SIAM J. Optim.
Vol. 10, No. 4, pp. 963–981

ON THE CONSTANT POSITIVE LINEAR DEPENDENCE CONDITION AND ITS APPLICATION TO SQP METHODS*

LIQUN QI† AND ZENGXIN WEI‡

Abstract. In this paper, we introduce a constant positive linear dependence condition (CPLD), which is weaker than the Mangasarian–Fromovitz constraint qualification (MFCQ) and the constant rank constraint qualification (CRCQ). We show that a limit point of a sequence of approximating Karush–Kuhn–Tucker (KKT) points is a KKT point if the CPLD holds there. We show that a KKT point satisfying the CPLD and the strong second-order sufficiency conditions (SSOSC) is an isolated KKT point. We then establish convergence of a general sequential quadratical programming (SQP) method under the CPLD and the SSOSC. Finally, we apply these results to analyze the feasible SQP method proposed by Panier and Tits in 1993 for inequality constrained optimization problems. We establish its global convergence under the SSOSC and a condition slightly weaker than the Mangasarian–Fromovitz constraint qualification, and we prove superlinear convergence of a modified version of this algorithm under the SSOSC and a condition slightly weaker than the linear independence constraint qualification.

Key words. constrained optimization, KKT point, constraint qualification, feasible SQP method, global convergence, superlinear convergence

AMS subject classifications. 90C30, 60K05

PII. S1052623497326629

1. Introduction.

Consider the *constrained optimization problem*

$$\min\{f(x) \mid x \in X\}, \tag{1.1}$$

where $X = \{x \in \Re^n \mid g(x) \leq 0, \quad h(x) = 0\}$, $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, and $h : \Re^n \to \Re^p$ are continuously differentiable functions. Assume that $X \neq \emptyset$. Let $I = \{1, \dots, m\}$ and $J = \{1, \dots, p\}$. For a vector $d \in \Re^q$, we let $\mathrm{supp}(d) = \{j \mid d_j \neq 0\}$.

Let $x \in X$ be a given *feasible point* of (1.1). Let

$$I(x) = \{j \in I \mid g_j(x) = 0\},$$

$$S(x) = \{\nabla g_j(x) \mid j \in I(x)\},$$

and

$$T(x) = \{\nabla h_j(x) \mid j \in J\}.$$

We call a feasible point x a *Karush–Kuhn–Tucker (KKT) point* of (1.1) if there exist vectors $u \in \Re^m$ and $v \in \Re^p$ such that the following requirements are simultaneously

*Received by the editors August 29, 1997; accepted for publication (in revised form) October 12, 1999; published electronically June 6, 2000. This work was supported by the Australian Research Council.

http://www.siam.org/journals/siopt/10-4/32662.html

†School of Mathematics, The University of New South Wales, Sydney, New South Wales 2052, Australia (L.Qi@unsw.edu.au). Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong.

‡Department of Mathematics, Guangxi University, Nanning, Guangxi, China (zxwei@gxu.edu.cn).

satisfied:

$$\begin{cases} \nabla f(x) + \sum_{j \in I} u_j \nabla g_j(x) + \sum_{j \in J} v_j \nabla h_j(x) = 0; \\ u \geq 0; \\ u^T g(x) = 0. \end{cases} \tag{1.2}$$

We call the pair (u, v) a *Lagrange multiplier* at x and denote the set of all possible Lagrange multipliers associated with x by $M(x)$. For a given $x \in X$, if we regard (1.2) as the constraints of a linear program with (u, v) as variables, we see that if x is a KKT point, there is a $(u, v) \in M(x)$ such that the vectors $\{\nabla g_j(x) \mid j \in \text{supp}(u)\} \cup \{\nabla h_j(x) \mid j \in \text{supp}(v)\}$ are linearly independent. We call such a (u, v) a *regular Lagrange multiplier* of x.

For convenience, we set $M(x) = \emptyset$ if x is not a KKT point. We say that x is an *isolated KKT point* of (1.1) if there is a neighborhood of x such that x is the only KKT point in this neighborhood. Note that an isolated KKT point may have more than one Lagrange multiplier.

In *sequential quadratic programming (SQP) methods* [7, 10, 23, 24, 9] and KKT equation methods [26] for solving (1.1), at each step, an approximate KKT point of (1.1) is found. Is any limit point of a sequence of approximate KKT points a KKT point of (1.1)? If it is, will the whole sequence converge to it? Under which conditions is a KKT point stable with respect to perturbations? In the next section, we formally define an *approximate KKT point sequence* and introduce a regularity condition called the *constant positive linear dependence condition (CPLD).* The CPLD is weaker than the well-known Mangasarian–Fromovitz constraint qualification (MFCQ) [16] and the constant rank constraint qualification (CRCQ); moreover, the MFCQ and the CRCQ together are weaker than the linear independence constraint qualification (LICQ). We show that a limit point x^* of an approximate KKT point sequence is a KKT point of (1.1) if the CPLD holds at x^*. In section 3, we show that if a KKT point x satisfies the CPLD and the strong second-order sufficiency conditions (SSOSC) [31], then it is an isolated KKT point. Hence, a limit point x^* of an approximate KKT point sequence is a KKT point of (1.1) and the whole sequence will converge to it if both the CPLD and the SSOSC hold at x^*. We state in section 3 a Kojima theorem on perturbed KKT points under the MFCQ and the SSOSC. The Kojima theorem will be used in section 6.

SQP methods constitute an important class of methods for solving (1.1). They enjoy local superlinear convergence under mild conditions [7, 10, 23, 24, 9]. The superlinear convergence of SQP methods was first established [7, 10] under a set of conditions: the LICQ, the second-order sufficiency conditions, and the strict complementarity slackness. This set of conditions was first studied in [5] and is called the Jacobian uniqueness condition [10]. Robinson [31] reduced the second-order sufficiency conditions and the strict complementarity slackness to the SSOSC. Robinson's condition has been used for classical SQP methods and KKT equations methods in [1, 9, 26]. What is difficult is to relax the LICQ. The relaxation of the LICQ may result in multiple Lagrange multipliers. Only recently, several authors [6, 28, 35] began to study the convergence of algorithms on problems with nonunique Lagrange multipliers. In section 4, we apply the results in sections 2 and 3 to a general SQP method and establish its convergence under the CPLD and the SSOSC. In sections 5 and 6, we further apply these results to a feasible SQP method. For classical SQP methods, the iteration points may be infeasible, while feasible SQP methods take special precautions to guarantee that the iteration points are feasible. Panier and Tits [17, 18] proposed two feasible SQP methods in 1987 and 1993. They established global and superlinear convergence of their feasible SQP methods under the classical Jacobian uniqueness condition. In section 5, we establish global convergence of the 1993 Panier–Tits method [18] under the condition that the CPLD and the SSOSC hold for a limit point of the primal iterative sequence and the MFCQ holds at all non-KKT points in X. In section 6, we first modify the 1993 Panier–Tits algorithm slightly; then, with the help of the Kojima theorem, we establish superlinear convergence of the modified algorithm under the SSOSC and a condition slightly weaker than the LICQ. In this way, both the strict complementarity condition and the LICQ, assumed in [18], are relaxed. These results can be extended to the 1987 Panier–Tits method.

Throughout the paper, we denote the Euclidean norm of a vector v by $\|v\|$, the corresponding induced norm of a matrix H by $\|H\|$, and the cardinality of a finite set J by $|J|$ and let $\mathcal{N} \equiv \{1, 2, \ldots\}$.

2. Limiting point of an approximate KKT point sequence.

We first review the concept of positive linear independence for vectors [21, 32, 33, 29].

