Crash Course Optimization: Basic statements on necessary optimality conditions and stability

Bernd Kummer, kummer@mathematik.hu-berlin.de ; Sept. 2005

Abstract. This paper summarizes - as a working paper- basic facts in both finite and infinite dimensional optimization in view of optimality conditions and stability of solutions to perturbed problems. Begining with section 4, partially new and unpublished results are included, elaborated in joint work with D. Klatte, Univ. Zuerich.

Key words. Existence of solutions, (strong) duality, Karush-Kuhn-Tucker points, Kojima-function, generalized equation, MFCQ, subgradient, subdifferential, conjugate function, vector-optimization, Pareto-optimality, properly efficient points, stability, Ekeland's variational principle, Lyusternik theorem, modified successive approximation, multifunctions, Aubin property, metric regularity, calmness, upper and lower Lipschitz, variational inequality, implicit functions, Clarke's generalized Jacobian $\partial^c f$, generalized derivatives TF, CF, D^*F, strict differentiability, perturbed solutions, solution methods, penalization, (nonsmooth, quasi-) Newton method.

Contents

1 Some basic notions and statements

Given $M \subset X$ (usually at least a Banach space) and $f : M \to \mathbb{R}$ the main problem of optimization consists in finding the value

$$v = \inf_{x \in M} f(x) \tag{1.1}$$

and, if it exists, in finding some minimizer $\hat{x} \in M$.

This requires to know necessary and sufficient optimality conditions like $Df(x) = 0$ and $D^2 f(x)$ positive (semi-)definite for free minima onto $X = \mathbb{R}^n$ and $f \in C^2$.
In many situations, they attain the form
(i) $Df(\hat{x}) \in \mathcal{N}_M(\hat{x})$ where $\mathcal{N}_M(\hat{x})$ denotes some "normal cone" to M at $\hat{x}$.
(ii) $D^2 f(\hat{x})$ positive (semi-) definite on some "tangent cone" to M at $\hat{x}$.

Local solutions $\hat{x}$ are solutions to (1.1) with new $\hat{M} = M \cap \Omega$ where Ω is a nbhd of $\hat{x}$. Solutions in the original sense solutions are *global* solutions.

Nobody can solve (1.1) without supposing some analytical description of the feasible domain M which makes it more or less difficult to determine the "right" normal or tangent cones. Moreover, if f is not differentiable then something else has to replace the Frechet derivatives $Df(x)$ or $D^2 f(x)$ if f is not twice differentiable. These objects are often called *generalized derivatives* and do (usually) not represent linear functions and bilinear forms, respectively.
We shall write $f \in C^k$ if f is k-times continuously differentiable (in some nbhd of the points of interest) and $f \in C^{k,1}$ if $f \in C^k$ and $D^k f$ is locally a Lipschitz function (loc. Lipsch.). In particular, $C^{0,1}$ denotes loc. Lipsch. functions.
Particular classes of problems:
1. Linear programming

$$f(x) = \langle c, x \rangle; \quad M := \{x \in \mathbb{R}^n | Ax \le b\}; \ A = (m, n) \text{ matrix }, \ b \in \mathbb{R}^m. \tag{1.2}$$

Here, $Ax \le b$ stands for m constraints. M is a *(convex) polyhedron.*

2. Mixed integer Linear programming: As above but

$$A \text{ is a rational matrix and some } x_1, ..., x_p \text{ are required to be integer.} \tag{1.3}$$

3. Mixed integer quadratic programming: As above but

$$f(x) = \langle c, x \rangle \ + \ \langle x, Qx \rangle; \quad Q \text{ is a rational matrix.} \tag{1.4}$$

In these cases, it holds the *existence theorem:* v finite $\Rightarrow$ some $\hat{x} \in M$ realizes the infimum. (quadratic without integer variables: Evans and Gould, quadratic with integer variables: Hansel). This statement still holds if f is an n-dimensional polynom of degree 3 with rational coefficients (Belousov); it fails to hold for polynoms of degree 4 on convex polyhedrons M. In what follows, we do not deal with integer variables.

4. Classical nonlinear problems in finite dimension

$$M := \{x \in X = \mathbb{R}^n \mid g_i(x) \le 0 \ \forall i = 1, ..., m_1 \text{ and } h_\nu(x) = 0 \ \forall \nu = 1..., m_2\}. \tag{1.5}$$

5. Classical convex problems in finite dimension: As above with f, g_i convex [i.e. for f : $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \ \forall x, y \in X, \lambda \in (0,1)$] and h_ν affine-linear.

6. Classical nonlinear problems in Banach spaces

$$\begin{array}{l} X, Y, Z \text{ are B-spaces}, \ g : X \to Y, \ h : X \to Z \\ M := \{x \in X \mid g(x) \in K \text{ and } h(x) = 0\}; \\ K \text{ is a closed, convex cone in } Y \end{array} \tag{1.6}$$

7. Classical convex problems in Banach spaces
as above with f convex, h affine-linear, $\operatorname{int} K \ne \emptyset$ and g convex w.r. to K; i.e.,

$$g(\lambda x + (1-\lambda)y) - [\lambda g(x) + (1-\lambda) g(y)] \in K \quad \forall x, y \in X, \ \lambda \in (0,1). \tag{1.7}$$

Here, K replaces the non-positive orthant, appearing under 4 and 5.

Examples of nonsmooth (read: not enough differentiable) problems:

Example 1 Tschebyschev- approximation: Given a continuous function $q = q(t)$ on a real intervall $[a, b]$, find a real polynom of degree n

$$p_x(t) = x_0 + x_1 t + ... + x_n t^n$$

that makes $f(x) = \|p_x - q\|_{\max} := \max_{a \le t \le b} |p_x(t) - q(t)|$ minimal. $\diamond$

Example 2 Nash-equilibrium: Suppose that $x_1 \in X_1, ..., x_n \in X_n$ are strategies of n players und that player i obtains payoff $g_i(x) = g_i(x_1, ..., x_n)$ if each player k applies the related x_k. Assume that player i has no influence on the choices of the remaining players (no cooperation). Then, he must be satisfied with a strategy-vector x if his own choice x_i satisfies

$$\max \{g_i(x_1, .., x_{i-1}, \xi, x_{i+1}, ..., x_n) \mid \xi \in X_i\} = g_i(x).$$

The left side is never smaller than the right one. Hence *all* players must be satisfied with a strategy-vector $\hat{x}$ (then $\hat{x}$ is called a Nash-equilibrium) provided that the non-differentiable function

$$f(x) = \sum_i [\, \max \{g_i(x_1, .., x_{i-1}, \xi, x_{i+1}, ..., x_n) \mid \xi \in X_i\} - g_i(x)\,]$$

attains its minimum at $\hat{x}$ and $f(\hat{x}) = 0$. This is one of the basic models in the game theory (a first existence theorem for such games - under strong hypotheses- has been shown by J. v. Neumann; solutions of matrix games). $\diamond$

1.1 Basic optimality conditions; general interrelations

For the linear problem (1.2), it holds the
Duality theorem:
$\hat{x}$ solves (1.2) $\Leftrightarrow \hat{x} \in M$ and $\exists \hat{y} \in \mathbb{R}^m$ such that $\langle \hat{y}, A\hat{x} - b\rangle = 0,\ \hat{y} \ge 0$ and $c + A^T\hat{y} = 0$.

The latter is equivalent to the fact that $\hat{y}$ solves the
Dual problem $\max \{\langle b, y\rangle \mid A^T y = -c,\ y \ge 0\}$ and $\langle b, \hat{y}\rangle = \langle c, \hat{x}\rangle$, $\hat{x} \in M$.
This statement can be understood or reformulated in various languages - in nearly all which appeared when the tower of Babel has been build.

The main reason for this multiplicity lies in the close relation between duality, subdifferentials and stability which will be explained next.

Subgradient, subdifferential, conjugate function: Given a function $f : X \to \mathbb{R} \cup \{+-\infty\}$ (X is a B-space), put $\operatorname{dom} f = \{x \mid f(x) \in \mathbb{R}\}$. Given any $x^0 \in \operatorname{dom} f$, some $x^* \in X^*$ is said to be a subgradient of f at x^0 if (with the canonical bilinear form $\langle ., .\rangle$),

$$f(x) \ge f(x^0) + \langle x^*, x - x^0\rangle \quad \forall\, x \in X. \tag{1.8}$$

The convex set $\partial f(x^0)$ of all subgradients at x^0 is called the *(classical, convex) subdifferential of* f at x^0.

Evidently, $\partial f(x^0) = \emptyset$ is possible and $\partial f(x^0) \ne \emptyset$ implies that f is lower semi-continuous at x^0. Trivial but important:

$$x^* \in \partial f(x^0) \quad \Leftrightarrow \quad x^0 \text{ is a global minimizer of } f(x) - \langle x^*, x\rangle + \langle x^*, x^0\rangle.$$

The (improper) concave function $f^*(x^*) := \inf_{x \in X} f(x) - \langle x^*, x\rangle$ is called the *conjugate* of f. Often, one defines f^* with opposite sign as $\sup_{x \in X}\langle x^*, x\rangle - f(x)$.

Duality and subgradient: Next define, for problem 6, the (perturbation) function

$$\phi(y, z) = \inf \{f(x) \mid g(x) \in y + K \text{ and } h(x) = z\}, \quad y \in Y, z \in Z \tag{1.9}$$

with possibly improper values $+-\infty$. Suppose $(0,0) \in \operatorname{dom} \phi$ (then $\phi(0,0) = v = \inf_{x\in M} f(x) \in \mathbb{R}$), define the Lagrangian

$$L(x, y^*, z^*) = f(x) + \langle y^*, g(x)\rangle + \langle z^*, h(x)\rangle \tag{1.10}$$

and

$$H(y^*, z^*) = \inf_{x\in X} L(x, y^*, z^*) \quad (\in \mathbb{R} \cup \{-\infty\}). \tag{1.11}$$

Then, it holds (basically shown already by Ioffe and Tichomirov) the key relation

$$-(y^*, z^*) \in \partial\phi(0,0) \;\Leftrightarrow\; H(y^*, z^*) = \phi(0,0) \text{ and } \langle y^*, k\rangle \le 0 \;\forall k \in K. \tag{1.12}$$

Notice that nowhere continuity or convexity has been supposed in this context. After defining the *polar cone* $K^* = \{y^* \in Y^* \mid \langle y^*, k\rangle \le 0 \;\forall k \in K\}$ and the *dual problem*

$$\max_{y^*\in K^*, z^*\in Z^*} H(y^*, z^*) \tag{1.13}$$

one can equivalently say that $-(y^*, z^*) \in \partial\phi(0,0)$ iff (y^*, z^*) solves the dual problem with optimal value $\phi(0,0)$.

Strong duality: In the second situation, one says that strong duality holds true for the related problems. Using the definitions only, (1.12) implies

$$\partial\phi(0,0) \neq \emptyset \;\Leftrightarrow\; \max_{y^*\in K^*, z^*\in Z^*} \inf_{x\in X} L(x, y^*, z^*) \;=\; \inf_{x\in X} \sup_{y^*\in K^*, z^*\in Z^*} L(x, y^*, z^*). \tag{1.14}$$

Finally, denoting a solution of (1.13) by $(\hat{y}^*, \hat{z}^*)$ and assuming that $\hat{x}$ solves the original problem, the right-hand side of (1.14) becomes the saddle point condition

$$L(\hat{x}, y^*, z^*) \;\le L(\hat{x}, \hat{y}^*, \hat{z}^*) \le\; L(x, \hat{y}^*, \hat{z}^*) \;\; \forall x \in X,\; y^* \in K^*,\; z^* \in Z^*. \tag{1.15}$$

and yields $L(\hat{x}, \hat{y}^*, \hat{z}^*) = f(\hat{x})$.

Normal cone: The left-hand inequality yields, with $\hat{y} = g(\hat{x}) \in K$, that

$$\langle \hat{y}^*,\; k - \hat{y}\rangle \le 0 \;\forall k \in K$$

and means (by definition) that $\hat{y}^*$ belongs to the *normal cone* (in the sense of convex analysis) $\mathcal{N}_K(\hat{y})$ of K at $\hat{y} \in K$.

Next let $f, g, h \in C^1$. Then the right-hand inequality yields the necessary (Lagrange) condition

$$D_x L(\hat{x}, \hat{y}^*, \hat{z}^*) = 0 \in X^*,$$

i.e., all together in terms of *adjoint operators*: *If strong duality holds true then every solution $\hat{x}$ of the original problem satisfies*

$$Df(\hat{x}) + Dg(\hat{x})^*\hat{y}^* + Dh(\hat{x})^*\hat{z}^* = 0 \text{ for some } \;\hat{y}^* \in \mathcal{N}_K(g(\hat{x}) \text{ and } \hat{z}^* \in Z^*. \tag{1.16}$$

Corresponding elements $\hat{y}^*, \hat{z}^*$ are called Lagrange multipliers to $\hat{x}$.

For linear problems (1.2), put $K = \mathbb{R}^{m-} := \{y \in \mathbb{R}^m \mid y_i \le 0 \;\forall i\}$ whereafter $K^* = \mathbb{R}^{m+}$ and show that $\partial\phi(0) \neq \emptyset$ since ϕ is *piecewise linear* (= both continuous and affine on polyhedrons which define a finite partition of $\mathbb{R}^m$).

In the theory of optimal control, (1.16) leads to the adjugate (and Hamilton) system, $\hat{z}^*$ is assigned to the differential equation for the trajectories, $\hat{y}^*$ corresponds to "phase constraints, e.g. $g(x(t), t) \le 0$" and both have to satisfy the adjoint equation (along with an optimal trajectory).

Karush-Kuhn-Tucker points: For problems 4 in $\mathbb{R}^n$, (1.16) is the key part of the Karush-Kuhn-Tucker (KKT-) conditions, imposed on a triple $(x, y, z) \in \mathbb{R}^{n+m1+m2}$:

$$\begin{array}{l} Df(x) + \sum_i \; y_i Dg_i(x) + \sum_\nu \; z_\nu Dh_\nu(x) = 0 \\ g(x) \le 0,\;\; h(x) = 0, \;\;\; y \ge 0 \text{ and } \;\; y_i g_i(x) = 0 \;\forall i. \end{array} \tag{1.17}$$

Conditions like $y_i g_i(x) = 0$ or $u_i(x)v_i(y) = 0$ are called *complementarity conditions.*

Example 3 The problem $\min \{x \in \mathbb{R} \mid x^2 = 0\}$ shows: Even if the involved functions are convex and arbitrarily smooth, the statements in (1.12) as well as the necessary optimality conditions (1.16) or (1.17) may fail to hold. $\diamond$

Hence, to derive necessary optimality conditions of the mentioned type, additional hypotheses are required either for the whole problem or the particular minimizer $\hat{x}$ under consideration. Such hypotheses, cf. (iii),(iv) below, are usually called *regularity conditions* or *constraint qualifications.*

1.2 Basic optimality conditions; convexity and linearizations

Strong duality: Suppose for the "convex" problem 7
(i) v finite, f, g, h continuous, the affine function h maps onto Z,
(ii) there is some $x \in M$ such that $g(x) \in \operatorname{int} K$.
Then $\partial \phi(0,0) \neq \emptyset$ (strong duality holds true).
For problems 5, the point in (ii) is said to be a *Slater point*: $g_i(x) < 0 \ \forall i$.

Linear approximations
For the more general problem 6 and $f, g, h \in C^1$, assume that $\hat{x}$ is a (local) minimizer and put $f_L(x) = f(\hat{x}) + Df(\hat{x})(x - \hat{x})$ as well as $g_L(x)$ and $h_L(x)$. The linearizations define a particular convex problem P_L of type 7 (by using (iv) below).

Necessary optimality condition for problem 6: If
(iii) $Dh(\hat{x})$ maps onto Z and
(iv) some x fulfills $g_L(x) \in \operatorname{int} K$ and $h_L(x) = 0$,
then P_L satisfies the hypotheses (i), (ii) for strong duality and $\hat{x}$ solves P_L. Hence the (necessary optimality) condition (1.16) is satisfied.

MFCQ [56] for problem 4:
For problem 4, condition (iii) means *rank* $Dh(\hat{x}) = m_2$, *and (iv) attains the form:*
Some $u \in \mathbb{R}^n$ *(namely* $u = x - \hat{x}$*) fulfills* $Dh(\hat{x})u = 0$ *and* $Dg_i(\hat{x})u < 0$ *whenever* $g_i(\hat{x}) = 0$.
All together, this is the *Mangasarian-Fromovitz constraint qualification* (MFCQ).

Hence, provided MFCQ is satisfied at a local minimizer $\hat{x}$ *of a* C^1 *problem 4, there are* y, z *(Lagrange multipliers) such that* $(\hat{x}, y, z)$ *is a KKT point.*

Weaker conditions for $\hat{x} \in M$
Lagrange multipliers exist under weaker conditions at optimal $\hat{x}$. Really, one only needs:

If, for some $\hat{u} \in \mathbb{R}^n$, $\varepsilon > 0$ and certain $t_k \downarrow 0$, $k = 1, 2, ...$ it holds

$$f(\hat{x} + t_k\hat{u}) < f(\hat{x}) - t_k\varepsilon, \ \|h(\hat{x} + t_k\hat{u})\| \leq o(t_k) \text{ and } \max_i g_i(\hat{x} + t_k\hat{u}) \leq o(t_k)$$

($\hat{u}$ exists - by linear-progr. duality - iff Lagrange multipliers do not exist for $\hat{x}$),
then

$$\text{there are } x^k \in M \text{ satisying } f(x^k) < f(\hat{x}) \text{ and } x^k \to \hat{x} \tag{1.18}$$

i.e., $\hat{x}$ is not locally minimal.

