On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers

A. F. Izmailov · M. V. Solodov

Abstract It has been previously demonstrated that in the case when a Lagrange multiplier associated to a given solution is not unique, Newton iterations [e.g., those of sequential quadratic programming (SQP)] have a tendency to converge to special multipliers, called critical multipliers (when such critical multipliers exist). This fact is of importance because critical multipliers violate the second-order sufficient optimality conditions, and this was shown to be the reason for slow convergence typically observed for problems with degenerate constraints (convergence to noncritical multipliers results in superlinear rate despite degeneracy). Some theoretical and numerical validation of this phenomenon can be found in Izmailov and Solodov (Comput Optim Appl 42:231–264, 2009; Math Program 117:271–304, 2009). However, previous studies concerned the basic forms of Newton iterations. The question remained whether the attraction phenomenon still persists for relevant modifications, as well as in professional implementations. In this paper, we answer this question in the affirmative by presenting numerical results for the well known MINOS and SNOPT software

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

packages applied to a collection of degenerate problems. We also extend previous theoretical considerations to the linearly constrained Lagrangian methods and to the quasi-Newton SQP, on which MINOS and SNOPT are based. Experiments also show that in the stabilized version of SQP the attraction phenomenon still exists but appears less persistent.

Keywords Constrained optimization · Degenerate constraints · Second-order sufficiency · Newton method · SQP · MINOS · SNOPT

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

1 Introduction

Consider the problem

$$
\begin{aligned}
&\text{minimize } f(x)\\
&\text{subject to } F(x) = 0,
\end{aligned}
\tag{1}
$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function and $F : \mathbf{R}^n \to \mathbf{R}^l$ is a smooth mapping. Specifically, we assume that f and F 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 Lagrange optimality system

$$
0 = \left(\frac{\partial L}{\partial x}(x, \lambda), \; F(x)\right),
\tag{2}
$$

where

$$
L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(x, \lambda) = f(x) + \langle \lambda, \; F(x)\rangle
$$

is the Lagrangian function of problem (1).

The set of Lagrange multipliers $\Lambda(\bar{x})$ associated with a stationary point $\bar{x}$ of (1) is given by

$$
\Lambda(\bar{x}) = \{\lambda \in \mathbf{R}^l \,|\, (F'(\bar{x}))^{\mathrm{T}}\lambda = -f'(\bar{x})\}.
\tag{3}
$$

We are interested in behavior of the dual part of the sequence generated by numerical algorithms for solving (1), and its influence on the primal rate of convergence. At issue, therefore, is the case when there is more than one possible limit for the (reasonably behaving) dual sequence, i.e., $\Lambda(\bar{x})$ is not a singleton (in particular, the classical regularity condition rank $F'(\bar{x}) = l$ does *not* hold). As had been shown in [10,11], in such cases properties of the dual limit are much more relevant for the speed of convergence of Newton-type methods than degeneracy of constraints as such. To this end, the following definition plays a central role.

Definition 1 A multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is called *critical* if

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

and *noncritical* otherwise.

Since $\text{im}(F'(\bar{x}))^{\mathrm{T}} = (\ker F'(\bar{x}))^{\perp}$, it is immediate that criticality implies that

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

In particular, the second-order sufficient condition for optimality (SOSC)

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \ker F'(\bar{x}) \backslash \{0\} \tag{4}$$

does *not* hold for critical multipliers. Evidently, critical multipliers form a special subclass within the multipliers violating SOSC. We note that an alternative definition would be to say that $\bar{\lambda} \in \Lambda(\bar{x})$ is a critical multiplier if, and only if, the reduced Hessian of the Lagrangian at the point $(\bar{x}, \bar{\lambda})$ is a singular matrix (see [11] or Sect. 4).

In [11], an iterative scheme was considered that includes the Newton method for the Lagrange system (2) globalized by a linesearch procedure (in the primal-dual space) for its squared residual, and the sequential quadratic programming method (SQP) globalized by a linesearch procedure (in the primal space) for some nonsmooth penalty function. It was shown in [11] that when critical multipliers exist, the generated sequence is, in a certain sense, unlikely to approach noncritical multipliers. Moreover, convergence to critical multipliers is actually *the* reason for slow convergence—if the dual limit were to be noncritical then the rate of primal convergence would have been superlinear. Those conclusions were also confirmed by numerical experiments for a simple implementation of SQP [10,11].

A natural and important question (mentioned, among others, by one of the referees of [10]) is whether the attraction phenomenon shows not only for simple implementations of the basic Newton method but also for its relevant modifications, and for sophisticated professional implementations. And if so, whether it still causes lack of superlinear convergence. In this paper, we give an affirmative answer for two well known and widely used algorithms. One is the linearly constrained Lagrangian method [5,13,17], on which the MINOS package [14] is based. The other is a class of quasi-Newton SQP methods related, in particular, to the SNOPT package [6]. For both packages, convincing numerical results on a collection of degenerate problems from [10] will be reported. Numerical results also indicate that for the stabilized version of SQP [2,4,8,19] attraction to critical multipliers still exists but appears to be somewhat less persistent.

The rest of the paper is organized as follows. In Sect. 2, we show that the linearly constrained Lagrangian method, as well as quasi-Newton and stabilized SQP, can all be treated within a unified inexact SQP framework under different assumptions on

the perturbation terms. Our theoretical analysis starts with Sect. 3, where we consider pure SQP for a problem with quadratic objective and constraints in the case of full degeneracy. This is done in order to show what can be considered, in some sense, an ideal theoretical statement concerning convergence to critical multipliers. Analysis of the general inexact SQP scheme is given in Sect. 4. It is much more complex and leads to formally somewhat weaker statements (which are, nevertheless, enough to make our point). Section 5 explains how the notion of critical multipliers, stated in Definition 1 for the case of equality constraints, extends to problems that involve inequality constraints as well. The paper concludes with Sect. 6 that presents numerical results for MINOS, SNOPT, and stabilized SQP.

In our theoretical analysis, we shall make use of a version of the Implicit Function Theorem (e.g., [12, Theorem 1.2.2 and Remark 1.2.2]), which we state next for future reference.

Theorem 1 *Let* $\Psi : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ *be a mapping such that* $\Psi(\bar{\sigma}, \bar{\xi}) = 0$ *for some fixed* $\bar{\sigma} \in \mathbf{R}^s$ *and* $\bar{\xi} \in \mathbf{R}^n$, *and suppose that the following assumptions are satisfied*:

(a)   *There exists a linear operator* $\bar{H} : \mathbf{R}^n \to \mathbf{R}^n$ *such that for any* $\varepsilon > 0$

$$\|\Psi(\sigma, \xi^1) - \Psi(\sigma, \xi^2) - \bar{H}(\xi^1 - \xi^2)\| \leq \varepsilon \|\xi^1 - \xi^2\|$$

   for all $\sigma \in \mathbf{R}^s$ *close enough to* $\bar{\sigma}$, *and all* $\xi^1, \xi^2 \in \mathbf{R}^n$ *close enough to* $\bar{\xi}$;
(b) $\bar{H}$ *is nonsingular*;
(c)   *The mapping* $\sigma \to \Psi(\sigma, \bar{\xi}) : \mathbf{R}^s \to \mathbf{R}^n$ *is continuous at the point* $\bar{\sigma}$.

   Then there exist neighbourhoods $\mathcal{U}$ *of* σ_0 *in* $\mathbf{R}^s$ *and* U *of* $\bar{\xi}$ *in* $\mathbf{R}^n$, *and a constant* $c > 0$, *such that* $\forall \sigma \in \mathcal{U}$ *there exists the unique* $\xi = \xi(\sigma) \in U$ *such that* $\Psi(\sigma, \xi) = 0$, *and*

$$\|\xi - \bar{\xi}\| \leq c \|\Psi(\sigma, \bar{\xi})\|.$$

Note that, by necessity, under assumption (a) the mapping $\xi \to \Psi(\bar{\sigma}, \xi) : \mathbf{R}^n \to \mathbf{R}^n$ is strictly differentiable at $\bar{\xi}$, and $\frac{\partial \Psi}{\partial \xi}(\bar{\sigma}, \bar{\xi}) = \bar{H}$.

2 Inexact SQP framework

Our theoretical considerations will concern a perturbation of the Newtonian process for the Lagrange system (2). Specifically, let $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$ be the current iterate. The next iterate is then given by

$$(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) + (\xi^k, \eta^k), \tag{5}$$

with the direction of change (ξ^k, η^k) satisfying

$$0 = \frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi^k + (F'(x^k))^{\mathrm{T}}\eta^k + \theta^k, \tag{6}$$

$$0 = F(x^k) + F'(x^k)\xi^k + \omega^k, \tag{7}$$

where the terms $\theta^k \in \mathbf{R}^n$ and $\omega^k \in \mathbf{R}^l$ correspond to a perturbation of the Newton equation for the Lagrange system (2). It is worth to mention that the perturbed iterative process (5)–(7) is closely related to the inexact SQP framework (iSQP) of [18]. However, our assumptions on the perturbations, and the role of perturbations themselves, are different.

We shall interpret the linearly constrained Lagrangian methods (e.g., MINOS) and quasi-Newton SQP methods (e.g., SNOPT) within the framework of (5)–(7), under two different sets of assumptions on the perturbation term θ^k (for both of those methods it holds that $\omega^k = 0$). Regarding the modified version of SQP suggested in [3], as well as the stabilized SQP (sSQP; see [19] and also [2,4,8]), those two methods can also be put in the framework of (5)–(7), with $\theta^k = 0, \omega^k \neq 0$ (see [18] for details). As indicate our numerical experiments for sSQP in Sect. 6, the attraction phenomenon appears less present for perturbations of this kind.

