JOTA manuscript No.
(will be inserted by the editor)

Perturbed Augmented Lagrangian Method Framework with Applications to Proximal and Smoothed Variants

A.F. Izmailov · M.V. Solodov

Received: January 26, 2021 / Accepted: date

Abstract We introduce a perturbed augmented Lagrangian methods framework, which is a convenient tool for local analyses of convergence and rates of convergence of some modifications of the classical augmented Lagrangian algorithm. One example to which our development applies is the proximal augmented Lagrangian method. Previous results for this version required twice differentiability of the problem data, the linear independence constraint qualification, strict complementarity, and second-order sufficiency; or the linear independence constraint qualification and strong second-order sufficiency. We obtain a set of convergence properties under significantly weaker assumptions: once (not twice) differentiability of the problem data, uniqueness of the Lagrange multiplier and second-order sufficiency (no linear independence constraint qualification and no strict complementarity); or even second-order sufficiency only. Another version to which the general framework applies, is the smoothed augmented Lagrangian method, where the plus-function associated to penalization of inequality constraints is approximated by a family of smooth functions (so that the subproblems are twice differentiable if the problem data is). Furthermore, for all the modifications, inexact solution of subproblems

Communicated by Boris S. Mordukhovich.

Dedicated to Professor Franco Giannessi on the occasion of his 85th birthday, and in appreciation of his longstanding selfless service for the success of JOTA.

A.F. Izmailov, Corresponding author
Lomonosov Moscow State University, MSU
Uchebniy Korpus 2, VMK Faculty, OR Department
Leninskiye Gory, 119991 Moscow, Russia
izmaf@ccas.ru

M.V. Solodov
IMPA – Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina 110, Jardim Botânico
Rio de Janeiro, RJ 22460-320, Brazil
solodov@impa.br

is handled naturally. The presented framework also subsumes the basic augmented Lagrangian method, both exact and inexact.

Keywords Augmented Lagrangian · Proximal method of multipliers · Smoothing · Linear convergence · Superlinear convergence · Strong metric regularity · Semistability · Upper-Lipschitz stability · Second-order sufficient optimality conditions

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

1 Introduction

We consider the constrained optimization problem

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

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable near a solution in question, but not necessarily twice differentiable.

Let $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ be the usual Lagrangian of problem (1), i.e.,

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

The classical augmented Lagrangian $L_c : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$, parameterized by the penalty parameter $c > 0$, is given by

$$L_c(x,\,\lambda,\,\mu) = f(x) + \frac{1}{2c}(\|\lambda + ch(x)\|^2 + \|\max\{0,\,\mu + cg(x)\}\|^2), \tag{2}$$

where the max-operation is applied component-wise. Throughout the paper, the inner product and the norm are always Euclidian.

Given the current dual iterate $(\lambda^k,\,\mu^k) \in \mathbb{R}^l \times \mathbb{R}^m$, the penalty parameter $c_k > 0$, and the error-tolerance parameter $\tau_k \ge 0$, the basic (inexact for $\tau_k > 0$) augmented Lagrangian method [14, 25, 26, 1, 2], known also as the method of multipliers, generates the next primal-dual iterate as follows. The next primal iterate $x^{k+1} \in \mathbb{R}^n$ is a τ_k-stationary point of the unconstrained optimization problem

$$\text{minimize } L_{c_k}(x,\,\lambda^k,\,\mu^k), \quad x \in \mathbb{R}^n, \tag{3}$$

that is, x^{k+1} satisfies

$$\left\| \frac{\partial L_{c_k}}{\partial x}(x^{k+1},\,\lambda^k,\,\mu^k) \right\| \le \tau_k. \tag{4}$$

Then, the next dual iterate $(\lambda^{k+1},\,\mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^m$ is given by

$$\lambda^{k+1} = \lambda^k + c_k h(x^{k+1}), \quad \mu^{k+1} = \max\{0,\,\mu^k + c_k g(x^{k+1})\}. \tag{5}$$

To put in perspective contributions of the current paper, we first give a brief survey concerning local convergence and rates of convergence results for the basic augmented Lagrangian scheme outlined above. For results on global convergence properties, see [4] and references therein. Global convergence is not at issue in this paper. Classical results affirming that the described iterations (with $\tau_k = 0$) are locally well-defined and converge at a linear rate, required the linear independence constraint qualification, strict complementarity and the second-order sufficient optimality condition; see [2, Propositions 3.2 and 2.7], [24, Theorem 17.6], [8], [29, Theorem 6.16]. Strict complementarity is not assumed in [15,9], but a stronger version of second-order sufficiency is employed. The linear independence constraint qualification was needed always. Linear convergence of the dual sequence was established in the quotient sense, but for the primal sequence only in the weaker root sense.

Improved local convergence theories for the augmented Lagrangian algorithm have been developed relatively recently, in [13,16,19,21]. In particular, the assumptions required are now much weaker, while the assertions are stronger. Specifically, no constraint qualifications are needed at all, and the only assumption is the usual (not strong) second-order sufficient optimality condition. In the case when there are no inequality constraints, even the yet weaker assumption of noncriticality of Lagrange multiplier [20, Definition 1.41] is enough [19]. Moreover, linear convergence in the quotient sense is established not only for the dual sequence, but for the primal as well. Some recent extensions along those lines beyond the usual nonlinear programming problems like (1), such as cone-constrained optimization, can be found in [6,10,22].

All the results mentioned above concern the basic augmented Lagrangian scheme (4), (5). In this paper, we aim at extending these theories to some relevant modifications.

One of those modifications is the proximal augmented Lagrangian method of [27], which is also classical. In this method, (3) is replaced by

$$\text{minimize } L_{c_k}(x,\, \lambda^k,\, \mu^k) + \frac{1}{2c_k}\|x - x^k\|^2, \quad x \in \mathbb{R}^n, \tag{6}$$

and accordingly, (4) is replaced by

$$\left\| \frac{\partial L_{c_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) + \frac{1}{c_k}(x^{k+1} - x^k) \right\| \leq \tau_k. \tag{7}$$

It is worth to point out that any extension of linear convergence results known for the basic augmented Lagrangian method to its proximal variant is certainly not obvious or automatic, since the primal step given by (6) is shorter than for (3). The available literature does show linear convergence of the proximal version (see [30,31]), but under the same strong assumptions as were traditionally employed for the basic method: the linear independence constraint qualification, strict complementarity and the second-order sufficient optimality condition (or the linear independence constraint qualification and strong second-order sufficiency). In this paper, we give the complete picture under three separate sets of assumptions coming from [16], among which are much weaker ones (the strict Mangasarian–Fromovitz condition and usual second-order sufficiency, or even second-order sufficiency alone). In addition, we consider the case when the problem data in (1) may be once but not twice differentiable, as well as inexact solution of subproblems.

To perform the analysis, we introduce a perturbed augmented Lagrangian methods framework, conceptually related to the perturbed sequential quadratic programming framework [20, Chapter 4.3]. The latter proved to be useful for analyzing, in a unified manner, a good number of popular algorithms for constrained optimization.

In addition to the proximal variant, another modification that we show to fit our perturbed augmented Lagrangian methods framework is given by smoothing the max-operation (the "plus-function") in (2). The motivation for this comes from the fact that even when the problem data in (1) is twice differentiable, the objective function in (2) is not. This precludes applying second-order (Newtonian) methods for solving subproblems (2). And even if quasi-Newton methods are used, superlinear convergence in solving the subproblems cannot be guaranteed, in general. For this reason, different (using other than quadratic penalty functions) augmented Lagrangians, which are twice differentiable, have been proposed in the literature. This leads to the exponential and modified barrier methods of multipliers; see [3, Chapter 4.2.5] and references therein. Here, motivated by the same consideration of making the subproblems twice differentiable if the data is so, we consider smoothing (in the sense of [7], [12, Chapter 11.8]) the plus-function $\max\{0, \cdot\}$ in (2) by a family of functions $s_\theta : \mathbb{R} \to \mathbb{R}$, where $\theta > 0$ is a smoothing parameter. Properties of the smoothing needed for our analysis will be specified in the sequel, as well as some examples satisfying these properties. Substituting (componentwise) $s_\theta(\cdot)$ for $\max\{0, \cdot\}$ in (2), we obtain the family of smoothed augmented Lagrangians $L_{c,\,\theta} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$,

$$L_{c,\,\theta}(x,\, \lambda,\, \mu) = f(x) + \frac{1}{2c}(\|\lambda + ch(x)\|^2 + \|s_\theta(\mu + cg(x))\|^2). \tag{8}$$

The corresponding modification of the basic augmented Lagrangian method is then obtained by replacing (4) with

$$\left\| \frac{\partial L_{c_k,\,\theta_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) \right\| \leq \tau_k. \tag{9}$$

The dual iterates are updated as before, by (5). Of course, the sequence $\{\theta_k\}$ of the smoothing parameters and the error tolerances $\{\tau_k\}$ must be controlled in an appropriate way, to be stated in due course.

As already mentioned, to obtain comprehensive local convergence theories for the specified algorithms, we first develop a general perturbation framework for the basic augmented Lagrangian method, which makes use of parametric abstract Newtonian schemes of [16]. Thus, we start in Section 2 with adapting to the current setting some results from [16]. Our perturbation framework for the augmented Lagrangian method, allowing not only for inexactness in solving subproblems, but also "structural" perturbations to the subproblems, is given in Section 3. This framework is then applied to the proximal augmented Lagrangian method in Section 4, and to the method with smoothing in Section 5.

We finish the current section with description of some notation. Throughout the paper, the distance from a point $u \in \mathbb{R}^p$ to a set $S \subset \mathbb{R}^p$ is defined as

$$\operatorname{dist}(u,\, S) = \inf_{v \in S} \|u - v\|,$$

and $B(u,\, \varepsilon)$ stands for the closed ball centered at u, of radius $\varepsilon > 0$.

If $S \subset \mathbb{R}^p$ is convex, the normal cone to S at a point $u \in \mathbb{R}^p$ is given by

$$N_S(u) = \begin{cases} \{w \in \mathbb{R}^p \mid \langle w, v - u\rangle \leq 0 \;\forall\, v \in S\} & \text{if } u \in S, \\ \emptyset & \text{otherwise.} \end{cases}$$

When well-defined, the B-differential of a mapping $F : \mathbb{R}^p \to \mathbb{R}^q$ at $u \in \mathbb{R}^p$ is the set

$$\partial_B F(u) = \{J \in \mathbb{R}^{q \times p} \mid \exists\, \{u^k\} \subset \mathcal{S}_F :\, \{u^k\} \to u,\; \{F'(u^k)\} \to J\},$$

where $\mathcal{S}_F$ is the set of points at which F is differentiable. Then the Clarke generalized Jacobian of F at u is given by

$$\partial F(u) = \operatorname{conv} \partial_B F(u),$$

where $\operatorname{conv} S$ stands for the convex hull of a set S.

Furthermore, for a mapping $F : \mathbb{R}^p \times \mathbb{R}^r \to \mathbb{R}^q$, the partial generalized Jacobian $\partial_u F(u,\, v)$ of F at $(u,\, v) \in \mathbb{R}^p \times \mathbb{R}^r$ with respect to u is the generalized Jacobian of the mapping $F(\cdot,\, v)$ at u.

For an index set $J \subset \{1,\, \ldots,\, m\}$, we write $|J|$ for its cardinality, and define the set $\setminus J$ by $\setminus J = \{1,\, \ldots,\, m\} \setminus J$. Furthermore, z_J stands for the subvector of $z \in \mathbb{R}^m$, with components z_i, $i \in J$, while by $\mathbb{R}^J$ we denote the linear space spanned by coordinates in J.

2 Preliminaries

Stationary points and associated Lagrange multipliers of problem (1) are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0, \qquad (10)$$

with respect to primal variables $x \in \mathbb{R}^n$ and dual variables $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$.

