Mathematical programs with equilibrium constraints: Theory and numerical methods

Jiří V. Outrata
Institute of Information Theory and Automation
Czech Academy of Science, Prague, Czech Republic

Abstract The lecture notes deal with optimization problems, where a generalized equation (modeling an equilibrium) arises among the constraints. The main attention is paid to necessary optimality conditions and methods to the numerical solution of such problems. The applications come from continuum mechanics.

Introduction

These lecture notes are devoted to a relatively new class of optimization problems which play an important role in mechanics. This class is now commonly entitled *mathematical programs with equilibrium problems* (MPECs) and our knowledge about it has grown remarkably fast especially in the last two decades. Nevertheless, this area still offers a considerable number of open questions motivated by both theoretical considerations as well as real-life problems. An MPEC can be written in an abstract form

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \\ & y \in S(x) \\ & (x,y) \in \kappa, \end{array} \tag{0.0.1}$$

where x is the *control* or *design* variable, y is the *state* variable, f is the objective, κ is a closed set of feasible control-state pairs and S is a closed-valued multifunction. The relation

$$y \in S(x) \tag{0.0.2}$$

says that y is a solution of an equilibrium problem, parameterized by x. Problems of this kind can be found already in von Stackelberg (1934), where the equilibrium is another (lower-level) optimization problem (in variable y). One speaks then about a *bilevel* optimization problem (Dempe (2002)) and we frequently meet such MPECs eg in shape optimization. Relation (0.0.2) can model, however, completely different equilibria governed by variational inequalities, fixed point problems etc., and it is exactly this relation (called also *equilibrium constraint*) which makes MPECs so specific.

The literature devoted to MPECs consists currently of three monographs (Luo et al. (1996),Outrata et al. (1998),Dempe (2002)) and hundreds of papers dealing with a broad spectrum of questions arising in this context. These notes follow approximately the setup of Outrata et al. (1998), but they employ another tools from nonsmooth analysis and handle a substantially broader spectrum of equilibria. Similarly as in Outrata et al. (1998), however, the considered equilibria are modeled by *generalized equations* (GEs)

$$0 \in F(x, y) + Q(y), \tag{0.0.3}$$

where F is a single-valued and Q is a set-valued map. All considered MPECs are finite-dimensional and so the obtained results are applicable in continuum mechanics only after a suitable discretization. In case of proposed numerical methods this is not a drawback. Concerning optimality conditions, the available infinite-dimensional results are substantially less sharp than their finite-dimensional counterparts and so we decided to keep all problems of these notes finite-dimensional. We omit completely the existence questions associated with (0.0.1), because the known results (based on various simple modifications of the Weierstrass Theorem) are not very deep and can easily be found in Luo et al. (1996). So our main attention is concentrated on 1st-order necessary optimality conditions and methods for the numerical solution of (0.0.1) with S given by various equilibria arising in mechanics.

The outline is as follows: In Section 1 we collect the fundamentals from nonsmooth and set-valued analysis needed throughout the whole sequel. Section 2 deals with the modeling of basic types of equilibria and some simple existence and uniqueness questions. With the constraint system in (0.0.1) we can associate multifunctions playing a crucial role in optimality conditions and stability analysis. Their Lipschitzian properties are investigated in Section 3. The results of Section 1 and 3 enable then to derive easily rather sharp necessary optimality conditions, which is done in Section 4. Section 5 and the first part of Section 6 are then focused on numerical methods. More precisely, Section 5 concerns the *implicit programming approach* (ImP) that, in connection with bundle methods of nonsmooth optimization, leads to an effective numerical procedure. In Section 6.1 we then describe briefly the *nonlinear programming* (NLP) approach which leads to several effective methods capable to cope with properly multi-valued maps S. Section 6.2 concentrates on MPECs, where the underlying equilibrium has an evolutionary nature; such problems arise eg in connection with delamination.

We employ the following notation which is rather standard in MPEC literature and nonsmooth analysis. Concretely, $\mathbb{R}^n_+, \mathbb{R}^n_-$ is the nonnegative and the nonpositive orthant of $\mathbb{R}^n$, respectively, and $\overline{\mathbb{R}}$ is the extended real line. For an $[m \times n]$ matrix A and index sets $I \subset \{1, 2, \ldots, m\}, J \subset \{1, 2, \ldots, n\}, A_{I,J}$ denotes the submatrix of A with rows and columns specified by I, J, respectively. A_I is the submatrix of A with rows specified by I. Similarly, for a vector $d \in \mathbb{R}^n, d_I$ is the subvector composed from the components $d^i, i \in I$. Furthermore, epi$f$ is the epigraph of a real-valued function f, for a multifunction Φ, domΦ and GphΦ denote its domain and its graph, and for a set Ω, intΩ, clΩ, convΩ and $\delta_\Omega(\cdot)$ mean its interior, closure, convex hull and indicatory function, respectively. $\mathbb{B}$ is the unit ball, E is the unit matrix and for a finite set $J, |J|$ denotes its cardinality. D^0 is the negative polar to a cone D and, for a continuously differentiable map $F[\mathbb{R}^n \to \mathbb{R}^m], \nabla F(z)$ denotes the Jacobian of F at $z \in \mathbb{R}^n$. If $m = 1, \nabla F(z)$ is the gradient (and thus a column vector). If F is only directionally differentiable at $z, F'(z; d)$ is its directional derivative at z in the direction d. Having two vectors u, v

of the same dimension, $\max\{u, v\}$ and $\min\{u, v\}$ denote their componentwise maximum and minimum. Finally, $o[\mathbb{R}_+ \to \mathbb{R}]$ is a function with the property

$$\lim_{t \downarrow 0} \frac{o(t)}{t} = 0$$

and "$\bullet$" denotes the Hadamard product.

The reference list by far does not contain all works, relevant with respect to this topic. An interested reader can find further useful references in the above mentioned monographs and in the commented bibliography Dempe (2003). The following acronyms will be used throughout the whole text.

MPEC	mathematical program with equilibrium constraints
MPCC	mathematical program with complementarity constraints
ImP	implicit programming approach
NLP	nonlinear programming
KKT	Karush-Kuhn-Tucker
CQ	constraint qualification
MFCQ	Mangasarian-Fromowitz CQ
GMFCQ	generalized MFCQ
SCQ	Slater CQ
ESCQ	extended SCQ
LICQ	linear independence CQ
ELICQ	extended LICQ
MPEC - LICQ	LICQ for MPEC
VI	variational inequality
QVI	quasi-variational inequality
HVI	hemivariational inequality
CP	complementarity problem
NCP	nonlinear complementarity problem
ICP	implicit complementarity problem
GE	generalized equation
NE	nonsmooth equation
SRC	strong regularity condition
SSOSC	strong second-order sufficient condition
BT	bundle-trust
QP	quadratic programming
SQP	sequential QP
PC^1	piecewise continuously differentiable
lsc	lower semicontinuous
usc	upper semicontinuous

1 Basic tools from nonsmooth analysis

Nonsmooth analysis represents a powerful tool enabling us to work with single- and set-valued mappings without classical differentiability assumptions. In some problems already the data exhibit a lack of differentiability, in another problems the nondifferentiability is inherent even if all data are sufficiently smooth (eg when dealing with the so-called value functions). In a fully convex framework the main part of this theory has been completed already in the sixties. Without convexity assumptions, however, the situation becomes substantially more difficult and some approaches can be applied only to certain special classes of problems. As a result, we dispose currently with several approaches with various degrees of generality and various advantages and drawbacks. Nevertheless, as shown in the excellent monograph by Rockafellar and Wets (1998), most of them can be derived on basis of a general framework which has been gradually built up in the last three decades of the past century. Following this framework, we describe the local behaviour of sets, real-valued functions and set-valued mappings (multifunctions) by means of notions defined either in the original space (at multifunctions it is the space, where their graphs are living), or in its topological dual. In this way one describes the local behaviour of sets via tangent and normal cones, the local behaviour of real-valued functions by means of directional derivatives and subdifferentials and the local behaviour of multifunctions through graphical derivatives and coderivatives. One speaks about six pillars of nonsmooth analysis. Not all existing approaches provide us with all of them, but within the same approach all available "pillars" are based on the same principle and closely related. In this series of lectures we will make use of two approaches that grew up from the classical notion of strict derivative; it is the approach of F.H. Clarke (Clarke (1983)) and B.S. Mordukhovich (Mordukhovich (1988)). In the latter one disposes merely with the dual notions (normal cones, subdifferentials and coderivatives) but they enjoy a number of excellent properties. We will introduce, however, only those notions which are indispensable for the understanding of the main part of these lectures. Thereafter we will turn our attention to Lipschitzian properties of multifunctions, playing a crucial role both in optimality conditions as well as in the numerical solution of optimization problems. The first lecture will be closed by stating necessary optimality conditions for several types of frequently arising optimization problems.

1.1 Approach of B. Mordukhovich

Consider a closed set $\Pi \in \mathbb{R}^l$ and a point $\bar{z} \in \Pi$. The cone

$$\widehat{N}_\Pi(\bar{z}) := \{\xi \in \mathbb{R}^l | \langle \xi, z - \bar{z} \rangle \leq o(\|z - \bar{z}\|) \forall z \in \Pi\}$$

is called the *Fréchet normal cone* to Π at $\bar{z}$. This cone is convex and closed and it is the negative polar to the *contingent (Bouligand)* cone

$$T_\Pi(\bar{z}) := \limsup_{\lambda \downarrow 0} \frac{\Pi - \bar{z}}{\lambda}. \tag{1.1.1}$$

If Π is convex, $\widehat{N}_\Pi(\bar{z})$ coincides with the classical normal cone in the sense of convex analysis.

Definition 1.1 The cone

$$N_\Pi(\bar{z}) := \limsup_{z \xrightarrow{\Pi} \bar{z}} \widehat{N}_\Pi(z)$$

is called the *limiting normal cone* to Π at $\bar{z}$.

The " lim sup " in the above definitions is the upper limit of multifunctions in the sense of Painléve-Kuratowski, cf. Aubin and Frankowska (1990). $N_\Pi(\bar{z})$ is also closed, but not necessarily convex. By definition, one always has the inclusion

$$\widehat{N}_\Pi(\bar{z}) \subset N_\Pi(\bar{z}).$$

If this inclusion becomes equality, we say that Π is *normally regular* at $\bar{z}$. Each convex set is normally regular at all its points. Frequently, Π amounts to the intersection $g^{-1}(\Lambda)\cap\Xi$, where Ξ, Λ are closed sets in $\mathbb{R}^l, \mathbb{R}^p$, respectively, and $g[\mathbb{R}^l \to \mathbb{R}^p]$ is a continuously differentiable operator. In this case the following statement can be useful.

Theorem 1.1 (Rockafellar and Wets, 1998, Thm.6.14). Let $\bar{z} \in \Pi = g^{-1}(\Lambda)\cap\Xi$. Then one has

$$\widehat{N}_\Pi(\bar{z}) \supset (\nabla g(\bar{z}))^T \widehat{N}_\Lambda(g(\bar{z})) + \widehat{N}_\Xi(\bar{z}). \tag{1.1.2}$$

Under the qualification condition

$$\left.\begin{array}{c} (\nabla g(\bar{z}))^T u^* \in -N_\Xi(\bar{z}) \\ u^* \in N_\Lambda(g(\bar{z})) \end{array}\right\} \Rightarrow u^* = 0 \tag{1.1.3}$$

it holds further that

$$N_\Pi(\bar{z}) \subset (\nabla g(\bar{z}))^T N_\Lambda(g(\bar{z})) + N_\Xi(\bar{z}). \tag{1.1.4}$$

By combining of (1.1.2) and (1.1.4) it follows that

$$N_\Pi(\bar{z}) = \widehat{N}_\Pi(\bar{z}) = (\nabla g(\bar{z}))^T N_\Lambda(g(\bar{z})) + N_\Xi(\bar{z}) \tag{1.1.5}$$

whenever condition (1.1.3) is fulfilled, Ξ is normally regular at $\bar{z}$ and Λ is normally regular at $g(\bar{z})$. We also immediately observe that condition (1.1.3) holds true provided $\Xi = \mathbb{R}^l$ and $\nabla g(\bar{z})$ is surjective. Equation (1.1.5) is valid without any assumptions, provided Ξ, Λ are convex polyhedra and g is affine.

In some cases $\Pi = \Pi_1 \times \Pi_2$ with $\Pi_1 \in \mathbb{R}_1^{l_1}, \Pi_2 \in \mathbb{R}^{l_2}, l_1 + l_2 = l$. Accordingly $z = (z_1, z_2)$ with $z_1 \in \Pi_1$ and $z_2 \in \Pi_2$. Then one has (Mordukhovich, 1988, Prop.1.6)

$$N_\Pi(z) = N_{\Pi_1}(z_1) \times N_{\Pi_2}(z_2). \tag{1.1.6}$$

We turn now our attention to real-valued functions.

Definition 1.2 Consider a lsc function $f[\mathbb{R}^l \to \bar{\mathbb{R}}]$ and assume that $z \in \mathrm{dom} f$. The sets

$$\widehat{\partial} f(z) = \{\xi \in \mathbb{R}^n | (\xi, -1) \in \widehat{N}_{\mathrm{epi} f}(z, f(z))\},$$
$$\partial f(z) = \{\xi \in \mathbb{R}^n | (\xi, -1) \in N_{\mathrm{epi} f}(z, f(z))\}$$

are called the *Fréchet subdifferential* and the *(lower) limiting subdifferential* of f at z, respectively.

$\partial f(z)$ is a closed, possibly nonconvex set which is bounded iff f is Lipschitz on a neighborhood of z. Putting $f = \delta_\Pi$, we observe that $\mathrm{epi}\delta_\Pi = \Pi \times \mathbb{R}_+$ and consequently, by virtue of (1.1.6),

$$\partial\delta_\Pi(z) = \{\xi \in \mathbb{R}^n | (\xi, -1) \in N_{\mathrm{epi}\delta_\Pi}(z, 0)\} = \{\xi \in \mathbb{R}^n | (\xi, -1) \in N_\Pi(z) \times N_{\mathbb{R}_+}(0)\} = N_\Pi(z)$$

whenever $z \in \Pi$. So, we have established a two-way relation between the limiting normal cone and the limiting subdifferential by means of the epigraph of the considered function and the indicatory function of the considered set. The set $\partial f(z)$ is a singleton iff f is strictly differentiable at z. In such a case $\partial f(z)$ amounts to the strict derivative of f at z.

Consider now a multifunction $\Phi[\mathbb{R}^l \rightsquigarrow \mathbb{R}^p]$ having a closed graph and assume that $b \in \Phi(a)$, i.e., $(a, b) \in \mathrm{Gph}\Phi$. The local behaviour of Φ around (a, b) can be described via another multifunction, denoted $D^*\Phi(a, b)$, defined by

$$D^*\Phi(a, b)(\eta) := \{\xi \in \mathbb{R}^l | (\xi, -\eta) \in N_{\mathrm{Gph}\Phi}(a, b)\},\ \eta \in \mathbb{R}^p. \tag{1.1.7}$$

This new multifunction is entitled the *coderivative* of Φ at (a, b). If the limiting normal cone in (1.1.7) is replaced by the Fréchet normal cone, we write $\widehat{D}^*\Phi$ and speak about the *regular coderivative.* If Φ is single-valued, we write simply $D^*\Phi(a)(\cdot)$, because $b = \Phi(a)$. If Φ is even continuously differeentiable at a, then one has $D^*\Phi(a)(\eta) = (\nabla\Phi(a))^T\eta\ \forall\, \eta \in \mathbb{R}^p$. In the single-valued case we can consider the composition

$$\varphi(z) := f(\Phi(z)), \tag{1.1.8}$$

where the outer function f maps $\mathbb{R}^p$ into $\overline{\mathbb{R}}$. Consider $\bar{z} \in \mathbb{R}^l$ such that $\Phi(\bar{z}) \in \mathrm{dom} f$. Then one has, cf. Rockafellar and Wets (1998, Thm.10.49),

$$\partial\varphi(\bar{z}) \subset D^*\Phi(\bar{z})(\partial f(\Phi(\bar{z}))), \tag{1.1.9}$$

provided Φ is Lipschitz around $\bar{z}$ and f is Lipschitz around $\Phi(\bar{z})$. If f is even continuously differentiable around $\Phi(\bar{z})$, it holds

$$\partial\varphi(\bar{z}) \supset \widehat{D}^*\Phi(\bar{z})(\nabla f(\Phi(\bar{z}))). \tag{1.1.10}$$

1.2 Approach of F. Clarke

In this section we confine ourselves only to two notions: The Clarke's subdifferential and the generalized Jacobian. Moreover, we will consider only Lipschitz mappings which play a crucial role in applications, where these notions will be employed.

Definition 1.3. Let $F[\mathbb{R}^l \to \mathbb{R}^p]$ be Lipschitz near $\bar{z} \in \mathbb{R}^l$ and let ω_F denote the set, where F fails to be differentiable. Let

$$\partial_B F(\bar{z}) := \left\{ \lim_{i\to\infty} \nabla F(z_i) | z_i \to \bar{z},\ z_i \notin \omega_F \right\}.$$

Then the set $\bar{\partial}F(\bar{z}) := \mathrm{conv}\partial_B F(\bar{z})$ is called the *Clarke's subdifferential* of F at $\bar{z}$, provided $p = 1$, and the *generalized Jacobian* of F at $\bar{z}$ otherwise.

The Clarke's subdifferential can also be defined via the directional derivative

$$F^0(\bar{z}; h) := \limsup_{z \to \bar{z}, \lambda \downarrow 0} \frac{F(z + \lambda h) - F(z)}{\lambda}.$$

Indeed, it holds that

$$\bar{\partial} F(\bar{z}) = \{\xi \in \mathbb{R}^l | \langle \xi, h \rangle \leq F^0(\bar{z}; h) \forall h \in \mathbb{R}^l\}.$$

Observe that this set is nonempty, closed and convex. Further, one has

$$\bar{\partial} F(\bar{z}) = \operatorname{conv} \partial F(\bar{z}). \tag{1.2.11}$$

If $p > 1$, then for all $y^* \in \mathbb{R}^p$

$$(\bar{\partial} F(\bar{z}))^T y^* = \operatorname{conv} D^* F(\bar{z})(y^*). \tag{1.2.12}$$

So, in the Lipschitz case the Clarke's notions are closely related to their limiting counterparts. Using generalized Jacobians, one can enrich the statements concerning the composition (1.1.8). According to (Clarke, 1983, Thm.2.6.6) one has namely that

$$\bar{\partial} \varphi(\bar{z}) = (\bar{\partial} \Phi(\bar{z}))^T \nabla f(\Phi(\bar{z})), \tag{1.2.13}$$

whenever f is continuously differentiable around $\Phi(\bar{z})$. The Clarke's notions can very well be applied to the so-called piecewise $C^1(PC^1)$-maps.

Definition 1.4 Let U be an open subset of $\mathbb{R}^l$. A function $F[U \to \mathbb{R}^p]$ is called a $PC^1 - function$ on U if it is continuous and if for every $z_0 \in U$ there exists an open neighborhood $\mathcal{O} \subset U$ and a finite number of continuously differentiable functions $F_i[\mathcal{O} \to \mathbb{R}^p],\ i = 1, 2, \ldots, k$, such that for every $z \in \mathcal{O}$ one has $F(z) \in \{F_1(z), \ldots, F_k(z)\}$. The single functions F_i are called *selections* or *pieces* for F at z_0, and the set

$$I_F(z_0) := \{i \in \{1, 2, \ldots, k\} \mid F(z_0) = F_i(z_0)\}$$

is the *active index set* at z_0. The selections F_i, $i \in I_F(z_0)$, are called *active selections (pieces)* for F at z_0. A selection F_i is *essentially active* at z_0, provided

$$z_0 \in \operatorname{cl}(\operatorname{int}\{z \in \mathcal{O} \mid F(z) = F_i(z)\}).$$

The respective index set of essentially active selections at z_0 is denoted by $I^e_F(z_0)$ and plays an important role in connection with the generalized Jacobian of F at z_0. Indeed, in Scholtes (1994) it was proved that every PC^1-function is locally Lipschitz and

$$\partial F(z_0) = \operatorname{conv}\{\nabla F_i(z_0) \mid i \in I^e_F(z_0)\}. \tag{1.2.14}$$

When deriving necessary optimality conditions for nonsmooth optimization problems, it is, by virtue of (1.2.11),(1.2.12), definitely better to work with the limiting notions than with the Clarke's subdifferentials and generalized Jacobians. The resulting conditions are then generally sharper, i.e., more selective. On the other hand, when solving nondifferentiable problems numerically, then the convexity of the Clarke's notions plays an important role. In fact, numerical methods based on the Clarke's subdifferential are dominating in nonsmooth minimization problems, whereas the generalized Jacobian is widely used in the solution of nonsmooth equations.

1.3 Lipschitz properties of multifunctions

When dealing with MPECs, one has to work with various multifunctions exhibiting various Lipschitz-like properties. It is one of the tasks of nonsmooth analysis to detect these properties, because they are of a great importance both in optimality conditions as well as in numerical procedures. Moreover, it is also an interesting issue to learn, how stable is a computed solution with respect to small perturbations of the problem data. For our purposes, the three notions defined below are the most relevant.

Definition 1.5 Consider a multifunction $\Gamma[\mathbb{R}^l \rightsquigarrow \mathbb{R}^p]$.

