Nonsmooth Equations in Optimization

Nonconvex Optimization and Its Applications

Volume 60

The titles published in this series are listed at the end of this volume.

Nonsmooth Equations in Optimization

Regularity, Calculus, Methods and Applications

by

Diethard Klatte

Institute for Operations Research and Mathematical Methods of Economics,
University of Zurich, Switzerland

and

Bernd Kummer

Institute of Mathematics, Faculty of Mathematics and Natural Sciences II,
Humboldt University Berlin, Germany

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-47616-9
Print ISBN: 1-4020-0550-4

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Contents

Introduction

Many questions dealing with solvability, stability and solution methods for variational inequalities or equilibrium, optimization and complementarity problems lead to the analysis of certain (perturbed) equations. This often requires a reformulation of the initial model being under consideration. Due to the specific of the original problem, the resulting equation is usually either not differentiable (even if the data of the original model are smooth), or it does not satisfy the assumptions of the classical implicit function theorem.

This phenomenon is the main reason why a considerable analytical instrument dealing with generalized equations (i.e., with finding zeros of multivalued mappings) and nonsmooth equations (i.e., the defining functions are not continuously differentiable) has been developed during the last 20 years, and that under very different viewpoints and assumptions.

In this theory, the classical hypotheses of convex analysis, in particular, monotonicity and convexity, have been weakened or dropped, and the scope of possible applications seems to be quite large. Briefly, this discipline is often called *nonsmooth analysis,* sometimes also *variational analysis.* Our book fits into this discipline, however, our main *intention* is to develop the analytical theory in close connection with the needs of applications in optimization and related subjects.

Main Topics of the Book

1. Extended analysis of Lipschitz functions and their generalized derivatives, including "Newton maps" and regularity of multivalued mappings.
2. Principle of successive approximation under metric regularity and its application to implicit functions.
3. Characterization of metric regularity for intersection maps in general spaces.
4. Unified theory of Lipschitzian critical and stationary points in $C^{1,1}$ optimization, in variational inequalities and in complementarity problems via a particular nonsmooth equation.
5. Relations between this equation and reformulations by penalty, barrier and so-called NCP functions.
6. Analysis of Newton methods for Lipschitz equations based on linear and

nonlinear approximations, in particular, for functions having a dense set of C^1 points.

7. Consistent interpretation of hypotheses and methods in terms of original data and quadratic approximations.
8. Collection of basic examples and exercises.

Motivations and Intentions

For sufficiently smooth functions, it is clear that many questions discussed in this field become trivial or have classical answers. Even the way of dealing with equations defined by nonsmooth functions seems to be evident:

(1) Define a reasonable derivative and prove a related inverse function theorem. Then, like in the Fréchet-differentiable case,
(2) derive statements about implicit functions, successive approximation and Newton's method,
(3) and develop conditions for characterizing critical points in extremal problems.

Of course, this calls for a deeper discussion.

First

of all, one has to specify the notion of a derivative. This should be a sufficiently nice function L that approximates the original one, say f, locally at least of first order like the usual linearization at some argument ξ in the Fréchet concept

$$f(x) - L(x) = o(x - \xi), \quad \|o(x - \xi)\|/\|x - \xi\| \to 0 \text{ as } x \to \xi.$$

However, there are many problems when going into the details.

Example 0.1 (piecewise linear **$o(\cdot)$–approximation**). Consider the real function $f(x) = x$ if $x < 0$, $f(x) = x^2$ if $x \geq 0$. For $\xi = 0$, the function $L_0(x) = \min\{x, 0\}$ satisfies $|f(x) - L_0(x)| \leq (x - \xi)^2 = o(x - \xi)$. Our "linearization" L_0 of f at the origin is nonlinear, but has still a simple piecewise linear structure. Taking $\xi \neq 0$ the linearization L_ξ of f should be the usual one, namely, $L_\xi(x) = f(\xi) + Df(\xi)(x - \xi)$. ◇

Evidently, in this example, we found a $o(\cdot)$–**approximation** L_0 of f near $\xi = 0$ which is simpler than the original function f. But in view of differentiation and inverse mappings, there arise already three new problems:

What about inverse maps of piecewise linear functions ?
What about continuity of derivatives or of "linearizations" in terms of ξ ?
Which kind of singularity (critical point) appears at the origin ?

Example 0.2 (no piecewise linear **$o(\cdot)$–approximation**). For the Euclidean norm on $\mathbb{R}^n$, one cannot find any piecewise linear $o(\cdot)$–**approximation** L_0 at the origin. ◇

The functions f and $\|\cdot\|$ of these examples are not only of academic interest because they are typical optimal value functions in parametric optimization:

$$f(x) = \min_y xy + x^2(1-y) \quad \text{with respect to } 0 \le y \le 1 \text{ (for } -1 < x < 1)$$
$$\|x\| = \max_y \langle x, y\rangle \quad \text{with respect to } \|y\| = 1.$$

They may occur, e.g., as objectives or as constraint functions in other optimization problems.

Next,
it may happen that one needs *different derivatives for different purposes*. To illustrate this we note that there exists a real, strictly monotonous directionally differentiable Lipschitz function f, such that

(i) f is C^1 on $\mathbb{R} \setminus N_D$, where N_D is a countable set.
(ii) The inverse f^{-1} is well-defined and globally Lipschitz.
(iii) Newton's method (to determine a zero of f) with start at any point $x \in X := \mathbb{R} \setminus N_D$ generates an alternating sequence and uses only points in X. Notice that X has *full* Lebesgue measure.

Concerning the construction and further properties of such a function f we refer to Chapter 12, Basic Example BE.1.

So, the existence of a Lipschitzian inverse on the one hand and local convergence of Newton's method on the other hand are different things. Indeed, we have to expect and to accept that there are generalized derivatives which allow (for certain nonsmooth functions) the construction of Newton-type solution methods without saying anything about uniqueness and Lipschitz behavior of the inverse, whereas other "derivatives", which characterize the inverse function, are rather inappropriate for Newton-type solution methods.

Moreover,
the power of the classical differential calculus lies in the possibility of *computing derivatives* for the functions of interest. The latter is based on several chain rules. Related rules for composed generalized derivatives of functions or multifunctions are often not true or hold in a weaker form only. Even for rather simple mappings in finite dimensional spaces, it may be quite difficult to determine the limits appearing in an actual derivative-definition. This means an increase in the technical effort.

In addition,
everybody has an idea about what tangency is or what a normal cone is. This had the effect that various more or less useful notions of generalized derivatives have been introduced in the literature, and many relations have been shown between them. Each of these derivatives has its own history and own motivation by geometrical aspects or by some statement, say by an application. However, these applications and motivations often play a second (or no) role in subsequent publications, which are devoted to technical refinements of the calculus, generalization and unification. So, the reader may easily gain the impression

that "nonsmooth analysis" is a graph the vertices of which are definitions of generalized derivatives and the edges are interrelations between them. It is hard to see that the graph is indeed something like a network of electric power because the lamps that can be switched on are hidden.

In the present book,
as far as general concepts are concerned, we motivate why this or another concept is appropriate (or not) for answering a concrete question, we develop a related theory and indicate possible applications in the context of optimization. We also try to use as few notions of generalized derivatives as possible (only those mentioned below), and we describe necessary assumptions mainly in terms of well–known upper and lower semi–continuity properties.

In this way, we hope that every reader who is familiar with basic topological and analytical notions and who is interested in the parametric behavior of solutions to equations and optimization problems (smooth or nonsmooth) or in the theory and methods of nonsmooth analysis itself will easily understand our statements and constructions.

As a basic general instrument, we apply Ekeland's variational principle.

A second tool consists in a slight generalization of successive approximation, which opens the same applications (by the same arguments) as successive approximation in the usual (single-valued) case, namely implicit function theorems and Newton-type solution methods.

Further, as a specific topic of our monograph, we use so-called Kojima-functions (having a nice, simple structure for analytical investigations) in order to characterize crucial points in variational models. For several reasons, but mainly in order to establish tools for studying variational problems with non-C^1 data and, closely related, stationary points in non–C^2 optimization, we summarize and extend the calculus of generalized derivatives for locally Lipschitz functions.

Finally, we connect generalized Newton-type methods with the continuity of (generalized) differentiability, as in the classical differentiable case; see the concept of *Newton maps.* Via perturbed Kojima systems, we establish relations to other standard optimization techniques, in particular, to penalty and barrier type methods.

However, the most important tool for understanding nonsmooth analysis with its various definitions and constructions, is the knowledge of several concrete functions and examples which show the difficulties and "abnormalities" in comparison with smooth settings. Such examples will be included in all parts of this monograph. The most important ones as well as answers to various exercises are collected in the last chapter.

We envision that our book is useful for researchers, graduate students and practitioners in various fields of applied mathematics, engineering, OR and economics, but we think that it is also of interest for university teachers and advanced students who wish to get insights into problems, potentials and recent developments of this rich and thriving area of nonlinear analysis and optimization.

Structure of the Book

In ***Chapter 1***, we start with some basic notation, in particular, with the presentation of certain desirable stability properties: pseudo- (or metric) regularity, strong and upper regularity. We try to find intrinsic conditions, equivalent or sufficient, which (as we hope) make the properties in question more transparent and indicate the relations to other types of stability.

In ***Chapter 2***, we present various conditions for certain Lipschitz properties of multivalued maps and the related types of regularity, we investigate interrelations between them and discuss classical applications as, e.g., (necessary) optimality conditions and "stability" in parametric optimization.

A great part of this chapter is devoted to pseudo-regularity of multifunctions in Banach spaces, where the use of generalized derivatives is avoided. This approach is based on the observation that the concepts of generalized derivatives which are usually applied for describing this important regularity-type (contingent derivatives as well as Mordukhovich's co-derivatives) lead us to conditions that are not necessary even for level set maps of monotone Lipschitzian functionals in separable Hilbert spaces, cf. Example BE.2. Therefore, we present characterizations which directly use Ekeland's variational principle as well as the family of assigned inverse functions. They allow characterizations of pseudo-regularity for the intersection of multifunctions and permit weaker assumptions concerning the image- and pre-image space as well.

In particular, we reduce the question of pseudo-regularity to the two basic classical problems:

(i) Show the existence of solutions to an equation after small constant perturbations, i.e., provided that $f(x) = y$ and $\|y' - y\|$ is small, verify that some x' satisfying $f(x') = y'$ exists.

(ii) Estimate the distance $\|x' - x\|$ for some solution x' in a Lipschitzian way, i.e., show that there is x' with $f(x') = y'$ such that $\|x' - x\| \leq L\|y' - y\|$.

Pseudo-regularity requires that, for certain neighborhoods U and V of x^0 and y^0, respectively, one finds a constant L such that both requirements can be satisfied whenever $x \in U$ and $y, y' \in V$.

We will demonstrate that, under weak hypotheses, it is enough to satisfy (i) and (ii) for all $x \in U$, $y \in V$ and for y' satisfying $0 < \|y' - y\| < \delta_{x,y}$, where $\delta_{x,y}$ is some constant depending on x and y.

Chapter 3 is devoted to characterizations of regularity by the help of (generalized) derivatives and may be seen as justification of the derivatives investigated in the current book.

We also intend to motivate why the regularity concepts introduced in the first two chapters are really important. In particular, this will be done in ***Chapter 4*** by showing persistence of regularity with respect to small Lipschitzian perturbations which has several interesting consequences (e.g. in Section 11.1). Note that we do not aim at presenting a complete perturbation theory for optimization problems and nonsmooth equations, our selection of results is subjective and essentially motivated by the applications mentioned above.

Many general regularity statements can be considerably sharpened for closed multifunctions in finite dimension and for continuous or locally Lipschitz function sending $\mathbb{R}^n$ into itself. So ***Chapter 5*** is devoted to specific properties of such mappings and functions where we pay attention to statements that are mainly of *topological* nature and independent on usual derivative concepts.

As an essential tool for locally Lipschitz functions, we apply here Thibault's limit sets. In contrast to Clarke's generalized Jacobians, the latter provide us with sufficient and necessary conditions for strong regularity and, even more important, they satisfy intrinsic chain rules for inverse and implicit functions.

Basic tools for dealing with several generalized derivatives will be developed in ***Chapter 6***. Our calculus of generalized derivatives includes chain rules and mean-value statements for contingent derivatives and Thibault's limit sets under hypotheses that are appropriate for critical point theory of optimization problems, where the involved problem-functions are not necessarily C^2. Here, we write coderivatives by means of contingent derivatives (which will be computed in Chapter 7), and we also introduce some derivative-like notion called *Newton function.* It represents linear operators that are of interest in relation to Newton-type solution methods for Lipschitzian equations and describes, in a certain sense, continuous differentiability for non-differentiable functions. Derivatives for so-called semismooth functions are included in this approach.

Chapter 7 is devoted to stable solution behavior of *generalized Kojima-functions*. By this approach, we cover in a unified way Karush-Kuhn-Tucker points and stationary points in parametric optimization, persistence and stability of local minimizers and related questions in the context of generalized equations, of complementarity problems and equilibrium in games, as well. The notation Kojima-function has its root in Kojima's representation of Karush-Kuhn-Tucker points as zeros of a particular nonsmooth function. We will see that basic generalized derivatives of such functions can be determined by means of usual chain rules. The properties of these derivatives determine, in a clear analytical way (based on results of Chapter 6), the stable behavior of critical points. In contrast to descriptions of critical points by generalized equations, our approach via Lipschitz equations has three advantages:

(i) The invariance of domain theorem and Rademacher's theorem may be used as additional powerful tools (not valid for multifunctions),

(ii) The classical approach via generalized equations is mainly restricted to systems of the type $f(x) \in \Gamma(x)$ where f varies in C^1 and the multifunction Γ is fixed. This means for optimization problems: The involved functions are C^2, and the perturbed problem has to have the same structure as the original one. By our approach, for instance, stationary points of the original problem and of assigned penalty or barrier functions may be studied and estimated as zeros of the same perturbed Kojima function (even for involved $C^{1,1}$ **-functions).**

(iii) The necessary approximations of the multifunction Γ - when speaking about generalized equations - are now directly determined by the type of derivative we are applying to the assigned Lipschitz equation.

In ***Chapter 8***, the regularity characterizations for zeros of generalized Kojima functions are applied to critical points, stationary solutions and local minimizers of parametric nonlinear $C^{1,1}$ programs in finite dimension, the specializations to the case of C^2 data – which is well-studied in the optimization literature – are explicitly discussed. In particular, we present second order characterizations of strong regularity, pseudo-regularity and upper Lipschitz stability in this context, give geometrical interpretations, and derive representations of derivatives of (stationary) solution maps. Finally, Taylor expansion formulas for critical value functions are obtained.

In ***Chapter 9***, we regard generalized derivatives of other mappings that are important for the analysis of optimization models, namely of

(i) positively homogeneous functions g and
(ii) Clarke subdifferentials $\partial_c f$ for the maximum f of finitely many C^2 functions F^k.

In particular, $g(u) = f'(x; u)$ may be a directional derivative or a so-called NCP-function, used for rewriting complementarity problems in form of nonsmooth equations. We study the latter more extensively in order to show how the properties of g determine the behavior of first and second order methods for solving the assigned equations and how related iteration steps can be interpreted in an intrinsic way. The simple derivative $D^\circ g$, defined below, plays an essential role in this context. In view of (ii), it turns out that $C\partial_c f$ (which may be seen as a proto-derivative, too) depends on the first and second derivatives of the functions F^k at the reference point only. We will determine the concrete form of $C\partial_c f$ in a direct way and establish the relations to C-derivatives of generalized Kojima-functions.

Solution methods for general Lipschitzian equations are the subject of ***Chapter 10***. Here, we summarize crucial conditions for superlinear convergence, based on linear and nonlinear auxiliary problems and present typical examples. In this chapter, our subsequent definitions of Newton maps, derivative D° and of locally PC^1**-functions** will be justified from the algorithmic point of view. Moreover, the relations between the regularity conditions, needed for Newton's method, as well as upper, pseudo- and strong regularity shall be clarified.

In ***Chapter 11***, (generalized) Newton methods will be applied in order to determine Karush-Kuhn-Tucker points of C^2-optimization problems. Depending on the reformulation as (nonsmooth) equation $F(z) = 0$ (via NCP- or Kojima-functions) and on the used generalized derivative "*DF*" as well, we formulate the related Newton steps in terms of assigned SQP- models and of (quadratic) penalty and (logarithmic) barrier settings. The connection of these different solution approaches becomes possible by considering the already mentioned perturbed Kojima functions and by studying the properties of their zeros. Taking the results of Chapter 4 into account, one obtains Lipschitz estimates for solutions, assigned to different methods of the mentioned type. From Chapter 10 it is obvious that the C^2-assumptions are only important for the interpretations in terms of quadratic problems, not for solving $F(z) = 0$ according to Chapter 10.

Chapter 12 contains *Basic Examples* which are used throughout the book at several places, while all numbered *Exercises* occurring in the first 11 chapters are once more compiled, now accompanied with the answers. In the ***Appendix,*** we prove some known basic tools for convenience of the reader.

Acknowledgements

We would like to thank several colleagues for stimulating discussions and valuable hints during the work on this project, in particular Asen Dontchev, Helmut Gfrerer, Peter Kall, Janos Mayer, Jiri Outrata, Stefan Scholtes and Alexander Shapiro, but especially Peter Fusek, Bert Jongen and Alexander Kruger, who in addition read parts of the manuscript.

Our thanks go to Jutta Kerger for her linguistic support and to Ramona Klaass who typed a great part of the manuscript.

We are grateful to Panos Pardalos for welcoming our work into the series *Nonconvex Optimization and Its Applications* and to John Martindale from *Kluwer* for his editorial help.

List of Results

Introduction

1. Basic Concepts

2. Regularity and Consequences

3. Characterizations of Regularity by Derivatives

4. Nonlinear Variations and Implicit Functions

5. Closed Mappings in Finite Dimension

6. Analysis of Generalized Derivatives

7. Critical Points and Generalized Kojima–Functions

8. Parametric Optimization Problems

9. Derivatives and Regularity of Further Nonsmooth Maps

10. Newton's Method for Lipschitz Equations

11. Particular Newton Realizations and Solution Methods

12. Basic Examples and Exercises

Appendix

Basic Notation

(in the order of first occurrence in the text)

Section 1.1

B_X (or B): the closed unit ball in X
B_X° (or B°): the open unit ball in X
X^*: dual of X
$X \times Y$ (or (X, Y)): product of sets X, Y
$\langle \cdot, \cdot \rangle$: canonical bilinear form of $X^* \times X$
$\mathbb{R}$: the reals
$C + rD$: Minkowski operations for $C, D \subset X$, $r \in \mathbb{R}$
$a + rD := \{a\} + rD$
$x + rB := \{x' \in X \mid d_X(x, x') \leq r\}$: convention even if (X, d_X) is a metric space
bd M, cl M, int M, conv M: boundary, closure, interior, convex hull of M
$(x, D) := \{(x, d) \mid d \in D\}$, similarly: $\langle x, D \rangle$, $D \cdot x$, etc.
dist $(x, A) := \inf_{a \in A} d_X(x, a)$ point-to-set distance in a metric space (X, d_X)
near x: (the statement holds) in some neighborhood of x
$f \in C^{0,1}(X, Y)$: f is a locally Lipschitz function from X to Y
$f \in C^1$: f has continuous first (Fréchet-) derivatives
$f \in C^{1,1}$: f has locally Lipschitz first (Fréchet-) derivatives
$f \in C^2$: has continuous first and second (Fréchet-) derivatives
$r^+ := \max\{r, 0\}$, $r^- := \min\{r, 0\}$ for $r \in \mathbb{R}$
r^+, r^-: defined componentwise for $r \in \mathbb{R}^m$
$\mathbb{R}^m$: the real m-vectors
$o(\cdot)$, $O(\cdot)$: o-type and O-type functions

Section 1.2

$F : X \rightrightarrows Y$: multi-valued map (multifunction) from X to Y
gph F: graph of F
dom F: domain of F
$F(A)$: image of A under F
$F^{-1} : Y \rightrightarrows X$: inverse multifunction of F

$CF(x,y)(u)$: contingent derivative of F at $(x,y) \in \operatorname{gph} F$ in direction $u \in X$
$TF(x,y)(u)$: Thibault's limit set of F at $(x,y) \in \operatorname{gph} F$ in direction $u \in X$
$D^*F(x,y)(v^*)$: coderivative of F at $(x,y) \in \operatorname{gph} F$ in direction $v^* \in Y^*$
$\rightharpoonup^*$: the weak* convergence
$F'(x;u)$: (one-sided) directional derivatives of $F : X \to Y$ at x in direction u
$f^c(x^0;u)$: Clarke's directional derivative of a functional f at x^0 in direction u
$\partial f(x^0)$: (usual convex) subdifferential of a functional f at x^0
$\partial_c f(x^0)$: Clarke-subdifferential of a functional f at x^0

Section 1.3

$Df(x)$: (Fréchet-) derivative of f at x
$D^2f(x)$: second (Fréchet-) derivative of f at x
$\partial_\circ f$: B-subdifferential for $f : \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitz
∂f: Clarke's generalized Jacobian for $f : \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitz
$D^\circ f$: another subdifferential
PC^1: piecewise C^1
$f = PC^1[f^1,\dots,f^N]$: f is a PC^1 function constructed by $f^1,\dots,f^N \in C^1$
NCP: nonlinear complementarity problem

Sections 1.4–1.6

MFCQ: Mangasarian-Fromovitz constraint qualification
l.s.c.: lower semicontinuous
u.s.c.: upper semicontinuous
$\limsup M_k$: upper Hausdorff limit of a set sequence $\{M_k\}$
$\liminf M_k$: lower Hausdorff limit of $\{M_k\}$
$\limsup_{y\to y^0} S(y), \liminf_{y\to y^0} S(y)$: upper and lower limits of a multifunction S
CLM set: complete local minimizing set
epi g: epigraph of a functional g

Section 2.1

$p \doteq S(y^0, X^0)$: p describes S near (y^0, X^0)
$\partial_n f(x^0) := \partial f'(x^0;\cdot)(0)$
$\mathbb{R}^m_+$: the non-negative real m-vectors
$A^\top$: A transposed
$D_x h$ or h_x: partial (Fréchet-) derivative of h with respect to x
$D_{xy} h$ or h_{xy}: partial second derivative of h with respect to x and y

Section 2.2

$E_f(p)$: set of all local Ekeland points with factor p
dim Y: dimension of Y
$\mathbb{R}^m_- := -\mathbb{R}^m_+$

$\text{DIST}(x, A)$: Gfrerer's DIST function

Section 4.0

$\sup(g, \Omega) := \sup\{\|g(x)\|_Y \mid x \in \Omega\}$
$\text{Lip}\,(g, \Omega) := \inf\{r > 0 \mid \|g(x) - g(x')\|_Y \leq r d(x, x') \;\; \forall x, x' \in \Omega\}$
$f \circ g$: composition of f and g

Section 6.1

$C(z, Z)$: contingent cone (Bouligand cone)
$T_c(z, Z)$: Clarke's tangent cone
$T(z, Z)$: tangent cone related to the Thibault derivative
$\pi(x, P)$: projection map
$N_P(x)$: map of normals

Sections 6.3–6.6

$A \in \text{ex}\, M$: A is exposed element of M
$T_x h$: partial Thibault derivative of h with respect to x
$f \in \text{loc}PC^1$: f is a locally PC^1 function
ker G: kernel of G

Sections 7.1–7.3

$F = (F_1, F_2, F_3) = NM$: generalized Kojima-function
LICQ: Linear Independence constraint qualification
SMFCQ: strict Mangasarian-Fromovitz constraint qualification
$\mathcal{R}_T(y^0) := \{\ r \in [0,1]^m \mid r_i = 1 \text{ if } y_i^0 > 0,\ \ r_i = 0 \text{ if } y_i^0 < 0\ \}$

$$\mathcal{R}_C(y^0, v) := \left\{ r \in \{0,1\}^m \;\middle|\; \begin{array}{l} r_i = 1 \text{ if } y_i^0 > 0 \text{ or if } y_i^0 = 0,\ v_i \geq 0 \\ r_i = 0 \text{ if } y_i^0 < 0 \text{ or if } y_i^0 = 0,\ v_i < 0 \end{array} \right\}$$

$$\mathcal{J}_T(y^0) := \left\{ \ (\alpha, \beta) \in \mathbb{R}^{2m} \;\middle|\; \begin{array}{rcll} \alpha_i \beta_i & \geq & 0 & \text{if } y_i^0 = 0 \\ \beta_i & = & 0 & \text{if } y_i^0 > 0 \\ \alpha_i & = & 0 & \text{if } y_i^0 < 0 \end{array} \right\}$$

$\mathcal{J}_C(y^0) := \{(\alpha, \beta) \in \mathcal{J}_T(y^0) \mid \alpha_i \geq 0 \geq \beta_i \text{ if } y_i^0 = 0\ \}$
$Q_T(u) := T_x F_1(s^0)(u)$ or $:= T_x[D_x L](s^0)(u)$, respectively
$Q_C(u) := C_x F_1(s^0)(u)$ or $:= C_x[D_x L](x^0, y^0)(u)$, respectively
U_h, U^T, $U^C(y^0)$, U^0: several critical cones
$K^T(u)$, $K^C(u)$: cones related to the T– and C–stability systems, respectively

Section 8.1

$I^0(y) := \{\, i \mid y_i = 0\, \}$ for $y \in \mathbb{R}^m$

$I^+(y) := \{ i \,|\, y_i > 0 \}$ for $y \in \mathbb{R}^m$
$I^-(y) := \{ i \,|\, y_i < 0 \}$ for $y \in \mathbb{R}^m$
$\widetilde{S}(\cdot)$ or $S(\cdot)$: critical point map
$\widetilde{X}(\cdot)$ or $X(\cdot)$: stationary solution map
$\widetilde{Y}(\cdot)$ or $Y(\cdot)$: multiplier map

Section 8.2

(SOC): second-order condition
(SSOC): strong second-order condition

Section 9.2

pNCP: cone of pNCP functions

Chapter 1

BASIC CONCEPTS

1.1 Formal Settings

Given a (real) normed space X, we denote by B_X and B_X° the *closed and open unit ball,* respectively. If the space is obvious, we omit the subscript. The normed spaces under consideration are always *real* normed spaces. The canonical bilinear form on the product space $X^* \times X$ is denoted by $\langle \cdot, \cdot \rangle$, where X^* denotes the dual of X.

For $C, D \subset X$ and $r \in \mathbb{R}$, we write $C + rD := \{c + rd \,|\, c \in C,\ d \in D\}$ in the sense of the *Minkowski operations.* We also often identify a set consisting of a single element with its element. So, for $a \in X$, $r \in \mathbb{R}$ and $D \subset X$, we write $a + rD$ instead of $\{a\} + rD$. In particular, $x + rB$ is the closed ball with centre x and radius r. This notation will be *also used for metric spaces.* In the suitable context, we denote by $\operatorname{bd} M$, $\operatorname{cl} M$, $\operatorname{int} M$, $\operatorname{conv} M$ the *boundary, closure, interior and the convex hull* of a given set M, respectively.

For compact writing sets which result from certain operations, we use symbols of the kind (x, D), $\langle x, D\rangle$, $D \cdot x$ etc. to denote the sets $\{(x,d) \,|\, d \in D\}$, $\{\langle x, d\rangle \,|\, d \in D\}$, $\{Q \cdot x \,|\, Q \in D\}$ etc. (for some set D under consideration and in a well–defined setting).

Given a metric space $(X, d_X(\cdot, \cdot))$, the *point–to–set distance* is denoted by $\operatorname{dist}(x, A) = \inf_{a \in A} d_X(x, a)$ with the convention $\operatorname{dist}(x, \emptyset) = \infty$.

We say that some statement holds *near* x if it holds for all x' in some neighborhood of x.

For metric spaces X and Y, we indicate by the symbol $f \in C^{0,1}(X, Y)$ that $f : X \to Y$ is a locally Lipschitz function, i.e., for each $x \in X$, there are a neighborhood $U \ni x$ and a constant L such that $d_Y(f(x'), f(x'')) \le L\, d_X(x', x'')\ \forall x', x'' \in U$. The constant L *is* said to be a *Lipschitz rank* (or *Lipschitz modulus*) of f near x. For Banach spaces X and Y, $f \in C^1(X, Y)$ [$f \in C^{1,1}(X, Y)$] indicates that $f : X \to Y$ is a function having continuous [locally Lipschitz] first Fréchet-derivatives. Similarly, $f \in C^2(X, Y)$ means that f is a function having continuous first and second Fréchet-derivatives. An optimization problem defined by C^k functions is said to be a ***C^k-problem.***

Often, we will assign, to some sequence of real t_k, certain elements $x^k \in M \subset X$. To indicate that the elements $x^k \in M$ converge to x, we write $x^k \to x$ in M. In order to avoid unnecessary indices, we will also speak of sequences of real, converging $t \to 0$ and assigned points $x_t \in X$. So the symbol t is *not reserved* for a continuous quantity, a priori.

For real r, we put $r^+ = \max\{0, r\}$ and $r^- = \min\{0, r\}$. For $r \in \mathbb{R}^m$, we define r^+ and r^- componentwise.

As usually, o-type functions $o(\cdot)$ are assumed to satisfy $o(u)/\|u\| \to 0$ as $u \to 0$ and $o(0) = 0$, while $O(\cdot)$ denotes a vanishing function $O(u) \to 0$ as $u \to 0$ and $o(0) = 0$.

1.2 Multifunctions and Derivatives

The symbol $F : X \rightrightarrows Y$ says that F is a *multi-valued map (multifunction)*, defined on X with $F(x) \subset Y$.

We abbreviate: $\operatorname{gph} F = \{(x, y) | y \in F(x),\ x \in X\}$, the *graph* of F, $\operatorname{dom} F = \{x \in X | F(x) \neq \emptyset\}$, the *domain* of F, and $F(A) = \cup_{a \in A} F(a)$, the *image* of $A \subset X$. Most of the multifunctions considered in this monograph will assign, to certain parameters, feasible sets of optimization problems or solutions of equations. If, for some neighborhood U of x, $F(U)$ is contained in a compact (bounded) set C, then F is said to be *locally compact* (*locally bounded*) near x. If $\operatorname{gph} F$ is closed in the product space $X \times Y$, then F is said to be *closed.* The *inverse* $F^{-1} : Y \rightrightarrows X$ is given by $F^{-1}(y) = \{x \in X | y \in F(x)\}$. For normed spaces X and Y and $(x, y) \in \operatorname{gph} F$, we associate with F the following maps:

> $CF(x, y) : X \rightrightarrows Y$, defined by $v \in CF(x, y)(u)$ if there are certain (discrete) $t = t_k \downarrow 0$ and assigned elements $(u_t, v_t) \to (u, v)$ such that $y + tv_t \in F(x + tu_t)$.
>
> $TF(x, y) : X \rightrightarrows Y$, defined by $v \in TF(x, y)(u)$ if there are certain (discrete) $t = t_k \downarrow 0$, assigned points $(x_t, y_t) \in \operatorname{gph} F$ with $(x_t, y_t) \to (x, y)$ and elements $(u_t, v_t) \to (u, v)$ such that $y_t + tv_t \in F(x_t + tu_t)$.
>
> $D^*F(x, y) : Y^* \rightrightarrows X^*$, defined by $u^* \in D^*F(x, y)(v^*)$ if there are certain (discrete) $t = t_k \downarrow 0$, $r_t > 0$, assigned points $(x_t, y_t) \to (x, y)$ in $\operatorname{gph} F$ and dual elements $(u_t^*, v_t^*) \rightharpoonup^* (u^*, v^*)$ in $X^* \times Y^*$ such that $\langle u_t^*, \xi\rangle + \langle v_t^*, \eta\rangle \leq t\|\xi\|_X + t\|\eta\|_Y$ if $\|\xi\|_X + \|\eta\|_Y < r_t$ and $(x_t + \xi, y_t + \eta) \in \operatorname{gph} F$,

where $\rightharpoonup^*$ is the weak* convergence.

Notice that $0 \notin D^*F(x, y)(v^*)$ is an existence condition: For all sequences $t = t_k \downarrow 0$, $r_t \downarrow 0$, $(x_t, y_t) \to z$ in $\operatorname{gph} F$ and $(u_t^*, v_t^*) \rightharpoonup^* (0, v^*)$ there are ξ_t, η_t with $\|\xi_t\| + \|\eta_t\| < r_t$ and $(x_t + \xi_t, y_t + \eta_t)$ in $\operatorname{gph} F$ such that, for sufficiently large k, $\langle u_t^*, \xi_t \rangle + \langle v_t^*, \eta_t \rangle > t\|\xi_t\| + t\|\eta_t\|$.

The mapping

> $CF(x, y)$ is the *contingent derivative* [AE84], also called *graphical derivative* or *Bouligand derivative* (since its graph is the contingent cone introduced by Bouligand [Bou32]),
>
> $D^*F(x, y)$ is (up to a sign) the *coderivative* in the sense of Mordukhovich [Mor93], and
>
> $TF(x, y)$ is *Thibault's limit set,* it was defined in [RW98] and was called *strict graphical derivative* there.

Note that we prefer to use the name Thibault's limit set (or *Thibault derivative*) for $TF(x, y)$, since this derivative has been first considered (however, for $F \in C^{0,1}(X, Y)$ and with another notation) by Thibault [Thi80] and [Thi82]. To unify the terminology, we call all these mappings *generalized derivatives.*

Remark 1.1 (derivatives of the inverse). For each of these generalized derivatives, the symmetric definitions induce that the inverse of the derivative is just the derivative of the inverse at corresponding points.

As usually, we will say that a derivative is *injective* if the origin belongs only to the image of $u = 0$ or $v^* = 0$, respectively.

For functions F, we have $y = F(x)$ and may write $CF(x)$, $TF(x)$ and $D^*F(x)$. Nevertheless, the images of the derivatives as well as the pre-images $F^{-1}(y)$ may be multivalued or empty.

If the (one-sided) limit $\lim_{t\downarrow 0} t^{-1}(F(x + tu) - F(x))$ exists uniquely for a function F and all sequences $t \downarrow 0$, then it is called the *directional derivative* of F at x in direction u, and denoted by $F'(x; u)$.

Further, for $f : X \to \mathbb{R}$, *Clarke's directional derivative* of f at x^0 in direction $u \in X$ is defined by the usual limes superior

$$f^c(x^0; u) = \limsup_{t\downarrow 0,\, x\to x^0} t^{-1}(f(x + tu) - f(x)).$$

which is obviously finite for locally Lipschitz functions.

The set $\partial f(x^0)$ of all $x^* \in X^*$ such that

$$f(x^0 + u) \geq f(x^0) + \langle x^*, u \rangle \; \forall u \in X$$

is called the (usual convex) *subdifferential* of f at x^0.

The set $\partial_c f(x^0)$ of all $x^* \in X^*$ such that

$$f^c(x^0; u) \geq \langle x^*, u \rangle \; \forall u \in X$$

is called the *Clarke-subdifferential* of f at x^0. It coincides with the subdifferential of $f^c(x^0; .)$ at $u^0 = 0$, and $\partial_c f(x^0) = \partial f(x^0)$ holds for convex f.

1.3 Particular Locally Lipschitz Functions and Related Definitions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz.

Generalized Jacobians of Locally Lipschitz Functions

By Rademacher's theorem (for proofs see, e.g., [Fed69, Har79, Zie89]), the set

$$\Theta = \{x \in \mathbb{R}^n \mid \text{the Fréchet derivative of } f \text{ exists at } x\}$$

has full Lebesgue measure, i.e., $\mu(\mathbb{R}^n \setminus \Theta) = 0$. Moreover, for $x' \in \Theta$ and x' near x, the norm of $Df(x')$ is bounded by a local Lipschitz rank L of f. These facts ensure that the mapping $\partial_\circ f : \mathbb{R}^n \rightrightarrows \mathbb{R}^{mn}$ defined by

$$\partial_\circ f(x) = \{A \mid A = \lim Df(x') \text{ for certain } x' \to x,\ x' \in \Theta\},$$

has non-empty images. In addition, one easily sees that $\partial_\circ f$ is closed and locally compact. The same properties are induced for the map ∂f, defined by the *generalized Jacobian* $\partial f(x) = \operatorname{conv} \partial_\circ f(x)$ of f at x.

These observations, along with an elaborated calculus for ∂f which includes an inverse-function theorem, cf. Theorem 5.13 as well as close connections to (several) directional derivatives, in particular $\partial f(x) = \partial_c f(x)$ for real-valued f, are the fundamentals of F.H. Clarke's [Cla83] concept of nonsmooth analysis. The latter equation induces that, for $m = 1$ and convex f, there is no difference between the classical subdifferential of convex analysis $\partial f(x)$ and the generalized Jacobian $\partial f(x)$.

In the literature, the mapping $\partial_\circ f$ is often called the *B-subdifferential* and also denoted by $\partial_B f$.

Pseudo-Smoothness and D°f

Next let us copy Clarke's definition. We put

$$\Theta^\circ = \{x \in \mathbb{R}^n \mid Df \text{ exists and is continuous near } x\}.$$

and

$$D^\circ f(x) = \{A \mid A = \lim Df(x') \text{ for certain } x' \to x,\ x' \in \Theta^\circ\}.$$

Evidently, $D^\circ f(x) \subset \partial_\circ f(x) \subset \partial f(x)$. In contrast to the pair $(\partial_\circ f, \Theta)$, the pair $(D^\circ f, \Theta^\circ)$ fulfills $D^\circ f \equiv Df$ on the open set Θ°.

However, Θ° and $D^\circ f(x)$ may be empty, cf. the real Lipschitz function G in Example BE.0 where also $\partial G(x)$ is a constant, proper interval. If Θ° is *dense* in $\mathbb{R}^n$, we call f *pseudo-smooth.*

The function f of Example BE.1 obeys this property and satisfies additionally $m = n = 1$ as well as

$$Df(0) = 1,\ D^0 f(0) = \{\tfrac{1}{2}, 2\},\ \partial_\circ f(0) = \{\tfrac{1}{2}, 1, 2\} \neq \partial f(0) = [\tfrac{1}{2}, 2].$$

Piecewise C^1 Functions

The class of ***PC^1-functions*** (*piecewise C^1*) is defined in the following way. Given a locally Lipschitz function f from $\mathbb{R}^n$ to $\mathbb{R}^m$, one says that f belongs to PC^1 if there is a finite family of functions $f^s \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ such that, for all $x \in \mathbb{R}^n$, the set

$$I(x) = \{s \mid f(x) = f^s(x)\}$$

is not empty.

We will also write $f = PC^1[f^1, \ldots, f^N]$. The set $I(x)$ characterizes the set of *active* functions at x. The generalized Jacobian of f has the representations

$$\partial f(x) = \operatorname{conv}\{Df^s(x) \mid x \in \operatorname{cl}\operatorname{int} I^{-1}(s)\},$$

see [Sch94], and

$$\partial f(x) = \operatorname{conv}\{Df^s(x) \mid x \in I^{-1}(s) \text{ and } Df(x') = Df^s(x') \text{ for certain } x' \to x\},$$

see [Kum88a]. Note that $x \in \operatorname{int} I^{-1}(s)$ if and only if f coincides with f^s near x, so the first index set of related s may be smaller.

Note. If one defines **PC^1–functions** in the same way, but requiring the weaker assumption that f is only continuous (instead of being locally Lipschitz), then one obtains the same class of functions, see [Hag79]. ◇

Obviously, the maximum-norm of $\mathbb{R}^n$ is a PC^1 function, not so the Euclidean norm.

Every **PC^1-function** is pseudo-smooth since Θ^o contains the open and dense set $\Omega = \cup_s \operatorname{int} I^{-1}(s)$, cf. the proof of Lemma 6.17.

Convex functions are not necessarily pseudo-smooth, cf. Example BE.6. Nevertheless, in many applications, they are even PC^1.

NCP Functions

An *NCP function* is any function $G : \mathbb{R}^2 \to \mathbb{R}$ such that

$$G^{-1}(0) = \{(s,t) \in \mathbb{R}^2 \mid s \geq 0,\ t \geq 0,\ st = 0\}.$$

Such functions are connected with *nonlinear complementarity problems (NCPs):* Given $u, v : \mathbb{R}^n \to \mathbb{R}^n$ find $x \in \mathbb{R}^n$ such that

$$u_i(x) \geq 0,\ v_i(x) \geq 0,\ u_i(x)v_i(x) = 0 \quad (i = 1, \ldots, n).$$

Using G, the NCP can be written as an equation $f(x) = 0$ by setting

$$f_i(x) := G(u_i(x), v_i(x)).$$

We will say that the NCP is (strongly) monotone if

$$\langle u(y) - u(x), v(y) - v(x)\rangle \geq \lambda \|y - x\|^2 \ \forall x, y \in \mathbb{R}^n,$$

where $\lambda \geq 0$ ($\lambda > 0$) is a fixed constant. A *standard* NCP is defined by $v(x) = x$.

1.4 Definitions of Regularity

Our main subject is the equation

$$F(x) = 0, \text{ where } F : X \subset \mathbb{R}^n \to Y = \mathbb{R}^m \text{ is a locally Lipschitz function.}$$

Its inverse $S = F^{-1}$ is a multifunction with possibly empty images $S(y) \subset X$. We shall be interested in the local properties of $S = F^{-1}$ near a pair $(x^0, F(x^0))$.

Generally, we will speak of *regularity* whenever $F^{-1}(y)$ is non-empty for y near $y^0 := F(x^0)$. The *type* of regularity (*strong, pseudo, upper*) will be concerned with Lipschitz properties of F^{-1} only.

So it does not make any difference whether F is a function or a multifunction having images $F(x) \subset Y$ and the inverse $F^{-1}(y) = \{x \in X | y \in F(x)\}$. Moreover, the requirements related to F^{-1} make sense for any multifunction F acting between *metric* spaces X and Y. In particular, Y may be a subset of $\mathbb{R}^m$, e.g. $Y = F(X) = \operatorname{dom} F^{-1}$which already ensures that $F^{-1}(y)$is non–empty. Therefore, we present the corresponding definitions - as usually - in this generality.

Definitions of Lipschitz Properties

Let X and Y be metric spaces, $S : Y \rightrightarrows X$ and $(y^0, x^0) \in \operatorname{gph} S$.

(D1) The mapping S is said to be *pseudo-Lipschitz* (with rank L) at (y^0, x^0) if there are neighborhoods U and V of x^0 and y^0, respectively, such that, given any points $(y, x) \in (V \times U) \cap \operatorname{gph} S$ and $y' \in V$,

$$\textbf{there exist } x' \in S(y') \textbf{ satisfying } d_X(x', x) \leq L\, d_Y(y', y). \qquad (1.1)$$

(D2) Similarly, if U, V and L exist in such a manner that for $y' \in V$,

$$x' \in S(y') \cap U \Rightarrow d_X(x', x^0) \leq L\, d_Y(y', y^0), \qquad (1.2)$$

then S is called *locally upper Lipschitz* (briefly *locally u.L.)* at (y^0, x^0) with rank L.

In many papers, condition (1.1) is written in a weaker form, namely as

$$\operatorname{dist}(x, S(y')) \leq L\, d_Y(y', y) \;\; \forall y' \in V. \qquad (1.3)$$

Here, $\operatorname{dist}(x, S(y')) = \inf_{x' \in S(y')} d_X(x, x')$ is the point-to-set distance in X, as defined above. Having (1.3) one can satisfy (1.1) with any $L' > L$. In this sense, the conditions (1.1) or (1.3) are equivalent. We will prefer the condition (1.1).

If S is a function, then definition (Dl) simply claims Lipschitz continuity on some neighborhood of y^0.

The notion *pseudo-Lipschitz* was introduced in [Aub84, AE84], it is also called *Aubin property* [RW98].

It is well-known from Robinson's [Rob76a] work that a finite-dimensional system

$$g(x) \le y, \quad h(x) = z, \tag{1.4}$$

has, at $s^0 = (x^0, y^0, z^0)$ and for $(g, h) \in C^1(\mathbb{R}^n, \mathbb{R}^{m+k})$, a pseudo-Lipschitzian solution map $S(y, z)$ if and only if the *Mangasarian–Fromovitz constraint qualification* (MFCQ) [MF67] is satisfied:

(MFCQ) $Dh(x^0)$ has full rank and there is some u such that $Dh(x^0)u = 0$ and $g(x^0) + Dg(x^0)u < y^0$,

see also §2.2.4 below. The solution map S defined by (1.4) is crucial for various properties of an optimization problem $\min\{f(x) | x \in S(0, 0)\}$.

Regularity Definitions

Let $S = F^{-1}$ be the inverse of a given multifunction $F : X \rightrightarrows Y$.

If S is pseudo-Lipschitz at (y^0, x^0), then F is called *pseudo-regular* at (x^0, y^0). If, additionally, neighborhoods U and V of x^0 and y^0, respectively, exist in such a way that $U \cap F^{-1}(y)$ is single-valued for $y \in V$, then we call F *strongly regular* at (x^0, y^0).

Finally, if S is locally upper Lipschitz at (y^0, x^0) and $S(y') \cap U$ is non-empty for all $y' \in V$, then F is said to be *upper regular* at (x^0, y^0).

In every case, one says that L is a *rank of* (the related) *regularity*. To distinguish the defining neighborhoods assigned to different regular maps F and G at points (x^0, y^0) and (ξ^0, η^0) we write $U_F(x^0)$, $V_F(y^0)$ and $U_G(\xi^0)$, $V_G(\eta^0)$, respectively, and to quantify these neighborhoods, we denote by δ_{UF} (similarly δ_{VF}, δ_{UG}, δ_{VG}) some positive constant such that

$$x^0 + \delta_{UF} B \subset U_F(x^0) \tag{1.5}$$

is satisfied, where we recall the convention $x^0 + rB := \{x \in X | d(x, x^0) \le r\}$ if X is a metric space.

Pseudo-regularity means that, locally around (x^0, y^0), a Lipschitzian *error estimate* holds true. Having a solution x to $y \in F(x)$, one finds some x' satisfying the perturbed inclusion $y' \in F(x')$ with a (small) distance $d(x', x) \le L d(y', y)$. Identifying the mappings $F = S^{-1}$, S from (1.4), this ensures

$$U \cap S(y, z) \subset S(y', z') + L d((y', z'), (y, z)) B \quad \forall (y, z), (y', z') \in V.$$

Evidently, condition (D1) remains true after changing the point $(y^0, x^0) \in (V \times U) \cap \operatorname{gph} S$. So, (D1) is a property which concerns the Lipschitz behavior of S near (y^0, x^0). Moreover, as a direct application of the definition only, one sees that pseudo-regularity is persistent *with respect to composition* of maps, and upper regularity shows the same property after a natural modification.

Lemma 1.2 (composed maps).

(i) If $G : X \rightrightarrows Y$ and $F : Y \rightrightarrows Z$ are pseudo-regular at (x^0, y^0) and (y^0, z^0), respectively, then $H = F \circ G : X \rightrightarrows Z$ as $H(x) = F(G(x))$ is pseudo-regular at (x^0, z^0).

(ii) If G and F are upper regular at the given points, then $H = F \circ (G \cap \Omega) : X \rightrightarrows Z$ as $H(x) = F(G(x) \cap \Omega)$ is upper regular at (x^0, z^0) for sufficiently small neighborhoods Ω of y^0.

In addition, the following estimates hold. Let $U_G(x^0), V_G(y^0)$ and $U_F(y^0), V_F(z^0)$ be the assigned neighborhoods with related constants δ_{UG}, δ_{VG}, δ_{UF}, δ_{VF} according to (1-5), and let L_G, L_F be related ranks of regularity. Then, in both cases, $L_H = L_G L_F$ is a rank of regularity for H, and related neighborhoods may be defined as follows:

$$U_H(x^0) = U_G(x^0), \quad \Omega = V_G(y^0) \text{ and } V_H(z^0) = z^0 + rB^0$$

provided that $0 < r \le r_0 := \min\{\delta_{VF}, (3L_F)^{-1}\delta_{VG}, (L_G L_F)^{-1}\delta_{UG}\}$. ◇

Proof. (i) By the choice of r we ensured that

$$z^0 + rB \subset V_F(z^0), \; y^0 + 3L_F rB \subset V_G(y^0) \text{ and } x^0 + L_G L_F rB \subset U_G(x^0).$$

Let $z \in H(x)$, $x \in U_G(x^0)$ and $z, z' \in z^0 + rB$ be given. We show that some x' fulfills $z' \in H(x')$ and $d(x', x) \le L_G L_F d(z', z)$.

Since $z^0 \in F(y^0)$ and $z \in V_F(z^0)$, we find $y \in F^{-1}(z)$ in such a way that $d(y, y^0) \le L_F d(z, z^0) \le L_F r$. So we have $y \in V_G(y^0)$. Since $z' \in V_F(z^0)$, one finds some $y' \in F^{-1}(z')$ satisfying $d(y', y) \le L_F d(z', z) \le 2L_F r$. Hence,

$$y, y' \in y^0 + 3L_F rB \subset V_G(y^0).$$

Using $y^0 \in G(x^0)$ next, some $x \in G^{-1}(y)$ fulfills $d(x, x^0) \le L_G d(y, y^0) \le L_G L_F r$. This yields $x \in U_G(x^0)$. By pseudo-regularity of G we finally obtain the existence of $x' \in G^{-1}(y')$ satisfying $d(x', x) \le L_G d(y', y) \le L_G L_F d(z', z)$.

Therefore, H is pseudo-regular with neighborhoods $U_H = U_G(x^0)$, $V_H = z^0 + rB^0$ and rank $L_H = L_G L_F$.

(ii) Let $z \in z^0 + rB$. Since $z \in V_F(z^0)$ and F is upper regular, we have

$$\begin{aligned} \emptyset \ne Y(z) := F^{-1}(z) \cap U_F(y^0) \; &\subset \; y^0 + L_F d(z, z^0)B \\ &\subset \; y^0 + L_F rB \subset \Omega = V_G(y^0). \end{aligned}$$

Selecting $y \in Y(z)$, this ensures $y \in \Omega, d(y, y^0) \le L_F d(z, z^0)$ and, due to upper regularity of G and $y \in V_G(y^0)$,

$$\emptyset \ne G^{-1}(y) \cap U_G(x^0) \subset x^0 + L_G d(y, y^0)B \subset x^0 + L_G L_F d(z, z^0)B.$$

So every $x \in G^{-1}(y) \cap U_G(x^0)$ belongs to $H^{-1}(z) \cap U_G(x^0) \cap [x^0 + L_G L_F d(z, z^0)B]$. Since we restricted F to Ω, the points $y \in F^{-1}(z) \setminus \Omega$ do not belong to the image of $(G \cap \Omega)$. Therefore, upper regularity of H with rank $L_G L_F$ now follows from

$$H^{-1}(z) \cap U_G(x^0) = \cup_{y \in \Omega \cap Y(z)} G^{-1}(y) \cap U_G(x^0).$$

□

Functions and Multifunctions

Given any closed multifunction $F : X \rightrightarrows Y$ define $f(x,y) = \text{dist}((x,y), \text{gph } F)$, say with distance $d((x',y'),(x,y)) = \max\{d_X(x',x), d_Y(y',y)\}$ *on* $X \times Y$. Then, condition (Dl) for $S = F^{-1}$ becomes a typical implicit-function requirement for the (globally) Lipschitz function $f : X \times Y \to \mathbb{R}$, namely:

> Given $(x,y) \in U \times V$ with $f(x,y) = 0$ and $y' \in V$, there is some x' such that $f(x',y') = 0$ and $d_X(x',x) \le L d_Y(y',y)$.

Similarly, (D2) requires:

> For all $(x,y) \in U \times V$ with $f(x,y) = 0$, it holds $d_X(x,x^0) \le L d_Y(y,y^0)$.

Each multifunction $S : Y \rightrightarrows X$ is the *inverse* of the map $F : X \rightrightarrows Y$, $F = S^{-1}$. Thus, there is no principal difference therein whether we investigate F or F^{-1} and speak about pseudo-regularity or the pseudo-Lipschitz condition (Dl). However, in any case, our assumptions should concern the given mapping F,

If F is a *function,* then $S = F^{-1}$ satisfies

$$S(y) \cap S(y') = \emptyset \text{ if } y \neq y'. \tag{1.6}$$

Conversely, if S satisfies (1.6), then $F = S^{-1}$ is a function, defined on dom S^{-1}. This fact has consequences for extending statements concerning inverse functions to inverse multifunctions: *If (1.6) has been nowhere* used, *then the related statement on F^{-1} is immediately true for multivalued $F = S^{-1}$, too. On the other hand, one cannot expect to obtain specific results for inverse functions from the theory of multifunctions, as long as (1.6) has been not exploited.*

Example 1.3 (regularity for C^1 functions). If $F : X = \mathbb{R}^n \to Y = \mathbb{R}^n$ is a continuously differentiable function, then all these regularity definitions coincide - due to usual implicit function theorem - with the requirement $\det DF(x^0) \neq 0$. ◇

Example 1.4 (pseudo-regular, but not strongly regular). The complex function $F(z) = z^2/|z|$ for $z \neq 0$, F(0) = 0, is a Lipschitz function which is pseudo-regular and upper regular without being strongly regular at the origin. ◇

Example 1.5 (strong regularity for continuous functions). For a continuous function $F : \mathbb{R}^n \to \mathbb{R}^m$, strong regularity at (x^0, y^0) induces that F is a homeomorphism between certain neighborhoods $U \ni x^0$ and $V \ni y^0$. Hence, $m = n$ is necessarily true due to Brouwer's famous invariance of domain theorem. This is an essential fact being true for functions, but not for multifunctions. ◇

Example 1.6 (pseudo-regularity for linear operators). Let $F : X \to Y$ be a *linear operator* onto Y where X and Y are normed spaces. Pseudo-regularity now requires that, given y', x and $y = F(x)$, there is some x' such that $F(x') = y'$ and $\|x' - x\| \le L\|y' - y\|$. In other words, F^{-1} is bounded as a mapping in the factor space $X/F^{-1}(0)$. Conversely, one may say that pseudo-regularity is just a nonlinear, local version of this property. ◇

Example 1.7 (Graves-Lyusternik theorem). Let $F : X \to Y$ be continuously *differentiable near* x^0, X and Y be Banach spaces and $DF(x^0)X = Y$. Then, F is pseudo-regular at $(x^0, F(x^0))$ (for references, proof and modifications, see Chapter 4, Theorem 4.11). One may state that pseudo-regularity is the basic topological property of F^{-1} near $(x^0, F(x^0))$. ◇

Example 1.8 (subdifferential of the Euclidean norm). A relevant *multifunction* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ being strongly regular at (0,0): Take the subdifferential (in the sense of convex analysis) $F(x) = \partial f(x)$ of the Euclidean norm $f(x) = \|x\|$: Then,

$$F^{-1}(y) = \{x \mid x \textbf{ minimizes } f(\xi) - \langle y, \xi\rangle\} = \{0\} \quad \forall y, \|y\| < 1.$$

◇

1.5 Related Definitions

Let us recall some common notions concerned with multifunctions $S : Y \rightrightarrows X$ for metric spaces X and Y.

Types of Semicontinuity

If $x \in S(y)$ and $\text{dist}\,(x, S(y')) \to 0$ for each sequence $y' \to y$, then S is said to be *lower semicontinuous* (l.s.c.) at (y, x). In the situation of definition (Dl), there is even a Lipschitzian estimate

$$\text{dist}\,(x, S(y')) \leq L d_Y(y, y') \text{ for all } y' \text{ in some neighborhood } V(x, y) \text{ of } y.$$

In the latter case, S is called *Lipschitz l.s.c.* at (y, x) with rank L. If, given y, S is l.s.c. at all (y, x), $x \in S(y)$, then S is said to be *l.s.c. at* y.

If $\text{dist}\,(x', S(y)) \to 0$ for each sequence $y' \to y$, and arbitrary $x' \in S(y')$ then S is said to be *upper semicontinuous* (u.s.c.) at y.

This coincides with C. Berge's [Ber63] u.s.c. - definition if $S(y)$ is compact (see, e.g. [BGK+82, Chapter 2]), where S is called *Berge-u.s.c.* at y if for any open set $Q \supset S(y)$ there is some neighborhood O of y such that $S(y') \subset Q$ for all $y' \in O$.

The map S is said to be *Lipschitz u.s.c.* at y (with rank L) if

$$\text{dist}\,(x', S(y)) \leq L d_Y(y, y') \text{ for all } x' \in S(y')$$
$$\text{and all } y' \text{ in some neighborhood } V(y).$$

If S is a function, we will then also say that S is *pointwise Lipschitz* at y. In comparison with the *local upper Lipschitz* property (D2), one considers now the whole set $S(y')$ and does not claim that the elements of $S(y')$ converge to a single point as $y' \to y$.

If the requirements of definition (DI) are satisfied for $U = X$, then we have

$$\begin{array}{l}\operatorname{dist}(x, S(y')) \leq L d_Y(y, y') \text{ for all } x \in S(y)\\ \text{and all } y,\, y' \text{ in some neighborhood } V \text{ of } y^0.\end{array}$$

In this case, S is said to be *Lipschitz-continuous* around y^0.

Example 1.9 (F^{-1} is u.s.c., but not l.s.c.). Assign, to each $x \in \mathbb{R}^n$, the line-segment $F(x) = [\frac{1}{2}x, x] \in \mathbb{R}^n$. The inverse is

$$F^{-1}(y) = [y, 2y],$$

and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is pseudo-regular at (0,0).

Setting similarly $F(x) = [0, x]$, then F^{-1} becomes

$$F^{-1}(0) = \mathbb{R}^n \text{ and } F^{-1}(y) = \{\lambda y | \lambda \geq 1\} \text{ for } y \neq 0.$$

Now, F^{-1} is Lipschitz u.s.c. with each L at $y = 0$ as well as Lipschitz l.s.c. with $L = 1$ at the origin $(y, x) = (0, 0)$; but F^{-1} is not l.s.c. at $y = 0$. ◇

Finally, for any given sequence of sets $M_k \subset X$ ($k = 1,2,\ldots$), one defines

$$\begin{aligned}\limsup M_k &= \{x \mid \liminf_{k\to\infty} \operatorname{dist}(x, M_k) = 0\},\\ \liminf M_k &= \{x \mid \limsup_{k\to\infty} \operatorname{dist}(x, M_k) = 0\}.\end{aligned}$$

These sets are often called the *upper and lower Hausdorff-limits* of the sequence M_k, respectively; sometimes the *Kuratovski-Painlevt limits.* Trivially,

$$\liminf M_k \subset \limsup M_k.$$

Similarly, the limits $\limsup_{y\to y^0} S(y)$ and $\liminf_{y\to y^0} S(y)$ are defined for multifunctions S:

$$\begin{aligned}\limsup_{y\to y^0} S(y) &= \{x \mid \liminf_{y\to y^0} \operatorname{dist}(x, S(y)) = 0\},\\ \liminf_{y\to y^0} S(y) &= \{x \mid \limsup_{y\to y^0} \operatorname{dist}(x, S(y)) = 0\}.\end{aligned}$$

Note that lim inf, in the bracket, has the following meaning: First take any sequence $y^k \to y^0$, next consider the usual lower limit $r = \liminf r_k$ for the related sequence of extended reals $r_k = \operatorname{dist}(x, S(y^k)) \in \mathbb{R}^+ \cup \{\infty\}$. Clearly, r depends on the selected y^k.Now, $\liminf_{y\to y^0} \operatorname{dist}(x, S(y))$ denotes the infimum of r over all sequences $y^k \to y^0$. Analogously, one has to read *limsup.*

We continue this chapter by clarifying some relations between pseudo-regularity and other regularity notions.

Metric, Pseudo-, Upper Regularity; Openness with Linear Rate

Let us first mention a connection between pseudo- and upper regularity at isolated pre-images.

Lemma 1.10 (pseudo-regularity at isolated zeros). *If $F : X \rightrightarrows Y$ (metric spaces) is pseudo-regular at $z^0 = (x^0, y^0)$ and if x^0 is isolated in $F^{-1}(y^0)$, then F is upper regular at z^0 with the same rank.* ◇

Proof. Indeed, starting with neighborhoods $U' \ni x^0$ and $V' \ni y^0$ related to pseudo-regularity, one may exploit (setting $y' = y^0$) that

$$\operatorname{dist}(x, F^{-1}(y^0)) \leq Ld(y, y^0) \quad \forall x \in F^{-1}(y) \cap U', \ \forall y \in V'.$$

Decreasing U' (then x moves to x^0), $\operatorname{dist}(x, F^{-1}(y^0))$ will be attained at the isolated point $x^0 \in F^{-1}(y^0)$. Thus (D2) holds true for $S = F^{-1}$ with $V = V'$ and with a new neighborhood U. Decreasing V' we can further arrange, by pseudo-regularity, that $F^{-1}(y) \cap U \neq \emptyset$ for all $y \in V'$. Therefore, upper regularity of F at z^0 holds with rank L. □

Remark 1.11 (pseudo-regularity and Lipschitz continuity). Under the assumptions of Lemma 1.10, one shows analogously the existence of neighborhoods U and V of x^0 and y^0, respectively, such that the multifunction $y \mapsto F^{-1}(y) \cap U$ is Lipschitz on V. So, if x^0 is isolated in $F^{-1}(y^0)$, *pseudo-regularity* of F at (x^0, y^0) and *Lipschitz continuity* (in the Hausdorff-distance) near y^0 of the map $y \mapsto F^{-1}(y) \cap (x^0 + \varepsilon B)$ (for fixed small $\varepsilon > 0$) mean exactly the same. ◇

One says that $F : X \rightrightarrows Y$ is *metrically regular* (with rank $L > 0$) at $z^0 = (x^0, y^0) \in \operatorname{gph} F$ if, for certain neighborhoods U and V of x^0 and y^0, respectively, the estimate

$$\operatorname{dist}(x, F^{-1}(y')) \leq L \operatorname{dist}(y', F(x)) \quad \forall x \in U, \ y' \in V \tag{1.7}$$

holds true. For completeness, we present a proof of the well-known fact that metric regularity and pseudo-regularity describe the same property. Basically, this statement is known from [Iof81, BZ88, Pen89].

Lemma 1.12 (metrically regular = pseudo-regular). *F is pseudo-regular at z^0 if and only if F is metrically regular at z^0.* ◇

Proof. Writing down the both definitions and using condition (1.3) instead of (1.1) one obtains

metric regularity: $\operatorname{dist}(x, F^{-1}(y')) \leq Ld(y', y) \ \forall x \in U, y \in F(x), y' \in V$,
pseudo-regularity: $\operatorname{dist}(x, F^{-1}(y')) \leq Ld(y', y) \ \forall x \in U, y \in F(x), y, y' \in V$.

To see that pseudo-regularity ensures (1.7) (the reverse is trivial), note that in case of $y \notin V$, the distance $d(y', y)$ becomes large if one restricts y' to a new, smaller neighborhood of y^0. We show (1.7) after taking sufficiently small neighborhoods $U_r = x^0 + rB$ and $V_r = y^0 + rB$.

Obviously, $\text{dist}\,(x^0, F^{-1}(y')) \le Ld(y', y^0)$ is valid by pseudo-regularity. Hence, for $x \in U_r$, $y' \in V_r$ and $y \in F(x)$, we have $\text{dist}\,(x, F^{-1}(y')) \le d(x, x^0) + Ld(y', y^0)$. So (1.7) is true if $d(x, x^0) + Ld(y', y^0) \le Ld(y', y)$. But otherwise, the inequality

$$Ld(y', y) < d(x, x^0) + Ld(y', y^0) \le r + Lr$$

tells us that $y \in V$ for small $r > 0$. So (1.7) follows again from pseudo-regularity. □

Openness of F with linear rate around $z^0 \in \text{gph}\, F$ means by definition the existence of $s > 0$, $L > 0$ and of some neighborhood Ω of z^0, such that

$$y + rB \subset F(x + LrB)\ \forall (x, y) \in \Omega \cap \text{gph}\, F,\ \forall r \in (0, s).$$

In other words, F^{-1} is Lipschitz l.s.c. with uniform rank L at all (y, x) such that $(x, y) \in \Omega \cap \text{gph}\, F$, and the related neighborhoods $V(x, y) = y + sB^0$ of y (the l.s.c. Lipschitz estimate holds on which) have again an *uniform radius s*. Evidently, this is pseudo-regularity of F at z^0, too.

Calmness and Upper Regularity at a Set

Our local upper Lipschitz property (D2) for $S : Y \rightrightarrows X$ was used in [Don95] for instance. More generally, Robinson [Rob81] defined (D2) *with respect to a set* $\emptyset \ne X^0 \subset S(y^0)$ by

$$S(y) \cap U \subset X^0 + Ld(y, y^0)B_X\ \forall y \in V. \tag{1.8}$$

Now, the neighborhood U of x^0 in (1.2) is replaced by an

open set U containing a set $X^0 + \lambda B := \{x | \text{dist}\,(x, X^0) \le \lambda\},\ \lambda > 0.$

As before, we call F *upper regular* at (X^0, y^0) if both F^{-1} is locally upper Lipschitz at (y^0, X^0) and $F^{-1}(y) \cap U \ne \emptyset\ \forall y \in V$.

Another variation of (D2), called *calmness* of S at $(y^0, x^0) \in \text{gph}\, S$ means the existence of some L and neighborhoods U, V of x^0 and y^0, respectively, such that

$$S(y) \cap U \subset S(y^0) + Ld(y, y^0)B_X\ \forall y \in V. \tag{1.9}$$

If $x^0 \in X^0$, then the local upper Lipschitz condition (1.8) implies (1.9) immediately. Calmness also means that the pseudo-Lipschitz condition (D1) has to hold for particular $y' = y^0$ only, and has been applied and investigated e.g. in [Cla83] for deriving optimality conditions. Under this respect, calmness can be similarly used as the local upper Lipschitz property at a set, cf. Section 2.1 (optimality conditions) and Theorem 2.10. An interesting calmness condition for multifunctions can be found in [HO01]. It is applicable to the models in [Out00] and many models in [LPR96].

Example 1.13 (pseudo-Lipschitz, but not locally upper Lipschitz). Let $s(y) = 1+\sqrt{|y|}$, and let $S(y)$ be the interval $[-s(y), s(y)]$ for real y. Then, if $\emptyset \neq X^0 \subset S(0)$, the mapping S is not locally upper Lipschitz at $(0, X^0)$ because, for each set $U = X^0 + \varepsilon B$ and each $L > 0$, one finds points $x(y) \in S(y) \cap U$ such that $\operatorname{dist}(x(y), X^0) > L|y|$ and $|y| < \varepsilon$.

Further, S is not calm at $(y^0, x^0) = (0, 1)$. On the other hand, S is pseudo-Lipschitz at each point $(0, x^0)$, $x^0 \in \operatorname{int} S(0)$. ◇

Example 1.14 (the inverse of Dirichlet's function). For the real function

$$h(x) = 0 \text{ if } x \text{ is rational}; \quad h(x) = 1 \text{ otherwise},$$

the inverse $S = h^{-1}$ is calm at $(y^0, x^0) = (0, 0)$ and locally upper Lipschitz at $(0, S(0))$ since $\operatorname{cl} S(0) = \mathbb{R}$. The mapping $S(y) = \{x | h(x) \geq y\}$ is even pseudo Lipschitz at (0,0) since $h(x) = 1 \geq y$ holds for all irrational x and all y near 0. ◇

The second example indicates that the usual construction of penalties for calm equations, $F(x) = f(x) + \alpha \|h(x)\|$, may lead to terrible auxiliary functions F. For related questions we refer to §2.1 and Lemma 2.1.

1.6 First Motivations

Property (D1) is closely related to continuity statements on parametric optimization models

$$\inf f(x, y) \text{ with respect to } x \in S(y). \tag{1.10}$$

Here, $S : Y \rightrightarrows X$ (metric spaces) and $f : X \times Y \to \mathbb{R}$ are given, and y plays the role of a parameter. If x^0 is a local solution for y^0, feasible points $x' \in S(y')$ can be assigned to $x \in S(y)$ in an uniform Lipschitzian manner provided that x as well as y and y' are close to x^0 and y^0, respectively. Under (Lipschitz) continuity of f, this allows estimates of the related infima

$$\varphi(y) = \inf_{x \in S(y)} f(x, y),$$

and of solution sets

$$\Psi(y) = \{x \in S(y) | f(x, y) = \varphi(y)\}.$$

The properties (D1) and (D2) also ensure the validity of well-known necessary optimality conditions and help to estimate related Lagrange multipliers in terms of the parameter distance $d(y, y^0)$ even if primal-dual solutions are not unique. These facts, which become clearer below, explain the great interest in (D1) and (D2) as well as in the other types of regularity and semicontinuity for multifunctions. As examples and basic results, we mention the following classical statements.

Parametric Global Minimizers

Theorem 1.15 (Berge–Hogan stability). *Let f be continuous and S be u.s.c. at y^0 and l.s.c. at some $(y^0, x^0), x^0 \in \Psi(y^0)$. Then one has:*

(i) (C. Berge [Ber63]) *If $S(y^0)$ is compact then, at y^0, φ is continuous and Ψ is u.s.c.*

(ii) (W. Hogan [Hog73]) *Let $X = \mathbb{R}^n$, $f(\cdot, y)$ be convex, $\Psi(y^0)$ be compact and $S(y^0)$ be closed and convex. Then, at y^0, φ is continuous and Ψ is u.s.c. If, in addition, all sets $S(y)$ are closed, then $\Psi(y) \neq \emptyset$ for y near y^0.* ◇

Proof. (i) Let $x^0 \in \Psi(y^0)$ and let $y \to y^0$ denote any sequence that realizes $\limsup \varphi(y)$. Since S is l.s.c., one finds $x \in S(y)$, $x \to x^0$. Hence,φ is upper semicontinuous due to

$$\varphi(y^0) = f(x^0, y^0) = \lim f(x, y) \geq \limsup \varphi(y).$$

On the other hand, to any $x' \in S(y')$, $y' \to y^0$, there corresponds some $x'' \in S(y^0)$ with $d(x'', x') \to 0$. By compactness of $S(y^0)$, there exists some common accumulation point $x_0 \in S(y^0)$ of all x' and x''. Hence, given $y \to y^0$, one may first select a subsequence y' such that $\liminf \varphi(y) = \lim \varphi(y')$ and next choose certain $x' \in S(y')$ such that $\lim \varphi(y') = \lim f(x', y')$. So we obtain continuity of φ:

$$\liminf \varphi(y) = \lim f(x', y') = f(x_0, y^0) \geq \varphi(y^0).$$

Finally, considering the (existing) accumulation points x_0 of any $x \in \Psi(y)$ as $y \to y^0$, one finds first $x_0 \in S(y^0)$ and next $f(x_0, y^0) = \lim \varphi(y) = \varphi(y^0)$. Thus $x_0 \in \Psi(y^0)$ yields that Ψ is u.s.c. at y^0.

(ii) Again, φ is u.s.c. due to the arguments from (i). Therefore, the rest will follow as above by continuity and compactness, provided that $\|x_y\|$ is bounded for every sequence of $x_y \in S(y)$ satisfying $y \to y^0$ and $\limsup f(x_y, y) \leq \varphi(y^0)$. To show the latter, choose r large enough such that $\Psi(y^0) \cap (x^0 + r\,\mathrm{bd}\, B) = \emptyset$. Next assume that certain x_y diverge. Then there are points z_y on the line segment $[x^0, x_y]$ with $\|z_y - x^0\| = r$. Since $\mathbf{dist}\,(x_y, S(y^0)) \to 0$ holds by upper semicontinuity, every accumulation point z^0 of the bounded elements z_y belongs to the closed and convex set $S(y^0)$. Because of $X = \mathbb{R}^n$ such a point z^0 exists. Since $f(\cdot, y)$ is (quasi-) convex, it holds additionally that

$$f(z^0, y^0) \leq \limsup_y f(z_y, y) \leq \limsup_y \max\{f(x^0, y), f(x_y, y)\} \leq \varphi(y^0).$$

Thus, we obtain $z^0 \in \Psi(y^0) \cap (x^0 + r\,\mathrm{bd}\, B)$, in contradiction to the choice of r. □

The statements of the foregoing theorem have been generalized under several points of view: with respect to continuity and compactness, and by investigating also **ε-solutions** x, i.e., $x \in S(y)$ and $f(x, y) \leq \varphi(y) + \varepsilon$, where $\varepsilon \downarrow 0$ and $y \to y^0$, see, e.g., [BGK+82, RW98]. Even our formulation (ii) is a slight generalization of Hogan's original result. Nevertheless, the basic arguments of the original proofs remained valid.

It should be also mentioned that the hypotheses concerning S were investigated for several important mappings in finite dimension, e.g., for the mappings

$$S_1(y) = \{x \in \mathbb{R}^n \mid g_i(x) \le y_i,\ i = 1, \ldots, m\},\ g_i \text{ analytic and convex on } \mathbb{R}^n;$$

$$S_2(y) = S_1(y),\quad g_i \text{ (quasi-) convex polynoms, rational coefficients;}$$

$$S_3(y) = \{x \in S_2(y) \mid x_k \text{ integer for } k = 1, \ldots, \kappa\},\quad 1 \le \kappa < n.$$

These investigations are (with appropriate objectives f) closely related to duality and existence theorems for problems of type (1.10), cf., e.g., [Roc71, Lau72, Roc74, Kum81, BGK+82, BM88, BA93, Kla97, Sha98].

We further note that stability results of the type presented in Theorem 1.15 may be also formulated in terms of (classical) convergence of functions and sets, see, for example, [DFS67, Fia74, Kum77].

Parametric Local Minimizers

In the case $X = \mathbb{R}^n$, the Berge–Hogan theorem may be extended to certain sets of parametric local minimizers of the problem (1.10). Following Robinson [Rob87] (see also [FM68, Kla85]), a nonempty set $Z \subset \mathbb{R}^n$ is said to be a *complete* (or *strict*) *local minimizing set* (*CLM set*) for $f(\cdot, y^0)$ on $S(y^0)$ if there is an open set $Q \supset Z$ such that

$$Z = \Psi_{\text{cl}\,Q}(y^0) := \operatorname{argmin}_x \{f(x, y^0) \mid x \in S(y^0) \cap \text{cl}\,Q\},$$

where cl Q is the closure of Q, and "argmin" is written for the set of global minimizers. Note that $Q \supset Z$ is supposed being open, hence each element of a CLM set is a local minimizer for $f(\cdot, y^0)$ on $S(y^0)$. In particular, $\{x^0\}$ is a CLM set if x^0 is a strict local minimizer for $f(\cdot, y^0)$ on $S(y^0)$, and $\Psi(y^0)$ is a CLM set provided it is not empty. Moreover, certain sets of local minimizers satisfying a linear or quadratic growth condition (sometimes called sets of *weak sharp minimizers*) are CLM sets, see, e.g., [War94, Kla94a, BS00].

Theorem 1.16 (stability of CLM sets [Kla85, Rob87]). *Consider (1.10) in the case $X = \mathbb{R}^n$. Given $y^0 \in Y$, let Z be a compact CLM set for $f(\cdot, y^0)$ on $S(y^0)$, and let $S(y)$ be closed for each y in some neighborhood of y^0. Further, suppose that f is continuous on $X \times Y$ and that S is both u.s.c. at y^0 and l.s.c. at some (y^0, x^0), $x^0 \in Z$. Then there are a neighborhood O of y^0 and an open bounded set $Q \supset Z$ such that*

(i) $\Psi_{\text{cl}\,Q}(y) \ne \emptyset$ *($\forall y \in O$) and $\Psi_{\text{cl}\,Q}$ is u.s.c. at y^0 with $\Psi_{\text{cl}\,Q}(y^0) = Z$,*

(ii) for each $y \in O$, $\Psi_{\text{cl}\,Q}(y)$ is a CLM set for $f(\cdot, y)$ on $S(y)$, i.e., in particular, any element of $\Psi_{\text{cl}\,Q}(y)$ is a local minimizer for $f(\cdot, y)$ on $S(y)$.

◇

Proof. By definition, there is some open bounded set $Q \supset Z$ such that $Z = \Psi_{\mathrm{cl}\,Q}(y^0)$. Hence, since S is l.s.c. at some (y^0, x^0), $x^0 \in Z$, the sets $S(y) \cap \mathrm{cl}\, Q$ are nonempty and compact for y near y^0. Since f is continuous, assertion (i) follows from Weierstrass' theorem and part (i) of Theorem 1.15. Moreover, when applying that $\Psi_{\mathrm{cl}\,Q}(y^0)$ is compact, assertion (i) gives the Berge u.s.c. of $\Psi_{\mathrm{cl}\,Q}$ at y^0. Hence, for the open set Q containing $\Psi_{\mathrm{cl}\,Q}(y^0)$, there is some neighborhood $\mathcal{O}$ of y^0 such that $\Psi_{\mathrm{cl}\,Q}(y) \subset Q$ for all y in $\mathcal{O}$, i.e., by definition, these sets are CLM sets. □

Epi-Convergence

The parametric optimization problem (1.10)

$$\inf f(x,y) \text{ with respect to } x \in S(y)$$

can be reformulated by introducing an (improper) function g as

$$g(x,y) = f(x,y) \text{ if } x \in S(y); \quad g(x,y) = \infty \text{ otherwise}$$

and studying the "free" parametric extremal problem

$$\inf g(x,y) \text{ with respect to } x \in X. \tag{1.11}$$

Conversely, having an improper function $g = g(x,y)$ then, after setting $S(y) := \{x | g(x,y) < \infty\}$ (the domain of g), we are just studying problem (1.10) with an objective $f(\cdot, y) = g(\cdot, y)$ defined on $S(y)$.

Similarly, to obtain an objective that is everywhere finite, one can put

$$\Gamma(y) = \mathrm{epi}\, g(\cdot, y) := \{(t,x) \mid t \geq g(x,y)\},$$

whereupon

$$\phi(y) = \inf\{t | (t,x) \in \Gamma(y)\}, \psi(y) = \{x | (\phi(y), x) \in \Gamma(y)\}. \tag{1.12}$$

Of course, the different formulations (1.10), (1.11), (1.12) of the same subject alone cannot present new insides for the analysis of parametric optimization problems. However, since the suppositions for (1.11) are usually written by means of the epigraphs $\mathrm{epi}\, g(\cdot, y)$ and their convergence properties (types of epi-convergence) as $y \to y^0$, related conditions have often another (shorter) form. On the other hand, they must be re-interpreted in terms of $\mathrm{dom}\, g(\cdot, y)$.

Here, we will prefer the classical parametric formulation (1.10), whereas e.g. in [RW84, Rob87, Att84, RW98] just (1.11) has been favored. Note that, in the context of approximations to optimization problems, the close relations of the arguments in the epi-convergence approach to those of the classical theory of functions were discussed by Kall [Kal86].

Chapter 2

REGULARITY AND CONSEQUENCES

In this chapter, we present conditions for certain Lipschitz properties of multivalued maps and the related types of regularity, we investigate interrelations between them and discuss classical applications as, e.g., (necessary) optimality conditions and stability in optimization. A great part of this chapter is devoted to pseudo-regularity of multifunctions in Banach spaces, where we do not utilize generalized derivatives. We directly use Ekeland's variational principle as well as the family of assigned inverse functions. They lead to characterizations of pseudo- regularity for the intersection of multifunctions and permit rather weak assumptions concerning the image- and pre-image space as well.

2.1 Upper Regularity at Points and Sets

Characterization by Increasing Functions

Let X, Y be metric spaces, and let $y^0 \in Y$, $S : Y \rightrightarrows X$, $\emptyset \neq X^0 \subset S(y^0)$ and $p : X \to \mathbb{R}$. We call p *Lipschitzian increasing* near X^0 if $p \equiv 0$ on X^0 and there are $c > 0$, $\delta > 0$ such that

$$p(x) \geq c \operatorname{dist}(x, X^0) \text{ whenever } \operatorname{dist}(x, X^0) < \delta. \tag{2.1}$$

Further, we say that p *describes S near* (y^0, X^0) (or p is *a describing function for S near* (y^0, X^0)), briefly

$$p \doteq S(y^0, X^0),$$

if

$$\begin{aligned} &S \text{ is locally upper Lipschitz at } (y^0, X^0) \\ &\Leftrightarrow p \text{ is Lipschitzian increasing near } X^0. \end{aligned}$$

We will see by Theorem 2.6 that describing functions can play the role of penalty functions in optimality conditions. So the structure of possible "candidates" becomes interesting. By the next statement, there is always a describing *Lipschitz* function, globally defined with rank 1 and not depending on X^0. Let us agree that the metric in product spaces $Y \times X$ is defined as

$$d((y,x),(y',x')) = \max\{d_Y(y,y'), d_X(x,x')\}.$$

Lemma 2.1 (upper Lipschitz and describing Lipschitz functionals). *Given* $S : Y \rightrightarrows X$ *and* $\emptyset \neq X^0 \subset S(y^0)$, *the distance function*

$$p_S(x) = \text{dist}\,((y^0,x), \text{gph}\, S) \tag{2.2}$$

satisfies $p_S \doteq S(y^0,X^0)$, *i.e., S is locally upper Lipschitz at* (y^0,X^0) *if and only if* p_S *is Lipschitzian increasing near* X^0. ◇

Proof. For simplicity, we write $d(\cdot,\cdot)$ both for $d_X(\cdot,\cdot)$ and $d_Y(\cdot,\cdot)$. Evidently, p_S vanishes on X^0. Let p_S be Lipschitzian increasing near X^0. Then S is locally upper Lipschitz with rank $L = c^{-1}$, since for $\text{dist}\,(x,X^0) < \delta$,

$$x \in S(y) \Rightarrow c\, d(x,X^0) \leq p_S(x) \leq d((x,y^0),(x,y)) = d(y,y^0).$$

Conversely, let p_S be *not* Lipschitzian increasing near X^0. This is, for each $\varepsilon > 0$, there is some x such that $\text{dist}\,(x,X^0) < \varepsilon$ and

$$\text{dist}\,((y^0,x),\text{gph}\, S) < t := \varepsilon\, \text{dist}\,(x,X^0).$$

Select any $(y_t,x_t) \in \text{gph}\, S$ with

$$d((y^0,x),(y_t,x_t)) = \max\{d(x,x_t), d(y^0,y_t)\} < t.$$

Then,

$$\begin{aligned} &\varepsilon + t > \text{dist}\,(x_t,X^0) \geq \text{dist}\,(x,X^0) - t = (1-\varepsilon)\,\text{dist}\,(x,X^0) \\ &\text{and } d(y^0,y_t) < \varepsilon\, \text{dist}\,(x,X^0). \end{aligned}$$

Thus, both $\text{dist}\,(x_t,X^0)$ and $d(y^0,y_t)/\text{dist}\,(x_t,X^0)$ vanish (as $\varepsilon \downarrow 0$); so S is not locally upper Lipschitz at (y^0,X^0). □

More examples
of describing functions for $S : Y \rightrightarrows X$ near X^0 are the following ones.

(i) Equations
In the case of functions $h : X \to Y$ and $S = h^{-1}$, one easily sees that $p(x) = d(h(x),y^0)$ fulfills $p \doteq S(y^0,X^0)$ and $p \geq p_S$, too.

In the form considered here, i.e., as an equation $h(x) = 0$ and with $S = h^{-1}$, one can study all maps Σ that were originally given by a function $G : (X,T) \to \mathbb{R}$ and a multifunction $H : X \rightrightarrows T$ via

$$\Sigma(y) = \{x \mid G(x,t) \leq y \ \forall t \in H(x)\} \qquad (y \in \mathbb{R});$$

this form occurs in models of *generalized semi-infinite optimization* where $X = \mathbb{R}^n$. One has only to put

$$h(x) = \max\{y^0, \sup_{t \in H(x)} G(x,t)\}, \tag{2.3}$$

whereupon

$$\begin{array}{ll} \Sigma(y^0) \subset \Sigma(y) = h^{-1}(y) & \text{for } y \geq y^0 \text{ and } h^{-1}(y) \neq \emptyset \\ \Sigma(y) \subset \Sigma(y^0) \text{ and } h^{-1}(y) = \emptyset & \text{for } y < y^0. \end{array}$$

Therefore, the map $S = \Sigma$ fulfills (1.8) iff so does h^{-1}, too.

(ii) Cone constraints

Let Y be a linear normed space, X be a metric space, $g : X \to Y$, $K \subset Y$ be a convex cone, $\eta \in \operatorname{int} K \setminus \{0\}$ and $S(y) = \{x \in X \mid g(x) \in y + K\}$.

Lemma 2.2 (Cone constraints). *Let $\emptyset \neq X^0 \subset S(0)$ and $X_r = X^0 + rB$. Then, if g is Lipschitz on X_β for some $\beta > 0$, the function*

$$p(x) = \inf\{\lambda > 0 \mid g(x) + \lambda\eta \in K\} \tag{2.4}$$

fulfills

$$c_1\, p_S(x) \;\leq\; p(x) \leq c_2\, p_S(x) \quad \forall x \in X_r \tag{2.5}$$

with certain constants $0 < c_1 \leq c_2$ and $r = \beta/3$. Hence $p \doteq S(0, X^0)$. ◇

Proof. Let L_g be some Lipschitz rank of g on X_β and $\eta + \alpha B_Y \subset K$, $\alpha > 0$. Then, one obtains for all $\lambda > 0$ and $x \in X_\beta$,

$$\lambda\eta + g(x) \in K \;\text{ if }\; \|g(x)\| \leq \lambda\alpha.$$

Hence $p(x) \leq \alpha^{-1}\|g(x)\|$ and

$$\begin{array}{rcl} p_S(x) & = & \operatorname{dist}((0_Y, x), \operatorname{gph} S) \\ & \leq & \inf_{\lambda > p(x)} \operatorname{dist}((0_Y, x), (\lambda\eta, x)) \\ & = & p(x)\|\eta\|. \end{array} \tag{2.6}$$

Next, fix any $r \in (0, \frac{1}{2}\beta)$. We verify

$$p(x) \leq (1 + L_g)\alpha^{-1} p_S(x) \quad \forall x \in X_r. \tag{2.7}$$

Due to $(0_Y, X^0) \subset \operatorname{gph} S$, it holds $p_S(x) \leq \operatorname{dist}(x, X^0) \leq r < \frac{1}{2}\beta$. So one finds some ε satisfying $p_S(x) < \varepsilon < \frac{1}{2}\beta$ as well as some (y', x') such that

$$g(x') \in y' + K, \quad \|y'\| < \varepsilon \quad \text{and} \quad d(x', x) < \varepsilon.$$

From $g(x') - y' \in K$ we conclude (by adding points of a convex cone) that

$$g(x') - y' + \lambda\eta + \lambda\alpha B_Y \subset K \quad \forall \lambda > 0.$$

Therefore, the inclusion
$$g(x) + \lambda\eta \in K \tag{2.8}$$
holds true, whenever
$$g(x) \in g(x') - y' + \lambda\alpha B_Y.$$
Because of $\|y'\| < \varepsilon$, the latter can be guaranteed by $\|g(x) - g(x')\| + \varepsilon \leq \lambda\alpha$. Now (by the choice of r), x and x' belong to X_β and fulfill $\|g(x) - g(x')\| \leq L_g\varepsilon$, whereupon (2.8) is ensured whenever $(L_g + 1)\varepsilon \leq \lambda\alpha$. Considering $\inf \lambda$, this yields
$$p(x) \leq \alpha^{-1}(1 + L_g)p_S(x) \quad \text{via} \quad \varepsilon \to p_S(x).$$
The assertion now follows from (2.6), (2.7) and Lemma 2.1. □

The convex cone K had not to be closed, indeed. If also X is a linear normed space and g is linear and continuous, then one easily shows that p is convex. In addition, p is bounded on some neighborhood of $x \in \operatorname{int} X_r$ due to (2.5). So it is locally Lipschitz on $\operatorname{int} X_r$, too. Needless to say that p is simpler than p_Σ from the viewpoint of computation. For $Y = \mathbb{R}^m$, $K = \{y \in Y | y_i \leq 0 \ \forall i\}$ and $\eta = -(1, ..., 1)$ one obtains the usual penalty term $p(x) = \max_i\{0, g_i(x)\}$.

(iii) Cone constraints and equations

Let
$$\Sigma(y, z) = S(y) \cap T(z),$$
where S satisfies the assumptions of Lemma 2.2, $h : X \to Z$ sends X into a linear normed space Z, $T(z) = h^{-1}(z)$ and $\emptyset \neq X^0 \subset \Sigma(0_{Y \times Z})$.
Writing Σ in form of cone constraints with the cone $K' = (K, \{0_Z\})$ in the product space, the interior of K' is empty and Lemma 2.2 cannot be applied. In addition, the describing distance-function p_Σ according to Lemma 2.1 only satisfies

$$\begin{aligned}
&p_\Sigma(x)\\
&= \inf_{(y',z',x') \in \operatorname{gph} \Sigma} \max\{d((0_Y, x), (y', x')), d((0_Z, x), (z', x'))\}\\
&\geq \max\{\inf_{(y',x') \in \operatorname{gph} S} d((0_Y, x), (y', x')), \ \inf_{(z'',x'') \in \operatorname{gph} T} d((0_Z, x), (z'', x''))\}\\
&= \max\{\ p_S(x)\ ,\ \ p_T(x)\ \}.
\end{aligned}$$

So we only know, by the previous statements, that the maximum function
$$q(x) = \max\{\ p_S(x)\ ,\ p_T(x)\ \},\ T = h^{-1} \tag{2.9}$$
fulfills
$$p_\Sigma(x) \ \geq \ q(x)\ , \tag{2.10}$$
and q is Lipschitzian increasing near X^0 iff so is
$$Q(x) = \max\{\ p(x)\ ,\ \|h(x)\|\ \},\quad p \text{ from Lemma 2.2.}$$

However, due to the gap between $p_\Sigma(x)$ and $q(x)$, the function p_Σ may Lipschitzian increase near X^0 while q does not (Then S and h^{-1} violate (L1), but not so Σ). In this situation, q and Q are no describing functions for Σ.

On the other hand, the maximum q turns out to be a describing function under all classical regularity assumptions that ensure, as in the subsequent Lemma, that Σ is pseudo-Lipschitzian (or only calm) at $(0_Y, 0_Z, x^0)$.

Lemma 2.3 (the max-form under calmness). *Suppose: X, Y,Z are Banach spaces,* $g, h \in C^1$, $Dh(x^0)X = Z$, *some* u *satisfies* $Dh(x^0)u = 0$ *and* $g(x^0) + Dg(x^0)u \in int\, K$, *and* $x^0 \in X^0 \subset \Sigma(0_Y, 0_Z)$. *Moreover, let* X^0 *be contained in a sufficiently small (by diameter) neighborhood* Ω *of* x^0.

Then, q *in (2.9) is a describing function for* Σ *near* $(0_Y, 0_Z, X^0)$. ◇

Proof. Our suppositions are nothing but well-known regularity conditions for optimization problems in Banach spaces, cf. [Rob76a, Rob76c, ZK79], which ensure that the map Σ is pseudo-Lipschitz at $(0_Y, 0_Z, x^0)$; see also the discussion after Theorem 2.22. So, the lemma will follow from Theorem 2.4 below because $T^{-1} = h$ is locally Lipschitz. □

(iv) Arbitrary Intersections

More general, let *X, Y, Z* be metric spaces, $S : Y \rightrightarrows X$, $T : Z \rightrightarrows X$ and $\Sigma(y, z) = S(y) \cap T(z)$.

Theorem 2.4 (the max-form for intersections). *Let* $x^0 \in X^0 \subset \Sigma(y^0, z^0)$, Σ *be calm at* (y^0, z^0, x^0) *and* T^{-1} *be pseudo-Lipschitz at* (x^0, z^0). *Moreover, suppose that* X^0 *is contained in a sufficiently small (by diameter) neighborhood* Ω *of* x^0. *Then, the function*

$$q(x) = \max\{\, p_S(x)\ ,\ p_T(x)\ \} \tag{2.11}$$

fulfills $q \doteq S(y^0, z^0, X^0)$. ◇

Proof. The inequality (2.10) follows as above without any assumptions, we estimate q in opposite direction. First notice that q is Lipschitz, so it holds $q(x) \downarrow 0$ as $x \to x^0$. In consequence, for sufficiently small neighborhoods Ω, we find arbitrarily small $\delta > q(x)$. Now, for x near $X^0 \subset \Omega$ and (small) $\delta > q(x)$, there are (by definition of p_S and p_T) points $(y', x') \in \operatorname{gph} S$ and $(z'', x'') \in \operatorname{gph} T$ such that

$$\max\{d(x', x), d(y', y^0)\} < \delta \text{ and } \max\{d(x'', x), d(z'', z^0)\} < \delta. \tag{2.12}$$

Next we apply that T^{-1} is pseudo-Lipschitz at (x^0, z^0), say with rank K. Since $z'' \in T^{-1}(x'')$, there exists, for small δ and Ω, some $z' \in T^{-1}(x')$ satisfying

$$d(z', z'') \le\ K d(x'', x') \le\ 2K\delta. \tag{2.13}$$

We thus obtain $(y', z', x') \in \operatorname{gph} \Sigma$ and

$$d((y', z', x'), (y^0, z^0, x^0)) \le \max\{\delta,\ \delta + 2K\delta,\ \delta + d(x, x^0)\}.$$

So, since (y', z', x') is close to (y^0, z^0, x^0), we may use calmness of Σ, say with rank L at (y^0, z^0, x^0). By (2.12) and (2.13) this ensures the existence of some $\xi \in \Sigma(y^0, z^0)$ such that

$$d(\xi, x') \leq L \max\{d(y', y^0), d(z', z^0)\} \leq L(1 + 2K)\delta.$$

Finally, $p_\Sigma(x) \leq d(\xi, x)$ implies the upper estimate

$$p_\Sigma(x) \leq d(\xi, x) \leq d(\xi, x') + d(x', x) \leq L(1 + 2K)\delta \ + \ \delta$$

and yields (as $\delta \downarrow q(x)$) $p_\Sigma(x) \ \leq \ (L(1 + 2K) + 1) \ q(x)$. Recalling (2.10) and Lemma 2.1, the latter tells us that q is a describing function for Σ near (y^0, z^0, X^0) because so is p_Σ. □

Notice that Lemma 2.3 and Theorem 2.4 do not assert the upper Lipschitz property of Σ at (y^0, z^0, X^0), itself. The relation between the upper and pseudo-Lipschitz properties as well as calmness will be investigated under Theorem 2.10. Next, we inspect the hypothesis of Σ being calm in the previous theorem and reduce calmness of the intersection of two mappings to the intersection of one mapping with a constant set (a new space *X)* only.

Theorem 2.5 (calm intersections). *Let S be calm at* (y^0, x^0)*, T be calm at* (z^0, x^0) *and* T^{-1} *be pseudo-Lipschitz at* (x^0, z^0). *Moreover, let* $H(z) = S(y^0) \cap T(z)$ *be calm at* (z^0, x^0). *Then* $\Sigma(y, z) = S(y) \cap T(z)$ *is calm at* (y^0, z^0, x^0). ◇

Proof. Let $(y, z, x) \in \operatorname{gph} \Sigma$ be close to (y^0, z^0, x^0). Since S and T are calm (say with rank L), there are $x' \in S(y^0)$ and $x'' \in T(z^0)$ such that

$$\max\{d(x, x'), d(x, x'')\} \leq L \max\{d(y, y^0), d(z, z^0)\}.$$

Since T^{-1} is pseudo-Lipschitz (rank K), $z^0 \in T^{-1}(x'')$ and x', x'' are close to x^0, we find z' such that

$$z' \in T^{-1}(x') \text{ and } d(z', z^0) \leq K d(x', x'').$$

Next observe that $x' \in H(z')$. Therefore, there exists also some $\xi \in H(z^0)$ satisfying

$$d(\xi, x') \leq L_H \ d(z', z^0).$$

Using these inequalities, we directly obtain the required Lipschitz estimate

$$\begin{aligned} &d(x, \xi) \leq d(x, x') + d(x', \xi) \\ &\leq L \max\{d(y, y^0), d(z, z^0)\} + L_H \, d(z', z^0) \\ &\leq L \max\{d(y, y^0), d(z, z^0)\} + L_H \, K \, d(x', x'') \\ &\leq L \max\{d(y, y^0), d(z, z^0)\} + 2 L_H \, K \, L \max\{d(y, y^0), d(z, z^0)\}. \end{aligned}$$

□

(v) Set-Constraints
Assume that $\Sigma(y) = S(y) \cap M$ and M is a fixed, closed subset of X. Clearly, then one may study S on the new metric space $X := M$ which leads us to a new function p_S. However, let us also regard two usual descriptions of $x \in M$ via functions under the viewpoint of the pseudo-Lipschitz assumption for T^{-1} in the theorem.

(a) Setting $h(x) = \text{dist}\,(x, M), Z = \mathbb{R}^+, z^0 = 0, T = h^{-1}$,the mapping $T^{-1} = h$ is pseudo-Lipschitz, and $p_T(x) = \inf\,\{\,\max\{z',\ d(x', x)\,\}\mid \text{dist}\,(x', M) = z'\}$. If S is already calm (w.r. to the space X) then the theorem allows us to study, instead of $S(y) \cap T(z)$, the calmness of the mapping

$$H(z) = S(y^0) \cap T(z) = S(y^0) \cap h^{-1}(z).$$

If H is calm at $(0, x^0)$, then so is $S \cap T$ at $(y^0, 0, x^0)$, hence also the original map $\Sigma(y) = S(y) \cap M$ at (y^0, x^0). This way, one may replace (for the calmness investigation) the fixed set M by $S(y^0)$ and the mapping S by h^{-1}.

(b) Setting $h(x) = 0$ if $x \in M$, $h(x) = 1$ otherwise, and Z, z^0, T as above, the function $T^{-1} = h$ is discontinuos and it holds $p_T(x) = \text{dist}\,(x, M)$ for $\text{dist}\,(x, M) < 1$. The theorem cannot be applied. Indeed, for small $z > 0$, we would obtain the trivial constant map

$$H(z) = S(y^0) \cap T(z) = S(y^0) \cap M$$

which tells us nothing about Σ.

Optimality Conditions

The local upper Lipschitz property of feasible set maps S ensures optimality conditions for constrained minimization in terms of free (i.e., unconstrained) local minimizers of an auxiliary function. To study an optimization problem $\min\{f(x) \mid x \in X^0\}$, consider any map $S : Y \rightrightarrows X$ (between metric spaces) as a parametric family of constraints satisfying $\emptyset \neq X^0 \subset S(y^0)$ for some $y^0 \in Y$. The following statement, though more general, applies basically the same simple arguments as the related proposition in [Cla83] for calm constraints.

Theorem 2.6 (free local minima and upper Lipschitz constraints). *Given metric spaces X, Y, let $S : Y \rightrightarrows X$ be locally upper Lipschitz at (y^0, X^0) with rank L, $f : X \to \mathbb{R}$ be Lipschitz near x^0 with rank K, and let $p \doteq S(y^0, X^0)$. Further, let x^0 be a local minimizer of f on X^0. Then x^0 is a local minimizer of*

$$P(x) = f(x) + \alpha p(x),$$

whenever $\alpha > Kc^{-1}$ with c from (2.1). ◇

Proof. Let $\mu > 0$, let U be the open set in (1.8) and K be some Lipschitz rank for f near x^0. Given $x \in U$, select some $\pi_x \in X^0$ with

$$d(x, \pi_x) \leq \text{dist}\,(x, X^0) + \mu.$$

Then, $d(x,\pi_x) \leq d(x,x^0)+\mu$. For $d(x,x^0) < \delta$ and small δ and μ, we know that $d(\pi_x,x^0) \leq d(\pi_x,x) + d(x,x^0)$ is small enough to apply the Lipschitz estimate $f(x) \geq f(\pi_x) - Kd(x,\pi_x)$ and $f(\pi_x) \geq f(x^0)$. Further, since $p \doteq S(y^0, X^0)$, we have $-p(x) \leq -c\,\text{dist}\,(x, X^0)$. So, it holds

$$\begin{aligned} f(x) &\geq f(\pi_x) - Kd(x,\pi_x) \\ &\geq f(x^0) - Kd(x,\pi_x) \\ &\geq f(x^0) - K\,[\,\text{dist}\,(x,X^0) + \mu\,] \\ &\geq f(x^0) - K\,c^{-1}p(x) - K\mu. \end{aligned}$$

After passing to the limit $\mu \downarrow 0$, the latter ensures the assertion due to

$$P(x) \geq f(x) + \alpha p(x) \geq P(x^0) = f(x^0) \quad \text{if } \alpha > Kc^{-1}.$$

□

It is trivial but useful to note that the function $p \doteq S(y^0, X^0)$ may be replaced, in Theorem 2.6, by any function p^+ satisfying $p^+ \geq p$ and $p^+(x^0) = 0$. Applying the function $p = p_S$ of Lemma 2.1, the new objective P turns out to be even Lipschitz near x^0.

Provided that X and Y are normed spaces, now all necessary optimality conditions for free local minimizers x^0 of P induce necessary conditions for the originally constrained problem. In particular, if directional derivatives $P'(x^0;u)$ of P at x^0 in direction u exist, then it must hold

$$P'(x^0;u) \geq 0 \qquad \forall u \in X; \tag{2.14}$$

and

$$\inf\{v \mid v \in CP(x^0)(u)\} \geq 0 \ \forall u \in X$$

follows for the contingent derivative CP.

Dual Conditions

Let us mention only two basic approaches for obtaining dual conditions; various other approaches and more involved results can be found in [Roc70, Gol72, Roc74, IT74, LMO74, War75, Iof79a, Ben80, KM80, Roc81, BBZ81, BZ82, Pen82, Cla83, Stu86, Mor88, Cha89, Sha98, BS00].

Dual conditions via directional derivatives

If f and p are directionally differentiable (below, we see that the directional derivatives may be generalized) and satisfy

$$(f + \alpha p)'(x^0; u) \leq f'(x^0; u) + \alpha p'(x^0; u), \tag{2.15}$$

then (2.14) yields a condition for the sum

$$\inf_u \; (f'(x^0; u) + \alpha p'(x^0; u)) \geq 0. \tag{2.16}$$

Let, in addition, the directional derivatives be *continuous and sublinear* in u (which is evident for locally Lipschitz convex functions).

Then, applying the Hahn-Banach theorem, see e.g. [KA64], to the sublinear function

$$Q(u, v) := f'(x^0; u) + \alpha p'(x^0, v) \text{ in the product space } \Pi = X \times X,$$

the supporting functional $L_0(u, v) = 0$ of Q on the subspace Π_0 defined by $\Pi_0 = \{(u, v) | \; u = v\}$ can be extended to an additive and homogeneous functional $L(u, v) = L_1(u) + L_2(v)$ on Π that supports Q everywhere. Thus,

$$L_1(u) + L_2(u) = 0 \quad \text{and} \quad Q(u, v) \geq \; L_1(u) + L_2(v)$$

hold for all $u, v \in X$. The latter implies (since Q is continuous by assumption) that L_1, L_2 are bounded, and

$$\inf_u \; (f'(x^0; u) - L_1(u)) \; + \; \inf_v \; (\alpha p'(x^0, v) - L_2(v)) \geq 0.$$

So one obtains the existence of some $x^* = L_1 \in X^*$ satisfying the (conjugate duality) inequality

$$\inf_u \; (f'(x^0; u) - x^*(u)) \; + \; \inf_u \; (x^*(u) + \; \alpha p'(x^0; u)) \geq 0.$$

Since the involved directional derivatives are positively homogenous, the infima are zero and x^* belongs (by definition) just to the usual, convex subdifferential $\partial f'(x^0; .)(0)$. Similarly, one obtains $-x^* \in \partial(\alpha p')(x^0; .)(0)$.

In other words, after defining a new subdifferential ∂_n for the non-convex function f at x^0 as

$$\partial_n f(x^0) = \partial f'(x^0; .)(0), \tag{2.17}$$

(and applying it to αp, too) some $x^* \in X^*$ satisfies the inclusions

$$x^* \in \partial_n f(x^0) \quad \text{and} \quad - x^* \in \partial_n(\alpha p)(x^0) = \alpha \partial_n p(x^0),$$

which is the generalized Lagange condition

$$0 \in \; \partial_n f(x^0) \; + \; \alpha \partial_n p(x^0) \tag{2.18}$$

or simply

$$Df(x^0) + \alpha Dp(x^0) = 0$$

for Fréchet differentiable functions.

Recalling the concrete form of p for h in (2.3) or q in Lemma 2.3, one sees that directional derivatives and contingent derivatives of maximum functions play a crucial role, in this context.

Further, one observes that several concepts of directional derivatives f' may be applied to derive (2.18) for the subdifferential (2.17) in the above way, provided that

(i) condition (2.16) remains valid for local minimizers x^0 of P, and
(ii) the existence of directional derivatives as well as sublinearity and continuity with respect to the directions u can be guaranteed.

For locally Lipschitz functions $f : X \to \mathbb{R}$ on linear normed spaces X, these hypotheses are satisfied by Clarke's directional derivatives f^c and his subdifferential $\partial_c f(x^0)$ which coincides with $\partial_n f(x^0)$ after identifying f' and f^c, cf. Sections 1.2 and 1.3. For $X = \mathbb{R}^n$, the equation $\partial_c f(x^0) = \partial f(x^0)$ in terms of generalized Jacobians $\partial f(x^0)$, cf. [Cla76], increases the analytical tools for computing the derivatives in question.

Dual conditions via generalized subdifferentials

Without applying directional derivatives, one may restrict the functions f, p to the set $X_\varepsilon = x^0 + \varepsilon B$, $\varepsilon > 0$ sufficiently small, whereafter

$$0 \in \partial_g (f + \alpha p)(x^0)$$

holds true for the minimizer x^0 and all types of generalized subdifferentials ∂_g. Then, provided that a chain rule

$$\partial_g (f + \alpha p)(x^0) \subset \partial_g f(x^0) + \partial_g (\alpha p)(x^0) \tag{2.19}$$

is valid, one directly obtains (2.18) with respect to the subdifferential under consideration. We refer the reader who is interested in recent results devoted to inclusion (2.19) for particular subdifferentials, to [MS97a, Kru00, NT01, Kru01]. For the related subdifferential-theory (mainly of certain limiting Fréchet subdifferentials), the Lipschitz property of f and p as well as the fact that X is an Asplund space play an important role, see also ε-Fréchet subdifferentials in §2.2.2.

Linear Inequality Systems with Variable Matrix

The upper Lipschitz property is particularly important for linear inequalities and polyhedral multifunctions. Basic results on this subject go back to Hoffman [Hof52], Walkup and Wets [WW69] and Nožička [NGHB74]. A complete theory of upper Lipschitz properties in the polyhedral case has been elaborated by Robinson [Rob76b, Rob81], for some (incomplete) summary see Theorem 6.4 below. Various results concerning the nonlinear case have been shown in [Rob76c] for the first time.

Fixed Matrix

The solution set of a parametric, finite dimensional linear inequality system

$$S(y) = \{x \in \mathbb{R}^n \mid Ax \leq y\},\ y \in Y \subset \mathbb{R}^m,\ A \text{ is an } (m,n) \text{ matrix},$$

is Lipschitz u.s.c. at y^0. This is a consequence of the famous *Hoffman's lemma*

Lemma 2.7 (Hoffman [Hof52]). *There is some constant L depending only on A such that the inequality*

$$\text{dist}\,(x, S(y)) \;\leq\; L \max_{1 \leq i \leq m} (A_{i\cdot}\, x \;-\; y_i)^+$$

holds for all $x \in \mathbb{R}^n$ and all y with $S(y) \neq \emptyset$, where $A_{i\cdot}$ is the i^{th} row of A. ◇

The lemma has initiated various investigations and proofs in order to find error bounds and best estimates (constants *L)* for linear inequalities or linear programs/complementarity problems (see, e.g., {Rob73, Rob76b, Man81a, Kla87, Man90, Li93, KT96]) as well as global error bounds for convex multifunctions or convex inequalities (see, e.g., [Rob75, Man85, LP97, BT96, Pan97, Kla98, LS98, KL99]). One of the first extensions of the lemma to Banach spaces can be found in [Iof79b].

If there is some x^0 with $Ax^0 < y^0$ (component–wise), then $S(y)$ is nonempty for y near y^0. Hence S^{-1}, i.e., $F(x) = \{y | Ax \leq y\}$, is *upper regular* at $(S(y^0), y^0)$. Moreover, using $Ax^0 < y^0$, one easily shows that F is even *pseudo-regular* at each pair (x, y^0) with $Ax \leq y^0$.

Having only $y^0 \in \text{dom}\, S$ then, after identifying *Y with* dom *S*, these statements remain true, and $\text{dom}\, S = A\mathbb{R}^n + \mathbb{R}^m_+$ is an unbounded, convex polyhedral set.

Lipschitzian Matrix

If $S(y) = \{x \in \mathbb{R}^n | Ax \leq b(y)\}$, where $b \in \mathbb{R}^m$ depends (locally) Lipschitzian on $y \in Y \subset \mathbb{R}^k$, then analogue statements as above can be immediately derived.

If also A depends Lipschitzian on y, then dom S is no longer closed and S is not locally upper Lipschitz, in general. Even if $S(y^0)$ is *non-empty and bounded,* the map S is not necessarily l.s.c. But the *upper Lipschitz* behavior remains valid. The next lemma is well-known (see Robinson [Rob77]) and applies to the continuous and Lipschitz continuous situation in the same manner.

Lemma 2.8 (Lipschitz u.s.c. linear systems). *Let*

$$S(y) = \{x \in \mathbb{R}^n \mid A(y)x \leq b(y)\},\ y \in Y \subset \mathbb{R}^k,$$

let A and b be pointwise Lipschitz at y^0, and let $S(y^0)$ be non-empty and bounded. Then S is Lipschitz u.s.c. at y^0. ◇

Proof. Let $x(y) \in S(y)$ and $\|y - y^0\|$ be small. Writing

$$A(y) = A(y^0) + C(y) \quad \text{and} \quad b(y) = b(y^0) + c(y)$$

we have, with some L and related norms,

$$\|C(y)\| \le L\|y - y^0\| \quad \text{and} \quad \|c(y)\| \le L\|y - y^0\|.$$

If, for certain $y \to y^0$, the elements $x(y)$ diverge, then division by $\|x(y)\|$ yields (for some subsequence of y)

$$x(y)/\|x(y)\| \to u \quad \text{and} \quad A(y^0)u \le 0.$$

In this case, $S(y^0)$ contains the ray $\{x^0 + \lambda u \mid \lambda \ge 0\}$ and is unbounded.

Hence, if $\|y - y^0\|$ is small enough, there is an upper bound K for $\|x(y)\|$. Then, setting

$$q = c(y) - C(y)x(y),$$

it holds

$$\|q\| \le (1 + K)L\|y - y^0\| \quad \text{and} \quad A(y^0)x(y) \le b(y^0) + q.$$

Since, for the fixed matrix $A(y^0)$, S is Lipschitz u.s.c. at y^0 with some rank L_A, now the estimate

$$x(y) \in S(y^0) + L_A\|q\|B \subset S(y^0) + L_A(1 + K)L\|y - y^0\|B$$

completes the proof. □

The reader will easily see that S is still u.s.c. at y^0 if both A and b are continuous at y^0 and $S(y^0)$ is non-empty and bounded.

Application to Lagrange Multipliers

Let $\Lambda(x,t) \subset \mathbb{R}^{m+\kappa}$ be the (possibly empty) set of Lagrange multipliers, assigned to a feasible point x and to some parameter $t \in \mathbb{R}^p$ of an optimization problem

$$\min_x\{f(x,t) \mid g_i(x,t) \le 0,\ h_k(x,t) = 0,\ i = 1,\dots,m;\ k = 1,\dots,\kappa\},$$

and suppose that

$$f, g_i, h_k \in C^{1,1}.$$

Then we have $(y,z) \in \Lambda(x,t)$ if and only if

$$D_x f(x,t) + D_x g(x,t)^{\mathsf{T}} y + D_x h(x,t)^{\mathsf{T}} z = 0,\ y \ge 0 \quad \text{and} \quad \langle y, g(x,t)\rangle = 0.$$

Let $I(x,t) = \{i \mid g_i(x,t) = 0\}$ (the index set assigned to active inequalities) and put, for fixed (x^0,t^0), $I^0 := I(x^0,t^0)$ and

$$\Lambda^0(x,t) := \left\{ (y,z) \in \mathbb{R}^{m+\kappa} \;\middle|\; \begin{array}{l} D_x g(x,t)^{\mathsf T} y + D_x h(x,t)^{\mathsf T} z = -D_x f(x,t), \\ y \geq 0 \text{ and } y_i = 0 \text{ if } i \notin I^0 \end{array} \right\}.$$

Due to $I(x,t) \subset I^0$ for (x,t) near (x^0,t^0), we observe that

$$\Lambda(x,t) \subset \Lambda^0(x,t) \quad \text{and} \quad \Lambda(x^0,t^0) = \Lambda^0(x^0,t^0).$$

Hence, Lemma 2.8 immediately ensures (by setting there $y := (x,t)$, $x := (y,z)$ and $S = \Lambda$) the following well-known result (cf., e.g., [Kla91, BS00]).

Corollary 2.9 (Lipschitz u.s.c. multipliers). *Provided that* $\Lambda(x^0,t^0)$ *is non-empty and bounded, the multiplier maps* Λ^0 *and* Λ *are Lipschitz u.s.c. at* (x^0,t^0).

If, in addition, $\operatorname{card}\Lambda(x^0,t^0) = 1$, *then the restricted map* $\Lambda|_{\operatorname{dom}\Lambda}$ *is Lipschitz l.s.c. at* (x^0,t^0). ◇

Note that, by Gauvin's theorem [Gau77], $\Lambda(x^0,t^0)$ is non-empty and bounded if and only if x^0 is a stationary point satisfying *MFCQ*, while $\operatorname{card}\Lambda(x^0,t^0) = 1$ means in algebraic formulation just the so-called *strict MFCQ* condition [Kyp85], for both results see also Lemma A.7 in the Appendix. The requirement of Λ^0 being Lipschitz l.s.c. at (x^0,t^0) with respect to $\operatorname{dom}\Lambda^0$ implies rank conditions concerning submatrices of $(D_x g(x,t), D_x h(x,t))$ on $\operatorname{dom}\Lambda^0$ which are often written as the so-called *constant rank condition* [Jan84]. Having only $f, g_i, h_k \in C^1$, the same arguments ensure that Λ^0 and Λ are at least u.s.c. at (x^0,t^0).

Upper Regularity and Newton's Method

Let $F : X \rightrightarrows Y$ be upper regular at (x^0,y^0) with rank L and neighborhoods U, V, and let $g : U \to Y$ be a (pointwise) Lipschitz function with

$$g(x^0) = y^0 \quad \text{and} \quad d(g(x), g(x^0)) \leq \beta d(x,x^0) \quad \text{for } x \text{ near } x^0.$$

Then (evidently) the map $H(x) = F^{-1}(g(x)) \cap U$ is locally upper Lipschitz at (x^0,y^0) with rank βL. Supposing

$$\Theta := \beta L < 1 \quad \text{and } x^1 \text{ close to } x^0, \tag{2.20}$$

the iteration process

$$x^{k+1} \in F^{-1}(g(x^k)) \cap U; \quad k \geq 1 \tag{2.21}$$

generates a (possibly not unique) sequence satisfying

$$d(x^{k+1},x^0) \leq \Theta d(x^k,x^0); \text{ in particular, } x^k \to x^0 \text{ and } g(x^k) \to y^0.$$

The same is true if F^{-1} is only locally upper Lipschitz at (y^0, x^0) with rank L and if *one knows* that x^{k+1} exists.

To obtain a *standard application* of the process (2.21), let $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, $f(x^0) = 0$, and let $Df(x^0)$ be a regular matrix. Setting

$$g(x) = x \quad \text{and} \quad F(x) = \{y \mid f(y) = Df(y)(y - x)\},$$

the process (2.21) describes just *Newton's method.* Indeed, we obtain $\beta = 1$, $x^0 = F^{-1}(x^0)$ and, considering x, y on small neighborhoods U, V of x^0, respectively,

$$\begin{aligned} F^{-1}(y) \cap U &= \{x \in U \mid y \in F(x)\} \\ &= \{x \in U \mid f(y) = Df(y)(y - x)\} \\ &= \{y - Df(y)^{-1} f(y)\}. \end{aligned}$$

So, it holds that

$$\xi \in F^{-1}(g(x)) \quad \Leftrightarrow \quad \xi = x - Df(x)^{-1} f(x)$$

and (2.21) describes Newton's method as asserted. Since

$$\begin{aligned} F^{-1}(y) - F^{-1}(x^0) &= y - Df(y)^{-1} f(y) - (x^0 - Df(x^0)^{-1} f(x^0)) \\ &= (y - x^0) - Df(y)^{-1}(f(y) - f(x^0)) = o(y - x^0), \end{aligned}$$

the assumption $\beta L < 1$ of (2.20) is valid with *arbitrarily small* $L > 0$. The latter ensures (locally) superlinear convergence.

2.2 Pseudo-Regularity

Pseudo-regularity is the most interesting and most complicated stability property we are dealing with in the present book and, in fact, there are still several open questions (mainly of topological nature) concerning this property even for Lipschitz functions. For instance, it was a big step ahead to know that, if $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is pseudo-regular at $(x^0, 0)$ and directionally differentiable at x^0, the zero x^0 is necessarily isolated (cf. [Fus99, Fus01] and Theorem 5.12). Notice that this statement is nearly trivial for $f \in C^1$. If f is not directionally differentiable at x^0, then the same question is completely open.

Pseudo-regularity may be (and has been) characterized by different means. For mappings in finite dimensional spaces, contingent and coderivatives describe the problem sufficiently well. In more general spaces, our Example BE.2 restricts these approaches essentially. For this reason, limits of certain Ekeland points, which describe the "pseudo-singular" situation, will be taken into consideration. In addition, we investigate the intersection of pseudo-regular maps by means of assigned inverse families. Such families come into the play if one studies the inverses of pseudo- regular maps in detail.

First of all, we establish a general connection between calm, upper, lower and pseudo-Lipschitzian maps S and the optimality condition of Theorem 2.6.

