Math. Program., Ser. A (2013) 142:591–604
DOI 10.1007/s10107-012-0586-z

SHORT COMMUNICATION

A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems

Alexey F. Izmailov · Alexey S. Kurennoy · Mikhail V. Solodov

Received: 2 May 2011 / Accepted: 11 August 2012 / Published online: 2 September 2012

Abstract We prove a new local upper Lipschitz stability result and the associated local error bound for solutions of parametric Karush–Kuhn–Tucker systems corresponding to variational problems with Lipschitzian base mappings and constraints possessing Lipschitzian derivatives, and without any constraint qualifications. This property is equivalent to the appropriately extended to this nonsmooth setting notion of noncriticality of the Lagrange multiplier associated to the primal solution, which is weaker than second-order sufficiency. All this extends several results previously known only for optimization problems with twice differentiable data, or assuming some constraint qualifications. In addition, our results are obtained in the more general variational setting.

Keywords KKT system · Error bound · Second-order sufficiency · Critical multipliers

---

Research of the first two authors is supported by the Russian Foundation for Basic Research Grant 10-01-00251. The third author is supported in part by CNPq Grant 302637/2011-7, by PRONEX-Optimization, and by FAPERJ.

A. F. Izmailov, A. S. Kurennoy
Uchebniy Korpus 2, OR Department, VMK Faculty,
Moscow State University, MSU,
Leninskiye Gory, 119991 Moscow, Russia
e-mail: izmaf@ccas.ru

A. S. Kurennoy
e-mail: alex-kurennoy@yandex.ru

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

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

1 Introduction

Given the base mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, consider the variational problem

$$x \in D, \quad \langle \Phi(x), \xi \rangle \geq 0 \quad \forall \xi \in T_D(x), \tag{1}$$

where

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

and $T_D(x)$ is the usual tangent (contingent) cone to D at $x \in D$. This problem setting is fairly general. In particular, it contains optimality conditions: for a given smooth function $f : \mathbb{R}^n \to \mathbb{R}$, any local solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \; g(x) \leq 0, \end{array} \tag{2}$$

necessarily satisfies (1) with the base mapping defined by

$$\Phi(x) = f'(x), \quad x \in \mathbb{R}^n. \tag{3}$$

Assuming that the constraints mappings h and g are smooth, for any local solution $\bar{x}$ of problem (1) under the appropriate constraint qualifications there exists a multiplier $(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ satisfying the Karush–Kuhn–Tucker (KKT) system

$$\begin{gathered} \Phi(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \\ h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, \, g(x) \rangle = 0 \end{gathered} \tag{4}$$

for $x = \bar{x}$. In particular, if (3) holds then (4) is the KKT system of optimization problem (2). Let $\mathcal{M}(\bar{x})$ stand for the set of such multipliers associated with $\bar{x}$.

Define the mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Psi(x, \, \lambda, \mu) = \Phi(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu. \tag{5}$$

Let

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

be the sets of indices of active and inactive constraints at $\bar{x}$ and, for $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, let

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

be the sets of indices of strongly active and weakly active constraints, respectively. If Φ is differentiable at $\bar{x}$, and h and g are twice differentiable at $\bar{x}$, then from the results and discussions in [7–9,17] the following three properties are known to be equivalent.

Property 1 (*Upper Lipschitz stability of the solutions of KKT system under canonical perturbations*) There exist a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and $\ell > 0$ such that for any $\sigma = (a, b, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(0, 0, 0)$, any solution $(x(\sigma), \lambda(\sigma), \mu(\sigma)) \in \mathcal{U}$ of the perturbed KKT system

$$\Psi(x, \lambda, \mu) = a, \quad h(x) = b, \quad \mu \geq 0, \quad g(x) \leq c, \quad \langle \mu, g(x) - c \rangle = 0 \tag{6}$$

satisfies the estimate

$$\|x(\sigma) - \bar{x}\| + \operatorname{dist}((\lambda(\sigma), \mu(\sigma)), \mathcal{M}(\bar{x})) \leq \ell \|\sigma\|. \tag{7}$$

Property 2 (*Error bound for KKT system*) There exist a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and $\ell > 0$ such that for all $(x, \lambda, \mu) \in \mathcal{U}$ it holds that

$$\|x - \bar{x}\| + \operatorname{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) \leq \ell \left\| \begin{pmatrix} \Psi(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\|. \tag{8}$$

Property 3 (*The multiplier $(\bar{\lambda}, \bar{\mu})$ is noncritical*) There is no triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, with $\xi \neq 0$, that satisfies the system

$$\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \tag{9}$$

$$\zeta_{A_0} \geq 0, \quad g'_{A_0}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}), \xi \rangle = 0, \; i \in A_0, \quad \zeta_N = 0. \tag{10}$$

The equivalence between Properties 1 and 2 is valid in the more general context of mixed complementarity problems that contain KKT systems as a special case and, as will be discussed in Sect. 2, this equivalence actually does not require differentiability of Φ and twice differentiability of h and g. Observe also that Property 1 (dealing with canonical perturbations) implies the corresponding upper Lipschitz stability property for arbitrary parametric perturbations satisfying the Lipschitz-continuity assumption specified in item (b) of Theorem 1 below.

