To appear in *Computational Optimization and Applications*

Extension of Quasi-Newton Methods to Mathematical Programs with Complementarity Constraints[1]

Houyuan Jiang [a] and Daniel Ralph [b]

[a] CSIRO Mathematical and Information Sciences
GPO Box 664, Canberra ACT 2601, Australia
Houyuan.Jiang@csiro.au

[b] The Judge Institute of Management
University of Cambridge, Trumpington St, CB2 1AG, UK
danny.ralph@jims.cam.ac.uk

Abstract. Quasi-Newton methods in conjunction with the piecewise sequential quadratic programming are investigated for solving mathematical programming with equilibrium constraints, in particular for problems with complementarity constraints. Local convergence as well as superlinear convergence of these quasi-Newton methods can be established under suitable assumptions. In particular, several well-known quasi-Newton methods such as BFGS and DFP are proved to exhibit the local and superlinear convergence.

Key Words. Mathematical programs with equilibrium constraints, MPEC, MPCC, complementarity problem, complementarity constraints, piecewise sequential quadratic programming, quasi-Newton method, superlinear convergence.

Dedication: We dedicate this paper to Professor Elijah (Lucien) Polak, a great optimizer, on the occasion of his 72nd birthday. In addition to Lucien's energy, shown by his many papers and books, his outward looking nature has yielded contact with diverse topics such as problems with maxmin constraints [21], which in fact generalise mathematical programs with complementarity constraints. Typical of his engaging style, Lucien is now working with the second author on optimization of maxmin problems. We congratulate him on his birthday, on his wonderful career, and on future work!

1 Introduction

Mathematical programs with complementarity constraints (MPCC) are optimization problems whose constraints are defined by traditional equalities and inequalities as well as parametric complementarity problems. We study MPCC with smooth data functions, i.e. nonlinear programs with parametric complementarity constraints:

$$\begin{array}{ll} \min & f(x,y) \\ \text{subject to} & 0 \le g(x,y), \\ & 0 \le F(x,y) \ \perp \ y \ge 0, \end{array} \tag{1}$$

[1] This work is supported by the Australian Research Council.

where $f : \Re^{n+m} \to \Re$, $g : \Re^{n+m} \to \Re^l$ and $F : \Re^{n+m} \to \Re^m$ are C^2 (twice continuously differentiable), and $a \perp b$ denotes orthogonality of the vectors a and b. The orthogonality condition is called complementarity because orthogonality of nonnegative vectors a and b implies that the ith components a_i and b_i satisfy either $a_i = 0$ or $b_i = 0$. The two vectors x and y play different roles. Often, x is called the leader variable, and y is called the follower variable. We omit equality constraints $h(x, y) = 0$ in the formulation (1) since these smooth equalities pose no fundamental mathematical difficulties but burden the notation.

Mathematical programs with equilibrium constraints (MPEC) generalise MPCC by allowing so-called equilibrium constraints such as parametric variational inequalities. MPEC, and especially MPCC, have found many applications in engineering and economics. Many of these applications are collected in [10, 25, 28]. If the equilibrium constraints of an MPEC happen to be first-order optimality conditions, such as the classical Karush-Kuhn-Tucker (KKT) conditions, of a parametric convex mathematical program, then the MPEC is equivalent to a bilevel program, which represents another important field of optimization [1, 2, 40].

We also mention that MPCCs form a subclass of nonlinear programs with maxmin constraints [21]. The connection is made by using a "min" formulation of the complementarity constraints. Retaining the inequalities $F(x, y)$, $y \geq 0$, the complementarity (orthogonality) of $F(x, y)$ and y is equivalent to $\min\{F_i(x, y), y_i\} \leq 0$ for each component i, or $\max_i \min\{F_i(x, y), y_i\} \leq 0$. For further connections in this direction see [38], which proposes a decomposition method for problems with combinatorial constraints that can be applied to MPCCs, and which is related to the piecewise programming approach described in Section 4.

Among others, the development of efficient numerical methods for solving MPCC includes specialized techniques such as [9, 13, 14, 15, 18, 25, 26, 27, 28, 29, 36, 37, 39]. Standard nonlinear program (NLP) solvers have also been applied directly to MPCC in [2, 12, 17, 24], e.g., by writing the complementarity condition of (1) as the smooth constraint $y^T F(x, y) = 0$. Looking at (1) as a nonlinear program is not as straightforward as it seems however. For instance it has been observed [6, 41] that the Mangasarian-Fromowitz constraint qualification, and hence the linear independence constraint qualification, is violated at every feasible point, which suggests that theoretical and numerical difficulties might be encountered when applying standard nonlinear programming techniques. See [12, 17] for further discussion and numerical results.

We propose a quasi-Newton version of the piecewise sequential quadratic programming (PSQP) method, a decomposition algorithm for MPEC first proposed in [33], extended in [25, 26], and studied computationally for quadratic programs with equilibrium constraints in [17]. PSQP is a direct generalization of the sequential quadratic programming (SQP) method from nonlinear programming to MPEC. Like SQP for nonlinear programs, PSQP requires the evaluation of second-order derivatives, which might be unavailable or expensive to obtain, and may not work well if certain convexity (second-order sufficient) conditions are not satisfied. The quasi-Newton approach to NLP is a very successful remedy for the first of these two potential difficulties, see, e.g., [11].

The aim of this paper is to show that the quasi-Newton PSQP method, QN-PSQP, achieves local superlinear convergence under reasonable conditions. Apart from previous work on PSQP mentioned above, we build directly on the convergence analysis of quasi-Newton methods for variational inequalities and nonlinear programs due to Bonnans [4]. PSQP decomposes the MPCC feasible set into branches which have better stability properties than the feasible set as a whole. It is well known that a linear independence condition, MPCC-LICQ, at a local minimizer z^* of the MPCC is sufficient to ensure that each branch shares the same KKT multiplier. This implies that the same second-order information, i.e. Hessian of the Lagrangean, is appropriate for each branch. This motivates the success of QN-PSQP: at each iteration a single approximation of the Hessian of the Lagrangean is shown to be equally appropriate for each relevant branch, leading to local Q-superlinear convergence.

The paper is structured as follows. In the next Section, we review the definition of MPEC, and recall some basic facts associated with decompositions, MPEC constraint qualifications, and first-order and second-order conditions. In Section 3, the SQP method for nonlinear programs is reviewed, convergence results are collected, and some new results are established. Section 4 is devoted to the study of local linear convergence of quasi-Newton methods for MPCC. A sufficient condition for local superlinear convergence is proved in Section 5. In Section 6, we present two concrete quasi-Newton methods, BFGS and DFP. The results of Sections 4 and 5 are applied here to establish local and superlinear convergence rates of quasi-Newton PSQP methods. Section 7 is concerned with quasi-Newton methods for MPCC with parametric linear complementarity constraints.

Notation used in the sequel is explained in order. If $f(x)$ is a twice continuously differentiable function from $\Re^N$ to $\Re$, then ∇f and $\nabla^2 f$ denote its first derivative (gradient) and second derivative (Hessian) mappings, respectively; partial derivatives are denoted using an appropriate subscript. If $F : \Re^N \to \Re^M$ is a vector-valued continuously differentiable function, then F' indicates its Jacobian.

2 Preliminaries

2.1 Decomposition of the MPCC feasible set

The quasi-Newton methods to be presented are based on the so-called `decomposition` or `disjunctive` or `piecewise` technique, which has been used for studying first- and second-order optimality conditions and algorithms for MPEC, e.g. [25, 26]. We give a brief review.

We begin with notation for the feasible set of (1):

$$\begin{aligned} \mathcal{F} \;=\; \{(x,y) \;:\; & g(x,y) \geq 0, \\ & 0 \leq F(x,y) \;\perp\; y \geq 0\}. \end{aligned}$$

Now, for any $z = (x,y)$, decompose the index set $\{1,2,\ldots,m\}$ into three disjoint subsets

$$\begin{aligned} \alpha(z) &= \{1 \leq i \leq m : F_i(z) < y_i\} \\ \beta(z) &= \{1 \leq i \leq m : F_i(z) = y_i\} \\ \gamma(z) &= \{1 \leq i \leq m : F_i(z) > y_i\} \end{aligned}$$

and define the decomposition index set $\mathcal{A}(z)$ at z by

$$\begin{aligned} \mathcal{A}(z) \;=\; \{(\mathcal{J},\mathcal{K}) :\; & \mathcal{J} \supseteq \alpha(z), \mathcal{K} \supseteq \gamma(z), \\ & \mathcal{J} \cup \mathcal{K} = \{1,2,\ldots,m\}, \mathcal{J} \cap \mathcal{K} = \emptyset\}. \end{aligned}$$

Any $i \in \beta(z)$ is called a degenerate index at z. Clearly, the cardinality of the decomposition index set $\mathcal{A}(z)$, denoted A, is determined by the cardinality b of the degenerate index $\beta(z)$: $A = 2^b$. If strict complementarity holds at z, i.e., $\beta(z) = \emptyset$, then $\mathcal{A}(z)$ contains exactly one element.

If $z^* = (x^*, y^*)$ happens to be a feasible point of the MPCC then $F_i(z^*) y_i^* = 0$ for each $i \in \{1,2,\ldots,m\}$. Consequently, the index sets $\alpha(z^*)$, $\beta(z^*)$ and $\gamma(z^*)$ at z^* satisfy

$$\begin{aligned} \alpha(z^*) &= \{1 \leq i \leq m : F_i(z^*) = 0 < y_i^*\} \\ \beta(z^*) &= \{1 \leq i \leq m : F_i(z^*) = 0 = y_i^*\} \\ \gamma(z^*) &= \{1 \leq i \leq m : F_i(z^*) > 0 = y_i^*\}. \end{aligned}$$

For each $(\mathcal{J},\mathcal{K}) \in \mathcal{A}(z^*)$, we may define a constraint set:

$$\begin{aligned}\mathcal{F}_{(\mathcal{J},\mathcal{K})} = \{z : \quad & 0 \le g(x,y), \\ & F_i(z) = 0 \le y_i, \quad \forall i \in \mathcal{J}, \\ & F_i(z) \ge 0 = y_i, \quad \forall i \in \mathcal{K} \ \}.\end{aligned}$$

Each such $\mathcal{F}_{(\mathcal{J},\mathcal{K})}$ is called a `branch` or `piece` of $\mathcal{F}$ at z^*.

The decompositions at a given point and the nearby points have the following relationship which will be used in establishing convergence of the quasi-Newton method. These facts are well known, e.g. [25], but we give the proof since it is immediate.

Lemma 2.1

(i) Let $z^* \in \Re^{n+m}$. Then $\mathcal{A}(z) \subset \mathcal{A}(z^*)$ for z near z^*.

