FULL LENGTH PAPER

On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions

A. F. Izmailov · M. V. Solodov

Abstract   Assuming that the primal part of the sequence generated by a Newton-type (e.g., SQP) method applied to an equality-constrained problem converges to a solution where the constraints are degenerate, we investigate whether the dual part of the sequence is attracted by those Lagrange multipliers which satisfy second-order sufficient condition (SOSC) for optimality, or by those multipliers which violate it. This question is relevant at least for two reasons: one is speed of convergence of standard methods; the other is applicability of some recently proposed approaches for handling degenerate constraints. We show that for the class of damped Newton methods, convergence of the dual sequence to multipliers satisfying SOSC is unlikely to occur. We support our findings by numerical experiments. We also suggest a simple auxiliary procedure for computing multiplier estimates, which does not have this

A. F. Izmailov
Moscow State University, Faculty of Computational Mathematics and Cybernetics,
Department of Operations Research, Leninskiye Gori, 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

undesirable property. Finally, some consequences for the case of mixed equality and inequality constraints are discussed.

1 Introduction

Consider the problem

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

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. Stationary points of problem (1) and the associated Lagrange multipliers are characterized by the Lagrange optimality system

$$\Phi(x, \lambda) = 0, \tag{2}$$

$$\Phi : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n \times \mathbf{R}^l, \quad \Phi(x, \lambda) = \left(\frac{\partial L}{\partial x}(x, \lambda), \, F(x) \right),$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$, $L(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle$ 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 the solution set of the system of linear equations with respect to $\lambda \in \mathbf{R}^l$. Specifically,

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

Evidently, $\Lambda(\bar{x})$ is a singleton if and only if the linear independence constraint qualification (LICQ) holds: rank $F'(\bar{x}) = l$. Otherwise $\Lambda(\bar{x})$ is an affine set parallel to $\ker(F'(\bar{x}))^{\mathrm{T}} = (\mathrm{im}\, F'(\bar{x}))^{\perp}$, where $\ker A$ stands for the null subspace of a linear operator A, while $\mathrm{im}\, A$ stands for the image subspace.

We note that the situation when LICQ is violated but multipliers exist is not unusual and deserves special consideration. For example, it is known that for degenerate problems of the form of (1) multipliers associated with a local solution exist generically, provided the number of variables is large enough with respect to the number of constraints (see Lemma 1 and subsequent discussion in [4,5]). The case of violation of classical constraint qualifications has been a subject of considerable interest in the past decade, both in the general case (e.g., [1,4,8,11,12,15,16,18,32]) and in the special case of equilibrium or complementarity constraints (e.g., [2,3,9,17,22,26–28]).

Not assuming LICQ for problem (1) means the absence of strong regularity in the sense of [25], and consequently, the absence of nice sensitivity properties and of nice

convergence properties of Newton-type methods. In particular, we face the situation where any solution of the Lagrange system (2) of the form $(\bar{x}, \lambda)$, $\lambda \in \Lambda(\bar{x})$, is degenerate.

To tackle the problem of degeneracy, various local stabilization and regularization methods have been proposed in [8,11,18,21,32] (see also [7,29–31]). Despite lack of LICQ, some of those methods do achieve superlinear or quadratic convergence from points $(\tilde{x}^0, \tilde{\lambda}^0)$ sufficiently close to pairs $(\bar{x}, \bar{\lambda})$ with $\bar{\lambda} \in \Lambda(\bar{x})$ satisfying the second-order sufficient condition (SOSC) for optimality

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

Typically, underlying local algorithms for degenerate problems is the following philosophy (see, for example [18] and especially [32], where this is explicit). It is assumed that an outer strategy generates a primal–dual sequence $\{(x^k, \lambda^k)\}$ approaching $(\bar{x}, \bar{\lambda})$ with the needed property (4), which guarantees that the local phase is activated at a certain iteration k, with $(\tilde{x}^0, \tilde{\lambda}^0) = (x^k, \lambda^k)$. This local phase forces fast convergence of the sequence $\{(\tilde{x}^k, \tilde{\lambda}^k)\}$ despite the point $\bar{x}$ violating LICQ.

When the primal trajectory converges to $\bar{x}$, the question of whether or not the dual trajectory converges to $\bar{\lambda} \in \Lambda(\bar{x})$ satisfying SOSC is also relevant for speed of convergence of the primal sequence in standard algorithms (i.e., not specifically designed to handle degeneracy). It has been observed in [31, Sect. 6] that when SQP converges slowly, the apparent reason is precisely convergence of the dual sequence to multipliers violating SOSC (see also Sect. 6 and Proposition 2 below with comments preceding it). This is also consistent with all the other examples of slow convergence of SQP in the degenerate case that we found in the literature and examined. Of course, this paper is not the first one where it is realized that slow convergence of SQP in the degenerate case has to do with "bad" dual behavior. In fact, some stabilized versions of SQP achieve better convergence properties precisely by adjusting in this or that way the multipliers estimates, e.g., [7,31]. The goal of this paper is to make it clear what exactly is the problem, to define with precision the set of troublesome multipliers attracting dual iterates (these are not simply multipliers violating SOSC!), and to show that convergence to troublesome multipliers is not only something that *may* happen but, in fact, is something that *should* be expected to happen.

We note that assuming SOSC (4) for some $\bar{\lambda} \in \Lambda(\bar{x})$, the point $\bar{x}$ is an isolated local minimizer of (1) and convergence of $\{x^k\}$ to $\bar{x}$ can be expected for (good implementations of) good algorithms, even in the degenerate cases (of course, the speed of convergence is another matter, as already mentioned). Here, one can even think of $\bar{x}$ as being the unique stationary point of (1). Convergence of $\{\lambda^k\}$ to the needed $\bar{\lambda}$ satisfying SOSC, on the other hand, is by no means a given (SOSC usually holds for some multipliers but not for all of them). Without making far-reaching conclusions for other formats of optimization problems and/or other types of algorithms, we show that, at least for the class of damped Newton methods for (1) (or for (2)), the conclusion is actually negative: the dual sequence is "very unlikely" to approach multipliers satisfying the needed SOSC. Moreover, we define a special subclass of multipliers

violating SOSC, called critical multipliers, and show that noncritical multipliers cannot attract Newton iterations.

Definition 1 The multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is referred to as *critical* if

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

and *noncritical* otherwise.

Since $(\operatorname{im} F'(\bar{x}))^{\mathrm{T}} = (\ker F'(\bar{x}))^{\perp}$, 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, SOSC (4) is violated. Evidently, critical multipliers form a special subclass within the multipliers violating SOSC.

We consider the following iterative process. Let $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$ be the current iterate. The next iterate has the form

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

where (ξ^k, η^k) satisfies the Newton equation for (2), i.e.,

$$\Phi\left(x^k, \lambda^k\right) + \Phi'\left(x^k, \lambda^k\right)\left(\xi^k, \eta^k\right) = 0, \tag{6}$$

and $\{\alpha_k\}$ and $\{\beta_k\}$ are sequences of real numbers such that

$$\hat{\alpha} \geq \alpha_k \geq \check{\alpha} > 0, \quad \beta_k \geq \check{\beta} > 0 \quad \forall k. \tag{7}$$

The above iterative scheme includes the Newton method for Lagrange system (2), globalized by a linesearch procedure for its squared residual

$$\Psi : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad \Psi(x, \lambda) = \frac{1}{2} \|\Phi(x, \lambda)\|^2. \tag{8}$$

In this case, we have $\alpha_k = \beta_k \; \forall k$ in (5), and stepsizes satisfy (7) under standard assumptions.

The sequential quadratic programming (SQP) is also a special case of our scheme. In SQP, ξ^k is computed as a stationary point of

$$\begin{aligned} \text{minimize} \quad & \left\langle f'\left(x^k\right), \xi \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\left(x^k, \lambda^k\right)\xi, \xi \right\rangle \\ \text{subject to} \quad & F\left(x^k\right) + F'\left(x^k\right)\xi = 0, \end{aligned}$$

and λ^{k+1} is set to be an associated multiplier. For SQP, we have in (5) that $\beta_k = 1 \; \forall k$, while α_k is computed by a linesearch procedure for some nonsmooth penalty function. Under standard assumptions, (7) holds.

The effect of attraction of the dual sequence $\{\lambda^k\}$ to critical multipliers was first reported in [13] for a special case of the iterative scheme (5)–(7). To expose the effect in question, we start with the following simple example.

Example 1 Let $n = l = 1$, $f(x) = x^2$, $F(x) = x^2$.

We have that $\bar{x} = 0$, $\Lambda(\bar{x}) = \mathbf{R}$, the SOSC (4) is satisfied for any $\bar{\lambda} > -1$, and the *unique* critical multiplier is $\bar{\lambda} = -1$. The Newton equation (6) takes the form $(1 + \lambda^k)\xi^k + x^k\eta^k = -(1 + \lambda^k)x^k$, $2x^k\xi^k = -(x^k)^2$.

Suppose that $\lambda^k \neq -1$. Then $\xi^k = -x^k\left(1 + \eta^k/(1 + \lambda^k)\right)$, and if $x^k \neq 0$, it follows that $\eta^k = -(1 + \lambda^k)/2$, $\xi^k = -x^k/2$. Hence,

$$x^{k+1} = x^k + \alpha_k\xi^k = (1 - \alpha_k/2)\,x^k,$$
$$\lambda^{k+1} + 1 = \lambda^k + \beta_k\eta^k + 1 = (1 - \beta_k/2)\left(\lambda^k + 1\right).$$

Thus, if (7) holds with $\hat{\alpha} < 4$ and $\hat{\beta} < 4$, and $\alpha_k \neq 2$, $\beta_k \neq 2 \,\forall k$, and if $x^0 \neq 0$ and $\lambda^0 \neq -1$, then $x^k \neq 0$ and $\lambda^k \neq -1 \,\forall k$, and the trajectory $\{(x^k, \lambda^k)\}$ is correctly defined and converges linearly to $(0, -1)$. In particular, $\{\lambda^k\}$ converges to the (unique!) critical multiplier.

It turns out that the phenomenon observed in Example 1 is rather "persistent". In fact, this example was not specially constructed with some goal in mind. It is easy to see taking other small examples of problem (1) with nonunique multipliers that the dual sequence $\{\lambda^k\}$, obtained by means of the Newton equation, inevitably tends to a critical multiplier. This is quite remarkable, because the set of critical multipliers is typically very small (just one point in Example 1!), while the set of other multipliers is large (the whole space except for one point in Example 1!). It is therefore of interest to investigate and explain the observed phenomenon. In this sense, it is important to make the following observation. It cannot be possible to *prove* that $\{\lambda^k\}$ *must* converge to a critical multiplier. Indeed, if the sequence $\left\{(x^k, \lambda^k)\right\}$ happens to hit an exact solution of the Lagrange system so that $\Phi(x^{\bar{k}}, \lambda^{\bar{k}}) = 0$ for some iteration $\bar{k}$ (for example, if we start with $x^0 = \bar{x}$ and $\lambda^0 \in \Lambda(\bar{x})$), the sequence obviously would be "attracted" to this solution, in the sense that $(x^k, \lambda^k) = (x^{\bar{k}}, \lambda^{\bar{k}})$ for all $k \geq \bar{k}$. Of course, here $\lambda^{\bar{k}}$ may be noncritical (for example, if we start at $x^0 = \bar{x}$ and a noncritical $\lambda^0 \in \Lambda(\bar{x})$). This situation of "finite termination" is certainly atypical and one can disregard it as not very interesting. But since it is possible in principle, it indicates that there cannot exist a proof showing that $\{\lambda^k\}$ converges to critical multipliers. Instead, the goal should be to show that if convergence to noncritical multipliers occurs, then this implies something unlikely. This would in turn justify the observed convergence to critical multipliers. For those reasons, the required analysis is somewhat subtle and unusual.

It can be said that the phenomenon discussed in this paper joins the list of other possible negative effects that must be taken into account when Newton-type methods are implemented. For example, the well-known Maratos effect [23] in SQP, or the effect of false numerical convergence of some generalized Newton methods [24]. For satisfactory numerical performance, certain care should be taken when implementing the related algorithms. Similarly, Newton-type methods for degenerate problems also require some safeguards in order to avoid convergence to critical multipliers.

Let $\bar{x} = 0$, and assume (without loss of generality) that $f(0) = 0$. Since convergence to $\bar{x}$ is part of our setting and we are investigating properties at the limit, we can locally represent twice continuously differentiable 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 $r(0) = 0$, $r'(0) = 0$, $r''(0) = 0$, $R(0) = 0$, $R'(0) = 0$, $R''(0) = 0$. Note that under these assumptions,

$$r(x) = o\left(\|x\|^2\right), \quad r'(x) = o\left(\|x\|\right), \quad r''(x) \to 0 \text{ as } x \to 0, \tag{9}$$

$$R(x) = o\left(\|x\|^2\right), \quad R'(x) = o\left(\|x\|\right), \quad R''(x) \to 0 \text{ as } x \to 0. \tag{10}$$

For this setting, the Lagrange system (2) takes the form

$$a + B_1^{\mathrm{T}}\lambda + H(\lambda)x + r'(x) + \left(R'(x)\right)^{\mathrm{T}}\lambda = 0, \quad B_1 x + \frac{1}{2}B_2[x, x] + R(x) = 0, \tag{11}$$

where $H(\lambda)$ is the (symmetric) matrix of the quadratic form

$$x \to \frac{\partial^2 L}{\partial x^2}(0, \lambda)[x, x] = \langle Ax, x \rangle + \langle \lambda, B_2[x, x] \rangle : \mathbf{R}^n \to \mathbf{R},$$

that is,

$$H(\lambda)x = Ax + (B_2[x])^{\mathrm{T}}\lambda, \quad x \in \mathbf{R}^n,$$

with $B_2[x]$ being the linear operator from $\mathbf{R}^n$ to $\mathbf{R}^l$ ($l \times n$ matrix), defined by

$$B_2[x]\xi = B_2[x, \xi], \quad \xi \in \mathbf{R}^n.$$

In what follows, we consider $a \in \operatorname{im} B_1^{\mathrm{T}}$, so that the point $(\bar{x}, \lambda)$ with $\bar{x} = 0$ is a solution of the system (11) for any $\lambda \in \Lambda(\bar{x})$, where $\Lambda(\bar{x})$ is an affine set parallel to $\ker B_1^{\mathrm{T}}$.