Let Π be the orthogonal projector onto $\ker F'(\bar{x})$ in $\mathbf{R}^n$, and let I_n be the identity operator in $\mathbf{R}^n$. In Sect. 2.1, we shall show that an iteration of the linearly constrained Lagrangian method can be put in the form of iSQP with $\omega^k = 0$ and $\theta^k = \theta(x^k, \lambda^k, \xi^k)$, where $\theta : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$ is a mapping such that $\theta(\bar{x}, \bar{\lambda}, 0) = 0$, and for all $\bar{\lambda} \in \mathbf{R}^l$, the following holds:

(LCL1) There exists a constant $c > 0$ such that the inequality

$$\|(I_n - \Pi)(\theta(x, \lambda, \xi^1) - \theta(x, \lambda, \xi^2))\| \leq c\|\xi^1 - \xi^2\|$$

holds for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close to $(\bar{x}, \bar{\lambda})$ and all $\xi^1, \xi^2 \in \mathbf{R}^n$ close to 0, and $\|(I_n - \Pi)\theta(x, \lambda, 0)\|$ tends to zero as (x, λ) tends to $(\bar{x}, \bar{\lambda})$.

(LCL2) For each $\varepsilon > 0$, the inequalities

$$\|\Pi(\theta(x, \lambda, \xi^1) - \theta(x, \lambda, \xi^2))\| \leq \varepsilon\|\xi^1 - \xi^2\|$$

and

$$\|\Pi\theta(x, \lambda, 0)\| \leq \varepsilon\|x - \bar{x}\|$$

hold for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ and all $\xi^1, \xi^2 \in \mathbf{R}^n$ close enough to 0.

In Sect. 2.2, we discuss that iterations of quasi-Newton SQP methods can be put in the form of iSQP with $\omega^k = 0$ and $\theta^k = \Theta_k \xi^k$, where $\Theta_k : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator for each k, and it is natural to expect that the following assumptions are satisfied, at least along some appropriate subsequences:

(QN1) The sequence $\{(I_n - \Pi)\Theta_k\}$ is bounded.
(QN2) It holds that $\|\Pi\Theta_k\| \to 0$ as $k \to \infty$.

Then Sect. 4 shows that under any of the two sets of assumptions on θ^k, if the dual part of the sequence satisfying (5)–(7) approaches a noncritical multiplier then the primal rate of convergence is superlinear. Thus, the lack of superlinear convergence that is typical for problems with nonunique multipliers is an indication of convergence

to critical ones. This is confirmed by numerical results for MINOS and SNOPT in Sect. 6.

2.1 Linearly constrained Lagrangian methods

Given an iterate (x^k, λ^k), linearly constrained Lagrangian methods [5,13,17] generate the next iterate (x^{k+1}, λ^{k+1}) as follows: x^{k+1} is a stationary point of the subproblem

$$
\begin{aligned}
&\text{minimize } f(x) + \langle \lambda^k, F(x) \rangle + \frac{c_k}{2} \| F(x) \|^2 \\
&\text{subject to } F(x^k) + F'(x^k)(x - x^k) = 0,
\end{aligned}
\tag{8}
$$

while $\lambda^{k+1} = \lambda^k + \eta^k$, where η^k is a Lagrange multiplier of (8) associated to x^{k+1}. In the original method [17], $c_k = 0$ for all k. In practice, it is often important to employ $c_k > 0$ [14]. For asymptotic analysis, one can consider that $c_k \geq 0$ is fixed for all k sufficiently large, which happens under natural assumptions (see, e.g., discussion in [5]).

Let $\xi^k = x^{k+1} - x^k$. By the optimality conditions for (8), we have that

$$
0 = F(x^k) + F'(x^k)\xi^k,
\tag{9}
$$

and

$$
\begin{aligned}
0 &= f'(x^{k+1}) + (F'(x^{k+1}))^{\mathrm{T}}\lambda^k + c_k(F'(x^{k+1}))^{\mathrm{T}}F(x^{k+1}) + (F'(x^k))^{\mathrm{T}}\eta^k \\
&= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) + c_k(F'(x^{k+1}))^{\mathrm{T}}F(x^{k+1}) + (F'(x^k))^{\mathrm{T}}\eta^k \\
&= \frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi^k + (F'(x^k))^{\mathrm{T}}\eta^k + \theta^k,
\end{aligned}
$$

where we have defined

$$
\begin{aligned}
\theta^k &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi^k + c_k(F'(x^{k+1}))^{\mathrm{T}}F(x^{k+1}) \\
&= \frac{\partial L}{\partial x}(x^k + \xi^k, \lambda^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi^k \\
&\quad + c_k(F'(x^k + \xi^k))^{\mathrm{T}}(F(x^k + \xi^k) - F(x^k) - F'(x^k)\xi^k),
\end{aligned}
$$

with the last equality following from (9). This shows that the step (ξ^k, η^k), given by solving the subproblem (8), can be written in the iSQP form (6) and (7), with the specified θ^k and with $\omega^k = 0$.

Assuming that $c_k = c$ for all k large enough, we can write $\theta^k = \theta(x^k, \lambda^k, \xi^k)$, where $\theta : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$,

$$
\begin{aligned}
\theta(x, \lambda, \xi) &= \frac{\partial L}{\partial x}(x + \xi, \lambda) - \frac{\partial L}{\partial x}(x, \lambda) - \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi \\
&\quad + c(F'(x + \xi))^{\mathrm{T}}(F(x + \xi) - F(x) - F'(x)\xi).
\end{aligned}
$$

By the Mean-Value Theorem, one can directly verify that for any $\bar{\lambda} \in \mathbf{R}^l$, this mapping $\theta(\cdot)$ satisfies the assumptions (LCL1), (LCL2) (and in particular, it holds that $\theta(x, \lambda, 0) = 0 \forall (x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$).

2.2 Quasi-Newton SQP

Quasi-Newton SQP methods generate the next iterate (x^{k+1}, λ^{k+1}) by computing a stationary point and an associated Lagrange multiplier of the subproblem

$$
\begin{aligned}
&\text{minimize } \langle f'(x^k), x - x^k \rangle + \tfrac{1}{2} \langle H_k(x - x^k), x - x^k \rangle \\
&\text{subject to } F(x^k) + F'(x^k)(x - x^k) = 0,
\end{aligned}
\tag{10}
$$

where H_k is a symmetric matrix approximating the Hessian of the Lagrangian in a certain sense (e.g., see discussion in [1, Ch. 18]).

Let $\xi^k = x^{k+1} - x^k$ and $\eta^k = \lambda^{k+1} - \lambda^k$. By the optimality conditions for (10), we immediately obtain that

$$
0 = F(x^k) + F'(x^k)\xi^k,
$$

and

$$
\begin{aligned}
0 &= f'(x^k) + H_k \xi^k + (F'(x^k))^{\mathrm{T}} \lambda^{k+1} \\
&= \frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi^k + (F'(x^k))^{\mathrm{T}} \eta^k + \theta^k,
\end{aligned}
$$

where

$$
\theta^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \right) \xi^k.
$$

In particular, the iteration can be put in the iSQP form of (6) and (7), with the specified θ^k and with $\omega^k = 0$.

Observe that θ^k depends linearly on ξ^k and we can write $\theta^k = \Theta_k \xi^k$, where $\Theta_k : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator defined by the relation

$$
\Theta_k \xi =
\begin{cases}
(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k))\xi, & \text{if } \xi \in \text{span}\{\xi^k\}, \\
0, & \text{if } \xi \in (\text{span}\{\xi^k\})^{\perp}.
\end{cases}
\tag{11}
$$

It is natural to assume that $\{H_k\}$ is a bounded sequence. Then, along any bounded subsequence of $\{(x^k, \lambda^k)\}$, it holds that the corresponding subsequence of $\{\Theta^k\}$ is bounded. This implies the assumption (QN1). Furthermore, let us assume that the following Dennis–Moré-type condition holds (cf. [15, (18.62)]):

$$
\Pi \left(H_k - \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) \right) \xi^k = o(\|\xi^k\|).
\tag{12}
$$

Evidently, (11′′) and (12) imply (QN2) along any subsequence of $\{(x^k, \lambda^k)\}$ convergent to $(\bar{x}, \bar{\lambda})$.

2.3 Stabilized SQP

In sSQP [2,4,8,19], the next iterate (x^{k+1}, λ^{k+1}) is computed as a stationary point of the subproblem

$$\text{minimize } \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k), x - x^k \right\rangle + \frac{\sigma_k}{2} \|\lambda\|^2$$
$$\text{subject to } F(x^k) + F'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0,$$

where $\sigma_k > 0$ is the dual stabilization parameter. The choice of this parameter is based on computing violation of optimality conditions by the point (x^k, λ^k). It can be easily checked that this iteration fits the iSQP framework (5)–(7) with $\theta^k = 0$ and $\omega^k = -\sigma_k \eta^k$.

3 Basic SQP for the quadratic case

We start our theoretical considerations with the simplest case of SQP without perturbation terms, applied to the problem with quadratic objective and quadratic constraints that are fully degenerate at a solution. This is done in order to derive a statement which is, in a sense, the strongest possible for our purposes. Specifically, under some natural assumptions, in this case we can prove that if the dual sequence approaches a noncritical multiplier then SOSC does not hold at $\bar{x}$ for *any* $\lambda \in \Lambda(\bar{x})$. Thus, if SOSC holds for *some* multiplier, convergence to a noncritical one is simply impossible (under the given assumptions). This statement is stronger than what has been established previously for this case [11, Proposition 1].

Consider

$$f(x) = \frac{1}{2}\langle Ax, x \rangle, \quad F(x) = \frac{1}{2}B[x, x],$$

where $A(= f''(0))$ is a symmetric $n \times n$-matrix, and $B(= F''(0)) : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping. The stationary point of interest is $\bar{x} = 0$, and $\Lambda(\bar{x}) = \mathbf{R}^l$.

