A TRUNCATED SQP METHOD BASED ON INEXACT INTERIOR-POINT SOLUTIONS OF SUBPROBLEMS*

A. F. IZMAILOV† AND M. V. SOLODOV‡

Abstract. We consider sequential quadratic programming (SQP) methods applied to optimization problems with nonlinear equality constraints and simple bounds. In particular, we propose and analyze a truncated SQP algorithm in which subproblems are solved approximately by an infeasible predictor-corrector interior-point method, followed by setting to zero some variables and some multipliers so that complementarity conditions for approximate solutions are enforced. Verifiable truncation conditions based on the residual of optimality conditions of subproblems are developed to ensure both global and fast local convergence. Global convergence is established under assumptions that are standard for linesearch SQP with exact solution of subproblems. The local superlinear convergence rate is shown under the weakest assumptions that guarantee this property for pure SQP with exact solution of subproblems, namely, the strict Mangasarian–Fromovitz constraint qualification and second-order sufficiency. Local convergence results for our truncated method are presented as a special case of the local convergence for a more general perturbed SQP framework, which is of independent interest and is applicable even to some algorithms whose subproblems are not quadratic programs. For example, the framework can also be used to derive sharp local convergence results for linearly constrained Lagrangian methods. Preliminary numerical results confirm that it can be indeed beneficial to solve subproblems approximately, especially on early iterations.

Key words. sequential quadratic programming, inexact sequential quadratic programming, truncated sequential quadratic programming, interior-point method, superlinear convergence

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

DOI. 10.1137/090758015

1. Introduction. In this paper, we are concerned with truncated and perturbed versions of sequential quadratic programming (SQP) methods [5, 27] for constrained optimization. We refer to an algorithm as a truncated SQP method if the solver for the SQP subproblem may be terminated early, producing an inexact solution. This is, in fact, a special case of the more general class of perturbed SQP methods, which we define as any approach where the iterates can be viewed, perhaps a posteriori, as approximate solutions to relevant SQP subproblems. We thus regard truncation as a special way of inducing perturbations. Our local convergence analysis allows other forms of perturbations as well and is also applicable to some methods that are not modifications of SQP as such. One example is the linearly constrained (augmented) Lagrangian method [42, 39, 18, 33].

We shall consider problems with equality constraints and simple bounds:

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

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraint mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ are differentiable. Not including general inequality constraints does not reduce the applicability of our development, as they can be converted into equalities introducing nonnegative slack variables. It should also be noted that many well-recognized solvers transform inequalities into equalities and bounds at the preprocessing stage anyway (for example, SNOPT [21] and MINOS [40]). In addition, in our proposal quadratic programming (QP) subproblems within SQP would be solved by an interior-point (IP) method, and commonly used IP methods convert subproblems to the form with equality constraints and simple bounds in any case. This is another reason why it makes sense in our context to consider the setting of (1.1) from the beginning.

SQP methods for problem (1.1) solve subproblems of the form

(1.2)
$$\begin{aligned} \text{minimize} \quad & f(x^k) + \langle f'(x^k),\, x - x^k \rangle + \tfrac{1}{2} \langle H_k(x - x^k),\, x - x^k \rangle \\ \text{subject to} \quad & h(x^k) + h'(x^k)(x - x^k) = 0, \quad x \geq 0, \end{aligned}$$

where the $n \times n$ matrix H_k aims to approximate, in some sense, the Hessian of the Lagrangian or of the augmented Lagrangian at the current point. SQP methods are known to be very efficient on small- to medium-sized problems, and much less so in the large-scale case; see, e.g., the discussions in [25, 26]. (There are some exceptions for certain special large-scale applications; also, when constraints have some separable structure, it may allow one to split subproblems into smaller ones [44].) One difficulty in the large-scale case is the cost of accurately solving large quadratic programs (QPs). This becomes a particularly serious issue when inequality constraints are present; equality-constrained QPs are computationally much cheaper. In fact, regardless of the size, it is clear that solving subproblems exactly, or very accurately, can be quite wasteful, especially on early iterations. Those considerations call for the development of approximation or truncation conditions for solving the subproblems.

Recall that in the classical *truncated Newton* approach [13] for solving a system of nonlinear equations

$$\Phi(u) = 0,$$

where $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is smooth, one iterates on the linear system of the subproblem of the Newton method

$$\Phi(u^k) + \Phi'(u^k)(u - u^k) = 0$$

until a point u^{k+1} satisfying the truncation condition

$$\| \Phi(u^k) + \Phi'(u^k)(u^{k+1} - u^k) \| \leq \theta_k \| \Phi(u^k) \|$$

is obtained ($\{\theta_k\}$ is the so-called forcing sequence satisfying $\{\theta_k\} \subset [0, \theta),\, 0 < \theta < 1$). Similar ideas are applicable to equality-constrained optimization problems since then the optimality conditions form a system of equations. However, a number of important modifications are required in that context, especially if the constraints are nonlinear and/or the objective function is nonconvex.

Some inexact/truncated SQP methods for problems with equality constraints are given in [17, 34, 30, 10, 11]. In [17], only local convergence is considered. A globally convergent linesearch reduced inexact SQP method is presented in [34], although some difficult to estimate parameters are used in the algorithm formulation. A readily implementable composite-step trust-region inexact SQP method is given in [30],

where inexactness can come from approximate solution of subproblems as well as from approximation of derivatives. Composite-step trust-region SQP methods with inexact computation of the step are also considered in [35, 12]. Other related work is found in [29, 47]. The approach of [11] concerns linesearch SQP and accepts an approximate SQP direction when it decreases the local approximation of the merit function, at the same time satisfying bounds on primal and dual components of the residual of the subproblem optimality conditions.

When inequality constraints are present (even if in the form of simple bounds!), the situation is very different from the equality-constrained case because of the complementarity part in the optimality conditions, and what might be a sound truncation rule is not immediately clear. It appears (see the discussions in [25, 26]) that very few approaches to truncated SQP for problems with inequality constraints have been proposed so far, namely, [6, 38, 23, 36]. Our developments are rather different and do not seem directly comparable with those just cited. Nevertheless, some comments are in order. In [38, 23], subproblems are solved by an active-set QP method, and the augmented Lagrangian is used as a merit function. In [6], an IP "optimal subspace" method is applied to solve subproblems; i.e., each minor iteration consists of solving exactly a reduced-space QP by an IP method, and a special merit function is introduced. In [6, 38, 23], truncation of QP iterations is based on the descent condition for the merit function employed. Rate of convergence results in [6, 38] require the linear independence constraint qualification, strict complementarity, and second-order sufficiency, among other assumptions. The closest to our development is probably [36], which also uses an IP method to solve the subproblems, truncating the iterations when the residual of the mixed complementarity formulation of optimality conditions for the given subproblem is sufficiently reduced. However, [36] presents local analysis only with no attempt at globalization, and in the global phase our truncation conditions are not related to those of [36]. Also, local analysis in [36] requires Robinson's strong regularity, which is stronger than our assumptions, to be discussed below.

In this paper, we suggest solving QP subproblems (1.2) approximately, by an infeasible predictor-corrector IP method. We give verifiable approximation conditions based on the residual of the Karush–Kuhn–Tucker (KKT) optimality system for subproblem (1.2) and the directional derivative of the l_1-penalty function that guarantee both global and local superlinear convergence. Global convergence is guided by linesearch for the l_1-penalty function and is established under the assumptions that are also standard for linesearch SQP with exact solution of subproblems. Our local analysis assumes the strict Mangasarian–Fromovitz constraint qualification, which is weaker than the linear independence constraint qualification, and the second-order sufficiency. Strict complementarity is not required. These are, in fact, the weakest set of assumptions needed for pure SQP with exact solution of subproblems to converge superlinearly [7]. The key to the construction is that the KKT residual of subproblems is checked not for the iterates produced by the IP method but for "purified" estimates obtained from the IP iterates by setting to zero some of the variables and some of the multipliers, so that complementarity holds. The other parts of KKT conditions are not satisfied, of course. The feature of enforcing complementarity is a subtle issue, especially for the local analysis. It appears important that the perturbed KKT conditions should be *exact* KKT conditions for some underlying *perturbed* optimization problem. This is discussed in detail in section 4.2; see in particular (4.18) and the associated (4.19). Note that if complementarity conditions do not hold, the perturbed KKT system cannot be a KKT system for *any* optimization problem.

The rest of the paper is organized as follows. In section 2 we formally state our truncated SQP (tSQP) algorithm and give some details of the IP method used to solve the QP subproblems. Section 3 shows that the method is well defined and globally convergent under the assumptions typical for linesearch SQP methods. In section 4, we first present a rather general perturbed SQP framework and its local analysis and then apply it to obtain local superlinear convergence of the proposed tSQP method under weak assumptions. It should be emphasized that the general framework also has applications other than inexact/truncated versions of SQP. For example, it is applicable to linearly constrained (augmented) Lagrangian (LCL) methods [42, 39, 18, 33], as discussed in section 4.4. We conclude the paper with some preliminary numerical results in section 5, demonstrating that the suggested truncation ideas are indeed potentially useful.

Some words about our notation are in order. We denote $\mathbf{R}^s_+ = \{z \in \mathbf{R}^s \mid z_i \geq 0, \ i = 1, \ldots, s\}$ and $\mathbf{R}^s_{++} = \{z \in \mathbf{R}^s \mid z_i > 0, \ i = 1, \ldots, s\}$. For our purposes, it is convenient to define the (partial) Lagrangian of problem (1.1) that involves the equality constraints only, omitting the bounds. Specifically, we define $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ by

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

Then stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$(1.3) \qquad \frac{\partial L}{\partial x}(x, \lambda) = \mu, \quad h(x) = 0, \quad \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0,$$

or, equivalently, by zeros of its natural residual $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$,

$$(1.4) \qquad \rho(x, \lambda, \mu) = \left\| \left(\frac{\partial L}{\partial x}(x, \lambda) - \mu, \ h(x), \ \min\{\mu, x\} \right) \right\|.$$

If the norm is not specified, it is always the 2-norm. Otherwise, we use the explicit notation $\| \cdot \|_\infty$ and $\| \cdot \|_1$.

2. The tSQP algorithm. In this section, we present our global tSQP algorithm and discuss some details of its possible implementations. In section 2.1, we state the generic algorithm without discussing any specifics of IP methods that can be used to generate approximate solutions of subproblems that satisfy the proposed truncation conditions. Relevant details concerning a specific class of IP methods constitute section 2.2.

2.1. The generic algorithm. Given the current primal iterate x^k and a symmetric $n \times n$ matrix H_k, the SQP subproblem to solve is stated in (1.2), and its approximate KKT conditions that are relevant in our setting are

$$
\begin{aligned}
& f'(x^k) + H_k(x - x^k) + (h'(x^k))^\top \lambda - \mu + \omega_1^k = 0, \\
(2.1) \qquad & h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0, \\
& \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0,
\end{aligned}
$$

where $\omega_1^k \in \mathbf{R}^n$ and $\omega_2^k \in \mathbf{R}^l$ are perturbation terms accounting for inexactness. As already mentioned, the iterates produced by an IP method applied to (1.2) would be "purified," so that complementarity conditions in (2.1) would always hold exactly; see

step 2 of Algorithm 2.1 below. Once conditions (2.1) hold with (ω_1^k, ω_2^k) satisfying certain prescribed requirements, the obtained approximate primal solution is refined by linesearch aimed at decreasing the l_1-penalty function

$$\psi_{c_k}(x) = f(x) + c_k \|h(x)\|_1,$$

where $c_k > 0$ is the penalty parameter. The bound constraints are not penalized, because they are satisfied along the iterations. Sufficient descent of ψ_{c_k} with respect to its value at the point x^k (see (2.7)) is measured using the estimate of the directional derivative of ψ_{c_k} in the obtained direction, defined in (2.5). Our truncation rules guarantee that this directional derivative is sufficiently negative; see (2.4).

Specifically, the method is the following.

ALGORITHM 2.1 (tSQP).
Choose some $(x^0, \lambda^0, \mu^0) \in \mathbf{R}_+^n \times \mathbf{R}^l \times \mathbf{R}_+^n$, $c_{-1} > 0$, $\bar{c} > 0$, $\varepsilon \in (0, 1)$, $\tau \in (0, 1)$, and $\theta_1 \geq 0$, $\theta_2 \in (0, 1)$, $\theta_3 > 1$. Set $k = 0$.

