SIAM J. OPTIM.
Vol. 30, No. 2, pp. 1555–1581

ADJUSTING DUAL ITERATES IN THE PRESENCE OF CRITICAL LAGRANGE MULTIPLIERS*

ANDREAS FISCHER†, ALEXEY F. IZMAILOV‡, AND WLADIMIR SCHECK†

Abstract. It is a well-known phenomenon that the presence of critical Lagrange multipliers in constrained optimization problems may cause a deterioration of the convergence speed of primal-dual Newton-type methods. Regardless of the method under consideration, we develop a new local technique for avoiding convergence to critical Lagrange multipliers of equality-constrained optimization problems. This technique consists of replacing dual iterates of the methods by a special function of primal iterates. Under some natural assumptions, this function yields an approximation of a Lagrange multiplier, whose quality agrees with the distance from the primal iterate to the respective stationary point, while at the same time staying away from the critical multiplier in question. The accelerating effect of this technique is demonstrated by numerical experiments for stabilized sequential quadratic programming, the Levenberg–Marquardt method, and the LP-Newton method.

Key words. constrained optimization, degenerate constraints, critical multipliers, stabilized sequential quadratic programming, Levenberg–Marquardt method, LP-Newton method

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

DOI. 10.1137/19M1255380

1. Introduction.

Critical Lagrange multipliers, when they exist, are known to be especially attractive for dual sequences generated by primal-dual Newton-type methods for constrained optimization and variational problems, and this phenomenon is the reason for typically slow convergence in such circumstances (see [17] and [16, section 7.1], and references therein, as well as recent related developments and extensions of the criticality concept in [3, 4, 19, 20, 24]). The existence of critical multipliers is quite common when the constraints are degenerate, i.e., violate some standard constraint qualifications, and in particular, when the set of Lagrange multipliers associated to the primal solution in question is not a singleton.

Moreover, even for special modifications of the basic Newtonian schemes, developed intentionally for tackling cases of possible degeneracy, the effect of attraction to critical multipliers still exists, even though the domains of attraction usually become somewhat smaller (see [9, 13]). Indeed, the stabilized sequential quadratic programming (sSQP) method introduced originally in [21] (see also [15, 22] and [16, section 7.2.2]), the Levenberg–Marquardt (LM) method with a reasonably controlled regularization [23] (see also [6, 8]), and the LP-Newton (LP-N) method [5] possess very strong local convergence properties assuming that they are initialized near a noncritical multiplier: in this case, they converge superlinearly or even quadratically to the stationary point in question and to some noncritical Lagrange multiplier, despite degeneracy of constraints. However, the closer a noncritical multiplier is to a critical

*Received by the editors April 10, 2019; accepted for publication (in revised form) March 29, 2020; published electronically June 8, 2020.
https://doi.org/10.1137/19M1255380
Funding: The first author is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), 409756759. The second author is supported by the Russian Foundation for Basic Research, grants 19-51-12003 NNIO_a and 20-01-00106. The third author is supported by the Volkswagen Foundation.
†Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany (andreas.fischer@tu-dresden.de, wladimir.scheck@tu-dresden.de).
‡Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia (izmaf@ccas.ru).

one, the smaller is the convergence domain that arises from the noncritical multiplier, and eventually the effect of attraction to critical multipliers often does not allow for those nice local convergence properties to show up.

In this paper, for equality-constrained optimization problems, and for a given approximation of the primal solution in question, we present a universal (i.e., not related to any specific algorithm) local technique allowing us to obtain an approximation of a Lagrange multiplier, of the same "quality" as the primal approximation, while at the same time staying away from the critical multiplier in question. When combined with various stabilized/regularized Newton-type methods, this allows us to further reduce the attraction domain of this critical multiplier, which may serve for future development of effective practical globalizations of such algorithms.

The structure of the paper is as follows. In section 2, we give the formal problem setting and some necessary preliminaries. Section 3 presents our new technique for adjusting approximations of Lagrange multipliers and shows that it indeed possesses the properties highlighted above. Then, in section 4, we give an example of how this technique can be incorporated into an algorithm. Specifically, we consider the primal version of sSQP, with dual estimates satisfying some requirements, and demonstrate that our dual estimates fulfill these requirements, in some sense. In section 5, we consider the case when there are no critical multipliers associated to the stationary point in question and demonstrate that our technique is usually not harmful. Section 6 contains numerical results confirming the theoretical conclusion of section 3. We compare the local performance of the sSQP method, the LM method, and the LP-N method, with and without the technique for adjusting dual iterates. Finally, section 7 summarizes the obtained results and outlines some directions for future work.

2. Problem setting and preliminaries. We consider the equality-constrained optimization problem

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad h(x) = 0, \tag{2.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint mapping $h : \mathbb{R}^n \to \mathbb{R}^l$ are at least twice differentiable.

The primal-dual first-order optimality conditions for problem (2.1), characterizing its stationary points and associated Lagrange multipliers, are given by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad h(x) = 0, \tag{2.2}$$

where $L : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ is the Lagrangian of problem (2.1), i.e.,

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

We note that, throughout the paper, scalar products and norms are Euclidean.

According to (2.2), stationarity of $\bar{x} \in \mathbb{R}^n$ means that it is feasible in (2.1), and the set of associated Lagrange multipliers

$$\Lambda(\bar{x}) = \left\{ \lambda \in \mathbb{R}^l \;\middle|\; \frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0 \right\}$$

is nonempty. It is well-known that every local solution $\bar{x}$ of problem (2.1), satisfying the constraint qualification

$$\operatorname{rank} h'(\bar{x}) = l, \tag{2.3}$$

is a stationary point of this problem, and moreover, in this case $\Lambda(\bar{x})$ is necessarily a singleton. However, in this work we are mostly interested in those cases when $\bar{x}$ is stationary, but (2.3) does not hold, implying that $\Lambda(\bar{x})$ is an affine manifold of a positive dimension.

Recall that a multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is called critical if there exists $\bar{\xi} \in \ker h'(\bar{x}) \setminus \{0\}$ such that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\bar{\xi} \in \operatorname{im}(h'(\bar{x}))^\top \tag{2.4}$$

and noncritical otherwise (see [16, Definition 1.41]). Criticality of $\bar{\lambda}$ is equivalent to singularity of the reduced (to $\ker h'(\bar{x})$) Hessian of the Lagrangian, which is the symmetric matrix $\mathcal{H}(\bar{x}, \bar{\lambda})$ of the quadratic form

$$\xi \to \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \, \xi \right\rangle : \ker h'(\bar{x}) \to \mathbb{R}. \tag{2.5}$$

Throughout the paper, I stands for the identity matrix whose dimensions are always clear from the context. By im and ker we denote the range space and the null space of a linear operator, respectively. In addition to the standard use of big-O and little-o notation (see, e.g., [16, Appendix A.2]), we employ the following. By $O(1)$ ($o(1)$) we denote any function whose upper limit is finite (respectively, whose limit is zero) as its arguments tend to specified values. Furthermore, when writing $O(1)x$ ($O(t)x$, $o(t)x$), we always mean the product of a matrix and x, where the norm of the matrix is $O(1)$ ($O(t)$, $o(t)$, respectively).

3. Adjusting dual iterates. At the beginning of this section, let us mention that a procedure for avoiding convergence to critical multipliers has already been proposed in [14]. Unfortunately, this procedure employing gradient steps does not help to accelerate the overall process.

Now, we introduce the new technique for obtaining an approximation of a Lagrange multiplier with the properties described in section 1. For a given primal point $x \in \mathbb{R}^n$, consider the following optimization problem with quadratic objective function and linear equality constraints:

$$\underset{\xi}{\text{minimize}} \quad \langle f'(x), \xi\rangle + \frac{1}{2}\|\xi\|^2 \quad \text{subject to} \quad h(x) + h'(x)\xi = 0. \tag{3.1}$$

Here, minimization is with respect to $\xi \in \mathbb{R}^n$, while x serves as a parameter. The impetus for using the auxiliary program (3.1) stems from [7]. There, for inequality-constrained optimization problems, a related subproblem was employed for a correction of dual multipliers. This correction aimed at saving quadratic convergence of the Wilson method for circumstances which allow nonunique but noncritical multipliers.

Assuming that $\operatorname{rank} h'(x) = l$, problem (3.1) has the unique solution $\widetilde{\xi}(x)$, and by direct computation it can be seen that the unique Lagrange multiplier associated to $\widetilde{\xi}(x)$ is

$$\widetilde{\lambda}(x) = \bar{\lambda}(x) + \widehat{\lambda}(x), \tag{3.2}$$

where

$$\bar{\lambda}(x) = -(H(x))^{-1}h'(x)f'(x) \quad \text{and} \quad \widehat{\lambda}(x) = (H(x))^{-1}h(x) \tag{3.3}$$

with

$$H(x) = h'(x)(h'(x))^\top. \tag{3.4}$$

This $\widetilde{\lambda}(x)$ is exactly what we suggest to use as an approximation of a Lagrange multiplier.

In the remainder of this section, we demonstrate that this choice of $\widetilde{\lambda}(x)$ satisfies the desired properties. Specifically, for a given stationary point $\bar{x}$ and a critical Lagrange multiplier $\bar{\lambda}$ associated with it, satisfying certain assumptions, we construct a "large" set of points $x \in \mathbb{R}^n$ (starlike with respect to $\bar{x}$, and with nonempty interior), such that for x from this set, on one hand, $\widetilde{\lambda}(x)$ stays separated away from $\bar{\lambda}$, while on the other hand, $\operatorname{dist}(\widetilde{\lambda}(x),\, \Lambda(\bar{x}))$ behaves like $O(\|x - \bar{x}\|)$.

In section 3.1, we consider the fully degenerate case, in order to deliver the main idea without too many technicalities. The general case is considered in section 3.2.

3.1. Fully degenerate case. Let $\bar{x}$ be a stationary point of problem (2.1). Here, we suppose that $h'(\bar{x}) = 0$ (i.e., the constraints are fully degenerate). Then, stationarity of $\bar{x}$ subsumes that $f'(\bar{x}) = 0$, and $\Lambda(\bar{x}) = \mathbb{R}^l$.

Let $\bar{\lambda}$ be a critical Lagrange multiplier associated to $\bar{x}$, which in the fully degenerate case means the existence of $\bar{\xi} \in \mathbb{R}^n$ such that $\|\bar{\xi}\| = 1$ and

$$f''(\bar{x})\bar{\xi} + (h''(\bar{x})[\bar{\xi}])^\top \bar{\lambda} = 0. \tag{3.5}$$

Assume further that h is 2-regular in the direction $\bar{\xi}$, i.e.,

$$\operatorname{rank} h''(\bar{x})[\bar{\xi}] = l. \tag{3.6}$$

Then, for any $\varepsilon > 0$ and $\delta > 0$, define the set

$$K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi}) = \left\{ x \in \mathbb{R}^n \setminus \{\bar{x}\} \;\middle|\; \|x - \bar{x}\| \le \varepsilon,\; \left\| \frac{x - \bar{x}}{\|x - \bar{x}\|} - \bar{\xi} \right\| \le \delta \right\}. \tag{3.7}$$

Since

$$h'(x) = h''(\bar{x})[x - \bar{x}] + o(\|x - \bar{x}\|) \tag{3.8}$$

as $x \to \bar{x}$, it can be easily seen that, by (3.6), there exist $\varepsilon = \varepsilon(\bar{\xi}) > 0$ and $\delta = \delta(\bar{\xi}) > 0$ such that

$$\operatorname{rank} h'(x) = l \quad \text{for all } x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi}) \tag{3.9}$$

and, further,

$$\left\|(H(x))^{-1}\right\| = O(\|x - \bar{x}\|^{-2}) \text{ as } x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi}) \text{ tends to } \bar{x}. \tag{3.10}$$

Therefore, $\bar{\lambda}(x)$ and $\widehat{\lambda}(x)$ are well-defined by (3.3) for any $x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi})$. Moreover, since $f'(\bar{x}) = 0$, employing (3.5) and (3.8)–(3.10) yields, for $x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi})$,

$$\begin{aligned}
\bar{\lambda}(x) &= -(H(x))^{-1} h'(x) f'(x) \\
&= -(H(x))^{-1} h'(x)(f''(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|)) \\
&= -(H(x))^{-1} h'(x) f''(\bar{x})(\|x - \bar{x}\|\bar{\xi}) + O(\delta) + o(1) \\
&= (H(x))^{-1} h'(x)(h''(\bar{x})[\|x - \bar{x}\|\bar{\xi}])^\top \bar{\lambda} + O(\delta) + o(1) \\
&= (H(x))^{-1} H'(x)\bar{\lambda} + O(\delta) + o(1) \\
&= \bar{\lambda} + O(\delta) + o(1)
\end{aligned}$$

as $\delta \to 0$ and $x \to \bar{x}$. This implies that by choosing $\varepsilon > 0$ and $\delta > 0$ small enough, $\bar{\lambda}(x)$ can be made arbitrarily close to $\bar{\lambda}$ for all $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$.