Denote $U = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ and $Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$. The KKT system (10) is a mixed complementarity problem, which can be written as the generalized equation (GE)

$$\Phi(u) + N_Q(u) \ni 0, \qquad (11)$$

where $u = (x, \lambda, \mu) \in U$, the mapping $\Phi : U \to U$ is given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), -g(x) \right), \tag{12}$$

and $N_Q(u)$ is the normal cone to $Q \subset U$ at $u \in U$. We shall refer to (11) as GE-KKT.

The abstract parametric Newtonian framework for GE in [16], adapted to the current context, can be described as follows. Given a current iterate $u^k \in U$, and choosing a parameter value $c_k > 0$, the next iterate u^{k+1} is obtained as a solution of the subproblem

$$\mathcal{A}(c_k, u^k, u) + N_Q(u) \ni 0, \tag{13}$$

where for any $c > 0$ and any $\widetilde{u} \in U$, the (generally) set-valued mapping $\mathcal{A}(c, \widetilde{u}, \cdot)$ from U to the subsets of U is some kind of approximation of Φ around $\widetilde{u}$. The results in [16] provide a toolkit for local convergence analysis of the specified framework under three different sets of hypotheses. Naturally, weaker hypotheses lead to weaker (but still fully reasonable) convergence properties. We next discuss certain adaptations and simplifications of the results from [16] for the purposes of this work.

The strongest set of hypotheses in [16] relies on the assumption of strong metric regularity [11, Section 3.7] of a solution, and the approximation mapping $\mathcal{A}$ must be single-valued. Below is a version of [16, Theorem 2.1], which is, on the one hand, simplified to be convenient for tackling the augmented Lagrangian methods specifically, while on the other hand, its assumptions are in some respect weaker than those in [16, Theorem 2.1]. To explain those differences, we note the following. Re-examining the proof in [16, Theorem 2.1], one can see that assumption (ii)(a) therein can be replaced by (15) below, which makes the entire set of assumptions weaker. Also, (16) below is a somehow cruder counterpart of assumption (ii)(b) in [16, Theorem 2.1], resulting in less detailed convergence rate estimates. Finally, the asymptotic nature of (15) allows not to pre-fix any set of appropriate parameter values (Π in [16]), but rather to define it in the assertion as the set of all sufficiently large c, and to obtain superlinear convergence rate when $c \to \infty$.

Proposition 2.1 *Let $\bar{u} \in U$ be a strongly metrically regular solution of the GE-KKT* (11)*, i.e., for any $w \in U$ close enough to* 0*, the perturbed problem*

$$\Phi(u) + N_Q(u) \ni w \tag{14}$$

has near $\bar{u}$ the unique solution $u(w)$, and the mapping $u(\cdot)$ is Lipschitz-continuous near 0 *with constant $\ell > 0$. Let $\mathcal{A} : (\mathbb{R}_+ \setminus \{0\}) \times U \times U \to U$ satisfy the following assumptions characterizing precision of approximation:*

$$\Phi(u) - \mathcal{A}(c, \widetilde{u}, u) = o(\|u - \widetilde{u}\| + \|\widetilde{u} - \bar{u}\|) \tag{15}$$

as $c \to +\infty$ and $\widetilde{u}, u \in U$ tend to $\bar{u}$, and there exists $\varkappa > 0$ such that $\ell\varkappa < 1$ and

$$\|(\Phi(u^1) - \mathcal{A}(c, \widetilde{u}, u^1)) - (\Phi(u^2) - \mathcal{A}(c, \widetilde{u}, u^2))\| \le \varkappa \|u^1 - u^2\| \tag{16}$$

for all c large enough and all $\widetilde{u}$, u^1, $u^2 \in U$ close enough to $\bar{u}$.

Then there exists $\bar{\delta} > 0$ such that for any $\delta \in (0, \bar{\delta}]$, there exist $\bar{c} = \bar{c}(\delta) \geq 0$ and $\bar{\varepsilon} = \bar{\varepsilon}(\delta) > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, any $\varepsilon \in (0, \bar{\varepsilon}]$, and any starting point $u^0 \in B(\bar{u}, \varepsilon)$, there exists the unique sequence $\{u^k\} \subset U$ such that u^{k+1} solves (13) *and*

$$\|u^{k+1} - u^k\| \leq \delta \tag{17}$$

for all $k = 0, 1, \ldots$; this sequence is contained in $B(\bar{u}, \varepsilon)$ and converges to $\bar{u}$ at a linear rate, and moreover, this rate is superlinear if $c_k \to +\infty$ as $k \to \infty$, and quadratic if in addition

$$\Phi(u^{k+1}) - \mathcal{A}(c_k, u^k, u^{k+1}) = O(\|u^k - \bar{u}\|\|u^{k+1} - u^k\|)$$

as $k \to \infty$.

The next result, which is an adapted version of [16, Theorem 3.1], relies on the weaker (than strong metric regularity) assumption of semistability ([5], [20, Definition 1.29]) of a solution, and it already allows for a set-valued approximation mapping $\mathcal{A}$. But, as in the case of strong metric regularity, semistability still subsumes that the solution in question must be isolated. On the other hand, generated sequence converging to this solution need not be unique, unlike was the case for strong metric regularity in Proposition 2.1.

Proposition 2.2 *Let $\bar{u} \in U$ be a semistable solution of the GE-KKT* (11), *i.e., for any $w \in U$, any solution $u(w)$ of the perturbed problem* (14), *close enough to $\bar{u}$, satisfies the estimate*

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

as $w \to 0$. Let mapping $\mathcal{A}$ from $(\mathbb{R}_+ \setminus \{0\}) \times U \times U$ to the subsets of U satisfy the following assumption characterizing precision of approximation:

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(c, \widetilde{u}, u)\} = o(\|u - \widetilde{u}\| + \|\widetilde{u} - \bar{u}\|) \tag{18}$$

as $c \to +\infty$ and $\widetilde{u}$, $u \in U$ tend to $\bar{u}$. Assume finally that for any $\varepsilon > 0$, the subproblem

$$\mathcal{A}(c, \widetilde{u}, u) + N_Q(u) \ni 0, \tag{19}$$

has a solution $u(c, \widetilde{u})$ satisfying $\|u(c, \widetilde{u}) - \bar{u}\| \leq \varepsilon$, for all c large enough and all $\widetilde{u} \in U$ close enough to $\bar{u}$.

Then there exists $\bar{\delta} > 0$ such that for any $\delta \in (0, \bar{\delta}]$, there exist $\bar{c} = \bar{c}(\delta) \geq 0$ and $\bar{\varepsilon} = \bar{\varepsilon}(\delta) > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, any $\varepsilon \in (0, \bar{\varepsilon}]$, and any starting point $u^0 \in B(\bar{u}, \varepsilon)$, there exists a sequence $\{u^k\} \subset U$ such that u^{k+1} solves (13) *and satisfies* (17) *for all $k = 0, 1, \ldots$; any such sequence $\{u^k\}$ is contained in $B(\bar{u}, \varepsilon)$ and converges to $\bar{u}$ at a linear rate, and moreover, this rate is superlinear if $c_k \to +\infty$ as $k \to \infty$, and quadratic if in addition*

$$\sup\{\|w\| \mid w \in \Phi(u^{k+1}) - \mathcal{A}(c_k, u^k, u^{k+1})\} = O(\|u^k - \bar{u}\|\|u^{k+1} - u^k\|) + O(\|u^k - \bar{u}\|^2)$$

as $k \to \infty$.

We complete this section with the following adaptation of the result in [16, Theorem 4.1]. This result employs an even weaker regularity requirement than semistability, in particular allowing for nonisolated solutions (which was not the case in the two propositions above).

Proposition 2.3 *Let U^* be the solution set of the GE-KKT* (11)*, and let $\bar{u} \in U^*$ satisfy the following upper-Lipschiz stability condition: for any $w \in U$, any solution $u(w)$ of the perturbed problem* (14)*, close enough to $\bar{u}$, satisfies the estimate*

$$\operatorname{dist}(u(w),\, U^*) = O(\|w\|) \tag{20}$$

as $w \to 0$. Let $\sigma > 0$ be fixed. Let mapping $\mathcal{A}$ from $(\mathbb{R}_+ \setminus \{0\}) \times U \times U$ to the subsets of U satisfy the following assumption characterizing precision of approximation:

$$\sup \left\{ \|w\| \;\middle|\; \begin{array}{l} w \in \Phi(u) - \mathcal{A}(c,\, \widetilde{u},\, u), \\ u \in U,\ \|u - \widetilde{u}\| \leq \sigma \operatorname{dist}(\widetilde{u},\, U^*) \end{array} \right\} = o(\operatorname{dist}(\widetilde{u},\, U^*)) \tag{21}$$

as $c \to +\infty$ and $\widetilde{u} \in U$ tends to $\bar{u}$. Assume finally that for all c large enough and all $\widetilde{u} \in U$ close enough to $\bar{u}$, the subproblem (19) *has a solution $u(c,\, \widetilde{u})$ satisfying*

$$\|u(c,\, \widetilde{u}) - \widetilde{u}\| \leq \sigma \operatorname{dist}(\widetilde{u},\, U^*). \tag{22}$$

Then for any $\varepsilon > 0$, there exists $\bar{c} \geq 0$ such that for any sequence $\{c_k\} \subset [\bar{c},\, +\infty)$ and any starting point $u^0 \in U$ close enough to $\bar{u}$, there exists a sequence $\{u^k\} \subset U$ such that u^{k+1} solves (13) *and satisfies*

$$\|u^{k+1} - u^k\| \leq \sigma \operatorname{dist}(u^k,\, U^*)$$

for all $k = 0,\, 1,\, \ldots$; any such sequence $\{u^k\}$ is contained in $B(\bar{u},\, \varepsilon)$ and converges to some $u^* \in U^*$, and the rates of convergence of $\{u^k\}$ to u^* and of $\operatorname{dist}(u^k,\, U^*)$ to 0 are linear, and moreover, both these rates are superlinear if $c_k \to +\infty$ as $k \to \infty$, and quadratic if in addition

$$\sup\,\{\|w\| \mid w \in \Phi(u^{k+1}) - \mathcal{A}(c_k,\, u^k,\, u^{k+1})\} = O((\operatorname{dist}(u^k,\, U^*))^2)$$

as $k \to \infty$.

3 Perturbed Augmented Lagrangian Methods: A General Framework

We now consider the framework with (4) replaced by the more general condition

$$\frac{\partial L_{c_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) + \Omega(c_k,\, (x^k,\, \lambda^k,\, \mu^k),\, x^{k+1} - x^k) \ni 0, \tag{23}$$

where Ω is a multifunction from $(\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \times \mathbb{R}^n$ to the subsets of $\mathbb{R}^n$, characterizing various kinds of perturbations. Dual variables are updated by (5), as before. Of course, the basic scheme (4)–(5) is one special case within this framework.

Similarly to the corresponding reasoning in [13,16], one can see that the iteration subproblem of the perturbed augmented Lagrangian method can be written as subproblem's GE (13) with

$$\mathcal{A}(c,\, \widetilde{u},\, u) = \begin{pmatrix} \dfrac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \\ h(x) - \dfrac{1}{c}(\lambda - \widetilde{\lambda}) \\ -g(x) + \dfrac{1}{c}(\mu - \widetilde{\mu}) \end{pmatrix} + \begin{matrix} \Omega(c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, x - \widetilde{x}) \\ \times \\ \{0\} \\ \times \\ \{0\}, \end{matrix} \tag{24}$$

where $\widetilde{u} = (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu})$. It then can be readily seen that for Φ defined in (12), and for all $c > 0$ and $\widetilde{u},\, u \in U$, we have that

$$\Phi(u) - \mathcal{A}(c,\, \widetilde{u},\, u) = \begin{pmatrix} 0 \\ \dfrac{1}{c}(\lambda - \widetilde{\lambda}) \\ -\dfrac{1}{c}(\mu - \widetilde{\mu}) \end{pmatrix} - \begin{matrix} \Omega(c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, x - \widetilde{x}) \\ \times \\ \{0\} \\ \times \\ \{0\}. \end{matrix} \tag{25}$$

For a feasible point $\bar{x}$ of problem (1), let

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

stand for the set of indices of inequality constraints active at $\bar{x}$. Furthermore, for a stationary point $\bar{x}$ of problem (1), let $\mathcal{M}(\bar{x})$ be the set of the associated Lagrange multipliers, i.e., the set of points $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ such that $(\bar{x}, \lambda, \mu)$ solves the KKT system (10). For any $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$, define the index sets

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

of strongly and weakly active constraints, respectively.

