ABSTRACT NEWTONIAN FRAMEWORKS AND THEIR APPLICATIONS

A. F. Izmailov[†] and A. S. Kurennoy[‡]

December 5, 2012

ABSTRACT

We unify and extend some Newtonian iterative frameworks developed earlier in the literature, which results in a collection of convenient tools for local convergence analysis of various algorithms under various sets of assumptions including strong metric regularity, semistability, or upper-Lipschizt stability, the latter allowing for nonisolated solutions. These abstract schemes are further applied for deriving sharp local convergence results for some constrained optimization algorithms under the reduced smoothness hypotheses. Specifically, we consider applications to the augmented Lagrangian method and to the linearly constrained Lagrangian method for problems with Lipschitzian derivatives but possibly without second derivatives, and our local convergence analysis for these methods improves all the existing theories of this kind.

Key words: generalized equation, abstract Newton scheme, augmented Lagrangian, multiplier method, local convergence, strong metric regularity, semistability, upper-Lipschitz stability.
AMS subject classifications. 49M15, 90C30, 90C33.

[*] This research is supported by the Russian Foundation for Basic Research Grants 10-01-00251, 12-01-31025, 12-01-33023.

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

1 Introduction

In this paper we extend the Newtonian iterative frameworks developed in [34, 7, 16] and in [12, Section 6C]. This development gives a convenient toolkit for local convergence analysis of various algorithms under various sets of assumptions. In particular, these schemes are further applied for deriving sharp local convergence results for some constrained optimization methods under the reduced smoothness assumptions, and in particular, to those methods which are usually not regarded as Newtonian. Specifically, in this work we concentrate on the augmented Lagrangian method and the linearly constrained Lagrangian method, though the list of potential applications is by no means limited to these algorithms.

Augmented Lagrangian methods date back to [18] and [30]; some other key references are [6, 4, 3]. These methods serve as a basis for successful software such as LANCELOT [2] and ALGENCAN [1], and their global and local convergence properties remain the subject of active research; see, e.g., [3, 14] and references therein. Traditionally, local convergence of these methods has been analyzed under the assumption of twice differentiability of the data. We extend the existing local convergence results for the augmented Lagrangian method to the case of problems with locally Lipschitzian derivatives. In particular, we extend the sharpest known result establishing local convergence under the sole second-order sufficient optimality condition, without assuming any constraint qualifications, recently derived in [14]. Observe that extensions of this kind are not always straightforward: avoiding the assumption of twice differentiability gives rise to certain peculiarities, some of which are well-known, and some will be exposed below.

Linearly constrained Lagrangian methods [31, 28, 17] are the basis of another successful solver MINOS [29]. So far, the sharpest local convergence analysis of these methods has been the one in [23], where for the twice differentiable case it was shown that the method locally converges superlinearly if the Lagrange multiplier is unique and the second order sufficient optimality condition holds. The approach taken in [23] consists of treating the method in question as a particular case of the perturbed Josephy–Newton method for solving generalized equations. Under the weaker smoothness assumptions, the perturbed semismooth Josephy–Newton method was developed in [22]. However, notably, the linearly constrained Lagrangian method does not fit this framework: this method, as well as the augmented Lagrangian method, cannot be regarded as a perturbation of the semismooth Josephy–Newton method, and in particular, semismoothness is never employed in this paper. It is also worth mentioning that the framework we develop in this paper allows not only to relax the smoothness assumptions in the result from [23], but also to show that the superlinear convergence rate estimate for the linearly constrained Lagrangian method can be naturally replaced by the quadratic one.

Observe that avoiding twice differentiability is particularly natural for the augmented Lagrangian and linearly constraint Lagrangian methods, since the iterative subproblems of both do not involve neither second derivatives of the problem data, nor any substitutes or approximations of second derivatives.

Consider the generalized equation (GE)

$$\Phi(u) + N(u) \ni 0, \tag{1.1}$$

where $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ is a continuous (single-valued) base mapping, and $N(\cdot)$ is a set-valued field multifunction from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$.

Let Π be any set of parameter values. We consider the class of methods that, given a current iterate $u^k \in \mathbb{R}^\nu$ and choosing a parameter $\pi^k \in \Pi$, generate the next iterate u^{k+1} by solving the subproblem of the form

$$\mathcal{A}(\pi^k,\, u^k,\, u) + N(u) \ni 0, \tag{1.2}$$

where for any $\pi \in \Pi$ and any $\tilde{u} \in \mathbb{R}^\nu$, the set-valued mapping $\mathcal{A}(\pi,\, \tilde{u},\, \cdot)$ from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$ is some kind of approximation of Φ around $\tilde{u}$. The required properties of this approximation will be specified below.

Several abstract Newtonian frameworks are well-known by now; see [7, 34, 26, 16, 13, 35] and [12, Section 6C]. These developments can be compared to each other and to our approach by various criteria. The frameworks in [34, 26, 13, 35] are designed for solving usual equations (that is, (1.1) with the field multifunction identically equal to $\{0\}$), while [7, 16, 12] deal with generalized equations. The convergence theorems in [34, 16, 12, 35] require continuity of the equation operator/base mapping; [26, 13] employ the the assumption of local Lipschitz continuity; [7] considers continuously differentiable and locally Lipschitzian data, but for the latter case, only a posteriori (rate of convergence) result is given.

Each of the schemes mentioned above involves some kind of approximation of the base mapping, depending on a current point and serving as a substitute of the base mapping in the subproblems of the method. In [34, 12, 35] such an approximation is a single-valued mapping; in [26, 16] it is a multifunction; and in [13] it is a family of mappings.

The local convergence analysis in [34, 35, 13, 12] relies on the uniform (with respect to the current point) local Lipschitz invertibility of the (multi)function in the approximating subproblem. In [13] this assumption (in a somewhat stronger form) appears explicitly, while in [34, 12, 35] it is implied by some other less technical assumptions concerned with solution regularity and quality of approximation. The framework in [26] is much different. In particular, it allows for solving the subproblems approximately, and this is an essential element of this development. The regularity condition is different from uniform local Lipschitz invertibility of the approximations: being global in nature, this condition is implied by the latter property in a positively homogeneous case. However, the framework in [26] does not guarantee neither the existence of the iterative sequence when the subproblems are solved exactly, nor its uniqueness.

In [34, 26, 16, 13, 35], the way of how the approximation depends on the current point is fixed, while in [7, 12] it may change from one iteration to another. In this work, we take an even more flexible approach which consists of parametrization of the iterative scheme. The resulting abstract frameworks turn out to be convenient for the analysis of optimization methods involving parameters, such as the augmented Lagrangian methods.

The framework presented in Section 2 may be viewed as an extension of the one from [12]. The development of Section 3 extends the ideas of [7]. Finally, the framework in Section 4 is a direct generalization of the one in [16].

Some words about our notation are in order. The symbol $B(u,\, \delta)$ stands for the closed ball of radius δ centered at u. The distance between a point $u \in \mathbb{R}^\nu$ and a set $U \subset \mathbb{R}^\nu$ is defined as

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

The B-differential of a mapping $F : \mathbb{R}^\nu \to \mathbb{R}^r$ at $u \in \mathbb{R}^\nu$ is the set

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

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.

According to [26, (6.6)], for a mapping $F : \mathbb{R}^\nu \to \mathbb{R}^r$ which is locally Lipschitz-continuous at $u \in \mathbb{R}^\nu$, the contingent derivative of F at u is the multifunction $CF(u)$ from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^r$, given by

$$CF(u)(v) = \{w \in \mathbb{R}^r \mid \exists\, \{t_k\} \subset \mathbb{R}_+,\ \{t_k\} \to 0+ : \{(F(u + t_k v) - F(u))/t_k\} \to w\}.$$

In particular, if F is directionally differentiable at u in the direction v then $CF(u)(v)$ is single-valued and coincides with the directional derivative of F at u in the direction v.

Furthermore, for a mapping $F : \mathbb{R}^\nu \times \mathbb{R}^p \to \mathbb{R}^r$, the partial contingent derivative (partial B-differential, partial generalized Jacobian) of F at $(u,\, v) \in \mathbb{R}^\nu \times \mathbb{R}^p$ with respect to u is the contingent derivative (B-differential, generalized Jacobian) of the mapping $F(\cdot,\, v)$ at u, which we denote by $C_u F(u,\, v)$ ($(\partial_B)_u F(u,\, v)$, $\partial_u F(u,\, v)$, respectively).

Finally, $N_U(u)$ stands for the normal cone to a closed convex set $U \subset \mathbb{R}^\nu$ at $u \in \mathbb{R}^\nu$.

2 Strongly metrically regular solutions

In this section we consider the variant of the framework from [12, Section 6C], which takes its origins in [34], and give some applications of it.

For each $\pi \in \Pi$ and $\tilde{u} \in \mathbb{R}^\nu$ define the set

$$U(\pi,\, \tilde{u}) = \{u \in \mathbb{R}^\nu \mid \mathcal{A}(\pi,\, \tilde{u},\, u) + N(u) \ni 0\}, \tag{2.1}$$

so that $U(\pi^k,\, u^k)$ is the solution set of the iteration subproblem (1.2). Since this set may contain remote points for $\tilde{u}$ arbitrarily close to a solution in question, we have to specify which of the solutions of (1.2) are allowed to be the next iterate (solutions "far away" must clearly be discarded from local analysis). In other words, we have to restrict the distance from the current iterate u^k to the next one, i.e., to an element of $U(\pi^k,\, u^k)$ that can be declared to be u^{k+1}. In the nondegenerate setting of this and the next section, it is sufficient to assume that

$$\|u^{k+1} - u^k\| \le \delta, \tag{2.2}$$

where $\delta > 0$ is fixed and small enough.

Consider the iterative scheme

$$u^{k+1} \in U(\pi^k,\, u^k) \cap B(u^k,\, \delta), \quad k = 0,\, 1,\, \ldots, \tag{2.3}$$

for some sequence $\{\pi^k\} \subset \Pi$.

The next result assumes that $\mathcal{A}$ is single-valued, and that it approximates Φ in a rather strong sense: the difference $\Phi(\cdot) - \mathcal{A}(\pi,\, \tilde{u},\, \cdot)$ is supposed to be locally Lipschitz-continuous at the solution in question with a small Lipschitz constant uniformly in $\pi \in \Pi$ and in $\tilde{u} \in \mathbb{R}^\nu$ close to this solution. This is a version of [12, Exercise 6C.4].

Theorem 2.1 *Let a mapping $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be continuous in a neighborhood of $\bar{u} \in \mathbb{R}^\nu$, and let $N(\cdot)$ be a set-valued mapping from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$. Let $\bar{u} \in \mathbb{R}^\nu$ be a solution of GE* (1.1)*. Let a set Π and a mapping $\mathcal{A} : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}^\nu$ be given.*

Assume that the following properties hold:

(i) (Strong metric regularity of the solution) There exists $\ell > 0$ such that for any $r \in \mathbb{R}^\nu$ close enough to 0*, the perturbed GE*

$$\Phi(u) + N(u) \ni r \tag{2.4}$$

has near $\bar{u}$ the unique solution $u(r)$, and the mapping $u(\cdot)$ is locally Lipschitz-continuous at 0 *with constant ℓ.*

(ii) (Precision of approximation) There exists $\bar{\varepsilon} > 0$ such that

(a) $\mathcal{A}(\pi,\, \tilde{u},\, \tilde{u}) = \Phi(\tilde{u})$ for all $\pi \in \Pi$ and all $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$.

(b) There exists a function $\omega : \Pi \times \mathbb{R}^\nu_+ \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}$ satisfying the inequality

$$q := \ell \sup \left\{\omega(\pi,\, \tilde{u},\, u^1,\, u^2) \mid \pi \in \Pi,\ \tilde{u},\, u^1,\, u^2 \in B(\bar{u},\, \bar{\varepsilon})\right\} < \frac{1}{2},$$

and such that the estimate

$$\|(\Phi(u^1) - \mathcal{A}(\pi,\, \tilde{u},\, u^1)) - (\Phi(u^2) - \mathcal{A}(\pi,\, \tilde{u},\, u^2))\| \leq \omega(\pi,\, \tilde{u},\, u^1,\, u^2)\|u^1 - u^2\| \tag{2.5}$$

holds for all $\pi \in \Pi$ and all $\tilde{u},\, u^1,\, u^2 \in B(\bar{u},\, \bar{\varepsilon})$.