(i) Γ is called *(locally) upper Lipschitz* at $\bar{z} \in \mathrm{dom}\Gamma$ (with modulus L), if there is a neighborhood $\mathcal{U}$ of $\bar{z}$ and a real $L \geq 0$ such that

$$\Gamma(z) \subset \Gamma(\bar{z}) + L\|z - \bar{z}\|\mathbb{B} \text{ for all } z \in \mathcal{U}.$$

(ii) We say that Γ has the *Aubin property* around $(a, b) \in \mathrm{Gph}\Gamma$ (with modulus L), if there are neighborhoods $\mathcal{U}$ of a, $\mathcal{V}$ of b and a real $L \geq 0$ such that

$$\Gamma(a') \cap \mathcal{V} \subset \Gamma(a'') + L\|a' - a''\|\mathbb{B} \text{ for all } a', a'' \in \mathcal{U}.$$

(iii) We say that Γ has a *Lipschitz single-valued localization* at $(a, b) \in \mathrm{Gph}\Gamma$, provided there are neighborhoods $\mathcal{U}$ of a, $\mathcal{V}$ of b and a Lipschitz single-valued map $\sigma[\mathcal{U} \to \mathbb{R}^p]$ such that $b = \sigma(a)$ and $\Gamma(a') \cap \mathcal{V} = \{\sigma(a')\}$ for all $a' \in \mathcal{U}$.

As stated in the next theorem, property (i) is possessed by a large class of the so-called polyhedral multifunctions. A multifunctions $P[\mathbb{R}^l \rightsquigarrow \mathbb{R}^p]$ is called *polyhedral*, provided its graph is the union of finitely many convex polyhedra, termed *components*.

Theorem 1.2 (Robinson (1981)) Let $P[\mathbb{R}^l \rightsquigarrow \mathbb{R}^p]$ be polyhedral. Then there is a constant λ such that P is (locally) upper Lipschitz with modulus λ at each $z \in \mathrm{dom}P$. In particular, if $\Omega \subset \mathbb{R}^l$ is a convex polyhedron, then the map

$$P : z \mapsto \begin{cases} N_\Omega(z) \text{ if } z \in \Omega \\ \emptyset \text{ otherwise,} \end{cases}$$

being polyhedral, is (locally) upper Lipschitz at each $z \in \Omega$.

From Theorem 1.2 it follows that a single-valued polyhedral map P is necessarily Lipschitz in the standard sense on $\mathrm{dom}P$. For property (ii) we dispose with a characterization in terms of coderivatives. Indeed, Γ has the Aubin property around $(a, b) \in \mathrm{Gph}\Gamma$ iff $D^*\Gamma(a, b)(0) = \{0\}$, cf. Mordukhovich (1994). This characterization is known as Mordukhovich criterion. If Γ has a concrete structure, then this characterization, together with rules of the coderivative calculus enables to obtain a sufficient condition ensuring the Aubin property. Consider eg the multifunction $\Sigma[\mathbb{R}^p \rightsquigarrow \mathbb{R}^l]$ defined by

$$\Sigma(u) := \{z \in \Xi \mid g(z) + u \in \Lambda\}, \tag{1.3.15}$$

where g, Ξ, Λ fulfill the assumptions posed before Theorem 1.1.

Theorem 1.3. Let $\bar{z} \in \Sigma(\bar{u})$ and assume that the qualification condition (1.1.3) holds true. Then Σ has the Aubin property around $(\bar{u}, \bar{z})$.

Property (iii) is closely related to the implicit programming approach investigated in Section 5. Also for this property nonsmooth analysis provides us with a suitable characterization, cf. Klatte and Kummer (2003). Instead of it, however, we will apply later two other criteria which are tailored to the concrete structure of Γ being investigated.

1.4 Necessary optimality conditions

In this section we will present a general framework suitable for deriving optimality conditions in MPECs. We start with an abstract optimization problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & \\ & z \in \Pi, \end{array} \tag{1.4.16}$$

where $z \in \mathbb{R}^l$, $f[\mathbb{R}^l \to \mathbb{R}]$ is locally Lipschitz and $\Pi \subset \mathbb{R}^l$ is closed. If $\widehat{z}$ is a local minimizer in (1.4.16), then, following Mordukhovich (1988), one has

$$0 \in \partial f(\widehat{z}) + N_\Pi(\widehat{z}), \tag{1.4.17}$$

and the main problem consists in the computation of the limiting normal cone $N_\Pi(\widehat{z})$ (or an upper approximation of it). In MPECs the set Π has frequently the structure investigated in Theorem 1.1, i.e.,

$$\Pi = \{z \in \Xi \,|\, g(z) \in \Lambda\}. \tag{1.4.18}$$

Theorem 1.4. Let $\widehat{z}$ be a local minimizer in (1.4.16), where Π is given by (1.4.18) and g, Ξ, Λ fulfill the assumptions posed before Theorem 1.1. Further assume that the qualification condition (1.1.3) holds true (with $\bar{z}$ replaced by $\widehat{z}$). Then there exists a Karush-Kuhn-Tucker (KKT) vector $u^* \in N_\Lambda(g(\widehat{z}))$ such that

$$0 \in \partial f(\widehat{z}) + (\nabla g(\widehat{z}))^T u^* + N_\Xi(\widehat{z}). \tag{1.4.19}$$

If g is affine and Ξ, Λ are convex polyhedral, then relation (1.4.19) holds true without any qualification conditions.

Proof. Under the posed assumptions one has that

$$N_\Pi(\widehat{z}) \subset (\nabla g(\widehat{z}))^T N_\Lambda(g(\widehat{z})) + N_\Xi(\widehat{z}).$$

The existence of a KKT vector $u^* \in N_\Lambda(g(\widehat{z}))$ thus follows from (1.4.17). □

If $\Lambda = \mathbb{R}^p_+$ condition (1.1.3) can be written down in the form

$$\left.\begin{array}{ll} 0 \in & \sum\limits_{i=1}^{p} u^{*i} \nabla g^i(\bar{z}) + N_\Xi(\bar{z}) \\ & u^{*i} \in \mathbb{R}^p_+,\ u^{*i} = 0 \text{ if } g^i(\bar{z}) < 0 \end{array}\right\} \Rightarrow u^* = 0,$$

where we recognize the dual form of the Mangasarian-Fromowitz constraint qualification (MFCQ), known from mathematical programming. In equilibrium problems, Λ is usually a complicated nonconvex set and therefore we will use the acronym GMFCQ. If f is continuously differentiable, then, following Hiriart-Urruty (1979), it is possible to replace condition (1.4.17) by a sharper one, namely

$$0 \in \nabla f(\widehat{z}) + \widehat{N}_\Pi(\widehat{z}). \tag{1.4.20}$$

The hurdle consists now in the computation of $\widehat{N}_\Pi(\widehat{z})$. In this regard, however, additional structural assumptions on Π can be helpful.

Theorem 1.5. (Flegel et al. (2004)) Let $\widehat{z}$ be a local minimizer in (1.4.16), where f is continuously differentiable. Let $\Pi = g^{-1}(\Lambda)$ with g like in Theorem 1.3 and assume that Λ is the union of polyhedral convex sets $\Lambda_i, i = 1, 2, \dots, q$. Finally suppose that the constraint qualification

$$T_{g^{-1}(\Lambda)}(\widehat{z}) = \{h \in \mathbb{R}^l \mid \nabla g(\widehat{z})h \in T_\Lambda(g(\widehat{z}))\} \tag{1.4.21}$$

holds true and with $I(\widehat{z}) := \{i \in \{1, 2, \dots, q\} \mid g(\widehat{z}) \in \Lambda_i\}$ the condition

$$\bigcap_{i \in I(\widehat{z})} (\nabla g(\widehat{z}))^T \widehat{N}_{\Lambda_i}(g(\widehat{z})) = (\nabla g(\widehat{z}))^T \bigcap_{i \in I(\widehat{z})} \widehat{N}_{\Lambda_i}(g(\widehat{z})) \tag{1.4.22}$$

is fulfilled. Then there exists a KKT vector $u^* \in \widehat{N}_\Lambda(g(\widehat{z}))$ such that

$$0 \in \nabla f(\widehat{z}) + (\nabla g(\widehat{z}))^T u^*. \tag{1.4.23}$$

Proof. By our assumptions $\Lambda = \bigcup\limits_{i=1}^{q} \Lambda_i$ with all sets Λ_i being convex polyhedra. Thus, by (1.4.21), one has

$$T_{g^{-1}(\Lambda)}(\widehat{z}) = \{h \in \mathbb{R}^l \mid \nabla g(\widehat{z})h \in T_{\bigcup\limits_{i=1}^{q} \Lambda_i}(g(\widehat{z}))\} = \bigcup_{i \in I(\widehat{z})} \{h \in \mathbb{R}^l \mid \nabla g(\widehat{z})h \in T_{\Lambda_i}(g(\widehat{z}))\}.$$

Indeed, the contingent cone to the union $\bigcup\limits_{i=1}^{q} \Lambda_i$ at $g(\widehat{z})$ amounts to the union of the contingent cones to the single sets Λ_i for $i \in I(\widehat{z})$ (at $g(\widehat{z})$). Since the polar to a union is the intersection of the polars, we infer that

$$\widehat{N}_{g^{-1}(\Lambda)}(\widehat{z}) = \bigcap_{i \in I(\widehat{z})} \left[(\nabla g(\widehat{z}))^{-1}(T_{\Lambda_i}(g(\widehat{z})))\right]^0.$$

The polars on the right-hand side of the preceding equation can be computed due to the polyhedrality of Λ_i; one obtains that

$$\left[(\nabla g(\widehat{z}))^{-1}(T_{\Lambda_i}(g(\widehat{z})))\right]^0 = (\nabla g(\widehat{z}))^T \widehat{N}_{\Lambda_i}(g(\widehat{z})).$$

It remains to observe that

$$\widehat{N}_\Lambda(g(\widehat{z})) = \bigcap_{i\in I(\widehat{z})} \widehat{N}_{\Lambda_i}(g(\widehat{z})),$$

employ condition (1.4.22), and we are done. □

It is easy to see that condition (1.4.22) is fulfilled whenever $\nabla g(\widehat{z})$ is surjective or $|I(\widehat{z})| = 1$. Surjectivity of $\nabla g(\widehat{z})$ guarantees simultaneously the satisfaction of condition (1.4.21). In classical mathematical programming Λ amounts to the carthesian product of zeros (for equality constraints) and a nonpositive orthant (for inequality constraints). In this case condition (1.4.21) is called the *Abadie* constraint qualification and is known to be one of the weakest constraint qualifications at all. We will keep this terminology also in the considered general situation. It has been proved eg in Henrion and Outrata (to appear) that (1.4.21) is fulfilled whenever the map

$$u \mapsto \left\{z \in \mathbb{R}^l \mid g(z) + u \in \Lambda\right\}$$

possesses the Aubin property around $(0, \widehat{z})$. Since (GMFCQ) implies the Aubin property by Theorem 1.3, we have the implication

$$\text{(GMFCQ)} \Rightarrow \text{(Abadie)}$$

exactly like in the classical situation. It follows that under (GMFCQ) the optimality conditions (1.4.19) can be sharpened to the form (1.4.23) whenever f is continuously differentiable, $\Xi = \mathbb{R}^l, \Lambda$ has the structure specified above and condition (1.4.22) is fulfilled.

Some MPECs can be converted to the form

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \\ & y = S(x) \\ & x \in \omega, \end{array} \tag{1.4.24}$$

where $x \in \mathbb{R}^n, y \in \mathbb{R}^m, f[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ and $S[\mathbb{R}^n \to \mathbb{R}^m]$ are locally Lipschitz and $\omega \subset \mathbb{R}^n$ is closed. In such a case one can apply the following statement.

Theorem 1.6. Let $(\widehat{x}, \widehat{y})$ be a locally optimal pair in (1.4.24). Then there exist vectors $(\xi, \eta) \in \partial f(x,y)$ such that

$$0 \in \xi + D^*S(\widehat{x})(\eta) + N_\omega(\widehat{x}). \tag{1.4.25}$$

Proof. Problem (1.4.24) can be converted to the form

$$\begin{array}{ll} \text{minimize} & h(x) \\ \text{subject to} & \\ & x \in \omega, \end{array}$$

where $h := f \circ \Phi$ with $\Phi(x) := \begin{bmatrix} x \\ S(x) \end{bmatrix}$. Therefore, for an arbitrary pair $(a, b) \in \mathbb{R}^n \times \mathbb{R}^m$ one has that

$$D^* \Phi(x)(a, b) = a + D^* S(x)(b).$$

The statement follows now from inclusion (1.1.9) and optimality condition (1.4.17). $\square$

In some MPECs considered in Chapters 4 and 5, f is continuously differentiable, but we are able to compute only an upper approximation of the coderivative $D^* S(\widehat{x})(\nabla_y f(\widehat{x}, \widehat{y}))$. This can be the set $(\bar{\partial} S(\widehat{x}))^T \nabla_y f(\widehat{x}, \widehat{y})$ or even its upper approximation. In this way we arrive naturally at weaker (less selective) optimality conditions.

2 Mathematical modeling of discretized equilibrium problems

Variational inequalities (VIs) and complementarity problems (CPs) provide a convenient and elegant tool for modeling of manifold equilibria. In numerous situations, however, it is advantageous to work with another models, above all with the so-called generalized and nonsmooth equations. Following Robinson, under *generalized equation* (GE) we understand the relation

$$0 \in F(y) + Q(y),$$

where F is a single- and Q is a set-valued mapping. In our models F is usually continuously differentiable. If F is merely locally Lipschitz and Q disappears, we speak about a *nonsmooth equation* (NE). In this chapter we will

(i) reformulate several fundamental types of VIs to the form of GEs and NEs;

(ii) state some basic existence and uniqueness results, and

(iii) describe briefly an efficient numerical method, associated with the NE model.

2.1 Basic reformulations

This section is devoted to the modeling of equilibria that are standardly described via variational inequalities of the first and second kind, hemivariational inequalities (HVIs) and quasi-variational inequalities (QVIs). As a particular case of VIs of the first kind we will introduce complementarity problems and mixed complementarity problems. A special attention is paid to (discretized) contact problems with Coulomb friction which can be modeled in various ways. Let $y \in \mathbb{R}^m$, $F[\mathbb{R}^m \to \mathbb{R}^m]$ be continuous and $\Omega \subset \mathbb{R}^m$ be convex and closed. The problem

$$\left.\begin{array}{ll} \text{find} & \bar{y} \in \Omega \text{ such that} \\ & \langle F(\bar{y}), y - \bar{y} \rangle \geq 0 \ \forall\, y \in \Omega \end{array}\right\} \tag{2.1.1}$$

is called the *variational inequality of the first kind*. Ω is termed the *constraint set*. If $\Omega = \mathbb{R}^m_+$, (2.1.1) amounts to a *complementarity problem* and if Ω is a box in $\mathbb{R}^m$, we speak about a *mixed complementarity problem*. In dependence on the nature of F these CPs can be linear (if F is affine) or nonlinear (otherwise). If F is monotone, continuously differentiable and the Jacobian ∇F is symmetric, then (2.1.1) is equivalent to the convex program

$$\begin{array}{ll} \text{minimize} & f(y) \\ \text{subject to} & \\ & y \in \Omega, \end{array} \tag{2.1.2}$$

where $F = \nabla f$. Equilibria of this type arise frequently in connection with (discretized) contact problems in linear elasticity. One readily deduces that (2.1.1) amounts to the GE

$$0 \in F(y) + N_\Omega(y) \tag{2.1.3}$$

and to the NE

$$\mathrm{Proj}_\Omega(y - F(y)) - y = 0. \tag{2.1.4}$$

Since $\mathrm{Proj}_{\mathbb{R}^m_+}(v) = \max\{v, 0\}$, equation (2.1.4) attains in case of a CP a particularly simple form

$$\min\{F(y), y\} = 0. \tag{2.1.5}$$

Consider now instead of (2.1.1) the problem

$$\left.\begin{array}{l} \text{find } \bar{y} \in \mathbb{R}^m \text{ such that} \\ \langle F(\bar{y}), y - \bar{y}\rangle + \varphi(y) - \varphi(\bar{y}) \geq 0 \ \forall \ y \in \mathbb{R}^m, \end{array}\right\} \tag{2.1.6}$$

where $\varphi[\mathbb{R}^m \to \overline{\mathbb{R}}]$ is proper convex and lsc. This problem is entitled the *variational inequality of the 2nd kind* and generalizes the model (2.1.1). Indeed, it suffices to put $\varphi = \delta_\Omega$ to obtain (2.1.1). With other choices of φ, however, we can model rather complicated equilibria arising eg in (discretized) contact problems with given friction. By virtue of the convexity of φ, (2.1.6) can also be written down in an equivalent form:

$$\left.\begin{array}{l} \text{Find } y \in \mathbb{R}^m \text{ such that} \\ \langle F(\bar{y}), y - \bar{y}\rangle + \varphi'(\bar{y}, y - \bar{y}) \geq 0 \ \forall \ y \in \mathbb{R}^m. \end{array}\right\} \tag{2.1.7}$$

From (2.1.7) we can now immediately derive the corresponding GE

$$0 \in F(y) + \partial\varphi(y). \tag{2.1.8}$$

If φ in (2.1.6) is not convex, but locally Lipschitz, it is possible to replace the classical directional derivative in (2.1.7) by the directional derivative of Clarke. In this way we arrive at a *hemivariational inequality*. In fact, in the literature (eg Panagiotopoulos (1993)) HVI has been introduced in a more general form. In the sequel, however, we will come out with this simplified formulation. As the counterpart to (2.1.7) we obtain now the GE

$$0 \in F(y) + \bar{\partial}\varphi(y). \tag{2.1.9}$$

By *constrained hemivariational inequality* we with mean the problem:

$$\left.\begin{array}{l} \text{Find } \bar{y} \in \Omega \text{ such that} \\ \langle F(\bar{y}), y - \bar{y}\rangle + \varphi^0(\bar{y}; y - \bar{y}) \geq 0 \ \forall \ y \in \Omega \end{array}\right\}, \tag{2.1.10}$$

where $\varphi[\mathbb{R}^m \to \mathbb{R}]$ is locally Lipschitz and the constraint set Ω is closed and convex. In this case the equivalent GE attains the form

$$0 \in F(y) + \bar{\partial}\varphi(y) + N_\Omega(y). \tag{2.1.11}$$

In this reasoning the convexity of Ω plays an important role. HVIs are frequently used to the modeling of the behaviour of laminates or sandwiches. If F is continuously differentiable and ∇F is symmetric, problem (2.1.10) is a first-order necessary optimality condition for the nonsmooth program

$$\begin{array}{ll} \text{minimize} & (f+\varphi)(y) \\ \text{subject to} & \\ & y \in \Omega, \end{array} \tag{2.1.12}$$

where again $F = \nabla f$. To be precise, the solutions of (2.1.10) are then *C(larke)- stationary* points of the program (2.1.12). Since in some cases not all these points are acceptable equilibria from the physical point of view, it is then reasonable to replace (2.1.11) by the GE

$$0 \in F(y) + \partial\varphi(y) + N_\Omega(y), \tag{2.1.13}$$

where the Clarke's subdifferential is replaced by the (smaller) limiting one. The GE (2.1.13), however, does not have a corresponding HVI due to the nonconvexity of the limiting subdifferential.

Let $\Gamma[\mathbb{R}^m \rightsquigarrow \mathbb{R}^m]$ be a closed- and convex-valued multifunction, F be like in (2.1.1) and consider the problem:

$$\left.\begin{array}{l} \text{Find } \bar{y} \in \Gamma(\bar{y}) \text{ such that} \\ \langle F(\bar{y}), y - \bar{y}\rangle \geq 0 \, \forall y \in \Gamma(\bar{y}) \end{array}\right\}. \tag{2.1.14}$$

Since the constraint set (counterpart to Ω in (2.1.1)) depends now on y, we speak about a *quasi-variational inequality.* If $\Gamma(y) = g(y) + \mathbb{R}^m_+$ with all components $g^i[\mathbb{R}^m \to \mathbb{R}], i = 1, 2, \ldots, m$, continuous, we arrive at the *implicit complementarity problem* (ICP):

$$\left.\begin{array}{l} \text{Find } \bar{y} \in \mathbb{R}^m \text{ such that} \\ F(\bar{y}) \geq 0, \bar{y} \geq g(\bar{y}), \langle F(\bar{y}), \bar{y} - g(\bar{y})\rangle = 0. \end{array}\right\} \tag{2.1.15}$$

Problems of the type (2.1.15) arise eg in (discretized) contact problems with compliant obstacles or in filtration through pourous media. (2.1.14) amounts evidently to the GE

$$0 \in F(y) + N_{\Gamma(y)}(y) \tag{2.1.16}$$

(with a rather difficult set-valued part) or to the NE

$$\mathrm{Proj}_{\Gamma(y)}(y - F(y)) - y = 0. \tag{2.1.17}$$

In ICPs one has

$$\mathrm{Proj}_{\Gamma(y)}(v) = \max\{v, g(y)\}.$$

Eq. (2.1.17) attains then the form

$$\max\{y - F(y), g(y)\} - y = 0$$

which is equivalent to

$$\min\{F(y), y - g(y)\} = 0. \tag{2.1.18}$$

From eq. (2.1.18) one immediately deduces the original formulation (2.1.15). Due to its simplicity, eq. (2.1.18) is frequently used in various contexts. Naturally, for $g \equiv 0$ we obtain the standard CP. As we have seen, most equilibria can be modeled as variational inequalities and generalized equations. By using of duality theory and artificial variables the modeling possibilities can be further enriched. This holds typically for contact problems with Coulomb friction. In Haslinger and Panagiotopoulos (1984) one finds a model in form of a QVI, whereas in A.M.Al-Fahed et al. (1991) the same problem is modeled as a linear CP. In the context of MPECs it is convenient to start with the GE (2.1.8) corresponding to the underlying (discretized) contact problem with given friction. In this problem $y = (y_1, y_2, y_3)^T, F = (F_1, F_2, F_3)^T$, φ depends on y_1 and a nonnegative parameter ν (coefficient of the given friction) and the constraint set Ω concerns only y_3. Correspondingly, the respective GE (2.1.8) attains the form

$$\begin{array}{rcl} 0 & \in & F_1(y) + \partial\varphi_\nu(y_1) \\ 0 & = & F_2(y) \\ 0 & \in & F_3(y) + N_\Omega(y_3), \end{array} \tag{2.1.19}$$

where $\varphi_\nu(y_1) = \nu \bullet \widetilde{\varphi}(y_1)$ with a convex continuous function $\widetilde{\varphi}$. If we now put in (2.1.19) ν equal to y_3 (multiplier responsible for the nonpenetrability condition), we obtain a GE describing a (discretized) contact problem with Coulomb friction. This procedure demonstrates the usefulness of the GE modeling framework.