We now make the additional assumption that

$$h''(\bar{x})[\bar{\xi},\,\bar{\xi}] \neq 0. \tag{3.11}$$

Observe further that

$$h(x) = \frac{1}{2}h''(\bar{x})[x-\bar{x},\,x-\bar{x}] + o(\|x-\bar{x}\|^2) \tag{3.12}$$

as $x \in \mathbb{R}^n$ tends to $\bar{x}$. Moreover, from (3.8) and the definition of $H(x)$ in (3.4), we obtain the existence of $\Gamma > 0$ such that

$$\|H(x)\| \le \Gamma\|x-\bar{x}\|^2$$

holds for all $x \in \mathbb{R}^n$ close enough to $\bar{x}$. Taking this into account, from the definition of $\widehat{\lambda}(x)$ in (3.3), and from (3.7) and (3.12), we have that, for $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$,

$$\begin{aligned}\|\widehat{\lambda}(x)\| &\ge \|H(x)\|^{-1}\|h(x)\|\\ &\ge \frac{1}{2\Gamma}\|x-\bar{x}\|^{-2}\|h''(\bar{x})[x-\bar{x},\,x-\bar{x}]\| + o(1)\\ &= \frac{1}{2\Gamma}\|h''(\bar{x})[\bar{\xi},\,\bar{\xi}]\| + O(\delta) + o(1)\end{aligned}$$

as $\delta \to 0$ and $x \to \bar{x}$. Employing (3.11), the latter implies that by choosing $\varepsilon > 0$ and $\delta > 0$ small enough, $\|\widehat{\lambda}(x)\|$ can be kept separated from zero by a positive constant for all $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$.

Example 3.1. Consider the case when (2.1) is a fully quadratic problem with a single constraint: let $l = 1$,

$$f(x) = \frac{1}{2}\langle Ax,\,x\rangle, \quad h(x) = \frac{1}{2}\langle Bx,\,x\rangle,$$

where A and B are symmetric $n\times n$ matrices. Then, for $\bar{x} = 0$, it holds that $h(\bar{x}) = 0$, $f'(\bar{x}) = h'(\bar{x}) = 0$, and the analysis above applies with any $\bar{\lambda} \in \mathbb{R}$ and $\bar{\xi} \in \mathbb{R}^n$ such that $\|\bar{\xi}\| = 1$ and

$$(A+\bar{\lambda}B)\bar{\xi} = 0, \quad \langle B\bar{\xi},\,\bar{\xi}\rangle \neq 0, \tag{3.13}$$

according to (3.5) and (3.11). And indeed, by (3.3), we have that, for $x \in \mathbb{R}^n$ with $Bx \neq 0$,

$$\bar{\lambda}(x) = -\frac{\langle Ax,\,Bx\rangle}{\|Bx\|^2} \quad \text{and} \quad \widehat{\lambda}(x) = \frac{1}{2}\frac{\langle Bx,\,x\rangle}{\|Bx\|^2} \tag{3.14}$$

hold. Hence, taking, e.g., $x = t\bar{\xi} + o(t)$, $t \in \mathbb{R}\setminus\{0\}$, we obtain by (3.13) that

$$\bar{\lambda}(x) = \frac{\langle \bar{\lambda}B\bar{\xi},\,B\bar{\xi}\rangle}{\|B\bar{\xi}\|^2} + o(1) = \bar{\lambda} + o(1),$$

while

$$\widehat{\lambda}(x) = \frac{1}{2}\frac{\langle B\bar{\xi},\,\bar{\xi}\rangle}{\|B\bar{\xi}\|^2} + o(1),$$

as $t \to 0$. Clearly, $\widehat{\lambda}(x)$ is separated from zero for t close enough to zero, thus keeping $\widetilde{\lambda}(x) = \bar{\lambda}(x) + \widehat{\lambda}(x)$ away from the critical multiplier $\bar{\lambda}$.

In particular, if $n = 1$, then (3.13) reduces to

$$\bar{\lambda} = -A/B, \quad B \neq 0,$$

and for every $x \neq 0$ it holds that

$$\bar{\lambda}(x) = -\frac{ABx^2}{B^2x^2} = \bar{\lambda}, \quad \widehat{\lambda}(x) = \frac{1}{2}\frac{Bx^2}{B^2x^2} = \frac{1}{2B}. \tag{3.15}$$

In this case, $\widetilde{\lambda}(x) = \bar{\lambda}(x) + \widehat{\lambda}(x) = \bar{\lambda} + 1/(2B)$ does not depend on x and does not coincide with the unique critical multiplier $\bar{\lambda} = -A/B$.

For $n = 2$, let us consider

$$A = \begin{pmatrix} a_1 & a \\ a & a_2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

where a_1, a_2, and a are real parameters. From (3.14), it then follows that

$$\bar{\lambda}(x) = -a_1 - a\frac{x_2}{x_1}, \quad \widehat{\lambda}(x) = 1/2,$$

whenever $x_1 \neq 0$, while otherwise, both $\bar{\lambda}(x)$ and $\widetilde{\lambda}(x)$ are not well-defined.

Furthermore, assuming that $a_2 \neq 0$, the unique critical multiplier is

$$\bar{\lambda} = \frac{a^2}{a_2} - a_1,$$

and (3.13) and the 2-regularity assumption (3.6) hold with $\bar{\xi} = (\theta, -\theta a/a_2)$ for every $\theta \neq 0$. And indeed, if we take, e.g., $x = t\bar{\xi} + o(t)$, then

$$\widetilde{\lambda}(x) = \bar{\lambda}(x) + \widehat{\lambda}(x) = \bar{\lambda} + 1/2 + o(1),$$

which again agrees with the theory above.

In contrast to this, consider now $x = t\xi$ with the direction ξ distinct from $\bar{\xi}$, in which case x does not need to belong to $K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$, no matter how close to zero t is. Observe first that if $\xi_1 = 0$, then both $\bar{\lambda}(x)$ and $\widetilde{\lambda}(x)$ are not well-defined, demonstrating that taking x within $K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$ with appropriate $\varepsilon > 0$ and $\delta > 0$ is essential for this reason already.

Furthermore, let $\xi = (1, \tau)$ with some real τ. Then, $\bar{\lambda}(x)$ and $\widetilde{\lambda}(x)$ are well-defined, and $\widetilde{\lambda}(x) = \bar{\lambda}(x) + \widehat{\lambda}(x) = -a_1 - a\tau + 1/2$. If $a = 0$, then $\widetilde{\lambda}(x) = \bar{\lambda} + 1/2$, which is a successful outcome. Otherwise, $\widetilde{\lambda}(x)$ can be *whatever*, depending on τ. In particular, if we take $\tau = -a/a_2 + 1/(2a)$, then $\widetilde{\lambda}(x) = \bar{\lambda}$, meaning that no adjusting of the dual iterate takes effect. This further supports the claim that taking x within $K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$ is essential in the analysis above (and hence, in Theorem 3.4 below).

Suppose now that $a_2 = 0$, and let $a = 0$ (since otherwise, there are no critical multipliers). Then, all multipliers are critical, and hence, we cannot expect any use of trying to adjust the dual iterate. Nevertheless, let us try to apply our technique with, say, $\bar{\lambda} = -1$ (which is critical of order 2, in the terminology of [13, 18]). The equality in (3.13) is satisfied with any $\bar{\xi} \in \mathbb{R}^2$, but if we take, say, $\bar{\xi} = (0, 1)$, then the 2-regularity assumption (3.6) does not hold, which in this case amounts to saying that $B\bar{\xi} = 0$. Now if we take $x = t\bar{\xi}$, then both $\bar{\lambda}(x)$ and $\widehat{\lambda}(x)$ in (3.14), and hence, $\widetilde{\lambda}(x)$, are not well-defined. This demonstrates that the 2-regularity assumption (3.6) in the analysis above (and hence, the corresponding assumption for the general case in Theorem 3.4 below) cannot be dropped.

Let us now consider

$$A = B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then, there is the unique critical multiplier $\bar{\lambda} = -1$ (also of order 2), and both the equality in (3.13) and 2-regularity condition (3.6) hold for any $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$, including, say, $\bar{\xi} = (1, -1)$, for which the inequality in (3.13) fails. Therefore, if we take $x = t\bar{\xi}$, then by (3.14) it holds that $\widehat{\lambda}(x) = 0$, and $\widetilde{\lambda}(x) = \bar{\lambda}(x) = \bar{\lambda}$, again meaning that no any adjusting of the dual iterate takes effect. This demonstrates that assumption (3.11) in the analysis above (and hence, the corresponding assumption (3.23) for the general case in Theorem 3.4 below) cannot be dropped.

Finally, we show an example where all the needed assumptions are satisfied, all the conclusions above are valid, and x is taken appropriately, but this does not help to escape criticality, as getting far from some critical multiplier, one may get close to another. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1/2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then, there are exactly two critical multipliers: $\bar{\lambda} = \bar{\lambda}^1 = -1$ and $\bar{\lambda} = \bar{\lambda}^2 = -1/2$, and for the former the first equality in (3.13) holds with every $\bar{\xi} \in \mathbb{R}^2$ satisfying $\bar{\xi}_2 = 0$, while for the latter it holds with every $\bar{\xi} \in \mathbb{R}^2$ satisfying $\bar{\xi}_1 = 0$. Since B is nonsingular, we obtain from (3.14) that

$$\bar{\lambda}(x) = -\frac{x_1^2 + x_2^2/2}{x_1^2 + x_2^2}, \quad \widehat{\lambda}(x) = \frac{1}{2}\frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} \quad \text{for all } x \in \mathbb{R}^n \setminus \{0\}.$$

Suppose, e.g., that $x_2 = 0$; then $\bar{\lambda}(x) = -1 = \bar{\lambda}^1$ (as it should be, according to the theory above), while $\widehat{\lambda}(x) = 1/2$, and hence, $\widetilde{\lambda}(x) = -1/2 = \bar{\lambda}^2$. Similarly, let $x_1 = 0$ be satisfied; then $\bar{\lambda}(x) = -1/2 = \bar{\lambda}^2$ (as it should be), while $\widehat{\lambda}(x) = -1/2$, and hence, $\widetilde{\lambda}(x) = -1 = \bar{\lambda}^1$. Therefore, in the former case, $\widetilde{\lambda}(x)$ is kept away from $\bar{\lambda}^1$ (as it should be, according to the theory above), but it coincides with the other critical multiplier $\bar{\lambda}^2$. Similarly, in the latter case, $\widetilde{\lambda}(x)$ is kept away from $\bar{\lambda}^2$ (as it should be), but it coincides with the critical multiplier $\bar{\lambda}^1$.

The final part of Example 3.1 demonstrates that when there is more than one critical multiplier, the goal of keeping the constructed dual estimate away from them may not be achieved: moving it away from one critical multiplier can make it close to another. Theoretical justification of our dual modification procedure consists of showing that under some relevant assumptions on the primal approximations, relating them to a specific critical multiplier $\bar{\lambda}$ (through $\bar{\xi}$ associated with the latter), the adjusted dual approximation will be kept away from $\bar{\lambda}$ while being still close to the set of multipliers. In other words, each critical multiplier $\bar{\lambda}$ gives rise to some domain of x where our $\widetilde{\lambda}(x)$ possesses the needed properties. In particular, if x belongs to the intersection of all such domains, $\widetilde{\lambda}(x)$ will be separated from all critical multipliers. However, in general, none of these domains needs to be a full neighborhood of $\bar{x}$. Therefore, the adjusted dual approximation computed in the domain related to a given critical multiplier can be close to any other critical multiplier with a different associated $\bar{\xi}$, and such undesirable behavior is indeed possible, as demonstrated by an artificial example above. However, in general, there is no special reason for this to happen, and this is confirmed by the numerical experiments in section 6.

3.2. General case. Let $\bar{x}$ be a stationary point of problem (2.1). We do not assume anymore that $h'(\bar{x}) = 0$. In this general case, 2-regularity of h at $\bar{x}$ in a direction $\xi \in \mathbb{R}^n$ consists of saying that

$$\operatorname{rank}(h'(\bar{x}) + Ph''(\bar{x})[\xi]) = l,$$

where P stands for the orthogonal projector in $\mathbb{R}^l$ onto $(\operatorname{im} h'(\bar{x}))^\perp$. For the use of this concept in nonlinear analysis and optimization theory see, e.g., [1] and references therein. 2-regularity of h at $\bar{x}$ in a direction ξ is further equivalent to saying that the linear operator $H_2(\xi) : \ker h'(\bar{x}) \to (\operatorname{im} h'(\bar{x}))^\perp$, $H_2(\xi)x = Ph''(\bar{x})[\xi, x]$, is onto. Evidently, this property is automatic for every $\xi \in \mathbb{R}^n$ (including $\xi = 0$) provided (2.3) holds. Observe further that by the result on small perturbation of the surjective linear operator (see, e.g., [16, Lemma A.5]), and by homogeneousness of $H_2(\cdot)$, if the surjectivity property holds with $\xi = \bar{\xi}$ for some $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$, then it also holds for every $\xi \in \mathbb{R}^n \setminus \{0\}$ such that $\xi/\|\xi\|$ is close enough to $\bar{\xi}$, and moreover, there exists $N_2 > 0$ such that for all such ξ, and all $\eta \in (\operatorname{im} h'(\bar{x}))^\perp$, the equation

$$H_2(\xi)x = \eta$$

has a solution $x(\xi, \eta)$ satisfying

$$\|x(\xi, \eta)\| \le N_2\|\xi\|^{-1}\|\eta\|.$$

Let Π stand for the orthogonal projector in $\mathbb{R}^n$ onto $\ker h'(\bar{x})$.