According to [17, Proposition 3, Remark 1], if the derivatives of f, h and g are Lipschitz-continuous near $\bar{x}$, then the strong metric regularity of a solution $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ of the GE-KKT (11) is implied by the combination of the linear independence constraint qualification (LICQ)

$$\operatorname{rank} \begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix} = l + |A|, \tag{26}$$

and the strong second-order sufficient optimality condition (SSOSC)

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\,, \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{27}$$

where

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\; g'_{A_+}(\bar{x})\xi = 0\}.$$

In the context of strong metric regularity we consider the case when Ω is single-valued, having in mind that in this case it characterizes "structural" perturbations only: no inexactness in solving the subproblems is allowed (clearly, as there are many points satisfying, say, (4) for $\tau_k > 0$, inexact solution of subproblems induces set-valued perturbations).

We would like to emphasize the following subtle issue. In Theorem 3.1 below, and all the other subsequent results, the "localization conditions" like (31) concern the primal variables only. In fact, this is very natural in the augmented Lagrangians context, as the next dual iterates are uniquely defined by the primal one, via (5). This is not so in various other primal-dual optimization methods, except under strong assumptions. This (new, purely primal) approach to localization conditions yields certain improvements of convergence results even for the basic augmented Lagrangian method, corresponding to $\Omega = B(0, \tau)$ with an appropriate control of the error tolerance $\tau \geq 0$. But, since in all the results in Section 2 such conditions were on all the variables of the problem (see (17), for example), to use the results of Section 2 in the sequel, as part of the proof we must show that in the given context primal localization conditions imply primal-dual. For discussions of localization conditions, and why they are natural and in fact unavoidable, see [21].

Theorem 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, satisfying LICQ* (26)*, and let SSOSC* (27) *hold for the unique associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. Let $\Omega : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \times \mathbb{R}^n \to \mathbb{R}^n$ be such that*

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), x - \widetilde{x}) = o(\|x - \widetilde{x}\| + \|(\widetilde{x} - \bar{x}, \widetilde{\lambda} - \bar{\lambda}, \widetilde{\mu} - \bar{\mu})\|) \tag{28}$$

and

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), x^1 - \widetilde{x}) - \Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), x^2 - \widetilde{x}) = o(\|x^1 - x^2\|) \tag{29}$$

as $c \to +\infty$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$, and $x, x^1, x^2 \in \mathbb{R}^n$ tend to $\bar{x}$.

Then there exists $\bar{\delta} > 0$ such that for any $\delta \in (0, \bar{\delta}]$, there exist $\bar{c} = \bar{c}(\delta) > 0$ and $\bar{\varepsilon} = \bar{\varepsilon}(\delta) > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, any $\varepsilon \in (0, \bar{\varepsilon}]$, and any starting point $(x^0, \lambda^0, \mu^0) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$, there exists the unique sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that x^{k+1} satisfies

$$\frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \lambda^k, \mu^k) + \Omega(c_k, (x^k, \lambda^k, \mu^k), x^{k+1} - x^k) = 0, \tag{30}$$

$$\|x^{k+1} - x^k\| \leq \delta, \tag{31}$$

with the pair $(\lambda^{k+1}, \mu^{k+1})$ given by (5)*, for all $k = 0, 1, \ldots$; this sequence is contained in $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ and converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ at a linear rate, and moreover, this rate is superlinear if $c_k \to +\infty$, and quadratic provided*

$$\frac{1}{c_k} = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|) \tag{32}$$

and

$$\Omega(c_k, (x^k, \lambda^k, \mu^k), x^{k+1} - x^k) = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\| \|x^{k+1} - x^k\|).$$

as $k \to \infty$.

Proof For Φ and $\mathcal{A}$ defined by (12) and (24), respectively, from (25) and (28)–(29) it follows that conditions (15) and (16) are satisfied with any prescribed $\varkappa > 0$.

Taking into account that the combination of LICQ (26) and SSOSC (27) implies the strong metric regularity of solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE-KKT (11), the needed result would follow from Proposition 2.1, if we could complement the condition (31) by

$$\|(\lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \widetilde{\delta} \tag{33}$$

with some $\widetilde{\delta} > 0$ small enough. Therefore, it remains to show that (33) with $(\lambda^{k+1},\, \mu^{k+1})$ uniquely defined by (5) actually holds automatically under the stated assumptions, for any fixed $\widetilde{\delta} > 0$ and any choice of x^{k+1} satisfying (30)–(31), provided $\delta > 0$ and $\bar{\varepsilon} > 0$ are taken small enough, while $\bar{c}$ is taken large enough. This will be demonstrated in the proof of Theorem 3.2 below, under even weaker assumptions. □

We proceed with considering other sets of assumptions. From [16, Proposition 3.2] and [18, Remark 1], it follows that semistability of a solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE-KKT (11) is implied by the combination of the strict Mangasarian–Fromovitz constraint qualification (SMFCQ) [20, Remark 1.15], which can be formulated by saying that the Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated with $\bar{x}$ exists and is unique, and the following second-order sufficient optimality condition (SOSC), first introduced in [23]:

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}), \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{34}$$

where

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0,\ g'_{A_0}(\bar{x})\xi \leq 0\}$$

is the critical cone of problem (1) at $\bar{x}$.

Since Proposition 2.2 allows for multi-valued approximation mapping $\mathcal{A}$, we can now consider augmented Lagrangian methods with multi-valued perturbation mapping Ω, thus covering inexact solution of subproblems.

Theorem 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, satisfying SMFCQ, and let SOSC* (34) *hold with the unique associated with $\bar{x}$ Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. Let Ω be a multifunction from $(\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \times \mathbb{R}^n$ to the subsets of $\mathbb{R}^n$, such that*

$$\Omega(c,\, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}),\, x - \widetilde{x}) \ni 0 \tag{35}$$

for all c large enough, all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and all x close enough to $\bar{x}$, and

$$\sup\{\|\omega\| \mid \omega \in \Omega(c,\, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}),\, x - \widetilde{x})\} = o(\|x - \widetilde{x}\| + \|(\widetilde{x} - \bar{x},\, \widetilde{\lambda} - \bar{\lambda},\, \widetilde{\mu} - \bar{\mu})\|) \tag{36}$$

as $c \to +\infty$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$, *and* $x \to \bar{x}$.

Then there exists $\bar{\delta} > 0$ *such that for any* $\delta \in (0, \bar{\delta}]$*, there exist* $\bar{c} = \bar{c}(\delta) > 0$ *and* $\bar{\varepsilon} = \bar{\varepsilon}(\delta) > 0$ *such that for any sequence* $\{c_k\} \subset [\bar{c}, +\infty)$*, any* $\varepsilon \in (0, \bar{\varepsilon}]$*, and any starting point* $(x^0, \lambda^0, \mu^0) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$*, there exists a sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *such that* x^{k+1} *satisfies* (23) *and* (31)*, while the pair* $(\lambda^{k+1}, \mu^{k+1})$ *is given by* (5)*, for all* $k = 0, 1, \ldots$*; any such sequence is contained in* $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ *and converges to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *at a linear rate, and moreover, this rate is superlinear if* $c_k \to +\infty$*, and quadratic provided* (32) *and*

$$\begin{aligned} &\sup\{\|\omega\| \mid \omega \in \Omega(c_k, (x^k, \lambda^k, \mu^k), x^{k+1} - x^k)\} \\ &= O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\| \|x^{k+1} - x^k\|) \\ &\quad + O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|^2) \end{aligned} \tag{37}$$

hold as $k \to \infty$.

Proof For Φ and $\mathcal{A}$ defined by (12) and (24), respectively, from (25) and (36) it follows that condition (18) is satisfied.

Furthermore, as demonstrated in [16, Propositions 3.3(b), A.2] (see also Remarks 3.1 and 3.2 there), under the stated assumptions, if we consider $\mathcal{A}$ defined by (24) with Ω identically equal to $\{0\}$, then for any $\varepsilon > 0$, the subproblem (19) (which in this case corresponds to the basic augmented Lagrangian method) has a solution $u(c, \widetilde{u})$ satisfying $\|u(c, \widetilde{u}) - \bar{u}\| \leq \varepsilon$, for all c large enough and all $\widetilde{u} \in U$ close enough to $\bar{u}$. But then the same property evidently holds with any Ω satisfying (35).

Recall now that the combination of SMFCQ and SOSC (34) implies semistability of solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE-KKT (11). So, the needed result would follow from Proposition 2.2, if we complement the condition (31) by (33). Therefore, it remains to show that (33) with $(\lambda^{k+1}, \mu^{k+1})$ uniquely defined by (5) holds automatically under the stated assumptions, for any fixed $\widetilde{\delta} > 0$ and any choice of x^{k+1} satisfying (23) and (31), provided $\delta > 0$ and $\bar{\varepsilon} > 0$ are taken small enough, while $\bar{c}$ is taken large enough. We show this next.

Let $\bar{\delta} > 0$ and $\bar{c}(\delta) > 0$, $\bar{\varepsilon}(\delta) > 0$ for a given $\delta \in (0, \bar{\delta}]$ be chosen according to Proposition 2.2. For given $c \geq \bar{c}(\delta)$, $\varepsilon \in (0, \bar{\varepsilon}(\delta)]$, and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$, let $x(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ be any solution of

$$\frac{\partial L_c}{\partial x}(x, \widetilde{\lambda}, \widetilde{\mu}) + \Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), x - \widetilde{x}) \ni 0, \tag{38}$$

satisfying

$$\|x(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| \leq \delta. \tag{39}$$

Observe that such $x(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ exists, by the restrictions on δ, c, and ε. Set

$$\lambda(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = \widetilde{\lambda} + ch(x(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))), \tag{40}$$

$$\mu(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = \max\{0, \widetilde{\mu} + cg(x(\delta, c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})))\}. \tag{41}$$

Taking into account the fact that sequences generated according to Proposition 2.2 stay within $B((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, \varepsilon)$, we only need to show that for any fixed $\widetilde{\delta} > 0$ it holds that

$$\|(\lambda(\delta,\, c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu})) - \widetilde{\lambda},\, \mu(\delta,\, c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu})) - \widetilde{\mu})\| \leq \widetilde{\delta}, \tag{42}$$

perhaps with smaller $\delta > 0$ and $\bar{\varepsilon}(\delta) > 0$, and larger $\bar{c}(\delta)$.