Under MFCQ, the implicit function theorem applied to h ensures, with small $\delta > 0$, that (1.18) holds for certain points

$$x^k = \hat{x} + t_k(\hat{u} + \delta u) + o(t_k) \quad \text{(with new } o\text{)}.$$

The same conclusion, even directly with $\delta = 0$, is possible under so-called *calmness* (cf. 3.5) at $(0, 0, \hat{x})$ of the constraint map

$$M(y, z) = \{x \mid g(x) \leq y, \ h(x) = z\}. \tag{1.19}$$

(1) Thus MFCQ can be cancelled for affine functions h and g; put $x^k = \hat{x} + t_k\hat{u}$.

(2) If h is affine, then replace $\mathbb{R}^n$ by $h^{-1}(0)$ to weaken MFCQ (no rank condition) or put $x^k = \hat{x} + t_k(\hat{u} + \delta u)$.
(3) If h is only piecewise linear then $h^{-1}(0)$ becomes a *union* of a finite number of polyhedrons P_μ (described by affine systems $A_\mu x \le b_\mu$). The problems/values

$$v_\mu = \inf \{ \ f(x) \mid x \in P_\mu, \ g(x) \le 0 \ \}$$

fulfill $v = \min_\mu \ v_\mu$ and allow again a reduction to simpler problems with simpler optimality conditions since:

$\hat{x}$ is optimal for (1.1) iff $\hat{x}$ is optimal for each μ with $v_\mu = v$ and $\hat{x} \in P_\mu$.

So it suffices to study these simpler problems separately.
(4) Similarly, subsystems of piecewise linear g_i and h_ν can be handled by studying

$$v_\mu = \inf \{ \ f(x) \mid x \in P_\mu, \ \ h_{\nu'}(x) = 0 \ \forall \nu', \ g_{i'}(x) \le 0 \ \forall i' \ \} \tag{1.20}$$

where ν' and i' denote the functions which are not piecewise linear.

MFCQ and Aubin property
Condition $MFCQ$ ensures that the topological behavior of the map M (1.19) near $(0, 0, \hat{x})$ is locally the same as for (proper) hyperplanes $H(r) := \{x \mid \langle c, x\rangle = r\}$:

Given $x \in M(y, z)$ and (y', z') (close to $\hat{x}$ and $(0, 0)$ respectively) there exists $x' \in M(y', z')$ satisfying a Lipschitz condition $\|x' - x\| \ \le \ L \ \|(y', z') - (y, z)\|$.

MFCQ is even equivalent (for $g, h \in C^1$) to this property (called Aubin property of M, cf. 3.4). In addition, MFCQ at a (local) minimizer $\hat{x}$ is equivalent to the fact that the set of assigned Lagrange multipliers is nonempty and bounded.

For Banach space problems 6 with involved functions $g, h \in C^{0,1}$, the Aubin property of M can be written in terms of a MFCQ- like condition, too. In this case, however, the fixed direction u becomes a family of directions (which are functions) $u_x = u_x(t) \in X, \ t > 0$, where x denotes feasible points near $\hat{x}$ [49]. Having $g, h \in C^1$ then (iii) and (iv) form the analogy to MFCQ.

1.3 The standard second order condition

For the finite-dimensional C^2 problem 4, let $(\hat{x}, \hat{y}, \hat{z})$ be a KKT point. Define the index sets $I^0 = \{i \mid g_i(\hat{x}) = 0\}$, $I^+ = \{i \in I^0 \mid \hat{y}_i > 0\}$ and the tangent cone

$$U = \{u \in \mathbb{R}^n \mid Df(\hat{x})u = 0, \ Dg_i(\hat{x})u \le 0 \forall i \in I^0, \ Dg_i(\hat{x})u = 0 \forall i \in I^+, \ Dh(\hat{x})u = 0\}. \tag{1.21}$$

Suppose that

$$\langle u, \ D_x^2 L(\hat{x}, \hat{y}, \hat{z})u\rangle > c \quad \forall u \in U, \ \|u\| = 1. \tag{1.22}$$

Then, it holds for sufficiently small $\delta > 0$

$$f(x) \ge f(\hat{x}) + \tfrac{1}{2}c\|x - \hat{x}\|^2 \quad \forall x \in M \cap B(\hat{x}, \delta). \tag{1.23}$$

For problems in Banach spaces, the situation is less obvious, we refer to [3]. Notice, that the question of second order conditions for classical problems of variational calculus

$$\min \ J(y) := \int_a^b f(y, y', x)dx, \quad y(a) = A, \ y(b) = B$$

leads to the Legendre- Jacobi conditions in terms of second-order differential equations.

1.4 Modifications for vector- optimization

In all considered problems 1, ... , 7, one can ask for Pareto-optimal (also called efficient) points $\hat{x}$. This means that a finite number of objectives f_j is given, and $\hat{x} \in M$ has to satisfy:

$$\text{there is no } x \in M \text{ such that } f_j(x) \le f_j(\hat{x}) \ \forall j \text{ and } f_j(x) < f_j(\hat{x}) \text{ for some } j. \tag{1.24}$$

Let E be the set of efficient points. To generate such points one can use (in the local and global sense) that every minimizer $\hat{x}$ of f_λ on M satisfies (1.24), provided that

$$f_\lambda(x) = \sum_j \lambda_j f_j(x) \text{ and } \lambda_j > 0 \ \forall j. \tag{1.25}$$

Hence the union of all minimizer to $f_\lambda,\ \lambda > 0$ is a subset $E' \subset E$. This set can be characterized as follows (we consider only points in M): $\hat{x} \in E'$ iff there is some $\varepsilon > 0$ such that

$$\text{if } f_j(x) < f_j(\hat{x}) \text{ for some } j, \text{ then } f_r(x) \geq f_r(\hat{x}) + \varepsilon(f_j(\hat{x}) - f_j(x)) \text{ for some } r \neq j. \tag{1.26}$$

Such points are also called *properly efficient*. They are similarly reasonable for defining a "cooperative" solution and exclude efficient points like

$$f_j(x) < f_j(\hat{x}) \ \forall j \neq r \text{ and } f_r(x) - f_r(\hat{x}) = o(\sum_{j \neq r}(f_j(\hat{x}) - f_j(x)) > 0. \tag{1.27}$$

For the "smooth" finite-dimensional vector optimization problem 4 one can also show:
To every $\hat{x} \in E$ which satisfies MFCQ, there is a nontrivial $\lambda \geq 0$ such that the KKT conditions can be satified with $f = f_\lambda$.

Main tools for proving the listed statements:
Separation of convex sets; implicit function theorem in B-spaces, in particular the Lyusternik (also called Lyusternik/Graves-) theorem.

2 Other descriptions of KKT points

For analyzing points (x, y, z) which satisfy the first-oder conditions for a classical NLP in finite dimension

$$\min f(x) \quad s.t. \ \ x \in M := \{x \in \mathbb{R}^n \mid g(x) \leq 0 \in \mathbb{R}^{m_1} \ \text{ and } h(x) = 0 \ \in \mathbb{R}^{m_2}\}, \tag{2.1}$$

several reformulations of the related KKT- conditions (1.17),

$$Df(x) + \textstyle\sum_i y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x) = 0; \quad h(x) = 0,\ \ g(x) \leq 0,\ \ y \geq 0,\ \ \langle y, g(x) \rangle = 0$$

are possible. The common idea consists in appropriate descriptions of the involved complementarity conditions

$$g(x) \leq 0,\ \ y \geq 0,\ \ \langle y, g(x) \rangle = 0 \tag{2.2}$$

in form of equations or generalized equations.

2.1 NCP- functions

One well-known equivalent description of (2.2) consists in requiring

$$\sigma(y_i, -g_i(x)) := \min\{y_i, -g_i(x)\} = 0 \ \forall i$$

or more general

$$\sigma(y_i, -g_i(x)) = 0 \ \forall i$$

where $\sigma : \mathbb{R}^2 \to \mathbb{R}$ is any function satisfying $\sigma(u, v) = 0 \ \Leftrightarrow \ u \geq 0, v \geq 0, uv = 0$; a so-called NCP function which should be sufficiently simple. An often used and in many respects useful example is the so-called Fischer-Burmeister function $\sigma(u, v) = u + v - \sqrt{u^2 + v^2}$. Setting

$$\begin{array}{rcl} \Theta_1 & = & Df(x) + \sum_i y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x), \\ \Theta_{2i} & = & \sigma(y_i, -g_i(x)) \\ \Theta_3 & = & h(x), \end{array} \tag{2.3}$$

the conditions (1.17) and $\Theta(x,y,z)=0$ (where $\Theta:\mathbb{R}^\mu\to\mathbb{R}^\mu, \mu=n+m_1+m_2$) are equivalent. The equation $\Theta(x,y,z)=(a,b,c)^T$ is connected with the canonically perturbed problem

$$P(a,b,c):\quad \min_{x\in G(b,c)} f(x)-\langle a,x\rangle \ \text{ where } G(b,c) \ = \ \{x \mid g(x)\le b,\ h(x)=c\} \tag{2.4}$$

but the transformations of the related solutions are more complicated than for the subsequent reformulations. Also a similar product representation as below is not true for Θ. This shrinks the value of Θ for stability investigations, but not in view of solution methods. For more details we refer to [14, 15, 36, 78].

2.2 Kojima's function

Similarly, system (1.17) can be written in terms of Kojima's [39] function $\Phi:\mathbb{R}^\mu\to\mathbb{R}^\mu$ which has the components

$$\begin{array}{lllll} \Phi_1 & = & Df(x) + \sum_i y_i^+ Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x), & & y_i^+=\max\{0,y_i\},\\ \Phi_{2i} & = & g_i(x) - y_i^-, & & y_i^-=\min\{0,y_i\},\\ \Phi_3 & = & h(x). & & \end{array} \tag{2.5}$$

Then the zeros of Φ are related to the KKT- points via the transformations

$$\begin{array}{lll} (x,y.z)\in\Phi^{-1}(0) & \Rightarrow & (x,u,z)=(x,y+g(x),z) \text{ is a KKT-point}\\ (x,u,z) \text{ is a KKT-point} & \Rightarrow & (x,y,z)=(x,u+g(x),z)\in\Phi^{-1}(0) \end{array} \tag{2.6}$$

and Φ is, for smooth f,g,h, one of the simplest nonsmooth function. Moreover, Φ can be written as a (separable) product $\Phi(x,y,z)=\mathcal{M}(x)N(y,z)$ where

$$N=(1,\ \ y_1^+,...,y_{m1}^+,\ \ y_1^-,...,y_{m1}^-,\ z)^T\in\mathbb{R}^{1+2m_1+m_2},$$

$$\mathcal{M}(x)=\begin{pmatrix} Df(x) & Dg_1(x)... & Dg_{m1}(x) & 0... & 0... & 0 & Dh_1(x)... & Dh_{m2}(x)\\ g_i(x) & 0\ \ ... & 0 & 0... & -1... & 0 & 0\ \ ... & 0\\ h(x) & 0\ \ ... & 0 & 0... & 0... & 0 & 0\ \ ... & 0 \end{pmatrix} \tag{2.7}$$

with $i=1,...,m_1$ and -1 at position i in the related block.

Writing, e.g., $(y_i^3)^+,(y_i^3)^-$ at the place of y_i^+ and y_i^-, KKT-points are even zeros of a smooth function. However, now $y_i=0$ leads to a zero-column in the Jacobian of this modified function Φ. So the standard tools for computing a zero or analyzing critical points via implicit functions fail again and the assignment (2.6) is no longer loc. Lipsch. in both directions for $g\in C^1$.

Using Φ (2.5), the points of interest are zeros of a (continuous, piecewise smooth) $\mathbb{R}^\mu\to\mathbb{R}^\mu$ function, and the equation

$$\Phi(x,y,z)=(a,b,c)^T \tag{2.8}$$

permits a canonical interpretation: It describes by (2.6) the KKT-points of the elementary (canonically) perturbed problem

$$P(a,b,c):\qquad \inf\ \{\ f(x)-\langle a,x\rangle \mid g(x)\le b \text{ and } h(x)=c\ \}. \tag{2.9}$$

For fixed a, this corresponds to the problems which already appeared for defining ϕ (1.9).
Due to the product structure of Φ and the simple type of non-differentiability, several generalized derivatives (see below) can be really determined for $f,g,h\in C^{1,1}$.
For $f,g,h\in C^2$, nonsmoothness is only implied by the components

$$\mu(y_i)=(y_i^+,\ y_i^-)=(y_i^+,\ y_i-y_i^+)=\tfrac{1}{2}\ (y_i+|y_i|,\ y_i-|y_i|), \tag{2.10}$$

of N. So, discussions on generalized derivatives can be reduced to defining a "derivative" of the *absolute value* at the origin.

2.3 Generalized equations

There is another, quite popular possibility of describing KKT-points, namely as solutions of inclusions. The simplest one is the system

$$\begin{array}{lcl} Df(x) + \sum_i y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x) & = 0 & \\ g(x) & \in & N_K(y), \\ h(x) & = 0, & \end{array} \tag{2.11}$$

where $N_K(y)$ denotes, for $K = \mathbb{R}^{m+}$ and $y \in K$, the normal cone of K at y. Then

$$N_K(y) = \{y^* | \langle y^*, k - y\rangle \leq 0 \ \forall k \in K\} = \{y^* |\ y_i^* = 0 \text{ if } y_i > 0\ ;\ y_i^* \leq 0 \text{ if } y_i = 0\}. \tag{2.12}$$

If $y \notin K$ put $N_K(y) = \emptyset$. Defining $\hat{K} = \mathbb{R}^n \times K \times \mathbb{R}^{m2}$ and similarly $N_{\hat{K}}(x, y, z)$, system (2.11) can be written, with left-hand side H and $s = (x, y, z)$, as a *generalized equation*

$$H(s) \in N_{\hat{K}}(s), \tag{2.13}$$

where H is a function and $N_{\hat{K}}$ a multivalued mapping (multifunction). Such systems have been introduced by S. Robinson who noticed (during the 70th) that the relations between system (2.13) and its linearization

$$H(\hat{s}) + DH(\hat{s})(s - \hat{s}) \in N_{\hat{K}}(s), \tag{2.14}$$

are (locally, and in view of inverse and implicit functions) the same as for usual equations. This was the starting point for various investigation of generalized equations (2.13) in different spaces and with arbitrary multifunctions $\mathcal{N}$ (based on the same story). Again, the solutions of the perturbed system

$$H(s) \in (a, b, c)^T + N_{\hat{K}}(s) \tag{2.15}$$

describe the KKT points of problem (2.9), the same perturbation in (2.14) describe the KKT points of a related quadratic problem. In general, the standard hypothesis of the inverse function theorem, "$DH(\hat{s})$ is regular" now attains the form: "The solutions of the *perturbed linearized problem* are locally unique and Lipschitz".

In terms of Kojima's function Φ, system (2.14) corresponds to linearization of $\mathcal{M}$ while N remains unchanged

$$[\mathcal{M}(\hat{x}) + D\mathcal{M}(\hat{x})(x - \hat{x})]N(y, z) = 0. \tag{2.16}$$

Additional approximations of N (being less obvious in model (2.13)) can be applied for solution methods. Needless to say that neither the stability theory of Kojima functions nor the one of generalized equations makes explicitly use of the particular structure of $\mathcal{M}(x)$ or $H(s)$.

3 Multiphase problems and stability

A deeper analysis of critical points in optimization problems is mainly required for hierarchic optimization models which arise as "multiphase problems" if solutions of some or several problems, say of $P(a, b, c)$ in (2.9), are involved in a next one, e.g.,

$$\inf_{a,b,c,x} \ F(a, b, c, x) \ \text{ where } x = x(a, b, c) \text{ is a (local) solution to } P(a, b, c). \tag{3.1}$$

Here, also further conditions can be required for a, b, c, x or for some other involved parameter p in $P(a, b, c)$. For various more concrete models and related solution methods we refer to [13] and [62]. Continuity results for optimal values and related solutions (along with instructive counterexamples) in view of $\mathbb{R}^n$- problems can be found in [2].

Even if $x = x(.)$ is unique and continuous then, as a rule, the solution map has kinks (and the optimal value $v = f(x(.))$ is not C^2) whenever x changes the "faces" of the parametric feasible set. In terms of the Kojima equation $\Phi(x, y, z) = (a, b, c)^T$, then certain y_i change the sign.

Obviously, these difficulties are less hard if local solutions move on smooth manifolds defined by a regular (rank $Dh = \dim c$) system $h(x) = c$ only, though minimizer may be transformed into saddle points or maximizer, depending on the Hessian on the tangent space. Hence inequality constraints or non-regular equations are the main reason for various difficulties.

The behavior of stationary solutions in $\mathbb{R}^n$ has been already precisely described by characterizing the possible singularities if the involved functions are of type C^3 and the whole problem belongs to some generic class, cf. [27, 28]. Here, we study less smooth problems, in general.