1. Stop if $\rho(x^k, \lambda^k, \mu^k) = 0$.

2. Choose an $n \times n$ symmetric positive definite matrix H_k and apply to the QP (1.2) the IP method (described in section 2.2) to generate the sequence $\{(x^{k,i}, \lambda^{k,i}, \mu^{k,i})\} \subset \mathbf{R}_{++}^n \times \mathbf{R}^l \times \mathbf{R}_{++}^n$. For each $i = 0, 1, \ldots$, define $\hat{x}^{k,i} \in \mathbf{R}_+^n$ and $\hat{\mu}^{k,i} \in \mathbf{R}_+^n$ by

$$(2.2) \qquad \hat{x}_j^{k,i} = \begin{cases} x_j^{k,i} & \text{if } x_j^{k,i} \geq \mu_j^{k,i}, \\ 0 & \text{if } x_j^{k,i} < \mu_j^{k,i}, \end{cases} \qquad \hat{\mu}_j^{k,i} = \begin{cases} 0 & \text{if } x_j^{k,i} \geq \mu_j^{k,i}, \\ \mu_j^{k,i} & \text{if } x_j^{k,i} < \mu_j^{k,i}, \end{cases}$$

$j = 1, \ldots, n$, and set $p^{k,i} = \hat{x}^{k,i} - x^k$.

Stop the IP iterations once the conditions

$$(2.3) \qquad\qquad\qquad\qquad \|\omega_1^{k,i}\|_1 \leq \theta_1 \|p^{k,i}\|_1,$$

$$(2.4) \qquad\qquad\qquad \Delta_{k,i} \leq -\theta_2 \langle H_k p^{k,i}, p^{k,i} \rangle - \bar{c} \|h(x^k)\|_1$$

hold for $\omega_1^{k,i} = \omega_1^k$ and $\omega_2^{k,i} = \omega_2^k$ defined by (2.1) with $(x, \lambda, \mu) = (\hat{x}^{k,i}, \lambda^{k,i}, \hat{\mu}^{k,i})$, where

$$(2.5) \qquad\qquad \Delta_{k,i} = \langle f'(x^k), p^{k,i} \rangle - c_{k,i}(\|h(x^k)\|_1 - \|\omega_2^{k,i}\|_1)$$

and

$$(2.6) \qquad c_{k,i} = \begin{cases} c_{k-1} & \text{if } h(x^k) = 0 \text{ or } c_{k-1} \geq \theta_3(\|\lambda^{k,i}\|_\infty + \bar{c}), \\ \theta_3(\|\lambda^{k,i}\|_\infty + 2\bar{c}) & \text{if } h(x^k) \neq 0 \text{ and } c_{k-1} < \theta_3(\|\lambda^{k,i}\|_\infty + \bar{c}). \end{cases}$$

Set $(\lambda^{k+1}, \mu^{k+1}) = (\lambda^{k,i}, \hat{\mu}^{k,i})$, $\hat{x}^k = \hat{x}^{k,i}$, $p^k = \hat{x}^k - x^k$, $c_k = c_{k,i}$, $\Delta_k = \Delta_{k,i}$.

3. Compute $\alpha_k = \tau^s$, where s is the smallest nonnegative integer such that

$$(2.7) \qquad\qquad\qquad \psi_{c_k}(x^k + \tau^s p^k) \leq \psi_{c_k}(x^k) + \varepsilon \tau^s \Delta_k.$$

4. Set $x^{k+1} = x^k + \alpha_k p^k$. Set $k = k + 1$ and go to step 1.

The key to the construction of Algorithm 2.1 is our ability to satisfy the truncation conditions (2.3) and (2.4) after a finite number of purified IP iterations, and the fact that (2.4) ensures that the obtained direction is that of descent for the penalty

function with the penalty parameter chosen according to (2.6). To make this clear, we immediately show that (2.4) would hold once the residual terms in the approximate KKT conditions (2.1) of subproblems are small enough to satisfy

$$
(2.8) \qquad \|\omega_2^{k,\,i}\|_1 \leq
\begin{cases}
\left(1 - \dfrac{1}{\theta_3}\right)\|h(x^k)\|_1 & \text{if } h(x^k) \neq 0, \\[2ex]
\dfrac{\delta}{c_{k,\,i}}\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle & \text{if } h(x^k) = 0,
\end{cases}
$$

$$
(2.9) \qquad \langle \lambda^{k,\,i},\, \omega_2^{k,\,i}\rangle - \langle \omega_1^{k,\,i},\, p^{k,\,i}\rangle \leq (1 - \theta_2 - \delta)\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle,
$$

where $\delta \in (0,\, 1 - \theta_2)$. We note that the two conditions above are sufficient for (2.4) but not necessary.

PROPOSITION 2.1. *Conditions* (2.8)–(2.9) *imply* (2.4).

Proof. Using (2.1), we derive that

$$
\begin{aligned}
\langle f'(x^k),\, p^{k,\,i}\rangle &= -\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle - \langle \lambda^{k,\,i},\, h'(x^k)p^{k,\,i}\rangle + \langle \hat{\mu}^{k,\,i} - \omega_1^{k,\,i},\, p^{k,\,i}\rangle \\
&= -\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle + \langle \lambda^{k,\,i},\, \omega_2^{k,\,i}\rangle - \langle \omega_1^{k,\,i},\, p^{k,\,i}\rangle + \langle \lambda^{k,\,i},\, h(x^k)\rangle \\
&\quad + \langle \hat{\mu}^{k,\,i},\, \hat{x}^{k,\,i}\rangle - \langle \hat{\mu}^{k,\,i},\, x^k\rangle \\
&\leq -(\theta_2 + \delta)\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle + \|h(x^k)\|_1\|\lambda^{k,\,i}\|_\infty,
\end{aligned}
$$

where the inequality follows from (2.9) and the facts that $\langle \hat{\mu}^{k,\,i},\, \hat{x}^k\rangle = 0$ by the construction of the purified IP iterates, and $\hat{\mu}^{k,\,i} \geq 0$, $x^k \geq 0$.

Using further definition (2.5), we obtain that

$$
\begin{aligned}
\Delta_{k,\,i} &= \langle f'(x^k),\, p^{k,\,i}\rangle - c_{k,\,i}(\|h(x^k)\|_1 - \|\omega_2^{k,\,i}\|_1) \\
&\leq -(\theta_2 + \delta)\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle + \|h(x^k)\|_1\|\lambda^{k,\,i}\|_\infty - c_{k,\,i}(\|h(x^k)\|_1 - \|\omega_2^{k,\,i}\|_1) \\
&\leq
\begin{cases}
-(\theta_2 + \delta)\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle - \left(\dfrac{c_{k,\,i}}{\theta_3} - \|\lambda^{k,\,i}\|_\infty\right)\|h(x^k)\|_1 & \text{if } h(x^k) \neq 0, \\[2ex]
-\theta_2\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle & \text{if } h(x^k) = 0,
\end{cases}
\end{aligned}
$$

where (2.8) has been used to derive the last inequalities. Since (2.6) implies that

$$
\frac{c_{k,\,i}}{\theta_3} - \|\lambda^{k,\,i}\|_\infty \geq \bar{c},
$$

it then follows that

$$
\Delta_{k,\,i} \leq -\theta_2\langle H_k p^{k,\,i},\, p^{k,\,i}\rangle - \bar{c}\|h(x^k)\|_1,
$$

whether $h(x^k) = 0$ or $h(x^k) \neq 0$. $\quad\square$

Some further remarks are in order.

Remark 2.1. In Algorithm 2.1, feasibility of subproblems is assumed. This is, in fact, a strong assumption unless h is affine. However, dealing with infeasibility is an issue completely unrelated to truncation of QP subproblems. It should be dealt with in exactly the same way as in any other SQP method, e.g., using "elastic mode" [20]. Since our focus here is truncation, we shall not enter into further details.

Remark 2.2. We believe that some other merit functions, linesearch rules, and penalty parameter updates can be used as well. This is again not related to truncation as such, although truncation conditions would of course need to be adjusted to any changes.

Remark 2.3. Assuming that f and h are twice differentiable, for fast local convergence we shall make the augmented Lagrangian [41, p. 539], [8, p. 330] choice of the matrix H_k, i.e.,

$$(2.10) \qquad H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \tilde{\lambda}^k) + \beta (h'(x^k))^{\mathrm{T}} h'(x^k),$$

where $\beta \geq 0$ is the parameter and

$$(2.11) \qquad \tilde{\lambda}^0 = \lambda^0, \quad \tilde{\lambda}^k = \lambda^k - \beta h(x^{k-1}), \quad k = 1, 2, \ldots.$$

As usual for local analysis, β is assumed to be fixed for all k large enough. The value $\beta = 0$ is allowed; it corresponds to the classical choice of the Hessian of the standard Lagrangian: it gives $\tilde{\lambda}^k = \lambda^k$ and

$$(2.12) \qquad H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \tilde{\lambda}^k) = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k).$$

However, as is well known, the Hessian of the augmented Lagrangian has a better chance of being locally positive definite. For example, if

$$(2.13) \qquad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\},$$

then the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) + \beta (h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$ is positive definite for all β large enough. In such a case, the matrix H_k defined by (2.10) is also positive definite for all (x^k, λ^k) close enough to $(\bar{x}, \bar{\lambda})$. Condition (2.13) can be thought of as a strong second-order condition for problem (1.1) in the sense that it implies the classical second-order sufficient condition for (1.1). This is because the matrix of the quadratic form in (2.13) is identical to the Hessian of the Lagrangian of (1.1) due to linearity of the omitted constraints, but the set $\ker h'(\bar{x})$ in (2.13) is in general larger than the critical cone of (1.1).

The SQP subproblem (1.2) with the choice (2.10) of H_k essentially corresponds to the classical SQP subproblem for the penalized version of problem (1.1), namely,

$$(2.14) \qquad \begin{aligned} &\text{minimize} && f(x) + \tfrac{\beta}{2}\|h(x)\|^2 \\ &\text{subject to} && h(x) = 0, \quad x \geq 0. \end{aligned}$$

The Hessian of the Lagrangian of this subproblem contains the extra term of the form $\beta (h''(x)[\cdot])^{\mathrm{T}} h(x)$, which vanishes at any point satisfying the equality constraints and is dropped in the definition of H_k in (2.10). As for the multiplier update rule (2.11) (relating λ^k and $\tilde{\lambda}^k$), it is standard for SQP applied to problem (2.14).

Another interpretation of (2.10) and (2.11) is the following: these choices essentially correspond to the standard Newton method applied to the optimality system based on the augmented Lagrangian, namely,

$$\frac{\partial}{\partial x}\left(L(x, \lambda) + \frac{\beta}{2}\|h(x)\|^2\right) = 0, \quad h(x) = 0.$$

This Newton method was studied, e.g., in [22]. In particular, the multiplier update rule (2.11) corresponds to [22, relation (19)].

For local superlinear convergence we shall also impose the following truncation requirement in optimality conditions (2.1) of subproblems (1.2):

$$(2.15) \qquad \max\{\|\omega_1^k\|, \|\omega_2^k\|\} \leq \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)),$$

where $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$ is the natural residual of KKT conditions defined in (1.4) and $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$, $\varphi(t) = o(t)$, is a forcing function.

2.2. Solving QP subproblems by an IP algorithm. In this section, we summarize some necessary details concerning solving QP subproblems with the structure given in (1.2) by IP methods. In fact, for global convergence we only need the IP method to generate a sequence $\{(x^{k,\,i}, \lambda^{k,\,i}, \mu^{k,\,i})\}$ such that the associated sequence $\{(\hat{x}^{k,\,i}, \lambda^{k,\,i}, \hat{\mu}^{k,\,i})\}$ defined in Algorithm 2.1, if continued infinitely, would have an accumulation point which is a solution of the exact KKT system of the subproblem— that is, (2.1) with $\omega_1^k = 0$ and $\omega_2^k = 0$. If this is the case, the terms $\omega_1^{k,\,i}$ and $\omega_2^{k,\,i}$ can be reduced as needed along some convergent subsequence of $\{(\hat{x}^{k,\,i}, \lambda^{k,\,i}, \hat{\mu}^{k,\,i})\}$, while the corresponding subsequence of $\{p^{k,\,i}\}$ can tend to zero only if the current iterate x^k is a stationary point of problem (1.1). The latter is because this is the only case when the exact KKT system of the SQP subproblem may have a solution with the primal component equal to x^k. Thus, the conditions in (2.3) and (2.8)–(2.9) would eventually be satisfied, with (2.8)–(2.9) implying (2.4); see Proposition 2.1. Naturally, for local rate of convergence analysis some further estimates are needed. To be specific, we shall refer to the infeasible predictor-corrector algorithm given in [8, Algorithm 22.8]. However, most certainly, many other IP methods have the needed properties as well; see the references in [8, section 22.4].

For the current iteration $k = 0, 1, \ldots$, our QP subproblem (1.2) has the form

$$(2.16) \qquad \begin{aligned} &\text{minimize} \quad \langle a^k,\, x\rangle + \tfrac{1}{2}\langle H_k x,\, x\rangle \\ &\text{subject to} \quad h'(x^k)x = b^k, \quad x \geq 0, \end{aligned}$$

where $a^k = f'(x^k) - H_k x^k$, $b^k = -h(x^k) + h'(x^k)x^k$. As is standard in IP methods, it is assumed that the constraint matrix $h'(x^k)$ has full rank, which can be achieved by preprocessing. We also make the assumption that H_k is positive semidefinite, which is quite common as well. Stationary points and Lagrange multipliers of this problem are characterized by the KKT system

$$(2.17) \qquad \begin{aligned} a^k + H_k x + (h'(x^k))^{\mathrm{T}}\lambda - \mu = 0, \quad h'(x^k)x = b^k, \\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu,\, x\rangle = 0 \end{aligned}$$

with respect to $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$, which is (2.1) with $\omega_1^k = 0$ and $\omega_2^k = 0$. Let $\mathcal{S}_k$ stand for the solution set of (2.17).