Suppose the contrary, i.e., there exist $\widetilde{\delta} > 0$, sequences of positive reals $\{\delta_k\}$, $\{c_k\}$, and $\{\varepsilon_k\}$, and a sequence $\{(\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that $\delta_k \to 0$, $c_k \to +\infty$, $\varepsilon_k \to 0$ as $k \to \infty$, and for all k it holds that $c_k \in [\bar{c}(\delta_k),\, +\infty)$, $\varepsilon_k \in (0,\, \bar{\varepsilon}(\delta_k)]$, $(\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k) \in B((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, \varepsilon_k)$, and

$$\|(\lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\| > \widetilde{\delta}, \tag{43}$$

where

$$\lambda^k = \lambda(\delta_k,\, c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k)), \quad \mu^k = \mu(\delta_k,\, c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k)). \tag{44}$$

For each k, set

$$x^k = x(\delta_k,\, c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k)). \tag{45}$$

From (39), and since $\delta_k \to 0$ and $\varepsilon_k \to 0$ (implying that $(\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$), we have that $\{x^k\} \to \bar{x}$. Furthermore, by the same reasoning as the one showing that the iteration subproblem of the perturbed augmented Lagrangian method can be written as (13) with $\mathcal{A}$ defined in (24), from (38) and (40)–(41) we obtain that for all k

$$\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \Omega(c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k),\, x^k - \widetilde{x}^k) \ni 0, \tag{46}$$

and since $c_k \to +\infty$, $(\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and $\{x^k\} \to \bar{x}$, form (36) for any accumulation point $(\widehat{\lambda},\, \widehat{\mu})$ of $\{(\lambda^k,\, \mu^k)\}$ we have that

$$\frac{\partial L}{\partial x}(\bar{x},\, \widehat{\lambda},\, \widehat{\mu}) = 0. \tag{47}$$

Moreover, from (41) we have that

$$\mu^k \geq 0 \tag{48}$$

for all k, and hence

$$\widehat{\mu} \geq 0. \tag{49}$$

Finally, for every $i \in \setminus A$ it holds that $g_i(\bar{x}) < 0$, and since $c_k \to +\infty$, $\{x^k\} \to \bar{x}$, and $\{\widetilde{\mu}^k\} \to \bar{\mu}$ (hence it is bounded), from (41) we obtain that

$$\mu^k_{\setminus A} = 0 \tag{50}$$

for all k large enough. Therefore,

$$\widehat{\mu}_{\setminus A} = 0, \tag{51}$$

and putting together (47), (49), and (51), we conclude that $(\widehat{\lambda}, \widehat{\mu}) \in \mathcal{M}(\bar{x})$. By SMFCQ, we conclude that $(\widehat{\lambda}, \widehat{\mu}) = (\bar{\lambda}, \bar{\mu})$.

It remains to show that $\{(\lambda^k, \mu^k)\}$ is bounded, because then, by the above, it converges to $(\bar{\lambda}, \bar{\mu})$, yielding a contradiction with (43). Set $t_k = \|(\lambda^k, \mu^k)\|$, and suppose without loss of generality that $t_k \to +\infty$ as $k \to \infty$. Then, again without loss of generality, we may assume that $\{(\lambda^k, \mu^k)/t_k\}$ converges to some $(\eta, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$, $\|(\eta, \zeta)\| = 1$. From (46) we then have that for all k

$$\frac{1}{t_k} f'(x^k) + (h'(x^k))^\top \frac{\lambda^k}{t_k} + (g'(x^k))^\top \frac{\mu^k}{t_k} + \frac{1}{t_k} \Omega(c_k, (\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k), x^k - \widetilde{x}^k) \ni 0.$$

Passing onto the limit as $k \to \infty$, and employing again (36), this yields

$$(h'(\bar{x}))^\top \eta + (g'(\bar{x}))^\top \zeta = 0.$$

Moreover, (48) implies that $\zeta \geq 0$, while (50) implies that $\zeta_{\setminus A} = 0$. The existence of nonzero (η, ζ) with these properties contradicts the Mangasarian–Fromovitz constraint qualification (its dual form, see [20, Remark 1.1]), while the latter is implied by the SMFCQ assumption. This contradiction completes the proof of (42). □

Finally, we consider the case when the Lagrange multiplier set $\mathcal{M}(\bar{x})$ need not be a singleton. According to [18, Corollary 1, Remark 1], for a given $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, the upper-Lipschitz stability of a solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE-KKT (11) is implied by SOSC (34), and with U^* in (20) replaced by its subset $\{\bar{x}\} \times \mathcal{M}(\bar{x})$. Moreover, the upper-Lipschitz stability is further equivalent to the error bound

$$\Delta(x, \lambda, \mu) = O\left(\left\|\left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), \min\{\mu, -g(x)\}\right)\right\|\right)$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$, where we set

$$\Delta(x, \lambda, \mu) = \|x - \bar{x}\| + \operatorname{dist}((\lambda, \mu), \mathcal{M}(\bar{x})). \tag{52}$$

This error bound implies, in particular, that near $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the set U^* coincides with $\{\bar{x}\} \times \mathcal{M}(\bar{x})$.

Theorem 3.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being locally Lipschitz-continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, and let SOSC* (34) *hold for an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. Let Ω be a multifunction from $(\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \times \mathbb{R}^n$ to the subsets of $\mathbb{R}^n$, such that* (35) *holds for all c large enough, all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and all x close enough to $\bar{x}$, and for any fixed $\sigma > 0$ it holds that*

$$\sup\left\{ \|\omega\| \;\middle|\; \begin{array}{l} \omega \in \Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), x - \widetilde{x}),\ x \in \mathbb{R}^n, \\ \|x - \widetilde{x}\| \leq \sigma \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \end{array} \right\} = o(\Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) \tag{53}$$

as $c \to +\infty$ and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$.

Then for any $\varepsilon > 0$, and for any $\sigma > 0$ and $\Sigma > 0$ large enough, there exists $\bar{c} > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$ and any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that x^{k+1} satisfies (23) and

$$\|x^{k+1} - x^k\| \leq \sigma \Delta(x^k, \lambda^k, \mu^k), \tag{54}$$

$$\|x^{k+1} - \bar{x}\|^2 \leq \Sigma \left(\left\| \begin{pmatrix} h(x^{k+1}) \\ \max\{-\mu^k/c_k,\, g(x^{k+1})\} \end{pmatrix} \right\| + \frac{1}{c_k} \right) \Delta(x^k, \lambda^k, \mu^k), \tag{55}$$

while the pair $(\lambda^{k+1}, \mu^{k+1})$ is given by (5), for all $k = 0, 1, \ldots$; any such sequence is contained in $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ and converges to $(\bar{x}, \lambda^*, \mu^*)$ with some $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$, and the rates of convergence of $\{(x^k, \lambda^k, \mu^k)\}$ to $(\bar{x}, \lambda^*, \mu^*)$ and of $\Delta(x^k, \lambda^k, \mu^k)$ to 0 are linear, and moreover, both these rates are superlinear if $c_k \to +\infty$, and quadratic provided

$$\frac{1}{c_k} = O(\Delta(x^k, \lambda^k, \mu^k)) \tag{56}$$

and

$$\sup\{\|\omega\| \mid \omega \in \Omega(c_k, (x^k, \lambda^k, \mu^k), x^{k+1} - x^k)\} = O((\Delta(x^k, \lambda^k, \mu^k))^2)$$

as $k \to \infty$.

Proof For Φ and $\mathcal{A}$ defined by (12) and (24), respectively, from (25) and (53) it follows that condition (21) is satisfied.

Furthermore, by [16, Proposition 3.3(a), Remark 3.1], and by the reasoning in [13, Corollary 3.2, Proposition 3.3], one can see that under the stated assumptions, if we consider $\mathcal{A}$ defined by (24) with Ω identically equal to $\{0\}$, then for sufficiently large σ and c, and for $\widetilde{u} = (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in U$ close enough to $\bar{u}$, the subproblem (19) has a solution $u(c, \widetilde{u}) = (x(c, \widetilde{u}), \lambda(c, \widetilde{u}), \mu(c, \widetilde{u}))$ satisfying (22), and moreover, [13, Corollary 3.2] implies that this solution can be chosen in such a way that for any fixed Σ large enough

$$\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\|^2 \leq \Sigma \left\| \begin{pmatrix} h(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) \\ \max\{-\widetilde{\mu}/c,\, g(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})))\} \end{pmatrix} \right\| \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}). \tag{57}$$

Evidently, the same properties hold with any Ω satisfying (35).

Finally, since SOSC (34) implies the upper-Lipschitz stability of a solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of GE-KKT (11), the needed result would follow from Proposition 2.3, if we complement (54) by

$$\|(\lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leq \widetilde{\sigma} \Delta(x^k, \lambda^k, \mu^k) \tag{58}$$

with some fixed $\widetilde{\sigma} > 0$. We shall show that (58) with $(\lambda^{k+1}, \mu^{k+1})$ uniquely defined by (5) holds automatically under the stated assumptions, for a sufficiently large $\widetilde{\sigma} > 0$, and for any choice of x^{k+1} satisfying (23) and (54)–(55) with any fixed $\sigma > 0$ and $\Sigma > 0$, provided $\bar{c}$ is taken large enough, and assuming that (x^0, λ^0, μ^0) is taken close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Let $\bar{c} > 0$ be chosen according to Proposition 2.3. Then for a given $c \geq \bar{c}$ and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, let $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ be a solution of the inclusion (38), satisfying

$$\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| \leq \sigma \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \tag{59}$$

and

$$\begin{aligned} &\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\|^2 \\ &\leq \Sigma \left(\left\| \begin{pmatrix} h(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))) \\ \max\{-\widetilde{\mu}/c,\, g(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})))\} \end{pmatrix} \right\| + \frac{1}{c} \right) \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}). \end{aligned} \tag{60}$$

Observe that such $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ exists by the restrictions on c and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$, and by the discussion above regarding (57) (which evidently implies (60)), provided σ and Σ are taken large enough (observe that increasing these constants only relaxes the requirements on x^{k+1}). Taking into account that sequences generated according to Proposition 2.3 stay within $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ with any prescribed $\varepsilon > 0$, provided $\bar{c}$ is large enough and (x^0, λ^0, μ^0) is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, we only need to prove that

$$\|(\lambda(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{\lambda},\, \mu(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{\mu})\| \leq \widetilde{\sigma} \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$$

for all $\widetilde{\sigma}$ and c large enough, and all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, where

$$\lambda(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = \widetilde{\lambda} + ch(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))), \tag{61}$$

$$\mu(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = \max\{0,\, \widetilde{\mu} + cg(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})))\}. \tag{62}$$

Suppose the contrary, i.e., that there exist sequences of positive reals $\{\widetilde{\sigma}_k\}$ and $\{c_k\}$ such that $\widetilde{\sigma}_k \to +\infty$ and $c_k \to +\infty$, and a sequence $\{(\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, such that for all k

$$D_k \geq \widetilde{\sigma}_k \Delta_k, \tag{63}$$

where

$$D_k = \|(\lambda^k - \widetilde{\lambda}^k,\, \mu^k - \widetilde{\mu}^k)\|, \tag{64}$$

$$\Delta_k = \Delta(\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k), \tag{65}$$

$$\lambda^k = \lambda(c_k, (\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k)), \quad \mu^k = \mu(c_k, (\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k)). \tag{66}$$

For all k we have that

$$(\lambda^k - \widetilde{\lambda}^k,\, \mu^k - \widetilde{\mu}^k) = (c_k h(x^k),\, \max\{-\widetilde{\mu}^k,\, c_k g(x^k)\}), \tag{67}$$

where

$$x^k = x(c_k, (\widetilde{x}^k, \widetilde{\lambda}^k, \widetilde{\mu}^k)). \tag{68}$$

Observe that the estimate (59) guarantees that $\{x^k\} \to \bar{x}$.

Without loss of generality we can assume that $\{(\lambda^k - \widetilde{\lambda}^k,\, \mu^k - \widetilde{\mu}^k)/D_k\}$ converges to some $(\eta,\, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$, $\|(\eta,\, \zeta)\| = 1$.