DEFINITION 2.1. *Let $A = \{a^1, \ldots, a^l\}$ and $B = \{b^1, \ldots, b^r\}$ be two finite subsets of $\Re^n$ such that $A \cup B \neq \emptyset$. We say that (A, B) is positive-linearly dependent if there are $\alpha \in \Re^l$ and $\beta \in \Re^r$ such that $\alpha \geq 0, (\alpha, \beta) \neq 0$, and*

$$\sum_{j=1}^{l} \alpha_j a^j + \sum_{j=1}^{r} \beta_j b^j = 0.$$

Otherwise, we say that (A, B) is positive-linearly independent. *If $B = \emptyset$, we simply say that A is positive-linearly dependent or independent.*

Clearly, just as linearly independent and dependent sets, a subset pair of a positive-linearly independent set pair is always positive-linearly independent and a set pair with a positive-linearly dependent subset pair is always positive-linearly dependent.

PROPOSITION 2.2. *Let $G_j : \Re^n \to \Re^n, j = 1, \ldots, l$, and $H_j : \Re^n \to \Re^n, j = 1, \ldots, r$, be continuous functions. If $(\{G_j(x)\}_{j=1}^l, \{H_j(x)\}_{j=1}^r)$ is positive-linearly independent for $x \in \Re^n$, then there is a neighborhood $N(x)$ of x such that for any $y \in N(x)$, $(\{G_j(y)\}_{j=1}^l, \{H_j(y)\}_{j=1}^r)$ is positive-linearly independent.*

Proof. If such a neighborhood does not exist, then there is a sequence $\{y^k\}_{k=1}^{\infty} \subset \Re^n$ with $y^k \to x$ as $k \to +\infty$ and $\alpha^k \geq 0, \|(\alpha^k, \beta^k)\| \equiv 1$, such that

$$\sum_{j=1}^{l} \alpha_j^k G_j(y^k) + \sum_{j=1}^{r} \beta_j^k H_j(y^k) = 0.$$

Without loss of generality, we may assume that $\alpha^k \to \alpha^*$ and $\beta^k \to \beta^*$ as $k \to +\infty$. Clearly,

$$\begin{cases} \sum_{j=1}^{l} \alpha_j^* G_j(x) + \sum_{j=1}^{r} \beta_j^* H_j(x) = 0; \\ \alpha^* \geq 0; \\ \|(\alpha^*, \beta^*)\| = 1. \end{cases}$$

This gives a contradiction. ◻

This proposition will be used in later sections.

PROPOSITION 2.3. *For any given $x \in X$, assume that $\nabla f(x)$, $\nabla g_j(x), j \in I(x)$, and $\nabla h_j(x), j \in J$, are not all zero. Then x is a KKT point of* (1.1), *i.e.*, $M(x) \neq \emptyset$, *if and only if there is a subset $S_0(x) \subseteq S(x)$ and a subset $T_0(x)$ of $T(x)$ such that $(S_0(x), T_0(x))$ is positive-linearly independent while $(S_0(x) \bigcup \{\nabla f(x)\}, T_0(x))$ is positive-linearly dependent.*

Proof. The case when $\nabla f(x) = 0$ is trivial. Thus we assume that $\nabla f(x) \neq 0$.

[$\Rightarrow$]. If $M(x) \neq \emptyset$, then there exist vectors $u = u^1 \in \Re^m$ and $v = v^1 \in \Re^p$ such that (1.2) holds. Let $I_1 = \mathrm{supp}(u)$, $J_1 = \mathrm{supp}(v)$, $S_1 = \{\nabla g_j(x) \mid j \in I_1\}$, and $T_1 = \{\nabla h_j(x) \mid j \in J_1\}$. Since $\nabla f(x) \neq 0$, by the first equality of (1.2), $S_1 \cup T_1 \neq \emptyset$. If (S_1, T_1) is positive-linearly independent, then let $S_0(x) = S_1$ and $T_0(x) = T_1$, and the first equality in (1.2) implies that $(S_0(x) \bigcup \{\nabla f(x)\}, T_0(x))$ is positive-linearly dependent. If (S_1, T_1) is positive-linearly dependent, then we have $\alpha_j \geq 0, j \in I_1$, and $\beta_j, j \in J_1$, such that not all of α_j and β_j are zero and

$$\sum_{j \in I_1} \alpha_j \nabla g_j(x) + \sum_{j \in J_1} \beta_j \nabla h_j(x) = 0.$$

If some $\alpha_j \neq 0$, let $\lambda = \min\{\frac{u_j}{\alpha_j} \mid j \in I_1, \alpha_j \neq 0\}$; otherwise, there is $\bar{j} \in J_1$ such that $\beta_{\bar{j}} \neq 0$, and we then let $\lambda = \frac{v_{\bar{j}}}{\beta_{\bar{j}}}$. Let $u_j^2 = u_j - \lambda \alpha_j$ for $j \in I_1$, $u_j^2 = 0$ for $j \notin I_1$, $v_j^2 = v_j - \lambda \beta_j$ for $j \in J_1$, and $v_j^2 = 0$ for $j \notin J_1$. Then $(u, v) = (u^2, v^2)$ still satisfies (1.2) but its support set is strictly contained in $I_1 \cup J_1$, the support sets of S_1 and T_1. Repeat this process. Finally, we have a subset $S_0(x)$ of $S(x)$ and a subset $T_0(x)$ of $T(x)$, which satisfy the requirements.

[$\Leftarrow$]. Assume that $I_0 \subseteq I(x)$ and $J_0 \subseteq J$ such that $S_0(x) = \{\nabla g_j(x) \mid j \in I_0\}$ and $T_0(x) = \{\nabla h_j(x) \mid j \in J_0\}$ satisfy the requirements. The fact that $(S_0(x) \bigcup \{\nabla f(x)\}, T_0(x))$ is positive-linearly dependent implies that there are $\gamma \in \Re$, $\alpha \in \Re^{|I_0|}$, and $\beta \in \Re^{|J_0|}$ such that $\gamma \geq 0, \alpha \geq 0, (\gamma, \alpha, \beta) \neq 0$, and

$$\gamma \nabla f(x) + \sum_{j \in I_0} \alpha_j \nabla g_j(x) + \sum_{j \in J_0} \beta_j \nabla h_j(x) = 0.$$

These and the assumption that $(S_0(x), T_0(x))$ is positive-linearly independent imply that $\gamma > 0$. Let $u_j = \frac{\alpha_j}{\gamma}$ for $j \in I_0$, $u_j = 0$ for $j \notin I_0$, $v_j = \frac{\beta_j}{\gamma}$ for $j \in J_0$, and $v_j = 0$ for $j \notin J_0$. Then (u, v) satisfies (1.2). Hence, $M(x) \neq \emptyset$. ◻

A given feasible point $x \in X$ is said to satisfy the MFCQ [16] if $T(x)$ is linearly independent and there is a vector $z \in \Re^n$ such that

$$(\nabla g_{I(x)}(x))^T z < 0$$

and

$$(\nabla h(x))^T z = 0.$$

The following proposition was given in section 1.8 of [21].

PROPOSITION 2.4. *For any given $x \in X$, assume that $I(x) \cup J \neq \emptyset$. Then the MFCQ holds at x if and only if $(S(x), T(x))$ is positive-linearly independent.*

Proof. If $I(x) = \emptyset$, the conclusion is obvious; otherwise, the conclusion follows Motzkin's theorem of the alternative [16].

We now define an approximate KKT point sequence of (1.1).

DEFINITION 2.5. *We say that $\{x^k\}_{k=1}^{\infty} \subset \Re^n$ is an approximate KKT point sequence of* (1.1) *if there is a sequence $\{(u^k, v^k, \epsilon^k, \delta^k, \lambda^k)\}_{k=1}^{\infty} \subset \Re^m \times \Re^p \times \Re^n \times \Re^m \times \Re$ such that the following requirements are simultaneously satisfied for each k:*

$$(2.1) \qquad \begin{cases} \nabla f(x^k) + \sum_{j \in I} u_j^k \nabla g_j(x^k) + \sum_{j \in J} v_j^k \nabla h_j(x^k) = \epsilon^k; \\ g(x^k) \leq \delta^k; \\ u^k \geq 0; \\ (u^k)^T (g(x^k) - \delta^k) = 0; \\ \|h(x^k)\| \leq \lambda^k; \end{cases}$$

and $\{(\epsilon^k, \delta^k, \lambda^k)\}_{k=1}^{\infty}$ converges to zero as $k \to \infty$.