Then there exists $\delta > 0$ and $\varepsilon_0 > 0$ such that for any starting point $u^0 \in B(\bar{u},\, \varepsilon_0)$ and any sequence $\{\pi^k\} \subset \Pi$, there exists the unique sequence $\{u^k\} \subset \mathbb{R}^\nu$ satisfying (2.3)*; this sequence converges to $\bar{u}$, and for all k it holds that*

$$\|u^{k+1} - \bar{u}\| \leq \frac{\ell\omega(\pi^k,\, u^k,\, u^k,\, u^{k+1})}{1 - \ell\omega(\pi^k,\, u^k,\, u^k,\, u^{k+1})}\|u^k - \bar{u}\| \leq \frac{q}{1-q}\|u^k - \bar{u}\|.$$

In particular, the rate of convergence of $\{u^k\}$ is linear. Moreover, this rate is superlinear provided $\omega(\pi^k,\, u^k,\, u^k,\, u^{k+1}) \to 0$ as $k \to \infty$, and quadratic provided $\omega(\pi^k,\, u^k,\, u^k,\, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

This specific statement follows from Theorem 3.1 below, and from [11, Theorem 1.4] saying essentially that strong metric regularity is stable subject to single-valued perturbations with a sufficiently small Lipschitz constant.

Theorem 2.1 reveals that superlinear rate of convergence is achieved if $\Phi(u^{k+1})$ is approximated by $\mathcal{A}(\pi^k, u^k, u^{k+1})$ more and more precisely as k goes to infinity, and the latter can be driven by two different factors: $\omega(\pi^k, u^k, u^k, u^{k+1})$ can be reduced either naturally, as u^k and u^{k+1} approach $\bar{u}$, or artificially, by an appropriate choice of the parameter values π^k. For instance, if Φ is differentiable near $\bar{u}$, with its derivative being continuous at $\bar{u}$, then Φ can be approximated by its linearization $\mathcal{A}(\tilde{u}, u) = \Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u})$ without any parameters. In this case, by the mean-value theorem, assumption (ii) of Theorem 2.1 holds with $\omega(\tilde{u}, u^1, u^2) = \sup_{t\in[0,1]} \|\Phi'(tu^1 + (1-t)u^2) - \Phi'(\tilde{u})\|$ which naturally vanishes as $\tilde{u}$, u^1, and u^2 tend to $\bar{u}$. With this particular choice of $\mathcal{A}$, the iterative scheme (2.3) corresponds to the commonly known Josephy–Newton method, and Theorem 2.1 covers the local convergence analysis of this method in [24, 25].

As will be seen below, the linearly constrained Lagrangian method also gains the needed approximation quality in a natural way, while on the other hand, for the augmented Lagrangian method superlinear convergence rate is achieved by an appropriate control of relevant parameters.

As demonstrated by the main result of [20], if Φ is locally Lipschitz-continuous at the solution $\bar{u}$ of GE (1.1), strong metric regularity of this solution is implied by the property defined there as CD-regularity of this solution: for each $J \in \partial\Phi(\bar{u})$ the solution $\bar{u}$ of the GE

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni 0$$

is strongly metrically regular, or equivalently, is strongly regular in the classical sense of [33].

The assumptions (i) and (ii) of Theorem 2.1 will be relaxed in Section 3, but this will not come for free: we will have to explicitly require solvability of the subproblems, and the iterative sequence will no more be necessarily unique. The solution regularity assumption will be further relaxed in Section 4, in particular allowing for nonisolated solutions, but at the price of replacing the localization condition (2.2) by a stronger one.

In the rest of this section we consider some applications of Theorem 2.1 in the case of a mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{2.6}$$

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 smooth, 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 (2.6), i.e.,

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

Then stationary points and associated Lagrange multipliers of problem (2.6) are characterized by the Karush–Kuhn-Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0, \tag{2.7}$$

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

Recall that the KKT system (2.7) can be written as the GE (1.1) with the mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), -g(x) \right), \tag{2.8}$$

and with

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+, \tag{2.9}$$

where $u = (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$.

For a feasible point $\bar{x}$ of problem (2.6), let

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

stand for the set of indices of inequality constraints active at $\bar{x}$. Furthermore, for a stationary point $\bar{x}$ of problem (2.6), let $\mathcal{M}(\bar{x})$ be the set of associated Lagrange multipliers. For any $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, define

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

Given a penalty parameter $c > 0$, the augmented Lagrangian $L_c : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ for this problem is defined by

$$L_c(x, \lambda, \mu) = f(x) + \frac{1}{2c}(\|\lambda + ch(x)\|_2^2 + \|\max\{0, \mu + cg(x)\}\|_2^2),$$

where the max operation is applied componentwise.

Given the current dual iterate $(\lambda^k, \mu^k) \in \mathbb{R}^l \times \mathbb{R}^m$, the current penalty parameter $c_k > 0$, and the current tolerance parameter $\tau_k \geq 0$, the augmented Lagrangian method generates the next primal dual-iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ as follows: x^{k+1} satisfies

$$\left\| \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \lambda^k, \mu^k) \right\| \leq \tau_k, \tag{2.10}$$

and

$$\lambda^{k+1} = \lambda^k + c_k h(x^{k+1}), \quad \mu^{k+1} = \max\{0, \mu^k + c_k g(x^{k+1})\}. \tag{2.11}$$

Throughout this section we consider the exact version of the method which corresponds to the choice $\tau_k = 0$ for all k. Therefore, the exact augmented Lagrangian method generates x^{k+1} as a stationary point of the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & L_{c_k}(x, \lambda^k, \mu^k) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \tag{2.12}$$

and it holds that

$$\begin{aligned} 0 &= \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \lambda^k, \mu^k) \\ &= f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^k + c_k h(x^{k+1})) + \sum_{i=1}^m \max\{0, c_k g_i(x^{k+1}) + \mu_i^k\} g_i'(x^{k+1}) \\ &= f'(x^{k+1}) + (h'(x^{k+1}))^{\mathrm{T}} \lambda^{k+1} + (g'(x^{k+1}))^{\mathrm{T}} \mu^{k+1} \\ &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}), \end{aligned} \tag{2.13}$$

$$0 = h(x^{k+1}) - \frac{1}{c_k}(\lambda^{k+1} - \lambda^k),$$

$$0 = \max\{-\mu^{k+1},\, c_k g(x^{k+1}) - (\mu^{k+1} - \mu^k)\} = -\min\{\mu^{k+1},\, -c_k g(x^{k+1}) + (\mu^{k+1} - \mu^k)\}.$$

The above implies that the iteration subproblem of the method can be written as (1.2) with $\mathcal{A} : (\mathbb{R}_+ \setminus \{0\}) \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}^\nu$,

$$\mathcal{A}(c,\, \tilde{u},\, u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, h(x) - \frac{1}{c}(\lambda - \tilde{\lambda}),\, -g(x) + \frac{1}{c}(\mu - \tilde{\mu}) \right),$$

where $\nu = n + l + m$, $\tilde{u} = (\tilde{x},\, \tilde{\lambda},\, \tilde{\mu})$, and with N defined in (2.9). It can be readily seen that $\mathcal{A}(c,\, \tilde{u},\, \tilde{u}) = \Phi(\tilde{u})$ and

$$\Phi(u) - \mathcal{A}(c,\, \tilde{u},\, u) = \left(0,\, \frac{1}{c}(\lambda - \tilde{\lambda}),\, -\frac{1}{c}(\mu - \tilde{\mu}) \right),$$

and hence

$$\|(\Phi(u^1) - \mathcal{A}(c,\, \tilde{u},\, u^1)) - (\Phi(u^2) - \mathcal{A}(c,\, \tilde{u},\, u^2))\| = \frac{1}{c}\|(\lambda^1 - \lambda^2,\, \mu^1 - \mu^2)\|,$$

implying that (2.5) holds with $\omega(c,\, \tilde{u},\, u^1,\, u^2) = 1/c$ and $u^1 = (x^1,\, \lambda^1,\, \mu^1)$, $u^2 = (x^2,\, \lambda^2,\, \mu^2)$ for any $c > 0$ and any $x^1,\, x^2 \in \mathbb{R}^n$, $\lambda^1,\, \lambda^2 \in \mathbb{R}^l$, $\mu^1,\, \mu^2 \in \mathbb{R}^m$. Therefore, the local convergence and rate of convergence result for the generic augmented Lagrangian method follows readily from Theorem 2.1 provided the solution $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ of the GE corresponding to the KKT system (2.7) is strongly metrically regular.

Furthermore, according to [22, Proposition 2.3, Remark 2.1], if the derivatives of f, h and g are locally Lipschitz-continuous at $\bar{x}$, then CD-regularity (and hence, strong metric regularity) of $\bar{u}$ is implied by the combination of the linear independence constraint qualification (LICQ), which consists of saying that the matrix

$$\begin{pmatrix} h'(\bar{x}) \\ g'_{A(\bar{x})}(\bar{x}) \end{pmatrix}$$

has full row rank, and the strong second-order sufficient optimality condition (SSOSC)

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi \rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{2.14}$$

where

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\}. \tag{2.15}$$

For purely equality-constrained problems CD-regularity of $(\bar{x},\, \bar{\lambda})$ is equivalent to the combination of the regularity condition $\operatorname{rank} h'(\bar{x}) = l$ and the condition that for any $H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})$ there is no $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ such that $H\xi \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}}$. The latter is stronger than the condition that $\bar{\lambda}$ is a noncritical multiplier (see the definition in Section 3), and equivalent to it in the twice differentiable case.

Putting all the ingredients together, we obtain the following.

Theorem 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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* (2.6)*, satisfying LICQ, and let SSOSC* (2.14) *hold for the associated unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then there exists $\bar{c} > 0$ and $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and any sequence $\{c_k\} \subset [\bar{c}, +\infty)$ 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 for all $k = 0, 1, \ldots$ it holds that x^{k+1} is a stationary point of problem (2.12)*, the pair $(\lambda^{k+1}, \mu^{k+1})$ satisfies* (2.11)*, and*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta; \tag{2.16}$$

this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is linear. Moreover, the rate of convergence is superlinear provided $c_k \to +\infty$, and quadratic provided $1/c_k = O(\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$.

To the best of our knowledge, this theorem is the first local convergence result for augmented Lagrangian methods without assuming twice differentiability.

As another application of Theorem 2.1, we next provide local convergence analysis for the so-called linearly constrained Lagrangian (LCL) methods which are traditionally stated for problem (2.6) with $m = n$ and $g(x) = -x$, $x \in \mathbb{R}^n$, i.e., for the case of equality constraints and simple bounds [31, 28, 17]. Therefore, we now consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ x \geq 0, \end{array} \tag{2.17}$$

and as above, neither f nor h is assumed twice differentiable.

Given the current iterate $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and the current penalty parameter $c_k \geq 0$, the next primal iterate x^{k+1} of LCL method is computed as a stationary point of the subproblem with linearized constraints

$$\begin{array}{ll} \text{minimize} & f(x) + \langle h(x),\, \lambda^k\rangle + \dfrac{c_k}{2}\|h(x)\|_2^2 \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ x \geq 0, \end{array} \tag{2.18}$$

and the next dual iterate $(\lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^n$ is defined in such a way that $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is a Lagrange multiplier corresponding to x^{k+1}. The objective function of (2.18) is the augmented Lagrangian involving the equality constraints only. In the analysis of local convergence, it is natural to assume that $c_k = c$ for all k (see the discussion in [17]).

The KKT system of problem (2.18) and the rule defining λ^{k+1} give the relations

$$\begin{gathered} \frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) + c(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1}) - (h'(x^{k+1}) - h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = 0, \\ h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0, \quad \mu^{k+1} \geq 0, \quad x^{k+1} \geq 0, \quad \langle \mu^{k+1}, x^{k+1}\rangle = 0. \end{gathered} \tag{2.19}$$

This implies that the LCL method is a particular case of (1.2) with $\mathcal{A} : \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}^\nu$,

$$\mathcal{A}(\tilde{u}, u) = \left(\begin{array}{c} \dfrac{\partial L}{\partial x}(x, \lambda, \mu) + c(h'(x))^{\mathrm{T}} h(x) - (h'(x) - h'(\tilde{x}))^{\mathrm{T}}(\lambda - \tilde{\lambda}) \\ h(\tilde{x}) + h'(\tilde{x})(x - \tilde{x}) \\ x \end{array}\right), \tag{2.20}$$

and with N defined in (2.9).