(ii) If $z^* \in \mathcal{F}$ then there exists a neighbourhood Z of z^* such that

$$\mathcal{F} \cap Z = \bigcup_{(\mathcal{J},\mathcal{K}) \in \mathcal{A}(z^*)} \mathcal{F}_{(\mathcal{J},\mathcal{K})} \cap Z.$$

Proof. Part (ii) follows directly from part (i). For part (i), observe that continuity of the function F gives a neighbourhood Z of z^* such that for any $z \in Z$ we have

$$\alpha(z^*) \subseteq \alpha(z), \quad \gamma(z^*) \subseteq \gamma(z).$$

By the definition, for $z \in Z$ it follows that any $(\mathcal{J},\mathcal{K}) \in \mathcal{A}(z)$ is also a member of $\mathcal{A}(z^*)$. □

2.2 Stationarity conditions and constraint qualifications

Let $z^* \in \mathcal{F}$. The tangent cone, or set of limiting feasible directions, to $\mathcal{F}$ at z^* is the set defined by

$$\mathcal{T}(z^*, \mathcal{F}) \ = \ \{\lim (z^k - z^*)/\tau^k : \{z^k\} \subset \mathcal{F}, \, z^k \to z^*, \tau^k \downarrow 0\}.$$

Let $dz = (dx, dy) \in \Re^{n+m}$. A natural first-order necessary condition (or stationarity condition) for the feasible point z^* to be a local minimizer of (1) is

$$\nabla f(x^*, y^*)^T (dx, dy) \ge 0, \quad \forall dz \in \mathcal{T}(z^*, \mathcal{F}). \tag{2}$$

This is referred to as a `primal` stationarity condition in [25].

As studied in [25], under some MPEC constraint qualifications, the tangent cone $\mathcal{T}(z^*, \mathcal{F})$ coincides with a certain linearized cone, formulated by linearizing the active inequalities corresponding to g, F and y in the definition of $\mathcal{F}$. These cones are generally nonconvex, though they can be written as the union of a finite but possibly huge number of convex polyhedral cones. In spite of this combinatorial structure, under a linear independence assumption on the active constraints of g, F and y, checking stationarity of (1) is equivalent to checking stationarity of a related nonlinear program [26, 35].

The next result, Proposition 2.1, is a formal statement of some easy facts from [25] which are related to Lemma 2.1, for instance part (i) of the Proposition is a corollary of part 2 of the Lemma. Proposition 2.1 says that stationarity conditions of (1) can be characterized in terms of traditional the nonlinear programs associated with each branch at z^*.

We need a little more notation. A `piecewise constraint qualification` at z^* [25] requires a traditional NLP constraint qualification on each branch of $\mathcal{F}$ at z^*, which in turn assures the existence of KKT multipliers if z^* is a local minimizer of the associated NLP, see (3) below. The `MPCC-Lagrangean` is the function of $z = (x, y) \in \Re^{n+m}$ and $\lambda = (\xi, \eta, \pi) \in \Re^{l+m+m}$,

$$L(z, \lambda) \;=\; f(x, y) - \xi^T g(x, y) - \eta^T F(x, y) - \pi^T y.$$

Proposition 2.1 ([25]) *Let* $z^* \in \mathcal{F}$.

(i) The point z^* *is a local minimizer of (1) if and only if, for each* $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$, z^* *is a local minimizer of the nonlinear program*

$$\begin{array}{ll} \min & f(z) \\ \text{subject to} & z \in \mathcal{F}_{(\mathcal{J},\mathcal{K})}. \end{array} \tag{3}$$

(ii) Under a piecewise constraint qualification at z^**, the point* z^* *is a primal stationary point of (1) if and only if, for each* $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$*, it is stationary for the NLP (3), i.e. (in addition to feasibility* $z^* \in \mathcal{F}_{(\mathcal{J},\mathcal{K})}$*), there exist multipliers* $\xi \in \Re^l$, $\eta \in \Re^m$ *and* $\pi \in \Re^m$ *such that*

$$\begin{array}{lll} \nabla_z L(z^*, \xi, \eta, \pi) = 0 & & \\ \xi \geq 0, & g(x^*, y^*)^T \xi = 0, & \\ \pi_i \geq 0, & y_i^* \pi_i = 0, & \forall i \in \mathcal{J}, \\ \eta_i \geq 0, & F_i(x^*, y^*) \eta_i = 0, & \forall i \in \mathcal{K}. \end{array} \tag{4}$$

The stationarity condition in part (ii) of Proposition 2.1 is known as `piecewise stationarity` [25] or `B-stationarity` [35].

In nonlinear programming, the Linear Independence Constraint Qualification (LICQ) is sufficient for existence and uniqueness of the KKT multiplier vector at a local minimizer. At a stationary point of an NLP, the Strict Mangasarian Fromowitz Constraint Qualification (SMFCQ) of [23] is equivalent to uniqueness of the KKT multiplier and is, in general, weaker than the LICQ. These ideas also have meaning for MPCC.

Let $z^* \in \mathcal{F}$. Let $I_g(z^*)$ denote the active index set of the constraint $g(x, y) \geq 0$ at z^*,

$$I_g(z^*) \;=\; \{i : g_i(x^*, y^*) = 0, i = 1, \ldots, l\}.$$

The `MPCC Linear Independence Constraint Qualification` (MPCC-LICQ) is said to hold at z^* if the active constraints at z^* of (1), ignoring the complementarity condition, have linearly independent gradients, that is the vectors

$$\begin{array}{ll} & \{(0, e^i) \in \Re^n \times \Re^m : i \in \beta(z^*) \cup \gamma(z^*)\} \\ \cup & \{F_i'(z^*) : i \in \beta(z^*) \cup \alpha(z^*)\} \\ \cup & \{g_i'(z^*) : i \in I_g(z^*)\} \end{array}$$

are linearly independent, where $e^i \in \Re^m$ indicates the vector of zeros except for a one in its ith component.

It can easily be seen that the MPCC-LICQ holds at z^* if and only if the LICQ holds for any or every branch (3) at z^*. It is also known [25, 26, 35] at a local minimizer z^* of an MPCC, that the LICQ is sufficient, to ensure that

(i) for each $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$, there exists a unique multiplier $\lambda = (\xi, \eta, \pi)$ satisfying (4), and

(ii) λ is independent of the branch index $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$.

A multiplier λ with property (ii) is said to be a `common` multiplier of the MPCC, since it is a KKT multiplier for each branch NLP (3) at z^*. The existence of a common multiplier is easily seen to be equivalent to `strong B-stationarity`, [35] (actually, [35] coins the term "strongly stationary"). See [41] for weaker conditions for the existence of common multipliers.

The `MPCC strict Mangasarian-Fromowitz Constraint Qualification` (MPCC-SMFCQ) is said to hold at a feasible point $z^* \in \mathcal{F}$ if there exists a multiplier λ satisfying (i) and (ii) above. The preceding remarks make it clear that the MPCC-LICQ implies the MPCC-SMFCQ. However the converse is not true, as can be verified from the following example in which $z^* = (0, 0)$ and the unique KKT multiplier for each of the two branches at z^* is the zero vector, whereas the MPCC-LICQ is violated since there are three active gradients in a 2-dimensional space:

$$\begin{array}{ll} \displaystyle\min_{z=(x,y)\in\Re^2} & x^2 + y^2 \\ \text{subject to} & 0 \le -x + 2y \\ & 0 \le y - x \;\perp\; y \ge 0. \end{array}$$

We also point out that a slightly different SMFCQ for MPCC was proposed by Scheel and Scholtes [35], by applying the usual SMFCQ to a "tightened NLP". The SMFCQ for MPCC in [35] is strictly stronger than the above MPCC-SMFCQ as can be checked using the above example.

2.3 Piecewise second-order sufficient conditions

Parts (i) and (ii) of Proposition 2.1 give first-order necessary optimality conditions for the MPCC (1). Second-order necessary and sufficient optimality conditions that use standard NLP tools, such as the critical cone reviewed in Section 3, have also been developed; see [25] for example. For instance, suppose the MPCC-SMFCQ holds for the MPCC at a given feasible point $z^* = (x^*, y^*)$, and denote by $\lambda = (\xi, \eta, \pi)$ its unique common multiplier. For any $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$, denote the critical cone of $\mathcal{F}_{(\mathcal{J},\mathcal{K})}(z^*)$ by $\mathcal{C}(\mathcal{J}, \mathcal{K})$ (See Section 3 for its definition). The usual second-order sufficient condition (SOSC) for the z^* to be a local minimizer of the branch NLP (3) is

$$d^T \nabla^2_{zz} L(z^*; \lambda) d > 0, \quad \forall d \in \mathcal{C}(\mathcal{J}, \mathcal{K}) \text{ such that } d \neq 0,$$

where $\lambda = (\xi, \eta, \pi)$. The `piecewise second-order sufficient condition` (piecewise SOSC) is said to be satisfied for the MPCC at z^* if the SOSC holds for each branch of the decomposition at z^*. Given the decomposition in part (ii) of Lemma 2.1, it is obvious that a feasible point z^* satisfying both the MPCC-SMFCQ and the piecewise SOSC is a local minimizer of MPCC (1).

3 The SQP Method for Nonlinear Programs

In this section we review quasi-Newton methods for nonlinear programs, from the usual viewpoint of sequential quadratic programming. Some new results will also be established. Without confusion, let us abuse our notation temporarily in this section for convenience of exposition. This means that the notation used in this section in the context of NLP may have different meanings in other sections in the context of MPCC. Consider the nonlinear program

$$\begin{array}{ll} \min & f(z) \\ \text{subject to} & g(z) \geq 0 \\ & h(z) = 0, \end{array} \tag{5}$$

where $f : \Re^n \to \Re$, $g : \Re^n \to \Re^{m_1}$, $h : \Re^n \to \Re^{m_2}$ are twice continuously differentiable

Assume that z^* is a stationary point of the above nonlinear program, and that the strict Mangasarian-Fromowitz constraint qualification (SMFCQ) and the second-order sufficient condition (SOSC) hold at z^*. Denote by $\lambda^* = (\lambda_g^*, \lambda_h^*) \in R^{m_1+m_2}$ the unique KKT multiplier at z^*. For definitions of the LICQ and the SOSC and other unexplained concepts and results regarding NLP, see the textbook [11]. We remark that all these concepts should coincide with the corresponding parts in the last section if the complementarity constraint in (1) is omitted. As a consequence of the LICQ, the multiplier λ^* is unique.