Theorem 2.10 (selection maps and optimality condition). *Let $S : Y \rightrightarrows X$ be pseudo-Lipschitz at (y^0, x^0) with rank L and neighborhoods U, V. Let $\varepsilon > 0$ be fixed such that $x^0 + \varepsilon B \subset U$, and define*

$$\Gamma(y) = S(y) \cap \{x | \ d(x, x^0) < \varepsilon - Ld(y, y^0)\}.$$

Then,

(i) Γ *is Lipschitz u.s.c. at* $(y^0, \Gamma(y^0))$
(ii) Γ *is Lipschitz l.s.c. at all* (y, x), $y \in V$, $x \in \Gamma(y)$ *(both with rank L);*
(iii) The functions

$$p_S(x) = \textbf{dist}\,((y^0, x), \text{gph}\, S) \ \textit{and} \ p_\Gamma(x) = \textbf{dist}\,((y^0, x), \text{gph}\, \Gamma)$$

coincide for x *near* x^0;
(iv) If $f : X \to \mathbb{R}$ *is locally Lipschitz and* x^0 *(locally) minimizes* f *on* $S(y^0)$ *then, provided that* α *is large enough,* x^0 *is a free local minimizer of*

$$P(x) := f(x) + \alpha p_S(x).$$

Moreover, the statements (i), (iii) and (iv) remain true if $S : Y \rightrightarrows X$ *is only calm at* (y^0, x^0) *with rank L.* ◇

Proof. (i) Let $x \in \Gamma(y)$, $y \in V$. Setting $y' = y^0$ in (1.1) there is some $x' \in S(y^0)$ with $d(x', x) \leq Ld(y, y^0)$. Notice that x' also exists (by definition) if S is calm at (y^0, x^0) with related neighborhoods U, V and rank L. Moreover, since

$$d(x', x^0) \leq d(x', x) + d(x, x^0) < Ld(y, y^0) + (\varepsilon - Ld(y, y^0)) = \varepsilon,$$

we obtain $x' \in \Gamma(y^0)$ and $\textbf{dist}\,(x, \Gamma(y^0)) \leq Ld(y, y^0)$.

(ii) Let $y \in V$ and $x \in \Gamma(y)$. Then,

$$d(x, x^0) = -p + \varepsilon - Ld(y, y^0) \ \text{ for some } p > 0.$$

Consider any $y' \in V$ such that $r := d(y', y)$ satisfies $0 < r < p/(2L)$. By (1.1), there is some $x' \in S(y')$ such that $d(x', x) \leq Ld(y', y) \leq Lr$. To show that $x' \in \Gamma(y')$ we estimate $d(x', x^0)$ by applying $2Lr < p$ and $-d(y, y^0) < r - d(y', y^0)$:

$$\begin{aligned}
d(x', x^0) &\leq d(x', x) + d(x, x^0) \ \leq Lr + d(x, x^0) \\
&= Lr - p + \varepsilon - Ld(y, y^0) \\
&< Lr - p + \varepsilon + Lr - Ld(y', y^0) \\
&\leq 2Lr - p + \varepsilon - Ld(y', y^0) \ < \varepsilon - Ld(y', y^0).
\end{aligned}$$

So $x' \in \Gamma(y')$ holds as required.

(iii) Clearly, $p_S(x^0) = p_\Gamma(x^0)$ is evident and $p_S \leq p_\Gamma$ follows from $\text{gph}\,\Gamma \subset \text{gph}\, S$. To verify $p_\Gamma(x) \leq p_S(x)$ for x near x^0, we consider any $x \in U$ such that $x \neq x^0$ and

$$(3 + 2L)\ d(x, x^0) < \varepsilon. \tag{2.22}$$

Let $(y', x') \in \operatorname{gph} S$ realize the distance $p_S(x)$ up to an error $\lambda d(x, x^0)$, $\lambda \in (0,1)$. We show that $(y', x') \in \operatorname{gph} \Gamma$. Indeed, since $(y^0, \Gamma(y^0)) \subset \operatorname{gph} S$, it holds

$$\begin{aligned} \max\{d(x',x), d(y',y^0)\} &\le p_S(x) + \lambda d(x,x^0) \\ &\le \operatorname{dist}(x, \Gamma(y^0)) + \lambda d(x,x^0) \\ &\le (1+\lambda) d(x,x^0) \\ &< 2d(x,x^0). \end{aligned}$$

Thus, the inequalities

$$d(x',x^0) \le d(x',x) + d(x,x^0) < 3d(x,x^0) \text{ and } d(y',y^0) < 2d(x,x^0)$$

are valid and ensure $(y', x') \in \operatorname{gph} \Gamma$ whenever

$$3d(x,x^0) < \varepsilon - 2Ld(x,x^0).$$

The latter holds true due to (2.22) and yields $p_\Gamma(x) \le p_S(x) + \lambda d(x,x^0)$ as well as $p_\Gamma(x) \le p_S(x)$ via $\lambda \downarrow 0$.

(iv) For sufficiently small $\varepsilon > 0$, x^0 minimizes f on $\Gamma(y^0)$. By (i), Γ is Lipschitz u.s.c., so x^0 is, by Lemma 2.1 and Theorem 2.6, a local minimizer of $P(x) := f(x) + \alpha p_\Gamma(x)$ whenever α is sufficiently large. Using (iii), this is the assertion. □

2.2.1 The Family of Inverse Functions

Let us consider the point x' satisfying the requirements (1.1) of pseudo-regularity, namely

$$x' \in S(y') = F^{-1}(y') \text{ and } d_X(x', x) \le L d_Y(y', y), \tag{2.23}$$

as a function of $z = (x,y)$ and y' for $x \in U$, $y, y' \in V$.

Then, F is pseudo-regular at z^0 iff there is a family Ψ of functions $\psi_z : V \to X$ such that $x' := \psi_z(y') \in F^{-1}(y')$ and, for all $z \in \Omega := (U,V) \cap \operatorname{gph} F$, one has $d(\psi_z(y'), x) \le Ld(y', y)$ whenever $y' \in V$.

So Ψ is a special family of selections ψ_z for F^{-1} which tells us that

F^{-1} is Lipschitz lower semicontinuous (l.s.c.) near z^0 with uniform rank L, i.e., given $z \in \Omega$, one finds some neighborhood V_z of y such that $\operatorname{dist}(x, F^{-1}(y')) \le Ld(y', y) \;\; \forall y' \in V_z$, (2.24)

and, in addition,

$$\text{the neighborhoods } V_z \text{ are fixed: } V_z \equiv V \;\forall z \in \Omega. \tag{2.25}$$

We will say that ψ_z is a *local inverse* of F, and Ψ is an *inverse family*.

The functions ψ_z are not unique a priori and may be discontinuous at $y' \in V\backslash\{y\}$. So, an inverse family, assigned to a pseudo-regular mapping F, may consist (at least theoretically) of more or less complicated selections of F^{-1}.

To get a close connection between inverse families and cones of tangents or normals, we consider the directions $x' - x$ and $y' - y$ for (linear) normed spaces $(X, \|\cdot\|)$ and $(Y, |\cdot|)$, respectively. Setting $y' = y + tv$, $t > 0$, $|v| = 1$ and $x' = x + tu$, we have $\psi_z(y') = x + tu$. Now, each $u = u_z(v, t)$ is bounded since $t\|u\| = \|x' - x\| \leq L|y' - y| = Lt$. So one may identify Ψ and a related family Φ of uniformly bounded functions u_z, defined for $v \in \text{bd}\, B_Y$ and for all t in some interval $(0, \delta)$, such that

$$y + tv \in F(x + tu_z(v, t)) \quad \text{if} \quad z = (x, y) \in \Omega.$$

Indeed, having Φ, one easily sees (by "decreasing" U and V) that F is pseudo-regular at z^0. We call Φ an *inverse family of directions*. By our definitions, we have

Remark 2.11 (inverse families and pseudo-regularity). The following conditions are equivalent to each other:

1. An inverse family Ψ exists.
2. An inverse family of directions Φ exists.
3. F is pseudo-regular at z^0. ◇

Up to now, the domain of all $\psi_z \in \Psi$ was a constant neighborhood V of y^0, containing the second component y of $z = (x, y)$, while all functions $u_z(v, \cdot)$ were defined on the same interval $(0, \delta)$. If these domains are replaced by different neighborhoods V_z of y and intervals $(0, \delta(z))$, respectively, the existence of Ψ (or Φ) describes the l.s.c. property (2.24) of F^{-1}. To indicate that we understand the families Ψ and Φ in this weaker sense, we denote them by Ψ^w and Φ^w, respectively. Our Theorem 2.17 will say that the existence of Ψ and Ψ^w - or pseudo-regularity and the l.s.c. property (2.24) - are equivalent for quite general mappings F.

Particular Local Inverses

Knowing that, for certain maps, there are *particular* local inverses, gives us extra information about *where or how* one can find some x' satisfying (1.1). It can also mean that not all variations y' of y must be regarded. In this way the question of whether F is pseudo-regular or not can be simplified. Let us regard some examples.

Simple Cases

1. Let X be a linear normed space, let $f : X \to \mathbb{R}$ be locally Lipschitz and $F(x) = \{y \in \mathbb{R} \mid f(x) \leq y\}$.

Provided that Clarke's directional derivative $f^c(x^0; p^0)$ is negative for some direction p^0, one may put

$$(x' =)\, \psi_z(y') = x + |y' - y| p^0 \tag{2.26}$$

in order to see that F is pseudo-regular at $z^0 = (x^0, f(x^0))$. The related inverse family consists of functions that move $x = \psi_z(y)$ "Lipschitzian far" in a fixed direction p^0 only.

2. If, moreover, the locally Lipschitz function f is convex or continuously differentiable, then one may even state:

Either F is not pseudo-regular at z^0 or there is an inverse family that consists of functions ψ_z having the form (2.26).

Indeed, if $f \in C^1$, then either F is not pseudo-regular at z^0, namely if $Df(x^0) = 0$, or one can take any direction p^0 with $\langle Df(x^0), p^0\rangle < -1$.

If f is convex and continuous, put $p^0 = x^* - x^0$ where x^* is any point with $f(x^*) < f(x^0)$ as far as x^* exists. Otherwise, F cannot be pseudo-regular at z^0 since $F^{-1}(f(x^0) - \varepsilon)$ is empty for $\varepsilon > 0$.

3. The reader will easily confirm that, by setting $p^0 = (1, ..., 1) \in \mathbb{R}^n$, the *either–or–statement* of 2. is true for each mapping

$$F(x) = \{y \in \mathbb{R}^m \mid f(x) \geq y\}, \quad f : \mathbb{R}^n \to \mathbb{R}^m,$$

provided that the (possibly discontinuous) components f_i are *non-decreasing in each component x_k.*

In fact, any point x' satisfying $f(x') \geq y'$ can be replaced by $x'' \geq x'$ without violating this inequality. So one may replace x' in (2.23) by $x'' = x + \|x' - x\| p^0 \geq x'$ after taking the maximum-norm. Having $\|x' - x\| \leq L|y' - y|$, now x'' can be again replaced by $\xi = x + (L|y' - y|)p^0 \geq x''$ whereafter $f(\xi) \geq y'$ remains true and the required Lipschitz estimate holds with another constant (or equivalent norm) only.

For instance, f_i may be an *extreme value function* (also called *marginal function*) of the type

$$f_i(x) = \inf\{h_i(z) \mid g_{ki}(z) \geq x_k;\ k = 1, \ldots, n\}$$

with arbitrary functions h_i and g_{ki}, or f_i is a probability distribution function or an utility function.

More Complicated Cases

1. *Discontinuous functions ψ_z* are needed for those pseudo-regular functions $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ which are not strongly regular at $z^0 = (x^0, f(x^0))$.

Indeed, in that case, the inverse f^{-1} does not possess a (single-valued) selection $s(\cdot) \in f^{-1}(\cdot)$, that is continuous on some neighborhood V of $f(x^0)$ and satisfies $s(f(x^0)) = x^0$, cf. Theorem 5.10. Thus, every local inverse ψ_z for

$z = z^0$ is *necessarily* discontinuous at some point of V.

2. *Pseudo-Lipschitzian level sets without descent directions in Hilbert spaces:* Our Example BE.2, where $F(x) = \{y \in \mathbb{R} \mid f(x) \leq y\}$ and $F^{-1}(y)$ is a lower level set map for $f(x) = \inf_k x_k$ on $X = l^2$, is helpful in order to understand pseudo-regularity in general spaces. It shows that locally inverse functions may have (necessarily) rather bad properties.

In particular, one has to take into account that the bounded directions $u_z(v,t)$ of every inverse family of directions Φ do not necessarily converge as $t \downarrow 0$. Even accumulation points may fail to exist, and directional derivatives, for certain x arbitrarily close to x^0, can satisfy the first-order minimum condition $f'(x;u) \geq 0\ \forall u \in X$.

Note that, in our example, f is one of the simplest nonsmooth, non-convex functions on a Hilbert space: f is globally Lipschitz, directionally differentiable and concave.

2.2.2 Ekeland Points and Uniform Lower Semicontinuity

In this section, we characterize pseudo-regularity by two topological means:

(i) by so–called Ekeland points, related to the distance functions $d_y(x) = \operatorname{dist}(y, F(x))$,

(ii) by Lipschitz lower semicontinuity of F^{-1} near the reference point.

The first characterization is our basic tool, the second one will help to understand the content of pseudo-regularity. We will require that Lipschitz l.s.c. holds with uniform rank L at the points in question. But the neighborhoods where the l.s.c. estimates are true, may have different size. For this fact, the notion *uniform lower semicontinuity* will be used.

We start with the formal *negation* of pseudo-regularity: A multifunction $F : X \rightrightarrows Y$ (between metric spaces) is *not pseudo-regular* at $z^0 = (x^0, y^0) \in \operatorname{gph} F$ iff

$$\begin{array}{l}\text{there are sequences } x^k \to x^0\ (k \to \infty) \text{ and } \eta^k,\ y^k \to y^0 \\ \text{such that } \eta^k \in F(x^k) \text{ and } \operatorname{dist}(x^k, F^{-1}(y^k)) > k d(\eta^k, y^k).\end{array} \tag{2.27}$$

Here, we have identified L, x, y and y' appearing in the definition with k, x^k, η^k and y^k, respectively. For instance, (2.27) holds for $f(x) = x^3$ with $x^k = \eta^k = 0$ and $y^k \downarrow 0$.

In order to show that such points cannot exist, we want to obtain additional information about possible sequences. This is the purpose of our next considerations where we replace x^k in (2.27) by an Ekeland-point z^k of the function $\phi(x) = \operatorname{dist}(y^k, F(x))$.

We apply *Ekeland's variational principle* [Eke74] in the following form.

Theorem 2.12 (Ekeland). *Let X be a complete metric space and $\phi : X \to \mathbb{R} \cup \{\infty\}$ be a l.s.c. function having a finite infimum. Let ε and α be positive, and let $\phi(x) \leq \varepsilon + \inf_X \phi$. Then there is some $z \in X$ such that*

$$d(z,x) \leq \alpha, \ \phi(z) \leq \phi(x) \ \text{ and } \ \phi(\xi) + (\varepsilon/\alpha) d(\xi, z) \geq \phi(z) \ \forall \xi \in X.$$

$\diamond$

A proof of the above theorem will be added in the *appendix.*

As usually, a function ϕ with values in $\mathbb{R} \cup \{\infty\}$ is called l.s.c. if all lower level sets $\{x \mid \phi(x) \leq r\}$ are closed. In applications, X is often a closed subset of a Banach space. Points x satisfying $\phi(x) \leq \varepsilon + \inf_X \phi$ are also said to be ***ε–optimal.***

We say that $z \in X$ is a *local Ekeland-point* of a functional ϕ with factor p, for short $z \in E_\phi(p)$, if $\phi(z)$ is finite and

$$\phi(\xi) + pd(\xi, z) \geq \phi(z) \quad \forall \xi \in X, \ \xi \text{ near } z. \tag{2.28}$$

If (2.28) holds for all $\xi \in X$ we call z a *global Ekeland-point.*

Via $[\phi(\xi) - \phi(z)]/d(\xi, z) \geq -p$, property (2.28) ensures that $-p$ is a lower bound for several generalized directional derivatives of ϕ at z.

Note that in Example 0.1 we had $f(x) = x$ (if $x \leq 0$) and $f(x) = x^2$ (if $x > 0$). For each $p \in (0,1)$, one finds here local Ekeland-points with $z > 0$, but not with $z < 0$.

If X is a normed space and $f : X \to \mathbb{R} \cup \{\infty\}$, one may introduce the mapping (of approximate local minimizers)

$$M_p(x^*) = \{x \in X \mid x \in E_\phi(p) \text{ for } \phi = f - x^*\}, \ x^* \in X^*.$$

Its inverse M_p^{-1} assigns, to $x \in X$, some subset of X^*, and defines via

$$\partial_\varepsilon^F f(x) = \cap_{p > \varepsilon} M_p^{-1}(x)$$

just the so-called ***ε-Fréchet*** *subdifferential* of f at x. This subdifferential can be explicitly defined by saying that $x^* \in \partial_\varepsilon^F f(x)$ if

$$\liminf_{\xi \to x, \xi \neq x} [f(\xi) - f(x) - \langle x^*, \xi - x \rangle]/\|\xi - x\| \geq -\varepsilon.$$

It has been extensively studied in the literature during the last years. For its behavior as $x \to x^0$ and/or $\varepsilon \downarrow 0$ and applications as well, we refer to [Kru85, Fab86, Fab89, Kru96, Kru97, NT01] and [Iof00] where the reader finds not only a comprehensive overview, but also various further references. As an introduction into the rich world of ideas for generalized derivatives, subdifferentials and their applications, one has to mention (even now) the paper [Iof79a].

The idea of dealing with pseudo-regularity by applying Ekeland's variational principle, was a basic one and goes back to J.P. Aubin and I. Ekeland. The proof of Theorem 4, §7.5, in [AE84] is a typical example for realizing this idea. A. Ioffe [Iof79b] and A. Auslender [Aus84] also used the Ekeland principle very early as a crucial tool in a more special context devoted to optimality conditions and finite systems of (in)equalities, respectively. There, sufficient conditions for F being pseudo-regular have been derived in terms of generalized derivatives.

Here, we add a condition which is both necessary and sufficient for important special cases which will be summarized in Lemma 2.13 below.

We say that $F : X \rightrightarrows Y$ is *proper* near y^0 if, for fixed y near y^0, the function $\text{dist}\,(y, F(\cdot))$ - with values in $\mathbb{R}^+ \cup \{\infty\}$ - is l.s.c. on X.

Lemma 2.13 (proper multifunctions). *A multifunction $F : X \rightrightarrows Y$ is proper (everywhere) under each of the following assumptions.*

(i) F is a continuous function;

(ii) F is closed and $Y = \mathbb{R}^n$;

(iii) $F(x) = f(x) - K$, $K \neq \emptyset$ is a closed convex set in a real Hilbert space Y and $f : X \to Y$ is a continuous function.

(iv) F is closed and locally compact.

(v) $F(x) = f(x) + \Phi(x)$ where X, Y are Banach spaces, $f : X \to Y$ is continuous and Φ satisfies assumption (iv).

In each of these cases, the sets $F(x)$ are closed and $\text{dist}\,(y, F(x))$ *will be attained, i.e.,* $\text{dist}\,(y, F(x)) = d(y, f)$ *for some $f \in F(x)$, provided that $F(x) \neq \emptyset$.* ◇

Proof. (i) is trivial.
(ii) and (iv) follow by compactness arguments in $\mathbb{R}^n$ or $\text{cl}\,F(\Omega_x)$.
(v): Note that $\text{dist}\,(y, F(x)) = \text{dist}\,(y - f(x), \Phi(x))$, and apply (iv).
(iii): The existence and the lower semicontinuity follow from

$$\begin{aligned} \text{dist}\,(y, F(x)) &= \inf\{\|f(x) - y - k\| \mid k \in K\} \\ &= \|f(x) - y - \pi_K(f(x) - y)\|, \end{aligned}$$

where π_K is the non-expansive projection map onto K. □

Lemma 2.14 (pseudo-regularity for proper mappings). *Let X and Y be metric spaces, let X be complete, and let $F : X \rightrightarrows Y$ be proper near $y^0 \in F(x^0)$. Then, F is not pseudo-regular at (x^0, y^0) if and only if for each $p > 0$, there exist $z \in \text{dom}\,F \cap (x^0 + pB_X)$, $y \in y^0 + pB_Y$ and $r > 0$ (all depending on p) such that z is a global Ekeland-point of* $\text{dist}\,(y, F(\cdot))$ *with factor p, and the inequalities*

$$d(\eta, y^0) < p \text{ and } \text{dist}\,(z, F^{-1}(y)) > p^{-1} d(y, \eta) \tag{2.29}$$

are true whenever

$$\eta \in F(z) \quad \text{and} \quad d(y, \eta) < \text{dist}\,(y, F(z)) + r. \tag{2.30}$$

◇

Note: Let the given condition be satisfied. Then, η in (2.30) exists since $z \in \operatorname{dom} F$. So, applying (2.29), it holds $y \notin F(z)$ and $\operatorname{dist}(y, F(z)) \le d(y, \eta) < 2p$. ◇

Proof of Lemma 2.14. ($\Leftarrow$) For this direction, neither the l.s.c.-assumption nor any Ekeland property of z is needed. It suffices to know that $F(z) \neq \emptyset$. According to our previous note, we may put $p = 1/k$, $x^k = z$, $y^k = y$ and $\eta^k = \eta$, where η satisfies (2.30).

Now (2.29) ensures $\eta^k \to y^0$ and $\operatorname{dist}(x^k, F^{-1}(y^k)) > kd(y^k, \eta^k)$. Thus (2.27) holds true.

($\Rightarrow$) Let $p > 0$ be given. We put $C = p^{-1}$, assume that (2.27) is true, and fix any $k > 3C$ with

$$\max\{Cd(\eta^k, y^k), 2d(\eta^k, y^k), d(x^k, x^0), d(y^k, y^0)\} < \tfrac{1}{2}p.$$

Setting $y = y^k$ and $\varepsilon = d(\eta^k, y)$, ε is positive by (2.27), and $\varepsilon < p/4$, $d(y, y^0) < p$ are true by the choice of k. From $\eta^k \in F(x^k)$, we have $\operatorname{dist}(y, F(x^k)) \le d(y, \eta^k) = \varepsilon$. Hence x^k is ε–optimal for the l.s.c. functional $\phi(\cdot) = \operatorname{dist}(y, F(\cdot))$ on X. To replace x^k by an Ekeland point, we put $\alpha = C\varepsilon$ in Theorem 2.12: There exists a global Ekeland point z with factor p such that

$$\phi(z) = \operatorname{dist}(y, F(z)) \le \varepsilon \ \text{ and } \ d(z, x^k) \le C\varepsilon < \tfrac{1}{2}p. \tag{2.31}$$

In particular, we observe that

$$d(x^0, z) < d(x^0, x^k) + d(x^k, z) < p, \quad \text{and} \quad z \in \operatorname{dom} F.$$

Next consider any η satisfying (2.30) with $r = \varepsilon$ (η should not be confused with the already fixed η^k). Using (2.31) we observe $\varepsilon \ge \operatorname{dist}(y, F(z)) > d(y, \eta) - r$, hence

$$2\varepsilon > d(y, \eta). \tag{2.32}$$

Now the estimate

$$d(\eta, y^0) \le d(\eta, y) + d(y, y^0) < 2\varepsilon + \tfrac{1}{2}p < p$$

verifies the first inequality in (2.29). On the other hand, it holds – due to (2.27) and by the choice of k –

$$\operatorname{dist}(x^k, F^{-1}(y)) = \operatorname{dist}(x^k, F^{-1}(y^k)) > kd(\eta^k, y^k) = k\varepsilon > 3C\varepsilon.$$

By (2.31), the latter ensures the estimate

$$\operatorname{dist}(z, F^{-1}(y)) \ge \operatorname{dist}(x^k, F^{-1}(y)) - d(x^k, z) > 3C\varepsilon - C\varepsilon = 2\varepsilon C.$$

Taking again (2.32) into account, we thus obtain

$$\operatorname{dist}(z, F^{-1}(y)) > Cd(y, \eta) = p^{-1}d(y, \eta).$$

So (2.30) implies (2.29), which completes the proof. □

From the proof of direction ($\Leftarrow$) one easily sees that Lemma 2.14 remains true after replacing the notion "global Ekeland-point" by "local Ekeland point".

In order to see what happens if gph F is not closed, consider

Example 2.15 (F is not pseudo-regular). For $x \in \mathbb{R}$, put

$$F(x) = \{y \in \mathbb{R} \mid y \text{ rational}\}.$$

Then, $F^{-1}(y) = \emptyset$ for irrational y, $F^{-1}(y) = \mathbb{R}$ for rational y.

Clearly, F is not pseudo-regular at $(0,0)$, and dist $(y, F(x)) = 0$ for all x. Thus each pair (y, z) trivially satisfies the inequality $\text{dist}(y, F(x)) + pd(x, z) \geq \text{dist}(y, F(z))$.

To fulfill the implication (2.30) $\Rightarrow$ (2.29), the pair (z, y) must be taken near the origin, say $|z| + |y| < \frac{1}{2}p$ with irrational y. With $r = \frac{1}{2}p$, now (2.29) follows from (2.30) due to

$$td(\eta, 0) \leq d(\eta, y) + d(y, 0) < p \quad \text{and} \quad \text{dist}(z, F^{-1}(y)) = \infty.$$

◇

The situation becomes simpler if F is supposed to have dosed images.

Theorem 2.16 (pseudo-regularity of proper mappings with closed ranges). *Let F be a closed-valued map satisfying the assumptions of Lemma 2.14- Then, F is* not *pseudo-regular at (x^0, y^0) if and only if for each $p > 0$, there exist $z \in \text{dom}\, F \cap (x^0 + pB_X)$ and $y \in y^0 + pB_Y$ such that z is a global Ekeland-point of* $\text{dist}(y, F(\cdot))$ *with factor p as well as* $0 < \text{dist}(y, F(z)) < 2p$. ◇

Proof. Let $\mu = \text{dist}(y, F(z))$ for some (y, z) under consideration.

The direction $(\Rightarrow)$ follows as for Lemma 2.14; concerning $0 < \mu < 2p$, see the Note following Lemma 2.14 and notice that $\mu > 0$ follows from $y \notin F(z)$ and the closedness of $F(z)$.

To show $(\Leftarrow)$ assume that F is pseudo-regular, contrarily to the assertion. Then, using the equivalence of both properties, F is metrically regular at (x^0, y^0) with some rank L. Since we may assume that (z, y) is close to (x^0, y^0), this yields $\text{dist}(z, F^{-1}(y)) \leq L\mu$. Because of $F(z) \neq \emptyset$ we have $\mu < \infty$ and obtain particularly $F^{-1}(y) \neq \emptyset$. On the other hand, it holds

$$\text{dist}(y, F(x)) = 0 \;\forall x \in F^{-1}(y) \quad \text{and} \quad \text{dist}(y, F(x)) + pd(x, z) \geq \mu.$$

So we derive

$$d(z, x) \geq p^{-1}\mu$$

as well as

$$L\mu \geq \text{dist}(z, F^{-1}(y)) = \inf\{d(z, x) | x \in F^{-1}(y)\} \geq p^{-1}\mu.$$

Due to $\mu > 0$, this yields $L \geq p^{-1} \to \infty$ as $p \downarrow 0$. So F cannot be pseudo-regular. □

Partial Inverses

Using Theorem 2.16, the notion of an inverse family (of directions) in §2.2.1 may be weakened without violating the equivalence with pseudo-regularity.

Let X be a metric space, $(Y, |\cdot|)$ be normed and $z^0 \in \operatorname{gph} F$. A *partial inverse* (with rank L) *at* $z = (x, y) \in \operatorname{gph} F$ is a mapping π_z that assigns, to $v \in \operatorname{bd} B_Y$, some sequence of $t \downarrow 0$ and related elements $x_t \in X, v_t \in Y$ such that

$$y + tv_t \in F(x_t), \quad d(x_t, x) \le tL \quad \text{and} \quad v_t \to v.$$

We say that F is *partially invertible* near z^0 if, for some neighborhood Ω of z^0 and some L, there is a partial inverse at each $z \in \Omega \cap \operatorname{gph} F$ with uniform rank L.

Recalling the convention $x + rB_X = \{x' \mid d(x', x) \le r\}$ in metric spaces, one can equivalently define

> F is partially invertible near z^0 if
>
> for each $z \in \operatorname{gph} F$ near z^0, all $v \in \operatorname{bd} B_Y$ and some fixed $L > 0$, it holds $0 = \liminf_{t\downarrow 0} t^{-1} \operatorname{dist}(y + tv, F(x + LtB_X))$.

In particular, F is partially invertible near z^0 if F^{-1} is Lipschitz l.s.c. near z^0 with uniform rank K, cf. condition (2.24). In fact, with $L = K + 1$, a partial inverse π_z can be defined by taking small t and setting $v_t \equiv v$ and $x_t \in F^{-1}(y + tv) \cap (x + tLB_X)$.

For showing partial invertibility of the lower level set map $F(x) = \{y \mid f(x) \le y\}$ of a continuous and directionally differentiable functional f on a Banach space, it is enough to assign, to x near x^0, some uniformly bounded direction $u(x)$ such that $f'(x; u(x)) < -1$.

Even if $u(\cdot)$ is discontinuous, now $\pi_z(v)$ may consist of all t in some interval $(0, \tau)$ (where τ depends on z and v) and of $v_t \equiv v, x_t = x + tu(x)$.

Theorem 2.17 (basic equivalences, proper mappings). *Let X be a complete metric space, Y be a normed space and $F : X \rightrightarrows Y$ be closed-valued and proper near $y^0 \in F(x^0)$. Moreover, let* $\operatorname{dist}(y, F(x)) = d(y, f)$ *hold for some $f \in F(x)$ whenever $F(x) \ne \emptyset$. Then, the following properties are equivalent to each other:*

(i) F is pseudo-regular at $z^0 = (x^0, y^0)$.

(ii) F is partially invertible near z^0.

(iii) For some neighborhood Ω of (y^0, x^0), F^{-1} is Lipschitz l.s.c. with uniform rank L at each $(y, x) \in \operatorname{gph} F^{-1} \cap \Omega$.

◇

Note: Concerning our assumptions we refer to Lemma 2.13,

Proof of Theorem 2.17. If F is pseudo-regular then there is an inverse family (see §2.2.1), so F is partially invertible and F^{-1} fulfills the l.s.c. condition. Thus, we have to show that F is pseudo-regular if F is partially invertible. Let us assume, contrarily to the assertion, that F is not pseudo-regular.

Consider any Ekeland pair (z, y), assigned to some $p > 0$ as in Theorem 2.16. Since $F(z) \neq \emptyset$, there is some $\eta \in F(z)$ such that $\operatorname{dist}(y, F(z)) = d(y, \eta) > 0$. Next put $v = (y - \eta)/d(y, \eta)$ and use our assumption. Hence, for certain $t = t_k \downarrow 0$, there are x_t and v_t with $\eta + tv_t \in F(x_t)$, $d(x_t, z) \leq tL$ and $v_t \to v$. If t is small then one finds

$$d(y, \eta + tv_t) \leq d(y, \eta) - \tfrac{1}{2}t. \tag{2.33}$$

By the Ekeland property of (z, y) with factor p, it holds for small t that

$$\operatorname{dist}(y, F(x_t)) + pd(x_t, z) \geq \operatorname{dist}(y, F(z)) = d(y, \eta).$$

The left-hand side can be estimated with the already shown relations

$$d(y, \eta) - \tfrac{1}{2}t \geq d(y, \eta + tv_t) \geq \operatorname{dist}(y, F(x_t)) \text{ and } tL \geq d(x_t, z),$$

thus

$$d(y, \eta) - \tfrac{1}{2}t + ptL \geq d(y, \eta)$$

yields

$$p \geq \tfrac{1}{2}L^{-1}.$$

So, the singularity condition $p \downarrow 0$ of Theorem 2.16 cannot be satisfied. □

It is worth noting that the key inequality (2.33) of the preceding proof is already true if $v_t \in \operatorname{bd} B$ and $\|v_t - v\| < \frac{1}{2}$. This means that the claim $v_t \to v$ in the definition of a partial inverse could be even weakened. The foregoing theorem also holds for closed F and Banach spaces X and Y, we refer to our discussion following corollary 3.3 below.

2.2.3 Special Multifunctions

Here, we apply Theorem 2.16 to particular forms of the map F.

Recall that $E_f(p)$ denotes the set of all local Ekeland points of $f : X \to \mathbb{R} \cup \{\infty\}$ with factor p, cf. (2.28).

Level Sets of L.s.c. Functions

Lemma 2.18 (pseudo-singular level sets of l.s.c, functions). *Let X be a complete metric space, $f : X \to \mathbb{R}$ be l.s.c., $F(x) = \{y \in \mathbb{R} \mid y \geq f(x)\}$ and $f(x^0) \leq y^0$. Then, F is not pseudo-regular at (x^0, y^0) if and only if $(x^0, y^0) = \lim(z(p), f(z(p)))$ for certain $z(p) \in E_f(p)$, $p \downarrow 0$.* ◇

Note: For continuous f, the condition takes the form: $f(x^0) = y^0$ and $x^0 \in \limsup_{p \downarrow 0} E_f(p)$. ◇

Proof of Lemma 2.18. It holds $\text{dist}(y, F(x)) = (f(x) - y)^+$ where $r^+ = \max\{0, r\}$. We may apply Theorem 2.16 since $(f(\cdot) - y)^+$ is l.s.c.

($\Rightarrow$) Let z and y be assigned to p as in Theorem 2.16. Then $0 < f(z) - y < 2p$ and, due to $z \to z^0$ and $y \to y^0$, one observes $\lim f(z) = y^0$. The local Ekeland-property of z yields

$$(f(\xi) - y)^+ + pd(\xi, z) \geq (f(z) - y)^+ > 0 \quad \forall \xi \text{ near } z \in X.$$

For ξ near z, it holds $f(\xi) > y$ because f is l.s.c. and $f(z) > y$. So we may write

$$f(\xi) - y + pd(\xi, z) \geq f(z) - y \text{ which ensures } z \in E_f(p).$$

($\Leftarrow$) Due to $z \in E_f(p)$ we have, with some $\alpha > 0$: $f(x) + pd(x, z) \geq f(z)\ \forall x \in z + \alpha B$. For each $\delta > 0$, now the both conditions $f(x) \leq f(z) - \delta$ and $x \in z + \alpha B$ yield

$$f(z) - \delta + pd(x, z) \geq f(x) + pd(x, z) \geq f(z);$$

hence $d(x, z) \geq p^{-1}\delta$.

Thus, the map F may be pseudo-regular at $(z, f(z))$ only with rank $L_z \geq p^{-1}$. Because $(z, f(z))$ converge to (x^0, y^0) as $p \downarrow 0$, F cannot be pseudo-regular at (x^0, y^0). □

Cone Constraints

Next, we consider case (iii) of Lemma 2.13 more in detail. Suppose that

$$\begin{aligned} &X \text{ is a complete metric space and } Y \text{ is a real Hilbert space with norm } |\cdot|, \\ &f \in C^{0,1}(X, Y),\ K \subset Y \text{ is nonempty, closed and convex,} \\ &F(x) = f(x) - K \text{ and } (x^0, y^0) \in \text{gph } F. \end{aligned} \tag{2.34}$$

We wrote *cone constraints* because K is a convex cone in many applications. The Lipschitz assumption is crucial for forthcoming estimates, the existence of inner points of K is not needed.

Let $\pi(x, y)$ be the projection of $f(x)$ onto the closed convex set $y + K$. Equivalently, this means that $f(x) - \pi(x, y)$ belongs to the normal cone of $y + K$ at $\pi(x, y)$. Clearly, $\pi(\cdot, \cdot)$ is locally Lipschitz. So pseudo- regularity of F can be reduced to the study of global Ekeland points z for the locally Lipschitz functionals $x \to \text{dist}(y, F(x)) = |f(x) - \pi(x, y)|$. Instead of norm-functionals, let us now use dual functions of the form $x \to \langle y^*, f(x) - \pi(x, y)\rangle$.

Lemma 2.19 (Ekeland-points of norm-functionals in a real Hilbert space). *Under (2.34), let* $g \in C^{0,1}(X, Y)$, $|g(z)| > 0$ *and* $y^* = |g(z)|^{-1} g(z)$.

(i) If $z \in E_h(p)$ *for* $h = |g|$, *then* $z \in E_u(p')$ *for* $u(\cdot) = \langle y^*, g(\cdot)\rangle$ *and all* $p' > p$.

(ii) Conversely, if $z \in E_u(p)$, *then* $z \in E_h(p)$.

◇

Proof. (i) We have, for some $\alpha > 0$, $|g(x)| + pd(x,z) \geq |g(z)| \forall x \in z + \alpha B$. Taking the square of both sides, we obtain

$$\langle g(x), g(x)\rangle + 2pd(x,z)|g(x)| + p^2 d(x,z)^2 \geq \langle g(z), g(z)\rangle.$$

With the notation

$$w = g(x) - g(z) \text{ and } A = \langle w, w\rangle + 2pd(x,z)|g(x)| + p^2 d(x,z)^2$$

this yields

$$\begin{aligned} 0 &\leq \langle g(z) + w, g(z) + w\rangle - \langle g(z), g(z)\rangle + 2pd(x,z)|g(x)| + p^2 d(x,z)^2 \\ &= 2\langle g(z), w\rangle + \langle w, w\rangle + 2pd(x,z)|g(x)| + p^2 d(x,z)^2 \\ &= 2|g(z)|\langle y^*, w\rangle + A. \end{aligned}$$

Let L_g be a Lipschitz constant for g near x^0. For z near x^0 and small α then w fulfills a Lipschitz estimate $|w| \leq L_g d(x,z)$. So, $|w|^2 = o(d(x,z))$, and the term A becomes

$$A = 2pd(x,z)|g(x)| + o(d(x,z)).$$

Hence,

$$0 \leq 2|g(z)|\langle y^*, w\rangle + 2pd(x,z)|g(x)| + o(d(x,z)). \tag{2.35}$$

Given any $\beta > 0$, we may further restrict x to a smaller neighborhood of z (if necessary), such that

$$2|g(x)| < (2+\beta)|g(z)| \text{ and } o(d(x,z)) \leq \beta p|g(z)|d(x,z).$$

Now (2.35) ensures

$$0 \leq 2|g(z)|\langle y^*, w\rangle + (2+\beta)pd(x,z)|g(z)| + \beta p|g(z)|d(x,z)$$

which is

$$-(2+2\beta)pd(x,z) \leq 2\langle y^*, w\rangle = 2\langle y^*, g(x) - g(z)\rangle.$$

So we see that z is a local Ekeland-point of

$$u(\cdot) = \langle y^*, g(\cdot)\rangle \text{ with factor } (1+\beta)p.$$

(ii) By the suppositions, it holds with some $\alpha > 0$ for all $x \in z + \alpha B$,

$$|g(z)| = \langle y^*, g(z)\rangle \leq \langle y^*, g(x)\rangle + pd(x,z) \leq |g(x)| + pd(x,z).$$

□

Returning to the particular (cone-) mapping F under consideration, we have to put $g(x) = f(x) - \pi(x,y)$ in order to obtain a characterization of pseudo-regularity by normalized functionals y^*. In view of Lemma 2.19, we define

$$v(\cdot, y^*, y) = \langle y^*, f(\cdot) - \pi(\cdot, y)\rangle.$$

To abbreviate we write $B^* = B_{Y^*}$ and $u = v(\cdot, y^*, y)$, being aware that y and y^* are fixed. Recall that $F(x) = f(x) - K$, and that $\pi(x,y)$ is the projection of $f(x)$ onto $y + K$. For basic techniques of dealing with projections in Hubert spaces we refer to [Har77] and [Sha88b].

Lemma 2.20 (pseudo-singular cone constraints). *Under the hypotheses (2.84), one has:*

(i) F is not pseudo-regular at (x^0, y^0) if and only if there are points z, y depending on $p \downarrow 0$ such that $(z, y) \to (x^0, y^0)$, $f(z) \neq \pi(z, y)$ and $z \in E_u(p)$, where u is defined as $u = v(\cdot, y^*, y)$ with $y^* = (f(z) - \pi(z, y))|f(z) - \pi(z, y)|^{-1}$.

(ii) If $\dim Y < \infty$ and F is not pseudo-regular at (x^0, y^0), then the functional y^* (as $p \downarrow 0$) have an accumulation point η^*, and $x^0 \in \limsup_{p \downarrow 0} E_{v(\cdot, \eta^*, y)}(p)$.

◇

Note: The inclusion $z \in E_u(p)$ means explicitly that

$$\langle y^*, f(x) - \pi(x, y) - (f(z) - \pi(z, y))\rangle \geq -pd(x, z) \text{ for } x \text{ near } z = z(p).$$

Proof of Lemma 2.20.

(i), ($\Leftarrow$) By Lemma 2.19 (ii), the given points z are Ekeland-points with factor p for $h(x) = |f(x) - \pi(x, y)| = \textbf{dist}\,(y, F(x))$. Thus, F is not pseudo-regular at (x^0, y^0) by Theorem 2.16.

(i), ($\Rightarrow$) Apply first Theorem 2.16, next Lemma 2.19 (i) .

(ii) The existence of η^* is evident. We consider z and y from condition in (i) and assume that $y^* = y^*(p) \to \eta^*$ for a certain sequence $p \downarrow 0$. By (2.34), f is locally Lipschitz. So there is a uniform Lipschitz constant L for the functions $g(\cdot) = f(\cdot) - \pi(\cdot, y)$ near (x^0, y^0). Then, we obtain for small p and small distance $d(x, z)$:

$$\begin{aligned} \langle \eta^*, g(x) - g(z)\rangle &= \langle y^*, g(x) - g(z)\rangle + \langle \eta^* - y^*, g(x) - g(z)\rangle \\ &\geq -pd(x, z) - |\eta^* - y^*| L d(x, z) \\ &= -(p + |\eta^* - y^*| L_f) d(x, z). \end{aligned}$$

This ensures $z \in E_{u(\cdot, \eta^*, y)}(p')$ with $p' = p + |\eta^* - y^*|L$.

Since $p' \downarrow 0$ and $(z, y) \to (x^0, y^0)$, we thus obtain

$$x^0 \in \limsup_{p' \downarrow 0,\ y \to y^0} E_{v(\cdot, \eta^*, y)}(p')$$

with the fixed element $\eta^* \in \text{bd}\, B^*$. □

Lipschitz Operators with Images in Hilbert Spaces

Having Lipschitzian operators with images in a Hilbert space, we are now in the situation of cone constraints with $K = \{0\}$ and $\pi(x, y) = y$. Given $y^* \in B^* := B_{Y^*}$ we put $\phi(x) = \langle y^*, f(x)\rangle$.

Lemma 2.21 (pseudo-singular equations). *Let $f \in C^{0,1}(X, Y)$, let X be a complete metric space and Y be a real Hilbert space, put $F \equiv f$, and consider some point $z^0 = (x^0, f(x^0))$. Then,*

(i) *F is not pseudo-regular at* z^0
if and only if $x^0 \in \limsup_{p\downarrow 0}(\cup_{y^* \in \text{bd}\, B^*} E_\phi(p))$.
(ii) *If* $\dim Y < \infty$, *then F is not pseudo-regular at* z^0
if and only if $x^0 \in \cup_{y^* \in \text{bd}\, B^*} \limsup_{p\downarrow 0} E_\phi(p)$.

◇

Exercise 1. The proof of Lemma 2.21 is left as exercise. ◇

By the Lemma, components of vector-valued functions may be aggregated by nontrivial linear functionals y^*. Let $f: X \to \mathbb{R}^m$ be a locally Lipschitz function, X be a complete metric space, $f(x^0) = y^0$. Then we obtain, due to norm-equivalence in $\mathbb{R}^m$:

f is pseudo-regular at (x^0, y^0)
$\Leftrightarrow \quad \forall \mu \in \mathbb{R}^m \setminus \{0\} : \phi(\cdot) = \langle \mu, f(\cdot) \rangle$ is pseudo-regular at $(x^0, \phi(x^0))$.

Necessary Optimality Conditions

To demonstrate how standard optimality conditions may be directly derived from the equivalences in the last lemmas, we consider only the case of $\dim Y < \infty$ where X is a B-space and all functions belong to C^1. Recall that, for more abstract problems with pseudo-Lipschitz contraint maps, the reduction to upper Lipschitz constraints is possible via Theorem 2.10, whereafter necessary optimality conditions can be derived as for free minimizers of $C^{0,1}$ functions, provided the objective is locally Lipschitz, too; see the part "Optimality Conditions" of Section 2.1.

Equality Constraints

Let x^0 be a (local) solution of the problem

$$\min f_0(x) \quad \text{s.t.} \quad f_k(x) = 0, \; k = 1, \ldots, m.$$

Then, $F(x) = \{y \mid y_k = f_k(x), \; k = 0, 1, \ldots, m\}$ is not pseudo-regular at (x^0, y^0), $y^0 := (f_0(x^0), 0, ..., 0)$. So, by Lemma 2.21 (ii), there is some nontrivial y^* and a sequence $x(p) \to x^0$ such that $x(p)$ is a local Ekeland point with factor $p \downarrow 0$ to the functional $\phi(x) := y_0^* f_0(x) + ... + y_m^* f_m(x)$.

Clearly, then $\|D_x \phi(x(p))\| \le p$ must be satisfied. Thus, via $p \downarrow 0$, it follows the (Fritz-John condition)

$$y_0^* D f_0(x^0) + ... + y_m^* D f_m(x^0) = 0.$$

If the constraint map is pseudo-Lipschitz at $(0, x^0)$, i.e. if $G(x) := \{y | y_k = f_k(x), \; k = 1, ..., m\}$ is pseudo- regular at $(x^0, 0, ..., 0)$, then $y_0^* = 0$ is impossible. Indeed, otherwise the singularity condition (ii) for G is fulfilled. So, division by y_0^* leads us (with new multipliers) to the Lagrange condition

$$D f_0(x^0) + y_1^* D f_0(x^0) + ... + y_m^* D f_m(x^0) = 0.$$

Inequality Constraints

Let x^0 be a (local) solution of the problem

$$\min f_0(x) \;\text{ s.t. }\; f_k(x) \le 0, k = 1, ..., m.$$

Now $F(x) = \{y | y_k \ge f_k(x),\, k = 0,1,\ldots,m\} = f(x) - \mathbb{R}^{1+m}_-$ is not pseudo-regular at (x^0, y^0), $y^0 := (f_0(x^0), 0, ..., 0)$.

By Lemma 2.20 (ii), there is a nontrivial vector $\eta^* = (\eta^*_0, ..., \eta^*_m)$ along with related sequences $z(p) \to x^0$, $y(p) \to y^0$ such that $z(p)$ is a local Ekeland point with factor p to the functional $\phi(x) := \langle \eta^*, f(x) - \pi(x, y(p))\rangle$. Here, $\pi(x, y)$ denotes the Euclidean projection of $f(x)$ onto the cone $y + \mathbb{R}^{1+m}_-$, i.e.,

$$f(x) - \pi(x, y) = ((f_0(x) - y_0)^+, ..., (f_m(x) - y_m)^+).$$

Let us write $y = y(p)$ and $z = z(p)$. Since $f(z) \ne \pi(z, y)$, the set $I^+(p) = \{k | f_k(z) > y_k\}$ is not empty. For some subsequence of $p \downarrow 0$, the finite set $I^+(p) =: J$ is constant. By the construction of ϕ, and because of

$$\eta^* = \lim y^*(p), \text{ where } y^* = (f(z) - \pi(z, y))|f(z) - \pi(z, y)|^{-1},$$

we have:

$$\eta^* \ge 0,\ \eta^* \ne 0,\ \eta^*_k > 0 \text{ only for } k \in J$$
$$\text{and } \phi(x) = \Sigma_{k\in J}\eta *_k (f_k(x) - y_k)^+.$$

For $k \in J$, we further obtain $(f_k(z) - y_k)^+ = f_k(z) - y_k$. Since z is a local Ekeland point of ϕ with factor p, it holds $\|D\phi(z)\| \le p$. This yields, as $p \downarrow 0$, $\Sigma_{k\in J}\eta^*_k Df_k(x^0) = 0$. By definition of J, it holds $f_k(x^0) \ge y^0_k\, \forall k \in J$.

Further, $f(x^0) \le y^0$ is valid since $y^0 \in F(x^0)$. Therefore, the Fritz-John optimality conditions are satisfied. If $\eta^*_0 = 0$ then, due to $\Sigma_{k\in J}\eta^*_k Df_k(x^0) = 0$, now $z(p) = x^0$ is a local Ekeland point with factor $p > 0$ for

$$\phi(x) := \Sigma_{k\in J\setminus\{0\}}\eta^*_k(f_k(x) - y_k(p))^+ \quad\text{where}\quad y_k(p) = y^0_k - p.$$

Thus, again by Lemma 2.20 (i), direction $(\Leftarrow)$, the map $G_J(y) := \{x \,|\, f_k(x) \le y_k \forall k \in J\setminus\{0\}\}$ is not pseudo-Lipschitz at $(0, x^0)$. So, even more, the original constrained map $G(y) := \{x \,|\, f_k(x) \le y_k,\ k = 1, \ldots, m\}$ has the same property. In other words, if G is pseudo-Lipschitz at $(0, x^0)$, now $\eta^*_0 > 0$ yields the Karush-Kuhn-Tucker conditions.

Exercise 2. How can the situation of mixed constraints (equations *and* inequalities) be handled in a similar manner? ◇

Exercise 3. Verify that, for $m > n$, every function $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is nowhere pseudo-regular. Hint: Apply Rademacher's theorem. ◇

2.2.4 Intersection Maps and Extension of MFCQ

It was already noticed in Section 1.4 that Robinson [Rob76c] proved the following: a finite-dimensional system

$$f(x) = y, \; g(x) \leq z \quad \text{with } (f, g) \in C^1(\mathbb{R}^n, \mathbb{R}^{k+m}),$$

has, at $s^0 = (x^0, y^0, z^0)$, a pseudo-Lipschitzian solution map $S(y, z)$ iff the *Mangasarian-Fromovitz constraint qualification* [MF67](MFCQ),

$$\begin{array}{l} Df(x^0) \text{ has full rank, and there is some } u \\ \text{such that } Df(x^0)u = 0 \text{ and } g(x^0) + Dg(x^0)u < z^0, \end{array}$$

is satisfied. The rank condition means that f is pseudo-regular at (x^0, y^0) whereas the *MFCQ direction* u is both a tangent direction of the (regular) manifold $f = y^0$ at x^0 and a "descent direction" for the cone-mapping $G(x) = g(x) + \mathbb{R}^m_+$ at (x^0, z^0).

Let us recall the idea of the sufficiency proof in a non-technical way because we intend to use it under more general settings. To show that $S(y', z')$ contains a point x' close enough to some given $x \in S(y, z)$, move x (relatively) sufficiently far in direction u. Then, the obtained point p fulfills the constraint $g(\cdot) \leq z$ with a big slack and violates the equation $f = y$ only by a little one. Next, using pseudo- regularity of f near $(p, f(p))$, one can replace p by a (close) solution x' of the equation $f = y'$. Due to the big slack $z - g(p)$, the point x' will also satisfy the inequalities $g(x') \leq z'$.

In the C^1**–situation,** the existence and suitable estimates for x' are ensured by the usual implicit function theorem. Now, this tool can be replaced by direct estimates as already done in [Kum00b]. However, the fixed direction u must be exchanged by (discontinuously) moving directions.

Intersection with a Quasi-Lipschitz Multifunction

Let *X*, *Y*, *Z* be normed spaces, and let $F : X \rightrightarrows Y$, $G : X \rightrightarrows Z$ be any multifunctions. The mapping

$$H(x) := \{(y, z) \mid y \in F(x), z \in G(x)\}$$

with

$$S(y, z) = H^{-1}(y, z) = \{x \mid y \in F(x), z \in G(x)\}$$

represents as above the intersection of independent (with respect to y and z) constraints. We ask for the objects playing the role of the MFCQ direction u now. Note that we cannot use fixed u due to our Example BE.2.

The following definitions enable us to deal with the quite general constraint $z \in G(x)$, as in the case of inequalities.

A multifunction $G : X \rightrightarrows Z$ is said to be *quasi-Lipschitz* near (x^0, z^0), if there is some (small) constant $q > 0$ such that, for (x, z) near (x^0, z^0) and all sufficiently small $\delta > 0$, the inclusion $z \in G(x')$ holds true, provided that $z + \delta B \subset G(x)$ and $\|x' - x\| \le q\delta$.

Needless to say that we are not interested in the trivial case of $\operatorname{int} G(x) \equiv \emptyset$.

Examples

If $g \in C^{0,1}(X, \mathbb{R}^m)$ and $G(x) = g(x) + \mathbb{R}^m_+$ then G is everywhere quasi-Lipschitz. Indeed, it holds $z + \delta B \subset G(x) \Leftrightarrow g_i(x) \le z_i - \delta$ (with maximum norm in $\mathbb{R}^m$). Using some Lipschitz constant L_g of g near x^0, this implies $g_i(x') \le z_i$ whenever $\|x' - x\| \le L_g^{-1}\delta$. Here, $q = L_g^{-1}$.

Similarly, G may describe *standard cone constraints,* i.e., $z \in G(x)$ if and only if $g(x) \in z + C$, where $g \in C^{0,1}(X, Z)$, $C \subset Z$ is a convex cone, and $\operatorname{int} C \neq \emptyset$. Now, $z + \delta B \subset G(x) \Leftrightarrow g(x) + \delta B \subset z + C$.

Pseudo-Regular Intersections

To quantify the distance of inner points to the boundary (our "slack"), we adopt an idea due to H. Gfrerer, by defining a function DIST with possibly negative values:

$$\operatorname{DIST}(x, A) := \begin{cases} \operatorname{dist}(x, A) & \text{if } x \in X\backslash \operatorname{int} A \\ -r, \text{ where } r = \sup\{s \mid x + sB \subset A\} & \text{if } x \in \operatorname{int} A \end{cases}$$

Of course, DIST may take the values $-\infty$ and ∞ and is therefore not a *distance function* in the standard sense. Nevertheless, as noticed in [Gfr98], the function DIST turns out to be useful when dealing with optimality conditions.

Convention. In the remainder of this subsection, the symbol $|\cdot|$ is used for the norms in Y and Z, while $\|\cdot\|$ is the norm in X. In product-spaces, we take the max-norm.

Theorem 2.22 (intersection theorem). *Suppose that* X, Y, Z *are normed (real) spaces,* $F : X \rightrightarrows Y$ *and* $G : X \rightrightarrows Z$, $s^0 = (x^0, y^0, z^0) \in \operatorname{gph} H$. *Further, let* G *be quasi-Lipschitz near* (x^0, z^0). *Then, the intersection map* H *is pseudo-regular at* s^0 *if the following conditions hold:*

(i) F *is pseudo-regular at* (x^0, y^0),

and there are elements $u = u(s,t) \in B \subset X$, *defined for* $s = (x, y, z) \in \operatorname{gph} H$ *near* s^0 *and for* t *in some interval* $(0, \tau)$, *such that, uniformly with respect to all sequences* $t \downarrow 0$ *and* $s \to s^0$ *in* gph H,

(ii) $\limsup\, t^{-1} \operatorname{DIST}(z, G(x + tu)) = -\gamma < 0$

(iii) $\limsup\, t^{-1} \operatorname{dist}(y, F(x + tu)) = 0.$

If G *describes standard cone constraints, these conditions are necessary for pseudo-regularity of* H *at* s^0, *and (iii) may be replaced by* $y \in F(x + tu)$.

If G *describes level sets of a locally Lipschitz functional* g, *i.e.,* $G(x) = g(x) + \mathbb{R}_+$, *then one may restrict all* $s = (x, y, z)$ *to* $s' = (x, y, g(x))$ *everywhere, without violating any of the above statements.* ◇

Before proving Theorem 2.22 we quantify the estimates.

Remark 2.23 (estimates). Let G be quasi-Lipschitz near (x^0, z^0) with constant q, let F be pseudo-regular at (x^0, y^0) with rank $L > 0$. Then, in a weaker form, the conditions (ii) and (iii) in Theorem 2.22 may be written as follows: For some $\varepsilon \in (0, \min\{\gamma, \frac{1}{2}qL^{-1}\})$, there are elements $u = u(s,t) \in B \subset X$, denned for $s = (x, y, z) \in \operatorname{gph} H$ near s^0 and for $0 < t < \tau$, such that

(ii)' $\operatorname{DIST}(z, G(x + tu)) < -\varepsilon\, t$ and
(iii)' $\operatorname{dist}(y, F(x + tu)) < \varepsilon^2\, t$.

These are just the properties we shall need in the proof of the sufficiency part, and it will turn out that H satisfies the pseudo-regularity condition with the estimate

$$\|x' - x\| \le 2L(1 + \varepsilon^{-1}q^{-1})|y' - y| + (q + 2\varepsilon^{-1})|z' - z|.$$

◇

Proof of Theorem 2.22.

Sufficiency: We consider small $t > 0$ and points $s := (x, y, z) \in \operatorname{gph} H$ close to s^0. We further agree that u denotes $u(s,t)$. In product-spaces, as mentioned above, we take the max-norm. Because of (ii) and (iii) we find, for any $\varepsilon \in (0, \gamma)$, some $\delta > 0$ such that, whenever $t + d(s, s^0) < \delta$, we have

$$z + \varepsilon t B \subset G(x + tu) \tag{2.36}$$

and

$$d(y, f) \le t\varepsilon^2 \quad \text{for certain} \quad f \in F(x + tu). \tag{2.37}$$

Let $\varepsilon \in (0, \min\{\gamma, \frac{1}{2}qL^{-1}\})$ be fixed. Setting

$$c = L/q$$

we note that $2\varepsilon L < q$ and

$$L(1 + 2\varepsilon c) < L + qc = 2qc = 2L. \tag{2.38}$$

From now on, we consider any points (x, y, z) and (y', z') satisfying

$$(x, y, z) \in \operatorname{gph} H \cap [s^0 + \alpha B] \quad \text{and} \quad (y', z') \in (y, z) + 2\alpha B, \tag{2.39}$$

with some $\alpha \in (0, \delta)$. Here, B is the unit ball in the related (product-) space.

We have to show that, for small α and some constant K, (2.39) ensures the existence of some x' such that

$$(y', z') \in H(x') \quad \text{and} \quad \|x' - x\| \le K(|y' - y| + |z' - z|). \tag{2.40}$$

The appropriate α will be specified during the proof by taking such α which satisfy the subsequent conditions. To find x', we move x first sufficiently far in (the moving) direction u: We put

$$p = x + tu \quad \text{with} \quad t = 2\varepsilon^{-1}(c|y' - y| + |z' - z|).$$

If α is small enough, we may apply (2.36) to see that

$$z' + (|z' - z| + 2c|y' - y|)B \subset z + (2|z' - z| + 2c|y' - y|)B = z + \varepsilon tB \subset G(p).$$

Hence, with

$$\beta = |z' - z| + 2c|y' - y|,$$

we obtain

$$z' + \beta B \subset G(p). \tag{2.41}$$

Applying (2.37), the existence of some $f \in F(p)$ with $|y - f| \le t\varepsilon^2$ is ensured. Next we "solve" $y' \in F(\cdot)$. Since y' is still close to f (after decreasing α once again if necessary) we obtain, due to $f \in F(p)$ and by applying pseudo-regularity of F at (x^0, y^0), the existence of some x' such that

$$y' \in F(x') \quad \text{and} \quad \|x' - p\| \le L|y' - f|.$$

So we may estimate

$$\begin{aligned} \|x' - p\| &\le L(|y' - y| + |y - f|) \\ &\le L(|y' - y| + t\varepsilon^2) \\ &= L(|y' - y| + 2\varepsilon(c|y' - y| + |z' - z|)) \\ &= L((1 + 2\varepsilon c)|y' - y| + 2\varepsilon|z' - z|). \end{aligned}$$

Recalling (2.38) this yields

$$\|x' - p\| \le 2L|y' - y| + 2\varepsilon L|z' - z| \le q(2cy' - y + z' - z) = q\beta.$$

Since G is quasi-Lipschitz we conclude that $z' \in G(x')$. Indeed, this follows now directly from $\|x' - p\| \le q\beta$ and (2.41). Therefore, the obtained point x' satisfies

$$x' \in x + tu + q\beta B \text{ and } (y', z') \in H(x'). \tag{2.42}$$

Finally, taking $\|p - x\| = \|tu\| \le t$ into account, we estimate

$$\|x' - x\| \le \|x' - p\| + \|p - x\| \le 2qc|y' - y| + q|z' - z| + t.$$

By our definition of $t = 2\varepsilon^{-1}(c|y' - y| + |z' - z|)$, the sum on the right-hand side has the form

$$K_1|y' - y| + K_2|z' - z|,$$

where K_1 and K_2 are constants depending on L, q and on the fixed ε only. This gives us the Lipschitz estimate in (2.40) with $K = \max\{K_1, K_2\}$ and verifies the sufficiency of the given conditions.

For an explicit estimate, let us summarize that, if $0 < \varepsilon < \min\{\gamma, \frac{1}{2}qL^{-1}\}$, we have

$$\begin{aligned} \|p-x\| &= t = 2\,\varepsilon^{-1}(Lq^{-1}|y'-y| + |z'-z|), \\ \|x'-p\| &\leq 2L|y'-y| + q|z'-z|, \\ K_1 &= Lq^{-1}(2L + 2\varepsilon^{-1}), \\ K_2 &= q + 2\varepsilon^{-1}. \end{aligned}$$

Our construction thus presented some x' satisfying (2.40) and belonging to the set

$$x + 2\varepsilon^{-1}(Lq^{-1}|y'-y| + |z'-z|)u + (2L|y'-y| + q|z'-z|)B.$$

This yields

$$\|x'-x\| \leq 2L|y'-y|\,(1 + \varepsilon^{-1}q^{-1}) \; + |z'-z|\,(q + 2\varepsilon^{-1})$$

which is the claimed estimate.

Necessity for standard cone constraints: Let $z \in G(x) \Leftrightarrow g(x) \in z + C$.

Obviously, condition (i) is necessary without any assumption. Let $c \in \operatorname{int} C$, $\|c\| = 1$, and fix any $\lambda > 0$ with $c + \lambda B \subset C$. We consider

$$s = (x, y, z) \in \operatorname{gph} H \quad \text{near} \quad s^0 \quad \text{(later we specialize } s\text{)}.$$

Pseudo-regularity of H at s^0 with rank L_H provides us, for small τ and $t \in (0, \tau)$, with x' such that

$$y \in F(x'),\; g(x') - z - tc \in C \text{ and } \|x'-x\| \leq L_H t.$$

Since $tc + \lambda tB \subset C$, we obtain, by adding points of a convex cone,

$$g(x') - z + \lambda tB \subset C,$$

hence

$$\operatorname{DIST}(z, G(x')) \leq -\lambda t.$$

Moreover, due to $y \in F(x')$, we have $\operatorname{dist}(y, F(x')) = 0$.

Writing $x' = x + tu(s,t)$, we thus obtain (ii) and (iii). Indeed, the norm of u is uniformly bounded by L_H, and so we may now replace u and t by $u' = u/L_H \in B$ and $t' = L_H t$, respectively. Then

$$\operatorname{DIST}(z, G(x + t'u')) \leq -L_H^{-1}\lambda.$$

Notice that (iii) is evident because of $y \in F(x + tu)$.

Level sets: Let $z \in G(x) \Leftrightarrow g(x) \leq z \in \mathbb{R}$.

Now we are in the situation of standard cone constraints with $C = R_-$. The theorem becomes trivial if $g(x^0) < z^0$, because one may put $u \equiv 0$. Let $g(x^0) = z^0$.

Necessity. We have

$$s' = (x, y, g(x)) \text{ is near } s^0 \text{ with } s' \in \operatorname{gph} H$$

if and only if

$$(x, y) \text{ is near } (x^0, y^0) \text{ with } (x, y) \in \operatorname{gph} F.$$

Therefore, the conclusions of the necessity part for standard cone constraints are particularly true for the points s'. Due to $z = g(x)$, the above constructed point x' does not depend on z and neither does $u(s', t) = (x' - x)/t$.

Sufficiency. Using (ii) and (iii) for the special points s', the general sufficiency proof provides the estimate (2.40) under the additional hypothesis that $z = z_x := g(x)$. Having other points satisfying (2.39), i.e.

$$(x, y, z) \in \operatorname{gph} H \cap [s^0 + \alpha B] \quad \text{and} \quad (y', z') \in (y, z) + 2\alpha B,$$

it holds $z \geq z_x = g(x)$. Using (2.40) for z_x, there is some x'' related to y' and to $z'' := z_x - |z' - z|$ such that

$$(y', z'') \in H(x'') \text{ and } \|x'' - x\| \leq K(|y' - y| + |z'' - z_x|) = K(|y' - y| + |z' - z|).$$

Since x'' fulfills $y' \in F(x'')$ and

$$g(x'') \leq z_x - |z' - z| \leq z - |z' - z| \leq z',$$

we may put $x' = x''$, in order to satisfy (2.40) with the given z, again. So the remark and all the statements of the theorem have been shown. □

Special Cases

To interpret the conditions of Theorem 2.22, we consider particular cases and suppose that

> G describes level sets of a locally Lipschitz function $g : X \to \mathbb{R}$, i.e., now the intersection is $H(x) = \{(y, z) \,|\, y \in F(x),\ g(x) \leq z\}$.

Corollary 2.24 (intersection with level set). *Let X and Y be normed spaces, $s^0 = (x^0, y^0, z^0) \in \operatorname{gph} H$. Further, let g be locally Lipschitz and $g(x^0) = z^0$. Then, H is pseudo-regular at s^0 if and only if*

(i) F is pseudo-regular at (x^0, y^0)

and there are elements $u = u(x, y, t) \in \operatorname{bd} B \subset X$ which satisfy, uniformly with respect to all $t \downarrow 0$ and $(x, y) \to (x^0, y^0)$ in gph *F,*

(ii) $\limsup t^{-1}(g(x + tu) - g(x)) < 0$ *and*
(iii) $\limsup t^{-1} \operatorname{dist}(y, F(x + tu)) = 0.$ ◇

Proof. In comparison with Theorem 2.22, only $u \in \operatorname{bd} B$ appears as a new topic. To obtain the necessity of this condition, note first that (ii) must be true since $g(x^0) = z^0$ (see the necessity part of the above proof). Then, $u \geq \mu$ holds with some $\mu > 0$ since g is locally Lipschitz. So we may replace u by $u' = u/\|u\|$. □

By (ii) in the foregoing corollary we know that G is pseudo- regular at (x^0, z^0) because there is an inverse family Ψ of selections $\psi_{x,z}$ of the form

$$x' = \psi_{x,z}(z') = x + t\,u(x, y, t); \quad t = |z' - z|,$$

whenever the appearing additional parameter $y \in F(x)$ is sufficiently close to y^0. Condition (iii) says - but only with dependence on t - that $u(x, y, t)$ is an object like a "horizontal-tangent direction" to gph F at (x, y). The inverse families of F are implicitly included in the hypothesis (i). Therefore, we are requiring that F is pseudo- regular and that, in addition, a horizontal tangent direction of gph F is a strict descent direction of g, *provided that we interpret directions as bounded functions.*

Explicitly, the conditions (ii) and (iii) of the corollary may be written as:

$$\begin{array}{l} \exists \delta > 0 \ \forall \varepsilon > 0 \ \exists r > 0 \ \forall t \in (0, r] \ \forall (x, y) \in [(x^0, y^0) + rB] \cap \operatorname{gph} F \\ \exists u \in \operatorname{bd} B : \ g(x + tu) - g(x)) < -\delta t \ \text{ and } \ \operatorname{dist}(y, F(x + tu)) < \varepsilon^2 t. \end{array}$$

Remark 2.23 tells us that we may even require the following (formally weaker) condition:

$$\begin{array}{ll} \textit{(iv)} & \exists \varepsilon_0 \ \forall \varepsilon \in (0, \varepsilon_0) \ \exists r > 0 \ \forall t \in (0, r] \ \forall (x, y) \in [(x^0, y^0) + rB] \cap \operatorname{gph} F \\ & \exists u \in \operatorname{bd} B : \ g(x + tu) - g(x)) < -\varepsilon t \ \text{ and } \ \operatorname{dist}(y, F(x + tu)) < \varepsilon^2 t. \end{array}$$

Let us continue specializing the assumptions of Corollary 2.24.

Case 1:

$F = f : X \to Y$ is a function, continuous at x^0.

Then $y = f(x)$, so $u = u(x, t)$ does not depend on y, and (ii) and (iii) in Corollary 2.24 attain the simpler form

$$\begin{array}{ll} \textit{(v)} & \limsup_{t \downarrow 0, x \to x^0} t^{-1}(g(x + tu) - g(x)) < 0 \\ \textit{(vi)} & \limsup_{t \downarrow 0, x \to x^0} t^{-1} f(x + tu) - f(x) = 0. \end{array}$$

Case 2:

$F = f : X \to Y$ is locally Lipschitz, $\dim X < \infty$.

Now it suffices to consider directions $u = u(x, t)$ in a finite subset of $\operatorname{bd} B_X$ only.

Corollary 2.25 (finite sets of directions). *Let* $X = \mathbb{R}^n$, $H(x) = \{(y, z) \mid y = f(x),\ g(x) \le z\}$, $s^0 = (x^0, y^0, z^0) \in \operatorname{gph} H$. *Let* f *and* g *be a locally Lipschitz and* $g(x^0) = z^0 \in \mathbb{R}$.

Then, H is pseudo-regular at s^0 *if and only if* f *is pseudo-regular at* (x^0, y^0) *and, for each sufficiently small* $\varepsilon > 0$, *there exists a finite subset* $U \subset \operatorname{bd} B_X$ *such that, for all* x *near* x^0 *and all sufficiently small* $t > 0$, *the inequalities*

$$g(x + tu) - g(x) < -\varepsilon t \quad \textit{and} \quad |f(x + tu) - f(x)| < \varepsilon^2 t$$

can be satisfied with some $u = u(x, t) \in U$. ◇

Proof. By Theorem 2.22 and its estimates, the conditions (ii) and (iii) of Corollary 2.24 are equivalent with condition (iv). Therefore, the direction $\Leftarrow$ is evident, we show the other one. Let the requirements (iv) be satisfied for some small $\varepsilon > 0$, and let K be a common Lipschitz rank of f and g near x^0. Then, after replacing $u = u(x,t)$ by some $u' \in u + \alpha B_X$ where $\alpha = \varepsilon^2/K$, we have

$$|t^{-1}(g(x+tu) - g(x)) - t^{-1}(g(x+tu') - g(x))| \leq K\alpha.$$

The same inequality holds for the norm of f. Therefore, because of $K\alpha \leq \varepsilon^2 < \varepsilon$, the elements u' will again satisfy the related conditions of (iv); we only have to take larger $\varepsilon' = 2\varepsilon$. Since $\dim X < \infty$, one may select a finite α-net U of $\operatorname{bd} B_X$, i.e. a finite set $U \subset \operatorname{bd} B_X$ such that $U + \alpha B_X \supset \operatorname{bd} B_X$. So we find new elements $u' = u'(x,t) \in U$ satisfying (iv) for $\varepsilon' = 2\varepsilon$, too. □

Note. Already $g(x) = -|x|$ shows that card $U = 1$ cannot be expected. The cardinality of U may increase while ε is vanishing. But, by the estimates of Theorem 2.22, the conditions of Corollary 2.25 must be satisfied only for some sufficientlysmall ε which has been already estimated in terms of L, K, γ and $q = L(g)^{-1}$.

Case 3:

$F = f$ is locally Lipschitz and $u = u(x,t)$ is fixed.

By definition, condition (v) just says that Clarke's directional derivative

$$g^c(x^0; u) = \limsup_{t \downarrow 0,\, x \to x^0} t^{-1}(g(x+tu) - g(x))$$

is negative.

Case 4:

Let $X = \mathbb{R}^n$, $F = f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$. **Moreover, suppose that**

$$g \text{ has directional derivatives such that } g'(\cdot; u) \text{ is u.s.c. near } x^0. \tag{2.43}$$

We show that these properties are sufficient to put

$$U = \{u^0\}$$

in Corollary 2.25, i.e., while there (in a more general situation) for each ε, $u(x,t)$ could attain finitely many values, now only one direction in $\operatorname{bd} B$ has to be considered.

To prove this, let u^0 be an accumulation point of $u(x^0, t)$ as $t \downarrow 0$. Now (v) and (vi) yield, with some $\varepsilon > 0$,

$$g'(x^0; u^0) < -\varepsilon < 0 \quad \text{and} \quad Df(x^0)u^0 = 0. \tag{2.44}$$

The first inequality follows from the existence of the directional derivatives g' and from the Lipschitz property of $g'(x^0;\cdot)$. Due to $Df(x^0)u^0 = 0$ and $f \in C^1$, condition (vi) is satisfied for $u = u^0$:

$$\limsup_{t\downarrow 0, x\to x^0} t^{-1}|f(x+tu^0) - f(x)| = 0.$$

Since $g'(\cdot;u^0)$ is u.s.c., (2.44) yields that $g'(x+su^0;u^0) < -\varepsilon$ is true for x near x^0 and $s > 0$ small enough. Using the estimate

$$g(x+tu^0) - g(x) \le t \sup_{0<s<t} g'(x+su^0;u^0)$$

which can be easily shown for all directionally differentiable functions $g \in C^{0,1}$, cf. Lemma 6.24, we obtain, for any $t \downarrow 0$ and $x \to x^0$ that

$$\limsup t^{-1}(g(x+tu^0) - g(x)) \le -\varepsilon.$$

Consequently, we have derived (v) and (vi) for fixed $u = u^0$, indeed.

Moreover, we have verified that (v) and (vi) together are *equivalent* with (2.44). Along with the equivalences

$$\text{(i)} \quad \Leftrightarrow \quad Df(x^0) \text{ is surjective} \quad \Leftrightarrow \quad \text{rank } Df(x^0) = m,$$

this leads to the *equivalent pseudo-regularity condition*

$$\text{rank}\, Df(x^0) = m \;\text{ and }\; \exists u^0 : \; Df(x^0)u^0 = 0, \;\; g'(x^0;u^0) < 0. \tag{2.45}$$

Case 5:

Let $X = \mathbb{R}^n$, $F = f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ and g be a max–function of a semi-infinite optimization problem.

For a semi-infinite optimization problem with parametric constraints

$$f(x) = y \;\text{ and }\; h(x,s) \le z \;\forall s \in S,$$

put $g(x) = \max_{s\in S} h(x,s)$. If S is a compact topological space and $h : \mathbb{R}^n \to \mathbb{R}$ is continuous and has continuous derivatives $D_x h(\cdot,\cdot)$, then (2.43) follows from the well-known formula for the directional derivatives

$$g'(x;u) = \max_s \{D_x h(x,s)u \,|\, h(x,s) = g(x)\}$$

which also implies that

$$\partial_c g(x) = \text{conv}\,\{D_x h(x,s) \,|\, h(x,s) = g(x)\}.$$

In this case, the necessary and sufficient condition (2.45) is said to be the *extended Mangasarian-Fromovitz constraint qualification* (*extended MFCQ*), see [HZ82, JTW92, HK94, Kla94b, Sha94, JRS98].

If S is finite, then our circle is closed: Condition (2.45) coincides with MFCQ, see §1.4.

Exercise. Let $M(a,b) = \{x | f(x) = a,\ g(x) \le b\}$ with vector-valued g and f in C^1. Show that $x^0 \in M(0,0)$ satisfies MFCQ already under the formally weaker (but equivalent, because of the C^1-assumption) hypothesis that M is Lipschitz l.s.c. at $((0,0), x^0)$.

Intersections with Hyperfaces

Again, let Y be a normed space, but X has to be a Banach space, now. We further suppose that

$$\begin{aligned}
&F = f : X \to Y \text{ is continuous at } x^0 \text{ and has closed pre-images } f^{-1}(y),\\
&g \in C^{0,1}(X,\mathbb{R}),\ H(x) := \{(y,z) \,|\, y = f(x),\ z = g(x)\},\\
&s^0 = (x^0, y^0, z^0) \in \operatorname{gph} H.
\end{aligned}$$

So, gph H is the intersection of gph f with a (Lipschitz) hyperface.

Theorem 2.26 (intersection with hyperfaces). *Under the above assumptions, H is pseudo-regular at s^0 if the following conditions are satisfied:*

(i) f is pseudo-regular at (x^0, y^0),

and there are uniformly bounded elements $u(x,t)^+$ and $u(x,t)^-$ in X such that

(ii) $\limsup_{t\downarrow 0, x\to x^0} t^{-1}[\, g(x + tu(x,t)^+) - g(x)] < 0,$

$\limsup_{t\downarrow 0, x\to x^0} t^{-1}[-g(x + tu(x,t)^-) + g(x)] < 0,$

(iii) for $u = u(x,t)^+$ and $u = u(x,t)^-$, there holds

$\limsup_{t\downarrow 0, x\to x^0} t^{-1}|f(x + tu) - f(x)| = 0.$

◇

Note: For $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, $g \in C^1(\mathbb{R}^n, \mathbb{R})$, the conditions (i), (ii), (iii) together are nothing else but the full rank condition for the $(n, m+1)$-matrix $(Df(x^0), Dg(x^0))$. Setting $g(x_1, x_2) = |x_1 - x_2|$, one needs directions $u^+ \neq -u^-$, indeed. ◇

Proof of Theorem 2.26. By Corollary 2.24, the mappings

$$H^+(x) = (f(x), g(x) + \mathbb{R}^+)$$

and

$$H^-(x) = (f(x), g(x) + \mathbb{R}^-)$$

are both pseudo-regular at s^0, say with rank $K > 0$. Let L_g be a Lipschitz constant for g near x^0 such that $c = L_g K > 1$. Next consider any points

$$s = (x, y, z) \in \operatorname{gph} H \cap [s^0 + \alpha B],\quad (y', z') \in (y, z) + 2\alpha B, \tag{2.46}$$

where $\alpha > 0$ is small (again fitted during the proof). We have to find some $x^* \in H^{-1}(y', z')$ satisfying a (uniform) Lipschitz estimate

$$\|x^* - x\| \quad \leq L(H)(|y' - y| + |z' - z|). \tag{2.47}$$

Without loss of generality, let $z' < z$. Due to regularity of H^+ (provided that α is sufficiently small) there exists x' such that

$$(y', z') \in H^+(x') \quad \text{and} \quad \|x' - x\| \leq K(|y' - y| + |z' - z|).$$

Put $x^1 = x'$. By definition of H^+ we have $g(x^1) \leq z'$. Clearly, $g(x^1) < z'$ is the essential case, otherwise we already found $x^* = x^1$. In what follows we construct a converging sequence x^k such that $g(x^k)$ tends to z'. By the Lipschitz property of g, we have

$$|g(x^1) - z| \leq L_g\|x' - x\| \leq L_g K(|y' - y| + |z' - z|) = c(|y' - y| + |z' - z|).$$

With $\Theta_1 = (z' - g(x^1))/c$, this yields

$$\Theta_1 \leq (z - g(x^1))/c \leq |y' - y| + |z' - z|. \tag{2.48}$$

For small α, we know (since f is continuous at x^0) that $\Theta_1 = \Theta_1(\alpha)$ and $\|x' - x\|$ are small enough such that all points in $x' + rB$ with

$$r = K\Theta_{1\gamma}/(1 - \gamma) \text{ and } \gamma := 1 - 1/c \tag{2.49}$$

are sufficiently close to x^0, in order to apply pseudo- regularity of H^-. We keep now such α and r fixed.

Then, there is some x^2 such that

$$f(x^2) = y'(= f(x^1)), \; g(x^2) \geq g(x^1) + \Theta_1 \text{ and } \|x^2 - x^1\| \leq K\Theta_1.$$

The Lipschitz property of g yields

$$g(x^2) \leq g(x^1) + L_g K\Theta_1 = g(x^1) + c\Theta_1 = z'.$$

Beginning with $m = 2$ we may now put

$$\Theta_m = (z' - g(x^m))/c$$

in order to find, again by pseudo-regularity of H^-, some x^{m+1} such that

$$f(x^{m+1}) = y', \quad g(x^{m+1}) \geq g(x^m) + \Theta_m$$

and

$$\|x^{m+1} - x^m\| \leq K\Theta_m. \tag{2.50}$$

This yields

$$g(x^{m+1}) \leq g(x^m) + L_g K\Theta_m = z'.$$

Moreover, as long as $\Theta_{m-1} \neq 0$ (the nontrivial case) we have

$$\begin{aligned}
\Theta_m/\Theta_{m-1} &= (z' - g(x^m))/(z' - g(x^{m-1})) \\
&\leq (z' - g(x^{m-1}) - \Theta_{m-1})/(z' - g(x^{m-1})) \\
&= 1 - \Theta_{m-1}/(z' - g(x^{m-1})) \\
&= 1 - 1/c = \gamma < 1.
\end{aligned}$$

So, we have defined a fundamental sequence x^m in $x' + rB$ with r as in (2.49). Since X is complete, $x^* = \lim x^m$ exists in $x' + rB$.

By the closeness of $f^{-1}(y')$ and because of $(z' - g(x^m))/c = \Theta_m \to 0$, it holds $f(x^*) = y'$ and $g(x^*) = z'$. Finally, (2.48), (2.49) and (2.50) yield the regularity estimate with rank

$$L(H) = K(2-\gamma)/(1-\gamma)$$

due to

$$\begin{aligned}
\|x^{m+1} - x\| &\leq \|x^1 - x\| + \|x^2 - x^1\| + \ldots + \|x^{m+1} - x^m\| \\
&\leq K(|y' - y| + |z' - z|) + K\Sigma_{n\geq 1}\Theta_n \\
&\leq K(|y' - y| + |z' - z|) + K\Theta_1/(1-\gamma) \\
&\leq K[1 + 1/(1-\gamma)](|y' - y| + |z' - z|).
\end{aligned}$$

□

Combining Corollary 2.25 and Theorem 2.26 one obtains by induction arguments a characterization of pseudo-regular Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}^m$ in terms of finite sets of normalized directions.

Corollary 2.27 (Lipschitz equations). *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ and $f(x^0) = 0$. Then, f is pseudo-regular at $(x^0, 0)$ if and only if for sufficiently small $\varepsilon > 0$, there are a finite subset $U \subset \operatorname{bd} B$ and some $\delta > 0$ such that, whenever $\|x - x^0\| < \delta$, $0 < t < \delta$ and $k \in \{1, ..., m\}$, the conditions*

$$|f_i(x + tu^+) - f_i(x)| + |f_i(x + tu^-) - f_i(x)| < \varepsilon^2 t \qquad \text{if } i < k$$

as well as

$$f_s(x + tu^+) - f_s(x) < -\varepsilon t \ \text{ and } f_s(x + tu^-) - f_s(x) > \varepsilon t \qquad \text{if } s \geq k$$

can be satisfied by taking certain $u^+ = u^+(x,t) \in U$ and $u^- = u^-(x,t) \in U$. ◇

If f has directional derivatives, then the directions $u^+(x,t)$ and $u^-(x,t)$ (which in general may jump in U as $t \downarrow 0$) can be regarded as being fixed at least for small $t < t(x,k)$.

Exercise 4. How Theorem 2.26 may be extended to the case of a closed multifunction $F : X \rightrightarrows Y$. What about necessity of the conditions in Theorem 2.26 (similar to Theorem 2.22) ? ◇

Chapter 3

Characterizations of Regularity by Derivatives

This chapter is devoted to characterizations of regularity by the help of (generalized) derivatives and may be seen as justification of the derivatives investigated in the current book.

Let $F : X \rightrightarrows Y$ be a multifunction, X, Y be normed spaces, $z^0 = (x^0, y^0) \in \operatorname{gph} F$.

3.1 Strong Regularity and Thibault's Limit Sets

According to the definition, strong regularity is pseudo- regularity along with a uniquely defined (local) inverse function. In what follows, we characterize this property by means of the generalized derivative TF. As in the context of pseudo- regularity, we start with the negation.

Assume that F is *not strongly regular* at z^0. Then, equivalently,

$$\begin{array}{l}\text{there is some sequence } y^k \to y^0 \text{ such that}\\ \operatorname{dist}(x^0, F^{-1}(y^k)) > kd(y^k, y^0)\end{array} \tag{3.1}$$

or

$$\begin{array}{l}\text{there are } (x^k, y^k),\ (\xi^k, \eta^k) \to z^0 \text{ in } \operatorname{gph} F\\ \text{such that } d(\xi^k, x^k) > kd(\eta^k, y^k).\end{array} \tag{3.2}$$

(3.1) says that F^{-1} is not Lipschitz l.s.c. at (y^0, x^0).

(3.2) holds, in particular, if certain pre-images $F^{-1}(y^k)$ are multivalued near x^0. Let us put $t_k = \xi^k - x^k$ and $u^k = (\xi^k - x^k)/t_k$. Now (3.2) just means equivalently the existence of sequences satisfying

$$\begin{array}{l}\eta^k \in F(x^k + t_k u^k),\ t_k \downarrow 0,\ (x^k, y^k) \to z^0 \text{ in } \operatorname{gph} F,\\ \|u^k\| = 1 \text{ and } v := \lim(\eta^k - y^k)/t_k = 0.\end{array} \tag{3.3}$$

For $\dim X < \infty$ the sequence u^k has an accumulation point u, hence, in this case, the origin belongs, for $u \neq 0$, to the set V of all limits

$$\begin{aligned} &v = \lim(\eta^k - y^k)/t_k, \text{ where } \eta^k \in F(x^k + t_k u^k),\ t_k \downarrow 0,\\ &\operatorname{gph} F \ni (x^k, y^k) \to z^0 \text{ and } u^k \to u. \end{aligned} \tag{3.4}$$

This set V is exactly $TF(z^0)(u)$, defined in Section 1.2. In terms of the upper Hausdorff limit, we have

$$\begin{aligned} &TF(z^0)(u) = \limsup t_k^{-1}(F(x^k + t_k u^k) - y^k),\\ &\text{with respect to } t_k \downarrow 0,\ (x^k, y^k) \to z^0 \text{ in } \operatorname{gph} F \text{ and } u^k \to u. \end{aligned}$$

By the construction of TF, we obtain

$$\begin{aligned} &0 \in TF(z^0)(u) \text{ for some } u \neq 0 \ \Rightarrow\ (3.2), \text{ and}\\ &(3.2)\ \Rightarrow\ 0 \in TF(z^0)(u) \text{ for some } u \neq 0, \text{ provided that } X = \mathbb{R}^n. \end{aligned} \tag{3.5}$$

We summarize this first consequence of the definition in a Lemma. As introduced in Section 1.2, we say that $TF(z^0)$ is *injective* if the origin does not belong to $TF(z^0)(u)$ for $u \neq 0$.

Lemma 3.1 (strong regularity for multifunctions). *Let* $F : X \rightrightarrows Y$ *(normed spaces),* $z^0 = (x^0, y^0) \in \operatorname{gph} F$. *Then, injectivity of* $TF(z^0)$ *is necessary for F being strongly regular at* z^0.

If $X = \mathbb{R}^n$ *, then F is strongly regular at* z^0 *if and only if* $TF(z^0)$ *is injective and* F^{-1} *is Lipschitz l.s.c. at* (y^0, x^0). ◇

Proof. Immediately by the above discussion. □

For locally Lipschitz functions F, the *limit sets* $TF(z^0)(u)$ have been studied by Thibault [Thi80] (to construct certain subdifferentials) and were denoted there by $D_F(x^0; u)$. According to §1.2, we call $TF(z^0)(u)$ also the *Thibault derivative* of F at z^0 in direction u. If F is a function, we write $TF(x^0)(u)$ because $y^0 = F(x^0)$ is unique.

For locally Lipschitz *functionals* f, these sets were already considered by F.H. Clarke [Cla76, Cla83], since his directional derivative are

$$f^c(x^0; u) = \limsup_{t \downarrow 0,\, x \to x^0} t^{-1}[f(x + tu) - f(x)] = \sup TF(x^0)(u).$$

Concerning strong regularity of Lipschitz functions, the value of the use of TF and the relations to Clarke's generalized Jacobians [Cla76, Cla83] have been shown in [Kum91b] and becomes clear in Chapter 6 below.

For multifunctions, $TF(z^0)(u)$ was defined in [RW98]. There, the T-operation was applied to F^{-1}, and the necessary condition of strong regularity took the equivalent form $\{0\} = TF^{-1}(y^0, x^0)(0)$.

3.2 Upper Regularity and Contingent Derivatives

By definition (see Section 2.1), F is *upper regular* at $z^0 = (x^0, y^0)$ if there exist $L > 0$ and neighborhoods U and V of x^0 and y^0, respectively, such that

$$\emptyset \neq F^{-1}(y) \cap U \subset x^0 + Ld(y, y^0)B_X \qquad \forall y \in V.$$

This requires (like strong and pseudo-regularity) in particular that F^{-1} is Lipschitz l.s.c. at (y^0, x^0). On the other hand, the local upper Lipschitz condition

$$F^{-1}(y) \cap U \subset x^0 + Ld(y, y^0)B_X \qquad \forall y \in V, \tag{3.6}$$

cannot be satisfied (for each choice of L, U, V) if and only if there are sequences

$$(x^k, y^k) \to z^0 \text{ in gph}\, F \text{ such that } t_k := d(x^k, x^0) > kd(y^k, y^0). \tag{3.7}$$

Now the quotients $d(y^k, y^0)/t_k$ are vanishing. Having an accumulation point u of the bounded sequence $u^k = (x^k - x^0)/t_k$, the latter can be written by means of the contingent derivative CF as

$$0_Y \in CF(z^0)(u) \text{ for some } u \neq 0. \tag{3.8}$$

So we have obtained the following well-known result [KR92].

Lemma 3.2 (upper regularity). *Let $F : X \rightrightarrows Y$ (normed spaces), and let $z^0 = (x^0, y^0) \in \text{gph}\, F$. Then, injectivity of $CF(z^0)$ is necessary for F^{-1} to be locally upper Lipschitz at (y^0, x^0).*

If $X = \mathbb{R}^n$, then it holds

$$CF(z^0) \text{ is injective} \Leftrightarrow F^{-1} \text{ is locally upper Lipschitz at } (y^0, x^0)$$

and

$$F \text{ is upper regular at } z^0 \Leftrightarrow \begin{array}{l} CF(z^0) \text{ is injective and} \\ F^{-1} \text{ is Lipschitz l.s.c. at } (y^0, x^0). \end{array}$$

◇

Proof. Immediately by the above discussion. □

Exercise 5. Show that, in the Lemmas 3.1 and 3.2, one may replace *Lipschitzian l.s.c.* by l.s.c. for F^{-1}. ◇

3.3 Pseudo-Regularity and Generalized Derivatives

To characterize pseudo-regularity, different generalized derivatives have been used in the literature, in particular contingent derivatives (Aubin & Ekeland) and coderivatives (Mordukhovich). These concepts (see Chapter 1 for the definitions) lead us, for closed F and finite dimension, to (primal and dual) criteria for pseudo-regularity. Having infinite dimension, additional assumptions must be imposed for getting equivalent conditions.

Contingent Derivatives

To keep the technical effort small, we will first use the suppositions of Theorem 2.17. So we assume

$$\begin{array}{l} X \text{ is complete, } Y \text{ is normed, } z^0 = (x^0, y^0) \in X \times Y, \\ F : X \rightrightarrows Y \text{ is closed–valued, proper near } y^0 \in F(x^0), \\ \text{and dist}\,(y, F(x)) = d(y, f) \text{ for some } f \in F(x) \text{ whenever } F(x) \neq \emptyset. \end{array} \tag{3.9}$$

Concerning the classical hypotheses (F closed and X, Y Banach spaces), we refer to Theorem 3.4.

Proper Mappings

To apply Theorem 2.17 in the framework of contingent derivatives, let X in (3.9) be a Banach space.

The *contingent derivative* $CF(z)(u)$ (see §1.2), has been successfully applied to describe locally stable behavior of F in [AE84]. In a related concept of tangent cones for sets in normed spaces, the set gph $CF(z)$ coincides with the *contingent cone* of gph$F$ at z, cf. Section 6.1 below. If F is a locally Lipschitz function having directional derivatives, then $CF(x, F(x)) = F'(x; \cdot)$. For chain rules and further properties, see again Section 6 below.

We say that CF is *linearly surjective near* z^0 if, for some L and some neighborhood Ω of z^0,

$$B_Y \subset \cup_{\|u\| \leq L} CF(z)(u) \quad \forall z \in \Omega \cap \text{gph}\, F. \tag{3.10}$$

Corollary 3.3 (pseudo-regularity if CF is linearly surjective 1). *Let (3.9) be true, X be a Banach space and* $z^0 \in \text{gph}\, F$. *Then, F is pseudo-regular at* z^0 *if CF is linearly surjective near* z^0. *For* $X = \mathbb{R}^n$, *the "only if" direction is also true.* ◇

Proof. Let $z \in \Omega \cap \text{gph}\, F$ and $v \in \text{bd}\, B$. Using (3.10) we find $u \in L\mathbb{B}_X$ such that $v \in CF(z)(u)$. So we know, by definition of CF (see Section 1.2), that $y + tv_t \in F(x + tu_t)$ for some sequence $t \downarrow 0$, $u_t \to u$ and $v_t \to v$. Therefore, F is partially invertible near z^0, and Theorem 2.17 guarantees pseudo-regularity.

Conversely, let F be pseudo-regular. We consider the sequence x_t, assigned to z and v by the partial inverse. Now, the bounded sequence $u_t := (x_t - x)/t$ has an accumulation point u since $X = \mathbb{R}^n$. Thus, $v \in CF(z)(u)$ and (3.10) are true. □

Closed Mappings

The assumptions of Theorem 2.17 and Corollary 3.3 are not the "classical" ones: In various papers, one imposes the hypotheses

$$X \text{ and } Y \text{ are Banach spaces and } F : X \rightrightarrows Y \text{ is closed.} \tag{3.11}$$

Theorem 3.4 (basic equivalences, closed maps). *Under (3.11), the equivalences of Theorem 2.17 remain true, i.e. the following statements are equivalent:*

(i) F is pseudo-regular at $z^0 = (x^0, y^0)$.
(ii) F is partially invertible near z^0.
(iii) For some neighborhood Ω of (y^0, x^0), F^{-1} is Lipschitz l.s.c. with uniform rank L at each $(y, x) \in \operatorname{gph} F^{-1} \cap \Omega$.

◇

For proving Theorem 3.4, it suffices to study the original proof of Corollary 3.3 under assumption (3.11) in [AE84] and to note that the hypothesis (3.10) may be replaced by the assumption of F being partially invertible. We repeat first the basic arguments given by Aubin and Ekeland, cf. [AE84, Thm. 4, §7.5].

Theorem 3.5 (pseudo-regularity if *CF* is linearly surjective 2). *Let (3.11) be satisfied and $z^0 = (x^0, y^0) \in \operatorname{gph} F$. Then, F is pseudo-regular at z^0 if there exist L > 0 and a neighborhood Ω of z^0 such that (3.10) holds true.*

Proof. (by contradiction) Given L from condition (3.10), take $r > 0$ such that $4r(L+1) < 1$. Next, using points from (2.27), choose k such that $k > r^{-1}$, set $\varepsilon = d(\eta^k, y^k)$ and introduce the functional ϕ as $\phi(x, f) = rd(x, x^k) + d(f, y^k)$, $(x, f) \in \operatorname{gph} F$. Then $\phi(x^k, \eta^k) = \varepsilon$, so (x^k, η^k) is a **ε-minimizer** of ϕ.

Put $\alpha = \varepsilon r^{-1}$ and apply Ekeland's theorem to ϕ on the complete metric space gph F. Now there is some $z = (p, q) \in \operatorname{gph} F$ such that

$$d((p,q),(x^k,\eta^k)) \le \alpha, \quad \phi(p,q) = rd(p,x^k) + d(q,y^k) \le \phi(x^k,\eta^k) = \varepsilon$$

and, by definition of ϕ,

$$rd(x,x^k) + d(f,y^k) + r[d(x,p) + d(f,q)] \ge rd(p,x^k) + d(q,y^k) \;\; \forall (x,f) \in \operatorname{gph} F. \tag{3.12}$$

Consider $v = y^k - q$.

If $v = 0$ then $y^k = q \in F(p)$ and $\phi(p,q) = rd(p,x^k) \le \varepsilon$. Hence, $d(p,x^k) \le \varepsilon r^{-1}$. By (2.27), we get

$$r^{-1}\varepsilon \ge d(x^k, p) \ge \operatorname{dist}(x^k, F^{-1}(y^k)) > k\varepsilon.$$

Since $k > r^{-1}$, the latter yields a contradiction.

Therefore, we have $v \ne 0$. The point (p,q) depends on k. Increasing k if necessary, we have $(p,q) \in \Omega$. By (3.10), one finds now a direction u and a sequence $t \downarrow 0$ such that special points (x, f_x) belong to gph F:

$$f_x = q + tv + o_Y(t), \;\; x = p + tu + o_X(t), \;\; d(x,p) \le 2t\|u\| \le 2Lt\|v\|. \tag{3.13}$$

This is the crucial inequality for the proof.

With $o(t) = o_Y(t)$ we obtain

$$\begin{aligned} d(f_x, y^k) &= \|q + tv + o(t) - y^k\| \\ &= \|(1-t)(q - y^k) + o(t)\| \\ &\le (1-t)\|v\| + \|o(t)\|. \end{aligned}$$

So (3.12) and (3.13) yield

$$rd(x, x^k) - t\|v\| + \|o(t)\| + r(2Lt\|v\| + t\|v\| + \|o(t)\|) \geq rd(p, x^k).$$

Using here

$$rd(x, x^k) \leq rd(x,p) + rd(p, x^k) \leq 2rLt\|v\| + rd(p, x^k),$$

we have

$$2rLt\|v\| - t\|v\| + \|o(t)\| + r(2Lt\|v\| + t\|v\| + \|o(t)\|) \geq 0.$$

Due to $t\|v\| > 0$, this leads us to

$$4rL + r - 1 + (1+r)\|v\|^{-1}\|o(t)/t\| \geq 0,$$

which cannot hold for small t, because $4r(L+1) < 1$. We arrived at a contradiction, which completes the proof. □

Proof of Theorem 3.4. In the above proof, the direction u in (3.13) only had only to ensure

$$d(x,p) \leq 2Lt\|v\|$$

to hold. The latter, however, is already guaranteed (by definition) if F *is* partially invertible near z^0, and L is the related constant. □

Hence, under (3.11), Corollary 3.3 remains valid, too.

Coderivatives

For normed spaces X and Y and $F : X \rightrightarrows Y$, one may define a pair $(x^*, y^*) \in (X^*, Y^*)$ to be an *ε- normal to* gph F *locally around* $z = (x, y) \in$ gph F, if there is a neighborhood Ω of z such that

$$\langle x^*, x' - x\rangle + \langle y^*, y' - y\rangle \leq \varepsilon d(x', x) + \varepsilon\|y' - y\| \quad \forall (x', y') \in \Omega \cap \operatorname{gph} F. \tag{3.14}$$

This definition corresponds with the approximate *ε-Fréchet normals* in [KM80]. For $\varepsilon = 0$, condition (3.14) yields a usual definition of normals at z to the set $\Omega \cap \operatorname{gph} F$.

The concept of *coderivatives* (see §1.2) requires to put

$$\begin{array}{l} u^* \in D^*F(z^0)(v^*) \text{ if } (u^*, v^*) \text{ is a weak}^* \text{–limit of} \\ \varepsilon\text{–normals } (x^*, y^*) \text{ for some pair of sequences} \\ \varepsilon \downarrow 0,\ (x,y) \to z^0 \text{ in } \operatorname{gph} F. \end{array} \tag{3.15}$$

Remark 3.6 The inequality in (3.14) implies

$$\langle u^*, x' - x\rangle + \langle v^*, y' - y\rangle \leq (\varepsilon + \|x^* - u^*\|)\ \|x' - x\| + (\varepsilon + \|y^* - v^*\|)\ \|y' - y\|.$$

Therefore, one may simplify the definition by setting $x^* = u^*$ and $y^* = v^*$, respectively, provided that the weak* convergence ensures norm-convergence in the related dual space (due to finite dimension or because of particular properties of F). ◇

The injectivity of $D^*F(z^0)$ is an elegant type of a pseudo- regularity condition, elaborated by Kruger and Mordukhovich.

Theorem 3.7 (injectivity of coderivatives and pseudo-regularity). *If* $\dim X + \dim Y < \infty$, *then injectivity of* $D^*F(z^0)$ *is necessary and sufficient for pseudo-regularity of F at* z^0. *If X is an Asplund space and* $\dim Y < \infty$, *then injectivity is still a sufficient condition.* ◇

Proof. The first statement has been shown in [Mor93] (in terms of openness one finds this statement already in [KM80, Mor88]), the second one in [MS97b]. □

We obtain Theorem 3.7, for $\dim X + \dim Y < \infty$, as a consequence of Theorem 3.11, too. For our Example BE.2 which concerns pseudo-regularity, the sufficient conditions of Corollary 3.3 and Theorem 3.7 are not satisfied due to the properties (iii) and (iv) of this example. A Banach space X is an *Asplund space* if every continuous convex function $f : X \to \mathbb{R}$ is Fréchet-differentiable on a dense subset of X, cf. [Asp68].

Vertical Normals

To derive sharper conditions via **ε-normals,** one has to avoid the requirement of weak* convergence for defining D^*F. To weaken the assumptions concerning X, one may modify the definition of **ε-normals** by considering the bilinear form $\langle y^*, y\rangle$ only. If $x^* = 0$ in (3.14), we say that y^* is a ***zero-ε*-normal**. Obviously, **zero-ε-normals** may be defined for metric spaces X by requiring that

$$\langle y^*, y' - y\rangle \le \varepsilon d(x', x) + \varepsilon\|y' - y\| \quad \forall (x', y') \in \Omega \cap \operatorname{gph} F. \tag{3.16}$$

For brevity, let us further say:

F has no *vertical normals near* z^0 if there is some $\varepsilon > 0$ such that, for all $z \in \operatorname{gph} F$ with $d(z, z^0) < \varepsilon$, there is no ***ε-normal*** to gph F locally around z which satisfies $\|x^*\| < \varepsilon$ and $\|y^*\| = 1$ (with $\|\cdot\|$ denoting here the norms in X^*, Y^*).

In this context, X has to be a normed space. If X is a metric space, we say:

F has *no vertical zero-normals near* z^0 if there is some $\varepsilon > 0$ such that, for all $z \in \operatorname{gph} F$ with $d(z, z^0) < \varepsilon$, there is no ***zero-ε-normal*** to gph F locally around z which satisfies $\|y^*\| = 1$.

We intend to show that the simpler **zero-ε-normals** are just the right objects to characterize pseudo-regularity in a dual manner.

Remark 3.8 (equivalence for normed spaces). For normed spaces X and Y, the two conditions of having no vertical (zero-) normals are equivalent. ◇

Proof. Indeed, if F has vertical normals, then the inequalities under remark 3.6 show that F has vertical zero normals, too. Conversely, **zero-ε-normals** are **ε-normals** with $x^* = 0$; this already completes the equivalence. □

In the case of finite dimension, weak* and strong convergence coincide. So one obtains:

Lemma 3.9 (vertical normals 1). *Let* $z^0 \in \operatorname{gph} F$.

(i) If $\dim X < \infty$ *then*
F has no vertical normals near $z^0 \Rightarrow D^*F(z^0)$ *is injective.*

(ii) If $\dim Y < \infty$ *then*
$D^*F(z^0)$ *is injective* $\Rightarrow$ *F has no vertical normals near* z^0.

◇

Next observe that F has never vertical normals if it is pseudo- regular.

Lemma 3.10 (vertical normals 2). *Let* $z^0 \in \operatorname{gph} F$.

(i) If X and Y are normed spaces then
F is pseudo-regular $\Rightarrow$ *F has no vertical normals.*

(ii) Similarly, if X is a metric space and Y is normed, then
F is pseudo-regular $\Rightarrow$ *F has no vertical zero-normals.*

◇

Proof. Let F be pseudo-regular with rank L, and let (x^*, y^*) be an **ε–normal** to $\operatorname{gph} F$ locally around z with $\|y^*\| = 1$. We put $\theta = \frac{1}{2}$ and choose $v \in \operatorname{bd} B_Y$ satisfying $\langle y^*, v\rangle \geq \theta$. Next consider points $z = (x,y) \in \operatorname{gph} F$ with $d(z, z^0) < \varepsilon$.

If ε is small then, for $y' = y + tv$ and small $t > 0$, one finds (by pseudo-regularity) some $x' \in F^{-1}(y')$ such that $d(x', x) \leq Lt$. Let Ω be the neighborhood, related to (x, y) in (3.14). Decreasing t if necessary we have $(x', y') \in \Omega$ and

$$\langle x^*, x' - x\rangle + \langle y^*, tv\rangle \leq \varepsilon d(x', x) + \varepsilon\|tv\|.$$

Dividing by t and using $t^{-1}d(x', x) \leq L$ yields the key inequality

$$\begin{aligned} -L\|x^*\| &\leq t^{-1}\langle x^*, x' - x\rangle \\ &\leq \varepsilon t^{-1} d(x', x) + \varepsilon - \langle y^*, v\rangle \\ &\leq \varepsilon(L+1) - \theta. \end{aligned}$$

(ii) For proving the second statement, we have to regard a **zero-ε–normal**, which is formally $(0, y^*)$ and yields the same formula without the terms including x^*. So we obtain

$$0 \leq \varepsilon(L+1) - \theta,$$

hence ε cannot tend to zero, and F has no vertical zero-normals.

(i) Concerning the first statement we obtain $\|x^*\| \geq L^{-1}(\theta - \varepsilon(L+1))$. The right-hand side tends to $L^{-1}\theta$ as $\varepsilon \downarrow 0$. So $\|x^*\|$ cannot vanish, which yields again the assertion. □

The condition in terms of **zero-ε-normals** is motivated by

Theorem 3.11 (vertical normals and regularity). *Let* $z^0 \in \operatorname{gph} F$.

(i) If X and F satisfy (3.9) and Y is a real Hilbert space, then
F is pseudo-regular at $z^0 \Leftrightarrow F$ *has no vertical zero-normals near* z^0.
(ii) If, in addition, X is a Banach space, then
F is pseudo-regular at $z^0 \Leftrightarrow F$ *has no vertical normals near* z^0.

◇

Proof. ($\Rightarrow$) For both statements, see Lemma 3.10.

($\Leftarrow$) (i) We suppose that F has no vertical zero-normals. The norm in Y is denoted by $|\cdot|$. Assuming that F is not pseudo-regular then, by Theorem 2.16, there is a sequence of related Ekeland points to $p \downarrow 0$. Using the existence of $\eta \in F(z)$ with

$$0 < |\eta - y| = \text{dist}\,(y, F(z)) < 2p$$

and setting $v = \eta - y$, the Ekeland property becomes

$$|f - y| + pd(x, z) \geq |v| \quad \forall (x, f) \in \text{gph}\, F.$$

Taking the square of both sides, one has

$$\langle f - y, f - y\rangle + 2p|f - y|d(x, z) + p^2 d(x, z)^2 \geq \langle v, v\rangle.$$

Writing here $h = f - \eta$ and $f - y = h + v$, one obtains

$$0 \leq 2\langle v, h\rangle + \langle h, h\rangle + 2p|v + h|d(x, z) + p^2 d(x, z)^2.$$

Now restrict $(x, f) \in \text{gph}\, F$ to $\Omega = (z + \varepsilon B_X) \times (\eta + \varepsilon B_Y)$, where $\varepsilon = p|v|$. Then

$$|h| = |f - \eta| \leq p|v|, \langle h, h\rangle \leq p|h||v| \text{ and } |v + h| \leq (1 + p)|v|.$$

Hence, we may proceed with

$$0 \leq 2\langle v, h\rangle + p|h||v| + 2p(1 + p)|v|d(x, z) + p^3|v|d(x, z),$$

i.e.,

$$0 \leq \langle v/|v|, h\rangle + \tfrac{1}{2}p|h| + [p(1 + p) + \tfrac{1}{2}p^3]d(x, z).$$

After re-substituting and setting $y^* = -v/|v|$, (3.16) holds true with $\varepsilon = p(1 + p) + \frac{1}{2}p^3$, namely,

$$\langle y^*, f - \eta\rangle \leq \varepsilon d(x, z) + \varepsilon|f - \eta| \quad \forall (x, f) \in \text{gph}\, F \cap \Omega.$$

Thus F has vertical zero-normals near z^0 which is impossible by the supposition.

(ii) If X is a Banach space, then the vertical zero-normals are vertical normals with $x^* = 0$. □

The equivalence in Theorem 3.11 (ii) has been already shown for Asplund spaces X, Y and closed F, cf. Theorem 3.4 in [MS98]. Due to this result and the remark 3.8, Theorem 3.11 (i) holds for Asplund spaces and closed F, too.

The new topic of Theorem 3.11 consists in the fact that, by applying the simpler **zero-ε- normals,** the geometry of the unit ball in X is completely out of discussion. Concerning assumption (3.9), we once again refer to Lemma 2.13.

Concluding Remarks

The Lemmas 3.1 and 3.2 indicate that, at least in finite dimension, a deeper study of the "derivatives" TF and CF could be valuable, even more since CF is also important with respect to pseudo-regularity. We investigate these derivatives in Chapter 6 below.

Until now, we have established characterizations of stability properties by more or less simple other conditions. However, at this point, we neither know any practicable criteria for checking some of these properties nor the main analytical reason why we should do it. Therefore, we will deal in the next chapter with nonlinear variations of (multi-)functions, with implicit functions as well as with successive approximation.

In this context, the three regularity notions under consideration shall play a key role, indeed.

As technical tools in finite dimension, the derivatives CF and TF will considerably help to analyze particular interesting mappings F. If $\dim X = \infty$, the value of these derivative-concepts is essentially limited by our Example BE.2, while Example BE.5 presents a Lipschitz function with empty contingent derivatives. In consequence, we will work here directly with the definitions.

Chapter 4

Nonlinear Variations and Implicit Functions

In this chapter, we will (at least) suppose that

$$\begin{array}{l} X \text{ is a complete metric space, } Y \text{ is a normed space,} \\ F : X \rightrightarrows Y \text{ is a multifunction, } z^0 = (x^0, y^0) \in \operatorname{gph} F, \\ \text{and the sets } F^{-1}(y) \text{ are closed for } y \text{ near } y^0. \end{array} \tag{4.1}$$

The notions of regularity, introduced so far, concern the solution sets $S(y)$ of the inclusion $g(x) \in F(x)$ where $g(\cdot) \equiv y \in Y$ is a constant function. To study nonlinear variations of F, we consider now inclusions

$$g(x) \in F(x), \ \ x \in \Omega \tag{4.2}$$

and their

$$\text{solution sets } S(g) \subset \Omega,$$

where Ω is some neighborhood of x^0 and $g : \Omega \to Y$ is supposed to be Lipschitz on Ω. We put

$$\begin{aligned} \sup(g, \Omega) &= \sup\{\|g(x)\|_Y \mid x \in \Omega\}, \\ \operatorname{Lip}(g, \Omega) &= \inf\{r > 0 \mid \|g(x) - g(x')\|_Y \leq r d(x, x') \ \forall x, x' \in \Omega\}, \end{aligned}$$

and we equip the space $G = C^{0,1}(\Omega, Y)$ of our variations g with the norm

$$|g| = \max\{\sup(g, \Omega), \operatorname{Lip}(g, \Omega)\}. \tag{4.3}$$

Since (4.2) means that

$$x \in F^{-1}(g(x)), x \in \Omega,$$

we study fixed points of $F^{-1} \circ g$.

The following approach is essentially based on [Kum99]. To modify the regularity definitions for variations in

$$G = C^{0,1}(\Omega, Y),$$

let

$$g^0 \in G \text{ be such that } x^0 \in S(g^0).$$

We call F *pseudo-regular with respect to* G at (x^0, g^0) with rank L if there are neighborhoods U of $x^0, U \subset \Omega$ and V of g^0 such that, given $g, g' \in V(\subset G)$ and $x \in S(g) \cap U$, there is some $x' \in S(g')$ satisfying

$$d(x', x) \le L|g' - g|.$$

The former neighborhood V of y^0 in Y is now a neighborhood of g^0 in G. Notice that $x' \in \Omega$ holds by definition of S via (4.2).

Further, we say that *pseudo-regularity* of F at (x^0, y^0) *is persistent with respect to* $C^{0,1}$*-perturbations,* if there is some neighborhood Ω of x^0 such that F is pseudo-regular with respect to $G = C^{0,1}(\Omega, Y)$ at (x^0, g^0), *provided that* $g^0(x) \equiv y^0$. The related Lipschitz ranks, say L and L_G, as well as the neighborhoods U and U_G of x^0 may differ from each other.

In the same manner, we understand *strong and upper regularity with respect to* G as well as the related persistence.

4.1 Successive Approximation and Persistence of Pseudo-Regularity

In what follows we verify that pseudo- and strong regularity are persistent under small $C^{0,1}$-variations of maps F satisfying (4.1), and we derive estimates for the assigned solutions. Without these estimates, the result of this chapter reads as follows (in fact, it will become a corollary of Theorem 4.3 below).

Theorem 4.1 (persistence under $C^{0,1}$ variations). *Suppose that F satisfies (4.1) and is pseudo- [strongly] regular at (x^0, y^0), $y^0 = 0$, Ω is some neighborhood of x^0 and $g^k \in C^{0,1}(\Omega, Y)$ fulfill $|g^k| < \varepsilon$ $(k = 1, 2)$. Then, if ε is small enough, there exist a second neighborhood $U \subset \Omega$ of x^0 and a constant K such that, to each zero x^1 of $F - g^1$ in U, there is a [unique] zero x^2 of $F - g^2$ in Ω satisfying $\|x^1 - x^2\| \le K\|g^1(x^1) - g^2(x^1)\| \le K \sup(g^1 - g^2, \Omega)$.* ◇

Our concept of *persistence of pseudo-regularity* does not only say that a multifunction $g + F$ is again pseudo-regular (this was shown by Cominetti [Com90]), it even requires that we have to estimate the distance of solutions in terms of $C^{0,1}$-norms.

Persistence of *strong regularity* was originally studied by Robinson [Rob80] for small C^1-functions g. In this context, the mapping $F^{-1} \circ g$ may be directly investigated by applying Banach's fixed point theorem. Having pseudo- regularity, F^{-1} is neither a contraction mapping nor convex-valued. This makes a direct application of known fixed point theorems more difficult. As a basic tool, we construct a solution to $g \in F$ directly by successive approximation.

Supposing (4.1), we want to solve (4.2) by a modification of Banach's fixed point approach, i.e., by selecting $x^{k+1} \in F^{-1}(g(x^k))$ such that $d(x^{k+1}, x^k)$ is "sufficiently small". However, the following algorithm simply generates (more general) elements $x^k \in X$ and $v^k \in Y$, independently of any function $g \in G$. In order to start, we need $(x^0, y^0) \in \operatorname{gph} F$, $v^0 \in Y$ and constants λ, $\beta > 0$. For realizing the initial step, assigned to $k = 0$, we define $v^{-1} = y^0$.

Process (4.4)

Step $k \geq 0$:

$$\begin{array}{l} \textbf{Determine } x^{k+1} \in F^{-1}(v^k) \textbf{ such that} \\ d(x^{k+1}, x^k) \leq \operatorname{dist}(x^k, F^{-1}(v^k)) + \lambda \|v^k - v^{k-1}\|. \\ \textbf{Choose } v^{k+1} \textbf{ such that } \|v^{k+1} - v^k\| \leq \beta\, d(x^{k+1}, x^k). \\ \textbf{Put } k := k+1 \textbf{ and proceed.} \end{array} \tag{4.4}$$

Notes

1. Generally, x^{k+1} may not exist, and the procedure becomes stationary if one selects $v^{k+1} = v^k$.
2. To solve (4.2), put $v^{k+1} = g(x^{k+1})$ which yields $\|v^{k+1} - v^k\| \leq \beta d(x^{k+1}, x^k)$ with $\beta = \operatorname{Lip}(g, \Omega)$ as far as x^{k+1} and x^k belong to Ω.
3. Let $F = \partial f$ be the subdifferential of a convex function f given on $X = \mathbb{R}^n$. Put $g(x) = \beta x$. Then $x \in F^{-1}(g(x^k))$ means $\beta x^k \in \partial f(x)$ and $0 \in -\beta x^k + \partial f(x)$.
   Hence, given x^k, we require $x^{k+1} \in \operatorname{argmin}_{x \in X} f(x) - \beta \langle x^k, x \rangle$. A solution $g(x^*) \in F(x^*)$ now satisfies $x^* \in \operatorname{argmin}_{x \in X} f(x) - \beta \langle x^*, x \rangle$.
4. Put $F(x) = \beta x + \partial f(x)$ and $g(x) = \beta x$. Then $x \in F^{-1}(g(x^k))$ means $g(x^k) \in F(x)$ and $0 \in \beta(x - x^k) + \partial f(x)$.
   Hence, x^{k+1} minimizes $f(x) + \frac{1}{2}\beta \|x - x^k\|^2$, and $g(x^*) \in F(x^*) \Leftrightarrow x^* \in \operatorname{argmin}_{x \in X} f(x)$.
   In this case, the algorithm minimizes f by a proximal point method.
5. To solve $H(x) \cap F(x) \neq \emptyset$ for closed $H : X \rightrightarrows Y$, assume $v^0 \in H(x^0)$ and select $v^{k+1} \in H(x^{k+1})$ with $\|v^{k+1} - v^k\| \leq \beta d(x^{k+1}, x^k)$.
   The latter is possible if H is pseudo-Lipschitz with rank β on Ω.

Before dealing with the convergence of the process (4.4), some comments are appropriate.

