SHARP PRIMAL SUPERLINEAR CONVERGENCE RESULTS FOR SOME NEWTONIAN METHODS FOR CONSTRAINED OPTIMIZATION[*]

D. FERNÁNDEZ[†], A. F. IZMAILOV[‡], AND M. V. SOLODOV[§]

Abstract. As is well known, Q-superlinear or Q-quadratic convergence of the primal-dual sequence generated by an optimization algorithm does not, in general, imply Q-superlinear convergence of the primal part. Primal convergence, however, is often of particular interest. For the sequential quadratic programming (SQP) algorithm, local primal-dual quadratic convergence can be established under the assumptions of uniqueness of the Lagrange multiplier associated to the solution and the second-order sufficient condition. At the same time, previous primal Q-superlinear convergence results for SQP required strengthening of the first assumption to the linear independence constraint qualification. In this paper, we show that this strengthening of assumptions is actually not necessary. Specifically, we show that once primal-dual convergence is assumed or already established, for primal superlinear rate one needs only a certain error bound estimate. This error bound holds, for example, under the second-order sufficient condition, which is needed for primal-dual local analysis in any case. Moreover, in some situations even second-order sufficiency can be relaxed to the weaker assumption that the multiplier in question is noncritical. Our study is performed for a rather general perturbed SQP framework which covers, in addition to SQP and quasi-Newton SQP, some other algorithms as well. For example, as a byproduct, we obtain primal Q-superlinear convergence results for the linearly constrained (augmented) Lagrangian methods for which no primal Q-superlinear rate of convergence results were previously available. Another application of the general framework is sequential quadratically constrained quadratic programming methods. Finally, we discuss some difficulties with proving primal superlinear convergence for the stabilized version of SQP.

Key words. Newton methods, sequential quadratic programming, linearly constrained Lagrangian methods, superlinear convergence, critical multipliers

AMS subject classifications. 90C30, 90C33, 90C55, 65K05

DOI. 10.1137/090776664

1. Introduction. Consider the mathematical programming problem

$$
\text{(1.1)} \qquad \begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && h(x) = 0, \ g(x) \le 0, \end{aligned}
$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function and $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings. Specifically, we assume that f, h, and g are twice differentiable near the point of interest $\bar{x} \in \mathbf{R}^n$, and their second derivatives are continuous at $\bar{x}$. Stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$
\text{(1.2)} \qquad \frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \ge 0, \quad g(x) \le 0, \quad \langle \mu, g(x) \rangle = 0,
$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (1.1), i.e.,

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle.$$

We shall consider the class of algorithms for solving (1.1) (or (1.2)) described by a *perturbed sequential quadratic programming* (pSQP) framework [24], to some extent related to inexact sequential quadratic programming in [37]; see also [21]. To this end, recall that given a primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the SQP subproblem [4] has the form

(1.3)   minimize $f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle$
          subject to $h(x^k) + h'(x^k)(x - x^k) = 0, \; g(x^k) + g'(x^k)(x - x^k) \le 0,$

where H_k is a symmetric $n \times n$ matrix. The choice

$$(1.4) \qquad\qquad H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$$

corresponds to the basic form of SQP. By pSQP, we shall generally mean the class of methods that generate iterates satisfying a certain perturbed version of KKT conditions for (1.3)–(1.4). Different forms of perturbations give rise to different algorithms, and some of the algorithms in question are not modifications of SQP itself. Specifically, for the given primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ in the pSQP framework must satisfy the following relations in the variables $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$:

$$\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k)$$
$$+ (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) + \omega_1^k = 0,$$
$$h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0,$$
$$\mu \ge 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \le 0, \quad \langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \rangle = 0,$$

(1.5)

where $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, and $\omega_3^k \in \mathbf{R}^m$ are perturbation terms defining specific algorithms. In particular, when $\omega_1^k = 0$, $\omega_2^k = 0$, and $\omega_3^k = 0$, then (1.5) is precisely the KKT system of the basic SQP subproblem (1.3)–(1.4). In addition, for certain specific forms of perturbations, pSQP includes also quasi-Newton versions of SQP (e.g., [31, 5]), linearly constrained (augmented) Lagrangian (LCL) methods [32, 27, 14, 22], sequential quadratically constrained quadratic programming [2, 15, 34, 11], and stabilized SQP [36, 16, 13, 38, 12]. It is worth emphasizing once again that for some forms of perturbations, the pSQP framework includes algorithms which may not be modifications of SQP per se, in the sense that subproblems of those algorithms are not even quadratic. The point is that iterates of all the methods in the considered class can be related to a perturbation of SQP given by (1.5) a posteriori. Specific relations would be made clear in the applications section 4 below.

Before surveying the available primal-dual and primal convergence results for the methods in question, we need some notation. By $\mathcal{M}(\bar{x})$ we denote the set of Lagrange multipliers associated with $\bar{x} \in \mathbf{R}^n$, that is,

$$\mathcal{M}(\bar{x}) = \{(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m \mid (\lambda, \mu) \text{ satisfies } (1.2) \text{ for } x = \bar{x}\}.$$

Thus $\bar{x}$ is a stationary point of problem (1.1) if $\mathcal{M}(\bar{x}) \ne \emptyset$. Let

$$A = A(\bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{x}) = 0\}$$

be the set of indices of active constraints at a stationary point $\bar{x}$ of problem (1.1),

$$N = N(\bar{x}) = \{1, \ldots, m\} \setminus A$$

be the set of indices of inactive constraints, and

$$A_+ = A_+(\bar{x}, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu}_i > 0\}, \quad A_0 = A_0(\bar{x}, \bar{\mu}) = A \setminus A_+$$

be the sets of indices of strongly and weakly active constraints, respectively. The linear independence constraint qualification (LICQ) at $\bar{x}$ consists of saying that the gradients of equality constraints together with the gradients of inequality constraints active at $\bar{x}$ form a linearly independent set in $\mathbf{R}^n$. Obviously, LICQ implies that the multiplier set $\mathcal{M}(\bar{x})$ is a singleton. The strict Mangasarian–Fromovitz constraint qualification (SMFCQ) consists of saying that the multiplier associated to $\bar{x}$ is unique. Thus SMFCQ is a weaker assumption than LICQ.

The critical cone of problem (1.1) at its stationary point $\bar{x}$ is given by

$$
\begin{aligned}
C &= C(\bar{x}) \\
&= \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \ g'_A(\bar{x})\xi \leq 0, \ \langle f'(\bar{x}), \xi \rangle \leq 0\} \\
&= \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \ g'_{A_+}(\bar{x})\xi = 0, \ g'_{A_0}(\bar{x})\xi \leq 0\}.
\end{aligned}
$$

We say that the second-order sufficient condition (SOSC) is satisfied at $\bar{x}$ with $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ if

$$(1.6) \qquad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C \setminus \{0\}.$$

The weakest assumptions under which Q-superlinear/quadratic primal-dual convergence of basic SQP (1.3)–(1.4) had been established in the literature are SMFCQ and SOSC [6]. Other results require, in addition to SOSC, the stronger LICQ (e.g., [7, Theorem 15.4]) or even strict complementarity (e.g., [33], [7, Theorem 15.2]). Recall that Q-superlinear/quadratic convergence of the primal-dual sequence does not automatically guarantee any Q-rate of convergence of the primal part of the sequence (e.g., [7, Exercise 14.8]). The issue of primal rate of convergence, assuming (or having established) primal-dual convergence, is thus studied separately, often in combination with quasi-Newton considerations. For SQP iterates given by (1.3), the following is known [7, Theorem 15.7] (see also [5, 6], [30, Theorem 18.5] for related statements). Assuming that the primal-dual sequence $\{(x^k, \lambda^k, \mu^k)\}$ converges and that LICQ and SOSC hold at the limit $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the rate of convergence of the primal sequence $\{x^k\}$ is superlinear if and only if the following Dennis–Moré-type condition holds:

$$(1.7) \qquad \pi_C\left(\left(\frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) - H_k\right)(x^{k+1} - x^k)\right) = o(\|x^{k+1} - x^k\|),$$

where C is the critical cone defined above, and by $\pi_S(x)$ we denote the Euclidean projection of $x \in \mathbf{R}^n$ onto the closed convex set $S \subset \mathbf{R}^n$. For the basic choice (1.4) of H_k, in case of convergence of the primal-dual sequence the above condition (1.7) is satisfied automatically, of course. However, compared to what is needed for superlinear primal-dual convergence, note that superlinear primal convergence required replacing SMFCQ by the stronger LICQ. In this paper, we shall obtain sharper results. Specifically, we shall show that under SMFCQ and SOSC needed

for superlinear primal-dual convergence of SQP, superlinear primal convergence is automatic. Moreover, if primal-dual convergence is assumed, as in typical analyses under the Dennis–Moré-type conditions, then the only further assumption needed for superlinear primal convergence is a certain error bound. In particular, no constraint qualifications are needed at all. The needed error bound is equivalent to saying that the relevant multiplier is noncritical (see Definition 2.1 below), which is implied by SOSC, for example.

