

Stabilized SQP revisited

A. F. Izmailov & M. V. Solodov

Mathematical Programming

A Publication of the Mathematical
Optimization Society

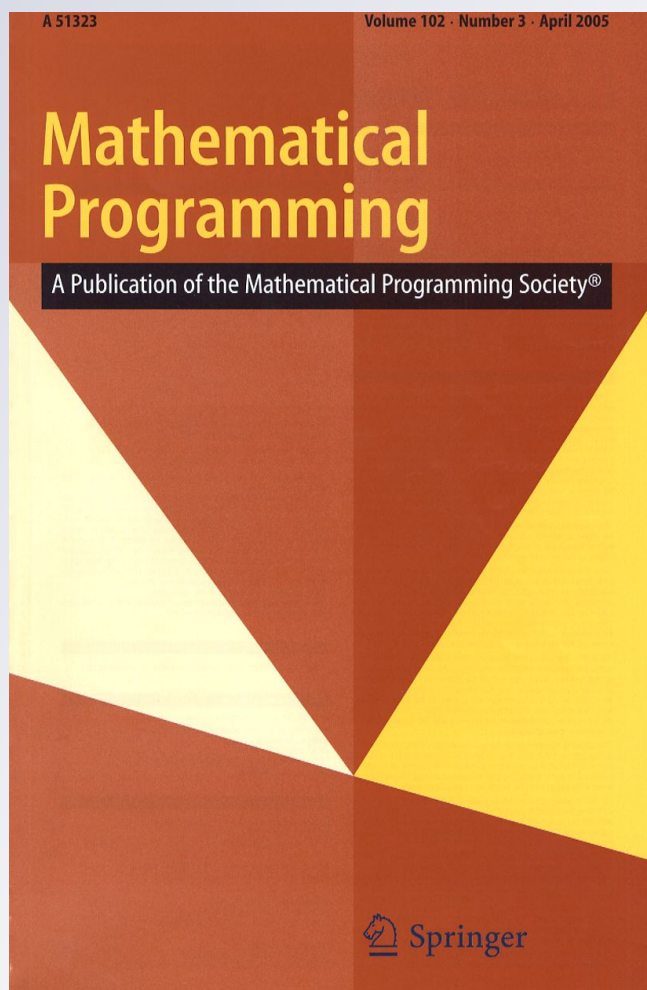
ISSN 0025-5610

Volume 133

Combined 1-2

Math. Program. (2012) 133:93-120

DOI 10.1007/s10107-010-0413-3



Your article is protected by copyright and all rights are held exclusively by Springer and Mathematical Optimization Society. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Stabilized SQP revisited

A. F. Izmailov · M. V. Solodov

Received: 9 September 2009 / Accepted: 11 August 2010 / Published online: 26 September 2010
© Springer and Mathematical Optimization Society 2010

Abstract The stabilized version of the sequential quadratic programming algorithm (sSQP) had been developed in order to achieve superlinear convergence in situations when the Lagrange multipliers associated to a solution are not unique. Within the framework of Fischer (Math Program 94:91–124, 2002), the key to local superlinear convergence of sSQP are the following two properties: upper Lipschitzian behavior of solutions of the Karush-Kuhn-Tucker (KKT) system under canonical perturbations and local solvability of sSQP subproblems with the associated primal-dual step being of the order of the distance from the current iterate to the solution set of the unperturbed KKT system. According to Fernández and Solodov (Math Program 125:47–73, 2010), both of these properties are ensured by the second-order sufficient optimality condition (SOSC) without any constraint qualification assumptions. In this paper, we state precise relationships between the upper Lipschitzian property of solutions of KKT systems, error bounds for KKT systems, the notion of critical Lagrange multipliers (a subclass of multipliers that violate SOSC in a very special way), the second-order necessary condition for optimality, and solvability of sSQP subproblems. Moreover,

Research of the first author is supported by the Russian Foundation for Basic Research Grant 10-01-00251, and by RF President's Grant NS-693.2008.1 for the support of leading scientific schools. The second author is supported in part by CNPq Grants 300513/2008-9 and 471267/2007-4, by PRONEX–Optimization, and by FAPERJ.

A. F. Izmailov

Department of Operations Research, Faculty of Computational Mathematics and Cybernetics,
Moscow State University, Leninskiye Gory, GSP-2, 119992 Moscow, Russia
e-mail: izmaf@ccas.ru

M. V. Solodov (✉)

IMPA, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110,
Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: solodov@impa.br

for the problem with equality constraints only, we prove superlinear convergence of sSQP under the assumption that the dual starting point is close to a noncritical multiplier. Since noncritical multipliers include all those satisfying SOSC but are not limited to them, we believe this gives the first superlinear convergence result for any Newtonian method for constrained optimization under assumptions that do not include any constraint qualifications and are weaker than SOSC. In the general case when inequality constraints are present, we show that such a relaxation of assumptions is not possible. We also consider applying sSQP to the problem where inequality constraints are reformulated into equalities using slack variables, and discuss the assumptions needed for convergence in this approach. We conclude with consequences for local regularization methods proposed in (Izmailov and Solodov *SIAM J Optim* 16:210–228, 2004; Wright *SIAM J. Optim.* 15:673–676, 2005). In particular, we show that these methods are still locally superlinearly convergent under the noncritical multiplier assumption, weaker than SOSC employed originally.

Keywords Constrained optimization · Degenerate constraints · Second-order sufficiency · Critical multipliers · Error bound · Stabilized SQP · Superlinear convergence · Local regularization

Mathematics Subject Classification (2000) 90C30 · 90C33 · 90C55 · 65K05

1 Introduction

Consider the mathematical programming (MP) problem

$$\begin{aligned} & \text{minimize}_x f(x) \\ & \text{subject to } h(x) = 0, \quad g(x) \leq 0, \end{aligned} \tag{1}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is a smooth function and $h : \mathbf{R}^n \rightarrow \mathbf{R}^l$ and $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$ are smooth mappings. Specifically, we assume that f, h and g are twice differentiable near the point of interest $\bar{x} \in \mathbf{R}^n$, and their second derivatives are continuous at \bar{x} .

Stationary points of problem (1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0, \tag{2}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}$ is the Lagrangian of problem (1), i.e.,

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

We denote by $\mathcal{M} = \mathcal{M}(\bar{x})$ the set of Lagrange multipliers associated with $\bar{x} \in \mathbf{R}^n$, that is, the pairs $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying (2) for $x = \bar{x}$. Thus \bar{x} is a stationary point of problem (1) if $\mathcal{M} \neq \emptyset$. We say that for a given stationary point \bar{x} of problem (1) and for an associated pair $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$ the second-order sufficient optimality condition

(SOSC) holds if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C \setminus \{0\}, \tag{3}$$

where

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

is the critical cone of problem (1) at \bar{x} , with

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

being the set of indices of active constraints at \bar{x} and

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

being the sets of indices of strongly active and weakly active constraints, respectively.

Given the current primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, stabilized sequential quadratic programming method (sSQP, see [9, 11, 12, 24, 25]) generates the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ as a stationary point of the quadratic programming (QP) subproblem in the primal-dual space:

$$\begin{aligned} &\text{minimize}_{(x, \lambda, \mu)} \left\{ \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) (x - x^k), x - x^k \right\rangle \right. \\ &\quad \left. + \frac{\sigma_k}{2} (\|\lambda\|^2 + \|\mu\|^2) \right\} \\ &\text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0, \\ &\quad g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0, \end{aligned} \tag{4}$$

where $\sigma_k > 0$ is the dual stabilization parameter, the choice of which is based on computing the violation of the KKT optimality conditions (2) by the point (x^k, λ^k, μ^k) . To be specific, in what follows we choose in (4) $\sigma_k = \sigma(x^k, \lambda^k, \mu^k)$, where $\sigma : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}_+$ is the natural residual of the KKT system (2), i.e.,

$$\sigma(x, \lambda, \mu) = \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\|, \tag{5}$$

with the minimum applied componentwise.

An iteration of the classical SQP method [1] can be formally thought as solving (4) with $\sigma_k = 0$. Thus minimization in the SQP subproblem is in primal variables,

while dual variables are given by multipliers associated to the primal solution. We emphasize that the sharpest superlinear convergence result for SQP [2] assumes that $\mathcal{M} = \{(\bar{\lambda}, \bar{\mu})\}$ (i.e., that there exists the unique multiplier associated to \bar{x} , also known as the strict Mangasarian–Fromovitz constraint qualification [20], SMFCQ) and that SOSC (3) holds. In particular, superlinear convergence of SQP requires certain regularity of constraints (specifically, SMFCQ, which implies the classical Mangasarian–Fromovitz constraint qualification [22], MFCQ). To deal with the case of possibly degenerate constraints, when multipliers associated to the solution are not unique, the stabilized version of SQP was first proposed in [24] (for inequality-constrained problems). It was further demonstrated in [21] that the min-max sSQP subproblems suggested in [24] are equivalent to QP subproblems with respect to the primal-dual variables stated in (4). Numerical experiments with sSQP are reported in [23]. Some results on sSQP still required at least MFCQ [24–26] or a strong version of SOSC [11, 12]. Recently [9], superlinear convergence of sSQP had been shown in a more general framework of a stabilized Newton-type method for variational problems using an appropriate second-order condition only, which reduces to SOSC (3) in the case of optimization. In principle, all the previous work on sSQP deals with inequality constraints. But since [9] does not require constraint qualifications, superlinear convergence for the case when equality constraints are present can be easily derived from [9] under SOSC (3) by replacing the equality $h(x) = 0$ by the two inequalities $h(x) \leq 0$ and $-h(x) \leq 0$. However, as will be shown in Sect. 2, when there are only equality constraints sharper results can be obtained by a different analysis. On the other hand, as will be shown in Sect. 3, the same is not possible when inequality constraints are present.

Convergence result for sSQP [9, Theorem 5] asserts that if $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ satisfying SOSC (3), then the sSQP iterates are well-defined and converge superlinearly to $(\bar{x}, \lambda^*, \mu^*)$, with some $(\lambda^*, \mu^*) \in \mathcal{M}$ close to $(\bar{\lambda}, \bar{\mu})$. The proof is by invoking [11, Theorem 1] after verifying assumptions of the latter by establishing the two key properties of the given stationary point \bar{x} and the associated multiplier $(\bar{\lambda}, \bar{\mu})$, stated immediately below.

