LOCAL CONVERGENCE OF THE METHOD OF MULTIPLIERS FOR VARIATIONAL AND OPTIMIZATION PROBLEMS UNDER THE SOLE NONCRITICALITY ASSUMPTION

A. F. Izmailov[†], A. S. Kurennoy[‡], and M. V. Solodov[§]

August 18, 2013

ABSTRACT

We present local convergence analysis of the method of multipliers for equality-constrained variational problems (in the special case of optimization, also called the augmented Lagrangian method) under the sole assumption that the dual starting point is close to a noncritical Lagrange multiplier (which is weaker than second-order sufficiency). Local superlinear convergence is established under the appropriate control of the penalty parameter values. For optimization problems, we demonstrate in addition local linear convergence for sufficiently large fixed penalty parameters. Both exact and inexact versions of the method are considered. Contributions with respect to previous state-of-the-art analyses for equality-constrained problems consist in the extension to the variational setting, in using the weaker noncriticality assumption instead of the usual second-order sufficient optimality condition, and in relaxing the smoothness requirements on the problem data. In the context of optimization problems, this gives the first local convergence results for the augmented Lagrangian method under the assumptions that do not include any constraint qualifications and are weaker than the second-order sufficient optimality condition. We also show that the analysis under the noncriticality assumption cannot be extended to the case with inequality constraints, unless the strict complementarity condition is added (this, however, still gives a new result).

Key words: variational problem, Karush–Kuhn–Tucker system, augmented Lagrangian, method of multipliers, noncritical Lagrange multiplier, superlinear convergence, generalized Jacobian.
AMS subject classifications. 65K05, 65K15, 90C30.

---

* Research of the first author is supported by the Russian Foundation for Basic Research Grant 12-01-33023. The second author is supported by the Russian Foundation for Basic Research Grants 12-01-31025 and 12-01-33023. The third author is supported in part by CNPq Grant 302637/2011-7, by PRONEX–Optimization, and by FAPERJ.

[†] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: izmaf@ccas.ru

[‡] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: alex-kurennoy@yandex.ru

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

1 Introduction

In this paper we are concerned with local convergence and rate of convergence properties of the augmented Lagrangian (multiplier) methods for optimization, and their extensions to the more general variational context. Augmented Lagrangian methods for optimization date back to [13] and [27]; some other key references are [5, 7, 8, 2]. Methods of this class are the basis for some successful software such as LANCELOT [25] and ALGENCAN [1] (the latter still under continuous development). Their global and local convergence properties remain a subject of active research; some significant theoretical advances rely on novel techniques and are therefore rather recent; see [2, 11, 15, 3, 22, 4], discussions therein, and some comments in the sequel.

Given the mappings $F\colon \mathbb{R}^n \mapsto \mathbb{R}^n$ and $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$, consider the variational problem

$$x \in D, \quad \langle F(x),\, \xi\rangle \geq 0 \quad \forall\, \xi \in T_D(x), \tag{1.1}$$

where

$$D = \{x \in \mathbb{R}^n \mid h(x) = 0\},$$

and $T_D(x)$ is the contingent (tangent in the sense of Bouligand) cone to the feasible set D at $x \in D$ (see, e.g., [9, 30]). Throughout the paper we assume that h is differentiable and that F and h' are Lipschitz-continuous near the point of eventual interest. Associated to (1.1) is solving the primal-dual system

$$F(x) + (h'(x))^{\mathrm{T}}\lambda = 0, \quad h(x) = 0, \tag{1.2}$$

in the variables $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$. In the context of multiplier methods we naturally assume this system to have solutions, which is guaranteed under appropriate constraint qualifications (CQs) [30], but may also be the case regardless of any CQs. No CQs will be assumed in our developments. Any $\lambda \in \mathbb{R}^l$ satisfying (1.2) for some $x = \bar{x}$ will be referred to as a Lagrange multiplier associated with the primal solution $\bar{x}$; the set of all such multipliers will be denoted by $\mathcal{M}(\bar{x})$.

The problem setting (1.1) covers, in particular, the necessary optimality conditions for equality-constrained optimization problems

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{1.3}$$

where $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ is a given function. Specifically, every local solution $\bar{x} \in \mathbb{R}^n$ of the problem (1.3), such that f is smooth near $\bar{x}$, necessarily satisfies (1.1) with the mapping F defined by

$$F(x) = f'(x) \tag{1.4}$$

for all $x \in \mathbb{R}^n$ close enough to $\bar{x}$. We start our discussion with this optimization setting.

Define the Lagrangian $L\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}$ of problem (1.3) by

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

and the augmented Lagrangian $L_\sigma : \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}$ by

$$L_\sigma(x,\,\lambda) = L(x,\,\lambda) + \frac{1}{2\sigma}\|h(x)\|^2,$$

where $\sigma > 0$ is the (inverse of) penalty parameter. Note that in this setting the first equation in (1.2) becomes

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

and (1.2) is then the standard Lagrange optimality system for the optimization problem (1.3).

Given the current estimate $\lambda^k \in \mathbb{R}^l$ of Lagrange multipliers and $\sigma_k > 0$, an iteration of the augmented Lagrangian method applied to (1.3) consists of computing the primal iterate x^{k+1} by solving

$$\begin{array}{ll} \text{minimize} & L_{\sigma_k}(x,\,\lambda^k) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array}$$

perhaps to approximate stationarity only, in the sense that

$$\left\| \frac{\partial L_{\sigma_k}}{\partial x}(x^{k+1},\,\lambda^k) \right\| \le \tau_k \tag{1.5}$$

for some error tolerance $\tau_k \ge 0$, and then updating the multipliers by the explicit formula

$$\lambda^{k+1} = \lambda^k + \frac{1}{\sigma_k} h(x^{k+1}).$$

In the optimization setting, the sharpest known results on local convergence of the augmented Lagrangian method are those in [11] (for problems with twice differentiable data) and in [15] (for problems with Lipschitzian first derivatives). Both these works establish (super)linear convergence (for general equality and inequality constraints) under the sole assumption that the multiplier estimate is close to a multiplier satisfying an appropriate form of second-order sufficient optimality condition. We point out that the earlier convergence rate statements all assumed, in addition, the linear independence CQ (and in the presence of inequality constraints, usually also strict complementarity). Various versions of such statements can be found, e.g., in [5, Prop. 3.2 and 2.7], [26, Thm. 17.6], [28, Thm. 6.16]. It is interesting to mention that in the case of twice differentiable data, the so-called stabilized sequential quadratic programming (sSQP) method, and its counterpart for variational problems, also require second-order sufficiency only [10], with no CQs, just like the augmented Lagrangian method. Moreover, for the special case of equality-constrained optimization, local convergence of sSQP is established under the assumption that the Lagrange multiplier in question is noncritical [21], which is weaker than second-order sufficiency. (We shall recall definitions of all the relevant notions in Section 2 below.) In this paper we show that for the exact and inexact versions of the multiplier method applied to problems with equality constraints, second-order sufficiency can also be relaxed to the noncriticality assumption. In addition, we perform the analysis of the method of multipliers in the more general variational setting, and relax smoothness assumptions on the problem data.

We next state our framework of the method of multipliers for the variational setting of the problem (1.1). Define the mapping $G\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}^n$,

$$G(x,\, \lambda) = F(x) + (h'(x))^{\mathrm{T}}\lambda, \tag{1.6}$$

and consider the following iterative scheme, which we shall refer to as the method of multipliers for solving the variational problem (1.1). If the current primal-dual iterate $(x^k,\, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfies (1.2), stop. Otherwise, choose the inverse penalty parameter $\sigma_k > 0$ and the error tolerance parameter $\tau_k \geq 0$, and compute the next primal-dual iterate $(x^{k+1},\, \lambda^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^l$ as any pair satisfying

$$\left\| G(x^{k+1},\, \lambda^k) + \frac{1}{\sigma_k}(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1}) \right\| \leq \tau_k, \tag{1.7}$$

$$\lambda^{k+1} = \lambda^k + \frac{1}{\sigma_k} h(x^{k+1}). \tag{1.8}$$

Clearly, in the optimization setting, (1.7) corresponds to the usual approximate stationarity condition (1.5) for the augmented Lagrangian $L_{\sigma_k}(\cdot,\, \lambda^k)$. More generally, the iterative scheme given by (1.7), (1.8) also makes sense: the constrained variational problem (1.1) is replaced by solving (approximately) a sequence of unconstrained equations, still in the primal space. A similar variational framework for multiplier methods was used in [3], but in the context of global convergence analysis.

The rest of the paper is organized as follows. In Section 2 we briefly review the abstract iterative framework developed in [12], which is the basis for our convergence analysis. This section also recalls some notions of generalized differentiation and the definition of noncritical Lagrange multipliers. In Section 3, we establish local superlinear convergence of the method of multipliers for equality-constrained variational problems under the sole assumption that the dual starting point is close to a Lagrange multiplier which is noncritical, and provided that the inverse penalty parameter is appropriately managed. For equality-constrained optimization problems, we also prove local linear convergence for sufficiently large fixed penalty parameters. As discussed above, these are the first convergence and rate of convergence results for methods of the type in consideration which employ an assumption weaker than second-order sufficiency and do not require any CQs. The analysis under the noncriticality assumption cannot be extended to the case when inequality constraints are present, as demonstrated in Section 4. However, the assertions hold if the strict complementarity condition is added to noncriticality. This still gives a new result: compared to [11, 15] noncriticality is used instead of second-order sufficiency (though at the price of adding strict complementarity), while compared to the already cited classical results in [5, 26, 28] the linear independence CQ is dispensed with and second-order sufficiency is relaxed to noncriticality (strict complementarity is needed in both approaches). Finally, in Section 5 we compare the obtained results with the related local convergence theory of sSQP, and summarize some remaining open questions. The Appendix contains lemmas concerning nonsingularity of matrices of certain structure, some of independent interest, that are used in our analysis.

Our notation is mostly standard, and would be introduced where needed. Here, we mention that throughout the paper $\|\cdot\|$ is the Euclidean norm, and $B(u,\, \delta)$ is the closed ball of radius

$\delta > 0$ centered at $u \in \mathbb{R}^\nu$. The distance from a point $u \in \mathbb{R}^\nu$ to a set $U \subset \mathbb{R}^\nu$ is defined by

$$\mathrm{dist}(u,\, U) = \inf_{v \in U} \|u - v\|.$$

2 Preliminaries

In this section, we outline the general iterative framework that would be used to derive local convergence of the method of multipliers. We also recall some notions of generalized differentiation, the definition of noncritical Lagrange multipliers, and their relations with second-order sufficiency conditions.