IP methods are presented in [8] for a monotone linear complementarity problem (LCP) in variables (x, μ). But under our assumptions regarding H_k and $h'(x^k)$, the KKT system (2.17) can be reduced to a monotone LCP in (x, μ). Then λ would be uniquely defined by

$$\lambda = -(h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h'(x^k)(a^k + H_k x - \mu).$$

With these observations, we can employ the results of [8, Chapter 22], while keeping the discussion around the QP (2.16) and its KKT system (2.17).

Let a starting point $(x^{k,\,0},\,\lambda^{k,\,0},\,\mu^{k,\,0}) \in \mathbf{R}^n_{++} \times \mathbf{R}^l \times \mathbf{R}^n_{++}$ and a starting value $\sigma_{k,\,0} > 0$ of the path-following parameter σ be such that

$$\mu_j^{k,\,0} x_j^{k,\,0} = \sigma_{k,\,0}, \quad j = 1, \dots, n.$$

Set

$$\xi^k = \frac{1}{\sigma_{k,\,0}} (a^k + H_k x^{k,\,0} + (h'(x^k))^{\mathrm{T}} \lambda^{k,\,0} - \mu^{k,\,0}), \quad \eta^k = \frac{1}{\sigma_{k,\,0}} (h'(x^k) x^{k,\,0} - b^k).$$

Then the point $(x^{k,\,0},\,\lambda^{k,\,0},\,\mu^{k,\,0})$ satisfies the system

(2.18)
$$a^k + H_k x + (h'(x^k))^{\mathrm{T}} \lambda - \mu = \sigma \xi^k, \quad h'(x^k) x = b^k + \sigma \eta^k,$$
$$\mu > 0, \quad x > 0, \quad \mu_j x_j = \sigma, \quad j = 1, \dots, n,$$

for $\sigma = \sigma_{k,\,0}$. Note that for $\sigma = 0$, with strict inequalities replaced by the nonstrict ones, (2.18) turns into the KKT system (2.17) for the QP (2.16).

IP methods we refer to in this section employ Newtonian directions computed for the equations in (2.18). This implies that the linear equations in (2.18) remain satisfied along the trajectories $\{(\sigma_{k,\,i},\,x^{k,\,i},\,\lambda^{k,\,i},\,\mu^{k,\,i})\}$ generated in this way. In other words, the iterates belong to the *perturbed feasible set*

(2.19)

$$\mathcal{F}_k = \left\{ (\sigma,\,x,\,\lambda,\,\mu) \in \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \left| \begin{array}{l} \sigma > 0,\ x > 0,\ \mu > 0, \\ a^k + H_k x + (h'(x^k))^{\mathrm{T}} \lambda - \mu = \sigma \xi^k, \\ h'(x^k) x = b^k + \sigma \eta^k \end{array} \right. \right\}.$$

Define the *perturbed central path* by

$$\mathcal{C}_k = \{(\sigma,\,x,\,\lambda,\,\mu) \in \mathcal{F}_k \mid \mu_j x_j = \sigma,\ j = 1, \dots, n\},$$

and the *small neighborhood* of the perturbed central path by

(2.20) $$\mathcal{V}_k(\theta) = \{(\sigma,\,x,\,\lambda,\,\mu) \in \mathcal{F}_k \mid |\mu_j x_j - \sigma| \leq \theta \sigma,\ j = 1, \dots, n\},$$

where $\theta \in (0,\,1)$.

The IP algorithm [8, Algorithm 22.8] generates a sequence $\{(\sigma_{k,\,i}, x^{k,\,i}, \lambda^{k,\,i}, \mu^{k,\,i})\}$ that stays in $\mathcal{V}_k(\theta)$. According to [8, Theorem 22.9], if there exists a solution $(\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k)$ of system (2.17) such that $\bar{x}^k \leq x^{k,\,0}$ and $\bar{\mu}^k \leq \mu^{k,\,0}$, then the sequence $\{\sigma_{k,\,i}\}$ generated according to the rules in [8, Algorithm 22.8] converges to zero. Furthermore, according to [8, Lemma 22.1], the sequence $\{(x^{k,\,i},\,\lambda^{k,\,i},\,\mu^{k,\,i})\}$ is bounded. Hence, this sequence has accumulations points, and from (2.19) and (2.20) it follows that each of these accumulation points is a solution of (2.17). This is essentially all that is needed for establishing global convergence of our tSQP method.

Indeed, for each $i = 0, 1, \dots$, let $(\hat{x}^{k,\,i},\,\hat{\mu}^{k,\,i})$ be the "purified" estimates defined in step 2 of Algorithm 2.1 (see (2.2)). Then by the inclusion $(\sigma_{k,\,i},\,x^{k,\,i},\,\lambda^{k,\,i},\,\mu^{k,\,i}) \in \mathcal{V}_k(\theta)$ and by [8, Lemma 22.2], it holds that

(2.21) $$\hat{\mu}_{A_+^k}^{k,\,i} = \mu_{A_+^k}^{k,\,i}, \quad \|\hat{x}_{A_+^k}^{k,\,i} - x_{A_+^k}^{k,\,i}\| = \|x_{A_+^k}^{k,\,i}\| = O(\sigma_{k,\,i}),$$

(2.22) $$\|\hat{\mu}_{N^k}^{k,\,i} - \mu_{N^k}^{k,\,i}\| = \|\mu_{N^k}^{k,\,i}\| = O(\sigma_{k,\,i}), \quad \hat{x}_{N^k}^{k,\,i} = x_{N^k}^{k,\,i}$$

for all i large enough, and

(2.23) $\quad \|\hat{\mu}_{A_0^k}^{k,\,i} - \mu_{A_0^k}^{k,\,i}\| \leq \|\mu_{A_0^k}^{k,\,i}\| = O(\sigma_{k,\,i}^{1/2}), \quad \|\hat{x}_{A_0^k}^{k,\,i} - x_{A_0^k}^{k,\,i}\| \leq \|x_{A_0^k}^{k,\,i}\| = O(\sigma_{k,\,i}^{1/2}),$

where

$$A_+^k = \{j = 1, \ldots, n \mid \exists\, (\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k) \in \mathcal{S}_k \text{ such that } \bar{\mu}_j^k > 0\},$$

$$N^k = \{j = 1, \ldots, n \mid \exists\, (\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k) \in \mathcal{S}_k \text{ such that } \bar{x}_j^k > 0\},$$

$$A_0^k = \{1, \ldots, n\} \setminus (A_+^k \cup N^k).$$

(Lemma 22.2 in [8] deals with an LCP in the so-called *canonical form*, that is, with $A_+^k = \emptyset$. However, any LCP can be reduced to the canonical form by permutation of the respective components of $\mu_{A_0^k}$ and $x_{A_0^k}$, without affecting monotonicity. This observation combined with Lemma 22.2 in [8] gives the above-stated estimates.)

From (2.21), (2.22) we then obtain that

(2.24) $\quad \|\hat{x}^{k,\,i} - x^{k,\,i}\| = O(\sigma_{k,\,i}^{1/2}), \quad \|\hat{\mu}^{k,\,i} - \mu^{k,\,i}\| = O(\sigma_{k,\,i}^{1/2}).$

Since $(\sigma_{k,\,i}, x^{k,\,i}, \lambda^{k,\,i}, \mu^{k,\,i}) \in \mathcal{F}_k$, by (2.19) and (2.24) we then derive that

$$\begin{aligned}
0 &= a^k + H_k x^{k,\,i} + (h'(x^k))^{\mathrm{T}} \lambda^{k,\,i} - \mu^{k,\,i} - \sigma_{k,\,i} \xi^k \\
&= a^k + H_k \hat{x}^{k,\,i} + (h'(x^k))^{\mathrm{T}} \lambda^{k,\,i} - \hat{\mu}^{k,\,i} - \sigma_{k,\,i} \xi^k \\
&\quad + O(\|\hat{x}^{k,\,i} - x^{k,\,i}\|) + O(\|\hat{\mu}^{k,\,i} - \mu^{k,\,i}\|) \\
&= a^k + H_k \hat{x}^{k,\,i} + (h'(x^k))^{\mathrm{T}} \lambda^{k,\,i} - \hat{\mu}^{k,\,i} + O(\sigma_{k,\,i}^{1/2})
\end{aligned}$$

(2.25)

and

$$\begin{aligned}
0 &= h'(x^k) x^{k,\,i} - b^k - \sigma_{k,\,i} \eta^k \\
&= h'(x^k) \hat{x}^{k,\,i} - b^k - \sigma_{k,\,i} \eta^k + O(\|\hat{x}^{k,\,i} - x^{k,\,i}\|) \\
&= h'(x^k) \hat{x}^{k,\,i} - b^k + O(\sigma_{k,\,i}^{1/2}).
\end{aligned}$$

(2.26)

Hence, by the convergence of $\{\sigma_{k,\,i}\}$ to zero, we conclude that accumulation points of the "purified" sequence $\{(\hat{x}^{k,\,i}, \lambda^{k,\,i}, \hat{\mu}^{k,\,i})\}$ are the same as those of the IP sequence $\{(x^{k,\,i}, \lambda^{k,\,i}, \mu^{k,\,i})\}$, and that they are solutions of the KKT system (2.17). This justifies that truncation conditions (2.3) and (2.4) of Algorithm 2.1 hold after a finite number of purified IP iterations, where (2.4) holds at least when (2.8)–(2.9) are achieved; see section 3 for a formal statement. The local truncation condition (2.15) would eventually be satisfied for the same reasons.

We finish this discussion of IP methods in the context of our tSQP proposal with some observations which are not central to our analysis but are nevertheless worth making. If there exists a solution $(\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k)$ satisfying the strict complementarity condition $\bar{\mu}^k + \bar{x}^k > 0$, then the rate of convergence of $\{\sigma_{k,\,i}\}$ to zero is quadratic [8, Theorem 22.9]. Moreover, the estimates (2.24) can be improved as follows:

(2.27) $\quad \|\hat{x}^{k,\,i} - x^{k,\,i}\| = O(\sigma_{k,\,i}), \quad \|\hat{\mu}^{k,\,i} - \mu^{k,\,i}\| = O(\sigma_{k,\,i}).$

Also, if the estimates (2.25), (2.26) hold uniformly with respect to major iterations of the tSQP method, they allow one to prescribe a priori the level $\bar{\sigma}_k > 0$ to be achieved by $\sigma_{k,i}$ in order to satisfy the local truncation condition (2.15). Observe that (2.15) has the form

$$\max\{\|a^k + H_k\hat{x}^{k,i} + (h'(x^k))^{\mathrm{T}}\lambda^{k,i} - \hat{\mu}^{k,i}\|, \|h'(x^k)\hat{x}^{k,i} - b^k\|\} \le \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)).$$

Now, if we take $\bar{\sigma}_k = o((\rho(x^k, \tilde{\lambda}^k, \mu^k))^2)$, e.g., by setting $\bar{\sigma}_k = (\rho(x^k, \tilde{\lambda}^k, \mu^k))^q$ with $q > 2$, the condition above holds with function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ satisfying $\varphi(t) = o(t)$.

Observe also that the estimates (2.25), (2.26) are uniform, provided that the strict complementarity condition holds for the QP subproblems in a uniform way, that is, the numbers $\inf_{(\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k) \in \mathcal{S}_k} \max\{\bar{\mu}_j^k, \bar{x}_j^k\}$ are bounded away from zero (along the major iterations k). Moreover, in this case, employing the improved estimates in (2.27), one can take, a priori, $\bar{\sigma}_k = o(\rho(x^k, \tilde{\lambda}^k, \mu^k))$, e.g., by setting $\bar{\sigma}_k = (\rho(x^k, \tilde{\lambda}^k, \mu^k))^q$ with $q > 1$.

For other details such as modified fields aimed at improving the performance in the case when strict complementarity is violated, complexity analysis, etc., as well as for other IP methods for QP problems, see [8, Part IV] and the references therein. We believe that application of some of those other methods for solving subproblems within our truncated SQP framework can be analyzed along the same lines. Another important issue in this context is warmstarting, i.e., employing information from the previous major iteration to initialize a new set of minor iterations. Though this is known to not be an easy task in IP methods, some approaches to warmstarting have been suggested recently; see, e.g., [24] and a survey therein. Also, one may conjecture that inexact solution of subproblems can actually help to define good interior starting points for the next iteration.

3. Global convergence. We first show that Algorithm 2.1 is well defined, i.e., that after a finite number of inner iterations purified IP iterates produce a point satisfying the truncation conditions (2.3) and (2.4), unless the current iterate x^k is already a stationary point of problem (1.1).