Property 1 (*upper Lipschitzian behavior of the solutions of KKT system under canonical perturbations*). There exists a neighborhood \mathcal{U} of $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and $L > 0$ such that for any $p = (a, b, c) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(0, 0, 0)$, any solution $(x(p), \lambda(p), \mu(p)) \in \mathcal{U}$ of the perturbed KKT system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = a, \quad h(x) = b, \quad \mu \geq 0, \quad g(x) \leq c, \quad \langle \mu, g(x) - c \rangle = 0 \quad (6)$$

satisfies the estimate

$$\|x(p) - \bar{x}\| + \text{dist}((\lambda(p), \mu(p)), \mathcal{M}) \leq L\|p\|,$$

where

$$\text{dist}((\lambda, \mu), \mathcal{M}) = \inf_{(\hat{\lambda}, \hat{\mu}) \in \mathcal{M}} \|(\lambda, \mu) - (\hat{\lambda}, \hat{\mu})\|.$$

Property 2 (solvability of sSQP subproblems). There exists $M > 0$ such that for all $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the QP problem

$$\begin{aligned} &\text{minimize}_{(x, \lambda, \mu)} \left\{ \langle f'(x), x - \tilde{x} \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(\tilde{x}, \tilde{\lambda}, \tilde{\mu})(x - \tilde{x}), x - \tilde{x} \right\rangle \right. \\ &\quad \left. + \frac{\sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu})}{2} (\|\lambda\|^2 + \|\mu\|^2) \right\} \tag{7} \\ &\text{subject to} \quad h(\tilde{x}) + h'(\tilde{x})(x - \tilde{x}) - \sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu})(\lambda - \tilde{\lambda}) = 0, \\ &\quad g(\tilde{x}) + g'(\tilde{x})(x - \tilde{x}) - \sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu})(\mu - \tilde{\mu}) \leq 0, \end{aligned}$$

has a stationary point $(\hat{x}, \hat{\lambda}, \hat{\mu})$ satisfying the estimate

$$\|(\hat{x} - \tilde{x}, \hat{\lambda} - \tilde{\lambda}, \hat{\mu} - \tilde{\mu})\| \leq M(\|\tilde{x} - \bar{x}\| + \text{dist}((\tilde{\lambda}, \tilde{\mu}), \mathcal{M})).$$

Remark 1 According to [11, Theorem 2], Property 1 is further equivalent to the following error bound: there exists $L > 0$ such that for all $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, it holds that

$$\|\tilde{x} - \bar{x}\| + \text{dist}((\tilde{\lambda}, \tilde{\mu}), \mathcal{M}) \leq L\sigma(\tilde{x}, \tilde{\lambda}, \tilde{\mu}). \tag{8}$$

We note that any assumptions implying the two properties above guarantee local superlinear convergence of sSQP. In particular, [9] shows that SOSC (3) does the job.

It is worth to clarify whether SOSC in the form of (3) could be relaxed or not; for example, because of the following considerations. As is well-known, when multipliers are not unique the existence of a multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$ satisfying

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \geq 0 \quad \forall \xi \in C \tag{9}$$

is not necessary for optimality of \bar{x} (for example, in the case when MFCQ holds the choice of the multiplier that verifies the inequality in (9) depends on ξ in general; see, e.g., [3, Theorem 3.45]). In particular, when multipliers are not unique, SOSC (3) does not have any necessary counterpart. Thus, some sort of relaxation of SOSC (3) would be desirable. We shall show that in the case of equality constraints SOSC can be relaxed to the assumption that the relevant multiplier is noncritical, while in the general case such a relaxation is not possible.

To be specific, our objectives in the present paper are as follows. Our principal contribution consists of proving in Sect. 2 that for the problem with equality constraints only, sSQP converges superlinearly under the single assumption that the dual starting point is close to a noncritical multiplier. Since noncritical multipliers include all those satisfying SOSC but are not limited to them, we believe this gives the first superlinear convergence result for any Newtonian method for constrained optimization under assumptions that do not include constraint qualifications and are weaker than SOSC. Second, in Sect. 3, for general problems with equality and inequality constraints we

investigate the relationships between the two principal properties stated above, that are needed for superlinear convergence of sSQP according to the framework of [11]. In particular, we show that Property 1 does not imply Property 2 in general. But if \bar{x} satisfies the second-order necessary optimality condition, we show that the upper Lipschitzian property is in fact equivalent to SOSC and, thus, implies solvability of subproblems (this is because SOSC is sufficient for both properties, according to [9]). Furthermore, we point out that the upper Lipschitzian property is in fact equivalent to the multiplier being noncritical, as defined below. In Sect. 4, we consider applying sSQP to the equality-constrained reformulation where inequality constraints are converted into equalities using squared slack variables, and show that the assumptions for convergence in this case are different from (not related to) SOSC for the original inequality-constrained problem. In Sect. 5, we discuss consequences of our results for local regularization methods proposed in [16,27] for dealing with nonunique multipliers. In particular, we show that for both methods SOSC employed originally can be replaced by the weaker assumption that the relevant multiplier is noncritical.

Some words about our notation. Given the notation for active constraints introduced above, we denote inactive ones by $N = N(\bar{x}) = \{1, \dots, m\} \setminus A$. For an index set J , by y_J we denote the components of the vector y indexed by J . Analogously, for a matrix H , H_J stands for the matrix composed by rows of H indexed by J . The identity matrix is denoted by I , and $\text{diag } y$ stands for the diagonal matrix with the entries given by the components of the vector y . We finally note that under our assumptions, the natural residual $\sigma(\cdot)$ defined by (5) is Lipschitz-continuous near $\{\bar{x}\} \times \mathbf{R}^l \times \mathbf{R}^m$, $\sigma(\bar{x}, \lambda, \mu) = 0$ for all $(\lambda, \mu) \in \mathcal{M}$, and hence

$$\sigma(x, \lambda, \mu) = O(\|x - \bar{x}\| + \text{dist}((\lambda, \mu), \mathcal{M})). \tag{10}$$

2 Local convergence in the case of equality constraints

Consider the equality-constrained problem

$$\begin{aligned} & \text{minimize}_x f(x) \\ & \text{subject to } h(x) = 0. \end{aligned} \tag{11}$$

In this case, the KKT system (2) reduces to the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad h(x) = 0, \tag{12}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \rightarrow \mathbf{R}$,

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

is the Lagrangian of problem (11). Accordingly, \mathcal{M} now stands for the set of those $\lambda \in \mathbf{R}^l$ satisfying (12) for $x = \bar{x}$, and the critical cone is $C = \ker h'(\bar{x})$. For problem (11), MFCQ, SMFCQ and LICQ all reduce to the classical regularity condition $\text{rank } h'(\bar{x}) = l$.

For problem (11), the sSQP subproblem (4) is given by the equality-constrained QP

$$\begin{aligned} & \text{minimize}_{(x, \lambda)} \left\{ \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) (x - x^k), x - x^k \right\rangle \right. \\ & \quad \left. + \frac{\sigma_k}{2} \|\lambda\|^2 \right\} \\ & \text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0, \end{aligned} \tag{13}$$

with $\sigma_k = \sigma(x^k, \lambda^k)$, where $\sigma(\cdot)$ is the residual of the Lagrange system (12), i.e., $\sigma : \mathbf{R}^n \times \mathbf{R}^l \rightarrow \mathbf{R}_+$,

$$\sigma(x, \lambda) = \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda) \\ h(x) \end{pmatrix} \right\|. \tag{14}$$

The Lagrange optimality system for subproblem (13) can be written in the form

$$\begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) & (h'(x^k))^T \\ h'(x^k) & -\sigma_k I \end{pmatrix} \begin{pmatrix} x - x^k \\ \lambda - \lambda^k \end{pmatrix} = - \begin{pmatrix} \frac{\partial L}{\partial x}(x^k, \lambda^k) \\ h(x^k) \end{pmatrix}, \tag{15}$$

and this linear system will now be regarded as the sSQP subproblem for problem (11).

Recall that according to Remark 1, which in the case of equality constraints can also be easily seen directly, Property 1 is equivalent to the following error bound:

Property 1-ECP. There exists $L > 0$ such that for any $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, it holds that

$$\|\tilde{x} - \bar{x}\| + \text{dist}(\tilde{\lambda}, \mathcal{M}) \leq L\sigma(\tilde{x}, \tilde{\lambda}). \tag{16}$$

As for Property 2, in the case of equality constraints it takes the following form:

Property 2-ECP. There exists $M > 0$ such that for all $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, the linear system

$$\begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(\tilde{x}, \tilde{\lambda}) & (h'(\tilde{x}))^T \\ h'(\tilde{x}) & -\sigma(\tilde{x}, \tilde{\lambda})I \end{pmatrix} \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix} = - \begin{pmatrix} \frac{\partial L}{\partial x}(\tilde{x}, \tilde{\lambda}) \\ h(\tilde{x}) \end{pmatrix} \tag{17}$$

has a solution $(\hat{x}, \hat{\lambda})$ satisfying the estimate

$$\|(\hat{x} - \tilde{x}, \hat{\lambda} - \tilde{\lambda})\| \leq M(\|\tilde{x} - \bar{x}\| + \text{dist}(\tilde{\lambda}, \mathcal{M})). \tag{18}$$

For equality-constrained problems, the definition of a critical multiplier was originally introduced in [14]. Special numerical and analytical properties of critical multipliers were further studied in [15, 17–19].

Definition 1 A multiplier $\bar{\lambda} \in \mathcal{M}$ is called *critical* if

$$\exists \xi \in \ker h'(\bar{x}) \setminus \{0\} \text{ such that } \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \text{im}(h'(\bar{x}))^T, \tag{19}$$

and *noncritical* otherwise.

Note that since $\text{im}(h'(\bar{x}))^T = (\ker h'(\bar{x}))^\perp$, if $\bar{\lambda} \in \mathcal{M}$ is critical then

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

In particular, SOSC (3) does not hold. Thus critical multipliers form a subclass of multipliers violating SOSC, while noncritical multipliers include all those that satisfy SOSC and more.

We first note that it follows from [15, Theorem 2.3] (and the discussion following that theorem) that Property 1 is actually equivalent to saying that the multiplier in question is noncritical. For convenience of a reader, in the case of equality constraints we shall provide a simple proof of this fact.

Proposition 1 For a Lagrange multiplier $\bar{\lambda}$ associated with a stationary point \bar{x} of problem (11), Property 1-ECP is equivalent to saying that $\bar{\lambda}$ is noncritical.