2.1 Noncritical Lagrange multipliers

According to [23, (6.6)], for a mapping $\Psi\colon \mathbb{R}^p \mapsto \mathbb{R}^r$ which is locally Lipschitz-continuous at $u \in \mathbb{R}^p$, the contingent derivative of Ψ at u is the multifunction $C\Psi(u)$ from $\mathbb{R}^p$ to the subsets of $\mathbb{R}^r$, given by

$$C\Psi(u)(v) = \left\{ w \in \mathbb{R}^r \;\middle|\; \exists\, \{t_k\} \subset \mathbb{R}_+,\ \{t_k\} \to 0+ : \left\{ \frac{\Psi(u + t_k v) - \Psi(u)}{t_k} \right\} \to w \right\}.$$

In particular, if Ψ is directionally differentiable at u in the direction v then $C\Psi(u)(v)$ is single-valued and coincides with the directional derivative of Ψ at u in the direction v. The B-differential of $\Psi : \mathbb{R}^p \mapsto \mathbb{R}^r$ at $u \in \mathbb{R}^p$ is the set

$$\partial_B \Psi(u) = \{J \in \mathbb{R}^{r \times p} \mid \exists\, \{u^k\} \subset \mathcal{S}_F \text{ such that } \{u^k\} \to u,\ \{\Psi'(u^k)\} \to J\},$$

where $\mathcal{S}_\Psi$ is the set of points at which Ψ is differentiable. Then the Clarke generalized Jacobian (see [6]) of Ψ at u is given by

$$\partial \Psi(u) = \mathrm{conv}\, \partial_B \Psi(u),$$

where $\mathrm{conv}\, V$ stands for the convex hull of the set V. Observe that according to [23, (6.5), (6.6), (6.16)],

$$\forall\, w \in C\Psi(u)(v) \quad \exists\, J \in \partial\Psi(u) \text{ such that } w = Jv. \tag{2.1}$$

Furthermore, for a mapping $\Psi : \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^r$, the partial contingent derivative (partial Clarke generalized Jacobian) of Ψ at $(u,\, v) \in \mathbb{R}^p \times \mathbb{R}^q$ with respect to u is the contingent derivative (Clarke generalized Jacobian) of the mapping $\Psi(\cdot,\, v)$ at u, which we denote by $C_u\Psi(u,\, v)$ (by $\partial_u \Psi(u,\, v)$).

Let $(\bar{x},\, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ be a solution of the system (1.2). As defined in [16], a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is called *noncritical* if

$$C_x G(\bar{x},\, \bar{\lambda})(\xi) \cap \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} = \emptyset \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{2.2}$$

We shall call $\bar{\lambda}$ a *strongly noncritical* multiplier if

$$\forall\, J \in \partial_x G(\bar{x},\, \bar{\lambda}) \quad \text{it holds that} \quad J\xi \notin \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{2.3}$$

From (2.1) it is evident that the property (2.3) is no weaker than noncriticality (2.2), and in fact it is strictly stronger; see [16, Remark 3]. If the mappings F and h' are differentiable near $\bar{x}$, with their derivatives continuous at $\bar{x}$, then the above two properties become the same, and can be stated as

$$\frac{\partial G}{\partial x}(\bar{x},\,\bar{\lambda})\xi \notin \operatorname{im}(h'(\bar{x}))^{\mathrm{T}} \quad \forall\,\xi \in \ker h'(\bar{x}) \setminus \{0\}, \tag{2.4}$$

which corresponds to the definition of a noncritical multiplier in [14, 18]. We refer the reader to [18, 19, 20, 21, 16] for the role this notion plays in convergence properties of algorithms, stability, error bounds, and other issues.

Here, we emphasize that as can be easily seen, essentially observing that $\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, the strong noncriticality property (2.3) (and hence noncriticality (2.2)) is implied by the second-order condition

$$\forall\, J \in \partial_x G(\bar{x},\,\bar{\lambda}) \quad \text{it holds that} \quad \langle J\xi,\,\xi\rangle > 0 \quad \forall\,\xi \in \ker h'(\bar{x}) \setminus \{0\}, \tag{2.5}$$

but not vice versa. In the optimization setting, i.e., when (1.4) holds, the condition (2.5) is the second-order sufficient optimality condition (SOSC) introduced in [24]. Moreover, for sufficiently smooth problem data, (2.5) is just the usual SOSC for equality-constrained optimization. It should be stressed again, however, that SOSC is much stronger than noncriticality. For example, in the case when f and h are twice differentiable near $\bar{x}$, with their second derivatives continuous at $\bar{x}$, noncritical multipliers, if they exist, form a relatively open and dense subset of the multiplier set $\mathcal{M}(\bar{x})$, which is of course not the case for multipliers satisfying SOSC.

2.2 Fischer's iterative framework

We next recall the abstract iterative framework from [12] for superlinear convergence in case of non-isolated solutions. This framework was designed for generalized equations; here we state its simplification for the usual equations, which is sufficient for our purposes. At the same time, we also make a modification to include the linear rate of convergence in addition to superlinear.

To this end, defining the mapping $\Phi\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}^n \times \mathbb{R}^l$,

$$\Phi(u) = (G(x,\,\lambda),\,h(x)), \tag{2.6}$$

where $u = (x,\,\lambda)$, the system (1.2) can be written in the form

$$\Phi(u) = 0. \tag{2.7}$$

Note also that by (1.6) and (1.7), (1.8), it follows that in the exact case of $\tau_k = 0$, the iterate $u^{k+1} = (x^{k+1},\,\lambda^{k+1})$ of the method of multipliers satisfies the system of equations

$$\Phi_{\sigma_k}(\lambda^k,\,u) = 0, \tag{2.8}$$

where $\Phi_\sigma\colon \mathbb{R}^l \times (\mathbb{R}^n \times \mathbb{R}^l) \mapsto \mathbb{R}^n \times \mathbb{R}^l$ is the family of mappings defined by

$$\Phi_\sigma(\tilde{\lambda},\,u) = (G(x,\,\lambda),\,h(x) - \sigma(\lambda - \tilde{\lambda})), \tag{2.9}$$

with $\sigma \geq 0$. Observe that $\Phi_\sigma(\tilde{\lambda}, \cdot)$ can be regarded as a perturbation of Φ defined in (2.6). Therefore, the iteration subproblem (2.8) of the method of multipliers is a perturbation of the original system (2.7) to be solved.

Consider the class of methods for (2.7) that, given the current iterate $u^k \in \mathbb{R}^\nu$, generate the next iterate $u^{k+1} \in \mathbb{R}^\nu$ as a solution of the subproblem of the form

$$\mathcal{A}(u^k, u) \ni 0, \tag{2.10}$$

where for any $\tilde{u} \in \mathbb{R}^\nu$, the multifinction $\mathcal{A}(\tilde{u}, \cdot)$ from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$ is some kind of approximation of Φ around $\tilde{u}$. For each $\tilde{u} \in \mathbb{R}^\nu$ define the set

$$U(\tilde{u}) = \{u \in \mathbb{R}^\nu \mid \mathcal{A}(\tilde{u}, u) \ni 0\}, \tag{2.11}$$

so that $U(u^k)$ is the solution set of the iteration subproblem (2.10). Of course, without additional (extremely strong) assumptions this set in principle may contain points arbitrarily far from relevant solutions of the original problem (2.7), even for u^k arbitrarily close to those solutions. As usual in local convergence studies, such far away solutions of subproblems must be discarded from the analysis. In other words, it must be specified which of the solutions of (2.10) are allowed to be the next iterate. Specifically, one has to restrict the distance from the current iterate u^k to the next one, i.e., to an element of $U(u^k)$ that can be declared to be u^{k+1} (the so-called localization condition). To this end, define

$$U^c(\tilde{u}) = \{u \in U(\tilde{u}) \mid \|u - \tilde{u}\| \leq c \operatorname{dist}(\tilde{u}, \bar{U})\}, \tag{2.12}$$

where $c > 0$ is arbitrary but fixed, and $\bar{U}$ is the solution set of the equation (2.7). Consider the iterative scheme

$$u^{k+1} \in U^c(u^k), \quad k = 0, 1, \ldots. \tag{2.13}$$

The following statement is essentially [12, Theorem 1], modified to include the case of linear convergence in addition to superlinear. A proof can be obtained by a relatively straightforward modification of that of [12, Theorem 1].

Theorem 2.1 *Let a mapping $\Phi \colon \mathbb{R}^\nu \mapsto \mathbb{R}^\nu$ be continuous in a neighborhood of $\bar{u} \in \mathbb{R}^\nu$. Let $\bar{U}$ be the solution set of the equation (2.7), and let $\bar{u} \in \bar{U}$. Let $\mathcal{A}$ be a set-valued mapping from $\mathbb{R}^\nu \times \mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$. Assume that the following properties hold with some fixed $c > 0$:*

(i) (Upper Lipschitzian behavior of solutions under canonical perturbations) There exists $\ell > 0$ such that for $r \in \mathbb{R}^\nu$, any solution $u(r) \in \mathbb{R}^\nu$ of the perturbed equation

$$\Phi(u) = r,$$

close enough to $\bar{u}$, satisfies the estimate

$$\operatorname{dist}(u(r), \bar{U}) \leq \ell \|r\|.$$

(ii) (Precision of approximation of Φ in subproblems) There exists $\bar{\varepsilon} > 0$ and a function $\omega\colon \mathbb{R}^\nu \times \mathbb{R}^\nu \mapsto \mathbb{R}_+$ such that for

$$q = \ell \sup\{\omega(\tilde{u},\, u) \mid \tilde{u} \in B(\bar{u},\, \bar{\varepsilon}),\ \|u - \tilde{u}\| \le c\,\mathrm{dist}(\tilde{u},\, \bar{U})\}$$

it holds that

$$q < \frac{1}{1+c},$$

and the estimate

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\tilde{u},\, u)\} \le \omega(\tilde{u},\, u)\,\mathrm{dist}(\tilde{u},\, \bar{U})$$

holds for all $(\tilde{u},\, u) \in \mathbb{R}^\nu \times \mathbb{R}^\nu$ satisfying $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$ and $\|u - \tilde{u}\| \le c\,\mathrm{dist}(\tilde{u},\, \bar{U})$.

(iii) (Solvability of subproblems with localization condition) For any $\tilde{u} \in \mathbb{R}^\nu$ close enough to $\bar{u}$ the set $U^c(\tilde{u})$ defined by (2.11), (2.12) is nonempty.

Then for any starting point $u^0 \in \mathbb{R}^\nu$ close enough to $\bar{u}$ there exists a sequence $\{u^k\} \subset \mathbb{R}^\nu$ satisfying (2.13); every such sequence converges to some $u^* \in \bar{U}$, and for all k the following estimates are valid:

$$\|u^{k+1} - u^*\| \le \frac{c\ell\omega(u^k,\, u^{k+1})}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}) \le \frac{cq}{1-q}\,\mathrm{dist}(u^k,\, \bar{U}),$$

$$\mathrm{dist}(u^{k+1},\, \bar{U}) \le \ell\omega(u^k,\, u^{k+1})\,\mathrm{dist}(u^k,\, \bar{U}) \le q\,\mathrm{dist}(u^k,\, \bar{U}).$$

In particular, the rates of convergence of $\{u^k\}$ to u^* and of $\{\mathrm{dist}(u^k,\, \bar{U})\}$ to zero are linear. Moreover, they are superlinear provided that $\omega(u^k,\, u^{k+1}) \to 0$ as $k \to \infty$. In addition, for any $\varepsilon > 0$ it holds that $\|u^* - \bar{u}\| < \varepsilon$ provided u^0 is close enough to $\bar{u}$.

To use the above theorem for the analysis of the multiplier method, we set $\nu = n + l$ and define Φ by (1.6), (2.6). Furthermore, suppose that the inverse penalty parameter σ_k and the tolerance parameter τ_k in (1.7), (1.8) are chosen depending on the current iterate only:

$$\sigma_k = \sigma(x^k,\, \lambda^k), \quad \tau_k = \tau(x^k,\, \lambda^k), \tag{2.14}$$

with some functions $\sigma\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ and $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$. Then the multiplier method can be viewed as a particular case of the iterative scheme (2.13) with $\mathcal{A}$ given by

$$\mathcal{A}(\tilde{u},\, u) = (G(x,\, \lambda) + B(0,\, \tau(\tilde{x},\, \tilde{\lambda})),\, h(x) - \sigma(\tilde{x},\, \tilde{\lambda})(\lambda - \tilde{\lambda})), \tag{2.15}$$

where $\tilde{u} = (\tilde{x},\, \tilde{\lambda})$.