In the case of LCL methods [32, 27, 14], the strongest local superlinear primal-dual convergence result again assumes SMFCQ and SOSC; see [22] (other cited literature on LCL methods requires LICQ instead of SMFCQ and, in addition, strict complementarity). To the best of our knowledge, no Q-superlinear primal rate of convergence results for LCL methods are available in the literature. We shall obtain the first results of this kind. Those results are also sharp, in the sense that no strengthening of assumptions are required with respect to primal-dual convergence.

Finally, primal rate of convergence of sequential quadratically constrained quadratic programming methods and of the stabilized version of SQP will be discussed.

We next describe our notation. For $y, z \in \mathbf{R}^m$, we use the notation $\min\{y, z\}$ for the componentwise minimum and $|z|$ for the componentwise absolute value. We shall also use the notation $|I|$ for cardinality of a finite set I, but this is always clear from the context and does not induce confusion. By z_I we denote the subvector of z with components z_i, $i \in I$. The image (range space) of a linear operator Λ is denoted by $\operatorname{im} \Lambda$, and its kernel (null space) is denoted by $\ker \Lambda$.

The polar (negative dual) cone to a convex cone K is

$$K^\circ = \{x \in \mathbf{R}^n \mid \langle x, \xi \rangle \leq 0 \ \forall \xi \in K\}.$$

As can be easily seen, it holds that

$$(1.8) \qquad \{x \in \mathbf{R}^n \mid \pi_K(x) = 0\} = K^\circ$$

and

$$(1.9) \qquad \pi_K(x - \pi_K(x)) = 0 \quad \forall x \in \mathbf{R}^n.$$

When S is a linear subspace in $\mathbf{R}^n$, $S^\perp$ denotes its orthogonal complement in $\mathbf{R}^n$.

In what follows, we shall also employ the linear subspace

$$C_+ = C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \ g'_{A_+}(\bar{x})\xi = 0\},$$

which contains the critical cone C.

2. Error bound for primal solution. As is typical in rate of convergence analyses, we shall need some estimates on the distance to the solution. For primal rate of convergence, we need estimates on the distance to the primal solution. Furthermore, the form of the Dennis–Moré condition (1.7) that arises in the context of SQP indicates that the estimates should involve the projection onto the critical cone C (or, perhaps, onto the subspace C_+).

Recall first that under SOSC (1.6), it follows from [17, Lemma 2] and [13, Theorem 2] that the estimate

$$(2.1) \qquad \|x - \bar{x}\| + \operatorname{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) = O\left(\left\|\begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix}\right\|\right)$$

holds for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. This estimate, however, cannot be used for our purposes, as it does not involve projections (and it seems that estimates with projections cannot be derived from it).

We start with some considerations that motivate the kind of primal error bounds that would be derived in this section. By the definition of $\mathcal{M}(\bar{x})$ (which is a polyhedral set) and by Hoffman's Lemma (see, e.g., [8, Theorem 2.200]), there exists $c > 0$ such that the following estimate is valid for all $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$:

$$
(2.2) \qquad \operatorname{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) \leq c \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(\bar{x}, \lambda, \mu) \\ \min\{0, \mu_A\} \\ \mu_N \end{pmatrix} \right\|.
$$

For each $\mu \in \mathbf{R}^m$, define

$$
(2.3) \qquad \tilde{\mu}_i = \begin{cases} \max\{0, \mu_i\} & \text{if } i \in A, \\ 0, & \text{if } i \in N, \end{cases}
$$

i.e., $\tilde{\mu}$ is the Euclidian projection of μ onto the set $\{\mu \in \mathbf{R}^m \mid \mu_A \geq 0, \ \mu_N = 0\}$. Then for any $\xi \in C$ it evidently holds that

$$
\left\langle \frac{\partial L}{\partial x}(\bar{x}, \lambda, \tilde{\mu}), \xi \right\rangle = \langle f'(\bar{x}), \xi \rangle + \langle \lambda, h'(\bar{x})\xi \rangle + \langle \tilde{\mu}_A, g'_A(\bar{x})\xi \rangle \leq 0,
$$

which means that $\frac{\partial L}{\partial x}(\bar{x}, \lambda, \tilde{\mu}) \in C^\circ$. Therefore, for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, it holds that

$$
\begin{aligned}
\frac{\partial L}{\partial x}(\bar{x}, \lambda, \mu) &= \frac{\partial L}{\partial x}(\bar{x}, \lambda, \tilde{\mu}) + O(\|\mu - \tilde{\mu}\|) \\
&= \pi_{C^\circ}\left(\frac{\partial L}{\partial x}(\bar{x}, \lambda, \tilde{\mu}) \right) + O(\|\mu - \tilde{\mu}\|) \\
&= \pi_{C^\circ}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu) \right) + O(\|x - \bar{x}\|) + O(\|\mu - \tilde{\mu}\|) \\
(2.4) \qquad &= \pi_{C^\circ}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu) \right) + O(\|x - \bar{x}\|) + O(\|\min\{0, \mu_A\}\| + \|\mu_N\|),
\end{aligned}
$$

where the last equality is by (2.3). Note that if for some $i \in A$ it holds that $\mu_i < 0$, then

$$
|\min\{0, \mu_i\}| = -\mu_i \leq \max\{-\mu_i, g_i(x)\} = -\min\{\mu_i, -g_i(x)\},
$$

and hence

$$
\|\min\{0, \mu_A\}\| \leq \|\min\{\mu_A, -g_A(x)\}\|.
$$

On the other hand, since $g_N(\bar{x}) < 0$ and $\bar{\mu}_N = 0$, it holds that

$$
\mu_N = \min\{\mu_N, -g_N(x)\}
$$

for all (x, μ) close enough to $(\bar{x}, \bar{\mu})$. Combining the last two relations with (2.2) and (2.4), we derive the estimate

$$
\operatorname{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) = O\left(\left\| \begin{pmatrix} \pi_{C^\circ}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu) \right) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\| \right) + O(\|x - \bar{x}\|).
$$

Comparing this with (2.1) and taking into account that $x = \pi_{C^\circ}(x) + \pi_C(x)$ for all $x \in \mathbf{R}^n$, it is natural to conjecture that in the estimate for $\|x - \bar{x}\|$ one might replace $\frac{\partial L}{\partial x}(x, \lambda, \mu)$ by the term $\pi_C(\frac{\partial L}{\partial x}(x, \lambda, \mu))$ of generally smaller norm. And this is indeed the case, as will be demonstrated in Theorem 2.3 below.

It is important to point out that the error bound (2.1) turns out to be equivalent to saying that the multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ associated to the stationary point $\bar{x}$ is noncritical in the sense of [23] (the latter is a weaker assumption than SOSC (1.6)). Specifically, the following notion was introduced in [23].

DEFINITION 2.1. *A multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *is called* critical *if there exists a triple* $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, *with* $\xi \neq 0$, *satisfying the system*

$$(2.5) \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0,$$

$$(2.6) \quad \zeta_{A_0} \geq 0, \quad g'_{A_0}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}), \xi \rangle = 0, \, i \in A_0, \quad \zeta_N = 0,$$

and noncritical *otherwise.*

The above is a natural extension of the definition for a problem with equality constraints only, which states that a multiplier $\bar{\lambda}$ is critical if

$$\exists \xi \in \ker h'(\bar{x}) \setminus \{0\} \text{ such that } \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}},$$

and noncritical otherwise; see [19, 20, 21, 23] where special properties of critical/ noncritical multipliers have been investigated.

It can be easily observed that (2.5)–(2.6) is the KKT system for the quadratic programming problem

$$(2.7) \quad \begin{array}{c} \text{minimize} \quad \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \\[2mm] \text{subject to} \quad \xi \in C. \end{array}$$

Therefore, $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ being noncritical is equivalent to saying that $\xi = 0$ is the unique stationary point of problem (2.7). Furthermore, it then follows that $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ being noncritical is also equivalent to saying that the Hessian $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})$ and the critical cone C are an R_0-pair in the sense of [9].

By multiplying the first equality in (2.5) by ξ and using the definition of C and the other relations in (2.5)–(2.6), we see that for any critical multiplier $(\bar{\lambda}, \bar{\mu})$ it must hold that

$$\exists \xi \in C \setminus \{0\} \text{ such that } \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle = 0.$$

Therefore, multipliers that satisfy SOSC (1.6) cannot be critical. Likewise, multipliers that satisfy the "symmetric" condition

$$(2.8) \quad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle < 0 \quad \forall \xi \in C \setminus \{0\}$$

also cannot be critical. There may, however, exist noncritical multipliers that satisfy neither (1.6) nor (2.8). But if SMFCQ holds at $\bar{x}$ for $(\bar{\lambda}, \bar{\mu})$ and $\bar{x}$ is a local solution of problem (1.1), then this multiplier is noncritical if and only if SOSC (1.6) holds (note that the multiplier in question in this particular case is actually unique); see [23].