2.2 Some elementary existence and uniqueness results

In this section we confine ourselves to VIs of the first kind and to HVIs. The presented results can be found in standard literature so that our aim is rather to illustrate the advantages of various reformulations presented in the previous section.
Consider first the VI (2.1.1).

Theorem 2.1. Assume that F is continuous and either
(i) Ω is compact;
(ii) F is *coercive* on Ω, i.e., $\exists y_0 \in \mathbb{R}^m$ such that

$$\frac{\langle F(y) - F(y_0), y - y_0 \rangle}{\|y - y_0\|} \to \infty \text{ for } \|y\| \to \infty,\ y \in \Omega.$$

Then the VI (2.1.1) possesses a solution.

Sketch of the proof. Let (i) hold true. Then we can consider the NE (2.1.4) which can be written down as the equation

$$\Phi(y) = y$$

with $\Phi(y) := \text{Proj}_\Omega(y - F(y))$. Since the projection onto a convex set is continuous, Φ is a continuous map from Ω into Ω and the statement follows from the Brouwer's Fixed Point Theorem. To prove the statement under assumption (ii), we show first that the existence of a solution to the VI (2.1.1) is equivalent to the existence of a positive real r and a solution y_r to the GE

$$0 \in F(y) + N_{\Omega \cap r\mathbb{B}}(y) \tag{2.2.20}$$

such that $\|y_r\| < r$. The existence of r and y_r can then be ensured eg via the coercivity assumption (ii), cf. Ekeland and Temam (1976). □

Of course, one could construct also a combination of conditions (i), (ii) which would generalize the above statement. Assumption (ii) is fulfilled provided F is *strongly monotone* on Ω, i.e., $\exists \alpha > 0$ such that

$$\langle F(v) - F(w), v - w \rangle \geq \alpha \|v - w\|^2 \ \forall\ v, w \in \Omega.$$

Concerning the existence of solutions to the HVI (2.1.10), we can make use of the GE (2.1.11) and employ another strong tool from nonlinear analysis, namely the Ky Fan's Inequality.

Theorem 2.2. Assume that F is continuous and Ω is compact. Then the HVI (2.1.10) possesses a solution.

For the proof it is important to note that the multifunction

$$y \mapsto F(y) + \bar{\partial}\varphi(y)$$

is usc and nonempty-, convex- and compact-valued due to the properties of the Clarke's subdifferential. This enables to apply (Aubin, 1993, Thm.9.9) to the GE (2.1.11). If F is continuously differentiable and ∇F is symmetric, then one can employ the reformulation (2.1.12). In this optimization problem the objective is continuous and the constraint set is compact, whence the existence by virtue of the Weierstrass Theorem. This argumentation remains valid also for the GE (2.1.13) for which the preceding idea of the proof does not work.

Concerning the question of uniqueness of the solutions to equilibrium problems analyzed above, we restrict ourselves merely to the VIs of the first kind, because there is no chance to ensure uniqueness at HVIs by means of realistic assumptions. On the basis of the original formulation (2.1.1) it is easy to prove that this VI has at most one solution, provided F is *strictly monotone* on Ω, i.e.,

$$\langle F(v) - F(w), v - w \rangle > 0 \ \forall v, w \in \Omega, v \neq w.$$

Since the strong monotonicity apparently implies the strict monotonicity, the strong monotonicity of F on Ω ensures both the existence as well as the uniqueness of solutions to the VI (2.1.1). It is thus an important issue to find verifiable criteria for this property under, possibly, some additional assumptions. If F happens to be continuously differentiable, then one can enforce its strong monotonicity on Ω via a suitable condition imposed on the Jacobian of F. One usually requires that ∇F is *uniformly positive definite* on Ω, i.e., $\exists \alpha > 0$ such that

$$\langle d, \nabla F(y) d \rangle \geq \alpha \|d\|^2 \ \ \forall\, y \in \Omega \ \ \forall d \in \mathbb{R}^m.$$

For affine F, however, this condition leads to positive definiteness of the defining matrix which is unduly strong in some cases.

Theorem 2.3. Let $F(y) = Ay + b$ with an $m \times m$ matrix A and $b \in \mathbb{R}^m$. Assume that Ω is a cone and A is *strictly copositive* with respect to $\Omega - \Omega$, i.e.,

$$\langle d, Ad \rangle > 0 \;\forall\; d \in \Omega - \Omega, d \neq 0.$$

Then the respective VI (2.1.1) has a unique solution.

If, additionally, $\Omega = \mathbb{R}^m_+$, we dispose with a full characterization of the case, when the respective linear CP possesses a unique solution for each $b \in \mathbb{R}^m$: This happens iff A is a P-matrix (Murty (1988)). In many cases

$$\Omega = \{y \in \mathbb{R}^m | H(y) = 0,\ G(y) \leq 0\}, \tag{2.2.21}$$

where $H[\mathbb{R}^m \to \mathbb{R}^l]$ is an affine map and all components G^i of $G[\mathbb{R}^m \to \mathbb{R}^s]$ are convex and continuously differentiable. Let the classic Slater constraint qualification

(SCQ) : $\exists \bar{y} \in \mathbb{R}^m$ such that $H(\bar{y}) = 0$ and $G(\bar{y}) \in \text{int}\mathbb{R}^s_+$;

be valid. Then we can compute the normal cone $N_\Omega(y)$ at $y \in \Omega$ and obtain that

$$N_\Omega(y) = \left\{ (\nabla H(y))^T \mu + (\nabla G(y))^T \lambda | \mu \in \mathbb{R}^l, \lambda \in \mathbb{R}^s_+, \langle \lambda, G(y) \rangle = 0 \right\}.$$

Consequently, the GE (2.1.3) attains then the form

$$\left.\begin{array}{ll} 0 & = \mathcal{L}(y, \mu, \lambda) \\ 0 & = H(y) \\ 0 & \in -G(y) + N_{\mathbb{R}^s_+}(\lambda), \end{array}\right\} \tag{2.2.22}$$

where

$$\mathcal{L}(y, \mu, \lambda) := F(y) + (\nabla H(y))^T \mu + (\nabla G(y))^T \lambda$$

is the *Lagrangian* associated with the analyzed VI. In (2.2.22) we have a substantially simpler constraint set than Ω, but we have to do with additional variables μ, λ. The GE (2.2.22) is called the KKT form of a VI and plays an important role in the sequel. Concerning the existence and uniqueness of its solutions, we have the following simple statement, where

$$I(y) := \{i \in \{1, 2, \ldots, s\} | G^i(y) = 0\}$$

denotes the *set of active indices* at y.

Theorem 2.4. Let F be strongly monotone on Ω given by (2.2.21) and $\bar{y}$ be the unique solution of the corresponding GE (2.1.3). Assume that the linear independence constraint qualification (LICQ) holds true at $\bar{y}$, i.e., the matrix

$$\begin{bmatrix} \nabla H(\bar{y}) \\ (\nabla G(\bar{y}))_{I(\bar{y})} \end{bmatrix}$$

has the full row rank. Then the GE (2.2.22) possesses a unique solution $(\widehat{y}, \widehat{\mu}, \widehat{\lambda})$ with $\widehat{y} = \bar{y}$.

2.3 Nonsmooth Newton's method

In section 2.1 we have seen that various equilibrium problems can be written down as the NE

$$C(z) = 0, \tag{2.3.23}$$

where the function $C[\mathbb{R}^m \to \mathbb{R}^m]$ arises at the left-hand side of (2.1.4) or (2.1.17). If C enjoys some properties specified below, we can compute the respective equilibria by a "nonsmooth" variant of the Newton's method. Naturally such Newton's methods are based on first-order approximations of C (directional derivatives, generalized Jacobians, limit sets of Thibault); see eg Kummer (1992),Qi (1993). We shortly sketch the variant due to Qi and Sun (1993) which has been used in numerous mechanical applications.

We assume throughout this section that C is *semismooth* on $\mathbb{R}^m$, i.e., C is locally Lipschitz and for each $z \in \mathbb{R}^m$ the limit

$$\lim_{\substack{V \in \bar{\partial} C(z+th') \\ h' \to h, t \downarrow 0}} \{Vh'\} \tag{2.3.24}$$

exists for all $h \in \mathbb{R}^m$. This implies in particular that C is directionally differentiable on $\mathbb{R}^m$. Moreover, by virtue of the Lipschitz continuity, if $z \in \mathbb{R}^m$ is fixed and $h \in \mathbb{R}^m$ is variable, then one has

$$C(z+h) - C(z) - C'(z;h) = o(\|h\|), \tag{2.3.25}$$

i.e., the positively homogeneous map $C'(z;\cdot)$ has the same approximation property as the Fréchet derivative; cf. Shapiro (1990). We speak of *Bouligand differentiability* and make use of (2.3.25) in the main convergence statement (Theorem 2.6. below). From semismoothness it further follows that for any $V \in \bar{\partial} C(z+h)$

$$Vh - C'(z;h) = o(\|h\|). \tag{2.3.26}$$

Besides semismoothness, we will require from C still another property.

Definition 2.1. We say that C is *strongly BD-regular* at z if all matrices $V \in \partial_B C(z)$ are nonsingular.

Lemma 2.5. Assume that C is strongly BD-regular at z. Then there exist a neighbourhood $\mathcal{U}$ of z and a constant $c \geq 0$ such that all matrices $V \in \partial_B C(v)$ with $v \in \mathcal{U}$ are nonsingular and

$$\|V^{-1}\| \leq c. \tag{2.3.27}$$

Proof. First we show the existence of a neighbourhood $\mathcal{U}$ and a constant c such that the Jacobians $\nabla C(v)$ are nonsingular for all $v \in \mathcal{U} \setminus \omega_C$ (the set where C is not differentiable) and

$$\|\nabla C(v)\|^{-1} \leq c. \tag{2.3.28}$$

If this claim is not true then there is a sequence $z_k \to z$, $z_k \notin \omega_C$ such that either all Jacobians $\nabla C(z_k)$ are singular or $\|(\nabla C(z_k))^{-1}\| \to \infty$. Since C is Lipschitz near z, the Jacobians $\nabla C(z_k)$ are bounded in a neighbourhood of z. Hence, there is a subsequence of $\{z_k\}$, say $\{z_{k'}\}$, such that $\{\nabla C(z_{k'})\}$ converges to a matrix V. This V must be singular. By the definition, $V \in \partial_B C(z)$ and this contradicts the strong BD-regularity of C at z. Hence, inequality (2.3.27) holds for all $v \in \mathcal{U} \setminus \omega_C$. At the points $v \in \mathcal{U} \cap \omega_C$ it suffices to observe that each $V \in \partial_B C(v)$ is the limit of a sequence $\nabla C(z_i)$ with $z_i \to v$ and $z_i \in \mathcal{U} \setminus \omega_C$ for all i. □

The Newton iteration based on generalized Jacobians is defined by

$$z_{k+1} = z_k - V_k^{-1} C(z_k), \quad k = 0, 1, \ldots \tag{2.3.29}$$

where $V_k \in \partial C(z_k)$. In Qi and Sun (1993) a variant is studied, where the V_k in (2.3.29) is taken from $\partial_B C(z_k)$. This allows to weaken the assumptions ensuring local superlinear convergence. Recall that a sequence $\{x_k\}$ is said to converge *superlinearly* to x^*, if

$$\|x_{k+1} - x^*\| = o\left(\|x_k - x^*\|\right).$$

The resulting algorithm can be written in the form:

Nonsmooth Newton's Method: *Choose a starting point* (2.3.30)
$z_0 \in \mathbb{R}^m, \varepsilon > 0$ *and put* $k := 0$.

(1) *If* $\|C(z_k)\| \leq \varepsilon$, *then STOP.*

(2) *Compute an arbitrary matrix* $V_k \in \partial_B C(z_k)$ *and put*

$$z_{k+1} = z_k - V_k^{-1} C(z_k).$$

(3) *Put* $k := k + 1$ *and go back to (1).*

The next theorem gives a convergence result for the above method in the case $\varepsilon = 0$.

Theorem 2.6. Suppose that z^* is a solution of (2.3.23) and C is strongly BD-regular at z^*. Then the Newton's method (2.3.30) is well-defined and converges locally superlinearly to z^*.

Proof. Let $\mathcal{U}$ be a neighborhood of z^* specified by Lemma 2.5. (with $z = z^*$) and assume that $z_k \in \mathcal{U}$. Then, by Lemma 2.5, the kth Newton iteration in step (2) is well defined, and

$$\begin{aligned}
\|z_{k+1} - z^*\| &= \|z_k - z^* - V_k^{-1} C(z_k)\| \\
&= \left\|V_k^{-1} \left[V_k(z_k - z^*) - C(z_k) + C(z^*) + C'(z^*; z_k - z^*) - C'(z^*; z_k - z^*)\right]\right\| \\
&\leq \left\|V_k^{-1} \left[C(z_k) - C(z^*) - C'(z^*; z_k - z^*)\right]\right\| \\
&\qquad\qquad + \left\|V_k^{-1} \left[V_k(z_k - z^*) - C'(z^*; z_k - z^*)\right]\right\| \\
&\leq \|V_k^{-1}\| \left[\|C(z_k) - C(z^*) - C'(z^*; z_k - z^*)\| + \|V_k(z_k - z^*) - C'(z^*; z_k - z^*)\|\right].
\end{aligned}$$

By (2.3.25)

$$C(z_k) - C(z^*) - C'(z^*; z_k - z^*) = o(\|z_k - z^*\|)$$

and, due to (2.3.26),

$$V_k(z_k - z^*) - C'(z^*; z_k - z^*) = o(\|z_k - z^*\|).$$

Consequently, there is a function $\varphi(t) = o(t)$ (not dependent on k) such that

$$\|z_{k+1} - z^*\| \leq \varphi(\|z_k - z^*\|).$$

This implies the superlinear convergence of $\{z_k\}$ to z^*, provided z_0 is sufficiently close to z^* and all iterates z_k stay in $\mathcal{U}$. To this purpose, however, it suffices to start in such a ball $\mathcal{B} := z^* + \rho\mathbb{B}$, which satisfies the following conditions:

(i) $\mathcal{B} \subset \mathcal{U}$;

(ii) for all $\nu \in [0, \rho]$ one has $\varphi(\nu) \leq k\nu$ with a $k \in (0, 1)$.

The statement has been proved. □

Under additional assumptions on C, even local quadratic convergence can be proved for (2.3.30). Although we had no difficulties with the convergence of this method in our mechanical problems, the method may not converge at all, similarly as the classic Newton's method in the smooth case. To achieve global convergence, various modifications of the presented method (and also other nonsmooth Newton variants) are available. These questions go, however, beyond the scope of these lecture notes.

3 Lipschitzian properties of solution maps

In this type of analysis it is very convenient to work with the GE model, introduced in the previous section. The model has to be enriched, however, by a parameter $x \in \mathbb{R}^n$ which may generally enter both the single-valued as well as the set-valued part of the respective GE. Later this parameter will take over the role of a control variable. We confine ourselves here only to equilibria, where this parameter arises merely in the single-valued part. So, we will investigate the *solution map* $S[\mathbb{R}^n \rightsquigarrow \mathbb{R}^m]$ defined by

$$S(x) := \{y \in \vartheta | 0 \in F(x, y) + Q(y)\}, \tag{3.0.1}$$

where the closed set $\vartheta \subset \mathbb{R}^m$ specifies the "nonequilibrial" constraints imposed on y. Throughout this chapter $F[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m]$ is continuously differentiable and $Q[\mathbb{R}^m \rightsquigarrow \mathbb{R}^m]$ has a closed graph. Let $(\bar{x}, \bar{y}) \in \mathrm{Gph}S$ be a *reference point.* In Section 3.1 we will derive conditions ensuring the Aubin property of S around $(\bar{x}, \bar{y})$, whereas Section 3.2 is devoted to the existence of a Lipschitz single-valued localization of S at $(\bar{x}, \bar{y})$. The former property is important above all in stability considerations. Indeed, if S possesses the Aubin property around $(\bar{x}, \bar{y})$ and x is close to $\bar{x}$, then there is an equilibrium $y \in S(x)$ and $\|y - \bar{y}\| \leq L\|x - \bar{x}\|$ for some $L \geq 0$. The Lipschitz single-valued localizations play an important role above all in ImP.

3.1 Aubin property

Clearly, $S(x) = \{y \in \vartheta | g(x, y) \in \Lambda\}$, where

$$g(x, y) = \begin{bmatrix} y \\ -F(x, y) \end{bmatrix} \qquad \text{and} \qquad \Lambda = \mathrm{Gph}Q. \tag{3.1.2}$$

Therefore we can proceed in a similar way as in the case of the map (1.3.15), but this time our analysis will be finer. We say that a multifunction $\Gamma[\mathbb{R}^l \rightsquigarrow \mathbb{R}^p]$ is *calm* at $(a, b) \in \mathrm{Gph}\Gamma$ (with modulus L), provided there exist neighborhoods $\mathcal{U}$ of a, $\mathcal{V}$ of b and a real $L \geq 0$ such that

$$\Gamma(a') \cap \mathcal{V} \subset \Gamma(a) + L\|a' - a\|\mathbb{B} \ \forall\ a' \in \mathcal{U}.$$

By comparing this property with Definition 1.5 we immediately conclude that calmness is a weaker property than both the (local) upper Lipschitz continuity at a as well as the Aubin property around (a, b). To be able to apply the Mordukhovich criterion to S, we have first to compute an upper approximation of $D^*S(\bar{x}, \bar{y})(0)$. To this purpose we introduce an auxiliary map $P[\mathbb{R}^m \rightsquigarrow \mathbb{R}^n \times \mathbb{R}^m]$ defined by

$$P(\xi) := \{(x, y) \in \mathbb{R}^n \times \vartheta | \xi + \begin{bmatrix} y \\ -F(x, y) \end{bmatrix} \in \mathrm{Gph}Q\}. \tag{3.1.3}$$

Clearly, $\mathrm{Gph}S = P(0)$.

Lemma 3.1. Let P be calm at $(0, \bar{x}, \bar{y})$. Then one has for all $y^* \in \mathbb{R}^m$ the inclusion

$$D^*S(\bar{x}, \bar{y})(y^*) \subset \left\{ x^* \in \mathbb{R}^n \,\middle|\, \begin{bmatrix} x^* \\ -y^* \end{bmatrix} \in \begin{bmatrix} 0 & -(\nabla_x F(\bar{x}, \bar{y}))^T \\ E & -(\nabla_y F(\bar{x}, \bar{y}))^T \end{bmatrix} N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y})) + \begin{bmatrix} 0 \\ N_\vartheta(\bar{y}) \end{bmatrix} \right\}.$$

The Mordukhovich criterion leads now directly to the following statement.

Theorem 3.2. Let P be calm at $(0, \bar{x}, \bar{y})$ and the qualification condition

$$\begin{bmatrix} x^* \\ 0 \end{bmatrix} \in \begin{bmatrix} 0 & -(\nabla_x F(\bar{x}, \bar{y}))^T \\ E & -(\nabla_y F(\bar{x}, \bar{y}))^T \end{bmatrix} N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y})) + \begin{bmatrix} 0 \\ N_\vartheta(\bar{y}) \end{bmatrix} \Rightarrow x^* = 0 \tag{3.1.4}$$

is fulfilled. Then S has the Aubin property around $(\bar{x}, \bar{y})$.

Unfortunately, it is not easy to ensure the calmness of P at $(0, \bar{x}, \bar{y})$ except of the case when F is affine, ϑ is a convex polyhedron and Q is polyhedral. Then, namely, one can employ Theorem 1.2 which implies that P is even (locally) upper Lipschitz at 0. A standard way consists in the application of the Mordukhovich criterion also to P along the lines of Theorem 1.3. In this way we arrive at the following statement.

Theorem 3.3. Assume that the qualification condition

$$\left.\begin{array}{c} 0 \in u - (\nabla_y F(\bar{x}, \bar{y}))^T v + N_\vartheta(\bar{y}) \\ (u, v) \in N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y})) \end{array}\right\} \Rightarrow u = 0, v = 0 \tag{3.1.5}$$

is fulfilled. Then S has the Aubin property around $(\bar{x}, \bar{y})$.