The Newton direction (ξ^k, η^k) for (11) is then given by the following relations:

$$H_k\xi^k + B_1^{\mathrm{T}}\eta^k + \left(B_2[x^k]\right)^{\mathrm{T}}\eta^k + r''(x^k)\xi^k + \left(R''(x^k)[\xi^k]\right)^{\mathrm{T}}\lambda^k + \left(R'(x^k)\right)^{\mathrm{T}}\eta^k$$

$$= -a - B_1^{\mathrm{T}}\lambda^k - H_k x^k - r'(x^k) - \left(R'(x^k)\right)^{\mathrm{T}}\lambda^k, \tag{12}$$

$$B_1 \xi^k + B_2\left[x^k,\ \xi^k\right] + R'(x^k)\xi^k = -B_1 x^k - \frac{1}{2}B_2\left[x^k,\ x^k\right] - R(x^k), \quad (13)$$

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 $y = y_1 + y_2$, $y_1 \in \operatorname{im} B_1$, $y_2 \in (\operatorname{im} B_1)^\perp = \ker B_1^{\mathrm{T}}$. Let Π be the orthogonal projector onto $\ker B_1$ in $\mathbf{R}^n$, and P be the orthogonal projector 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,

$$\hat{H}(\lambda)x_2 = \Pi H(\lambda)x_2, \quad x_2 \in \ker B_1. \quad (14)$$

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.

The rest of the paper is organized as follows. We start our analysis in Sect. 2, with the case where the derivatives of f and F vanish at a solution. In Sect. 3, we extend the analysis to the general case by means of the Liapunov–Schmidt procedure. In Sect. 4 we suggest an auxiliary scheme for computing the multiplier estimates, which does not have the exhibited effect of attraction to critical multipliers. Some numerical experiments are presented in Sect. 5. Consequences for the general case of mixed equality and inequality constraints are discussed in Sect. 6. Concluding remarks are given in Sect. 7.

2 The case of complete degeneracy

In this section, we deal with the case when $a = 0$ and $B_1 = 0$, so that $f'(\bar{x}) = 0$ and $F'(\bar{x}) = 0$. In this case, $\operatorname{im}(F'(\bar{x}))^{\mathrm{T}} = (\ker F'(\bar{x}))^\perp = \{0\}$, and Definition 1 says that $\bar{\lambda} \in \Lambda(\bar{x}) = \mathbf{R}^l$ is critical if, and only if, the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x},\ \bar{\lambda})$ is singular. Or, equivalently, $H(\bar{\lambda})$ is singular (recall (14), where $\ker B_1 = \mathbf{R}^n$ and Π is the identity). We first state our result, then discuss our assumptions and conclusions, and then give the proof.

Proposition 1 *Suppose that the iterative process given by* (5)–(7) *correctly generates the trajectory* $\{(x^k,\ \lambda^k)\}$, *and 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* $\bar{\lambda} \in \Lambda(\bar{x})(= \mathbf{R}^l)$.

(A2) *The subsequence* $\{\lambda^{k+1} \mid k \in K\}$ *converges to the same* $\bar{\lambda}$.
Then either $\bar{\lambda}$ *is a critical multiplier or*

$$\|B_2[x^k, x^k]\| = o(\|x^k\|^2), \quad k \in K, \tag{15}$$

and if $x^k \neq 0 \, \forall \, k \in K$ *then*

$$B_2[\chi, \chi] = 0 \tag{16}$$

for any accumulation point $\chi \in \mathbf{R}^n \setminus \{0\}$ *of the sequence* $\{x^k/\|x^k\| \mid k \in K\}$.
Assuming (A1) and (A2), suppose that $x^k \neq 0 \, \forall \, k \in K$ *and, in addition, that*
$r(\cdot) \equiv 0, \, R(\cdot) \equiv 0,$ *and that*
(A3) *there exists an infinite subset* $\tilde{K}$ *of* K *such that* $x^{k+1} - (1 - \alpha_k)x^k \neq 0 \, \forall \, k \in$
$\tilde{K}$ *and the subsequence* $\{(x^{k+1} - (1 - \alpha_k)x^k)/\|x^{k+1} - (1 - \alpha_k)x^k\| \mid k \in$
$\tilde{K}\}$ *converges either to* χ *or to* $-\chi$, *where* χ *is the limit of the subsequence*
$\{x^k/\|x^k\| \mid k \in \tilde{K}\}$.

Then either $\bar{\lambda}$ *is a critical multiplier or* (16) *holds and the matrix* $B_2[\chi](H(\bar{\lambda}))^{-1}$
$(B_2[\chi])^{\mathrm{T}}$ *is singular.*

We next explain the result stated above.

The Assumptions (A1) and (A2) are trivial with respect to the questions we are investigating; essentially they just say that there is convergence in some sense. If $\{x^k\}$ converges to $\bar{x}$ and $\{\lambda^k\}$ converges to some compact subset of $\Lambda(\bar{x})$ and $\|\lambda^{k+1} - \lambda^k\| \to 0$, they are satisfied automatically on some subsequence K. We note that even the first assertion, obtained under those minimal assumptions, already shows quite special behavior of the primal sequence, if $\bar{\lambda}$ would be a noncritical multiplier. Indeed, if $\bar{\lambda}$ is noncritical, then (15) shows that the primal trajectory enters $\bar{x}$ *tangentially* to the null set of the quadratic mapping associated to B_2. Apparently, there are no particular reasons for this to happen, and we do not observe this behavior in our examples and experiments. We emphasize that the first assertion is obtained without "technical" assumptions, such as (A3).

The Assumption (A3) may appear somewhat questionable. Nevertheless, it is not unreasonable and can be justified for the kind of analysis we are doing. First of all, note that the required property is asked to hold only on some further subsequence $\tilde{K}$ of K, and not for all of K. Let $r(\cdot) \equiv 0$ and $R(\cdot) \equiv 0$ (this is a handy simplification but it is not all that necessary, see Remark 1 below).

We first discuss the first part of (A3), i.e., $x^{k+1} - (1 - \alpha_k)x^k \neq 0, \, k \in \tilde{K}$. From (5) to (13) it follows that

$$\begin{aligned}
2B_2\left[x^k, x^{k+1} - (1 - \alpha_k)x^k\right] &= 2\alpha_k B_2\left[x^k, \xi^k\right] + 2\alpha_k B_2\left[x^k, x^k\right] \\
&= -\alpha_k B_2\left[x^k, x^k\right] + 2\alpha_k B_2\left[x^k, x^k\right] \\
&= \alpha_k B_2\left[x^k, x^k\right].
\end{aligned}$$

Hence, according to (7), if for some k it holds that $x^{k+1} - (1 - \alpha_k)x^k = 0$ then necessarily $B_2[x^k, x^k] = 0$. In the case under consideration, this means that x^k is feasible, which is already an unusual situation for a method linearizing equality constraints with curvature. Furthermore, if the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ exists and is non-singular, and $k \in K$ is large enough, then it can be seen that $\eta^k = 0$ (this follows from (7) and from (22); the latter to be established soon, within the proof of Proposition 1). Hence, according to (12), $\xi^k = -x^k$ (note that by (A1), the existence of $(H(\bar{\lambda}))^{-1}$ implies nonsingularity of H_k for all $k \in K$ large enough). But from (5) we then obtain that $B_2[x^{k+1}, x^{k+1}] = (1 - \alpha_k)^2 B_2[x^k, x^k] = 0$, $\lambda^{k+1} = \lambda^k$. This means that $B_2[x^k, x^k] = 0$ and $\lambda^k = \bar{\lambda}$ for all $k \in K$ large enough. This is, of course, an exceptional situation (the linearization method somehow stays feasible and the multiplier is never updated). Therefore, assuming that $x^{k+1} - (1 - \alpha_k)x^k \neq 0$ for $k \in K$ (or at least on a subset $\tilde{K}$ of K) appears more than reasonable.

We now discuss the second part of (A3). The point $\tilde{x}^{k+1} = x^{k+1} - (1 - \alpha_k)x^k = \alpha_k(x^k + \xi^k)$ is obtained as a result of the full Newton step from x^k, multiplied by α_k. By the Assumption (A3), there should exist some subsequence $\tilde{K}$ of K such that $\{\tilde{x}^{k+1} \mid k \in \tilde{K}\}$ enters $\bar{x}$ tangentially to the same direction χ as $\{x^k \mid k \in \tilde{K}\}$, or tangentially to the opposite direction $-\chi$. If $\alpha_k = 1$ for k large enough (full Newton step), then $\tilde{x}^{k+1} = x^{k+1}$ for k large enough, and the required property holds if, for example, the subsequence $\{x^{k+1} \mid k \in \tilde{K}\}$ enters $\bar{x}$ tangentially to the same direction χ as the subsequence $\{x^k \mid k \in \tilde{K}\}$. Note that this holds automatically if, say, the entire sequence $\{x^k\}$ converges to $\bar{x}$ tangentially to some direction χ. While this is a technical assumption, it is certainly completely reasonable (for a reasonable algorithm in a situation when it does converge to a solution). At the very least, there is no good reason for it not to hold, at least on *some subsequence*, which is all that is needed in our analysis. We also checked whether the Assumption (A3) holds in our numerical experiments in Sect. 5, and observed that this appears to be always the case when the primal sequence converges (we examined a good portion of test problems with respect to (A3), although not all of them).

We now discuss the second assertion of Proposition 1, which is the main result. According to this assertion, if $\bar{\lambda}$ is noncritical, then it holds that the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ is singular for $\chi \in \mathbf{R}^n \setminus \{0\}$ satisfying $B_2[\chi, \chi] = 0$. This is an exceptional/unlikely situation for the following reasons. Clearly, the matrix in question cannot be singular if $H(\bar{\lambda})$ is positive or negative definite and

$$\mathrm{rank}\ B_2[\chi] = l. \tag{17}$$

Now, SOSC (4) means that $H(\bar{\lambda})$ is positive definite. In the terminology of [19], (17) means that the quadratic mapping $\xi \to B_2[\xi, \xi] : \mathbf{R}^n \to \mathbf{R}^l$ is 2-regular in the direction χ. The above always holds if this quadratic mapping is 2-regular in any nonzero direction in its null set, which is a generic property for quadratic mappings! (See [19, Proposition 1.3.7], [14, Proposition 1]).

We conclude that the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ is generically nonsingular if SOSC holds. In other words, its singularity is extremely unlikely to occur, and thus convergence to a noncritical multiplier must be an unlikely event. In addition,

by Example 3 below, the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ can be nonsingular even if $H(\bar{\lambda})$ is not positive or negative definite. Moreover, Example 2 shows that even if the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ happens to be singular (however unlikely this is in general), convergence to critical multipliers can still be observed. Putting together all those considerations, examples and numerical experiments of Sect. 5, we believe that one obtains a rather conclusive picture.

Before passing to the proof of Proposition 1, we note that

$$\|\eta^k\| = \beta_k^{-1}\|\lambda^{k+1} - \lambda^k\| \le \hat{\beta}^{-1}\|\lambda^{k+1} - \lambda^k\| \to 0, \quad k \in K, \tag{18}$$

by (5), the second condition in (7), and by (A1) and (A2).

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

From (12), where $a = 0$ and $B_1 = 0$, taking into account (9), (10), and (18), we obtain

$$\begin{aligned}
\xi^k &= -x^k - H_k^{-1}\left(B_2[x^k]\right)^{\mathrm{T}}\eta^k \\
&\quad - H_k^{-1}\left(r'(x^k) + \left(R'(x^k)\right)^{\mathrm{T}}\lambda^k + r''(x^k)\xi^k \right. \\
&\quad \left. + \left(R''(x^k)[\xi^k]\right)^{\mathrm{T}}\lambda^k + \left(R'(x^k)\right)^{\mathrm{T}}\eta^k\right) \\
&= -x^k - H_k^{-1}\left(B_2[x^k]\right)^{\mathrm{T}}\eta^k + S_k\xi^k + o\left(\|x^k\|\right),
\end{aligned}$$

where $S_k : \mathbf{R}^n \to \mathbf{R}^n$ is a linear operator satisfying $\|S_k\| \to 0$ as $k \to \infty$. Then we easily obtain that

$$\xi^k = -x^k - H_k^{-1}\left(B_2[x^k]\right)^{\mathrm{T}}\eta^k + o\left(\|x^k\|\right). \tag{19}$$

(Note that, according to (18), the second term in the right-hand side is $o(\|x^k\|)$, and thus, could be dropped here. However, we prefer to keep this term since it will play its role in the future; see below.)

