OPTIMIZATION and VARIATIONAL INEQUALITIES
Basic statements and constructions

Bernd Kummer; kummer@mathematik.hu-berlin.de ; May 2011

Abstract.
This paper summarizes basic facts in both finite and infinite dimensional optimization and for variational inequalities. In addition, partially new results (concerning methods and stability) are included. They were elaborated in joint work with D. Klatte, Univ. Zürich and J. Heerda, HU-Berlin.

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

Bemerkung.
Das Material wird ständig aktualisiert. Es soll Vorlesungen zur Optimierung und zu Variationsungleichungen (glatt, nichtglatt) unterstützen. Es entstand aus Scripten, die teils in englisch teils in deutsch aufgeschrieben wurden. Daher gibt es noch Passagen in beiden Sprachen sowie z.T. verschiedene Schreibweisen wie etwa $c^T x$ und $\langle c, x\rangle$ für das Skalarprodukt und die kanonische Bilinearform.

Hinweise auf Fehler sind willkommen.

Contents

Chapter 1

Introduction

1.1 The intentions of this script

This script presents an elementary introduction into the current theory of optimization and (formally more general) variational inequalities. We focus our attention to *basic ideas and statements* concerning *optimality conditions and solution methods* as well. In particular, we present classical proofs for first and second order optimality conditions and discuss basic methods like successive projection, proximal points, cutting planes, penalty- and barrier methods in the traditional framework.

On the other hand, we demonstrate how recent approaches (based on modified optimality conditions) are related to Newton-type methods in the smooth and non-smooth version as well as to stability theory of solutions.

Since optimality conditions and methods are closely related to the behavior of solutions and feasible points under small variations of the initial data [cf. stability, regularity, sensitivity or parametric optimization], we pay particular attention to this fact. We describe stability by known conditions in terms of generalized derivatives and by the help of Ekeland's variational principle. In addition, we show how "stability" can be (1-to-1) characterized by (linear) convergence of certain solution methods; a topic which is quite new and useful for several applications.

1.2 Notations

We write $A \subset B$ with the convention that $A = B$ is permitted, and $f \in C^k$ if f is k-times continuously differentiable. A function $f : X \to Y$ (metric spaces) is called locally Lipschitz ($f \in C^{0,1}$) if, for each $x \in X$, there is a neighborhood (nbhd) Ω of x and a constant L such that

$$d_Y(f(x'), f(x'')) \leq L\ d_X(x', x'') \quad \forall x', x'' \in \Omega.$$

To say that the $k-$th derivative of f exists and is locally Lipschitz, we also write $f \in C^{k,1}$.

Let $M \subset X,\ K \subset Y$. A mapping Γ which assigns, to each $x \in M$, any subset $\Gamma(x) \subset K$ (empty or not), is denoted by $\Gamma : M \rightrightarrows K$ and called *multifunction or (multivalued) mapping*. The set

$$\operatorname{gph}\Gamma := \{(x, y) \mid x \in M,\ y \in \Gamma(x)\}$$

is the *graph of* Γ, while $\operatorname{dom}\Gamma := \{x \in M \mid \Gamma(x) \neq \emptyset\}$ is its *domain*. One says that Γ is closed if $\operatorname{gph}\Gamma$ is a closed set in $X \times Y$.

All considered linear spaces are assumed to be linear spaces over $\mathbb{R}$. If A, B are sets in the same linear space, $A \pm B = \{a \pm b \mid a \in A, b \in B\}$ denotes the element-wise sum. Similarly, the notations $a \pm B = \{a \pm b \mid b \in B\}$ and $a + rB$ are used.

$\operatorname{dist}(x, M) = \inf_{y \in M} d(x, y)$ is the point-to-set distance; $\operatorname{dist}(x, \emptyset) = \infty$.

$B(x, r)$ denotes the closed ball around x with radius r, sometimes also $x + rB$ where B stands for the closed unit ball.

The topological dual space X^* of a linear normed space X consists of all additive, homogeneous and continuous functions $x^* : X \to \mathbb{R}$. We write $\langle x^*, x\rangle$ also for the image $x^*(x)$, and affine stands for affine-linear. In $X = \mathbb{R}^n$, we use the Euclidean norm and the canonical scaler product $\langle x, y\rangle = x^{\mathsf{T}} y = \sum x_i y_i$ if nothing else is said .

Some property holds *near* x if it holds for all x' in some nbhd of x. We say that x is a C^k point of a function f if f is C^k near x.

By $x_k \overset{M}{\to} x$ $(k = 1, 2...)$ we denote convergence $x_k \to x$ where $x_k \in M$. For $S_k \subset X$, the set $S = \text{Limsup}_{k\to\infty} S_k$ consists of all accumulation points x of sequences $x_k \in S_k$ (the so-called upper Hausdorff or Kuratovski limit).

Very often in this paper, assumptions like $f \in C^k$ can be replaced by *f is C^k near the point under consideration*. In addition, several statements, formulated for $f \in C^1$, need differentiability at the current point only. For sake of simple formulations we shall not explicitly pay attention to these (mostly obvious) facts.

Limiting negation, approach direction: Let us negate a statement of the type:
" All $\xi \neq x$ near $x \in \mathbb{R}^n$ satisfy condition (C) ".
Clearly, negation means $\exists \xi_k \to x$ $(k = 1, 2...)$ such that all ξ_k satisfy $\xi_k \neq x$ and (not C). Setting $t_k = \|\xi_k - x\|$, $u_k = (\xi_k - x)/t_k$ we obtain $\xi_k = x + t_k u_k$ where $\|u_k\| = 1$, $u_k \in \mathbb{R}^n$ and there is a converging (infinite) subsequence of certain $u_{k'}$, say $u_{k'} \to u$. Choosing the related subsequence of ξ_k, we thus obtain a useful form of the negation:
" $\exists u_k \to u, t_k \downarrow 0 :$ $\|u_k\| = 1$ and all $\xi_k = x + t_k u_k$ satisfy (not C)" .
Analogously, "All ξ in some sufficiently big ball $B(x, r)$ satisfy (C)" yields, in $\mathbb{R}^n$, the negation
" $\exists u_k \to u, t_k \to \infty :$ $\|u_k\| = 1$ and all $\xi_k = x + t_k u_k$ satisfy (not C)" .
Due to lack of a better phrase we call such statements *limiting negations*. The convergence $u_k \to u$ is often helpful. The direction u will be called *approach direction* (of all ξ_k).

In a separable, reflexive B-space, one can similarly use weak convergence of $u_k \rightharpoonup u$ (for elements or linear functionals), but possible conclusions of this fact are much weaker.

Lower level sets
For $g : X \to Y = \mathbb{R}^m$ with components g_i let $g_{\leq}(b) = \{x \in X \mid g_i(x) \leq b_i \forall i\}$ denote the lower level sets, and for any Y, $g_{=}(y) = g^{-1}(y) = \{x \mid g(x) = y\}$ be the set of pre-images.

1.3 Was ist (mathematische) Optimierung ?

Mathematische Optimierung beschäftigt sich mit Extremalwertaufgaben, d.h., eine (reellwertige) Funktion $f = f(x)$ soll auf einer Menge X minimal (maximal) gemacht werden.

1. Ist $X = \mathbb{R}$, sind solche Aufgaben aus der Schule wohlbekannt.

2. Ist X die Menge aller $x \in \mathbb{R}^n$, die einem Gleichungssystem $g(x) = 0$ genügen, kennt man aus der Analysisvorlesung die Lagrange Bedingung: Für extremales x gibt es ein λ mit $Df(x) = \sum_i \lambda_i \, Dg_i(x)$, sofern $Dg(x)$ maximalen Rang besitzt.

3. Ist $X \subset \mathbb{R}^n$ durch endlich viele lineare Ungleichungen definiert (also ein Polyeder) und f ebenfalls linear, hat man es mit einer linearen Optimierungaufgabe zu tun, die man mit Hilfe vieler fertiger Programme lösen kann (sagen die Vertreiber).

Optimierungsaufgaben sind vielfältig und treten überall in Theorie und Praxis auf:

A1. In Approximationsaufgaben: Sei ein reelles Polynom

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

vom Grade $p \leq n$ gesucht, das eine gegebene Funktion $q = q(t)$ über einem Intervall T "am besten" approximiert. Dann soll eine Norm $f(x) = \|p_x - q\|$ durch passende Wahl von $x = (x_0, ..., x_n) \in \mathbb{R}^{1+n}$ minimiert werden. Ist etwa

$$f(x) = \sup_{t \in T} |p_x(t) - q(t)|,$$

erfordert die Lösung schon einige Arbeit (Tschebyschev-Approximation). Weitere Bedingungen an die Koeffizienten x könnten hinzu kommen.

A2. Im klassischen Problem nichtlinearer Optimierung: Es soll f unter endlich vielen (Neben-) Bedingungen $g_i(x) \leq 0, \quad h_k(x) = 0,\ x \in \mathbb{R}^n$ minimiert werden; z.B. beschreibt f Kosten, die bei der Produktion eines *Warenbündels* x entstehen. Müssen gewisse Variable x_j zusätzlich ganzzahlig sein, erhält man eine *gemischt- ganzzahlige Aufgabe* (etwa *Optimierung eines Verkehrsflusses*, wenn man Einbahnstrassen Schilder aufstellen muss), die neben analytischen Mitteln auch solche der diskreten Mathematik erfordert.

A3. In Gleichgewichtsmodellen (sie spielen in ökonomischen Modellen, etwa Markt-Gleichgewichten unterschiedlicher Typen eine zentrale Rolle): Z.B. beschreibt x Strategien von n Spielern und

$$f_1(x),\ ...,\ f_n(x) \quad \text{mit} \quad x = (x_1, ..., x_n),\ \ x_i \in X_i$$

deren (aus irgendwelchen Spielregeln resultierende) Gewinne, wenn Spieler i jeweils x_i benutzt. Hat Spieler i keinen Einfluss auf die Strategiewahl der anderen Spieler, muss er zufrieden sein, wenn seine Wahl x_i die Gleichung

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

erfüllt. Die linke Seite ist nie kleiner als die rechte. Deshalb müssen *alle* Spieler zufrieden sein - bzw. es besteht ein gewisses *(Nash-) Gleichgewicht* - wenn x die Summe

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

minimiert und der Extremwert gerade Null ist. Wir sind einerseits in der *Spieltheorie* gelandet und haben eine i.a. nicht differenzierbare Funktion ϕ zu minimieren *(nichtglatte Optimierung, nonsmooth optimization)*. Oft ist es sinnvoll, Lösungen x als Fixpunkt zu interpretieren und Kakutani's Fixpunktsatz [58] zu benutzen, sect. 3.2, 3.3.

A4. In Extremalprinzipien in der Physik: Z.B. nimmt das Licht durch unterschiedliche Medien gerade den Weg, der es am schnellsten von A nach B kommen lässt. Man sucht

dann eine Kurve $x = x(t)$, die den Weg des Lichtes abhängig von der Zeit t beschreibt, und $f(x)$ ist ein Integral, in das die Ableitung von x eingeht,

$$f(x) = \int_0^T g(\ x(t),\ \dot{x}(t),\ t\ )\ dt.$$

Aufgaben dieses Typs werden im Rahmen der *Variationsrechnung* behandelt, ein klassischer Teil der Analysis, in dem die Wurzeln der heutigen *Theorie von Extremalaufgaben in allgemeinen Räumen* liegen.
A5. In Aufgaben der optimalen Steuerung: Z.B. soll eine Rakete weich auf dem Mond (kein Luftwiderstand) landen. Zur Zeit t habe sie die Höhe $h(t)$, die Geschwindigkeit $v(t)$ und erfahre die Beschleunigung $b(t) = \dot{v}(t)$ (alles "nach oben gemessen"). Dabei ist $b(t) = -g_{mond} + r(t)$ und g_{mond} die Gravitationskonstante des Mondes. Der Rest $r(t)$ sei die Beschleunigung in entgegengesetzte Richtung, die zur Vereinfachung proportional zur momentan (für das Bremsen) eingesetzten Treibstoffmenge $u(t)$ mit $0 \le u(t) \le c$ sei (tatsächlich spielt auch die Masse $M(t)$ der Rakete mit). Der Flug wird dann vereinfacht beschrieben durch die lineare DGL

$$\dot{h}(t) = v(t), \quad \dot{v}(t) = -g_{mond} + \rho\ u(t); \qquad h(0) = h_0,\ v(0) = v_0.$$

Gesucht wird die jeweils einzusetzende Treibstoffmenge $u(t)$, so dass

(a) die Funktion u noch hinreichend "vernünftig" ist (stückw. stetig mit endlich vielen Sprüngen im Interesse der Besatzung; wer sitzt gern drin bei $u \in L^p$?) und

(b) zu einer gewissen Zeit T die Bedingung der weichen Landung $h(T) = 0 = v(T)$ (erstmals !) erfüllt wird.

Zu minimieren ist etwa die Zeit T oder der verbrauchte Treibstoff $f(u, T) = \int_0^T u(t)dt$.

Allgemeine *Aufgaben der optimalen Steuerung* besitzen wie im Beispiel ein DGL-System mit Randbedingungen als Nebenbedingung

$$\dot{x}(t) = g(\ x(t),\ u(t),\ t), \quad x(0) = x_0,\ x(T) = x_1,$$

und eine Funktion u wird (in einer gewissen Menge) so gesucht, dass ein Integral $f(x, u, T) = \int_0^T h(x(t), u(t), t)\ dt$ oder eine andere Funktion $f(x, u, T)$ minimal wird.

Hängen die gesuchten Funktionen von mehreren Variablen ab, kommen partielle Diff.-Gleichungen ins Spiel. Gehen die Lösungen (oder Extrema) gewisser Aufgaben in andere ein, *Multi-level Probleme*, dann spielt (unter A3) die Abhängigkeit der Lösungen von den Parametern der Aufgabe eine zentrale Rolle (*parametrische Optimierung, Sensitivität*).

Für alle diese Aufgabentypen führen Fragen nach der Existenz von Lösungen und nach notwendigen und hinreichenden Optimalitäts-Bedingungen zu interessanten analytischen Problemen. Die Konzipierung und Umsetzung effektiver Lösungsmethoden auf der Grundlage *verwertbarer* Optimalitätsbedingungen ist deshalb eine permanente, zumeist nichttriviale Herausforderung und zugleich "konstruktive Mathematik".

1.4 Particular classes of problems

Given $M \subset X$ ($X = \mathbb{R}^n$ or X is a Banach space) and $f : M \to \mathbb{R}$ the main problem of optimization consists in finding the value

$$v = \inf\ \{\ f(x) \mid x \in M\} \tag{1.4.1}$$

and, if possible, some *minimizer* $\bar{x} \in M$. *Local solutions* $\bar{x}$ are solutions to (1.4.1) with new $M_\Omega = M \cap \Omega$ where Ω is a nbhd of $\bar{x}$. Solutions in the original sense are *global* solutions. The set M of all feasible points is often called *constraint set*. Particular problems:

P1. Linear programming

$$f(x) = \langle c, x\rangle; \quad M := \{x \in \mathbb{R}^n \mid Ax \leq b\}; \quad A = (m, n) \text{ matrix }, \ b \in \mathbb{R}^m. \tag{1.4.2}$$

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

P2. Mixed integer linear programming:
As above but A is a rational matrix and certain $x_1, ..., x_p$ are required to be integer.

P3. Mixed integer quadratic programming:
As above but $f(x) = \langle c, x\rangle \ + \ \langle x, Qx\rangle$; Q is a rational matrix.

Remark 1.4.1 In these cases, it holds the *existence theorem:* v finite $\Rightarrow$ some $\bar{x} \in M$ realizes the infimum. (quadratic without integer variables: Evans and Gould, quadratic with integer variables: Hansel). This statement still holds if f is an n-dimensional polynomial of degree 3 with rational coefficients [9]; it fails to hold for polynomials of degree 4 on convex polyhedrons M, cf. example 2.1.1. $\diamond$

In this script, integer variables will only play some role in sect. 6.8.

P4. Classical nonlinear problems in finite dimension

$$M := \{x \in X = \mathbb{R}^n \mid g_i(x) \leq 0 \ \forall i = 1, ..., m \text{ and } h_\nu(x) = 0 \ \forall \nu = 1..., m_h\}. \tag{1.4.3}$$

P5. Convex problems in finite dimension: As above with f, g_i convex and h_ν affine.

P6. Classical nonlinear problems in Banach spaces

$$\begin{gathered} X, Y, Z \text{ are B-spaces}, \ \ g : X \to Y, \ h : X \to Z \\ M := \{x \in X \mid g(x) \in K \text{ and } h(x) = 0\}; \ K \text{ is a closed, convex cone in } Y. \end{gathered} \tag{1.4.4}$$

X may describe the pairs (x, u) of functions in an optimal control problem, $h(x) = 0$ stands for a differential equation and $g(x) \in K$ for constraints like $x(t) \geq 0 \ \forall t$ or $u(t) \in [0, 1]$.

P7. Classical convex problems in Banach spaces
as above with f convex, h affine, $\operatorname{int} K \neq \emptyset$ and g convex w.r. to K; cf. sect. 5.4.1.
Here, K replaces the non-positive orthant, appearing under 4 and 5.

P8. Examples of nonsmooth (read: not enough differentiable) problems
8.1 Tschebyschev- approximation: see sect. 1.3: A1 and 3.6.
8.3 Multilevel problems: see sect. 7.1
8.2 Nash-equilibrium: see sect. 1.3: A3, sect. 3.3 and [109, 110]:
Nash-equilibrium is one of the basic solution concepts for non-cooperative games. A first existence theorem for such solution has been shown by J. v. Neumann; cf. [113].

1.5 Crucial questions

In order to find a solution for (1.4.1) or to check optimality of a given point, we need certain necessary and sufficient optimality conditions like "$Df(x) = 0$" and "$D^2 f(x)$

positive (semi-)definite" for free minima onto $X = \mathbb{R}^n$ and $f \in C^2$. In many papers, such conditions are written in the form

(i) $Df(\bar{x}) \in \mathcal{N}_M(\bar{x})$ where $\mathcal{N}_M(\bar{x})$ denotes some "normal cone" to M at $\bar{x}$.

(ii) $D^2f(\bar{x})$ positive (semi-) definite on some "tangent cone" $\mathcal{T}_M(\bar{x})$ to M at $\bar{x}$.

The simple task $\min \{x_1^2 + x_2 \mid x \in M\}$ with $M = \{x \in \mathbb{R}^2 \mid x_2 \geq 1 + x_1^2\}$ and $M = \{x \in \mathbb{R}^2 \mid x_2 \geq 1 + |x_1|\}$ gives a first impression on the kind of the needed cones $\mathcal{N}_M(\bar{x})$ and $\mathcal{T}_M(\bar{x})$.

Nobody can solve (1.4.1) without supposing some analytical description of the set M. Depending on this description, it may be more or less difficult to determine the "right" normal or tangent cones in (i) and (ii).

If f is not (once or twice) differentiable then something else has to replace the Fréchet derivatives $Df(x)$ and $D^2f(x)$. These *generalized derivatives*, cannot be applied in the same universal manner as Fréchet derivatives and do (usually) not represent linear functions and bilinear forms, respectively.

For these reasons, one finds various definitions of normals, tangents and generalized derivatives in the recent literature on optimization, and many crucial statements are formulated - unfortunately - in terms of such notions only. Here, we shall translate or check the abstract conditions in terms of the original data.
It turns out that, for many "non-optimization problems", conditions like (i), say

$$\gamma(x) \in \Gamma(x) \tag{1.5.1}$$

where γ is a function (not necessarily a derivative) and Γ a multifunction, play a role in order to describe the solution. Such systems are usually called
generalized equations and - if $\Gamma(x)$ is still some "normal cone" - *variational inequalities*. In both cases, $\Gamma(x)$ is a subset of the image space of γ.

Chapter 2

Lineare Optimierung

Lineare Optimierung ist grundlegend für nichtlineare Probleme. Daher erfolgt hier eine Zusammenfassung wichtiger Aussagen. Das grundlegende Problem ist

$$\min_{x \in M} c^T x \quad \text{mit } M = \{x \in \mathbb{R}^n \mid Ax \leq b\} \tag{2.0.1}$$

wobei $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ und eine (m, n) Matrix A gegeben sind. Die Menge M wird, den Zeilen A_i und Komponenten b_i entsprechend, durch m Ungleichungen $A_i x \leq b_i \ \forall i$ beschrieben. Sie ist also der Durchschnitt endlich vieler *affiner Halbräume*.

Mengen dieser Form (einschliesslich der ganze Raum) heissen auch (konvexe) Polyeder.

Oft wird unter einem Polyeder auch nur die konvexe Hülle $C := \operatorname{conv}\{x_1, .., x_N\} \subset \mathbb{R}^n$ endlich vieler Punkte verstanden. Hier also nicht. Es ist aber nicht-trivial (und eine Sternchen Uebungsaufgabe), dass C ein Polyeder in unserem Sinne darstellt.

Lemma 2.0.1 *Die konvexe Hülle $C := \operatorname{conv}\{x_1, .., x_N\}$ ist ein Polyeder.*

Proof. *Zu zeigen: Es gilt $C = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ mit geeigneter Matrix A und einem passenden b.*
Man kann voraussetzen, dass $\dim C = n$. Sonst betrachte man alles im kleinsten affinen Teilraum, der C enthält. Für affine Teilräume wissen wir schon, dass sie sich in Gleichungsform schreiben lassen.
Zu jedem $a \in \mathbb{R}^n,\ a \neq 0$ sei $L(a)$ die Lösungsmenge von

$$\max\{a^T x \mid x \in C\}$$

und $v(a)$ der Extremalwert. Dann ist

$$C \subset \{x \in \mathbb{R}^n \mid a^T x \leq v(a)\ \}\ \forall a. \tag{2.0.2}$$

Man sieht leicht mittels der (nicht notwendig eindeutigen) Darstellung $x = \sum_i \lambda_i x_i,\ \lambda_i \geq 0, \sum_i \lambda_i = 1$: Jedes $L(a)$ ist die konvexe Hülle endlich vieler x_i; nämlich derjeniger, die

$$a^T x_i = v(a)$$

erfüllen. Es gibt also nur endlich viele verschiedene Lösungsmengen. Die Durchschnitte $C \cap L(a)$ heissen *Seiten* von C.

Weiter liegt jeder Randpunkt x von C in einer Seite. Er lässt sich nämlich von $M = \operatorname{int} C$ trennen (Trennungssatz). Damit gibt es ein $a \neq 0$ mit $a^T x \geq a^T y\ \forall y \in C$. Mit $v(a) = \max\{a^T y \mid y \in C\}$, haben wir also eine solche Seite gefunden.

Hilfsaussage:
Wir zeigen nun als entscheidendes Hilfmittel, dass jeder Randpunkt x von C in (wenigstens) einer Seite der Dimension $n-1$ liegt.

Sei dazu $L(a_1)$ so fixiert, dass $x \in L(a_1)$ (Existenz s.o.). Wir können annehmen, dass

$$\dim L(a_1) = d_1 < n-1 \quad \text{und (OBdA)} \quad L(a_1) = \operatorname{conv}\{x_1, ..., x_q\}, \quad x_{q+1}, ..., x_N \notin L(a_1)$$

Der Fall $q = N$ scheidet aus wegen $a_1 \neq 0$ and $\dim C = n$. Man betrachte die Menge

$$Z(a_1) := \{a \mid L(a) = L(a_1)\}.$$

Sie hat die Form

$$Z(a_1) = \{\ a \mid \ a^T y = a^T x_1 \quad \forall y \in L(a_1) \text{ and } a^T y < a^T x_1 \ \forall y = x_{q+1}, ..., x_N\ \}.$$

Die Ungleichungen werden (Stetigkeit) von allen a nahe a_1 erfüllt. Damit folgt wegen $\dim L(a_1) = d_1$,

$$d_1' := \dim Z(a_1) = n - d_1 > 1.$$

Also gibt es ausser a_1 einen weiteren (lin. unabh.) Vektor $c \in Z(a_1)$. Mit jedem $t \in \mathbb{R}$ folgt deshalb

$$(\ a_1 + t(c - a_1)\ )^T y = v(a) \quad \forall\ y \in L(a_1). \tag{2.0.3}$$

Gilt auch noch

$$(\ a_1 + t(c - a_1)\ )^T y < v(a) \quad \text{für alle } y = x_{q+1}, ..., x_N \text{ und für alle } t,$$

so müssen alle Elemente aus C orthogonal zu $c - a_1 \neq 0$ sein. Das widerspricht $\dim C = n$. Also findet sich ein spezielles $t = t_1$ (mit kleinstem Betrag) so dass

$(a_1 + t_1(c - a_1))^T y \leq v(a)\ \forall y = x_{q+1}, ..., x_N$ und Gleichheit für ein $y = x_p$ mit $q < p \leq N$ gilt.

Mit $a_2 := a_1 + t_1(c - a_1) \neq 0$ folgt deshalb:
$L(a_2)$ enthält ein x_i mehr als $L(a_1)$, $L(a_1) \subset L(a_2)$ und $d_2 := \dim L(a_2) \geq d_1$. Wiederholt man diesen Schluss im Falle von $d_2 < n-1$, gelangt man so nach weniger als N Schritten zu einer Menge $L(a)$ mit

$$a \neq 0, \quad L(a_1) \subset L(a) \quad \text{und} \quad \dim L(a) = n - 1.$$

Das ist eine Seite grösster Dimension, die $L(a_1)$ enthält. Unsere Hilfsaussage ist bewiesen. □

Seien nun $a_1, ...a_m$ so gewählt, dass jede der endlich vielen Lösungsmengen $L(a)$ mit einem $L(a_k)$ zusammenfaellt, $1 \leq k \leq m$. Dann sind insbesondere alle $L(a)$ maximaler Dimension $n-1$ dabei (Tatsächlich brauchen wir nur diese). Die Menge

$$P = \{\ x \in \mathbb{R}^n \mid \ a_k^T\ x \leq v(a_k)\ \ \forall k = 1, ..., m\ \} \tag{2.0.4}$$

enthält C. Wir zeigen, dass sie gleich C ist. Andernfalls gibt es ein $\bar{x} \in P \setminus C$. Wir betrachten die Punkte

$$z(\lambda) = \lambda \bar{x} + (1 - \lambda) x^S$$

der Strecke zwischen $\bar{x} \notin C$ und einem festen Punkt $x^S \in \operatorname{int} C$. Ein $z(\lambda)$ muss im Rand von C liegen. Dabei ist $0 < \lambda < 1$. Als Randpunkt liegt $z(\lambda)$ auch in einer Seite der Dimension $n-1$. Mit dem entsprechenden a_k gilt dort $a_k^T z(\lambda) = v(a_k)$. Weil ausserdem $a_k^T x^S < v(a_k)$ (da $x^S \in \operatorname{int} C$) und $\lambda \in (0, 1)$ gelten, muss deshalb $a_k^T \bar{x} > v(a_k)$ sein. Dieser Widerspruch zu $\bar{x} \in P$ beweist das Lemma. □

2.1 Dualität und Existenzsatz für LOP

Die Aufgabe (2.0.1) lässt sich leicht in die folgende Gleichungsform überführen (mit neuem A und x). Mit ihr beweist man am einfachsten per vollständige Induktion

Theorem 2.1.1 *(Existenzsatz der lin. Optimierung). Die Aufgabe*

$$\max \{c^T x \mid x \in M\} \quad mit \quad M = \{x \in \mathbb{R}^n \mid \ Ax = b,\ x \geq 0\}$$

besitzt eine Lösung, wenn $M \neq \emptyset$ *and* $v := \sup_{x \in M} c^T x < \infty$. ◇

Proof. Für $n = 1$ ist der Satz richtig. Er gelte für $n < p$, und es sei $n = p$. Gelte $x_k \in M,\ c^T x_k \to v\ (k = 1, 2, ...)$. Besitzt die Folge einen Häufungpunkt x, löst er das Problem. Andernfalls folgt $\|x_k\| \to \infty$, und für eine unendliche Teilfolge konvergieren die beschränkten Vektoren $u_k = \frac{x_k}{\|x_k\|} \to u$. O.B.d.A. sei das schon die Ausgangsfolge. Wegen $Ax_k = \|x_k\| Au_k = b$ und $c^T x_k = \|x_k\| c^T u_k \to v$ muss nun $Au = 0$ und $c^T u = 0$ gelten. Damit folgt $u \geq 0, u \neq 0,\ A(x_k + tu) = b,\ c^T(x_k + tu) = c^T x_k\ \forall t \in \mathbb{R}$. Mit geeignetem $t = t_k \leq 0$ liegt der Punkt $h_k = x_k + t_k u$ auf dem Rand des nichtnegat. Orthanten. Damit ist eine seiner Komponenten, etwa die letzte, unendlich oft Null. Die Aufgabe $\max \{c^T x \mid x \in M,\ x_n = 0\}$ besitzt somit ebenfalls das Supremum v, was nach Ind. Vorauss. in einem $\bar{x} \in M$ angenommen wird. □

Analog zu diesem Beweis lässt sich (Indukt. über n) zeigen

Lemma 2.1.2 *Für jede* (m, n) *Matrix* A *ist die Menge* $K = \{z \in \mathbb{R}^m \mid z = Ax,\ x \geq 0\}$ *abgeschlossen.* ◇

Remark 2.1.3 Satz 2.1.1 gilt auch für quadratische Zielfunktionen $f(x) = c^T x + x^T Dx$ und den Fall von Polynomen f dritten Grades mit n reellen Variablen. Er wird falsch für Polynome vom Grade 4 und $x \in M = \mathbb{R}^2$.

Example 2.1.1 Man untersuche das Infimum von $f(x, t) = t^2 x^2 - 2t(1 - t)x$ für $x \geq 0$. ◇

Lemma 2.1.4 *(Farkaš) Die Ungleichung* $c^T u \leq 0$ *gilt für alle* u *mit* $Au \leq 0$ *genau dann, wenn* c *im Kegel* $K = \{z \mid z = A^T y,\ y \geq 0\}$ *liegt.* ◇

Proof. Wenn $c \in K$, gilt für die Zeilen von A im Falle $Au \leq 0$: $c^T u = \sum_i y_i \langle A_i, u\rangle \leq 0$. Wenn $c \notin K$, kann man nach den Lemmata 2.1.2 und 3.1.4 trennen: Ein $u \in \mathbb{R}^n$ erfüllt $\langle u, c\rangle > \langle u, z\rangle\ \forall z \in K$. *Bemerkung: Ohne direkt Lemma 3.1.4 anzuwenden, kann man auch die Euklidische Projektion* π *von* c *auf* K *nutzen und* $u = c - \pi$ *setzen.*

Da $0 \in K$ and $\lambda z \in K\ \forall \lambda > 0$ (falls $z \in K$), folgt also auch

$$\langle u, c\rangle > 0 \geq \langle u, A^T y\rangle = y^T Au \quad \forall y \geq 0.$$

Damit muss $Au \leq 0$ gelten; aus $Au \leq 0$ folgt also nicht $c^T u \leq 0$. □

Theorem 2.1.5 *(Dualitätssatz) Ist eine der Aufgaben*

$$\begin{array}{ll} (P) & \max\ c^T x\ ,\ x \in \mathbb{R}^n\ mit \quad Ax \leq b,\ x \geq 0 \\ (D) & \min\ b^T y\ ,\ y \in \mathbb{R}^m\ mit \quad A^T y \geq c,\ y \geq 0 \end{array} \tag{2.1.1}$$

lösbar, so auch die andere. Dann gilt ausserdem $v_P = v_D$ *für die Extremwerte.* ◇

Proof. Sei (P) lösbar und $\bar{x}$ eine Lösung für (P). Wir bilden $I(\bar{x}) = \{i \mid A_i\bar{x} = b_i\}$ und $J(\bar{x}) = \{j \mid \bar{x}_j = 0\}$. Weiter erfülle $u \in \mathbb{R}^n$ das System

$$A_i u \leq 0 \ \forall i \in I(\bar{x}), \quad u_j \geq 0 \ \forall j \in J(\bar{x}). \tag{2.1.2}$$

Dann ist $x(t) = \bar{x} + tu$ in (P) zulässig für kleine $t > 0$. Ihr Zielfunktionswert erfüllt daher wegen Optimalität von $\bar{x}$: $c^T\bar{x} \geq c^T x(t) = c^T\bar{x} + tc^T u$. Also folgt $c^T u \leq 0$ aus (2.1.2). Nach Lemma 2.1.4 exisieren dann $y_i \geq 0 (i \in I(\bar{x}))$ und $\mu_j \geq 0 (j \in J(\bar{x}))$ mit

$$c^T = \sum y_i A_i - \sum \mu_j e_j^T; \quad e_j = j\text{ter Einheitsvektor } \in \mathbb{R}^n. \tag{2.1.3}$$

Setzt man die restlichen Komponenten von y Null (um $y \in \mathbb{R}^m$ zu erreichen), wird y zulässig für (D), und man erhält wegen $\bar{x}_j = 0 \forall j \in J(\bar{x})$:

$$c^T\bar{x} = \sum_i y_i A_i\bar{x} - \sum_{j\in J(\bar{x})} \mu_j e_j^T\bar{x} = \sum_i y_i b_i = b^T y. \tag{2.1.4}$$

Schliesslich ist y optimal für (D), denn jedes (andere) zulässige y von (D) erfüllt

$$\sum_j c_j\bar{x}_j \ \leq \ \sum_j (\sum_i y_i a_{i,j})\bar{x}_j = \sum_i \sum_j y_i(a_{i,j}\bar{x}_j) \ \leq \ \sum_i y_i b_i, \tag{2.1.5}$$

kurz $c^T\bar{x} \leq y^T A\bar{x} \leq b^T y$. Also sichert die Lösbarkeit von (P) die von (D) sowie $v_P = v_D$. Die entsprechende Aussage unter Lösbarkeit von (D) zeigt man analog. □

Weil (2.1.5) für alle zul. Punkte $\bar{x}, y$ beider Aufgaben gilt, folgt aus Thm. 2.1.1: (P) und (D) sind lösbar, wenn beide Probleme zulässige Punkte besitzen. Da stets $c^T x \leq b^T y$ für ihre zul. Punkte gilt, hat man zugleich eine Abschätzung für die Extremalwerte v_P, v_D. Weiter folgt:

Corollary 2.1.6 *Zwei zulässige Punkte $\bar{x}, \bar{y}$ der Aufgaben (P) bzw. (D) sind genau dann optimal, wenn $c^T\bar{x} = b^T\bar{y}$ gilt, was dasselbe ist wie*

$$\bar{y}^T(A\bar{x} - b) = 0 = \bar{x}^T(A^T\bar{y} - c). \qquad \diamond \tag{2.1.6}$$

Wegen $\bar{y}_i \geq 0$ und $A_i\bar{x} - b_i \leq 0$, bedeutet die *Komplementaritätsbedingung* (2.1.6):

Gilt eine Ungleichung nicht als Gleichung, muss (in Lösungen)
die entsprechende duale Variable $\bar{y}_i$ bzw. $\bar{x}_j$ stets Null sein.

Damit sind die Aufgaben (P) und (D) zu einem Ungleichungssystem äquivalent.

Corollary 2.1.7 *Zwei Punkte $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ lösen (P) bzw. (D) genau dann, wenn*

$$Ax \leq b, \quad A^T y \geq c, \quad c^T x \geq b^T y, \quad x \geq 0, \ y \geq 0. \tag{2.1.7}$$

Aus (2.1.7) folgt ausserdem $c^T x = b^T y$. $\diamond$

Schreibt man (D) als Maximum Problem (P') in der Form von (P) und betrachtet die Dualaufgabe (D') zu (P'), so beschreiben (D') und (P) dieselben Aufgaben. Also sind (P) und (D) "gleichberechtigt" in dem Sinne, dass jede dual zur anderen ist.

Die allgemeine Dualaufgabe: Für die lineare Aufgabe (2.0.1), i.e.,

$$\min_{x\in M} c^T x \quad \text{mit } M = \{x \in \mathbb{R}^n \mid \ Ax \leq b\} \tag{2.1.8}$$

konstruiert man eine Dualaufgabe, indem man sie äquivalent in der Form (P) oder (D) aufschreibt. Z.B. kann man, mit Vorzeichenwechsel des Extremwertes, die Aufgabe

$$\max\ -c^T x; \quad Ax \le b$$

betrachten und diese weiter umschreiben (mittels $x = u - v,\ u \ge 0,\ v \ge 0$) als

$$\max\ -c^T u + c^T v; \quad Au - Av \le b,\ u \ge 0,\ v \ge 0.$$

Diese Aufgabe besitzt die Form von (P) mit $x = (u, v)^T$. Ihre Dualaufgabe ist demnach

$$\min\ b^T y; \quad A^T y \ge -c,\ -A^T y \ge c,\ y \ge 0 \quad \text{bzw.}$$

$$\min\ b^T y; \quad A^T y = -c,\ y \ge 0$$

und nach Übergang zum Maximum (erneut Vorzeichenwechsel des Extremalwertes) erhalten wir eine Form der Dualaufgabe zu (2.1.8) als

$$\max\ -b^T y; \quad A^T y = -c,\ y \ge 0, \quad \text{bzw. mit } z = -y$$

$$\max\ b^T z; \quad A^T z = c,\ z \le 0. \tag{2.1.9}$$

Vergleicht man (2.1.8) mit der Aufgabe (D) des dualen Paares, so unterscheidet sich (2.1.9) von (P) (abgesehen von der Umbenennung der Variablen und von b und c) durch die folgende Regel:

Remark 2.1.8 Hat man, im Vergleich zum symmetrischen Paar (P), (D), umgekehrte Ungleichungen, so werden die Vorzeichenbedingungen der entsprechenden Dualvariablen zu "≤ 0". Hat man Gleichungen, sind die entsprechenden Dualvariablen nicht vorzeichenbeschränkt. Dabei gilt jeweils auch das Umgekehrte. ◇

Diese Regel ist für die Dualisierung beliebiger linearer Aufgaben richtig (nach analogen Umformungen wie oben).
Man zeige durch äquivalente Umformungen wie oben: Die Aufgaben

$$\max_{x \in M} c^T x \quad \text{mit } M = \{x \in \mathbb{R}^n \mid\ Ax \ge b\} \tag{2.1.10}$$

$$\min_{y \in M} b^T y \quad \text{mit } M = \{y \in \mathbb{R}^m \mid\ A^T y = c,\ y \le 0\} \tag{2.1.11}$$

sind dual zueinander.

2.2 Die Simplexmethode

Für die folgende klassische Lösungsmethode betrachtet man Aufgaben *in Gleichungsform*:

$$\max_{x \in M} c^T x \quad \text{mit } M = \{x \in \mathbb{R}^n |\ Ax = b, x \ge 0\}. \tag{2.2.1}$$

Dabei sind $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und eine (m, n) Matrix A gegeben. Wir unterstellen weiter rank $A = m$ (sonst wären Zeilen des GL-Systems redundant) und $m < n$.
Vorbemerkungen:
Das System $Ax = b$ beschreibt einen $n - m$- dimensionalen affinen Unterraum U des $\mathbb{R}^n$. Er kann in unterschiedlicher Weise dargestellt werden. Seien dazu eine Zerlegung von A und x in der Form

$$A = (A_B, A_N), \quad x = (x_B, x_N)^T$$

gegeben, so dass A_B eine reguläre (m,m)- Matrix ist und A_N die Restmatrix aus den übrigen Spalten kennzeichnet. Wir verstehen B als Indexmenge der in A_B aufgenommenen Spalten von A (sie bilden eine Basis des Spaltenraumes von A) und N als entsprechende Komplementärmenge. Die Menge M ist dann äquivalent darstellbar in der Form

$$A_B x_B + A_N x_N = b, \quad x_B \geq 0, \ x_N \geq 0. \tag{2.2.2}$$

Schreibt man analog $c^T x = c_B^T x_B + c_N^T x_N$, und benutzt $Q = A_B^{-1}$, kann man mittels (2.2.2)

$$x_B = Qb - QA_N x_N \tag{2.2.3}$$

substituieren und die Aufgabe mit $n-m$ Variablen schreiben: Man finde

$$\max c_B^T(Qb - QA_N x_N) + c_N x_N \text{ wobei } Qb - QA_N x_N \geq 0, \ x_N \geq 0. \tag{2.2.4}$$

Mit entsprechenden Abkürzungen hat das Problem die Form

$$\max d_{0,0} - d_{0,N} x_N \text{ wobei } d_{B,0} - d_{B,N} x_N \geq 0, \ x_N \geq 0. \tag{2.2.5}$$

Hier ist

$$\begin{array}{lll} d_{0,0} & = c_B^T Qb \in \mathbb{R}, & d_{0,N} = c_B^T QA_N - c_N \in \mathbb{R}^{n-m}, \\ d_{B,0} & = Qb \in \mathbb{R}^m, & d_{B,N} = QA_N \text{ eine } (m, n-m) \text{ Matrix.} \end{array} \tag{2.2.6}$$

In der Folge betrachten wir spezielle Punkte (die "Ecken") aus M. Ein Punkt $\bar{x} \in \mathbb{R}^n$ heisst *zulässiger Basispunkt* zur "Basis" B, wenn er (2.2.2) mit $\bar{x}_N = 0$ erfüllt. Dann gilt mit den transformierten Grössen: $d_{B,0} = \bar{x}_B \geq 0$, $c^T \bar{x} = d_{0,0}$, d.h., $\bar{x}_N = 0$ ist zulässig für (2.2.5). Wegen der Forderung $x_N \geq 0$ ist $\bar{x}_N = 0$ gewiss Lösung von Aufgabe (2.2.5), wenn

$$d_{0,N} \geq 0 \tag{2.2.7}$$

(in jeder Komponente) gilt. Wegen der durch (2.2.3) vermittelten Äquivalenz von (2.2.1) und (2.2.5) ist dann $\bar{x}$ auch Lösung von (2.2.1).

Simplexmethode:
Die Methode besteht darin, *zulässige Basispunkte* mit wachsenden Werten der Zielfunktion zu konstruieren. Der Einfachkeit halber nehmen wir an, die Aufgabe sei *nicht entartet* im folgenden Sinne: Für alle (der endlich vielen) zul. Basispunkte gelte $d_{i,0} = \bar{x}_i > 0 \ \forall i \in B$.

Weiter möge ein zul. Basispunkt $\bar{x}$ mit Basis B und ausgerechneter Matrix D gegeben sein (Wie man ihn findet falls $M \neq \emptyset$, zeigen wir später). Gilt $d_{0,N} \geq 0$, sind wir fertig; also sei

$$d_{0,k0} < 0 \text{ für ein } k0 \in N.$$

Wir halten irgendein solches $k0$ fest und sehen anhand von (2.2.5), dass die Punkte

$$x_N(t) = (0, .., 0, t, 0, .., 0)^T \text{ mit } t > 0 \text{ in Komponente } k0$$

die Zielfunktion vergrössern. Sie bleiben, wegen $d_{B,0} > 0$, für kleine t zulässig mit

$$x_B(t) = d_{B,0} - d_{B,N} x_N(t) = d_{B,0} - d_{B,k0} t. \tag{2.2.8}$$

Offenbar darf t unbeschränkt wachsen (und hat (2.2.1) ein unbeschränktes Supremum), wenn keine Komponente von $d_{B,k0}$ positiv ist (Unlösbarkeit). Andernfalls gibt

$$\tau = \min_i \{d_{i,0}/d_{i,k0} \mid d_{i,k0} > 0\} \tag{2.2.9}$$

das grösste t an, mit dem $x_N(t)$ noch zulässig in (2.2.5) bleibt. Werde τ in $i = i0$ angenommen. Für $t = \tau$ wird dann Komponente $x_{i0}(t)$ aus (2.2.8) Null. Die Matrix $A_{B'}$ aus den Spalten zu

$$B' = \{k0\} \cup (B \setminus \{i0\})$$

ist erneut regulär (wieso ?), und nach Definition ist der Punkt

$$x' = x(\tau) = (x_B(\tau), x_N(\tau))$$

ein passender zul. Basispunkt. Wegen Nicht- Entartung sind die Komponenten von $x'_{B'}$ alle positiv, also war das Element $i0$ aus (2.2.9) eindeutig. Schliesslich gehört zu x' ein besserer Zielfunktionswert als zu $\bar{x}$. Bei Entartung wäre hier einfach $\tau = 0$ und der Wert würde sich nicht ändern. Wohl aber die Menge B' und damit auch $A_{B'}$.

Zusammenfassung:
Wenn die hinr. Optimalitätsbeding. (2.2.7) verletzt ist, entdecken wir entweder Unlösbarkeit der Aufgabe ($c^T x$ wächst unbeschränkt) oder wir können den aktuellen zul. Basispunkt $\bar{x}$ durch einen besseren x' (mit neuer Basis B') ersetzen. Im zweiten Fall setze man $\bar{x} := x', B := B'$ und wiederhole die Prozedur.
Gehört zur Original-Aufgabe ein endliches Supremum und *gibt es einen zul. Basispunkt*, so erhalten wir eine Folge zul. Basispunkte mit wachsenden Werten der Zielfunktion. Da es nur endlich viele $BasismatritzenA_B$ gibt, ist so nach endlichen vielen Schritten die Bedingung (2.2.7) erfüllt und die Methode bricht mit einer (Basis-) Lösung ab. Damit hätte man zugleich den Existenzsatz der lin. Opt. bewiesen. Man braucht nur:

Lemma 2.2.1 *Es existiert ein zul. Basispunkt für Aufgabe (2.2.1), wenn $M \neq \emptyset$ und rank $A = m$ ist.*

Proof. Man wähle $\bar{x} \in M$ mit maximal vielen Nullkomponenten und bilde

$$B1 = \{i \mid \bar{x}_i > 0\}, \quad B2 = \{k \mid \bar{x}_k = 0\}.$$

Wenn die, sagen wir p, Spalten der zugeordeten Matrix $A1$ linear abhängig sind, nehme man ein $u \neq 0$, so dass $A1u = 0$. Ergänzt man u durch $n-p$ Nullen zu einem Element aus $\mathbb{R}^n$, bleibt $A(\bar{x}+tu) = b\ \forall t$ richtig. Da ein u_i (zu $B1$) nicht Null ist, kann man mit passendem t eine weitere Komponente von $(\bar{x}+tu)_{B1}$ zu Null machen, was der Maximalität von $B2$ widerspricht. Also bilden die Spalten der Matrix $A1$ ein linear unabhängiges System und können mit $m-p$ geeigneten Spalten zu $B2$ zu einer Basis des m-dimensionalen Spaltenraumes der Matrix A ergänzt werden (Basis-Ergänzungs-Satz). Die zusammengesetze (m, m) Matrix ist daher regulär und folglich eine Basismatrix zu $\bar{x}$. □

Folgerung:
Ist die Aufgabe (2.2.1) nicht entartet, $M \neq \emptyset$ und rank $A = m$, so besitzt (2.2.1) entweder eine unbeschränkte Zielfunktion oder es gibt unter ihren endlich vielen zul. Basispunkten eine Lösung. Letztere kann man mit der angegeben Methode bestimmen.

Die Methode heisst Simplexmethode und wurde in den 40-ger Jahren entwickelt (Dantzig und Kuratovski). Um sie vollständig zu beschreiben, müssen wir noch klären, wie die zum neuen Punkt x' gehörige Matrix D' möglichst billig aus D bestimmt werden kann. Dazu denken wir uns D und die zugeordneten Variablen in der folgenden Form aufgeschrieben

$(r = n - m)$:

$$\begin{array}{cccccccc} & & x_{k1} & ... & x_{kq} & ... & x_{kr} \\ & d_{0,0} & d_{0,1} & ... & d_{0,q} & ... & d_{0,r} \\ x_{i1} & d_{1,0} & d_{1,1} & ... & d_{1,q} & ... & d_{1,r} \\ & .... & & & & & \\ x_{ip} & d_{p,0} & d_{p,1} & ... & d_{p,q} & ... & d_{p,r} \\ & .... & & & & & \\ x_{im} & d_{m,0} & d_{m,1} & ... & d_{m,q} & ... & d_{m,r} \end{array} \tag{2.2.10}$$

Die Indizes $i1, ..., im$ bilden B, die Indizes $k1, ..., kr$ bilden N, und die Zeile zu ip bedeutet die p-te Gleichung

$$x_{ip} = d_{p,0} - d_{p,1}x_{k1}... - d_{p,q}x_{kp}... - d_{p,r}x_{kr} \tag{2.2.11}$$

aus Darstellung (2.2.3) von x_B bzw. aus $x_B = d_{B,0} - d_{B,N}x_N$. Die Zahl $d_{0,0}$ ist der Wert der Zielfunktion im aktuellen Punkt, die restlichen D Komponenten der 0-ten Zeile bzw. Spalte bilden die Vektoren $d_{0,N}$ (zugeordnet der Zielfunktion) und $d_{B,0}$ (zugeordnet dem Wert von $\bar{x}_B$). Für den neuen Punkt x' wollen wir die Daten D' analog und am selben Platz aufschreiben.
Seien die kritischen Indizes $i0, k0$ gerade ip und kq in unserem *Simplextableau* (2.2.10). Man beachte dass $q > 0$ nur

$$d_{0,q} < 0$$

erfüllen musste (gibt es so ein q nicht, ist die Aufgabe gelöst), während Index p über den minimalem Quotienten

$$\frac{d_{i,0}}{d_{i,q}} \qquad \text{mit } d_{i,q} > 0,\ i > 0$$

mittels der Spalten 0 und q festgelegt war, siehe (2.2.9). Gibt es kein solches $d_{i,q} > 0$, ist die Aufgabe unlösbar. In der nächsten Tabelle vertauschen einzig x_{kq} und x_{ip} ihren Platz, weil nun gerade $kq \in B'$ und $ip \in N'$ gilt. Um D' zu bestimmen, haben wir nur (2.2.11) nach der neuen "Basisvariablen"

$$x_{kq} = \frac{d_{p,0}}{d_{p,q}} - \frac{d_{p,1}}{d_{p,q}}x_{k1} \ ... \ - \frac{1}{d_{p,q}}x_{ip} \ ... \ - \frac{d_{p,r}}{d_{p,q}}x_{kr}$$

aufzulösen und in die restlichen Gleichungen des Systems einzusetzen. In Zeile p zu ip (und neu zu kq) stehen dann also

$$\begin{array}{ll} x_{kq} & D'_{p,0} \quad D'_{p,1} \quad ... \ D'_{p,q} \quad ... \ D'_{p,r} \\ \text{mit} & D'_{p,k} = \frac{d_{p,k}}{d_{p,q}} \ \ (0 \le k \le r,\ k \ne q) \quad \text{und} \quad D'_{p,q} = \frac{1}{d_{p,q}}. \end{array}$$

Für die neue Tabelle der Form

$$\begin{array}{cccccccc} & & x_{k1} & ... & x_{ip} & ... & x_{kr} \\ & D'_{0,0} & D'_{0,1} & ... & D'_{0,q} & ... & D'_{0,r} \\ x_{i1} & D'_{1,0} & D'_{1,1} & ... & D'_{1,q} & ... & D'_{1,r} \\ & .... & & & & & \\ x_{kq} & D'_{p,0} & D'_{p,1} & ... & D'_{p,q} & ... & D'_{p,r} \\ & .... & & & & & \\ x_{im} & D'_{m,0} & D'_{m,1} & ... & D'_{m,q} & ... & D'_{m,r} \end{array} \tag{2.2.12}$$

wissen wir so, wie die sogenannte "Pivot-Zeile" zum "Pivot-Element" $d_{p,q}$ am Platz (p, q) aussieht. Nach der Einsetz-Methode bestimmt man die übrigen Elemente. Das Resultat

sieht so aus: Die restlichen Elemente der "Pivot-Spalte" q zu ip und kq entstehen aus Division durch $-d_{p,q}$:

$$D'_{i,q} = -\frac{d_{i,q}}{d_{p,q}} \qquad (0 \le i \le m,\ i \ne p).$$

Die übrigen Elemente werden mittels "Kreuzregel" bestimmt:

$$D'_{i,k} = \ d_{i,k} \ - \frac{d_{i,q}\ d_{p,k}}{d_{p,q}} \qquad (0 \le i \le m,\ i \ne p;\ 0 \le k \le r,\ k \ne q).$$

Der Punkt x' zu (2.2.12) ist optimal, wenn alle Elemente der Zeile 0, ausser $D'_{0,0}$, nichtnegativ sind. Andernfalls wähle man eine Spalte k mit ... und wiederhole.

Bemerkungen zu den Voraussetzungen:
Die Rangvoraussetzung kann für grosse Aufgaben lästig sein, da das Bestimmen redundanter Gleichungen vorausgeschickt werden muss, was allerdings nicht prinzipiell schwierig ist. Wenn zul. Basispunkte entartet sind, kann man wie angegeben von $\bar{x}$ zu x' übergehen, allerdings ist $\tau = 0$ möglich, wonach sich Basismatrizen A_B wiederholen könnten (nach mehreren Schritten, nicht unmittelbar) und das Verfahren so nicht abbricht. Die Wiederholung von Basismatrizen lässt sich durch verfeinerte Regeln für die Auswahl des Elements $i0$ (der Pivot-Zeile) vermeiden. Falls es mehrere Kandidaten für $i0$ gibt, muss dann die Auswahl anhand weiterer Quotienten erfolgen. Man sichert so, dass der Vektor $d_{0,N}$ (ergänzt durch weitere Komponenten) in einer gewissen Ordnung (der lexikographischen) wächst, was erneut das Auftreten von Zyklen verhindert. Man findet diese Algorithmen in der Literatur unter dem Stichwort "lexikographische Simplexmethode".

Es gibt weitere Modifikationen dieser Methode wie "duale" oder "revidierte" Simplexmethode. Oft werden lineare Aufgaben aber auch völlig anders, mittels speziell angepasster Methoden vom Newton-Type gelöst (siehe "Innere Punkt Methoden", "Ellipsoid Methoden"). Prinzipiell ist wegen Corollary 2.1.7 jedes Verfahren geeignet, das ein lineares Ungleichungssystem lösen kann oder, wie wir in sect. 2.7 sehen, ein Matrixspiel.

Finden eines Startpunktes:
Um die Methode anzuwenden, braucht man einen zul. Basispunkt für Aufgabe (2.2.1)

$$\max_{x \in M} c^T x \quad \text{mit } M = \{x \in \mathbb{R}^n \mid Ax = b, x \ge 0\}$$

oder man stellt fest, dass es keinen zul. Punkt gibt. Hierzu kann man zunächst die Gleichungen $A_i x = b_i$ ($b_i < 0$) mit -1 multiplizieren, so dass dann $b \ge 0$ gilt. Anschliessend beschäftigt man sich mit der Aufgabe

$$\min \{ \sum_i u_i \mid Ax + Eu = b,\ x, u \ge 0\}. \tag{2.2.13}$$

Hier findet man einen zulässigen Startpunkt sofort: $u = b, x = 0, A_B = E$. Also kann man die Simplexmethode anwenden, um (2.2.13) zu lösen. Ist der Optimalwert v grösser als Null, hat (2.2.1) keinen zul. Punkt. Ist $v = 0$, folgt $u = 0$ in der Lösung. Im Falle der Nichtentartung (den wir wieder voraussetzen), kann dann keine Variable u_i (in der letzten Simplex-Tabelle) Basisvariable sein. Diese Tabelle liefert uns also eine Darstellung des Systems $Ax + Eu = b$ in der Form

$$x_B = d_{B,0} - d_{B,N}\ x_N - Ru$$

mit einer (m, m)-Restmatrix R. Nach Streichen von Ru haben wir die erforderliche Darstellung für x in Aufgabe (2.2.1). Es bleibt einzig,

$$x_B = d_{B,0} - d_{B,N} x_N$$

in $c^T x$ zu substituieren, um Zeile $d_{0,N}$ und $d_{0,0}$ aus (2.2.5) zu erhalten, d.h. die Zielfunktion in der Form

$$c^T x = d_{0,0} - d_{0,N} x_N$$

zu schreiben. Damit hat man nun auch die erste Zeile (Nr. 0) der Simplextabelle. Die angegebene Methode zur Bestimmung eines Startpunktes heisst auch *Methode der künstlichen Variablen* (hier u).
Wir kommen nun zu Interpretationen und speziellen lin. Optimierungsaufgaben.

2.3 Schattenpreise

Die Aufgaben

$$\begin{array}{ll} (P) & \max\ c^T x\ ,\ x \in \mathbb{R}^n \text{ mit } \quad Ax \le b,\ x \ge 0 \\ (D) & \min\ b^T y\ ,\ y \in \mathbb{R}^m \text{ mit } \quad A^T y \ge c,\ y \ge 0 \end{array}$$

sind wie folgt interpretierbar. Die Zielfunktion $\sum c_j x_j$ beschreibt einen zu maximierenden Gewinn bei der Produktion von $x_j \ge 0$ Einheiten eines Produktes P_j. Die Summen $\sum_j a_{ij} x_j$ stellen den Verbrauch einer Ressource R_i beim aktuellen Produktionsvektor x dar, die vorhandene Menge an Ressource R_i sei b_i.
Sei $v(b)$ der Extremalwert dieser Aufgabe (P) in Abhängigkeit von den Ressourcen b_i. Wir variieren b durch einen normkleinen Vektor β und nehmen an, dass die resultierenden Aufgaben alle lösbar bleiben. Dann ändert sich v um $\delta v = v(b + \beta) - v(b)$.
Die Preise der Ressourcen R_i auf dem Markt seien p_i pro Einheit. Durch Kauf/Verkauf von β_i Einheiten R_i lässt sich also ein Ertrag $E = \sum_i p_i\ \beta_i$ erzielen. Ist $E > \delta v$, wäre der Verkauf sogar sinnvoll. Umgekehrt ist es bei $E < \delta v$ vernünftig, Ressourcen zu kaufen und dann mit $b' = b + \beta$ optimal zu produzieren. Daher wird die Frage interessant, bei welchen Preisen $p_i \ge 0$ die Identität

$$v(b + \beta) - v(b) = \sum_i p_i\ \beta_i \quad \text{für kleine } \|\beta\| \tag{2.3.1}$$

gilt. Mittels Dualität lässt sich leicht zeigen (ÜA):
Wenn (P) und (D) eindeutig lösbar sind, bedeutet (2.3.1), dass $y = p$ die Aufgabe (D) löst. Die Lösung der Dualaufgabe hat also eine ökonomisch Bedeutung. Die Komponenten y_i für optimale y heissen in der Ökonomie auch *"Schattenpreise"*.