The concrete, general formulation of our algorithm may be new, not so the idea of applying successive approximation schemes for showing solvability of (pseudo-Lipschitzian) equations or inclusions. This idea can be found already in [Lyu34, Gra50] as well as in [Com90] and, in a more general form, in [DMO80] and [Sle96].

One can also find extensions of Newton's method which use linearization (in the proper Fréchet sense or a generalized one) $\phi(x) := g(\xi) + Dg(\xi)(x - \xi)$ of the function g at some iteration point ξ, and solve the auxiliary problems $\phi(x) \in F(x)$. Some solution ξ' now replaces ξ in the next step. Methods of this type were studied in [AC95] and [Don96] and were applied to show persistence of solutions like here, based on Kantorovich-type statements [KA64]. Similar approaches are also known from [Kum92, Rob94] and [Kum95a] where the "derivatives", however, had to satisfy conditions which lead to strongly regular and upper regular solutions, respectively.

Here, we intend to show that zeros of pseudo-regular mappings F after small Lipschitzian perturbations g can be determined and estimated via a procedure like Banach's successive approximation: not depending on the linear structure of X and without hypotheses concerning derivatives. So, algorithm (4.4) can be used in the same manner as successive approximation for functions; in particular for deriving implicit-function theorems.

The linear structure of Y will not be used, so Y may be a metric space. Then G is no longer normed; $\sup(g, \Omega)$ and $\mathrm{Lip}\,(g, \Omega)$ must be defined via y^0 and the metric in Y. In this form, the algorithm and Theorem 4.2 apply to multifunctions $F : X \rightrightarrows X$ whereafter the relation to Banach's fixed point theorem is even better visible.

Theorem 4.2 (successive approximation). *Suppose (4.1), let F be pseudo-regular at z^0 with rank L and neighborhoods $U \supset x^0 + B^\circ_X$, $V \supset y^0 + B^\circ_Y$, and assume that*

$$\theta := \beta(L + \lambda) < 1 \text{ and } \alpha := \|v^0 - y^0\| < \delta(1 - \theta)/\max\{1, L + \lambda\}. \tag{4.5}$$

Then,

(i) The process (4.4) generates in U a (geometrically) convergent sequence $x^k \to x^*$ with $d(x^*, x^0) \le (1 - \theta)^{-1}(L + \lambda)\alpha$.

(ii) It holds $v^k \in V$ and $\|v^k - y^0\| \le (1 - \theta)^{-1}\alpha$.

(iii) If $v^k = g(x^k)$ for $k \ge 0$ in (4.4), where g is Lipschitz with rank β on U, then $g(x^*) \in F(x^*)$.

(iv) If one can satisfy $v^k \in H(x^k)$ where $H : X \rightrightarrows Y$ is a closed mapping and Y is complete, then $\lim v^k = v^* \in H(x^*) \cap F(x^*)$.

◇

Note: To simplify we will put in later applications and will require that $\theta = \beta(L + 1) < \frac{1}{2}$ which then will lead to $d(x^*, x^0) \le 2(L + 1)\alpha$ whenever $\beta < \frac{1}{2}(L + 1)^{-1}$ and $\alpha < \frac{1}{2}\delta(L + 1)^{-1}$. ◇

Proof of Theorem 4.2. (i), (ii) Let us first assume that the points under consideration belong to the regions U, V of pseudo-regularity.

Having $x^k \in F^{-1}(v^{k-1})$ (as for $k = 0$) we see that some point $x^{k+1} \in F^{-1}(v^k)$ satisfying

$$d(x^{k+1}, x^k) \leq \text{dist}\,(x^k, F^{-1}(v^k)) + \lambda\|v^k - v^{k-1}\|$$

really exists. Indeed, if $v^k = v^{k-1}$ then we put $x^{k+1} = x^k$, otherwise x^{k+1} exists by the definition of the point-to-set distance. So, by pseudo-regularity of F, we observe that

$$d(x^{k+1}, x^k) \leq (L + \lambda)\|v^k - v^{k-1}\|.$$

If $k = 0$ then

$$d(v^k, v^{k-1}) = \|v^0 - y^0\| = \alpha \text{ and } d(x^1, x^0) \leq (L + \lambda)\alpha.$$

If $k > 0$ then v^k has already been defined according to one or more previous steps, hence

$$\|v^k - v^{k-1}\| \leq \beta d(x^k, x^{k-1})$$

and

$$d(x^{k+1}, x^k) \leq (L + \lambda)\beta d(x^k, x^{k-1}) = \theta d(x^k, x^{k-1}).$$

Moreover, the points v^{k+1} fulfill

$$\|v^{k+1} - v^k\| \leq \beta d(x^{k+1}, x^k) \leq \beta(L + \lambda)\|v^k - v^{k-1}\| = \theta\|v^k - v^{k-1}\|.$$

This ensures

$$\begin{aligned} d(x^{k+1}, x^0) &\leq d(x^1, x^0) + \Sigma_{1\leq i\leq k} d(x^{i+1}, x^i) \\ &\leq (L + \lambda)\alpha + (L + \lambda)\alpha(\theta + ... + \theta^k) \\ &\leq (1 - \theta)^{-1}(L + \lambda)\alpha. \end{aligned}$$

and

$$\|v^{k+1} - y^0\| \leq \alpha + \Sigma_{0\leq i\leq k}\|v^{i+1} - v^i\| \leq (1 - \theta)^{-1}\alpha.$$

Therefore, we generate Cauchy sequences to U and V whenever

$$(1 - \theta)^{-1}\alpha \max\{1, L + \lambda\} < \delta,$$

which is ensured, provided that α and β are small enough, namely if (4.5) holds true. The sequence x^k then converges in the complete space X, and the limit x^* fulfills

$$d(x^*, x^0) \leq (1 - \theta)^{-1}(L + \lambda)\alpha.$$

(iii) To show $g(x^*) \in F(x^*)$ even if Y is not complete, note first that

$$g(x^*) = \lim g(x^k) = \lim v^k \in V.$$

Pseudo-regularity and $x^{k+1} \in F^{-1}(v^k)$ now yield

$$\text{dist}\,(x^{k+1}, F^{-1}(g(x^*))) \leq L\|g(x^*) - v^k\|.$$

Thus, dist $(x^*, F^{-1}(g(x^*))) = 0$ and, since $F^{-1}(g(x^*))$ is closed by (4.1), we have $g(x^*) \in F(x^*)$.

(iv) The existence of $g^* = \lim v^k$ and $g^* \in H(x^*)$ is ensured since H is closed and Y is complete. As under (iii), we obtain the assertion from dist $(x^*, F^{-1}g^*)) = 0$. □

Theorem 4.3 (estimates for variations in $C^{0,1}$). *Suppose* (4.1), *and let* F, δ, L *satisfy the assumptions of Theorem* 4.2. *For some* $r \in (0, \frac{1}{2}\delta)$, *let certain functions* g *and* g' *fulfill on* $U(r) = x^0 + rB_X^\circ$ *the following relations:*

$$\begin{array}{l} g, g' \in C^{0,1}(U(r), Y),\ \mathrm{Lip}\,(g', U(r)) < \frac{1}{2}(L+1)^{-1} \\ \sup_{x \in U(r)} \|g(x) - y^0\| < \frac{r}{8}(L+1)^{-1}, \\ \sup_{x \in U(r)} \|g'(x) - y^0\| < \frac{r}{8}(L+1)^{-1}. \end{array}$$

Then, if $\xi \in x^0 + \frac{1}{2}rB$ *solves* $g(x) \in F(x)$, *there exists a solution* ξ' *to* $g'(x) \in F(x)$ *such that* $d(\xi', \xi) \le 2(L+1)\|g'(\xi) - g(\xi)\| < \frac{1}{2}r$. ◇

Proof. Having $\xi \in x^0 + \frac{1}{2}rB$ such that $g(\xi) \in F(\xi)$, we put $\eta = g(\xi)$ and $\delta' = \frac{1}{2}r$. Then

$$(\xi, \eta) \in \mathrm{gph}\, F, \quad \xi + \delta' B^\circ \subset U(\delta)$$

and, since $\|g(\xi) - y^0\| < \frac{r}{8}(L+1)^{-1} < \frac{r}{8}$, it holds

$$\eta + \delta' B^\circ \subset V(\delta).$$

Thus F is pseudo-regular at (ξ, η) with rank L and neighborhoods $\xi + \delta' B_X^\circ$, $\eta + \delta' B_Y^\circ$.

By our assumptions, it is easy to see that

$$\begin{array}{rcl} a := \|g'(\xi) - \eta\| & \le & \|g'(\xi) - y^0\| + \|y^0 - \eta\| \\ & < & 2\frac{r}{8}(L+1)^{-1} \\ & = & \frac{1}{2}\delta'(L+1)^{-1} \end{array}$$

and

$$b := \mathrm{Lip}\,(g', \xi + \delta' B^\circ) < \frac{1}{2}(L+1)^{-1}.$$

Thus, Theorem 4.2 may be applied to (ξ, η, g', δ') instead of (x^0, y^0, g, δ). Denoting the fixed point x^* by ξ' we observe that both $g(\xi') \in F(\xi')$ and

$$d(\xi', \xi) \le 2(L+1)a = 2(L+1)\|g'(\xi) - g(\xi)\| < \delta'.$$

□

The next direct consequences of Theorem 4.3 extend pseudo- and strong regularity explicitly to small Lipschitzian perturbations.

Corollary 4.4 (pseudo- and strong regularity w.r. to $C^{0,1}$). *Let (4.1) be satisfied and let F be pseudo- [strongly] regular at z^0 with constant L.*

(i) Setting $g^0(\cdot) \equiv y^0$, then F is pseudo- [strongly] regular at (x^0, g^0) with respect to $G = C^{0,1}(U, Y)$ with rank $2(L + 1)$.

(ii) In particular, if $g \in G$ satisfies $\sup(g, x^0 + rB) = o(r)$ and $\mathrm{Lip}\,(g, x^0 + rB) = O(r)$, then F is pseudo- [strongly] regular at z^0 if and only if so is $g + F$. ◇

Proof. Concerning pseudo-regularity, see Theorem 4.3. Concerning strong regularity note that the function $h := F^{-1} \circ g$ is a contraction on U if $|g - g^0|$ is small enough. Thus fixed points x and ξ of h are unique on U due to

$$d(x, \xi) = d(h(x), h(\xi)) \leq \theta d(x, \xi)$$

for some $\theta \in (0, 1)$. □

In a slightly more special form and by direct application of Banach's fixed point theorem, the strong-regularity-version of Corollary 4.4 was a main result of [Rob80] while the second statement of Corollary 4.4 can be found in [Com90].

4.2 Persistence of Upper Regularity

An extension of upper regularity at (X^0, y^0) (X^0 may be a set) to the related upper regularity with respect to G is not possible in the generality of §4.1 in spite of the fact that the upper *Lipschitz estimate* is very simple:

$$g(x) \in F(x) \Rightarrow y \in F(x) \text{ with } y = g(x) \Rightarrow \mathrm{dist}\,(x, X^0) \leq L\|g(x) - y^0\|.$$

The problem arises from the requirement that $g(x) \in F(x)$ should be *solvable.* This cannot be ensured by the weak topological hypothesis of upper regularity at (X^0, y^0), alone. So one needs more structure of the mappings, e.g., convexity in order to apply Kakutani's fixed-point theorem.

Persistence Based on Kakutani's Fixed Point Theorem

The next hypotheses permit the consideration of quite general perturbations of F. The latter is desirable, e.g., if we are interested in minimizing a perturbed convex function f on $\mathbb{R}^n$. Then, $F = \partial f$ is the basic multifunction, but studying $g(x) \in F(x)$ allows us to deal with *differentiable* perturbations φ of f only: $D\varphi(x) = g(x)$. So, if one is interested, e.g., in a homotopy between two *arbitrary* convex functions, $h_t(x) = tf_1(x) + (1 - t)f_2(x)$, the mapping $F_t(\cdot) = \partial h_t = t\partial f_1 + (1 - t)\partial f_2$ becomes the key for analyzing minimizers. But F_t, being a map "near F", is no longer a continuous translation of a multifunction F.

We consider this situation for the case of $X = Y = \mathbb{R}^n$ where the typical ideas already apply. For similarly perturbed inclusions $0 \in F(x)$ in Banach spaces, we refer to [Klu79, Kum84, Kum87].

Theorem 4.5 (persistence of upper regularity). *Let $F, G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be closed and $X^0 \subset F^{-1}(0)$ be non-empty, convex and compact. Let F be upper regular at $(X^0, 0)$ with rank L and neighborhoods $U = X^0 + \beta B^\circ_X$, $V = \beta B^\circ_Y$, let the ranges of G and F^{-1} (on U and V, respectively) be non-empty and convex and, in addition, let $G(x) \cap K \neq \emptyset$ $(\forall x \in U)$ for some bounded set K. Finally, suppose $0 < \delta < \min\{\beta, \beta L^{-1}\}$.*

Then, G has a zero in $X^0 + \delta L B$ if

$$\delta B \cap F(x) \subset G(x) + \delta B \quad \forall x \in X^0 + \delta L B \tag{4.6}$$

is satisfied. ◇

Notes: The additional hypothesis $G(x) \subset F(x) + \delta B$ $\forall x \in U$ ensures the estimate

$$0 \in G(x) \text{ and } x \in U \;\Rightarrow\; y \in F(x) \text{ for some } y \in \delta B \;\Rightarrow\; \operatorname{dist}(x, X^0) \leq L\delta.$$

For *continuous functions* G, F, the suppositions of the theorem hold true, if F is upper regular at $(X^0, 0)$, the sets $F^{-1}(y)$ $(y \in V)$ are convex and non-empty (as for strongly regular F) and $\delta := \sup_{x \in U} \|G(x) - F(x)\|$ is sufficiently small.

Our statement then follows already from [Rob79] where inclusions $g(x) \in F(x)$ have been studied (g is a function, F a multifunction). Robinson's setting allows the direct application of Kakutani's theorem to the map $F^{-1}(g(\cdot))$. Here, we have to prepare this application by partition of unity or by Michael's selection theorem. ◇

Proof of Theorem 4.5. For $y \in \delta B$, upper regularity of F yields

$$\emptyset \neq U \cap F^{-1}(y) \subset X^0 + L\|y\|B \subset X^0 + L\delta B \subset U.$$

Therefore, the convex and compact set $C = X^0 + L\delta B \subset U$ fulfills

$$\emptyset \neq C \cap F^{-1}(y) \subset X^0 + L\|y\|B \text{ if } \|y\| \leq \delta.$$

In what follows, we consider only points $x \in C$. Condition (4.6) ensures that, if $y \in \delta B \cap F(x)$, there is some $y' \in G(x)$ with $y \in y' + \delta B$. Then $\|y'\| \leq 2\delta \leq 2\beta$. Hence, if G satisfies (4.6), so does the map $G_r(x) := rB \cap G(x)$ for $r \geq 2\beta$. Using K we find some $r \geq 2\beta$ such that $G_r(x) \neq \emptyset$ for all x. The mapping G_r is closed, has uniformly bounded, convex images and fulfills (4.6).

To show that $C \cap G_r^{-1}(0) \neq \emptyset$, we assume the contrary,

$$0 \notin G_r(x) \forall x \in C.$$

Since $G_r(x)$ is non-empty, closed and convex, one finds some $z \in \mathbb{R}^n$ such that

$$t(z, x) := \inf\{\langle z, y\rangle \mid y \in G_r(x)\} > 0.$$

Since G_r is closed and uniformly bounded, one sees that $t(z,x) > 0$ implies $t(z,x') > 0$ for x' near x. Therefore, the map $Z(x) := \{z \mid t(z,x) > 0\}$ is l.s.c. and has non-empty convex ranges. By Michael's selection theorem [Mic56], there is a continuous selection $\gamma(\cdot) \in Z(\cdot)$, defined on C. Since $\gamma(x) \neq 0$, we have after normalization,

$$\gamma(x) = 1 \text{ and } \inf\{\langle\gamma(x), y\rangle \mid y \in G_r(x)\} > 0 \; \forall x \in C. \tag{4.7}$$

Now consider the mapping $\Gamma(x) = C \cap F^{-1}(-\delta\gamma(x))$. It fulfills $\Gamma(x) \neq \emptyset$ on C and, due to our hypothesis, $\Gamma : C \rightrightarrows C$ is closed (since γ is continuous) and convex-valued. Thus, by Kakutani's fixed point theorem, there is some $x^* \in C$ with $x^* \in \Gamma(x^*)$. The latter means

$$x^* \in C \cap F^{-1}(-\delta\gamma(x^*)), \quad \text{hence} \quad -\delta\,\gamma(x^*) \in F(x^*).$$

On the other hand, since $x^* \in C$ and $\|-\delta\gamma(x^*)\| = \delta$, $G_r(x^*)$ contains, due to (4.6), some element g which satisfies $\|-\delta\gamma(x^*) - g\| \leq \delta$. This contradicts (4.7) because of

$$\inf\{\langle\gamma(x^*), y\rangle \mid y \in G_r(x^*)\} \leq \langle\gamma(x^*), g\rangle \leq \langle\gamma(x^*), -\delta\gamma(x^*)\rangle + \delta = 0.$$

So $C \cap G_r^{-1}(0) \neq \emptyset$ must be true. □

For making clear the necessity of the imposed assumptions, it is useful to regard the following (Lipschitz continuous) multifunctions

$$F_\delta, F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$$

and sets X^0 where $G = F_\delta$ has no zero near X^0 for $\delta > 0$:

$$\begin{array}{llll} (i) & F_\delta(x) = B \setminus \delta B^\circ & (\delta > 0), \quad F(x) = B, & X^0 = B. \\ (ii) & F_\delta(x) = \operatorname{conv}\{\delta x, x\} & (\delta \geq 0), \quad F = F_0, & X^0 = \operatorname{bd} B. \end{array}$$

Persistence Based on Growth Conditions

Solution sets of optimization problems form a further special class of mappings. Let us point out here a situation which is simple because the feasible set is fixed and is interesting because we get a direct relation to quadratic growth conditions.

In Section 5.1.2 we apply the subsequent theorem to the subdifferential of convex functions. Suppose

$$\begin{array}{l} M \text{ is a non-empty subset of a real Hilbert space } X, \\ f : M \to \mathbb{R}, \text{ and let} \\ \varphi(y) := \inf\{f(x) - \langle y, x\rangle \mid x \in M\} \text{ for } y \in X, \\ \Psi(y) := \operatorname{argmin}_{x \in M}\,[f(x) - \langle y, x\rangle]. \end{array} \tag{4.8}$$

Lemma 4.6 (lsc. and isolated solutions). *Under the assumptions (4.8), the map* Ψ *is l.s.c. at* $(0, x^0) \in \operatorname{gph}\Psi$ *only if* $\Psi(0) = \{x^0\}$. ◇

Proof. Assume (conversely) the existence of

$$\xi \in \Psi(0), \ \|\xi - x^0\| = t > 0.$$

Let $\varepsilon > 0$ and let, according to the l.s.c. assumption,

$$x \in \Psi(\varepsilon u) \text{ for fixed } u = (\xi - x^0)/t.$$

Then, due to optimality of x and ξ for the assigned parameters, we notice that

$$f(x) - \varepsilon\langle u, x\rangle \le f(\xi) - \varepsilon\langle u, \xi\rangle$$

and

$$-f(x) \le -f(\xi).$$

Adding both inequalities yields

$$\langle u, \xi - x\rangle \le 0.$$

But ξ also fulfills, by definition of u,

$$\langle u, \xi - x^0\rangle = t > 0.$$

Since $\|u\| = 1$, the both inequalities imply $\|x - x^0\| \ge t$. Recalling that $x \in \Psi(\varepsilon u)$ was arbitrarily taken, we obtain

$$\text{dist}\,(x^0, \Psi(\varepsilon u)) \ge t \quad \text{for all} \quad \varepsilon > 0.$$

Hence Ψ is not l.s.c. at $(0, x^0)$ whenever $t > 0$. □

As an immediate consequence we obtain

Corollary 4.7 (pseudo-Lipschitz and isolated solutions). *Under the assumptions (4.8), the solution mapping Ψ is pseudo-Lipschitz at $(y^0, x^0) \in \text{gph}\,\Psi$ only if it is single-valued near y^0.* ◇

Proof. Indeed, considering the neighborhoods U, V in Definition (D1) of §1.4, we have $\Psi(y) \cap U \neq \emptyset \ \forall y \in V$, and Ψ must be l.s.c. at all $(y, x) \in \text{gph}\,\Psi \cap V \times U$. So $\Psi(y) = \{x\}$ follows. □

Next we consider f on small neighborhoods of x^0. Given any subset Ω of X, we define

$$\Gamma(y) = \{x \in M \cap \text{cl}\,\Omega \mid f(x) - \langle y, x\rangle \le f(x^0) - \langle y, x^0\rangle\}$$

and

$$\Psi_\Omega(y) = \text{argmin}\,\{f(x) - \langle y, x\rangle \mid x \in M \cap \Omega\}.$$

Theorem 4.8 (growth and upper regularity of minimizers). *Let the assumptions (4.8) be satisfied and $x^0 \in \Psi(0)$. Consider the following statements:*

(i) Ψ *is Lipschitz l.s.c. at* $(0, x^0)$ *with rank L;*
(ii) $\Psi(0) = \{x^0\}$ *and, in addition, there is a neighborhood* Ω *of* x^0 *such that*

$$f(\xi) - f(x^0) \geq (L+1)^{-2}\|\xi - x^0\|^2 \; \forall \xi \in \text{cl}\,\Omega \cap M; \qquad (4.9)$$

(iii) the mappings $\Gamma(\cdot)$, $\Psi(\cdot) \cap \text{cl}\,\Omega$ *and* $\Psi_{\text{cl}\,\Omega}(\cdot)$ *are Lipschitz u.s.c. at* $(0, x^0)$ *with rank* $(L+1)^2$*, and, if* $f \in C(X, \mathbb{R})$ *and* $X = \mathbb{R}^n$*, it holds* $\Psi_{\text{cl}\,\Omega}(y) = \Psi_\Omega(y) \neq \emptyset$ *for small* $\|y\|$.

Then, the following implications are true: (i) $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii).* ◇

Note: Condition (4.9) is called a *quadratic growth* condition (of f at x^0). ◇

Proof of Theorem 4.8. (i) $\Rightarrow$ (ii) Let $\xi \in M$, $t = \|\xi - x^0\| > 0$, $u = (\xi - x^0)/t$ and $\alpha = f(\xi) - f(x^0)$. For small $\varepsilon > 0$, consider any (existing by l.s.c.) $x \in \Psi(\varepsilon u)$ with $\|x - x^0\| \leq L\varepsilon$. Then

$$f(x) - \varepsilon\langle u, x\rangle \leq f(\xi) - \varepsilon\langle u, \xi\rangle = f(x^0) + \alpha - \varepsilon\langle u, \xi\rangle,$$

hence

$$0 \leq f(x) - f(x^0) \leq \alpha - \varepsilon\langle u, \xi - x\rangle.$$

This yields

$$\langle u, \xi - x\rangle \leq \alpha/\varepsilon \quad \text{as well as} \quad \langle u, \xi - x^0\rangle = t, \; \|u\| = 1.$$

So we obtain

$$L\varepsilon \geq \|x - x^0\| \geq \langle u, x - x^0\rangle = \langle u, \xi - x^0 + x - \xi\rangle \geq t - \alpha/\varepsilon$$

and

$$L\varepsilon^2 \geq \varepsilon t - \alpha,$$

as well as, recalling $t = \|\xi - x^0\|$,

$$\varepsilon\|\xi - x^0\| - L\varepsilon^2 \leq \alpha = f(\xi) - f(x^0).$$

In particular, this inequality holds for all $\xi \in M$ with $\|\xi - x^0\| = (L+1)\varepsilon$. Thus,

$$f(\xi) - f(x^0) \geq \varepsilon^2 = \|\xi - x^0\|^2(L+1)^{-2}.$$

(ii) $\Rightarrow$(iii)

Let $f(x) \geq f(x^0) + \delta\|x - x^0\|^2$ whenever $x \in \text{cl}\,\Omega \cap M$, where $\delta > 0$ is fixed. Notice that

$$\Psi(y) \cap \text{cl}\,\Omega \subset \Psi_{\text{cl}\,\Omega}(y) \subset \Gamma(y).$$

Let $x \in \Gamma(y)$. Using

$$f(x^0) - \langle y, x^0\rangle \geq f(x) - \langle y, x\rangle,$$

we observe that

$$-\delta\|x - x^0\|^2 \geq f(x^0) - f(x) \geq \langle y, x^0 - x\rangle \geq \| - y\| \, \|x - x^0\|.$$

Hence $\delta\|x - x^0\| \le \|y\|$ yields the upper Lipschitz properties with rank δ^{-1}. If $\|y\|$ is small enough, we have

$$\Psi_{\mathrm{cl}\,\Omega}(y) \subset \Gamma(y) \subset x^0 + \|y\|\delta^{-1}B \subset \Omega.$$

Thus, it holds $\Psi_{\mathrm{cl}\,\Omega}(y) = \Psi_\Omega(y)$. Finally, if f is continuous and $X = \mathbb{R}^n$, then it suffices to minimize $f(x) - \langle y, x\rangle$ on the compact set $\Gamma(y)$ (which contains x^0), in order to find some element of $\Psi_{\mathrm{cl}\,\Omega}(y)$. □

We notice that under (ii) in Theorem 4,8, the set $\Psi_{\mathrm{cl}\,\Omega}(0)$ does not necessarily cover *all* local minimizers of f with respect to M near x^0, see the discussion of so-called *local minimizing sets*. It seems further worth to mention that the growth condition

$$f(x) - f(x^0) - Df(x^0)(x - x^0) \ge \delta\|x - x^0\|^2$$

is persistent with respect to small $C^{1,1}$ perturbations of f (cf. Corollary 6.21 and formula (6.43)).

4.3 Implicit Functions

Knowing that a certain "nice behavior" of solutions $x(y)$ to $f(x) = y$ can be extended to solutions $x(\phi)$ of equations $f(x) = \phi(x)$ for small functions ϕ in some class G, we have several means for studying solutions $x(t)$ of the equation

$$h(x, t) = 0, \tag{4.10}$$

for t near some critical parameter t^0, say $t^0 = 0$. As one of the simplest, define

$$f(x) = h(x, 0), \; g(x, t) = h(x, 0) - h(x, t).$$

Now equation 4.10 becomes $g(x, t) = f(x)$. It remains to ensure that $\phi := g(\cdot, t)$ becomes sufficiently small in G as $t \to 0$.

For inclusions of the form

$$0 \in h(x, t) + \Gamma(x), \tag{4.11}$$

there is only a formal difference after the same settings because inclusion 4.11 takes the form

$$g(x, t) \in F(x) := f(x) + \Gamma(x).$$

"Nice behavior" must be extended from the inclusion $y \in F(x)$ to $\phi(x) \in F(x)$, $\phi \in G$. As before, $\phi = g(\cdot, t)$ should be small in G provided that t is close to 0.

The "nice behavior" we have in mind is the pseudo-Lipschitz property of solutions depending on y and ϕ , respectively. This has been clarified by Theorem 4.3. The unchanged multi-valued term Γ in (4.11) makes the hypothesis of (strong, pseudo-) regularity for F more or less hard.

However, if this supposition is satisfied, we have to deal with F directly, and the mapping Γ appears as a formal appendix in the well-known context of usual implicit-function theorems only.

Thus, as a technical problem, we have to ensure that

$$g(\cdot,t) = h(\cdot,0) - h(\cdot,t)$$

is indeed sufficiently small (in the sense of Theorem 4.3 for t near 0, as far as x is restricted to a small ball U around x^0.

This is the content of the rest of this chapter where we formulate the consequences of Theorem 4.3 in the current terminology and discuss the needed suppositions for sufficiently smooth h.

We impose the general assumption

$$\begin{array}{l} h : X \times T \to Y,\ h(\cdot,t) \text{ is continuous,} \\ \Gamma : X \rightrightarrows Y \text{ is a closed multifunction,} \\ X \text{ is a Banach space, } Y \text{ and } T \text{ are normed spaces,} \\ \text{and } x^0 \text{ solves (4.11) for } t = 0. \end{array} \tag{4.12}$$

To indicate that in (4.12), one additionally supposes that

Y is a *Banach space*, we write (4.12)'.

As before, we put

$$U(r) = x^0 + rB_X^\circ \quad \text{and} \quad g(x,t) = h(x,0) - h(x,t), \quad x \in U(r).$$

For $\varphi : U \to Y$, recall that Lip (φ, U) denotes the smallest Lipschitz rank, and $\sup(\varphi, U)$ denotes the sup–norm of φ on U.

Theorem 4.9 (estimate of solutions). *Let the assumption (4.12) be fulfilled, and let $F(\cdot) = h(\cdot,0) + \Gamma(\cdot)$ be pseudo-regular at $(x^0,0) \in \operatorname{gph} F$ with rank L and neighborhoods $U(\delta) = x^0 + \delta B_X^\circ$, $V(\delta) = \delta B_Y^\circ$. Suppose that, for some $r_0 \in (0, \frac{1}{2}\delta)$, for all $r \in (0,r_0)$, and for all t in some ball $\tau(r)B_T$, it holds*

$$\begin{array}{l} a(r) := \sup(g(\cdot,t), U(r)) < \frac{r}{8}(L+1)^{-1}, \\ b(r) := \operatorname{Lip}(g(\cdot,t), U(r)) < \frac{1}{2}(L+1)^{-1}. \end{array} \tag{4.13}$$

Then, if $r \in (0,r_0)$ and $s,t \in \tau(r)B_T$, there exists, to each solution $x(t) \in x^0 + \frac{1}{2}rB_X^\circ$ of inclusion $(4.11)_t$ some solution $x(s)$ of $(4.11)_s$ for parameter s such that

$$\|x(s) - x(t)\| \le 2(L+1)\|h(x(t),s) - h(x(t),t)\| \le 4(L+1)a(r).$$

◇

Proof. Solutions of(4.11)$_t$ are solutions to $g(x,t) \in F(x)$. Because of (4.12), the pre-images $F^{-1}(y) = \{x \mid y - h(x,0) \in \Gamma(x)\}$ are closed. In addition, (4.13) ensures the estimate

$$\|h(x(t),s) - h(x(t),t)\| = \|g(x(t),s) - g(x(t),t)\| \leq 2a(r).$$

Since $U(\delta)$ and $V(\delta)$ are related to pseudo- regularity with rank L, we have only to apply Theorem 4.3 with $y^0 = 0$, $g' := g(\cdot,s)$ and $g := g(\cdot,t)$. □

Assumption (4.13) is fulfilled, e.g. for $g(x,t) = \|t\|\omega(x)$ where $\omega : X \to Y$ is locally Lipschitz. Setting $t = 0$ and $x(0) = x^0$ the existence of a solution $x(s)$ is guaranteed if $|g(\cdot,s)|$ is sufficiently small as $C^{0,1}$-norm on $U(r)$.

Replacing "pseudo-regular" by "strongly regular", the solutions are unique in some neighborhood of x^0.

In order to apply Theorem 4.9, there are two essential tasks, namely: *simplify,* if possible, the condition of pseudo-regularity imposed on F, and verify whether $g(\cdot,t) = h(\cdot,0) - h(\cdot,t)$ is small enough in $C^{0,1}(U(r),Y)$ (for small t and r) such that (4.13) can be satisfied.

Practically (since L is usually unknown), we have to ensure that $g(\cdot,t)$ fulfills

$$\begin{aligned} a(r) &:= \sup(g(\cdot,t),U(r)) = o(r) \\ b(r) &:= \mathrm{Lip}\,(g(\cdot,t),U(r)) = O(r) \end{aligned} \tag{4.14}$$

whenever t belongs to a small neighborhood $\tau(r)B_T$ of the origin. To check (4.14), the function $h(\cdot,0)$ may be replaced, in the definition of $g(\cdot,t)$ by any (simpler) continuous function ϕ such that $h(\cdot,0) - \phi$ fulfills

$$\sup(h(\cdot,0) - \phi, U(r)) = o(r) \text{ and } \mathrm{Lip}\,(h(\cdot,0) - \phi, U(r)) = O(r). \tag{4.15}$$

Using a C^1-property of h, condition (4.15) is well-known to hold for the linearization

$$\phi(\xi) = h(x^0,0) + D_x h(x^0,0)(\xi - x^0)$$

of $h(\cdot,0)$ at x^0.

The next lemmas as well as the estimates in the proof of Theorem 4.11 below will summarize these facts and are the key for deriving classical implicit function theorems based on the contraction principle, cf. Zeidler [Zei76]. We need here (4.12)' instead of (4.12) because integrals in Y are needed for proving the used mean-value theorem.

Lemma 4.10 (variations in C^1). *Let (4.12)' be true and* $h(\cdot,0)$ *be continuously Frechet differentiable w.r. to* x. *Then the function* $p(\xi) = h(\xi,0) - \phi(\xi)$ *fulfills*

$$a(r) := \sup_{\xi,\xi' \in U(r)} \|p(\xi) - p(\xi')\|_Y \leq k(r)\|\xi - \xi'\| \quad \textit{with } k(r) \to 0 \textit{ as } r \downarrow 0;$$

and the multifunction

$$F = h(\cdot,0) + \Gamma$$

is pseudo-regular at $(x^0, 0_Y)$ *iff so is* $\Phi = \phi(\cdot) + \Gamma$. ◇

Proof. It holds

$$\begin{array}{rcl} p(\xi) - p(\xi') & = & h(\xi,0) - h(\xi',0) - D_x h(x^0,0)(\xi - \xi') \\ & = & \int_0^1 [D_x h(\theta\xi + (1-\theta)\xi',0) - D_x h(x^0,0)](\xi - \xi')d\theta. \end{array}$$

Thus,

$$\begin{array}{rcl} \|p(\xi) - p(\xi')\| & \leq & \sup_\theta \|D_x h(\theta\xi + (1-\theta)\xi',0) - D_x h(x^0,0)\| \, \|\xi - \xi'\| \\ & = & k(r)\|\xi - \xi'\| \end{array}$$

is true with $k(r) \to 0$ as $r \downarrow 0$ due to the continuity of the derivative. For the *second statement,* now Corollary 4.4 may be applied. □

The following statement can be also shown via the Newton-approaches in [AC95, Don96] or by using such assumptions on h which guarantee that $g(\cdot,t)$ fulfills (4.14) because of the so-called *(strong) B-differentiability,* cf. [Rob91]. In [Com90], condition(4.14) is rewritten as a strict (partial) differentiability condition. There, one finds also a brief and precise characterization of the relations to the *Graves-Lyusternik theorem* [Gra50, Lyu34] and to the (Robinson-Ursescu) *open mapping theorem in* [Rob76a]. By our approach, the statement may be verified, under pseudo-regularity, as an identical copy of the usual implicit-function proof (based on strong regularity).

Theorem 4.11 (the classical parametric form). *Suppose that (4.12)' holds,*
(i) $D_x h(\cdot,\cdot)$ *exists and is continuous near* $(x^0,0)$, *and*
(ii) $h(x^0,\cdot)$ *is continuous at* 0.
Moreover, let $F(\cdot) = h(\cdot,0) + \Gamma(\cdot)$ *be pseudo-regular at* $(x^0,0) \in \operatorname{gph} F$ *with constant L. Then, for sufficiently small* $r > 0$ *there exists* $\tau(r) > 0$ *such that, whenever* $s,t \in \tau(r)B_T$ *and* $x(t) \in x^0 + \frac{1}{2}rB_X^\circ$ *solves* $(4.11)_t$ *, there is a solution* $x(s)$ *of* $(4.11)_s$ *satisfying*

$$\|x(s) - x(t)\| \leq 2(L+1)\|h(x(t),s) - h(x(t),t)\| \leq \tfrac{1}{2}r.$$

◇

Proof. We have to estimate the quantities $a(r)$ and $b(r)$ of Theorem 4.9. For small $\delta > 0$, $\xi, x \in U(\delta)$, $t \in \delta B_T$ and $\xi \neq x$, the mean-value theorem implies that

$$\begin{array}{rcl} k(\delta,t) & := & \|g(\xi,t) - g(x,t)\| / \|\xi - x\| \\ & = & \|h(\xi,0) - h(\xi,t) - (h(x,0) - h(x,t))\| \,/\, \|\xi - x\| \\ & = & \|h(\xi,0) - h(x,0) - (h(\xi,t) - h(x,t))\| \,/\, \|\xi - x\| \\ & \leq & \sup\limits_{x' \in U(\delta)} \|D_x h(x',0) - D_x h(x',t)\|. \end{array}$$

Continuity of $D_x h(\cdot,\cdot)$ along with $t \in \delta B_T$ yields $k(\delta,t) \leq O(\delta)$. Thus, for sufficiently small r and for $t \in rB_T$, it holds

$$b(r) := \operatorname{Lip}(g(\cdot,t),U(r)) \leq O(r) < \tfrac{1}{2}(L+1)^{-1}.$$

To estimate $a := \sup_{x\in U(r)} \|g(x,t)\|$ one may write

$$\begin{aligned}\|g(x,t)\| &= \|g(x^0,t) + (g(x,t) - g(x^0,t))\| \\ &\le \|g(x^0,t)\| + b(r)\|x - x^0\| \\ &\le \|h(x^0,0) - h(x^0,t)\| + O(r)r.\end{aligned}$$

With small r, such that $o(r) := O(r)r < \frac{r}{16}(L+1)^{-1}$, and small $\tau(r) > 0$ such that

$$\tau(r) < r \text{ and } \|h(x^0,0) - h(x^0,t)\| < o(r)\ \forall t \in \tau(r)B_T,$$

this inequality yields $a < 2o(r) < \frac{r}{8}(L+1)^{-1}$. □

Let us add an approximation, applied in §6.6.2, where the existence of $D_x h$ is not supposed. For this reason, we assume that also t varies in a Banach space and we write, for continuous h,

$$h(x,t) = h(x,0) + D_t h(x^0,0)t + \alpha(x,t).$$

provided that the partial derivative $D_t h$ exists on some convex neighborhood of $(x^0,0)$ which contains (x,t).

Lemma 4.12 (linearization w.r. to parameters). *Let $x \in x^0 + rB$ and $t \in r\varrho B$ where $\varrho > 0$ is a fixed constant. If $D_t h$ exists and is continuous near $(x^0,0)$ then it holds*

$$\|t\|^{-1}\|\alpha(x,t)\| \to 0 \quad as \quad x \to x^0 \ and\ \|t\| \downarrow 0$$

as well as $\sup\|\alpha(x,t)\| = o(r)$. *If $D_t h$ is even Lipschitz near $(x^0,0)$, then* $\operatorname{Lip}(\alpha(\cdot,t), x^0 + rB) = O(r)$. ◇

Proof. By the mean-value theorem, one obtains

$$\alpha(x,t) = h(x,t) - h(x,0) - D_t h(x^0,0)t = \int_0^1 [D_t h(x,\theta t) - D_t h(x^0,0)]\, t\, d\theta$$

which yields the first assertion immediately. The second one follows from

$$\sup(\alpha(\cdot,t), x^0 + rB) \le \varrho r O(r) = o(r).$$

If $D_t h$ is Lipschitz near $(x^0,0)$ then it even holds, with some constant K,

$$\begin{aligned}\|\,\alpha(x',t) - \alpha(x,t)\,\| &= \|\int_0^1 [D_t h(x',\theta t) - D_t h(x,\theta t)]\, t\, d\theta\,\| \\ &\le K\|t\|\,\|x' - x\|\end{aligned}$$

for x, x' near x^0 and t near the origin. □

In the present case, solutions of

$$0 \in h(x,0) + D_t h(x^0,0)t + \Gamma(x) \quad \text{and} \quad 0 \in h(x,t) + \Gamma(x)$$

can be compared: With

$$S = [h(\cdot,0) + \Gamma(\cdot)]^{-1},$$

the inclusions become

$$x \in S(-D_t h(x^0,0)t) \quad \text{and} \quad x \in S(-D_t h(x^0,0)\, t - \alpha(x,t)),$$

respectively.

Concluding Remarks

1. If F *is* pseudo-regular then the possibly multi-valued solution map $g \to X(g)$ defined by the solutions of $g(x) \in F(x)$ behaves even pseudo-Lipschitz with respect to $g \in G = C^{0,1}$. In addition, solution estimates in terms of $\sup(g''-g', U)$ can be derived from Theorem 4.3 and lead to implicit-function statement for systems $h(x,t) \in F(x)$.

2. The estimates (4.13) in Theorem 4.9 do not only hold for small smooth perturbations of a smooth original function $h(\cdot,0)$. Other simple possibilities are linear homotopies $h(x,t) = (1-t)p(x) + tq(x)$ where p and q are locally Lipschitz, cf. §11.1.

3. All the solutions, whose existence was claimed, can be determined (theoretically) by using successive approximation, see §4.1.

Chapter 5

Closed Mappings in Finite Dimension

In this chapter, we regard only closed multifunctions $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, suppose $y^0 \in F(x^0)$, and investigate, in particular, regularity of locally Lipschitz functions.

5.1 Closed Multifunctions in Finite Dimension

In finite dimension, the regularity conditions derived up to now may be simplified and allow additional conclusions that are not true, in general.

5.1.1 Summary of Regularity Conditions via Derivatives

Theorem 5.1 (regularity of multifunctions, summary).
Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be dosed and $z^0 = (x^0, y^0) \in \operatorname{gph} F$. Then:

$$\begin{aligned} &F \text{ is upper regular at } z^0 \\ &\Leftrightarrow\ CF(z^0) \text{ is injective and } F^{-1} \text{ is l.s.c. at } (y^0, x^0). \end{aligned} \tag{5.1}$$

$$\begin{aligned} &F \text{ is strongly regular at } z^0 \\ &\Leftrightarrow\ TF(z^0) \text{ is injective and } F^{-1} \text{ is l.s.c. at } (y^0, x^0). \end{aligned} \tag{5.2}$$

$$\begin{aligned} &F \text{ is pseudo-regular at } z^0 \\ &\Leftrightarrow\ \exists \varepsilon > 0 : \varepsilon B \subset CF(z)(B) \text{ for all } z \in \operatorname{gph} F \cap (z^0 + \varepsilon B) \\ &\Leftrightarrow\ D^*F(z^0) \text{ is injective.} \end{aligned} \tag{5.3}$$

$$\begin{aligned} &\text{If } F^{-1} \text{ is Lipschitz l.s.c. at } (y^0, x^0) \\ &\text{then there exists } r > 0 \text{ such that } B \subset CF(z^0)(rB). \end{aligned} \tag{5.4}$$

If X is a normed space, the conditions (5.1) and (5.2) are still necessary for the related regularity. ◇

Proof.
For (5.1) see Lemma 3.2 and the related exercise.
For (5.2) see Lemma 3.1 and the related exercise.
For (5.3) see Corollary 3.3 and Theorem 3.7 (§3.3).
For (5.4) see the arguments of the proof to Corollary 3.3 for $\dim X < \infty$.
For normed *X*, the necessity of the conditions (5.1) and (5.2) follows from the same lemmas. □

The (pointwise) condition (5.4) is indeed not sufficient for F^{-1} to be Lipschitz *l.s.c.* at (y^0, x^0), cf. Exercise 9 below.

The l.s.c. condition under (5.1) and (5.2) is, in general, not ensured by the already imposed injectivity of *CF* and *TF*, respectively.

On the other hand, the condition (5.3) can be replaced by a formally weaker surjectivity condition. To show this, we establish first a relation between *CF* and *D*F* without considering limits of the dual elements.

Theorem 5.2 (*CF* and *D*F*). *Under the assumptions of Theorem 5.1, it holds* $u^* \in D^*F(x^0, y^0)(v^*)$ *if and only if*
$\liminf_{(x,y)\to(x^0,y^0),\ y\in F(x)} \ \sup_{(u,v)\in B,\ v\in CF(x,y)(u)} \left(\langle u^*, u\rangle + \langle v^*, v\rangle\right) \le 0.$ ◇

Proof. By Remark 3.6, we have $u^* \in D^*F(x^0, y^0)(v^*)$ iff there are $(x, y) \to (x^0, y^0)$ in gph F and $\varepsilon \downarrow 0$, $r \downarrow 0$ such that

$$\begin{array}{l}\langle u^*, x' - x\rangle + \langle v^*, y' - y\rangle \le \varepsilon(\|x' - x\| + \|y' - y\|) \\ \text{if } \|x' - x\| + \|y' - y\| < r \text{ and } (x', y') \in \operatorname{gph} F\end{array} \tag{5.5}$$

Writing here $x' - x = tu_t$ and $y' - y = tv_t$ with $(u_t, v_t) \in B$ (the sum-norm ball in $\mathbb{R}^{n+m}$) this becomes

$$\begin{array}{l}\langle u^*, u_t\rangle + \langle v^*, v_t\rangle \le \varepsilon(\|u_t\| + \|v_t\|) \\ \text{whenever t is small enough and } (x + tu_t, y + tv_t) \in \operatorname{gph} F\end{array} \tag{5.6}$$

Since every sequence of $(u_t, v_t) \in B$ (as $t = t_k \downarrow 0$) possesses a convergent subsequence, and the accumulation points (u, v) form just gph $CF(x, y)$, one easily sees that (5.6) yields both directions of the assertion. □

The next statement can be shown by using the both pseudo-regularity conditions of Theorem 5.1 and Theorem 5.2 as well. Our proof applies the Ekeland points of Theorem 2.16 and *CF* only.

Theorem 5.3 (conv*CF*). *Under the assumptions of Theorem 5.1,*

(i) the mapping F is pseudo-regular at (x^0, y^0) *if and only if*

$$\exists \varepsilon > 0: \ \varepsilon B \subset \operatorname{conv}[CF(x, y)(B)] \ \forall (x, y) \in \operatorname{gph} F \cap ((x^0, y^0) + \varepsilon B). \tag{5.7}$$

(ii) F is not pseudo-regular at (x^0, y^0) *if and only if*

$$\begin{array}{l}\textit{there exist } y^* \in \mathbb{R}^n \setminus \{0\} \textit{ and } (x^k, y^k) \in \operatorname{gph} F \\ \textit{such that } (x^k, y^k) \to (x^0, y^0) \textit{ as } k \to \infty \textit{ and} \\ \limsup_{k\to\infty} \ \sup_{\zeta \in CF(x^k, y^k)(B)} \langle y^*, \zeta\rangle \le 0.\end{array} \tag{5.8}$$

◇

Note. Evidently, condition (5.8) is equivalent to the (also necessary and sufficient) coderivative condition $0 \in D^*F(x^0, y^0)(y^*)$, but (5.8) requires the study of contingent derivatives, multiplied with fixed y^*, only. In this way, the present theorem indicates that the condition $0 \in D^*F(x^0, y^0)(y^*)$ (in finite dimension) is nothing but a limit-condition for contingent derivatives. ◇

Proof of Theorem 5.3. (i) By Theorem 5.1, only the sufficiency of the condition must be shown. We apply Theorem 2.16. Assume that F is not pseudo-regular at (x^0, y^0) though ε exists in the given way. Then one finds, for each $p > 0$, certain $z \in \operatorname{dom} F \cap (x^0 + pB_X)$ and $y \in y^0 + pB_Y$ such that both $0 < \operatorname{dist}(y, F(z)) < 2p$ and $z \in X$ is a global Ekeland-point of $\operatorname{dist}(y, F(\cdot))$ with factor p. Recall that the latter means

$$\operatorname{dist}(y, F(x)) + p\, d(x, z) \;\geq\; \operatorname{dist}(y, F(z)) \;\; \forall x. \tag{5.9}$$

Let p be fixed such that $p + d(y, y^0) + \|z - x^0\| < \varepsilon/3$. Since $F(z) \subset \mathbb{R}^m$ is nonempty and closed, some $\eta \in F(z)$ realizes the distance $d(y, \eta) = \operatorname{dist}(y, F(z))$. Setting $v^0 = (y - \eta)/d(y, \eta)$ and applying condition (5.7) to $\varepsilon v^0 \in \varepsilon B$, there exist $u \in B$ and $v \in CF(z, \eta)$ satisfying

$$\varepsilon = \langle v^0, \varepsilon v^0 \rangle \leq \langle v^0, v \rangle.$$

Next consider any $t = t_k \downarrow 0$ such that assigned points $x_t = z + tu_t,\;\; y_t = \eta + tv_t$ satisfy $u_t \to u,\;\; v_t \to v$ and $y_t \in F(x_t)$. One easily determines, by using the Euclidean norm and $\sqrt{a + w} = \sqrt{a} + \frac{1}{2}\frac{w}{\sqrt{a}} + o(w)$ for $a > 0$ and small $|w|$ that

$$\begin{aligned}
&\|y - y_t\| = \|y - \eta - tv_t\| \\
&= \|y - \eta\| - t\|y - \eta\|^{-1}\langle y - \eta, v_t \rangle + o(t) \\
&= \|y - \eta\| - t\langle v^0, v_t \rangle + o(t) \leq \|y - \eta\| - \varepsilon t/2.
\end{aligned}$$

From (5.9), we thus obtain

$$\|y - \eta\| - \varepsilon t/2 + pt\|u_t\| \geq \operatorname{dist}(y, F(x_t)) + pd(x_t, z) \geq \|y - \eta\|$$

and $-\varepsilon t/2 + pt\|u_t\| \geq 0$ as well as $p \geq \varepsilon/2$. Since $p \downarrow 0$ is impossible, the assertion now follows from Theorem 2.16.

(ii) This statement can be easily derived by negation of (5.7) along with the separation theorem. □

Recall that, for locally Lipschitz functions, Corollary 2.27 gave another criterion without using derivatives.

Exercise 6. Show that, for $f \in C^1$, one obtains $D^*f(x) = -Df(x)^{\mathsf{T}}$. ◇

Note: To get here $D^*f(x) = Df(x)^{\mathsf{T}}$, Mordukhovich [Mor88] defined the coderivative with the opposite sign.

5.1.2 Regularity of the Convex Subdifferential

Here, we discuss the content of the regularity definitions for the subdifferential of a convex function on $\mathbb{R}^n$ by showing how they are related to growth conditions. Let f be a convex functional on $\mathbb{R}^n$ and $\partial f(x)$ be its subdifferential at x,

$$\partial f(x) = \{y \in \mathbb{R}^n \mid f(\xi) - f(x) \geq \langle y, \xi - x\rangle \ \forall \xi \in \mathbb{R}^n\}.$$

Then, $x \in (\partial f)^{-1}(y)$ means that x is a solution of

$$\min\{f(\xi) - \langle y, \xi\rangle \mid \xi \in \mathbb{R}^n\},$$

and the related infimum value is just the value of the (concave) *conjugate function* f^* at y. Let us put

$$h_{(y)}(x) = f(x) - \langle y, x\rangle.$$

The inverse $(\partial f)^{-1}(y)$ is the mapping $\Psi(y)$ of §4.2, now with $C = \mathbb{R}^n$. *Strong regularity* of $F = \partial f$ at $(x^0, 0) \in \operatorname{gph} \partial f$ simply means that $h_{(y)}(\cdot)$ has unique and locally Lipschitz minimizers $x(\cdot)$ for y near the origin.

Next we demonstrate that our three regularity properties can be completely characterized by growth at a point x^0 or by (uniform) growth on a neighborhood, respectively. For this reason we introduce the following conditions:

(CG.1) $\exists\, \varepsilon > 0$ such that $\|g' - g''\| \geq \varepsilon\|x' - x''\|$ whenever $x', x'' \in x^0 + \varepsilon B, g', g'' \in \varepsilon B$ and $g' \in \partial f(x'), g'' \in \partial f(x'')$,

(CG.2) $\exists\, \varepsilon > 0$ such that, for all $y \in \varepsilon B$, a minimizer x of $h_{(y)}(\cdot)$ exists and satisfies $h_{(y)}(x+u) \geq h_{(y)}(x) + \varepsilon\|u\|^2 \ \forall u \in \varepsilon B$.

(CG.3) $\exists\, \varepsilon > 0$ such that $f(x^0 + u) \geq f(x^0) + \varepsilon\|u\|^2 \quad \forall u \in \varepsilon B$.

Taking into account that $T\partial f(x^0, 0)$ is a closed and positively homogeneous map, condition (CG.1) is nothing else but *injectivity* of $T\partial f(x^0, 0)$.

Condition (CG.2) has been used in [LS97], in a pointwise context. Here, (CG.2) is an uniform growth condition for all $h_{(y)}$ near the minimizers $x(y)$. For $f \in C^2$, this condition simply means that $D^2 f(x^0)$ is positively definite.

Condition (CG.3) is the *standard quadratic growth condition* with respect to f at some point x^0.

Theorem 5.4 (regularity of the convex subdifferential). *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be convex, and let* x^0 *be a point such that* $0 \in \partial f(x^0)$. *Then*

∂f *is pseudo-regular at* $(x^0, 0)$
$\Leftrightarrow$ ∂f *is strongly regular at* $(x^0, 0)$
$\Leftrightarrow$ *(CG.1)*
$\Leftrightarrow$ *(CG.2),*

and

∂f *is upper regular at* $(x^0, 0) \Leftrightarrow$ *(CG.3).* ◇

Notes:
(i) Compared with Theorem 5.1, the l.s.c.-condition of (5.2) does not appear.
(ii) The local monotonicity condition (CG.1) holds true if f is *strongly convex*; but the reverse is not true; take $f(x) = |x|$. This shows that the local situation differs from the global one, i.e., if $(\partial f)^{-1}$ is *globally* unique and Lipschitz, then f is strongly convex, cf. [LS96, Thm. 2.2].
(iii) For $f \in C^{1,1}$, the equivalence "strong regularity $\Leftrightarrow$ (CG.1)" will also follow from Theorem 5.14 below. ◇

Proof of Theorem 5.4. The claimed equivalences essentially follow from previously proved regularity conditions.

pseudo-regular $\Leftrightarrow$ *strongly regular* follows from Corollary 4.7.

strongly regular $\Rightarrow$*(CG.1):* see Theorem 5.1.

(CG.1) $\Rightarrow$*strongly regular:* By Theorem 5.1, it is only to show that $(\partial f)^{-1}$ is l.s.c. at $(0, x^0)$. Setting $x'' = x^0$ and $g'' = 0$ in (CG.1), one observes

$$\|g'\| \geq \varepsilon \|x' - x^0\| \quad \text{if } x' \in x^0 + \varepsilon B, g' \in \partial f(x') \text{ and } \|g'\| \leq \varepsilon.$$

Taking $x' \in \operatorname{argmin} f, x'$ near x^0 and $g' = 0$, it follows that x^0 is isolated in $(\partial f)^{-1}(0)$, hence, in particular, the set $\operatorname{argmin} f = (\partial f)^{-1}(0)$ is non-empty and bounded. So, by Theorem 1.15, the solution sets $(\partial f)^{-1}(y) = \operatorname{argmin} h_{(y)}$ are nonempty for sufficiently small $\|y\|$, and $(\partial f)^{-1}$ is u.s.c. at 0. Since x^0 is isolated in $(\partial f)^{-1}(0)$, this u.s.c. mapping is l.s.c. at $(0, x^0)$, too.

(CG.2) $\Rightarrow$ *strongly regular.* Due to (CG.2) and Theorem 4.8, $\Psi = (\partial f)^{-1}$ is locally Lipschitz u.s.c. with uniform rank ε^{-1} at the unique minimizers $x = x(y)$ for y near 0. By uniqueness of the solutions, it is uniformly Lipschitz l.s.c., too. But this is pseudo-regularity due to Theorem 2.17 and strong regularity due to uniqueness.

strongly regular $\Rightarrow$ *(CG.2)* see Theorem 4.8.

upper regular $\Leftrightarrow$ *(CG.3)* see Theorem 4.8 and recall again (for $\Leftarrow$) that $(\partial f)^{-1}(y)$ is not empty for small $\|y\|$ because $(\partial f)^{-1}(0)$ -as a singleton- is non-empty and bounded. □

5.2 Continuous and Locally Lipschitz Functions

In what follows, we intend to elaborate certain deeper, specific properties of continuous and pseudo-regular functions $f : \mathbb{R}^n \to \mathbb{R}^m$. Let us first note that we simultaneously speak of mixed systems of equations and inequalities, in this context.

For $(f,g) \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^{k+m})$, consider the mapping

$$M(a,b) = \{x \in \mathbb{R}^n \mid f(x) = a, g(x) \le b\}.$$

For $y \in \mathbb{R}^m$, we write

$$\begin{array}{rcllll} h(x,y) & = & (f(x), g(x) + y^+) & \in & \mathbb{R}^{k+m}, \\ H(a,b) & = & \{(x,y) \mid h(x,y) = (a,b)\} & \subset & \mathbb{R}^{n+m}. \end{array} \tag{5.10}$$

Since $H = h^{-1}$, pseudo-regularity of h is the pseudo- Lipschitz property of H, and M is pseudo-Lipschitz at $((a^0, b^0), x^0)$ iff so is H at $((a^0, b^0), (x^0, y^0))$ for $y^0 = b^0 - g(x^0)$.

5.2.1 Pseudo-Regularity and Exact Penalization

Let

$$f \in C(\mathbb{R}^n, \mathbb{R}^m), \ f(x^0) = 0.$$

On some bounded neighborhood Ω of x^0, we choose any small function $g \in C^{0,1}(\Omega, \mathbb{R}^m)$; recall that

$$|g| = \max\{\sup(g, \Omega), \text{Lip}\,(g, \Omega)\}$$

was defined in (4.3).

Given x near x^0 we want to find some x' (Lipschitzian close to x) satisfying the equation

$$f(x') - g(x') = f(x).$$

By the results of Section 4.1, x' exists, provided that f is pseudo-regular at $(x^0, 0)$ and $|g|$ is small. Here, we show that x' can be found by an exact penalty approach if and only if f is pseudo-regular. For this reason, we define for $\alpha > 0$

$$p_\alpha(x') = \|x' - x\| + \alpha\|f(x) + g(x') - f(x')\| \qquad (x' \in \Omega).$$

The function p_α is a *penalty function* (with parameters x and g) for the problem

$$\text{(P)} \qquad \min_{x'}\{\|x' - x\| \mid x' \in \Omega, f(x') - g(x') = f(x)\}.$$

Let $P_\alpha(x,g) = \arg\min_\Omega p_\alpha$.

Lemma 5.5 *Given* $\alpha > 0$, *it holds* $\emptyset \ne \operatorname{argmin}_{x^0+rB} p_\alpha = P_\alpha(x,g)$ *and* $P_\alpha(x,g) \subset x + \alpha\|g(x)\|B$, *whenever* $|g|$ *and* $\|x - x^0\|$ *are small enough such that* $x^0 + rB \subset \Omega$ *for* $r = \alpha|g| + \|x - x^0\|$. ◇

Proof. It holds $x \in x^0 + rB \subset \Omega$. So we have

$$P_\alpha(x,g) = \operatorname{argmin}\{p_\alpha(x') | x' \in \Omega, p_\alpha(x') \le p_\alpha(x)\}.$$

If $p_\alpha(x') \le p_\alpha(x)$, then

$$\|x' - x\| \le p_\alpha(x') \le p_\alpha(x) = \alpha\|g(x)\| \le \alpha|g|. \tag{5.11}$$

Thus,

$$\|x' - x^0\| \le \alpha|g| + \|x - x^0\| = r \text{ and } P_\alpha(x,g) = \operatorname{argmin}_{x^0+rB} p_\alpha.$$

Since $x^0 + rB$ is compact and nonempty, this yields $\operatorname{argmin}_{x^0+rB} p_\alpha \neq \emptyset$. The inclusion $P_\alpha(x,g) \subset x + \alpha\|g(x)\|B$ now follows from (5.11). □

The statement (ii) of the next theorem characterizes pseudo- regularity by the fact that, for large α, p_α is a penalty function of problem (P) provided that $\|x - x^0\| + |g|$ is small enough. The notion *exact penalty* indicates here that the minimizers are feasible points for the original problem (P),

Theorem 5.6 (pseudo-regularity and exact penalization). *Let x^0 be a zero of $f \in C(\mathbb{R}^n, \mathbb{R}^m)$, let Ω be a bounded neighborhood of x^0, and put $G = C^{0,1}(\Omega, \mathbb{R}^m)$. Then the following statements are equivalent:*

(i) f is pseudo-regular at $(x^0, 0)$.

(ii) $\exists\, \alpha, \beta > 0 :\ \xi \in P_\alpha(x,g) \Rightarrow f(\xi) = f(x) + g(\xi)\ \ \forall x \in x^0 + \beta B, g \in \beta B_G$,

where B_G is the unit ball in G with respect to $|\cdot|$ given in (4.3). ◇

Proof.

$(i) \Rightarrow (ii)$: Given $x \in x^0 + \beta B$ and $g \in \beta B_G$, let $\xi \in P_\alpha(x,g)$ and assume that (in contrast to the assertion)

$$q := g(\xi) + f(x) - f(\xi) \neq 0.$$

For small $\beta > 0$, continuity ensures that q has small norm, and by Lemma 5.5, the norm $\|\xi - x^0\|$ is small, too. From Theorem 4.3 we thus obtain the existence of a constant C such that, to each small $t > 0$, there corresponds some $x(t)$ with

$$g(x(t)) + f(x) - f(x(t)) = q - tq \text{ and } \|x(t) - \xi\| \le C\|q\|t.$$

Now choose any $\alpha > C$. Then

$$\begin{aligned} p_\alpha(x(t)) &= \|x(t) - x\| + \alpha(1-t)\|q\| \\ &\le \|x(t) - \xi\| - t\alpha\|q\| + \|\xi - x\| + \alpha\|q\| \\ &\le Ct\|q\| - t\alpha\|q\| + p_\alpha(\xi) < p_\alpha(\xi). \end{aligned}$$

So, since $x(t) \in \Omega$ for small t and β, the point $\xi \in P_\alpha(x,g)$ cannot minimize p_α. Hence $q = 0$.

$(ii) \Leftarrow (i)$: Decreasing β if necessary, we know that $r := \alpha|g| + \|x - x^0\|$ fulfills

$$x^0 + rB \subset \Omega \text{ whenever } x \in x^0 + \beta B \text{ and } g \in \beta B_G.$$

Next put g constant: $g(\cdot) = y' - f(x)$ for $y' \in \delta B, x \in x^0 + \delta B$ and $0 < \delta < \beta$. If δ is small enough, continuity of f ensures $g \in \beta B_G$. Thus, Lemma 5.5 yields the existence of some $x' \in P_\alpha(x, g)$ and guarantees the estimate

$$\|x' - x\| \le \alpha \|g(x)\| = \alpha \|y' - f(x)\|.$$

Since $f(x') = f(x) + g(x') = y'$, we obtain (i) with pseudo-Lipschitz rank $L = \alpha$ and neighborhoods U, V having radius δ. □

Let f be pseudo-regular at $(x^0, 0)$ and Ω be any fixed bounded neighborhood of x^0. From the above theorem we know that, for small $\|x - x^0\|$, small $|g|$ in $C^{0,1}(\Omega, \mathbb{R}^m)$ and for sufficiently large α, there is a minimizer x' of $p_\alpha(\xi) = \|\xi - x\| + \alpha\|f(x) + g(\xi) - f(\xi)\|$ with respect to $\xi \in \Omega$. Each such x' fulfills the perturbed equation $f(x) + g(x') = f(x')$ and the estimate $p_\alpha(x') = \|x' - x\| \le \alpha\|g(x)\| \le \alpha|g|$.

This provides us (at least theoretically) with an *exact penalty approach* for computing a solution x' of $f(x) + g(\cdot) = f(\cdot)$ being "Lipschitzian close" to x. In particular, we may identify f with the function H in (5.10). Then g takes the place of a perturbation of H, and we are speaking about solutions to a perturbed system of constraints.

What About the Infinite Dimensional Case ?

We imposed the hypothesis $x \in \mathbb{R}^n$ to obtain $P_\alpha(x, g) \neq \emptyset$ in Lemma 5.5. Concerning the image space, one may *formally* permit that $\mathbb{R}^m$ is replaced by a Banach space Y. But, if $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$, the pseudo-regularity of f already implies $m \le n$ (see the Exercise 3 in Section 2.2.3). So, at least for the locally Lipschitz case, the assumption $\dim Y = \infty$ would be an empty generalization because the remaining hypotheses cannot be satisfied even if $\dim Y > n$.

5.2.2 Special Statements for m = n

Specific properties of continuous functions $f : \mathbb{R}^n \to \mathbb{R}^m$ can be used if $m = n$ or/and f is locally Lipschitz. As a main motivation for particularly investigating such functions, we mention that (generalized) Kojima-functions are just of the considered type, provided the functions - involved in the underlying optimization problem- have Lipschitzian derivatives. Therefore, our statements concerning zeros (or level sets) of functions are closely related to critical points of optimization and variational problems. Let us start with a consequence of *Rademacher's theorem.* The latter states for any locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$ that the set N_f of those x where the Fréchet-derivative $Df(x)$ does not exist, has Lebesgue measure $\mu(N_f) = 0$. This yields, for $m = n$, a close connection between upper regularity, pseudo-regularity and isolated pre-images.

Theorem 5.7 (pseudo-regular & upper regular). *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be pseudo-regular at $(x^0, f(x^0))$. Then there are neighborhoods $U \ni x^0, V \ni f(x^0)$ and some $c > 0$ such that*

(i) For almost all $x \in U$, in particular for all points of Fréchet differentiability, f is upper regular at $(x, f(x))$.
(ii) For almost all $y \in V$, the sets $f^{-1}(y) \cap U$ are finite.

◇

Proof. (i) If $Df(x)$ exists, one has $Cf(x) = Df(x)$. By Theorem 5.1, the Jacobians (for x near x^0) are regular matrices with uniformly bounded inverses $Df(x)^{-1}$. Therefore, the point x is an *isolated solution* of $f(\cdot) = f(x)$. The upper regularity at x follows again from Theorem 5.1.

(ii) With some neighborhood U from (i), let $M = N_f \cap U$. Since $\mu(M) = 0$ and f is Lipschitz on U, also $\mu(f(M)) = 0$ is true (note that $m = n$). By pseudo-regularity, $V = f(U)$ is a neighborhood of $f(x^0)$. Let $y \in V \setminus f(M)$. Using (i), the pre-images $x \in f^{-1}(y) \cap U$ are isolated, so they are isolated for almost all $y \in V$. Taking some closed ball $U_r = x^0 + rB \subset U$, the set $f^{-1}(y) \cap U_r$ must be finite for such y. Finally, identifying U with int U_r, the assertion has been shown. □

Selections

Given any $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ with $f(x^0) = 0$, we ask now for the existence of a *continuous* function h that assigns, to y near 0, pre-images $h(y) \in f^{-1}(y)$ and satisfies $h(0) = x^0$. Evidently, h exists trivially if, for some neighborhood U of x^0, the map $y \mapsto f^{-1}(y) \cap U$ is already single-valued and continuous near the origin. The next lemma tells us that h exists *only* in this trivial case.

Lemma 5.8 (continuous selections of the inverse map). *Let $f \in C(\mathbb{R}^n, \mathbb{R}^n)$, $0 \in V \subset \mathbb{R}^n$, V be open and bounded and $h : V \to \mathbb{R}^n$ be any continuous selection of f^{-1} on V. Then $U := h(V)$ is open, $f^{-1}(v) \cap U = \{h(v)\}$ for all $v \in V$ and $h^{-1} = f|_U$.* ◇

Proof. Since $f^{-1}(v') \cap f^{-1}(v'') = \emptyset$ $(v' \neq v'')$, the selection $h : V \to U$ of f^{-1} has the inverse

$$h^{-1} : U \to V \quad \text{with} \quad h^{-1} = f|_U.$$

The function $f|_U$ has the inverse h; hence it is one-to-one, which tells us that

$$f^{-1}(v) \cap U = \{h(v)\} \text{ for all } v \in V.$$

By our hypotheses, h and $h^{-1} = f|_U$ are continuous; hence V and U are homeomorphic sets. Since V is open and bounded, so U is open, too. Notice that this conclusion is just the statement of Brouwer's *invariance of domain theorem,* cf. [AH35, Kap. X.2, Satz IX). □

Note. Consider, for the situation of the lemma above, the special case of $x^0 = h(0)$. Then x^0 is an isolated zero of f, and U is a neighborhood of x^0 such that the map $v \mapsto f^{-1}(v) \cap U$ is single-valued and continuous on V. If there exist such open sets V and U, we will also say that f^{-1} is *locally unique and continuous* near $(x^0, 0)$. ◇

Lemma 5.9 (convex pre-images). *Let $f \in C(\mathbb{R}^n, \mathbb{R}^n), V \subset \mathbb{R}^n$ be open and bounded and suppose that, for some subset $X \subset \mathbb{R}^n$, the map $v \mapsto F(v) := f^{-1}(v) \cap X$ is l.s.c. on V. Further, suppose that F has nonempty convex images for all $v \in V$. Then F is single-valued and continuous on V.* ◇

Proof. The application of E. Michael's [Mic56] selection theorem yields the existence of a continuous selection h of F on V. By Lemma 5.8, it holds that $\{h(v)\} = f^{-1}(v) \cap h(V)$. Since $U = h(V)$ is open and $f^{-1}(v) \cap X$ is convex, we easily obtain from

$$\text{card}\,(f^{-1}(v) \cap U) = 1$$

that $f^{-1}(v) \cap X$ is single-valued, too. □

Projections

Let $f \in C(\mathbb{R}^n, \mathbb{R}^n), f(x^0) = 0$ and f be pseudo- regular at $(x^0, 0)$.
Put

$$\phi(y) = \inf\{\|x - x^0\| \mid x \in f^{-1}(y)\} \quad \text{and} \quad \Psi(y) = \operatorname{argmin} \phi(y).$$

With any norm $|.|$ of $\mathbb{R}^n$, we know that, for sufficiently small $r > 0$ and $y, y' \in rB$, the closed set $f^{-1}(y)$ is nonempty, and

$$\emptyset \neq \Psi(y) \subset x^0 + L|y|B,$$

where L may depend on the fixed norm. Moreover, given $\xi \in \Psi(y)$, there exists $x' \in f^{-1}(y')$ such that $\|x' - \xi\| \leq L|y' - y|$, hence

$$\phi(y') \leq \phi(y) + L|y' - y|.$$

So the distance ϕ is Lipschitz on rB. This yields the well-known observation:

If card $\Psi(y) = 1$ on rB, then the projection Ψ is continuous.

For fixed y, the number of elements in $\Psi(y) \subset f^{-1}(y)$ may depend on the used norm.

Theorem 5.10 (equivalence of pseudo- and strong regularity, bifurcation). *Let $f \in C(\mathbb{R}^n, \mathbb{R}^n), f(x^0) = 0$ and f be pseudo-regular at $(x^0, 0)$.*
Then the following properties are equivalent.

(i) f^{-1} has, on some neighborhood of the origin, a continuous selection h with $h(0) = x^0$.
(ii) The projection map Ψ is single-valued near the origin.
(iii) f is strongly regular at $(x^0, 0)$.

Moreover, if f is not strongly regular at $(x^0, 0)$, then $(0, x^0)$ is a bifurcation point of f^{-1} such that, near the origin, f^{-1} has no (single-valued) continuous selection s satisfying $s(0) = x^0$. ◇

Proof. The statements follow immediately from the Selection Lemma 5.8 along with continuity of Ψ for card $\Psi(\cdot) \equiv 1$. □

By the theorem, the local uniqueness of the projection Ψ *near the origin* does not depend on the used norm, *provided that* $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ *is pseudo-regular.* Recall that Example 1.4 presented just a pseudo-regular function which was not strongly regular.

5.2.3 Continuous Selections of Pseudo-Lipschitz Maps

Suppose that

$$f \in C(\mathbb{R}^n, \mathbb{R}^n) \text{ is pseudo-regular at an } \textit{isolated} \text{ zero } x^0 \text{ of } f.$$

Then one easily sees that, if $r > 0$ is small enough, the multifunction H defined by

$$H(y) = f^{-1}(y) \cap (x^0 + 2L|y|B) \text{ (with L from pseudo-regularity)}$$

has compact images, is Lipschitz (upper and lower) on rB and fulfills $H(0) = \{x^0\}$.

With the same properties, H can be defined for isolated zeros of any pseudo-regular mapping. However, if x^0 is not isolated, the existence of a nontrivial *continuous compact-valued selection* $H \subset f^{-1}$ seems to be an open problem even for Lipschitz functions f. Nevertheless, the existence of H is ensured for piecewise C^1 **-functions.**

Lemma 5.11 (isolated zeros of PC^1**-functions**). *Let* $f \in PC^1(\mathbb{R}^n, \mathbb{R}^n)$ *be pseudo-regular at* $(x^0, 0)$. *Then* x^0 *is an isolated zero of* f. *Thus, near the origin,* f^{-1} *has a Lipschitz continuous, compact-valued selection* $H \subset f^{-1}$ *with* $H(0) = \{x^0\}$. ◇

Exercise 7. Verify Lemma 5.11. ◇

The previous lemma is a special case of a powerful theorem, which was recently shown by P. Fusek.

Theorem 5.12 (isolated zeros of Lipschitz-functions, $m = n$). *Suppose that* $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ *is pseudo-regular at* $(x^0, 0)$ *and directionally differentiable at* x^0. *Then* x^0 *is an isolated zero of* f *and, in addition,* $f'(x^0; u) \neq 0$ *holds for all* $u \neq 0$.

Moreover, if f *is even directionally differentiable for* x *near* x^0, *then there is some* $\varepsilon > 0$ *such that* $\|f'(x; u)\| \geq \varepsilon$ *for all* $u \in \text{bd } B$ *and* $x \in x^0 + \varepsilon B$. ◇

Proof. See [Fus99, Fus01). □

Exercise 8. Let $F : B \subset \mathbb{R}^2 \to \mathbb{R}^2$ be defined as follows:

$$F(y) = \{x \in B | \ \|x - y\| \geq \frac{1}{2}\}.$$

Analyze the continuity properties for F with the Euclidean norm, with polyhedral norms and with " > " instead of " ≥ " in the definition of F. Does there exist a continuous function $f : B \to B$ such that $f(\cdot) \in F(\cdot)$ on B ? ◇

Exercise 9. Find a counterexample $(n = m = 2)$ showing that the pointwise condition (5.4) in Theorem 5.1 is not sufficient for the Lipschitz l.s.c. of F^{-1}.

◇

5.3 Implicit Lipschitz Functions on $\mathbb{R}^n$

For locally Lipschitz functions f in finite dimension, the derivative Tf describes precisely (local) uniqueness and Lipschitz behavior of inverse and implicit functions. This will follow from the subsequent Theorems 5.14 and 5.15, shown in [Kum91b] (including Example BE.3, too) and [Kum91a], respectively.

Let us first recall Clarke's basic inverse function theorem in [Cla76].

Theorem 5.13 (inverse functions and ∂f). *If $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ and all matrices in $\partial f(x^0)$ are regular, then f is strongly regular at $(x^0, f(x^0))$.* ◇

This statement follows from the next Theorem and Theorem 6.6. Replacing now ∂f by Tf, we obtain a *necessary* and sufficient condition and a clear description for the T-derivative of the inverse. In Chapter 6, we shall see that Tf fulfills also several chain rules which are important for computing them in relevant special cases.

Theorem 5.14 (inverse functions and *Tf*). *A function $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x^0, f(x^0))$ if and only if*

$$\exists c > 0: \ \|f(x') - f(x)\| \geq c\|x' - x\| \ \forall x, x' \in x^0 + cB. \tag{5.12}$$

Moreover, if f even belongs to $C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$, and Ω is some neighborhood of x^0, then the following statements are equivalent.

(i) f is strongly regular at $(x^0, f(x^0))$.

(ii) $Tf(x^0)$ is injective.

(iii) f is strongly regular at $(x^0, f(x^0))$ with respect to $G = C^{0,1}(\Omega, \mathbb{R}^n)$.

If $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x^0, f(x^0))$, then the locally unique (and Lipschitz) inverse f^{-1} satisfies the equivalence

$$u \in Tf^{-1}(f(x^0))(v) \Leftrightarrow v \in Tf(x^0)(u).$$

◇

Note. In comparison with Theorem 5.1, the requirement "f^{-1} is Lipschitz l.s.c. at (y^0, x^0)" does not appear and is, on the contrary, a key consequence of (ii). The first statement of the theorem was already a footnote in [Cla76].

Proof. Let (5.12) be true. Then the open ball $U = x^0 + cB^\circ$ and its f-image $V = f(U)$ are homeomorphic because (5.12) ensures that f^{-1}, as a mapping of the type $V \to U$, is well-defined and Lipschitz with constant c^{-1}. By the invariance of domain theorem, V is open if so is U. Since $f(x^0) \in V$, we thus obtain $f(x^0) \in \text{int } V$. Hence f is strongly regular at x^0.

Conversely, (5.12) follows immediately from strong regularity of f at x^0 : Put $y = f(x), y' = f(x')$ and apply that f^{-1} is locally Lipschitz. Thus the first statement is true; (i) $\Leftrightarrow$ (5.12).

Now, let f be locally Lipschitz, say with some constant K on Ω. (i)$\Rightarrow$ (ii) follows from Theorem 5.1. (ii) $\Rightarrow$ (5.12) can be seen as follows: If (5.12) is not true, then there are sequences $x \to x^0, x' \to x^0, c \downarrow 0$ with

$$\|f(x') - f(x)\| < c\|x' - x\|.$$

Setting $u' = (x' - x)/\|x' - x\|$ the sequence of $u' \in \mathbb{R}^n$ has a cluster point $u^0 \neq 0$. With $t = \|x' - x\|$ we thus obtain $x' = x + tu'$ and, for some appropriate subsequence,

$$0 = \lim t^{-1}(f(x + tu') - f(x)) \in Tf(x^0)(u^0);$$

hence (ii) is violated. Thus, using (i) $\Leftrightarrow$ (5.12), we also have (i) $\Leftrightarrow$ (ii). For the equivalence (i) $\Leftrightarrow$ (iii), we refer to Corollary 4.4.

The formula for Tf^{-1} is an evident consequence of the definitions. □

The implication (i) $\Rightarrow$ (iii) is not restricted to Lipschitz functions in finite dimension, only. As already mentioned in the proof of Corollary 4.4, one may consider the function $x \mapsto f^{-1}(g(x))$ which maps a small neighborhood of x^0 into itself and is *contractive* as far as f is *strongly* regular at x^0 and $|g|$ is small in $C^{0,1}(\Omega, \mathbb{R}^n)$.

Example BE.3 indicates that the conditions of Theorem 5.14 are weaker than F.H. Clarke's requirement of all matrices in $\partial f(x^0)$ being regular because there is a Lipschitz-homeomorphism of $\mathbb{R}^2$ (piecewise linear) such that $\partial f(x^0)$ contains the zero matrix. So, Theorem 5.13 presents a sufficient condition for strong regularity, which is not necessary even for piecewise linear functions ($n >$ 1). Equipped with chain rules for computing Tf, Theorem 5.14 turns out to be a powerful tool for strong stability analysis of critical points in (finite dimensional) $C^{1,1}$ optimization and for regularity of generalized Kojima functions.

Additional Nonlinear Perturbations

Theorem 5.14 can be extended to systems of the form

$$f(x, p) = y, \quad \text{where } f \in C^{0,1}(\mathbb{R}^{n+m}, \mathbb{R}^n) \text{ and } f(x^0, 0) = 0, \tag{5.13}$$

and to the related implicit function

$$x = x(p, y) \text{ for } x \text{ near } x^0 = x(0, 0).$$

In order to exploit the equivalence of (i) and (iii) in Theorem 5.14, we can consider, as in §4.3, the nonlinear perturbation $g(x, p) = f(x, 0) - f(x, p)$ of the original function. *But one has to be careful:* In §4.1 and §4.3 we needed small $C^{0,1}$-norms of $g(\cdot, p)$ on some neighborhood Ω of x^0 as $p \to 0$.

The latter *is not necessarily ensured* by supposing f to be locally Lipschitz, see Example 6.7 in §6.4. We only know that the sup-norm on Ω fulfills

$$\sup_{x\in\Omega} \|f(x,0) - f(x,p)\| \to 0 \text{ as } p \to 0.$$

If $f(\cdot,0)$ is strongly regular at $(x^0,0) \in \mathbb{R}^{2n}$, Theorem 4.5 ensures the *existence and local upper Lipschitz behavior* of solutions to (5.13), but not the uniqueness. Nevertheless, the Thibault derivative $Tf(x^0,0)$ describes again the implicit-function situation near $(x^0,0,0)$.

Theorem 5.15 (implicit Lipschitz functions). *The implicit functio $x = x(p,y)$ to (5.13) locally exists as a uniquely defined Lipschitz function (that maps a neighborhood V of* (0,0) into some *neighborhood U of x^0) if and only if* $0 \notin Tf(x^0,0)(u,0)$ *holds for each* $u \neq 0$. ◇

Proof.

($\Leftarrow$) For $p = 0$, the function $f(\cdot,0) : \mathbb{R}^n \to \mathbb{R}^n$ has an inverse φ being locally Lipschitz near $(0, x^0)$ by Theorem 5.14. So there are positive α and β such that the solutions $x = \varphi(y)$ in $x^0 + \alpha B$ of $f(x,0) = y$ are well defined and Lipschitz for $\|y\| \le \beta$. By Theorem 4.5, we find positive α' and β' such that

$$f(x,p) = y, \qquad x \in x^0 + \alpha' B$$

is solvable whenever $\|p\| + \|y\| \le \beta'$, and related solutions $x(p,y)$ satisfy

$$\|x(p,y) - x^0\| \le C\|(p,y)\| \text{ with some (local) upper Lipschitz rank } C.$$

Therefore, if ($\Leftarrow$) is wrong, then there exist sequences $p_k, p'_k \to 0, y_k, y'_k \to 0$ and related solutions $\xi_k = x(p_k, y_k)$ and $\xi'_k = x(p'_k, y'_k)$, both tending to x^0, such that

$$\|\xi'_k - \xi_k\| / \|(p'_k, y'_k) - (p_k, y_k)\| \to \infty.$$

Then, $f(\xi_k, p_k) = y_k$, $f(\xi'_k, p'_k) = y'_k$ and $\|f(\xi'_k, p_k) - f(\xi'_k, p'_k)\| \le L_f \|p'_k - p_k\|$. Setting $t_k = \|\xi'_k - \xi_k\|$ and $w_k = (\xi'_k - \xi_k)/t_k \in B$, we obtain

$$\begin{aligned} &\|f(\xi_k + t_k w_k, p_k) - f(\xi_k, p_k)\| \\ &\le \| f(\xi'_k, p'_k) - f(\xi_k, p_k)\| + \|f(\xi'_k, p_k) - f(\xi'_k, p'_k)\| \\ &\le \| y'_k - y_k\| + L_f \| p'_k - p_k\|. \end{aligned}$$

Because of $t_k^{-1}\|(p'_k, y'_k) - (p_k, y_k)\| \to 0$, this tells us that

$$\begin{aligned} &\lim t_k^{-1} \|f(\xi_k + t_k w_k, p_k) - f(\xi_k, p_k)\| \\ &\le \lim t_k^{-1} (\|y'_k - y_k\| + L_f \|p'_k - p_k\|) = 0. \end{aligned}$$

After selecting a subsequence such that w_k converges to some u, this just means

$$0 \in Tf(x^0,0)(u,0), u \neq 0.$$

($\Rightarrow$) If $0 \in Tf(x^0, 0)(u, 0)$ and $u \neq 0$, then we find

$$\lim t_k^{-1}(f(\xi_k + t_k w_k, p_k) - f(\xi_k, p_k)) \to 0$$

for appropriate sequences converging as noted above. Setting

$$y'_k = f(\xi_k + t_k w_k, p_k) \quad \text{and} \quad y_k = f(\xi_k, p_k)$$

one easily sees that the inverse $x(p, y)$, if it exists at all, cannot be locally Lipschitz. □

Relations between Tf and "partial derivatives" $T_x f$ and $T_p f$ will be studied in Section 6.4.

Exercise 10. Verify: If $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x^0, f(x^0))$ and directionally differentiable for x near x^0 then the local inverse f^{-1} is directionally differentiable for y near $f(x^0)$. ◇

Chapter 6

Analysis of Generalized Derivatives

In this chapter, we study properties of selected generalized derivatives which will play a crucial role in the subsequent chapters. We mainly focus on the contingent derivative *CF* and the Thibault derivative *TF* of some given (multi-) function *F*, both generalized derivatives were introduced in Section 1.2. The presented properties of *CF* can be found in [AE84], the related statements for *TF* are often similar, we refer, e.g., to [Thi80, Kum91b]. Moreover, we examine the relations between Thibault derivatives and Clarke's [Cla83] generalized Jacobians with respect to locally Lipschitzian functions, and we discuss so–called *Newton maps* [Kum00a] which are set-valued first–order approximations of nonsmooth functions and are of interest in the local convergence analysis of Newton–type methods.

6.1 General Properties for Abstract and Polyhedral Mappings

Suppose that

$$X \text{ and } Y \text{ are normed spaces, } F : X \rightrightarrows Y \text{ and } z = (x, y) \in \operatorname{gph} F. \tag{6.1}$$

We recall the definitions of $CF(z) : X \rightrightarrows Y$ and $TF(z) : X \rightrightarrows Y$ given in Section 1.2 above. One has $v \in CF(z)(u)$ if there exists a sequence $t \downarrow 0$ and an associated sequence $(u_t, v_t) \to (u, v)$ such that $y + tv_t \in F(x + tu_t)$, while $v \in TF(z)(u)$ means that there exists some sequence $t \downarrow 0$ and associated sequences $(u_t, v_t) \to (u, v)$ and $(x_t, y_t) \to (x, y) \in \operatorname{gph} F$ such that $y_t + tv_t \in F(x_t + tu_t)$.

We will also use the characterization that $v \in TF(z)(u)$ iff there are $t \downarrow 0$, $(x_t, y_t) \to z$ in gph *F* and *o*-type functions o_1, o_2 such that $y_t + tv + o_2(t) \in F(x_t + tu + o_1(t))$. Clearly, $o_1(t) = t(v_t - v)$ and $o_2(t) = t(u_t - u)$.

By the definitions only, it holds

$$\limsup_{z' \to z \text{ in gph } F,\, u' \to u} CF(z')(u') \subset TF(z)(u).$$

In general, the inclusion in that relation cannot be replaced by an equation, see the simple example $F(x) = |x|$.

Tangents

Derivatives of multifunctions are closely related to *tangents of sets*, cf., e.g., [Cla83, AE84, RW98, BL00]. To see this, let $Z \subset X$ and $z \in Z$. Note that convergence of sequences is written index–free according to the convention of §1.1.

One says that $w \in X$ belongs to the *contingent cone* $C(z, Z)$ of Z at z (also called *Bouligand cone*) if, for some sequence $t \downarrow 0$ and some related sequence $w_t \to w$, there holds $z + tw_t \in Z$.

More restrictive, w belongs to *Clarke's tangent cone* $T_c(z, Z)$ if, whenever $t \downarrow 0$ and $z_t \to z$ in Z, there are related $w_t \to w$ such that $z_t + tw_t \in Z$.

Another cone can be defined by saying that w belongs to $T(z, Z)$ if there exist certain sequences $t \downarrow 0, w_t \to w$ and $z_t \to z$ in Z such that $z_t + tw_t \in Z$. Evidently,

$$T_c(z, Z) \subset C(z, Z) \subset T(z, Z).$$

If Z is the closure of an open set, then $T(z, Z)$ is the whole space. If Z is a convex polyhedral set in finite dimensions, then $T_c(z, Z)$, $C(z, Z)$ and $T(z, Z)$ are convex polyhedral cones.

Returning to the map F and setting $Z = \text{gph}\, F \subset X \times Y$, the definitions of tangents in the product space (X, Y) yield that

$$\text{gph}\, CF(z) = C(z, \text{gph}\, F) \qquad \text{and} \qquad \text{gph}\, TF(z) = T(z, \text{gph}\, F).$$

Elementary Properties

Because of the definition via limits, the sets $TF(z)(u)$ and $CF(z)(u)$ are closed in Y. If the images of F are convex then $CF(z)(u)$ or $TF(z)(u)$ are not necessarily convex again. This fact explains one type of difficulties for establishing a related differential calculus for CF and TF as well.

Another type of difficulties comes from the fact that, for writing some element $v \in TF(z)(u)$ in limit-form, it may happen that one needs a particular sequence $t = t_k \downarrow 0$. In other words, some sequence $t \downarrow 0$ already assigned to the derivative of another function may be inappropriate to represent $v \in TF(z)(u)$ in limit-form. This leads to difficulties if we want to show the additivity $TF(z) + TG(z) = T(F + G)(z)$ or other chain rules.

On the other hand, many proofs concerning chain rules for generalized derivatives are only straightforward consequences of the definitions: One has to select appropriate converging subsequences from a given one (which exists by the definitions only). If this is possible, we will say that the related rule can be shown *in an elementary way.*

As a **drastic example of an invalid chain rule** we regard the following conjecture: Given real functions f, g that are continuous at z, it holds

$$C(f+g)(z)(u) = \{0\} \quad \text{provided that} \quad Cf(z)(u) = Cg(z)(u) = \{0\}.$$

Put $f(x) = \sqrt{|x|}$ if x is rational, $f = 0$ otherwise; $g(x) = \sqrt{|x|}$ if x irrational, $g = 0$ otherwise. Then $(f+g)(x) = \sqrt{|x|}$ and $C(f+g)(z)(u) = \emptyset$ for $z = 0$ and $u = 1$. Hence, the conjecture is false.

Lemma 6.1 (*TF, CF* are homogeneous; TF^{-1}, CF^{-1}). *Suppose (6.1). Then*

$$\begin{array}{lcll} TF(z)(ru) & = & rTF(z)(u) & \forall r \in \mathbb{R} \quad \forall u \in X, \\ CF(z)(\lambda u) & = & \lambda CF(z)(u) & \forall \lambda \geq 0 \quad \forall u \in X. \\ v \in TF(x,y)(u) & \Leftrightarrow & u \in T(F^{-1})(y,x)(v), & \\ v \in CF(x,y)(u) & \Leftrightarrow & u \in C(F^{-1})(y,x)(v). & \end{array}$$

◇

Proof. Given $v \in TF(z)(u)$ one may reverse the role of $y_t + tv_t$ and y_t as well as of $x_t + tu_t$ and x_t in the definition of *TF*. Then one obtains $-v \in TF(z)(-u)$. Thus, $TF(z)$ is a *homogeneous* mapping. Similarly, one sees that $CF(z)$ is only *positively homogeneous.* As already noticed in Remark 1.1, the equivalences for the inverse multifunction F^{-1} are evident due to the symmetric definitions of the derivatives. □

Lemma 6.2 (variation by C^1 functions). *Suppose that X and Y are normed spaces,* $f \in C^1(X,Y)$, $F : X \rightrightarrows Y$ *and* $G = f + F$. *Then*

$$\begin{array}{lcl} TG(x, f(x) + y)(u) & = & Df(x)u + TF(x,y)(u), \\ u \in T(G^{-1})(y + f(x), x)(v) & \Leftrightarrow & u \in T(F^{-1})(y,x)(v - Df(x)u). \end{array}$$

The mappings G and Γ, *where* Γ *is defined by*

$$\Gamma(\xi) = f(x^0) + Df(x^0)(\xi - x^0) + F(\xi),$$

have the same derivatives $TG(z^0) = T\Gamma(z^0)$ *at* $z^0 = (x^0, y^0) \in \operatorname{gph} G$.

All these statements are also valid for the contingent derivatives CG, CG^{-1} *and* $C\Gamma$. ◇

Proof. By writing down the related limits, one directly sees that

$$\begin{array}{l} TG(x, f(x) + y)(u) = Df(x)u + TF(x,y)(u), \\ CG(x, f(x) + y)(u) = Df(x)u + CF(x,y)(u). \end{array}$$

Taking Lemma 6.1 into account we thus obtain that

$$\begin{array}{lcl} u \in T(G^{-1})(y + f(x), x)(v) & \Leftrightarrow & v \in TG(x, f(x) + y)(u) \\ & \Leftrightarrow & v - Df(x)u \in TF(x,y)(u) \\ & \Leftrightarrow & u \in T(F^{-1})(y,x)(v - Df(x)u). \end{array}$$

The analogous arguments hold true for contingent derivatives. Moreover, if in particular $Df(x) = 0$ and $f(x) = 0$, then

$$TG(x,y) = TF(x,y) \text{ and } T(G^{-1})(y,x) = T(F^{-1})(y,x).$$

Applying this fact to the "difference" of G and Γ at $x = x^0$, G and Γ have the same T-derivatives at z^0. Again, the analogous arguments are valid for CG and $C\Gamma$. □

Similarly, one shows by elementary means

Lemma 6.3 (small variations of given mappings). *Let X and Y be normed spaces, $F, G : X \rightrightarrows Y$, $x^0 \in X$, $f(x^0) = 0$ and $z^0 = (x^0, y^0) \in \operatorname{gph} F \cap \operatorname{gph} G$. Then the following properties hold.*

1. *If* $\operatorname{Lip}(f, x^0 + rB) = O(r)$, *then* $T(f+F)(z^0) = TF(z^0)$.
2. *If* $f(x) = o(x - x^0)$, *then* $C(f+F)(z^0) = CF(z^0)$.
3. *If* $Tf(x^0)(u) \equiv \{0\}$ *and* $f \in C^{0,1}$, *then* $T(f+F)(z^0) = TF(z^0)$.
4. *If* $d_H(F(x), G(x)) = o(x - x^0)$, *then* $CF(z^0) = CG(z^0)$.

◇

Polyhedral Maps

An important special class of multifunctions $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ was introduced and studied by S.M. Robinson [Rob81]. One says that F is *polyhedral,* if gph F is a union of a finite number of convex polyhedral sets P_k (bounded or not). Such a union will also be called a *polyhedral set.* Clearly, F is polyhedral if and only if so is F^{-1}.

Examples

Given a convex polyhedron $P \subset \mathbb{R}^n$, the following mappings are polyhedral and send $\mathbb{R}^n$ into itself:

(i) the *projection map* $\pi(x, P)$ of x onto P, which is multivalued for polyhedral norms and piecewise linear for the Euclidean norm;

(ii) the *map of normals* $N_P(x) = \{y \mid \langle y, p - x\rangle \le 0 \ \forall\, p \in P\}$ (if $x \in P$) and $N_P(x) = \emptyset$ (if $x \in \mathbb{R}^n \setminus P$), which is related to the Euclidean projection π by $y \in N_P(x) \Leftrightarrow \pi(x + y, P) = x$.

Further important examples are the solution maps of parametric linear inequalities and parametric linear complementarity problems, respectively,

(iii) $S(y) = \{x \in \mathbb{R}^n \mid Ax \le y\}$, $y \in \mathbb{R}^m$,

(iv) $F(z) = \{(x,y) \in \mathbb{R}^{2n} \mid Ax + By = z, x \ge 0, y \ge 0, \langle x, y\rangle = 0\}$, $z \in \mathbb{R}^m$,

where A, B are given (m,n)–matrices, as well as linear transformations of convex hulls

(v) $M(x) = \operatorname{conv}\{p^k + L_k x \mid k = 1, ..., N\}$ with given points $p^k \in \mathbb{R}^m$, and linear functions $L_k : \mathbb{R}^n \to \mathbb{R}^m$.

Properties

Having a finite representation of the graph *by closed sets,*

$$\operatorname{gph} F = \cup_k P_k,$$

the submappings F_k, defined by $\operatorname{gph} F_k = P_k$, still describe the contingent-tangent cones and the contingent derivatives via

$$C(z, \operatorname{gph} F) = \bigcup_{k:\, z \in P_k} C(z, P_k) \quad \text{and} \quad CF(z)(u) = \bigcup_{k:\, z \in P_k} CF_k(z)(u), \tag{6.2}$$

respectively. Clearly, some of these sets $CF_k(z)(u)$ may be empty. For polyhedral mappings, the cones $C(z, P_k)$ are solution sets of linear inequality systems $A_k v \le 0$, hence they are computable and have a nice structure. This is the key for showing the next well-known statements.

Given a linear transformation $A : \mathbb{R}^m \to \mathbb{R}^d$ we define

$$AF(\cdot) = \{Ay \mid y \in F(\cdot)\}.$$

Clearly, *AF* is polyhedral if *F* is so.

Theorem 6.4 (polyhedral mappings). *Let* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a polyhedral multifunction, and let* $z = (x, y) \in \operatorname{gph} F$. *Then:*

(i) *The contingent derivative* $CF(z)(\cdot)$ *is again polyhedral.*

(ii) *(exact approximation). For sufficiently small* $\varepsilon > 0$, *it holds*
$(z + \varepsilon B) \cap \operatorname{gph} F = (z + \varepsilon B) \cap (z + \operatorname{gph} CF(z))$, *and*
$CF(z)(u) = \{v \in \mathbb{R}^m | y + tv \in F(x + tu)\ \forall\, t \in (0, \varepsilon)\}$ *if* $\|u\| \le 1$.

(iii) *(linear transformation). Under linear transformations* $A : \mathbb{R}^m \to \mathbb{R}^d$, *it holds*
$A[CF(x, y)(u)] \subset C(AF)(x, Ay)(u)$
Conversely, if $c \in C(AF)(x, Ay)(u)$, *then one has*

$$c \in A[CF(x, y)(u)] \ \Leftrightarrow\ t \mapsto \Gamma(t),\ t \ge 0, \text{ is l.s.c. at } (0, y), \tag{6.3}$$

where Γ *is defined by* $\Gamma(t) = F(x + tu) \cap \{\eta | A\eta = Ay + tc\}$. ◇

Note. In particular, Γ is l.s.c. at $(0, y)$ provided that

(i) the matrix *A* is regular (due to $\eta = y + tA^{-1}c$), or

(ii) *F* is upper regular at (x, y) and $\Gamma(t) \ne \emptyset$ for small $t > 0$.

In general, the l.s.c. condition for Γ may fail to hold: Let $\operatorname{gph} F = P \cup Q$ be given by $P = \{(\xi, 0) | \xi \in R\} \subset \mathbb{R}^2$ and $Q = \{(0, y)\}$ with fixed $y \ne 0$. Then $CF(0, y)(u)$ is empty for $u \ne 0$, but $0 \in C(AF)(0, Ay)(u)$ holds for $Ay = 0$. ◇

Proof of Theorem 6.4. The statements (i) and (ii) are left as Exercise 11. We consider the statements (ii).

($\subset$) Let $v \in CF(x, y)(u)$. Then, for small $t > 0$, it holds $y + tv \in F(x + tu)$ and, by definition,

$$Ay + tAv \in (AF)(x + tu).$$

Hence,

$$Av \in C(AF)(x, Ay)(u).$$

($\supset$) The supposition $c \in C(AF)(x, Ay)(u)$ means that, for certain $t \downarrow 0$, there are elements η satisfying

$$Ay + tc = A\eta \quad \text{and} \quad \eta \in F(x + tu). \tag{6.4}$$

We have to verify that $c = Av$ holds for some $v \in CF(x, y)(u)$.

If the latter holds true then, as shown in the first part, $\eta = y + tv$ belongs to $\Gamma(t)$ and fulfills $\|\eta - y\| \le t\|v\|$. Thus Γ is necessarily l.s.c. for $c \in A[CF(x, Ay)(u)]$.

We show the sufficiency. If Γ is l.s.c. at $(0, y)$, then the optimization problem

$$\min_\eta\{\|\eta - y\| \mid \eta \text{ satisfies (6.4)}\}$$

has solutions $y(t)$ for small $t \downarrow 0$, and $y(t) \to y$. Because F is polyhedral, we find a subsequence of t and a fixed convex polyhedron $P \subset \text{gph}\, F$ such that the points $(x + tu, y(t))$ belong to P. Let P have the implicit description

$$(\xi, \eta) \in P \Leftrightarrow M\xi + N\eta \le w$$

with appropriate fixed matrices M, N and vector w. Then, for certain $t \downarrow 0$, our minimum problem reads

$$(6.4)_t \qquad \min_\eta \|\eta - y\| \text{ s.t. } A\eta = Ay + tc, \; N\eta \le w - Mx - tMu.$$

The linear constraints depend on a real parameter $t > 0$ on the right-hand sides, and the related feasible set map is Lipschitz l.s.c. on its closed *domain*, by Hoffman's lemma given in Section 2.1 above. Therefore, $\|y(t) - y\| \le Lt$ holds for some L and small t. After selecting an accumulation point v of $(y(t) - y)/t$ for $t \downarrow 0$, one sees that $v \in CF(x, y)(u)$. Along with (6.4) for $\eta = y(t) = y + tv + o(t)$, this yields the claim $c = Av$. □

Exercise 11. Prove the statements (i) and (ii) of Theorem 6.4. ◇

6.2 Derivatives for Lipschitz Functions in Finite Dimension

In this section, we suppose that

$$f : \mathbb{R}^n \to \mathbb{R}^m \text{ is a locally Lipschitz function with rank } L \text{ near } x.$$

Then the differences after replacing u by a sequence $u^k \to u$ in the definition of Tf,

$$d(k) := t_k^{-1}[f(x^k + t_k u^k) - f(x^k)] - t_k^{-1}[f(x^k + t_k u) - f(x^k)],$$

are vanishing due to $\|d(k)\| \le L\|u^k - u\| \to 0$.

The same holds in view of Cf. Hence,

$$Tf(x)(u) = \left\{ v \,\middle|\, \begin{array}{l} v = \lim t_k^{-1}[f(x^k + t_k u) - f(x^k)] \\ \text{for certain } t_k \downarrow 0 \text{ and } x^k \to x \end{array} \right\}, \tag{6.5}$$

$$Cf(x)(u) = \left\{ v \,\middle|\, \begin{array}{l} v = \lim t_k^{-1}[f(x + t_k u) - f(x)] \\ \text{for certain } t_k \downarrow 0 \end{array} \right\}. \tag{6.6}$$

These sets are

$$\text{non-empty, closed and bounded } (\subset L\|u\|B)$$

because f is locally Lipschitz and $Y = \mathbb{R}^m$.

For $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, it holds

$$Cf = Tf = Df.$$

For the absolute value $f(x) = |x|$ we observe that

$$Cf(0)(u) = \{f'(0; u)\} \quad \text{(the usual directional derivative)},$$

and

$$Tf(0)(u) = [-|u|, |u|] \quad \text{(a closed interval)}.$$

So Cf and Tf are different even for elementary functions.

Further Properties

$$Tf(x)(u) \text{ and } Cf(x)(u) \text{ are connected sets.} \tag{6.7}$$

Proof. We consider Tf. Assume there are some x^0 and disjoint open sets Ω_1 and Ω_2 such that $M := Tf(x^0)(u)$ meets Ω_1 and Ω_2 , and $M \subset \Omega_1 \cup \Omega_2$. Then the same is true for the (larger) set

$$M(\varepsilon) := \{v \mid v = t^{-1}(f(x + tu) - f(x)),\ d(x, x^0) < \varepsilon,\ 0 < t < \varepsilon\}$$

as far as $\varepsilon > 0$ is sufficiently small. Indeed, it holds $M = \limsup_{\varepsilon \downarrow 0} M(\varepsilon)$, and all the sets under consideration are uniformly bounded. Next, let elements $v^i \in M(\varepsilon) \cap \Omega_i$ be written with related t_i and x^i. Setting, for $0 \le \lambda \le 1$, $x(\lambda) = \lambda x^1 + (1 - \lambda)x^2$ and $t(\lambda) = \lambda t_1 + (1 - \lambda)t_2$, the points $v(\lambda) = t(\lambda)^{-1}[f(x(\lambda) + t(\lambda)u) - f(x(\lambda))]$ belong to $M(\varepsilon)$ and form a continuous curve, connecting v^1 and v^2. So $M(\varepsilon) \subset \Omega_1 \cup \Omega_2$ cannot be true; M is connected. For Cf, put $x = x^0$. □

Both mappings are Lipschitz in u, i.e.

$$Tf(x)(u') \subset Tf(x)(u) + L\|u' - u\|B \tag{6.8}$$

$$Cf(x)(u') \subset Cf(x)(u) + L\|u' - u\|B. \tag{6.9}$$

This follows from the Lipschitz estimate

$$\|t_k^{-1}[f(x^k+t_ku')-f(x^k)]-t_k^{-1}[f(x^k+t_ku)-f(x^k)]\|\le L\|u'-u\|,$$

by assumption. □

The mapping Tf has a useful *subadditivity* property which was already shown in [Thi80].

$$Tf(x)(u'+u'')\subset Tf(x)(u')+Tf(x)(u''). \tag{6.10}$$

Setting $\xi^k=x^k+t_ku''$, (6.10) follows due to

$$\begin{aligned}&f(x^k+t_k(u'+u''))-f(x^k)\\&=f(x^k+t_ku''+t_ku')-f(x^k+t_ku'')+f(x^k+t_ku'')-f(x^k),\\&=f(\xi^k+t_ku')-f(\xi^k)+f(x^k+t_ku'')-f(x^k)\end{aligned}$$

and by passing to the limits as $k\to\infty$. □

The statements

$$Tf \text{ is a closed mapping of both arguments,} \tag{6.11}$$

$$T(f+g)(x)(u)\subset Tf(x)(u)+Tg(x)(u) \text{ if } f \text{ and } g \text{ are locally Lipschitz,} \tag{6.12}$$

follow directly from the definition by selecting appropriate subsequences. □

For locally Lipschitz functions, the relation between D^*f and Cf is very close. To see this, we prove a further characterization of D^*f. Let

$$\sup\langle v^*,\ Cf(\xi)(u)\rangle \text{ denote } \sup_{v\in Cf(\xi)(u)}\langle v^*,\ v\rangle.$$

Theorem 6.5 (Cf and D^*f). For $f\in C^{0,1}(\mathbb{R}^n,\mathbb{R}^m)$, *onehas:*

(i) $u^*\in D^*f(x)(v^*)$ *if and only if*
$0\in D^*g(x)(1)$ *for* $g=\langle u^*,\cdot\rangle+\langle v^*,f(\cdot)\rangle$

(ii) $u^*\in D^*f(x)(v^*)$ *if and only if*
$\liminf_{\xi\to x}\ \sup_{u\in B}\ (\langle u^*,u\rangle\ +\sup\langle v^*,Cf(\xi)(u)\rangle)\le 0.$ ◇

Proof. (i) By Remark 3.6, the relation $u^*\in D^*f(x^0)(v^*)$ means that

$$\begin{aligned}&\langle u^*,u\rangle+s^{-1}\langle v^*,f(x_t+su)-f(x_t)\rangle\le t+ts^{-1}\|f(x_t+su)-f(x_t)\|\\&\text{whenever } 0<s\le r_t \text{ and } u\in B \text{ hold for certain sequences}\\&t=t_k\downarrow 0, x_t\to x \text{ and } r_t\downarrow 0.\end{aligned} \tag{6.13}$$

The left-hand side of (6.13) does not depend on t explicitly. The right-hand side fulfills

$$t+ts^{-1}\|f(x_t+su)-f(x_t)\|\le t(1+L_f)$$
with some Lipschitz rank L_f of f near x^0

and vanishes.

So $u^* \in D^*f(x)(v^*)$ means equivalently that

$$\begin{aligned}&\sup_{u\in B, 0<s\le r(t)}\langle u^*,u\rangle + s^{-1}\langle v^*, f(x_t+su)-f(x_t)\rangle \le \varepsilon(t)\\ &\text{with } \varepsilon(t)\downarrow 0 \text{ for certain } t=t_k\downarrow 0, x_t\to x, r(t)\downarrow 0.\end{aligned} \tag{6.14}$$

The same condition appears for $0\in D^*g(x)(1)$ for the given g. This verifies (i).

(ii) We show first $(\Leftarrow)$. Let

$$\lim_{\xi\to x}\sup_{u\in B}(\langle u^*,u\rangle + \sup\langle v^*, Cf(\xi)(u)\rangle)\le 0,$$

for certain $\xi\to x$. The contingent derivatives fulfill by definition

$$Cg(\xi)(u) = \langle u^*,u\rangle + \langle v^*, Cf(\xi)(u)\rangle,$$

hence $c(\xi) := \sup_{u\in B}\sup Cg(\xi)(u)$ satisfies $\lim_{\xi\to x} c(\xi)\le 0$. For fixed ξ, we may estimate

$$g(x)\le g(\xi)+c(\xi)\|x-\xi\| + o(x-\xi),$$

see Lemma A.2. So, given any $\varepsilon(\xi)\downarrow 0$, one finds $r(\xi)>0$ such that

$$s^{-1}[g(\xi+su)-g(\xi)]\le c(\xi)+\varepsilon(\xi) \qquad \forall u\in B \text{ and } 0<s\le r(\xi).$$

This ensures, for certain $\xi\to x$,

$$\lim_{\xi\to x}\sup_{u\in B,0<s\le r(\xi)}\ s^{-1}[g(\xi+su)-g(\xi)]\le 0,$$

i.e., $0\in D^*g(1)$. From (i), we thus obtain the assertion $u^*\in D^*f(x)(v^*)$.
(ii) $(\Rightarrow)$. Using

$$\sup\langle v^*, Cf(x_t)(u)\rangle \le \sup_{0<s\le r(t)}\ s^{-1}\langle v^*, f(x_t+su)-f(x_t)\rangle$$

which is true (for all $r(t)>0$) by definition of Cf, we obtain from (6.14)

$$\begin{aligned}&\sup_{u\in B}(\langle u^*,u\rangle + \sup\langle v^*, Cf(x_t)(u)\rangle)\\ &\le \sup_{u\in B,0<s\le r(t)}\langle u^*,u\rangle + s^{-1}\langle v^*, f(x_t+su)-f(x_t)\rangle \le \varepsilon(t)\end{aligned}$$

with $\varepsilon(t)\downarrow 0$, again for certain sequences of $t=t_k\downarrow 0, x_t\to x, r(t)\downarrow 0$. Setting $\xi = x_t$, the latter yields the assertion. □

6.3 Relations between Tf and ∂f

Again, let $f:\mathbb{R}^n\to\mathbb{R}^m$ be locally Lipschitz.

Generalized Jacobians

We briefly recall the needed definition, see Chapter 1:

$$\Theta = \{x \mid Df(x) \text{ exists as Fréchet derivative}\},$$

$$\partial_\circ f(x) = \{A \mid A = \lim Df(x^k) \text{for certain } x^k \to x \text{ in } \Theta\}$$

and

$$\partial f(x) = \operatorname{conv} \partial_\circ f(x)$$

is the *generalized Jacobian* of f at x in Clarke's sense.

Since $\operatorname{gph} \partial_\circ f$ is closed (in finite dimension), $\partial f : \mathbb{R}^n \to \mathbb{R}^{mn}$ is also a closed mapping, by Caratheodory's theorem concerning convex hulls. We show that

$$\partial_\circ f(x)u \subset Tf(x)(u).$$

Let $A \in \partial_\circ f(x)$. Considering points $x^k \to x$ such that $x^k \in \Theta$ and $Df(x^k) \to A$, one finds $t_k > 0$ such that

$$f(x^k + t_k u) - f(x^k) = t_k Df(x^k)u + r^k, \text{ where } \|r^k\| < t_k/k.$$

Hence,

$$\lim Df(x^k)u = A\,u \in Tf(x)(u).$$

After replacing $\partial_\circ f(x)$ by its convex hull, one obtains $\partial f(x)u \subset \operatorname{conv}(Tf(x)(u))$. Further, if $A \in \partial f(x)$ is an *exposed matrix* (i.e., A is not a proper convex combination of elements in $\partial f(x)$), then $A \in \partial_\circ f(x)$, and, consequently, the related set fulfills

$$(\operatorname{ex} \partial f(x))u \subset \partial_\circ f(x)u \subset Tf(x)(u) \quad \text{and} \quad \partial f(x)u \subset \operatorname{conv} Tf(x)(u), \tag{6.15}$$

where the symbol "ex" refers to the set of exposed elements.

Conversely, the (deeper) relation

$$Tf(x)(u) \subset \partial f(x)u. \tag{6.16}$$

holds. To verify this inclusion, one may apply the mean value theorem

$$f(x^k + t_k u) - f(x^k) \in \operatorname{conv}[\cup_{0 \le \theta \le 1} \partial f(x^k + \theta t_k u)(t_k u)],$$

shown in [Cla83] and the fact that ∂f is closed and locally bounded. Moving $x^k \to x$ and $t_k \downarrow 0$ here, and taking into account that (obviously)

$$t_k^{-1} \partial f(x^k + \theta t_k u)(t_k u) = \partial f(x^k + \theta t_k u)(u),$$

the inclusion (6.16) follows immediately.

Theorem 6.6 (Tf and generalized Jacobians)

$$\partial f(x)u = \text{conv}\,(Tf(x)(u)). \tag{6.17}$$

◇

Proof. Now, the assertion is a consequence of (6.15) and (6.16) because $\partial f(x)u$ is convex. □

The inclusion (6.16) may be strict even for piecewise linear functions $f : \mathbb{R}^2 \to \mathbb{R}^2$, see Example BE.3.

For $m = 1$, (6.17) shows $\partial f(x)u = Tf(x)(u)$ because $Tf(x)(u)$ is convex as a connected subset in R.

The listed properties concerning Tf, including also chain rules for composed functions, have been shown basically in [Thi80], while (6.7) and (6.17) were proved in [Kum91b]. Concerning properties of ∂f we refer to [Cla83], and concerning Cf to [AE84].

6.4 Chain Rules of Equation Type

6.4.1 Chain Rules for Tf and Cf with f ∈ $C^{0,1}$

In what follows we derive chain rules for functions in finite dimension and impose the **general assumptions:**

$$\begin{aligned}
&g : \mathbb{R}^n \to \mathbb{R}^m,\\
&f : \mathbb{R}^m \to \mathbb{R}^p \text{ is locally Lipschitz},\\
&h = f(g(\cdot)) \text{ and } y = g(x).
\end{aligned}$$

Then one has

$$\begin{aligned}
Th(x)(u) &\subset Tf(y)(Tg(x)(u)) := \cup_{v \in Tg(x)(u)} Tf(y)(v),\\
Ch(x)(u) &\subset Cf(y)(Cg(x)(u)) := \cup_{v \in Cg(x)(u)} Cf(y)(v).
\end{aligned} \tag{6.18}$$

The proof is elementary: Let $w \in Th(x)(u)$ and $w = \lim t^{-1}[h(x_t+tu) - h(x_t)]$ for certain $t \downarrow 0$ and $x_t \to x$. One can select a subsequence such that $v_t := t^{-1}[g(x_t + tu) - g(x_t)]$ has a limit $v \in Tg(x)(u)$. Setting now $y_t = g(x_t)$, one obtains $y_t \to y$ and $w = \lim\ t^{-1}[f(y_t + tv_t) - f(y_t)]$, i.e. $w \in Tf(y)(v)$.

If $x_t = x$ and $y_t = y$ are fixed, our arguments remain valid, so one obtains the assertion for Ch, too. □

For the reverse inclusions, one needs extra assumptions, because elements $v \in Cg(x)(u)$ and $w \in Cf(y)(v)$ may require different sequences $t \downarrow 0$ for the related limit representations. If the limits do not depend on the particular sequences, the difficulties vanish. So we leave the following chain rule as an

Exercise 12. Verify the following stronger versions of (6.18):

$$\begin{array}{ll} \text{(i)} & \text{If } f \text{ or } g \text{ is directionally differentiable,} \\ & \text{then } Ch(x)(u) = Cf(y)(Cg(x)(u)). \\ \text{(ii)} & \text{If } f \in C^1, \text{ then } Th(x)(u) = Df(y)(Tg(x)(u)). \\ \text{(iii)} & \text{If } g \in C^1 \text{ and } g^{-1} \text{ is l.s.c. at } (y,x), \\ & \text{then } Th(x)(u) = Tf(y)(Dg(x)(u)). \end{array} \tag{6.19}$$

◇

If f and g are not continuously differentiable, then it still holds:

$$\begin{array}{l} \text{If } g \text{ is pseudo-regular at } (x,y), \\ \text{then } Tf(y)(v) \subset Th(x)(Tg^{-1}(y,x)(v)). \end{array} \tag{6.20}$$

Proof. Let $w \in Tf(y)(v)$ be written as $w = \lim t^{-1}[f(y_t + tv) - f(y_t)]$, with some $t \downarrow 0$, $y_t \to y$. Since g is pseudo-regular, we find first certain $x_t \in g^{-1}(y_t)$ tending to x, and next related points $x'_t \in g^{-1}(y_t + tv)$ satisfying a Lipschitz estimate $\|x'_t - x_t\| \le Lt$. Thus, the points $a' := t^{-1}(x'_t - x_t)$ form a sequence with some cluster a. Then $a \in Tg^{-1}(y,x)(v)$ and, due to $w = \lim t^{-1}[f(g(x'_t)) - f(g(x_t))]$ and $x'_t = x_t + ta'$, it holds also $w \in Th(x)(a)$ as required. □

Note. We have shown that $w \in Tf(y)(v)$ belongs to $Th(x)(a)$ for some $a \in Tg^{-1}(y,x)(v)$, and we already know that $v \in Tg(x)(u) \Leftrightarrow u \in Tg^{-1}(y,x)(v)$. Hence, if $Tg^{-1}(y,x)$ is *single-valued*, then $a = u$ and

$$Th(x)(u) = \cup_{v \in Tg(x)(u)} Tf(y)(v)$$

hold true. ◇

However, even if g is linear, the reverse inclusion $\supset$ may fail to hold for Th in (6.18).