The local version of the traditional sequential quadratic programming method for NLP is usually expressed as follows: Given the current iterate z and the current estimate of the KKT multipliers (λ_g, λ_h), generate the next iterate by

$$z^+ = z + d \tag{6}$$

where d is a solution of the following quadratic program

$$\begin{array}{ll} \min\limits_{d} & \nabla f(z)^T d + \frac{1}{2} d^T B d \\ \text{subject to} & g(z) + g'(z) d \geq 0 \\ & h(z) + h'(z) d = 0, \end{array} \tag{7}$$

and B is a symmetric (possibly positive definite) matrix, which is an approximation of the second-order derivative with respect to z of the Lagrangean function

$$L(z, \lambda_g, \lambda_h) \;=\; f(z) - \lambda_g^T g(z) - \lambda_h^T h(z).$$

One immediate consequence associated with the above algorithm is the following useful lemma.

Lemma 3.1 *Suppose that the SMFCQ and the SOSC hold at a local solution z^* of the above nonlinear program. Let $w \equiv (z, \lambda) = (z, \lambda_g, \lambda_h)$. Then for any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|w - w^*\| \leq \varepsilon(r)$ and $\|B - \nabla^2_{zz} L(w^*)\| \leq \delta(r)$, then*

$$\|w^+ - w^*\| \leq r \|w - w^*\|$$

where $w^+ = (z + d, \lambda^+)$ and λ^+ is a multiplier of the QP (7).

The conclusion of the above lemma was established by Han [16] under the stronger assumptions of the LICQ, the SOSC and the strict complementarity condition (SCC) at the solution z^*. Josephy's remarkable results [19, 20] on Newton's method for generalized equations make it possible to remove the SCC at z^* under the strong SOSC [34]. Furthermore, the LICQ can be replaced by the SMFCQ. Indeed, the above lemma is implicitly proved by Bonnans [4]. See Theorem 2.3 and Lemma 2.2 and their proofs of [4]. A similar result to Lemma 3.1 is also established in [32] in the framework of nonsmooth equation reformulations of nonlinear programs.

In fact, [4] established more general results on superlinear convergence of Newton-type methods for variational inequality problems by using his concepts of semistablility and hemistability which are equivalent to the SMFCQ and the SOSC in the context of nonlinear programs. Similar results were obtained independently by Pang [30].

Superlinear convergence is an important feature of quasi-Newton methods for nonlinear programs. The characterizations of superlinear convergence of quasi-Newton methods for unconstrained optimization and smooth nonlinear equations are first obtained by Dennis and Moré [7]. Later on, this theory has been extended to general nonlinear programs [16, 3], generalized equations [20], variational inequalities [4] and nonsmooth equations [22, 31, 32]. Before recalling such a characterization given for NLP in [4, Theorem 6.2], we introduce some notation for NLP at the solution z^* as follows.

The active, strongly active and weakly active index sets associated with (z^*, λ^*) are

$$\begin{array}{rcl} I &=& \{i : g_i(z^*) = 0, i = 1, \ldots, m_1\}, \\ I^+ &=& \{i \in I : \lambda_i^* > 0\}, \\ I^0 &=& I \setminus I^+ = \{i \in I : \lambda_i^* = 0\}, \end{array}$$

The critical cone for this problem at the stationary point z^* is defined to be

$$\begin{array}{rcll} \mathcal{C}(z^*) &=& \{d \in \Re^n &:& g_i'(z^*)d = 0, i \in I^+, \\ &&&& g_i'(z^*)d \geq 0, i \in I^0, \\ &&&& h_j'(z^*)d = 0, j = 1, \ldots, m_2\}. \end{array}$$

At iteration k of the method, given z^k, the following cone is defined in [4]:

$$\begin{array}{rcll} \mathcal{C}_k &=& \{d \in \Re^n &:& g_i'(z^k)d = 0, i \in I^+, \\ &&&& g_i'(z^k)d \geq 0, i \in I^0, \\ &&&& h_j'(z^k)d = 0, j = 1, \ldots, m_2\}. \end{array} \tag{8}$$

We call this the `approximate critical cone` for later use.

Lemma 3.2 ([4]) *Suppose that z^* is a local solution of the NLP (5), that the LICQ and the SOSC hold at z^*. Let $\{w^k = (z^k, \lambda^k)\}$ be generated by the SQP method and converge to $w^* = (z^*, \lambda^*)$. Then $\{z^k\}$ converges superlinearly if and only if*

$$\|P_k[(\nabla^2_{zz} L(z^*, \lambda^*) - B_k)d^k]\| = o(\|d^k\|),$$

where P_k indicates the Euclidean projection on the set $\mathcal{C}_k$, and $\lambda = (\lambda_g, \lambda_h)$ are Lagrange multipliers associated with the constraints.

The above characterization of superlinear convergence for quasi-Newton methods for NLP reduces to the well-known characterization of Boggs, Tolle and Wang [3] if there are no inequality constraints. We show that the above lemma is still valid if the LICQ is weakened to the SMFCQ. We state this claim in the following proposition and give a proof in the appendix. Note that our proof is similar to that of [4] but with some technical differences since we do not assume the LICQ.

Proposition 3.1 *Lemma 3.2 still holds when the LICQ is replaced by the SMFCQ.*

Remark 3.1 For convenience of the subsequent exposition, we illustrate Lemma 3.2 as follows. The fact that z^k converges superlinearly to z^* can be equivalently stated as

$$\|z^{k+1} - z^*\| \ = \ o(\|z^k - z^*\|),$$

i.e., there exists a function $\Delta : [0, \infty) \to [0, \infty)$ satisfying $\lim_{t \downarrow 0} \frac{\Delta(t)}{t} = 0$ such that for all sufficiently large k

$$\|z^{k+1} - z^*\| \leq \Delta(\|z^k - z^*\|).$$

When all equality and inequality constraints are linear, the assumptions in Lemma 3.1 , 3.2 and Proposition 3.1 can be weakened further.

Lemma 3.3 ([4]) *Suppose g and h are linear, and the SOSC holds at the local solution z^* of the NLP (5). For any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|z - z^*\| \le \varepsilon(r)$ and $\|B - \nabla^2 f(z^*)\| \le \delta(r)$, then*

$$\|z^+ - z^*\| \le r\|z - z^*\|$$

where $z^+ = z + d$ and d is a solution of (7).

Proof. It is known that linearly constrained nonlinear programs have KKT conditions that are equivalent to variational inequalities over polyhedral convex sets. Thus the conclusion follows directly from [4]. This is because the SOSC at a point implies both semistability and hemistability at the same point. □

The following proposition presents a characterization for superlinear convergence of the SQP method for linearly constrained nonlinear programs. A proof can be found in the Appendix, which is again similar to that given in [4]. Also, Josephy's early investigations [19, 20] can be applied to generalized equations with polyhedral constraints.

Proposition 3.2 *Suppose g and h are linear, and the SOSC holds at the local solution of z^* of the nonlinear program (5). Let $\{z^k\}$ converge to z^*. Then $\{z^k\}$ converges superlinearly if and only if*

$$\|P_k[(\nabla^2 f(z^*) - B_k)d^k]\| = o(\|d^k\|),$$

where P_k indicates the Euclidean projection on the set $\mathcal{C}_k$ defined in (8).

4 Quasi-Newton Methods and Local Linear Convergence

PSQP is an extension of SQP methods of nonlinear programs to MPCC. Applying the decomposition technique to MPCC, Proposition 2.1(i) shows that solving (1) is equivalent to solving finitely many nonlinear programs. Consequently, SQP methods can be applied to some of nonlinear programs (3). Therefore we may propose the following quasi-Newton methods in conjunction with PSQP.

Algorithm: QN-PSQP

Step 1. Choose a starting point $z^0 = (x^0, y^0)$ and a positive symmetric matrix $B_0 \in \Re^{(m+n)\times(m+n)}$. Let $k = 0$

Step 2. Choose any $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^k)$.

Step 3. Let $d = d^k$ solve the quadratic program:

$$\begin{array}{ll} \min\limits_{d} & \nabla f(x^k, y^k)^T d + \frac{1}{2} d^T B_k d \\ \text{subject to} & d \in \mathcal{LF}_{(\mathcal{J},\mathcal{K})}(z^k), \end{array} \tag{9}$$

where $\mathcal{LF}_{(\mathcal{J},\mathcal{K})}(z^k)$ is the linearization of $\mathcal{F}_{(\mathcal{J},\mathcal{K})}(z^k)$ at z^k:

$$\begin{array}{rcl}\mathcal{LF}_{(\mathcal{J},\mathcal{K})}(z^k) & = & \{d=(dx,dy):\\ & & g(z^k)+\nabla g(z^k)^T d \geq 0\\ & & F_i(z^k)+F_i'(z^k)d=0,\ y_i^k+dy_i\geq 0, \quad \forall i\in\mathcal{J}\\ & & F_i(z^k)+F_i'(z^k)d\geq 0,\ y_i^k+dy_i=0, \quad \forall i\in\mathcal{K}\}.\end{array}$$

Step 4. Let $z^{k+1} = z^k+d^k$, $k = k+1$. Choose a positive symmetric matrix $B_k \in \Re^{(m+n)\times(m+n)}$. Go to step 2.

Remark. When $\lambda^k = (\xi^k,\eta^k,\pi^k)$ is the KKT multiplier vector corresponding to d^k in solving (9), and B_k is defined as the second derivative with respect to z of the associated Lagrangean at (z^k,λ^k), the above algorithm reduces to the PSQP method proposed in either [33] for MPCC with affine constraint functions g and F (c.f. Section 7 to follow), or [25] for nonlinear MPCC. For the case of affine constraint functions, if z^k is stationary for the current branch $(\mathcal{J},\mathcal{K})\in\mathcal{A}(z^k)$ but not B-stationary for the MPCC, and the MPCC-LICQ holds at z^k, the papers [39] and later [37] show how to use the multipliers to determine the next branch in which z^k is feasible but non-stationary. Local convergence and superlinear convergence of PSQP are established respectively under certain conditions [25, 26, 33].

We make some standing assumptions which will be used throughout this section.

Assumption 1 The point $z^*=(x^*,y^*)$ is a primal stationary or B-stationary point for the above MPCC (1) at which the MPCC-SMFCQ holds and the associated multiplier is λ^*. Let $w^*=(z^*,\lambda^*)$.

Assumption 2 The piecewise SOSC holds at z^* for (1).

Proposition 4.1 *Let Assumptions 1 and 2 hold. For any $r\in(0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|w-w^*\|\leq\varepsilon(r)$ and $\|B-\nabla^2_{zz}L(w^*)\|\leq\delta(r)$, then*

$$\|w^+-w^*\|\leq r\|w-w^*\|, \tag{10}$$

where $w^+=(z+d,\lambda^+)$, d is a solution of the QP (9) corresponding to any branch of the decomposition of the MPCC at w, and λ^+ is one of its multipliers.