The notion of critical and noncritical multipliers had been discussed in [15,16], and the equivalence of noncriticality to Properties 1 and 2 in [17]. Note also that, as can be easily seen, when $\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a symmetric matrix [e.g., in optimization setting (3)], (9)–(10) is the KKT system for the quadratic programming problem

$$\begin{aligned} &\text{minimize}_\xi \quad \frac{1}{2}\left\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \\ &\text{subject to} \quad \xi \in C(\bar{x}), \end{aligned} \tag{11}$$

where

$$\begin{aligned} C(\bar{x}) &= \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi \le 0,\ \langle \Phi(\bar{x}),\ \xi \rangle \le 0\} \\ &= \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0,\ g'_{A_0}(\bar{x})\xi \le 0\} \end{aligned}$$

is the critical cone of the KKT system (4) at $\bar{x}$. Therefore, in this case $(\bar{\lambda},\ \bar{\mu})$ being noncritical is equivalent to saying that $\xi = 0$ is the unique stationary point of problem (11). Using the terminology of [4], the multiplier $(\bar{\lambda},\ \bar{\mu})$ being noncritical also means that the cone-matrix pair $(C(\bar{x}),\ \frac{\partial \Psi}{\partial x}(\bar{x},\ \bar{\lambda},\ \bar{\mu}))$ has the so-called R_0 property (see the discussion following (3.3.18) in [4]). Finally, note that multiplying the first equality in (9) by ξ and using the other two equalities in (9) and the relations in (10), it can be seen that a sufficient condition for $(\bar{\lambda},\ \bar{\mu})$ to be noncritical is the following second-order sufficiency condition (SOSC):

$$\left\langle \frac{\partial \Psi}{\partial x}(\bar{x},\ \bar{\lambda},\ \bar{\mu})\xi,\ \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x})\backslash\{0\}. \tag{12}$$

In particular, SOSC (12) is sufficient for the equivalent Properties 1–3, but is stronger in general.

Most error bounds for KKT systems would require at least some regularity-type assumptions about the constraints; see [13] for a summary. The exception is precisely the bound (8) first established under SOSC (12), which follows from [7,8]; and which is actually equivalent to the weaker assumption of noncriticality of the multiplier, as discussed above. Speaking about the predecessor of these results, we first mention [27] where under the additional Mangasarian–Fromovitz constraint qualification (MFCQ) it was proved that in optimization setting (3), SOSC (12) for all $(\bar{\lambda},\ \bar{\mu}) \in \mathcal{M}(\bar{x})$ implies the stronger version of Property 1, with $\mathcal{U}$ being a neighborhood of the entire set $\{\bar{x}\} \times \mathcal{M}(\bar{x})$. In [2,22], SOSC (12) was replaced by a condition close to noncriticality, but still for all $(\bar{\lambda},\ \bar{\mu}) \in \mathcal{M}(\bar{x})$, and still under MFCQ. Note that the works cited above deal with upper Lipschitzian stability of the primal solution only but, as will be demonstrated in the proof of Theorem 1 below, the needed property of dual solutions can then be easily derived by an application of Hoffman's error bound.

Furthermore, in [3,18,19] these results were extended to the case when the functions in (2) may not be twice differentiable, but only possess Lipschitz-continuous first derivatives. This problem setting with restricted smoothness requirements has multiple applications: e.g., in stochastic programming and optimal control (the so-called extended linear-quadratic problems [26,28,29]), and in semi-infinite programming and in primal decomposition procedures (see [21,25] and references therein). Once but not twice differentiable functions arise also when reformulating complementarity constraints as in [12] or in the lifting approach [10,11,30]. Other possible sources are subproblems in penalty or augmented Lagrangian methods with lower-level constraints treated directly and upper-level inequality constraints treated via quadratic penalization or via augmented Lagrangian, which gives rise to certain terms that are not twice differentiable in general; see, e.g., [1]. Theoretical difficulties concerned with the lack of smoothness have already been highlighted in [31] where, among other things, an example of an unconstrained optimization problem was provided showing that, unlike in the twice differentiable case, a point satisfying the quadratic growth condition may not be an isolated stationary point, and hence, may not be upper Lipschitz stable or satisfy the error bound (8).

We also mention the related work concerned with characterization of Lipschitz stability properties for generalized equations with possible nonsmooth base mappings by means of generalized differentiation [23]. However, the much stronger pseudo-Lipschitzian property of the solution mapping is investigated in that work.

Apart from theoretical significance, all this has implications for convergence of numerical algorithms. For example, SOSC (12) was the only assumption needed to prove local convergence of the stabilized sequential quadratic programming method in [5] and of the augmented Lagrangian algorithm in [6], with the error bound (8) playing a key role. When there are equality constraints only, the error bound itself (equivalently, noncriticality of the multiplier) is enough in the case of stabilized sequential quadratic programming [17]. The error bound (8) was also the key in active set identification [3] and constructing local regularization methods for problems with degenerate constraints in [14,32]. For other roles and applications of error bounds, see, e.g., [24].