Lemma 3.2. *Let $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$, and assume that it is 2-regular at $\bar{x}$ in a direction $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$.*

Then, there exist $\varepsilon = \varepsilon(\bar{\xi}) > 0$ and $\delta = \delta(\bar{\xi}) > 0$ such that (3.9) *holds with $K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$ defined in* (3.7) *and, moreover, there exists $N > 0$ such that for all $x \in K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$ and $y \in \mathbb{R}^l$, the equation*

$$h'(x)\xi = y \tag{3.16}$$

has a solution $\xi(x, y)$ satisfying

$$\|(I - \Pi)\xi(x, y)\| \le N\|y\|, \quad \|\Pi\xi(x, y)\| \le N(\|(I - P)y\| + \|x - \bar{x}\|^{-1}\|Py\|). \tag{3.17}$$

Proof. The proof is essentially by applying the Lyapunov–Schmidt procedure (see, e.g., [11, Chapter VII], in a somewhat simplified form as the system of equations to which we apply it here is linear, and there is no need to involve any implicit function theorems). Setting $\xi_1 = (I - \Pi)\xi$, $\xi_2 = \Pi\xi$, we can write (3.16) as

$$(I - P)y = (I - P)h'(x)\xi = H_1\xi_1 + (I - P)O(\|x - \bar{x}\|)(\xi_1 + \xi_2), \tag{3.18}$$

$$Py = Ph'(x)\xi = Ph''(\bar{x})[x - \bar{x}, \xi_1 + \xi_2] + Po(\|x - \bar{x}\|)(\xi_1 + \xi_2), \tag{3.19}$$

where $H_1 : (\ker h'(\bar{x}))^\perp \to \operatorname{im} h'(\bar{x})$, $H_1\xi_1 = h'(\bar{x})\xi_1$, is an invertible linear operator. By the standard result on small perturbation of an invertible linear operator (see, e.g., [16, Lemma A.6]), we then have that for any fixed ξ_2, if $\varepsilon > 0$ is small enough, (3.18) with respect to ξ_1 has the unique solution $\xi_1(x, y, \xi_2)$, and this solution satisfies

$$\xi_1(x,\, y,\, \xi_2) = O(1)(I-P)y + O(\|x-\bar{x}\|)\xi_2. \tag{3.20}$$

Substituting this expression into (3.19), we come to the linear equation

$$(H_2(x-\bar{x}) + Po(\|x-\bar{x}\|))\xi_2 = PO(\|x-\bar{x}\|)(I-P)y + Py$$

with respect to ξ_2. By the result on small perturbation of a surjective linear operator, and by the observations accompanying the definition of 2-regularity above, we conclude that this equation has a solution $\xi_2(x,\, y)$ satisfying

$$\xi_2(x,\, y) = O(1)(I-P)y + O(\|x-\bar{x}\|^{-1})Py$$

provided $\varepsilon > 0$ and $\delta > 0$ are small enough. Substituting this solution into (3.20), we obtain a solution $\xi(x,\, y) = \xi_1(x,\, y,\, \xi_2(x,\, y)) + \xi_2(x,\, y)$ of (3.16), satisfying (3.17). $\square$

LEMMA 3.3. *Under the assumptions of Lemma 3.2, there exist $\varepsilon = \varepsilon(\bar{\xi}) > 0$, $\delta = \delta(\bar{\xi}) > 0$, and $N > 0$ such that $H(x)$ defined in (3.4) is nonsingular for every $x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi})$, and for every $y \in \mathbb{R}^l$ it holds that*

$$\|(I-P)(H(x))^{-1}y\| \le N(\|(I-P)y\| + \|x-\bar{x}\|^{-1}\|Py\|), \tag{3.21}$$

$$\|P(H(x))^{-1}y\| \le N(\|x-\bar{x}\|^{-1}\|(I-P)y\| + \|x-\bar{x}\|^{-2}\|Py\|). \tag{3.22}$$

The proof again relies on the Lyapunov–Schmidt procedure and is similar to the one of Lemma 3.2. We skip the proof since it is much more technical and does not contain new ideas.

THEOREM 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$, and let $\bar{x}$ be a stationary point of problem (2.1), with an associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$ which is critical, i.e., there exists $\bar{\xi} \in \ker h'(\bar{x})$ such that $\|\bar{\xi}\| = 1$ and (2.4) holds. Assume that h is 2-regular at $\bar{x}$ in a direction $\bar{\xi}$, and*

$$h''(\bar{x})[\bar{\xi},\, \bar{\xi}] \notin \operatorname{im} h'(\bar{x}). \tag{3.23}$$

Then, there exist $\varepsilon = \varepsilon(\bar{\xi}) > 0$, $\delta = \delta(\bar{\xi}) > 0$, and $\gamma = \gamma(\bar{\xi}) > 0$, such that $\widetilde{\lambda}(x)$ is well-defined by (3.2)–(3.4) for any $x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi})$, and

$$\|\widetilde{\lambda}(x) - \bar{\lambda}\| \ge \gamma, \tag{3.24}$$

$$\operatorname{dist}(\widetilde{\lambda}(x),\, \Lambda(\bar{x})) = O(\|x-\bar{x}\|) \tag{3.25}$$

as $x \to \bar{x}$.

Proof. From (2.4) we have that there exists $\bar{\eta} \in \mathbb{R}^l$ such that

$$f''(\bar{x})\bar{\xi} + ((h''(\bar{x})[\bar{\xi}])^\top \bar{\lambda} = (h'(\bar{x}))^\top \bar{\eta}. \tag{3.26}$$

For any $\varepsilon > 0$ and $\delta > 0$, and for any $x \in K_{\varepsilon,\,\delta}(\bar{x},\, \bar{\xi})$, employing (3.7) we have

$$\begin{aligned} h'(x) &= h'(\bar{x}) + h''(\bar{x})[x-\bar{x}] + o(\|x-\bar{x}\|) \\ &= h'(\bar{x}) + h''(\bar{x})[\|x-\bar{x}\|\bar{\xi}] + O(\delta\|x-\bar{x}\|) + o(\|x-\bar{x}\|). \end{aligned} \tag{3.27}$$

Hence, by the inclusion $\bar{\lambda} \in \Lambda(\bar{x})$ and by (3.26), it follows that

$$
\begin{aligned}
f'(x) &= f'(\bar{x}) + f''(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|) \\
&= f'(\bar{x}) + f''(\bar{x})(\|x - \bar{x}\|\bar{\xi}) + O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|) \\
&= -(h'(\bar{x}))^\top \bar{\lambda} - (h''(\bar{x})[\|x - \bar{x}\|\bar{\xi}])^\top \bar{\lambda} + \|x - \bar{x}\|(h'(\bar{x}))^\top \bar{\eta} \\
&\quad + O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|) \\
&= -(h'(x))^\top \bar{\lambda} + \|x - \bar{x}\|(h'(\bar{x}))^\top \bar{\eta} + O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|) \\
&= -(h'(x))^\top \bar{\lambda} + \|x - \bar{x}\|(h'(x))^\top \bar{\eta} + O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|) \qquad (3.28)
\end{aligned}
$$

as $\delta \to 0$ and $x \to \bar{x}$.

Assume now that $\varepsilon > 0$ and $\delta > 0$ are chosen according to Lemma 3.3. Then, for any $x \in K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$, the matrix $H(x)$ is invertible and $\bar{\lambda}(x)$ and $\widehat{\lambda}(x)$ are well-defined by (3.3). Thus, using (3.28), we obtain

$$
\begin{aligned}
\bar{\lambda}(x) &= -(H(x))^{-1}h'(x)f'(x) \\
&= (H(x))^{-1}H(x)\bar{\lambda} - \|x - \bar{x}\|(H(x))^{-1}H(x)\bar{\eta} \\
&\quad - (H(x))^{-1}h'(x)(O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|)) \\
&= \bar{\lambda} - \|x - \bar{x}\|\bar{\eta} - (H(x))^{-1}h'(\bar{x})(O(\delta\|x - \bar{x}\|) + o(\|x - \bar{x}\|)) \\
&\quad - (H(x))^{-1}(O(\delta\|x - \bar{x}\|^2) + o(\|x - \bar{x}\|^2))
\end{aligned}
$$

for any $x \in K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$. By estimates (3.21) and (3.22) from Lemma 3.3, we now have

$$(I - P)\bar{\lambda}(x) = (I - P)\bar{\lambda} + O(\|x - \bar{x}\|), \qquad (3.29)$$

as $x \to \bar{x}$, and

$$P\bar{\lambda}(x) = P\bar{\lambda} + O(\delta) + o(1) \qquad (3.30)$$

as $\delta \to 0$ and $x \to \bar{x}$.

We will further need the following relations which can be easily derived from the first equality in (3.27), and the definition of $H(x)$ in (3.4), employing symmetry of P:

$$(I - P)H(x)(I - P) = h'(\bar{x})(h'(\bar{x}))^\top + O(\|x - \bar{x}\|), \qquad (3.31)$$

(3.32)

$$(I - P)H(x)P = (h'(\bar{x}) + O(\|x - \bar{x}\|))(Ph''(\bar{x})[x - \bar{x}] + o(\|x - \bar{x}\|))^\top = O(\|x - \bar{x}\|),$$

(3.33)

$$PH(x)(I - P) = (Ph''(\bar{x})[x - \bar{x}] + o(\|x - \bar{x}\|))(h'(\bar{x}) + O(\|x - \bar{x}\|))^\top = O(\|x - \bar{x}\|),$$

$$PH(x)P = Ph''(\bar{x})[x - \bar{x}](Ph''(\bar{x})[x - \bar{x}])^\top + o(\|x - \bar{x}\|^2) = O(\|x - \bar{x}\|^2) \qquad (3.34)$$

as $x \to \bar{x}$. Moreover, observe that

$$(I - P)h(x) = h'(\bar{x})(x - \bar{x}) + O(\|x - \bar{x}\|^2), \qquad (3.35)$$

$$Ph(x) = \frac{1}{2}Ph''(\bar{x})[x - \bar{x}, x - \bar{x}] + o(\|x - \bar{x}\|^2) \qquad (3.36)$$

as $x \to \bar{x}$. Then, taking into account the equality $P^2 = P$, the definition of $\widehat{\lambda}(x)$ in (3.3), relations (3.32) and (3.35), the definition of $K_{\varepsilon,\delta}(\bar{x},\bar{\xi})$ in (3.7), and the inclusion $\bar{\xi} \in \ker h'(\bar{x})$, it follows that

$$\begin{aligned}(I-P)H(x)(I-P)\widehat{\lambda}(x) &= (I-P)H(x)\widehat{\lambda}(x) - (I-P)H(x)P^2\widehat{\lambda}(x)\\ &= h'(\bar{x})(x-\bar{x}) + O(\|x-\bar{x}\|^2) + O(\|x-\bar{x}\|)P\widehat{\lambda}(x)\\ &= \|x-\bar{x}\|h'(\bar{x})\bar{\xi} + O(\delta\|x-\bar{x}\|) + O(\|x-\bar{x}\|^2)\\ &\quad + O(\|x-\bar{x}\|)P\widehat{\lambda}(x)\\ &= O(\delta\|x-\bar{x}\|) + O(\|x-\bar{x}\|^2) + O(\|x-\bar{x}\|)P\widehat{\lambda}(x) \end{aligned} \tag{3.37}$$

for any $x \in K_{\varepsilon,\delta}(\bar{x},\bar{\xi})$. In addition, for these x, exploiting (3.7) and (3.36), we get

$$\begin{aligned}PH(x)(I-P)^2\widehat{\lambda}(x) + PH(x)P^2\widehat{\lambda}(x) &= PH(x)(I-P)\widehat{\lambda}(x) + PH(x)P\widehat{\lambda}(x)\\ &= PH(x)\widehat{\lambda}(x)\\ &= \frac{1}{2}Ph''(\bar{x})[x-\bar{x},\, x-\bar{x}] + o(\|x-\bar{x}\|^2)\\ &= \frac{1}{2}\|x-\bar{x}\|^2 Ph''(\bar{x})[\bar{\xi},\,\bar{\xi}]\\ &\quad + O(\delta\|x-\bar{x}\|^2) + o(\|x-\bar{x}\|^2). \end{aligned} \tag{3.38}$$

Thanks to (3.31), the linear operator $L_x : \operatorname{im} h'(\bar{x}) \to \operatorname{im} h'(\bar{x})$ defined by

$$L_x\xi = (I-P)H(x)\xi = (I-P)H(x)(I-P)\xi$$

is invertible for all $x \in \mathbb{R}^n$ in some neighborhood of $\bar{x}$, with the inverse being uniformly bounded. Therefore, since the left-hand side of (3.37) is nothing else than $L_x(I-P)\widehat{\lambda}(x)$, (3.37) leads to

$$(I-P)\widehat{\lambda}(x) = O(\|x-\bar{x}\|)P\widehat{\lambda}(x) + O(\delta\|x-\bar{x}\|) + O(\|x-\bar{x}\|^2). \tag{3.39}$$

Now, substituting this expression into (3.38), and employing (3.33)–(3.34), we obtain

$$O(\|x-\bar{x}\|^2)P\widehat{\lambda}(x) = \frac{1}{2}\|x-\bar{x}\|^2 Ph''(\bar{x})[\bar{\xi},\,\bar{\xi}] + O(\delta\|x-\bar{x}\|^2) + o(\|x-\bar{x}\|^2) \tag{3.40}$$

as $\delta \to 0$ and $x \to \bar{x}$.

