Thus, by Lemma 5.3, the entire sequence $\{x^k, y^k, z^k\}$ converges to the optimal primal-dual solution (x^*, y^*, z^*) of problem (2.5). However, this contradicts the assumption that the KKT error $\|\mathcal{R}(x^k, y^k, z^k)\|$ is bounded away from zero. Thus, it follows that $\{\|\mathcal{R}(x^k, y^k, z^k)\|\}_{\mathcal{K}} \to 0$.

Now by Lemma 4.7 we obtain the convergence of $\{y^k, \lambda^k\}_{\mathcal{K}}$ to (y^*, λ^*), which is the unique multiplier vector associated with x^*. Since (5.1) holds, letting $k \to \infty$ in (3.19) yields $\{z^k\}_{\mathcal{K}} \to z^*$. Clearly, $\|\mathcal{R}(x^*, y^*, z^*)\| = 0$. Notice that for any $s \in \mathcal{S}(x^*)$, where $\mathcal{S}(x^*)$ is given in Definition 2.6, we have

$$s^\top \nabla^2((x_1^*)^\top x_2^*) s = 2 \sum_{i \in \mathcal{A}} s^\top e_{1,i} e_{2,i}^\top s = 2 \sum_{i \in \mathcal{A}} s_{1,i} s_{2,i} = 0.$$

This implies $s^\top \nabla_{xx}^2 \bar{\mathcal{L}}(x^*, y^*, z^*) s = s^\top \nabla^2 f(x^*) s$ for $s \in \mathcal{S}(x^*)$, where $\bar{\mathcal{L}}(\cdot)$ is given by (2.7). Notice the null space of the active constraint gradients for problem (2.5) at x^* is also $\mathcal{S}(x^*)$. Hence, given that the MPLCC-SOSC hold at x^*, the SOSC for problem (2.5) hold with (x^*, y^*, z^*). Thus, by Theorem 2 in [8], there is a constant $\chi > 0$ such that

$$\|x^k - x^*\| \le \chi \|\mathcal{R}(x^k, y^k, z^k)\|$$

for all $k \in \mathcal{K}$ large enough. This together with (3.1) and the fact that $\eta \in (0,1)$ imply $\mathcal{I}^k = \mathcal{I}(x^*)$, $\mathcal{A}_1^k = \mathcal{A}_1(x^*)$ and $\mathcal{A}_2^k = \mathcal{A}_2(x^*)$ for all $k \in \mathcal{K}$ large enough. Hence we conclude from Lemma 5.3 that the entire sequence $\{x^k, y^k, z^k, \lambda^k\}$ converges to $(x^*, y^*, z^*, \lambda^*)$. Finally, we again have from Theorem 2 in [8] and (3.1) that $\mathcal{I}^k = \mathcal{I}(x^*)$, $\mathcal{A}_1^k = \mathcal{A}_1(x^*)$ and $\mathcal{A}_2^k = \mathcal{A}_2(x^*)$ for all k large enough. □

Let

$$E^* = [e_{0,i}, i \in \mathcal{I}(x^*);\ e_{1,i}, i \in \mathcal{A}_1(x^*);\ e_{2,i}, i \in \mathcal{A}_2(x^*)]. \tag{5.3}$$

The next result shows the quadratic convergence of a full Newton step.

LEMMA 5.5. *Suppose Assumptions A1-A6 and (5.1) hold. Then for all k large enough, (i) $\varpi^{k,0} = -(E^*)^\top x^k$; (ii) $\mathcal{H}^k = \nabla^2 f(x^k)$; (iii) $\|x^k + \triangle x^{k,0} - x^*\| = \mathcal{O}(\|x^k - x^*\|^2)$; (iv) there is a constant $\bar{\mu} > 0$ such that for all k large enough, $\nabla f(x^k)^\top \triangle x^{k,0} \le -\bar{\mu} \|\triangle x^{k,0}\|^2$.*