Proof. The proof of this proposition is a direct application of Lemma 3.1. Clearly, the following statements hold:

- There are only finitely many branches $(\mathcal{J},\mathcal{K})$ of the decomposition of the MPCC at the
- Lemma 2.1 says that any branch $(\mathcal{J},\mathcal{K})\in\mathcal{A}(z)$ is a also a branch of the decomposition at z^* as long as z is sufficiently close to z^*.

Note that z^* is a solution of each NLP associated with a branch of the decomposition of the MPCC at z^* by (i) of Proposition 2.1. Moreover, the MPCC-SMFCQ and the piecewise SOSC imply that the SMFCQ and the SOSC hold for each NLP at z^*. Applying the result in Lemma 3.1 to every NLP at z^*, there exist $\varepsilon_{(\mathcal{J},\mathcal{K})}(r)$ and $\delta_{(\mathcal{J},\mathcal{K})}(r)$ such that if $\|w-w^*\|\leq\varepsilon_{(\mathcal{J},\mathcal{K})}(r)$ and $\|B-\nabla^2_{zz}L(w^*)\|\leq\delta_{(\mathcal{J},\mathcal{K})}(r)$, then

$$\|w^+-w^*\|\leq r\|w-w^*\|.$$

Define

$$\varepsilon(r) = \min_{(\mathcal{J},\mathcal{K})\in\mathcal{A}(z^*)} \varepsilon_{(\mathcal{J},\mathcal{K})}(r), \quad \delta(r) = \min_{(\mathcal{J},\mathcal{K})\in\mathcal{A}(z^*)} \delta_{(\mathcal{J},\mathcal{K})}(r).$$

Then we have common positive constants $\varepsilon(r)$ and $\delta(r)$ for all branches of the decomposition at z^* such that if $\|w - w^*\| \leq \varepsilon(r)$ and $\|B - \nabla^2_{zz}L(w^*)\| \leq \delta(r)$, then

$$\|w^+ - w^*\| \leq r\|w - w^*\|,$$

no matter what branch of the decomposition of the MPCC at z^k is chosen. This completes the proof. □

Based on the above proposition, we are able to establish the following proposition which is essentially a generalization of local convergence results for classical SQP [16, 4]. Let $\|\cdot\|'$ denote any matrix norm which may be different from the matrix norm $\|\cdot\|$.

Proposition 4.2 *Let Assumptions 1 and 2 hold. Assume there exist two positive constants μ_1 and μ_2 such that*

$$\|B_{k+1} - \nabla^2_{zz}L(w^*)\|' \leq (1 + \mu_1\sigma^k)\|B_k - \nabla^2_{zz}L(w^*)\|' + \mu_2\sigma^k,$$

with $\sigma^k = \max\{\|w^{k+1} - w^*\|, \|w^k - w^*\|\}$. Then for any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|w - w^*\| \leq \varepsilon(r)$ and $\|B_0 - \nabla^2_{zz}L(w^*)\| \leq \delta(r)$ are satisfied, then the sequence $\{w^k\}$ generated by QN-PSQP with respect to a sequence of positive symmetric matrices $\{B_k\}$ is well defined and converges Q-linearly to w^*.

Proof. The proof is an easy exercise of the induction method in view of Proposition 4.1. We refer the reader to [16] for a detailed proof for the SQP method for NLP. □

The last proposition states that the sequence $\{w^k\}$ generated by QN-PSQP converges locally to w^* under suitable conditions provided that the approximate Jacobian matrix $\{B_k\}$ is updated in a reasonably controlled way, which is usually called the bounded deterioration condition [8]. The quasi-Newton approach can also be implemented via the inverse of the Jacobian, as in NLP [8]. We express the latter idea in the following result for QN-PSQP.

Proposition 4.3 *Suppose that the MPCC-SMFCQ and the piecewise SOSC hold at the local solution $z^* = (x^*, y^*)$ of the MPCC, and that λ^* is the unique multiplier associated with z^*. Let $w^* = (z^*, \lambda^*)$. Assume that $\nabla^2_{zz}L(w^*)$ is nonsingular, and there exist two positive constants μ_1 and μ_2 such that*

$$\|H_{k+1} - (\nabla^2_{zz}L(w^*))^{-1}\|' \leq (1 + \mu_1\sigma^k)\|H_k - (\nabla^2_{zz}L(w^*))^{-1}\|' + \mu_2\sigma^k,$$

with $\sigma^k = \max\{\|w^{k+1} - w^*\|, \|w^k - w^*\|\}$. Then for any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|w - w^*\| \leq \varepsilon(r)$ and $\|H_0 - (\nabla^2_{zz}L(w^*))^{-1}\| \leq \delta(r)$ are satisfied, then the sequence $\{w^k\}$ generated by QN-PSQP with respect to a sequence of positive symmetric matrices $\{H_k\}$ is well defined and converges Q-linearly to w^*.

5 Superlinear Convergence

In this section we give a sufficient condition for local superlinear convergence of QN-PSQP for MPCC.

Suppose z^* is a local solution of the MPCC at which Assumption 1 holds and the sequence $\{z^k\}$ generated by QN-PSQP converges to z^*. Let $\lambda^k = (\xi^k, \eta^k, \pi^k)$ be the KKT multiplier associated with the solution d^k of (9), and $\lambda^* = (\xi^*, \eta^*, \pi^*)$ denote the unique MPCC multiplier associated with z^*.

We apply the notation of Section 3, which deals with nonlinear programs, to each of the relevant branch NLPs. Let $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$. Let $\mathcal{C}_k(\mathcal{J}, \mathcal{K})$ denote the approximate critical cone, as defined for a general NLP in (8), for the branch NLP (3) at z^k. Let $P_k(\mathcal{J}, \mathcal{K})$ denote the Euclidean projection on the set $\mathcal{C}_k(\mathcal{J}, \mathcal{K})$.

In order to simultaneously deal with all branches of the local decomposition about a feasible point z^* of (1), we recall the `relaxed NLP` [25]:

$$\begin{array}{lll} \min & f(x,y) & \\ \text{subject to} & 0 \le g(x,y), & \\ & F_i(z) = 0 \le y_i, & \forall i \in \alpha(z^*), \\ & F_i(z) \ge 0 \le y_i, & \forall i \in \beta(z^*) \\ & F_i(z) \ge 0 = y_i, & \forall i \in \gamma(z^*). \end{array}$$

Now define the `relaxed approximate critical cone`, $\mathcal{C}_k^{\mathrm{r}}$, as the approximate critical cone of the relaxed NLP at z^k. That is,

$$\begin{array}{lllll} \mathcal{C}_k^{\mathrm{r}} & = & \{\ d = (dx, dy) & : & g_i'(z^k)d \ge 0, \quad i \in I_g(z^*) \text{ and } \xi_i^* = 0 \\ & & & & g_i'(z^k)d = 0, \quad \xi_i^* > 0 \\ & & & & F_i'(z^k)d \ge 0, \quad i \in \beta(z^*) \text{ and } \eta_i^* = 0 \\ & & & & F_i'(z^k)d = 0, \quad i \in \alpha(z^*) \text{ or } \eta_i^* > 0 \\ & & & & (dy)_i \ge 0, \quad i \in \beta(z^*) \text{ and } \pi_i^* = 0 \\ & & & & (dy)_i = 0, \quad i \in \gamma(z^*) \text{ or } \pi_i^* > 0 \quad \}. \end{array}$$

Let P_k^{r} denote the Euclidean projection onto $\mathcal{C}_k^{\mathrm{r}}$.

Observe that the feasible set of the relaxed NLP contains each branch $\mathcal{F}_{(\mathcal{J},\mathcal{K})}$ of $\mathcal{F}$ at z^*, hence that $\mathcal{C}_k^{\mathrm{r}} \supset \mathcal{C}_k(\mathcal{J}, \mathcal{K})$. Since each of $\mathcal{C}_k^{\mathrm{r}}$ and $\mathcal{C}_k(\mathcal{J}, \mathcal{K})$ is a nonempty closed convex cone, the previous inclusion immediately yields the following technical result, which is purely a fact of convex analysis. A proof is found in the Appendix.

Lemma 5.1 *Let $z^* \in \mathcal{F}$, $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$ and z be any point of $\Re^{n+m}$. Then*

$$\|P_k(\mathcal{J}, \mathcal{K})(z)\| \ \le \ \|P_k^{\mathrm{r}}(z)\|.$$

We are now ready to present a general result on local superlinear convergence of QN-PSQP. It is analogous to the characterisation of Boggs, Tolle and Wang [3] for equality-constrained NLPs, and Bonnans' generalisation [4] for general NLPs which is quoted above as Lemma 3.2. The proof of result also owes a debt to the analysis of Kojima and Shindo [22]. There, quadratic (superlinear respectively) convergence of a Newton (quasi-Newton resp.) method for finding a zero of a piecewise smooth function was shown by examining the action of the method on each "piece" of the function adjacent to a given solution. Here we make a similar argument, though each piece is a branch of an MPCC that we want to minimise. Note that our sufficient condition below is not generally necessary, however, because we replace (that is, bound) each branch projection operator $P_k(\mathcal{J}, \mathcal{K})$ by P_k^{r}.