By substituting (19) into (13), where $B_1 = 0$, we obtain that

$$B_2\left[x^k, \, x^k\right] + B_2\left[x^k\right]H_k^{-1}\left(B_2\left[x^k\right]\right)^{\mathrm{T}}\eta^k = \frac{1}{2}B_2\left[x^k, \, x^k\right] + o\left(\|x^k\|^2\right),$$

that is,

$$B_2\left[x^k, \, x^k\right] = -2B_2\left[x^k\right]H_k^{-1}\left(B_2[x^k]\right)^{\mathrm{T}}\eta^k + o\left(\|x^k\|^2\right).$$

Using (18), we now obtain (15). Then (16) follows immediately, and the first assertion is thus proved.

Throughout the rest of the proof, we shall make the assumption that the matrix $B_2[\chi](H(\bar\lambda))^{-1}(B_2[\chi])^{\mathrm{T}}$ is nonsingular, since otherwise the second assertion is obtained.

Set $\chi^k = x^k/\|x^k\|$. It is easy to see that if $r(\cdot) \equiv 0$, $R(\cdot) \equiv 0$ then there is no little-o term in (19). Then by (5), we obtain that

$$x^{k+1} - (1 - \alpha_k)x^k = -\alpha_k H_k^{-1}(B_2[x^k])^{\mathrm{T}}\eta^k. \tag{20}$$

Thus, by boundedness of $\{\alpha_k\}$ and of the tail of $\{H_k^{-1} \mid k \in K\}$, we conclude that there exists an independent constant $\gamma_1 > 0$ such that

$$\|x^k\|\|\eta^k\| \geq \gamma_1 \left\| x^{k+1} - (1 - \alpha_k)x^k \right\|. \tag{21}$$

From (20) it further follows that

$$B_2[x^k, \ x^{k+1} - (1 - \alpha_k)x^k] = -\alpha_k B_2[x^k] H_k^{-1}(B_2[x^k])^{\mathrm{T}}\eta^k. \tag{22}$$

Hence, by the first condition in (7) and by the nonsingularity of the matrix $B_2[\chi](H(\bar\lambda))^{-1}(B_2[\chi])^{\mathrm{T}}$, and by (21), there exists an independent constant $\gamma_2 > 0$ such that

$$\begin{aligned}
\left\| B_2\left[\chi^k, \ x^{k+1} - (1 - \alpha_k)x^k \right] \right\| &= \alpha_k \|x^k\| B_2[\chi^k] H_k^{-1} \left(B_2[\chi^k] \right)^{\mathrm{T}} \eta^k \| \\
&\geq \check\alpha \gamma_2 \|x^k\|\|\eta^k\| \\
&\geq \check\alpha \gamma_1\gamma_2 \left\| x^{k+1} - (1 - \alpha_k)x^k \right\|.
\end{aligned} \tag{23}$$

On the other hand, by the Assumption (A3), by (16), and by the continuity of B_2,

$$\left\| B_2\left[\chi^k, \ \frac{x^{k+1} - (1 - \alpha_k)x^k}{\|x^{k+1} - (1 - \alpha_k)x^k\|} \right] \right\| \to 0, \quad k \in \tilde K,$$

contradicting (23). This completes the proof. □

Remark 1 In the second assertion of Proposition 1, assumptions $r(\cdot) \equiv 0$ and $R(\cdot) \equiv 0$ can be dropped. The argument becomes somewhat less strict, but it still gives an explanation of the effect in question. Specifically, assume that r and R are three times differentiable at 0. Then it is easy to see that (9) and (10) transform into

$$r(x) = O(\|x\|^3), \quad r'(x) = O(\|x\|^2), \quad r''(x) = O(\|x\|), \tag{24}$$
$$R(x) = O(\|x\|^3), \quad R'(x) = O(\|x\|^2), \quad R''(x) = O(\|x\|), \tag{25}$$

and instead of (20), one obtains

$$x^{k+1} - (1 - \alpha_k)x^k = -\alpha_k H_k^{-1} \left(B_2[x^k] \right)^{\mathrm{T}} \eta^k + O\left(\|x^k\|^2 \right). \qquad (26)$$

Then by boundedness of $\{\alpha_k\}$ and of the tail of $\{H_k^{-1} \mid k \in K\}$, we conclude that

$$\left\| x^{k+1} - (1 - \alpha_k)x^k \right\| = O\left(\|x^k\| \|\eta^k\| \right) + O\left(\|x^k\|^2 \right). \qquad (27)$$

From (26) it further follows that

$$B_2 \left[x^k, \, x^{k+1} - (1 - \alpha_k)x^k \right] = -\alpha_k B_2[x^k] H_k^{-1} \left(B_2[x^k] \right)^{\mathrm{T}} \eta^k + O\left(\|x^k\|^3 \right).$$

By the first condition in (7) and by nonsingularity of $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$, it still holds that there exists an independent constant $\tilde{\gamma} > 0$ such that the norm of the first term in the right-hand side is not less than $\tilde{\gamma} \|x^k\|^2 \|\eta^k\|$. At the same time, the big-O term in the right-hand side depends on the properties of r and R, and generally, there is no reason for this big-O term to be little-o, or for it to be somehow related to the first term in the right-hand side. Thus, we can expect that there exists $\gamma > 0$ such that

$$\left\| B_2 \left[x^k, \, x^{k+1} - (1 - \alpha_k)x^k \right] \right\| \geq \gamma \max \left\{ \|x^k\|^2 \|\eta^k\|, \, \|x^k\|^3 \right\}. \qquad (28)$$

In particular, this is always the case when $\|x^k\| = o(\|\eta^k\|)$.

On the other hand, by the Assumption (A3), by (16), by the continuity of B_2 and by (27), we obtain

$$\left\| B_2 \left[x^k, \, x^{k+1} - (1 - \alpha_k)x^k \right] \right\|$$
$$= \left\| B_2 \left[\chi^k, \, \frac{x^{k+1} - (1 - \alpha_k)x^k}{\|x^{k+1} - (1 - \alpha_k)x^k\|} \right] \right\| \|x^k\| \|x^{k+1} - (1 - \alpha_k)x^k\|$$
$$= o\left(\max \left\{ \|x^k\|^2 \|\eta^k\|, \, \|x^k\|^3 \right\} \right), \quad k \in \tilde{K},$$

which contradicts (28).

We next present some more subtle examples, to complement what has been already observed in Example 1 and established in Proposition 1. Example 2 demonstrates that convergence to critical multipliers may still persist even when the matrix $B_2[\chi](H(\bar{\lambda}))^{-1}(B_2[\chi])^{\mathrm{T}}$ happens to be singular (this is the situation not covered by Proposition 1). Example 3 shows that this matrix can be nonsingular even without SOSC (or more generally, when $H(\bar{\lambda})$ is not sign-defined).

Example 2 Let $n = 2$, $l = 1$, $\langle Ax, x \rangle = x_1^2 - x_2^2$, $B_2[x, x] = x_1^2 - x_2^2$, $r(\cdot) \equiv 0$, $R(\cdot) \equiv 0$. Here, the unique critical multiplier is -1. Moreover, for any $\lambda \in \Lambda(\bar{x}) = \mathbf{R}$ the matrix $H(\lambda)$ is not positive or negative definite. Furthermore,

$$B_2[\chi]\,(H(\lambda))^{-1}\,(B_2[\chi])^{\mathrm{T}} = \left(\chi_1^2 - \chi_2^2\right)/(1+\lambda)$$

is singular (equals to 0) for each $\lambda \neq -1$ and each $\chi \in \mathbf{R}^n$ such that $B_2[\chi,\,\chi] = 0$. The Newton system (12), (13) takes the form

$$\left(1+\lambda^k\right)\xi_1^k + x_1^k\eta^k = -\left(1+\lambda^k\right)x_1^k, \quad \left(1+\lambda^k\right)\xi_2^k - x_2^k\eta^k = -\left(1+\lambda^k\right)x_2^k,$$
$$2\left(x_1^k\xi_1^k - x_2^k\xi_2^k\right) = -\left((x_1^k)^2 - (x_2^k)^2\right).$$

Suppose that $\lambda^k \neq -1$ and $B_2[x^k,\,x^k] \neq 0$. Then $\xi_1^k = -x_1^k(1 + \eta^k/(1+\lambda^k))$, $\xi_2^k = -x_2^k(1 + \eta^k/(1+\lambda^k))$, $\eta^k = -(1+\lambda^k)/2$. Hence, $\xi_1^k = -x_1^k/2$, $\xi_2^k = -x_2^k/2$, and

$$x_1^{k+1} = x_1^k + \alpha_k\xi_1^k = (1 - \alpha_k/2)\,x_1^k, \quad x_2^{k+1} = x_2^k + \alpha_k\xi_2^k = (1 - \alpha_k/2)\,x_2^k,$$
$$\lambda^{k+1} + 1 = \lambda^k + \beta_k\eta^k + 1 = (1 - \beta_k/2)\left(\lambda^k + 1\right).$$

Thus, if (7) holds with $\hat{\alpha} < 4$ and $\hat{\beta} < 4$, and $\alpha_k \neq 2$, $\beta_k \neq 2 \,\forall k$, and if $B_2[x^0, x^0] \neq 0$ and $\lambda^0 \neq -1$, then $B_2[x^k,\,x^k] \neq 0$ and $\lambda^k \neq -1 \,\forall k$, and moreover, the trajectory $\{(x^k,\,\lambda^k)\}$ is correctly defined and converges linearly to $(0,\,-1)$. In particular, convergence to (the unique) critical multiplier is still observed.

Example 3 Let $n = 2, l = 1, \langle Ax,\,x \rangle = x_1^2, B_2[x,\,x] = x_1^2 - x_2^2, r(\cdot) \equiv 0, R(\cdot) \equiv 0$. Here, the critical multipliers are -1 and 0. Moreover, SOSC holds with $\lambda \in (-1,\,0)$, while for $\lambda < -1$ or $\lambda > 0$ the matrix $H(\lambda)$ is not positive or negative definite. Nevertheless,

$$B_2[\chi]\,(H(\lambda))^{-1}\,(B_2[\chi])^{\mathrm{T}} = \chi_1^2/(1+\lambda) - \chi_2^2/\lambda,$$

is nonsingular (not equal to 0) for each $\lambda \notin \{-1,\,0\}$ and each $\chi \in X \setminus \{0\}$ such that $B_2[\chi,\,\chi] = 0$. The Newton system (12), (13) takes the form

$$\left(1+\lambda^k\right)\xi_1^k + x_1^k\eta^k = -\left(1+\lambda^k\right)x_1^k, \quad -\lambda^k\xi_2^k - x_2^k\eta^k = \lambda^k x_2^k,$$
$$2\left(x_1^k\xi_1^k - x_2^k\xi_2^k\right) = -\left((x_1^k)^2 - (x_2^k)^2\right).$$

Suppose that $\lambda^k \neq -1$, $\lambda^k \neq 0$, and $\lambda^k(x_1^k)^2 \neq (1+\lambda^k)(x_2^k)^2$. Then $\xi_1^k = -x_1^k(1 + \eta^k/(1+\lambda^k))$, $\xi_2^k = -x_2^k(1 + \eta^k/\lambda^k)$,

$$\eta^k = -\frac{1}{2}\lambda^k\left(1+\lambda^k\right)\frac{(x_1^k)^2 - (x_2^k)^2}{\lambda^k(x_1^k)^2 - (1+\lambda^k)(x_2^k)^2}.$$

If, e.g., $x_2^k = 0$ then it is easy to see that $\eta^k = -(1 + \lambda^k)/2$, $\xi_1^k = -x_1^k/2$, $\xi_2^k = 0$. Hence,

$$x_1^{k+1} = x_1^k + \alpha_k \xi_1^k = (1 - \alpha_k/2) x_1^k, \quad x_2^{k+1} = x_2^k + \alpha_k \xi_2^k = 0,$$
$$\lambda^{k+1} + 1 = \lambda^k + \beta_k \eta^k + 1 = (1 - \beta_k/2)\left(\lambda^k + 1\right).$$

We thus conclude that if the trajectory hits the line $x_2 = 0$ (e.g., if $x_2^0 = 0$), then the trajectory stays on this line, and moreover, its behavior with respect to variables x_1 and λ is similar to that in Example 1: under the appropriate assumptions on $\{\alpha_k\}$ and $\{\beta_k\}$, the sequence $\{(x_1^k, \lambda^k)\}$ converges linearly to $(0, -1)$.

By a similar argument it can be seen that if the trajectory hits the line $x_1 = 0$ (e.g., if $x_1^0 = 0$), then the trajectory stays on this line, and the sequence $\{(x_2^k, \lambda^k)\}$ converges linearly to $(0, 0)$.