Proof. First note that for every k, $\mathcal{A}_1^{k,0} = \mathcal{A}_1^k$ and $\mathcal{A}_2^{k,0} = \mathcal{A}_2^k$. Hence, it follows from (3.14) that $\mathcal{B}_l^{k,0} = \emptyset$ $(l = 10, 11, 12, 13)$ for every k. Since strict complementarity holds at x^*, we obtain from Lemma 5.4 and (3.14) that for all k large enough, $\mathcal{I}_+^k = \mathcal{I}(x^*)$, $\mathcal{I}_-^k = \emptyset$, $\mathcal{B}_1^{k,0} = \mathcal{A}_{12}^k = \mathcal{A}_{12}(x^*)$ and $\mathcal{B}_l^{k,0} = \emptyset$ for $l = 2, ..., 9$. Now result (i) follows from (3.15).

From (3.8) and Lemma 5.4, we have $\mathcal{S}^k = \mathcal{S}(x^*)$ for all k large enough. Hence, the satisfaction of the MPLCC-SOSC at x^* implies that (3.7) eventually holds with the exact Hessian $\nabla^2 f(x^k)$. Thus, by Steps 1.2 and 4 of Algorithm 3.1, we have $\mathcal{H}^k = \nabla^2 f(x^k)$ for all k large enough.

Combining the results above, we obtain from (3.16), (2.2), (2.3) and (2.4) that for $j = 0$ and all k large enough,

$$\begin{aligned} \mathcal{M}^k \begin{bmatrix} x^k + \triangle x^{k,0} - x^* \\ y^{k+1,0} - y^* \\ \bar{\lambda}^{k+1,0} - (E^*)^\top \lambda^* \end{bmatrix} &= \begin{bmatrix} -\nabla f(x^k) + \nabla^2 f(x^k)(x^k - x^*) - A^\top y^* + E^*(E^*)^\top \lambda^* \\ A(x^k - x^*) \\ (E^*)^\top (x^k - x^*) + \varpi^{k,0} \end{bmatrix} \\ &= \begin{bmatrix} \nabla f(x^*) - \nabla f(x^k) + \nabla^2 f(x^k)(x^k - x^*) \\ 0 \\ 0 \end{bmatrix}. \end{aligned} \tag{5.4}$$

By Taylor's theorem, we have

$$\|\nabla f(x^*) - \nabla f(x^k) + \nabla^2 f(x^k)(x^k - x^*)\| = \mathcal{O}(\|x^k - x^*\|^2). \tag{5.5}$$

Hence, we know from Lemma 4.2 and (5.4) that result (iii) holds for all large enough k.

From (3.16) for $j = 0$ and k large enough we have that

$$\nabla f(x^k)^\top \triangle x^{k,0} = -(\triangle x^{k,0})^\top \nabla^2 f(x^k) \triangle x^{k,0} + (\varpi^{k,0})^\top \bar{\lambda}^{k+1,0}. \tag{5.6}$$

Also from Definition 2.6 of the MPLCC-SOSC we have using a modified version of a lemma in [4; pp. 296] that there exists a $\delta_0 > 0$ such that for any μ, $0 < \mu < \varrho$, where ϱ is specified in Definition 2.6,

$$d^\top (\nabla^2 f(x^*) + \delta E^* (E^*)^\top) d \geq \mu \|d\|^2 \tag{5.7}$$

for any d such that $Ad = 0$, whenever $\delta > \delta_0$. Finally, we have for k large enough, since $(E^*)^\top \triangle x^{k,0} = (E^k)^\top \triangle x^{k,0} = \varpi^{k,0} = -(E^*)^\top x^k$, $\{\bar{\lambda}^{k+1,0}\} \to (E^*)^\top \lambda^*$ by (5.4) and (5.5), $\mathcal{B}_l^{k,0} = \emptyset$ for $l = 2, ..., 13$, and $\mathcal{I}_-^k = \emptyset$, and $\varpi_{1,i}^{k,0} = 0$ for $i \in \mathcal{A}_1^{k,0} \backslash \mathcal{A}_{12}^{k,0}$ and $\varpi_{2,i}^{k,0} = 0$ for $i \in \mathcal{A}_2^{k,0} \backslash \mathcal{A}_{12}^{k,0}$, that