Let $(\bar{x},\, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ be a solution of the system (1.2). It follows from [16, Corollary 1] that assumption (i) of Theorem 2.1 with $\bar{u} = (\bar{x},\, \bar{\lambda})$ is implied by noncriticality of the multiplier $\bar{\lambda}$, i.e., by the property defined in (2.2). Hence, the same implication holds also under the strong noncriticality property defined in (2.3). Moreover [16], noncriticality is equivalent to the error bound

$$\mathrm{dist}((x,\, \lambda),\, \{\bar{x}\} \times \mathcal{M}(\bar{x})) = O(\rho(x,\, \lambda)) \tag{2.16}$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$, where the residual function $\rho\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}$ of the system (1.2) is given by

$$\rho(x, \lambda) = \|(G(x, \lambda), h(x))\|. \tag{2.17}$$

In particular, in this case the solution set $\bar{U}$ of (2.7) locally (near $\bar{u}$) coincides with $\{\bar{x}\} \times \mathcal{M}(\bar{x})$.

Concerning assumption (ii), suppose that the function $\tau(\cdot)$ satisfies

$$\tau(x, \lambda) = o(\mathrm{dist}((x, \lambda), \{\bar{x}\} \times \mathcal{M}(\bar{x}))) \tag{2.18}$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$. In this case, for any function $\sigma(\cdot)$ such that $\sigma(x, \lambda)$ is sufficiently small when (x, λ) is close to $(\bar{x}, \bar{\lambda})$, and for $\mathcal{A}$ defined in (2.15), assumption (ii) of Theorem 2.1 holds, at least when $\bar{U}$ locally coincides with $\{\bar{x}\} \times \mathcal{M}(\bar{x})$. As discussed above, the latter is automatic when $\bar{\lambda}$ is a noncritical multiplier. Note that constructive and practically relevant choices of a function $\tau(\cdot)$ with the needed properties can be based on the residual ρ. Specifically, if

$$\tau(x, \lambda) = o(\rho(x, \lambda))$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$, then (2.18) holds.

As usual (see [10, 21], where this framework of analysis is used in the context of sSQP), the main difficulties are concerned with verification of assumption (iii). This will be the central issue in Section 3, where in particular we have to consider more specific rules for choosing the penalty parameters σ_k.

3 Main results

In this section we prove local convergence of the method of multipliers under the assumption of the dual starting point being close to a noncritical multiplier. For the general case of variational problems, we establish superlinear convergence if the inverse penalty parameter σ_k is controlled in a special way suggested below. Restricting our attention to the case of equality-constrained optimization, we prove in addition local convergence at a linear rate assuming that the inverse penalty parameter is fixed at a sufficiently small value.

Our developments use results concerning nonsingularity of matrices of certain structure, that are collected in the Appendix.

Lemma 3.1 *Let $F\colon \mathbb{R}^n \mapsto \mathbb{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbb{R}^n$, and let $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ be differentiable in some neighbourhood of $\bar{x}$ with its derivative being locally Lipschitz-continuous at $\bar{x}$. Let $\bar{\lambda} \in \mathcal{M}(\bar{x})$ be a strongly noncritical multiplier.*

Then for every $M > 0$ there exists $\gamma > 0$ such that for every sufficiently small $\sigma > 0$, every $\lambda \in \mathbb{R}^l$ close enough to $\bar{\lambda}$, and every $x \in \mathbb{R}^n$ satisfying $\|x - \bar{x}\| \leq \sigma M$, it holds that

$$\forall\, J \in \partial_x G(x, \lambda) \quad \left\| \left(J + \frac{1}{\sigma}(h'(x))^{\mathrm{T}} h'(x) \right) \xi \right\| \geq \gamma \|\xi\| \quad \forall\, \xi \in \mathbb{R}^n,$$

where the mapping $G\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}^n$ is defined according to (1.6).

Proof. Assume the contrary, i.e., that for some $M > 0$ there exist sequences $\{\sigma_k\} \subset \mathbb{R}$ of positive reals, $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$, $\{J_k\}$ of $n \times n$-matrices, and $\{\xi^k\} \subset \mathbb{R}^n$ such that $\|x^k - \bar{x}\| \leq \sigma_k M$ and $J_k \in \partial_x G(x^k, \lambda^k)$ for all k, $\{\sigma_k\} \to 0$, $\{\lambda^k\} \to \bar{\lambda}$, and

$$\left(J_k + \frac{1}{\sigma_k}(h'(\bar{x}) + (h'(x^k) - h'(\bar{x})))^{\mathrm{T}} h'(x^k)\right)\xi^k = \left(J_k + \frac{1}{\sigma_k}(h'(x^k))^{\mathrm{T}} h'(x^k)\right)\xi^k = o(\|\xi^k\|)$$

as $k \to \infty$. Since F and h' are locally Lipschitz-continuous at $\bar{x}$, the mapping $G(\cdot, \lambda^k)$ is locally Lipschitz-continuous at x^k for all sufficiently large k, and moreover, due to the boundedness of $\{\lambda^k\}$, the corresponding Lipschitz constant can be chosen the same for all such k. Since the norms of all matrices in the generalized Jacobian are bounded by the Lipschitz constant of the mapping in question, it then follows that the sequence $\{J_k\}$ is bounded, and therefore, we can assume that it converges to some $n \times n$-matrix J. Then by means of [17, Lemma 2], and by the upper-semicontinuity of the generalized Jacobian, we conclude that $J \in \partial_x G(\bar{x}, \bar{\lambda})$. (For the properties of Clarke's generalized Jacobian see [6].)

Furthermore, due to local Lipschitz-continuity of h' at $\bar{x}$,

$$\|h'(x^k) - h'(\bar{x})\| = O(\|x^k - \bar{x}\|) = O(\sigma_k),$$

implying, in particular, that the sequence $\{(h'(x^k) - h'(\bar{x}))/\sigma_k\}$ is bounded.

A contradiction now follows from Lemma A.1 (in the Appendix) applied with $H = J$, $B = h'(\bar{x})$, $\tilde{H} = J_k$, $\tilde{B} = h'(x^k)$, $\Omega = (h'(x^k) - h'(\bar{x}))$, and $t = 1/\sigma_k$. ■

Lemma 3.1 says, in particular, that if $\lambda \in \mathbb{R}^l$ is close enough to $\bar{\lambda}$, then for any sufficiently small $\sigma > 0$ there exists a neighbourhood of $\bar{x}$ such that

$$\forall\, J \in \partial_x G(x, \lambda) \quad \text{it holds that} \quad \det\left(J + \frac{1}{\sigma}(h'(x))^{\mathrm{T}} h'(x)\right) \neq 0 \tag{3.1}$$

for all x in this neighborhood. The following simple example demonstrates that, generally, this neighbourhood indeed depends on σ.

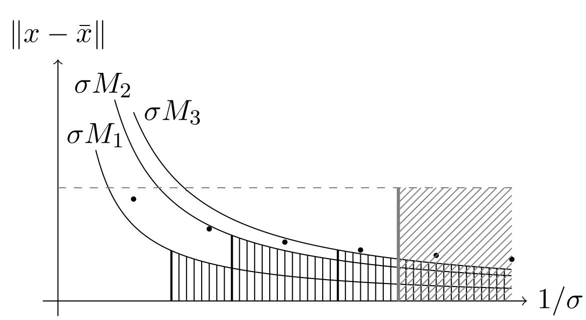

Figure 1: Nonsingularity areas.

Example 3.1 Let $n = l = 1$, $h(x) = x^2/2$, and let $F\colon \mathbb{R} \mapsto \mathbb{R}$ be an arbitrary function differentiable in some neighbourhood of $\bar{x} = 0$, with its derivative being continuous at this point, and such that $F(0) = 0$. Then $\mathcal{M}(0) = \mathbb{R}$, and any $\bar{\lambda} \in \mathcal{M}(0) \setminus \{-F'(0)\}$ is (strongly) noncritical.

Fix any $\bar{\lambda} < -F'(0)$ and an arbitrary sequence $\{x^k\} \subset \mathbb{R}$ convergent to 0, and set $\sigma_k = -(x^k)^2/(F'(x^k) + \bar{\lambda}) > 0$ for all k large enough. Clearly, $\sigma_k \to 0$. However, (3.1) does not hold with $\lambda = \bar{\lambda}$, $\sigma = \sigma_k$, and $x = x^k$ for all k. Therefore, the radius of the neighbourhood in which (3.1) is valid cannot be chosen the same for all sufficiently small $\sigma > 0$ even if $\lambda = \bar{\lambda}$.

In contrast, as can be easily shown by contradiction, if $\bar{\lambda}$ satisfies the SOSC (2.5), then (3.1) holds for all sufficiently small $\sigma > 0$ and for all $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$. The situation is illustrated by Figure 1. The black dots correspond to a sequence of points from Example 3.1. Vertically hatched are the areas of nonsingularity (i.e. where (3.1) holds provided that λ is close enough to $\bar{\lambda}$) given by Lemma 3.1 applied with three different values of M: $M_1 < M_2 < M_3$. Finally, the slope hatching demonstrates the rectangular nonsingularity area that would exist if $\bar{\lambda}$ were to satisfy the SOSC (2.5).

3.1 Superlinear convergence

We first consider the case of the penalty parameter controlled as in (2.14), where the function $\sigma(\cdot) = \sigma_\theta(\cdot)$ is of the form

$$\sigma_\theta(x, \lambda) = (\rho(x, \lambda))^\theta, \tag{3.2}$$

with $\rho(\cdot)$ being the problem residual defined in (2.17), and $\theta \in (0, 1]$ being fixed.

Remark 3.1 For any $\sigma > 0$, any $\tilde{\lambda} \in \mathbb{R}^l$, and any $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ such that F and h' are locally Lipschitz-continuous at x, for the mapping Φ_σ defined in (2.9) it holds that

$$\partial_u \Phi_\sigma(\tilde{\lambda}, u) = \left\{ \left(\begin{array}{cc} J & (h'(x))^{\mathrm{T}} \\ h'(x) & -\sigma I \end{array} \right) \middle| \; J \in \partial_x G(x, \lambda) \right\},$$

where I is the $l \times l$ identity matrix. Indeed, from [17, Lemma 2] it follows that the left-hand side is contained in the right-hand side. The converse inclusion is by the fact that a mapping of two variables, which is differentiable with respect to one variable and affine with respect to the other, is necessarily differentiable with respect to the aggregated variable (cf. [17, Remark 1]).

Making use of Lemma 3.1, we obtain the following

Corollary 3.1 *Under the assumptions of Lemma 3.1, for any $c > 0$ and any $\theta \in (0, 1]$, for the function $\sigma_\theta(\cdot)$ defined in (3.2) it holds that all matrices in $\partial_u \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)$ are nonsingular if $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ is close enough to $(\bar{x}, \bar{\lambda})$, $\tilde{x} \neq \bar{x}$ or/and $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$, and if $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfies*

$$\|(x - \tilde{x}, \lambda - \tilde{\lambda})\| \leq c(\mathrm{dist}((\tilde{x}, \tilde{\lambda}), \{\bar{x}\} \times \mathcal{M}(\bar{x}))). \tag{3.3}$$

Proof. Fix any $c > 0$ and $\theta \in (0, 1]$. According to the error bound (2.16) which is valid under (strong) noncriticality, for all $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ and such that $x \neq \bar{x}$ or/and $\lambda \notin \mathcal{M}(\bar{x})$ it holds that $\rho(\tilde{x}, \tilde{\lambda}) > 0$, and hence, according to (3.2), $\sigma_\theta(\tilde{x}, \tilde{\lambda}) > 0$. Moreover, $\sigma_\theta(\tilde{x}, \tilde{\lambda}) \to 0$ as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$.