If $\vartheta = \mathbb{R}^m$, condition (3.1.5) reduces to the form

$$\left.\begin{array}{c} 0 = u - (\nabla_y F(\bar{x}, \bar{y}))^T v \\ (u, v) \in N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y})) \end{array}\right\} \Rightarrow v = 0.$$

Proof. P has the same structure as the multifunction Σ in (1.3.15). Consequently, by virtue of Theorem 1.3, P has the Aubin property around $(0, \bar{x}, \bar{y})$, whenever the implication

$$\left.\begin{array}{c} 0 \in \begin{bmatrix} 0 & -(\nabla_x F(\bar{x}, \bar{y}))^T \\ E & -(\nabla_y F(\bar{x}, \bar{y}))^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} 0 \\ N_\vartheta(\bar{y}) \end{bmatrix} \\ (u, v) \in N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y})) \end{array}\right\} \Rightarrow u = 0, v = 0 \tag{3.1.6}$$

holds true. Thus, condition (3.1.5) ensures both condition (3.1.6) (and hence the calmness of P at $(0, \bar{x}, \bar{y})$) as well as condition (3.1.4). The result follows from Theorem 3.2. The simplification of (3.1.5) in the case $\vartheta = \mathbb{R}^m$ is evident. $\square$

The condition (3.1.5) is, however, sometimes too restrictive and excludes many situations, where S does have the Aubin property. Therefore it is important to develop specific calmness criteria, weaker than condition (3.1.6), cf. Henrion et al. (2002), Henrion and Outrata (to appear). Conditions (3.1.4),(3.1.5) are helpful only in the case, when we are able to express the limiting normal cone $N_{\mathrm{Gph}Q}(\bar{y}, -F(\bar{x}, \bar{y}))$ in terms of the original data, associated with the considered equilibrium. This can be a difficult task as shown in Dontchev and Rockafellar (1996) and Mordukhovich and Outrata (2001). To illustrate the procedure, consider the case with $Q(y) = N_{\mathbb{R}^m_+}(y)$ corresponding to a complementarity problem. Then we can exploit a simple result from Outrata (1999) (Lemma 3.4 below). In its formulation and also in numerous other results, presented in the sequel, we will associate with each pair $(y, z) \in \mathrm{Gph}N_{\mathbb{R}^m_+}$ the index sets

$$\begin{array}{ll} L(y, z) := & \{i \in \{1, 2, \ldots, m\} | y^i > 0\} \\ I_+(y, z) := & \{i \in \{1, 2, \ldots, m\} | y^i = 0, z^i < 0\} \\ I_0(y, z) := & \{i \in \{1, 2, \ldots, m\} | y^i = 0, z^i = 0\}. \end{array}$$

They are related to the constraint $y \geq 0$, and therefore we speak about the index set of *inactive, strongly* and *weakly active* inequalities, respectively. The set $I_0(y,z)$ is also sometimes called the *degenerate* index set. The arguments at L, I_+, I_0 are omitted whenever these sets arise as subscripts or whenever it cannot cause any confusion.

Lemma 3.4. Consider a pair $(a,b) \in \mathrm{Gph} N_{\mathbb{R}^m_+}$ Then one has

$$N_{\mathrm{Gph} N_{\mathbb{R}^m_+}}(a,b) = \mathop{\mathsf{X}}_{i=1}^{m} N_{\mathrm{Gph} N_{\mathbb{R}_+}}(a^i, b^i),$$

where

$$N_{\mathrm{Gph} N_{\mathbb{R}_+}}(a^i, b^i) = \begin{cases} 0 \times \mathbb{R} \text{ for } i \in L(a,b) \\ \mathbb{R} \times 0 \text{ for } i \in I_+(a,b) \\ (0 \times \mathbb{R}) \cup (\mathbb{R} \times 0) \cup (\mathbb{R}_- \times \mathbb{R}_+) \text{ for } i \in I_0(a,b). \end{cases}$$

If $\vartheta = \mathbb{R}^m$, condition (3.1.5) attains in this situation the form

$$\left.\begin{array}{l} 0 = u - ((\nabla_y F(\bar{x},\bar{y}))_{L\cup I_0})^T v_{L\cup I_0} \\ u_L = 0 \\ \text{for } i \in I_0 \text{ either } u^i v^i = 0 \\ \qquad\qquad \text{or } u^i < 0 \text{ and } v^i > 0 \end{array}\right\} \Rightarrow v_{L\cup I_0} = 0. \tag{3.1.7}$$

In the above relations all index sets L, I_+, I_0 have the argument $(\bar{y}, -F(\bar{x},\bar{y}))$. Of course, also the verification of condition (3.1.7) is not easy; one has to employ the structural properties of the partial Jacobian $\nabla_y F(\bar{x},\bar{y})$.

3.2 Lipschitz single-valued localization

We immediately infer that S possesses a Lipschitz single-valued localization at $(\bar{x},\bar{y})$, whenever it has the Aubin property around $(\bar{x},\bar{y})$ and for each x close to $\bar{x}$ the GE

$$0 \in F(x,y) + Q(y) \tag{3.2.8}$$

possesses a unique solution in $\vartheta \cap \mathcal{O}$, where $\mathcal{O}$ is a neighborhood of $\bar{y}$. Another way is to employ a powerful result from the landmark paper (Robinson (1980)). In this context, however, we have to assume that $\vartheta = \mathbb{R}^m$.

Definition 3.1. We say that the GE (3.2.8) satisfies the *Strong Regularity Condition* (SRC) at $(\bar{x},\bar{y})$ provided the multifunction $\sum[\mathbb{R}^m \rightsquigarrow \mathbb{R}^m]$, defined by

$$\sum(\xi) = \{y \in \mathbb{R}^m | \xi \in F(\bar{x},\bar{y}) + \nabla_y F(\bar{x},\bar{y})(y-\bar{y}) + Q(y)\}$$

possesses a Lipschitz single-valued localization at $(0,\bar{y})$.

Theorem 3.5. Assume that the GE (3.2.8) satisfies (SRC) at $(\bar{x}, \bar{y})$. Then the corresponding solution map S possesses a Lipschitz single-valued localization at $(\bar{x}, \bar{y})$.

In this way we have converted the analysis of S to the analysis of a simpler map $\sum$, where the single-valued part of the underlying GE is affine. Thus, $\sum$ is (locally) upper Lipschitz at 0 whenever Q is polyhedral and then it suffices to show that the GE

$$\xi \in F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y})(y - \bar{y}) + Q(y) \tag{3.2.9}$$

possesses a unique solution y for all ξ from a neighborhood of 0. If $Q(y) = N_\Omega(y)$ with Ω being a convex polyhedron, the GE (3.2.9) can be significantly simplified. We recall that the cone

$$K(y, z) := T_\Omega(y) \cap z^\perp$$

is termed the *critical* cone of Ω with respect to y and z.

Theorem 3.6. Let Ω be a convex polyhedron. Then the GE

$$\xi \in F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y})(y - \bar{y}) + N_\Omega(y) \tag{3.2.10}$$

possesses a unique solution y for all ξ from a neighborhood of 0 iff the GE

$$\xi \in \nabla_y F(\bar{x}, \bar{y})h + N_{K(\bar{y}, F(\bar{x}, \bar{y}))}(h) \tag{3.2.11}$$

possesses a unique solution h for all $\xi \in \mathbb{R}^m$.

Proof. (3.2.10) can clearly be rewritten to the form

$$y = \mathrm{Proj}_\Omega(\xi - F(\bar{x}, \bar{y}) - \nabla_y F(\bar{x}, \bar{y})(y - \bar{y}) + y) \tag{3.2.12}$$

If we choose y sufficiently close to $\bar{y}$ and ξ sufficiently close to 0, the norm of the vector

$$\eta := \xi - \nabla_y F(\bar{x}, \bar{y})(y - \bar{y}) + y - \bar{y}$$

can be made arbitrarily small. Since

$$\bar{y} = \mathrm{Proj}_\Omega(\bar{y} - F(\bar{x}, \bar{y})),$$

we infer that

$$\mathrm{Proj}_\Omega(\xi - F(\bar{x}, \bar{y}) - \nabla_y F(\bar{x}, \bar{y})(y - \bar{y}) + y) = \mathrm{Proj}_\Omega(\bar{y} - F(\bar{x}, \bar{y}) + \eta) = \bar{y} + \mathrm{Proj}_{K(\bar{y}, F(\bar{x}, \bar{y}))}(\eta),$$

where the last equality follows from the directional differentiability of the projection onto convex polyhedra (Robinson (1991)). We conclude that with $h := y - \bar{y}$ equation (3.2.12) amounts to

$$h = \mathrm{Proj}_{K(\bar{y}, F(\bar{x}, \bar{y}))}(\xi - \nabla_y F(\bar{x}, \bar{y})h + h)$$

which is exactly the GE (3.2.11). It remains to observe that, since (3.2.11) is positively homogeneous in h, the restriction of ξ to a neighborhood of 0 has no importance. □

In the considered case the problem of the Lipschitz single-valued localization has thus been reduced to the question of existence and uniqueness of solutions to (3.2.11). The application of the results from Section 2 leads now a number of useful results. We confine ourselves, however, only to the following two situations playing an important role in the sequel.

Theorem 3.7. Consider the NCP governed by the GE

$$0 \in F(x, y) + N_{\mathbb{R}^m_+}(y). \tag{3.2.13}$$

This GE fulfills the (SRC) at the reference point $(\bar{x}, \bar{y})$ iff the matrix $(\nabla_y F(\bar{x}, \bar{y}))_{L,L}$is nonsingular and its Schur complement in the matrix

$$\begin{bmatrix} (\nabla_y F(\bar{x}, \bar{y}))_{L,L} & (\nabla_y F(\bar{x}, \bar{y}))_{L,I_0} \\ (\nabla_y F(\bar{x}, \bar{y}))_{I_0,L} & (\nabla_y F(\bar{x}, \bar{y}))_{I_0,I_0} \end{bmatrix}$$

is a P-matrix. (Again, the index sets L, I_0 have the argument $(\bar{y}, -F(\bar{x}, \bar{y}))$).

Proof. We have to show that the above condition is equivalent to the unique solvability of the GE

$$\xi \in \nabla_y F(\bar{x}, \bar{y})h + N_K(h)$$

for each $\xi \in \mathbb{R}^m$; thereby

$$\begin{aligned} K = T_{\mathbb{R}^m_+}(\bar{y}) \cap (F(\bar{x}, \bar{y}))^\perp &= \{h \in \mathbb{R}^m | h_{I_+} \geq 0,\, h_{I_0} \geq 0\} \cap \{h \in \mathbb{R}^m | h_{I_+} = 0\} \\ &= \{h \in \mathbb{R}^m | h_{I_+} = 0,\, h_{I_0} \geq 0\}. \end{aligned}$$

Consequently,

$$N_K(h) = \{\eta \in \mathbb{R}^m | \eta_L = 0,\ \ \eta_{I_0} \leq 0,\, \langle \eta_{I_0}, h_{I_0} \rangle = 0\},$$

and the respective GE (3.2.11) reduces to the form

$$\begin{aligned} \xi_L &= (\nabla_y F(\bar{x}, \bar{y}))_{L,L} h_L + (\nabla_y F(\bar{x}, \bar{y}))_{L,I_0} h_{I_0} \\ \xi_{I_0} &= (\nabla_y F(\bar{x}, \bar{y}))_{I_0,L} h_L + (\nabla_y F(\bar{x}, \bar{y}))_{I_0,I_0} h_{I_0} + N_{\mathbb{R}^a_+}(h_{I_0}), \end{aligned}$$

where $a = |I_0|$. The result follows now from the characterization of the unique solvability of a linear CP, mentioned in the previous section, cf. Robinson (1980). $\square$

Next we turn our attention to the case, where Ω is given by (2.2.21) and apply the technique leading to the GE (2.2.22). This technique enables even to generalize the original model and allows that x enters also the functions H, G defining Ω. So,we start with the GE

$$0 \in F(x, y) + N_{\Omega(x)}(y) \tag{3.2.14}$$

where the multifunction $\Omega[\mathbb{R}^n \rightsquigarrow \mathbb{R}^m]$ is defined by

$$\Omega(x) := \{y \in \mathbb{R}^m | H(x, y) = 0, G(x, y) \leq 0\},$$

and assume that $H[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l], G[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^s]$ are continuously differentiable and for all $x \in \mathbb{R}^n, H(x, \cdot)$ is affine and all components $G^i(x, \cdot)$ are convex. Moreover, for the reference parameter value $\bar{x}$ we modify the (SCQ) as follows:

(ESCQ): $\exists\ \bar{y}_{\bar{x}} \in \mathbb{R}^m$ such that $H(\bar{x}, \bar{y}_x) = 0$ and $G(\bar{x}, \bar{y}_x) \in \operatorname{int} \mathbb{R}^s_+$.

In this way, for x close to $\bar{x}$, the GE (3.2.14) admits the KKT form

$$\left.\begin{array}{rcl} 0 & = & \mathcal{L}(x,y,\mu,\lambda) \\ 0 & = & H(x,y) \\ 0 & \in & -G(x,y)+N_{\mathbb{R}^s_+}(\lambda), \end{array}\right\} \tag{3.2.15}$$

where $\mathcal{L}(x,y,\mu,\lambda) := F(x,y)+(\nabla_y H(x,y))^T\mu+(\nabla_y G(x,y))^T\lambda$. Note that in (3.2.15) the parameter x does not enter the multi-valued part but, in contrast to (3.2.14), we have to do with additional variables μ,λ. As in case of NCPs we will need the index sets

$$\begin{array}{ll} L(x,y,\lambda) := & \{i \in \{1,2,\ldots,s\} | \lambda^i > 0\} \\ I_+(x,y,\lambda) := & \{i \in \{1,2,\ldots,s\} | \lambda^i = 0, G^i(x,y) < 0\} \\ I_0(x,y,\lambda) := & \{i \in \{1,2,\ldots,s\} | \lambda^i = 0, G^i(x,y) = 0\}, \end{array}$$

related to the constraint $\lambda \geq 0$. Again, their arguments are omitted whenever these sets arise as subscripts or whenever it cannot cause any confusion. To be able to present the next result, we also have to associate with the reference point $(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$ the matrix

$$D(\bar{x},\bar{y},\bar{\mu},\bar{\lambda}) := \left[\begin{array}{lcc} \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda}) & (\nabla_y H(\bar{x},\bar{y}))^T & (\nabla_y G(\bar{x},\bar{y}))^T \\ \nabla_y H(\bar{x},\bar{y}) & 0 & 0 \\ -\nabla_y G(\bar{x},\bar{y}) & 0 & 0 \end{array}\right],$$

and, for an arbitrary subset N of $I_0(\bar{x},\bar{y},\bar{\lambda})$, the matrix

$$D_{(N)}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda}) := \left[\begin{array}{lcc} \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda}) & (\nabla_y H(\bar{x},\bar{y}))^T & ((\nabla_y G(\bar{x},\bar{y}))_N)^T \\ \nabla_y H(\bar{x},\bar{y}) & 0 & 0 \\ (\nabla_y G(\bar{x},\bar{y}))_N & 0 & 0 \end{array}\right].$$

Theorem 3.8. The GE (3.2.15) fulfills the (SRC) at $(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$ iff the matrix $D_{(L)}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$ is nonsingular and its complement in $D_{(L\cup I_0)}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$ is a P-matrix.

This statement can be proved along the same lines as Theorem 3.7., cf. eg (Outrata et al., 1998, Theorem 5.7). It remains to ensure the above requirements by means of reasonable assumptions imposed on the problem data.

Theorem 3.9. Suppose that the constraint qualification

(ELICQ): The gradients $\nabla_y H^i(\bar{x},\bar{y}), i=1,2,\ldots,l$, and
$\nabla_y G^i(\bar{x},\bar{y}), i \in L(\bar{x},\bar{y}) \cup I_0(\bar{x},\bar{y},\bar{\lambda})$, are linearly independent;

is fulfilled. Further assume that $\nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$ is strictly copositive on

$$\ker \nabla_y H(\bar{x},\bar{y}) \cap \ \ker((\nabla_y G(\bar{x},\bar{y}))_L).$$

Then the requirement of Theorem 3.8 is satisfied and, consequently, the GE (3.2.15) fulfills the (SRC) at $(\bar{x},\bar{y},\bar{\mu},\bar{\lambda})$.

The proof follows directly from Theorem 2.3. If the GE (3.2.15) represents first-order optimality conditions of an optimization problem, then the second condition of Theorem 3.9 is just the *strong second-order sufficient condition*, well-known from mathematical programming. Therefore it is sometimes also referred to as the (SSOSC) at $(\bar{x}, \bar{y}, \bar{\mu}, \bar{\lambda})$.

If σ is a Lipschitz single-valued localization of S at $(\bar{x}, \bar{y})$, one could describe its local behaviour by $\bar{\partial}\sigma(\bar{x})$ or an upper approximation of this set. The computation of an upper approximation of $\bar{\partial}\sigma(\bar{x})$ is particularly simple, provided σ is a PC^1-function. We explain this technique again by means of an NCP and put $\vartheta = \mathbb{R}^m$ for the sake of simplicity. From a continuity argument it is clear that in a neighborhood of the reference point one definitely has

$$\begin{array}{lll} F^i(x,y) \geq 0 & y^i = 0 & \text{for } i \in I_+ \cup M_1 \\ F^i(x,y) = 0 & y^i \geq 0 & \text{for } i \in L \cup M_2 \end{array} \tag{3.2.16}$$

for some partition(s) M_1, M_2 of $I_0(\bar{y}, -F(\bar{x}, \bar{y}))$. Ignoring the inequalities in (3.2.16), we arrive at a system of nonlinear equations

$$\begin{array}{rcll} F^i(x,y) & = & 0 & \text{for } i \in L \cup M_2 \\ y^i & = & 0 & \text{for } i \in I_+ \cup M_1 \end{array} \tag{3.2.17}$$

to which the classical implicit function theorem can be applied whenever the matrix (after a suitable reorganization)

$$\begin{bmatrix} (\nabla_y F(\bar{x}, \bar{y}))_{L \cup M_2} \\ 0 \qquad E_{I_+ \cup M_1} \end{bmatrix} \tag{3.2.18}$$

is nonsingular. If matrices (3.2.18) are nonsingular for all possible partitions M_1, M_2, we immediately infer that σ is a PC^1-function. Indeed, let us introduce the index set $\mathbb{K}(\bar{x}, \bar{y})$ of all possible partitions of I_0 so that to each pair M_1, M_2 we assign an index $j \in \mathbb{K}(\bar{x}, \bar{y})$. Further, denote by $\sigma_j, j \in \mathbb{K}(\bar{x}, \bar{y})$, the implicit function specified by the corresponding system (3.2.17). Clearly, $\bar{y} = \sigma_j(\bar{x})$ for all $j \in \mathbb{K}(\bar{x}, \bar{y})$, the single functions σ_j are continuously differentiable and there exists a neighborhoods $\mathcal{U}$ of $\bar{x}$ such that $\sigma(x) = \sigma_j(x)$ for some $j \in \mathbb{K}(\bar{x}, \bar{y})$ and all $x \in \mathcal{U}$. Hence, σ is a PC^1-function on $\mathcal{U}$. On the basis of (1.2.14) we are now able to compute an upper approximation of $\bar{\partial}\sigma(\bar{x})$ in the form

$$\bar{\partial}\sigma(\bar{x}) \subset \operatorname{conv}\{\nabla\sigma_j(\bar{x}) | j \in \mathbb{K}(\bar{x}, \bar{y})\}. \tag{3.2.19}$$

This so-called PC^1 approach to the computing of $\bar{\partial}\sigma(\bar{x})$ is exploited above all in cases, where the existence of a Lipschitz single-valued localization is ensured by means of the strong regularity condition. Then we benefit from the fact that (SRC) implies the regularity for all systems (3.2.17). We return to this procedure in Section 5, devoted to ImP.

4 Optimality conditions

Consider the MPEC

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \\ & 0 \in F(x,y) + Q(y) \\ & (x,y) \in \kappa, \end{array} \tag{4.0.1}$$

where $f[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is an objective function, the equilibrium constraint is modeled by a GE studied in the previous section and $\kappa \subset \mathbb{R}^n \times \mathbb{R}^m$ is a closed set which comprises all other constraints imposed on x and y. In particular, κ can be the Carthesian product $\omega \times \vartheta$, where $\omega \subset \mathbb{R}^n$ is the set of *feasible controls* and $\vartheta \subset \mathbb{R}^m$ is the set of *feasible state variables.* To derive optimality conditions, we can put $z = (x, y)$ and rewrite the constraints in (4.0.1) to the form

$$\{z \in \kappa | g(z) \in \Lambda\}$$

where g and Λ are defined in (3.1.2). This enables a direct application of Theorems 1.4, 1.5 which will be done in Section 4.1. If the GE in (4.0.1) admits a Lipschitz single-valued localization at a (locally) optimal pair $(\widehat{x}, \widehat{y})$ and the nonequilibrium constraint κ admits the product form $\omega \times \vartheta$, we can make use of Theorem 1.6 and arrive at optimality conditions of a similar type. This technique, however, combined with the PC^1 approach from the previous section, leads to different optimality conditions, closely related with ImP. This is the topic of Section 4.2. All optimality conditions will be illustrated by means of MPCC, i.e. the problem (4.0.1), where the GE models an NCP.