We shall next derive two different estimates for the distance to the primal solution involving projections. The first one assumes the weaker property of noncriticality of the multiplier but makes projection onto the larger cone C_+, while the second assumes the stronger SOSC (1.6) but makes projection onto the smaller cone C. We also provide an example showing that one cannot use C instead of C_+ under the weaker assumption of noncriticality.

THEOREM 2.2. *Let $\bar{x}$ be a stationary point of problem (1.1), and let $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated noncritical Lagrange multiplier.*

Then the estimate

$$(2.9) \qquad \|x - \bar{x}\| = O\left(\left\|\left(\begin{array}{c} \pi_{C_+}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\} \end{array}\right)\right\|\right)$$

holds for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Proof. We argue by contradiction. Suppose that (2.9) does not hold. Then there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, with $x^k \neq \bar{x}$ for all k, and such that

$$\|x^k - \bar{x}\|\left\|\left(\begin{array}{c} \pi_{C_+}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) \\ h(x^k) \\ \min\{\mu^k, -g(x^k)\} \end{array}\right)\right\|^{-1} \to \infty \text{ as } k \to \infty,$$

which is equivalent to saying that

$$(2.10) \qquad \pi_{C_+}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|),$$

$$(2.11) \qquad h(x^k) = o(\|x^k - \bar{x}\|),$$

$$(2.12) \qquad \min\{\mu^k, -g(x^k)\} = o(\|x^k - \bar{x}\|).$$

By (2.11), using the Taylor formula, we have that

$$0 = h(x^k) + o(\|x^k - \bar{x}\|)$$
$$= h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|)$$
$$(2.13) \qquad = h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|).$$

Moreover, since $g_{A_+}(\bar{x}) = 0 < \bar{\mu}_{A_+}$, for all k large enough, from (2.12) we derive that

$$0 = \min\{\mu^k_{A_+}, -g_{A_+}(x^k)\} + o(\|x^k - \bar{x}\|)$$
$$= -g_{A_+}(x^k) + o(\|x^k - \bar{x}\|)$$
$$= -g_{A_+}(\bar{x}) - g'_{A_+}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|)$$
$$(2.14) \qquad = -g'_{A_+}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|),$$

and similarly since $g_N(\bar{x}) < 0 = \bar{\mu}_N$,

$$0 = \min\{\mu^k_N, -g_N(x^k)\} + o(\|x^k - \bar{x}\|)$$
$$(2.15) \qquad = \mu^k_N + o(\|x^k - \bar{x}\|).$$

Since the number of different partitions of the set A_0 is finite, passing onto a subsequence if necessary, we can assume that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0$, $I_1 \cap I_2 = \emptyset$, and for each k it holds that

$$(2.16) \qquad \mu^k_{I_1} \geq -g_{I_1}(x^k), \quad \mu^k_{I_2} < -g_{I_2}(x^k).$$

Similar to the above, from (2.12) we then obtain that

$$
\begin{aligned}
0 &= \min\{\mu_{I_1}^k, -g_{I_1}(x^k)\} + o(\|x^k - \bar{x}\|) \\
&= -g_{I_1}(x^k) + o(\|x^k - \bar{x}\|) \\
&= -g_{I_1}(\bar{x}) - g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\
&= -g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

(2.17)

$$
\begin{aligned}
0 &= \min\{\mu_{I_2}^k, -g_{I_2}(x^k)\} + o(\|x^k - \bar{x}\|) \\
&= \mu_{I_2}^k + o(\|x^k - \bar{x}\|).
\end{aligned}
$$

(2.18)

Moreover, by (2.16) and the second equality in (2.17) we obtain that

$$
\begin{aligned}
\mu_{I_1}^k &\geq -g_{I_1}(x^k) \\
&= o(\|x^k - \bar{x}\|).
\end{aligned}
$$

(2.19)

Finally, from (2.16) it also follows that

$$
\begin{aligned}
-\mu_{I_2}^k &> g_{I_2}(x^k) \\
&= g_{I_2}(\bar{x}) + g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\
&= g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

(2.20)

and hence, by (2.18),

(2.21)
$$
g'_{I_2}(\bar{x})(x^k - \bar{x}) \leq o(\|x^k - \bar{x}\|).
$$

Furthermore, employing (1.9) and (2.10), we derive that

$$
\begin{aligned}
0 &= \pi_{C_+}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \pi_{C_+}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right)\right) \\
&= \pi_{C_+}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|)\right).
\end{aligned}
$$

The latter, by (1.8), implies that

$$
\begin{aligned}
C_+^\perp = C_+^\circ \\
&\ni \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|) \\
&= \frac{\partial L}{\partial x}(x^k, \bar{\lambda}, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|) \\
&= \frac{\partial L}{\partial x}(x^k, \bar{\lambda}, \bar{\mu}) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\
&\quad + o(\|x^k - \bar{x}\|) \\
&= \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

(2.22)

where the first equality is by the fact that C_+ is a linear subspace, the third is by $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, and the last is by the Taylor formula. Since

$$
C_+^\perp = (\ker h'(\bar{x}) \cap \ker g'_{A_+}(\bar{x}))^\perp = \operatorname{im}(h'(\bar{x}))^{\mathrm{T}} + \operatorname{im}(g'_{A_+}(\bar{x}))^{\mathrm{T}},
$$

taking into account that $\bar\mu_{A_0 \cup N} = 0$, (2.22) implies the inclusion

$$-\operatorname{im}(h'(\bar x))^{\mathrm T} - \operatorname{im}(g'_{A_+}(\bar x))^{\mathrm T} \ni \frac{\partial^2 L}{\partial x^2}(\bar x,\,\bar\lambda,\,\bar\mu)(x^k - \bar x) + (g'_{A_0 \cup N}(\bar x))^{\mathrm T}\mu^k_{A_0 \cup N} + o(\|x^k - \bar x\|).$$
(2.23)

Without loss of generality, again passing onto a subsequence if necessary, we can assume that there exist index sets J_1 and J_2 such that $J_1 \cup J_2 = I_1$, $J_1 \cap J_2 = \emptyset$, and for each k it holds that

$$\mu^k_{J_1} \geq 0, \quad \mu^k_{J_2} < 0.$$

Then from (2.23) and (2.19) we derive the estimates

$$-\operatorname{im}(h'(\bar x))^{\mathrm T} - \operatorname{im}(g'_{A_+}(\bar x))^{\mathrm T} - (g'_{J_1}(\bar x))^{\mathrm T}\mathbf{R}^{|J_1|}_+ \ni \frac{\partial^2 L}{\partial x^2}(\bar x,\,\bar\lambda,\,\bar\mu)(x^k - \bar x)$$
$$+ (g'_{J_2 \cup I_2 \cup N}(\bar x))^{\mathrm T}\mu^k_{J_2 \cup I_2 \cup N}$$
$$+ o(\|x^k - \bar x\|)$$
(2.24)

and

(2.25) $$\mu^k_{J_2} = o(\|x^k - \bar x\|).$$

Note that the left-hand side of (2.24) is closed (as the sum of linear subspaces and a polyhedral cone).

Again passing onto a subsequence, if necessary, we can assume that $\{(x^k - \bar x)/\|x^k - \bar x\|\}$ converges to some $\xi \in \mathbf{R}^n$ ($\|\xi\| = 1$). Dividing (2.24) and (2.13), (2.14), (2.17), (2.21) by $\|x^k - \bar x\|$, passing onto the limit as $k \to \infty$, and employing the relations (2.15), (2.18), and (2.25), we obtain the following:

(2.26) $$\frac{\partial^2 L}{\partial x^2}(\bar x,\,\bar\lambda,\,\bar\mu)\xi \in -\operatorname{im}(h'(\bar x))^{\mathrm T} - \operatorname{im}(g'_{A_+}(\bar x))^{\mathrm T} - (g'_{J_1}(\bar x))^{\mathrm T}\mathbf{R}^{|J_1|}_+,$$

(2.27) $$h'(\bar x)\xi = 0, \quad g'_{A_+}(\bar x)\xi = 0,$$

(2.28) $$g'_{I_1}(\bar x)\xi = 0, \quad g'_{I_2}(\bar x)\xi \leq 0.$$

Inclusion (2.26) means that there exists $(\eta,\,\zeta) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying the first equality in (2.5) and such that

$$\zeta_{J_1} \geq 0, \quad \zeta_{J_2 \cup I_2 \cup N} = 0.$$

Combining this with (2.27), (2.28), we obtain that $(\xi,\,\eta,\,\zeta)$ satisfies (2.5), (2.6), which contradicts the assumption that $(\bar\lambda,\,\bar\mu)$ is noncritical because $\xi \neq 0$. $\qquad\square$

We now provide an example demonstrating that, under the noncriticality assumption, in the estimate (2.9) one cannot replace C_+ by the (generally smaller) cone C, even if $\bar x$ satisfies LICQ.