From now on, we employ assumption (3.23) or, in other terms, $Ph''(\bar{x})[\bar{\xi},\,\bar{\xi}] \neq 0$. Then, by choosing $\varepsilon > 0$ and $\delta > 0$ small enough, it follows from (3.40) that $\widehat{\gamma} > 0$ exists with

$$\|P\widehat{\lambda}(x)\| \geq \widehat{\gamma} \quad \text{for all } x \in K_{\varepsilon,\delta}(\bar{x},\bar{\xi}).$$

Using (3.2) and (3.30), we then have

$$\|\widetilde{\lambda}(x) - \bar{\lambda}\| \geq \|P(\widetilde{\lambda}(x) - \bar{\lambda})\| \geq \|P\widehat{\lambda}(x)\| - \|P(\bar{\lambda}(x) - \bar{\lambda})\| \geq \frac{1}{2}\widehat{\gamma}$$

for any $x \in K_{\varepsilon,\delta}(\bar{x},\bar{\xi})$ provided $\varepsilon > 0$ and $\delta > 0$ are chosen sufficiently small. Thus, (3.24) must hold with $\gamma = \widetilde{\gamma}/2$.

Observe now that (3.22) from Lemma 3.3, together with (3.35) and (3.36), implies that $P\widehat{\lambda}(\cdot)$ is bounded on $K_{\varepsilon,\delta}(\bar{x},\bar{\xi})$ whenever $\varepsilon > 0$ is small enough. This and (3.39) further yield

$$(I-P)\widehat{\lambda}(x) = O(\|x-\bar{x}\|) \tag{3.41}$$

as $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ tends to $\bar{x}$. Combining this with (3.29) and (3.30), we conclude that $\widetilde{\lambda}(\cdot)$ is bounded on $K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ with the specified $\varepsilon > 0$ and $\delta > 0$. Moreover, from (3.29) and (3.41) we obtain

$$\widetilde{\lambda}(x) = \bar{\lambda} + P(\widetilde{\lambda}(x)-\bar{\lambda}) + (I-P)(\widetilde{\lambda}(x)-\bar{\lambda}) = \bar{\lambda} + P(\widetilde{\lambda}(x)-\bar{\lambda}) + O(\|x-\bar{x}\|)$$

and, since $\bar{\lambda} + P(\widetilde{\lambda}(x)-\bar{\lambda}) \in \Lambda(\bar{x})$, we finally conclude that (3.25) holds as $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ tends to $\bar{x}$. ∎

Remark 3.5. If the subspace of $\bar{\xi} \in \ker h'(\bar{x})$ satisfying (2.4) has dimension higher than 1 (meaning that $\bar{\lambda}$ is critical of order higher than 1 [13, 18]), and if h is 2-regular at $\bar{x}$ in at least one direction $\bar{\xi}$ in this subspace, satisfying (3.23), then the set Ξ of such $\bar{\xi}$ with $\|\bar{\xi}\| = 1$ is open and dense in the intersection of this subspace with the unit sphere, as its complement is the intersection of null sets of some nontrivial homogeneous polynomials. In this case, Theorem 3.4 is applicable with all $\bar{\xi} \in \Xi$. In particular, if we take any closed $\overline{\Xi} \subset \Xi$, then the assertion of Theorem 3.4 will be valid with some $\varepsilon = \varepsilon(\overline{\Xi}) > 0$, $\delta = \delta(\overline{\Xi}) > 0$, and $\gamma = \gamma(\overline{\Xi}) > 0$, for all x from the union of $K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ over $\bar{\xi} \in \overline{\Xi}$. This allows us to further enlarge, in this case, the set of points $x \in \mathbb{R}^n$ for which $\widetilde{\lambda}(x)$ possesses the needed properties, still keeping this set starlike with respect to $\bar{x}$.

If there exists more than one critical multiplier associated with $\bar{x}$, Theorem 3.4 can be applicable with any of them, with associated $\bar{\xi}$ satisfying the needed requirements. Recall, however, the final part of Example 3.1.

4. Primal stabilized sequential quadratic programming. One of the algorithms being tested in section 6 is sSQP, where, for a given primal iterate x^k, the corresponding dual iterate is every time redefined as $\widetilde{\lambda}(x^k)$ from section 3. In this section, we consider a class of such algorithms with a more general choice of $\widetilde{\lambda}(\cdot)$ and specify the requirements needed for quadratic convergence.

For a current primal-dual iterate $(x,\,\lambda)$, the sSQP method, if applied to an equality-constrained program, defines the next iterate as $(x+\xi,\,\lambda+\eta)$, where $(\xi,\,\eta)$ solves the linear system

$$\frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi + (h'(x))^\top\eta = -\frac{\partial L}{\partial x}(x,\,\lambda), \quad h'(x)\xi - \sigma\eta = -h(x). \tag{4.1}$$

Here, the stabilization parameter σ is typically defined as

$$\sigma = \sigma(x,\,\lambda) = \left\| \left(\frac{\partial L}{\partial x}(x,\,\lambda),\, h(x) \right) \right\|, \tag{4.2}$$

and we will restrict ourselves to this choice. Observe that $\sigma(x,\,\lambda) = 0$ if and only if $(x,\,\lambda)$ solves the Lagrange system (2.2).

Assuming that $\sigma(x,\,\lambda) > 0$, we can express η from the second equation in (4.1) and substitute it in the first equation; this yields

$$\left(\frac{\partial^2 L}{\partial x^2}(x,\,\lambda) + \frac{1}{\sigma(x,\,\lambda)}(h'(x))^\top h'(x) \right)\xi = -\frac{\partial L}{\partial x}(x,\,\lambda) - \frac{1}{\sigma(x,\,\lambda)}(h'(x))^\top h(x).$$

Now, if we fix some mapping $\widetilde{\lambda}: \mathbb{R}^n \to \mathbb{R}^l$, and everywhere substitute λ by $\widetilde{\lambda}(x)$, we obtain the fully primal process with the next iterate defined as $x+\xi$, where ξ solves the linear system

$$\left(\frac{\partial^2 L}{\partial x^2}(x,\,\widetilde{\lambda}(x)) + \frac{1}{\sigma(x,\,\widetilde{\lambda}(x))}(h'(x))^\top h'(x)\right)\xi = -\frac{\partial L}{\partial x}(x,\,\widetilde{\lambda}(x)) - \frac{1}{\sigma(x,\,\widetilde{\lambda}(x))}(h'(x))^\top h(x). \tag{4.3}$$

Algorithm 4.1 Primal stabilized SQP

1: Fix a mapping $\widetilde{\lambda}:\mathbb{R}^n \to \mathbb{R}^l$. Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.
2: If $\sigma(x^k,\,\widetilde{\lambda}(x^k)) = 0$, where $\sigma(\cdot)$ is defined in (4.2), then stop.
3: Compute $\xi^k \in \mathbb{R}^n$ as a solution of the linear system (4.3) with $x = x^k$.
4: Set $x^{k+1} = x^k + \xi^k$, increase k by 1, and go to step 2.

THEOREM 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable near $\bar{x} \in \mathbb{R}^n$, with their second derivatives being Lipschitz-continuous with respect to $\bar{x}$, i.e.,*

$$f''(x) - f''(\bar{x}) = O(\|x - \bar{x}\|), \quad h''(x) - h''(\bar{x}) = O(\|x - \bar{x}\|)$$

as $x \to \bar{x}$. Let $\bar{x}$ be a stationary point of problem (2.1)*. Moreover, with the mapping $\widetilde{\lambda}:\mathbb{R}^n \to \mathbb{R}^l$ used in Algorithm* 4.1*, suppose that there exists a mapping $\pi:\mathbb{R}^n \to \mathbb{R}^l$ satisfying the following assumptions:*

(A1) $\widetilde{\lambda}(x) - \pi(x) = O(\|x - \bar{x}\|)$ *as $x \to \bar{x}$.*

(A2) *There exists a compact subset of $\Lambda(\bar{x})$, consisting of noncritical multipliers, such that $\pi(x)$ belongs to this subset for all x close enough to $\bar{x}$.*

Then, for every $x^0 \in \mathbb{R}^n$ close enough to $\bar{x}$, Algorithm 4.1 *either finitely terminates at $\bar{x}$ or generates an infinite sequence $\{x^k\}$ convergent to $\bar{x}$, and the rate of convergence is quadratic.*

Assumptions (A1) and (A2) imply that $\widetilde{\lambda}(\cdot)$ must be bounded near $\bar{x}$.

Proof. We first show that assumptions (A1) and (A2) imply

$$\|x - \bar{x}\| = O(\sigma(x,\,\widetilde{\lambda}(x))) \tag{4.4}$$

as $x \to \bar{x}$. To this end suppose the contrary, i.e., there exists a sequence $\{x^k\}$ convergent to $\bar{x}$, such that $x^k \neq \bar{x}$ for all k and

$$\frac{\sigma(x^k,\,\widetilde{\lambda}(x^k))}{\|x^k - \bar{x}\|} \to 0 \tag{4.5}$$

as $k \to \infty$. According to (A2), passing to a subsequence, if necessary, we can suppose that $\{\pi(x^k)\}$ converges to some noncritical $\lambda^* \in \Lambda(\bar{x})$ and, due to (A1), $\{\widetilde{\lambda}(x^k))\}$ converges to the same λ^*. Since λ^* is noncritical, the latter implies the error bound

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

as $(x,\,\lambda) \to (\bar{x},\,\lambda^*)$ (see [15, Proposition 1], [16, Proposition 1.43]), and hence,

$$\|x^k - \bar{x}\| = O(\sigma(x^k,\,\widetilde{\lambda}(x^k)))$$

as $k \to \infty$, contradicting (4.5).

In particular, if $\sigma(x,\,\widetilde{\lambda}(x)) = 0$ for x close enough to $\bar{x}$, then from (4.4) it follows that $x = \bar{x}$.

Now, recall the matrix $\mathcal{H}(\bar{x}, \bar{\lambda})$ introduced in section 2 as the symmetric matrix of the quadratic form defined in (2.5), where $\bar{\lambda}$ denotes a multiplier in $\Lambda(\bar{x})$. This matrix is nonsingular if and only if $\bar{\lambda}$ is noncritical. Therefore, we observe that, for x sufficiently close to $\bar{x}$, assumption (A2) implies nonsingularity of $\mathcal{H}(\bar{x}, \pi(x))$ and the uniform boundedness of its inverse. The latter follows by the compactness of the set of noncritical multipliers to which $\pi(x)$ belongs.