Für nichtlineare, konvexe Aufgaben folgt die analoge Bedeutung der dualen Lösung aus der Tatsache, dass sie (bis aufs Vorzeichen) ein Subgradient der Extremwertfunktion ist; im differenzieraren Fall also ein Gradient, was (2.3.1) bis auf einen Fehler $o(\beta)$ sichert.

2.4 Klassische Transportaufgabe

Gegeben seien m Lieferanten L_i und n Kunden K_j für ein beliebig teilbares Gut (Kies). Lieferant L_i besitze a_i Tonnen des Gutes, der Bedarf von Kunde K_j seien b_j Tonnen. Aufgrund der Entfernungen möge der Transport einer Tonne des Gutes von L_i zu K_j gerade c_{ij} Euro kosten (und im übrigen linear sein). Es gelte sogar die Bilanzgleichung $\sum_i a_i = \sum_j b_j$.

Gesucht sind Liefermengen $x_{ij} \geq 0$ (von L_i zu K_j), so dass der Bedarf gedeckt wird und die Kapazitäten der Lieferanten nicht überschritten werden. Dabei sollen die Gesamtkosten $G(x) = \sum_i \sum_j c_{ij}\ x_{ij}$ minimal werden. Nebenbedingungen sind

$$\sum_j x_{ij} = a_i\ , \quad \sum_i x_{ij} = b_j; \quad x \geq 0.$$

Zielfunktion und Nebenbedingungen sind linear. Der Variablenvektor hat Dimension $d = mn$. Die Aufgabe könnte mittels Simplexmethode gelöst werden (eine Gleichung ist redundant), nur bekommt die Matrix D sehr grosse Dimension. Deshalb sind hier andere Methoden zweckmässiger, die den Dualitätssatz und die Tatsache nutzen, dass die Dualaufgabe jetzt nur $n + m$ Variablen besitzt (z.B. "Potentialmethode" mittels dualer Variabler u_i, v_j; siehe Vorles.).

Wir geben hier nur einen Beweis für die zentrale Aussage, dass stets ein Zyklus (Kreis) existiert, der ein "Nicht-Kästchenelement" einbezieht.

Zyklenbeweis:

Nach k=1 Iterationen (Starttabelle) ist das klar wegen der Treppenstruktur der Kästchenelemente, die aus der NW-Eckenregel folgt.

Es sei richtig für alle Tabellen nach k Iterationen; wir betrachten die Situation für $k + 1$.

1. Bezeichne $C:\quad c_1, ...,\ c_p,\ ..., c_{m+n-1}$ die vorhandenen Kästchen (einschliesslich deren Position in der Matrix) und möge auf c_0 das einzubeziehende "Kringel-" Element stehen. Die vorhandenen Kästchen wurden im k -ten Schritt ueber einen Kreis verandert, der aus einem Kringel (er sei oben an Position c_p) ein Kästchen machte und dafür eines der im Schritt k vorhandenen Kästchen löschte.

2. Dieser "Lösch" -Kreis bestehe (in der angegebenen Reihenfolge) aus

$$A:\quad (c_p =)\ a_0,\ a_1, ...,\ a_{q-1},\ a_q,\ ..., a_r = a_0, \quad \text{wobei in Schritt } k+1,\ a_q \text{ gelöscht ist,}$$

also oben nicht auftritt, und a_0 neues Kästchen wird.

Wenn $c_0 = a_q$, bezieht Kreis A das Element c_0 wieder ein (Trivialfall). Sei also $c_0 \neq a_q$.

3. Mit den Kästchen im Schritt k (einschliesslich a_q) hätte man nach Ind.-Vorauss. das in $k + 1$ aktuelle Kringelelement c_0 in einen Kreis einbeziehen koennen. Dieser Kreis sei

$$B:\quad (c_0 =)\ b_0,\ b_1, ...,\ b_{q'-1},\ b_{q'},\ ..., b_s = b_0, \quad \text{wobei in Schritt } k+1,\ a_q = b_{q'} \text{ gelöscht ist.}$$

Dafür gibt es jetzt (in Schritt $k + 1$) das Element $a_0 = c_p$ als Kästchenelement.

4. Bis auf die genannten Elemente sind die Kreiselemente von A und B aus derselben Menge von Kästchenelementen: Unter A und B stellen wir uns nur die Folge der im Kreis durchlaufenen Elemente vor. Hierbei ist es gleichgültig, was wir als Anfangselement des Kreises auffassen und wie herum er orientiert ist. Wichtig ist jedoch, dass zwei Nachbarn abwechselnd in derselben Zeile/Spalte stehen.

Wir nutzen nun, dass $a_q = b_{q'}$.

5. Wir dürfen annehmen, dass die Elemente a_{q-1}, a_q in derselben Zeile stehen. Denn andernfalls stehen (wegen der Knickform eines Kreises) a_q, a_{q+1} in derselben Zeile. "Laufen" wir den Kreis andersherum (orientieren wir ihn also um), wird a_q, a_{q+1} das Vorgaengerpaerchen zu a_{q-1}, a_q. Nach Umnumerierung der Elemente entsprechend der Orientierung haben wir dann die Annahme erfüllt. Ebenso können wir voraussetzen, dass die Elemente $b_{q'-1}, b_{q'}$ in derselben Zeile stehen. Dann stehen auch a_{q-1} und $b_{q'-1} = a_q$ in derselben Zeile (nämlich in der von $b_{q'} = a_q$).

6. Wir bilden nun den gesuchten neuen Kreis:

Beginnend mit c_0 folgen wir dem Kreis B bis $b_{q'-1}$.

Von $b_{q'-1}$ "springen" wir - statt waagerecht zu $b_{q'}$ - zu a_{q-1}, das in derselben Zeile steht.

Von a_{q-1} folgen wir dem Kreis A (rueckwärts, um a_q auszusparen) bis zu a_{q+1}, was in der Spalte von a_q steht. Wegen $a_q = b_{q'}$ steht in dieser Spalte auch $b_{q'+1}$. Wir "springen" zu $b_{q'+1}$.
Von $b_{q'+1}$ aus gelangen wir (im B-Kreis vorwärts, ohne $b_{q'}$ zu benutzen) bis zu b_0.
Das ist der gesuchte Kreis. □

Eine besondere Transport-Aufgabe entsteht, wenn $m = n$ und $a_i = b_j = 1$. Sie heisst *Zuordnungsproblem* und ist "unangenehm", weil viele ihrer zulässigen Basispunkte entartet sind.

2.5 Maximalstromproblem

Es seien n Punkte p_i und n^2 Zahlen $c_{ij} \geq 0$ gegeben. Sie mögen die Kapazität einer Pipeline von p_i nach p_j in Tonnen pro Stunde für den Strom einer Flüssigkeit (oder auch einer Ladungsmenge in einem elekrischen Netzwerk) beschreiben. Gibt es keine Verbindung von p_i nach p_j, ist $c_{ij} = 0$. Weiter seien zwei Punkte, etwa p_1 und p_n als "Quelle" bzw. "Senke" ausgezeichnet. Es soll maximal viel Flüssigkeit von p_1 nach p_n geschickt werden, wobei in den übrigen Punkten nichts hinzukommt oder verschwindet. Bezeichnet x_{ij} den Fluss von p_i nach p_j (Tonnen pro Stunde), so besteht die Aufgabe darin, die Summe $S(x) = \sum_j x_{1,j}$ zu maximieren. Die Nebenbedingungen besitzen nun die Form

$$\sum_j x_{ij} - \sum_j x_{ji} = 0 \ \forall i: \ 1 < i < n; \quad 0 \leq x_{ij} \leq c_{ij}.$$

Diese lineare Optimierungsaufgabe heisst *Maximalstromproblem*. Ein erster und noch immer aktueller Algorithmus wurde für sie von Ford und Fulkerson in den 50-ger Jahren entwickelt. Er nutzt das Dualproblem (über den Begriff des Schnittes):
Die Zielfunktion ist nach oben beschränkt, ein zulässiger Fluss existiert stets, $x = 0$. Die Aufgabe besitzt also eine Lösung. Wie beim Transportproblem gibt es einen speziellen effektiven Algorithmus. Wir gehen hier nicht den Umweg über Dualisierung, sondern benutzen gleich die grundlegende Idee für solche Aufgaben.
Eine Zerlegung der Menge $P = \{p_1, ..., p_n\}$ in zwei (disjunkte) Teilmengen S, T mit $p_1 \in S, p_n \in T$ heisse Schnitt. Die Zahl

$$C(S,T) := \sum_{s \in S, \ t \in T} c_{st}$$

wird Kapazität des Schnittes genannt. Für jeden zul. Fluss folgt nun $F(x) \leq C(S,T)$ denn die in p_n ankommende Flüssigkeitsmenge muss notwendig aus S nach T fliessen:

$$F(x) \leq \sum_{s \in S, t \in T} x_{st} \leq \sum_{s \in S, t \in T} c_{st} = C(S,T).$$

Damit ist ein zul. Strom x optimal, wenn $F(x) = C(S,T)$ für einen Schnitt gilt. Von Ford und Fulkerson stammt die Verschärfung dieser Aussage als notwendige und hinreichende Bedingung und die Konstruktionsmethode für eine Lösung:
Ein zul. Strom x ist genau dann optimal, wenn $F(x) = C(S,T)$ für einen Schnitt gilt.
Beweis und Konstruktion:
Sei x ein zul. Fluss, seien (der Einfachkeit halber) alle c_{ij} ganzzahlig (bzw. rational). Man bilde $S_0 = \{p_0\}$ und setze $z(p_o) = \infty$.
Hat man $S_k \subset P$ und $z : S_k \to \mathbb{R} \cup \{\infty\}$ bilde man

$$S_{k+1} = S_k \cup \{p_j\}$$

wenn mit einem Paar $p_m \in S_k,\ p_j \in P \setminus S_k$ entweder gilt
(1) $x_{mj} < c_{mj}$; dann setze man $z(p_j) = \min\{z(p_m),\ c_{mj} - x_{mj}\}$,
oder
(2) $x_{jm} > 0$; dann setze man $z(p_j) = \min\{z(p_m),\ x_{jm}\}$.
Dazu merke man sich das Paar (m, j) sowie den Typ (1) bzw. (2) der erfolgten Konstruktion von S_{k+1}. Wenn beides möglich ist, entscheiden wir uns für einen der Schritte.

Es gibt nun 2 Fälle:

1. Die Konstruktion bricht nach k Schritten ab ohne dass $p_n \in S_k$ ist.
   Dann bilden $S = S_k$ and $T = P \setminus S$ einen Schnitt mit $F(x) = C(S, T)$,
   wonach x optimal ist.
2. Nach k Schritten ist $p_n \in S_k$. Dann gibt es eine Kette von Knotenindizes,

$$(1, n_1)\ (n_1, n_2)\ ...\ (n_{k-1}, n_k), \quad n_k = n,$$

wobei jeweils p_j aufgrund von (1) oder (2) aufgenommen wurde. Die Kette entsteht dadurch, dass wir rueckwärts, beginnend mit p_n aufschreiben, aufgrund welchen Knotens $p_{\nu-1}$ der Knoten p_ν in S aufgenommen wurde. Mit dem kleinsten z- Wert d bzgl. aller obigen Knoten, kann man dann x in den jeweiligen Bögen ändern:
$x'_{mj} = x_{mj} + d$ im Fall (1),
$x'_{jm} = x_{jm} - d$ im Fall (2), $\qquad x' = x$ in allen restlichen Knoten.
Nach Konstruktion ist x' wieder zulässig und F wächst um das ganzzahlige $d > 0$. Daher muss der Algorithmus nach endlich vielen Schritten eine Lösung liefern. □

2.6 The core of a game

Duality helps to solve games. Given a set $I = \{1, ..., n\}$ of "players" and $S \subset I$ (this allows $S = I$ by our convention) let $v(S)$ denote some payoff which the players of the *coalition* S can (at least) obtain, even if the others in $I \setminus S$ work against S. Assume that v is superlinear, i.e.,

$$v(S) + v(T) \le v(S \cup T) \ \text{ if } S \cap T = \emptyset.$$

Then $v(S)+v(I\setminus S) \le v(I),\ \sum_i v(\{i\}) \le v(I)$ and the total payoff $v(I)$ can be distributed such that player i gets x_i and $\sum_{i\in I} x_i \le v(I)$. Clearly, some coalition S is dissatisfied with x if $\sum_{i\in S} x_i < v(S)$. Let $\emptyset \ne S \ne I$ in the following. So we have the problem

$$\text{Find } x \in \mathbb{R}^n \text{ such that } v(S) \le \sum_{i\in S} x_i \ \ \forall S \quad \text{ and } \ \sum_{i\in I} x_i = v(I). \tag{2.6.1}$$

The set of these x is called the *core of the game* (I, v).

We are interested in conditions for getting a nonempty core. To simplify the interpretation one may suppose that all $v(\{i\})$ (hence also x_i and $v(S)$) are nonnegative.

Idea: Read (2.6.1) as a constraint system of a linear problem with objective $\min \sum_{i\in I} 0x_i$. The dual problem has variables y_S, y_I and constraints, assigned to all $x_i \in \mathbb{R}$, namely

$$\max\ d := v(I)y_I + \sum_S v(S)y_S \ \text{ where } \ y_I + \sum_{S:\ i\in S} y_S = 0\ \forall i, \quad y_S \ge 0. \tag{2.6.2}$$

There is no feasible y with $y_I > 0$. So we substitute $z = -y_I \ge 0$ which gives equivalently

$$\max\ d := -v(I)z + \sum_S v(S)y_S \text{ where } \sum_{S:\ i\in S} y_S = z\ \forall i, \quad y_S \ge 0, z \ge 0 \tag{2.6.3}$$

We investigate solvability of this problem. If $d > 0$ for feasible y, z, so multiply with $\lambda > 0$ to see that $d \to \infty$ as $\lambda \to \infty$. Hence for solvability, it has to follow $\max d = 0$. In other words, the constraints in (2.6.3) have to imply $\sum_S v(S)y_S \leq v(I)z$. This is trivial for $z = 0$ and equivalent for all positive z, thus - by duality - solvability of (2.6.3) and (2.6.1) means

$$\sum_{S:\ i\in S} y_S = 1\ \forall i \ \text{ and } y_S \geq 0 \ \Rightarrow\ \sum_S v(S)y_S \leq v(I). \tag{2.6.4}$$

Of course, only positive y_S are of interest. So define:

A set (sometimes called family or system) σ of subsets S of I is said to be *balanced* if positive y_S, $S \in \sigma$ exist such that $\sum_{S\in\sigma:\ i\in S}\ y_S = 1\ \forall i$.

Setting $y_S = 0$ for all $S \notin \sigma$ (if such S exist), y is feasible. Conversely, if y satisfies our constraints then all S with $y_S > 0$ form a balanced family σ. Finally, define

The *game* (I, v) *is called balanced* if for each balanced family σ and related y_S the inequality $\sum_S v(S)y_S \leq v(I)$ is valid.

Then, being balanced is the same as (2.6.4), and we have already shown the main

Theorem 2.6.1 *(on balanced games): The core is not empty iff the game is balanced.* $\diamond$

How to interpret a balanced family ? The sets in σ are certain clubs (collecting stamps, playing chess, cooking, ...) of the players, and a single player may be a member of a given club $S \in \sigma$ or not. If he is a member of S, he has to pay some fee $y_S > 0$ (per month). The fees of the family σ are fair if the total amount $a_i := \sum_{S\in\sigma:\ i\in S}\ y_S$ is the same for all players ($a_i = 1$ without loss of generality), and a family σ is balanced if fair fees exist (if e.g. σ consists of all S without player 1 then this is impossible).
If the fee y_S is large then so is the term $v(S)y_S$ in the sum, and the value $v(I)$ has to be sufficiently large.
Example 1: $n = 3, v(S) = 1\ if\ |S| \geq 2,\ v(S) = 0$ otherwise; core $= \emptyset$.
Example 2: $n = 3, v(I) = 3/2,\ v(S) = 1\ if\ |S| = 2,\ v(S) = 0$ otherwise; core $\neq \emptyset$.
Exercise: Find balanced and not balanced sets σ for $n = 4$.

2.7 Äquivalenz: Matrixspiel und Lineare Optimierung

2.7.1 Matrixspiel

Es sei eine $(m, n)-$ Matrix $A = (a_{ij})$ gegeben. Zwei Spieler (1) und (2) mögen voneinander unabhängig den Index i einer Zeile bzw. den einer Spalte j wählen. Im Ergebnis erhalte (1) den Gewinn a_{ij} von (2).
Bei Wahl von j durch (2) kann (1) offenbar höchstens den Gewinn $max_i\ a_{ij}$ erzielen. Realisiert die aktuelle Strategie $i = i0$ schon das Maximum, hätte Spieler (1) -bei Wahl von j durch Spieler (2)- keinen Grund abzuweichen.
Dasselbe gilt für Strategien in Bezug auf $min_j\ a_{ij}$ aus der Sicht von Spieler (2).
Also sind Paare von Strategien $(i0, j0)$ interessant, die gerade

$$\max_i\ a_{i,j0} = a_{i0,j0} = \min_j\ a_{i0,j}$$

erfüllen. Sie heissen in der Spieltheorie (Nash-) Gleichgewichtssituationen, kurz GGS, cf. sect. 3.3. Inhaltlich drücken sie eine gewisse Stabilität aus: Kein Spieler kann seinen Gewinn erhöhen, indem er allein von dem gegebenen Strategientupel abweicht. Allerdings gibt es (Matrix-) Spiele ohne eine solche Lösung.

Gemischte Erweiterung:
Angenommen, das Spiel wird oft gespielt. Dann kann Spieler (1) seine Strategien mit gewissen Wahrscheinlichkeiten (in diesem Moment seien das einfach relative Häufigkeiten) $x_1, ..., x_m$ wählen, wobei $x_i \geq 0$ und $\sum_i x_i = 1$ gilt; kurz $x \in X$. Gegen Strategie j von (2) wird sein (zu erwartender) Gewinn dann $g(x, j) = \sum_i x_i a_{ij}$. Analog kann Spieler (2) seine Strategien mischen mittels $y = (y_1, ..., y_n)$, wobei $y_j \geq 0$ und $\sum_j y_i = 1$ gilt; kurz $y \in Y$. Bei Anwendung von x und y ergibt sich so der Gewinn

$$G(x, y) = \sum_j y_j \ (\sum_i x_i a_{ij}) = x^T A y$$

für Spieler (1). Eine GGS $(\bar{x}, \bar{y})$ in diesen neuen, gemischten Strategien, ist nun definiert durch die Bedingung

$$\max_{x \in X} G(x, \bar{y}) = G(\bar{x}, \bar{y}) = \min_{y \in Y} G(\bar{x}, y). \tag{2.7.1}$$

Sie impliziert offenbar über $\eta = \bar{y}, \xi = \bar{x}$,

$$\min_{\eta \in Y} \max_{x \in X} G(x, \eta) \leq G(\bar{x}, \bar{y}) \leq \max_{\xi \in X} \min_{y \in Y} G(\xi, y). \tag{2.7.2}$$

Für beliebige $\xi \in X, \eta \in Y$ gilt ferner trivial $\max_{x \in X} G(x, \eta) \geq G(\xi, \eta) \geq \min_{y \in Y} G(\xi, y)$. Geht man in der Ungleichung $\max_{x \in X} G(x, \eta) \geq \min_{y \in Y} G(\xi, y)$ links zum Minimum bzgl. η über, rechts zum Maximum bzgl. ξ, folgt also auch

$$\min_{\eta \in Y} \max_{x \in X} G(x, \eta) \geq \max_{\xi \in X} \min_{y \in Y} G(\xi, y), \tag{2.7.3}$$

wonach (2.7.2), wenn überhaupt, als Gleichung gilt und gerade (2.7.1) bedeutet mit Punkten $\xi = \bar{x}$ und $\eta = \bar{y}$, die die äusseren Extrema realisieren. Der folgende Zusammenhang zwischen Matrixspiel und linearer Optimierung geht zurück auf [37].

2.7.2 Matrixspiel als LOP; Bestimmen einer GGS

Wir suchen nun ein $\xi \in X$, welches das Maximum

$$\max_{\xi \in X} \phi(\xi) \text{ mit } \phi(\xi) = \min_{y \in Y} G(\xi, y)$$

auf der rechten Seite von (2.7.2) realisiert. Für festes $\xi \in X$ ist G linear in y. Also wird $\min_{y \in Y} G(\xi, y)$ in einem der n Einheitsvektoren $e_j \in Y$ angenommen (wieso ?). So folgt

$$\phi(\xi) = \min_{y \in Y} G(\xi, y) = \min_j G(\xi, e_j) = \min_j \xi^T A e_j = \min_j \xi^T A_{\cdot,j}$$

$$\text{und} \quad \phi(\xi) = \min_j \xi^T A_{\cdot,j} = \max\{ \ t \in \mathbb{R} \mid t \leq \xi^T A_{\cdot,j} \ \forall j\}.$$

Damit hat unser Problem die Form

$$\max \ t \ \text{ so dass } \xi \in X \text{ und } \xi^T A_{\cdot,j} - t \geq 0 \ \forall j. \tag{2.7.4}$$

Das ist eine (stets lösbare) lineare Aufgabe mit gesuchtem (t, ξ). Analog kann man ein $\eta \in Y$ suchen, welches das Minimum

$$\min_{\eta \in Y} \psi(\eta) \ \text{ mit } \psi(\eta) = \max_{x \in X} G(x, \eta)$$

auf der linken Seite von (2.7.2) realisiert. Man stellt dann fest, dass die resultierende Aufgabe gerade dual zu (2.7.4) ist. Daher sind beide Aufgaben lösbar, und ihre Lösungen erfüllen (2.7.2) wegen Gleichheit der Extremalwerte.

In dieser Weise sichert der Dualitätssatz die Existenz einer GGS (für gemischte Strategien) in einem Matrixspiel (das ist der Hauptsatz für Matrixspiele (J. v. Neumann). Zugleich kann eine GGS, wie angegeben, mittels linearer Optimierung bestimmt werden.

2.7.3 LOP als Matrixspiel; Bestimmen einer LOP-Lösung

Umgekehrt kann man lineare Optimierung als ein Matrixspiel verstehen. Sei dazu

$$s^T = (\xi, \eta, t) \in \mathbb{R}^{n+m+1}$$

eine (gemischte) Strategie im Matrixspiel G mit der aus dem dualen Paar (P), (D) aus (2.1.1) gebildeten quadratischen (und schiefsymmetrischen, d.h. $g_{ij} = -g_{ji}$) Matrix

$$G = G(A, b, c) := \begin{pmatrix} 0 & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & 0 \end{pmatrix}. \tag{2.7.5}$$

Dann gilt stets $s^T G s = 0$, womit für jede Gleichgewichtssituation (s, σ) von G folgt, dass $s^T G \sigma = 0$ gilt. Beweisen Sie damit und mittels Dualitätssatz der linearen Optimierung (es reicht Corollary 2.1.7):

(1) Ein Strategienpaar (σ, s) ist eine Gleichgewichtssituation des Spiels G genau dann, wenn s und σ das Ungleichungssystem $Gs \leq 0$ erfüllen.

(2) Ist s eine Strategie mit $Gs \leq 0$ und $t > 0$, so lösen $x = \xi/t$ und $y = \eta/t$ die Aufgaben (P) bzw. (D).

(3) Lösen x und y die Aufgaben (P) bzw. (D), so ist $s = (\xi, \eta, t)$ mit

$$\xi = tx,\ \eta = ty,\ t = (1 + \sum x_j + \sum y_i)^{-1}$$

eine Strategie des Spiels G, die $Gs \leq 0$ und $t > 0$ erfüllt.

Remark 2.7.1 Matrixspiele sind also lineare Optim.Aufgaben und umgekehrt. ◇

Ein erster Algorithmus zur Lösung von Matrixspielen stammt von Julia Robinson [123] (1951). Er benötigt viele Iterationen mit sehr einfachen Schritten. Wir beschreiben ihn unter der Voraussetzung, dass A (wie oben G) schiefsymmetrisch ist.

2.8 The Julia Robinson Algorithm

We describe the J. Robinson algorithm for a *skew-symmetric* (n, n) matrix A (because of the above transformations, this is the basic case). We want to find some x in the simplex of mixed strategies such that $Ax \leq 0$.

(i) Put $Y^0 = 0 \in \mathbb{R}^n$ and choose any $Z^0 \in \mathbb{R}^n$ such that $\max Z^0 := \max_i Z^0_i = 0$.

(ii) Step $s \geq 0$: Determine some maximal component of Z^s (say its index is $i = i(s)$) and add the ith unit vector and the ith column of A, respectively:

$$Y^{s+1} = Y^s + e^i, \qquad Z^{s+1} = Z^s + A_{\cdot i}; \qquad \text{put } s := s + 1 \text{ repeat.} \quad \diamond \tag{2.8.1}$$

The simple operations do not require any matrix-multiplication and keep A fixed. The index $i = i(s)$ is not unique, in general. The assigned elements in the simplex X of mixed strategies are given by

$$x^s = \frac{Y^s}{s}\ (s > 0) \text{ and satisfy } Ax^s = \frac{AY^s}{s} = \frac{Z^s - Z^0}{s} \text{ and } x^{s+1} = \frac{sx^s + e^i}{s+1}.$$

Hence, up to the vanishing term $\max_k |Z^0_k|/s$ for $s \to \infty$, the error $\varepsilon(s) = \max_k A_{k\cdot} x^s$ of x^s coincides with $\max Z^s/s$, and $\max Z^s/s < \delta$ is a stopping rule.

Theorem 2.8.1 *[123] It holds* $\lim_{s\to\infty} \max Z^s/s = 0$. ◇

Hence every accumulation point of the bounded sequence $\{x^s\}$ solves the game. So the statement is also an existence theorem. The requirement $\max Z^0 = 0$ is a skilful device needed for proving the theorem. Clearly, one may start with $Z^0 = 0$ in the algorithm. Then the index $i = i(s)$ is just given by

$$A_{i.}x^s \geq A_{k.}x^s \quad \forall k \in \{1, ..., n\}.$$

For numerical questions and an essential improvement of this algorithm in view of large scale problems (games and linear programs), see [89]. A proof of the theorem can be also found in the script on games and fixed points.

Chapter 3

Elements of convex analysis

Convex Analysis bildet den Kern der Theorie konvexer und das Fundament nichtglatter Optimierung. Alle Definitionen und die meisten der folgenden Aussagen findet man, für den endlich-dimensionalen Fall, in der grundlegenden Monographie von R.T. Rockafellar [131]. Natürlich standen dafür die Trennungsaussagen bzgl. konvexer Mengen schon zur Verfügung. Für ein erstes Durchlesen stelle man sich alle Mengen in $X = \mathbb{R}^n$ vor, auch wenn viele Konstruktionen im linearen normierten Raum X analog erfolgen können.

3.1 Separation of convex sets

Sei $M \subset X$. M heisst *konvex*, wenn

$$\lambda x + (1-\lambda)y \in M \quad \text{if } x, y \in M \text{ and } 0 < \lambda < 1.$$

Wenn $x, y \in M$, so liegt auch die Strecke zwischen x und y in M. Man zeigt leicht: Der Durchschnitt (beliebig vieler) konvexer Mengen ist wieder konvex; ebenso die Abschliessung und das Innere einer konvexen Menge (UebAufg.). Ausserdem folgt für konvexes M (Induktion über m)

$$z := \sum_{i=1}^{m} \lambda_i \, x^i \in M \quad \text{sofern } x^i \in M, \ \lambda_i \geq 0 \text{ und } \sum_{i=1}^{m} \lambda_i = 1. \tag{3.1.1}$$

M heisst *Kegel*, wenn aus $x \in M$ folgt, dass auch alle Punkte $\lambda x, \ \lambda \geq 0$ (der Strahl durch 0 und x) aus M sind.

Der für alle $M \subset X$ konvexe und abgeschlossene Kegel

$$M^* = \{x^* \in X^* \mid \langle x^*, x \rangle \leq 0 \ \forall x \in M\}$$

heisst *Polarkegel von* M. Er spielt eine zentrale Rolle in der Optimerung.

For $A \subset X$ bezeichnet $\operatorname{conv} A$ die *konvexe Hülle* von A. Sie ist die kleinste konvexe Menge M mit $A \subset M$. Analog definiert man die konische Hülle con A. Damit gilt auch

$$z \in \operatorname{conv} A \quad \Leftrightarrow \quad \exists \, m, \ \lambda_i \geq 0, \ a^i \in A \ \text{ mit } \ \sum_{i=1}^{m} \lambda_i = 1 \text{ und } \ z = \sum_{i=1}^{m} \lambda_i a^i. \tag{3.1.2}$$

Ist $A = \{a^1, a^2, ..., a^m\}$ und folgt aus ($\sum_{i=1}^m r_i a^i = 0$ and $\sum_{i=1}^m r_i = 0$), dass $r_i = 0 \ \forall i$, so heissen die m Punkte a^i (bzw. das System dieser Punkte) affin unabhänging. Äquivalent: lineare Unabhängigk. der m Vektoren $(a^1, 1), ..., (a^m, 1)$ bzw. der $m-1$ Vektoren $a^2 - a^1, ..., a^m - a^1$. (Nachrechnen !). In diesem Fall heisst $\operatorname{conv} A$ auch *Simplex* (der Dimension $m-1$). Für jedes $z \in \operatorname{conv} A$ sind dann die Koeffizienten λ_i aus (3.1.2) eindeutig bestimmt; gäbe es weitere, etwa λ_i', folgte mit $r_i = \lambda_i' - \lambda_i$ ein Widerspruch.

Theorem 3.1.1 *(Caratheodory) Für alle $A \subset \mathbb{R}^n$ gilt (3.1.2) mit $m = n + 1$.* $\diamond$

Proof. Man schreibe $z \in \operatorname{conv} A$ mit minimaler Anzahl von Elementen $a^i \in A,\ i = 1, ..., m$ als Konvexkombination:

$$z = \sum_{i=1}^{m} \lambda_i a^i \quad \text{mit } \lambda_i > 0,\ \sum_{i=1}^{m} \lambda_i = 1. \tag{3.1.3}$$

Wenn $m > n + 1$, so sind die m Vektoren $(a^i, 1) \in \mathbb{R}^{n+1},\ (i \geq 1)$ linear abhängig; also existieren nichttriviale $r_i,\ (i \geq 1)$ mit $0 = \sum_{i=1}^{m} r_i a^i = 0$ und $\sum_{i=1}^{m} r_i = 0$. Mit beliebigem $t \in \mathbb{R}$ gilt daher auch

$$z = \sum_{i=1}^{m} (\lambda_i + t r_i) a^i \quad \text{and} \quad \sum_{i=1}^{m} (\lambda_i + t r_i) = 1.$$

O.B.d.A. sei $r_1 \neq 0$. Ist $r_1 < 0$, dann gibt es ein positives t, sodass $(\lambda_i + t r_i) \geq 0 \forall i$ und $\lambda_i + t r_i = 0$ für wenigstens ein i gilt (Man erhoehe t bis die erste Funktion $t \mapsto \lambda_i + t r_i$ Null wird). Ist $r_1 > 0$, findet man analog ein negatives t. Die Darstellung (3.1.3) war also nicht minimal. $\square$

Damit ist $\operatorname{conv} A$ kompakt, wenn $A \subset \mathbb{R}^n$ kompakt ist (wieso ?).

Theorem 3.1.2 *(Projektionssatz) Sei $\emptyset \neq M \subset \mathbb{R}^n$ konvex und abgeschlossen und $\pi_M(x)$ die Projektion von $x \in \mathbb{R}^n$ auf M, d.h. $\pi_M(x) \in M$ erfülle (mit Euklidischer Norm)*

$$\|\pi_M(x) - x\| \leq \|y - x\| \quad \forall y \in M. \tag{3.1.4}$$

(i) Dann ist $p = \pi_M(x)$ eindeutig bestimmt und erfüllt

$$(p - x)^T (y - p) \geq 0 \quad \forall y \in M. \tag{3.1.5}$$

(ii) Umgekehrt, erfüllt $p \in M$ Bedingung (3.1.5), so ist $\pi(x) = p$ Lösung von (3.1.4).

(iii) Die Projektion π_M ist nicht-expansiv, d.h. $\|\pi_M(x') - \pi_M(x)\| \leq \|x' - x\|$. Dabei gilt für $\pi_M(x') \neq \pi_M(x)$ die Gleichung, genau dann wenn $\pi_M(x') - \pi_M(x) = x' - x$. $\diamond$

Proof. Ein Minimalpunkt $p = \pi_M(x)$ für (3.1.4) existiert, weil die Norm stetig ist und es - mit irgendeinem $y^0 \in M$ - ausreicht, die Norm über der kompakten Menge $M^0 = M \cap \{y \mid \|y - x\| \leq \|y^0 - x\|\}$ zu minimieren.
(i) Mit beliebigem $y \in M$ setze man $\quad z(t) = p + t(y - p), \quad 0 < t < 1.$
Dann folgt $z(t) = ty + (1 - t)p \ \in M$ (Konvexität) und

$$\|z(t) - x\|^2 = \|p - x + t(y - p))\|^2 \ = \|p - x\|^2 + 2t(p - x)^T (y - p) \ + t^2 \|y - p\|^2.$$

Wegen $\|z(t) - x\|^2 \geq \|p - x\|^2$, muss deshalb (3.1.5) gelten (sonst entsteht mit kleinen $t > 0$ ein Widerspruch da $2t\ (p - x)^T (y - p) \ + t^2\ \|y - p\|^2 < 0$).
(ii) Erfüllen zwei Punkte $p, p' \in M$ die Bedingung (3.1.5), erhalten wir

$$(p - x)^T (p' - p) \geq 0 \ \text{ and } \ (p' - x)^T (p - p') \geq 0$$

und nach Addition $(p - p')^T (p' - p) = -\|p' - p\|^2 \geq 0,$ also $p = p'$. Damit gilt (3.1.5) nur für $p = \pi_M(x)$, und $\pi_M(x)$ aus (3.1.4) muss eindeutig sein.
(iii) Man wende (3.1.5) auf beide Projektionen an:

$$\langle \pi_M(x) - x,\ \pi_M(x') - \pi_M(x) \rangle \geq 0 \qquad , \quad \langle \pi_M(x') - x',\ \pi_M(x) - \pi_M(x') \rangle \geq 0.$$

Addition und Ausmultipliz. liefert

$$\begin{array}{ll} \langle\ \pi_M(x) - x\ + x' - \pi_M(x')\ ,\ \ \pi_M(x') - \pi_M(x)\rangle & \geq 0, \\ \langle\ x' - x,\ \pi_M(x') - \pi_M(x)\rangle - \|\pi_M(x') - \pi_M(x)\|^2 & \geq 0. \end{array}$$

Damit gilt auch

$$\|\pi_M(x') - \pi_M(x)\|^2 \leq \langle\ \pi_M(x') - \pi_M(x),\ x' - x\rangle \leq \|x - x'\|\ \|\pi_M(x') - \pi_M(x)\|. \quad (3.1.6)$$

Division durch $\|\pi_M(x') - \pi_M(x)\| > 0$ liefert Nicht-Expansivität. Ausserdem: Die rechte Ungleichung gilt mit $\pi_M(x') \neq \pi_M(x)$ als Gleichung $\Leftrightarrow\ \pi_M(x') - \pi_M(x) = x' - x$. □

Remark 3.1.3 (Proj. in Hilbert spaces:) Thm. 3.1.2 bleibt im Hilbert Raum (statt $\mathbb{R}^n$) richtig. ◇

Beweis: Man hat nur die Existenz von $\pi_M(x)$ anders zu zeigen, z.B. mittels Parallellogrammgleichung

$$\|a + b\|^2 + \|a - b\|^2 = 2(\|a\|^2 + \|b\|^2), \quad (3.1.7)$$

die durch direktes Ausrechnen folgt.

Sei nun $d = \operatorname{dist}(x, M)$ und $\{z_n\}$ eine Folge in M mit $\|x - z_n\|^2 < d^2 + n^{-2}$. Man zeigt Konvergenz der $\{z_n\}$ durch Abschätzen von $\|z_m - z_n\|^2$. Dazu wird (3.1.7) auf

$$a = x - z_n, \qquad b = x - z_m$$

angewandt:

$$\|2x - (z_m + z_n)\|^2 + \|z_m - z_n\|^2 = 2\ (\|x - z_n\|^2 + \|x - z_m\|^2).$$

Das erste Quadrat ist gerade 4 $\|x - \frac{z_m + z_n}{2}\|^2$. Ausserdem ist $\frac{z_m + z_n}{2} \in M$. Damit gilt $4\ \|x - \frac{z_m + z_n}{2}\|^2 \geq 4\ d^2$, und folglich

$$\begin{array}{l} \|z_m - z_n\|^2 \quad \leq \quad 2\ (\ \|x - z_n\|^2 + \|x - z_m\|^2\ ) - 4d^2 \\ \leq\ 2\ (d^2 + n^{-2} + d^2 + m^{-2}) - 4d^2\ =\ 2(\ n^{-2} + m^{-2}). \end{array}$$

Die Folge z_n ist also eine Cauchy- Folge und $z = \lim z_n$ ist der gesuchte Punkt. □

Lemma 3.1.4 *Trennungssatz 1: Sei $\emptyset \neq M \subset \mathbb{R}^n$ konvex, abgeschlossen und $x \notin M$. Dann existiert ein $u \in \mathbb{R}^n$ mit $u^T x < u^T y \quad \forall y \in M$.* ◇

Proof. Es sei p die Projektion von x auf M (Projektionssatz) und $u = p - x$. Dann ist

$$u^T(y - x) = u^T(y - p + p - x) = u^T(y - p + u) = u^T(y - p) + \|u\|^2 > 0\ \forall y \in M. \quad \square$$

Analog kann man im Hilbert Raum schliessen.
UeAufg. Zeigen Sie mit Lemma 3.1.4,

Theorem 3.1.5 *(Bipolarsatz) Für jeden abgeschlossenen konvexen Kegel $\emptyset \neq K \subset \mathbb{R}^n$ und dessen Polarkegel K^* gilt $(K^*)^* = K$.* ◇

Proof. ($\subset$): Sei $u \in (K^*)^*$. Das heisst

$$\langle u, x^*\rangle \leq 0 \quad \forall x^* \in K^*. \quad (3.1.8)$$

Wenn $u \notin K$, kann man trennen: $\exists y^* \neq 0 : \langle y^*, k\rangle < \langle y^*, u\rangle \quad \forall k \in K$ (abgeschl. konvex). Weil K ein Kegel ist, folgt damit $\langle y^*, k\rangle \leq 0 < \langle y^*, u\rangle\ \forall k \in K$. Also ist $y^* \in K^*$ ein Element, das (3.1.8) verletzt.
($\supset$): $u \in K$. Angenommen $u \notin (K^*)^*$. Dann gibt es - da K^* konvex und abgeschlossen ist - ein $x^* \in K^*$ mit $\langle u, x^*\rangle > 0$. Weil $x^* \in K^*$, kann daher nur $u \notin K$ gelten. □

Lemma 3.1.6 *Trennungssatz 2: Es sei $M \subset \mathbb{R}^n$ konvex, abgeschlossen, nichtleer und $x \notin \operatorname{int} M$. Dann gibt es ein $u \in \mathbb{R}^n,\ u \neq 0$, so dass $u^T x \leq u^T y \quad \forall y \in M$ gilt.* ◇

Proof. Da $x \notin \operatorname{int} M$, gibt es Punkte $x_k \to x\ (k \to \infty)$, so dass $x_k \notin M$. Auf diese lässt sich Lemma 3.1.4 anwenden. Sei u_k ein entsprechender Vektor. O.B.d.A. gelte $\|u_k\| = 1$. (Division durch die Norm). Für eine (unendl.) Teilfolge konvergieren die beschränkten u_k gegen ein u mit $\|u\| = 1$. Mit jedem $y \in M$ folgt so wegen $u_k^T x_k \leq u_k^T y$ und Stetigkeit des Skalarprodukts auch $u^T x \leq u^T y$. □

Remark 3.1.7 Ist $\emptyset \neq M \subset \mathbb{R}^n$ nur konvex, beweist man Lemma 3.1.6 analog, indem man x von der Abschliessung $\operatorname{cl} M$ trennt, deren Inneres x ebenfalls nicht enthalten kann. ◇

! Dieser Schluss ist für konvexe Mengen in B-Räumen falsch. Wieso ? Und wie zeigt man $x \notin \operatorname{int}\operatorname{cl} M$, wenn $\emptyset \neq M \subset \mathbb{R}^n$ konvex ist und $x \notin \operatorname{int} M$?

Theorem 3.1.8 *(Trennung allgemein) Seien A, B konvexe Teilmengen eines linearen normierten Raumes X (über $\mathbb{R}$), A und $\operatorname{int} B$ nichtleer, und $A \cap \operatorname{int} B = \emptyset$. Dann gibt es ein $x^* \in X^* \setminus \{0\}$ mit $\langle x^*, a\rangle \leq \langle x^*, b\rangle \quad \forall\ a \in A,\ b \in B$.* ◇

Proof. Man bilde die (konvexe) Menge $M = \{x \mid x = b - a,\ a \in A,\ b \in B\}$. Dann folgt $0 \notin \operatorname{int} M \neq \emptyset$. Trennung von 0 und M liefert die Behauptung.

Für $X = \mathbb{R}^n$ kann man dazu Lemma 3.1.6, Remark 3.1.7 benutzen. Im normierten Raum benötigt der Beweis der Trennbarkeit von 0 und M das Zornsche Lemma oder eine dazu äquivalente Aussage, s. Thm. 5.1.1. □

From the Theorem, the subsequent version of a Hahn-Banach Theorem [6] can be derived. Notice that p is *sublinear* iff p is convex and positively homogeneous.

Corollary 3.1.9 *Let $p : X \to \mathbb{R}$ be sublinear and continuous, U be a subspace of X (not necessarily closed) and $l : U \to \mathbb{R}$ be additive and homogeneous with $l(u) \leq p(u)\ \forall u \in U$. Then: $\exists\ x^* \in X^*$ such that $\langle x^*, x\rangle \leq p(x)\ \forall x \in X$ and $x^* \equiv l$ on U.* ◇

Proof. Separation of $A = \{\ (u,\ l(u)) \mid u \in U\}$ and $B = \operatorname{epi} p$ can be applied since $\operatorname{int} B \neq \emptyset$ by continuity of p, and $A \cap \operatorname{int} B = \emptyset$. □

3.2 Brouwer's and Kakutani's Fixpunktsatz

Die folgenden Fixpunktsätze sind fundamental für viele Existenzsätze in Optimierung, Spieltheorie und mathematischer Ökonomie und wurden in vieler Hinsicht verallgemeinert. Wir betrachten zuerst die klassischen Aussagen. *Die beiden folgenden Skizzen wurden im Mathe-Lager "Lust auf Mathematik", Blossin 2006, erstellt von Thomas Bünger, Erik Esche, Felix Günther, Arne Müller und Andrej Stepanchuk.*

3.2.1 The theorems in $\mathbb{R}^n$

Theorem 3.2.1 *(L.E.J. Brouwer's Fixpunktsatz) Let $\emptyset \neq S \subset \mathbb{R}^n$ be convex and compact and $f : S \to S$ be continuous. Then: $\exists\ \bar{x} \in S$ such that $f(\bar{x}) = \bar{x}$.* ◇

Proof. Man beweist den Satz zuerst für ein Simplex $S = \operatorname{conv}\{P_1, ..., P_n\},\ P_i \in \mathbb{R}^n$

$$S = \{x \in \mathbb{R}^n \mid x = \sum \alpha_i P_i,\ \sum \alpha_i = 1, \alpha_i \geq 0\}.$$

Man beachte, dass die Koeffizienten α_i eindeutig sind und glm. stetig von $x \in S$ abhängen.
Man definiere eine *Simplexunterteilung* induktiv:
Für ein Intervall kanonisch durch Halbierung.
1. einfache Unterteilung: über die konvexe Hülle einer einfach unterteilten Seite und dem Schwerpunkt des Simplex (siehe Bildchen).
2. mehrfache Unterteilung: Wiederholung der Prozedur auf Teilsimplizes.
Damit ist klar, was eine N-fache Unterteilung sein soll. Knoten der Unterteilung seien die Ecken der Teilsimplizes. Der Durchmesser der Teilsimplizes (= max Abstand zweier Punkte daraus) geht mit der Ordnung N der Unterteilung gegen Null (Induktion).

Indexfunktion: Eine Funktion $j = j(x) \in \{1, ..., n\}$ heisst zulässige Indexfunktion, sofern: $j(x) \in \{\pi 1, ..., \pi k\}$ if $x \in conv\{P_{\pi 1}, ...P_{\pi k}\}$ $(x \in S)$. Zunächst beweisen wir

Lemma 3.2.2 *(Sperner's Lemma) Sei j(.) eine, auf den Knoten (Ecken) einer N-fachen Unterteilung definierte, zulässige Indexfunktion. Dann gibt es ein Teilsimplex T der Unterteilung, so dass den Ecken $p_1, ...p_n$ von T alle Nummern $1, ..., n$ zugeordnet sind.* $\diamond$

Proof. (Lemma) Betrachte Teilsimplizes $T(k)$ und definiere: Eine ausgezeichnete Seite ist eine Seite eines Teilsimplex' $T(k)$ mit zugeordneten j-Werten $1, ..., n-1$. "Normales" $T(k)$: zugeordnete j-Werte sind $1, ..., n$. Sperner zeigte per Induktion:

Die Anzahl normaler Teilsimplizes ist ungerade.

Für $\dim S = 0, d.h., S = \text{conv}\,\{P_1\}$ ist die Aussage trivial. Sie sei für $S = \text{conv}\,\{P_1, ..., P_{n-1}\}$ bereits bewiesen.
Idee: Sehen und gesehen werden !
1) sehen: Wir laufen durch alle $T(k)$ und zählen die dort zu sehenden ausgezeichneten Seiten σ. Das seien $t(k)$ Stück. Sei

$$a = \sum t(k).$$

Wenn $T(k)$ normal, ist offenbar $t(k) = 1$. Sonst: $t(k) = 0$ oder $t(k) = 2$. Also gilt

a ist gerade $\Leftrightarrow$ die Anzahl normaler $T(k)$ ist gerade.

2) gesehen werden:
Wir fragen nun, wie oft eine ausgezeichnete Seite σ gezählt wurde.

Fall 1: σ liegt in der Original- Simplexseite der Knoten $P_1, ..., P_{n-1}$: Genau einmal.

Fall 2: σ liegt nicht in dieser Simplexseite: Dann liegt sie wegen der Zulässigk. Bedingung auch in keiner anderen Seite des Originalsimplex S. Sie wird deshalb von genau zwei Teilsimplizes $T(k), T(k')$ aus gesehen (die an der Seite σ gespiegelt sind). Bezeichnet b_1 und b_2 die Anzahl der ausgezeichneten Seiten zu beiden Fällen, folgt also

$$a = b_1 + 2b_2.$$

Damit ist a gerade $\Leftrightarrow b_1$ gerade.

Nun ist b_1 zugleich die Anzahl normaler Teilsimplizes in einem Simplex kleinerer Dimension. Daher ist b_1 ungerade, folglich auch a, was zu zeigen war. $\square$

Zum Beweis des FP-Satzes wähle man eine spezielle zulässige Indexfunktion.

Da $f(x) \in S$, gibt es eindeutige und stetige $\lambda_i = \lambda_i(x) \geq 0, \sum \lambda_i = 1$, so dass

$$f(x) = \sum \lambda_i \; P_i.$$

Dann sieht man leicht: x ist Fixpunkt von $f \Leftrightarrow \; \alpha_i \geq \lambda_i \; \forall i$.
Ausserdem gilt:
Wenn $x \in \text{conv}\,\{P_1, ...P_k\}$, so gibt es ein $i \in \{1, ..., k\}$ mit $\alpha_i \geq \lambda_i$, denn sonst wäre die

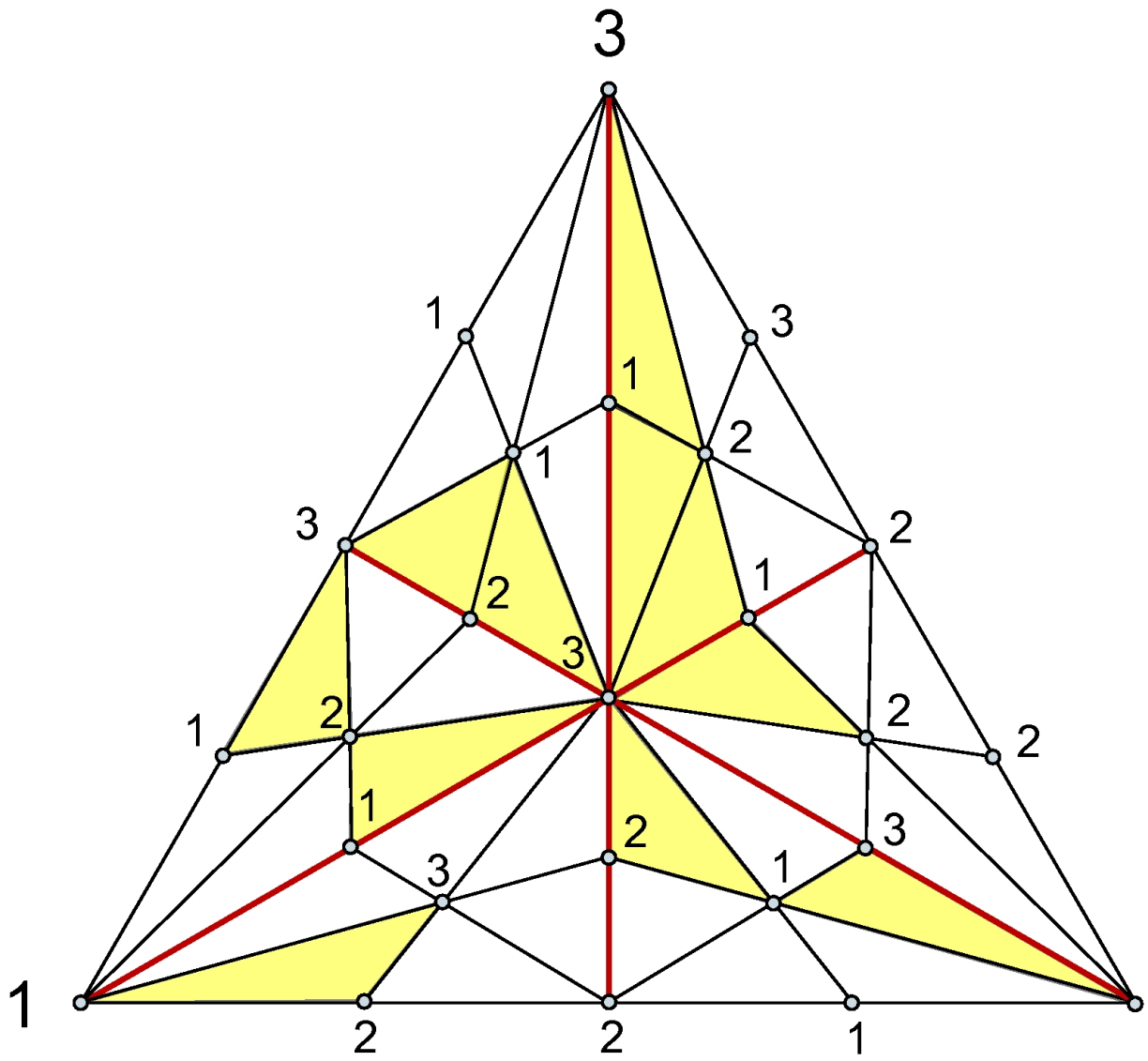


Figure 3.1: Unterteilung des Simplex und 11 normale Teilsimplizes

Summenbedingung für die λ_i verletzt.
Also kann man jedem $x \in \text{conv}\{P_1, ...P_k\}$ ein $i = i(x) \in \{1, ..., k\}$ mit $\alpha_i \geq \lambda_i$ zuordnen. Das gilt analog für jede Auswahl $\{P_{\pi 1}, ..., P_{\pi k}\}$ von k anderen Ecken.

Die Zuordnung $i = i(x)$ ist deshalb eine spezielle zulässige Indexfunktion j. Nach dem Lemma gibt es zu jeder N-fachen Unterteilung von S ein normales Teilsimplex $T(k)$. Für beliebiges $x(N) \in T(k)$ folgt dann, weil $\lambda = \lambda(x), \alpha = \alpha(x)$ gleichmässig stetig sind, dass mit gewissen $\varepsilon(N) \to 0$ (bei $N \to \infty$) gilt

$$\alpha_i \geq \lambda_i - \varepsilon(N) \ \forall i,$$

denn die Ungleichung gilt (ohne ε) für jeweils eine Ecke x von $T(k)$ und

$$\|\lambda(x) - \lambda(x(N))\| + \|\alpha(x) - \alpha(x(N))\|$$

strebt für $N \to \infty$ gleichmässig gegen Null, weil die Teilsimplizes beliebig klein werden.

Nimmt man nun einen Häuf.-Punkt $\bar{x}$ der $x(N)$, für $N \to \infty$, erfüllt er damit

$$\alpha_i(\bar{x}) \geq \lambda_i(\bar{x}) \ \forall i.$$

Also ist der Brouwersche FP Satz für das Simplex S richtig. Die Erweiterung auf bel. konvexe, kompakte Mengen C wird nun (fast) eine UeA. Man wähle ein Simplex S, das C enthält. Sei $p(x)$ die Projektion von x auf C. Sie ist stetig. Durch $g(x) = f(p(x))$ für $x \in S$ lässt sich f stetig auf S erweitern. Da g und f dieselben Fixpunkte besitzen, folgt die Behauptung. □

Remark 3.2.3 Mit einer anderen Indexfunktion kann ein normales Teilsimplex *konstruktiv* gefunden werden (Methode von Scarf [135] und Hoang Tuy). Die Methode wird mit Fig. 2 verdeutlicht. Knotennummer i bedeutet hier die umgekehrte Ungleichung $\alpha_i \leq \lambda_i$. Punkte im "Innern" der Seite, die p_i gegenüber liegt, erhalten die Nummer i. ◇

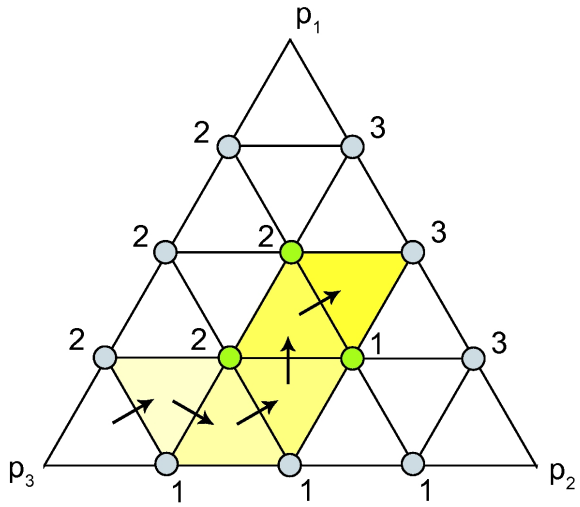


Figure 3.2: Weg durch das Simplex zu einem normalen Teilsimplex

Recall that $F : S \rightrightarrows S$ is closed iff so is $\operatorname{gph} F := \{(x, y) \mid x \in S, y \in F(x)\}$.