For each k, let $(\widehat{\lambda}^k,\, \widehat{\mu}^k)$ be the projection of $(\widetilde{\lambda}^k,\, \widetilde{\mu}^k)$ onto $\mathcal{M}(\bar{x})$. Since $\bar{\mu}_{\setminus A} = 0$ and $\widehat{\mu}^k_{\setminus A} = 0$, and since $c_k \to +\infty$, employing (63) and (52), (65), for all k large enough we obtain that

$$\| \max\{-\widetilde{\mu}^k_{\setminus A},\, c_k g_{\setminus A}(x^k)\}\| = \|\widetilde{\mu}^k_{\setminus A}\| = \|\widetilde{\mu}^k_{\setminus A} - \widehat{\mu}^k_{\setminus A}\| \leq \varDelta_k \leq D_k/\widetilde{\sigma}_k.$$

Hence, taking into account (67), the assumption $\widetilde{\sigma}_k \to +\infty$ implies that

$$\zeta_{\setminus A} = 0. \tag{69}$$

Furthermore, since $x(c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}))$ solves (38), from (61)–(62), (66), (68), we derive that

$$\begin{aligned}
0 \in\ & \frac{\partial L}{\partial x}(x^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k) + (h'(x^k))^\top c_k h(x^k) + (g'(x^k))^\top \max\{-\widetilde{\mu}^k,\, c_k g(x^k)\} \\
& +\varOmega(c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k),\, x^k - \widetilde{x}^k) \\
=\ & (h'(x^k))^\top c_k h(x^k) + (g'(x^k))^\top \max\{-\widetilde{\mu}^k,\, c_k g(x^k)\} \\
& +\varOmega(c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k),\, x^k - \widetilde{x}^k) + \frac{\partial L}{\partial x}(\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k) + O(\|x^k - \widetilde{x}^k\|) \\
=\ & (h'(x^k))^\top c_k h(x^k) + (g'(x^k))^\top \max\{-\widetilde{\mu}^k,\, c_k g(x^k)\} \\
& +\varOmega(c_k,\, (\widetilde{x}^k,\, \widetilde{\lambda}^k,\, \widetilde{\mu}^k),\, x^k - \widetilde{x}^k) + O(\varDelta_k),
\end{aligned}$$

where the last equality is by (59) and (52), (65). Employing (53) and (63)–(65) we then obtain

$$(h'(x^k))^\top c_k h(x^k) + (g'(x^k))^\top \max\{-\widetilde{\mu}^k,\, c_k g(x^k)\} = O(\varDelta_k) = O(D_k/\widetilde{\sigma}_k).$$

Since $\widetilde{\sigma}_k \to +\infty$, according to (67) and (69) this implies that

$$\begin{pmatrix} \eta \\ \zeta_A \end{pmatrix} \in \ker \begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix}^\top . \tag{70}$$

Let P stand for the orthogonal projector in $\mathbb{R}^l \times \mathbb{R}^A$ on the null space

$$\ker \begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix}^\top = \left(\mathrm{im} \begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix} \right)^\perp ,$$

appearing in the right-hand side of (70). Since

$$\begin{pmatrix} h(x^k) \\ g_A(x^k) \end{pmatrix} = \begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix} (x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2),$$

employing (60), (64)–(68), we obtain that

$$\begin{aligned}
\left\| P \begin{pmatrix} h(x^k) \\ g_A(x^k) \end{pmatrix} \right\| &= O(\|x^k - \bar{x}\|^2) \\
&= O\left(\left(\left\| \begin{pmatrix} h(x^k) \\ \max\{-\widetilde{\mu}^k/c_k,\, g(x^k)\} \end{pmatrix} \right\| + \frac{1}{c_k} \right) \Delta_k \right) \\
&= O\left(\frac{1}{c_k} (D_k + 1) \Delta_k \right) \\
&= O\left(\frac{1}{c_k} D_k \left(\Delta_k + 1/\widetilde{\sigma}_k \right) \right)
\end{aligned}$$

as $k \to \infty$, where the last estimate is by (63). Therefore, by the obvious property $\Delta_k \to 0$, and since $\widetilde{\sigma}_k \to +\infty$, we have that

$$\left\{ \frac{c_k}{D_k} P \begin{pmatrix} h(x^k) \\ g_A(x^k) \end{pmatrix} \right\} \to 0. \tag{71}$$

Since the set A is finite, we can assume, without loss of generality, that there exist $I_1,\, I_2 \subset A$ such that $I_1 \cup I_2 = A$, $I_1 \cap I_2 = \emptyset$, and for all k

$$\max\{-\widetilde{\mu}^k_{I_1},\, c_k g_{I_1}(x^k)\} = c_k g_{I_1}(x^k), \tag{72}$$

$$\begin{aligned}
\max\{-\widetilde{\mu}^k_{I_2},\, c_k g_{I_2}(x^k)\} &= -\widetilde{\mu}^k_{I_2} \\
&= -\widehat{\mu}^k_{I_2} + (\widehat{\mu}^k_{I_2} - \widetilde{\mu}^k_{I_2}) \\
&= -\widehat{\mu}^k_{I_2} + O(\Delta_k) \\
&= -\widehat{\mu}^k_{I_2} + o(D_k),
\end{aligned} \tag{73}$$

where the next-to-last equality is by (52), (65), while the last one is by (63) and the assumption that $\widetilde{\sigma}_k \to +\infty$. Taking again into account that $\widehat{\mu}^k \geq 0$, the latter chain of equalities yields

$$\zeta_{I_2} \leq 0. \tag{74}$$

Define the polyhedral cone

$$K = \{(y,\, z_A) \in \mathbb{R}^l \times \mathbb{R}^A \mid y = 0,\, z_{I_1} = 0,\, z_{I_2} \geq 0\}.$$

Observe that according to (67) and (72), and the first equality in (73), for all k it holds that

$$(\lambda^k - \widetilde{\lambda}^k) - c_k h(x^k) = 0, \quad (\mu^k - \widetilde{\mu}^k)_{I_1} - c_k g_{I_1}(x^k) = 0,$$

$$(\mu^k - \widetilde{\mu}^k)_{I_2} - c_k g_{I_2}(x^k) = -\widetilde{\mu}^k_{I_2} - c_k g_{I_2}(x^k) \geq 0,$$

implying that

$$\begin{pmatrix} \lambda^k - \widetilde{\lambda}^k \\ (\mu^k - \widetilde{\mu}^k)_A \end{pmatrix} - c_k \begin{pmatrix} h(x^k) \\ g_A(x^k) \end{pmatrix} \in K.$$

Hence,

$$\frac{1}{D_k} P \begin{pmatrix} \lambda^k - \widetilde{\lambda}^k \\ (\mu^k - \widetilde{\mu}^k)_A \end{pmatrix} - \frac{c_k}{D_k} P \begin{pmatrix} h(x^k) \\ g_A(x^k) \end{pmatrix} \in PK.$$

Passing onto the limit as $k \to \infty$, and making use of (70) and (71), and of the fact that a linear map transforms a polyhedral set into a polyhedral (hence, closed) set, this yields

$$\begin{pmatrix} \eta \\ \zeta_A \end{pmatrix} = P \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix} \in PK,$$

i.e., there exists $(y,\, z_A) \in K$ such that

$$\begin{pmatrix} \eta \\ \zeta_A \end{pmatrix} = P \begin{pmatrix} y \\ z_A \end{pmatrix}. \tag{75}$$

By (74), and by the definition of K, we further obtain that

$$\begin{aligned} 0 \geq \langle \zeta_{I_2},\, z_{I_2} \rangle &= \left\langle \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix}, \begin{pmatrix} y \\ z_A \end{pmatrix} \right\rangle \\ &= \left\langle P \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix}, \begin{pmatrix} y \\ z_A \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix}, P \begin{pmatrix} y \\ z_A \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix}, \begin{pmatrix} \eta \\ \zeta_A \end{pmatrix} \right\rangle = \|(\eta,\, \zeta_A)\|^2 , \end{aligned}$$

where the second equality is by (70), the third one is by the symmetry of P, and the fourth is by (75). Combined with (69), this implies that $(\eta,\, \zeta) = 0$, contradicting $\|(\eta,\, \zeta)\| = 1$. □

4 Local Convergence of the Proximal Augmented Lagrangian Method

We start with the exact proximal augmented Lagrangian method, defined by (5) and (7) with $\tau_k = 0$ for all k. It fits the framework of the perturbed augmented Lagrangian methods developed in Section 3, by taking in (23)

$$\Omega(c,\, (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, \xi) = \frac{1}{c}\, \xi \tag{76}$$

for $c > 0$, $(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, $\xi \in \mathbb{R}^n$, and using (5) for updating dual variables. I.e., for now, we consider only the "structural" perturbation to the basic scheme, which in this case is the proximal regularization. The framework of the perturbed augmented Lagrangian methods immediately gives the following result, which improves on that in [31] by not requiring twice differentiability of the problem data. The subsequent results go further beyond, in various directions.

Theorem 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, satisfying LICQ* (26)*, and let SSOSC* (27) *hold for the unique associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then for any $\delta > 0$ small enough, there exists $\bar{c} > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, and any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists the unique sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that x^{k+1} is a stationary point of problem (6)*, satisfying* (31)*, and the pair $(\lambda^{k+1}, \mu^{k+1})$ is given by* (5)*, for all $k = 0, 1, \ldots$; this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ at a linear rate, and moreover, this rate superlinear if $c_k \to +\infty$, and quadratic provided* (32) *holds as $k \to \infty$.*

Proof The function Ω defined in (76) evidently satisfies (28) and (29). The needed result then follows by applying Theorem 3.1. □

Let now the tolerance parameter τ_k in (7) be taken as a function of the penalty parameter c_k and of the current iterate:

$$\tau_k = \tau(c_k, (x^k, \lambda^k, \mu^k)), \tag{77}$$

with some fixed $\tau : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$. Of course, this function would be required to satisfy certain conditions ((79) or (91) below). A computationally implementable way of choosing a suitable τ will be specified in Remark 5.1, along with all the other parameters that appear in the algorithms.

The inexact proximal augmented Lagrangian method, defined by (5) and (7), can fit the perturbed framework of (23) and (5), now with multi-valued Ω, taking

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \xi) = \frac{1}{c}\xi + B\left(0, \tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))\right).$$

However, writing this way, it can be seen that the assumption (35) of the abstract framework is not satisfied. This is why we replace this formulation by a larger Ω, as follows:

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \xi) = B\left(0, \frac{1}{c}\|\xi\| + \tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))\right). \tag{78}$$

Of course, this does not lead to an equivalent interpretation of the inexact proximal augmented Lagrangian iteration: (78) allows for different iterations as well. For instance, the iteration of the basic inexact augmented Lagrangian method given by (4)–(5) also fits the framework (23), (5), with Ω specified in (78). Nevertheless, this imprecise description is enough to derive the needed convergence properties from the abstract framework that indeed should be considered as a theoretical framework, i.e., a technical tool needed for (some aspects of the) analysis only.

Theorem 4.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, satisfying SMFCQ, and let SOSC* (34) *hold with the unique associated with $\bar{x}$ Lagrange* *multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$*. Let* $\tau : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$ *be a function satisfying*

$$\tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = o(\|(\widetilde{x} - \bar{x}, \widetilde{\lambda} - \bar{\lambda}, \widetilde{\mu} - \bar{\mu})\|) \tag{79}$$

as $c \to +\infty$ *and* $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$*.*

Then for any $\delta > 0$ *small enough, there exists* $\bar{c} > 0$ *such that for any sequence* $\{c_k\} \subset [\bar{c}, +\infty)$*, and any starting point* $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*, there exists a sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *such that* x^{k+1} *satisfies* (7) *with* τ_k *defined according to* (77)*,* x^{k+1} *also satisfies* (31)*, the pair* $(\lambda^{k+1}, \mu^{k+1})$ *is given by*(5)*, for all* $k = 0, 1, \ldots$*; any such sequence converges to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *at a linear rate, and moreover, this rate is superlinear if* $c_k \to +\infty$*, and quadratic provided* (32) *and*

$$\tau(c_k, (x^k, \lambda^k, \mu^k)) = O(\|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\|^2), \tag{80}$$

hold as $k \to \infty$*.*

Proof The multifunction Ω defined in (78) with $\tau(\cdot)$ satisfying (79) evidently satisfies (35) and (36). The needed result will follow applying Theorem 3.2, if we show that for any $\delta > 0$, and for each (x^k, λ^k, μ^k) generated according to that theorem (and in particular, staying within $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon))$, there exists x^{k+1} satisfying (7) with $\tau_k = 0$ (hence, also for any $\tau_k > 0$) and (31), provided $\bar{c}$ is taken large enough, while $\bar{\varepsilon} > 0$ is taken small enough.