For any $x \in \mathbb{R}^n$, by assumptions (A1) and (A2), we further have

$$\begin{aligned} \frac{\partial L}{\partial x}(x, \widetilde{\lambda}(x)) &= \frac{\partial L}{\partial x}(x, \pi(x)) + (h'(x))^\top (\widetilde{\lambda}(x) - \pi(x)) \\ &= \frac{\partial^2 L}{\partial x^2}(\bar{x}, \pi(x))(x - \bar{x}) + (h'(\bar{x}))^\top O(\|x - \bar{x}\|) + O(\|x - \bar{x}\|^2) \end{aligned} \tag{4.6}$$

and

$$\frac{\partial^2 L}{\partial x^2}(x, \widetilde{\lambda}(x)) = \frac{\partial^2 L}{\partial x^2}(x, \pi(x)) + O(\|\widetilde{\lambda}(x) - \pi(x)\|) = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \pi(x)) + O(\|x - \bar{x}\|) \tag{4.7}$$

as $x \to \bar{x}$. Moreover, assuming that $\sigma(x, \widetilde{\lambda}(x)) > 0$, setting

$$\Delta(x) = \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(x) - h'(\bar{x}))^\top,$$

and employing (4.4), we have

$$\begin{aligned} \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(x))^\top h(x) &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(\bar{x}))^\top h(x) + \Delta(x)h(x) \\ &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}\big((h'(\bar{x}))^\top h'(\bar{x})(x - \bar{x}) + (h'(\bar{x}))^\top O(\|x - \bar{x}\|^2)\big) \\ &\quad + \Delta(x)h'(\bar{x})(x - \bar{x}) + \Delta(x)O(\|x - \bar{x}\|^2) \\ &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(\bar{x}))^\top h'(\bar{x})(x - \bar{x}) + (h'(\bar{x}))^\top O(\|x - \bar{x}\|) \\ &\quad + \Delta(x)h'(\bar{x})(x - \bar{x}) + O(\|x - \bar{x}\|^2) \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(x))^\top h'(x) &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(\bar{x}))^\top h'(x) + \Delta(x)h'(x) \\ &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}\big((h'(\bar{x}))^\top h'(\bar{x}) + (h'(\bar{x}))^\top O(\|x - \bar{x}\|)\big) \\ &\quad + \Delta(x)h'(\bar{x}) + \Delta(x)O(\|x - \bar{x}\|) \\ &= \frac{1}{\sigma(x, \widetilde{\lambda}(x))}(h'(\bar{x}))^\top h'(\bar{x}) + (h'(\bar{x}))^\top O(1) \\ &\quad + \Delta(x)h'(\bar{x}) + O(\|x - \bar{x}\|) \end{aligned} \tag{4.9}$$

as $x \to \bar{x}$.

As in section 3.2, let Π denote the orthogonal projector in $\mathbb{R}^n$ onto $\ker h'(\bar{x})$. Again, we set $x_1 = (I - \Pi)x$ and $x_2 = \Pi x$ for any $x \in \mathbb{R}^n$. Using (4.6)–(4.9), we now decompose the left- and right-hand sides of (4.3) by multiplying them with $I-\Pi$ and Π as follows:

$$
\begin{aligned}
(I-\Pi)\left(\frac{\partial^2 L}{\partial x^2}(x,\,\widetilde{\lambda}(x))+\frac{(h'(x))^\top h'(x)}{\sigma(x,\,\widetilde{\lambda}(x))}\right)\xi &= (I-\Pi)\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x))\xi\\
&\quad+\frac{(h'(\bar{x}))^\top h'(\bar{x})\xi_1}{\sigma(x,\,\widetilde{\lambda}(x))}+(h'(\bar{x}))^\top O(1)\xi\\
&\quad+(I-\Pi)\Delta(x)h'(\bar{x})\xi_1\\
&\quad+(I-\Pi)O(\|x-\bar{x}\|)\xi, \qquad (4.10)
\end{aligned}
$$

$$
\begin{aligned}
(I-\Pi)\left(-\frac{\partial L}{\partial x}(x,\,\widetilde{\lambda}(x))-\frac{(h'(x))^\top h(x)}{\sigma(x,\,\widetilde{\lambda}(x))}\right) &= -(I-\Pi)\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x))(x-\bar{x})\\
&\quad-\frac{(h'(\bar{x}))^\top h'(\bar{x})(x_1-\bar{x}_1)}{\sigma(x,\,\widetilde{\lambda}(x))}\\
&\quad-(h'(\bar{x}))^\top O(\|x-\bar{x}\|)\\
&\quad-(I-\Pi)\Delta(x)h'(\bar{x})(x_1-\bar{x}_1)\\
&\quad+(I-\Pi)O(\|x-\bar{x}\|^2), \qquad (4.11)
\end{aligned}
$$

and

$$
\begin{aligned}
\Pi\left(\frac{\partial^2 L}{\partial x^2}(x,\,\widetilde{\lambda}(x))+\frac{(h'(x))^\top h'(x)}{\sigma(x,\,\widetilde{\lambda}(x))}\right)\xi &= \Pi\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x))\xi\\
&\quad+\Pi\Delta(x)h'(\bar{x})\xi_1+\Pi O(\|x-\bar{x}\|)\xi, \qquad (4.12)
\end{aligned}
$$

$$
\begin{aligned}
\Pi\left(-\frac{\partial L}{\partial x}(x,\,\widetilde{\lambda}(x))-\frac{(h'(x))^\top h(x)}{\sigma(x,\,\widetilde{\lambda}(x))}\right) &= -\Pi\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x))(x-\bar{x})\\
&\quad-\Pi\Delta(x)h'(\bar{x})(x_1-\bar{x}_1)\\
&\quad-\Pi O(\|x-\bar{x}\|^2) \qquad (4.13)
\end{aligned}
$$

as $x\to\bar{x}$.

After multiplying both sides of each of the equalities (4.10)–(4.11) by $\sigma(x,\,\widetilde{\lambda}(x))$, and taking into account (4.3), a lengthy reordering of terms yields

$$
\begin{aligned}
\left((h'(\bar{x}))^\top h'(\bar{x})+(I-\Pi)O(\|x-\bar{x}\|)\right)\xi_1 \;=\; &-(h'(\bar{x}))^\top h'(\bar{x})(x_1-\bar{x}_1)\\
&-(I-\Pi)O(\|x-\bar{x}\|)\xi_2\\
&-(I-\Pi)O(\|x-\bar{x}\|^2), \qquad (4.14)
\end{aligned}
$$

where the linear operator $\xi_1\to(h'(\bar{x}))^\top h'(\bar{x})\xi_1:(\ker h'(\bar{x}))^\perp\to(\ker h'(\bar{x}))^\perp$ is nonsingular. The latter implies that for all x close enough to $\bar{x}$, and for any ξ_2, the linear equation (4.14) with respect to ξ_1 has a unique solution, and it satisfies

$$\xi_1=-(x_1-\bar{x}_1)+O(\|x-\bar{x}\|)\xi_2+O(\|x-\bar{x}\|^2) \qquad (4.15)$$

as $x\to\bar{x}$.

Furthermore, again taking into account (4.3), from (4.12)–(4.13) we obtain

$$\begin{aligned}\Pi\left(\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x)) + O(\|x-\bar{x}\|)\right)\xi_2 = &-\Pi\left(\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x)) + O(\|x-\bar{x}\|)\right)\xi_1\\ &-\Pi\Delta(x)h'(\bar{x})\xi_1 - \Pi\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\pi(x))(x-\bar{x})\\ &-\Pi\Delta(x)h'(\bar{x})(x_1-\bar{x}_1)\\ &-\Pi O(\|x-\bar{x}\|^2).\end{aligned}$$

Substituting ξ_1 in the right-hand side of this equality by the expression from (4.15), we get

$$(\mathcal{H}(\bar{x},\,\pi(x)) + \Pi O(\|x-\bar{x}\|))\,\xi_2 = -\mathcal{H}(\bar{x},\,\pi(x))(x_2-\bar{x}_2) - \Pi O(\|x-\bar{x}\|^2). \tag{4.16}$$

In the first part of this proof, we showed that the matrix $\mathcal{H}(\bar{x},\,\pi(x))$ is invertible with a uniformly bounded inverse for x close enough to $\bar{x}$. Therefore, the linear equation (4.16) has a unique solution and it satisfies

$$\xi_2 = -(x_2-\bar{x}_2) + O(\|x-\bar{x}\|^2). \tag{4.17}$$

Substituting this into (4.15), we conclude that

$$\xi_1 = -(x_1-\bar{x}_1) + O(\|x-\bar{x}\|^2) \tag{4.18}$$

as $x \to \bar{x}$. Combining (4.17) and (4.18) we finally obtain

$$x + \xi - \bar{x} = O(\|x-\bar{x}\|^2).$$

This clearly implies the desired assertion. ◻

For one (rather theoretical) possibility of choosing $\widetilde{\lambda}(\cdot)$, suppose that by some chance we are aware of a noncritical multiplier $\bar{\lambda} \in \Lambda(\bar{x})$. Then, we might take $\widetilde{\lambda}(\cdot) = \pi(\cdot) \equiv \bar{\lambda}$.

Ideally, we would be able to show that $\widetilde{\lambda}(\cdot)$ defined according to section 3 satisfies the assumptions of Theorem 4.1, but it seems hardly possible to do this to the full extent. On the one hand, Theorem 3.4 with the assertions (3.25) and (3.24) guarantees assumptions (A1) and (A2), respectively. On the other hand, these assertions are valid only under the conditions in the theorem which, in particular, restrict x to belong to set(s) $K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ with $\bar{\xi}$ related to critical multipliers(s) associated with $\bar{x}$. The union of such sets may not be a full neighborhood of $\bar{x}$.

Therefore, Theorem 4.1 cannot serve for full justification of Algorithm 4.1 with our choice of $\widetilde{\lambda}(\cdot)$, but it somehow explains its numerical success demonstrated in section 6. Observe finally that convergence of the sequence $\{\widetilde{\lambda}(x^k)\}$ is not claimed in Theorem 4.1, and indeed, with our choice of $\widetilde{\lambda}(\cdot)$, convergence of these sequences is often not observed.

5. When there are no critical multipliers. The procedure developed in section 3 is intended to avoid attraction to critical multipliers associated with a stationary point in question. However, one cannot know in advance if such multipliers exist or not, and therefore, it is important to verify that this procedure is at least not harmful in the absence of critical multipliers. This is the subject of the current section.

We start with the case when the constraint qualification (2.3) holds, implying that the Lagrange multiplier associated to a stationary point $\bar{x}$ is unique: $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, where $\bar{\lambda} \in \mathbb{R}^l$ necessarily satisfies $\bar{\lambda} = \bar{\lambda}(\bar{x}) = \widetilde{\lambda}(\bar{x})$ with $\bar{\lambda}(\cdot)$ and $\widetilde{\lambda}(\cdot)$ defined in (3.2)–(3.3). Assume further that this $\bar{\lambda}$ is noncritical. Then, one can easily check that Theorem 4.1 is applicable with the specified $\widetilde{\lambda}(\cdot)$, which in this case is well-defined on a whole neighborhood of $\bar{x}$, and with $\pi(\cdot) \equiv \bar{\lambda}$: both assumptions (A1) and (A2) are satisfied. Therefore, Algorithm 4.1 with this choice of $\widetilde{\lambda}(\cdot)$ locally uniquely defines sequences $\{x^k\}$ converging to $\bar{x}$, and the rate of convergence is quadratic. Moreover, since $\widetilde{\lambda}(\cdot)$ is continuous at $\bar{x}$, it holds that $\{\widetilde{\lambda}(x^k)\}$ converges to $\bar{\lambda}$.

We next consider the more general case when (2.3) may not hold, and hence, $\Lambda(\bar{x})$ need not be a singleton, but all $\bar{\lambda} \in \Lambda(\bar{x})$ are noncritical. We will need the following error bound result for the case of linear constraints, which can be readily derived from [12, Theorem 2] and Hoffman's lemma (see, e.g., [2, Theorem 2.200]).

Proposition 5.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$ and let $h : \mathbb{R}^n \to \mathbb{R}^l$ be affine: $h(x) = Ax + b$, $A \in \mathbb{R}^{l\times n}$, $b \in \mathbb{R}^l$. Let $\bar{x}$ be a stationary point of problem* (2.1)*, and assume that*

$$\langle f''(\bar{x})\xi, \xi\rangle > 0 \quad \text{for all } \xi \in \ker A \setminus \{0\} \tag{5.1}$$

(which is equivalent to saying that the standard second-order sufficient optimality condition holds, as in the case of linear constraints, it does not depend on the choice of a Lagrange multiplier associated to $\bar{x}$).

Then, there exists $M > 0$ such that

$$\|x - \bar{x}\| + \operatorname{dist}(\lambda, \Lambda(\bar{x})) \le M\sigma(x, \lambda) \tag{5.2}$$

for all $x \in \mathbb{R}^n$ close enough to $\bar{x}$ and for all $\lambda \in \mathbb{R}^l$.

The specificity of this result when compared to, say, [16, Proposition 1.43] is that there are no restrictions on λ in (5.2), and in particular λ is not assumed to be close to any given multiplier, or even to the set $\Lambda(\bar{x})$.

Proposition 5.1 allows us to show that our choice of $\widetilde{\lambda}(\cdot)$ satisfies estimate (3.25) from Theorem 3.4 under the sole assumption of 2-regularity of constraints and, in particular, without any criticality or noncriticality assumptions on multipliers.