Theorem 3.2.4 *(Kakutani's [58] FP Theorem) Let $S \subset \mathbb{R}^n$ be convex and compact, let $F : S \rightrightarrows S$ be a closed mapping such that all $F(x)$ are non-empty convex subsets of S. Then: $\exists\, \bar{x} \in S$ such that $\bar{x} \in F(\bar{x})$.* ◇

Proof. Diesmal nehmen wir an, $S = B$ sei die Euklidische Einheitskugel des $\mathbb{R}^n$, und es gelte $x \notin F(x) \forall x \in S$. Stets sei $x \in S$. Dann gibt es (Trennung), zu jedem x ein $z \in \mathbb{R}^n$ als Trennungsvektor mit

$$t(z, x) := \inf_{y \in F(x)} \langle z, y - x \rangle > 0.$$

Sei $W(x)$ die Menge dieser z und $W^{-1}(z) = \{x \mid t(z, x) > 0\}$. Weil $\operatorname{gph} F$ abgeschlossen ist, folgert man leicht, dass $W^{-1}(z)$ (relativ) offen in S ist. Da nun $S \subset \cup_z W^{-1}(z)$, gibt es endlich viele (Überdeckungssatz) Elemente z_p $(p = 1, ..., N)$ mit

$$S \subset \cup_p W^{-1}(z_p) \tag{3.2.1}$$

Sei $A_p = S \setminus W^{-1}(z_p)$ and $d_p(x) = \operatorname{dist}(x, A_p)$. Wegen (3.2.1) und Abgeschlossenheit der A_p gilt $d(x) := \sum_p d_p(x) > 0\ \forall x$. Damit kann man stetige Funktionen

$$\lambda_p(x) = \frac{d_p(x)}{d(x)} \quad \text{und} \quad z(x) = \sum_p \lambda_p(x)\ z_p \quad \text{definieren.}$$

Bemerkung: Stetige Funktionen λ_p, die in dieser (und allgemeinerer) Weise einer offenen Überdeckung (wie hier (3.2.1)) zugeordnet sind, heissen Zerlegung der 1 (partition of unity), wenn

$$\sum_p \lambda_p(x) = 1 \text{ gilt und ausserdem } \lambda_p(x) > 0 \;\Leftrightarrow\; x \in W^{-1}(z_p).$$

Um $z(x) \in W(x)\ \forall x \in S$ zu zeigen, sei $y \in F(x)$. Es gibt ein positives $\lambda_p(x)$, und aus $\lambda_p(x) > 0$ folgt $x \in W^{-1}(z_p)$ und $\lambda_p(x)\langle z_p,\ y - x\rangle > 0$. Also ist auch

$$\langle z(x), y - x\rangle = \sum_p \lambda_p(x)\langle z_p,\ y - x\rangle > 0.$$

Da $y \in F(x)$ beliebig war, folgt so $z(x) \in W(x)$. Dann ist auch $z(x) \neq 0$ richtig. Die Funktion $f(x) = \frac{z(x)}{\|z(x)\|}$ bildet daher die Kugel $S = B$ stetig auf den Rand $\operatorname{bd} B$ ab und erfüllt ebenfalls $f(x) \in W(x)$. Nach Brouwer besitzt sie einen Fixpunkt: $\bar{x} = f(\bar{x}) \in W(\bar{x})$. Er erfüllt

$$0 < \inf_{y \in F(\bar{x})} \langle \bar{x}, y - \bar{x}\rangle \text{ und } \|\bar{x}\| = 1.$$

Zusammen liefert das $F(\bar{x}) \cap S = \emptyset$ im Gegensatz zu $\emptyset \neq F(\bar{x}) \subset S$. Das beweist den Satz für die Einheitskugel. Damit gilt er für jede Kugel $S = B(0, r)$ (wieso ?).

Ist schlielich $S = C$ konvex und kompakt, wählen wir ein r mit $C \subset B(0, r)$ und betrachten $G(x) = F(p(x)),\ \ x \in B(0, r)$ mit $p(x) =$ Proj. von x auf C. □

Remark 3.2.5 Brouwer's FP Theorem is the single-valued version of Kakutani's Theorem since continuity of functions $F : S \to S$ (compact) follows from closeness. ◇

3.2.2 The theorems in normed spaces

If $S \subset \mathbb{R}^n$ is replaced by $S \subset X$ (a linear normed space) in Kakutani's theorem, the statement remains valid.

Proof: Let $\varepsilon > 0$. Since S is compact, there exists a finite set $M_\varepsilon \subset S$ such that $\forall x \in S\ \exists y \in M_\varepsilon : \|y - x\| \leq \varepsilon$ (a finite ε-net in S). The set $C = \operatorname{conv} M_\varepsilon$ belongs to a finite-dimensional subspace U of X (generated by the elements of M_ε) and is compact in U (with induced topology). To every $u \in C$, assign the projection $\Gamma(u)$ of $F(u)$ onto C. Then we obtain $\Gamma(u) \subset F(u) + \varepsilon B$ and, due to convexity of $F(u)$, also $\operatorname{conv}\Gamma(u) \subset F(u) + \varepsilon B$. Since F is closed, so are Γ and $\operatorname{conv}\Gamma$ on C (see Thm 3.1.1). Applying Katutani's theorem to $\operatorname{conv}\Gamma$ on C, one finds some $y_\varepsilon \in \operatorname{conv}\Gamma(y_\varepsilon),\ y_\varepsilon \in C$. This entails $\operatorname{dist}(y_\varepsilon, F(y_\varepsilon)) \leq \varepsilon$. Considering any $\varepsilon_k \to 0$ and selecting an accumulation point $\bar{x}$ of $y_{\varepsilon_k} \subset S$, the assertion now follows by arguments of closeness and compactness. □

Brouwer's theorem still holds by the same reasonings. However, we have again to apply Kakutani's theorem for proving it, since the projection $\Gamma(u)$ of $F(u)$ onto C is not necessarily single-valued if so is $F(u)$.

3.3 Nash Equilibria and Minimax Theorem

We consider a non-cooperative N-person game with "strategy vectors"

$$x = (x_1, ..., x_p, ...x_N); \qquad x_p \in \mathbb{R}^{n_p},$$

payoff functions $f_p(x)$ for player p who chooses x_p and want to maximize f_p under some constraints. Let $(x||\xi_p) = (x_1, ..., \xi_p, ...x_N)$.

The next statements follow from Kakutani's theorem.

3.3.1 The basic Existence Theorem

Suppose that $x_p \in X_p \subset \mathbb{R}^{n_p}$ is required. Then, a (Nash-) equilibrium x is defined by

$$x_p \in X_p \quad \text{and} \quad f_p(x) \geq f_p(x||\xi_p) \quad \forall \xi_p \in X_p; \quad \text{for } p = 1, ..., N. \tag{3.3.1}$$

Theorem 3.3.1 *(Nash) The model (3.3.1) has a solution if all $X_p \neq \emptyset$ are convex and compact, all f_p are continuous and, in addition, concave in the players variable x_p.* $\diamond$

Proof. Define the sets of maximizers ξ_p, depending on x,

$$F_p(x) = \text{argmax } \{ \ f_p(x||\xi_p) \ \mid \ \xi_p \in X_p \ \}.$$

By our assumptions, $F_p(x) \neq \emptyset$ is compact and convex, and the multifunctions $x \mapsto F_p(x)$ are closed. This remains true for the product mapping

$$F(x) := F_1(x) \times ... \times F_N(x) \subset M := X_1 \times ... \times X_N. \tag{3.3.2}$$

Now $F : M \rightrightarrows M$ satisfies the hypotheses of Kakutani's Theorem 3.2.4. The fixed point x fulfills obviously (3.3.1). $\square$

Minimax: Setting particularly $N = 2$, $f_2 = -f_1$ and $\phi = f_1$, ϕ is a continuous concave-convex function and Thm. 3.3.1 attains (with $X_i \neq \emptyset$ convex, compact) the form

Corollary 3.3.2 *(Minimax-Theorem) There is some $(u, v) \in X_1 \times X_2$ such that*

$$\phi(x_1, v) \leq \phi(u, v) \leq \phi(u, x_2) \quad \forall x_1 \in X_1, x_2 \in X_2. \qquad \diamond$$

A typical convex-concave function is the Langrangian $L(x, y)$, Def. 4.1.1, in convex optimization. One can establish the duality theory of convex optimization by using the Minimax Theorem, though the sets X_i are there not compact a priory.

3.3.2 Normalized Nash Equilibria

Now suppose there exist joint constraints $x \in M$ for the strategies

$$M = \{x \mid g_i(x) \leq 0, \ i = 1, ..., m\}. \tag{3.3.3}$$

Interpretation: If like above, only $x_p \in X_p$ is required, we can write, e.g., $g_p(x_p) := \text{dist}\,(x_p, X_p) \leq 0$. Clearly, g_p is convex if so is X_p. If nation p likes to fish out x_p tons of cod in the North Sea, it makes sense to claim (among others) that $g_1(x) := \sum x_p - C \leq 0$. Now, a (more general) equilibrium x is defined by the requirements

$$x \in M \ \text{ and } \ f_p(x) \geq f_p(x||\xi_p) \quad \text{whenever } (x||\xi_p) \in M; \quad \text{for } p = 1, ..., N. \tag{3.3.4}$$

Theorem 3.3.3 *The model (3.3.4) has a solution if $M \neq \emptyset$ is convex and compact, and -as above - all f_p are continuous and, in addition, concave in the players variable x_p.* $\diamond$

Proof. For $x \in M$, define the function $\phi(\xi, x) = \sum_p \ f_p(x||\xi_p)$ which is concave in ξ, and put (as proposed by Rosen [133])

$$F(x) = \text{argmax } \{ \ \phi(\xi, x) \ \mid \ \xi \in M \ \}. \tag{3.3.5}$$

Again, $F : M \rightrightarrows M$ satisfies the assumptions of Kakutani's Theorem. So F has a fixed point, and $x \in F(x)$ implies (3.3.4), now. $\square$

The fixed points of (3.3.5) are particular equilibria, also called normalized equilibria. They are closely connected with solutions of optimization problems, and many statements of optimization remain true (with the same proofs) for them.

To see this, let $f, g \in C^1$ and suppose a condition, called constraint qualification CQ below. Then, since x solves the optimization problem (3.3.5), we may use that x fulfills the related necessary KKT-conditions, cf. Thm. 4.1.2: There is some $y \in \mathbb{R}^m$ such that

$$\begin{array}{c} \bar{f}(x) + \sum_i y_i Dg_i(x) = 0, \quad y \geq 0, \quad \langle y, g(x)\rangle = 0 \\ \text{where} \\ \bar{f} = -(\frac{\partial f_1}{\partial x_1}, ..., \frac{\partial f_N}{\partial x_N}) \ \text{ replaces } Df \text{ in (usual) optimization models.} \end{array} \tag{3.3.6}$$

So we obtain *variational inequalities* where Df is simply replaced by some $\bar{f}$ of related dimension and smoothness. To study stability of such solutions, now second derivatives come into the play, and

$$D\bar{f}(x) + \sum_i \ y_i \ D^2 g_i(x) \quad \text{ stands for } \quad D^2 L_x(x,y) \tag{3.3.7}$$

in the statements on stability of critical points in optimization, chapter 12. However, the related sets of all (or normalized) equilibria have, generally, a complicated structure [4] even under the made assumptions and $N = 2$.

Example 3.3.1 For $X_1 = X_2 = [0,1]$ and $f_1 = f_2 = -(x_1 - x_2)^2$, all feasible points of the diagonal $x_2 = x_1$ are Nash-equilibria.

3.4 Classical subdifferentials, normals, conjugation

Here, we derive key statements of convex analysis which can be summarized as follows, provided that two basic notions are defined:
1. The subdifferential ∂f of a (non-differentiable) convex function, (a generalization of the derivative) and
2. a normal cone $N_M(x)$ of a closed convex set at some $x \in M$ (which generalizes normal directions of smooth manifolds).

Then one may prove, for closed convex sets $\emptyset \neq M \subset \mathbb{R}^n$ and convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{array}{ll} x^* \in \partial f(x) & \Leftrightarrow \ x \in \operatorname{argmin} f(.) - \langle x^*, . \ \rangle \ \Leftrightarrow \ (x^*, -1) \in N_{\operatorname{epi} f}(x, f(x)) \\ 0 \ \in \partial f(x) + N_M(x) & \Leftrightarrow \ x \in \operatorname{argmin}_{x \in M} f(x). \end{array}$$

In addition, the statements $\partial f(x) \neq \emptyset, \quad \partial(f+g)(x) = \partial f(x) + \partial g(x)$ simplify the calculus with these objects. By the help of the *indicator function*

$$i_M(x) = \left\{ \begin{array}{cc} 0 & \text{if } \ x \in M \\ \infty & otherwise \end{array} \right. \tag{3.4.1}$$

the basic notions are connected each other by $N_M(x) = \partial i_M(x)$. By a cone of tangents $C_M(x)$, another connection can be formulated:

$$N_M(x) = C_M(x)^* \ \text{ (the polar cone of } C_M(x)).$$

All applications depend on the ability to determine these objects for concrete M and f.

The mappings N_D and N_{D^*}:

Let $D \subset \mathbb{R}^n$ be a closed convex cone. Then $N_D(y)$ consists of the normals y^* to the cone D at y. They are contained in the polar cone $D^* = N_D(0)$ of D for y. Then we have $N_D(y) \subset D^*$. Moreover, the conditions

$$y^* \in N_D(y) \text{ which means } y \in D \text{ and } \langle y^*, y' - y\rangle \leq 0 \ \forall y' \in D$$

and

$$y \in N_{D^*}(y^*) \text{ which means } y^* \in D^* \text{ and } \langle y, d^* - y^*\rangle \leq 0 \ \forall d^* \in D^*$$

are equivalent.
Proof by contradiction: Assume $y^* \in N_D(y)$. Then, it follows $y^* \in D^*$ i.e., $\langle y^*, D\rangle \leq 0$ since D is a cone, and $\langle y^*, 0-y\rangle \leq 0$ since $0 \in D$ as well as $\langle y^*, y\rangle \leq 0$ since $y \in D$. Hence $\langle y^*, y\rangle = 0$.

On the other hand, if $y \notin N_{D^*}(y^*)$ then, due to $y^* \in D^*$, there is some $p \in D^*$ with $\langle y, p\rangle > \langle y, y^*\rangle$. Because of $y \in D$ and $p \in D^*$, so a contradiction follows: $0 \geq \langle y, p\rangle > \langle y, y^*\rangle = 0$. In consequence, $y^* \in N_D(y)$ implies $y \in N_{D^*}(y^*)$. The reverse direction can be similarly shown. □

3.4.1 Definitions and existence

Given $f : X \to \mathbb{R} \cup \{\pm\infty\}$ (X a linear normed space), put $\operatorname{dom} f = \{x \mid f(x) \in \mathbb{R}\}$.
Def. f is said to be *convex* if the *epigraph* $\operatorname{epi} f := \{(x,t) \in X \times \mathbb{R} \mid f(x) \leq t\}$ is a convex set in $X \times \mathbb{R}$.
Def. f is said to be *proper* if $f(x) \neq -\infty \ \forall x$. If f is proper, convexity of f means that $\operatorname{dom} f$ is convex and

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall\, x, y \in \operatorname{dom} f \text{ and } 0 < \lambda < 1.$$

Def. f is concave if $-f$ is convex, i.e., the *hypograph* $\operatorname{hypo} f := \{(x,t) \mid f(x) \geq t\}$ is convex.

Remark 3.4.1 If $f(x) \in \mathbb{R}$ and $f(x') = -\infty$ then, under convexity, it follows

$$f(x + \lambda(x' - x)) = -\infty \ \forall \lambda \in (0,1) \quad \text{and} \quad f(x + \lambda(x' - x)) = +\infty \ \forall \lambda < 0$$

(consider f on the line through x, x'). Thus $f(x') = -\infty$ is impossible if f is bounded above on some open set. ◇

Lemma 3.4.2 *If f is convex then every local minimizer is also a global one, and the set* $\operatorname{argmin} f$ *of all minimizers is convex (possibly empty).* ◇

Theorem 3.4.3 *(continuity of convex functions) Let $f : X \to \mathbb{R} \cup \{\infty\}$ be convex.*
(i) If $X = \mathbb{R}^n$ then f is continuous on $\operatorname{int} \operatorname{dom} f$.
(ii) f is continuous on open sets Ω with $\sup_\Omega f < \infty$. ◇

Proof. Preparation: Given $y \in \operatorname{int} \operatorname{dom} f$ choose a (small) simplex $S \subset \operatorname{dom} f$ with $y \in \operatorname{int} S$. Using its vertices p_k, one obtains for $x \in S$, via $x = \sum_k \lambda_k p_k$, $\sum_k \lambda_k = 1, \lambda_k \geq 0$, that $f(x) \leq \sum_k \lambda_k f(p_k)$. Hence f is bounded above on S: $f \leq C$ on S. Clearly, then $f \leq C$ also holds on some small closed ball $B_y = B(y, \varepsilon)$. Besides convexity, this is the crucial (and only) property we shall need below.
(i), (ii) Next let $x \to y,\ x \neq y$. Choose $z \in \operatorname{bd} B_y$ such that, with some $\lambda \in (0,1)$,

$$x = \lambda y + (1-\lambda) z, \quad x \text{ between } y \text{ and } z, \quad \lambda = \lambda(x).$$

Then, since $\lambda \to 1$ follows from $x \to y$, convexity and $f(z) \le C$ yield

$$f(x) \le \lambda f(y) + (1-\lambda) f(z) \le \lambda f(y) + (1-\lambda) C \quad \text{and } \limsup_{x \to y} f(x) \le f(y).$$

Now choose $z \in \operatorname{bd} B_y$ such that, with some $\lambda \in (0,1)$,

$$y = \lambda x + (1-\lambda) z, \quad y \text{ between } x \text{ and } z, \quad \lambda = \lambda(x).$$

Since $\lambda \to 1$ follows from $x \to y$, convexity and $f(z) \le C$ yield

$$f(y) \le \lambda f(x) + (1-\lambda) f(z) \le \lambda f(x) + (1-\lambda) C \quad \text{and } \liminf_{x \to y} f(x) \ge f(y). \quad \square$$

Remark 3.4.4 The assignment $x \mapsto \lambda(x)$ was a Lipschitzian one for lim sup and lim inf as well. This can be exploited for verifying that f is even locally Lipschitz near y. $\diamond$

Subgradients, subdifferential

Definition 3.4.1 Some $x^* \in X^*$ is a subgradient of $f : X \to \mathbb{R} \cup \{\infty\}$ at $x \in \operatorname{dom} f$, if

$$f(\xi) \ge f(x) + \langle x^*, \xi - x \rangle \quad \forall\, \xi \in X. \tag{3.4.2}$$

The closed convex set $\partial f(x)$ of all subgradients is the *subdifferential of* f at x. $\diamond$

If $X = \mathbb{R}^n$, we identify $X^* \equiv \mathbb{R}^n$ and $\langle x^*, x \rangle = x^{*T} x = \sum_i x_i^* x_i$. If $x^* \in \partial f(x)$, the affine function $l(\xi) = f(x) + \langle x^*, \xi - x \rangle$ supports f, and $f(\xi) = -\infty$ is impossible.

Example 3.4.1 The convex function $f = -\sqrt{x}$ if $x \ge 0$; $f = \infty$ else, has $\partial f(0) = \emptyset$. For $f(x) = |x|$, we have $\partial f(0) = [-1, 1]$, for $f(x) = \|x\|_2$, $\partial f(0) =$ Euclidean unit ball. $\diamond$

Evidently, $\partial f(x) \ne \emptyset \Rightarrow$ f is *lower semi-continuous at* x (even weakly l.s.c.).
Moreover, $\partial f(x)$ is bounded provided that f is bounded above near x. Indeed, if $x + \varepsilon B \subset \operatorname{dom} f$ and $x^* \in \partial f(x)$, choose $u \in \varepsilon B$ in such a way that $\langle x^*, u \rangle \ge \frac{1}{2} \varepsilon \, \|x^*\|$. This yields $f(x+u) \ge f(x) + \frac{1}{2} \varepsilon \|x^*\|$ and ensures the assertion.

Subdifferentials, solution sets and conjugation:
Trivial but important

$$0 \in \partial f(x) \quad \Leftrightarrow \; x \text{ is a global minimizer of } f.$$

More general, the inverse map $(\partial f)^{-1}(x^*) := \{x \mid x^* \in \partial f(x)\}$ assigns, to the function $h(\xi) = f(\xi) - \langle x^*, \xi \rangle$ the set of minimizers $\operatorname{argmin} h$, i.e.,

$$(\partial f)^{-1}(x^*) \;=\; \operatorname{argmin} (f - x^*) \tag{3.4.3}$$

So ∂f is inverse to the solution map of a linearly perturbed optimization problem, and $\partial f(x) \ne \emptyset$ means $x \in \operatorname{argmin}(f - x^*)$ for some x^*. The function f^* of the extreme values

$$f^*(x^*) := \inf_{\xi \in X} \; f(\xi) - \langle x^*, \xi \rangle \qquad (\text{possibly } -\infty)$$

is called the *conjugate* of f. Though f^* is mostly applied to *convex* f, it holds

Lemma 3.4.5 *For arbitrary f, the conjugate $f^* : X^* \to \mathbb{R} \cup \{-\infty\}$ is concave.* $\diamond$

Proof. Indeed, for $\lambda \in (0,1)$, it holds

$$\begin{aligned} & f^*(\lambda x^* + (1-\lambda)y^*) \\ = \; & \inf_\xi \, [\, \lambda f(\xi) - \lambda\langle x^*, \xi\rangle \; + \; (1-\lambda) f(\xi) - (1-\lambda)\langle y^*, \xi\rangle \,] \\ \geq \; & \inf_\xi \, [\, \lambda f(\xi) - \lambda\langle x^*, \xi\rangle \,] \; + \; \inf_\xi \, [\, (1-\lambda) f(\xi) - (1-\lambda)\langle y^*, \xi\rangle \,] \\ = \; & \lambda f^*(x^*) + (1-\lambda) f^*(y^*). \qquad \square \end{aligned}$$

To obtain a convex conjugate, one often defines f^* with opposite sign. The subsequent statements are consequences of the separation theorem.

Theorem 3.4.6 *(existence of subgrad.)* *If $f : X \to \mathbb{R} \cup \{\infty\}$ is convex and f is continuous at $\bar{x} \in \operatorname{int}\operatorname{dom} f$ then $\partial f(\bar{x}) \neq \emptyset$.* $\diamond$

For continuity, see Thm. 3.4.3.

Proof. Define $M := \operatorname{epi} f = \{(t,x) \mid t \geq f(x)\}$. Then $(f(\bar{x}), \bar{x}) \notin \operatorname{int} M$ follows from $(f(\bar{x}) - \varepsilon, \bar{x}) \notin M$. On the other hand, $(f(\bar{x}) + \varepsilon, \bar{x}) \in \operatorname{int} M$ holds by continuity of f at $\bar{x}$ (it suffices upper semi-continuity; but this implies continuity). Thus M and $(f(\bar{x}), \bar{x})$ can be separated by Thm. 3.1.8: Some $(r,u) \in \mathbb{R} \times X^*$ (not zero) satisfies

$$\langle\, (r,u), (f(\bar{x}), \bar{x})\rangle \leq \langle\, (r,u), (t,x)\rangle \quad \forall\, (t,x) \in M. \tag{3.4.4}$$

The structure of M implies $r \geq 0$ (via $t \to \infty$). Furthermore, $\bar{x} \in \operatorname{int}\operatorname{dom} f$ shows that $r = 0$ would yield the contradiction $u \neq 0$ and $\langle u, \bar{x}\rangle \leq \langle u, x\rangle \;\; \forall\; x \in \operatorname{dom} f$. Thus, it follows $r > 0$; and (3.4.4) ensures that $-u/r$ is a subgradient (divide by r and put $t = f(x)$). $\square$

As an application, we regard the relation between $(\partial f)^{-1}$ and $\partial(-f^*)$ and verify Corollary 1.4.4. in [48], part I.

Lemma 3.4.7 *For convex $f : \mathbb{R}^n \to \mathbb{R}$, it holds $(\partial f)^{-1}(u) = \partial(-f^*)(u)$.* $\diamond$

Proof. Let $z \in (\partial f)^{-1}(u)$. By definition of this set, it holds $f^*(u) = f(z) - u^T z$, and by definition of $f^*(\xi) = \inf_x f(x) - \xi^T x$, it holds $f^*(\xi) \leq f(z) - \xi^T z \;\forall \xi$. Hence one obtains $f^*(\xi) - f^*(u) \leq -(\xi - u)^T z$ or equivalently

$$-f^*(\xi) \geq -f^*(u) + (\xi - u)^T z, \quad \textit{i.e.}, \quad z \in \partial(-f^*)(u).$$

Let $z \in \partial(-f^*)(u)$. Again by definition, this means $f^*(\xi) + z^T\xi \leq f^*(u) + z^T u \;\forall \xi$. Taking any $\xi \in \partial f(z)$ (not empty !), it follows $f^*(\xi) + z^T\xi = f(z)$. Summarizing, the assertion follows: $f(z) - z^T u = f^*(\xi) + z^T\xi - z^T u \leq f^*(u)$, i.e., $z \in (\partial f)^{-1}(u)$. $\square$

3.4.2 Directional derivative, subdifferential, optimality condition

Theorem 3.4.8 *For the directional derivative of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at x in direction u, it holds*

$$f'(x;u) := \lim_{t \downarrow 0} \frac{f(x+tu) - f(x)}{t} \;=\; \max\{\langle x^*, u\rangle \mid x^* \in \partial f(x)\,\}. \qquad \diamond \tag{3.4.5}$$

Note: For the same statement in normed spaces, suppose that f is continuous.
For directional derivatives, the limit has to be unique; in contrast to limits which will be called contingent derivatives or generalized directional derivatives, below.

Proof. By convexity, the quotients $h(t) = t^{-1}(f(x+tu) - f(x))$ are decreasing as $t \downarrow 0$, and $h(t) \geq \langle x^*, u\rangle$ holds for all $x^* \in \partial f(x) \neq \emptyset$. Thus, $f'(x;u) = \lim h(t)$ exists and $f'(x;u) \geq \max\{\langle x^*, u\rangle \mid x^* \in \partial f(x)\}$.
Next, given x and u, consider $f(x+tu)$ on the real line, and put $h = f'(x;u)$. Since $(\forall t > 0)$ we have $f(x+tu) - f(x) \geq th$ and $f(x-tu) - f(x) \geq -th$, the line

$$G = \{(x+tu,\ f(x)+th) \mid t \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$$

does not meet $\operatorname{int}\operatorname{epi} f$. By separation of G and $\operatorname{int}\operatorname{epi} f$, some pair $(w,r) \neq (0,0)$ exists such that, for all $t \in \mathbb{R}, p > 0$ and $y \in \mathbb{R}^n$,

$$\langle w, x+tu\rangle + r(f(x)+th) \ \leq\ \langle w, x+y\rangle + r(f(x+y)+p).$$

Via $p \to \infty$, this implies $r \geq 0$, and $r = 0$ can be excluded by setting $y = -w \neq 0$ and considering small $t > 0$. Hence it holds $r > 0$ and, after division by r it follows with $v = \frac{w}{r}$ (and by $p \downarrow 0$),

$$\begin{array}{lll} \langle v, x+tu\rangle + f(x) + th & \leq & \langle v, x+y\rangle + f(x+y) \quad \forall t \in \mathbb{R}, y \in \mathbb{R}^n, \\ \langle v, tu - y\rangle + t\ h & \leq & f(x+y) - f(x) \qquad\quad \forall t \in \mathbb{R}, y \in \mathbb{R}^n. \end{array}$$

Setting $y = 0$ we obtain $h = -\langle v, u\rangle$. Therefore, it holds

$$\langle -v, y\rangle \ \leq\ f(x+y) - f(x)\ \forall y \in \mathbb{R}^n \ \text{ and } \ x^* := -v \in \partial f(x).$$

Due to $\langle x^*, u\rangle = h$, so also $f'(x;u) \leq \max\{\langle x^*, u\rangle \mid x^* \in \partial f(x)\}$ is valid. □

Sublinearity:
For directions u, v and $\lambda \geq 0$, it follows immediately $f'(x, \lambda u) = \lambda f'(x,u)$ and by the Theorem

$$f'(x, u+v) = \max_{x^* \in \partial f(x)} \langle x^*, u+v\rangle \leq \max_{x^* \in \partial f(x)} \langle x^*, u\rangle + \max_{x^* \in \partial f(x)} \langle x^*, v\rangle = f'(x,u) + f'(x,v).$$

So f' is sublinear (and convex) in the direction.

Theorem 3.4.9 *(Moreau / Rockafellar) Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be convex. Then*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x) \qquad \textit{(element-wise sum)} \qquad \diamond \tag{3.4.6}$$

Proof. This theorem is once more an application of the separation theorem. Show first that one may (by adding a linear function) suppose that x is a minimizer of $f+g$. Next, by separating epi f and hypo $(-g)$, one obtains the crucial inclusion $0 \in \partial f(x) + \partial g(x)$. □

For any set $M \subset X$ and $x \in M$ one can define 3 important cones

Definition 3.4.2
1. The (classical) normal cone of M at x is $N_M(x) = \{x^* \mid \langle x^*,\ y - x\rangle \leq 0\ \forall y \in M\}$. This is the polar cone $(M-x)^*$ of $M - x$.
2. The contingent (or Bouligand) cone of M at x is defined as

$$C_M(x) = \{u \in X \mid \liminf_{t \downarrow 0} \frac{\operatorname{dist}(x+tu, M)}{t} = 0\}.$$

3. The third one is the polar cone $\mathcal{N}_M(x) = C_M(x)^*$ of the contingent cone. ◇

The cone $C_M(x)$ is a (closed) cone of "tangent directions". Equivalently, it holds

$$\begin{array}{lll} u \in C_M(x) \;\Leftrightarrow\; \exists t_k \downarrow 0 :\; \mathrm{dist}\,(x + t_k u, M) = o(t_k) & & \\ \quad \Leftrightarrow\; \exists t_k \downarrow 0 :\quad x + t_k u_k \;\in M & \text{for certain } u_k \to u & \\ \quad \Leftrightarrow\; \exists t_k \downarrow 0 :\; x + t_k u + w^k \;\in M & \text{for certain } w^k \text{ with } \|w^k\|/t_k \to 0. & \end{array} \tag{3.4.7}$$

Clearly, $C_M(x)$ is convex if so is M. By definition, $N_M(x)$ is polar to $M - x$, is closed, convex and satisfies $N_M(x) = N_{\mathrm{conv}\, M}(x)$. Alternatively, the indicator function i_M can be used to *define* $N_M(x) = \partial i_M(x)$. Hence N_M is also a subdifferential.

Example 3.4.2 For the non-convex set $M = \{x \in \mathbb{R}^2 \mid x_2 \geq -x_1^2\}$, it holds

$$\begin{array}{ll} N_M(0) & = \{0\}, \\ C_M(0) & = \{u \in \mathbb{R}^2 \mid u_2 \geq 0\}, \\ C_M(0)^* & = \{x^* \in \mathbb{R}^2 \mid x_1^* = 0,\; x_2^* \leq 0\} \end{array}$$

Here, $N_M(x) \neq \mathcal{N}_M(x) = C_M(x)^*$. ◇

On the other hand, one easily shows $N_M(x) = \mathcal{N}_M(x)$ if M is closed and convex.

Theorem 3.4.10 *(Optimality condition for convex problems) Let $M \subset \mathbb{R}^n$ be closed and convex, $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $x \in M$.*
Then, x solves $\min\{f(\xi) \mid \xi \in M\}$ $\Leftrightarrow$ x satisfies the minimum-condition

$$0 \in \partial f(x) + \mathcal{N}_M(x). \qquad \diamond \tag{3.4.8}$$

Proof. By convexity, x solves the min-problem $\Leftrightarrow$ $f'(x;\; y - x) \geq 0\ \forall y \in M$. From Thm. 3.4.8 we know $f'(x;\; y - x) = \max_{x^* \in \partial f(x)} \langle x^*,\; y - x\rangle$. (3.4.8) tells us

$$\exists\; x^* \in \partial f(x) \;\text{ with }\; -x^* \in \mathcal{N}_M(x), \tag{3.4.9}$$

Thus optimality follows from (3.4.8) due to

$$f'(x;\; y - x) \geq \langle x^*,\; y - x\rangle \geq 0 \quad \forall y \in M.$$

Conversely, if (3.4.8) is violated, the compact convex set $\partial f(x)$ and the closed convex cone $-\mathcal{N}_M(x)$ can be separated:

$$\exists u :\; \langle u, \partial f(x)\rangle < \langle u, -\mathcal{N}_M(x)\rangle$$

(the inequality holds for any selected elements). Since $0 \in \mathcal{N}_M(x)$, this yields

$$f'(x; u) < 0 \;\text{ and }\; u \in \mathcal{N}_M(x)^*.$$

Thm. 3.1.5 (bipolar) ensures $\mathcal{N}_M(x)^* = [\, C_M(x)^* \,]^* = C_M(x)$. Hence we have $u \in C_M(x)$, after which $x + t_k u_k \in M$ follows for certain $u_k \to u,\; t_k \downarrow 0$. Since f is locally Lipschitz by Remark 3.4.4, so also $f(x + t_k u_k) < f(x)$ is true. Thus x was not optimal. □

Notice: The proof shows that condition (3.4.8) means nothing else than

$$f'(x, u) \geq 0 \quad \forall u \in C_M(x). \tag{3.4.10}$$

3.4.3 Particular subdifferentials in analytical form

Subdifferentials (like derivatives) are only helpful if they can be determined. For this purpose the next Lemmas are of importance.

Lemma 3.4.11 *(maximum-function I) For*

$$f(x) = \max_i g_i(x); \quad g_i : \mathbb{R}^n \to \mathbb{R} \text{ convex, } g_i \in C^1 \ (i = 1, ..., m) \tag{3.4.11}$$

it holds $\partial f(\bar{x}) = \text{conv}\,\{Dg_i(\bar{x}) \mid g_i(\bar{x}) = f(\bar{x})\}$. ◇

Proof. Put $I(\bar{x}) = \{i \mid g_i(\bar{x}) = f(\bar{x})\}$ and $M = \text{conv}\,\{Dg_i(\bar{x}) \mid i \in I(\bar{x})\}$.

$M \subset \partial f(\bar{x})$: Since g_i is convex, we have for all x,

$$\begin{array}{lll} g_i(x) & \geq & g_i(\bar{x}) + \langle Dg_i(\bar{x}), x - \bar{x}\rangle \qquad \forall i \in I(\bar{x}) \quad \text{and} \\ f(x) & \geq & g_i(x) \ \geq \ f(\bar{x}) + \langle Dg_i(\bar{x}), x - \bar{x}\rangle \quad \forall i \in I(\bar{x}). \end{array}$$

Thus, $Dg_i(\bar{x}) \in \partial f(\bar{x})$ and, by convexity of $\partial f(\bar{x})$, also $M \subset \partial f(\bar{x})$ holds true.

$\partial f(\bar{x}) \subset M$: Assume, in contrary, that some $x^* \in \partial f(\bar{x}) \setminus M$ exists. Define

$$\hat{g}_i(x) = g_i(x) - \langle x^*, x\rangle \text{ and } \hat{f}(x) = f(x) - \langle x^*, x\rangle.$$

Then, $\hat{f}(x) = \max_i \hat{g}_i(x)$ and $0 \in \partial\hat{f}(\bar{x}) \setminus \text{conv}\,\{D\hat{g}_i(\bar{x}) \mid i \in I(\bar{x})\}$. By separation of 0 and $\text{conv}\,\{D\hat{g}_i(\bar{x}) \mid i \in I(\bar{x})\}$, some $u \in \mathbb{R}^n$ exists such that

$$0 < \langle D\hat{g}_i(\bar{x}), u\rangle \ \forall i \in I(\bar{x}).$$

Hence, for $i \in I(\bar{x})$ and small $t > 0$, we also have $\hat{g}_i(\bar{x} - tu) < \hat{g}_i(\bar{x}) = \hat{f}(\bar{x})$.
For $i \notin I(\bar{x})$, the inequality $\hat{g}_i(\bar{x} - tu) < \hat{f}(\bar{x})$ follows from $\hat{g}_i(\bar{x}) < \hat{f}(\bar{x})$.
Summarizing, this ensures $\hat{f}(\bar{x} - tu) < \hat{f}(\bar{x})$, in contradiction to $0 \in \partial\hat{f}(\bar{x})$. □

Lemma 3.4.12 *(maximum-function II) Suppose that*

$$f(x) = \max_{t \in T} g(x, t) \qquad \textit{where } T \textit{ is a compact metric space,}$$

$g(., t) : \mathbb{R}^n \to \mathbb{R}$ *is convex,* g_x *exists and* g *and* g_x *are continuous in both variables. Then,* $\partial f(\bar{x}) = \text{conv}\,\{g_x(\bar{x}, t) \mid \ g(\bar{x}, t) = f(\bar{x})\}$. ◇

Proof. Put $T^0 = \{t \in T \mid g(\bar{x}, t) = f(\bar{x})\}$ and $M = \text{conv}\,\{g_x(\bar{x}, t) \mid t \in T^0\}$.

(i) $M \subset \partial f(\bar{x})$: This follows as for Lemma 3.4.11.

(ii) $\partial f(\bar{x}) \subset M$: Assume that some $x^* \in \partial f(\bar{x}) \setminus M$ exists. As in the former proof, we may add $-\langle x^*, x\rangle$ to all functions under consideration. Thus we may assume that $x^* = 0$. Since M is convex and compact, separation from 0 provides us with some u such that

$$g_x(\bar{x}, t)u < -2\delta < 0 \quad \forall t \in T^0.$$

By continuity of g_x and compactness of T then some $\alpha > 0$ exists such that

$$g_x(x, s)u < -\delta \quad \forall s \in T \cap (T^0 + \alpha B) \text{ and } x \in \bar{x} + \alpha B. \tag{3.4.12}$$

In addition, some $\beta > 0$ exists such that $g(\bar{x}, t) < f(\bar{x}) - 2\beta \ \forall t \in T \setminus (T^0 + \alpha B)$ and, by continuity,

$$g(x, t) < f(\bar{x}) - \beta \quad \forall t \in T \setminus (T^0 + \alpha B) \text{ and } x \text{ near } \bar{x}. \tag{3.4.13}$$

With $x = \bar{x} + \lambda u$ for small $\lambda > 0$, we thus obtain from (3.4.12) and (3.4.13), a contradiction to $0 \in \partial f(\bar{x})$, namely $f(\bar{x} + \lambda u) = \max_{t \in T} g(x + \lambda u, t) < f(\bar{x})$. □

Exercise: Show that $\partial f(0) = B$ holds for the Euclidean norm $f(x)$.

3.5 Normal cones and Variational Inequalities

3.5.1 Definitions and Basic motivations

Necessary optimality condition for the (possibly non-convex) problem

$$\min\ f(x), \quad x \in M \subset X; \qquad f \in C^1,\ M \text{ closed}. \tag{3.5.1}$$

Consider any $\bar{x} \in M$, $u \in C_M(\bar{x})$ and elements $u_k \in X,\ t_k > 0$ such that

$$x_k := \bar{x} + t_k\ u_k \in M \quad \text{where } u_k \to u \text{ and } t_k \downarrow 0 \text{ as } k \to \infty. \tag{3.5.2}$$

Then $Df(\bar{x})u < 0$ implies $f(x_k) < f(\bar{x})$ for large k. Hence we obtain immediadly,

Lemma 3.5.1 *(NLO-Optimality primal) If $\bar{x}$ is a local solution to (3.5.1) then*

$$Df(\bar{x})u \geq 0\ \forall u \in C_M(\bar{x}). \tag{3.5.3}$$

After formalizing once more, this obvious necessary optimality condition reads (by the definitions only) as

Lemma 3.5.2 *(NLO-Optimality dual) If $\bar{x}$ is a local solution to (3.5.1) then*

$$-Df(\bar{x}) \in \mathcal{N}_M(\bar{x}) \qquad \text{where } \mathcal{N}_M(\bar{x}) = C_M(\bar{x})^*. \qquad \diamond\ \square \tag{3.5.4}$$

This condition, often written as

$$0 \in Df(\bar{x}) + \mathcal{N}_M(\bar{x}),$$

will be modified for locally Lipschitz functions f in chapter 9.
Recall: If M is convex then $\mathcal{N}_M(\bar{x}) = N_M(\bar{x})$. If also f is a convex function, the necessary condition (3.5.3) of Lemma 3.5.1 becomes

$$-Df(\bar{x}) \in N_M(\bar{x}) \ \text{ or explicitly } \ \langle -Df(\bar{x}),\ y - \bar{x}\rangle \leq 0\ \forall y \in M \tag{3.5.5}$$

and is even necessary and sufficient for $\bar{x} \in \ \operatorname{argmin}\{f(x) \mid x \in M\}$. In dual form, this is Thm. 3.4.10.

If $f \in C^{0,1}$ has directional derivatives $f'(\bar{x}; u)$, condition (3.5.3) of Lemma 3.5.1 remains (obviously) necessary for (loc.) optimality after replacing $Df(\bar{x})u$ by $f'(\bar{x}; u)$.

(Quasi-)Variational Inequality (Quasi VI).
Replacing $-Df(x)$ in (3.5.5) by any continuous function $g : M \subset X \to X^*$, the inclusion

$$g(x) \in N_M(x), \quad \text{or explicitly } \ \langle g(x), y - x\rangle \leq 0\ \forall y \in M \tag{3.5.6}$$

is called a variational inequality (VI), cf. sect. 3.3.2 where $g = \hat{f}$. Several modifications of such systems are possible; so the condition

$$\langle g(x), y - x\rangle \leq 0\ \forall y \in \Gamma(x) \ \text{ where } \Gamma : M \rightrightarrows X \text{ is multivalued} \tag{3.5.7}$$

is often called a Quasi-Variational Inequality (Quasi VI). Similarly, one speaks about variational inequalities if the cones N_M and $\mathcal{N}_M$ are exchanged. Notice that (3.5.7) can be written (by the definitions only) as $g(x) \in N_{\operatorname{conv}\{x,\Gamma(x)\}}(x)$.

In order to obtain a closed mapping $x \mapsto N_{\operatorname{conv}\{x,\Gamma(x)\}}(x)$ we assume that Γ is l.s.c., i.e., $\Gamma^{-1}(\Omega) := \{x \mid \Gamma(x) \cap \Omega \neq \emptyset\}$ is open for all open sets $\Omega \subset X$.

3.5.2 VI's and Fixed-points

We are now going to show that, whenever M is convex and compact and the (multi) functions are well-defined on M, the VI (3.5.6) is equivalent to a Brouwer- FP- problem while the quasi VI (3.5.7) is equivalent to a Kakutani FP- problem. Let us continue in considering (3.5.7). Setting

$$C(x) = \text{conv}\,(\ \{x\} \cup \Gamma(x)\ ) \quad \text{and} \quad G(x) = N_{C(x)}(x),$$

(3.5.7) is equivalent to

$$\langle g(x),\ u - x\rangle \leq 0\ \forall u \in C(x),$$

hence also to the *generalized equation*

$$g(x) \in G(x) \tag{3.5.8}$$

and, with $F(x) := x - g(x) + G(x)$ for $x \in M$, to the FP- problem

$$x \in F(x), \qquad x \in M. \tag{3.5.9}$$

By the construction, the sets $F(x)$, depending on g and Γ, are non-empty, convex and closed. Notice also that, for solving (3.5.9), only the sets $\hat{F}(x) = F(x) \cap M$ are of interest. They are compact if so is M. So (3.5.9) shows that Kakutani's Theorem is considerably important for all quasi VI's since,

Corollary 3.5.3 *The quasi VI (3.5.7) is solvable, if $\emptyset \neq M \subset \mathbb{R}^n$ is convex and compact, $g : M \to \mathbb{R}^n$ is continuous, Γ is closed and l.s.c., and the sets $F(x) = x - g(x) + G(x)$ have ($\forall x \in M$) a non-empty intersection $\hat{F}(x)$ with M.* ◇

Like Kakutani's FP- theorem for quasi VI's, Brouwer's FP-theorem is crucial for VI's.

Corollary 3.5.4 *The VI (3.5.6) is solvable if $\emptyset \neq M \subset \mathbb{R}^n$ is convex and compact and $g : M \to \mathbb{R}^n$ is continuous.* ◇

Proof. Since g is continuous and M is compact, there is some (closed) ball such that

$$M \cup g(M) \cup (M + g(M)) \subset C = B(0, r). \tag{3.5.10}$$

Using the projection π_M onto M, define $f(z) = \pi_M(z) + g(\pi_M(z)) \quad \forall z \in C$.
By Brouwer's FP theorem, the continuous function $f : C \to C$ has a fixed-point:

$$\bar{z} = f(\bar{z}) = \pi_M(\bar{z}) + g(\pi_M(\bar{z})) \quad \text{for some } \bar{z} \in C. \tag{3.5.11}$$

Then $\bar{x} = \pi_M(\bar{z})$ belongs to M, (3.5.11) entails $g(\bar{x}) = \bar{z} - \bar{x}$, and $\bar{z} - \bar{x} \in N_M(\bar{x})$ follows from the projection property (3.1.5). □

Corollary 3.5.5 *If $\emptyset \neq M \subset \mathbb{R}^n$ is convex and compact and $g : M \to \mathbb{R}^n$ continuous, the solutions of VI (3.5.6) coincide - via $x = \pi_M(z)$ - with the solutions to the FP- equation $f(z) = z$ where $z \in C$ (3.5.10) and $f = \pi_M + g \circ \pi_M$* ◇

Proof. Indeed, if z is a fixed-point then $\pi_M(z)$ solves (3.5.6), as just shown. Conversely, if x solves (3.5.6) and $g(x) = \mu \in N_M(x)$ then $g(x) + x = x + \mu$. The point $z = x + \mu$ is a fixed point since $\mu \in N_M(x)$ implies, by Thm. 3.1.2, $\pi_M(z) = \pi_M(\mu + x) = x$. □

Alternatively, one may replace VI (3.5.6) with the equation $x = \pi_M(x + g(x))$.
For $g \in C^1$, there is some technical difference between the nondifferentiable functions

$$F_1(x) = \pi(g(x)) \quad \text{and} \quad F_2(x) = g(\pi(x))$$

the type of which is used now or in the corollary: Several (multivalued) generalized derivatives $\mathcal{D}$ for the composed function F satisfy the "exact" chain rule

$$\mathcal{D}F_2(x) = Dg(\pi(x))\mathcal{D}\pi(x), \quad \text{but only} \quad \mathcal{D}F_1(x) \subset \mathcal{D}\pi(g(x))Dg(x),$$

cf. sect. 7.3.2. More skillful existence theorems for generalized equations and VI's can be found in [81, 124, 143] and, under the "almost weakest" conditions for B-spaces, in [76]. Often, conditions like $x \in M$ are "generalized" in the literature by writing $u(x) \in M$ with some function u. However, whenever M can be given in an analytical form, say $M = \{y \mid v(y) \le 0\}$ the difference vanishes since $u(x) \in M \Leftrightarrow v(u(x)) \le 0$.

3.5.3 Monotonicity

Definition 3.5.1 (monotonicity)
Let $g : M \subset X \to X^*$. We require the next conditions for all $x, x' \in M,\ x \neq x'$.

(1) g is monotone if $\quad \langle g(x') - g(x),\ x' - x\rangle \ge 0.$

(2) g is strictly monotone if $\quad \langle g(x') - g(x),\ x' - x\rangle > 0.$

(3) g is strongly monotone if $\exists c > 0$: $\langle g(x') - g(x),\ x' - x\rangle \ge c\|x' - x\|^2.$ $\quad \diamond$

These properties remain valid after adding a monotone function h, $g_h = g + h$ like $h(x) = \lambda x, \lambda > 0$ for $X = \mathbb{R}^n$.

Similarly, one defines monotonicity for multifunctions $G : M \subset X \rightrightarrows X^*$: The requirements have to hold for all $g(x) \in G(x)$ and $g(x') \in G(x')$, respectively.

Usually, M coincides with $\operatorname{dom} G$ or $\operatorname{dom} g$ for mappings which are formally defined on X. In all other situations, one should better say that g or G are monotone on M.

Lemma 3.5.6 *(monotonicity of ∂f) If $f : X \to \mathbb{R} \cup \{\infty\}$ is convex then*

$$[\ x^* \in \partial f(x) \text{ and } y^* \in \partial f(y)\ ] \ \Rightarrow\ \langle y^* - x^*,\ y - x\rangle \ge 0. \qquad \diamond$$

Proof. Let $p = \frac{1}{2}(x + y)$. Then $f(p) \ge f(x) + \langle x^*, p - x\rangle,\ f(p) \ge f(y) + \langle y^*, p - y\rangle$ and by convexity $2f(p) \le f(x) + f(y)$. Adding the first inequalities yields the assertion:

$$f(x) + \langle x^*, p - x\rangle + f(y) + \langle y^*, p - y\rangle \le\ 2f(p)\ \le f(x) + f(y),$$

and, in consequence, $0 \ge \langle x^*, p - x\rangle + \langle y^*, p - y\rangle = -\frac{1}{2}\ \langle x^* - y^*,\ y - x\rangle$. $\square$

Due to $N_M = \partial i_M$ so also the normal cone map N_M for a convex set $M \subset X$ is monotone.

Lemma 3.5.7 *For convex $M \neq \emptyset$, consider the solutions to*

$$0 \ \in\ g(x) + N_M(x). \tag{3.5.12}$$

(i) If g is strictly monotone then there is at most one solution.
(ii) If g is strongly monotone and $M \subset \mathbb{R}^n$ is closed then there is a unique solution. $\diamond$

Proof. Let us first estimate two solutions of $p \in g(x)+N_M(x)$ and $p' \in g(x')+N_M(x')$. We may write, with ν, ν' in the related cones, $p = g(x) + \nu,\ p' = g(x') + \nu'$, as well as

$$p' - p = g(x') - g(x) + \nu' - \nu \ \text{ and } \ \langle p' - p, x' - x\rangle = \langle g(x') - g(x) + \nu' - \nu,\ x' - x\rangle.$$

By monotonicity of N_M (Lemma 3.5.6) this yields via $\langle \nu' - \nu,\ x' - x\rangle \geq 0$ the crucial estimate

$$\|p' - p\|\ \|x' - x\| \geq \langle p' - p, x' - x\rangle \geq \langle g(x') - g(x),\ x' - x\rangle. \tag{3.5.13}$$

(i) For two solutions x, x' of $0 \in g(x) + N_M(x)$ this implies $0 \geq \langle g(x') - g(x),\ x' - x\rangle$. If g is strictly or strongly monotone, so $x' = x$ follows.
(ii) Under strong monotonicity, inequality (3.5.13) implies

$$\|p' - p\|\ \|x' - x\| \geq c\|x' - x\|^2. \tag{3.5.14}$$

Hence, x' and x satisfy

$$\|x' - x\| \leq \ \frac{\|p' - p\|}{c} \ \text{ and } \ \|x'\| < r := \|x\| + 1 + \ \frac{\|p' - p\|}{c}. \tag{3.5.15}$$

Next, with fixed $x \in M \neq \emptyset$, choose some $p \in g(x) + N_M(x)$ and consider any solution $x' \in M;\ 0 \in g(x') + N_M(x')$. Then x' satisfies

$$\|x' - x\| \leq \ \frac{\|p\|}{c} \quad \text{ and } \ \|x'\| < r := \|x\| + 1 + \ \frac{\|p\|}{c}.$$

These estimates depend on g only and remain true if we replace M by $M_r := M \cap B(0, r)$. Due to (3.5.15), we have $x, x' \in \operatorname{int} B(0, r)$. Thus the solution sets for M and M_r coincide. With M_r, the hypotheses of Cor. 3.5.4 are satisfied. Hence (3.5.12) has exactly one solution. □

The case of **coercive** g: Similarly, one may suppose (instead of strong monotonicity) that g is *coercive*, i.e., for some fixed $x \in M$, it holds

$$\frac{\langle g(x') - g(x),\ x' - x\rangle}{\|x' - x\|} \ \to \infty \quad \text{ if } \|x' - x\| \to \infty. \tag{3.5.16}$$

This can be used like formula (3.5.13) in the proof above.

Corollary 3.5.8 *If $M \neq \emptyset$ is closed and convex and g is coercive then (3.5.12) is solvable. (uniqueness is no longer ensured)* ◇

Monotonicity alone plays no important role. The appropriate notion is that of maximal monotonicity of a multifunction G. This means besides monotonicity that there is no other monotone mapping H such that $\operatorname{gph} G \subset \operatorname{gph} H$. In particular, then G is closed.

3.5.4 Analytical formulas for C_M and $\mathcal{N}_M = C_M^*$ and hypo-monotonicty

Having all these relations, it becomes desirable to derive formulas for the cones in question provided that some concrete description of M is known. For sets of the form

$$M = \{x \in \mathbb{R}^n \mid h_k(x) = 0,\ k = 1, .., m_h, \ \ g_i(x) \leq 0\ i = 1, 2, ..., m\},\ h, g \in C^1, \tag{3.5.17}$$

Corollary 4.1.3 can be applied: If $\bar{x} \in M$ satisfies a *Constraint Qualification*, it holds

$$C_M(\bar{x}) = \{u \in \mathbb{R}^n \mid Dh(\bar{x})u = 0 \text{ and } \ Dg_i(\bar{x})u \leq 0 \ \forall i \in I(\bar{x})\} \tag{3.5.18}$$

where $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$.
Then, the cone $\mathcal{N}_M(\bar{x}) = C_M(\bar{x})^*$ (Def. 3.4.2) has - by LOP duality Thm. 2.1.5 - the explicit analytical form

$$\mathcal{N}_M(\bar{x}) = \{x^* \in \mathbb{R}^n \mid \exists (y,z) \in \mathbb{R}^{m+m_h} : x^* = Dh(\bar{x})^T z \ + \sum_{i \in I(\bar{x})} y_i Dg_i(\bar{x}), \ y \geq 0\}. \tag{3.5.19}$$

Hence $\mathcal{N}_M(\bar{x})$ is the sum of the subspace, spanned by all $Dh_k(\bar{x})$, and the polyhedral cone, generated by the "active" gradients $Dg_i(\bar{x})$. Though direct proofs of (3.5.19) could be given here, based on "constraint qualifications" and on the already summarized facts, we give the proofs within the framework of the classical KKT-theory in chapter 4. We only notice that (3.5.19) along with Lemma 3.5.1 implies the important KKT-Theorem 4.1.2.
Exercise:
Let $M = \{x \in \mathbb{R}^2 \mid x_1 \leq 0, \ x_2 \leq 0, \ x_1 x_2 = 0\}$ be the boundary of the negative orthant and $\bar{x} = 0$. Determine $N_M(\bar{x})$ and $\mathcal{N}_M(\bar{x})$ and check equation (3.5.19) for these cones.

Hypo-monotonicity of $\mathcal{N}_M(x)$
Under MFCQ at $\bar{x}$ (defined below) and $h, g \in C^2$, the map $\mathcal{N}_M$ of normals has, even for x near $\bar{x}$, $x \in M$, the form (3.5.19). Using $y_i^+ = \max\{0, y_i\}, \ y_i^- = \min\{0, y_i\}$ this is

$$\mathcal{N}_M(x) = \{x^* \mid x^* = Dh(x)^T z + Dg(x)^{\mathsf{T}} y^+, \ \ g(x) = y^-\}. \tag{3.5.20}$$

Since convexity is not supposed, $\mathcal{N}_M$ is no longer monotone, but has a local quadratic miniorant, i.e.,

Lemma 3.5.9 *There are $c \in \mathbb{R}$ (possibly < 0) and $\delta > 0$ such that (with $x, x' \in M$),*

$$\langle x^* - x'^*, x - x' \rangle \geq c \, \|x' - x\|^2 \quad \forall \, x'^* \in \mathcal{N}_M(x'), \ x^* \in \mathcal{N}_M(x), \ x', x \in B(\bar{x}, \delta). \ \diamond \tag{3.5.21}$$

Proof. Consider $(x, x^*), (x', x'^*)$ near $(\bar{x}, x^{0*})$; all in $\operatorname{gph} \mathcal{N}_M$. For small $\delta > 0$, all related elements y, y', z, z' are bounded by the hypothesis MFCQ (cf. remark 4.1.7). The products have the form

$$\langle x^* - x'^*, \ x - x' \rangle = [z^T Dh(x) - z'^T Dh(x')](x - x') \ + \sum_i [y_i^+ Dg_i(x) - y'^+_i Dg_i(x')](x - x').$$

With $t = \|x - x'\|$ and $u = \frac{x - x'}{t}$ we estimate the sum $R = \sum ...$ on the right-hand side

$$\frac{R}{t^2} \ = \ \sum y'^+_i \ \frac{Dg_i(x) - Dg_i(x')}{t} u \ + \ \sum (y_i^+ - y'^+_i) \frac{Dg_i(x) u}{t}$$

and show that $\frac{R}{t^2}$ is bounded below. The first term $\sum y'^+_i \ \frac{Dg_i(x) - Dg_i(x')}{t} u$ is bounded since so are the LM y', and Dg_i is loc. Lipsch. Let us turn to the next one.
Because of $x, x' \in M$, now $y_i > 0$ yields $g_i(x) = 0$ and, applying $x = x' + tu$ as well as

$$g_i(x) - g_i(x') = t Dg_i(x')u + \tfrac{1}{2} t^2 u^T D^2 g_i(\theta_i) u \tag{3.5.22}$$

and $\qquad t Dg_i(x')u = t Dg_i(x)u + o(t^2), \qquad$ we have

$$Dg_i(x)u \ + \frac{o_i(x - x')}{t} = \frac{g_i(x) - g_i(x')}{t} \geq 0.$$

Similarly, $y_i' > 0$ yields $g_i(x') = 0$ and $Dg_i(x)u + \frac{o_i(x-x')}{t} = \frac{g_i(x)-g_i(x')}{t} \le 0$. Thus

$$\sum (y_i^+ - y_i'^+) Dg_i(x)u \ \ge \ \frac{o(x'-x)}{t}.$$

Due to $g \in C^2$, the functions $|o_i|$ and $|o|$ have quadratic majorants, $|o + o_i| \le Kt^2$. This entails the crucial estimate for the second term of R/t^2,

$$\sum (y_i^+ - y_i'^+) \frac{Dg_i(x)u}{t} \ge -K.$$

Handling the sum $[z^T Dh(x) - z'^T Dh(x')](x - x')$ in the same way, (3.5.21) is shown. □

In consequence, we see that $F(x) = \mathcal{N}_M(x)$ is a multifunction with the property that, for some $\lambda > 0$,

$$x \mapsto \lambda x + F(x)$$

is strongly monotone near $\bar{x}$ (put $\lambda = |c| + 1$). Such mappings F are called *max-hypomonotone*, provided the graph is closed as here. Also this property can be used for existence statements. For more details see [118]. With the Taylor formula in sect. 11.2.6 instead of (3.5.22), the estimate (3.5.21) also follows for $g, h \in C^{1,1}$.

