ACCELERATING CONVERGENCE OF A GLOBALIZED SEQUENTIAL QUADRATIC PROGRAMMING METHOD TO CRITICAL LAGRANGE MULTIPLIERS

A.F. Izmailov[1]

May 9, 2021

ABSTRACT

This paper concerns the issue of asymptotic acceptance of the true Hessian and the full step by the sequential quadratic programming algorithm for equality-constrained optimization problems. In order to enforce global convergence, the algorithm is equipped with a standard Armijo linesearch procedure for a nonsmooth exact penalty function. The specificity of considerations here is that the standard assumptions for local superlinear convergence of the method may be violated. The analysis focuses on the case when there exist critical Lagrange multipliers, and does not require regularity assumptions on the constraints or satisfaction of second-order sufficient optimality conditions. The results provide a basis for application of known acceleration techniques, such as extrapolation, and allow the formulation of algorithms that can outperform the standard SQP with BFGS approximations of the Hessian on problems with degenerate constraints. This claim is confirmed by some numerical experiments.

Key words: equality-constrained optimization; Lagrange optimality system; critical Lagrange multiplier; 2-regularity; Newton-type methods; sequential quadratic programming; linesearch globalization of convergence; merit function; nonsmooth exact penalty function; true Hessian; unit stepsize; extrapolation.
AMS subject classifications: 49M05, 49M15, 65K05, 65K10, 90C30, 90C55.

[1] Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, IO Department, Leninskiye Gory, 119991 Moscow, Russia.

1 Introduction

When standard constraint qualifications are violated at a stationary point of an optimization problem, and in particular, when the set of associated Lagrange multipliers is not a singleton, this set may naturally contain the so-called critical instances. It is now well-known that critical Lagrange multipliers strongly attract dual iterative sequences generated by various primal-dual optimization algorithms, and this is the main reason for the lack of superlinear convergence rate in the presence of constraint degeneracy; see [11, Section 7.1], [13], and references therein. Even the algorithms equipped with dual stabilization mechanisms intended for suppressing the attraction effect locally, still typically have large domains of convergence to critical multipliers; see, e.g., the analysis in [8], covering the stabilized sequential quadratic programming method (for the latter, see [17, 19], as well as [12], and [11, Section 7.2.2]). Therefore, the phenomenon of attraction to critical multipliers remains an issue, especially when globalization of convergence is concerned.

One possible approach to the specified issue is to develop further techniques for decreasing the chances of convergence to critical multipliers, e.g., as it was done very recently in [3]. Another approach, pursued here, can be summarized as follows: once it is difficult to avoid convergence to critical multipliers, the understanding of the special pattern of this convergence can be employed in order to accelerate it. To explain the point, consider first the basic Newton method for generic equations.

It is well known that when initialized close to a nonsingular solution of a nonlinear equation

$$\Phi(u) = 0 \tag{1.1}$$

with a smooth mapping $\Phi : \mathbb{R}^p \to \mathbb{R}^p$, the Newton method converges to this solution superlinearly (see, e.g., [11, Section 2.1.1]). Here a solution $\bar{u}$ of (1.1) is refereed to as (non)singular if $\Phi'(\bar{u})$ is a (non)singular matrix. Moreover, convergence can be globalized by the standard Armijo linesearch procedure for the (squared) residual of this equation, and this procedure accepts the unit stepsize near nonsingular solutions, thus ensuring superlinear asymptotic convergence rate (see [11, Section 5.1.1]).

This paper mainly concerns the cases of singular and possibly even nonisolated solutions. The local linear convergence of the Newton method from a set of starting points that is starlike and asymptotically dense with respect to a singular solution can still be guaranteed under reasonable assumptions. These assumptions do not imply that the solution in question is isolated, and the pattern of convergence is rather special [5, 6, 7]. Moreover, employing this pattern summarized in Proposition 2.1 below, convergence can be accelerated by extrapolation or overrelaxation techniques [5, 7]. However, when these methods are combined with linesearch globalization, ultimate acceptance of the unit stepsize is not at all automatic/evident, and this becomes the key issue for potential acceleration, and even for ensuring linear convergence of the basic Newton method.

This issue has been studied in detail in [4], where ultimate acceptance of the unit stepsize has been established under the same assumptions as those needed for local linear convergence of the basic Newton method, and for all starting points in a set that is starlike and asymptotically dense with respect to the solution in question.

This work deals with a special system of nonlinear equations arising as first-order necessary optimality conditions for an equality-constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{1.2}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint mapping $h : \mathbb{R}^n \to \mathbb{R}^l$ are sufficiently smooth. The analysis from [4] is certainly applicable to this special system. Observe, however, that the analysis in [4] is developed for the residual of the equation serving as a merit function, which is not fully adequate in optimization context: other merit functions are usually preferred, reflecting the intention to find a solution of the optimization problem rather than just any stationary point of it; see the details in Section 2. Moreover, the algorithms with linesearch for those other merit functions often perform better in practice than those employing the residual of the first-order optimality system. One typical choice of a merit function in this context is a nonsmooth exact penalty function, and the issues in question for such merit functions become more involved.

Our notation is fairly standard. The Euclidian (l_2), l_1, and l_∞ norms will be denoted by $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, respectively. For a function $\varphi : \mathbb{R}^n \to \mathbb{R}$, the notation $\varphi'(x;\, \xi)$ will be used for the standard directional derivative of φ at $x \in \mathbb{R}^n$ in a direction $\xi \in \mathbb{R}^n$.

The structure of the paper is as follows. After providing some necessary preliminaries in Section 2, we proceed in Section 3 with the main part of this work, namely, establishing asymptotic acceptance of the true Hessian and of the unit stepsize by a linesearch Newton method applied to the Lagrange optimality system of problem (1.2). In Section 4, we provide some numerical results demonstrating, in particular, the effect of supplying the method in question with extrapolation technique, justification of which in this context relies on the main result of Section 3. Finally, Section 5 provides some concluding remarks, and specifies directions for future research.

2 Preliminaries

Some more notation and terminology are now in order. For a given $\bar{u} \in \mathbb{R}^p$, let Π stand for the orthogonal projector onto $(\operatorname{im} \Phi'(\bar{u}))^\perp$ in $\mathbb{R}^p$. For every $v \in \mathbb{R}^p$, let the linear operator $\mathcal{B}(v) : \ker \Phi'(\bar{u}) \to (\operatorname{im} \Phi'(\bar{u}))^\perp$ be defined as the restriction of $\Pi\Phi''(\bar{u})[v]$ to $\ker \Phi'(\bar{u})$. The mapping Φ is said to be 2-regular at $\bar{u}$ in a direction v if $\mathcal{B}(v)$ is nonsingular. Furthermore, assuming that $\|v\| = 1$, for any scalars $\varepsilon > 0$ and $\delta > 0$

define the set

$$K_{\varepsilon,\,\delta}(\bar{u},\,v)=\left\{u\in\mathbb{R}^p\setminus\{\bar{u}\}\,:\,\|u-\bar{u}\|\le\varepsilon,\ \left\|\frac{u-\bar{u}}{\|u-\bar{u}\|}-v\right\|\le\delta.\right\}$$

The following is [4, Proposition 1] characterizing the local convergence properties of the basic Newton method for a generic equation, when initialized near $\bar{u}+\ker\Phi'(\bar{u})$. This result can be considered as a version of [6, Lemma 5.1] and [5, Lemma 2.2].

Proposition 2.1 *Let $\Phi:\mathbb{R}^p\to\mathbb{R}^p$ be twice differentiable near $\bar{u}\in\mathbb{R}^p$, with its second derivative Lipschitz-continuous with respect to $\bar{u}$, that is,*

$$\Phi''(u)-\Phi''(\bar{u})=O(\|u-\bar{u}\|)$$

as $u\to\bar{u}$. Let $\bar{u}$ be a solution of equation (1.1)*, and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v}\in\ker\Phi'(\bar{u})$, $\|\bar{v}\|=1$.*

Then, for every $\bar{\varepsilon}>0$ and $\bar{\delta}>0$, there exist $\varepsilon=\varepsilon(\bar{v})>0$ and $\delta=\delta(\bar{v})>0$ such that, for every starting point $u^0\in K_{\varepsilon,\,\delta}(\bar{u},\,\bar{v})$, a unique sequence $\{u^k\}\subset\mathbb{R}^p$ exists such that $v^k=u^{k+1}-u^k$ solves

$$\Phi(u^k)+\Phi'(u^k)v=0 \tag{2.1}$$

for each k, and this sequence possesses the following properties: $\{u^k\}\subset K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\,\bar{v})$, $\{u^k\}$ converges to $\bar{u}$,

$$\lim_{k\to\infty}\frac{\|u^{k+1}-\bar{u}\|}{\|u^k-\bar{u}\|}=\frac{1}{2}, \tag{2.2}$$

and the sequence $\{(u^k-\bar{u})/\|u^k-\bar{u}\|\}$ converges to some $v\in\ker\Phi'(\bar{u})$.

Proposition (2.1) specifies the convergence pattern of the basic Newton method to a solution satisfying 2-regularity in some direction in the null space. This pattern, and in particular, relation (2.2), suggest the idea of duplicating the Newton step in order to obtain much better approximation of the solution. This is the essence of the simplest extrapolation procedure intended to accelerate the process, to be formally introduced and tested in Section 4.

We now get back to optimization problems. Let $L:\mathbb{R}^n\times\mathbb{R}^l\to\mathbb{R}$ be the Lagrangian of problem (1.2), i.e.,

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

Then the primal-dual first-order optimality conditions for problem (1.2), characterizing its stationary points and associated Lagrange multipliers, are given by (1.1), where one should take $p=n+l$, $\Phi:\mathbb{R}^n\times\mathbb{R}^l\to\mathbb{R}^n\times\mathbb{R}^l$,

$$\Phi(u)=\left(\frac{\partial L}{\partial x}(x,\,\lambda),\,h(x)\right), \tag{2.3}$$

with $u = (x, \lambda)$. If $\bar{x}$ is a local solution of (1.2), satisfying the constraints regularity condition

$$\operatorname{rank} h'(\bar{x}) = l, \tag{2.4}$$

then there exists the unique $\bar{\lambda} \in \mathbb{R}^l$ such that $\bar{u} = (\bar{x}, \bar{\lambda})$ satisfies (1.1). However, here we are mostly interested in the case when (2.4) does not hold, but $\bar{x}$ is stationary for (1.2) with some (nonunique) associated Lagrange multiplier $\bar{\lambda}$.

For Φ defined in (2.3), the iteration system (2.1) of the Newton method for (1.1) at the current iterate $u^k = (x^k, \lambda^k)$ takes the form

$$\frac{\partial L}{\partial x}(x^k, \lambda^k) + H_k\xi + (h'(x^k))^\top \eta = 0, \quad h(x^k) + h'(x^k)\xi = 0, \tag{2.5}$$

with respect to $v = (\xi, \eta)$. Here we allow for choices of an $n \times n$ symmetric matrix H_k different from the basic choice

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \tag{2.6}$$

corresponding exactly to (2.1). This method can be seen as the sequential quadratic programming (SQP) algorithm, since (2.5) is the Lagrange optimality system for the corresponding quadratic programming subproblem; see, e.g., [11, Section 4.2].

The need to use something other than (2.6) may arise from globalization techniques, and especially those natural in optimization context, i.e., developed with intention for finding solutions of (1.2) rather than its stationary points.

Specifically, the well-established globalization techniques for Newton-type methods applied to constrained optimization problems consist of linesearch for l_1-penalty function $\varphi_c : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi_c(x) = f(x) + c\|h(x)\|_1, \tag{2.7}$$

where $c > 0$ is a penalty parameter (see [1, Section 17], [18, Section 18.4], [11, Section 6.2]). If $v^k = (\xi^k, \eta^k)$ solves (2.5), then it can be seen (e.g., [11, Lemma 6.8]) that the directional derivative of φ_c at x^k in the direction ξ^k satisfies

$$\varphi_c'(x^k; \xi^k) = \langle f'(x^k), \xi^k\rangle - c\|h(x^k)\|_1 \leq -\langle H_k\xi^k, \xi^k\rangle + (\|\lambda^k + \eta^k\|_\infty - c)\|h(x^k)\|_1. \tag{2.8}$$

Therefore, if, say, H_k is positive definite, $\xi^k \neq 0$, and $c = c_k$ is chosen satisfying $c \geq \|\lambda^k + \eta^k\|_\infty$, then

$$\varphi_c'(x^k; \xi^k) < 0, \tag{2.9}$$

and in particular, ξ^k is a direction of descent for φ_c at x^k, and hence, linesearch can be performed in this direction.

The problem, however, is that H_k defined according to (2.6) cannot be expected to be positive definite, even for (x^k, λ^k) close to a primal-dual solution, even satisfying all the reasonable assumptions need for local superlinear convergence of the full-step method. Therefore, when (2.9) is violated, H_k defined in (2.6) should be replaced by some other matrix ensuring (2.9) (or more precisely, ensuring the "quantified" property like (2.12) below).