Specifically, we shall show the existence of $\bar{c} > 0$ and $\bar{\varepsilon} > 0$ such that for any $c \geq \bar{c}$ and any $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \bar{\varepsilon})$, there exists a solution $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ of the equation

$$\frac{\partial L_c}{\partial x}(x, \widetilde{\lambda}, \widetilde{\mu}) + \frac{1}{c}(x - \widetilde{x}) = 0 \tag{81}$$

(whose iterative counterpart is (7) with $\tau_k = 0$), satisfying

$$\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| \leq \delta \tag{82}$$

(whose iterative counterpart is (31)).

According to [16, Proposition 3.3 (b)], there exist $\bar{\rho} > 0$ and $\bar{c} > 0$ possessing the following property: for any $\rho \in (0, \bar{\rho}]$ there exists $\bar{\varepsilon} = \bar{\varepsilon}(\rho) > 0$ such that for any $c \geq \bar{c}$ and any $(\widetilde{\lambda}, \widetilde{\mu}) \in B((\bar{\lambda}, \bar{\mu}), \bar{\varepsilon}(\rho))$, any global solution $\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu}))$ of the optimization problem

$$\text{minimize } L_c(x, \widetilde{\lambda}, \widetilde{\mu}) \text{ subject to } x \in B(\bar{x}, \bar{\rho}) \tag{83}$$

satisfies

$$\|\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\| \leq \rho. \tag{84}$$

Without loss of generality we may also assume that

$$\bar{\varepsilon}(\rho) \leq \rho. \tag{85}$$

Suppose now that $c \geq \bar{c}$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \bar{\varepsilon}(\rho))$. Let $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ be a global solution of the optimization problem

$$\text{minimize } L_c(x, \widetilde{\lambda}, \widetilde{\mu}) + \frac{1}{2c}\|x - \widetilde{x}\|^2 \text{ subject to } x \in B(\bar{x}, \bar{\rho}). \tag{86}$$

Then

$$\begin{aligned} L_c(\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})), \widetilde{\lambda}, \widetilde{\mu}) + \frac{1}{2c}\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\|^2 &\leq L_c(x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})), \widetilde{\lambda}, \widetilde{\mu}) \\ &\quad + \frac{1}{2c}\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\|^2 \\ &\leq L_c(\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})), \widetilde{\lambda}, \widetilde{\mu}) \\ &\quad + \frac{1}{2c}\|\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\|^2. \end{aligned} \tag{87}$$

The latter implies that

$$\begin{aligned} \|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| &\leq \|\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| \\ &\leq \|\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\| + \|\widetilde{x} - \bar{x}\| \\ &\leq \rho + \bar{\varepsilon}(\rho) \\ &\leq 2\rho, \end{aligned} \tag{88}$$

where the last two inequalities are by (84) and (85), respectively. Then, similarly,

$$\begin{aligned} \|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\| &\leq \|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| + \|\widetilde{x} - \bar{x}\| \\ &\leq 2\rho + \bar{\varepsilon}(\rho) \\ &\leq 3\rho. \end{aligned} \tag{89}$$

If we take $\rho \in (0, \min\{\delta/2, \bar{\rho}\}]$, then (88) yields (82). If, in addition, $\rho < \bar{\rho}/3$, then (89) yields

$$\|x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\| < \bar{\rho}.$$

Therefore, $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ is a local solution of the unconstrained optimization problem

$$\text{minimize } L_c(x, \widetilde{\lambda}, \widetilde{\mu}) + \frac{1}{2c}\|x - \widetilde{x}\|^2, \quad x \in \mathbb{R}^n, \tag{90}$$

and hence, a stationary point of this problem, i.e., a solution of the equation (81). □

We complete this section with the case of possibly nonunique Lagrange multipliers.

Theorem 4.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being locally Lipschitz-continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, and let*

SOSC (34) *hold for an associated Lagrange multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. *Let* $\tau : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$ *be a function satisfying*

$$\tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = o(\Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) \tag{91}$$

as $c \to +\infty$ *and* $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$, *where* Δ *is defined in* (52).

Then for any $\varepsilon > 0$, *and for any* $\sigma > 0$ *large enough, there exists* $\bar{c} > 0$ *such that for any sequence* $\{c_k\} \subset [\bar{c}, +\infty)$, *and any* $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *there exists a sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *such that* x^{k+1} *satisfies* (7) *with* τ_k *defined according to* (77), x^{k+1} *also satisfies* (54) *and* (55), *the pair* $(\lambda^{k+1}, \mu^{k+1})$ *is given by* (5), *for all* $k = 0, 1, \ldots$; *any such sequence is contained in* $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ *and converges to* $(\bar{x}, \lambda^*, \mu^*)$ *with some* $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$, *and the rates of convergence of* $\{(x^k, \lambda^k, \mu^k)\}$ *to* $(\bar{x}, \lambda^*, \mu^*)$ *and of* $\Delta(x^k, \lambda^k, \mu^k)$ *to* 0 *are linear, and moreover, both these rates are superlinear if* $c_k \to +\infty$, *and quadratic provided* (56) *and*

$$\tau(c_k, (x^k, \lambda^k, \mu^k)) = O((\Delta(x^k, \lambda^k, \mu^k))^2) \tag{92}$$

hold as $k \to \infty$.

Proof The multifunction Ω defined in (78) with $\tau(\cdot)$ satisfying (91), evidently satisfies (35) and (53). The needed result would follow from Theorem 3.3, if we show that for $\sigma > 0$ and $\Sigma > 0$ large enough, and for each (x^k, λ^k, μ^k) generated according to that theorem (and in particular, staying close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$), there exists x^{k+1} satisfying (7) with $\tau_k = 0$ (hence, with any $\tau_k > 0$ also), as well as (54), (55), provided $\bar{c}$ is taken large enough, and (x^0, λ^0, μ^0) is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

By the chain of inequalities (87) from the proof of Theorem 4.2, one can see that for any $\bar{\rho} > 0$, $c > 0$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, and any global solutions $\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu}))$ and $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ of problems (83) and (86), respectively, the first inequality in (88) is satisfied.

Appealing again to the reasoning in [13, Corollary 3.2, Proposition 3.3], and employing also [16, Proposition 3.3 (b), Remark 3.1], it can be seen that

$$\|\widehat{x}(c, (\widetilde{\lambda}, \widetilde{\mu})) - \widetilde{x}\| \leq \sigma \Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$$

for all σ and c large enough, and all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Then the first inequality in (88) implies (59) (whose iterative counterpart is (54)).

Furthermore, from [16, Proposition 3.3 (a), Remark 3.1] it follows that there exist $\bar{c} > 0$ and $\gamma > 0$ such that

$$L_c(x, \widehat{\lambda}, \widehat{\mu}) \geq L_c(\bar{x}, \widehat{\lambda}, \widehat{\mu}) + \gamma \|x - \bar{x}\|^2$$

for all $c \geq \bar{c}$, all $x \in \mathbb{R}^n$ close enough to $\bar{x}$, and all $(\widehat{\lambda}, \widehat{\mu}) \in \mathcal{M}(\bar{x})$ close enough to $(\bar{\lambda}, \bar{\mu})$. Let $(\widehat{\lambda}, \widehat{\mu})$ be the projection of $(\widetilde{\lambda}, \widetilde{\mu})$ onto $\mathcal{M}(\bar{x})$. Following the reasoning in [13, Corollary 3.2], one can see that

$$\begin{aligned}\gamma\|x(c,\,(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu}))-\bar{x}\|^2 \le\, & \langle\widehat{\lambda}-\widetilde{\lambda},\,h(x(c,\,(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu})))\rangle \\ & +\langle\widehat{\mu}-\widetilde{\mu},\,\max\{-\widetilde{\mu}/c,\,g(x(c,\,(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu})))\}\rangle \\ & +\frac{1}{2c}\|\widetilde{x}-\bar{x}\|^2. \end{aligned} \tag{93}$$

The argument involves concavity of the function

$$(\lambda,\,\mu)\mapsto L_c(x,\,\lambda,\,\mu)-\|(\lambda,\,\mu)\|^2/(2c):\mathbb{R}^l\times\mathbb{R}^m\to\mathbb{R}$$

for all $x\in\mathbb{R}^n$ and $c>0$, following from [28, Proposition 7.1], and the equality

$$L_c(\bar{x},\,\lambda,\,\mu)=f(\bar{x})+\|(\lambda,\,\mu)\|^2/(2c),$$

for all $c>0$ and all $(\lambda,\,\mu)\in\mathbb{R}^l\times\mathbb{R}^m$. Recalling (52), from (93) we readily obtain (60) (whose iterative counterpart is (55)) with $\Sigma=1/\gamma$.

It remains to observe that according to (93), $x(c,\,(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu}))\to\bar{x}$ as $c\to+\infty$ and $(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu})\to(\bar{x},\,\bar{\lambda},\,\bar{\mu})$. Therefore, for all c large enough, and all $(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu})$ close enough to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, the point $x(c,\,(\widetilde{x},\,\widetilde{\lambda},\,\widetilde{\mu}))$ is a local solution of the unconstrained optimization problem (90), and hence, a stationary point of this problem, i.e., a solution of the equation (81) (whose iterative counterpart is (7) with $\tau_k=0$). □

5 Local Convergence of the Augmented Lagrangian Method with Smoothing

The proofs in Section 4 are by relating properties of primal iterates of the proximal augmented Lagrangian method to those of the basic one. Proofs in the current section need to be more direct, as there appears to be no evident relations of this kind for primal iterates of the method with smoothing, given by (5), (9). The counterparts of these direct proofs might be developed in the context of Section 4 as well, allowing, in particular, to cover the local convergence theories for the basic method of multipliers itself, rather than to refer to them. The latter can also be achieved within the current section, adapting the natural convention that $s_0(\cdot)\equiv\max\{0,\,\cdot\}$, the property fully agreeing with the requirements on s_θ for $\theta>0$, stated below.

We can employ any family of smoothing functions s_θ satisfying the following properties: for every $\theta>0$ the function is twice continuously differentiable on $\mathbb{R}$, and there exists $\Gamma>0$ such that

$$|s_\theta(t)-\max\{0,\,t\}|\le\Gamma\theta,\quad |s_\theta(t)s'_\theta(t)-\max\{0,\,t\}|\le\Gamma\theta, \tag{94}$$

for all $t\in\mathbb{R}$. The first property in (94) is standard in the smoothing literature, starting with [7]. The second inequality in (94) we introduce here for our purposes. Both properties in (94) hold for (at least) the following two examples, both coming from [7]:

$$s_\theta(t) = \frac{1}{2}(t + \sqrt{t^2 + 4\theta^2}), \tag{95}$$

and

$$s_\theta(t) = t + \theta \log\left(1 + e^{-t/\theta}\right). \tag{96}$$

Let the smoothing parameter in (9) be controlled by

$$\theta_k = \theta(c_k, (x^k, \lambda^k, \mu^k)), \tag{97}$$

with some function $\theta : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$. Again, of course, this θ would be required to satisfy certain conditions ((101) or (107) below). An example of a computationally implementable choice will be given in Remark 5.1, together with all the other parameters.

To place the method with smoothing, given by (5) and (9), within the perturbed framework of Section 3, define

$$r(c, \theta, \widetilde{\mu}, x) = \frac{1}{2c}(\|s_\theta(\widetilde{\mu} + cg(x))\|^2 - \|\max\{0, \widetilde{\mu} + cg(x)\}\|^2), \tag{98}$$

$$\begin{aligned} R(c, \theta, \widetilde{\mu}, x) &= \frac{\partial r}{\partial x}(c, \theta, \widetilde{\mu}, x) \\ &= (g'(x))^\top (s_\theta(\widetilde{\mu} + cg(x)) \circ s'_\theta(\widetilde{\mu} + cg(x)) - \max\{0, \widetilde{\mu} + cg(x)\}), \end{aligned} \tag{99}$$

where s'_θ is also applied componentwise, and $\circ$ stands for the Hadamard product.