The goal of this note is to prove that under weaker smoothness assumptions Properties 1 and 2 are equivalent to the appropriately defined notion of noncriticality, without any constraint qualifications. Not employing constraint qualifications is crucial for the applications of error bounds discussed above. Also, avoiding MFCQ is what makes the results presented below meaningful for problems with pure equality constraints and nonunique multipliers associated to a solution. Our smoothness hypotheses reduces to local Lipschitz-continuity of the mapping Ψ, and in particular, the results presented below are new and fully relevant in the case of a piecewise smooth Ψ.

Some words about our notation and terminology are in order. According to [20, (6.6)], for a mapping $F : \mathbb{R}^p \to \mathbb{R}^r$ which is locally Lipschitz-continuous at $u \in \mathbb{R}^p$ (that is, Lipschitz-continuous in some neighborhood of u), the contingent derivative of F at u is the multifunction $CF(u)$ from $\mathbb{R}^p$ to the subsets of $\mathbb{R}^r$, given by

$$CF(u)(v) = \{w \in \mathbb{R}^r \mid \exists \{t_k\} \subset \mathbb{R}_+,\ \{t_k\} \to 0+ : \\ \{(F(u + t_k v) - F(u))/t_k\} \to w\}. \tag{13}$$

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

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

where $\mathcal{S}_F$ is the set of points at which F is differentiable (this set is dense in the contexts of our interest). Then the Clarke generalized Jacobian of F at u is given by

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

where conv S stands for the convex hull of the set S. Observe that according to [20, (6.5), (6.6), (6.16)],

$$\forall\, w \in CF(u)(v) \quad \exists\, J \in \partial F(u) \;\text{ such that }\; w = Jv. \tag{14}$$

Moreover, if in addition F has the property

$$\sup_{J\in\partial F(u+v)} \|F(u+v) - F(u) - Jv\| = o(\|v\|), \tag{15}$$

then for any $v \in \mathbb{R}^p$ and any sequence $\{t_k\}$ corresponding to a given $w \in CF(u)(v)$ due to (13), it holds that

$$F(u + t_k v) - F(u) = t_k J_k v + o(t_k) \tag{16}$$

for any choice of $J_k \in \partial_B F(u + t_k v)$. Since the multifunction corresponding to the B-differential of a locally Lipschitz-continuous mapping is locally bounded and upper semicontinuous, we may assume without loss of generality that $\{J_k\}$ converges to some $J \in \partial_B F(u)$. Then from (16) we derive the equality $w = Jv$. It is thus proved that assuming (15), property (14) can be strengthened as follows:

$$\forall\, w \in CF(u)(v) \quad \exists\, J \in \partial_B F(u) \;\text{ such that }\; w = Jv. \tag{17}$$

Property (15) is one of the ingredients of the widely used notion of semismoothness of F at u (see [4, Section 7.4]). Specifically, semismoothness combines local Lipschitz continuity of F at u and property (15) with directional differentiability of F at u in any direction. The latter is not needed in our development.

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

Finally, a mapping $F : \mathbb{R}^p \to \mathbb{R}^r$ is said to be locally Lipschitz-continuous with respect to $u \in \mathbb{R}^p$ if there exist a neighborhood U of u and $\ell > 0$ such that

$$\|F(v) - F(u)\| \le \ell \|v - u\| \quad \forall\, v \in U.$$

2 Lipschitz-continuous KKT systems

The following is an extension of the notion of a noncritical multiplier to the case when the mapping Ψ defined in (5) is locally Lipschitz-continuous at the reference point but may not be differentiable.

Definition 1 Given a solution $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ of (4), we say that the multiplier $(\bar{\lambda},\, \bar{\mu})$ is *noncritical* if 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

$$d + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+}(\bar{x})\xi = 0, \tag{18}$$

$$\zeta_{A_0} \geq 0, \quad g'_{A_0}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}),\, \xi\rangle = 0,\ i \in A_0, \quad \zeta_N = 0. \tag{19}$$

with some $d \in C_x\Psi(\bar{x},\, \bar{\lambda},\, \bar{\mu})(\xi)$. Otherwise $(\bar{\lambda},\, \bar{\mu})$ is *critical*.

Observe that system (18)–(19) corresponds to system (35) in [19].

Remark 1 Employing (14) it is immediate that the multiplier $(\bar{\lambda},\, \bar{\mu})$ is necessarily noncritical if 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 (18)–(19) with $d = H\xi$ and some $H \in \partial_x\Psi(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Now one can readily see that the second-order sufficiency condition

$$\forall\, H \in \partial_x\Psi(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x})\backslash\{0\} \tag{20}$$

implies that the multiplier $(\bar{\lambda},\, \bar{\mu})$ is necessarily noncritical. It was established in [21] that in the optimization setting (3), condition (20) is indeed sufficient for local optimality of $\bar{x}$ in the problem (2).