Once (2.9) is satisfied, it can be shown in a standard way that for any fixed $\sigma \in (0,\, 1)$, the Armijo inequality

$$\varphi_c(x^k + \alpha\xi^k) \leq \varphi_c(x^k) + \sigma\alpha\varphi'_c(x^k;\, \xi^k) \tag{2.10}$$

holds for all $\alpha > 0$ small enough. Therefore, $\alpha = \alpha_k$ satisfying (2.10) can be obtained from the starting trial value $\alpha = 1$ after a finite number of backtracking steps. Once this is done, the next primal iterate is defined as $x^{k+1} = x^k + \alpha_k\xi^k$, thus completing the iteration.

For recent discussions of this kind of algorithms, including their global convergence properties, and rate of convergence properties under "standard" assumptions, see [14] and references therein. Recall, however, that here we are interested in what happens not under "standard" assumptions but rather in cases of convergence to singular primal-dual solutions. In these circumstances, we observe again that for the acceleration techniques mentioned in Section 1 to take effect, we need the globalized algorithm to follow the convergence pattern of the pure Newton method. For the globalization technique considered in this section, the latter means not only the asymptotic acceptance of the unit stepsize, but also the asymptotic absence of the need to modify H_k defined according to (2.6). These are the issues to be addressed in the rest of this work.

In the development below, we will refer to the following model algorithm implementing the considerations above but intended for local analysis only. "True" implementations are supposed to invoke appropriate modifications of H_k when the current algorithm stops with failure; see Algorithm 4.1 below. Another issue that will not be addressed in this paper is possible infeasibility of subproblems. This is a general issue concerned with practical implementations of SQP methods, and there exist known tools for tackling it; see, e.g., [11, Section 6.2].

Algorithm 2.1 Choose $u^0 = (x^0,\, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ and set $k = 0$. Fix the parameters $\bar{c} > 0$, $\rho > 0$ and $\sigma,\, \theta \in (0,\, 1)$.

1. Compute $v^k = (\xi^k,\, \eta^k)$ solving (2.5) with H_k given by (2.6). If (2.5) cannot be solved, stop with failure.

2. Choose c satisfying

$$c \geq \|\lambda^k + \eta^k\|_\infty + \bar{c}. \tag{2.11}$$

3. If

$$\varphi'_c(x^k;\, \xi^k) \leq -\rho\|\xi^k\|^2 \tag{2.12}$$

is violated, stop with failure.

4. Set $\alpha = 1$. If the Armijo inequality (2.10) holds, set $\alpha_k = \alpha$ and go to Step 5. Otherwise, keep replacing α by $\theta\alpha$ until (2.10) is satisfied.

5. Set $u^{k+1} = u^k + \alpha_k v^k$, increase k by 1, and go to Step 1.

Observe that instead of the primal-dual update rule in Step 5, it is quite typical to use $x^{k+1} = x^k + \alpha_k \xi^k$ but $\lambda^{k+1} = \lambda^k + \eta^k$; see, e.g., [11, Algorithm 6.7]. However, the proofs of Lemmas 3.6 and 3.7 below require using the stepsize parameter in the dual update as well. This variant is also not at all uncommon; see, e.g., [18, Algorithm 18.3]. Moreover, such modifications do not alter the related global convergence results like the one in [11, Theorem 6.9].

3 Asymptotic acceptance of the true Hessian and of the unit stepsize

This section is the main part of the paper. Some further necessary preliminaries are presented in Section 3.1. Section 3.2 investigates the behavior of iteration sequences when they get close to the shifted null space of the Jacobian of Φ, while Section 3.3 deals with starting points away from the shifted null space. This analysis culminates in the main result in Section 3.4.

3.1 2-regularity issues and characterization of the full Newton step

According to [8, Proposition 1] and [9, Proposition 2], the assumptions in Proposition 2.1 for Φ defined in (2.3) may only hold at a solution $\bar{u} = (\bar{x}, \bar{\lambda})$ of (1.2) if $\bar{\lambda}$ is a critical Lagrange multiplier associated to $\bar{x}$, i.e., the linear subspace

$$Q(\bar{x}, \bar{\lambda}) = \left\{ \xi \in \ker h'(\bar{x}) \;\middle|\; \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^\top \right\} \tag{3.1}$$

is nontrivial (see [11, Definition 1.41]).

In this paper, we will restrict our attention to the case of criticality of order 1, which means that $\dim Q(\bar{x}, \bar{\lambda}) = 1$, or, in other terms,

$$Q(\bar{x}, \bar{\lambda}) = \operatorname{span}\{\bar{\xi}\} \tag{3.2}$$

with some $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$. Multipliers critical of order higher than 1 are certainly of interest, and give rise to wide possibilities for applicability of Proposition 2.1. However, the subsequent exposition demonstrates that even for criticality of order 1, the required analysis is quite involved.

Similarly to considerations in [9], it can be seen that under (3.2), the mapping Φ defined in (2.3) is 2-regular at $\bar{u}$ in a direction $v = (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ if and only if

$$\operatorname{rank} h'(\bar{x}) = l - 1 \tag{3.3}$$

and

$$h''(\bar{x})[\bar{\xi}, \xi] \notin \operatorname{im} h'(\bar{x}). \tag{3.4}$$

For every $u = (x, \lambda) \in \mathbb{R}^p$ we will make use of the decomposition $u = u_1 + u_2$, where $u_1 = (x_1, \lambda_1) \in (\ker \Phi'(\bar{u}))^\perp$ and $u_2 = (x_2, \lambda_2) \in \ker \Phi'(\bar{u})$ are uniquely defined.

The next result follows from [4, Lemma 1] applied with $v/\|v\|$ instead of $\bar{v}$, employing considerations above. It characterizes a single Newton step for the Lagrange optimality system.

Lemma 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be three times differentiable near $\bar{x} \in \mathbb{R}^n$, with their third derivatives Lipschitz-continuous with respect to $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.2)*, with an associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$, and assume that* (3.2)*–*(3.3) *hold with $Q(\bar{x}, \bar{\lambda})$ defined in* (3.1)*, and with some $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$. Set $\bar{u} = (\bar{x}, \bar{\lambda})$.*

Then for any $v = (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ with ξ satisfying (3.4)*, there exist $\bar{\varepsilon} = \bar{\varepsilon}(v) > 0$ and $\bar{\delta} = \bar{\delta}(v) > 0$ such that for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, v/\|v\|)$ there exists the unique $v^k = (\xi^k, \eta^k)$ solving* (2.5) *with H_k given by* (2.6)*, and this v^k satisfies*

$$u_1^k + v_1^k - \bar{u}_1 = O(\|u^k - \bar{u}\| \|u_1^k - \bar{u}_1\|) + O(\|u^k - \bar{u}\|^3), \tag{3.5}$$

$$u_2^k + v_2^k - \bar{u}_2 = \frac{1}{2}\pi(u^k) + O(\|u^k - \bar{u}\|^2), \tag{3.6}$$

where $\pi(u^k) \in \ker \Phi'(\bar{u})$, $\pi(u^k)$ depends homogeneously on u^k, and

$$\pi(u^k) = u_2^k - \bar{u}_2 + O(\|u_1^k - \bar{u}_1\|) \tag{3.7}$$

as $u^k \to \bar{u}$.

3.2 Behavior near the null space

Assuming (3.2), we have

$$\ker \Phi'(\bar{u}) = \{v = (t\bar{\xi}, t\bar{\eta} + \widehat{\eta}) \mid t \in \mathbb{R}, \widehat{\eta} \in \ker(h'(\bar{x}))^\top\}, \tag{3.8}$$

where $\bar{\eta}$ is the unique in $\operatorname{im} h'(\bar{x}) = (\ker(h'(\bar{x}))^\top)^\perp$ (i.e., least-length) solution of the linear system

$$(h'(\bar{x}))^\top \eta = -\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\bar{\xi}. \tag{3.9}$$

Moreover, according to the discussion in Section 3.1, Φ is 2-regular at $\bar{u}$ in every direction $\bar{v} = (t\bar{\xi},\, t\bar{\eta} + \widehat{\eta}) \in \ker \Phi'(\bar{u})$ with $t \neq 0$ if and only if (3.3) is satisfied and

$$h''(\bar{x})[\bar{\xi},\, \bar{\xi}] \notin \operatorname{im} h'(\bar{x}). \tag{3.10}$$

Therefore, the combination of (3.2)–(3.3) and (3.10) (agreeing with what appears in [8, Proposition 4]) allows us to apply Proposition 2.1 in order to characterize convergence of the basic Newton method for the Lagrange optimality system from starting points close to $\bar{u} + \ker \Phi'(\bar{u})$. Our goal in this section is to see how this behavior transmits to Algorithm 2.1 equipped with linesearch.

Observe that since $\ker(h'(\bar{x}))^\top = (\operatorname{im} h'(\bar{x}))^\perp$, under the requirement (3.3) we have that $\dim \ker(h'(\bar{x}))^\top = 1$, and hence, according to (3.8), $\dim \ker \Phi'(\bar{u}) = 2$.

Lemma 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be three times differentiable near $\bar{x} \in \mathbb{R}^n$, with their third derivatives Lipschitz-continuous with respect to $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.2)*, with an associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$, and assume that* (3.2) *with $Q(\bar{x}, \bar{\lambda})$ defined in* (3.1)*, as well as* (3.3) *and* (3.10) *hold with some $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$. Set $\bar{u} = (\bar{x}, \bar{\lambda})$. Fix any $\widehat{\eta} \in (\operatorname{im} h'(\bar{x}))^\perp$, and define $\bar{v} = (\bar{\xi},\, \bar{\eta} + \widehat{\eta})$, where $\bar{\eta}$ is the unique in $\operatorname{im} h'(\bar{x})$ solution of the linear system* (3.9)*.*

Then there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$ there exists the unique $v^k = (\xi^k, \eta^k)$ solving (2.5) *with H_k given by* (2.6)*, and this v^k satisfies*

$$u^k - \bar{u} = O(\|x_2^k - \bar{x}_2\|), \tag{3.11}$$

$$v^k = O(\|x_2^k - \bar{x}_2\|) \tag{3.12}$$

as $x_2^k \to \bar{x}_2$, and if $\bar{\delta} < 1$, then

$$u_1^k - \bar{u}_1 = O(\bar{\delta}\|x_2^k - \bar{x}_2\|) \tag{3.13}$$

as $\bar{\delta} \to 0$ and $x_2^k \to \bar{x}_2$.

Proof. Without loss of generality assume that $\bar{u} = (\bar{x}, \bar{\lambda}) = (0, 0)$. Let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0,\, 1/(2\|\bar{v}\|))$ be first chosen according to Lemma 3.1. This choice implies the existence of the unique $v^k = (\xi^k, \eta^k)$ solving (2.5) with H_k given by (2.6) for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$, and the relations (3.5)–(3.7) as $u^k \to 0$.

Furthermore, inclusion $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$ yields

$$\left\| \frac{x^k}{\|u^k\|} - \frac{\bar{\xi}}{\|\bar{v}\|} \right\| \le \left\| \frac{u^k}{\|u^k\|} - \frac{\bar{v}}{\|\bar{v}\|} \right\| \le \bar{\delta},$$

implying, in particular, that $x^k \neq 0$ provided $\bar{\delta} > 0$ is small enough, and that

$$\left| \frac{\|x^k\|}{\|u^k\|} - \frac{1}{\|\bar{v}\|} \right| \le \bar{\delta} \tag{3.14}$$

(recall that $\|\bar{\xi}\| = 1$). Therefore,

$$\left\| \frac{x^k}{\|x^k\|} - \bar{\xi} \right\| \leq \left\| \frac{\|\bar{v}\| x^k}{\|u^k\|} - \bar{\xi} \right\| + \left\| \frac{x^k}{\|x^k\|} - \frac{\|\bar{v}\| x^k}{\|u^k\|} \right\| \leq \bar{\delta}\|\bar{v}\| + \bar{\delta}\|\bar{v}\| = 2\bar{\delta}\|\bar{v}\|.$$

Observe that $x_2^k = \|x_2^k\|\bar{\xi}$ or $x_2^k = -\|x_2^k\|\bar{\xi}$. Then, arguing similarly to the proof of [4, Lemma 2], but with u^k, u_1^k, u_2^k, and $\bar{\delta}$ replaced by x^k, x_1^k, x_2^k, and $2\bar{\delta}\|\bar{v}\|$, respectively, we obtain that

$$x^k = O(\|x_2^k\|) \tag{3.15}$$

as $x^k \to 0$.

Moreover, from (3.14) it follows that

$$\frac{\|x^k\|}{\|u^k\|} \geq \frac{1}{\|\bar{v}\|} - \bar{\delta}.$$

Hence,

$$u^k = O(\|x^k\|) \tag{3.16}$$

and employing (3.5)–(3.7)

$$v^k = O(\|u^k\|) = O(\|x^k\|) \tag{3.17}$$

as $x^k \to 0$. Combining (3.16) and (3.17) with (3.15) yields (3.11) and (3.12), respectively.

Finally, (3.13) follows from [4, (19)] and (3.11). ∎

Lemma 3.3 *Let the assumptions of Lemma 3.2 be satisfied.*

Then

(a) For every $\bar{c} > 0$ there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$, $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$, and $\rho = \rho(\bar{v}) > 0$, such that for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ there exists the unique $v^k = (\xi^k, \eta^k)$ solving (2.5) with H_k given by (2.6), and for all real c satisfying (2.11) inequality (2.12) holds.