Now suppose that $B_2[x^k, x^k] = (x_1^k)^2 - (x_2^k)^2 = 0$. Then $\eta^k = 0$, $\xi_1^k = -x_1^k$, $\xi_2^k = -x_2^k$, and hence,

$$x_1^{k+1} = x_1^k + \alpha_k \xi_1^k = (1 - \alpha_k)x_1^k, \quad x_2^{k+1} = x_2^k + \alpha_k \xi_2^k = (1 - \alpha_k)x_2^k,$$
$$\lambda^{k+1} = \lambda^k + \beta_k \eta^k = \lambda^k.$$

Thus, if (7) holds with $\hat{\alpha} < 2 \; \forall k$, the primal trajectory converges linearly to (or finitely terminates at) 0 staying within the set defined by $B_2[x, x] = 0$, while the dual trajectory does not leave λ^k *whatever this λ^k is*. (Finite termination happens in the case of full Newton step, that is, when $\alpha_k = 1$ for all k).

Note that the primal trajectory can hit the set defined by $B_2[x, x] = 0$ even if $B_2[x^0, x^0] \neq 0$. Indeed, let, e.g.,

$$\left(x_1^0 \lambda^0\right)^2 = \left(x_2^0 (1 + \lambda^0)\right)^2. \tag{29}$$

It is easy to see that in this case $(x_1^1)^2 = (x_2^1)^2$, and $\lambda^2 = \lambda^1 = -1/2$ is a noncritical multiplier.

Of course, the above situation of convergence to a noncritical multiplier is very special. In fact, numerical experiments for this problem demonstrate the following. If x^0 and/or λ^0 are slightly perturbed, so that (29) is "almost" satisfied, and if $\alpha_k = 1$ $\forall k$, then the first two steps are indeed in the direction of the point $(0, -1/2)$, but afterwards, the primal trajectory keeps on approaching 0, while the dual trajectory leaves the vicinity of $-1/2$ and eventually converges to a critical multiplier.

3 The general case

Let the Assumptions (A1) and (A2) hold, and let r and R be three times differentiable at 0. We first show that under these assumptions either $\bar{\lambda}$ is a critical multiplier or

$$\left\| P B_2 \left[x^k, x^k\right] \right\| = o\left(\|x^k\|^2\right) \tag{30}$$

for $k \in K$, where P is the orthogonal projector in $\mathbf{R}^l$ onto $\ker B_1^{\mathrm{T}}$. Recall that (A1) and (A2) imply (18).

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.

To reduce the present setting to the case of complete degeneracy, we shall use the Liapunov–Schmidt procedure (e.g., [10, Chap. VII]).

Let I_n be the identity operator in $\mathbf{R}^n$. Applying $(I_n - \Pi)$ and Π to both sides of (12), and taking into account (18), (24), (25) and the inclusion $a \in \operatorname{im} B_1^{\mathrm{T}}$, we obtain

$$
\begin{aligned}
B_1^{\mathrm{T}} \eta_1^k + (I_n - \Pi) &\left((H_k + M_k)\xi^k + (B_2[x^k])^{\mathrm{T}} \eta^k \right) \\
&= -a - B_1^{\mathrm{T}} \lambda_1^k - (I_n - \Pi) \left(H_k x^k + O(\|x^k\|^2) \right),
\end{aligned}
\tag{31}
$$

$$
\Pi \left((H_k + M_k)\xi^k + (B_2[x^k])^{\mathrm{T}} \eta^k \right) = -\Pi \left(H_k x^k + O(\|x^k\|^2) \right),
\tag{32}
$$

where we defined the linear operator

$$
M_k : \mathbf{R}^n \to \mathbf{R}^n, \quad M_k \xi = r''(x^k)\xi + \left(R''(x^k)[\xi] \right)^{\mathrm{T}}.
$$

Note that $\|M_k\| = O(\|x^k\|)$.

Let I_l be the identity operator in $\mathbf{R}^l$. Applying $(I_l - P)$ and P to both sides of (13), and taking into account (25), we obtain that

$$
\begin{aligned}
B_1 \xi_1^k + (I_l - P) &\left(B_2[x^k, \xi^k] + R'(x^k)\xi^k \right) \\
&= -B_1 x_1^k - (I_l - P) \left(\frac{1}{2} B_2[x^k, x^k] + O(\|x^k\|^3) \right),
\end{aligned}
\tag{33}
$$

$$
P \left(B_2[x^k, \xi^k] + R'(x^k)\xi^k \right) = -P \left(\frac{1}{2} B_2[x^k, x^k] + O(\|x^k\|^3) \right).
\tag{34}
$$

Recall that $\|R'(x^k)\| = O(\|x^k\|^2)$. Clearly, the linear operators $\mathcal{B} : (\ker B)^{\perp} = \operatorname{im} B_1^{\mathrm{T}} \to \operatorname{im} B_1$, defined by $\mathcal{B} x_1 = B_1 x_1$, and $\mathcal{B}^* : \operatorname{im} B_1 \to \operatorname{im} B_1^{\mathrm{T}} = (\ker B_1)^{\perp}$, defined by $\mathcal{B}^* y_1 = B_1^{\mathrm{T}} y_1$, are invertible. It follows that for each k large enough, the linear operators

$$
\mathcal{B}_k : \operatorname{im} B_1^{\mathrm{T}} \to \operatorname{im} B_1, \quad \mathcal{B}_k x_1 = B_1 x_1 + (I_l - P) \left(B_2[x^k, x_1] + R'(x^k)x_1 \right),
$$

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

are invertible, and

$$\mathcal{B}_k^{-1} = \mathcal{B}^{-1} + O\left(\|x^k\|\right), \quad (\mathcal{B}_k^*)^{-1} = (\mathcal{B}^*)^{-1} + O\left(\|x^k\|\right). \tag{35}$$

The relation (33) evidently implies that

$$\mathcal{B}_k \xi_1^k + (I_l - P)\left(B_2[x^k, \xi_2^k] + R'(x^k)\xi_2^k\right) = -\mathcal{B}_k x_1^k + O\left(\|x^k\|^2\right).$$

Applying $\mathcal{B}_k^{-1}$ to both sides of the latter equality, we now obtain that

$$\xi_1^k = -x_1^k + \tilde{M}_k \xi_2^k + O\left(\|x^k\|^2\right), \tag{36}$$

where we defined the linear operator

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

Note that $\|\tilde{M}_k\| = O(\|x^k\|)$. Furthermore, the relation (31) can be written in the form

$$\mathcal{B}_k^* \eta_1^k + (I_n - \Pi)\left((H_k + M_k)\xi^k + (B_2[x^k])^{\mathrm{T}}\eta_2^k\right)$$
$$= -a - B_1^{\mathrm{T}}\lambda_1^k - (I_n - \Pi)\left(H_k x^k + O(\|x^k\|^2)\right).$$

Applying $(\mathcal{B}_k^*)^{-1}$ to both sides of the latter equality, and taking into account the second relation in (35), and (36), we derive

$$\eta_1^k = -(\mathcal{B}_k^*)^{-1}\left(a + B_1^{\mathrm{T}}\lambda_1^k\right)$$
$$- (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\left((B_2[x^k])^{\mathrm{T}}\eta_2^k + (H_k + M_k)\xi^k + H_k x^k + O(\|x^k\|^2)\right)$$
$$= -(\mathcal{B}_k^*)^{-1}\mathcal{B}^*\left((\mathcal{B}^*)^{-1}a + \lambda_1^k\right)$$
$$- (\mathcal{B}_k^*)^{-1}(I_n - \Pi)\left((B_2[x^k])^{\mathrm{T}}\eta_2^k + (H_k + M_k)(x^k + \xi^k)\right.$$
$$\left. - M_k x^k + O(\|x^k\|^2)\right)$$
$$= \left(I_l + \tilde{M}_k^1\right)\left(\hat{\lambda} - \lambda_1^k\right) + \tilde{M}_k^2 \eta_2^k$$
$$- (\mathcal{B}_k^*)^{-1}(I_n - \Pi)(H_k + M_k)\left(x_2^k + \xi_2^k + \tilde{M}_k \xi_2^k\right) + O\left(\|x^k\|^2\right),$$

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 we defined the linear operators

$$\tilde{M}_k^1 = (\mathcal{B}_k^*)^{-1}\mathcal{B}^* - I_l,$$
$$\tilde{M}_k^2 = -(\mathcal{B}_k^*)^{-1}(I_n - \Pi)(B_2[x^k])^{\mathrm{T}}.$$

Note that $\|\tilde{M}_k^1\| = O(\|x^k\|)$, $\|\tilde{M}_k^2\| = O(\|x^k\|)$, and that by the Assumption (A1), $\{\lambda_1^k\} \to \hat{\lambda}$. We then further obtain that

$$
\begin{aligned}
\eta_1^k &= \left(I_l + \tilde{M}_k^1\right)\left(\hat{\lambda} - \lambda_1^k\right) + \tilde{M}_k^2 \eta_2^k \\
&\quad - (\mathcal{B}_k^*)^{-1}(I_n - \Pi)(H_k + M_k)\left((I_n + \tilde{M}_k)(x_2^k + \xi_2^k) - \tilde{M}_k x_2^k\right) + O(\|x^k\|^2) \\
&= \left(I_l + \tilde{M}_k^1\right)\left(\hat{\lambda} - \lambda_1^k\right) + \tilde{M}_k^2 \eta_2^k + C_k\left(x_2^k + \xi_2^k\right) + O\left(\|x^k\|^2\right),
\end{aligned}
\tag{37}
$$

where we defined the linear operator

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

Note that the sequence $\{C_k \mid k \in K\}$ is bounded. By substituting (36) and (37) into (32), and by taking into account (18), we obtain

$$
\begin{aligned}
&\Pi\left((H_k + M_k)\left(\xi_2^k - x_1^k + \tilde{M}_k \xi_2^k\right)\right. \\
&\quad \left. + (B_2[x^k])^{\mathsf{T}}\left(\eta_2^k + (I_l + \tilde{M}_k^1)(\hat{\lambda} - \lambda_1^k) + \tilde{M}_k^2 \eta_2^k + C_k(x_2^k + \xi_2^k)\right)\right) \\
&= -\Pi\left(H_k x^k + O(\|x^k\|^2)\right),
\end{aligned}
$$

and hence,

$$
\begin{aligned}
&\Pi\left(\left((H_k + M_k)(I_n + \tilde{M}_k) + (B_2[x^k])^{\mathsf{T}} C_k\right)\xi_2^k + (B_2[x^k])^{\mathsf{T}} \eta_2^k\right) \\
&= -\Pi\left(H_k\left(x^k - x_1^k\right) + \left(B_2[x^k]\right)^{\mathsf{T}}\left(\hat{\lambda} - \lambda_1^k\right) + O\left(\|x^k\|^2\right)\right),
\end{aligned}
$$

which can be written in the form

$$
\begin{aligned}
&\Pi\left((H_k + \hat{M}_k)\xi_2^k + (B_2[x^k])^{\mathsf{T}} \eta_2^k\right) \\
&= -\Pi\left(H_k x_2^k + (B_2[x^k])^{\mathsf{T}}(\hat{\lambda} - \lambda_1^k) + O(\|x^k\|^2)\right),
\end{aligned}
\tag{38}
$$

where we defined the linear operator

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

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

Similarly, by substituting (36) into (34), we obtain

$$
\begin{aligned}
&P\left(B_2\left[x^k, \xi_2^k - x_1^k + \tilde{M}_k \xi_2^k\right] + R'(x^k)\left(\xi_2^k - x_1^k + \tilde{M}_k \xi_2^k\right)\right) \\
&= -P\left(\frac{1}{2}B_2[x^k, x^k] + O(\|x^k\|^3)\right),
\end{aligned}
$$

and hence,

$$P\left(B_2[x^k, \xi_2^k] + B_2[x^k, \tilde{M}_k\xi_2^k] + R'(x^k)(I_n + \tilde{M}_k)\xi_2^k\right)$$
$$= -P\left(\frac{1}{2}B_2[x^k, x_1^k + x_2^k] - \frac{1}{2}B_2[x^k, 2x_1^k] + O(\|x^k\|^3)\right),$$

which can be written in the form

$$P\left(B_2[x^k, \xi_2^k] + N_k\xi_2^k\right) = -P\left(\frac{1}{2}B_2[x^k, x_2^k - x_1^k] + O(\|x^k\|^3)\right), \qquad (39)$$

where we defined the linear operator

$$N_k = B_2[x^k]\tilde{M}_k + R'(x^k)(I_n + \tilde{M}_k).$$

Note that $\|N_k\| = O(\|x^k\|^2)$.