PROPOSITION 3.1. *If x^k is not a stationary point of problem* (1.1), *then the purified IP iterates of Algorithm* 2.1, *generated according to section* 2.2, *stop after a finite number of steps.*

Proof. Recalling that H_k is positive definite, if x^k is not a stationary point of problem (1.1), then, as is well known, the unique solution and the unique stationary point of the SQP subproblem (1.2) is some $\bar{x}^k \ne x^k$. As justified in section 2.2, all accumulation points of IP iterates $\{(x^{k,i}, \lambda^{k,i}, \mu^{k,i})\}$, as well as those of the purified iterates $\{(\hat{x}^{k,i}, \lambda^{k,i}, \hat{\mu}^{k,i})\}$, are stationary point-multiplier triples of (1.2), which implies that $\{\hat{x}^{k,i}\} \to \bar{x}^k \ne x^k$, i.e., $\{p^{k,i}\} \to \bar{p}^k = \bar{x}^k - x^k \ne 0$, as $i \to \infty$. It follows that $(\omega_1^{k,i}, \omega_2^{k,i}) \to 0$ as $i \to \infty$, while the right-hand sides in all the inequalities in (2.3) and (2.8)–(2.9) are bounded away from zero (recall that $\{\lambda^{k,i}\}$, and hence $\{c_{k,i}\}$, are bounded, as discussed in section 2.2). Hence, these conditions would be satisfied for some iteration index i large enough, at which time (2.4) would hold as well, by Proposition 2.1. $\square$

We also need to rule out that $p^{k,i} = 0$ may satisfy the truncation conditions (2.3) and (2.4), as in such a case no progress would be made.

PROPOSITION 3.2. *If x^k is not a stationary point of problem* (1.1), *then Algorithm* 2.1 *generates some $p^k \ne 0$.*

Proof. Suppose $p^{k,i} = 0$ satisfies (2.3) and (2.4). Then (2.3) immediately gives that $\omega_1^{k,i} = 0$, and also from the second equation in (2.1) we have that $\omega_2^{k,i} = h(x^k)$.

Then (2.4) gives

$$-\bar{c}\|h(x^k)\|_1 \geq \Delta_{k,\,i} = -c_{k,\,i}(\|h(x^k)\|_1 - \|\omega_2^{k,\,i}\|_1) = 0,$$

so that $h(x^k) = 0$. Thus $\omega_2^{k,\,i} = h(x^k) = 0$. Together with $p^{k,\,i} = 0$ and $\omega_1^{k,\,i} = 0$, this means that (2.1) reduces to the exact KKT conditions for problem (1.1). □

From now on, it is assumed that x^k is not a stationary point of problem (1.1) for any index k, so that the algorithm generates nonzero directions p^k. We next show that the linesearch procedure is well defined.

PROPOSITION 3.3. *Let the derivatives of f and of h be Lipschitz-continuous. Then the linesearch procedure of Algorithm* 2.1 *is well defined and generates a stepsize $\alpha_k > 0$ after a finite number of backtrackings.*

Proof. Let $\ell > 0$ be a common Lipschitz constant for the gradients of f and of all components of h. By [4, Proposition A.24], for any $\alpha \in [0, 1]$ we have that

$$f(x^k + \alpha p^k) \leq f(x^k) + \alpha \langle f'(x^k), p^k \rangle + \frac{\ell \alpha^2}{2}\|p^k\|^2,$$

and similarly, for each $i = 1, \dots, l$,

$$
\begin{aligned}
|h_i(x^k + \alpha p^k)| &= |h_i(x^k + \alpha p^k) - \alpha(h_i(x^k) + \langle h_i'(x^k), p^k \rangle + (\omega_2^k)_i)| \\
&\leq |h_i(x^k + \alpha p^k) - h_i(x^k) - \alpha \langle h_i'(x^k), p^k \rangle| \\
&\quad + (1 - \alpha)|h_i(x^k)| + \alpha|(\omega_2^k)_i| \\
&\leq (1 - \alpha)|h_i(x^k)| + \alpha|(\omega_2^k)_i| + \frac{\ell \alpha^2}{2}\|p^k\|^2,
\end{aligned}
$$

where the second equality in (2.1) was used. Combining the two relations above, we obtain

$$
\begin{aligned}
\psi_{c_k}(x^k + \alpha p^k) &\leq f(x^k) + \alpha \langle f'(x^k), p^k \rangle + c_k(1 - \alpha)\|h(x^k)\|_1 + c_k\alpha\|\omega_2^k\|_1 \\
&\quad + \frac{\ell(1 + lc_k)\alpha^2}{2}\|p^k\|^2 \\
&= \psi_{c_k}(x^k) + \alpha \langle f'(x^k), p^k \rangle - \alpha c_k(\|h(x^k)\|_1 - \|\omega_2^k\|_1) \\
&\quad + \frac{\ell(1 + lc_k)\alpha^2}{2}\|p^k\|^2 \\
&= \psi_{c_k}(x^k) + \alpha\left(\Delta_k + \frac{\ell(1 + lc_k)\alpha}{2}\|p^k\|^2\right),
\end{aligned}
$$
(3.1)

where

$$\Delta_k = \langle f'(x^k), p^k \rangle - c_k(\|h(x^k)\|_1 - \|\omega_2^k\|_1)$$

is defined according to (2.5).

From (2.4) it follows that $\Delta_k < 0$. In particular, this shows that the linesearch rule (2.7) is satisfied whenever s is large enough so that for $\alpha = \tau^s$ it holds that

$$\Delta_k + \frac{\ell(1 + lc_k)\alpha}{2}\|p^k\|^2 \leq \varepsilon\Delta_k,$$

i.e,

$$\alpha \leq \frac{2(\varepsilon - 1)\Delta_k}{\ell + (1 + lc_k)\|p^k\|^2}.$$
(3.2)

The assertion follows. □

We conclude with a statement which shows that tSQP Algorithm 2.1 retains the usual global convergence properties of linesearch SQP methods. We make the assumption of boundedness of the generated sequence, which ensures that the penalty parameter c_k settles at some fixed value from some iteration on. This boundedness assumption, or some other related assumption with the same effect on c_k, is typical for SQP with linesearch (see, e.g., [8, Theorem 17.2], [3, Proposition 4.14(a)–(b)]), except for some more special problems and with some modifications to the algorithm (see, e.g., [46]).

THEOREM 3.4. *Let the derivatives of f and of h be Lipschitz-continuous. Let matrices H_k be uniformly bounded and uniformly positive definite; i.e, there exist γ_1, γ_2, $0 < \gamma_1 \leq \gamma_2$, such that*

$$(3.3) \qquad \gamma_1 \|\xi\|^2 \leq \langle H_k \xi, \, \xi \rangle \leq \gamma_2 \|\xi\|^2 \quad \forall \xi \in \mathbf{R}^n, \quad \forall k = 0, 1, \ldots.$$

Suppose, finally, that the sequence $\{(x^k, \, \lambda^k, \, \mu^k)\}$ generated by Algorithm 2.1 is bounded.

Then

$$(3.4) \qquad \begin{aligned} \left\{ \tfrac{\partial L}{\partial x}(x^k, \, \lambda^{k+1}) - \mu^{k+1} \right\} \to 0, & \quad \{h(x^k)\} \to 0, \\ \mu^{k+1} \geq 0, \quad x^k \geq 0, & \quad \langle \mu^{k+1}, \, x^k \rangle \to 0, \end{aligned}$$

and each accumulation point of $\{(x^k, \, \lambda^k, \, \mu^k)\}$ is a stationary point-multiplier triple of problem (1.1).

Proof. By (2.6), at each minor iteration either $c_{i, \, k} = c_{k-1}$ or $c_{i, \, k} > c_{k-1} + \theta_3 \bar{c} > c_{k-1}$. It follows that the sequence $\{c_k\} \subset \mathbf{R}_{++}$ is nondecreasing. If $c_k > c_{k-1}$, then the value of this parameter is increased by at least $\theta_3 \bar{c} > 0$. If this happens an infinite number of times, we have that $c_k \to +\infty$ and the second case in (2.6) then shows that $\|\lambda^k\| \to +\infty$, contradicting the boundedness of the generated sequence. It follows that $c_k = c > 0$ for all k sufficiently large, say $k \geq \bar{k}$ for some $\bar{k} \geq 0$.

Using (2.4), (3.2), and (3.3), we then have that for all $k \geq \bar{k}$ the linesearch rule (2.7) would be satisfied no later than $\alpha = \tau^s$ becomes such that

$$\alpha \leq \frac{2\theta_2(1 - \varepsilon)\langle H_k p^k, \, p^k \rangle}{\ell(1 + lc)\|p^k\|^2} \leq \frac{2\gamma_2 \theta_2(1 - \varepsilon)}{\ell(1 + lc)},$$

which by standard argument implies that there exists $\bar{\alpha} > 0$ such that

$$\alpha_k \geq \bar{\alpha} \quad \forall k \geq \bar{k}.$$

Then from (2.7) we have that

$$-\varepsilon \bar{\alpha} \Delta_k \leq -\varepsilon \alpha_k \Delta_k \leq \psi_c(x^k) - \psi_c(x^{k+1}).$$

In particular, since $\{\psi_c(x^k)\}$ is bounded below by the boundedness of $\{x^k\}$, we have that $\{\psi_c(x^k)\}$ converges, and it follows that

$$\Delta_k \to 0.$$

Taking into account (2.4) and (3.3), this implies that

$$\{p^k\} \to 0, \quad \{h(x^k)\} \to 0, \quad \{H_k p^k\} \to 0$$

and, taking into account (2.3), that

$$\{\omega_1^k\} \to 0.$$

Passing now to the limit in the approximate KKT system (2.1) with $(x, \lambda, \mu) = (x^k + p^k, \lambda^{k+1}, \mu^{k+1})$, we obtain the first two relations in (3.4).

Finally, by the construction of Algorithm 2.1, $\mu^{k+1} \geq 0$ and $x^k \geq 0$. Furthermore, by the construction of the purified sequence, $\langle \mu^{k+1}, \hat{x}^k \rangle = 0$. Therefore, using the boundedness of $\{\mu^k\}$, we obtain that

$$\langle \mu^{k+1}, x^k \rangle = \langle \mu^{k+1}, \hat{x}^k \rangle + \langle \mu^{k+1}, x^k - \hat{x}^k \rangle = -\langle \mu^{k+1}, p^k \rangle \to 0,$$

establishing the last relation in (3.4).

We conclude that each accumulation point of $\{(x^k, \lambda^{k+1}, \mu^{k+1})\}$ is a stationary point-multiplier triple of problem (1.1). Since $\{p^k\} \to 0$, if $\{x^{k_j}\} \to \bar{x}$ as $j \to \infty$, then also $\{x^{k_j+1}\} \to \bar{x}$. Hence, accumulation points of $\{(x^k, \lambda^k, \mu^k)\}$ are also stationary point-multiplier triples of problem (1.1).   □

4. Local convergence of the perturbed SQP framework and of tSQP and LCL methods. We start this section by considering the general constrained optimization problem

$$(4.1) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \ g(x) \leq 0, \end{array}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraint mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice-differentiable. The reason for stating some results for this general problem is that the perturbed SQP framework proposed in section 4.2 and the corresponding general convergence theorem, Theorem 4.3, have interesting applications beyond the tSQP method for problem (1.1) considered thus far. Section 4.1 contains a sensitivity result for the parametric version of problem (4.1), which is a combination of some known sensitivity results for optimization problems and optimality systems, and the local convergence result for the inexact Josephy–Newton method for an abstract generalized equation, considered in [33]. In section 4.2, these two basic facts will be used to derive a very general local convergence result for a certain perturbed SQP framework containing as particular cases various practical Newton-type schemes for problem (4.1). In section 4.3, this general result will serve for local convergence analysis of the tSQP algorithm proposed above. As an example of another application, in section 4.4 we briefly discuss the use of the perturbed SQP framework for local analysis of LCL methods.

Recall that stationary points of problem (4.1) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$(4.2) \qquad \frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0,$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (4.1):

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

For a given feasible point $\bar{x}$ of problem (4.1), we define the set of indices of inequality constraints active at $\bar{x}$:

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

The Mangasarian–Fromovitz constraint qualification (MFCQ) at $\bar{x}$ consists of saying that

$$\operatorname{rank} h'(\bar{x}) = l \text{ and } \exists \bar{\xi} \in \mathbf{R}^n \text{ such that } h'(\bar{x})\bar{\xi} = 0, \ g'_{A(\bar{x})}(\bar{x})\bar{\xi} < 0.$$

The strict Mangasarian–Fromovitz constraint qualification (SMFCQ) is said to be satisfied at a stationary point $\bar{x}$ of problem (4.1) if the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ is unique. Recall that the latter certainly implies MFCQ at $\bar{x}$, and is weaker than linear independence of gradients of active constraints (LICQ).