Furthermore, employing again the error bound (2.16), for any $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying (3.3) we obtain the estimate

$$\|x - \bar{x}\| \leq \|x - \tilde{x}\| + \|\tilde{x} - \bar{x}\| = O(\rho(\tilde{x}, \tilde{\lambda})) = O\left(\sigma_\theta(\tilde{x}, \tilde{\lambda})(\rho(\tilde{x}, \tilde{\lambda}))^{1-\theta}\right) = O(\sigma_\theta(\tilde{x}, \tilde{\lambda}))$$

as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$. Finally, (3.3) implies that $\lambda \to \bar{\lambda}$ as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$.

Then, applying Lemma 3.1, we conclude that whenever $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ is close enough to $(\bar{x}, \bar{\lambda})$ and $\tilde{x} \neq \bar{x}$ or/and $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$, for any $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying (3.3) the matrix

$$J + \frac{1}{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(h'(x))^{\mathrm{T}} h'(x)$$

is nonsingular for all $J \in \partial_x G(x, \lambda)$. According to Remark 3.1, the latter implies that every matrix in $\partial_u \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)$ has a nonsingular submatrix of the form $-\sigma_\theta(\tilde{x}, \tilde{\lambda})I$ with nonsingular Schur complement, and hence, it is nonsingular (see, e.g., [29, Prop. 3.9]). Therefore, all matrices in $\partial_u \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)$ are nonsingular. ∎

For a given $c > 0$, define the function $\delta_c\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$,

$$\delta_c(x, \lambda) = c(\mathrm{dist}((x, \lambda), \{\bar{x}\} \times \mathcal{M}(\bar{x}))). \tag{3.4}$$

For any $\lambda \in \mathbb{R}^l$, let $\pi(\lambda)$ be the orthogonal projection of λ onto the affine set $\mathcal{M}(\bar{x})$.

Lemma 3.2 *Let $F\colon \mathbb{R}^n \mapsto \mathbb{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbb{R}^n$, and let $h\colon \mathbb{R}^n \mapsto \mathbb{R}^l$ be differentiable in some neighbourhood of $\bar{x}$ with its derivative being locally Lipschitz-continuous at $\bar{x}$. Let $\bar{\lambda} \in \mathcal{M}(\bar{x})$ be a noncritical multiplier.*

Then for any $c > 0$, any $\theta \in (0, 1]$, and any $\gamma \in (0, 1)$, there exists $\varepsilon > 0$ such that for the function $\sigma_\theta(\cdot)$ defined in (3.2) and for all $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying

$$\|(\tilde{x} - \bar{x}, \tilde{\lambda} - \bar{\lambda})\| \leq \varepsilon, \tag{3.5}$$

the inequality

$$\|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u) - \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, (\bar{x}, \pi(\tilde{\lambda})))\| \geq \gamma \sigma_\theta(\tilde{x}, \tilde{\lambda}) \|(x - \bar{x}, \lambda - \pi(\tilde{\lambda}))\| \tag{3.6}$$

holds for all $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying

$$\|(x - \tilde{x}, \lambda - \tilde{\lambda})\| \leq \delta_c(\tilde{x}, \tilde{\lambda}),$$

with $\delta_c(\cdot)$ defined in (3.4).

Proof. Arguing by contradiction, suppose that there exist $c > 0$, $\theta \in (0, 1]$, $\gamma \in (0, 1)$, and sequences $\{(\tilde{x}^k, \tilde{\lambda}^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ and $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ such that $\{(\tilde{x}^k, \tilde{\lambda}^k)\} \to (\bar{x}, \bar{\lambda})$, and for each k it holds that $(x^k, \lambda^k) \in B((\tilde{x}^k, \tilde{\lambda}^k), \delta_k)$ and

$$\|\Phi_{\sigma_k}(\tilde{\lambda}^k, (x^k, \lambda^k)) - \Phi_{\sigma_k}(\tilde{\lambda}^k, (\bar{x}, \pi(\tilde{\lambda}^k)))\| < \gamma\sigma_k \|(x^k - \bar{x}, \lambda^k - \pi(\tilde{\lambda}^k))\|, \tag{3.7}$$

where $\sigma_k = \sigma_\theta(\tilde{x}^k, \tilde{\lambda}^k)$, $\delta_k = \delta_c(\tilde{x}^k, \tilde{\lambda}^k)$.

Set $t_k = \|(x^k - \bar{x}, \lambda^k - \pi(\tilde{\lambda}^k))\|$. The inequality (3.7) implies that $\sigma_k > 0$ and $t_k > 0$. Observe further that $\sigma_k \to 0$ as $k \to \infty$, and according to (1.6) and (2.9), it holds that

$$\|\Phi_{\sigma_k}(\tilde{\lambda}^k, (x^k, \lambda^k)) - \Phi_{\sigma_k}(\tilde{\lambda}^k, (\bar{x}, \pi(\tilde{\lambda}^k)))\| = \left\| \Big(G(x^k, \lambda^k), \, h(x^k) - \sigma_k(\lambda^k - \pi(\tilde{\lambda}^k)) \Big) \right\|.$$

Therefore, (3.7) implies that

$$\|G(x^k, \lambda^k)\| < \gamma\sigma_k t_k = o(t_k) \tag{3.8}$$

and

$$\|h(x^k) - h(\bar{x}) - \sigma_k(\lambda^k - \pi(\tilde{\lambda}^k))\| < \gamma\sigma_k t_k = o(t_k), \tag{3.9}$$

as $k \to \infty$.

Set $\xi^k = (x^k - \bar{x})/t_k$ and $\eta^k = (\lambda^k - \pi(\tilde{\lambda}^k))/t_k$. Without loss of generality, we may assume that the sequence $\{(\xi^k, \eta^k)\}$ converges to some $(\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ such that

$$\|(\xi, \eta)\| = 1. \tag{3.10}$$

From (3.9) it then easily follows that

$$\xi \in \ker h'(\bar{x}). \tag{3.11}$$

Moreover, again taking into account (1.6), we obtain that

$$\begin{aligned} G(x^k, \lambda^k) &= G(x^k, \pi(\tilde{\lambda}^k)) - G(\bar{x}, \pi(\tilde{\lambda}^k)) + G(x^k, \lambda^k) - G(x^k, \pi(\tilde{\lambda}^k)) \\ &= G(x^k, \bar{\lambda}) - G(\bar{x}, \bar{\lambda}) + (h'(x^k) - h'(\bar{x}))^{\mathrm{T}}(\pi(\tilde{\lambda}^k) - \bar{\lambda}) + (h'(x^k))^{\mathrm{T}}(\lambda^k - \pi(\tilde{\lambda}^k)) \\ &= G(\bar{x} + t_k\xi_k, \bar{\lambda}) - G(\bar{x}, \bar{\lambda}) + t_k(h'(x^k))^{\mathrm{T}}\eta^k + o(t_k\|\eta^k\|) \\ &= G(\bar{x} + t_k\xi, \bar{\lambda}) - G(\bar{x}, \bar{\lambda}) + t_k(h'(x^k))^{\mathrm{T}}\eta^k + O(t_k\|\xi^k - \xi\|) + o(t_k\|\eta^k\|), \end{aligned} \tag{3.12}$$

where in the last transition we have taken into account that under our assumptions the mapping $G(\cdot, \bar{\lambda})$ is locally Lipschitz-continuous at $\bar{x}$. Combining (3.8) and (3.12) we derive the existence of $d \in C_xG(\bar{x}, \bar{\lambda})(\xi)$ satisfying

$$d + (h'(\bar{x}))^{\mathrm{T}}\eta = 0. \tag{3.13}$$

Since $\bar{\lambda}$ possesses the property (2.2), relations (3.11) and (3.13) imply that $\xi = 0$, and in particular, $\{\xi^k\} \to 0$.

Furthermore, from (3.9) we have that

$$\eta^k = \frac{1}{\sigma_k t_k}(h(x^k) - h(\bar{x})) + \zeta^k, \tag{3.14}$$

where $\zeta^k \in \mathbb{R}^l$ satisfies $\|\zeta^k\| \leq \gamma$. Note also that since $(x^k,\, \lambda^k) \in B((\tilde{x}^k,\, \tilde{\lambda}^k),\, \delta_k)$, by (3.4) we obtain that

$$\|x^k - \bar{x}\| \leq \|x^k - \tilde{x}^k\| + \|\tilde{x}^k - \bar{x}\| \leq (c+1)(\operatorname{dist}((\tilde{x}^k,\, \tilde{\lambda}^k),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))).$$

Employing again the error bound (2.16), we then obtain the estimate

$$\|x^k - \bar{x}\| = O(\rho(\tilde{x}^k,\, \tilde{\lambda}^k)). \tag{3.15}$$

Let P be the orthogonal projector onto $(\operatorname{im} h'(\bar{x}))^\perp$ in $\mathbb{R}^l$. Applying P to both sides of (3.14), making use of the mean-value theorem and taking into account the assumption that h' is locally Lipschitz-continuous at $\bar{x}$ and (3.2), (3.15), we obtain that

$$\begin{aligned}
\|P\eta^k\| &\leq \frac{1}{\sigma_k} \sup_{\tau \in [0,\,1]} \|P(h'(\bar{x} + \tau(x^k - \bar{x})) - h'(\bar{x}))\| \|\xi^k\| + \|P\zeta^k\| \\
&= \|P\zeta^k\| + O\left(\frac{\|x^k - \bar{x}\|\|\xi^k\|}{\sigma_k}\right) \\
&= \|P\zeta^k\| + O((\rho(\tilde{x}^k,\, \tilde{\lambda}^k))^{1-\theta}\|\xi^k\|).
\end{aligned} \tag{3.16}$$

As $\|\zeta^k\| \leq \gamma$ for all k, passing onto a subsequence if necessary, we can assume that the sequence $\{\zeta^k\}$ converges to some $\zeta \in \mathbb{R}^l$ satisfying $\|\zeta\| \leq \gamma$. Then, since $\{\xi^k\} \to 0$, passing onto the limit in (3.16) yields

$$\|P\eta\| = \|P\zeta\| \leq \|\zeta\| \leq \gamma < 1. \tag{3.17}$$

However, since $\xi = 0$ it follows that $d \in C_x G(\bar{x},\, \bar{\lambda})(0)$, and hence, $d = 0$. Then by (3.13) it holds that $\eta \in \ker(h'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} h'(\bar{x}))^\perp$. Therefore, $P\eta = \eta$ and hence, by (3.17), $\|\eta\| < 1$. Since $\xi = 0$, this contradicts (3.10). ■

We are now in position to prove that the subproblems (2.8) of the exact (and hence, of the inexact) multiplier method have solutions possessing the needed localization properties if the inverse penalty parameter is chosen according to (2.14) with $\sigma(\cdot) = \sigma_\theta(\cdot)$ defined in (3.2), and if the current point is close enough to a solution $(\bar{x},\, \bar{\lambda})$ of the system (1.2), such that $\bar{\lambda}$ is a strongly noncritical multiplier.