Observe that for each k large enough, the linear operator

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

is invertible, and

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

where (14) was taken into account. Applying $\mathcal{H}_k^{-1}$ to both sides of (38) and using (18), we now obtain

$$\xi_2^k = -x_2^k - \hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\left(\hat{\lambda} - \lambda_1^k + \eta_2^k\right) + O(\|x^k\|^2)$$
$$= -x_2^k - \hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k + O\left(\|x^k\|\|\lambda_1^k - \hat{\lambda}\|\right) + O(\|x^k\|^2). \qquad (40)$$

In particular,

$$\xi_2^k = O(\|x^k\|). \qquad (41)$$

By substituting (40) into (39), we then derive the equality

$$P B_2\left[x^k, x_2^k + \hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k\right]$$
$$= P\left(\frac{1}{2}B_2\left[x^k, x_2^k - x_1^k\right] - N_k\left(x_2^k + \hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k\right)\right)$$
$$\quad + O\left(\|x^k\|^3\right) + O\left(\|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|\right)$$
$$= \frac{1}{2}P B_2\left[x^k, x_2^k - x_1^k\right] + O\left(\|x^k\|^3\right) + O\left(\|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|\right),$$

which can be written in the form

$$PB_2[x^k, x^k] = -2PB_2[x^k]\hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k + O(\|x^k\|^3) + O(\|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|).$$

From (18) and from the convergence of $\{\lambda_1^k\}$ to $\hat{\lambda}$, we now obtain (30). And consequently,

$$PB_2[\chi, \chi] = 0. \tag{42}$$

Furthermore, by (5), (36), (40) and (41), and by the boundedness of $\{\alpha_k\}$ and of the tail of $\{H_k^{-1} \mid k \in K\}$, we conclude that

$$
\begin{aligned}
x^{k+1} - (1 - \alpha_k)x^k &= \alpha_k\left(x^k + \xi^k\right) = \alpha_k\left(x_1^k + \xi_1^k\right) + \alpha_k\left(x_2^k + \xi_2^k\right) \\
&= -\alpha_k\hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k \\
&\quad + O\left(\|x^k\|^2\right) + O\left(\|x^k\|\|\lambda_1^k - \hat{\lambda}\|\right),
\end{aligned} \tag{43}
$$

and in particular

$$\left\|x^{k+1} - (1 - \alpha_k)x^k\right\| = O\left(\|x^k\|\|\eta_2^k\|\right) + O\left(\|x^k\|^2\right) + O\left(\|x^k\|\|\lambda_1^k - \hat{\lambda}\|\right). \tag{44}$$

From (43) it further follows that

$$
\begin{aligned}
&PB_2\left[x^k, x^{k+1} - (1 - \alpha_k)x^k\right] \\
&= -\alpha_k PB_2[x^k]\hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k + O\left(\|x^k\|^3\right) + O\left(\|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|\right).
\end{aligned}
$$

If we suppose that the linear operator $PB_2[\chi]\hat{H}_k^{-1}\Pi(B_2[\chi])^{\mathrm{T}}$ is nonsingular on $\ker B_1^{\mathrm{T}}$ (that is, the null space of this operator on this subspace is trivial) then by the same reasons as in Remark 1 we may expect that there exists $\gamma > 0$ such that

$$
\begin{aligned}
&\left\|PB_2\left[x^k, x^{k+1} - (1 - \alpha_k)x^k\right]\right\| \\
&\geq \gamma \max\left\{\|x^k\|^2\|\eta_2^k\|, \|x^k\|^3, \|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|\right\}.
\end{aligned} \tag{45}
$$

On the other hand, by the Assumption (A3), by (42), by the continuity of B_2, and by (44) we obtain

$$
\begin{aligned}
&\left\|PB_2\left[x^k, x^{k+1} - (1 - \alpha_k)x^k\right]\right\| \\
&= \left\|PB_2\left[\chi, \frac{x^{k+1} - (1 - \alpha_k)x^k}{\|x^{k+1} - (1 - \alpha_k)x^k\|}\right]\right\| \|x^k\|\|x^{k+1} - (1 - \alpha_k)x^k\| \\
&= o\left(\max\left\{\|x^k\|^2\|\eta_2^k\|, \|x^k\|^3, \|x^k\|^2\|\lambda_1^k - \hat{\lambda}\|\right\}\right),
\end{aligned}
$$

which contradicts (45).

Summarizing, under the Assumptions (A1)–(A3), the multiplier $\bar{\lambda}$ may be expected to be noncritical only if (42) holds and the linear operator $PB_2[\chi](\hat{H}(\bar{\lambda}))^{-1}\Pi(B_2[\chi])^{\mathrm{T}}$ is singular on $\ker B_1^{\mathrm{T}}$, which is quite a special situation. In particular, by the definitions of Π, P and $\hat{H}(\bar{\lambda})$, for each $\eta \in \ker B_1^{\mathrm{T}}$ it holds that

$$\left\langle PB_2[\chi](\hat{H}(\bar{\lambda}))^{-1}\Pi(B_2[\chi])^{\mathrm{T}}\eta, \, \eta \right\rangle$$
$$= \left\langle PB_2[\chi]\Pi(\hat{H}(\bar{\lambda}))^{-1}\Pi(B_2[\chi])^{\mathrm{T}}P\eta, \, \eta \right\rangle$$
$$= \left\langle (\hat{H}(\bar{\lambda}))^{-1}\Pi(PB_2[\chi])^{\mathrm{T}}\eta, \, \Pi(PB_2[\chi])^{\mathrm{T}}\eta \right\rangle,$$

and the matrix in question cannot be singular if $\hat{H}(\bar{\lambda})$ is positive or negative definite and

$$\ker B_1^{\mathrm{T}} \cap \ker \Pi(PB_2[\chi])^{\mathrm{T}} = \{0\}. \tag{46}$$

Recall that SOSC (4) means that $\hat{H}(\bar{\lambda})$ is positive definite. Furthermore, (46) can be expressed in the form

$$\mathrm{im}\, PB_2[\chi]|_{\ker B_1} = \ker B_1^{\mathrm{T}},$$

which means that the mapping $\xi \to B_1\xi + \frac{1}{2}B_2[\xi, \xi] : \mathbf{R}^n \to \mathbf{R}^l$ is 2-regular at 0 in the direction χ ([19, Remark 1.3.2]), which is a generic property. Hence, if SOSC holds, then the linear operator $PB_2[\chi](\hat{H}(\bar{\lambda}))^{-1}\Pi(B_2[\chi])^{\mathrm{T}}$ is generically nonsingular on $\ker B_1^{\mathrm{T}}$. As a consequence, convergence to noncritical multipliers is an extremely unlikely event.

We finish with the following statement, which highlights one of the two possible reasons for lack of superlinear convergence of the Newton or SQP method in the degenerate case. The first reason is fairly obvious: the dual sequence may not converge at all. However, if the dual sequence converges, the reason becomes indeed convergence to critical multipliers. If the dual sequence were to converge to noncritical multipliers, the rate of primal convergence would have been superlinear. But, as demonstrated above, convergence to noncritical multipliers is extremely atypical.

Proposition 2 *Suppose that the iterative process given by (5)–(7), with $\alpha_k = 1$ for all k sufficiently large, correctly generates a sequence $\{(x^k, \lambda^k)\}$ converging to $(\bar{x}, \bar{\lambda})$. Suppose that $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical.*

Then $\{x^k\}$ converges to $\bar{x}$ superlinearly.

Proof Under the current assumptions, (A1) and (A2) hold trivially. Summing up (36) and (40), we obtain that

$$\xi^k = -x^k + \tilde{M}_k\xi_2^k - \hat{H}_k^{-1}\Pi(B_2[x^k])^{\mathrm{T}}\eta_2^k + O\left(\|x^k\|\|\lambda_1^k - \hat{\lambda}\|\right) + O\left(\|x^k\|^2\right).$$

Now taking into account that $\alpha_k = 1$, $\{\lambda_1^k\} \to \hat{\lambda}$, $\{\eta_2^k\} \to 0$, $\|\tilde{M}_k\| = O(\|x^k\|)$, $\xi_2^k = O(\|x^k\|)$ (for the latter, see (41)), and that the tail of $\{\hat{H}_k^{-1}\}$ is bounded, we obtain that

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

i.e., primal convergence is superlinear. □

4 An auxiliary procedure for estimating multipliers (dual stabilization)

The effect exposed above shows that even when λ^0 is close to some noncritical multiplier (or even when it *is* a noncritical multiplier but $x^0 \neq \bar{x}$), the sequence $\{\lambda^k\}$ tends to move almost parallel to the set $\Lambda(\bar{x})$ towards a multiplier which is critical. Typically, the set of critical multipliers is thin within the set of all multipliers (is of Lebesgue measure zero), while the set of noncritical ones is thick. We can therefore expect that the projection of some (say, random) λ^0 onto $\Lambda(\bar{x})$ should produce a multiplier which is not critical and, thus, has a much better chance of satisfying SOSC than the limit of $\{\lambda^k\}$. The purpose of dual stabilization suggested below is to compute an estimate of some multiplier which is not too far from the projection of λ^0 onto $\Lambda(\bar{x})$ and is expected to be noncritical. It is proposed having in mind mainly globalization issues (outer strategies) for special local methods designed for tackling degeneracy, e.g., [18,32]. This procedure generates an auxiliary dual sequence $\{\tilde{\lambda}^k\}$ but, by itself, does not interfere with the original algorithm generating the sequence $\{(x^k, \lambda^k)\}$. Therefore, the (outer) algorithm's primal behavior (convergence and rate of convergence to $\bar{x}$) is unchanged. The auxiliary sequence $\{\tilde{\lambda}^k\}$ is expected to approach $\Lambda(\bar{x})$ not too far from the projection of its starting point onto $\Lambda(\bar{x})$. We can therefore expect that applying a regularization/stabilization method for degenerate problems from $(x^k, \tilde{\lambda}^k)$, with k large enough, should be successful in forcing fast convergence of the local phase, e.g., [18,32].

Of course, the set $\Lambda(\bar{x})$ is not known and we cannot compute the projection of λ^0 onto it. However, the following observations appear useful.

As is easy to see, the squared residual $\Psi(\bar{x}, \lambda)$ (given by (8)) of the Lagrange system (2), for fixed $x = \bar{x}$, is a convex quadratic function with respect to λ, whose set of unconstrained minimizers is $\Lambda(\bar{x})$. The steepest descent method for a convex quadratic function generates iterates which move *orthogonally* to its set of minimizers. For completeness, we show this below.

Consider $\psi(\lambda) = \frac{1}{2}\langle Q\lambda, \lambda \rangle + \langle q, \lambda \rangle$, where Q is a symmetric positive semidefinite $l \times l$ matrix and $q \in \mathbf{R}^l$. Let $\bar{\lambda}$ belong to the (nonempty) set Λ of unconstrained minimizers of ψ. Then $Q\bar{\lambda} + q = 0$ and $\Lambda = \bar{\lambda} + \ker Q$. Let $\tilde{\lambda}^0$ be arbitrary and consider the steepest descent step $\tilde{\lambda}^1 = \tilde{\lambda}^0 - \gamma_0(Q\tilde{\lambda}^0 + q)$, $\gamma_0 > 0$. Evidently, $\tilde{\lambda}^1 - \tilde{\lambda}^0 = -\gamma_0(Q\tilde{\lambda}^0 + q) = -\gamma_0 Q(\tilde{\lambda}^0 - \bar{\lambda}) \in \text{im } Q = (\ker Q)^\perp$. Applying this reasoning inductively, we obtain that $\tilde{\lambda}^k \in \tilde{\lambda}^0 + (\ker Q)^\perp$ for all k. It easily follows that $\{\tilde{\lambda}^k\}$ converges (if the stepsizes $\{\gamma_k\}$ are chosen appropriately) to the orthogonal projection of $\tilde{\lambda}^0$ onto Λ.

Of course, we also cannot implement steepest descent steps for $\Psi(\bar{x}, \cdot)$, since $\bar{x}$ is unknown. The idea is that when x^k becomes close to $\bar{x}$, the steepest descent step for $\Psi(x^k, \cdot)$ is a small perturbation of the steepest descent step for $\Psi(\bar{x}, \cdot)$. Therefore, if we use $\Psi(x^k, \cdot)$ instead, we should be moving almost orthogonally to $\Lambda(\bar{x})$, the set of minimizers of $\Psi(\bar{x}, \cdot)$. This gives rise to the following strategy.

As before, suppose that the damped Newton scheme considered above generates a sequence $\{(x^k, \lambda^k)\}$ such that $\{x^k\} \to \bar{x}$. The idea is, starting from some iteration k, to produce (in a cheap way) an auxiliary sequence $\{\tilde{\lambda}^k\}$, which approaches $\Lambda(\bar{x})$ and is expected to move almost orthogonally to this set (instead of along this set, as is the case for the Newton iterates).

Let k be an arbitrary iteration index (perhaps, large enough). Set $\tilde{\lambda}^k = \lambda^k$. On subsequent iterations, proceed as follows. Having computed (x^{k+1}, λ^{k+1}), compute some $\tilde{\lambda}^{k+1}$ satisfying

$$\tilde{\lambda}^{k+1} = \tilde{\lambda}^k - \gamma_k \frac{\partial \Psi}{\partial \lambda}\left(x^{k+1}, \tilde{\lambda}^k\right), \tag{47}$$

$$0 < c_1 \leq \gamma_k < \min\left\{c_2 \,;\, \|(F'(x^k))^{\mathrm{T}}\|^{-2}\right\}, \quad c_2 > 0. \tag{48}$$

The above conditions include various possibilities for computing the stepsize. For example, we can perform the standard steepest descent step with linesearch for $\Psi(x^{k+1}, \cdot)$ from the point $\tilde{\lambda}^k$. Another possibility is the primal-dual steepest descent step with linesearch for $\Psi(\cdot, \cdot)$ from the point $(x^{k+1}, \tilde{\lambda}^k)$. In this second scheme, an additional auxiliary sequence $\{\tilde{x}^k\}$ would be generated, which can be discarded for our purposes (we are only interested in the dual sequence here).