Such an approximate KKT point sequence is produced by SQP methods, KKT equations methods, and some other methods for solving (1.1). If x^* is a limit point of $\{x^k\}$, or without loss of generality, if $\{x^k\}$ converges to x^*, is x^* a KKT point of (1.1)? To answer this question, we introduce a regularity condition.

DEFINITION 2.6. *A given feasible point $x \in X$ is said to satisfy the CPLD if for any $I_0 \subseteq I(x)$ and $J_0 \subseteq J$, whenever $(\{\nabla g_j(x) \mid j \in I_0\}, \{\nabla h_j(x) \mid j \in J_0\})$ is positive-linearly dependent, there is a neighborhood $N(x)$ of x such that for any $y \in N(x)$, $(\{\nabla g_j(y) \mid j \in I_0\}, \{\nabla h_j(y) \mid j \in J_0\})$ is linearly dependent.*

Note that in the definition we do not require that $(\{\nabla g_j(y) \mid j \in I_0\}, \{\nabla h_j(y) \mid j \in J_0\})$ be positive-linearly dependent, which is stronger than our requirement here. By Propositions 2.2 and 2.4, the CPLD is weaker than the MFCQ.

It is said that the CRCQ [11, 15, 20, 34, 22] holds at $x \in X$ if there is a neighborhood $N(x)$ of x such that for every $I_0 \subseteq I(x)$ and $J_0 \subseteq J$, the family of gradient vectors

$$\{\nabla g_j(y) \mid j \in I_0\} \bigcup \{\nabla h_j(y) \mid j \in J_0\}$$

has the same rank (which depends on I_0 and J_0) for all vectors $y \in N(x)$. It is not difficult to see that the CRCQ holds at x if and only if for any $I_0 \subseteq I(x)$ and $J_0 \subseteq J$, whenever $\{\nabla g_j(x) \mid j \in I_0\} \bigcup \{\nabla h_j(x) \mid j \in J_0\}$ is linearly dependent, there is a neighborhood $N(x)$ of x such that for any $y \in N(x)$, $\{\nabla g_j(y) \mid j \in I_0\} \bigcup \{\nabla h_j(y) \mid j \in J_0\}$ is linearly dependent. Hence, the CPLD is also weaker than the CRCQ. Note [11] that neither the CRCQ implies the MFCQ nor the MFCQ implies the CRCQ. Furthermore, even the MFCQ and the CRCQ together are weaker than the LICQ. This can be seen from the following example: $n = m = 2, p = 0, g_1(x) = x_1 + x_2, g_2(x) = 2x_1 + 2x_2$, at $x = (0, 0)^T$. If x is a local minimum point of (1.1) and the CPLD holds at x, is x always a KKT point of (1.1)? If so, we may also call the CPLD a constraint qualification, but at this moment we only use the CPLD to derive the following result.

THEOREM 2.7. *If an approximate KKT point sequence $\{x^k\}_{k=1}^{\infty}$ converges to x^* as $k \to \infty$ and the CPLD holds at x^*, then x^* is a KKT point of* (1.1)*, i.e., there are a $u^* \in \Re^m$ and a $v^* \in \Re^p$ such that (x^*, u^*, v^*) satisfies* (1.2).

Proof. By using the theory of linear programming, we may assume, without loss of generality, for any given k, there is $(\bar{u}^k, \bar{v}^k)$ satisfying (2.1) such that

$$\{\nabla g_j(x^k) \mid j \in \mathrm{supp}(\bar{u}^k)\} \bigcup \{\nabla h_j(x^k) \mid j \in \mathrm{supp}(\bar{v}^k)\}$$

is linearly independent. Let $I_k = \text{supp}(\bar{u}^k)$ and $J_k = \text{supp}(\bar{v}_k)$. Without loss of generality, we may assume that $I_0 \equiv I_k$ and $J_0 \equiv J_k$. Then $I_0 \subseteq I(x^*)$ and $J_0 \subseteq J$. If $\{(\bar{u}^k, \bar{v}^k)\}_{k=1}^{\infty}$ has a bounded subsequence, then, without loss of generality, we may assume that there are $u^* \in \Re^m$ and $v^* \in \Re^p$ such that $\bar{u}^k \to u^*$ and $\bar{v}^k \to v^*$ as $k \to \infty$. Letting k tend to infinity in (2.1), we see that (x^*, u^*, v^*) satisfies (1.2), and hence the conclusion holds for this case. We assume now that

$$\lim_{k\to\infty} \|(u^k, v^k)\| = +\infty.$$

Without loss of generality, we may assume that

$$\lim_{k\to\infty} \frac{(u^k, v^k)}{\|(u^k, v^k)\|} = (\alpha, \beta).$$

Then $\|(\alpha, \beta)\| = 1$, $\text{supp}(\alpha) \subseteq I_0$, $\text{supp}(\beta) \subseteq J_0$, and $\alpha \geq 0$. Dividing both sides of

$$\nabla f(x^k) + \sum_{j\in I} \bar{u}_j^k \nabla g_j(x^k) + \sum_{j\in J} \bar{v}_j^k \nabla h_j(x^k) = \epsilon^k$$

by $\|(u^k, v^k)\|$ and letting k tend to infinity in the above equality, we obtain

$$\sum_{j\in I_0} \alpha_j \nabla g_j(x^*) + \sum_{j\in J_0} \beta_j \nabla h_j(x^*) = 0.$$

This implies that $(\{\nabla g_j(x^*) \ \mid \ j \in I_0\}, \{\nabla h_j(x^*) \ \mid \ j \in J_0\})$ is positive-linearly dependent. By the assumptions that the CPLD holds at x^* and $x^k \to x^*$, we have that for all large k, $(\{\nabla g_j(x^k) \mid j \in I_0\}, \{\nabla h_j(x^k) \mid j \in J_0\})$ are linearly dependent. This contradicts the fact that $\{\nabla g_j(x^k) \mid j \in I_0\} \bigcup \{\nabla h_j(x^k) \mid j \in J_0\}$ are linearly independent for all k. $\square$.

3. Isolated and stable KKT points. We now assume that f, g, and h are twice continuously differentiable.

For any $x \in \Re^n, u \in \Re^m$, and $v \in \Re^p$, we denote the Lagrange function of (1.1) by

$$L(x, u, v) = f(x) + u^T g(x) + v^T h(x).$$

By Robinson [31], a triplet (x, u, v) is said to satisfy the SSOSC if it satisfies the KKT conditions (1.2) and $\nabla_{xx} L(x, u, v)$ is positive definite on the subspace

$$\begin{aligned} G(x, u, v) = \{d \in \Re^n \ \mid \ \nabla f(x)^T d = 0, \ \ \nabla g_j(x)^T d = 0 \ \text{ for } \ \ j \in \text{supp}(u), \\ \nabla h_j(x)^T d = 0 \ \text{ for } \ j \in J\}. \end{aligned}$$

Note that even under the second-order sufficiency conditions, x will be a strict local minimum of (1.1).

DEFINITION 3.1. *Suppose that x is a KKT point of* (1.1). *If for all Lagrange multipliers (u, v) of x, (x, u, v) satisfies the SSOSC, then we say that x satisfies the SSOSC.*

THEOREM 3.2. *Suppose that x^* is a KKT point of* (1.1). *If x^* satisfies the CPLD and the SSOSC, then x^* is an isolated KKT point of* (1.1).