Corollary 5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$, and let $\bar{x}$ be a stationary point of problem* (2.1)*. Assume that h is* 2*-regular at $\bar{x}$ in a direction $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$.*

Then, there exist $\varepsilon = \varepsilon(\bar{\xi}) > 0$ and $\delta = \delta(\bar{\xi}) > 0$ such that $\widetilde{\lambda}(x)$ is well-defined by (3.2)–(3.3) *for any $x \in K_{\varepsilon,\delta}(\bar{x}, \bar{\xi})$, $\widetilde{\lambda}(\cdot)$ is bounded near $\bar{x}$, and* (3.25) *holds as $x \to \bar{x}$.*

Proof. Since $\bar{x}$ is a stationary point of problem (2.1), we have that problem (3.1) with $x = \bar{x}$ has the unique solution at 0, which is also the unique stationary point of this problem, and the set of associated Lagrange multipliers coincides with $\Lambda(\bar{x})$. Observe that (5.1) holds with f substituted by the objective function of problem (3.1). Therefore, Proposition 5.1 is applicable to problem (3.1) with $x = \bar{x}$, yielding the existence of $M > 0$ such that

$$\|\xi\| + \operatorname{dist}(\lambda, \Lambda(\bar{x})) \le M\bar{\sigma}(\xi, \lambda) \tag{5.3}$$

for all $\xi \in \mathbb{R}^n$ close enough to 0, and all $\lambda \in \mathbb{R}^l$, where $\bar{\sigma} : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}_+$,

$$\bar{\sigma}(\xi, \lambda) = \left\| \left(\frac{\partial L}{\partial x}(\bar{x}, \lambda) + \xi,\, h'(\bar{x})\xi \right) \right\|. \tag{5.4}$$

Let $\varepsilon > 0$, $\delta > 0$, $N > 0$, and $\xi(\cdot,\,\cdot)$ be defined according to both Lemmas 3.2 and 3.3. Then, for every $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$, problem (3.1) has a feasible point $\xi(x,\,-h(x))$, for which from (3.17) we derive the estimate

$$\|\xi(x,\,-h(x))\| = O(\|x - \bar{x}\|)$$

as $x \to \bar{x}$. This implies, in particular, that for the optimal value $v(x)$ of problem (3.1), it holds that

$$\limsup_{x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi}),\, x \to \bar{x}} v(x) \leq 0. \tag{5.5}$$

We next show that for $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$, for the unique solution $\widetilde{\xi}(x)$ of problem (3.1) it holds that $\widetilde{\xi}(x) \to 0$ as $x \to \bar{x}$. Suppose the contrary, i.e., there exist $\gamma > 0$ and a sequence $\{x^k\} \subset K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ convergent to $\bar{x}$ such that

$$\|\widetilde{\xi}(x^k)\| \geq \gamma \tag{5.6}$$

is satisfied for all k.

If $\{\widetilde{\xi}(x^k)\}$ has an accumulation point $\widetilde{\xi}$, then by the constraint in (3.1) we get that $\widetilde{\xi} \in \ker h'(\bar{x})$. Therefore, by the stationarity of $\bar{x}$ in (2.1), it holds that $\langle f'(\bar{x}),\, \widetilde{\xi}\rangle = 0$. Passing to a subsequence, if necessary, we may assume that the entire sequence $\{\widetilde{\xi}(x^k)\}$ converges to $\widetilde{\xi}$. Hence, by (5.6),

$$v(x^k) = \langle f'(x^k),\, \widetilde{\xi}(x^k)\rangle + \frac{1}{2}\|\widetilde{\xi}(x^k)\|^2 \geq \frac{1}{4}\gamma^2$$

follows for all k large enough, which contradicts (5.5).

It remains to consider the case when $\|\widetilde{\xi}(x^k)\| \to \infty$ as $k \to \infty$. But this immediately implies that

$$v(x^k) = \langle f'(x^k),\, \widetilde{\xi}(x^k)\rangle + \frac{1}{2}\|\widetilde{\xi}(x^k)\|^2 \to +\infty$$

as $k \to \infty$, which again contradicts (5.5).

We have thus demonstrated that by further reducing $\varepsilon > 0$, if necessary, one can force $\widetilde{\xi}(x)$ to be arbitrarily close to 0 for all $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$.

Observe now that by the choice of $\varepsilon > 0$, $\delta > 0$, and $N > 0$ according to Lemma 3.3, it follows that $\widetilde{\lambda}(x)$ is well-defined by (3.2)–(3.3) for any $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$, and the estimates (3.21)–(3.22) hold. We next show that $\widetilde{\lambda}(\cdot)$ is bounded near $\bar{x}$.

Indeed, fix any $\bar{\lambda} \in \Lambda(\bar{x})$ (which is nonempty since $\bar{x}$ is assumed stationary in problem (2.1)). Then,

$$f'(x) = f'(\bar{x}) + O(\|x - \bar{x}\|) = -(h'(\bar{x}))^\top\bar{\lambda} + O(\|x - \bar{x}\|) = -(h'(x))^\top\bar{\lambda} + O(\|x - \bar{x}\|)$$

and, hence,

$$h'(x)f'(x) = -H(x)\bar{\lambda} + h'(x)O(\|x - \bar{x}\|) = -H(x)\bar{\lambda} + h'(\bar{x})O(\|x - \bar{x}\|) + O(\|x - \bar{x}\|^2)$$

as $x \to \bar{x}$. Assuming now that $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$, we derive that

$$(H(x))^{-1}h'(x)f'(x) = -\bar{\lambda} + (H(x))^{-1}h'(\bar{x})O(\|x - \bar{x}\|) + (H(x))^{-1}O(\|x - \bar{x}\|^2), \tag{5.7}$$

where the right-hand side can readily be checked to be bounded according to (3.21) and (3.22). At the same time,

$$(H(x))^{-1}h(x) = (H(x))^{-1}h'(\bar{x})(x-\bar{x}) + (H(x))^{-1}O(\|x-\bar{x}\|^2), \tag{5.8}$$

and (3.21) and (3.22) imply that the right-hand side of this equality is bounded as well.

Combining (3.2)–(3.4) with boundedness of the right-hand sides in (5.7)–(5.8), we conclude that $\widetilde{\lambda}(x)$ remains bounded for all $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ close to $\bar{x}$.

With this property at hand, and employing (5.3) and (5.4), for all $x \in K_{\varepsilon,\,\delta}(\bar{x},\,\bar{\xi})$ we finally derive the estimates

$$\begin{aligned}
\|\widetilde{\xi}(x)\| + \operatorname{dist}(\widetilde{\lambda}(x),\,\Lambda(\bar{x})) &\le M\bar{\sigma}(\widetilde{\xi}(x),\,\widetilde{\lambda}(x)) \\
&= M\left\|\left(\frac{\partial L}{\partial x}(\bar{x},\,\widetilde{\lambda}(x)) + \widetilde{\xi}(x),\; h'(\bar{x})\widetilde{\xi}(x)\right)\right\| \\
&\le M\left\|\left(\frac{\partial L}{\partial x}(x,\,\widetilde{\lambda}(x)) + \widetilde{\xi}(x),\; h(x) + h'(x)\widetilde{\xi}(x)\right)\right\| \\
&\quad + M\left(\left\|\frac{\partial L}{\partial x}(x,\,\widetilde{\lambda}(x)) - \frac{\partial L}{\partial x}(\bar{x},\,\widetilde{\lambda}(x))\right\|\right. \\
&\qquad\qquad \left. + \|h(x) + (h'(x) - h'(\bar{x}))\widetilde{\xi}(x)\|\right) \\
&= O(\|x-\bar{x}\|)
\end{aligned}$$

as $x \to \bar{x}$. This yields (3.25). ∎

Assuming now that all Lagrange multipliers are noncritical, Corollary 5.2 implies the existence of a set $U \subset \mathbb{R}^n$ which is starlike with respect to $\bar{x}$, with only excluded directions being those in which h is not 2-regular at $\bar{x}$, and such that our $\widetilde{\lambda}(\cdot)$ is well-defined on U, and satisfies both assumptions (A1) and (A2) of Theorem 4.1 for $x \in U$ with, say, $\pi(x)$ defined as the projection of $\widetilde{\lambda}(x)$ onto $\Lambda(\bar{x})$. As in section 4, U may not be a full neighborhood of $\bar{x}$. Observe, however, that if there exists at least one direction of 2-regularity of h at $\bar{x}$, then the set of such directions is open and dense in the unit sphere, and hence, U is "large" (asymptotically dense at $\bar{x}$).

When h is not 2-regular at $\bar{x}$ in any direction, Algorithm 4.1 with our choice of $\widetilde{\lambda}(\cdot)$ may have no reasonable local convergence properties, even in the absence of critical multipliers.

Example 5.3. The problem

$$\underset{x}{\text{minimize}} \quad x^2 \quad \text{subject to} \quad x^3 = 0$$

has the solution $\bar{x} = 0$. Any $\lambda \in \mathbb{R}$ is an associated noncritical Lagrange multiplier. For any $x \neq 0$, (3.2)–(3.3) yield $\widetilde{\lambda}(x) = -5/(9x)$, and then it can be seen that for x^k close to 0, Algorithm 4.1 defines $x^{k+1} \approx 1.25x^k$. Therefore, convergence of the primal sequence to the solution $\bar{x} = 0$ is not possible, and this is due to the lack of 2-regularity of constraints at this solution.

Observe, however, that the constraints in this example are atypical in the class of degenerate constraints, as the second derivative of h vanishes at $\bar{x}$.

6. Numerical results. The algorithms being tested in this section are stabilized Newton-type methods mentioned in section 1. Specifically, we consider the following:

- sSQP, the stabilized sequential quadratic programming method,
- LM1, the Levenberg–Marquardt method with regularization parameter chosen as the Euclidean residual of the Lagrange system (2.2),

- LM2, the Levenberg–Marquardt method with regularization parameter chosen as the squared Euclidean residual, and
- LP-N, the LP-Newton method.

For comparison, these methods are used with or without the technique for adjusting approximations of Lagrange multipliers as introduced in section 3. As a shortcut, we will refer to this technique by DM (which stands for "dual modification").

The experiments were performed in a MATLAB environment. In particular, the subproblems occurring in each step of sSQP, LM1, and LM2 were solved by standard MATLAB tools for systems of linear equations, while the linear programming problems in LP-N were solved by `cplexlp` from CPLEX Studio IDE 12.7.1, with standard settings except for the following:

- `LPoptions.simplex.tolerances.feasibility` = 1e-9 (default is 1e-6).
- `LPoptions.simplex.tolerances.optimality` = 1e-9 (default is 1e-6).
- `LPoptions.barrier.convergetol` = 1e-10 (default is 1e-8).
- `options.LPoptions.lpmethod` = 2 (dual simplex-method; default is automatic selection of an optimizer).

When the DM technique is not well-defined, i.e., when solving the linear system

$$H(x^k)\widetilde{\lambda} = h(x^k) - h'(x^k)f'(x^k)$$

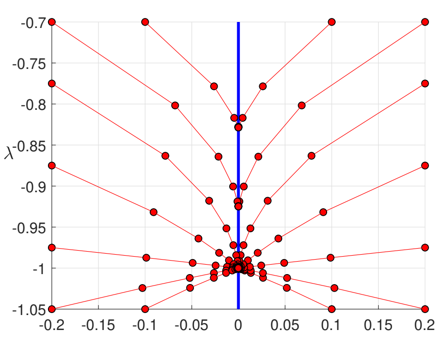
defining $\widetilde{\lambda}(x^k)$ ends up with a failure, we resort to the minimum norm least squares solution of this system.

The algorithms were terminated if the Euclidean residual of the Lagrange system (2.2) was smaller than 1e-12, if solving a subproblem failed, if the Euclidean norm of a computed step was smaller than 1e-14, or if the number of iterations exceeded 200.