In practical implementations, in addition to (47)–(48), one should also ensure descent for $\Psi(x^{k+1}, \cdot)$, of course. We do not impose this condition explicitly, because doing so does not seem to strengthen what can be formally claimed concerning the convergence of $\{\tilde{\lambda}^k\}$. We show below that this sequence has an accumulation point which belongs to $\Lambda(\bar{x})$. Moreover, if the primal sequence converges at a linear rate (which is a frequent situation for Newton and SQP methods in the degenerate case), then all accumulation points of $\{\tilde{\lambda}^k\}$ belong to $\Lambda(\bar{x})$. Also, it is absolutely clear that $\{\tilde{\lambda}^k\}$ does not have any attraction preference for critical multipliers. All this is confirmed by our numerical experiments, see Sect. 5. In particular, when there is convergence of the primal sequence, the auxiliary dual sequence generated by the outlined procedure with linesearch for $\Psi(x^{k+1}, \cdot)$ never failed to converge to a valid multiplier. Moreover, in the vast majority of cases this limit is a noncritical multiplier, as expected.

We start with the following characterization of the multiplier set $\Lambda(\bar{x})$.

Lemma 1 *If $\Lambda(\bar{x}) \neq \emptyset$ then for each $\lambda \in \mathbf{R}^l$ satisfying*

$$F'(\bar{x})\frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0, \tag{49}$$

it holds that $\lambda \in \Lambda(\bar{x})$.

Proof Since $\Lambda(\bar{x}) \neq \emptyset$, we have that

$$f'(\bar{x}) \in \text{im}\left(F'(\bar{x})\right)^{\mathrm{T}}.$$

Consequently, for any $\lambda \in \mathbf{R}^l$, we obtain that

$$\frac{\partial L}{\partial x}(\bar{x}, \lambda) = f'(\bar{x}) + \left(F'(\bar{x})\right)^{\mathrm{T}} \lambda \in \text{im}\left(F'(\bar{x})\right)^{\mathrm{T}} = \left(\ker F'(\bar{x})\right)^{\perp}.$$

Combining the latter relation with (49), we conclude that $\frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0$, i.e., that $\lambda \in \Lambda(\bar{x})$. $\qquad\square$

Observe that $\frac{\partial \Psi}{\partial \lambda}(\bar{x}, \lambda) = F'(\bar{x})\frac{\partial L}{\partial x}(\bar{x}, \lambda)$, so that we can write

$$\Lambda(\bar{x}) = \left\{\lambda \in \mathbf{R}^l \;\middle|\; \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \lambda) = 0\right\}. \tag{50}$$

We proceed to analyze the sequence $\{\tilde{\lambda}^k\}$.

Proposition 3 *Suppose* $\{x^k\} \to \bar{x}$ *and* $\{\tilde{\lambda}^k\}$ *satisfying conditions* (47)–(48) *is bounded.*

Then the sequence $\{\tilde{\lambda}^k\}$ *has an accumulation point which belongs to* $\Lambda(\bar{x})$.

If it holds that $\sum_{k=0}^{\infty} \|x^{k+1} - \bar{x}\| < +\infty$ *(for example, if* $\{x^k\} \to \bar{x}$ *at a linear rate), then all accumulations points of* $\{\tilde{\lambda}^k\}$ *belong to* $\Lambda(\bar{x})$.

Proof Observe that

$$\tilde{\lambda}^{k+1} = \tilde{\lambda}^k - \gamma_k \frac{\partial \Psi}{\partial \lambda}\left(\bar{x}, \tilde{\lambda}^k\right) + \delta^k,$$

where

$$\delta^k = \gamma_k \left(\frac{\partial \Psi}{\partial \lambda}\left(\bar{x}, \tilde{\lambda}^k\right) - \frac{\partial \Psi}{\partial \lambda}\left(x^{k+1}, \tilde{\lambda}^k\right)\right) \to 0 \text{ as } k \to \infty, \tag{51}$$

by $\{x^k\} \to \bar{x}$, the boundedness of $\{\tilde{\lambda}^k\}$ and (48). We obtain that

$$
\begin{aligned}
&\Psi(\bar{x}, \tilde{\lambda}^k) - \Psi(\bar{x}, \tilde{\lambda}^{k+1}) \\
&= \frac{1}{2}\left\|\frac{\partial L}{\partial x}(\bar{x}, \tilde{\lambda}^k)\right\|^2 - \frac{1}{2}\left\|\frac{\partial L}{\partial x}(\bar{x}, \tilde{\lambda}^{k+1})\right\|^2 \\
&= -\left\langle\frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k), \tilde{\lambda}^{k+1} - \tilde{\lambda}^k\right\rangle - \frac{1}{2}\|(F'(\bar{x}))^{\mathrm{T}}(\tilde{\lambda}^{k+1} - \tilde{\lambda}^k)\|^2 \\
&\geq \gamma_k\left\|\frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k)\right\|^2 - \|\delta^k\|\left\|\frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k)\right\|
\end{aligned}
$$

$$-\frac{1}{2}\|(F'(\bar{x}))^{\mathrm{T}}\|^2 \left(\gamma_k^2 \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\|^2 + \|\delta^k\|^2 + 2\gamma_k \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\| \|\delta^k\| \right)$$

$$= \gamma_k \left(1 - \frac{\gamma_k}{2}\|(F'(\bar{x}))^{\mathrm{T}}\|^2 \right) \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\|^2 - \varepsilon_k$$

$$\geq c_1 \left(1 - \frac{\|(F'(\bar{x}))^{\mathrm{T}}\|^2}{2\|(F'(x^k))^{\mathrm{T}}\|^2} \right) \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\|^2 - \varepsilon_k$$

$$\geq \frac{c_1}{4} \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\|^2 - \varepsilon_k, \tag{52}$$

where

$$\varepsilon_k = \|\delta^k\| \left(\left(1 + \gamma_k\|(F'(\bar{x}))^{\mathrm{T}}\|^2\right) \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\| + \frac{1}{2}\|\delta_k\|\|(F'(\bar{x}))^{\mathrm{T}}\|^2 \right) \to 0,$$

as $k \to \infty$, by (51). For any $j > k$, we then obtain that

$$\Psi(\bar{x}, \tilde{\lambda}^k) \geq \Psi(\bar{x}, \tilde{\lambda}^k) - \Psi(\bar{x}, \tilde{\lambda}^j) = \sum_{i=k}^{j-1} \left(\Psi(\bar{x}, \tilde{\lambda}^i) - \Psi(\bar{x}, \tilde{\lambda}^{i+1}) \right)$$

$$\geq \sum_{i=k}^{j-1} \left(\frac{c_1}{4} \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^i) \right\|^2 - \varepsilon_i \right). \tag{53}$$

Suppose that there exists $\nu > 0$ such that $\left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\| \geq \nu \ \forall k$. Fix any iteration index k such that $\varepsilon_i \leq c_1 \nu^2/8$ for all $i > k$. By (53), for $j > k$ we have that

$$\Psi(\bar{x}, \tilde{\lambda}^k) \geq \sum_{i=k}^{j-1} \left(c_1 \nu^2/4 - c_1 \nu^2/8 \right) = (j-k)c_1\nu^2/8,$$

which results in a contradiction when $j \to \infty$. Hence,

$$\liminf_{k \to \infty} \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\| = 0.$$

Recalling (50), we conclude that $\{\tilde{\lambda}^k\}$ has an accumulation point in $\Lambda(\bar{x})$.

Suppose finally that $\sum_{k=0}^{\infty} \|x^{k+1} - \bar{x}\| < +\infty$. Taking into account (48), the boundedness of $\{\tilde{\lambda}^k\}$, and the Lipschitz-continuity of $\frac{\partial \Psi}{\partial \lambda}(\cdot, \lambda)$, it follows that $\|\delta^k\| \leq M\|x^{k+1} - \bar{x}\|$ for some $M > 0$. Consequently, $\sum_{k=0}^{\infty} \varepsilon_k < +\infty$. By (53), we then obtain that for any $j = 0, 1, \ldots,$ it holds that

$$\Psi(\bar{x}, \tilde{\lambda}^0) + \sum_{k=0}^{j} \varepsilon_k \geq \frac{c_1}{4} \sum_{k=0}^{j} \left\| \frac{\partial \Psi}{\partial \lambda}(\bar{x}, \tilde{\lambda}^k) \right\|^2.$$

Letting $j \to \infty$ gives

$$+\infty > \sum_{k=0}^{\infty} \left\| \frac{\partial \Psi}{\partial \lambda} (\bar{x}, \tilde{\lambda}^k) \right\|^2,$$

implying that

$$\frac{\partial \Psi}{\partial \lambda} \left(\bar{x}, \tilde{\lambda}^k \right) \to 0 \text{ as } k \to \infty.$$

Hence, all accumulation points of $\{\tilde{\lambda}^k\}$ are in $\Lambda(\bar{x})$. $\qquad \square$

As already mentioned, dual stabilization suggested above is proposed having in mind mainly globalization issues (outer strategies) for special local methods designed for tackling degeneracy, e.g., [18,32]. In particular, by itself, this procedure does not interfere with the sequence $\{x^k\}$. One can also think of different (local) dual stabilization procedures, which *do interfere* with local primal behavior in order to improve it. Actually, stabilized SQP method which was suggested in [29] and further investigated in [8,11,21,30,31], as well as the approaches developed in [7,18,32], can be regarded as this second kind of dual stabilization.

5 Numerical experiments

In this section, we report on some numerical experiments, which complement and confirm our theoretical analysis. Our experiments deal with randomly generated quadratically constrained quadratic problems. The stationary point of interest is always $\bar{x} = 0$. Note that for randomly generated problems of the type we consider here, this stationary point is not always a solution (even local), and moreover, is not necessarily the unique stationary point. Thus, naturally, convergence to other points happens sometimes, as well as failure of convergence. For analysis, we select the runs for which convergence to 0 occurs and base our statistics regarding critical multipliers on these runs only. As convergence to 0 happens often enough, the analysis of those runs is sufficient to draw conclusions.

Our experiments are performed in Matlab environment. For each given triple (n, l, r), where $r < l$ stands for rank of $F'(\bar{x})$, we generate 100 problems of the form

$$\text{minimize } \langle a, x \rangle + \frac{1}{2} \langle Ax, x \rangle \text{ subject to } B_1 x + \frac{1}{2} B_2[x, x] = 0.$$

Each problem is constructed as follows. First, we generate an $l \times n$-matrix B_1 such that rank $B_1 = r$. The first r rows of this matrix are generated as random vectors in $\mathbf{R}^n$ with components in $[-10, 10]$. Each of the rest $l - r$ rows is defined as a linear combination of the first r rows with random coefficients taken from $[-1, 1]$. We further define $a \in \mathbf{R}^n$ as the linear combination of the rows of B_1 with random coefficients in $[-1, 1]$. Finally, we generate random symmetric $n \times n$-matrices A and

$B_2^i, i = 1, \ldots, l$, with entries taken from $[-10, 10]$. Matrices $B_2^i, i = 1, \ldots, l$, define the symmetric bilinear mapping B_2 by $B_2[x^1, x^2] = (\langle B_2^1 x^1, x^2 \rangle, \ldots, \langle B_2^l x^1, x^2 \rangle)$, $x^1, x^2 \in \mathbf{R}^n$.

For each of the problems, we run the damped Newton method starting from 10 random points (x^0, λ^0) whose components are in $[-100, 100]$ (thus, we perform 1,000 runs for each triple (n, l, r)). In the process, we also generate an auxiliary dual sequence by the stabilization procedure suggested in Sect. 4.

The damped Newton method is defined by (5), (6) with $\beta_k = 1 \; \forall \, k$ and α_k computed by the standard SQP linesearch procedure. Specifically, at each iteration k the initial trial value $\alpha_k = 1$ is halved until

$$\psi_k \left(x^k + \alpha_k \xi^k \right) \leq \psi_k(x^k) + \varepsilon \alpha_k \left(\langle f'(x^k), \xi^k \rangle - c_k \| F(x^k) \|_1 \right),$$

where $\varepsilon = 0.1$ and

$$\psi_k : \mathbf{R}^n \to \mathbf{R}, \quad \psi_k(x) = f(x) + c_k \| F(x) \|_1,$$

is the l_1-penalty function with the penalty parameter $c_k > 0$. We use the following simple update rule for penalty parameters: $c_0 = \| \lambda^1 \|_\infty + 1$, and then for each $k = 1, 2, \ldots$, we set $c_k = c_{k-1}$ if $c_{k-1} \geq \| \lambda^{k+1} \|_\infty$, and $c_k = \| \lambda^{k+1} \|_\infty + 1$ otherwise. We note that even though the multiplier set is unbounded, we did not have any trouble with bounding the penalty parameter c_k (i.e., the generated dual sequence stays bounded).

The additional dual sequence $\{ \tilde{\lambda}^k \}$ is generated starting from $\tilde{\lambda}^0 = \lambda^0$ according to (47), with γ_k computed as follows. At each iteration k the initial trial value $\gamma_k = 1$ is halved until the Armijo inequality

$$\Psi \left(x^{k+1}, \tilde{\lambda}^k - \gamma_k \frac{\partial \Psi}{\partial \lambda} \left(x^{k+1}, \tilde{\lambda}^k \right) \right) \leq \Psi \left(x^{k+1}, \tilde{\lambda}^k \right) - \varepsilon \gamma_k \left\| \frac{\partial \Psi}{\partial \lambda} \left(x^{k+1}, \tilde{\lambda}^k \right) \right\|^2$$

is satisfied (with the same $\varepsilon = 0.1$).