$$\begin{aligned} (\varpi^{k,0})^\top \bar{\lambda}^{k+1,0} + \delta \triangle x^{k,0} E^* (E^*)^\top \triangle x^{k,0} &= -\textstyle\sum_{i \in \mathcal{I}_+^k} (\lambda_{0,i}^* - \delta x_{0,i}^k) x_{0,i}^k \\ &\quad - \textstyle\sum_{i \in \mathcal{B}_1^{k,0}} \left[(\lambda_{1,i}^* - \delta x_{1,i}^k) x_{1,i}^k + (\lambda_{2,i}^* - \delta x_{2,i}^k) x_{2,i}^k \right] \leq 0. \end{aligned} \tag{5.8}$$

The nonnegativity of the above follows from the facts that $\lambda_{0,i}^k \geq x_{0,i}^k$ for $i \in \mathcal{I}_+^k$, $\lambda_{1,i}^k \geq x_{1,i}^k$ and $\lambda_{2,i}^k \geq x_{2,i}^k$ for $i \in \mathcal{B}_1^{k,0}$. Combining (5.6), (5.7) and (5.8) and using the continuity of $\nabla^2 f(x)$ yields

$$\begin{aligned} & \nabla f(x^k)^\top \triangle x^{k,0} \\ = \; & -(\triangle x^{k,0})^\top \nabla^2 f(x^*) \triangle x^{k,0} + (\varpi^{k,0})^\top \bar{\lambda}^{k+1,0} \\ & +(\triangle x^{k,0})^\top (\nabla^2 f(x^*) - \nabla^2 f(x^k)) \triangle x^{k,0} \\ \leq \; & -\mu \|\triangle x^{k,0}\|^2 + \delta (\triangle x^{k,0})^\top E^* (E^*)^\top \triangle x^{k,0} + (\varpi^{k,0})^\top \bar{\lambda}^{k+1,0} \\ & +(\triangle x^{k,0})^\top (\nabla^2 f(x^*) - \nabla^2 f(x^k)) \triangle x^{k,0} \\ \leq \; & -\tfrac{\mu}{2} \|\triangle x^{k,0}\|^2. \end{aligned}$$

Part (iv) of the Lemma follows by choosing $\bar{\mu} = \frac{\mu}{2}$. □

LEMMA 5.6. *Suppose Assumptions A1-A6 and (5.1) hold. Then* $x^{k+1} = x^k + \triangle x^{k,0}$ *for all* k *large enough.*

Proof. Since $\alpha_{k,0} = 1$ for every k by Step 3.2 of Algorithm 3.1, it suffices to show that (3.17) and (3.18) hold for $j = 0$ when k is large enough. Clearly, $\{\triangle x^{k,0}\} \to 0$ by Lemma 5.5 (iii). Hence, for all large enough k, $x_{0,i}^k + \triangle x_{0,i}^{k,0} > 0$, $\forall i \in \mathcal{I} \backslash \mathcal{I}(x^*)$, $x_{1,i}^k + \triangle x_{1,i}^{k,0} > 0$, $\forall i \in \mathcal{A} \backslash \mathcal{A}_1(x^*)$ and $x_{2,i}^k + \triangle x_{2,i}^{k,0} > 0$, $\forall i \in \mathcal{A} \backslash \mathcal{A}_2(x^*)$. Moreover, by Lemma 5.5 (i), $x_{0,i}^k + \triangle x_{0,i}^{k,0} = x_{0,i}^k + \varpi_{0,i}^{k,0} = 0$, $\forall i \in \mathcal{I}(x^*)$, $x_{1,i}^k + \triangle x_{1,i}^{k,0} = x_{1,i}^k + \varpi_{1,i}^{k,0} = 0$, $\forall i \in \mathcal{A}_1(x^*)$ and $x_{2,i}^k + \triangle x_{2,i}^{k,0} = x_{2,i}^k + \varpi_{2,i}^{k,0} = 0$, $\forall i \in \mathcal{A}_2(x^*)$. Therefore, (3.17) holds for $j = 0$ and all large enough k.