Example 6.7 (chain rule, counterexample). Let $g(x) = (x,0)$ and $h(x) := f(g(x))$, where

$$f(y_1,y_2) = \begin{cases} 0 & \text{if } y_1 \le 0; \\ y_1 & \text{if } 0 \le y_1 \le |y_2| \\ |y_2| & \text{otherwise.} \end{cases}$$

Then $f(y_1,0) \equiv 0$, $h(x) \equiv 0$, $Th(0)(1) = \{0\}$, but $1 \in Tf(g(0))(1,0)$. ◇

Partial Thibault Derivatives

The next chain rules appear to be the key for our later applications. We consider

$$f(x,y) = h(x,g(y)); \quad h : \mathbb{R}^{n+q} \to \mathbb{R}^p,\ g : \mathbb{R}^m \to \mathbb{R}^q,\ f : \mathbb{R}^{n+m} \to \mathbb{R}^p.$$

Again, h and g are supposed to be locally Lipschitz. We are interested in the formula

$$Tf(x,y)(u,v) = T_x h(x,g(y))(u) + T_g h(x,g(y))(Tg(y)(v)), \tag{6.21}$$

where $T_x h$ and $T_g h$ denote the *partial T-derivatives,* defined - as usually - by fixing the remaining arguments. In general, (6.21) is not true, we need a special property of g.

"Simple" Lipschitz Functions

A locally Lipschitz function $g : \mathbb{R}^m \to \mathbb{R}^q$ is said to be *simple* at y if, for all $v \in \mathbb{R}^m$, $w \in Tg(y)(v)$ and each sequence $t_k \downarrow 0$, there is a sequence $y^k \to y$ such that

$w = \lim t_k^{-1}[g(y^k + t_k v) - g(y^k)]$ holds
at least for some subsequence of $k \to \infty$.

It is remarkable that neither all functions $g \in C^{0,1}(\mathbb{R}, \mathbb{R}^2)$ nor all PC^1**-functions** g are simple. On the other hand, all $g \in C^{0,1}(\mathbb{R}^m, \mathbb{R})$ are simple (cf. [Kum91b]). Further simple functions are $y \mapsto y^+$ and $y \mapsto (y^+, y^-)$, see Lemma 7.4 below. Detailed investigations of simple functions and relations to the following chain rule may be found in [Fus94].

Theorem 6.8 (partial derivatives for Tf). *Let g and h be locally Lipschitz, $f = h(x,g(y))$, and let $D_g h(\cdot,\cdot)$ exist and be locally Lipschitz, too. Then*

$$Tf(x,y)(u,v) \subset T_x h(x,g(y))(u) + T_g h(x,g(y))(Tg(y)(v)).$$

Let, additionally, g be simple at y. Then the equation (6.21) holds true. Moreover, given any

$$a \in T_x h(x,g(y))(u), \quad q \in Tg(y)(v) \quad \textit{and} \quad b \in T_g h(x,g(y))(q),$$

there are sequences $x^k \to x, y^k \to y$ and $t_k \downarrow 0$ such that

$$\begin{array}{lll} a+b & = \lim t_k^{-1}[f(x^k + t_k u, y^k + t_k v) - f(x^k, y^k)] & \textit{as well as} \\ q & = \lim t_k^{-1}[g(y^k + t_k v) - g(y^k)], & \\ b & = \lim t_k^{-1}[h(x, g(y^k) + t_k q) - h(x, g(y^k))], & \\ a & = \lim t_k^{-1}[h(x^k + t_k u, g(y)) - h(x^k, g(y))]. & \end{array}$$

◇

Note. Clearly, $T_g h = D_g h$. ◇

Proof of Theorem 6.8. All sequences t_k, s_k, r_k will be supposed to be positive and vanishing. Let $(x^k, y^k) \to (x, y)$.

Proof of "$\subset$": Let $w \in Tf(x,y)(u,v)$ be given by

$$w = \lim w^k \quad \text{and} \quad w^k := t_k^{-1}[f(x^k + t_k u, y^k + t_k v) - f(x^k, y^k)].$$

We put $\xi^k = x^k + t_k u$, $\eta^k := y^k + t_k v$ and analyze the right-hand side of (6.21). The set $T_x h(x, g(y))(u)$ contains the accumulation points of

$$\begin{aligned} a_k &= t_k^{-1}[h(x^k + t_k u, g(y)) - h(x^k, g(y))] \\ &= t_k^{-1}[h(\xi^k, g(y)) - h(x^k, g(y))]. \end{aligned}$$

Since the sequence $\{a_k\}$ is bounded, convergence $a_k \to a$ may be assumed (at least for some subsequence). Next consider

$$q^k = t_k^{-1}[g(\eta^k) - g(y^k)].$$

Again, convergence $q^k \to q \in Tg(y)(v)$ may be assumed (again for some subsequence if necessary). Thus, $T_g h(x, g(y))(q)$ contains the limits of

$$\begin{aligned} b_k &= t_k^{-1}[h(x, g(y^k + t_k q^k) - h(x, g(y^k))] \\ &= t_k^{-1}[h(x, g(\eta^k)) - h(x, g(y^k))]. \end{aligned}$$

Since also $\{b_k\}$ is bounded, now $b_k \to b \in T_g h(x, g(y))(q)$ may be assumed. Therefore, the right-hand side in (6.21) contains $a + b = \lim(a_k + b_k)$. It remains to show that

$$c^k := w^k - a_k - b_k$$

is vanishing. Explicitly, we have

$$\begin{aligned} t_k c^k = \; & [\; h(\xi^k, g(\eta^k)) && - \quad h(x^k, g(y^k)) \;] \\ & -[\; h(\xi^k, g(y)) - h(x^k, g(y)) \;] && - \quad [h(x, g(\eta^k)) - h(x, g(y^k)) \;]. \end{aligned}$$

Adding elements of "vertical groups" as well as $0 = h(x^k, g(\eta^k)) - h(x^k, g(\eta^k))$, this yields

$$\begin{aligned} t_k c^k &= [\; \{h(\xi^k, g(\eta^k)) - h(\xi^k, g(y))\} - \{h(x^k, g(\eta^k)) - h(x^k, g(y))\} \;] \\ &\quad + [\; \{h(x^k, g(\eta^k)) - h(x^k, g(y^k))\} - \{h(x, g(\eta^k)) - h(x, g(y^k))\} \;] \\ &= A + B, \end{aligned}$$

where A and B denote the two squared brackets.

The term

$$A_1 := h(\xi^k, g(\eta^k)) - h(\xi^k, g(y))$$

may be written as

$$A_1 = \int_0^1 D_g h(\xi^k, g(y) + s[g(\eta^k) - g(y)])\, ds.$$

Similarly, we may write the other three differences:

$$A_2 = \int_0^1 D_g h(x^k, g(y) \;+ s[g(\eta^k) - g(y)])\, ds,$$

$$B_1 = \int_0^1 D_g h(x^k, g(y^k) + s[g(\eta^k) - g(y^k)])\, ds,$$

$$B_2 = \int_0^1 D_g h(x, g(y^k) \; + s[g(\eta^k) - g(y^k)])\, ds.$$

Since the derivatives $D_g h$ as well as g are locally Lipschitz, we thus obtain estimates of the form:

$$\|A\| = \|A_1 - A_2\| \le L\|\xi^k - x^k\|\|\eta^k - y\|,$$

$$\|B\| = \|B_1 - B_2\| \le L\|x^k - x\|\|\eta^k - y\|,$$

for some $L > 0$. Finally, recalling that $\|\xi^k - x^k\| = t_k\|u\|$ and $\|\eta^k - y^k\| = t_k\|v\|$, we see that

$$\|c^k\| \le t_k^{-1}(A + B)$$

tends to zero, indeed.

Proof of "$\supset$". Suppose that g is simple. Now let $a + b$ be any element of the right-hand side in (6.21),

$$a \in T_x h(x, g(y))(u), \quad b \in T_g h(x, g(y))(q) \text{ with } q \in Tg(y)(v).$$

We have to verify that there are sequences $x^k \to x$, $y^k \to y$ and $t_k \downarrow 0$ such that

$$a + b = w := \lim t_k^{-1}[f(x^k + t_k u, y^k + t_k v) - f(x^k, y^k)].$$

The limit expressions of q, a and b will be a by-product of the construction. Due to our assumptions we may write

$$a = \lim a^k,\ a^k = t_k^{-1}[h(x^k + t_k u, g(y)) - h(x^k, g(y))],$$

$$b = \lim b^k,\ b^k = r_k^{-1}[h(x, g^k + r_k q) - h(x, g^k)],$$

with $g^k \to g(y)$. Since g is simple, $q \in Tg(y)(v)$ can be written as a limit (of a subsequence), where the already given sequence of t_k occurs:

$$q = \lim q^k,\ q^k = t_k^{-1}[g(y^k + t_k v) - g(y^k)],$$

provided that $y^k \to y$ has been suitably taken. Notice that

$$g(y^k + t_k v) = g(y^k) + t_k q^k.$$

Using next that $D_g h(\cdot, \cdot)$ exists and is locally Lipschitz, the limit

$$b = \lim b^k = D_g h(x, g(y))\, q$$

does not depend on the selected sequences g^k and r_k. So we may change them and put $g^k = g(y^k)$ and $r_k = t_k$. Additionally, the terms $r_k q$ may be replaced (without changing) by $t_k q^k$ since $q^k \to q$. In this way we obtain $b = \lim \beta^k$ with

$$\beta^k = t_k^{-1}[h(x, g(y^k) + t_k q^k) - h(x, g(y^k))] = t_k^{-1}[h(x, g(y^k + t_k v)) - h(x, g(y^k))].$$

Setting now

$$\xi^k = x^k + t_k u, \ \eta^k = y^k + t_k v$$

as well as

$$w^k = t_k^{-1}[f(\xi^k, \eta^k) - f(x^k, y^k)],$$

we observe that (again for some subsequence)

$$w = \lim w^k \in Tf(x, y)(u, v).$$

Moreover, the elements a_k and b_k of the first part "$\subset$" are now

$$a_k = t_k^{-1}[h(\xi^k, g(y)) - h(x^k, g(y))] = a^k$$

and

$$b_k = t_k^{-1}[h(x, g(\eta^k)) - h(x, g(y^k))] = \beta^k.$$

So we can estimate as above and obtain $w^k - (a^k + \beta^k) \to 0$. This proves the theorem. □

Corollary 6.9 (standard partial derivative). *Suppose that* $f = f(x,y)$ *is locally Lipschitz,* $D_y f(\cdot,\cdot)$ *exists and is locally Lipschitz as well. Then*

$$Tf(x, y)(u, v) = T_x f(x, y)(u) + D_y f(x, y)v.$$

◇

Proof. The function $g(y) = y$ is simple. □

The next conclusion will be our key for dealing with *generalized Kojima functions* in the subsequent chapter.

Corollary 6.10 (product rule). *Let* $F(x,y) = M(x)N(y)$, *where* $M(\cdot)$ *and* $N(\cdot)$ *are locally Lipschitz matrix-valued functions of related size. Suppose that one of them is simple. Then the product rule of differentiation holds for TF, i.e.,*

$$TF(x, y)(u, v) = [TM(x)(u)]N(y) + M(x)[TN(y)(v)].$$

◇

Proof. If, for example, N is simple, put $g = N$ and $h(x, g(y)) = M(x)g(y)$. □

Note that the result of the operation $TM(x)(u)$ is a set S of matrices A having the size of $M(x)$. To get the first set $[TM(x)(u)]N(y)$ of the sum, one has to multiply all $A \in S$ by $N(y)$. Needless to say, Corollary 6.10 holds (by the same arguments) for sums

$$F(x, y) = M(x) + N(y),$$

too. The latter is formally needed if one writes $g(x) \leq 0$ as the equivalent equation $g(x) + y^+ = 0$.

Partial Contingent Derivatives

What about the contingent derivatives under similar assumptions? To study this problem, we must put $x^k \equiv x$ and $y^k \equiv y$ and, of course, replace the derivatives under consideration. We have to begin with the definition of a *simple* locally Lipschitz function g *with respect to* Cg. To be *simple* at y, now means the following:

> For each $v \in \mathbb{R}^m$, $w \in Cg(y)(v)$ and every sequence $t_k \downarrow 0$, there holds $w = \lim t_k^{-1}[g(y + t_k v) - g(y)]$ at least for some subsequence of t_k.

This is nothing else but the existence of the directional derivative at y, i.e.,

$$Cg(y)(v) = \{g'(y; v)\}.$$

The proof of Theorem 6.8 now remains valid step by step in the present context, it becomes only shorter due to the fixed sequences $(x^k, y^k) = (x, y)$, and the notion *simple at* y *with repect to* Cg may be replaced by *directionally differentiable at* y. The results are the following analogous statements.

Theorem 6.11 (partial derivatives for Cf). *Let g and h be locally Lipschitz, $f(x, y) = h(x, g(y))$, and suppose $D_g h(\cdot, \cdot)$ exists and is locally Lipschitz, too. Then*

$$Cf(x, y)(u, v) \subset C_x h(x, g(y))(u) + D_g h(x, g(y))(Cg(y)(v)).$$

If, additionally, g is directionally differentiable at y, then the inclusion holds as equation:

$$Cf(x, y)(u, v) = C_x h(x, g(y))(u) + D_g h(x, g(y)) g'(y; v).$$

◇

As a product rule this yields

Corollary 6.12 *Let $F(x, y) = M(x)N(y)$, where $M(\cdot)$ and $N(\cdot)$ are locally Lipschitz matrix-valued functions of related size. Suppose that one of them is directionally differentiable. Then*

$$CF(x, y)(u, v) = [CM(x)(u)]N(y) + M(x)[CN(y)(v)].$$

◇

6.4.2 Newton Maps and Semismoothness

Newton Functions

Let $f : X \to Y$ be any function and X, Y be normed spaces. If f is *continuously* differentiable and x^* is fixed, the two approximations

$$f(x) - f(x^*) = Df(x^*)(x - x^*) + o_1(x - x^*) \qquad \text{and}$$
$$f(x) - f(x^*) = Df(x\ )(x - x^*) + o_2(x - x^*)$$

may replace each other because both, o_1 and o_2 satisfy $o_k(u)/\|u\| \to 0$.

For $f(x) = x^2 \sin(1/x)$ if $x \neq 0$ and $f(0) = 0$, o_1 exists but o_2 does not so. For $f(x) = |x|$, the reverse situation occurs. When applying solution methods, we need (or have) Df at points x *near* a solution x^*. So the second approximation becomes important and, if $f \notin C^1$, the condition must be specified for multivalued derivatives.

Let $Rf : X \to \text{Lin}(X,Y)$ be locally bounded. We say that Rf is a *Newton function of* f at x^* if

$$Rf(x^* + u)u \in f(x^* + u) - f(x^*) + o(u)B. \tag{6.22}$$

Our notation will be motivated by Newton's method, see Lemma 10.1. At this moment, one may regard the actual property as being a generalization of continuous differentiability for nonsmooth functions.

Notice that in (6.22), $o(u) \geq 0$ may be replaced by the u.s.c. function

$$o_{\sup}(u) := \limsup_{u' \to u} o(u') \geq o(u) \tag{6.23}$$

without violating this condition. So $o(\cdot)$ may be supposed to be u.s.c. (or continuous as well).

Further, the function Rf may be arbitrary at x^* and is not uniquely defined at $x \neq x^*$. If Rf satisfies (6.22), then it is also a Newton function of g at x^*, whenever $g(x) = f(x) + o(x - x^*)$. Here, $g - f$ is not necessarily small in the $C^{0,1}$ - norm, cf. (4.3).

Newton functions at x^* are (single-valued) selections of locally bounded maps $M : X \rightrightarrows \text{Lin}(X,Y)$ such that

$$\emptyset \neq M(x^* + u)u := \{Au \mid A \in M(x^* + u)\} \subset f(x^* + u) - f(x^*) + o(u)B. \tag{6.24}$$

Accordingly, we call M a *Newton map* of f at x^*.

This property is obviously invariant if one forms the union or the convex hull of two Newton maps (the set on the right-hand side of (6.24) is convex).

Example 6.13 (examples of Newton functions).
1. If $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ and B^{nm} denotes the unit ball of (n,m)-matrices, then

$$M(x) = Df(x) + \|f(x) - f(x^*)\| B^{nm}$$

is a Newton map at x^*.

Indeed, since $f \in C^1$ is locally Lipschitz with rank L_f, every matrix $A_u \in \|f(x^* + u) - f(x^*)\| B^{nm}$ satisfies

$$\|A_u\| \leq L_f \|u\|.$$

Thus, $(Df(x^* + u) + A_u)u = Df(x^* + u)u + A_u u$ with $\|A_u u\| \leq L_f \|u\|^2$, and we may write

$$(Df(x^* + u) + A_u)u = f(x^* + u) - f(x^*) + o(u) + r(u),$$

where $\|r(u)\| \leq L_f\|u\|^2$. Using the new o-type function

$$\|o(u)\| + L_f\|u\|^2$$

in (6.24), this is the assertion.

2. For $f = PC^1[f^1, ..., f^N]$ and $f(x^*) = 0$ one may put

$$M(x) = \{Df^i(x) | i \in J(x)\}, \quad \text{where } J(x) = \{i \mid \|f^i(x) - f(x)\| \leq \|f(x)\|^2\}.$$

Indeed, for $\|u\|$ sufficiently small, the index sets fulfill $J(x^* + u) \subset J(x^*)$. Thus, with some Lipschitz rank L of f near x^* , we obtain

$$\begin{array}{lllll} f(x^* + u) - f(x^* - Df^i(x^* + u)u & & & & \\ \in & f^i(x^* + u) - f^i(x^*) - Df^i(x^* + u)u & + & \|f(x^* + u)\|^2 B \\ \subset & o_i(u)B & + & L^2\|u\|^2 B. \end{array}$$

Hence, $o(u) = L^2\|u\|^2 + \max_i o_i(u)$ satisfies (6.24). ◇

If $\dim X + \dim Y < \infty$, then, due to

$$f(x^* + u) - f(x^*) \subset Cf(x^*)(u) + o(u)B,$$

cf. Lemma A2, one easily confirms that (6.24) implies (with a possibly new o-type function),

$$\begin{array}{rll} M(x^* + u)u & \subset & Cf(x^*)(u) + o(u)B \\ & \subset & Tf(x^*)(u) + o(u)B \\ & \subset & \partial f(x^*)(u) + o(u)B. \end{array} \tag{6.25}$$

However, f is not necessarily directionally differentiable (see the next theorem), and M does not have a so-called approximate Jacobian [JLS98] which would require

$$\langle y^*, Cf(x)(u)\rangle \subset \text{conv}\,\langle y^*, M(x + u)u + o(u)B\rangle \;\forall y^* \in \text{bd}\, B_{Y^*}.$$

Surprisingly, condition (6.22) is a weak one, and Newton functions satisfy a common chain rule.

Theorem 6.14 (existence and chain rule for Newton functions).

(i) Every locally Lipschitz function $f : X \to Y$ (X, Y Banach spaces) possesses, at each x^*, a Newton function Rf being locally bounded by a Lipschitz constant L for f near x^*.

(ii) Let $h : X \to Y$ and $g : Y \to Z$ be $C^{0,1}$ with Newton functions Rh at x^* and Rg at $h^* := h(x^*)$. Then $Rf(x) = Rg(h(x))Rh(x)$ defines a Newton function of $f(\cdot) = g(h(\cdot))$ at x^*.

◇

Proof. (i) Given $u \in X\backslash\{0\}$, there is a linear operator $\Phi_u : X \to Y$ with

$$\Phi_u(u) = f(x^* + u) - f(x^*).$$

By Hahn-Banach arguments, Φ_u even exist with bounded norm

$$\|\Phi_u\| \le \|f(x^* + u) - f(x^*)\|/\|u\|.$$

For small $\|u\|$, this yields $\|\Phi_u\| \le L$. So it suffices to put $Rf(x^* + u) = \Phi_u$ and $o(u) = 0$.

(ii) The straightforward proof is basically the same as for Fréchet derivatives. We put $v = h(x^* + u) - h(x^*)$ and $x = x^* + u$. Our assumptions yield

$$v = Rh(x)u + r_h, \quad \text{where } r_h \in o_h(u)B_Y,$$

and

$$g(h^* + v) - g(h^*) = Rg(h^* + v)v + r_g, \quad \text{where } r_g \in o_g(v)B_Z.$$

Thus,

$$\begin{aligned} f(x^* + u) - f(x^*) &= g(h^* + v) - g(h^*) \\ &= Rg(h^* + v)v + r_g \\ &= Rg(h(x))v + r_g \\ &= Rg(h(x))Rh(x)u + Rg(h(x))r_h + r_g \\ &= Rf(x)u + Rg(h(x))r_h + r_g. \end{aligned}$$

Now $Rg(h(x))r_h$ is of type $o(u)$ since $Rg(h(x))$ is uniformly bounded for x near x^*. If $v = 0$ then $r_g = 0$. Otherwise, we obtain from

$$\|v\| = \|h(x^* + u) - h(x^*)\| \le L_h\|u\|$$

that $o_g(v)/\|u\| = (o_g(v)/\|v\|)(\|v\|/\|u\|)$ vanishes as $\|u\| \downarrow 0$. Hence $f(x^* + u) - f(x^*) \in Rf(x^* + u)u + o(u)B_Z$. □

By Theorem 6.14, it turns out that, having Newton maps Mg and Mh at the related points $h(x^*)$ and x^*, then the canonically composed map $M = Mg(h(\cdot))Mh(\cdot)$ is a Newton map for the composed function $f = g \circ h$. However, the function Rf, defined under (i) in the previous theorem, does not use local behavior of f near x and depends on x^* which is often an unknown solution. So one cannot directly apply statement (i) of Theorem 6.14 for solution methods since first one has to find Rf satisfying (6.22) *without using* x^*.

Nevertheless, having Rf, it can be applied for Newton's method exactly like Df under the usual boundedness condition of the inverse, see Section 10.1. To investigate convergence of Newton's method for $f \in C^{0,1}(X, Y)$, maps M satisfying (6.24) and particular realizations have been considered in [Kum88b].

Semismoothness

A function $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is said to be *semismooth* at x^* if $M = \partial f$ is a Newton map at x^*. This notion, based on Mifflin [Mif77], has been introduced and used for Newton's method by [PQ93] and [QS93] and in many subsequent papers. Some modifications of semismoothness are mentioned in Section 10.1.

One well-known class of semismooth functions is the class $PC^1[f^1, ...f^N]$ since

$$\partial f(x) \subset \mathbf{conv}\,\{Df^i(x)|f^i(x) = f(x)\}$$

and each f^i is trivially semismooth (everywhere). Another class consists of certain NCP functions, which will be considered in Section 9.2. The real, globally Lipschitz function h in Example BE.0 is nowhere semismooth.

Before showing how Newton maps may be applied to the class locPC^1, defined below, we recall conditions for semismoothness given by [Mif77, Prop. 3, Thm. 2].

Theorem 6.15 (semismoothness; Mifflin). *Convex functions* $f : \mathbb{R}^n \to \mathbb{R}$ *and maximum functions* $f(x) = \max_{y\in Y} g(x, y)$ *over compact sets Y are semismooth, provided that* g *is continuous and* $D_x g(\cdot, \cdot)$ *exists and is continuous, too.* ◇

Proof. We present a proof for completeness and in order to show how the inclusion $I(x) \subset I(x^*)$ which holds for x near x^* in the case of a finite set Y (where $I(x) := \{y \in Y | f(x) = g(x, y)\}$) will be modified in the current case of a compact index set Y. For seek of simplicity, let $x^* = 0$.

(i) Let f be convex. The inclusion

$$\langle y_x, x\rangle \in [f(x) - f(0) - o(x)\ ,\ f(x) - f(0) + o(x)]$$

must be shown for all $y_x \in \partial f(x)$. By the definition of ∂f, we already have

$$f(0) - f(x) \geq \langle y_x, -x\rangle, \text{ i.e., } \langle y_x, x\rangle \geq f(x) - f(0).$$

Since ∂f is u.s.c., it holds $\partial f(x) \subset \partial f(0) + O(x)B$. Thus, one finds some $y_{0x} \in \partial f(0)$ such that $\|y_{0x} - y_x\| \leq O(x)$. This ensures

$$\langle y_x, x\rangle \leq \langle y_{0x}, x\rangle + O(x)\|x\| \leq f(x) - f(0) + o(x)$$

as reqired.

(ii) For the maximum function, it holds

$$\partial f(x) = \mathbf{conv}\,\{D_x g(x, y_x)|y_x \in \Phi(x)\}$$

where $\Phi(x) = \mathbf{argmax}_{y\in Y}\ g(x, y)$. The semismoothness condition becomes by definition

$$D_x g(x, y_x)x \in f(x) - f(0) + o(x)B\ \forall y_x \in \Phi(x). \tag{6.26}$$

We will see that this is equivalent to

$$\inf\{g(0,y_x) | y_x \in \Phi(x)\} \geq f(0) - o(x) \tag{6.27}$$

which is just the mentioned inclusion $I(x) \subset I(0)$ for x near x^* if Y is finite.

Using compactness of Y and continuity of $D_x g$ one estimates uniformly

$$D_x g(x,y_x)x \in g(x,y_x) - g(0,y_x) + o(x)B. \tag{6.28}$$

By upper semicontinuity of Φ (cf. Theorem 1.15), one has $\Phi(x) \subset \Phi(0)+O(x)B$. So let $y_{0x} \in \Phi(0)$ fulfill $\|y_x - y_{0x}\| \leq O(x)$. Notice that

$$g(0,y_x) \leq g(0,y_{0x}) \text{ and } g(x,y_x) \geq g(x,y_{0x}). \tag{6.29}$$

Thus, taking (6.26) and (6.28) into account, we have to show that (6.27) is valid, indeed. Suppose contrarily that

$$g(0,y_x) \leq g(0,y_{0x}) - c\|x\|, \ c > 0$$

holds for certain $x \to 0$, and related y_x, y_{0x}. Via the mean-value theorem, we obtain points z_x, z_{0x} (between 0 and x) satisfying

$$\begin{aligned} g(x,y_x) &= g(0,y_x) + D_x g(z_x,y_x)x \\ g(x,y_{0x}) &= g(0,y_{0x}) + D_x g(z_{0x},y_{0x})x. \end{aligned}$$

It follows

$$g(0,y_x) + D_x g(z_x,y_x)x \geq g(0,y_{0x}) + D_x g(z_{0x},y_{0x})x,$$

$$g(0,y_{0x}) - c\|x\| + D_x g(z_x,y_x)x \geq g(0,y_{0x}) + D_x g(z_{0x},y_{0x})x$$

and

$$D_x g(z_x,y_x)x \geq D_x g(z_{0x},y_{0x})x + c\|x\|.$$

Passing to a subsequence if necessary, there exists a common accumulation point y_0 of y_x and y_{0x}. So, with some accumulation point u of $x/\|x\|$, $x \to 0$, we finally arrive at a contradiction

$$D_x g(0,y_0)u \geq D_x g(0,y_0)u + c.$$

□

Note. In consequence, the function $f_Y(x) = \min\{\|x-y\|^2 \mid y \in Y\}$ is semismooth for Euclidean norm and compact, non-empty $Y \subset \mathbb{R}^n$. Further, each *DC*-functional f (difference of convex functions) is semismooth. The same holds (by Theorem 6.14), if $f : \mathbb{R}^n \to \mathbb{R}^m$ has *DC* components since $\emptyset \neq \partial f(x) \subset (\partial f_1(x), ..., \partial f_m(x))$. ◇

Pseudo-Smoothness and D°f

We call $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ *pseudo-smooth* if f is a C^1-function on an open and dense subset $\Omega \subset \mathbb{R}^n$. These functions appear in many applications, cover the class PC^1 by Lemma 6.17, have locally bounded derivatives on Ω, and obey nonempty sets

$$D^\circ f(x) := \limsup_{y \to x, y \in \Omega} \{Df(y)\} \quad \text{(as upper Hausdorff-limit).} \tag{6.30}$$

Nevertheless, in Example BE.6, §12.1, we present a real convex function that is not pseudo-smooth.

Let $\Theta^1(f)$ be the set of all C^1 points of f. It makes no difficulties to see that $D^\circ f(x)$ does not depend on the choice of Ω in (6.30). One could even replace Ω by any dense subset Ω_d of $\Theta^1(f)$, in (6.30).

In addition, it holds

$$D^\circ f(x) = Df(x) \;\; \forall x \in \Theta^1(f). \tag{6.31}$$

The single-valued selections of $D^\circ f$ are natural candidates for being Newton functions, because $D^\circ f(\cdot) \subset M(\cdot)$ holds necessarily for all closed maps M satisfying

$$M(x) = \{Df(x)\} \;\; \forall x \in \Theta^1(f),$$

hence also for all closed Newton maps M which assign, to $x \in \Theta^1(f)$, as usually, the Jacobian $M(x) = \{Df(x)\}$.

In order to check whether $D^\circ f$ is a Newton map for a pseudo-smooth function f at x^*, it suffices to consider all points $x^* + u$ in a dense subset Ω_d of $\Theta^1(f)$, and to investigate whether

$$f(x^* + u) - f(x^*) - Df(x^* + u)u \;\in\; o(u)B \tag{6.32}$$

holds true. In this case, the contingent derivative can be estimated by $D^\circ f(x^*)$, too.

Lemma 6.16 (selections of $D^\circ f$). *If f is pseudo-smooth and some selection Rf of $D^\circ f$ is a Newton function, then $M = D^\circ f$ is a Newton map at the same fixed x^*, and it holds*

$$Cf(x^*)(u) \subset D^\circ f(x^*)u. \tag{6.33}$$

◇

Proof. Let $Sf \in D^\circ f$ be a second selection. By (6.31), Sf satisfies (6.22) for pouits $x^* + u' \in \Omega$. If $o(\cdot)$ in (6.22) is not u.s.c., we replace it by $o_{\sup}(\cdot)$ in (6.23) which is again of little-o-type. Then, since each $x^* + u$ is a limit of elements in Ω, condition (6.22) holds by continuity arguments for $u' \to u$ at $x^* + u$, too. Thus, every selection of $D^\circ f$ fulfills (6.22), so (6.24) is true.
Inclusion (6.33): For small $t > 0$, due to pseudo-smoothness, the quotients

$$a(t) := t^{-1}[f(x^* + tu) - f(x^*)]$$

can be approximated (with error $< t$) by

$$b(t) := t^{-1}[f(x^* + tu(t)) - f(x^*)],$$

such that $\|u(t) - u\| < t$ and $x^* + tu(t) \in \Omega$. Then (6.31) and (6.22) guarantee that $b(t) \in Df(x^*+tu(t))u(t)+t^{-1}o(tu(t))B$, which yields the assertion because $\lim b(t) = \lim a(t) \in D^\circ f(x^*)u$ as $t \downarrow 0$. □

In Example BE.1, f is pseudo-smooth and directionally differentiable with $D^\circ f(x^*) \neq \partial_\circ f(x^*)$, (6.33) fails to hold though $Df(x^*)$ exists, and neither $D^\circ f$ nor $\partial_\circ f$ contain a Newton function at $x^* = 0$. Nevertheless, there are pseudo-smooth functions outside PC^1 such that $D^\circ f$ is always a Newton map.

Locally PC^1 Functions

Let f be pseudo-smooth. We call f *locallyPC^1* (and write $f \in \text{loc}PC^1$) if there is an open and dense subset $\Omega \subset \mathbb{R}^n$ such that f is C^1 on Ω and the following holds:

There exists a finite family of open sets $U^s \subset \mathbb{R}^n$ and continuous functions $f^s : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

(i) f^s is C^1 on U^s and $Df^s(\cdot)$ is uniformly continuous on $U^s \cap K$ for each bounded set K, *and*

(ii) for each $x \in \mathbb{R}^n$, there exists an $r = r(x) > 0$ such that, given $y \in \Omega \cap (x + rB)$, one finds some s with $\text{rel int conv}\{x, y\} \subset U^s$, $f^s(x) = f(x)$, $f^s(y) = f(y)$ and $Df^s(y) = Df(y)$.

In comparison with (proper) PC^1 functions, *we do not claim* that f^s is C^1 on the whole space. The set Ω in the previous definition will be also called *C^1-set of f*.

Lemma 6.17 (special locally PC^1 functions). *The Euclidean norm of a linear function $f(y) = \|Ay\|$ and all functions $f \in PC^1$ are locally PC^1.*

A pseudo-smooth function f is locally PC^1 if there is a finite covering $\{P^s | s = 1, ..., N\}$ of $\mathbb{R}^n$ by convex polyhedra P^s such that f is C^1 and Df is uniformly continuous on $\text{int}\,P^s$.

In addition, if g and h are locally PC^1 and $\Phi \in C^1$, then $f(x) = \Phi(g(x), h(x))$ is again locally PC^1 (provided that g, h, Φ are of appropriate dimension). ◇

Proof. *Euclidean Norm:* If $A \neq 0$, put $\Omega = \mathbb{R}^n \setminus \ker A, U^1 = \Omega, f^1 = f, r = 1$ if $x \in \ker A$ and $r = \frac{1}{2}$ dist $(x, \ker A)$ otherwise.

PC^1 : Let $f = PC^1(f^1, ..., f^N)$ and $I(y) = \{s | f^s(y) = f(y)\}$. Note that f^s coincides with f *near* y if $y \in \text{int}\, I^{-1}(s)$. We put

$$U^s = \mathbb{R}^n \text{ and } \Omega = \cup_s \text{ int}\, I^{-1}(s).$$

The set Ω is dense in $\mathbb{R}^n$. We prove this known fact for completeness: Otherwise, one finds some ball $V = y^0 + \varepsilon B$ with $V \cap \Omega = \emptyset$. Then the relatively open sets $N^s = \{x \in V | f^s(x) \neq f(x)\}$ are dense in V, by definition of Ω. So the intersection $D = \cap_s N^s$ is again dense. But $y \in D$ means $f^s(y) \neq f(y)\ \forall s$, which contradicts $f \in PC^1$. Thus, Ω is dense, indeed.

Next, given x, one finds $r > 0$ such that

$$I(y) \subset I(x) \quad \forall y \in x + rB.$$

Now assign some s to $y \in \Omega \cap (x + rB)$ such that $y \in \text{ int } I^{-1}(s)$. Then the equations $f^s(x) = f(x)$ and $f^s(y) = f(y)$ follow from the choice of s and r, and $Df(y) = Df^s(y)$ is valid because f and f^s coincide near y. Finally, $Df^s(\cdot)$ is uniformly continuous on $U^s \cap K$ since Df^s is continuous on $\mathbb{R}^n$.

Covering: Define $f^s = f$, $U^s = \text{int } P^s$, $\Omega = \cup U^s$ and take r small enough such that, for $0 < \varepsilon < r$, the set $S(\varepsilon) := \{s | (x + \varepsilon B) \cap U^s \neq \emptyset\}$ is constant. The existence of r is ensured since all P^s are polyhedral sets.

$\Phi(g(\cdot), h(\cdot))$: With the related sets and radii assigned to g and h, one may put

$$\Omega = \Omega(g) \cap \Omega(h), \quad U^{s\sigma} = U^s(g) \cap U^\sigma(h), \quad f^{s\sigma} = \Phi(g^s, h^\sigma)$$

and $r(x) = \min\{r(x, g), r(x, h)\}$. □

We are now ready to present the motivation for the above definitions.

Theorem 6.18 (Newton maps of $f \in \text{loc}PC^1$). *Let f be a locally PC^1 function and $x^* \in \mathbb{R}^n$. Then*

(i) $M = D^\circ f$ is a Newton map of f at x^*.

(ii) The function $o(\cdot)$ in (6.24) can be taken as $o(u) = \|u\| O(\|u\|)$ provided that both $O(\|u\|)$ is a modulus of uniform continuity for all functions $Df^s(\cdot)$ on U^s near x^* and $O(\cdot)$ is continuous.

(iii) For the composition $f = g(h(x))$ of locally PC^1 functions g and h, the mapping $M(x) = D^\circ g(h(x)) D^\circ h(x)$ is a Newton map of f at x^*.

◇

Note. *Modulus of uniform continuity means* that

$$\|Df^s(x') - Df^s(x'')\| \le O(\|x' - x''\|) \text{ for all } x', x'' \in U^s \text{ near } x^*.$$

If all Df^s are globally Lipschitz on U^s, then $o(u) \le K\|u\|^2$ holds for small $\|u\|$.
◇

Proof of Theorem 6.18.
Proof of (i) and (ii). Given x^* we find some r that defines the ball $x^* + rB$ in the definition of $\text{loc}PC^1$. Let $y = x^* + u \in \Omega \cap (x^* + rB)$. With some s and U^s, f^s according to the definition of $\text{loc}PC^1$, we obtain

$$\text{int conv}\{x^*, y\} \subset U^s, \ f^s(x^*) = f(x^*), \ f^s(y) = f(y), \ Df(y) = Df^s(y).$$

This allows us to integrate and to estimate

$$\begin{aligned}
&f(y) - f(x^*) \\
&= f^s(y) - f^s(x^*) \\
&= \int_0^1 Df^s(x^* + tu)u\,dt \\
&\in \int_0^1 Df^s(y)u\,dt + \|u\| \sup_{0\le t\le 1} \|Df^s(x^* + tu) - Df^s(x^* + u)\|B.
\end{aligned}$$

Due to uniform continuity, the supremum is bounded by $O(\|u\|)$. Using $Df(y) = Df^s(y)$, this guarantees

$$f(x^*+u)-f(x^*)-Df(x^*+u)u \in \|u\|O(\|u\|)B \ \forall x^*+u \in \Omega\cap(x^*+rB). \quad (6.34)$$

Let Rf be any selection of $D^\circ f$. Since $Rf \equiv Df$ on Ω, now (6.22) holds true, provided that $x^* + u$ belongs to the dense subset

$$\Omega \cap (x^* + rB)$$

of $x^* + rB$. Because $Df(x^* + u)$ remains bounded, continuity arguments then yield that (6.34) also holds for the upper Hausdorff limit:

$$\begin{aligned}
&A := \limsup_{u\to u^*} [f(x^* + u) - f(x^*) - Df(x^* + u)u] \subset \|u^*\|O(\|u^*\|)B \\
&\text{if } x^* + u \in \Omega \cap (x^* + rB).
\end{aligned} \quad (6.35)$$

Replacing here Ω by *any* open and dense set $\Omega(f)$ where f is C^1 on, we obtain the same set A on the left side of (6.35) because Df and f are continuous on $\Omega(f)$ and each $x^* + u' \in \Omega(f)$ can be approximated by $x \in \Omega\cap\Omega(f)$ arbitrarily close. Thus, by (6.35) and definition of $D^\circ f$, we obtain for all u^*,

$$f(x^* + u^*) - f(x^*) - D^\circ f(x^* + u^*)u^* \ \subset \ \|u^*\|O(\|u^*\|)B$$

which verifies (i) and (ii).

Proof of (iii), Knowing (i), statement (iii) follows from Theorem 6.14. □

Remark 6.19 Combining (6.25) and (6.33) one obtains upper and lower estimates of Cf for $f \in \text{loc}PC^1$, with certain o-type functions o_1, o_2:

$$D^\circ f(x^* + u)u \subset Cf(x^*)(u) + o_1(u)B \subset D^\circ f(x^*)(u) + o_2(u)B. \quad (6.36)$$

Here, o_1, o_2 will depend on x^*, too.

However, for $f \in PC^1(f^1, .., f^m)$, it becomes obvious that these functions are uniformly bounded for all x^* in some compact set K and $v \in B$, by

$$r(x^*, v) := \sum_k |f^k(x^* + v) - f^k(x^*) - Df^k(x^*)v|$$

as $o_1(v) = \sup\{r(x^*, v) \mid x^* \in K \text{ and } v \in B\}$ and $o_2(v) = 2o_1(v)$. ◇

Exercise 13. Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be strongly regular at $(x^*, 0)$. Show that f^{-1} is semismooth at 0 if f is semismooth at x^*. Note that you *cannot* apply invertibility of the matrices in $\partial f(x^*)$ due to Example BE.3. ◇

6.5 Mean Value Theorems, Taylor Expansion and Quadratic Growth

In this section, we establish the Taylor expansion of a $C^{1,1}$ -**function** f in finite dimension in terms of $T(Df)$ as in the smooth case. We start with the simplest case of a mean value theorem. There holds:

$$\begin{array}{l}\text{If } h \in C^{0,1}(\mathbb{R},\mathbb{R}) \text{ and } h(0)=h(1)\\ \text{then } 0 \in Th(\theta)(1) \text{ for some } \theta \in (0,1).\end{array} \tag{6.37}$$

Notice that (6.37) fails to hold with Ch in place of Th.

Proof of (6.37): Assume first that some $\theta \in (0,1)$ realizes max $h(t)$ on [0,1]. Then

$$(h(t)-h(\theta))(t-\theta) \leq 0 \quad \text{for small } t-\theta > 0$$

yields $\underline{\alpha} \in Th(\theta)(1)$ for some $\underline{\alpha} \leq 0$ and

$$(h(\theta)-h(t))(\theta-t) \geq 0 \quad \text{for } \theta - t > 0$$

yields $\overline{\alpha} \in Th(\theta)(1)$ for some $\overline{\alpha} \geq 0$. Since $Th(\theta)(1)$ is connected, we obtain $0 \in Th(\theta)(1)$. If there is no maximizer of h on [0,1] in (0,1), then $h(0) = h(1)$ implies that there is a minimizer of h on [0,1] in (0,1), and, by similar arguments, one comes to the same inclusion. □

By using a linear transformation and Lemma 6.2, we have the statement:

$$\text{If } h \in C^{0,1}(\mathbb{R},\mathbb{R}), \text{ then } h(t)-h(0) \in Th(\theta)(t) \text{ for some } \theta \in (0,t). \tag{6.38}$$

For $f \in C^{0,1}(\mathbb{R}^n,\mathbb{R}^m)$, a rather large set must be used for estimates. There holds

$$f(x+u)-f(x) \in C := \text{cl conv} \cup_{\theta\in(0,1)} Tf(x+\theta u)(u). \tag{6.39}$$

Proof. Indeed, otherwise one may separate $\{f(x+u)-f(x)\}$ and C, i.e., there is some $v \in \mathbb{R}^m$ with

$$\langle v, f(x+u)-f(x)\rangle < \langle v, c\rangle \quad \forall c \in C.$$

Put $h(t) = \langle v, f(x+tu)\rangle$. Then property (6.38) yields, with some $\theta \in (0,1)$,

$$\langle v, f(x+u)-f(x)\rangle = h(1)-h(0) \in Th(\theta)(1) \subset \langle v, Tf(x+\theta u)(u)\rangle.$$

The last inclusion holds true due to property (6.18), where $\langle v, Q\rangle$ denotes the set $\{\langle v,q\rangle | q \in Q\}$. Considering all $c \in Tf(x+\theta u)(u) \subset C$, one obtains a contradiction. □

Statement (6.39), in terms of ∂f instead of Tf, has been shown and is a main theorem in [Cla83]. Our set C is formally smaller than the related set

$$C' = \text{cl conv} \cup_{\theta\in(0,1)} \partial f(x+\theta u)(u),$$

since $Tf \subset \partial f$. So (6.39) looks even stronger. But recall that, for showing $Tf \subset \partial f$, we had already applied the **∂f-version** of the mean value theorem (based on Rademacher's theorem and Fubini's theorem in [Cla83]). Really, the two versions are equivalent because of the convex-hull operation and Theorem 6.4.

The next theorem extends well-known facts from C^2 to $C^{1,1}$ functionsby using the some devices concerning the proof.

Theorem 6.20 ($C^{1,1}$-Taylor expansion). *Let $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$. Then*

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

holds for same $\theta \in (0,1)$ and some $q \in TDf(x+\theta u)(u)$. ◇

Proof. Without loss of generality let $x = 0$, $f(0) = 0$ and $u \neq 0$. Moreover, replacing f by $\varphi(\cdot) = f(\cdot) + \langle c, \cdot\rangle$, we do not change the statements, since this transformation does not change the set TDf. So we may assume that $f(u) = 0$. Next put $g(t) = f(tu), t \in \mathbb{R}$. Then

$$\begin{array}{l} g(0) = g(1) = 0,\ g'(t) = \langle u, Df(tu)\rangle,\ Df(0)u = g'(0) \\ \textbf{and} \quad Tg'(s)(1) \subset \langle u, T(Df)(su)(u)\rangle. \end{array} \tag{6.40}$$

The last inclusion is justified by (6.18). Due to the transformations (6.40), it suffices to show that

$$-g'(0) \in \tfrac{1}{2}Tg'(\theta)(1) \quad \text{for some } \theta \in (0,1).$$

For this reason we define the real function $r(t) = g(t) + g'(0)(t - \frac{1}{2})^2$. Then

$$\begin{array}{l} r(0) = r(1),\ r'(t) = g'(t) + 2g'(0)(t - \frac{1}{2}),\ r'(0) = 0, \\ \text{and} \quad Tr'(\theta)(1) = Tg'(\theta)(1) + 2g'(0). \end{array}$$

Applying the usual mean value theorem, there is some $s \in (0,1)$ such that

$$0 = r(1) - r(0) = r'(s).$$

Next we use property (6.38) to obtain

$$0 = r'(s) - r'(0) \in Tr'(\theta)(s) \qquad \text{for some } \theta \in (0, s).$$

With our settings, this means that

$$0 \in s\,Tr'(\theta)(1) = s\,(Tg'(\theta)(1) + 2g'(0)),$$

and, since $s > 0$, $-g'(0) \in \frac{1}{2}Tg'(\theta)(1)$ as required. □

For the related statement based on Clarke's generalized Jacobian $\partial[Df]$, we refer to [HUSN84].

Quadratic Growth

The following statements are of particular interest if the subsequent constant c is positive, because then f is locally growing in comparison with its linearization. Nevertheless, c may be any real.

Corollary 6.21 (quadratic growth on a neighborhood). *Let $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$, $U \subset \mathbb{R}^n$ be a cone and c be a constant such that*

$$c < \inf\{\langle q, u\rangle | q \in TDf(x^0)(u)\} \quad \forall u \in U \cap \operatorname{bd} B. \tag{6.41}$$

Then there exist a neighborhood Ω of x^0 and some $\varepsilon > 0$ such that, for all $\xi \in \Omega$,

$$\begin{aligned} &f(x) - f(\xi) \geq Df(\xi)(x-\xi) + \tfrac{1}{2}c\,\|x-\xi\|^2 \\ &\forall x \in \Omega \text{ with } \operatorname{dist}(x-\xi, U) \leq \varepsilon\|x-\xi\|. \end{aligned} \tag{6.42}$$

◇

Note. In particular, one may put $x = \xi + u, u \in U$ (as long as $x, \xi \in \Omega$) in (6.42). Since $TDf(x^0)$ is homogeneous, (6.41) holds also for $-u$, hence for a "double cone". ◇

Proof of Corollary 6.21. The mapping $TDf(\cdot)(\cdot)$ is closed and locally bounded, and $TDf(\xi)(\cdot)$ is Lipschitz in u. So, if $\varepsilon > 0$ is small, (6.41) remains valid for ξ near x^0 and $u \in \operatorname{bd} B$ with $\operatorname{dist}(u, U) \leq \varepsilon$. Applying Theorem 6.20 to ξ and $u = x - \xi$ (and using $\lambda u \in \operatorname{bd} B + \varepsilon B$) now proves the corollary. □

Even for convex $C^{1,1}$ functions, quadratic growth at a point, i.e. (6.42) for $\xi = x^0$ only, does not induce that (6.41) holds with some $c > 0$.

Example 6.22 (counterexample). Take a real, monotone Lipschitz function g such that $g(0) = 0, g(x) \geq x$ and g is constant on intervals of the form $2^{-(1+k)} < x < 2^{-k}$ for odd k. Setting

$$f(t) = \int_0^t g(x)dx,$$

one obtains a convex $C^{1,1}$ function f satisfying (6.42) for $U = R$ and $\xi = x^0 = 0$. Simultaneously, $\inf\{\langle q, u\rangle | q \in TDf(x^0)(u)\} = 0$ holds. ◇

Condition (6.42) for $\xi = x^0$ requires a less strong assumption, hence we may replace TDf by CDf there.

Theorem 6.23 (quadratic growth at a point). *Let $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$, $U \subset \mathbb{R}^n$ be a cone and c be a constant such that*

$$\text{(6.41)'} \qquad c < \inf\{\langle u, q\rangle | q \in CDf(x^0)(u)\} \qquad \forall u \in U \cap \operatorname{bd} B.$$

Then there exist a neighborhood Ω *of* x^0 *and some* $\varepsilon > 0$ *such that (6.42) holds for* $\xi = x^0$, *i.e.,*

$$\text{(6.42)'} \qquad \begin{array}{l} f(x) - f(x^0) \geq Df(\xi)(x - x^0) + \frac{1}{2}c\|x - x^0\|^2 \\ \forall x \in \Omega \text{ with } \operatorname{dist}(x - x^0, U) \leq \varepsilon\|x - x^0\|. \end{array}$$

◇

Proof. We assume $Df(x^0) = 0$, otherwise consider $g(x) = f(x) - Df(x^0)(x - x^0)$. Given any sequence $x \to x^0$ there exists a subsequence such that x can be written as

$$x = x(\lambda) = x^0 + u(\lambda) \text{ where, for certain } \lambda \downarrow 0, u(\lambda) = \lambda u^0 + o(\lambda).$$

If (6.42)' is false -with $\xi = x^0$ - then it does not hold for related x with $u^0 \in U \cap \operatorname{bd} B$. We consider such $x = x(\lambda)$. Let

$$c < \gamma < \inf\{\langle u^0, q\rangle | q \in CDf(x^0)(u^0)\}$$

and

$$\delta(t, \lambda) = Df(x^0 + t\lambda u^0 + to(\lambda)) - Df(x^0 + t\lambda u^0).$$

Since $f \in C^{1,1}$, there is some K depending on f only, such that

$$\|\delta(t, \lambda)\| \leq K\|o(\lambda)\| \text{ for } 0 < t < 1. \tag{6.43}$$

Now it holds

$$\begin{aligned} &f(x(\lambda)) - f(x^0) \\ &= \textstyle\int_0^1 Df(x^0 + tu(\lambda))u(\lambda)dt \\ &= \textstyle\int_0^1 Df(x^0 + t\lambda u^0 + to(\lambda))u(\lambda)dt \\ &= \textstyle\int_0^1 [Df(x^0 + t\lambda u^0) + \delta(t, \lambda)](\lambda u^0 + o(\lambda))dt \\ &= \lambda \textstyle\int_0^1 Df(x^0 + t\lambda u^0)u^0 dt + \int_0^1 [Df(x^0 + t\lambda u^0)o(\lambda) + \delta(t, \lambda)(\lambda u^0 + o(\lambda))]dt. \end{aligned}$$

Due to (6.43) and $Df(x^0) = 0$, the second integral turns out to be of type $\lambda o(\lambda)$. By (6.41)' and by the choice of γ we have, for $\lambda > 0$ sufficiently small and $0 < t < 1$,

$$Df(x^0 + t\lambda u^0)u^0 = [Df(x^0 + t\lambda u^0) - Df(x^0)]u^0 \geq t\lambda\gamma.$$

Therefore,

$$f(x(\lambda)) - f(x^0) \geq \lambda \int_0^1 t\lambda\gamma\, dt - \lambda o(\lambda) = \frac{1}{2}\gamma\lambda^2 - \lambda o(\lambda).$$

Recalling that $\gamma > c$, this shows, for sufficiently small $\lambda > 0$ of the selected sequence, the inequalities

$$f(x(\lambda)) - f(x^0) \geq \frac{1}{2}\gamma\lambda^2 - \lambda o(\lambda) \geq \frac{1}{2}c\|x(\lambda) - x^0\|^2.$$

This verifies the assertion. □

Further, the contingent derivative can be applied in order to derive a first-order estimate for functionate.

Lemma 6.24 (mean values via Cf). *If* $f \in C(\mathbb{R}^n, \mathbb{R})$ *and* $\inf Cf(x+tu)(u) \le c\ \forall t \in (0,1)$ *then* $f(x+u) \le f(x) + c$. ◇

Proof. Let $q > c$ be fixed and let $T \subset [0,1]$ be the set of t satisfying

$$f(x+tu) \le f(x) + qt.$$

Then T is closed and $0 \in T$. Let s be the maximal element of T. To show that $s = 1$, assume $s < 1$. Since

$$\inf Cf(x+su)(u) < q \text{ and } s \in T,$$

we know that

$$f(x+su+\varepsilon_k u) < f(x+su) + q\varepsilon_k \le f(x) + qs + q\varepsilon_k,$$

for certain $\varepsilon_k \downarrow 0$, i.e., $s + \varepsilon_k \in T$. This contradiction shows $1 \in T$. Because $q > c$ was arbitrary, the lemma is shown. □

Corollary 6.25 (Lipschitz condition). *Let* $f \in C(\mathbb{R}^n, \mathbb{R}^m)$ *and suppose, with some* $\alpha > 0$ *and max-norm in* $\mathbb{R}^m$, *that* $Cf(x)(u) \cap \alpha B \ne \emptyset$ *holds for all* x *near* x^* *and* $u \in \operatorname{bd} B$. *Then* f *is Lipschitz near* x^* *with rank* α. ◇

Proof. Otherwise there are x' and x'' near x^* such that, with some i,

$$f_i(x'') - f_i(x') > \alpha\|x'' - x'\|.$$

Put $x'' = x' + su, u \in \operatorname{bd} B, s > 0$. Since, by assumption,

$$Cf_i(x' + t\,su)(su) \cap [-s\alpha, s\alpha] \ne \emptyset \quad \forall t \in (0,1),$$

Lemma 6.24 yields the opposite inequality

$$f_i(x'') - f_i(x') = f_i(x'+su) - f_i(x') \le s\alpha = \alpha\|x'' - x'\|,$$

which completes the proof. □

Exercise 14. Show that $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is C^1 on an open set Ω if $Tf(x)$ is single-valued $\forall x \in \Omega$. ◇

6.6 Contingent Derivatives of Implicit (Multi–) Functions and Stationary Points

Let us consider the set $\Phi(p)$ of solutions x to the equation

$$F(x,p) = 0, \tag{6.44}$$

under the assumptions that F maps $\mathcal{X} \times \mathcal{P}$ to $\mathcal{Z}$, and

$$\begin{array}{l}\mathcal{X}, \mathcal{P}, \mathcal{Z} \text{ are Banach spaces, some point } x^0 \in \mathcal{X} \text{ satisfies } F(x^0,0) = 0\,, \\ F_p \text{ exists and is continuous on some neighborhood } (\Omega, \Pi) \text{ of } (x^0,0),\end{array} \tag{6.45}$$

where F_p denotes the partial derivative with respect to p. In order to avoid confusion in notation of the present section, we use calligraphic letters to denote the spaces under consideration. Let

$$S = \Omega \cap F(\cdot,0)^{-1},$$

i.e.,

$$S(z) = \{x \in \Omega \mid F(x,0) = z\}.$$

Under the hypotheses of the usual implicit function theorem for $F \in C^1$, Φ and S are single–valued and locally C^1 (with derivatives $D\Phi, DS$), and there holds

$$D\Phi(0) = -DS(0)F_p(x^0,0) = -F_x(x^0,0)^{-1}F_p(x^0,0). \tag{6.46}$$

In this section, we are interested in a similar characterization of the contingent derivative $C\Phi(0,x^0)(\cdot)$ for the multivalued map Φ under the assumption (6.45). Moreover, we shall derive contingent derivative formulas for mappings of the (projection-) type $p \mapsto X(p) = \{\xi | \exists \eta : F(\xi,\eta,p) = 0\}$. In particular settings, this may be interpreted as a mapping of *stationary solutions* ξ with associated multipliers η.

A crucial motivation to study these questions comes from the sensitivity analysis of solutions to nonlinear programs. A.V. Fiacco and G.P. McCormick [Fia76, FM68] were the first who derived sufficient conditions for the validity of the formula (6.46) when $\mathcal{X}, \mathcal{P}, \mathcal{Z}$ are finite–dimensional spaces and Φ is the (single–valued and differentiable) primal–dual critical point mapping of a perturbed equality/inequality-constrained C^2 program. However, it is well–known that even in this special case, the assumptions to guarantee formula (6.46) are very restrictive, involving LICQ and the technical assumption of strict complementarity. The situation is similar for the stationary solution mapping. So, for non-C^2 optimization problems or for smooth programs under weaker assumptions one may only expect the existence of either one or the other *generalized* (directional) derivative of critical point or stationary/optimal solution maps, including related formulas. For this reason, such generalized differentiability properties play some rule in the literature, see the bibliographical note at the end of the section.

Our purpose is to point out which assumptions ensure (different levels of) generalizations of the relation (6.46) to contingent derivatives if Φ is defined by an *equation,* and this under a smooth parameter–dependence (6.45). For several instances of (6.44) which are related to nonlinear programs or complementarity problems (for example, if F *is* a so–called generalized Kojima function, see [KK99a]), the contingent derivative of $S = \Omega \cap F(\cdot, 0)^{-1}$ has an explicit representation – and this carries over to $C\Phi$ provided that a suitable extension of formula (6.46) is true.

6.6.1 Contingent Derivative of an Implicit (Multi-) Function

In this subsection, we are concerned with the contingent derivative of the solution set mapping $\Phi(p)$ of the equation $F(x,p) = 0$. Assume that (6.45) holds. By definition, the contingent derivative $C\Phi(0,x^0)$ contains all limits of usual difference quotients:

$$\xi \in C\Phi(0,x^0)(q) \Leftrightarrow \left[\begin{array}{l} \text{there exists some sequence } t \downarrow 0 \text{ along with} \\ \text{related } o\text{–type functions } o_1(\cdot),\ o_2(\cdot) \\ \text{such that } x^0 + t\xi + o_1(t) \in \Phi(tq + o_2(t)) \end{array}\right],$$

i.e., $C\Phi(0,x^0)(q)$ consists of all limits $\xi = \lim t^{-1}(x(t) - x^0)$ that can be obtained for *certain* sequences $t \downarrow 0$ and $x(t) \in \Phi(tq + o_2(t))$. Note that here and in the following, o–type functions will often be equipped with a subscript o_k in order to distinguish them. In any case, writing o_k we are saying that $\|o_k(\cdot)\|/\|\cdot\| \downarrow 0$ as $\|\cdot\| \downarrow 0$.

As only the local behavior of Φ near $(x^0, 0)$ is of importance, we may identify $\Phi(p)$ with $\Phi(p) \cap \Omega$, where Ω is the neighborhood of x^0 appearing in assumption (6.45). Notice further, that one could similarly regard points $p(t) = tq + o_2(t)$ in a fixed subset $P_0 \subset P$ only.

As it is standard in the smooth case, we consider, for (x,p) near $(x^0, 0)$, the function

$$r(x,p) = F(x,p) - F(x,0) - F_p(x^0,0)p. \tag{6.47}$$

By the mean-value theorem, one obtains

$$r(x,p) \;=\; \int_0^1 [F_p(x,\theta p) - F_p(x^0,0)]\, p\, d\theta,$$

where $\alpha(x,p,\theta) = \|F_p(x,\theta p) - F_p(x^0,0)\|$ can be estimated (uniformly for $0 < \theta < 1$) by

$$\alpha(x,p,\theta) \le O(x,p) \text{ with } O(x,p) \downarrow 0 \text{ as } x \to x^0 \text{ and } \|p\| \downarrow 0.$$

Due to $\|r(x,p)\| \le O(x,p)\|p\|$, one easily sees that

$$\|p\|^{-1}\|r(x,p)\| \to 0 \text{ as } x \to x^0 \text{ and } \|p\| \downarrow 0, \tag{6.48}$$

and, moreover,

$$r(x(t),p(t)) = o_2(t) \quad \text{if } x(t) \to x^0 \text{ and } p = tq + o_1(t) \text{ with some } q. \tag{6.49}$$

Further, for (x,p) near $(x^0,0)$, we have that

$$F(x,p) = 0 \Leftrightarrow F(x,0) = -F_p(x^0,0)p - r(x,p),$$

i.e.,

$$x \in \Phi(p) \Leftrightarrow x \in S(-F_p(x^0,0)p - r(x,p)). \tag{6.50}$$

If , moreover, F_p is locally Lipschitz near $(x^0,0)$ with rank K, we even know that the difference

$$r(x',p) - r(x,p) = \int_0^1 [F_p(x',\theta p) - F_p(x,\theta p)]\, p\, d\theta$$

satisfies

$$\|r(x',p) - r(x,p)\| \leq K\|x' - x\|\|p\| \tag{6.51}$$

for (x',p), (x,p)near $(x^0,0)$.

From now on, $r(\cdot,\cdot)$ according to (6.47) remains the same function, only x will be replaced by (x,y) in the next section. Note that we consider the mapping $S = F(\cdot,0)^{-1}$, because, due to Remark 1.1, the contingent derivative CS is known if and only if so is CF:

$$u \in CS(0,x^0)(\zeta) \Leftrightarrow \zeta \in CF(\cdot,0)(x^0)(u).$$

In the following we intend to exploit (6.48) and (6.50) under different *topological* assumptions concerning Φ and S, while, in the chapter on parametric optimization, we will additionally apply algebraic properties of F, which are available when stationary points of optimization problems in $\mathbb{R}^n$ come into the play.

Note that the (simple) inclusion (i) in the following theorem could be derived from [Lev96, Thm. 3.1]. To abbreviate, we denote the linear map $F_p(x^0,0)$ by F_p^o.

Theorem 6.26 (C-derivative of the implicit function). *Let (6.45) be satisfied. Then*

(i) (inclusion) $C\Phi(0,x^0)(q) \subset CS(0,x^0)(-F_p^o q)$.

(ii) (existence) Let $\dim \mathcal{X} < \infty$. *If* Φ *is l.s.c. at* $(0,x^0)$ *and* $S = F(\cdot,0)^{-1}$ *is locally Lipschitz u.s.c. at* $(0,x^0)$, *then* $\emptyset \neq C\Phi(0,x^0)(q)$.

(iii) (best case of a Lipschitzian inverse) Let $\dim \mathcal{X} + \dim \mathcal{Z} < \infty$. *If* Φ *is l.s.c. at* $(0,x^0)$ *and* $S = F(\cdot,0)^{-1}$ *is a locally Lipschitz function near* $(0,x^0)$, *then* $\emptyset \neq C\Phi(0,x^0)(q) = CS(0,x^0)(-F_p^o q)$.

◇

Proof, (i) Let $\xi \in C\Phi(0,x^0)(q)$. By definition of $C\Phi$, there are some sequence $t \downarrow 0$ and related $o_1(t)$ and $o_2(t)$ such that

$$x(t) := x^0 + t\xi + o_1(t) \in \Phi(tq + o_2(t)).$$

Setting $p(t) = tq + o_2(t)$ and taking (6.50) into account, we obtain

$$\begin{aligned} x^0 + t\xi + o_1(t) &\in S(-F_p^o(tq + o_2(t)) - r(x(t), p(t))) \\ &= S(-tF_p^o q - F_p^o o_2(t) - r(x(t), p(t))). \end{aligned}$$

After subtracting $x^0 \in S(0)$ and division by t, now (6.49) ensures that

$$\xi \in CS(0,x^0)(-F_p^o q).$$

(ii) We have to show that $C\Phi(0,x^0)(q) \neq \emptyset$. For $q = 0$ this is trivial, so let $q \neq 0$ and $t \downarrow 0$. We put $p = tq$ and use the assumption that S is locally Lipschitz u.s.c. with rank L. Since Φ is l.s.c. at $(0, x^0)$ there exist $x(p) \in \Phi(p)$ such that $x(p) \to x^0$. Applying (6.48) and (6.50), we observe for sufficiently small $\tau > 0$ and $0 < t < \tau$ that

$$\|r(x(p), p)\| < Lt\|q\|$$

and

$$x(p) \in S(-tF_p^o q - r(x(p), p)) \subset x^0 + t(L + \|F_p^o\|)\|q\|)B_{\mathcal{X}}.$$

Thus the sequence $t^{-1}\|(x(p) - x^0)\|$ remains bounded since $\dim \mathcal{X} < \infty$. So $\xi \in C\Phi(0,x^0)(q)$ holds for every accumulation point ξ of this sequence as t vanishes.

(iii) It remains to show that $CS(0,x^0)(-F_p^o q) \subset C\Phi(0,x^0)(q)$. Let $\xi \in CS(0,x^0)(-F_p^o q)$. Then for a certain sequence $t \downarrow 0$ and some related $o_1(t)$, $o_2(t)$, one has

$$x^0 + t\xi + o_1(t) \in S(-tF_p^o q - o_2(t)).$$

Due to the lower semicontinuity of Φ we find, for the same sequence $t \downarrow 0$, points

$$y(t) \in \Phi(tq) \text{ such that } y(t) \to x^0 \text{ as } t \downarrow 0.$$

Once more we apply (6.50) and obtain

$$y(t) = S(-tF_p^o q - r(y(t), tq)).$$

Since S is a locally Lipschitz function near $(0, x^0)$, say with rank L, we conclude

$$\|y(t) - (x^0 + t\xi + o_1(t))\| \leq L\|r(y(t), tq) + o_2(t)\| =: o_3(t).$$

This ensures, for the current sequence, $(y(t) - x^0)/t \to \xi$ and $\xi \in C\Phi(0,x^0)(q)$.
□

Note that indeed $C\Phi(0,x^0)(q)$ may be empty if any of the assumptions of (ii) does not hold. Consider, as a first example, $\Phi(p) = S(p) = \{x \in \mathbb{R} | g(x) = p\}$ with $g(x) := \min\{x, x^2\}$. Then $\Phi(p) = \{\sqrt{p}\}$ for $0 \le p \le 1$, and so $C\Phi(0,0)(1) = \emptyset$. Here Φ is l.s.c. at (0,0), but the local Lipschitz upper semicontinuity fails. A second example gives a locally Lipschitz u.s.c, mapping 5, but the l.s.c. of Φ fails. Consider $F(x,p) = (F_1(x,p), F_2(x,p))$ with $F_1(x,p) = x - p$ and $F_2(x,p) = x$. Obviously $\Phi(p) := \{x | F(x,p) = 0\}$ is empty for $p \neq 0$ and is equal to {0} for $p = 0$. Hence $C\Phi(0,0)(q) = \emptyset$ for $q \neq 0$. However $S(z) := \{x | F(x,0) = z\}$ is empty for $z = (z_1, z_2)$ with $z_1 \neq z_2$, but coincides with $\{s\}$ for $z_1 = z_2 = s$, hence S is locally Lipschitz u.s.c. at the origin.

Remark. From the proof of the preceding theorem, one observes that in (ii) as well as in (iii) the l.s.c. assumption on Φ may be weakened: Given some direction q, one has only to suppose that the multifunction $t \in \mathbb{R}_+ \mapsto \Phi(tq)$ is l.s.c. at $(0, x^0)$. In the latter case, we shall say that Φ is *l.s.c. at* $(0, x^0)$ *in direction q*. In the following, we construct a problem for which this weaker l.s.c. assumption on Φ is satisfied, and property (ii) holds but property (iii) fails, i.e.,

$$\emptyset \neq C\Phi(0,x^0)(q) \subsetneqq CS(0,x^0)(-F_p^o q).$$

In this example, F is even a locally Lipschitzian function satisfying (6.45). Consider

$$\Phi(p) := \left\{ (x,y) \in \mathbb{R}^2 \,\middle|\, F(x,y,p) = \begin{pmatrix} |x| + |y| - p \\ x - px + y - p \end{pmatrix} = 0 \right\}, \quad p \in \mathbb{R},$$

in direction $q = 1$. Then $\Phi(p) = \{(0,p)\}$ for $0 \le p < 2$ and $\Phi(p) = \emptyset$ for $p < 0$, hence Φ is l.s.c. at 0 in direction $q = 1$ (but not l.s.c. at 0). Moreover, $C\Phi(0)(1) = \{(0,1)\}$ and $-F_p(0) \cdot 1 = (1,1)$. One easily verifies that $S = F(\cdot,\cdot,0)^{-1}$ is locally Lipschitz u.s.c. at 0. However, $CS(0)(1,1)$ is the convex hull of (1,0) and (0,1). ◇

To show the equation in assertion (iii) of Theorem 6.26 but allowing that the related sets are empty, it suffices to know that S is pseudo–Lipschitz and that $F_p(\cdot,\cdot)$ is locally Lipschitz. As a basic tool we use the implicit (multi–) function estimate of Theorem 4.9 in the case of a pseudo–Lipschitz inverse mapping. Note that $F(x,0) = g(x,p)$ with $g(x,p) = -F_p(x^0,0)p - r(x,p)$ defines a nonlinear perturbation of the initial equation $F(x,0) = 0$, so the next theorem does *not immediately* follow from the pseudo–Lipschitz continuity of the solution set mapping $S(z)$ of $F(x,0) = z$.

Theorem 6.27 (the case of pseudo-Lipschitz S). *Let (6.45) be satisfied, and let $F_p(\cdot,\cdot)$ be locally Lipschitz near $(x^0, 0)$. If S is pseudo–Lipschitz near $(0, x^0)$, then $C\Phi(0,x^0)(q) = CS(0,x^0)(-F_p^o q)$. If, moreover, $\dim \mathcal{X} < \infty$, then $C\Phi(0,x^0)(q) \neq \emptyset$.*

Proof. By assumption, $F(x^0, 0) = 0$. Moreover, we note that now the Lipschitz estimate (6.51)

$$\|r(x', p) - r(x, p)\| \le K\|p\|\|x' - x\|$$

holds locally with some K. The inclusion (i) of Theorem 6.26 is true as shown above. Hence, we have only to verify that $C\Phi(0, x^0)(q) \supset CS(0, x^0)(-F_p^o q)$. Let $p = tq$ and $u \in CS(0, x^0)(-F_p^o q)$. So, for some sequence $t \downarrow 0$ and related $o_1(t)$ and $o_2(t)$ the equations

$$F(x(t), 0) = -tF_p^o q + o_2(t), \qquad x(t) = x^0 + tu + o_1(t)$$

are valid. By (6.50), we have $F(x,p) = 0 \;\Leftrightarrow\; F(x,0) = -tF_p^o q - r(x, tq)$. Since $S = F(\cdot, 0)^{-1}$ is pseudo-Lipschitz (say, with constant L) at $(0, x^0)$, we may apply Theorem 4.9 which states the following result:

Let $\varepsilon > 0$ be sufficiently small, and suppose that any locally Lipschitz functions g and $\tilde{g}$ with values in $\mathcal{Z}$ fulfill the estimates $\mathbf{Lip}\,(\tilde{g}, U(\varepsilon)) < \frac{1}{2}(L+1)^{-1}$,

$$\sup_{x \in U(\varepsilon)} \|g(x)\| < \frac{\varepsilon}{8}(L+1)^{-1} \text{ and } \sup_{x \in U(\varepsilon)} \|\tilde{g}(x)\| < \frac{\varepsilon}{8}(L+1)^{-1},$$

where $U(\varepsilon) := x^0 + \varepsilon B_{\mathcal{X}}$, and $\mathbf{Lip}\,(\tilde{g}, U(\varepsilon))$ denotes the smallest Lipschitz constant of $\tilde{g}$ on $U(\varepsilon)$. Then, if $\xi \in U(\frac{\varepsilon}{2})$ solves $F(x,0) = g(x)$, there exists a solution $\tilde{\xi}$ to $F(x,0) = \tilde{g}(x)$ such that $\|\xi - \tilde{\xi}\| \le 2(L+1)\|\tilde{g}(\xi) - g(\xi)\|$.

Setting now $\xi = x(t)$, $g(x) = -tF_p^o q + o_2(t)$ and $\tilde{g}(x) = -tF_p^o q - r(x, tq)$, we observe that $\|x(t) - x^0\| < t(\|u\| + 1) =: \frac{\varepsilon}{2}$, and the Lipschitz rank of $r(\cdot, tq)$ becomes arbitrarily small for sufficiently small t. Thus, there is a solution $\tilde{\xi} = y(t)$ to $F(x,0) = \tilde{g}(x)$ such that

$$\|\tilde{\xi} - \xi\| \le 2(L+1)\|\tilde{g}(\xi) - g(\xi)\| =: o_3(t).$$

This yields $\|y(t) - (x^0 + tu + o_1(t))\| \le o_3(t)$ and $F(y(t), tq) = 0$. So the current sequence fulfills $(y(t) - x^0)/t \to u$ and $u \in C\Phi(0, x^0)(q)$, and the main assertion of the theorem is shown.

For the second statement $C\Phi(0, x^0)(q) \neq \emptyset$ under $\dim \mathcal{X} < \infty$, it suffices now to verify that $CS(0, x^0)(\zeta) \neq \emptyset$ for all $\zeta \in \mathcal{Z}$. By assumption, S is pseudo–Lipschitz at $(0, x^0)$ with some rank L. Hence, given any sequence $t \downarrow 0$, there are $x_t \in S(t\zeta)$ such that $\|x_t - x^0\| \le tL\|\zeta\|$ provided that $t > 0$ is small enough. If $x_t = x^0$ then $0 \in CS(0, x^0)(\zeta)$ is evident, otherwise an accumulation point ξ of the sequence $(x_t - x^0)/\|x_t - x^0\|$ exists and fulfills $\xi \in CS(0, x^0)(\zeta)$. □

6.6.2 Contingent Derivative of a General Stationary Point Map: Topological Assumptions

In contrast to the former settings, let F depend on a third variable $y \in \mathcal{Y}$, and suppose that the spaces

$$\mathcal{X}, \mathcal{Y}, \mathcal{Z} \text{ are finite–dimensional.}$$

We are now interested in the *projection X(p) of the solution set*

$$\Phi(p) = \{(x,y) | F(x,y,p) = 0\} \tag{6.52}$$

onto the x-space, i.e. we put

$$x \in X(p) \text{ if } (x,y) \in \Phi(p) \text{ for some } y. \tag{6.53}$$

In several settings, e.g., if (x,y) is a primal-dual solution of the KKT system to a parametric optimization problem, then x is said to be a *stationary point* for the parameter p. This explains the title of this subsection.

In accordance with the previous subsection, we suppose

$$\begin{array}{l} F : \mathcal{X} \times \mathcal{Y} \times \mathcal{P} \to \mathcal{Z}, \ F(x^0, y^0, 0) = 0 \text{ holds for some } (x^0, y^0), \\ \text{and the partial derivative } F_p \text{ exists and is continuous on } (\Omega, \mathcal{Y}, \Pi), \end{array} \tag{6.54}$$

where again, (Ω, Π) denotes a neighborhood of $(x^0, 0)$. To abbreviate, let

$$s^0 = (x^0, y^0) \text{ and } F_p^o = F_p(s^0, 0).$$

We want to characterize the contingent derivative $CX(0, x^0)$, where we only use topological properties which are typical for standard stationary point mappings. A crucial special realization – devoted to stationary solutions of nonlinear programs – will be regarded in the chapter on parametric optimization. By definition only, it holds

$$\xi \in CX(0,x^0)(q) \Leftrightarrow \left[\begin{array}{l} \text{for some sequence } t \downarrow 0 \text{ there are functions} \\ o_1(\cdot) \text{ and } o_2(\cdot) \text{ such that, setting} \\ p(t) = tq + o_2(t),\ x(t) = x^0 + t\xi + o_1(t), \\ \text{the points } x(t) \text{ with some related } y(t) \\ \text{fulfill } (x(t), y(t)) \in \Phi(p(t)). \end{array} \right] \tag{6.55}$$

The former mapping S now reads

$$S = F(\cdot, \cdot, 0)^{-1},$$

and is given by the solutions (x,y) to $F(x,y,0) = z$. Then it holds for directions ζ being assigned to $z \in \mathcal{Z}$,

$$(u,v) \in CS(0,s^0)(\zeta) \Leftrightarrow \zeta \in CF(\cdot,\cdot,0)(s^0)(u,v).$$

Again, CS is known if so is CF. We summarize these facts by

$$(\xi,\eta) \in CS(0,s^0)(-F_p^o q) \ \Leftrightarrow \ -F_p^o q \in CF(\cdot,\cdot,0)(s^0)(\xi,\eta) \tag{6.56}$$

and

$$(\xi,\eta) \in C\Phi(0,s^0)(q) \Rightarrow \xi \in CX(0,x^0)(q) \tag{6.57}$$

Further, we may identify (x,y) with x of the previous subsection and set

$$r(x,y,p) = F(x,y,p) - F(x,y,0) - F_p^o p$$

in order to obtain (by the same arguments) the same estimates, in particular,

$$\|p\|^{-1}\|r(x,y,p)\| \to 0 \text{ as } (x,y) \to s^0 \text{ and } \|p\| \downarrow 0, \tag{6.58}$$

and

$$(x,y) \in \Phi(p) \;\Leftrightarrow\; x \in S(-F_p(s^0,0)p - r(x,y,p)). \tag{6.59}$$

In the following, we intend to obtain a statement saying that $\xi \in CX(0,x^0)(q)$ for some $s^0 = (x^0,y^0) \in \Phi(0)$ yields the existence of some η such that

$$(\xi,\eta) \in C\Phi(0,s^0)(q). \tag{6.60}$$

Having this statement, it is quite natural to write $C\Phi(0,s^0)(q)$ by using the set $CS(0,s^0)(-F_p^o q)$ and, finally, to write it via (6.56) in terms of *CF* for the given *F*. The inclusion (6.60) means that the equation

$$F(x,y,p) = 0 \quad \text{or, equivalently,} \quad F(x,y,0) = -F_p^o p - r(x,y,p)$$

has solutions satisfying (for some sequence $t \downarrow 0$ and related $o_1(t)$, $o_2(t)$, $o_3(t)$)

$$x = x^0 + t\xi + o_1(t),\;\; y = y^0 + t\eta + o_2(t),\;\; p = tq + o_3(t). \tag{6.61}$$

So (6.58) ensures, with some $o_4(\cdot)$, $F(x,y,0) = -tF_p^o q - o_4(t)$ and $(\xi,\eta) \in CS(0,s^0)(-F_p^o q)$ as well as

$$C\Phi(0,s^0)(q) \subset CS(0,s^0)(-F_p^o q). \tag{6.62}$$

Moreover, if $\|\xi\| + \|q\| > 0$ then it is easy to see that the sequences (6.61) *necessarily* satisfy a l.s.c. condition (with respect to y) of Lipschitz-type, namely,

$$\|y - y^0\| = \|t\eta + o_2(t)\| \le L\|(x - x^0, p)\| \quad \text{for small } t > 0,$$

where L is chosen such that $\eta \le \frac{1}{2}L(\|\xi\| + \|q\|)$. So, to simplify the further presentation, we impose this l.s.c. assumption which is needed anyway if $\|\xi\| + \|q\| > 0$ holds:

> **If $x(p) \in X(p)$ is any given sequence with $(x(p),p) \to (x^0,0)$, then, at least for some subsequence of it, there exist $y(p)$ and $L > 0$ in such a way that both $(x(p),y(p)) \in \Phi(p)$ and an estimate $\|y(p) - y^0\| \le L\|(x(p) - x^0, p)\|$ hold true.** (6.63)

We will use this condition for points of the form $p = tq + o_3(t)$. Recall that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are in general supposed to have finite dimension.

Theorem 6.28 (C–derivatives of stationary points, general case). *Let (6.54) and (6.63) be satisfied, let $s^0 = (x^0,y^0)$ and $F_p^o = F_p(x^0,y^0,0)$. Then one has:*

(i) (inclusion) For each $\xi \in CX(0,x^0)(q)$ there is some η such that

$$(\xi,\eta) \in C\Phi(0,s^0)(q). \tag{6.64}$$

(ii) *(existence) If X is l.s.c. at* $(0, x^0)$ *and S is locally Lipschitz u.s.c. at* $(0, s^0)$, *then* $CX(0, x^0)(q) \neq \emptyset$.

(iii) *(Lipschitzian inverse) If X is l.s.c. at* $(0, x^0)$ *and S is a locally Lipschitz function near* $(0, s^0)$, *then for each q, one has both* $CX(0, x^0)(q) \neq \emptyset$ *and* $\xi \in CX(0, x^0)(q) \Leftrightarrow (\xi, \eta) \in CS(0, s^0)(-F_p^o q)$ *with some* η.

◇

Proof. (i) Let $\xi \in CX(0, x^0)(q)$. Then, according to (6.55), there are some sequence $t \downarrow 0$ and functions $o_1(\cdot)$ and $o_2(\cdot)$ such that by setting $p(t) = tq + o_2(t)$, $x(t) = x^0 + t\xi + o_1(t)$, the points $x(t)$ with some related $y(t)$ fulfill $F(x(t), y(t), p(t)) = 0$. By (6.63), the points $y(t)$ can be taken in such a way that $\|y(t) - y^0\| \leq Lt$ with some fixed L. Thus, since an accumulation point η of $(y(t) - y^0)/t \in \mathcal{Y}$ exists, (6.64) follows by definition of $C\Phi$.

(ii) Now Φ is l.s.c. at $(0, s^0)$ due to (6.63), and Theorem 6.26 (ii) tells us that $C\Phi(0, s^0)(q)$ contains some element (ξ, η). So $\xi \in CX(0, x^0)(q)$ follows from (6.57).

(iii) Again by Theorem 6.26(iii), we observe that

$$\emptyset \neq C\Phi(0, s^0)(q) = CS(0, s^0)(-F_p^0 q)$$

holds, and (ii) shows $CX(0, x^0)(q) \neq \emptyset$. Using (i), this is the assertion. □

From the proof of the preceding theorem, one observes that in (ii) and (iii) the l.s.c. assumption on Φ may be replaced by the l.s.c. of Φ *in direction q* (see the remark following Theorem 6.26). Further, notice that Theorem 6.27 can be applied to the current maps Φ and S if one is interested in obtaining the assertions under (iii) without supposing (6.63). In the particular setting of optimality conditions, this relates metric regularity and the local Lipschitz property of 5, we shall discuss this in the chapter devoted to parametric programs.

On Assumption (6.63)

In the next theorem we shall substitute (6.63) by a *transversality* condition. To do this we first suppose local boundedness of $y(p)$ as follows:

If $(x(p), p) \to (x^0, 0)$, $x(p) \in X(p)$, is any given sequence, then at least for some subsequence of the original one, there exists a bounded sequence $y(p)$ such that $(x(p), y(p)) \in \Phi(p)$. (6.65)

Then some accumulation point y^0 of the sequence $y(p) \in \mathcal{Y}$ exists, since $\mathcal{Y}$ is finite–dimensional by assumption, and – after selecting another subsequence if necessary – we may assume that

$$w = \lim \|y(p) - y^0\|^{-1}(y(p) - y^0) \tag{6.66}$$

exists, too. We also consider the partial inverse mapping

$$S^0(z) = \{y | F(x^0, y, 0) = z\}$$

and suppose that S^0 is *upper Lipschitz at the origin* in the following sense: For every bounded set $V \subset \mathcal{Y}$ there is a constant $K > 0$ such that

$$V \cap S^0(z) \subset S^0(0) + K\|z\|B \quad \forall z \in K^{-1}B.$$

We note that for X being the stationary point mapping of a standard C^1 program, assumption (6.65) is satisfied if the Mangasarian–Fromovitz constraint qualification (MFCQ) holds at x^0 (for $p = 0$). In this special case, it is obvious that condition (6.63) is stronger than MFCQ; in our general setting (6.53), (6.54), a transversality condition will be added to the boundedness assumption (6.65).

Let

$$Y^0 := \{y | F(x^0, y, 0) = 0\},$$

and let $C(y^0, Y^0)$ be the contingent (or Bouligand) tangent cone of Y^0 at y^0, i.e.,

$$C(y^0, Y^0) = \limsup_{\lambda \downarrow 0} \lambda^{-1}(Y^0 - y^0).$$

Finally, let $C_y F(x^0, y^0, 0)$ denote the (partial) contingent derivative of F with respect to y at $(x^0, y^0, 0)$, and let $\ker C_y F(x^0, y^0, 0) = \{w | 0 \in C_y F(x^0, y^0, 0)w\}$ be its kernel.

Theorem 6.29 (transversality condition). *Let F satisfy (6.54) and (6.65). In addition, let F be locally Lipschitz and* S^0 *be upper Lipschitz at the origin. Let* y^0 *be some accumulation point of y(p) where* $(x(p), y(p)) \in \Phi(p)$ *and* $(p, x(p)) \to (0, x^0)$. *Then* $\|y(p) - y^0\| \le L\|(x(p) - x^0, p)\|$ *holds for some subsequence and some L whenever the transversality condition*

$$\{0\} = C(y^0, Y^0) \cap \ker C_y F(x^0, y^0, 0) \tag{6.67}$$

is satisfied. ◇

Proof. Assume that (6.63) does not hold. Then, by (6.65), for given $(x(p), p) \to (x^0, 0)$, $x(p) \in X(p)$ and some related subsequence $y(p)$ considered under (6.66), it holds

$$\|y(p) - y^0\| / \|(x(p) - x^0, p)\| \to \infty \tag{6.68}$$

We show that the nontrivial vector w from (6.66) belongs to

$$C(y^0, Y^0) \cap \ker C_y F(x^0, y^0, 0).$$

With $x = x(p)$, $y = y(p)$ and $F(x^0, y^0, 0) = 0$, we obtain that

$$\begin{aligned} 0 &= F(x, y, p) \\ &= (F(x^0, y, 0) - F(x^0, y^0, 0)) + (F(x, y, p) - F(x^0, y, 0)) \end{aligned}$$

where, since F is locally Lipschitz with some rank L_F, one has

$$\|F(x,y,p) - F(x^0,y,0)\| \leq L_F\|(x - x^0,p)\|. \tag{6.69}$$

Further, in finite dimension and for locally Lipschitz F, the differences

$$F(x^0,y,0) - F(x^0,y^0,0)$$

satisfy

$$F(x^0,y,0) - F(x^0,y^0,0) \in C_yF(x^0,y^0,0)(y - y^0) + o_1(y - y^0)B_Z.$$

So we obtain from (6.69)

$$0 \in C_yF(x^0,y^0,0)(y - y^0) + (o_1(y - y^0) + L_F\|(x - x^0,p)\|)B_Z.$$

Taking (6.68) into account, division by $\|y - y^0\|$ and passing to the limit $p \to 0$ yield

$$0 \in C_yF(x^0,y^0,0)\, w, \quad \text{i.e.,} \quad w \in \ker C_yF(x^0,y^0,0)$$

Hence there exists some sequence $\lambda \downarrow 0$ such that $F(x^0,y^0 + \lambda w,0) = o_2(\lambda)$. Since S^0 is upper Lipschitz we derive from $y^0 + \lambda w \in S^0(o_2(\lambda))$ that there exist y_λ and a constant K such that both $F(x^0,y_\lambda,0) = 0$ and $\|y^0 + \lambda w - y_\lambda\| \leq K\|o_2(\lambda)\|$. Thus, it holds $y_\lambda \in Y^0$ as well as $w = \lim(y_\lambda - y^0)/\lambda$. So w belongs to $C(y^0,Y^0)$, too. □

If Y^0 is a singleton then (6.67) holds trivially. For standard nonlinear programs, this means uniqueness of the Lagrange multiplier to x^0, and is often written in an algebraic manner as *strict MFCQ* at x^0 (for $p = 0$). It is worth noting that the particular structure of F is not needed for showing (6.63) under this condition.

Nevertheless, a complete description of CX as in Theorem 6.28 (iii),

$$\xi \in CX(0,x^0)(q) \Leftrightarrow (\xi,\eta) \in CS(0,s^0)(-F_p^o q),$$

needed hard suppositions due to the following facts.

Having some pair (y^0,η) such that $(\xi,\eta) \in C\Phi(0,s^0)(q)$ holds, where $s^0 = (x^0,y^0)$, we know that $(\xi,\eta) \in CS(0,s^0)(-F_p^o q)$, but the reverse conclusion is not true, in general. Indeed, the latter inclusion only says that certain sequences (as $t \downarrow 0$) satisfy

$$x_t = x^0 + t\xi + o_1(t), \; y_t = y^0 + t\eta + o_2(t) \;\text{ and }\; F(x_t,y_t,0) = -tF_p^o q - o_3(t).$$

But it must be shown that (as in the proof of Theorem 6.27), the particular nonlinear equation

$$F(x,y,0) = -tF_p^o q - r(x,y,tq)$$

has solutions close enough to (x_t,y_t). This requires basically implicit function arguments. To avoid them, often the *extended* stationary point map $\widetilde{X}$ defined as

$$\widetilde{X}(p,z) = \{x | F(x,y,p) = z \text{ for some } y\} \tag{6.70}$$

has been investigated. We will follow this line in our application to nonlinear programs, see Chapter 8.1.

The contingent derivatives of $\tilde{X}$ and X are (using the definitions only) related to each other by

$$CX(0_{\mathcal{P}}, x^0)(q) \subset C\tilde{X}((0_{\mathcal{P}}, 0_{\mathcal{Z}}), x^0)(q, 0_{\mathcal{Z}}), \tag{6.71}$$

where the subscript of 0 refers to the corresponding space.

Bibliographical Note. The approach presented in Section 6.6 is based on the authors' paper [KK01]. In the context of *inclusions,* contingent derivatives of general implicit multifunctions were studied, e.g., in [AF90, LR94, Lev96, RW98].

In the context of nonlinear programming, generalized differentiability properties have been investigated in the literature quite often. By using the theory of second–order optimality and stability conditions, the existence and representation of *standard* directional derivatives of stationary or optimal solutions to C^2–programs in $\mathbb{R}^n$ were studied, for example, in the papers [GD82, GJ88, Sha88b, RD95]. For a recent survey of this approach, we refer to [BS98] and [BS00, Chapter 3].

In [LR95] the existence of the proto–derivative (and, hence, of the contingent derivative) of the stationary solution map of a parametric C^2 optimization problem in $\mathbb{R}^n$ was shown to hold under the Mangasarian–Fromovitz constraint qualification, and a derivative formula was given. The approach in [LR95] was based on the study of proto-derivatives in the context of subgradient mappings, see [PR94, LR96, RW98] and Section 9.3. More results on the existence and representation of proto–derivatives or B–derivatives of the stationary solution mapping (or of the solution sets of parametric nonsmooth equations), can be found, e.g., in [Rob91, Pan93, Lev96, RW98]. For nonlinear programs with $C^{1,1}$ data, an extension of formula (6.46) to Thibault derivatives of the critical point map was given in [Kum91a] under the assumption of strong regularity. ◇

Chapter 7

Critical Points and Generalized Kojima–Functions

In order to study critical points of optimization problems as well as solutions of generalized equations and complementarity problems in a unified way, we prefer to use a direct, analytical approach for characterizing such points: namely as zeros of some nonsmooth function F sending $\mathbb{R}^d$ into itself. Various functions are suitable for this purpose, and later we will deal with several of them, indeed. In the present chapter we consider systems of equations which are defined by locally Lipschitz functions of a special structure and are adapted from Kojima's [Koj80] form of the KKT conditions for C^2–optimization problems. This leads to the notion *(generalized) Kojima–function.* We shall investigate different regularity concepts for such systems, or, equivalently, different Lipschitz properties of critical point and stationary solution mappings.

7.1 Motivation and Definition

As a starting point let us consider the nonlinear optimization problem

$$\min f(x) \quad \text{s.t.} \quad x \in \mathcal{M}, \tag{7.1}$$

where

$$\mathcal{M} = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ccc} g_i(x) & \leq & 0 \\ h_k(x) & = & 0 \end{array} \quad \begin{array}{c} (i = 1, \dots, m) \\ (k = 1, \dots, \kappa) \end{array} \right\},$$

and the functions $f, g_i, h_k : \mathbb{R}^n \to \mathbb{R}$ $(\forall i\ \forall k)$ are (at least) continously differentiable near some point of interest.

In the following, however, we will mainly deal with the case that the functions f, g_i and h_k belong to the class $C^{1,1}$. The latter hypothesis opens the use of several technical tools and allows us to weaken the standard smoothness assumption of critical point theory in nonlinear programming. The weaker supposition makes sense, for example, in two–level optimization, decomposition approaches and semi–infinite optimization, where $C^{1,1}$ data appear in a natural way. Note that even for problems without constraints, the gap between $f \in C^{1,1}$ and $f \in C^2$ is very large. This will become clear in several situations below.

KKT Points and Critical Points in Kojima's Sense

Using the *standard Lagrange function* $L(x,y,z) := f(x) + \langle y, g(x)\rangle + \langle z, h(x)\rangle$, the classical necessary optimality conditions to problem (7.1) in the sense of Karush, Kuhn and Tucker have the form

$$D_x L(x,y,z) = 0,\ h(x) = 0,\ g(x) \le 0,\ y \ge 0,\ \langle y, g(x)\rangle = 0, \tag{7.2}$$

a solution (x,y,z) of this system is called a *Karush–Kuhn–Tucker point (KKT point)* of (7.1). If (y,z) exists such that (x,y,z) satisfies (7.2), then x is said to be a *stationary solution* to (7.1). It is well–known that a local minimizer of (7.1) is necessarily a stationary solution, provided that some constraint qualification holds.

Following Kojima [Koj80], one may assign to (7.1) the function $F : \mathbb{R}^d \to \mathbb{R}^d$, $d = n + m + \kappa$, with the components

$$\begin{array}{lll} F_1 = Df(x) + \sum_{i=1}^m y_i^+ Dg_i(x) + \sum_{k=1}^\kappa z_k Dh_k(x) & \textbf{Lagrange term,} & \\ F_2 = g(x) - y^-, & \textbf{feasibility term,} & (7.3) \\ F_3 = h(x), & \textbf{feasibility term,} & \end{array}$$

where $y_i^+ := \max\{y_i, 0\}$ and $y_i^- := \min\{y_i, 0\}$, and the vectors y^+ and y^- are defined by the components y_i^+ and y_i^- $(\forall i)$, respectively. Evidently, $y^- \perp y^+$ and $y = y^+ + y^-$. Notice that y^+ and y^- are the Euclidean projections of y to the nonnegative and non-positive orthant of $\mathbb{R}^m$, respectively. As projections, they are connected with the related normals in an evident way, in particular, $y^- = y - y^+ \in N_{\mathbb{R}^m_+}(y^+)$.

The function F is called the (usual) *Kojima–function* of the program (7.1). If $s = (x,y,z) \in \mathbb{R}^d$ is a zero of F, then we say that s is a *critical point* and x is a *stationary solution of the system* $F(x,y,z) = 0$ (or simply *of F*). Moreover, if $\phi = f(x)$ for some stationary solution x of F, then we say that ϕ is a *critical value* of (7.1).

Since, in general, y may have negative components, the *Lagrangian L* is now

$$L(x,y,z) = f(x) + \sum_{i=1}^m y_i^+ g_i(x) + \sum_{k=1}^\kappa z_k h_k(x),$$

and $F_1 = D_x L$. We do not need a new symbol because y was nonnegative in the former settings. One immediately sees that

$$\begin{array}{lll} (x,y,z) \text{ KKT point} & \Rightarrow & (x, y + g(x), z) \text{ critical} \\ (x,y,z) \text{ critical} & \Rightarrow & (x, y^+, z) \text{ KKT point.} \end{array} \tag{7.4}$$

Note that both transformations are locally Lipschitz, this is important in view of our regularity notions.

One may even directly identify critical points and KKT points: By continuity, inactive constraints – i.e., such with $g_i(x) < 0$ – remain inactive for x' near x. So these constraints do not play any role for the local analysis of KKT points near $s = (x, y, z)$, and we may delete g_i in such a context. Hence, one might suppose that $g(x) = 0$ at the given stationary solution x of interest (but not at x' near x), whereupon s is a KKT point iff s is critical. Nevertheless, we will not always make use of this fact, in order to show also certain symmetries concerning *active* and *non–active* constraints in several statements.

Generalized Kojima–Functions – Definition

It was first observed in [Kum98] that several regularity results with respect to the Kojima–function F defined in (7.3) do not require the concrete form of the first component F_1 of F, but only the affine–linear structure of F_1 with respect to (y^+, z), i.e., in (7.3), Df, Dg_i and Dh_k can be replaced by arbitrary (continuous) functions Φ, Ψ_i and Γ_k. This extension maintains the nice separable form of F, namely, $F(x, y, z)$ can be written as the product of the ***row*** vector

$$N(y, z) = (1, y^+, y^-, z), \tag{7.5}$$

and a specially structured matrix $M(x)$. This is our ***key observation*** and suggests the following definition. In the next section, we shall see that this product describes besides Kojima's function (7.3) several further objects of interest in the context of optimization. Let again $d = n + m + \kappa$.

Definition. A function $F : \mathbb{R}^d \to \mathbb{R}^d$, is said to be a *generalized Kojima-function* if it has the representation

$$F(x, y, z) = N(y, z)M(x),$$

where $N(y, z)$, $(y, z) \in \mathbb{R}^{m+\kappa}$,is given by (7.5), $M(x)$, $x \in \mathbb{R}^n$, is defined by

$$M(x) = \begin{pmatrix} \Phi(x) & g(x)^{\mathsf T} & h(x)^{\mathsf T} \\ \Psi(x) & 0 & 0 \\ 0 & -E & 0 \\ \Gamma(x) & 0 & 0 \end{pmatrix}, \tag{7.6}$$

and E is the (m, m)–identity matrix, $\Phi(x) \in \mathbb{R}^n$ is a row vector, $\Psi(x)$ and $\Gamma(x)$ are matrices formed by the row vectors $\Psi_1(x), \ldots, \Psi_m(x), \Gamma_1(x), \ldots, \Gamma_\kappa(x)$ in $\mathbb{R}^n$, and $g(x)$, $h(x)$ are regarded as *column vectors* of length m and κ, respectively. ◇

Convention. To avoid the frequent use of the transition symbol, we often will omit it if the context is clear. In particular, we sometimes write (a, b) to denote a column vector composed by the column vectors a and b, and we identify uA

and $A^{\mathsf{T}}u$ for the product of a vector u and a matrix A. Further, we will agree in several applications of generalized Kojima functions to

$$\text{write } F = MN \text{ instead of } F^{\mathsf{T}} = M^{\mathsf{T}}N^{\mathsf{T}} \tag{7.7}$$

and to omit the transition symbol in $\Phi(x)$, $\Psi_i(x)$, and so on. ◇

Obviously, F is the usual Kojima–function if $\Phi(x) = Df(x)$, $\Psi_i(x) = Dg_i(x)$ and $\Gamma_k(x) = Dh_k(x)$. In view of a "second–order" analysis, we shall often suppose in (7.6) that $\Phi, \Psi, \Gamma \in C^{0,1}$ and $g, h \in C^{1,1}$, but formally this is not required in the above definition.

The notions *critical point* and *stationary solution* are used similarly to those for the usual Kojima–function. A generalized Kojima–function F with components F_1, F_2 and F_3 has the explicit form

$$\begin{aligned} F_1 &= \Phi(x) + \textstyle\sum_{i=1}^{m} y_i^+ \Psi_i(x) + \sum_{k=1}^{\kappa} z_k \Gamma_k(x), \\ F_2 &= g(x) - y^-, \\ F_3 &= h(x). \end{aligned} \tag{7.8}$$

Of course, the components F_2 and F_3 are still the same as in (7.3). The equation $F(x, y, z) = 0$ is sometimes called a (generalized) *Kojima system.*

The separable structure of $F = NM$ and the special form of N are very helpful for computing generalized derivatives. The function N is locally Lipschitz and has, in addition, a special simple structure (as a projection onto a polyhedron). We will see that N is *simple* in the sense of Section 6.4. Exploiting this fact and supposing that M is locally Lipschitz, one can determine the Thibault derivative TF and the contingent derivative CF (recall that they are crucial for strong, upper and pseudo-regularity of F, cf. Theorem 5.1 and Theorem 5.14) by the product rule of differentiation, cf. the properties (6.5) ... (6.8) of Section 6.4. We shall see below that these derivatives can be (more or less explicitly) described and interpreted in terms of the original functions involved. This is the most hard and interesting task for applying abstract stability statements formulated in *any* equivalent setting.

Clearly, if one of the input functions in M is a complicated Lipschitz function, the remaining problem may be serious. On the other hand, if M is even continuously differentiable (in particular, if M corresponds to the classical Kojima–function (7.3) with $f, g, h \in C^2$), then only the function $\tilde{N}(y) = (y^+, y^-)$ is nonsmooth, and F becomes piecewise smooth. If, in addition, at some zero $s^0 = (x^0, y, z)$ the *strict complementarity condition* $y_i \neq 0\ \forall i$ holds true, also $\tilde{N}(\cdot)$ is smooth near y, and the usual implicit function theorem is applicable in order to study regularity of F: In fact, then all our regularities (strong, pseudo, upper) coincide and are satisfied if and only if $DF(s^0)$ is nonsingular. However, in the following we are just interested in the opposite (non–standard) case.

Having (pseudo, strong) regularity of F at $s^0 = (x^0, y^0, z^0)$, we may modify all the functions in (7.8) as long as the variations remain sufficiently $C^{0,1}$–small (see the Theorems 4.1 and 4.3 concerning existence and behavior of related zeros). If we vary only M, but not the fixed elements 0 and 1 in (7.6), then we are still in the classical framework of parametric optimization. *But what happens* after changing the functions y^+ and y^- in $N(\cdot)$ or the other elements of M? The result and interpretations of such perturbations will be studied in Chapter 11 for usual critical points.

For many stability statements, the assumed *constraint qualifications* are of crucial importance. In the regularity context, they clarify the local behavior of the feasible points under right–hand side perturbations or, from the dual point of view, the behavior of the related normal cone map. Mostly, one supposes MFCQ or the more restrictive *Linear Independence constraint qualification (LICQ)* which says that the gradients of active constraints are linearly independent.

Therefore, before we consider particular cases and apply rules for differentiating F, let us state that LICQ is a necessary consequence of pseudo-regularity for generalized Kojima functions under quite weak assumptions. Our proof directly indicates which perturbations disturb the pseudo–Lipschitz property, provided LICQ is violated.

Lemma 7.1 (necessity of LICQ for pseudo-regularity). *Let $F = NM$ be a generalized Kojima function, where Φ, Ψ and Γ are continuous and $g, h \in C^1$. Then, F is pseudo-regular at some zero $s^0 = (x^0, y^0, z^0)$ only if the gradients of active constraints $\{Dh_k(x^0) | 1 \le k \le \kappa\} \cup \{Dg_i(x^0) | g_i(x^0) = 0\}$ are linearly independent.* ◇

Proof. Without loss of generality we may assume $g(x^0) = 0$. For simplicity, put $g_{m+k} := h_k$, $\Psi_{m+k} := \Gamma_k$ and $y_{m+k} := z_k$ for $1 \le k \le \kappa$, and omit the summation index $i = 1, ..., l := m + \kappa$. For $\mathbf{1} := (1, ..., 1)^\mathsf{T} \in \mathbb{R}^l$ and $\varepsilon > 0$, we set

$$a(\varepsilon) = \varepsilon \sum \Psi_i(x^0).$$

Then, the point $(x^0, y^0 + \varepsilon \mathbf{1})$ solves $F_1 = a(\varepsilon)$, $(F_2, F_3) = 0$.

Next let any α satisfy

$$\sum \alpha_i Dg_i(x^0) = 0.$$

We will show that $\alpha = 0$. To this end, with $\delta > 0$, consider solutions (x, y) to

$$F_1 = a(\varepsilon), (F_2, F_3) = \delta\, \alpha.$$

For small fixed ε, and for δ tending to zero, now there are such solutions (x, y) which satisfy the pseudo-Lipschitz inequality

$$\|(x, y) - (x^0, y^0 + \varepsilon \mathbf{1})\| \le L\delta \|\alpha\|. \tag{7.9}$$

The first m components of $y^0 + \varepsilon \mathbf{1}$ are not smaller than ε. Hence, it holds $y_i > 0$ for $1 \le i \le m$ and small δ. This implies

$$\delta \alpha_i = g_i(x) \text{ for } 1 \le i \le m.$$

For equality constraints $g_i, i > m$, $\delta\alpha_i = g_i(x)$ is trivially true. By the mean-value theorem and due to $g(x^0) = 0$, we derive the identities

$$g_i(x) = Dg_i(\theta_i)(x - x^0), \quad \text{where } \theta_i \to x^0 \text{ (as } \delta \to 0).$$

Therefore,

$$\delta\|\alpha\|^2 = \langle \alpha, \delta\alpha \rangle = \langle \alpha, g(x) \rangle = \sum \alpha_i Dg_i(\theta_i)(x - x^0).$$

Since Dg is continuous and $\delta \to 0$, there holds

$$\sum \alpha_i Dg_i(\theta_i) \to \sum \alpha_i Dg_i(x^0) = 0.$$

Recalling (7.9) one thus observes that

$$\|x - x^0\| \le L\delta\|\alpha\|$$

and hence

$$\delta\|\alpha\|^2 = \|\sum \alpha_i Dg_i(\theta_i)(x - x^0)\| \le \|\sum \alpha_i Dg_i(\theta_i)\|\|x - x^0\| = o(\delta)$$

for arbitrarily small δ. But this inequality yields $\alpha = 0$. □

7.2 Examples and Canonical Parametrizations

In this section, we discuss relations to other settings of critical point systems. In particular, we give typical examples of generalized Kojima–functions and show which kinds of parametrization appear if the system $F(x) = 0$ is perturbed in the right–hand side.

The Subdifferential of a Convex Maximum Function

The two basic notions of convex analysis, the conjugate f^* and the subdifferential ∂f of a given convex functional f on $\mathbb{R}^n$, just describe key quantities of some parametric optimization problem. So $f^*(x^*)$ yields the infimum value, and $(\partial f)^{-1}(x^*)$ the minimizers of the perturbed function $f(\xi) - \langle x^*, \xi \rangle$.

Strong regularity of $\partial f(\cdot)$ at $(x^0, 0)$ means Lipschitz continuity and uniqueness of $(\partial f)^{-1}$ near $(0, x^0)$. Further, we know that strong and pseudo-regularity here coincide, cf. Theorem 5.4. The inclusion $x^* \in \partial f(x)$ cannot be modeled by Kojima functions, in general. But for solving it (or for any analysis of related solutions), one needs information concerning f. At this stage, Kojima functions come into the play. To see what happens we regard one of the simplest cases.

Let f be a maximum function on $\mathbb{R}^n$, i.e., $f = \max_{1 \le i \le m} f^i$, where all f^i are convex functions belonging to the class $C^{1,1}$. Then

$$\partial f(x) = \operatorname{conv}\{Df^i(x) | f^i(x) = f(x)\}$$

and

$$0 \in \partial f(x) \quad \Leftrightarrow\ 0 = \sum_{i=1}^m y_i^+ Df^i(x),\ f(x) + y_i^- = f^i(x),\ \sum_{i=1}^m y_i^+ = 1.$$

Setting

$$h(x,z) := z \in \mathbb{R} \qquad \text{and} \qquad g_i(x,z) := f^i(x) - z,$$

now we have $D_z g_i(x,z) = -1$, and our conditions attain the form

$$0 = \sum_{i=1}^m y_i^+ D_x g_i(x,z),\ \ -1 = \sum_{i=1}^m y_i^+ D_z g_i(x,z),\ \ y_i^- = g_i(x,z),$$

i.e.,

$$Dh(x,z) + \sum_{i=1}^m y_i^+ Dg_i(x,z) = 0,\ \ g_i(x,z) - y_i^- = 0. \tag{7.10}$$

These equations form the Kojima system of the problem

$$\min_{(x,z)}\{z | f^i(x) - z \le 0,\ i = 1,\dots,m\}, \tag{7.11}$$

this is a program with convex and smooth $C^{1,1}$ data.

Strong regularity of the system (7.10) means local Lipschitz continuity of the *unique* solutions to

$$Dh(x,z) + \sum_{i=1}^m y_i^+ Dg_i(x,z) = (a,\alpha),\ \ g_i(x,z) - y_i^- = b_i \tag{7.12}$$

for the parameter (α, a, b) having small norm. The latter is the Lipschitz property of *unique primal–dual* solutions to the parametric problem

$$\min\{(1-\alpha)z - \langle a, x\rangle \mid f^i(x) - z \le b_i,\ i = 1,\dots,m\}. \tag{7.13}$$

According to Lemma 7.1, strong regularity of system (7.10) at (x^0, y^0, z^0) requires LICQ. In the current case for x^0 and $z^0 = f(x^0)$, LICQ means: The vectors $\{Df^i(x^0) | i \in I^0\}$ have to be affinely independent, where

$$I^0 = \{i | f^i(x^0) = f(x^0)\}.$$

It is worth noting that the weaker MFCQ requires the existence of $(u,\zeta) \in \mathbb{R}^{n+1}$ such that $Df^i(x^0)u - \zeta < 0\ \forall i \in I^0$ and is always satisfied.

In contrast to the initial situation when studying the subdifferential mapping ∂f, in system (7.12) additional parameters α, b_i appear, and we are speaking about dual solutions, too. By variation of b_i , we may modify the functions f_i separately such that $f_{\text{new}}(x) = \max_i(f^i(x) - b_i)$. This was an impossible variation as long as we considered f without any inner structure. Keeping $b = 0$ fixed we just return to our initial question of strong regularity of ∂f.

The question arises whether the two forms of strong regularity are indeed different, or not. The answer is yes.

Remark 7.2 (strong regularity of ∂f).

1. Clearly, if the Kojima function in (7.10) is strongly regular then so is ∂f.
2. On the other hand, ∂f may be strongly regular while the Kojima system (7.10) does not so. This case happens if and only if ∂f is strongly regular at x^0 and $\{Df^i(x^0) | i \in I^0\}$ are affinely dependent. ◇

To see [2.], we first recall that under strong regularity of (7.10), LICQ holds because of Lemma 7.1. Conversely, let ∂f be strongly regular at x^0, and, in addition, let LICQ be true. Then x^0 is isolated in $(\partial f)^{-1}(0)$, and the (uniform) growth condition holds, see Lemma 3.1 and Theorem 4.8. So (7.13) is still solvable with solutions $(x(a, \alpha, b),\ z(a, \alpha, b))$ near $(x^0, z^0) = (x^0, f(x^0))$ for small parameters (because $(\partial f)^{-1}$ is l.s.c. at 0). By LICQ, the duals $y(a, \alpha, b, x, z)$ in (7.12) uniquely exist and are Lipschitz in the indicated variables. This is just strong regularity of (7.10).

Summary. Strong regularity of ∂f means only uniform growth of f (and can evidently hold even if all f^i coincide). Strong regularity of system (7.10) means just both, uniform growth and LICQ at the solution to problem (7.11).

Complementarity Problems

Given locally Lipschitz functions $u, v : \mathbb{R}^n \to \mathbb{R}^n$, find x such that

$$u(x) \geq 0,\ v(x) \geq 0 \quad \text{and} \quad \langle u(x), v(x)\rangle = 0. \tag{7.14}$$

With $y \in \mathbb{R}^n$, we rewrite the conditions as $u(x) = y^+$, $v(x) = -y^-$ which yields the equation

$$F(x, y) = 0, \qquad \text{where } F_1(x, y) = u(x) - y^+,\ F_2(x, y) = -v(x) - y^-.$$

In fact, F is a generalized Kojima-function on $\mathbb{R}^{2n}$:

$$\Phi(x) = u(x),\ \Psi(x) = -E,\ g(x) = -v(x),$$

where E is again the (n, n) identity matrix. Clearly, z and F_3 do not appear. The perturbed system $F_1 = a$, $F_2 = b$ means

$$u(x) - y^+ = a \quad \text{and} \quad -v(x) - y^- = b \tag{7.15}$$

and describes a natural parametrization of the original problem with parameters a and b: Find x such that

$$u(x) \geq a,\ v(x) \geq -b \quad \text{and} \quad \langle u(x) - a\, , v(x) + b\rangle = 0. \tag{7.16}$$

In the *standard* case $v(x) = x$, we obtain a generalized Kojima–function where $\Psi = Dg$, i.e., in comparison with the usual Kojima–function (7.3) only Df is replaced by $\Phi = u$. Explicitly,

$$F_{1i} = u_i - y_i^+,\quad F_{2i} = -x_i - y_i^-.$$

The special relation $\Psi = Dg$ becomes interesting, e.g., in Lemma 7.18.

Several other descriptions of the complementarity problem are possible, for example, via an optimization problem. Instead of (7.16), one may investigate the optimization problem

$$\min \langle u(x), v(x)\rangle \ \text{ s.t. } \ -u(x) \le 0, \ -v(x) \le 0, \tag{7.17}$$

provided the original problem (7.14) is solvable and $u, v \in C^1$. The right–hand side perturbations of the related Kojima-system lead us to

$$\langle u, Dv\rangle + \langle v, Du\rangle - y^+ - z^+ = c, \ -u - y^- = a, \ -v - z^- = b.$$

The system describes the critical points of the perturbed problem

$$\min \langle u(x), v(x)\rangle - \langle c, x\rangle, \quad \text{s.t.} \quad u(x) \ge a, \ v(x) \ge b$$

with parameters a, b, c. Compared with (7.16), an interpretation in the context of complementarity is now less obvious and the analytical form of the Kojima-system is more complicated. *Concerning approaches via so-called NCP functions we refer to Chapter 9.*

Generalized Equations

Given any closed, convex set $C \subset \mathbb{R}^d$ and any continuous function $H : \mathbb{R}^d \to \mathbb{R}^d$, a *generalized equation* (written in traditional form as variational inequality) claims to find some $s \in C$ such that

$$\langle H(s), c - s\rangle \le 0 \qquad \forall c \in C.$$

After introducing the (contingent) *normal cone*

$$N_C(s) = \{\zeta \in \mathbb{R}^d | \langle \zeta, c - s\rangle \le 0 \ \forall c \in C\} \quad (s \in C),$$

this means

$$H(s) \in N_C(s), \ s \in C. \tag{7.18}$$

The introduction of generalized equations for the unified study of KKT systems, complementarity problems and equilibrium problems is due to S.M. Robinson (see, e.g., [Rob80, Rob82]).

The general equation (7.18) becomes an *equation* by writing $N_C(\cdot)$ in terms of the Euclidean projection onto C. But more interesting, let us suppose that some analytic description of C is given, say C is polyhedral,

$$C = \{s | As \le \alpha\},$$

with some suitable matrix A and vector α. Taking the particular form of the normal cone into account, (7.18) is equivalent to

$$H(s) = y^+ A, \quad As - \alpha = y^- \tag{7.19}$$

for some y. With $g(s) = As - \alpha$, this is a generalized Kojima system with $\Phi = -H$, $\Psi = Dg$, and F_3 does not appear. The related parametric equation $F = (F_1, F_2) = (a, b)$ now becomes

$$-H(s) + y^+ A = a, \quad As - \alpha - y^- = b \tag{7.20}$$

This system characterizes, by putting $C(b) := \{s | As - \alpha \leq b\}$, the solutions of the parametric generalized equation

$$H(s) + a \in N_{C(b)}(s), \quad s \in C(b) \tag{7.21}$$

with parameters a, b. Here, the feasible set C is no longer constant.

When studying (7.21) *directly in a general framework* (e.g. by the tools of Chapter 2), new difficulties will appear because N_C depends on b. However, knowing that (7.21) is only a perturbed generalized Kojima-system in the sense that

$$s \text{ satisfies (7.21)} \quad \Leftrightarrow \quad s \text{ satisfies (7.20) with some } y$$

keeps the things simpler since the consideration of multifunctions can be avoided at all.

On the other hand, the traditional form of parametrizing (7.18) according to Robinson's work [Rob80, Rob82] leads to a problem where C remains fixed:

$$H(s) + p \in N_C(s), \quad s \in C \qquad \text{(parameter } p). \tag{7.22}$$

So the parametrizations (7.21) and (7.22) are dealing with different subjects, both closely related to the original problem. A similar situation was discussed above concerning subdifferentials. Again, with the same arguments as for the subdifferential, *linear independence* of all active gradients (i.e., of all A_i satisfying $A_i s^0 = \alpha_i$ for the solution (s^0, y^0) to (7.19) under consideration) is the crucial condition which ensures that strong regularity in Robinson's sense (i.e., the existence of locally unique and Lipschitz solutions to (7.22)) implies strong regularity with respect to the parametrization (7.21).

In particular, if the generalized equation (7.18) describes KKT points of the nonlinear program (7.1), i.e. if (7.18) has the particular form

$$\begin{array}{lcl} Df(x) + \sum_{i=1}^m \eta_i Dg_i(x) + \sum_{k=1}^\kappa z_k Dh_k(x) & = & 0 \\ g(x) & \in & N_{\mathbb{R}^m_+}(y) \\ h(x) & = & 0, \end{array} \tag{7.23}$$

then the parameterization (7.22) with $p = -(a, b, c)$ coincides with the right–hand side perturbation of the Kojima function (7.3) due to the special kind of H and C. Both parameterizations now describe exactly the KKT points of

$$\min f(x) - \langle a, x \rangle \;\; \text{s.t.} \;\; g(x) \leq b, \; h(x) = c, \tag{7.24}$$

which follows easily from the particular form of C. Besides, strong regularity with respect to (7.22) implies here strong regularity with respect to (7.21) because the linear independence condition concerning A holds true; not that $C = \{(x, \eta, z) | -E\eta \le 0\}$.

Finally, if C is not polyhedral (and/or non–convex), but say C is described by $C = \{s | g(s) \le 0\}$, $g : \mathbb{R}^d \to \mathbb{R}^k$ with $g \in C^1$, the situation is similar. The interesting set $N_C(s)$ in (7.18) is now the normal cone of the contingent cone to C at s. Under a constraint qualification like MFCQ, there holds

$$N_C(s) = \{y^+ Dg(s) \,|\, y \text{ such that } g(s) = y^-\}.$$

So, setting $F = (-H(s) + y^+ Dg(s), g(s) - y^-)$, equation (7.20) passes into

$$-H(s) + y^+ Dg(s) = a, \quad g(s) - y^- = b. \tag{7.25}$$

Clearly, F is again of generalized Kojima-type, where, $\Phi := K$ and $\Psi = Dg$, i.e., in comparison with the usual Kojima function only Df is replaced by $-H$.

Nash Equilibria

Consider the following problem of a *(Nash) equilibrium* of the *non-antagonistic 2-person game* (u, v, X, Y): Given continuously differentiable functions $u, v : \mathbb{R}^{n+m} \to \mathbb{R}$ and compact convex sets X, Y in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, find $(x^0, y^0) \in X \times Y$ such that

$$\begin{array}{ll} u(x^0, y^0) \ge u(x, y^0) & \forall x \in X \\ v(x^0, y^0) \ge v(x^0, y) & \forall y \in Y. \end{array} \tag{7.26}$$

Writing down the first order necessary optimality conditions for both optimization problems, we get the generalized equation

$$\begin{array}{rcl} D_x u(x^0, y^0) & \in & N_X(x^0) \\ D_y v(x^0, y^0) & \in & N_Y(y^0) \end{array}$$

where N_X, N_Y are again the usual normal cone maps.