(b) For every $\bar{c} > 0$ and $\sigma \in (0, 1)$ one can choose $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ in item (a), and $\bar{\Gamma} = \bar{\Gamma}(\bar{v}) > 0$ in such a way that for every $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ satisfying

$$\|x_1^k - \bar{x}_1\| \geq \bar{\Gamma}\|x_2^k - \bar{x}_2\|^2, \tag{3.18}$$

and for all real c satisfying (2.11), it holds that

$$\varphi_c(x^k + \xi^k) \leq \varphi_c(x^k) + \sigma\varphi_c'(x^k; \xi^k). \tag{3.19}$$

(c) *For every* $\bar{c} > 0$ *and* $\sigma \in (0,\, 3/4)$ *one can choose* $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ *and* $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ *in item (a), and* $\bar{\gamma} = \bar{\gamma}(\bar{v}) > 0$ *in such a way that for every* $u^k \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$ *satisfying*

$$\|x_1^k - \bar{x}_1\| \le \bar{\gamma}\|x_2^k - \bar{x}_2\|^2, \tag{3.20}$$

inequality (3.19) *holds for all real c satisfying*

$$c \ge \frac{4(1-\sigma)\|\lambda^k + \eta^k\|_\infty + \|\lambda^k\|_\infty}{3 - 4\sigma} + \bar{c}. \tag{3.21}$$

Proof. Assume again for simplicity that $\bar{u} = (\bar{x},\, \bar{\lambda}) = (0,\, 0)$, and let $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ be first chosen according to Lemma 3.2. Let $u^k = (x^k,\, \lambda^k) \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$.

Observe that according to (2.6), (2.8), and (2.11),

$$\varphi_c'(x^k;\, \xi^k) \le -\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k,\, \xi^k \right\rangle - \bar{c}\|h(x^k)\|_1. \tag{3.22}$$

By the inclusion $\bar{\xi} \in \ker h'(0)$ and the definition of $\bar{\eta}$, it holds that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)\bar{\xi},\, \bar{\xi} \right\rangle = -\langle \bar{\eta},\, h'(0)\bar{\xi}\rangle = 0, \tag{3.23}$$

and hence, according to (3.5)–(3.7),

$$\begin{aligned}
\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k,\, \xi^k \right\rangle &= \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)\xi^k,\, \xi^k \right\rangle + O(\|u^k\|\|\xi^k\|^2)\\
&= \frac{1}{4}\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)x_2^k,\, x_2^k \right\rangle + O(\|u_1^k\|\|x_2^k\|) + O(\|u_1^k\|^2)\\
&\quad + O(\|u^k\|^3) + O(\|u^k\|\|\xi^k\|^2)\\
&= \frac{1}{4}\|x_2^k\|^2\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)\bar{\xi},\, \bar{\xi} \right\rangle + O(\|u_1^k\|\|x_2^k\|) + O(\|u_1^k\|^2)\\
&\quad + O(\|u^k\|^3) + O(\|u^k\|\|\xi^k\|^2)\\
&= O(\bar{\delta}\|x_2^k\|^2) + O(\bar{\delta}^2\|x_2^k\|^2) + O(\|x_2^k\|^3),
\end{aligned} \tag{3.24}$$

where the last equality employs (3.11)–(3.13).

Furthermore, similarly to (3.24) we obtain that

$$\begin{aligned}
h(x^k) &= h'(0)x^k + \frac{1}{2}h''(0)[x^k,\, x^k] + O(\|x^k\|^3)\\
&= h'(0)x_1^k + \frac{1}{2}h''(0)[x_2^k,\, x_2^k] + O(\bar{\delta}\|x_2^k\|^2) + O(\bar{\delta}^2\|x_2^k\|^2) + O(\|x_2^k\|^3)\\
&= h'(0)x_1^k + \frac{1}{2}\|x_2^k\|^2 h''(0)[\bar{\xi},\, \bar{\xi}] + O(\bar{\delta}\|x_2^k\|^2) + O(\bar{\delta}^2\|x_2^k\|^2) + O(\|x_2^k\|^3).
\end{aligned} \tag{3.25}$$

Let P be the orthogonal projector onto $(\operatorname{im} h'(0))^\perp$ in $\mathbb{R}^l$. Then by (3.10) we have that

$$Ph''(0)[\bar{\xi}, \bar{\xi}] \neq 0, \tag{3.26}$$

and hence, from (3.25) it follows that there exists $\gamma = \gamma(\bar{v}) > 0$ such that

$$\begin{aligned} \|h(x^k)\| &\geq \|Ph(x^k)\| \\ &= \frac{1}{2}\|x_2^k\|^2\|Ph''(0)[\bar{\xi}, \bar{\xi}]\| + O(\bar{\delta}\|x_2^k\|^2) + O(\bar{\delta}^2\|x_2^k\|^2) + O(\|x_2^k\|^3) \\ &\geq \gamma\|x_2^k\|^2 \end{aligned} \tag{3.27}$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough.

Combining (3.22), (3.24), and (3.27), and using again (3.12), after further reducing $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$, if necessary, we obtain the existence of $\rho = \rho(\bar{v}) > 0$ such that (2.12) holds. This completes the proof of item (a).

In order to establish (b) and (c), we start with the following chain of relations, employing (2.7) and (2.8) for $v^k = (\xi^k, \eta^k)$ solving (2.5) with H_k given by (2.6):

$$\begin{aligned} \varphi_c(x^k + \xi^k) - \varphi_c(x^k) - \sigma\varphi_c'(x^k;\, \xi^k) &= f(x^k + \xi^k) - f(x^k) \\ &\quad + c\|h(x^k + \xi^k)\|_1 - c\|h(x^k)\|_1 - \sigma\varphi_c'(x^k;\, \xi^k) \\ &= \langle f'(x^k),\, \xi^k\rangle + \frac{1}{2}\langle f''(x^k)\xi^k,\, \xi^k\rangle \\ &\quad + c\|h(x^k + \xi^k)\|_1 - c\|h(x^k)\|_1 - \sigma\varphi_c'(x^k;\, \xi^k) \\ &\quad + O(\|\xi^k\|^3) \\ &= (1-\sigma)\varphi_c'(x^k;\, \xi^k) + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k,\, \xi^k\right\rangle \\ &\quad + c\|h(x^k + \xi^k)\|_1 - \frac{1}{2}\langle \lambda^k,\, h''(x^k)[\xi^k,\, \xi^k]\rangle \\ &\quad + O(\|\xi^k\|^3) \\ &\leq \left(\sigma - \frac{1}{2}\right)\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\xi^k,\, \xi^k\right\rangle \\ &\quad + (1-\sigma)(\|\lambda^k + \eta^k\|_\infty - c)\|h(x^k)\|_1 \\ &\quad + c\|h(x^k + \xi^k)\|_1 - \frac{1}{2}\langle \lambda^k,\, h''(x^k)[\xi^k,\, \xi^k]\rangle \\ &\quad + O(\|\xi^k\|^3). \end{aligned} \tag{3.28}$$

Since ξ^k solves the second equation in (2.5), we have

$$\begin{aligned} h(x^k + \xi^k) &= h(x^k) + h'(x^k)\xi^k + \frac{1}{2}h''(x^k)[\xi^k,\, \xi^k] + O(\|\xi^k\|^3) \\ &= \frac{1}{2}h''(x^k)[\xi^k,\, \xi^k] + O(\|\xi^k\|^3), \end{aligned} \tag{3.29}$$

and combining this with (3.11), (3.12), (3.24), and (3.28), yields

$$\begin{aligned}\varphi_c(x^k+\xi^k)-\varphi_c(x^k)-\sigma\varphi_c'(x^k;\,\xi^k) &\leq (1-\sigma)(\|\lambda^k+\eta^k\|_\infty-c)\|h(x^k)\|_1\\ &+\frac{1}{2}\left(c+\|\lambda^k\|_\infty\right)\|h''(x^k)[\xi^k,\,\xi^k]\|\\ &+O(\bar{\delta}\|x_2^k\|^2)+O(\bar{\delta}^2\|x_2^k\|^2)+O(\|x_2^k\|^3).\end{aligned} \tag{3.30}$$

We are now in a position to prove item (b). Suppose that (3.18) is satisfied. Let $\bar{\nu}>0$ be the smallest positive singular value of $h'(0)$. Then from (3.25) we have that

$$\|h(x^k)\|\geq\bar{\nu}\|x_1^k\|+O(\|x_2^k\|^2)\geq\nu\bar{\Gamma}\|x_2^k\|^2$$

for every pre-fixed $\nu\in(0,\,\bar{\nu})$ provided $\bar{\varepsilon}>0$ is small enough while $\bar{\Gamma}>0$ is large enough. Combining this with (2.11), (3.12), and (3.30), yields (3.19) provided $\bar{\varepsilon}>0$ is small enough while $\bar{\Gamma}>0$ is large enough.

In order to establish item (c), suppose now that (3.20) is satisfied. From (3.25) we then have

$$h(x^k)=\frac{1}{2}\|x_2^k\|^2h''(0)[\bar{\xi},\,\bar{\xi}]+O((\bar{\gamma}+\bar{\delta})\|x_2^k\|^2)+O(\bar{\delta}^2\|x_2^k\|^2)+O(\|x_2^k\|^3). \tag{3.31}$$

Furthermore, by (3.5)–(3.7), (3.11), and by (3.20), it holds that

$$\xi^k=-\frac{1}{2}x_2^k+O(\|x_2^k\|^2).$$

Then

$$h''(x^k)[\xi^k,\,\xi^k]=\frac{1}{4}h''(0)[x_2^k,\,x_2^k]+O(\|x_2^k\|^3)=\frac{1}{4}\|x_2^k\|^2h''(0)[\bar{\xi},\,\bar{\xi}]+O(\|x_2^k\|^3). \tag{3.32}$$

Employing (3.31) and (3.32), from (3.30) we now obtain

$$\begin{aligned}\varphi_c(x^k+\xi^k)-\varphi_c(x^k)-\sigma\varphi_c'(x^k;\,\xi^k) &\leq \frac{1}{2}\Bigg((1-\sigma)(\|\lambda^k+\eta^k\|_\infty-c)\\ &+\frac{1}{4}\left(c+\|\lambda^k\|_\infty\right)\Bigg)\|x_2^k\|^2\|h''(0)[\bar{\xi},\,\bar{\xi}]\|\\ &+O((\bar{\gamma}+\bar{\delta})\|x_2^k\|^2)+O(\bar{\delta}^2\|x_2^k\|^2)+O(\|x_2^k\|^3).\end{aligned}$$

Since from (3.21) it follows that

$$\left((1-\sigma)(\|\lambda^k+\eta^k\|_\infty-c)+\frac{1}{4}\left(c+\|\lambda^k\|_\infty\right)\right)\leq-\frac{3-4\sigma}{4}\bar{c},$$

we then have that

$$\begin{array}{rcl}\varphi_c(x^k+\xi^k)-\varphi_c(x^k)-\sigma\varphi_c'(x^k;\,\xi^k) & \leq & -\dfrac{3-4\sigma}{8}\bar{c}\|x_2^k\|^2\|h''(0)[\bar{\xi},\,\bar{\xi}]\| \\ & & +O((\bar{\gamma}+\bar{\delta})\|x_2^k\|^2)+O(\bar{\delta}^2\|x_2^k\|^2)+O(\|x_2^k\|^3).\end{array}$$

Employing again (3.26) we now conclude that (3.19) holds provided $\bar{\varepsilon}>0$, $\bar{\delta}>0$, and $\bar{\gamma}>0$, are small enough. ■

We next demonstrate by examples that neither assumption (3.3) nor (3.10) can be dropped in Lemma 3.3, and hence, in subsequent developments. We also show that the restriction $\sigma\leq 3/4$ is essential in item (c) of Lemma 3.3. Recall again that assuming (3.2), both (3.3) and (3.10) are not only sufficient but also necessary for 2-regularity of Φ at $\bar{u}$ in the direction $\bar{v}=(\bar{\xi},\,\bar{\eta}+\widehat{\eta})\in\ker\Phi'(\bar{u})$. This implies that when any of these assumptions is violated, $v^k=(\xi^k,\,\eta^k)$ solving (2.5) may not exist at $u^k=(x^k,\,\lambda^k)\in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\,\bar{v}/\|\bar{v}\|)$, or may be not unique, no matter how small are $\bar{\varepsilon}>0$ and $\bar{\delta}>0$. Therefore, assertion (a) may not hold in these circumstances.

Example 3.1 Problem (1.2) with $n=l=2$,

$$f(x)=-\|x\|^2,\quad h(x)=(x_1^2,\,x_2^2),$$

has the unique solution $\bar{x}=0$. We have

$$h'(0)=0,\quad h''(0)[\xi]=\left(\begin{array}{cc}2\xi_1 & 0\\ 0 & 2\xi_2\end{array}\right).$$

In particular, $\operatorname{rank}h'(0)=l-2$, and hence, (3.3) is violated. One can verify that critical multipliers are those satisfying $\lambda_1=1$ or $\lambda_2=1$, and they are all of order 1, except for (1, 1) that is of order 2.