Now consider (3.18). Using Taylor's theorem, we have

$$\begin{aligned} & f(x^k + \triangle x^{k,0}) \\ = \; & f(x^k) + \nabla f(x^k)^\top \triangle x^{k,0} + \tfrac{1}{2} (\triangle x^{k,0})^\top \nabla^2 f(x^k) \triangle x^{k,0} + o(\|\triangle x^{k,0}\|^2) \\ = \; & f(x^k) + \tfrac{1}{2} \nabla f(x^k)^\top \triangle x^{k,0} + \tfrac{1}{2} (\triangle x^{k,0})^\top E^* \bar{\lambda}^{k+1,0} + o(\|\triangle x^{k,0}\|^2) \\ \leq \; & f(x^k) + \tfrac{1}{2} \nabla f(x^k)^\top \triangle x^{k,0} + o(\|\triangle x^{k,0}\|^2) \\ \leq \; & f(x^k) + \sigma \nabla f(x^k)^\top \triangle x^{k,0} - \bar{\mu} (\tfrac{1}{2} - \sigma) \|\triangle x^{k,0}\|^2 + o(\|\triangle x^{k,0}\|^2) \\ \leq \; & f(x^k) + \sigma \nabla f(x^k)^\top \triangle x^{k,0}, \end{aligned} \tag{5.9}$$

where the second equality follows from (5.6), the first inequality follows from (5.8) and the fact that $\varpi^{k,0} = (E^*)^\top \triangle x^{k,0}$ and the last two inequalities follow from Lemma 5.5 (iv) and the fact that $\sigma \in (0, \frac{1}{2})$. By (5.9), (3.18) holds eventually for $j = 0$. Hence, the lemma follows. □

Combining the results in Lemma 5.5 (iii) and Lemma 5.6, we directly obtain that Algorithm 3.1 converges quadratically.

THEOREM 5.7. *Suppose Assumptions A1-A6 and (5.1) hold. Then the sequence* $\{x^k\}$ *generated by Algorithm 3.1 converges quadratically, i.e.,* $\|x^{k+1} - x^*\| = \mathcal{O}(\|x^k - x^*\|^2)$.

6. Conclusion. We have proposed an active set method for solving MPLCCs with smooth objective functions. Under the MPLCC-LICQ, we have proved that the proposed method converges to a S-stationary solution of the MPLCC. As far as we are aware, this result is stronger than any that have been obtained for NLP based regularization or decomposition methods, which only guarantee convergence to C- or M-stationary points under the MPLCC-LICQ. Assuming additional MPLCC-SOSC and strict complementarity, we have shown the asymptotic rate of convergence of our method is quadratic.

Our method has several distinguishing features compared with existing methods. First, it employs a projected Newton step with respect to the working set that is maintained at every iteration to estimate the final active set. The working set defines a subspace for the primal variables. Elements of the step direction corresponding to the subspace are determined according to the current dual iterate. Other elements are computed through a Newton system. Our method is implementable at a reasonable cost, involving mainly one matrix factorization for each major iteration. The line search used in our method is a variant of a traditional backtracking line search, that not only guarantees the feasibility of every iterate, but also ensures that at each iteration a sufficient reduction is achieved in the objective function.

There are several issues that require further research. First, can fast local convergence occur without strict complementarity? Also, fast local convergence was established under condition (5.1). Although we can always choose parameter $z_3^{\max}$ large enough so that (5.1) holds, it is important in practice to update $z_3^{\max}$ dynamically to ensure (5.1). Finally, we would like to extend the proposed active set approach to handle nonlinear constraints.