Proposition 5.1 *Let Assumptions 1 and 2 hold. Suppose that* $\{w^k\}$ *is generated by QN-PSQP with respect to a sequence of matrices* $\{B_k\}$*, and that* $\{w^k\}$ *converges to* w^**. Then a sufficient condition for the sequence* $\{z^k\}$ *to converge Q-superlinearly to* z^* *is*

$$\lim_{k\to\infty} \frac{\|P_k^{\mathrm{r}}[(\nabla^2_{zz}L(w^*) - B_k)(z^{k+1} - z^k)]\|}{\|z^{k+1} - z^k\|} = 0.$$

Proof. Under the condition given in the statement of the Proposition, Lemma 5.1 gives

$$\lim_{k\to\infty} \frac{\|P_k^{\mathrm{r}}(\mathcal{J},\mathcal{K})[(\nabla^2_{zz}L(w^*) - B_k)(z^{k+1} - z^k)]\|}{\|z^{k+1} - z^k\|} = 0$$

for each $(\mathcal{J},\mathcal{K}) \in \mathcal{A}(z^*)$. Since $z^k \to z^*$, by Lemma 2.1 we may assume k is large enough such that any branch chosen in the kth iteration, which is a branch of the decomposition at z^k, is also a branch of the decomposition at z^*. By Proposition 3.1, the superlinear convergence of $\{z^k\}$ follows if $\{z^k\}$ is generated by the same branch of the decomposition. However, the algorithm QN-PSQP may visit different branches of the decomposition at the different iterations. Let $\Delta_{(\mathcal{J},\mathcal{K})}$ denote the "small-order" function defined in Remark 3.1 corresponding to any such branch $\mathcal{F}_{(\mathcal{J},\mathcal{K})}$, and define

$$\Delta_{\max}(\cdot) \;=\; \max\{\Delta_{(\mathcal{J},\mathcal{K})}(\cdot) : \; (\mathcal{J},\mathcal{K}) \in \mathcal{A}(z^*)\}.$$

Clearly, $\Delta_{\max}$ is still a small-order function due to the finite cardinality of the set $\mathcal{A}(z^*)$. It follows from Proposition 3.1 that

$$\|z^{k+1} - z^*\| \le \Delta_{\max}(\|z^k - z^*\|)$$

which gives the desired result. □

6 BFGS, DFP and Convergence

By Proposition 4.1, QN-PSQP converges locally if $B_k \equiv B_0$ and B_0 is sufficiently close to $\nabla^2_{zz}L(w^*)$. However, one may not expect any fast convergence of QN-PSQP if $B_k \equiv B_0$ in general. In order to achieve fast convergence, B_k needs to be updated by fully exploiting first order derivatives of the Lagrangean function L in recent iterates. One of well-known conditions in the context of NLP is the so-called quasi-Newton equation that the matrix B_{k+1} has to satisfy:

$$B_{k+1}(w^{k+1} - w^k) = \nabla_z L(w^{k+1}) - \nabla_z L(w^k).$$

Then different well-known updates can be derived from minimizing the distance in some norm space between the matrices B_k and B_{k+1}. Indeed, the paper [4] derived the local convergence as well as superlinear convergence of some special quasi-Newton methods such as the Powell-Symmetric-Broyden (PSB) method and the Broyden method for NLP. However, we may find that the well-known updates like BFGS and DFP are not included in the framework of [4] since BFGS and DFP do not satisfy the general quasi-Newton update conditions used there (see (2.9), (2.10) and (2.11), or (5.20), (5.21) and (5.22) in [4]). Therefore, the results in [4] are not directly applicable to BFGS and DFP which use symmetric positive definite matrices. This suggests that we need further results to show that local and superlinear convergence of QN-PSQP holds for BFGS and DFP for MPCC.

The main goal of this section is to apply the general convergence and superlinear convergence results to the two well-known concrete updates BFGS and DFP. These special updates are defined as follows respectively.

$$
\begin{aligned}
(\mathbf{BFGS}) \quad B_{k+1} = &\ B_k + \frac{t_k(t_k)^T}{(t_k)^T d^k} - \frac{B_k^T d^k (d^k)^T B_k}{(d^k)^T B_k d^k}. \\
(\mathbf{DFP}) \quad H_{k+1} = &\ H_k + \frac{(t^k - H_k d^k)(t^k)^T + t^k(t^k - H_k d^k)^T}{(t^k)^T d^k} - \frac{(d^k)^T(t^k - H_k d^k)t^k(t^k)^T}{((t^k)^T d^k)^2} \\
= &\ \left(I - \frac{t^k(d^k)^T}{(t^k)^T d^k}\right) H_k \left(I - \frac{d^k(t^k)^T}{(t^k)^T d^k}\right) + \frac{t^k(t^k)^T}{(t^k)^T d^k}.
\end{aligned}
$$

where

$$
\begin{aligned}
d^k &= z^{k+1} - z^k, \\
t^k &= \nabla_z L(z^{k+1}, \lambda^{k+1}) - \nabla_z L(z^k, \lambda^{k+1}).
\end{aligned}
$$

Our convergence analysis of BFGS and DFP for MPCC parallels that of Dennis and Moré [8] for unconstrained minimization. For the sake of completeness, we present some results from [8]. It will be seen that some of fundamental lemmas developed for unconstrained and constrained nonlinear programs are needed in establishing convergence of the BFGS and the DFP methods. The first of the following lemmas is a generalization of Lemma 3.2 in [16], which can be proved from the Mean-Value Theorem.

Lemma 6.1 *There exist a neighborhood $N(z^*, \varepsilon^*)$ of z^* and positive constants μ_1 and μ_2 such that for any $z, z^+ \in N(z^*, \varepsilon^*)$ and any $\lambda \in \Re^{l+m+m}$, we have*

$$
\begin{aligned}
&\|\nabla_z L(z^+, \lambda) - \nabla_z L(z, \lambda) - \nabla^2_{zz} L(z^*, \lambda^*)(z^+ - z)\| \\
\le\ &(\mu_1 \max\{\|z^+ - z^*\|, \|z - z^*\| + \mu_2 \|\lambda - \lambda^*\|)\ \|z^+ - z\|.
\end{aligned}
$$

Lemma 6.2 ([8]) *Let u, v belong to $\Re^n$ with $u \ne 0$, $v \ne 0$, let $\mu \in (0,1)$. If $\|u - v\| \le \mu\|u\|$, then $u^T v$ is positive and*

$$
1 - (\frac{u^T v}{\|u\|\|v\|})^2 \le \mu^2.
$$

Let $\|\cdot\|_F$ denote the Frobenius matrix norm on square matrices, that is $\|A\|_F = \sqrt{\sum_i \sum_j (A_{ij})^2}$. Define two matrix norms for the matrix Q associated with the symmetric square matrix A:

$$
\begin{aligned}
\|Q\|_{\text{DFG}} &= \|A^{-1/2} Q A^{-1/2}\|_{\text{F}}, \\
\|Q\|_{\text{BFGS}} &= \|A^{1/2} Q A^{1/2}\|_{\text{F}}.
\end{aligned}
$$

Lemma 6.3 ([8]) *Let*

$$
\theta = \frac{t^T d}{\|A^{-1/2} t\| \|A^{1/2} d\|} = \frac{(A^{-1/2} t)^T A^{1/2} d}{\|A^{-1/2} t\| \|A^{1/2} d\|}.
$$

Then

$$
\|B_+ - A\|_{\text{DFP}} \le \frac{1}{\theta^2} \|B - A\|_{\text{DFP}} + \frac{2}{\theta^2} \frac{\|A^{-1/2} t - A^{1/2} d\|}{\|A^{1/2} d\|} \tag{11}
$$

if B_+ is updated by the DFP method, and

$$
\|H_+ - A^{-1}\|_{\text{BFGS}} \le \frac{1}{\theta^2} \|H - A^{-1}\|_{\text{BFGS}} + \frac{2}{\theta^2} \frac{\|A^{1/2} d - A^{-1/2} t\|}{\|A^{-1/2} t\|} \tag{12}
$$

if H_+ is updated by the BFGS method.

We are now ready to present the convergence result of the BFGS and the DFP methods for MPCC. We strengthen Assumption 2, the piecewise SOSC at z^*, by positive definiteness of the Hessian of the Lagrangean.

Assumption 3 Given $w^* = (z^*, \lambda^*)$, $\nabla^2_{zz} L(w^*)$ is positive definite.

Proposition 6.1 *Let Assumptions 1 and 3 hold.*

(i) There exist constants $\varepsilon > 0$ and $\delta > 0$ such that if $\|w - w^*\| \le \varepsilon$ and B_0 is a symmetric matrix satisfying $\|B_0 - \nabla^2_{zz} L(w^*)\| \le \delta$, then the sequence $\{w^k\}$ generated by QN-PSQP with the matrix B updated by the DFP method converges Q-linearly to w^*.

(ii) There exist constants $\varepsilon > 0$ and $\delta > 0$ such that if $\|w - w^*\| \le \varepsilon$ and $\|H_0 - (\nabla^2_{zz} L(w^*))^{-1}\| \le \delta$, then the sequence $\{w^k\}$ generated by QN-PSQP with the matrix H updated by the BFGS method converges Q-linearly to w^*.

Proof. Consider the DFP method. For the ease of presentation, we drop the iteration index k and replace the index $k+1$ by $+$. Let $A = \nabla^2_{zz} L(w^*)$. It follows that

$$\begin{aligned} \|A^{-1/2}t - A^{1/2}d\|\|d\| &= \|A^{-1/2}(t - Ad)\|\|d\| \\ &= \|A^{-1/2}(t - Ad)\|\|A^{-1/2}A^{1/2}d\| \\ &\le \|A^{-1/2}\|^2\|(t - Ad)\|\|A^{1/2}d\|, \end{aligned}$$

i.e.,

$$\|A^{-1/2}t - A^{1/2}d\| \le \|A^{-1}\|\frac{\|t - Ad\|}{\|d\|}\|A^{1/2}d\|. \tag{13}$$

By Lemma 6.1, there exist a neighborhood $N(z^*, \varepsilon^*)$ of z^* and positive constants μ_1 and μ_2 such that if $z, z^+ \in N(z^*, \varepsilon^*)$, then

$$\begin{aligned} \|t - Ad\| &\le [\mu_1 \max\{\|z^+ - z^*\|, \|z - z^*\|\} + \mu_2\|\lambda^+ - \lambda^*\|]\|z^+ - z\| \\ &\le \mu_3 \max\{\|w^+ - w^*\|, \|w - w^*\|\}\|d\|, \end{aligned}$$

where $\mu_3 > 0$ is a suitable positive number. Proposition 4.1 shows that $z^+ \in N(z^*, \varepsilon^*)$ and $\|w^+ - w^*\| < \|w - w^*\|$ provided that $z \in N(z^*, \varepsilon^*)$ and $\varepsilon^* \in (0,1)$ is sufficiently small. This implies that there exists a sufficiently small positive constant κ independent of the iteration index such that

$$\|t - Ad\| \le \mu_3\|w - w^*\|\|d\| \le \kappa\|d\|. \tag{14}$$

From (13), the condition of Lemma 6.2 holds with

$$\mu = \|A^{-1}\|\frac{\|t - Ad\|}{\|d\|} \le \mu_3\|w - w^*\|\|A^{-1}\| < 1,$$

$u = A^{1/2}d$ and $v = A^{-1/2}t$. Lemmas 6.2, 6.3 and (13) imply that

$$1 - \theta^2 \le \left[\|A^{-1}\|\frac{\|t - Ad\|}{\|d\|}\right]^2 \le [\kappa\|A^{-1}\|]^2.$$

It can easily be verified that

$$\theta^2 \ge \frac{1}{2},$$

$$\frac{1}{\theta^2} = 1 + \frac{1-\theta^2}{\theta^2} \le 1 + 2(\mu_3\|A^{-1}\|\|w - w^*\|)^2 \le 1 + \mu_4\|A^{-1}\|\|w - w^*\|,$$

with μ_4 a suitable positive number. Therefore, Lemma 6.3, (13) and (14) show that

$$\|B_+ - A\|_{\mathrm{DFP}} \le (1 + \mu_4\|A^{-1}\|\|w - w^*\|)\|B - A\|_{\mathrm{DFP}} + 4\mu_3\|A^{-1}\|\|w - w^*\|, \tag{15}$$

which obviously implies the bounded deterioration condition in Proposition 4.2. By an analogous argument, the bounded deterioration condition is also satisfied for the BFGS method. Then convergence follows from Proposition 4.2. □

Proposition 6.2 *Let Assumptions 1 and 3 hold.*

(i) There exist constants $\varepsilon > 0$ and $\delta > 0$ such that if $\|w - w^*\| \le \varepsilon$ and B_0 is a symmetric matrix satisfying $\|B_0 - \nabla^2_{zz}L(w^*)\| \le \delta$. Let the sequence $\{w^k = (z^k, \lambda^k)\}$ be generated by QN-PSQP with the matrix B updated by the DFP method. Then $\{z^k\}$ converges Q-superlinearly to z^*.