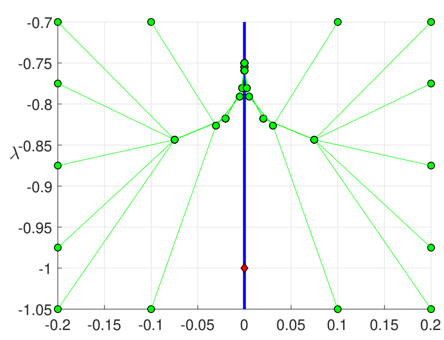
We first show the effect of the procedure for adjusting dual iterates on a very simple example, namely, Example 3.1 with $n = l = 1$ and $A = B = 2$. This means we consider the problem

$$\underset{x}{\text{minimize}} \quad x^2 \quad \text{subject to} \quad x^2 = 0.$$

Figure 6.1 demonstrates the primal-dual solution set $\{0\} \times \mathbb{R}$ as the thick vertical line, the primal-dual solution corresponding to the unique critical Lagrange multiplier $\bar{\lambda} = -1$, and some iterative sequences generated by the sSQP, without the technique for adjusting the dual iterate (Figure 6.1(a)), and with this technique applied at each iteration (Figure 6.1(b)), initialized at the same points. It is evident that adjusting the dual iterate prevents convergence to the critical multiplier. In fact, all sequences

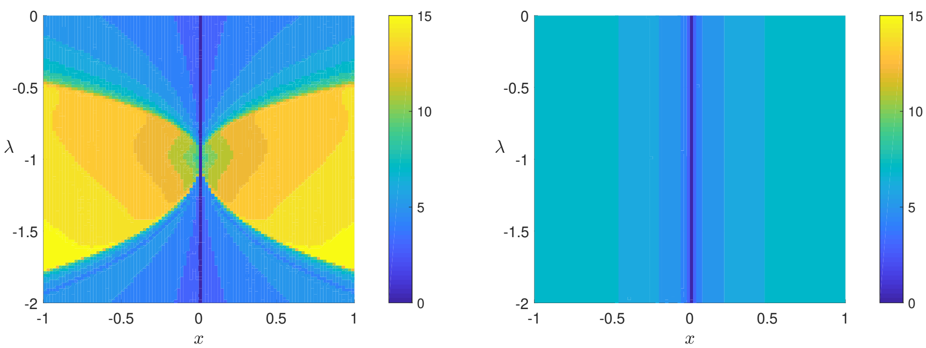


(a) sSQP



(b) sSQP with DM technique

FIG. 6.1. *Iterative sequences.*



(a) sSQP (b) sSQP with DM technique

FIG. 6.2. *Iteration count.*

in Figure 6.1(b) converge to the same limit $(0, -3/4)$, which agrees with (3.15): in this particular case, $\widetilde{\lambda}(x) = -3/4$ for all x. Different values of dual starting points are used in Figure 6.1(b) only to emphasize that they are actually irrelevant: for a given primal starting point, they all give rise to the same next iterate.

Furthermore, Figure 6.2 shows how the iteration count of the sSQP until successful termination depends on starting points, without the technique for adjusting the dual iterate (Figure 6.2(a)) and with this technique applied at each iteration (Figure 6.2(b)). The positive effect of the latter is evident. Moreover, Figure 6.2(a) shows very clearly the domain of attraction to the critical multiplier (the set with bright colors), and it fully agrees with convergence results obtained in [13]. At the same time, there is no such domain in Figure 6.2(b): it is eliminated by the dual adjusting technique. Figure 6.2(b) also shows that values of dual starting points do not affect the performance in any way.

We now proceed with systematic testing for a set of randomly generated optimization problems with quadratic objective functions and quadratic equality constraints, employing the generator described in [14]. The entries of all randomly generated arrays take values in $[-10, 10]$. For each triple (n, l, r) of nonnegative integers, we generated 100 problems with n variables and l constraints, such that $\bar{x} = 0$ is a stationary point of each problem, and rank $h'(\bar{x}) = r$ (with $r < l$).

Observe further that problems generated this way are nonconvex, can be unbounded below, and may have stationary points other than the one of interest, namely other than $\bar{x} = 0$.

Therefore, convergence to other stationary points is certainly a possibility, as well as failure of convergence. Convergence to $\bar{x} = 0$ was declared when at termination the Euclidean norm of the last iterate was not larger than 1e-5. The statistics on criticality and convergence rate estimates, reported below, are concerned with such cases only; other runs were disregarded. Interestingly, if the dual modification technique is applied, the primal sequences usually converge to 0 more often or at least as frequently as without dual modification; see Figure 6.3.

Since we are mostly interested in the behavior of the algorithms near critical multipliers, we have selected for our analysis only those problems in which critical multipliers do exist. The latter was verified by running the pure SQP method for every generated problem, which, as mentioned in section 1, has a strong tendency to produce dual sequences convergent to critical multipliers when they exist. In cases when SQP failed or did not demonstrate convergence to $\bar{x} = 0$ (see above), or when the smallest absolute value of the eigenvalues of $\mathcal{H}(x^k, \lambda^k)$ at termination was greater

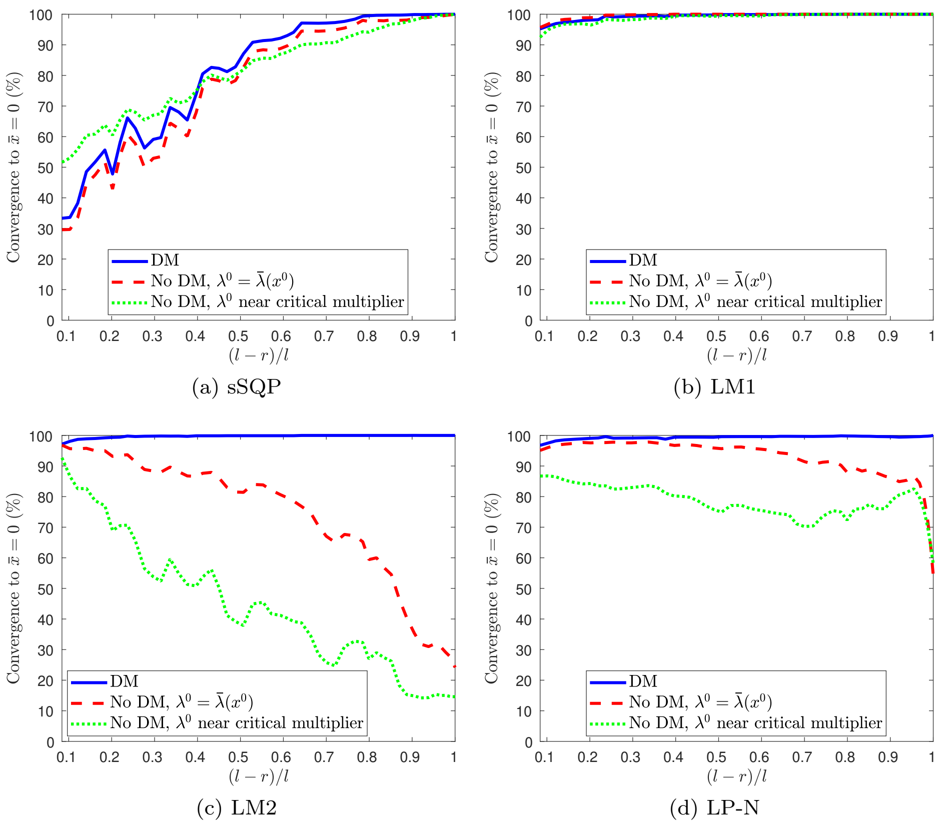


FIG. 6.3. *Convergence to $\bar{x} = 0$.*

than 1e-5, the problem was disregarded, and the next one was tried, until 100 problems (presumably) possessing critical multipliers were generated. Moreover, the dual points $\bar{\lambda}$ at termination were stored for these problems as approximate critical multipliers associated to $\bar{x} = 0$. We emphasize that when performing this procedure, we have encountered very few problem instances for which the existence of critical multipliers has not been detected, and therefore, this procedure did not affect much the set of test problems. However, it gave us a possibility to initialize algorithms near approximate critical multipliers $\bar{\lambda}$ which were not known otherwise.

The same test for approximate criticality of the dual point at termination (employing the eigenvalues of $\mathcal{H}(x^k, \lambda^k)$ as described above) was used for all algorithms involved in the experiments.

For each problem generated as described above, we ran the algorithms from 10 starting points (x^0, λ^0) with random x^0 satisfying $\|x^0\| = 0.1$. Therefore, we performed 1000 runs in total for each triple (n, l, r). As for λ^0, we used two different choices:

- $\lambda^0 = \bar{\lambda}(x^0)$, where $\bar{\lambda}(\cdot)$ is defined according to (3.3). This is quite a typical choice of a dual approximation from a known primal one, as it corresponds to the least squares solution of the first equation in (2.2) with $x = x^0$.
- λ^0 satisfying $\|\lambda^0 - \bar{\lambda}\| = 0.1$ is taken randomly.

Our third option for each of the algorithms is to use the technique for adjusting the dual iterate at every iteration, i.e., every iterate (x^k, λ^k) obtained by an algorithm was replaced by $(x^k, \widetilde{\lambda}(x^k))$. Observe that in this case, dual starting points are not needed.

The data we report on is as follows. For each triple (n, l, r), we count the number C_x of detected cases of convergence to $\bar{x} = 0$. Out of the latter, we compute the number C_λ of cases when convergence to a critical multiplier was detected, the number $E_x^{1.1}$ of cases when the experimental convergence rate E was greater than 1.1, and the number $E_x^{1.5}$ of cases when E was greater than 1.5. The experimental convergence rate is defined as

$$E = \max\left\{\frac{\log\|x^k\|}{\log\|x^{k-1}\|}, \frac{\log\|x^{k-1}\|}{\log\|x^{k-2}\|}\right\},$$

where k is the iteration number at termination. For a related definition of an experimental convergence rate see [10]. In order to better cope with numerical problems close to the stationary point $\bar{x} = 0$, the rate E now relies on values for (x^k, x^{k-1}) and (x^{k-1}, x^{k-2}). If E is greater than 1, this gives a sign that the primal convergence rate is superlinear (of approximate order E). Therefore, one might expect $E_x^{1.1}$ to be close to $C_x - C_\lambda$, and, indeed, this often can be observed. However, we note that E is only an approximation of the real convergence rate.

We report on the results obtained for triples (n, l, r) with $n = 25$, $l \in \{1, \ldots, n-1\}$, and $r \in \{0, \ldots, l-1\}$. The picture for the other values of n which we tried looks quite similar.

The results are shown in Figures 6.4–6.6 in the following form. For each of the two specified options for defining the dual starting point, and for DM used at each step, we present the average values of C_λ (Figure 6.4), $E_x^{1.1}$ (Figure 6.5), and $E_x^{1.5}$ (Figure 6.6)