Example 2.1. Let $n = m = 1$, $l = 0$, $f(x) = -x^2/2$, $g(x) = x$. Then $\bar x = 0$ is a stationary point of problem (1.1) satisfying LICQ, and $\bar\mu = 0$ is the unique associated multiplier. Note that $C_+ = \mathbf{R}$, $C = \{\xi \in \mathbf{R} \mid \xi \leq 0\}$, and the multiplier is evidently noncritical ((2.8) holds).

Take any $x < 0$ and $\mu = 0$. Then

$$\frac{\partial L}{\partial x}(x,\,\mu) = -x + \mu = -x > 0,$$

and the projection of this element onto C is zero. Also,

$$\min\{\mu, -x\} = \mu = 0.$$

Hence, if C_+ were to be replaced by C in (2.9), the "residual" in the right-hand side would be equal to zero at points in question, failing to provide the needed estimate. At the same time, (2.9) with C_+ is evidently valid.

We next show that if noncriticality of the multiplier is replaced by the stronger SOSC (1.6), then C_+ in (2.9) can be replaced by C.

THEOREM 2.3. *Let $\bar{x}$ be a stationary point of problem* (1.1)*, and let* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *be an associated Lagrange multiplier satisfying SOSC* (1.6)*.*

Then the estimate

$$(2.29) \qquad \|x - \bar{x}\| = O\left(\left\| \begin{pmatrix} \pi_C\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\| \right)$$

holds for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Proof. The argument is along the lines of the proof of Theorem 2.2. Assuming that (2.29) does not hold, we obtain that there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and satisfying

$$(2.30) \qquad \pi_C\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|)$$

and (2.11), (2.12). Passing onto a subsequence, if necessary, we can assume that the sequence $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some $\xi \in \mathbf{R}^n$ ($\|\xi\| = 1$) and that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0$, $I_1 \cap I_2 = \emptyset$, and that relations (2.15), (2.18), (2.27), (2.28) are satisfied.

Using the same argument as the one leading to (2.22), employing (1.8) and (1.9) we derive from (2.30) that

$$C^{\circ} \ni \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|)$$

$$= \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|).$$

(2.31)

Note that, according to (2.27), (2.28), it holds that $\xi \in C$. Therefore, taking into account (2.15), (2.18), (2.27), the first relation in (2.28), and the equality $\bar{\mu}_{A_0 \cup N} = 0$, from (2.31) we derive that

$$0 \geq \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}), \xi \right\rangle + \langle \lambda^k - \bar{\lambda}, h'(\bar{x})\xi \rangle + \langle \mu^k - \bar{\mu}, g'(\bar{x})\xi \rangle + o(\|x^k - \bar{x}\|)$$

$$= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}), \xi \right\rangle + \langle \mu^k_{I_2 \cup N}, g'_{I_2 \cup N}(\bar{x})\xi \rangle + o(\|x^k - \bar{x}\|)$$

$$= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}), \xi \right\rangle + o(\|x^k - \bar{x}\|).$$

Dividing the latter relation by $\|x^k - \bar{x}\|$ and passing onto the limit, we conclude that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \leq 0,$$

which contradicts SOSC (1.6) because $\xi \in C \setminus \{0\}$. $\square$

Note that unlike the primal-dual error bound (2.1), which is valid according to [17, Lemma 2] and [13, Theorem 2], estimates (2.9) and (2.29) measure the distance to the primal solution $\bar{x}$ only. The distance to the multiplier set $\mathcal{M}(\bar{x})$ is not estimated. This is the price paid for deriving sharper error bounds for the primal part, with generally smaller right-hand sides. Note also that unlike (2.1), the right-hand sides in the estimates (2.9) and (2.29) are not computable, as they involve C_+ and C, which are defined at the unknown solution. Thus the purpose for deriving our new estimates is for convergence analysis only, with no intended use within algorithms themselves.

Remark 2.1. Under the "symmetric" second-order condition (2.8), we can obtain the following estimate:

$$(2.32) \qquad \|x - \bar{x}\| = O\left(\left\| \begin{pmatrix} \pi_C\left(-\frac{\partial L}{\partial x}(x,\ \lambda,\ \mu)\right) \\ h(x) \\ \min\{\mu,\ -g(x)\} \end{pmatrix} \right\| \right).$$

The proof is analogous to that of Theorem 2.3, noting that for this case in (2.31) we have that

$$C^\circ \ni -\frac{\partial L}{\partial x}(x^k,\ \lambda^k,\ \mu^k) + o(\|x^k - \bar{x}\|).$$

3. Primal superlinear convergence in pSQP framework.
We start with necessary conditions for superlinear primal convergence of pSQP iterates given by (1.5). After that we establish sufficient conditions for primal superlinear rate and apply those results to specific algorithms that fall within our general pSQP framework.

PROPOSITION 3.1. *Let $\bar{x}$ be a stationary point of problem (1.1), and let $(\bar{\lambda},\ \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated Lagrange multiplier. Let the sequence $\{(x^k,\ \lambda^k,\ \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be convergent to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, and assume that for each k the triple $(x^{k+1},\ \lambda^{k+1},\ \mu^{k+1})$ satisfies the system (1.5) with some $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, and $\omega_3^k \in \mathbf{R}^m$.*
If

$$(3.1) \qquad \{(\omega_3^k)_N\} \to 0 \ as \ k \to \infty$$

and the rate of convergence of $\{x^k\}$ is superlinear, then it holds that

$$(3.2) \qquad \pi_C(-\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$
$$(3.3) \qquad \omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$
$$(3.4) \qquad (\omega_3^k)_{A_+} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

Proof. From (1.5), we obtain that

$$\begin{aligned}
\omega_1^k &= -\frac{\partial L}{\partial x}(x^k,\ \lambda^k,\ \mu^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\ \lambda^k,\ \mu^k)(x^{k+1} - x^k) \\
&\quad - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) - (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= -\frac{\partial L}{\partial x}(x^{k+1},\ \lambda^k,\ \mu^k) + o(\|x^{k+1} - x^k\|) \\
&\quad - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) - (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= -\frac{\partial L}{\partial x}(x^{k+1},\ \lambda^{k+1},\ \mu^{k+1}) + o(\|x^{k+1} - x^k\|) \\
&= -\frac{\partial L}{\partial x}(\bar{x},\ \lambda^{k+1},\ \mu^{k+1}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\ \lambda^{k+1},\ \mu^{k+1})(x^{k+1} - \bar{x})
\end{aligned}$$

$$+ o(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|)$$
$$= -\frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) - \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^{k+1} - \bar{x}) - (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda})$$
$$- (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|)$$
$$= -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu})$$
$$\text{(3.5)} \qquad + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|),$$

where the second and fourth equalities follow from Taylor expansions and convergence of $\{(\lambda^k, \mu^k)\}$ to $(\bar{\lambda}, \bar{\mu})$ and superlinear convergence of $\{x^k\}$ to $\bar{x}$ are used in the last two equalities.

Similarly, from (1.5) it also follows that

$$\omega_2^k = -h(x^k) - h'(x^k)(x^{k+1} - x^k)$$
$$= -h(x^{k+1}) + o(\|x^{k+1} - x^k\|)$$
$$= -h(\bar{x}) - h'(\bar{x})(x^{k+1} - \bar{x}) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|)$$
$$= o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|).$$

The latter relation gives (3.3). Moreover, since $\bar{\mu}_{A_+} > 0$, we have that $\mu_{A_+}^k > 0$ for all k large enough, and it then follows from the last line in (1.5) that

$$(\omega_3^k)_{A_+} = -g_{A_+}(x^k) - g'_{A_+}(x^k)(x^{k+1} - x^k)$$
$$= -g_{A_+}(x^{k+1}) + o(\|x^{k+1} - x^k\|)$$
$$= -g_{A_+}(\bar{x}) - g'_{A_+}(\bar{x})(x^{k+1} - \bar{x}) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|)$$
$$= o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|),$$

which gives (3.4).