Take, e.g., a critical multiplier $\bar{\lambda}=(1,\,0)$. Then (3.2) with $Q(0,\,\bar{\lambda})$ defined in (3.1) holds with $\bar{\xi}=(1,\,0)$, and $h''(0)[\bar{\xi},\,\xi]=(2\xi_1,\,0)$, implying that (3.4) holds for all $\xi\in\mathbb{R}^2$ with $\xi_1\neq 0$, and in particular, (3.10) holds.

Consider $x=t\bar{\xi}$, and $u=(t\bar{\xi},\,\bar{\lambda})$ that belongs to $K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\,\bar{v}/\|\bar{v}\|)$ for any $\bar{\varepsilon}>0$ and $\bar{\delta}>0$ provided $t>0$ is small enough. Then

$$h'(x)=\left(\begin{array}{cc}2t & 0\\ 0 & 0\end{array}\right),$$

implying that the last row of $\Phi'(u)$ for Φ defined in (2.3) consists of zeroes, and hence, $\Phi'(u)$ is singular. Moreover, if we modify h by setting, say, $h(x)=(x_1^2,\,x_2^2+x_1^3)$, then all the conclusions above remain unchanged, but the second equality in (2.5) with $x^k=x$ and $\lambda^k=\bar{\lambda}$ takes the form

$$t^2+2t\xi_1=0,\quad t^3+3t^2\xi_1=0,$$

and there exists no ξ satisfying these two equations simultaneously.

Example 3.2 Problem (1.2) with $n = l = 1$,

$$f(x) = x^p, \quad h(x) = x^q,$$

where p and q are positive integers, has the unique solution $\bar{x} = 0$. If $q = 1$, (2.4) holds, and this is not the case of interest in this work. Therefore, let $q \geq 2$, in which case $\bar{x}$ is stationary if and only if $p \geq 2$. Then

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

implying that (3.3) holds, and the set of Lagrange multipliers associated with $\bar{x} = 0$ is the entire $\mathbb{R}$.

Consider first the case when $p = q = 2$.

$$h''(0)[\xi] = 2\xi, \quad \frac{\partial^2 L}{\partial x^2}(0,\, \lambda) = 2(1+\lambda),$$

implying that there is the unique critical multiplier associated with $\bar{x} = 0$, namely, $\bar{\lambda} = -1$, and (3.2) with $Q(0,\, \bar{\lambda})$ defined in (3.1) holds with $\bar{\xi} = \pm 1$. Then

$$h''(0)[\bar{\xi},\, \xi] = \pm 2\xi,$$

implying that (3.4) holds for all $\xi \neq 0$ (and in particular, (3.10) holds).

System (2.5) with index k dropped has the form

$$2x(1+\lambda)x + 2(1+\lambda)\xi + 2x\eta = 0, \quad x^2 + 2x\xi = 0,$$

and assuming that $x \neq 0$, this gives $\xi = -x/2$, $\eta = -(1+\lambda)/2$. Substituting ξ in the equality in (2.8) (again with index k dropped), we get

$$\varphi_c'(x;\, \xi) = -(1+c)x^2 = -4(1+c)\xi^2,$$

and hence, (2.12) always holds with, say, $\rho = 4$, whatever is taken as $c \geq 0$. Furthermore, from (2.7) we get

$$\varphi_c(x+\xi) = \frac{1}{4}(1+c)x^2, \quad \varphi_c(x) = (1+c)x^2,$$

and hence,

$$\varphi_c(x+\xi) - \varphi_c(x) - \sigma\varphi_c'(x;\, \xi) = \left(\sigma - \frac{3}{4}\right)(1+c)x^2 \leq 0$$

precisely when $\sigma \leq 3/4$. Therefore, this restriction on σ cannot be avoided in order to ensure (3.19).

Furthermore, if $p = 3$, $q = 2$, then the unique critical multiplier associated with $\bar{x} = 0$ is $\bar{\lambda} = 0$, and one can directly verify that all the assumptions of Lemma 3.2 are satisfied.

Consider finally the case when $p \geq 3$ and $q \geq 3$. We have

$$h''(0) = 0, \quad \frac{\partial^2 L}{\partial x^2}(0, \lambda) = 0,$$

implying that every $\lambda \in \mathbb{R}$ is a critical Lagrange multiplier associated with $\bar{x} = 0$, and (3.2) holds with $\bar{\xi} = \pm 1$. However, (3.4) is violated for all $\xi \neq 0$, and in particular, (3.10) does not hold.

System (2.5) with index k dropped has the form

$$px^{p-1} + q\lambda x^{q-1} + (p(p-1)x^{p-2} + q(q-1)\lambda x^{q-2})\xi + qx^{q-1}\eta = 0, \quad x^q + qx^{q-1}\xi = 0,$$

and assuming that $x \neq 0$, this gives

$$\xi = -x/q, \quad \eta = -\frac{2q-1}{q}\lambda - \frac{p(p+q-1)}{q^2}x^{p-q}. \tag{3.33}$$

Substituting ξ in the equality in (2.8) (again with index k dropped), we get

$$\varphi_c'(x;\, \xi) = -\frac{p}{q}x^p - c|x^q|.$$

Then for any fixed $\rho > 0$, (2.12) with $x^k = x$, $\xi^k = \xi$ holds if and only if

$$c \geq \frac{\rho}{q^2}|x^{2-q}| - \frac{p}{q}\operatorname{sign}(x^p)|x^{p-q}|.$$

According to the assumptions $p \geq 3$ and $q \geq 3$, and to the second equality in (3.33), for any fixed $\bar{c} > 0$, this cannot be assured by (2.11) with $\lambda^k = \lambda$, $\eta^k = \eta$, for x close to 0, no matter how small is ρ.

Furthermore, even if $\xi^k = \xi$ is accepted by (2.12), inequality (2.10) with $x^k = x$ takes the form

$$\left(\left(\frac{q-1}{q}\right)^p + \sigma\frac{p}{q} - 1\right)x^p + c\left(\left(\frac{q-1}{q}\right)^q + \sigma - 1\right)|x^q| \leq 0.$$

If $p = q$ and $x^p \geq 0$, this reduces to

$$\left(\frac{q-1}{q}\right)^q + \sigma - 1 \leq 0,$$

and hence,

$$\sigma \leq 1 - \left(1 - \frac{1}{q}\right)^q,$$

where the right-hand side equals $3/4$ for $q = 2$, and monotonically decreases to $1 - e^{-1}$ as $q \to +\infty$. In particular, for $q \geq 3$, it is not enough to assume that $\sigma \leq 3/4$ in order to ensure (3.19).

The next lemma demonstrates that once an iterate gets close enough to $\bar{u}$ and to a direction of $\bar{v}$, the next iterate necessarily satisfies (3.20) with any pre-fixed $\bar{\gamma} > 0$.

Lemma 3.4 *Under the assumptions of Lemma 3.2, for every $\bar{\gamma} > 0$ there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ there exists the unique $v^k = (\xi^k, \eta^k)$ solving (2.5) with H_k given by (2.6), and it holds that*

$$\|x_1^k + \xi_1^k - \bar{x}_1\| \leq \bar{\gamma}\|x_2^k + \xi_2^k - \bar{x}_2\|^2. \tag{3.34}$$

Proof. Assume again that $\bar{u} = (\bar{x}, \bar{\lambda}) = (0, 0)$, and let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0, 1)$ be first chosen according to Lemma 3.2.

For $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ we then get from (3.5)–(3.7), (3.11) and (3.13) that

$$x_1^k + \xi_1^k = O(\bar{\delta}\|x_2^k\|^2) + O(\|x_2^k\|^3),$$

$$x_2^k + \xi_2^k = \frac{1}{2}\|x_2^k\| + O(\bar{\delta}\|x_2^k\|) + O(\|x_2^k\|^2),$$

which evidently implies (3.34) provided that $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. ∎

Combining Proposition 2.1 with assertion (c) of Lemma 3.3 and with Lemma 3.4, we come to the following statement.

Proposition 3.1 *Let the assumptions of Lemma 3.2 be satisfied.*

Then, for every $\sigma \in (0, 3/4)$, one can choose $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ according to Proposition 2.1 in such a way that, in addition to its assertion, if c is chosen satisfying (3.21) for each k, and if (3.19) holds for $k = 0$, then (3.19) holds for all k as well.

The next lemma deals with the issue of keeping the penalty parameter bounded.

Lemma 3.5 *Let the assumptions of Lemma 3.2 be satisfied.*

Then for every $\bar{c} > 0$ and $\sigma \in (0, 1)$, and every $\bar{C} > 0$, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$, $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$, and $\bar{\alpha} = \bar{\alpha}(\bar{v}) > 0$ such that, for every $u^k = (x^k, \lambda^k) \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ there exists the unique $v^k = (\xi^k, \eta^k)$ solving (2.5) with H_k given by (2.6), and (2.10) holds for all $c \in [\|\lambda^k + \eta^k\|_\infty + \bar{c}, \bar{C}]$ and $\alpha \in (0, \bar{\alpha}]$. In addition,

(a) If

$$\bar{C} > \|\bar{\lambda}\|_\infty + \bar{c}, \tag{3.35}$$

then

$$\|\lambda^k + \eta^k\|_\infty + \bar{c} \leq \bar{C} \tag{3.36}$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are taken small enough.

(b) If

$$\bar{C} > \frac{5-4\sigma}{3-4\sigma}\|\bar{\lambda}\|_\infty + \bar{c}, \tag{3.37}$$

then

$$\frac{4(1-\sigma)\|\lambda^k + \eta^k\|_\infty + \|\lambda^k\|_\infty}{3-4\sigma} \leq \bar{C}$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are taken small enough.

Proof. Let $\bar{\varepsilon} > 0$, $\bar{\delta} > 0$, and $\rho > 0$ be chosen according to both Lemma 3.2 and item (a) of Lemma 3.3.

Take any $\alpha \in (0, 1]$. Under the smoothness assumptions of Lemma 3.2, f' and h' are Lipschitz-continuous near $\bar{x}$ with some constant $\ell > 0$. By [11, Lemma A.11] it then follows that

$$f(x^k + \alpha\xi^k) \leq f(x^k) + \alpha\langle f'(x^k),\, \xi^k\rangle + \frac{\ell\alpha^2}{2}\|\xi^k\|^2, \tag{3.38}$$

and by the second equality in (2.5)

$$|h(x^k+\alpha\xi^k)| = |h(x^k+\alpha\xi^k) - \alpha h(x^k) - \alpha h'(x^k)\xi^k| \leq (1-\alpha)|h(x^k)| + \frac{\ell\alpha^2}{2}\|\xi^k\|^2, \tag{3.39}$$

where the absolute value is taken componentwise.

Using (2.3), and combining the relations (3.38)–(3.39), we obtain that

$$\begin{aligned} \varphi_c(x^k + \alpha\xi^k) &= f(x^k + \alpha\xi^k) + c\|h(x^k + \alpha\xi^k)\|_1 \\ &\leq f(x^k) + \alpha\langle f'(x^k),\, \xi^k\rangle + c(1-\alpha)\|h(x^k)\|_1 \\ &= \varphi_c(x^k) + \alpha\langle f'(x^k),\, \xi^k\rangle - \alpha c\|h(x^k)\|_1 + C\alpha^2\|\xi^k\|^2 \\ &= \varphi_c(x^k) + \alpha\varphi_c'(x^k;\, \xi^k) + C\alpha^2\|\xi^k\|^2, \end{aligned} \tag{3.40}$$

where

$$C = \frac{\ell}{2}(1 + lc) \leq \frac{\ell}{2}(1 + l\bar{C}). \tag{3.41}$$

Employing (3.40) we observe that the inequality (2.10) is satisfied if

$$\varphi_c'(x^k;\, \xi^k) + C\alpha\|\xi^k\|^2 \leq \sigma\varphi_c'(x^k;\, \xi^k).$$

According to (2.12) and (3.41), this is automatic if $\alpha \leq \bar{\alpha}$, where we have set

$$\bar{\alpha} = \frac{2(1-\sigma)\rho}{\ell(1+l\bar{C})} \leq \frac{(1-\sigma)\rho}{C}.$$

It remains to prove items (a) (showing that under (3.35), $[\|\lambda^k+\eta^k\|_\infty+\bar{c},\, \bar{C}] \neq \emptyset$, and hence the assertion just proven is not vacuous) and (b).

From (3.5)–(3.7) and (3.13) it follows that by further reducing $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ if necessary, for any $u^k = (x^k,\, \lambda^k) \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$ we have

$$\|u^k + v^k - \bar{u}\| \leq \|u^k - \bar{u}\| \leq \bar{\varepsilon}.$$

Hence, employing (3.35), by further reducing $\bar{\varepsilon} > 0$ if necessary, we can ensure that

$$\|\lambda^k + \eta^k\|_\infty + \bar{c} \leq \|\bar{\lambda}\|_\infty + \bar{c} + \|u^k + v^k - \bar{u}\|_\infty \leq \bar{C},$$

i.e., (3.36) holds.