For any $\delta \geq 0$, $\sigma \geq 0$, $\tilde{x} \in \mathbb{R}^n$, and $\tilde{\lambda} \in \mathbb{R}^l$, let $U_\delta(\sigma,\, \tilde{x},\, \tilde{\lambda})$ stand for the (evidently nonempty) solution set of the optimization problem in the variable $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ given by

$$\begin{array}{ll} \text{minimize} & \|\Phi_\sigma(\tilde{\lambda},\, u)\|^2 \\ \text{subject to} & \|(x - \tilde{x},\, \lambda - \tilde{\lambda})\| \leq \delta. \end{array} \tag{3.18}$$

Proposition 3.1 *Under the assumptions of Lemma 3.1, for any $c > 3$, any $\theta \in (0, 1]$, and each $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ sufficiently close to $(\bar{x}, \bar{\lambda})$, the equation*

$$\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u) = 0 \tag{3.19}$$

with the function $\sigma_\theta(\cdot)$ defined in (3.2) has a solution $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying (3.3).

Proof. Observe first that if $\tilde{x} = \bar{x}$ and $\tilde{\lambda} \in \mathcal{M}(\bar{x})$, then the needed assertion is evidently valid taking $x = \tilde{x}$ and $\lambda = \tilde{\lambda}$. In the rest of the proof we assume that $\tilde{x} \neq \bar{x}$ or/and $\tilde{\lambda} \notin \mathcal{M}(\bar{x})$.

Fix any $\gamma \in (2/(c-1), 1)$. From Lemma 3.2 it follows that there exists $\varepsilon > 0$ such that for all $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying (3.5), the inequality (3.6) holds for all $u = (x, \lambda) \in U_{\delta_c(\tilde{x}, \tilde{\lambda})}(\sigma_\theta(\tilde{x}, \tilde{\lambda}), \tilde{x}, \tilde{\lambda})$ with $\delta_c(\cdot)$ defined in (3.4). According to Corollary 3.1, reducing ε if necessary we can assure that the set $\partial_u \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)$ does not contain singular matrices. We now show that for all $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying (3.5) with the specified ε, any $u = (x, \lambda) \in U_{\delta_c(\tilde{x}, \tilde{\lambda})}(\sigma_\theta(\tilde{x}, \tilde{\lambda}), \tilde{x}, \tilde{\lambda})$ is a solution of (3.19). From the definition of $U_{\delta_c(\tilde{x}, \tilde{\lambda})}(\sigma_\theta(\tilde{x}, \tilde{\lambda}), \tilde{x}, \tilde{\lambda})$ it then will follow that (3.3) holds as well.

If $\|(x - \tilde{x}, \lambda - \tilde{\lambda})\| = \delta_c(\tilde{x}, \tilde{\lambda})$, then by (3.4) it holds that

$$\begin{aligned} \|(x - \bar{x}, \lambda - \pi(\tilde{\lambda}))\| &\geq \|(x - \tilde{x}, \lambda - \tilde{\lambda})\| - \|(\tilde{x} - \bar{x}, \tilde{\lambda} - \pi(\tilde{\lambda}))\| \\ &= (c-1)\|(\tilde{x} - \bar{x}, \tilde{\lambda} - \pi(\tilde{\lambda}))\| \\ &\geq (c-1)\|\tilde{\lambda} - \pi(\tilde{\lambda})\| \\ &= (c-1)\,\mathrm{dist}(\tilde{\lambda}, \mathcal{M}(\bar{x})). \end{aligned}$$

Employing (3.6), we then derive

$$\begin{aligned} \|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u) - \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, (\bar{x}, \pi(\tilde{\lambda})))\| &\geq \gamma\sigma_\theta(\tilde{x}, \tilde{\lambda})\|(x - \bar{x}, \lambda - \pi(\tilde{\lambda}))\| \\ &\geq \gamma(c-1)\sigma_\theta(\tilde{x}, \tilde{\lambda})\,\mathrm{dist}(\tilde{\lambda}, \mathcal{M}(\bar{x})) \\ &> 2\sigma_\theta(\tilde{x}, \tilde{\lambda})\,\mathrm{dist}(\tilde{\lambda}, \mathcal{M}(\bar{x})), \end{aligned} \tag{3.20}$$

where the choice of γ was taken into account.

On the other hand, by (3.4), $\|(\bar{x} - \tilde{x}, \pi(\tilde{\lambda}) - \tilde{\lambda})\| \leq \delta_c(\tilde{x}, \tilde{\lambda})$, and since u is a solution of the problem (3.18) with $\sigma = \sigma_\theta(\tilde{x}, \tilde{\lambda})$ and $\delta = \delta_c(\tilde{x}, \tilde{\lambda})$, it holds that

$$\begin{aligned} \|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u) - \Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, (\bar{x}, \pi(\tilde{\lambda})))\| &\leq \|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)\| + \|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, (\bar{x}, \pi(\tilde{\lambda})))\| \\ &\leq 2\|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, (\bar{x}, \pi(\tilde{\lambda})))\| \\ &= 2\sigma_\theta(\tilde{x}, \tilde{\lambda})\|\tilde{\lambda} - \pi(\tilde{\lambda})\| \\ &= 2\sigma_\theta(\tilde{x}, \tilde{\lambda})\,\mathrm{dist}(\tilde{\lambda}, \mathcal{M}(\bar{x})), \end{aligned}$$

which contradicts (3.20).

Therefore, $\|(x - \tilde{x}, \lambda - \tilde{\lambda})\| < \delta_c(\tilde{x}, \tilde{\lambda})$, and hence, u is an unconstrained local minimizer of the objective function in (3.18). According to [6, Proposition 2.3.2], this implies that

$$0 \in \partial_u \left(\|\Phi_{\sigma_\theta(\tilde{x}, \tilde{\lambda})}(\tilde{\lambda}, u)\|^2 \right),$$

and according to the chain rule in [6, Theorem 2.6.6], the latter means the existence of $\mathcal{J} \in \partial_u \Phi_{\sigma\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\, u)$ such that

$$\mathcal{J}^{\mathrm{T}} \Phi_{\sigma\theta(\tilde{x},\,\tilde{\lambda})}(\tilde{\lambda},\, u) = 0.$$

By the choice of ε, the matrix $\mathcal{J}$ is nonsingular, and hence, u is a solution of (3.19). ■

Combining Proposition 3.1 with the considerations in Section 2 concerning assumptions (i) and (ii) of Theorem 2.1, we conclude that all the assumptions of Theorem 2.1 are satisfied. This gives the main result of this section.

Theorem 3.1 *Under the assumptions of Lemma 3.1, let* $\tau \colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ *be any function satisfying* (2.18)*.*

Then for any $c > 3$ *and any* $\theta \in (0,\, 1]$*, for any starting point* $(x^0,\, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ *close enough to* $(\bar{x},\, \bar{\lambda})$ *there exists a sequence* $\{(x^k,\, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ *generated by the multiplier method with* σ_k *and* τ_k *computed according to* (2.14)*,* (3.2)*, satisfying*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k)\| \leq c\, \mathrm{dist}((x^k,\, \lambda^k),\, \bar{U}) \tag{3.21}$$

for all k*; every such sequence converges to* $(\bar{x},\, \lambda^*)$ *with some* $\lambda^* \in \mathcal{M}(\bar{x})$*, and the rates of convergence of* $\{(x^k,\, \lambda^k)\}$ *to* $(\bar{x},\, \lambda^*)$ *and of* $\{\mathrm{dist}((x^k,\, \lambda^k),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ *to zero are superlinear. In addition, for any* $\varepsilon > 0$ *it holds that* $\|\lambda^* - \bar{\lambda}\| < \varepsilon$ *provided* $(x^0,\, \lambda^0)$ *is close enough to* $(\bar{x},\, \bar{\lambda})$*.*

3.2 Fixed penalty parameters, optimization case

We now turn our attention to the optimization setting of (1.4), and consider the case when the parameter σ_k is fixed at some value $\sigma > 0$, that is,

$$\sigma(x,\, \lambda) = \sigma \quad \forall\, (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l.$$

The motivation for this additional study of the optimization case is that in computational implementations, boundedness of the penalty parameters is considered important to avoid ill-conditioning in the subproblems of minimizing augmented Lagrangians.

Proposition 3.2 *Let* $f \colon \mathbb{R}^n \mapsto \mathbb{R}^n$ *and* $h \colon \mathbb{R}^n \mapsto \mathbb{R}^l$ *be differentiable in some neighbourhood of* $\bar{x}$ *with their derivatives being locally Lipschitz-continuous at* $\bar{x}$*. Let* $\bar{x}$ *be a solution of the problem* (1.1) *with the mapping* $F \colon \mathbb{R}^n \mapsto \mathbb{R}^n$ *given by* (1.4)*, and let* $\bar{\lambda} \in \mathcal{M}(\bar{x})$ *be a strongly noncritical multiplier.*

Then for any $c > 2$ *it holds that for any sufficiently small* $\sigma > 0$ *there exists a neighbourhood of* $(\bar{x},\, \bar{\lambda})$ *such that if* $(\tilde{x},\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ *belongs to that neighbourhood, the equation*

$$\Phi_\sigma(\tilde{\lambda},\, u) = 0 \tag{3.22}$$

has a solution $u = (x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ *satisfying* (3.3)*.*

Proof. For any $\sigma > 0$ the point $\bar{u} = (\bar{x}, \bar{\lambda})$ is a solution of the equation

$$\Phi_\sigma(\bar{\lambda}, u) = 0.$$

Furthermore, by Remark 3.1 and Lemma 3.1 we conclude that if σ is small enough, then every matrix in the set $\partial_u \Phi_\sigma(\bar{\lambda}, \bar{u})$ has a nonsingular submatrix with nonsingular Schur complement, and therefore, every matrix in $\partial_u \Phi_\sigma(\bar{\lambda}, \bar{u})$ is nonsingular. Then Clarke's inverse function theorem [6, Theorem 7.1.1] guarantees that for any such σ there exist neighbourhoods U_σ of $\bar{u}$ and V_σ of zero such that for every $r \in V_\sigma$ the equation

$$\Phi_\sigma(\bar{\lambda}, u) = r \tag{3.23}$$

has in U_σ the unique solution $u_\sigma(r)$, and the function $u_\sigma(\cdot)\colon V_\sigma \mapsto U_\sigma$ is Lipschitz-continuous with some constant ℓ_σ.