For each $\lambda \in \mathbf{R}^l$ set

$$H(\lambda) = \frac{\partial^2 L}{\partial x^2}(0, \lambda).$$

Note that in the present setting (since $\ker F'(0) = \mathbf{R}^n$), a multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is critical if, and only if, $H(\bar{\lambda})$ is a singular matrix.

The SQP iteration (i.e., (5)–(7) with $\theta^k = 0$ and $\omega^k = 0$) can be written in the form

$$H(\lambda^k)x^{k+1} + \left(B[x^k]\right)^{\mathrm{T}}\left(\lambda^{k+1} - \lambda^k\right) = 0, \tag{13}$$

$$B\left[x^k,\, x^{k+1}\right] = \frac{1}{2}B\left[x^k,\, x^k\right], \tag{14}$$

with $B[x]$ standing for the linear operator from $\mathbf{R}^n$ to $\mathbf{R}^l$ ($l \times n$ matrix), defined by

$$B[x]\xi = B[x,\, \xi],\quad \xi \in \mathbf{R}^n.$$

Proposition 1 *Let* $\{(x^k,\, \lambda^k)\}$ *be a sequence satisfying* (13) *and* (14). *Suppose further that there exists an infinite set* $K \subset \{0,\, 1,\, \ldots\}$ *possessing the following properties*:

(A1) *The subsequence* $\{(x^k,\, \lambda^k) \mid k \in K\}$ *converges to* $(\bar{x},\, \bar{\lambda})$ *with* $\bar{x} = 0$ *and some* $\bar{\lambda} \in \Lambda(\bar{x})$.

(A2) *The subsequence* $\{\lambda^{k+1} \mid k \in K\}$ *converges to the same* $\bar{\lambda}$.

Then for $k \in K$ *it holds that*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})x^k,\, x^{k+1}\right\rangle = o(\|x^k\|\|x^{k+1}\|), \tag{15}$$

and if $\bar{\lambda}$ *is a noncritical multiplier, then*

$$x^{k+1} = o(\|x^k\|), \tag{16}$$

$$B[x^k,\, x^k] = o(\|x^k\|^2). \tag{17}$$

Proof First note that, by (A1) and (A2), we have that

$$\eta^k = \lambda^{k+1} - \lambda^k \to 0 \text{ as } k \to \infty,\quad k \in K. \tag{18}$$

Multiplying both sides of (13) by x^k, we obtain the estimate

$$\begin{aligned}
\left\langle H(\lambda^k)x^k,\, x^{k+1}\right\rangle &= -\left\langle \lambda^{k+1} - \lambda^k,\, B[x^k,\, x^k]\right\rangle \\
&= -2\left\langle \lambda^{k+1} - \lambda^k,\, B[x^k,\, x^{k+1}]\right\rangle \\
&= o\left(\|x^k\|\|x^{k+1}\|\right),
\end{aligned}$$

where the second equality is by (14) and the third is by (18). This proves (15).

Furthermore, if $\bar{\lambda}$ is noncritical then $H(\bar{\lambda})$ is a nonsingular matrix. Then $H(\lambda^k)$ is nonsingular for all $k \in K$ large enough, and the tail of the sequence $\{(H(\lambda^k))^{-1}\}$ is well-defined and bounded. For such k, from (13) we derive

$$\begin{aligned}
x^{k+1} &= -(H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&= o(\|x^k\|),
\end{aligned}$$

where the last equality is again by (18). This proves (16). By substituting the latter into (14), we finally obtain (17). □

The assumptions (A1) and (A2) are natural for the question being investigated — the influence of the dual limit on the primal convergence rate. Essentially they just say that there is convergence in some specific sense. They hold trivially if the whole primal-dual sequence converges. In the latter case, Proposition 1 implies that if the dual limit $\bar{\lambda}$ is noncritical, then the primal trajectory converges to $\bar{x} = 0$ superlinearly (see (16)). In addition, two asymptotic relations (15) and (17) hold. The second relation implies that $\{x^k\}$ converges to $\bar{x} = 0$ tangentially to the null set of the quadratic mapping corresponding to B.

We next explain why the behaviour characterized by these two relations is highly unlikely to occur.

Remark 1 The following argument needs an additional technical assumption: Suppose that there exists an infinite subset $\tilde{K}$ of K such that $x^k \neq 0$ and $x^{k+1} \neq 0 \, \forall k \in \tilde{K}$, the subsequence $\{x^k / \|x^k\| \mid k \in \tilde{K}\}$ converges to some $\xi \in \mathbf{R}^n$, and the subsequence $\{x^{k+1} / \|x^{k+1}\| \mid k \in \tilde{K}\}$ converges either to ξ or to $-\xi$. Note that this is automatic if the entire sequence $\{x^k\}$ converges to $\bar{x} = 0$ tangentially to a direction $\xi \neq 0$, which is quite reasonable numerical behaviour that we found to be typical in our experiments. Then (15) and (17) imply that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0, \bar{\lambda})\xi, \, \xi \right\rangle = 0, \quad B[\xi, \, \xi] = 0.$$

We then further obtain that

$$\begin{aligned}\langle f''(0)\xi, \, \xi \rangle &= \langle f''(0)\xi, \, \xi \rangle + \langle \bar{\lambda}, \, F''(0)[\xi, \, \xi]\rangle \\ &= \left\langle \frac{\partial^2 L}{\partial x^2}(0, \, \bar{\lambda})\xi, \, \xi \right\rangle \\ &= 0,\end{aligned}$$

and hence,

$$\begin{aligned}\left\langle \frac{\partial^2 L}{\partial x^2}(0, \lambda)\xi, \, \xi \right\rangle &= \langle f''(0)\xi, \, \xi \rangle + \langle \lambda, \, F''(0)[\xi, \, \xi]\rangle \\ &= 0 \quad \forall \lambda \in \Lambda(\bar{x}).\end{aligned}$$

The latter means that SOSC (4) does not hold for *any* multiplier associated with $\bar{x}$. Moreover, the following weaker SOSC also does not hold at $\bar{x}$:

$$\forall \xi \in \ker F'(\bar{x}) \backslash \{0\} \, \exists \lambda \in \Lambda(\bar{x}) \text{ such that } \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \lambda)\xi, \, \xi \right\rangle > 0.$$

Moreover, the corresponding weak SOSC for maximizers cannot hold as well. The conclusion is that, under the stated assumptions, if SOSC holds (even in the weaker

form than (4)!) then convergence to a noncritical multiplier cannot occur. Of course, if the assumptions are violated (which therefore must be the case if critical multipliers do not exist, for example) then convergence to noncritical multipliers becomes a possibility. However, this situation is unlikely and is indeed very rare in practice.

However, if we perturb the data by higher-order terms, the argument given in Remark 1 becomes less clean. More precisely, asymptotic relations (16) and (17) in Proposition 1 remain valid (in the case of convergence to a noncritical multiplier), but (15) takes the cruder form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0, \bar{\lambda})x^k, x^{k+1} \right\rangle = o(\|x^k\|\|x^{k+1}\|) + o(\|x^k\|^2),$$

and generally, this does not lead to the needed formal conclusions. The same happens for other kinds of perturbations, and in particular, for those associated with linearly constrained Lagrangian methods and quasi-Newton SQP. Furthermore, passing to the general case of degeneracy leads more-or-less to the same effect, and in our numerical experiments for random problems with linear-quadratic data (see [11]), some examples of convergence to noncritical multipliers were indeed encountered beyond the fully degenerate case. That is why in the analysis of iSQP below, we concentrate mainly on superlinear rate of primal convergence in the case of dual convergence to a noncritical multiplier, without attempting the argument along the lines of Remark 1.

4 Theoretical analysis of inexact SQP

We can assume, without loss of generality, that $\bar{x} = 0$ and $f(\bar{x}) = 0$. Since convergence to $\bar{x} = 0$ is part of our setting and we are investigating properties at the limit, we can locally represent f and F in the following form:

$$f(x) = \langle a, x \rangle + \frac{1}{2}\langle Ax, x \rangle + r(x), \quad F(x) = B_1 x + \frac{1}{2}B_2[x, x] + R(x),$$

where $a \in \mathbf{R}^n$; A is a symmetric $n \times n$ matrix; B_1 is an $l \times n$ matrix; $B_2 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping; function $r : \mathbf{R}^n \to \mathbf{R}$ and mapping $R : \mathbf{R}^n \to \mathbf{R}^l$ are twice differentiable near 0, their second derivatives are continuous at 0, and $r(0) = 0$, $r'(0) = 0, r''(0) = 0, R(0) = 0, R'(0) = 0, R''(0) = 0$. For this setting, the Lagrange system (2) takes the form

$$a + B_1^{\mathrm{T}}\lambda + H(\lambda)x + r'(x) + (R'(x))^{\mathrm{T}}\lambda = 0, \quad B_1 x + \frac{1}{2}B_2[x, x] + R(x) = 0. \quad (19)$$

In what follows, we consider $a \in \mathrm{im}\, B_1^{\mathrm{T}}$, so that the point $(\bar{x}, \lambda)$ with $\bar{x} = 0$ is a solution of the system (19) for any $\lambda \in \Lambda(\bar{x})$, where $\Lambda(\bar{x})$ is an affine set parallel to $\ker B_1^{\mathrm{T}}$.

The step (ξ^k, η^k) in the iSQP framework (6) and (7) is then given by the following relations:

$$H_k\xi^k + B_1^{\mathrm{T}}\eta^k + (B_2[x^k])^{\mathrm{T}}\eta^k + r''(x^k)\xi^k + (R''(x^k)[\xi^k])^{\mathrm{T}}\lambda^k + (R'(x^k))^{\mathrm{T}}\eta^k$$
$$= -a - B_1^{\mathrm{T}}\lambda^k - H_k x^k - r'(x^k) - (R'(x^k))^{\mathrm{T}}\lambda^k - \theta^k, \qquad (20)$$

$$B_1\xi^k + B_2[x^k, \xi^k] + R'(x^k)\xi^k = -B_1x^k - \frac{1}{2}B_2[x^k, x^k] - R(x^k) - \omega^k, \quad (21)$$

where $H_k = H(\lambda^k)$.