In order to apply Theorem 2.1, we have to redefine Φ by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) + c(h'(x))^{\mathrm{T}} h(x),\, h(x),\, x\right). \tag{2.21}$$

Observe that with this definition, the KKT system of problem (2.17) is still equivalent to GE (1.1) with N defined in (2.9), for any fixed c. This reflects the well-known fact that the augmented Lagrangian can be used in KKT conditions instead of the usual Lagrangian. Note that the GE in question corresponds to the following optimization problem

$$f(x) + \frac{c}{2}\|h(x)\|_2^2 \to \min, \quad h(x) = 0,\ x \geq 0. \tag{2.22}$$

It is evident that if $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ is a solution of KKT system of (2.17), satisfying LICQ, the same is true for problem (2.22). Moreover, due to the rules of Clarke's calculus [9] and the equality $h(\bar{x}) = 0$, the following relation is valid for the Lagrangian $\tilde{L} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ of problem (2.22):

$$\partial_x \frac{\partial \tilde{L}}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x}). \tag{2.23}$$

Since the sets $C_+(\bar{x},\, \bar{\mu})$ defined according to (2.15) for problems (2.17) and (2.22) coincide, relation (2.23) implies that SSOSC (2.14) also remains valid when passing from problem (2.17) to (2.22). Thus, by [22, Proposition 2.3, Remark 2.1] and [20], LICQ and SSOSC (2.14) imply that $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ is a strongly metrically regular solution of GE (1.1) with Φ defined by (2.21), and with N defined by (2.9).

It is immediate that $\mathcal{A}(\tilde{u},\, \tilde{u}) = \Phi(\tilde{u})$, and

$$\begin{aligned} \|(\Phi(u^1) - \mathcal{A}(\tilde{u},\, u^1)) - (\Phi(u^2) - \mathcal{A}(\tilde{u},\, u^2))\| &\leq \|(h'(x^1) - h'(\tilde{x}))^{\mathrm{T}}(\lambda^1 - \lambda^2)\| \\ &\quad + \|(h'(x^1) - h'(x^2))^{\mathrm{T}}(\lambda^2 - \tilde{\lambda})\| \\ &\quad + \|h(x^1) - h(x^2) - h'(\tilde{x})(x^1 - x^2)\| \\ &\leq \omega(\tilde{u},\, u^1,\, u^2)\|u^1 - u^2\|, \end{aligned}$$

where the last inequality follows by the mean-value theorem with

$$\omega(\tilde{u},\, u^1,\, u^2) = O\left(\sup_{t \in [0,\, 1]} \|h'(tx^1 + (1-t)x^2) - h'(\tilde{x})\| + \|\lambda^2 - \tilde{\lambda}\|\right). \tag{2.24}$$

In order to complete verification of assumption (ii) of Theorem 2.1, it remains to observe that $\omega(\tilde{u},\, u^1,\, u^2) \to 0$ as $\tilde{u},\, u^1,\, u^2 \to \bar{u}$, by continuity of h'.

Moreover, from (2.24), and from local Lipschitz continuity of h' at $\bar{x}$, it follows that

$$\omega(\tilde{u},\, \tilde{u},\, u) = O(\|u - \tilde{u}\|).$$

With this estimate at hand, if $\{u^k\} \subset \mathbb{R}^\nu$ is a sequence convergent to $\bar{u}$ (super)linearly, then $\|u^{k+1} - u^k\| = O(\|u^k - \bar{u}\|)$, and hence, $\omega(u^k,\, u^k,\, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

Applying Theorem 2.1, we now obtain the following result.

Theorem 2.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ 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 local solution of problem* (2.17)*, satisfying LICQ, and let SSOSC* (2.14) *hold for the associated unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.*

Then for any fixed $c > 0$ there exists $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ there exists the unique sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ such that for all $k = 0, 1, \ldots$ it holds that x^{k+1} is a stationary point of problem (2.18)*, the pair $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is an associated Lagrange multiplier, and* (2.16) *holds; this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and the rate of convergence is quadratic.*

3 Semistable solutions

The next result extends the analysis in [7].

Theorem 3.1 *Let a mapping $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be continuous in a neighborhood of $\bar{u} \in \mathbb{R}^\nu$, and let $N(\cdot)$ be a set-valued mapping from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$. Let $\bar{u} \in \mathbb{R}^\nu$ be a solution of GE* (1.1)*. Let $\mathcal{A}$ be a set-valued mapping from $\Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$, where Π is a given set. Assume that the following properties hold:*

(i) (Semistability of the solution) There exists $\ell > 0$ such that for any $r \in \mathbb{R}^\nu$, any solution $u(r)$ of the perturbed GE (2.4) *close enough to $\bar{u}$ satisfies the estimate*

$$\|u(r) - \bar{u}\| \leq \ell \|r\|.$$

(ii) (Precision of approximation) There exist $\bar{\varepsilon} > 0$ and a function $\omega : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}_+$ such that

$$q := \ell \sup \{\omega(\pi, \tilde{u}, u) \mid \pi \in \Pi, \ \tilde{u}, u \in B(\bar{u}, \bar{\varepsilon})\} < \frac{1}{3}, \tag{3.1}$$

and the estimate

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\pi, \tilde{u}, u)\} \leq \omega(\pi, \tilde{u}, u)(\|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\|) \tag{3.2}$$

holds for all $\pi \in \Pi$ and all $\tilde{u}, u \in B(\bar{u}, \bar{\varepsilon})$.

(iii) (Solvability of subproblems) For any $\varepsilon > 0$ there exists $\tilde{\varepsilon} > 0$ such that

$$U(\pi, \tilde{u}) \cap B(\bar{u}, \varepsilon) \neq \emptyset \quad \forall \, \pi \in \Pi, \ \forall \, \tilde{u} \in B(\bar{u}, \tilde{\varepsilon}). \tag{3.3}$$

Then there exists $\delta > 0$ and $\varepsilon_0 > 0$ such that for any starting point $u^0 \in B(\bar{u}, \varepsilon_0)$ and any sequence $\{\pi^k\} \subset \Pi$, the iterative scheme (2.3) *(with an arbitrary choice of u^k satisfying* (2.3)*) generates a sequence $\{u^k\} \subset \mathbb{R}^\nu$; any such sequence converges to $\bar{u}$, and for all k the following estimate is valid:*

$$\|u^{k+1} - \bar{u}\| \leq \frac{2\ell\omega(\pi^k, u^k, u^{k+1})}{1 - \ell\omega(\pi^k, u^k, u^{k+1})} \|u^k - \bar{u}\| \leq \frac{2q}{1 - q} \|u^k - \bar{u}\|.$$

In particular, the rate of convergence of $\{u^k\}$ is linear. Moreover, the rate is superlinear provided $\omega(\pi^k,\, u^k,\, u^{k+1}) \to 0$ as $k \to \infty$, and quadratic provided $\omega(\pi^k,\, u^k,\, u^{k+1}) = O(\|u^k - \bar{u}\|)$.

If, in addition, for any $\varepsilon > 0$ small enough one can choose $\tilde{\varepsilon} > 0$ in assumption (iii) in such a way that the set in (3.3) is a singleton then δ and ε_0 can be chosen in such a way that for any $u^0 \in B(\bar{u},\, \varepsilon_0)$ and any $\{\pi^k\} \subset \Pi$ the sequence $\{u^k\}$ satisfying (2.3) is unique.

Proof. According to (2.1), for all $\pi \in \Pi$ and all $\tilde{u} \in \mathbb{R}^\nu$, any $u \in U(\pi,\, \tilde{u})$ satisfies GE (2.4) with some

$$r \in \Phi(u) - \mathcal{A}(\pi,\, \tilde{u},\, u), \tag{3.4}$$

and moreover, by assumption (ii), there exist $\bar{\varepsilon} > 0$ such that

$$\|r\| \le \omega(\pi,\, \tilde{u},\, u)(\|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\|)$$

provided $\tilde{u},\, u \in B(\bar{u},\, \bar{\varepsilon})$. Moreover, by assumption (i), reducing $\bar{\varepsilon} > 0$, if necessary, we obtain that for all $\pi \in \Pi$, $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$ and $u \in B(\bar{u},\, \bar{\varepsilon}) \cap U(\pi,\, \tilde{u})$ it holds that

$$\|u - \bar{u}\| \le \ell\|r\| \le \ell\omega(\pi,\, \tilde{u},\, u)(\|u - \bar{u}\| + 2\|\tilde{u} - \bar{u}\|).$$

Taking into account (3.1), the latter implies the estimate

$$\|u - \bar{u}\| \le \frac{2\ell\omega(\pi,\, \tilde{u},\, u)}{1 - \ell\omega(\pi,\, \tilde{u},\, u)}\|\tilde{u} - \bar{u}\|. \tag{3.5}$$

By assumption (iii), there exist $\tilde{\varepsilon},\, \varepsilon \in (0,\, \bar{\varepsilon}/3]$, such that (3.3) holds. Set $\delta = \varepsilon + \tilde{\varepsilon}$. Then for all $\tilde{u} \in B(\bar{u},\, \tilde{\varepsilon})$ and $u \in B(\bar{u},\, \varepsilon)$ we have

$$\|u - \tilde{u}\| \le \|u - \bar{u}\| + \|\tilde{u} - \bar{u}\| \le \varepsilon + \tilde{\varepsilon} = \delta,$$

and therefore, (3.3) implies that

$$U(\pi,\, \tilde{u}) \cap B(\tilde{u},\, \delta) \neq \emptyset \quad \forall\, \pi \in \Pi,\ \forall\, \tilde{u} \in B(\bar{u},\, \tilde{\varepsilon}). \tag{3.6}$$

Moreover, for any $\tilde{u} \in B(\bar{u},\, \tilde{\varepsilon})$ and any $u \in B(\tilde{u},\, \delta)$ it holds that

$$\|u - \bar{u}\| \le \|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\| \le \delta + \tilde{\varepsilon} = \varepsilon + 2\tilde{\varepsilon} \le \bar{\varepsilon},$$

and hence, according to (3.5),

$$\|u - \bar{u}\| \le \frac{2\ell\omega(\pi,\, \tilde{u},\, u)}{1 - \ell\omega(\pi,\, \tilde{u},\, u)}\|\tilde{u} - \bar{u}\| \le \frac{2q}{1-q}\|\tilde{u} - \bar{u}\| \tag{3.7}$$

holds for all $\pi \in \Pi$, all $\tilde{u} \in B(\bar{u},\, \tilde{\varepsilon})$, and all $u \in U(\pi,\, \tilde{u}) \cap B(\tilde{u},\, \delta)$, where, according to (3.1), $2q/(1-q) < 1$.

Relations (3.6) and (3.7) evidently imply the needed conclusions with $\varepsilon_0 = \tilde{\varepsilon}$. In particular, if ε and $\tilde{\varepsilon}$ are chosen in such a way that $\tilde{\varepsilon} \le \varepsilon$ and $U(\pi,\, \tilde{u}) \cap B(\bar{u},\, \varepsilon)$ is a singleton for all

$\pi \in \Pi$ and all $\tilde{u} \in B(\bar{u}, \tilde{\varepsilon})$, one can additionally guaranty that the set in (3.6) contains only one element for all such π and $\tilde{u}$. ∎

One can see from the proof of Theorem 3.1 that assumption (ii) can be somewhat weakened: (3.1) can be replaced by

$$q := \ell \sup \{\omega(\pi, \tilde{u}, u) \mid \pi \in \Pi, \ u \in U(\pi, \tilde{u}), \ \tilde{u}, u \in B(\bar{u}, \varepsilon)\} < \frac{1}{3},$$

and instead of (3.2), it is sufficient to assume that the estimate

$$\|w\| \leq \omega(\pi, \tilde{u}, u)(\|u - \tilde{u}\| + \|\tilde{u} - \bar{u}\|)$$

holds for all $\pi \in \Pi$, all $w \in \Phi(u) - \mathcal{A}(\pi, \tilde{u}, u) \cap N(u)$, and all $\tilde{u}, u \in B(\bar{u}, \varepsilon)$. Taking this into account, Theorem 3.1 covers the result obtained in [23, Theorem 2.1] for the perturbed Josephy–Newton method, corresponding to $\mathcal{A}(\tilde{u}, u) = \Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u}) + \Omega(\tilde{u}, u - \tilde{u})$, where inexactness is characterized by a set-valued mapping Ω from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$.

Observe that unlike in Theorem 2.1, the weaker assumptions of Theorem 3.1 do not guaranty uniqueness of the sequence generated by the method, in general.