We will also make use of the following second-order sufficient optimality condition (SOSC):

$$(4.3) \qquad \left\langle \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\},$$

where

$$(4.4) \qquad C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \ g'_{A(\bar{x})}(\bar{x})\xi \leq 0, \ \langle f'(\bar{x}), \xi \rangle \leq 0\}$$

is the critical cone of problem (4.1) at $\bar{x}$. As is well known, if SOSC (4.3) holds at a stationary point $\bar{x}$ of problem (4.1) for some associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$, then $\bar{x}$ is a strict local minimizer of problem (4.1).

4.1. Auxiliary results. Consider the parametric mathematical programming problem

$$(4.5) \qquad \begin{array}{ll} \text{minimize} & f(\sigma, x) \\ \text{subject to} & h(\sigma, x) = 0, \ g(\sigma, x) \leq 0, \end{array}$$

where $\sigma \in \mathbf{R}^s$ is a parameter and $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$, $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$. Let $L : \mathbf{R}^s \times \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ be the Lagrangian of problem (4.5):

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

THEOREM 4.1. *Let the function $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$ and mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ possess the following properties:*
   (i) f, h, and g are twice-differentiable at $(\bar{\sigma}, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$.
   (ii) $f(\cdot, x)$, $h(\cdot, x)$, and $g(\cdot, x)$ are continuous at $\bar{\sigma}$ for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$.
   (iii) $f(\sigma, \cdot)$ is continuous near $\bar{x}$ for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$.
   (iv) h and g are differentiable with respect to x near $(\bar{\sigma}, \bar{x})$, $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ are continuous at $(\bar{\sigma}, \bar{x})$, and $\frac{\partial h}{\partial x}(\sigma, \cdot)$ and $\frac{\partial g}{\partial x}(\sigma, \cdot)$ are continuous near $\bar{x}$ for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$.
Let $\bar{x}$ be a local solution of problem (4.5) with $\sigma = \bar{\sigma}$, satisfying SMFCQ and SOSC of the form

$$(4.6) \qquad \left\langle \frac{\partial L}{\partial x}(\bar{\sigma}, \bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{\sigma}, \bar{x}) \setminus \{0\}$$

for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$, where

$$C(\bar{\sigma}, \bar{x}) = \left\{ \xi \in \mathbf{R}^n \ \middle| \ \frac{\partial h}{\partial x}(\bar{\sigma}, \bar{x})\xi = 0, \ \frac{\partial g_{A(\bar{x})}}{\partial x}(\bar{\sigma}, \bar{x})\xi = 0, \ \left\langle \frac{\partial h}{\partial x}(\bar{\sigma}, \bar{x}), \xi \right\rangle \leq 0 \right\}$$

is the critical cone of this problem at $\bar{x}$.

Then for any $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, problem (4.5) has a stationary point $x(\sigma)$ such that it is a local solution of this problem, and for any associated Lagrange multiplier $(\lambda(\sigma), \mu(\sigma))$ the following estimate holds:

$$(4.7) \qquad \|x(\sigma) - \bar{x}\| + \|\lambda(\sigma) - \bar{\lambda}\| + \|\mu(\sigma) - \bar{\mu}\| = O(\|\sigma - \bar{\sigma}\|).$$

Proof. As mentioned above, SOSC (4.6) implies that $\bar{x}$ is a strict local minimizer of problem (4.5) with $\sigma = \bar{\sigma}$, and SMFCQ subsumes MFCQ at $\bar{x}$. It then follows (e.g., from Robinson's stability theorem [43] and from [2, Theorem 3.1]) that for each $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, problem (4.5) has a local solution $x(\sigma)$ such that $x(\sigma) \to \bar{x}$ as $\sigma \to \bar{\sigma}$. Since MFCQ is stable subject to small perturbations (see, e.g., [9, Remark 2.88]), we conclude that for σ close enough to $\bar{\sigma}$, MFCQ holds at $x(\sigma)$, and hence, by the standard first-order necessary optimality conditions (see, e.g., [9, Theorem 3.9]), $x(\sigma)$ is a stationary point of problem (4.5) with some Lagrange multiplier $(\lambda(\sigma), \mu(\sigma))$. The estimate (4.7) now follows from [9, Theorem 5.9]. $\square$

As is well known, SQP and some other Newton-type methods for optimization are closely related to the Josephy–Newton method for generalized equations (GEs); see, e.g., [7, 33]. To that end, consider GEs of the form

$$(4.8) \qquad \Phi(u) + N(u) \ni 0,$$

where $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is a smooth single-valued mapping and N is a set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$. According to [7], solution $\bar{u}$ of GE (4.8) is said to be semistable if for any $r \in \mathbf{R}^n$ close enough to 0, any solution $u(r)$ of the perturbed GE

$$\Phi(u) + N(u) \ni r$$

close enough to $\bar{u}$ satisfies the estimate

$$\|u(r) - \bar{u}\| = O(\|r\|).$$

Following [33], we shall consider the inexact Josephy–Newton method with iteration subproblems of the form

$$(4.9) \qquad \Phi(u^k) + \Phi'(u^k)(u - u^k) + \Omega(u^k, u - u^k) + N(u) \ni 0,$$

where Ω is a set-valued mapping from $\mathbf{R}^\nu \times \mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$, characterizing inexactness. The next theorem was established in [33].

THEOREM 4.2. *Let a mapping $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ be differentiable in a neighborhood of $\bar{u} \in \mathbf{R}^\nu$, with its derivative being continuous at $\bar{u}$. Let $\bar{u}$ be a semistable solution of GE (4.8). Let $\Omega : \mathbf{R}^\nu \times \mathbf{R}^\nu \to 2^{\mathbf{R}^\nu}$ be a multifunction satisfying the following assumptions: for each $u \in \mathbf{R}^\nu$ close enough to $\bar{u}$, the GE*

$$(4.10) \qquad \Phi(u) + \Phi'(u)v + \Omega(u, v) + N(u + v) \ni 0$$

has a solution $v(u)$ such that $v(u) \to 0$ as $u \to \bar{u}$, and the estimate

$$(4.11) \qquad \omega = o(\|v\| + \|u - \bar{u}\|)$$

holds uniformly for $\omega \in \Omega(u, v)$, $u \in \mathbf{R}^\nu$, and $v \in \mathbf{R}^\nu$ satisfying

$$(4.12) \qquad \Phi(u) + \Phi'(u)v + \omega + N(u + v) \ni 0.$$

Then there exists $\delta > 0$ such that for any starting point $u^0 \in \mathbf{R}^\nu$ close enough to $\bar{u}$, there exists a trajectory $\{u^k\} \subset \mathbf{R}^\nu$ such that u^{k+1} is a solution of GE (4.9) for each $k = 0, 1, \ldots,$ satisfying

$$\|u^{k+1} - u^k\| \leq \delta;$$

any such trajectory converges to $\bar{u}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic, provided that the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{u}$, and provided that (4.11) can be replaced by the estimate

$$\omega = O(\|v\|^2 + \|u - \bar{u}\|^2).$$

As is well known, the KKT system (4.2) can be written as the GE (4.8) with the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

(4.13) $$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), g(x) \right),$$

and with

(4.14) $$N(u) = N(\mu) = \begin{cases} \{0\} \times \{0\} \times \{y \in \mathbf{R}_+^m \mid \langle \mu, y \rangle = 0\} & \text{if } \mu \geq 0, \\ \emptyset & \text{otherwise}, \end{cases}$$

where $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Then an iteration of pure SQP based on the Hessian of the Lagrangian is nothing more than (4.9) with $\Omega = \{0\}$.

For a stationary point $\bar{x}$ of problem (4.1) and for an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ the following key facts were established in [7]:

- If the solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE (4.8) with $\Phi(\cdot)$ and $N(\cdot)$ defined according to (4.13) and (4.14) is semistable, then $\bar{x}$ necessarily satisfies SMFCQ for $(\bar{\lambda}, \bar{\mu})$.
- If SMFCQ holds at $\bar{x}$ for $(\bar{\lambda}, \bar{\mu})$, and if SOSC (4.3) holds as well, then $\bar{u}$ is semistable.
- If $\bar{x}$ is a local solution of problem (4.1), then SOSC (4.3) is also necessary for the semistability of $\bar{u}$.

4.2. The perturbed SQP framework. Recall that for a current iterate $x^k \in \mathbf{R}^n$, the generic SQP subproblem for problem (4.1) has the form

(4.15)
$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), x - x^k \rangle + \frac{1}{2}\langle H_k(x - x^k), x - x^k \rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0, \quad g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array}$$

where $H_k \in \mathbf{R}^{n \times n}$ is some symmetric matrix. The basic SQP algorithm corresponds to the choice

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

The KKT system of problem (4.15) has the form

(4.17)
$$\begin{array}{c} f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) \leq 0, \quad \langle \mu, g(x^k) + g'(x^k)(x - x^k) \rangle = 0. \end{array}$$

By the *perturbed SQP* (pSQP) *framework*, we mean the following. 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 $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ must satisfy the following perturbed version of the KKT system (4.17) with H_k defined according to (4.16):

(4.18)

$$\frac{\partial L}{\partial x}(x^k, \lambda, \mu) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) + \omega_1^k = 0,$$
$$h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0,$$
$$\mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \leq 0, \quad \langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \rangle = 0,$$

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. It is important to emphasize that while perturbation in (4.18) can be induced by inexact/truncated solution of SQP subproblems, the framework is much more general. Specifically, perturbations in (4.18) may result from solving subproblems which are not even quadratic, as will be clear in what follows; see also [32, 33].

Any specific method within pSQP can be regarded as an inexact Josephy–Newton method for GE (4.8) with $\Phi(\cdot)$ and $N(\cdot)$ defined according to (4.13) and (4.14), respectively, as considered in [33]. We next establish local superlinear convergence of pSQP, based on Theorems 4.1 and 4.2, where the former will be used in order to establish solvability of subproblems. To that end, we first need to consider the pSQP iteration system (4.18) with more structured perturbation terms, corresponding to some optimization problem which can be tackled by Theorem 4.1. We start with the latter. Consider the iteration subproblem of the form

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

(4.19)
$$\text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k, \lambda^k, \mu^k), x - x^k) = 0,$$
$$g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \leq 0,$$

with some function $\psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}$ and mappings $\omega_2 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^l$ and $\omega_3 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^m$, which are assumed to be smooth with respect to the last variable.

Define the function $\Psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \to \mathbf{R}$:

(4.20)
$$\Psi((x, \lambda, \mu), (\xi, \eta, \zeta)) = \psi((x, \lambda, \mu), \xi) + \langle \lambda + \eta, \omega_2((x, \lambda, \mu), \xi) \rangle$$
$$+ \langle \mu + \zeta, \omega_3((x, \lambda, \mu), \xi) \rangle.$$

For $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, set

(4.21)
$$\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = \frac{\partial \Psi}{\partial \xi}((x, \lambda, \mu), (\xi, \eta, \zeta)).$$

Then we can write the KKT system of problem (4.19) as follows:

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) + (h'(x^k))^{\mathrm{T}} \lambda + (g'(x^k))^{\mathrm{T}} \mu$$
$$+ \omega_1((x^k, \lambda^k, \mu^k), (x - x^k, \lambda - \lambda^k, \mu - \mu^k)) = 0,$$

(4.22)
$$h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k, \lambda^k, \mu^k), x - x^k) = 0,$$
$$\mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \leq 0,$$
$$\langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \rangle = 0,$$

which corresponds to (4.18) with

$$(4.23) \qquad \omega_1^k = \omega_1((x^k, \lambda^k, \mu^k), (x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)),$$

$$(4.24) \qquad \omega_j^k = \omega_j((x^k, \lambda^k, \mu^k), x^{k+1} - x^k), \quad j = 2, 3.$$

In what follows, the terms defined by (4.23), (4.24) (employing (4.20), (4.21)) will correspond to *structural perturbations*, i.e., perturbations with respect to basic SQP, characterizing various specific Newton-type methods for problem (4.1). In particular, in addition to tSQP our framework includes quasi-Newton SQP, LCL methods [42, 39, 18], sequential quadratically constrained quadratic programming [1, 19, 45, 15], and stabilized SQP [48, 28, 49, 16]; see [33] and section 4.4 for details. Note that for some of those methods (4.19) is no longer a QP problem, and it is therefore even more natural to expect that the subproblems can be solved only approximately. For this reason, it makes sense to consider the scheme defined by (4.22) with additional perturbation terms that are induced by truncated solution of subproblems. Note that for the tSQP method proposed above, even though subproblems are QPs, this truncation is actually the essence of the corresponding algorithm. Also, if H_k is the exact Hessian of the Lagrangian, then truncation is actually the only source of perturbation. For methods other than SQP, this is not the case because of the presence of structural perturbations with respect to SQP itself.