4.1 Optimality conditions via mathematical programming

Throughout the sequel $(\widehat{x}, \widehat{y})$ is a (locally) optimal pair in (4.0.1). Moreover, we will assume that the objective f is Lipschitz near $(\widehat{x}, \widehat{y})$.

Theorem 4.1. Let the qualification condition

$$\left.\begin{array}{c} \begin{bmatrix} 0 & -(\nabla_x F(\widehat{x}, \widehat{y}))^T \\ E & -(\nabla_y F(\widehat{x}, \widehat{y}))^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \in -N_\kappa(\widehat{x}, \widehat{y}) \\ (u, v) \in N_{\mathrm{Gph}Q}(\widehat{y}, -F(\widehat{x}, \widehat{y})) \end{array}\right\} \Rightarrow \left\{ \begin{array}{lcl} u & = & 0 \\ v & = & 0 \end{array}\right. \tag{4.1.2}$$

be fulfilled. Then there exists a pair $(\widehat{\xi}, \widehat{\eta}) \in \partial f(\widehat{x}, \widehat{y})$, a KKT vector $\widehat{v} \in \mathbb{R}^m$ and a pair $(\widehat{w}, \widehat{z}) \in N_\kappa(\widehat{x}, \widehat{y})$ such that

$$\begin{array}{l} 0 = \widehat{\xi} \;\; + \;\; (\nabla_x F(\widehat{x}, \widehat{y}))^T \widehat{v} \;\; + \;\; \widehat{w} \\ 0 = \widehat{\eta} \;\; + \;\; (\nabla_y F(\widehat{x}, \widehat{y}))^T \widehat{v} \;\; + \;\; D^* Q(\widehat{y}, -F(\widehat{x}, \widehat{y}))(\widehat{v}) \; + \; \widehat{z}. \end{array} \tag{4.1.3}$$

Proof. (4.1.2) is exactly condition (1.1.3) for the constraint structure being considered. It enables to express an upper approximation of the limiting normal cone to the whole constraint system in (4.0.1) in the form

$$\begin{bmatrix} 0 & -(\nabla_x F(\widehat{x}, \widehat{y}))^T \\ E & -(\nabla_y F(\widehat{x}, \widehat{y}))^T \end{bmatrix} N_{\mathrm{Gph}Q}(\widehat{y}, -F(\widehat{x}, \widehat{y})) + N_\kappa(\widehat{x}, \widehat{y}) =$$
$$\left\{ \begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))^T v + w \\ (\nabla_y F(\widehat{x}, \widehat{y}))^T v + D^* Q(\widehat{y}, -F(\widehat{x}, \widehat{y}))(v) + z \end{bmatrix} \middle| v \in \mathbb{R}^m, (w, z) \in N_\kappa(\widehat{x}, \widehat{y}) \right\}.$$

It remains thus to apply relation (1.4.17). $\square$

Instead of the qualification condition (4.1.2) it suffices to require that the map $\widetilde{P}[\mathbb{R}^m \rightsquigarrow \mathbb{R}^n \times \mathbb{R}^m]$, defined by

$$\widetilde{P}(\xi) := \left\{ (x,y) \in \kappa \middle| \xi + \begin{bmatrix} y \\ -F(x,y) \end{bmatrix} \in \mathrm{Gph} Q \right\},$$

is calm at $(0, \widehat{x}, \widehat{y})$, cf. Henrion et al. (2002). From this it follows that the qualification condition (4.1.2) can be omitted, whenever F is affine (in both variables), κ is a convex polyhedron and Q is polyhedral. It is also important to point out that the qualification condition (4.1.2) is fulfilled whenever $\kappa = \omega \times \mathbb{R}^m$ and the GE in (4.0.1) satisfies (SRC) at $(\widehat{x}, \widehat{y})$, cf. Outrata (2000). Points $(\widehat{x}, \widehat{y})$ satisfying (4.1.3) are called *M(ordukhovich)-stationary.* Similarly as in the statement of Theorem 3.3, relations (4.1.2),(4.1.3) have to be expressed in terms of the problem data. We show next the form of these relations provided $Q(y) = N_{\mathbb{R}^m_+}(y)$, i.e., the considered equilibrium is an NCP.

Theorem 4.2. Suppose that the qualification condition

$$\left. \begin{array}{l} \begin{bmatrix} 0 & -((\nabla_x F(\widehat{x},\widehat{y}))_{L\cup I_0})^T \\ E & -((\nabla_y F(\widehat{x},\widehat{y}))_{L\cup I_0})^T \end{bmatrix} \begin{bmatrix} u \\ v_{L\cup I_0} \end{bmatrix} \in -N_\kappa(\widehat{x},\widehat{y}) \\ u_L = 0 \text{ and for } i \in I_0 \text{ either } u^i v^i = 0 \\ \text{or } u^i < 0 \text{ and } v^i > 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{lcl} u & = & 0 \\ v_{L\cup I_0} & = & 0 \end{array} \right. \tag{4.1.4}$$

holds true. Then there exists a pair $(\widehat{\xi}, \widehat{\eta}) \in \partial f(\widehat{x}, \widehat{y})$, a pair of KKT vectors $(\widehat{u}, \widehat{v}) \in \mathbb{R}^m \times \mathbb{R}^m$ and a pair $(\widehat{w}, \widehat{z}) \in N_\kappa(\widehat{x}, \widehat{y})$ such that $\widehat{u}_L = 0, \widehat{v}_{I_+} = 0$,

$$\begin{array}{lcl} 0 & = & \widehat{\xi} - ((\nabla_x F(\widehat{x},\widehat{y}))_{L\cup I_0})^T \widehat{v}_{L\cup I_0} + \widehat{w} \\ 0 & = & \widehat{\eta}_{L\cup I_0} + \widehat{u}_{L\cup I_0} - ((\nabla_y F(\widehat{x},\widehat{y}))_{L\cup I_0, L\cup I_0})^T \widehat{v}_{L\cup I_0} + \widehat{z}_{L\cup I_0}, \end{array} \tag{4.1.5}$$

and for $i \in I_0$ one has either $\widehat{u}^i \widehat{v}^i = 0$ or $\widehat{u}^i < 0$ and $\widehat{v}^i > 0$.

The index sets L, I_+, I_0 in the above statement are related to the argument $(\widehat{y}, -F(\widehat{x}, \widehat{y}))$. Its proof follows immediately from Theorem 4.1 and Lemma 3.4. Condition (4.1.4) is generally not easy to verify. In Outrata (1999) this condition has been suitably reformulated and the resulting condition seems to be more easily tractable, at least in low dimensional problems. Furthermore, there is a number of easily verifiable conditions which imply (4.1.4), cf. Ye (2000). One of them is the MPEC-LICQ (linear independence constraint qualification) which will be presented in the next part of this section. If f is continuously differentiable, we can think about the strengthening of the above statements (Theorems 4.1, 4.2) along the lines of Theorem 1.5. To this purpose we will assume that

$$\kappa := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m | q(x,y) \le 0\}, \tag{4.1.6}$$

where the map $q[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^t]$ is continuously differentiable. This enables to write down the overall constraint system from (4.0.1) in the compact form

$$\left\{ z \in \mathbb{R}^n \times \mathbb{R}^m \middle| \begin{bmatrix} g(z) \\ q(z) \end{bmatrix} \in \mathrm{G}phQ \times \mathbb{R}^t_- \right\}. \tag{4.1.7}$$

Let $J(\widehat{x}, \widehat{y})$ be the index set of inequalities from (4.1.6) that are active at $(\widehat{x}, \widehat{y})$.

Theorem 4.3. Assume that Q is polyhedral and the matrix

$$\begin{bmatrix} 0 & -(\nabla_x F(\widehat{x},\widehat{y}))^T & ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \\ E & -(\nabla_y F(\widehat{x},\widehat{y}))^T & ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \end{bmatrix} \tag{4.1.8}$$

has the maximum column rank. Then there are multipliers $(\widehat{u},\widehat{v},\widehat{\mu}) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^a_+$, $a = |J(\widehat{x},\widehat{y})|$ such that

$$\begin{aligned} &0 = \nabla_x f(\widehat{x},\widehat{y}) - (\nabla_x F(\widehat{x},\widehat{y}))^T \widehat{v} + ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \widehat{\mu} \\ &0 = \nabla_y f(\widehat{x},\widehat{y}) + \widehat{u} - (\nabla_y F(\widehat{x},\widehat{y}))^T \widehat{v} + ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \widehat{\mu} \\ &(\widehat{u},\widehat{v}) \in \widehat{N}_{\mathrm{Gph}Q}(\widehat{y}, -F(\widehat{x},\widehat{y})). \end{aligned}$$

Proof. Since Q is polyhedral, the constraints have the structure investigated in Theorem 1.5. The rank condition, imposed on the matrix (4.1.8) ensures condition (1.4.22). Moreover, it ensures also condition (1.4.21), cf. (Rockafellar and Wets, 1998, Exercise 6.7). Therefore, it suffices to observe that (with $\widehat{z} = (\widehat{x},\widehat{y})$)

$$\widehat{N}_{(\mathrm{Gph}Q \times \mathbb{R}^t_-)}(g(\widehat{z}), q(\widehat{z})) = \widehat{N}_{\mathrm{Gph}Q}(g(\widehat{z})) \times \widehat{N}_{\mathbb{R}^t_-}(q(\widehat{z})),$$

and

$$\widehat{N}_{\mathbb{R}^t_-}(q(\widehat{z})) = \{\mu \in \mathbb{R}^t_+ | \mu^i = 0 \text{ for } i \notin J(\widehat{z})\}.$$

Thus the Fréchet normal cone to the set (4.1.7) amounts to

$$\left\{ \begin{bmatrix} 0 & -(\nabla_x F(\widehat{z}))^T & ((\nabla_x q(\widehat{z}))_J)^T \\ E & -(\nabla_y F(\widehat{z}))^T & ((\nabla_y q(\widehat{z}))_J)^T \end{bmatrix} \begin{bmatrix} u \\ v \\ \mu \end{bmatrix} \middle| (u,v) \in \widehat{N}_{\mathrm{Gph}Q}(\widehat{y}, -F(\widehat{x},\widehat{y})), \mu \geq 0 \right\},$$

and we are done. □

Condition imposed on the matrix (4.1.8) is the MPEC-LICQ mentioned above. Let us illustrate the difference between the limiting and the Fréchet normal cones by means of the MPCC, investigated in Theorem 4.2.

Theorem 4.4. Suppose that the matrix

$$\begin{bmatrix} 0 & -((\nabla_x F(\widehat{x},\widehat{y}))_{L\cup I_0})^T & ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \\ (E_{I_+ \cup I_0})^T & -((\nabla_y F(\widehat{x},\widehat{y}))_{L\cup I_0})^T & ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \end{bmatrix} \tag{4.1.9}$$

has the maximum column rank. Then there are multipliers $(\widehat{u},\widehat{v},\widehat{\mu}) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^a_+$ such that $\widehat{u}_L = 0, \widehat{v}_{I_+} = 0$,

$$\begin{aligned} &0 = \nabla_x f(\widehat{x},\widehat{y}) - ((\nabla_x F(\widehat{x},\widehat{y}))_{L\cup I_0})^T \widehat{v}_{L\cup I_0} + (\nabla_x q(\widehat{x},\widehat{y}))_J)^T \widehat{\mu} \\ &0 = (\nabla_y f(\widehat{x},\widehat{y}))_{L\cup I_0} + \widehat{u}_{L\cup I_0} - ((\nabla_y F(\widehat{x},\widehat{y}))_{L\cup I_0, L\cup I_0,})^T \widehat{v}_{L\cup I_0} + (\nabla_y q(\widehat{x},\widehat{y}))_{J, L\cup I_0})^T \widehat{\mu} \end{aligned}$$

and for $i \in I_0$ one has $\widehat{u}^i \leq 0$, $\widehat{v}^i \geq 0$.

Proof. Q is polyhedral and so, again, Theorem 1.5 can be applied. The condition, imposed on the matrix (4.1.9) implies in particular the linear independence of the vectors $\nabla q^i(\widehat{x},\widehat{y}), i \in J$, so that

$$N_\kappa(\widehat{x},\widehat{y}) = \{((\nabla q(\widehat{x},\widehat{y}))_J)^T \mu | \mu \geq 0\}.$$

Therefore the condition, imposed on the matrix (4.1.9) implies the satisfaction of condition (4.1.4). The latter is exactly the (GMFCQ) for the considered MPEC; therefore the respective Abadie CQ (1.4.21) holds true as well. Next we observe that for a pair $(a,b) \in \mathrm{Gph}\widehat{N}_{\mathbb{R}^m_+}$ one has

$$\widehat{N}_{\mathrm{Gph}N_{\mathbb{R}^m_+}}(a,b) = \sum_{i=1}^{m} \widehat{N}_{\mathrm{Gph}N_{\mathbb{R}_+}}(a^i,b^i), \tag{4.1.10}$$

$$\widehat{N}_{\mathrm{Gph}N_{\mathbb{R}_+}}(a^i,b^i) = \begin{cases} 0 \times \mathbb{R} & \text{for} \quad i \in L(a,b) \\ \mathbb{R} \times 0 & \text{for} \quad i \in I_+(a,b) \\ \mathbb{R}_- \times \mathbb{R}_+ & \text{for} \quad i \in I_0(a,b). \end{cases}$$

Thus $u_L = 0, v_{I_+} = 0$ for all $(u,v) \in \widehat{N}_{\mathrm{Gph}N_{\mathbb{R}^m_+}}(\widehat{y}, -F(\widehat{x},\widehat{y}))$ and we claim that the condition, imposed on that matrix (4.1.9), implies also condition (1.4.22). Indeed, the equation

$$\begin{bmatrix} 0 & -((\nabla_x F(\widehat{x},\widehat{y}))_{L\cup I_0})^T & ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \\ (E_{I_+\cup I_0})^T & -((\nabla_y F(\widehat{x},\widehat{y}))_{L\cup I_0})^T & ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \end{bmatrix} \begin{bmatrix} u_{I_+\cup I_0} \\ v_{L\cup I_0} \\ \mu \end{bmatrix} = 0$$

implies immediately that $u_{I_+\cup I_0}, v_{L\cup I_0}, \mu = 0$ so that altogether the linear mapping

$$\begin{bmatrix} 0 & -(\nabla_x F(\widehat{x},\widehat{y}))^T & ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \\ E & -(\nabla_y F(\widehat{x},\widehat{y}))^T & ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

is injective on the Fréchet normal cone to each (convex polyhedral) component of $\mathrm{Gph}N_{\mathbb{R}^m_+} \times \mathbb{R}^t_-$. This is sufficient for condition (1.4.22) to hold. It remains to modify formulas (4.1.5) in such a way that $\widehat{\xi} = \nabla_x f(\widehat{x},\widehat{y}), \widehat{\eta} = \nabla_y f(\widehat{x},\widehat{y})$,

$$\begin{bmatrix} \widehat{w} \\ \widehat{z} \end{bmatrix} = \begin{bmatrix} ((\nabla_x q(\widehat{x},\widehat{y}))_J)^T \mu \\ ((\nabla_y q(\widehat{x},\widehat{y}))_J)^T \mu \end{bmatrix},$$

and $(\widehat{u},\widehat{v}) \in \widehat{N}_{\mathrm{Gph}N_{\mathbb{R}^m_+}}(\widehat{y}, -F(\widehat{x},\widehat{y}))$. □

When comparing Theorems 4.2 and 4.4, the most important difference concerns the multipliers $\widehat{u}^i, \widehat{v}^i, i \in I_0$. In Theorem 4.4 they have necessarily certain signs whereas in Theorem 4.2 they merely satisfy a weaker condition of combinatorial nature. The points satisfying the optimality conditions from Theorem 4.4 are called *strongly stationary*, cf. Scheel and Scholtes (2000). It is also important to note that under strong stationarity there cannot be a direction h in the contingent cone to the set (4.1.7) at $\widehat{z}$, for which

$$\langle \nabla f(\widehat{z}), h\rangle < 0.$$

This is not true in case of the M-stationarity.

4.2 Optimality conditions via ImP

ImP is applicable under the following two principal assumptions:

(i) $\kappa = \omega \times \vartheta$;

(ii) the map S defined in (3.0.1) possesses a Lipschitz single-valued localization σ at $(\widehat{x}, \widehat{y})$.

Typically, $\vartheta = \mathbb{R}^m$ and condition (ii) is implied by (SRC) at $(\widehat{x}, \widehat{y})$, cf. Outrata et al. (1998). Another possibility is to impose condition (3.1.5) and the local single-valuedness of S around $\widehat{x}$. Under (i),(ii) it is possible to rewrite (4.0.1) locally to the form

$$\begin{array}{ll} \text{minimize} & \Theta(x) \\ \text{subject to} & \\ & x \in \omega, \end{array} \tag{4.2.11}$$

where $\Theta(x) := f(x, \sigma(x))$. Reformulation (4.2.11) is the essential step in ImP. Concerning optimality conditions, the only hurdle consists now in the computation of $\partial\Theta(\widehat{x})$.

Theorem 4.5. Let assumptions (i),(ii) be fulfilled and suppose that the multifunction P (given by (3.1.3)) is calm at $(0, \widehat{x}, \widehat{y})$. Then there exists a pair $(\widehat{\xi}, \widehat{\eta}) \in \partial f(\widehat{x}, \widehat{y})$, a KKT vector $\widehat{v} \in \mathbb{R}^m$ and vectors $\widehat{w} \in N_\omega(\widehat{x}), \widehat{z} \in N_\vartheta(\widehat{y})$ such that equations (4.1.3) hold true.

Proof. By definition, $D^*\sigma(\widehat{x}) = D^*S(\widehat{x})$. Due to incl. (1.1.9) one has under assumption (ii) that

$$\partial\Theta(\widehat{x}) \subset \{\xi + D^*S(\widehat{x})(\eta) | (\xi, \eta) \in \partial f(\widehat{x}, \widehat{y})\}.$$

By virtue of Lemma 3.1 under assumption (i) and the imposed calmness condition

$$D^*S(\widehat{x})(\eta) \subset \{(\nabla_x F(\widehat{x}, \widehat{y}))^T v | 0 \in \eta + (\nabla_y F(\widehat{x}, \widehat{y}))^T v + D^*Q(\widehat{x}, -F(\widehat{x}, \widehat{y}))(v) + N_\vartheta(\widehat{y})\}.$$

The (local) optimality of $(\widehat{x}, \widehat{y})$ yields

$$0 \in \partial\Theta(\widehat{x}) + N_\omega(\widehat{x}),$$

and it suffices to combine all three relations together. □

Observe that the calmness of P is a different condition than the calmness of $\widetilde{P}$ required (essentially) in Theorem 4.1. If $\vartheta = \mathbb{R}^m$ and assumption (ii) is ensured via (SRC) at $(\widehat{x}, \widehat{y})$, then the respective map P is calm at $(0, \widehat{x}, \widehat{y})$ ((Outrata, 2000, Proposition 3.2)). The relation

$$0 \in \eta + (\nabla_y F(\widehat{x}, \widehat{y}))^T v + D^*Q(\widehat{x}, -F(\widehat{x}, \widehat{y}))(v) + N_\vartheta(\widehat{y}) \tag{4.2.12}$$

is called the *adjoint generalized equation.* ImP can also be suitably combined with the PC^1 approach, described at the end of Section 3. This combination leads to a number of optimality conditions for problem (4.0.1) with manifold equilibria, cf. Outrata et al. (1998). We illustrate the nature of these conditions by means of the MPCC, discussed already in Theorems 4.2, 4.4.

Theorem 4.6. Consider the problem (4.0.1), where f is continuously differentiable, $\kappa = \omega \times \mathbb{R}^m$ and $Q(y) = N_{\mathbb{R}^m_+}(y)$. Suppose that (SRC) is fulfilled at $(\widehat{x}, \widehat{y})$ and for all $j \in \mathbb{K}(\widehat{x}, \widehat{y})$ the vectors $\widehat{p}_j$ are the (unique) solutions of the adjoint equations

$$\begin{bmatrix} (\nabla_y F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \qquad E_{I_+ \cup M_1} \end{bmatrix}^T p_j + \nabla_y f(\widehat{x}, \widehat{y}) = 0, \tag{4.2.13}$$

where M_1, M_2 is the partition of $I_0(\widehat{y}, -F(\widehat{x}, \widehat{y}))$, corresponding to j. Then one has

$$0 \in \nabla_x f(\widehat{x}, \widehat{y}) + \operatorname{conv} \left\{ \begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \end{bmatrix}^T \widehat{p}_j \,\middle|\, j \in \mathbb{K}(\widehat{x}, \widehat{y}) \right\} + N_\omega(\widehat{x}). \tag{4.2.14}$$

Proof. By virtue of the generalized Jacobian Chain Rule (Clarke (1983)) and incl. (3.2.19) it suffices to show that

$$\begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \end{bmatrix}^T \widehat{p}_j = (\nabla \sigma_j(\widehat{x}))^T \nabla_y f(\widehat{x}, \widehat{y}) \text{ for all } j \in \mathbb{K}(\widehat{x}, \widehat{y}),$$

where σ_j is the implicit function specified by the equation system

$$\begin{aligned} F^i(x, y) &= 0 \quad \text{for} \quad i \in L \cup M_2 \\ y^i &= 0 \quad \text{for} \quad i \in I_+ \cup M_1. \end{aligned}$$

Due to the classical Implicit Function Theorem one has

$$\nabla \sigma_j(\widehat{x}) = - \begin{bmatrix} (\nabla_y F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \qquad E_{I_+ \cup M_1} \end{bmatrix}^{-1} \begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \end{bmatrix},$$

where the existence of the inverse is ensured by (SRC) at $(\widehat{x}, \widehat{y})$. Thus,

$$\begin{aligned} (\nabla \sigma_j(\widehat{x}))^T \nabla_y f(\widehat{x}, \widehat{y}) &= - \begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \end{bmatrix}^T \left(\begin{bmatrix} (\nabla_y F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \qquad E_{I_+ \cup M_1} \end{bmatrix}^{-1} \right)^T \nabla_y f(\widehat{x}, \widehat{y}) \\ &= \begin{bmatrix} (\nabla_x F(\widehat{x}, \widehat{y}))_{L \cup M_2} \\ 0 \end{bmatrix} \widehat{p}_j, \end{aligned}$$

which was to be proved. $\square$

The terminology “adjoint equations”, used for eqs.(4.2.13), comes from optimal control. Indeed, due to assumption (ii), problem (4.0.1) can be viewed as a standard optimal control problem. By the technique, illustrated in Theorem 4.6, it is possible to derive optimality conditions for MPECs with substantially more difficult equilibria than NCP, cf. Beremlijski et al. (2002). Due to the relation between coderivatives and generalized Jacobians, however, these conditions are less sharp than the conditions involving coderivatives. We speak about the *C(larke)-stationarity*, which is weaker than the *M*-stationarity. On the other hand, conditions of the type (4.2.13),(4.2.14) are intimately connected with a nonsmooth approach to the numerical solution of problems (4.0.1), described in the next section.