Proof We start with showing that if $\bar{\lambda}$ is a noncritical multiplier then Property 1-ECP holds. To demonstrate that (16) is satisfied with some $L > 0$ for all $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, we actually need to prove the following two estimates:

$$\|\tilde{x} - \bar{x}\| = O(\sigma(\tilde{x}, \tilde{\lambda})) \tag{20}$$

and

$$\text{dist}(\tilde{\lambda}, \mathcal{M}) = O(\sigma(\tilde{x}, \tilde{\lambda})) \tag{21}$$

as $(\tilde{x}, \tilde{\lambda}) \rightarrow (\bar{x}, \bar{\lambda})$. Note that if $\tilde{x} = \bar{x}$ and $\tilde{\lambda} \in \mathcal{M}$ then both (20) and (21) hold trivially. Taking this into account, we now prove (20) by contradiction: suppose that there exist sequences $\{\tilde{x}^k\} \subset \mathbf{R}^n$ and $\{\tilde{\lambda}^k\} \subset \mathbf{R}^l$ such that $\{\tilde{x}^k\} \rightarrow \bar{x}$, $\{\tilde{\lambda}^k\} \rightarrow \bar{\lambda}$, $\tilde{x}^k \neq \bar{x}$ or $\tilde{\lambda}^k \notin \mathcal{M}$ for each k , and

$$\frac{\|\tilde{x}^k - \bar{x}\|}{\sigma(\tilde{x}^k, \tilde{\lambda}^k)} \rightarrow \infty$$

or, equivalently,

$$\sigma(\tilde{x}^k, \tilde{\lambda}^k) = o(\|\tilde{x}^k - \bar{x}\|). \tag{22}$$

Employing (14), (22) and our smoothness hypothesis, we then derive the estimates

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

Assuming that the entire sequence $\{(\tilde{x}^k - \bar{x}) / \|\tilde{x}^k - \bar{x}\|\}$ converges to some $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$ (passing onto a subsequence, if necessary), the last two estimates imply the relations

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \text{im}(h'(\bar{x}))^T, \quad h'(\bar{x})\xi = 0.$$

Since $\xi \neq 0$, the latter contradicts the assumption that $\bar{\lambda}$ is a noncritical multiplier (cf. (19)). Therefore, (20) is valid.

Furthermore, since \mathcal{M} is defined by a linear system, estimate (21) follows immediately from the Hoffman's Lemma (see, e.g., [3, Theorem 2.200]), and from (14) and (20):

$$\begin{aligned} \text{dist}(\tilde{\lambda}, \mathcal{M}) &= O\left(\left\|\frac{\partial L}{\partial x}(\bar{x}, \tilde{\lambda})\right\|\right) \\ &= O\left(\left\|\frac{\partial L}{\partial x}(\tilde{x}, \tilde{\lambda})\right\|\right) + O\left(\left\|\frac{\partial L}{\partial x}(\tilde{x}, \tilde{\lambda}) - \frac{\partial L}{\partial x}(\bar{x}, \tilde{\lambda})\right\|\right) \\ &= O(\sigma(\tilde{x}, \tilde{\lambda})) + O(\|\tilde{x} - \bar{x}\|) \\ &= O(\sigma(\tilde{x}, \tilde{\lambda})). \end{aligned}$$

This completes the proof of the first implication: if $\bar{\lambda}$ is a noncritical multiplier then Property 1-ECP holds.

To establish the converse implication, we shall assume that the multiplier $\bar{\lambda}$ is critical and explicitly construct $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ arbitrarily close to $(\bar{x}, \bar{\lambda})$ and violating (16) for any pre-fixed $L > 0$. Since $\bar{\lambda}$ is critical, according to (19) there exist $\xi \in \mathbf{R}^n \setminus \{0\}$ and $\eta \in \mathbf{R}^l$ such that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi + (h'(\bar{x}))^T\eta = 0, \quad h'(\bar{x})\xi = 0. \tag{23}$$

Setting $\tilde{x} = \tilde{x}(t) = \bar{x} + t\xi$, $\tilde{\lambda} = \tilde{\lambda}(t) = \bar{\lambda} + t\eta$, $t \in \mathbf{R}$, one immediately obtains from (23) that under our smoothness hypothesis

$$\begin{aligned} \frac{\partial L}{\partial x}(\tilde{x}, \tilde{\lambda}) &= \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}) + t \left(\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi + (h'(\bar{x}))^T \eta \right) + o(t) = o(t), \\ h(\tilde{x}) &= h(\bar{x}) + th'(\bar{x})\xi + o(t) = o(t). \end{aligned}$$

According to (14), this implies the estimate $\sigma(\tilde{x}, \tilde{\lambda}) = o(t)$, while $\|\tilde{x} - \bar{x}\| = t\|\xi\|$, where $\xi \neq 0$. Hence, the estimate (16) cannot hold for t close enough to zero. \square

We next prove that for the equality-constrained problem (11), if $\bar{\lambda}$ is a noncritical multiplier then Property 2-ECP holds automatically. This implies that noncriticality of the multiplier (which, in particular, is weaker than SOSC) is the only assumption needed for the local superlinear convergence of sSQP in the case of equality-constrained problems.

We first need to prove that the matrix in (17) is locally nonsingular. We believe this can be done in several ways. One line of proof can be based on introducing the singular-value decomposition of the constraints Jacobian, along the lines of [24, Theorem 3.2]; however, we cannot quote [24, Theorem 3.2] since its proof uses the stronger SOSC assumption. We pursue here a different line of analysis, based on the following result close in nature to the Finsler–Debreu Lemma [5, 10], which states that if for some matrices H and B it holds that $\langle H\xi, \xi \rangle > 0$ for all $\xi \in \ker B \setminus \{0\}$, then $H + tB^T B$ is positive definite for all $t > 0$ large enough. We feel that Lemma 1 below is of independent interest and may also find applications other than the one in this paper.

Lemma 1 *Let H be an $n \times n$ -matrix, B be an $l \times n$ -matrix, and assume that*

$$H\xi \notin \text{im } B^T \quad \forall \xi \in \ker B \setminus \{0\}. \tag{24}$$

Then for any $c > 0$, any $n \times n$ -matrix \tilde{H} close enough to H , and any $l \times n$ -matrix \tilde{B} close enough to B , the matrix $\tilde{H} + t(B + \Omega)^T \tilde{B}$ is nonsingular for all $t \in \mathbf{R}$ such that $|t|$ is large enough, and for all $l \times n$ -matrices Ω satisfying $\|\Omega\| \leq c/|t|$.

Proof Suppose the contrary, i.e., that there exist sequences $\{H_k\}$ of $n \times n$ -matrices, $\{B_k\}$ and $\{\Omega_k\}$ of $l \times n$ -matrices, $\{t_k\} \subset \mathbf{R}$ and $\{\xi^k\} \subset \mathbf{R}^n \setminus \{0\}$ such that $\{H_k\} \rightarrow H$, $\{B_k\} \rightarrow B$, $|t_k| \rightarrow \infty$, for all k it holds that $\|\Omega_k\| \leq c/|t_k|$ and

$$H_k \xi^k + t_k(B + \Omega_k)^T B_k \xi^k = 0. \tag{25}$$

Clearly, we can assume, without loss of generality, that $\|\xi^k\| = 1$ for all k , and $\{\xi^k\} \rightarrow \xi \neq 0$. Then it must hold that $B^T B \xi = 0$, since

$$B^T B_k \xi^k = -\frac{1}{t_k} H_k \xi^k - \Omega_k^T B_k \xi^k \rightarrow 0 \text{ as } k \rightarrow \infty.$$

We thus proved that $B\xi \in \ker B^T$, and since $B\xi \in \text{im } B = (\ker B^T)^\perp$, this shows that $B\xi = 0$. Thus, $\xi \in \ker B \setminus \{0\}$.

On the other hand, (25) implies that the inclusion

$$H_k \xi^k + t_k \Omega_k^T B_k \xi^k = -t_k B^T B_k \xi^k \in \text{im } B^T$$

holds for all k , where the second term in the left-hand side tends to zero as $k \rightarrow \infty$ because $\{t_k \Omega_k\}$ is bounded and $\{B_k \xi^k\} \rightarrow B \xi = 0$. Hence, $H \xi \in \text{im } B^T$ by closedness of $\text{im } B^T$. This completes a contradiction with (24). \square

Lemma 2 *If $\bar{\lambda}$ is a noncritical Lagrange multiplier associated with a stationary point \bar{x} of problem (11) then the matrix in the left-hand side of (17) is nonsingular for all $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, such that $\tilde{x} \neq \bar{x}$ or $\tilde{\lambda} \notin \mathcal{M}$.*

Proof By Proposition 1, the assumption that $\bar{\lambda}$ is noncritical in the sense of Definition 1 is equivalent to Property 1-ECP. It is then evident that $\sigma(\tilde{x}, \tilde{\lambda}) > 0$ if $\tilde{x} \neq \bar{x}$ or $\tilde{\lambda} \notin \mathcal{M}$.

Suppose that there exist sequences $\{(\tilde{x}^k, \tilde{\lambda}^k)\} \subset (\mathbf{R}^n \times \mathbf{R}^l) \setminus (\{\bar{x}\} \times \mathcal{M})$ and $\{(\xi^k, \eta^k)\}$ such that $\{(\tilde{x}^k, \tilde{\lambda}^k)\} \rightarrow (\bar{x}, \bar{\lambda})$, and for each k , (ξ^k, η^k) belongs to the null space of the matrix in question, that is,

$$\frac{\partial^2 L}{\partial x^2}(\tilde{x}^k, \tilde{\lambda}^k) \xi^k + (h'(\tilde{x}^k))^T \eta^k = 0, \quad h'(\tilde{x}^k) \xi^k - \sigma(\tilde{x}^k, \tilde{\lambda}^k) \eta^k = 0.$$

The second equation implies that

$$\eta^k = \frac{1}{\sigma(\tilde{x}^k, \tilde{\lambda}^k)} h'(\tilde{x}^k) \xi^k, \tag{26}$$

and by substituting this into the first equation, we obtain

$$\left(\frac{\partial^2 L}{\partial x^2}(\tilde{x}^k, \tilde{\lambda}^k) + \frac{1}{\sigma(\tilde{x}^k, \tilde{\lambda}^k)} (h'(\tilde{x}^k))^T h'(\tilde{x}^k) \right) \xi^k = 0. \tag{27}$$

Set $H = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})$, $B = h'(\bar{x})$, $H_k = \frac{\partial^2 L}{\partial x^2}(\tilde{x}^k, \tilde{\lambda}^k)$, $B_k = h'(\tilde{x}^k)$, $\Omega_k = h'(\tilde{x}^k) - h'(\bar{x})$, $t_k = 1/\sigma(\tilde{x}^k, \tilde{\lambda}^k)$. Observing that, by Property 1-ECP (see 16), it holds that

$$\Omega_k = O(\|\tilde{x}^k - \bar{x}\|) = O(\sigma(\tilde{x}^k, \tilde{\lambda}^k)) = O(1/t_k),$$

and applying Lemma 1 with $\tilde{H} = H_k$, $\tilde{B} = B_k$, $\Omega = \Omega_k$ and $t = t_k$, we obtain that for all k large enough, the matrix in the left-hand side of (27) is nonsingular. Then (27) implies that $\xi^k = 0$. By (26), we then have that $\eta^k = 0$.