To include truncation of subproblems within a pSQP framework, let the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ satisfy the following inexact version of system (4.22):

$$(4.25) \quad \begin{aligned} & \left\| f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu \right. \\ & \left. + \omega_1((x^k, \lambda^k, \mu^k), (x - x^k, \lambda - \lambda^k, \mu - \mu^k)) \right\| \le \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ & \|h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k, \lambda^k, \mu^k), x - x^k)\| \le \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ & \mu \ge 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \le 0, \\ & \langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \rangle = 0. \end{aligned}$$

Here, $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is some forcing function controlling truncation, and $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ is the natural residual of the KKT system (4.2) of the original problem (4.1):

$$(4.26) \qquad \rho(x, \lambda, \mu) = \left\| \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), \min\{\mu, -g(x)\} \right) \right\|.$$

We note that one can actually use any other residual which is locally Lipschitz-continuous with respect to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ under natural smoothness assumptions.

Note that the conditions in (4.25) that refer to the inequality constraints of (4.19) allow structural perturbation only and do not allow for any additional inexactness: these are precisely the corresponding conditions in (4.22).

For $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, set

$$(4.27) \quad \omega(u, v) = (\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)), \omega_2((x, \lambda, \mu), \xi), \omega_3((x, \lambda, \mu), \xi)),$$

$$(4.28) \qquad \Theta_1(u) = \{\theta_2 \in \mathbf{R}^n \mid \|\theta_2\| \le \varphi(\rho(x, \lambda, \mu))\},$$

$$(4.29) \qquad \Theta_2(u) = \{\theta_3 \in \mathbf{R}^l \mid \|\theta_3\| \le \varphi(\rho(x, \lambda, \mu))\},$$

$$(4.30) \qquad \Theta(u) = \Theta_1(u) \times \Theta_2(u) \times \{0\},$$

$$(4.31) \qquad \Omega(u, v) = \omega(u, v) + \Theta(u).$$

The system (4.25) can then be written as the iteration system of the inexact Josephy–Newton method considered in Theorem 4.2, that is, as the GE (4.9), where $u^k = (x^k, \lambda^k, \mu^k)$, and $\Phi(\cdot)$ and $N(\cdot)$ are defined according to (4.13) and (4.14), respectively.

Separating the perturbation into the single-valued part $\omega(u, v)$ and the set valued-part $\Theta(u, v)$ is instructive, because the two parts correspond to inexactness of different kinds: as explained above, $\omega(u, v)$ stands for structural perturbation with respect to the basic SQP, while $\Theta(u, v)$ stands for additional inexactness allowed when solving subproblems.

We next present a rather general local convergence result for the pSQP framework stated above. The statement of Theorem 4.3 below is rather cumbersome, and this theorem should of course be regarded as a technical tool, to be applied to some more specific algorithms. For example, the local superlinear convergence results for LCL methods [42, 39, 18], sequential quadratically constrained quadratic programming [1, 19, 45, 15], and stabilized SQP [48, 28, 49, 16], obtained in [33], can be derived in a simpler unified way by applying Theorem 4.3 (rather than Theorem 4.2, as done in [33]). Since in this work our focus is on another application of Theorem 4.3, namely, local convergence analysis of tSQP developed in section 2, we shall give full details for this method only. Nevertheless, to illustrate the generality of Theorem 4.3 a brief discussion of LCL methods is provided in section 4.4.

THEOREM 4.3. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^l$ be twice-differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.1), satisfying SMFCQ and SOSC (4.3) for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$. Furthermore, let the function $\psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}$ and mappings $\omega_2 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^l$ and $\omega_3 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^m$ possess the following properties:*

(i) *$\psi(\cdot, \xi)$, $\omega_2(\cdot, \xi)$, and $\omega_3(\cdot, \xi)$ are continuous at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ for all $\xi \in \mathbf{R}^n$ close enough to zero.*

(ii) *$\psi((x, \lambda, \mu), \cdot)$ is continuous near zero for all $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

(iii) *ψ, ω_2, and ω_3 are differentiable with respect to ξ near $((\bar{x}, \bar{\lambda}, \bar{\mu}), 0)$, the derivatives $\frac{\partial \omega_2}{\partial \xi}$ and $\frac{\partial \omega_3}{\partial \xi}$ are continuous at $((\bar{x}, \bar{\lambda}, \bar{\mu}), 0)$, and the mappings $\frac{\partial \omega_2}{\partial \xi}((x, \lambda, \mu), \cdot)$ and $\frac{\partial \omega_3}{\partial \xi}((x, \lambda, \mu), \cdot)$ are continuous near zero for all $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

(iv) *ψ, ω_2, and ω_3 are twice differentiable at $((\bar{x}, \bar{\lambda}, \bar{\mu}), 0)$.*

(v) *The equalities*

$$\omega_2((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0, \quad \omega_3((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0,$$

$$\frac{\partial \psi}{\partial \xi}((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0, \quad \frac{\partial \omega_2}{\partial \xi}((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0, \quad \frac{\partial \omega_3}{\partial \xi}((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0$$

hold, and for the function Ψ defined by (4.20), it holds that

$$(4.32) \qquad \left\langle \frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x}, \bar{\lambda}, \bar{\mu}), (0, 0, 0))\xi, \xi \right\rangle \geq 0 \quad \forall \xi \in C(\bar{x}),$$

where $C(\bar{x})$ is the critical cone of problem (4.1) at $\bar{x}$, defined in (4.4). Assume further that $\varphi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ is a function such that $\varphi(t) = o(t)$, and that the estimates

$$(4.33) \qquad \omega_j((x, \lambda, \mu), \xi) = o(\|\xi\| + \|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|), \quad j = 2, 3,$$

(4.34) $\dfrac{\partial \Psi}{\partial \xi}((x,\,\lambda,\,\mu),\,(\xi,\,\eta,\,\zeta)) = o(\|(\xi,\,\eta,\,\zeta)\| + \|(x - \bar{x},\,\lambda - \bar{\lambda},\,\mu - \bar{\mu})\|)$

hold for $(x,\,\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and for $(\xi,\,\eta,\,\zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to zero and satisfying

$$\left\| f'(x) + \dfrac{\partial^2 L}{\partial x^2}(x,\,\lambda,\,\mu)\xi + (h'(x))^{\mathrm{T}}(\lambda + \eta) + (g'(x))^{\mathrm{T}}(\mu + \zeta) \right.$$
$$\left. + \dfrac{\partial \Psi}{\partial \xi}((x,\,\lambda,\,\mu),\,(\xi,\,\eta,\,\zeta)) \right\| \le \varphi(\rho(x,\,\lambda,\,\mu)),$$

(4.35) $\|h(x) + h'(x)\xi + \omega_2((x,\,\lambda,\,\mu),\,\xi)\| \le \varphi(\rho(x,\,\lambda,\,\mu)),$

$\mu \ge 0, \quad g(x) + g'(x)\xi + \omega_3((x,\,\lambda,\,\mu),\,\xi) \le 0,$

$\langle \mu + \zeta,\, g(x) + g'(x)\xi + \omega_3((x,\,\lambda,\,\mu),\,\xi) \rangle = 0.$

Then there exists $\delta > 0$ such that for any starting point $(x^0,\,\lambda^0,\,\mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, there exists a trajectory $\{(x^k,\,\lambda^k,\,\mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each $k = 0, 1, \ldots$, the triple $(x^{k+1},\,\lambda^{k+1},\,\mu^{k+1})$ satisfies system (4.22) (and, in particular, satisfies system (4.25)) with ω_1 defined in (4.21) and Ψ defined in (4.20), and also satisfies the inequality

(4.36) $\|(x^{k+1} - x^k,\,\lambda^{k+1} - \lambda^k,\,\mu^{k+1} - \mu^k)\| \le \delta;$

any such trajectory converges to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic, provided that the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$, $\varphi(t) = O(t^2)$, and (4.33), (4.34) can be replaced by the estimates

(4.37) $\|\omega_j((x,\,\lambda,\,\mu),\,\xi)\| = O(\|\xi\|^2 + \|(x - \bar{x},\,\lambda - \bar{\lambda},\,\mu - \bar{\mu})\|^2), \quad j = 2,\,3,$

(4.38) $\dfrac{\partial \Psi}{\partial \xi}((x,\,\lambda,\,\mu),\,(\xi,\,\eta,\,\zeta)) = O(\|(\xi,\,\eta,\,\zeta)\|^2 + \|(x - \bar{x},\,\lambda - \bar{\lambda},\,\mu - \bar{\mu})\|^2).$

Proof. As discussed in section 4.1, under the assumptions of this theorem, the solution $\bar{u} = (\bar{x},\,\bar{\lambda},\,\bar{\mu})$ of GE (4.8) is automatically semistable.

We proceed with verifying the assumptions of Theorem 4.2. For a given $u = (x,\,\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, consider the system

$$f'(x) + \dfrac{\partial^2 L}{\partial x^2}(x,\,\lambda,\,\mu)\xi + (h'(x))^{\mathrm{T}}(\lambda + \eta) + (g'(x))^{\mathrm{T}}(\mu + \zeta)$$
$$+ \omega_1((x,\,\lambda,\,\mu),\,(\xi,\,\eta,\,\zeta)) = 0,$$

(4.39) $h(x) + h'(x)\xi + \omega_2((x,\,\lambda,\,\mu),\,\xi) = 0,$

$\mu + \zeta \ge 0, \quad g(x) + g'(x)\xi + \omega_3((x,\,\lambda,\,\mu),\,\xi) \le 0,$

$\langle \mu + \zeta,\, g(x) + g'(x)\xi + \omega_3((x,\,\lambda,\,\mu),\,\xi) \rangle = 0,$

with respect to $v = (\xi,\,\eta,\,\zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Since system (4.39) is equivalent to the GE

(4.40) $\Phi(u) + \Phi'(u)v + \omega(u,\,v) + N(u + v) \ni 0,$

and since any solution of (4.40) evidently satisfies (4.10) (see (4.28)–(4.31)), we need to show that for each $u \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $\bar{u}$, system (4.39) has a

solution $v(u) = (\xi(u), \eta(u), \zeta(u))$ which tends to zero as u tends to $\bar{u}$. To that end, consider the optimization problem

$$\text{minimize} \quad f(x) + \langle f'(x), \xi \rangle + \tfrac{1}{2} \left\langle \tfrac{\partial^2 L}{\partial x^2}(x, \lambda, \mu)\xi, \xi \right\rangle + \psi((x, \lambda, \mu), \xi)$$
(4.41) $\quad \text{subject to} \quad h(x) + h'(x)\xi + \omega_2((x, \lambda, \mu), \xi) = 0,$
$$g(x) + g'(x)\xi + \omega_3((x, \lambda, \mu), \xi) \leq 0,$$

with $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ regarded as a parameter with the base value $\bar{u}$. One can easily check that $\xi = 0$ is a stationary point of this problem for $u = \bar{u}$, and this point satisfies SMFCQ and SOSC for this problem with the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$. Under the stated assumptions, the needed assertion now readily follows from Theorem 4.1, taking into account that (4.39) with $\omega_1((x, \lambda, \mu), \cdot)$ defined according to (4.20), (4.21) is the KKT system of problem (4.41) (with the dual variables of the form $(\lambda + \eta, \mu + \zeta)$).

Assume now that $\omega \in \Omega(u, v)$, $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, and $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ satisfy (4.12) or, equivalently, satisfy (4.35) (see (4.21), (4.27)–(4.31)). Then

(4.42) $$\omega - \omega(u, v) \in \Theta(u)$$

(see (4.31)), and by the assumptions of the theorem, estimates (4.33) and (4.34) hold or, equivalently,

(4.43) $$\omega(u, v) = o(\|v\| + \|u - \bar{u}\|)$$

(see (4.21)).

Since $\varphi(t) = o(t)$, we have that

$$\varphi(\rho(x, \lambda, \mu)) = o(\rho(x, \lambda, \mu))$$
$$= o(\rho(x, \lambda, \mu) - \rho(\bar{x}, \bar{\lambda}, \bar{\mu}))$$
(4.44) $$= o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|),$$

where the last equality is by local Lipschitz-continuity of ρ with respect to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ (the latter can be easily shown by the mean-value theorem). Then from (4.28), (4.29) we derive that for each $\theta_2 \in \Theta_1(u)$ and $\theta_3 \in \Theta_2(u)$ it holds that

(4.45) $$\theta_2 = o(\|u - \bar{u}\|), \quad \theta_3 = o(\|u - \bar{u}\|).$$

Combining this with (4.43), and employing (4.30) and (4.42), we obtain the needed estimate (4.11).

Thus, all the assumptions of Theorem 4.2 are satisfied, and the needed result follows from this theorem.