For seek of simplicity, let X be the unit simplex of $\mathbb{R}^n$, i.e., $X = \{x | 1_n{}^{\mathsf T} x = 1,\ x \ge 0\}$ (where $1_n{}^{\mathsf T} := (1, \ldots, 1) \in \mathbb{R}^n$), and let Y be the unit simplex of $\mathbb{R}^m$. Further, let r, s be the dual vectors associated with $-x \le 0$ and $-y \le 0$, respectively, and let $p, q \in \mathbb{R}$ be the dual variables with respect to the equations. Then the critical point system takes the form

$$\begin{array}{lcrcl} -D_x u(x, y) & - & r^+ + p\, 1_n & = & 0, \\ -D_y v(x, y) & - & s^+ + q\, 1_m & = & 0, \\ -x - r^- = 0, & & 1_n{}^{\mathsf T} x - 1 & = & 0, \\ -y - s^- = 0, & & 1_m{}^{\mathsf T} y - 1 & = & 0. \end{array} \tag{7.27}$$

This system can be rewritten as $F = 0$ by means of a generalized Kojima-function F with

$$\Phi(x, y) = (-D_x u(x, y), -D_y v(x, y)).$$

In (7.6), the matrix M depends now on the *strategies* x and y, and the vector N depends on the dual vector (r, s, p, q). The related system $F = c$ with right-hand side perturbations $c = (a, b, \alpha, \beta, \xi, \eta)$ leads us to linear perturbations of the *utilities* in the player's variables as

$$u(x,y) - a^{\mathsf{T}}x, \quad v(x,y) - b^{\mathsf{T}}y$$

and claims perturbations of the simplexes X and Y as well: $x \geq \xi$, $1_n^{\mathsf{T}}x = 1+\alpha$, $y \geq \eta$, $1_m^{\mathsf{T}}y = 1 + \beta$. Similarly, one can handle games of more players.

Piecewise Affine Bijections

If F is the usual Kojima–function of a quadratic optimization problem, say

$$\min\{x^{\mathsf{T}}Qx + a^{\mathsf{T}}x \mid Ax \leq b, \; Bx = c\},$$

then Ψ and Γ are constant and Φ, g, h are linear. The investigation of F^{-1} then leads us to piecewise linear systems studied by Kuhn and Löwen [KL87]. Basic extensions of their results to the case of PC^1-equations can be found in [JP88, Sch94, PR96, RS97].

7.3 Derivatives and Regularity of Generalized Kojima–Functions

The possibilities for computing relevant derivatives are important for any analysis of nonsmooth functions, in particular also for generalized Kojima–functions. Because of the crucial role which the Thibault derivative TF and the contingent derivative CF are playing for strong, pseudo– and upper regularity, it is desirable to have an explicit and intrinsic description of these mappings in terms of the original functions. We shall derive such descriptions and utilize them both for characterizing regularity and solving Kojima systems.

Properties of N

Recall that $F(x, y, z) = N(y, z)M(x)$, where $N(y, z)$ is the row vector

$$N(y,z) = (1, y^+, y^-, z),$$

and $M(x)$ is defined according to (7.6), i.e.,

$$M(x) = \begin{pmatrix} \Phi(x) & g(x)^{\mathsf{T}} & h(x)^{\mathsf{T}} \\ \Psi(x) & 0 & 0 \\ 0 & -E & 0 \\ \Gamma(x) & 0 & 0 \end{pmatrix}.$$

Below we shall show that N is simple, hence, we can apply Theorem 6.8 and the Corollaries 6.9 and 6.10 in order to represent TF and CF. We will see that

$CN \neq TN$, and that, even if M is arbitrarily smooth, we will get $CF \neq TF$ in general.

However, the only part interesting in this context is the piece-wise linear function

$$y \mapsto \tilde{N}(y) := (y^+, y^-) = (y^+, y - y^+) \in \mathbb{R}^{2m}.$$

Its derivatives $T\tilde{N}(y^0)(v)$ and $C\tilde{N}(y^0)(v) = \tilde{N}'(y^0; v)$ can be written down by considering the components separately.

Obviously, by defining

$$\mathcal{R}_T(y^0) := \{\ r \in [0,1]^m \ \mid\ r_i = 1 \text{ if } y_i^0 > 0, \quad r_i = 0 \text{ if } y_i^0 < 0\ \}, \tag{7.28}$$

the set $T\tilde{N}(y^0)(v)$ consists of all vectors

$$(\alpha, v - \alpha) \text{ satisfying } \alpha_i = r_i v_i\ (\forall i) \text{ with some } r \in \mathcal{R}_T(y^0).$$

The reader easily sees that $\mathcal{R}_T(y^0) = \partial(y^+)(y^0)$ and hence $TN(y^0, z^0)(\cdot)$ coincides with the multivalued directional derivative $\partial N(y^0, z^0)(\cdot)$ in terms of Clarke's generalized Jacobian. Setting similarly,

$$\mathcal{R}_C(y^0, v) := \left\{ r \in \{0,1\}^m \ \middle|\ \begin{array}{l} r_i = 1 \text{ if } y_i^0 > 0 \text{ or if } y_i^0 = 0,\ v_i \geq 0; \\ r_i = 0 \text{ if } y_i^0 < 0 \text{ or if } y_i^0 = 0,\ v_i < 0 \end{array} \right\}, \tag{7.29}$$

we obtain a singleton and

$$\tilde{N}'(y^0; v) = (\alpha, v - \alpha), \quad \text{where } \alpha_i = r_i v_i\ (\forall i) \text{ and } \mathcal{R}_C(y^0, v) = \{r\}.$$

For $y_i^0 = 0$, now r_i is simply the directional derivative of $y_i \mapsto y_i^+$ at 0 in direction $\operatorname{sign} v_i$. Hence, *CN* is the directional derivative of *N*. Trivially, $\mathcal{R}_C(y^0, v) \subset \mathcal{R}_T(y^0)$.

Some Transformation

For several reasons, it makes sense to rewrite the representations (7.28) and (7.29) by using the transformation

$$\alpha_i = r_i v_i, \quad \beta_i = (1 - r_i) v_i \quad (\forall i) \tag{7.30}$$

with $r \in \mathcal{R}_T(y^0)$ or $r \in \mathcal{R}_C(y^0, v)$. The reader easily sees that

(I) if $r \in \mathcal{R}_T(y^0)$ then $(\alpha, \beta) \in \mathcal{J}_T(y^0)$,

(II) if $r \in \mathcal{R}_C(y^0, v)$ then $(\alpha, \beta) \in \mathcal{J}_C(y^0)$,

where

$$\mathcal{J}_T(y^0) := \left\{ (\alpha, \beta) \in \mathbb{R}^{2m} \ \middle|\ \begin{array}{lcll} \alpha_i \beta_i & \geq & 0 & \text{if } y_i^0 = 0 \\ \beta_i & = & 0 & \text{if } y_i^0 > 0 \\ \alpha_i & = & 0 & \text{if } y_i^0 < 0 \end{array} \right\} \tag{7.31}$$

and

$$\mathcal{J}_C(y^0) := \left\{(\alpha,\beta) \in \mathcal{J}_T(y^0) \,\middle|\, \alpha_i \ge 0 \ge \beta_i \text{ if } y_i^0 = 0 \right\}. \tag{7.32}$$

Conversely, if α and β satisfy the conditions under (I) or (II), we define v and r by

$$v = \alpha + \beta, \quad r_i = a_i/v_i \text{ if } v_i \ne 0, \quad r_i = 1 \text{ if } v_i = 0. \tag{7.33}$$

Then, one concludes in an elementary way that $r \in \mathcal{R}_T(y^0)$ and $r \in \mathcal{R}_C(y^0, v)$, respectively. Moreover, in both cases, the one-to-one correspondences $(v, r) \leftrightarrow (\alpha, \beta)$ satisfy

$$v = \alpha + \beta \qquad \text{and} \qquad v = 0 \Leftrightarrow (\alpha,\beta) = 0. \tag{7.34}$$

The latter will be important for studying the injectivity of the derivatives *TF* and *CF* of the generalized Kojima-function.

Derivatives of N

In the following lemma, we summarize representations of *TN* and *CN* which immediately follow from the discussion above.

Lemma 7.3 *(TN, CN). Let* $\lambda^0 = (y^0, z^0)$ *and* $\mu = (v, w) \in \mathbb{R}^{m+\kappa}$.
Then, $TN(\lambda^0)(\mu)$ *has the following representations:*

$$TN(\lambda^0)(\mu) = \{(0, \alpha, v - \alpha, w) \mid \alpha_i = r_i v_i \, (\forall i), \ r \in \mathcal{R}_T(y^0)\},$$
$$TN(\lambda^0)(\mu) = \{(0, \alpha, \beta, w) \mid (\alpha, \beta) \in \mathcal{J}_T(y^0), \ v = \alpha + \beta\},$$
$$TN(\lambda^0)(\mu) = \mu\, \partial N(\lambda^0) = \{\mu J \mid J \in \partial N(\lambda^0)\}.$$

Moreover, $CN(\lambda^0)(\mu)$ *has the following representations:*

$$CN(\lambda^0)(\mu) = \{(0, \alpha, v - \alpha, w)\} \text{ with } \alpha_i = r_i v_i \ (\forall i), \ \mathcal{R}_C(y^0, v) = \{r\},$$
$$CN(\lambda^0)(\mu) = \{(0, \alpha, \beta, w)\} \text{ with } \mathcal{J}_C(y^0) = \{(\alpha, \beta)\}, \ v = \alpha + \beta,$$
$$CN(\lambda^0)(\mu) = N'(\lambda^0; \mu).$$

◇

Basic Lemma on N

Now we make sure that (the locally Lipschitz functions) $\tilde{N}$ and N are simple in the sense of Section 6.4 and have further useful properties which are based on their special structure.

For real y, the functions y^+ and y^- are monotone and satisfy $y^+ = \frac{1}{2}(y + |y|)$, $y^- = \frac{1}{2}(y - |y|)$. So the related derivatives of $\tilde{N} = (y^+, y^-)$ are component-wisely given by the "derivatives" of the absolute value function $y \mapsto |y|$. Let $\tilde{N}_i$ denote the $\mathbb{R}^2$ components (y_i^+, y_i^-) of $\tilde{N}$. Since $\tilde{N}(y)$ is component-wisely defined via independent variables, all the following statements must be shown only for each component. In addition, they are evident for $y_i^0 \ne 0$ (where $\tilde{N}_i$ is locally linear), and they also hold for N which differs from $\tilde{N}$ only by additional linear and independent components.

Lemma 7.4 (*N* simple, and further properties).

1. *The functions* $\tilde{N}$ *and N are simple everywhere.*

2. *Given* (y^0, z^0) *and* (v, w) *with* $\|v\| \le 1$, *one has for all* z *and small* $\varepsilon > 0$ *that* $0 < t < \varepsilon$ *and* $\|y - y^0\| < \varepsilon$ *imply the inclusions*

$$\begin{array}{llcll} \text{(a)} & t^{-1}[N(y+tv, z+tw) & - & N(y,z)] & \in & TN(y^0,z^0)(v,w), \\ \text{(b)} & t^{-1}[N(y^0+tv, z^0+tw) & - & N(y^0,z^0)] & \in & N'((y^0,z^0);(v,w)). \end{array}$$

◇

Proof. 1. Obviously, N is simple if and only if so is $\tilde{N} = \tilde{N}(y)$. Given arbitrary y^0 and v, let $(\alpha, v - \alpha)$ belong to $\tilde{N}(y^0)(v)$ and let $t = t_k \downarrow 0$ be any given sequence. To show that $\tilde{N}$ is simple, we have to construct elements $y = y(t) \to y^0$ in such a way that

$$(\alpha, v - \alpha) = \lim t^{-1}(\tilde{N}(y + tv) - \tilde{N}(y)), \tag{7.35}$$

with the chosen sequence of t (or with some infinite subsequence). To do this, write $\alpha_i = r_i v_i$ with $r \in \mathcal{R}_T(y^0)$ and define $y_i = y_i^0 - t\lambda_i v_i$, where

$$\lambda_i := \begin{cases} 0 & \text{if (i)} \quad y_i^0 \neq 0, \\ 1 - r_i & \text{if (ii)} \quad y_i^0 = 0 \text{ and } v_i > 0, \\ r_i & \text{if (iii)} \quad y_i^0 = 0 \text{ and } v_i \le 0. \end{cases}$$

Now one easily determines in case of

$$\begin{array}{ll} \text{(i)} & (y_i + tv_i)^+ - y_i^+ = t\alpha_i, \\ \text{(ii)} & (y_i + tv_i)^+ - y_i^+ = r_i t v_i - 0 = t\alpha_i, \\ \text{(iii)} & (y_i + tv_i)^+ - y_i^+ = ((1 - r_i)tv_i)^+ - (-r_i t v_i)^+ = 0 - (-r_i t v_i) = t\alpha_i. \end{array}$$

Therefore, even

$$t^{-1}[\tilde{N}(y + tv) - \tilde{N}(y)] = (\alpha, v - \alpha)$$

holds for the constructed sequence of $y \to y^0$.

2. The proof of (a) and (b) is ensured by piecewise linearity and is left to the reader. □

By construction, the points y defined under (i), (ii), (iii) in the previous proof belong to a line given by r and v.

Conventions

Throughout the rest of this section, we suppose that

$$s^0 = (x^0, y^0, z^0) \text{ and } \sigma = (u, v, w) \text{ are fixed in } \mathbb{R}^d, \tag{7.36}$$

where $d = n + m + \kappa$. Further, we shall use the abbreviations

$$\Phi^\circ := \Phi(x^0), \quad \Psi^\circ := \Psi(x^0), \quad \Gamma^\circ := \Gamma(x^0), \tag{7.37}$$

and we write

$$I^0 := \{i|y_i^0 = 0\},\ I^+ := \{i|y_i^0 > 0\},\ I^- := \{i|y_i^0 < 0\} \tag{7.38}$$

and

$$\textstyle\sum_i y_i = \sum_{i=1}^m y_i, \qquad \sum_k z_k = \sum_{k=1}^\kappa z_k. \tag{7.39}$$

Moreover, we put for $u \in \mathbb{R}^n$,

$$Q_T(u) := T_x F_1(s^0)(u) \text{ and } Q_C(u) := C_x F_1(s^0)(u), \tag{7.40}$$

where F_1 is the first component of F, i.e., $F_1(x,y,z) = \Phi(x) + \sum_{i=1}^m y_i^+ \Psi_i(x) + \sum_{k=1}^\kappa z_k \Gamma_k(x)$, while $T_x F_1(s^0)(u)$ denotes the partial Thibault derivative of F_1 at s^0 with respect to x in direction u. Analogously, $C_x F_1(s^0)(u)$ denotes the corresponding partial contingent derivatives.

Recall again that for the optimization model (7.1), s^0 corresponds to a primal–dual vector, while $\Phi(x) = Df(x)$, $\Phi° = Df(x^0)$, $\Psi_i° = Dg_i(x^0)$ and $\Gamma_k° = Dh_k(x^0)$.

Formulas for Generalized Derivatives

Product Rules

From Lemma 7.4 and Lemma 7.3 we know that N is simple and directionally differentiable. Hence, if $M \in C^{0,1}$ then results of §6.4.1 immediately imply the partial differentiation and product rules for TF and CF. For completeness, we repeat here the product rules.

Theorem 7.5 (*TF, CF;* product rules). *Let $F = NM$ be a generalized Kojima–function according to (7.8), $s^0 = (x^0, y^0, z^0)$ and $\sigma = (u, v, w)$. Suppose $M \in C^{0,1}$.*

Then the Thibault derivative TF of F has the representation

$$TF(s^0)(\sigma) = N(y^0,z^0)TM(x^0)(u) + TN(y^0,z^0)(v,w)M(x^0). \tag{7.41}$$

Moreover, given $M_0 \in TM(x^0)(u)$ and $N_0 \in TN(y^0,z^0)(v,w)$, the three conditions

$$\begin{aligned} N(y^0,z^0)[M_0 + N_0 M(x^0)] &= \lim t^{-1}[F((x,y,z)+t\sigma) - F(x,y,z)], \\ M_0 &= \lim t^{-1}[M(x+tu) - M(x)], \\ N_0 &\equiv t^{-1}[N((y,z)+t(v,w)) - N(y,z)] \end{aligned} \tag{7.42}$$

can be satisfied with the same sequences $s^t = (x,y,z) \to s^0$ and $t = t_k \downarrow 0$, where $z \equiv z^0$, all y are located on a line and $\alpha_i = t^{-1}[(y_i + tv_i)^+ - y_i^+]$.
Finally, for the contingent derivative CF, the same statements are true, one has only to replace "T" with "C" and $s^t = (x,y,z)$ with s^0. ◇

Proof. Since N is a simple function according to Lemma 7.4, the statements concerning TF follow from Theorem 6.8 and Corollary 6.10. The contingent derivative can be determined similarly, one has to apply Corollary 6.12 instead. □

Explicit Formulas

Theorem 7.6 (*TF, CF;* explicit formulas). *Assume the hypotheses of Theorem 7.5, and suppose in addition that g, h are C^1-functions.*

Then, $TF(s^0)(\sigma)$ consists exactly of all vectors $p = (\xi, \eta, \zeta)$ such that

$$\begin{array}{lcll} \xi & \in & Q_T(u) + \sum_i r_i v_i \Psi_i{}^\circ + \sum_k w_k \Gamma_k{}^\circ & \\ \eta_i & = & Dg_i(x^0)u - (1 - r_i)v_i & \forall i \\ \zeta_k & = & Dh_k(x^0)u & \forall k \end{array} \tag{7.43}$$

holds with some $r \in \mathcal{R}_T(y^0)$.

Equivalently, $TF(s^0)(\sigma)$ consists exactly of all vectors $p = (\xi, \eta, \zeta)$ such that

$$\begin{array}{lcll} \xi & \in & Q_T(u) + \sum_i \alpha_i \Psi_i{}^\circ + \sum_k w_k \Gamma_k{}^\circ & \\ \eta_i & = & Dg_i(x^0)u - \beta_i & \forall i \\ \zeta_k & = & Dh_k(x^0)u & \forall k, \end{array} \tag{7.44}$$

holds with some $(\alpha, \beta) \in \mathcal{J}_T(y^0)$ satisfying $v = \alpha + \beta$.

Further, $CF(s^0)(\sigma)$ consists exactly of all vectors $p = (\xi, \eta, \zeta)$ such that (7.43) or (7.44) hold after replacing $Q_T(u)$ with $Q_C(u)$ as well as $\mathcal{R}_T(y^0)$ and $\mathcal{J}_T(y^0)$ with $\mathcal{R}_C(y^0, v)$ and $\mathcal{J}_C(y^0)$, respectively. In this case, the elements $r \in \mathcal{R}_C(y^0, v)$ and $(\alpha, \beta) \in \mathcal{J}_C(y^0)$ with $\alpha + \beta = v$ are unique. ◇

Note. The vectors $(u, v, w, \xi, \eta, \zeta, \alpha, \beta)$ satisfying (7.44) with $v = \alpha + \beta$ describe exactly the set gph $TF^L(s^0)$, where F^L is defined by $F^L(s) := (F(s), y^+, y^-)$ for $s = (x, y, z)$. The analogous result for gph $CF^L(s^0)$ holds after replacing $Q_T(u)$ with $Q_C(u)$ and $\mathcal{J}_T(y^0)$ with $\mathcal{J}_C(y^0)$. ◇

Proof of Theorem 7.6. To show (7.43), we have still to determine the terms of the product rule (7.41) by using that $g, h \in C^1$. Recall that M has the structure (7.6). Since $g, h \in C^1$, the derivative $TM(x^0)(u)$ yields (independently of concrete choices of $x \to x^0$ and $t \downarrow 0$) with respect to the columns 2 and 3 of M the submatrix

$$\begin{pmatrix} Dg(x^0)u & Dh(x^0)u \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The related limits of column 1 are just given by all elements

$$(\phi, \psi, 0, \gamma) \in T(\Phi, \Psi, 0, \Gamma)(x^0)(u).$$

So, the vectors of the product $N(y^0, z^0)TM(x^0)(u)$, in (7.41), have exactly the form

$$(\phi + \langle y^{0+}, \psi \rangle + \langle z^0, \gamma \rangle, \, Dg(x^0)u \, , \, Dh(x^0)u),$$

where the three blocks are assigned to the components F_1, F_2 and F_3 of F. By the chain rule (6.19), the elements

$$\phi + \langle y^{0+}, \psi \rangle + \langle z^0, \gamma \rangle$$

are also forming the set $Q_T(u) = TF_1(\cdot, y^0, z^0)(x^0)(u)$, since

$$F_1(x, y^0, z^0) = \Phi(x) + \textstyle\sum_i y_i^{0+} \Psi_i(x) + \sum_k z_k^0 \gamma_k(x).$$

By the structure of TN according to Lemma 7.3, the second term

$$TN(y^0, z^0)(v, w)M(x^0)$$

assigns to F_1, $F_{2\nu}$ and F_3 all triples

$$(\textstyle\sum_i r_i v_i \Psi_i{}^\circ + \sum_k w_k \Gamma_k{}^\circ \,,\, -(1 - r_\nu)v_\nu \,,\, 0) \in \mathbb{R}^{n+1+\kappa},$$

where $r \in \mathcal{R}_T(y^0)$. Thus, the explicit formula (7.43) follows from the chain rules.

The equivalent formula (7.44) follows from (7.43) and Lemma 7.3 based on the transformations (7.30). The explicit formulas for CF can be shown similarly, according to Theorem 7.5. □

The main point of the above proof was the product rule $TF = T(NM) = N(TM) + (TN)M$, Since our way via the rather general Theorem 6.8 and its corollaries is quite long we add a direct proof in the appendix as Lemma A.6 which is valid for the actual product rule only.

Let us emphasize that in order to apply the explicit formulas, we supposed (at least) that $M \in C^{0,1}$ and $g, h \in C^1$. The foregoing theorem will allow us to show that TF and CF are related in a very simple way, provided that $M(\cdot)$ is even a C^1 mapping. This will be done in §7.4.1.

Further notice that in the context of C^2**–optimization,** $Q_T(u)$ coincides with $Q_C(u)$ and reduces to a singleton, namely to Hu, where H is the Hessian $H = D^2_{xx}L(x^0, y^0, z^0)$.

Subspace Property

The following lemma will be important for deriving marginal value formulas. Suppose $M \in C^{0,1}$ and $g, h \in C^1$. Further, let us agree that

> T^- and C^- are the maps which assign to (ξ, η, ζ) all (u, v, w, r) satisfying the explicit formula (7.43) concerning TF and its modification concerning CF, respectively.

Then,

$$TF^{-1}(0, s^0)(\xi, \eta, \zeta) = \{(u, v, w) | (u, v, w, r) \in T^-(\xi, \eta, \zeta)\}.$$

If F is C^1 near s^0 and strongly regular, the map $(\xi, \eta, \zeta) \mapsto (u, v, w)$ plays the role of $DF(s^0)^{-1}$.

Generally, the component u, which characterizes the movement of the *stationary* solutions under variations of F, is not uniquely determined by T^-. However, in the essential case of $\Psi = Dg$ and $\Gamma = Dh$, the inner product $\langle\Phi(x^0), u\rangle$ is fix at critical points.

Lemma 7.7 (subspace property of TF^{-1}). *Let* $M \in C^{0,1}$, $g, h \in C^1$ *and* $F(s^0) = 0$. *Further, let* $\Psi = Dg$ *and* $\Gamma = Dh$. *Then,*

$$\langle\Phi(x^0), u\rangle = -\Big(\sum_{i=1}^{m} y_i^{0+}\eta_i \; + \; \sum_{k=1}^{\kappa} z_k^0\zeta_k\Big)$$

holds for all (u, v, w) *in* $TF^{-1}(0, s^0)(\xi, \eta, \zeta)$.

◇

Proof. Since $F(s^0) = 0$, we obtain from the explict formula that

$$\begin{aligned} 0 &= \langle\Phi(x^0), u\rangle + \textstyle\sum_{i=1}^m y_i^{0+}\langle\Psi_i{}^\circ, u\rangle + \sum_{k=1}^\kappa z_k\langle\Gamma_k{}^\circ, u\rangle \\ &= \langle\Phi(x^0), u\rangle + \big(\textstyle\sum_{i=1}^m y_i^{0+}\eta_i \; + \; \sum_{k=1}^\kappa z_k^0\zeta_k\big) \end{aligned}$$

is true. □

Evidently, the lemma similarly also holds for $CF^{-1} \subset TF^{-1}$.

Regularity Characterizations by Stability Systems

Using the explicit representations of *TF* and *CF* derived in the previous subsection, we shall characterize strong regularity of a generalized Kojima–function $F = NM$ and local upper Lipschitz continuity of its critical point mapping. From Theorem 5.14 we know that the generalized Kojima–function F is strongly regular at a zero s^0 of F if and only if $TF(s^0)$ is injective, i.e., if there is no nontrivial direction $\sigma = (u, v, w)$ such that $0 \in TF(s^0)(\sigma)$ holds. Further, the specialization of Lemma 3.2 to F says that the critical point mapping F^{-1} is locally upper Lipschitz at $(0, s^0)$ if and only if $CF(s^0)$ is injective, i.e., if $0 \in CF(s^0)(\sigma)$ implies $\sigma = 0$.

Hence, these injectivity properties can be verified by using the explicit formulas derived in Theorem 7.6: Given $s^0 = (x^0, y^0, z^0) \in F^{-1}(0)$, we have to look for solutions (u, α, β, w) of the *T–stability system*

$$\begin{aligned} &Q_T(u) \; + \; \textstyle\sum_i \alpha_i\Psi_i{}^\circ \; + \; \sum_k w_k\Gamma_k{}^\circ \; \ni \; 0, \\ &Dg(x^0)u - \beta = 0, \;\; Dh(x^0)u = 0, \;\; (\alpha, \beta) \in \mathcal{J}_T(y^0), \end{aligned} \tag{7.45}$$

or of the *C–stability system*

$$\begin{aligned} &Q_C(u) \; + \; \textstyle\sum_i \alpha_i\Psi_i{}^\circ \; + \; \sum_k w_k\Gamma_k{}^\circ \; \ni \; 0, \\ &Dg(x^0)u - \beta = 0, \;\; Dh(x^0)u = 0, \;\; (\alpha, \beta) \in \mathcal{J}_C(y^0). \end{aligned} \tag{7.46}$$

In the analysis of strong regularity, we shall also consider the problem of finding solutions (u, v, w, r) of the system related to (7.43),

$$\begin{array}{l} Q_T(u) \;+\; \sum_{i=1}^m r_i v_i \Psi_i{}^\circ \;+\; \sum_{k=1}^\kappa w_k\, \Gamma_k{}^\circ \;\ni\; 0, \\ Dg_i(x^0)u - (1-r_i)v_i = 0\,(\forall i),\; Dh_k(x^0)u = 0\,(\forall k),\; r \in \mathcal{R}_T(y^0), \end{array} \tag{7.47}$$

which is equivalent to the T–stability system. Now, the next theorem immediately follows.

Theorem 7.8 $(M(\cdot) \in C^{0,1})$. *Let $F = NM$ be a generalized Kojima–function according to Definition (7.8), and let s^0 be a zero of F. Suppose that $M \in C^{0,1}$ and $g, h \in C^1$. Then the following properties are equivalent:*

(i) F is strongly regular at s^0.

(ii) The T–stability system has only the trivial solution (i.e., $TF(s^0)$ is injective).

(iii) If (u, v, w, r) solves (7.47), then $(u, v, w) = (0, 0, 0)$.

Moreover, F^{-1} is locally upper Lipschitz at $(0, s^0)$ if and only if the C–stability system has only the trivial solution (i.e., $CF(s^0)$ is injective). ◇

Geometrical Interpretation

Under the assumptions of Theorem 7.8, *the T-stabsility system* permits a geometrical interpretation of strong regularity, first given in [Kum91b, §5] for usual Kojima functions. Set

$$U_h = \{u \in \mathbb{R}^n | Dh(x^0)u = 0\},$$

and let X_Γ be the subspace of $\mathbb{R}^n$, generated by the vectors $\Gamma_k{}^\circ$, $1 \le k \le \kappa$, where we put $X_\Gamma = \{0\}$ if no equations are required. Define the *(large) tangent space*

$$U^T = U_h \cap \{u \in \mathbb{R}^n | \langle Dg_i(x^0), u\rangle = 0 \;\forall i \in I^+\},$$

and introduce the polyhedral cone

$$K^T(u) = X_\Gamma + \left\{ \sum_{i \in I^+ \cup I^0} \lambda_i \Psi_i{}^\circ \mid \lambda_i \in \mathbb{R}\; \forall i,\; \lambda_j \langle Dg_j(x^0), u\rangle \le 0 \;\; \forall j \in I^0 \right\}.$$

Remark 7.9 (nontrivial solution of the T-stability system). The T-stability system has a nontrivial solution (u, α, β)

(i) with $u = 0 \;\Leftrightarrow\;$ some pair $(\alpha, w) \neq 0$ satisfies $\sum_{i \in I^+ \cup I^0} \alpha_i \Psi_i{}^\circ \;+\; \sum_{k=1}^\kappa w_k \Gamma_k{}^\circ \;=\; 0$;

(ii) with $u \neq 0 \;\Leftrightarrow\;$ $Q_T(u) \cap K^T(u) \neq \emptyset$ and $u \in U^T$.

◇

Indeed, using $\beta = Dg(x^0)u$, the assertion of the foregoing remark follows from the condition $(\alpha, \beta) \in \mathcal{J}_T(y^0)$ in (7.45).

If $\Gamma^0 = Dh(x^0)$ and $\Psi^0 = Dg(x^0)$ hold in Remark 7.9(i), then (i) means precisely that LICQ is violated. Therefore, in the general case, we say that the *generalized LICQ* holds (at x^0) if such a pair $(\alpha, w) \neq 0$ does not exist. Notice, however, that the generalized LICQ *does not coincide* with the linear independence of related vectors Dg_i and Dh_j (at x^0), considered in Lemma 7.1. So one obtains, for $s^0 = (x^0, y^0, z^0)$,

Remark 7.10 (*TF* injective). $TF(s^0)$ is injective if and only if
(i) the generalized LICQ holds (at x^0) and
(ii) $Q_T(u) \cap K^T(u) = \emptyset \;\; \forall u \in U^T \setminus \{0\}$.

◇

For the special case of **C^2-optimization,** (i) and (ii) take the form

$$\text{LICQ and } Hu \notin K^T(u) \; \forall u \in U^T \setminus \{0\}.$$

Similarly, the C-stability system permits a geometrical interpretation of F^{-1} being locally upper Lipschitz at s^0. Let X_Γ and U_h be as above, and put

$$U^C = U_h \cap \{u \in \mathbb{R}^n | \langle Dg_i(x^0), u\rangle = 0 \; \forall i \in I^+, \langle Dg_j(x^0), u\rangle \leq 0 \; \forall j \in I^0\}.$$

Further, introduce the cone

$$K^C(u) = X_\Gamma + \{ \textstyle\sum_{i \in I^+ \cup I^0} \lambda_i \Psi_i{}^\circ \mid \lambda_j \langle Dg_j(x^0), u\rangle = 0 \;\; \forall j \in I^0 \}.$$

Now we obtain

Remark 7.11 (nontrivial solution of the C-stability system). The C-stability system has a nontrivial solution (u, α, β)

(i) with $u = 0 \;\Leftrightarrow\;$ some pair $(\alpha, w) \neq 0$ satisfies $\alpha_j \geq 0 \; \forall j \in I^0$ and $\sum_{i \in I^+ \cup I^0} \alpha_i \Psi_i{}^\circ + \sum_{k=1}^{\kappa} w_k \Gamma_k{}^\circ = 0$;

(ii) with $u \neq 0 \;\Leftrightarrow\;$ $Q_C(u) \cap K^C(u) \neq \emptyset$ for some $u \in U^C \setminus \{0\}$.

◇

To prove this, one has again to put $\beta = Dg(x^0)u$ and to apply that $(\alpha, \beta) \in \mathcal{J}_C(y^0)$.

Having $\Gamma^0 = Dh(x^0)$ and $\Psi^0 = Dg(x^0)$ in (i), this condition means precisely that the strict MFCQ is violated. In the current case, we say that the *generalized strict MFCQ* holds (at x^0) if such a pair $(\alpha, w) \neq 0$ does not exist. So one obtains for $s^0 = (x^0, y^0, z^0)$,

Remark 7.12 (*CF* injective). $CF(s^0)$ is injective if and only if
(i) the generalized strict MFCQ holds (at x^0) and
(ii) $Q_C(u) \cap K^C(u) = \emptyset \;\; \forall u \in U^C \setminus \{0\}$. ◇

For the special case of **C^2-optimization,** (i) and (ii) take the form

$$\text{strict MFCQ and } Hu \notin K^C(u) \; \forall u \in U^C \setminus \{0\}.$$

7.4 Discussion of Particular Cases

In this section, we specialize the results of the previous section to the case of $M(\cdot) \in C^1$ and apply them to nonlinear complementarity problems. In particular, we discuss consequences for Newton–type methods. However, the application to usual Kojima functions related to nonlinear programs will be postponed to Chapter 8.

7.4.1 The Case of Smooth Data

Let $s^0 = (x^0, y^0, z^0)$ be a zero of the Kojima function F. We again use the abbreviations (7.37) ... (7.40). To avoid the transposition symbol, we agree that the vectors

$$u,\ v,\ w,\ \sigma,\ \Phi,\ \Psi_i,\ \Gamma_k \text{ are columns}$$
$$Dg_i(x^0),\ Dh_k(x^0) \text{ are rows}$$

and use the convention (7.7). Suppose that $M(\cdot) \in C^1$. For $F_1(s) = \Phi(x) + \sum_{i=1}^m y_i^+ \Psi_i(x) + \sum_{k=1}^{\kappa} z_k \Gamma_k(x)$, now one has

$$Q_T(u) = Q_C(u) = \{Hu\} \quad \text{with} \quad H := D_x F_1(s^0). \tag{7.48}$$

For example, if F is the Kojima function of the nonlinear program (7.1), H is the (partial) Hessian with respect to x at s^0 of the Lagrangian

$$L(x, y, z) = f(x) + \sum_{i=1}^{m} y_i^+ g_i(x) + \sum_{k=1}^{\kappa} z_k h_k(x).$$

Matrix Representation

From Theorem 7.6 we know that the Thibault derivative of F at s^0 in direction $\sigma = (u, v, w)$ can be written as

$$TF(s^0)(\sigma) = \{P(r)\sigma \,|\, r \in \mathcal{R}_T(y^0)\},$$

where $P(r)$, $r \in \mathcal{R}_T(y^0)$, is the quadratic matrix of order $d = n + m + \kappa$,

$$P(r) = \begin{pmatrix} H & r_1\Psi_1{}^\circ & \cdots & r_m\Psi_m{}^\circ & \Gamma_1{}^\circ & \cdots & \Gamma_\kappa{}^\circ \\ Dg_1(x^0) & -(1-r_1) & & & 0 & & \\ \vdots & & \ddots & & & \ddots & \\ Dg_m(x^0) & & & -(1-r_m) & & & 0 \\ Dh(x^0) & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{7.49}$$

and $\mathcal{R}_T(y^0)$ is given according to (7.28).

Specializations of Regularity Characterizations

We know that F is strongly regular at $(s^0, 0)$ if and only if $0 \in TF(s^0)(\sigma)$ implies $\sigma = 0$, this obviously yields

$$F \text{ is strongly regular at } s^0 \Leftrightarrow \det P(r) \neq 0 \ \forall r \in \mathcal{R}_T(y^0)$$

in our special case. Now let Ξ be the set of all matrices $P(r)$ under consideration. Since Ξ is (arc–wise) connected in the space of (d, d)–**matrices,** one obtains

$$\begin{aligned} & F \text{ is strongly regular at } s^0 \\ \Leftrightarrow \ & \det P(r) \text{ has the same } \mathbf{sign} \neq 0 \ \forall r \in \mathcal{R}_T(y^0). \end{aligned}$$

Because each r_i appears in exactly one column and is affine–linear, the function $r \mapsto P(r)$ is affine–linear in each r_i, too. So (by induction arguments) it suffices to check only all determinants $\det P(r)$ for $r \in \mathcal{R}_T(y^0)$ with $r_i \in \{0, 1\}$.

Moreover, since only $N(\cdot)$ includes nondifferentiable terms and

$$TN(y^0, z^0)(v, w) = \partial N(y^0, z^0) \begin{pmatrix} v \\ w \end{pmatrix},$$

one easily sees that the set of matrices Ξ is just Clarke's [Cla83] generalized Jacobian $\partial F(s^0)$. Clarke's sufficient condition for local Lipschitz invertibility (i.e., strong regularity) of F says that all matrices in $\partial F(s^0)$ have to be non–singular. *Hence, in our particular case, this sufficient condition is also a necessary condition for strong regularity of F.*

Finally, along with F, let us also study the linearized Kojima function $LF = (LF_1, LF_2, LF_3)$

$$\begin{aligned} LF_1(s) &= \Phi(x^0) + H(x - x^0) + \textstyle\sum_{i=1}^m y_i^+ \Psi_i{}^\circ + \sum_{k=1}^\kappa z_k \Gamma_k{}^\circ, \\ LF_2(s) &= \ g(x^0) + Dg(x^0)(x - x^0) - y^-, \\ LF_3(s) &= \ h(x^0) + Dh(x^0)(x - x^0). \end{aligned} \tag{7.50}$$

Clearly, it holds $LF(s^0) = F(s^0)$, and the related derivatives of F and LF evidently coincide at the point s^0.

Now we summarize in the smooth case several characterizations of strong regularity and local upper Lipschitz continuity of F.

Corollary 7.13 ($M \in C^1$; strong regularity and local u.L. behavior). *Let $F = MN$ be a generalized Kojima–function according to (7.8), and let $F(s^0) = 0$. Suppose that $M \in C^1$.*

Then the following properties are mutually equivalent:

(i) $TF(s^0)$ *is injective (i.e., F is strongly regular at s^0).*

(ii) *LF is strongly regular at s^0.*

(iii) $\det P(r) \neq 0$ *for all* $r \in R$.

(iv) $\det P(r)$ *has the same* $\mathbf{sign} \neq 0$ *for all* $r \in R$ *with* $r_i \in \{0, 1\}$ $(\forall i)$.

(v) $J \in \partial F(s^0)$ *are non-singular matrices.*

Moreover, the following properties are mutually equivalent:

(a) $CF(s^0)$ *is injective (i.e.,* F^{-1} *is locally u.L. at* $(0, s^0)$*).*
(b) $C(LF)\ (s^0)$ *is injective.*
(c) $(LF)^{-1}$ *is locally u.L. at* $(0, s^0)$. ◇

Proof. Both sets of equivalences are consequences of Theorem 7.8 according to the above discussion. □

Note. With respect to C^2–optimization problems, Corollary 7.13 ensures, that the injectivity condition for *TF* does not change if, instead of the original problem, we study its quadratic approximation at s^0, namely,

$$\text{(PQ)} \qquad \begin{array}{ll} \min & Df(x^0)(x-x^0) + \frac{1}{2}\langle x-x^0, H(x-x^0)\rangle \\ \text{s.t.} & g(x^0) + Dg(x^0)(x-x^0) \le 0,\ h(x^0) + Dh(x^0)(x-x^0) = 0. \end{array}$$

Hence, strong stability of the original problem and its quadratic approximation (PQ) at s^0 coincide. This reduction to quadratic problems was a key result of Robinson's paper [Rob80]. By Corollary 7.13, the same may be said about the locally upper Lipschitz property of KKT–points at $(0, s^0)$.

Concerning necessity of LICQ under strong regularity, let us add the following argumentation, based on Corollary 7.13 and proposed in [Kum91b]: Strong regularity implies that $0 \notin TF(s^0)(\sigma)$ for all $\sigma = (0, v, w) \neq 0$, and this again implies LICQ via $r_i = 1$ if $y_i^0 \geq 0$.

Some Historical Notes on Strong Regularity

The history of the strong regularity conditions presented in Corollary 7.13 (in the context of C^2–optimization problems) is quite long. The conditions are basically known from Kojima's and Robinson's work in 1980, see [Koj80, Rob80]. Robinson wrote the condition in a different algebraic way by means of Schur complements. Kojima proved the characterization (i) ⇔ (iv), however, LICQ was still an additional assumption. Again LICQ extra assuming, Jongen et al. [JMRT87] proved in 1987 that Robinson's and Kojima's matrix conditions are equivalent.

A little gap was remaining after the mentioned papers: the proof that LICQ is a (simple !) consequence of strong regularity at critical points. In our knowledge this gap was first time closed in 1990-91 in the papers [KT90, Thm. 2.3] and [Kum91b, Thm. 5.1]. Further, in [JKT90], the equivalence between Clarke's regularity condition for $\partial F(s^0)$ and strong regularity has been shown, so it became again evident, now by means of nonsmooth analysis, that the earlier conditions imposed by Robinson and Kojima are both sufficient and necessary for strong regularity.

The strong regularity conditions of Corollary 7.13 can be also derived from the stability results for PC^1 equations, presented in [RS97] in terms of the so-called *coherent orientation property,* note that *F* is PC^1 in this special case. For recent self-contained presentations of characterizing strong regularity in C^2 optimization, we refer, e.g., to [Don98, KK99b, BS00].

Additionally, by using the injectivity condition of coderivatives (see our Theorem 3.7), characterizations in terms of intersections of polar cones have been also derived via generalized equations, see [DR96]. There, as a new final result, it has been shown that (in our terminology) strong and pseudo-regularity of generalized Kojima functions coincide provided that M is a C^1 function. An alternative proof and an example for the fact that this statement does not hold for $C^{1,1}$ optimization without constraints and piecewise linear Df, has been given in [Kum98], we will present this result in §7.5 and Example BE.4.

The present approach to strong regularity of $C^{1,1}$-optimization problems (via injectivity and computing *TF*) has been developed in [Kum91b, Kum91a], extensions to generalized Kojima systems and conditions for upper regularity have been presented in [KK99a, KK99b]. The reader will easily find various other approaches and remarkable contributions devoted to stability of critical and stationary points for optimization problems and related variational problems, we refer to [KR92, Mor94, PR96]. However, by our opinion, the applied techniques in these papers or in the book [LPR96] are essentially restricted to $M \in C^1$ and $f, g, h \in C^2$, respectively.

In this historical series, we have to include A.V. Fiacco's pioneering work concerning sensitivity in parametric optimization, see [FM68, Fia83], though he studied even differentiability properties of solutions. But, at the crucial points, there strict complementarity has been supposed which allows to apply the usual C^1-implicit function theorem to the KKT-system for deriving a local stability theory. Nevertheless, our desire of extending his clear analytical approach to stability in optimization, was just a key idea for writing this book at all.

Difference between TF and CF

Let us return to the derivatives TF and CF and regard its difference. Actually, only the replacement of $\mathcal{R}_T(y^0)$ and $\mathcal{R}_C(y^0, v)$ is of importance for comparing them.

Corollary 7.14 (difference between *TF* and *CF*). *Let* $s^0 = (x^0, y^0, z^0)$, $\sigma = (u, v, w)$ *, and let* $M(\cdot) \in C^1$*. Then*

$$TF(s^0)(\sigma) = CF(s^0)(\sigma) + P,$$

where, with $d = n + m + \kappa$ *and* $e_i :=$ *i-th unit vector of* $\mathbb{R}^m$*,*

$$P = P(x^0, y^0, v) := \left\{ -\sum_{i=1}^{m} \lambda_i |v_i| (\Psi_i{}^\circ, e_i, 0) \in \mathbb{R}^d \;\middle|\; \begin{matrix} \lambda_i \in [0,1] \text{ if } y_i^0 = 0, \\ \lambda_i = 0 \quad \text{else} \end{matrix} \right\}.$$

◇

Note. The statement says that in the smooth case, $TF(s^0)(\sigma)$ is a translation of the polyhedron P along the vector $CF(s^0)(\sigma)$.

Proof of Corollary 7.14. We apply the explicit formulas. Since $F_1 \in C^1$, we have $Q_T(u) = Q_C(u) = \{Hu\}$, where

$$H := D_x F_1(x^0, y^0, z^0).$$

The only difference concerns r_i when $y_i^0 = 0$. TF permits full variation $r_i \in [0,1]$, CF restricts r_i to 0 or 1, depending on the selected v_i . This leads to different elements ξ and η_i of the derivatives. First, to see what happens, let us suppose that $I^0 := \{i | y_i^0 = 0\}$ has only one element. Having $(\xi, \eta, \mu) \in CF(s^0)(\sigma)$ and $(\xi', \eta', \mu) \in TF(s^0)(\sigma)$, we observe that if $v_i^0 \geq 0$ then

$$\begin{aligned} \xi' - \xi &= r_i v_i \Psi_i{}^\circ - v_i \Psi_i{}^\circ &&= (r_i - 1) v_i \Psi_i{}^\circ, \\ \eta_i' - \eta_i &= -(1 - r_i) v_i + 0 v_i &&= (r_i - 1) v_i; \end{aligned}$$

if $v_i < 0$ then

$$\begin{aligned} \xi' - \xi &= r_i v_i \Psi_i{}^\circ - 0 &&= r_i v_i \Psi_i{}^\circ, \\ \eta_i' - \eta_i &= -(1 - r_i) v_i + (1 - 0) v_i &&= r_i v_i, \end{aligned}$$

where $r_i \in [0,1]$ in both cases. Unifying both cases, we have

$$(\xi' - \xi, \eta_i' - \eta_i) = -\lambda_i |v_i| (\Psi_i{}^\circ, 1),$$

and each such difference with $\lambda_i \in [0,1]$ may occur. Considering now the case $\operatorname{card} I^0 > 1$, the difference $(\xi' - \xi, \eta' - \eta)$ can be written as a sum:

$$(\xi' - \xi, \eta' - \eta) = -\sum_{I^0} \lambda_i |v_i| (\Psi_i{}^\circ, e_i), \ \lambda_i \in [0,1].$$

Again, all right-hand sides may appear. Finally, to send the interesting sets into $\mathbb{R}^d$, we defined just the polyhedron P. □

So, even for C^2**–optimization** problems, TF and CF are different if $I^0 \neq \emptyset$. On the other hand (see Section 7.2), by Robinson's approach via generalized equations (cf. §7.1), there is only one crucial generalized equation, based on the linearization of H for such problems. In this framework, one may understand our derivatives TN and CN as different approximations of the normal map.

The injectivity check for TF as well as computing all elements in the set $TF^{-1}(s^0)(\xi, \eta, \zeta)$ requires to solve a finite number of linear equations assigned to the matrices $P(r)$ for $r \in \mathcal{R}_T(y^0)$, $r_i \in \{0,1\}$. In particular, for CF, the nontrivial solutions of the C–stability system are of interest. This leads us, by definition of $\mathcal{J}_C(y^0)$, to a (generally non-monotone) linear complementarity system.

Consequences for Newton Methods

For computing solutions of the generalized Kojima systems by a *Newton method* (based on linear auxiliary problems), one may fix some matrix $P(r, s)$, assigned to $s = (x, y, z)$ and $r \in \mathcal{R}_T(y)$, in order

$$\begin{array}{l} \text{to find } \sigma \text{ with } F(s) + P(r,s)\sigma = 0 \\ \text{and to put } s_{\text{new}} = s + \sigma, \end{array} \tag{7.51}$$

where $P(r,s)$ denotes the matrix (7.49) with s in place of s^0. Concrete methods then differ each to each other by the selection of $r \in \mathcal{R}_T(y)$ at a current iteration point s.

Lemma 7.15 (Newton's method under strong regularity). *Under strong regularity at the solution s^0 and for $M(\cdot) \in C^1$, these methods converge (locally) superlinearly to s^0* for all *selections of $r \in \mathcal{R}_T(y)$.* ◇

Note. Whether strong regularity is really needed for convergence, depends on the choice of r.

Proof of Lemma 7.15. Put $P = P(r,s)$. We have

$$\begin{aligned} s_{\text{new}} - s^0 &= s - s^0 - P^{-1}(F(s) - F(s^0)) \\ &= -P^{-1}(F(s) - F(s^0) - P(s - s^0)). \end{aligned}$$

By strong regularity, all matrices $P \in \partial F(s)$ have uniformly bounded inverses for s near s^0. Since $M(\cdot) \in C^1$, it holds that F is a PC^1 function. So F is semismooth, cf. §6.4.2, and the matrices $P(r,s) \in \partial F(s)$ form a Newton function of F at s^0. But this yields $s_{\text{new}} - s^0 \in o(s - s^0)B$. □

In Chapter 10 we will see that this statement remains even true if F is less smooth, and we will interpret the auxiliary Newton-systems.

7.4.2 Strong Regularity of Complementarity Problems

Given $u, v \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, the nonlinear complementarity problem (NCP) claims to find some x such that

$$u(x) \geq 0, \; v(x) \geq 0 \quad \text{and} \quad \langle u(x), v(x) \rangle = 0. \tag{7.52}$$

To this problem, we assign the generalized Kojima function

$$F_1 := u(x) - y^+, \quad F_2 := -v(x) - y^-, \tag{7.53}$$

F_3 does not appear. It holds $y^* = u(x^*) - v(x^*)$ at any solution x^* of the NCP. *Strong regularity of the NCP* (i.e., by definition, strong regularity of F) at (x^*, y^*) means regularity of all matrices $P(r)$ in (7.49). In the present context, the rows of $P(r)$, $r \in \mathcal{R}_T(y^*)$, attain the form

$$\begin{array}{llcll} Du_i(x^*) & 0 & \ldots \quad -r_i \quad \ldots & 0 & (\text{row } i, \qquad -r_i \text{ at column } n+i), \\ -Dv_i(x^*) & 0 & \ldots -(1-r_i)\ldots & 0 & (\text{row } n+i, \; -(1-r_i) \text{ at column } n+i). \end{array} \tag{7.54}$$

In complementarity theory, one often uses matrices combined by Du and Dv, so let us transform the system again.

Lemma 7.16 (strong regularity of an NCP). *Let u, $v \in C^1$, and let x^* be a solution of the NCP (7.52). Denote by $C(r)$ the matrix formed by the rows $C_i(r) = (1-r_i)Du_i(x^*) + r_iDv_i(x^*)$. Then, for every fixed $r \in \mathbb{R}^n$, the matrix $P(r)$ in (7.54) is singular if and only if so is the matrix $C(r)$. Moreover, the NCP is strongly regular at x^* if and only if the matrices $C(r)$ are non-singular for all $r \in \mathcal{R}_T(u(x^*) - v(x^*))$.* ◇

Proof. We verify the first statement, the second one is a consequence of Corollary 7.13.

($\Rightarrow$) Let $P(r)\binom{\xi}{\eta} = 0$, $\binom{\xi}{\eta} \neq 0$. Then $\xi = 0$ yields $\eta = 0$, hence $\xi \neq 0$. The two equations

$$Du_i(x^*)\xi - r_i\eta_i = 0, \qquad Dv_i(x^*)\xi + (1 - r_i)\eta_i = 0, \tag{7.55}$$

assigned to lines i and $n + i$, now yield $C_i(r)\xi = 0$ due to $(1 - r_i)Du_i(x^*)\xi = (1 - r_i)r_i\eta_i = -r_iDv_i(x^*)\xi$.

($\Leftarrow$) Let $C(r)\xi = 0$, $\xi \neq 0$. Then one finds η such that $P(r)\binom{\xi}{\eta} = 0$ by setting

$$\begin{array}{ll} \eta_i = Du_i(x^*)\xi & \text{if } r_i = 1, \\ \eta_i = -Dv_i(x^*)\xi & \text{if } r_i = 0, \\ \eta_i = Du_i(x^*)\xi / r_i & \text{otherwise} \end{array}$$

If $r_i \in \{0, 1\}$, (7.55) follows elementary. In the last case, it holds $Du_i(x^*)\xi = r_i\eta_i$ by definition and $Dv_i(x^*)\xi = -(1 - r_i)r_i^{-1}Du_i(x^*)\xi = -(1 - r_i)\eta_i$, again (7.55) is shown. □

To characterize the Newton equations (7.51) for the actual case, let

$$[Lu]_i(x, \xi) := u_i(x) + Du_i(x)\xi$$

be the linearization of u_i at x.

Lemma 7.17 (transformed Newton solutions). *Let $r \in \mathcal{R}_T(y)$, and let $\sigma := (\xi, \eta)$ solve (7.51) for $s = (x,y) = (x,u(x) - v(x))$. Then,*

$$0 = (1 - r_i)[Lu]_i(x, \xi) + r_i[Lv]_i(x, \xi). \tag{7.56}$$

Conversely, if ξ satisfies (7.56) for $r \in \mathcal{R}_T(y)$ and $y = u(x) = v(x)$, then (7.51) is satisfied with η as

$$\begin{array}{ll} \eta_i = F_i(x, y) \quad + Du_i(x)\xi & \text{if } r_i = 1, \\ \eta_i = F_{n+i}(x, y) - Dv_i(x)\xi & \text{if } r_i = 0, \\ \eta_i = (F_i(x, y) \quad + Du_i(x)\xi)/r_i & \text{otherwise} \end{array}$$

◇

Proof. To abbreviate we omit the arguments x,y of F_i and F_{n+i}. By (7.51) it holds

$$F_i + Du_i(x)\xi = r_i\eta_i \quad \text{and} \quad F_{n+i} - Dv_i(x)\xi = (1 - r_i)\eta_i.$$

Then we have $r_iy_i^- = (1 - r_i)y_i^+$ due to $r \in \mathcal{R}_T(y)$ and

$$(1 - r_i)F_i + (1 - r_i)Du_i(x)\ \xi = (1 - r_i)r_i\eta_i = r_iF_{n+i} - r_iDv_i(x)\xi.$$

Thus,

$$\begin{aligned} 0 &= (1-r_i)F_i - r_iF_{n+i} + [(1-r_i)Du_i(x) + r_iDv_i(x)]\xi \\ &= (1-r_i)(u_i(x) - y_i^+) + r_i(v_i(x) + y_i^-) + C_i(r_i)\xi \\ &= (1-r_i)u_i(x) + r_iv_i(x) + [(1-r_i)Du_i(x) + r_iDv_i(x)]\xi. \end{aligned}$$

So (7.56) is valid. Conversely, we have to discuss the three possible cases, namely,

$r_i = 1$:

Then $y_i^- = 0$, $F_i + Du_i(x)\xi = \eta_i = r_i\eta_i$ and $0 = v_i(x) + Dv_i(x)\xi$. Thus, $F_{n+i} - Dv_i(x)\xi = F_{n+i} + v_i(x) = -v_i(x) - y_i^- + v_i(x) = 0 = (1-r_i)\eta_i$.

$r_i = 0$:

Then $y_i^+ = 0$, $F_{n+i} - Dv_i(x)\xi = \eta_i = (1-r_i)\eta_i$ and $0 = u_i(x) + Du_i(x)\xi$. Thus $F_i + Du_i(x)\xi = F_i - u_i(x) = u_i(x) - y_i^+ - u_i(x) = 0 = r_i\eta_i$.

$0 < r_i < 1$:

Now we have $y_i = 0$ and $F_i + Du_i(x)\xi = r_i\eta_i$ as well as $F_{n+i} - Dv_i(x)\xi = -v_i(x) - Dv_i(x)\xi = [Lv]_i(x,\xi) = r_i^{-1}(1-r_i)[Lu]_i(x,\xi)$ $= r_i^{-1}(1-r_i)(F_i + Du_i(x)\xi) = (1-r_i)\eta_i$.

This completes the proof. □

Having $y_i = 0$, all choices of $r_i \in [0,1]$ are possible in (7.56), and $F_i = u_i$, $F_{n+i} = -v_i$. The equations (7.56) are crucial if one solves the complementarity problem by means of certain positively homogeneous NCP-functions g. The coefficients r_i in (7.56) are then defined as normalized partial derivatives of g at $(u_i(x), v_i(x))$. For details we refer to Section 9.1.

7.4.3 Reversed Inequalities

If M is only a $C^{0,1}$ mapping, the equation $Q_T(u) = Q_C(u) = \{Hu\}$ (see (7.48)) is no longer true. The linear operator $P(r)$ of (7.49) now includes the multifunction $u \mapsto Q_T(u)$ at the place of H and becomes a nonlinear, set–valued operator which is defined in accordance with the explicit formula of Theorem 7.6 for each $r \in \mathcal{R}_T(y^0)$. This operator $P(r)$ is still linear in the dual directions v and w.

Let us next again assume that

$$\Psi(x) = Dg(x) \quad \text{and} \quad \Gamma(x) = Dh(x),$$

i.e., in contrast to Kojima systems for standard nonlinear programs, only Df is replaced by Φ. Suppose that s^0 is a zero of F such that strong regularity at s^0 is *violated*, i.e.,

$$0 \in TF(s^0)(u,v,w) \text{ for some } (u,v,w) \neq 0.$$

Further, let i be some index such that $y_i^0 = 0$.

Since $g_i(x^0) = 0$, the point s^0 is also a zero of the function F^i which differs from F only by changing the sign of both g_i and $\Psi_i = Dg_i(x)$. This procedure simply means that the original inequality $g_i(x) \le 0$ has been reversed. Changing the sign of v_i in the vector (u, v, w), we see that F^i is not strongly regular at s^0 , too. Therefore, we may state

Lemma 7.18 (invariance when reversing constraints). *Suppose* $\Phi \in C^{0,1}$, $g, h \in C^1$, $\Psi(x) = Dg(x)$ *and* $\Gamma(x) = Dh(x)$, *and let* s^0 *be a zero of* F. *Then, strong regularity of* F *at* s^0 *is invariant with respect to multiplication of any* g_i *with* $\mu_i \neq 0$ *provided that* $y_i^0 = 0$. ◇

The previous lemma explains why characterizations of strongly regular critical points may differ by sign, compare, e.g., differences in such conditions given in [Kum91b] and [DR96], It is worth to mention that the lemma fails to hold for the related injectivity of CF.

7.5 Pseudo–Regularity versus Strong Regularity

In Chapter 5 we gave a characterization of pseudo–regularity for continuous functions. Instead of specializing these results to generalized Kojima functions (which would not give much new insight), we shall discuss the close connections to strong regularity in various situations.

Throughout this section, let F be a generalized Kojima function, and suppose that

$$\Phi,\ \Psi,\ \Gamma,\ g \text{ and } h \text{ are } C^{0,1} \text{ functions.}$$

In the first two lemmas we shall show that *zero–Lagrange–multipliers (zero LM)* do not play an essential role for the relation between pseudo– and strong regularity. For the given F, define $F^{(m)}$ by removing both, the m–th component of F_2 and the product $y_m^+ \Psi_m$ in F_1. Obviously, $F^{(m)}$ is again a generalized Kojima–function. Let as above $d = n + m + \kappa$.

Lemma 7.19 (deleting constraints with zero LM, pseudo-regular). *If* F *is pseudo–regular at some zero* (x^0, y^0, z^0) *of* F *with* $y_m^0 = g_m(x^0) = 0$, *then* $F^{(m)}$ *is again pseudo–regular at* $(x^0, y_1^0, \ldots, y_{m-1}^0, z^0)$. ◇

Proof. Specify the norms used in $\mathbb{R}^d$ and $\mathbb{R}^{d-1}$ to be the maximum–norm, and put $S := (F^{(m)})^{-1}$. We use the letters a, x for elements of $\mathbb{R}^n$, β, η for elements of $\mathbb{R}^m$, b, y for elements of $\mathbb{R}^{m-1}$, and c, z for elements of $\mathbb{R}^\kappa$. In the definition of pseudo-regularity of F, let L be the Lipschitz constant, let V be the open ball $\delta B_d^0 \subset \mathbb{R}^d$ and let $U := s^0 + \varepsilon B_d^0 \subset \mathbb{R}^d$, where $\varepsilon > 0$ is already small enough such that

$$\|x' - x^0\| \le \varepsilon \quad \Longrightarrow \quad g_m(x') < \delta/3.$$

Again by continuity, there is some $\varrho \in (0, \delta)$ such that

$$\|x'' - x^0\| \le \varepsilon + L\varrho \quad \Longrightarrow \quad g_m(x'') < \delta/2.$$

Now let

$$(a', b', c'), (a'', b'', c'') \in \tfrac{1}{2}\varrho B^0_{d-1} \text{ and } (x', y', z') \in (s^0 + \varepsilon B^0_{d-1}) \cap S(a', b', c')$$

be arbitrarily fixed. We have to show that there is some point $(x'', y'', z'') \in S(a'', b'', c'')$ such that

$$\|(x', y', z') - (x'', y'', z'')\| \le L\|(a', b', c') - (a'', b'', c'')\|. \tag{7.57}$$

For this reason, we define the vector β' with $(a', \beta', c') \in \delta B^0_d$, which differs from b' by the additional component $\beta'_m := \frac{1}{2}\delta$ only. Similarly put $\beta''_m := \frac{1}{2}\delta$ for defining (a'', β'', c'') by using (a', b', c').

Because of $g_m(x') < \delta/3 < \beta'_m$, the m–th inequality is not active at x'. Hence the point $(x', \eta', z') := (x', y', g_m(x') - \beta'_m, z')$ belongs to $F^{-1}(a', \beta', c')$. The pseudo–regularity of F provides us with some $(x'', \eta'', z'') \in F^{-1}(a'', \beta'', c'')$ satisfying

$$\|(x', \eta', z') - (x'', \eta'', z'')\| \le L\|(a', \beta', c') - (a'', \beta'', c'')\| \le L\varrho.$$

Hence, $\|x'' - x^0\| \le \varepsilon + L\varrho$, and so, the choice of r yields $g_m(x'') < \frac{1}{2}\delta = \beta''_m$. This gives $\eta''_m = g_m(x'') - \beta''_m < 0$. Therefore, the point (x'', y'', z'') defined by deleting η''_m in (x'', η'', z'') belongs to $S(a'', b'', c'')$ and satisfies (7.57). □

Recall that a (multi-valued) selection $\Delta(s) \subset \Omega(s)$ $(\forall s)$ of a given multifunction $\Omega : \mathbb{R}^d \to \mathbb{R}^d$ is said to be *continuous* if it is upper and lower semicontinuous.

Lemma 7.20 (deleting constraints with zero LM, not strongly regular). *Let F be pseudo–regular at some zero $s^0 = (x^0, y^0, z^0)$ of F with $y^0_m = g_m(x^0) = 0$. Further, suppose that F^{-1} has a closed–valued and continuous selection Δ such that $\Delta(0) = \{s^0\}$. Then, if F is not strongly regular at s^0, so $F^{(m)}$ is also not strongly regular at $(x^0, y^0_1, \dots, y^0_{m-1}, z^0)$.* ◇

Proof. Consider the mapping $p = (a, b, c) \mapsto \Sigma(p) := \text{argmin}\{y_m | (x, y, z) \in \Delta(p)\}$ and suppose first that Σ is single–valued on some ball εB. Then, by the properties of Δ, Σ is a continuous selection of F^{-1}, $\Sigma(0) = \{s^0\}$, and, in accordance with Theorem 5.10, F is strongly regular at s^0. Hence, if F is not strongly regular, then for some sequence $p \to 0$, there exist certain elements $s' = (x', y', z')$, $s'' = (x'', y'', z'') \in \Sigma(p)$ satisfying

$$\sigma = (u, v, w) := (x'' - x', y'' - y', z'' - z') \ne 0 \quad \text{and} \quad y''_m = y'_m. \tag{7.58}$$

Our assumptions concerning Δ ensure that $\sigma \to 0$ and $y_m \to y^0_m = 0$. Next, consider the inverse map S of $F^{(m)}$ at the parameter points

$$p' = (a - (y'_m)^+\Psi_m(x'), \pi(b), c) \quad \text{and} \quad p'' = (a - (y''_m)^+\Psi_m(x''), \pi(b), c),$$

where $\pi(b)$ is the projection of b onto $\mathbb{R}^{m-1}$. Deleting the y_m-coordinates of the points in (7.58), we define points (x', η', z') and (x'', η'', z'') in $S(p')$ and $S(p'')$, respectively. Due to $v_m = 0$, the difference of the parameters becomes

$$\|p' - p''\| = \|(y''_m)^+ \Psi_m(x'') - (y'_m)^+ \Psi_m(x')\| = (y'_m)^+ \|\Psi_m(x'') - \Psi_m(x')\|.$$

To show that $F^{(m)}$ is not strongly regular at s^0, we assume the contrary is true (with rank L). With some Lipschitz constant K for Ψ_m near x^0, and $\sigma_m = (u, v_1, \ldots, v_{m-1}, w)$, one then obtains

$$\begin{aligned} \|\sigma_m\| &\leq L\|p' - p''\| \\ &= L(y'_m)^+ \|\Psi_m(x'') - \Psi_m(x')\| \\ &\leq L(y'_m)^+ K\|u\| \leq L(y'_m)^+ K\|\sigma_m\|. \end{aligned}$$

Since $y'_m \to 0$ and $\sigma_m \to 0$, the estimate implies $\sigma_m = 0$ and, due to $v_m = 0$, even $\sigma = 0$. This contradicts (7.58) and so completes the proof. □

Recall that, by Lemma 5.11, s^0 is an isolated zero of a PC^1–function F from $\mathbb{R}^d$ in itself, provided that F is pseudo–regular at s^0. In this case, near the origin, F^{-1} has a continuous multivalued selection Δ with compact images and $\Delta(0) = \{s^0\}$. This result, together, with the foregoing lemmas allow for special cases a simple reduction procedure, which has an interesting application: for the usual Kojima–function of a C^2–program, regularity and pseudo–regularity coincide.

Theorem 7.21 (reduction for PC^1 data). *Let $F = NM$ be a generalized Kojima–function, and suppose that $g_i, h_k, \Phi, \Psi_i, \Gamma_k$ are PC^1–functions. Let $s^0 = (x^0, y^0, z^0)$ be a zero of F and $I^+ = \{i | y_i^0 > 0\}$. Define the reduced generalized Kojima–function F^r by deleting from F all components F_{2i} and all products $y_i^+ \Psi_i$ in F_1 with $i \notin I^+$. If F is pseudo-regular but not strongly regular at s^0, then the same is true for F^r at the reduced part s^r of s^0.* ◇

Proof. The function F is PC^1, hence F^{-1} has a compact–valued and continuous selection Δ with $\Delta(0) = \{s^0\}$ as long as F is pseudo–regular at s^0. Due to the Lemmata 7.19 and 7.20, one may successively remove all constraints with $y_i^0 = 0$ (which automatically includes that $g_i(x^0) = 0$). After deleting the related components, the corresponding reduced part s^r of s^0 is still a zero of F^r, which remains pseudo–regular but not strongly regular. By standard continuity arguments, this reduction may be continued for the components with $y_i^0 < 0$ without affecting the desired properties. □

With respect to F^r the reduced zero s^r now fulfills the strict complementarity condition. So, if all data in M are even C^1–functions (which means for the usual Kojima-function that $f, g, h \in C^2$), then the system $F^r = 0$ is locally a C^1–equation and $TF^r(s^r)(\sigma) = CF^r(s^r)(\sigma) = \{DF^r(s^r)\sigma\}$ holds. Therefore, Theorem 5.1 yields that pseudo–regularity of F^r at s^r implies non–singularity of the Jacobian $DF^r(s^r)$ and hence strong regularity of F^r at s^r. So we have proved

Corollary 7.22 ($M \in C^1$; pseudo-regular = strongly regular). *If $F = NM$ is a generalized Kojima–function with g, h, Φ, Ψ, Γ being C^1–functions, then, at any zero s^0, F is strongly regular if and only if F is pseudo–regular.* ◇

The results of this subsection on pseudo–regularity were taken from [Kum98]. Note that in the context of C^2 optimization problems and related variational inequalities, Corollary 7.22 also appears in [DR96]. However, even for stationary points of a $C^{1,1}$ function (unconstrained), both regularity concepts do not coincide, see Example BE.4.

Chapter 8

Parametric Optimization Problems

In this chapter, we study the local Lipschitz behavior of critical points and critical values as well as stationary and (local) optimal solutions for parametric nonlinear optimization problems *in finitely many variables*. We do not aim at a comprehensive or even complete presentation of all aspects of sensitivity and local stability analysis in nonlinear optimization. Our purpose is to derive (Lipschitz) stability results for programs involving $C^{1,1}$ data, and for that to apply largely the results of the previous chapter on regularity and Kojima–functions.

It will turn out that our approach also yields several known (or new) basic results for perturbed nonlinear programs with C^2 data. The statements shall concern strong regularity, pseudo regularity and upper Lipschitz stability, second order characterizations, geometrical interpretations, as well as representations of derivatives of solution and marginal value maps.

Note that there is a well-developed perturbation theory for programs with smooth (i.e., usually at least C^2) data, for a book reflecting the state of the art of this theory we refer to Bonnans and Shapiro [BS00]. Basic monographs in the field of parametric nonlinear optimization are, e.g., [**BGK$^+$82**, Fia83, Mal87, DZ93, Lev94], crucial aspects and applications of this field are systematically handled, e.g., in the books [Gol72, BM88, GGJ90, BA93, Gau94, RW98]. Moreover, for programs with C^3-**data**, there exists a powerful and deeply developed singularity theory based on the characterization of generically appearing singular cases of Kojima's system, we mainly refer to the basic work of Jongen, Jonker and Twilt [JJT86] and to [JJT83, JJT88, JJT91].

As a starting point, we introduce the *parametric nonlinear program*

$$\mathrm{P}(t),\ t \in \mathcal{T}: \qquad \min_x f(x,t) \quad \text{s.t.} \quad g(x,t) \le 0,\ h(x,t) = 0, \tag{8.1}$$

where $\mathcal{T}$ is a subset of $\mathbb{R}^r$, and f, g and h map $\mathbb{R}^n \times \mathcal{T}$ to $\mathbb{R}$, $\mathbb{R}^m$ and $\mathbb{R}^\kappa$, respectively. If $f(\cdot,t)$, $g(\cdot,t)$ and $h(\cdot,t)$ belong to the class C^k, then the problem (8.1) is called a *parametric C^k program.* In particular, we are interested in the classes $C^{1,1}$ and C^2. Recall that the related parametric Kojima system defining the critical points of $\mathrm{P}(t)$, $t \in \mathcal{T}$, then becomes $F(s,t) = 0$, where $s = (x,y,z)$ and $F = (F_1, F_2, F_3)$,

$$\begin{aligned} F_1(s,t) &= D_x f(x,t) + D_x g(x,t)^\top y^+ + D_x h(x,t)^\top z \\ F_2(s,t) &= g(x,t) - y^-, \\ F_3(s,t) &= h(x,t). \end{aligned} \tag{8.2}$$

The associated *Lagrangian* $L(\cdot,t)$ of $\mathrm{P}(t)$ is defined by

$$L(x,y,z,t) = f(x,t) + \sum_{i=1}^{m} y_i^+ g_i(x,t) + \sum_{k=1}^{\kappa} z_k h_k(x,t), \tag{8.3}$$

and we have $F_1 = D_x L$ in (8.2).

An Illustrative Example

Even if $\mathcal{T}$ is a subset of $\mathbb{R}$ and if all data functions are arbitrarily smooth, then critical points and critical values considered as functions of t may behave rather badly (in particular, discontinuity may hold).

Example 8.1 (see [BGK$^+$82]). Consider the real convex, quadratic one–parametric problem

$$\mathrm{P}(t): \qquad \min_x \; t^2x^2 - 2t(1-t)x \quad \text{s. t.} \quad -x \le 0.$$

For $0 < t < 1$ the stationary (= optimal) solutions $x(t)$ and the critical (= optimal) values $\varphi(t)$ are unique, namely,

$$x(t) = \frac{1-t}{t}, \quad \varphi(t) = -(1-t)^2.$$

For $t = 0$ all feasible x are critical, and $\varphi(0) = 0$. For $t \ge 1$ and $t < 0$ we have $x(t) = 0$ and $\varphi(t) = 0$.

The special discontinuity of φ at $t = 0$ (note that φ is not lower semicontinuous at 0) has a strange consequence: If P(0) should be solved but (because of a former computational error) really one solves $\mathrm{P}(t)$ with some $t > 0$, then the error $|\varphi(t) - \varphi(0)|$ becomes as larger as better t approximates the true value 0. On the other hand, if the t-error is large enough $(t \ge 1)$, then one gets again the exact critical value. ◇

Thus, even under the practical point of view, the preceding example illustrates the need of some "stable behavior" of problem (8.1) with respect to parameters describing the involved functions.

Results on Parametric Optimization from Previous Chapters

Many results of the first chapters of the present book (in particular, those of Chapter 7) can be considered as contributions to stability and parametric analysis of feasible and stationary point sets to optimization problems. We compile here a list of propositions explicitly devoted to parametric optimization problems: Theorem 1.15 (Berge/Hogan stability), Theorem 1.16 (stability of complete local minimizing sets), Theorem 2.6 (free local minima and upper Lipschitz constraints), Lemma 2.7 (Hoffman's lemma), Lemma 2.8 (Lipschitz u.s.c. linear systems), Theorem 2.10 (selection maps and optimality conditions), Lemma 4.6 (lsc. and isolated optimal solutions), Corollary 4.7 (pseudo-Lipschitz and isolated optimal solutions), Theorem 4.8 (growth and upper regularity of minimizers).

8.1 The Basic Model

Throughout the present chapter, our basic model is

$$\mathrm{P}(t,p): \qquad \min_x \{f(x,t) - a^{\mathsf T}x \mid g(x,t) \le b\}, \tag{8.4}$$

where t varies in $\mathcal{T} \subset \mathbb{R}^r$, $p = (a,b)$ varies in $\mathbb{R}^{n+m}$, and f, g are given as above. This is a parametric program with *additional canonical perturbations* which are particularly needed in showing that certain sufficient conditions for Lipschitz stability (in the one or the other sense) are also necessary ones.

Since we study the *local* stability behavior of solutions, throughout we associate with the *unperturbed problem* a fixed element of t^0, where we shall often identify $f = f(\cdot,t^0)$, $g = g(\cdot,t^0)$, $F = F(\cdot,t^0)$. The related problem

$$(\mathrm{P})(p): \quad \min_x \{f(x) - a^{\mathsf T}x \mid g(x) \le b\},\ p = (a,b) \in \mathbb{R}^{n+m}, \tag{8.5}$$

is called a parametric program *with canonical perturbations,* i.e., the perturbations of the corresponding Kojima system are only in the right-hand side.

To get a compact and brief description of our results, we have omitted the equality constraints $h(x) = 0$ (or $h(x,t) = 0$). As far as we apply results on Kojima functions, it can be easily and directly seen by the assumptions and proofs below that the equalities play the same role as inequalities with positive multiplier components y_i^0 of the critical point (x^0, y^0) under consideration. This becomes also formally clear from the explicit representations of the derivatives *CF* and *TF* in Theorem 7.6, since there $\beta_i = 0$ or $r_i = 1$ if $y_i^0 > 0$.

We are mainly interested in a local stability analysis of $C^{1,1}$ (or C^2) programs around some given stationary solution x^0 of (P)=P$(t^0,0)$ and for small perturbations (t,p). So, the following general assumptions are supposed to hold:

$$\begin{array}{l}\textbf{x^0 is a stationary solution of (P)=P$(t^0,0)$, $t^0 \in \mathcal{T}$,}\\ \textbf{f, $g_i \in C^{1,1}(\Omega,\mathbb{R})$, where Ω is a neighborhood of (x^0,t^0).}\end{array} \tag{8.6}$$

Note that many stability results derived in the present chapter can be extended to more general parameter spaces and less restrictive assumptions on the data functions, in particular, when taking the results of the Chapters 2 and 3 into account.

Recall that the parameterized Kojima function has the product representation $F(x,y,t) = M(x,t)N(y)$ with

$$M(x,t) = \begin{pmatrix} D_x f(x,t) & D_x g(x,t)^{\mathsf{T}} & 0 \\ g(x,t) & 0 & -E \end{pmatrix} \tag{8.7}$$

and $N(y) = (1, y^{+^{\mathsf{T}}}, (y-y^+)^{\mathsf{T}})^{\mathsf{T}}$, where E means the (m,m) identity matrix. Here, the convention (7.7) is used, i.e., $D_x f$, g and $(D_x g_i)^{\mathsf{T}}$ are considered as column vectors. We again put

$$\mathcal{J}_T(y) := \left\{ (\alpha,\beta) \in \mathbb{R}^{2m} \,\middle|\, \begin{array}{l} \alpha_i = 0 \ (i \in I^-(y)), \ \beta_i = 0 \ (i \in I^+(y)) \\ \alpha_i\beta_i \ \geq \ 0, \quad (i \in I^0(y)) \end{array} \right\},$$

see (7.31), and

$$\mathcal{J}_C(y) := \left\{ (\alpha,\beta) \in \mathbb{R}^{2m} \,\middle|\, \begin{array}{l} \alpha_i = 0 \ (i \in I^-(y)), \ \beta_i = 0 \ (i \in I^+(y)) \\ \alpha_i\beta_i \ = \ 0, \quad \alpha_i \geq 0 \geq \beta_i \ (i \in I^0(y)) \end{array} \right\},$$

see (7.32), where

$$I^0(y) := \{\, i \,|\, y_i = 0 \,\}, \ I^+(y) := \{\, i \,|\, y_i > 0 \,\}, \ I^-(y) := \{\, i \,|\, y_i < 0 \,\}.$$

The sets of *critical points, stationary solutions* and *multipliers* related to $\mathrm{P}(t,p)$ and $(\mathrm{P})(p)$, respectively, are denoted by

$$\begin{array}{rcll} \widetilde{S}(t,p) & := & \{(x,y)\,|\,F(x,y,t) = p\}, & S(p) := \widetilde{S}(t^0,p), \\ \widetilde{X}(t,p) & := & \{x\,|\,\exists y : F(x,y,t) = p\}, & X(p) := \widetilde{X}(t^0,p), \\ \widetilde{Y}(x,t,p) & := & \{y\,|\,F(x,y,t) = p\}, & Y(x,p) := \widetilde{Y}(x,t^0,p), \end{array} \tag{8.8}$$

with $p = (a,b)$. If the multiplier y^0 associated with x^0 is fixed, we use the abbreviations

$$I^+ = I^+(y^0),\ I^0 = I^0(y^0),\ I^- = I^-(y^0), \tag{8.9}$$

and

$$A := D_x g(x^0,t^0) \text{ with rows } A_i := D_x g_i(x^0,t^0). \tag{8.10}$$

Further, since $F_1 = D_x L$, the sets $Q_T(u)$ and $Q_C(u)$ defined in (7.40) become "generalized Hessians" of L, namely

$$\begin{array}{l} Q_T(u) = T_x[D_x L](x^0,y^0,t^0)(u), \\ Q_C(u) = C_x[D_x L](x^0,y^0,t^0)(u), \end{array} \tag{8.11}$$

where $T_x[D_x L](x^0,y^0,t^0)(u)$ $(C_x[D_x L](x^0,y^0,t^0)(u))$ is the partial Thibault-derivative (contingent derivative) of $D_x L$ with respect to x at (x^0,y^0,t^0) in direction u.

To have a concise formulation of stability conditions, we will sometimes suppose that $g(x^0, t^0) = 0$ holds at some initial stationary solution x^0 of the program (P)(0)= $P(t^0, 0)$, i.e., $I^-(y) = \emptyset$ for all $y \in \tilde{Y}(x^0, t^0, 0)$. Of course, under the general assumptions (8.6), $\tilde{X}$ is locally upper Lipschitz (or locally single-valued and Lipschitz) at $((t^0, 0), x^0)$ if and only if the stationary solution set map of the parametric program

$$\min_x\{f(x,t) - \langle a, x\rangle | g_i(x,t) \le b_i \ (\forall i : g_i(x^0, t^0) = 0)\}$$

has this property. The same argument applies to the critical point mapping $\tilde{S}$ at some $((t^0, 0), (x^0, y^0)) \in \operatorname{gph}\tilde{S}$.

8.2 Critical Points under Perturbations

In this section we shall discuss local Lipschitz continuity and local upper Lipschitz behavior of critical points to parametric $C^{1,1}$ and C^2 programs. In the case of canonical perturbations, these properties are defined via the Kojima function $F = F(\cdot, t^0)$. In the case of nonlinear perturbations, the implicit function theorems of previous chapters are helpful. Of course, the regularity characterizations given in Chapter 7 apply to the situation of the present section by putting there $\Phi = D_x f$ and $\Psi := D_x g$.

8.2.1 Strong Regularity

We shall say that the optimization problem (P)=$P(t^0, 0)$ given in (8.4) is *strongly (pseudo) regular* at a critical point $s^0 = (x^0, y^0)$ (or, synonymously, s^0 is a *strongly (pseudo) regular* critical point) if the associated Kojima function $F = F(\cdot, t^0)$ has this property.

Theorem 8.2 (strongly regular critical points). *Let (x^0, y^0) be a critical point of the problem (P). For f, $g_i \in C^{1,1}$, with the notation (8.9)–(8.11), the following properties (i) – (iii) are all equivalent to each other, and each of them implies that LICQ holds at x^0:*

(i) The problem (P) is strongly regular at (x^0, y^0).

(ii) For each solution (r, u, v) of the system

$$\begin{array}{lcll} Q_T(u) & + & \sum_{i=1}^m r_i v_i A_i^T & \ni \ 0, \\ A_i u & - & (1 - r_i) v_i & = \ 0, \ \forall i, \\ \multicolumn{4}{l}{r \in [0,1]^m, \ r_i = 1 \, (i \in I^+), \ r_i = 0 \, (i \in I^-),} \end{array} \qquad (8.12)$$

one has $(u, v) = (0, 0)$.

(iii) The T–stability system

$$\begin{array}{llll} Q_T(u) & + & A^{\mathsf T}\alpha & \ni & 0, \\ Au & - & \beta & = & 0, \\ (\alpha,\beta) \in \mathcal{J}_T(y^0), \end{array} \tag{8.13}$$

has only the trivial solution $(u,\alpha,\beta)=0$.

If f, $g_i \in C^2$, *then each of the conditions (i) – (iii) is equivalent to each of the following conditions:*

(iv) The problem (P) is pseudo-regular at (x^0, y^0).

(v) The determinants of all matrices

$$\begin{pmatrix} H & r_1 A_1^{\mathsf T} & \dots & r_m A_m^{\mathsf T} \\ A_1 & -(1-r_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_m & 0 & \dots & -(1-r_m) \end{pmatrix}$$

with $r \in [0,1]^m$, $r_k = 0$ *if* $k \in I^-$ *and* $r_k = 1$ *if* $k \in I^+$ *have the same non–vanishing sign, where* $H = D^2_{xx}L(x^0,y^0)$. ◇

Proof. Apply Lemma 7.1, Theorem 7.8, Corollary 7.22 and Corollary 7.13. □

Remark 8.3 (necessity of **LICQ** variation of a). In the previous theorem, we used that **LICQ** is a consequence of pseudo-regularity for *generalized* Kojima-functions, see Lemma 7.1. For the present situation, one also knows that if the critical point map $a \mapsto S(a) := S(a,0)$ of the particularly perturbed program

$$\mathrm{P}(a,0): \quad \min_x \{f(x) - \langle a,x\rangle \mid g(x) \le 0\}, \qquad f,g \in C^1, \tag{8.14}$$

is locally single-valued near $(0,x^0,y^0)$, then **LICQ** has necessarily to hold, see [KT90]. For completeness, we give the proof: Assume that $S(\cdot)\cap\mathcal{N}$ is single-valued on some open neighborhood $\mathcal{O}$ of 0, where $\mathcal{N}$ is an open neighborhood of (x^0,y^0), but **LICQ** fails at x^0. Then there is some $\mu \neq 0$ such that $\sum_{i\in I}\mu_i A_i = 0$, where $I := I^+ \cup I^0$ and $A_i = Dg_i(x^0)$. Let $a := -\sum_{i\in I} A_i$ and $y_i(\delta) := y_i^0 + \delta$ if $i \in I$ and $= 0$ if $i \notin I$ for given $\delta \in \mathbb{R}$. Hence, for sufficiently small $\varepsilon > 0$, one has $-\varepsilon a \in \mathcal{O}$, and the point $(x^0, y(\varepsilon)) \in \mathcal{N}$ satisfies

$$F_1(x^0,y(\varepsilon)) = -\varepsilon a, \quad F_2(x^0,y(\varepsilon)) = 0.$$

However, for any sufficiently small number $\theta > 0$, one also has $F(x^0, y(\varepsilon+\theta\mu)) = (-\varepsilon a, 0)$, and again $(x^0, y(\varepsilon+\theta\mu)) \neq (x^0, y(\varepsilon))$ belongs to $\mathcal{N}$, which yields a contradiction. ◇

Corollary 8.4 (nonlinear variations, strongly regular). *Consider the parametric program* $P(t,p)$: $\min_x\{f(x,t) - \langle a,x\rangle | g(x,t) \le b\}$, $t \in \mathcal{T}$, $p = (a,b) \in \mathbb{R}^{n+m}$. *Suppose* $\mathcal{T}$ *is an open subset of* $\mathbb{R}^r$, *and* f, g_i *are real-valued* $C^{1,1}$ *functions defined on* $\mathbb{R}^n \times \mathcal{T}$. *Let* $s^0 = (x^0,y^0)$ *be some critical point of* $(P) = P(t^0, 0)$. *Then,*

1. *the critical point mapping* $\tilde{S}$ *is locally single-valued and Lipschitz around* $(t^0, 0, s^0)$ *if and only if* $0 \in TF(s^0, t^0)(u, 0)$ *holds for each* $u \neq 0$. *Moreover,*
2. *if, in addition,* $D_t F(\cdot, \cdot)$ *exists and is Lipschitzian on some neighborhood of* (s^0, t^0), *then* $\tilde{S}$ *is locally single-valued and Lipschitz around* $(t^0, 0, s^0)$ *if and only if (P)=P($t^0, 0$) is strongly regular at* s^0. ◇

Proof. The first assertion immediately follows from Theorem 5.15, and it implies the second assertion by using Theorem 8.2 and the standard partial derivative formula for Thibault derivatives (cf. Corollary 6.9). □

Remark 8.5 (strong stability in Kojima's sense). Adapting Kojima's [Koj80] definition, we say that a critical point $s^0 = (x^0, y^0)$ of $(P)_{f,g} := (P)$ is *strongly stable* with respect to some perturbation class $\Phi \subset C^2(\mathbb{R}^n, \mathbb{R}^{n+m})$ if (i) there is a constant $\varepsilon > 0$ such that the equation $F(s) + \Delta F(s) = 0$, $s \in B(s^0, \varepsilon)$, has a unique solution $s = s(\Delta f, \Delta g)$ (where $F + \Delta F$ is the Kojima function of $(P)_{(f+\Delta f, g+\Delta g)}$), whenever $(\Delta f, \Delta g)$ is small in the C^2 norm on $B(s^0, \varepsilon)$, and (ii) $s(\cdot)$ is continuous at the zero map 0 with $s(0) = s^0$. To get a relation to strong regularity, one has to ensure that the perturbation class is rich enough. If Φ contains all *small* perturbations of the type $[\Delta f](x) = \frac{1}{2}x^{\mathsf{T}}Dx + a^{\mathsf{T}}x$ (D symmetric) and $[\Delta g](x) = b$, then, by Kojima [Koj80], s^0 is strongly stable with respect to Φ if and only if (P) (i.e. F) is strongly regular at s^0. For a discussion of this relationship see also [KT90, KK99b]. ◇

In view of the previous corollary and remark, we now concentrate ourselves to *canonically perturbed programs.* The next discussions specialize several facts known from Section 7.3.

Geometrical Interpretation

Here we adapt the geometrical interpretation of strong regularity given in Remark 7.10. Put there $\Psi = Dg$, and let (x^0, y^0) be a given critical point of (P)=(P)(0). Then we obtain

$$U^T = \{u \in \mathbb{R}^n \mid A_i u = 0,\ i \in I^+\},$$

and

$$K^T(u) = \{\sum_i \lambda_i A_i \mid \lambda \in \mathbb{R}^m,\ \lambda_i A_i u \le 0,\ i \in I^0, \lambda_k = 0,\ k \in I^-\}.$$

Hence, it holds the

Corollary 8.6 (geometrical interpretation, strongly regular).

$s^0 = (x^0, y^0)$ *is a strongly regular critical point of (P)* *if and only if*

$$\begin{array}{lll} \textit{(LICQ)} & \{A_i,\ i \in I^+ \cup I^0\} \textit{ are linearly independent, and} & \\ \textit{(SOC)} & Q_T(u) \cap K^T(u) = \emptyset \ \textit{ for each } u \in U^T,\ u \neq 0 & \end{array} \qquad (8.15)$$

are valid.

Since $\xi \in Q_T(u) \cap K^T(u)$ and $u \in U^T$ imply $\langle \xi, u \rangle = \sum_{i \in I^0} \lambda_i A_i u \leq 0$, the *second–order condition* (SOC) is always true if

$$\text{(SSOC)} \quad \langle \xi, u \rangle > 0 \text{ for all } \xi \in Q_T(u) \text{ and } u \in U^T,\ u \neq 0 \qquad (8.16)$$

is fulfilled. In the C^2 case, it holds $\langle \xi, u \rangle = \langle u, D^2_{xx} L(x^0, y^0) u \rangle$, i.e., (SSOC) is the so–called *strong second-order condition.*

Direct Perturbations for the Quadratic Approximation

For **C^2–problems,** the T-stability system (8.13) indicates *that and where,* near a KKT point (x^0, y^0) of the initial problem and for certain (a, b) with small norm, there are two different KKT points to the quadratic problem, with $H := D^2_{xx} L(x^0, y^0)$,

$$(\mathrm{PQ})(a,b): \quad \begin{array}{l} Df(x^0)(x - x^0) + \frac{1}{2}\langle x - x^0, H(x - x^0)\rangle - \langle a, x \rangle \ \rightarrow \ \min \\ \text{subject to } g(x^0) + A(x - x^0) \leq b \end{array}$$

such that strong regularity fails to hold. This can be formulated precisely as follows, where $g(x^0) = 0$, i.e., $I^- = \emptyset$, is assumed without loss of generality. Note that the rule of the quadratic approximation $\mathrm{PQ}(a, b)$ in the analysis of strong regularity of (P) was pointed out by Robinson in [Rob80]. The following lemma was first given in [Kum98].

Lemma 8.7 (two close critical points, quadratic problems). *Let* (x^0, y^0) *be a KKT point for (PQ)*$(0,0)$ *with* $g(x^0) = 0$, *i.e.,* $I^- = \emptyset$, *and suppose* $f, g \in C^2$.

(i) If (x, y) *and* (x', y') *are KKT points for (PQ)*(a, b), *sufficiently close to* (x^0, y^0), *then* $u = x' - x$, $\alpha = y' - y$ *and* $\beta = Au$ *solve (8.13).*

(ii) If (u, α, β) *solves (8.13) and has small norm, then, setting* $a = -A^{\mathsf{T}}\alpha^-$ *and* $b = \beta^+$, *the points* $(x, y) = (x^0, y^0 - \alpha^-)$ *and* $(x', y') = (x, y) + (u, \alpha) = (x^0 + u, y^0 + \alpha^+)$ *are KKT points for (PQ)*(a, b). ◇

Proof. One has to put the given points in the related equations. The proof is similar to that of Lemma 8.17 below, we omit the details. □

Corollary 8.8 *Let* (x^0, y^0) *be a KKT point for (PQ)*$(0,0)$ *and* $g(x^0) = 0$, *and suppose* $f, g \in C^2$. *Then problem (P) is* not *strongly regular at* (x^0, y^0) *if and only if for some sequences of vanishing vectors* $\gamma = (\alpha, \beta)$ *and corresponding perturbations* $a = -A^{\mathsf{T}}\alpha^-$ *and* $b = \beta^+$, *the quadratic problem (PQ)*(a, b) *has two different KKT points* $(x', y')(\gamma)$ *and* $(x, y)(\gamma)$ *converging to* (x^0, y^0) *as* $\gamma \to 0$. ◇

Proof. Indeed, if such KKT points exist then we obtain a nontrivial solution of (8.13) via Lemma 8.7. Hence, (P) is not strongly regular.

On the other hand, if (P) is not strongly regular at (x^0, y^0), then some non–trivial (u, α, β) solves (8.13). Multiplying this point by some small $\lambda > 0$, then it remains a solution, and we obtain two different KKT points for $PQ(a, b)$ by Lemma 8.7(ii). □

So, in order to characterize strong regularity of critical points, KKT point of problems involving *particular* canonical perturbations $a = A^{\mathsf{T}}\mu$ with $\mu \geq 0$ and $b \geq 0$ must be considered only.

Lemma 8.7 cannot be applied for $C^{1,1}$**–problems** since the quadratic program (PQ) is not defined. Nevertheless, if the Kojima–function is still piecewise linear,problem $PQ(a, b)$ may be perturbed in a similar way, only $x = x^0$ is not longer true in (ii).

Lemma 8.9 (two close critical points, piecewise quadratic problems). *Let the functions involved in (P) of (8.4) have piecewise linear first derivatives, and let* (x^0, y^0) *be a KKT-point for (P) with* $g(x^0) = 0$.

(i) If (x, y) *and* (x', y') *are KKT points for (P)*(a, b) *in* $(x^0, y^0) + \varepsilon B$ *and* $\varepsilon + \|(a, b)\|$ *is small enough, then* $u = x' - x$, $\alpha = y' - y$ *and* $\beta = Au$ *solve the system (8.13).*

(ii) If (u, α, β) *solves (8.13) and has small norm, then given any* δ, *one finds* x *and* $t > 0$ *such that both* $\|x - x^0\| + t < \delta$ *holds and the point* q *defined by* $q := \theta^{-1}(D_xL(x + tu, y^0) - D_xL(x, y^0))$ *satisfies* $q \in Q_T(u)$ *as well as* $q + A^{\mathsf{T}}\alpha = 0$. *Moreover, by setting* $a = -A^{\mathsf{T}}\alpha^-$ *and* $b = \beta^+$, *the points* $(x, y) = (x, y^0 - \alpha^-)$ *and* $(x, y) + (u, \alpha)$ *are KKT points for (P)*(a, b). ◇

Proof. The representation of q as a difference quotient now follows from the piecewise linear structure of the matrix M in the Kojima function $F = MN$. The rest again requires only the direct calculation. □

So the particular perturbations which must be considered have the same form as in Lemma 8.7 above.

Strong Regularity of Local Minimizers under LICQ

We finish this subsection by specializing Theorem 8.2 to the case of critical points (x, y) of C^2 programs such that the x**-parts** are local minimizers. Recall that $U^T := \{u | A_i u = 0, i \in I^+\}$.

Theorem 8.10 (strongly regular local minimizers). *Suppose that* f *and* g *are* C^2 *functions. Let* $s^0 = (x^0, y^0)$ *be a critical point of the program (P), and suppose that* x^0 *is a local minimizer of this program. Then (P) is strongly regular at* $s^0 = (x^0, y^0)$ *if and only if* x^0 *satisfies LICQ and* $H = D^2_{xx}L(x^0, y^0)$ *is positive definite on* U^T *(i.e., (SSOC) holds). In this case, if* $(x(a, b), y(a, b))$ *denotes the critical point of the canonically perturbed program (P)*(a, b), *then* $x(a, b)$ *is a local minimizer of (P)*(a, b) *whenever* (a, b) *is sufficiently small.* ◇

Proof. Before proving the equivalence, we note that the additional proposition on persistence of the local minimizing property is an immediate consequence of Theorem 1.16.

"If "–direction of the equivalence: This immediately follows from the characterization (8.15) and the sufficient condition (8.16) discussed above.

"Only if" –direction of the equivalence: Suppose that F is strongly regular at $s^0 = (x^0, y^0)$. Then, as shown above, x^0 satisfies **LICQ**. Note that, by convention, **LICQ** is also satisfied if $I^+ \cup I^0 = \emptyset$. To prove (SSOC), we consider

$$\mathrm{P}\,(\varepsilon): \qquad \min\{f(x)|g(x) \le b(\varepsilon)\},\ \varepsilon \ge 0, \quad \text{where } b(\varepsilon)_i := \begin{cases} \varepsilon & \text{if } y_i^0 = 0, \\ 0 & \text{if } y_i^0 \ne 0, \end{cases}$$

and first show that

$$\exists \varepsilon > 0: \qquad x^0 \text{ is a local minimizer of } \mathrm{P}\,(\varepsilon). \tag{8.17}$$

Indeed, strong regularity at (x^0, y^0) particularly includes that the local minimizer x^0 of P (0) is a strict one. Hence, Theorem 1.16 applies and so, for some δ, $\varepsilon' > 0$ and all $\varepsilon \in (0, \varepsilon')$, we have

$$\emptyset \ne \Psi_\delta(\varepsilon) := \operatorname{argmin}\{f(x)|g(x) \le b(\varepsilon), \|x - x^0\| \le \delta\} \subset \{x \,|\, \|x - x^0\| < \delta\},$$

and therefore, each element of $\Psi_\delta(\varepsilon)$ is a local minimizer of $\mathrm{P}\,(\varepsilon)$. Assume that δ, ε' were already small enough such that for $\varepsilon \in (0, \varepsilon')$, both LICQ holds on $\Psi_\delta(\varepsilon)$ – thus, each $x \in \Psi_\delta(\varepsilon)$ is a stationary solution of $\mathrm{P}\,(\varepsilon)$ – and $\mathrm{P}\,(\varepsilon)$ has a unique stationary solution in $B(x^0, \delta)$, which is implied by strong regularity. Since, by construction, x^0 is a stationary solution of $\mathrm{P}\,(\varepsilon)$ even for all $\varepsilon > 0$, we then finally conclude that $\Psi_\delta(\varepsilon) = \{x^0\}$, $\varepsilon \in (0, \varepsilon')$. So (8.17) is shown.

Because of (8.17), the relation

$$\langle u, Hu\rangle \ge 0 \qquad \forall u \in U^T, \tag{8.18}$$

holds by a classical necessary optimality condition. Theorem 8.2 implies that, in particular, the matrix of (v) in that theorem with $r_i = 0$ if $y_i^0 \le 0$ and $r_i = 1$ if $y_i^0 > 0$ is nonsingular, and so

$$\begin{pmatrix} H & Dg_{I^+}(x^0)^\mathsf{T} \\ Dg_{I^+}(x^0) & O \end{pmatrix}$$

is also nonsingular, where g_{I^+} is the vector function built by g_i, $i \in I^+$. Hence, by a known fact from linear algebra, we then obtain that the strong inequality in (8.18) is satisfied for $u \in U^T \setminus \{0\}$, i.e., (SSOC) holds, which completes the proof. □

Note. *In the case of* $C^{1,1}$ *programs*, strong regularity of (P) at (x^0, y^0) for a local minimizer x^0 in general *does not imply* the corresponding second-order condition (SSOC) defined in (8.16), see the counterexample of a simple unconstrained $C^{1,1}$ program presented in Example 6.22. ◇

8.2.2 Local Upper Lipschitz Continuity

Consider again the basic model (P) of (8.4) and the parametric pendants $P(t,(a,b))$ and (P)(a,b), where t and (a,b) vary. Let, as above, $F(s,t)$ (or $F(s) := F(s,t^0)$ if $t = t^0$ is fixed) denote the associated Kojima function. In this subsection, we are interested in necessary and sufficient conditions for the local upper Lipschitz continuity of the critical point mappings $\tilde{S}$ and $S = F(\cdot,t^0)^{-1}$, respectively. Similarly to the case of strong regularity, we again discuss quadratic approximations and geometrical interpretations, however, now the Thibault derivative TF is replaced by the contingent derivative CF.

Recall that a multifunction Σ from $\mathbb{R}^\nu$ to $\mathbb{R}^\lambda$ is said to be *locally upper Lipschitz* (briefly *locally u.L.*) at $(q^0, s^0) \in \operatorname{gph}\Sigma$ if there are positive real numbers $\varrho, \delta, \varepsilon$ such that

$$\Sigma(q) \cap B(s^0,\varepsilon) \subset B(s^0, \varrho\|q - q^0\|) \qquad \text{whenever } \|q - q^0\| \le \delta. \tag{8.19}$$

Again we note that definition (8.19) includes that s^0 is isolated in $\Sigma(q^0)$, but it does not include that $\Sigma(q) \cap B(s^0,\varepsilon)$ is nonempty for q near q^0.

As above, we use the notation

$$\mathcal{J}_C(y^0) = \left\{ (\alpha,\beta) \,\middle|\, \begin{array}{l} \alpha_i = 0\,(i \in I^-),\ \beta_i = 0\,(i \in I^+), \\ \alpha_i\beta_i = 0,\ \ \alpha_i \ge 0 \ge \beta_i\ (i \in I^0) \end{array} \right\} \tag{8.20}$$

for given $y^0 \in \mathbb{R}^m$ and with I^-, I^+, I^0 according to (8.9). Now, Theorem 7.8 which was proved in the context of generalized Kojima functions immediately gives the following characterization theorem.

Theorem 8.11 (locally u.L. $F(\cdot,t^0)^{-1}$). *Let (x^0,y^0) be a critical point of the problem (P)=P(t^0,0). For $f = f(\cdot,t^0)$, $g_i = g_i(\cdot,t^0) \in C^{1,1}$, and with the notation (8.9)–(8.11) and (8.20), the critical point map $S = F(\cdot,t^0)^{-1}$ is locally upper Lipschitz at $(0,s^0)$ if and only if the C–stability system*

$$\begin{array}{llll} Q_C(u) & + & A^{\mathsf T}\alpha & \ni 0, \\ Au & - & \beta & = 0, \quad (\alpha,\beta) \in \mathcal{J}_C(y^0), \end{array} \tag{8.21}$$

has only the trivial solution $(u,\alpha,\beta) = 0$. ◇

It is worth noting that *for unconstrained programs,* the condition of Theorem 8.11 is reduced to the second-order criterion

$$0 \notin C[Df](x^0)(u) \qquad \forall u \ne 0,$$

which would imply for $f \in C^2$ even strong regularity. However, in the case of $C^{1,1}$ programs, this criterion in general does not imply *upper regularity*, i.e., the existence of critical points of *slightly* perturbed problems cannot be guaranteed, see the following simple example.

Example 8.12 (no upper regularity for $C^{1,1}$ programs). Put $f(x) = \frac{1}{2}x|x|$. Here, $f'(x) = |x|$, $C[f'](0)(u) = |u|$ and $u\,C[f'](0)(u) = u|u|$. The origin is not a minimizer, but a stable critical point in the sense of local upper Lipschitz behavior. Indeed, for $a \geq 0$, stationary points $\xi(a)$ to $f(x) - ax$ exist (not uniquely) near $x^0 = 0$ and $|\xi(a) - x^0| \leq |a|$. If $a < 0$ then $S(a) = \emptyset$. Moreover, replacing the generalized derivative we obtain $0 \in T[f'](0)(u) = [-|u|, |u|]\ \forall u$.
◇

If all data in the problem $P(t,p)$, (t,p) near $(t^0,0)$, are supposed to be $C^{1,1}$ functions with respect to a finite-dimensional (s,t), the Kojima function (8.2) is locally Lipschitzian with respect to (s,t). Hence, with $s^0 = (x^0,y^0)$, $F(s^0,t^0) = 0$, one has

$$F(s,t) = p \;\Leftrightarrow\; F(s,t^0) = p + G(s,t), \quad \text{where } G(s,t) := F(s,t^0) - F(s,t),$$

and there exist $\mu > 0$ and $\gamma > 0$ such that $\|F(s,t) - F(s,t^0)\| \leq \gamma\|t - t^0\|$ for all $(s,t) \in (s^0,t^0) + \mu B$. Thus, if

$$q \mapsto S(q) = \{s | F(s,t^0) = q\}$$

is locally u.L. at $(0,s^0)$ with constants $\varrho, \delta, \varepsilon$ according to (8.19) (put there $q^0 = 0$ and $\Sigma = S$), then each $s \in \widetilde{S}(t,p) = S(p + G(s,t))$ with $(s,t) \in (s^0,t^0) + \mu B$, $\|s - s^0\| < \varepsilon$ and $\|p\| + \gamma\|t - t^0\| < \delta$ satisfies the estimate

$$\|s - s^0\| \;\leq\; \varrho\|p + G(s,t)\| \;\leq\; \varrho(\|p\| + \gamma\|t - t^0\|).$$

This means that $\widetilde{S}$ is locally u.L. at (t^0,s^0). From these observations, we immediately obtain

Corollary 8.13 (nonlinear variations, u.L.). *Let $\mathcal{T} \subset \mathbb{R}^r$ be open, and suppose that $f, g_i : \mathbb{R}^n \times \mathcal{T} \to \mathbb{R}$ belong to the class $C^{1,1}$. Then the critical point map $\widetilde{S}$ of the parametric program $P(t,p)$, (t,p), near $(t^0,0)$, is locally upper Lipschitz at $(t^0,0,s^0)$ if and only if the critical point map S of the canonically perturbed program $(P)(p)$=$P(t^0,p)$, p near 0, is locally upper Lipschitz at $(0,s^0)$.*
◇

In view of the previous corollary, we now again concentrate ourselves to *canonically perturbed* programs. The next discussions partially specialize facts known from Section 7.3.

Reformulation of the C-Stability System

In what follows, we give a reformulation of the C-stability system (8.21), which will allow to reduce in the C^2 case the characterization of a locally u.L. critical point to the question whether the reference point is an *isolated* critical point (i.e., the unique critical point in some neighborhood of it) to the assigned quadratic problem $(PQ)_0$=$(PQ)(0,0)$ defined in §8.2.1. It will be also seen how the quadratic problems can be substituted for $(f,g) \in C^{1,1}$.

Replacing TF by CF, we may apply similar arguments as above in the case of strong regularity. As above, let

$$Y^0 = \{y \mid (x^0, y) \text{ is critical for (P)}\},$$

$$U^C(y^0) = \{u \mid A_i u \le 0 \text{ if } y_i^0 = 0 \text{ and } A_i u = 0 \text{ if } y_i^0 > 0\}.$$

Further, we define

$$U^0 = \{u \mid A_i u \le 0 \text{ if } g_i(x^0) = 0 \text{ and } Df(x^0)u \le 0\}.$$

For some given $y^0 \in Y^0$, the cones coincide, since for each vector u satisfying $A_i u \le 0 \; \forall i \in I^+(y^0) \cup I^0(y^0)$, one has

$$u \in U^C(y^0) \;\Rightarrow\; Df(x^0)u = -\Sigma_i y_i^{0+} A_i u = 0 \;\Rightarrow\; u \in U^0,$$

and

$$u \in U^0 \;\Rightarrow\; Df(x^0)u = -\Sigma_i y_i^{0+} A_i u \le 0 \;\Rightarrow\; A_i u = 0 \text{ if } y_i^0 > 0 \;\Rightarrow\; u \in U^C(y^0).$$

Remark 8.14 (reformulation of the C-stability system). By Theorem 8.11, the critical point map $S = F^{-1}$ of the canonically perturbed program $(P)(a, b)$, $(a, b) \in \mathbb{R}^{n+m}$, is locally u.L. at $((0, 0), (x^0, y^0)) \in \operatorname{gph} S$ if and only if the system (8.21) has only the trivial solution $(u, \alpha, \beta) = 0$.

Using the structure of $\mathcal{J}_C(y^0)$, some point (u, α, β) solves (8.21) if and only if $Au = \beta$, $\alpha_i = 0$ $(i \in I^-)$, and, for some $q(u) \in Q_C(u)$, the point u is an optimal solution of the linear program

$$(L)_{q(u)} \qquad \min_x \{\langle q(u), x\rangle \mid A_i x \le 0 \; \forall i \in I^0 \text{ and } A_i x = 0 \; \forall i \in I^+\},$$

where again $I^0 = I^0(y^0)$ and $I^+ = I^+(y^0)$. If $(f, g) \in C^2$, we have $q(u) = D^2_{xx}L(x^0, y^0)u = Hu$, and so we arrive at the auxiliary problem

$$\min_x \{\tfrac{1}{2}\langle x, Hx\rangle \mid A_i x \le 0 \; \forall i \in I^0 \text{ and } A_i x = 0 \; \forall i \in I^+\}. \tag{8.22}$$

From Theorem 7.6 we know that the set $Z = CF^{-1}((0,0), (x^0, y^0))(a, b)$ is just defined by all points

$$(u, \alpha + Au - b)$$

satisfying the perturbed system

$$Q_C(u) + A^{\mathsf{T}}\alpha \ni a, \quad (\alpha, Au - b) \in \mathcal{J}_C(y^0),$$

i.e., if $g(x^0) = 0$ (without loss of generality), Z is given by all those KKT points (u, α) of the perturbed linear program

$$(L)_{q(u),a,b} \quad \min_x \{\langle q(u) - a, x\rangle \mid A_i x \le b_i \; \forall i \in I^0 \text{ and } A_i x = b_i \; \forall i \in I^+\},$$

which satisfy, in addition, $q(u) \in Q_C(u)$. In the C^2 case, this is the quadratic program (8.22), *perturbed* by a and b. ◇

The remark tells us that the analysis of system (8.21) is nothing else than the analysis of a family of linear optimization problems and of quadratic problems, respectively. So it is not surprising that the roots of the following statements are basically quadratic parametric optimization.

Lemma 8.15 (auxiliary problems). *Some point* (u, α, β) *solves (8.21) if and only if* $x = u$ *is a stationary point of*

$$\begin{aligned} &\min_x\{\langle q(u), x\rangle \mid x \in U^0\} \text{ with some } q(u) \in Q_C(u) && \text{for } (f,g) \in C^{1,1};\\ &\min_x\{\tfrac{1}{2}\langle x, D^2_{xx}L(x^0, y^0)x\rangle \mid x \in U^0\} && \text{for } (f,g) \in C^2, \end{aligned} \tag{8.23}$$

respectively. ◇

Proof. Recall that $U^0 = U^C(y^0)$ for every $y^0 \in Y^0$. So, given any objective function $h \in C^1$, the *stationary* points of the problems

$$\min\{h(x) \mid x \in U^0\}$$

and

$$\min\{h(x) \mid x \in U^C(y^0)\}$$

(defined as x-**part** of KKT-points) coincide because all constraints in $U^0 = U^C(y^0)$ are linear. In particular, this holds both for

$$h(x) = \langle q(u), x\rangle \text{ and } h(x) = \tfrac{1}{2}\langle x, D^2_{xx}L(x^0, y^0)x\rangle.$$

So our Remark 8.14 finishes the proof. □

Geometrical Interpretation

Next we adapt the geometrical interpretation of F^{-1} being locally u.L., which was given for *generalized Kojima–functions* in the Remarks 7.11 and 7.12. The cone U^C considered there now becomes

$$U^C = U^C(y^0) = U^0,$$

(see the previous proof), while

$$K^C(u) = \{ \sum_{i \in I^+ \cup I^0} \lambda_i A_i \mid \lambda_k \in \mathbb{R},\ k \in I^+;\ \lambda_i \geq 0 \text{ and } \lambda_i A_i u = 0,\ i \in I^0\}.$$

The cone $K^C(u)$ is just polar to the tangent cone $C(u, U^C)$. Since $U^C = U^0$, so $K^C(u)$ also coincides with the polar cone $K(u)$ of $C(u, U^0)$. By putting $J(x^0) = \{i | g_i(x^0) = 0\}$, the latter can be written as

$$K(u) = \left\{ \lambda_0 Df(x^0) + \sum_{i \in J(x^0)} \lambda_i A_i \,\middle|\, \begin{array}{l} \lambda_0 \geq 0, \lambda_0 Df(x^0)u = 0, \\ \lambda_i \geq 0, \lambda_i A_i u = 0,\ i \in J(x^0) \end{array} \right\}. \tag{8.24}$$

So we obtain from the Remarks 7.11 and 7.12 the following corollary.

Corollary 8.16 (geometrical interpretation, u.L.). *The critical point map $S = F^{-1}$ of (P)((a,b)), $(a,b) \in \mathbb{R}^{n+m}$, is locally u.L. at $((0,0),(x^0,y^0)) \in$ gph S if and only if*

(i) *strict MFCQ and*
(ii) $Q_C(u) \cap K(u) = \emptyset \ \forall u \in U^0 \setminus \{0\}$

are satisfied. Moreover,

(i) *is violated if and only if (8.21) has a nontrivial solution with $u = 0$, and*
(ii) *is violated if and only if (8.21) has a solution with $u \neq 0$.*

◇

Now $\xi \in Q_C(u) \cap K^C(u)$ and $u \in U^C$ imply $\langle \xi, u \rangle = 0$. Thus, (ii) holds true under the sufficient condition

$$\text{(SSOC)'} \qquad \langle \xi, u \rangle \neq 0 \ \ \forall u \in U^0 \setminus \{0\} \text{ and } \xi \in Q_C(u).$$

Direct Perturbations for the Quadratic Approximation

We show for C^2 problems, how the system (8.21) indicates *where*, near (x^0, y^0), there is a second KKT-point for $\text{PQ}((0,0))$. In contrast to the corresponding geometrical interpretation of strong regularity, here we have only to deal with the unperturbed quadratic approximation, i.e., we study with $H = D^2_{xx}L(x^0, y^0)$,

$$(\text{PQ})_0 \quad \min\{Df(x^0)(x - x^0) + \tfrac{1}{2}\langle x - x^0, H(x - x^0)\rangle \mid A(x - x^0) \le 0\},$$

where we again assume, without loss of generality, $g(x^0) = 0$ (i.e., $I^- = \emptyset$).

Lemma 8.17 *(two critical points, quadratic problems). Let $(f,g) \in C^2, g(x^0) = 0$ and (x^0, y^0) be a KKT-point for $(PQ)_0$ (or, equivalently, for (P)). Let $\theta(y^0) = \min\{y_i^0 | i \in I^+\}$ if $I^+ \neq \emptyset$ and $\theta(y^0) = \infty$ if $I^+ = \emptyset$.*

(i) *If (x,y) is a critical point for $(PQ)_0$ and $|y_i - y_i^0| \le \theta(y^0) \ \forall i \in I^+$, then $(u, \alpha, \beta) = (x - x^0, y^+ - y^{0+}, Au)$ solves (8.21).*
(ii) *If (u, α, β) solves (8.21) and $|\alpha_i| \le \theta(y^0) \ \forall i \in I^+$, then $(x,y) = (x^0 + u, y^0 + \alpha)$ is a KKT-point for $(PQ)_0$.*

Proof. Some point $(x,y) = (x^0 + u, y)$ is critical for $(\text{PQ})_0$ if and only if

$$Df(x^0) + Hu + A^{\mathsf T} y^+ = 0 \text{ and } Au - y^- = 0.$$

Setting $\alpha = y^+ - y^{0+}$, $\beta = y^- - y^{0-}$ and using $Df(x^0) + A^{\mathsf T} y^{0+} = 0$ and $y^{0-} = 0$, this is

$$Hu + A^{\mathsf T}\alpha = 0, \ Au = \beta. \tag{8.25}$$

Proof of (i). Let (x,y) be critical for $(\text{PQ})_0$. So we already know that (8.25) holds, and we have only to show that $(\alpha, \beta) = (y^+ - y^{0+}, y^- - y^{0-}) \in \mathcal{J}_C(y^0)$, provided that the condition $|y_i - y_i^0| \le \theta(y^0)\ \forall i \in I^+$ holds. Indeed, from $0 \le y^+ = y^{0+} + \alpha$ and $0 \ge y^- = y^{0-} + \beta$ we obtain that

$$\begin{array}{lll} \text{if } i \in I^0 & \text{then} & \alpha_i = y_i^+ \ge 0 \text{ and } 0 \ge \beta_i, \\ \text{if } \alpha_i > 0 & \text{then} & y_i^+ = (y^{0+} + \alpha)_i > 0, \\ & & \text{hence } 0 = y_i^- = y_i^{0-} + \beta_i = \beta_i. \end{array}$$

Thus, in any case, $\alpha_i \beta_i = 0$ if $i \in I^0$. Moreover,

$$\begin{array}{lll} \text{If } i \in I^+ & \text{then} & |y_i - y_i^0| \le \theta(y^0) \text{ and } y_i^0 > 0 \text{ yield } y_i \ge 0 \\ & & \text{as well as } 0 = y_i^- = y_i^{0-} + \beta_i = \beta_i. \end{array}$$

Summarizing, this is $(\alpha, \beta) \in \mathcal{J}_C(y^0)$, so (u, α, β) solves (8.21).

Proof of (ii). Let (u, α, β) solve (8.21) at (0,0). Using $(x, y) = (x^0+u, y^0+\alpha)$ and $(\alpha, \beta) \in \mathcal{J}_C(y^0)$, we see that

$$\begin{array}{lll} \text{for } i \in I^+, & \text{it holds} & \beta_i = 0 \text{ and } y_i = (y^0 + \alpha)_i \ge 0 \\ & & \text{since } y_i^0 > 0 \text{ and } |\alpha_i| \le \theta(y^0), \text{while} \\ \text{for } i \in I^0, & \text{it holds} & \beta_i \le 0, 0 \le \alpha_i = (y^0 + \alpha)_i = y_i \\ & & \text{as well as } 0 = \alpha_i \beta_i = y_i A_i u. \end{array}$$

Hence, since also (8.25) follows from (8.21), (x, y) is a KKT-point for $(\mathrm{PQ})_0$. □

Analogously to the discussion following Lemma 8.7, one has

Corollary 8.18 *Let* $(f, g) \in C^2$, $g(x^0) = 0$ *and* (x^0, y^0) *be a KKT-point for (P). Then, the critical point map* $S = F^{-1}$ *of (P)(a,b),* $(a, b) \in \mathbb{R}^{n+m}$*, is locally u.L. at* $((0,0), (x^0, y^0))$ *if and only if the point* (x^0, y^0) *is an isolated KKT-point of problem* $(PQ)_0$. ◇

Proof. If (8.21) has a nontrivial solution (u, α, β) then, using the solution $\lambda \cdot (u, \alpha, \beta)$ for small $\lambda > 0$, we find KKT-points $(x, y) \ne (x^0, y^0)$ arbitrarily close to (x^0, y^0) by Lemma 8.17(ii). Conversely, having such KKT-points for $(\mathrm{PQ})_0$, we find related $(u, \alpha, \beta) \ne 0$ by Lemma 8.17(i). Taking Theorem 8.11 into account, this yields the assertion. □

8.3 Stationary and Optimal Solutions under Perturbations

In this section, we characterize Lipschitz properties of the *stationary solution set maps* $\tilde{X}(t, p)$ and $X(p)$ of the parametric problems $\mathrm{P}(t, p)$ and(P)(p) introduced in §8.1. Let the general assumptions (8.6) be satisfied, i.e., x^0 is a stationary solution of $(\mathrm{P})=\mathrm{P}(t^0,0)$, $t^0 \in \mathcal{T}$, and $f, g_i \in C^{1,1}(\Omega,\mathbb{R})$ for some neighborhood Ω of (x^0,t^0).