(ii) There exist constants $\varepsilon > 0$ and $\delta > 0$ such that if $\|w - w^*\| \le \varepsilon$ and $\|H_0 - (\nabla^2_{zz}L(w^*))^{-1}\| \le \delta$. Let the sequence $\{w^k = (z^k, \lambda^k)\}$ be generated by QN-PSQP with the matrix H updated by the BFGS method. Then $\{z^k\}$ converges Q-superlinearly to z^*.

Proof. First note that $\{z^k\}$ converges Q-linearly to z^* by the last proposition. Similar to (4.10) of Broyden, Dennis and Moré [5], the following more stringent inequality than (15) holds for some positive constants $\mu \in (0,1)$, μ_1 and μ_2 if DFP is used and B_0 is a symmetric matrix

$$b_{k+1} \le [(1 - \mu\theta_k^2)^{1/2} + \mu_1\sigma^k]b_k + \mu_2\sigma^k, \tag{16}$$

where

$$\begin{aligned} b_k &= \|B_{k+1} - \nabla^2_{zz}L(w^*)\|_{\mathrm{DFP}}, \\ \theta_k &= \begin{cases} \dfrac{\|M(\nabla^2_{zz}L(w^*) - B_k)d^k\|}{b_k\|M^{-1}d^k\|} & \text{if } B_k \ne \nabla^2_{zz}L(w^*), \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and $M^2 = (\nabla^2_{zz}L(w^*))^{-1}$. It can be verified that $\lim_{k\to\infty} b_k$ exists. Since

$$(1 - \mu\theta_k^2)^{1/2} \le 1 - \frac{\mu}{2}\theta_k^2,$$

the above inequality (16) can be rewritten as follows

$$\frac{\mu\theta_k^2}{2}b_k \le b_k - b_{k+1} + [\mu_1 b_k + \mu_2]\sigma^k.$$

Summing both sides of the above inequality for all k, we have

$$\sum_{k=1}^{\infty} \theta_k^2 b_k < \infty. \tag{17}$$

If $b_k \to 0$, Proposition 5.1 implies that $\{z^k\}$ converges Q-superlinearly to z^*. Otherwise, $\{b_k\}$ is bounded away from zero. It follows from (17) that

$$\lim_{k\to\infty} \frac{\|M(B_k - \nabla^2_{zz}L(w^*))d^k\|}{\|M^{-1}d^k\|} = 0,$$

which together with Proposition 5.1 implies the Q-superlinear convergence of $\{z^k\}$ to z^*. The Q-superlinear convergence of the BFGS can be proved analogously. □

7 Linear Complementarity Constraints

One of important special cases of MPCC is the so-called LCP-MP in which equilibrium constraints are parametric linear complementarity problems, and the upper level constraints are characterized by linear equalities and inequalities. Namely, in the formulation (1), both g and F are linear affine. When $g = Gx + Hy + b$ and $F = Nx + My + q$ with $G \in \Re^{n\times l}$, $H \in \Re^{m\times l}$, $b \in \Re^{l}$, $N \in \Re^{n\times m}$, $M \in \Re^{m\times m}$ and $q \in \Re^{m}$, the LCP-MP in the form of (1) reads

$$\begin{array}{cc} \min & f(x,y) \\ \text{subject to} & Gx + Hy + b \geq 0 \\ & 0 \leq (Nx + My + q) \ \perp \ y \geq 0. \end{array} \tag{18}$$

The LCP-MP not only arises from practical applications but also is used to convert nonlinearly constrained MPECs into LCP-MPs through partial penalty and reformulation techniques. Therefore, it is important to study how quasi-Newton methods with PSQP work for the LCP-MP. In fact, the first local superlinear convergence for PSQP was established for LCP-MP in [33]. Parallel to the case of SQP applied to linearly constrained NLP [4], fewer assumptions are needed for the superlinear convergence results of [33] than for the case of nonlinearly constrained MPCC. It is no surprise that fewer assumptions are also needed to show superlinear convergence of quasi-Newton PSQP methods for LCP-MP.

The results established in the previous sections for general MPCCs remain valid for the LCP-MP after dropping the MPCC-SMFCQ assumption.

Proposition 7.1 *Suppose that the piecewise SOSC holds at a local solution $z^* = (x^*, y^*)$ of the LCP-MP (18). The the following conclusions hold.*

(i) For any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|z - z^*\| \leq \varepsilon(r)$ and $\|B - \nabla^2 f(x^*, y^*)\| \leq \delta(r)$, then

$$\|z^+ - z^*\| \leq r\|z - z^*\|,$$

where $z^+ = z + d$, d is a solution of the QP (9) corresponding to a branch of the decomposition of the LCP-MP at z.

(ii) Assume there exist two positive constants μ_1 and μ_2 such that

$$\|(B_{k+1} - \nabla^2 f(x^*, y^*)\|' \leq (1 + \mu_1\sigma^k)\|B_k - \nabla^2 f(x^*, y^*)\|' + \mu_2\sigma^k,$$

or

$$\|H_{k+1} - (\nabla^2 f(x^*, y^*))^{-1}\|' \leq (1 + \mu_1\sigma^k)\|H_k - \nabla^2 f(x^*, y^*)\|' + \mu_2\sigma^k,$$

with $\sigma^k = \max\{\|z^{k+1} - z^*\|, \|z^k - z^*\|\}$. Then for any $r \in (0,1)$, there exist two positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|z - z^*\| \leq \varepsilon(r)$ and $\|B_0 - \nabla^2 f(x^*, y^*)\| \leq \delta(r)$ or $\|H_0 - (\nabla^2 f(x^*, y^*))^{-1}\| \leq \delta(r)$ are satisfied, then the sequence $\{z^k\}$ generated by QN-PSQP with respect to a sequence of matrices $\{B_k\}$ or $\{H_k\}$ is well defined and converges to Q-linearly to z^*.

(iii) If z^k converges to z^*, then a sufficient condition for the sequence $\{z^k\}$ to converge Q-superlinearly to z^* is

$$\lim_{k\to\infty} \frac{\|P_k^{\mathrm{r}}[\nabla^2 f(x^*, y^*) - B_k)(z^{k+1} - z^k)]\|}{\|z^{k+1} - z^k\|} = 0,$$

where P_k^{r} is the Euclidean projection on the relaxed approximate critical cone $\mathcal{C}_k^{\mathrm{r}}$.

(iv) If $\nabla^2 f(x^*, y^*)$ is positive definite, then the sequence $\{z^k\}$ generated by QN-PSQP with either DFP or BFGS updates converges Q-superlinearly to z^*.

Proof. (i) Similar to Proposition 4.1, the desired result follows from Lemma 3.3(i).

(ii) It is similar to Propositions 4.2 and 4.3.

(iii) It is a consequence of Proposition 3.2(ii).

(iv) Note that the results in Proposition 6.1 are still valid from (ii) of the current proposition. Proposition 6.2 is also true by the last equation of the proof of the same proposition and (iii) of the current proposition. □

References

[1] G. Anandalingam and T.L. Friesz, eds., Hierarchical Optimization, *Annals of Operations Research*, 1992.

[2] J. Bard, An algorithm for solving the general bilevel programming problem, *Mathematics of Operations Research* 8 (1983) 260–272.

[3] P.T. Boggs, J.W. Tolle and P. Wang, On the local convergence of quasi-Newton methods for constrained optimization, *SIAM Journal on Control and Optimization* 20 (1982) 161–171.

[4] J.F. Bonnans, Local analysis of Newton-type methods for variational inequalities and nonlinear programming, *Applied Mathematics and Optimization* 29 (1994) 161–186.

[5] G.G. Broyden, J.E. Dennis and J.J. Moré, On the local and superlinear convergence of quasi-Newton methods, *Journal of Institute Mathematics and Applications* 12 (1973) 223–246.

[6] Y. Chen and M. Florian, The nonlinear bilevel programming problem: Formulation, regularity and optimality conditions, *Optimization* 32 (1995) 193–209.

[7] J.E. Dennis and J.J. Moré, A characterization of superlinear convergence and its applications to quasi-Newton methods. *Mathematics Computation* 28 (1974) 549–560.

[8] J.E. Dennis and J.J. Moré, Quasi-Newton methods: Motivation and Theory. *SIAM Review* 19 (1977) 46–89.

[9] F. Facchinei, H. Jiang and L. Qi, A smoothing method for mathematical programs with equilibrium constraints, *Mathematical Programming* 85 (1999) 81–106.

[10] M.C. Ferris and J.S. Pang, Engineering and economics applications of complementarity problems, *SIAM Review* 39 (1997) 669-713.

[11] R. Fletcher, *Practical Methods of Optimization* John Wiley, 2nd Edition, 1987.

[12] R. Fletcher, S. Leyffer, D. Ralph and S. Scholtes, Local convergence of SQP methods for Mathematical Programs with Equilibrium Constraints, Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, 2002.

[13] M. Fukushima, Z.-Q. Luo and J.S. Pang, A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints, *Computational Optimization and Applications* 10 (1998) 5–34.

[14] M. Fukushima and J.S. Pang, Convergence of a smoothing continuation method for mathematical programs with complementarity constraints, in *Ill-posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems*, Vol. 447, M. Théra and R. Tichatschke, eds., Springer-Verlag, Berlin/Heidelberg, 1999, 99–110.

[15] M. Fukushima and P. Tseng, An implementable active-set methods for computing a B-stationary point of a mathematical program with linear complementarity constraints, *SIAM J Optimization*, to appear.