Moreover, assuming that the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$, that the estimates (4.33) and (4.34) can be replaced by (4.37) and (4.38), and that $\varphi(t) = O(t^2)$, the right-hand sides of (4.43) and (4.45) can be replaced by $O(\|v\|^2 + \|u - \bar{u}\|^2)$ and $O(\|u - \bar{u}\|^2)$, respectively. Instead of (4.11), we then obtain a sharper estimate

$$\omega = O(\|v\|^2 + \|u - \bar{u}\|^2),$$

and the assertion regarding quadratic convergence rate follows from the corresponding assertion of Theorem 4.2. $\quad \square$

We note that, as can be seen from the proof, condition (4.32) in Theorem 4.3 can actually be relaxed as follows:

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

4.3. Local convergence of tSQP. In this section, we discuss the local convergence and rate of convergence properties of the tSQP algorithm developed in section 2. We shall derive our local convergence result for the augmented Lagrangian choice of H_k, given by (2.10), (2.11). But note that the value $\beta = 0$ corresponding to the usual Lagrangian is also allowed. As discussed in Remark 2.3, the Hessian of the augmented Lagrangian has a much better chance of being positive definite, which provides a good connection between our global and local convergence results. But, of course, positive definiteness is not needed for our local analysis itself. It is also worth mentioning that, for the purposes of this section, subproblems may also be solved by means other than the IP methods described in section 2.2, for example, if H_k is not positive semidefinite, as long as approximate solutions conforming to the truncation rule (2.15) are produced. For convenience, we shall restate the latter in the following form (recall also (2.1)):

$$
\begin{aligned}
&\left\| f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda - \mu \right\| \leq \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)), \\
(4.46) \qquad &\left\| h(x^k) + h'(x^k)(x - x^k) \right\| \leq \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)), \\
&\qquad \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0,
\end{aligned}
$$

where $\tilde{\lambda}^k$ is given by (2.11), $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is a forcing function, and $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$ is the natural residual defined in (1.4).

Observe that (4.46) is precisely the system (4.25) with $g : \mathbf{R}^n \to \mathbf{R}^n$, $g(x) = -x$, where

$$
\begin{aligned}
\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = \omega_1(x, \xi) &= \beta(h'(x))^{\mathrm{T}}(h(x) + h'(x)\xi), \\
\omega_2(\cdot) \equiv 0, &\quad \omega_3(\cdot) \equiv 0
\end{aligned}
$$

for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $\xi \in \mathbf{R}^n$. Moreover, (4.46) is the "truncated" optimality system for the pSQP subproblem (4.19), where one should take

$$\psi((x, \lambda, \mu), \xi) = \psi(x, \xi) = \frac{\beta}{2}\|h(x) + h'(x)\xi\|^2.$$

With these definitions of ψ, ω_1, ω_2, and ω_3, it holds that Ψ coincides with ψ (see (4.20)), and the equality (4.21) is valid.

Applying Theorem 4.3 now, we obtain a local convergence result for the full-step tSQP method, i.e., assuming that the unit stepsize is locally accepted (see, however, Remark 4.1).

THEOREM 4.4. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice-differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (1.1), satisfying SMFCQ and SOSC for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^n$. Assume further that $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is a function such that $\varphi(t) = o(t)$.*

Then for any fixed $\beta \geq 0$, there exists $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a trajectory $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ such that for each $k = 0, 1, \ldots$, the triple $(x^{k+1},$

$\lambda^{k+1}, \mu^{k+1})$ *satisfies (4.36) and (4.46) with H_k and $\tilde{\lambda}^k$ defined by (2.10) and (2.11), respectively; any such trajectory converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the trajectory $\{\tilde{\lambda}^k\}$ converges to $\bar{\lambda}$, and the rate of convergence of $\{(x^k, \tilde{\lambda}^k, \mu^k)\}$ is superlinear. Moreover, the rate of convergence of $\{(x^k, \tilde{\lambda}^k, \mu^k)\}$ is quadratic, provided that the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and provided that $\varphi(t) = O(t^2)$.*

Proof. The proof is by direct verification of the assumptions of Theorem 4.3. First, note that

$$\frac{\partial \psi}{\partial \xi}(\bar{x}, 0) = \omega_1(\bar{x}, 0) = 0.$$

Also, according to (4.4), for all $\xi \in C(\bar{x})$ the equality $h'(\bar{x})\xi = 0$ holds, and hence

$$\frac{\partial^2 \psi}{\partial \xi^2}(\bar{x}, 0)\xi = \beta(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})\xi = 0,$$

implying (4.32). Furthermore, if the second relation in (4.35) holds for some $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ with ω_2 identically equal to zero, and with φ satisfying $\varphi(t) = o(t)$, then by (4.44) it follows that

$$\begin{aligned}
\|h(x) + h'(x)\xi\| &\le \varphi(\rho(x, \lambda, \mu)) \\
(4.47) &= o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|).
\end{aligned}$$

This implies the estimate

$$\begin{aligned}
\frac{\partial \psi}{\partial \xi}(x, \xi) &= \omega_1(x, \xi) \\
&= \beta(h'(x))^{\mathrm{T}}(h(x) + h'(x)\xi) \\
(4.48) &= o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|),
\end{aligned}$$

which gives (4.34). Moreover, provided that $\varphi(t) = O(t^2)$, the estimate (4.47) can be sharpened as follows:

$$\|h(x) + h'(x)\xi\| = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|^2).$$

We then obtain the sharper version of estimate (4.48):

$$\frac{\partial \psi}{\partial \xi}(x, \xi) = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|^2),$$

which gives (4.38).

All the assertions now follow from Theorem 4.3. ∎

Remark 4.1. It should be noted that, as is well known, linesearch for a nonsmooth penalty function may not accept the unit stepsize even close to a solution with all the desirable properties (Maratos effect [37]). But this is another issue which is in essence unrelated to truncation as such and should be dealt with as in any other SQP method. Some options are second-order correction, watchdog steps, nonmonotone linesearch; see, e.g., [41]. In our local rate of convergence analysis, we therefore assume that the unit stepsize is locally accepted, by one device or another. That said, it is of some interest to point out that truncation can sometimes help to *prevent* the Maratos effect. Specifically, it may happen that while the unit stepsize along

the exact SQP direction is rejected, there are approximate directions that allow the unit stepsize to be accepted, yielding superlinear convergence. This can be observed, for example, considering the problem of minimizing $f(x) = x_1 + x_1^2 + x_2^2$ subject to $h(x) = x_1^2 + x_2^2 - 1 = 0$, which has the solution $\bar{x} = (-1, 0)$ satisfying LICQ and SOSC. Then the exact SQP step from any point $x^k \in \mathbf{R}^2$ such that $h(x^k) = 0$ and $x^k \neq \bar{x}$ fails to decrease the penalty function for any value of the penalty parameter. At the same time, there exist approximate SQP directions which admit unit stepsize and result in superlinear convergence. But, of course, we are not claiming that truncation is a guaranteed remedy for the Maratos effect.

4.4. Local convergence of LCL method. As another application of our general theorem, Theorem 4.3, in this section we briefly discuss local superlinear convergence of the LCL method; see [42, 39, 18]. For the current primal-dual iterate (x^k, λ^k, μ^k), the next primal iterate x^{k+1} of the LCL method is computed as a stationary point of the subproblem of minimizing the (augmented) Lagrangian subject to bounds and linearized equality constraints:

(4.49)
$$\begin{aligned}\text{minimize} \quad & L(x, \lambda^k) + \tfrac{c_k}{2}\|h(x)\|^2 \\ \text{subject to} \quad & h(x^k) + h'(x^k)(x - x^k) = 0, \quad x \geq 0,\end{aligned}$$

where $c_k \geq 0$ is the penalty parameter. The next dual iterate $(\lambda^{k+1}, \mu^{k+1})$ is of the form $(\lambda^k + \eta^k, \mu^{k+1})$, where (η^k, μ^{k+1}) is a Lagrange multiplier of the subproblem (4.49), associated to the stationary point x^{k+1}.

The LCL subproblem does not actually fit the assumptions of Theorem 4.3 directly, but the method can still be analyzed by considering its modification which generates the *same iterates*. To this end, assuming that $c_k = c$ for k large enough, which is standard for asymptotic analysis, consider the following modification of (4.49):

(4.50)
$$\begin{aligned}\text{minimize} \quad & L(x, \lambda^k) - \langle \lambda^k, h'(x^k)(x - x^k)\rangle + \tfrac{c}{2}\|h(x)\|^2 \\ \text{subject to} \quad & h(x^k) + h'(x^k)(x - x^k) = 0, \quad x \geq 0.\end{aligned}$$

Stationary points x^{k+1} of the two subproblems (4.49) and (4.50) coincide, and the associated multipliers are of the form (η^k, μ^{k+1}) and $(\lambda^{k+1}, \mu^{k+1})$ with $\lambda^{k+1} = \lambda^k + \eta^k$, respectively. Thus, if we prove convergence results for the method based on modified subproblems (4.50), they immediately translate into convergence results for LCL based on (4.49).

Observe that constraints of the modified LCL subproblem (4.50) are exactly the same as in the SQP subproblem (4.15). Structural perturbation that defines the method within the pSQP framework is therefore given by the objective function of the subproblem (4.50). Specifically, this subproblem (4.50) can be seen as a particular case of pSQP subproblem (4.19) where we take

$$\psi((x, \lambda, \mu), \xi) = L(x + \xi, \lambda) - \langle \lambda, h'(x)\xi\rangle + \frac{c}{2}\|h(x + \xi)\|^2$$

$$- f(x) - \langle f'(x), \xi\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi, \xi \right\rangle,$$

$$\omega_2((x, \lambda, \mu), \xi) = 0, \quad \omega_3((x, \lambda, \mu), \xi) = 0.$$

Applying Theorem 4.3, we easily obtain the following theorem.

THEOREM 4.5. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice-differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local*

solution of problem (1.1), *satisfying SMFCQ and SOSC for the associated Lagrange multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^n$.

Then for any fixed $c \geq 0$, there exists $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ such that for each $k = 0, 1, \ldots$, the point x^{k+1} is a stationary point of problem (4.49) *and $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is an associated Lagrange multiplier satisfying* (4.36); *any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic, provided the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$.*

We note that other superlinear convergence results in the literature on LCL require the stronger LICQ instead of SMFCQ, as well as strict complementarity [42, 18]. The result stated in Theorem 4.5 was first directly derived in [33]. Applying Theorem 4.3 gives an easier path.

Moreover, taking into account that the LCL subproblem (4.49) is not a QP, it makes good sense to consider its approximate solution. In accordance with our pSQP framework, we can introduce for this purpose the extra perturbation associated to inexact solution of subproblems (truncation). Local convergence of this truncated version of LCL can be analyzed along the same lines as was done above for tSQP.

5. Numerical results. The purpose of our experiments is to confirm that one can expect to achieve some gain using truncation (tSQP), as compared to solving subproblems "exactly" (SQP). No attempt is made to compare either of the algorithms with other solvers. Moreover, our implementation of SQP itself, and therefore of tSQP, is rather basic. Nevertheless, we believe some cautious conclusions can be drawn, as one can expect that benefits of more sophisticated SQP implementations should improve in a similar way the corresponding tSQP modification.

Specifically, we compare the following three algorithms:

- SQP, that is, Algorithm 2.1 with the IP iterations stopped not according to truncation conditions but by the internal stopping rule of the IP solver, uniformly along the SQP iterations;
- tSQP with global truncation rules (2.3) and (2.4), with the following modification: truncation is switched off if it happens that the IP solver is unable to satisfy the truncation tests while its internal stopping test can be achieved;
- tSQP with global and local truncation rules, that is, the same algorithm as the previous algorithm but with global truncation conditions (2.3) and (2.4) complemented by the local truncation condition (2.15) with $\varphi(t) = t^{1.2}$.

Our strategy for switching the truncation off is quite natural. When the internal stopping rule of the IP solver is satisfied, this indicates that the IP solver is unlikely to achieve any further significant progress. In particular, we cannot expect that it would manage to satisfy the truncation conditions which were not achieved up to that point. By stopping the truncation and accepting the generated direction, we prevent the algorithm from the almost inevitable immediate failure and give it a chance to proceed as a basic SQP algorithm. We note that switching the truncation off is not needed very often, and, when needed, it usually happens at the end of the run when significant benefit from truncating has already been accumulated.