Proof. Suppose that x^* is not an isolated KKT point of (1.1). Then there is a KKT point sequence $\{x^k\}_{k=1}^{\infty}$ such that $x^k \neq x^*$ and $\lim_{k\to\infty} x^k = x^*$. It follows from the theory of linear programming that for each k, there is a regular Lagrange multiplier (u^k, v^k) for x^k. Let $I_k = \text{supp}(u^k)$ and $J_k = \text{supp}(v_k)$. Without loss of generality, we may assume that $I_0 \equiv I_k$ and $J_0 \equiv J_k$ for all k. Then

$$\{\nabla g_j(x^k) \ \mid \ j \in I_0\} \bigcup \{\nabla h_j(x^k) \ \mid \ j \in J_0\}$$

is linearly independent for all k. By the CPLD at x^*,

$$(\{\nabla g_j(x^*) \ \mid \ j \in I_0\}, \{\nabla h_j(x^*) \ \mid \ j \in J_0\})$$

is positive-linearly independent.

If $\{(u_k, v_k)\}_{k=1}^{\infty}$ is unbounded, without loss of generality, we may assume that

$$\lim_{k \to \infty} \|(u^k, v^k)\| = +\infty,$$

$$\lim_{k \to \infty} \frac{(u^k, v^k)}{\|(u^k, v^k)\|} = (\alpha, \beta),$$

$\|(\alpha, \beta)\| = 1, \alpha \geq 0, \text{supp}(\alpha) \subseteq I_0$, and $\text{supp}(\beta) \subseteq J_0$. Then dividing

$$\nabla f(x^k) + \sum_{j \in I} u_j^k \nabla g_j(x^k) + \sum_{j \in J} v_j^k \nabla h_j(x^k) = 0$$

by $\|(u^k, v^k)\|$ and letting $k \to \infty$, we have

$$\sum_{j \in I_0} \alpha_j \nabla g_j(x^*) + \sum_{j \in J_0} \beta_j \nabla h_j(x^*) = 0.$$

This contradicts the fact that

$$(\{\nabla g_j(x^*) \ \mid \ j \in I_0\}, \{\nabla h_j(x^*) \ \mid \ j \in J_0\})$$

is positive-linearly independent.

Hence $\{(u^k, v^k)\}_{k=1}^{\infty}$ is bounded. Without loss of generality, we may assume that $u^k \to u^*$ and $v^k \to v^*$. Then $(u^*, v^*) \in M(x^*)$ is a Lagrange multiplier of x^*, $\text{supp}(u^*) \subseteq I_0$, and $\text{supp}(v^*) \subseteq J_0$. We may assume that

$$\lim_{k \to \infty} \frac{x^k - x^*}{\|x^k - x^*\|} = d.$$

Then $\|d\| = 1$. Since

$$g_j(x^k) - g_j(x^*) = 0, \qquad j \in I_0,$$

and

$$h_j(x^k) - h_j(x^*) = 0, \qquad j \in J,$$

we have, by Taylor's theorem, that

$$g_j(x^k) - g_j(x^*) = \nabla g_j(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|), \qquad j \in I_0, \tag{3.1}$$

and

$$h_j(x^k) - h_j(x^*) = \nabla h_j(x^*)^T(x^k - x^*) + o(\|x^k - x^*\|), \qquad j \in J. \tag{3.2}$$

Dividing (3.1) and (3.2) by $\|x^k - x^*\|$ and letting $k \to \infty$, we have

$$\nabla g_j(x^*)^T d = 0, \qquad j \in I_0, \tag{3.3}$$

and

$$\nabla h_j(x^*)^T d = 0, \qquad j \in J. \tag{3.4}$$

On the other hand, since $(u^*, v^*) \in M(x^*)$, we have

$$\nabla f(x^*) + \sum_{j \in \mathrm{supp}(u^*)} u_j^* \nabla g_j(x^*) + \sum_{j \in \mathrm{supp}(v^*)} v_j^* \nabla h_j(x^*) = 0.$$

This formula, combined with (3.3), (3.4) and observing that $\mathrm{supp}(u^*) \subseteq I_0$ and $\mathrm{supp}(v^*) \subseteq J$, yields

$$\nabla f(x^*)^T d = 0. \tag{3.5}$$

From (3.3), (3.4), and (3.5), we have $d \in G(x^*, u^*, v^*)$. For any given k and $t \in [0,1]$, let

$$(x^t, u^t, v^t) = (1-t)(x^*, u^*, v^*) + t(x^k, u^k, v^k).$$

Then, Robinson's function [31, 8] is defined by

$$\begin{aligned} s(t) = (x^k - x^*)^T &\left[\nabla f(x^t) + \sum_{j \in I_0} u_j^t \nabla g_j(x^t) + \sum_{j \in J_0} v_j^t \nabla h_j(x^t)\right] \\ &- (u^k - u^*)^T g(x^t) - (v^k - v^*)^T h(x^t). \end{aligned}$$

The function $s : [0,1] \to \Re$ is clearly continuous on $[0,1]$ and continuously differentiable on $(0,1)$. Moreover, $s(0) = 0 = s(1)$. By the mean-value theorem, for any given k, there exists $t_k \in (0,1)$ such that $s'(t_k) = 0$, i.e.,

$$(x^k - x^*)^T \nabla_{xx} L(x^{t_k}, u^{t_k}, v^{t_k})(x^k - x^*) = 0.$$

Dividing this inequality by $\|x^k - x^*\|^2$ and passing to the limit $k \to \infty$, we obtain

$$d^T \nabla_{xx} L(x^*, u^*, v^*) d = 0.$$

This formula, combined with the facts that $d \in G(x^*, u^*, v^*)$ and $\nabla_{xx} L(x^*, u^*, v^*)$ is positive definite in $G(x^*, u^*, v^*)$, implies that $d = 0$, which contradicts the fact that $\|d\| = 1$. This proves the theorem. □

Remark. It is possible to reduce the requirement of twice differentiability of f, g, and h to semismoothness of $\nabla f, \nabla g$, and ∇h. Such an optimization problem is called an SC^1 optimization problem. For SC^1 optimization and its applications, see [24, 19, 4, 9, 3, 12, 27, 13, 2, 25].

THEOREM 3.3. *Suppose that x^* is a limit point of an approximate KKT point sequence $\{x^k\}_{k=1}^\infty$ of* (1.1) *and the CPLD and the SSOSC hold at x^*. If*

$$\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0, \tag{3.6}$$

then $\lim_{k\to\infty} x^k = x^*$.

Proof. By Theorem 2.7, we have that x^* is a KKT point of (1.1) and every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is a KKT point of (1.1). The assumptions that the CPLD and the SSOSC hold at x^* and Theorem 3.2 imply that x^* is an isolated KKT point of (1.1), i.e., there is $\epsilon > 0$ such that the ball $O(x^*, \epsilon) = \{x \in \Re^n, \ | \ \|x - x^*\| \le \epsilon\}$ does not contain any KKT point other than x^*. On the other hand, (3.6) implies that for k large enough, $\|x^{k+1} - x^k\| < \frac{\epsilon}{4}$ and there exists a subsequence $\{x^k\}_{k\in\mathcal{K}}$ such that $\|x^k - x^*\| < \frac{\epsilon}{4}$ on $\mathcal{K}$. It is then impossible to leave $O(x^*, \epsilon)$ without creating another cluster point and hence a KKT point in that ball. □

In the remaining part of this section, as in [14], $\|\cdot\|$ is for the infinity norm instead of the Euclidean norm. Let $N(x, \delta) = \{y \in \Re^n : \|y - x\| \le \delta\}$.

Consider the perturbed form of (1.1)

$$\min\{f(x) + \bar{f}(x) \ \mid \ x \in \bar{X}\}, \tag{3.7}$$

where $\bar{X} = \{x \in \Re^n \ \mid \ g(x) + \bar{g}(x) \le 0, \quad h(x) + \bar{h}(x) = 0\}$, $f, \bar{f} : \Re^n \to \Re$, $g, \bar{g} : \Re^n \to \Re^m$, and $h, \bar{h} : \Re^n \to \Re^p$ are twice continuously differentiable functions.