The results will be used to give conditions for the Lipschitz behavior (in the one or the other sense) of perturbed local minimizers near a strict local minimizer of the initial problem.

8.3.1 Contingent Derivative of the Stationary Point Map

From the general theory developed in the chapters before, we know that generalized derivatives of the mappings X and $\tilde{X}$ play an essential role for characterizing Lipschitz stability. In this subsection we describe the contingent derivative of $\tilde{X}$ in terms of the contingent derivative of Kojima's function F.

Throughout Subsection 8.3.1, we put without loss of generality $t^0 = 0$ for the initial parameter, and we assume that

$$\begin{array}{l} x^0 \text{ is a stationary solution of } \mathrm{P}(0,0), \\ f, g_i \in C^{1,1}(\Omega,\mathbb{R}), \text{ where } \Omega \text{ is a neighborhood of } (x^0,0) \in \mathbb{R}^{n+r}, \\ \text{and } f, g_i \text{ have continuous second derivatives } f_{xt}, (g_i)_{xt} \text{ on } \Omega. \end{array} \tag{8.26}$$

By these assumptions, F is the product of a locally Lipschitz matrix-function $M(x,t)$ (cf. (8.7)) with the piecewise linear vector $N(y) = (1, y^{+\mathsf{T}}, (y-y^+)^\mathsf{T})^\mathsf{T}$. This yields, for fixed $t = 0$ and contingent-derivative $C_{(x,y)}F$ of F with respect to (x,y), the elementary product (or partial derivative) rule

$$C_{(x,y)}F(x^0,y^0,0)(u,v) = [C_xM(x^0,0)(u)]N(y^0) + M(x^0,0)[CN(y^0)(v)]$$

holds, which provides us CS via (6.56). For sufficiently small $\|(u,v,\tau)\|$ and some y^0 satisfying $F(x^0,y^0,0) = 0$, the formula

$$\begin{aligned} F(x^0+u, y^0+v, \tau) \in\ & [C_xM(x^0,0)(u) + M_t(x^0,0)\tau]N(y^0) \\ & + M(x^0,0)[CN(y^0)(v)] + o(u,\tau)B \end{aligned} \tag{8.27}$$

can also be easily shown, where here and in the following subscripts denote partial derivatives in the contingent and Fréchet sense, respectively, and B denotes a closed unit ball independently of the space under consideration. Indeed, the inclusion

$$M(x^0+u,\tau) \in C_xM(x^0,0)(u) + M_t(x^0,0)\tau + o_1(u,\tau)B$$

is true since M is locally Lipschitz and continuously differentiable with respect to t. The identity

$$N(y^0+v) = N(y^0) + CN(y^0)(v)$$

holds for small $\|v\|$ since N is piecewise linear. Rewriting F as a product by using these terms, (8.27) follows immediately.

Nevertheless, even for arbitrarily smooth involved data in (8.26) there is (up to now) no complete formula which describes for the stationary point mapping

$t \mapsto \mathcal{X}(t) := \{x | \exists y : F(x,y,t) = 0\}$ the contingent derivative $C\mathcal{X}$ under **MFCQ** *only*. However, the contingent derivative can be computed, under **MFCQ**, when including *additional canonical perturbations*, i.e., for the stationary point map $\tilde{X}$ defined in (8.8),

$$\tilde{X}(t,p) = \{x | F(x,y,t) = p \text{ for some } y\}.$$

The importance of adding such canonical (also called 'tilt') perturbations for characterizing stability properties of optimization and variational problems is well–established in the literature. In our context, a map similar to $C\tilde{X}((0,0),x^0)$ has been considered in [LR95]. There the authors assumed that f, g are C^2 functions and were able to put the additional tilt perturbation only in the objective function. So, their results are, in this smooth case, stronger than those of the following Theorem 8.19. On the other hand, our result applies to $C^{1,1}$**-problems** in which quadratic approximations of the input functions cannot be applied.

The Case of Locally Lipschitzian F

The following approach was recently presented in [KK01] and is based on the results obtained in [KK99a] for parametric $C^{1,1}$ programs under canonical perturbations only. In both papers, **MFCQ** was a crucial assumption. Note that this assumption may be slightly weakened, see [Kla00] and the remark following Theorem 8.19. To compare with the usual case of $F \in C^1$, let us, for the present, assume that M and N are C^1 functions, $F = M(x,t)N(y)$, and (x^0, y^0) is a regular zero of $F(x,y,0)$. Then we obtain by the classical implicit function theorem that

$$\begin{aligned} &\xi \in D\tilde{X}((0,0),x^0)(\tau,\pi) \\ &\Leftrightarrow \pi \in [DM(x^0,0)(\xi,\tau)]N(y^0) + M(x^0,0)DN(y^0)\mathbb{R}^m. \end{aligned}$$

Now we derive the same formula for Kojima's function in terms of contingent derivatives by considering all y^0 in the set $Y^0 = \{y \in \mathbb{R}^m | F(x^0,y,0) = 0\}$ which is bounded under **MFCQ**. Let M_t denote the partial derivative of M with respect to t.

Theorem 8.19 (*CX* under **MFCQ**). *Let* $F = M(x,t)N(y)$ *be the Kojima function of problem (8.26), and let* **MFCQ** *be satisfied at* $x^0 \in \tilde{X}(0,0)$. *Then, it holds*

$$\xi \in C\tilde{X}((0,0),x^0)(\tau,\pi) \tag{8.28}$$

if and only if

$$\pi \in [C_xM(x^0,0)(\xi) + M_t(x^0,0)\tau]N(y^0) + M(x^0,0)[CN(y^0)(\mathbb{R}^m)] \tag{8.29}$$

for some $y^0 \in Y^0$. ◇

Proof. Condition (8.28) means that, for some sequence $\theta = \theta_k \downarrow 0$ and certain o-type functions o_j , the points

$$x = x^0 + \theta\xi + o_1(\theta),\ t = \theta\tau + o_2(\theta),\ p = \theta\pi + o_3(\theta) \text{ and related } y = y(\theta)$$

fulfill

$$F(x, y, t) = p.$$

Due to **MFCQ** we find (for some subsequence of θ) some

$$y^0 \in Y^0 \text{ such that } y = y^0 + v(\theta) \text{ where } \|v(\theta)\| \text{ vanishes.} \tag{8.30}$$

Then (8.27) tells us that

$$\begin{aligned} \theta\pi + o_3(\theta) &= F(x, y^0 + v(\theta), t) \\ &\in \theta[C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau]N(y^0) \\ &\quad + M(x^0, 0)[CN(y^0)(v(\theta))] + o_4(\theta)B, \end{aligned}$$

and after division by θ,

$$\begin{aligned} \pi + \theta^{-1}o_3(\theta) &\in [C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau]N(y^0) \\ &\quad + M(x^0, 0)[CN(y^0)(\theta^{-1}v(\theta))] + \theta^{-1}o_4(\theta)B. \end{aligned}$$

Since N is piecewise linear and $M(x^0, 0)$ is a fixed matrix, the set

$$\Psi := M(x^0, 0)[CN(y^0)(\mathbb{R}^m)],$$

is closed as a finite union of polyhedral cones, and Ψ contains, by the above inclusion, certain elements

$$\psi(\theta) \in \pi + \theta^{-1}o_3(\theta) - [C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau]N(y^0) - \theta^{-1}o_4(\theta)B.$$

On the other hand, the set $C_x M(x^0, 0)(\xi)$ is compact because M is locally Lipschitz. So there is some accumulation point ψ of the sequence $\psi(\theta)$ under consideration and

$$\psi \in \Psi \cap (\pi - [C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau]N(y^0)).$$

Thus, since the intersection is not empty, (8.29) is true.

Conversely, let (8.29) hold and note that N is directionally differentiable. Then one finds certain

$$\mu \in C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau, \ v \in \mathbb{R}^m$$

as well as some sequence $\theta = \theta_k \downarrow 0$ and related $o_j(\theta)$ such that, with $\eta = CN(y^0)(v)$,

$$\pi = \mu N(y^0) + M(x^0, 0)\eta$$

and

$$\mu = \lim \theta^{-1}(M(x^0 + \theta\xi + o_1(\theta), t\tau + o_2(\theta)) - M(x^0, 0)).$$

For small θ (of the given sequence), we know that

$$N(y^0 + \theta v) - N(y^0) = \theta\eta.$$

After an elementary calculation, this tells us, because of $F(x^0, y^0, 0) = 0$ and $F = MN$, that

$$F(x^0 + \theta\xi + o_1(\theta), y^0 + \theta v, \theta\tau + o_2(\theta)) = \theta\pi + o_3(\theta).$$

The latter gives (8.28). □

Remark 8.20 (selection property). The crucial assumption **MFCQ** was only used in the proof of the "only if" direction in order to show that assertion (8.30) holds. Indeed, it is immediately clear from this proof (see also [Kla00]) that **MFCQ** may be replaced by the following *selection property*:

> Given a sequence $(x^k, t^k, p^k) \to (x^0, t^0, p^0)$ (put above $p^0 = 0$) with $x^k \in \tilde{X}(t^k, p^k)$, there exists a sequence of associated multipliers $y^k \in \tilde{Y}(x^k, t^k, p^k)$ such that $\{y^k\}$ has an accumulation point $y^0 \in Y^0$,

where $Y^0 = \tilde{Y}(x^0, t^0, p^0)$. In particular, Corollary 2.9 says that under **MFCQ**, one may take any sequence $\{y^k\}$ of associated multipliers. Further, the so–called constant *rank condition* guarantees the selection property, see [Kla00, Lemma 5]. In the case of fixed $t = t^0$, a linearly constrained program also fulfills the selection property, this will be shown in Lemma 8.29 below. ◇

The Smooth Case

In specializing the result of Theorem 8.19 to C^2**–problems**, we again suppose $g(x^0, 0) = 0$ and hence $y^0 \geq 0$. This convention is used in the following discussion.

For C^2**- problems**, the multivalued term $C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau$ becomes (single-valued and) linear in ξ and τ, namely,

$$C_x M(x^0, 0)(\xi) + M_t(x^0, 0)\tau = \{M_x(x^0, 0)\xi + M_t(x^0, 0)\tau\}.$$

Explicitly, the term $DM(x^0, 0)(\xi, \tau) = [M_x(x^0, 0)\xi + M_t(x^0, 0)\tau]$ for $M(x, t)$ according to (8.7) is a matrix of the form

$$\begin{pmatrix} f_{xx}\xi + f_{xt}\tau & (g_{xx}\xi + g_{xt}\tau)^\mathsf{T} & 0 \\ g_x\xi + g_t\tau & 0 & 0 \end{pmatrix}, \tag{8.31}$$

where all first and second–order partial derivatives are taken at $(x^0, 0)$. So, (8.29) requires, writing $CN(y^0)(v) = \{(0, \alpha, \beta)\}$,

$$\begin{aligned} \pi_1 &= f_{xx}\xi + f_{xt}\tau + (y_g^0)^+(g_{xx}\xi + g_{xt}\tau) + \alpha g_x \\ \pi_2 &= g_x\xi + g_t\tau - \beta \end{aligned}$$

for some $y^0 \in Y^0$ and $v \in \mathbb{R}^m$.

Next consider, for comparison, the following quadratic approximation of problem (8.26) at $(x^0, 0)$:

$$\begin{array}{lll} \min & [f_x^0 + f_{xt}t](x - x^0) + \frac{1}{2}(x - x^0)^\mathsf{T} f_{xx}(x - x^0) & \\ \text{s.t.} & [g_x^0 + g_{xt}t](x - x^0) + \frac{1}{2}(x - x^0)^\mathsf{T} g_{xx}(x - x^0) & \leq 0 \end{array} \tag{8.32}$$

where $f_x^0 = f_x(x^0, 0)$, similarly g_x^0. When specializing (8.7), the related matrix $M = M_Q$ of Kojima's function, assigned to (8.32), attains the form

$$\begin{pmatrix} f_x^0 + f_{xt}t + f_{xx}\Delta x & [g_x^0 + g_{xt}t + g_{xx}\Delta x]^T & 0 \\ a_g(x,t) & 0 & -E \end{pmatrix},$$

where $a_g(x,t)$ is the constraint function of the problem (8.32), and $\Delta x = x - x^0$. So, the derivative $DM_Q(x^0,0)$ coincides with $DM(x^0,0)$. Moreover, at $(x^0,0)$, the problems (8.26) and (8.32) have the same set Y^0 of dual vectors. By Theorem 8.19, this yields

Corollary 8.21 *If $f, g \in C^2$ and MFCQ holds at $(x^0,0)$, then the derivative $C\tilde{X}((0,0),x^0)$ coincides for the problems (8.26) and (8.32).*

Note that in the quadratic–quadratic program (8.32), the parameter t appears now only linearly in the first–order terms with respect to x.

8.3.2 Local Upper Lipschitz Continuity

In this paragraph, we characterize local upper Lipschitz continuity of the stationary point maps X and $\tilde{X}$ of the parametric problems $(\mathrm{P})(p)$ and $\mathrm{P}(t,p)$, respectively, and this by supposing the data belong to the class $C^{1,1}$ (or C^2) and MFCQ holds at the initial stationary solution. It turns out that the (generalized) second-order approximation developed in §8.2.2 carries over to stationary and optimal solutions, where the representation of CX given in the previous paragraph plays an essential role. Moreover, we discuss these second-order type conditions with respect to local minimizers.

As above, we consider the parametric problems $\mathrm{P}(t,p)$ and $(\mathrm{P})(p)=\mathrm{P}(t^0,p)$ according to (8.4) and (8.5), and we suppose at least that

$$\begin{array}{l} x^0 \text{ is a stationary solution of } (\mathrm{P})(0)=\mathrm{P}(t^0,0), \\ f,\, g_i \in C^{1,1}(\Omega,\mathbb{R}),\ \ \Omega \text{ is a neighborhood of } (x^0,t^0) \in \mathbb{R}^{n+r}. \end{array} \tag{8.33}$$

Then, for the Kojima function $F(x,y,t) = M(x,t)N(y)$, M is Lipschitz on Ω. Note once more that equality constraints could be included without any problems, we have not done this for brevity of presentations.

The presentation is based on the authors' papers [KK99a, KK01] and is related (in the C^2 case) to [Lev96, DR98, HG99, LPR00].

Two Illuminating Examples

To illustrate typical difficulties, we start by two interesting examples. The first example illustrates that even for smooth data standard first and second–order optimality conditions do not imply that the solutions behave locally upper Lipschitz. Some "stronger" second–order sufficient condition is needed. Note that this example was first given in [KK85], it modifies a proposal made in [GT77].

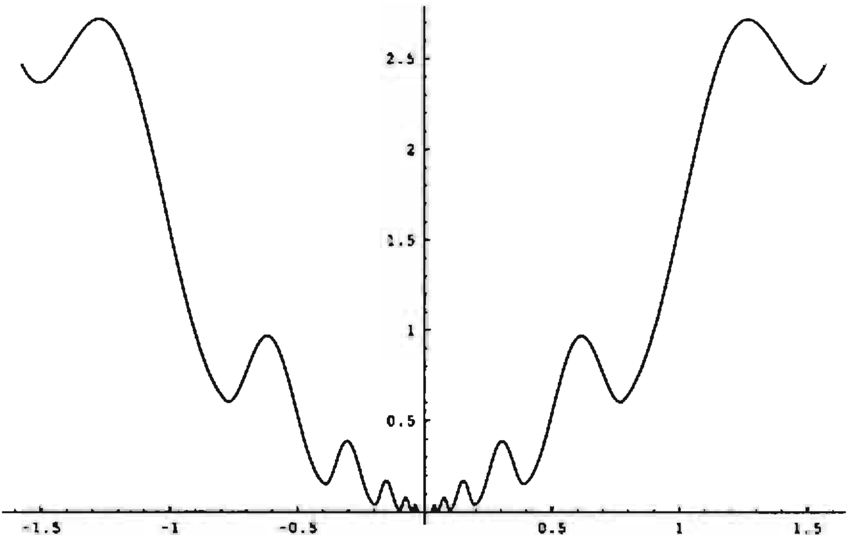

Figure 8.1: Strict local minimizer is not isolated for $f \in C^{1,1}$.

Example 8.22 Minimize x_2 subject to $x_2 \geq x_1^2$, $x_2 \geq b$, where $b \in \mathbb{R}$ is a parameter. Then the optimal solution set (= set of stationary points) is $X(b) = \{(0,0)\}$ if $b \leq 0$ and $X(b) = \{(x_1, b) | -\sqrt{b} \leq x_1 \leq \sqrt{b}\}$ if $b > 0$. In this example of a perturbed convex quadratic program, the solution set mapping is not locally upper Lipschitz. Note that MFCQ holds, and (0,0) is a strict local minimizer of order 2. ◇

The next example illuminates an essential difference between C^2**–optimization** problems and $C^{1,1}$**–problems**, see Figure 8.1.

Example 8.23 (Ward [War94, Ex. 3.1]) Let Z denote the set of integers, and define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) := \begin{cases} 0, & if\ x = 0 \\ x^2 + x(\cos[2^{n+2}x - 3\pi] + 1)/2^n, & if\ x \in I_n,\ n \in \mathbb{Z}, \\ f(-x), & if\ x < 0, \end{cases}$$

where $I_n := \{x | \pi/2^{n+1} < x \leq \pi/2^n\}$. f is a $C^{1,1}$ function, and $x^0 = 0$ is a strict local minimizer of order 2. For the derivative of f at $x \in I_n$, $n \in \mathbb{Z}$, we have

$$2^n f'(x) = (\cos[2^{n+2}x - 3\pi] + 1) + 2^{n+1}x(1 - 2\sin[2^{n+2}x - 3\pi]).$$

It is easy to see that for each n, f' has two zeros in I_n and

$$2^n f'(x) = \begin{cases} 3\pi/2 + 2 & > & 0, & if\ x = 3\pi/2^{n+2}, \\ -7\pi/4 + 1 & < & 0, & if\ x = 7\pi/2^{n+3}, \\ 2\pi & > & 0, & if\ x = \ \pi/2^n. \end{cases}$$

Hence, each of the intervals I_n, $n \in \mathbb{Z}$, contains a local minimizer and a local maximizer of f. This example shows: for a $C^{1,1}$ function it may happen that in any neighborhood of a strict local minimizer of order two there are infinitely many other stationary points of this function. Hence, in particular, $X = S$ is not locally u.L. at (0,0). For C^2 functions, unconstrained strict local minimizers of order 2 are automatically isolated stationary points. Under C^2 constraints, to get this property, one has additionally to assume that MFCQ holds, see Robinson [Rob82]. ◇

The latter example was originally given in [War94] to illustrate that some second-order optimality condition [Kla92] in terms of Clarke's generalized Hessian is not equivalent to strict local minimality of order 2. Since this example concerns the unconstrained minimization of a $C^{1,1}$ function, it applies also to "most regular" constrained problems.

Injectivity and Second-Order Conditions

Combining Theorem 8.19 on the representation of CX and the Theorems 7.5 and 7.6 on representations of CF, we now will easily obtain that the stationary solution set mappings X and $\tilde{X}$ are locally upper Lipschitz (locally u.L.) if and only if a restricted injectivity condition on the Kojima function F is satisfied.

Given a stationary solution x^0 of (P)=$P(t^0, 0)$, we say that (x^0, Y^0) satisfies the *injectivity condition for* $CF = CF(\cdot,\cdot,t^0)$ *with respect to u* if for all $y \in Y^0$ and all $u \neq 0$, one has $0 \notin CF(x^0, y)(u, \mathbb{R}^m)$, where $Y^0 = \tilde{Y}(x^0, t^0, 0)$ is the set of multipliers associated with x^0. If the included parameters are obvious, we will also say that (x^0, Y^0) satisfies *CF-injectivity w.r. to u*. If the same sense, *TF-injectivity w.r. to u* will be used.

If the original problem (P) is a linear program, then one easily observes: (x^0, Y^0) satisfies CF-injectivity w.r. to u iff x^0 is the unique solution of (P)(0). (x^0, Y^0) satisfies TF-injectivity w.r. to u iff x^0 is, for each $y \in Y^0$, the unique solution of

$$\min f(x) \quad \text{s.t.} \quad g_i(x) = 0 \ \forall i \in I^+(y).$$

Theorem 8.19, specialized to X, says that for a stationary solution x^0 of (P),

$$\begin{aligned} &\xi \in CX(0, x^0)(\pi) \\ \Leftrightarrow\ &\exists y^0 \in Y^0 : \ \pi \in CM(x^0)(\xi)N(y^0) + M(x^0)[CN(y^0)(\mathbb{R}^m)]. \end{aligned} \tag{8.34}$$

The basic notation is used according to §8.1, in particular, x^0 is a fixed stationary solution of (P). Moreover, as defined in §8.2.2, consider the cones,

$$U^C(y) = \{u \mid A_i u \leq 0 \text{ if } y_i = 0 \text{ and } A_i u = 0 \text{ if } y_i > 0\},$$

and

$$U^0 = \{u \mid A_i u \leq 0 \text{ if } g_i(x^0) = 0 \text{ and } Df(x^0)u \leq 0\}.$$

For some given $y \in Y^0$, these cones coincide, see the dicussion prior to Remark 8.14. We also recall the definition (8.24) of the polar cone $K(u)$ of $C(u, U^0)$, namely, with $J(x^0) = \{i | g_i(x^0) = 0\}$,

$$K(u) = \left\{ \lambda_0 Df(x^0) + \sum_{i \in J(x^0)} \lambda_i A_i \;\middle|\; \begin{array}{l} \lambda_0 \geq 0, \lambda_0 Df(x^0)u = 0, \\ \lambda_i \geq 0, \lambda_i A_i u = 0, \ i \in J(x^0) \end{array} \right\}.$$

Given any $y \in Y^0$ and any direction $u \in \mathbb{R}^n$, we write

$$Q_C(y)(u) = C_x[D_x L](x^0, y)(u),$$

because in the following y may vary.

Theorem 8.24 (locally u.L. stationary points). *Consider the parametric programs (P)(p) and P(t, p), t near t^0, p near 0, and let f, g_i be $C^{1,1}$ functions with respect to (x, t). Suppose that $x^0 \in X(0) = \tilde{X}(t^0, 0)$ satisfies MFCQ. Then the following properties are mutually equivalent.*

(i) The stationary point map X of (P)($\cdot$) is locally u.L. at $(0, x^0)$.
(ii) The stationary point map $\tilde{X}$ of P($\cdot,\cdot$) is locally u.L. at $((t^0, 0), x^0)$.
(iii) (x^0, Y^0) satisfies the injectivity condition for $CF(\cdot, \cdot, t^0)$ with respect to u.
(iv) For each $y \in Y^0$, the C-stability system $\{Q_C(y)(u) + A^T\alpha \ni 0,\ Au - \beta = 0,\ (\alpha, \beta) \in \mathcal{J}_C(y)\}$ has no solution (u, α, β) with $u \neq 0$.
(v) For each $u \in U^0 \setminus \{0\}$ and all $y \in Y^0$, one has $Q_C(y)(u) \cap K(u) = \emptyset$.

◇

Proof. The equivalence of (iii) and (iv) is a direct consequence of Theorem 7.6 (see also Remark 8.14). Because of the product rule of Theorem 7.5

$$C[F(\cdot, \cdot, t^0)](x^0, y)(\xi, \eta) = [CM(x^0)(\xi)]N(y) + M(x^0)[CN(y)(\eta)],$$

Theorem 8.19 applied to the mapping $X = \tilde{X}(t^0, \cdot)$ and (8.34) immediately yield the equivalence of (i) and (iii). The equivalence of (iv) and (v) follows from the second part of Corollary 8.16.

In order to show the equivalence of (i) and (ii), we use, similarly to the proof of Corollary 8.13, the reformulation

$$F(x, y, t) = p \Leftrightarrow F(x, y, t^0) = p + G(x, y, t), \tag{8.35}$$

where $G(x, y, t) := F(x, y, t^0) - F(x, y, t)$.

Since (ii) $\Rightarrow$ (i) is trivially fulfilled, it suffices to prove the direction (i) $\Rightarrow$ (ii). Suppose $X = \tilde{X}(\cdot, t^0)$ is locally u.L. at $(0, x^0)$, i.e., there exist $\varepsilon, \delta, c_X > 0$ such that

$$\tilde{X}(t^0, p) \cap B(x^0, \varepsilon) \subset \{x^0\} + c_X \|p\| B \quad \forall p \in B(0, 2\delta). \tag{8.36}$$

Assuming on the contrary that (ii) is not true, we conclude that there is a sequence $\{(x^k, t^k, p^k)\}$ with $x^k \in \tilde{X}(t^k, p^k)$ (and associated $y^k \in \tilde{Y}(x^k, t^k, p^k)$) such that $(x^k, t^k, p^k) \to s^0 = (x^0, t^0, 0)$ and

$$\|x^k - x^0\| / (\|t^k - t^0\| + \|p^k\|) \to \infty \tag{8.37}$$

By MFCQ, Corollary 2.9 particularly yields that $\tilde{Y}$ is u.s.c. at s^0, and so, $\{y^k\}$ has an accumulation point $y^0 \in \tilde{Y}(s^0)$. By $F \in C^{0,1}$, $\varepsilon > 0$ may be regarded being so small that

$$F \text{ is Lipschitz on } B((x^0, y^0, t^0), \varepsilon) \text{ with some rank } c_F. \tag{8.38}$$

Hence, $G(x, y, t) \in B(0, \delta)$ if (x, y, t) near (x^0, y^0, t^0). Then (8.36) and (8.38) imply that for sufficiently large k,

$$\|x^k - x^0\| \le c_X \|p^k + G(x^k, y^k, t^k)\| \le c_X (\|p^k\| + c_F \|t^k - t^0\|)$$

is satisfied, which contradicts (8.37) and so completes the proof. □

Note. By a slight modification of the proof, one obtains that in the last theorem, MFCQ may be replaced by the selection property defined in Remark 8.20. In this form, the result was given in [Kla00]. ◇

Corollary 8.25 (second–order sufficient condition). *Suppose the assumptions of Theorem 8.24 are satisfied, in particular, let x^0 be an element of $X(0) = \tilde{X}(t^0, 0)$. If for each $y \in Y^0$, the condition*

$$SOCL: \quad 0 \notin \langle u, Q_C(y)(u) \rangle \ \forall u \in U^0 \setminus \{0\}$$

holds, then X and $\tilde{X}$ are locally upper Lipschitz at $(0, x^0)$ and $(t^0, 0, x^0)$, respectively. ◇

Proof. If X is not locally u.L. at $(0, x^0)$, then, by Theorem 8.24, there are some $y \in Y^0$ and a solution (u, α, β) of

$$Q_C(y)(u) + A^\mathsf{T}\alpha \ni 0,\ Au - \beta = 0,\ (\alpha, \beta) \in \mathcal{J}_C(y) \tag{8.39}$$

with $u \neq 0$. This implies $u \in U^0 \setminus \{0\}$. After scalar multiplication of the inclusion in (8.39) by u and of the equation in (8.39) by α, we obtain $0 = -\alpha^\mathsf{T}\beta \in u^\mathsf{T} Q_C(y)(u)$. By contraposition, this completes the proof. □

Remark 8.26 (illustration by examples). The injectivity condition of Theorem 8.24 is obviously not fulfilled in both Example 8.22 and Example 8.23.

Choose in Example 8.22 the multipliers $y_1 = 0$ and $y_2 = 1$ and the nontrivial direction $u = (1, 0)$. Then $Dg(x^0)_1 = Dg(x^0)_2 = (0, -1)$ implies that (8.39) can be satisfied with u and $\alpha = \beta = 0$.

In Example 8.23 no constraints appear, and (8.39) reduces to $0 \in Cf'(u)$ which is trivially satisfied for $u \neq 0$, by definition of f. ◇

Conditions via Quadratic Approximation

The quadratic approximations introduced in §8.2.2 with respect to the upper Lipschitz behavior of *critical* points to C^2 or $C^{1,1}$ programs are now be studied under the viewpoint of upper Lipschitz behavior of *stationary* points. By Theorem 8.24, we may restrict our analysis without loss of generality to *canonically perturbed* programs.

Theorem 8.27 (quadratic approximations). *Let x^0 be a stationary point of (P)=(P)(0), and suppose that x^0 satisfies MFCQ. Then, the following statements are equivalent:*

(i) The stationary point map X of (P)($\cdot$) is locally u.L. at $(0, x^0)$.

(ii) For all $y^0 \in Y^0$, the origin u = 0 is the unique point which, for some $q \in Q_C(y^0)(u)$, solves $\min\{\langle q, x\rangle \mid x \in U^0\}$.

For $(f, g) \in C^2$, condition (ii) coincides with the following one:

(iii) For all $y^0 \in Y^0$, the origin is the unique stationary point of the quadratic program

$$\min\{\langle x, D^2_{xx}L(x^0, y^0)x\rangle \mid x \in U^0\}. \tag{8.40}$$

◇

Note. To have, for $(f, g) \in C^2$, a relation to the second quadratic approximation $(PQ)_0$: $\min\{Df(x^0)(x - x^0) + \frac{1}{2}\langle x - x^0, H(x - x^0)\rangle \mid A(x - x^0) \le 0\}$, with $H = D^2_{xx}L(x^0, y^0)$, we have to apply Lemma 8.17: Suppose $g(x^0) = 0$, then condition (iii) equivalently means that

(iv) for all $y^0 \in Y^0$, with $\theta(y^0)$ defined according to Lemma 8.17, it holds that if (x, y) is critical for $(PQ)_0$ and $|y_i - y_i^0| \le \theta(y^0)\ \forall i \in I^+(y^0)$ then $x = x^0$.

◇

Proof. By Theorem 8.24, (i) holds if and only if for all $y^0 \in Y^0$ one has that each solution (u, α, β) of the C-stability system (8.21) satisfies $u = 0$. Then the Lemmas 8.15 and 8.17 establish correspondences between solutions (u, α, β) of (8.21) and the stationary points x of the auxiliary optimization problems in such a manner that $u = 0$ iff $x = 0$ for the problems (8.23), and $u = 0$ iff $x = x^0$ for $(PQ)_0$. Using Remark 8.14, so nothing remains to prove. □

To derive an alternative form of characterization (iii) in the preceding theorem, we consider the function

$$\phi(y) = \min\{\langle u, D^2_{xx}L(x^0, y)u\rangle \mid u \in U^0 \cap \operatorname{bd} B\}.$$

Here x^0 is again a given stationary solution of (P)=(P)(0). Note that the set Y^0 of multipliers, assigned to x^0, may be unbounded.

Lemma 8.28 (uniform sign of the Hessian form). *Let $(f, g) \in C^2$ and $x^0 \in X(0)$. Suppose that for each $y^0 \in Y^0$, the origin is an isolated stationary point of (8.40). Then ϕ has no zero in Y^0, and so, by continuity, $\operatorname{sign}\phi$ is constant on Y^0.* ◇

Proof. Indeed, if $\phi(y) = 0$ for some $y \in Y^0$, then the related minimizer u^0 of $u^\mathsf{T} D^2_{xx} L(x^0, y)u$ over $U^0 \cap \mathrm{bd}\, B$ fulfills

$$\langle u, D^2_{xx}L(x^0, y)u\rangle \geq \langle u^0, D^2_{xx}L(x^0, y)u^0\rangle \quad \forall u \in U^0.$$

So we obtain stationary points λu^0 ($\lambda > 0$) of (8.40) arbitrarily close to the origin, a contradiction. □

The previous two theorems particularly imply that under MFCQ at x^0, sign ϕ is necessarily constant on Y^0 if the stationary solution set mapping X is locally u.L. at $(0, x^0)$.

Linearly Constrained Programs

For *canonically perturbed* programs under affine-linear constraints, our characterizations of the local upper Lipschitz behavior of X become simpler and hold without assuming MFCQ [Kla00]. To obtain this, one essentially uses that the selection property of Remark 8.20 for the multiplier mapping Y is automatically satisfied in this case:

Lemma 8.29 (selection property, linear constraints). *Let A be an* (m, n)*-matrix, $b^0 \in \mathbb{R}^m$ and $f \in C^1(\mathbb{R}^n, \mathbb{R})$, consider the parametric program*

$$\min_x \{f(x) - \langle a, x\rangle \,|\, Ax - b^0 \leq b\}, \quad p = (a, b) \text{ varies,} \tag{8.41}$$

and let x^0 be a stationary solution of this program for $p = 0$. Then, for any sequence $\{(x^k, p^k)\} \subset \mathrm{gph}\, X$ with $(x^k, p^k) \to (x^0, 0)$, there exists a sequence of multipliers $y^k \in Y(x^k, p^k)$ such that $\{y^k\}$ has an accumulation point $y^0 \in Y^0 = Y(x^0, 0)$. ◇

Proof. Because of the Lipschitzian one–to–one correspondence between critical points and KKT points, we may assume that $Y(x, p)$ is the (standard) Lagrange multiplier set, i.e., for $(x, p) \in \mathrm{gph}\, X$, $p = (a, b)$,

$$Y(x, p) = \{y \in \mathbb{R}^m \,|\, Df(x) + A^\mathsf{T} y = a, \ y \geq 0, \ y^\mathsf{T}(Ax - b^0 - b) = 0\}.$$

For $I \subset \{1, \dots, m\}$ and $q \in \mathbb{R}^n$, let

$$H^I(q) \ := \ \{\eta \in \mathbb{R}^m \,|\, A^\mathsf{T}\eta = q, \ \eta_i \geq 0 \text{ if } i \in I, \ \eta_j = 0 \text{ if } j \notin I\},$$

and for $(x, b) \in \mathbb{R}^n \times \mathbb{R}^m$, we put

$$I(x, b) := \{i \in \{1, \dots, m\} \,|\, (Ax - b^0)_i = b_i\}.$$

Hence, for $(x, p) \in \mathrm{gph}\, X$, $p = (a, b)$, one has $Y(x, p) = H^{I(x,b)}(a - Df(x))$.

Now let $\{(x^k, p^k)\}$, $p^k = (a^k, b^k)$, be any sequence in gph X such that $(x^k, p^k) \to (x^0, 0)$. Then there are some $I \subset \{1, \dots, m\}$ and some infinite set $K \subset \{1, 2, \dots\}$ such that

$$I \equiv I(x^k, b^k) \quad \forall k \in K.$$

By Lemma 2.7 (Hoffman's lemma), the multifunction H^I is Lipschitz on its effective domain $\operatorname{dom} H^I$. Therefore,

$$\begin{array}{l}\Gamma^I(q) := \operatorname{argmin}\{\eta^{\mathsf T}\eta \,|\, \eta \in H^I(q)\} \\ \textbf{is single–valued and continuous on } \operatorname{dom} H^I.\end{array} \tag{8.42}$$

Since H^I is a closed multifunction, one has

$$H^I(q^0) = \emptyset \;\Rightarrow\; H^I(q) = \emptyset \text{ for all } q \text{ near } q^0.$$

Hence, $(x^k, a^k)\,(k \in K) \to (x^0, 0)$, the continuity of $Df(\cdot)$ and $H^I(a^k - Df(x^k)) \neq \emptyset$ $(\forall k \in K)$ imply that $H^I(-Df(x^0)) \neq \emptyset$. Now, (8.42) applied to the sequence $q^k := a^k - Df(x^k)\,(k \in K) \to -Df(x^0)$ yields the desired result. □

Theorem 8.30 (locally u.L. X, linear constraints). *For the stationary point mapping X of the parametric problem (8-41) and given $x^0 \in X(0)$, the following statements are equivalent:*

(i) X is locally upper Lipschitz at $(0, x^0)$.

(ii) There exists at least one *multiplier $y \in Y(x^0, 0)$ such that the system*

$$\begin{array}{l} C[Df](x^0)(u) \;+\; \alpha A \;\ni\; 0, \\ Au - \beta = 0, \;\; Bu = 0, \;\; (\alpha, \beta) \;\in\; \mathcal{J}_C(y) \end{array} \tag{8.43}$$

has no solution (u, α, β) with $u \neq 0$.

(iii) The origin $u = 0$ is the unique point which, for some $q \in C[Df](x^0)(u)$, solves $\min\{\langle q, x\rangle \,|\, x \in U^0\}$.

◇

Proof. By the note following Theorem 8.24, the characterization (i) ⇔ (iv) of that theorem also holds if MFCQ is replaced by the selection property for Y which is, by Lemma 8.29, automatically satisfied under linear constraints. Since $Q_C(y)(u) \equiv C[Df](x^0)(u)$ does not depend on the multiplier y, Theorem 8.27 yields the claimed result (note again that in the proof of Theorem 8.27 the assumption MFCQ can be replaced by the selection property on Y). □

The specializations to the C^2 case and to a second–order condition similar to Corollary 8.25 are now obvious, again MFCQ is not needed, and the phrase "for all multipliers y" may be replaced by "for at least one multiplier y". By the previous theorem, related results in [HG99] are covered.

8.3.3 Upper Regularity

In general, the local upper Lipschitz property of the stationary point mapping to a parametric program does not include *persistence* of the existence of stationary solutions. However, starting with a *strict local minimizer* of the initial problem, Theorem 1.16 on stability of complete local minimizing sets guarantees the existence of local minimizers of slightly perturbed problems under natural assumptions. In the present subsection, we derive sufficient conditions for the stationary or (locally) optimal solution set mapping to be *locally nonempty-valued and upper Lipschitz* (briefly, this is called *upper regularity*).

Again we consider the parametric programs $\mathrm{P}(t,p)$ and $(\mathrm{P})(p)$ (cf. §8.1), t near t^0, $p=(a,b)$ near (0,0), and its stationary solution multifunctions $\tilde{X}$ and X, respectively. We suppose that some stationary solution x^0 of the initial problem $(\mathrm{P})=\mathrm{P}(t^0,0)$ is given, and f, g_i are at least in $C^1(\Omega,\mathbb{R})$, where Ω is a neighborhood of $(x^0,t^0)\in\mathbb{R}^{n+r}$.

Lemma 8.31 (upper regularity implies MFCQ). *For the parametric C^1 program (P)(0, b), b varies near* 0, *let x^0 be a stationary solution of (P)=(P)(0,0). If $X(0,\cdot)$ is locally nonempty-valued and upper Lipschitz at $(0,x^0)$ then MFCQ holds at x^0.* ◇

Proof. By assumption, for some constant $L>0$ and each sequence $\varepsilon\downarrow 0$ there exist some sequence $x^\varepsilon\in X(0,b(\varepsilon))$, $b_i(\varepsilon)=-\varepsilon$ *if* $g_i(x^0)=0$, $b_i(\varepsilon)=0$ if $g_i(x^0)<0$, and ε sufficiently small, such that

$$\|x^\varepsilon-x^0\| \;\le\; L\varepsilon \text{ for small } \varepsilon>0.$$

Hence, one has that, with $J(x^0)=\{i|g_i(x^0)=0\}$,

$$-\frac{1}{L}\|x^\varepsilon-x^0\|\ge-\varepsilon>g_i(x^\varepsilon)=Dg_i(x^0)(x^\varepsilon-x^0)+o(\|x^\varepsilon-x^0\|),\ i\in J(x^0),$$

holds for small positive ε. Therefore, division by $\|x^\varepsilon-x^0\|$ and passing to the limit u of a suitable subsequence of $(x^\varepsilon-x^0)/\|x^\varepsilon-x^0\|$ yield

$$Dg_i(x^0)u\le-\frac{1}{L},\ i\in J(x^0),$$

which means that MFCQ is fulfilled at x^0. □

Upper Regularity of Isolated Minimizers

Denote by $\Psi_{\mathrm{loc}}(t,p)$ the set of all local minimizers of $\mathrm{P}(t,p)$ for fixed (t,p), $p=(a,b)$, and put

$$\begin{array}{rcl}\mathcal{M}(t,b) & := & \{x\in\mathbb{R}^n|g(x,t)\le b\}\\ \Psi_Q(t,p) & := & \operatorname{argmin}_x\{f(x,t)-\langle a,x\rangle\,|\,x\in\mathcal{M}(t,b)\cap Q\},\end{array} \tag{8.44}$$

where Q is a given nonempty subset of $\mathbb{R}^n$. In the following lemma, we recall conditions (cf. [Kla86], and [Rob82] for the C^2 case) which ensure that

$$\emptyset\ne\Psi_{\mathrm{loc}}(t,p)\cap Q\subset\tilde{X}(t,p)\cap Q$$

is fulfilled for all (t,p) near $(t^0,0)$ and for some neighborhood Q of an initial minimizer x^0. In the C^2 case, Robinson [Rob82, Thm. 3.2] has essentially used the fact that a strict local minimizer of order 2 under MFCQ is automatically an isolated stationary solution. Example 8.23 shows that this is no longer true for $C^{1,1}$ programs, we have to make an extra assumption.

Recall that x is said to be an *isolated stationary solution* of $\mathrm{P}(t^0)$ if $\{x\} = \tilde{X}(t^0) \cap \mathcal{N}$ holds for some neighborhood $\mathcal{N}$ of x.

Lemma 8.32 (u.s.c. of stationary and optimal solutions). *For the parametric C^1 program P(t,p), let x^0 be a stationary solution of (P)=P$(t^0,0)$. Suppose that MFCQ holds at x^0 with respect to $\mathcal{M}(t^0,0)$. Then one has:*

(i) $\tilde{X}(\cdot) \cap \mathcal{N}$ is u.s.c. at $(t^0,0)$ for some neighborhood $\mathcal{N}$ of x^0.

(ii) If, in addition, x^0 is both a local minimizer and an isolated stationary solution to (P), then $\emptyset \neq \Psi_{loc}(t,p) \cap \mathrm{cl}\,\mathcal{Q} \subset \tilde{X}(t,p) \cap \mathrm{cl}\,\mathcal{Q}$ $(\forall (t,p) \in \mathcal{O})$ holds for some neighborhoods $\mathcal{O}$ of $(t^0,0)$ and $\mathcal{Q}$ of x^0. Further, $\Psi_{loc}(\cdot) \cap \mathrm{cl}\,\mathcal{Q}$ and $\tilde{X}(\cdot) \cap \mathrm{cl}\,\mathcal{Q}$ are u.s.c. at $(t^0,0)$.

◇

Proof. By MFCQ, we have from Corollary 2.9 and from persistence of MFCQ under small perturbations that there are neighborhoods $\mathcal{N}$ of x^0 ($\mathcal{N}$ may be assumed to be compact), $\mathcal{U}$ of t^0 and $\mathcal{W}$ of 0 such that $\tilde{Y}$ is u.s.c on $\mathcal{N} \times \mathcal{U} \times \mathcal{W}$. Hence, in particular, there are a compact set Z and some $\varepsilon > 0$ such that

$$\tilde{Y}(x,t,p) \subset Z \quad \forall (x,t,p) \in (\mathcal{N} \times \mathcal{U} \times \mathcal{W}) + \varepsilon B.$$

It suffices to show that the multifunction $\tilde{X}(\cdot) \cap \mathcal{N}$ is closed at $(t^0,0)$. Taking any sequence $\{(x^k,t^k,p^k)\}$ satisfying $x^k \in \tilde{X}(t^k,p^k) \cap \mathcal{N}$ (with related $y^k \in \tilde{Y}(x^k,t^k,p^k)$) and $(x^k,t^k,p^k) \to (x^*,t^0,0)$, we have $x^* \in \mathcal{N}$ and the existence of some accumulation point $y^* \in Z$ of $\{y^k\}$. Hence, $y^* \in \tilde{Y}(x^*,t^0,0)$ because $\tilde{Y}$ is closed at $(x^*,t^0,0)$. This yields $x^* \in \tilde{X}(t^0,0) \cap \mathcal{N}$, and so, (i) is shown.

To show (ii), we first use that x^0 is an isolated point in $\tilde{X}(t^0,0)$. Because of the MFCQ, then x^0 is also an isolated local minimizer. Thus, with $\mathcal{N}$ from (i), there is a neighborhood $V \subset \mathcal{N}$ satisfying

$$\{x^0\} = \tilde{X}(t^0,0) \cap \mathrm{cl}\,V = \Psi_{\mathrm{loc}}(t^0,0) \cap \mathrm{cl}\,V = \Psi_{\mathrm{cl}\,V}(t^0,0). \tag{8.45}$$

Since MFCQ persists under small perturbations, we may assume with no loss of generality that MFCQ holds at each point in $\mathcal{M}(t,b) \cap \mathrm{cl}\,V$, (t,b) near $(t^0,0)$, and so $\Psi_{\mathrm{loc}}(t,p) \cap \mathrm{cl}\,V$ is a subset of $\tilde{X}(t,p) \cap \mathrm{cl}\,V$ for all (t,p) near $(t^0,0)$. By Theorem 1.16, there exist a neighborhood $\mathcal{O}$ of $(t^0,0)$ and a neighborhood $\mathcal{Q} \subset V$ such that $\emptyset \neq \Psi_{\mathrm{cl}\,\mathcal{Q}}(t,p) \subset \Psi_{\mathrm{loc}}(t,p)$ is true for $(t,p) \in \mathcal{O}$. Hence, from (i) and (8.45), we obtain that $\Psi_{\mathrm{loc}}(\cdot) \cap \mathrm{cl}\,\mathcal{Q}$ and $\tilde{X}(\cdot) \cap \mathrm{cl}\,\mathcal{Q}$ are also u.s.c. at $(t^0,0)$.

Let again

$$Y^0 = \tilde{Y}(x^0,t^0,0).$$

Note that, in particular, the injectivity condition at $(x^0, Y^0) \subset \widetilde{S}(t^0, 0)$ for CF with respect to u implies that x^0 is an *isolated* stationary solution of (P). Now, Theorem 8.24 and Lemma 8.32 immediately yield the following result.

Theorem 8.33 (upper regular minimizers, $C^{1,1}$). *Consider the parametric $C^{1,1}$ program $P(t,p)$, and let $x^0 \in \widetilde{X}(t^0, 0)$ be a local minimizer of $(P)=P(t^0,0)$. Then $\widetilde{X}$ is locally nonempty–valued and u.L. at $((t^0,0),x^0)$ if and only if both MFCQ holds at x^0 with respect to $\mathcal{M}(t^0,0)$ and (x^0, Y^0) satisfies the injectivity condition for $CF(\cdot,\cdot,t^0)$ with respect to u. Further, if $\widetilde{X}$ is locally nonempty–valued and u.L. at $((t^0,0),x^0)$, then Ψ_{loc} has this property, too.* ◇

Remark 8.34 (isolated minimizing sets). In the case of replacing x^0 by an isolated compact set X^0 of local minimizers of (P), characterizations of upper regularity in the sense of Theorem 8.33 are still not known.

However, under MFCQ on X^0 and under a growth condition of order $\varrho \geq 1$ imposed on $f(\cdot, t^0)$ with respect to an open bounded set Q containing X^0, local upper Hölder continuity of order ϱ^{-1} for $\Psi_{\mathrm{cl}\,Q}$ has been shown in the literature; details may be found in Klatte [Kla94a] for a general setting including Lipschitzian programs, for $\varrho = 1, 2$ in Bonnans and Shapiro [BS00, Prop. 4.41] concerning C^2 optimization problems and Ioffe [Iof94] concerning Lipschitzian programs with fixed constraints. For the case $X^0 = \{x^0\}$, results of that type are well-known already from the 80ies, see, e.g., [Alt83, Don83, Aus84, Kla85, Gfr87].

Local upper Lipschitz continuity of $\Psi_{\mathrm{cl}\,Q}$ holds under LICQ (on X^0) already if quadratic growth of $f(\cdot, t^0)$ with respect to $Q \supset X^0$ is assumed; this was shown for C^2 programs first in [Sha88a], and for $C^{1,1}$ programs in [Kla94a] and [BS00, Thm. 4.81]. Related results under a different set of assumptions can be found in [Iof94]. In the case $X^0 = \{x^0\}$, the mentioned condition reduces to that of Robinson [Rob82]. ◇

Remark 8.35 (some consequences of Theorem 8.33). Recall that, by Corollary 8.25, the second order condition SOCL implies that (x^0, Y^0) satisfies the injectivity condition for CF with respect to u. Hence, by Theorem 8.33, SOCL on (x^0, Y^0) and MFCQ at x^0 together ensure that $\widetilde{X}$ and Ψ_{loc} are locally nonempty–valued and u.L. at $((t^0,0),x^0)$, provided that x^0 is a local minimizer of (P). Below we shall derive second–order optimality conditions for $C^{1,1}$ programs which guarantee that a stationary solution x^0 of (P) is a (strict) local minimizer and also satisfies SOCL at some $(x^0, y^0) \in (x^0, Y^0)$.

If x^0 is a global minimizer of (P), then one may replace in statement (ii) of Theorem 8.33 the local minimizing set mapping Ψ_{loc} by the global optimizing set mapping Ψ, provided that Ψ is locally bounded near that point. ◇

We finish this paragraph by a complete characterization of upper regularity of the stationary solution set mapping for parametric C^2 programs, provided that x^0 is a local minimizer of (P). Because of Lemma 8.31, MFCQ may be supposed without restriction of generality. Note that the proof of the inclusion (i) $\Rightarrow$ (iii) in the following theorem essentially uses an idea proposed by Gfrerer [Gfr00].

Theorem 8.36 (upper regular minimizers, C^2). *Let x^0 be a local minimizer of (P)=P(t^0,0), and suppose that f, g_i belong to $C^2(\Omega, \mathbb{R})$, Ω a neighborhood of (x^0, t^0). If x^0 satisfies MFCQ, then the following properties are equivalent to each other and imply that Ψ_{loc} is locally nonempty-valued and u.L. at $((t^0,0),x^0)$:*

(i) $\tilde{X}$ is locally nonempty-valued and u.L. at $((t^0,0),x^0)$.

(ii) $\tilde{X}$ is locally u.L. at $((t^0,0),x^0)$.

(iii) x^0 satisfies the quadratic growth property

$$\min\{\langle u, D^2_{xx}L(x^0,y)u\rangle \mid y \in Y^0, u \in U^0 \cap \operatorname{bd} B\} > 0.$$

(iv) For each $y \in Y^0$, SOCL is satisfied on (x^0, Y^0), i.e.,

$$\langle u, D^2_{xx}L(x^0,y)u\rangle \neq 0 \quad \forall y \in Y^0 \, \forall u \in U^0 \cap \operatorname{bd} B$$

holds true.

◇

Proof. Without loss of generality, let $g(x^0) = 0$. Then, in particular, KKT points and critical points in Kojima's form coincide, and L is the usual Lagrange function. Note that x^0 is a stationary solution of (P) due to MFCQ.

The equivalence (i) ⇔ (ii) follows from Theorem 8.33 together with Theorem 8.24. The inclusion (iii) ⇒ (iv) is trivial, while (iv) ⇒ (ii) under MFCQ follows from Corollary 8.25. Put

$$Q_y := D^2_{xx}L(x^0,y), \; y \in Y^0 := \tilde{Y}(x^0,t^0,0).$$

If (iii) does not hold, then for some $\tilde{u} \in U^0 \cap \operatorname{bd} B$ and some $y \in Y^0$, one has $\langle \tilde{u}, Q_y \tilde{u}\rangle \leq 0$. Then, by a known second-order necessary optimality condition, there is some $\eta \in Y^0$ with $\langle \tilde{u}, Q_\eta \tilde{u}\rangle \geq 0$, hence, for some $\tilde{y} \in \operatorname{conv}\{y, \eta\} \subset Y^0$, it follows $\langle \tilde{u}, Q_{\tilde{y}} \tilde{u}\rangle = 0$. Therefore (iv) is not true, and we have shown that (iv) ⇒ (iii).

It remains to prove that (ii) ⇒ (iii) is true. Note that for fixed u, the function $y \mapsto \langle u, Q_y u\rangle$ attains its maximum on Y^0 in vertices of the bounded convex polyhedron Y^0. Let $Y^0 = \operatorname{conv}\{y^1, \ldots, y^d\}$. Putting $Q_j := Q_{y^j}$, we define the continuous functions

$$\begin{aligned}
\mu(u) &:= \max_{1 \leq j \leq d}\langle u, Q_j u\rangle, \; u \in U^0, \\
\nu(u) &:= \langle u, Q_1 u\rangle, \; u \in U^0, \\
\varphi(\lambda) &:= \min\{\lambda\mu(u) + (1-\lambda)\nu(u) | u \in U^0 \cap \operatorname{bd} B\}, \; \lambda \in [0,1].
\end{aligned}$$

If (iii) does not hold, then $\varphi(0) = \min\{\nu(u) | u \in U^0 \cap \operatorname{bd} B\} \leq 0$ since otherwise already Lemma 8.28 yields the assertion. Further, the second-order necessary optimality condition used above gives $\varphi(1) = \min\{\mu(u) | u \in U^0 \cap \operatorname{bd} B\} \geq 0$. By continuity of φ, it follows $\varphi(\lambda) = 0$ for some fixed $\lambda \in [0,1]$. Since $\mu(\theta u) = \theta^2\mu(u)$ and $\nu(\theta u) = \theta^2\nu(u)$, this means

$$0 = \lambda\mu(u^0) + (1-\lambda)\nu(u^0) = \varphi(\lambda) = \min\{\lambda\mu(u) + (1-\lambda)\nu(u) | u \in U^0\}$$

for some $u^0 \in U^0 \cap \mathrm{bd}\, B$. Hence, for all directions $u - u^0$, $u \in U^0$, the directional derivatives of $\lambda\mu(\cdot) + (1-\lambda)\nu(\cdot)$ at u^0 are non-negative. Setting $J = \{j \in \{1,\dots,d\} | \langle u^0, Q_j u^0\rangle = \mu(u^0)\}$ the latter means

$$\begin{aligned} 0 &\le \lambda \max\nolimits_{j\in J}\langle u - u^0, Q_j u^0\rangle + (1-\lambda)\langle u - u^0, Q_1 u^0\rangle \\ &= \max\nolimits_{j\in J}\langle u - u^0, (\lambda Q_j + (1-\lambda)Q_1)u^0\rangle \\ &\le \max\nolimits_{y\in Y^0}\langle u - u^0, D^2_{xx}L(x^0, y)u^0\rangle. \end{aligned}$$

So we obtain, for some compact, convex neighborhood W of u^0

$$\min_{u\in U^0\cap W}\max_{y\in Y^0}\langle u - u^0, D^2_{xx}L(x^0,y)u^0\rangle \ge 0$$

and the minimax theorem ensures the existence of a multiplier $y^0 \in Y^0$ such that

$$\langle u - u^0, D^2_{xx}L(x^0, y^0)u^0\rangle \ge 0 \quad \forall u \in U^0 \cap W.$$

Thus, $u^0 \in U^0 \cap \mathrm{bd}\, B$ is a stationary solution of the quadratic auxiliary program

$$\min\{\langle x, D^2_{xx}L(x^0, y^0)x\rangle \,|\, x \in U^0\}.$$

By Theorem 8.27, this contradicts (ii), and so, (ii) $\Rightarrow$ (iii) is shown. This completes the proof. □

Corollary 8.87 (necessary condition for strong regularity, C^2 case). *Let the assumptions of Theorem 8.36 be satisfied. If $\tilde{X}$ is locally single-valued and Lipschitz near $((t^0, 0), x^0)$, then x^0 satisfies the strong growth property*

$$\min\{\langle u, D^2_{xx}L(x^0, y)u\rangle \,|\, y \in Y^0,\ u \in U^+(y) \cap \mathrm{bd}\, B\} > 0.$$

◇

Proof. In particular, $\tilde{X}$ is locally u.L. at each $((t,p),x) \in \mathrm{gph}\,\tilde{X}$ in some neighborhood of $((t^0,0),x^0)$. Considering for each $y \in Y^0$ the particular right-hand side perturbation $b(\varepsilon)$ with $b_i = \varepsilon > 0$ for $i \in I^0(y)$ and $b_j \equiv 0$ for $j \notin I^0(y)$, we then have the result immediately from Theorem 8.36. □

The opposite direction (under MFCQ) is not true, cf. the counterexample given by Robinson [Rob82]. From the proof of Theorem 8.36 and Corollary 8.37 one sees that the strong growth property already follows if only perturbations (a, b) appear.

Second–Order Conditions for $C^{1,1}$ Programs

Second-order optimality conditions for $C^{1,1}$ programs were given in terms of Clarke's generalized Hessian [HUSN84, KT88] and second-order (tangential) directional derivatives [War94] of the Lagrangian. Here we present necessary and sufficient optimality conditions in terms of the contingent derivative of

D_xL (cf. [Kla00]). In particular, this is of interest with respect to the regularity assumptions discussed in Remark 8.35. Throughout we consider a given stationary solution x^0 of the unperturbed $C^{1,1}$ problem (P) $=$ P(t^0), and we again write $f = f(\cdot, t^0)$, $g = g(\cdot, t^0)$, and so on. Recall that the set $Q_C(y)(u) = C_x[D_xL](x^0, y)(u)$ is connected and compact. We denote by $\mathcal{M}$ the constraint set of (P), and the abbreviation SMFCQ means the *strict MFCQ.*

Theorem 8.38 (local minimizer and quadratic growth, $C^{1,1}$ case). *Consider the $C^{1,1}$ program (P). Suppose that (x^0, y) is a critical point of (P).*

(i) If $\min_q\{\langle u, q\rangle : q \in Q_C(y)(u)\} > c$ holds for some $c > 0$ and for each $u \in U^0$, $\|u\| = 1$, then there exists a neighborhood $\mathcal{Q}$ of x^0 such that for all $x \in \mathcal{M} \cap \mathcal{Q}$, the quadratic growth condition $f(x) - f(x^0) \geq \frac{1}{2} c\|x - x^0\|^2$ *is fulfilled.*

(ii) If x^0 is a local minimizer of (P), and SMFCQ is satisfied at x^0, then there holds $\max_q\{\langle u, q\rangle : q \in Q_C(y)(u)\} \geq 0$ for every $u \in U^0$. ◇

Proof. Assertion (i) was proved in Theorem 6.23.

It remains to prove (ii). Since y is a fixed multiplier vector, we write $L(\cdot) := L(\cdot, y)$. All constructions in the proof can be restricted to points in some neighborhood $\mathcal{N}$ of x^0. Assume $\mathcal{N}$ being small enough such that DL is Lipschitzian on $\mathcal{N}$ with constant ϱ_{DL}.

By assumption, SMFCQ holds at x^0, hence statement (i) of Corollary 8.16 and Gordan's theorem of the alternative (see (A.8) or, e.g., [Man81b, Man94]) imply that MFCQ holds at x^0 even with respect to the constraint set

$$\widetilde{\mathcal{M}} := \{x | g(x) \leq 0,\ g_i(x) = 0,\ i \in I^+(y)\}.$$

Let $u \in U^0$ be any vector with $u \neq 0$, with no loss of generality suppose that $\|u\| = 1$. Note that U^0 is the linearization cone of $\widetilde{\mathcal{M}}$. Hence, by the classical theory of constraint qualifications, there exists a sequence

$$\{x^k\} \subset \widetilde{\mathcal{M}} \setminus \{x^0\} : \quad x^k \to x^0 \quad \text{and} \quad u^k := \|x^k - x^0\|^{-1}(x^k - x^0) \to u.$$

By the assumption of (ii), x^0 is a local minimizer of (P). Note that $\widetilde{\mathcal{M}} \subset M$. Thus, for sufficently large k, there holds with some $\varepsilon \geq 0$ that

$$f(x^k) - f(x^0) \geq \varepsilon\|x^k - x^0\|^2. \tag{8.46}$$

Since $\{x^k\} \subset \widetilde{\mathcal{M}}$, we have $L(x^k) = f(x^k)$ for all k. Hence, there exist $0 < \tau_k < 1$ such that with $\theta_k := \|x^k - x^0\|$,

$$f(x^k) - f(x^0) = L(x^k) - L(x^0) = DL(x^0 + \tau_k\theta_k u^k)(x^k - x^0), \tag{8.47}$$

by the mean value theorem. Using that $DL \in C^{0,1}$ and $DL(x^0) = 0$, we find that some subsequence of $(\tau_k\theta_k)^{-1}DL(x^0 + \tau_k\theta_k u^k)$ converges to a limit q which must belong to $Q_C(y)(u)$. Then, after dividing (8.47) by $\tau_k\theta_k\|x^k - x^0\|$ and passing to the limit for the corresponding subsequence, we see that (8.46) and (8.47) imply $u^{\mathsf{T}}q \geq \varepsilon$. This completes the proof of (ii). □

Note. A local minimizer x^0 which satisfies the *quadratic growth condition* of Theorem 8.38 (i) is called a *strict local minimizer of order 2 to* (P). If x^0 is a strict local minimizer of order 2 to (P) and SMFCQ holds at x^0, then

$$\max_q\{\langle u,q\rangle : q \in Q_C(y)(u)\} > 0 \quad \text{for every } u \in U^0 \setminus \{0\},$$

this is easy to see from part (ii) of the above proof. ◇

We learn from Example 8.23 that the condition in (i) is not necessary for the quadratic growth property at x^0. However, statement (ii) of Theorem 8.38 can be considered as a compromise. Moreover, we mention that for a C^2 program, statement (i) of this theorem carries over into [Rob82, Thm. 2.2].

8.3.4 Strongly Regular and Pseudo-Lipschitz Stationary Points

In this subsection, we are interested in characterizations of strong regularity and of the pseudo-Lipschitz property of stationary solutions if LICQ fails to hold.

Strong Regularity

A necessary condition for strong regularity when supposing that x^0 is a local minimizer of a C^2 program was given in Corollary 8.37. To find even a *characterization,* we know from Lemma 3.1 that a suitable representation of the Thibault derivative of the stationary solution set mapping might be helpful. Indeed, Lemma 3.1 and Exercise 5 imply that the stationary point map $X = X(a,b)$ of the parametric $C^{1,1}$ program (P)(a,b), $(a,b) \in \mathbb{R}^{n+m}$ is a locally Lipschitz function near $(y^0,x^0) \in \operatorname{gph} X$ if and only if both the Thibault derivative satisfies $\{0\} = TX(y^0,x^0)(0)$ and X is a l.s.c. multifunction.

This gives rise to look for a suitable representation of TX in terms of the problem data for (P)(a,b). It will turn out that under MFCQ, this is a more difficult task than the description of the contingent derivative CX. We were only able to give a limit representation of TX. It is an open question whether there exists an explicit form of description, or not.

For seek of simplicity, we again assume that $g(x^0) = 0$ holds for the given stationary solution x^0 of (P)(0). Hence, under MFCQ at x^0, the multiplier set $Y^0 = \{y|F(x^0,y) = 0\}$ is a bounded subset of $\mathbb{R}^m_+$.

Theorem 8.39 *(TX under MFCQ). Suppose that x^0 is a stationary point of (P)(0), x^0 satisfying MFCQ and $g(x^0) = 0$. Then there holds $u \in TX(0,x^0)(a,b)$ if and only if there exist $\theta = \theta_k \downarrow 0$ and related points $x = x^k \to x^0$, $y = y^k \to y^0 \in Y^0$, $\tilde{y} = \tilde{y}^k \to y^1 \in Y^0$ such that*

(i) $\theta^{-1}(M(x+\theta u) - M(x)) \to M_0 \in TM(x^0)(u)$ and

(ii) with $\tilde{\delta} = \tilde{y} - y^1$, $\delta = y - y^0$, the vectors

$$h_k = \theta^{-1}M(x)[N(y^1) - N(y^0) + N'(y^1;\tilde{\delta}) - N'(y^0;\delta)]$$

fulfill $M_0N(y^1) + h_k \to (a,b)$.
In terms of original data, the latter means that $\alpha = \widetilde{y}^+ - y^+$ *and* $\beta = \widetilde{y}^- - y^-$ *satisfy*

$$\begin{aligned} q_\theta := \theta^{-1}(L'(x+\theta u) - L'(x)) &\to q^0 \in TL'(x^0)(u), \\ q^0 + \theta^{-1}Dg(x)^\mathsf{T}\alpha &\to a, \\ Dg(x^0)u - \theta^{-1}\beta &\to b, \end{aligned}$$

where $L'(\cdot) := DL(\cdot, y^1)$. ◇

Proof. ($\Rightarrow$) Let $u \in TX((0,0),x^0)(a,b)$. Then, for some sequence $\theta = \theta_k \downarrow 0$, assigned sequences of parameters $p = p(\theta) \to (0,0)$, $x = x(\theta) \to x^0$ and appropriate o-type functions $o_x(\theta), o_p(\theta)$, we have

$$x = x(\theta) \in X(p(\theta)) \quad \text{and} \quad x' + \theta u + o_x(\theta) \in X(p(\theta) + \theta(a,b) + o_p(\theta)).$$

Due to MFCQ, Y^0 is compact, so there exist dual variables $y = y(\theta)$, $\tilde{y} = \tilde{y}(\theta) = y + v(\theta)$, assigned to x and $\widetilde{x}$, respectively, such that

$$F(\widetilde{x},\widetilde{y}) - F(x,y) = \theta(a,b) + o_p(\theta)$$

and, for some subsequence, the points y and $\widetilde{y}$ converge: $y \to y^0 \in Y^0$, $\widetilde{y} \to y^1 \in Y^0$. Since $\|\delta\|$ and $\|\widetilde{\delta}\|$ vanish, $N(y)$ and $N(\widetilde{y})$ can be written by the help of N',

$$N(y) = N(y^0) + N'(y^0;\delta), \;\; N(\widetilde{y}) = N(y^1) + N'(y^1;\widetilde{\delta}).$$

This yields

$$\begin{aligned} &\theta(a,b) \\ &= F(\widetilde{x},\widetilde{y}) - F(x,y) \\ &= M(\widetilde{x})[N(y^1) + N'(y^1;\widetilde{\delta})] - M(x)[N(y^0) + N'(y^0;\delta)] \\ &= [M(\widetilde{x}) - M(x)]N(y^1) + M(x)[N(y^1) + N'(y^1;\widetilde{\delta}) - N(y^0) - N'(y^0;\delta)] \\ &\quad + [M(\widetilde{x}) - M(x)]N'(y^1;\widetilde{\delta}). \end{aligned}$$

After division by θ (and after selecting an appropriate subsequence), the bounded matrices $\theta^{-1}[M(\widetilde{x}) - M(x)]$ converge to some element $M_0 \in TM(x^0)(u)$ while the term $\theta^{-1}[M(\widetilde{x}) - M(x)]N'(y^1;\widetilde{\delta})$ vanishes. Thus, the limit of

$$h_k = \theta^{-1}M(x)[N(y^1) - N(y^0) + N'(y^1;\widetilde{\delta}) - N'(y^0;\delta)]$$

exists and satisfies $h_k \to (a,b) - M_0N(y^1)$. So we derived ($\Rightarrow$).

($\Leftarrow$) Conversely, assume that such sequences have been found. Setting $\widetilde{x} = x + \theta u$, $p = F(x,y)$ and $\widetilde{p} = F(\widetilde{x},\widetilde{y})$, one easily determines (by the above calculation) that

$$F(\widetilde{x},\widetilde{y}) - F(x,y) = M(\widetilde{x})N(\widetilde{y}) - M(x)N(y) = \theta(a,b).$$

So, we know that

$$x \in X(p),\ \tilde{x} \in X(\tilde{p}),\ \tilde{x} - x = \theta u, \text{ and } \theta^{-1}(\tilde{p} - p) \to (a,b).$$

Using also

$$x \to x^0,\ y \to y^0,\ \tilde{y} \to y^1, \text{ and } F(x,y^0) = F(x,y^1) = 0,$$

we obtain $\tilde{p}, p \to 0$. This yields $u \in TX(0,x^0)(a,b)$ by definition of *TX*.

The second description follows by direct calculation from the first one. This completes the proof. □

Remark 8.40 If $f,g \in C^2$ then we obtain in Theorem 8.39 that

$$\begin{array}{lcl} D^2_{xx}L(x^0,y^1)u + \theta^{-1}Dg(x)^{\mathsf{T}}\alpha & \to & a \\ Dg(x^0)u - \theta^{-1}\beta & \to & b. \end{array}$$

Notice that $\theta^{-1}Dg(x)^{\mathsf{T}}\alpha$ and $-\theta^{-1}\beta$ are assigned to the part $M(x^0)TN(y^0)(v)$, $v = \tilde{y} - y$, in the explicit formula for the derivative *TF*, see Theorem 7.5. However, now $\|\tilde{y} - y\|$ is only bounded and, even more important, both x^0 and x (near x^0) appear in the formula. For this reason, the description of $TX(0,x^0)(a,b)$ in terms of first and second derivatives of f,g at x^0 makes difficulties even for arbitrarily smooth functions. We are not sure that such a description exists at all. For linear problems (P), the condition attains the form

$$(a,b) \in M(x^0)N'(y,\mathbb{R}^m) \text{ for some } y \in Y^0.$$

◇

On the other hand, the derivative *TF* presents a necessary condition for *X* being locally unique and Lipschitz.

Lemma 8.41 (*TF*-injectivity w.r. to u). *Let X be locally single–valued and Lipschitz at* $(0,x^0) \in \operatorname{gph} X$. *Then,* (x^0,Y^0) *satisfies TF-injectivity w.r. to u.* ◇

Proof. Otherwise, it holds $0 \in TF(x^0,y^0)(u,v)$ with some $y^0 \in Y^0$ and $u \neq 0$. Then, by definition of *TF*, there are sequences $t = t_k \downarrow 0$, related $(x_t,y_t) \to (x^0,y^0), (u_t,v_t) \to (u,v)$ and $w_t \to 0$ such that

$$tw_t = F(x_t + tu_t, y_t + tv_t) - F(x_t,y_t).$$

Setting

$$p^1_t = F(x_t,y_t) \text{ and } p^2_t = F(x_t + tu_t, y_t + tv_t),$$

one sees that $x_t \in X(p^1_t)$, $x_t + tu_t \in X(p^2_t)$, and $\|p^2_t - p^1_t\| \,/\, \|(x_t + tu_t) - x_t\| = \|w_t\|/\|u_t\|$ vanishes. So *X* cannot be locally single–valued and Lipschitz near $(0,x^0)$. □

Pseudo-Lipschitz Property

As before, we investigate the map $X = X(a,b)$ at $(0,x^0)$ under MFCQ at x^0. Due to MFCQ, X is closed (near $(0, x^0)$), so we know from Theorem 5.3 that X is not pseudo-Lipschitz at $(0, x^0)$ if and only if $\Gamma = X^{-1}$ satisfies the ("pseudo-singularity") condition of Theorem 5.2. In the current case, this means

$$\begin{aligned}&\exists\ (u^*,v^*) \in \mathbb{R}^{n+m}\ (u^*,v^*) \neq 0 \text{ and } (p^k,x^k) \in \operatorname{gph} X\\ &\text{such that } (p^k,x^k) \to (0,x^0) \text{ as } k \to \infty \text{ and}\\ &\sup_{(a,b)\,\in\, CX^{-1}(x^k,p^k)(B)} \langle (u^*,v^*),(a,b)\rangle\ <\ \varepsilon_k \downarrow 0.\end{aligned} \tag{8.48}$$

We close this subsection with a direct relation between this condition and the derivative TF.

Theorem 8.42 (TF and pseudo-regularity of X). ***Let** $x^0 \in X(0)$ fulfill MFCQ.*

(i) If the constraints g_i are (affine) linear then (8.48) yields $u^* \neq 0$.

(ii) Let (8.48) be satisfied and $f, g \in C^2$. Then the point (u^*, v^*, Au^*) is a solution of the T-stability system (8.13) for some $y^0 \in Y^0$ and $(a,b) = 0$.

(iii) If the constraints g_i are (affine) linear, $f \in C^2$ and (x^0, Y^0) satisfies TF-injectivity w.r. to u, then X is pseudo Lipschitz at $(0, x^0)$.

◇

Proof. From Theorem 8.19, we have $CX^{-1}(0,x^0)(u) = CF(x^0,Y^0)(u,\mathbb{R}^m)$. So the set $\langle (u^*,v^*), CX^{-1}(0,x^0)(B(0,1))\rangle$ coincides, by Theorem 7.6, with the set of all sums

$$\begin{aligned}S &= \langle u^*, q(u)\ +\ A^{\mathsf T}\alpha\rangle + \langle v^*, Au - \beta\rangle\\ &= \langle u^*, q(u)\rangle\ + \langle v^*, Au\rangle + \textstyle\sum\ (\alpha_i A_i u^* - v_i^*\beta_i)\end{aligned} \tag{8.49}$$

such that $q(u) \in Q_C(u)$, $(\alpha,\beta) \in \mathcal{J}_C(y^0)$, $y^0 \in Y^0$ and u is restricted to the unit ball.

Now we apply these equivalences to the points (p^k,x^k) in (8.48) and take into account that the index sets I^+, I^0, I^- and the allowed variations of $(\alpha,\beta) \in \mathcal{J}_C(y)$ in (7.32) depend on the selection of $y = y^k \in Y^k = \{y|F(x^k,y) = p^k\}$. Similarly, $A = A(x^k)$ and $Q_C(u) = C_x[D_xL](x^k,y^k)(u)$ depend on x^k and y^k; we will write $Q_C(u) = Q(x^k,y^k)(u)$. Next observe that (8.48) and (8.49) imply

$$\sum\ (\alpha_i A_i u^* - v_i^*\beta_i) \le 0, \quad (A = A(x^k))$$

for all (α,β) in the polyhedral cone $\mathcal{J}_C(y)$ and all $y \in Y^k$.

Due to the possible variations of (α_i,β_i) in $(\mathbb{R},0)$ (if $y_i > 0$), $(0,\mathbb{R})$ (if $y_i < 0$) and in $(0,\mathbb{R}^-) \cup (\mathbb{R}^+,0)$ (if $y_i = 0$), this restricts (u^*,v^*) (with $A = A(x^k)$) by

$$\begin{aligned}&A_i u^* = 0 \ \text{ if } y_i > 0 \text{ for some } y \in Y^k\\ &v_i^* \quad\ = 0 \ \text{ if } y_i < 0 \text{ for some } y \in Y^k\\ &A_i u^* \le 0 \text{ and } v_i^* \le 0\ \text{ if } y_i = 0 \text{ for some } y \in Y^k.\end{aligned} \tag{8.50}$$

The conditions (8.50) define a polyhedral cone $K^*(x^k, Y^k) \subset \mathbb{R}^{n+m}$ and ensure

$$\max\{\alpha_i A_i u^* - v_i^* \beta_i \mid (\alpha, \beta) \in \mathcal{J}_C(y^k)\} = 0.$$

So, the crucial inequality of (8.48) becomes

$$\sup_{u \in B(0,1),\ q(u) \in Q(x^k,y)(u),\ y \in Y^k} \langle u^*, q(u) \rangle + \langle v^*, Dg(x^k)u \rangle < \varepsilon_k \tag{8.51}$$

along with (8.50), i.e., $(u^*, v^*) \in K^*(x^k, Y^k)$. Selecting an appropriate subsequence, the three index sets in (8.50) are constant for all k, we denote them by J^+, J^-, J^0. Then $(u^*, v^*) \in K^*(x^k, Y^k)$ if and only if

$$\begin{array}{l} A(x^k)_i u^* = 0 \ (i \in J^+) \ \ v_i^* = 0 \ (i \in J^-) \\ A(x^k)_i u^* \leq 0 \ \text{ and } v_i^* \leq 0 \ (i \in J^0) \end{array} \tag{8.52}$$

For the non-empty set $Y^{oS} = \limsup Y^k \subset Y^0$, one so derives

$$\sup_{u \in B(0,1),\ q(u) \in Q(x^0,y)(u),\ y \in Y^{oS}} \langle u^*, q(u) \rangle + \langle v^*, A(x^0)u \rangle \leq 0. \tag{8.53}$$

We consider now the particular cases mentioned in the theorem.

(i) If $u^* = 0$ and all g_i are affine-linear, then one can show that$v^* = 0$. To do this, note that now $v^* A = 0$ and the matrices A do not depend on x. Next we verify that, for $i \in J^{++} := J^+ \setminus J^0$, the components v_i^* must vanish. Indeed, the relation $i \in J^{++}$ means that $y^i > 0 \ \forall y \in Y^k$. Having $v_i^* > 0$ for some $i \in J^{++}$, then one finds, with appropriate $\lambda > 0$, that $y(\lambda) := y - \lambda v^*$ belongs to Y^k and fulfills $y(\lambda)_\nu = 0$ for some $\nu \in J^{++}$, a contradiction by definition of J^{++}. Thus, $v_i^* \leq 0$ for all $i \in J^{++}$. The latter yields $y(\lambda) \in Y^k \ \forall \lambda > 0$. Since Y^k is bounded, we conclude $v_i^* = 0 \ \forall i \in J^{++}$. Multiplying finally$v^* A = 0$ with a MFCQ direction u $(Au < 0)$, it follows $v^* = 0$ as claimed. So the non-trivial vector (u^*, v^*) vanishes, i.e., $u^* = 0$ is impossible.

(ii) Since $f, g \in C^2$, we have $q(u) = D^2_{xx} L(x^k, y)u$, hence (8.53) means

$$u^* D^2_{xx} L(x^0, y) + v^* A(x^0) = 0 \ \ \forall y \in Y^{oS}$$

and (8.52) holds with $x^k = x^0$, too. For each $y^0 \in Y^{oS}$ and $(a, b) = 0$, so the point $(u^*, v^*, A(x^0)u^*)$ is a solution of system (8.13).

(iii) If (x^0, Y^0) satisfies TF-injectivity w.r. to u, then it follows $u^* = 0$ from (ii), and $v^* = 0$ from (i). So we obtain $(u^*, v^*) = 0$, i.e., the pseudo-Lipschitz property must hold. □

8.4 Taylor Expansion of Critical Values

In the present section, we derive formulas for the Taylor expansion of the critical value function φ with second-order terms $TD\varphi$ and $CD\varphi$, supposing different regularity properties of the critical point map. These results are understood as supplements to the well–developed theory of first- and second-order directional differentiability of the *optimal value function,* for which we refer to the books [Gol72, DM74, Fia83, Gau94, Lev94, BS00] with many references to the field.

8.4.1 Marginal Map under Canonical Perturbations

Again we regard the canonically perturbed standard problem

$$\mathrm{P}(a,b): \quad \min\{f(x) - \langle a, x\rangle \mid g(x) \le b\}, \ (f,g) \in C^{1,1}(\mathbb{R}^n, \mathbb{R}^{1+m}),$$

at some critical point $s^0 = (x^0, y^0)$ for $(a,b) = (0,0)$.

Supposing strong regularity, the critical points $(x,y) \in F^{-1}(a,b)$ are locally unique and Lipschitz for small parameters, and so is the *marginal map* (or *critical value function*)

$$\varphi = \varphi(a,b) \quad \text{defined as } \varphi = f(x) - \langle a,x\rangle \ , \ (x,y) \in F^{-1}(a,b). \tag{8.54}$$

Under convexity and/or C^2 hypotheses, the structure of φ is well-known, for basic studies we refer to the literature just mentioned.

Because $f \in C^1$ and $F^{-1} \in C^{0,1}$ (of course, locally), we may apply chain rule (6.19) to determine the T-derivative of φ:

$$T\varphi(0,0)(\alpha,\beta) = \left\{ \psi \,\middle|\, \begin{array}{l} \psi = Df(x^0)u - \langle \alpha, x^0\rangle \text{ with some} \\ (u,v) \in TF^{-1}((0,0),s^0)(\alpha,\beta) \end{array} \right\}.$$

The elements of the set $S := TF^{-1}((0,0),s^0)(\alpha,\beta)$ are known by Theorem 7.6 and Lemma 6.1: One has $(u,v) \in S$ iff

$$Q_T(u) + \sum r_i v_i A_i \ \in \ \alpha, \tag{8.55}$$

$$A_i u \ - (1-r_i)v_i \ = \ \beta_i, \qquad r \in \mathcal{R}_T(y^0), \tag{8.56}$$

where $Q_T(u) = T[D_xL](s^0)(u)$ and $\mathcal{R}_T(y^0)$ are defined according to Chapter 7. Recall that the set $T^-(\alpha,\beta)$ collects all (u,v,r) satisfying (8.55) and (8.56), and $C^-(\alpha,\beta)$ is the corresponding set for the contingent derivative, where one has to replace in (8.55) $Q_T(u)$ by $Q_C(u) = C[D_xL](s^0)(u)$ and in (8.56) $\mathcal{R}_T(y^0)$ by $\mathcal{R}_C(y^0)$.

In what follows, we make sure that $D\varphi$ exists and study the explicit form of the $C^{1,1}$**- derivatives** $TD\varphi$ and $CD\varphi$. Throughout this section, we often use the convention to

$$\text{write } Cg(z)(u) = Dg(z)u \text{ instead of } Cg(z)(u) = \{Dg(z)u\}$$

if $Dg(z)$ exists, the same for $Tg(z)$.

Theorem 8.43 ($C^{1,1}$ derivatives of marginal maps). *Under strong regularity of F at a zero $s^0 = (x^0, y^0)$, the map φ belongs to $C^{1,1}$, it holds*

$$D\varphi(0,0) = -(x^0, y^{0+})$$

and

$$\begin{array}{rcl} TD\varphi(0,0)(\alpha,\beta) & = & \{-(u, r_1 v_1, ..., r_m v_m) | (u,v,r) \in T^-(\alpha,\beta)\}, \\ CD\varphi(0,0)(\alpha,\beta) & = & \{-(u, r_1 v_1, ..., r_m v_m) | (u,v,r) \in C^-(\alpha,\beta)\}, \end{array}$$

and moreover,

$$\begin{array}{rcl} \Pi_T & = & \{-\langle Q_T(u), u\rangle - 2\sum_i \beta_i r_i v_i - \sum_i r_i(1-r_i)v_i^2 | (u,v,r) \in T^-(\alpha,\beta)\}, \\ \Pi_C & = & \{-\langle Q_C(u), u\rangle - 2\sum_i \beta_i r_i v_i \qquad | (u,v,r) \in C^-(\alpha,\beta)\}, \end{array}$$

where $\Pi_T := \langle(\alpha,\beta), TD\varphi(0,0)(\alpha,\beta)\rangle$ *and* $\Pi_C := \langle(\alpha,\beta), CD\varphi(0,0)(\alpha,\beta)\rangle$. ◇

Note: Under strict complementarity of a C^2 problem and with $I^+ = \{i | y^0 > 0\}$, this yields

$$\Pi_T = \Pi_C = \{-\langle u, Hu\rangle - 2\sum_{i\in I^+} v_i\beta_i \,|\, Hu + \sum_{i\in I^+} v_i A_i = \alpha \text{ and } A_i u = \beta_i,\ i \in I^+\},$$

which is a singleton. ◇

Proof. (Representation of $D\varphi$). We show that

$$D\varphi(a,b) = -(x, y^+), (x,y) = F^{-1}(a,b). \tag{8.57}$$

Since $F(s^0) = 0$, we have by Lemma 7.7,

$$Df(x^0)\,u = -\sum_i (y_i^0)^+ \beta_i. \tag{8.58}$$

Therefore, *independently* of the choice of $(u,v) \in S$, the set $T\varphi(0,0)(\alpha,\beta)$ consists only of the element

$$\langle -(x^0, y^{0+}), (\alpha,\beta)\rangle,$$

i.e., the inner product of the negative KKT point and the direction under consideration. Since strong regularity is persistent, the same applies to small parameters (a,b). Hence, $T\varphi$ is single-valued near the origin. So (see Exercise 14) $D\varphi$ locally exists and has the form (8.57).

(Representation of $TD\varphi$). Up to the term y^+ (in place of y), $D\varphi$ is $-F^{-1}$ where $S := TF^{-1}((0,0), s^0)(\alpha,\beta)$ is given by the components (u,v) of $T^-(\alpha,\beta)$. By Lemma 7.4, the terms $r_i v_i$ form just the T-derivative of the proper Lagrange multiplier, i.e.,

$$\theta^{-1}[(y+\theta v)^+ - y^+] = (r_1 v_1, ..., r_m v_m), \quad r \in \mathcal{R}_T(y^0) \quad \text{(for small } \theta).$$

Taking into account that for each $(u,v,r) \in T^-(\alpha,\beta)$, we can find sequences $\theta \downarrow 0$ and $(x,y) \to s^0$ which realize all limit-relations of Theorem 7.6 at once, we obtain that *all* vectors $-(u, r_1v_1, ..., r_mv_m)$ belong to $TD\varphi(0,0)(\alpha,\beta)$. So the representation of $TD\varphi$ is correct.

(Representation of Π_T). By definition, we have

$$\Pi_T = \{-\langle(\alpha,\beta),(u,r_1v_1,...,r_mv_m)\rangle \mid (u,v,r) \in T^-(\alpha,\beta)\}.$$

To see how $q \in \Pi_T$ depends on $Q_T(u)$ – which reduces to $D_x^2L(x^0,y^{0+})u$ for C^2 problems –, we use (8.55) and (8.56) in order to replace α and β:

$$\begin{aligned} -q &= \langle(\alpha,\beta),(u,r_1v_1,...,r_mv_m)\rangle \\ &\in \langle Q_T(u) + \sum_i r_iv_iA_i, u\rangle + \sum_i \beta_ir_iv_i \\ &= \langle Q_T(u), u\rangle + 2\sum_i \beta_ir_iv_i + \sum_i r_i(1-r_i)v_i^2. \end{aligned}$$

For the *contingent derivatives* of $D\varphi$, the same system (8.55), (8.56) is crucial after the replacements mentioned above. The formulas follow now by analogue arguments, and $r_i(1-r_i) = 0$ holds due to the definition of $\mathcal{R}_C(y^0, v)$. □

In accordance with Theorem 6.20, the set Π_T provides us with a second-order approximation of φ near the origin: For fixed (α,β) and $\theta \downarrow 0$ it holds

$$\varphi(\theta(\alpha,\beta)) - \varphi(0,0) = \theta D\varphi(0,0)(\alpha,\beta) + \frac{1}{2}\theta^2 q(\theta)$$

with some

$$q(\theta) \in \langle(\alpha,\beta), TD\varphi(\varrho\theta(\alpha,\beta))(\alpha,\beta)\rangle,\ \varrho \in (0,1).$$

Clearly, $q(\theta)$ has a cluster point q^0 in Π_T. Thus,

$$q^0 = -[\langle\alpha,u\rangle + \sum_i \beta_ir_iv_i] \text{ for some } (u,v,r) \in T^-(\alpha,\beta).$$

Having Π_C, Theorem 6.23 gives us a condition for *growth at* some point (like Corollary 6.21 for growth near a point via Π_T).

Corollary 8.44 (lower estimates). *Under strong regularity of F at a zero s^0, it holds:*

(i) If $\lambda < \inf \Pi_C$, then one has for θ sufficiently small,

$$\varphi(\theta(\alpha,\beta)) - \varphi(0,0) \geq \theta D\varphi(0,0)(\alpha,\beta) + \frac{1}{2}\theta^2\lambda.$$

(ii) If $\lambda < \inf \Pi_T$, then this estimate is locally persistent, i.e., there exists $\varepsilon > 0$ such that

$$\varphi((a,b) + \theta(\alpha,\beta)) - \varphi(a,b) \geq \theta D\varphi(a,b)(\alpha,\beta) + \frac{1}{2}\theta^2\lambda,$$

whenever $0 < \theta < \varepsilon$ and $\|(a,b)\| < \varepsilon$

◇

8.4.2 Marginal Map under Nonlinear Perturbations

For parametric problems of the kind

$$\tilde{\mathrm{P}}(t) \qquad \min\{f(x,t)|g(x,t) \le 0\}, \ (f,g) \in C^{1,1}(\mathbb{R}^{n+\mu},\mathbb{R}^{1+m}),$$

the Kojima-function F depends on t. For any t, we denote by

$\tilde{X}(t)$ the set of stationary points,
$\tilde{S}(t)$ the set of critical points.

We assume that

$$t \text{ varies near } 0, \text{ and } s^0 = (x^0, y^0) \in \tilde{S}(0).$$

Throughout this subsection we suppose at least that

$$\begin{array}{l}\textbf{the partial derivative } D_tF \textbf{ (w.r. to all arguments) exists}\\ \textbf{and belongs to } C^{0,1}\textbf{, and } \tilde{X} \textbf{ is upper regular at } (0,x^0),\\ \textbf{say with rank } K \textbf{ and neighborhood } \Omega \ni x^0,\end{array} \tag{8.59}$$

where the latter means that for some $\delta > 0$,

$$\emptyset \ne \tilde{X}(t) \cap \Omega \subset B(x^0, K\|t\|) \text{ whenever } \|t\| < \delta$$

We consider

$$\tilde{\varphi}(t) = \{f(x,t)|x \in \tilde{X}(t) \cap \Omega\}$$

and are interested in the interplay of the different regularity assumptions for obtaining more or less detailed characterizations of the map $\tilde{\varphi}$.

Formulas under Upper Regularity of Stationary Points

Note that, by (8.59),

$$\tilde{\varphi}(0) = \{f(x^0,0)\},$$

but $\tilde{\varphi}$ may be multivalued for $t \ne 0$. Further, the set $C\tilde{X}(0,x^0)$ has non-empty images by (8.59). Let us first observe that, due to $f \in C^1$ and again (8.59), it holds the usual chain rule formula

$$C\tilde{\varphi}(0)(\tau) = D_xf(x^0,0)C\tilde{X}(0,x^0)(\tau) + D_tf(x^0,0)\tau \qquad \forall \tau \in \mathbb{R}^\mu. \tag{8.60}$$

Indeed, to each sequence $x(t) \in \tilde{X}(t) \cap \Omega$, $t = \theta\tau + o(\theta)$, $\theta = \theta_k \downarrow 0$, there corresponds, by (8.59), at least one approach direction $\xi = \lim \theta^{-1}(x(t) - x^0) \in C\tilde{X}(0,x^0)(\tau)$ (for a certain subsequence of $\theta \downarrow 0$). Since every element of $\tilde{\varphi}(t)$ coincides with some $f(x(t),t)$ and because

$$\begin{array}{l} f(x(t),t) = f(x^0,t) + \theta D_xf(x^0,t)\tau + o_1(\theta) \text{ holds with}\\ t^{-1}\|o_1(\theta)\| \downarrow 0 \text{ as } \theta \downarrow 0 \textit{ uniformly} \text{ for } \tau \in B,\end{array}$$

we obtain the inclusion "$\subset$". The reverse inclusion becomes evident after writing any $\xi \in C\widetilde{X}(0, x^0)(\tau)$ in limit-form.

Having (8.59) and (8.60), we can apply all the arguments of §6.6.2 given there for the contingent derivative $C\widetilde{X}$. There, the hypothesis (6.63) had turned out to be crucial. In the present context, (6.63) attains the following form:

$$\begin{array}{l}\text{For given points } x \in \widetilde{X}(t) \text{ with } (x,t) \to (x^0,0) \text{ and } t \to 0,\\ \text{there exist } y \text{ in such a way that both } (x,y) \in \widetilde{S}(t) \text{ and a}\\ \text{Lipschitz estimate } \|y - y^0\| \le L\|(x - x^0, t)\| \text{ holds true.}\end{array} \tag{8.61}$$

Theorem 8.45 ($C\tilde{\varphi}$ for nonlinear perturbations I). *Let f and g belong to $C^{1,1}$, and suppose (8.59) and (8.61). Then, for $s^0 = (x^0, y^0) \in \widetilde{S}(0)$, one has*

$$C\tilde{\varphi}(0)(\tau) = \{D_t f(x^0,0)\tau + \langle (y^0)^+, D_t g(x^0,0)\tau \rangle\} = \{D_t L(s^0,0)\tau\}, \tag{8.62}$$

where $L(x,y,t) = f(x,t) + \langle y^+, g(x,t)\rangle$ is the related Lagrangian. ◇

Remark 8.46 In (8.62), $C\tilde{\varphi}(0)(\cdot)$ is a linear map and approximates the (locally upper Lipschitzian) multifunction $\tilde{\varphi}$ in the form

$$\tilde{\varphi}(\tau) - \tilde{\varphi}(0) = C\tilde{\varphi}(0)(\tau) + o(\tau),$$

cf. Lemma A.3 and the subsequent remark. So $C\tilde{\varphi}(0)$ can be identified with the Fréchet derivative of $\tilde{\varphi}$ at the origin.

Note that Theorem 8.45 even holds if f, g_i are only C^1 functions (in particular, without supposing the smooth parameter dependence of (8.59)). This follows from [JMT86, Lemma 2.1], where Fréchet differentiability at the origin was proven straightforwardly when assuming the existence of a pointwise Lipschitz (at 0) selection of $\widetilde{S}$. ◇

Proof of Theorem 8.45. By Theorem 6.28 (i), it holds with $(\alpha, \beta) = -D_t F(s^0, 0)\tau$,

$$C\widetilde{X}(0,x^0)(\tau) \subset U_C(\alpha,\beta) := \{u | \exists v : (u,v) \in CS((0,0),s^0)(\alpha,\beta)\}, \tag{8.63}$$

where the images of S are the *critical* points of problem P(0)(a,b), i.e., $S(a,b) = F^{-1}(a,b,0)$. Thus,

$$u \in U_C(\alpha,\beta) \Leftrightarrow (u,v,r) \in C^-(\alpha,\beta) \text{ for some } (v,r).$$

Since $F(s^0,0) = 0$, Lemma 7.7 ensures

$$D_x f(x^0,0)u = -\sum_i (y_i^0)^+\beta_i = -\sum_i (y_i^0)^+(-D_t g_i(s^0,0)\tau) \quad \forall u \in U_C(\alpha,\beta).$$

Again, this term does not depend on the selection of u. Hence (8.60) and (8.63) ensure (8.62). □

Note. Writing canonical perturbations as $t = (a,b)$, $f(x,t) := f(x) - \langle a, x\rangle$, $g(x,t) := g(x) - b$, and splitting τ into (τ_a, τ_b), formula (8.62) yields for the corresponding marginal function φ that

$$C\varphi(0)\tau = D_t L(s^0, 0)\tau = -\langle x^0, \tau_a\rangle - \langle y^{0+}, \tau_b\rangle,$$

this is again again formula (8.57). ◇

Condition (8.61) is valid under several constraint qualifications which have been already discussed in Chapter 5. Let us consider two special cases of the foregoing theorem.

(i) In particular, (8.61) holds true if $S = F(\cdot,\cdot,0)^{-1}$ is *pseudo-Lipschitz* at $((0,0), s^0)$, because the latter yields LICQ due to Lemma 7.1. Then, (8.63) is even valid as equation. To see this, one can use the same arguments as under Theorem 6.28 (iii), now supported by Theorem 6.27. So, formula (8.62) holds true though the set $\tilde{X}(\tau)$ of stationary points is not necessarily single-valued.

(ii) If the Kojima function $F(\cdot,\cdot,0)$ of the unperturbed problem $\tilde{\text{P}}(0)$ is even *strongly* regular (i.e., $F(\cdot,\cdot,0)$ is locally Lipschitz invertible), then the critical point mapping $t \mapsto \tilde{S}(t) = \{s(t)\}$ is Lipschitz near 0, and (8.62) is true not only for $t = 0$ but also for parameters t *near* 0, i.e.

$$C\tilde{\varphi}(t)(\tau) = D_t L(s(t), t)\tau.$$

In the present situation, we again easily see that

$$D\tilde{\varphi}(t) = D_t L(s(t), t) \tag{8.64}$$

locally exists as a Lipschitz function.

Formulas under Strong Regularity and Smooth Parametrization

Suppose that P(0) is strongly regular at s^0 and hence the critical point map $\tilde{S}$ is locally single–valued and Lipschitz near $(0, s^0)$. Further suppose that

$$D_t L(\cdot,\cdot) \text{ is continuously differentiable near } (s^0, 0). \tag{8.65}$$

Then the Thibault derivative of $D\tilde{\varphi}$ can be computed via formula (6.19). Indeed, let us first put

$$H(s,t) = D_t L(s,t) \quad \text{and} \quad G(t) = (s(t), \text{id}\,(t)),$$

where id $(\cdot)$ denotes the identity mapping. Then, by (8.64),

$$D\tilde{\varphi}(t) = D_t L(s(t), t) = H(G(t)). \tag{8.66}$$

Since G is locally Lipschitz and $H \in C^1$, it holds

$$TD\tilde{\varphi}(0)(\tau) = DH(G(0))[TG(0)(\tau)].$$

We show that the crucial component $Ts(t)$ of $TG(0)$ fulfills the natural chain rule

$$Ts(0)(\tau) = TF_o^{-1}(-D_tF(s^0,0)\tau)$$

with the local inverse F_o^{-1} of $F(\cdot,0)$. Indeed, by putting

$$r_s(t) := F(s,t) - F(s,0) + D_tF(s^0,0),$$

the function G assigns to the point t the pair $(s(t),t)$, where $s = s(t)$ is the critical point given by

$$0 = F(s,t) = F(s,0) + D_tF(s^0,0)t + r_s(t).$$

Using F_o^{-1}, this is

$$s(t) = F_o^{-1}(-D_t(s^0,0)t - r_s(t)).$$

In order to estimate $r_s(t)$, let $(s,t) \in (s^0 + \delta K\,B, \delta\,B)$. Then we obtain (uniformly) for the quantities of the $C^{0,1}$ norm of $r_s(\cdot)$ on δB,

$$\text{Lip}\,(r_s, \delta B) \le O(\delta) \text{ and } \sup_{t\in\delta B} \|r_s(t)\| \le o(\delta) \text{ if } s \in s^0 + \delta KB.$$

So, since F_o^{-1} is locally Lipschitz by assumption and $O(\delta)$ vanishes, we obtain in fact

$$Ts(0)(\tau) = TF_o^{-1}(-D_tF(s^0,0)\tau). \tag{8.67}$$

The latter set is given by Theorem 7.6 and the rule of the inverse derivative. The same formula for the contingent derivative

$$Cs(0)(\tau) = CF_o^{-1}(-D_tF(s^0,0)\tau) \tag{8.68}$$

follows (under strong regularity) by completely analogous arguments, we omit the details.

Theorem 8.47 ($C\tilde{\varphi}$ for nonlinear perturbations II). *Let f and g belong to $C^{1,1}$, and suppose (8.65). If $s^0 = (x^0, y^0)$ is a strongly regular critical point of P(0), then*

$$TD\tilde{\varphi}(0)(\tau) = D^2_{st}L(s^0,0)TF_o^{-1}(-D_tF(s^0,0)\tau) + D^2_{tt}L(s^0,0)\tau,$$
$$CD\tilde{\varphi}(0)(\tau) = D^2_{st}L(s^0,0)CF_o^{-1}(-D_tF(s^0,0)\tau) + D^2_{tt}L(s^0,0)\tau.$$

◇

Proof. With $H(s,t) = D_tL(s,t)$and $G(t) = (s(t),t) = (s(t), \text{id}\,(t))$, one has, as shown above,

$$D\tilde{\varphi}(t) = D_tL(s(t),t) = H(G(t)) \text{ and } TD\tilde{\varphi}(0)(\tau) = DH(G(0))[TG(0)(\tau)].$$

Hence, Theorem 6.8 yields

$$\begin{aligned} TD\tilde{\varphi}(0)(\tau) &= TD_tL(s(\cdot), \text{id}\,(\cdot))(s^0,0)(\tau) \\ &= T_sD_tL(s^0,0)Ts(0)(\tau) + T_tD_tL(s(0),0)(\tau) \\ &= D_sD_tL(s^0,0)Ts(0)(\tau) + D_tD_tL(s^0,0)\tau, \end{aligned}$$

i.e., by using (8.67), the assertion for $TD\tilde{\varphi}$ is shown. The second assertion follows analogously. □

Note. Writing canonical perturbations as $t = (a,b)$, $f(x,t) := f(x) - \langle a,x\rangle$, $g(x,t) := g(x) - b$, and splitting τ into (τ_a, τ_b), the term $D^2_{tt}L(s^0,0)\tau$ vanishes, and we have for $F = (F_1, F_2)$ and $s^0 = (x^0, y^0)$,

$$-D_tF_1(s^0,0)\tau = \tau_a \text{ and } -D_tF_2(s^0,0)\tau = \tau_b.$$

Further,

$$\begin{aligned} D^2_{st}L(s^0,0)(\xi,\eta) &= D^2_{xt}L(s^0,0)\xi + D^2_{yt}L(s^0,0)\eta \\ &= [D^2_{xt}f(x^0,0) + \textstyle\sum_{i=1}^m (y_i^0)^+ D^2_{xt}g_i(x^0,0)]\xi \\ &\quad +[D^2_{yt}\textstyle\sum_{i=1}^m (y_i^0)^+ D^2_{yt}g_i(x^0,0)]\eta \end{aligned}$$

holds true. ◇

Formulas in Terms of the Critical Value Function Given under Canonical Perturbations

One can directly compare $C\tilde{\varphi}$ with the contingent derivative $C\varphi$ of the marginal map φ of the canonically perturbed problem at $t = 0$,

$$(\mathrm{P})(a,b): \quad \min\{f(x,0) - \langle a,x\rangle \mid g(x,0) \le b\}.$$

To do this we suppose again that

s^0 is a strongly regular critical point of (P)(0,0),
and $D_tL(\cdot,\cdot)$ is locally Lipschitz near $(s^0,0)$.

We consider

$$F(s,t) = F(s,0) + D_tF(s,0)t + o(s,t), \; s = (x,y),$$

and use that $s \in \Omega \cap \tilde{S}(t)$ holds if and only if $s \in \Omega$ is a critical point of (P)(a,b) for the perturbation $(a,b) = -[D_tF(s,0)t + o(s,t)]$.

Setting $t = \theta\pi + o_1(\theta)(\theta \downarrow 0)$, we may write

$$(a,b) = -[\theta D_tF(s,0)\pi + o_2(\theta)].$$

Here, o_2 depends also on s, but it holds $\|o_2(\theta)\|/\theta \downarrow 0$ for $\theta \downarrow 0$ uniformly with respect to $s \to s^0$. Due to strong regularity and $t \to 0$, the points $s \in \Omega \cap \Sigma(\theta\pi + o_1(\theta))$ converge to s^0, indeed. We thus obtain with some new $o(\cdot)$ that

$$(a,b) = \theta(\alpha,\beta) + o(\theta), \text{ where } (\alpha,\beta) = -D_tF(s^0,0)\pi.$$

Hence,

$$\tilde{\varphi}(\theta\pi + o_1(\theta)) - \tilde{\varphi}(0) = \varphi(\theta(\alpha,\beta) + o(\theta)) - \varphi(0,0).$$

Due to $C\tilde{\varphi}(0)(\pi) \ne \emptyset$, this yields

$$C\tilde{\varphi}(0)(\pi) = D\varphi_0(0,0)(\alpha,\beta) = D\varphi(0,0)(-D_tF(s^0,0))\pi.$$

By persistence of strong regularity, the obtained formula

$$D\tilde{\varphi}(0) = -D\varphi_0(0,0)D_tF(s^0,0) \tag{8.69}$$

then holds once more also for parameters $t \neq 0$ near the origin. By Theorem 6.11, this yields the contingent derivative of $D\tilde{\varphi}$ as

$$CD\tilde{\varphi}(0) = -[CD\varphi(0,0)D_tF(s^0,0) + D\varphi(0,0)D_t^2F(s^0,0)], \tag{8.70}$$

and the set Π_C becomes

$$\begin{aligned} \Pi_C &= \{\langle \pi, CD\tilde{\varphi}(0)(\pi)\rangle\} \\ &= -\langle \pi, CD\varphi(0,0)D_tF(s^0,0)\pi\rangle - \langle \pi, (x^0, y^{o+})D_{tt}^2F(s^0,0)\pi\rangle. \end{aligned} \tag{8.71}$$

Note that only differentiability properties of F along with strong regularity played any role in the present context. So, in particular, *we did not utilized* that the parameter t appeared only in the matrix M of the Kojima function $F = M N$.

Remark. The set Π_C is interesting if the original problem appears in a decomposition setting: a and b are parameters given by the "master" to some (or more than one) follower who solves his problem $\mathbf{P}(a,b)$ with primal-dual-solution $x(a,b)$, $y(a,b)$. The objective of the master consists in minimizing a function

$$H(a,b,\varphi(a,b),x(a,b),y(a,b))$$

with certain constraints concerning a, b, x and y. In the simplest case, we have $H = -\varphi(a,b)$ without constraints which yields a max-min problem, namely,

$$\max_{a,b} \varphi(a,b) = \max_{a,b} \min_x \{f(x) - \langle a, x\rangle \mid g(x) \leq b\}$$

Clearly, the master is interested in (stationary) points where $\varphi(a,b) = 0$. To show that such a point is a local solution, one may consider $CD\varphi(a,b)$ in all directions (α,β) (or under constraints in the "feasible" directions at (a,b)).

Chapter 9

Derivatives and Regularity of Further Nonsmooth Maps

9.1 Generalized Derivatives for Positively Homogeneous Functions

Several practically important functions are positively homogeneous, e.g., Euclidean projections onto a closed convex cone, among of them many NCP functions, the function $N(y, z)$, which appears in Kojima functions and the directional derivative $\varphi(\cdot) = f'(x; \cdot)$ of a directionally differentiable, locally Lipschitz function f. Injectivity of $T\varphi(0)$ for the directional derivative then means: There is a unique and (globally) Lipschitzian assignment $u = u(v)$ such that $v = f'(x; u)$.

In the current subsection we will investigate the derivatives of such functions while, in the next one, those properties of NCP-functions will be studied that are important for solving the related NCP-equations, cf. Section 1.3.

Accordingly, we suppose that

$$g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m) \text{ is positively homogeneous,}$$

and want to determine the derivatives $Tg(y)$, $Cg(y)$ and $D^\circ g(y)$ at the origin.

First of all we observe that, for all $p, q \in \mathbb{R}^n$ and positive t, λ,

$$\tfrac{1}{t}[g(p + tq) - g(p)] = \tfrac{\lambda}{t}\,[g(\tfrac{1}{\lambda}p + \tfrac{t}{\lambda}q) - g(\tfrac{1}{\lambda}p)].$$

This tells us that the derivatives Tf and Cf are norm-invariant, i.e.,

$$Tg(p) = Tg(\lambda p)\ \forall \lambda > 0;\ \text{similarly for } Cg. \tag{9.1}$$

Next, we immediately see that

$$Cg(0)(r) = \{g(r)\}.$$

In consequence, (9.1) along with closenness of $Tg(\cdot)(r)$ yields

$$Tg(y)(r) \subset Tg(0)(r) \quad \forall y. \tag{9.2}$$

We will show that, for the set $Tg(0)(r)$, the following collection of difference quotients

$$P(r) = \{\alpha^{-1}[g(p+\alpha r) - g(p)] \mid \alpha > 0 \text{ and } p \in \text{bd}\, B\}$$

which is $\{g(q+r) - g(q) \mid q \neq 0\}$, plays a key role. Let Θ^1 denote the set of all C^1 points of g.

Lemma 9.1 ($Tg(0)$ and $D^\circ g$ for positively homogeneous functions). *Let $g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ be positively homogeneous. Then, g is simple at the origin, and it holds*

(i) $Tg(0)(r) = \text{cl}\, P(r) = \text{cl}\,(\, P(r) \,\cup\, Tg(\mathbb{R}^n \setminus \{0\})(r)\,)$.

(ii) If $m = 1$, one has

$$\begin{aligned} Tg(0)(r) &= \text{cl}\,(\text{conv}\,\{g(r), -g(-r)\} \cup Tg(\mathbb{R}^n \setminus \{0\})(r)\,) \\ &= \text{cl conv}\, Tg(\mathbb{R}^n \setminus \{0\})(r). \end{aligned}$$

(iii) If g is pseudo-smooth then $D^\circ g(0) = \text{cl}\, Dg(\Theta^1)$.

◇

Proof. (iii) This statement is a direct consequence of formula (9.1) and the definition of $D^\circ g(0)$ as being the set of all limits of sequences $Dg(y)$ for $y \to 0$ in Θ^1.

(i) To determine all limits L of terms

$$Q = t^{-1}\,[g(y+tr) - g(y)]\ (t \downarrow 0\,,\ y \to 0),$$

we put $h = \|y\|/t$ and distinguish three cases.

CASE 1: $h \to 0$.
Then $Q = t^{-1}g(y+tr) - t^{-1}g(y) = g(y/t + r) - g(y/t)$ and $L = g(r)$.

CASE 2: $h \to \gamma > 0$.
Now $Q = t^{-1}\|y\|[\,g(\|y\|^{-1}y + t\|y\|^{-1}r) - g(\|y\|^{-1}y)\,]$. Without loss of generality,

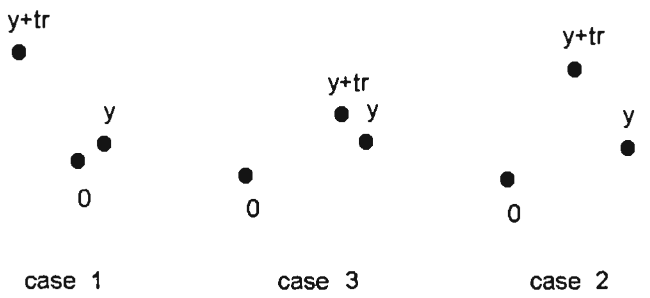

Figure 9.1: Limit situations in the proof of Lemma 9.1.

we may assume that $q := y\|y\|^{-1} \to p \in$ bd B and $s = t\|y\|^{-1} \to \alpha := \gamma^{-1}$. Then

$$Q = [g(q+sr) - g(q)]/s \text{ tends to } L = \alpha^{-1}[g(p+\alpha r) - g(p)].$$

CASE 3: $h \to \infty$.
With the settings of case 2, we now obtain

$$Q = [g(q+sr) - g(q)]/s \text{ where } s \downarrow 0.$$

Thus, $L \in Tg(p)(r)$.

In the cases 2 and 3, *each pair* of $\alpha \geq 0$ and $p \in \text{bd}\, B$ can be written as

$$\alpha = \lim t\|y\|^{-1} \text{ and } p = \lim \|y\|^{-1}y \text{ by suitable choice of } t \text{ and } y.$$

This yields $\{g(r)\} \cup P(r) \subset Tg(0)(r)$.

Moreover, $Tg(p)(r) \subset Tg(0)(r)$ is true due to (9.2). So $Tg(0)(r)$ is just the closed set which contains $\{g(r)\} \cup P(r)$ as well as $\cup_{y \neq 0} Tg(y)(r)$. Taking into account that

$$g(r) = \lim_{\alpha \to \infty} \alpha^{-1}[g(p+\alpha r) - g(p)] \in \text{cl}\, P(r),$$

we now obtain assertion (i).

(ii) If $m = 1$, the terms $d = \alpha^{-1}[g(p+\alpha r) - g(p)]$ may be split into two groups G_1 and G_2.

For group G_1, the origin is not contained in the line-segment S connecting $p + \alpha r$ and p. By the mean-value statement (6.38), we can write $d \in Tg(y)(r)$ with some $y \in S$. Because $y \neq 0$, we get $d \in \cup_{y \neq 0} Tg(y)(r)$.

For group G_2, it holds $0 \in S$. Considering g on S one easily sees that d is a convex combination of $g(r)$ and $-g(-r)$. Conversely, every such convex combination can be written as a quotient $d = \alpha^{-1}[g(p + \alpha r) - g(p)]$. So it follows from (i) that

$$Tg(0)(r) = \mathrm{cl}\,(\mathrm{conv}\,\{g(r), -g(-r)\} \cup \bigcup_{y \neq 0} Tg(y)(r)).$$

Moreover, as a connected set in $\mathbb{R}$, $Tg(0)(r)$ is convex. Hence,

$$C := \mathrm{cl\,conv} \cup_{y \neq 0} Tg(y)(r) \text{ is contained in } Tg(0)(r).$$

Since the elements $g(r) - g(0)$ and $g(0) - g(-r)$ belong to C (again due to (6.38)), the same holds for conv $\{g(r), -g(-r)\}$. This yields $C = Tg(0)(r)$.

So, (i), (ii) and (iii) have been verified. The proof of the simple-property is left as Exercise 15. □

Exercise 15. Verify that positively homogeneous $g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ are *simple* at the origin. ◇

Exercise 16. Show that, for $m = 1$, the situation conv $\{g(r), -g(-r)\} \not\subset$ cl $\cup_{y \neq 0} Tg(y)(r)$ must be taken into account. ◇

Difficulties for Compositions

We are now going to study $Tf(0)(u)$ for a composed function $f(x) = g(z(x))$, where we suppose that

$$g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m) \text{ is positively homogeneous,}$$
$$\text{and } z \in C^1(\mathbb{R}^k, \mathbb{R}^n), z(0) = 0.$$

We even suppose that

$$g \text{ is } C^1 \text{ on } \mathbb{R}^m \setminus \{0\},$$

so only zeros of z can make any difficulties.

Clearly, it holds $Tf(0)(u) \subset Tg(0)(Dz(0)u)$, but not any of our chain rules in Section 6.4 guarantees the equality. So the present function is a good example for discussing the related problems in detail. We have to regard all limits

$$L = \lim t^{-1}[g(z(x_t + tu)) - g(z(x_t))] \text{ for certain } t = t_k \downarrow 0 \text{ and } x_t \to 0.$$

Setting $r = Dz(0)u$ we may write $L = \lim t^{-1}[g(z(x_t) + tr) - g(z(x_t))]$. For $x_t \equiv 0$ we obtain $L = g(r)$; the same is true if $z(x_t) \equiv 0$.

So let us turn to the crucial case of $z(x_t) \neq 0$ for all t under consideration. With $y = z(x_t)$, the possible limits L have been considered in the proof of Lemma 9.1. So we know that L depends on

$$\gamma = \lim \|z(x_t)\|/t \quad \text{and} \quad p = \lim z(x_t)/\|z(x_t)\|$$

according to the cases considered under Lemma 9.1:

$$L = g(r^i) \quad \text{if} \quad \gamma = 0, \tag{9.3}$$
$$L = \alpha^{-1}[g(p+\alpha r) - g(p)] \quad \text{if} \quad \alpha = \gamma^{-1} \in \mathbb{R} \tag{9.4}$$
$$L = Dg(p)r \quad \text{if} \quad \gamma = \infty. \tag{9.5}$$

Therefore, we put

$$s(t) = \|x_t\|, \qquad w(t) = x_t/\|x_t\|$$

and select a (further) subsequence of $t = t_k$ such that the limits

$$\beta = \lim s(t)/t \in \mathbb{R}^+ \cup \{\infty\}, \; w = \lim w(t)$$

exist.

The vector w plays, for the sequence $x_t \to 0$, the role of a normalized approach direction. So the directional derivatives $Dz(0)w$ become important. We consider the simple cases first.

(i) If $Dz(0)w \neq 0$, the limits γ and p are uniquely determined, namely,

$$\begin{aligned} \gamma &= \lim \|s(t)Dz(0)w + o(s(t))\|/t \\ &= \|Dz(0)w\| \lim s(t)/t \\ &= \|Dz(0)w\|\beta. \end{aligned}$$

This yields $p = Dz(0)w/\|Dz(0)w\|$. Hence γ and p are uniquely defined by β and w. So also L is well-defined by (9.3), (9.4) and (9.5).

(ii) Let $Dz(0)w = 0$ and $0 \le \beta < \infty$. Then $\gamma = 0$. Indeed, writing each component of z, say z_1, by using the mean value theorem, $z_1(x) = s(t)\; Dz_1(\Theta s(t)w(t))\; w(t)$, $\Theta \in (0,1)$, one sees that $\lim \|z(x_t)\|/s(t) = 0$. So $\beta < \infty$ yields $\gamma = 0$ and $L = g(r)$.

(iii) The crucial case consists of $Dz(0)w = 0$ and $\beta = \infty$. Now

$\|x_t\|$ is much greater than t,

For this reason, the quotients

$$\|z(x_t)\|/t = o(x_t)/t$$

may have limits γ which depend on the high-order term $o(\cdot)$.

In the last case, γ and p cannot be written by means of the approach directions w, u and first derivatives only. This makes the explicit computation of $Tf(0)(u)$ hard. To see that the unpleasant situation may really appear, let $Dz(0)w = 0$ and $w \neq 0$.

Setting, e.g., $x = sw + o_1(s)$ and $t = s^2$, we get $z(x) = o_2(s)$ (which may be non-zero due to $o_1(s)$ even for linear z) and $\beta = \infty$. Now $\gamma = \lim \|z(x)\|/t$ depends on $o_2(s)/s^2$.

9.2 NCP Functions

NCP-functions are functions g: $\mathbb{R}^2 \to \mathbb{R}$ satisfying

$$g^{-1}(0) = \{ (s,t) \in \mathbb{R}^2_+ \mid st = 0 \},$$

which is supposed for all functions g appearing in this section. They are used in order to formulate the NCP (or the subsequent complementarity conditions in some more complex system)

$$u(x) \geq 0, \qquad v(x) \geq 0, \qquad \langle u(x), v(x)\rangle = 0,$$

cf. Section 1.3, as an equation

$$h(x) = 0, \quad \text{where } h_i(x) := g(z_i(x)), \text{ and } z = (u,v) : \mathbb{R}^n \to \mathbb{R}^{2n}. \tag{9.6}$$

Because of the composed structure of h and the difficulties for computing Th (as mentioned above) and similarly for its convex hull map ∂h, the application of the derivatives $D^\circ g$ and $D^\circ z$ is more convenient in the present situation.