Due to the fact that $\mathcal{A}$ in Theorem 3.1 is allowed to be multivalued, this theorem can serve as a tool for the analysis of the inexact versions of the augmented Lagrangian method, with positive tolerance parameters τ_k. We now verify the assumptions of Theorem 3.1 for this method applied to problem (2.6). To that end, let Φ and N be defined by (2.8) and (2.9), respectively. Semistability of the solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of the corresponding GE (1.1) (assumption (i)) admits the following characterization which can be easily derived from [26, Theorem 8.11].

Proposition 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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* (2.6)*, and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier.*

Then $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ is a semistable solution of GE (1.1) *with Φ and N be defined by* (2.8) *and* (2.9)*, respectively, if and only if there is no nontrivial triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying the system*

$$\begin{gathered} d + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0, \\ \zeta_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g_i'(\bar{x}), \xi\rangle = 0, \ i \in A_0(\bar{x}, \bar{\mu}), \\ \zeta_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0 \end{gathered} \tag{3.8}$$

with some $d \in C_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(\xi)$.

From Proposition 3.1 it evidently follows that semistability is equivalent to the combination of two properties: the strict Mangasarian–Fromovitz constrain qualification (SMFCQ), which consists of saying that $(\bar{\lambda}, \bar{\mu})$ is the unique Lagrange multiplier associated with the stationary point $\bar{x}$, and the so-called *noncriticality* of the multiplier $(\bar{\lambda}, \bar{\mu})$, which consists of saying that there is no triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, with $\xi \neq 0$, satisfying the system (3.8) with some $d \in C_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(\xi)$. Furthermore, as mentioned in [21, Remark 2.1], noncriticality of the multiplier is implied by the following SOSC first introduced in [27]:

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{3.9}$$

where

$$\begin{aligned} C(\bar{x}) &= \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}), \xi\rangle \leq 0\} \\ &= \left\{\xi \in \mathbb{R}^n \,\middle|\, h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0\right\} \end{aligned} \tag{3.10}$$

is the critical cone of problem (2.6) at $\bar{x}$. At the same time, unlike in the twice differentiable case, SOSC (3.9) is not necessary for semistability of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ even if $\bar{x}$ is a local solution of (2.6) (see, e.g., the example in [21, Remark 3]).

Suppose that the tolerance parameter τ_k is chosen as a function of the current iterate:

$$\tau_k = \tau(x^k, \lambda^k, \mu^k) \tag{3.11}$$

with some $\tau : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$. Then the iteration of the inexact augmented Lagrangian method given by (2.10), (2.11), can be written as (1.2) with $\mathcal{A} : (\mathbb{R}_+ \setminus \{0\}) \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to 2^{\mathbb{R}^\nu}$,

$$\mathcal{A}(c, \tilde{u}, u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu) + B(0, \tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu})),\ h(x) - \frac{1}{c}(\lambda - \tilde{\lambda}),\ -g(x) + \frac{1}{c}(\mu - \tilde{\mu})\right). \tag{3.12}$$

Suppose further that τ satisfies

$$\tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leq \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\|(\tilde{x} - \bar{x}, \tilde{\lambda} - \bar{\lambda}, \tilde{\mu} - \bar{\mu}))\| \quad \forall\, (\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m, \tag{3.13}$$

where $\theta : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ is a function satisfying $\theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to 0$ as $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. In this case, for $\mathcal{A}$ defined in (3.12), assumption (ii) of Theorem 3.1 holds with $\Pi = [\bar{c}, +\infty)$ for a sufficiently large $\bar{c} > 0$, and with $\omega(c, \tilde{u}, u) = \max\{1/c, \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\}$. Constructive and practically relevant choices of a function τ with the needed properties can be based on residuals of the KKT system (2.7). For instance, assuming Lipschitz continuity of the derivatives of f, h and g, one can take any τ such that

$$\tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) = o\left(\left\|\left(\frac{\partial L}{\partial x}(\tilde{x}, \tilde{\lambda}, \tilde{\mu}),\ h(\tilde{x}),\ \min\{\tilde{\mu}, -g(\tilde{x})\}\right)\right\|\right). \tag{3.14}$$

Verifying assumption (iii) is a more subtle issue; to establish it we make use of the following fact.

Proposition 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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* (2.6)*, and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier satisfying*

$$\liminf_{t \to +0} \frac{L(\bar{x} + t\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})}{t^2/2} > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{3.15}$$

Then there exist $\bar{c} > 0$ and $\bar{\delta} > 0$ such that the following assertions are valid:

(a) There exist $\bar{\gamma} > 0$ and $\bar{\varepsilon} > 0$ such that for all $c \geq \bar{c}$

$$\begin{array}{c} L_c(x,\, \lambda,\, \mu) \geq L_c(\bar{x},\, \lambda,\, \mu) + \bar{\gamma}\|x - \bar{x}\|^2 \\ \forall\, x \in B(\bar{x},\, \bar{\delta}),\ \forall\, (\lambda,\, \mu) \in \mathcal{M}(\bar{x}) \text{ such that } \|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leq \bar{\varepsilon}. \end{array} \tag{3.16}$$

(b) For any $\delta \in (0,\, \bar{\delta}]$ there exists $\varepsilon(\delta) > 0$ such that such for any $c \geq \bar{c}$ and any $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ satisfying $\|(\tilde{\lambda} - \bar{\lambda},\, \tilde{\mu} - \bar{\mu})\| \leq \varepsilon(\delta)$, any global solution x of the optimization problem

$$\begin{array}{ll} \text{minimize} & L_c(x,\, \tilde{\lambda},\, \tilde{\mu}) \\ \text{subject to} & x \in B(\bar{x},\, \bar{\delta}) \end{array} \tag{3.17}$$

satisfies $\|x - \bar{x}\| \leq \delta$.

Proof. We first prove assertion (a). Both the assertion and its proof extend their counterparts in [14, Proposition 3.1] to the case when the data is not assumed twice differentiable.

For any fixed $c > 0$, any $x \in \mathbb{R}^n$, and any $(\lambda,\, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$, it holds

$$L_c(x,\, \lambda,\, \mu) = L(x,\, \lambda,\, \mu) + \frac{1}{2c}\|\lambda\|_2^2 + \frac{c}{2}\|h(x)\|_2^2 + \frac{1}{2c}\|\max\{0,\, \mu + cg(x)\}\|_2^2 - \langle \mu,\, g(x)\rangle. \tag{3.18}$$

If $x \in \mathbb{R}^n$ is close enough to $\bar{x}$, and if $(\lambda,\, \mu)$ is close enough to $(\bar{\lambda},\, \bar{\mu})$, for any $i \in A_+(\bar{x},\, \bar{\mu})$ we have that $\mu_i > -cg_i(x)$ and

$$\frac{1}{2c}\left(\max\{0,\, \mu_i + cg_i(x)\}\right)^2 - \mu_i g_i(x) = \frac{1}{2c}\mu_i^2 + \frac{c}{2}(g_i(x))^2, \tag{3.19}$$

This relation also holds for any $i \in A_0(\bar{x},\, \bar{\mu})$ such that $\mu_i \geq -cg_i(x)$, further implying

$$\frac{1}{2c}\left(\max\{0,\, \mu_i + cg_i(x)\}\right)^2 - \mu_i g_i(x) \geq \frac{1}{2c}\mu_i^2 + \frac{c}{2}\left(\max\{0,\, g_i(x)\}\right)^2. \tag{3.20}$$

On the other hand, if $\mu \geq 0$, for any $i \in A_0(\bar{x},\, \bar{\mu})$ such that $\mu_i < -cg_i(x)$, it holds that $g_i(x) < 0$, and hence,

$$\frac{1}{2c}\left(\max\{0,\, \mu_i + cg_i(x)\}\right)^2 - \mu_i g_i(x) = -\mu_i g_i(x) \geq \frac{1}{2c}\mu_i^2 = \frac{1}{2c}\mu_i^2 + \frac{c}{2}\left(\max\{0,\, g_i(x)\}\right)^2. \tag{3.21}$$

Finally, for $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$, if $(\lambda,\, \mu) \in \mathcal{M}(\bar{x})$ it holds that $\mu_i = 0$, and hence, for all x sufficiently close to $\bar{x}$

$$\frac{1}{2c}\left(\max\{0,\, \mu_i + cg_i(x)\}\right)^2 - \mu_i g_i(x) = 0. \tag{3.22}$$

At the same time, for any $i \in A(\bar{x})$ and any $\mu \geq 0$ we have

$$\frac{1}{2c}\left(\max\{0,\, \mu_i + cg_i(\bar{x})\}\right)^2 - \mu_i g_i(\bar{x}) = \frac{1}{2c}\mu_i^2. \tag{3.23}$$

From (3.18)–(3.23) we obtain that for any fixed $c > 0$ the following chain of relations is valid for all $x \in \mathbb{R}^n$ close enough to $\bar{x}$, and all $(\lambda,\, \mu) \in \mathcal{M}(\bar{x})$ close enough to $(\bar{\lambda},\, \bar{\mu})$:

$$\begin{aligned}
L_c(x,\, \lambda,\, \mu) - L_c(\bar{x},\, \lambda,\, \mu) &= L(x,\, \lambda,\, \mu) - L(\bar{x},\, \lambda,\, \mu) + \frac{c}{2}\|h(x)\|_2^2 \\
&\quad + \frac{1}{2c}\|\max\{0,\, \mu + cg(x)\}\|_2^2 - \langle \mu,\, g(x)\rangle \\
&\quad - \left(\frac{c}{2}\|h(\bar{x})\|_2^2 + \frac{1}{2c}\|\max\{0,\, \mu + cg(\bar{x})\}\|_2^2 - \langle \mu,\, g(\bar{x})\rangle\right) \\
&\geq L(x,\, \lambda,\, \mu) - L(\bar{x},\, \lambda,\, \mu) + \frac{c}{2}\|h(x)\|_2^2 \\
&\quad + \frac{c}{2}\sum_{i \in A_+(\bar{x},\, \bar{\mu})} (g_i(x))^2 + \frac{c}{2}\sum_{i \in A_0(\bar{x},\, \bar{\mu})} (\max\{0,\, g_i(x)\})^2 .
\end{aligned} \tag{3.24}$$

Now, combining (3.24) with Lemma 5.2 in Appendix, we obtain the existence of $\bar{c} > 0$, $\bar{\delta} > 0$ and $\bar{\varepsilon} > 0$ such that

$$\begin{gathered}
L_{\bar{c}}(x,\, \lambda,\, \mu) \geq L_{\bar{c}}(\bar{x},\, \lambda,\, \mu) + \bar{\gamma}\|x - \bar{x}\|^2 \\
\forall\, x \in B(\bar{x},\, \bar{\delta}),\ \forall\, (\lambda,\, \mu) \in \mathcal{M}(\bar{x}) \text{ such that } \|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \leq \bar{\varepsilon}.
\end{gathered}$$

The needed property (3.16) for $c \geq \bar{c}$ now follows from the fact that for any fixed $x \in \mathbb{R}^n$ and $(\lambda,\, \mu) \in \mathcal{M}(\bar{x})$ the difference $L_c(x,\, \lambda,\, \mu) - L_c(\bar{x},\, \lambda,\, \mu)$ is nondecreasing in $c > 0$. The latter can be checked directly.

We proceed with proving (b). Let $\bar{c} > 0$ and $\bar{\delta} > 0$ be chosen as in (a). Consider any sequences $\{c_k\} \subset \mathbb{R}_+$, $\{x^k\} \subset \mathbb{R}^n$ and $\{(\tilde{\lambda}^k,\, \tilde{\mu}^k)\} \subset \mathbb{R}^l \times \mathbb{R}^m$ such that for all k the inequality $c_k \geq \bar{c}$ is valid, x^k is a global solution of problem (3.17) with $c = c_k$ and $(\lambda,\, \mu) = (\tilde{\lambda}^k,\, \tilde{\mu}^k)$, and $\{(\tilde{\lambda}^k,\, \tilde{\mu}^k)\}$ converges to $(\bar{\lambda},\, \bar{\mu})$. Since $\{x^k\}$ is contained in the compact set $B(\bar{x},\, \bar{\delta})$, this sequence has a limit point $\hat{x} \in B(\bar{x},\, \bar{\delta})$, and we need to show that $\hat{x} = \bar{x}$; this will mean that $\{x^k\}$ converges to $\bar{x}$.