Item (b) is established by similar considerations. ■

In the next lemma we need an extra restriction on the choice of values of the penalty parameter on Step 2 of Algorithm 2.1, since for application of Lemma 3.5 it is essential to keep the sequence of these values bounded. Specifically, suppose that c is taken equal to the value from the previous iteration if $k \geq 1$ and this value satisfies (2.11), and equal to $\|\lambda^k + \eta^k\|_\infty + \bar{c}$ otherwise. This specification of Algorithm 2.1 will be referred to as **Algorithm 2.1***.

Lemma 3.6 *Let the assumptions of Lemma 3.2 be satisfied.*

Then for every $\bar{c} > 0$, $\rho > 0$, $\sigma \in (0,\, 1)$, and $\theta \in (0,\, 1)$, and for every $\bar{\gamma} > 0$, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that, if a sequence $\{u^k\} \subset K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$ is generated by Algorithm 2.1, then (3.20) holds for some k.*

Proof. Let us assume again that $\bar{u} = (\bar{x},\, \bar{\lambda}) = (0,\, 0)$, and let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0,\, 1)$ be chosen according to both Lemma 3.2 and Lemma 3.5 applied with any $\bar{C}$ satisfying (3.35).

For every $u^k \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u},\, \bar{v})$ by (3.5)–(3.7), (3.11) and (3.13) we have

$$x_1^{k+1} = x_1^k + \alpha_k \xi_1^k = (1 - \alpha_k) x_1^k + O(\bar{\delta}\|x_2^k\|^2) + O(\|x_2^k\|^3), \tag{3.42}$$

and

$$x_2^{k+1} = x_2^k + \alpha_k \xi_2^k = \left(1 - \frac{1}{2}\alpha_k\right) x_2^k + O(\bar{\delta}\|x_2^k\|) + O(\|x_2^k\|^2). \tag{3.43}$$

(This is where we need the dual updating rule of the form $\lambda^{k+1} = \lambda^k + \alpha_k \eta^k$, since with $\lambda^{k+1} = \lambda^k + \eta^k$ instead, say, $x_1^k + \alpha_k \xi_1^k$ need not be equal to x_1^{k+1}.)

Let us now suppose that x^k violates (3.20) for some $k \in \mathbb{N}$. Then (3.42) and (3.43) imply the existence of $C > 0$ and $c > 0$ such that

$$\|x_1^{k+1}\| \leq (1 - \alpha_k)\|x_1^k\| + C(\bar{\delta}\|x_2^k\|^2 + \|x_2^k\|^3) \leq \left(1 - \alpha_k + \frac{C}{\bar{\gamma}}(\bar{\delta} + \|x_2^k\|)\right) \|x_1^k\|$$

and

$$\|x_2^{k+1}\| \geq \left(1 - \frac{1}{2}\alpha_k\right)\|x_2^k\| - c(\bar{\delta}\|x_2^k\| + \|x_2^k\|^2) \geq \left(1 - \frac{1}{2}\alpha_k - c\left(\bar{\delta} + \|x_2^k\|\right)\right)\|x_2^k\|.$$

Therefore, we obtain

$$\frac{\|x_1^{k+1}\|}{\|x_2^{k+1}\|^2} \leq \frac{1 - \alpha_k + \dfrac{C}{\bar{\gamma}}(\bar{\delta} + \|x_2^k\|)}{\left(1 - \dfrac{1}{2}\alpha_k - c\left(\bar{\delta} + \|x_2^k\|\right)\right)^2}\frac{\|x_1^k\|}{\|x_2^k\|^2}. \tag{3.44}$$

According to Lemma 3.5, there exists $\widehat{\alpha} \in (0,\, 1]$ independent of k, and such that $\alpha_k \in [\widehat{\alpha},\, 1]$. It can be easily seen that the function $\varphi : [0,\, 1] \to \mathbb{R}$, $\varphi(\alpha) = (1-\alpha)/(1-\alpha/2)^2$, is monotonically decreasing with $\varphi(0) = 1$, $\varphi(1) = 0$. Therefore, $\widehat{q} = \varphi(\widehat{\alpha}) \in [0,\, 1)$, and $\varphi(\alpha) \leq \widehat{q}$ for all $\alpha \in [\widehat{\alpha},\, 1]$. Then, by the uniform continuity argument, we obtain that for every $q \in (\widehat{q},\, 1)$ it holds that

$$\frac{1 - \alpha_k + \dfrac{C}{\bar{\gamma}}(\bar{\delta} + \|x_2^k\|)}{\left(1 - \dfrac{1}{2}\alpha_k - c\left(\bar{\delta} + \|x_2^k\|\right)\right)^2} \in (0,\, q],$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. Inequality (3.44) now yields

$$\frac{\|x_1^{k+1}\|}{\|x_2^{k+1}\|^2} \leq q\frac{\|x_1^k\|}{\|x_2^k\|^2}.$$

Since $q \in (0,\, 1)$, this implies that if x^k violates (3.20) for all sufficiently large k, then $\|x_1^k\|/\|x_2^k\|^2 \to 0$ as $k \to \infty$. But then $\|x_1^k\|/\|x_2^k\|^2 \leq \bar{\gamma}$ for all sufficiently large k, which contradicts the violation of (3.20). ∎

Let **Algorithm 2.1**** be defined the same way as Algorithm 2.1*, but with (2.11) replaced by the stronger condition (3.21). It can be easily seen from its proof that Lemma 3.6 remains valid with Algorithm 2.1* in it substituted by Algorithm 2.1**. The next result corresponds to [4, Lemma 6]. We provide here a full proof precisely because only a sketch of it was given in [4]; otherwise the specificity of a linesearch rule is irrelevant in these results demonstrating that once initialized at u^0 close enough to $\bar{u}$, and with the direction of $u^0 - \bar{u}$ close enough to that of $\bar{v}$, the subsequent iterates of Algorithm 2.1** are well-defined and keep satisfying these proximity properties.

Lemma 3.7 *Let the assumptions of Lemma 3.2 be satisfied.*

*Then, for every $\bar{\varepsilon} > 0$, $\bar{\delta} > 0$, $\bar{c} > 0$, $\sigma \in (0,\, 1)$, and $\theta \in (0,\, 1)$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$, $\delta = \delta(\bar{v}) > 0$, and $\rho = \rho(\bar{v}) > 0$, such that, for every starting point $u^0 \in K_{\varepsilon,\delta}(\bar{u},\, \bar{v}/\|\bar{v}\|)$, Algorithm 2.1** uniquely defines the sequence $\{u^k\}$, and $\{u^k\} \subset K_{\bar{\varepsilon},\bar{\delta}}(\bar{u},\, \bar{v}/\|\bar{v}\|)$.*

Proof. Assume again that $\bar{u} = (\bar{x}, \bar{\lambda}) = (0, 0)$. Without loss of generality, we can further assume that $\bar{\varepsilon} > 0$, $\bar{\delta} > 0$, and $\rho > 0$ are chosen according to Lemmas 3.1, 3.3, and 3.5, applied with any $\bar{C}$ satisfying (3.37) (an explanation of why this assumption does not reduce generality of the choice of $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ can be found in the proof in [8, Theorem 1]). Then, for every $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ we have that the test (2.12) is passed, and relations (3.5)–(3.7) hold.

Considering that $\bar{v}_1 = 0$ (since $\bar{v} \in \ker \Phi'(\bar{u})$), observe first that if $u \in K_{\varepsilon, \delta}(\bar{u}, \bar{v}/\|\bar{v}\|)$ with some $\varepsilon > 0$ and $\delta \in (0, 1)$, then

$$\frac{\|u_1\|}{\|u\|} = \left\| \frac{u_1}{\|u\|} - \frac{\bar{v}_1}{\|\bar{v}\|} \right\| \leq \left\| \frac{u}{\|u\|} - \frac{\bar{v}}{\|\bar{v}\|} \right\| \leq \delta.$$

This implies that

$$\|u_1\| \leq \delta \|u\|, \tag{3.45}$$

and hence,

$$\|u\| \leq \|u_1\| + \|u_2\| \leq \delta \|u\| + \|u_2\|,$$

so that

$$(1 - \delta)\|u\| \leq \|u_2\|. \tag{3.46}$$

Then

$$\begin{aligned}
\left\| \frac{u_2}{\|u_2\|} - \frac{\bar{v}}{\|\bar{v}\|} \right\| &\leq \left\| \frac{u_2}{\|u\|} - \frac{\bar{v}_2}{\|\bar{v}\|} \right\| + \left\| \frac{u_2}{\|u_2\|} - \frac{u_2}{\|u\|} \right\| \\
&\leq \left\| \frac{u}{\|u\|} - \frac{\bar{v}}{\|\bar{v}\|} \right\| + \frac{\|u\| - \|u_2\|}{\|u\|} \\
&\leq \delta + 1 - \frac{\|u_2\|}{\|u\|} \\
&\leq 2\delta.
\end{aligned} \tag{3.47}$$

For any $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ with $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ specified above, from (3.5)–(3.7) we have

$$u_1^{k+1} = u_1^k + \alpha_k v_1^k = (1 - \alpha_k) u_1^k + O(\|u_1^k\| \|u^k\| + \|u^k\|^3)$$

and

$$u_2^{k+1} = u_2^k + \alpha_k v_2^k = \left(1 - \frac{1}{2}\alpha_k\right) u_2^k + O(\|u_1^k\| + \|u^k\|^2)$$

as $u^k \to 0$. Therefore, there exists $C > 0$ such that

$$\|u_1^{k+1}\| \leq (1 - \alpha_k)\|u_1^k\| + C\|u^k\|^2, \tag{3.48}$$

$$\left\| u_2^{k+1} - \left(1 - \frac{1}{2}\alpha_k\right) u_2^k \right\| \leq \left\| u^{k+1} - \left(1 - \frac{1}{2}\alpha_k\right) u_2^k \right\| \leq C(\|u_1^k\| + \|u^k\|^2), \tag{3.49}$$

and hence,

$$
\begin{aligned}
\left(1-\frac{1}{2}\alpha_k\right)\|u_2^k\| - C(\|u_1^k\| + \|u^k\|^2) &\leq \|u_2^{k+1}\| \\
&\leq \|u^{k+1}\| \\
&\leq \left(1-\frac{1}{2}\alpha_k\right)\|u_2^k\| + C(\|u_1^k\| + \|u^k\|^2).
\end{aligned}
$$

Employing (3.45) and (3.46) with $\delta = \bar{\delta}$, we then further derive that

$$
\begin{aligned}
\left(\left(1-\frac{1}{2}\alpha_k\right)(1-\bar{\delta}) - C(\bar{\delta}+\bar{\varepsilon})\right)\|u^k\| &\leq \|u_2^{k+1}\| \\
&\leq \|u^{k+1}\| \\
&\leq \left(1-\frac{1}{2}\alpha_k + C(\bar{\delta}+\bar{\varepsilon})\right)\|u^k\|.
\end{aligned}
\tag{3.50}
$$

By further reducing $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ if necessary, we can ensure the inequalities

$$
\frac{\bar{\delta}}{2} + C(\bar{\delta}+\bar{\varepsilon}) < \frac{1}{2}, \quad C(\bar{\delta}+\bar{\varepsilon}) < \frac{1}{2}\widehat{\alpha}.
$$

Then setting

$$
q_- = \frac{1-\bar{\delta}}{2} - C(\bar{\delta}+\bar{\varepsilon}), \quad q_+ = 1 - \frac{1}{2}\widehat{\alpha} + C(\bar{\delta}+\bar{\varepsilon}),
$$

from (3.50) we obtain that

$$
q_-\|u^k\| \leq \|u_2^{k+1}\| \leq \|u^{k+1}\| \leq q_+\|u^k\|, \tag{3.51}
$$

where

$$
0 < q_- < q_+ < 1. \tag{3.52}
$$

Arguing similarly to the proof of Lemma 3.6, but employing the function $\varphi : [0,\,1] \to \mathbb{R}$, $\varphi(\alpha) = (1-\alpha)/(1-\alpha/2)$, we conclude that $\varphi(\widehat{\alpha}) \in (0,\,1)$, and for every $q \in (\varphi(\widehat{\alpha}),\,1)$ it holds that

$$
\frac{1-\alpha_k}{\left(1-\frac{1}{2}\alpha_k\right)(1-\bar{\delta}) - C(\bar{\delta}+\bar{\varepsilon})} \in (0,\,q]
$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. Then from (3.48) and (3.50) we have

$$
\frac{\|u_1^{k+1}\|}{\|u^{k+1}\|} \leq q\frac{\|u_1^k\|}{\|u^k\|} + \frac{C}{q_-}\|u^k\|. \tag{3.53}
$$

By (3.49) and the left inequality in (3.51) we have that

$$\begin{aligned}
\left\| \frac{u^{k+1}}{\|u^{k+1}\|} - \frac{u_2^k}{\|u_2^k\|} \right\| &= \frac{\| \|u_2^k\| u^{k+1} - \|u^{k+1}\| u_2^k \|}{\|u_2^k\| \|u^{k+1}\|} \\
&\leq \frac{\|u_2^k\| \|u^{k+1} - (1-\alpha_k/2)u_2^k\| + |\|u^{k+1}\| - (1-\alpha_k/2)\|u_2^k\|| \|u_2^k\|}{\|u_2^k\| \|u^{k+1}\|} \\
&\leq \frac{2C(\|u_1^k\| + \|u^k\|^2)}{q_- \|u^k\|} \\
&= \frac{2C}{q_-} \left(\frac{\|u_1^k\|}{\|u^k\|} + \|u^k\| \right), \qquad (3.54)
\end{aligned}$$

and similarly,

$$\left\| \frac{u_2^{k+1}}{\|u_2^{k+1}\|} - \frac{u_2^k}{\|u_2^k\|} \right\| \leq \frac{2C}{q_-} \left(\frac{\|u_1^k\|}{\|u^k\|} + \|u^k\| \right). \qquad (3.55)$$

Now choose $\varepsilon \in (0, \bar{\varepsilon}]$ and $\delta \in (0, \bar{\delta}]$ satisfying

$$2\delta + \frac{2C}{q_-} \left(\frac{\delta}{1-q} + \frac{C\varepsilon}{q_-(1-q)(1-q_+)} + \frac{\varepsilon}{1-q_+} \right) \leq \bar{\delta}, \qquad (3.56)$$

and assume that $u^0 \in K_{\varepsilon,\delta}(\bar{u}, \bar{v}/\|\bar{v}\|)$. Suppose that iterates $u^j \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$ are generated by Algorithm 2.1** for all $j = 1, \ldots, k$. Then by the choice of $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$, the algorithm uniquely defines u^{k+1} satisfying (3.51)–(3.55).