For each k, define

$$\text{(3.6)} \qquad \tilde{\omega}_1^k = -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}).$$

Then, by (3.5), it holds that

$$\text{(3.7)} \qquad \|\omega_1^k - \tilde{\omega}_1^k\| = o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|).$$

Since $g_N(x^k) \to g_N(\bar{x}) < 0$ while $(\omega_3^k)_N \to 0$ (see (3.1)), the last line in (1.5) implies that $\mu_N^k = 0$ for all k large enough. Taking this into account, we obtain from (3.6) that for such k, for any $\xi \in C$ it holds that

$$\langle -\tilde{\omega}_1^k, \xi \rangle = \langle \lambda^{k+1} - \bar{\lambda}, h'(\bar{x})\xi \rangle + \langle \mu^{k+1} - \bar{\mu}, g'(\bar{x})\xi \rangle$$
$$= \langle \mu_{A_0}^{k+1}, g'_{A_0}(\bar{x})\xi \rangle$$
$$\leq 0,$$

where the inequality $\mu^{k+1} \geq 0$ was also employed. This means that $-\tilde{\omega}_1^k \in C^\circ$, and hence, by (1.8), $\pi_C(-\tilde{\omega}_1^k) = 0$. Employing now (3.7) and the fact that $\pi_C(\cdot)$ is nonexpansive, this immediately implies (3.4). $\square$

We proceed with sufficient conditions for primal superlinear convergence of pSQP framework. The goal is, of course, to keep those conditions as close as possible to the necessary conditions of Proposition 3.1. We start with the following general fact. It

is stated for an arbitrary Lipschitz-continuous function $\pi : \mathbf{R}^n \to \mathbf{R}^n$, which later on will be chosen as either the projector π_C or the projector π_{C_+}.

PROPOSITION 3.2. *Let $\pi : \mathbf{R}^n \to \mathbf{R}^n$ be a Lipschitz-continuous function. Let $\bar{x}$ be a stationary point of problem (1.1), and let $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated Lagrange multiplier such that the estimate*

$$(3.8) \qquad \|x - \bar{x}\| = O\left(\left\|\begin{pmatrix} \pi\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix}\right\|\right)$$

holds for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Let $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and assume that for each k the triple $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ satisfies the system (1.5) with some $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, and $\omega_3^k \in \mathbf{R}^m$, and

$$(3.9) \qquad \{\pi(-\omega_1^k)\} \to 0, \ \{\omega_2^k\} \to 0, \ \{\omega_3^k\} \to 0 \ as \ k \to \infty.$$

Then

$$(3.10) \qquad \|x^{k+1} - \bar{x}\| = O(\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_A)\|) + o(\|x^k - \bar{x}\|).$$

In particular, if

$$(3.11) \qquad \pi(-\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

and conditions (3.3), (3.4), and

$$(3.12) \qquad (\omega_3^k)_{A_0} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

hold, then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. For each k, we have that

$$\begin{aligned}
\frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&\quad + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x^{k+1} - x^k) \\
&\quad + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&\quad + o(\|x^{k+1} - x^k\|) \\
(3.13) \qquad &= -\omega_1^k + o(\|x^{k+1} - x^k\|),
\end{aligned}$$

where the Taylor formula and convergence of $\{(x^k, \lambda^k, \mu^k)\}$ to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ were used in the second equality and the first relation of (1.5) was used in the last one.

Similarly, using the second relation of (1.5), we have that

$$\begin{aligned}
h(x^{k+1}) &= h(x^k) + h'(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|) \\
(3.14) \qquad &= -\omega_2^k + o(\|x^{k+1} - x^k\|).
\end{aligned}$$

Furthermore, from the last line in (1.5) we have that

$$(3.15) \qquad \min\{\mu^{k+1}, -g(x^k) - g'(x^k)(x^{k+1} - x^k) - \omega_3^k\} = 0.$$

Since $g_N(x^k) \to g_N(\bar{x}) < 0$ while $\omega_3^k \to 0$ (see (3.9)), this evidently implies that for each k large enough $\mu_N^{k+1} = 0$, and hence

$$(3.16) \qquad \min\{\mu_N^{k+1}, -g_N(x^{k+1})\} = 0.$$

Finally, using the obvious property

$$|\min\{a, b\} - \min\{a, c\}| \le |b - c| \quad \forall\, a, b, c \in \mathbf{R},$$

and again employing (3.15), for each $i \in A$ we obtain

$$\begin{aligned} |\min\{\mu_A^{k+1}, -g_A(x^{k+1})\}| = {} & |\min\{\mu_A^{k+1}, -g_A(x^k) - g_A'(x^k)(x^{k+1} - x^k) \\ & + o(\|x^{k+1} - x^k\|)\} \\ & - \min\{\mu_A^{k+1}, -g_A(x^k) - g_A'(x^k)(x^{k+1} - x^k) - (\omega_3^k)_A\}| \\ (3.17) \qquad \le {} & |(\omega_3^k)_A| + o(\|x^{k+1} - x^k\|). \end{aligned}$$

From (3.13), (3.14), (3.16), and (3.17), by the error bound (3.8) we derive the estimate

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= O(\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_A)\|) + o(\|x^{k+1} - x^k\|) \\ &= O(\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_A)\|) + o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|), \end{aligned}$$

which means the existence of a constant $c > 0$ and of a sequence $\{t_k\} \in \mathbf{R}$ such that $\{t_k\} \to 0$, and for all k

$$\|x^{k+1} - \bar{x}\| \le c\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_A)\| + t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|).$$

The latter implies that for all k large enough

$$\begin{aligned} \frac{1}{2}\|x^{k+1} - \bar{x}\| &\le (1 - t_k)\|x^{k+1} - \bar{x}\| \\ &\le c\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_A)\| + t_k\|x^k - \bar{x}\|, \end{aligned}$$

which evidently gives the estimate (3.10).

Furthermore, combining (3.10) with conditions (3.11), (3.3), (3.4), and (3.12), we conclude that

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \\ &= o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|), \end{aligned}$$

i.e.,

$$\begin{aligned} 0 &= \lim_{k \to \infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|} \\ &= \lim_{k \to \infty} \frac{1}{1 + \|x^k - \bar{x}\|/\|x^{k+1} - \bar{x}\|}. \end{aligned}$$

This implies that

$$\frac{\|x^k - \bar{x}\|}{\|x^{k+1} - \bar{x}\|} \to \infty \text{ as } k \to \infty,$$

i.e.,

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|),$$

which gives superlinear convergence rate. $\square$

We can now combine Proposition 3.2 with the error estimates under the noncriticality and second-order sufficiency assumptions obtained in section 2. Specifically, combining Theorem 2.2 and Proposition 3.2 gives the following.

THEOREM 3.3. *Let $\bar{x}$ be a stationary point of problem* (1.1), *and let* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *be an associated noncritical Lagrange multiplier. Let* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *be convergent to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *and assume that for each k the triple* $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ *satisfies the system* (1.5) *with some* $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, *and* $\omega_3^k \in \mathbf{R}^m$.

If

$$(3.18) \qquad \pi_{C_+}(\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

and conditions (3.1), (3.3), (3.4), (3.12) *hold, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Example 2.1 demonstrates that in (3.18) one cannot replace C_+ by the (generally smaller) cone C, even if $\bar{x}$ satisfies LICQ. Indeed, in this example, for any $\omega_1^k < 0$ and $\omega_3^k = 0$, the point $(x^{k+1}, \mu^{k+1}) = (\omega_1^k, 0)$ satisfies the system (1.5). If the sequence $\{\omega_1^k\}$ converges to zero, then $\{(x^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\mu})$. However, $\pi_C(-\omega_1^k) = 0$, and condition (3.18) with C_+ replaced by C does not impose any restrictions on the rate of convergence of $\{\omega_1^k\}$ and hence on the rate of convergence of $\{x^k\}$.

On the other hand, combining Theorem 2.3 and Proposition 3.2, we obtain that C_+ can be replaced by C if the noncriticality assumption is replaced by the (stronger) SOSC (1.6). Note that, unlike C_+, the critical cone C is not necessarily a linear subspace. Therefore, one cannot remove the minus sign in the left-hand side of (3.2).

THEOREM 3.4. *Let $\bar{x}$ be a stationary point of problem* (1.1), *and let* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *be an associated Lagrange multiplier satisfying SOSC* (1.6). *Let* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *be convergent to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *and assume that for each k the triple* $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ *satisfies the system* (1.5) *with some* $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, *and* $\omega_3^k \in \mathbf{R}^m$.

If conditions (3.1)–(3.4), (3.12) *hold, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Remark 3.1. In each of the conditions (3.2)–(3.4), (3.12) of Propositions 3.1 and 3.2 and Theorems 3.3 and 3.4, the right-hand side can of course be replaced by either $o(\|x^{k+1} - x^k\|)$ or $o(\|x^k - \bar{x}\|)$. The conditions modified this way are generally stronger than (3.2)–(3.4) and (3.12). However, if $\{x^k\}$ is superlinearly convergent to $\bar{x}$ (which is assumed in Proposition 3.1 and established in Proposition 3.2 or Theorems 3.3 and 3.4), all these pairs of conditions become equivalent. Thus, the results modified this way would not be any weaker.