The assumption $f, g \in C^1$ with loc. Lipsch. derivatives (i.e. $f, g \in C^{1,1}$), but $f, g \notin C^2$ is typical for problems which involve optimal-value functions of other (sufficiently regular) optimization models like in design- or semi-infinite optimization or for multi-level problems.

Stability: In order to analyze and solve problems like (3.1), one is mostly interested in some kinds of Lipschitz-continuity (summarized as *stability*) of the solution map S, assigned to (2.9) or more general problems. Here, "solution" may be taken in the local and global sense, and often it also denotes the couple of points satisfying the first order necessary conditions.

To have a sufficiently general model that covers all these variants in view of optimization and related fields, let us consider an inclusion of the form (2.11)

$$p \in F(x) := h(x) + \mathcal{N}(x); \qquad h : X \to P,\ \mathcal{N} : X \rightrightarrows P; \quad \text{(B- spaces)} \tag{3.2}$$

with a multifunction $\mathcal{N}$, element-wise sum for $h + \mathcal{N}$ and the solution set $S(p) = F^{-1}(p)$. In this form, hierarchic or multilevel problems may be equations, variational inequalities (i.e. $\mathcal{N}(x)$ is some normal cone of a given set M at a point $x \in M$), games, control problems et cet., too. In the sequel, suppose that $\operatorname{gph} S := \{(p, x) \mid x \in S(p)\}$ is a closed set. The set $\operatorname{dom} S = \{p \mid S(p) \neq \emptyset\}$ is the (effective) domain of S.

Notions of (local) Lipschitz stability.
Let $S : P \rightrightarrows X$ be a (closed) multifunction and $z^0 = (p^0, x^0) \in \operatorname{gph} S$. We write ζ^0 in place of (x^0, p^0) and say that some property holds *near* x if it holds for all points in some neighborhood of x. Further, let B denote the closed unit ball in the related space and

$$S_\varepsilon(p) := S(p) \cap (x^0 + \varepsilon B) := S(p) \cap \{x \mid \|x - x^0\| \le \varepsilon\}.$$

The following definitions generalize typical properties of the multivalued inverse $S = f^{-1}$ or of level sets $S(p) = \{x \mid f(x) \le p\}$ for functions $f : M \subset X \to \mathbb{R}$ at the origin. After each definition, we add an example such that S obeys the claimed property, but (if possible) not the remaining ones.
Definitions:

$$\begin{array}{l} S \text{ is said to be } \textit{strongly Lipschitz} \text{ at } z^0 \text{ if} \\ \exists \varepsilon, \delta, L\ (>0) \text{ such that } d(x', x) \le L\|p' - p\| \text{ for all} \\ p, p' \in (p^0 + \delta B) \cap \operatorname{dom} S_\varepsilon,\ x' \in S_\varepsilon(p') \text{ and } x \in S_\varepsilon(p), \end{array} \tag{3.3}$$

i.e., S_ε is locally single-valued and Lipschitz on $\operatorname{dom} S_\varepsilon(p)$ near z^0; $S_\varepsilon(p) = \emptyset$ is allowed. ($M = \mathbb{R}^+, X = \mathbb{R}, f(x) = x, S = f^{-1}$).
If also $p^0 \in \operatorname{int}\operatorname{dom} S_\varepsilon$ is required then S_ε is locally a Lipschitz function, and we call S *strongly Lipschitz stable* (s.L.s.) at z^0. ($M = X = \mathbb{R}, f(x) = 2x - |x|, S = f^{-1}$).

$$\begin{array}{l} S \text{ is said to be } \textit{pseudo–Lipschitz} \text{ at } z^0 \text{ if } \exists \varepsilon, \delta, L\ (>0) \\ \text{such that } S_\varepsilon(p) \subset S(p') + L\|p' - p\| B\ \forall p, p' \in p^0 + \delta B. \end{array} \tag{3.4}$$

($M = X = \mathbb{R}^2, f(x, y) = x + y, S = f^{-1}$).
Other notations (or equivalent notions) for the same fact are: S^{-1} is metrically regular resp. pseudo–regular or S has the *Aubin property* [76].

Setting $p = p^0$ in (3.4), one obtains $S(p') \neq \emptyset$ due to $x^0 \in S_\varepsilon(p)$. Thus $p^0 \in \operatorname{int}\operatorname{dom} S_\varepsilon$ is always ensured under (3.4).

$$\begin{array}{l} S \text{ is said to be } \textit{calm} \text{ at } z^0 \text{ if (3.4) holds for } p' = p^0, \quad \text{i.e.,} \\ \exists \varepsilon, \delta, L\ (>0) \text{ such that } S_\varepsilon(p) \subset S(p^0) + L\|p - p^0\| B\ \forall p \in p^0 + \delta B. \end{array} \tag{3.5}$$

($M = X = \mathbb{R}, f(x) \equiv 0, S = f^{-1}$).

$$\begin{array}{l} S \text{ is said to be } \textit{upper Lipschitz} \text{ at } z^0 \text{ if } \exists \varepsilon, \delta, L\ (>0) \\ \text{such that } S_\varepsilon(p) \subset x^0 + L\|p - p^0\| B\ \ \forall p \in p^0 + \delta B. \end{array} \tag{3.6}$$

($M = X = \mathbb{R}, f(x) = |x|, S = f^{-1}$).
If, in addition, $p^0 \in \operatorname{int} \operatorname{dom} S_\varepsilon$ we call S *upper Lipschitz stable (u.L.s.)* at z^0. ($M = X = \mathbb{R}, f$ constant on $[\frac{1}{2^{k+1}}, \frac{1}{2^k}], k > 0$ integer, and $Df(x) \equiv 1$ otherwise; $S = f^{-1}$). Finally,

$$\begin{array}{l} S \text{ is said to be } \textit{lower Lipschitz} \text{ at } z^0 \text{ if } \exists \delta, L\ (> 0) \\ \text{such that } S(p) \cap (x^0 + L\|p - p^0\|B) \neq \emptyset\ \forall p \in p^0 + \delta B. \end{array} \tag{3.7}$$

($M = X = \mathbb{R}, f(x) = x$ if $x \leq 0$, $f(x) = x^2$ if $x \geq 0$, $S(p) = \{x | f(x) \leq p\}$.

Comments.
In some of these definitions, one may put $\varepsilon = \delta$. We used different constants for different spaces. If $S = f^{-1}$ is the inverse of a C^1 function $f : \mathbb{R}^n \to \mathbb{R}^n$ with $S(p^0) = \{x^0\}$, all these properties coincide with $\det Df(x^0) \neq 0$. If f is only loc. Lipsch., they are quite different.

The constant L is called a *rank* of the related stability. The requirement $p^0 \in \operatorname{int} \operatorname{dom} S_\varepsilon$ means that solutions to $p \in F(x)$ are (locally) *persistent*, and the lower Lipschitz property quantifies this persistence in a Lipschitzian manner.

The notions concerning stability or regularity differ in the literature. So "s.L.s." and "strongly regular" mean often the same, and our "upper Lipschitz" is "locally upper Lipschitz" in [8] while "u.L.s." is "upper regular" in [36]. Further, "regularity" of multifunctions has been also defined in an alternative manner via local linearizations in [69].

Remark 1 For fixed $z^0 \in \operatorname{gph} S$, one easily sees by the definitions:
(i) S is u.L.s. iff S is both upper and lower Lipschitz.
(ii) S is calm if S is upper Lipschitz.
(iii) S is upper Lipschitz iff S is calm and x^0 is isolated in $S(p^0)$.
(iv) S is s.L.s. iff S is pseudo-Lipschitz and $\operatorname{card} S_\varepsilon(p) \leq 1$ for p near p^0.
(v) S is pseudo-Lipschitz iff S is lower Lipschitz at all points $z \in \operatorname{gph} S$ near z^0 with fixed constants ε, δ and L.
(vi) S is pseudo-Lipschitz iff S is both calm at all $z \in \operatorname{gph} S$ near z^0 with fixed constants ε, δ, L and lower Lipschitz at z^0. ◇

Composed mappings and intersections
(i) Often, $S(p) = U(V(p))$ is a composed map where $V : P \rightrightarrows Y$ and $U : Y \rightrightarrows X$. In many situations, then the related Lipschitz properties at z^0, where now $y^0 \in V(p^0)$ and $x^0 \in U(y^0)$, are consequences of the related properties for V and U at the corresponding points. One has, however, to shrink the image of V to the $\varepsilon-$ nbhds of y^0 which appear in the related definitions, i.e., one has to study composed maps of the form

$$S(p) = U(\ V(p) \cap (y^0 + \varepsilon B)\ ).$$

Further, local solvability plays an important role in this context; for details we refer to [35], Lemma 1.2.

(ii) If $S = \Gamma^{-1}$ with $\Gamma(x) := G(F(x))$ where $F : X \rightrightarrows Y$, $G : Y \rightrightarrows P$ then $x \in S(p) \Leftrightarrow p \in G(F(x))$ $\Leftrightarrow x \in F^{-1}(G^{-1}(p))$. This is situation (i) with $U = F^{-1}$ and $V = G^{-1}$.

(iii) The map S under (ii) can be written by intersections and projections: Define $g : (X, Y) \rightrightarrows P$ as $g(x, y) = G(y)$ and $f : (X, Y) \rightrightarrows Y$ as $y \in f(x', y')$ iff $y = y' \in F(x')$ (otherwise $f(x', y') = \emptyset$). This yields $g^{-1}(p) = (X, G^{-1}(p))$, $f^{-1}(y) = (F^{-1}(y), \{y\})$ and

$$x' \in S(p) \Leftrightarrow (x', y') \in f^{-1}(y') \cap g^{-1}(p) \text{ for some } y'.$$

(iv) Intersections $S(p_1, p_2) = F(p_1) \cap G(p_2)$ where $F : P_1 \rightrightarrows X$, $G : P_2 \rightrightarrows X$. This is the typical situation for studying solutions x of (in)equality systems: $F(p_1) = \{x | f(x) = p_1\}$ and $G(p_2) = \{x | g(x) \in p_2 + K\}$, $K \subset P_2$.

If f and g are loc. Lipsch. then calmness of F and G at (p_1^0, x^0) and (p_2^0, x^0), respectively, ensures calmness of S at (p_1^0, p_2^0, x^0), provided that one of the mappings

$$S_1(p_1) = \{x | f(x) = p_1,\ g(x) \in p_2^0 + K\},$$

$$S_2(p_2) = \{x | f(x) = p_1^0,\ g(x) \in p_2 + K\}$$

is calm at the corresponding point, cf. [35], Thm. 3.6. The statement is helpful, e.g., for discussing calmness in case 4, section 1.2, MFCQ.

For calmness, the upper Lipschitz and the Aubin property of such intersections and their interrelations we refer to [35] and [49]. For upper Lipschitz properties of composed maps we refer to [53], [54], [52], [64], [65] and [36]. In what follows we do not exploit special structures as above.

In the current literature on generalized equations, variational conditions and related fields, it seems to be standard to reformulate the listed stabilities in terms of certain generalized derivatives. The results are statements which look, after replacing the notion of the derivative, similar to corresponding inverse and implicit function theorems for smooth functions. However, unlike the smooth case, methods of computing these derivatives in terms of original data do often not exist. This motivates why other characterizations are desirable, in particular characterizations via (slightly simpler) Lipschitz functions or directly via the main applications of stability statements, the behavior of solutions methods.

4 Stability in terms of Lipschitz functions

Next we show that, though we are speaking on *multifunctions*, the required stability properties are classical properties of *non-expansive, real-valued functions* only.
In many publication, the (improper) function $\phi(x,p) = \text{dist}\,(x, S(p))$ has been used to describe Lipschitz behavior of S. There is, however, a second function important for studying S, namely

$$\psi(x,p) = \text{dist}\,((p,x), \text{gph}\, S) \quad (\leq \phi(x,p)).$$

Unlike ϕ, the function ψ is well-defined and non-expansive whenever $\text{gph}\, S \neq \emptyset$.

Calmness: In terms of ψ, *calmness* of S at $(p',x') \in \text{gph}\, S$ means that

$$\exists \varepsilon > 0, \lambda > 0 \text{ such that } \psi(x,p') \geq \lambda\ \text{dist}\,(x, S(p'))\ \ \forall x \in X'_\varepsilon := x' + \varepsilon B. \tag{4.1}$$

Details, direct applications of ψ for penalty approaches and duality, estimates of ψ for particular systems and other consequences can be found in [35]. Condition (4.1) requires that $\psi(.,p')$ increases in a Lipschitzian way if x moves away from $S(p')$.

Aubin property at z^0: From (4.1) and Remark 1(vi), one obtains

$$\begin{array}{l} S \text{ is pseudo-Lipschitz at } z^0 \text{ iff it is lower Lipschitz at } z^0 \text{ and (4.1)} \\ \text{holds true for all } (p',x') \in \text{gph}\, S \text{ near } z^0 \text{ with the same } \varepsilon \text{ and } \lambda. \end{array} \tag{4.2}$$

Remark 2 Hence calmness is a monotonicity property with respect to two canonically assigned Lipschitz functions. The first one is the distance to $\text{gph}\, S$, the second one is the distance to the image set at the crucial point. The same for points near z^0 combined with the lower Lipschitz property at z^0 characterizes the Aubin property. ◇

Strong Lipschitz stability of S at z^0 is the Aubin property along with $\text{card}\, S_\varepsilon(p) \equiv 1$. The latter means that $\psi(.,p)$ is locally injective on $\psi^{-1}(0)$:

$$\psi(x,p) = 0 = \psi(x',p)\ \Rightarrow\ x = x'\ \ \forall x, x' \in x^0 + \varepsilon B \ \text{ and } p \text{ near } p^0. \tag{4.3}$$

The upper Lipschitz property at z^0 requires equivalently that S is calm and x^0 is isolated in $S(p^0)$. This can be summarized by

$$\exists \varepsilon > 0, \lambda > 0 \text{ such that } \psi(x,p^0) \geq \lambda\ d(x,x^0)\ \ \forall x \in x^0 + \varepsilon B. \tag{4.4}$$

In other words, x^0 has to be a *minimizer of order 1* for $\psi(.,p^0)$.

Solutions of penalty problems.
The function ψ can be applied for both *characterizing optimality and computing* solutions in optimization models via penalization. In fact, let S be calm at (p', x') and $X'_\varepsilon := x' + \varepsilon B$. Further, suppose:

$$\begin{array}{l} f \text{ is a Lipschitz function with rank } L_f \text{ on some nbhd of} \\ S(p') \cap X'_\varepsilon \quad \text{and } x' \text{ is a minimizer of } \ f \text{ on } S(p') \cap X'_\varepsilon. \end{array} \tag{4.5}$$

Then, x' is a (free) local minimizer of $f(x) + k\ \psi(x, p')$ for large k (this is well-known and can be immmediately seen). To find x' or some $x^* \in S(p') \cap X'_\varepsilon$ with $f(x^*) = f(x')$ by means of a penalty method, several techniques can be applied. In particular, one may replace ψ by ψ^s $(s > 0)$, in order to solve

$$\text{minimize} \quad Q_k(x) := f(x) + k\ \psi(x, p')^s \quad \text{on } X'_\varepsilon \quad (k \to \infty). \tag{4.6}$$

In this respect, it is important that, if $\dim P + \dim X < \infty$, the function ψ^2 is *semismooth*, a useful property for various solution methods based on Newton techniques, cf. [57] or [36], chapter 6. We give a short proof for the convergence of the minimizers to (4.6) in order to discuss the role of s and the possibility of removing the constraint $d(x, x') \le \varepsilon$.
Convergence of minimizers x^k for given p':
Let $x^k \in \operatorname{argmin} Q_k$ and $u^k \in S(p')$ satisfy $d(x^k, u^k) = \operatorname{dist}(x^k, S(p'))$. Then

$$\begin{array}{rl} f(x') & \ge f(x^k) + k\psi(x^k, p')^s \\ & \ge f(u^k) - L_f d(x^k, u^k) + k\psi(x^k, p')^s \\ & \ge f(x') \ - L_f d(x^k, u^k) + k\psi(x^k, p')^s. \end{array} \tag{4.7}$$

Hence $L_f d(x^k, u^k) \ge k\psi(x^k, p')^s$. Now calmness (4.1) permits to continue

$$L_f d(x^k, u^k) \ge k\lambda^s d(x^k, u^k)^s. \tag{4.8}$$

For $s > 1$, this ensures

$$\operatorname{dist}(x^k, S(p'))^{s-1} = d(x^k, u^k)^{s-1} \ \le \ \frac{L_f}{k\lambda^s} \to 0 \quad \text{and} \quad \psi(x^k, p') \to 0. \tag{4.9}$$

Next let $0 < s \le 1$. Now $0 < d(x^k, u^k) \le 1$ implies $d(x^k, u^k) \le d(x^k, u^k)^s$. Thus $k\lambda^s > L_f$ ensures $x^k = u^k$. On the other hand, if $d(x^k, u^k) \ge 1$ then $\varepsilon \ge \|x^k - u^k\| \ge \|x^k - u^k\|^s \ge 1$. Hence this case cannot appear whenever k is large enough such that $\varepsilon L_f < k\lambda^s$. In consequence, $x^k = u^k$ holds again for sufficiently large k. Summarizing, so every cluster point x^* of x^k is feasible and satisfies $f(x^*) = f(x')$, independent of the choice of $s > 0$.
Local and global minimizers:
If x' was even the *unique* global minimizer of f on $S(p') \cap X'_\varepsilon$ (such points are called strict local minimizers) then (4.9) and $f(x^*) = f(x')$ imply $x^* = x'$ and $x^k \to x'$. Hence $x^k \in \operatorname{int} X_\varepsilon$ holds for large k, and implies, as typical for penalty methods, that the related x^k are *free local minimizer of* $Q_k(x)$. □

Deleting calmness:
Looking once more at the above estimates, one sees that calmness of S at p' can be immediately replaced, in the present penalty context, by the weaker (Hoelder) property

$$\exists \varepsilon > 0, \lambda > 0, r > 0 \text{ such that } \psi(x, p') \ge \lambda \operatorname{dist}(x, S(p'))^r \quad \forall x \in X'_\varepsilon := x' + \varepsilon B. \tag{4.10}$$

Then only the critical value $s^* = 1$ from above changes: $s^* = 1/r$. Of course, now Q_k (4.6) is not differentiable (like under calmness for $s = 1$) and even more, existing derivatives may be unbounded. So they must be made bounded artificially if they are "too large". However, this is standard in every numerical program, it can be avoided by several smoothing techniques and becomes necessary only if x^k is already "almost" feasible.