From (3.51)–(3.52) we have

$$0 < \|u^{k+1}\| \leq q_+ \|u^k\| \leq q_+^2 \|u^{k-1}\| \leq \ldots \leq q_+^{k+1} \|u^0\| \leq q_+^{k+1}\varepsilon \leq \varepsilon \leq \bar{\varepsilon}. \qquad (3.57)$$

Furthermore, employing (3.45), (3.53), (3.57), we obtain the estimate

$$\begin{aligned}
\sum_{j=0}^{k} \frac{\|u_1^j\|}{\|u^j\|} &\leq \frac{\|u_1^0\|}{\|u^0\|} + \sum_{j=0}^{k-1} \left(q \frac{\|u_1^j\|}{\|u^j\|} + \frac{C}{q_-} \|u^j\| \right) \\
&\leq \frac{\|u_1^0\|}{\|u^0\|} + q \frac{\|u_1^0\|}{\|u^0\|} + \sum_{j=0}^{k-2} \left(q^2 \frac{\|u_1^j\|}{\|u^j\|} + q \frac{C}{q_-} \|u^j\| \right) + \frac{C}{q_-} \sum_{j=0}^{k-1} \|u^j\| \\
&\leq \frac{\|u_1^0\|}{\|u^0\|} + q \frac{\|u_1^0\|}{\|u^0\|} + q^2 \frac{\|u_1^0\|}{\|u^0\|} + \sum_{j=0}^{k-3} \left(q^3 \frac{\|u_1^j\|}{\|u^j\|} + q^2 \frac{C}{q_-} \|u^j\| \right) \\
&\quad + q \frac{C}{q_-} \sum_{j=0}^{k-2} \|u^j\| + \frac{C}{q_-} \sum_{j=0}^{k-1} \|u^j\| \\
&\leq \dots \\
&\leq \frac{\|u_1^0\|}{\|u^0\|} \sum_{j=0}^{k} q^j + \frac{C}{q_-} \sum_{i=0}^{k-1} q^i \sum_{j=0}^{k-i-1} \|u^j\| \\
&\leq \delta \sum_{j=0}^{k} q^j + \frac{C}{q_-} \varepsilon \sum_{i=0}^{k-1} q^i \sum_{j=0}^{k-i-1} q_+^j \\
&\leq \frac{\delta}{1-q} + \frac{C\varepsilon}{q_-(1-q)(1-q_+)}.
\end{aligned}$$

With this estimate at hand, and employing (3.47), (3.54)–(3.55), and (3.57), we finally conclude that

$$
\begin{aligned}
\left\|\frac{u^{k+1}}{\|u^{k+1}\|}-\bar v\right\| &\le \left\|\frac{u_2^k}{\|u_2^k\|}-\bar v\right\| + \left\|\frac{u^{k+1}}{\|u^{k+1}\|}-\frac{u_2^k}{\|u_2^k\|}\right\| \\
&\le \left\|\frac{u_2^{k-1}}{\|u_2^{k-1}\|}-\bar v\right\| + \left\|\frac{u_2^k}{\|u_2^k\|}-\frac{u_2^{k-1}}{\|u_2^{k-1}\|}\right\| + \left\|\frac{u^{k+1}}{\|u^{k+1}\|}-\frac{u_2^k}{\|u_2^k\|}\right\| \\
&\le \dots \\
&\le \left\|\frac{u_2^0}{\|u_2^0\|}-\bar v\right\| + \sum_{j=1}^{k}\left\|\frac{u_2^j}{\|u_2^j\|}-\frac{u_2^{j-1}}{\|u_2^{j-1}\|}\right\| + \left\|\frac{u^{k+1}}{\|u^{k+1}\|}-\frac{u_2^k}{\|u_2^k\|}\right\| \\
&\le 2\delta + \sum_{j=1}^{k}\frac{2C}{q_-}\left(\frac{\|u_1^{j-1}\|}{\|u^{j-1}\|}+\|u^{j-1}\|\right) + \frac{2C}{q_-}\left(\frac{\|u_1^k\|}{\|u^k\|}+\|u^k\|\right) \\
&\le 2\delta + \frac{2C}{q_-}\sum_{j=0}^{k}\left(\frac{\|u_1^j\|}{\|u^j\|}+\|u^j\|\right) \\
&\le 2\delta + \frac{2C}{q_-}\left(\frac{\delta}{1-q}+\frac{C\varepsilon}{q_-(1-q)(1-q_+)}+\sum_{j=0}^{k}\|u^j\|\right) \\
&\le 2\delta + \frac{2C}{q_-}\left(\frac{\delta}{1-q}+\frac{C\varepsilon}{q_-(1-q)(1-q_+)}+\sum_{j=0}^{k}q_+^j\varepsilon\right) \\
&\le 2\delta + \frac{2C}{q_-}\left(\frac{\delta}{1-q}+\frac{C\varepsilon}{q_-(1-q)(1-q_+)}+\frac{\varepsilon}{1-q_+}\right) \\
&\le \bar\delta, \qquad (3.58)
\end{aligned}
$$

where the last inequality is by (3.56).

Combining (3.57) and (3.58), we obtain that $u^{k+1} \in K_{\bar\varepsilon,\bar\delta}(\bar u, \bar v/\|\bar v\|)$, leading to the needed conclusion. ∎

Proposition 3.2 *Let the assumptions of Lemma 3.2 be satisfied.*

*Then, for every $\bar\varepsilon > 0$, $\bar\delta > 0$, $\bar c > 0$, $\sigma \in (0, 3/4)$, and $\theta \in (0, 1)$, there exist $\varepsilon = \varepsilon(\bar v) > 0$, $\delta = \delta(\bar v) > 0$, and $\rho = \rho(\bar v) > 0$, such that for every starting point $u^0 \in K_{\varepsilon,\delta}(\bar u, \bar v/\|\bar v\|)$, Algorithm 2.1** uniquely defines the sequence $\{u^k\}$, $\{u^k\} \subset K_{\bar\varepsilon,\bar\delta}(\bar u, \bar v/\|\bar v\|)$, and* (3.19) *holds for all k large enough.*

Proof. We only need to prove that $\varepsilon > 0$ and $\delta > 0$ can be taken so small that (3.19) holds for all k large enough, as the preceding part of the statement of this proposition literarily repeats the statement of Lemma 3.7.

First select $\bar\varepsilon > 0$, $\bar\delta > 0$, and $\bar\gamma > 0$ according to item (c) in Lemma 3.3. (These $\bar\varepsilon > 0$ and $\bar\delta > 0$ have nothing to do with those in the statement of proposition being proved; the latter were taken arbitrarily.) Then reduce $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$, if necessary, so that the assertion of Proposition 3.1 holds with $\varepsilon = \bar{\varepsilon}$ and $\delta = \bar{\delta}$, and the assertion of Lemma 3.6 holds as well. Finally, for thus defined $\bar{\varepsilon}$ and $\bar{\delta}$, select $\varepsilon > 0$ and $\delta > 0$ according to Lemma 3.7.

With these choices, Lemma 3.7 implies that $\{u^k\} \subset K_{\bar{\varepsilon},\bar{\delta}}(\bar{u}, \bar{v}/\|\bar{v}\|)$. Then:

- According to Lemma 3.6, (3.20) holds for some k.
- Therefore, according to item (c) in Lemma 3.3, (3.19) holds for this k.
- Therefore, according to Proposition 3.1, (3.19) holds for all subsequent k.

■

3.3 Behavior away from the null space

Complementing the analysis in Section 3.2, we now investigate the effect of a single step of our algorithms from starting points close to $\bar{u}$, but in a sense staying away from $\bar{u} + \ker \Phi'(\bar{u})$. The next lemma considers the case when $x^0 - \bar{x} \notin \ker h'(\bar{x})$.

Lemma 3.8 *Let the assumptions of Lemma 3.1 be satisfied.*

Then for any $\rho > 0$, $\bar{c} > 0$, and $\sigma \in (0, 1)$, and any $v = (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ with ξ satisfying (3.4)*, and such that $\xi \notin \ker h'(\bar{x})$, there exists $\tau = \tau(v) > 0$ such that for every $t \in (0, \tau]$ there exists the unique $v^0 = (\xi^0, \eta^0)$ solving* (2.5) *with $k = 0$, H_0 given by* (2.6)*, and $u^0 = (x^0, \lambda^0) = \bar{u} + tv$, and* (2.12) *and* (3.19) *hold for this v^0 with $k = 0$, and with any real c satisfying* (2.11).

Proof. We again assume for simplicity that $\bar{u} = 0$. For $t \in (0, \tau]$ with $\tau > 0$ small enough, the existence and uniqueness of $v^0 = (\xi^0, \eta^0)$ solving (2.5) with $k = 0$ and H_0 given by (2.6) follows by Lemma 3.1, and this v^0 satisfies (3.5) and (3.6) with $\pi(u^k) \in \ker \Phi'(\bar{u})$ depending homogeneously on u^k and satisfying (3.7). Then according to (3.8), the x-component of $\pi(u^k)$ has the form $t\beta(v)\bar{\xi}$ with some real $\beta(v)$, and

$$\xi^0 = -t\xi + \frac{1}{2}t\beta(v)\bar{\xi} + O(t^2) = O(t), \quad \eta^0 = O(t) \tag{3.59}$$

as $t \to 0$.

According to (3.22),

$$\varphi_c'(x^0; \xi^0) \leq -\left\langle \frac{\partial^2 L}{\partial x^2}(x^0, \lambda^0)\xi^0, \xi^0 \right\rangle - \bar{c}\|h(x^0)\|_1, \tag{3.60}$$

where

$$h(x^0) = th'(0)\xi + \frac{1}{2}t^2 h''(0)[\xi, \xi] + O(t^3) \tag{3.61}$$

as $t \to 0$.

Since $\xi \notin \ker h'(0)$, by (3.61) there exists $\gamma > 0$ such that

$$\|h(x^0)\| \geq \gamma t \tag{3.62}$$

for all $t > 0$ small enough, and therefore, by (3.59)–(3.60),

$$\varphi_c'(x^0;\, \xi^0) \leq -\bar{c}\gamma t + O(t^2)$$

as $t \to 0$. Then for every $\theta \in (0,\, \bar{c}\gamma)$ it holds that

$$\varphi_c'(x^0;\, \xi^0) \leq -\theta t \tag{3.63}$$

for all $t > 0$ small enough. Condition $\xi \notin \ker h'(0)$ also implies that ξ and $\bar{\xi}$ are linearly independent, and hence, the first equality in the first relation in (3.59) also yields the estimate $t = O(\|\xi^0\|)$. Therefore, from (3.63) we get (2.12) with $k = 0$ for any fixed $\rho > 0$, and for all $t \in (0,\, \tau]$ provided $\tau > 0$ is taken small enough.