We have thus established that the null space of the matrix in the left-hand side of (17) contains the zero vector only. The conclusion follows. \square

It remains to establish the estimate (18).

Proposition 2 *If $\bar{\lambda}$ is a noncritical Lagrange multiplier associated with a stationary point \bar{x} of problem (11) then Property 2-ECP holds.*

Proof If $\tilde{x} = \bar{x}$ and $\tilde{\lambda} \in \mathcal{M}$ then the right-hand side of system (17) is equal to zero, and hence, this system has a solution of the form $(\hat{x}, \hat{\lambda}) = (\tilde{x}, \tilde{\lambda})$, trivially satisfying (18) with any $M \geq 0$.

Now, according to Lemma 2, for all $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ and such that $\tilde{x} \neq \bar{x}$ or $\tilde{\lambda} \notin \mathcal{M}$, the linear system (17) has the (unique) solution $(\hat{x}, \hat{\lambda})$. Suppose that there exists no $M > 0$ such that the estimate (18) is valid. The latter means that there exist sequences $\{(\tilde{x}^k, \tilde{\lambda}^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ and $\{(\hat{x}^k, \hat{\lambda}^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ such that for each k , $\tilde{x}^k \neq \bar{x}$ or $\tilde{\lambda}^k \notin \mathcal{M}$, $(\hat{x}^k, \hat{\lambda}^k)$ solves system (17) with $(\tilde{x}, \tilde{\lambda}) = (\tilde{x}^k, \tilde{\lambda}^k)$, and

$$\frac{\|(\hat{x}^k - \tilde{x}^k, \hat{\lambda}^k - \tilde{\lambda}^k)\|}{\|\tilde{x}^k - \bar{x}\| + \text{dist}(\tilde{\lambda}^k, \mathcal{M})} \rightarrow \infty \text{ as } k \rightarrow \infty.$$

Setting $\xi^k = \hat{x}^k - \tilde{x}^k$, $\eta^k = \hat{\lambda}^k - \tilde{\lambda}^k$, the last relation above can be written as

$$\frac{\|\tilde{x}^k - \bar{x}\| + \text{dist}(\tilde{\lambda}^k, \mathcal{M})}{\|(\xi^k, \eta^k)\|} \rightarrow 0 \text{ as } k \rightarrow \infty, \tag{28}$$

and from (17) we obtain that

$$\begin{aligned} \frac{\partial^2 L}{\partial x^2}(\tilde{x}^k, \tilde{\lambda}^k) \xi^k + (h'(\tilde{x}^k))^T \eta^k &= -\frac{\partial L}{\partial x}(\tilde{x}^k, \tilde{\lambda}^k), \\ h'(\tilde{x}^k) \xi^k - \sigma(\tilde{x}^k, \tilde{\lambda}^k) \eta^k &= -h(\tilde{x}^k). \end{aligned} \tag{29}$$

Furthermore, employing (10), (28) implies that

$$\frac{\left\| \left(\frac{\partial L}{\partial x}(\tilde{x}^k, \tilde{\lambda}^k), h(\tilde{x}^k) \right) \right\|}{\|(\xi^k, \eta^k)\|} = \frac{\sigma(\tilde{x}^k, \tilde{\lambda}^k)}{\|(\xi^k, \eta^k)\|} \rightarrow 0 \text{ as } k \rightarrow \infty. \tag{30}$$

Passing onto a convergent subsequence if necessary, we can assume that $\{(\xi^k, \eta^k)/\|(\xi^k, \eta^k)\|\} \rightarrow (\xi, \eta) \neq 0$. Then dividing the relations in (29) by $\|(\xi^k, \eta^k)\|$, passing onto the limit and using also (30), we obtain that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi + (h'(\bar{x}))^T \eta = 0, \quad h'(\bar{x})\xi = 0. \tag{31}$$

If $\xi \neq 0$ then the latter contradicts the assumption that $\bar{\lambda}$ is noncritical. Thus $\xi = 0$, and the first equation in (31) reduces to

$$(h'(\bar{x}))^T \eta = 0. \tag{32}$$

Furthermore, denote by P the orthogonal projector onto $(\text{im } h'(\bar{x}))^\perp$. Observe that $Ph'(\bar{x}) = 0$. Then by the second equation in (29), by twice differentiability of h at \bar{x} , and by Lipschitz-continuity of $h'(\cdot)$ near \bar{x} , we obtain

$$\begin{aligned} \sigma(\tilde{x}^k, \tilde{\lambda}^k)P\eta^k &= P \left(h(\tilde{x}^k) + h'(\tilde{x}^k)\xi^k \right) \\ &= P(h(\bar{x}) + h'(\bar{x})(\tilde{x}^k - \bar{x}) + O(\|\tilde{x}^k - \bar{x}\|^2) \\ &\quad + h'(\bar{x})\xi^k + (h'(\tilde{x}^k) - h'(\bar{x}))\xi^k) \\ &= O(\|\tilde{x}^k - \bar{x}\|^2) + O(\|\tilde{x}^k - \bar{x}\|\|\xi^k\|) \\ &= O((\sigma(\tilde{x}^k, \tilde{\lambda}^k))^2) + O(\sigma(\tilde{x}^k, \tilde{\lambda}^k)\|\xi^k\|), \end{aligned}$$

where the last equality is by Property 1-ECP (equivalent to the assumption that $\bar{\lambda}$ is noncritical). Thus,

$$P\eta^k = O\left(\sigma(\tilde{x}^k, \tilde{\lambda}^k)\right) + O\left(\|\xi^k\|\right).$$

Dividing both sides of the latter relation by $\|(\xi^k, \eta^k)\|$ and passing onto the limit, we obtain that

$$P\eta = 0,$$

where (30) and the fact that $\xi = 0$ were used. This shows that $\eta \in \ker P = \text{im } h'(\bar{x})$. At the same time, (32) means that $\eta \in \ker(h'(\bar{x}))^T = (\text{im } h'(\bar{x}))^\perp$. Hence, $\eta = 0$, which gives a contradiction with $(\xi, \eta) \neq 0$. \square

By Proposition 2, the discussion above, and [11, Theorem 1], we finally obtain the following result.

Theorem 1 *Let $\bar{\lambda}$ be a noncritical Lagrange multiplier associated with a stationary point \bar{x} of problem (11).*

If $(x^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ is close enough to $(\bar{x}, \bar{\lambda})$, then the sSQP iterates $\{(x^k, \lambda^k)\}$ given by (15), with $\sigma_k = \sigma(x^k, \lambda^k)$ defined according to (14), are well-defined and converge superlinearly to (\bar{x}, λ^) for some $\lambda^* \in \mathcal{M}$ such that $\lambda^* \rightarrow \bar{\lambda}$ as $(x^0, \lambda^0) \rightarrow (\bar{x}, \bar{\lambda})$.*

3 Equality and inequality constraints

We now turn our attention to the general problem (1) and proceed to investigate the relationships between the properties relevant for sSQP convergence, second-order

necessary/sufficient optimality conditions, critical multipliers (and some others) in more detail.

It is worth to point out that Properties 1 and 2 have natural counterparts in the analysis of SQP and, more generally, of (Josephy–)Newton-type methods for generalized equations in [2]. Specifically, Property 1 is a weakened version of semistability of a solution [2] (the latter subsumes that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is an isolated solution of the KKT system (2), in which case semistability and Property 1 are the same). The property of semistability has various equivalent characterizations; see [2, 4, 7] for details. Property 2 is related to the so-called hemistability [2] (although the latter concerns solvability of subproblems of the usual SQP rather than sSQP). Note that solvability of subproblems is, of course, a necessary condition for the corresponding iterates to be well-defined. In [2] it was shown that generally, semistability does not imply hemistability, and that both assumptions are needed for the analysis of Newton-type methods for generalized equations. However, [2] also shows that semistability implies hemistability provided \bar{x} is a local solution of problem (1). Our primary goal in this section is to clarify whether similar relationships can be derived in the context of sSQP, i.e., for Properties 1 and 2 above, and establish their connections to other relevant concepts.

We first extend the notion of critical/noncritical multipliers to the problem with equality and inequality constraints. To this end, observe that the definition (19) of a critical multiplier for equality-constrained problems is equivalent to saying that the linear system (23) has a solution $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^l$, with $\xi \neq 0$. An extension of this definition to problem (1) when inequality constraints are present is given by the following.

Definition 2 A multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$ is called *critical* if there exists a triple $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, with $\xi \neq 0$, satisfying the system

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^T \eta + (g'(\bar{x}))^T \zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \tag{33}$$

$$\zeta_{A_0} \geq 0, \quad g'_{A_0}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}), \xi \rangle = 0, \quad i \in A_0, \quad \zeta_N = 0, \tag{34}$$

and *noncritical* otherwise.

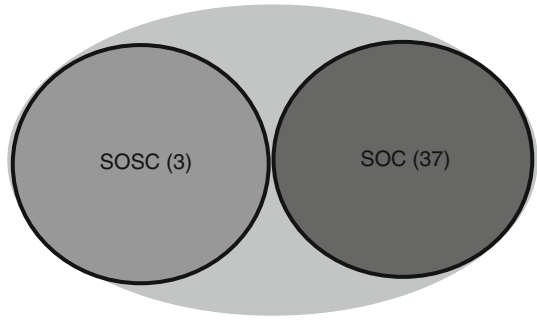
As can be easily seen, (33)–(34) is the KKT system for the QP problem

$$\begin{aligned} & \text{minimize}_{\xi} \quad \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \\ & \text{subject to} \quad \xi \in C. \end{aligned} \tag{35}$$

Thus, $(\bar{\lambda}, \bar{\mu})$ being noncritical is equivalent to saying that $\xi = 0$ is the unique stationary point of problem (35). It follows that the multiplier $(\bar{\lambda}, \bar{\mu})$ being noncritical is further equivalent to saying that the pair $(C, \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu}))$ has the so-called R_0 property; see the discussion following (3.3.18) in [7].