For each $x \in \mathbf{R}^n$, define the decomposition $x = x_1 + x_2$, $x_1 \in (\ker B_1)^\perp = \operatorname{im} B_1^{\mathrm{T}}$, $x_2 \in \ker B_1$. Similarly, for each $y \in \mathbf{R}^l$, define the decomposition $y = y_1 + y_2$, $y_1 \in \operatorname{im} B_1$, $y_2 \in (\operatorname{im} B_1)^\perp = \ker B_1^{\mathrm{T}}$. Along with the orthogonal projector Π onto $\ker B_1$ in $\mathbf{R}^n$, define the orthogonal projector P onto $\ker B_1^{\mathrm{T}}$ in $\mathbf{R}^l$. For each $\lambda \in \mathbf{R}^l$, define $\hat{H}(\lambda)$ as the (symmetric) matrix of the quadratic form

$$x_2 \to \frac{\partial^2 L}{\partial x^2}(\bar{x}, \lambda)[x_2, x_2] = \langle Ax_2, x_2 \rangle + \langle \lambda, B_2[x_2, x_2] \rangle : \ker B_1 \to \mathbf{R},$$

that is,

$$\hat{H}(\lambda)x_2 = \Pi H(\lambda)x_2, \quad x_2 \in \ker B_1. \qquad (22)$$

The matrix $\hat{H}(\lambda)$ can be regarded as the reduced Hessian of the Lagrangian function. In particular, SOSC (4) means precisely that $\hat{H}(\lambda)$ is positive definite. With this notation, $\bar{\lambda} \in \Lambda(\bar{x})$ is a critical multiplier according to Definition 1 if, and only if, the matrix $\hat{H}(\bar{\lambda})$ is singular.

We note that the assumption of minimality of the primal step in Theorem 2 below (or, more generally, sufficient closeness of the next iterate to the current one) is completely standard (and unavoidable) in local analysis of any SQP-type method (e.g., [15, Theorem 18.1], [1, Theorems 15.2 and 15.4]). In practice, it is essentially ignored, of course (it is natural to expect that if the solver applied to the subproblem takes the current iterate as its starting point, then the next iterate should not be too far from it).

Theorem 2 *Let $\{\theta^k\} \subset \mathbf{R}^n$. Let $\{(x^k, \lambda^k)\}$ be a trajectory generated by the iterative process (5) where, for each k, the primal-dual step (ξ^k, η^k) is computed as a solution of (20) and (21) with $\omega^k = 0$, and with minimal-norm primal component. Suppose further that there exists an infinite set $K \subset \{0, 1, \ldots\}$ satisfying (A1), (A2) and*

(A3) *For each $k \in K$, either $\theta^k = \theta(x^k, \lambda^k, \xi^k)$, where $\theta : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$ is a mapping such that $\theta(0, \bar{\lambda}, 0) = 0$ and assumptions (LCL1), (LCL2) hold (with $\bar{x} = 0$ and $\bar{\lambda}$ from (A1), (A2)), or $\theta^k = \Theta_k\xi^k$, where $\Theta_k : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator satisfying assumptions (QN1), (QN2).*

Then if $\bar{\lambda}$ is a noncritical multiplier, for $k \in K$ it holds that

$$x^{k+1} = o(\|x^k\|) \qquad (23)$$

and

$$PB_2[x^k, x^k] = o(\|x^k\|^2). \tag{24}$$

Proof We provide the proof for the case when $r(\cdot) \equiv 0$ and $R(\cdot) \equiv 0$, strictly in order to simplify the presentation, as it is rather cumbersome even with this simplification. The assertions remain true under the general assumptions regarding r and R that are stated above.

Suppose that $\bar{\lambda}$ is a noncritical multiplier, that is, $\hat{H}(\bar{\lambda})$ is nonsingular. For each k set $\hat{H}_k = \hat{H}(\lambda^k)$. Evidently, $\{H_k \mid k \in K\}$ converges to $H(\bar{\lambda})$, and hence, $\{\hat{H}_k \mid k \in K\}$ converges to $\hat{H}(\bar{\lambda})$, and the tail of the sequence $\{\hat{H}_k^{-1} \mid k \in K\}$ is correctly defined and bounded. From now on, we consider $k \in K$ large enough.

In order to estimate (ξ^k, η^k), we shall make use of the Liapunov–Schmidt procedure, well-known in the bifurcation theory (e.g., [7, Ch. VII]). Applying $(I_n - \Pi)$ and Π to both sides of (20), and taking into account the inclusion $a \in \text{im } B_1^T$, we obtain

$$B_1^T \eta_1^k + (I_n - \Pi)\left(H_k \xi^k + \left(B_2[x^k]\right)^T \eta^k\right) = -a - B_1^T \lambda_1^k - (I_n - \Pi)(H_k x^k + \theta^k), \tag{25}$$

$$\Pi\left(H_k \xi^k + \left(B_2[x^k]\right)^T \eta^k\right) = -\Pi\left(H_k x^k + \theta^k\right). \tag{26}$$

Furthermore, applying $(I_l - P)$ and P to both sides of (21), we obtain that

$$B_1 \xi_1^k + (I_l - P)B_2[x^k, \xi^k] = -B_1 x_1^k - \frac{1}{2}(I_l - P)B_2[x^k, x^k], \tag{27}$$

$$PB_2[x^k, \xi^k] = -\frac{1}{2}PB_2[x^k, x^k]. \tag{28}$$

Clearly, the linear operators $\mathcal{B} : (\ker B_1)^\perp = \text{im } B_1^T \to \text{im } B_1$, defined by $\mathcal{B}x_1 = B_1 x_1$, and $\mathcal{B}^* : \text{im } B_1 \to \text{im } B_1^T = (\ker B_1)^\perp$, defined by $\mathcal{B}^* y_1 = B_1^T y_1$, are invertible. It follows that for each k large enough, the linear operators

$$\mathcal{B}_k : \text{im } B_1^T \to \text{im } B_1, \quad \mathcal{B}_k x_1 = B_1 x_1 + (I_l - P)B_2[x^k, x_1],$$

and

$$\mathcal{B}_k^* : \text{im } B_1 \to \text{im } B_1^T, \quad \mathcal{B}_k^* y_1 = B_1^T y_1 + (I_n - \Pi)(B_2[x^k])^T y_1,$$

are invertible, and

$$\mathcal{B}_k^{-1} = \mathcal{B}^{-1} + O(\|x^k\|), \quad (\mathcal{B}_k^*)^{-1} = (\mathcal{B}^*)^{-1} + O(\|x^k\|). \tag{29}$$

The relation (27) evidently implies that

$$\mathcal{B}_k \xi_1^k + (I_l - P)B_2[x^k, \xi_2^k] = -\mathcal{B}_k x_1^k - \frac{1}{2}(I_l - P)B_2[x^k, x^k].$$

Applying $\mathcal{B}_k^{-1}$ to both sides of the latter equality, we now obtain that, for a fixed ξ_2^k, there exists the unique ξ_1^k satisfying (27), and

$$\xi_1^k = -x_1^k + \tilde{M}_k \xi_2^k + O(\|x^k\|^2), \tag{30}$$

where we defined the linear operator

$$\tilde{M}_k = \tilde{M}(x^k) = -\mathcal{B}_k^{-1}(I_l - P)B_2[x^k].$$

Note that $\|\tilde{M}(x)\| = O(\|x\|)$. Note also that one can write $\xi_1^k = \xi_1(x^k, \xi_2^k)$, where $\xi_1 : \mathbf{R}^n \times \ker B_1 \to (\ker B_1)^\perp$ is the affine mapping such that

$$\xi_1(x, \xi^1) - \xi_1(x, \xi^2) = \tilde{M}(x)(\xi^1 - \xi^2) \tag{31}$$

holds for all $x \in \mathbf{R}^n$ and $\xi^1, \xi^2 \in \ker B_1$, and

$$\xi_1(x, 0) = O(\|x\|). \tag{32}$$

Furthermore, the relation (25) can be written in the form

$$\mathcal{B}_k^* \eta_1^k + (I_n - \Pi)(H_k \xi^k + (B_2[x^k])^{\mathrm{T}} \eta_2^k) = -a - B_1^{\mathrm{T}} \lambda_1^k - (I_n - \Pi)(H_k x^k + \theta^k).$$

Applying $(\mathcal{B}_k^*)^{-1}$ to both sides of the latter equality, and taking into account (18), the second relation in (29), and (30), we derive that for fixed ξ_2^k and η_2^k, there exists the unique η_1^k satisfying (25) (with uniquely defined $\xi_1^k = \xi_1(x^k, \xi_2^k)$), and

$$
\begin{aligned}
\eta_1^k &= -(\mathcal{B}_k^*)^{-1}(a + B_1^{\mathrm{T}} \lambda_1^k) - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\left(\left(B_2[x^k]\right)^{\mathrm{T}} \eta_2^k + H_k(x^k + \xi^k) + \theta^k \right) \\
&= -(\mathcal{B}_k^*)^{-1} \mathcal{B}^* \left(\left(\mathcal{B}^*\right)^{-1} a + \lambda_1^k \right) \\
&\quad - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\left(\left(B_2[x^k]\right)^{\mathrm{T}} \eta_2^k + H_k(x^k + \xi^k) + \theta^k \right) \\
&= \hat{\lambda} - \lambda_1^k - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\left(H_k(x_2^k + \xi_2^k + \tilde{M}_k \xi_2^k) + \theta^k \right) + o(\|x^k\|), \tag{33}
\end{aligned}
$$

where $\hat{\lambda} = -(\mathcal{B}^*)^{-1} a \in \Lambda(\bar{x}) \cap \operatorname{im} B_1$ is the uniquely defined *normal multiplier* (the one with the smallest norm), and that by assumption (A1), $\{\lambda_1^k\} \to \hat{\lambda}$. We then further obtain that

$$
\begin{aligned}
\eta_1^k &= -(\lambda_1^k - \hat{\lambda}) - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)(H_k((I_n + \tilde{M}_k)(x_2^k + \xi_2^k) - \tilde{M}_k x_2^k) \\
&\quad + \theta^k) + o(\|x^k\|) \\
&= -(\lambda_1^k - \hat{\lambda}) + C_k(x_2^k + \xi_2^k) - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\theta^k + o(\|x^k\|), \tag{34}
\end{aligned}
$$

where we defined the linear operator

$$C_k = -(\mathcal{B}_k^*)^{-1}(I_n - \Pi)H_k(I_n + \tilde{M}_k).$$

Note that the sequence $\{C_k \mid k \in K\}$ is bounded.