Define $r(\tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ by

$$r(\tilde{\lambda}) = \begin{pmatrix} 0 \\ -\sigma(\tilde{\lambda} - \bar{\lambda}) \end{pmatrix}.$$

If $\tilde{\lambda} \in \mathbb{R}^l$ is close enough to $\bar{\lambda}$, the vector $r(\tilde{\lambda})$ belongs to V_σ, and therefore, the equation (3.23) with $r = r(\tilde{\lambda})$ has in U_σ the unique solution $u_\sigma(r(\tilde{\lambda}))$. Observe that

$$u = u_\sigma(r(\tilde{\lambda})) \tag{3.24}$$

satisfies (3.22). Moreover, since $u_\sigma(0) = \bar{u}$, it holds that

$$\|u - \bar{u}\| = \|u_\sigma(r(\tilde{\lambda})) - u_\sigma(0)\| \le \ell_\sigma \|r(\tilde{\lambda})\| = \ell_\sigma \sigma \|\tilde{\lambda} - \bar{\lambda}\|. \tag{3.25}$$

We now show that for any sufficiently small $\sigma > 0$ there exists a neighbourhood of $(\bar{x}, \bar{\lambda})$ such that for every $\tilde{u} = (\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ from that neighbourhood, u defined by (3.24) satisfies the estimate (3.3). Suppose that this is not the case. Then there exist $c > 2$, $M > 0$, and sequences $\{\sigma_k\}$ of positive reals and $\{\tilde{u}^k\} \subset \mathbb{R}^n \times \mathbb{R}^l$, $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, such that $\sigma_k \to 0$, $\{\tilde{u}^k\} \to \bar{u}$, and for all k it holds that $\ell_{\sigma_k} \|\tilde{u}^k - \bar{u}\| \le M$, and $u^k = u_{\sigma_k}(r(\tilde{\lambda}^k))$ violates (3.3), that is,

$$\|u^k - \tilde{u}^k\| > c \operatorname{dist}(\tilde{u}^k, \{\bar{x}\} \times \mathcal{M}(\bar{x})). \tag{3.26}$$

Observe that (3.25) then implies that for all k

$$\|x^k - \bar{x}\| \le \sigma_k M. \tag{3.27}$$

Furthermore, taking into account that $\Phi_{\sigma_k}(\tilde{\lambda}^k, u^k) = 0$ and $\Phi(\bar{x}, \pi(\tilde{\lambda}^k)) = 0$ (recall that π is a projector onto $\mathcal{M}(\bar{x})$), we can write

$$\Phi_{\sigma_k}(\tilde{\lambda}^k, u^k) - \Phi_{\sigma_k}(\tilde{\lambda}^k, (\bar{x}, \pi(\tilde{\lambda}^k))) = \left(0, -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k))\right).$$

Employing the mean-value theorem (see, e.g., [9, Proposition 7.1.16]) and Remark 3.1, we derive the existence of $u^{k,\,i}$ in the line segment connecting u^k and $(\bar{x},\, \pi(\tilde{\lambda}^k))$, $\alpha_{k,\,i} \geq 0$, and matrices $J_{k,\,i} \in \partial_x G(x^{k,\,i},\, \lambda^{k,\,i})$, $i = 1,\, \ldots,\, n$, such that $\sum_{i=1}^n \alpha_{k,\,i} = 1$ and

$$\begin{pmatrix} \displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, J_{k,\,i} & \left(\displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, h'(x^{k,\,i})\right)^{\mathrm{T}} \\ \displaystyle\sum_{i=1}^n \alpha_{k,\,i}\, h'(x^{k,\,i}) & -\sigma_k I \end{pmatrix} \begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} 0 \\ -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)) \end{pmatrix}$$

for all sufficiently large k. Since $\{\sigma_k\} \to 0$, $\{u^k\} \to \bar{u}$, and (3.27) holds, from Lemma 3.1 it follows that for all sufficiently large k the matrix in the left-hand side of the above equation is nonsingular (as a matrix containing a nonsingular submatrix with nonsingular Schur complement). Then

$$\begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} J_k & B_k^{\mathrm{T}} \\ B_k & -\sigma_k I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ -\sigma_k(\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)) \end{pmatrix},$$

where J_k and B_k stand for $\sum_{i=1}^n \alpha_{ik} J_{ik}$ and $\sum_{i=1}^n \alpha_{ik} h'(x^{ik})$, respectively. Writing the inverse matrix in the above formula in terms of the inverse of the Schur complement of $-\sigma_k I$ (see, e.g., [5, Section 1.2]), we obtain that

$$\begin{pmatrix} x^k - \bar{x} \\ \lambda^k - \pi(\tilde{\lambda}^k) \end{pmatrix} = \begin{pmatrix} \left(J_k + \dfrac{1}{\sigma_k} B_k^{\mathrm{T}} B_k\right)^{-1} B_k^{\mathrm{T}} \\ -I + \dfrac{1}{\sigma_k} B_k \left(J_k + \dfrac{1}{\sigma_k} B_k^{\mathrm{T}} B_k\right)^{-1} B_k^{\mathrm{T}} \end{pmatrix} (\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)). \tag{3.28}$$

For each $i = 1,\, \ldots,\, n$ we have that, since f' and h' are locally Lipschitz-continuous at $\bar{x}$, the mapping $G(\cdot,\, \lambda^{k,\,i})$ is locally Lipschitz-continuous at $x^{k,\,i}$ for all sufficiently large k, and moreover, due to the boundedness of $\{\lambda^{k,\,i}\}$ the corresponding Lipschitz constant can be chosen the same for all such k. Since the norms of all matrices in the generalized Jacobian are bounded by the Lipschitz constant of the mapping in question, it then follows that the sequences $\{J_{k,\,i}\}$ are bounded, and therefore, we can assume that they converge to some $n \times n$-matrices J_i as $k \to \infty$ (passing onto a subsequence if necessary). Then by means of [17, Lemma 2], and by the upper-semicontinuity of the generalized Jacobian, we conclude that $J_i \in \partial_x G(\bar{x},\, \bar{\lambda})$. Furthermore, without loss of generality we may assume that $\alpha_{k,\,i}$ tend to some $\alpha_i \geq 0$. Then $\sum_{i=1}^n \alpha_i = 1$, and setting $J = \sum_{i=1}^n \alpha_i J_i$, we have that $\{J_k\} \to J$ and $J \in \partial_x G(\bar{x},\, \bar{\lambda})$.

Furthermore, due to the fact that h' is locally Lipschitz-continuous at $\bar{x}$, and using also (3.27), we obtain that

$$\begin{aligned}
\|B_k - h'(\bar{x})\| &= \left\| \sum_{i=1}^{n} \alpha_{ik}(h'(x^{k,\,i}) - h'(\bar{x})) \right\| \\
&\leq \sum_{i=1}^{n} \alpha_{k,\,i} \|h'(x^{k,\,i}) - h'(\bar{x})\| \\
&= O\left(\sum_{i=1}^{n} \alpha_{k,\,i} \|x^{k,\,i} - \bar{x}\| \right) \\
&= O\left(\sum_{i=1}^{n} \alpha_{k,\,i} \|x^{k} - \bar{x}\| \right) \\
&= O(\sigma_k)
\end{aligned}$$

as $k \to \infty$. Now, applying Lemma A.2 (in the Appendix) with $H = J$, $B = h'(\bar{x})$, $\tilde{H} = J_k$, $\tilde{B} = B_k$, $\Omega = B_k - h'(\bar{x})$, and $t = 1/\sigma_k$, we conclude that

$$\left\{ \left(J_k + \frac{1}{\sigma_k} B_k^{\mathrm{T}} B_k \right)^{-1} B_k^{\mathrm{T}} \right\} \to 0. \tag{3.29}$$

Finally, note that since (1.4) holds, the matrices J and J_k for all k are symmetric. Then employing Lemma A.3 (in the Appendix) with H, B, $\tilde{H}$, Ω and t set the same as for the application of Lemma A.2, we obtain that

$$\left\| \frac{1}{\sigma_k} B_k \left(J_k + \frac{1}{\sigma_k} B_k^{\mathrm{T}} B_k \right)^{-1} B_k^{\mathrm{T}} \right\| \to 1. \tag{3.30}$$

Combining (3.28) with (3.29) and (3.30), we conclude that for any constant $\tilde{c} > 1$ it holds that

$$\|u^k - (\bar{x},\, \pi(\tilde{\lambda}^k))\| \leq \tilde{c} \|\tilde{\lambda}^k - \pi(\tilde{\lambda}^k)\| \leq \tilde{c} \|\tilde{u}^k - (\bar{x},\, \pi(\tilde{\lambda}^k))\|$$

for all sufficiently large k. Then due to the error bound (2.16), we conclude that for all k large enough

$$\|u^k - \tilde{u}^k\| \leq \|u^k - (\bar{x},\, \pi(\tilde{\lambda}^k))\| + \|\tilde{u}^k - (\bar{x},\, \pi(\tilde{\lambda}^k))\| \leq (1 + \tilde{c})\, \mathrm{dist}((\tilde{x}^k,\, \tilde{\lambda}^k),\, \{\bar{x}\} \times \mathcal{M}(\bar{x})).$$

This gives a contradiction with (3.26), completing the proof. ∎

We mention, in passing, that if f and h are twice differentiable with their second derivatives being continuous at $\bar{x}$, it can be shown that the assertion of Proposition 3.2 holds for any $c > 1$ (instead of $c > 2$).

Assumption (iii) of Theorem 2.1 is therefore verified for the augmented Lagrangian method with fixed penalty parameter, under the stated conditions. Combining this with the discussion in Section 2 of assumptions (i) and (ii), we obtain the following.

Theorem 3.2 *Under the assumptions of Proposition 3.2, let* $\tau\colon \mathbb{R}^n \times \mathbb{R}^l \mapsto \mathbb{R}_+$ *be any function satisfying* (2.18).

Then for any $c > 2$ *there exists* $\bar{\sigma} > 0$ *such that for any* $\sigma \in (0, \bar{\sigma})$ *the following assertion is valid: for every starting point* $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ *close enough to* $(\bar{x}, \bar{\lambda})$ *there exists a sequence* $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ *generated by the multiplier method with* $\sigma_k = \sigma$ *for all* k*, and with* τ_k *computed according to* (2.14)*, satisfying* (3.21) *for all* k*; every such sequence converges to* $(\bar{x}, \lambda^*)$ *with some* $\lambda^* \in \mathcal{M}(\bar{x})$*, and the rates of convergence of* $\{(x^k, \lambda^k)\}$ *to* $(\bar{x}, \lambda^*)$ *and of* $\{\mathrm{dist}((x^k, \lambda^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ *to zero are linear. In addition, for any* $\varepsilon > 0$ *it holds that* $\|\lambda^* - \bar{\lambda}\| < \varepsilon$ *provided* (x^0, λ^0) *is close enough to* $(\bar{x}, \bar{\lambda})$*.*

4 Inequality constraints

In this section we exhibit that the results above under the noncriticality assumption cannot be extended to problems with inequality constraints, even in the optimization case with arbitrarily smooth data. That said, the extension is possible if the strict complementarity condition is added.

Consider the variational problem (1.1) with D defined by

$$D = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \le 0\}, \tag{4.1}$$

where $g\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ and h as specified before. If h and g are smooth, associated to this problem is the primal dual Karush–Kuhn–Tucker (KKT) system

$$F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \quad \mu \ge 0, \quad g(x) \le 0, \quad \langle \mu, g(x)\rangle = 0 \tag{4.2}$$

in the variables $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$. Define $G\colon \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \mapsto \mathbb{R}^n$ by

$$G(x, \lambda, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu.$$

For a solution $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ of the KKT system (4.2) define the index sets

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

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

of active, strongly active and weakly active constraints, respectively.

Using again the terminology introduced in [16], a multiplier $(\bar{\lambda}, \bar{\mu})$ is said to be noncritical if there is no triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, with $\xi \ne 0$, satisfying the system

$$\begin{gathered} d + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \\ \zeta_{A_0} \ge 0, \quad g'_{A_0}(\bar{x})\xi \le 0, \quad \zeta_i\langle g'_i(\bar{x}), \xi\rangle = 0,\ i \in A_0, \quad \zeta_{\{1,\ldots,m\}\setminus A} = 0 \end{gathered} \tag{4.3}$$

with some $d \in C_x G(\bar{x}, \bar{\lambda}, \bar{\mu})(\xi)$. The multiplier $(\bar{\lambda}, \bar{\mu})$ is said to be strongly noncritical if for each matrix $J \in \partial_x G(\bar{x}, \bar{\lambda}, \bar{\mu})$, there is no triple (ξ, η, ζ), with $\xi \ne 0$, satisfying (4.3) with $d = J\xi$. In the case when there are no inequality constraints, these properties are equivalent to their counterparts stated previously; see (2.2) and (2.3). Again, it can be easily verified that (strong) noncriticality is strictly weaker than second-order sufficiency for the problem at hand.