3.6 An Application: Tschebyschev-approximation

Find a real polynomial $p_x(t) = x_0 + x_1 t + ... + x_n t^n$ of maximal degree $\le n$ such that a given continuous function $q = q(t)$ on a compact interval T is approximated in the best way by the max- norm

$$\textit{minimize} \ \ f(x) = \max_{t \in T} \ |p_x(t) - q(t)| \ , \quad x = (x_0, ..., x_n).$$

Here, f is convex and satisfies the assumptions of Lemma 3.4.12 with

$$g_1(x,t) = p_x(t) - q(t), \quad g_2 = -g_1 \quad \text{and} \ \ f(x) = \max_{t \in T} \{g_1(x,t), \ g_2(x,t)\}.$$

Assume that $\bar{x}$ is a solution with $f(\bar{x}) > 0$ and put

$$T_1^0 = \{t \in T \mid g_1(\bar{x}, t) = f(\bar{x})\} \ \text{ and } \ T_2^0 = \{s \in T \mid g_2(\bar{x}, s) = f(\bar{x})\}.$$

By Lemma 3.4.12 it holds

$$0 \in \partial f(\bar{x}) = \ \text{conv} \ (\ \{D_x g_1(\bar{x}, t) \mid t \in T_1^0\} \cup \{D_x g_2(\bar{x}, s) \mid s \in T_2^0\} \ ).$$

Due to Thm. 3.1.1 there are m_1 elements $t_i \in T_1^0$ and m_2 elements $s_j \in T_2^0$ such that

$$m_1 + m_2 = n + 2 \ \text{ and } \ 0 \in \ \text{conv} \ (\ \{D_x g_1(\bar{x}, t_i)\} \cup \{D_x g_2(\bar{x}, s_j)\} \ ).$$

Hence there exist $n + 2$ non-negative λ_i and μ_j satisfying the $n + 2$ conditions

$$\sum_{i=1}^{m_1} \lambda_i + \sum_{j=1}^{m_2} \mu_j = 1 \quad \text{and} \quad \sum_{i=1}^{m_1} \lambda_i \ (1, t_i, t_i^2, ..., t_i^n) \ = \ \sum_{j=1}^{m_2} \mu_j \ (1, s_j, s_j^2, ..., s_j^n)$$

or equivalently

$$\sum_{i=1}^{m_1} \lambda_i = \frac{1}{2} = \sum_{j=1}^{m_2} \mu_j \quad \text{and} \quad \sum_{i=1}^{m_1} \lambda_i \ (t_i, t_i^2, ..., t_i^n) \ = \ \sum_{j=1}^{m_2} \mu_j \ (s_j, s_j^2, ..., s_j^n).$$

Tschebyschev: The signs of $p_x(r) - q(r)$ alternate for increasing $r \in T_1^0 \cup T_2^0$. Why ?

Chapter 4

Nonlinear Problems in $\mathbb{R}^n$

We investigate the standard problems

$$\min\{\ f(x)\ |\ \ g_i(x) \le 0,\ i = 1, 2, ..., m\} \quad f, g \in C^1, \tag{4.0.1}$$

$$\begin{array}{l} \min\{\ f(x)\ |\ x \in M\ \} \quad \text{where} \quad f, g, h \in C^1 \quad \text{and} \\ M = \{\ x \in \mathbb{R}^n\ |\ \ g_i(x) \le 0\ \forall i = 1, 2, ..., m, \quad h_j(x) = 0\ \forall j = 1, .., m_h\} \end{array} \tag{4.0.2}$$

and shall consider the feasible sets after small perturbations of the right-hand sides:

$$S(b) = \{x\ |\ g(x) \le b\}, \qquad S(b, c) = \{x\ |\ g(x) \le b,\ h(x) = c\}.$$

Clearly, since all equations $h_j = 0$ can be written as $h_j \le 0, -h_j \le 0$, the second problem is not more general. But if we need that $S(b) \neq \emptyset$ for small $\|b\|$, equations should not be written as inequalities since $h_j \le -\varepsilon,\ -h_j \le -\varepsilon$ is unsolvable.

We derive now necessary and sufficient conditions for optimality. The set M has exactly the form (3.5.17), section 3.5.1, and the formulas (3.5.18), (3.5.19) for tangents and normals will be verified under an assumption called constraint qualification (CQ, Regularitätsbedingung).

4.1 Notwendige Optimalitätsbedingungen

4.1.1 KKT- Bedingungen, Lagrange- Multiplikatoren

Dieser Abschnitt liefert einen direkten Weg zu den Lagrange Bedingungen für Aufgabe (4.0.2), ohne auf Zusammenhänge zu Subdifferential und Störungen einzugehen (siehe dazu Kapitel 5). Die Bedingungen charakterisieren optimalverdächtige Punkte analytisch. Sie ersetzen die notw. Bedingung $Df(x) = 0$ für freie Extrema.

We begin with the well-known case of constraints in equation form

$$\begin{array}{l} \min\{\ f(x)\ |\ x \in M\ \} \quad \text{where} \quad f, h \in C^1 \quad \text{and} \\ M = \{\ x \in \mathbb{R}^n\ |\ \quad h_j(x) = 0\ \forall j = 1, .., m\} \end{array} \tag{4.1.1}$$

Theorem 4.1.1 *Let $m \le n$ and rank $Dh(\bar{x}) = m$ at a solution $\bar{x}$. Then there are (unique) μ_j such that*

$$Df(\bar{x}) + \sum_j \mu_j Dh_j(\bar{x}) = 0. \tag{4.1.2}$$

Proof. Without loss of generality, assume that the first m columns of the (m,n)-matrix $Dh(\bar{x})$ are linearly independent and write $x = (x^1, x^2)$, $x^1 \in \mathbb{R}^m, x^2 \in \mathbb{R}^{n-m}$. Now we can define, in terms of partial derivatives,

$$\mu := -f_{x^1}(\bar{x})\ [h_{x^1}(\bar{x})]^{-1}. \tag{4.1.3}$$

Then

$$\sum_j \mu_j\ \frac{\partial h_j}{\partial x^1}(\bar{x}) = \mu\ h_{x^1}(\bar{x}) = -f_{x^1}(\bar{x}). \tag{4.1.4}$$

On the other hand, the implicit function theorem ensures, for points near $\bar{x}$, that equation $h(x^1, x^2) = 0$ defines a unique C^1- function $x^1 = \phi(x^2)$ such that $h(\phi(x^2), x^2) = 0$ and $D\phi(\bar{x}^2) = -[h_{x^1}(\bar{x})]^{-1} h_{x^2}(\bar{x})$. Hence $\bar{x}_2$ minimizes

$$F(x^2) := f(\ \phi(x^2),\ x^2\ ) \qquad \text{and}$$

$$0 = DF(\bar{x}_2) = f_{x^1}(\bar{x})\ D\phi(\bar{x}^2) + f_{x^2}(\bar{x}\ ) = -f_{x^1}(\bar{x})\ [h_{x^1}(\bar{x})]^{-1} h_{x^2}(\bar{x}) + f_{x^2}(\bar{x}\ ).$$

The latter ensures

$$\sum_j \mu_j\ \frac{\partial h_j}{\partial x^2}(\bar{x}) = \mu\ h_{x^2}(\bar{x}) = -f_{x^2}(\bar{x}) \tag{4.1.5}$$

after which (4.1.4) and (4.1.5) verify (4.1.2). □

Condition (4.1.2) says that $-Df(\bar{x})$ belongs to the normal cone of the manifold $h = 0$ at $\bar{x}$. In other words, the projection of $-Df(\bar{x})$ onto the tangent space of M at $\bar{x}$ has to vanish. In what follows, the optimality conditions have the same geometrical meaning. Differences arise only by the facts, that

1. the tangent space becomes a cone,
2. the Lagrange multipliers μ_j are non-negative for inequalities $h_j \le 0$ and
3. the regularity condition $rank\ Dh(\bar{x}) = m$ must be changed.

Now we consider the case of equations and inequalities in the constraints.

Theorem 4.1.2 *Sei $\bar{x}$ eine lokal optimale Lösung von (4.0.2). Erfüllt $\bar{x}$ eine gewisse* **"Regularitätsbedingung"**, *so gibt es Zahlen $\bar{y}_i$ und $\bar{z}_j$, dass gilt*

$$\begin{array}{l} Df(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i Dg_i(\bar{x}) + \sum_{j=1}^{m_h} \bar{z}_j Dh_j(\bar{x}) = 0, \\ h(\bar{x}) = 0, \quad g(\bar{x}) \le 0, \quad \bar{y} \ge 0, \quad \bar{y}_i\ g_i(\bar{x}) = 0\ \forall i. \quad \diamond \end{array} \tag{4.1.6}$$

Vor dem Beweis einige Bemerkungen:
Die Zahlen $\bar{y}_i, \bar{z}_j$ heissen wieder **Lagrange-Multiplikatoren**, die Bedingungen selbst Karush-Kuhn-Tucker Bedingungen. Der Punkt $(\bar{x}, \bar{y}, \bar{z})$ wird auch *KKT-Punkt* genannt. Die schon aus $\bar{x} \in M$ folgenden Bedingungen $h(\bar{x}) = 0, g(\bar{x}) \le 0$ nimmt man der Vollständigkeit halber mit auf.
Wir haben $n+m+m_h$ Gleichungen mit derselben Zahl reeller Variabler zu erfüllen. Dazu die Ungleichungen $y \ge 0$, $g(\bar{x}) \le 0$. Daher kann man hoffen, dass das System nicht mit jedem zulässigen $\bar{x}$ erfüllbar ist.
Die Bedingungen (4.1.6) kann man mit Hilfe der *Lagrange Funktion* zur Aufgabe (4.0.2),

Definition 4.1.1

$$L(x,y,z) = f(x) + \sum_{i}^{m} y_i\ g_i(x) + \sum_{j=1}^{m_h} z_j\ h_j(x) = f(x) + \langle y, g(x)\rangle + \langle z, h(x)\rangle \qquad \diamond$$

und ihrer partiellen Ableitungen in $(\bar{x}, \bar{y}, \bar{z})$ kompakter schreiben:

$$L_x = 0,\ L_z = 0, \quad L_y \le 0,\ y \ge 0,\ \langle y, L_y \rangle = 0. \tag{4.1.7}$$

Die Bedingungen $L_z = 0,\ L_y \le 0$ bedeuten $\bar{x} \in M$. Die drei letzten Bedingungen sagen: $(\bar{y}, \bar{z})$ *maximiert* $L(\bar{x}, y, z)$ bzgl. $y \in \mathbb{R}^{m+}, z \in \mathbb{R}^{m_h}$.

Hat man keine Ungleichungen, bleibt in (4.1.7) nur $L_x = 0$ und $L_z = 0$. Das ist im Sinne der Optimierungstheorie (mit nötiger Differenzierbarkeit) der uninteressante Fall von Theorem 4.1.1, weil - wie für freie Minima - nur ein Gradient zu Null gemacht werden muss, um optimalverdächtige Punkte zu finden. Auch die Regularitätsbedingung ist dann einfacher als jetzt.

Der konvexe Fall: *Sind alle f, g_i konvex und die h_j affine,* kann die erste Bedingung ebenfalls als Extremalbedingung interpretiert werden:

$\bar{x}$ minimiert $L(x, \bar{y}, \bar{z})$ bzgl. $x \in \mathbb{R}^n$.

Denn dann ist L in x konvex, wonach Minimalität aus $L_x(\bar{x}, \bar{y}, \bar{z}) = 0$ folgt. Das liefert insgesamt die *Sattelpunktbedingung*

$$L(\bar{x}, y, z) \le L(\bar{x}, \bar{y}, \bar{z}) \le L(x, \bar{y}, \bar{z}) \quad \forall\ x \in \mathbb{R}^n,\ y \in \mathbb{R}^{m+},\ z \in \mathbb{R}^{m_h}. \tag{4.1.8}$$

Die Sattelpunktbedingung ist so stark, dass aus ihr die Optimalität von $\bar{x}$ auch ohne Konvexität oder Stetigkeit der beteiligten Funktionen folgt; denn die linke Ungl. liefert Zulässigkeit, die rechte dann Optimalität von $\bar{x}$.

Für lineare Aufgaben kann man Thm. 4.1.2 mittels LP duality Thm. 2.1.5 (ohne zusätzliche Regularität) beweisen; UebAufg.

The general case: In (4.1.6) sind offenbar nur die Ungleichungen zu $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$ von Interesse, weil sonst einfach $\bar{y}_i = 0$ verlangt wird.

Definition 4.1.2 *Eine Regularitätsbedingung (= Constraint Qualification, CQ) in $\bar{x} \in M$ ist eine Bedingung, die sichert, dass aus*

$$Dh(\bar{x})u = 0 \quad \textit{and} \quad Dg_i(\bar{x})u \le 0\ \forall i \in I(\bar{x}) \tag{4.1.9}$$

folgt: Es gibt Punke $x_k = \bar{x} + t_k u + o(t_k)$, die $h(x_k) = 0$ und $g_i(x_k) \le 0\ \forall i \in I(\bar{x})$ mit $\lim \frac{o(t_k)}{t_k} \to 0$ für gewisse $t_k \downarrow 0$ erfüllen. ◇

Für $i \notin I(\bar{x})$ gilt $g_i(\bar{x}) < 0$. Das bleibt (Stetigkeit) richtig für x_k und kleine $t_k > 0$. Also folgt - unter Regularität - nun sogar $x_k \in M$ (für gewisse $t_k \downarrow 0$). Das bedeutet:
CQ is äquivalent zu

$$\liminf_{t \downarrow 0}\ t^{-1}\ \mathrm{dist}\,(\bar{x} + tu, M) = 0 \quad \text{bzw.} \quad u \in C_M(\bar{x}) \qquad \forall\ u \text{ in (4.1.9)}. \tag{4.1.10}$$

s. contingent cone C_M, Def. 3.4.2. Umgekehrt, CQ ist genau dann verletzt, wenn

$$\mathrm{dist}\,(\bar{x} + tu, M) \ge \delta t \ \text{ gilt für ein } u \text{ aus (4.1.9), ein } \delta > 0 \text{ und alle kleinen } t > 0. \tag{4.1.11}$$

Dann gilt auch noch (für kleine $t > 0$)

$$\mathrm{dist}\,(\bar{x} + tw + o(t),\ M) \ge \tfrac{1}{2}\delta t \ \text{ für alle } w \text{ mit } \|w - u\| < \frac{\delta}{2}. \tag{4.1.12}$$

Proof. (Thm. 4.1.2) Man betrachtet das durch y_0 ergänzte System (4.1.6); es heisst Fritz-John-System:

$$y_0 Df(\bar{x}) + \sum_{i=1}^{m} y_i Dg_i(\bar{x}) + \sum_{j=1}^{m_h} z_j Dh_j(\bar{x}) = 0, \quad y_0, y_i \geq 0, \quad y_i\ g_i(\bar{x}) = 0\ \forall i > 0. \quad (4.1.13)$$

Es ist stets lösbar ($y = 0, z = 0$). Gibt es eine Lösung mit $y_0 > 0$, kann man durch y_0 teilen und erhält eine Lösung von (4.1.6), andernfalls nicht. Also gelte stets $y_0 = 0$. O.B.d.A. sei $I(\bar{x}) = \{1, ..., p\}$. System (4.1.13) bekommt so die Form

$$y_0 Df(\bar{x}) + \sum_{j=1}^{m_h} z_j Dh_j(\bar{x}) + \sum_{i=1}^{p} y_i Dg_i(\bar{x}) = 0, \quad y_i \geq 0\ \forall i:\ 0 \leq i \leq p \quad (4.1.14)$$

und die LO-Aufgabe $\max\{\ y_0 \mid\ y$ erfüllt (4.1.14) $\}$ hat den Optimalwert Null. Dasselbe gilt nach Thm. 2.1.5 für die lösbare Dualaufgabe. Sie hat die Form

$$(D) \quad \min\{\ 0^T v \mid Df(\bar{x})v \geq 1,\ Dh(\bar{x})v = 0,\ Dg_i(\bar{x})v \geq 0\ \forall i \in I(\bar{x});\quad v \in \mathbb{R}^n\}.$$

Also ist ein v zulässig für (D). Mit $u = -v$ folgt so

$$Df(\bar{x})u \leq -1, \quad Dh(\bar{x})u = 0,\ Dg_i(\bar{x})u \leq 0\ \forall i \in I(\bar{x}). \quad (4.1.15)$$

Wegen

$$f(\bar{x} + tu) = f(\bar{x}) + tDf(\bar{x})u + r(t), \quad \text{mit } \frac{r(t)}{t} \to 0 \text{ wenn } t \downarrow 0,$$

gilt dann $f(\bar{x} + tu) < f(\bar{x}) + \frac{1}{2}tDf(\bar{x})u < f(\bar{x})$ für kleine $t > 0$. Das bleibt richtig für Punkte x_k der Form

$$x_k = \bar{x} + t_k u + o(t_k), \quad t_k \downarrow 0. \quad (4.1.16)$$

Wegen Regularität sind aber gewisse x_k dieser Form zulässig, was der lokalen Optimalität von $\bar{x}$ widerspricht und so den Satz beweist; es ist $y_0 > 0$ für eine Lösung von (4.1.13). □

Forderung (4.1.10) heisst auch Abadie CQ. Oft kann man hier lim inf durch lim ersetzen.

Corollary 4.1.3 *(C_M,$\mathcal{N}_M$ and CQ)*
(i) The Constraint Qualification (4.1.10) requires, for M in (4.0.2), nothing else but the validity of formula (3.5.18) in section 3.5.1, i.e.,

$$C_M(\bar{x}) = \{u \mid Dh(\bar{x})u = 0,\ Dg_i(\bar{x})u \leq 0\ \forall i \in I(\bar{x})\}. \quad (4.1.17)$$

(ii) In this case, the KKT conditions (4.1.6) obviously coincide with $0 \in Df(\bar{x}) + \mathcal{N}_M(\bar{x})$; compare with (3.5.4), Lemma 3.5.1. ◇

Proof. Recall that $\mathcal{N}_M = (C_M)^*$ was defined in Def. 3.4.2.
(i) Because of (4.1.10), only $u \in C_M(\bar{x}) \Rightarrow$ (4.1.9) remains to show. This follows from the mean-value theorem since $u \in C_M(\bar{x})$ implies, with certain $t_k \downarrow 0$,

$$h(\bar{x} + t_k u) - h(\bar{x}) = o(t_k) \quad \text{and} \quad g_i(\bar{x} + t_k u) - g_i(\bar{x}) \leq o_i(t_k)\ \forall i \in I(\bar{x}). \quad (4.1.18)$$

Apply $t_k Dg_i(\bar{x} + \theta_{i,k}u)u = g_i(\bar{x} + t_k u) - g_i(\bar{x})$ for some $\theta_{i,k} \in (0, t_k)$; similarly for h_ν. Division by t_k and passing to the limit then yields the assertion $Dh(\bar{x})u = 0$ and $Dg_i(\bar{x})u \leq 0\ \forall i \in I(\bar{x})$.
(ii) The analytical form (3.5.19) sect. 3.5.4, of $\mathcal{N}_M$, namely

$$\mathcal{N}_M(\bar{x}) = \{x^* \in \mathbb{R}^n \mid x^* = Dh(\bar{x})^T z + \sum_{i \in I(\bar{x})} y_i Dg_i(\bar{x}),\quad y \geq 0\}. \quad (4.1.19)$$

then follows from (i) and LOP duality Thm. 2.1.5. □

Remark 4.1.4 (Linearisierung in $\bar{x}$) Bedingung (4.1.15) - also Nichtexistenz von Lagrange-Multiplikatoren zu $\bar{x}$ - gilt offenbar genau dann, wenn $u = 0$ *nicht* die linearisierten Originalaufgabe

$$\min_u \{Df(\bar{x})u \mid h(\bar{x}) + Dh(\bar{x})u = 0,\ g(\bar{x}) + Dg(\bar{x})u \leq 0\} \tag{4.1.20}$$

löst. Genau das soll also eine CQ für lokale Lösungen $\bar{x}$ von (4.0.2) ausschliessen. ◇

Example 4.1.1 The problem min $\{x \in \mathbb{R} \mid x^2 \leq 0\}$ shows: Even for convex problems, Thm. 4.1.2 may fail to hold without CQ. ◇

4.1.2 Concrete Constraint Qualifications

The role of all these conditions consists in ensuring (4.1.17) which allows us to pass from the geometric conditions of Lemma 3.5.1 to the analytic KKT-formulation Thm. 4.1.2. Next we show how regularity, Def. 4.1.2, can be satisfied.

(i) Linear problems: All h_j, g_i are (affin) linear. ◇

Then one can directly put $x(t) = \bar{x} + tu$ in Def. 4.1.2. □

(ii) Convex problems with Slater points: h affin, all g_i convex and $\exists x^S \in M$ such that $g_i(x^S) < 0\ \forall i$. ◇

Put $u^0 = x^S - \bar{x}$. Then $Dh(\bar{x})u^0 = 0$ and, for $i \in I(\bar{x})$, also $Dg_i(\bar{x})u^0 < 0$ holds true, since for small $t > 0$, we have

$$g_i(\bar{x} + tu^0) = g_i((1-t)\bar{x} + tx^S) \leq (1-t)0 + tg_i(x^S) \text{ and}$$

$$Dg_i(\bar{x})u^0 = \lim_{t \downarrow 0} t^{-1}[g_i(\bar{x} + tu^0) - g_i(\bar{x})] \leq g_i(x^S) < 0.$$

Now assume (4.1.11). Then, with small $\varepsilon > 0$ and for δ from (4.1.11), it follows $\|v - u\| < \delta/2$ with $v = u + \varepsilon u^0$. Because of $Dg_i(\bar{x})u \leq 0$, we also obtain $Dg_i(\bar{x})v < 0$, thus $\bar{x} + tv \in M$ for small $t > 0$. In consequence, (4.1.11) cannot be true. □

The condition "$\exists x^S \in M:\ g(x^S) < 0$" is called *Slater Bedingung, Slater condition* [138]. For convex g_i and if no equations are required, this condition coincides with MFCQ ($u^0 = x^S - \bar{x}$).

(iii) MFCQ: The next condition from Mangasarian and Fromovitz [102], also called MFCQ, is one of the most important constraint qualifications:

$$rank\ Dh(\bar{x}) = m_h \text{ and } \exists\ u^0:\ Dh(\bar{x})u^0 = 0 \text{ and } Dg_i(\bar{x})u^0 < 0\ \forall i \in I(\bar{x}). \quad ◇$$

Beweis, dass MFCQ eine CQ ist: Wir nehmen das Gegenteil an. Dann gilt (4.1.11). Mit entsprechenden u und δ, sei wie oben $v = u + \varepsilon u^0$ mit kleinem $\varepsilon > 0$ gebildet, so dass $\|v - u\| < \frac{\delta}{2}$. Wir konstruieren Punkte $x(t) = \bar{x} + tv + o(t) \in M$. Beachte dazu, dass

$$Dg_i(\bar{x})v < 0\ \forall i \in I(\bar{x}) \text{ and } Dh(\bar{x})v = 0$$

und studiere die Lösungen $\xi = \xi(t)$ des Systems

$$h(\bar{x} + tv + \xi) = 0, \quad \langle c^\mu, \xi \rangle = 0\ (\mu = m_h + 1, ..., n) \tag{4.1.21}$$

wobei $n - m_h$ Vektoren $c^\mu \in \mathbb{R}^n$ so gewählt seien, dass die Matrix A, bestehend aus allen c^μ und den m_h (linear unabhängigen !) Gradienten $Dh_j(\bar{x}) \in \mathbb{R}^n$ vollen Rang besitzt. Das System (4.1.21) hat die Form

$$F(t, \xi) = 0 \ (\in \mathbb{R}^n) \quad t \in \mathbb{R}, \ \xi \in \mathbb{R}^n \tag{4.1.22}$$

und genügt in $(0, 0)$ den Voraussetzungen des Satzes über implizite Funktionen: $F(0, 0) = 0$, $D_\xi F(0, 0) = A$ ist regulär. Damit existieren (für kleine $|t|$ und nahe $0 \in \mathbb{R}^n$) eindeutige Lösugen $\xi = \xi(t)$ von (4.1.21), wobei $\xi(.)$ die Ableitung

$$\frac{d\xi(0)}{dt} = -D_\xi F(0, 0)^{-1} D_t F(0, 0)$$

besitzt. Mit den Funktionen in (4.1.21) gilt $D_t F(0, 0) = 0$, denn $Dh(\bar{x})v = 0$, und c^μ ist unabhängig von t. Also ist

$$\frac{d\xi(0)}{dt} = 0, \quad d.h., \quad \frac{\|\xi(t)\|}{t} \to 0 \ \text{ if } t \to 0.$$

Wir wissen nun, dass $x(t) = \bar{x} + tv + \xi(t)$ die Gleichungen erfüllt. Für $i \in I(\bar{x})$ gilt $g_i(\bar{x}) = 0$ und $Dg_i(\bar{x})v < 0$; also folgt (für kleine $t > 0$) auch $g_i(x(t)) < 0$. Für $i \in \{1, .., m\} \setminus I(\bar{x})$ folgt $g_i(x(t)) < 0$ aus $g_i(\bar{x}) < 0$ und Stetigkeit. Damit ist $x(t) \in M$, im Widerspruch zur Wahl von v (4.1.11). □

(iv) LICQ (Linear Independence Constraint Qualification): The gradients $Dh_j(\bar{x})$ and $Dg_i(\bar{x}), \ i \in I(\bar{x})$ form a linearly independent system. ◇

Then $Dh(\bar{x})$ has full rank and the system $Dh(\bar{x})v = 0, \ Dg_i(\bar{x})v = -1 \ \forall i \in I(\bar{x})$ has a solution u^0. Therefore, MFCQ is satisfied. □

(v) Calmness: Man betrachte M unter kleinen Störungen p der rechten Seiten,

$$S(p) := g_{\leq}(b) \cap h_{=}(c), \quad p = (b, c); \quad S(0) = M. \tag{4.1.23}$$

Definition 4.1.3 S heisst *calm in* $(0, \bar{x})$, wenn es $\varepsilon, \delta, L > 0$ gibt, so dass $\forall p \in B(0, \delta)$ und $x \in B(\bar{x}, \varepsilon) \cap S(p)$ gilt: $\text{dist}\,(x, \ S(0)) \leq L \ \|p\|$.

Auch calmness ist eine Constraint Qualification. Für u aus (4.1.9) folgt nämlich (4.1.18) (mit bel. $t_k \downarrow 0$). Damit gilt $\bar{x} + tu \in S(p)$ mit einem p der Grösse $o(t)$, und calmness liefert, wie in (4.1.10) verlangt, $\text{dist}\,(\bar{x}+tu, \ S(0)) \leq L \ \|p\| \leq L \ o(t)$. □

Calmness drückt zugleich ein stabiles Verhalten allgemeiner Abbildungen S aus. Eine Charakterisierung erfolgt in section 10.7, speziell Thm. 10.7.6.

(vi) Weaker conditions: Sometimes, the structure of the constraints can be directly used. (1) If h is affine, then replace $\mathbb{R}^n$ by $h_{=}(0)$ to weaken MFCQ (no rank condition).

(2) If h is continuous and piecewise affine then $h_{=}(0)$ is a *union* of a finite number of polyhedrons P_μ (described by affine systems $A_\mu x \leq b_\mu$). The problems/values

$$(Pr)_\mu \qquad \min \ \{ \ f(x) \mid x \in P_\mu, \ g(x) \leq 0 \ \}$$

allow again a reduction to simpler problems since

Remark 4.1.5 $\bar{x}$ is optimal for (4.0.2) $\Leftrightarrow$ $\bar{x}$ is optimal $\forall \ (Pr)_\mu$ with $\bar{x} \in P_\mu$. ◇

So it suffices to study the problems $(Pr)_\mu$ and their optimality conditions separately.

(3) Similarly, subsystems of piecewise linear g_i and h_ν can be handled by studying

$$(Pr)_\mu \qquad v_\mu = \inf \ \{ \ f(x) \mid x \in P_\mu, \ \ h_{j'}(x) = 0 \ \forall j', \ g_{i'}(x) \leq 0 \ \forall i' \ \}$$

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

4.1.3 Calmness, MFCQ and Aubin-Property

Ist MFCQ erfüllt, kann man ein in Def. 4.1.3 gesuchtes $x' \in S(0)$ mit $d(x', x) \leq L\|p\|$ per Satz über implizite Funktionen konstruieren (mit einer Modifikation des Beweises unter (iii)). Calmness fordert daher nicht mehr als MFCQ. Tatsächlich ist calmness die schwächste der hier genannten Bedingungen. Calmness linearer Systeme (i) wurde zuerst in [50] gezeigt (Hoffman's Lemma) und ist eine gute Übungsaufgabe.

MFCQ, Aubin property and bounded LM
Condition MFCQ does not only guarantee the existence of Lagrange multipliers for local minimizers. It also ensures that the behavior of the mapping S (4.1.23) near $(0, \bar{x})$ is locally the same as for hyperplanes $H(r) := \{x \mid \langle c, x\rangle = r\}$:

$$\begin{array}{c}\text{There is some } L \text{ such that, given } p' \text{ and } x \in S(p)\\ (x \text{ close to } \bar{x},\ p, p' \text{ close to } 0),\ \text{ it holds } \ \operatorname{dist}(x,\ S(p')) \leq\ L\ \|p' - p\|.\end{array}$$

MFCQ is even equivalent (for $g, h \in C^1$) to this property, called Aubin property of S, cf. section 10.1. In addition, MFCQ at a local minimizer $\bar{x}$ of problem (4.0.2) also implies that the (non-empty) set of all Lagrange multipliers is bounded.

Theorem 4.1.6 *(Gauvin) [38] Let $\bar{x}$ be a local minimizer of problem (4.0.2). Then, MFCQ holds at $\bar{x}$ $\Leftrightarrow$ the set of Lagrange multipliers*

$$\Lambda(\bar{x}) = \{(y, z) \mid (\bar{x}, y, z) \text{ is a KKT point}\}$$

is nonempty and bounded. ◇

Proof. $\Lambda(\bar{x})$ is a polyhedron. If it contains an unbounded sequence λ^k then $\bar{\lambda} = \lim \frac{\lambda^k}{\|\lambda^k\|}$ exists for some subsequence. With $\bar{\lambda} = (\bar{y}, \bar{z})$ then

$$\bar{z}^T Dh(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{y}_i Dg_i(\bar{x}) = 0, \quad (\bar{y}, \bar{z}) \neq (0, 0), \quad \bar{y} \geq 0 \text{ and } \bar{y}_i = 0 \text{ if } i \notin I(\bar{x}) \tag{4.1.24}$$

follows. Conversely, if $\Lambda(\bar{x}) \neq \emptyset$ and (4.1.24) has a solution, then $\Lambda(\bar{x})$ is unbounded since $\lambda + t\bar{\lambda} \in \Lambda(\bar{x})$ for $\lambda \in \Lambda(\bar{x})$ and all $t > 0$.

Now suppose MFCQ. We already know that $\Lambda(\bar{x}) \neq \emptyset$. If $\Lambda(\bar{x})$ is unbounded then multiplying the first equation in (4.1.24) with the MFCQ-direction u^0 yields first $\bar{y} = 0$. Due to $\bar{\lambda} \neq 0$ so $\bar{z} \neq 0$ has to hold. But then $Dh(\bar{x})$ has no full rank since $\bar{z}^T Dh(\bar{x}) = 0$. Thus ($\Rightarrow$) is valid.

Conversely, let $\Lambda(\bar{x}) \neq \emptyset$ be bounded. Then (4.1.24) is unsolvable. Since (4.1.24) is unsolvable with $\bar{y} = 0, \bar{z} \neq 0$, so $Dh(\bar{x})$ has full rank. Moreover, since (4.1.24) is unsolvable, also the LP problem

$$\max\{\sum_{i \in I(\bar{x})} y_i \mid z^T Dh(\bar{x}) + \sum_{i \in I(\bar{x})} y_i Dg_i(\bar{x}) = 0,\ y \geq 0 \text{ and } y_i = 0 \text{ if } i \notin I(\bar{x})\}$$

is solvable with max = 0. Then, the existence of a MFCQ vector u^0 follows from LP-duality. Thus ($\Leftarrow$) is valid, too. □

Remark 4.1.7 (Persistence of MFCQ) Having MFCQ at $\bar{x}$, one easily obtains MFCQ for points x near $\bar{x}$. Even if h and g are slightly modified within C^1, MFCQ is still valid and the related LM's (if they exist for x) remain uniformly bounded.

4.1.4 Calmness and Complementarity

Die Bedingung

$$x_1 \geq 0,\ x_2 \geq 0,\ x_1 x_2 = 0$$

heisst Komplementaritätsbedingung (auch bei anderen Vorzeichenbedingungen). Für das Problem

$$(Pr) \quad \min \{\ f(x) \mid x_1 \leq 0,\ x_2 \leq 0,\ x_1 x_2 = 0\}$$

ist die (complementarity-) Abbildung

$$S(p) = \{x \in \mathbb{R}^2 \mid x_1 \leq p_1,\ x_2 \leq p_2,\ \ x_1 x_2 = p_3\}$$

nicht calm im Ursprung $\bar{x} = 0$; wähle $x = (-\varepsilon, -\varepsilon)$ und $p = (0, 0, \varepsilon^2)$.

Doch $M = S(0)$, bestehend aus $\bar{x}$ und den beiden negativen Halbachsen, kann alternativ beschrieben werden, z.B. als $S(0) = \{x \in \mathbb{R}^2 \mid h(x) := \max\{x_1, x_2\} = 0\}$. Nun liegt Situation (vi)(2) aus 4.1.2 vor. Man betrachte die Probleme $(Pr)_1, (Pr)_2$ mit den Restriktionen

$$M_1 = \{x \mid \ x_1 = 0, x_2 \leq 0\} \ \text{ bzw. } M_2 = \{x \mid \ x_1 \leq 0, x_2 = 0\}$$

(und derselben Zielfunkt. f). Analog gehe man vor, wenn es weitere Variable und Nebenbed. gibt; dann werden in M_1 und M_2 auch alle diese Bedingungen aufgenommen. Natürlich darf man hier vereinfachen, indem man $x_1 = 0$ bzw. $x_2 = 0$ direkt einsetzt.

Offenbar ist $\bar{x}$ optimal bzgl. $M = M_1 \cup M_2$ genau dann, wenn $\bar{x}$ optimal bzgl. M_1 und M_2 ist. Also muss ein optimales $\bar{x}$ zwar nicht immer die KKT-Bedingungen der Originalaufgabe erfüllen, aber beide KKT Systeme zu $(Pr)_1, (Pr)_2$, wenn letztere "regulär" sind. Speziell also, *wenn die Teilabbildungen S_1 und S_2 (Variation der rechten Seiten) zu M_1 und M_2 in $(0, \bar{x})$ calm sind.* Letzteres ist z.B. erfüllt, wenn auch die restlichen Bedingungen (stückw.) affin sind.

Example 4.1.2 Die Aufgabe $\min\{x_1 \mid x_1 \leq 0,\ x_2 \leq 0,\ x_1 x_2 = 0,\ \ x_1 - x_2 = 0\}$ mit der Lösung $\bar{x} = 0$ und dem zu $\bar{x}$ gehörenden KKT System

$$1 + y_1 + z_1 x_2 + z_2 = 0, \qquad 0 + y_2 + z_1 x_1 - z_2 = 0, \qquad y \geq 0$$

sei zu behandeln. Die Abbildung der gestörten Restriktionen

$$S(p) = \{x \mid x_1 \leq p_1,\ x_2 \leq p_2,\ x_1 x_2 = p_3,\ x_1 - x_2 = p_4\}$$

ist -wie oben- *nicht calm* in $(0, \bar{x})$. Wir wissen also nicht, ob ein Lagrange Vektor zur Lösung $\bar{x}$ existiert. In der Tat gibt es keinen, weil sonst

$$1 + y_1 = -z_2 = -y_2 \ \text{ mit } \ y \geq 0$$

folgen würde, was offenbar unmöglich ist. Man schreibe nun $M = S(0)$ als $M_1 \cup M_2$:

$$M_1 = \{x \mid x_1 = 0, x_2 \leq 0,\ x_1 - x_2 = 0\},\ M_2 = \{x \mid x_1 \leq 0, x_2 = 0,\ x_1 - x_2 = 0\}.$$

Offenbar löst $\bar{x}$ sowohl $(Pr)_1 : \min\{x_1 \mid x \in M_1\}$ als auch $(Pr)_2 : \min\{x_1 \mid x \in M_2\}$. Die beiden Abbildungen zu M_1 und M_2 (hier umständlich aufgeschrieben ohne $x_1 = 0$ bzw. $x_2 = 0$ direkt zu benutzen),

$$S_1(p) = \{x \mid x_1 = p_1,\ x_2 \leq p_2,\ x_1 - x_2 = p_3\},$$
$$S_2(p) = \{x \mid x_1 \leq p_1,\ x_2 = p_2,\ x_1 - x_2 = p_3\}$$

sind calm. Damit müssen die zwei KKT Systeme

$$\begin{array}{ll} zu\ (Pr)_1: & 1+z_1+z_2=0, \quad 0+y_2-z_2=0, \\ zu\ (Pr)_2: & 1+y_1'+z_2'=0, \quad 0+z_1'-z_2'=0. \end{array}$$

beide lösbar sein (was sie offenbar auch sind). Umgekehrt, sind mit einem $\bar{x} \in M$ die KKT-Systeme zu $(Pr)_1, (Pr)_2$ lösbar und die Aufgaben $(Pr)_1, (Pr)_2$ konvex, so löst $\bar{x}$ beide Aufgaben und somit auch die Originalaufgabe. $\diamond$

Gibt es weitere komplementäre Paare, etwa x_3, x_4 mit $\bar{x}_3 = \bar{x}_4 = 0$, hat man es analog mit 4 KKT Systemen und Abbildungen S_k zu tun usw. An der prinzipiellen Vorgehensweise ändert sich nichts. Die beschriebene Situation tritt auf, wenn KKT Punkte in die Nebenbedingungen anderer Aufgaben einfliessen. Mit

$$\xi_i = -y_i,\ \eta_i = g_i(x) \ \text{ und } \ \xi_i \le 0,\ \eta_i \le 0,\ \xi\eta = 0$$

wird dann offenbar Komplementarität zwischen y_i und g_i beschrieben.

Thus
complementarity conditions among the constraints do *not* require any new theory of local optimality conditions. It increases, however, the number of subsystems which must be studied at points z which violate strict complementarity. This drawback (from the practical point of view) initiated several other descriptions of the related systems, in particular by using so-called NCP functions; we refer to [28], [140] and the cited literature therein.
In brief, though weaker than MFCQ, calmness of S (4.1.23) may be a (too) strong condition for ensuring the existence of Lagrange multipliers to (4.0.2). This fact reduces the meaning of calmness for this purpose.
Nevertheless, calmness does not only describe useful error estimates for $\operatorname{dist}(x, S(0))$ for points $x \in S(p)$ near $\bar{x}$. It is also closely connected with methods for solving the systems in question, see below.

Exercises
1) Quadratische Optimierung und Komplementarität: Es sei $\bar{x}$ globaler Minimalpunkt der Aufgabe

$$\min\{p^T x + \tfrac{1}{2} x^T Q x \mid Ax \le b\} \ \text{ mit } Q^T = Q.$$

Zeigen Sie, dass es dann gewisse y^0 und u^0 (entsprechender Dimension) gibt, so dass $(\bar{x}, y^0, u^0)$ global die Aufgabe (genannt linear complementarity problem)

$$\min\{p^T x - b^T y \mid \quad Qx + A^T y = -p,\ Ax + u = b,\ y \ge 0,\ u \ge 0, \ \ u^T y = 0\} \quad \text{löst.}$$

2) Diskutieren Sie $\min\{\ ax^2 + by^2 \mid x + 2y \ge 1\}$ für alle reellen a, b.
3) Calmness and Slaterpunkt: Zeigen Sie, dass calmness für konvexe Probleme mit Slaterpunkt (ii) gilt.

Hilfe: Ordnen Sie jedem $x \ne M$ (mit $\|x - \bar{x}\| \le 1$) zunächt $x' = \pi_{h^{-1}(0)}(x)$ als Projektion zu und dann $x'' = \lambda x^S + (1-\lambda)x' \in M$ mit minimalem $\lambda \in [0,1]$. Es bleibt, $\|x'' - x\|$ abzuschätzen. Differenzierbarkeit von g_i wird nicht gebraucht.

4.2 The standard second order condition

Let $\bar{s} = (\bar{x}, \bar{y}, \bar{\mu})$ be a KKT point for problem (4.0.2). We are interested in a condition which ensures that $\bar{x}$ is a local minimizer for (4.0.2). With the set $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$, define the tangent cone

$$V = \{v \in \mathbb{R}^n \mid Df(\bar{x})v = 0,\ Dh(\bar{x})v = 0,\ Dg_i(\bar{x})v \le 0 \text{ if } i \in I(\bar{x})\}. \tag{4.2.1}$$

Theorem 4.2.1 *Let $\bar{s}$ be a KKT point for (4.0.2) and*

$$\langle u,\ D_x^2 L(\bar{s})u\rangle > c \quad \forall u \in V,\ \|u\| = 1. \tag{4.2.2}$$

Then

$$f(x) \geq f(\bar{x}) + \tfrac{1}{2}\ c\ \|x-\bar{x}\|^2 \qquad \forall x \in M \cap B(\bar{x},\delta)\ \textit{and}\ \ \textit{some}\ \delta > 0. \quad \diamond \tag{4.2.3}$$

The *sufficient second-order condition (SSOC)* requires that (4.2.2) holds for some $c > 0$.

Proof. Since $\bar{s}$ is a KKT point, we have $Df(\bar{x})v + \sum_i \bar{y}_i Dg_i(\bar{x})v + \sum \bar{z}_j Dh_j(\bar{x})v = 0$. For this reason, $v \in V$ yields that $Dg_i(\bar{x})v = 0$ if $\bar{y}_i > 0$.

Next we delete the equations $h_j = 0$ for sake of simplicity. It will be seen that they play the same role as an inequality with $\bar{y}_i > 0$. We also write $\bar{y}Dg(\bar{x})v$ for $\langle \bar{y}, Dg(\bar{x})v\rangle$.

Assume the theorem fails to hold. Then there is some sequence $x_k \to \bar{x}$ such that

$$x_k \in M \text{ and } f(x_k) < f(\bar{x}) + \tfrac{1}{2}\ c\ \|x_k - \bar{x}\|^2. \tag{4.2.4}$$

Write $x_k = \bar{x} + t_k u_k$, such that $\|u_k\| = 1,\ t_k \downarrow 0$. Without loss of generality, we may assume that $u_k \to v$ (in V or not). Notice that we applied limiting negation of (4.2.3), cf. sect. 1.2.

Knowing that x, t and u depend on k, let us omit this index in the following. By second order expansion of f (and the KKT-property), it holds

$$\begin{array}{l} f(x) - f(\bar{x}) \\ =\quad t\ \ Df(\bar{x})u \quad +\ \tfrac{1}{2}\ t^2\ u^T D^2 f(\bar{x}+\theta\ tu)u \\ =\ -t\ \bar{y}Dg(\bar{x})u \quad +\ \tfrac{1}{2}\ t^2\ u^T D^2 f(\bar{x}+\theta\ tu)u \quad \text{with some } \theta = \theta(t) \in (0,1). \end{array} \tag{4.2.5}$$

Feasibility of x yields, by the mean value theorem:

If $g_i(\bar{x}) = 0$ then $g_i(\bar{x}+tu) \leq g_i(\bar{x})$ and, passing to the limit, $Dg_i(\bar{x})v \leq 0$. (For h, it follows similarly $Dh(\bar{x})v = 0$.)

If $g_i(\bar{x}) < 0$ then $\bar{y}_i = 0$.

Hence $r := \bar{y}Dg(\bar{x})v \leq 0$. Furthermore, $r = -Df(\bar{x})v$ holds for each KKT-point.
Next consider the two possible cases.

Case 1: $r = \bar{y}Dg(\bar{x})v < 0$. Since $u^T D^2 f(\bar{x}+\theta\ tu)u$ remains bounded, we obtain from (4.2.5) for small t, $f(x) - f(\bar{x}) > -\frac{1}{2}\ t\ r > \frac{1}{2}\ c\ t^2$. This contradicts (4.2.4).

Case 2: $r = \bar{y}Dg(\bar{x})v = 0$. Then $Df(\bar{x})v = -r$ vanishes, too. Hence $Df(\bar{x})v = 0$ holds true and yields $v \in V$.
Because of $\langle \bar{y}, g(x)\rangle \leq 0$ and $D_x L(\bar{x},\bar{y}) = 0$, it holds with certain $\theta = \theta(t) \in (0,1)$, the crucial estimate in terms of the Lagrangian

$$f(x) \geq f(x) + \langle \bar{y}, g(x)\rangle = L(x,\bar{y}) = L(\bar{x},\bar{y}) + \tfrac{1}{2}\ t^2\ u^T D_x^2 L(\bar{x}+\theta tu, \bar{y})u.$$

Finally, we have $L(\bar{x},\bar{y}) = f(\bar{x})$ and $u^T D_x^2 L(\bar{x}+\theta tu,\bar{y})u \ \to\ v^T D_x^2 L(\bar{x},\bar{y})v > c$. Thus, for small t, we obtain again a contradiction to (4.2.4), namely

$$f(x) \geq L(x,\bar{y}) = f(\bar{x}) + \frac{c}{2}t^2.$$

□

For problems in Banach spaces, the situation is less obvious (strong convergence $u_k \to v$ cannot be ensured), we refer to [10]. Notice, that the question of second order conditions for classical problems of variational calculus, cf. sect. 1.3, 4 leads to the Legendre-Jacobi conditions in terms of linear second-order differential equations.

4.3 Eine Anwendung: Brechungsgesetz

Seien 2 Punkte (a,b) und (A,B) in der Ebene gegeben und links bzw. rechts der Geraden $x+y=1$ gelegen. Im linken Teil möge sich das Licht mit der Geschwindigkeit v_1 ausbreiten können, im rechten (ein anderes Medium) mit Geschwindigkeit v_2. Wie sieht der optimale (schnellste) Weg des Lichtes von (a,b) nach (A,B) aus ?
Wir denken uns zwei Strecken von (a,b) nach (x,y) (auf der Geraden) und von (x,y) nach (A,B). Ihre Längen sind

$$l_1 = \sqrt{(x-a)^2+(y-b)^2} \quad , \quad l_2 = \sqrt{(x-A)^2+(y-B)^2}.$$

Die erforderlichen Zeiten werden so $T_1 = \frac{l_1}{v_1}$, $T_2 = \frac{l_2}{v_2}$. Also ist zu lösen

$$\min\ f(x,y) := \frac{l_1}{v_1} + \frac{l_2}{v_2} \text{ unter der Nebenbeding. } h(x,y) := x+y-1 = 0.$$

Angenommen (x,y) ist optimal gewählt. Dann sind die KKT-Bedingungen mit einem Lagrange-Multiplikator μ erfüllt, weil h affin-linear ist. Also gilt mit Ableitungen in (x,y)

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \mu\frac{\partial h}{\partial x} = 0, \qquad \frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} + \mu\frac{\partial h}{\partial y} = 0, \qquad \text{d.h.}$$

$$\frac{\partial L}{\partial x} = \frac{x-a}{v_1 l_1} + \frac{x-A}{v_2 l_2} + \mu = 0, \qquad \frac{\partial L}{\partial y} = \frac{y-b}{v_1 l_1} + \frac{y-B}{v_2 l_2} + \mu = 0 \tag{4.3.1}$$

und somit auch

$$\frac{x-a}{v_1 l_1} + \frac{x-A}{v_2 l_2} = \frac{y-b}{v_1 l_1} + \frac{y-B}{v_2 l_2}. \tag{4.3.2}$$

Nimmt man ϕ als Winkel zwischen den Strecken $(a,b)(x,y)$ und $(a,b)(x,b)$, wird $\frac{x-a}{l_1} = \cos\phi$, $\frac{y-b}{l_1} = \sin\phi$. Mit dem analogen Winkel ψ zwischen $(A,B)(x,y)$ und $(A,B)(x,B)$, wird $\frac{A-x}{l_2} = \cos\psi$, $\frac{B-y}{l_2} = \sin\psi$ (siehe Bild aus Vorlesung). Dann gilt also auch

$$\frac{\cos\phi}{v_1} - \frac{\cos\psi}{v_2} = \frac{\sin\phi}{v_1} - \frac{\sin\psi}{v_2}, \qquad \frac{1}{v_1}(\cos\phi - \sin\phi) = \frac{1}{v_2}(\cos\psi - \sin\psi).$$

Setzt man hier $\gamma = \frac{\pi}{4}$, $\phi = \gamma + \phi'$ und $\psi = \gamma + \psi'$, vereinfacht sich die erste Differenz mittels $\cos\gamma = \sin\gamma = \frac{\sqrt{2}}{2}$ und der bekannten Additionstheoreme

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta, \quad \sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta :$$

$$\cos\phi - \sin\phi = \frac{\sqrt{2}}{2}\cos\phi' - \frac{\sqrt{2}}{2}\sin\phi' - \frac{\sqrt{2}}{2}\cos\phi' - \frac{\sqrt{2}}{2}\sin\phi' = -\sqrt{2}\sin\phi'.$$

Dasselbe für ψ liefert $\cos\psi - \sin\psi = -\sqrt{2}\sin\psi'$ und somit

$$\frac{v_1}{v_2} = \frac{\sin\phi'}{\sin\psi'}.$$

Das ist das Brechungsgesetz. Der Weg des Lichtes ist also zeitoptimal. Machen Sie sich klar, wo die Winkel ϕ', ψ', in der Physik auch Eintritts- und Austrittswinkel (des Lichtes in ein Prisma) genannt, im Bild auftauchen !

Um die *sufficient second order condition SOSC* zu prüfen (besser argumentiert man hier direkt, dass man ein Minimum hat), kann man die 2. Abl. aus (4.3.1) ausrechnen,

$$\frac{\partial^2 L}{\partial x \partial y} = \frac{\partial^2 L}{\partial y \partial x} = -\frac{1}{v_1}\,\frac{(x-a)(y-b)/l_1}{l_1^2} - \frac{1}{v_2}\,\frac{(x-A)(y-B)/l_2}{l_2^2}, \tag{4.3.3}$$

$$\frac{\partial^2 L}{\partial x^2} = \frac{1}{v_1}\ \frac{l_1 - (x-a)^2/l_1}{l_1^2}\ + \frac{1}{v_2}\ \frac{l_2 - (x-A)^2/l_2}{l_2^2} \tag{4.3.4}$$

$$\frac{\partial^2 L}{\partial y^2} = \frac{1}{v_1}\ \frac{l_1 - (y-b)^2/l_1}{l_1^2}\ + \frac{1}{v_2}\ \frac{l_2 - (y-B)^2/l_2}{l_2^2}. \tag{4.3.5}$$

Der Kegel V in Thm. 4.2.1 (hier U, weil v schon besetzt) ist jetzt durch

$$U = \{u \in \mathbb{R}^2 \mid Df(\bar{x},\bar{y})u = 0,\ Dh(\bar{x},\bar{y})u = 0\} \tag{4.3.6}$$

gegeben, also durch $(\frac{\bar{x}-a}{v_1 l_1} + \frac{\bar{x}-A}{v_2 l_2})\ u_1 +\ (\frac{\bar{y}-b}{v_1 l_1} + \frac{\bar{y}-B}{v_2 l_2})\ u_2 = 0$ und $u_1 + u_2 = 0$. Wenn $u \in U, u \neq 0$, folgt also $u_2 = -u_1$. Die andere Bedingung schränkt u nicht ein, da sie zu $\frac{\bar{x}-a}{v_1 l_1} + \frac{\bar{x}-A}{v_2 l_2} = \frac{\bar{y}-b}{v_1 l_1} + \frac{\bar{y}-B}{v_2 l_2}$ führt, was nach (4.3.2) in der Lösung erfüllt ist. Die SOSC ist also für ein nichttriviales u, etwa $u = (1,-1)$ zu prüfen:

$$0 < u^T D_x^2 L u = \frac{\partial^2 L}{\partial x^2} - 2\frac{\partial^2 L}{\partial x \partial y} + \frac{\partial^2 L}{\partial y^2} \quad ?$$

Nach Multiplikation mit $v_1 v_2\ l_1^2 l_2^2$ heisst das

$$\begin{array}{lll} 0 < -v_2 l_2^2\ (2(x-a)(y-b)/l_1) & -\ v_1 l_1^2\ (2(x-A)(y-B)/l_2) \\ +\ \ v_2 l_2^2\ \ (l_1 - (x-a)^2/\ l_1) & +\ v_1 l_1^2\ \ (l_2 - (x-A)^2/\ l_2) \\ +\ \ v_2 l_2^2\ \ (l_1 - (y-b)^2/\ l_1) & +\ v_1 l_1^2\ \ (l_2 - (y-B)^2/\ l_2). \end{array}$$

Für die Summanden der ersten Spalte folgt (bis auf den Faktor)

$$[l_1 - (x-a)^2/l_1]\ +\ [l_1 - (y-b)^2/l_1] -\ [2(x-a)(y-b)/l_1] =\ \ldots\ = l_1 - l_1 \sin(2\phi).$$

Die Summe ist positiv $\Leftrightarrow \phi \neq \frac{\pi}{4}$. Analog ist die zweite Spaltensumme positiv $\Leftrightarrow \psi \neq \frac{\pi}{4}$. Die SOSQ ist also nur im Falle $\phi = \psi = \frac{\pi}{4}$ verletzt. Dann ist allerdings $v_1 = v_2$, womit das Problem trivial wird.

Chapter 5

Duality and perturbations in Banach spaces

Optimality conditions and duality are closely related to perturbations of the constraints in the original problem. We already mentioned such relations in sect. 4.1. Here, we investigate these interrelations for nonlinear B-space problems P6 and P7, sect. 1.4.

5.1 Separation in a normed space

In sect. 4.1, we applied LP-duality (based on a separation theorem) to derive the KKT conditions. In order to guarantee similar optimality and duality conditions for problems in normed spaces, one needs a stronger separation theorem. Therefore, it's now time to note that Thm. 3.1.8 can be derived, with $B - A = M$ from

Theorem 5.1.1 *Let X be a linear normed space on $\mathbb{R}$, let $M \subset X$ be convex, and let $0 \notin \operatorname{int} M \neq \emptyset$. Then there is some $x^* \in X^* \setminus \{0\}$ such that $0 \leq \langle x^*, x\rangle \ \forall x \in M$.* $\diamond$

Proof. We fix some $\bar{x} \in \operatorname{int} M$ and consider the 1-dimensional subspace $U = \{t\bar{x} \mid t \in \mathbb{R}\}$. Setting $g^0(t\bar{x}) = t$, we ensure that

$$0 \leq g^0(x) \quad \forall\, x \in M \cap U. \tag{5.1.1}$$

This way g^0 satisfies the imposed requirements at x^* for all $x \in M \cap U$. We shall extend g^0, defined on U, to a lin. mapping g defied on X. Then $g \not\equiv 0$ holds trivially.

Let $G \supset U$ be a subspace of X (closed or not) and $g : G \to R$ be additive und homogeneous with $g \equiv g^0$ on U and $0 \leq g(x)\ \forall x \in M \cap G$. Among these pairs (g, G) define a half-ordering:

$$(g, G) < (g', G') \quad \text{if} \quad G \subset G',\ G \neq G' \text{ and } g' \equiv g\ \forall x \in G.$$

The set of these (feasible) pairs is not empty, i.e., $(g, G) = (g^0, U)$.

For any increasing sequence (g_α, G_α) with $(g_\alpha, G_\alpha) < (g_\beta, G_\beta)$ if $\alpha < \beta$ (these are ordinal numbers) there exists some upper bound (g', G'), namely $G' = \cup G_\alpha$ and $g'(x) = g_\alpha(x)$ if $x \in G_\alpha$. Therefore, there exists (Zorn's Lemma) some maximal element for the half-ordering. This is, by definition, a pair (g, G), such that $(g, G) < (g', G')$ is never true. We show by contradiction that the latter entails $G = X$. Assume there is some $p \in X \setminus G$. Then we obtain

$$G \subset G' := \operatorname{lin}(G \cup \{p\}), \quad G' \neq G,$$

and each $x' \in G'$ has a unique representation $x' = x + tp$ ($x \in G,\ t \in \mathbb{R}$). Uniqueness: If $x' = y + sp$ ($y \in G, s \in R$) then $0 = y - x + (s-t)p$. Since 0 and $y - x$ belong to G we see that $(s-t)p \in G$. The latter yields $s = t$ since $p \notin G$. Thus also $y = x$ follows.

We define now $g'(x') = g(x) + at$ with some real a. To obtain a feasible pair (g', G') we need some a, such that

$$0 \le g(x) + at \quad \forall (x,t) \in G \times \mathbb{R} \ \text{ with } \ x + tp \in M. \tag{5.1.2}$$

To find a, we consider positive and negative t, so (5.1.2) attains the form:

$$0 \le g(x) + at \quad \forall x + tp \in M \cap G',\ t > 0. \tag{5.1.3}$$

$$0 \le g(y) + as \quad \forall y + sp \in M \cap G',\ s < 0. \tag{5.1.4}$$

The existence of such pairs (x,t) and (y,s) is ensured by $\bar{x} \in G \cap \operatorname{int} M$ which yields that $\bar{x} + \varepsilon p \in M,\ \bar{x} - \varepsilon p \in M$ for small $\varepsilon > 0$ (Note: this needs the algebraic interior only). Now let (x,t) and (y,s) be fixed as above. We define $(z,0)$ as a convex-combination with $0 = \lambda t + (1-\lambda)s, \quad z = \lambda x + (1-\lambda)y$. This implies

$$\lambda = -s/(t-s), \quad 1 - \lambda = t/(t-s) \quad \text{and} \quad z = (t-s)^{-1}(ty - sx).$$

Obviously, $z \in G$. For the convex- combination of points in M, also $z = z + 0p \ \in M$ holds true. Thus we conclude

$$0 \le g(z) = (t-s)^{-1}(-sg(x) + tg(y)) \quad \text{and} \quad sg(x) \le tg(y)$$

as well as (divide by $-st > 0$)

$$-\frac{g(x)}{t} \ \le \ -\frac{g(y)}{s}.$$

Hence there is some a which satisfies the crucial inequality

$$\sup_{(x,t)} \ -\frac{g(x)}{t} \le a \le \inf_{(y,s)} \ -\frac{g(y)}{s}.$$

In consequence, we obtain $-g(x) \le ta$ and $sa \ge -g(y)$ ($s < 0$), such that (5.1.3) and (5.1.4) are really true. This contradicts maximality of (g, G). Therefore, $G = X$.