5 Calmness and Aubin property for strongly closed maps

We proceed with negating the Aubin property of a (closed) map $S = F^{-1} : P \rightrightarrows X$. The map S is not pseudo-Lipschitz with rank L at $z^0 = (p^0, x^0)$ iff

$$\begin{array}{l}\exists\ (p^k,x^k)\to z^0 \text{ and } \pi_k\to p^0 \text{ such that } (p^k,x^k)\in \operatorname{gph} S\\ \text{and } \operatorname{dist}(x^k,S(\pi_k))>L\|\pi_k-p^k\|>0\ (\forall k>0,k\to\infty).\end{array}\tag{5.1}$$

The inequality in (5.1) allows $S(\pi_k)=\emptyset$ and involves interesting particular cases.

case 1: With $(p^k,x^k)\equiv(p^0,x^0)$ (5.1) is the negation of S to be lower Lipschitz with rank L.

case 2: With $\pi_k\equiv p^0$, (5.1) is the negation of S to be calm with rank L.

Below, we shall need a further distance function, namely $\phi_F(x)=\operatorname{dist}(\pi,F(x))$ for given $\pi\in P$. We call S *strongly closed* if ϕ_F is l.s.c. and some $p\in F(x)$ realizes $\operatorname{dist}(\pi,F(x))$ whenever $F(x)\neq\emptyset$ and $\pi\in P$.

Since $\operatorname{gph} F$ is closed by assumption, also $F(x)$ is closed. Hence S is strongly closed if F is locally compact or $\dim P<\infty$. Further, by the projection theorem, S is strongly closed if P is a Hilbert space and all $F(x)$ are closed and convex. For B-spaces P, our requirement is clearly a strong restriction to F, but notice that even for continuous $F:X\to P$ (then S is trivially strongly closed) the Lipschitz behavior of F^{-1} is nowhere completely characterized.

5.1 Refinements via Ekeland's principle for strongly closed maps

Let $f:X\to\mathbb{R}\cup\{+\infty\}$. We say that $z\in X$ is a (local) $\varepsilon-$ Ekeland point of f if $f(z)$ is finite and $f(x)+\varepsilon\, d(x,z)\geq f(z)\quad\forall x\in X$ ($\forall x$ near z). Recall

Ekeland's variational principle [11]: Let X be a complete metric space, $f:X\to\mathbb{R}\cup\{+\infty\}$ be lower semi-continuous and $v:=\inf_X f>-\infty$. Then, given $\hat{x}$ with $f(\hat{x})\leq v+\varepsilon$ and $\alpha>0$, there is an $\frac{\varepsilon}{\alpha}-$ Ekeland point z of f such that $f(z)\leq f(\hat{x})$ and $d(z,\hat{x})\leq\alpha$. $\diamond$

For strongly closed S, all (p^k,x^k) in (5.1) can be replaced by "better" pairs (p^k_E,x^k_E) via Ekeland's principle. This was a basic tool in [1] and in various other papers dealing with the Aubin property. The following particular replacement has been used in [36, 49] and is our key for deriving all subsequent conditions in an intrinsic manner. The proof is added since it demonstrates a typical application of Ekeland's principle.

Lemma 1 *Let S be strongly closed.*
(i) Condition (5.1) implies, with $\lambda=L$ and new points $(p^k,x^k)=(p^k_E,x^k_E)$:

$$\begin{array}{l}\exists\ (p^k,x^k)\to z^0 \textit{ in } \operatorname{gph} S \textit{ and } \pi_k\to p^0 \textit{ such that } p^k\neq\pi_k\ (\forall k>0)\\ \textit{and } (p^k,x^k) \textit{ minimizes } H_\lambda(p,x):=\|p-\pi_k\|+\frac{1}{\lambda}d(x,x^k) \textit{ on } \operatorname{gph} S.\end{array}\tag{5.2}$$

(ii) Condition (5.2) implies (5.1) for each $L\in(0,\lambda)$.
(iii) In the same way, only with $\pi_k\equiv p^0\ \forall k>0$, calmness can be characterized.
(iv) For all $t\in(0,1]$ and $p_t=p^k_E+t(\pi_k-p^k_E)$, it holds $\operatorname{dist}(x^k_E,S(p_t)\geq\lambda\|p_t-p^k_E\|$. $\diamond$

Notes: By (iv), even continuous parameter changes on a fixed line show that every lower Lipschitz rank L of S at (p^k_E,x^k_E) fulfills $L\geq\lambda$. With $v^k=p^k-\pi_k$ $(\neq 0)$, the minimum condition in (5.2) means

$$\|v^k+\eta\|-\|v^k\|\geq-\frac{1}{\lambda}\|\xi\|\quad\text{ whenever } (p^k+\eta,x^k+\xi)\in\operatorname{gph} S.\tag{5.3}$$

Proof. (i) Let (5.1) be true. The current part of the proof is the same for fixed $\pi_k\equiv p^0$ and $\pi_k\to p^0$, respectively. It remains also valid for $(p^k,x^k)\equiv(p^0,x^0)$ in (5.1), which corresponds to the negation of being lower Lipschitz with rank L.

For fixed k, define the l.s.c. function $\phi(x)=\operatorname{dist}(\pi_k,F(x))$, put $\varepsilon_k=\|\pi_k-p^k\|$ and note that $p^k\in F(x^k$ yields

$$0\leq\inf_x\phi(x)\leq\phi(x^k)=\operatorname{dist}(\pi_k,F(x^k))\leq\varepsilon_k.$$

Setting $\alpha_k=L\varepsilon_k$, we have $\frac{\varepsilon_k}{\alpha_k}=\frac{1}{L}$, and by Ekeland's principle some x^k_E fulfills

$$\phi(x)+\frac{1}{L}d(x,x^k_E)\geq\phi(x^k_E)\ \forall x\in X,\quad d(x^k_E,x^k)\leq\alpha_k\ \text{ and }\ \phi(x^k_E)\leq\phi(x^k).\tag{5.4}$$

Explicitly, the main condition requires

$$\operatorname{dist}(\pi_k,F(x))+\frac{1}{L}d(x,x^k_E)\geq\operatorname{dist}(\pi_k,F(x^k_E))\ \forall x\in X.\tag{5.5}$$

Since $F(x^k_E) \neq \emptyset$ and S is strongly closed, some $p^k_E \in F(x^k_E)$ fulfills $\text{dist}\,(\pi_k, F(x^k_E)) = \|\pi_k - p^k_E\|$. Obviously, $(p^k_E, x^k_E) \to z^0$ as $k \to \infty$.
If $p^k_E = \pi_k$ then a contradiction follows from $x^k_E \in S(p^k_E)$ and (5.1):

$$\alpha_k \geq d(x^k, x^k_E) \geq \text{dist}\,(x^k, S(p^k_E)) = \text{dist}\,(x^k, S(\pi_k)) \; > \; L\|\pi_k - p^k\| = L\varepsilon_k.$$

Hence we have $p^k_E \neq \pi_k$ whereafter (5.5) yields (5.2 with the new points $(p^k, x^k) = (p^k_E, x^k_E)$.
(ii) Conversely, (5.2) implies, after setting there $p = \pi_k$,

$$\frac{1}{\lambda} d(x, x^k) \geq \|p^k - \pi_k\| \quad \forall x \in S(\pi_k),$$

hence $\text{dist}\,(x^k, S(\pi_k)) \geq \lambda\|p^k - \pi_k\| > 0$. So (5.2) yields (5.1) for every positive $L < \lambda$.

(iv) For $x \in S(p_t)$, it follows $\|p_t - \pi_k\| + \frac{1}{\lambda} d(x, x^k_E) \geq \|p^k_E - \pi_k\|$. Due to the special choice of p_t, this is $\frac{1}{\lambda} d(x, x^k_E) \geq \|p^k_E - \pi_k\| - \|p_t - \pi_k\| = \|p_t - p^k_E\|$. □

Condition (5.2) permits a characterization of calmness and the Aubin property by the simpler lower Lipschitz property. It also permits a formulation of Lemma 1 in terms of (3.4), where now (p, x) takes the place of (p^k_E, x^k_E).

Theorem 1 *For strongly closed S, the following statements are equivalent:*
(i) S is pseudo-Lipschitz at $z^0 = (p^0, x^0)$
(ii) $\exists L > 0$ and $\varepsilon > 0$ such that

$$\begin{array}{l} \textit{for all } x \in S_\varepsilon(p) \;\; \textit{and} \;\; \pi, p \in p^0 + \varepsilon B, \; \pi \neq p, \; \textit{it holds} \\ L\|p' - \pi\| + \text{dist}\,(x, \; S(p')) < L\|p - \pi\| \;\; \textit{for some } p' \in P. \end{array} \tag{5.6}$$

(iii) S is lower Lipschitz at all $(p, x) \in \text{gph}\, S$ near z^0 with uniform rank λ.
In addition, S is calm at z^0 iff (5.6) holds for $\pi \equiv p^0$. ◇

Proof. (i) ⇔ (ii) Clearly, S is pseudo-Lipschitz iff (5.2) cannot hold for some (large) λ. The latter is (by formal negation) equivalent to condition (5.6). With respect to calmness, the same remains true after setting $\pi = p^0$.
(iii) ⇒ (i) We prove that (ii) is valid if (iii) holds with some $\lambda < L$. Indeed, setting $p' = p + t(\pi - p)$, there exists, for small $t > 0$, some $x' \in S(p')$ satisfying $d(x', x) \leq \lambda\|p' - p\| = \lambda t(\pi - p)$. So we obtain

$$\begin{aligned} \text{dist}\,(x, S(p')) + L\|p' - \pi\| \; &\leq t\|\pi - p\|\lambda + L(1-t)\|p - \pi\| \\ &\leq (\; t\lambda + L(1-t) \;)\; \|p - \pi\| < L\|\pi - p\|. \end{aligned}$$

(i) ⇒ (iii) Since this is evident, nothing remains to prove. □

Note that the size of δ, included in the lower Lipschitz definition, may depend on the point $(p, \; x)$ now, and that the original definitions (3.4) and (3.5) claim (5.6) stronger with $p' = \pi$ and $p' = p^0$, respectively (up to arbitrarily small changes of the rank L).

5.2 The Aubin property via weakly stationary points

Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be any function and $f(x)$ be finite. We call x stationary if there are $\varepsilon_k \downarrow 0$ such that x is a local ε_k- Ekeland point of f. Clearly, for differentiable f, this means $Df(x) = 0$. Further, we call x weakly stationary if there are $\varepsilon_k \downarrow 0$ and $x^k \to x$ such that x^k is a local ε_k- Ekeland point of f.

Theorem 2 *Let $\dim X + \dim P < \infty$ and $F \to P$ be a loc. Lipsch. function. Then $S = F^{-1}$ is pseudo Lipschitz at (p^0, x^0) iff there is no $y^* \neq 0$ such that x^0 is weakly stationary for f as $f(x) = \langle y^*, F(x)\rangle$.* ◇

Proof. Let S be not pseudo Lipschitz at (p^0, x^0). We first verify formula (5.8) by supposing only that F is a closed multifunction.

With every fixed λ and $v^k = p^k - \pi_k$, condition (5.3) holds true. Setting $y_k^* = v^k/\|v^k\|$ and using Euclidean norms, (5.3) implies

$$\|v^k + \eta\|^2 \geq (\|v^k\| - \frac{1}{\lambda}\|\xi\|)^2 \quad \text{if } \|\xi\| < \lambda\|v^k\| \text{ and } (p^k + \eta, x^k + \xi) \in \operatorname{gph} S$$

and after division by $2\|v^k\|$,

$$\frac{\|\eta\|^2}{2\|v^k\|} + \langle y_k^*, \eta\rangle \geq \frac{\|\xi\|^2}{2\lambda^2\|v^k\|} - \frac{\|\xi\|}{\lambda} \geq -\frac{\|\xi\|}{\lambda}.$$

Thus, given any $\alpha_k \downarrow 0$, one finds small $\delta_k > 0$ such that

$$\langle y_k^*,\ \eta\rangle + \frac{\alpha_k}{\lambda}\|\eta\| \geq -\frac{1}{\lambda}\|\xi\| \ \text{ if } \|(\xi,\eta)\| < \delta_k \text{ and } (x^k + \xi, p^k + \eta) \in \operatorname{gph} F. \tag{5.7}$$

Since S is not pseudo-Lipschitz, (5.7) hold for $\lambda = \lambda_\nu \to \infty$. We select, to $\nu > 1$, some sufficiently large index $k = k(\nu) > k(\nu - 1)$ and the related points in (5.7). Then $y_{k(\nu)}^* \to y^* \neq 0$ may be assumed (otherwise pass to a related subsequence), and (5.7) tells us that, with vanishing $\varepsilon_\nu = \frac{1}{\lambda_\nu}$, $\beta_\nu = \|y_{k(\nu)}^* - y^*\| + \alpha_{k(\nu)}/\lambda_\nu$ and $\delta'_\nu = \delta_{k(\nu)}$,

$$\langle y^*,\ \eta\rangle + \ \beta_\nu\|\eta\| \geq -\varepsilon_\nu\|\xi\| \ \text{ if } \|(\xi,\eta)\| < \delta'_\nu \text{ and } (x^{k(\nu)} + \xi, p^{k(\nu)} + \eta) \in \operatorname{gph} F. \tag{5.8}$$

Using that F is loc. Lipsch. with rank L_F near x^0, we have

$$p^{k(\nu)} = F(x^{k(\nu)}), \ \ p^{k(\nu)} + \eta = F(x^{k(\nu)} + \xi) \ \text{ and } \ \|\eta\| \leq L_F\|\xi\|.$$

Thus $x^{k(\nu)}$ is a local $(\varepsilon_\nu + \ \beta_\nu L_F)-$ Ekeland point for $\langle y^*, F(x)\rangle$.

On the other hand, if x^0 is weakly stationary for $y^* \neq 0$ and $f(x) = \langle y^*, F(x)\rangle$ then, considering the equation $F(x^k + \xi) = F(x^k) - ty^*$ for small $t > 0$ at ε_k- Ekeland points x^k of f, it follows that S cannot be pseudo-Lipschitz. □

6 Stability and algorithms

Again we consider (in B-spaces) strongly closed maps S only. Given some $(p, x) \in \operatorname{gph} S$ close to (p^0, x^0) and π close to p^0, we want to determine some $x_\pi \in S(\pi)$ with $d(x_\pi, x) \leq L\|\pi - p\|$ by algorithms. Evidently, it suffices to solve

$$\min\ d(x_\pi, x) \quad \text{s.t. } x_\pi \in S(\pi), \tag{6.1}$$

but we are interested in an iterative procedure for this (generally) nonlinear problem. By Thm. 1, the Aubin property of S at (p^0, x^0) is equivalent to condition (5.6):

$$\begin{gathered}\exists L > 0,\ \varepsilon > 0 \text{ such that } \forall x \in S_\varepsilon(p) \text{ and } \pi \neq p \in p^0 + \varepsilon B \text{ it holds} \\ \operatorname{dist}(x, S(p')) + L\|p' - \pi\| < L\|p - \pi\| \ \text{ for some } p'.\end{gathered}$$

Therefore, if x, p and π belong to the related neighborhoods, (5.6) can be satisfied even with the particular point $p' = \pi$. This p' satisfies, for each given $\theta \in (0, 1)$,

$$\|p' - \pi\| \leq \theta\|p - \pi\|. \tag{6.2}$$

6.1 The general scheme

Next we require (6.2) for a sequence of parameters p^k. For the subsequent algorithm, which should be seen as being a framework for several more concrete procedures, we suppose that some $\lambda > 0$ and $\theta \in (0, 1)$ are given.
ALG1

Put $(p^1, x^1) = (p, x) \in \operatorname{gph} S$ and choose $(p^{k+1}, x^{k+1}) \in \operatorname{gph} S$ in such a way that

$$\begin{aligned}&(i)\quad \lambda^{-1}d(x^{k+1}, x^k) + \|p^{k+1} - \pi\| \leq \|p^k - \pi\| \quad \text{and} \\ &(ii)\quad \|p^{k+1} - \pi\| \leq \theta\,\|p^k - \pi\|.\end{aligned} \tag{6.3}$$

Lemma 2 *If related points* (p^{k+1}, x^{k+1}) *exist in each step then convergence follows.*

$$x^k \to x_\pi, \quad p^k \to \pi, \text{ where } \quad x_\pi \in S(\pi) \text{ and } d(x_\pi, x) \le \lambda \|\pi - p\|. \qquad \diamond \tag{6.4}$$

Proof. Beginning with $n = 1$, the estimate

$$\|x^{n+1} - x\| \le \sum_{k=1}^{n} d(x^{k+1}, x^k) \le \lambda \left(\|p^1 - \pi\| - \|p^{n+1} - \pi\| \right) \tag{6.5}$$

follows from (6.3)(i) by complete induction. So, a Cauchy sequence $\{x^k\}$ will be generated; let $x_\pi = \lim x^k$. Then (6.5) ensures $d(x_\pi, x) \le \lambda \|\pi - p\|$. Finally, (6.3)(ii) yields $p^k \to \pi$ whereafter $x_\pi \in S(\pi)$ holds due to closeness of S. $\square$

We call the algorithm applicable if related (p^{k+1}, x^{k+1}) exist in each step. Under calmness, we apply the same algorithm with fixed $\pi \equiv p^0$.