Moreover, if in addition Ψ has the property

$$\sup_{H \in \partial_x\Psi(\bar{x}+\xi,\, \bar{\lambda},\, \bar{\mu})} \|\Psi(\bar{x}+\xi,\, \bar{\lambda},\, \bar{\mu}) - \Psi(\bar{x},\, \bar{\lambda},\, \bar{\mu}) - H\xi\| = o(\|\xi\|)$$

(in particular, if $\Psi(\cdot,\, \bar{\lambda},\, \bar{\mu})$ is semismooth at $\bar{x}$), then employing (17) it is immediate that the multiplier $(\bar{\lambda},\, \bar{\mu})$ is necessarily noncritical if 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 (18)–(19) with $d = H\xi$ and some $H \in (\partial_B)_x\Psi(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

Consider now the following parametric version of the KKT system (4):

$$\begin{gathered} \Phi(\sigma,\, x) + \left(\frac{\partial h}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}}\lambda + \left(\frac{\partial g}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}}\mu = 0, \\ h(\sigma,\, x) = 0, \quad \mu \geq 0, \quad g(\sigma,\, x) \leq 0, \quad \langle \mu,\, g(\sigma,\, x)\rangle = 0, \end{gathered} \tag{21}$$

where $\sigma \in \mathbb{R}^s$ is a (perturbation) parameter, and $\Phi : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^l$, $g : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^m$ are the given mappings such that h and g are differentiable with respect to x.

The following sensitivity result generalizes [9, Theorem 2.3] with respect to its smoothness assumptions (see also the related result in [8]).

Theorem 1 *Let $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ be a solution of system (21) for $\sigma = \bar{\sigma}$, where $\Phi : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^m$ are such that h and g are differentiable with respect to x near $(\bar{\sigma},\, \bar{x})$. Let the following assumptions be satisfied:*

1. *$\Phi(\bar{\sigma},\, \cdot)$, $\frac{\partial h}{\partial x}(\bar{\sigma},\, \cdot)$ and $\frac{\partial g}{\partial x}(\bar{\sigma},\, \cdot)$ are locally Lipschitz-continuous at $\bar{x}$.*
2. *There exist a neighborhood $\mathcal{U}$ of $\bar{\sigma}$ and $\ell > 0$ such that for any fixed x close enough to $\bar{x}$ the mappings $\Phi(\cdot,\, x)$, $g(\cdot,\, x)$, $h(\cdot,\, x)$, $\frac{\partial h}{\partial x}(\cdot,\, x)$ and $\frac{\partial g}{\partial x}(\cdot,\, x)$ are Lipschitz-continuous with respect to $\bar{\sigma}$ on $\mathcal{U}$ with the Lipschitz constant ℓ.*

If $(\bar{\lambda},\ \bar{\mu})$ *is a noncritical multiplier associated with* $\bar{x}$ *for the KKT system* (21) *for* $\sigma = \bar{\sigma}$*, then for each* σ *close enough to* $\bar{\sigma}$ *and every solution* $(x(\sigma),\ \lambda(\sigma),\ \mu(\sigma))$ *of* (21) *close enough to* $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$*, it holds that*

$$\|x(\sigma) - \bar{x}\| + dist((\lambda(\sigma),\ \mu(\sigma)),\ \mathcal{M}(\bar{\sigma},\ \bar{x})) = O(\|\sigma - \bar{\sigma}\|),$$

where $\mathcal{M}(\bar{\sigma},\ \bar{x})$ *is the set of* $(\lambda,\ \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ *such that* $(\bar{x},\ \lambda,\ \mu)$ *is a solution of* (21) *for* $\sigma = \bar{\sigma}$.

Proof We first establish that

$$\|x(\sigma) - \bar{x}\| = O(\|\sigma - \bar{\sigma}\|). \tag{22}$$

Suppose that (22) does not hold, which means that there exist some sequences $\{\sigma^k\} \subset \mathbb{R}^s$ and $\{(x^k,\ \lambda^k,\ \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that $\{\sigma^k\} \to \bar{\sigma}$, $\{(x^k,\ \lambda^k,\ \mu^k)\} \to (\bar{x},\ \bar{\lambda},\ \bar{\mu})$, for each k the point $(x^k,\ \lambda^k,\ \mu^k)$ is a solution of (21) for $\sigma = \sigma^k$, and $\|x^k - \bar{x}\| > \gamma_k \|\sigma^k - \bar{\sigma}\|$ with some $\gamma_k > 0$ such that $\gamma_k \to \infty$. It then holds that

$$\|\sigma^k - \bar{\sigma}\| = o(\|x^k - \bar{x}\|). \tag{23}$$

Define the index sets

$$\begin{aligned}
&A = A(\bar{\sigma},\ \bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{\sigma},\ \bar{x}) = 0\},\\
&N = N(\bar{\sigma},\ \bar{x}) = \{1, \ldots, m\} \backslash A,\\
&A_+ = A_+(\bar{\sigma}, \bar{x}, \bar{\mu}) = \{i \in A \mid \bar{\mu}_i > 0\},\\
&A_0 = A_0(\bar{\sigma}, \bar{x}, \bar{\mu}) = A \backslash A_+.
\end{aligned}$$