We shall not consider the strongly regular case in the smoothing setting, because satisfying the assumption (29) in Theorem 3.1 would require further assumptions on s_θ, to ensure Lipschitz-continuity of $R(c, \theta, \widetilde{\mu}, \cdot)$ with an (arbitrarily) small constant. Such assumptions might be at least questionable, as they do not hold for s_θ defined in (95) or (96), for example.

We proceed with the other two sets of assumptions considered in Section 3, which are less demanding not only with respect to the constraint qualifications and second-order conditions being used, but also with respect to the required quality of approximations. In particular, the two smoothing options (95) and (96) are readily accommodated.

Using (2) and (8), from (99) we obtain that the augmented Lagrangian method with smoothing, given by (5) and (9) with τ_k defined by (77), can be embedded into the framework of Section 3 by setting

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \xi) = R(c, \theta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \widetilde{\mu}, \widetilde{x} + \xi) + B\left(0, \tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))\right),$$

for $c > 0$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, $\xi \in \mathbb{R}^n$. However, similarly to the case of the inexact proximal augmented Lagrangian method considered above, such a formulation will not satisfy (35) in the general framework. This is why we replace this formulation by a larger Ω, as follows:

$$\Omega(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \xi) = B\left(0, \|R(c, \theta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \widetilde{\mu}, \widetilde{x}+\xi)\| + \tau(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))\right) \tag{100}$$

for $c > 0$, $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, $\xi \in \mathbb{R}^n$.

One comment is in order concerning the differentiability assumptions in this section. As already mentioned in Section 1, as for other approaches described in [3, Chapter 4.2.5], the motivation for smoothing is to preserve in the subproblems of the algorithm twice differentiability, when the data of the original problem (1) is twice differentiable. While this is a point to keep in mind, as the subsequent proofs themselves do not need twice differentiability, we shall employ the same weaker smoothness assumptions as are used throughout the paper.

Theorem 5.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, satisfying SMFCQ, and let SOSC* (34) *hold with the unique associated with $\bar{x}$ Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. For every $\theta > 0$, let $s_\theta : \mathbb{R} \to \mathbb{R}$ be a function satisfying* (94) *with some $\Gamma > 0$. Let $\tau, \theta : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$ be functions satisfying* (79) *and*

$$\theta(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = o(\|(\widetilde{x}-\bar{x}, \widetilde{\lambda}-\bar{\lambda}, \widetilde{\mu}-\bar{\mu})\|) \tag{101}$$

as $c \to +\infty$ and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$.

Then for any $\delta > 0$ small enough, there exists $\bar{c} > 0$ such that for any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, and any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that x^{k+1} satisfies (9) *with τ_k and θ_k defined according to* (77) *and* (97)*, respectively, as well as* (31)*, while the pair $(\lambda^{k+1}, \mu^{k+1})$ is given by* (5)*, for all $k = 0, 1, \ldots$; any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ at a linear rate, and moreover, this rate is superlinear if $c_k \to +\infty$, and quadratic provided* (32)*,* (80)*, and*

$$\theta(c_k, (x^k, \lambda^k, \mu^k)) = O(\|(x^k-\bar{x}, \lambda^k-\bar{\lambda}, \mu^k-\bar{\mu})\|^2) \tag{102}$$

hold as $k \to \infty$.

Proof From the second inequality in (94), (99) and (101), we have that

$$R(c, \theta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), \widetilde{\mu}, \widetilde{x}+\xi) = O(\theta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = o(\|(\widetilde{x}-\bar{x}, \widetilde{\lambda}-\bar{\lambda}, \widetilde{\mu}-\bar{\mu})\|)$$

as $c \to +\infty$ and $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Then the multifunction Ω defined in (100) with $\tau(\cdot)$ satisfying (79) satisfies (35) and (36).

The needed result would now follow applying Theorem 3.2, if we show that for any $\delta > 0$, and for each (x^k, λ^k, μ^k) generated according to that theorem (and in particular, staying within $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$), there exists x^{k+1} satisfying

(9) with $\tau_k = 0$ (hence, also for any $\tau_k > 0$) and also satisfying (31), provided $\bar{c}$ is taken large enough, while $\bar{\varepsilon} > 0$ is taken small enough.

Fix $\bar{\rho} > 0$ such that $\bar{x}$ is the unique global solution of problem (1) with the additional constraint $x \in B(\bar{x}, \bar{\rho})$ (the existence of such $\bar{\rho}$ follows from [16, Remarks 3.1, 3.2]). Consider any sequences of positive reals $\{c_k\}$ and $\{\theta_k\}$ such that $c_k \to +\infty$ and $\theta_k \to 0$ as $k \to \infty$, and any sequences $\{(\widetilde{\lambda}^k, \widetilde{\mu}^k)\} \subset \mathbb{R}^l \times \mathbb{R}^m$ convergent to 0, and $\{x^k\} \subset \mathbb{R}^n$ such that for each k, the point x^k is a global solution of the optimization problem

$$\text{minimize } L_{c,\theta}(x, \widetilde{\lambda}, \widetilde{\mu}) \text{ subject to } x \in B(\bar{x}, \bar{\rho}), \tag{103}$$

with $c = c_k$, $\theta = \theta_k$, and $(\widetilde{\lambda}, \widetilde{\mu}) = (\widetilde{\lambda}^k, \widetilde{\mu}^k)$. Since $\{x^k\}$ is contained in the compact set $B(\bar{x}, \bar{\rho})$, this sequence has an accumulation point $\widehat{x} \in B(\bar{x}, \bar{\rho})$. We are going to show that $\widehat{x} = \bar{x}$, implying that $\{x^k\} \to \bar{x}$.

Without loss of generality, suppose that the entire $\{x^k\}$ converges to $\widehat{x}$. Employing (8), the first inequality in (94), and (98), for each k large enough we then obtain that

$$\begin{aligned}
&f(x^k) + \frac{1}{2c_k}(\|\widetilde{\lambda}^k + c_k h(x^k)\|^2 + \|s_{\theta_k}(\widetilde{\mu}^k + c_k g(x^k))\|^2) \\
&= L_{c_k,\theta_k}(x^k, \widetilde{\lambda}^k, \widetilde{\mu}^k) \\
&\le L_{c_k,\theta_k}(\bar{x}, \widetilde{\lambda}^k, \widetilde{\mu}^k) \\
&= f(\bar{x}) + \frac{1}{2c_k}(\|\widetilde{\lambda}^k + c_k h(\bar{x})\|^2 + \|s_{\theta_k}(\widetilde{\mu}^k + c_k g(\bar{x}))\|^2) \\
&= f(\bar{x}) + \frac{1}{2c_k}(\|\widetilde{\lambda}^k\|^2 + \|\max\{0, \widetilde{\mu}_A^k\}\|^2) \\
&\quad + r(c_k, \theta_k, \widetilde{\mu}^k, \bar{x}) \\
&= f(\bar{x}) + \frac{1}{2c_k}(\|\widetilde{\lambda}^k\|^2 + \|\max\{0, \widetilde{\mu}_A^k\}\|^2) \\
&\quad + \frac{1}{2c_k}(2\Gamma\|\max\{0, \widetilde{\mu}_A^k\}\|\theta_k + \Gamma^2\theta_k^2).
\end{aligned}$$

Hence,

$$f(\widehat{x}) + \limsup_{k\to\infty} \frac{1}{2c_k}(\|\widetilde{\lambda}^k + c_k h(x^k)\|^2 + \|s_{\theta_k}(\widetilde{\mu}^k + c_k g(x^k))\|^2) \le f(\bar{x}). \tag{104}$$

From the latter, we readily conclude that $h(\widehat{x}) = 0$. Moreover, if there exists an index $i \in \{1, \ldots, m\}$ such that $g_i(\bar{x}) > 0$, then for all k large enough it holds that $\widetilde{\mu}_i^k + c_k g_i(x^k) > 0$, and hence, by the first inequality in (94),

$$s_{\theta_k}(\widetilde{\mu}_i^k + c_k g_i(x^k)) \ge \max\{0, \widetilde{\mu}_i^k + c_k g_i(x^k)\} - \Gamma\theta_k = \widetilde{\mu}_i^k + c_k g_i(x^k) - \Gamma\theta_k.$$

Since $c_k \to +\infty$ and $\theta_k \to 0$, the latter inequality implies that

$$\frac{1}{2c_k}(s_{\theta_k}(\widetilde{\mu}_i^k + c_k g_i(x^k))\})^2 \to +\infty,$$

giving a contradiction with (104). Therefore, $g(\widehat{x}) \leq 0$, and hence, $\widehat{x}$ is feasible in problem (1) with the additional constraint $x \in B(\bar{x},\, \bar{\rho})$. Then, by the choice of $\bar{\rho} > 0$, it must hold that

$$f(\bar{x}) \leq f(\widehat{x}). \tag{105}$$

From (104) we then derive that

$$\limsup_{k\to\infty} \frac{1}{2c_k}(\|\widetilde{\lambda}^k + c_k h(x^k)\|^2 + \|s_{\theta_k}(\widetilde{\mu}^k + c_k g(x^k))\|^2) \leq 0,$$

and hence,

$$\lim_{k\to\infty} \frac{1}{2c_k}(\|\widetilde{\lambda}^k + c_k h(x^k)\|^2 + \|s_{\theta_k}(\widetilde{\mu}^k + c_k g(x^k))\|^2) = 0.$$

Then (104) yields

$$f(\widehat{x}) \leq f(\bar{x}),$$

and combining the latter with (105) gives the equality $f(\widehat{x}) = f(\bar{x})$, which is only possible if $\widehat{x} = \bar{x}$.

We have thus proven that for any $\rho \in (0,\, \bar{\rho}]$, any solution $x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu}))$ of problem (103) satisfies

$$\|x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu})) - \bar{x}\| \leq \rho,$$

provided c is large enough, $\theta > 0$ is small enough, and $(\widetilde{\lambda},\, \widetilde{\mu})$ is close enough to $(\bar{\lambda},\, \bar{\mu})$. Taking $\rho \in (0,\, \bar{\rho})$ such that $\rho \leq \delta/2$, and assuming that in addition $\|\widetilde{x} - \bar{x}\| \leq \delta/2$, we then have that $x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu}))$ solves the equation

$$\frac{\partial L_{c,\,\theta}}{\partial x}(x,\, \widetilde{\lambda},\, \widetilde{\mu}) = 0$$

(whose iterative counterpart is (9) with $\tau_k = 0$), and this point satisfies

$$\|x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu})) - \widetilde{x}\| \leq \|x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu})) - \bar{x}\| + \|\widetilde{x} - \bar{x}\| \leq \delta$$

(whose iterative counterpart is (31)). This gives the needed conclusions. □

Finally, the case of possibly nonunique Lagrange multipliers is treated in the following result. Note that in addition to (94), the smoothing function s_θ is required here to satisfy the additional condition (106). Again, two examples of suitable functions are given by (95) and (96).

Theorem 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being Lipschitz-continuous near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1)*, and let SOSC* (34) *hold for an associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. For every $\theta > 0$, let $s_\theta : \mathbb{R} \to \mathbb{R}$ be a function satisfying* (94) *with some $\Gamma > 0$, and in addition satisfying the condition*

$$s_\theta(t) \geq \max\{0,\, t\}, \tag{106}$$

for all $t \in \mathbb{R}$. *Let* $\tau, \theta : (\mathbb{R}_+ \setminus \{0\}) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}_+$ *be functions satisfying* (91) *and the conditions*

$$\theta(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = o(\Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})), \quad \theta(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})) = O(c(\Delta(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))^2), \tag{107}$$

as $c \to +\infty$ *and* $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$*, where* Δ *is defined by* (52).