Without loss of generality, suppose that the entire $\{x^k\}$ converges to $\hat{x}$. If the sequence $\{c_k\}$ is unbounded, the needed result follows immediately from Proposition 5.1 in Appendix, since either of the conditions 1 or 2 implies that $\bar{x}$ is a strict local solution of problem (2.6). Therefore, we can suppose that $\{c_k\}$ converges to some $\hat{c} \geq \bar{c}$.

Now consider (3.17) as a parametric optimization problem, with c and $(\tilde{\lambda},\, \tilde{\mu})$ playing the role of parameters with the base values $\hat{c}$ and $(\bar{\lambda},\, \bar{\mu})$, respectively. According to the choice of $\bar{c} > 0$ and $\bar{\delta} > 0$, the point $\bar{x}$ is a strict (global) solution of problem (3.17) for these base values of parameters. The needed result now follows from the well-known facts regarding stability of strict local solutions of optimization problems (see, e.g., [5, Theorem 3.1]). ∎

Remark 3.1 The expression in the left-hand side of (3.15) is the lower second-order directional derivative of the function $L(\cdot,\, \bar{\lambda},\, \bar{\mu})$ at $\bar{x}$ in a direction ξ (recall that $\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = 0$). Condition (3.15) is always valid under SOSC (3.9). Indeed, due to the compactness of $\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, SOSC (3.9) implies the existence of $\gamma > 0$ such that

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle \geq \gamma\|\xi\|^2 \quad \forall\, \xi \in C(\bar{x}).$$

Consider any $\xi \in C(\bar{x}) \setminus \{0\}$ and any sequence $\{t_k\} \to 0+$. By an appropriate mean-value theorem (e.g., [19, Theorem 2.3]),

$$\frac{L(\bar{x} + t_k\xi,\, \bar{\lambda},\, \bar{\mu}) - L(\bar{x},\, \bar{\lambda},\, \bar{\mu})}{t_k^2/2} = \langle H_k\xi,\, \xi\rangle, \tag{3.25}$$

where $H_k \in \partial_x \frac{\partial L}{\partial x}(\bar{x} + \tilde{t}_k\xi,\, \bar{\lambda},\, \bar{\mu})$ for some $\tilde{t}_k \in (0,\, t_k)$. Since all matrices in the generalized Jacobian are bounded by the Lipschitz constant of the mapping in question, the sequence $\{H_k\}$ is bounded, and by upper semicontinuity of generalized Jacobian, any accumulation point of this sequence belongs to $\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Therefore, any accumulation point of the bounded sequence in the right-hand (and hence, left-hand) side of (3.25) is no less than $\gamma\|\xi\|^2$, and therefore, (3.15) holds.

The converse implication is not valid as can be seen from Example 3.1 below.

Remark 3.2 It can be easily seen that condition (3.15) is equivalent to the following quadratic growth condition of the Lagrangian over the critical cone: there exist $\gamma > 0$ and $\rho > 0$ such that

$$L(\bar{x} + \xi,\, \bar{\lambda},\, \bar{\mu}) \geq L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \gamma\|\xi\|^2 \quad \forall\, \xi \in C(\bar{x}) \cap B(0,\, \rho).$$

Moreover, under SMFCQ, condition (3.15) is implied by the following customary quadratic growth condition for problem (2.6) at $\bar{x}$: there exist $\gamma > 0$ and $\rho > 0$ such that

$$f(x) \geq f(\bar{x}) + \gamma\|x - \bar{x}\|^2 \quad \forall\, x \in B(\bar{x},\, \rho) \text{ such that } h(x) = 0,\, g(x) \leq 0. \tag{3.26}$$

To see this, consider any $\xi \in C(\bar{x}) \setminus \{0\}$ and any sequence $\{t_k\} \to 0+$. Recall that SMFCQ is equivalent to MFCQ at $\bar{x}$ for the constraint system

$$h(x) = 0, \quad g_{A_+(\bar{x},\, \bar{\mu})}(x) = 0, \quad g_{A_0(\bar{x},\, \bar{\mu})}(x) \leq 0.$$

Therefore, according to (3.10), ξ is (inner) tangential to the set given by this constraint system at $\bar{x}$ (see, e.g., [8, Corollary 2.91]), and in particular, there exists a sequence $\{\xi^k\} \subset \mathbb{R}^n$ convergent to ξ and such that for all k

$$h(\bar{x} + t_k\xi^k) = 0, \quad g_{A_+(\bar{x},\, \bar{\mu})}(\bar{x} + t_k\xi^k) = 0, \quad g_{A_0(\bar{x},\, \bar{\mu})}(\bar{x} + t_k\xi^k) \leq 0.$$

Since $\bar{\mu}_i = 0$ for all $i \in \{1,\, \ldots,\, m\} \setminus A_+(\bar{x},\, \bar{\mu})$, for all k large enough it then holds that

$$L(\bar{x} + t_k\xi^k,\, \bar{\lambda},\, \bar{\mu}) - L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = f(\bar{x} + t_k\xi^k) - f(\bar{x}) \geq \gamma\|\xi^k\|^2 t_k^2,$$

where the last inequality is by (3.26). Employing Lemma 5.1 in Appendix, this implies (3.15).

According to [27, Theorem 1], SOSC (3.9) implies the quadratic growth condition (3.26) with some $\gamma > 0$ and $\rho > 0$. Observe, however, that unlike in the twice differentiable case (see, e.g., [8, Theorem 3.70]), the converse implication is not valid even for the case of pure equality constraints satisfying the regularity condition, and even if the associated unique Lagrange multiplier is noncritical, as we show by the following example.

Example 3.1 Let $n = 2$, $l = 1$, $m = 0$, $f(x) = (x_1 + x_2 - \sqrt{x_1^2 + x_2^2})^2/2$, and $h(x) = x_1 + x_2$. Then $\bar{x} = (0,\, 0)$ is the unique solution of optimization problem (2.6), and $\bar{\lambda} = 0$ is the unique associated Lagrange multiplier which is noncritical, as can be verified directly.

Taking any sequence $\{x^k\} \subset \mathbb{R}^2$ such that $\{x^k\} \to 0$ and $x_1^k = x_2^k > 0$ for all k, we have that $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda})$ is differentiable at x^k and

$$\frac{\partial^2 L}{\partial x^2}(x^k,\, \bar{\lambda}) = \frac{\sqrt{2}-1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{\sqrt{2}-1}{2} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Therefore, this matrix, which we denote by H, belongs to $\partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})$. At the same time, for $\xi = (1,\, -1) \in \ker h'(\bar{x})$ we have

$$\langle H\xi,\, \xi \rangle = -2(\sqrt{2} - 1) < 0.$$

Thus, SOSC (3.9) does not hold, although the quadratic growth condition is evidently valid.

Combining assertion (b) of Proposition 3.2 with Proposition 5.2 in Appendix and taking in account Remark 3.2, we see that assumption (iii) of Theorem 3.1 holds for the exact (and even more so, inexact) augmented Lagrangian method if SMFCQ and the quadratic growth condition are satisfied.

Theorem 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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* (2.6)*, satisfying SMFCQ, let the associated unique Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be noncritical, and let there exist $\gamma > 0$ and $\rho > 0$ such that the quadratic growth condition* (3.26) *holds. Let $\tau : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ be a function satisfying $\tau(x,\, \lambda,\, \mu) = o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|)$.*

Then there exists $\bar{c} > 0$ and $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and any sequence $\{c_k\} \subset [\bar{c},\, +\infty)$, there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying for all $k = 0,\, 1,\, \ldots$ condition (2.10) *with τ_k defined according to* (3.11)*, and conditions* (2.11) *and* (2.16)*; any such sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is linear. Moreover, the rate of convergence is superlinear provided $c_k \to +\infty$. If, in addition, $\tau_k = O(\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|^2)$, then the rate of convergence is quadratic provided $1/c_k = O(\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$.*

For purely equality-constrained problems with twice differentiable data, semistability is equivalent to strong regularity, which is further equivalent to the constraints regularity condition and noncriticality of the multiplier, and all this is covered by the development in Section 2 (apart from the possibility to solve the subproblems inexactly). However, in the absence of twice differentiability, semistability can be weaker than strong metric regularity, and Theorem 3.2 is fully relevant even in the special case of pure equality constraints. Indeed, for any $t > 0$, the Lagrange optimality system for the problem in Example 3.1 with the right-hand side-perturbed constraint $x_1 + x_2 = t$ has two solutions $(x,\, \lambda) = ((t,\, 0),\, 0)$ and $(x,\, \lambda) = ((0,\, t),\, 0)$, and therefore, the primal-dual solution $(\bar{x},\, \bar{\lambda})$ of the unperturbed problem cannot be strongly metrically regular.

We proceed with LCL method applied to problem (2.17). Therefore, let Φ and $\mathcal{A}$ be defined by (2.21) and (2.20), respectively. Similarly to the discussion following (2.22), it can be seen that SMFCQ and SOSC (3.9) imply semistability of $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ as a solution of GE (1.1) with such Φ and N defined by (2.9).

Furthermore, by the mean-value theorem,

$$\|\Phi(u) - \mathcal{A}(\tilde{u}, u)\| = O\left(\sup_{t\in[0,1]} \|h'(tx + (1-t)\tilde{x}) - h'(\tilde{x})\|\right) \|u - \tilde{u}\|.$$

Therefore, assumption (ii) of Theorem 3.1 is satisfied, and it remains to verify assumption (iii).

Proposition 3.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (2.17)*, satisfying SMFCQ, and let SOSC* (3.9) *hold for the associated unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.*

Then for any $c \geq 0$ and any $\varepsilon > 0$ there exists $\tilde{\varepsilon} > 0$ such that for any $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ satisfying $\|(\tilde{x} - \bar{x}, \tilde{\lambda} - \bar{\lambda})\| \leq \tilde{\varepsilon}$ there exists a stationary point x of the optimization problem

$$\begin{array}{ll} \textit{minimize} & f(x) + \langle h(x), \tilde{\lambda}\rangle + \dfrac{c}{2}\|h(x)\|_2^2 \\ \textit{subject to} & h(\tilde{x}) + h'(\tilde{x})(x - \tilde{x}) = 0, \; x \geq 0, \end{array} \tag{3.27}$$

satisfying $\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\| \leq \varepsilon$ with any pair $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^n$ such that $(\lambda - \tilde{\lambda}, \mu)$ is a Lagrange multiplier associated with x.

Proof. Problem (3.27) can be viewed as a parametric optimization problem in which $\tilde{x}$ and $\tilde{\lambda}$ play the role of parameters with the base values $\bar{x}$ and $\bar{\lambda}$, respectively. It can be checked directly that $\bar{x}$ is a stationary point of problem

$$\begin{array}{ll} \text{minimize} & f(x) + \langle h(x), \bar{\lambda}\rangle + \frac{c}{2}\|h(x)\|_2^2 \\ \text{subject to} & h(\bar{x}) + h'(\bar{x})(x - \bar{x}) = 0, \; x \geq 0, \end{array} \tag{3.28}$$

which is (3.27) for these base values of parameters. Also, it is easy to see that SMFCQ holds at $\bar{x}$ as a stationary problem of (3.28), and that the associated unique Lagrange multiplier is $(0, \bar{\mu})$. Letting $\bar{L} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}$ be the Lagrangian of problem (3.28), that is,

$$\bar{L}(x, \lambda, \mu) = f(x) + \langle h(x), \bar{\lambda}\rangle + \frac{c}{2}\|h(x)\|_2^2 + \langle \lambda, h(\bar{x}) + h'(\bar{x})(x - \bar{x})\rangle - \langle \mu, x\rangle,$$

similarly to (2.23) we derive

$$\partial_x \frac{\partial \bar{L}}{\partial x}(\bar{x}, 0, \bar{\mu}) = \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x}).$$

Since the critical cones of problems (2.17) and (3.28) coincide, the last equality implies that SOSC holds at a stationary point $\bar{x}$ of problem (3.28) for the associated Lagrange multiplier $(0, \bar{\mu})$. In particular, $\bar{x}$ is a strict local solution of (3.28). Since SMFCQ implies the usual Mangasarian–Fromovitz constraint qualification (MFCQ), it then follows (e.g., from Robinson's stability theorem [32] and from [5, Theorem 3.1]) that for any $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, problem (3.27) has a local solution $x(\tilde{x}, \tilde{\lambda})$ such that $x(\tilde{x}, \tilde{\lambda}) \to \bar{x}$ as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$. Furthermore, since MFCQ is stable under small perturbations (see, e.g., [8, Remark 2.88]), for all $(\tilde{x}, \tilde{\lambda})$ close enough to $(\bar{x}, \bar{\lambda})$ the local solution $x = x(\tilde{x}, \tilde{\lambda})$ of problem (3.27) satisfies MFCQ and consequently, is a stationary point of this problem (see, e.g., [8, Theorem 3.9]).