Now g is additive, homogeneous and satisfies $0 \le g(x)\ \forall x \in M$. Recalling $\bar{x} \in \operatorname{int} M$ (only here, the topological interior is used), let $B(\bar{x}, \varepsilon) \subset M$. Then

$$0 \le g(\bar{x}) + \varepsilon g(b) \quad \forall\ b \in B(0,1),$$

shows that g is bounded below on $B(0,1)$. Clearly, this implies that g is bounded at all and that g is continuous. Hence $x^* = g$ is the functional in question. □

Now also Thm. 3.4.6 and 3.1.8 (existence of subgradients, separation general) are verified.

5.2 Strong duality and subgradients in B-spaces

We consider the classical nonlinear B-space problem, i.e.,

$$\begin{array}{lll} (P) & \inf\ \{\ f(x) \mid\ g(x) \in K \text{ and } h(x) = 0\} & \text{where} \\ & X, Y, Z \ \text{ are B-spaces,} & \\ & f : X \to \mathbb{R}, \ \ g : X \to Y, \ \ h : X \to Z, & \\ & K \neq \emptyset \ \text{ is a closed, convex cone in } Y. & \end{array} \tag{5.2.1}$$

For our standard $\mathbb{R}^n$ problems, we had $K = \mathbb{R}^m_-,\ Y = \mathbb{R}^m$.
We will also suppose that the extremal value v_P of (P) is finite.

Remark 5.2.1 As long as K is any closed convex cone (possibly with $\operatorname{int} K = \emptyset$), we may equivalently put/require

$$\hat{g} = (g, h) : X \to Y \times Z, \quad \hat{K} = K \times \{0\} \subset \hat{Y} := Y \times Z, \quad \hat{g}(x) \in \hat{K}.$$

Now, the equation is hidden in $\hat{g} \in \hat{K}$. Hence, to simplify the subsequent proofs, we may suppose that h is not present. ◇

Remark 5.2.2 Until formula (5.3.3), we shall nowhere need continuity or convexity of f, g, h, and X may be any set. ◇

Remark 5.2.3 One may define a half-ordering by setting $y \leq_K 0$ if $y \in K$; and $y \leq_K y'$ if $y - y' \leq_K 0$. Then, one may read $g(x) \in K$ like a usual inequality constraint. ◇

We put

$$S(y, z) = \{\ x \in X \mid g(x) \in y + K \text{ and } h(x) = z\} \text{ where } y \in Y, z \in Z \tag{5.2.2}$$

and study the perturbation function

$$\phi(y, z) = \inf\ \{f(x) \mid x \in S(y, z)\} \tag{5.2.3}$$

with possibly improper values $\pm\infty$.
Suppose $(0, 0) \in \operatorname{dom} \phi$ (the original infimum v_P is finite), and define the Lagrangian

$$L(x, y^*, z^*) = f(x) + \langle y^*, g(x)\rangle + \langle z^*, h(x)\rangle \qquad \text{and} \tag{5.2.4}$$

$$H(y^*, z^*) = \inf_{x \in X}\ L(x, y^*, z^*) \qquad (\in \mathbb{R} \cup \{-\infty\}). \tag{5.2.5}$$

We also define the *polar cone* $K^* = \{y^* \in Y^* \mid \langle y^*, k\rangle \leq 0\ \forall k \in K\}$ and the *dual problem*

$$(D) \qquad \sup\ \{H(y^*, z^*) \mid y^* \in K^*,\ z^* \in Z^*\} \tag{5.2.6}$$

with the finite or infinite extreme value v_D.

Lemma 5.2.4 *It holds* $\phi(0, 0) = \inf_{x \in X}\ \sup_{y^* \in K^*,\ z^* \in Z^*}\ L(x, y^*, z^*)\ \geq\ H(y^*, z^*)$. ◇

Proof. We apply Remark 5.2.1 to simplify the proof by deleting h, z and z^*.
First, consider any x with $g(x) \notin K$. Since K is closed, some $\varepsilon > 0$ fulfills

$$K \cap B(g(x), \varepsilon) = \emptyset.$$

Convexity of $K \subset Y$ and separation in accordance with Thm. 3.1.8 yields the existence of $y^* \in Y^* \setminus \{0\}$ with

$$(y^*, y) \geq (y^*, k) \quad \forall y \in B(g(x), \varepsilon) \text{ and } k \in K.$$

Because of $0 \in K$ and $y^* \neq 0$ it follows $\langle y^*, g(x)\rangle > 0$.
Since $k \in K \Rightarrow \lambda k \in K\ \forall \lambda > 0$, it follows $\langle y^*,\ k\rangle \leq 0$. Thus $y^* \in K^*$.
Finally, multiplying y^* with big λ shows that $\sup_{y^* \in K^*} L(x, y^*) = \infty$ if $g(x) \notin K$.
For $g(x) \in K$ we have $\max_{y^* \in K^*} \langle y^*, g(x)\rangle = 0$ and $\sup_{y^* \in K^*} L(x, y^*) = f(x)$.
Thus, since $\inf \sup L \geq \sup \inf L$ is always true, the assertion follows

$$\phi(0) = \inf_x \sup_{y^* \in K^*} L(x, y^*) \geq \sup_{y^* \in K^*} \inf_x L(x, y^*) = v_D. \qquad \square$$

Theorem 5.2.5 *(strong duality and subgradient) It holds the key relation*

$$-(y^*, z^*) \in \partial\phi(0,0) \;\Leftrightarrow\; (y^*, z^*) \text{ solves (D) with } v_D = v_P. \quad \diamond \tag{5.2.7}$$

Definition 5.2.1 *Strong/weak duality:* In the second situation, one says strong duality holds true for the pair of problems (P), (D). If only $v_D = v_P$, one often speaks on weak duality. $\diamond$

Proof. (of the Thm.) Again we apply Remark 5.2.1 to simplify the proof by deleting h, z and z^*.

1. Let $u^* \in \partial\phi(0)$.

For all y and all k in the *cone* K, we observe that $g(x) - y \in K$ yields $k + g(x) - y \in K$. This implies

$$S(y) \subset S(y-k) \;\text{ and }\; \phi(y-k) \le \phi(y).$$

In particular, it holds $\phi(-k) \le \phi(0)$. This and the subdifferential property

$$\phi(0) \ge \phi(-k) \ge \phi(0) + \langle u^*, -k\rangle$$

ensure $\langle u^*, -k\rangle \le 0$ (otherwise multiply k with big $\lambda > 0$). Thus, $y^* := -u^* \in K^*$. Due to $u^* \in \partial\phi(0)$ and $0 \in K$, it also holds

$$\phi(0) \le \inf_y\,[\phi(y) + \langle y^*, y\rangle] = \inf_y\,[\inf_{x:g(x)\in y+K} f(x) + \langle y^*, y\rangle\,] \le \inf_y\,[\inf_{x:g(x)=y} f(x) + \langle y^*, y\rangle\,].$$

In consequence, we derived $\phi(0) \le \inf_y\,[\,\inf_x\,\{\,f(x) + \langle y^*, g(x)\rangle \mid g(x) = y\,\}\,]$ and, since nothing is required for y,

$$\phi(0) \le \inf_x f(x) + \langle y^*, g(x)\rangle = H(y^*).$$

Since $H(y^*) \le \phi(0)$ is always true, cf. (2.7.3), sect. 2.7, we showed that $H(y^*) = \phi(0)$ is maximal.

2. Let $H(y^*) = \phi(0)$ and $y^* \in K^*$.

Then

$$\inf_x\; f(x) + \langle y^*, g(x)\rangle = H(y^*) \ge \phi(0). \tag{5.2.8}$$

Recalling $y^* \in K^*$ the inclusion $g(x) - y \in K$ implies

$$\langle y^*, g(x)\rangle \le (y^*, y).$$

Thus we may replace, on the left-hand side of (5.2.8), the term $\langle y^*, g(x)\rangle$ by $\langle y^*, y\rangle$, provided that $g(x) \in y + K$. Since $0 \in K$, such y exist, i.e., $y = g(x)$. This way we obtain

$$\begin{array}{lll} \phi(0) \le & \inf_x & \inf_{g(x)\in y+K} \quad f(x) \;+\; \langle y^*, y\rangle \qquad\qquad \text{and} \\ \phi(0) \le & \inf_y\,(\,[\inf_{x:g(x)\in y+K} & f(x)\,] + \langle y^*, y\rangle\,) = \inf_y\; \phi(y) + \langle y^*, y\rangle. \end{array}$$

So $u^* = -y^*$ is a subgradient for ϕ at 0. $\square$

5.3 Minimax and saddle points

Using the definitions only, (5.2.7) implies the "minimax-relation"

$$\partial\phi(0,0) \ne \emptyset \;\Leftrightarrow\; \max_{y^*\in K^*, z^*\in Z^*}\; \inf_{x\in X}\; L(x, y^*, z^*) = \inf_{x\in X}\; \sup_{y^*\in K^*, z^*\in Z^*} L(x, y^*, z^*). \tag{5.3.1}$$

Denoting a solution of (5.2.6) by $(\bar{y}^*, \bar{z}^*)$ and assuming that $\bar{x}$ solves the original problem, the right-hand side of (5.3.1) becomes the *saddle point condition*

$$L(\bar{x}, y^*, z^*) \ \leq L(\bar{x}, \bar{y}^*, \bar{z}^*) \leq \ L(x, \bar{y}^*, \bar{z}^*) \ \ \forall x \in X, \ y^* \in K^*, \ z^* \in Z^*. \tag{5.3.2}$$

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

Normal cone: The left-hand inequality yields that $\langle \bar{y}^*, \ k - g(\bar{x}) \rangle \leq 0 \ \forall k \in K$ and means (by definition only) that

$$\bar{y}^* \in N_K(g(\bar{x})). \tag{5.3.3}$$

Additional hypotheses: Let $f, g, h \in C^1$. Then the right-hand inequality of (5.3.2) yields the necessary (Lagrange) condition

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

In terms of *adjoint operators* this yields:

Corollary 5.3.1 *Under strong duality, every solution $\bar{x}$ of the original problem satisfies*

$$Df(\bar{x}) + Dg(\bar{x})^* \bar{y}^* + Dh(\bar{x})^* \bar{z}^* = 0 \in X^* \ \ \textit{for some} \ \ \bar{y}^* \in N_K(g(\bar{x})) \ \textit{and} \ \bar{z}^* \in Z^*. \ \diamond \tag{5.3.4}$$

Again the elements $\bar{y}^*, \bar{z}^*$ are called Lagrange multipliers to $\bar{x}$. For $\mathbb{R}^n$- problems, (5.3.4) represents the KKT conditions. In the theory of optimal control, (5.3.4) leads to the adjugate (and Hamilton) system, $\bar{z}^*$ is assigned to the differential equation for the trajectories, $\bar{y}^*$ corresponds to "phase constraints, e.g. $g(x(t), t) \leq 0$" and all together have to satisfy the adjoint equation.

5.4 Ensuring strong duality and existence of LM's

The above statements do *not* ensure that $\partial\phi \neq \emptyset$ or (D) has the required solutions. Therefore, it needs *additional conditions for verifying strong duality* !

5.4.1 The convex case

Theorem 5.4.1 *(Convex problems in B-spaces)*
Let problem (5.2.1) be a "convex" problem with Slater point:
(i) f, g, h *are contin.,* f *is convex,* h *is affine and* $h(X) = Z$, g *is convex w.r. to* K*;*
i.e., $g(\lambda x + (1-\lambda)x') \ - \ [\lambda g(x) + (1-\lambda) g(x')] \in K \quad \forall x, x' \in X, \ \lambda \in (0,1)$
(ii) $g(x^S) \in \operatorname{int} K$ *for some* $x^S \in M$.
Then, if v_P *is finite, it holds* $\quad \partial\phi(0,0) \neq \emptyset \quad$ *(= strong duality).* $\diamond$

The point x^S can be identified with a *Slater point* for $\mathbb{R}^n$-problems.

Proof. Having Thm. 3.4.6 in mind, we obtain $\partial\phi(0,0) \neq \emptyset$ if
(1) $\phi(y,z) \leq C < \infty$ for (y,z) near the origin and (2) convexity of ϕ can be shown.

(1) Since h maps onto Z and X, Z are Banach spaces, it follows from Banach's theorem on the inverse linear operator that, for each $z \in Z$, there is some

$$x_z = x^S + \xi \in h^{-1}(z) \text{ satisfying } \|\xi\| \leq L\|z\|$$

with some constant L. Let $\delta > 0$ satisfy $B(g(x^S), \delta) \subset K$. Since g is continuous, there is some $\alpha > 0$ such that $\| g(\xi + x^S) - g(x^S) \| < \frac{\delta}{2}$ if $\|\xi\| < \alpha$. This entails

$$g(\xi + x^S) - y \ \in \ B(g(x^S), \delta) \subset K \ \text{ if } \|\xi\| < \alpha \text{ and } \|y\| < \frac{\delta}{2}.$$

Therefore, it follows the existence of $x_z \in S(y, z)$ provided that $L\|z\| < \alpha$ and $\|y\| < \frac{\delta}{2}$. Since $f(x_z) < f(x^S) + 1$ for small $\|z\| + \|y\|$ is ensured by continuity of f, we obtain that ϕ is bounded above near the origin, say for $(y, z) \in B(0, \gamma)$.

(2) Let $x \in S(y, z)$ and $x' \in S(y', z')$ realize the infimum up to some $\varepsilon > 0$, i.e.,

$$f(x) \le \phi(y, z) + \varepsilon, \qquad f(x') \le \phi(y', z') + \varepsilon \tag{5.4.1}$$

and let $\lambda \in (0, 1)$ (if $\phi = -\infty$ read $\phi + \varepsilon = -1/\varepsilon$). Then

$$x_\lambda := \lambda x + (1 - \lambda)x' \in S(\ \lambda(y, z) + (1 - \lambda)(y', z')\ )$$

follows from affine linearity of h and convexity of g since

$$g(x_\lambda) - \ [\lambda g(x) + (1 - \lambda)g(x')] \in K \ \text{ and } \ \lambda g(x) + (1 - \lambda)g(x') \in K$$

yields that $g(x_\lambda)$ is again in K. By convexity of f, also

$$f(x_\lambda) \le \lambda f(x) + (1 - \lambda)f(x') \le \lambda(\phi(y, z) + \varepsilon) + (1 - \lambda)(\phi(y', z') + \varepsilon)$$

holds true and implies as $\varepsilon \downarrow 0$, for finite values,

$$\phi(\ \lambda(y, z) + (1 - \lambda)(y', z')\ ) \le f(x_\lambda) \le \lambda\phi(y, z) + (1 - \lambda)\phi(y', z')$$

and $\qquad \phi(\ \lambda(y, z) + (1 - \lambda)(y', z')\ ) \ = -\infty$ otherwise.

The second case only happens if $\phi(y', z') = -\infty$ or $\phi(y, z) = -\infty$. We assume the latter and consider a point $-t(y, z)$ which belongs to $B(0, \gamma)$ for small $t > 0$. Connecting $-t(y, z)$ and (y, z), now the above estimate yields $\phi(0, 0) = -\infty$, too. This contradicts $v_P \in \mathbb{R}$. Thus the second case is impossible; ϕ is convex. □

Note:
In accordance with this proof, the hypotheses of Thm. 5.4.1 had only to ensure that ϕ is convex and locally bounded above. This is valid after several modifications of the hypotheses. We restrict us to the above situation.

5.4.2 The C^1- case and linear approximations

For the more general non-convex problem (5.2.1), we suppose

$$\text{let } f, g, h \in C^1, \text{ and let } \bar{x} \text{ be a (local) minimizer.} \tag{5.4.2}$$

Put $f_L(x) = f(\bar{x}) + Df(\bar{x})(x - \bar{x})$, similarly $g_L(x)$ and $h_L(x)$. The linearizations define a particular convex problem P_L.

$$P_L: \qquad \inf\{\ f_L(x) \ \mid \ g_L(x) \in K \text{ and } h_L(x) = 0\ \}. \tag{5.4.3}$$

Theorem 5.4.2 *(C^1- problems in B-spaces) For problem (5.2.1), suppose (5.4.2) and*
(iii) $Dh(\bar{x})$ maps onto Z and
(iv) some x^S fulfills $g_L(x^S) \in \operatorname{int} K$ and $h_L(x^S) = 0$.
Then $\bar{x}$ solves P_L which satisfies the hypotheses of Thm. 5.4.1. Hence the optimality condition (5.3.4) is satisfied. ◇

Proof. We have only to show that $\bar{x}$ solves P_L. Otherwise some $u \in X$ exists such that

$$Df(\bar{x})u < 0,\ Dh(\bar{x})u = 0,\ g(\bar{x}) + Dg(\bar{x})u \in K.$$

With small $\varepsilon > 0$, the point $v = u + \varepsilon(x^S - \bar{x})$, satisfies the same condition. In addition, also $g(\bar{x}) + Dg(\bar{x})v \in \operatorname{int} K$ follows. This implies, with some $\delta > 0$,

$$g(\bar{x}) + tDg(\bar{x})v + y \in K \quad \text{if} \quad \|y\| \le t\delta. \tag{5.4.4}$$

Because of $Dh(\bar{x})v = 0$, it holds $h(\bar{x} + tv) = o(t)$. Now, instead of the implicit function theorem in finite dimension, one can apply the Lyusternik/Graves Theorem (Thm. 10.8.2), see also [20]. It ensures that there exist a constant L and certain $\xi(t),\ t \downarrow 0$ such that

$$h(\bar{x} + tv + \xi(t)) = 0 \ \text{ and } \ \|\xi(t)\| \le L\|o(t)\|.$$

Applying $g \in C^1$ and (5.4.4), we still obtain $g(\bar{x} + tv + \xi(t)) \in K$. Hence we find feasible points $x(t) = \bar{x} + tv + \xi(t) \to \bar{x}$ for (P) such that $f(x(t)) < f(\bar{x})$ though $\bar{x}$ is, by (5.4.2), a local minimizer for problem (5.2.1). This contradiction proves the assertion. □

Comparison with MFCQ for the $\mathbb{R}^n$ case:
The proof of Thm. 5.4.2 used, with another implicit function theorem, the arguments of the proof for MFCQ being a CQ in sect. 4.1.2. Also the hypotheses are similar (put $K = \mathbb{R}^{m-}$). In fact, condition (iii) means that *rank $Dh(\bar{x}) = m_h$ is maximal, and (iv) attains the form: Some $u \in \mathbb{R}^n$ ($u = x^S - \bar{x}$) fulfills $Dh(\bar{x})u = 0$ and $Dg_i(\bar{x})u < 0$ whenever $g_i(\bar{x}) = 0$.* This is again MFCQ.

Hence (iii) and (iv) represent a generalized MFCQ for B-space problems. However, this generalized MFCQ is no longer necessary for the Aubin property, details in [88].

Remark 5.4.3 By the reasoning of Thm. 5.4.2, we also see:
Provided strong duality holds for the linearized problem P_L, the weak hypotheses (5.4.2) entail the Lagrange condition (5.3.4), *if $\bar{x}$ also solves the linearization P_L.* The latter - solving P_L - is already guaranteed under *calmness of the constraint map S (5.2.2) at* $(0, \bar{x})$ (exactly by the arguments (v), sect. 4.1.2). ◇

Roughly speaking, a sharp characterization of the needed "regularity conditions CQ" requires a sharp characterization of our weakest CQ, namely calmness for perturbed feasible sets. This becomes a subject of sect. 10.

5.5 Modifications for vector- optimization

In all considered problems P1, ... , P7, sect. 1.4, one can ask for Pareto-optimal (also called efficient) points $\bar{x}$.

This means that a finite number of objectives f_j is given, and $\bar{x} \in M$ has to satisfy:

$$\text{there is no } x \in M \text{ such that } f_j(x) \le f_j(\bar{x})\ \forall j \text{ and } f_j(x) < f_j(\bar{x}) \text{ for some } j. \tag{5.5.1}$$

Let E be the set of efficient points.

To generate such points one can use (in the local and global sense) that every minimizer $\bar{x}$ of f_λ on M satisfies (5.5.1), provided that

$$f_\lambda(x) = \sum_j \lambda_j f_j(x) \text{ and } \lambda_j > 0\ \forall j. \tag{5.5.2}$$

Hence: The union of all minimizer to $f_\lambda,\ \lambda > 0$ is a subset $E' \subset E$.

This set can be characterized as follows (we consider only points in M):

$\bar{x} \in E' \ \Leftrightarrow\ \exists \varepsilon > 0$ such that if $f_j(x) < f_j(\bar{x})$ for some j, then

$$f_r(x) \geq f_r(\bar{x}) + \varepsilon(f_j(\bar{x}) - f_j(x)) \text{ for some } r \neq j. \tag{5.5.3}$$

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

$$f_j(x) < f_j(\bar{x})\ \forall j \neq r \text{ and } f_r(x) - f_r(\bar{x}) = o(\sum_{j \neq r}(f_j(\bar{x}) - f_j(x)) > 0 \tag{5.5.4}$$

where the variation of f_r is very small in comparison with $f_j(\bar{x}) - f_j(x)$.

For the "smooth" finite-dimensional vector optimization problem P4, it holds

Theorem 5.5.1 *To every $\bar{x} \in E$ which satisfies MFCQ (or some other CQ), there is a nontrivial $\lambda \geq 0$ such that the KKT conditions can be satisfied with $f = f_\lambda$.* ◇

Proof. The proof can similarly exploit LP-duality and CQ as Thm. 4.1.2. Now, we have only more Lagrange multipliers, assigned to the objective functions. □

Chapter 6

Lösungsverfahren; NLO in endl. Dimension

Let us start with free minima.

6.1 Freie Minima, Newton Meth. und variable Metrik

The basic model:

$$\min f(x)$$

where $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$ (locally Lipschitzian derivatives) is supposed to have bounded level sets

$$F(c) = \{x \mid f(x) \leq c\}$$

for all c. The goal consists in finding some $\bar{x}$ with $Df(\bar{x}) = 0$.

The standard iteration scheme to find some x with $\|Df(x)\| \leq \varepsilon$ consists in choosing, for given x_k, some descent direction u_k and a stepsize $t_k > 0$ such that $f(x_k + t_k u_k) < f(x_k)$ and to put

$$x_{k+1} = x_k + t_k u_k.$$

Of course, in order to ensure convergence, the descent has to be big enough.

6.1.1 "Fixed stepsize" or Armijo- Golstein rule

Let $0 < \lambda < 1$ and $\varepsilon > 0$.
ALG0: Given some x_k (with $\|Df(x_k)\| > \varepsilon$) and a stepsize parameter $t_k > 0$,
put $u_k = -Df(x_k)$, $\xi = x_k + t_k u_k$ and check whether

$$f(\xi) < f(x_k) - \tfrac{1}{2}\, t_k\, \|Df(x_k)\|^2. \tag{6.1.1}$$

(1) If Yes then put $x_{k+1} := \xi$, $t_{k+1} := t_k$.
(2) Otherwise put $x_{k+1} := x_k$, $t_{k+1} := \lambda\, t_k$.
$k := k + 1$, repeat.

Notice that, for linear f, we would always have situation (1)

$$f(\xi) - f(x_k) = t_k\, Df(x_k)u_k = -\, t_k\, \|Df(x_k)\|^2 < -\tfrac{1}{2} t_k\, \|Df(x_k)\|^2.$$

Lemma 6.1.1 *ALG0 finds (for all $\varepsilon, t_0 > 0$, $x_0 \in \mathbb{R}^n$) some x_k with $\|Df(x_k)\| \leq \varepsilon$.* $\diamond$

Proof. Assume that $t_k \geq \tau > 0 \ \forall k$. Then, it holds for all $k > k0$;

$$f(x_{k+1}) < f(x_k) - t_k \ \|Df(x_k)\|^2 \leq f(x_k) - \tau \ \|Df(x_k)\|^2.$$

Since f is bounded below, this implies $\|Df(x_k)\| \to 0$.
Now assume $t_k \to 0$ and consider the steps of case (2). They satisfy

$$f(x_k + t_k u_k) - f(x_k) \geq -\tfrac{1}{2} \ t_k \|Df(x_k)\|^2.$$

By the mean-value theorem, we may write

$$f(x_k + t_k u_k) - f(x_k) = \ t_k Df(x_k + \theta t_k u_k) u_k$$

with some $\theta = \theta_k \in (0,1)$. Hence, it follows

$$Df(x_k + \theta t_k u_k) u_k \geq -\tfrac{1}{2} \ \|Df(x_k)\|^2.$$

All x_k belong to the compact set F_c for $c = f(x_0)$. Thus $Df(x_k)$ and u_k remain bounded. So one may estimate, using a Lipschitz constant L for Df,

$$\|Df(x_k + \theta t_k u_k) - Df(x_k)\| \leq L \ t_k \|u_k\|$$

and conclude that

$$Df(x_k) u_k + L \ t_k \|u_k\|^2 \geq -\tfrac{1}{2} \ \|Df(x_k)\|^2. \tag{6.1.2}$$

In consequence, we obtain from $u_k = -Df(x_k)$,

$$-\|Df(x_k)\|^2 + L \ t_k \|u_k\|^2 \geq -\tfrac{1}{2} \ \|Df(x_k)\|^2,$$

$$L \ t_k \|u_k\|^2 \geq \tfrac{1}{2} \ \|Df(x_k)\|^2 \quad \text{and} \quad t_k \geq \frac{1}{2L}$$

thus $t_k \to 0$ cannot be true. □

Comments: The test (6.1.1) can be modified, e.g., by

$$f(\xi) - f(x_k) < -t_k \gamma \ \|Df(x_k)\|^2$$

with fixed $\gamma \in (0,1)$. Instead of $u_k = -Df(x_k)$ we can use some direction which still fulfills

$$Df(x_k) u_k < -p \ \|Df(x_k)\|^2, \quad p \in (\gamma, 1) \text{ fixed.}$$

6.1.2 Line search

Given $u_k = -Df(x_k)$ one may minimize the real function $g(t) = f(x_k + t u_k)$ for $t > 0$. Then the related minimizing t_k satisfies

$$0 = \frac{dg(t_k)}{dt} = Df(x_k + t_k u_k) \ u_k, \quad \text{and we put } x_{k+1} = x_k + t_k u_k.$$

To show again

$$Df(x_k) \to 0$$

we regard the opposite case: $\|Df(x_k)\| = \|u_k\| \geq \varepsilon > 0$ for some subsequence. Since $f(x_{k+1}) < f(x_k)$, the level-set assumption yields that t_k remains bounded and that, for some subsequence (of the subsequence), there are limits of x_k, t_k and u_k, say x, t and $u \neq 0$.

Hence $0 = Df(x+tu)\ u$. If $t = 0$ this yields the contradiction $0 = Df(x+tu)\ u = -\|u\|^2$. Let $t > 0$. At x, we know from $u = -Df(x) \neq 0$ that

$$f(x + su) - f(x) < \delta(x, s) < 0 \quad \text{for small } s > 0.$$

We fix some s with $s < t$ and put $\delta_0 = \delta(x, s)$. By continuity and line-search arguments, then also

$$f(x_k + t_k u_k) - f(x_k) \leq f(x_k + su_k) - f(x_k) < \delta_0 < 0$$

holds for the elements of our subsequence whenever k is large enough. For all other k, we have $f(x_k + t_k u_k) - f(x_k) \leq 0$. Thus we obtain the contradiction $f(x_k) \to -\infty$. In consequence, $\lim Df(x_k) = 0$ holds for the whole sequence as asserted.

6.1.3 Variable metric

The direction u_k can be determined in such a way that $\langle u, Df(x_k)\rangle$ is minimal w.r. to all $u \in B(0, \|Df(x_k)\|)$. With Euclidean norm, this gives again $u_k = -Df(x_k)$, but with other norms we obtain other u_k, in general. In particular, the norm may depend on k, then one speaks about *methods of variable metric*.

Suppose $f \in C^2$ has a positive definite Hessian $D^2 f(x_k)$ everywhere. Furthermore, determine u_k by

$$\min \{\ \langle u, Df(x_k)\rangle \mid u^T D^2 f(x_k) u \leq \|Df(x_k)\|\ \}.$$

Applying the KKT conditions to this convex problem with Slater points, one obtains as a necessary and sufficient optimality condition

$$Df(x_k) + 2yD^2 f(x_k)u = 0; \quad u = -\frac{1}{2y}[D^2 f(x_k)]^{-1} Df(x_k).$$

This tells us that u_k coincides, up to the factor $\frac{1}{2y}$, with the Newton direction

$$v = -[D^2 f(x_k)]^{-1} Df(x_k)$$

which solves the Newton equation for $Df(x) = 0$, namely $Df(x_k) + D^2 f(x_k)v = 0$.
Hence variable metric methods may connect methods of first and second order.

6.2 The nonsmooth Newton method

Here, we present properties of h which are necessary and sufficient for solving an equation

$$h(x) = 0, \qquad h : \mathbb{R}^n \to \mathbb{R}^n \text{ loc. Lipschitz}$$

by a Newton-type method.

Such methods can be applied to KKT-systems below (after any reformulation as a non-smooth equation). We shall see that our approach works more general and in the same manner for $h : X \to Y$ (normed spaces). Then, however, it is more difficult to find practically relevant mappings A_k considered next.

First of all notice that Newton methods cannot be applied to all locally Lipschitz functions, even if $n = 1$ (provided the steps have the usual form at C^1-points of h), cf. Example BE.1 in [70]. In this example, one obtains alternating sequences for almost all

initial points. Study also $h(x) = x^q,\ q \in (0,1)$, which shows the difficulties if h is everywhere C^1 except the origin, or if h is not locally Lipschitz.

The crucial conditions

Newton's method for computing a zero $\hat{x}$ of $h : X \to Y$ (normed spaces) is determined by the iterations

$$x_{k+1} = x_k - A^{-1}h(x_k), \tag{6.2.1}$$

where $A = Dh(x_k)$ is supposed to be invertible. The formula means that x_{k+1} solves

$$h(x_k) + A(x - x_k) = 0, \qquad A = Dh(x_k). \tag{6.2.2}$$

Forgetting differentiability let us replace A by any invertible linear operator $A_k : X \to Y$ (even if Dh exists).

Next, to replace regularity of $Dh(\hat{x})$ for the usual C^1- Newton method, suppose that there are constants K^+ and K^- such that

$$\|A\| \le K^+ \text{ and } \|A^{-1}\| \le K^- \quad \text{for all } A = A_k \text{ and small } \|x_k - \hat{x}\|. \tag{6.2.3}$$

The *locally superlinear convergence* of Newton's method means that, for some o-type function r and initial points near $\hat{x}$, we have

$$x_{k+1} - \hat{x} = z_k \quad \text{with} \quad \|z_k\|_X \le r(x_k - \hat{x}). \tag{6.2.4}$$

Substituting x_{k+1} from (6.2.1) and applying A_k to both sides, this requires

$$h(x_k) = h(x_k) - h(\hat{x}) = A_k\,(x_k - x_{k+1}) = A_k\,[(x_k - \hat{x}) - z_k] \ \text{ with } \ \|z_k\| \le r(x_k - \hat{x}). \tag{6.2.5}$$

Condition (6.2.5) claims equivalently (with $A = A_k$)

$$A(x_k - \hat{x}) = h(x_k) - h(\hat{x}) + Az_k, \quad \|z_k\|_X \le r(x_k - \hat{x}) \tag{6.2.6}$$

and yields *necessarily*, with

$$o(u) = K^+ r(u), \tag{6.2.7}$$

$$A(x_k - \hat{x}) = h(x_k) - h(\hat{x}) + v^k \quad \text{for some } v^k \in B(0,\ o(x_k - \hat{x})) \subset Y. \tag{6.2.8}$$

Conversely, having (6.2.8), it follows

$$x_k - \hat{x} = A^{-1}(h(x_k) - h(\hat{x})) + A^{-1}v^k \quad \text{for some } v^k \in B(0,\ o(x_k - \hat{x})). \tag{6.2.9}$$

So the solutions of the Newton equation (6.2.1) fulfill

$$\begin{aligned} z_k := x_{k+1} - \hat{x} \quad &= (x_{k+1} - x_k) + (x_k - \hat{x}) \\ &= -A^{-1}h(x_k) + A^{-1}(h(x_k) - h(\hat{x})) + A^{-1}v^k \\ &= A^{-1}v^k. \end{aligned}$$

Hence $\|z_k\|_X = \|A^{-1}v^k\| \le K^- o(x_k - \hat{x})$ yields the convergence (6.2.4) with

$$r(u) = K^- o(u) \tag{6.2.10}$$

for all initial points near $\hat{x}$. So we have shown

Theorem 6.2.1 *(Convergence of Newton's method) Under the regularity condition (6.2.3), method (6.2.1) fulfills the condition (6.2.4) iff the assignment $x_k \mapsto A_k$ satisfies (6.2.8).* ◇

Hence we may use any $A = A(x) \in Lin(X, Y)$ whenever the conditions

$$(CI) \qquad \|A(x)\| \le K^+ \text{ and } \|A(x)^{-1}\| \le K^- \text{ (injectivity) as well as} \tag{6.2.11}$$

$$(CA) \qquad A(x)(x - \hat{x}) \in h(x) - h(\hat{x}) + o(x - \hat{x})B \text{ (approximation)} \tag{6.2.12}$$

are satisfied for sufficiently small $\|x - \hat{x}\|$.
Notice that the quantities $o()$ and $r()$ are directly connected by (6.2.7) and (6.2.10).

Remark 6.2.2 If (6.2.11) and (6.2.12) hold true with $o(u) \le p\|u\|^2$ for all $\|u\| \le \delta$, then the method converges quadratically for all initial points $\bar{x}$ satisfying

$$\|\bar{x} - \hat{x}\| = q \quad \text{with } q < q_o := \min\{\delta,\ (K^- p)^{-1}\}$$

since, by induction, $\|x_{k+1} - \hat{x}\| \le K^- o(x_k - \hat{x}) \le K^- p\ q\|x_k - \hat{x}\|$ and $K^- p\ q < 1$. ◇

More important:
In the current context, the function $h : X \to Y$ may be arbitrary (for normed spaces X, Y) as long as $A(x)$ consists of linear (continuous) bijections between X and Y.

Nevertheless, outside the class of locally Lipschitz functions we cannot suggest any reasonable definition for $A(x)$ since (6.2.11) and (6.2.12) already imply the (pointwise) Lipschitz estimate

$$\|h(x) - h(\hat{x})\| \le (1 + K^+)\|x - \hat{x}\|$$

where the zero $\hat{x}$ is usually unknown.

On the other hand, linearity of $A(x - x_k)$ in the Newton equation (6.2.2) is not essential. By the estimates above and some additional devices [70] one can even use directional derivatives $h'(x_k, x - x_k)$ and other terms defined by generalized derivatives instead of $A(x - x_k)$. Then, one has to require, for points near $\hat{x}$:

$$(CI) \qquad \|h'(x_k, x - x_k)\| \ge K^- \|x - x_k\| \qquad \text{(injectivity)} \tag{6.2.13}$$

$$(CA) \qquad h'(x_k, x - x_k) \in h(x) - h(\hat{x}) + o(x - \hat{x})B \qquad \text{(approximation)} \tag{6.2.14}$$

while the existence of K^+ follows from the Lipschitz property of h.

Semismoothness:
Condition (6.2.12) appears in various versions in the literature. Let h be locally Lipschitz near $\hat{x}$ and $X = Y = \mathbb{R}^n$: If all $A(x)$ in Clarke's generalized Jacobian $\partial^c h(x)$ (a set of matrices, cf. sect. 7.3.2) satisfy (6.2.12) with the same $o(.)$, then h is called *semismooth* at $\hat{x}$, [103] ; sometimes – if $o(\cdot)$ is even quadratic – also *strongly semismooth*. In others papers, A is a mapping that approximates $\partial^c h$; and the functions h satisfying the related conditions (6.2.12) are called weakly semismooth.

However, neither a relation between A and $\partial^c h$ nor the existence of directional derivatives is essential for the interplay of the conditions (6.2.4), (6.2.11) and (6.2.12) in Thm. 6.2.1. The main problem is the characterization of those functions h which allow us to *find practically relevant Newton maps* $A = A(x)$ satisfying (6.2.12). These function classes are not very big in the class of loc. Lipschitz functions (cf. [70], locally PC^1-functions).

Particular Newton maps:

(i) Let $h = PC^1(\alpha_1, ..., \alpha_k)$ be a piecewise C^1 function, i.e., h is continuous, $\alpha_j \in C^1(X, Y)$, and for each x there is some (*active*) $j = j(x)$ such that $h(x) = \alpha_j(x)$. In this case, all

$$A(x) \in \{D\alpha_j(x) \mid \alpha_j(x) = h(x)\}$$

fulfill (6.2.12); this is an exercise. So one may assign, to x, any $A(x) = D\alpha_{j(x)}(x)$.

(ii) For an NCP functions $\sigma : \mathbb{R}^2 \to \mathbb{R}$ considered below, one may put (and easily determine)

$$A(z) \in \text{Limsup } D\sigma(z_k)$$

where $z_k \to z$ are arguments such that $D\sigma(z_k)$ exists and *Limsup* is the set of their possible limits (in $\mathbb{R}^{n*n}$). For real C^1 functions u and v and the auxiliary function

$$h(x) = \sigma(z), \quad \text{with } z = (u(x), v(x))$$

all chain-rule-matrices $A(z)\ (Du(x), Dv(v))^T$ satisfy (6.2.12).

For other approaches to such methods, see [43], [78], [82], [85], [128].

6.3 Cyclic Projections; Feijer Methode

Let $\emptyset \neq M_i \subset \mathbb{R}^n$ $(i = 1, ..., m)$ be closed, convex sets and let π_i denote the projection onto M_i. Given any $x_0 \in \mathbb{R}^n$ generate a sequence by cyclic projections onto M_i

$$x_1 = \pi_1(x_0),\ x_2 = \pi_2(x_1), ..,\ x_m = \pi_m(x_{m-1}), \quad x_{m+1} = \pi_1(x_m), ...$$

Theorem 6.3.1 *If $D := \cap_i M_i \neq \emptyset$ then this sequence converges to some $x \in D$.* ◇

Proof. Let $\hat{x} \in D$ be arbitrarily fixed, let π be the composed map $\pi_m \circ \pi_{m-1} \circ ... \circ \pi_1$ and $y_0 = x_0$. The projection Thm. 3.1.2, (iii) can be obviously extended to π, so we have for all steps $s = 0, 1, ...$ with $y_{s+1} = \pi(y_s)$,

$$\|\pi(y_s) - \hat{x}\| \leq \|y_s - \hat{x}\|. \tag{6.3.1}$$

Hence the sequence y_s is bounded and has accumulation points in $\mathbb{R}^n$. Next consider any $y \in \mathbb{R}^n \setminus D$. Then

$$\|\pi(y) - \hat{x}\| < \|y - \hat{x}\| - \delta, \quad \delta > 0$$

holds for some δ. By non-expansivity of the projection, so also

$$\|\pi(x) - \hat{x}\| < \|x - \hat{x}\| - \tfrac{1}{2}\delta$$

is valid for all $x \in B(y, \frac{\delta}{4})$. In consequence, $y \in \mathbb{R}^n \setminus D$ is never an accumulation point of the generated sequence y_s. So all accumulation points y of y_s belong to D. Having two of them, say x, y, then (6.3.1), with $\hat{x} = x$ entails $y = x$. Finally, the convergence $x_k \to x$ follows from the fact that

$$\|y_{s+1} - x\| \leq \|\pi_i(x_k) - x\| \leq \|y_s - x\| \tag{6.3.2}$$

holds for certain $s = s(k)$. □

6.4 Proximal Points, Moreau-Yosida approximation

For minimizing a convex function f on $\mathbb{R}^n$, one may consider the so-called (strongly convex) Moreau-Yosida approximation of f

$$F_y(x) = f(x) + \tfrac{1}{2}\|x - y\|^2.$$

Its minimizer $x = x(y)$ is unique since F_y is strongly convex, and is characterized by

$$0 \in \partial F_y(x) \ = \ x - y + \partial f(x). \tag{6.4.1}$$

Hence, the original solutions $\hat{x} \in \text{argmin } f$ are just the fixed points of the function

$$y \mapsto x(y).$$

The proximal point method generates a sequence by setting

$$x_{k+1} = \operatorname{argmin} F_{x_k} \qquad x_0 \text{ arbitrary.}$$

This is of some advantage if f is, like $f = x^8$ "very flat" near a minimizer; then direct solution methods create several difficulties in view of possible errors and speed of convergence.

Lemma 6.4.1 *If* $\operatorname{argmin} f \neq \emptyset$ *then the sequence* x_k *converges to a minimizer of* f.

Proof. Every x_{k+1} is the unique solution x to (6.4.1) for $y = x_k$. Since ∂f is monotone, it holds for related solutions to y and y'

$$y - x \in \partial f(x), \quad y' - x' \in \partial f(x')$$

$$0 \leq \langle y' - x' - (y - x),\ x' - x\rangle = \langle y' - y,\ x' - x\rangle - \|x' - x\|^2.$$

This is the same estimate as in formula (3.1.6), sect. 3.1, and entails non-expansivity as there, due to $\langle y' - y,\ x' - x\rangle \leq \|y' - y\|\ \|x' - x\|$, $0 \leq \|y' - y\|\ \|x' - x\| - \|x' - x\|^2$ and

$$\|x' - x\| \leq \|y' - y\|$$

Discussing here the equation, it should be evident, that convergence follows in the same manner as for cyclic projection. □

If $\mathbb{R}^n$ is replaced by a Hilbert space, one obtains still weak convergence of x_k by the same proof.

6.5 Schnittmethoden; Kelley-Schnitte

6.5.1 The original version

Die Aufgabe (4.0.1) $\min\{\ f(x) \mid\ g_i(x) \leq 0\ i = 1, 2, ..., m\}$ $f, g_i \in C^1$ kann man umformulieren als

$$\min\{\ t \mid\ f(x) - t \leq 0,\ g_i(x) \leq 0\ i = 1, 2, ..., m\}.$$

Hier ist die Zielfunktion linear in der neuen Variablen (t, x), und Konvexität der Input-Funktionen bleibt erhalten. Wir dürfen also die Zielfunktion in (4.0.1) gleich als linear annehmen und betrachten Problem

$$\min\{\ c^T x \mid g_i(x) \leq 0\ i = 1, 2, ..., m\}\ \ g_i \in C^1,\ g_i \text{ konvex,} \tag{6.5.1}$$

wobei die Restriktionsmenge $M \neq \emptyset$ kompakt sei.

Sei P ein kompaktes Polyeder, das die Menge M enthält. Dann ist die Aufgabe

$$\min\{\ c^T x \mid x \in P\} \tag{6.5.2}$$

lösbar. Hat man eine Lösung $x(P)$ von (6.5.2), so folgt im Falle $x(P) \in M$, dass $x(P)$ auch (6.5.1) löst. Im interessanten Fall $x(P) \notin M$ setzt Kelleys Idee [61] ein:
Man finde einen (affinen) Halbraum $H = \{x \mid\ a^T x \leq b\}$, so dass

$$M \subset H \text{ und } x(P) \notin H \qquad (H \text{ heisst dann Schnitt}).$$

Setzt man $P' = P \cap H$, folgt wieder $M \subset P'$, aber auch $x(P) \notin P'$. Die Lösung von (6.5.2) mit P' statt P liefert also einen neuen Punkt $x(P')$. Wiederholt man den Prozess, entsteht eine Folge P_k von Polyedern mit

$$M \subset P_{k+1} \subset P_k$$

und von Lösungen $x(P_k)$ der entsprechenden linearen Probleme.
Bei "vernünftiger" Definition der Halbräume $H = H_k$, löst jeder Häufungspunkt der Folge $x(P_k)$ das Problem (6.5.1). Kelley gab eine konkrete Möglichkeit an:

Sei $x_k := x(P_k) \notin M$. Dann ist ein $g_i(x_k)$ positiv. Man fixiere $i = i(k)$, so dass $g_i(x_k)$ maximal ist, und bilde

$$H_k = \{x \mid g_i(x_k) + Dg_i(x_k)(x - x_k) \leq 0\}; \quad P_{k+1} = P_k \cap H_k. \tag{6.5.3}$$

Dann sind beide Schnittbedingungen erfüllt: $x_k \notin P_{k+1}$ ist trivial. Für $x \in M$ folgt $g_i(x) \leq 0$, aber auch

$$g_i(x) \geq g_i(x_k) + Dg_i(x_k)(x - x_k) \tag{6.5.4}$$

wegen Konvexität. Also ist $M \subset H_k$. Die für konvexe Funktionen wichtige Ungleichung (6.5.4) folgt (unter anderem) aus $Dg_i(x_k) \in \partial g_i(x_k)$ (und Monotonie des Subdiff).

Lemma 6.5.1 *Jeder Häufungspunkt $\hat{x}$ der so konstruierten Folge löst (6.5.1).* ◇

Proof. Sei $\hat{x} = \lim x(P_{k(\nu)})$, $\nu \to \infty$. Weil stets $c^T x(P_k) \leq \inf_{x \in M} c^T x$ gilt, muss nur $\hat{x} \in M$ gezeigt werden. Andernfalls gibt es ein $i0$ und ein $\delta > 0$ mit $g_{i0}(\hat{x}) > \delta$, also auch $g_{i0}(x(P_{k(\nu)})) > \delta$ für grosse ν. Sei zur Abkürzung $x(\nu) = x(P_{k(\nu)})$. Für den in (6.5.3) (in Schritt $k(\nu)$) benutzten Index $i = i(\nu)$ (den man nach Auswahl einer geeigneten Teilfolge als konstant annehmen darf) gilt erst recht

$$g_i(x(\nu)) > \delta \text{ and } H_{k(\nu)} = \{x \mid g_i(x(\nu)) + Dg_i(x(\nu))(x - x(\nu)) \leq 0\}. \tag{6.5.5}$$

Nun ist $\|Dg_i(x(\nu))\| \leq C$ beschränkt (Dg_i ist auf P stetig). Somit hat $g_i(x(\nu)) > \delta$ und $x \in H_{k(\nu)}$ zur Folge, dass

$$\|x - x(\nu)\| \geq \frac{\delta}{C}$$

sein muss. Für grosse ν und $\nu' > \nu$ ist diese Ungleichung mit $x = x(\nu')$ wegen des gemeinsamen HPunktes $\hat{x}$ nicht erfüllt. Dieser Widerspruch beweist das Lemma. □

Deleting C^1: Sind die g_i nur konvex auf $\mathbb{R}^n$, kann man die Schnitte (6.5.3) analog mit einem beliebigen Subgradienten $s^k \in \partial g_i(x_k)$ anstelle von $Dg_i(x_k)$ definieren,

$$H_k = \{x \mid g_i(x_k) + \langle s^k, x - x_k \rangle \leq 0\}; \quad P_{k+1} = P_k \cap H_k. \tag{6.5.6}$$

Der Konvergenzbeweis bleibt derselbe.

6.5.2 Minimizing a convex function by linear approximation

Angenommen, es soll $\min f(x)$, $x \in P$ (kompaktes Polyeder, f konvex, C^1) bestimmt werden. Die Aufgabe kann man wieder umformulieren:

$$\min\{t \mid f(x) - t \leq 0, \quad x \in P, \ |t| \leq C\}.$$

Dabei sei C hinr. gross, so dass $|f(x)| < C \ \forall x \in P$ gilt. Mit

$$(x,t) \in \mathbb{R}^{n+1}, \quad g(x,t) = f(x) - t \ \text{ and } P^0 = P \times [-C, C]$$

wende man nun die Kelley Methode an. Der erste Punkt $(\bar{x}, t_0) \in P^0$ löst also

$$\min\{t \mid \ (x,t) \in P^0\}.$$

Der allgemeine Schritt lässt sich leicht mit Hilfe von f und P allein interpretieren. Zeigen Sie: Beginnend mit

$$L_1(x) = f(\bar{x}) + Df(\bar{x})(x - \bar{x}),$$

ist in Schritt $k \geq 1$ jeweils eine stückweise lineare Funktion L_k über P zu minimieren. In der jeweiligen Lösung x_k kommt eine weitere lineare Funktion hinzu, die Linearisierung $l_k(x) = f(x_k) + Df(x_k)(x - x_k)$ von f und es wird $L_{k+1}(x) = \max\{l_k(x), \ L_k(x)\}$. Die Grösse $\delta_k := f(x_k) - L_k(x_k)$ ist hierbei ein Mass für den Fehler.
Deleting C^1: Falls f nur konvex ist, ersetze man $Df(x_k)$ durch $s^k \in \partial f(x_k)$.

Es gibt viele Modifikationen dieses Algorithmus, vor allem indem man

$$L_k = \max\{l_1, ..., l_k\}$$

ersetzt durch $\hat{L}_k = \max\{l_{k-m}, ..., l_k\}$ oder eine analoge quadratische Funktion, sofern gewisse Fehlergrössen klein bleiben. Sie werden zumeist *bundle bzw. trust region methods* genannt.

Eine Uebersicht ueber die vielfaeltigen Ideen für Verfahren der nicht-differenzierbaren konvexen Optimierung einschliesslich ihrer Grundlagen findet man in [48].

6.6 Strafmethoden

Um die Aufgabe (4.0.1) $\min\{ \ f(x) \mid g_i(x) \leq 0 \ i = 1, 2, ..., m\} \ \ f, g_i \in C^1$ zu lösen, wendet man oft folgende *Strafmethode* an: Für grosse p minimiere man

$$H_p(x) = f(x) \ + \frac{1}{2} \ p \ \sum_i \ [\ \max(0, g_i(x)) \ ]^2 \qquad \text{bzgl. } \ x \in \mathbb{R}^n.$$

Der Term $S_p(x) = \frac{1}{2} p \sum_i [\max(0, g_i(x))]^2$ ist ein *Strafterm*: Er wird gross, wenn x die Nebenbedingungen verletzt, ansonsten ist er Null. Durch das Quadrieren von $\max(0, g_i(x))$ verschwindet die Nicht-Differenzierbarkeit dieser Funktion an Punkten mit $g_i(x) = 0$,

$$DH_p(x) = Df(x) + \ p \ \sum_i \ [\ \max(0, g_i(x)) \ ] \ Dg_i(x).$$

Man hofft nun, dass man Minimalpunkte $x(p)$ für H_p ausrechnen kann und diese für wachsendes $p = p_k \to \infty$ gegen eine Lösung von (4.0.1) konvergieren (zumindest nach Auswahl einer geeigneten Teilfolge). Das kann falsch sein.

Example 6.6.1 $f(x) = x^3, \ g(x) = g_1(x) = -x. \ \ H_p$ besitzt keine Minimum. $\diamond$

Man braucht unter anderem Konvexität.

Theorem 6.6.1 *Sei die Restriktionsmenge M von (4.0.1) nichtleeer und beschränkt und seien f, g_i konvex. Dann exist. $x(p) \in \operatorname{argmin} H_p$, und aus jeder Folge $\{x(p_k)\}$ für $p_k \to \infty$ kann man eine konverg. Teilfolge auswählen. Ihr Häufungspunkt löst (4.0.1).* $\diamond$

Proof. Zunächst ist (4.0.1) lösbar, da $M \neq \emptyset$ kompakt und f stetig ist. Sei $\hat{x}$ eine Lösung. Sei weiter $r > 0$ hinreichend gross, so dass M ganz im Innern der (Euklidischen) Kugel $K_r = \{x \mid \|x\| \leq r\}$ liegt. Mit einem $\delta > 0$ gilt dann für alle z mit $\|z\| = r$

$$\max_i [\max(0, g_i(z))] > \delta.$$

(i) Minima von $H_p,\ p > 0$ fest:
Angenommen $x_k \in \mathbb{R}^n$ erfüllen $\lim H_p(x_k) = \inf H_p$ (endlich oder nicht). Bleibt eine Teilfolge aller x_k beschränkt, so ist jeder ihrer Häufungspunkte Minimalpunkt von H_p. Also gelte $\|x_k\| \to \infty$. Dann schneidet (für grosse k) die Strecke zwischen $\hat{x}$ und x_k den Rand der Kugel K_r. Der Schnittpunkt z_k bleibt beschränkt und hat die Form

$$z_k = \lambda\hat{x} + (1-\lambda)x_k; \quad 0 < \lambda = \lambda(k) \to 1 - 0.$$

Also findet man ein $i = i(k)$, so dass

$$\delta < g_i(z_k) \leq \lambda g_i(\hat{x}) + (1-\lambda)g_i(x_k) \leq (1-\lambda)g_i(x_k).$$

Damit divergiert $S_p(x_k)$ in der Form $S_p(x_k) \geq \frac{1}{2}p\ \ [\frac{\delta}{1-\lambda}]^2$, während

$$f(z_k) \leq \lambda f(\hat{x}) + (1-\lambda)f(x_k) \text{ zeigt:} \quad f(x_k) \geq \frac{f(z_k) - \lambda f(\hat{x})}{1-\lambda}.$$

Der Zähler bleibt nach unten beschränkt durch ein $C < 0$, wonach gilt

$$H_p(x_k) \geq \frac{C}{1-\lambda} + \frac{1}{2}p\ \ [\frac{\delta}{1-\lambda}]^2 \to \infty.$$

Andererseits ist stets $\inf H_p \leq f(\hat{x})$, was $\lim H_p(x_k) = \inf H_p$ widerspricht.
(ii) Beschränktheit von $x(p)$:
Bleibt eine Teilfolge aller $x(p_k)$ beschränkt, folgt wieder über Stetigkeit, dass jeder Häufungspunkt (4.0.1) löst. Gelte also $\|x(p_k)\| \to \infty$. Mit $x_k = x(p_k)$ sind dann dieselben Abschätzungen wie oben richtig, nur divergiert jetzt auch $p = p_k$. Damit erhalten wir aber erneut den Widerspruch $H_p(x_k) \to \infty$. □

Der Beweis bleibt für zahlreiche Typen von Straftermen richtig, z.B. für

$$S_p(x) = p \sum_i [\ \max(0, g_i(x))\ ]^q\ ; \qquad q > 1 \text{ fest.}$$

Mit $q = 1$ ist S_p nicht mehr differenzierbar und die Existenz von $x(p)$ folgt nur noch (wie oben) für hinreichend grosse p. Dafür hat diese Funktion andere Vorteile. So gilt sogar $x(p) \in M$ bei linearen Nebenbedingungen und grossen p, was für $q > 1$ i.a. falsch ist.

Der Konvergenzbeweis benötigte $f, g_i \in C^1$ *nicht*. Allerdings wird die Hilfsfunktion H_p andernfalls nicht-differenzierbar.

Probleme bei der Anwendung von Strafmethoden: Prinzipiell kann man jede unbeschränkte Folge p_k benutzen, um zugehörige Lösungen x_k zu konstruieren. Will man aber x_k als guten Startpunkt zur Berechnung von x_{k+1} benutzen (was zweckmässig wäre), darf sich H_p in den Parameterwerten p_k, p_{k+1} nicht sehr unterscheiden. Also sollte p_k langsam wachsen. Andererseits hat dann x_k wenig mit den Lösungen von (4.0.1) zu tun, und man braucht viele Schritte, um in die Nähe der Restriktionsmenge M zu gelangen.

6.7 Barrieremethoden

Bei der Strafmethode hat man es i.a. mit unzulässigen Punkten der Aufgabe (4.0.1) zu tun, während der folgende Ansatz für dieselbe Aufgabe mit SlaterPunkten arbeitet und deren Existenz voraussetzt; $M_0 := \{x \mid g_i(x) < 0 \ \forall i = 1, 2, ..., m\} \neq \emptyset$.
Für $p > 0$ und SlaterPunkte $x \in M_0$ definiere man die Funktion

$$B_p(x) = f(x) \ - p \sum_i \ln(-g_i(x)).$$

Sie wird beliebig gross für Punkte nahe dem Rand von M_0,

$$B_p(x) \to \infty \text{ wenn } g_i(x) \to -0 \tag{6.7.1}$$

und ist auf M_0 differenzierbar:

$$DB_p(x) = Df(x) \ + \ p \sum_i \frac{Dg_i(x)}{-g_i(x)}.$$

Nun bestimmt man Minimalpunkte $x(p)$ für B_p und hofft, dass diese für

$$p = p_k \to +0$$

gegen eine Lösung von (4.0.1) konvergieren (wieder zumindest nach Auswahl einer passenden Teilfolge). Die Zahlen $y_i = -\frac{p}{g_i(x)}$ approximieren dann die Lagrange Multiplikatoren.

Erneut ist Konvexität von f und g_i die entscheidend. Z.B. schon für den Nachweis dafür, dass jedes $\hat{x} \in M$ (also insbesondere Lösungen von (4.0.1)) ein Limes von Punkten aus M_0 ist, d.h., $M \subset \operatorname{cl} M_0$. Hinzu kommt die Vorauss., dass M beschränkt ist.
Auf die Aufgabe

$$\min \{B_p(x) \mid x \in M_0\}, \quad p > 0 \ \text{ fest}$$

kann man wieder Verfahren der freien Minimierung anwenden, falls sie wie üblich Punkte mit fallenden Funktionswerten erzeugen. Dadurch wird (bei kleinen Schrittweiten) wegen der *Barriere-Eigenschaft* (6.7.1) verhindert, dass die Iterationspunkte (bei jedem festem $p > 0$) gegen den Rand von M_0 streben. Methoden diese Typs heissen auch *Innere Punkt Methoden*.