DEFINITION 3.4. *Let x^* be a KKT point of* (1.1). *We call x^* a strongly stable KKT point of* (1.1) *if for some $\delta^* > 0$ and each $\delta \in (0, \delta^*]$ there exists an $\alpha > 0$ such that whenever twice continuously differentiable functions $\bar{f}, \bar{g}$, and $\bar{h}$ satisfy*

$$\sup_{\substack{\|x-x^*\|\le\delta^* \\ i\in I, j\in J}} \{|\bar{f}(x)|, |\bar{g}_i(x)|, |\bar{h}_j(x)|, \|\nabla \bar{f}(x)\|, \|\nabla \bar{g}_i(x)\|, \|\nabla \bar{h}_j(x)\|, \|\nabla^2 \bar{f}(x)\|, \|\nabla^2 \bar{g}_i(x)\|, \|\nabla^2 \bar{h}_j(x)\|\} \le \alpha,$$

$N(x^*, \delta)$ contains a solution $\bar{x}^*$ of (3.7), *which is unique in $N(x^*, \delta^*)$.*

The following theorem is Theorem 7.2 of [14]. We will use it in section 6.

THEOREM 3.5 (by Kojima [14]). *Suppose that x^* is a KKT point of* (1.1) *and that the MFCQ holds at x^*. Then x^* is a strongly stable KKT point of* (1.1) *if and only if for all $(u, v) \in M(x^*)$, (x^*, u, v) satisfies the SSOSC.*

Remark. The Kojima theorem can be regarded as an alternative to Robinson's perturbation theorem in [30]. Theorem 4.1 of [30] (together with Theorem 2.1 and Corollary 2.2 of the same paper) shows that under the SSOSC and the LICQ one has Lipschitzian behavior of the solution and the multipliers, with respect to perturbations, while the Kojima theorem shows that under the SSOSC and the MFCQ one has continuity of the solution and the multipliers, with respect to perturbations (but a counterexample in [31] shows that we cannot prove Lipschitz continuity in this situation). It is thus not surprising that in section 6 we must add the CRCQ to get our superlinear convergence result for a modified version of the 1993 Panier–Tits method. Note that the example in [31] does not satisfy the CRCQ. A question is, Is the Kojima theorem still true if the MFCQ is replaced by the CRCQ?

4. A general SQP method.

We describe a general SQP method as follows.

ALGORITHM A.

Let $C > 0$.

Data. $x^0 \in X, H_0 \in \Re^{n\times n}$, symmetric positive definite.

Step 0. (Initialization.) Set $k = 0$.

Step 1. (Computation of a search direction.) Compute d^k by solving the quadratic program

$$(QP)\begin{cases} \min \;\; \frac{1}{2}d^T H_k d + \nabla f(x^k)^T d \\ \\ \text{s.t.} \;\; g_j(x^k) + \nabla g_j(x^k)^T d \le 0, \quad j \in I, \\ \text{s.t.} \;\; h_j(x^k) + \nabla h_j(x^k)^T d = 0, \quad j \in J. \end{cases}$$

If $d^k = 0$ stop.
Step 2. (Line search and additional correction.) Determine the steplength $\alpha_k \in (0,1)$ and a correction direction $\bar{d}^k$ such that

$$\|\bar{d}^k\| \le C\|d^k\|. \tag{4.1}$$

Step 3. (Updates.) Compute a new symmetric positive definite approximation H_{k+1} to the Hessian of the Lagrangian. Set $x^{k+1} = x^k + \alpha_k d^k + \bar{d}^k$ and $k = k+1$. Go back to Step 1.

Algorithm A is a general model for SQP methods. For a specific SQP method, the rules for determining $\alpha_k, \bar{d}^k$, and H_k must be given. For classical SQP methods [23], $\bar{d}^k = 0$. We assume that the quadratic program (QP) is always solvable. This is obvious for feasible SQP methods since 0 is a feasible solution of (QP) in that case. Checking the KKT conditions of (QP) for $d = 0$, we have the following proposition.

PROPOSITION 4.1. *If Algorithm A stops in Step* 1*, then x^k is a KKT point of* (1.1).

Hence, we need only consider the case where Algorithm A generates an infinite sequence.

THEOREM 4.2. *Assume that Algorithm A generates an infinite sequence $\{x^k\}_{k=1}^{\infty}$ and that this sequence has an accumulation point x^*. Let $\mathcal{K}$ be a subsequence of $\mathcal{N}$ such that*

$$\lim_{k\in\mathcal{K}} x^k = x^*.$$

Suppose that the CPLD holds at x^* and that the Hessian estimates $\{H_k\}_{k=0}^{\infty}$ are bounded, i.e., there exists a scalar $C_1 > 0$ such that for all k

$$\|H_k\| \le C_1. \tag{4.2}$$

If

$$\liminf_{k\in\mathcal{K}} \|d^k\| = 0, \tag{4.3}$$

then x^* is a KKT point of (1.1).

Proof. Without loss of generality, by passing to a subsequence if necessary, we may assume that

$$\lim_{k\in\mathcal{K}} \|d^k\| = 0. \tag{4.4}$$

By the KKT conditions of (QP), we have

$$\begin{cases} H_k d^k + \nabla f(x^k) + \nabla g(x^k)^T \bar{u}^k + \nabla h(x^k)^T u^k = 0; \\ g(x^k) + \nabla g(x^k)^T d^k \le 0; \\ u^k \ge 0; \\ (u^k)^T (g(x^k) - \nabla g(x^k)^T d^k) = 0; \\ h(x^k) + \nabla h(x^k)^T d^k = 0. \end{cases}$$

By (4.2) and (4.4), as $k \to \infty$ for $k \in \mathcal{K}$, we have that

$$\epsilon_k \equiv -H_k d^k \to 0,$$

$$\delta_k \equiv -\nabla g(x^k)^T d^k \to 0,$$

and

$$\lambda_k \equiv \|\nabla h(x^k)^T d^k\| \to 0.$$

Then, by Theorem 2.7, x^* is a KKT point of (1.1). ◻

THEOREM 4.3. *Assume that the conditions of Theorem* 4.2 *hold. If, furthermore,* f, g, *and* h *are twice continuously differentiable, the SSOSC holds at* x^*, *and*

$$\lim_{k \to \infty} d^k = 0, \tag{4.5}$$

then $\lim_{k\to\infty} x^k = x^*$.

Proof. This follows from (4.1) and Theorem 3.3. ◻

We can establish superlinear convergence of the general SQP method by following, step by step, with minor modifications, the proofs of Lemma 3 to Theorem 1 of [23] and replacing $\nabla^2_{xx} L(x^*, u^*, v^*)$ with $\nabla^2_{xx} L(x^k, u^k, v^k)$ in (3.10) of [23]. We will see this more clearly in section 6.

To establish (4.3) or (4.5) one must use the properties of specific SQP methods. In the next section, we will establish these two conditions for a feasible SQP method.

5. Global convergence of a Panier–Tits method. In this section, we establish the global convergence of the 1993 Panier–Tits feasible SQP method [18] under the SSOSC and a condition slightly weaker than the MFCQ. The global convergence of the 1987 Panier–Tits method [17] can be established in the same way. First of all, we describe the algorithm given in [18]. Keep in mind that the Panier–Tits methods are for inequality constrained optimization problems. Therefore, in this section and the next section, problem (1.1) becomes

$$\min\{f(x) \mid x \in X\}, \tag{5.1}$$

where $X = \{x \in \Re^n \mid g(x) \le 0\}$.