Since there is only a finite number of ways to decompose the index set A_0 into two non-intersecting subsets, without loss of generality we can assume that for each k,

$$\mu_i^k > 0 \quad \forall\, i \in I_1, \quad \mu_i^k = 0 \quad \forall\, i \in I_2, \tag{24}$$

where I_1 and I_2 are some fixed index sets such that $I_1 \cup I_2 = A_0$, $I_1 \cap I_2 = \emptyset$. Furthermore, under the assumptions (a) and (b), g is continuous at $(\bar{\sigma},\ \bar{x})$. Therefore, by the complementarity condition in (21), and by convergence of $\{(x^k,\ \lambda^k,\ \mu^k)\}$ to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, without loss of generality we can assume that for all k

$$\mu_i^k > 0 \quad \forall\, i \in A_+, \quad \mu_i^k = 0 \quad \forall\, i \in N. \tag{25}$$

Define the mapping $\Psi : \mathbb{R}^s \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Psi(\sigma,\ x,\ \lambda,\ \mu) = \Phi(\sigma,\ x) + \left(\frac{\partial h}{\partial x}(\sigma,\ x)\right)^{\mathrm{T}} \lambda + \left(\frac{\partial g}{\partial x}(\sigma,\ x)\right)^{\mathrm{T}} \mu.$$

Passing onto a subsequence if necessary, we can assume that $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some $\xi \in \mathbb{R}^n$, $\|\xi\| = 1$. Then, setting $t_k = \|x^k - \bar{x}\|$ and passing to a further subsequence, if necessary, by (13) and by local Lipschitz continuity of $\Psi(\bar{\sigma}, \cdot, \bar{\lambda}, \bar{\mu})$ at $\bar{x}$ (implied by local Lipschitz continuity of $\Phi(\bar{\sigma}, \cdot)$, $\frac{\partial h}{\partial x}(\bar{\sigma}, \cdot)$ and $\frac{\partial g}{\partial x}(\bar{\sigma}, \cdot)$ at $\bar{x}$) we have that there exists $d \in C_x\Psi(\bar{\sigma}, \bar{x}, \bar{\lambda}, \bar{\mu})(\xi)$ such that

$$\begin{aligned}
\Psi(\bar{\sigma}, x^k, \bar{\lambda}, \bar{\mu}) &= \Psi(\bar{\sigma}, \bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}) - \Psi(\bar{\sigma}, \bar{x}, \bar{\lambda}, \bar{\mu}) \\
&\quad + \Psi(\bar{\sigma}, x^k, \bar{\lambda}, \bar{\mu}) - \Psi(\bar{\sigma}, \bar{x} + t_k\xi, \bar{\lambda}, \bar{\mu}) \\
&= t_k d + o(t_k) + O(\|x^k - \bar{x} - t_k\xi\|) \\
&= t_k d + o(t_k) + O\left(t_k \left\| \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} - \xi \right\|\right) \\
&= \|x^k - \bar{x}\| d + o(\|x^k - \bar{x}\|).
\end{aligned} \tag{26}$$

The first line in (21) can be written in the form

$$\Psi(\sigma, x, \lambda, \mu) = 0.$$

Therefore, using (24)–(26) and assumptions (a) and (b), as well as (23), we obtain that

$$\begin{aligned}
0 &= \Psi(\sigma^k, x^k, \lambda^k, \mu^k) \\
&= \Psi(\sigma^k, x^k, \bar{\lambda}, \bar{\mu}) + \left(\frac{\partial h}{\partial x}(\sigma^k, x^k)\right)^{\mathrm{T}} (\lambda^k - \bar{\lambda}) + \left(\frac{\partial g}{\partial x}(\sigma^k, x^k)\right)^{\mathrm{T}} (\mu^k - \bar{\mu}) \\
&= \Psi(\bar{\sigma}, x^k, \bar{\lambda}, \bar{\mu}) + \left(\frac{\partial h}{\partial x}(\bar{\sigma}, x^k)\right)^{\mathrm{T}} (\lambda^k - \bar{\lambda}) + \left(\frac{\partial g}{\partial x}(\bar{\sigma}, x^k)\right)^{\mathrm{T}} (\mu^k - \bar{\mu}) \\
&\quad + O(\|\sigma^k - \bar{\sigma}\|) \\
&= \|x^k - \bar{x}\| d + \left(\frac{\partial h}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} (\lambda^k - \bar{\lambda}) \\
&\quad + \left(\frac{\partial g_{A_+ \cup I_1}}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} (\mu^k_{A_+ \cup I_1} - \bar{\mu}_{A_+ \cup I_1}) + o(\|x^k - \bar{x}\|) \\
&= \|x^k - \bar{x}\| d + \left(\frac{\partial h}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} (\lambda^k - \bar{\lambda}) \\
&\quad + \left(\frac{\partial g_{A_+}}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} (\mu^k_{A_+} - \bar{\mu}_{A_+}) + \left(\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} \mu^k_{I_1} + o(\|x^k - \bar{x}\|).
\end{aligned}$$