Abschätzungen der Lösungen für Straf-und Barriere-Methoden können einheitlich über Stabilität des KKT-Systems erfolgen. Das ist in der Standardliteratur nicht bekannt.

s. sect. 8.4 and [44], [74], [86].

6.8 Ganzzahligkeit und "Branch and Bound"

Die Standardaufgabe für nichtlineare Optimierung in endlicher Dimension hat die Form

$$\min\{ \ f(x) \mid x \in M\}, \tag{6.8.1}$$

wobei $M = \{ \ x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ i = 1, 2, ..., m\}$ und $f, g_i \in C^1(\mathbb{R}^n, \mathbb{R})$.
Dieses Modell beinhaltet auch Gleichungs- Nebenbedingungen; für $h(x) = 0$ schreibe $h(x) \leq 0$ und $-h(x) \leq 0$.
Es kann sogar Ganzzahligkeit gewisser Variabler, etwa x_1, dadurch beschreiben, dass man die Bedingung $1 - \cos(2\pi x_1) \leq 0$ oder auch $x_1(1 - x_1) = 0$ hinzu nimmt. Dadurch löst

man zwar keins der inhaltlichen Probleme der diskreten Optimierung, doch sieht man, dass trotz der Beschreibung mittels C^∞ Funktionen die Menge M eine komplizierte Struktur haben kann. Deshalb ist es meist sinnvoller, solche Bedingungen direkt auszunutzen. Eine oft angewandte Methode besteht in sogenannten "branch and bound" Algorithmen.

Angenommen, in (6.8.1) sollen die Variablen $x_1, ..., x_p$ ganzzahlig sein. M sei beschränkt, so dass zur wir Beschreibung von x_j höchstens q von Ziffern $0, 1$ brauchen. Wir können dann weiter annehmen, dass wir es mit pq Variablen aus $\{0, 1\}$ zu tun haben. Und zur Vereinfachung, werden wir so tun, als wäre schon für die p Variablen x_j jeweils $x_j \in \{0, 1\}$ verlangt. Schliesslich nehmen wir an, dass die Aufgabe ohne Ganzzahligkeit, aber mit den (Intervall-) Bedingungen $x_j \in [0, 1], 1 \leq j \leq p$ gelöst werden kann.

Idee der branch and bound Methode:
Für jede $0-1$ Kombination P von $x_1, ..., x_p$ sei $F(P)$ der optimale Zielfunktionswert des Problems mit $x_j = P_j$, d.h.,

$$F(P) = \min\{\ f(x) \mid x \in M,\ x_j = P_j \text{ für } 1 \leq j \leq p\}, \tag{6.8.2}$$

so dass unser Ziel die Minimierung von $F(P)$ in Bezug auf alle $0-1$ Vektoren P ist.

Wir betrachten nun die Kombinationsmöglickeiten der ersten p Variablen in (heuristisch) sinnvoller Weise.

Wir setzen zuerst $x_1 = 0$, dann $x_1 = 1$ und berechnen die dazu gehörenden Minimalwerte $v(0)$ und $v(1)$ von Problem (6.8.1) mit Zusatzforderung $x_j \in [0, 1]$ (also im ganzen Intervall) für $1 < j \leq p$. Offenbar sind $v(0)$ und $v(1)$ untere Schranken für den Optimalwert des gemischt-ganzzahligen Originalproblems.

Ist $v(0) < v(1)$, scheint es sinnvoll zu sein, zunächst $x_1 = 0$ beizubehalten und die beiden Möglichkeiten für x_2 damit zu kombinieren.

Also berechnen wir die entsprechenden Optimalwerte $v(0, 0), v(0, 1)$, für welche $x_1 = 0$ und x_2 fixiert sind und für $2 < j \leq p$ wieder das ganze Intervall zugelassen ist. Mit anderen Worten, wir *verzweigen* die Möglichkeit $x_1 = 0$.
Damit kennen wir $v(0), v(1), v(0, 0), v(0, 1)$ und wissen, dass

$$v(0) \leq v(0, 0),\ v(0) \leq v(0, 1)$$

gilt. Die Zahlen $v(0, 0), v(0, 1)$ verbessern also unsere Abschätzung des Optimalwertes bzgl. $x_1 = 0$, während $v(1)$ in keiner Beziehung zu den anderen Grössen steht.

Im weiteren streichen wir die verbesserte Schranke $v(0)$, wählen die kleinste der Grössen $v(1), v(0, 0), v(0, 1)$ und variieren dort die entsprechende nächste Variable, *verzweigen* also erneut.
"Gewinnt" $v(1)$, würden wir $v(1, 0), v(1, 1)$ berechnen und $v(1)$ aus der Liste unserer v-Werte streichen, "gewinnt" $v(0, 0)$, müssten wir $v(0, 0, 0), v(0, 0, 1)$ berechnen und $v(0, 0)$ streichen; analog für $v(0, 1)$.

Allgemein passiert also folgendes: In der aktuellen Liste der v-Werte wählen wir einen (den) kleinsten, streichen ihn aus der Liste und verzweigen wie oben mittels der nächste ganzzahlige Variablen. Die Liste wird so um einen Eintrag grösser.

Der Prozess bricht ab, wenn in einem minimalen v-Wert schon *alle* $0-1$-Variablen x_j irgendwie fixiert sind; $x_j = P_j$. Dann ist die entsprechende untere Schranke identisch mit dem optimalen Zielfunktionswert für den zugehörenden $0-1$ Vektor, $v(P) = F(P)$. Dieser Vektor ist optimal, weil schon die unteren Schranken v der Optimalwerte für alle übrigen Kombinationen höchstens grösser als $F(P)$ sind.

Das Vorgehen heisst auch *Branch and Bound* (= Verzweigung und Schranke) Methode. Unsere Argumente hingen nur von drei Dingen ab:
1) Die Zahlen v bilden untere Schranken der jeweiligen Optimalwerte (die durch Minimierung von $F(P)$ entstehen, wenn man die ersten Komponenten von P auf das Argument von v festlegt).
2) sie sind nicht-fallend bei weiterer Verzweigung und
3) sie stimmen mit dem wirklichen Wert $F(P)$ der zu minimierenden Funktion überein, wenn alle $0-1$ Variablen fixiert sind.

Die Methode kann unter obigen Bedingungen für die Minimierung jeder Funktion F, definiert für $0-1$ Vektoren fester Dimension, angewandt werden. Allerdings braucht man dazu eine Funktion v (möglichst guter) unterer Schranken.

Im ungünstigsten Fall muss man alle $0-1$ Kombinationen durchmustern, bevor man eine Lösung findet, andererseits ist die B-B-Vorgehensweise oft zweckmässig.

6.9 Second-order methods

As in the case of free minimizers, methods of second order can be applied to problems with involved constraints. Then we are looking for solutions of the KKT- system. Due to the involved inequalities such methods are often described with considerable formal effort.

We shall deal with them after certain useful reformulations of the KKT-system in form of equations and show that (nonsmooth) Newton methods are applicable as in the standard case

[under the canonical suppositions (6.2.11) and (6.2.12) in section 6.2]
after the "right reformulations", see chapter 8, in particular sect. 8.5.

Chapter 7

Stability: Motivations and first results

In many situations, one needs informations about the behavior of solutions to problems which depend on certain parameters. So assume that we study, with some parameter $t \in T$ (some Euclidean space), problems $P(t)$ with feasible sets $M(t)$ as

$$P(t): \qquad \min f(x,t) \quad s.t. \;\; x \in \mathbb{R}^n, \; g(x,t) \leq 0 \in \mathbb{R}^m, \; h(x,t) = 0 \;\in \mathbb{R}^{m_h} \tag{7.0.1}$$

[$f, g, h \in C^1$ or better] and the related KKT-system (Thm. 4.1.2)

$$\begin{array}{rl} D_x f(x,t) + D_x g(x,t)^T y + D_x h(x,t)^T z & = 0; \\ h(x,t) = 0, \; g(x,t) \leq 0, \; y \geq 0, \; \langle y, g(x,t) \rangle & = 0. \end{array} \tag{7.0.2}$$

Let $s(t)$ denote a solution we are interested in (extreme-value, minimizer or KKT point) and suppose that s is involved in some other problem. This means we have to study

7.1 Two-stage optimization and the main problems

Let a solutions $s(t)$ of $P(t)$ be involved in some other optimization problem, say in

$$\min \{ \; H(s,t) \;\; | \;\; G(s,t) \leq 0, \quad s = s(t) \text{ solves } P(t) \; \}. \tag{7.1.1}$$

This is one of various examples for *hierarchic optimization models* which arise as so-called multi-level (or multi-phase, multi-stage) problems. For other models (which appear in design- or semi-infinite optimization) and related solution methods we refer to [28], [114] and [18]. Continuity results for optimal values and solutions of $P(t)$ under weaker hypotheses than here can be found in [7].
With more than one problem $P(t)$, we obtain more variables in (7.1.1) only. If also P includes variables which are solutions of other problems, one speaks about multi-level problems. This does not change the general situation. So we restrict us to the current model.

7.1.1 The simplest two-stage problem

Let H and G be sufficiently smooth in what follows. Problem (7.1.1) requires to minimize

$$\phi(t) = H(s(t),t) \quad s.t. \;\; \gamma(t) := G(s(t),t) \leq 0 \tag{7.1.2}$$

where $s = s(t)$ is some more or less complicated implicit function, given by the KKT conditions. If all $s(t)$ are unique and $s(.) \in C^1$, then

$$\frac{d\phi}{dt} = H_s \circ Ds + H_t, \qquad \frac{d\gamma}{dt} = G_s \circ Ds + G_t \tag{7.1.3}$$

and the optimality conditions (KKT-system) for (7.1.1) attain the form - with derivatives and values at the (searched) arguments (s, t),

$$H_s \circ Ds + H_t \ + \ (G_s \circ Ds + G_t)^T \lambda = 0, \ \lambda \geq 0, \ G \leq 0, \quad \langle \lambda, G \rangle = 0 \tag{7.1.4}$$

and with the additional requirement that s solves $P(t)$. Suppose the latter can be written as equation

$$F(s, t) = 0, \quad F \in C^1 \tag{7.1.5}$$

where s and $F(s, t)$ have the same dimension and satisfy the assumption of the implicit function theorem. Then we obtain $Ds = -(F_s)^{-1} F_t$ and

$$\begin{array}{rcl} -(H_s + G_s^T \lambda) \ (F_s)^{-1} D_t F \ + \ H_t + G_t^T \lambda & = & 0, \\ F = 0, \quad \lambda \geq 0, \quad G \leq 0, \quad \langle \lambda, G \rangle & = & 0. \end{array} \tag{7.1.6}$$

Thus, if $Ds(.)$ is smooth enough, problem (7.1.1) is a usual optimization problem with KKT system (7.1.6).

However, even if the original solution $s := x = x(.)$ is unique and continuous then, as a rule, this function has kinks (the same for the whole KKT-vector s), and the critical value $s := f(x(.))$ is not C^2. The bad situations occur when x changes the "faces" of the parametric feasible set or, in other words, if the set of active inequalities changes.

The behavior of KKT points $s(t)$ has been described by characterizing the possible singularities if both the involved functions are C^3 and the whole problem belongs to some generic class (then certain singularities cannot occur), cf. [55, 56]. But even linear problems are not automatically in this class, and we study less smooth problems, in general.

7.1.2 The general case

If (7.1.5) describes the solutions of $P(t)$, but the crucial matrix F_s is singular, then we are in an unpleasant situation which, however, has been studied by the classical analysis under various aspects.
The main problems occur if we have not such descriptions at all or F is not C^1. Then neither the Newton method nor the usual implicit function theorem can be directly applied. The difficulty consists in the fact that the standard formula

$$Ds(0)\tau = -(F_s(\bar{s}, 0))^{-1} F_t(\bar{s}, 0)\tau$$

must be replaced. If parameter t varies, we may have many or no solutions $s(t)$, and their dependence on directions τ is no longer linear, in general. So $s(t)$ becomes a set $S(t)$

$$\Gamma := -(D_s F(\bar{s}, 0))^{-1}$$

becomes a multifunction, and the old directional derivative $Ds(0)\tau$ is composed by this multifunction Γ (which is now a "derivative" of S at the origin) and a function, namely $\tau \mapsto F_t(\bar{s}, 0)\tau$ (if the inputs are smooth enough).

Knowing this, also the formulation of problem (7.1.2) is not evident, even more its necessary (and sufficient) optimality condition.

The analysis of Γ and of the (now multivalued) mapping S is a main reason for dealing with stability (= continuity properties) of multifunctions and of nonsmooth functions and their minimizers, cf. chapter 9. The best we can hope is to obtain a locally Lipschitz or piecewise smooth *function* S and *piecewise linear* Γ. Then ϕ and γ in (7.1.2) are C^1 transformations of S, too.

Thus we need appropriate optimality conditions and tools to handle functions which are Lipschitz or piecewise smooth (which is better) or are even multifunctions. Before, however, let us have a look at the best situation.

7.2 KKT under strict complementarity

Consider the KKT conditions as in Thm. 4.1.2.

Definition 7.2.1 Strict complementarity holds at a KKT point $\bar{s} = (\bar{x}, \bar{y}, \bar{z})$ if (for all i), $\bar{y}_i = 0 \;\Rightarrow\; g_i(\bar{x}) < 0$. ◇

In other words, $g_i(\bar{x}) = 0$ yields $\bar{y}_i > 0$; if constraint g_i is active at $\bar{x}$ then $\bar{y}_i > 0$.

Let strict complementarity be satisfied at $\bar{s}$ and put $I_0^+ = \{i \mid \bar{y}_i > 0\}$.

Then, for all KKT points $s = (x, y, z)$ sufficiently close to $\bar{s}$, the same y_i remain positive and the same $g_i(x)$ are negative. This means that KKT points are - locally around $\bar{s}$ - simply the solutions of the system

$$\begin{array}{lll} Df(x) + \sum_{i \in I_0^+} y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x) & = 0 & \\ g_i(x) & = 0 & \forall i \in I_0^+ \\ h(x) & = 0 & \end{array} \tag{7.2.1}$$

($y_i = 0$ if $i \notin I_0^+$). This is a usual nonlinear system $F(x, y, z) = 0$ with the same number of variables and conditions. For

$$p = (a, b, c)^T \in \mathbb{R}^{n+m+m_h}$$

with small norm (where we may delete all the b_i, $i \notin I_0^+$ due to $g_i(x) < b_i$), the system

$$F(x, y, z) = p \tag{7.2.2}$$

describes, for $s = (x, y, z)$ near $\bar{s}$, exactly the KKT points of the (canonically) perturbed problem

$$P(a, b, c): \qquad \min \{\, f(x) - \langle a, x\rangle \mid g(x) \le b \text{ and } h(x) = c \,\}. \tag{7.2.3}$$

In consequence, if the involved functions are C^2, then F is C^1 and we may

(i) solve $F = 0$ by Newton's method if some initial point sufficiently close to $\bar{s}$ is known,

(ii) study the KKT points of $P(a, b, c)$ by applying the usual inverse-function theorem to (7.2.2).

(iii) extend the analysis of (7.2.2) to system $F(x, y, z, t) = 0$ by the implicit function theorem, to analyze the KKT points of nonlinearly perturbed problems $P(t)$, (7.0.1).

In view of (i) note that, if $s = (x, y, z)$ is sufficiently close to $\bar{s}$, the set $\{i \mid y_i + g_i(x) > 0\}$ coincides with I_0^+. Hence I_0^+ is known though $\bar{s}$ may be unknown. The resulting Newton method is also called the Wilson-method.

For every purpose, we need only one crucial hypothesis, namely regularity of the Jacobian [which may also depend on a - now fixed - parameter t]

$$DF(x,y,z) = \begin{pmatrix} D_x^2 L(x,y,z) & Dg_{I^+}(x)^T & Dh(x)^T \\ Dg_{I^+}(x) & 0 & 0 \\ Dh(x) & 0 & 0 \end{pmatrix} \tag{7.2.4}$$

at $\bar{s} = (\bar{x}, \bar{y}, \bar{z})$;

Dg_i is taken as a row, Dh is of type (m_h, n), I^+ denotes the components of I_0^+.

These observations have been used for a first systematic local sensitivity analysis of NLO by A.V. Fiacco [30], based on the usual implicit function theorem.

Theorem 7.2.1 *(canonical perturbations) Suppose that $f, g, h \in C^2$, strict complementarity holds at a KKT point $\bar{s}$, and $DF(\bar{s})$ is regular. Then there are $\varepsilon, \delta > 0$ such that, whenever $\|(a,b,c)\| < \delta$:*

The problems $P(a,b,c)$ have unique KKT points $s(a,b,c)$ in the ball $B(\bar{s}, \varepsilon)$,

$s(.)$ is C^1 on a ball around the origin and

$$s(0) = \bar{s}, \quad Ds(0)(\alpha,\beta,\gamma)^T = -DF(\bar{s})^{-1}(\alpha,\beta,\gamma)^T. \qquad \diamond$$

Proof. The statements are direct consequences of the implicit function theorem. □

Then $s(.)$ is also loc. Lipschitz since the norm of the Jacobian of $s(.)$ remains bounded by some C which implies $\|s(p') - s(p'')\| \leq C \ \|p' - p''\|$ (mean-value theorem) on the ball in question.

Theorem 7.2.2 *(nonlinear perturbations) The assumptions of Thm. 7.2.1 (with derivative for fixed $t = 0$) ensure for problem $P(t)$ that $s(.)$ is C^1 near the origin and*

$$s(0) = \bar{s}, \quad Ds(0)\tau = -F_s(\bar{s}, 0)^{-1} F_t(\bar{s}, 0)\tau. \qquad \diamond$$

Hence (α, β, γ) is replaced by $F_t(\bar{s}, 0)\tau$ only.

Remark 7.2.3 Singularity of $A := DF(\bar{s})$, (7.2.4):
Let V be the space, generated by all gradients $Dh_j(\bar{x}), Dg_i(\bar{x})$, $i \in I^+$.
Then, A is not regular $\Leftrightarrow$ The listed gradients are linearly dependent or $\exists u \neq 0, \ u \perp V$ such that $D_x^2 L(\bar{s})u \in V$.

Hence, *LICQ - even more MFCQ - is necessary* for regularity of A under strict complementarity at $\bar{s}$. $\diamond$

Since LICQ is necessarily satisfied, the Lagrange multipliers are unique and (if they exist for x near $\bar{x}$ and small variations of the parameters) are also Lipschitz functions of x and the parameters.

The remainder of the current and next chapter is devoted to the less pleasant (but for problems with inequality constraints typical) situation when strict complementarity does not hold. We consider approaches which allow us to investigate S and Γ in detail and shall (later) extend Newtons method to non-smooth functions. Special conclusions, possible by applying Kojima's description of KKT-points, will finish the last sections of chapter 8.

7.3 Basic generalized derivatives

7.3.1 CF and TF

Definition 7.3.1 Below, we shall use certain "directional limits" of a function $f : X \to Y$ at x in direction $u \in X$. They collect certain limits v of difference quotients, namely

$$Cf(x;u) = \{v \mid \exists u_k \to u, \quad t_k \downarrow 0 : \quad v = \lim t_k^{-1}\ [f(\ x + t_k u_k) - f(\ x)\ ]\ \},$$
$$Tf(x;u) = \{v \mid \exists (x_k, u_k) \to (x, u), \quad t_k \downarrow 0 : \quad v = \lim t_k^{-1}\ [f(x_k + t_k u_k) - f(x_k)\ ]\ \}.$$

The mapping Cf is said to be the *contingent derivative of* f. Alternatively, one can define

$$v \in Cf(x;u) \quad \text{ if } \ (u,v) \in C_{\operatorname{gph} f}(x, f(x)). \tag{7.3.1}$$

where $C_{\operatorname{gph} f}$ denotes the contingent cone to $\operatorname{gph} f$.
The limits of Tf were introduced by Thibault in [141, 142] (to define other objects) and called *limit sets*. They appeared in [67, 70, 83] (to study inverse Lipschitz functions) as $\Delta-$ or T- *derivatives*.
Evidently, $Cf(x;u) \subset Tf(x;u)$; and $Cf(x;u) = Tf(x;u) = \{Df(x)u\}$ if $f \in C^1$.

For multifunctions $F : X \rightrightarrows Y$ and $y \in F(x)$, define CF as: $v \in CF(x,y)(u)$ if

$\exists (u_k, v_k) \to (v, u)$ and $t_k \downarrow 0$ such that $(x + t_k u_k, y + t_k v_k) \in \operatorname{gph} F$. This is the same as $(u,v) \in C_{\operatorname{gph} F}(x,y)$ and corresponds to Cf for functions where $y + t_k v_k = f(x + t_k u_k)$. The same relation with Tf is valid, if one defines: $v \in TF(x,y)(u)$ if

$\exists (u_k, v_k, x_k, y_k) \to (v, u, x, y)$ and $t_k \downarrow 0$ such that $(x_k, y_k) \in \operatorname{gph} F$ and $(x_k + t_k u_k, y_k + t_k v_k) \in \operatorname{gph} F$. This defines a (generally larger) set called *strict graphical derivative* in [134]. ◇

The general definition of TF has been applied (up to now) only to mappings which can be linearly transformed into functions $f \in C^{0,1}$, cf. sect. 12.3 and [91], [92]. Contingent derivatives, successfully applied in [5], will play a role for all questions below; Tf plays a crucial role for Lipschitz properties of f^{-1}.

Remark 7.3.1 If $f \in C^{0,1}(X, \mathbb{R}^n)$ (locally Lipschitz), $Cf(x;u)$ and $Tf(x;u)$ are nonempty and compact, and one may put $u_k \equiv u$ in Def. 7.3.1, without changing these sets. ◇

In what follows, we shall also write $Cf(x;u) = Cf(x)(u)$ and $Tf(x;u) = Tf(x)(u)$.

7.3.2 Co-derivatives and generalized Jacobians

Let $F : X = \mathbb{R}^n \rightrightarrows Y = \mathbb{R}^m$ and $\bar{y} \in F(\bar{x})$.

Definition 7.3.2 *Co-derivative* D^*F: Explicitly, $D^*F(\bar{x}, \bar{y}) : Y^* \rightrightarrows X^*$ is defined by $x^* \in D^*F(\bar{x}, \bar{y})(y^*)$ if $\exists\ \varepsilon_k \downarrow 0,\ \delta_k \downarrow 0$ and points $(x_k, y_k) \to (\bar{x}, \bar{y})$ in $\operatorname{gph} F$ such that

$$\langle y^*, \eta \rangle + \varepsilon_k \|(\xi, \eta)\| \geq \langle x^*, \xi \rangle \text{ whenever } (x_k + \xi, y_k + \eta) \in \operatorname{gph} F \text{ and } \|(\xi, \eta)\| \leq \delta_k. \tag{7.3.2}$$

cf. Mordukhovich [105, 106]. ◇

If (x_k, y_k) is isolated in $\operatorname{gph} F$ then (7.3.2) holds trivially for small δ_k. Replacing (x^*, y^*) by any $(x_k^*, y_k^*) \to (x^*, y^*)$ is possible, but does not change the definition.

Using sect. 9.1.4, the pairs $(x^*, -y^*)$ with $x^* \in D^*F(\bar{x}, \bar{y})(y^*)$ are limiting F-normals or, being the same by Thm. 9.1.4, limiting ε-normals. Recall: $(x^*, -y^*)$ is (locally) an *approximate* ε_k- *normal* at (x_k, y_k) to $\operatorname{gph} F$ if (put $x = x_k + \xi,\ y = y_k + \eta$):

$$\langle\ (x^*, -y^*), (x - x_k, y - y_k)\rangle\ \leq\ \varepsilon_k\ \|(x,y) - (x_k, y_k)\|\ \ \forall (x,y) \in \operatorname{gph} F \text{ near } (x_k, y_k).$$

For $y^* = e_1$, only component y_1 of $(y,x) \in \operatorname{gph} F$ near $(\bar{y},\bar{x})$ plays any role: $y_1 - \bar{y}_1$ should be miniorized (approximately) by the related elements $\langle x^*, x - x_k\rangle$. For $y^* \neq e_1$, the same holds with some new basis of Y, cf. the note after Thm. 11.1.5.

Remark 7.3.2 D^*F can be written in terms of CF since (7.3.2) means with new $\varepsilon_k \downarrow 0$,

$$\langle y^*,\ v\rangle + \varepsilon_k \geq \langle x^*, u\rangle \ \text{ whenever } \ v \in CF(x_k,y_k)(u) \text{ and } \|(v,u)\| \leq 1. \tag{7.3.3}$$

Thus (in finite dimension) D^*F is not based on normals being *independent of tangents*, as said in some papers. ◇

Definition 7.3.3 *Generalized Jacobian* $\partial^c f(x)$ (Clarke) [12, 13]: For $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$, put

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

Then $M \neq \emptyset$ holds by Rademacher's theorem (f is a.e. F-differentiable; for a proof see [29]). The convex hull $\partial^c f(x) = \operatorname{conv} M$ is Clarke's generalized Jacobian of f at x. The set M is sometimes called B-derivative of f at x. ◇

Example 7.3.1 For both functions, $f = |x|$ and $f = -|x|$, one obtains $\partial^c f(0) = [-1,1]$.

Definition 7.3.4 *Strict differentiability*: If $f : X \to Y$ and $Tf(\bar{x})(u)$ is single-valued for all directions u, one says that f is *strictly differentiable* at $\bar{x}$. ◇

In other words, there exists a unique limit for all difference quotients

$$t_k^{-1}\ [\ f(x_k + t_k u_k) - f(x_k)\ ] \quad \text{as } x_k \to \bar{x},\ u_k \to u \text{ and } t_k \downarrow 0.$$

Then $Tf(\bar{x})(u) = \{Df(\bar{x})u\}\ \forall u$. Every C^1 function is strictly differentiable. The reverse is not true: f may have (vanishing) kinks at certain points $x_k \to \bar{x}$.

Before justifying these definitions by statements on stability, some technical comments concerning chain rules are useful.

7.4 Some chain rules

The defined generalized derivatives do not change after adding a function h with

$$h(\bar{x}) = 0 \ \text{ and } \ Th(\bar{x})(u) = \{0\}\ \forall u. \tag{7.4.1}$$

Furthermore, if

$$F, G : X \rightrightarrows P$$

are two mappings such that the *Hausdorff-distance* of the images

$$d_H(F(x), G(x)) := \inf\{\alpha > 0 \mid F(x) \subset G(x) + \alpha B \text{ and } G(x) \subset F(x) + \alpha B\}$$

satisfies

$$d_H(F(x), G(x)) \leq \|h(x)\|, \quad h \text{ from (7.4.1)} \tag{7.4.2}$$

then the introduced derivatives of F and G at $(\bar{x},\bar{y})$ remain the same (replace the elements $y_k, v_k, \eta, ...$ which appear in the related derivative for F by corresponding (nearby) elements of the $G-$images and vice versa).

7.4.1 Adding a function

$h \in C^1$ (or if $Dh(\bar{x})$ exists as strict derivative) one easily shows (directly) that $G = h + F$ fulfills

$$\begin{array}{lcl} C\ G(\bar{x}, h(\bar{x}) + \bar{y}) & = & Dh(\bar{x})\ +\ CF(\bar{x}, \bar{y}), \qquad \text{the same for } TF, \\ D^*G(\bar{x}, h(\bar{x}) + \bar{y}) & = & Dh(\bar{x})^* + D^*F(\bar{x}, \bar{y}). \end{array} \tag{7.4.3}$$

7.4.2 Inverse mappings

Due to the symmetry with respect to images and pre-images, the derivative TS or CS of the inverse $S = F^{-1}$ is just the inverse of TF or CF, respectively. For D^*S, one has $y^* \in D^*S(\bar{z})(x^*)$ if the elements in (7.3.2) satisfy $\varepsilon_k \|(\xi, \eta)\| \geq \langle y^*, \eta \rangle - \langle x^*, \xi \rangle$. So, compared with D^*F, the elements y^* and x^* change the place and the sign. This tells us

$$\begin{array}{rcl} -x^* \in D^*F(\bar{x}, \bar{y})(-y^*) & \Leftrightarrow & y^* \in D^*S(\bar{y}, \bar{x})(x^*), \\ v \in CF(\bar{x}, \bar{y})(u) & \Leftrightarrow & u\ \in CS(\bar{y}, \bar{x})(v), \qquad \text{the same for } TF. \end{array} \tag{7.4.4}$$

Thus the "derivative" of the inverse can be determined if the original "derivative" is known. However, from linear functions we know that computing $v = Au$ is trivial, not so computing u for given v. In our context, computing u will often require to solve a linear complementarity (or quadratic optimization) problem. Notice also that, even for functions, the inverse is generally a multifunction.

7.4.3 Derivatives of the inverse $S = (h + F)^{-1}$

This mapping assigns, to $y \in Y$, the solutions x of $y \in h(x) + F(x)$. Again, if $Dh(\bar{x})$ exists as strict derivative, the above formulas can be combined in order to obtain, for $\bar{x} \in S(\bar{y})$, $\bar{y} - h(\bar{x}) = q^0$ and $q^0 \in F(\bar{x})$, $\bar{z} = (\bar{y}, \bar{x})$,

$$\begin{array}{lcl} u \in CS(h(\bar{x}) + \bar{y}, \bar{x})(v) & \Leftrightarrow & v\ \in\ [Dh(\bar{x})\ +\ CF(\bar{x}, \bar{y})](\ u), \quad \text{the same for } TF, \\ y^* \in D^*S(h(\bar{x}) + \bar{y}, \bar{x})(x^*) & \Leftrightarrow & -\,x^*\ \in\ [D^*h(\bar{x}) + D^*F(\bar{x}, \bar{y})](-y^*). \end{array} \tag{7.4.5}$$

7.4.4 Composed mappings

For a composed mapping $G(x) = g(F(x))$ (and related dimensions) we have

$$TG(x)(u) = Dg(y) \circ TF(x, y)(u) \qquad (g \in C^1, \text{ the same for } CF). \tag{7.4.6}$$

Conversely, for $G(x) = F(g(x))$, it holds generally only

$$TG(x)(u) \subset TF(x, g(x))(Dg(x)u) \qquad (g \in C^1, \text{ the same for } CF). \tag{7.4.7}$$

Example 7.4.1 Put $g(x_1, x_2) = (x_1, 0)$, $F(x_1, x_2) = \sqrt{|x_2|}$, $G = F \circ g$. Then $G \equiv \{0\}$, but the limit

$$1 = \lim t^{-1}(\sqrt{0 + 0t + |t^2|} - \sqrt{0})$$

belongs to $CF(0,0)(1,0)$ and $(1,0) = Dg(0)u$ for $u = (1,1)$. ◇

7.4.5 Linear transformations or diffeomorphisms of the space

Regular linear transformations (or diffeomorphisms) of the image- or pre-image space change the derivatives C, T, D^* in the same manner as DF and the usual adjoint map D^*F, respectively. So, it holds for any linear function $A : P \to Y$ (and for $A \in C^1$ with derivatives at the related points)

$$v \in CF(\bar{x}, \bar{y})(u) \Rightarrow Av \in C(AF)(\bar{x}, A\bar{p})(u), \quad \text{similarly for } T(AF). \tag{7.4.8}$$

Direction ($\Leftarrow$) is valid if A^{-1} exists: apply (7.4.8) with A^{-1} to the right-hand side. □
Regularity of A (or existence of $A^{-1} \in C^1$) plays a role for direction ($\Leftarrow$), indeed.

Example 7.4.2 $F(x) = 1/x,\ F(0) = 0$: With $A = 0$, it holds $0 \in C(AF)(0,0)(1)$ while $CF(0,0)(1) = \emptyset$. The same effect appears for TF. ◇

The given formulas ensure, under strict differentiability of $h : X \to P$ or regular linear transformations in X and P, that the derivatives of $S = h + F$ and S^{-1} are available iff they are known for F.

Remark 7.4.1 The derivatives C, T, D^* do not satisfy partial derivative rules for $f(x, y)$ or additivity of the derivatives to $f + g$ (put e.g. $f = |x|,\ g = -|x|$).

A deeper analysis of possible chain rules can be found in [36], [68] and [134].

7.5 Loc. Lipschitzian inverse and implicit functions

Let $\phi : X \to P$ (Banach spaces) be locally Lipschitz.

Definition 7.5.1 We call ϕ locally Lipschitzian invertible at $\bar{x}$ if $\exists\ \varepsilon, \delta > 0$ such that

$$\begin{gathered}\forall p \in B(\phi(\bar{x}), \delta)\ \exists x = s(p) \in B(\bar{x}, \varepsilon) \text{ (unique) satisfying } \phi(x) = p, \\ \text{and if, in addition, } s \text{ is Lipschitz on } B(\phi(\bar{x}), \delta).\end{gathered} \tag{7.5.1}$$

In other words, s is locally a Lipschitzian inverse to ϕ near $(\phi(\bar{x}), \bar{x})$. This is the same as ϕ^{-1} being strongly Lipschitz stable (s.L.s.), cf. Def. 10.1.

A function $\phi \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ is loc. Lipsch. invertible at $\bar{x} \Leftrightarrow D\phi(\bar{x})$ is injective (= regular or bijective in this case).

For $\phi \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$, it turns out that ϕ is loc. Lipsch. invertible at $\bar{x} \Leftrightarrow T\phi(\bar{x})$ is injective.
It is important that $T\phi(\bar{x})$ can be determined and the related condition can be directly applied to the KKT points of a C^2 optimization problem, Thm. 8.5.2 and sect. 12.1.

7.5.1 Inverse Lipschitz functions

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be loc. Lipschitz, $S = f^{-1}$ and $f(0) = 0$. By our definitions, f is loc. Lipsch. invertible at the origin iff S is s.L.s. at $(0, 0)$.

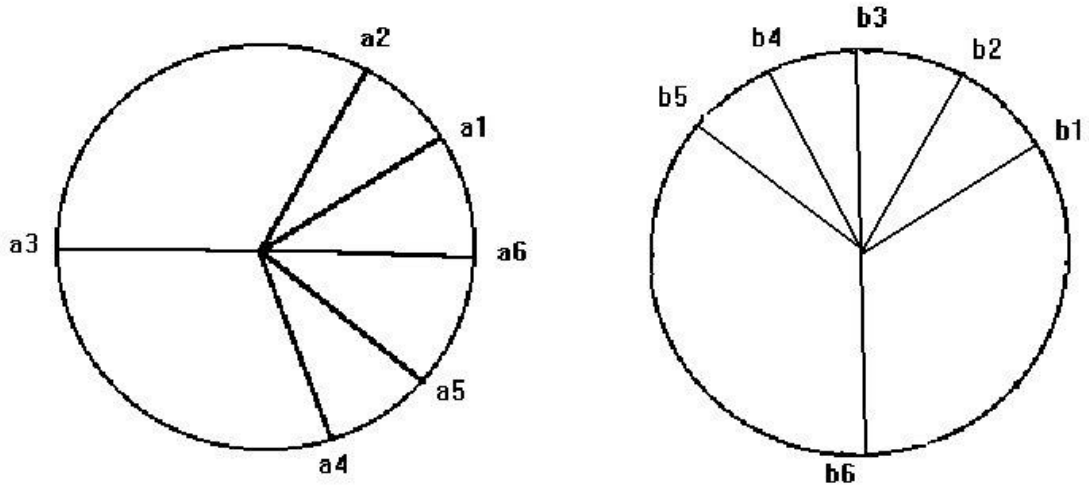
Theorem 7.5.1 *[12], [13]: S is s.L.s. at $(0,0)$ if all $A \in \partial^c f(0)$ are regular.* ◇

This can be shown by

Theorem 7.5.2 *[83], [84]: S is s.L.s. at $(0,0)$ $\Leftrightarrow 0 \notin Tf(0)(u)\ \forall u \in \mathbb{R}^n \setminus \{0\}$.* ◇

Proof. ($\Rightarrow$) Since S is s.L.s., it is strongly Lipschitz (which means that (7.5.1) holds at least for $p \in f(X)$). By Thm. 11.2.2 (ii), then $u \neq 0$ yields $0 \notin Tf(0)(u)$.
($\Leftarrow$) Again by Thm. 11.2.2 (ii), S strongly Lipschitz. Hence, if $\varepsilon > 0$ is small enough, then every $y \in Y_\varepsilon := f(B(0,\varepsilon))$ has exactly one pre-image $x = g(y)$ in $B(0,\varepsilon)$ and, moreover, g is Lipschitz on Y_ε. The restriction $f_\varepsilon = f|_{B(0,\varepsilon)}$ of f onto the ε ball is now continuous and has a continuous inverse $g = f_\varepsilon^{-1} : Y_\varepsilon \to B(0,\varepsilon)$. In other words, $B(0,\varepsilon)$ and Y_ε are homeomorphic. By Brouwer's theorem on invariance of domains, cf. [1], then $f(0) \in \operatorname{int} Y_\varepsilon$ holds true. Thus g is defined and Lipschitz for small $\|y\|$, so S is s.L.s. at the origin. □

Remark 7.5.3 For $n = 2$, there is a function f such that S is s.L.s. at $(0,0)$ and $0 \in \partial^c f(0)$, [83] or [70], BE3. f consists of 6 linear pieces which map the cones spanned by a_1, a_2 in the figure below onto the cone spanned by b_1, b_2 and so on. Hence the condition of Thm. 7.5.1 (Clarke) is not necessary for piecewise linear functions. ◇



Remark 7.5.4 [83, 70] $Tf(x)(u)$ is connected and $\{Au \mid A \in \partial^c f(x)\} = \operatorname{conv} Tf(x)(u)$. Thus Thm. 7.5.1 follows from Thm. 7.5.2.

For Kojima's function Φ, assigned to optimization problems in section 8.3 and $f, g, h \in C^{1,1}$, the derivatives $T\Phi$ and $C\Phi$ *can be determined* by a product rule, Thm. 8.3.2. $\diamond$

7.5.2 Implicit Lipschitz functions

Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ be loc. Lipschitz, $S(t,p) = \{x \in \mathbb{R}^n \mid f(x,t) = p\}$, $f(0,0) = 0$. Now (by definition), S is s.L.s. at $(0,0,0)$ iff the solutions x to $f(x,t) = p$ are locally unique and Lipschitz near the origin.

Theorem 7.5.5 *[84]: S is s.L.s. at $(0,0,0) \Leftrightarrow 0 \notin Tf(0,0)(u,0)\ \forall u \in \mathbb{R}^n \setminus \{0\}$. Moreover, in the s.L.s. case, it holds*

$$u \in TS(0,0)(\tau, v) \ \Leftrightarrow\ v \in Tf(0,0)(u,\tau). \qquad \diamond \tag{7.5.2}$$

With generalized Jacobians, the " $\Leftrightarrow$ characterizations" are not true. For the chain rule in terms of partial derivatives (as usually defined)

$$Tf(0,0)(u,\tau) = T_x f(0,0)(u) + T_t f(0,0)(\tau), \tag{7.5.3}$$

which needs additional hypotheses, and for $Tf(x, g(z))$, we refer again to [83] and [70].

The implicit function
Let $\phi(\bar{x}) = 0$. As in the context of the usual implicit function theorem, invertibility of ϕ can be used to investigate solutions x_u of

$$\phi(x) = u(x) \tag{7.5.4}$$

provided that $u : X \to P$ is a *small Lipschitz function* near $\bar{x}$; i.e., $\exists \gamma > 0$ with

$$\|u(x') - u(x)\| \le \alpha\ d(x', x)\ \ \forall x, x' \in B(\bar{x}, \gamma)\ \text{ and }\ \|u(\bar{x})\| \le \beta \tag{7.5.5}$$

where $\alpha + \beta$ is sufficiently small.

Theorem 7.5.6 *Let ϕ be loc. Lipsch. invertible at $\bar{x}$ and $\gamma \in (0, \min\{\varepsilon, \delta\})$. Then, if u fulfills (7.5.5) with sufficiently small $\alpha+\beta$, (7.5.4) has a unique solution $x_u \in B(\bar{x}, \gamma)$.* $\diamond$

Proof. Let L be a Lipschitz constant of s and let α, β satisfy $\beta + \alpha\gamma \le \min\{\varepsilon, \delta\}$, $L\alpha < \frac{1}{2}$ and $L\beta < \frac{1}{2}\gamma$. For $x \in B(\bar{x}, \gamma)$ then $\|u(x)\| \le \delta$ follows. Thus the function

$$T(x) = s(u(x)), \qquad x \in B(\bar{x}, \gamma)$$

is well-defined, has Lipschitz constant $L\alpha < \frac{1}{2}$ and maps $B(\bar{x}, \gamma)$ into itself since

$$\begin{aligned} &\|T(x) - \bar{x}\| = \|s(u(x)) - \bar{x}\| \ \le\ \|s(u(x)) - s(u(\bar{x}))\| + \|s(u(\bar{x})) - \bar{x}\| \\ &\le\ \ \alpha L\gamma \ +\ \|s(u(\bar{x})) - s(0)\| \ <\ \tfrac{1}{2}\gamma \ +\ L\beta \ <\ \gamma. \end{aligned}$$

The fixed points of T in $B(\bar{x}, \gamma)$ and the solutions of (7.5.4) coincide. So Banach's fixed point theorem yields the assertion. $\square$

Clearly, with the Lipschitz constant L for s, then also an estimate holds true:

$$\|x_u - \bar{x}\| \le L\|u(x_u)\| \le L\ (\beta + \gamma\alpha). \tag{7.5.6}$$

Extensions of these statements for multivalued $S = \phi^{-1}$ will be given chapter 11. Further conditions for loc. Lipsch. invertibility can be found (for KKT points) in 8.5 and 12.1.

Corollary 7.5.7 *To obtain the classical implicit function theorem for* $f(x,t)=0,\ f\in C^1$ *at* $(\bar{x},0)$, *put* $\phi(x)=D_xf(\bar{x},0)(x-\bar{x})$ *and* $u(x)=\phi(x)-f(x,t)$. $\diamond$

Proof. Indeed, one may estimate u via

$$\begin{aligned}-u(x) &= [\ f(x,t)-f(\bar{x},t)\ ]\ \ +\ \ f(\bar{x},t)-\phi(x)\\ &= \textstyle\int_0^1\ D_xf(\bar{x}+\theta(x-\bar{x}),\ t)\ (x-\bar{x})\ d\theta\ +\ f(\bar{x},t)-\phi(x)\\ &= \textstyle\int_0^1\ [D_xf(\bar{x}+\theta(x-\bar{x}),\ t)-D_xf(\bar{x},0)]\ (x-\bar{x})\ d\theta\ +\ f(\bar{x},t).\end{aligned}$$

Here, we have $\beta=\|f(\bar{x},t)\|$ and

$$\alpha:=\sup_{x\in B(\bar{x},\gamma)\ \theta\in[0,1]}\|D_xf(\bar{x}+\theta(x-\bar{x}),\ t)-D_xf(\bar{x},0)\|\to 0\quad \text{if } (\gamma,t)\to 0.$$

So, for γ sufficiently small, the hypotheses are fulfilled if $\|t\|$ is small enough in comparison to γ (in order to satisfy $L\beta<\frac{1}{2}\gamma$). $\square$

It was S.M. Robinson who had the splendid idea to see [124], [125] that, for $\phi\in C^1$,

(i) the classical interplay between the local inverse s(.) and the implicit function remains true if we replace (7.5.1) by a generalized equation $p\in\phi(x)+N(x)$ (with locally unique solution $s(p)$) and (7.5.4) by $u(x)\in\phi(x)+N(x)$ where N is multivalued.

(ii) with appropriate ϕ and N, one may equivalently describe the KKT points (of C^2 problems).

Chapter 8

KKT points as zeros of equations

In order to analyze the KKT points (x, y, z) for a NLP in finite dimension

$$\min f(x) \quad s.t. \;\; x \in M := \{x \in \mathbb{R}^n \mid g(x) \leq 0 \in \mathbb{R}^m \;\text{ and } h(x) = 0 \;\in \mathbb{R}^{m_h}\} \tag{8.0.1}$$

[where $f, g, h \in C^1$], several reformulations of the KKT- conditions (Thm. 4.1.2)

$$Df(x) + Dg(x)^T y + Dh(x)^T z = 0; \;\; h(x) = 0, \; g(x) \leq 0, \; y \geq 0, \; \langle y, g(x)\rangle = 0$$

are useful and possible for computing them as well as for studying its dependence on parameters.

By the subsequent approaches, KKT- points will be written as zeros of Lipschitzian equations or by the help of generalized equations. The common idea consists in appropriate descriptions of the complementarity conditions in the KKT system

$$g(x) \leq 0, \quad y \geq 0, \quad \langle y, g(x)\rangle = 0. \tag{8.0.2}$$

The common problem is the question of how local Lipschitzian invertibility can be ensured by *verfiable conditions* in terms of the original data.

8.1 (S. M. Robinson's) Generalized equations

There is a quite popular possibility of describing KKT-points, namely as solutions of inclusions (called generalized equations). The simplest one is the system

$$\begin{array}{rcl} Df(x) + \sum_i y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x) & = & 0 \\ g(x) & \in & N_K(y) \\ h(x) & = & 0, \end{array} \tag{8.1.1}$$

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

$$\begin{array}{rl} N_K(y) & = \{y^* \mid \langle y^*, k - y\rangle \leq 0 \;\forall k \in K\} \\ & = \{y^* \mid \; y_i^* = 0 \text{ if } y_i > 0 \;;\; y_i^* \leq 0 \text{ if } y_i = 0\}. \end{array} \tag{8.1.2}$$

If $y \notin K$ put $N_K(y) = \emptyset$. Defining

$$\hat{K} = \mathbb{R}^n \times K \times \mathbb{R}^{m_h} \text{ and similarly } N_{\hat{K}}(x, y, z),$$

system (8.1.1) can be written, with the left-hand side H from (8.1.1) and $s = (x, y, z)$, as

$$H(s) \in N_{\hat{K}}(s), \tag{8.1.3}$$

where H is a function and $N_{\hat{K}}$ a multifunction. Such systems have been introduced by S.M. Robinson who noticed (during the 70th) that the relations between system (8.1.3) and its linearization

$$H(\hat{s}) + DH(\hat{s})(s - \hat{s}) \in N_{\hat{K}}(s), \tag{8.1.4}$$

are (locally, and in view of inverse and implicit functions) the same as for usual equations. This was the start for various investigation of generalized equations (8.1.3) in different spaces with arbitrary multifunctions $\mathcal{N}$ (based on the same story). In the current situation, the solutions of the perturbed system

$$H(s) \in p + N_{\hat{K}}(s), \qquad p = (a, b, c)^T \tag{8.1.5}$$

describe just the KKT-points of the canonically perturbed problem (7.2.3). The same perturbation p in (8.1.4), i.e., system

$$H(\hat{s}) + DH(\hat{s})(s - \hat{s}) \in p + N_{\hat{K}}(s) \tag{8.1.6}$$

describes the KKT points of a related quadratic problem, namely

$$\min\{\tfrac{1}{2}(x - \bar{x})^T Q(x - \bar{x}) - \langle a, x - \bar{x}\rangle \mid g(\bar{x}) + Dg(\bar{x})(x - \bar{x}) \leq b, Dh(\bar{x})(x - \bar{x}) = c$$

with $Q = D_x^2 L(\bar{x}, \bar{y}, \bar{z})$.
As already used for Thm. 7.5.6, the usual hypothesis of the inverse function theorem,

"$DH(\hat{s})$ is regular"

now attains the form:

"The solutions $s(p)$ of the *perturbed linearized problem* (8.1.6)
are locally unique and Lipschitz".

This cannot be checked in the same simple manner as regularity of a matrix. For possible descriptions of the needed conditions, see the Thms. 8.5.2 and 12.1.1.
Locally unique and Lipschitz solutions are also called strongly stable, cf. sect. 10.1.

8.2 NCP- functions

Another well-known equivalent description of (8.0.2) consists in requiring

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

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

$$\begin{array}{rcc} \Theta_1 & = & Df(x) + \sum_i y_i Dg_i(x) + \sum_\nu z_\nu Dh_\nu(x), \\ \Theta_{2i} & = & \sigma(y_i, -g_i(x)) \\ \Theta_3 & = & h(x), \end{array} \tag{8.2.1}$$

the KKT- conditions (4.1.6) and

$$\Theta(x, y, z) = 0$$

(where $\Theta : \mathbb{R}^\mu \to \mathbb{R}^\mu$ and $\mu = n + m + m_h$) are equivalent. The equation

$$\Theta(x, y, z) = (a, b, c)^T$$

is connected with problem (7.2.3) but the transformations of the related solutions are more complicated than for the other reformulations. Also a product representation as below is not true for Θ. This shrinks the value of Θ for stability investigations, but not in view of solution methods. For more details we refer to [27, 31, 70, 140].

8.3 Kojima's function

The KKT-System (4.1.2) can be also written in terms of Kojima's [77] function $\Phi : \mathbb{R}^\mu \to \mathbb{R}^\mu$ which has the components

$$\begin{array}{rclcccll} \Phi_1 & = & Df(x) & + & \sum_i y_i^+ Dg_i(x) & + \sum_\nu z_\nu Dh_\nu(x), & y_i^+ = \max\{0, y_i\}, & \\ \Phi_{2i} & = & g_i(x) & - & y_i^-, & & y_i^- = \min\{0, y_i\}, & (8.3.1) \\ \Phi_3 & = & h(x). & & & & & \end{array}$$

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

$$\begin{array}{ll} (x, y, z) \in \Phi^{-1}(0) & \Rightarrow\ (x, u, z) = (x, y - g(x), z) = (x, y^+, z) \text{ is KKT-point} \\ (x, u, z) \text{ a KKT-point} & \Rightarrow\ (x, y, z) = (x, u + g(x), z) \in \Phi^{-1}(0) \end{array} \tag{8.3.2}$$

and Φ is, for $f, g, h \in C^2$, one of the simplest nonsmooth functions.

8.3.1 The product form

Moreover, Φ can be written as a (separable) product

$$\Phi(x, y, z) = \ \mathcal{M}(x)\ N(y, z) \tag{8.3.3}$$

$$\text{where} \qquad N = \ (1, \ \ y_1^+, ..., y_m^+, \ \ y_1^-, ..., y_m^-, \ z)^T \in \mathbb{R}^{1+2m+m_h}, \tag{8.3.4}$$

$$\mathcal{M}(x) = \begin{pmatrix} Df(x) & Dg_1(x)... & Dg_m(x) & 0... & 0... & 0 & Dh_1(x)... & Dh_{m_h}(x) \\ g_i(x) & 0 \ \ ... & 0 & 0... & -1... & 0 & 0 \ \ ... & 0 \\ h(x) & 0 \ \ ... & 0 & 0... & 0... & 0 & 0 \ \ ... & 0 \end{pmatrix} \tag{8.3.5}$$

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

Remark 8.3.1 Writing, e.g., $(y_i^3)^+, (y_i^3)^-$ instead of y_i^+ and y_i^-, KKT-points are even zeros of a smooth function. However, now $y_i = 0$ leads to a zero-column in the Jacobian of this modified function Φ. So standard tools for computing a zero or analyzing critical points via implicit functions fail again. Also the adapted formula (8.3.2) is no longer loc. Lipsch. in both directions. $\diamond$

Using Φ (8.3.1), the points of interest are zeros of a $\mathbb{R}^\mu \to \mathbb{R}^\mu$ function, and equation

$$\Phi(x, y, z) = (a, b, c)^T \tag{8.3.6}$$

permits a familiar interpretation: It describes by (8.3.2) the KKT-points of problem (7.2.3). Due to the structure of Φ and the simple type of non-differentiability, the generalized derivatives $T\Phi$ and $C\Phi$ (Def. 7.3.1 can be exactly determined for $f, g, h \in C^{1,1}$ (then Φ is loc. Lipschitz) by the product rule Thm. 8.3.2

$$\mathcal{D}\Phi(s)(u, v, w) = [\mathcal{D}M(x)(u)]\ N(y, z)\ +\ M(x)\ [\mathcal{D}N(y, z)(v, w)] \tag{8.3.7}$$

($\mathcal{D} = T$ or $\mathcal{D} = C$) and the right-hand side is defined by the element-wise operations:

$$\begin{array}{ll} [\mathcal{D}M(x)(u)]\ N(y,z) & = \{\ \mu\ N(y,z) \mid \mu \in \mathcal{D}M(x)(u)\ \}, \\ M(x)\ [\mathcal{D}N(y,z)(v,w)] & = \{\ \ M(x)\ \nu\ \ \mid \nu \in \mathcal{D}N(y,z)(v,w)\ \}. \end{array} \tag{8.3.8}$$

The situation $f, g, h \in C^{1,1} \setminus C^2$ is typical for multi-level problems which involve optimal values or solutions of other (sufficiently "regular") optimization models.
For $f, g, h \in C^2$, non-smoothness is *only* implied by the components of N:

$$\phi(y_i) = (y_i^+,\ y_i^-) = (y_i^+,\ y_i - y_i^+) = \tfrac{1}{2}\ (y_i + |y_i|,\ y_i - |y_i|). \tag{8.3.9}$$

So, Φ is a PC^1 function, and discussions on generalized derivatives can be reduced to "generalized derivatives" of the *absolute value* at the origin. Questions on stability of solutions can be reduced as in Thm. 11.2.2 to injectivity of $\mathcal{D}\Phi$ which means (as in the differentiable case) that

$$0 \in \mathcal{D}\Phi(\bar{s})(u,v,w)\ \Rightarrow (u,v,w) = 0 \tag{8.3.10}$$

holds at the crucial point $\bar{s} = (\bar{x}, \bar{y}, \bar{z})$. The final results in terms of the given functions are summarized in sect. 12.

Theorem 8.3.2 *(Product rule) Let $\Phi = MN$ be the Kojima function for $f, g, h \in C^{1,1}$. Then*

$$T\Phi(\bar{x},\bar{y},\bar{z})(u,v,w) = [TM(\bar{x})(u)]\ N(\bar{y},\bar{z})\ +\ M(\bar{x})\ [TN(\bar{y},\bar{z})(v,w)]. \tag{8.3.11}$$

The same holds for $C\Phi$. ◇

Proof. To begin with we set $\delta\Phi = \Phi(x+u', y+v', z+w') - \Phi(x,y,z)$ (similarly δN and δM are defined) and observe that

$$\Phi + \delta\Phi = (M + \delta M)(N + \delta N) = MN + M\ \delta N + \delta M\ N + \delta M\ \delta N$$

and

$$\delta\Phi = \ \delta M\ N(y,z) + \ M(x)\ \delta N + \ \delta M\ \delta N.$$

Now specify $(u', v', w') = t(u, v, w)$ for any given sequence $t = t_k \downarrow 0$. Then $\delta\Phi, \delta M$ and δN depend on t and

$$\frac{\delta\Phi(t)}{t} = \ \frac{\delta M(t)}{t} N(y,z)\ +\ M(x)\frac{\delta N(t)}{t}\ +\ \delta M(t)\frac{\delta N(t)}{t}$$

If, moreover, $(x,y,z) \to \bar{s} = (\bar{x},\bar{y},\bar{z})$ then - since M and N are loc. Lipsch. - the bounded sequences $\delta M(t)/t$ and $\delta N(t)/t$ have accumulation points $M_0 \in TM(\bar{x})(u)$ and $N_0 \in TN(\bar{y},\bar{z})(v,w)$, respectively. The third term $\delta N(t)\ \delta M(t)/t$ is vanishing. So we obtain, for all converging subsequences $\delta\Phi(t)/t$, that the limit can be written as

$$\lim\ \frac{\delta\Phi(t)}{t} = N(\bar{y},\bar{z})M_0 + N_0 M(\bar{x}).$$

Thus (8.3.11) holds as inclusion "$\subset$". To show "$\supset$", the structure of N is important. Let

$$M_0 \in TM(\bar{x})(u) \quad \text{and} \quad N_0 \in TN(\bar{y},\bar{z})(v,w)$$

be given, and let $t = t_k \downarrow 0,\ x = x(t) \to \bar{x}$ be sequences such that

$$M_0 = \lim t^{-1}(M(x+tu) - M(x)).$$

The existence of such sequences is ensured by definition of TM.

To show that $N(\bar{y},\bar{z})M_0+N_0M(\bar{x}) \in T\Phi(\bar{s})(u,v,w)$, it suffices to find elements $(y,z) = (y(t),z(t)) \to (\bar{y},\bar{z})$ such that N_0 can be written as

$$N_0 = \lim\ t^{-1}(N(y+tv,z+tw) - N(y,z)). \tag{8.3.12}$$

with the already given sequence of t. If this is possible then, considering $\delta\Phi(t)$ for $(x,y,z)+t(u,v,w)$ as above, we obtain

$$M_0\ N(\bar{y},\bar{z}) + M(\bar{x})\ N_0 = \lim\ \frac{\delta\Phi(t)}{t}\ \in\ T\Phi(\bar{s})(u,v,w),$$

which proves (8.3.11). 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 $N_0 \in TN(\bar{y},\bar{z})(v,w)$ belong to the map $z \to (1,z)$ and are obviously 0 and w, respectively. So one may put $z(t) = \bar{z}$. The other components of N_0 are the T-derivatives of the functions

$$y_i \to \phi(y_i) = (y_i^+, y_i^-) = (y_i^+, y_i - y_i^+)$$

at $\bar{y}_i$ in direction v_i. For $\bar{y}_i = 0$ (otherwise ϕ is C^1 near $\bar{y}_i$ and nothing must be verified) we have

$$T\phi(0)(v_i) = \{\ v_i\ (r_i,\ 1-r_i)\ |\ 0 \le r_i \le 1\}.$$

Hence the related components of N_0, say $(\alpha_i, v_i - \alpha_i)$, belong to this set. Consider now

$$c(y_i) := t^{-1}[\phi(y_i+tv_i) - \phi(y_i)] \in \mathbb{R}^2$$

for given $t > 0$. If $v_i > 0$ and y_i moves from $-tv_i$ to 0, the first component of c moves continuously from 0 to v_i. Hence $c_1(y_i) = \alpha_i$ holds for some $y_i(t,\alpha_i)$. The same holds for $v_i < 0$ if y_i moves from 0 to $t|v_i|$. Accordingly, there are $y(t)$ such that all quotients c_i attain the values $(\alpha_i, v_i - \alpha_i)$, required by N_0. Therefore, (8.3.12) can be satisfied (for the given t-sequence), and (8.3.11) is true.