We now proceed with establishing (3.19). From (2.7) and (2.8), similarly to (3.28) and (3.29) we derive

$$\begin{aligned}
\varphi_c(x^0 + \xi^0) - \varphi_c(x^0) - \sigma\varphi_c'(x^0;\, \xi^0) \;\leq\; & \left(\sigma - \frac{1}{2}\right) \left\langle \frac{\partial^2 L}{\partial x^2}(x^0,\, \lambda^0)\xi^0,\, \xi^0 \right\rangle \\
& +(1-\sigma)(\|\lambda^0 + \eta^0\|_\infty - c)\|h(x^0)\|_1 \\
& +c\|h(x^0+\xi^0)\|_1 - \frac{1}{2}\langle \lambda^0,\, h''(x^0)[\xi^0,\, \xi^0]\rangle \\
& +O(\|\xi^0\|^3),
\end{aligned}$$

$$h(x^0 + \xi^0) = \frac{1}{2}h''(x^0)[\xi^0,\, \xi^0] + O(\|\xi^0\|^3).$$

Combining these relations with (3.59) and (3.62) we get

$$\varphi_c(x^0 + \xi^0) - \varphi_c(x^0) - \sigma\varphi_c'(x^0;\, \xi^0) \leq (1-\sigma)(\|\lambda^0 + \eta^0\|_\infty - c)\gamma t + O(ct^2) + O(t^2)$$

as $t \to 0$. Arguing by contradiction, it can be easily verified that the right-hand side of this inequality is negative for all $t > 0$ small enough, and all c satisfying (2.11). (One should employ the second estimate in (3.59), implying that $\lambda^0 + \eta^0$ stays bounded for bounded t, and consider separately the cases when the values of c stay bounded, and when they are unbounded.) ∎

If $h'(\bar{x}) \neq 0$, Lemma 3.8 allows to establish ultimate acceptance of the unit stepsize from an asymptotically dense starlike set of starting points; see Theorem 3.1 below. However, the case of starting points with $x^0 - \bar{x} \in \ker h'(\bar{x})$ is much more problematic, and in fact, these difficulties are related to the phenomenon known as Maratos effect that may show up even in nonsingular cases [11, Section 6.2.2]. We note in passing that the well-established techniques for avoiding the Maratos effect, such as second-order corrections [11, Section 6.2.2], may have no positive effect in the current setting, as the assumptions needed for justification of these techniques are not satisfied. Moreover, Example 3.3 below can be used to demonstrate that second-order corrections may indeed not ensure the acceptance of the unit stepsize in a large domain near the solution in question.

As for the sufficient descent test (2.12), the following additional result is also valid.

Lemma 3.9 *In addition to the assumptions of Lemma 3.1, suppose that the second-order necessary optimality condition for problem* (1.2) *holds at* $\bar{x}$ *with the multiplier* $\bar{\lambda}$, *i.e.,*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle \geq 0 \quad \forall\, \xi \in \ker h'(\bar{x}). \tag{3.64}$$

Then for any $\bar{c} > 0$ *and any* $v = (\xi,\, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ *with* ξ *satisfying* (3.4), *and such that* $v \notin \ker \Phi'(\bar{u})$, *there exists* $\tau = \tau(v) > 0$ *and* $\rho = \rho(v) > 0$ *such that for every* $t \in (0,\, \tau]$ *there exists the unique* $v^0 = (\xi^0,\, \eta^0)$ *solving* (2.5) *with* $k = 0$, H_0 *given by* (2.6), *and* $u^0 = (x^0,\, \lambda^0) = \bar{u} + tv$, *and* (2.12) *holds for this* v^0 *with* $k = 0$, *and with any real* c *satisfying* (2.11).

Proof. Observe that under the assumption (3.2), condition (3.64) implies that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \operatorname{span}\{\bar{\xi}\}. \tag{3.65}$$

We again assume for simplicity that $\bar{u} = 0$. The case when $\xi \notin \ker h'(0)$ was considered in Lemma 3.8.

Assume now that $\xi \in \ker h'(0) \setminus \operatorname{span}\{\bar{\xi}\}$. Then from the first equality in the first relation in (3.59) it follows that $\xi^0 = t\widetilde{\xi} + O(t^2)$, where

$$\widetilde{\xi} = -\xi + \frac{1}{2}\beta(v)\bar{\xi} \in \ker h'(0) \setminus \operatorname{span}\{\bar{\xi}\}.$$

Therefore, from (3.60) and (3.65) we have that there exists $\gamma > 0$ such that

$$\varphi_c'(x^0;\, \xi^0) \leq -\left\langle \frac{\partial^2 L}{\partial x^2}(x^0,\, \lambda^0)\xi^0,\, \xi^0 \right\rangle \leq -t^2 \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)\widetilde{\xi},\, \widetilde{\xi} \right\rangle + O(t^3) \leq -\gamma t^2 + O(t^3)$$

for all $t > 0$ small enough. This and the first relation in (3.59) imply the existence of $\rho > 0$ such that (2.12) holds for all $t \in (0,\, \tau]$ provided $\tau > 0$ is small enough.

It remains to consider the case when $\xi = \theta\bar{\xi}$ with some real θ. Then (3.4) implies (3.10), and in particular, $h''(0)[\bar{\xi},\, \bar{\xi}] \neq 0$. From (3.23) and (3.59)–(3.61) we then have that there exists $\gamma > 0$ such that

$$\begin{aligned}\varphi_c'(x^0;\, \xi^0) &\leq \left(\theta - \frac{1}{2}\beta(v)\right)^2 t^2 \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, 0)\bar{\xi},\, \bar{\xi}\right\rangle - \frac{1}{2}t^2\bar{c}\|h''(0)[\bar{\xi},\, \bar{\xi}]\|_1 + O(t^3)\\ &\leq -\gamma t^2 + O(t^3)\end{aligned}$$

for all $t > 0$ small enough. As above, this implies the needed assertion regarding (2.12).
■

In [14], it has been demonstrated that (2.12) is satisfied with a sufficiently small $\rho > 0$ near multipliers satisfying the second-order sufficient optimality condition

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi\right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{3.66}$$

Lemma 3.9 can be regarded as a generalization of that result under the weaker than (3.66) condition (3.65).

Unfortunately, the full-step condition (3.19) cannot be established even under the assumptions of Lemma 3.9: there may exist large domains away from $\bar{u} + \ker \Phi'(\bar{u})$ where full step is not accepted. This is demonstrated by the next example that is essentially `ralph2` in MacMPEC test set [16], or 20216 in DEGEN [10], but with dropped nonnegativity constraints.

Example 3.3 Problem (1.2) with $n = 2$, $l = 1$,

$$f(x) = \frac{1}{2}(x_1^2 - 4x_1x_2 + x_2^2), \quad h(x) = x_1x_2,$$

has the unique solution $\bar{x} = 0$. We have

$$h'(0) = 0, \quad h''(0)[\xi] = (\xi_2,\, \xi_1),$$

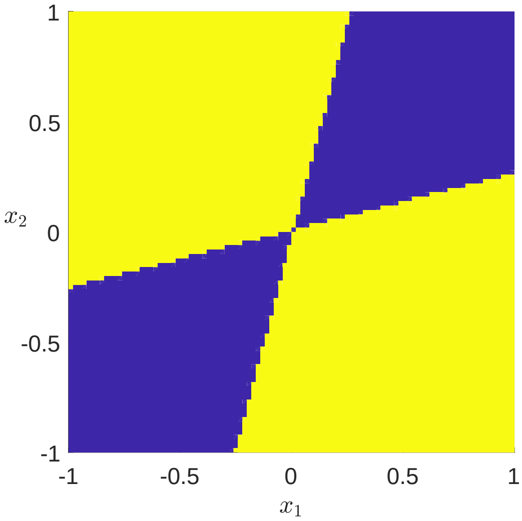
$$\frac{\partial L}{\partial x}(0,\, \lambda) = 0, \quad \frac{\partial^2 L}{\partial x^2}(0,\, \lambda) = \begin{pmatrix} 1 & -2+\lambda \\ -2+\lambda & 1 \end{pmatrix},$$

implying that there are two critical multipliers associated with $\bar{x} = 0$: 1 and 3. To be specific, consider $\bar{\lambda} = 3$. Then

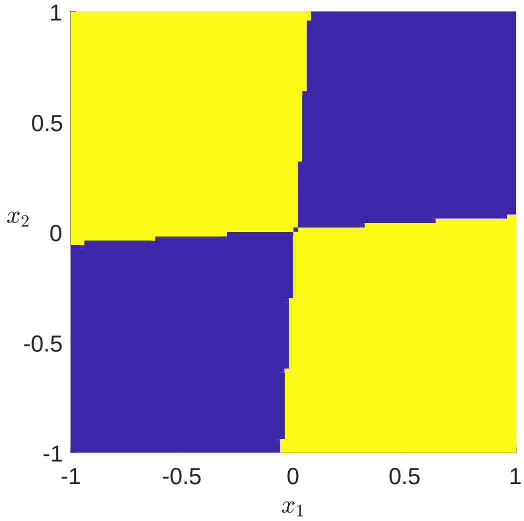
$$\frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda}) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

is positive semidefinite, and hence, (3.64) holds. Furthermore, (3.2) with $Q(0,\, \bar{\lambda})$ defined in (3.1) holds with $\bar{\xi} = (\pm\sqrt{2}/2,\, \mp\sqrt{2}/2)$, and hence,

$$h''(0)[\bar{\xi},\, \xi] = \pm\frac{\sqrt{2}}{2}(\xi_1 - \xi_2),$$



(a) $\sigma = 0.1$



(b) $\sigma = 0.5$

Figure 1: Example 3.3.

implying that (3.4) holds for all $\xi \in \mathbb{R}^2$ with $\xi_1 \neq \xi_2$ (and in particular, (3.10) holds).

System (2.5) with index k dropped has the form

$$x_1+(-2+\lambda)x_2+\xi_1+(-2+\lambda)\xi_2+x_2\eta = 0, \quad (-2+\lambda)x_1+x_2+(-2+\lambda)\xi_1+\xi_2+x_1\eta = 0,$$

$$x_1x_2 + x_2\xi_1 + x_1\xi_2 = 0.$$

Let $\lambda = \bar{\lambda} = 3$, and assume that $x_1 \neq x_2$, as otherwise, the matrix of this system is singular (agreeing with violation of (3.4) for $\xi = x$). Then it can be easily seen that

$$\xi_1 = -\frac{x_1^2}{x_1 - x_2}, \quad \xi_2 = \frac{x_2^2}{x_1 - x_2}, \quad \eta = 0.$$

Substituting these ξ_1 and ξ_2 into the equality in (2.8) (again with index k dropped), we get

$$\varphi_c'(x; \xi) = -((x_1 - x_2)^2 + x_1x_2) - c|x_1x_2|,$$

and hence, (2.12) with $x^k = x$, $\xi^k = \xi$ holds with some $\rho > 0$ provided $c \geq 1$. Furthermore, from (2.7) we get

$$\varphi_c(x + \xi) = (3 + c)\frac{x_1^2x_2^2}{(x_1 - x_2)^2}, \quad \varphi_c(x) = \frac{1}{2}(x_1^2 - 4x_1x_2 + x_2^2) + c|x_1x_2|.$$

Let $x = (t, (1 + \theta)t)$ with real t and θ. It can be directly verified that $\varphi_c(x + \xi) > \varphi_c(x)$ provided θ is close enough to 0, whatever is taken as c, and hence, (3.19) with $x^k = x$, $\xi^k = \xi$ cannot hold, whatever is taken as σ.

At the same time, if $x = (t, -(1+\theta)t)$, and assuming that $\sigma \in (0, 3/4)$, one can see that (3.19) holds provided θ is close enough to 0. This agrees with the theory in Section 3.2.

In Figure 1, we show the domains where the full step is accepted (bright) and not accepted (dark), for $c = \|\bar{\lambda}+\eta\|+1 = 4$. Figures 1a and 1b demonstrate the dependence of these domains on σ.

3.4 Main result

When complemented by Proposition 3.2, the statements in [4, Lemma 7] and Lemma 3.8 lead to the following final result, also employing Proposition 2.1.

Theorem 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be three times differentiable near $\bar{x} \in \mathbb{R}^n$, with their third derivatives Lipschitz-continuous with respect to $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.2)*, with an associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$, and assume that* (3.2) *with $Q(\bar{x}, \bar{\lambda})$ defined in* (3.1)*,* (3.3)*, and* (3.10) *hold with some $\bar{\xi} \in \mathbb{R}^n$, $\|\bar{\xi}\| = 1$. Let $\bar{\eta}$ be the unique in $\operatorname{im} h'(\bar{x})$ solution of the linear system* (3.9)*. Set $\bar{u} = (\bar{x}, \bar{\lambda})$.*

Then for every $\bar{c} > 0$, $\sigma \in (0, 3/4)$, and $\theta \in (0, 1)$, there exist $\rho > 0$ and a set $U \subset \mathbb{R}^n \times \mathbb{R}^l$ starlike with respect to $\bar{u}$, with possibly excluded directions being only those $v = (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ for which $\xi \in \ker h'(\bar{x})$, or (3.4) *is violated, or the x-component of $\pi(v)$ equals* 0 *(with π from Lemma 3.1), and such that for every starting point $u^0 \in U$, Algorithm 2.1** uniquely defines the sequence $\{u^k\}$, and* (3.19) *holds for all k large enough. Moreover, $\{u^k\}$ converges to $\bar{u}$ with the rate of convergence characterized by* (2.2)*, and the sequence $\{(u^k - \bar{u})/\|u^k - \bar{u}\|\}$ converges to $v = (t\bar{\xi}, t\bar{\eta} + \widehat{\eta})$ with some real t and $\widehat{\eta} \in \ker(h'(\bar{x}))^\top$.*

The set of excluded direction specified in this theorem is thin provided $h'(\bar{x}) \neq 0$; see the argument in [4] after Theorem 4.1, with understanding that the thin set of excluded directions specified there must now be complemented by an also thin set $\ker h'(\bar{x})$. If $h'(\bar{x}) = 0$, this theorem is vacuous as one can take $U = \{\bar{u}\}$, and Example 3.3 demonstrates that indeed, an asymptotically dense set U with needed properties may not exist. This theoretical collision can always be formally resolved by introducing an extra variable x_{n+1} and an extra constraint $x_{n+1} = 0$, though a practical impact of such manipulation when using remote starting points is of course doubtful. Anyway, assumption (3.3) allows for $h'(\bar{x}) = 0$ only provided $l = 1$, i.e., in a very special case of a single constraint.