Hence,

$$\begin{aligned}
&- \operatorname{im}\left(\frac{\partial h}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} - \operatorname{im}\left(\frac{\partial g_{A_+}}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} - \left(\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma}, \bar{x})\right)^{\mathrm{T}} \mathbb{R}^{|I_1|}_+ \\
&\qquad \ni \|x^k - \bar{x}\| d + o(\|x^k - \bar{x}\|),
\end{aligned} \tag{27}$$

where the set in the left-hand side is a closed cone (as a sum of a linear subspace and a polyhedral cone).

Using the second line in (21), and assumption (b), as well as (23), we obtain that

$$\begin{aligned} 0 &= h(\sigma^k,\, x^k) \\ &= \frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + O(\|\sigma^k - \bar{\sigma}\|) + o(\|x^k - \bar{x}\|) \\ &= \frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \end{aligned} \tag{28}$$

Similarly, making use of (24), (25), we derive the estimates

$$0 = g_{A_+ \cup I_1}(\sigma^k,\, x^k) = \frac{\partial g_{A_+ \cup I_1}}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \tag{29}$$

$$0 \geq g_{I_2}(\sigma^k,\, x^k) = \frac{\partial g_{I_2}}{\partial x}(\bar{\sigma},\, \bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \tag{30}$$

Dividing (27)–(30) by $\|x^k - \bar{x}\|$ and taking the limit as $k \to \infty$, we obtain that

$$d \in -\mathrm{im} \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} - \mathrm{im} \left(\frac{\partial g_{A_+}}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} - \left(\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \mathbb{R}_+^{|I_1|}, \tag{31}$$

$$\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \quad \frac{\partial g_{A_+}}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \tag{32}$$

$$\frac{\partial g_{I_1}}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \quad \frac{\partial g_{I_2}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leq 0. \tag{33}$$

The inclusion (31) means that there exists $(\eta,\, \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$ satisfying

$$d + \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \eta + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \zeta = 0 \tag{34}$$

and such that

$$\zeta_{I_1} \geq 0, \quad \zeta_{I_2 \cup N} = 0.$$

Then, taking also into account (33), the triple $(\xi,\, \eta,\, \zeta)$ satisfies

$$\begin{gathered} \zeta_{A_0} \geq 0, \quad \frac{\partial g_{A_0}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leq 0, \quad \zeta_i \left\langle \frac{\partial g_i}{\partial x}(\bar{\sigma},\, \bar{x}),\, \xi \right\rangle = 0,\ i \in A_0, \\ \zeta_N = 0. \end{gathered} \tag{35}$$

As $\xi \neq 0$, relations (32), (34), (35) contradict the assumption that the multiplier $(\bar{\lambda},\, \bar{\mu})$ is noncritical. This proves (22).

Considering σ close enough to $\bar{\sigma}$, and a solution $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ of system (21) close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, we have that

$$\Psi(\sigma,\, x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) = 0, \quad \mu(\sigma) \geq 0, \quad \mu_N(\sigma) = 0,$$

where the last equality is by the continuity of g at $(\bar{\sigma}, \bar{x})$. Since $\mathcal{M}(\bar{\sigma}, \bar{x})$ is the solution set of the linear system

$$\Psi(\bar{\sigma}, \bar{x}, \lambda, \mu) = 0, \quad \mu_A \geq 0, \quad \mu_N = 0,$$

by Hoffman's error bound (e.g., [4, Lemma 3.2.3]) we obtain that

$$\begin{aligned} \operatorname{dist}((\lambda(\sigma), \mu(\sigma)), \mathcal{M}(\bar{\sigma}, \bar{x})) &= O\Bigg(\|\Psi(\bar{\sigma}, \bar{x}, \lambda(\sigma), \mu(\sigma))\| \\ &\quad + \sum_{i \in A} \min\{0, \mu_i(\sigma)\} + \|\mu_N(\sigma)\|\Bigg) \\ &= O(\|\Psi(\sigma, x(\sigma), \lambda(\sigma), \mu(\sigma)) \\ &\quad - \Psi(\bar{\sigma}, \bar{x}, \lambda(\sigma), \mu(\sigma))\|) \\ &= O(\|\sigma - \bar{\sigma}\|) + O(\|x(\sigma) - \bar{x}\|), \end{aligned}$$

where assumptions (a) and (b) were also used. Together with (22), this gives the assertion of the theorem. □

In the optimization setting (3), *and* if MFCQ and noncriticality of all the multipliers are assumed, Theorem 1 essentially recovers the sufficiency part of Theorem 4 in [19] (see also [20, Theorem 8.24]). Regarding the necessity part, the following comments are in order.