5 Numerical solution of MPECs via ImP

There is a number of approaches for the numerical solution of MPECs with various degrees of generality. In these lecture notes we will pay our attention only to two rather general methods which seem to be perspective in problems coming from mechanics. The first one relies on the connection of ImP with bundle methods of nonsmooth optimization. In fact, to the numerical solution of (4.2.11) various optimization techniques can be used, cf. eg Luo et al. (1996). The connection with bundle methods, however, proved to be especially efficient in a large variety of shape optimization problems, cf. Outrata et al. (1998),Beremlijski et al. (2002). In Section 5.1 we will discuss the essential question of the computation of subgradients required by the bundle methods. Section 5.2 is then devoted to a brief description of the classical bundle idea and Section 5.3 illustrates the proposed technique by a difficult MPEC: A shape optimization of contact problems with Coulomb friction.

5.1 Computation of subgradients

Consider problem (4.2.11) and assume that ω (the set of feasible controls) is of a simple nature (eg a convex polyhedron or a box). Then a standard bundle method of nonsmooth optimization can be applied to its numerical solution, provided we are able to compute to each $x \in \omega$ the corresponding value $\Theta(x)$ and one arbitrary vector from $\bar{\partial}\Theta(x)$. To facilitate the explanation, we will assume that either our problem does not have any state constraints ($\vartheta = \mathbb{R}^m$), or it is possible to augment them to the objective by means of a locally Lipschitz penalty. Moreover, we will suppose that the solution map

$$S(x) = \{y \in \mathbb{R}^m | 0 \in F(x, y) + Q(y)\}$$

is single-valued on ω. We start with the situation, where the Lipschitz continuity of S over ω is ensured by (SRC), valid at each point $\bar{x} \in \omega$.

Theorem 5.1. Let $\bar{y} = S(\bar{x}), (\bar{\xi}, \bar{\eta}) \in \widehat{\partial} f(\bar{x}, \bar{y})$ and $\bar{p}$ is a solution of the "regular" adjoint GE

$$0 \in \bar{\eta} + (\nabla_y F(\bar{x}, \bar{y}))^T p + \widehat{D}^* Q(\bar{y}, -F(\bar{x}, \bar{y}))(p). \tag{5.1.1}$$

Then one has

$$\bar{\xi} + (\nabla_x F(\bar{x}, \bar{y}))^T \bar{p} \in \partial\Theta(\bar{x}) \subset \bar{\partial}\Theta(\bar{x}). \tag{5.1.2}$$

If f is regular at $(\bar{x}, \bar{y})$ and Gph$Q$ is normally regular at $(\bar{y}, -F(\bar{x}, \bar{y}))$, then it holds even

$$\begin{aligned} \partial\Theta(\bar{x}) \quad = \quad & \{\xi + (\nabla_x F(\bar{x}, \bar{y}))^T p | 0 \in \eta + (\nabla_y F(\bar{x}, \bar{y}))^T p + D^* Q(\bar{y}, -F(\bar{x}, \bar{y}))(p), \\ & (\xi, \eta) \in \partial f(\bar{x}, \bar{y})\}. \end{aligned}$$

Proof. As in the proof of Theorem 1.6 we can invoke formulas (1.1.9),(1.1.10) which yield

$$\{\xi+\widehat{D}^*S(\bar{x})(\eta)|(\xi,\eta)\in\bar{\partial}f(\bar{x},\bar{y})\}\subset\widehat{\partial}\Theta(\bar{x})\subset\partial\Theta(\bar{x})\subset\{\xi+D^*S(\bar{x})(\eta)|(\xi,\eta)\in\partial f(\bar{x},\bar{y})\}.$$

According to (Rockafellar and Wets, 1998, Thm.10.6) it holds

$$\begin{aligned}\widehat{D}^*S(\bar{x})(\eta)\supset\{(\nabla_x F(\bar{x},\bar{y}))^T p\ \ |0\in\eta+(\nabla_y F(\bar{x},\bar{y}))^T p+\widehat{D}^*Q(\bar{y},-F(\bar{x},\bar{y}))(p),\\ (\xi,\eta)\in\widehat{\partial}f(\bar{x},\bar{y})\text{ with some }\xi\in\mathbb{R}^n\}\end{aligned}\tag{5.1.3}$$

and, under a suitable qualification condition, one has

$$\begin{aligned}D^*S(\bar{x})(\eta)\subset\{(\nabla_x F(\bar{x},\bar{y}))^T p\ \ |0\in\eta+(\nabla_y F(\bar{x},\bar{y}))^T p+D^*Q(\bar{y},-F(\bar{x},\bar{y}))(p),\\ (\xi,\eta)\in\partial f(\bar{x},\bar{y})\text{ with some }\xi\in\mathbb{R}^n\},\end{aligned}\tag{5.1.4}$$

cf. also Lemma 3.1. The first statement follows now from (5.1.3), whereas the second one results from the equality of the right-hand sides at (5.1.3),(5.1.4) under the assumed strong regularity and additional conditions. □

We conclude that one either has to solve the "regular" adjoint GE (5.1.1) or require that f is regular at $(\bar{x},\bar{y})$ and Gph$Q$ is normally regular at $(\bar{y},-F(\bar{x},\bar{y}))$. In the latter case (5.1.1) amounts to the usual adjoint GE (4.2.12) (at $(\bar{x},\bar{y})$). Since (4.2.12) is mostly easier to solve than (5.1.1), it is important to realize that incl. (5.1.4) becomes equality whenever $\nabla_x F(\bar{x},\bar{y})$ is surjective. Indeed, one can then invoke a result from Rockafellar and Wets (1998), according to which

$$N_{\text{Gph}S}(\bar{x},\bar{y})=\begin{bmatrix}0 & -(\nabla_x F(\bar{x},\bar{y}))^T\\ E & -(\nabla_y F(\bar{x},\bar{y}))^T\end{bmatrix}N_{\text{Gph}Q}(\bar{y},-F(\bar{x},\bar{y})),$$

whenever the matrix

$$\begin{bmatrix}0 & E\\ -\nabla_x F(\bar{x},\bar{y}) & -\nabla_y F(\bar{x},\bar{y})\end{bmatrix}$$

has the full row rank. This helps in the case when f is continuously differentiable. On the basis of (1.2.13) one has then

$$\nabla_x f(\bar{x},\bar{y})+(\nabla_x F(\bar{x},\bar{y}))^T\bar{p}\in\bar{\partial}\Theta(\bar{x})$$

with $\bar{p}$ being a solution of (4.2.12) (at $(\bar{x},\bar{y})$).

The above idea can sometimes be applied also in the case, when the considered equilibrium is governed by the GE

$$0\in F(x,y)+Q(y),\tag{5.1.5}$$

with $F(x,y)=\begin{bmatrix}F_1(x,y)\\ F_2(x,y)\end{bmatrix}$, $y=(y_1,y_2)\in\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}$, $Q(y)=Q_1(y_1)\times Q_2(y_2)$; thereby F_i maps $\mathbb{R}^n\times\mathbb{R}^m$ into $\mathbb{R}^{m_i}$ and Q_i maps $\mathbb{R}^{m_i}$ into (subsets of) $\mathbb{R}^{m_i}$, $i=1,2$. Then, namely, GphQ_1 is often rather simple and the relation

$$0\in F_1(x,y)+Q_1(y_1)$$

can be used to the elimination of y_1 from the other one, i.e.

$$0 \in F_2(x, y) + Q_2(y_2).$$

The surjectivity argument needs then to be applied only to a "reduced" GE. We illustrate now the above theory again by means of an NCP, i.e., we put $Q(y) = N_{\mathbb{R}^m_+}(y)$.

Theorem 5.2. Consider a reference point $(\bar{x}, \bar{y}) \in \mathrm{Gph} S$ and suppose that $(\bar{\xi}, \bar{\eta}) \in \widehat{\partial} f(\bar{x}, \bar{y})$.

(i) Let the adjoint variable $\bar{p} \in \mathbb{R}^m$ be a feasible point of the system

$$\left.\begin{array}{l} p_{I_+} = 0, p_{I_0} \geq 0 \\ 0 = \bar{\eta}_L + (((\nabla_y F(\bar{x}, \bar{y}))_{L \cup I_0})^T p_{L \cup I_0} \\ \bar{\eta}_{I_0} + (((\nabla_y F(\bar{x}, \bar{y}))_{L \cup I_0})^T p_{L \cup I_0} \leq 0. \end{array}\right\} \tag{5.1.6}$$

Then one has

$$\xi + (\nabla_x F(\bar{x}, \bar{y}))^T \bar{p} \in \bar{\partial} \Theta(\bar{x}). \tag{5.1.7}$$

(ii) Let f be continuously differentiable, the matrix

$$\begin{bmatrix} 0 & E_{I_+ \cup I_0} \\ -\nabla_x F(\bar{x}, \bar{y})_{L \cup I_0} & -\nabla_y F(\bar{x}, \bar{y})_{L \cup I_0} \end{bmatrix}$$

have the maximum row rank and M_1, M_2 be an arbitrary partition of I_0. Then the relation (5.1.7) is fulfilled whenever $\bar{p}$ is a solution of the linear system

$$\left.\begin{array}{l} 0 = p_{I_+ \cup M_1} \\ 0 = \bar{\eta}_{L \cup M_2} + ((\nabla_y F(\bar{x}, \bar{y}))_{L \cup M_2, L \cup M_2})^T p_{L \cup M_2} \end{array}\right\}. \tag{5.1.8}$$

Proof. Statement (i) follows from the first assertion of Theorem 5.1, because (5.1.6) amounts exactly to (5.1.1) by virtue of (4.1.10). To prove statement (ii), note that the adjoint equation (4.2.12) attains here the form

$$\left.\begin{array}{l} 0 = p_{I_+}, 0 = u_L \\ 0 = \bar{\eta}_{L \cup I_0} + ((\nabla_y F(\bar{x}, \bar{y}))_{L \cup I_0, L \cup I_0})^T p_{L \cup I_0} + u_{L \cup I_0} \\ \text{for } i \in I_0 \text{ either } p^i u^i = 0 \text{ or } p^i > 0 \text{ and } u^i < 0. \end{array}\right\} \tag{5.1.9}$$

We can thus partition I_0 into M_1, M_2 and put $p^i = 0$ for $i \in M_1, u^i = 0$ for $i \in M_2$. Therefore u^i is free for $i \in M_1$ and we obtain exactly the equation system (5.1.8). □

As in Section 3, the index sets L, I_+, I_0 are related to the argument $(\bar{y}, -F(\bar{x}, \bar{y}))$. One immediately sees that (5.1.8) is substantially simpler than the system in statement (i). For each choice of M_1, M_2, the solvability of (5.1.8) (for each $\bar{\eta}_{L \cup I_0}$) is ensured by (SRC) at $(\bar{x}, \bar{y})$. The same procedure as in (ii) can be derived by means of the PC^1 approach.

It remains to comment on the situation, when we are not able to fulfill the assumptions needed for the computation of an element from $\bar{\partial}\Theta(\bar{x})$ via the standard adjoint GE (in (5.1.4)). In such a case it is possible that we supply the used bundle method a false subgradient information. Theoretically, the convergence of a general bundle method in this case has been investigated in Mikhalevich et al. (1987) and especially in Dempe (2002). It turned out that the method provides then a sequence converging to a point which stationarity is weaker than the C-stationarity. This is, in a sense, a positive result, because it justifies the combination of ImP with a bundle method without too restrictive assumptions.

To ensure the convergence of bundle method, one also needs the co-called weak semismoothness of the minimized function which excludes some pathological functions of the type $x^2 \sin \frac{1}{x}$. We say that a function $f[\mathbb{R}^p \to \mathbb{R}]$ is *weakly semismooth* at $\bar{x} \in \mathbb{R}^p$, provided it is Lipschitz near $\bar{x}$ and the limit

$$\lim_{\substack{\xi_i \in \bar{\partial} f(\bar{x}+\theta_i h)\\ \theta_i \downarrow 0}} \langle \xi_i, h \rangle \tag{5.1.10}$$

exists for all $h \in \mathbb{R}^p$. This assumption is automatically fulfilled, provided S is a PC^1 -map and the objective is smooth. Otherwise, this assumption should be tested but its violation is very unlikely.

5.2 Bundle idea in nonsmooth optimization

Bundle methods have been originally designed for unconstrained minimization of convex continuous functions. Substantial further developments have lead, however, to variants which are applicable also to the minimization of (nonconvex) locally Lipschitz functions subject to various constraints (including, eg, nonsmooth inequalities). In the following we explain briefly the essential idea in the convex framework and thereafter point out some substantial changes indispensable for the minimization of nonconvex functions. Finally, we mention some developments which essentially contributed to the considerable performance of modern implementations.

Assume first that $f[\mathbb{R}^n \to \mathbb{R}]$ is convex continuous and consider a sequence of points $x_1, x_2, \ldots, x_k$ and the associated bundle of subgradients $g_1, g_2, \ldots, g_k$ with $g_i \in \partial g(x_i), i = 1, 2, \ldots, k$. Put

$$\alpha_{k,i} := f(x_k) - f(x_i) - \langle g_i, x_k - x_i \rangle \tag{5.2.11}$$

so that $\alpha_{ki} \geq 0$ due to the convexity of f. This enables to introduce a new function $\widehat{f}[\mathbb{R}^n \to \mathbb{R}]$ defined by

$$\begin{aligned} \widehat{f}(x) &:= \max_{i=1\ldots k} \{f(x_i) + \langle g_i, x - x_i \rangle\} = f(x_k) + \max_{i=1\ldots k} \{-f(x_k) + f(x_i) + \langle g_i, x - x_k \rangle \\ &+ \langle g_i, x_k - x_i \rangle\} = f(x_k) + \max_{i=1\ldots k} \{-\alpha_{k,i} + \langle g_i, x - x_k \rangle\}. \end{aligned} \tag{5.2.12}$$

$\widehat{f}$ is piecewise affine and serves as a lower approximation (model) of f on a neighborhood of x_k. Note that if $\alpha_{k,i}$ is large, then the respective affine component $-\alpha_{k,i} + \langle g_i, x - x_k \rangle$ can hardly be active at x close to x_k. We set now

$$d := x - x_k$$

and employ the model function $\widehat{f}$ in the *direction finding subproblem*, where one attempts to compute the "best" descent direction $\widehat{d}$. Since $\widehat{f}$ can be quite different from f at points, far from x_k, it is reasonable not to permit too large directions d. Therefore we compute $\widehat{d}$ as the second component of the (unique) solution $(\widehat{v}, \widehat{d})$ of the quadratic program

$$\begin{array}{ll} \text{minimize} & v + \frac{1}{2t}\|d\|^2 \\ \text{subject to} & \\ & \langle g_i, d\rangle - \alpha_{k,i} \leq v, \;\; i = 1, 2 \ldots, k, \end{array} \tag{5.2.13}$$

where the parameter $t > 0$ specifies the neighborhood of x_k on which we consider $\widehat{f}$ to be an acceptable approximation of f. The first component $\widehat{v}$ estimates the difference $f(x_k + \widehat{d}) - f(x_k)$. It is an easy task to verify that the optimization problem (in variable $\lambda \in \mathbb{R}^k$)

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\sum_{i=1}^{k} \lambda^i g_i\|^2 + \frac{1}{t}\sum_{i=1}^{k} \lambda^i \alpha_{k,i} \\ \text{subject to} & \\ & \sum_{i=1}^{k} \lambda^i = 1, \;\; \lambda^i \geq 0 \end{array} \tag{5.2.14}$$

is the standard Lagrangian dual to (5.2.13). To its numerical solution various efficient techniques can be used and one has

$$\begin{array}{l} \widehat{d} = -t \sum_{i=1}^{k} \widehat{\lambda}^i g_i \\ \widehat{v} = -\frac{1}{t}\|\widehat{d}\|^2 - \sum_{i=1}^{k} \widehat{\lambda}^i \alpha_{k,i}, \end{array} \tag{5.2.15}$$

where $\widehat{\lambda}$ is a solution of (5.2.14). When k (the dimension of the bundle) is not too high, it is reasonable to solve (5.2.13) via its dual (5.2.14) and compute its solution by (5.2.15). As the next step in classical bundle methods one performs a special line search (Lemaréchal (1981)) along the direction $\widehat{d}$ which answers a new point $y_{k+1} = x_k + \theta\widehat{d}, \theta \in [0, 1]$, and a new subgradient $g^+ \in \partial f(y_{k+1})$ such that

$$\langle g^+, \widehat{d}\rangle \geq m_1 \widehat{v} \tag{5.2.16}$$

and either

$$f(y_{k+1}) \leq f(x_k) + m_2 \theta \widehat{v} \tag{5.2.17}$$

or

$$\alpha_{k,k+1} \leq m_3 \sum_{i=1}^{k} \widehat{\lambda}^i \alpha_{k,i}. \tag{5.2.18}$$

In (5.2.16)-(5.2.18) the given parameters m_1, m_2, m_3 are positive and fulfill the inequalities $m_2 < m_1, m_1 + m_3 < 1$. Ineq. (5.2.17) ensures a "sufficient" decrease of the objective value when moving from x_k to y_{k+1}; so, if it holds, we set $x_{k+1} := y_{k+1}$ and include g^+ to the bundle. This is called a *serious step*. Otherwise we perform a *null step* which means that we do not leave x_k but improve our model $\widehat{f}$ by including a new affine component generated by y_{k+1} and g^+. The above conditions ensure that in both cases the new direction finding subproblem will produce a descent direction, different from $\widehat{d}$. Indeed, in case of the serious step

$$\langle g^+, \widehat{d}\rangle - \alpha_{k+1,k+1} = \langle g^+, \widehat{d}\rangle \geq m_1\widehat{v} > \widehat{v}$$

and in case of the null step

$$\langle g^+, \widehat{d}\rangle - \alpha_{k,k+1} \geq m_1\widehat{v} - m_3 \sum_{i=1}^{k} \widehat{\lambda}^i \alpha_{k,i} = m_1(-\frac{1}{t}\|\widehat{d}\|^2)$$

$$+(m_1 + m_3)(-\sum_{i=1}^{k} \widehat{\lambda}^i \alpha_{k,i}) > \widehat{v}.$$

Consequently, the pair $(\widehat{v}, \widehat{d})$ is infeasible with respect to the respective direction finding subproblems. The line search from Lemaréchal (1981) is consistent in the sense that it produces a serious step or a null step after a finite number of computations of function values and subgradients. The whole method is now organized in such a way that the stabilizing parameter t is changed in the course of iterations. The main convergence statement claims that after a finite number of null steps we either end up with a serious step or verify an approximate optimality of the current iterate (x_k) in the sense that

$$\left\|\sum_{i=1}^{k} \widehat{\lambda}^i g_i\right\| \leq \mathcal{E} \qquad \text{and} \qquad \sum_{i=1}^{k} \widehat{\lambda}^i \alpha_{k,i} \leq \mathcal{E}, \tag{5.2.19}$$

where $\mathcal{E}$ is an a priori specified accuracy. Note that inequalities (5.2.19) imply that

$$f(x_k) \leq f(x) + \mathcal{E}\|x - x_k\| + \mathcal{E} \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{5.2.20}$$

Indeed, on the basis of (5.2.12),

$$f(x) \geq f(x_k) + \langle g_i, x - x_k\rangle - \alpha_{k,i} \quad \text{for all} \quad i.$$

For an arbitrary set of coefficients $\lambda^i \geq 0$, $\sum\limits_{i=1}^{k} \lambda_i = 1$ one has

$$\sum_{i=1}^{k} \langle \lambda_i g_i, x - x_k\rangle \leq f(x) - f(x_k) + \sum_{i=1}^{k} \lambda_i \alpha_{k,i},$$

which immediately implies (5.2.20).