Remark 3.2. Theorems 3.3 and 3.4 rely on the assumptions that the dual sequence converges to a multiplier which satisfies SOSC or, more generally, is noncritical. This deserves some comments. According to the analysis and numerical results in [19, 20, 21], when there exist critical multipliers, they often serve as attractors for some of the Newton-type methods under consideration, such as (quasi-Newton) SQP and LCL methods. This should be kept in mind, especially in the case of equality constraints only (in the case when there are also inequality constraints, the number of typical scenarios for dual behavior is larger). On the other hand, there are certainly problems that do not have any critical multipliers at all, and then this need not be an issue. For example, in the case of SMFCQ the multiplier is unique, and it is, therefore, the unique natural attractor for the dual sequence. This unique multiplier may very well be noncritical, of course, and even satisfy the stronger SOSC. Also, in the case of inequality constraints while attraction to critical multipliers still exists, it is somewhat

less persistent. For example, the multipliers violating strict complementarity, which may well be noncritical, also appear to attract iterates of Newton-type methods [20].

Finally, it is important to emphasize that even under SMFCQ (in which case the attraction issues mentioned above are not relevant) our analysis applied to specific algorithms in section 4 below still gives stronger results than those in the literature. Furthermore, under SMFCQ and SOSC, the algorithms in question are known to converge, and so primal-dual convergence does not need to be stated as an assumption. In this case, our assertions hold whenever a starting point is close enough to the primal-dual solution.

Remark 3.3. Combining the estimate (2.32) from Remark 2.1 with Proposition 3.2, we immediately obtain primal superlinear convergence under the assumptions of Theorem 3.4, but with SOSC (1.6) replaced by the "symmetric" condition (2.8) and with (3.2) replaced by

$$\pi_C(\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

Observe that in Example 2.1, condition (2.8) is satisfied. If $\omega_1^k < 0$, then $\pi_C(\omega_1^k) = \omega_1^k$, and the assertion stated in Remark 3.3 is evidently valid.

4. Applications to specific methods. We next show how our general analysis above improves primal superlinear convergence results for some specific algorithms.

4.1. SQP and quasi-Newton SQP. Consider again the SQP subproblem (1.3), where the symmetric $n \times n$–matrix H_k given by (1.4) corresponds to the pure SQP iteration, and otherwise H_k is some quasi-Newton approximation of $\frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$ (see, e.g., the discussion in [7, Chapter 18]). Then the KKT system of SQP subproblem (1.3) is a special case of pSQP framework (1.5) with

$$\omega_1^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) \right)(x^{k+1} - x^k), \quad \omega_2^k = 0, \quad \omega_3^k = 0.$$

The characterization of primal Q-superlinear convergence of both pure and quasi-Newton SQP methods readily follows from Proposition 3.1, Theorems 3.3 and 3.4, and Remark 3.1.

THEOREM 4.1. *Let $\bar{x}$ be a stationary point of problem* (1.1), *and let $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated Lagrange multiplier. Let $\{H_k\}$ be a sequence of $n \times n$ symmetric matrices, and let the sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be generated in the following way: for each k, x^{k+1} is a stationary point of problem* (1.3) *with $(\lambda^{k+1}, \mu^{k+1})$ being an associated Lagrange multiplier. Suppose that $\{(x^k, \lambda^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

If the rate of convergence of $\{x^k\}$ is superlinear, then the Dennis–Moré condition (1.7) *holds.*

Conversely, if $(\bar{\lambda}, \bar{\mu})$ satisfies SOSC (1.6) *and the Dennis–Moré condition* (1.7) *holds, then the rate of convergence of $\{x^k\}$ is superlinear.*

Alternatively, under the weaker assumption that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier, if

$$(4.1) \qquad \pi_{C_+}\left(\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) \right)(x^{k+1} - x^k) \right) = o(\|x^{k+1} - x^k\|),$$

then the rate of convergence of $\{x^k\}$ is also superlinear.

We emphasize that previous results in the literature (see, e.g., [5, 6], [7, Theorem 15.7] and [30, Theorem 18.5] for typical statements) assume LICQ and SOSC

for the sufficiency part of Theorem 4.1 above. By contrast, our result does not assume any constraint qualification at all. Also, for pure SQP with the choice of H_k as in (1.4), since the condition (4.1) becomes automatic, even SOSC can be further replaced by the weaker assumption that the relevant multiplier is noncritical. Finally, we emphasize that under the assumptions of SMFCQ and SOSC, SQP converges [6], and thus the assumption of primal-dual convergence is not needed in that case. Theorem 4.1 then implies primal Q-superlinear rate under SMFCQ and SOSC whenever the starting point is close enough to the primal-dual solution, and this also gives a new result.

Example 2.1 demonstrates that in the absence of SOSC (1.6), C_+ in (4.1) cannot be replaced by C. Indeed, taking, e.g., $H_k = -2$ for all k, for any $x^k < 0$ we obtain that $x^{k+1} = x^k/2 < 0$ is a stationary point of problem (1.3). Thus, for any $x^0 < 0$, the corresponding primal trajectory converges to $\bar{x} = 0$ but only linearly. At the same time, the left-hand side of (1.7) (but not (4.1)!) equals zero for all k.

Employing Remark 3.3, it can be easily seen that the Dennis-Moré condition (1.7) can be replaced in Theorem 4.1 by the "symmetric" condition

$$\pi_C\left(\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\,\lambda^k,\,\mu^k)\right)(x^{k+1} - x^k)\right) = o(\|x^{k+1} - x^k\|),$$

provided SOSC (1.6) is also replaced by the "symmetric" condition (2.8). In particular, this gives a sufficient condition for primal superlinear convergence of a quasi-Newton SQP method in Example 2.1: $H_k \to -1$ as $k \to \infty$.

4.2. LCL methods. LCL methods are traditionally stated for optimization problems with equality constraints and simple bounds, which means that general inequality constraints are reformulated as equality constraints introducing slack variables; see [32, 27, 14]. This approach is adopted, in particular, in the MINOS software package [28]. We therefore consider problem (1.1) with bound constraints given by $g(x) = -x$, $x \in \mathbf{R}^n$, i.e.,

$$
\begin{aligned}
(4.2) \qquad & \text{minimize} \quad f(x) \\
& \text{subject to} \quad h(x) = 0,\ x \geq 0.
\end{aligned}
$$

Let $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ be the Lagrangian of problem (4.2), including only the equality constraints, i.e.,

$$L(x,\,\lambda) = f(x) + \langle \lambda,\,h(x)\rangle.$$

Let $(x^k,\,\lambda^k,\,\mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ be the current iterate. Subproblems of the LCL method for (4.2) consist in minimizing the (augmented) Lagrangian subject to bounds and linearized equality constraints:

$$
\begin{aligned}
(4.3) \qquad & \text{minimize} \quad L(x,\,\lambda^k) + \frac{c_k}{2}\|h(x)\|^2 \\
& \text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) = 0,\ x \geq 0,
\end{aligned}
$$

where $c_k \geq 0$ is a penalty parameter. The next primal iterate $x^{k+1} \in \mathbf{R}^n$ is defined as a stationary point of problem (4.3). Taking an associated Lagrange multiplier $(\eta^k,\,\mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^n$, the next dual iterate is defined by $(\lambda^{k+1},\,\mu^{k+1}) = (\lambda^k + \eta^k,\,\mu^{k+1})$.

In the original LCL method proposed in [32], c_k was set equal to zero for all k (that is, the subproblems of the method involve the usual Lagrangian rather than the augmented Lagrangian). However, in practice, it is often important to employ $c_k > 0$ [28]. For asymptotic analysis, one can consider that $c_k = c \geq 0$ is fixed for all k sufficiently large, which can be ensured under natural assumptions (see, e.g., discussion in [14]).

The sharpest local convergence result for LCL methods had been obtained in [22]. It affirms quadratic primal-dual convergence under SMFCQ and SOSC. Other results in the literature require the stronger LICQ and strict complementarity in addition [32, 27, 14]. To the best of our knowledge, no primal Q-rate of convergence has been previously available. Applying our general results for pSQP framework, we obtain the following theorem.