Remark 2 Being applied to the case of canonical perturbations, Theorem 1 reveals that noncriticality of the multiplier in question is sufficient for Property 1 to hold. It turns out that it is necessary as well. In order to show this, take any triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying the system (18)–(19) with some $d \in C_x\Psi(\bar{x}, \bar{\lambda}, \bar{\mu})(\xi)$, and fix a sequence $\{t_k\}$ corresponding to this d due to (13). It then can be directly checked that for all k large enough the triple $(\bar{x} + t_k\xi, \bar{\mu} + t_k\eta, \bar{\lambda} + t_k\zeta)$ satisfies (6) with some $a = o(t_k)$, $b = o(t_k)$ and $c = o(t_k)$. Therefore, if $\xi \neq 0$, we would get a contradiction with (7).

Remark 3 The properties specified in Remark 1 are indeed strictly stronger than noncriticality introduced in Definition 1, and hence, strictly stronger than Property 1.

For example, let $n = m = 1$, $l = 0$, $\Phi(x) = \max\{0, x\}$, $g(x) = -x$. The point $(\bar{x}, \bar{\mu}) = (0, 0)$ is the only solution of system (4). It can be easily checked that the multiplier $\bar{\mu}$ is noncritical, and Property 1 is valid for the specified $(\bar{x}, \bar{\mu})$. Indeed, system (6) takes the form

$$\max\{0, x\} - \mu = a, \quad \mu \geq 0, \quad -x \leq c, \quad \mu(x + c) = 0.$$

This implies that either $\mu = 0$, in which case x satisfies at least one of the relations $x = a$ or $-c \leq x < 0$; or $\mu > 0$, in which case $x = -c$ and μ equals to at least one of $-a$ or $-a - c$. Therefore, $|x| + |\mu| \leq |a| + 2|c|$, giving (7) with $\mathcal{U} = \mathbb{R} \times \mathbb{R}$ and an appropriate $\ell > 0$ (depending on the choice of the norm in the right-hand side of (7)).

However, $(\partial_B)_x \Psi(\bar{x}, \bar{\mu}) = \{0, 1\}$ contains zero, and any $\xi \geq 0$ satisfies (18)–(19) with $d = 0$ and $\eta = 0$.

Note that the example above corresponds to the parametric optimization problem

$$\begin{aligned} &\text{minimize } \frac{1}{2}(\max\{0, x\})^2 \\ &\text{subject to } x \geq 0. \end{aligned}$$

Therefore, the conditions from Remark 1 are stronger than noncriticality and not necessary for Property 1 to hold even for optimization problems. This is different from the case when the problem data is twice differentiable.

Remark 4 As already mentioned in Sect. 1, Properties 1 an 2 are equivalent under the only assumptions that Φ and the derivatives of h and g are locally Lipschitz-continuous with respect to $\bar{x}$. The fact that Property 2 implies Property 1 is immediate. The converse implication was established in [7, Theorem 2]. The essence of the argument is the following. For any triple $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ one can construct $\tilde{\mu} \in \mathbb{R}^m$ and $\sigma = (a, b, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that

$$\begin{gathered} \Phi(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\tilde{\mu} = a, \\ h(x) = b, \quad \tilde{\mu} \geq 0, \quad g(x) \leq c, \quad \langle \tilde{\mu}, g(x) - c \rangle = 0 \end{gathered}$$

[cf. (6)], and

$$\|\tilde{\mu} - \mu\| = O(\|\min\{\mu, -g(x)\}\|), \quad \|\sigma\| = O\left(\left\| \begin{pmatrix} \Psi(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\|\right).$$

In the case when there are no inequality constraints, the construction is obvious, as one simply takes $a = \Phi(x) + (h'(x))^{\mathrm{T}}\lambda$ and $b = h(x)$. We refer the reader to [7, Theorem 2] for details in the case when inequality constraints are present.

Combining Remarks 2 and 4, we finally obtain the following result.

Corollary 1 *Let $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ be a solution of system* (4)*, where $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are such that h and g are differentiable near $\bar{x}$, and Φ, h' and g' are locally Lipschitz-continuous at $\bar{x}$.*

Then the following three properties are equivalent:

1. *Property 1 (Upper Lipschitz stability of the solutions of KKT system under canonical perturbations)*
2. *Property 2 (Error bound for KKT system).*
3. *The multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is noncritical* (*in the sense of Definition* 1).

We complete this note with another example illustrating the setting in question and the results obtained. Note that the derivatives of the data of the reformulated optimization problem in this example are locally Lipschitz-continuous but not piecewise smooth.