Moreover, multiplying the first equality in (33) by ξ and using the second equality in (33) and the relations in (34), it can be seen that a sufficient condition for $(\bar{\lambda}, \bar{\mu})$ to be

Fig. 1 Critical multipliers



noncritical (i.e., for (33)–(34) not to possess any solution with nonzero ξ component) is the following:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \neq 0 \quad \forall \xi \in C \setminus \{0\}. \tag{36}$$

Note that, by the convexity of the cone C , the condition above means that the quadratic form in question has either the positive sign or the negative sign on C , i.e., either SOSC (3) holds or the following second-order condition (SOC) holds:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle < 0 \quad \forall \xi \in C \setminus \{0\}. \tag{37}$$

Thus, all multipliers satisfying SOSC (3) are noncritical, as well as those satisfying SOC (37) (or, in other words, any critical multiplier necessarily violates SOSC (3) and SOC (37)). However, it is important to emphasize that there may exist noncritical multipliers that do not satisfy either (3) or (37), as (36) is evidently merely a sufficient condition, not an equivalent definition of noncriticality. See also Example 3.

Figure 1 is intended to provide some intuition regarding the structure of the set of critical multipliers. Gray oval represents the set \mathcal{M} of all multipliers, while two circles represent the sets of multipliers satisfying SOSC (3) and SOC (37). The boundary of these circles corresponds to critical multipliers (but there may exist other critical multipliers as well).

As already mentioned, it follows from the analysis in [15] that Property 1 is actually equivalent to saying that the multiplier in question is noncritical. In the case when inequality constraints are present the proof is rather involved when compared to the proof for the equality constrained case given above. We shall not give details here, referring the reader to [15].

Proposition 3 For a Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated with a stationary point \bar{x} of problem (1), Property 1 is equivalent to saying that $(\bar{\lambda}, \bar{\mu})$ is noncritical.

Remark 2 It is worth to point out that, according to Remark 1 and Proposition 3, the error bound property (8) is equivalent to the assumption that the multiplier in question is noncritical. Consequently, the weaker noncriticality assumption could be employed

everywhere where SOSC (3) had been invoked for purposes of the error bound (8). Some examples are [13, 16, 27]; see also the discussion in Sect. 5.

We next point out that Property 1 does not imply Property 2 in general. To show this, we shall employ the example that was used in [2] to illustrate that semistability does not imply hemistability (even in the simplest nondegenerate case). Example 1 shows, in particular, that for problems with inequality constraints the weaker assumption that the multiplier is noncritical cannot replace SOSC (3) in [9, Theorem 5] to establish superlinear convergence of sSQP. However, for problems with equality constraints only such a relaxation is possible, as was shown in Sect. 2.

Example 1 Let $n = 1, l = 0, m = 1, f(x) = -x^2/2 + x^3/6, g(x) = -x$. Problem (1) with this data has a global minimizer $x^* = 2$, but it also has a stationary point $\bar{x} = 0$, which is not even a local minimizer. The point \bar{x} satisfies LICQ, and the associated unique multiplier is $\bar{\mu} = 0$. Thus, $A = A_0, A_+ = \emptyset$, and system (33), (34) takes the form

$$-\xi - \zeta = 0, \quad \zeta \geq 0, \quad \xi \geq 0, \quad \zeta\xi = 0,$$

which has the unique solution $(\xi, \eta) = (0, 0)$. Thus, $\bar{\mu}$ is a noncritical multiplier, and hence, Property 1 holds (which in this case reduces to semistability).

We next show that sSQP subproblems may not be solvable, i.e., Property 2 does not hold. For a given $(\tilde{x}, \tilde{\mu}) \in \mathbf{R} \times \mathbf{R}$, problem (7) takes the form

$$\begin{aligned} &\text{minimize}_{(x, \mu)} - \left(\tilde{x} - \frac{1}{2}\tilde{x}^2 \right) (x - \tilde{x}) - \frac{1}{2}(1 - \tilde{x})(x - \tilde{x})^2 + \frac{\sigma(\tilde{x}, \tilde{\mu})}{2}\mu^2 \\ &\text{subject to} \quad x + \sigma(\tilde{x}, \tilde{\mu})(\mu - \tilde{\mu}) \geq 0, \end{aligned}$$

where

$$\sigma(\tilde{x}, \tilde{\mu}) = \left\| \begin{pmatrix} -\tilde{x} + \frac{1}{2}\tilde{x}^2 - \tilde{\mu} \\ \min\{\tilde{\mu}, \tilde{x}\} \end{pmatrix} \right\|.$$

One can directly check that for any $(\tilde{x}, \tilde{\mu}) \neq (0, 0)$ close enough to $(0, 0)$ and satisfying $\tilde{\mu} \geq 0$, the latter problem does not have any stationary points. Indeed, the KKT system of this problem gives the relations

$$\begin{aligned} &\frac{1}{2}\tilde{x}^2 + (1 - \tilde{x})x + \mu = 0, \quad \mu \geq 0, \quad x + \sigma(\tilde{x}, \tilde{\mu})(\mu - \tilde{\mu}) \geq 0, \\ &\mu(x + \sigma(\tilde{x}, \tilde{\mu})(\mu - \tilde{\mu})) = 0. \end{aligned} \tag{38}$$

If we suppose that $\mu = 0$ then the first equation in (38) gives

$$x = -\frac{\tilde{x}^2}{2(1 - \tilde{x})},$$

and hence,

$$x + \sigma(\tilde{x}, \tilde{\mu})(\mu - \tilde{\mu}) = -\frac{\tilde{x}^2}{2(1 - \tilde{x})} - \sigma(\tilde{x}, \tilde{\mu})\tilde{\mu} < 0,$$

which means that the third relation in (38) cannot be satisfied.

On the other hand, if $\mu > 0$ then according to the complementary slackness condition (the last relation in (38)), the third relation in (38) must hold as an equality. Combined with the first equation in (38), this gives

$$-\frac{\tilde{x}^2 + 2\mu}{2(1 - \tilde{x})} + \sigma(\tilde{x}, \tilde{\mu})(\mu - \tilde{\mu}) = 0.$$

The latter further implies for $(\tilde{x}, \tilde{\mu})$ close to $(0, 0)$ that

$$\mu = -\frac{\frac{1}{2}\tilde{x}^2 + \sigma(\tilde{x}, \tilde{\mu})(1 - \tilde{x})\tilde{\mu}}{1 - \sigma(\tilde{x}, \tilde{\mu}) + \sigma(\tilde{x}, \tilde{\mu})\tilde{x}} < 0,$$

which means that the second relation in (38) cannot be satisfied.

We next show that the situation changes if condition (9) holds, i.e., in that case Property 1 does imply Property 2. The argument is similar to proving that semistability implies hemistability when \bar{x} is a local solution of problem (1) [2]. That said, note that semistability implies SMFCQ, and hence MFCQ, so that in that case local optimality of \bar{x} implies (9) with the unique associated multiplier $(\bar{\lambda}, \bar{\mu})$. In our case optimality of \bar{x} and Property 1 do *not* imply (9), in general.

Proposition 4 *Suppose that (9) holds for a stationary point \bar{x} and an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ of problem (1).*

Then Property 1 (which is equivalent to saying that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier) is equivalent to SOSC (3), and thus implies Property 2.

Proof As discussed above, SOSC (3) always implies that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier.

Suppose now that $(\bar{\lambda}, \bar{\mu})$ is noncritical but SOSC (3) does not hold. Under the assumed condition (9), this means that there exists $\bar{\xi} \in C \setminus \{0\}$ such that the inequality in (9) holds as an equality for $\xi = \bar{\xi}$. Then (9) implies that $\bar{\xi}$ is a solution of problem (35), and hence, a stationary point of this problem (due to the linearity of its constraints, which is a CQ). However, according to the discussion above, such $\bar{\xi} \neq 0$ does not exist in the case of noncritical $(\bar{\lambda}, \bar{\mu})$, which gives a contradiction. Hence, SOSC (3) does hold.

The fact that SOSC (3) implies Property 2 follows from [9, Theorem 3] (the error bound needed for the latter is implied by Property 1, see Remark 1). □

To complete the study of relevant relationships, it remains to consider the case when \bar{x} is a local solution of problem (1), but condition (9) does not hold. The next example demonstrates that in this case, Property 1 does not necessarily imply Property 2.

Example 2 Let $n = 1, l = 0, m = 2, f(x) = -x^2/2, g(x) = (-x, x^3/6)$. Problem (1) with this data has the unique feasible point (hence, the unique solution) $\bar{x} = 0$. This point is stationary, and $\mathcal{M} = \{\mu \in \mathbf{R}_+^2 \mid \mu_1 = 0\}, C = \mathbf{R}_+$. Consider the multiplier $\bar{\mu} = 0$. It can be easily seen that condition (9) does not hold with this multiplier. At the same time, condition (37) holds. Hence, $\bar{\mu}$ is a noncritical multiplier, and thus, Property 1 holds (but note that semistability does not hold, since \mathcal{M} is not a singleton!).

For a given $(\tilde{x}, \tilde{\mu}) \in \mathbf{R} \times \mathbf{R}^2$, problem (7) takes the form

$$\begin{aligned} & \text{minimize}_{(x, \mu)} \quad -\tilde{x}(x - \tilde{x}) - \frac{1}{2}(x - \tilde{x})^2 + \frac{\sigma(\tilde{x}, \tilde{\mu})}{2}(\mu_1^2 + \mu_2^2) \\ & \text{subject to} \quad x + \sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1) \geq 0, \\ & \quad \quad \quad -\frac{1}{3}\tilde{x}^3 + \frac{1}{2}\tilde{x}^2x - \sigma(\tilde{x}, \tilde{\mu})(\mu_2 - \tilde{\mu}_2) \leq 0, \end{aligned} \tag{39}$$

where

$$\sigma(\tilde{x}, \tilde{\mu}) = \left\| \begin{pmatrix} -\tilde{x} - \tilde{\mu}_1 + \frac{1}{2}\tilde{x}^2\tilde{\mu}_2 \\ \min\{\tilde{\mu}_1, \tilde{x}\} \\ \min\{\tilde{\mu}_2, -\tilde{x}^3/6\} \end{pmatrix} \right\|.$$

Take $\tilde{x} > 0, \tilde{\mu}_1 > 0, \tilde{\mu}_2 = 0$. Then

$$\sigma(\tilde{x}, \tilde{\mu}) = \left\| \begin{pmatrix} \tilde{x} + \tilde{\mu}_1 \\ \min\{\tilde{\mu}_1, \tilde{x}\} \\ -\tilde{x}^3/6 \end{pmatrix} \right\|,$$

and there evidently exists a constant $\gamma > 0$ (not depending on \tilde{x} and $\tilde{\mu}_1$) such that the inequality

$$\sigma(\tilde{x}, \tilde{\mu}) \geq \gamma(\tilde{x} + \tilde{\mu}_1) \tag{40}$$

holds for all such \tilde{x} and $\tilde{\mu}$.