By substituting (30) and (34) into (26) we obtain

$$\Pi(H_k(\xi_2^k - x_1^k + \tilde{M}_k\xi_2^k) + (B_2[x^k])^{\mathrm{T}}(\eta_2^k - (\lambda_1^k - \hat{\lambda}) + C_k(x_2^k + \xi_2^k)$$
$$-(\mathcal{B}_k^*)^{-1}(I_n - \Pi)\theta^k))$$
$$= -\Pi(H_kx^k + \theta^k + o(\|x^k\|)),$$

and hence, taking into account (18),

$$\Pi((H_k(I_n + \tilde{M}_k) + (B_2[x^k])^{\mathrm{T}}C_k))\xi_2^k = -\Pi(H_k(x^k - x_1^k) - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}$$
$$\times (I_n - \Pi)\theta^k + \theta^k + o(\|x^k\|)),$$

which can be written in the form

$$\Pi(H_k + \hat{M}_k)\xi_2^k = -\Pi\left(H_kx_2^k - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I_n - \Pi)\theta^k + \theta^k + o(\|x^k\|)\right), \quad (35)$$

where we defined the linear operator

$$\hat{M}_k = H_k\tilde{M}_k + (B_2[x^k])^{\mathrm{T}}C_k.$$

Note that $\|\hat{M}_k\| = O(\|x^k\|)$.

Observe that for each k large enough, the linear operator

$$\mathcal{H}_k : \ker B_1 \to \ker B_1, \quad \mathcal{H}_kx_2 = \Pi(H_k + \hat{M}_k)x_2,$$

is invertible, and

$$\mathcal{H}_k^{-1} = \hat{H}_k^{-1} + O(\|x^k\|),$$

where (22) was taken into account.

Consider first the case when the perturbation term $\theta^k = \theta(x^k, \lambda^k, \xi^k)$, where $\theta : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$, satisfies the assumptions (LCL1), (LCL2). We next show that for all k large enough, Eq. (26) uniquely defines ξ_2^k, and

$$\xi_2^k = O(\|x^k\|). \quad (36)$$

Observe that by the first equality in (33), one can write $\eta_1^k = \eta_1(x^k, \lambda^k, \xi_2^k, \eta_2^k)$, where $\eta_1 : \mathbf{R}^n \times \mathbf{R}^l \times \ker B_1 \times (\mathrm{im}\, B_1)^\perp \to \mathrm{im}\, B_1$ is the affine mapping such that

$$\eta_1(x, \lambda, \xi^1, \eta_2) - \eta_1(x, \lambda, \xi^2, \eta_2) = O(\|\xi^1 - \xi^2\|)$$

holds for all (x, λ) close enough to $(0, \bar{\lambda})$, all $\xi^1, \xi^2 \in \ker B_1$ close enough to 0, and all $\eta_2 \in (\text{im } B_1)^\perp$, and the value $\|\eta_1(x, \lambda, 0, \eta_2)\|$ is bounded (by an independent constant) for all (x, λ) close to $(\bar{x}, \bar{\lambda})$ and all $\eta_2 \in (\text{im } B_1)^\perp$ close to 0. (Here, we have employed the assumption (LCL1) and (31).)

Consider the mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times (\text{im } B_1)^\perp \times \ker B_1 \to \ker B_1$,

$$\Psi(x, \lambda, \eta_2, \xi_2) = \Pi(H(\lambda)\xi_2 + H(\lambda)(x + \xi_1(x, \xi_2)) + (B_2[x])^{\mathrm{T}}(\eta_1(x, \lambda, \xi_2, \eta_2) + \eta_2) + \theta(x, \lambda, \xi_1(x, \xi_2) + \xi_2)).$$

Then (25) is equivalent to

$$\Psi(x, \lambda, \eta_2, \xi_2) = 0 \tag{37}$$

with $x = x^k, \lambda = \lambda^k, \eta_2 = \eta_2^k, \xi_2 = \xi_2^k$. Note that $(0, \bar{\lambda}, 0, 0)$ is a solution of (37). We shall apply Theorem 1 to (37) at this point, treating (x, λ, η_2) as a parameter.

Let $\bar{H} = \hat{H}(\bar{\lambda})$. Employing the assumption (LCL2) and the above-established properties of $\xi_1(\cdot)$ and $\eta_1(\cdot)$, we derive that for each $\varepsilon > 0$

$$\begin{aligned}
&\|\Psi(x, \lambda, \eta_2, \xi^1) - \Psi(x, \lambda, \eta_2, \xi^2) - \bar{H}(\xi^1 - \xi^2)\| \\
&\leq \|(\hat{H}(\lambda) - \bar{H})(\xi^1 - \xi^2)\| + \|\Pi H(\lambda)(\xi_1(x, \xi^1) - \xi_1(x, \xi^2))\| \\
&\quad + \|\Pi(B_2[x])^{\mathrm{T}}(\eta_1(x, \lambda, \xi^1, \eta_2) - \eta_1(x, \lambda, \xi^2, \eta_2))\| \\
&\quad + \|\theta(x, \lambda, \xi_1(x, \xi^1) + \xi^1) - \theta(x, \lambda, \xi_1(x, \xi^2) + \xi^2)\| \\
&\leq \varepsilon \|\xi^1 - \xi^2\|
\end{aligned}$$

holds for all (x, λ) close enough to $(0, \bar{\lambda})$, all $\eta_2 \in (\text{im } B_1)^\perp$, and all $\xi^1, \xi^2 \in \ker B_1$ close enough to 0. Thus, Ψ satisfies near $(0, \bar{\lambda}, 0, 0)$ assumption (a) of Theorem 1. Moreover,

$$\frac{\partial \Psi}{\partial \xi_2}(0, \bar{\lambda}, 0, 0) = \bar{H}$$

is nonsingular, which is assumption (b) of Theorem 1. Finally, the mapping $(x, \lambda, \eta_2) \to \Psi(x, \lambda, \eta_2, 0) : \mathbf{R}^n \times \mathbf{R}^l \times (\text{im } B_1)^\perp \to \mathbf{R}^n$ is continuous at $(0, \bar{\lambda}, 0)$, which is assumption (c) of Theorem 1.

Theorem 1 now implies that for all $(x, \lambda, \eta_2) \in \mathbf{R}^n \times \mathbf{R}^l \times (\text{im } B_1)^\perp$ close enough to $(0, \bar{\lambda}, 0)$, there exists the unique $\xi_2 = \xi_2(x, \lambda, \eta_2)$ close enough to 0 and satisfying (37), and

$$\begin{aligned}
\|\xi_2\| &= O(\|\Psi(x, \lambda, \eta_2, 0)\|) \\
&= O(\Pi(H(\lambda)(x + \xi_1(x, 0)) + (B_2[x])^{\mathrm{T}}(\eta_1(x, \lambda, 0, \eta_2) + \eta_2) \\
&\quad + \theta(x, \lambda, \xi_1(x, 0)))) \\
&= O(\|x\|),
\end{aligned}$$

where we have again employed assumption (LCL2) and the above-established properties of $\xi_1(\cdot)$ and $\eta_1(\cdot)$. Since $(\xi^k, \eta^k) = ((\xi_1(x^k, \xi_2^k), \xi_2^k), \eta^k)$ is a solution of (20),

(21) with a minimal-norm ξ^k-component, it follows that ξ_2^k is uniquely defined, and satisfies (36).

Applying now the linear operator $\mathcal{H}_k^{-1}$ to both sides of (35) and using assumptions (LCL1), (LCL2) and (30), (36), we obtain that

$$
\begin{aligned}
\xi_2^k &= -x_2^k + \hat{H}_k^{-1}\Pi((B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I_n - \Pi)\theta(x^k, \lambda^k, \xi^k) - \theta(x^k, \lambda^k, \xi^k)) \\
&\quad + o(\|x^k\|) \\
&= -x_2^k + O(\|x^k\|\|\xi^k\|) + o(\|\xi^k\|) + o(\|x^k\|) \\
&= -x_2^k + o(\|x^k\|).
\end{aligned}
\tag{38}
$$

Before proceeding, we prove the same relation (38) under the second set of assumptions on θ^k, i.e., when $\theta^k = \Theta_k \xi^k$, where $\Theta_k : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator for each k, satisfying (QN1), (QN2).