For the contingent derivative $C\Phi$, this proof is simpler. The points (x,y,z) now coincide with $(\bar{x},\bar{y},\bar{z})$, and N is directional differentiable, so $CN(\bar{y},\bar{z})(v,w)$ is a singleton. □

8.3.2 Implicit KKT-points

For Kojima's function Φ and $f,g,h \in C^2$, all mentioned derivatives can be computed and the conditions of the Theorems 7.5.1 and 7.5.2 are equivalent. Thm. 7.5.5 is important for studying the problem

$$\min\ \{f(x,t) - \langle a,x\rangle\ |\ g(x,t) \le b,\ h(x,t) = c\}, \qquad p = (a,b,c). \tag{8.3.13}$$

with some parameter $t \in \mathbb{R}^{n'}$ for t near $t^0(=0)$ and small norm of p.
Let the involved functions be C^2 with loc. Lipsch. second derivatives. Then, (7.5.3) after Thm. 7.5.5 may be applied to $\Phi = \Phi(x,y,z,t) = M(x,t)N(y,z)$. Now $s = (x,y,z)$ stands for x in the statements on inverse functions.

This yields with derivatives and arguments at a zero $(\bar{s},t^0) = (\bar{x},\bar{y},\bar{z},t^0)$ of Φ:
The matrices A_r with the assigned feasible r in formula (8.5.1) below represent $T_{(x,y,z)}\Phi$, and for the derivative w.r.to the parameter, one has $D_t\Phi\ \tau = [D_tM\ \tau]\ N$.

If the KKT map of the canonically perturbed problem, $p \mapsto S(t^0,p)$ is s.L.s. at $(0,\bar{s})$ (i.e. all A_r are regular), we thus obtain

$$\begin{array}{l}(u,v,w) \in TS\ (\tau,\hat{p})\ \Leftrightarrow \hat{p} \in T_{(x,y,z)}\Phi\ (u,v,w) + D_t\Phi\ \tau \\ \Leftrightarrow (u,v,w) = A_r^{-1}\ (\hat{p} - D_t\Phi\ \tau)\ \text{ for some feasible } r\end{array} \tag{8.3.14}$$

which gives us concrete formulas since all A_r are know.
For the contingent derivative CS, the same formulas are valid (replace T by C), but now A_r^{-1} is no longer linear, cf. [18, 68]. For computing the derivative, now linear complementarity problems must be solved.

8.4 Relations to penalty-barrier functions

To reduce the formal effort for the next modification of Φ, let us once more delete the equations $h = 0$ in M. We consider the perturbed Kojima system

$$\Phi^t(x, y) = 0 \tag{8.4.1}$$

where only the terms y_i^- in $N(y)$ (8.3.4) are replaced by $y_i^- + t_i y_i^+$ with some perturbation parameter $t \in \mathbb{R}^m$ (small norm). The former equations

$$\Phi_{2i} = g_i(x) - y_i^- = 0$$

of the system $\Phi = 0$ are now perturbed by small Lipschitz functions $t_i y_i^+$

$$\Phi_{2i}^t = g_i(x) - (y_i^- + t_i y_i^+) = 0.$$

If $y_i > 0$ and $t_i \neq 0$ then $y_i = y_i^+ = \frac{g_i(x)}{t_i}$ follows trivially and can be substituted into the first (Lagrange) equation.
If $y_i \leq 0$ then feasibility $g_i(x) = y_i^- \leq 0$ and $y_i^+ = 0$ follow. This leads us to some interesting consequences.

First of all notice that, due to Thm. 7.5.6 and formula (7.5.6), convergence and estimates of the solutions $(x, y)(t)$ - Lipschitzian w.r. to small $\|t\|$ - are ensured (at least) if Φ is locally Lipschitzian invertible at the KKT point in question.

Moreover, knowing that certain penalty or barrier points can be one-to-one assigned to $(x, y)(t)$, we have immediately estimates for the related errors and the speed of convergence in penalty and barrier methods, too.

Finally, we see that these methods require nothing else but to solve a slightly modified KKT system where the modifying parameter t vanishes and the type of the method depends only on its sign.

8.4.1 Quadratic Penalties

Suppose $t_i > 0 \ \forall i$. Then any zero (x, y) of Φ^t has the following property:

$$\begin{gathered}\text{The point } x \text{ is stationary for the well-known penalty function} \\ P_t(x) = f(x) + \tfrac{1}{2} \textstyle\sum_i t_i^{-1}[g_i(x)^+]^2.\end{gathered} \tag{8.4.2}$$

(Stationary means $D_x P_t(x) = 0$). Conversely, if x is stationary for $P_t(x)$, then (x, y) with

$$y_i = t_i^{-1} g_i(x) \text{ for } g_i(x) > 0 \ \text{ and } \ y_i = g_i(x) \text{ for } g_i(x) \leq 0$$

solves (8.4.1). Thus applying the penalty method based on $P_t(x)$ for $t_i \downarrow 0$ or solving $\Phi^t = 0$ for the same t, mean exactly the same.

8.4.2 Quadratic and logarithmic barriers

Suppose $t_i < 0\ \forall i$. Let (x, y) solve $\Phi^t = 0$ under the perturbation (8.4.1) and put $J(y) = \{i \mid\ y_i > 0\}$. Then (x, y) has the following properties:

$$\begin{array}{c}\text{The point } x \text{ is feasible for the original problem, fulfills } g_i(x) < 0\ \forall i \in J(y)\\ \text{and is stationary for the function}\\ Q_t(x) = f(x) + \frac{1}{2}\ \sum_{i\in J(y)}\ t_i^{-1}[g_i(x)^-]^2.\end{array} \tag{8.4.3}$$

Conversely, having some x with the properties (8.4.3), imposed for any index set $J \subset \{1, ..., m\}$, the point (x, y) with

$$y_i = t_i^{-1} g_i(x)\ (i \in J)\ \text{ and }\ y_i = g_i(x)\ (i \in \{1, ..., m\} \setminus J)$$

is a zero of Φ^t. Moreover, the zeros (x, y) of Φ^t, for $t_i < 0\ \forall i$, can be also (equivalently) characterized by logarithmic barriers:

$$\begin{array}{c}\text{The point } x \text{ is feasible for the original problem, fulfills } g_i(x) < 0\ \forall i \in J(y)\\ \text{and is stationary for the logarithmic barrier function}\\ B_{t,y}(x) = f(x) + \ \sum_{i\in J(y)}\ t_i\ (y_i^+)^2\ \ln(-g_i(x)).\end{array} \tag{8.4.4}$$

To see the equivalence between (8.4.3) and (8.4.4), note that $i \in J(y)$ yields

$$\begin{array}{ll} D_x[\ln(-g_i(x))] = -\frac{Dg_i(x)}{-g_i(x)} = \frac{Dg_i(x)}{t_i y_i} & \text{and}\\ \frac{1}{2} D_x[g_i^2(x)] = g_i(x) Dg_i(x) = t_i y_i Dg_i(x) = (t_i y_i)^2 \frac{Dg_i(x)}{t_i y_i}, & \end{array}$$

hence

$$\frac{1}{2} t_i^{-1} D_x[g_i^2(x)] = \ t_i y_i^2 \frac{Dg_i(x)}{t_i y_i} = \ t_i\ (y_i^+)^2\ D_x[\ln(-g_i(x))].$$

For $t_i \to -0$, the factors $s_i = t_i\ (y_i^+)^2$ vanish (whenever y remains bounded as under MFCQ or under loc. Lipschitzian invertibility). This is the required behavior of these factors in usual log-barrier settings. However, for inactive constraints $g_i(\hat{x}) < 0$, we obtain $y_i^t < 0$ if convergence of the perturbed Kojima points is ensured. Then constraint i is not included in the sum $B_{t,y}(x)$ which improves the speed of convergence in comparison to the usual barrier method.

8.4.3 Modifications and estimates

If some component of t, say t_1, is zero, then the line $g_1(x) - y_1^-$ in Kojima's function (8.3.1) remains unchanged. So the first constraint $g_1(x) \le 0$ is still explicitly required and the term $y_1^+ Dg_1(x)$ (as usually) appears in the Lagrange condition. Similarly, one can discuss the situation when t has both positive and negative components.
Estimates:
Let $t_i = s\ \forall i$. Knowing that $(x, y)(t)$ is Lipschitzian w.r. to small $\|t\|$ and that $p = s^{-1} > 0$ is the penalty- factor of the usual penalty method, we can estimate the penalty solutions

$$\|x(p) - x(p')\| \le L\|s - s'\| = L\|\ 1/p - \ 1/p'\ \|.$$

For the log-barrier method with factors $s_i = t_i\ (y_i^+)^2 < 0\ (i \in J)$, this similarly yields

$$\|x(s) - x(s')\| \le L\|t - t'\|.$$

8.5 Regularity and Newton Meth. for KKT points and VI's

The stability theory of Kojima functions or of generalized equations makes nowhere explicitly use of the particular structure of $\mathcal{M}(x)$ or $H(s)$ in (8.3.5) and (8.1.3), respectively.

Thus $Df(x)$ can be replaced by some (more or less) smooth function $\hat{f} \in C(\mathbb{R}^n, \mathbb{R}^n)$.

The related model then corresponds to the variational inequality

$$0 \in \hat{f}(x) + \mathcal{N}_M(x)$$

where, under a CQ for $M = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$, we have

$$\mathcal{N}_M(x) = \{w \mid \ w = Dh(x)^T z + \sum y_i^+ Dg_i(x), \ \ g(x) = y^- \ \},$$

cf. Corollary 4.1.3. Nevertheless, we consider optimization models with $f, g, h \in C^2$.

The assigned functions Φ, Φ^t (being PC^1) and Θ (from NCP) then satisfy hypothesis (6.2.12) for the (nonsmooth) Newton method with the listed particular Newton maps. The remaining supposition (6.2.11) is satisfied (not only) if Φ is loc. Lipsch. invertible at $(0, \bar{s})$.

Hence the Newton method can be applied to these equations after the mentioned small modifications.

A "nonsmooth Newton step", applied to Φ at $s^k = (x_k, y_k, z_k)$, requires to solve

$$\Phi(s^k) \ + \ \mathcal{D}\Phi(s^k)(u, v, w)^T = 0$$
$$\text{and to put } s^{k+1} = s^k + (u, v, w)$$

where the matrix $\mathcal{D}\Phi(s^k)$ can be computed via the rule (8.3.7) with $\mathcal{DM} = D\mathcal{M}$ and

$$\mathcal{D}N(y_k, z_k) = (0, \ r_1, ..., r_m, \ (1 - r_1), ..., (1 - r_m), \ 1, ..., 1)^T$$

$$\text{where} \qquad r_i = 1 \text{ if } y_{k,i} \geq 0 \ \text{ and } \ r_i = 0 \ \textit{otherwise}.$$

This setting of r describes only one of several possibilities and corresponds to the definition $\mathcal{D}(|y_i|) = 1$ at $y_i = 0$, cf. (8.3.9). Explicitly, the whole matrix is given by

$$\mathcal{D}\Phi(x, y, z) = \begin{pmatrix} D_x^2 L(x, y^+, z) & r_1 Dg_1(x)^T & ... \ r_m Dg_m(x)^T & Dh(x)^T \\ Dg_1(x) & -(1 - r_1) & 0 & 0 \\ ... & ... & ... & ... \\ Dg_m(x) & 0 & -(1 - r_m) & 0 \\ Dh(x) & 0 & 0 & 0 \end{pmatrix}. \qquad (8.5.1)$$

Here, Dg_i is a row, Dh is of type (m_h, n), compare with (7.2.4).
The regularity condition for the Newton method (6.2.11) requires that these matrices are regular near a solution.
Condition (6.2.12) (requiring the approximation to be good enough) is always satisfied since Φ is PC^1, and our definition of r corresponds to the selection of $D\alpha_j$ for some "active" C^1 function α_j.

Remark 8.5.1 (Settings for penalty or barrier methods)
For solving the PC^1 (penalty-barrier) equation $\Phi^t = 0$, one obtains the related matrix $\mathcal{D}\Phi^t(x,y,z)$ very simple:

$$\text{Replace, in (8.5.1), the terms } \ (1-r_i) \ \text{ with } \ (1-r_i)+t_i r_i$$

(this stands for a derivative of $\ y_i \mapsto y_i^- + t_i y_i^+$). ◇

More details can be found in [74]. This paper contains, in particular, Newton-type methods which solve (for small $\|t\|$)

$$\Phi^t(s^k)+\mathcal{D}\Phi^t(s^k)(u,v,w)^T=0 \ \text{ or better } \ \Phi(s^k)+\mathcal{D}\Phi^t(s^k)(u,v,w)^T=0$$
$$\text{in order to put} \qquad s^{k+1}=s^k+(u,v,w),$$

as well as direct relations of these Newton methods to the NCP approach and to sequentially quadradic approximation methods (SQP methods).

Theorem 8.5.2 *(loc. Lip. invertibility) The set of all matrices (8.5.1), defined by the rule*

$$r_i = 0 \ \textit{if}\ y_i<0, \quad r_i\in[0,1] \ \textit{if}\ y_i=0, \quad r_i=1 \ \textit{if}\ y_i>0$$

forms the generalized Jacobian $\partial^c\Phi(x,y,z)$, and Φ is loc. Lipsch. invertible at (x,y,z) $\Leftrightarrow$ all these matrices are regular. ◇

Proof. This follow from the Thms. 7.5.1 and 7.5.2, which present equivalent conditions in the current C^2 situation, and from the product rule Thm. 8.3.2 for $T\Phi$. □

Another form of this condition will be given below by Thm. 12.1.1. To see the equivalence of both conditions, assign to the feasible pairs $(r_i v_i, (1-r_i)v_i)$ (1-to-1) the pairs

$$(\alpha_i,\ \beta_i) = (r_i v_i,\ (1-r_i)v_i). \tag{8.5.2}$$

Remark 8.5.3 (Comparing the approaches)
Though the discussed reformulations of the KKT conditions are equivalent under certain aspects, the useful variations of the factor N (8.3.4) by $t_i y_i^+$ and the related interrelation to penalty-barrier methods are less obvious in model (8.1.3) and in the NCP approach.

Also the calculus of generalized derivatives is more intrinsic and simpler for Kojima's function Φ due to the product rule (8.3.7), cf. Thm. 8.3.2. A similar exact chain rule for $C^{1,1}$ and even C^2 problems does not hold with other descriptions of KKT.

On the other hand, the NCP version of KKT-points is a famous idea for applying directly methods of (generalized) Newton-type. ◇

Chapter 9

More elements of nonsmooth analysis

We already know that convex analysis is a powerful tool due to the following facts concerning minimizers, normals and subgradients. They hold true (at least) for closed convex sets $\emptyset \neq M \subset \mathbb{R}^n$ and convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$ (we repeat the big points):

$$\begin{array}{c} x^* \in \partial f(x) \;\Leftrightarrow\; x \in \operatorname{argmin} f(.) - \langle x^*, . \rangle \;\Leftrightarrow\; (x^*, -1) \in N_{\operatorname{epi} f}(x, f(x)) \\ N_M(x) = \partial i_M(x), \\ 0 \in \partial f(x) + N_M(x) \;\Leftrightarrow\; x \in \operatorname{argmin}_{x \in M} f(x). \\ \partial f(x) \neq \emptyset, \qquad \partial(f+g)(x) = \partial f(x) + \partial g(x). \end{array} \tag{9.0.1}$$

The so-called *nonsmooth analysis or variational analysis* has been developed to extend the theory of convex analysis [131, 41] to non-convex situations, but also in order to unify several results of (local) parametric optimization by appropriate analytical tools. Let us mention only books in chronological manner [7, 13, 55, 5, 104, 134, 114, 10, 70, 18, 107] and the pioneering articles [26, 12, 125] (Ekeland, Clarke, Robinson).

Extending convex analysis to the non-convex case is only possible with weaker statements. In fact, let $\partial f(x)$ be any "subdifferential" which coincides with the usual convex subdifferential for convex f. Put

$$h(x) = 0 = |x| + (-|x|) = f(x) + g(x).$$

Then $\partial h(0) = \{0\},\ \partial f(0) = [-1,\ 1]$ and, after defining $\partial g(0)$ in any way, additivity of $\partial(f+g)$ cannot hold. Similarly, it is evident that our sharp optimality conditions

$$0 \in \partial f(x) \text{ if } x \in \operatorname{argmin} f, \quad \text{even more} \quad 0 \in \partial f(x) + N_M(x) \;\text{ if }\; x \in \operatorname{argmin}_M f$$

are only necessary conditions for non-convex f. All specified generalized subdifferentials below coincide with Df for $f \in C^1$ and with the usual subdifferential ∂f for convex f, except $\partial^{a,\varepsilon} f$. To explain the main problems, we study below only optimality conditions for the basic case of

$$\emptyset \neq M \subset X = \mathbb{R}^n,\ M \text{ closed and } f \in C^{0,1}(X, \mathbb{R}) \text{ (loc. Lipschitz)}, \tag{9.0.2}$$

though many conclusions remain true under weaker hypotheses.

9.1 Abstract normals, subgradients, coderivatives, tangents

In the literature, various definitions of the mentioned objects exist. Let us first demonstrate the close connections between the related definitions.

9.1.1 Subdifferentials and optimality conditions derived from normals

For $x \in M$, let some set $N_M^a(x) \subset X^*$ be defined the elements of which we call (abstract) normals to M at x. We only require, in order to obtain (9.1.3), that

$$(M-x)^* \subset N_M^a(x) \quad \text{or explicitly} \quad x^* \in N_M^a(x) \quad \text{if} \quad \langle x^*, \xi - x\rangle \le 0 \ \forall \xi \in M. \tag{9.1.1}$$

Having $N_M^a(x)$ other objects can be defined.

1. Def. $\partial^a f(x)$: Given f, consider $\operatorname{epi} f \subset X \times \mathbb{R}$, $x \in \operatorname{dom} f$ and put

$$\partial^a f(x) = \{x^* \mid (x^*, -1) \in N^a_{\operatorname{epi} f}(x, f(x))\} \quad \text{(empty or not)}. \tag{9.1.2}$$

This defines a multifunction $x \mapsto \partial^a f(x) \subset X^*$ and implies, by (9.1.1), a

2. Necessary optimality condition:

$$x \in \operatorname{argmin} f \ \Rightarrow \ 0 \in \partial^a f(x). \tag{9.1.3}$$

Proof of necessity: $x \in \operatorname{argmin} f$ and $(\xi, t) \in \operatorname{epi} f$ yield $f(x) \le f(\xi) \le t$, thus

$$\langle 0, \xi - x\rangle - (t - f(x)) \le 0, \ \textit{i.e.}, \quad (0,-1) \in [\operatorname{epi} f - (x, f(x))]^* \subset N^a_{\operatorname{epi} f}(x, f(x)). \quad \square$$

Additionally, we may define a so-called co-derivative $D^{a*}F(x,y) : Y^* \rightrightarrows X^*$ for multifunctions $F : X \rightrightarrows Y$:

3. Def. co-derivative , for $(x,y) \in \operatorname{gph} F \subset X \times Y$, put

$$D^{a*}F(x,y)(y^*) = \{x^* \mid (x^*, -y^*) \in N^a_{\operatorname{gph} F}(x,y)\}. \tag{9.1.4}$$

Particular cases:

If $F(\xi) = f(\xi) + \mathbb{R}^+ \ \forall \xi \in X$, then $\operatorname{gph} F = \operatorname{epi} f$, and $0 \in D^{a*}F(x, f(x))(1)$ says nothing but $0 \in \partial^a f(x)$ (our optimality condition).

If $F(\xi) = A\xi$ is linear, then $x^* \in D^{a*}F(x, f(x))(y^*)$ says $x^* = A^*y^*$ (the adjoint operator).

It depends of our goals whether some definition of N_M^a makes sense or not. Next we summarize basic examples.

9.1.2 (i) Usual localized normals

The simplest case: To characterize *local* minimizer of f in a sharp manner

$$\textbf{define} \quad x^* \in N_M^a(x) \ \Leftrightarrow \ \langle x^*, \xi - x\rangle \le 0 \quad \forall\, \xi \in M \text{ near } x. \tag{9.1.5}$$

The relation $x^* \in \partial^a f(x)$, i.e. $(x^*, -1) \in N^a_{\operatorname{epi} f}(x, f(x))$ by (9.1.2), now means

$$\langle x^*, \xi - x\rangle + f(x) \le t \quad \forall\ (\xi, t) \text{ near } (x, f(x)) \text{ with } f(\xi) \le t.$$

This yields, as desired, the optimality conditions

$$0 \in \partial^a f(x) \ \Leftrightarrow \ f(x) \le f(\xi) \qquad \forall \xi \text{ near } x \qquad \text{and} \tag{9.1.6}$$

$$x^* \in \partial^a f(x) \ \Leftrightarrow \ \langle x^*, \xi - x\rangle + f(x) \le f(\xi) \ \ \forall \xi \text{ near } x. \tag{9.1.7}$$

We obtained a local form of the usual subdifferential, and $\partial^a f = \partial f$ holds for convex f. Since we are interested in non-convex settings, these definitions of N_M^a and $\partial^a f$ will not play any role next.

9.1.3 (ii) Fréchet-normals

To describe *stationary points* x of f in the usual way by the obviously necessary condition for a local minimim at x,

$$f(\xi) \geq f(x) - o(\xi - x), \tag{9.1.8}$$

$$\textbf{define} \qquad x^* \in N_M^a(x) \;\Leftrightarrow\; \langle x^*, \xi - x\rangle \leq o(\xi - x) \quad \forall\, \xi \in M \text{ near } x \tag{9.1.9}$$

Fréchet-normals; we also write F-normals and (below) F-subdifferentials as well as N_M^F and ∂^F. This definition yields like under (i) by definition (9.1.2),

$$x^* \in \partial^F f(x) \;\Leftrightarrow\; \langle x^*, \xi - x\rangle + f(x) \leq t + o(\ (\xi, t) - (x, f(x))\ ) \ \text{ if } f(\xi) \leq t.$$

Again, for proving it, put $t = f(\xi)$. Since $|f(\xi) - f(x)| \leq L\|\xi - x\|$, the above o-term is of type $o(\xi - x)$. We thus obtain the desired statement

$$0 \in \partial^F f(x) \;\Leftrightarrow\; f(x) \leq f(\xi) + o(\xi - x) \Leftrightarrow \ (9.1.8) \qquad \text{and} \tag{9.1.10}$$

$$x^* \in \partial^F f(x) \;\Leftrightarrow\; \langle x^*, \xi - x\rangle + f(x) \leq f(\xi) + o(\xi - x) \Leftrightarrow 0 \in \partial^F(f - x^*)(x). \tag{9.1.11}$$

By (9.1.11), one defines the *Fréchet-subdifferential* also for arbitrary f. The cone (9.1.9) already appeared.

Lemma 9.1.1 *Because of $X = \mathbb{R}^n$, the F-normal cone $N_M^F(x)$ (9.1.9) is polar to the contingent cone: $N_M^F(x) = \mathcal{N}_M(x) := C_M(x)^*$ with C_M from Def. 3.4.2.* ◇

Proof. Implication $(x^* \in \mathbb{R}^n \setminus N_M^F(x) \Rightarrow x^* \notin C_M^*(x))$ can be seen by negation of $x^* \in N_M^F(x)$ and (9.1.9): $\exists \varepsilon > 0,\ \xi_k \overset{M}{\to} x$ such that $\langle x^*, \xi_k - x\rangle \geq \varepsilon\|\xi_k - x\|$. Setting $t_k = \|\xi_k - x\|,\ u_k = (\xi_k - x)/t_k$, we may assume $u_k \to u$ (choose a subsequence; cf. limiting negation). This yields $u \in C_M(x)$ and $t_k \langle x^*, u_k\rangle \geq \varepsilon\ t_k$ as well as $\langle x^*, u\rangle \geq \varepsilon$ and $x^* \notin C_M^*(x)$. Similarly, $N_M^F(x) \subset C_M^*(x)$ follows by contradiction via limiting negation of $x^* \in C_M^*(x)$. □

Again by considering approach directions for $\xi_k \to x$, we see that the optimality condition $0 \in \partial^F f(x)$ can be written by means of the contingent derivative Cf (7.3.1) as $\inf Cf(x; u) \geq 0 \forall u$. This yields the (necessary)

$$\text{optimality condition} \qquad x \in \operatorname{argmin} f \Rightarrow \inf Cf(x; u) \geq 0 \qquad \forall u \in \mathbb{R}^n. \tag{9.1.12}$$

The function $\inf Cf(x; .)$ is often called the *lower Dini-derivative at x*. Since $f \in C^{0,1}$ (locally Lipschitz), $Cf(x; u)$ is a non-empty, compact interval which depends Lipschitzian on u and inf may be replaced by min. It also follows - by the definitions only -

$$(u, \tau) \in C_{\operatorname{epi} f}(x, f(x)) \;\Leftrightarrow\; \min Cf(x; u) \leq \tau. \tag{9.1.13}$$

In addition we obtain a simple necessary condition for constraint minima (known already from [5], chapter 7).

Lemma 9.1.2 $x \in \operatorname{argmin}_{x \in M} f(x) \;\Rightarrow\; \max Cf(x; u) \geq 0\ \forall u \in C_M(x).$ ◇

Proof. Indeed, otherwise we find $u \in C_M(x)$ with $\max Cf(x; u) = -2\delta < 0$. Selecting any $t_k \downarrow 0, u_k \to u$ such that $x + t_k u_k \in M$ (which exist by definition of C_M) so the contradiction $f(x + t_k u_k) \leq f(x) - \delta t_k < f(x)$ follows (for large k). □

In chapter 11, we apply Cf to more general situations (multifunctions).
From the viewpoint of computation, Cf is the simplest generalized derivative of Lipschitz analysis. It can be even determined for critical and stationary point maps, cf. Thm. 8.3.2.

In both cases (i), (ii), also $N_M^a(x) = \partial^a i_M(x)$ holds for the indicator function i_M of M.

Why other concepts ?
The minimum conditions (9.1.10) and Lemma 9.1.2 are as sharp as $0 \in \partial f(x)$ for convex f, $Df(x) = 0$ for $f \in C^1$ and Thm. 3.4.10, respectively.
So there is absolutely no reason to reject these conditions.
The only drawback (in contrast to the convex case) consists in the *formal fact that*

$$\partial^a f(x) = \emptyset$$

may happen for $f \in C^{0,1}$ under (i) and (ii).

Example 9.1.1 For $f(x) = \min\{x, -x^2\}$ (useful for checking many statements), only the origin is a Fréchet normal to epi f at $(0,0)$. So it follows $\partial^F f(0) = \emptyset$. This tells us by (9.1.11) that $0 \notin \partial^F(f - x^*)(0)\ \forall x^*$ and, equivalently, that 0 is even never a stationary point after adding any linear function $\pm\langle x^*, x\rangle$ to f. ◇

To obtain nonempty subdifferentials, also other generalized derivatives (and normals or coderivatives) have been introduced; basically Clarkes subdifferential, limiting normals, limiting subdifferentials and Mordukhovichs coderivative.
However, all these generalizations lead us to weaker (not stronger) optimality conditions.
In particular, all of them will tell us that the origin fulfills the related necessary minimum condition $0 \in \partial^a f(0)$ for f of example 9.1.1, in spite of the fact that $u = -1$ is a proper descent direction.

Thus a decision is required:
Do we need sharp optimality conditions or non-empty subdifferentials ?

Let us look at constructions of non-empty subdifferentials and the related optimality conditions.

9.1.4 (iii) Limiting ε- normals.

Given $\varepsilon > 0$ define certain (approximate) local ε- normals

$$x^* \in N_M^{a,\varepsilon}(x) \ \text{ if } \langle x^*, \xi - x\rangle \le \varepsilon\|\xi - x\| \quad \forall \ \xi \in M \text{ near } x. \tag{9.1.14}$$

Now, $o(\xi - x)$ of case (ii) is replaced by $\varepsilon\|\xi - x\|$ only. The same replacement occurs for the related, i.e. by (9.1.2) assigned, ε-subdifferentials in (9.1.11):

$$x^* \in \partial^{a,\varepsilon} f(x) \ \Leftrightarrow \ \langle x^*, \xi - x\rangle + f(x) \le f(\xi) + \varepsilon\|\xi - x\| \ \text{ (for } \xi \text{ near x).} \tag{9.1.15}$$

(where the ε of normals and subdifferentials may differ by some factor depending on the Lipschitz constant for f). Next consider sequences

$$x_k \overset{M}{\to} x, \ \varepsilon_k \downarrow 0 \text{ and } x_k^* \in N_M^{a,\varepsilon_k}(x_k)$$

and define $N_M^a(x)$ (*the limiting ε-normal cone*) to consists of all limits $x^* = \lim x_k^*$ which can be obtained in this way. A (often used) standard notation of the same fact (cf. sect. 1.2) is $N_M^a(x) = \text{Limsup}_{x_k \overset{M}{\to} x,\ \varepsilon_k \downarrow 0} N_M^{a,\varepsilon_k}(x_k)$. Explicitly, one requires for each k,

$$\langle x_k^*, \xi - x_k\rangle \le \varepsilon_k\|\xi - x_k\| \qquad \forall \xi \in M \text{ in some ball } B(x_k, \delta_k). \tag{9.1.16}$$

If $\dim X = \infty$, the type of convergence $x_k^* \to x^*$ can be modified; but we are assuming $X = \mathbb{R}^n$. Then, due to $|\langle x_k^* - x^*, \xi - x_k\rangle| \le \|x_k^* - x^*\|\ \|\xi - x_k\|$, one may **fix** $x_k^* \equiv x^*$ to obtain slightly simpler

$$x^* \in N_M^a(x) \Leftrightarrow \exists x_k \overset{M}{\to} x,\ \varepsilon_k \downarrow 0 \text{ such that } \langle x^*, \xi - x_k\rangle \le \varepsilon_k \|\xi - x_k\|\ \forall \xi \in M \text{ near } x_k.$$

Comparing with (ii), we now obtain

$$\begin{array}{l} x^* \in \partial^a f(x) \quad \Leftrightarrow \quad \exists\ (x_k, x_k^*) \to (x, x^*),\quad \varepsilon_k, \delta_k \downarrow 0 : \\ \langle x_k^*, \xi - x_k\rangle + f(x_k) \le f(\xi) + \varepsilon_k \|\xi - x_k\| \quad \forall \xi \in B(x_k, \delta_k). \end{array} \tag{9.1.17}$$

which can be again simplified by fixing $x_k^* \equiv x^*$ (but not x_k).

9.1.5 (iv) Limiting Fréchet-normals.

Similarly as under (iii), one may study

$$\text{Limits of F-normals (9.1.9)}\quad x^* = \lim x_k^* \quad \text{for } x_k \overset{M}{\to} x \quad \text{and} \quad x_k^* \in N_M^a(x_k)$$

which define the so-called *limiting F-normal cone* $N_M^a(x)$. This cone and the subdifferential $\partial^a f(x)$ - defined by (9.1.2) - are formally smaller than $N_M^a(x)$ and $\partial^a f(x)$ under (iii) since $o(\xi - x) < \varepsilon\|\xi - x\|$ for ξ near x. For the related limiting subdifferential one obtains as above

$$\begin{array}{l} x^* \in \partial^a f(x) \quad \Leftrightarrow \quad \exists (x_k, x_k^*) \to (x, x^*) : \\ \langle x_k^*, \xi - x\rangle + f(x_k) \le f(\xi) + o_k(\xi - x_k). \end{array} \tag{9.1.18}$$

To see that the defined subdifferentals are not empty, the next statements are helpful.

Lemma 9.1.3 *A limiting F-normal $x^* \neq 0$ exists for every $x \in \operatorname{bd} M$.* ◇

Proof. Indeed, take $x_k \in \mathbb{R}^n \setminus M$, $x_k \to x$ and select, with Euclidean norm, any projection π_k of x_k onto M (π_k is not unique, in general. It exists since $M \neq \emptyset$ is closed in $\mathbb{R}^n$). Now $x_k^* = (\pi_k - x_k)/\|\pi_k - x_k\|$ is a F-normal at π_k, and $\pi_k \to x$. For some subsequence, $x_k^* \to x^* \neq 0$ in $\mathbb{R}^n$ is ensured. So $x^* \in N_M^a(x)$ follows. □

Since we suppose $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R})$, a limiting F-normal (x^*, τ^*) at the epigraph cannot be horizontal at $(x, f(x))$; i.e., $\tau^* \neq 0$. Since $\tau^* \le 0$ holds for any epigraph, $\tau^* < 0$ follows. Thus division by $|\tau^*|$ also yields $\partial^a f(x) \neq \emptyset$.

9.1.6 Equivalence of the limiting concepts

The next statement implies that both concepts (iii) and (iv) define the same normals and, via (9.1.2), the same subdifferentials. $\partial^{(iii)} f(x) = \partial^{(iv)} f(x)$.

Theorem 9.1.4 *The limiting normal cones $N_M^{lim}(x)$ under (iii) and (iv) coincide, and $x^* \in N_M^{lim}(x) \ \Leftrightarrow\ x^* = \lim x_k^*$ with $x_k^* \in C_M(x_k)^*$ and $x_k \overset{M}{\to} x$.* ◇

Other notation, $N_M^{lim}(x) = \operatorname{Limsup}_{x_k \overset{M}{\to} x} C_M(x_k)^*$.

Proof. Consider first a local ε-normal x^* with $\varepsilon < 1$ and $\|x^*\| = 1$ at x, and points $x(t) = x + tx^*$, $t > 0$. Choose any Euclidean projection π_t of $x(t)$ onto the closed set M. Then

$$r_t := \|x(t) - \pi_t\| \ \le\ \|x(t) - x\| = t = \langle x^*, x(t) - x\rangle.$$

If $\pi_t = x(t)$, it follows $x(t) \in M$ and, for small t, we also have from $x^* \in N_M^{a,\varepsilon}(x)$, $t = \langle x^*, x(t)-x\rangle \le \varepsilon\|x(t)-x\| = \varepsilon\, t < t$. Thus $\pi_t \neq x(t)$, and $u_t = (x(t)-\pi_t)\,\|x(t)-\pi_t\|^{-1}$ is a local Fréchet-normal for M at π_t. Since $x^* \in N_M^{a,\varepsilon}(x)$, we also have (for small t)

$$\langle x^*, \pi_t - x\rangle \le \varepsilon\,\|\pi_t - x\| \le \varepsilon\,(\,\|\pi_t - x(t)\| + \|x(t) - x\|\,) = \varepsilon(r_t + t) \;\le\; 2\varepsilon t.$$

Calling x^* the south direction, x is located above the north-pole $x^n = x + (t - r_t)\,x^*$ of the ball $B_t = x(t) + r_t B$, and π_t belongs to the upper cap of B_t due to

$$0 \le \langle x^*, \pi_t - x^n\rangle = \langle x^*, \pi_t - x - (t - r_t)\,x^*)\rangle \le 2\varepsilon t + r_t - t = r_t - (1 - 2\varepsilon)t.$$

Since

$$\frac{r_t - (1-2\varepsilon)t}{r_t} \to 0 \quad as \quad \varepsilon \to 0,$$

we obtain $\lim_{\varepsilon\downarrow 0}\ \|\pi_t - x^n\|\ \|\pi_t - x(t)\|^{-1} = 0$ after which $\|u_t - x^*\|$ is arbitrarily small if ε is small enough. Now let $x^* = \lim x_k^*$ be any limiting normal (iii):

$$x_k \overset{M}{\to} x, \quad \varepsilon_k \downarrow 0 \ \text{ and } \ x_k^* \in N_M^{a,\varepsilon_k}(x_k).$$

If $x^* = 0$ then $x^* \in N_M^a(x)$ (iv) is trivial. Let $\|x^*\| = c > 0$. We may suppose $\|x_k^*\| = c$, otherwise choose new $y_k^* = cx_k^*\|x_k^*\|^{-1}$ and slightly bigger $\varepsilon_k \downarrow 0$. By the same devise, $c = 1$ can be assumed. Finally, passing (for each large k) to one of the above found F-normal u_t and setting $u(k) = u_t$, we obtain $x^* = \lim u(k)$. In consequence, $N_M^{(iii)} = N_M^{(iv)}$.
The form of $N_M^{lim}(x)$ as Limsup of cones $C_M(x_k)^*$ follows from Lemma 9.1.1. □

9.2 (v) Clarke's approach

Clarke's concept [13] represents a direct generalization of convex analysis for loc. Lipschitz functions on $\mathbb{R}^n$.

9.2.1 Basic definitions and interrelations

Before considering the related normals, let us begin with a set of tangents:

$$u \in T_M^{Cl}(x) \ \text{ if } \ \forall x_k \overset{M}{\to} x,\ t_k \downarrow 0\ \ \exists u_k \to u \text{ such that } \ x_k + t_k u_k \in M\ \forall k. \tag{9.2.1}$$

Setting here $x_k \equiv x$, we see that

$$T_M^{Cl}(x) \subset C_M(x). \tag{9.2.2}$$

Remark 9.2.1
(i) The definition of $T_M^{Cl}(x)$ does not change if we require the inclusion $x_k + t_k u_k \in M$ in (9.2.1) only for $k \ge k_0$. Indeed, for $k \le k_0$, we could put $u_k = (u - x_k)/t_k$ in order to ensure $x_k + t_k u_k = x \in M$ without changing $u_k \to u$.
(ii) It is even sufficient that the inclusion always holds for some infinite subsequence. To see this, note that by negation

$$u \notin T_M^{Cl}(x) \Leftrightarrow \exists x_k \overset{M}{\to} x,\ t_k \downarrow 0 \ \text{ such that } \forall u_k \text{ with } \ x_k + t_k u_k \in M$$
$$\exists \varepsilon > 0 \text{ with } \|u_k - u\| > \varepsilon \text{ for some infinite subsequence of } k = k_\nu.$$

Considering this subsequence, one obtains

$$u \notin T_M^{Cl}(x) \Leftrightarrow \exists x_k \overset{M}{\to} x,\ t_k \downarrow 0 \ \text{ such that } \forall u_k \text{ with } \ x_k + t_k u_k \in M$$
$$\exists \varepsilon > 0 \text{ with } \|u_k - u\| > \varepsilon\ \forall k.$$

◇

Theorem 9.2.2 *(properties of T^{Cl})* $T_M^{Cl}(x)$ *is a closed, convex cone.* ◇

Proof. We abbreviate $T^{Cl} = T_M^{Cl}(x)$.

Cone property: Let $u \in T^{Cl}$ and $\lambda \geq 0$. Given $x_k \overset{M}{\to} x,\ t_k \downarrow 0$, we find $u_k \to u$ such that $x_k + (\lambda t_k)u_k \in M$. Thus $\lambda u \in T^{Cl}$ follows from $u'_k := \lambda u_k \to \lambda u$ and $x_k + t_k u'_k \in M$.

Closeness: Let $v_m \to u$, $v_m \in T^{Cl}$. Given $x_k \overset{M}{\to} x$ and $t_k \downarrow 0$ we have to find $u_k \to u$ such that $x_k + t_k u_k \in M$. By assumption there are - for fixed m - certain $w_{m,k}$ such that $w_{m,k} \to v_m$ (as $k \to \infty$) and $x_k + t_k\ w_{m,k} \in M$. For $m = 1$ put $k(1) = 1$, otherwise we find inductively some $k(m) > k(m-1)$ such that $\|w_{m,k(m)} - v_m\| < 1/m$. Then $x_{k(m)} + t_{k(m)} w_{m,k(m)} \in M$ holds true by supposition, and $w_{m,k(m)} \to u$ as $k \to \infty$ follows from $v_m \to u$. Hence, for some infinite subsequence of all k given by $k' = k(m)$, the required elements $u_k = w_{m,k(m)} \to u$ with $x_k + t_k u_k \in M$ exist. This is already sufficient for $u \in T^{Cl}$, c.f. Remark 9.2.1 (ii).

Convexity: Let $u, v \in T^{Cl}$. To verify $u + v \in T^{Cl}$, consider any $x_k \overset{M}{\to} x$ and $t_k \downarrow 0$. By $u \in T^{Cl}$ there are $u_k \to u$ such that $x'_k := x_k + t_k u_k \in M$. Moreover, since $x'_k \in M$ there are also $v_k \to v$ such that $x'_k + t_k v_k \in M$. Due to $x_k + t_k(u_k + v_k) \in M$, this yields $u + v \in T^{Cl}$. Therefore, the cone T^{Cl} is convex. □

Lemma 9.2.3 *If M is closed and convex then* $T_M^{Cl}(x) = C_M(x)\ \forall x \in M$. ◇

Proof. First of all choose some $p \in \operatorname{int} M$ (if $\operatorname{int} M = \emptyset$, study M in the smallest affine subspace containing M). Having $B(p, \delta) \subset M$ then, changing the norm, we may assume $\delta = 1$ to simplify. By convexity this yields for all $\lambda \in (0, 1)$,

$$B(y, \lambda) \subset M \qquad \text{if } \ y = \lambda p + (1 - \lambda)\zeta \ \text{ and } \zeta \in M. \tag{9.2.3}$$

Due to (9.2.2) it remains to show $C_M(x) \subset T_M^{Cl}(x)$. Thus, let $x = 0$ (without loss of generality) and $u \in C_M(x)$, $x_k \overset{M}{\to} x, t_k \downarrow 0$ be given. We have to verify that certain $u_k \to u$ satisfy $x_k + t_k u_k \in M$ (for large k). Because of $u \in C_M(0)$ we know: There are $v_\nu \to u$ and $s_\nu \downarrow 0$ such that $z_\nu := s_\nu\ v_\nu \in M$. Since $\|x_k\| + t_k \to 0$, there exists, for large k, some smallest $s_{\nu(k)}$ such that

$$\|x_k\| + t_k < s_{\nu(k)}^2 < s_{\nu(k)}. \tag{9.2.4}$$

Though $\nu(k) \leq \nu(k+1)$ is not excluded for certain k, we observe that $s_{\nu(k)} \downarrow 0$. Setting $w_k = v_{\nu(k)}$ and $\tau_k := s_{\nu(k)}$ this ensures

$$z_{\nu(k)} = \tau_k\ w_k \in M \text{ and } w_k \to u. \tag{9.2.5}$$

Now connect $\zeta = \tau_k\ w_k$ and p. Applying (9.2.3), it follows for any $\lambda_k \in (0, 1)$

$$B(\ \lambda_k p + (1 - \lambda_k)\ \tau_k\ w_k,\ \ \lambda_k) \subset M. \tag{9.2.6}$$

Thus, if $\lambda_k \geq \|x_k\|$, we have $x_k + \lambda_k p + (1 - \lambda_k)\tau_k\ w_k \in M$. Using (9.2.4), inequality $\lambda_k \geq \|x_k\|$ holds for $\lambda_k = \tau_k^2$. Hence $x_k + \tau_k^2 p + (1 - \tau_k^2)\tau_k w_k \in M$. This ensures

$$x_k + \tau_k u_k \in M \text{ and } u_k \to u \quad \text{ with } \ u_k = \tau_k p + (1 - \tau_k^2) w_k.$$

From $x_k \in M$ and $t_k < \tau_k$ so also $x_k + t_k u_k \in M$ follows from convexity. □

Clarke's normal cone is defined as $N_M^{Cl}(x) = T_M^{Cl}(x)^*$.
Since $T_M^{Cl}(x) \subset C_M(x)$, so $N_M^{Cl}(x)$ contains $C_M(x)^*$.
Clarke's subdifferential [13] (we define it first directly, later via (9.1.2)) is based on his directional derivative; put

$$f^{Cl}(x;u) = \limsup_{t_k \downarrow 0,\ x_k \to x} t_k^{-1}(f(x_k + t_k u) - f(x_k)) \qquad (< \infty). \tag{9.2.7}$$

The notation says that $f^{Cl}(x;u)$ is the largest limit of the difference quotients that can be obtained by sequences $x_k \to x,\ t_k \downarrow 0$. Recall that $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R})$. So $f^{Cl}(x;u)$ is bounded and we may replace u by arbitrary $u_k \to u$ in the limit without changing the definition. This yields that

$$f^{Cl}(x;u) = \max Tf(x;u), \quad (Tf \text{ from 7.3.1 is the set of all possible limits}).$$

One easily shows that

$$f^{Cl}(x;u'+u'') \le f^{Cl}(x;u') + f^{Cl}(x;u'') \ \text{ and } f^{Cl}(x;\lambda u) = \lambda f^{Cl}(x;u) \text{ if } \lambda \ge 0. \tag{9.2.8}$$

Positively homogeneous is evident. Sublinearity: Write

$$\begin{aligned} & t_k^{-1}[f(x_k + t_k(u'+u'')) - f(x_k)] \\ = \ & t_k^{-1}[\, f(x_k + t_k(u'+u'') - f(x_k + t_k u')\,] + t_k^{-1}[\, f(x_k + t_k u') - f(x_k)\,] \\ = \ & t_k^{-1}[\, f((x_k + t_k u') + t_k u'') - f(x_k + t_k u')\,] + t_k^{-1}[\, f(x_k + t_k u') - f(x_k)\,]. \end{aligned}$$

The first difference quotient corresponds to $f^{Cl}(x;u'')$ (x_k is replaced by $x_k + t_k u'$), the second one to $f^{Cl}(x;u')$. Passing now to lim sup, the assertion follows.

Hence, for fixed x, the function $u \mapsto g(u) = f^{Cl}(x;u)$ is sublinear and convex. The nonempty and compact subdifferential of convex analysis $\partial g(0)$ defines Clarke's subdifferential

Definition 9.2.1 $\partial^{Cl} f(x) := \partial g(0)\ , \quad \partial g = \text{convex subdifferential}.$ ◇

Now, every (unconstrained) local minimizer x of f satisfies the condition

$$f^{Cl}(x;u) \ge 0 \ \forall u \in \mathbb{R}^n \tag{9.2.9}$$

which explicitly requires $\limsup_{x_k \to x,\ t_k \downarrow 0} t_k^{-1}(f(x_k + t_k u) - f(x_k)) \ge 0 \ \forall u \in \mathbb{R}^n$ and has to hold even if $x_k \equiv x$.

Particular cases:
If $f \in C^1$ then $\partial^{Cl} f(x) = \{Df(x)\}$.
If f is convex then $\partial^{Cl} f(x) = \partial f(x)$.

Proof. If $f \in C^1$ then each difference can be written as

$$f(x_k + t_k u) - f(x_k) = t_k Df(\theta_k)u \text{ where } \theta_k \to x$$

which gives

$$t_k^{-1}(f(x_k + t_k u) - f(x_k)) \to Df(x)u \quad \text{for all } x_k \to x,\ t_k \downarrow 0$$

and

$$\partial^{Cl} f(x) = \{Df(x)\}.$$

Convex f: We first note that *the usual subdifferential ∂f is a closed mapping, i.e., if $x_k \to x$, $s_k \to s$ and $s_k \in \partial f(x_k)$ then $s \in \partial f(x)$.*

Indeed, if $s \notin \partial f(x)$ is outside, we have $f(x+w) < f(x) + \langle s, w\rangle - \delta$ for some w. So also $f(x_k + w) < f(x_k) + \langle s_k, w\rangle - \delta$ holds true by continuity (k big). Hence $s_k \in \partial f(x_k)$ cannot hold.

Next, setting $x_k = x$, it follows

$$\limsup_{x_k \to x,\ t_k \downarrow 0} t_k^{-1}(f(x_k + t_k u) - f(x_k)) \geq f'(x,u) = \max_{s \in \partial f(x)} \langle s, u\rangle.$$

On the other hand let, x_k, t_k be sequences which realize the lim sup. Each difference fulfills, with $s_k \in \partial f(x_k + t_k u)$,

$$f(x_k) - f(x_k + t_k u) \geq -t_k \langle s_k, u\rangle \quad \text{i.e.,} \quad f(x_k + t_k u) - f(x_k) \leq t_k \langle s_k, u\rangle.$$

Passing to the limit, the s_k have some cluster point $s \in \partial f(x)$, thus

$$\limsup_{x_k \to x,\ t_k \downarrow 0} t_k^{-1}(f(x_k + t_k u) - f(x_k)) \leq \langle s, u\rangle \text{ for some } s \in \partial f(x).$$

Summarizing, this means $f^{Cl}(x;u) = f'(x,u)$ and implies $\partial^{Cl} f(x) = \partial f(x)$. □

Next we show that, by separation as in the convex version Thm. 3.4.10, Clarke's optimality condition follows.

Theorem 9.2.4 *(Optimality condition, Clarke) For closed M and $f \in C^{0,1}$, it holds*

$$x \in \operatorname{argmin}_M f \text{ (locally)} \ \Rightarrow \ \text{(Min. condit.)} \ \ 0 \in \partial^{Cl} f(x) + N_M^{Cl}(x). \qquad \diamond \quad (9.2.10)$$

Proof. If the Min. condition is violated, the compact convex set $\partial^{Cl} f(x)$ and the closed convex cone $-N_M^{Cl}(x)$ can be separated:

$$\exists u: \ \langle u, \partial^{Cl} f(x)\rangle < \langle u, -N_M^{Cl}(x)\rangle$$

(the inequality holds for any selected elements of the sets). Since $0 \in N_M^{Cl}(x)$, this yields both

$$\langle u, \partial^{Cl} f(x)\rangle < 0 \ \text{ and } \ u \in N_M^{Cl}(x)^*.$$

Thm. 3.1.5 (bipolar) ensures $N_M^{Cl}(x)^* = T_M^{Cl}(x)$. By Def. 9.2.1, we have $\partial^{Cl} f(x) = \partial g(0)$ and, recalling Thm. 3.4.8 (dir. deriv.),

$$f^{Cl}(x,u) = g(u) = g'(0;u) = \max_{x^* \in \partial g(0)} \langle u, x^*\rangle = \mu < 0. \qquad (9.2.11)$$

Equation $g(u) = g'(0,u)$ is valid since $g(.) = f^{Cl}(x,.)$ is positively homogeneous. Because of $u \in T_M^{Cl}(x))$ we find, for any sequence $t_k \downarrow 0$, elements $u_k \to u$ such that $x + t_k u_k \in M$. By (9.2.11) we know that $f(x + t_k u_k) - f(x) < \frac{1}{2} t_k \mu < 0$ for large k. Thus x is not optimal. □

Compare with the proof of Thm. 3.4.10 and notice that the minimum conditions

$$\text{(d)} \quad 0 \in \partial^{Cl} f(x) + N_M^{Cl}(x), \qquad \text{(p)} \quad f^{Cl}(x;u) \geq 0 \quad \forall u \in T_M^{Cl}(x) \qquad (9.2.12)$$

are telling the same (in the dual and primal language, respectively).
If (p) is violated, there is some $u \in T_M^{Cl}(x)$ such that $f^{Cl}(x;u) < -\delta < 0$.

Alternative definitions:
Let us write $x^* \in \partial_g f(x)$ if the optimality condition (9.2.9) is valid for $h := f - x^*$. Then

one easily shows that $\partial_g f(x) = \partial^{Cl} f(x)$.
The reverse way of defining $N_M^{Cl}(x)$ by the help of $\partial^{Cl} i_M$ is less obvious. Indeed, provided we put $x^* \in N_M(x)$ if $x^* \in \partial_g i_M(x)$, we require formally for i_M and x^*

$$\limsup_{(x_k,u_k)\to(x,u),\ t_k\downarrow 0} t_k^{-1}(\ i_M(x_k+t_k u_k) - \ i_M(x_k)\ ) \geq \langle x^*, u\rangle. \tag{9.2.13}$$

Now possible ∞-values make difficulties. However, as above, one can use the normal cone in order to define the subdifferential by (9.1.2).

Using Clarke's normals to define $\partial^{Cl} f(x)$.
Having the cone

$$N_M^a(x) := N_M^{Cl}(x),$$

define a subdifferential once more via (9.1.2):

$$x^* \in \partial^a f(x) \Leftrightarrow (x^*, -1) \in N^a_{epi f}(x, f(x)).$$

Let us show that

$$\partial^a f(x) = \partial^{Cl} f(x).$$

In other words, also $\partial^{Cl} f(x)$ can be defined via the normal cone $N^{Cl}_{\mathrm{epi}\, f}(x, f(x))$ and (9.1.2), though this is not the usual way: We have

$$(x^*, -1) \in N^a_{epi f}(x, f(x)) \Leftrightarrow \langle x^*, u\rangle - 1v \leq 0\ \forall (u, v) \in T^{Cl}_{epi f}(x, f(x)).$$

By definition of T^{Cl}, we may equivalently transform (abbreviate $\mathcal{D} = \partial^a f(x)$)

$$\begin{gathered} x^* \in \mathcal{D} \Leftrightarrow \langle x^*, u\rangle \leq v \ \text{ if } \ \forall t_k \downarrow 0,\ (x_k, \tau_k) \overset{\mathrm{epi}\, f}{\to} (x, \tau) \\ \exists (u_k, v_k) \to (u, v) :\ (x_k + t_k u_k,\ \tau_k + t_k v_k) \in \mathrm{epi}\, f. \end{gathered}$$

$$\begin{gathered} x^* \in \mathcal{D} \Leftrightarrow \langle x^*, u\rangle \leq v \ \text{ if } \ \forall t_k \downarrow 0,\ (x_k, \tau_k) \to (x, \tau)\ with\ f(x_k) \leq \tau_k \\ \exists (u_k, v_k) \to (u, v) :\ f(x_k + t_k u_k) \leq \tau_k + t_k v_k. \end{gathered}$$

If we take v_k slightly larger (add $L\|u_k - u\|$) we may write $u_k = u$

$$\begin{gathered} x^* \in \mathcal{D} \Leftrightarrow \langle x^*, u\rangle \leq v \ \text{ if } \ \forall t_k \downarrow 0,\ (x_k, \tau_k) \to (x, \tau)\ with\ f(x_k) \leq \tau_k \\ \exists v_k \to v : f(x_k + t_k u) \leq \tau_k + t_k v_k. \end{gathered}$$

This condition does not change if we consider $\tau_k = f(x_k)$ only:

$$x^* \in \mathcal{D} \Leftrightarrow \langle x^*, u\rangle \leq v \text{ if } \forall t_k \downarrow 0, x_k \to x\ \exists v_k \to v : f(x_k + t_k u) \leq f(x_k) + t_k v_k.$$

This tells us finally, $x^* \in \mathcal{D} \Leftrightarrow \langle x^*, u\rangle \leq \limsup_{x_k\to x,\ t_k\downarrow 0} t_k^{-1}(\ f(x_k+t_k u) - f(x_k)\ ) = g(u)$ and $x^* \in \partial g(0) = \partial^{Cl} f(x)$. Notice that all steps can be reversed. □

9.2.2 Mean-value theorems and Generalized Jacobian

Before finishing this excursion into Clarke's theory, let us mention that $\partial^{Cl} f(x)$ can be determined (for "reasonable" f) by the help of generalized Jacobians, $\partial^{GenJac} f(x)$ cf. Def. 7.3.3 and the related chapter. For real valued f as here, then

$$\partial^{Cl} f(x) = \partial^{GenJac} f(x) \tag{9.2.14}$$

holds true. For this reason, the same abbreviation $\partial^{Cl} f(x)$ is used for both in the literature. To prove (9.2.14) we need

Theorem 9.2.5 *Mean-value theorem for $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R})$:*

$$\begin{array}{l} f(b) - f(a) \in \partial^{Cl} f(\theta)(b-a) \text{ holds for some } \theta \in (a,b) \\ \text{where} \\ \partial^{Cl} f(\theta)(b-a) := \{\ \langle x^*, b-a\rangle \mid\ x^* \in \partial^{Cl} f(\theta)\ \}. \\ \text{The same is true with } \partial^{GenJac} f \text{ and } \theta \in [a,b]. \end{array} \tag{9.2.15}$$

Proof. Part one: $f : \mathbb{R} \to \mathbb{R}$.
$\partial^{Cl} f$.
Suppose first $f(a) = f(b),\ a < b$. Then, there is a minimizer/maximizer $\theta \in (a,b)$ of f. Thus, the necessary optimality condition $0 \in \partial^{Cl} f(\theta)$ holds true. If $f(a) \neq f(b)$ we add

$$L(x) = -\frac{f(b)-f(a)}{b-a}(x-a); \quad \text{where } L(a) = 0,\ L(b) = f(a) - f(b).$$

Since $(f+L)(a) = f(a) = (f+L)(b)$, there is some $\theta \in (a,b)$ such that

$$0 \in \partial^{Cl}(f+L)(\theta)(b-a) = -\frac{f(b)-f(a)}{b-a} + \partial^{Cl} f(\theta).$$

Thus (9.2.15) holds true.