5.1. A Panier–Tits method. In [18], a continuous map $d_1 : \Re^n \to \Re^n$ is needed in the algorithm such that

$$d_1(x) = 0 \text{ if } x \text{ is a KKT point of (5.1),} \tag{5.2}$$

$$\nabla f(x)^T d_1(x) < 0 \text{ if } x \text{ is not a KKT point of (5.1),} \tag{5.3}$$

and

$$\nabla g_j(x)^T d_1(x) < 0 \text{ if } x \text{ is not a KKT point of (5.1) and } j \in I(x). \tag{5.4}$$

As indicated in [18], if the LICQ holds at x, then the continuous map $d_1(x)$ satisfying (5.2), (5.3), and (5.4), for example, can be obtained as the solution of

$$\min \ \frac{1}{2}\|d\|^2 + \max\{\nabla f(x)^T d; \ \max\{g_j(x) + \nabla g_j(x)^T d \ | \ j \in I\}\}. \tag{5.5}$$

We see from (5.4) that the existence of such a $d_1(x)$ implies that the MFCQ holds at all non-KKT points. On the other hand, if the MFCQ holds at all non-KKT points, then such a continuous map still exists (see section 2.6 of [21]). However, this does not require that the MFCQ hold at KKT points.

In the method of [18], it is necessary to have a map $\rho : \Re^n \to [0, 1]$ that is bounded away from zero outside every neighborhood of zero, and for v small

$$\rho(v) = O(\|v\|^2).$$

Since the existence of the map ρ is independent of problem (5.1), for sake of simplicity, we choose

$$\rho(v) = \frac{\|v\|^2}{1 + \|v\|^2}.$$

Establishing the convergence properties of the algorithm presents no difficulty when choosing other such maps.

The 1993 Panier–Tits method is as follows.

ALGORITHM B.

Let $C > 0, \tau_1 \in (0, \frac{1}{2}), \tau_2 \in (0, 1), \tau_3 \in (2, 3)$.
Data. $x^0 \in X, H_0 \in \Re^{n\times n}$, symmetric positive definite.
Step 0. (Initialization.) Set $k = 0$.
Step 1. (Computation of a search arc.)

(i) Compute d_0^k by solving the quadratic program

$$(QP_1) \begin{cases} \min \ \frac{1}{2} d^T H_k d + \nabla f(x^k)^T d \\ \text{s. t. } \ g_j(x^k) + \nabla g_j(x^k)^T d \le 0, \quad j \in I. \end{cases}$$

If $d_0^k = 0$, stop.

(ii) Let d_1^k be the solution of (5.5), $\rho_k = \rho(d_0^k)$, and $d^k = (1 - \rho_k)d_0^k + \rho_k d_1^k$.

(iii) Compute a correction $\tilde{d}^k$ as the solution of the problem

$$(QP_2) \begin{cases} \min \ \frac{1}{2}(d^k + d)^T H_k(d + d^k) + \nabla f(x^k)^T (d + d^k) \\ \text{s.t. } \ g_j(x^k + d^k) + \nabla g_j(x^k)^T d \le -\|d^k\|^{\tau_3}, \quad j \in I, \end{cases}$$

if it exists and has norm less than $\min\{\|d^k\|, C\}$ and $\tilde{d}^k = 0$ otherwise. Hence, in any case, we have

$$\|\tilde{d}^k\| \leq \min \{\|d^k\|, C\}. \tag{5.6}$$

Step 2. (Arc search.)

Compute t_k, the first number t of the sequence $\{1, \tau_2, \tau_2^2, \dots\}$ satisfying

$$f(x^k + td^k + t^2\tilde{d}^k) \leq f(x^k) + \tau_1 t \nabla f(x^k)^T d^k$$

and

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq 0, \qquad j \in I.$$

Step 3. (Updates.) Compute a new symmetric positive definite approximation H_{k+1} to the Hessian of the Lagrangian. Set $x^{k+1} = x^k + t_k d^k + t_k^2 \tilde{d}^k$ and $k = k+1$. Go back to Step 1.

We see that Algorithm B is a special case of Algorithm A with d_0^k in Algorithm B playing the role of d^k in Algorithm A. The following two propositions show that Algorithm B is well defined and either stops at a KKT point of (5.1) or generates a sequence $\{x^k\}_{k=1}^{\infty}$.

Proposition 5.1 (Proposition 3.1 of [18]). *If Algorithm B stops at Step* 1(i)*, then x^k is a KKT point of* (5.1)*. If x^k is not a KKT point of* (5.1)*, d_0^k satisfies*

$$\nabla f(x^k)^T d_0^k < 0 \tag{5.7}$$

and

$$\nabla g_j(x^k)^T d_0^k \leq 0 \;\; \textit{for all} \;\; j \in I(x^k).$$

Proposition 5.2 (Proposition 3.2 of [18]). *The line search yields a step $t_k = \tau_2^i$ for some finite $i = i(k)$.*

5.2. Global convergence of Algorithm B. In order to prove the convergence properties of Algorithm B, we assume that

(H1) the Hessian estimates $\{H_k\}_{k=0}^{\infty}$ are bounded, i.e., there exists a scalar $C_1 > 0$ such that for all k, $\|H_k\| \leq C_1$;

(H2) the MFCQ holds at all non-KKT points in X.

As discussed in subsection 5.1, (H2) implies that (5.5) has a continuous solution for x. By Proposition 4.1 or Proposition 5.1, we may assume that Algorithm B generates an infinite sequence $\{x^k\}_{k=1}^{\infty}$ and $\{x^k\}_{k=1}^{\infty}$ has an accumulation point x^*. Furthermore, we assume that

(H3) the CPLD holds at x^*.

(H2) and (H3) together are slightly weaker than the condition that the MFCQ holds at all points in X.

Theorem 5.3. *Assume that the hypotheses* (H1)–(H3) *hold. Then x^* is a KKT point of* (5.1)*.*

Proof. We assume that there is $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}} x^k = x^*.$$

By Theorem 4.2, we only need to prove that

$$\liminf_{k \in \mathcal{K}} \|d_0^k\| = 0.$$

Assume that this does not hold. Then there exists a subsequence $\mathcal{K}' \subset \mathcal{K}$ and a scalar $c > 0$ such that for all $k \in \mathcal{K}'$, $\|d_0^k\| \geq c$. Suppose, by contradiction, that x^* is not a KKT point of (5.1). Then from the definitions of ρ_k and ρ, there exists a number $c_0 > 0$ such that for all $k \in \mathcal{K}'$, $\rho_k \geq c_0$. Therefore, using (5.3), (5.4), (5.7), and the definition of d^k in Step 1(ii) of Algorithm B, we have

$$\nabla f(x^k)^T d^k \leq c_0 \nabla f(x^k)^T d_1^k. \tag{5.8}$$

Similarly, for $j \in I$, we have

$$\nabla g_j(x^k)^T d^k \leq -g_j(x^k) + c_0 \nabla g_j(x^k)^T d_1^k. \tag{5.9}$$

Since x^* is not a KKT point, we may assume that

$$\lim_{k \in \mathcal{K}'} d_1^k = d_1^*, \tag{5.10}$$

$$\nabla f(x^*)^T d_1^* \leq -3c_1, \tag{5.11}$$

and

$$\nabla g_j(x^*)^T d_1^* \leq -3c_1 \quad \text{for} \quad j \in I(x^*) \tag{5.12}$$

for some $c_1 > 0$. (5.10) and (5.11) imply that, for $k \in \mathcal{K}'$ large enough,

$$\nabla f(x^k)^T d_1^k \leq -2c_1. \tag{5.13}$$

Similarly, from (5.10) and (5.12), we have, for $k \in \mathcal{K}'$ large enough, that

$$\nabla g_j(x^k)^T d_1^k \leq -2c_1 \quad \text{for} \quad j \in I(x^*). \tag{5.14}$$

Therefore, by viewing (5.8) and (5.13), (5.9) and (5.14), we have $c_2 > 0$ such that, for all $k \in \mathcal{K}'$ large enough,

$$\nabla f(x^k)^T d^k < -c_2,$$

$$\nabla g_j(x^k)^T d^k < -c_2 \quad \text{for} \quad j \in I(x^*),$$

and, by continuity of g,

$$g_j(x^k) \leq -c_2 \quad \text{for} \quad j \in I \setminus I(x^*).$$

From the definitions of ρ and d^k, we see that $\{d^k\}_{k=1}^\infty$ is bounded. From (5.6), $\{\tilde{d}^k\}_{k=1}^\infty$ is also bounded. The argument used in the proof of Proposition 3.2 of [17] implies that, in this case, the step performed by the line search is bounded away from zero. This and the monotonic decrease of $f(x^k)$ imply therefore that $\{f(x^k)\}_{k\in\mathcal{K}'}$ is unbounded, which contradicts the facts that $x^k \to x^*$ as $k \in \mathcal{K}'$ and $k \to \infty$ and the continuity of f. Hence the proof of this theorem is complete. □

In addition to (H1)–(H3), we further assume that

(H4) f and g are twice continuously differentiable;

(H5) there exists a scalar $C_2 > 0$ such that, for all k, the Hessian estimates satisfy

$$d^T H_k d \geq C_2 \|d\|^2 \quad \text{for any} \quad d \in \Re^n; \tag{5.15}$$

(H6) x^* satisfies the SSOSC.