Applying $\mathcal{H}_k^{-1}$ to both sides of (35) and using (18), (30) and assumptions (QN1) and (QN2), we now obtain

$$
\begin{aligned}
\xi_2^k &= -x_2^k + \hat{H}_k^{-1}\Pi((B_2[x^k])^{\mathrm{T}}((\mathcal{B}_k^*)^{-1}(I_n - \Pi)\Theta^k \xi^k) - \Theta^k \xi^k) + o(\|x^k\|) \\
&= -x_2^k - \hat{H}_k^{-1}\Pi(I_n - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I_n - \Pi))\Theta^k(\xi_2^k - x_1^k + \tilde{M}_k \xi_2^k) \\
&\quad + o(\|x^k\|) \\
&= -x_2^k + S_k \xi_2^k + o(\|x^k\|),
\end{aligned}
$$

where we defined the linear operator

$$
S_k = \hat{H}_k^{-1}\Pi(I_n - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I_n - \Pi)(I_n + \tilde{M}_k)\Theta^k.
$$

Note that $\|S_k\| \to 0$ as $k \to \infty, k \in K$. Applying now the linear operator $(I - S_k)^{-1}$ to both sides in

$$
(I_n - S_k)\xi_2^k = -x_2^k + o(\|x^k\|),
$$

we obtain (38), and in particular, (36) holds as well.

To conclude the proof (now, for both sets of assumptions on θ^k), combining (5), (30), (36) and (38), we obtain that

$$
\begin{aligned}
x^{k+1} &= x^k + \xi^k \\
&= (x_1^k + \xi_1^k) + (x_2^k + \xi_2^k) \\
&= \tilde{M}_k \xi_2^k + (x_2^k + \xi_2^k) + O(\|x^k\|^2) \\
&= o(\|x^k\|),
\end{aligned}
$$

which is (23). Moreover, by (5), we drive from (28), (30) and (36) that

$$
\frac{1}{2}PB_2[x^k, x^k] = PB_2[x^k, x^{k+1}].
$$

Combined with (23), this implies the estimate (24).   □

Assumption (A3) on the perturbation term θ^k has been justified in Sect. 2 for the methods under consideration. Theorem 2 says that under the given assumptions, if the dual limit $\bar{\lambda}$ is noncritical then, according to (23), primal steps from the points of the subsequence $\{x^k \mid k \in K\}$ give superlinear decrease of the distance to the primal solution. Since we usually do not observe superlinear convergence in practice, the reason must be precisely convergence to critical multipliers.

The following considerations also show that convergence to a noncritical multiplier can only be possible for a rather special behaviour of the iterative process, which should not be common.

Remark 2 Estimate (24) means that the primal trajectory approaches $\bar{x}$ tangentially to the null set of the quadratic mapping associated to PB_2, which is a rather special behaviour. Moreover, assuming that $r(\cdot)$ and $R(\cdot)$ are three times differentiable at 0, we can derive the estimate

$$\frac{1}{2} PB_2[x^k, x^k] = PB_2[x^k, x^{k+1}] + O(\|x^k\|^3). \tag{39}$$

Imposing again the additional technical assumption stated in Remark 1, by (24), we obtain that

$$PB_2[\xi, \xi] = 0, \tag{40}$$

and by the continuity of B_2 it then follows that for $k \in \tilde{K}$

$$PB_2 \left[\frac{x^k}{\|x^k\|}, \frac{x^{k+1}}{\|x^{k+1}\|} \right] \to 0 \text{ as } k \to \infty.$$

Combining the latter with (39), we derive the following improvement of estimate (24) for $k \in \tilde{K}$:

$$PB_2[x^k, x^k] = o(\|x^k\|\|x^{k+1}\|) + O(\|x^k\|^3). \tag{41}$$

This shows behaviour that is even more special than (24): relation (41) implies that $\{x^k/\|x^k\| \mid k \in \tilde{K}\}$ must not only converge to some $\xi \neq 0$ satisfying (40), but also the rate of this convergence must be rather high.

Theorem 2 highlights one of the two possible reasons for lack of superlinear convergence of the Newton-type methods in the degenerate case. The first reason is fairly obvious: the dual sequence may not converge at all. However, in our experience, convergence of the dual sequence is much more common than nonconvergence. And when the dual sequence converges, the reason for loss of superlinear rate becomes indeed attraction to critical multipliers: if the dual sequence were to converge to noncritical multipliers, the rate of primal convergence would have been superlinear. But, according to the discussion above, dual convergence to a noncritical multiplier can happen only for a very special behavior of the process, and numerical results in Sect. 6 demonstrate that this behaviour is completely atypical for MINOS and SNOPT.

We conclude with the analysis of sSQP [2,4,8,19]. We note that the assumption (42) below on the choice of the regularization parameters σ_k can be satisfied by a number of computable choices of measures of violation of optimality conditions that are based on so-called error bounds [16] (see also [9] for a detailed comparison of error bounds for optimality systems).

Theorem 3 *Let* $\{(x^k,\ \lambda^k)\}$ *be a trajectory generated by the iterative process* (5) *where, for each k, the primal-dual step* $(\xi^k,\ \eta^k)$ *satisfies* (20) *and* (21) *with* $\theta_k = 0$, $\omega^k = -\sigma_k\eta^k$, *and*

$$\sigma_k = O(\text{dist}(\lambda^k,\ \Lambda(\bar{x}))). \tag{42}$$

Suppose further that there exists an infinite set $K \subset \{0,\ 1,\ \ldots\}$ *satisfying (A1), (A2). Then if* $\bar{\lambda}$ *is a noncritical multiplier, for* $k \in K$ *it holds that*

$$\|x^{k+1}\| + \text{dist}(\lambda^{k+1},\ \Lambda(\bar{x})) = o(\|x^k\|) + o(\text{dist}(\lambda^k,\ \Lambda(\bar{x}))). \tag{43}$$

Proof Again, we provide the proof for the case when $r(\cdot) \equiv 0$ and $R(\cdot) \equiv 0$. Suppose that $\bar{\lambda}$ is a noncritical multiplier, and let $k \in K$ be large enough.

The argument follows the lines of the proof of Theorem 2. From (20) and (21) we obtain

$$B_1^{\mathrm{T}}\eta_1^k + (I_n - \Pi)(H_k\xi^k + (B_2[x^k])^{\mathrm{T}}\eta^k) = -a - B_1^{\mathrm{T}}\lambda_1^k - (I_n - \Pi)H_kx^k, \tag{44}$$

$$\Pi(H_k\xi^k + (B_2[x^k])^{\mathrm{T}}\eta^k) = -\Pi H_kx^k, \tag{45}$$

$$B_1\xi_1^k + (I_l - P)B_2[x^k, \xi^k] = -B_1x_1^k - \frac{1}{2}(I_l - P)B_2[x^k, x^k] + \sigma_k\eta_1^k. \tag{46}$$

From (46) it can be derived that

$$\begin{aligned}\xi_1^k &= -x_1^k + \tilde{M}_k\xi_2^k + O(\|x^k\|^2) + O(\sigma_k\|\eta_1^k\|) \\ &= -x_1^k + \tilde{M}_k\xi_2^k + O(\|x^k\|^2) + o(\sigma_k),\end{aligned} \tag{47}$$

where the last equality is by (18). Furthermore, from (18) and the relations (44) and (47) it can be derived that

$$\eta_1^k = -(\lambda_1^k - \hat{\lambda}) + C_k(x_2^k + \xi_2^k) + o(\|x^k\|) + o(\sigma_k). \tag{48}$$

Finally, by substituting (47) and (48) into (45), we then obtain

$$\xi_2^k = -x_2^k + o(\|x^k\|) + o(\sigma_k). \tag{49}$$

Combining (5), (47)–(49), we obtain that

$$
\begin{aligned}
x^{k+1} &= x^k + \xi^k \\
&= (x_1^k + \xi_1^k) + (x_2^k + \xi_2^k) \\
&= \tilde{M}_k \xi_2^k + (x_2^k + \xi_2^k) + O(\|x^k\|^2) + o(\sigma_k) \\
&= o(\|x^k\|) + o(\sigma_k), \\
\mathrm{dist}(\lambda^{k+1},\, \Lambda(\bar{x})) &= \lambda_1^{k+1} - \hat{\lambda} \\
&= \lambda_1^k + \eta_1^k - \hat{\lambda} \\
&= C_k(x_2^k + \xi_2^k) + o(\|x^k\|) + o(\sigma_k) \\
&= o(\|x^k\|) + o(\sigma_k).
\end{aligned}
$$

Employing (42), we thus obtain (43).   □

5 Extensions to mixed constraints

We next explain how the notion of critical multipliers extends to the case of problems with equality and inequality constraints. We shall follow the view of critical multipliers for mixed-constrained problems suggested in [10,11] in the context of SQP. Here we explain how this definition applies to the specific solvers we deal with in this paper, and why the analysis for equality-constrained problems provided above remains relevant in the mixed-constrained case as well.

Consider, for simplicity, the purely inequality-constrained problem

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & G(x) \le 0,
\end{aligned}
\tag{50}
$$

where $G : \mathbf{R}^n \to \mathbf{R}^m$ is a smooth mapping. Let $\bar{x}$ be a stationary point of this problem, that is, there exists a Lagrange multiplier $\bar{\mu} \in \mathbf{R}^m$ satisfying the Karush–Kuhn–Tucker optimality conditions

$$
\frac{\partial L}{\partial x}(\bar{x},\, \bar{\mu}) = 0,
$$
$$
\bar{\mu} \ge 0, \quad G(\bar{x}) \le 0, \quad \langle \bar{\mu},\, G(\bar{x}) \rangle = 0,
$$

where

$$
L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}, \quad L(x,\, \mu) = f(x) + \langle \mu,\, G(x) \rangle
$$

is the Lagrangian function of problem (50). Define the set of indices of inequality constraints active at $\bar{x}$: $I(\bar{x}) = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\}$. The following notion was suggested in [10, Definition 2.2].

Definition 2 A multiplier $\bar{\mu}$ associated with a stationary point $\bar{x}$ of problem (50) is referred to as *critical* with respect to a given index set $I \subset I(\bar{x})$, if $\bar{\mu}_i = 0$

$\forall i \in \{1, \ldots, m\}\backslash I$, and the multiplier $\bar{\mu}_I$ associated with stationary point $\bar{x}$ of the equality-constrained problem

$$\begin{array}{ll} \text{minimize } & f(x) \\ \text{subject to } & G_I(x) = 0 \end{array} \tag{51}$$

is critical for this problem in the sense of Definition 1.