Theorem 3 *Let S be strongly closed.*
(i) The Aubin property of S holds at z^0 iff ALG1 is applicable, for some pair of $\theta \in (0,1)$ and $\lambda > 0$, whenever $\|\pi - p^0\| + d((p,x), z^0)$ is small enough.
(ii) The same statement, with $\pi \equiv p^0$, holds in view of calmness of S at z^0. $\diamond$

Proof.
(i) Let the Aubin property be satisfied. Then (5.6) holds for all sufficiently large $L = \lambda$ and ensures via Thm. 1 (as explained above) and (6.5) the existence of the next iterates whenever $\|\pi - p^0\| + d((p,x), z^0)$ was small enough, e.g., if

$$\lambda(\|p - p^0\| + \|\pi - p^0\|) < \tfrac{1}{2}\varepsilon \text{ and } d(x, x^0) < \tfrac{1}{2}\varepsilon$$

with ε from (5.6).
Conversely, if the Aubin property is violated then, for each $\lambda > 0$, one finds points $(p,x) \in \operatorname{gph} S$ arbitrarily close to z^0, and related π, namely (p^k, x^k) and π_k from (5.2), (if (5.2) is applied to some $\lambda' > \lambda$) such that already the first step of the procedure does not work.
(ii) In view of calmness, the same arguments can be applied with $\pi \equiv p^0$ since Thm. 1 holds in the same manner. $\square$

Remark 3 Property (6.3)(ii) follows from both (6.3)(i) and the stepsize rule

$$d(x^{k+1}, x^k) \ge \tau_k := \lambda(1 - \theta)\ \|p^k - \pi\|, \tag{6.6}$$

since (6.3)(i) and (6.6) yield $(1-\theta)\|p^k - \pi\| + \|p^{k+1} - \pi\| \le \|p^k - \pi\|$. Further, the theorem still holds after replacing (6.3)(ii) by any other condition which ensures $p^k \to \pi$ and $\|p^{k+1} - \pi\| < \|p^k - \pi\|$ if $p^k \ne \pi$. $\diamond$

The formally similar statements concerning calmness and the Aubin property do not imply that algorithm (6.3) runs in the same way under these conditions.

Aubin property: If (6.3) is applicable for all inititial points near z^0 in $\operatorname{gph} S$ then we can first fix any p^{k+1} with $\|p^{k+1} - \pi\| \le \theta \|p^k - \pi\|$ and next find, since S is pseudo-Lipschitz at z^0 and the points under consideration are close to z^0, some $x^{k+1} \in S(p^{k+1})$ satisfying the required inequality. Accordingly, related x^{k+1} exist for each sequence $p^k \to \pi$ satisfying $\|p^{k+1} - \pi\| \le \theta \|p^k - \pi\|$.
Calmness: Though every feasible sequence in (6.3) leads us to some $x_\pi \in S(\pi)$ we are only sure that some feasible x^{k+1} exists if $p^{k+1} = \pi = p^0$. In other words, the sequence (p^k, x^k) could be trivial, $(p^k, x^k) = (\pi, x_\pi)\ \forall k \ge k_0$. The most simple example: $F(x) \equiv \{0\}$, $S = F^{-1}$.

6.2 Particular realizations

6.2.1 Descent method

For minimizing $f = f(x)$, $f \in C^1(\mathbb{R}^n, \mathbb{R})$ suppose that $\pi = \min f$ exists and f has compact level sets. Given any x^1 the gradient is uniformly continuous and bounded on $M = \{x | f(x) \le f(x^1) + 1\}$. So, for each $\delta > 0$ there exists $t(\delta) > 0$ such that, if $|t| < t(\delta)$ and $f(x^k) \le f(x^1)$, the conditions

$$x^k - tDf(x^k) \in M \text{ and } \|Df(x^k - tDf(x^k)) - Df(x^k)\| < \frac{\delta}{2}$$

are satisfied.

We consider a descent method with start at x^1: $x^{k+1} = x^k - t_k Df(x^k)$, $p^k = f(x^k)$, where $t_k \in (\tau,\ t(\delta))$, $\tau > 0$, and verify that

$$\|Df(x^k)\| < \delta \text{ holds for some } k. \tag{6.7}$$

Writing (as usual) $p^{k+1} - p^k = -t_k \langle Df(\xi^k), Df(x^k)\rangle$ by the mean value theorem, and replacing $Df(\xi^k)$ by $Df(x^k) - (Df(x^k) - Df(\xi^k))$ yields

$$\langle Df(\xi^k), Df(x^k)\rangle \geq \|Df(x^k)\|^2 - \frac{\delta}{2}\|Df(x^k)\| = \|Df(x^k)\|\ (\|Df(x^k)\| - \frac{\delta}{2}).$$

This gives the standard estimate

$$p^{k+1} - p^k \leq -t_k\ \|Df(x^k)\|\ (\|Df(x^k)\| - \tfrac{\delta}{2}) \leq -\ (\|Df(x^k)\| - \tfrac{\delta}{2})\ \|x^{k+1} - x^k\|. \tag{6.8}$$

Now let (6.7) be violated. Then (6.8) yields (6.3)(i) for $\lambda = \frac{2}{\delta}$ as well as (6.6) for each $\theta > 0$ satisfying $1 - \theta \leq \frac{\delta\ \tau}{\lambda\ \|p^1 - \pi\|}$, due to

$$d(x^{k+1}, x^k) \geq \delta\tau \geq \lambda(1-\theta)\ \|p^1 - \pi\| \geq \lambda(1-\theta)\ \|p^k - \pi\|. \tag{6.9}$$

Lemma 2 would yield $x^k \to x_\pi \in \operatorname{argmin} f$ and $\|Df(x_\pi)\| \geq \delta$ by continuity of Df. The latter fails to hold at a minimizer. Hence (6.7) holds indeed.

6.2.2 Generalized (non-smooth) Newton methods

The algorithm is known if S is the inverse of a loc. Lipsch. function $f : X \to P,\ X, P$ (B-spaces). To show this we simplify:

$$\pi = p^0 = 0.$$

Then (6.3) requires with $p^k = f(x^k)$,

$$\frac{\|f(x^{k+1})\| - \|f(x^k)\|}{\|x^{k+1} - x^k\|} < -\frac{1}{\lambda} \text{ and } \|f(x^{k+1})\| \leq \theta\|f(x^k)\|. \tag{6.10}$$

These are key properties for convergence of *(generalized) Newton methods*

$$x^{k+1} = x^k - A_k^{-1} f(x^k), \quad A_k = Rf(x^k),\ k \geq 1$$

to solve $f(x) = 0$ with x^1 close to x^0, under the standard hypotheses at a zero x^0, namely:
(1) For x near x^0, let uniformly bounded linear operators $Rf(x)$ fulfill the approximation condition

$$f(x) - f(x^0) - Rf(x)(x - x^0) = o(x - x^0)$$

(2) and have uniformly bounded inverses $Rf(x)^{-1}$ (injectivity condition).
Then (cf. [36] for details and related references and [20] for Broyden-type modifications) there is some $\gamma > 0$ such that

$$\|f(x^k)\| \geq \gamma\ d(x^k, x^0). \tag{6.11}$$

Further, given $\alpha > 0$ then, taking $d(x^1, x^0)$ small enough, it holds

$$d(x^{k+1}, x^0) \leq \alpha\ d(x^k, x^0) \quad \text{and} \quad \|f(x^{k+1})\| \leq \alpha\ d(x^k, x^0).$$

With $0 < \alpha < \min\{\gamma, 1\}$, this yields

$$d(x^k, x^{k+1}) \leq d(x^k, x^0) + d(x^0, x^{k+1}) \leq (1+\alpha)\ d(x^k, x^0) \leq 2d(x^k, x^0)$$

and ensures the first inequality of (6.10)

$$\|f(x^k)\| - \|f(x^{k+1})\| \geq \gamma\ d(x^k, x^0) \geq \tfrac{1}{2}\ \gamma\ d(x^k, x^{k+1}), \tag{6.12}$$

while the second one follows from (6.11): $\|f(x^{k+1})\| \leq \alpha\ d(x^k, x^0) \leq \frac{\alpha}{\gamma}\|f(x^k)\|$. □
In the general version of algorithm (6.3), the conditions (6.10) are required for appropriate selections $f(x^k) \in F(x^k)$ only.

6.2.3 The projection method

If dist $((p,z), \operatorname{gph} S)$ will be attained for all (p,z), Thm. 3 can be written by means of the following projection method which presents a particular possibility for realizing the steps in ALG1.
ALG2

$$\begin{array}{l}\text{Put } (p^1, x^1) = (p,x) \in \operatorname{gph} S \text{ and choose } (p^{k+1}, x^{k+1}) \in \operatorname{gph} S \text{ in such a way that} \\ (p^{k+1}, x^{k+1}) \text{ minimizes } \ \|x' - x^k\| + \lambda \|p' - \pi\| \quad s.t. \ (p', x') \in \operatorname{gph} S.\end{array} \tag{6.13}$$

Theorem 4 *Assume the minimum exist in each step. Then*
(i) The Aubin property of S holds at z^0 iff ALG2 generates, for sufficiently large λ and small $d((p,x), z^0) + \|\pi - p^0\|$, a sequence satisfying

$$\lambda^{-1} \|x^{k+1} - x^k\| \ + \ \|p^{k+1} - \pi\| \ \le \ \theta \|p^k - \pi\| \tag{6.14}$$

with some constant $\theta < 1$.
(ii) The same statement, with $\pi \equiv p^0$, holds in view of calmness of S at z^0. ◇

Note that (6.14) ensures again convergence $x^k \to x_\pi \in S(\pi)$ with $\|x_\pi - x\| \le \lambda \|\pi - p\|$

Proof. (i) Suppose the Aubin property with rank L, and fix $\lambda > L$. Considering again points near (p^0, x^0) one may apply the existence of $\hat{x} \in S(\pi)$ with $\|\hat{x} - x^k\| \le L \|\pi - p^k\|$. This yields for the minimizer in (6.13)

$$\|x^{k+1} - x^k\| + \lambda \|p^{k+1} - \pi\| \le \|\hat{x} - x^k\| + \lambda \|\pi - \pi\| \le L \|p^k - \pi\| \tag{6.15}$$

and implies that $\theta = \frac{L}{\lambda} < 1$ fulfills (6.14).

Conversely, let (6.14) be true for certain $\lambda > 0$, $\theta \in (0,1)$ and all related initial points. Then also (6.3)(i) and (6.3)(ii) are valid for the current sequences. By Thm. 3 so the Aubin property must be satisfied.

(ii) Applying the corresponding modification for calmness in the same manner, the assertion follows. □

6.2.4 Interpretations of ALG2 as Feijer and Penalty method

ALG2 as Feijer method:
The construction of the sequence can be understood as a Feijer method w.r.t. to the norm $\|.\|_X + \lambda \|.\|_P$ and the two subsets $M_1 = (\pi, X)$, $M_2 = \operatorname{gph} S$ of (P, X).

Given $z^k = (p^k, x^k)$, find first the point $z(1)^k = (\pi, x^k)$ by projection of z^k onto M_1 and next $z(2)^k$ by projection of $z(1)^k$ onto M_2. Write $z^{k+1} = z(2)^k = (p^{k+1}, x^{k+1})$ and repeat.

ALG2 as penalty method:
The term $\lambda \|p - \pi\|$ in the objective of ALG2 can be understood as penalization of the requirement $p = \pi$. So we simply solve

$$\min \|x - x^k\| \quad s.t. \ (p,x) \in \operatorname{gph} S, \ p = \pi$$

by partial penalization and know that p^k is the current value of p, assigned to x^k. Condition (6.14) requires *linear convergence*.

Summarizing, this ensures (at least) for $\dim X + \dim P < \infty$:

Corollary 1 *Calmness and the Aubin property at z^0 are equivalent to local (linear) convergence of the penalty method for suffic. large penalization factor λ and initial points in $\operatorname{gph} S$ near z^0 where one has to require $\pi = p^0$ (calm) und π near p^0 (Aubin property), respectively.*

6.3 Modified successive approximation and perturbed mappings

Modified successive approximation is the typical method for showing the following statement for closed mappings $F : X \rightrightarrows P$ (B-spaces) and $\Gamma = (h + F)^{-1}$.

The key observation consists in the fact that, if $T : X \rightrightarrows X$ obeys the Aubin property with rank $\gamma < 1$ at $(x^1, x^2) \in \operatorname{gph} T$ and $d(x^2, x^1)$ is sufficiently small (compared with γ and ε, δ in definition D3), there exist, for $k > 1$, elements x^{k+1} such that

$$x^{k+1} \in T(x^k) \text{ and } d(x^{k+1}, x^k) \ \leq \ \gamma \, d(x^k, x^{k-1}). \tag{6.16}$$

The sequence then fulfills $x^k \to \hat{x} \in T(\hat{x})$ and behaves as in Banach's fixed point theorem:

$$\begin{aligned} d(x^{n+1}, x^1) \ &\leq d(x^{n+1}, x^n) + ... + d(x^2, x^1) \\ &\leq (\gamma^{n-1} + ... + \gamma^0) \ d(x^2, x^1) \leq \tfrac{1}{1-\gamma} \ d(x^2, x^1). \end{aligned} \tag{6.17}$$

In more special settings, this been already applied in [6] for verifying persistence of the Aubin property under small C^1 perturbations.

Theorem 5 *Let $S = F^{-1}$ obey the Aubin property with rank L at $z^0 = (p^0, x^0)$ and let $h : X \to P$ be a function with*

$$\|h(x') - h(x)\| \leq \alpha \, d(x', x) \textit{ for all } x', x \textit{ near } x^0 \quad \textit{and} \quad \|h(x^0)\| \leq \beta.$$

Then, if $\|\pi - p^0\| + \alpha + \beta$ is sufficiently small, the mapping $\Gamma = (h+F)^{-1}$ obeys the Aubin property at $(p^0 + h(x^0), \ x^0)$ and, moreover, there exists some x_π with

$$\pi \in h(x_\pi) + F(x_\pi) \textit{ and } \ d(x_\pi, x^0) \leq \beta L + \frac{L}{1 - L\alpha} \ (\ \|\pi - p^0\| + \alpha\beta L \ ). \quad \diamond$$

Proof. Let $\|\pi - p^0\| + \alpha + \beta$ be small enough such that

$$(L + \frac{L}{1 - L\alpha}) \ (\ \|\pi - p^0\| + \beta \ ) \ + \ L\alpha < \min\{1, \varepsilon/2\}$$

holds with $\varepsilon = \delta$ in definition (3.4). Note that

$$x \in \Gamma(p) \Leftrightarrow p \in (h + F)(x) \Leftrightarrow p - h(x) \in F(x) \Leftrightarrow x \in S(p - h(x)). \tag{6.18}$$

For small $\|p - h(x) - p^0\| + \|x - x^0\|$, the mapping

$$T_p(x) = S(p - h(x)) \tag{6.19}$$

obeys the Aubin property with rank $\gamma = L\alpha < 1$. We show the Aubin property of Γ. Given $(p, x) \in \operatorname{gph} \Gamma$ close to $(p^0 + h(x^0), \ x^0)$, let $(p^1, x^1) = (p, x)$. Then we have $x^1 \in S(p^1 - h(x^1))$ and find, by the Aubin property of S at (p^0, x^0), some

$$x^2 \in S(\pi - h(x^1)) = T_\pi(x^1) \ \text{ with } \ d(x^2, x^1) \leq L \ \|\pi - p^1\|.$$

Next one may apply the Aubin property of $T := T_\pi$ with rank $\gamma < 1$ as in (6.16) and (6.17) since the latter estimate yields

$$d(x^{n+1}, x^1) \ \leq \tfrac{1}{1-\gamma} \ d(x^2, x^1) \leq \tfrac{L}{1-L\alpha} \ \|p^1 - \pi\|. \tag{6.20}$$

Denoting the fixed point $\hat{x}$ by x_π we obtain

$$\pi \in h(x_\pi) + F(x_\pi) \ \text{ and } \ d(x_\pi, x^1) \ \leq \frac{L}{1 - L\alpha} \ \|\pi - p^1\|.$$

Hence $\Gamma = (h + F)^{-1}$ obeys the Aubin property at $(p^0 + h(x^0), \ x^0)$ with rank $\frac{L}{1-L\alpha}$.