PROPOSITION 5.4. *Assume that* (H1)–(H6) *hold and* $\{x^k\}_{k=1}^{\infty}$ *is generated by Algorithm B. Then*

$$\lim_{k\to\infty} x^k = x^*. \tag{5.16}$$

Proof. The argument used in the proof of Proposition 3.4 in [18] showed that

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0,$$

which combined with Theorem 3.3 yields that (5.16) holds. □

PROPOSITION 5.5. *Assume that* (H1)–(H6) *hold. Then*

$$\lim_{k\to\infty} d_0^k = 0.$$

Proof. By Proposition 4.3, $\{x_k\}_{k=1}^{\infty}$ is bounded. From

$$-\|\nabla f(x^k)\|\|d_0^k\| \le \nabla f(x^k)^T d_0^k \le -\frac{1}{2}(d_0^k)^T H_k d_0^k \le -\frac{1}{2}C_2\|d_0^k\|^2$$

we have

$$\|d_0^k\| \le \frac{2}{C_2}\|\nabla f(x^k)\|^2,$$

which implies that $\{d_0^k\}_{k=1}^{\infty}$ is bounded.

Suppose, by contradiction, that there exists a subsequence $\{d_0^k\}_{k\in\mathcal{K}}$ such that

$$\lim_{k\in\mathcal{K}} d_0^k = d_0^* \ne 0. \tag{5.17}$$

Since x^* is a KKT point of (1.1), there is a $u^* \ge 0$ such that

$$\begin{cases} \nabla f(x^*) + \sum_{j\in I(x^*)} u_j^* \nabla g_j(x^*) = 0, \\ u_j^* g_j(x^*) = 0 \quad \text{for} \quad j \in I(x^*), \end{cases}$$

which combined with the facts that $\nabla f(x^*)^T d_0^* \le 0$, $\nabla g_j(x^*)^T d_0^* \le 0$ for $j \in I(x^*)$ and $u^* \ge 0$ implies that

$$\nabla f(x^*)^T d_0^* = 0.$$

On the other hand, from $\nabla f(x^k)^T d_0^k \le -\frac{1}{2}C_2\|d_0^k\|^2$, $C_2 > 0$, Proposition 4.3, and $\lim_{k\in\mathcal{K}} d_0^* = d_0^*$, we have

$$0 = \nabla f(x^*)^T d_0^* \le -\frac{1}{2}C_2\|d_0^*\|^2,$$

which contradicts (5.17). This completes the proof. □

6. Superlinear convergence of a modified Panier–Tits method. We begin by modifying Algorithm B in subsection 6.1 to enable us to prove its superlinear convergence. Then we establish the superlinear convergence of the modified algorithm in subsection 6.2.

6.1. A modified Panier–Tits method. In Algorithm B, let u^k be a regular Lagrange multiplier of d_0^k with respect to (QP_1). Let $\hat{I}_k$ be the active constraint set of (QP_1). Then there is a subset $I_k \subset \hat{I}_k$ such that $u_j^k = 0$ if $j \notin I_k$ and $\{\nabla g_j(x^k) \mid j \in I_k\}$ is a maximum linearly independent subset of $\{\nabla g_j(x^k) \mid j \in \hat{I}_k\}$. We now replace (QP_2) in Algorithm B by

$$(QP_3)\begin{cases} \min \ \frac{1}{2}(d^k+d)^T H_k (d+d^k) + \nabla f(x^k)^T (d+d^k) \\ \\ \text{s.t.} \ g_j(x^k+d^k) + \nabla g_j(x^k)^T d = -\|d^k\|^{\tau_3}, \quad j \in I_k, \\ \qquad g_j(x^k+d^k) + \nabla g_j(x^k)^T d \le -\|d^k\|^{\tau_3}, \quad j \in I \setminus I_k. \end{cases}$$

We call the resulting algorithm Algorithm C. This modification forces I_k to be a part of the active constraints of (QP_3), which is necessary for the proof of superlinear convergence without assuming the LICQ and the strict complementarity slackness. Checking the proofs of subsection 5.2, we see that this modification does not affect the global convergence of the algorithm since only (5.6) is required for $\tilde{d}^k$ in the global convergence analysis in subsection 5.2. We did not make this modification in section 5 since there it was not needed.

Let R_k be the $n \times |I_k|$ matrix whose columns consist of $\nabla g_j(x^k)$ for $j \in I_k$. Note that $R_k^T R_k$ is invertible in view of the definition of the regular Lagrange multiplier. Let

$$P_k = I - R_k (R_k^T R_k)^{-1} R_k^T$$

and

$$\nabla_{xx}^2 L(x^k, u^k) = \nabla^2 f(x^k) + \sum_{j \in I_k} u_j^k \nabla^2 g_j(x^k).$$

6.2. Superlinear convergence of Algorithm C. In the following analysis, we assume that $\{x^k\}_{k=1}^\infty$ converges to a point x^*. It follows from the preceding discussion that x^* is a KKT point of (5.1). In addition to (H1)–(H6), we assume that the following hypotheses hold:

(H7) x^* satisfies the CRCQ;

(H8) whenever $B \subset I(x^*)$ and vectors in $\{\nabla g_j(x^*) | j \in B\}$ are linearly independent, $(\{\nabla g_j(x^*) | j \in I(x^*) \setminus B\}, \{\nabla g_j(x^*) | j \in B\})$ is positive-linearly independent;

(H9)

$$\frac{\|P_k(H_k - \nabla_{xx}^2 L(x^k, u^k)) P_k d^k\|}{\|d^k\|} \to 0 \text{ as } k \to \infty.$$

Note that the LICQ implies both (H7) and (H8). Thus, even (H7) and (H8) together are slightly weaker than the LICQ at x^*.

PROPOSITION 6.1. *Assume that* (H1)–(H9) *hold and that* $\{x^k\}_{k=1}^\infty$ *is generated by Algorithm C. Then for* k *large enough, the step size* t_k *is one.*

Proof. Since I_k are finite sets for all k, we may partition $\mathcal{N} \equiv \{1, 2, \dots\}$ into $l+1$ disjoint subsets $\mathcal{K}_i$ for $i = 0, 1, \dots, l$ such that $\mathcal{K}_0$ is finite, while other $\mathcal{K}_i$ are infinite and $I_k \equiv \bar{I}_i$ if $k \in \mathcal{K}_i$ and $i > 0$. For $i = 1, \dots, l$, let $\bar{R}_i$ be the $n \times |\bar{I}_i|$ matrix whose columns consist of $\nabla g_j(x^*)$ for $j \in \bar{I}_i$. Note that by (H7), $\bar{R}_i^T \bar{R}_i$ is also invertible. By the equality part of the KKT conditions for (QP_1),

$$u_{I_k}^k = -(R_k^T R_k)^{-1} R_k^T (H_k d_0^k + \nabla f(x^k)).$$

Then, as $k \to \infty$ for $k \in \mathcal{K}_i$,

$$u^k_{I_k} \to u^*_{\bar{I}_i} = -(\bar{R}_i^T \bar{R}_i)^{-1} \bar{R}_i^T \nabla f(x^*).$$