Therefore, we will apply the results of Section 6.4.2 and suppose throughout that

$$\begin{array}{l} g \in \text{loc}PC^1 \text{ with } C^1\text{-set } \Theta^1(g), \\ z \in \text{loc}PC^1 \text{ with } C^1\text{-set } \Theta^1(z). \end{array} \tag{9.7}$$

We further recall that the NCP is said to be *(strongly) monotone* if

$$\langle u(y) - u(x), v(y) - v(x)\rangle \geq \lambda \|y - x\|^2 \qquad \forall x, y \in \mathbb{R}^n$$

holds true, where $\lambda \geq 0$ ($\lambda > 0$) is a fixed constant. *A standard NCP* is defined by $v(x) = x$. Finally, let g_s and g_t denote the partial derivatives of g on $\Theta^1(g)$.

If $x \in \Theta^1(z)$ then monotonicity yields (via $y = x + w$ and first-order approximation):

$$\lambda \|w\|^2 \leq \sum_i (Dv_i(x)w)(Du_i(x)w).$$

The same remains true (consider limits for $x' \to x$ in $\Theta^1(z)$) if the pairs $(Du_i(x), Dv_i(x)) = (R_i u(x), R_i v(x))$ are components of a Newton function $Rz(x) \in D^\circ z(x)$, i.e.,

$$\lambda \|w\|^2 \leq \sum_i (R_i u(x)w)(R_i v(x)w). \tag{9.8}$$

In order to find some zero of h, several NCP functions g can be (and have been) used, cf. [SQ99] for some overview. Necessary and desirable properties of g may depend on z, but also on the method one is aiming to apply. So we will regard two principal possibilities of solving (9.6).

(5) minimize a so-called merit function, e.g.,

$$q(x) = \tfrac{1}{2} \sum_i h_i(x)^2 \tag{9.9}$$

by a descent method or

(ii) solve (9.6) directly by a Newton method.

Though also combinations of both ideas are possible, we study these cases separately because they require different properties of the NCP function g. We will see that, having satisfied these properties, the concrete definition of g plays a less important role.

CASE (i): Descent Methods

Having $z \in C^1$, the function g should ensure that $q \in C^1$. This is true if g satisfies

$$\Theta^1(g) \cup g^{-1}(0) = \mathbb{R}^2. \tag{9.10}$$

As a second requirement, $Dq(x) = 0$ should imply $q(x) = 0$. The latter cannot be ensured for all problems, but at least for *monotone standard NCP's*. Clearly, then g has to be monotone in a certain sense, too.

We call an NCP function g *strongly monotone* if

$$ab > 0 \text{ for all } (a,b) \in D^\circ g(s,t) \text{ with } (s,t) \in \mathbb{R}^2 \backslash g^{-1}(0).$$

Lemma 9.2 (NCP: minimizers and stationary points). *Let g fulfill (9.10) and be strongly monotone. Further, let the NCP be monotone, $z \in C^1$ and $Dv(x)$ be a regular matrix. Then $Dq(x) = 0$ implies $q(x) = 0$.* ◇

Proof. Given $\sigma = z(x)$, define w by

$$Dv_i(x)w = h_i(x)g_s(\sigma_i) \qquad (\text{if } h_i = 0 \text{ then put } h_i g_s = 0).$$

Then

$$\begin{aligned} Dq(x)w &= \sum h_i(x)g_s(\sigma_i)Du_i(x)w + \sum h_i(x)\, g_t(\sigma_i)Dv_i(x)w \\ &= \sum (Dv_i(x)w)(Du_i(x)w) + \sum h_i(x)^2 g_t(\sigma_i)g_s(\sigma_i). \end{aligned} \tag{9.11}$$

The first sum is non-negative by (9.8), the second one is positive iff $q(x) > 0$. □

Notes

(i) For *strongly monotone NCP's,* the same is true if g is *monotone* in the weaker sense

$$ab \geq 0 \text{ and } a \neq 0 \qquad \forall (a,b) \in D^\circ g(s,t) \;\&\; (s,t) \in \mathbb{R}^2 \backslash g^{-1}(0), \tag{9.12}$$

because now (9.11), (9.12) and $h(x) \neq 0$ ensure $w \neq 0$ and

$$0 < \lambda \|w\|^2 \leq \sum (Du_i(x)w)(Dv_i(x)w).$$

(ii) For $z \in$ locPC^1, one may replace Du and Dv by a Newton function as in (9.8) and can define

$$Rq(x) := \sum h_i(x)[g_s(\sigma_i)R_iu(x) + g_t(\sigma_i)R_iv(x)].$$

Then $Rq(x) = 0$ implies $q(x) = 0$ by the same arguments.

(iii) Without supposing the smoothness (9.10), one may also replace the pair $(g_s(\sigma_i), g_t(\sigma_i))$ by pairs $(a_i, b_i) \in D^\circ g(\sigma_i)$ and comes to the same conclusion.

Knowing that $Dq = 0$ implies $q = 0$, all first order methods for minimizing a C^1– or a $C^{1,1}$– function may be applied to q.

NCP-functions g satisfying the assumptions of the lemma can be chosen arbitrarily smooth. Nevertheless, one may also apply methods of nonsmooth convex optimization (cf. [SZ88, SZ92, HUL93, OKZ98]) for minimizing

$$Q(x) := \sum_i |h_i(x)|$$

as long as $G := |g|$ is sublinear and the NCP is monotone. Then we have at C^1 points that

$$Q(x) = \sum_i \langle DG(z_i(x)), z_i(x)\rangle, \qquad DQ(x) = \sum_i DG(z_i(x))Dz_i(x)$$

and

$$Q(x) + DQ(x)w = \sum_i \langle\, DG(z_i(x)),\ z_i(x) + Dz_i(x)w\,\rangle$$

hold true. Directions w satisfying

$$\langle\, DG(z_i(x)),\ z_i(x) + Dz_i(x)w\,\rangle = 0 \text{ for all } i$$

will just appear as Newton directions in the next subsection.

CASE (ii): Newton Methods

Having Newton's method in mind, we require that the NCP function g satisfies with the C^1-set $\Theta^1 = \Theta^1(g)$,

$$g \in \text{loc}PC^1 \text{ and } 0 \notin \Theta^1, \tag{9.13}$$

$$g \text{ is positively homogeneous}, \tag{9.14}$$

$$e^1 = (1,0),\ e^2 = (0,1) \in \Theta^1 \tag{9.15}$$

$$Dg(\sigma) \geq 0 \text{ and } Dg(\sigma) \neq 0\ \forall \sigma \in \Theta^1. \tag{9.16}$$

Let us discuss these conditions.

If for $g \in \mathrm{loc}PC^1$, in contrast to (9.13), $0 \in \Theta^1$, then $Dg(0) = 0$, hence $Dh_i(x) = 0$ if both $z_i(x) = 0$ and $z \in C^1$. So system (9.6) degenerates whenever strict complementarity $(z_i(x^*) \neq 0 \; \forall i)$ does not hold.

By (9.14), g belongs to the simplest class of functions satisfying $0 \notin \Theta^1(g)$.

Condition (9.15) guarantees that g is C^1 on $g^{-1}(0)\backslash\{0\}$ and h is C^1 at strictly complementary solutions x^* as long as u and v are continuously differentiable, too.

Condition (9.16) is consistent with the assumption of Lemma 9.2 and avoids singular derivatives of h for strongly monotone NCP's, cf. Theorem 10.6.

Properties and Construction of $g \in pNCP$

Let

$pNCP$ be the cone of NCP-functions g satisfying (9.13) ... (9.16),

such functions are called ***pNCP** functions.* Due to (9.14), we have

$$Dg(\sigma) = Dg(\lambda\sigma) \quad \forall\lambda > 0 \quad \forall\sigma \in \Theta^1. \tag{9.17}$$

Hence one easily derives that

$$D^\circ g(0) = \mathrm{cl}\, Dg(\Theta^1), \; g(\sigma) = Dg(\sigma)\sigma \; \forall\sigma \in \Theta^1 \;\; \text{and} \;\; g(\sigma) = D^\circ g(\sigma)\sigma. \tag{9.18}$$

In consequence, there is a positive lower bound for all gradient norms:

$$\exists p > 0 \text{ such that } \|Dg(\sigma)\| \geq p \; \forall\sigma \in \Theta^1 \text{ and } \inf \|D^\circ g(0)\| \geq p. \tag{9.19}$$

Moreover, $Dg(e^1) = \lambda e^2$ and $Dg(e^2) = \mu e^1$ hold with certain $\lambda, \mu > 0$.

Taking into account that Dg is norm-invariant and locally bounded, we obtain the basic properties

$$\begin{aligned} g(y') &= Dg(y')y' \quad \forall y' \in \Theta^1, \\ g(y) &\in D^\circ g(y)y \quad \forall y \in \mathbb{R}^n \end{aligned} \tag{9.20}$$

The first equation is just $0 - g(y') = Dg(y')(0 - y')$.

Examples of pNCP functions

(i) Put $g = g_{min}(s,t) := \min\{s,t\}$, this is an often used concave standard function, or

(ii) $g = g_{dist}(s,t) := \mathrm{dist}\,((s,t), g^{-1}(0))$, and, to satisfy (9.16), change the sign of g on $\mathbb{R}^2\backslash\mathbb{R}^2_+$.

(iii) One can define g via any norm of $\mathbb{R}^2$, such that its unit sphere $\mathrm{bd}\, B$ is piecewise smooth, has no kinks at the positive axes and fulfills

$$e^1 + e^2 \notin B, \quad B \subset e^1 + e^2 - \mathbb{R}^{2+} \quad \text{and } \{e^1, e^2\} \subset \mathrm{bd}\, B.$$

Setting $\Psi(p) = e^1 + e^2 - p$ for $p \in \text{bd}\, B$ and $g(\lambda p) = \lambda\langle\Psi(p), p\rangle$ for $\lambda \geq 0$, one easily infers that g belongs to $pNPC$.

With the Euclidean ball, one obtains the strongly monotone concave function

$$g_2(s,t) = s + t - \sqrt{s^2 + t^2},$$

used e.g. in [Kap76] (for penalization), [Fis97] and [KYF97].

(iv) In addition, g can be defined (and each $g \in pNCP$ can be written) by means of a real 2π-periodic locPC^1 function ϕ with zeros at 0 and $\frac{1}{2}\pi$

$$g(s,t) = r\phi(\omega), \text{ where } (r,\omega) \text{ are the polar coordinates of } (s,t).$$

Then, by well known derivative transformations,

$$Dg(s,t) = r^{-1}(s\phi(\omega) - tD\phi(\omega), t\phi(\omega) + sD\phi(\omega)) \quad \text{for radius } r > 0 \text{ at } (s,t) \in \Theta^1.$$

In particular, the natural setting

$$\phi(\omega) = \sin(2\omega) \text{ for } 0 \leq \omega \leq \tfrac{1}{2}\pi$$

with the symmetric extension

$$\phi(\omega) = -3\phi(\frac{1}{3}(2\pi - \omega)) \text{ for } \tfrac{1}{2}\pi \leq \omega \leq 2\pi$$

defines a function g_ϕ which satisfies, like g_2, all the already mentioned conditions.

Lemma 9.3 (limits of Dg/g for pNCP). *For $g \in pNCP$, one has*

$$\begin{aligned} \lim g_s(\sigma)/g(\sigma) &= 0 \text{ as } \sigma \to e^1 \text{ in } \Theta^1, \\ \lim g_t(\sigma)/g(\sigma) &= 0 \text{ as } \sigma \to e^2 \text{ in } \Theta^1. \end{aligned}$$

◇

Proof. We apply the polar representation of Dg, put $\sigma = (s,t) = r(\cos\omega, \sin\omega)$ and study the first limit for $\omega \to 0, t \to 0$.

Due to (9.15), the function ϕ is C^1 near 0, so one may write

$$\phi(\omega) = D\phi(0)\omega + o(\omega), \text{ and } D\phi(\omega) = D\phi(0) + O(\omega), \quad \text{where } D\phi(0) \neq 0 \text{ by (9.16)}.$$

Hence $g_s(\sigma)/g(\sigma)$

$$\begin{aligned} &= r^{-2}(s\phi(\omega) - tD\phi(\omega))/\phi(\omega) \\ &= r^{-2}s - r^{-2}t(D\phi(0) + O(\omega))/(D\phi(0)\omega + o(\omega)) \\ &= r^{-2}s - r^{-1}\omega^{-1}\sin\omega(D\phi(0) + O(\omega)/(D\phi(0) + o(\omega)/\omega). \end{aligned}$$

Since $s \to 1, r \to 1$ (due to $\sigma \to e^1$) and $\omega^{-1}\sin\omega \to 1$, we obtain the first assertion, the second one is left to the reader. □

The previous lemma allows us to interpret a Newton step in terms of the original functions. The linearized equation (9.6), i.e.,

$$h(x) + Dh(x)w = 0, \tag{9.21}$$

means with

$$\sigma_i = (u_i(x), v_i(x)) \text{ at } C^1\text{-points } x,$$
$$g(\sigma_i) + g_s(\sigma_i)Du_i(x)w + g_t(\sigma_i)Dv_i(x)w = 0,$$

i.e.,

$$[a_i Du_i(x) + b_i Dv_i(x)]w = -1 \tag{9.22}$$

where g is only a vehicle for defining the coefficients a_i, b_i as

$$a_i = g_s(\sigma_i)/g(\sigma_i), \qquad b_i = g_t(\sigma_i)/g(\sigma_i).$$

If σ_i is close to some point $(0, p_i), p_i > 0$, then the previous lemma tells us that the weight a_i is large (near to some positive real) and b_i is small. So, roughly speaking, we solve, *independently* of the concrete choice of $g \in pNCP$,

$$[a_i Du_i(x) + 0]w = -1.$$

This is exactly the form of the Newton equation with $g = g_{min}$, where $(a_i, b_i) = (1, 0)$. If σ_i is close to $(p_i, 0)$, the symmetric situation appears.

Finally, if $\sigma_i = 0$ holds at the solution x^0 then we cannot predict how the derivatives of the functions u_i and v_i will be weighted by (a_i, b_i) for x near x^0. So we have to hope that the Newton equation has still uniformly bounded solutions for all limits of (a_i, b_i) as $\sigma_i \in \Theta^1$ vanishes ("Newton-regularity"), whereupon we can define the weights at the non-C^1 point (0,0) by some of these limits. Similarly, we may proceed at other non-C^1 points of g.

Depending on g, "Newton-regularity" may be a more or less strong assumption. Though g_{min} does not fulfill the requirements of Lemma 9.2, our Theorem 10.5 will indicate that it belongs to the best pNCP-functions in view of this regularity hypothesis for Newton's method.

9.3 The C-Derivative of the Max-Function Subdifferential

To illustrate the above settings and assumptions in the context of multifunctions, we study here the contingent derivative CF where $F = \partial_c f$ is Clarke's subdifferential (the generalized Jacobian) for a maximum function

$$f(x) = \max_k F^k(x), \quad F^k \in C^2(\mathbb{R}^n, \mathbb{R}), \ k = 1, ..., m. \tag{9.23}$$

We begin with some basic facts. It is well known that $\partial_c f(x)$ is the convex hull of all "active" gradients, i.e.,

$$\partial_c f(x) = \text{conv}\,\{DF^k(x),\ k \in I(x)\}, \text{ where } I(x) = \{k | F^k(x) = f(x)\}. \tag{9.24}$$

So we have by dualization

$$x^* \in \partial_c f(x) \Leftrightarrow \langle DF^k(x) - x^*, u\rangle \geq 0 \ \forall u \ \forall k \in I(x). \tag{9.25}$$

Further, for all x and $\xi = x + u$ in some compact set we obtain due to $F^k \in C^2$ that

$$F^k(\xi) = F^k(x) + DF^k(x)u + \tfrac{1}{2}\langle u, D^2F^k(x)u\rangle + r_{k,x}(u),$$

where

$$|r_{k,x}(u)| \leq O(u)\|u\|^2.$$

Let x^* be any convex-combination of related active gradients

$$x^* = \textstyle\sum_{J(x)} \alpha_k DF^k(x) \in \partial_c f(x), \ \textstyle\sum_{J(x)} \alpha_k = 1, \alpha \geq 0; \quad J(x) \subset I(x). \tag{9.26}$$

We denote

by $\lambda(x, x^*, J(x))$ the related set of weights $\alpha \in \mathbb{R}^m$

and put

$$\alpha_i = 0 \text{ if } i \notin J(x).$$

Then,

$$\begin{aligned} f(\xi) &\geq \max_{k \in J(x)} F^k(\xi) \\ &\geq \textstyle\sum_{J(x)} \alpha_k F^k(\xi) \\ &= \textstyle\sum_{J(x)} \alpha_k [F^k(x) + DF^k(x)u + \tfrac{1}{2}\langle u, D^2F^k(x)u\rangle + r_{k,x}(u)] \\ &= f(x) + \langle x^*, u\rangle + \tfrac{1}{2}\textstyle\sum_{J(x)} \alpha_k [\, \langle u, D^2F^k(x)u\rangle + r_{k,x}(u)\,], \end{aligned} \tag{9.27}$$

and equality holds if $J(x) \subset I(\xi)$. Hence (9.26) ensures, for all x, ξ in any compact set and with $u = \xi - x$

$$f(\xi) \geq f(x) + \langle x^*, u\rangle + \tfrac{1}{2}\textstyle\sum_{J(x)} \alpha_k \, \langle u, D^2F^k(x)u\rangle - O(u)\|u\|^2. \tag{9.28}$$

So the convex combination $\sum_{J(x)} \alpha_k D^2F^k(x)u$ plays, in view of lower estimates, the role of a second derivative. Reversing the role of ξ and $x = \xi - u$ yields

$$\begin{aligned} &f(x) \geq f(\xi) - \langle \xi^*, u\rangle + \tfrac{1}{2}\textstyle\sum_{J(\xi)} \beta_k \, \langle u, D^2F^k(\xi)u\rangle - O(u)\|u\|^2 \\ &\text{whenever } \xi^* \in \partial_c f(\xi), J(\xi) \subset I(\xi) \text{ and } \beta \in \lambda(\xi, \xi^*, J(\xi)). \end{aligned} \tag{9.29}$$

Adding (9.28) and (9.29), one obtains a monotonicity relation

$$\begin{aligned} &\langle \xi^* - x^*, u\rangle \\ &\geq \tfrac{1}{2}\textstyle\sum_{J(x)} \alpha_k \langle u, D^2F^k(x)u\rangle + \tfrac{1}{2}\textstyle\sum_{J(\xi)} \beta_k \langle u, D^2F^k(\xi)u\rangle - 2O(u)\|u\|^2. \end{aligned} \tag{9.30}$$

Contingent Limits

Now assume, in addition, that

$$\begin{array}{l} \xi = x + tu_t \text{ and } \xi^* = x^* + tu_t^* \in (\partial_c f)(\xi) \\ \text{for certain } t = t_v \downarrow 0,\ u_t \to u,\ u_t^* \to u^*, \\ \text{where } x, x^*, u \text{ and } u^* \text{ are fixed,} \\ \text{i.e., } u^* \in C(\partial_c f)(x, x^*)(u). \end{array} \tag{9.31}$$

With related coefficients $\beta^v \in \lambda(\xi, \xi^*, I(\xi))$ and for some subsequence, the index sets $J^v = \{i | \beta_i^v > 0\} \subset I(\xi)$ are constant $J^v \equiv J$ and β^v converge, say $\beta^v \to \alpha$. Hence (9.28) and (9.29) can be applied with $J(\xi) = J(x) \equiv J$, and (9.30) yields, up to terms of order $o(t^2)$:

$$t^2 \langle u^*, u \rangle \geq t^2 \sum\nolimits_J \alpha_k \langle u, D^2 F^k(x) u \rangle,$$

thus

$$\langle u^*, u \rangle \geq \sum\nolimits_J \alpha_k \langle u, D^2 F^k(x) u \rangle.$$

Moreover, since $J \subset I(\xi) \subset I(x)$ for small t, now (9.28), (9.29) and (9.30) hold as equations, so

$$\langle u^*, u \rangle = \sum\nolimits_J \alpha_k \langle u, D^2 F^k(x) u \rangle$$

and

$$\begin{aligned} f(\xi) &= f(x) + t\langle x^*, u \rangle + \tfrac{1}{2} t^2 \textstyle\sum_J \alpha_k \langle u, D^2 F^k(x) u \rangle + o(t^2) \\ &= (f(x) + t\langle x^*, u \rangle + \tfrac{1}{2} t^2 \langle u^*, u \rangle + o(t^2), \end{aligned}$$

and $C(\partial_c f)(x, x^*)(u)$ may be seen as second order derivative of f in direction u.

The analysis of $C\,\partial_c f$ from the viewpoint of second order derivatives has been developed in [Roc88] and can be applied even to composed function $f = h \circ F$ where $h \in PC^2(h^1, ..., h^N)$ is convex with polyhedral structure of the sets $I^{-1}(k) = \{y | h^k(y) = h(y)\}$ and F is C^2, cf. [Pol90], and to sensitivity analysis, as well, cf. [LR95].

The crucial point for this approach consists in the fact that (not only Clarke's subdifferential) ∂f can be interpreted as a composed map of the subdifferentials to the functions

$$h(y) = \max\ y_k \qquad \text{and } F = (F^1, ..., F^m).$$

Indeed,

$$y^* \in \partial h(y) \Leftrightarrow y^* \in \text{conv}\,\{e^k | y_k = h(y)\} = \text{conv}\,\{Dh^k(y) | k \text{ active at } y\}$$

and

$$x^* \in \partial f(x) \Leftrightarrow x^* = y^* DF(x) := \sum y_i^* DF^i(x) \text{ for some } y^* \in \partial h(F(x)),$$

briefly

$$\partial f(x) = \partial h(F(x)) \circ DF(x).$$

Here, ∂h is the usual subdifferential of a convex function, but $\partial f(x)$ must be a generalized one (since f is not convex). The crucial question then consists in the validity and interpretation of the chain rule

$$\begin{aligned} &C(\partial f)(x, x^*)(u) \\ &= \cup_{y^* \in (\partial h)(F(x))}[C(\partial h)(F(x), y^*)(u)] \circ DF(x) + \partial h(F(x)) \circ D^2F(x)u. \end{aligned}$$

In what follows, we determine

$$C\,\partial f \text{ (for } \partial f = \partial_c f)$$

directly, based on the C^2 property of F and Ekeland's principle only, and show the relations between $C\,\partial_c f$ and the contingent derivative CX of the stationary point map in Chapter 8. This way we obtain, for a particular case, known statements (including the above chain rule).

On the other hand, our proof is self-contained and does not require an extension of the tools by proto-derivatives, epi-convergence and approximate subgradients. In addition, we pay attention to those functions F^k that are active at ξ near x up to order $o(\|\xi - x\|^2)$ and show (cf. Corollary 9.6) that variations of F of this order keep $C(\partial_c f)(x, x^*)$ invariant though the (proper) active index sets $I(\xi)$ may switch around x.

Characterization of $C\,\partial_c f$ for Max-Functions: Special Structure

To determine $C(\partial_c f)(x, x^*)(u)$, several simplifications are possible.

We may assume $x = 0$, $f(0) = 0$ and $F^k(0) = 0\ \forall k$ since other F^k than those with $F^k(0) = 0$ are not maximal for ξ near $x = 0$ and could be deleted. After adding a (symmetric Q) quadratic function

$$G^k(x) = F^k(x) - \langle x^*, x\rangle + \tfrac{1}{2}\langle x, Qx\rangle,$$

and setting $g = \max G^k$, we have

$$\partial_c g(x) = \partial_c f(x) - x^* + Qx$$

and

$$y \in \partial_c g(x) - Qx \Leftrightarrow y + x^* \in \partial_c f(x).$$

The latter follows easily from the definition of $C\,\partial_c g$ and yields

$$u^* \in C(\partial_c f)(0, x^*)(u) \;\Leftrightarrow\; u^* + Qu \in C(\partial_c g)(0,0)(u). \tag{9.32}$$

Thus, we may assume that $x^* = 0$. Finally, taking any Q with $Qu = -u^*$, we must only investigate the situation $0 \in C(\partial_c g)(0,0)(u)$ which characterizes just singularity with respect to upper Lipschitz behavior of the stationary point map $(\partial_c g)^{-1}$ at the origin, cf. §4.2 and §5.1.2 where local minimizer and/or convex function have been regarded.

Therefore, we study the essential case of

$$(x,x^*)=(0,0),\ F^k(0)=f(0)=0,\ u^*=0\in C(\partial_c f)(0,0)(u),\ u\neq 0. \tag{9.33}$$

Let us abbreviate

$$c^k=DF^k(0),\ Q^k=D^2F^k(0),\ K(u)=\{k\in I(0)\,|\,c^k\perp u\}. \tag{9.34}$$

The normal cone of $\partial f(x^0)$ at $0\in\partial f(x^0)$ is denoted by $N^0_{\partial f}$. Clearly,

$$u\in N^0_{\partial_c f}\ \Leftrightarrow\ \langle c^k,u\rangle\le 0\ \ \forall k\in I(0). \tag{9.35}$$

Theorem 9.4 (particular structure of $C\,\partial_c f$ for max-functions). *Under the settings (9.34), it holds* $0\in C(\partial_c f)(0,0)(u)$ *if and only if both the direction u belongs to* $N^0_{\partial_c f}$ *and there are an index set* $\emptyset\neq J\subset K(u)$, *elements* $z^0\in\mathbb{R}^n$ *and* $\alpha,\eta\in\mathbb{R}^m_+$ *with* $\alpha_i+\eta_i=0$ *for* $i\notin J$ *such that, with* $\gamma_k:=\langle c^k,z^0\rangle+\frac{1}{2}\langle u,Q^k u\rangle$,

$$\gamma_i\ge\gamma_k\quad\forall i\in J,\ k\in K(u). \tag{9.36}$$

$$0=\textstyle\sum_J\alpha_i c^i,\quad \sum_J\alpha_i=1, \tag{9.37}$$

$$0=\textstyle\sum_J\alpha_i Q^i u\ +\ \sum_J\eta_i c^i. \tag{9.38}$$

◇

Before proving the theorem some comments are appropriate because the requirements (9.36) and (9.38) permit several equivalent descriptions. These comments basically apply facts from linear optimization.

Remark 9.5 (equivalent conditions). In Theorem 9.4, condition (9.36) may be replaced by

$$y=\alpha \text{ solves the linear problem } (\mathrm{P})_u \text{ with optimal value } 0, \tag{9.39}$$

where

$$(\mathrm{P})_u:\quad \max_y\left\{\textstyle\sum_{K(u)}\langle u,Q^k u\rangle\,y_k\ \middle|\ \begin{array}{l}\sum_{K(u)}y_k c^k=0,\ \sum_{K(u)}y_k=1,\\ y_k\ge 0,\ k\in K(u)\backslash J\end{array}\right\},$$

while (9.38) may be replaced by

$$0\ =\ \min_z\{\,\textstyle\sum_J\langle\alpha_k Q^k u,z\rangle\ |\ \langle c^k,z\rangle\le 0\ \forall k\in J\,\}. \tag{9.40}$$

◇

Proof of Remark 9.5. First note that one may identify $J=\{i|\alpha_i+\eta_i>0\}$ by removing other indices from J. Further, the condition $\eta_i\ge 0$ is not essential for i satisfying $\alpha_i>0$ because it can be satisfied (keeping the rest valid) by forming new η as $\eta + \lambda\alpha$ with large λ.

Condition (9.36): Under the remaining conditions, the real numbers $\gamma_k(z) := \langle c^k, z\rangle + \frac{1}{2}\langle u, Q^k u\rangle$ satisfy for all z,

$$\sum_J \alpha_i\gamma_i(z) = \sum_J \alpha_i\langle c^i, z\rangle + \tfrac{1}{2}\sum_J \alpha_i\langle u, Q^i u\rangle = 0 - \tfrac{1}{2}\sum_J \eta_i\langle c^i, u\rangle = 0.$$

So, (9.36) may be replaced by

$$\gamma_k \le 0 \ \forall k \in K(u), \quad \gamma_i = 0 \ \forall i \in J, \tag{9.41}$$

because the value $\mu = \max_{k\in K(u)} \gamma_k$ in (9.36) fulfills

$$\mu = \sum_J \alpha_i\mu = \sum_J \alpha_i\gamma_i = 0.$$

Condition (9.41) can be written (put $\zeta = 2z$) by

$$0 = \max\{\langle 0, \zeta\rangle \mid \langle c^i, \zeta\rangle = -\langle u, Q^i u\rangle \ \forall i \in J, \ \langle c^k, \zeta\rangle \le -\langle u, Q^k u\rangle \ \forall k \in K(u)\backslash J\}.$$

Hence, by duality of linear programming,

$$0 = \min\{-\sum_K y_k \langle u, Q^k u\rangle \mid \sum_K y_k c^k = 0, \ y_k \ge 0 \ \forall k \in K(u)\backslash J\}. \tag{9.42}$$

Further, having solvability of problem (9.42) the point $y = \alpha$ is a nontrivial solution because feasibility is obvious and (9.38) yields

$$-\sum_{K(u)} \alpha_k \langle u, Q^k u\rangle = -\sum_J \alpha_k \langle u, Q^k u\rangle = \sum_J \eta_k \langle c^k, u\rangle = 0.$$

Thus, instead of (9.36), one may equivalently claim that $y = \alpha$ solves

$$\max_y\{\sum_K y_k \langle u, Q^k u\rangle \mid \sum_K y_k c^k = 0, \ y_{K(u)\backslash J} \ge 0\}$$

with optimal value 0.

Then, the additional constraint $\sum_K y_k = 1$ (satisfied by α) does not change this condition and leads us to (9.39).

Condition (9.38): Again by duality, (9.38) means

$$\begin{aligned} 0 &= \max\{0^T\eta \mid \sum_J \eta_i c^i = -\sum_J \alpha_i Q^i u, \ \eta \ge 0\} \\ &= \min_y\{-\langle \sum_J \alpha_i Q^i u, y\rangle \mid \langle c^i, y\rangle \ge 0 \ \forall i \in J\} \end{aligned}$$

and

$$0 = \min_z\{\langle \sum_J \alpha_i Q^i u, z\rangle \mid \langle c^i, z\rangle \le 0 \ \forall i \in J\},$$

which is (9.40). □

Note that by *parametric* linear programming (via Hoffman's lemma and the global Lipschitz property of the linear objective), (9.40) means exactly that there exists a constant $\Theta > 0$ such that

$$-\Theta\varepsilon \leq \min_z\{\langle \sum_j \alpha_i Q^i u, z\rangle \mid \langle c^i, z\rangle \leq \varepsilon \,\forall i \in J\} \; (\forall \varepsilon \geq 0). \tag{9.43}$$

The latter will be important for the second part of the subsequent proof.

After replacing (9.36) by (9.39) and (9.38) by (9.40), respectively, the explicit variables z^0 and η disappear while α and J are connected by the relations $\emptyset \neq \{i|\alpha_i > 0\} \subset J \subset K(u)$.

Proof of Theorem 9.4. With $u = 0$, the statement becomes trivial, so let $u \neq 0$.

($\Rightarrow$) Let $0 \in C(\partial_c f)(0,0)(u)$. By definition of $C\,\partial_c f$ there are points

$$x = tu + o(t) \text{ and } g = \sum \beta_k DF^k(x) \in \partial_c f(x)$$

such that, for certain $t = t_\nu \downarrow 0$, one has $\beta_k \geq 0, \sum \beta_k = 1, \beta_k = 0$ if $k \notin I(x)$ and, in addition, $w := g/t \to 0$.

Clearly, x, g, w and β depend on t; for seek of simplicity, we avoid to write it explicitly.

(*Notice, in view of the following corollary, that the conclusions in this part of the proof are even valid if* $I(x)$ *contains all k such that* $F^k(x) \geq f(x) - o(t^2)$, *i.e.,* F^k *must be only "approximately active" at* x.)

Selecting an appropriate subsequence, the sets $J(t) = \{k|\beta_k > 0\}$ are constant:

$$J(t) \equiv J \subset I(x) \subset I(0).$$

Further, convergence of the bounded elements β may be assumed, say $\beta \to \alpha$. Next, each w can be written, as

$$\begin{aligned} w &= t^{-1}\sum \beta_k DF^k(x) = w_1 + w_2, \quad \text{where} \\ w_1 &= t^{-1}\sum \beta_k c^k, \\ w_2 &= t^{-1}\sum \beta_k[DF^k(x) - c^k] = t^{-1}\sum \beta_k[Q^k x + s_k(x)]. \end{aligned}$$

Here, $s_k(x)$ is the error of the first-order approximation of DF^k near $x^0 = 0$, so $s_k(x)/t$ vanishes. In consequence, w_2 converges:

$$w_2 \to w_s^* := \sum \alpha_k Q^k u.$$

Due to $w \to 0$, the limit $w_1^* = \lim w_1 = -w_2^*$ exists, too. This implies (9.37):

$$\sum \alpha_k c^k = \lim \sum \beta_k c^k = 0.$$

In addition, for $w_1 \to w_1^*$, there are solutions $y = t^{-1}\beta \geq 0$ of the linear system

$$\sum_{k \in J} y_k c^k = w_1.$$

This yields solvability for the right-hand side w_1^* :

$$\textstyle\sum_{k\in J}\eta_k c^k = w_1^* \text{ holds for certain } \eta_k \geq 0.$$

Therefore, (9.37) and (9.38) are valid.

We derive (9.36). Using second order expansion, we conclude that

$$\begin{aligned} F^k(x) &= \langle c^k, x\rangle + \tfrac{1}{2}\langle x, Q^k x\rangle + r_k(x) \quad (\text{where } r_k(x)/t^2 \to 0) \\ &= t\langle c^k, u\rangle + \langle c^k, o(t)\rangle + \tfrac{1}{2}t^2\langle u + o(t)/t, Q^k(u + o(t)/t)\rangle + r_k(x) \\ &= t\langle c^k, u\rangle + t^2[\,\langle c^k, o(t)/t^2\rangle + \tfrac{1}{2}\langle u, Q^k u\rangle\,] + R_k(x), \end{aligned} \tag{9.44}$$

where again $R_k(x)/t^2 \to 0$.

Let $i \in J$. Since $F^i(x)$ is maximal *(or maximal up to* $o(t^2)$*)* and t vanishes, we obtain $\langle c^i, u\rangle \geq \langle c^k, u\rangle$ for all k. Knowing that $0 \in \operatorname{conv}\{c^i \,|\, i \in J\}$, one has $\langle c^i, u\rangle = 0 \geq \langle c^k, u\rangle$ for all k. So we have $u \in N^0_{\partial f}$ and $\emptyset \neq J \subset K(u)$.

Next, considering any $k \in K(u)$, we obtain by (9.44) for $i \in J$,

$$\langle c^i, o(t)/t^2\rangle + \tfrac{1}{2}\langle u, Q^i u\rangle + R_i(x)/t^2 \;\geq\; \langle c^k, o(t)/t^2\rangle + \tfrac{1}{2}\langle u, Q^k u\rangle + R_k(x)/t^2.$$

Therefore, the linear systems (in ξ)

$$\langle c^i - c^k, \xi\rangle + \tfrac{1}{2}\langle u, (Q^i - Q^k)u\rangle \geq (R_k(x) - R_i(x))/t^2 \quad (i \in J, k \in K(u))$$

have solutions $\xi = o(t)/t^2$ for each t under consideration. The right-hand sides vanish. So also

$$\langle c^i - c^k, \xi\rangle + \tfrac{1}{2}\langle u, (Q^i - Q^k)u\rangle \geq 0 \;\forall i \in J, k \in K(u)$$

remains solvable, and every solution z^0 fulfills (9.36).

($\Leftarrow$) If the assertion does not hold then, by formal negation, there exist positive p_0 and q_0 such that

$$\|x^*\| \geq p_0 t \;\forall x^* \in \partial_c f(tu + tq_0 B), \tag{9.45}$$

whenever $t > 0$ is small enough. By the well-known relations between $f', \partial_c f$ and Clarke's directional derivative f^c, cf. [Cla83] and Chapter 6, it holds

$$f'(x; v) = f^c(x; v) = \max_{x^* \in \partial_c f(x)} \langle x^*, v\rangle.$$

Selecting $x_x^* \in \partial_c f(x)$ with minimal Euclidean norm and setting $v_x = -x_x^*/\|x_x^*\|$, this yields for $x \in tu + tq_0 B$,

$$f^c(x; v_x) = \max_{x^* \in \partial_c f(x)} \langle x^*, v_x\rangle = -\|x_x^*\| \leq -p_0 t.$$

Therefore, one obtains from (9.45),

$$f'(x; v) \leq -p_0 t \text{ for all } x \in tu + tq_0 B \text{ and certain } v = v_x \in \operatorname{bd} B. \tag{9.46}$$

We will show that (9.46) cannot hold. Let z^0, J and α satisfy the conditions of the theorem. We fix $\delta > \|z^0\|$ and put

$$\begin{array}{lcl} C_0 & > & \max\{\|c^v\| \mid 1 \le v \le m\}, \\ C_1 & = & \max\{\langle c^k, u\rangle \mid k \in I(0), \langle c^k, u\rangle < 0\} \text{ (only important if such } k \text{ exists)} \\ C_2 & > & \max\{\|Q^k\| \mid k \in K(u)\}. \end{array}$$

Let $q \in (0, 1)$ be sufficiently small such that, if C_1 is defined, $C_1 + q\delta C_0 < \frac{1}{2}C_1$ holds true. For any $z \in \delta B$, we put

$$x_{st} = tu + sz, \qquad |s| \le qt \qquad (t > 0 \text{ small })$$

and apply

$$\begin{array}{rcl} F^k(x_{st}) & = & t\langle c^k, u\rangle + s\langle c^k, z\rangle + \frac{1}{2}t^2\langle u, Q^k u\rangle \\ & & +st\langle z, Q^k u\rangle + \frac{1}{2}s^2\langle zQ^k z\rangle + r_k(x_{st}), \end{array} \tag{9.47}$$

where $|r_k(x_{st})| \le \frac{1}{2}O(t)^2$. If $\langle c^k, u\rangle < 0$ then

$$t\langle c^k, u\rangle + s\langle c^k, z\rangle \le tC_1 + tq\delta C_0 < \frac{1}{2}tC_1 < 0$$

implies that F^k (being smaller than F^i for $i \in J$) is not active for sufficiently small positive $t < t(C_1, C_2)$. So, in what follows, we must only regard $k \in K(u)$.

First we intend to show that, for sufficiently small $t > 0$,

$$f(x_{st}) \ge -Cq^2t^2 \text{ holds true with } C = 2\delta^2 C_2. \tag{9.48}$$

Due to $\sum_J \alpha_k c^k = \sum_J \alpha_k \langle u, Q^k u\rangle = 0$ (from (9.37), (9.38)), it holds

$$\begin{array}{rcl} f(x_{st}) & \ge & \max_{k\in J} F^k(x_{st}) \\ & \ge & \sum_J \alpha_k F^k(x_{st}) \\ & = & st\sum_J \alpha_k\langle z, Q^k u\rangle + \frac{1}{2}s^2\sum_J \alpha_k\langle z, Q^k z\rangle + \sum_J \alpha_k r_k(x_{st}) \\ & \ge & st\sum_J \alpha_k\langle z, Q^k u\rangle - \frac{1}{2}t^2(q\delta)^2 C_2 - \frac{1}{2}O(t)t^2. \end{array} \tag{9.49}$$

Again for small $t < t(C_1, C_2, q)$, we have $O(t) < (q\delta)^2 C_2$. Hence the crucial inequality

$$f(x_{st}) \ge st\sum_J \alpha_k\langle z, Q^k u\rangle - t^2(q\delta)^2 C_2 \tag{9.50}$$

is valid. We are now able to apply (9.50) for proving (9.48).

CASE 1
If s/t is small enough, namely if $s\,\delta\|u\|C_2 < t(q\delta)^2 C_2$, then (9.50) yields

$$|s\, t\sum_J \alpha_k\langle z, Q^k u\rangle| < t^2(q\delta)^2 C_2 \quad \text{and} \quad f(x_{st}) \ge -2t^2(q\delta)^2 C_2.$$

The latter is (9.48). So let the opposite hold,

$$s/t \geq q' := (q\delta)^2 \, \delta^{-1} \, \|u\|^{-1},$$

and put $\varepsilon(z) = \max_j c^i z$ for the current s, t and z.

CASE 2
If $\varepsilon(z)$ is small, namely if $\Theta\varepsilon(z) < q\delta^2 C_2$, we may use (9.43), i.e.,

$$\textstyle\sum_{k \in J} \alpha_k \langle z, Q^k u\rangle \geq -\Theta\varepsilon,$$

in order to obtain again

$$\begin{aligned} f(x_{st}) &\geq -st\Theta\varepsilon(z) - t^2(q\delta)^2 C_2 \geq -qt^2\Theta\varepsilon(z) - t^2(q\delta)^2 C_2 \\ &\geq -q^2 t^2 \delta^2 C_2 - t^2(q\delta)^2 C_2 = -2t^2(q\delta)^2 C_2. \end{aligned}$$

CASE 3
Finally, if $\varepsilon(z) = c^i \, z = \sigma \geq \Theta^{-1} q\delta^2 C_2$ for some $i \in J$, we estimate by the help of (9.47) and $s/t \geq q'$ in order to deduce (for small t not depending on z),

$$\begin{aligned} f(x_{st}) &\geq F^i(x_{st}) \geq s(\sigma + t\langle z, Q^i u\rangle) + \tfrac{1}{2}t^2 \langle u, Q^i u\rangle - t^2(q\delta)^2 C_2 \\ &\geq q' t(\sigma + t\langle z Q^i u\rangle) + \tfrac{1}{2}t^2 \langle u, Q^i u\rangle - t^2(q\delta)^2 C_2 > 0. \end{aligned}$$

Summarizing, (9.48) is true, provided that t is sufficiently small (depending on q only).

Next, setting $s = t^2$ and $z = z^0$, (9.47) ensures for all $k \in K(u)$

$$\begin{aligned} F^k(x_{st}) &= t^2(\langle c^k, z^0\rangle + \tfrac{1}{2}t^2\langle u, Q^k u\rangle) + t^3\langle z^0, Q^k u\rangle + \tfrac{1}{2}t^4\langle z^0, Q^k z^0\rangle + r_k(x_{st}) \\ &\leq t^3\langle z^0, Q^k u\rangle + \tfrac{1}{2}t^4\langle z^0, Q^k z^0\rangle + r_k(x_{st}) \\ &= R_k(t)t^2 \quad \text{with } R_k(t) \to 0. \end{aligned}$$

So, for the particular points $x(t) = x_{t^2 t}$, we have $f(x(t)) \leq R(t)t^2$, $R(t) \downarrow 0$.

It remains to apply Ekeland's principle for f on

$$X(t) = tu + tq\delta B.$$

Using (9.48) and $R(t)t^2 \leq Cq^2t^2$ for small t, the point $x(t) \in X(t)$ is ε-optimal for f on $X(t)$ with $\varepsilon = 2C\, q^2 \, t^2$ for small $t > 0$. Let $z(t) \in X(t)$ be a related Ekeland-point with $\alpha = \frac{1}{2}tq\delta$. This ensures

$$f(\xi) + (\varepsilon/\alpha)d(\xi, z(t)) \geq f(z(t)) \ \forall \xi \in X(t) \text{ and } d(z(t), x(t)) \leq \alpha.$$

Clearly, now $\varepsilon/\alpha = 4C\,\delta^{-1}\, q\, t$ vanishes (as $t \downarrow 0$).
Moreover, by the construction of $x(t)$ and α, the point $z(t)$ belongs to the interior of $X(t)$ for small t. So, the Ekeland- inequality yields necessarily

$$f'(z(t); v) \geq -\varepsilon/\alpha \qquad \text{for all } v \in \operatorname{bd} B.$$

With small q, such that $4C\,\delta^{-1}\, q\, t < p_0\, t$, this contradicts (9.46) and proves the assertion. □

Corollary 9.6 (approximations of high order). *Suppose, for any index set J and* $x = tu + o(t), t = t_v \downarrow 0,$ *that*

$$F^k(x) \geq f(x) - o(t^2) \;\forall k \in J \quad \textit{and} \quad 0 = \lim t^{-1} \text{dist}\,(0, \text{conv}\,\{DF^k(x)/k \in J\}).$$

Then $0 \in C(\partial_c f)(0,0)(u)$, *and the conditions of Theorem 9.4 are fulfilled with the given set J (though these* F^k *are not active at* x *in general).* ◇

Proof. By the first part $(\Rightarrow)$ of the proof to Theorem 9.4 we obtain the related conditions, the second one $(\Leftarrow)$ verifies $0 \in C(\partial_c f)(0,0)(u)$. □

Note. The part $(\Leftarrow)$ of the proof indicates (via $q = q(t) \downarrow 0$) that, for every given sequence $t_v \downarrow 0$, it holds $0 = \lim t_v^{-1} \text{dist}\,(0, \partial_c f(z^v))$ with certain $z^v = t_v w^v, w^v \to u$.

This means that if $0 \in C(\partial_c f)(0,0)(u)$, i.e., by definition,

$$0 \in \limsup_{t \downarrow 0} \limsup_{w \to u} t^{-1} \;\text{dist}\,(0, \partial_c f(tw)),$$

then even

$$0 \in \liminf_{t \downarrow 0} \limsup_{w \to u} t^{-1} \;\text{dist}\;(0, \partial_c f(t, w)).$$

is valid.

Mappings F possessing a contingent derivative that satisfies (like $\partial_c f$) this limsup = lim inf-equation are introduced in [Roc88] as being *proto-differentiable* multifunctions. This property is a multivalued version of "simple" in §6.4.1 and turns out to be similarly useful for establishing chain rules (because of the same technical reasons as for functions). ◇

Characterization of $C\,\partial_c f$ for Max-Functions: General Structure

Theorem 9.7 (general structure of $C\partial_c f$). *It holds* $u^* \in C(\partial_c f)(0, x^*)(u)$ *if and only if there exist* $\alpha, \eta \in \mathbb{R}^m_+$ *and* $z \in \mathbb{R}^n$ *such that* $\emptyset \neq J := \{i/\alpha_i + \eta_i > 0\} \subset I(0)$ *and*

(i) $\langle DF^k(0) - x^*, u\rangle \leq 0 \quad \forall k \in I(0),$

(ii) $\langle DF^k(0) - x^*, u\rangle = 0 \quad \forall k \in J,$

(iii) $\langle DF^k(0) - x^*, z\rangle + \langle u, D^2F^k(0)u\rangle \leq \langle u^*, u\rangle$ *if* $\langle DF^k(0) - x^*, u\rangle = 0$ *and* $k \in I(0)$, *and equation holds for* $k \in J$,

(iv) $x^* = \sum_J \alpha_i DF^i(0), \quad \sum_J \alpha_i = 1,$

(v) $u^* = \sum_J \alpha_i D^2F^i(0)u + \sum_J \eta_i (DF^i(0) - x^*).$

◇

Remark 9.8 (equivalent conditions). In the theorem, condition (iii) may be replaced by

$$y = \alpha \text{ solves the linear problem } (\text{LP})_u \text{ with optimal value } \langle u^*, u\rangle, \qquad (9.51)$$

where

$$(\text{LP})_u: \quad \max_y \left\{ \sum_K \langle u, D^2F^k(0)u\rangle\, y_k \;\middle|\; \begin{array}{l} \sum_K y_k(DF^k(0) - x^*) = 0, \\ \sum_K y_k = 1,\; y_{K\setminus J} \geq 0 \end{array} \right\}$$

and

$$K = \{k \in I(0) \mid \langle DF^k(0) - x^*, u\rangle = 0\}.$$

Condition (v) may be replaced by

$$0 = \min_z \{\langle \textstyle\sum_J \alpha_i (D^2F^i(0)u - u^*), z\rangle \mid \langle DF^i(0) - x^*, z\rangle \leq 0 \;\forall i \in J\}. \quad (9.52)$$

◇

Proof of Theorem 9.7. For characterizing $u^* \in C(\partial_c f)(0, x^*)(u)$ generally, we apply the transformations between f and g in (9.32):

$$u^* \in C(\partial_c f)(0, x^*)(u) \;\Leftrightarrow\; u^* + Qu \in C(\partial_c g)(0,0)(u) \text{ with } Qu = -u^*.$$

The quantities of Theorem 9.4 have now the form

$$DG^k(0) = DF^k(0) - x^* = c^k,\; D^2G^k(0) = D^2\, F^k(0) + Q = Q^k$$

Thus, the related conditions for $u^* \in C(\partial_c f)(0, x^*)(u)$ are as follows:

$$u \in N_{\partial_c f(0)}(x^*),\; \emptyset \neq J \subset K(u) = \{k \in I(0) \mid \langle DF^k(0) - x^*, u\rangle = 0\},$$

the real numbers $\gamma_k = \langle DF^k(0) - x^*, z^0\rangle + \frac{1}{2}\langle u, (D^2F^k(0) + Q)u\rangle$ fulfill

$$\gamma_i = 0 \geq \gamma_k \text{ for all } i \in J, k \in K(u),$$

$$0 = \textstyle\sum_J \alpha_i (DF^i(0) - x^*), \quad \sum_J \alpha_i = 1,$$

$$0 = \textstyle\sum_J \alpha_i (D^2F^i(0) + Q)u + \sum_J \eta_i (DF^i(0) - x^*).$$

These conditions are independent of the choice of Q in the equation $Qu = -u^*$.

Condition (9.38) becomes

$$u^* = \textstyle\sum_J \alpha_i D^2F^i(0)u + \sum_J \eta_i (DF^i(0) - x^*)$$

and γ_k in (9.36) attains the form

$$\gamma_k = \langle DF^k(0) - x^*, z\rangle + \tfrac{1}{2}\langle u, D^2F^k(0)u\rangle - \tfrac{1}{2}\langle u^*, u\rangle$$

Substituting finally z by $2z$, and using Remark 9.5, the theorem is verified. □

Application 1

Let us compare the conditions of Theorem 9.7 with those which describe the contingent derivative CX of the stationary point map X for the C^2-problem

$$\min\{z \mid x \in \mathbb{R}^n, f^i(x) - z \le 0 \ \forall i\} \qquad (z = f(x) = \max f^i(x)) \tag{9.53}$$

in Section 8.2. The map $X(a_0, a, b)$ is defined by the stationary points (z, x) of the (canonically perturbed) parametric problem

$$\min_{(x,z)}\{z - a_0 z - \langle a, x\rangle \mid f^i(x) - z \le b_i, \ i = 1, \ldots, m\} \tag{9.54}$$

where $a = (a_1, ... a_n)$. Since MFCQ holds everywhere for (9.53), we know that, in terms of the related Kojima function F and with $(u_0, u) \in \mathbb{R}^{1+n}$,

$$\begin{aligned}&(u_0, u) \in CX((0,0),(z^0, x^0))(\pi)\\ &\Leftrightarrow \exists\, y^0 \in Y^0 : \pi \in CF((z^0, x^0), y^0)((u_0, u), \mathbb{R}^m) \subset \mathbb{R}^{1+n+m}.\end{aligned}$$

Using the explicit form of CF (see Theorem 7.6), direction π is characterized by (we write here (p, q) instead of (α, β)),

$$\begin{aligned}\pi_1 &= D^2L(z^0, x^0, y^0)(u_0, u) + \textstyle\sum_i p_i(-1, Df^i(x^0))\\ \pi_{2i} &= -u_0 + Df^i(x^0)u \ - q_i\end{aligned}$$

for some $(p, q) \in \mathcal{J}_C(y^0)$ (for the definition see (7.32)), and the set Y^0 of dual solutions for (z^0, x^0) has the form

$$y \in Y^0 \ \Leftrightarrow \ \sum_i y_i^+ Df^i(x^0)) = 0, \ \sum_i y_i^+ = 1 \text{ and } f^i(x^0) - z^0 = y_i^-.$$

Here, D^2L denotes the second derivative of $L = z + \sum y_i^+(f^i - z)$ with respect to (z^0, x^0),

$$D^2L(z^0, x^0, y)(u_0, u) = (0, \sum_i y_i^+ D^2 f^i(x^0)u) \in \mathbb{R}^{1+n} \ \text{ for all } u_0.$$

To model the variations of f that correspond to $0 \in C(\partial_c f)(0,0)(u)$, we put

$$a_0 \equiv 0, b \equiv 0, \ x^* = 0, x_0^* = 0, \ \pi = (0, u^*), \ u^* = 0 \ \text{ and } \ \pi_2 = 0.$$

Then,

$$u_0 = Df^i(x^0)u - q_i \ \text{ and } \ 0 = \sum_i p_i$$

follow. Therefore,

$$u_0 = Df^i(x^0)u = 0 \ \text{ for all } i \text{ with } y_i > 0 \text{ and (in consequence) } u_0 = 0$$

follow, too.

Corollary 9.9 (reformulation 1). *For the stationary point map X of the C^2 program (9.53), the elements $(u_0, u) \in CX((0,0),(z^0,x^0))(0_1,u^*,0_m)$ are exactly characterized by the conditions of Theorem 9.7, except for condition (iii).* ◇

Proof. Indeed, the explicit formula for CX attains the form

$$\begin{aligned}&(u_0,u) \in CX((0,0),(z^0,x^0))(0_1,u^*,0_m)\\ \Leftrightarrow\ & u_0 = 0 \text{ and there exist } y \in Y^0 \text{ and } (p,q) \in \mathcal{J}_C(y) \text{ satisfying}\\ & (0=)\, u^* = \textstyle\sum_i y_i^+ D^2 f^i(x^0)u + \sum_i p_i Df^i(x^0) \text{ and } Df^i(x^0)u = q_i.\end{aligned}$$

So, the point (α,η) in Theorem 9.7 and the current vector (y^+,p) may be identified, while J becomes the set $J = \{i \mid y_i^+ + p_i > 0\}$. Here, we *do not* require $p_i \geq 0\ \forall i$ because $(p,q) \in \mathcal{J}_C(y)$ permits $p_i \in \mathbb{R}$ if $y_i^+ > 0$. However, as already noted after Theorem 9.4, the condition $\eta_i \geq 0$ is not essential for those i satisfying $\alpha_i > 0$. □

The absence of condition (iii) arises from different variations of the same problem (9.54).

The relation $u^* \in C(\partial_c f)(0,x^*)(u)$ describes the existence of stationary points (z,x) to problem (9.54) with

$$\begin{aligned}&(z,x) = (z^0,x^0) + t(0,u) + o(t)\\ &\text{for variations } a_0 \equiv 0, a = tu^* + o_1(t) \text{ and } b \equiv 0.\end{aligned}$$

The relation $(u_0,u) \in CX((0,0),(z^0,x^0))(0_1,u^*,0_m)$ describes the existence of stationary points (z,x) to problem (9.54) with

$$\begin{aligned}&(z,x) = (z^0,x^0) + t(0,u) + o(t)\\ &\text{for variations } (a_0,a) = t(0,u^*) + o_1(t) \text{ and } b = o_2(t).\end{aligned}$$

Here, the replacement of a_0 by $a_0 \equiv 0$ does not change the stationary points (z,x). However, the original constraints

$$f^i(x) - z \leq b_i \equiv 0$$

must be satisfied up to error $o_2(t)$ only. Thus, in comparison to $b \equiv 0$, the possible sets of active constraints may increase, and new stationary points may occur.

However, the requirements $b = o(t^2)$ and $b \equiv 0$ are equivalent in accordance with Corollary 9.6.

Application 2

Let X_1 be the mapping $a \mapsto X_1(a)$ given by the stationary points for

$$\min\{f(x) - \langle a,x\rangle \mid g(x) \leq 0\},\quad (f,g) \in C^2(\mathbb{R}^n,\mathbb{R}^{1+m}),$$

and let $F = MN$ be the associated Kojima function. Consider the contingent derivative

$$CX_1(0, x^0)(\pi_1)$$

at some stationary point $x^0 \in X_1(0)$, and suppose that

$$x^0 \text{ satisfies MFCQ}.$$

To simplify the calculations, we further suppose that

$$\begin{array}{l} g(x^0) = 0, \text{ and } f \text{ has small } C^1\text{-norm such that (by MFCQ) any} \\ \text{possible Lagrange multiplier vector}(\gamma_1, ..., \gamma_m) \text{ for } (x, a) \text{ near } (x^0, 0) \\ \text{fulfills } \sum \gamma_i < 1 \text{ (otherwise replace } f \text{ by } \theta f \text{ with some small } \theta > 0). \end{array} \tag{9.55}$$

In order to establish the correspondence between $C\partial_c f$ and the contingent derivative of Kojima's function, introduce the functions

$$g_0 \equiv 0 \text{ and } G(x) = \max\{g_i(x) | 0 \le i \le m\}.$$

Then one may write

$$x \in X_1(a) \Leftrightarrow -(Df(x) - a) \in \partial_c G(x) \text{ and } G(x) = 0.$$

Here,

$$\partial_c G(x) = \text{conv}\{Dg_i(x) | i \in I(x)\}, \quad I(x) = \{i | g_i(x) = G(x)\}.$$

If $-(Df(x) - a) = \sum_{I(x)} \gamma_i Dg_i(x)$ holds with weights $\gamma_0, ..., \gamma_m$ (put $\gamma_i = 0$ if $i \notin I(x)$), then $(\gamma_1, ..., \gamma_m)$ is a Lagrange multiplier for (x, a). This yields:

$$\begin{array}{ll} & u \in CX_1(0, x^0)(\pi_1) \\ \Leftrightarrow & \text{certain } x(t) = x^0 + tu + o_1(t) \text{ fulfill, for } t = t_v \downarrow 0, \; x(t) \in X_1(t\pi_1 + o_2(t)) \\ \Leftrightarrow & -[Df(x(t)) - t\pi_1 - o_2(t)] \in \partial_c G(x(t)) \text{ and } G(x(t)) = 0 \\ \Leftrightarrow & -Df(x^0) - t(D^2 f(x)u - \pi_1) + o(t) \in \partial_c G(x(t)) \text{ and } G(x(t)) = 0. \end{array}$$

Thus, setting $x^* = -Df(x^0)$ and $u^* = -[D^2 f(x^0)u - \pi_1]$, one has $u^* \in C(\partial_c G)(x^0, x^*)(u)$.

Corollary 9.10 (reformulation 2). *Under assumption (9.55), the application of Theorem 9.7 to* $u^* \in C(\partial_c G)(x^0, x^*)(u)$ *leads, via the transformation*

$$\mu = \sum\nolimits_J \eta_k, \; y_i^0 = \alpha_i, \; p_i = \eta_i + \mu\alpha_i, \; q_i = Dg_i(x^0)u \;\; (i > 0),$$

to a particular solution (u, p, q) *of the system*

$$\begin{array}{ll} (\pi_1, 0) = [DM(x^0)u]N(y^0) + M(x^0)(0, p, q)^T & \\ (0, p, q) \in N'(y^0)(\mathbb{R}^m), & F(x^0, y^0) = 0 \end{array}$$

such that (y^0, u, p, q) *in addition satisfies* $p \geq 0$ *and*

$$\begin{array}{l} \langle u, D^2_{xx}L(x^0, y)u\rangle \leq \langle u, D^2_{xx}L(x^0, y^0)u\rangle = \langle u, \pi_1\rangle \\ \textit{if } D_xL(x^0, y) = 0 \textit{ and } y_{K'\setminus J'} \geq 0, \end{array} \tag{9.56}$$

with $L = f + \sum_{i>0} y_i g_i$ *(not with* y_i^+*),* $K' = \{i | q_i = 0\}$, *and* $J' = \{i | p_i > 0\}$. ◇

Proof. Note that $I(x^0) = \{0, ..., m\}$ since $g(x^0) = 0$. By Theorem 9.7, we have with

$$c^k = Dg^k(x^0) - x^* \text{ and } K = \{k |\ c^k \perp u\} \subset \{0, ..., m\}$$

that $\langle c^v, u\rangle \leq 0\ \forall v$ and $\exists \alpha, \eta \in \mathbb{R}^{1+m}_+$ such that

$$\emptyset \neq J := \{i | \alpha_i + \eta_i > 0\} \subset K \text{ and}$$

$y = \alpha$ solves

$$\max_y \{\textstyle\sum_K \langle u, D^2 g^k(x^0)u\rangle y_k \mid \sum_K y_k c^k = 0, \sum_K y_k = 1, y_{K\setminus J} \geq 0\} \tag{9.57}$$

with optimal value $\langle u^*, u\rangle = -\langle D^2 f(x^0)u - \pi_1, u\rangle$, and one also has

$$x^* = \textstyle\sum_J \alpha_i Dg^i(x^0), \tag{9,58}$$

$$u^* = \textstyle\sum_J \alpha_i D^2 g^i(x^0)u \ + \ \sum_J \eta_i\, c^i. \tag{9.59}$$

By (9.58), $(\alpha_1, ... \alpha_m)$ is a Lagrange multiplier vector for $(x^0, 0)$. Since $\sum_{i>0} \alpha_i < 1$, we obtain from (9.58) that $\alpha_0 > 0$ and $0 \in J$. This yields, due to $c^0 \perp u$,

$$\langle Df(x^0), u\rangle = \langle x^*,\ u\rangle = 0, \langle Dg_v(x^0), u\rangle \leq 0\ \forall v,\ \langle Dg_k(x^0), u\rangle = 0\ \forall k \in K.$$

Because of $u^T D^2 f(x^0)u = \sum_K y_k \langle D^2 f(x^0)u, u\rangle$, the conditions (9.57) and

$$\begin{array}{l} \langle u, D^2_{xx}L(x^0, y)u\rangle \leq \langle u, D^2_{xx}L(x^0, y^0)u\rangle = \langle u, \pi_1\rangle \\ \text{if } \sum_K y_k c^k = 0, \sum_K y_k = 1, y_{K\setminus J} \geq 0 \end{array}$$

coincide. Using $c^k = Dg^k(x^0) + Df(x^0), g_0 \equiv 0$ and noting that $\eta_i + \alpha_i > 0$ iff $p_i > 0$, the latter is equivalent to (9.56).

The subset of non-negative feasible points y in (9.56) is just the set of Lagrange multiplier vectors for $(x^0, 0)$ (without component y_0). Since $\mu = \sum_J \eta_k$, we also have

$$\textstyle\sum_J \eta_i (Dg^i(x^0) - x^*) = \sum_J \eta_i Dg^i(x^0) + \mu \sum_J \alpha_i Dg^i(x^0) = \sum_J (\eta_i + \mu\alpha_i) Dg^i(x^0).$$

So (9.59) becomes:

$$\begin{array}{rcll} \pi_1 & = & D^2 f(x^0)u + \sum_J \alpha_i D^2 g^i(x^0)u & + \ \sum_J \eta_i (Dg^i(x^0) - x^*), \\ & = & D^2_{xx}L(x^0, y^0)u & + \ \sum_J (\eta_i + \mu\alpha_i) Dg^i(x^0). \end{array}$$

The latter verifies the assertion. □

Chapter 10

Newton's Method for Lipschitz Equations

For computing a zero of a locally Lipschitz function $h : X \to Y$, several Newton-type methods have been developed and investigated (from the theoretical and practical point of view as well) during the last 20 years. They have been applied to variational inequalities, generalized equations, Karush-Kuhn-Tucker systems or nonlinear complementarity problems, see, e.g., [KS87, Kum88b, HP90, Pan90, IK92, Kum92, PQ93, QS93, Rob94, Don96, BF97, Fis97, KYF97]. Accordingly, one finds various conditions for convergence of nonsmooth Newton methods (mainly written in terms of semismoothness) and many reformulations of identical problems by means of different equations. Especially for complementarity problems, a big number of so-called NCP functions have been applied in order to obtain such a description cf. [SQ99].

In this chapter, we elaborate those properties of h and related derivatives which are necessary and sufficient for solving $h = 0$ by a Newton method, and we compare the imposed assumptions in terms of the original data.

Before going into the details we suggest the reader to study Example BE.1, which indicates that Newton methods cannot be applied to the class of all locally Lipschitz functions, even if $X = Y = \mathbb{R}$ (provided that the Newton steps have the usual form at C^1-points of the given function). We also mention the well-known real function $h(x) = x^q$ for fixed $q \in (0, 1)$ which shows the difficulties if h is everywhere locally C^1 excepted the origin, and if h is not locally Lipschitz.

10.1 Linear Auxiliary Problems

Newton's method for computing a zero x^* of $h : X \to Y$ (Banach spaces) is determined by the iterations

$$x^{k+1} = x^k - A^{-1}h(x^k),$$

where $A = A_k = Dh(x^k)$ is supposed to be invertible.

The local superlinear convergence of this method means that, for some o-type function r and x^0 near x^*, we have

$$x^{k+1} - x^* = z^k \text{ with } \|z^k\| \leq r(x^k - x^*), \tag{10.1}$$

which is, after substituting x^{k+1} and applying A_k to both sides,

$$A_k[(x^k - x^*) - z^k] = h(x^k) - h(x^*), \text{ with } \|z^k\| \leq r(x^k - x^*). \tag{10.2}$$

The equivalence between (10.1) and (10.2) is still true if one defines, in a more general way,

$$x^{k+1} = x^k - A^{-1}h(x^k), \text{ with some } A \in M(x^k) \tag{10.3}$$

where $M(x^k) \neq \emptyset$ is a given set of invertible linear maps. A method of this type is often said to be a *generalized Newton method.*

The elements x^{k+1} in (10.3) and z^k in (10.2) now depend on the selected elements A. So we have to precise that the inequality in (10.1) should hold *independently* of the choice of $A \in M(x^k)$.

Next suppose that there are constants K^+ and K^- such that

$$\|A\| \leq K^+ \text{ and } \|A^{-1}\| \leq K^- \text{ for all } A \in M(x^* + u) \text{ and small } \|u\|. \tag{10.4}$$

Omitting the indices and setting $u = x - x^*$, $A \in M(x)$, the convergence condition (10.2) now attains the equivalent form

$$Au = h(x) - h(x^*) + Az, \ \|z\| \leq r(u) \tag{10.5}$$

and yields necessarily, with $o(u) = K^+r(u)$,

$$Au \in h(x^* + u) - h(x^*) + o(u)B \text{ for all } A \in M(x^* + u). \tag{10.6}$$

Conversely, having (10.6), i.e.,

$$Au = h(x^* + u) - h(x^*) + v \text{ for some } v \in o(u)B,$$

then, via $x^{k+1} - x^* = z = A^{-1}v \in K^-o(u)B$ and (10.5), one obtains the convergence

$$\begin{array}{l} \|x^{k+1} - x^*\| \leq K^-o(x^k - x^*) \\ \text{for all initial points } x^0 \text{ sufficiently close to } x^*. \end{array} \tag{10.7}$$

But (10.6) is condition (6.24) in §6.4.2: M has to be a *Newton map of* h at x^*. So we have shown

Lemma 10.1 (convergence of Newton's method - I). *Suppose the regularity condition (10.4). Then, the method (10.3) fulfills the convergence condition (10.7) if and only if M satisfies (10.6). The latter means that M is a* Newton map of h at x^*. ◇

Remark 10.2 If the conditions (10.4) and (10.6) hold true with

$$o(u) \le p\|u\|^2 \text{ for all } \|u\| \le \delta,$$

then the method converges for all initial points x^0 satisfying

$$\|x^0 - x^*\| = q \text{ with } q < q_o := \min\{\delta, (K^- p)^{-1}\}$$

since, by induction arguments,

$$\|x^{k+1} - x^*\| \le K^- o(x^k - x^*) \le K^- p\, q\|x^k - x^*\| \text{ and } K^- p\, q < 1.$$

◇

In the current context, the function $h : X \to Y$ may be arbitrary (even for normed spaces X, Y) as long as $M(x^k)$ consists of linear bijections between X and Y. Nevertheless, we will suppose that

$$h \text{ is locally Lipschitz near } x^*.$$

This is justified by two reasons:

(i) If h is only continuous, we cannot suggest any practically relevant definition for $M(x)$.

(ii) Having uniformly bounded $\|A\| \le K^+$ and writing $x = x^k$, (10.2) implies that h *satisfies a pointwise Lipschitz condition* at x^*, namely

$$\|h(x) - h(x^*)\| \le 2K^+\|x - x^*\| \quad \text{for } \|x - x^*\| < \varepsilon, \tag{10.8}$$

if ε is small enough such that $\|r(u)\| < \frac{1}{2}\|u\|$ for $\|u\| < \varepsilon$. Since the solution x^* is unknown, our assumptions should hold for all x^* near the solution. Then, the local Lipschitz property of h (near the solution) follows necessarily from (10.8).

Further, having *uniformly bounded* $\|A^{-1}\| \le K^-$, now (10.2) guarantees after applying A^{-1}, $x - x^* - z = A^{-1}(h(x) - h(x^*))$ and

$$\frac{1}{2}\|x - x^*\| \le \|x - x^* - z\| \le K^-\|h(x) - h(x^*)\| \text{ if } \|x - x^*\| < \varepsilon.$$

Therefore,

$$\|h(x) - h(x^*)\| \ge \tfrac{1}{2}(K^-)^{-1}\|x - x^*\| \quad \text{for } \|x - x^*\| < \varepsilon, \tag{10.9}$$

restricts h again in a canonical manner: h^{-1} is locally upper Lipschitz at $(0, x^*)$.

Condition (10.6) can be met in various versions in the literature. Let h be locally Lipschitz with rank L near x^* and $X = Y = \mathbb{R}^n$:

If Clarke's generalized Jacobian $M(x) = \partial h(x)$ is a Newton map at x^*, then h is called *semismooth* at x^*, sometimes – if $o(\cdot)$ is even quadratic – also *strongly semismooth.*

In several papers [PQ93], [QS93], [Fis97], semismoothness has been defined in a slightly different way by using directional derivatives $h'(x^*;u)$ at the place of $h(x^*+u) - h(x^*)$,

$$Au \in h'(x^*;u) + o(u)B \qquad \forall A \in \partial h(x^*+u). \tag{10.10}$$

For $M \subset \partial h$, condition (10.6) now follows from (10.10) due to the uniform approximation of h by directional derivatives (see Lemma A1):

$$h'(x^*;u) \in h(x^*+u) - h(x^*) + o(u)B.$$

In others papers, M is a mapping that approximates ∂h; and functions h satisfying the related conditions (10.6) are called weakly semismooth. However, neither the relation between M and ∂h nor the existence of $h'(x^*;\cdot)$ is essential for the interplay of the conditions (10.4), (10.6), (10.7) in accordance with Lemma 10.1. The main problem consists in the characterization of those functions which allow us to *find a practically* relevant Newton map. These function classes are not very big up to now. The biggest class of pseudo-smooth functions for which nonsmooth Newton methods have been successfully *applied* is -at least up to now and by our knowledge- the class of composed locally PC^1**-functions,** cf. §6.4.2.

10.1.1 Dense Subsets and Approximations of M

For $X = Y = \mathbb{R}^n$, the conditions (10.4) and (10.6) define M not uniquely and must hold on a dense subset of a neighborhood of x^* only.

Indeed, let $X = Y = \mathbb{R}^n$, and assume (without loss of generality) that the function $o(u)$ under consideration is upper semi-continuous. Then, in order to fulfill (10.4) and (10.6), it suffices to know some M such that (10.4) and (10.6) hold for all $x = x^* + u$ in a dense *subset* S of a neighborhood Ω of x^*. Having this, one may define

$$M_o(x) = \limsup_{s \to x, s \in S} M(s) \text{ for } x \in \Omega \backslash S \text{ as Hausdorff - limit}$$

and

$$M_o(x) = M(x) \text{ for } x \in S.$$

The map M_o has non-empty ranges (for this reason, finite dimension was required), and M_o fulfills (10.4) and (10.6) on Ω by continuity arguments.

Further, if M satisfies (10.4) and (10.6), then (10.4) holds for each M' with $\emptyset \neq M' \subset M$; and (10.6) holds for each map M' with $\emptyset \neq M' \subset \text{conv } M$.

Finally, one may replace M satisfying (10.4) and (10.6) by another map M' as far as

$$\emptyset \neq M'(x) \subset M(x) + O(x - x^*)B_{L(X,X)} \quad (B_{L(X,Y)} = \text{unit ball in } L(X,Y)).$$

In particular, consider

$$N(x) \subset M(x) + \|h(x)\|B_{L(X,Y)} \tag{10.11}$$

which permits to approximate the elements of $M(x)$ with accuracy $\|h(x)\|$.

Remark 10.3 When using N instead of M, condition (10.4) is still satisfied with each $K_N^- > K^-$. The function $o(\cdot)$ in (10.6) changes only by $L\|.\|^2$. Thus, the replacement (10.11) of M by N will not disturb locally quadratic (or worse) convergence. ◇

Indeed, let $A \in N(x)$ and $\|h(x)\| < 1/K^-$. Then $v = Au$ yields

$$\|v\| \geq ((1/K^-) - \|h(x)\|)\ \|u\|, \text{ and } \|A^{-1}\| \leq K^-(1 - K^-\|h(x)\|)^{-1}.$$

With some Lipschitz rank L of h near x^*, inclusion (10.6) ensures, for all linear functions $P \in B_{L(X,Y)}$ and $A \in M(x^* + u)$,

$$\begin{aligned} Au + \|h(x)\|Pu \ &\in h(x^* + u) - h(x^*) + (o(u) + \|h(x)\|\|u\|)B \\ &\subset h(x^* + u) - h(x^*) + (o(u) + L\|u\|^2)B. \end{aligned}$$

This provides us $o(\cdot)$ for N.

Remark 10.4 Similarly, the related Newton equation (10.3) may be replaced by

$$x^{k+1} = x^k - A^{-1}(h(x^k) + \mu(x^k)),\ A \in N(x^k)$$

as long as (10.4), $N(x) = M(x) + \|h(x)\|B_{L(X;Y)}$ and $\|\mu(x)\| \leq o(h(x))$ are satisfied. ◇

Note that, in this case, $\|\mu(x)\| \leq Co(x - x^*)$ is true with some constant C.

10.1.2 Particular Settings

If $X = Y = \mathbb{R}^n$, the following particular settings seem to be appropriate:

$$M(x) = \partial h(x) \text{ or } M(x) = \partial_0 h(x).$$

One could also define

$$A \in M(x) \text{ if the rows } A_i \text{ of } A \text{ belong to } \partial h_i(x)\ \forall i,$$

or one considers the sets A_{ij} of all difference quotients (with unit vectors e^j)

$$\begin{aligned} &[h_i(x + te^j) - h_i(x - se^j)]/(s + t), \quad \text{with } s, t \geq 0 \text{ and} \\ &s + t = \|h(x)\| \text{ if } h(x) \neq 0,\ 0 < s + t \leq \text{ const } (> 0) \text{ if } h(x) = 0 \end{aligned}$$

and defines

$$A \in M(x) \;\Leftrightarrow\; \text{the entrees } a_{ij} \text{ of } A \text{ belong to } \operatorname{cl} A_{ij} \text{ for all } i, j.$$

In the latter cases, $M(x) \subset \partial h(x)$ is not necessarily true.

If h is a PC^1-function generated by $h^1, ..., h^N \in C^1$, then one may put

$$M(x) = \{Dh^i(x) \mid h^i(x) = h(x)\}.$$

This is the setting of one of the first paper on "non-smooth Newton methods", cf. Kojima and Shindoh, [KS87].

In each of these cases, it is easy to see that M is closed and locally bounded, and that (10.4) holds true if and only if $\sup\{\|A^{-1}\| \mid A \in M(x^*)\} < \infty$.

The both conditions (10.4) and (10.6) however are not fulfilled a priory. They require different properties of h, depending on the choice of M.

10.1.3 Realizations for locPC^1 and NCP Functions

For locally PC^1 functions h, Theorem 6.18 yields that $M = D^\circ h$ is a Newton map, so only condition (10.4) becomes crucial.

We consider the case of a complementarity problem with locally PC^1 data in detail. Let $g \in pNCP$, $z = (u, v) \in \text{loc}PC^1$, and denote the related sets of C^1-points by Θ^1_g and Θ^1_z.

We already know that the maps $x \mapsto D^\circ z(x)$ and $y \mapsto D^\circ g(y)$ are Newton maps. Given some element $Rz(x) \in D^\circ z(x)$, let $R_i z(x) = (R_i u(x), R_i v(x))$ denote its i^{th} component.

Further, let $\Phi(x)$ consist of all matrices A the rows A_i of which satisfy

$$A \in \Phi(x) \text{ iff } A_i = G^i R_i z(x) \text{ with } G^i \in D^\circ g(z_i(x)) \text{ and } Rz(x) \in D^\circ z(x)\ \forall i. \tag{10.12}$$

This map Φ, contained in the carthesian product of all sets $D^\circ g(z_i(x)) D^\circ z_i(x)$, is a Newton map for the function h as

$$h_i(x) := g(z_i(x)),$$

cf. Theorem 6.18. Hence condition (10.4), namely the existence of K^-, remains the only problem for applying Newton's method to the NCP-equation $h = 0$ with Newton steps

$$g(z_i(x)) + A_i w = 0 \text{ and } x_{\text{new}} := x + w. \tag{10.13}$$

Moreover, method (10.13) just means to solve the "weighted equation"

$$\begin{aligned} &G^i_s(u_i(x) + R_i u(x) w) + G^i_t(v_i(x) w) + R_i v(x) w) = 0,\\ &(G^i_s, G^i_t) \in D^\circ g(u_i(x), v_i(x)). \end{aligned} \tag{10.14}$$

Indeed, by (9.20), it holds at any $(s, t) \in \Theta^1_g$, after setting $(G_s, G_t) = Dg(s, t)$,

$$g(s, t) = sG_s + tG_t.$$

This equation is still valid for the limits (G_s, G_t) in $D^\circ g$. The latter ensures for the i^{th} Newton equation

$$\begin{aligned} 0 &= g(z_i(x)) + A_i w \\ &= G_s^i u_i(x) + G_t^i v_i(x) + G_s^i R_i u(x) w + G_t^i R_i v(x) w. \end{aligned}$$

Theorem 10.5 (regularity condition (10.4) for NCP). *Let $g \in pNCP$, $z \in C^1$, and let x^* be a solution of the complementarity problem.*

(i) If condition (10.4) is satisfied for the settings of method (10.13), then (10.4) is also satisfied for the special NCP function $g_{\min}\{s,t\} = \min\{s,t\}$.
(ii) Condition (10.4) holds true if the NCP is strongly regular at x^*.
*(iii) Condition (10.4) is equivalent with strong regularity of the NCP at x^* if there is an arc in the C^1- **set** Θ_g^1 of g that connects the unit vectors of $\mathbb{R}^2$.* ◇

Proof. Recalling (9.19), it holds $\max\{G_s^i, G_t^i\} \geq p$ $(\forall i)$ for some $p > 0$. So we see that the matrices A in (10.12) are regular iff so are the matrices with rows

$$C_i(r,x) := (1 - r_i) R_i u(x) + r_i R_i v(x)$$

where

$$r_i = G_t^i [G_s^i + G_t^i]^{-1},\ 1 - r_i = G_s^i [G_s^i + G_t^i]^{-1}$$

and

$$(G_s^i, G_t^i) \in D^\circ g(u_i(x), v_i(x)).$$

For $z \in C^1$, these rows have the form

$$C_i(r,x) = (1 - r_i) D u_i(x) + r_i D v_i(x),$$

and the coefficients r_i form a subset $S_i(x) \subset [0,1]$. By continuity arguments, it suffices to consider $x = x^*$ only for showing (10.4). So, (10.4) holds true if and only if all matrices $C(r, x^*)$ (which form a compact set) are invertible. This condition is as weaker as smaller the sets $S_i(x^*)$ are.

To study $S_i(x^*)$, let $y^* = u(x^*) - v(x^*)$. The pairs (G_s^i, G_t^i) vary in $D^\circ g(u_i(x^*), v_i(x^*))$. If $y_i^* > 0$, we obtain that g is C^1 near $(u_i(x^*), v_i(x^*))$ and $G_s^i = 0, G_t^i > 0, r_i = 1$. Similarly, $y_i^* < 0$ yields $r_i = 0$.

Now let $y_i^* = 0$. Then (G_s^i, G_t^i) is any limit of derivatives $Dg(s', t')$ for $(s', t') \to (0,0)$ in Θ_g^1. By norm-invariance of Dg, we conclude that $D^\circ G(0,0) = \operatorname{cl} Dg(\Theta_g^1)$. So, the pairs (G_s^i, G_t^i) vary in the whole set $\operatorname{cl} Dg(\Theta_g^1)$, and

$$S_i(x^*) = \{G_t [G_s + G_t]^{-1} \mid G \in \operatorname{cl} Dg(\Theta_g^1)\}.$$

In the "smallest case", $S_i(x^*)$ contains 0 and 1. This is just the situation for $g = g_{\min}$. In the "largest one", the full interval [0,1] belongs to $S_i(x^*)$ whenever $y_i^* = 0$. Then, nonsingularity of all $C(r, x^*)$ coincides, by Lemma 7.16, just with strong regularity of the NCP at x^*. Clearly, having an (continuous) arc in Θ^1 which connects the unit vectors of $\mathbb{R}^2$, the set $S_i(x^*)$ is connected and contains 0 and 1. So the equation $S_i(x^*) = [0,1]$ is in fact true for $y_i^* = 0$. This proves the theorem. □

Theorem 10.6 (uniform regularity and monotonicity). *Let the NCP be strongly monotone. Then, for $g \in pNCP$ and $z \in locPC^1$, every matrix $A \in \Phi(x)$ according to (10.12) is regular and fulfills, for each bounded set $X \subset \mathbb{R}^n$,*

$$\|A^{-1}\|_\infty \le 2nC_x(\lambda p)^{-1} \qquad \forall x \in X,$$

where $C_x := \max_i \sup\{\|Dz_i(x)\|_\infty \,|\, x \in X \cap \Theta_z^1\}$, $\lambda = \lambda(u,v)$ is the strict monotonicity constant of NCP, and $p = p(g)$ is the constant from (9.19) taken with the max-norm. ◇

Proof. Suppose one finds $\varepsilon > 0$, $x \in X$, $A \in \Phi(x)$ and some $w \in \operatorname{bd} B$ (Euclidean sphere) such that

$$\varepsilon \ge |A_i w| \text{ for all rows } A_i \text{ of } A. \tag{10.15}$$

This corresponds to the fact that A *is* singular or $\|A^{-1}\|_\infty \ge 1/\varepsilon$ in terms of the maximum-norm. By definition of Φ, it holds for certain $G^i \in D^\circ g(z_i(x))$

$$A_i w = G^i_s R_i u(x) w + G^i_t R_i v(x) w.$$

We know that $Rz \in D^\circ z$ is a Newton function for z at every fixed x by Theorem 6.18. The strict monotonicity of NCP yields, setting $y = x + tw$, $t > 0$,

$$\lambda\|y - x\|^2 \le \langle u(y) - u(x), v(x) - v(y)\rangle = \langle tRu(y)w + o_u(t), tRv(y)w + o_v(t)\rangle.$$

This ensures, for small $0 < t < \tau(x)$,

$$\tfrac{1}{2}\lambda t^2 \le t^2\langle Ru(y)w, Rv(y)w\rangle$$

and

$$\tfrac{1}{2}\lambda \le \langle Ru(x + tw)w, Rv(x + tw)w\rangle.$$

For any $(U, V) \in \limsup_{x' \to x}(Ru(x'), Rv(x'))$, and in particular for each $(U, V) \in D^\circ z(x)$, we thus obtain

$$\tfrac{1}{2}\lambda \le \langle Uw, Vw\rangle.$$

Hence

$$\tfrac{1}{2}\lambda \le \textstyle\sum_i (R_i u(x)w)(R_i v(x)w) =: \sum_i P_i.$$

Let

$$\lambda' = \tfrac{1}{2}\lambda \text{ and } P_k = \max_i P_i.$$

Due to

$$P_k \ge \lambda'/n, \tag{10.16}$$

the factors $a = R_k u(x)w$ and $b = R_k v(x)w$ have the same (non-zero) sign. Further, $\max\{|a|, |b|\} \le C_x$. So, the inequality $ab = P_k \ge \lambda'/n$ ensures

$$\min\{|a|, |b|\} \ge \lambda' C_x^{-1}/n.$$

Returning to (10.15) for $i = k$, and taking into account that $G^k_s \ge 0$ and $G^k_s \ge 0$, the latter yields by (9.19),

$$\varepsilon \ge |G^k_s a + G^k_t b| \ge [\lambda' C_x^{-1}/n] \max\{G^k_s, G^k_t\} \ge [\lambda' C_x^{-1}/n]p.$$

Therefore, $\|A^{-1}\|_\infty \ge 1/\varepsilon$ implies $1/\varepsilon \le ([\lambda' C_x^{-1}/n]p)^{-1}$ as asserted. □

Taking any $g \in pNCP$ and $u \equiv 0$ as well as $v_i(x) = x_i > 0$, one sees that Theorem 10.6 fails to hold for a monotone standard *NCP*, since $A_i = 0$ follows from $Dg(z_i(x)) = (D_s g(0, x_i), 0)$ and $Dz_i(x) = (0, e^i)$.

On the other hand, the theorem still holds without strong monotonicity of NCP whenever (10.16) remains true for some $\lambda' = \lambda'(z) > 0$ and some $k = k(x, w)$.

Moreover, if g has *locally Lipschitz* derivatives on an open and dense subset of $\mathbb{R}^2$ (which is fulfilled for all the given examples of g in Section 9.1) and if $z \in C^{1,1}$, then one obtains quadratic convergence because $o(\cdot)$ in (10.6) now fulfills $o(\cdot) \le L\|\cdot\|^2$.

10.2 The Usual Newton Method for PC^1 Functions

Condition (10.6) also holds for all PC^1 - **functions** h, if we put

$$M(x) = \{Dh^s(x) \mid s \in I(x)\}; \quad I(x) = \{s \mid h^s(x) = h(x)\}.$$

The remaining condition (10.4) now means regularity of all matrices $Dh^s(x^*)$, $s \in I(x^*)$. In this case, x^* is obviously a strongly regular zero of each C^1 -function h^s, $s \in I(x^*)$. However, then one may apply the *usual* Newton method to *any* fixed generating function $g = h^s$, active at x^0, provided that (as usually supposed) x^0 is already close enough to the solution.

Notice that this simplification is possible, if all generating functions h^s are *explicitly* known. But the functions h^s are needed anyway in order to find some element of $M(x)$ and are known for many problems, e.g. for an *NCP* with $z = (u, v) \in C^1$ or for polyhedral generalized equations with C^1 - **functions**, cf. Section 7.1.

10.3 Nonlinear Auxiliary Problems

Solving linear auxiliary problems is only one possibility of dealing with a Lipschitzian equation, many other approaches are thinkable. In this section, we consider linear and non-linear auxiliary problems which may be solved only approximately. In contrast to the previous sections, now the existence of exact solutions (for the auxiliary problems) must not be required.

Let $h : X \to Y$ (normed spaces) be locally Lipschitz with rank L near $x^* \in X$, let $h(x^*) = 0$ and let

$$Gh : (X, X) \rightrightarrows Y$$

be a mapping satisfying the general supposition

$$\emptyset \neq Gh(x, u) \quad \text{and} \quad Gh(x, 0) = \{0\} \qquad (10.17)$$

of this section. Having x^k we want to solve an inclusion of the form

$$\emptyset \neq \alpha\|h(x^k)\|B \cap [h(x^k) + Gh(x^k, u)], \text{ and put } x^{k+1} := x^k + u. \tag{10.18}$$

The fixed parameter α prescribes the accuracy when solving

$$0 \in h(x^k) + Gh(x^k, u). \tag{10.19}$$

If (10.19) holds true, we call u and $x^{k+1} = x^k + u$ *exact solutions.*

One may identity $Gh(x^k, u)$ with any (suitable) multivalued generalized directional derivative of h at x^k. In particular, the settings of the former section are still possible,

$$Gh(x, u) := M(x)u := \{Au \mid A \in M(x)\}, \tag{10.20}$$

where

$$M(x) \text{ is a set of functions in } \operatorname{Lin}(X,Y).$$

Notice, that the existence of an inverse or even surjectivity are not explicitly required, now. This is a realistic assumption for equations arising from control problems. Basic ideas to this topic can be found, e.g., in [Alt90].

Feasibility: We call the triple (h, Gh, x^*) *feasible* if, for each $\varepsilon \in (0,1)$, there are positive r and α such that, whenever $\|x^0 - x^*\| \leq r$, process (10.18) has solutions and generates iterates satisfying

$$\|x^{k+1} - x^*\| \leq \varepsilon\|x^k - x^*\|.$$

To ensure feasibility of (h, Gh, x^*), we will impose the following conditions for x near x^* which now replace (10.4) and (10.6) in the previous section.

Condition (CI) (injectivity of the derivative).

$$\|v\| \geq c\|u\| \quad \forall v \in Gh(x,u)\ \forall u \in X \quad (c > 0 \text{ fixed}).$$

Condition (CA) (condition for the approximation).

$$h(x) - h(x^*) + Gh(x,u) \subset Gh(x, x + u - x^*) + o(x - x^*)B\ \forall u \in X.$$

Considering only the directions $u = x^* - x$ in (CA) we get a weaker condition by using $Gh(x, 0) = \{0\}$, namely,

Condition (CA)* (simplified condition CA).

$$h(x) - h(x^*) + Gh(x, x^* - x) \subset o(x - x^*)B.$$

This condition requires a good behavior of the "directional derivatives" $Gh(x, \cdot)$ with respect to difference quotients including x^*, provided that Gh is positively homogeneous in the second argument:

$$Gh(x, (x^* - x)/\|x - x^*\|) \subset [h(x^*) - h(x)]/\|x - x^*\| + O(x - x^*)B.$$

Since $h \in C^{0,1}(X, Y)$, (CA)* implies, for all $v \in Gh(x, x^* - x)$,

$$\|v\| \le \|h(x) - h(x^*)\| - o(x - x^*) \le L\|x - x^*\| - o(x - x^*) \le 2L\|x - x^*\|,$$

i.e.,

$$Gh(x, x^* - x) \subset 2L\|x - x^*\|B \text{ for small } \|x - x^*\|. \qquad (10.21)$$

It turns out that

$$(CA) \quad \Leftrightarrow \quad (CA)^*$$

holds for many relevant settings of Gh, cf. Theorem 10.8.

10.3.1 Convergence

Based on (CA) and (CI), let us summarize the convergence properties for the current method (10.18).

Theorem 10.7 (convergence of Newton's method - II).

(i) The triple (h, Gh, x^*) is feasible if there exist $c > 0, \delta > 0$ and a function $o(\cdot)$ such that, for all $x \in x^* + \delta B$, the conditions (CI) and (CA) are satisfied.

Moreover, having (CI) and (CA), let

$$\varepsilon \in (0,1), \ \alpha \in (0, \tfrac{1}{2}c\varepsilon L^{-1}], \text{ and let } r \in (0,\delta] \text{ be small enough such that } o(x - x^*) \le \tfrac{1}{2}c\|x - x^*\| \ \forall x \in x^* + rB. \qquad (10.22)$$

Under this condition, the convergence can be quantified as follows:

(ii) If r even satisfies

$$o(x - x^*) \le \tfrac{1}{2}\alpha c\|x^* - x\| \quad \forall x \in x^* + rB, \qquad (10.23)$$

then ε, α and r fulfill the requirements in the definition of feasibility. In particular, (10.18) remains solvable if $\|x^0 - x^*\| \le r$.

(iii) If there exists a solution u of (10.18) for every $x^k \in x^* + rB$, then

$$\|x^{k+1} - x^*\| \le \tfrac{1}{2}(1 + \varepsilon)\|x^k - x^*\|, \ \textit{provided that } \|x^0 - x^*\| \le r.$$

So (10.23) and $\|x^{k+1} - x^*\| \le \varepsilon\|x^k - x^*\|$ hold for large k.

(iv) If all x^{k+1} are exact solutions of (10.18), then they fulfill

$$c\|x^{k+1} - x^*\| \le o(x^k - x^*) \textit{ with } o(\cdot) \textit{ from (CA) if } \|x^0 - x^*\| \le r.$$

◇

Note: The conditions (CI), (CA) and (CA)*, respectively, must be imposed for $\|u\| \le 2\|x - x^*\|$ only. ◇

Proof. Given $\varepsilon \in (0,1)$, let r be taken as under (10.22), and let $x \in x^* + rB$.

(Preparation)
First we apply (CA)* and (CI) to elements $u \in \|x - x^*\|B$ only. With $u^* = x^* - x$, (CA)* and (10.17) yield

$$\emptyset \neq h(x) + Gh(x, u^*) \subset o(u^*)B,$$

and (CI) ensures

$$\|v\| \geq c\|u^*\| \ \forall v \in Gh(x, u^*).$$

Thus, since $o(u^*) \leq \frac{1}{2}c\|u^*\|$, it follows

$$\|h(x)\| \geq \tfrac{1}{2}c\|u^*\|. \tag{10.24}$$

This yields that, for each $\alpha > 0$,

$$\tfrac{1}{2}\alpha c\|x^* - x\| \leq \alpha\|h(x)\| \leq \alpha L\|x - x^*\| \tag{10.25}$$

holds true.

(ii, existence of a solution to $\alpha > 0$*)*
Now let $\alpha \in (0, \frac{1}{2}\varepsilon c L^{-1}]$ and r be small enough such that (10.23) holds true. Then $h(x) + Gh(x, u^*)$, contained in $o(u^*)B$, meets the larger balls

$$\tfrac{1}{2}\alpha c\|x^* - x\|B \subset \alpha\|h(x)\|B.$$

Hence $u^* = x^* - x$ solves (10.18) for $x^k = x$ with the current α.

Up to now we only applied the conditions (CA)* and (CI).

(ii, estimate of any solution to $\alpha > 0$*)*
Next, let u by any solution of the auxiliary problem (10.18) at $x = x^k$. By (CA) we observe

$$\begin{array}{rcl} \emptyset & \neq & \alpha\|h(x)\|B \cap [h(x) + Gh(x,u)] \\ & \subset & \alpha\|h(x)\|B \cap (Gh(x, x + u - x^*) + o(x - x^*)B), \end{array} \tag{10.26}$$

and by (CI), each $v \in Gh(x, x + u - x^*) + o(x - x^*)B)$ has at least the norm

$$\|v\| \geq c\|x + u - x^*\| - o(x - x^*).$$

Because some v belongs to $\alpha\|h(x)\|B$, this yields with (10.25) that the key inequalities

$$c\|x + u - x^*\| \leq o(x - x^*) + \alpha\|h(x)\| \leq o(x - x^*) + \alpha L\|x^* - x\| \tag{10.27}$$

hold true.

(iv, estimate of any exact solution)
For *exact solutions* u of (10.18), condition (CA) yields

$$0 \in h(x) + Gh(x,u) \subset Gh(x, x+u-x^*) + o(x-x^*)B,$$

so the estimates (10.26) and (10.27) hold even with $\alpha = 0$ and lead us to

$$c\|x+u-x^*\| \le o(x-x^*) \le \tfrac{1}{2}c\|x-x^*\| \le \tfrac{1}{2}c\,r.$$

Thus, exact solutions $x^{k+1} = x+u$ satisfy $\|x^{k+1}-x^*\| \le \frac{1}{2}r$, so $x^{k+1} \in x^* + \frac{1}{2}rB$ and (iv) is true.

(Note)
Recalling our basic settings in (10.22), namely

$$o(x-x^*) \le \tfrac{1}{2}c\|x-x^*\| \text{ and } \alpha \le \tfrac{1}{2}c\varepsilon L^{-1},$$

inequality (10.27) provides (for exact and "inexact" solutions as well) an estimate for u, namely,

$$\begin{aligned} \|u\| &\le \|x-x^*\| + \|x^*-x-u\| \\ &\le \|x-x^*\| + c^{-1}o(x-x^*) + \alpha c^{-1}L\|x-x^*\| \\ &\le \|x-x^*\| + \tfrac{1}{2}\|x-x^*\| + \tfrac{1}{2}\varepsilon\|x-x^*\| \\ &\le 2\|x-x^*\|. \end{aligned} \tag{10.28}$$

So, our assumptions (CI), (CA) and (CA)*, respectively, have to hold for

$$\|u\| \le 2\|x-x^*\|$$

only.

(iii)
Additionally, (10.27) ensures

$$\|x+u-x^*\| \le \tfrac{1}{2}\|x-x^*\| + \tfrac{1}{2}\varepsilon\|x-x^*\| = \tfrac{1}{2}(1+\varepsilon)\|x-x^*\|.$$

Thus, all x^{k+1} are again in $x^* + rB$ and converge to x^* as asserted under (iii).

(i) and *(ii)*
Finally, if also (10.23) holds true, then (10.27) ensures

$$\begin{aligned} c\|x+u-x^*\| &\le \tfrac{1}{2}\alpha c\|x^*-x\| + \alpha L\|x^*-x\| \\ &\le 2\,\alpha\,L\,\|x-x^*\| \\ &\le c\,\varepsilon\,L^{-1}\,L\,\|x-x^*\|. \end{aligned}$$

So (ii) (and hence (i)) is valid as desired: $\|x+u-x^*\| \le \varepsilon\|x-x^*\|$. □

The proof also shows that, with the constants of the theorem, the following holds:

Due to (10.24), the zero x^* of h is *isolated.* For $\alpha > 0$ and $x = x^k$ near x^*, the point $u^* = x^* - x$ is a solution of the auxiliary problem (10.18) satisfying

$$(0 =) \|x + u^* - x^*\| \leq \varepsilon \|x - x^*\|.$$

The *full condition* (CA) - not only (CA)* - was needed for showing that *all solutions* u of (10.18) satisfy this estimate

$$\|x + u - x^*\| \leq \varepsilon \|x - x^*\|,$$

thus they fulfill (10.28), too.

Finally, using (ii), the inequalities (10.24) and $\|x^{k+1} - x^*\| \leq \varepsilon \|x^k - x^*\|$ ensure

$$\|h(x^{k+1})\| \leq L\|x^{k+1} - x^*\| \leq L\varepsilon \|x^k - x^*\| \leq 2L\varepsilon c^{-1} \|h(x^k)\|. \tag{10.29}$$

So $\|h(\cdot)\|$ is decreasing provided that ε (depending on L and c) has been taken sufficiently small.

10.3.2 Necessity of the Conditions

Under several particular settings, the technical condition (CA) may be replaced by (CA)*.

Theorem 10.8 (the condition (CA)). *Suppose $X = Y = \mathbb{R}^n$, and let Gh denote any of the following generalized directional derivatives:*

$$\begin{aligned}
Gh(x,u) &= Th(x,u),\\
Gh(x,u) &= Ch(x,u),\\
Gh(x,u) &= \partial h(x)u \text{ (Clarke's Jacobian applied to } u),\\
Gh(x,u) &= h'(x,u) \text{ (usual directional derivatives, provided that they exist)}\\
Gh(x,u) &= \{Au \mid A = Dh^k(x) \text{ and } h^k(x) = h(x)\} \text{ if } h \in PC^1(h^1, ..., h^m).
\end{aligned}$$

Then, the conditions (CA) and (CA) are equivalent.* ◇

Proof. If $Gh(x,u) = M(x)u$ where $M(x)$ is a set of linear functions, condition (CA)* means

$$h(x) + A(x^* - x) \in o(x - x^*)B \qquad \forall A \in M(x).$$

One obtains (CA) by adding Au:

$$h(x) + Au \in A(x - x^* + u) + o(x - x^*)B \qquad \forall A \in M(x).$$

For $Gh = Th$, the proof follows from the subadditivity inclusion

$$Th(x, u' + u'') \subset Th(x, u') + Th(x, u''),$$

cf. (6.10) in Section 6.2, and is left to the reader.

Next we will only consider directional derivatives h', because the proof for Ch is basically the same (one has only to select appropriate subsequences). Thus, let h be directionally differentiable *near* x^*. We may suppose that $o = o(u)$ in (CA)* is u.s.c. and decreases for $\|u\| \to 0$. Further, $x^* = 0$ may be assumed. We have to show that, for arbitrarily fixed $u \in X$,

$$h(x) + h'(x,u) \in h'(x, x+u) + o(x)B. \tag{10.30}$$

Our assumption (CA)*, $h(x) + h'(x,-x) \in o(x)B$, allows us to write

$$\begin{aligned} h(x - \lambda x) &= h(x) + \lambda h'(x,-x) + \alpha(x,\lambda) \\ &= h(x) + \lambda(-h(x) + \varepsilon(x)) + \alpha(x,\lambda) \end{aligned} \tag{10.31}$$

where $\varepsilon(x) \in o(x)B$ and $\|\alpha(x,\lambda)\|/\lambda \to 0$ as $\lambda \to 0$. Regarding the assigned limits (for $t \downarrow 0$) we have

$$\begin{aligned} h'(x, x+u) &= \lim t^{-1}[h(x + t(u+x)) - h(x)] \\ &= h'(x,u) + \lim t^{-1}[h(x+tu+tx) - h(x+tu)] \end{aligned}$$

So we have to verify that

$$h(x) - \lim t^{-1}[h(x+tu+tx) - h(x+tu)] \in o(x)B.$$

After setting

$$s = 1/(1+t),\ r = t^2/(1+t) \text{ and } p = p(t) = x + tu + tx$$

we notice that

$$x + tu = sp + ru$$

and

$$h(x+tu+tx) - h(x+tu) = h(p) - h(sp + [t^2/(1+t)]u).$$

When computing the limit of the crucial quotients

$$t^{-1}[h(x+tu+tx) - h(x+tu)] = t^{-1}[h(p) - h(sp + [t^2/(1+t)]u)],$$

the term $[t^2/(1+t)]u$ may be omitted because h is locally Lipschitz.

Next, write $h(sp) = h(p - \lambda p)$ with

$$\lambda = t/(1+t)$$

and apply (10.31). Then,

$$[h(p - \lambda p) - h(p)]/\lambda = -h(p) + \varepsilon(p) + \alpha(p,\lambda)/\lambda.$$

For small t, it holds $\|\alpha(p,\lambda)\|/\lambda < o(x)$ and, since $\varepsilon(p) \in o(p)B$ and $o(\cdot)$ is norm-decreasing and continuous,

$$\|\varepsilon(p)\| < o(2x).$$

Thus the limit lim $[h(x+tu+tx) - h(x+tu)]\ /t$ (for vanishing t or λ) belongs to

$$-h(x) + (o(2x) + o(x))B.$$

This ensures the assertion (10.30) with a new function $o'(x) = 2o(2x)$. □

Using the approximation of a Lipschitz function by directional derivatives (if they exist near x^*), i.e., $h(x)-h(0) \in h'(0;x)+o(x)B$, our Theorem 10.8 verifies the following statement:

$$\text{If } h(x) - h(0) + h'(x;-x) \in o(x)B$$
$$\text{then } h'(0;x) + h'(x;u) \in h'(0+x;x+u) + o(x)B.$$

This is an interesting additivity property of directional derivatives for $C^{0,1}$ functions.

Having *normed* X and Y, the equivalence (CA) ⇔ (CA)* for directional derivatives remains true (by the proof just given). However, when dealing with contingent derivatives and $\dim Y = \infty$, one needs a strong *extra assumption* that replaces the existence of h': Given any sequence $t_k \downarrow 0$ there is always a (norm-)convergent subsequence of $t_k^{-1}[h(x+t_k u) - h(x)]$.

At the end of this section, we characterize the *necessity* of (CI) and (CA) for several relevant settings.

Theorem 10.9 (the condition (CI)). *Suppose that* $h \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ *and* $h(x^*) = 0$.

Let $Gh(x,u) = Th(x)(u)$.
Then (CI) holds at x^* ⇔ *(CI) holds for* x *near* x^* ⇔ h *is strongly regular at* x^*. *Having (CI), Condition (CA) is necessary and sufficient for* (h, Gh, x^*) *being feasible.*

Let $Gh(x,u) = \partial h(x)u$.
Then (CI) holds at x^* ⇔ *(CI) holds for* x *near* x^* ⇔ $\partial h(x^*)$ *non singular. This condition is stronger than strong regularity.*

Let $Gh(x,u) = Ch(x)(u)$.
Then (CI) holds at x^* ⇔ h^{-1} *is locally upper Lipschitz at* $(0, x^*)$.

Let $Gh(x,u) = h'(x;u)$, *provided that directional derivatives exist near* x^*. *Then, supposing strong regularity, (CA) is necessary and sufficient for* (h, Gh, x^*) *being feasible; supposing pseudo-regularity, (CI) is satisfied for* x *near* x^*. ◇

Proof. *Let* $Gh(x,u) = Th(x)(u)$.
The first assertions follow from Theorem 5.14 and closeness of $Th(\cdot)$. To show the necessity of (CA) [for sufficiency apply Theorem 10.7], we may assume that $x^* = 0$.

Let $v \in h(x) + Th(x)(-x)$. We have to show that $v \in o(x)B$. Using the inverse and subadditivity (6.10) of Th^{-1}, it holds

$$-x \in Th^{-1}(h(x))(v - h(x)) \subset Th^{-1}(h(x))(v) + Th^{-1}(h(x))(-h(x)).$$

We select

$$w \in Th^{-1}(h(x))(v) \text{ and } u \in Th^{-1}(h(x))(-h(x)) \text{ such that } -x = w + u.$$

Every $u \in Th^{-1}(h(x))(-h(x))$ solves $0 \in h(x) + Th(x)(u)$. Hence, due to feasibility, $-w = x + u$ belongs to $o(x)B$. Since $v \in Th(x)(w)$, this yield as desired $v \in Lo(x)B$ where $L =$ Lip (h).

Let $Gh(x,u) = \partial h(x)u$.
By Clarke's inverse function theorem, we have $\partial h(x^*)$ non singular $\Rightarrow$ (CI). By Example BE.3 we see that the reverse statement does not hold, in general. The rest follows from closeness of $\partial h(\cdot)$. Notice that the piecewise linear function of this example fulfills condition (CA); so the example is relevant in the present context.

Let $Gh(x,u) = Ch(x)(u)$.
The statement follows immediately from Theorem 5.1.

Let $Gh(x,u) = h'(x;u)$.
Let h be strongly regular. Then (CI) is obviously true for x near x^*. Moreover, under strong regularity one can show that h is directionally differentiable iff h^{-1} shows the same property (cf. Exercise 10). Then is also holds - since the next equivalence is valid for contingent derivatives,

$$0 = h(x) + h'(x;u) \Leftrightarrow u \in (h^{-1})'(h(x); -h(x)).$$

Because $(h^{-1})'$ exists, there is always some $u = u(x)$ satisfying this equation. Having feasibility of our triple (or superlinear convergence of the $(\alpha = 0)-$version), we obtain once more $x+u \in o(x)B$. Accordingly, we write $u = -x+o_x$.

Now, to show that (CA) is valid, assume that $v = h(x) + h'(x; -x)$. Then,

$$-h(x) \in h'(x; -x + o_x) \text{ and } -h(x) + v \in h'(x; -x).$$

Since h' is Lipschitz (with rank L) in the second argument, it follows

$$\|v\| \leq L\|o_x\| \text{ with } o_x \in o(x)B,$$

which gives (CA)* and – by Theorem 10.8 – even (CA).

Finally, if h is only pseudo-regular at x^*, the assertion is true due to Theorem 5.12. □

Chapter 11

Particular Newton Realizations and Solution Methods

After some problem has been modeled as a (nonsmooth) equation, the Newton-steps for solving the latter induce particular iteration steps (actions) and regularity requirements in the original problem; for instance, compare equation (9.21) and the intrinsic equivalent equation (9.22). We study these actions and requirements for Karush-Kuhn-Tucker systems (KKT) of optimization models, and want to demonstrate their dependence on the applied Newton techniques and the corresponding reformulations.

We will see that the related auxiliary problems (being linear or nonlinear equations a priori) describe solutions of quadratic optimization problems in all cases. So one can solve the auxiliary problems by several methods (of second or first order), in particular if certain numerical difficulties occur with the current one. In this way, connections to *SQP-methods* (sequential quadratic programming) become evident, but we also establish a bridge to methods of penalty-barrier type and will compare the hypotheses and actions in terms of the original problems.

Our tool consists in studying certain perturbed Kojima systems that describe, in a uniform way, stationary points of assigned penalty or barrier functions close to an original solution x^*. So the approach permits solution estimates, based on regularity assumptions at x^*. In addition, it makes also clear (by using general properties of pNCP functions) that reformulations of the KKT- complementarity condition by pNCP functions can be always modeled in form of particularly perturbed Kojima systems with perturbations that depend on $g_i(x)$ and y_i only.

11.1 Perturbed Kojima Systems

We consider perturbations of the Kojima function $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$, assigned to the $C^{1,1}$ problem

$$\min f(x) \text{ s. t. } g(x) \le 0, \tag{11.1}$$

namely,

$$\begin{aligned} F_1 &= Df(x) + \Sigma y_i^+ \, Dg_i(x), \\ F_{2i}^t &= g_i(x) - y_i^- - t_i \, y_i^+ \end{aligned}$$

and study solutions of the system

$$\begin{aligned} F_1 &= 0 \\ F_{2i}^t &= 0. \end{aligned} \tag{11.2}$$

Let us show that this system is closely related to penalty and barrier methods for problem (11.1). Note again that - in contrast to the common terminology in the literature - the whole auxiliary function (i.e., objective function + penalty/barrier term) is said to be a *penalty/barrier function.*

Quadratic Penalties

Suppose $t_i > 0 \; \forall i$. Let (x, y) solve (11.2). Then we know:

If $y_i \le 0$, then it follows $y_i^+ = 0$ and $g_i(x)^+ = 0$.
If $y_i > 0$, then it follows $g_i(x) = t_i y_i^+$ and $y_i^+ = t_i^{-1} g_i(x)^+$.

Hence, we obtain in both cases $0 = F_1 = Df(x) + \Sigma \, t_i^{-1} g_i(x)^+ Dg_i(x)$, i.e., x is a stationary point of the penalty function

$$P_t(x) = f(x) + \tfrac{1}{2} \textstyle\sum_i t_i^{-1} [g_i(x)^+]^2.$$

Conversely, if x is stationary for $P_t(x)$, then (x, y) with

$$y_i = t_i^{-1} g_i(x) \text{ for } g_i(x) > 0 \text{ and } y_i = g_i(x) \text{ for } g_i(x) \le 0$$

solves (11.2).

Logarithmic Barriers

Let $t_i < 0 \; \forall i$. Now, the second equation of (11.2), $g_i(x) = y_i^- + t_i y_i^+ \; (\le 0)$, implies feasibility of x in (11.1). Let (x, y) solve (11.2). Then:

If $y_i \le 0$, then $g_i(x) = y_i^-$ and $y_i^+ = 0$ does not appear in the Lagrangian.
If $y_i > 0$, then $g_i(x) = t_i y_i^+$ and $y_i^+ = t_i^{-1} g_i(x)^-$.

Setting $J = \{i | y_i > 0\}$ we thus observe

$$0 = F_1 = Df(x) + \textstyle\sum_{i \in J} t_i^{-1} g_i(x)^- Dg_i(x).$$

Hence, the point x has the following properties. It is feasible for (11.1), fulfills $g_i(x) < 0 \ \forall i \in J$, and is stationary (not necessarily minimal !) for the function

$$Q_t(x) = f(x) + \tfrac{1}{2} \textstyle\sum_{i \in J} t_i^{-1} [g_i(x)^-]^2.$$

Conversely, having some x with the latter properties, the point (x, y) with

$$y_i = t_i^{-1} g_i(x)^- (i \in J) \quad \text{and} \quad y_i = g_i(x) \ (i \notin J)$$

solves (11.2).

For $i \in J$, the terms $g_i(x)^- Dg_i(x)$ coincide with $g_i(x)^2 D(\ln(-g_i(x))$. So we see that

$$t_i^{-1} g_i(x)^- Dg_i(x) = t_i^{-1} g_i(x)^2 D(\ln(-g_i(x)) = t_i \, y_i^2 D(\ln(-g_i(x)).$$

Accordingly, the current point x is also stationary for the logarithmic barrier function

$$B_t(x) = f(x) - \textstyle\sum_{i \in J} |t_i| \, y_i^2 \, \ln(-g_i(x)).$$

Theorem 11.1 (perturbed Kojima-systems). *Under the above settings, zeros of the perturbed Kojima equation (11.2) and critical points of the well-known auxiliary functions P_t, Q_t, B_t correspond to each other.*

Under strong regularity of (11.1) at a critical point (x^*,y^*) of F, the solutions (x_t, y_t) of (11.2) are, for small $\|t\|$, locally unique and Lipschitz. So, it holds

$$\|(x_{t'}, y_{t'}) - (x_t, y_t)\| \le L\|t' - t\| \text{ for all } t, t' \text{ near the origin.} \tag{11.3}$$

◇

Proof. This follows directly from the given transformations and Corollary 4.4, since the maps $y_i \mapsto t_i y_i^+$ are small Lipschitz functions in the $C^{0,1}$-norm. □

Remark 11.2 (modifications). The inequality (11.3) now compares solutions of *different methods* in a Lipschitzian manner. Further, one may mix the signs of the t-components and obtains similarly stationary points for auxiliary functions containing both penalty and barrier terms. So, given some initial point x, it is quite natural to put

$$t_i < 0 \text{ if } g_i(x) < 0 \quad \text{and} \quad t_i > 0 \text{ if } g_i(x) > 0.$$

Moreover, for critical points (x_t, y_t) which are not locally unique, the same arguments including Corollary 4.4, present estimates of (x_t, y_t) under pseudo-regularity of F at (x^*, y^*) or ensure estimates of the difference $(x_t, y_t) - (x^*, y^*)$ under the upper Lipschitz property of F^{-1} at (x^*, y^*), even if f and g are only C^1 functions. ◇

If problem (11.1) includes also equality constraints

$$h(x) = 0 \text{ with related duals } z,$$

then additional perturbations of the type

$$h_k(x) = \tau_k\, z_k$$

change the functions P_t, Q_t and B_t only by additional terms

$$+\frac{1}{2}\sum_k \tau_k^{-1}\, h_k(x)^2.$$

Concerning other auxiliary problems and more details we refer to [Kum95b, Kum97].

11.2 Particular Newton Realizations and SQP-Models

Let us assume that h in Section 10.1 coincides with Kojima's function $F(x,y)$, assigned to our standard optimization problem (11.1),

$$F_1(x,y) = Df(x) + \sum y_i^+ Dg_i(x), \qquad F_{2j}(x,y) = g_j(x) - y_j^-.$$

For deriving relations to **SQP-methods**, we suppose $f, g \in C^2$. Then F is a PC^1 function, and all the mentioned derivatives are Newton maps (or satisfy condition (CA)). Again, we omit additional equality constraints only for seek of brevity.

Depending on the choice of a Newton map M (or of a generalized derivative) we investigate the kind of the related auxiliary problems and the meaning of the (Newton-) regularity condition (10.4), imposed for points $z = (x,y)$ near a zero z^*. In all subsequent cases, we assume that z is the current iteration point and (u,v) describes the movement $(x,y)_{new} - (x,y)$, defined by the Newton step.

It will be seen that (u,v) is a solution (primal-dual) of some quadratic optimization problem. So the (generalized) Newton-methods are at the same time **SQP-methods**, and differences between them arise from the different approximations of $h = F$ in the related Newton equations.

We are now going to study this interrelation for particular settings.

Case 1:

Apply the usual Newton method to *any* fixed generating function F^S of F being active at the *initial* point (x^0, y^0).

The functions F^S are defined by index sets $S \subset \{1,...,m\}$ as

$$\begin{array}{lcll} F_1^S(x,y) & = & Df(x) + \sum_{i\in S} y_i Dg_i(x) & \\ F_{2i}^S(x,y) & = & g_i(x) & \text{if } i \in S \\ F_{2j}^S(x,y) & = & g_j(x) - y_j & \text{if } j \in \{1,...,m\}\backslash S. \end{array}$$

Here, we assigned, to (y_i^+, y_i^-), the function $y_i \mapsto (y_i, 0)$if $i \in S$ and $y_i \mapsto (0, y_i)$ otherwise. The initial set S^0 has to be active at (x^0, y^0), i.e.,

$$i \in S^0 \text{ if } y_i^0 > 0 \text{ and } j \notin S^0 \text{ if } y_j^0 < 0.$$

If $y_i^0 = 0$, we may fix any of the two alternatives. Because $S = S^0$ remains fixed during all steps, the iterations require

$$F^S(x,y) + DF^S(x,y)(u,v)^T = 0.$$

The equations related to F_{2j}^S for $j \notin S$ have the form

$$g_j(x) + Dg_j(x)u = y_j + v_j \; (= y_j^{\text{new}}),$$

and v_j does not appear in any other equation. Thus, we have to solve the problem

$$p[S^0]: \qquad \min f(\xi) \text{ s.t. } g_i(\xi) = 0 \; \forall i \in S^0$$

by linearization of the related C^1- **Karush-Kuhn-Tucker** system at (x, y).

Condition (10.4) requires regularity of the Jacobians $DF^s(z^*)$ for all S, active at z^*. This is strong regularity of all related problems $p[S]$ at the solutions (x^*, y_S^*). So condition (10.4) is weaker than strong regularity of the original problem at (x^*, y^*).

Case 2:

With the Kojima-Shindoh approach (see §10.1.2), one selects some set S being active at the current point (x, y) and makes next a Newton step based on (changing) S as above. The condition (10.4) is the former one. The method is a classical " active index set" algorithm.

Case 3:

Applying the generalized Jacobian $M = \partial F$ (= TF, since $f, g \in C^2$) one may take any matrix $P(r) \in \partial F(z)$, cf. (7.49), for the Newton step

$$F(z) + P(r)(u,v)^T = 0.$$

Condition (10.4) requires more than above, namely just strong regularity of problem (11.1) (or of F) at (x^*, y^*).

We study the Newton steps for the original Kojima system and the perturbed equation (11.2) at once by considering any $t_i \in \mathbb{R}$ and dealing with the Newton equation for F^t

$$F^t(z) + P(r,t)(u,v)^T = 0, \quad P(r,t) \in \partial F^t(z). \tag{11.4}$$

Recall that this setting represents a mixed penalty-barrier approach, cf. Section 11.1, for solving (11.1).