When a problem contains inequality constraints, both MINOS and SNOPT reformulate them into equality constraints and simple bounds using slack variables. We therefore consider the problem in the following format:

$$\begin{array}{ll} \text{minimize } & f(x) \\ \text{subject to } & G(x) + u = 0, \ u \geq 0. \end{array} \tag{52}$$

Given $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$, the subproblem of the linearly constrained Lagrangian method applied to (52) is

$$\begin{array}{ll} \text{minimize } & f(x) + \langle \lambda^k, G(x) + u \rangle + \frac{c_k}{2}\|G(x) + u\|^2 \\ \text{subject to } & G(x^k) + G'(x^k)(x - x^k) + u = 0, \ u \geq 0, \end{array} \tag{53}$$

with optimality conditions given by

$$f'(x^{k+1}) + (G'(x^{k+1}))^{\mathrm{T}}\lambda^k + c_k(G'(x^{k+1}))^{\mathrm{T}}(G(x^{k+1}) + u^{k+1}) + (G'(x^k))^{\mathrm{T}}\eta^k = 0, \tag{54}$$

$$\lambda^k + c_k(G(x^{k+1}) + u^{k+1}) + \eta^k - \mu^{k+1} = 0, \tag{55}$$

$$G(x^k) + G'(x^k)(x^{k+1} - x^k) + u^{k+1} = 0, \tag{56}$$

$$\mu^{k+1} \geq 0, \quad u^{k+1} \geq 0, \quad \langle \mu^{k+1}, u^{k+1} \rangle = 0. \tag{57}$$

Suppose that the primal trajectory $\{x^k\}$ converges to a solution $\bar{x}$ of (52). It is quite natural to assume that the set

$$I_k = \{i = 1, \ldots, m \mid u_i^{k+1} = 0\}$$

of indices of nonnegativity constraints active at the computed stationary points of subproblems (53) remains unchanged for k sufficiently large. This is actually automatic when $\{\mu^k\}$ tends to a multiplier $\bar{\mu}$ satisfying strict complementarity, i.e., such that $\bar{\mu}_{I(\bar{x})} > 0$. In other cases, the assumption that the set I_k is asymptotically unchanged may not hold, but this still seems to be reasonable numerical behaviour, which should not be unusual, and this is confirmed by our numerical experiments. Note also that if this stabilization property does not hold, one should hardly expect convergence of the dual trajectory $\{(\lambda^k, \mu^k)\}$, and this situation falls into the other scenario of slow primal convergence.

Assuming that $I_k = I$ for all k large enough, by (57) we have that $\mu_i^k = 0$ $\forall i \in \{1, \ldots, m\} \backslash I$. Then, as can be seen from (54)–(57), a stationary point of subproblem (53) gives rise to a stationary point of the equality-constrained subproblem

$$\text{minimize } f(x) + \langle \lambda^k, \, G(x) + u \rangle - \langle \mu_I^k, \, u_I \rangle + \tfrac{c_k}{2} (\|G(x) + u\|^2 + \|u_I\|^2)$$
$$\text{subject to } G(x^k) + G'(x^k)(x - x^k) + u = 0, \; u_I = 0. \tag{58}$$

Indeed, setting $\zeta_i^k = 0$ for $i \in \{1, \ldots, m\} \backslash I$, optimality conditions for (58) can be written as

$$f'(x^{k+1}) + (G'(x^{k+1}))^{\mathrm{T}} \lambda^k + c_k (G'(x^{k+1}))^{\mathrm{T}} (G(x^{k+1}) + u^{k+1}) + (G'(x^k))^{\mathrm{T}} \eta^k = 0,$$
$$\lambda^k - \mu^k + c_k (G(x^{k+1}) + u^{k+1}) + \eta^k - \zeta^k = 0,$$
$$G(x^k) + G'(x^k)(x^{k+1} - x^k) + u^{k+1} = 0, \quad u_I^{k+1} = 0,$$

which are satisfied by a solution of (54)–(57) with $\zeta_I^k = \mu_I^{k+1} - \mu_I^k$.

Observe now that (58) is the subproblem of the linearly constrained Lagrangian method applied to the equality-constrained problem

$$\text{minimize } f(x)$$
$$\text{subject to } G(x) + u = 0, \; u_I = 0. \tag{59}$$

Thus, the trajectory $\{(x^k, \, \lambda^k, \, \mu_I^k)\}$ can be thought of as generated by the linearly constrained Lagrangian method for (59). Note finally that any critical with respect to I (in the sense of Definition 2) multiplier $\bar{\mu}$ of problem (50), associated with the stationary point $\bar{x}$, corresponds to the critical multiplier $(\bar{\mu}, \, \bar{\mu}_I)$ of problem (59), associated with the stationary point $(\bar{x}, \, -G(\bar{x}))$.

For SNOPT, the reduction to equality constraints goes the same way. The subproblems are of the form

$$\text{minimize } \langle f'(x^k), \, x - x^k \rangle + \tfrac{1}{2} \langle H_k(x - x^k), \, x - x^k \rangle$$
$$\text{subject to } G(x^k) + G'(x^k)(x - x^k) + u = 0, \; u \geq 0,$$

with optimality conditions given by

$$f'(x^k) + H_k(x^{k+1} - x^k) + (G'(x^k))^{\mathrm{T}} \lambda^{k+1} = 0,$$
$$\lambda^{k+1} = \mu^{k+1}$$
$$G(x^k) + G'(x^k)(x^{k+1} - x^k) + u^{k+1} = 0,$$
$$\mu^{k+1} \geq 0, \quad u^{k+1} \geq 0, \quad \langle \mu^{k+1}, \, u^{k+1} \rangle = 0.$$

By the same reasoning as above, if the index set $I_k = I$ is the same for all k sufficiently large, then the point $(x^{k+1}, \, \lambda^{k+1}, \, \mu_I^{k+1})$ satisfies optimality conditions for

$$\text{minimize } \langle f'(x^k), \, x - x^k \rangle + \tfrac{1}{2} \langle H_k(x - x^k), \, x - x^k \rangle$$
$$\text{subject to } G(x^k) + G'(x^k)(x - x^k) + u = 0, \; u_I = 0.$$

Evidently, the latter is the subproblem of SQP method applied to the equality-constrained problem (59).

Regarding sSQP [19], in our implementation in Sect. 6 it does not employ any transformation of the original problem. Given $(x^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m$, the subproblem of sSQP applied to (50) is

$$\text{minimize } \langle f'(x^k), x - x^k \rangle + \tfrac{1}{2} \left\langle \tfrac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x - x^k), x - x^k \right\rangle + \tfrac{\sigma_k}{2} \|\mu\|^2$$
$$\text{subject to } G(x^k) + G'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0. \tag{60}$$

The optimality system for (60) can be written in the form

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x^{k+1} - x^k) + (G'(x^k))^{\mathrm{T}}\mu^{k+1} = 0,$$
$$\mu^{k+1} \geq 0, \quad G(x^k) + G'(x^k)(x - x^k) - \sigma_k(\mu^{k+1} - \mu^k) \leq 0,$$
$$\langle \mu^{k+1}, G(x^k) + G'(x^k)(x - x^k) - \sigma_k(\mu^{k+1} - \mu^k) \rangle = 0.$$

Assuming that the set

$$I_k = \{i = 1, \ldots, m \mid G_i(x^k) + G'_i(x^k)(x^{k+1} - x^k) - \sigma_k(\mu_i^{k+1} - \mu_i^k) = 0\}$$

stabilizes, i.e., that $I_k = I$ for all k sufficiently large, we observe that the point (x^{k+1}, μ_I^{k+1}) satisfies optimality conditions for

$$\text{minimize } \langle f'(x^k), x - x^k \rangle + \tfrac{1}{2} \left\langle \tfrac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x - x^k), x - x^k \right\rangle + \tfrac{\sigma_k}{2} \|\mu_I\|^2$$
$$\text{subject to } G_I(x^k) + G'_I(x^k)(x - x^k) - \sigma_k(\mu_I - \mu_I^k) = 0.$$

The latter is the subproblem of sSQP method applied to the equality-constrained problem (51).

6 Numerical results

In this section, we report on numerical results for MINOS, SNOPT and sSQP for problems with degenerate constraints collected in [10], as well as some additional examples presented below. In order to avoid confusion, examples from [10] will be referred to as "Tests" (rather than "Examples"), but with their numbers exactly corresponding to those in [10].

In our experiments, we use the default versions of MINOS and SNOPT that come with AMPL [20] student edition; default values of parameters are used, except for some special cases specified below. Regarding sSQP, it was implemented in Matlab, and supplied with the globalization strategy based on linesearch for a nonsmooth exact penalty function. We refer the reader to [10] for details of the globalization strategy and the values of parameters; those details are not central for the questions at hand in any case. In our implementation, the use of the stabilization term of sSQP is allowed only when the residual of the KKT system becomes small enough (less then 1). In

our experience, attempts to use this term globally (far from solution) usually result in failure of convergence. We stop the process when either the distance from the current primal iterate to the solution of interest, or the residual of the KKT system, become smaller than 1e–6.

For each test problem, we performed 10 runs from randomly generated starting points such that the absolute value of each component is no greater than 20 (except for some special cases, as discussed below), and with the additional nonnegativity restrictions for multipliers corresponding to inequality constraints. Failures and the cases of convergence to stationary points distinct from those of interests were disregarded.