For an appropriate treatment of t (or other quantities with the same role) various strategies have been proposed, cf. Lemaréchal et al. (1981),Lemaréchal (1989).

To prevent a permanent increase of the number of subgradients in the bundle (and thus the dimension of problem (5.2.14)) one resets the bundle, mostly after each serious or null step. A successful resetting strategy consists in selecting a fixed number of subgradients with the smallest values of $\alpha_{k,i}$. From the rest one constructs the so-called aggregated subgradient by using the solution of (5.2.14). In this way it is possible to work with a bundle of a reasonable size without destroying the convergence of the method.

If we do not require f to be convex, the reals $\alpha_{k,i}$ need not be nonnegative and $\widehat{f}$ need not be a lower approximation of f around x_k. This hurdle can be overcome by the following trick: We replace $\alpha_{k,i}$ with different quantities $\beta_{k,i}$, defined by

$$\beta_{k,i} := \max\{\alpha_{k,i}, c_0\|x_k - x_i\|\},$$

where c_0 is a suitable positive number. Clearly, the reals $\beta_{k,i}$ are nonnegative by definition and one can introduce a modified model function

$$\widetilde{f}(x) := f(x_k) + \max\{-\beta_{k,i} + \langle g_i, x - x_k\rangle\}. \tag{5.2.21}$$

Note that whenever $\|x_k - x_i\|$ is too large, the corresponding affine component $-\beta_{k,i} + \langle g_i, x - x_k\rangle$ is very unlikely to be active at x close to x_k.

To ensure a proper behaviour of the line search in this case, one needs to assume that f is weakly semismooth, cf. (5.1.10). With this additional assumption all statements can be proved in a similar manner like in the convex case. We have to confess, however, that the model (5.2.21) is much less satisfactory than its convex counterpart. Also the inequalities (5.2.19) do not imply the approximate optimality (5.2.20). All we get from (5.2.19) in the nonconvex case is that 0 "lies up to $\mathcal{E}$" in the convex hull of certain subgradients computed at the points not too far away from x_k. This corresponds to "almost" stationarity in smooth optimization.

The numerical method described above proved to be rather efficient in numerous nonsmooth optimization problems. At the end of the eighties the bundle idea has been successfully coped with the trust region philosophy. This lead to an improved treatment of t during the iteration process and enabled to eliminate completely the line search in the convex case. The resulting Bundle-Trust (BT) algorithm (Schramm and Zowe (1992)) is able to minimize convex piecewise affine functions in a finite number of iterations. In the nonconvex case, however, the line search could not be completely eliminated: It has to be applied if the modified conditions (5.2.17), (5.2.18) do not permit to perform the serious step or the null step and also the trust region (specified by t) cannot be decreased. Nevertheless, this branch of the algorithm is only an emergency exit. The BT algorithm has been successfully applied to the numerical solution of all MPECs in Outrata et al. (1998). As further important developments let us mention a class of methods relying on the Yosida approximation of the minimized function (Lemaréchal and Sagastizábal (1997)) and the method of Lukšan and Vlček (1998) which replaces the model (5.2.12) by a piecewise quadratic model. This method proved its efficiency eg in a class of equilibria, governed by HVIs, where the piecewise affine model (5.2.12) is quite inadequate.

Concerning possible constraints we have to distinguish between affine and nonaffine equations and inequalities. The affine case can be treated by a suitable modification of the direction finding subproblems. The nonaffine one, however, requires a more sophisticated approach. In this respect, eg the *filter idea*, proposed in Fletcher and Leyffer (2002) seems to be rather efficient.

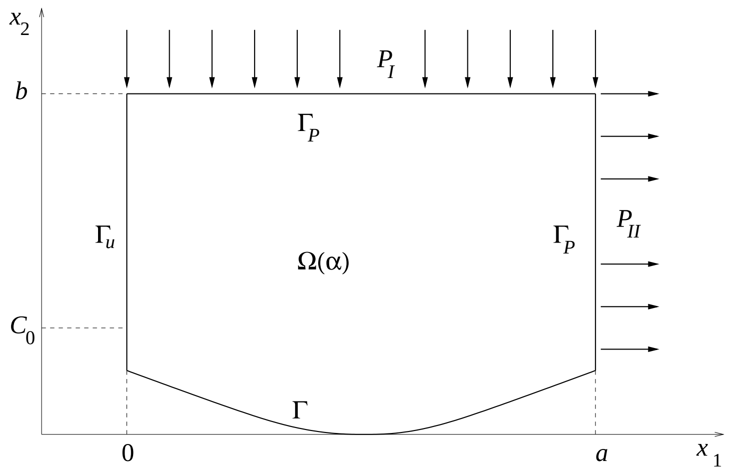

Figure 1. The elastic body and applied loads.

5.3 Shape optimization of contact problems with Coulomb friction

To demonstrate the performance of ImP combined with a bundle method, we have chosen a rather difficult MPEC with an equilibrium mentioned already in Section 2.1. This MPEC is thoroughly analyzed in Beremlijski et al. (2002); here we given only a brief description of the numerical realization of the method and provide the results of some test examples. The equilibrium is governed by the GE

$$\left.\begin{array}{lcl} 0 & \in & A_{\tau\tau}(x)u_\tau + A_{\tau\nu}(x)u_\nu - l_\tau(x) + \widetilde{Q}(u_\tau, \lambda_\nu) \\ 0 & = & A_{\nu\tau}(x)u_\tau + A_{\nu\nu}(x)u_\nu - l_\nu(x) \\ 0 & \in & u_\nu + x + N_{\mathbb{R}^p_+}(\lambda_\nu) \end{array}\right\} \qquad (5.3.22)$$

with the control variable $x \in \mathbb{R}^n$ specifying the shape of the contact boundary Γ and the state variable

$$y := (u_\tau, u_\nu, \lambda_\nu) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$$

consisting from the normal and the tangential component of the displacements on the contact boundary Γ and the normal component of the contact stress vector on Γ, see Fig. 1. This figure also shows the distribution of external loads P_I, P_{II} on the boundary Γ_P. Further, Γ_u is the part of the boundary with prescribed Dirichlet condition. $A_{\tau,\tau}, A_{\tau,\nu}, A_{\nu\tau}$ and $A_{\nu\nu}$ are blocks of the appropriate restriction of the *stiffness matrix* to Γ. These matrices depend on x in a continuously differentiable way. This holds true also for the vectors l_τ, l_ν reflecting the action of external forces on Γ. The multivalued part in the first line of (5.3.22) is given by (cf. 2.1.19)

$$\widetilde{Q}(u_\tau, \lambda_\nu) = \lambda_\nu \bullet \partial j(u_\tau), \;\; j(u_\tau) = \mathcal{F} \sum_{i=1}^{p} |u_\tau^i|,$$

where $\mathcal{F}$ is the *friction coefficient*.

The shape optimization problem is defined as follows:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \\ & y \text{ solves the GE } (5.3.22) \\ & x \in \omega \end{array}$$

with

$$\omega = \{x \in \mathbb{R}^n \quad | \; 0 \le x^i \le C_0, i = 0,1,\ldots,n; \\ |x^{i+1} - x^i| \le \tfrac{C_1}{n}, i = 0,1,\ldots,n-1; \sum_{i=0}^{n} x^i = C_2(n+1)\},$$

where C_0, C_1, C_2 are given positive constants. The equality constraint in the definition of ω has a physical meaning of preserving the body volume.

As it has been proved in Beremlijski et al. (2002), the respective solution map S is single-valued and locally Lipschitz on ω. Consequently, we are entitled to apply ImP. The resulting program (4.2.11) was solved by the BT algorithm from Schramm and Zowe (1992). In every BT iteration we have to solve the equilibrium problem which we formulate to this purpose as a fixed-point problem. For that, we use the splitting variant of the fixed-point method introduced in Haslinger et al. (2002). This is basically the method of successive approximations where, at each step, we solve the contact problem with a given friction. The iterative process then updates the coefficient of the given friction. The problem with the given friction is solved using the so-called reciprocal variational formulation that leads to a quadratic programming problem with simple box constraints. For its solution we used a special splitting technique, based on a version of the Gauss-Seidel algorithm.

To compute a subgradient information required by BT one can use the upper approximation of D^*S on the right-hand side of (5.1.4). Suitable elements from D^*Q are selected on the basis of index sets describing the position of the considered point on the graph of Q. Concerning $\widetilde{Q}$, we have to distinguish for each node in contact with the obstacle (superscript i) among the *slip*

$$u_\tau^i \neq 0,$$

the *strong stick*

$$u_\tau^i = 0, \; |(l_\tau(x) - A_{\tau\tau}(x)u_\tau - A_{\tau\nu}(x)u_\nu)^i| < \mathcal{F}\lambda^i$$

and the *weak stick*

$$u_\tau^i = 0, \; |(l_\tau(x) - A_{\tau\tau}(x)u_\tau - A_{\tau\nu}(x)u_\nu)^i| = \mathcal{F}\lambda^i.$$

The adjoint GE becomes then a linear equation systems whose solvability follows from the properties of the stiffness matrix. To verify that a vector computed in this way is indeed a subgradient of the respective composite objective Θ, an additional testing is needed; however, in the test runs we did not observe any difficulties of BT caused by a false subgradient information.

Next, we will present results of a numerical example. The shape of the unloaded elastic body $\mathcal{O}(x), x \in \omega$, is defined by

$$\mathcal{O}(x) = \{(\xi_1, \xi_2) \in \mathbb{R}^2 \,|\, \xi_1 \in (0, a), \mathcal{B}_x(\xi_1) < \xi_2 < b\},$$

where $\mathcal{B}_x$ is a suitable approximation of the shape of Γ, dependent on x, see Figure 2.

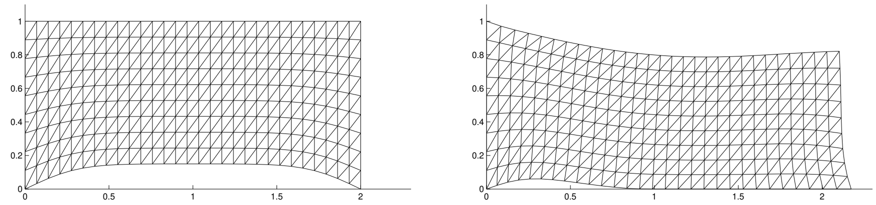

Figure 2. Initial design, unloaded and loaded case.

We try to identify the contact normal stress distribution with a given function $\bar{\lambda}_\nu$. The shape optimization problem can be written as

$$\begin{aligned} &\text{minimize} && \|\bar{\lambda}_\nu - \lambda_\nu\|^2 \\ &\text{subject to} && \\ & && x \in \omega. \end{aligned}$$

We discretized $\mathcal{O}(x)$ by a regular 29×9 mesh, i.e., we had 261 nodes and 522 unknowns in the state problem. The results are shown in Figure 3.

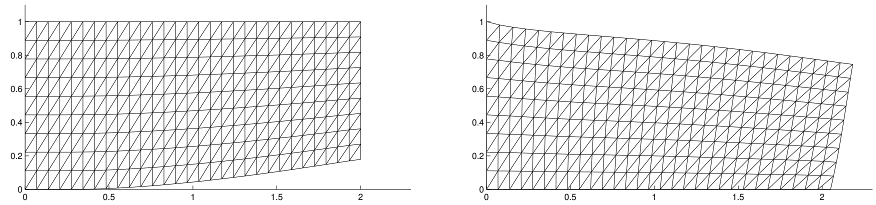

Figure 3. Optimal design, unloaded and loaded case.

6 Nonlinear programming approach and evolutionary equilibrium

Numerous MPECs do not fulfill the principal assumptions needed for the application of ImP, cf. Section 4.2. To their numerical solution, various alternative approaches have been proposed and tested. Among them we find methods based on the penalization of the lower-level constraint, on the penalization of the whole equilibrium constraint and on the PC^1 approach. Recently, however, it turned out that, at least for MPCCs, one can modify some standard NLP methods to enable them to cope with the original problem formulation

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & \\ & F(x,y) \geq 0,\ y \geq 0,\ \langle F(x,y), y\rangle = 0 \\ & (x,y) \in \kappa. \end{array} \tag{6.0.1}$$

In fact, in this so called NLP approach it is convenient to generalize the complementarity constraints from (6.0.1) to the form

$$F(x,y) \geq 0,\ H(x,y) \geq 0,\ \langle F(x,y), H(x,y)\rangle = 0, \tag{6.0.2}$$

where $H[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m]$ is continuously differentiable. In Section 6.1 we will briefly discuss the main issues associated with this recent development.

In some mechanical applications it is important to take into account the evolution nature of the considered process. One of the reasons is the necessity to respect the irreversible changes occuring during the process. Such a situation arises eg in delamination, where the modeling via a stationary HVI does not always lead to a physically acceptable solution. It seems to be better to replace a stationary HVI by a more sophisticated model (Kočvara et al.) which leads, after a suitable discretization, to a finite sequence of coupled GEs associated with the single time levels. MPECs with such evolutionary equilibria have been considered in Kocvara and Outrata (2004), where necessary optimality conditions are derived. Concerning the numerical solution of these problems, the NLP approach seems to be useful, but it is important to take the special structure, induced by the time discretization, into account. This work has not been done yet, but one can employ the ideas used in the adaptation of SQP methods to optimal control problems. In Section 6.2 we will present the evolution model from Kocvara and Outrata (2004) and discuss the optimality conditions.

6.1 NLP approach

Consider an MPCC with the equilibrium constraint given by (6.0.2). Put again $z := (x,y)$ and assume that κ is given by (4.1.6). Then we face a standard NLP with inequalities and one difficult equality

$$\langle F(z), H(z)\rangle = 0. \tag{6.1.3}$$

It is exactly this equality which is responsible for all hurdles associated with this type of MPCC. In particular, the KKT vectors can be unbounded, which destroys our estimates of the rate of convergence. In what follows we will employ SQP methods as a solver. Therefore we strengthen the smoothness requirements and assume that the functions f, F, H and q are twice continuously differentiable. The NLP approach has been initiated by the paper (Scholtes (2001)), where the author suggests to replace (6.1.3) by the harmless inequality

$$\langle F(z), H(z)\rangle \leq \tau$$

with some $\tau > 0$ and to approach a solution of the considered MPCC by solutions of the respective nonlinear programs with $\tau \downarrow 0$. Under certain conditions this sequence converges to various types of stationary points. In the subsequent works, however, an SQP method has been applied directly to the considered MPCC, where (6.1.3) has been handled as the inequality

$$\langle F(z), H(z) \rangle \leq 0. \tag{6.1.4}$$

Of course, this formal change has no influence on the constraint qualifications. Nevertheless, for modern SQP implementations this reformulation proved to be a useful step: both the failure quota as well as the convergence improved. When analysing this phenomenon, the mathematicians encountered a distinguished, so-called basic KKT vector to which SQP methods converge provided all QP subproblems remain consistent. This is, unfortunately, not always the case and so, various devices have been proposed to maintain their consistency or to return to a consistent QP after an inconsistency has been realized. In Anitescu (2000) the usage of the so-called *elastic mode* is recommended. This consists in relaxing of (6.1.4) and adding of $\langle F(z), H(z) \rangle$ to the objective with a suitable penalty parameter. This parameter is updated during the iteration process together with the relaxation parameter. Another way how to cope with the QP inconsistency is to relax only the linearization of (6.1.4). This way has been adopted in the successful code Filter MPEC together with a restoration mode which enforces consistency of the next QP subproblem, cf. Fletcher et al. (2002), Fletcher and Leyffer (2004).

Among the assumptions, needed to ensure a local superlinear convergence of SQP methods in the above setting we find also a respective variant of MPEC-LICQ which amounts to the requirement that the matrix

$$\left[((\nabla H(\widehat{z}))_{I_+ \cup I_0})^T - ((\nabla F(\widehat{z}))_{L \cup I_0})^T \;\; ((\nabla q(\widehat{z}))_J)^T \right]$$

has the maximum column rank. Thereby, $\widehat{z}$ is an accumulation point of the sequence generated by the used SQP algorithm, J is the active index set for the inequality $q(z) \leq 0$ (at $\widehat{z}$) and the index sets L, I_+, I_0 are this time associated with the point $(H(\widehat{z}), -F(\widehat{z}))$. It follows from Theorem 4.4 that, under MPEC-LICQ, $\widehat{z}$ is strongly stationary whenever it is a (local) minimum of the solved MPEC. In Scholtes (2001) it has been proved that MPEC-LICQ is generically fulfilled so that it cannot be considered as a too restrictive requirement.

Let us list the main advantages and drawbacks of the NLP approach.

+ The NLP approach can directly handle state constraints (unlike the ImP technique).
+ There are several well-developed, robust and sophisticated NLP codes, both academic and commercial. Also, some MPECs allow for formulations using tools like GAMS or AMPL. In this way, one can easily generate new problems and solve them using standard codes.

On the other hand,

- The fact that the NLP approach works on Cartesian product of the variables x and y prevents from using special solvers for the equilibrium problems (finite elements solvers, multigrid, etc.). Both variables, control and state, are treated in the same way and any structure in the problem is ignored (in the current methods).

- The NLP approach is limited to MPECs with equilibria governed by (generalized) NCPs.

It can be expected that the performance of modern NLP methods (not only SQP) with respect to MPECs will increase. This will happen in particular due to modifications respecting the concrete structure of the considered equilibrium. An example of a specially structured equilibrium can be found in the next section.

6.2 Evolutionary equilibria

A thorough study of the delamination process, modeled usually by an elliptic HVI, has lead to a more sophisticated model (Kočvara et al.) which is able to grasp better the irreversibility of changes occurring during the process (breakage of the adhesive). After a complete discretization this model can be embeded into a class of *evolutionary equilibria* defined below.

Let T be a natural number specifying the terminal time, $\bar{y}_0 \in \mathbb{R}^m$ be an initial condition and $x \in \mathbb{R}^n$ be a control vector. Further, we are given a family of continuously differentiable functions $F_i[\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m]$ and closed-graph multifunctions $Q_i[\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \rightsquigarrow \mathbb{R}^m]$, $i = 1, 2, \ldots, T$. With these data we associate a control-dependent *evolutionary equilibrium* via the solution map $S[\mathbb{R}^n \rightsquigarrow \mathbb{R}^m]$ defined by

$$S(x) := \{y_T \in \mathbb{R}^n \,|\, 0 \in F_i(x, y_{i-1}, y_i) + Q_i(x, y_{i-1}, y_i), i = 1, 2, \ldots, T, y_0 = \bar{y}_0\}. \quad (6.2.5)$$

So, the solution y_i of the GE

$$0 \in F_i(x, y_{i-1}, y_i) + Q_i(x, y_{i-1}, y_i) \quad (6.2.6)$$

depends both on the control x (which is the same for all time levels) as well as on the solution y_{i-1} of the preceding GE (or on $\bar{y}_0$). In fact, to each x we assign a set of *trajectories* $(y_1, y_2, \ldots, y_T)$ corresponding to the *evolution* of the considered equilibrium, but only the terminal vectors y_T are considered as the image $S(x)$.

The GE (6.2.6) could be the KKT system for a convex program, where both the objective as well as the functions in the constraints depend on x and on a KKT pair from the previous GE. If all objectives are strictly convex (with respect to the unknown variable) for all values of the parameters, then S becomes single-valued. In the delamination process we have indeed to do with convex programs, but they admit nonunique solutions and so ImP cannot be applied to an MPEC with such an evolutionary equilibrium. Observe that this equilibrium does not amount to a discrete-time optimal control problem, because we do not perform an optimization over the time interval $[0, T]$; instead of it we optimize at each time level separately.

Consider now the MPEC

$$\begin{array}{ll} \text{minimize} & \varphi(x, y_T) \\ \text{subject to} & \\ & 0 \in F_i(x, y_{i-1}, y_i) + Q_i(x, y_{i-1}, y_i), i = 1, 2, \ldots, T \\ & y_0 = \bar{y}_0 \\ & x \in \omega, \end{array} \quad (6.2.7)$$

where the objective $\varphi[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is locally Lipschitz and the set of admissible controls ω is nonempty and closed. In contrast to the previous notation y means now the trajectory $(y_1, y_2, \ldots, y_T)$. Further, we put $\bar{m} = Tm$ and denote by $\nabla_j F_i, j = 1, 2, 3$, the partial Jacobian of F_i with respect to the jth variable. Finally, $\widetilde{Q}_1(x, y_1) := Q_1(x, \bar{y}_0, y_1)$.