When $\{ (x^k, \lambda^k) \}$ achieves the stopping condition $\| \Phi(x^k, \lambda^k) \| < 10^{-10}$, we compute some information about x^k and λ^k described below. In the table of numerical results, this information is reported in the columns labeled "Without DS" ("DS" stands for "dual stabilization"). Similarly, when $\{ (x^k, \tilde{\lambda}^k) \}$ achieves the stopping condition $\| \Phi(x^k, \tilde{\lambda}^k) \| < 10^{-10}$, we report about this point in the group of columns labeled "With DS".

For each sequence, failures have been declared when the corresponding criterion was not satisfied after 1,000 iterations (we put a high limit on the number of iterations to maximize the number of cases for analysis). We also declare failure (for a sequence which did not yet satisfy the corresponding condition for collecting information) when at some iteration the Matlab linear solver did not manage to compute the Newton step (because the matrix of the system was close to degenerate), or when α_k became less than 10^{-15} in the process of backtracking.

Note that for our randomly generated problems, by no means H_k is positive definite for all (or even some) steps k. Hence, global convergence of the damped Newton

Table Damped Newton method

n	l	r	Without DS			With DS		
			CT0	C	CR	CT0	C	CR
5	2	0	474	300	469	474	0	7
5	2	1	146	56	139	110	0	0
5	3	0	698	415	694	719	0	6
5	3	1	279	109	277	294	0	0
5	3	2	106	25	99	101	0	0
10	5	0	738	430	710	744	0	2
10	5	1	369	181	356	366	0	0
10	5	2	191	72	176	192	0	0
10	5	3	72	20	70	66	0	0
10	5	4	29	5	24	13	0	0
10	6	0	843	484	813	826	0	0
10	6	2	250	98	239	225	0	0
10	6	3	131	43	124	114	0	0
10	6	4	63	13	58	58	0	0
10	8	0	943	587	929	942	0	0
10	8	3	272	58	241	253	0	0
10	8	4	139	24	119	156	0	0
10	8	6	48	8	42	48	0	0
10	8	7	16	1	14	14	0	0
25	12	0	904	259	559	909	0	0
25	13	0	926	206	525	906	0	2
25	13	6	88	20	70	72	0	0
25	13	7	32	4	27	29	0	0
25	14	0	933	231	536	930	0	0
25	17	0	964	245	605	962	0	0
25	20	0	982	337	783	984	0	0
25	20	10	58	6	40	54	0	0
25	20	11	40	3	17	32	0	0
25	23	0	999	563	960	999	0	0
25	23	12	53	0	12	64	0	0
25	23	13	42	0	14	33	0	0

CT0 number of problems selected for analysis (convergence to $\bar{x} = 0$ was obtained), *C and CR* number of times (out of CT0) convergence to a critical multiplier was detected by absolute and relative criteria, respectively

method is not a given: failures may occur, and they do occur indeed. Robustness, however, is not a concern in our experiments at all. We are merely interested in obtaining convergence to 0 in a statistically meaningful number of cases, so that conclusions can be drawn about convergence to critical multipliers.

If failure did not occur, we check whether $\|x^k\| < 10^{-5}$. If this is the case, we say that primal convergence to 0 is detected (reported as "CT0").

In the case of the sequence without stabilization, we then compute the smallest absolute value σ_k of the eigenvalues of the reduced Hessian $\hat{H}(\lambda^k)$. For the sequence with dual stabilization, σ_k is the smallest absolute value of the eigenvalues of the reduced Hessian $\hat{H}(\tilde{\lambda}^k)$. If $\sigma_k < 10^{-5}$, then λ^k (respectively, $\tilde{\lambda}^k$) is declared to be close to a critical multiplier (reported as "C"). If $\sigma_k/\sigma_0 < 10^{-5}$, then λ^k (respectively, $\tilde{\lambda}^k$) is also declared close to a critical multiplier but by the weaker relative ratio criterion (reported as "CR"). Each of the columns "CT0", "C" and "CR" reports the number of runs (out of total 1,000) which fall into each of these categories, as defined above.

We believe that the overall picture is very conclusive. Method without dual stabilization evidently has preference for critical multipliers. Results do not look as impressive for some runs with $n = 25$, but as far as we can tell, this is explained by instability of the process in our simple implementation, especially with respect to the dual trajectory (of course, instability increases as the dimensions get higher).

By contrast, and as expected, the stabilized scheme has no preference for critical multipliers whatsoever, as convergence to the latter is observed in a statistically negligible number of cases.

6 The general case of mixed equality and inequality constraints

We now turn our attention to the problem with mixed constraints

$$\text{minimize } f(x) \text{ subject to } F(x) = 0, \ G(x) \le 0, \tag{54}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $F : \mathbf{R}^n \to \mathbf{R}^l$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ are sufficiently smooth. We can report conclusions similar to the above, if the algorithm in question asymptotically behaves as a Newton-type method applied to some related equality-constrained problem. This may happen for different reasons, as discussed next.

Let the primal–dual trajectory $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be generated by the standard linesearch SQP algorithm, i.e., $x^{k+1} = x^k + \alpha_k \xi^k$, where $\alpha_k \ge 0$ is the stepsize parameter, ξ^k is a stationary point of the SQP subproblem

$$\begin{array}{ll} \text{minimize} & \langle f'(x^k), \xi \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)\xi, \ \xi \right\rangle \\ \text{subject to} & F(x^k) + F'(x^k)\xi = 0, \ G(x^k) + G'(x^k)\xi \le 0, \end{array} \tag{55}$$

$(\lambda^{k+1}, \mu^{k+1})$ is a Lagrange multiplier associated to ξ^k, and we denote by $L(x, \lambda, \mu) = f(x) + \langle \lambda, F(x) \rangle + \langle \mu, G(x) \rangle$, $x \in \mathbf{R}^n, \lambda \in \mathbf{R}^l, \mu \in \mathbf{R}^m$, the Lagrangian of problem (54).

Suppose that the primal trajectory $\{x^k\}$ converges to a solution $\bar{x}$ of (54). In the case of problem (54), LICQ has the form

$$\text{rank} \begin{pmatrix} F'(\bar{x}) \\ G'_{I(\bar{x})}(\bar{x}) \end{pmatrix} = l + |I(\bar{x})|,$$

where y_I stands for the subvector of vector y with components y_i, $i \in I$, and

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

is the set of indices of inequality constraints active at $\bar{x}$. The situation of interest is when LICQ is violated at $\bar{x}$, but the set $\mathcal{M}(\bar{x})$ of Lagrange multipliers associated with $\bar{x}$ is, nevertheless, nonempty. Thus, the Lagrange multipliers exist but are not unique.

As before, we are interested in possible scenarios for the dual trajectory $\{(\lambda^k, \mu^k)\}$. The first question is: What is a critical multiplier in this setting? It seems quite natural to assume that, for k large enough, the set

$$I_k = \left\{i = 1, \ldots, m \mid G_i(x^k) + \langle G_i'(x^k), \xi^k \rangle = 0\right\}$$

of indices of inequality constraints active at the computed stationary points ξ^k of SQP subproblems remains unchanged. According to the discussion below, this is automatic when $\{(\lambda^k, \mu^k)\}$ tends to a multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfying strict complementarity, that is, such that $\bar{\mu}_{I(\bar{x})} > 0$. In other cases, the assumption that the set I_k is asymptotically unchanged may not hold, of course. Nevertheless, this seems to be reasonable numerical behavior, which should not be unusual.

Assuming that $I_k = I$ for all k large enough, we have that $\mu_i^k = 0 \, \forall \, i \in \{1, \ldots, m\} \setminus I$, for each such k. Then, as is readily seen from (55), $(\xi^k, \lambda^{k+1}, \mu^{k+1})$ satisfies the equations

$$f'(x^k) + \frac{\partial^2 L_I}{\partial x^2}\left(x^k, \lambda^k, \mu_I^k\right)\xi^k + \left(F'(x^k)\right)^{\mathrm{T}}\lambda^{k+1} + \left(G_I'(x^k)\right)^{\mathrm{T}}\mu_I^{k+1} = 0,$$
$$F(x^k) + F'(x^k)\xi^k = 0, \quad G_I(x^k) + G_I'(x^k)\xi^k = 0, \tag{56}$$

where we defined $L_I(x, \lambda, \mu_I) = f(x) + \langle \lambda, F(x) \rangle + \langle \mu_I, G_I(x) \rangle$, $x \in \mathbf{R}^n$, $\lambda \in \mathbf{R}^l$ and $\mu_I \in \mathbf{R}^{|I|}$. Note that the latter is the Lagrangian of the following equality-constrained optimization problem:

$$\text{minimize } f(x) \text{ subject to } F(x) = 0, \ G_I(x) = 0. \tag{57}$$

Assuming that $\{\xi^k\}$ converges to 0 (which is an immediate consequence of the convergence of $\{x^k\}$ if the stepsize α_k stays bounded away from zero), and assuming that (λ^k, μ^k) is bounded, by passing onto the limit in (56) (along an appropriate subsequence) we conclude that $\bar{x}$ is a stationary point of (57). And, in particular, $I \subset I(\bar{x})$. Moreover, any $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfying $\bar{\mu}_i = 0 \, \forall \, i \in \{1, \ldots, m\} \setminus I$ corresponds to the multiplier $(\bar{\lambda}, \bar{\mu}_I)$ of (57) associated with $\bar{x}$, and we call such multiplier $(\bar{\lambda}, \bar{\mu})$ critical if $(\bar{\lambda}, \bar{\mu}_I)$ is critical for (57) according to the Definition 1 for equality-constrained problems. According to (56), the primal-dual trajectory $\{(x^k, \lambda^k, \mu_I^k)\}$ can be thought of as generated by the Newton–Lagrange method for (57). In fact, *this* is the motivation for the notion of critical multipliers that we currently suggest for the case of mixed constraints.

According to our theoretical and numerical results for equality-constrained problems, we conclude the following. If $\{(\lambda^k, \mu^k)\}$ converges to a noncritical multiplier

then the primal convergence rate is superlinear. This situation is quite possible when the constraints of (57) are regular at $\bar{x}$ (which of course may happen when I is strictly smaller than $I(\bar{x})$). In this case, the multiplier $(\bar{\lambda}, \bar{\mu}_I)$ of (57) associated with $\bar{x}$ is unique. Note that in this case, dual convergence is also superlinear. However, if the constraints of (57) are degenerate at $\bar{x}$ (for example, when $I = I(\bar{x})$), this situation is highly unlikely to occur. The expected behavior in this case is the following: either $\{(\lambda^k, \mu^k)\}$ does not converge or it converges to a critical multiplier, and primal convergence is slow. Example 6 below shows convergence to a critical multiplier, and slow rate of convergence, even when I is strictly smaller than $I(\bar{x})$.

We next give some further details and examples. Let $\{(\lambda^k, \mu^k)\}$ converge to some $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$. Define

$$I_+(\bar{x}, \bar{\mu}) = \{i \in I(\bar{x}) \mid \bar{\mu}_i > 0\},$$

which is the set of strongly active constraints. As is easy to see,

$$I_+(\bar{x}, \bar{\mu}) \subset I \subset I(\bar{x}). \tag{58}$$

Note that if strong second-order sufficient condition (SSOSC)

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \left(\ker F'(\bar{x}) \cap \ker G'_{I_+(\bar{x}, \bar{\mu})}(\bar{x})\right) \backslash \{0\} \tag{59}$$

for problem (54) holds, then SOSC holds for problem (57) at $\bar{x}$ with multiplier $(\bar{\lambda}, \bar{\mu}_I)$. Therefore, in this case $(\bar{\lambda}, \bar{\mu})$ cannot be a critical multiplier.

Consider first the case when dual trajectory converges to a strictly complementary multiplier. The specificity of this case is that $I = I(\bar{x})$, by (58). Hence, the constraints of (57) are degenerate at $\bar{x}$, and we expect dual convergence to a critical multiplier and low rate of convergence. Moreover, $\bar{\mu}_i = 0 \; \forall i \in \{1, \dots, m\} \backslash I$ and $\forall (\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, which means that any multiplier in $\mathcal{M}(\bar{x})$ can be classified as being critical or noncritical according to the notion introduced above. Furthermore, in this case, if SOSC for problem (54) holds, i.e., if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \backslash \{0\}, \tag{60}$$

where $C(\bar{x}) = \left\{\xi \in \ker F'(\bar{x}) \mid G'_{I(\bar{x})}(\bar{x})\xi \leq 0, \; \langle f'(\bar{x}), \xi \rangle \leq 0\right\}$ is the critical cone of (54) at $\bar{x}$, then SOSC holds for problem (57) at $\bar{x}$ with the multiplier $(\bar{\lambda}, \bar{\mu}_I)$. In particular, such $(\bar{\lambda}, \bar{\mu})$ cannot be critical.

The following example, coming from [6, Example 4.23], demonstrates that convergence to a strictly complementary critical multiplier is possible even if the Mangasarian–Fromovitz constraint qualification (MFCQ) holds at $\bar{x}$. Recall that MFCQ means that

$$\text{rank } F'(\bar{x}) = l \text{ and } \exists \bar{\bar{\xi}} \in \ker F'(\bar{x}) \text{ such that } G'_{I(\bar{x})}(\bar{x})\bar{\bar{\xi}} < 0,$$

and it is equivalent to the multiplier set $\mathcal{M}(\bar{x})$ being nonempty and bounded.