Finally, by the argument similar to the one in the proof of Proposition 5.2 in Appendix, it can be seen that for any choice of $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^n$ such that $(\lambda - \tilde{\lambda}, \mu)$ is a Lagrange multiplier associated with $x(\tilde{x}, \tilde{\lambda})$, it holds that $(\lambda, \mu) \to (\bar{\lambda}, \bar{\mu})$ as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$.

∎

Applying Theorem 3.1 and Proposition 3.3, similarly to Theorem 2.3 we obtain the following.

Theorem 3.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ 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 local solution of problem* (2.17)*, satisfying SMFCQ, and let SOSC* (3.9) *hold for the associated unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$.*

Then for any fixed $c > 0$ there exists $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfying for all $k = 0, 1, \ldots$ conditions (2.19), (2.16)*; any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is quadratic.*

4 Possibly nonisolated upper-Lipschitz stable solutions

In this section we extend the iterative framework from [16].

In order to tackle the case when a solution in question is not necessarily isolated, one has to drive δ in (2.2) to zero along the iterations, and with some specific speed. This is the essence of the stabilization mechanism needed in this case. Specifically, for an arbitrary but fixed $\sigma > 0$ define the set

$$U^\sigma(\pi, \tilde{u}) = \{u \in U(\pi, \tilde{u}) \mid \|u - \tilde{u}\| \leq \sigma \operatorname{dist}(\tilde{u}, \bar{U})\}, \tag{4.1}$$

where $\bar{U}$ is the solution set of GE (1.1), and consider the iterative scheme

$$u^{k+1} \in U^\sigma(\pi^k, u^k), \quad k = 0, 1, \ldots. \tag{4.2}$$

Theorem 4.1 *Let a mapping $\Phi : \mathbb{R}^\nu \to \mathbb{R}^\nu$ be continuous in a neighborhood of $\bar{u} \in \mathbb{R}^\nu$, and let $N(\cdot)$ be a set-valued mapping from $\mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$. Let $\bar{U}$ be the solution set of GE* (1.1)*, let $\bar{u} \in \bar{U}$, and assume that the set $\bar{U} \cap B(\bar{u}, \varepsilon)$ is closed for any $\varepsilon > 0$ small enough. For each k, let $\mathcal{A}$ be a set-valued mapping from $\Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu$ to the subsets of $\mathbb{R}^\nu$. Assume that the following properties hold with some fixed $\sigma > 0$:*

(i) (Upper Lipschitzian behavior of the solution under canonical perturbations) There exists $\ell > 0$ such that for any $r \in \mathbb{R}^\nu$, any solution $u(r)$ of the perturbed GE (2.4) close enough to $\bar{u}$ satisfies the estimate

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

(ii) (Precision of approximation) There exist $\bar{\varepsilon} > 0$ and a function $\omega : \Pi \times \mathbb{R}^\nu \times \mathbb{R}^\nu \to \mathbb{R}_+$ such that

$$q := \ell \sup\{\omega(\pi,\, \tilde{u},\, u) \mid \pi \in \Pi,\ \tilde{u} \in B(\bar{u},\, \bar{\varepsilon}),\ \|u - \tilde{u}\| \leq \sigma \operatorname{dist}(\tilde{u},\, \bar{U})\} < 1 \tag{4.3}$$

and the estimate

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\pi,\, \tilde{u},\, u)\} \leq \omega(\pi,\, \tilde{u},\, u) \operatorname{dist}(\tilde{u},\, \bar{U}) \tag{4.4}$$

holds for all $\pi \in \Pi$ and all $(\tilde{u},\, u) \in \mathbb{R}^\nu \times \mathbb{R}^\nu$ satisfying $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$, $\|u - \tilde{u}\| \leq \sigma \operatorname{dist}(\tilde{u},\, \bar{U})$.

(iii) (Solvability of subproblems and localization condition) For any $\pi \in \Pi$ and any $\tilde{u} \in \mathbb{R}^\nu$ close enough to $\bar{u}$ the set $U^\sigma(\pi,\, \tilde{u})$ defined by (2.1), (4.1) is nonempty.

Then for any $\varepsilon > 0$ there exists $\varepsilon_0 > 0$ such that for any starting point $u^0 \in B(\bar{u},\, \varepsilon_0)$ and any sequence $\{\pi^k\} \subset \Pi$, the iterative scheme (4.2) (with an arbitrary choice of u^k satisfying (4.2)) generates a sequence $\{u^k\} \subset B(\bar{u},\, \varepsilon)$; any such sequence converges to some $u^* \in \bar{U}$, and for all k the following estimates are valid:

$$\|u^{k+1} - u^*\| \leq \frac{\sigma \ell \omega(\pi^k,\, u^k,\, u^{k+1})}{1-q} \operatorname{dist}(u^k,\, \bar{U}) \leq \frac{\sigma q}{1-q} \operatorname{dist}(u^k,\, \bar{U}), \tag{4.5}$$

$$\operatorname{dist}(u^{k+1},\, \bar{U}) \leq \ell \omega(\pi^k,\, u^k,\, u^{k+1}) \operatorname{dist}(u^k,\, \bar{U}) \leq q \operatorname{dist}(u^k,\, \bar{U}). \tag{4.6}$$

In particular, the rates of convergence of $\{u^k\}$ to u^* and of $\{\operatorname{dist}(u^k,\, \bar{U})\}$ to zero are superlinear provided $\omega(\pi^k,\, u^k,\, u^{k+1}) \to 0$ as $k \to \infty$. Moreover, both these rates are quadratic provided $\omega(\pi^k,\, u^k,\, u^{k+1}) = O(\operatorname{dist}(u^k,\, \bar{U}))$.

Proof. According to (2.1), (4.1), for any $\pi \in \Pi$ and any $\tilde{u} \in \mathbb{R}^\nu$, any $u \in U^\sigma(\pi,\, \tilde{u})$ satisfies GE (2.4) with some $r \in \mathbb{R}^\nu$ satisfying (3.4), and moreover, if $\tilde{u} \in B(\bar{u},\, \bar{\varepsilon})$, assumption (ii) implies the estimate

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

Since for any $\varepsilon > 0$ it holds that $\|u - \bar{u}\| < \varepsilon$ provided $\tilde{u}$ is close enough to $\bar{u}$, assumptions (i) and (iii) further imply that for any $\varepsilon > 0$ there exist $\tilde{\varepsilon} \in (0,\, \min\{\varepsilon,\, \bar{\varepsilon}\}]$ and $\ell > 0$ such that for all $\pi \in \Pi$

$$U^\sigma(\pi,\, \tilde{u}) \neq \emptyset, \quad \operatorname{dist}(u,\, \bar{U}) \leq \ell \omega(\pi,\, \tilde{u},\, u) \operatorname{dist}(\tilde{u},\, \bar{U}) \quad \forall\, u \in U^\sigma(\pi,\, \tilde{u}),\ \forall\, \tilde{u} \in B(\bar{u},\, \tilde{\varepsilon}), \tag{4.7}$$

and the set $\bar{U} \cap B(\bar{u},\, 2\tilde{\varepsilon})$ is closed. Set

$$\varepsilon_0 = \frac{\tilde{\varepsilon}(1-q)}{\sigma + 1 - q}. \tag{4.8}$$

According to (4.3) it holds that $\varepsilon_0 > 0$. We now prove by induction that if $u^0 \in B(\bar{u},\, \varepsilon_0)$ then for any sequence $\{\pi^k\} \subset \Pi$ iterative scheme (4.2) successfully produces a sequence $\{u^k\} \subset \mathbb{R}^\nu$ regardless of the choice of $u^{k+1} \in U^\sigma(\pi^k,\, u^k)$ at each iteration, and that for any such sequence it holds that $u^k \in B(\bar{u},\, \tilde{\varepsilon}) \subset B(\bar{u},\, \varepsilon)$ for all k.

Indeed, since (4.8) evidently implies the inequality $\varepsilon_0 \leq \tilde{\varepsilon}$, we have that $u^0 \in B(\bar{u},\, \tilde{\varepsilon})$. Suppose that $u^j \in \mathbb{R}^n$, $j = 1,\, \ldots,\, k$, are such that

$$u^j \in U^\sigma(\pi^{j-1},\, u^{j-1}) \cap B(\bar{u},\, \tilde{\varepsilon}) \quad \forall\, j = 1,\, \ldots,\, k.$$

Then, $\operatorname{dist}(u^j,\, \bar{U}) \leq \|u^j - \bar{u}\| \leq \tilde{\varepsilon}$ for all $j = 0,\, 1,\, \ldots,\, k$, and according to (4.3) and (4.7), there exists $u^{k+1} \in U^\sigma(\pi^k,\, u^k)$, and for any such u^{k+1} it holds that

$$\operatorname{dist}(u^{j+1},\, \bar{U}) \leq \ell\omega(\pi^j,\, u^j,\, u^{j+1}) \operatorname{dist}(u^j,\, \bar{U}) \leq q \operatorname{dist}(u^j,\, \bar{U}) \quad \forall\, j = 0,\, 1,\, \ldots,\, k.$$

Therefore, taking into account the inequality in (4.1), we obtain

$$\begin{aligned} \|u^{k+1} - u^0\| &\leq \sum_{j=0}^{k} \|u^{j+1} - u^j\| \\ &\leq \sum_{j=0}^{k} \sigma \operatorname{dist}(u^j,\, \bar{U}) \\ &\leq \sigma \sum_{j=0}^{k} q^j \operatorname{dist}(u^0,\, \bar{U}) \\ &= \sigma \frac{1 - q^{k+1}}{1 - q} \operatorname{dist}(u^0,\, \bar{U}) \\ &\leq \frac{\sigma}{1 - q} \operatorname{dist}(u^0,\, \bar{U}) \\ &\leq \frac{\sigma}{1 - q} \|u^0 - \bar{u}\| \\ &\leq \frac{\sigma \varepsilon_0}{1 - q}. \end{aligned} \tag{4.9}$$

Hence,

$$\|u^{k+1} - \bar{u}\| \leq \|u^{k+1} - u^0\| + \|u^0 - \bar{u}\| \leq \left(\frac{\sigma}{1 - q} + 1 \right) \varepsilon_0 = \tilde{\varepsilon},$$

where the last equality is by (4.8). Therefore, $u^{k+1} \in B(\bar{u},\, \tilde{\varepsilon})$, which completes the induction argument.

Furthermore, according to (4.3) and (4.7), and the assertion just established, the relations

$$\operatorname{dist}(u^{k+1},\, \bar{U}) \leq \ell\omega(\pi^k,\, u^k,\, u^{k+1}) \operatorname{dist}(u^k,\, \bar{U}) \leq q \operatorname{dist}(u^k,\, \bar{U}) \quad \forall\, k \tag{4.10}$$

are valid for any $u^0 \in B(\bar{u},\, \varepsilon_0)$, any sequence $\{\pi^k\} \subset \Pi$, and any sequence $\{u^k\} \subset \mathbb{R}^\nu$ satisfying (4.2). In particular, by (4.3), the sequence $\{\operatorname{dist}(u^k,\, \bar{U})\}$ converges to zero, and the estimate (4.6) holds.

Similarly to (4.9), one can establish the estimate

$$\|u^{k+j} - u^k\| \leq \frac{\sigma}{1-q} \operatorname{dist}(u^k, \bar{U}) \quad \forall\, k, \forall\, j, \tag{4.11}$$

where the right-hand side tends to zero as $k \to \infty$. Therefore, $\{u^k\}$ is a Cauchy sequence. Hence, it converges to some $u^* \in \mathbb{R}^\nu$. Moreover, employing convergence of $\{\operatorname{dist}(u^k, \bar{U})\}$ to zero, the inclusion $\{u^k\} \subset B(\bar{u}, \tilde{\varepsilon})$, and closedness of $\bar{U} \cap B(\bar{u}, 2\tilde{\varepsilon})$, we conclude that $u^* \in \bar{U}$.