For proving the full theorem, observe that by the Aubin property of S,

$$p^0 - h(x^0) \in F(x_h) \text{ holds for some } x_h \in x^0 + \beta L \ B. \tag{6.21}$$

Setting $p^1 = p^0 - h(x^0) + h(x_h)$, it follows $p^1 \in h(x_h) + F(x_h)$ whereafter (6.18) yields $(p^1, x_h) \in \operatorname{gph} \Gamma$. Further we have

$$\|p^1 - p^0\| = \|h(x^0) - h(x_h)\| \leq \alpha \ d(x_h, x^0) \leq \alpha\beta L.$$

Thus we may start the above prescribed process with $(p^1, x^1) = (p^1, x_h)$. This ensures the existence of x_π with $\pi \in h(x_\pi) + F(x_\pi)$ as well as the required estimate

$$\begin{array}{l} d(x_\pi, x^0) \le d(x^0, x^1) + d(x^1, x_\pi) \\ \le \beta L + \frac{1}{1-L\alpha}\ L\ \|\pi - p^1\| \\ \le \beta L + \frac{1}{1-L\alpha}\ L\ (\ \|\pi - p^0\| + \|p^0 - p^1\|\ ) \\ \le \beta L + \frac{L}{1-L\alpha}\ (\ \|\pi - p^0\| + \alpha\beta L\ ). \qquad \square \end{array} \tag{6.22}$$

The Lyusternik Theorem. Let $g : X \to P$ be a C^1 function and $Dg(x^0)$ map X onto P (B-spaces). Since $Dg(x^0)^{-1} : P \to X$ is pseudo-Lipschitz by Banach's inverse mapping theorem, we may put

$$F(x) = g(x^0) + Dg(x^0)(x - x^0) \text{ and } h(x) = g(x) - F(x).$$

Then the suppositions of the Thm. 5 hold for all small positive α and β, in particular for small $\alpha < L/2$ which yields $\gamma < \frac{1}{2}$. Due to $\pi = h(x) + F(x) \Leftrightarrow \pi = g(x)$, this proves local solvability of the latter equation with related estimates which is Lyusternik's theorem.

Quasi-Newton method. In order to solve

$$g(x) = \pi$$

with initial point x^0, $p^0 = g(x^0)$ and π close to p^0, the first step of finding x_h in (6.21) allows to put $x_h = x^0$ since $h(x^0) = 0$. The iterations in (6.16) stand for solving

$$Dg(x^0)x = \pi - h(x^k) \quad \text{and} \quad d(x, x^k) \ \le\ \gamma\ d(x^k, x^{k-1}). \tag{6.23}$$

Thus the derivative at the initial point x^0 may be used in order to determine a solution of $g(x) = \pi$ whenever $\|\pi - p^0\|$ is small enough. The resulting method is a (very simple since no update) quasi-Newton method. In a Hilbert space, one could select a minimizer of $d(x, x^k)$ among the solutions of the linear system $Dg(x^0)x = \pi - h(x^k)$.

Note. For more special mappings under the Aubin property, methods like (6.3) have been investigated also in [7, 48, 36] (generalized Newton methods and successive approximation). A general approach and its relations to proximal point methods can be found in [36], too. To verify calmness for certain intersections of mappings, an algorithmic approach based on Newton's method for semismooth functions has been used in [23], too. To characterize the Aubin property equivalently for intersections by MFCQ-like conditions in B-spaces, an algorithmic approach has been applied in [49].

7 Stability and generalized derivatives

In various papers dealing with non-smooth and multivalued analysis (nowadays also called variational analysis), conditions of stability have been presented in terms of certain generalized derivatives. The latter are, in any case, sets of certain limits and have a more complicated structure than Fréchet- derivatives since, at least, double limits are involved. As already mentioned, this has restrictive consequences for computing them and for all calculus rules. On the other hand, these derivatives show which local properties are essential for the stability in question and which differences now occure in comparison with the known case of continuously differentiable functions. The most of the next statements can be found in several (quite distributed) papers. Our approach, based on Theorem 1, is self-contained, straightforward and establishes the bridge to the above-mentioned conditions.

In this section, we suppose that X and P are Euclidean spaces since otherwise not any of the following characterizations by generalied derivatives remains valid.

7.1 Some generalized derivatives

Let $F : X \rightrightarrows P,\ S = F^{-1}$ and $\zeta^0 = (x^0, p^0) \in \operatorname{gph} F$.

Definition 1 (*contingent derivative CF*). Let $\hat{x} \in X$. The set of all $\hat{p} \in P$ such that $p^0 + t_k\hat{p}^k \in F(x^0 + t_k\hat{x}^k)$ holds for certain $t_k \downarrow 0$ and $(\hat{x}^k, \hat{p}^k) \to (\hat{x}, \hat{p})$, forms the *contingent derivative, also called Bouligand derivative* $CF(\zeta^0)(\hat{x})$, cf. [1]. $\diamond$

Then $\text{gph}\, CF(\zeta^0) \subset (X, P)$ is *Bouligand's tangent cone* $T_{\text{gph}\, F}(\zeta^0)$ *of* gph F *at* ζ^0.

Definition 2 (*strict graphical derivative* TF). Let $\hat{x} \in X$. The set of all $\hat{p} \in P$ such that $p^k \in F(x^k)$ and $p^k + t_k\hat{p}^k \in F(x^k + t_k\hat{x}^k)$ hold for certain $t_k \downarrow 0$ and related sequences $(x^k, \hat{x}^k, p^k, \hat{p}^k) \to (x^0, \hat{x}, p^0, \hat{p})$ forms the derivative $TF(\zeta^0)(\hat{x})$. ◇

Sets of this form have been called *strict graphical derivative* in [76]. For loc. Lipsch. $F : X \to P$, $TF(\zeta^0)(\hat{x})$ consists just of all limits of the form

$$\hat{p} = \lim_{t_k \downarrow 0,\ x^k \to x^0} t_k^{-1}\ [F(x^k + t_k\hat{x}) - F(x^k)].$$

Such limits were introduced by Thibault (to define other derivatives) and called *limit sets* in [79, 80], and appeared in [32, 36, 43] (to study inverse Lipschitz functions) as $\Delta-$ or T- *derivatives.* For loc. Lipsch. $f : X \to \mathbb{R}$, the value $\sup Tf(\zeta^0)(\hat{x})$ is just Clarke's [5] directional derivative $f^{Cl}(x^0)(\hat{x})$.

The general definition has been applied (up to now) only to mappings which can be (linearly) transformed into loc. Lipsch. functions, cf. [50], [51].

Definition 3 (*co-derivative* D^*F). The map $D^*F(\zeta^0) : P^* \rightrightarrows X^*$ is defined by $x^* \in D^*F(\zeta^0)(p^*)$ if $\exists\ \varepsilon_k \downarrow 0,\ \delta_k \downarrow 0$ and points $(x^k, p^k) \to \zeta^0$ in gph F with

$$\langle p^*, \eta\rangle + \varepsilon_k\|(\xi, \eta)\| \geq \langle x^*, \xi\rangle \text{ whenever } \|(\xi, \eta)\| \leq \delta_k \text{ and } (x^k + \xi, p^k + \eta) \in \text{gph}\, F, \quad ◇ \qquad (7.1)$$

cf. [59, 60]. The latter requires that $(x^*, -p^*)$ is (locally) an *approximate normal* at (x^k, p^k) to gph F with *error* ε_k. The vector $(x^*, -p^*)$ is also called a limiting Fréchet- normal. In terms of CF, (7.1) means with new $\varepsilon_k \downarrow 0$,

$$\langle p^*,\ \hat{p}\rangle + \varepsilon_k \geq \langle x^*, \hat{x}\rangle \text{ whenever } \hat{p} \in CF(x^k, p^k)(\hat{x}) \text{ and } \|(\hat{p}, \hat{x})\| \leq 1. \qquad (7.2)$$

If (x^k, p^k) is isolated in gph F then (7.1) holds trivially for sufficiently small δ_k.

Definition 4 (*Generalized Jacobian* $\partial^c f(x)$). For $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$, put

$$M = \{A \mid A = \lim Df(x^k),\ x^k \to x,\ Df(x^k) \text{ exists}\}.$$

Then $M \neq \emptyset$ holds by Rademacher's theorem (f is a.e. F-differentiable). The convex hull $\partial^c f(x) = \text{conv}\, M$ is Clarke's generalized Jacobian of f at x, cf. [4, 5]. ◇

Strict differentiability
For $h : X \to P$, the point $p^0 = h(x^0)$ can be deleted from the description of the derivatives. If $Th(x^0)(\hat{x})$ is even single-valued for all directions $\hat{x}$, one says that h is *strictly differentiable* at x^0. Explicitly, this means that all difference quotients

$$t_k^{-1}\ (\ h(x^k + t_k\hat{x}^k) - h(x^k)\ ) \quad \text{as } x^k \to x^0,\ \hat{x}^k \to \hat{x} \text{ and } t_k \downarrow 0$$

have the same limit. Obviously, then $Th(x^0)(\hat{x}) = \{Dh(x^0)\hat{x}\}\forall \hat{x}$. Every C^1 function is strictly differentiable at x^0. The reverse is not true since h may have kinks at certain points $x^k \to x^0$.

7.2 First motivations of the definitions

Inverse functions:

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be loc. Lipschitz and $f(0) = 0$.
(i) f is s.L.s. at $(0, 0)$ (i.e. f^{-1} is locally well-defined and Lipschitz) if all matrices in $\partial^c f(0)$ are regular, cf. Clarke, [4, 5].
(ii) f is s.L.s. at $(0, 0) \Leftrightarrow\ 0 \notin Tf(0)(u)\forall u \in \mathbb{R}^n \setminus \{0\}$. For $n = 2$, there is a piecewise linear homeomorphism f which is s.L.s. at $(0, 0)$ and $0 \in \partial^c f(0)$.
Hence Clarke's condition is not necessary.
(iii) $Tf(x)(u)$ is connected and $\text{conv}\, Tf(x)(u) = \{Au \mid A \in \partial^c f(x)\ \}$.
(iv) For Kojima's function Φ and involved $C^{1,1}$ functions, the derivatives $T\Phi$ and $C\Phi$ *can be determined* by the usual product rule of differentiation.

Implicit functions:

Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ be loc. Lipsch. and $f(0,0) = 0$.
(v) Then, the solutions x to $f(x,p) = y$ are locally unique and Lipschitz (near the origin) $\Leftrightarrow 0 \notin Tf(0,0)(u,0)\ \forall u \in \mathbb{R}^n \setminus \{0\}$. For (ii) ... (v), see [43, 44].

Minimizer:

Let $f : \mathbb{R}^n \to \mathbb{R}$ be loc. Lipsch. and $\hat{x}$ be a local minimizer. Then:
(vi) $0 \in \partial^c f(\hat{x})$, [5].
(vii) $0 \in D^*F(\hat{x})(1)$ for $\operatorname{gph} F = \{(x,r) \mid r \geq f(x)\}$ (this implies (vi)), [58].
(viii) If $f(x) = \max_i g_i(x)$ where the g_i form a finite collection of C^1 functions, then $\partial^c f(x) = \operatorname{conv}\{Dg_i(x) \mid \text{over } i \text{ with } g_i(x) = f(x)\}$ [5]; similarly for $f(x) = \max_{0 \leq t \leq 1} g(x,t)$ if g and $D_x g$ are continuous.

Taylor expansion:

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $f \in C^{1,1}$. Then

$$\text{(ix)} \qquad f(x+u) - f(x) = Df(x)u + \tfrac{1}{2}\langle u, q\rangle$$

holds for some $\theta \in (0,1)$ and $q \in \partial^c Df(x+\theta u)u$ as well as for some $\theta \in (0,1)$ and $q \in TDf(x+\theta u)u$, cf. [22] and [44], respectively.

All mentioned derivativs can be computed for Kojima's function if $f, g, h \in C^2$. For other motivations, see Thm. 6.

7.3 Some chain rules

One easily sees that the above generalized derivatives do not change after adding a function h with

$$h(x^0) = 0 \text{ and } Th(x^0)(\hat{x}) = \{0\} \forall \hat{x}. \tag{7.3}$$

Even more, if F and $G : X \rightrightarrows P$ are two mappings such that the *Hausdorff-distance* of the images $d_H(F(x), G(x)) := \inf\{\alpha > 0 | F(x) \subset G(x) + \alpha B \text{ and } G(x) \subset F(x) + \alpha B\}$ satisfies

$$d_H(F(x), G(x)) \leq \|h(x)\|, \quad h \text{ from (7.3)} \tag{7.4}$$

then the introduced derivatives of F and G at $\zeta^0 = (x^0, p^0)$ remain the same (replace the elements $p^k, \hat{p}^k, \eta$ which appear in the related derivative for F by corresponding (nearby) elements of the $G-$images and vice versa).

Similarly, if $G(x) = h(x) + F(x) \forall x$ (near x^0) and if $Dh(x^0)$ exists as strict derivative, then (by direct substitutions) it follows with $q^0 = h(x^0) + p^0$,

$$\begin{array}{ll} CG\ (x^0, q^0) \ = Dh(x^0) \ + \ CF(\zeta^0), & TG(x^0, q^0) = Dh(x^0) + TF(\zeta^0), \\ D^*G(x^0, q^0) = \ Dh(x^0)^* + D^*F(\zeta^0). & \end{array} \tag{7.5}$$

In particular, this permits us to interpret the derivatives in terms of linear functions $A : X \to Z$ and zero- derivatives only (the latter indicate some singularity, cf. Thm. 6) where $q^0 = p^0 - Ax^0$:
(i) $p \in CF(\zeta^0)(u) \Leftrightarrow \exists A$ such that $0 \in C(F - A)(x^0, q^0)(u)$ and $p = Au$
(ii) the same for the derivative T
(iii) $u^* \in D^*F(\zeta^0)(p^*) \Leftrightarrow \exists A$ such that $0 \in D^*(F - A)(x^0, q^0)(p^*)$ and $u^* = A^*p^*$.
By considering related A with smallest norm one can study some range of stability for the mapping S, cf. [10].

Inverse mappings.
Due to the symmetry with respect to images and pre-images, the derivative TS or CS of the inverse $S = F^{-1}$ is just the inverse of TF or CF, respectively. For D^*S, one has $p^* \in D^*S(z^0)(x^*)$ if the elements in (7.1) satisfy $\varepsilon_k \|(\xi, \eta)\| \geq \langle p^*, \eta\rangle - \langle x^*, \xi\rangle$. So, compared with D^*F, p^* and x^* change the place and the sign. Summarizing, this tells us

$$\begin{array}{lcl} -x^* \in D^*F(\zeta^0)(-p^*) & \Leftrightarrow & p^* \in D^*F^{-1}(z^0)(x^*), \\ \hat{p} \in CF(\zeta^0)(\hat{x}) & \Leftrightarrow & \hat{x} \in C(F^{-1})(z^0)(\hat{p}), \\ \hat{p} \in TF(\zeta^0)(\hat{x}) & \Leftrightarrow & \hat{x} \in T(F^{-1})(z^0)(\hat{p}). \end{array} \tag{7.6}$$

C, T **and** D^* **for the inverse** $S = (h + F)^{-1}$.
This mapping assignes, to every $p \in P$, the solutions of the inclusion $p \in h(x) + F(x)$. If

$Dh(x^0)$ exists as strict derivative, the above formulas can be combined in order to obtain, for $x^0 \in S(p^0), p^0 - h(x^0) = q^0$ and $q^0 \in F(x^0)$,

$$\begin{array}{lcll} \hat{x} \in CS(z^0)(\hat{p}) & \Leftrightarrow & \hat{p} & \in [Dh(x^0) + CF(x^0, q^0)](\hat{x}), \\ \hat{x} \in TS(z^0)(\hat{p}) & \Leftrightarrow & \hat{p} & \in [Dh(x^0) + TF(x^0, q^0)](\hat{x}), \\ p^* \in D^*S(z^0)(x^*) & \Leftrightarrow & -x^* & \in [D^*h(x^0) + D^*F(x^0, q^0)](-p^*). \end{array} \tag{7.7}$$

Linear transformations

Regular linear transformations of the image- or pre-image space change the generalized derivatives C, T, D^* in the same manner as DF and the usual adjoint map D^*F, respectively. Further, it holds for any linear function A

$$\hat{p} \in CF(\zeta^0)(\hat{x}) \Rightarrow A\hat{p} \in C(AF)(x^0, Ap^0)(\hat{x}), \text{ similarly for } T(AF). \tag{7.8}$$

However, regularity plays a role for the reverse direction.

Example 4 $F(x) = 1/x,\ F(0) = 0$: With $A = 0$ it holds $0 \in C(AF)(0,0)(1)$ but $CF(0,0)(1) = \emptyset$. The same effect appears for TF. $\diamond$

The direction $\Leftarrow$ in (7.8) is valid if A^{-1} exists: apply (7.8) with A^{-1} to the right-hand side.

The formulas ensure, under strict differentiability of $h : X \to P$ and after regular linear transformations in X and P, that the derivatives of $S = h + F$ and S^{-1} are available if (and only if !) the related derivatives are known for F.

7.4 Conditions of stability

As before, we study closed maps $S : P \rightrightarrows X$ at a given point $z^0 = (p^0, x^0) \in \operatorname{gph} S$, put $F = S^{-1}$, suppose that P, X are Euclidean spaces and write ζ^0 in place of (x^0, p^0). There is a basic device for describing the desired Lipschitz properties by generalized derivatives: Negate the related stability.