Then for any $\varepsilon > 0$*, and for any* $\sigma > 0$ *large enough, there exists* $\bar{c} > 0$ *such that for any sequence* $\{c_k\} \subset [\bar{c}, +\infty)$*, and any* $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*, there exists a sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *such that* x^{k+1} *satisfies* (9) *with* τ_k *and* θ_k *defined according to* (77) *and* (97)*, respectively, as well as* (54), (55)*, while the pair* $(\lambda^{k+1}, \mu^{k+1})$ *is given by* (5)*, for all* $k = 0, 1, \ldots$*; any such sequence is contained in* $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$ *and converges to* $(\bar{x}, \lambda^*, \mu^*)$ *with some* $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$*, and the rates of convergence of* $\{(x^k, \lambda^k, \mu^k)\}$ *to* $(\bar{x}, \lambda^*, \mu^*)$ *and of* $\Delta(x^k, \lambda^k, \mu^k)$ *to* 0 *are linear, and moreover, both these rates are superlinear if* $c_k \to +\infty$*, and quadratic provided* (56), (92)*, and*

$$\theta(c_k, (x^k, \lambda^k, \mu^k)) = O((\Delta(x^k, \lambda^k, \mu^k))^2) \tag{108}$$

hold as $k \to \infty$.

Proof Similarly to the proof of Theorem 5.1, one can see from the second inequality in (94), and (99), that the multifunction Ω defined in (100), with $\tau(\cdot)$ satisfying (91) and $\theta(\cdot)$ satisfying the first condition in (107), verifies (53). So, the needed result would follow from Theorem 3.3, if we demonstrate the following: for any fixed $\bar{\rho} > 0$, any $\sigma > 0$ and $\Sigma > 0$ large enough, any c large enough, and any $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, any global solution $x(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ of (103) with $\theta = \theta(c, (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}))$ satisfies the estimates (59) (whose iterative counterpart is (54)) and (60) (whose iterative counterpart is (55)).

By the reasoning similar to that we referred to in the proof of Theorem 4.3, we derive the existence of $\gamma > 0$ such that for all c large enough, for any $\theta \geq 0$ and any $(\widetilde{\lambda}, \widetilde{\mu})$ close enough to $(\bar{\lambda}, \bar{\mu})$, and for any solution $x(c, \theta, (\widetilde{\lambda}, \widetilde{\mu}))$ of (103), it holds that

$$\begin{aligned}\gamma \|x(c, \theta, (\widetilde{\lambda}, \widetilde{\mu})) - \bar{x}\|^2 \leq{}& \langle \widehat{\lambda} - \widetilde{\lambda}, h(x(c, \theta, (\widetilde{\lambda}, \widetilde{\mu}))) \rangle \\ &+ \langle \widehat{\mu} - \widetilde{\mu}, \max\{-\widetilde{\mu}/c, g(x(c, \theta, (\widetilde{\lambda}, \widetilde{\mu})))\} \rangle \\ &+ \frac{1}{2c}(\|s_\theta(\widetilde{\mu} + cg(\bar{x}))\|^2 - \|\widetilde{\mu}\|^2) \\ &- r(c, \theta, \widetilde{\mu}, x(c, (\widetilde{\lambda}, \widetilde{\mu})))\end{aligned} \tag{109}$$

(cf. (93)), where $(\widehat{\lambda}, \widehat{\mu})$ again stands for the projection of $(\widetilde{\lambda}, \widetilde{\mu})$ on $\mathcal{M}(\bar{x})$, while $r(\cdot)$ is defined in (98). Since $g(\bar{x}) \leq 0$, it can be readily seen that for all $c \geq 0$ and $\widetilde{\mu} \in \mathbb{R}^m$, it holds that

$$\| \max\{0, \widetilde{\mu} + cg(\bar{x})\}\| \leq \| \max\{0, \widetilde{\mu}\}\| \leq \|\widetilde{\mu}\|.$$

Hence, by the first inequality in (94), we obtain that

$$\begin{aligned} & \|s_\theta(\widetilde{\mu}+cg(\bar{x}))\|^2 - \|\widetilde{\mu}\|^2 & (110)\\ &= \|s_\theta(\widetilde{\mu}+cg(\bar{x})) - \max\{0,\, \widetilde{\mu}+cg(\bar{x})\|^2 \\ &\quad +2\|s_\theta(\widetilde{\mu}+cg(\bar{x})) - \max\{0,\, \widetilde{\mu}+cg(\bar{x})\|\|\max\{0,\, \widetilde{\mu}+cg(\bar{x})\| \\ &\quad +\|\max\{0,\, \widetilde{\mu}+cg(\bar{x})\|^2 - \|\widetilde{\mu}\|^2 \\ &\leq \Gamma^2\theta^2 + 2\Gamma\|\widetilde{\mu}\|\theta. & (111)\end{aligned}$$

Then from (109), employing (106) and the first condition in (107), we obtain (60) with some $\Sigma > 0$.

Furthermore, for all c large enough, for any $\widetilde{\mu}$ close enough to $\bar{\mu}$, and for any x close enough to $\bar{x}$, employing that $\widehat{\mu}_A \geq 0$, $\widehat{\mu}_{\setminus A} = 0$, $g_A(\bar{x}) = 0$, we easily derive the following chain of estimates:

$$\begin{aligned} \|\max\{-\widetilde{\mu}/c,\, g(x)\}\| &\leq \|\max\{-\widehat{\mu}/c,\, g(x)\}\| \\ &\quad +\|\max\{-\widetilde{\mu}/c,\, g(x)\} - \max\{-\widehat{\mu}/c,\, g(x)\}\| \\ &\leq \|g_A(x)\| + \frac{1}{c}\|\widetilde{\mu} - \widehat{\mu}\| \\ &\leq \frac{1}{c}\Delta(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}) + O(\|x-\bar{x}\|)\end{aligned}$$

as $x \to \bar{x}$. Since $h(x) = O(\|x-\bar{x}\|)$, and employing (106), the second condition in (107), and (110), from (109) we obtain that

$$\|x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu})) - \bar{x}\|^2 = O(\|x(c,\, \theta,\, (\widetilde{\lambda},\, \widetilde{\mu})) - \bar{x}\|\Delta(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu})) + O((\Delta(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}))^2)$$

evidently implying (59) with some $\sigma > 0$. □

Remark 5.1 Under the smoothness assumptions in the theorems above, possible practical choices of $\tau(\cdot)$ and $\theta(\cdot)$ satisfying (79) and (91), and (101) and (107), respectively, are as follows: for a fixed $\nu > 1$,

$$\tau(c,\, \widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}) = \left\| \left(\frac{\partial L}{\partial x}(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, h(\widetilde{x}),\, \min\{\widetilde{\mu},\, -g(\widetilde{x})\} \right) \right\|^\nu,$$

$$\theta(c,\, \widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}) = \min\left\{ \tau(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, c\left\| \left(\frac{\partial L}{\partial x}(\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}),\, h(\widetilde{x}),\, \min\{\widetilde{\mu},\, -g(\widetilde{x})\} \right) \right\|^2 \right\},$$

Similarly, practical choices of c_k, $\tau(\cdot)$, and $\theta(\cdot)$, satisfying (32) or (56), (80) or (92), and (102) or (108), respectively, can be as follows:

$$\frac{1}{c_k} = O\left(\left\| \left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k),\, h(x^k),\, \min\{\mu^k,\, -g(x^k)\} \right) \right\| \right),$$

$$\begin{aligned} \tau(c_k,\, (x^k,\, \lambda^k,\, \mu^k)) &= \theta(c_k,\, (x^k,\, \lambda^k,\, \mu^k)) \\ &= O\left(\left\| \left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k),\, h(x^k),\, \min\{\mu^k,\, -g(x^k)\} \right) \right\|^2 \right).\end{aligned}$$

6 Conclusions

We developed a perturbed augmented Lagrangian methods framework, and demonstrated its use in local analysis for the exact and inexact versions of the proximal augmented Lagrangian method and the smoothed augmented Lagrangian method. This allowed to derive the local convergence and rate-of-convergence theories for these methods under assumptions weaker than those in previous literature. We believe that some other modifications of the basic augmented Lagrangian method can also be covered by our framework, and this will be a subject of future research.

Acknowledgements This research was supported by the Russian Foundation for Basic Research Grants 19-51-12003 NNIO_a and 20-01-00106, by CNPq Grant 303913/2019-3, by FAPERJ Grant E-26/202.540/2019, by PRONEX–Optimization, and by Volkswagen Foundation.

Data sharing not applicable to this article as no data sets were generated or analyzed during the current study.

References

1. Bertsekas, D.P.: Multiplier methods: A survey. Automatica J. IFAC 12, 133–145 (1976)
2. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, USA (1982)
3. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont, MA, USA (1995)
4. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. Fundamentals of Algorithms, vol. 10. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA (2014)
5. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. 29, 161–186 (1994)
6. Börgens, E., Kanzow, C., Steck, D.: Local and global analysis of multiplier methods for constrained optimization in Banach spaces. SIAM J. Control Optim. 57, 3694–3722 (2019)
7. Chen, C., Mangasarian, O.L.: A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. 5, 97–138 (1996)
8. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. SIAM, Philadelphia (2000)
9. Contesse-Becker, L.: Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. J. Optim. Theory Appl. 79, 273–310 (1993)
10. Cui, Y., Sun, D., Toh, K.-C.: On the R-superlinear convergence of the KKT residuals generated by the augmented Lagrangian method for convex composite conic programming. Math. Program. 178, 381–415 (2019)
11. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. Second Edition. Springer, New York, USA (2014)
12. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Series in Operations Research, vol. II. Springer, New York (2003)
13. Fernández, D., Solodov, M.V.: Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. 22, 384–407 (2012)

14. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. 4, 303–320 (1969)
15. Ito, K., Kunisch, K.: The augmented Lagrangian method for equality and inequality contraints in Hilbert spaces. Math. Program. 46 (1990), 341–360.
16. Izmailov, A.F., Kurennoy, A.S.: Abstract Newtonian frameworks and their applications. SIAM J. Optim. 23, 2369–2396 (2013).
17. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization. Set-Valued Variational Anal. 21, 17–45 (2013)
18. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. 142, 591–604 (2013)
19. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: Local convergence of the method of multipliers for variational and optimization problems under the noncriticality assumption. Comput. Optim. Appl. 60, 111–140 (2015)
20. Izmailov, A.F., Solodov, M.V.: Newton-type Methods for Optimization and Variational Problems. Springer, Cham, Switzerland (2014)
21. Izmailov, A.F., Solodov, M.V.: Newton-type methods: A broader view. J. Optim. Theory Appl. 164, 577–620 (2015)
22. Kanzow, C., Steck, D.: Improved local convergence results for augmented Lagrangian methods in C^2-cone reducible constrained optimization. Math. Program. 177, 425–438 (2019)
23. Klatte, D., Tammer, K.: On the second order sufficient conditions to perturbed $\mathrm{C}^{1,1}$ optimization problems. Optimization 19, 169–180 (1988)
24. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York (1999)
25. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.): Optimization, pp. 283–298. Academic Press, New York (1972)
26. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. J. Optim. Theory Appl. 12, 555–562 (1973)
27. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. 1, 97–116 (1976)
28. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. 35, 183–238 (1993)
29. Ruszczynski, A.: Nonlinear Optimization. Princeton University Press, Princeton, NJ (2006)
30. Zhang, Y., Wu, J., Zhang, L.: The rate of convergence of proximal method of multipliers for equality constrained optimization problems. Optim. Letters 14, 1599-1613 (2020)
31. Zhang, Y., Wu, J., Zhang, L.: The rate of convergence of proximal method of multipliers for nonlinear programming. Optim. Meth. Software 35, 1022–1049 (2020)