THEOREM 4.2. *Let $\bar{x}$ be a stationary point of problem* (1.1), *and let $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated noncritical Lagrange multiplier. Let $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be a sequence generated in the following way: for each k, x^{k+1} is a stationary point of problem* (4.3) *with $c_k = c$, and $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is an associated Lagrange multiplier. Suppose that $\{(x^k, \lambda^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. The KKT system defining stationary points and multipliers of subproblem (4.3) has the form

$$\frac{\partial L}{\partial x}(x, \lambda^k) + c(h'(x))^{\mathrm{T}} h(x) + (h'(x^k))^{\mathrm{T}} \eta - \mu = 0,$$

$$h(x^k) + h'(x^k)(x - x^k) = 0,$$

$$\mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0,$$

with the dual variables $\eta \in \mathbf{R}^l$ and $\mu \in \mathbf{R}^n$. Within our pSQP framework (1.5), this corresponds to setting

$$
\begin{aligned}
\omega_1^k &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) + c(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1}) + (h'(x^k))^{\mathrm{T}} \eta^k - \mu^{k+1} \\
&\quad - \frac{\partial L}{\partial x}(x^k, \lambda^k) + \mu^k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k) - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&\quad + (\mu^{k+1} - \mu^k) \\
&= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k) \\
&\quad - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k - \eta^k) + c(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1}) \\
&= c(h'(\bar{x}))^{\mathrm{T}} h(x^{k+1}) + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

(4.4)

$$\omega_2^k = 0, \quad \omega_3^k = 0.$$

Since $\mathrm{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, it follows that first term of the right-hand side of (4.4) is orthogonal to the subspace $\ker h'(\bar{x})$ containing C_+. This implies (3.18), and the needed result follows from Theorem 3.3. ∎

Note that Theorem 4.2 implies, in particular, that under the same assumptions needed for primal-dual Q-superlinear convergence of LCL methods (SMFCQ and SOSC), we have primal Q-superlinear convergence as well. Again, under these assumptions LCL methods converges, and so the corresponding assumption in Theorem 4.2 is not needed in that case.

4.3. Sequential quadratically constrained quadratic programming. The *sequential quadratically constrained quadratic programming* (SQCQP) method for problem (1.1) is the following algorithm (see [35, 25, 2, 15, 34, 11]). For the current iterate x^k, the next primal iterate $x^{k+1} \in \mathbf{R}^n$ is computed as a stationary point of the subproblem

$$\text{minimize} \quad \langle f'(x^k),\, x - x^k \rangle + \frac{1}{2} f''(x^k)[x - x^k,\, x - x^k]$$

(4.5) $\quad$ subject to $\quad h(x^k) + h'(x^k)(x - x^k) + \frac{1}{2} h''(x^k)[x - x^k,\, x - x^k] = 0,$

$$g(x^k) + g'(x^k)(x - x^k) + \frac{1}{2} g''(x^k)[x - x^k,\, x - x^k] \leq 0,$$

and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ is defined as an associated Lagrange multiplier.

In the convex case, subproblem (4.5) can be cast as a second-order cone program [26, 29], which can be solved efficiently by interior-point algorithms. Another possibility for the convex case is [18]. In [2], nonconvex subproblems were also handled quite efficiently by using other computational techniques. In the nonconvex case, one might also use the approaches from [1, 3] for solving the subproblems.

The primal superlinear convergence result in [2] refers to a trust-region version of SQCQP with exact values of second derivatives and assumes Mangasarian–Fromovitz constraint qualification (MFCQ) and a quadratic growth condition (a weak form of SOSC). Quadratic primal-dual convergence had been established in [22] under SMFCQ and SOSC. Superlinear primal convergence under Dennis–Moré-type conditions is shown in [11] (in a more general variational context) assuming LICQ and SOSC. Applying our general pSQP framework, we obtain that primal convergence is superlinear whenever there is primal-dual convergence and the relevant multiplier is noncritical.

THEOREM 4.3. *Let $\bar{x}$ be a stationary point of problem (1.1), and let $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$ be an associated noncritical Lagrange multiplier. Let $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be generated in the following way: for each k, x^{k+1} is a stationary point of subproblem (4.5), and $(\lambda^{k+1},\, \mu^{k+1})$ is an associated Lagrange multiplier. Suppose that $\{(x^k,\, \lambda^k,\, \mu^k)\}$ converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. It can be readily seen that the KKT conditions of subproblem (4.5) correspond to the pSQP framework (1.5) with

$$\omega_1^k = (h''(x^k)[x^{k+1} - x^k])^{\mathrm{T}} (\lambda^{k+1} - \lambda^k) + (g''(x^k)[x^{k+1} - x^k])^{\mathrm{T}} (\mu^{k+1} - \mu^k)$$
$$= o(\|x^{k+1} - x^k\|),$$

$$\omega_2^k = \frac{1}{2} h''(x^k)[x^{k+1} - x^k,\, x^{k+1} - x^k]$$
$$= O(\|x^{k+1} - x^k\|^2),$$

$$\omega_3^k = \frac{1}{2} g''(x^k)[x^{k+1} - x^k,\, x^{k+1} - x^k]$$
$$= O(\|x^{k+1} - x^k\|^2).$$

The needed result follows immediately from Theorem 3.3. □

4.4. Stabilized SQP is an open question. The stabilized version of SQP (sSQP) was proposed in [36] and further studied in [16, 13, 37, 38, 12, 23]. Given the current iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the next iterate is computed as a

stationary point of the following quadratic programming problem in the primal-dual space:

$$\text{minimize}_{(x,\lambda,\mu)} \quad \left\{ \langle f'(x^k),\, x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k),\, x - x^k \right\rangle \right.$$

(4.6)
$$\left. + \frac{\sigma_k}{2}(\|\lambda\|^2 + \|\mu\|^2) \right\}$$

$$\text{subject to} \qquad h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0,$$
$$g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \le 0,$$

where

$$\sigma_k = \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) \\ h(x^k) \\ \min\{\mu^k,\, -g(x^k)\} \end{pmatrix} \right\|.$$

The sharpest result for problems with inequality constraints assumes SOSC (no constraint qualifications of any kind) and affirms a local quadratic rate of convergence of the primal-dual sequence [12]. In the case when there are equality constraints only, SOSC can be further relaxed to the assumption that the multiplier in question is noncritical [23]. No primal convergence results are available in the literature, except for [10] that shows a kind of "two-step" primal superlinear estimate. It is thus tempting to try to derive primal superlinear convergence of sSQP using our general approach.

One can easily see that sSQP fits the pSQP framework (1.5) by taking

$$\omega_1^k = 0, \quad \omega_2^k = -\sigma_k(\lambda^{k+1} - \lambda^k), \quad \omega_3^k = -\sigma_k(\mu^{k+1} - \mu^k).$$

That said, our general results do not seem to be readily applicable in this case because conditions (3.3), (3.4), and (3.12) cannot be guaranteed for the form of perturbations at hand. The primal superlinear convergence of sSQP iterations, or lack thereof, remains an open question at this time. The following example indicates certain difficulties one faces in analyzing the primal rate of convergence of sSQP—it shows that even under LICQ and SOSC, there may exist points arbitrarily close to the solution from which the distance to the primal solution actually increases rather than decreases. That said, it is important to stress that this does not preclude *eventual* primal superlinear convergence of an sSQP sequence, since superlinear decrease can occur already from the next iteration on. And this is indeed the case in this particular example. Nevertheless, even though this example does not settle the question, it is a good illustration of the difficulties involved.

Example 4.1. Let $n = m = 1$, $l = 0$, $f(x) = (x-1)^2/2$, $g(x) = -\log(x)$. Then $\bar{x} = 1$ is the unique solution and the unique stationary point of problem (1.1). (To be consistent with the setting of the paper where the functions are defined on the whole space, we can set $g(x) = -\log(x)$ in an appropriate neighborhood of $\bar{x}$ and extend it smoothly to the rest of $\mathbf{R}$.) The solution $\bar{x}$ satisfies LICQ, and $\bar{\mu} = 0$ is the unique associated multiplier. Since $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu}) = 1 > 0$, SOSC holds.

Consider points of the form $(x_\varepsilon, \mu_\varepsilon) = (1 + \varepsilon^3, \varepsilon)$, $\varepsilon > 0$, which can be made arbitrarily close to the primal-dual solution $(\bar{x}, \bar{\mu})$ by taking ε small enough.