(i) **Strongly Lipschitz**: The map S is not strongly Lipschitz iff

$$\begin{array}{l} \exists x^k \in S(p^k), \xi^k \in S(\pi_k) \text{ with } x^k, \xi^k \to x^0 \text{ and } p^k, \pi_k \to p^0 \\ \text{such that } x^k \neq \xi^k \text{ and } \ \|\pi_k - p^k\|/\|\xi^k - x^k\| \to 0 \ (k \to \infty). \end{array} \tag{7.9}$$

Writing, in situation (7.9), $\xi^k = x^k + t_k\hat{x}^k$, where $\|\hat{x}^k\| = 1$ and $t_k > 0$, and selecting a subsequence such that $\hat{x}^k \to \hat{x}$, one obtains $\pi_k = p^k + t_k\hat{p}^k$ with $\hat{p}^k \to 0$ and

$$\text{some } \hat{x} \neq 0 \text{ belongs to } TS(z^0)(0), \tag{7.10}$$

and vice versa. Hence, (7.9) and (7.10) coincide. In terms of F, the negation of (7.10) (i.e., the strong Lipschitz property of S) is just injectivity of $TF(\zeta^0)$:

$$0 \in TF(\zeta^0)(\hat{x}) \Rightarrow \hat{x} = 0. \tag{7.11}$$

(ii) **Upper Lipschitz**: The negation of the upper Lipschitz property is just

$$\begin{array}{l} \exists x^k \in S(p^k) \text{ with } x^k \to x^0 \text{ and } p^k \to p^0 \text{ such that} \\ x^k \neq x^0 \text{ and } \ \|p^k - p^0\|/\|x^k - x^0\| \to 0 \ (k \to \infty). \end{array} \tag{7.12}$$

Writing then $\hat{x}^k = \frac{x^k - x^0}{\|x^k - x^0\|}$, and selecting a subsequence such that $\hat{x}^k \to \hat{x}$, one sees that

$$\text{some } \hat{x} \neq 0 \text{ belongs to } CS(z^0)(0), \tag{7.13}$$

and vice versa. Hence, (7.12) and (7.13) coincide, too. In terms of F, the negation of (7.13) (i.e., the upper Lipschitz property of S) requires exactly injectivity of $CF(\zeta^0)$:

$$0 \in CF(\zeta^0)(\hat{x}) \Rightarrow \hat{x} = 0. \tag{7.14}$$

(iii) **Lower Lipschitz**: If S is lower Lipschitz then, taking $x(p) \in S(p) \cap (x^0 + L\|p - p^0\|B)$, it holds $\frac{\|x(p) - x^0\|}{\|p - p^0\|} \leq L$ for all $p \neq p^0, p$ near p^0. Setting $p = p^0 + t\hat{p}$ for fixed $\hat{p} \in \operatorname{bd} B$ and passing to the limit $t \downarrow 0$, some accumulation point $\hat{x}$ of $\frac{x(p)-x^0}{t} \in LB$ exists and belongs to $CS(z^0)(\hat{p})$. Since $\text{gph}\, CS(z^0)$ is a cone, so

$$CS(z^0)(\hat{p}) \cap LB \neq \emptyset \quad \forall \hat{p} \in B \tag{7.15}$$

is *necessary* for S to be lower Lipschitz at z^0. In terms of F, (7.15) means *surjectivity* (with linear rate) of $CF(\zeta^0)$, i.e.,

$$B \subset CF(\zeta^0)(LB). \tag{7.16}$$

Though (7.16) is not sufficient for being lower Lipschitz even for $F \in C(\mathbb{R}^2, \mathbb{R}^2)$, cf. Ex.9 in [36], (7.16) plays a crucial role if it holds for all $(x,p) \in \text{gph}\, F$ *near* ζ^0.

(iv) Pseudo-Lipschitz: We are now going to apply Lemma 1 for deriving two equivalent characterizations of the Aubin property of $S = F^{-1}$ at z^0, namely:

$$\exists \lambda > 0: \; B \subset CF(x,p)(\lambda B) \; \text{ for all } (p,x) \in \text{gph}\, S \text{ near } z^0; \tag{7.17}$$

and

$$0 \in D^*F(\zeta^0)(p^*) \Rightarrow p^* = 0. \tag{7.18}$$

Our general hypothesis that X and P are Euclidean spaces, is important in this context. We refer to [35], example BE.2, where F is given by the level sets of a Lipschitz function, $X = l^2, P = \mathbb{R}$ and both conditions are not necessary (for $\dim X = \infty$, Def. 3 must be modified).
Necessity:
Condition (7.17) is necessary since the Aubin property implies that S is lower Lipschitz for $z \in \text{gph}\, S$ near z^0 with the same rank, cf. (7.16). We consider (7.18), assume that $x^* \in D^*F(\zeta^0)(p^*)$, $p^* \neq 0$ and verify: If $x^* \neq 0$ then the rank L of the Aubin property fulfills $L \geq \frac{\|p^*\|}{\|x^*\|}$. If $x^* = 0$ then S does not obey the Aubin property.

In fact, let the derivative condition hold with sequences ε_k, δ_k and $(p^k, x^k) \to z^0$. Then (7.1) ensures due to $|\langle x^*, \xi \rangle| \leq \|x^*\| \, \|\xi\|$,

$$\varepsilon_k \|(\xi, \eta^k)\| + \|x^*\| \, \|\xi\| \geq \|p^*\| \, \|\eta^k\| \quad \text{if } \|(\xi, \eta^k)\| < \delta_k \text{ and } p^k + \eta^k \in F(x^k + \xi). \tag{7.19}$$

Choose $t_k = \|\delta_k\|^2$ $(\to 0)$ and put $\eta^k := -t_k p^*$. Assume that any Lipschitz estimate $\|\xi\| \leq L\|\eta^k\|$ holds for certain solutions $\xi = \xi(k)$ to

$$\pi_k := p^k + \eta^k \in F(x^k + \xi).$$

Since $\|(\xi, \eta^k)\| < \delta_k$ follows for large k, (7.19) may be applied. If $x^* = 0$ this yields $\varepsilon_k \|(\xi, \eta^k)\| \geq \|p^*\| \, \|\eta^k\|$ which contradicts $p^* \neq 0$. If $x^* \neq 0$ then (7.19) yields, with every $\beta > 0$, that $\|\xi\| \geq \frac{\|p^*\| + \beta}{\|x^*\|} \|\eta^k\|$ for large k. This verifies $L \geq \frac{\|p^*\|}{\|x^*\|}$. □
Sufficiency:
Condition (7.17): As for Thm. 1, we verify (5.6) provided that (7.17) holds with $\lambda < L$: Given (p,x), choose some $\hat{x} \in \lambda B$ such that $\frac{\pi - p}{\|\pi - p\|} \in CF(x,p)(\hat{x})$. Then $x' \in S(p')$ holds for certain elements $p' = p + t(\pi - p) + o_1(t)$ and $x' = x + t\|\pi - p\|\hat{x} + o_2(t)$ where $t = t_k \downarrow 0$ and $o_1(t)/t \to 0, o_2(t)/t \to 0$. So we obtain for small $t = t_k$,

$$\begin{aligned} \text{dist}\,(x, S(p')) + L\|p' - \pi\| &\leq t\|\pi - p\|\lambda + \|o_2(t)\| + L(1-t)\|p - \pi\| + L\|o_1(t)\| \\ &\leq (\; t\lambda + L(1-t)\;)\; \|p - \pi\| + \|o(t)\| < L\|\pi - p\| \end{aligned}$$

which finishes the proof. □

The criterion (7.17) is known from [1].
Condition (7.18): If S is not pseudo-Lipschitz at z^0, then the proof of Thm. 2 shows that (5.8) holds for related sequences. This tells us by definition that $y^* \neq 0$ and $0 \in D^*F(\zeta^0)(y^*)$. Condition (7.18) excludes the existence of such y^*, hence it implies the Aubin property. □

The criterion (7.18) is known from [58] and also from [40] where the equivalent property of openess with linear rate has been investigated.

In consequence, the following stability conditions in terms of generalized derivatives are valid (for closed mappings in finite dimension).

Theorem 6 *It holds for* $z^0 \in \operatorname{gph} S$ *and* $F = S^{-1}$*:*
(i) The Aubin property of S is equivalent to each of the conditions (7.17), (7.18).
(ii) S is strongly Lipschitz iff $0 \in TF(\zeta^0)(\hat{x}) \Rightarrow \hat{x} = 0$.
(iii) S is upper Lipschitz iff $0 \in CF(\zeta^0)(\hat{x}) \Rightarrow \hat{x} = 0$.
(iv) For S to be lower Lipschitz, condition (7.16) is necessary, but not sufficient, in general. ◇

We emphasize once again that these facts were observed in many papers, e.g., [4, 43, 30, 76, 50, 36, 40, 58], and the statements have been modified for more general spaces, e.g., in [1, 61, 25, 41, 36]. Notice however that, under (ii) and (iii), local solvability is not ensured.

8 Fixed points and persistence of solvability

In finite dimension, the Aubin property of $S = F^{-1}$ at z^0 along with the identity

$$D_g G(\zeta^0) = D_g F(\zeta^0) \tag{8.1}$$

for $D_g = D^*$ ensures, by Thm. 6, local solvability of

$$p \in G(x). \tag{8.2}$$

We would like to show that upper and strong Lipschitz stability are similarly invariable if $CF = CG$ or $TF = TG$ coincide at ζ^0. But this is not true:

Example 5 Let F be the real, linear function $F(x) = x$, and define $G(x) = \{x\}$ if $(|x| = 1/k,\ k = 1, 2, \dots$ or $x = 0)$, and $G(x) = \emptyset$ otherwise. For both mappings at $(0,0)$, the $C-$ and $T-$derivative is just the identity, but $F^{-1} = F$ obeys the related stability property in contrast to $G^{-1} = G$. The related co-derivatives are:
$D^*F(0,0)(p^*) = \{p^*\},\ D^*G(0,0)(p^*) = \mathbb{R}$. ◇

Hence solvability of (8.2) does not only depend on CG or TG at the reference point and needs extra assumptions. In addition, solvability may disappear if $D_g F$ and $D_g G$ slightly differ at the reference point. On the other hand, solvability can be handled by the help of Thm. 5 and via standard fixed point techniques. The latter will be investigated now.

Suppose that the variation of F is given by a small loc. Lipsch. perturbation h as in Thm. 5

$$G = h + F$$

with small α and β. The inclusion (8.2), i.e., $p \in h + F$, leads us to fixed points after the setting (6.19), i.e.,

$$T_p(x) := S(p - h(x))$$

since

$$p \in h(x) + F(x) \quad \Leftrightarrow \quad x \in T_p(x). \tag{8.3}$$

The inner function $\gamma_p(x) = p - h(x)$ has Lipschitz rank α on $X_\delta = x^0 + \delta B$ for small $\delta > 0$. Moreover, if S is upper Lipschitz (or strongly Lipschitz) with rank λ then the estimate

$$\begin{aligned} T_{p,\,\varepsilon}(x) := (p^0 + \varepsilon B) \cap T_p(x) \quad &\subset x^0 + \lambda(\|p - p^0\| + \|h(x) - h(x^0)\|)B \\ &\subset x^0 + \lambda(\|p - p^0\| + \alpha\delta)B \end{aligned}$$

ensures for p near p^0 and small α, namely (e.g.) if

$$\|p - p^0\| < \tfrac{1}{2}\lambda^{-1} \min\{\varepsilon, \delta\} \ \text{ and } \ \alpha\lambda\delta < \tfrac{1}{2} \min\{\varepsilon, \delta\}, \tag{8.4}$$

that $T_{p,\,\varepsilon}$ maps the ball X_δ into itself.

Hence fixed point Theorems can be applied to verify solvability. We mention here only those approaches which are closely related to the stability notions of this paper.

(i) Studying the fixed points of (8.3) was the key idea in [68]: Apply Kakutani's Theorem to $T_{p,\,\varepsilon}$ if S is upper Lipschitz stable and has convex ranges.

(ii) If S is strongly Lipschitz stable, Banach's principle can be directly applied since $T_{p,\,\varepsilon}$ is contractive for $\alpha < \lambda^{-1}$. This was a crucial observation in [69].

(iii) Finally, under the Aubin property of S at z^0, solvability of (8.2) follows again from Thm. 5

In [36], Thm. 4.5 and Thm. 4.2, also perturbations by multifunctions h are allowed in view of (iii). In [6], [69], [48] and [36] the underlying spaces were Banach spaces. Recall that Thm. 5 then remains valid. In some of the mentioned papers, only the case of small C^1 functions h has been taken into account, but the proofs for small loc. Lipsch. functions use basically the same principles.

Persistence of upper Lipschitzian stability for more general variations G in (8.1) (where $D_g = C$), has been shown in [42] (S upper Lipschitz stable, S and G closed and convex-valed, $x^0 \in \operatorname{dom} G$). So one may summarize.

Theorem 7 *The Aubin property of S at z^0 as well as strong Lipschitz stability are persistent (at least) with respect to small perturbations h as in Thm. 5. The same holds true, in finite dimension, for upper Lipschitz stability, provided that $S = F^{-1}$ is closed and convex-valued.* ◇

Implicit mappings and invariances

Invariance w.r. to first-order approximation of involved functions:
In Thm. 7, one can put $F(x) = \mu(x) + \mathcal{N}(x)$ and $h(x) = \mu_L(x) - \mu(x)$ where $\mu_L(x) = \mu(x^0) + D\mu(x^0)(x - x^0)$ is a linearization of a C^1 function μ at a zero $x^0 \in F^{-1}(0)$. Then $p \in h(x) + F(x)$ means

$$p \in \mu_L(x) + \mathcal{N}(x) \tag{8.5}$$

and h becomes an arbitrarily small Lipschitz function in the sense of Thm. 5. This is a basic situation Thm. 7 can be applied to. Strong Lipschitz stability (which includes local solvability), the Aubin property as well as upper Lipschitz stability (which includes local solvability again), for convex-valued S is invariant with respect to replacing μ by its linearization. Moreover, derivative formulas for solution mappings then follow from the fixed point representation

$$x \in S(p - h(x)) = F^{-1}(p - h(x))$$

and can be computed, by the chain rules above, if (and only if) related derivatives for $\mathcal{N}$ are known. The validity of this equivalence between

$$p \in \mu(x) + \mathcal{N}(x) \text{ and } p \in \mu_L(x) + \mathcal{N}(x)$$

in view of being s.L.s. was shown by Robinson [69], 1980. In his paper, $\mathcal{N}$ had a particular (normal cone) structure, but the main proofs hold for any closed $\mathcal{N}$, too. Concerning the same principle for other stability notions, we refer to [68, 42, 9, 48, 35, 76, 36]. Notice however, that replacing μ by μ_L does not work in view of calmness: With $\mathcal{N} = \{0\}$ and $\mu = x^2$, calmness is violated at the origin though it holds true for $\mu_L \equiv 0$. This is a consequence of the possible "discontinuous change" of $S(p^0)$ when passing from μ to μ_L.

The invariance principle simplifies stability conditions only up to a certain level. If one cannot translate some condition including $\mathcal{N}$ or some "derivative" of $\mathcal{N}$ in original terms, nothing is known about stability of (8.5) as well. So the structure and description of $\mathcal{N}$ become important.

Invariance w.r. to second-order approximations inside $\mathcal{N}$:
If $\mathcal{N}$ has a concrete structure, defined by some function ψ and related systems of equations and inequalities or their polar cones, a similar invariance principle can be observed. In many stability conditions, ψ may be replaced by its *quadratic approximation* at the reference point, cf. e.g., [70, 64, 8, 76, 9, 36, 73]. This means for the original problem that the related stability is invariant w.r. to quadratic approximation of all involved functions near the reference point. Of course, this invariance principle fails if μ or ψ are not sufficiently smooth. Then also the stability results in [53, 54, 76, 50, 51] for solutions of optimization problems cannot be applied since second derivatives are decisive used.

However, depending on the stability we are aiming at, this principle may fail even if all involved functions are convex polynomials. A typical example will be presented now.

9 The form of concrete known stability conditions

To simplify we consider C^2 problems in finite dimension without equations (they make the statements only longer, not more difficult).

$$P(a,b): \quad \min f(x) - \langle a, x\rangle, \ s.t.\ x \in \mathbb{R}^n, \quad g_i(x) \le b \ \ \forall\, i = 1, ..., m. \tag{9.1}$$

Consider first *the map* $S(a,b)$ *defined by the KKT points* in Kojima's form (2.6). The *upper Lipschitz property at* $((0,0),(x^0,y^0))$ can be checked by solving the system

$$\begin{array}{lclr} D^2L_x(x^0,y^{0+})u & + & Dg(x^0)^{\mathsf T}\alpha & = 0,\\ Dg(x^0)\ u & - & \beta & = 0,\\ \alpha_i = 0 \ \text{ if } \ y_i^0 < 0, & & \beta_i = 0 \quad \text{if } \ y_i^0 > 0, & \end{array} \tag{9.2}$$

with variables $u \in \mathbb{R}^n$ and $(\alpha,\beta) \in \mathbb{R}^{2m}$ which have, in addition, to satisfy

$$\alpha_i\beta_i \ = \ 0, \quad \alpha_i \ \ge \ 0 \ \ge \ \beta_i \quad \text{ if } \ y_i^0 = 0. \tag{9.3}$$

Similarly, *strong Lipschitz stability s.L.s.* can be checked by solving (9.2) where (α,β) has, instead of (9.3), to satisfy the weaker condition

$$\alpha_i\beta_i \ge 0 \quad \text{ if } \ y_i^0 = 0. \tag{9.4}$$

In both cases, the related stability just means, that only the trivial solution exists.