Let $u^*_i = 0$ if $i \notin \bar{I}_i$. Then $u^* \in M(x^*)$. We see that 0 is a KKT point of

$$(\bar{QP}_i)\begin{cases} \min \ \frac{1}{2} d^T d + \nabla f(x^*)^T d \\ \text{s.t. } \ g_j(x^*) + \nabla g_j(x^*)^T d = 0, \quad j \in \bar{I}_i, \\ \qquad g_j(x^*) + \nabla g_j(x^*)^T d \le 0, \quad j \in I \setminus \bar{I}_i, \end{cases}$$

with a Lagrange multiplier u^*. Because of (H7), vectors in $\{\nabla g_j(x^*) | j \in \bar{I}_i\}$ are linearly independent. Then (H8) implies that the MFCQ holds at x^* for $(\bar{QP}_i)$. It is easy to see that the SSOSC holds at x^* for $(\bar{QP}_i)$ too. Applying the Kojima theorem (Theorem 3.5), we see that

$$(QP_4)\begin{cases} \min \ \frac{1}{2} d^T d + \nabla f(x^k)^T d \\ \text{s.t. } \ g_j(x^k + d^k) + \nabla g_j(x^k)^T d = -\|d^k\|^{\tau_3}, \quad j \in I_k, \\ \qquad g_j(x^k + d^k) + \nabla g_j(x^k)^T d \le -\|d^k\|^{\tau_3}, \quad j \in I \setminus I_k, \end{cases}$$

is feasible for k large enough, since (QP_4) is a perturbed form of $(\bar{QP}_i)$. Since (QP_3) has the same constraints as (QP_4), (QP_3) is also feasible for k large enough. Hence, and because of (H5) and (H7), and because of the fact that if the CRCQ holds at a point then it holds at a neighborhood of that point, (QP_3) has a KKT point $\tilde{d}^k$ for k large enough. Now, we may follow the proof of Proposition 3.6 of [18] step by step with minor modification for each i satisfying $1 \le i \le l$. Note that l is finite. The conclusion follows. ∎

Finally, two-step superlinear convergence follows. As in [18], the proof is not given as it follows step by step, with minor modifications, that of Lemma 3 to Theorem 1 in [23]. Note that with (H9), (H8), and (H7), we do not need to invoke Lemmas 1 and 2 in [23], which rely on the LICQ and the strict complementarity slackness.

THEOREM 6.2. *Under the stated assumptions, the convergence is two-step superlinear, i.e.,*

$$\lim_{k\to\infty} \frac{\|x^{k+2} - x^*\|}{\|x^k - x^*\|} = 0.$$

Remark. Similarly, as we mentioned in section 4, the conditions of Powell's theorem on the SQP method, Theorem 1 of [23], may be reduced to the SSOSC and the CPLD, by replacing (3.10) in [23] with (H9).

Again, the result in this section can also be extended to the 1987 Panier–Tits algorithm.

Acknowledgments. The authors are grateful to Professors Olvi Mangasarian, Jorge Nocedal, Jong-Shi Pang, Elijah Polak, and Stephen Robinson and two referees for their helpful comments.

REFERENCES

[1] J. F. BONNANS, *Local study of Newton type algorithms for constrained problems*, in Optimization—Fifth French-German Conference, S. Dolecki, ed., Springer-Verlag, Berlin, 1989, pp. 13–24.

[2] X. Chen, *Convergence of the BFGS methods for LC^1 convex constrained optimization*, SIAM J. Control Optim., 34 (1996), pp. 2051–2063.

[3] X. Chen, L. Qi, and R. Womersley, *Newton's method for quadratic stochastic programs with recourse*, J. Comput. Appl. Math., 60 (1995), pp. 29–46.

[4] F. Facchinei, *Minimization of SC^1 functions and the Maratos effect*, Oper. Res. Lett., 17 (1995), pp. 131–137.

[5] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[6] A. Fischer, *Modified Wilson Method for Nonlinear Programs with Nonunique Multipliers*, preprint, MATH-NM-04-1997, Institute for Numerical Mathematics, Technical University of Dresden, Dresden, Germany, February 1997.

[7] U. C. Garcia Palomares and O. L. Mangasarian, *Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems*, Math. Programming, 11 (1976), pp. 1–13.

[8] M. S. Gowda and J. S. Pang, *Stability analysis of variational inequalities and nonlinear complementarity problems, via the mixed linear complementarity problems and degree theory*, Math. Oper. Res., 19 (1994), pp. 831–879.

[9] J. Han and D. Sun, *Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity*, in Recent Advances in Nonsmooth Optimization, D. Du, L. Qi, and R. S. Womersley, eds., World Scientific, River Edge, NJ, 1995, pp. 141–158.

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

[11] R. Janin, *Direction derivative of the marginal function in nonlinear programming*, Math. Programming Study, 21 (1984), pp. 127–138.

[12] H. Jiang and L. Qi, *Globally and superlinearly convergent trust region algorithm for convex SC^1 minimization problems and its application to stochastic programs*, J. Optim. Theory Appl., 90 (1996), pp. 653–673.

[13] H. Jiang, L. Qi, X. Chen, and D. Sun, *Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations*, in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi, eds., Plenum, New York, 1996, pp. 197–212.

[14] M. Kojima, *Strongly stable stationary solutions in nonlinear programming*, in Analysis and Computation of Fixed Points, S. M. Robinson, ed., Academic Press, New York, 1980, pp. 93–138.

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

[16] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[17] E. R. Panier and A. L. Tits, *A superlinearly convergent feasible method for the solution of inequality constrained optimization problems*, SIAM J. Control Optim., 25 (1987), pp. 934–950.

[18] E. R. Panier and A. L. Tits, *On combining feasibility, descent and superlinear convergence in inequality constrained optimization*, Math. Programming, 59 (1993), pp. 261–276.

[19] J. S. Pang and L. Qi, *A globally convergent Newton method for convex SC^1 minimization problems*, J. Optim. Theory Appl., 85 (1995), pp. 633–648.

[20] J. S. Pang and D. Ralph, *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*, Math. Oper. Res., 21 (1996), pp. 401–426.

[21] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, New York, 1997.

[22] E. Polak and L. Qi, *A globally and superlinearly convergent scheme for minimizing a normal merit function*, SIAM J. Control Optim., 36 (1998), pp. 1005–1019.

[23] M. J. D. Powell, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in Nonlinear Programming 3, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, 1978, pp. 27–63.

[24] L. Qi, *Superlinearly convergent approximate Newton methods for LC^1 optimization problems*, Math. Programming, 64 (1994), pp. 277–294.

[25] L. Qi, *LC^1 functions and LC^1 optimization*, in Operations Research and Its Application, D. Z. Du, X. S. Zhang, and K. Chen, eds., World Publishing, Beijing, 1996, pp. 4–13.

[26] L. Qi and H. Jiang, *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and quasi-Newton methods for solving these equations*, Math. Oper. Res., 22 (1997), pp. 301–325.

[27] L. Qi and R. Womersley, *An SQP Algorithm for extended linear-quadratic problems in*

stochastic programming, Ann. Oper. Res., 56 (1995), pp. 251–285.

[28] D. Ralph and S. J. Wright, *Superlinear convergence of an interior-point method for monotone variational inequalities*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 345–385.

[29] S. M. Robinson, *First order conditions for general nonlinear optimization*, SIAM J. Appl. Math., 30 (1976), pp. 597–607.

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

[31] S. M. Robinson, *Generalized equations and their solutions, part* II*: Applications to nonlinear programming*, Math. Programming Study, 19 (1982), pp. 200–221.

[32] S. M. Robinson and R. R. Meyer, *Lower semicontinuity of multivalued linearization mappings*, SIAM J. Control, 11 (1973), pp. 525–533.

[33] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions.* I, Springer-Verlag, New York, 1970.

[34] D. Sun, M. Fukushima and L. Qi, *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 452–473.

[35] S. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275.