Let $z = (x, y)$ and t be fixed. Practically, t may depend on z (in each step). To obtain locally superlinear convergence, it suffices to ensure that

$$\|t\| = o(F(z)), \quad \text{e.g., } |t_i| = \|F(z)\|^2,$$

cf. the Remarks 10.3, 10.4 concerning approximation (10.11) in Section 10.1 and notice that not only ∂F but also the original Kojima function F has been changed by t.

We abbreviate $Df = Df(x), Dg_i = Dg_j(x)$ and $F = F(x,y)$. Given $r \in \mathcal{R}_T(y)$, cf. (7.28), we put

$$b_i = 1 - r_i + t_i r_i \quad \forall i$$

in accordance with the T-derivative of $y_i^- + t_i y_i^+$.

Below, $D^2_{xx}L(z)$ will stand for $D_x F_1(z)$, so the Lagrangian $L = f + \langle y^+, g \rangle$ does not depend on $y_i \le 0$ (in contrast to the case 4 following next). Finally, put

$$J = \{i|\ b_i \ne 0\}, \ K = \{k|\ b_k = 0\}.$$

If $y_i < 0$ then $r_i = 0$, $b_i = 1$; hence $i \in J$, and our weights w_i in the next statement are zero.

Lemma 11.3 (Newton steps with perturbed F). *In the current case, a Newton step (11.4) means to find a KKT- point (u,μ) of the problem*

$$\begin{aligned} &\min_u \ (Df + \textstyle\sum_{k\in K} y_k Dg_k)u + \frac{1}{2}u^T D^2_{xx}Lu + \frac{1}{2}\sum_{i\in J} w_i(g_i + Dg_i\, u)^2 \\ &\text{s.t.} \ \ g_k + Dg_k\, u = 0 \quad \forall\, k \in K; \end{aligned} \tag{11.5}$$

where $w_i = r_i b_i^{-1}$ and $L = f + \langle y^+, g \rangle$. The vector v in (11.4) is then given by

$$v_k = \mu_k(1 - t_k)\ (k \in K) \ \text{and}\ v_i = b_i^{-1}(g_i + Dg_i\, u - (y_i^- + t_i y_i^+))\ (i \in J).$$

◇

Proof. The linearized equations $F^t_{2i} = 0$ require (equivalently), by the product rule given in Corollary 6.10,

$$g_i + Dg_i\, u - b_i v_i = y_i^- + t_i y_i^+,$$

i.e.,

$$v_i = b_i^{-1}[g_i + Dg_i\, u - (y_i^- + t_i y_i^+)]\ (i \in J) \ \text{and}\ g_k + Dg_k\, u = 0\ (k \in K).$$

By substituting v_j in the linearized equation $F_1 = 0$, i.e., in

$$F_1 + D^2_{xx}Lu + \sum_{k\in K} r_k v_k Dg_k + \sum_{i\in J} r_i v_i Dg_i = 0,$$

and setting

$$\mu_k = r_k v_k = (1 - t_k)^{-1} v_k \quad \text{for } k \in K,$$

one obtains

$$\begin{aligned} 0 &= F_1 + D^2_{xx}Lu + \textstyle\sum_{k\in K} r_k v_k Dg_k + \sum_{i\in J} w_i[g_i + Dg_i u - (y_i^- + t_i y_i^+)]Dg_i \\ &= D^2_{xx}Lu + Df + \textstyle\sum_{k\in K}(y_k^+ + \mu_k)Dg_k \\ &\qquad + \textstyle\sum_{i\in J}(y_i^+ + w_i[g_i + Dg_i u - (y_i^- + t_i y_i^+)])Dg_i \\ &= D^2_{xx}Lu + Df + \textstyle\sum_{k\in K}(y_k^+ + \mu_k)Dg_k \\ &\qquad + \textstyle\sum_{i\in J}((1 - w_i t_i)y_i^+ - w_i y_i^- + w_i[g_i + Dg_i u])Dg_i. \end{aligned}$$

For $i \in J$, we have $(1 - w_i t_i)y_i^+ - w_i y_i^- = 0$. Indeed, if $y_i < 0$, we know that $y_i^+ = 0$, $w_i = 0$. If $y_i > 0$ we have $r_i = 1$, $w_i = b_i^{-1}$ and $b_i = t_i$.

So the F_1 - Newton equation becomes

$$0 = D^2_{xx}Lu + Df + \textstyle\sum_{k\in K} y_k Dg_k + \sum_{i\in J} w_i(g_i + Dg_i u)Dg_i + \sum_{k\in K}\mu_k Dg_k$$

and has the form

$$0 = D_u Q(u, r) + \textstyle\sum_{k\in K}\mu_k Dg_k,$$

where

$$Q = (Df + \textstyle\sum_{k\in K} y_k Dg_k)u + \dfrac{1}{2}u^T D^2_{xx}Lu + \dfrac{1}{2}\displaystyle\sum_{i\in J} w_i(g_i + Dg_i u)^2.$$

This proves the assertion. □

Note. The case of $K = \emptyset$ (no constraints in problem (11.5)) can be easily forced by setting $t_i \neq 0\, \forall i$ and $r_i = 1$ whenever $y_i \geq 0$.

Let $t_i > 0$. Then, if $y_i > 0$, the weights $w_i = t_i^{-1}$ are just the penalty factors. For $y_i = 0$ and $t_i = 0$, all choices of $r_i \in [0, 1]$ are allowed. So w_i may attain all non-negative values.

Let $t_i < 0$. If $y_i > 0$, now $w_i = t_i^{-1}$ is negative, and stationary u are not necessarily minimizer of problem (11.5). If $y_i = 0$, it holds $0 \geq w_i \geq t_i^{-1}$. ◇

Case 4:

Application of NCP functions. To solve the KKT-system of the C^2 - problem (11.1) by the help of some function $G \in pNCP$, require the usual Lagrange condition (without y_i^+ !)

$$\Phi_1(x, y) := D_x L(x, y) := Df(x) + \textstyle\sum_i y_i Dg_i(x) \quad = 0$$

and write the remaining conditions as

$$\Phi_{2i}(x, y) := G(-g_i(x), y_i) \qquad\qquad = 0.$$

Using the derivative $D^\circ G$ in accordance with (10.12) we have to solve

$$\begin{aligned} D_xL(x,y) + D^2_{xx}L(x,y)u + \textstyle\sum_i v_i Dg_i(x) &= 0 & (11.6)\\ -a_i(g_i(x) + Dg_i(x)u) + b_i(y_i + v_i) &= 0, & (11.7) \end{aligned}$$

with $(a_i, b_i) \in D^\circ G(-g_i(x), y_i)$. Let

$$J = \{i |\ b_i \neq 0\},\ K = \{k |\ b_k = 0\}.$$

Now the Newton equation has again the form of case 3, only a_i and r_i must be identified for $i \in J$ and L stands for the usual Lagrangian.

Lemma 11.4 (Newton steps with pNCP). *In the current case, a Newton step means to find a KKT-point (u,μ) of problem (11.5)*

$$\begin{aligned} &\min_u (Df + \textstyle\sum_{k\in K} y_k Dg_k)u + \frac{1}{2}u^T D^2_{xx}Lu + \frac{1}{2}\sum_{i\in J} w_i(g_i + Dg_i\, u)^2\\ &\text{s.t.}\ \ g_k + Dg_k\, u \qquad = 0 \quad \forall\, k \in K; \end{aligned}$$

where $w_i = a_i b_i^{-1} \geq 0$ and $L = f + \langle\, y, g\rangle$. The vector v in (11.6), (11.7) is then given by

$$v_k = \mu_k (k \in K) \text{ and } v_i = -y_i + w_i(g_i + Dg_i\, u) \quad (i \in J).$$

◇

Note. It holds $y^* \geq 0$, and non-zero coefficients $w_i = a_i b_i^{-1}$ $(i \in J, y_i \geq 0)$ coincide with $w_i = r_i(1 - r_i + t_i r_i)^{-1}$ of case 3 after setting $t_i = b_i a_i^{-1}$ and $r_i = 1$. ◇

Proof. Since $(a_i, b_i) \neq 0$, (11.7) yields

$$\begin{aligned} 0 &= g_k + Dg_k\, u\ (k \in K)\\ v_i &= -y_i + w_i(g_i + Dg_i\, u)\ (i \in J). \end{aligned}$$

Replacing v_J in (11.6) we obtain

$$\begin{aligned} 0 &= D^2_{xx}Lu + D_xL + \textstyle\sum_{k\in K} v_k Dg_k + \sum_{i\in J}[-y_i + w_i(g_i + Dg_i\, u)]Dg_i\\ &= D^2_{xx}Lu + Df(x) + \textstyle\sum_{k\in K}(y_k + v_k)Dg_k + \sum_{i\in J} w_i(g_i + Dg_i\, u)Dg_i. \end{aligned}$$

So the equivalence follows by the same arguments as under case 3. □

For $k \in K$, now $y_k < 0$ is possible. Further, the convergence $z \to z^*$ yields both

$$w_i \to \infty \text{ if } y_i^* > 0, \text{ and } w_i \downarrow 0 \text{ if } g_i(x^*) < 0.$$

So the method realizes basically a penalty approach.

Case 5:

Perturbed generalized Jacobians and unperturbed Kojima function. Let the Newton step be given by

$$F(z) + P(r,t)(u,v)^T = 0, \qquad P(r,t) \in \partial F^t(z), \tag{11.8}$$

where F^t belongs again to the perturbed equation (11.2), $t_i \in \mathbb{R}$. We are now using approximations of the (unperturbed) Newton map $M = \partial F(z)$ which is justified as long as we select t (assigned to z) in such a way that $\|t\| \leq \|F(z)\|$, cf. the Remarks 10.3 and 10.4

Compared with case 3, the terms $t_i y_i^+$ do not appear, and the above proof leads us via

$$(1 - w_i t_i) y_i^+ - w_i y_i^- = y_i^+ - w_i y_i^- = y_i^+$$

directly to the modified objective

$$(Df + \textstyle\sum_{i \in K \cup J} y_i^+ Dg_i) u + \frac{1}{2} u^T D^2_{xx} L u + \frac{1}{2} \sum_{i \in J} w_i (g_i + Dg_i\, u)^2.$$

All the other conclusions of case 3 remain true after setting $t_i\, y_i^+ = 0$, i.e.,

$$\begin{array}{ll} r \in \mathcal{R}_T(y), & b_i = 1 - r_i + t_i r_i \quad \forall i \\ J = \{i \mid b_i \neq 0\}, & K = \{k \mid b_k = 0\}. \end{array}$$

This way one obtains

Lemma 11.5 (Newton steps with perturbed ∂F). *In the current case, a Newton step (11.8) means to find* a *KKT-point* (u, μ) *of the problem*

$$\begin{array}{l} \min_u D_x L u + \frac{1}{2} u^T D^2_{xx} L u + \frac{1}{2} \sum_{i \in J} w_i (g_i + Dg_i\, u)^2 \\ \text{s.t.} \;\; g_k + Dg_k\, u = 0 \;\forall k \in K, \end{array} \tag{11.9}$$

where $w_i = r_i b_i^{-1}$ *and* $L = f + \langle y^+, g \rangle$. *The vector* v *in (11.8) is then given by*

$$v_k = \mu_k (1 - t_k)(k \in K) \;\; \textit{and} \;\; v_i = b_i^{-1}(g_i + Dg_i\, u - y_i^-)(i \in J).$$

◇

Note. In comparison with (11.5) now the first derivative of the full Lagrangian appears in the objective. Setting particularly $t_i \neq 0 \;\forall\, i$ and selecting $r \in \mathcal{R}_T(y)$ with $r_i = 1$ if $y_i \geq 0$, we obtain $K = \emptyset$ as well as:

A Newton step (11.8) means to find a stationary point of

$$D_x L u + \frac{1}{2} u^T D^2_{xx} L u + \frac{1}{2} \textstyle\sum_{i \in J} t_i^{-1} (g_i + Dg_i\, u)^2 \tag{11.10}$$

where $L = f + \langle\, y^+, g\, \rangle$.
The vector v is then given by $v_i = t_i^{-1}(g_i + Dg_i\, u - y_i^-)\; \forall i$. ◇

Case 6:

Solving auxiliary problems of Wilson- type, means to apply Newton's method by using directional (or contingent-) derivatives of F:

$$F(x,y) + CF(x,y)(u,v) = 0. \tag{11.11}$$

The solutions (u,v) fulfill the same conditions as in case 3 since $CF \subset TF$. The structure of $\mathcal{R}_C(y,v)$, cf. (7.29) implies additionally that $r_i \in \{0,1\}$ and $\mu_k = r_k v_k \geq 0$. So the constrains may be written as inequalities.

Since $f, g \in C^2$, it holds $CF = F'$ (directional derivative). Now we have to solve the system

$$\begin{array}{lclcl} D^2_{xx}L(z)u & + & \sum r_i v_i Dg_i & = & -F_1 \\ Dg_i\, u & - & (1-r_j)v_i & = & -(g_i - y_i^-),\ r \in \mathcal{R}_C(y,v), \end{array} \tag{11.12}$$

where again $L = f + \sum_i y_i^+ g_i$. With the transformations (7.30) $\alpha_i = r_i v_i$ and $\beta_i = (1-r_i)v_i$ the conditions become

$$\alpha_i = 0 \text{ for } y_i < 0, \beta_i = 0 \text{ for } y_i > 0, \text{ and } \alpha_i\beta_i = 0, \alpha_i \geq 0, \beta_i \leq 0 \text{ for } y_i = 0.$$

The left side in (11.12) is

$$\begin{array}{ll} D^2_{xx}L(z)u & +Dg^T\alpha \\ Dgu & -\beta. \end{array}$$

Therefore, we are solving a linear complementarity problem, and since $\alpha_i \geq 0$ and $\beta_i \leq 0$ for $y_i = 0$, the solutions are the critical points (u,α) of the quadratic problem (now with inequality constraints)

$$\begin{array}{lll} \min & D_x Lu + \frac{1}{2}u^T D^2_{xx} Lu & \\ \text{s.t.} & g_i + Dg_i\, u \leq 0 & \text{for } y_i = 0, \\ & g_k + Dg_k\, u = 0 & \text{for } y_k > 0, \end{array}$$

where $L = f + \langle y^+, g\rangle$. The vector v is then given by

$$\begin{array}{lcll} v_i & = & \alpha_i + g_i + Dg_i\, u & \text{for } y_i \geq 0, \\ v_i & = & -y_i + g_i + Dg_i\, u & \text{for } y_i < 0. \end{array}$$

So we are applying, as before, a method of sequentially quadratic approximation.

Basically, the present one is *Wilson's method* which has been originally developed under the strict complementarity assumption, **LICQ** and the strong sufficient second order condition (i.e., positive definiteness of the Hessian on the (y^+)- **tangent** space). We investigate condition (CI).

Let S have the same meaning as in case 1. Recall that S is active *near* z if $F^S \equiv F$ holds on some neighborhood of z. We put

$$S \in I^e(z) \Leftrightarrow S \text{ is active } \textit{near } z' \text{ for certain } z' \to z.$$

Lemma 11.6 (condition (CI) in Wilson's method). *Condition (CI) for method (11.11) means equivalently that*

$$DF^S(z^*) \textit{ is regular for each } S \in I^e(z^*).$$

◇

Proof. Assume (CI) holds true. We show first

$$\|DF^S(z)^{-1}\| \leq 1/c \text{ if } S \text{ is active near } z \in z^* + \varepsilon B^0 \text{ and } \varepsilon \text{ is small enough.} \tag{11.13}$$

Indeed, for $\|z - z^*\| < \varepsilon$, the terms $CF(z)(u, v)$ in (11.11) satisfy

$$\inf \|CF(z)(\text{bd}\, B)\| \geq c > 0$$

If S is active near $z \in z^* + \varepsilon B^0$, then from $CF(z) = DF^S(z)$, it follows $\|DF^S(z)^{-1}\| \leq 1/c$.

Conversely, we show that (11.13) ensures (CI). The set Ω of the points z under consideration in (11.13), is dense in $z^* + \varepsilon B^0$ (shown in the proof of Lemma 6.17). We thus conclude (by continuity arguments) that all $P \in D^\circ F(z')$ for $z' \in z^* + \varepsilon B$ satisfy $\|P^{-1}\| \leq 1/c$.

Formula (6.33) in Section 6.4.2 now yields $CF(z')(u, v) \subset D^\circ F(z')(u, v)$ and ensures (CI) because of

$$\inf \|CF(z')(\text{bd}\, B)\| \geq \inf \|D^\circ F(z')(\text{bd}\, B)\| \geq c \qquad \forall z' \in z^* + \varepsilon B.$$

So (CI) and (11.13) are equivalent.

Taking into account that $I^e(z) \subset I^e(z^*)$ for $z \in z^* + \varepsilon B$ and small ε, condition (CI) can be equivalently written as regularity of all $DF^S(z^*)$ for $S \in I^e(z^*)$.

This Page Intentionally Left Blank

Chapter 12

Basic Examples and Exercises

12.1 Basic Examples

Example BE.0
A pathological real Lipschitz function (lightning function).

We present a simple construction of a special real Lipschitz function G such that F.H. Clarke's subdifferential fulfills $\partial G(x) \equiv [-1, 1]$. The existence of such functions has been clarified in [BMX94].

It will be seen that the following sets are dense in $\mathbb{R}$:
the set $D_N = \{x|$ G is not directionally differentiable at $x\}$,
the set of local minimizers, and the set of local maximizers.

To begin with, let $U : [a, b] \to \mathbb{R}$ be any affine-linear function with Lipschitz rank $L(U) < 1$, and let $c = \frac{1}{2}(a + b)$. As the key of the following construction, we define a linear function V by

$$V(x) = \begin{cases} U(c) - a_k(x - c) & \text{if } U \text{ is increasing,} \\ U(c) + a_k(x - c) & \text{otherwise.} \end{cases}$$

Here,

$$a_k := \frac{k}{k+1},$$

and k denotes the step of the (later) construction. Given any $\varepsilon \in (0, \frac{1}{2}(b - a))$ we consider the following 4 points in $\mathbb{R}^2$:

$$p_1 = (a, U(a)),\ p_2 = (c - \varepsilon, V(c - \varepsilon)),\ p_3 = (c + \varepsilon, V(c + \varepsilon)),\ p_4 = (b, U(b)).$$

By connecting these points in natural order, a piecewise affine function

$$w(\varepsilon, U, V) : [a, b] \to \mathbb{R}$$

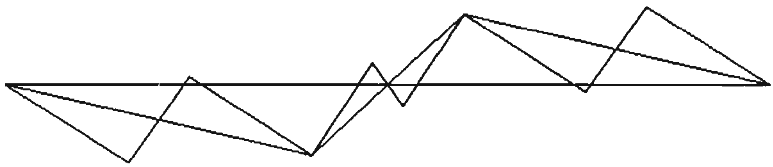

Figure 12.1: Lightning function: principle of successive construction.

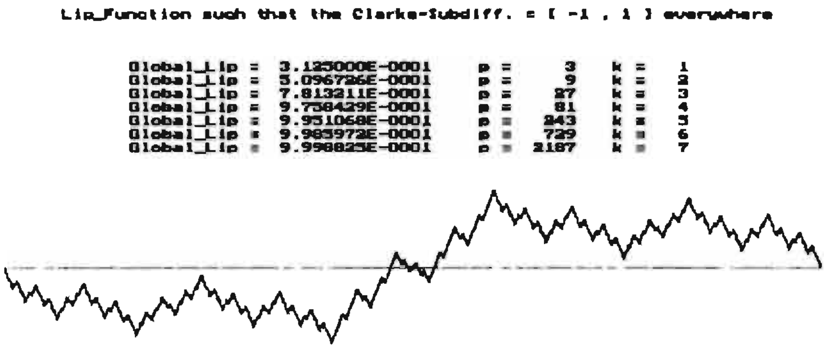

Figure 12.2: Lightning function: g_k for $k = 7$ ($x = 1$ lies outside).

is defined. It consists of 3 affine pieces on the intervals

$$[a, c - \varepsilon], [c - \varepsilon, c + \varepsilon], [c + \varepsilon, b].$$

By the construction of V and $p_1, \ldots, p_4$, it holds

$$\text{Lip}\,(w(\varepsilon, , U, V)) < 1 \text{ provided that } \varepsilon \text{ is small.}$$

After taking ε in this way, we may repeat our construction (like defining Cantor's set) with each of the related 3 pieces and larger k, see Figures 12.1 and 12.2.

Now, start this procedure on the interval [0, 1] with the initial function

$$U(x) = 0 \quad \text{and} \quad k = 1.$$

In the next step $k = 2$ we apply the construction to the 3 pieces just obtained, then with $k = 3$ to the now existing 9 pieces and so on.

The concrete choice of the (feasible) $\varepsilon = \varepsilon(k) > 0$ is not important in this context. We obtain a sequence of piecewise affine functions g_k on [0,1] with

Lipschitz rank < 1. This sequence has a cluster point g in the space $C[0,1]$ of continuous functions, and g has the Lipschitz rank $L = 1$. Let

$$N_k = \{y \in (0,1) \mid g_k \text{ has a kink at } y\} \quad \text{and } N \text{ be the union of all } N_k.$$

If $y \in N_k$, then the values $g_i(y)$ will not change during all forthcoming steps $i > k$. Hence $g(y) = g_k(y)$. The set N is dense in [0,1].

Connecting arbitrary 3 neighbored kink-points of g_k and taking into account that these points belong to the graph of g, one sees that g has a dense set of local minimizers (and maximizers).

Further, let D be the dense set of all centre points c belonging to some subinterval used during the construction. Then each $y \in D$ is again a centre point of some subinterval $I(k)$ for each step with sufficiently large k. Thus, $g(y) = g_k(y)$ is again true.

Moreover, for arbitrary $\delta \in (0,1)$, one finds points

$$\begin{aligned} &y', y'' \in (y, y+\delta) \text{ such that } y', y'' \in N \\ &\text{and } g(y') - g(y) > (1-\delta)(y'-y) \text{ as well as} \\ &g(y'') - g(y) < -(1-\delta)(y''-y); \end{aligned}$$

namely the nearest kinks of g_k on the right side of y where k is (large and) odd or even, respectively. This shows that directional derivatives $g'(y;1)$ cannot exist for $y \in D$. In addition, by the mean-value theorem for Lipschitz functions [Cla83], one obtains $\partial g(x) = [-1,1]\, \forall x \in (0,1)$.

To finish the construction define G on $\mathbb{R}$ by setting $G(x) = g(x - \text{integer}(x))$, where integer$(x)$ denotes the integer part of x. Needless to say that G is also nowhere semismooth.

Derived functions: Let $h(x) = \frac{1}{2}(x + G(x))$. Then $\partial h(x) = [0,1]$ for all x, h is strictly increasing, has a continuous inverse h^{-1} which is nowhere locally Lipschitz, and h is not directionally differentiable on a dense subset of $\mathbb{R}$. In the negative direction -1, h is strictly decreasing, but Clarke's directional derivative $h^c(x;-1)$ is identically zero. The integral

$$F(t) = \int_0^t h(x)\,dx$$

is a convex $C^{0,1}$ function with strictly increasing derivative h, such that

$$0 \in Th(t)(1) = [0,1]\ \forall t \text{ and } 0 \in Ch(t)(1) \text{ for all } t \text{ in a dense set}$$

holds true.

Example BE.1

Alternating *Newton sequences for real, Lipschitzian f with almost all initial*

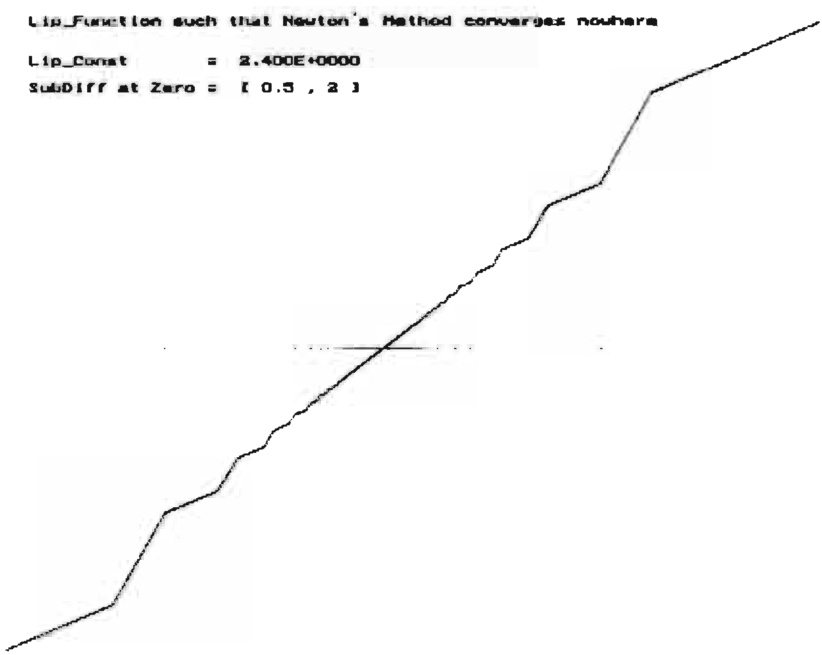

Figure 12.3: Alternating Newton sequences for Lipschitzian f.

points.

To construct $f : \mathbb{R} \to \mathbb{R}$, consider intervals $I(k) = [k^{-1}, (k-1)^{-1}] \subset \mathbb{R}$ for integers $K \geq 2$, and put

$$\begin{aligned} c(k) &= \tfrac{1}{2}[k^{-1} + (k-1)^{-1}] && \text{(the center of } I(k)\text{)} \\ c(2k) &= \tfrac{1}{2}[(2k)^{-1} + (2k-1)^{-1}] && \text{(the center of } I(2k)\text{).} \end{aligned}$$

In the (x, y)-plane, define

$g_k = g_k(x)$ to be the linear function through the points $((k-1)^{-1}, (k-1)^{-1})$ and $(-c(k), 0)$,

i.e.,

$$g_k(x) = a_k(x + c(k)), \text{ where } a_k = (k-1)^{-1}/[(k-1)^{-1} + c(k)].$$

Similarly, let

$h_k = h_k(x)$ be the linear function through the points (k^{-1}, k^{-1}) and $(c(2k), 0)$, i.e.,

$$h_k(x) = b_k(x - c(2k)), \text{ where } b_k = k^{-1}/[k^{-1} - c(2k)].$$

Evidently, $g_k = 0$ at $x = -c(k), h_k = 0$ at $x = c(2k)$. Now define f for $x > 0$ as

$$f(x) = \min\{g_k(x), h_k(x)\} \text{ if } x \in I(k) \quad \text{and} \quad f(x) = g_2(x) \text{ if } x > 1.$$

We finish the construction by setting $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0$.

The related properties can be seen as follows:

For $k \to \infty$, one obtains $\lim a_k = \frac{1}{2}$ and $\lim b_k = 2$. The assertion $Df(0) = 1$ can be directly checked. Again directly, one determines the global Lipschitz rank

$$L = \max b_k = b_2 = \frac{1}{2}/[\frac{1}{2} - \frac{1}{2}(\frac{1}{4} + \frac{1}{3})] = \frac{12}{5}.$$

One the left side of the interval $I(k), f$ coincides with h_k, one the right with g_k. Since $g_k(c(k)) < h_k(c(k)), f$ coincides with g_k on a small neighborhood of the center point $c(k)$.

Now, let us start Newton's method a some $x^0 \in \Theta^1$. Then the next iterate x^1 is some point $\pm c(k) \in \Theta^1$. There, it holds $Df = Dg_k$ (or $Df = -Dg_k$ for negative arguments). Hence, the method generates the alternating sequence $x^0, x^1, x^2 = -x^1, x^3 = x^1, ...$

Example BE.2

A function f which is one of the *simplest nonsmooth, nonconvex functions on a Hilbert space. Pseudo-regularity of the map* $F(x) = \{y \in \mathbb{R} \mid f(x) \le y\}$ *can be easily* shown. However, the sufficient conditions of Section 3.3 in terms of contingent derivatives and coderivatives will not be satisfied.

Let

$$X = l^2, \quad x = (x_1, x_2, ...) \text{ and } f(x) = \inf_k x_k.$$

Now $F^{-1}(y) = \{x \in X \mid f(x) \le y\}$ is the level set map of a globally Lipschitz functional. Since f is concave the directional derivatives $f'(x; u)$ exist everywhere. Further, f is monotone with respect to the natural vector ordering, and f is nowhere positive.

(i) The mapping F is (globally) pseudo-regular, e.g., with rank $L = 2$. Indeed, if $f(x) \le y$ and $y' < y$, there is some k such that $x_k < y + \frac{1}{2}|y' - y|$. Put $x' = x - 2|y' - y|e^k$ where e^k is k-th unit vector in l^2. Then, pseudo-regularity follows from $\|x' - x\| \le 2|y' - y|$ and $x' \in F^{-1}(y')$ since $f(x') \le x'_k \le y - \frac{3}{2}|y' - y| \le y'$.

(ii) Next we are going to show that, at each ξ with $\xi_k > f(\xi)\ \forall k$, it holds

$$f'(\xi; u) \ge 0 \forall u \in X \tag{12.1}$$

in spite of (uniform) pseudo-regularity of F. We show even more:

(iii) If $f(\xi + tu) \leq f(\xi) - t$ for certain $t \downarrow 0$ and bounded u, say for $\|u\| \leq C$, then $u = u(t)$ *necessarily* depends on t and there is no (strong) accumulation point of $u(t)$.

In fact, by the choice of ξ, we have $\xi_k > f(\xi) = \inf_n \xi_n = 0$ and $\xi_k + tu_k < -\frac{1}{2}t$ for some k. Due to $|u_k| \leq C$, the latter inequality can never hold for $t \downarrow 0$ if $k = k(t)$ is bounded. So one obtains $u_{k(t)} < -\frac{1}{2}$ for an infinite number of components.
Assuming u to be fixed, this yields the contradiction $u \notin l^2$. Hence u depends on t. Assuming convergence $u(t) \to u^0$ for certain $t \downarrow 0$, we obtain again a contradiction, namely

$$\liminf_{t\downarrow 0} u(t)_{k(t)} \leq -\frac{1}{2} \text{ for certain } k(t) \to \infty,$$
$$\text{though } u^0 \in l^2 \text{ yields } \lim_{k\to\infty} u^0_k = 0.$$

Finally, we consider ε-**normals** of F.

(iv) The point $(x^*, y^*) = (e^m, -1)$ is an ε-**normal** to gph F around $(x, y) = (-\varepsilon e^m, -\varepsilon)$.

Clearly, $(x, y) \in \operatorname{gph} F$ due to $f(x) = -\varepsilon$.
Setting $x' = x + tu, u \in \operatorname{bd} B, t \geq 0, y' = f(x') + s, s \geq 0$, the condition (3.14) for ε-**normals** requires

$$t\langle x^*, u\rangle \leq s + f(x + tu) - f(x) + t\varepsilon + \varepsilon|s + f(x + tu) - f(x)|$$

provided that $\|u\| = 1$ and t, s are small, say $\max\{s, t\} < \delta < \varepsilon$. But then we have $f(x + tu) - f(x) = tu_m$, and condition (3.14) becomes

$$t\langle x^*, u\rangle = tu_m \leq s + tu_m + t\varepsilon + \varepsilon|s + f(x + tu) - f(x)|.$$

Since $s \geq 0$ is small, this condition is always true, so our assertion is valid.

Further, since m was arbitrary, we may put $m > 1/\varepsilon$ as $\varepsilon \downarrow 0$ in order to obtain $x^* \to 0$ (weak*). Thus, $0 \in D^*F(0,0)(-1)$.

(v) For the points $z = (\xi, f(\xi)), \xi$ from (ii) and for $0 < \varepsilon < \frac{1}{2}$, there is no weak* or strong accumulation point (ξ^*, η^*) of ε-**normals** (x^*, y^*) to gph F around z with $|\eta^*| = 1$.
To verify this, we show that $|y^*| \leq 2\varepsilon$.

Due to $f(\xi) = 0$, condition (3.14) particularly requires that

$$t\langle x^*, u\rangle + y^*[s + f(\xi + tu)] \leq t\varepsilon + \varepsilon|s + f(\xi + tu)| \tag{12.2}$$

holds for small $s, t > 0$ and all $u \in 2\operatorname{bd} B$.
With $t = 0$ and small $s > 0$, this implies $y^* \leq \varepsilon$.
It remains to consider negative y^*. Then we select $u = -2e^k$ with large $k = k(t)$ such that $f(\xi + tu) < -t$. Now (12.2) yields the assertion (setting $s = 0$ and using that f has Lipschitz rank 1),

$$-2tx^*_{k(t)} + |y^*|t \leq t\varepsilon + \varepsilon|t|, \text{ hence } |y^*| \leq 2\varepsilon.$$

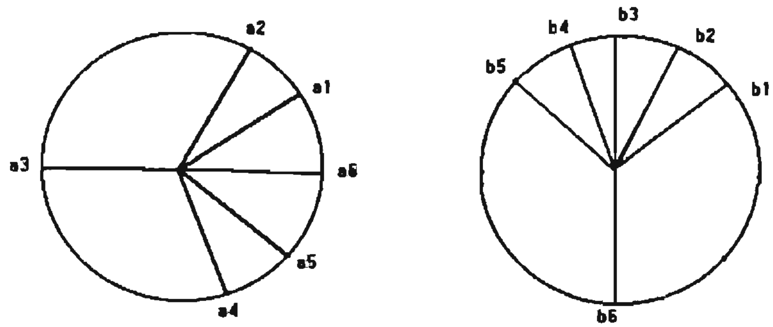

Figure 12.4: Piecewise linear bijection of $\mathbb{R}^2$ with $0 \in \partial f(0)$.

While (iv) says that D*F(0,0) is not injective, the property (v) indicates that the sets $D^*F(0,0)(\eta^*)$ for $|\eta^*| = 1$ tell us nothing about ε**−normals** at points $(\xi, f(\xi))$ from(ii).

Example BE.3
Piecewise linear bijection of $\mathbb{R}^2$ *with* $0 \in \partial f(0)$.

On the sphere of $\mathbb{R}^2$, let vectors a^k and $b^k (k = 1, 2, ..., 6)$ be arranged as follows:

Put $a^7 = a^1, b^7 = b^1$ and notice the following important properties:

(i) $a^1 = b^1, a^2 = b^2; a^4 = -b^4, a^5 = -b^5$.

(ii) The vectors a^k and b^k turn around the sphere in the same order.

(iii) The cones K_i generated by a^i and a^{i+1}, and P_i generated by b^i and b^{i+1}, are proper.

Let $L_i : \mathbb{R}^2 \to \mathbb{R}^2$ be the unique linear function satisfying $L_i(a^i) = b^i$ and $L_i(a^{i+1}) = b^{i+1}$. Setting $f(x) = L_i(x)$ if $x \in K_i$ we define a piecewise linear function which maps K_i onto P_i. By the construction, f is surjective and has a well-defined inverse; hence it is a (piecewise linear) Lipschitzian homeomorphism of $\mathbb{R}^2$. Moreover, $f = id$ on int K_1 and $f = -id$ on int K_4.

Thus, $\partial f(0)$ contains the unit-matrix *E as* well as *–E* and, by convexity, the zero-matrix, too.

Example BE.4
A piecewise quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ having *pseudo-Lipschitzian* stationary points being not unique.

We put $z = (x, y) \in \mathbb{R}^2$ in polar-coordinates,

$$z = r(\cos\phi + i \sin\phi),$$

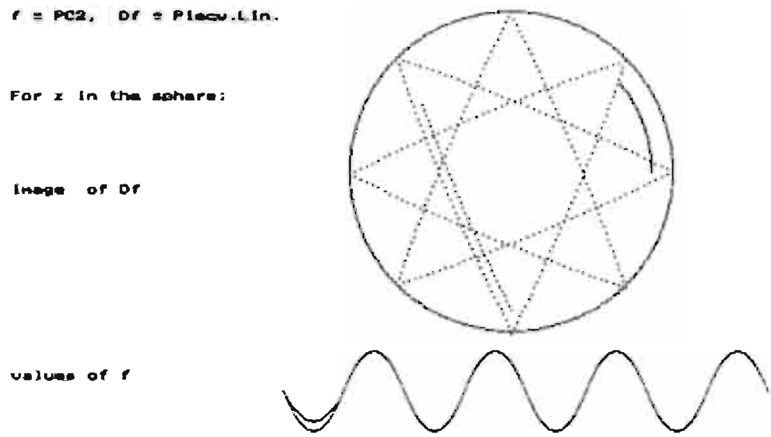

Figure 12.5: The image of Df and the behavior of the values $f(z)$ when z turns around the sphere.

and describe f as well as the partial derivatives $D_x f, D_y f$ over the 8 cones

$$C(k) = \{z \mid \phi \in [\frac{1}{4}(k-1)\pi, \frac{1}{4}k\pi]\}, \ (1 \le k \le 8),$$

by

cone	f	$D_x f$	$D_y f$
$C(1)$	$y\,(y-x)$	$-y$	$2y-x$
$C(2)$	$x\,(y-x)$	$-2x+y$	x
$C(3)$	$x\,(y+x)$	$+2x+y$	x
$C(4)$	$-y\,(y+x)$	$-y$	$-2y-x$

and on the remaining cones $C(k+4)$, $(1 \le k \le 4)$, f is defined as in $C(k)$.

Studying the Df-image of the sphere, it is not difficult to see (but needs some effort) that Df is continuous and Df^{-1} is pseudo-Lipschitz at the origin. For $a \in \mathbb{R}^2 \setminus \{0\}$, there are exactly 3 solutions of $Df(z) = a$.

Example BE.5

A Lipschitz function $f : [0, \frac{1}{2}) \to C$ such that directional derivatives f' nowhere exist, neither as strong nor weak (pointwise) limits; and contingent derivatives are empty.

For $x \in [0, \frac{1}{2})$ define a continuous function $h_x : [0,1] \to \mathbb{R}$ by

$$h_x(t) = \begin{cases} 0 & \text{for } 0 \le t < x \\ t - x & \text{for } x \le t < 2x \\ x & \text{for } 2x \le t \le 1 \end{cases}$$

The mapping $f(x) := h_x$ is a Lipschitz function from the interval $[0, \frac{1}{2})$ into C[0, 1]. For small $|\lambda| > 0$, consider the function

$$g(x, \lambda) = (f(x + \lambda) - f(x))/\lambda.$$

If $\lambda > 0$, then

$$g(x, \lambda)(2x) \leq 0 \text{ and } g(x, \lambda)(2x + 2\lambda) = 1.$$

Hence, the limit $\lim g(x, \lambda)$ (as $\lambda \downarrow 0$) cannot exist in $C[0,1]$ (neither in a strong nor in a weak sense). If $\lambda < 0$, then we obtain for $x > 0$ that

$$g(x, \lambda)(2x) \geq 0 \text{ and } g(x, \lambda)(2x + 2\lambda) = -1.$$

Thus $\lim g(x, \lambda)$ (as $\lambda \uparrow 0$) cannot exist, too.

This shows that f is a Lipschitz function without directional derivatives and with empty contingent derivatives for nontrivial directions.

Example BE.6

A convex function $f : \mathbb{R} \to \mathbb{R}$, non-differentiable on a dense set.

Consider all rational arguments $y = \frac{p}{q} \in (0, 1]$ such that p, q are positive integers, prime to each other, and put

$$h(y) = \frac{1}{q!}.$$

For fixed q, the sum $S(q)$ over all feasible $h(y)$ is bounded by

$$S(q) \leq \frac{q}{q!} \quad \text{and} \quad \sum_q S(q) = c < \infty.$$

Now define

$$g_1(x) \text{ by } g_1(0) = 0 \text{ and } g_1(x) = \textstyle\sum_{y \leq x} h(y) \text{ for } x \in (0, 1].$$

Then g_1 is increasing, bounded by c and has jumps of size $(q!)^{-1}$ at $x = y$.

Next extend g_1 on $\mathbb{R}_+$ by setting

$$g(x) = kg_1(1) + g_1(x - k) \text{ if } x \in [k, k + 1),\ k = 1, 2, \ldots$$

and put $g(x) = -g(-x)$ for $x < 0$. Since g is increasing, the function

$$f(t) = \int_0^t g(x)\, dx \text{ as Lebesgue integral}$$

is convex and for $t \downarrow y$ and $t \uparrow y$ (t irrational, y rational) one obtains different limits of $Df(t)$. Thus f is not differentiable at y.

12.2 Exercises

Exercise 1
Proof of Lemma 2.21.

(i) ($\Leftarrow$) Let F be pseudo-regular at (x^0, y^0) with rank L and neighborhoods U, V. For fixed $y^* \in \mathrm{bd}\, B^*$, then the mapping $F_\phi(x) = \{r \in \mathbb{R} \mid \phi(x) \le r\}$ is again pseudo-regular at $(x^0, \phi(x^0))$ with rank L and the same neighborhoods U, V.

The second part of the proof to Lemma 2.18 now shows

$$U \cap E_\phi(p) = \emptyset \quad \text{if } p < L^{-1} \text{ and } \phi(z) \in V.$$

Since $y^* \in \mathrm{bd}\, B^*$ was arbitrary, this yields

$$x^0 \notin \limsup_{p \downarrow 0} (\cup_{y^* \in \mathrm{bd}\, B^+} E_\phi(p)).$$

(i) ($\Rightarrow$) By Lemma 2.20, there are points z and y, depending on $p \downarrow 0$, such that $z \to x^0$, $y \to y^0$, $f(z) \ne \pi(z, y)$ and $z \in E_v(p)$, where $y^* \in \mathrm{bd}\, B^*$.

Since $\pi(x, y) = y$, now $z \in E_v(p)$ means:

$$\langle y^*, f(x) - y\rangle + pd(x, z) \ge \langle y^*, f(z) - y\rangle \text{ if } d(x, z) < \alpha(z),$$

hence $z \in E_\phi(p)$.

(ii) ($\Leftarrow$) F is not pseudo-regular at (x^0, y^0) because of (i).

(ii) ($\Rightarrow$) The condition holds true due to Lemma 2.20 (ii).

Exercise 2
How the situation of mixed constraints (equations and inequalities) can be handled in a similar manner?

Define the cone $C = \{y \mid y_i \ge 0$ if i corresponds to an inequality, $y_i = 0$ if i corresponds to an equation$\}$.

Exercise 3
Verify that, for $m > n$, every function $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is nowhere pseudo-regular. Hint: Apply Rademacher's theorem.

Assume the contrary. Take x near x^0 such that $Df(x)$ exists (Rademacher's theorem). Since $m > n$, there exists $\mu \ne 0$ such that $\sum_i \mu_i Df_i(x) = 0$. The function

$$\phi(\xi) = \sum_i \mu_i Df_i(\xi)$$

has derivative 0 at x. Thus, x is a local Ekeland-point for ϕ with each $p > 0$. Apply Lemma 2.21 (ii) to obtain a contradiction.

Exercise 4
Show how Theorem 2.26 may be extended to the case of a closed multifunction $F : X \rightrightarrows Y$. What about necessity of the conditions in Theorem 2.26?

One can repeat all the arguments of the necessity part for Theorem 2.22.

Exercise 5
Show that, in the Lemmas 3.1 and 3.2, one may replace "Lipschitzian "l.s.c." *by* "l.s.c.".

Proof: If F^{-1} is l.s.c, at (y^0, x^0) without being Lipschitz l.s.c. then, for certain $y^k \to y^0 (y^k \neq y^0)$ we find $x^k \in F^{-1}(y^k)$ with $t_k := \|x^k - x^0\| \to 0$ and $t_k / \|y^k - y^0\| \to \infty$.

Since $X = \mathbb{R}^n$ we obtain, via $u^k = (x^k - x^0)/t_k$ a convergent (sub)sequence which shows that $0 \in CF(z^0)(u)$ for some $u \neq 0$. So, already the necessary injectivity conditions for the related regularity are violated.

Exercise 6
Show that for $f \in C^1$, one has $D^* f(x) = -Df(x)^{\mathsf{T}}$.

The simplest way is to use Theorem 6.5. For $f \in C^1$, statement (ii) yields $\langle v^*, Df(x)(u)\rangle + \langle u^*, u\rangle \leq 0$ for all $u \in B$, so the linear function

$$u \to \langle v^*, Df(x)(u)\rangle + \langle u^*, u\rangle$$

is identical zero. But this means just $u^* = -D^* f(x)^{\mathsf{T}}(v^*)$.

Exercise 7
Proof of Lemma 5.11.

By definition, f is continuous, and $f(x) = h_k(x)$ is always true for some h_k where $k = k(x)$ and h_k belongs to a finite family F of C^1-functions. We may assume that the sets $X_i = \{x | f(x) = h_i(x)\}$ fulfill $x^0 \in \text{cl}\,(\text{int}\, X_i)$, otherwise the local representation of f (near x^0) by the family F would need less functions h_k. Applying Theorem 5.1 to $x \in \text{int}\, X_i \cap (x^0 + \alpha^{-1} B)$ we obtain $\|Dh_i(x)^{-1}\| \leq 1/\alpha$ for some large α. Hence, locally, $\text{gph}\, f^{-1}$ consists of arcs belonging to the *strongly regular,* inverse functions h_i^{-1} :

$$\exists \gamma > 0 : f^{-1}(y) \cap (x^0 + \gamma B) \subset \{h_i^{-1}(y) \,|\, i \in I\} \text{ for sufficiently small } \|y\|.$$

Thus, x^0 is an isolated zero, and H exists as required.

Exercise 8
Analyze the continuity properties for $F(y) = \{x \in B \subset \mathbb{R}^2 |\; \|x - y\| \geq \frac{1}{2}\}$ for

F with the Euclidean norm and with polyhedral norms, respectively, and *with "$>$" instead of "$\geq$". Does there exists a continuous function $f : B \to B$ such that $f(\cdot) \in F(\cdot)$ on B?*

The first part is left to the reader. A continuous function with $f(\cdot) \in F(\cdot)$ on B cannot exists because is would have a fixed point.

Exercise 9

Find a counterexample $(n = m = 2)$ showing that the pointwise condition (5.4) in Theorem 5.1 is not sufficient for the *Lipschitz l.s.c. of* F^{-1}.

We construct $f : \mathbb{R}^2 \to \mathbb{R}^2$ continuous with

$$f'(0; u) = u \;\; \forall u \in \mathbb{R}^2 \;\text{ and }\; 0 \notin \operatorname{int} f(\mathbb{R}^2).$$

Let

$$M = \{(x,y) \in \mathbb{R}^2 \mid |y| \geq x^2 \text{ if } x \geq 0,\, x^2 + y^2 \leq 1,\, x \leq \tfrac{1}{2}\}$$

and $G = \operatorname{conv} M$. For $(x,y) \in M$, let $f(x,y) = (x,y)$. For $(x,y) \in G \backslash M$ and $y > 0$ put $f(x,y) = (x,x^2)$.

In order to define f at $(x,y) \in G\backslash M$ with $0 > y > -x^2$, let D be the triangle given by the points

$$P^1 = (x,-x^2),\; P^2 = (0,0),\; P^3 = (x,x^2) \text{ and let } t = t(x,y) := \frac{-y}{x^2}.$$

Then $t \in (0,1)$. We shift the point $(x, t(-x^2) + (1-t)x^2)$ to the left boundary of D and define f to be the related point:

$$f(x,y) = \begin{cases} (2t-1)(x,-x^2) & \text{for } t \geq \frac{1}{2}, \\ (1-2t)(x,x^2) & \text{for } t \leq \frac{1}{2}. \end{cases}$$

So f becomes a continuous function of the type $\mathbb{R}^2 \to M$. Setting $g(z) = f(\pi(z))$ where $\pi(z)$ is the projection of z onto G, f can be continuously extended to the whole space. We identify f and g. Clearly, $f'(0;u) = u$ holds for all u, and $0 \notin \operatorname{int} f(\mathbb{R}^2)$.

Exercise 10

Show that if $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x^0, f(x^0))$ and directionally differentiable for x near x^0 then the local inverse f^{-1} is directionally differentiable for y near $f(x^0)$.

Otherwise one finds images $y = f(x)$ for x near x^0 and $v \in \mathbb{R}^n$ such that $Cf^{-1}(y)(v)$ contains at least two different elements p and q. Since f' exists and $p \in Cf^{-1}(y)(v)$ iff $v \in Cf(x)(p)$, one obtains $f'(x;p) = v = f'(x;q)$. For small $t > 0$, then the images

$$f(x+tp) - f(x+tq) = f(x+tp) - f(x) - (f(x+tq) - f(x))$$

differ by a quantity of type $o(t)$ while the pre-images differ by $t(p-q)$. Therefore, the local inverse f^{-1} cannot be Lipschitz near $(f(x^0), x^0)$ for $p \neq q$.

Exercise 11

Verify Theorem 6.4, first part "polyhedral".

The statements (i) and (ii) can be easily seen for each submapping F_k ,defined by gph $F_k = P_k$, since P_k is a convex polyhedron. From gph $F = \cup_k \text{gph}\, F_k$ and (6.2) then the assertions follow via selection of subsequences assigned to fixed F_k.

Exercise 12

Verify

(i) If f or g is directionally differentiable, then $Ch(x)(u) = Cf(y)(Cg(x)(u))$.
(ii) If $f \in C^1$ then $Th(x)(u) = Df(y)(Tg(x)(u))$.
(iii) If $g \in C^1$ and g^{-1} is l.s.c. at (y, x) then $Th(x)(u) = Tf(y)(Dg(x)(u))$.

Proof. Note that the functions are locally Lipschitz by assumption.

(i) g directionally differentiable:

Let $v = g'(x;u)$ and $w = \lim t^{-1}[f(y+tv) - f(y)]$ for certain $t \downarrow 0$. Write

$$g(x+tu) = g(x) + tv + o(t) = y + tv + o(t)$$

(possible *with given* t, since g is directionally differentiable). Then,

$$\begin{aligned} w &= \lim t^{-1}[f(g(x+tu) - o(t)) - f(g(x))] \\ &= \lim t^{-1}[f(g(x+tu) - f(g(x))] \in Ch(x)(u). \end{aligned}$$

f directionally differentiable:

Let $v \in Cg(x)(u)$ and $w = f'(y;v)$. We may write, with certain $t \downarrow 0$,

$$g(x+tu) = g(x) + tv + o(t) \text{ and } w = \lim t^{-1}[f(g(x) + tv) - f(g(x))].$$

Using again $g(x) + tv = g(x+tu) - o(t)$ we get $w \in Ch(x)(u)$ as above.

(ii) Let $w = Df(y)v$, and let $v \in Tg(x)(u)$ be written as

$$v = \lim v';\ v' = t^{-1}[g(x'+tu) - g(x')],\ x' \to x,\ t \downarrow 0.$$

Then $g(x'+tu) = g(x') + tv'$. Since $f \in C^1$, it holds

$$w = \lim s^{-1}[f(y'+sv') - f(y')] \text{ for all } s \downarrow 0 \text{ and } y' \to y.$$

Setting $y' = g(x')$ and $s = t$ this yields $w = \lim t^{-1}[f(g(x'+tu)) - f(g(x'))] \in Th(x)(u)$.

(iii) Let $v = Dg(x)u$, and let $w \in Tf(y)(v)$ be written as

$$w = \lim t^{-1}(f(y' + tv) - f(y')) \text{ with } y' \to y, t \downarrow 0.$$

Since g^{-1} is l.s.c. at (y, x), one finds $x' \to x$ such that $y' = g(x')$. Substituting, we obtain $w = \lim t^{-1}(f(g(x') + tv) - f(g(x')))$ and, since $g(x') + tv = g(x' + tu) + o(t)$, it follows $w \in Th(x)(u)$ as required.

Exercise 13
Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be strongly regular at $(x^*, 0)$. Show that f^{-1} is semi-smooth at 0 if so is f at x^*.

Otherwise, $\partial(f^{-1})$ is not a Newton map at 0. Then, due to $\operatorname{conv} Tf^{-1} = \partial(f^{-1})$ (cf. (6.17)), also Tf^{-1} is not a Newton map at 0.

So one finds some $c > 0$ and elements $u \in Tf^{-1}(y)(y - 0)$ such that

$$\|u - (f^{-1}(y) - f^{-1}(0))\| > c\|y\| \quad \text{where } u = u(y) \text{ and } y \to 0.$$

Setting $x = f^{-1}(y)$ and using that f and f^{-1} are locally Lipschitz, we obtain with some new positive constant C :

$$\|u - (x - x^*)\| \geq C\|x - x^*\|.$$

Since Tf is a Newton map at x^*, we may write (with different o– functions)

$$Tf(x)(x - x^*) \subset f(x) - f(x^*) + o(x - x^*)B = y + o(x - x^*)B.$$

Next apply that $u \in Tf^{-1}(y)(y) \Leftrightarrow y \in Tf(x)(u)$ By *subadditivity* of the homogeneous map Tf (cf. (6.10)), we then observe

$$\begin{aligned} y \in Tf(x)(u) &\subset Tf(x)(u + x^* - x) + Tf(x)(x - x^*) \\ &\subset Tf(x)(u + x^* - x) + y + o(x - x^*)B. \end{aligned}$$

Hence

$$0 \in Tf(x)(u + x^* - x) + w \quad \text{with certain } w \in o(x - x^*)B.$$

We read the latter as

$$u + x^* - x \in Tf^{-1}(y)(-w)$$

which yields, with some Lipschitz rank L of f^{-1} near the origin,

$$C\|x - x^*\| \leq \|u - (x - x^*)\| \leq L\|w\| \leq Lo(x - x^*).$$

This contradiction proofs the statement.

Exercise 14
Show that $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is C^1 on an open set Ω if card $Tf(x)(u) = 1$ for *all $x \in \Omega$ and $u \in \mathbb{R}^n$.*

By its basic properties, now the map $Tf(x)$ is additive and homogeneous for each $x \in \Omega$. In addition, as a locally bounded and closed mapping, Tf is continuous on Ω.

Exercise 15
Verify that positively homogeneous $g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ are simple *at the origin.*

Let $v \in Tg(0)(r)$ and $t_k \downarrow 0$ be given $(k = 1, 2, ...)$. We know by the structure of $Tg(0)(r)$ that there exist q_k such that $v_k := g(q_k + r) - g(q_k) \to v$.

Given k select some $\nu > k$ such that $\|t_\nu q_k\| < 1/k$ and put $p_\nu = t_\nu q_k$. Then

$$v_k = t_\nu^{-1}[g(t_\nu q_k + t_\nu r) - g(t_\nu q_k)] = t_\nu^{-1}[g(p_\nu + t_\nu r) - g(p_\nu)].$$

Next select $k' > \nu$ and choose a related $v' > k'$ in the same way as above. Repeating this procedure, the subsequence of all $s \in \{t_\nu, t_{\nu'}, t_{\nu''}, ...\}$ then realizes, with the assigned $p(s) \in \{p_\nu, p_{\nu'}, p_{\nu''}, ...\}, v = \lim s^{-1}[g(p(s) + sr) - g(p(s))]$ and $p(s) \to 0$.

Exercise 16
Show that, *for $m = 1$, the situation* $\text{conv}\,\{g(r), -g(-r)\} \not\subset \text{cl} \cup_{y \neq 0} Tg(y)(r)$ *must be taken into account.*

The situation $\text{conv}\,\{g(r), -g(-r)\} \not\subset \text{cl} \cup_{y \neq 0} Tg(y)(r)$ occurs once more for $g(y) = |y|$.

Appendix

In this section, we present proofs of often applied (and well-known) basic tools for convenience of the reader.

Ekeland's Variational Principle

Theorem A.1 (Ekeland's variational principle, appears also as Theorem 2.12). *Let X be a complete metric space and $\phi : X \to \mathbb{R} \cup \{\infty\}$ be a l.s.c. function having a finite infimum. Let ε and α be positive, and let $\phi(x) \le \varepsilon + \inf_X \phi$. Then there is some $z \in X$ such that*

$$d(z,x) \le \alpha, \ \ \phi(z) \le \phi(x) \ \ \text{and} \ \ \phi(\xi) + (\varepsilon/\alpha)d(\xi,z) \ge \phi(z) \ \forall \xi \in X.$$

$\diamond$

Proof. Put $h(z) = \inf_{\xi \in X}[\phi(\xi) + (\varepsilon/\alpha)d(\xi,z)]$. For arbitrary ξ, z and z' in X, we observe

$$\phi(\xi) + (\varepsilon/\alpha)d(\xi,z) \le \phi(\xi) + (\varepsilon/\alpha)[d(\xi,z') + d(z',z)].$$

Taking the infimum over all $\xi \in X$ on both sides, we obtain

$$h(z) \le h(z') + (\varepsilon/\alpha)d(z',z).$$

Therefore, h is a Lipschitz function; in particular, h is u.s.c. To construct a sequence z^k we set $z^0 = x$.

If $h(z^0) \ge \phi(z^0)$ then $z = z^0$ realizes all the assertions of the theorem. Thus, beginning with $k = 0$, assume that $h(z^k) < \phi(z^k)$. Then one finds some z^{k+1} such that

$$\phi(z^{k+1}) + (\varepsilon/\alpha)d(z^{k+1},z^k) < \phi(z^k) \tag{A.1}$$

and, in addition,

$$\phi(z^{k+1}) + (\varepsilon/\alpha)d(z^{k+1},z^k) < h(z^k) + 2^{-k}. \tag{A.2}$$

From (A.1) and $\phi(z^0) \le \inf_x \phi + \varepsilon$, we obtain for each k,

$$(\varepsilon/\alpha)\textstyle\sum_{s\le k} d(z^{s+1},z^s) < \sum_{s\le k}(\phi(z^s) - (\phi(z^{s+1})) = \phi(z^0) - \phi(z^{k+1}) \le \varepsilon.$$

This yields particularly

$$d(z^{k+1}, z^0) \le \sum_{s\le k} d(z^{s+1}, z^s) \le (\alpha/\varepsilon)[\phi(z^0) - \phi(z^{k+1}] \le \alpha. \tag{A.3}$$

By (A.1) and (A.3), $z = z^k$ is again the point in question whenever $\phi(z^k) \le h(z^k)$. Otherwise (A.3) shows that the Cauchy sequence z^k has a limit z^* in the complete space X. Since ϕ is l.s.c., we observe $\phi(z^*) \le \liminf \phi(z^k)$.

By (A.1), the sequence of $\phi(z^k)$ is decreasing, hence $\liminf \phi(z^k) \le \phi(z^0)$. Moreover, using (A.3) we even obtain

$$d(z^*, z^0) \le (\alpha/\varepsilon)[\phi(z^0) - \phi(z^*)].$$

Finally, recalling that h is u.s.c., we infer due to (A.2), the key relation

$$\phi(z^*) \le \liminf[\phi(z^{k+1}) + (\varepsilon/\alpha)d(z^{k+1}, z^k)] \le \limsup[h(z^k) + 2^{-k}] \le h(z^*).$$

The latter proves the theorem. □

Approximation by Directional Derivatives

The following lemma can be found in Shapiro's paper [Sha90] where a survey of concepts of directional differentiability and their interrelations is presented, see also [BS00].

Lemma A.2 (approximation by directional derivatives 1). *Let $f : \mathbb{R}^n \to Y$ be locally Lipschitz (Y normed), and let directional derivatives $f'(x^0; u)$ exist for all $u \in \mathbb{R}^n$. Then*

$$f(x^0 + u) - f(x^0) \in f'(x^0; u) + o(u)B.$$

◇

Proof. Otherwise one finds converging $u \to 0$ (a sequence) and some $c > 0$ such that

$$[f(x^0 + u) - f(x^0)]/\|u\| - f'(x^0; u/\|u\|) = y(u), \text{ where } \|y(u)\| \ge c.$$

The directions $v = u/\|u\| \in \mathbb{R}^n$ have some cluster point; so they converge for some subsequence, $v \to v^0$. Setting $t = \|u\|$ we obtain

$$[f(x^0 + tv) - f(x^0)]/t - f'(x^0; v) = y(u). \tag{A.4}$$

Since f is locally Lipschitz, say with rank L, we also observe (for all small t) that

$$\| [f(x^0 + tv^0) - f(x^0)]/t - [f(x^0 + tv) - f(x^0)]/t \| \le L\|v - v^0\|$$

and

$$\| f'(x^0; v) - f'(x^0; v^0) \| \le L\|v - v^0\|.$$

So, replacing v by v^0 in A.4, it holds for some subsequence,

$$\lim_{t\downarrow 0} \| [f(x^0 + tv^*) - f(x^0)]/t - f'(x^0; v^0)\| \geq c > 0,$$

in contradiction to the directional differentiability. □

Lemma A.3 (approximation by directional derivatives 2). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz. Then*

$$f(x^0 + u) - f(x^0) \in Cf(x^0)(u) + o(u)B.$$

◇

Proof. Otherwise one finds a sequence of converging directions $u \to 0$ and some $c > 0$ such that

$$\text{dist}\ ([f(x^0 + u) - f(x^0)]/\|u\|\ ,\ Cf(x^0)(u/\|u\|)) \geq c.$$

Setting $t = \|u\|$ and $v = u/t$, this is

$$\text{dist}\ ([f(x^0 + tv) - f(x^0)]/t,\ Cf(x^0)(v)) \geq c. \tag{A.5}$$

The directions $v \in \mathbb{R}^n$ have some cluster point. Thus, for certain $t \downarrow 0$, (belonging to some subsequence) the bounded quotients $[f(x^0 + tv) - f(x^0)]/t \in \mathbb{R}^m$ converge to an element $g \in Cf(x^0)(v^*)$. Since $f \in C^{0,1}$ the multifunction $Cf(x^0)(\cdot)$ is Lipschitz with respect to the Hausdorff-distance. In particular, it is lower semicontinuous, so dist $(g, Cf(x^0)(v))$ vanishes, in contradiction to A.5. □

Remark A.4 Analogously, one shows that if $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally upper Lipschitz at $(x^0, y^0) \in \text{gph}\, F$ and $CF(x^0, y^0)(\cdot)$ is lower semicontinuous, it holds

$$(F(x^0 + u) \cap \Omega) - y^0 \subset CF(x^0, y^0)(u) + o(u)B.$$

for some neighborhood Ω of y^0. ◇

The l.s.c. assumption is essential even for pointwise Lipschitz functions f: Put, for $x \in \mathbb{R}^2$ and $t \in \mathbb{R}$,

$$f(x) = t \text{ if } x = (t, t^2) \text{ and } f(x) = 0 \text{ otherwise.}$$

Lemma A.5 (descent directions). *Let $f : X \to \mathbb{R}$ be locally Lipschitz (X normed), and let directional derivatives $f'(x^0; u)$ exist for each u. Further, let $t \downarrow 0$ be some sequence such that related elements $u(t)$ fulfill $f(x^0 + tu(t)) - f(x^0) \leq ct$ with some fixed $c \in \mathbb{R}$. Then, it holds $f'(x^0; u^0) \leq c$ for each cluster point u^0 of the sequence $u(t)$.* ◇

Proof. If, for certain $s \downarrow 0$ we have $u(s) \to u^0$ then using some Lipschitz rank L for f near x^0, it follows

$$\begin{aligned} cs &\geq f(x^0 + su(s)) - f(x^0) \\ &= [f(x^0 + su^0) - f(x^0)] + [f(x^0 + su(s)) - f(x^0 + su^0)] \\ &\geq [f(x^0 + su^0) - f(x^0)] - Ls\|u(s) - u^0\|. \end{aligned}$$

Thus,

$$c \geq \limsup s^{-1}[f(x^0 + su^0) - f(x^0)]$$

holds for the particular sequence $s \downarrow 0$. Since f is directionally differentiable, this yields

$$c \geq \lim t^{-1}[f(x^0 + tu^0) - f(x^0)] = f'(x^0; u^0)$$

for every sequence $t \downarrow 0$. □

Proof of TF = T(NM) = N TM + TN M

The main point in the proof of Theorem 7.6 was the product rule $TF = T(NM) = NTM + TNM$. Since the way via the more general Theorem 6.8 is quite long we add a direct proof which is valid for the actual product rule only.

Lemma A.6 (direct proof of the product rule).

$$TF = T(NM) = N\,TM + TN\,M. \tag{A.6}$$

◇

Proof. To begin with we set

$$\delta F = F(x + u', y + v', z + w') - F(x, y, z)$$

(similarly δN and δM are defined), and observe that

$$\begin{aligned} \delta F &= N(y,z)(M(x+u') - M(x)) + (N(y+v', z+w') - N(y,z))M(x+u') \\ &= N(y,z)\delta M + \delta N M(x) + \delta N \delta M. \end{aligned}$$

Now put $(u', v', w') = t(u, v, w) = t\sigma$ for any given sequence $t \downarrow 0$.

Then $\delta F, \delta M$ and δN depend on t and

$$\delta F(t)/t = N(y,z)(\delta M(t)/t) + (\delta N(t)/t)M(x) + \delta N(t)\delta M(t)/t.$$

If, moreover, $(x, y, z) \to s^0 = (x^0, y^0, z^0)$ then - since M and N are locally Lipschitz - the bounded sequences $\delta M(t)/t$ and $\delta N(t)/t$ have accumulation points $M_0 \in TM(x^0)(u)$ and $N_0 \in TN(y^0, z^0)(v, w)$, respectively. The third term $\delta N(t)\delta M(t)/t$ is vanishing. So we obtain, for all converging subsequences $\delta F(t)/t$, that the limit can be written as

$$\lim \delta F(t)/t = N(y^0, z^0)M_0 + N_0 M(x^0).$$

This tells us $TF(s^0)(\sigma) \subset N(y^0,z^0)[TM(x^0)(u)] + [TN(y^0,z^0)(v,w)]M(x^0)$.

For showing the reverse inclusion, the special structure of N comes into the play. Let

$$M_0 \in TM(x^0)(u) \text{ and } N_0 \in TN(y^0,z^0)(v,w)$$

be arbitrarily given, and let $x = x(t) \to x^0$ and $t \downarrow 0$ be appropriate sequences such that $M_0 = \lim t^{-1}(M(x+tu) - M(x))$. The existence of such sequences is ensured by the definition of TM. To show that $N(y^0,z^0)M_0 + N_0M(x^0) \in TF(s^0)(\sigma)$, we have to find elements $(y,z) = (y(t),z(t)) \to (y^0,z^0)$ in such a way that N_0 can be written as

$$N_0 = \lim t^{-1}(N(y+tv,z+tw) - N(y,z)). \tag{A.7}$$

with the already given sequence of t (or with some infinite subsequence). If this is possible then, considering $\delta F(t)$ for $(x,y,z) + t(u,v,w)$ as above, we obtain

$$N(y^0,z^0)M_0 + N_0M(x^0) = \lim \delta F(t)/t \in TF(s^0)(\sigma),$$

which proves the lemma. We are now going to construct $(y(t),z(t))$ for given $t \downarrow 0$. By definition of $N = (1,y^+,y^-,z)$, the first and the last components of any element $N_0 \in TN(y^0,z^0)(v,w)$ belong to the map $z \to (1,z)$ and are obviously 0 and w, respectively. The remaining components of N_0 are formed by the T-derivative of the function $y \to c(y)$ at y^0 in direction v, which has been already studied in Lemma 7.4. Accordingly, we find $y = y(t)$ such that $t^{-1}[c(y+tv)-c(y)] = (\alpha, v-\alpha)$ for small $t > 0$, and (A.7) holds even as identity:

$$N_0 \equiv t^{-1}[N(y+tv,z+tw) - N(y,z)] \text{ for small } t > 0.$$

Hence the lemma is true, indeed. □

Constraint Qualifications

The following lemma compiles some basic facts on crucial constraint qualifications (CQ) for a nonlinear program

$$\text{(P)} \qquad \min\{f(x) | g(x) \le 0, h(x) = 0\},$$

with f, g, and h being C^1 functions defined from $\mathbb{R}^n$ to $\mathbb{R}$, $\mathbb{R}^m$ and $\mathbb{R}^\kappa$, respectively.

Recall that MFCQ (*Mangasarian-Fromovitz* CQ) is said to hold at some feasible point $\bar{x}$ if both $Dh(\bar{x})$ has full row rank and

$$\{u \,|\, Dg_i(\bar{x})u < 0 \ (i \in I(\bar{x})),\ Dh(\bar{x})u = 0\} \ne \emptyset,$$

while LICQ (*Linear Independence* CQ) is said to hold at $\bar{x}$ if

$$\{Dg_i(\bar{x}) \ (i \in I(\bar{x})),\ Dh_j(\bar{x}) \ (j = 1,\dots,\kappa)\}$$

is linearly independent, where $I(\bar{x}) = \{i | g(\bar{x}) = 0\}$.

Given a stationary solution $\bar{x}$ of (P), **SMFCQ** (strict **MFCQ**) is defined to hold at $\bar{x}$ if the set of Lagrange multipliers

$$\Lambda(\bar{x}) = \{(\bar{y}, \bar{z}) | (\bar{x}, \bar{y}, \bar{z}) \text{ is a KKT point of (P)}\}$$

is a singleton. To have a unified algebraic description of the above CQs, let us introduce, for any feasible point $\bar{x}$ of (P) and any index set $J \subset I(\bar{x})$, the polyhedral cone

$$A^J(\bar{x}) := \{(\alpha, w) | Dg(\bar{x})^\mathsf{T}\alpha + Dh(\bar{x})^\mathsf{T} w = 0,\ \alpha_i(x) \geq 0,\ i \in J,\ \alpha_j = 0,\ j \notin I(\bar{x})\}.$$

Let us also recall *Gordon's theorem of the alternative* [Man81b, Man94] which says that for matrices Q and R of suitable dimensions,

$$\begin{array}{l} \{u \mid Qu < 0,\ Ru = 0\} \neq \emptyset, \text{ and } R \text{ has full row rank} \\ \quad \Leftrightarrow \quad \{(v, w) \mid Q^\mathsf{T} v + R^\mathsf{T} w = 0,\ v \geq 0\} \ = \ \{0\}, \end{array} \tag{A.8}$$

and, by a standard argument from convex analysis (see, e.g., [Man81a, Man94, Roc70, SW70]), this equivalently means that for any right-hand side q, the linear system $\{Q^\mathsf{T} y + R^\mathsf{T} z = q,\ y \geq 0\}$ has a bounded (possibly empty) solution set.

Hence, it follows immediately that a feasible point $\bar{x}$ of (P) satisfies

$$\text{LICQ} \quad \text{if and only if} \quad A^J(\bar{x}) = \{0\} \text{ for } J = \emptyset, \tag{A.9}$$

$$\text{MFCQ} \quad \text{if and only if} \quad A^J(\bar{x}) = \{0\} \text{ for } J = I(\bar{x}). \tag{A.10}$$

Moreover, given a stationary solution $\bar{x}$ and $(y, z) \in \Lambda(\bar{x})$, the following lemma states that $\bar{x}$ satisfies

$$\text{SMFCQ} \quad \text{if and only if} \quad A^J(\bar{x}) = \{0\} \text{ for } J = I^0(y), \tag{A.11}$$

where again $I^0(y) := \{i | y_i = 0\} \subset I(\bar{x})$. From (A.9)–(A.11), the known implications

$$\text{LICQ} \Rightarrow \text{SMFCQ} \Rightarrow \text{MFCQ}$$

are obvious.

Lemma A.7 (Gauvin's theorem [Gau77] and Kyparisis' [Kyp85] theorem). *Let $\bar{x}$ be a stationary solution of (P). Then*

(i) $\Lambda(\bar{x})$ *is bounded if and only if MFCQ is satisfied at* $\bar{x}$.
(ii) $\Lambda(\bar{x})$ *is a singleton (i.e., SMFCQ is satisfied at* $\bar{x}$*) if and only if for some* $(y, z) \in \Lambda(\bar{x})$, *one has* $A^{I^0(y)}(\bar{x}) = \{0\}$. ◇

Proof. (i) follows from (A.8) and (A.10) according to the discussion above. To show the "only if"-direction of (ii), let $(y, z) \in \Lambda(\bar{x})$ and $(\alpha, w) \in A^{I^0(y)}(\bar{x})$ with $(\alpha, w) \neq 0$. Then, for all $\varepsilon > 0$ sufficiently small, one has

$$\begin{array}{l} Df(\bar{x}) + \sum_{i=1}^m (y_i + \varepsilon\alpha_i) Dg_i(\bar{x}) + \sum_{k=1}^\kappa (z_k + \varepsilon w_k) Dh_k(\bar{x}) = 0, \\ y_i + \varepsilon\alpha_i \geq 0,\ i \in I(\bar{x}) = I^0(y) \cup I^+(y), \\ y_j + \varepsilon\alpha_j = 0,\ j \notin I(\bar{x}), \end{array}$$

where $I^+(y) = \{i | y_i > 0\}$. Thus, $\Lambda(\bar{x})$ is not a singleton.

To show the "if"-direction of (ii), assume that $\Lambda(\bar{x})$ is not a singleton. Let $(\bar{y}, \bar{z})$ be any element of $\Lambda(\bar{x})$. Then there is a second element $(\tilde{y}, \tilde{z}) \in \Lambda(\bar{x})$ such that both $(\bar{y}, \bar{z})$ and $(\tilde{y}, \tilde{z})$ satisfy

$$Df(\bar{x}) + Dg(\bar{x})^\top y + Dh(\bar{x})^\top z = 0,$$
$$y_i \geq 0, i \in I(\bar{x}), \quad y_j = 0, j \notin I(\bar{x}).$$

Hence, $Dg(\bar{x})^\top(\tilde{y} - \bar{y}) + Dh(\bar{x})^\top(\tilde{z} - \bar{z}) = 0$ and

$$(\tilde{y} - \bar{y})_i \geq 0, i \in I^0(\bar{y}) \subset I(\bar{x}), \quad (\tilde{y} - \bar{y})_j = 0, j \notin I(\bar{x}),$$

i.e., $0 \neq (\tilde{y} - \bar{y}, \tilde{z} - \bar{z}) \in A^{I^0(\bar{y})}(\bar{x})$, which completes the proof. □

This Page Intentionally Left Blank

Bibliography

[AC95] D. Azé and C.C. Chou. On a Newton type iterative method for solving inclusions. *Mathematics of Operations Research,* 20:790–800, 1995.

[AE84] J.-P. Aubin and I. Ekeland. *Applied Nonlinear Analysis.* Wiley, New York, 1984.

[AF90] J.-P. Aubin and H. Frankowska. *Set–Valued Analysis.* Birkhäuser, Boston, 1990.

[AH35] P.S. Alexandroff and H. Hopf. *Topologie.* Springer, Berlin, 1935.

[Alt83] W. Alt. Lipschitzian perturbations of infinite optimization problems. In A.V. Fiacco, editor, *Mathematical Programming with Data Perturbations,* pages 7–21. M. Dekker, New York, 1983.

[Alt90] W. Alt. *Stability of Solutions and the Lagrange–Newton Method for Nonlinear Optimization and Optimal Control Problems.* Universität Bayreuth, Bayreuth, 1990. Habilitationsschrift.

[Asp68] E. Asplund. Fréchet differentiability of convex functions. *Acta Math.,* 121:31–47, 1968.

[Att84] H. Attouch. *Variational Convergence for Functions and Operators.* Applicable Mathematics Series. Pitman, London, 1984.

[Aub84] J.P. Aubin. Lipschitz behaviour of solutions to convex minimization problems. *Mathematics of Operations Research,* 9:87–111, 1984.

[Aus84] A. Auslender. Stability in mathematical programming with nondifferentiable data. *SIAM Journal on Control and Optimization,* 22:29–41, 1984.

[BA93] E.G. Belousov and V.G. Andronov. *Solvability and Stability for Problems of Polynomial Programming.* Moscow University Publishers, Moscow, 1993. in Russian.

[BBZ81] A. Ben-Israel, A. Ben-Tal, and S. Zlobec. *Optimality in Nonlinear Programming: A Feasible Direction Approach.* Wiley, New York, 1981.

[Ben80] A. Ben-Tal. Second–order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications,* 31:143–165, 1980.

[Ber63] C. Berge. *Topological Spaces.* Macmillan, New York, 1963.

[BF97] S.C. Billups and M.C. Ferris. QPCOMP: A quadratic programming based solver for mixed complementarity problems. *Mathematical Programming B,* 76:533–562, 1997.

[BGK+82] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

[BL00] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples.* CMS Books in Mathematics. Springer, New York, 2000.

[BM88] B. Bank and R. Mandel. *Parametric Integer Optimization.* Mathematical Research, Vol. 39. Akademie-Verlag, Berlin, 1988.

[BMX94] J.M. Borwein, W.B. Moors, and W. Xianfy. Lipschitz functions with prescribed derivatives and subderivatives. CECM Information Document 94-026, Simon Fraser Univ., Burnaby, 1994.

[Bou32] G. Bouligand. *Introduction à la Géométrie Infinitésimale Directe.* Gauthier-Villars, Paris, 1932.

[BS98] J.F. Bonnans and A. Shapiro. Optimization problems with perturbations: A guided tour. *SIAM Review,* 40:228–264, 1998.

[BS00] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer, New York, 2000.

[BT96] J.V. Burke and P. Tseng. A unified analysis of Hoffman's bound via Fenchel duality. *SIAM Journal on Optimization,* 6:265–282, 1996.

[BZ82] A. Ben-Tal and J. Zowe. A unified theory of first and second order conditions for extremum problems in topological vector spaces. *Mathematical Programming Study,* 19:39–76, 1982.

[BZ88] J.M. Borwein and D.M. Zhuang. Verifiable necessary and sufficient conditions for regularity of set-valued and single-valued maps. Journal *of Mathematical Analysis and Applications,* 134:441–459, 1988.

[Cha89] R. W. Chaney. Optimality conditions for piecewise C^2 nonlinear programming. *Journal of Optimization Theory and Applications,* 61:179–202, 1989.

[Cla76] F.H. Clarke. On the inverse function theorem. *Pacific Journal of Mathematics,* 64:97–102, 1976.

[Cla83] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983.

[Com90] R. Cominetti. Metric regularity, tangent sets and second-order optimality conditions. *Applied Mathematics and Optimization,* 21:265–287, 1990.

[DFS67] G.B. Dantzig, J. Folkman, and N. Shapiro. On the continuity of the minimum set of a continuous function. *Journal of Mathematical Analysis and Applications,* 17:519–548, 1967.

[DM74] V.F. Demyanov and V.N. Malozemov. *Introduction to Minimax.* Wiley, New York, 1974.

[DMO80] A.V. Dmitruk, A.A. Milyutin, and N.P. Osmolovski. Lyusternik's theorem and the theory of the extremum. *Uspekhy Mat. Nauk,* 35:11–46, 1980. in Russian.

[Don83] A. Dontchev. *Perturbations, Approximations and Sensitivity Analysis of Optimal Control Systems.* Lecture Notes in Control and Information Sciences 52. Springer, Berlin, 1983.

[Don95] A. Dontchev. Characterizations of Lipschitz stability in optimization. In R. Lucchetti and J. Revalski, editors, *Recent Developments in Well–Posed Variational Problems,* pages 95–116. Kluwer, 1995.

[Don96] A. Dontchev. Local convergence of the Newton method for generalized equations. *Comptes Rendus de l'Académie des Sciences de Paris,* 332, Ser. I:327–331, 1996.

[Don98] A. Dontchev. A proof of the necessity of linear independence constraint qualification and strong second–order sufficient optimality condition for Lipschitzian stability in nonlinear programming. *Journal of Optimization Theory and Applications,* 98:467–473, 1998.

[DR96] 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.

[DR98] 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, New York, 1998.

[DZ93] A. Dontchev and T. Zolezzi. *Well–Posed Optimization Problems,* Lecture Notes in Mathematics 1543. Springer, Berlin, 1993.

[Eke74] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications,* 47:324–353, 1974.

[Fab86] M. Fabian. Subdifferentials, local ε-supports and Asplund spaces. *Journal London Math. Soc.,* 34:568–576, 1986.

[Fab89] M. Fabian. Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss. *Acta Univ. Carolinae,* 30:51–56, 1989. 17th Winter School on Abstract Analysis, Srni 89.

[Fed69] H. Federer. *Geometric Measure Theory.* Springer, New York, Heidelberg, 1969.

[Fia74] A.V. Fiacco. Convergence properties of local solutions of sequences of mathematical programming problems in general spaces. *Journal of Optimization Theory and Applications,* 13:1–12, 1974.

[Fia76] A.V. Fiacco. Sensitivity analysis for nonlinear programming using penalty functions. *Mathematical Programming,* 10:287–311, 1976.

[Fia83] A.V. Fiacco. *Introduction to Sensitivity and Stability Analysis.* Academic Press, New York, 1983.

[Fis97] A. Fischer. Solutions of monotone complementarity problems with locally Lipschitzian functions. *Mathematical Programming, Series B,* 76:513–532, 1997.

[FM68] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New York, 1968.

[Fus94] P. Fusek. *Über Kettenregeln in Gleichungsform für Ableitungen nichtglatter Funktionen.* Diplomarbeit, Fachbereich Mathematik. Humboldt–Universitat zu Berlin, Berlin, 1994.

[Fus99] P. Fusek. *Eigenschaften pseudo-regulärer Funktionen und einige Anwendungen auf Optimierungsaufgaben.* Dissertation, Fachbereich Mathematik. Humboldt–Universität zu Berlin, Berlin, Februar 1999.

[Fus01] P. Fusek. Isolated zeros of Lipschitzian metrically regular $\mathbf{R}^n$ functions. *Optimization,* 49:425–446, 2001.

(Gau77] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming,* 12:136–138, 1977.

[Gau94] J. Gauvin. *Theory of Nonconvex Programming.* Les Publications CRM, Montreal, 1994.

[GD82] J. Gauvin and F. Dubeau. Differential properties of the marginal function in mathematical programming. *Mathematical Programming Study,* 19:101–119, 1982.

[Gfr87] H. Gfrerer. Holder continuity of solutions of perturbed optimization problems under Mangasarian-Fromovitz Constraint Qualification. In J. Guddat, H.Th. Jongen, B. Kummer, and F. **Nožička**, editors, Parametric *Optimization and Related Topics,* pages 113–124, Akademie-Verlag, Berlin, 1987.

[Gfr98] H. Gfrerer. Personal communication, 1998.

[Gfr00] H. Gfrerer. Personal communication, 2000.

[GGJ90] J. Guddat, F. Guerra, and H.Th. Jongen. *Parametric Optimization: Singularities, Pathfollowing and Jumps.* Wiley, Chichester, 1990,

[GJ88] J. Gauvin and R. Janin. Directional behaviour of optimal solutions in nonlinear mathematical programming. *Mathematics of Operations Research,* 13:629–649, 1988.

[Gol72] E.G. Golstein. *Theory of Convex Programming.* Transactions of Mathematical Monographs 36. American Mathematical Society, Providence, RI, 1972.

[Gra50] L.M. Graves. Some mapping theorems. *Duke Mathematical Journal,* 17:111–114, 1950.

[GT77] J. Gauvin and J.W. Tolle. Differential stability in nonlinear programming. *SIAM Journal on Control and Optimization,* 15:294–311, 1977.

[Hag79] W.W. Hager. Lipschitz continuity for constrained processes. *SIAM Journal on Control and Optimization,* 17:321–338, 1979.

[Har77] A. Haraux. How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities. *Journal of the Mathematical Society of Japan,* 29:615–631, 1977.

[Har79] R. Hardt. *An Introduction to Geometric Measure Theory,* Lecture Notes, Melbourne University, 1979.

[HG99] W.W. Hager and M.S. Gowda. Stability in the presence of degeneracy and error estimation. *Mathematical Programming,* 85:181–192, 1999.

(HK94) R. Henrion and D. Klatte. Metric regularity of the feasible set mapping in semi-infinite optimization. *Applied Mathematics and Optimization,* 30:103–109, 1994.

[HO01] R. Henrion and J. Outrata. A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications,* 258:110–130, 2001.

[Hof52] A.J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards,* 49:263–265, 1952.

[Hog73] W.W. Hogan. Point-to-set maps in mathematical programming. *SIAM Review,* 15:591–603, 1973.

[HP90] P.T. Harker and J.-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming,* 48:161–220, 1990.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I, II.* Springer, New York, 1993.

[HUSN84] 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.

[HZ82] R. Hettich and P. Zencke. *Numerische Methoden der Approximation und Semi–Infiniten Optimierung.* Teubner, Stuttgart, 1982.

[IK92] C.M. Ip and J. Kyparisis. Local convergence of quasi-Newton methods for B-differentiable equations. *Mathematical Programming,* 56:71–89, 1992.

[Iof79a] A.D. Ioffe. Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augmented duality. *SIAM Journal on Control and Optimization,* 17:266–288, 1979.

[Iof79b] A.D. Ioffe. Regular points of Lipschitz functions. *Transactions of the American Mathematical Society,* 251:61–69, 1979.

[Iof81] A.D. Ioffe. Nonsmooth analysis: differential calculus of nondifferentiable mappings. *Transactions of the American Mathematical Society,* 266:1–56, 1981.

[Iof94] A.D. Ioffe. On sensitivity analysis of nonlinear programs in Banach spaces: The approach via composite unconstrained optimization. *SIAM Journal on Optimization,* 4:1–43, 1994.

[Iof00] A.D. Ioffe. Codirectional compactness, metric regularity and subdifferential calculus. *American Math. Soc.*, Providence, RI:123–163, 2000. Conf. on Nonlin. Analysis, Limoges 1999.

[IT74] A.D. Ioffe and V.M. Tichomirov. *Theory of Extremal Problems.* Nauka, Moscow, 1974. in Russian.

[Jan84] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study,* 21:110–126, 1984.

[JJT83] H.Th. Jongen, P. Jonker, and F. Twilt. *Nonlinear Optimization in $\mathbb{R}^n$, I: Morse Theory, Chebychev Approximation.* Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork, 1983.

[JJT86] H.Th. Jongen, P. Jonker, and F. Twilt. *Nonlinear Optimization in $\mathbb{R}^n$, II: Transversality, Flows, Parametric Aspects.* Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork, 1986.

[JJT88] H.Th. Jongen, P. Jonker, and F. Twilt. The continuous, desingularized Newton method for meromorphic functions. *Acta Applicandae Mathematicae,* 13:81–121, 1988.

[JJT91] H.Th. Jongen, P. Jonker, and F. Twilt. On the classification of plane graphs representing structurally stable rational Newton flows. *Journal of Combinatorial Theory, Series B,* 51:256–270, 1991.

[JKT90] H.Th. Jongen, D. Klatte, and K. Tammer. Implicit functions and sensitivity of stationary points. *Mathematical Programming,* 49:123–138, 1990.

[JLS98] V. Jeyakumar, D.T. Luc, and S. Schaible. Characterization of generalized monotone nonsmooth continuous maps using approximate Jacobians. *Journal of Convex Analysis,* 5:119–132, 1998.

[JMRT87] H.Th. Jongen, T. Möbert, J. Rückmann, and K. Tammer. On inertia and Schur complement in optimization. *Linear Algebra and its Applications,* 95:97–109, 1987.

[JMT86] H.Th. Jongen, T. Möbert, and K. Tammer. On iterated minimization in nonconvex optimization. *Mathematics of Operations Research,* 11:679–691, 1986.

[JP88] H.Th. Jongen and D. Pallaschke. On linearization and continuous selections of functions. Optimization, 19:343–353, 1988.

[JRS98] H.Th. Jongen, J.-J. Rückmann, and O. Stein. Generalized semi–infinite optimization: A first order optimality condition and examples. *Mathematical Programming,* 83:145–158, 1998.

[JTW92] H.Th. Jongen, F. Twilt, and G.W. Weber. Semi-infinite optimization: structure and stability of the feasible set. *Journal of Optimization Theory and Applications,* 72:529–552, 1992.

[KA64] L.W. Kantorovich and G.P. Akilov. *Funktionalanalysis in normierten Räumen.* Akademie Verlag, Berlin, 1964.

[Kal86] P. Kall. Approximations to optimization problems: An elementary review. *Mathematics of Operations Research,* 11:9–18, 1986.

[Kap76] A. Kaplan. On the convergence of the penalty function method. *Soviet Math. Dokl.,* 17:1008–1012, 1976.

[KK85] D. Klatte and B. Kummer. Stability properties of infima and optimal solutions of parametric optimization problems. In V.F. Demyanov and D. Pallaschke, editors, *Nondifferentiable Optimization: Motivations and Applications,* pages 215–229. Springer, Berlin, 1985.

[KK99a] D. Klatte and B. Kummer. Generalized Kojima functions and Lipschitz stability of critical points. *Computational Optimization and Applications,* 13:61–85, 1999.

[KK99b] D. Klatte and B. Kummer. Strong stability in nonlinear programming revisited. *Journal of the Australian Mathematical Society, Series B,* 40:336–352, 1999.

[KK01] D. Klatte and B. Kummer. Contingent derivatives of implicit (multi–) functions and stationary points. *Annals of Operations Research,* 101:313–331, 2001.

[KL87] D. Kuhn and R. Löwen. Piecewise affine bijections of $\mathbb{R}^n$, and the equation $Sx^+ - Tx^- = y$. *Linear Algebra and its Applications,* 96:109–129, 1987.

[KL99] D. Klatte and W. Li. Asymptotic constraint qualifications and global error bounds for convex inequalities. *Mathematical Programming,* 84:137-160, 1999.

[Kla85] D. Klatte. On the stability of local and global optimal solutions in parametric problems of nonlinear programming. Part I: Basic results. Seminarbericht Nr. 75 der Sektion Mathematik, Humboldt-Universität Berlin, pages 1-21, 1985.

[Kla86] D. Klatte. On persistence and continuity of local minimizers of nonlinear optimization problems under perturbations. Seminarbericht Nr. 80 der Sektion Mathematik, Humboldt-Universität Berlin, pages 32-42, 1986.

[Kla87] D. Klatte. Lipschitz continuity of infima and optimal solutions in parametric optimization: The polyhedral case. In J. Guddat, H.Th. Jongen, B. Kummer, and F. **Nožička,** editors, *Parametric Optimization and Related Topics,* pages 229–248. Akademie–Verlag, Berlin, 1987.

[Kla91] D. Klatte. Strong stability of stationary solutions and iterated local minimization. In J. Guddat, H.Th. Jongen, B. Kummer, and F. **Nožička,** editors, *Parametric Optimization and Related Topics II,* pages 119–136. Akademie-Verlag, Berlin, 1991.

[Kla92] D. Klatte. Nonlinear optimization under data perturbations. In W. Krabs and J. Zowe, editors, *Modern Methods of Optimization,* pages 204–235. Springer, Berlin, 1992.

[Kla94a] D. Klatte. On quantitative stability for non-isolated minima. *Control and Cybernetics,* 23:183–200, 1994.

[Kla94b] D. Klatte. On regularity and stability in semi–infinite optimization. *Set-Valued Analysis,* 3:101-111, 1994.

[Kla97] D. Klatte. Lower semicontinuity of the minimum in parametric convex programs. *Journal of Optimization Theory and Applications,* 94:511–517, 1997.

[Kla98] D. Klatte. Hoffman's error bound for systems of convex inequalities. In A.V. Fiacco, editor, *Mathematical Programming with Data Perturbations,* pages 185–199. Marcel Dekker, New York, 1998.

[Kla00] D. Klatte. Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs. *Mathematical Programming,* 88:285–311, 2000.

[Klu79] R. Kluge. *Nichtlineare Variationsungleichungen und Extremalaufgaben.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1979.

[KM80] A.Y. Kruger and B.S. Mordukhovich. Extremal points and Euler equations in nonsmooth optimization. *Doklady Akad. Nauk BSSR,* 24:684–687, 1980. in Russian.

[Koj80] 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.

[KR92] A. King and R.T. Rockafellar. Sensitivity analysis for nonsmooth generalized equations. *Mathematical Programming,* 55:341–364, 1992.

[Kru85] A.Y. Kruger. Properties of generalized subdifferentials. *Sibirian Mathematical Journal,* 26:822–832, 1985.

[Kru96] A.Y. Kruger. On calculus of strict ε-**semidifferentials**. *Dokl. Akad. Nauk Belarus,* 40:4:34–39, 1996. in Russian.

[Kru97] A.Y. Kruger. Strict ε-**semidifferentials** and extremality conditions. *Dokl. Akad. Nauk Belarus,* 41:3:21–26, 1997. in Russian.

[Kru00] A.Y. Kruger. Strict (ε, δ)-**semidifferentials** and extremality of sets and functions. *Dokl. Nat. Akad. Nauk Belarus,* 44:4:21–24, 2000. in Russian.

[Kru01] A.Y. Kruger. Strict (ε, δ)-**semidifferentials** and extremality conditions. *Optimization,* 2001. to appear.

[KS87] M. Kojima and S. Shindoh. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of the Operational Research Society of Japan,* 29:352–372, 1987.

[KT88] D. Klatte and K. Tammer. On second–order sufficient optimality conditions for $C^{1,1}$–**optimization** problems. *Optimization,* 19:169–179, 1988.

[KT90] 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.

[KT96] D. Klatte and G. Thiere. A note on Lipschitz constants for solutions of linear inequalities and equations. *Linear Algebra and its Applications,* 244:365–374, 1996.

[Kum77] B. Kummer. Global stability of optimization problems. *Mathematische Operationsforschung und Statistik, Series Optimization,* 8:367–383, 1977.

[Kum81] B. Kummer. Stability and weak duality in convex programming without regularity. *Wissenschaftliche Zeitschrift der Humboldt–Universität zu Berlin, Mathematisch–Naturwissenschaftliche Reihe,* XXX:381–386, 1981.

[Kum84] B. Kummer. Generalized equations: Solvability and regularity. *Mathematical Programming Studies,* 21:199–212, 1984. as preprint Nr. 30, Sektion Mathematik Humboldt–Universität Berlin, 1982.

[Kum87] B. Kummer, Linearly and nonlinearly perturbed optimization problems in finite dimension. In J. Guddat et al., editor, *Parametric Optimization and Related Topics,* pages 249–267. Akademie Verlag, Berlin, 1987.

[Kum88a] B. Kummer. The inverse of a Lipschitz function in $\mathbb{R}^n$: Complete characterization by directional derivatives. Preprint no. 195, Humboldt-Universität Berlin, Sektion Mathematik, 1988.

[Kum88b] B. Kummer. Newton's method for non-differentiable functions. In J. Guddat et al., editor, *Advances in Math. Optimization,* pages 114–125. Akademie Verlag Berlin, (Ser. Math. Res. 45), Berlin, 1988.

[Kum91a] 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.

[Kum91b] B. Kummer. Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ **optimization**. *Journal of Optimization Theory and Applications,* 70:559–580, 1991.

[Kum92] 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.

[Kum95a] B. Kummer. Approximation of multifunctions and superlinear convergence. In R. Durier and C. Michelot, editors, *Recent Developments in Optimization,* volume 429 of Lecture *Notes in Economics and Mathematical Systems,* pages 243–251. Springer, Berlin, 1995.

[Kum95b] B. Kummer. On solvability and regularity of a parametrized version of optimality conditions. *ZOR Mathematical Methods of OR,* 41:215–230, 1995.

[Kum97] B. Kummer. Parametrizations of Kojima's system and relations to penalty and barrier functions. *Mathematical Programming, Series B,* 76:579–592, 1997.

[Kum98] 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.

[Kum99] B. Kummer. Metric regularity: Characterizations, nonsmooth variations and successive approximation. *Optimization,* 46:247–281, 1999.

[Kum00a] B. Kummer. Generalized Newton and NCP-methods: Convergence, regularity, actions. *Discussiones Mathematicae - Differential Inclusions,* 20:209–244, 2000.

[Kum00b] B. Kummer. Inverse functions of pseudo regular mappings and regularity conditions. *Mathematical Programming, Series B,* 88:313–339, 2000.

[KYF97] C. Kanzow, N. Yamashita, and M. Fukushima. New NCP-functions and their properties. *Journal of Optimization Theory and Applications,* 94:115–135, 1997.

[Kyp85] J. Kyparisis. On uniqueness of Kuhn–Tucker multipliers in nonlinear programming. *Mathematical Programming,* 32:242–246, 1985.

[Lau72] P.-J. Laurent. *Approximation et Optimisation.* Hermann, Paris, 1972.

[Lev94] E.G. Levitin. *Perturbation Theory in Mathematical Programming and its Applications.* Wiley, New York, 1994.

[Lev96] A.B. Levy. Implicit multifunction theorems for the sensitivity analysis of variational conditions. *Mathematical Programming,* 74:333–350, 1996.

[Li93] W. Li. The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program. *Linear Algebra and its Applications,* 187:15–40, 1993.

[LMO74] E.S. Levitin, A.A. Miljutin, and N.P. Osmolovski. On conditions for a local minimum in a problem with constraints. In B.S. Mitjagin, editor, *Mathematical Economics and Functional Analysis,* pages 139–202. Nauka, Moscow, 1974. in Russian.

[LP97] A.S. Lewis and J.-S. Pang. Error bounds for convex inequality systems. In J.P. Crouzeix, J.-E. Martinez-Legaz, and M. Volle, editors, *Generalized Convexity, Generalized Monotonicity: Recent Results,* pages 75–110. Kluwer Academic Publishers, Dordrecht, 1997.

[LPR96] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, Cambridge, 1996.

[LPR00] A.B. Levy, R.A. Poliquin, and R.T. Rockafellar. Stability of locally optimal solutions. *SIAM Journal on Optimization,* 10:580-604, 2000.

[LR94] A.B. Levy and R.T. Rockafellar. Sensitivity analysis for solutions to generalized equations. *Transactions of the American Mathematical Society,* 345:661-671, 1994.

[LR95] 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.

[LR96] A.B. Levy and R.T. Rockafellar. Variational conditions and the proto-differentiation of partial subgradient mappings. *Nonlinear Analysis: Theory, Methods & Applications,* 26:1951-1964, 1996.

[LS96] C. Lemaréchal and C. Sagastizabal. More than first-order developments of convex functions: Primal-dual relations. *Journal of Convex Analysis,* 3:1-14, 1996.

[LS97] C. Lemaréchal and C. Sagastizabal. Practical aspects of the Moreau-Yosida regularization: Theoretical premilinaries. *SIAM Journal on Optimization,* 7:367-385, 1997.

[LS98] W. Li and I. Singer. Global error bounds for convex multifunctions and applications. *Mathematics of Operations Research,* 23:443-462, 1998.

[Lyu34] L. Lyusternik. Conditional extrema of functions. *Math. Sbornik,* 41:390-401, 1934.

[Mal87] K. Malanowski. *Stability of Solutions to Convex Problems of Optimization.* Lecture Notes in Control and Information Sciences 92. Springer, Berlin, 1987.

[Man81a] O.L. Mangasarian. A condition number of linear inequalities and equalities. *Methods of Operations Research,* 43:3-15, 1981.

[Man81b] O.L. Mangasarian. A stable theorem of the alternative: An extension of the Gordan theorem. *Linear Algebra and its Applications,* 41:209-223, 1981.

[Man85] O.L. Mangasarian. A condition number for differentiable convex inequalities. *Mathematics of Operations Research,* 10:175-179, 1985.

[Man90] O.L. Mangasarian. Error bounds for nondegenerate monotone linear complementarity problems. *Mathematical Programming,* 48:437-446, 1990.

[Man94] O.L. Mangasarian. *Nonlinear Programming.* Classics in Applied Mathematics. SIAM, Philadelphia, 1994. republication of the work first published by McGraw-Hill Book Company, New York, 1969.

[MF67] 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.

[Mic56] E. Michael. Continuous selections I. *Annals of Mathematics,* 63:361-382, 1956.

[Mif77] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization,* 15:957–972, 1977.

[Mor88] B.S. Mordukhovich. *Approximation Methods in Problems of Optimization and Control.* Nauka, Moscow, 1988. in Russian.

[Mor93] B.S. Mordukhovich. Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions. *Transactions of the American Mathematical Society,* 340:1–35, 1993.

[Mor94] 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.

[MS97a] B.S. Mordukhovich and Y. Shao. Fuzzy calculus for coderivatives of multifunctions. *SIAM Journal on Control and Optimization,* 35:285–314, 1997.

[MS97b] B.S. Mordukhovich and Y. Shao. Stability of set-valued mappings in infinite dimensions: point criteria and applications. *SIAM Journal on Control and Optimization,* 35:285–314, 1997.

[MS98] B.S. Mordukhovich and Y. Shao. Mixed coderivatives of set-valued mappings in variational analysis. *Journal of Applied Analysis,* 4:269–294,1998.

[NGHB74] F. **Nožička,** J. Guddat, H. Hollatz, and B. Bank. *Theorie der linearen parametrischen Optimierung.* Akademie-Verlag, Berlin, 1974,

[NT01] H.V. Ngai and M. Théra. Metric inequality, subdifferential calculus and applications. *Set-Valued Analysis,* 9:187–216, 2001.

[OKZ98] J. Outrata, M. **Kočvara,** and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publ., Dordrecht-Boston-London, 1998.

[Out00] J. Outrata. A generalized mathematical program with equilibrium constraints. *SIAM Journal on Control and Optimization,* 38:1623–1638, 2000.

[Pan90] J.-S. Pang. Newton's method for B-differentiable equations. *Mathematics of Operations Research,* 15:311–341, 1990.

[Pan93] J.-S. Pang. A degree-theoretic approach to parametric nonsmooth equations with multivalued perturbed solution sets. *Mathematical Programming,* 62:359–383, 1993.

[Pan97] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming,* 79:299–332, 1997.

[Pen82] J.-P. Penot. On regularity conditions in mathematical programming. *Mathematical Programming Study,* 19:167–199, 1982.

[Pen89] J.-P. Penot. Metric regularity, openess and Lipschitz behavior of multifunctions. *Nonlinear Analysis: Theory, Methods & Applications,* 13:629–643, 1989.

[Pol90] R.A. Poliquin. Proto-differentiation of subgradient set-valued mappings. *Canadian Journal of Mathematics,* XLII, No.3:520–532, 1990.

[PQ93] J.-S. Pang and L. Qi. Nonsmooth equations: motivation and algorithms. *SIAM Journal on Optimization,* 3:443–465, 1993.

[PR94] R.A. Poliquin and R.T. Rockafellar. Proto-derivative formulas for basic subgradient mappings in mathematical programming. *Set–Valued Analysis,* 2:275–290, 1994.

[PR96] J.-S. Pang and D. Ralph, Piecewise smoothness, local invertibility, and parametric analysis of normal maps. *Mathematics of Operations Research,* 21:401–426, 1996.

[QS93] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming,* 58:353–367, 1993.

[RD95] D. Ralph and S. Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical Programming,* 70:159–172, 1995.

[Rob73] S.M. Robinson. Bounds for error in the solution set of a perturbed linear program. *Linear Algebra and its Applications,* 6:69–81, 1973.

[Rob75] S.M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal on Control and Optimization,* 13:271–273, 1975.

[Rob76a] S.M. Robinson. Regularity and stability for convex multivalued functions. *Mathematics of Operations Research,* 1:130–143, 1976.

[Rob76b] S.M. Robinson. Stability theorems for systems of inequalities. Part I: Linear systems. *SIAM Journal on Numerical Analysis,* 12:754–769, 1976.

[Rob76c] S.M. Robinson. Stability theorems for systems of inequalities. Part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis,* 13:497–513, 1976.

[Rob77] S.M. Robinson. A characterization of stability in linear programming. *Operations Research,* 25:435–447, 1977.

[Rob79] S.M. Robinson. Generalized equations and their solutions, Part I: Basic theory. *Mathematical Programming Study,* 10:128–141, 1979.

[Rob80] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research,* 5:43–62, 1980.

[Rob81] S.M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study,* 14:206–214, 1981.

[Rob82] S.M. Robinson. Generalized equations and their solutions. Part II: Applications to nonlinear programming. *Mathematical Programming Study,* 19:200–221, 1982.

[Rob87] S.M. Robinson. Local epi-continuity and local optimization. *Mathematical Programming,* 37:208–223, 1987.

[Rob91] S.M. Robinson. An implicit function theorem for a class of nonsmooth functions. *Mathematics of Operations Research,* 16:292–309, 1991.

[Rob94] S.M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis,* 2:291–305, 1994.

[Roc70] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, N.J., 1970.

[Roc71] R.T. Rockafellar. Ordinary convex programs without a duality gap. *Journal of Optimization Theory and Applications,* 7:143–148, 1971.

[Roc74] R.T. Rockafellar. *Conjugate Duality and Optimization.* Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1974.

[Roc81] R.T. Rockafellar. *The Theory of Subgradients and its Application to Problems of Optimization. Convex and Nonconvex Functions.* Heldermann, Berlin, 1981.

[Roc88] R.T. Rockafellar. First and second order epi-differentiability in nonlinear programming. *Transactions of the American Mathematical Society,* 207:75–108, 1988.

[RS97] D. Ralph and S. Scholtes. Sensivitity analysis of composite piecewise smooth equations. *Mathematical Programming, Series B,* 76:593–612, 1997.

[RW84] R.T. Rockafellar and R. J.-B. Wets. Variational systems, an introduction. In G. Salinetti, editor, *Multifunctions and Integrands: Stochastic Analysis, Approximation and Optimization,* Lecture Notes in Mathematics 1091, pages 1–54. Springer, Berlin, 1984.

[RW98] R.T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, Berlin, 1998.

[Sch94] S. Scholtes. *Introduction to Piecewise Differentiable Equations.* Preprint No. 53/1994. Institut für Statistik und Mathematische Wirtschaftstheorie, Universität Karlsruhe, 1994.

[Sha88a] A. Shapiro. Perturbation theory of nonlinear programs when the set of solutions is not a singleton. *Applied Mathematics and Optimization,* 18:215–229, 1988.

[Sha88b] A. Shapiro. Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. *SIAM Journal on Control and Optimization,* 26:628–645, 1988.

[Sha90] A. Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications,* 66:477–487, 1990.

[Sha94] A. Shapiro. On Lipschitzian stability of optimal solutions of parameterized semi–infinite programs. *Mathematics of Operations Research,* 19:743–752, 1994.

[Sha98] A. Shapiro. First and second order optimality conditions and perturbation analysis of semi–infinite programming problems. In R. Reemtsen and J.-J. Rückmann, editors, *Semi–Infinite Programming,* pages 103–133. Kluwer, Boston, Dordrecht, London, 1998.

[Sle96] B. Slezák. An inverse function theorem in topological groups. *Studia Scientiarum Mathematicarum Hungarica,* 31:415–421, 1996.

[SQ99] D. Sun and L. Qi. On NCP functions. *Computational Optimization and Applications,* 13:201–220, 1999.

[Stu86] M. Studniarski. Necessary and sufficient conditions for isolated local minima of nonsmooth functions. *SIAM Journal on Control and Optimization,* 24:1044–1049, 1986.

[SW70] J. Stoer and C. Witzgall. *Convexity and Optimization in Finite Dimensions I.* Springer, Berlin, 1970.

[SZ88] H. Schramm and J. Zowe. A combination of the bundle approach and the trust region concept. In J. Guddat and et al., editors, *Advances in Math. Optimization,* volume Ser. Mathem. Res., 45, pages 196–209. Akademie Verlag, Berlin, 1988.

[SZ92] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization,* 2:121–152, 1992.

[Thi80] L. Thibault. Subdifferentials of compactly Lipschitz vector-valued functions. *Annali di Matematica Pura ed Applicata,* 4:157–192, 1980.

[Thi82] L. Thibault. On generalized differentials and Subdifferentials of Lipschitz vector-valued functions. *Nonlinear Analysis: Theory, Methods & Applications,* 6:1037–1053, 1982.

[War75] J. Warga. Necessary conditions without differentiability assumptions in optimal control. *Journal of Differential Equations,* 15:13–46, 1975.

[War94] D.E. Ward. Characterizations of strict local minima and necessary conditions for weak sharp minima. *Journal of Optimization Theory and Applications,* 80:551–571, 1994.

[WW69] D.W. Walkup and R.J.-B, Wets. A Lipschitzian characterizations of convex polyhedra. *Proceedings of the American Mathematical Society,* 20:167–173, 1969.

[Zei76] E. Zeidler. *Vorlesungen über nichtlineare Funktionalanalysis I - Fixpunktsätze -*. Teubner Verlagsgesellschaft, Leipzig, 1976.

[Zie89] William P. Ziemer. *Weakly Differentiable Functions.* Springer, New York, 1989.

[ZK79] J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization,* 5:42–62, 1979.

Index

Nonconvex Optimization and Its Applications

1. D.-Z. Du and J. Sun (eds.): *Advances in Optimization and Approximation.* 1994 ISBN 0-7923-2785-3
2. R. Horst and P.M. Pardalos (eds.): *Handbook of Global Optimization.* 1995 ISBN 0-7923-3120-6
3. R. Horst, P.M. Pardalos and N.V. Thoai: *Introduction to Global Optimization* 1995 ISBN 0-7923-3556-2; Pb 0-7923-3557-0
4. D.-Z. Du and P.M. Pardalos (eds.): *Minimax and Applications.* 1995 ISBN 0-7923-3615-1
5. P.M. Pardalos, Y. Siskos and C. Zopounidis (eds.): *Advances in Multicriteria Analysis.* 1995 ISBN 0-7923-3671-2
6. J.D. Pintér: *Global Optimization in Action.* Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications. 1996 ISBN 0-7923-3757-3
7. C.A. Floudas and P.M. Pardalos (eds.): *State of the Art in Global Optimization.* Computational Methods and Applications. 1996 ISBN 0-7923-3838-3
8. J.L. Higle and S. Sen: *Stochastic Decomposition.* A Statistical Method for Large Scale Stochastic Linear Programming. 1996 ISBN 0-7923-3840-5
9. I.E. Grossmann (ed.): *Global Optimization in Engineering Design.* 1996 ISBN 0-7923-3881-2
10. V.F. Dem'yanov, G.E. Stavroulakis, L.N. Polyakova and P.D. Panagiotopoulos: *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics.* 1996 ISBN 0-7923-4093-0
11. B. Mirkin: *Mathematical Classification and Clustering.* 1996 ISBN 0-7923-4159-7
12. B. Roy: *Multicriteria Methodology for Decision Aiding.* 1996 ISBN 0-7923-4166-X
13. R.B. Kearfott: *Rigorous Global Search: Continuous Problems.* 1996 ISBN 0-7923-4238-0
14. P. Kouvelis and G. Yu: *Robust Discrete Optimization and Its Applications.* 1997 ISBN 0-7923-4291-7
15. H. Konno, P.T. Thach and H. Tuy: *Optimization on Low Rank Nonconvex Structures.* 1997 ISBN 0-7923-4308-5
16. M. Hajdu: *Network Scheduling Techniques for Construction Project Management.* 1997 ISBN 0-7923-4309-3
17. J. Mockus, W. Eddy, A. Mockus, L. Mockus and G. Reklaitis: *Bayesian Heuristic Approach to Discrete and Global Optimization.* Algorithms, Visualization, Software, and Applications. 1997 ISBN 0-7923-4327-1
18. I.M. Bomze, T. Csendes, R. Horst and P.M. Pardalos (eds.): *Developments in Global Optimization.* 1997 ISBN 0-7923-4351-4
19. T. Rapcsák: Smooth Nonlinear Optimization in $\boldsymbol{R}^n$. 1997 ISBN 0-7923-4680-7
20. A. Migdalas, P.M. Pardalos and P. Värbrand (eds.): *Multilevel Optimization: Algorithms and Applications.* 1998 ISBN 0-7923-4693-9
21. E.S. Mistakidis and G.E. Stavroulakis: *Nonconvex Optimization in Mechanics.* Algorithms, Heuristics and Engineering Applications by the F.E.M. 1998 ISBN 0-7923-4812-5

Nonconvex Optimization and Its Applications

22. H. Tuy: *Convex Analysis and Global Optimization.* 1998 ISBN 0-7923-4818-4
23. D. Cieslik: *Steiner Minimal Trees.* 1998 ISBN 0-7923-4983-0
24. N.Z. Shor: *Nondifferentiable Optimization and Polynomial Problems.* 1998 ISBN 0-7923-4997-0
25. R. Reemtsen and J.-J. Rückmann (eds.): *Semi-Infinite Programming.* 1998 ISBN 0-7923-5054-5
26. B. Ricceri and S. Simons (eds.): *Minimax Theory and Applications.* 1998 ISBN 0-7923-5064-2
27. J.-P. Crouzeix, J.-E. Martinez-Legaz and M. Volle (eds.): *Generalized Convexitiy, Generalized Monotonicity: Recent Results.* 1998 ISBN 0-7923-5088-X
28. J. Outrata, M. **Kočvara** and J. Zowe: *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* 1998 ISBN 0-7923-5170-3
29. D. Motreanu and P.D. Panagiotopoulos: *Minimax Theorems and Qualitative Properties of the Solutions of Hemivariational Inequalities.* 1999 ISBN 0-7923-5456-7
30. J.F. Bard: *Practical Bilevel Optimization.* Algorithms and Applications. 1999 ISBN 0-7923-5458-3
31. H.D. Sherali and W.P. Adams: *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems.* 1999 ISBN 0-7923-5487-7
32. F. Forgó, J. Szép and F. Szidarovszky: *Introduction to the Theory of Games.* Concepts, Methods, Applications. 1999 ISBN 0-7923-5775-2
33. C.A. Floudas and P.M. Pardalos (eds.): *Handbook of Test Problems in Local and Global Optimization.* 1999 ISBN 0-7923-5801-5
34. T. Stoilov and K. Stoilova: *Noniterative Coordination in Multilevel Systems.* 1999 ISBN 0-7923-5879-1
35. J. Haslinger, M. Miettinen and P.D. Panagiotopoulos: *Finite Element Method for Hemivariational Inequalities.* Theory, Methods and Applications. 1999 ISBN 0-7923-5951-8
36. V. Korotkich: *A Mathematical Structure of Emergent Computation.* 1999 ISBN 0-7923-6010-9
37. C.A. Floudas: *Deterministic Global Optimization: Theory, Methods and Applications.* 2000 ISBN 0-7923-6014-1
38. F. Giannessi (ed.): *Vector Variational Inequalities and Vector Equilibria.* Mathematical Theories. 1999 ISBN 0-7923-6026-5
39. D.Y. Gao: *Duality Principles in Nonconvex Systems.* Theory, Methods and Applications. 2000 ISBN 0-7923-6145-3
40. C.A. Floudas and P.M. Pardalos (eds.): *Optimization in Computational Chemistry and Molecular Biology.* Local and Global Approaches. 2000 ISBN 0-7923-6155-5
41. G. Isac: *Topological Methods in Complementarity Theory.* 2000 ISBN 0-7923-6274-8
42. P.M. Pardalos (ed.): *Approximation and Complexity in Numerical Optimization: Concrete and Discrete Problems.* 2000 ISBN 0-7923-6275-6
43. V. Demyanov and A. Rubinov (eds.): *Quasidifferentiability and Related Topics.* 2000 ISBN 0-7923-6284-5

Nonconvex Optimization and Its Applications

44. A. Rubinov: *Abstract Convexity and Global Optimization.* 2000 ISBN 0-7923-6323-X
45. R.G. Strongin and Y.D. Sergeyev: *Global Optimization with Non-Convex Constraints.* 2000 ISBN 0-7923-6490-2
46. X.-S. Zhang: *Neural Networks in Optimization.* 2000 ISBN 0-7923-6515-1
47. H. Jongen, P. Jonker and F. Twilt: *Nonlinear Optimization in Finite Dimensions.* Morse Theory, Chebyshev Approximation, Transversability, Flows, Parametric Aspects. 2000 ISBN 0-7923-6561-5
48. R. Horst, P.M. Pardalos and N.V. Thoai: *Introduction to Global Optimization.* 2nd Edition. 2000 ISBN 0-7923-6574-7
49. S.P. Uryasev (ed.): *Probabilistic Constrained Optimization.* Methodology and Applications. 2000 ISBN 0-7923-6644-1
50. D.Y. Gao, R.W. Ogden and G.E. Stavroulakis (eds.): *Nonsmooth/Nonconvex Mechanics.* Modeling, Analysis and Numerical Methods. 2001 ISBN 0-7923-6786-3
51. A. Atkinson, B. Bogacka and A. Zhigljavsky (eds.): *Optimum Design 2000.* 2001 ISBN 0-7923-6798-7
52. M. do Rosário Grossinho and S.A. Tersian: *An Introduction to Minimax Theorems and Their Applications to Differential Equations.* 2001 ISBN 0-7923-6832-0
53. A. Migdalas, P.M. Pardalos and P. Värbrand (eds.): *From Local to Global Optimization.* 2001 ISBN 0-7923-6883-5
54. N. Hadjisavvas and P.M. Pardalos (eds.): *Advances in Convex Analysis and Global Optimization.* Honoring the Memory of C. Caratheodory (1873-1950). 2001 ISBN 0-7923-6942-4
55. R.P. Gilbert, P.D. Panagiotopoulos† and P.M. Pardalos (eds.): *From Convexity to Nonconvexity.* 2001 ISBN 0-7923-7144-5
56. D.-Z. Du, P.M. Pardalos and W. Wu: *Mathematical Theory of Optimization.* 2001 ISBN 1-4020-0015-4
57. M.A. Goberna and M.A. López (eds.): *Semi-Infinite Programming. Recent Advances.* 2001 ISBN 1-4020-0032-4
58. F. Giannessi, A. Maugeri and P.M. Pardalos (eds.): *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models.* 2001 ISBN 1-4020-0161-4
59. G. Dzemyda, V. Šaltenis and A. Žilinskas (eds.): *Stochastic and Global Optimization.* 2002 ISBN 1-4020-0484-2
60. D. Klatte and B. Kummer: *Nonsmooth Equations in Optimization.* Regularity, Calculus, Methods and Applications. 2002 ISBN 1-4020-0550-4
61. S. Dempe: *Foundations of Bilevel Programming.* 2002 ISBN 1-4020-0631-4
62. P.M. Pardalos and H.E. Romeijn (eds.): *Handbook of Global Optimization, Volume 2.* 2002 ISBN 1-4020-0632-2

KLUWER ACADEMIC PUBLISHERS - DORDRECHT / BOSTON / LONDON