The KKT system of sSQP subproblem (4.6) for $(x^k, \mu^k) = (x_\varepsilon, \mu_\varepsilon)$ has the form

$$0 = x_\varepsilon - 1 + \left(1 + \frac{\mu_\varepsilon}{x_\varepsilon^2}\right)(x - x_\varepsilon) - \frac{\mu}{x_\varepsilon},$$

(4.7)
$$\mu \geq 0, \quad -\log(x_\varepsilon) - \frac{1}{x_\varepsilon}(x - x_\varepsilon) - \sigma_\varepsilon(\mu - \mu_\varepsilon) \leq 0,$$

$$\mu\left(-\log(x_\varepsilon) - \frac{1}{x_\varepsilon}(x - x_\varepsilon) - \sigma_\varepsilon(\mu - \mu_\varepsilon)\right) = 0,$$

where

$$\sigma_\varepsilon = \left\| \begin{pmatrix} \varepsilon^3 - \frac{\varepsilon}{1+\varepsilon^3} \\ \min\{\varepsilon, \log(1 + \varepsilon^3)\} \end{pmatrix} \right\| > 0.$$

For ε small enough, the unique solution of (4.7) is given by

(4.8) $\quad \bar{x}_\varepsilon = \frac{1}{\alpha_\varepsilon}(x_\varepsilon^2 + x_\varepsilon\mu_\varepsilon + x_\varepsilon\bar{\mu}_\varepsilon), \quad \bar{\mu}_\varepsilon = \frac{1}{\beta_\varepsilon}\left(\sigma_\varepsilon\mu_\varepsilon + \frac{x_\varepsilon(x_\varepsilon - 1)}{\alpha_\varepsilon} - \log(x_\varepsilon)\right),$

where $\alpha_\varepsilon = x_\varepsilon^2 + \mu_\varepsilon$ and $\beta_\varepsilon = 1 + \sigma_\varepsilon\alpha_\varepsilon$. Using that $\log(x) = x - 1 + o(|x - 1|)$, we obtain

$$\bar{x}_\varepsilon - 1 = \frac{1}{\alpha_\varepsilon}((x_\varepsilon - 1)\mu_\varepsilon + x_\varepsilon\bar{\mu}_\varepsilon),$$

$$x_\varepsilon\bar{\mu}_\varepsilon = \frac{1}{\beta_\varepsilon}\left(\sigma_\varepsilon x_\varepsilon\mu_\varepsilon + \frac{x_\varepsilon^2(x_\varepsilon - 1)}{\alpha_\varepsilon} - x_\varepsilon(x_\varepsilon - 1)\right) + o(|x_\varepsilon - 1|).$$

Thus

$$\frac{|\bar{x}_\varepsilon - 1|}{|x_\varepsilon - 1|} = \frac{\bar{x}_\varepsilon - 1}{x_\varepsilon - 1} = \frac{\mu_\varepsilon}{\alpha_\varepsilon} + \frac{1}{\alpha_\varepsilon\beta_\varepsilon}\left(\frac{x_\varepsilon^2}{\alpha_\varepsilon} - x_\varepsilon\right) + \frac{o(|x_\varepsilon - 1|)}{x_\varepsilon - 1} + \frac{1}{\alpha_\varepsilon\beta_\varepsilon}\frac{\sigma_\varepsilon x_\varepsilon\mu_\varepsilon}{x_\varepsilon - 1}.$$

It can be seen that for ε small enough, the first three terms in the right-hand side can be made arbitrarily small, while for the last one it holds that

(4.9)
$$\frac{\sigma_\varepsilon x_\varepsilon\mu_\varepsilon}{x_\varepsilon - 1} > \frac{\varepsilon}{2}\frac{x_\varepsilon\mu_\varepsilon}{x_\varepsilon - 1} = \frac{1}{2}\left(\frac{1}{\varepsilon} + \varepsilon^2\right) > \frac{1}{2\varepsilon},$$

where in the first relation we use that $\sigma_\varepsilon = \varepsilon + o(\varepsilon) > \varepsilon/2$ for ε small enough. Hence, for ε close to zero, the above relation shows that sSQP steps from points of the given form actually increase rather than decrease the distance to the primal solution. That said, it can be seen that the sSQP sequence does converge even from those starting points, and the primal rate of convergence is eventually superlinear.

5. Concluding remarks. We have derived sharp primal superlinear convergence results for a class of Newtonian methods for constrained optimization, which includes (quasi-Newton) SQP, LCL, and SQCQP. In particular, we have shown that for those methods the primal sequence converges superlinearly under the same assumptions needed to prove primal-dual superlinear convergence. More specifically, if the primal-dual convergence is assumed or already established, the only property needed to show primal superlinear rate is a certain error bound estimate. This estimate holds, for example, under second-order sufficiency, which is typically required to prove primal-dual convergence in any case. Moreover, in some situations second-order sufficiency can be further relaxed to the assumption that the relevant multiplier is noncritical. Open questions refer to the primal rate of convergence of the sSQP and possible extensions to other methods, like reduced Hessian methods, for example.

REFERENCES

[1] L.T.H. AN, *An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints*, Math. Program., 87 (2000), pp. 401–426.

[2] M. ANITESCU, *A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming*, SIAM J. Optim., 12 (2002), pp. 949–978.

[3] C. AUDET, P. HANSEN, B. JAUMARD, AND G. SAVARD, *A branch and cut algorithm for nonconvex quadratically constrained quadratic programming*, Math. Program., 87 (2000), pp. 131–152.

[4] B.T. BOGGS AND J.W. TOLLE, *Sequential quadratic programming*, Acta Numer., 4 (1996), pp. 1–51.

[5] P.T. BOGGS, J.W. TOLLE, AND P. WANG, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control Optim., 20 (1982), pp. 161–171.

[6] J.F. BONNANS, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Appl. Math. Optim., 29 (1994), pp. 161–186.

[7] J.F. BONNANS, J.CH. GILBERT, C. LEMARÉCHAL, AND C. SAGASTIZÁBAL, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer-Verlag, Berlin, 2006.

[8] J.F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[9] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[10] D. FERNÁNDEZ, *A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems*, submitted.

[11] D. FERNÁNDEZ AND M. SOLODOV, *On local convergence of sequential quadratically-constrained quadratic-programming type methods, with an extension to variational problems*, Comput. Optim. Appl., 39 (2008), pp. 143–160.

[12] D. FERNÁNDEZ AND M. SOLODOV, *Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems*, Math. Program., 125 (2010), pp. 47–73.

[13] A. FISCHER, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124.

[14] M.P. FRIEDLANDER AND M.A. SAUNDERS, *A globally convergent linearly constrained Lagrangian method for nonlinear optimization*, SIAM J. Optim., 15 (2005), pp. 863–897.

[15] M. FUKUSHIMA, Z.-Q. LUO, AND P. TSENG, *A sequential quadratically constrained quadratic programming method for differentiable convex minimization*, SIAM J. Optim., 13 (2003), pp. 1098–1119.

[16] W.W. HAGER, *Stabilized sequential quadratic programming*, Comput. Optim. Appl., 12 (1999), pp. 253–273.

[17] W.W. HAGER AND M.S. GOWDA, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192.

[18] Z.-H. HUANG, D. SUN, AND G. ZHAO, *A smoothing Newton-type algorithm of stronger convergence for the quadratically constrained convex quadratic programming*, Comput. Optim. Appl., 35 (2006), pp. 199–237.

[19] A.F. IZMAILOV AND M.V. SOLODOV, *On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions*, Math. Program., 117 (2009), pp. 271–304.

[20] A.F. IZMAILOV AND M.V. SOLODOV, *Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints*, Comput. Optim. Appl., 42 (2009), pp. 231–264.

[21] A.F. IZMAILOV AND M.V. SOLODOV, *On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers*, Math. Program., 2009, DOI 10.1007/s10107-009-0279-4.

[22] A.F. IZMAILOV AND M.V. SOLODOV, *Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization*, Comput. Optim. Appl., 46 (2010), pp. 347–368.

[23] A.F. IZMAILOV AND M.V. SOLODOV, *Stabilized SQP Revisited*, 2010, DOI 10.1007/s10107-010-0413-3.

[24] A.F. IZMAILOV AND M.V. SOLODOV, *A truncated SQP method based on inexact interior-point solutions of subproblems*, SIAM J. Optim., 20 (2010), pp. 2584–2613.

[25] S. KRUK AND H. WOLKOWICZ, *Sequential, quadratically constrained, quadratic programming for general nonlinear programming*, in Handbook of Semidefinite Programming, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Academic Publishers, Dordrecht, 2000, pp. 563–575.

[26] M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Applications of second-order cone programming*, Linear Algebra Appl., 284 (1998), pp. 193–228.
[27] B.A. Murtagh and M.A. Saunders, *A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints*, Math. Program. Study, 16 (1982), pp. 84–117.
[28] B.A. Murtagh and M.A. Saunders, *MINOS* 5.0 *User's Guide*, Technical report SOL 83.20, Stanford University, Palo Alto, CA, 1983.
[29] Y.E. Nesterov and A.S. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM Publications, Philadelphia, 1993.
[30] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[31] 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, London, 1978, pp. 27–63.
[32] S.M. Robinson, *A quadratically convergent algorithm for general nonlinear programming problems*, Math. Program., 3 (1972), pp. 145–156.
[33] S.M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class nonlinear-programming algorithms*, Math. Program., 7 (1974), pp. 1–16.
[34] M.V. Solodov, *On the sequential quadratically constrained quadratic programming methods*, Math. Oper. Res., 29 (2004), pp. 64–79.
[35] E.J. Wiest and E. Polak, *A generalized quadratic programming-based phase*-I–*phase*-II *method for inequality-constrained optimization*, Appl. Math. Optim., 26 (1992), pp. 223–252.
[36] S.J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275.
[37] S.J. Wright, *Modifying SQP for degenerate problems*, SIAM J. Optim., 13 (2002), pp. 470–497.
[38] S.J. Wright, *Constraint identification and algorithm stabilization for degenerate nonlinear programs*, Math. Program., 95 (2003), pp. 137–160.