An iteration of the multiplier method for the problem (1.1) with D defined in (4.1) is the following procedure. If the current primal-dual iterate $(x^k,\,\lambda^k,\,\mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfies (4.2), stop. Otherwise, choose the inverse penalty parameter $\sigma_k > 0$ and the tolerance parameter $\tau_k \geq 0$, and compute the next primal-dual iterate $(x^{k+1},\,\lambda^{k+1},\,\mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ as any triple satisfying

$$\left\| F(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}}\left(\lambda^k + \frac{1}{\sigma_k}h(x^{k+1})\right) + (g'(x^{k+1}))^{\mathrm{T}}\max\left\{0,\,\mu^k + \frac{1}{\sigma_k}g(x^{k+1})\right\}\right\| \leq \tau_k, \tag{4.4}$$

$$\lambda^{k+1} = \lambda^k + \frac{1}{\sigma_k}h(x^{k+1}), \quad \mu^{k+1} = \max\left\{0,\,\mu^k + \frac{1}{\sigma_k}g(x^{k+1})\right\}, \tag{4.5}$$

where maximum is taken componentwise. In the optimization case, this is again the usual augmented Lagrangian method with approximate solution of subproblems.

Similarly to the previous interpretations for the equality-constrained case it can be seen that if σ_k and τ_k are computed as functions of the current iterate, the method in question can be embedded into the framework of [12]. Moreover, it satisfies the counterparts of assumptions (i) and (ii) of Theorem 2.1, provided the multiplier in question is noncritical. However, no reasonable counterpart of assumption (iii) holds for the exact multiplier method in the inequality-constrained case (in general). Local solvability of subproblems is not guaranteed by noncriticality, as demonstrated by the following example. The problem in this example is taken from [21, Example 2].

Example 4.1 Let $n = 1$, $l = 0$, $m = 2$, $f(x) = -x^2/2$, $g(x) = (-x,\,x^3/6)$. The corresponding KKT system (4.2) with F defined according to (1.4) has the unique primal solution $\bar{x} = 0$, and the associated multipliers are $\mu \in \mathbb{R}^2$ such that $\mu_1 = 0$, $\mu_2 \geq 0$. The multiplier $\bar{\mu} = 0$ is noncritical.

For a current dual iterate $\mu^k = \tilde{\mu}$ and for $\sigma_k = \sigma > 0$, from (4.4), (4.5) it can be seen that the next iterate $(x^{k+1},\,\mu^{k+1})$ must satisfy the system

$$\begin{gathered} -x - \mu_1 + \frac{1}{2}x^2\mu_2 = 0, \\ \mu_1 \geq 0, \quad x + \sigma(\mu_1 - \tilde{\mu}_1) \geq 0, \quad \mu_1\,(x + \sigma(\mu_1 - \tilde{\mu}_1)) = 0, \\ \mu_2 \geq 0, \quad \frac{1}{6}x^3 - \sigma(\mu_2 - \tilde{\mu}_2) \leq 0, \quad \mu_2\left(\frac{1}{6}x^3 - \sigma(\mu_2 - \tilde{\mu}_2)\right) = 0. \end{gathered} \tag{4.6}$$

Let $\tilde{\mu}_1 > 0$ and $\tilde{\mu}_2 = 0$.

1. If $\mu_1 = \mu_2 = 0$ then from the first relation in (4.6) it follows that $x = 0$. But then the second line in (4.6) contradicts the assumption $\tilde{\mu}_1 > 0$.

2. If $\mu_1 > 0$, $\mu_2 = 0$, then the first relation in (4.6) implies that $x = -\mu_1$, and therefore, by the second line in (4.6),
$$-\mu_1(1 - \sigma) = \sigma\tilde{\mu}_1 > 0,$$
which can not be true if $\sigma \leq 1$.

3. If $\mu_1 = 0$, $\mu_2 > 0$, then by the first relation in (4.6) either $x = 0$ or $x\mu_2 = 2$. In the former case, the second line in (4.6) yields $-\sigma\tilde{\mu}_1 \geq 0$, which again contradicts the assumption $\tilde{\mu}_1 > 0$. On the other hand, the latter case is not possible whenever (x, μ) is close enough to $(\bar{x}, \bar{\mu})$.

4. Finally, if $\mu_1 > 0$ and $\mu_2 > 0$, then by the third line in (4.6), $\mu_2 = x^3/(6\sigma)$, and hence $x > 0$. Moreover from the first line in (4.6) it follows that $\mu_1 = -x + x^2\mu_2/2$, and therefore, $\mu_1 \leq -x + x^2/2 < 0$ whenever $\mu_2 \leq 1$ and $x \in (0, 2)$.

Therefore, in every neighbourhood of $\bar{\mu}$ there exists a point $\tilde{\mu}$ such that the system (4.6) does not have solutions in some fixed neighbourhood of $(\bar{x}, \bar{\lambda})$ if $\sigma > 0$ is small enough. Consequently, assumption (iii) of Theorem 2.1 cannot hold with any $c > 0$ for the exact multiplier method, neither if the penalty parameter σ_k is chosen in such a way that it tends to zero as the current primal-dual iterate tends to $(\bar{x}, \bar{\mu})$, nor if it is fixed at a sufficiently small value.

Observe, however, that in Example 4.1 the strict complementarity condition is violated: $\bar{\mu} = 0$, and in fact, all Lagrange multipliers have a zero component corresponding to an active constraint. Assuming the strict complementarity condition $\bar{\mu}_A > 0$, the phenomenon exhibited in Example 4.1 would not be possible, as we discuss next.

Under the strict complementarity assumption, the KKT system (4.2) reduces locally (near $(\bar{x}, \bar{\lambda}, \bar{\mu})$) to the system of equations

$$F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'_A(x))^{\mathrm{T}}\mu_A = 0, \quad h(x) = 0, \quad g_A(x) = 0, \tag{4.7}$$

with the additional equation $\mu_{\{1,\ldots,m\}\setminus A} = 0$. This primal-dual system corresponds to the equality-constrained variational problem (1.1) with

$$D = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g_A(x) = 0\}. \tag{4.8}$$

Observe that under strict complementarity, the multiplier $(\bar{\lambda}, \bar{\mu})$ is (strongly) noncritical for the original problem (1.1), (4.1) if, and only if, the multiplier $(\bar{\lambda}, \bar{\mu}_A)$ associated to the primal solution $\bar{x}$ of the system (4.7) is (strongly) noncritical.

Furthermore, under strict complementarity, iteration (4.4), (4.5) equipped with a reasonable localization condition not allowing $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ to be too far from (x^k, λ^k, μ^k), evidently subsumes that if the latter is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and if $\sigma_k > 0$ is small enough, then

$$\mu_A^{k+1} = \mu_A^k + \frac{1}{\sigma_k} g_A(x^{k+1}), \quad \mu_{\{1,\ldots,m\}\setminus A}^{k+1} = 0.$$

This means that the multiplier method for the general problem (1.1), (4.1) locally reduces to the multiplier method for the equality-constrained problem (1.1), (4.8). Employing this reduction, Theorems 3.1 and 3.2 can be extended to the case when inequality constraints are present, assuming strict complementarity. A formal exposition of this development and formal convergence statements would require certain technicalities, which we prefer to omit here.

We finish with recalling that even adding strict complementarity to noncriticality of the multiplier still gives new and meaningful results, as discussed in Section 1.

5 Concluding remarks and open questions

As mentioned in Section 1, contemporary local convergence theory of the augmented Lagrangian methods is closely related to that of sSQP. This is actually not surprising, as the two methods are indeed related: in some sense, sSQP can be regarded as a "linearization" of the exact augmented Lagrangian method. That said, there are also some subtle but remarkable differences in the results currently available. We highlight these next. For simplicity, we shall refer to the equality-constrained optimization problem (1.3) only.

Under the SOSC (2.5) and without any CQs, the (apparent) "ideal" for local convergence of the augmented Lagrangian methods for optimization was achieved in [11, 15]: the linear rate if the inverse penalty parameter σ_k is small enough, becoming superlinear if it tends to zero (in an arbitrary way!). For sSQP, Example 3.1 can be used to show that σ_k (for sSQP it plays the role of a stabilization parameter) cannot be driven to zero arbitrarily fast: if done so, iteration subproblems may have no solutions satisfying the localization condition of the kind (3.3). In other words, the augmented Lagrangian subproblem may possess some "good" solutions whose "counterparts" are missing for the sSQP subproblem.

The same Example 3.1 can be used to show that assuming noncriticality instead of SOSC, the iteration subproblems of sSQP may have *no solutions at all* if σ_k is driven to zero too fast, and that for a fixed sufficiently small value of this parameter, the neighborhood of appropriate starting points can be shrinking as this value tends to zero. We are not aware of such examples for the augmented Lagrangian methods. In particular, it is an open question whether the local superlinear convergence result of Theorem 3.1 remains valid if the inverse penalty parameter is driven to zero in an arbitrary way, and whether the local linear convergence result of Theorem 3.2 actually requires the neighborhood of appropriate starting points to be dependent on the fixed inverse penalty parameter value.

Finally, unlike for the augmented Lagrangian methods, all the existing results for sSQP assume twice differentiability of the problem data, and attempts to relax smoothness were not successful so far. The reason is that under the weaker smoothness hypotheses (Lipschitz-continuity of the first derivatives, for example), assumption (ii) of Theorem 2.1 (precision of approximation) cannot be established for sSQP. Possible relaxations for this assumption that might do the job are not clear.

Appendix

This appendix contains lemmas concerning nonsingularity of matrices of certain structure, used in the analysis above. The first one is a refined version of [21, Lemma 1].

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

$$H\xi \notin \operatorname{im} B^{\mathrm{T}} \quad \forall\, \xi \in \ker B \setminus \{0\}. \tag{a.1}$$

Then for any $M > 0$ there exists $\gamma > 0$ such that

$$\left\| \left(\tilde{H} + t(B + \Omega)^{\mathrm{T}} \tilde{B} \right) \xi \right\| \geq \gamma \|\xi\| \quad \forall\, \xi \in \mathbb{R}^n$$

for every $n \times n$-matrix $\tilde{H}$ close enough to H, every $l \times n$-matrix $\tilde{B}$ close enough to B, every $t \in \mathbb{R}$ such that $|t|$ is sufficiently large, and for every $l \times n$-matrix Ω satisfying $\|\Omega\| \leq M/|t|$.

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

$$H_k\xi^k + t_k(B + \Omega_k)^{\mathrm{T}} B_k\xi^k = o(\|\xi^k\|) \tag{a.2}$$

as $k \to \infty$. Without loss of generality we may assume that $\|\xi^k\| = 1$ for all k and that $\{\xi^k\} \to \xi \neq 0$. Then (a.2) means the existence of a sequence $\{w^k\} \subset \mathbb{R}^n$ such that $\{w_k\} \to 0$ and

$$H_k\xi^k + t_k(B + \Omega_k)^{\mathrm{T}} B_k\xi^k = w_k \tag{a.3}$$

for all k. Therefore, it must hold that $B^{\mathrm{T}}B\xi = 0$, since

$$B^{\mathrm{T}} B_k\xi^k = -\frac{1}{t_k} H_k\xi^k - \Omega_k^{\mathrm{T}} B_k\xi^k + \frac{1}{t_k} w^k$$

tends to 0 as $k \to \infty$. Consequently, $\xi \in \ker B$.