[16] S.P. Han, Superlinearly convergent variable metric algorithms for general nonlinear programming problems, *Mathematical Programming* 11 (1976) 263–282.

[17] H. Jiang and D. Ralph, QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints, *Computational Optimization and Applications* 13 (1999) 25–49.

[18] H. Jiang and D. Ralph, Smooth SQP methods for mathematical programs with nonlinear complementarity constraints, *SIAM Journal on Optimization* 10 (2000) 779–808.

[19] N.H. Josephy, Newton's method for generalized equations, Technical Summary Report #1965, Mathematics Research Center, University of Wisconsin, Madison, 1979.

[20] N.H. Josephy, Quasi-Newton methods for generalized equations, Technical Summary Report #1966, Mathematics Research Center, University of Wisconsin, Madison, 1979.

[21] C. Kirjner and E. Polak, On the conversion of optimization problems with maxmin constraints to standard optimization problems, *SIAM J. Optimization*, Vol. 8, No. 4 (1998), 887–915.

[22] M. Kojima and S. Shindo, Extensions of Newton and quasi-Newton methods to systems of PC^1 Equations, *Journal of Operations Research Society of Japan* 29 (1986) 352–374.

[23] J. Kyparisis, On uniqueness of Kuhn - Tucker multipliers in nonlinear programming, *Mathematical Programming* 32 (1985) 242–246.

[24] S. Leyffer, MPEC — much ado about nothing?, presentation at 17th International Symposium on Mathematical Programming, ISMP 2000, Atlanta, August 7–11, 2000.

[25] Z.-Q. Luo, J.S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, 1996.

[26] Z.-Q. Luo, J.S. Pang and D. Ralph, Piecewise sequential quadratic programming for mathematical Programs with nonlinear complementarity constraints, in: A. Migdalas *et al*, eds., *Multilevel Optimization: Algorithms, Complexity and Applications*, Kluwer Academic Publishers, 1998.

[27] Z.-Q. Luo, J.S. Pang, D. Ralph and S.-Q. Wu, Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints, *Mathematical Programming* 75 (1996) 19–76.

[28] J. Outrata, M. Kočvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Boston, 1998.

[29] J.V. Outrata and J. Zowe, A numerical approach to optimization problems with variational inequality constraints, *Mathematical Programming* 68 (1995) 105–130.

[30] J.S. Pang, Convergence of splitting and Newton methods for complementarity problems: an application of some sensitivity results, *Mathematical Programming* 58 (1993) 149–160.

[31] J.S. Pang and L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM Journal on Optimization* 3 (1993) 443–465.

[32] L. Qi and H. Jiang, Semismooth Karush–Kuhn–Tucker equations and convergence analysis of Newton methods and Quasi–Newton methods for solving these equations, *Mathematics of Operations Research* 22 (1997) 301–325.

[33] D. Ralph, Sequential quadratic programming for mathematical programs with linear complementarity constraints, in: R.L. May and A.K. Easton eds., *CTAC95 Computational Techniques and Applications,* World Scientific, 1996.

[34] S.M. Robinson, Strongly regular generalized equations, *Mathematics of Operations Research* 5 (1980) 43–62.

[35] H. Scheel and S. Scholtes, Mathematical programs with equilibrium constraints: Stationarity, optimality, and sensitivity, *Mathematics of Operations Research* 25 (2000).

[36] S. Scholtes, Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, *SIAM J Optimization* 11 (2001), 918–936.

[37] S. Scholtes, Active set methods for inverse linear complementarity problems, manuscript, Judge Institute of Management, Cambridge University, Trumpington St, Cambridge, CB2 1AG, UK, 1999.

[38] S. Scholtes, Combinatorial structures in nonlinear programming, manuscript, Judge Institute of Management, Cambridge University, Trumpington St, Cambridge, CB2 1AG, UK, 2002.

[39] S. Scholtes and M. Stöhr, Exact penalization of mathematical programs with equilibrium constraints, *SIAM Journal on Control and Optimization* 37 (1999) 617–652.

[40] L.N. Vicente and P. Calamai, Bilevel and Multilevel Programming: A bibliography review, *Journal of Global Optimization* 5 (1994) 291–306.

[41] J.J. Ye, Optimality conditions for optimization problems with complementarity constraints, *SIAM Journal on Optimization* 9 (1999) 374–387.

[42] E.H. Zarantonello, Projections on convex sets in Hilbert space and spectral theory, in E.H. Zarantonello, ed., *Contributions to Nonlinear Functional Analysis*, Academic Press, New York, 1971, 237-424.

8 Appendix: Proof of Propositions 3.1 and 3.2, and Lemma 5.1

Proof of Proposition 3.1:

Let $s^k = P_k[(\nabla^2_{zz}L(w^*) - B_k)d^k]$. By the definition of $\mathcal{C}_k$, for any k, there exist $\mu_g^k \in \Re^{m_1}$ and $\mu_h^k \in \Re^{m_2}$ such that

$$s^k - g'(z^k)^T\mu_g^k - h'(z^k)^T\mu_h^k = (\nabla^2_{zz}L(w^*) - B_k)d^k. \tag{19}$$

It is clear that $(\mu_g^k)_i = 0, \forall i \notin I$ and $(\mu_g^k)_i \geq 0, \forall i \in I^0$. By the KKT conditions of the quadratic program (7), we obtain

$$s^k - \nabla f(z^k) - \nabla^2_{zz}L(w^*)d^k + g'(z^k)^T(-\mu_g^k + \lambda_g^{k+1}) + h'(z^k)^T(-\mu_h^k + \lambda_h^{k+1}) = 0. \tag{20}$$

From the Taylor's expansion and the KKT conditions at z^*, for $(\lambda_g, \lambda_h) \in \Lambda$, where Λ is the set of Lagrangean multiplers of (5) at z^*, we have

$$\begin{aligned}\nabla f(z^k) &= \nabla f(z^*) + \nabla^2 f(z^*)(z^k - z^*) + o(\|z^k - z^*\|)\\ &= g'(z^*)^T\lambda_g + h'(z^*)^T\lambda_h + \nabla^2 f(z^*)(z^k - z^*) + o(\|z^k - z^*\|)\\ &= g'(z^k)^T\lambda_g + h'(z^k)^T\lambda_h + \nabla^2_{zz}L(w^*)(z^k - z^*) + o(\|z^k - z^*\|).\end{aligned}$$

Therefore, (20) implies that

$$s^k - \nabla^2_{zz}L(w^*)(z^k + d^k - z^*) + g'(z^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + h'(z^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1}) = o(\|z^k - z^*\|). \tag{21}$$

let $\delta^k = \|s^k\| + \|z^k - z^*\| + \|d^k\|$. Then the sequences $\{s^k/\delta^k\}$, $\{(z^k + d^k - z^*)/\delta^k\}$ and

$$\left\{\frac{g'(z^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + h'(z^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{\delta^k}\right\}$$

are all bounded. Dividing (21) by δ^k and taking the limit if necessary passing to a subsequence K, we obtain

$$s - \nabla^2_{zz}L(w^*)t + r = 0, \tag{22}$$

for some vectors s, t and r, where

$$r = \lim_{k\in K, k\to\infty} \frac{g'(z^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + h'(z^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{\delta^k},$$

and r is independent of $(\lambda_g, \lambda_h) \in \Lambda$.

From the KKT conditions of (5) at z^* and of (7) at z^k, and the Taylor expansion, we may prove that $t \in \mathcal{C}(z^*)$. Since $s^k \in \mathcal{C}_k$ for each k, $s \in \mathcal{C}(z^*)$.

Suppose $\{z^k\}$ superlinearly converges to z^*. Then $t = 0$. If $r = 0$, then $s = 0$ which implies that

$$s^k = o(\|z^k - z^*\|).$$

We next aim to prove that $r = 0$. In fact, this follows from the following arguments,

$$\begin{aligned}
-r^T r &= s^T r \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + (h'(z^k)s)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{\delta^k} \\
&\quad \text{(Definition of } r\text{)} \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + (h'(z^k)s^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{(\delta^k)^2} \\
&\quad \text{(Definition of } s\text{)} \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s^k)^T(-\lambda_g + \lambda_g^{k+1}) + (h'(z^k)s^k)^T(-\lambda_h + \lambda_h^{k+1})}{(\delta^k)^2} \\
&\quad \text{(Orthogonality of } s^k \text{ and } -g'(z^k)^T\mu_g^k - h'(z^k)^T\mu_h^k; \\
&\quad \text{See the comment in the proof of Lemma 5.1)} \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I_0} \frac{(g_i'(z^k)s^k)^T(-(\lambda_g)_i + (\lambda_g^{k+1})_i)}{(\delta^k)^2} \\
&\quad (s^k \in \mathcal{C}_k) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0} \frac{1}{(\delta^k)^2}(g_i'(z^k)s^k)(\lambda_g^{k+1})_i \\
&\quad ((\lambda_g)_i = 0, i \in I^0) \\
&\geq 0. \\
&\quad (g_i'(z^k)s^k \geq 0, (\lambda_g^{k+1})_i \geq 0, i \in I^0)
\end{aligned}$$

Conversely, suppose $s^k = o(\|d^k\|)$. Then $s = 0$, and (22) becomes

$$-\nabla^2_{zz} L(w^*)t + r = 0.$$

It suffices to demonstrate that $t = 0$ for the superlinear convergence of $\{z^k\}$. Note that $t \in \mathcal{C}(z^*)$ from above and the SOSC holds at z^* by the hypotheses. The above equation implies that it suffices to prove that $t^T r \leq 0$. We claim the following results hold:

(a) $g'(z^k)(z^{k+1} + d^k - z^*) = g'(z^k)d^k + g(z^k) - g(z^*) + O(\|z^k - z^*\|^2)$,
$h'(z^k)(z^{k+1} + d^k - z^*) = h'(z^k)d^k + h(z^k) - h(z^*) + O(\|z^k - z^*\|^2)$;

(b) $(\lambda_g^{k+1}, \lambda_h^{k+1}) \to (\lambda_g, \lambda_h)$;

(c) (μ_g^k, μ_h^k) converges to zero.