Example 4 Consider the problem

$$\text{minimize} - x_1 \text{ subject to } x_1 - x_2^2 \le 0, \ x_1 + x_2^2 \le 0.$$

Its solution is $\bar{x} = 0$. It satisfies MFCQ and $\mathcal{M}(\bar{x}) = \{\bar{\mu} \in \mathbf{R}_+^2 \mid \bar{\mu}_1 + \bar{\mu}_2 = 1\}$. Furthermore, SOSC (60) holds for any $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $\bar{\mu}_1 < \bar{\mu}_2$, and does not hold for other $\bar{\mu} \in \mathcal{M}(\bar{x})$. SQP subproblem (55) takes the form

$$\begin{aligned}
&\text{minimize} && -\xi_1 - (\mu_1^k - \mu_2^k)\xi_2^2 \\
&\text{subject to} && x_1^k - (x_2^k)^2 + \xi_1 - 2x_2^k\xi_2 \le 0, \ x_1^k + (x_2^k)^2 + \xi_1 + 2x_2^k\xi_2 \le 0.
\end{aligned}$$

For simplicity, let $x_1^k = 0$, $x_2^k \neq 0$. Suppose that μ^k is close enough to $\mathcal{M}(\bar{x})$. Then the point $\xi^{k+1} = (0, -x_2^k/2)$ is stationary in this subproblem, with both constraints being active. Hence, the primal SQP step (with $\alpha_k = 1$) is given by $x^{k+1} = x^k/2 = (0, x_2^k/2)$. In particular, both QP constraints remain active along the primal trajectory, and hence, $I = I(\bar{x}) = \{1, 2\}$.

The multiplier of SQP subproblem is given by $\mu^{k+1} = (1/2 + (\mu_1^k - \mu_2^k)/4, \ 1/2 - (\mu_1^k - \mu_2^k)/4)$. It follows that $\mu_1^{k+1} - \mu_2^{k+1} = \frac{1}{2}(\mu_1^k - \mu_2^k)$, and hence, $(\mu_1^k - \mu_2^k) \to 0$. The latter implies that $\{\mu^k\} \to (1/2, 1/2)$, which is a strictly complementary multiplier, and the unique critical multiplier.

As expected, the rate of convergence is only linear.

Consider now the case when dual trajectory converges to a multiplier violating strict complementarity. Note that in this case, multiplier satisfying SOSC (60) but not SSOSC (59) can be critical (see Example 6 further below).

There exist examples of this scenario both with convergence to critical multipliers and to noncritical ones. Thus, by itself, this scenario is not a reason for slow convergence. However, when convergence is slow, attraction to critical multipliers can be observed yet again. The following example is taken from [31, Sect. 6].

Example 5 Consider the problem

$$\text{minimize } x_1 \text{ subject to } - x_1 \le 0, \ (x_1 - 2)^2 + x_2^2 \le 4.$$

The solution is $\bar{x} = 0$, and it satisfies MFCQ. We have that $\mathcal{M}(\bar{x}) = \{\bar{\mu} \in \mathbf{R}^2 \mid \bar{\mu}_1 = 1 - 4\bar{\mu}_2, \ 0 \le \bar{\mu}_2 \le 1/4\}$, and SOSC (60) (and even SSOSC (59)) holds with all $\bar{\mu} \in \mathcal{M}(\bar{x})$, except for $\bar{\mu} = (1, 0)$.

According to numerical experience in [31], in this example SQP (with $\alpha_k = 1$) converges linearly, and the dual trajectory is attracted by the "troublesome" multiplier $\bar{\mu} = (1, 0)$. Let us examine this convergence in more detail.

SQP subproblem (55) takes the form

$$\begin{aligned}
&\text{minimize} && \xi_1 + \mu_2^k(\xi_1^2 + \xi_2^2) \\
&\text{subject to} && -x_1^k - \xi_1 \le 0, \ (x_1^k - 2)^2 + (x_2^k)^2 - 4 + 2(x_1^k - 2)\xi_1 + 2x_2^k\xi_2 \le 0.
\end{aligned}$$

For simplicity, let $x_1^k = 0$, $x_2^k \neq 0$. Suppose that μ^k is close enough to $\mathcal{M}(\bar{x})$. Then the point $\xi^{k+1} = (0, -x_2^k/2)$ is stationary in this subproblem, with both constraints being active. Hence, the primal SQP step (with $\alpha_k = 1$) is given by $x^{k+1} = x^k/2 = (0, x_2^k/2)$. In particular, both QP constraints remain active along the primal trajectory, and hence, $I = I(\bar{x}) = \{1, 2\}$.

Since we expect the dual convergence to $\bar{\mu} = (1, 0)$, we should consider the case when the first constraint of SQP subproblem is active: $\xi_1^k = -x_1^k$, and hence, $x_1^{k+1} = 0$.

Now we can deal solely with the case when $x_1^k = 0$. In this case, the point $\xi^k = (0, -x_2^k/2)$ is stationary in SQP subproblem, with both constraints being active. Hence, the primal SQP step is given by $x^{k+1} = x^k/2 = (0, x_2^k/2)$. In particular, both QP constraints remain active, and hence, $I = I(\bar{x}) = \{1, 2\}$.

The multiplier of SQP subproblem is given by $\mu^{k+1} = (1 + 2\mu_2^k, \mu_2^k/2)$. It follows that $\{\mu^k\} \to \bar{\mu} = (1, 0)$. The constraints of (57) are degenerate at $\bar{x}$, and $\bar{\mu} = (1, 0)$ is the unique associated critical multiplier. Hence, this $\bar{\mu}$ is a non-strictly complementary critical (with respect to the chosen set I) multiplier of (54).

In Examples 4 and 5, we had $I = I(\bar{x})$. We finish this discussion with the case when I is strictly smaller than $I(\bar{x})$. The following example is `ralph2` in MacMPEC [20].

Example 6 Consider the problem with complementarity constraints (MPCC)

$$\text{minimize } x_1^2 + x_2^2 - 4x_1x_2 \text{ subject to } x_1 \geq 0, \ x_2 \geq 0, \ x_1x_2 \leq 0.$$

Its solution is $\bar{x} = 0$. As any MPCC feasible point, it violates MFCQ. However, the multiplier set is nonempty (hence, unbounded), and it is given by $\mathcal{M}(\bar{x}) = \{\bar{\mu} \in \mathbf{R}^3 \mid \bar{\mu}_1 = \bar{\mu}_2 = 0, \ \bar{\mu}_3 \geq 0\}$. Moreover, SOSC (60) (but not SSOSC (59)) holds with all $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $\bar{\mu}_3 > 2$.

SQP subproblem (55) takes the form

$$\begin{aligned}
\text{minimize} \quad & 2x_1^k\xi_1 + 2x_2^k\xi_2 - 4x_2^k\xi_1 - 4x_1^k\xi_2 + \xi_1^2 + \xi_2^2 - (4 - \mu_3^k)\xi_1\xi_2 \\
\text{subject to} \quad & x_1^k + \xi_1 \geq 0, \ x_2^k + \xi_2 \geq 0, \ x_1^kx_2^k + x_2^k\xi_1 + x_1^k\xi_2 \leq 0.
\end{aligned}$$

Suppose that $0 \leq \mu_3^k < 6$, and let $x_1^k = x_2^k \neq 0$. It can be easily seen that the first two constraints cannot be active in this case, and that $\xi^k = -x^k/2$ is the unique stationary point of SQP subproblem, with the last constraint being active. Hence, the primal SQP step (with $\alpha_k = 1$) is given by $x^{k+1} = x^k/2$. In particular, only the last QP constraint remains active along the primal trajectory. Hence, $I = \{3\}$, while $I(\bar{x}) = \{1, 2, 3\}$. The multiplier of SQP subproblem is given by $\mu^{k+1} = (0, 0, 1 + \mu_3^k/2)$. It follows that $\{\mu^k\} \to \bar{\mu} = (0, 0, 2)$. The unique constraint of (57) is degenerate at $\bar{x}$, and $\bar{\mu}_3 = 2$ is an associated critical multiplier. Hence, this $\bar{\mu}$ is a non-strictly complementary critical (with respect to the chosen set I) multiplier of (54), and $I \neq I(\bar{x})$.

Note that in this example, (57) has another critical multiplier $\bar{\mu}_3 = 6$, despite of SOSC (60) (but not SSOSC (59)!) for (54) being valid with $\bar{\mu} = (0, 0, 6)$.

7 Concluding remarks

The purpose of this work was to raise the question of convergence properties of the dual part of the sequence in primal–dual optimization methods in the case where some multipliers associated to the solution satisfy the second-order sufficient condition for optimality (SOSC), while others do not. Whether one may expect (or not) the dual part of the sequence to approach those multipliers which satisfy SOSC is interesting theoretically, but also relevant numerically (for example, for applicability of methods which have been proposed to handle constraints degeneracy, as well as for speed of convergence of standard algorithms).

The overall picture (different methods, different classes of problems) requires further investigation. We have presented some evidence (both theoretical and numerical) for the damped Newton scheme, including SQP, applied to the problem with equality constraints. Our findings show that there is a real and persistent phenomenon of the iterates being attracted to multipliers violating SOSC. More than that, the iterates are attracted to a rather special subset of multipliers violating SOSC, which we call *critical multipliers*.

Our conclusions remain relevant also for the general case of mixed equality and inequality constraints, if the method in question asymptotically behaves as a Newton method for some associated equality-constrained problem.

Other type of methods to which our conclusions readily apply are those based on active-set strategies, i.e., methods that are supposed to switch to equality-constrained phase asymptotically, by the construction of the algorithm.

Generally speaking, in all examples violating LICQ but satisfying SOSC that we have examined so far (analytically or numerically), slow convergence of SQP is observed for one of the following two reasons: either the dual sequence converges to a multiplier which is critical for some related equality-constrained problem, or the dual sequence does not converge at all. Out of the two, the first scenario seems more common, at least when critical multipliers exist.

References

1. Anitescu, M.: Nonlinear programs with unbounded Lagrange multiplier sets. Preprint ANL/MCS-P796-0200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne (2000)
2. Anitescu, M.: On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. SIAM J. Optim. **15**, 1203–1236 (2005)
3. Anitescu, M., Tseng, P., Wright, S.J.: Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. Math. Program. (in press)
4. Arutyunov, A.V.: Optimality Conditions: Abnormal and Degenerate Problems.   Kluwer, Dordrecht (2000)
5. Arutyunov, A.V.: Optimality conditions of higher order for abnormal minimization problems. Siberian Math. J. **33**, 557–565 (1992)

6. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
7. Fischer, A.: Modified Wilson's method for nonlinear programs with nonunique multipliers. Math. Oper. Res. **24**, 699–727 (1999)
8. Fischer, A.: Local behaviour of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
9. Fletcher, R., Leyffer, S., Ralph, D., Scholtes, S.: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. **17**, 259–286 (2006)
10. Golubitsky, M., Schaeffer, D.G.: Singularities and Groups in Bifurcation Theory, vol. 1. Springer, Heidelberg (1985)
11. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**, 253–273 (1999)
12. Hager, W.W., Gowda, M.S.: Stability in the presence of degeneracy and error estimation. Math. Program. **85**, 181–192 (1999)
13. Izmailov, A.F.: On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. **45**, 930–946 (2005)
14. Izmailov, A.F.: 2-Regularity and reversibility of quadratic mappings. Contemporary Math. **272**, 127–134 (2000)
15. Izmailov, A.F., Solodov, M.V.: Optimality conditions for irregular inequality-constrained problems. SIAM J. Control Optim. **40**, 1280–1295 (2001)
16. Izmailov, A.F., Solodov, M.V.: The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. **27**, 614–635 (2002)
17. Izmailov, A.F., Solodov, M.V.: Complementarity constraint qualification via the theory of 2-regularity. SIAM J. Optim. **13**, 368–385 (2002)
18. Izmailov, A.F., Solodov, M.V.: Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim. **15**, 210–228 (2004)
19. Izmailov, A. F., Tretyakov, A. A.: 2-Regular Solutions of Nonlinear Problems. Theory and Numerical Methods (in Russian). Fizmatlit, Moscow (1999)
20. Leyffer, S.: MacMPEC: AMPL collection of MPECs. www.mcs.anl.gov/leyffer/MacMPEC/
21. Li, D.-H., Qi, L.: A stabilized SQP method via linear equations. Applied mathematics technical report AMR00/5, The University of New South Wales (2000)
22. Luo, Z.-Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
23. Maratos, N.: Exact penalty function algorithms for finite dimensional and control optimization problems. PhD thesis, Imperial College, London (1978)
24. Robinson, S.M.: False numerical convergence in some generalized Newton methods. In: Equilibrium problems and variational models (Erice, 2000), Nonconvex Optim. Appl., vol. 68, pp 401–416. Kluwer, Norwell (2003)
25. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
26. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)
27. Scholtes, S., Stöhr, M.: Exact penalization of mathematical programs with equilibrium constraints. SIAM J. Control Optim. **37**, 617–652 (1999)
28. Scholtes, S., Stöhr, M.: How stringent is the linear independence assumption for mathematical programs with complementarity constraints?. Math. Oper. Res. **26**, 851–863 (2001)
29. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
30. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**, 470–497 (2002)
31. Wright, S.J.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program. **95**, 137–160 (2003)
32. Wright, S.J.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**, 673–696 (2005)