On the other hand, (a.3) implies that

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

for all k, where the second term in the left-hand side tends to zero as $k \to \infty$ because $\{t_k\Omega_k\}$ is bounded and $\{B_k\xi^k\} \to B\xi = 0$. Hence, $H\xi \in \operatorname{im} B^{\mathrm{T}}$ by the closedness of $\operatorname{im} B^{\mathrm{T}}$. This completes a contradiction with (a.1). ■

Lemma A.2 *Under the assumptions of Lemma A.1, for any $M > 0$ and any $\varepsilon > 0$ it holds that for every $n \times n$-matrix $\tilde{H}$ close enough to H, every $l \times n$-matrix $\tilde{B}$ close enough to B, every real t such that $|t|$ is sufficiently large, and for all $l \times n$-matrices Ω satisfying $\|\Omega\| \leq M/|t|$, the matrix $\tilde{H} + t(B + \Omega)^{\mathrm{T}}\tilde{B}$ is nonsingular and*

$$\left\| \left(\tilde{H} + t(B + \Omega)^{\mathrm{T}} \tilde{B} \right)^{-1} (B + \Omega)^{\mathrm{T}} \right\| \leq \varepsilon. \tag{a.4}$$

Proof. Fix arbitrary $M > 0$ and $\varepsilon > 0$. The assertion regarding nonsingularity of $\tilde{H} + t(B + \Omega)^{\mathrm{T}}\tilde{B}$ follows directly from Lemma A.1. Therefore, we only have to prove that (possibly by making $\tilde{H}$ closer to H, $\tilde{B}$ closer to B, and $|t|$ larger) one can additionally ensure (a.4).

By contradiction, suppose first that there exist sequences $\{H_k\}$ of $n \times n$-matrices, $\{B_k\}$ and $\{\Omega_k\}$ of $l \times n$-matrices, $\{t_k\}$ of reals, and $\{\eta^k\} \subset \mathbb{R}^n$, such that $\{H_k\} \to H$, $\{B_k\} \to B$, $|t_k| \to \infty$, $\|\Omega_k\| \leq M/|t_k|$, $\|\eta^k\| = 1$ and $\det(H_k + t_k(B + \Omega_k)^{\mathrm{T}} B_k) \neq 0$ for all k, and for

$$\xi^k = (H_k + t_k(B + \Omega_k)^{\mathrm{T}} B_k)^{-1} (B + \Omega_k)^{\mathrm{T}} \eta^k \tag{a.5}$$

it holds that

$$\|\xi^k\| > \varepsilon \tag{a.6}$$

for all k. By (a.5) we have that

$$(B+\Omega_k)^{\mathrm{T}}\eta^k = H_k\xi^k + t_k(B+\Omega_k)^{\mathrm{T}}B_k\xi^k. \tag{a.7}$$

Due to (a.6), the sequence $\{\eta^k/\|\xi^k\|\}$ is bounded. Without loss of generality we may assume that the sequence $\{\xi^k/\|\xi^k\|\}$ converges to some $\xi \in \mathrm{I\!R}^n$ such that $\|\xi\| = 1$. Then dividing both sides of (a.7) by $t_k\|\xi^k\|$ and passing onto the limit as $k \to \infty$, we obtain that $B^{\mathrm{T}}B\xi = 0$, and hence, $\xi \in \ker B$.

Furthermore, by (a.7), it holds that

$$H_k\frac{\xi^k}{\|\xi^k\|} - \Omega_k^{\mathrm{T}}\frac{\eta^k}{\|\xi^k\|} + t_k\Omega_k^{\mathrm{T}}B_k\frac{\xi^k}{\|\xi^k\|} = \frac{1}{\|\xi^k\|}B^{\mathrm{T}}(\eta^k - t_kB_k\xi^k) \in \operatorname{im} B^{\mathrm{T}}$$

for all k. The second term in the left-hand side tends to zero because $\{\|\Omega_k\|\} \to 0$ while the sequence $\{\eta^k/\|\xi^k\|\}$ is bounded. Moreover, the third term in the left-hand side tends to zero as well, because $\{t_k\Omega_k\}$ is bounded while $\{B_k\xi^k/\|\xi^k\|\} \to B\xi = 0$. Therefore, by closedness of $\operatorname{im} B^{\mathrm{T}}$, it follows that $H\xi \in \operatorname{im} B^{\mathrm{T}}$, which contradicts (a.1). ■

Lemma A.3 *In addition to the assumptions of Lemma A.1, let H be symmetric.*

Then for any $M > 0$ and any $\varepsilon > 0$ it holds that for every symmetric $n \times n$-matrix $\tilde{H}$ close enough to H, every real t such that $|t|$ is sufficiently large, and for all $l \times n$-matrices Ω satisfying $\|\Omega\| \le M/|t|$, the matrix $\tilde{H} + t(B+\Omega)^{\mathrm{T}}(B+\Omega)$ is nonsingular and the following estimate is valid

$$\left\| t(B+\Omega)\left(\tilde{H} + t(B+\Omega)^{\mathrm{T}}(B+\Omega)\right)^{-1}(B+\Omega)^{\mathrm{T}} \right\| \le 1+\varepsilon. \tag{a.8}$$

Proof. Again, nonsingularity of $\tilde{H}+t(B+\Omega)^{\mathrm{T}}(B+\Omega)$ is given by Lemma A.1. If at the same time the estimate (a.8) does not hold, there must exist sequences $\{H_k\}$ of symmetric $n \times n$-matrices, $\{\Omega_k\}$ of $l \times n$-matrices, $\{t_k\}$ of reals, and $\{\eta^k\} \subset \mathrm{I\!R}^n$, such that $\{H_k\} \to H$, $|t_k| \to \infty$, and for all k it holds that $\|\Omega_k\| \le M/|t_k|$, $\|\eta^k\| = 1$, $\det(H_k + t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)) \ne 0$, and

$$\left\| t_k(B+\Omega_k)\left(H_k + t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1}(B+\Omega_k)^{\mathrm{T}}\eta^k \right\| > 1+\varepsilon. \tag{a.9}$$

For each k set

$$\begin{aligned} W_k &= (B+\Omega_k)\left(H_k + t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1} \\ &= \left(\left(H_k + t_k(B+\Omega_k)^{\mathrm{T}}(B+\Omega_k)\right)^{-1}(B+\Omega_k)^{\mathrm{T}}\right)^{\mathrm{T}}, \end{aligned} \tag{a.10}$$

where the symmetry of H_k was taken into account. Due to Lemma A.2 we have that $\{W_k\} \to 0$.

Furthermore, for each k the vector η^k can be decomposed into the sum

$$\eta^k = \eta_1^k + \eta_2^k,$$

where $\eta_1^k \in \ker B^{\mathrm{T}} = (\operatorname{im} B)^{\perp}$ and $\eta_2^k \in \operatorname{im} B$. Observe that $t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta_1^k = W_k (t_k \Omega_k^{\mathrm{T}}) \eta_1^k$, and since the sequences $\{\eta_1^k\}$ and $\{t_k \Omega_k\}$ are bounded, and $\{W_k\} \to 0$, we conclude that $\{t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta_1^k\} \to 0$. On the other hand, as $\eta_2^k \in \operatorname{im} B$, there exists $\xi_2^k \in \mathbb{R}^n$ such that $B\xi_2^k = \eta_2^k$ and the sequence $\{\xi_2^k\}$ is bounded. Therefore, employing (a.10),

$$\begin{aligned}
\left\| t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta_2^k \right\| &= \left\| W_k (t_k (B+\Omega_k)^{\mathrm{T}}) B \xi_2^k \right\| \\
&\leq \left\| W_k (H_k + t_k (B+\Omega_k)^{\mathrm{T}} (B+\Omega_k)) \xi_2^k \right\| \\
&\quad + \left\| W_k (H_k + t_k (B+\Omega_k)^{\mathrm{T}} \Omega_k) \xi_2^k \right\| \\
&= \left\| (B+\Omega_k) \xi_2^k \right\| + \left\| W_k (H_k + t_k (B+\Omega_k)^{\mathrm{T}} \Omega_k) \xi_2^k \right\| \\
&\leq \|\eta_2^k\| + \|\Omega_k \xi_2^k\| + \left\| W_k (H_k + t_k (B+\Omega_k)^{\mathrm{T}} \Omega_k) \xi_2^k \right\| \\
&\leq 1 + \|\Omega_k \xi_2^k\| + \left\| W_k (H_k + t_k (B+\Omega_k)^{\mathrm{T}} \Omega_k) \xi_2^k \right\|.
\end{aligned}$$

The last two terms in the right-hand side tend to zero because the sequences $\{\xi_2^k\}$ and $\{H_k + t_k (B+\Omega_k)^{\mathrm{T}} \Omega_k\}$ are bounded, while $\{\Omega_k\} \to 0$ and $\{W_k\} \to 0$. Therefore,

$$\limsup_{k\to\infty} \left\| t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta^k \right\| \leq \lim_{k\to\infty} \left\| t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta_1^k \right\| + \limsup_{k\to\infty} \left\| t_k W_k (B+\Omega_k)^{\mathrm{T}} \eta_2^k \right\| \leq 1,$$

which contradicts (a.9). ■

References

[1] ALGENCAN. `http://www.ime.usp.br/egbirgin/tango/`.

[2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On augmented Lagrangian methods with general lower-level constrints. SIAM J. Optim. 18 (2007), 1286–1309.

[3] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification Math. Program. 111 (2008), 5–32.

[4] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. 135 (2012), 255–273.

[5] D.P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, USA, 1982.

[6] F.H. Clarke. Optimization and Nonsmooth Analysis. John Wiley & Sons, New York, USA, 1983.

[7] A.R. Conn, N. Gould, and P.L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. 28 (1991), 545–572.

[8] A.R. Conn, N. Gould, A. Sartenaer, and P.L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM J. Optim. 6 (1996), 674–703.

[9] F. Facchinei and J.-S. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer-Verlag, New York, USA, 2003.

[10] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. 125 (2010), 47–73.

[11] D. Fernández and M.V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. 22 (2012), 384–407.

[12] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. 94 (2002), 91–124.

[13] M.R. Hestenes. Multiplier and gradient methods. J. Optim. Theory and Appl. 4 (1969), 303–320.

[14] A.F. Izmailov. On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. 45 (2005), 930–946.

[15] A.F. Izmailov and A.S. Kurennoy. Abstract Newtonian frameworks and their applications. 2012. Available at Optimization Online, `http://www.optimization-online.org/DB_HTML/2013/02/3760.html`.

[16] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. DOI 10.1007/s10107-012-0586-z.

[17] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization. Set-Valued Variational Anal. 21 (2013), 17-45.

[18] A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. 117 (2009), 271–304.

[19] A.F. Izmailov and M.V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comp. Optim. Appl. 42 (2009), 231–264.

[20] A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. 126 (2011), 231–257.

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

[22] A.F. Izmailov, M.V. Solodov, and E.I. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. 22 (2012), 1579–1606.

[23] D. Klatte and B. Kummer. Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. Kluwer Academic Publishers, Dordrecht, 2002.

[24] D. Klatte and K. Tammer. On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems. Optimization 19 (1988), 169–180.

[25] LANCELOT. `http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml`.

[26] J. Nocedal and S.J. Wright. Numerical Optimization. Springer–Verlag, New-York, 1999.

[27] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In Optimization, R. Fletcher, ed., Academic Press, New York, 1969, 283–298.

[28] A. Ruszczynski. Nonlinear Optimization. Princeton University Press, 2006.

[29] D. Serre. Matrices: Theory and Applications. Second Edition. Springer, New-York, 2010.

[30] M.V. Solodov. Constraint qualifications. In Wiley Encyclopedia of Operations Research and Management Science, James J. Cochran, et al. (editors), John Wiley & Sons, Inc., 2010.