Regarding the convergence rate of $\{u^k\}$, by (4.10) and (4.11) we derive

$$\|u^{k+j} - u^{k+1}\| \leq \frac{\sigma}{1-q} \operatorname{dist}(u^{k+1}, \bar{U}) \leq \frac{\sigma \ell \omega(\pi^k, u^k, u^{k+1})}{1-q} \operatorname{dist}(u^k, \bar{U}) \quad \forall\, k, \forall\, j \geq 1.$$

Passing onto the limit as $j \to \infty$ we further obtain

$$\|u^* - u^{k+1}\| \leq \frac{\sigma \ell \omega(\pi^k, u^k, u^{k+1})}{1-q} \operatorname{dist}(u^k, \bar{U}) \leq \frac{\sigma q}{1-q} \operatorname{dist}(u^k, \bar{U}) \quad \forall\, k.$$

This estimate proves (4.5). ∎

One can see from the proof of Theorem 4.1 that assumption (ii) can be replaced by the weaker one: (4.3) can be replaced by

$$q := \ell \sup\{\omega(\pi, \tilde{u}, u) \mid \pi \in \Pi,\ u \in U^\sigma(\pi, \tilde{u}),\ \tilde{u} \in B(\bar{u}, \bar{\varepsilon})\} < 1,$$

and (4.4) can be assumed to hold only under the additional requirement $u \in U(\pi, \tilde{u})$.

Consider now a stationary point $\bar{x} \in \mathbb{R}^n$ of the mathematical programming problem (2.6), and some particular $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ which is no more supposed to be unique. It was established in [21, Corollary 2.1] that in this setting assumption (i) of Theorem 4.1 with $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ and with $\bar{U}$ replaced by its subset $\{\bar{x}\} \times \mathcal{M}(\bar{x})$ is equivalent to saying that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier. In its turn, as we have already mentioned, the latter property is implied by SOSC (3.9).

As in the previous section, suppose that the approximation parameter τ_k in the inexact augmented Lagrangian method is chosen according to (3.11), where τ will now be assumed to satisfy

$$\tau(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leq \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \operatorname{dist}((\tilde{x}, \tilde{\lambda}, \tilde{\mu}), \{\bar{x}\} \times \mathcal{M}(\bar{x})) \quad \forall\, (\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m,$$

where $\theta : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ is again a function satisfying $\theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to 0$ as $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Then for any $\sigma > 0$ assumption (ii) of Theorem 4.1 holds for the inexact augmented Lagrangian method with $\Pi = [\bar{c}, +\infty)$ and $\omega(c, \tilde{u}, u) = \max\{\sigma/c, \theta(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\}$ for any sufficiently large $\bar{c} > 0$. Note that (3.14) still gives a relevant practical rule for choosing τ with the needed properties.

Finally, assumption (iii) of Theorem 4.1 (with Π defined above) can be verified for any $\sigma > 0$ and any sufficiently large $\bar{c} > 0$, employing assertion (a) of Proposition 3.2, Remark 3.1, and the reasoning in [14, Corollary 3.2, Proposition 3.3], assuming SOSC (3.9). Similarly to the previous section, SOSC can be replaced with SMFCQ and the quadratic growth condition.

However, this set of assumptions would not be relevant in this section, since in order to ensure assumption (i) of Theorem 4.1, we would have to require the multiplier in question to be noncritical, which together with SMFCQ would imply semistability, and hence, isolatedness of the primal-dual solution.

Theorem 4.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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* (2.6)*, and let SOSC* (3.9) *hold for an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. Let $\tau : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ be a function satisfying $\tau(x, \lambda, \mu) = o(\mathrm{dist}((x, \lambda, \mu), \{\bar{x}\} \times \mathcal{M}(\bar{x})))$.*

Then for any $\sigma > 0$ there exists $\bar{c} > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and any sequence $\{c_k\} \subset [\bar{c}, +\infty)$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying for all $k = 0, 1, \ldots$ condition (2.10) *with τ_k defined according to* (3.11)*, and conditions* (2.11) *and*

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leq \sigma(\|x^k - \bar{x}\| + \mathrm{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x}))); \tag{4.12}$$

any such sequence 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 $\{\|x^k - \bar{x}\| + \mathrm{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x}))\}$ to zero are linear. Moreover, both these rates are superlinear provided $c_k \to +\infty$, and, if $\tau_k = O((\mathrm{dist}((x^k, \lambda^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x})))^2)$, they are quadratic provided $1/c_k = O(\|x^k - \bar{x}\| + \mathrm{dist}((\lambda^k, \mu^k), \mathcal{M}(\bar{x})))$. In addition, for any $\varepsilon > 0$ it holds that $\|(\lambda^* - \bar{\lambda}, \mu^* - \bar{\mu})\| < \varepsilon$ provided (x^0, λ^0, μ^0) is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

For the case of twice differentiable data, Theorem 4.2 was obtained in [14]. When deriving it from Theorem 4.1, we assume the norm in $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ to be defined as the sum of norms in $\mathbb{R}^n$ and $\mathbb{R}^l \times \mathbb{R}^m$. Note also that when applying Theorem 4.1, we must choose $\varepsilon_0 > 0$ such that every sequence generated by the iterative scheme in question does not leave a neighborhood of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ where the localization condition in (4.1) is equivalent to (4.12).

The localization condition (4.12) in Theorem 4.2 is stronger (and, in a sense, "less practical") than its counterpart (2.16) in Theorems 2.2 and 3.2. Observe, however, that according to [21, Corollary 2.1], noncriticality of $(\bar{\lambda}, \bar{\mu})$ is also equivalent to the error bound

$$\|x - \bar{x}\| + \mathrm{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) = O\left(\left\|\left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), \min\{\mu, -g(x)\}\right)\right\|\right) \tag{4.13}$$

for $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. The computable residual of the KKT system (2.7), appearing in the right-hand side of (4.13), can be used in order to control $\{c_k\}$, and in (4.12).

5 Appendix

Proposition 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 continuous near $\bar{x} \in \mathbb{R}^n$, and let $\bar{x}$ be a strict local solution of problem* (2.6)*.*

Then there exists $\bar{\delta} > 0$ such that for any $\delta \in (0, \bar{\delta}]$ and any $M > 0$ there exists $c(\delta, M) > 0$ such that for any $c \geq c(\delta, M)$ and any $(\tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ satisfying $\|(\tilde{\lambda}, \tilde{\mu})\| \leq M$, any global solution x of the problem (3.17) *satisfies $\|x - \bar{x}\| \leq \delta$.*

For equality-constrained problems, this proposition is [6, Theorem 2.2]. Without any doubt, this property is well-understood in the general case as well, but we are not aware of an adequate reference, and therefore, we give a full proof.

Proof. Let $\bar{\delta} > 0$ be such that $\bar{x}$ is the unique global solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0,\ x \in B(\bar{x},\, \bar{\delta}). \end{array} \tag{5.1}$$

For any sequence $\{c_k\} \subset \mathbb{R}_+$ such that $c_k \to +\infty$, and any bounded sequence $\{(\tilde{\lambda}^k,\, \tilde{\mu}^k)\} \subset \mathbb{R}^l \times \mathbb{R}^m$, consider any sequence $\{x^k\} \subset \mathbb{R}^n$ such that for each k the point x^k is a global solution of problem (3.17) with $c = c_k$ and $(\lambda,\, \mu) = (\tilde{\lambda}^k,\, \tilde{\mu}^k)$. Since $\{x^k\}$ is contained in the compact set $B(\bar{x},\, \bar{\delta})$, this sequence has a limit point $\hat{x} \in B(\bar{x},\, \bar{\delta})$, and we need to show that $\hat{x} = \bar{x}$; this will mean that $\{x^k\}$ converges to $\bar{x}$.

Without loss of generality, suppose that the entire $\{x^k\}$ converges to $\hat{x}$. For each k large enough we then have

$$\begin{aligned}
f(x^k) + \frac{1}{2c_k}(\|\tilde{\lambda}^k + c_k h(x^k)\|_2^2 & \\
+\|\max\{0,\, \tilde{\mu}^k + c_k g(x^k)\}\|_2^2) &= L_{c_k}(x^k,\, \tilde{\lambda}^k,\, \tilde{\mu}^k) \\
&\leq L_{c_k}(\bar{x},\, \tilde{\lambda}^k,\, \tilde{\mu}^k) \\
&= f(\bar{x}) + \frac{1}{2c_k}(\|\tilde{\lambda}^k + c_k h(\bar{x})\|_2^2 + \|\max\{0,\, \tilde{\mu}^k + c_k g(\bar{x})\}\|_2^2) \\
&= f(\bar{x}) + \frac{1}{2c_k}\left(\|\tilde{\lambda}^k\|_2^2 + \sum_{i \in A(\bar{x})} (\max\{0,\, \tilde{\mu}_i^k\})^2\right),
\end{aligned}$$

implying that

$$f(\hat{x}) + \limsup_{k \to \infty} \frac{1}{2c_k}(\|\tilde{\lambda}^k + c_k h(x^k)\|_2^2 + \|\max\{0,\, \tilde{\mu}^k + c_k g(x^k)\}\|_2^2) \leq f(\bar{x}). \tag{5.2}$$

From the latter we get $h(\hat{x}) = 0$ and $g(\hat{x}) \leq 0$, and hence, $\hat{x}$ is feasible in problem (5.1). Therefore,

$$f(\bar{x}) \leq f(\hat{x}), \tag{5.3}$$

and from (5.2) we than derive that

$$\limsup_{k \to \infty} \frac{1}{2c_k}(\|\tilde{\lambda}^k + c_k h(x^k)\|_2^2 + \|\max\{0,\, \tilde{\mu}^k + c_k g(x^k)\}\|_2^2) \leq 0,$$

and hence,

$$\lim_{k \to \infty} \frac{1}{2c_k}(\|\tilde{\lambda}^k + c_k h(x^k)\|_2^2 + \|\max\{0,\, \tilde{\mu}^k + c_k g(x^k)\}\|_2^2) = 0.$$

From (5.2) we now get

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

and combining the latter with (5.3) gives the equality $f(\hat{x}) = f(\bar{x})$, further implying that $\hat{x}$ is a global solution of problem (5.1), and hence, $\hat{x} = \bar{x}$. ∎

Proposition 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 near $\bar{x} \in \mathbb{R}^n$, with their derivatives being continuous at $\bar{x}$, and let $\bar{x}$ be a stationary point of problem (2.6) satisfying SMFCQ with the unique associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then for any $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that for all $c > 0$ and all $(\tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ satisfying $|\tilde{\mu}_i| \leq \delta(\varepsilon)c$ for each $i \in \{1, \ldots, m\} \setminus A(\bar{x})$, for any stationary point x of the problem

$$\begin{array}{ll} \text{minimize} & L_c(x, \tilde{\lambda}, \tilde{\mu}) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \tag{5.4}$$

such that $\|x - \bar{x}\| \leq \delta(\varepsilon)$, it holds that the vectors

$$\lambda = \tilde{\lambda} + ch(x), \quad \mu = \max\{0, \tilde{\mu} + cg(x)\} \tag{5.5}$$

satisfy the inequality $\|(\lambda - \bar{\lambda}, \mu - \bar{\mu})\| \leq \varepsilon$.

Proof. Assume that the sequences $\{c_k\} \subset \mathbb{R}_+ \setminus \{0\}$, $\{x^k\} \subset \mathbb{R}^n$, $\{(\tilde{\lambda}^k, \tilde{\mu}^k)\} \subset \mathbb{R}^l \times \mathbb{R}^m$ satisfy the following: for each k the point x^k is a stationary point of (5.4) with $c = c_k$ and $(\tilde{\lambda}, \tilde{\mu}) = (\tilde{\lambda}^k, \tilde{\mu}^k)$, and

$$\{x^k\} \to \bar{x}, \quad \left\{\frac{\tilde{\mu}_i^k}{c_k}\right\} \to 0 \quad \text{as } k \to \infty \quad \forall\, i \in \{1, \ldots, m\} \setminus A(\bar{x}). \tag{5.6}$$

Consider the sequence $\{(\lambda^k, \mu^k)\}$ defined by (5.5) with $c = c_k$, $x = x^k$, and $(\tilde{\lambda}, \tilde{\mu}) = (\tilde{\lambda}^k, \tilde{\mu}^k)$. Similarly to (2.13) we have that for all k

$$\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) = 0, \tag{5.7}$$

and hence, from convergence of $\{x^k\}$ to $\bar{x}$ we get that any accumulation point $(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ of $\{(\lambda^k, \mu^k)\}$ satisfies

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

From (5.5) it readily follows that

$$\mu^k \geq 0 \quad \forall\, k. \tag{5.8}$$

Furthermore, for any $i \in \{1, \ldots, m\} \setminus A(\bar{x})$, since $\bar{\mu}_i^k = 0$ and $g_i(\bar{x}) < 0$, from (5.5) and (5.6) we have

$$\mu_i^k = \max\{0, \tilde{\mu}_i^k + c_k g_i(x^k)\} = c_k \max\left\{0, \frac{\tilde{\mu}_i^k}{c_k} + g_i(x^k)\right\} = 0 \quad \forall\, k \text{ large enough}. \tag{5.9}$$

By (5.8) and (5.9) we conclude that $\hat{\mu} \geq 0$, and that $\hat{\mu}_i = 0$ for all $i \in \{1, \ldots, m\} \setminus A(\bar{x})$. Therefore, $(\hat{\lambda}, \hat{\mu})$ is a Lagrange multiplier associated with $\bar{x}$, and hence, due to SMFCQ, $(\hat{\lambda}, \hat{\mu}) = (\bar{\lambda}, \bar{\mu})$.