The KKT system of problem (39) gives the relations

$$\begin{aligned} & -x - \mu_1 + \frac{1}{2}\tilde{x}^2\mu_2 = 0, \\ & \mu_1 \geq 0, \quad x + \sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1) \geq 0, \quad \mu_1(x + \sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1)) = 0, \\ & \mu_2 \geq 0, \quad -\frac{1}{3}\tilde{x}^3 + \frac{1}{2}\tilde{x}^2x - \sigma(\tilde{x}, \tilde{\mu})\mu_2 \leq 0, \\ & \mu_2 \left(-\frac{1}{3}\tilde{x}^3 + \frac{1}{2}\tilde{x}^2x - \sigma(\tilde{x}, \tilde{\mu})\mu_2 \right) = 0. \end{aligned} \tag{41}$$

If we suppose that $\mu_1 = \mu_2 = 0$ then the first equation in (41) gives $x = 0$, and hence,

$$x + \sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1) = -\sigma(\tilde{x}, \tilde{\mu})\tilde{\mu}_1 < 0,$$

which means that the second relation in the second line of (41) cannot be satisfied.

Furthermore, if $\mu_1 = 0, \mu_2 > 0$ then according to the complementary slackness condition (the last relation in the last line of (41)), the second relation in the last line of (41) must hold as an equality. Combined with the first equation in (41), this gives

$$\left(-\sigma(\tilde{x}, \tilde{\mu}) + \frac{1}{4}\tilde{x}^4\right)\mu_2 = \frac{1}{3}\tilde{x}^3. \tag{42}$$

By (40) we derive

$$-\sigma(\tilde{x}, \tilde{\mu}) + \frac{1}{4}\tilde{x}^4 \leq -\gamma(\tilde{x} + \tilde{\mu}_1) + \frac{1}{4}\tilde{x}^4 < 0$$

provided \tilde{x} is small enough, which means that (42) may hold only with $\mu_2 < 0$. Thus, the first relation in the last line of (41) cannot be satisfied.

If $\mu_1 > 0, \mu_2 = 0$ then the first equation in (41) gives $x = -\mu_1$, and hence,

$$x + \sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1) = -(1 - \sigma(\tilde{x}, \tilde{\mu}))\mu_1 - \sigma(\tilde{x}, \tilde{\mu})\tilde{\mu}_1 < 0$$

provided \tilde{x} and $\tilde{\mu}_1$ are small enough. Thus, the second relation in the second line of (41) cannot be satisfied in this case.

Finally, if $\mu_1 > 0$ and $\mu_2 > 0$ then employing the first equation in (41) and the complementary slackness conditions (the last relations in the last two lines of (41)) we obtain the equations

$$x = -\mu_1 + \frac{1}{2}\tilde{x}^2\mu_2, \quad x = -\sigma(\tilde{x}, \tilde{\mu})(\mu_1 - \tilde{\mu}_1), \quad x = \frac{2}{3}\tilde{x} + \frac{2\sigma(\tilde{x}, \tilde{\mu})}{\tilde{x}^2}\mu_2 = 0.$$

Solving this linear system, we obtain

$$\mu_2 = -\frac{\tilde{x}^2 \left(\frac{2}{3}\tilde{x} - \sigma(\tilde{x}, \tilde{\mu})\tilde{\mu}_1 - \sigma(\tilde{x}, \tilde{\mu})\tilde{x} \right)}{\sigma(\tilde{x}, \tilde{\mu}) \left(2(1 - \sigma(\tilde{x}, \tilde{\mu})) + \frac{1}{2}\tilde{x}^4 \right)}.$$

Let, e.g., $\tilde{\mu}_1 = \tilde{x}$. Then

$$\mu_2 = -\frac{\tilde{x}^2 \left(\frac{2}{3}\tilde{x} + o(\tilde{x}) \right)}{\sigma(\tilde{x}, \tilde{\mu}) \left(2(1 - \sigma(\tilde{x}, \tilde{\mu})) + \frac{1}{2}\tilde{x}^4 \right)} < 0$$

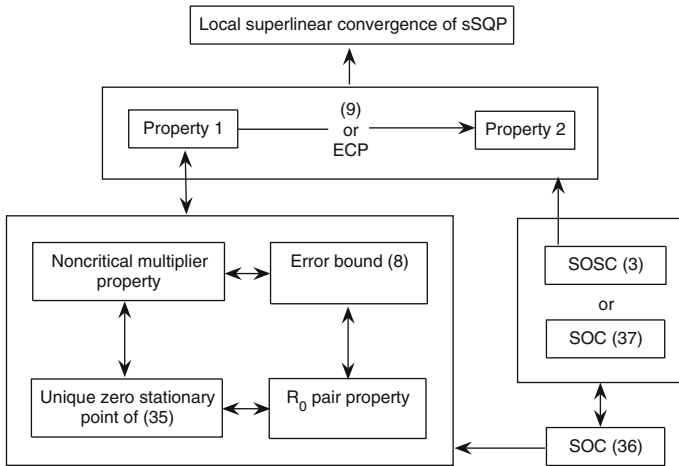


Fig. 2 Flowchart of relations between some relevant properties

provided \bar{x} is small enough, and thus, the first relation in the last line of (41) cannot be satisfied.

Summarizing, we demonstrated that for $\bar{x} = \tilde{\mu}_1 > 0$ small enough, and for $\tilde{\mu}_2 = 0$, problem (39) does not have any stationary points.

Recall again that violation of condition (9) for a local solution \bar{x} of problem (1) and an associated multiplier $(\bar{\lambda}, \bar{\mu})$ subsumes that the multiplier is not unique (i.e., SMFCQ and, even more so, LICQ do not hold at \bar{x}). But MFCQ may hold, as the latter does not imply (9) for a given (or even some) multiplier $(\bar{\lambda}, \bar{\mu})$.

The relationships discussed above are summarized in Fig. 2.

Remark 3 Some words are in order about the relation of Definition 2 with the slightly different notion of critical multipliers for problems with inequality constraints in [18].

In [18], a multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$ is called *critical with respect to an index set* $J \subset A$ if $\bar{\mu}_i = 0$ for all $i \in \{1, \dots, m\} \setminus J$, and the multiplier $(\bar{\lambda}, \bar{\mu}_J)$ associated with the stationary point \bar{x} of the equality-constrained problem

$$\begin{aligned} & \text{minimize}_x f(x) \\ & \text{subject to } h(x) = 0, \quad g_J(x) = 0 \end{aligned} \tag{43}$$

is critical for this problem in the sense of Definition 2.

One can easily check that if a multiplier $(\bar{\lambda}, \bar{\mu})$ is critical in the sense of Definition 2 then it is critical with respect to the index set $J = A_+ \cup \tilde{J}$ for some $\tilde{J} \subset A_0$ in the sense of [18].

4 Slacks reformulation of inequality constraints

Consider the approach of applying sSQP not directly to (1), but rather to the equality-constrained reformulation

$$\begin{aligned} & \text{minimize}_{(x,s)} f(x) \\ & \text{subject to} \quad h(x) = 0, \quad g(x) + s^2 = 0, \end{aligned} \tag{44}$$

where $s \in \mathbf{R}^m$ is the vector of slack variables and the square is meant componentwise. In the context of sSQP this “brute force” reformulation is not unreasonable, especially since constraint qualifications are not required for convergence. According to Theorem 1, the only assumption needed for local superlinear convergence of the corresponding version of sSQP consists of having the dual starting point close to a multiplier which is noncritical for problem (44). Understanding what the latter means in terms of the original problem (1) leads to assumptions different from [9, Theorem 5], i.e., different from SOSC (3) for (1). In particular, for a given stationary point \bar{x} of problem (1), and for a given $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$, the needed assumptions do not even subsume condition (9) for (1), see Example 3 (note again that (9) is not necessary for optimality of \bar{x} in problem (1) in the absence of constraint qualifications).

Defining the mapping

$$\tilde{h} : \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}^l \times \mathbf{R}^m, \quad \tilde{h}(x, s) = (h(x), g(x) + s^2),$$

the problem (44) can be written as

$$\begin{aligned} & \text{minimize}_{(x,s)} f(x) \\ & \text{subject to} \quad \tilde{h}(x, s) = 0, \end{aligned}$$

and the Lagrangian $\tilde{L} : (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m) \rightarrow \mathbf{R}$ of this problem has the form

$$\begin{aligned} \tilde{L}((x, s), (\lambda, \mu)) &= f(x) + \langle (\lambda, \mu), \tilde{h}(x, s) \rangle \\ &= L(x, \lambda, \mu) + \langle \mu, s^2 \rangle, \end{aligned}$$

where L is the Lagrangian of (1). Stationary points and Lagrange multipliers of problem (44) are thus characterized by the Lagrange optimality system

$$\frac{\partial \tilde{L}}{\partial (x, s)}((x, s), (\lambda, \mu)) = 0, \quad \tilde{h}(x, s) = 0, \tag{45}$$

where for any $(x, s) \in \mathbf{R}^n \times \mathbf{R}^m$ and $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ it holds that

$$\frac{\partial \tilde{L}}{\partial (x, s)}((x, s), (\lambda, \mu)) = \begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda, \mu) \\ 2(\text{diag } \mu)s \end{pmatrix}.$$

It is evident that stationarity of \bar{x} in problem (1) implies stationarity of (\bar{x}, \bar{s}) in problem (44), with $\bar{s} = \sqrt{-g(\bar{x})}$, where the square root is applied componentwise. Moreover, any $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}$ is a Lagrange multiplier of problem (44), associated with its stationary point (\bar{x}, \bar{s}) .