Due to the C^2 hypothesis, here the Aubin property and s.L.s. coincide [8]. This fails to holds for $f \in C^{1,1}$ [48]. These and the following statemens remain true for variational inequalities (replace Df in Kojima's function by any function of related dimension and smootheness).

Next let S *denote the map of stationary points,* i.e.,
$S(a,b) = \{x \mid \exists y:\ (x,y)$ is a KKT point for $P(a,b)\}$, and let $x^0 \in S(0,0)$ be the crucial point.

If $LICQ$ is satisfied at x^0 (the active constraint- gradients form a linear independent system) then there is exactly one Lagrange multiplier $y(x,a,b)$ and the function $y(.)$ is loc. Lipsch.. So S and the map of KKT points are locally "lipeomorph"; the above characterizations remain true.

Hence let only $MFCQ$ be satisfied (without MFCQ or related conditions in section 1, nearly nothing is known).

Now, the upper Lipschitz property can be checked by solving a finite number of quadratic systems, each defined by first and second derivativs of f, g at the reference point. Such systems are not known for the Aubin property and strong stability. Recently, it has been shown that they do not exist (even for convex, polynomial problems), cf. [37]. Let (without loss of generality) $g(x^0) = 0$.

Theorem 8 *The stationary point map* S *is not strongly Lipschitz at* $(0,0,x^0)$ *iff*

$$\begin{array}{l} \textit{There exist } u \in \mathbb{R}^n \setminus \{0\} \textit{ and a Lagrange vector } \ y \textit{ to } (x^0,0,0) \textit{ such that}\\ y_i Dg_i(x^0)u \ = \ 0\ \forall i, \ \textit{and with certain } x^k \to x^0 \textit{ and } \alpha^k \in \mathbb{R}^m, \textit{ one has}\\ \alpha_i^k\ Dg_i(x^0)u \ge 0\ \forall i \quad \textit{and} \quad \lim_{k\to\infty} \sum_i \alpha_i^k\ Dg_i(x^k) = -D_x^2L(x^0,y)u. \quad \diamond \end{array} \tag{9.5}$$

Examples demonstrate that the limit condition cannot be replaced by a condition in terms of derivatives (for f, g at x^0) up to a fixed order.

By similar limits, the T-derivative of S can be "determined" and the Aubin property at $(0,0,x^0)$ can be characterized: The Aubin property is violated iff there is a nontrivial pair $(u^*,\alpha^*) \in \mathbb{R}^{n+m}$ such that, for some sequence

$$(p^k,x^k) \to (0,x^0)\ ,\ (p^k,x^k) \in \text{gph}\, S, \tag{9.6}$$

the following conditions hold true (with $p = (a,b)$ and $Y(p,x)$ = set of Kojima- Lagr. multipliers, cf. (2.6)):

$$\begin{array}{lll} Dg_i(x^k)u^* = 0 & \text{if } i \in J^+, & \text{i.e. if } y_i > 0 \text{ for some } y \in Y(p^k,x^k),\\ \alpha_i^* \le 0 \text{ and } Dg_i(x^k)u^* \le 0 & \text{if } i \in J^0, & \text{i.e. if } y_i = 0 \text{ for some } y \in Y(p^k,x^k),\\ \alpha_i^* = 0 & \text{if } i \in J^-, & \text{i.e. if } y_i < 0 \text{ for some } y \in Y(p^k,x^k) \end{array} \tag{9.7}$$

and

$$\|D_x^2 L(x^0, y^+)u^* + Dg(x^0)^T \alpha^*\| < \varepsilon_k \downarrow 0 \quad \forall y \in Y(p^k, x^k). \tag{9.8}$$

A proof and specializations can be found in [36], Thm. 8.42. After choosing an appropriate subsequence the index sets in (9.7) are fix. However, replacing the points (p^k, x^k) by $(0, x^0)$ violates the equivalence for nonlinear g.

10 Future research

(1) All these characterizations of stability do nothing say about the *topological properties* of the solution sets, in particular (and important due to possible characterizations via Kojima functions) if $F : \mathbb{R}^n \to \mathbb{R}^n$ is a loc. Lipsch. function. Up to now, only the following is known:

If F^{-1} obeys the Aubin property at the origin without being strongly Lipschitz then there is no continuous function $s = s(p)$ such that $s(p) \in F^{-1}(p)$ for all p in some nbhd of 0. Hence bifurcation is necessary. Such F exists: identify $z \in \mathbb{R}^2$ with a complex number and put $F(0) = 0$ and $F(z) = z^2/|z|$ for $z \neq 0$.

If F^{-1} obeys the Aubin property at the origin and has directional derivatives (i.e. $card\, CF(x, u) \equiv 1$), then $x = 0$ is necessarily an isolated solution of $F(x) = 0$, cf. [16]. After deleting the hypothesis $card\, CF(x, u) \equiv 1$, the same statement or counterexamples are unknown.

(2) For stability of B- space problems, the direct approach via algorithms seems to be most appropriate to decrease the gap between stability theory and practical aspects of applications (basically) by refinements of the algorithms for particular classes of problems. The most relevant (and simplest) classes are those where the graph of the mapping is a union of finitely many smooth manifolds. Nevertheless, algorithmic approaches can be used quite general, [49].

(3) Already for $\mathbb{R}^n$ problems, it would be a big step ahead to characterize subclasses of problems which permit more convenient conditions in Thm. 8. Up to now, this has been done (without requiring $LICQ$) only for problems having linear constraints with at most one quadratic exception, cf. [37].

References

[1] J.-P. Aubin and I. Ekeland. *Applied Nonlinear Analysis.* Wiley, New York, 1984.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

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

[4] F.H. Clarke. On the inverse function theorem. *Pacific Journal of Mathematics*, 64: 97–102, 1976.

[5] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983.

[6] R. Cominetti. Metric regularity, tangent sets and second-order optimality conditions. *Applied Mathematics and Optimization*, 21: 265–287, 1990.

[7] A. Dontchev. Local convergence of the Newton method for generalized equations. *Comptes Rendus de l'Acédemie des Sciences de Paris*, 332: 327–331, 1996.

[8] A. Dontchev and R.T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization*, 6:1087–1105, 1996.

[9] A. Dontchev and R.T. Rockafellar. Characterizations of Lipschitz stability in nonlinear programming. In A.V. Fiacco, editor, *Mathematical Programming with Data Perturbations*, pages 65–82, Marcel Dekker, 1998.

[10] A. Dontchev and R.T. Rockafellar. Regularity and conditioning of solution mappings in variational analysis. *Set Valued Analysis*, 12: 79–109, 2004.

[11] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47: 324–353, 1974.

[12] A.V. Fiacco. Introduction to Sensitivity and Stability Analysis. Academic Press, New York, 1983.

[13] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Caomplementary Problems, Vol I and Vol II.* Springer, New York, 2003.

[14] F. Facchinei and C. Kanzow. A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. *Math.Progr. B* 76, No 3: 493-512, 1997.

[15] A. Fischer Solutions of monotone complementarity problems with locally Lipschitzian functions. *Math.Progr. B*, 76, No 3: 513– 532, 1997.

[16] P. Fusek. Isolated zeros of Lipschitzian metrically regular R^n functions. *Optimization*, 49: 425–446, 2001.

[17] P. Fusek, D. Klatte and B. Kummer. Examples and Counterexamples in Lipschitz Analysis. *SIAM J. Control and Cybernetics*, 31 (3): 471–492, 2002.

[18] H. Gfrerer. Hölder continuity of solutions of perturbed optimization problems under Mangasarian-Fromovitz Constraint Qualification. in Parametric Optimization and Related Topics, Akademie-Verlag, eds J. Guddat et al. Berlin, 113–124, 1987.

[19] L.M. Graves. Some mapping theorems. *Duke Mathematical Journal*, 17: 11–114, 1950.

[20] A. Griewank. The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces. SIAM J. Numer. Anal. 24: 684–705, 1987.

[21] J.-B. Hiriart-Urruty and C. Lemaréchal. Convex Analysis and Minimization Algorithms I, II. Springer, New York, 1993.

[22] J.-B. Hiriart-Urruty, J.J. Strodiot and V. Hien Nguyen. Generalized Hessian matrix and second order optimality conditions for problems with $C^{1,1}$-data. Applied Mathematics and Optimization, 11: 43–56, 1984.

[23] R. Henrion and J. Outrata. A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications*, 258: 110–130, 2001.

[24] A.D. Ioffe. On sensitivity analysis of nonlinear programs in Banach spaces: The approach via composite unconstraint optimization. *SIAM J. on Optimization*, 4: 1–43, 1994.

[25] A.D. Ioffe. Metric regularity and subdifferential calculus. *Russ. Math. Surveys*, 55: 501–558, 2000.

[26] A.D. Ioffe and V.M. Tichomirov. Theory of Extremal Problems. Nauka, Moscow, 1974, in Russian.

[27] H.Th. Jongen, P. Jonker and F. Twilt. Nonlinear Optimization in R^n, I: Morse Theory, Chebychev Approximation. Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork, 1983.

[28] H.Th. Jongen, P. Jonker and F. Twilt. Nonlinear Optimization in R^n, II: Transversality, Flows, Parametric Aspects. Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork, 1986.

[29] H.Th. Jongen, D. Klatte and K. Tammer. Implicit functions and sensitivity of stationary points. *Mathematical Programming*, 49:123–138, 1990.

[30] A. King and R.T. Rockafellar. Sensitivity analysis for nonsmooth generalized equations. *Mathematical Programming*, 55: 341–364, 1992.

[31] D. Klatte. Nonlinear optimization under data perturbations. *in Modern Methods of Optimization*, W.Krabs and J. Zowe, eds, Springer Verlag, New York 1992, 204–235.

[32] D. Klatte and B. Kummer. Generalized Kojima functions and Lipschitz stability of critical points. *Computational Optimization and Applications*, 13:61–85, 1999.

[33] D. Klatte and B. Kummer. Strong stability in nonlinear programming revisited. *Journal of the Australian Mathematical Society, Series B*, 40:336–352, 1999.

[34] D. Klatte and B. Kummer. Contingent derivatives of implicit (multi–) functions and stationary points. *Annals of Operations Research*, 101:313–331, 2001.

[35] D. Klatte and B. Kummer. Constrained minima and Lipschitzian penalties in metric spaces. *SIAM J. Optimization*, 13 (2): 619–633, 2002.

[36] D. Klatte and B. Kummer. *Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications.* Nonconvex Optimization and Its Applications. Kluwer Academic Publ., Dordrecht-Boston-London, 2002.

[37] D. Klatte and B. Kummer. *Strong Lipschitz Stability of Stationary Solutions for Nonlinear Programs and Variational Inequalities. SIAM Optimization*, 16: 96–119, 2005.

[38] D. Klatte and K. Tammer. Strong stability of stationary solutions and Karush-Kuhn-Tucker points in nonlinear optimization. *Annals of Operations Research*, 27:285–307, 1990.

[39] M. Kojima. Strongly stable stationary solutions in nonlinear programs. In S.M. Robinson, editor, *Analysis and Computation of Fixed Points*, pages 93–138. Academic Press, New York, 1980.

[40] A.Y. Kruger and B.S. Mordukhovich. Extremal points and Euler equations in nonsmooth optimization (in Russian). *Doklady Akad. Nauk BSSR*, 24: 684–687, 1980.

[41] A.Y. Kruger. Strict (ε, δ)-subdifferentials and extremality conditions. *Optimization*, 51: 539–554, 2002.

[42] B. Kummer. Generalized Equations: Solvability and Regularity. *Mathematical Programming Study*, 21: 199–212, 1984.

[43] B. Kummer. Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ optimization. *Journal of Optimization Theory and Applications*, 70:559–580, 1991.

[44] B. Kummer. An implicit function theorem for $C^{0,1}$-equations and parametric $C^{1,1}$-optimization. *Journal of Mathematical Analysis and Applications*, 158:35–46, 1991.

[45] B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In W. Oettli and D. Pallaschke, editors, *Advances in Optimization*, pages 171–194. Springer, Berlin, 1992.

[46] B. Kummer. On solvability and regularity of a parametrized version of optimality conditions. *ZOR Mathematical Methods of OR*, 41:215–230, 1995.

[47] B. Kummer. Parametrizations of Kojima's system and relations to penalty and barrier functions. *Mathematical Programming, Series B*, 76:579–592, 1997.

[48] B. Kummer. Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear programming. In A.V. Fiacco, editor, *Mathematical Programming with Data Perturbations*, pages 201–222. Marcel Dekker, New York, 1998.

[49] B. Kummer. Inverse functions of pseudo regular mappings and regularity conditions. Mathematical Programming, Series B, 88: 313–339, 2000.

[50] A.B. Levy. Solution sensitivity from general principles. *SIAM Journal on Optimization*, 40: 1–38, 2001.

[51] A.B. Levy. Lipschitzian Multifunctions and a Lipschitzian Inverse Mapping Theorem. Mathe-matics of Operations Research, 26: 105–118, 2001.

[52] A.B. Levy, R.A. Poliquin and R.T. Rockafellar. Stability of locally optimal solutions. *SIAM Journal on Optimization*, 10:580–604, 2000.

[53] A.B. Levy and R.T. Rockafellar. Sensitivity of solutions in nonlinear programs with nonunique multipliers. In D.-Z. Du, L. Qi, and R.S. Womersley, editors, *Recent Advances in Nonsmooth Optimization*, pages 215–223. World Scientific Press, Singapore, 1995.

[54] A.B. Levy and R.T. Rockafellar. Variational conditions and the proto-differentiation of partial subgradient mappings. *Nonlinear Analysis: Theory, Methods and Applications*, 26: 1951–1964, 1996.

[55] L. Lyusternik. Conditional extrema of functions. *Math. Sbornik*, 41: 390–401, 1934.

[56] O.L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17: 37–47, 1967.

[57] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. SIAM Journal on Control and Optimization, 15: 957–972, 1977.

[58] B.S. Mordukhovich. Approximation Methods in Problems of Optimization and Control (in Russian). *Nauka, Moscow*, 1988.

[59] B.S. Mordukhovich. Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions. Transactions of the American Mathematical Society. 340: 1–35, 1993.

[60] B.S. Mordukhovich. Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis. Transactions of the American Mathematical Society. 343: 609–657, 1994.

[61] B.S. Mordukhovich and Y. Shao. Mixed coderivatives of set-valued mappings in variational analysis. *Journal of Applied Analysis*, 4: 269–294, 1998.

[62] J. Outrata and M. Kočvara and J. Zowe. Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. *Kluwer Academic Publ.*, Dordrecht-Boston-London, 1998.

[63] R.A. Poliquin. Proto-differentiation of subgradient set-valued mappings. *Canadian Journal of Mathematics*, XLII, No. 3:520–532, 1990.

[64] R.A. Poliquin and R.T. Rockafellar. Proto–derivative formulas for basic subgradient mappings in mathematical programming. *Set–Valued Analysis*, 2: 275–290, 1994.

[65] R.A. Poliquin and R.T. Rockafellar. Prox–regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348: 1805–1838, 1995.

[66] D. Ralph and S. Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical Programming*, 70:159–172, 1995.

[67] D. Ralph and S. Scholtes. Sensivitity analysis of composite piecewise smooth equations. *Mathematical Programming, Series B*, 76:593–612, 1997.

[68] S.M. Robinson. Generalized equations and their solutions, Part I: Basic theory. *Mathematical Programming Study*, 10: 128–141, 1979.

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

[70] S.M. Robinson. Generalized equations and their solutions. Part II: Applications to nonlinear programming. *Mathematical Programming Study*, 19:200–221, 1982.

[71] S.M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17: 691–714, 1992.

[72] S.M. Robinson. Constraint nondegeneracy in variational analysis. *Mathematics of Operations Research*, 28: 201–232, 2003.

[73] S.M. Robinson. Variational conditions with smooth constraints: structure and analysis. *Mathematical Programming*, 97: 245–265, 2003.

[74] R.T. Rockafellar. Convex Analysis. Princeton University Press, Princeton, N.J., 1970.

[75] R.T. Rockafellar. First and second order epi-differentiability in nonlinear programming. *Transactions of the American Mathematical Society*, 207:75–108, 1988.

[76] R.T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, Berlin, 1998.

[77] S. Scholtes. Introduction to Piecewise Differentiable Equations. *Institut für Statistik und Mathematische Wirtschaftstheorie, Preprint No. 53*, Universität Karlsruhe, 1994.

[78] D. Sun and L. Qi. On NCP functions. *Computational Optimization and Appl.*, 13: 201-220, 1999

[79] L. Thibault. Subdifferentials of compactly Lipschitz vector-valued functions. *Annali di Matematica Pura ed Applicata*, 4:157–192, 1980.

[80] L. Thibault. On generalized differentials and subdifferentials of Lipschitz vector-valued functions. *Nonlinear Analysis: Theory, Methods and Applications*, 6:1037–1053, 1982.