It remains to prove that $\{(\lambda^k, \mu^k)\}$ is bounded; then from the above it will follow that this sequence converges to $(\bar{\lambda}, \bar{\mu})$. Without loss of generality, suppose that $t_k := \|(\lambda^k, \mu^k)\| \to \infty$

as $k \to \infty$, and that $\{(\lambda^k, \mu^k)/t_k\}$ converges to some $(\eta, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$, $\|(\eta, \zeta)\| = 1$. From (5.7) we have that for all k large enough

$$\frac{1}{t_k} f'(x^k) + (h'(x^k))^{\mathrm{T}} \frac{\lambda^k}{t_k} + (g'(x^k))^{\mathrm{T}} \frac{\mu^k}{t_k} = 0,$$

and passing onto the limit as $k \to \infty$ gives

$$(h'(\bar{x}))^{\mathrm{T}} \eta + (g'(\bar{x}))^{\mathrm{T}} \zeta = 0.$$

Moreover, from (5.8) and (5.9) it readily follows that $\zeta \geq 0$, and $\zeta_i = 0$ for all $i \in \{1, \ldots, m\} \setminus A(\bar{x})$. However, the existence of such (η, ζ) contradicts MFCQ. ■

Lemma 5.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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 (2.6), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier.*

Then for any $\xi \in C(\bar{x})$ and any sequences $\{t_k\} \to 0+$, $\{\xi^k\} \subset \mathbb{R}^n$ convergent to ξ, and $\{(\lambda^k, \mu^k)\} \subset \mathcal{M}(\bar{x})$ convergent to $(\bar{\lambda}, \bar{\mu})$, it holds that

$$\lim_{k\to\infty} \frac{L(\bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})}{t_k^2} = \lim_{k\to\infty} \frac{L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}, \lambda^k, \mu^k)}{t_k^2}$$

in the case of existence of these two limits which can exist only simultaneously.

Proof. For any k, it holds that

$$\begin{aligned} &(L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}, \lambda^k, \mu^k)) \\ &\quad -(L(\bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})) = L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x} + t_k\xi, \lambda^k, \mu^k) \\ &\qquad\qquad + L(\bar{x} + t_k\xi, \lambda^k, \mu^k) - L(\bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}), \end{aligned} \tag{5.10}$$

since $L(\bar{x}, \lambda^k, \mu^k) = L(\bar{x}, \bar{\lambda}, \bar{\mu}) = f(\bar{x})$, because both $(\bar{\lambda}, \bar{\mu})$ and (λ^k, μ^k) belong to $\mathcal{M}(\bar{x})$.

By the mean-value theorem, there exists $\theta_k \in [0, 1]$ such that

$$\begin{aligned} L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x} + t_k\xi, \lambda^k, \mu^k) &= L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x} + t_k\xi, \lambda^k, \mu^k) \\ &\quad - \left\langle \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k), t_k(\xi^k - \xi) \right\rangle \\ &= \left\langle \frac{\partial L}{\partial x}(\bar{x} + t_k(\theta_k\xi^k + (1-\theta_k)\xi), \lambda^k, \mu^k) \right. \\ &\quad \left. - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k), t_k(\xi^k - \xi) \right\rangle \\ &= O(t_k^2\|\xi^k - \xi\|) \\ &= o(t_k^2), \end{aligned} \tag{5.11}$$

where the next-to-the-last estimate employs local Lipschitz continuity of the derivatives of f, h and g at $\bar{x}$.

Furthermore, since $\xi \in C(\bar{x})$, by (3.10), and by the mean-value theorem, we derive

$$\begin{aligned} L(\bar{x}+t_k\xi,\, \lambda^k,\, \mu^k) - L(\bar{x}+t_k\xi,\, \bar{\lambda},\, \bar{\mu}) &= \langle \lambda^k-\bar{\lambda},\, h(\bar{x}+t_k\xi)\rangle + \langle \mu^k-\bar{\mu},\, g(\bar{x}+t_k\xi)\rangle \\ &= \langle \lambda^k-\bar{\lambda},\, h(\bar{x}+t_k\xi)-h(\bar{x})-t_kh'(\bar{x})\xi\rangle \\ &\quad +\langle \mu^k-\bar{\mu},\, g(\bar{x}+t_k\xi)-g(\bar{x})-t_kg'(\bar{x})\xi\rangle \\ &= \langle \lambda^k-\bar{\lambda},\, t_k(h'(\bar{x}+\theta_kt_k\xi)-h'(\bar{x}))\xi\rangle \\ &\quad +\langle \mu^k-\bar{\mu},\, t_k(g'(\bar{x}+\theta_kt_k\xi)-g'(\bar{x}))\xi\rangle \\ &= O(t_k^2\|\lambda^k-\bar{\lambda}\|)+O(t_k^2\|\mu^k-\bar{\mu}\|) \\ &= o(t_k^2). \end{aligned} \tag{5.12}$$

Employing (5.11) and (5.12), we obtain that the expression in the left-hand side of (5.10) is $o(t_k^2)$, implying the needed conclusion. ■

The next result can be regarded as a far-reaching generalization of the Finsler–Debreu lemma [15, 10].

Lemma 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^l$ 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 (2.6), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ be an associated Lagrange multiplier, satisfying (3.15).*

Then there exist $\bar{c} > 0$, $\bar{\delta} > 0$, $\bar{\gamma} > 0$ and $\bar{\varepsilon} > 0$ such that for all $c \geq \bar{c}$

$$\begin{aligned} & L(x,\, \lambda,\, \mu) - L(\bar{x},\, \lambda,\, \mu) + \frac{c}{2}\|h(x)\|_2^2 \\ & +\frac{c}{2}\sum_{i\in A_+(\bar{x},\,\bar{\mu})}(g_i(x))^2 + \frac{c}{2}\sum_{i\in A_0(\bar{x},\,\bar{\mu})}(\max\{0,\, g_i(x)\})^2 \;\geq\; \bar{\gamma}\|x-\bar{x}\|^2 \end{aligned}$$

for all $x \in B(\bar{x}, \bar{\delta})$ and all $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ such that $\|(\lambda-\bar{\lambda},\, \mu-\bar{\mu})\| \leq \bar{\varepsilon}$.

Proof. Arguing by contradiction, suppose that there exist a sequences $\{c_k\} \to +\infty$, $\{x^k\} \subset \mathbb{R}^n \setminus \{\bar{x}\}$ convergent to $\bar{x}$, and $\{(\lambda^k, \mu^k)\} \subset \mathcal{M}(\bar{x})$ convergent to $(\bar{\lambda}, \bar{\mu})$, such that

$$\begin{aligned} & L(x^k,\, \lambda^k,\, \mu^k) - L(\bar{x},\, \lambda^k,\, \mu^k) + \frac{c_k}{2}\|h(x^k)\|_2^2 \\ & +\frac{c_k}{2}\sum_{i\in A_+(\bar{x},\,\bar{\mu})}(g_i(x^k))^2 + \frac{c_k}{2}\sum_{i\in A_0(\bar{x},\,\bar{\mu})}\Big(\max\{0,\, g_i(x^k)\}\Big)^2 \;\leq\; o(\|x^k-\bar{x}\|^2). \end{aligned} \tag{5.13}$$

For each k, set $t_k = \|x^k-\bar{x}\|$ and $\xi^k = (x^k-\bar{x})/t_k$. Without loss of generality we may assume that $\{\xi^k\}$ converges to some $\xi \in \mathbb{R}^n \setminus \{0\}$. Dividing (5.13) by $c_kt_k^2$ and passing onto the limit as $k \to \infty$, we then obtain that

$$\|h'(\bar{x})\xi\|^2 + \|g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})\xi\|^2 + \|\max\{0,\, g'_{A_0(\bar{x},\,\bar{\mu})}(\bar{x})\xi\}\|^2 = 0,$$

and hence, $\xi \in C(\bar{x}) \setminus \{0\}$ (recall (3.10)).

On the other hand, since the last three terms in the left-hand side of (5.13) are nonnegative, it follows that that

$$L(\bar{x} + t_k\xi^k, \lambda^k, \mu^k) - L(\bar{x}, \lambda^k, \mu^k) \leq o(t_k^2).$$

Dividing this relation by t_k^2, and employing Lemma 5.1, we arrive at a contradiction with (3.15). ■

Acknowledgments

We thank Mikhail Solodov for his useful comments on the draft of this paper.

References

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

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

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

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

[5] A. V. Arutyunov and A. F. Izmailov. Sensitivity analysis for cone-constrained optimization problems under the relaxed constraint qualifications. *Math. Oper. Res.*, 30:333–353, 2005.

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

[7] J.F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.*, 29:161–186, 1994.

[8] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York, USA, 2000.

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

[10] G. Debreu. Definite and semidefinite quadratic forms. *Econometrica*, 20:295–300, 1952.

[11] A.L. Dontchev and R.T. Rockafellar. *Implicit Functions and Solution Mappings.* Springer, New York, USA, 2009.

[12] A.L. Dontchev and R.T. Rockafellar. Newton's method for generalized equations: a sequential implicit function theorem. *Math. Program.*, 123:139–159, 2010.

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

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

[15] P. Finsler. Über das vorkommen definiter und semidefiniter formen und scharen quadratischer formen. *Commentarii Matematici Helvetica*, 94:188–192, 1937.

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

[17] M.P. Friedlander and M.A. Saunders. A globally convergent linearly constrained Lagrangian method for nonlinear optimization. *SIAM J. Optim.*, 15:863–897, 2005.

[18] M.R. Hestenes. Multiplier and gradient methods. *J. Opt. Theory and Appl.*, 4:303–320, 1969.

[19] J.-B. Hiriart-Urruty, J.J. Strodiot, and V.H. Nguyen. Generalized Hessian matrix and second-order optimality conditions for problems with $\mathrm{C}^{1,1}$-data. *Appl. Math. Optim.*, 11:43–56, 1984.

[20] A.F. Izmailov. Strongly regular nonsmooth generalized equations. IMPA preprint serie D 096/2012, 2012.

[21] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. IMPA preprint serie A 694/2011, 2011.

[22] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization. *Set Valued and Variational Anal.*, 2012.

[23] A.F. Izmailov and M.V. Solodov. Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. *Comput. Optim. Appl.*, 46:347–368, 2010.

[24] N.H Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

[25] N.H. Josephy. Quasi-Newton methods for generalized equations. Technical Summary Report 1966, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

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

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

[28] B.A. Murtagh and M.A. Saunders. A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. *Math. Program. Study*, 16:84–117, 1982.

[29] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, 1983.

[30] M.J.D. Powell. *A method for nonlinear constraints in minimization problems*, pages 283–298. Academic Press, London and New York, 1969.

[31] S.M. Robinson. A quadratically convergent algorithm for general nonlinear programming problems. *Math. Program.*, 3:145–156, 1972.

[32] S.M. Robinson. Stability theory for systems of inequalities, Part II. Differentiable nonlinear systems. *SIAM J. Numer. Anal.*, 13:497–513, 1976.

[33] S.M. Robinson. Strongly regular generalized equations. *Math. Oper. Res.*, 5:43–62, 1980.

[34] S.M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis*, 2:291–305, 1994.

[35] S.M. Robinson. A point-of-attraction result for Newton's method with point-based approximations. *Optim.*, 60(1-2):89–99, 2011.