Example 1 Consider the mathematical program with model complementarity constraints

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(x_1^2 + x_2^2) \\ \text{subject to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1 x_2 = 0. \end{array}$$

Employing the squared Fischer–Burmeister complementarity function, we can reformulate the constraints of this problem as a single equality constraint with locally Lipschitz-continuous derivative:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(x_1^2 + x_2^2) \\ \text{subject to} & \frac{1}{2}\left(x_1 + x_2 - \sqrt{x_1^2 + x_2^2}\right)^2 = 0. \end{array}$$

The unique solution of this problem is $\bar{x} = 0$, and $\mathcal{M}(\bar{x}) = \mathbb{R}$. After some manipulations, one can see that the only critical multipliers are $-3 \pm 2\sqrt{2}$.

Take, e.g., the critical multiplier $\bar{\lambda} = -3 - 2\sqrt{2}$, and consider $x \in \mathbb{R}^2$ such that $x_1 = x_2 = t > 0$. Then by direct computation we obtain that for this problem data $\Psi(x, \bar{\lambda}) = 0$, $h(x) = (2 - \sqrt{2})^2 t^2/2$. Therefore, the error bound (8) does not hold with $\lambda = \bar{\lambda}$ for any $\ell > 0$ provided t is small enough.

On the other hand, and as established above, it can be seen that the error bound (8) holds around noncritical multipliers.

Acknowledgments The authors thank the anonymous referees and the Associate Editor for useful comments that led to an improved version of the original submission.

References

1. Andreani, R., Birgin, E., Martínez, J., Schuverdt, M.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2007)
2. Dontchev, A., Rockafellar, R.: Characterizations of Lipschitzian stability in nonlinear programming. In: Fiacco, A. (ed.) Mathematical Programming with Data Perturbations, pp. 65–82. Marcel Dekker, NY (1997)
3. Facchinei, F., Fischer, A., Kanzow, C.: On the accurate identification of active constraints. SIAM J. Optim. **9**, 14–32 (1999)
4. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
5. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**, 47–73 (2010)
6. Fernández, D., Solodov, M.: Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. **22**, 384–407 (2012)
7. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
8. Hager, W., Gowda, M.: Stability in the presence of degeneracy and error estimation. Math. Program. **85**, 181–192 (1999)
9. Izmailov, A.: Solution sensitivity for Karush–Kuhn–Tucker systems with nonunique Lagrange multipliers. Optimization **95**, 747–775 (2010)

10. Izmailov, A., Pogosyan, A., Solodov, M.: Semismooth SQP method for equality-constrained optimization problems with an application to the lifted reformulation of mathematical programs with complementarity constraints. Comput. Optim. Appl. **26**, 847–872 (2011)
11. Izmailov, A., Pogosyan, A., Solodov, M.: Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints. Comput. Optim. Appl. **51**, 199–221 (2012)
12. Izmailov, A., Solodov, M.: The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. **27**, 614–635 (2002)
13. Izmailov, A., Solodov, M.: Karush–Kuhn–Tucker systems: regularity conditions, error bounds, and a class of Newton-type methods. Math. Program. **95**, 631–650 (2003)
14. Izmailov, A., Solodov, M.: Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim. **16**, 210–228 (2004)
15. Izmailov, A., Solodov, M.: On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. **117**, 271–304 (2009)
16. Izmailov, A., Solodov, M.: On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. **126**, 231–257 (2011)
17. Izmailov, A., Solodov, M.: Stabilized SQP revisited. Math. Program. **133**, 93–120 (2012)
18. Klatte, D.: Nonlinear Optimization Under Data Perturbations, pp. 204–235. Springer, Berlin (1992)
19. Klatte, D.: Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs. Math. Program. **88**, 285–311 (2000)
20. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. Kluwer, Dordrecht (2002)
21. Klatte, D., Tammer, K.: On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems. Optimization **19**, 169–180 (1988)
22. Levy, A.: Errata in implicit multifunction theorems for the sensitivity analysis of variational conditions. Math. Program. **86**, 439–441 (1999)
23. Mordukhovich, B.: Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. Trans. Am. Math. Soc. **343**, 609–657 (1994)
24. Pang, J.: Error bounds in mathematical programming. Math. Program. **79**, 299–332 (1997)
25. Qi, L.: LC^1 functions and LC^1 optimization problems. Technical Report AMR 91/21, School of Mathematics, The University of New South Wales, Sydney (1991)
26. Qi, L.: Superlinearly convergent approximate Newton methods for LC^1 optimization problems. Math. Program. **64**, 277–294 (1994)
27. Robinson, S.: Generalized equations and their solution. Part II: applications to nonlinear programming. Math. Program. Study **19**, 200–221 (1982)
28. Rockafellar, R.: Computational schemes for solving large-scale problems in extended linear-quadratic programming. Math. Program. **48**, 447–474 (1990)
29. Rockafellar, R., Wets, J.: Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. SIAM J. Control Optim. **28**, 810–922 (1990)
30. Stein, O.: Lifting mathematical programs with complementarity constraints. Math. Program. **131**, 71–94 (2012)
31. Ward, D.: Characterizations of strict local minima and necessary conditions for weak sharp minima. Optim. Theory Appl. **80**, 551–571 (1994)
32. Wright, S.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**, 673–696 (2005)