Theorem 6.1 Let $(\widehat{x}, \widehat{y}_T)$ be a (local) solution of the MPEC (6.2.7) and $\widehat{y}$ be the corresponding trajectory. Put $\widehat{y}_0 = \bar{y}_0$, denote $\widehat{c}_i = -F_i(\widehat{x}, \widehat{y}_{i-1}, \widehat{y}_i), i = 1, 2, \ldots, T$, and define the multifunction $P[\mathbb{R}^{\bar{m}} \times \mathbb{R}^{nT} \times \mathbb{R}^{(m-1)T} \times \mathbb{R}^{\bar{m}} \rightsquigarrow \mathbb{R}^n \times \mathbb{R}^{\bar{m}}]$ by

$$
\begin{aligned}
&P(\xi_1, \xi_2, \ldots, \xi_T, \alpha_1, \alpha_2, \ldots, \alpha_T, \beta_2, \beta_3, \ldots, \beta_T, \gamma_1, \gamma_2, \ldots, \gamma_T) \\
&:= \{(x, y) \in \omega \times \mathbb{R}^{\bar{m}} \,|\xi_1 \in F_1(x, \widehat{y}_0, y_1) + \widetilde{Q}_1(x - \alpha_1, y_1 - \gamma_1),\ \xi_i \in F_i(x, y_{i-1}, y_i) \\
&+Q_i(x - \alpha_1, y_{i-1} - \beta_i, y_i - \gamma_i), i = 2, 3, \ldots, T\}.
\end{aligned}
$$

Assume that P is calm at $(0, \widehat{x}, \widehat{y})$. Then there exist four sequences of adjoint vectors $p_1, p_2, \ldots, p_T, q_2, q_3, \ldots, q_T, v_1, v_2, \ldots, v_T, w_1, w_2, \ldots, w_T$ and subgradients $(\kappa, \eta) \in \partial\varphi(\widehat{x}, \widehat{y}_T)$ such that

$$
\begin{aligned}
&(p_1, v_1) \in D^*\widetilde{Q}_1(\widehat{x}, \widehat{y}_1, \widehat{c}_1)(w_1) \\
&(p_i, q_i, v_i) \in D^*Q_i(\widehat{x}, \widehat{y}_{i-1}, \widehat{y}_i, \widehat{c}_i)(w_i), i = 2, 3, \ldots, T
\end{aligned}
\tag{6.2.8}
$$

and the adjoint equation system

$$
\left.
\begin{aligned}
&0 = \eta + (\nabla_3 F_T(\widehat{x}, \widehat{y}_{T-1}, \widehat{y}_T))^T w_T + v_T \\
&0 = (\nabla_3 F_{T-1}(\widehat{x}, \widehat{y}_{T-2}, \widehat{y}_{T-1}))^T w_{T-1} + v_{T-1} \\
&\qquad +(\nabla_2 F_T(\widehat{x}, \widehat{y}_{T-1}, \widehat{y}_T))^T w_T + q_T \\
&\ldots\ldots \\
&0 = (\nabla_3 F_1(\widehat{x}, \widehat{y}_0, \widehat{y}_1))^T w_1 + v_1 + (\nabla_2 F_2(\widehat{x}, \widehat{y}_1, \widehat{y}_2))^T w_2 + q_2
\end{aligned}
\right\}
\tag{6.2.9}
$$

is fulfilled. Moreover, one has

$$
0 \in \kappa + \sum_{i=1}^{T} \left[\nabla_1 F_i(\widehat{x}, \widehat{y}_{i-1}, \widehat{y}_i))^T w_i + p_i\right] + N_\omega(\widehat{x}). \tag{6.2.10}
$$

Proof. The constraints in (6.2.7) can be written down in the form

$$
\Phi(x, y) \in \Lambda,\ x \in \omega,
$$

where

$$\Phi(x,y) = \begin{bmatrix} x \\ y_1 \\ -F_1(x,\bar{y}_0,y_1) \\ x \\ y_1 \\ y_2 \\ -F_2(x,y_1,y_2) \\ \cdots\cdots \\ x \\ y_{T-1} \\ y_T \\ -F_T(x,y_{T-1},y_T) \end{bmatrix} \quad \text{and} \quad \Lambda = Gph\widetilde{Q}_1 \times \mathop{\mathrm{X}}_{i=2}^{T} GphQ_i.$$

Now one can invoke Theorem 1.4 with the CQ (1.1.3) replaced by the assumed calmness of P. This yields the existence of a KKT vector

$$b = (p_1, v_1, -w_1, p_2, q_2, v_2, -w_2, \ldots, p_T, q_T, v_T, -w_T,) \in N_\Lambda(\Phi(\widehat{x},\widehat{y}))$$

and subgradients $(\kappa,\eta) \in \partial\varphi(\widehat{x},\widehat{y})$ such that

$$0 \in \begin{bmatrix} \kappa \\ 0 \\ \vdots \\ 0 \\ \eta \end{bmatrix} + (\nabla\Phi(\widehat{x},\widehat{y}))^T b + \begin{bmatrix} N_\omega(\widehat{x}) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{6.2.11}$$

Since $N_\Lambda(\Phi(\widehat{x},\widehat{y})) = N_{Gph\widetilde{Q}_1}(\widehat{x},\widehat{y}_1,\widehat{c}_1) \times X_{i=2}^T N_{GphQ_i}(\widehat{x},\widehat{y}_{i-1},\widehat{y}_i,\widehat{c}_i)$, see (Mordukhovich, 1988, Prop.1.6), it follows that $(p_1, v_1, -w_1) \in N_{Gph\widetilde{Q}_1}(\widehat{x},\widehat{y}_1,\widehat{c}_1)$ and $(p_i, q_i, v_i, -w_i) \in N_{GphQ_i}(\widehat{x},\widehat{y}_{i-1},\widehat{y}_i,\widehat{c}_i), i = 2, 3, \ldots, T$. In this way relations (6.2.8) have been established. The first line of (6.2.11) leads now directly to relation (6.2.10), whereas the remaining T lines generate the adjoint system (6.2.9). □

As mentioned in the text below Theorem 3.2, the calmness assumption is automatically fulfilled provided ω is convex polyhedral, all functions F_i are affine and all maps Q_i are polyhedral. An other possibility is to impose the respective (GMFCQ) which attains the following form:

(GMFCQ): The system

$$\begin{aligned}
&0 = (\nabla_3 F_T(\widehat{x},\widehat{y}_{T-1},\widehat{y}_T))^T w_T + v_T \\
&0 = (\nabla_3 F_{T-1}(\widehat{x},\widehat{y}_{T-2},\widehat{y}_{T-1}))^T w_{T-1} + v_{T-1} + (\nabla_2 F_T(\widehat{x},\widehat{y}_{T-1},\widehat{y}_T))^T w_T + q_T \\
&\cdots\cdots \\
&0 = (\nabla_3 F_1(\widehat{x},\widehat{y}_0,\widehat{y}_1))^T w_1 + v_1 + (\nabla_2 F_2(\widehat{x},\widehat{y}_1,\widehat{y}_2))^T w_2 + q_2 \\
&0 \in \sum_{i=1}^{T} \left[(\nabla_1 F_i(\widehat{x},\widehat{y}_{i-1},\widehat{y}_i))^T w_i + p_i\right] + N_\omega(\widehat{x})
\end{aligned}$$

with $(p_1, v_1) \in D^*\widetilde{Q}_1(\widehat{x}, \widehat{y}_1, \widehat{c}_1)(w_1)$ and $(p_i, q_i, v_i) \in D^*Q_i(\widehat{x}, \widehat{y}_{i-1}, \widehat{y}_i, \widehat{c}_i)(w_i), i = 2, 3, \ldots, T$, possesses only the trivial solution $p_1 = p_2 = \ldots = p_T = 0, q_2 = q_3 = \ldots = q_T = 0, v_1 = v_2 = \ldots = v_T = 0$ and $w_1 = w_2 = \ldots = w_T = 0$.

In this way we ensure even the Aubin property of P around $(0, \widehat{x}, \widehat{y})$. The optimality conditions of Theorem 6.1 attain a simpler form provided the maps Q_i do not depend on all three variables x, y_{i-1}, y_i.

Corollary 6.2. Let all assumption of Theorem 6.1 be fulfilled and assume that the maps $Q_i, i = 1.2. \ldots, T$, do not depend on x. Then there exist three sequences of adjoint vectors $q_2, q_3, \ldots, q_T, v_1, v_2, \ldots, v_T, w_1, w_2, \ldots, w_T$ and subgradients $(\kappa, \eta) \in \partial\varphi(\widehat{x}, \widehat{y}_T)$ such that

$$\begin{array}{l} v_1 \in D^*\widetilde{Q}_1(\widehat{y}_1, \widehat{c}_1)(w_1) \\ (q_i, v_i) \in D^*Q_i(\widehat{y}_{i-1}, \widehat{y}_i, \widehat{c}_i)(w_i),\ i = 2, 3, \ldots, T, \end{array}$$

the adjoint equation system (6.2.9) is satisfied, and

$$0 \in \kappa + \sum_{i=1}^{T} (\nabla_1 F_i(\widehat{x}, \widehat{y}_{i-1}, \widehat{y}_i))^T w_i + N_\omega(\widehat{x}). \tag{6.2.12}$$

Corollary 6.3. Let all assumptions of Theorem 6.1 be fulfilled and assume that for $i = 1, 2, \ldots, T$, the maps Q_i depend exclusively on variables y_i, respectively. Then there exist two sequences of adjoint vectors $v_1, v_2, \ldots, v_T, w_1, w_2, \ldots, w_T$ and subgradients $(\kappa, \eta) \in \partial\varphi(\widehat{x}, \widehat{y}_T)$ such that

$$v_i \in D^*Q_i(\widehat{y}_i, \widehat{c}_i)(w_i),\ i = 1, 2, \ldots, T,$$

and the adjoint equation system

$$\left.\begin{array}{l} 0 = \eta + (\nabla_3 F_T(\widehat{x}, \widehat{y}_{T-1}, \widehat{y}_T))^T w_T + v_T \\ 0 = (\nabla_3 F_{T-1}(\widehat{x}, \widehat{y}_{T-2}, \widehat{y}_{T-1}))^T w_{T-1} + v_{T-1} + (\nabla_2 F_T(\widehat{x}, \widehat{y}_{T-1}, \widehat{y}_T))^T w_T \\ \ldots\ldots \\ 0 = (\nabla_3 F_1(\widehat{x}, \widehat{y}_0, \widehat{y}_1))^T w_1 + v_1 + (\nabla_2 F_2(\widehat{x}, \widehat{y}_1, \widehat{y}_2))^T w_2 \end{array}\right\} \tag{6.2.13}$$

is fulfilled. Moreover, relation (6.2.12) holds true.

The next example illustrates the nature of evolutionary equilibria by means of a simple academic problem, close to a delamination process.

Example 6.1. Consider a four-string mechanical system depicted on Figure 4. In reality, the strings are at the same horizontal position, here they are plotted with some gap for presentation reasons. The elasticity modulae of the strings are $e_1, \ldots, e_4$. Assume that $e_3 = e_4$ and denote it by e. The end-nodes of the strings are denoted n_0, n_1, n_2. The vertical displacements at the nodes are u^0, u^1, u^2. Node n_0 is fixed ($u^0 = 0$), node n_2 is subjected to nonzero Dirichlet boundary condition (prescribed displacements) $u^2 = \bar{u}$.

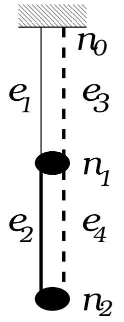

Figure 4. Four-string system.

The left-hand strings are elastic and "never" break. The right-hand strings are also elastic but can break when the relative displacement reaches a certain value.

The Dirichlet condition $\bar{u}$ depends on time. The equilibrium state at the ith time level is obtained by solving the optimization problem (in variables u_i, ζ_i)

$$
\begin{aligned}
&\text{minimize} \quad && g(e_1, e_2, \zeta_{i-1}, u_i, \zeta_i) := \\
& && \sum_{j=1}^{2} \left[e_j (u_i^j - u_i^{j-1})^2 + \zeta_i^j e (u_i^j - u_i^{j-1})^2 + (\zeta_{i-1}^j - \zeta_i^j)^2 ed \right] \\
&\text{subject to} && \\
& && u_i^2 \geq \bar{u}_i \\
& && u_i^j - u_i^{j-1} \geq 0, \ \ j = 1, 2 \\
& && \zeta_{i-1}^i \geq \zeta_i^j \geq 0, \ \ j = 1, 2,
\end{aligned}
\tag{6.2.14}
$$

where $\zeta_i \in [0,1] \times [0,1]$ is the *damage parameter* indicating the state of the breakable strings: 1 means that the respective string is intact, whereas 0 indicates that the string is broken. The values from $(0,1)$ do not have a physical meaning in this example, but can be interpreted as a partial damage of the adhesive in delamination. The first term in the objective is the strain energy of the elastic strings 1 and 2. The second term is the strain energy of the breakable strings 3 and 4 and the third term is the dissipation energy. Number d is the energy dissipated by breaking one string (this is given in our case). The first from the last two inequalities among the constraints reflects the fact that a breakage is irreversible.

We put now $x := (e_1, e_2)$ and introduce the notation

$$
G(\zeta_{i-1}, u_i, \zeta_i) := \begin{bmatrix} u_i^1 - u_i^2 \\ \zeta_i^1 - \zeta_{i-1}^1 \\ \zeta_i^2 - \zeta_{i-1}^2 \end{bmatrix},
$$

$$
\Omega_i = \left\{ a \in \mathbb{R}^4 \mid a^1 \geq 0, a^2 \geq \bar{u}_i, \, a^3 \geq 0, \, a^4 \geq 0 \right\}.
$$

This enables to write down the KKT system for (6.2.14) in the following form, where the partial derivatives are denoted by integers, indicating the positions of the respective variables.

Theorem 6.4. Let $x = \widetilde{x}$ and $\zeta_{i-1} = \widetilde{\zeta}_{i-1}$ be given and $(\widetilde{u}_i, \widetilde{\zeta}_i)$ be a solution of the respective problem (6.2.14). Then there exists a KKT vector $\widetilde{\lambda}_i \in \mathbb{R}^3_+$ such that

$$\begin{array}{l} 0 \in \nabla_{3,4} g(\widetilde{x}, \widetilde{\zeta}_{i-1}, \widetilde{u}_i, \widetilde{\zeta}_i) + (\nabla_{2,3} G(\widetilde{\zeta}_{i-1}, \widetilde{u}_i, \widetilde{\zeta}_i))^T \widetilde{\lambda}_i + N_{\Omega_i}(\widetilde{u}_i, \widetilde{\zeta}_i) \\ 0 \in -G(\widetilde{\zeta}_{i-1}, \widetilde{u}_i, \zeta_i) + N_{\mathbb{R}^3_+}(\widetilde{\lambda}_i). \end{array} \tag{6.2.15}$$

Note that in the above optimality conditions we do not need any constraint qualification, because all functions arising in the constraints are affine. The GE (6.2.15) is already in the required form (6.2.6); it suffices to put

$$\begin{array}{rcl} y_i : & = & (u_i, \zeta_i, \lambda_i), \\ F(x, y_{i-1}, y_i) : & = & \left[\begin{array}{l} \nabla_{3,4} g(x, \zeta_{i-1}, u_i, \zeta_i) + (\nabla_{2,3} G(\zeta_{i-1}, u_i, \zeta_i))^T \lambda_i \\ -G(\zeta_{i-1}, u_i, \zeta_i) \end{array} \right] \end{array}$$

and

$$Q_i(y_i) := \left[\begin{array}{l} N_{\Omega_i}(u_i, \zeta_i) \\ N_{\mathbb{R}^3_+}(\lambda_i) \end{array} \right].$$

Following Kocvara and Outrata (2004), we can now formulate an academic MPEC with the above equilibrium, where one aims to find the elasticity modulae e_1, e_2 in such a way that the dissipation energy at the terminal time T will be maximized. Since the objectives in (6.2.14) are not strictly convex, ImP cannot be applied to the numerical solution of this MPEC. On the other hand, due to the time discretization, the dimension $\bar{m} = 7T$ can be quite large even in this very simple example. So we see that the requirements, posed on an NLP method used to the numerical solution of MPECs with evolutionary equilibria, are rather demanding.

Acknowledgement.

The author is indebted to the Grant Agency of the Czech Academy of Sciences for the support in the framework of the grant A 107 5402.

Bibliography

A.M.Al-Fahed, G.E. Stavroulakis, and P.D. Panagiotopoulos. Hard and soft fingered robot grippers. The linear complementarity approach. *ZAMM-Zeit. Angew. Math. Mech.*, 71:257–265, 1991.

M. Anitescu. On solving mathematical programs with complementarity constraints as nonlinear programs. Technical Report ANL/MCS-P864-1200, Argonne National Laboratory, 2000.

J.-P. Aubin. *Optima and Equilibria.* Springer-Verlag, 1993.

J.P. Aubin and H. Frankowska. *Set-valued Analysis*. Birkhäuser, 1990.

R. Beremlijski, J. Haslinger, M. Kočvara, and J.V. Outrata. Shape optimization in contact problems with Coulomb friction. *SIAM J. Optimization*, 13:561–587, 2002.

F.H. Clarke. *Optimization and Nonsmooth Analysis*. J. Wiley & Sons, 1983.

S. Dempe. *Foundation of Bilevel Programming*. Kluwer Acad. Publ., 2002.

S. Dempe. Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization*, 52:333–359, 2003.

A.L. Dontchev and R.T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optimization*, 7:1087–1105, 1996.

I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland Publ. Comp., 1976.

M.L. Flegel, Ch. Kanzow, and J.V. Outrata. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. Technical Report 256, Math. Inst. of Julius-Maximilian University Wuerzburg, 2004.

R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239–269, 2002.

R. Fletcher and S. Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19:15–40, 2004.

R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. Technical Report NA/209, University of Dundee, 2002.

J. Haslinger, Z. Dostál, and R. Kučera. On a splitting type algorithm for the numerical realization of contact problems with Coulomb friction. *Comput. Methods Appl. Mech. Engrg.*, 191:2261–2281, 2002.

J. Haslinger and P.D. Panagiotopoulos. The reciprocal variational approach to the Signorini problem with friction. Approximation results. *Proc. of the Royal Society of Edinburgh, Sect.A*, 98:365–383, 1984.

R. Henrion, A. Jourani, and J.V. Outrata. On the calmness of a class of multifunctions. *SIAM J. Optimization*, 13:603–618, 2002.

R. Henrion and J.V. Outrata. Calmness of constraint systems with applications. *Mathematical Programming*, to appear.

J.-B. Hiriart-Urruty. Refinements of necessary optimality conditions in nondifferentiable programming I. *Appl. Math. Optim.*, 5:63–82, 1979.

D. Klatte and B. Kummer. Strong Lipschitz stability of stationary solutions for nonlinear programs and non-polyhedral variational inequalities under MFCQ. Technical Report 2003-9, Humboldt-Universitaet zu Berlin, 2003.

M. Kočvara, A. Mielke, and T. Roubíček. A rate-independent approach to the delamination problem. *Mathematics and Mechanics of Solids (to appear)*.

M. Kocvara and J.V. Outrata. On the modeling and control of delamination processes. In J. Cagnol and J.-P. Zolésio, editors, *Control and Boundary Analysis*. Marcel Dekker, 2004.

B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions: Convergence analysis. In W. Oettli and D. Pallaschke, editors, *Advances in Optimization*. Springer-Verlag, 1992.

C. Lemaréchal. A view of line searches. In W. Oettli and J. Stoer, editors, *Optimization and Optimal Control*. Springer-Verlag, 1981.

C. Lemaréchal. Nondifferentiable optimization. In G. Nemhauser, A.R. Kan, and M. Todd, editors, *Handbooks in Operations Research and Management Science, Volume I*. North-Holland Publ. Comp., 1989.

C. Lemaréchal and C. Sagastizábal. Variable metric bundle method. From conceptual to implementable forms. *Mathematical Programming*, 76:393–410, 1997.

C. Lemaréchal, J.-J. Strodiot, and A. Bihain. On a bundle algorithm for nonsmooth optimization. In O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, editors, *Nonlinear Programming 4*. Academic Press, 1981.

L. Lukšan and J. Vlček. A bundle Newton method for nonsmooth unconstrained minimization. *Mathematical Programming*, 83:373–391, 1998.

Z.Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge Univ. Press, 1996.

V.S. Mikhalevich, A.M. Gupal, and V.I. Norkin. *Methods of Nonconvex Optimization (in Russian)*. Nauka, 1987.

B.S. Mordukhovich. *Approximation Methods in Problems of Optimization and Control (In Russian)*. Nauka, 1988.

B.S. Mordukhovich. Lipschitzian stability of constraint systems and generalized equations. *Nonl. Anal.: Theory, Methods & Appl.*, 22:173–206, 1994.

B.S. Mordukhovich and J.V. Outrata. On second-order subdifferentials and their applications. *SIAM J. Optimization*, 12:139–169, 2001.

K.G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. Heldermann, 1988.

J.V. Outrata. Optimality conditions for a class of mathematical programs with equilibrium constraints. *Mathematics of Operations Research*, 24:627–644, 1999.

J.V. Outrata. A generalized mathematical program with equilibrium constraints. *SIAM J. Control Optimization*, 38:1623–1638, 2000.

J.V. Outrata, M. Kočvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Acad. Publ., 1998.

P.D. Panagiotopoulos. *Hemivariational Inequalities. Applications in Mechanics and Engineering*. Springer Verlag, 1993.

L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 17:227–244, 1993.

L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–368, 1993.

S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.

S.M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, 14:206–214, 1981.

S.M. Robinson. An implicit-function theorem for a class of nonsmooth functions. *Math. Oper. Res.*, 16:292–309, 1991.

R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer-Verlag, 1998.

H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25: 1–22, 2000.

S. Scholtes. Introduction to piecewise differentiable equations. Technical Report (Habilitation Thesis), Universitaet Karlsruhe, 1994.

S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optimization*, 11:918–936, 2001.

H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM J. Optimization*, 2:121–152, 1992.

A. Shapiro. On concepts of directional differentiability. *J. Optim. Theory Appl.*, 66: 477–487, 1990.

H. von Stackelberg. *Marktform und Gleichgewicht.* Springer-Verlag, 1934.

J.J. Ye. Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM J. Optimization*, 10:943–962, 2000.