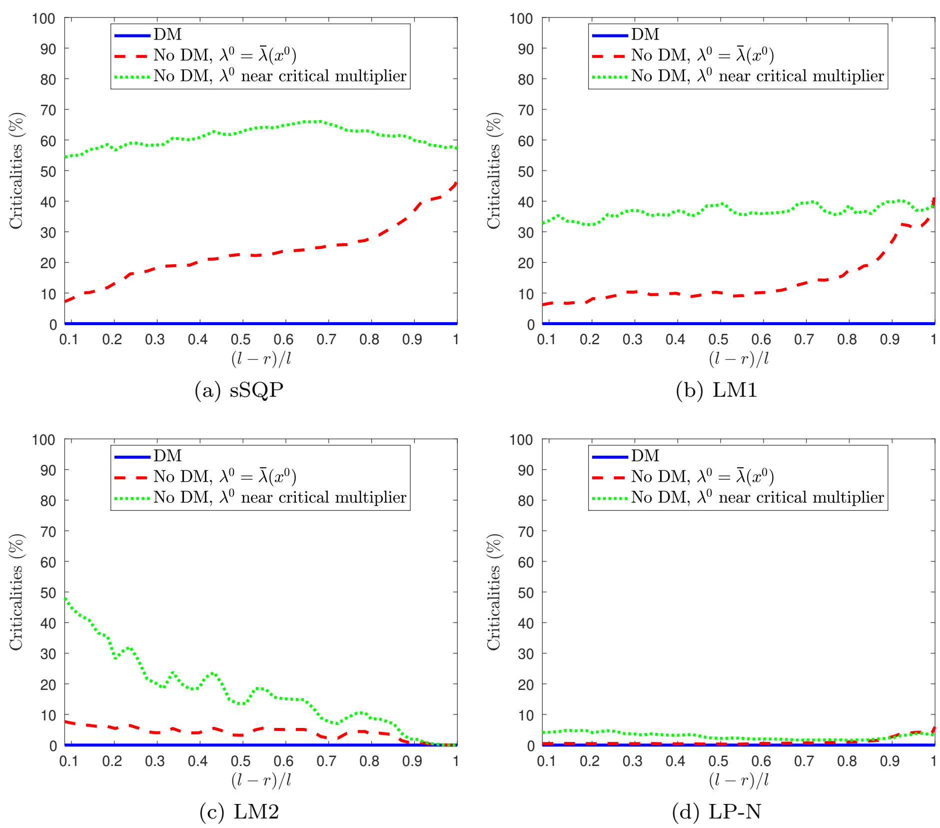


FIG. 6.4. *Cases of convergence to critical multipliers.*

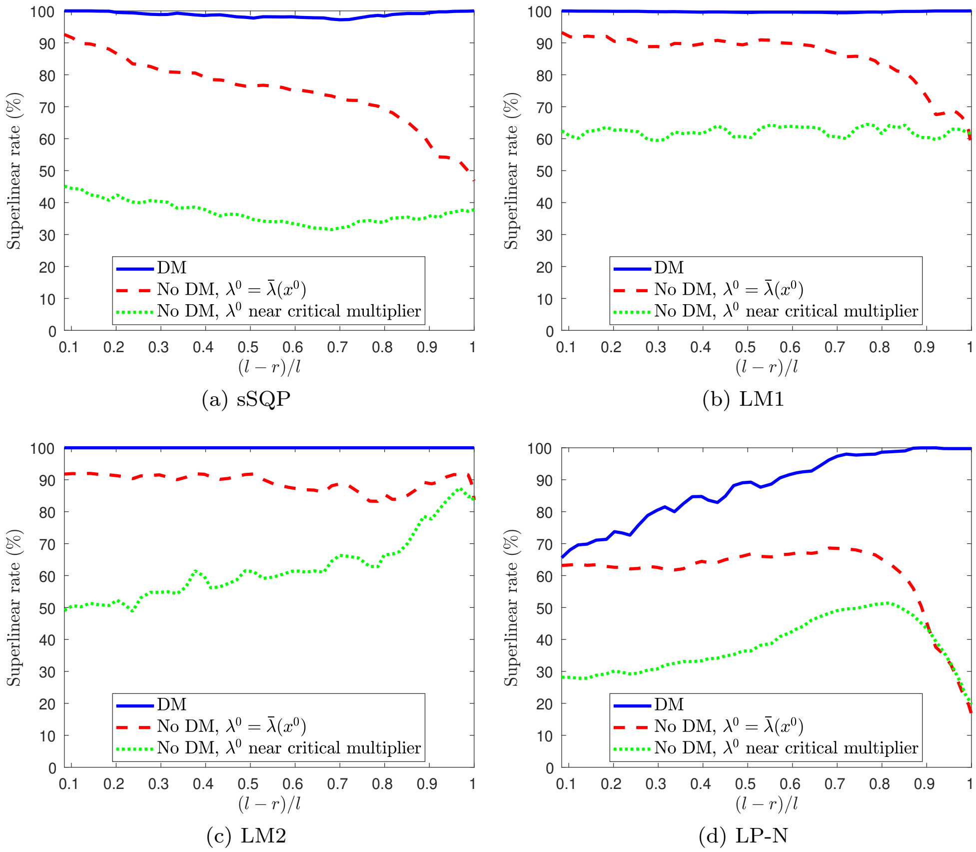
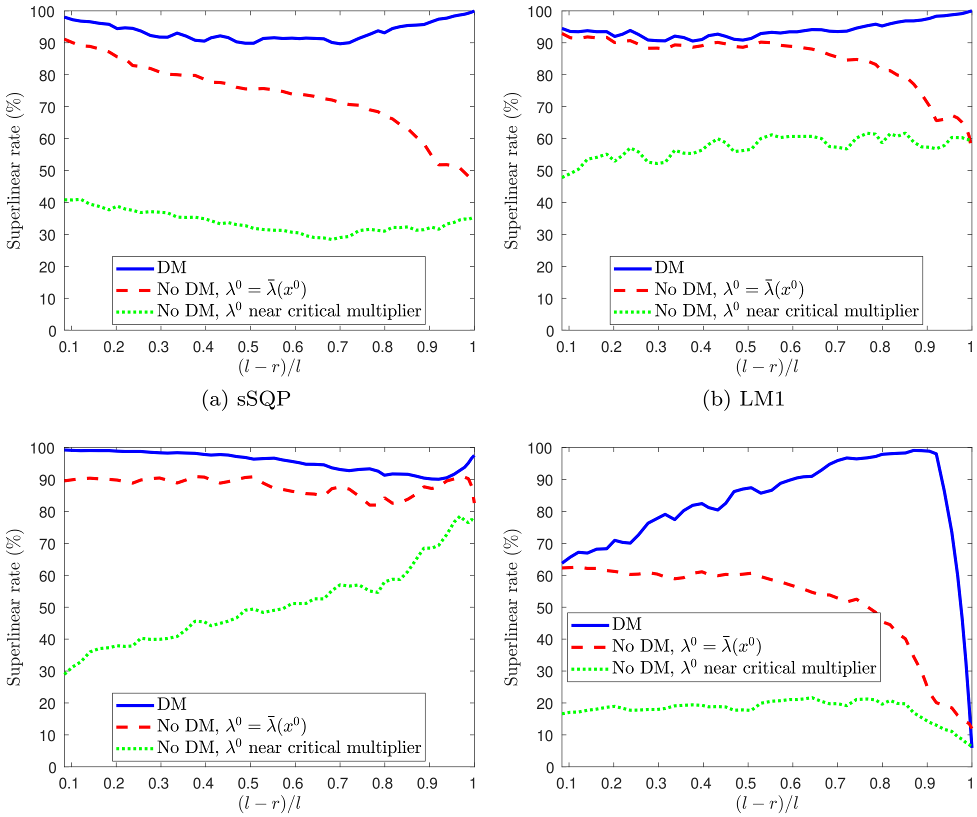


FIG. 6.5. *Superlinear rate: $E > 1.1$.*

as functions of the relative rank deficiency $rrd = (l-r)/l$. These graphs are based on all runs converging to $\bar{x} = 0$. Since C_λ, $E_x^{1.1}$, and $E_x^{1.5}$ show certain oscillations we used a smoothing procedure (some kind of a moving average) in order to improve the visibility of the principal dependency of these values on the relative rank deficiency rrd. This procedure is described in detail in the following paragraph. First, let us make a remark on the DM-curve in picture (d) of Figure 6.6. The steep descent of this curve is due to the fact that the solver `cplexlp` shows numerical problems if the relative rank deficiency $rrd = (l-r)/l$ is close to 1, i.e., if the rank of $h'(\bar{x})$ is close or equal to 0.

For a triple (n, l, r), we initially have some values of C_λ, $E_x^{1.1}$, and $E_x^{1.5}$. Then, we calculate the values of rrd, and sort them in the increasing order, keeping the correspondence with the values of C_λ, $E_x^{1.1}$, and $E_x^{1.5}$:

$$\begin{gathered}(rrd_1,\ C_{\lambda,1},\ E_{x,1}^{1.1},\ E_{x,1}^{1.5})\\ \vdots\\ (rrd_i,\ C_{\lambda,i},\ E_{x,i}^{1.1},\ E_{x,i}^{1.5})\\ \vdots\\ (rrd_N,\ C_{\lambda,N},\ E_{x,N}^{1.1},\ E_{x,N}^{1.5}),\end{gathered}$$



(a) sSQP (b) LM1

(c) LM2 (d) LP-N

FIG. 6.6. *Superlinear rate:* $E > 1.5$.

with $rrd_i \leq rrd_{i+1}$ for all $i = 1, \ldots, N-1$, where N is the total number of different triples (n, l, r) used (with $n = 25$ fixed). Now, for each $i = 1, \ldots, N-6$, the value of the "smoothed" relative rank deficiency $srrd_i$ is calculated as the average over ith and six consecutive values of rrd in the above sorting, and the same is done to calculate the values of "smoothed" sC_λ, $sE_x^{1.1}$, and $sE_x^{1.5}$:

$$srrd_i = \left(\sum_{j=i}^{i+6} rrd_j\right)/7,$$

$$sC_{\lambda,\,i} = \left(\sum_{j=i}^{i+6} C_{\lambda,\,j}\right)/7, \quad sE_{x,\,i}^{1.1} = \left(\sum_{j=i}^{i+6} E_{x,\,j}^{1.1}\right)/7, \quad sE_{x,\,i}^{1.5} = \left(\sum_{j=i}^{i+6} E_{x,\,j}^{1.5}\right)/7.$$

Since this smoothing procedure preserves the order in the new set of "smoothed" relative rank deficiency values, we can regard these new data as "smoothed" representation of C_λ, $E_x^{1.1}$, and $E_x^{1.5}$. For generating the graphs in Figure 6.3 with the C_x values, we have used the same smoothing procedure.

The numerical results reported above support quite convincingly the theoretical analysis in section 3. Specifically, the use of the procedure for adjusting dual iterates totally suppresses the tendency of convergence to critical multipliers (see Figure 6.4) and has an evident positive effect on the convergence rate.

7. Conclusions.

We have proposed the technique of adjusting dual iterates of primal-dual optimization algorithms. Theoretical and numerical results presented in this paper put in evidence that this technique often allows one to avoid attraction to critical Lagrange multipliers and, accordingly, to accelerate convergence of Newton-type methods. Further developments will include incorporating this technique into some globally convergent algorithms, as well as extensions to problems involving inequality constraints.

Acknowledgments. The authors would like to thank the anonymous referees for their helpful comments and, in particular, for a question that led to the results in section 5 added in the revised version.

REFERENCES

[1] A. V. Arutyunov, *Optimality Conditions: Abnormal and Degenerate Problems*, Kluwer, Dordrecht, The Netherlands, 2000, https://dx.doi.org/10.1007/978-94-015-9438-7.

[2] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000, https://doi.org/10.1007/978-1-4612-1394-9.

[3] Y. Cui, D. Sun, and K.-C. Toh, *On the Asymptotic Superlinear Convergence of the Augmented Lagrangian Method for Semidefinite Programming with Multiple Solutions*, preprint, https://arxiv.org/abs/1610.00875v1 [math.OC], 2016.

[4] H. Do, B. S. Mordukhovich, and M. E. Sarabi, *Criticality of Lagrange multipliers in extended nonlinear optimization* Optimization, (2020), https://doi.org/10.1080/02331934.2020.1723585.

[5] F. Facchinei, A. Fischer, and M. Herrich, *An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions*, Math. Program., 146 (2014), pp. 1–36, https://doi.org/10.1007/s10107-013-0676-6.

[6] J.-Y. Fan and Y.-X. Yuan, *On the quadratic convergence of the Levenberg-Marquardt method*, Computing, 74 (2005), pp. 23–39, https://doi.org/10.1007/s00607-004-0083-1.

[7] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Math. Oper. Res., 24 (1999), pp. 699–727, https://doi.org/10.1287/moor.24.3.699.

[8] A. Fischer, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124, https://doi.org/10.1007/s10107-002-0364-4.

[9] A. Fischer, A. F. Izmailov, and M. V. Solodov, *Local attractors of Newton-type methods for constrained equations and complementarity problems with nonisolated solutions*, J. Optim. Theory Appl., 180 (2019), pp. 140–169, https://doi.org/10.1007/s10957-018-1297-2.

[10] A. Fischer, P. K. Shukla, and M. Wang, *On the inexactness level of robust Levenberg–Marquardt methods*, Optimization, 59 (2010), pp. 273–287, https://doi.org/10.1080/02331930801951256.

[11] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 1, Springer, New York, 1984, https://doi.org/10.1007/978-1-4612-5034-0.

[12] W. W. Hager and M. S. Gowda, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192, https://doi.org/10.1007/s101070050051.

[13] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov, *Critical solutions of nonlinear equations: Local attraction for Newton-type methods*, Math. Program., 167 (2018), pp. 355–379, https://doi.org/10.1007/s10107-017-1128-5.

[14] A. F. Izmailov and M. V. Solodov, *On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions*, Math. Program., 117 (2009), pp. 271–304, https://doi.org/10.1007/s10107-007-0158-9.

[15] A. F. Izmailov and M. V. Solodov, *Stabilized SQP revisited*, Math. Program., 133 (2012), pp. 93–120, https://doi.org/10.1007/s10107-010-0413-3.

[16] A. F. Izmailov and M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Ser. Oper. Res. Financ. Eng., Springer, Cham, Switzerland, 2014, https://doi.org/10.1007/978-3-319-04247-3.

[17] A. F. Izmailov and M. V. Solodov, *Critical Lagrange multipliers: What we currently know about them, how they spoil our lives, and what we can do about it*, TOP, 23 (2015), pp. 1–26, https://doi.org/10.1007/s11750-015-0372-1.

[18] A. F. Izmailov and E. I. Uskov, *Attraction of Newton method to critical Lagrange multipliers: Fully quadratic case*, Math. Program., 152 (2015), pp. 33–73, https://doi.org/10.1007/s10107-014-0777-x.

[19] B. S. Mordukhovich and M. E. Sarabi, *Critical multipliers in variational systems via second-order generalized differentiation*, Math. Program., 169 (2018), pp. 605–648, https://doi.org/10.1007/s10107-017-1155-2.

[20] B. S. Mordukhovich and M. E. Sarabi, *Criticality of Lagrange multipliers in variational systems*, SIAM J. Optim., 29 (2019), pp. 1524–1557, https://doi.org/10.1137/18M1206862.

[21] S. J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275, https://doi.org/10.1023/A:1018665102534.

[22] *S. J. Wright*, *An algorithm for degenerate nonlinear programming with rapid local convergence*, SIAM J. Optim., 15 (2005), pp. 673–696, https://doi.org/10.1137/030601235.

[23] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg-Marquardt method*, in Topics in Numerical Analysis, G. Alefeld and X. Chen, eds., Springer, Vienna, 2001, pp. 239–249, https://doi.org/10.1007/978-3-7091-6217-0_18.

[24] T. Zhang and L. Zhang, *Critical multipliers in semidefinite programming*, Asia-Pac. J. Oper. Res., (2020), https://doi.org/10.1142/S0217595920400126.