REFERENCES

[1] M. ANITESCU, *On Solving Mathematical Programs with Complementarity Constraints as Nonlinear Programs*, SIAM J. Optim., 15 (2005), pp. 1203-1236.
[2] M. ANITESCU, *Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints,* SIAM J. Optim., 16 (2005), pp. 120-145.
[3] Y. CHEN, B. F. HOBBS, S. LEYFFER AND T. S. MUNSON, *Leader-follower equilibria for electric power and NOx allowances markets,* Computational Management Science, 3 (2006), pp. 307-330.
[4] G. DEBREU, *Definite and semidefinite quadratic forms*, Econometrica, 20 (1952), pp. 295-300.
[5] A. DEMIGUEL, M. FRIEDLANDER, F. NOGALES AND S. SCHOLTES, *A two-sided relaxation scheme for mathematical programs with equilibrium constraints*, SIAM J. Optim., 16 (2005), pp. 587-609.
[6] F. FACCHINEI, H. JIANG AND L. QI, *A smoothing method for mathematical programs with equilibrium constraints*, Math. Programming, 85 (1999), pp. 107-134.
[7] M. C. FERRIS AND J. S. PANG, *Engineering and economic applications of complementarity problems*, SIAM Review, 39 (1997), pp. 669-713.
[8] A. FISCHER, *Local behaviour of an iterative framework for generalized equations with nonisolated solutions.* Math. Programming, 94 (2002), 91-124.
[9] R. FLETCHER AND S. LEYFFER, *Solving mathematical program with complementarity constraints as nonlinear programs*, Optim. Methods and Softw., 19 (2004), pp. 15C40.
[10] R. FLETCHER, S. LEYFFER, D. RALPH AND S. SCHOLTES, *Local convergence of SQP methods for mathematical programs with equilibrium constraints*, SIAM J. Optim., 17 (2006), pp. 259-286.
[11] A. FORSGREN, *Inertia-controlling factorizations for optimization algorithms*, Appl. Numer. Math., 43 (2002), pp. 91-107.
[12] ———, *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints,* Ill-posed Variational Poroblems and Regularization Techniques, Edited by M. Théra and R. Tichatschke, Springer, New York, NY, pp. 99-110, 2000.
[13] M. FUKUSHIMA AND P. TSENG, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints*, SIAM J. Optim., 12 (2002), pp. 724-739.

[14] ———, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints: erratum*, to appear in SIAM J. Optim., 2006.

[15] G. Giallombardo and D. Ralph, *Multiplier convergence in trust-region methods with application to convergence of decomposition methods for MPECs*, Online first, Math. Programming, 2006.

[16] X. M. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with complementarity constraints*, J. Optim. Theory Appl., 123 (2004), pp. 365-390.

[17] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM J. Optim., 10 (2000), pp. 779-808.

[18] C. Kanzow and H. Qi, *A QP-free constrained Newton-type method for variational inequality problems,* Math. Programming, 85 (1999), pp. 81-106.

[19] J. J. Moré and D. C. Sorensen, *Computing a trust region step,* SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553-572.

[20] S. Leyffer, G. López-Calva and Jorge Nocedal, *Interior methods for mathematical programs with complementarity constraints*, SIAM J. Optim., 17 (2006), pp. 52-77.

[21] G. Lin and M. Fukushima, *New relaxation method for mathematical programs with complementarity constraints*, J. Optim. Theory Appl., 118 (2003), pp. 81-116.

[22] G. Liu and J. J. Ye, *A merit function piecewise SQP algorithm for solving mathematical programs with equilibrium constraints*, to appear in J. Optim. Theory Appl., 2006.

[23] X. Liu and J. Sun, *Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints*, Math. Programming, 101 (2004), pp. 231-261.

[24] Z. Luo, J. S. Pang and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, UK, 1996.

[25] ———, *Piecewise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints*, A. Migdalas, P. M. Pardalos and P. Värbrand, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 209-229.

[26] J. Outrata, M. Kocvara and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints*, Kluwer Academic Publisher, Dordrecht, 1998.

[27] A. U. Raghunathan and L. T. Biegler, *Interior point methods for mathematical programs with complementarity constraints*, SIAM J. Optim., 15 (2005), pp. 720-750.

[28] D. Ralph, *Sequential quadratic programming for mathematical programs with linear complementarity constraints*, Computational Techniques and Applications: CTAC95 (Melbourne, 1995), pp. 663-668, World Scientific, River Eege, NJ, 1996.

[29] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43-62.

[30] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1-22.

[31] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918-936.

[32] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints,* SIAM J. Control Optim., 37 (1999), pp. 617-652.

[33] D. F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: orderings and high-order methods*, Math. Programming, 87 (2000), pp. 303-316.

[34] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Programming, 106 (2006), pp. 25-57.