We start with examples where critical multipliers do exist; these are Tests 2.1–2.4, 3.2–3.5, 4.16–4.22, and the following two additional examples:

Example 1 (*communicated by C. Sagastizábal*) The equality-constrained problem

$$\text{minimize } \tfrac{1}{2}(x_1^2 + x_2^2)$$
$$\text{subject to } \tfrac{1}{2}(x_1^2 + x_2^2) - x_2 = 0, \ \tfrac{1}{2}(x_1^2 + x_2^2) + x_2 = 0$$

has the unique feasible point (hence, the unique solution) $\bar{x} = 0$, with $\mathcal{M}(\bar{x}) = \{\lambda \in \mathbf{R}^2 \mid \lambda_1 - \lambda_2 = 0\}$. The unique critical multiplier is $\bar{\lambda} = (-1/2, -1/2)$.

Example 2 (*modified Test 3.5*) The inequality-constrained problem

$$\text{minimize } x_1^2 + x_2^2 - 4x_1 x_2 + x_2^3$$
$$\text{subject to } x_1 + \tfrac{1}{2}x_2^2 \geq 0, \ x_2 - x_1^2 \geq 0, \ (x_1 + \tfrac{1}{2}x_2^2)(x_2 - x_1^2) \leq 0$$

has the unique solution $\bar{x} = 0$, with $\mathcal{M}(\bar{x}) = \{\bar{\mu} \in \mathbf{R}^3 \mid \bar{\mu}_1 = \bar{\mu}_2 = 0, \ \bar{\mu}_3 \geq 0\}$. The critical with respect to $I = \{3\}$ multipliers are $\bar{\mu} = (0, 0, 2)$ and $\bar{\mu} = (0, 0, 6)$.

Numerical results for this set of test problems are reported in Table 1. For each problem and each solver, we report three indicators, separated by slashes: the number of times superlinear convergence (of major iterates) to a noncritical multiplier was detected; the number of times superlinear convergence to a critical multiplier was detected; the number of times convergence to a critical multiplier was detected but it was not superlinear.

Table 1 puts in evidence that attraction of MINOS and SNOPT to critical multipliers is typical and, as a consequence, convergence is slow. Results for sSQP are more mixed.

We proceed with comments on the behavior of solvers for some specific test problems, where warranted. Test 2.2 has a nonisolated solution at 0; nonzero solutions are nondegenerate. Quite often MINOS major iterates converge superlinearly to nonzero solutions; since those solutions are nondegenerate, these cases were not analyzed. Otherwise, typical behavior is as follows: major iterates are attracted by 0, and approach it slowly up to the order 1e–4; then the process jumps in one step to a nearby nonzero solution (probably due to some heuristics). Anyway, the multiplier produced is very close to a critical one. SNOPT does not demonstrate any tendency of convergence to 0 for this problem: major iterates converge superlinearly to nonzero solutions, and that is why the results for this problem/solver are missing in Table 1. For Tests 2.4 and 4.20,

Table 1 Examples where there exist critical multipliers

Test problems	Algorithm		
	MINOS	SNOPT	sSQP
Test 2.1	0/0/10	0/0/10	5/0/5
Test 2.2	0/0/10		3/0/5
Test 2.3	0/0/10	0/0/10	7/0/3
Test 2.4	0/0/10	0/0/10	6/0/4
Test 3.2	0/0/10	0/0/10	6/0/4
Test 3.3	0/0/10	0/0/10	2/0/8
Test 3.4	0/0/10	0/0/10	3/0/7
Test 3.5	10/0/0	9/0/1	8/0/2
Test 4.16	0/0/10	0/0/10	6/0/4
Test 4.17	0/10/0	0/10/0	4/0/6
Test 4.18	0/0/10	0/0/10	6/0/4
Test 4.19	0/0/10	0/0/10	8/0/2
Test 4.20	2/0/8	1/0/9	4/0/6
Test 4.21	0/0/10	0/0/10	10/0/0
Test 4.22	1/2/5	8/0/2	9/0/0
Example 1	0/0/10	0/0/10	2/0/8
Example 2	0/0/10	4/0/6	0/0/0

The entries show, out of 10 random runs: superlinear convergence to a noncritical multiplier; superlinear convergence to a critical multiplier; nonsuperlinear convergence to a critical multiplier

the optimality tolerance of MINOS and SNOPT was reduced (from 1e–6 to 1e–8) in order to better separate the cases of criticality and noncriticality. Test 3.5: this is just finite termination, a very special case not appropriate for any conclusions concerning asymptotic behaviour. However, for SNOPT there are some rare cases of slow convergence to a critical multiplier. Test 4.17 is a very special problem, with nonisolated solution at 0, and with all solutions (including nonzero ones) being degenerate and possessing critical multipliers. Each time the primal solution produced by MINOS or SNOPT is different, and convergence to a critical multiplier is detected; however, convergence rate of major iterates is superlinear. For Tests 4.21 and 4.22, the size of the box where starting points are generated was reduced (from 20 to 2) to enforce convergence to the primal solution of interest. The cases of superlinear convergence for Test 4.22 look more like finite termination. There are also some cases of slow convergence of major iterates to the noncritical multiplier $(1, 2, 0, 0, 0)$. For Example 2, the size of the box where starting points are generated was reduced (from 20 to 0.2) to enforce convergence to the primal solution of interest. Otherwise MINOS usually fails, or declares that the problem is unbounded. Regarding SNOPT for this problem, there are multiple cases of superlinear convergence to noncritical multipliers.

The results for sSQP in Table 1 demonstrate that the effect of attraction to critical multipliers still exists (when evaluating the numbers, recall that critical multipliers are typically few; the usual situation is that they form a set of measure zero within the set of all multipliers), but the attraction is much less persistent. The runs clearly split into two groups. Sometimes the (globalized) process manages to enter the "good" primal-dual

Table 2 Examples with no critical multipliers: number of times superlinear convergence was detected (out of 10 random runs)

Examples	Algorithm		
	MINOS	SNOPT	sSQP
Test 3.1	10	10	10
Test 4.1	10	10	10
Test 4.2	5	5	10
Test 4.3	10	8	10
Test 4.4			10
Test 4.5	10	10	10
Test 4.6	10	10	10
Test 4.7	10	10	10
Test 4.8	10	10	10
Test 4.9	10	10	10
Test 4.11	10	10	10
Test 4.12	10	10	10
Test 4.13	10	10	10
Test 4.14	10	10	10
Test 4.15	10	10	10

region, where the stabilization term starts working properly (has the needed "size"), and then it converges superlinearly to a noncritical multiplier. However, in many cases, this does not happen, and then the process still converges slowly to a critical multiplier. Thus, by itself, sSQP does not seem to be a reliable tool for avoiding the effect of attraction.

Some comments on the behavior of sSQP for specific test problems are in order. For Test 4.16, the size of the box where starting points are generated was reduced (from 20 to 2) to enforce convergence (otherwise the primal trajectory gets stuck far from solution). For Example 2, the size of the box where starting points are generated was reduced (from 20 to 0.2) to enforce convergence to the primal solution of interest; otherwise the method usually fails. However, even for this reduced size, dual trajectories diverge to infinity (that is why there is 0/0/0 in Table 1).

We continue with test problems where there are no critical multipliers; these are Tests 3.1, 4.1–4.9, 4.11–4.15. Numerical results for these problems are reported in Table 2. For each problem and each solver, we report on the number of times superlinear convergence (of major iterates) was detected. Recall the the number of runs from different starting points for each problem and solver is 10.

Table 2 puts in evidence that when there are no critical multipliers, all three solvers usually converge superlinearly, despite degeneracy of constraints.

We conclude with comments on the behavior of solvers on some specific test problems. For Tests 3.1, 4.1, 4.3, and 4.5, superlinear rate of convergence of major iterates of MINOS and SNOPT shows up only on a small number of last steps. The same applies to some runs of MINOS for Tests 4.13 and 4.14: sometimes the superlinear rate of convergence shows up only after a very long slow run. For Test 4.4, superlinear convergence to local minimizers distinct from 0 is observed (not reported). For

Test 4.15, there are some strange cases of one-step termination of MINOS after a long slow run (immediately after singularity is encountered, resulting in severe violation of optimality criterion).

Regarding sSQP, for Test 4.14 the set of indices of active constraints of subproblems does not stabilize, and the dual sequence does not converge. For Test 4.15, the size of the box where starting points are generated was reduced (from 20 to 2) to avoid failures.

References

1. Bonnans, J.F., Gilbert, J.Ch., Lemaréchal, C., Sagastizábal, C.: Numerical Optimization: Theoretical and Practical Aspects, 2nd edn. Springer, Berlin (2006)
2. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. doi:10.1007/s10107-008-0255-4
3. Fischer, A.: Modified Wilson's method for nonlinear programs with nonunique multipliers. Math. Oper. Res. **24**, 699–727 (1999)
4. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
5. Friedlander, M.P., Saunders, M.A.: A globally convergent linearly constrained Lagrangian method for nonlinear optimization. SIAM J. Optim. **15**, 863–897 (2005)
6. Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM J. Optim. **12**, 979–1006 (2002)
7. Golubitsky, M., Schaeffer, D.G.: Singularities and Groups in Bifurcation Theory, vol. 1. Springer, New York (1985)
8. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**, 253–273 (1999)
9. Izmailov, A.F., Solodov, M.V.: Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. Math. Program. **95**, 631–650 (2003)
10. 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)
11. 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)
12. Izmailov, A.F., Tretyakov, A.A.: 2-Regular Solutions of Nonlinear Problems. Theory and Numerical Methods. Fizmatlit, Moscow (1999) (in Russian)
13. Murtagh, B.A., Saunders, M.A.: A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. Math. Program. Study **16**, 84–117 (1982)
14. Murtagh, B.A., Saunders, M.A.: MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, December (1983)
15. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
16. Pang, J.-S.: Error bounds in mathematical programming. Math. Program. **79**, 299–332 (1997)
17. Robinson, S.M.: A quadratically convergent algorithm for general nonlinear programming problems. Math. Program. **3**, 145–156 (1972)
18. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**, 470–497 (2002)
19. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
20. http://www.ampl.com/