Nevertheless, the issue of reducing the set of excluded directions with $\xi \in \ker h'(\bar{x})$ remains open. Of course, in Example 3.3, we do not claim that the iterative sequences will stay forever in the domain where the full step is not accepted, even if initialized there. However, this issue should be first investigated in the context of the Maratos effect, and under *standard* assumptions (regularity of constraints and SOSC). And anyway, this example shows that the reasoning used in [4] for the case when starting points are away from $\bar{u} + \ker \Phi'(\bar{u})$ does not work here for those starting points whose primal part is not away from $\bar{x} + \ker h'(\bar{x})$.

4 Numerical results

In our numerical experiments we use the following more practical version of the prototype Algorithm 2.1 (we provide here the exact implementation used in our testing).

Algorithm 4.1 Choose $u^0 = (x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ and set $k = 0$, $c_{-1} = 0$. Fix the parameters $\bar{c} > 0$, $\widetilde{c} > 0$, $\rho > 0$ and $\sigma \in (0, 3/4)$, $\theta \in (0, 1)$.

1. Define H_k according to (2.6).
2. Compute $v^k = (\xi^k, \eta^k)$ solving (2.5). If the second equation in (2.5) cannot be solved, stop with a failure. Otherwise, if (2.5) cannot be solved, go to Step 5.
3. Set
$$c_k = \max\left\{c_{k-1}, \frac{4(1-\sigma)\|\lambda^k + \eta^k\|_\infty + \|\lambda^k\|_\infty}{3 - 4\sigma} + \bar{c}\right\}. \tag{4.1}$$
If max in (4.1) is attained at the second argument, replace c_k by $c_k + \widetilde{c}$.
4. If (2.12) with $c = c_k$ is satisfied, go to Step 6.
5. Choose $\tau_k > 0$, replace H_k by $H_k + \tau_k I$, and go to Step 2.
6. Set $\alpha = 1$. If (2.10) with $c = c_k$ holds, set $\alpha_k = \alpha$ and go to Step 7. Otherwise, keep replacing α by $\theta\alpha$ until (2.10) is satisfied.
7. Set $u^{k+1} = u^k + \alpha_k v^k$, increase k by 1, and go to Step 1.

The following parameter values were adopted in our computations: $\bar{c} = \widetilde{c} = 1$, $\rho = 10^{-9}$, $\sigma = 0.01$, $\theta = 0.5$.

In what follows, we abbreviate Algorithm 4.1 as SQP, and we compare it and its version supplied with the simplest extrapolation procedure [5, 7], which will be abbreviated as SQP-EP. Specifically, this algorithm implements the following additional step: if Step 2 is visited for the first time for a given k, set

$$\widetilde{u}^{k+1} = u^k + 2v^k.$$

Therefore, SQP-EP produces, along with $\{u^k\}$, an auxiliary sequence $\{\widetilde{u}^k\}$ obtained by doubling the true Newton step when it exists (for formal correctness, for those k for which $\widetilde{u}^{k+1}$ is not defined because the true Newton step does not exist at u^k, we put

$\widetilde{u}^{k+1} = u^{k+1}$). Accelerating properties of extrapolation rely on the convergence pattern of the true Newton method specified in Proposition 2.1, and hence, the key to success of extrapolation is in preserving this pattern, which evidently requires asymptotic acceptance of the true Hessian, i.e., avoiding Step 5 of Algorithm 4.1, and asymptotic acceptance of the unit stepsize at Step 6 of Algorithm 4.1.

Different rules can be used for controlling τ_k at Step 5 of Algorithm 4.1. In our computations, we adopted the following: for every k, we first try $\tau_k = 1$, and then multiply it by 10 every time modification of H_k is required within the current iteration.

Another approach for dealing with possible lack of positive definiteness of H_k is to define these matrices as quasi-Newton approximations of the Hessian. Here we used BFGS approximations complemented by Powells correction; see [11, Section 4.1]. In the corresponding version of Algorithm 4.1, to be abbreviated as SQP-BFGS, Steps 4 and 5 are dropped.

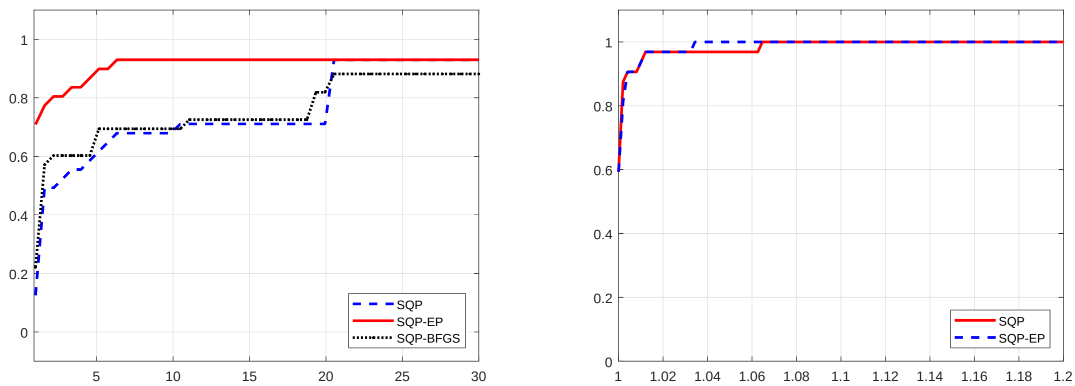
Algorithms SQP and SQP-BFGS terminate, with success declared, if a newly generated iterate u^k satisfies

$$\|\Phi(u^k)\| \leq 10^{-8}. \tag{4.2}$$

As for SQP-EP, for $k = 1, 2, \ldots$ we first compute $\widetilde{u}^k$, and terminate with success if

$$\|\Phi(\widetilde{u}^k)\| \leq 10^{-8};$$

otherwise, we proceed with computing u^k and verifying (4.2). If successful termination did not occur after 200 iterations, or the backtracking procedures in Algorithm 4.1 produced a trial value α such that $\alpha\|v^k\| \leq 10^{-10}$, the process was terminated declaring failure.

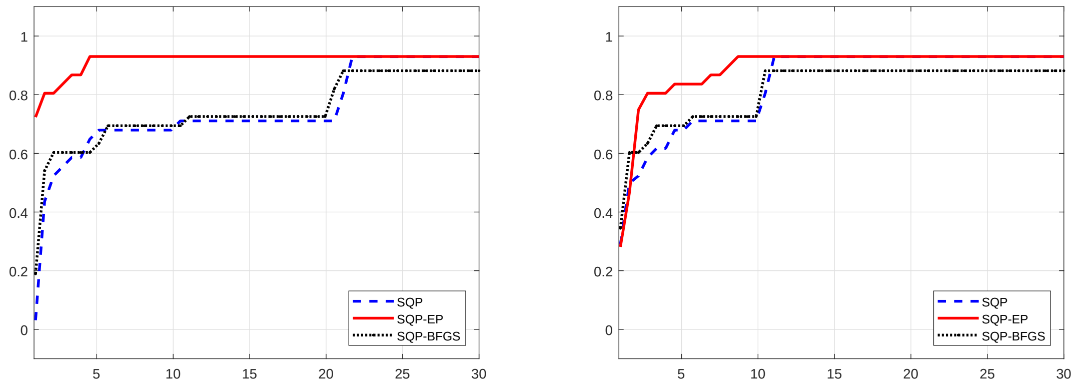


(a) Performance profile (b) Linear systems per iteration

Figure 2: Linear systems solved.

Algorithms were tested on all equality-constrained problems from DEGEN test set [10], except for those with no Lagrange multipliers associated to the solution of interest.

This leaves 32 test problems. For every test problem, all algorithm were initialized at the same 100 starting points generated randomly in the l_∞-ball of radius 100, centered at the primal solution of interest (which is known for each test problem in DEGEN) for the primal part, and at 0 for the dual part.



(a) Evaluations of f

(b) Evaluations of h

Figure 3: Performance profiles on evaluations.

As measures of efficiency, we used the average number of linear systems solved and the average number of evaluations per one successful run. These results are presented in the form of performance profiles that are a modification of the original proposal in [2], for the case of multiple runs for each test problems; see [15] for details.

Figure 2a demonstrates the performance profile for linear systems solved. One can see that SQP-EP is as robust as SQP, more robust than SQP-BFGS, and by far more efficient than both. In particular, this gives an indirect evidence that the true Hessian and the full step are typically asymptotically accepted, and this is indeed the case. In addition, the plots in Figure 2b show which portion of problems required solving no more than a given number of linear systems per iteration on the average (SQP-BFGS does not appear in thus figure as it always solves one linear systems per iteration). The behavior of SQP and SQP-EP in this respect is quite similar, and anyway, this figure demonstrates that iterations requiring solving more than one linear system were actually very few.

The performance profiles for evaluations of f and h are shown in Figures 3a and 3b, respectively. The picture for evaluations of the derivatives of f and h is very similar to that for h (which indicated that the unit stepsize is usually accepted in these experiments), while the number of evaluations of the Hessian in SQP and SQP-EP is the same as the iteration count, and hence, the corresponding picture would be quite similar to the one in Figure 2a. Therefore, we do not provide the corresponding performance profiles.

According to Figure 3a, the relative efficiency of SQP-EP by evaluations of f is more-or-less the same as that by the number of linear systems solved. The picture in Figure 3b by evaluations of h is quite different though. In particular, SQP-EP demonstrated the best result by this measure for about 25% of problems only. This can be explained by the following consideration: if, say, the unit stepsize is accepted at some iteration of SQP, and the corresponding iteration of SQP-EP does not end up with successful termination, the latter requires two evaluations of h, while the former only one. However, for 80% of problems the result of SQP-EP was no more than 2 times worse than the best one, while the performance of the other two algorithms is by far not so good in this respect.

5 Concluding remarks

We have established conditions ensuring that a sequential quadratic programming algorithm with linesearch asymptotically accepts the true Hessian of the Lagrangian and the unit stepsize when converging to a critical Lagrange multiplier of an equality-constrained optimization problem. The paper does not present any essentially new algorithmic techniques, but rather investigates the cases when certain combinations of known techniques can be successfully applied in "nonstandard" circumstances. One potential direction of further development of this material concerns problems with inequality constraints, and apparently, any successful analysis of this kind can be expected for such problems only by some reduction to the equality-constrained case, and by application of the results obtained in this paper. Other issues for possible investigation include possible further insight into the case when the direction from solution to a stating point belongs to the null space of the constraint Jacobian (which apparently would require better understanding of the Maratos effect), as well as the case of Lagrange multipliers critical of order higher than 1.

Acknowledgments

This research was supported in part by the Russian Foundation for Basic Research Grants 19-51-12003 NNIO_a and 20-01-00106_a, and by the Volkswagen Foundation. The author thanks Ivan Rodin for his assistance with numerical experiments.

References

[1] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. Numerical optimization. Theoretical and practical aspects. 2nd edition. Springer-Verlag, Berlin, 2006.

[2] E.D. Dolan and J.J. More. Benchmarking optimization software with performance profiles. Math. Program. 91 (2002), 201–213.

[3] A. Fischer, A.F. Izmailov, and W. Scheck. Adjusting dual iterates in the presence of critical Lagrange multipliers. 2020. SIAM J. Optim. 30 (2020), 1555–1581.

[4] A. Fischer, A.F. Izmailov, and M.V. Solodov. Unit stepsize for the Newton method close to critical solutions. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01496-z.

[5] A. Griewank. Analysis and modification of Mewton's method at singularities. PhD Thesis. Australian National University, Canberra, 1980.

[6] A. Griewank. Starlike domains of convergence for Newton's method at singularities. Numer. Math. 35 (1980), 95–111.

[7] A. Griewank. On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. 27 (1985), 537–563.

[8] 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), 355–379.

[9] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Critical solutions of nonlinear equations: Stability issues. Math. Program. 168 (2018), 475–507.

[10] A.F. Izmailov and M.V. Solodov. DEGEN (2009). http://w3.impa.br/~optim/DEGEN_collection.zip.

[11] A.F. Izmailov and M.V. Solodov. Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, Springer International Publishing, Switzerland, 2014.

[12] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. Math. Program. 133 (2012), 93–120.

[13] 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), 1–26.

[14] A.F. Izmailov and M.V. Solodov. Some new facts about sequential quadratic programming methods employing second derivatives. Optim. Meth. Software 31 (2016), 1111–1131.

[15] A.F. Izmailov, M.V. Solodov, and E.I. Uskov. Combining stabilized SQP with the augmented Lagrangian algorithm Comput. Optim. Appl. 62 (2015), 405–429.

[16] S. Leyffer. MacMPEC (2002).
http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC.

[17] D.-H. Li and L. Qi. Stabilized SQP method via linear equations. Applied Mathematics Technical Reptort AMR00/5. University of New South Wales, Sydney, 2000.

[18] J. Nocedal and S.J. Wright. Numerical Optimization. 2nd edition. Springer-Verlag, New York, Berlin, Heidelberg, 2006.

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