By a direct computation, for any $(x, s) \in \mathbf{R}^n \times \mathbf{R}^m$ and $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ it holds that

$$\tilde{h}'(x, s) = \begin{pmatrix} h'(x) & 0 \\ g'(x) & 2 \operatorname{diag} s \end{pmatrix},$$

$$\frac{\partial^2 \tilde{L}}{\partial(x, \sigma)^2}((x, s), (\lambda, \mu)) = \begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(x, \lambda, \mu) & 0 \\ 0 & 2 \operatorname{diag} \mu \end{pmatrix}.$$

The sSQP iteration system for problem (44)

$$\begin{pmatrix} \frac{\partial^2 \tilde{L}}{\partial(x, \sigma)^2}((x^k, s^k), (\lambda^k, \mu^k)) & (\tilde{h}'(x^k, s^k))^T \\ \tilde{h}'(x^k, s^k) & -\sigma_k I \end{pmatrix} \begin{pmatrix} x - x^k \\ s - s^k \\ \lambda - \lambda^k \\ \mu - \mu^k \end{pmatrix}$$

$$= - \begin{pmatrix} \frac{\partial \tilde{L}}{\partial x}((x^k, s^k), (\lambda^k, \mu^k)) \\ \tilde{h}(x^k, s^k) \end{pmatrix}$$

can thus be written in the form

$$\begin{aligned} & \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) + (h'(x^k))^T(\lambda - \lambda^k) + (g'(x^k))^T(\mu - \mu^k) \\ & = - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k), \\ & 2(\operatorname{diag} \mu^k)(s - s^k) + 2(\operatorname{diag} s^k)(\mu - \mu^k) = -2(\operatorname{diag} \mu^k)s^k, \\ & h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = -h(x^k), \\ & g'(x^k)(x - x^k) + 2(\operatorname{diag} s^k)(s - s^k) - \sigma_k(\mu - \mu^k) = -g(x^k) - (s^k)^2. \end{aligned} \tag{46}$$

As before, we take $\sigma_k = \sigma((x^k, s^k), (\lambda^k, \mu^k))$, with $\sigma : (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m) \rightarrow \mathbf{R}_+$ being the residual of the optimality system (45):

$$\begin{aligned} \sigma((x, s), (\lambda, \mu)) &= \left\| \begin{pmatrix} \frac{\tilde{\partial} L}{\partial(x, s)}((x, s), (\lambda, \mu)) \\ \tilde{h}(x, s) \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda, \mu) \\ 2(\operatorname{diag} \mu)s \\ h(x) \\ g(x) + s^2 \end{pmatrix} \right\|. \end{aligned} \tag{47}$$

Proposition 5 *The Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated to the stationary point $(\bar{x}, \sqrt{-g(\bar{x})})$ of problem (44) is critical if and only if either $\bar{\mu}_i = 0$ for some $i \in A$,*

or the Lagrange multiplier $(\bar{\lambda}, \bar{\mu}_A)$ associated to the stationary point \bar{x} of the equality-constrained problem

$$\begin{aligned} & \text{minimize}_x f(x) \\ & \text{subject to } h(x) = 0, \quad g_A(x) = 0 \end{aligned} \tag{48}$$

is critical.

Proof By rearranging the variables (if necessary), we obtain the equalities

$$\tilde{h}'(\bar{x}, \bar{\sigma}) = \begin{pmatrix} h'(\bar{x}) & 0 & 0 \\ g'_A(\bar{x}) & 0 & 0 \\ g'_N(\bar{x}) & 0 & 2 \operatorname{diag} \sqrt{-g_N(\bar{x})} \end{pmatrix}$$

and

$$\frac{\partial^2 \tilde{L}}{\partial(x, \sigma)^2}((\bar{x}, \bar{\sigma}), (\bar{\lambda}, \bar{\mu})) = \begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu}) & 0 & 0 \\ 0 & 2 \operatorname{diag} \bar{\mu}_A & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, the multiplier $(\bar{\lambda}, \bar{\mu})$ is critical for problem (44) if and only if there exist $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$, $(\xi, \zeta) \neq (0, 0)$, and $(y, z) \in \mathbf{R}^l \times \mathbf{R}^m$ such that

$$\begin{aligned} & \xi \in \ker h'(\bar{x}) \cap \ker g'_A(\bar{x}), \quad g'_N(\bar{x})\xi + 2 \left(\operatorname{diag} \sqrt{-g_N(\bar{x})} \right) \zeta_N = 0, \\ & \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi = (h'(\bar{x}))^T y + (g'_A(\bar{x}))^T z_A \\ & (\operatorname{diag} \bar{\mu}_A)\zeta_A = 0, \quad z_N = 0. \end{aligned} \tag{49}$$

If there exists $i \in A$ such that $\bar{\mu}_i = 0$, the latter system is evidently satisfied with $\xi = 0, y = 0, z = 0$, and with all components of ζ equal to zero, except for the i th component, which can be arbitrary. Thus, in this case, $(\bar{\lambda}, \bar{\mu})$ is critical.

Suppose now that $\bar{\mu}_A > 0$. Then the first equation in the last line of (49) is equivalent to $\zeta_A = 0$. Note also that for every $\xi \in \mathbf{R}^n$ there exists the unique ζ_N satisfying the second equation (49), and in particular, $\zeta_N = 0$ provided $\xi = 0$. It follows that in this case, $(\bar{\lambda}, \bar{\mu})$ is critical for problem (44) if and only if there exist $\xi \in \mathbf{R}^n \setminus \{0\}$ and $(y, z_A) \in \mathbf{R}^l \times \mathbf{R}^{|A|}$ such that

$$\xi \in \ker h'(\bar{x}) \cap \ker g'_A(\bar{x}), \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi = (h'(\bar{x}))^T y + (g'_A(\bar{x}))^T z_A,$$

which means that the multiplier $(\bar{\lambda}, \bar{\mu}_A)$ is critical for problem (48). □

The next result follows immediately from Proposition 5 and Theorem 1, but some comments are in order. First, observe that under the strict complementarity condition $\bar{\mu}_A > 0$, the multiplier $(\bar{\lambda}, \bar{\mu}_A)$ is noncritical for the equality-constrained problem

(48) if and only if the multiplier $(\bar{\lambda}, \bar{\mu})$ is noncritical for the original problem (1) (in the sense of Definition 2). Note also that even though sSQP is applied here to the equality-constrained reformulation (44), under the assumptions of Theorem 2 the limit (λ^*, μ^*) of the dual sequence is a multiplier not only for (44) but also for the original problem (1). This is because the multiplier sets in (1) and (44) differ only in the lack of nonnegativity of the μ -part in the latter. But here we have that $\mu^* \geq 0$, due to the proximity of the starting dual point to a multiplier satisfying strict complementarity. Hence, $(\lambda^*, \mu^*) \in \mathcal{M}$.

Theorem 2 *Let $(\bar{\lambda}, \bar{\mu})$ be a Lagrange multiplier associated with a stationary point \bar{x} of problem (1), and suppose that $\bar{\mu}_A > 0$ (strict complementarity), and the Lagrange multiplier $(\bar{\lambda}, \bar{\mu}_A)$ associated to the stationary point \bar{x} of the equality-constrained problem (48) is noncritical.*

If $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, then the sSQP iterates $\{((x^k, s^k), (\lambda^k, \mu^k))\}$ given by (46) with $s^0 = \sqrt{\max\{0, -g(x^0)\}}$, and with $\sigma_k = \sigma((x^k, s^k), (\lambda^k, \mu^k))$ defined according to (47), are well-defined and converge superlinearly to $((\bar{x}, \sqrt{-g(\bar{x})}), (\lambda^, \mu^*))$ for some $(\lambda^*, \mu^*) \in \mathcal{M}$ such that $(\lambda^*, \mu^*) \rightarrow (\bar{\lambda}, \bar{\mu})$ as $(x^0, \lambda^0, \mu^0) \rightarrow (\bar{x}, \bar{\lambda}, \bar{\mu})$.*

We next provide an example demonstrating that the assumptions of Theorem 2 may hold when condition (9) (and even more so SOSC (3)) is not satisfied with any multiplier.

Example 3 Let $n = 1, l = 0, m = 2, f(x) = -x^2/2, g(x) = (-x^3/6, x^3/6)$. Similarly to Example 2, problem (1) with this data has the unique feasible point (hence, the unique solution) $\bar{x} = 0$; this point is stationary, but unlike in Example 2, $\mathcal{M} = \mathbf{R}_+^2, C = \mathbf{R}$. It can be easily seen that for any $\bar{\mu} \in \mathcal{M}$, condition (9) does not hold. At the same time, the equality-constrained problem (48) has no critical multipliers, and hence, Theorem 2 is applicable with any $\bar{\mu} > 0$ and guarantees superlinear convergence of sSQP.

5 Consequences for local regularization methods

Two different CQ-free local algorithms for problem (1) with inequality constraints were developed in [16,27]. Instead of converting inequalities into equalities via slack variables, both of these methods are based on the following approach: first, identify the constraints active at the solution (along iterations of some generic globally convergent scheme), and second, apply a regularization or stabilization technique (different for the two methods) to the resulting equality-constrained problem. For both methods it was demonstrated that the two stages (identification of active constraints and the equality-constrained phase) can be constructed and justified under SOSC (3). The main message of this section is that these conclusions actually remain valid with SOSC replaced by the weaker assumption that the multiplier in question is noncritical.

Identification of active constraints in [16,27] is based on the error bound (8). According to Remark 1 and Proposition 3, if $(\bar{\lambda}, \bar{\mu})$ is a noncritical Lagrange multiplier associated with a stationary point \bar{x} of problem (1), then this error bound holds

with some $L > 0$ for all $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. According to [6], this in turn implies that for any $\theta \in (0, 1)$, the set

$$J = J(x^0, \lambda^0, \mu^0) = \left\{ i = 1, 2, \dots, m \mid \left(\sigma(x^0, \lambda^0, \mu^0) \right)^\theta \geq -g_i(x^0) \right\}$$

coincides with $A = A(\bar{x})$ 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 $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Thus, noncriticality of the multiplier takes care of active constraints identification, (at least) as a matter of theory.

The approach of [27] can essentially be seen as applying sSQP to the equality-constrained problem (43), starting from $(x^0, \lambda^0, \mu_j^0)$. According to Remark 3, since $(\bar{\lambda}, \bar{\mu})$ is noncritical, the multiplier (λ^0, μ_j^0) associated with the stationary point \bar{x} of problem (43) is noncritical as well (recall that for (x^0, λ^0, μ^0) close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, it holds that $J = A = A_+ \cup A_0$). Therefore, by Theorem 1, we obtain local superlinear convergence of the algorithm of [27] under the only assumption that $(\bar{\lambda}, \bar{\mu})$ is noncritical.