For our test set, we used equality-constrained problems, with or without bounds, from the Hock and Schittkowski collection [31] modified in the following way. If the solution of a given problem is not nonnegative, we transform the variables by shifting the negative components to enforce nonnegativity, and then add bound constraints to the problem. If the solution is nonnegative as is, we just add bounds. This gave 24 test problems. From this set we removed `hs8`, `hs40`, `hs63`, and `hs78` for the
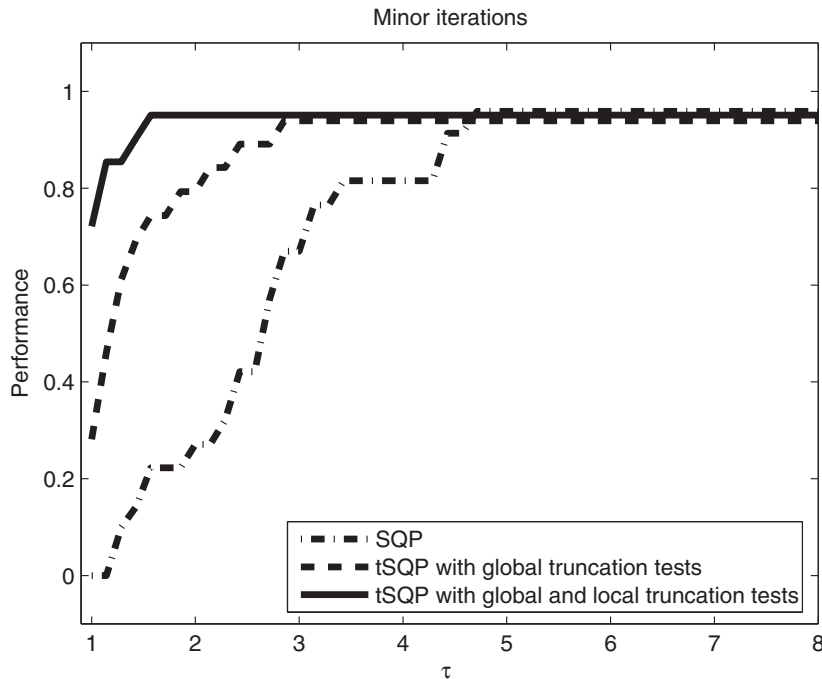
FIG. 5.1. *Performance profile.*

following reasons. Test `hs8` is a feasibility problem, and both SQP and tSQP behave identically and do not require any minor iterations. For `hs40`, `hs63`, and `hs78`, our implementations of both SQP and tSQP fail from too many starting points (because of infeasible subproblems, etc.), and not enough meaningful information is accumulated. Thus, we ended up with 20 test problems.

The experiments were performed in the MATLAB environment. We used the IP QP solver `spsolqp.m` listed in http://plato.asu.edu/sub/nlores.html#QP-problem, adapted to our purposes. We had to reduce the internal stopping tolerance of this solver from the default 1e-5 to 1e-8, because the default was not enough for robustness of our SQP code. Both SQP and tSQP use H_k defined by BFGS updates with Powell's modification (see, e.g., [41, p. 538]). The key truncation parameters in (2.3) and (2.4) were set at $\theta_1 = 1000$ and $\theta_2 = 0.5$. The other parameters of Algorithm 2.1 were as follows: $\theta_3 = 1.1$, $c_{-1} = 1$, $\bar{c} = 1$, $\varepsilon = 0.1$, $\tau = 0.5$. For each test problem, we performed 100 runs from randomly generated starting points (the same for all three algorithms) in the cube centered at the solution; the edge of the cube is the doubled distance from the starting point suggested in [31] to the solution. Failures were declared when the accuracy 1e-4 in the residual of the KKT system was not achieved after 500 iterations or when a given method failed to complete an iteration, for whatever reason. Other runs were regarded as successful.

In what follows we refer to the iterations of the SQP or tSQP method as "major iterations" and to the iterations of the IP QP solver, which is run for each SQP subproblem, as the "minor iterations" of the overall method. Since once a direction is computed by the IP QP solver, the structure of major iterations of SQP and tSQP is the same, the total number of minor iterations along the run until termination seems to be one relevant measure for comparing performance. Figure 5.1 reports on the

TABLE 5.1
The ratios between numbers of minor iterations of the IP QP solver needed by SQP and tSQP with global and local truncation tests to reduce the initial KKT residual by the given factor.

Problem	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}
hs6	2.5	2.5	3.0	2.0	2.3	2.0	1.7	1.7	1.6	1.5
hs7	8.0	7.0	7.0	5.9	5.1	5.0	4.7	4.6	4.4	4.3
hs9	6.0	6.5	4.3	4.7	3.8	3.8	3.2	2.8	2.9	2.9
hs26	5.0	5.7	4.5	5.0	4.4	4.2	3.5	2.8	2.5	2.0
hs27	12.3	6.3	6.6	5.5	5.6	5.3	4.8	4.9	4.6	4.5
hs28	7.0	6.2	5.7	5.0	4.1	3.6	3.5	2.6	2.3	1.9
hs39	9.0	9.5	7.7	6.8	6.0	5.5	4.7	4.2	3.9	3.6
hs42	5.0	4.8	4.4	4.2	3.3	3.1	2.9	2.7	2.6	2.5
hs47	5.3	5.8	5.0	5.4	4.3	4.3	3.9	3.7	3.6	3.2
hs48	6.0	7.0	5.8	5.0	5.4	4.8	5.0	4.4	4.0	4.0
hs49	7.0	7.5	5.3	6.0	5.0	5.5	4.8	4.2	3.5	3.1
hs50	9.0	9.0	9.5	6.7	7.3	6.0	6.3	5.4	5.6	4.8
hs51	5.3	6.0	5.0	5.3	4.4	4.6	4.0	4.0	3.6	3.1
hs52	6.6	6.0	5.7	5.3	4.8	4.4	4.2	3.9	3.2	2.9
hs53	7.8	7.2	6.3	6.1	5.8	4.9	4.3	3.6	3.1	2.7
hs56	5.0	4.5	2.5	2.2	1.8	1.6	1.5	1.4	1.5	1.8
hs60	5.0	5.7	5.0	4.4	4.2	3.9	3.6	3.3	3.2	3.0
hs61	7.8	7.0	6.8	5.6	4.4	4.5	4.1	3.6	3.9	3.8
hs77	6.0	5.7	5.4	4.2	3.7	2.5	1.6	1.5	1.5	1.4
hs79	5.7	5.0	5.5	4.8	3.7	3.2	2.9	2.8	2.8	2.5

total number of minor iterations, in the form of a performance profile [14]. For each algorithm, the value of the plotted function at $\tau \in [1, +\infty)$ corresponds to the part of runs for which the achieved result (the minor iteration count) was no more than τ times worse (bigger) than the best result. Of course, failure is regarded as infinitely many times worse than the best result. As can be seen, the truncated versions are generally more efficient and not less robust. Moreover, Figure 5.1 demonstrates the gain of using the local truncation condition as a complement of the global truncation conditions.

Another relevant piece of information is how many minor iterations a method needs to reduce the KKT residual by some factor. Table 5.1 reports ratios of the average numbers of minor iterations needed by SQP and tSQP employing both global and local truncation conditions to reduce the initial KKT residual by factors from 2 to 2^{10}. These averages were computed only over the runs for which the required accuracies were actually achieved. Thus, Table 5.1 gives some "qualitative" impression of the gain of truncation: it shows by how much the average number of minor iterations required by SQP to achieve a given accuracy exceeds the corresponding number for tSQP. Note that the ratios in Table 5.1 usually reduce as we approach a solution. This is only natural, since every time tighter and tighter tolerances would make the IP QP solver take the number of minor iterations closer to the "exact" method.

Overall, we believe that our experiments show the potential benefits of solving SQP subproblems inexactly, especially on early iterations.

REFERENCES

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

[2] A. V. Arutyunov and A. F. Izmailov, *Sensitivity analysis for cone-constrained optimization problems under the relaxed constraint qualifications*, Math. Oper. Res., 30 (2005), pp. 333–353.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[5] B. T. Boggs and J. W. Tolle, *Sequential quadratic programming*, in Acta Numerica, 1995, Acta Numer., Cambridge University Press, Cambridge, UK, 1995, pp. 1–51.

[6] B. T. Boggs, J. W. Tolle, and A. J. Kearsley, *A truncated SQP algorithm for large scale nonlinear programming problems*, in Advances in Optimization and Numerical Analysis, Math. Appl. 275, S. Gomez and J.-P. Hennart, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994, pp. 69–77.

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

[8] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer-Verlag, Berlin, 2006.

[9] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[10] R. H. Byrd, F. E. Curtis, and J. Nocedal, *An inexact Newton method for nonconvex equality constrained optimization*, Math. Program., 122 (2010), pp. 273–299.

[11] R. H. Byrd, F. E. Curtis, and J. Nocedal, An *inexact SQP method for equality constrained optimization*, SIAM J. Optim., 19 (2008), pp. 351–369.

[12] R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.

[13] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[14] E. Dolan and J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.

[15] 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.

[16] D. Fernández and M. Solodov, *Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems*, Math. Program., to appear.

[17] R. Fontecilla, *On inexact quasi-Newton methods for constrained optimization*, in Numerical Optimization 1984, Proceedings of the SIAM Conference on Numerical Optimization, SIAM, Philadelphia, 1985, pp. 102–118.

[18] 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.

[19] 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.

[20] P. E. Gill, W. Murray, and M. A. Saunders, *User's Guide for SNOPT* 5.3*: A Fortran Package for Large-Scale Nonlinear Programming*, Report NA 97-5, Department of Mathematics, University of California, San Diego, 1997.

[21] P. E. Gill, W. Murray, and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM J. Optim., 12 (2002), pp. 979–1006.

[22] S. T. Glad, *Properties of updating methods for the multipliers in augmented Lagrangian*, J. Optim. Theory Appl., 28 (1979), pp. 135–156.

[23] M. J. Goldsmith, *Sequential Quadratic Programming Methods Based on Indefinite Hessian Approximations*, Ph.D. thesis, Department of Management Science and Engineering, Stanford University, Stanford, CA, 1999.

[24] J. Gondzio and A. Grothey, *A new unblocking technique to warmstart interior point methods based on sensitivity analysis*, SIAM J. Optim., 19 (2008), pp. 1184–1210.

[25] N. I. M. Gould, *Some Reflections on the Current State of Active-set and Interior-point Methods for Constrained Optimization*, Numerical Analysis Group Internal Report 2003-1, Computational Science and Engineering Department, Rutherford Appleton Laboratory, Oxfordshire, UK, 2003.

[26] N. I. M. GOULD, D. ORBAN, AND PH. L. TOINT, *Numerical methods for large-scale nonlinear optimization*, Acta Numer., 14 (2005), pp. 299–361.

[27] N. I. M. GOULD AND PH. L. TOINT, *SQP methods for large-scale nonlinear programming*, in System Modeling and Optimization: Methods, Theory and Applications, M. J. D. Powell and S. Scholtes, eds., Kluwer, Dordrecht, The Netherlands, 2000, pp. 149–178.

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

[29] M. HEINKENSCHLOSS AND D. RIDZAL, *An inexact trust-region SQP method with applications to PDE-constrained optimization*, in Numerical Mathematics and Advance Applications, Proceedings of the 7th European Conference on Numerical Mathematics and Advanced Applications, Graz, Austria, 2007, K. Kunisch, O. Steinbach, and G. Of, eds., Springer-Verlag, Heidelberg, 2008, pp. 613–620.

[30] M. HEINKENSCHLOSS AND L. N. VICENTE, *Analysis of inexact trust-region SQP algorithms*, SIAM J. Optim., 12 (2001), pp. 283–302.

[31] W. HOCK AND K. SCHITTKOWSKI, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Econom. and Math. Systems 187, Springer-Verlag, Berlin, New York, 1981.

[32] 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., to appear.

[33] 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., to appear.

[34] H. JÄGER AND E. W. SACHS, *Global convergence of inexact reduced SQP methods*, Optim. Methods Softw., 7 (1997), pp. 83–110.

[35] M. LALEE, J. NOCEDAL, AND T. PLANTENGA, *On the implementation of an algorithm for large-scale equality constrained optimization*, SIAM J. Optim., 8 (1998), pp. 682–706.

[36] F. LEIBFRITZ AND E. W. SACHS, *Inexact SQP interior point methods and large scale optimal control problems*, SIAM J. Control Optim., 38 (1999), pp. 272–293.

[37] N. MARATOS, *Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems*, Ph.D. thesis, University of London, London, UK, 1978.

[38] W. MURRAY AND F. J. PRIETO, *A sequential quadratic programming algorithm using an incomplete solution of the subproblem*, SIAM J. Optim., 5 (1995), pp. 590–640.

[39] 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.

[40] B. A. MURTAGH AND M. A. SAUNDERS, *MINOS 5.0 User's Quide*, Technical report SOL 83.20, Stanford University, Stanford, CA, 1983.

[41] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[42] S. M. ROBINSON, *A quadratically convergent algorithm for general nonlinear programming problems*, Math. Program., 3 (1972), pp. 145–156.

[43] S. M. ROBINSON, *Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[44] C. A. SAGASTIZÁBAL AND M. V. SOLODOV, *Parallel variable distribution for constrained optimization*, Comput. Optim. Appl., 22 (2002), pp. 111–131.

[45] M. V. SOLODOV, *On the sequential quadratically constrained quadratic programming methods*, Math. Oper. Res., 29 (2004), pp. 64–79.

[46] M. V. SOLODOV, *Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences*, Math. Program., 118 (2009), pp. 1–12.

[47] S. ULBRICH, *Generalized SQP-methods with "parareal" time-domain decomposition for time-dependent PDE-constrained optimization*, in Real-Time PDE-Constrained Optimization, L. T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, eds., SIAM, Philadelphia, 2007, pp. 145–168.

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

[49] S. J. WRIGHT, *Modifying SQP for degenerate problems*, SIAM J. Optim., 13 (2002), pp. 470–497.