Apparently, (a) follows from the Taylor expansion and twice differentiability of g and h, and (b) from the SMFCQ. By the SMFCQ at z^*, the sequence $\{(\mu_g^k, \mu_h^k)\}$ must be bounded and its unique limit must be zero (otherwise, by suitable algebraic manipulations and taking limit, a contradiction would follow if either $\{(\mu_g^k, \mu_h^k)\}$ is not bounded or it has a nonzero accumulation point). Then the following arguments hold,

$$
\begin{aligned}
t^T r &= \lim_{k\in K, k\to\infty} \left[\frac{(g'(z^k)(z^k+d^k-z^*))^T(-\lambda_g-\mu_g^k+\lambda_g^{k+1})}{(\delta^k)^2} \right. \\
&\qquad \left. + \frac{(h'(z^k)(z^k+d^k-z^*))^T(-\lambda_h-\mu_h^k+\lambda_h^{k+1})}{(\delta^k)^2} \right] \\
&= \lim_{k\in K, k\to\infty} \left[\sum_{i\in I} \frac{(g_i'(z^k)(z^k+d^k-z^*))(-(\lambda_g)_i-(\mu_g^k)_i+(\lambda_g^{k+1})_i)}{(\delta^k)^2} \right. \\
&\qquad \left. + \frac{(h'(z^k)(z^k+d^k-z^*))(-\lambda_h-\mu_h^k+\lambda_h^{k+1})}{(\delta^k)^2} \right] \\
&\qquad ((\lambda_g)_i=(\mu_g^k)_i=(\lambda_g^{k+1})_i=0, i\notin I) \\
&= \lim_{k\in K, k\to\infty} \left[\sum_{i\in I^0} \frac{(g_i'(z^k)(z^k+d^k-z^*))(-(\lambda_g)_i-(\mu_g^k)_i+(\lambda_g^{k+1})_i)}{(\delta^k)^2} \right. \\
&\qquad + \sum_{i\in I^+} \frac{O(\|z^k-z^*\|^2)(-(\lambda_g)_i-(\mu_g^k)_i+(\lambda_g^{k+1})_i)}{(\delta^k)^2} \\
&\qquad \left. + \frac{O(\|z^k-z^*\|^2)\ \|-\lambda_h-\mu_h^k+\lambda_h^{k+1}\|}{(\delta^k)^2} \right] \\
&\qquad (\text{Claim } (a);\ g_i'(z^k)d^k+g_i(z^k)=g_i(z^*)=0, i\in I^+; \\
&\qquad\quad h'(z^k)d^k+h(z^k)=h(z^*)=0) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0} \frac{(g_i'(z^k)(z^k+d^k-z^*))(-(\lambda_g)_i-(\mu_g^k)_i+(\lambda_g^{k+1})_i)}{(\delta^k)^2} \\
&\qquad (\text{Claims } (b) \text{ and } (c);\ \|z^k-z^*\|\le\delta^k) \\
&= \lim_{k\in K, k\to\infty} \left[\sum_{i\in I^0,(\lambda_g^{k+1})_i=0} \frac{(g_i'(z^k)(z^k+d^k-z^*))(-(\lambda_g)_i-(\mu_g^k)_i)}{(\delta^k)^2} \right. \\
&\qquad \left. + \sum_{i\in I^0,(\lambda_g^{k+1})_i>0} \frac{(O\|z^k-z^*\|^2)(-(\lambda_g)_i-(\mu_g^k)_i+(\lambda_g^{k+1})_i)}{(\delta^k)^2} \right] \\
&\qquad (\text{Claim } (a)) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0,(\lambda_g^{k+1})_i=0} \frac{(g_i'(z^k)(z^k+d^k-z^*))(-(\mu_g^k)_i)}{(\delta^k)^2} \\
&\qquad (\text{Claims } (b) \text{ and } (c);\ \|z^k-z^*\|\le\delta^k;\ (\lambda_g)_i=0, i\in I^0) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0,(\lambda_g^{k+1})_i=0} \frac{(g_i'(z^k)d^k+g_i(z^k)+O(\|z^k-z^*\|^2))(-(\mu_g^k)_i)}{(\delta^k)^2} \\
&\qquad (\text{Claim } (a);\ g_i(z^*)=0, i\in I^0)
\end{aligned}
$$

$$
\begin{aligned}
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0, (\lambda_g^{k+1})_i=0} \frac{(g_i'(z^k)d^k + g_i(z^k))(-(\mu_g^k)_i)}{(\delta^k)^2} + \frac{O(\|z^k - z^*\|^2))(-(\mu_g^k)_i)}{(\delta^k)^2} \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0, (\lambda_g^{k+1})_i=0} \frac{(g_i'(z^k)d^k + g_i(z^k))(-(\mu_g^k)_i)}{(\delta^k)^2} \\
&\qquad \text{(Claim } (c);\ \|z^k - z^*\| \le \delta^k) \\
&\le 0. \\
&\qquad (g_i'(z^k)d^k + g_i(z^k) \ge 0, (\mu_g^k)_i \ge 0, i \in I^0)
\end{aligned}
$$

Therefore $tr \le 0$, which completes the proof. □

Proof of Proposition 3.2:

For convenience of exposition, we still use $g'(z)$ and $h'(z)$ to denote the Jacobians of g and h at z respectively though it is obvious that $g'(z)$ and $h'(z)$ are constant matrices independent of z.

Following the same argument in the proof of Proposition 3.1 and taking into account of the fact that g and h are linear, we are able to obtain equations below:

$$s^k - g'(z^k)^T \mu_g^k - h'(z^k)^T \mu_h^k = (\nabla^2 f(z^*) - B_k)d^k, \tag{23}$$

$$s - \nabla^2 f(z^*)t + r = 0, \tag{24}$$

for some vectors s, t and r.

From the KKT conditions of (5) at z^* and of (7) at z^k, we may prove that $t \in \mathcal{C}(z^*)$. Since $s^k \in \mathcal{C}_k$ for each k, $s \in \mathcal{C}(z^*)$. See Section 3 for definitions of $\mathcal{C}(z^*)$ and $\mathcal{C}_k$.

Suppose $\{z^k\}$ superlinearly converges to z^*. Then $t = 0$. If $r = 0$, then $s = 0$ which implies that

$$s^k = o(\|z^k - z^*\|).$$

We now prove that $r = 0$. In fact, this follows from the following arguments,

$$\begin{aligned}
-r^T r &= s^T r \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + (h'(z^k)s^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{\delta^k} \\
&\quad \text{(Definition of } r) \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s^k)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + (h'(z^k)s^k)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{(\delta^k)^2} \\
&\quad \text{(Definition of } s) \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)s^k)^T(-\lambda_g + \lambda_g^{k+1}) + (h'(z^k)s^k)^T(-\lambda_h + \lambda_h^{k+1})}{(\delta^k)^2} \\
&\quad \text{(Orthogonality of } s^k \text{ and } -g'(z^k)^T\mu_g^k - h'(z^k)^T\mu_h^k; \\
&\quad \text{See the comment in the proof of Lemma 5.1)} \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^*)s)^T(-\lambda_g + \lambda_g^{k+1}) + (h'(z^*)s)^T(-\lambda_h + \lambda_h^{k+1})}{\delta^k} \\
&\quad \text{(Definition of } s^k; g, h \text{ are linear)} \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0} \frac{1}{\delta^k}(g_i'(z^*)s)(-(\lambda_g)_i + (\lambda_g^{k+1})_i) \\
&\quad (s \in \mathcal{C}(z^*)) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I^0} \frac{1}{\delta^k}(g_i'(z^*)s)(\lambda_g^{k+1})_i \\
&\quad ((\lambda_g)_i = 0, i \in I^0) \\
&\geq 0. \\
&\quad (s \in \mathcal{C}(z^*);\ (\lambda_g^{k+1})_i \geq 0)
\end{aligned}$$

Conversely, suppose $s^k = o(\|d^k\|)$. Then $s = 0$ and (24) becomes

$$-\nabla^2 f(z^*)t + r = 0.$$

It suffices to demonstrate that $t = 0$ for the superlinear convergence of $\{z^k\}$. Note that $t \in \mathcal{C}(z^*)$ from above and the SOSC holds at z^* by the hypotheses. The above equation implies that it suffices to prove that $t^T r \le 0$. Similar to the proof of necessity, the following arguments hold,

$$\begin{aligned}
t^T r &= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)t)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1}) + (h'(z^k)t)^T(-\lambda_h - \mu_h^k + \lambda_h^{k+1})}{\delta^k} \\
&\quad \text{(Definition of } r\text{)} \\
&= \lim_{k\in K, k\to\infty} \frac{(g'(z^k)t)^T(-\lambda_g - \mu_g^k + \lambda_g^{k+1})}{\delta^k} \\
&\quad (t \in \mathcal{C}(z^*); g'(z^k) = g'(z^*)) \\
&\le \lim_{k\in K, k\to\infty} \frac{(g'(z^k)t)^T \lambda_g^{k+1}}{\delta^k} \\
&\quad (g'(z^k) = g'(z^*);\ t \in \mathcal{C}(z^*);\ \lambda_g \ge 0;\ (\mu_g^k)_i \ge 0, i \in I^0) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I} \frac{(g_i'(z^k)t)(\lambda_g^{k+1})_i}{\delta^k} \\
&\quad ((\lambda_g^{k+1})_i = 0, i \notin I) \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I} \frac{(g_i'(z^k)(z^k + d^k - z^*))(\lambda_g^{k+1})_i}{(\delta^k)^2} \\
&\quad \text{(Definition of } t\text{)} \\
&= \lim_{k\in K, k\to\infty} \sum_{i\in I} \frac{(g_i(z^k) + g_i'(z^k)d^k)(\lambda_g^{k+1})_i}{(\delta^k)^2} \\
&\quad (g_i'(z^k)(z^k - z^*) = g_i(z^k) - g(z^*);\ g_i(z^*) = 0, i \in I;\ g_i'(z^k) = g_i'(z^*)) \\
&= 0. \\
&\quad \text{(Complementarity condition of QP (7))}
\end{aligned}$$

□

Proof of Lemma 5.1: It is well known [42] for any $z \in \Re^{n+m}$ and any closed convex cone $\mathcal{K}$ in $\Re^{n+m}$, that if y is the Euclidean projection of z onto $\mathcal{K}$ then $z - y$ and y are orthogonal, hence $\|z\|^2 = \|z-y\|^2 + \|y\|^2$. Therefore

$$\|z - P_k(\mathcal{J},\mathcal{K})(z)\|^2 + \|P_k(\mathcal{J},\mathcal{K})(z)\|^2 = \|z - P_k^{\mathrm{r}}(z)\|^2 + \|P_k^{\mathrm{r}}(z)\|^2. \quad (25)$$

Also the distance from z to $\mathcal{C}_k(\mathcal{J},\mathcal{K})$ is not less than the distance from z to $\mathcal{C}_k^{\mathrm{r}}$, because the former cone is a subset of the latter, so

$$\|z - P_k(\mathcal{J},\mathcal{K})(z)\|^2 \ge \|z - P_k^{\mathrm{r}}(z)\|^2.$$

The result now follows from (25). □