In [16], instead of applying sSQP, it is suggested to apply the standard Newton method to the following regularized Lagrange optimality system of problem (43):

$$f'(x) + (h'(x))^T \lambda + (g'_j(x))^T \mu_j = 0, \quad \begin{pmatrix} h(x) \\ g_j(x) \end{pmatrix} + \Pi \begin{pmatrix} \lambda - \lambda^0 \\ \mu_j - \mu_j^0 \end{pmatrix} = 0. \quad (50)$$

Here $\Pi = \Pi(x^0, \lambda^0, \mu_j^0)$ is a linear operator approximating the orthogonal projector onto $(\text{im } \Lambda(\bar{x}))^\perp$, where for $x \in \mathbf{R}^n$ we set

$$\Lambda(x) = \begin{pmatrix} h'(x) \\ g'_j(x) \end{pmatrix}.$$

By ‘‘approximating’’ we mean that $\bar{\Pi} = \Pi(\bar{x}, \bar{\lambda}, \bar{\mu}_j)$ is the specified orthogonal projector, and the mapping $(x^0, \lambda^0, \mu_j^0) \rightarrow \Pi(x^0, \lambda^0, \mu_j^0)$ is continuous at $(\bar{x}, \bar{\lambda}, \bar{\mu}_j)$. Given $(x^0, \lambda^0, \mu_j^0)$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the procedure for computing Π relies on the singular-value decomposition of $\Lambda(x^0)$. We emphasize that once computed on the first iteration, Π (as well as (λ^0, μ_j^0)) remains fixed over the Newton iterations applied to (50). Of course, any globalization strategy when coupled with this local algorithm may require many singular-value decompositions before the local phase is activated. We refer the reader to [16] for details, and for a comparison of this approach with the one from [27]. The procedure for computing Π also requires an error bound similar to (16), but for problem (43). The latter is further equivalent to the assumption that $(\bar{\lambda}, \bar{\mu}_j)$ is a noncritical multiplier for problem (43) (see Remark 1), which is again implied by the assumption that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier for problem (1) (see Remark 3).

Furthermore, it was demonstrated in [16] that for Π defined above and any $(x^0, \lambda^0, \mu_j^0)$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu}_j)$, there exists the unique solution of system (50) near $(\bar{x}, \bar{\lambda}, \bar{\mu}_j)$. Moreover, this unique solution has the form $(\bar{x}, \hat{\lambda}, \hat{\mu}_j)$ for some multiplier $(\hat{\lambda}, \hat{\mu}_j)$ (depending on $(x^0, \lambda^0, \mu_j^0)$) associated to the stationary

point \bar{x} of problem (43), and $(\hat{\lambda}, \hat{\mu}_J) \rightarrow (\bar{\lambda}, \bar{\mu}_J)$ as $(x^0, \lambda^0, \mu_J^0) \rightarrow (\bar{x}, \bar{\lambda}, \bar{\mu}_J)$. It remains to show that the Jacobian of system (50) is nonsingular at $(\bar{x}, \hat{\lambda}, \hat{\mu}_J)$, so that the Newton method applied to this system would converge superlinearly.

Proposition 6 *If $(\bar{\lambda}, \bar{\mu}_J)$ is a noncritical multiplier associated with a stationary point \bar{x} of problem (43) then for any $(x^0, \lambda^0, \mu_J^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^{|J|}$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu}_J)$, the Jacobian of the corresponding system (50) is nonsingular at its solution $(\bar{x}, \hat{\lambda}, \hat{\mu}_J)$ (the latter being the unique solution of (50) close to $(\bar{x}, \bar{\lambda}, \bar{\mu}_J)$).*

Proof Suppose that there exists a sequence $\{(x^k, \lambda^k, \mu_J^k)\}$ convergent to $\{(\bar{x}, \bar{\lambda}, \bar{\mu}_J)\}$ and such that for each k , the Jacobian of system (50) is singular at $(\bar{x}, \hat{\lambda}, \hat{\mu}_J)$ for the corresponding $(\hat{\lambda}, \hat{\mu}_J) = (\hat{\lambda}^k, \hat{\mu}_J^k)$ (depending on $(x^k, \lambda^k, \mu_J^k)$). This means that for each k , there exists $(\xi^k, \eta^k, \zeta_J^k) \neq 0$ such that

$$f''(\bar{x})\xi^k + (h''(\bar{x})\xi^k)^T \hat{\lambda}^k + (g_J''(\bar{x})\xi^k)^T \hat{\mu}_J^k + (\Lambda(\bar{x}))^T \begin{pmatrix} \eta^k \\ \zeta_J^k \end{pmatrix} = 0,$$

$$\Lambda(\bar{x})\xi^k + \Pi \begin{pmatrix} \eta^k \\ \zeta_J^k \end{pmatrix} = 0.$$

Normalizing $(\xi^k, \eta^k, \zeta_J^k)$, and passing onto the limit along the appropriate subsequence, we obtain the existence of $(\xi, \eta, \zeta_J) \neq 0$ such that

$$f''(\bar{x})\xi + (h''(\bar{x})\xi)^T \bar{\lambda} + (g_J''(\bar{x})\xi)^T \bar{\mu}_J + (\Lambda(\bar{x}))^T \begin{pmatrix} \eta \\ \zeta_J \end{pmatrix} = 0,$$

$$\Lambda(\bar{x})\xi + \bar{\Pi} \begin{pmatrix} \eta \\ \zeta_J \end{pmatrix} = 0 \tag{51}$$

(recall that $(\hat{\lambda}^k, \hat{\mu}_J^k) \rightarrow (\bar{\lambda}, \bar{\mu}_J)$ as $\{(x^k, \lambda^k, \mu_J^k)\}$ tends to $\{(\bar{x}, \bar{\lambda}, \bar{\mu}_J)\}$). Since $\bar{\Pi}$ is the orthogonal projector onto $(\text{im } \Lambda(\bar{x}))^\perp$, the second equality in (51) is equivalent to saying that

$$\Lambda(\bar{x})\xi = 0, \quad \bar{\Pi} \begin{pmatrix} \eta \\ \zeta_J \end{pmatrix} = 0. \tag{52}$$

If $\xi \neq 0$ then the first equalities in (51) and (52) contradict the assumption that $(\bar{\lambda}, \bar{\mu}_J)$ is a noncritical multiplier (cf. (19)).

Assuming now that $\xi = 0$, the first equality in (51) yields

$$(\Lambda(\bar{x}))^T \begin{pmatrix} \eta \\ \zeta_J \end{pmatrix} = 0,$$

which means that $(\eta, \zeta_J) \in \ker(\Lambda(\bar{x}))^T = (\text{im } \Lambda(\bar{x}))^\perp$. At the same time, the second equality in (52) means that $(\eta, \zeta_J) \in \text{im } \Lambda(\bar{x})$. It follows that $(\eta, \zeta_J) = 0$, which again gives a contradiction. \square

The presented argument justifies local superlinear convergence of the algorithm proposed in [16] under the only assumption that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier.

6 Concluding remarks

We have put together precise relationships between upper Lipschitzian behavior of solutions of KKT systems under canonical perturbations, error bounds for KKT systems, critical Lagrange multipliers, second-order optimality conditions, and solvability of sSQP subproblems. It has been shown that for problems where inequality constraints are present, the second-order sufficient optimality condition in [9, Theorem 5] cannot be replaced by the weaker assumption that the Lagrange multiplier is noncritical for proving superlinear convergence of sSQP iterations. However, this relaxation of assumptions is possible for problems with equality constraints only. We have also established that noncriticality of the multiplier can be used instead of second-order sufficiency in the regularization methods of [16, 27] for achieving superlinear convergence despite nonuniqueness of multipliers. An interesting open question concerns primal (rather than primal-dual) convergence rate of sSQP iterates; see [8] for details.

Acknowledgments We are grateful to the two anonymous referees for useful comments that helped us to improve the original version of the manuscript.

References

1. Boggs, B.T., Tolle, J.W.: Sequential quadratic programming. *Acta Numerica* **4**, 1–51 (1996)
2. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* **29**, 161–186 (1994)
3. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer, New York (2000)
4. Daryina, A.N., Izmailov, A.F., Solodov, M.V.: A class of active-set Newton methods for mixed complementarity problems. *SIAM J. Optim.* **15**, 409–429 (2004/2005)
5. Debreu, G.: Definite and semidefinite quadratic forms. *Econometrica* **20**, 295–300 (1952)
6. Facchinei, F., Fischer, A., Kanzow, C.: On the accurate identification of active constraints. *SIAM J. Optim.* **9**, 14–32 (1999)
7. Facchinei, F., Pang, J.-S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)
8. Fernández, D., Izmailov, A.F., Solodov, M.V.: Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. *SIAM J. Optim.* (2009, to appear)
9. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program.* **125**, 47–73 (2010)
10. Finsler, P.: Über das vorkommen definiten und semidefiniten formen und scharen quadratischer formen. *Commentarii Mathematici Helvetica* **94**, 188–192 (1937)
11. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program.* **94**, 91–124 (2002)
12. Hager, W.W.: Stabilized sequential quadratic programming. *Comput. Optim. Appl.* **12**, 253–273 (1999)
13. Hager, W., Gowda, M.: Stability in the presence of degeneracy and error estimation. *Math. Program.* **85**, 181–192 (1999)
14. Izmailov, A.F.: On the analytical and numerical stability of critical Lagrange multipliers. *Comput. Math. Math. Phys.* **45**, 930–946 (2005)
15. Izmailov, A.F.: Solution sensitivity for Karush–Kuhn–Tucker systems with nonunique Lagrange multipliers. *Optimization* (2008). doi:[10.1080/02331930802434922](https://doi.org/10.1080/02331930802434922)

16. Izmailov, A.F., Solodov, M.V.: Newton-type methods for optimization problems without constraint qualifications. *SIAM J. Optim.* **16**, 210–228 (2004)
17. Izmailov, A.F., Solodov, M.V.: On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. *Math. Program.* **117**, 271–304 (2009)
18. Izmailov, A.F., Solodov, M.V.: Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. *Comput. Optim. Appl.* **42**, 231–264 (2009)
19. Izmailov, A.F., Solodov, M.V.: On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. *Math. Program.* (2009). doi:[10.1007/s10107-009-0279-4](https://doi.org/10.1007/s10107-009-0279-4)
20. Kyparisis, J.: On uniqueness of Kuhn–Tucker multipliers in nonlinear programming. *Math. Program.* **32**, 242–246 (1985)
21. Li, D.-H., Qi, L.: Stabilized SQP method via linear equations. Application on Mathematics. Technical Report AMR00/5. University of New South Wales, Sydney (2000)
22. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.* **7**, 37–47 (1967)
23. Mostafa, E.M.E., Vicente, L.N., Wright, S.J.: Numerical behavior of a stabilized SQP method for degenerate NLP problems. In: Blik, C., Jermann, C., Neumaier A. (eds.) *Lecture Notes in Computer Science, Global Optimization and Constraint Satisfaction*, vol. 2861, pp. 123–141. Springer, Berlin (2003)
24. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.* **11**, 253–275 (1998)
25. Wright, S.J.: Modifying SQP for degenerate problems. *SIAM J. Optim.* **13**, 470–497 (2002)
26. Wright, S.J.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. *Math. Program.* **95**, 137–160 (2003)
27. Wright, S.J.: An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM J. Optim.* **15**, 673–696 (2005)