$\partial^{GenJac} f$: Next we show that, for $f(a) = f(b)$, also $0 \in \partial^{GenJac} f(x)$ holds for some $x \in [a,b]$. Hence assume that all intervals $\partial^{GenJac} f(x) = [\alpha(x), \beta(x)]$ do not contain the origin. The mapping $x \mapsto \partial^{GenJac} f(x)$ is closed. So also

$$0 < c := \inf_x \ \text{dist}\,(0, \partial^{GenJac} f(x))$$

is true. If we find two points $x_1, x_2 \in [a,b]$ such that $0 < \alpha(x_1)$ and $0 > \beta(x_2)$ then, by a bisection method, we also find some $\xi \in [x_1, x_2]$ with $0 \in \partial^{GenJac} f(\xi)$ which finishes the proof. It remains, without loss of generality, the case of $0 < c \leq \alpha(x) \forall x$. Then $Df(x) \geq c$ holds for all x where f is differentiable, a set of full Lebesgue-measure. Setting $Dg(x') = 0$ for the remaining points, this gives as Lebesgue Integral

$$f(b) - f(a) = \int_a^b Df(x)dx \geq c(b-a) > 0,$$

a contradiction to $f(a) = f(b)$. For the general case, this yields as above via $f + L$,

$$f(b) - f(a) \in \partial^{GenJac} f(x)(b-a) \text{ for some } x \in [a,b].$$

Part two: $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R})$.
As for C^1 mean-value statements, define

$$h(t) = f(a + t(b-a)),\ t \in \mathbb{R}$$

and apply the related statements for h. There are, however two facts which must be taken into consideration.
∂^{Cl}:
We obtain

$$f(b) - f(a) = h(1) - h(0) \in \partial^{Cl} h(\theta), \quad 0 < \theta < 1.$$

In general, the set $\partial^{Cl} h(\theta)$ does not coincide with $\partial^{Cl} f(a+\theta(b-a))(b-a)$: For the related directional derivative with $x_k = a + t_k(b-a)$ in the interval, it holds

$$h^{Cl}(\theta, 1) = \limsup_{s_k \downarrow 0,\ x_k = a + t_k(b-a)\ \to\ \theta} s_k^{-1}[f(x_k + s_k(b-a)) - f(x_k)]$$

while we have any $x_k \to a+\theta(b-a)$ in $\mathbb{R}^n$ for

$$f^{Cl}(a+\theta(b-a))(b-a) = \limsup_{s_k\downarrow 0,\ x_k\to a+\theta(b-a)} s_k^{-1}[f(x_k+s_k(b-a))-f(x_k)].$$

But since $h^{Cl}(\theta,1) \le f^{Cl}(a+\theta(b-a))(b-a)$ and analogue

$$h^{Cl}(\theta,\tau) \le f^{Cl}(a+\theta(b-a),\tau(b-a)) \quad \forall \tau\in\mathbb{R},$$

we obtain the needed inclusion $\partial^{Cl}h(\theta) \subset \partial^{Cl}f((a+\theta(b-a))(b-a)$.
∂^{Cl}:
Let N_f be the set of all points where $Df(x)$ does not exist. Then, by Rademacher's Theorem, $\mu_n(N_f)=0$ holds for the n-dimensional Lebesgue-measure μ_n. Nevertheless, on the segment $[a,b]$ between a and b in $\mathbb{R}^n$, the one-dimensional Lebesgue-measure μ_1 of the set where Dh does not exist, may be positive. So we cannot directly apply the result for $\mathbb{R}^n=\mathbb{R}$. By a theorem of Fubini there are, however (due to $\mu_n(N_f)=0$), points $a_k\to a$ and $b_k\to b$ such that $\mu_1(N_f\cap[a_k,b_k])=0\ \forall k$. Considering $h_k(t)=f(a_k+t(b_k-a_k))$, related θ_k and passing to the limit, then the assertion follows. □

Decreasing the length of intervals with center x, so also formula (9.2.14) follows for the convex sets $\partial^{Cl}f(x)$ and $\partial^{GenJac}f(x)$.

Theorem 9.2.6 *Mean-value theorem for $f\in C^{0,1}(\mathbb{R}^n,\mathbb{R}^m)$:*

$$f(b)-f(a) \in \text{cl conv}(\ \cup_{x\in[a,b]}\ \partial^{GenJac}f(x)\ )\ (b-a). \tag{9.2.16}$$

Proof. Let $C=\text{cl conv}(\ \cup_{x\in[a,b]}\ \partial^{GenJac}f(x)\ )\ (b-a)$. This is a non-empty, compact, convex set in $\mathbb{R}^m$. Assume that $f(b)-f(a)\notin C$. Then (separation) there is some $v\in\mathbb{R}^m$ such that $\langle v,c\rangle < \langle v,f(b)-f(a)\rangle$ for all $c\in C$. The function

$$h(x)=\langle v,f(x)\rangle$$

maps $\mathbb{R}^n$ into the reals. Hence Thm. 9.2.5 ensures

$$\langle v,f(b)-f(a)\rangle \in \partial^{GenJac}h(x)(b-a) \text{ for some } x\in[a,b].$$

Moreover, $Dh(x)=v^TDf(x)$ holds true if $Df(x)$ exists, but $Dh(y_k)$ may exist if $Df(y_k)$ does not (e.g. if $v=e_1$ and $h=f_1$). However, for writing $f(b)-f(b)$ as L-integral in the former proof of real-valued f, we needed only any set of full measure where Df exists. Doing this now for h and $\mathbb{R}^n\setminus N_f$, we see that $\partial^{GenJac}h(x)(b-a)\subset C$ and obtain a contradiction to the separation property. □

9.3 Subdifferentials derived from optimality conditions

9.3.1 The elementary principle

Assume we know any necessary optimality condition (OC) for loc. minimizers of f. Then we can immediately define a generalized subdifferential,

$$x^*\in\partial_g f(x) \quad \text{if } x \text{ fulfills (OC) for } h:=f-x^*. \tag{9.3.1}$$

In consequence, the original condition (OC), which defines our *stationary points*, attains the well-known form $0\in\partial_g f(x)$ and, obviously, (OC) defines all properties of ∂_g. Similarly, an optimality condition can be used to define a normal cone:

$$x^*\in N_M^g(x) \ \text{ if } x \text{ satisfies (OC) for } \ \min\{-\langle x^*,x\rangle \mid x\in M\}.$$

The mapping $x^* \mapsto (\partial_g f)^{-1}(x^*)$ assigns, to x^*, all x which satisfy (OC) for $f - x^*$. Thus it characterizes (as in the convex case) the behavior of the related stationary points under linear perturbations of the objective. Moreover and more important,

also $\partial_g f(x) = \emptyset$ has a precise meaning:

It tells us that x will never satisfy the optimality condition (OC), even after any linear transformation of f by x^*.

This situation (which of course depends on optimality condition OC) is quite natural in Lipschitz analysis, cf. $f(x) = -|x|$ or example 9.1.1 after taking the canonical optimality condition (OC) of (9.1.10), namely $f(x) \le f(\xi) + o(\xi - x)$.

Limiting subdifferentials:
As above, one can define limiting subdifferentials by the Limsup of the original ones. Then

$$0 \in \partial_g^{lim} f(x) \;\Leftrightarrow\; \exists (x_k, x_k^*) \overset{\operatorname{gph} \partial_g f}{\longrightarrow} (x, 0)$$

tells us that x is a limit of stationary points after particular vanishing linear perturbations of f. For F-stationarity, this defines again $\partial^{lim} f$ of sect. 9.1.5 with $\operatorname{dom} \partial^{lim} f = \mathbb{R}^n$. However, corollary 9.3.1 ($0 \notin \partial_g f(x) \Leftrightarrow$ $f_\le$ is lower Lipschitz at $(f(x), x)$) does no longer hold for the bigger limiting F-subdifferential $\partial_g^{lim} f$.

9.3.2 The empty and limiting F-subdifferential

Next we consider this case of F-stationarity, $\partial_g f(x) = \partial^F f(x)$ (ii), in detail.
Negation of (OC):
The condition $0 \notin \partial_g f(x)$ negates the optimality condition and means: $\exists \varepsilon > 0$ such that certain $\xi_k \to x$ satisfy $f(\xi_k) - f(x) \le -\varepsilon \|\xi_k - x\|$. By *limiting negation* we even obtain

$$\exists \xi_k = x + t_k u_k,\; u_k \to u, \|u_k\| = 1, t_k \downarrow 0 \quad \text{with } f(x + t_k u_k) - f(x) \le -\varepsilon t_k. \tag{9.3.2}$$

Thus $0 \notin \partial_g f(x)$ means, for $f \in C^{0,1}$,

$$\min Cf(x; u) < 0 \quad \text{holds for some } \; u \in \operatorname{bd} B.$$

If f is directionally differentiable, (9.3.2) holds for all sufficiently small t, not only for *certain* $t_k \downarrow 0$. Then we see that the lower level sets $f_\le(p)$ satisfy, for small $|p|$, the estimate

$$\operatorname{dist}(x, f_\le(p)) \le L|p - f(x)| \;\text{ where } L = 1/\varepsilon.$$

In other words, $f_\le$ is (locally) lower Lipschitz at $z := (f(x), x)$; a useful stability property, cf. Def. 10.1.1. Conversely, if $f_\le$ is lower Lipschitz at z then $0 \in \partial_g f(x)$ cannot hold. Thus:

Corollary 9.3.1 *For directionally differentiable f ($\in C^{0,1}$), the both properties $0 \notin \partial_g f(x)$ and $f_\le$ being lower Lipschitz at $(f(x), x)$ are equivalent.* ◇

The set $\operatorname{dom} \partial_g f$, $\partial_g = \partial^F$:
This set has full Lebesgue measure since $Df(x) \in \partial_g f(x)$ and, by Rademachers theorem, the usual derivative $Df(x)$ exists, for each $f \in C^{0,1}$, on a set of full Lebesgue measure (see [29] for a proof of this non-trivial fact). The points $x \in \operatorname{bd} \operatorname{dom} \partial_g f$ are of particular

interest.
Having $0 \in \partial_g f(x)$ (possible after some linear transformation by x^*) then x is a stationary with the property $\partial_g f(x_k) = \emptyset$ for certain $x_k \to x$. This may happen even if

$$Df(x) = 0 \quad \text{and} \quad 0 \in \operatorname{int} \partial^{Cl} f(x) \qquad \text{(Sägefunktion mit } x^2, \text{ see the lecture)}$$

and demonstrates one specific of $f \in C^{0,1}$ which is completely different from $f \in C^1$.

Now we characterize the situation $\partial_g f(x) = \emptyset$ for F-stationarity, in general. The seemingly unpleasant case $\partial^F f(\bar{x}) = \emptyset$ provides some additional information, namely: The point $\bar{x}$ cannot satisfy the necessary optimality condition for $\min_{\mathbb{R}^n} f$ even if we change f by adding any linear function.

Theorem 9.3.2 *It holds* $\partial^F f(\bar{x}) = \emptyset \quad \Leftrightarrow$

$$\exists u_0, ..., u_{n+1} \in \mathbb{R}^n \text{ such that } \sum_{\nu=0}^{n+1} u_\nu = 0 \text{ and } \sum_{\nu=0}^{n+1} \min Cf(\bar{x})(u_\nu) = -1. \tag{9.3.3}$$

Proof. Let $q(u) := \min Cf(\bar{x})(u)$.
($\Leftarrow$) The condition (9.3.3) implies $0 \notin \partial^F f(\bar{x})$ since $q(u_\nu) < 0$ holds for some ν. Considering $\hat{f} := f - x^*$ and using that $C\hat{f}(\bar{x})(u) = Cf(\bar{x})(u) - \langle x^*, u\rangle$, (9.3.3) also holds for $\hat{f}$. Thus, it holds $0 \notin \partial^F \hat{f}(\bar{x})$ and, equivalently, $x^* \notin \partial^F f(\bar{x})$.
($\Rightarrow$) Let $\partial^F f(\bar{x}) = \emptyset$. This means $\forall x^* \ \exists u$ such that $q(u) - \langle x^*, u\rangle < 0$. Thus the set $H = \{\ x^* \mid \ \langle x^*, u\rangle \le q(u)\ \forall u\}$ is empty. Let $Q = \operatorname{epi} q \subset \mathbb{R}^{n+1}$, $Q^c = \operatorname{conv} Q$. Then $0 \in Q^c$. If $0 \notin \operatorname{int} Q^c$, we obtain a contradiction by separation as follows: Some $(x^*, \tau^*) \neq 0$ fulfills

$$\langle x^*, u\rangle + \tau^* t \le 0 \quad \forall (u,t) : \ t \ge q(u).$$

Since $q(u) < \infty \forall u$, then $\tau^* \ge 0$ is impossible. Hence $\tau^* < 0$ and, without loss of generality, $\tau^* = -1$. But this yields with $t = q(u)$ that $x^* \in H$, a contradiction. Hence $0_{n+1} \in \operatorname{int} Q^c$ and $(0_n, -\varepsilon) \in Q^c$ for some small $\varepsilon > 0$ (the subscript shows the dimension). Using Caratheodory's theorem there are $n+2$ elements $(u_\nu, t_\nu) \in Q \subset \mathbb{R}^{n+1}$ and $\lambda_\nu \ge 0$ such that

$$\sum_{\nu=0}^{n+1} \lambda_\nu = 1 \quad \text{and} \quad \sum_{\nu=0}^{n+1} \lambda_\nu (u_\nu, t_\nu) = (0, -\varepsilon).$$

Setting $u'_\nu = \lambda_\nu u_\nu$, this yields $q(u'_\nu) = \lambda_\nu q(u_\nu) \le \lambda_\nu t_\nu$ as well as

$$\sum_{\nu=0}^{n+1} u'_\nu = 0 \text{ and } s := \sum_{\nu=0}^{n+1} q(u'_\nu) \le -\varepsilon.$$

Multiplying all u'_ν with $1/|s|$ yields the assertion. □

A deeper investigation of F-stationarity for $f \in C^{0,1}$ remains as an important subject for the future. It is closely related, by Lemma 9.1.1, to the study of the cones $\mathcal{N}_M$ and C_M which permit an analytical representation under certain "regularity", cf. sect. 3.5.4.

In chapter 4 and 10 one can see how this regularity in connected with some stability called calmness, and that calmness - finally - means solvability of related systems with some linear rate of convergence.

Chapter 10

Stability (regularity) of solutions

"Stability of solutions" is often replaced by "regularity of solutions" in the literature. We prefer the current terminology since "regularity" stands also often for constraint qualifications only.

A deeper analysis of the solution behavior in optimization is mainly required for multilevel models, cf. the discussion at the end of sect. 7.2. We already know that the behavior of feasible points, in parametric constraint sets, is closely connected with optimality conditions. If one assigns, to some parameter, the optimal points, then we obtain a differentiable function only in particular cases. Usually, the result is a multifunction or a non-differentiable function.

10.1 Definitions of Stability

In order to analyze and solve hierarchic problems [like (7.1.1), sect. 7.2] one is mostly interested in some kinds of Lipschitz-continuity of the solution map S, assigned to parametric optimization or more general problems. Here, "solution" may be taken in the local and global sense, and often it also denotes the couple of points satisfying the first order necessary conditions. Details are contained in sect. 7.2 and 7.1.

To have a sufficiently general model that covers all these variants, one can study inclusions

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

with solution set $S(p) = F^{-1}(p)$. Such problems may be equations, (quasi-)VI's, games, control problems et cet. Below, we shall suppose that S is a closed (= $\operatorname{gph} S$ is closed).

Notions of (local) Lipschitz stability:
Let $S : P \rightrightarrows X$ be closed and $\bar{z} = (\bar{p}, \bar{x}) \in \operatorname{gph} S$. We write ζ^0 for $(\bar{x}, \bar{p})$. Let

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

The next definitions generalize properties of the multivalued inverse $S = f^{-1}$ or of level sets $S(p) = \{x \in M \mid f(x) \le p\}$ for $f : M \subset X \to \mathbb{R}$. After each definition, we add an example that obeys the claimed property at $\bar{z} = 0$, but (if possible) not the others.

Definition 10.1.1 The mapping S is - at $\bar{z} \in \operatorname{gph} S$ - called

strLip : *strongly Lipschitz* if $\exists \varepsilon, \delta, L\ (> 0)$ such that

$$d(x', x) \le L\|p' - p\| \ \forall\ p, p' \in (\bar{p} + \delta B) \cap \operatorname{dom} S_\varepsilon,\ x' \in S_\varepsilon(p'),\ x \in S_\varepsilon(p), \tag{10.1.1}$$

i.e., S_ε is locally single-valued and Lipschitz on $\operatorname{dom} S_\varepsilon(p)$ near $\bar z$; $S_\varepsilon(p) = \emptyset$ is allowed. $(M = \mathbb{R}^+, X = \mathbb{R}, f(x) = x, S = f^{-1})$.

s.L.s: If, in addition, $\bar p \in \operatorname{int}\operatorname{dom} S_\varepsilon$ then S_ε is locally a Lipschitz function, and we call S *strongly Lipschitz stable* (s.L.s.). $(M = X = \mathbb{R}, f(x) = 2x - |x|, S = f^{-1})$.

ps.Lip, Aubin prop. : *pseudo–Lipschitz* if $\exists \varepsilon, \delta, L\ (> 0)$ such that

$$S_\varepsilon(p) \subset S(p') + L\|p' - p\|B \ \forall p, p' \in \bar p + \delta B. \tag{10.1.2}$$

$(M = X = \mathbb{R}^2, f(x, y) = x + y, S = f^{-1})$. Other notations (or equivalent notions) for the same fact: S^{-1} is metrically regular resp. pseudo–regular or S has the *Aubin property* [134]. With $p = \bar p$ in (10.1.2), one obtains $S(p') \neq \emptyset$ from $\bar x \in S_\varepsilon(p)$. Thus (10.1.2) yields $\bar p \in \operatorname{int}\operatorname{dom} S_\varepsilon$.

calm: *calm* if (10.1.2) holds for fixed $p' \equiv \bar p$ $(M = X = \mathbb{R}, f(x) \equiv 0, S = f^{-1})$.

upperLip : *upper Lipschitz* if $\exists \varepsilon, \delta, L\ (> 0)$ such that

$$S_\varepsilon(p) \subset \bar x + L\|p - \bar p\|B \ \ \forall p \in \bar p + \delta B. \tag{10.1.3}$$

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

u.L.s.: If, in addition, $\bar p \in \operatorname{int}\operatorname{dom} S_\varepsilon$ we call S *upper Lipschitz stable (u.L.s.)*. $(M = X = \mathbb{R}, f$ constant on short intervals, e.g. $[\frac{1}{2^{4k+1}}, \frac{1}{2^{4k}}], k > 0$ integer, and increasing with $Df(x) \equiv 1$ otherwise; $S = f^{-1})$.

lowerLip : *lower Lipschitz* if $\exists \delta, L\ (> 0)$ such that

$$S(p) \cap (\bar x + L\|p - \bar p\|B) \neq \emptyset \ \forall p \in \bar p + \delta B. \tag{10.1.4}$$

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

10.2 Interrelations and composed mappings

In some of these definitions, one may put $\varepsilon = \delta$. We used different constants for different spaces. The constant L is called a *rank* of the related stability.

The requirement $\bar p \in \operatorname{int}\operatorname{dom} S_\varepsilon$ means that solutions to $p \in F(x)$ are (locally) *persistent*, and the lower Lipsch. property quantifies this persistence in a Lipschitzian manner.

The notions concerning stability or regularity differ in the literature. So "s.L.s." and "strongly regular" mean often the same, and our "upper Lipschitz" is "locally upper Lipschitz" in [23] while "u.L.s." is "upper regular" in [70]. Further, "regularity" of multifunctions has been also defined in an alternative manner via local linearizations in [125].

Remark 10.2.1 If S is the (multivalued) inverse of $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, all these properties (except calmness) coincide with $\det Df(\bar x) \neq 0$. If f is only loc. Lipschitz, they are quite different. Concerning calmness compare the level sets of $f = x^2$ and $f = 0$ near $(0, 0)$. ◇

Remark 10.2.2 For fixed $\bar z \in \operatorname{gph} S$, one easily sees by the definitions:
(i) S is u.L.s. $\Leftrightarrow$ S is both upper and lower Lipschitz.
(ii) S is calm if S is upper Lipschitz.
(iii) S is upper Lipschitz $\Leftrightarrow$ S is calm and $\bar x$ is isolated in $S(\bar p)$.
(iv) S is s.L.s. $\Leftrightarrow$ S is pseudo-Lipschitz and $\operatorname{card} S_\varepsilon(p) \leq 1$ for p near $\bar p$.
(v) S is pseudo-Lipschitz (= Aubin prop.) $\Leftrightarrow$ S is lower Lipschitz at all points $z \in \operatorname{gph} S$ near $\bar z$ with fixed constants δ and L.
(vi) S is pseudo-Lipschitz $\Leftrightarrow$ S is both calm at all $z \in \operatorname{gph} S$ near $\bar z$ with fixed constants ε, δ, L and lower Lipschitz at $\bar z$. ◇

Composed mappings and intersections
Multifunctions are often written in particular forms which are more or less preferred in different papers. Examples:

(i) Let $S(p) = U(V(p))$ be a composed map where $V : P \rightrightarrows Y$ and $U : Y \rightrightarrows X$.

In many situations, then the Lipschitz properties at $\bar{z}$ [now $y^0 \in V(\bar{p})$ and $x^0 \in U(\bar{y})$] are consequences of the related properties for V and U at the corresponding points. One has, however, to shrink the image of V to the $\varepsilon-$ nbhds of $\bar{y}$ which appear in the definitions, i.e., one has to study composed maps of the form

$$S(p) = U(\ V(p) \cap (\bar{y} + \varepsilon B)\ ).$$

(ii) Let $S = \Gamma^{-1}$ with $\Gamma(x) := G(F(x))$ where $F : X \rightrightarrows Y,\ G : Y \rightrightarrows P$.

Then $x \in S(p) \Leftrightarrow p \in G(F(x)) \Leftrightarrow x \in F^{-1}(G^{-1}(p))$. This is situation (i) with $U = F^{-1}$ and $V = G^{-1}$.

(iii) The map S under (ii) can be written by intersections and projections:

Define $g : (X, Y) \rightrightarrows P$ as $g(x, y) = G(y)$ and $f : (X, Y) \rightrightarrows Y$ as $y \in f(x', y') \Leftrightarrow y = y' \in F(x')$ (otherwise $f(x', y') = \emptyset$). This yields $g^{-1}(p) = (X, G^{-1}(p))$, $f^{-1}(y) = (F^{-1}(y), \{y\})$ and $x' \in S(p) \Leftrightarrow (x', y') \in f^{-1}(y') \cap g^{-1}(p)$ for some y'.

(iv) Intersections $S(p_1, p_2) = F(p_1) \cap G(p_2)$ where $F : P_1 \rightrightarrows X,\ G : P_2 \rightrightarrows X$.

This is typical for (in)equality systems:

$F(p_1) = \{x \mid f(x) = p_1\}$ and $G(p_2) = \{x \mid g(x) \in p_2 + K\}$, $K \subset P_2$.

If f and g are loc. Lipschitz then calmness of F and G at $(p_1^0, \bar{x})$ and $(p_2^0, \bar{x})$, respectively, yields calmness of S at $(p_1^0, p_2^0, \bar{x})$, provided that one of the mappings

$$S_1(p_1) = \{x \mid f(x) = p_1,\ g(x) \in p_2^0 + K\}$$
$$S_2(p_2) = \{x \mid f(x) = p_1^0,\ g(x) \in p_2 + K\}$$

(with fixed second argument) is calm at the corresponding point, [69], Thm. 3.6.

For calmness, the upper Lipschitz and Aubin property of intersections we refer to [69] and [88]. For upper Lipschitz properties of composed maps we refer to [94], [95], [93], [117], [118] and [70]. In what follows we do not exploit special structures as above.

10.3 Stability of multifunctions via Lipschitz functions

10.3.1 Basic transformations

Though we are speaking about closed *multifunctions* $S : P \rightrightarrows X$ which act between B-spaces, calmness is a *monotonicity property* w.r. to two canonically assigned *Lipschitz functions*: the distance of x to $S(\bar{p})$ and the graph-distance

$$\psi_S(x, p) = \operatorname{dist}((p, x),\ \operatorname{gph} S),$$

defined via the norm $\|(p, x)\| = \max\{\|p\|, \|x\|\}$ or some equivalent norm in $P \times X$. The statements of this section even hold for $S : P \rightrightarrows X$ if P and X are only normed spaces.

Lemma 10.3.1 *S is calm at $(\bar{p}, \bar{x}) \in \operatorname{gph} S$ if and only if*

$$\exists \varepsilon > 0,\ \alpha > 0 \text{ such that } \alpha \operatorname{dist}(x, S(\bar{p})) \le \psi_S(x, \bar{p}) \quad \forall x \in \bar{x} + \varepsilon B. \quad \diamond \qquad (10.3.1)$$

In other words, calmness at $(\bar{p}, \bar{x})$ is violated iff

$$0 < \psi_S(x_k, \bar{p}) = o(\ \mathrm{dist}\,(x_k, S(\bar{p}))\ ) \text{ holds for some sequence } x_k \to \bar{x}. \tag{10.3.2}$$

Proof. A proof is possible as for Lemma 3.2 in [69]; we verify Lemma 10.3.1 for completeness. Let (10.3.1) hold true. Then, given $x \in S(p) \cap (\bar{x} + \varepsilon B)$, it holds

$$\alpha\ \mathrm{dist}\,(x, S(\bar{p})) \le \psi_S(x, \bar{p}) \le d((p, x), (\bar{p}, x)) = \|p - \bar{p}\|$$

i.e., calmness with $L > \frac{1}{\alpha}$. Conversely, let (10.3.1) be violated, i.e., (10.3.2) be true. Given any positive $\delta_k < o(\ \mathrm{dist}\,(x_k, S(\bar{p}))\ )$, we find $(y_k, \xi_k) \in \mathrm{gph}\, S$ such that

$$d((p_k, \xi_k),\ (\bar{p}, x_k)) < \psi_S(x_k, \bar{p}) + \delta_k < \ b_k := 2\ o(\ \mathrm{dist}\,(x_k, S(\bar{p}))\ ).$$

In addition, the triangle inequality $\mathrm{dist}\,(x_k, S(\bar{p})) \le d(x_k, \xi_k) + \mathrm{dist}\,(\xi_k, S(\bar{p}))$ yields

$$\mathrm{dist}\,(\xi_k, S(\bar{p})) \ge \mathrm{dist}\,(x_k, S(\bar{p})) - d(\xi_k, x_k) > \mathrm{dist}\,(x_k, S(\bar{p})) - b_k.$$

Using also the evident inequality $\|p_k - \bar{p}\| < b_k$, we thus obtain for $\xi_k \in S(p_k)$,

$$\frac{\|p_k - \bar{p}\|}{\mathrm{dist}\,(\xi_k, S(0))} < \frac{b_k}{\mathrm{dist}\,(x_k, S(\bar{p})) - b_k} \ \to 0 \quad \text{as } k \to \infty.$$

Hence, since $\xi_k \to \bar{x}$, S cannot be calm at $(\bar{p}, \bar{x})$. □

Estimates of ψ_S, for composed systems, can be found in [69]. Condition (10.3.1) requires that $\psi_S(., \bar{p})$ increases in a Lipschitzian manner if x (near $\bar{x}$) leaves $S(\bar{p})$. Clearly, this property depends on the local structure of the boundaries of $\mathrm{gph}\, S$ and $S(\bar{p})$. For convex multifunctions (i.e. $\mathrm{gph}\, S$ is convex), ψ_S and $d(., S(\bar{p}))$ are convex functions.

Combined with Remark 10.2.2(iii), condition (10.3.1) characterizes the Aubin property, too. For similar characterizations of other "stabilities", we refer to [72]. The distance ψ_S can be applied also for both characterizing optimality and computing solutions in optimization models via penalization [88, 69] and [70, Chapt. 2]; for the particular context of exact penalization techniques, see also [21, 13, 11]. Detailed investigations of locally Lipschitz functions can be found in [13, 70]. Approximate minima of ψ_S play a role below.

Setting $G = \psi_S$ we obtain a (globally) Lipschitz function, assigned to S, such that

$$(p, x) \in \mathrm{gph}\, S \ \Leftrightarrow\ G(x, p) \le 0. \tag{10.3.3}$$

For every such description of $\mathrm{gph}\, S$, it follows

Lemma 10.3.2 *S is calm at $(\bar{p}, \bar{x})$ if*

$$\exists \varepsilon > 0,\ \alpha > 0 \text{ such that } \ \alpha\ \mathrm{dist}\,(x, S(\bar{p})) \le\ G(x, \bar{p}) \quad \forall x \in \bar{x} + \varepsilon B. \quad \diamond \tag{10.3.4}$$

Proof. Given any $\delta > 0$ choose $(p', x') \in \mathrm{gph}\, S$ with $d((p', x'), (\bar{p}, x)) < \psi_S(x, \bar{p}) + \delta$. Then $G(x', p') \le 0$ yields with some Lipschitz constant L, $G(x, \bar{p}) \le L\ (\psi_S(x, \bar{p}) + \delta)$ and $\frac{\alpha}{L}\ \mathrm{dist}\,(x, S(\bar{p})) \le \frac{G(x, \bar{p})}{L} \le \psi_S(x, \bar{p}) + \delta$. Hence one obtains, via $\delta \downarrow 0$, that (10.3.1) holds with some new $\alpha := \frac{\alpha}{L}$. □

The conditions (10.3.3) and (10.3.4) are only sufficient for calmness (put, e.g., $G = \psi_S^2$). Nevertheless, since one may put $G = \psi_S$, it holds

Corollary 10.3.3 *A multifunction S is calm at $(\bar{p}, \bar{x})$ $\Leftrightarrow$ there is some Lipschitz function $G : X \times P \to \mathbb{R}$ satisfying (10.3.3) and (10.3.4).* ◇

Finally, with any locally Lipschitz function $\phi : X \to \mathbb{R}$ such that

$$c_1\phi(x) \le \psi_S(x,\bar{p}) \le c_2\phi(x) \quad \text{for } x \text{ near } \bar{x} \text{ and certain constants } 0 < c_1 \le c_2 \tag{10.3.5}$$

and with the mapping

$$\Sigma(q) = \{x \in X \mid \phi(x) \le q\}, \tag{10.3.6}$$

condition (10.3.1) of Lemma 10.3.1 is equivalent to

$$\exists \varepsilon > 0,\ \alpha > 0 \text{ such that } \alpha \operatorname{dist}(x,\ \Sigma(0)) \le\ q\ \ \forall x \in \bar{x} + \varepsilon B \text{ with } \phi(x) = q > 0. \tag{10.3.7}$$

This verifies

Corollary 10.3.4 *Calmness for any multifunction S at $(\bar{p},\bar{x})$ can be reduced to calmness of a Lipschitzian inequality only, namely to calmness of Σ (10.3.6) at $(0,\bar{x})$. It suffices to take $\phi = \psi_S(.,\bar{p})$ or another Lipschitz function ϕ satisfying (10.3.5).* ◇

10.3.2 Solutions of (exact) penalty problems

The function $\psi = \psi_S$ can be applied for both *characterizing optimality and computing* solutions in optimization models via penalization:
Let S be calm at (p',x') and $X'_\varepsilon := x' + \varepsilon B$. Further, suppose

$$\begin{array}{l} f \text{ is a Lipschitz function with rank } L_f \text{ on some nbhd of} \\ S(p') \cap X'_\varepsilon \quad \text{and } x' \text{ is a minimizer of } f \text{ on } S(p') \cap X'_\varepsilon. \end{array} \tag{10.3.8}$$

Then, x' is a (free) local minimizer of $x \mapsto P_k(x) := f(x) + k\ \psi(x,p')$ for large k (this is well-known and can be immediately seen). The current penalty function P_k is also said to be an exact penalization since, for large k, the original solution x' (not only an approximation) solves the auxiliary problem.

To find x' or some $\hat{x} \in S(p') \cap X'_\varepsilon$ with $f(\hat{x}) = f(x')$ by means of a penalty method, several techniques can be applied. In particular, one may replace ψ by ψ^s $(s > 0)$, in order to solve

$$\text{minimize } \ Q_k(x) := f(x) + k\ \psi(x,p')^s \ \text{ on } X'_\varepsilon \ \ (k \to \infty). \tag{10.3.9}$$

In this respect, it is important that, if $\dim P + \dim X < \infty$, the function ψ^2 is *semismooth*, a useful property for Newton techniques, cf. [103], [70] or sect. 6.2.

Convergence of minimizers x_k to (10.3.9) for given p':
We prove this convergence, to discuss the role of s and the possibility of removing the constraint $x \in X'_\varepsilon$ in (10.3.9).
Let $x_k \in \operatorname{argmin} Q_k$ and let $u_k \in S(p')$ satisfy $d(x_k,u_k) = \operatorname{dist}(x_k, S(p'))$. Then

$$\begin{array}{rl} f(x') & \ge f(x_k) + k\psi(x_k,p')^s \\ & \ge f(u_k) - L_f d(x_k,u_k) + k\psi(x_k,p')^s \\ & \ge f(x') \ - L_f d(x_k,u_k) + k\psi(x_k,p')^s. \end{array} \tag{10.3.10}$$

Hence $L_f d(x_k,u_k) \ge k\psi(x_k,p')^s$. Now calmness (10.3.1) permits to continue

$$L_f d(x_k,u_k) \ge k\lambda^s d(x_k,u_k)^s. \tag{10.3.11}$$

For $s > 1$, this entails

$$\operatorname{dist}(x_k, S(p'))^{s-1} = d(x_k,u_k)^{s-1} \ \le\ \frac{L_f}{k\lambda^s} \ \to 0 \ \text{ and } \ \psi(x_k,p') \to 0. \tag{10.3.12}$$

Next let $0 < s \leq 1$. Now $0 < d(x_k, u_k) \leq 1$ implies $d(x_k, u_k) \leq d(x_k, u_k)^s$. Thus

$$k\lambda^s > L_f \;\Rightarrow\; x_k = u_k.$$

On the other hand, if $d(x_k, u_k) \geq 1$ then $\varepsilon \geq \|x_k - u_k\| \geq \|x_k - u_k\|^s \geq 1$. Hence this case cannot appear whenever k is large enough such that $\varepsilon L_f < k\lambda^s$. Thus

$$x_k = u_k \in S(p')$$

holds again for sufficiently large k.
In consequence, every cluster point $\hat{x}$ of x_k is feasible and satisfies $f(\hat{x}) = f(x')$, (under the supposed calmness) independent of the choice of $s > 0$. □

Local and global minimizers:
If x' was even the *unique* global minimizer of f on $S(p') \cap X'_\varepsilon$ (such points are called strict local minimizers) then (10.3.12) and $f(\hat{x}) = f(x')$ imply $\hat{x} = x'$ and $x_k \to x'$. Hence also

$$x_k \in \operatorname{int} X_\varepsilon \quad \text{holds for large } k,$$

and implies, as typical for penalty methods, that x_k is a *free local minimizer of* $Q_k(x)$. □

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

$$\begin{array}{l} \exists \varepsilon > 0, \lambda > 0, r > 0 \quad \text{such that} \\ \psi(x, p') \geq \lambda \operatorname{dist}(x, S(p'))^r \;\; \forall x \in X'_\varepsilon := x' + \varepsilon B. \end{array} \tag{10.3.13}$$

Then only the critical value $s^* = 1$ from above changes: $s^* = 1/r$. Of course, now Q_k (10.3.9) is not differentiable (like under calmness for $s = 1$) and existing derivatives may be unbounded. So they must be made bounded artificially if they are "too large". This is standard in every program, it can be avoided by several smoothing techniques and becomes necessary only if x_k is already "almost" feasible. Sufficient conditions for Hölder behavior of stationary points can be found in [40].

10.4 Stability and algorithms

The next sections is based on results, just published by J. Heerda, D. Klatte and B. Kummer. As above, let $S = F^{-1} : P \rightrightarrows X$ (Banach spaces). Given $(p, x) \in \operatorname{gph} S$ near $\bar{z} = (\bar{p}, \bar{x})$ and π near $\bar{p}$; in brief *given initial points* x, p, π *near* $\bar{z}$, we want to

$$\text{determine some } x(\pi) \in S(\pi) \text{ with } d(x(\pi), x) \leq L\|\pi - p\| \;\text{ by algorithms.}$$

The existence of $x(\pi)$ is claimed under the Aubin property (or under calmness if $\pi = \bar{p}$).

In the sequel, we are interested in procedures which find $x(\pi)$ with well-defined rate of convergence, *exactly under the Aubin property* (or under calmness, respectively). By saying that some algorithm has this specific property (for initial points near $\bar{z}$) we connect stability and solution methods in a direct and general manner. Due to the aimed generality, our methods ALG1 and PRO(γ) are quite simple. Nevertheless they involve several more or less fast local methods under additional assumptions.

10.5 The algorithmic framework

The subsequent algorithm is a scheme for particular procedures which compute $x(\pi) \in S(\pi)$. Let $\lambda \in (0,1)$ be given.

ALG1 Put $(p_1, x_1) = (p, x) \in \operatorname{gph} S$ and choose $(p_{k+1}, x_{k+1}) \in \operatorname{gph} S$ in such a way that

$$\begin{array}{lllll} (i) & \|p_{k+1}-\pi\| & - \|p_k-\pi\| & \leq & -\lambda\, d(x_{k+1},x_k) \\ (ii) & \|p_{k+1}-\pi\| & - \|p_k-\pi\| & \leq & -\lambda\, \|p_k - \pi\|. \end{array} \quad \text{and} \tag{10.5.1}$$

equiv. in $\operatorname{gph} S$: for (x,p) exists (p',x') with:

$$\begin{array}{lll} (i) & \|p'-\pi\| & + \lambda\, d(x',x) \leq \|p-\pi\| \\ (ii) & \|p'-\pi\| & \leq (1-\lambda)\|p-\pi\|. \end{array} \tag{10.5.2}$$

Definition 10.5.1 We call ALG1 *applicable* if (p_{k+1}, x_{k+1}) exist in each step (for some fixed $\lambda > 0$). Having calmness in mind, we apply the same algorithm with fixed $\pi \equiv \bar{p}$.

Interpretation: Identify p_k with some $f(x_k) \in F(x_k)$. Then (10.5.1)(i) requires more familiar

$$\frac{\|f(x_{k+1})-\pi\| - \|f(x_k)-\pi\|}{d(x_{k+1},x_k)} \leq -\lambda \quad \text{for } x_{k+1} \neq x_k, \tag{10.5.3}$$

and (10.5.1)(ii) is only one of various conditions which ensure $\|f(x_k)-\pi\| \to 0$ for this (non-increasing) sequence. In this interpretation, ALG1 is a *descent method* for $x \mapsto \|f(x)-\pi\|$.

Remark 10.5.1 (Reducing the stepsize) As in every method of this type, one may start with some $\lambda = \lambda_1 > 0$ and, if (p_{k+1}, x_{k+1}) satisfying (10.5.1) cannot be found, decrease λ by a constant factor, e.g., $\lambda_{k+1} = \frac{1}{2}\lambda_k$ while $(p_{k+1}, x_{k+1}) := (p_k, x_k)$ remains unchanged. In this form, *being applicable* coincides with $\inf \lambda_k \geq \alpha > 0$, and we need the same α w.r. to the possible starting points. This reduction of λ (like for the Armijo-Goldstein stepsize in free minimization) is possible for all algorithms we shall speak about, though we make explicitly use of it only for ALG2 and ALG3, cf. Thm. 10.7.6. ◇

Theorem 10.5.2 *Let $S : P \rightrightarrows X$ be closed. If ALG1 is applicable for given initial points x, p, π near $\bar{z}$, then the sequence converges $(x_k, p_k) \to (x(\pi), \pi) \in \operatorname{gph} S$, and*

$$d(x(\pi), x) \leq \frac{\|\pi - p\|}{\lambda}. \tag{10.5.4}$$

Moreover,

(i) The Aubin property of S holds at $\bar{z} = (\bar{p}, \bar{x}) \Leftrightarrow$ ALG1 is applicable, for some fixed $\lambda \in (0,1)$ and all initial points x, p, π near $\bar{z}$.

(ii) The same holds true, but with fixed $\pi \equiv \bar{p}$, in view of calmness of S at $\bar{z}$. ◇

Proof. If ALG1 is applicable then, beginning with $n = 1$ and $x_1 = x$, the estimate

$$d(x_{n+1}, x) \leq \sum_{k=1}^{n} d(x_{k+1}, x_k) \leq \frac{\|p_1 - \pi\| - \|p_{n+1} - \pi\|}{\lambda} \tag{10.5.5}$$

follows from (10.5.1)(i) by complete induction. So, a Cauchy sequence $\{x_k\}$ will be generated and (10.5.5) entails (10.5.4) for the limit $x(\pi) = \lim x_k$.
By (10.5.1)(ii), it follows $\|p_{k+1} - \pi\| \leq (1-\lambda)\, \|p_k - \pi\|$ and $p_k \to \pi$. Since S is closed, so also $x(\pi) \in S(\pi)$ is valid.

(i), (ii) ($\Rightarrow$) Let the Aubin property be satisfied with related constants L, ε, δ in (10.1.2). Then we obtain the existence of the next iterates whenever $0 < \lambda < L^{-1}$ and $\|\pi - \bar{p}\| + d((p,x),\bar{z})$ was small enough. Indeed, if $\hat{\varepsilon} := \min\{\varepsilon, \delta\}$ and

$$\max\{\ \frac{\|p-\bar{p}\| + \|\pi - \bar{p}\|}{\lambda}\ ,\ d(x,\bar{x})\ \} < \tfrac{1}{2}\hat{\varepsilon}$$

then $d(x_k,\bar{x}) < \hat{\varepsilon}$ and $\|p_k - \bar{p}\| < \hat{\varepsilon}$ follow from (10.5.5) by induction. Thus, for any $p_{k+1} \in \operatorname{conv}\{p_k, \pi\}$ satisfying (10.5.1)(ii) there is some $x_{k+1} \in S(p_{k+1})$ such that

$$d(x_{k+1}, x_k) \le L\|p_{k+1} - p_k\| \le \frac{\|p_{k+1} - p_k\|}{\lambda} = \frac{\|p_k - \pi\| - \|p_{k+1} - \pi\|}{\lambda}.$$

Hence also some x_{k+1} exists as required in (10.5.1)(i).

Having only calmness, the existence of a related element $x_{k+1} \in S(p_{k+1})$ is ensured by setting $p_{k+1} = \pi = \bar{p}$ (the sequence becomes constant).

(i), (ii) ($\Leftarrow$) If the Aubin property is violated and $\lambda > 0$, then (by definition only) one finds points $(p,x) \in \operatorname{gph} S$ arbitrarily close to $\bar{z}$, and π arbitrarily close to $\bar{p}$, such that $\operatorname{dist}(x, S(\pi)) > \frac{\|p-\pi\|}{\lambda}$. Consequently, it is also impossible to find some related $x(\pi)$ by ALG1.

In view of calmness, the same arguments apply to $\pi \equiv \bar{p}$. □

Remark 10.5.3 The following elementary consequences are worth to be mentioned.

(i) Thm. 10.5.2 still holds after replacing (10.5.1)(ii) by *any condition* which ensures, along with (10.5.1)(i), that $p_k \to \pi$. Hence, instead of (10.5.1)(ii), one can require (e.g.) that the stepsize is linearly bounded below by the current error

$$d(x_{k+1}, x_k) \ge c\ \|p_k - \pi\| \quad \text{for some } c > 0. \tag{10.5.6}$$

Evidently, (10.5.1)(i) and (10.5.6) yield

$$\|p_{k+1} - \pi\| - \|p_k - \pi\| \ \le -\lambda d(x_{k+1}, x_k)\ \le -\lambda c\ \|p_k - \ \pi\|.$$

This implies (10.5.1) with new λ since we may choose $\lambda' < \lambda,\ \lambda' \in (0, \lambda c)$.

(ii) Generally, (10.5.6) does not follow from (10.5.1), take the function $F(x) = \sqrt[3]{x}$. So requiring (10.5.1) is weaker than (10.5.1)(i) and (10.5.6).

(iii) Thm. 10.5.2 remains true (with the same proof) if one additionally requires $p_k \in \operatorname{conv}\{p_1, \pi\} \forall k$ in (10.5.1). ◇

Without considering sequences explicitly, the statements (i), (ii) of Thm. 10.5.2 can be written as stability criterions.

Corollary 10.5.4
(i) The Aubin property of S holds at $\bar{z} = (\bar{p}, \bar{x})$ $\Leftrightarrow$ For some $\lambda \in (0,1)$ and all x, p, π near $\bar{z}$ there exist $(p', x') \in \operatorname{gph} S$ such that

$$\begin{array}{lllll} (i) & \|p' - \pi\| \ - \ \|p - \pi\| & \le & -\lambda\ d(x', x) & \text{and} \\ (ii) & \|p' - \pi\| \ - \ \|p - \pi\| & \le & -\lambda\ \|p - \pi\|. & \end{array} \tag{10.5.7}$$

(ii) The same statement, with fixed $\pi \equiv \bar{p}$, holds in view of calmness of S at $\bar{z}$. ◇

Proof. It suffices to show that ALG1 is applicable under (10.5.7). Denoting (p', x') by $\phi(p,x)$, define

$$(p_1, x_1) = (p, x) \text{ and } (p_{k+1}, x_{k+1}) = \phi(p_k, x_k). \tag{10.5.8}$$

Due to (10.5.5), (p^n, x^n) belongs to an arbitrary small nbhd Ω of $\bar{z}$ for all initial points $(x,p), \pi$ sufficiently close to $\bar{z}$ and $\bar{p}$, respectively. Hence ALG1 is applicable. □

The similarity of the statements for calmness and the Aubin property does not imply that ALG1 runs in the same way under each of these properties.

Aubin property: If ALG1 is applicable for all initial points near $\bar{z} \in \operatorname{gph} S$, we can first fix any $p_{k+1} \in \operatorname{conv}\{p_k, \pi\}$ satisfying (10.5.1)(ii) and next find (since the Aubin property holds at $\bar{z}$ by Thm. 10.5.2 and (p_k, x_k) is close to $\bar{z}$) some $x_{k+1} \in S(p_{k+1})$ satisfying (10.5.1)(i).

In other words, $x(\pi)$ can be determined by small steps. This is not important for estimating $d(x, x(\pi))$, but for constructing concrete algorithms which use local information for F near (p_k, x_k) in order to find (p_{k+1}, x_{k+1}).

Calmness: Though every sequence in (10.5.1) leads us to $x(\pi) \in S(\pi)$, we can *guarantee* that some feasible x_{k+1} exists *for some already given* p_{k+1}, only if $p_{k+1} = \pi = \bar{p}$.

In other words, the sequence could be trivial, $(p_k, x_k) = (\pi, x(\pi)) \ \forall\ k \geq k_0$, since calmness allows (by definition) that $S(p) = \emptyset$ for $p \notin \{p_1, \bar{p}\}$.

In this case, local information for F near (p_k, x_k) cannot help to find x_{k+1} for given $p_{k+1} \in \operatorname{int} \operatorname{conv}\{p_1, \pi\}$.

However, for many mappings which describe constraint systems or solutions of variational inequalities, this is not the typical situation. In particular, if $\operatorname{gph} S$ is convex then $S(p_{k+1}) \neq \emptyset$ holds for each $p_{k+1} \in \operatorname{conv}\{p_1, \pi\}$ (since $S(\pi)$ and $S(p_1)$ are non-empty by assumption). This remains true if $\operatorname{gph} S$ is a finite union of closed convex sets C_i since $I(z) := \{i \mid z \in C_i\} \subset I(\bar{z})$ holds for all initial points $z = (p, x) \in \operatorname{gph} S$ near $\bar{z}$.

10.6 Stability via approximate projections

The following *approximate projection method* (onto $\operatorname{gph} S$) has, in contrast to ALG1, the advantage that iteration points throughout exist (for $\gamma > 0$). "Stability" is now characterized by linear order of convergence. Let $\gamma \geq 0$.

PRO(γ) Put $(p_1, x_1) = (p, x) \in \operatorname{gph} S$ and choose $(p_{k+1}, x_{k+1}) \in \operatorname{gph} S$ in such a way that

$$d(x_{k+1}, x_k) + \frac{\|p_{k+1} - \pi\|}{\lambda} \leq \inf_{(p', x') \in \operatorname{gph} S} \left[\ d(x', x_k) + \frac{\|p' - \pi\|}{\lambda}\ \right] + \gamma \|p_k - \pi\|. \quad (10.6.1)$$

Theorem 10.6.1

(i) The Aubin property of S holds at $\bar{z} = (\bar{p}, \bar{x})$ $\Leftrightarrow$ PRO(γ) generates, for some $\lambda > 0$ and all initial points x, p, π near $\bar{z}$, a sequence satisfying

$$\lambda\ d(x_{k+1}, x_k)\ +\ \|p_{k+1} - \pi\|\ \leq\ \theta \|p_k - \pi\| \quad \text{with some fixed } \theta < 1. \quad (10.6.2)$$

(ii) The same statement, with $\pi \equiv \bar{p}$, holds in view of calmness of S at $\bar{z}$. $\diamond$

Note. Obviously (10.6.2) means $\|p_{k+1} - \pi\| - \|p_k - \pi\| \leq -\lambda\ d(x_{k+1}, x_k) - (1 - \theta)\|p_k - \pi\|$ which again implies (10.5.1)(i) and convergence $x_k \to x(\pi) \in S(\pi)$ satisfying (10.5.4). Having the related stability property, our proof indicates that PRO(γ) can be applied with any $\gamma > 0$, provided that λ is sufficiently small, see the requirement $\lambda(L + \gamma) < 1$.

Proof. (i) ($\Rightarrow$) Suppose the Aubin property with rank L, and fix $\lambda \in (0,\ (L + \gamma)^{-1})$. Considering again points near $(\bar{p}, \bar{x})$ there exists $\hat{x} \in S(\pi)$ with $d(\hat{x}, x_k) \leq L\|\pi - p_k\|$. This yields for the approximate minimizer in (10.6.1)

$$d(x_{k+1}, x_k) + \frac{1}{\lambda}\|p_{k+1} - \pi\| \leq\ d(\hat{x}, x_k) + \frac{1}{\lambda}\|\pi - \pi\|\ + \gamma\|p_k - \pi\| \leq\ (L + \gamma)\|p_k - \pi\|$$

and implies $\lambda d(x_{k+1}, x_k) + \|p_{k+1} - \pi\| \le \lambda(L+\gamma)\|p_k - \pi\|$ as well as (10.6.2) with $\theta = \lambda(L+\gamma) < 1$.

($\Leftarrow$) Conversely, let PRO(γ) (or any algorithm) generate a sequence satisfying (10.6.2) with some $\lambda > 0,\ \theta \in (0,1)$ and all related initial points. Then also (10.5.1)(i) is valid for the current sequences and $\|p_{k+1} - \pi\|$ vanishes. By Thm. 10.5.2 and Remark 10.5.3(i) so the Aubin property holds true.

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

Combining condition (10.3.1) for calmness of S at $\bar{z}$ (with the norm $\lambda\|.\|_X + \|.\|_P$ in $X \times P$) and condition (10.6.2) with $\pi = \bar{p}$, one directly obtains the calmness estimate

$$\theta\ \|p_k - \bar{p}\| \ge\ \lambda\ d(x_{k+1}, x_k) + \|p_{k+1} - \pi\| \ \ge\ \alpha\ \mathrm{dist}\,(x_k, S(\bar{p})). \tag{10.6.3}$$

PRO(γ) as penalty method:

The term $\frac{1}{\lambda}\|p' - \pi\|$ in the objective of (10.6.1) can be understood as penalizing the requirement $p' = \pi$. So we simply solve (again approximately), for given x_k and π,

$$\min\ d(x', x_k) \quad \text{such that}\ \ (p', x') \in \mathrm{gph}\, S \text{ and } p' = \pi$$

by penalizing equation $p' = \pi$. Now p_k is the current approximation of π, assigned to x_k.

10.7 Stability of a locally Lipschitz operator

Next identify F with a function $f \in C^{0,1}(X, P)$. We shall see that, in this situation, condition (10.5.1) can be written (up to a possibly new constant λ) as

$$\|f(x_{k+1}) - \pi\| - \|f(x_k) - \pi\| \le -\lambda\ d(x_{k+1}, x_k)\ \text{ and }\ d(x_{k+1}, x_k) \ge \lambda\|f(x_k) - \pi\| \tag{10.7.1}$$

or equivalently as

$$\|f(x_k) - \pi\| - \|f(x_{k+1}) - \pi\| \ge \lambda d(x_{k+1}, x_k) \ge \lambda^2 \|f(x_k) - \pi\|.$$

This permits a stability characterizations in terms of minimizing sequences with a stepsize estimate as in Remark 10.5.3(i).

Corollary 10.7.1 *Let $f \in C^{0,1}(X, P)$. Then $S = f^{-1}$ obeys the Aubin property at $(0, \bar{x}) \Leftrightarrow\ \ \exists \lambda \in (0,1)$ such that, for each x_1 near $\bar{x}$ and π near the origin, there is a minimizing sequence $\{x_k\}_{k\ge 1}$ for $x \mapsto \|f(x) - \pi\|$ satisfying (10.7.1). With fixed $\pi = 0$, this condition describes calmness of S at $(0, \bar{x})$.* ◇

Proof. If ALG1 is applicable then convergence of $\{x_k\}$ and (10.5.1) yield with $p = f(x)$, since f is locally Lipschitz,

$$-Cd(x_{k+1}, x_k) \le \|f(x_{k+1}) - \pi\| - \|f(x_k) - \pi\| \le -\lambda\ \|f(x_k) - \pi\|$$

for some $C > 0$, hence (10.5.6) is necessarily satisfied. The latter implies, up to a new constant in (10.5.1)(ii), that (10.5.1) and the requirements

$$\|f(x_{k+1}) - \pi\| - \|f(x_k) - \pi\| \le -\lambda\ d(x_{k+1}, x_k) \text{ and } d(x_{k+1}, x_k) \ge\ c\ \|f(x_k) - \pi\|$$

(for $\lambda, c > 0$) are equivalent. Setting $\lambda := \min\{\lambda,\ c\}$, we need one constant only which gives (10.7.1). □

As in Corollary 10.5.4, to show the related stability, it suffices to verify (10.7.1) for x_1 near $\bar{x}$ and appropriate x_2 only. This yields

Corollary 10.7.2
(i) The Aubin property of $S = f^{-1}$ holds at $\bar{z} = (0, \bar{x})$ $\Leftrightarrow$ For some $\lambda \in (0,1)$ and all points (x, π) near $(\bar{x}, 0)$ there is some x' such that

$$\|f(x') - \pi\| - \|f(x) - \pi\| \le -\lambda\, d(x', x) \quad \text{and} \quad d(x', x) \ge \lambda \|f(x) - \pi\|. \tag{10.7.2}$$

(ii) The same statement, with fixed $\pi \equiv 0$, holds in view of calmness of S at $(0, \bar{x})$. $\diamond$

10.7.1 Calmness and the relative slack for inequality systems

In particular, Corollary 10.7.1 and 10.7.2 apply to system

$$g(x) \le p,\ h(x) = q, \quad g, h \in C^{0,1} \qquad (\textit{loc. Lipsch.})$$

after setting $f = (g^+, h)$. However, for the sake of simplicity we assume that the equations are written as inequalities, and study, first with $I = \{1, ..., m\}$, calmness of

$$\Sigma(p) = \{x \in X \mid g_i(x) \le p_i,\ \forall\, i \in I\}, \quad g \in C^{0,1} \tag{10.7.3}$$

at $(0, \bar{x})$ with a Banach space X. We write

$$g^m(x) = \max_i\ g_i(x)$$

and define, for $g^m(x) > 0$, some *relative slack of* g_i in comparison with g^m,

$$s_i(x) = \frac{g^m(x) - g_i(x)}{g^m(x)} \quad (\ge 0). \tag{10.7.4}$$

In the special case of $g \in C^1$, $X = \mathbb{R}^n$, the following condition (10.7.6) differs just by the additionally appearing quantities $s_i(x)$ from the MFCQ-condition for inequalities.

Theorem 10.7.3 *Let $g^m(\bar{x}) = 0$. Σ (10.7.3) is calm at $(0, \bar{x})$ $\Leftrightarrow$ There exist some $\lambda \in (0,1)$ and a nbhd Ω of $\bar{x}$ such that: For all $x \in \Omega$ with $g^m(x) > 0$ there exist $u \in \operatorname{bd} B$ and $t > 0$ satisfying*

$$\frac{g_i(x + tu) - g_i(x)}{t} \le \frac{g^m(x) - g_i(x)}{t} - \lambda\ \forall i \quad \textit{and} \quad \lambda g^m(x) \le t \le \frac{1}{\lambda} g^m(x). \tag{10.7.5}$$

Moreover, if $g \in C^1$, one may delete t and replace (10.7.5) by

$$Dg_i(\bar{x})u \ \le\ \frac{s_i(x)}{\lambda} - \lambda \quad \forall i. \qquad \diamond \tag{10.7.6}$$

Proof. We study the system $f(x) := (g^m)^+(x) = r$ which is calm at $(0, \bar{x})$ iff so is Σ. In accordance with Corollary 10.7.2, calmness means that some $\lambda \in (0,1)$ satisfies:

$$\forall x \text{ near } \bar{x} \text{ with } g^m(x) > 0\ \exists x' \text{ such that}$$

$$(g^m)^+(x') - g^m(x) \le -\lambda\, d(x', x) \ \text{ and } \ d(x', x) \ge \lambda\, g^m(x). \tag{10.7.7}$$

Writing here $x' = x + tu$ with $\|u\| = 1$ and $t > 0$, and defining $Q_i = \frac{g_i(x') - g_i(x)}{t}$ we have $g_i(x') = g_i(x) + tQ_i$. Then the first condition of (10.7.7) implies both

$$\begin{array}{lll} t & \le \frac{g^m(x)}{\lambda} & \text{and} \\ g_i(x) + tQ_i \quad (= g_i(x')\ ) & \le g^m(x) - \lambda t\ \forall\ i & \end{array} \tag{10.7.8}$$

Hence (10.7.7) claims exactly (10.7.5). It remains to study the case of $g \in C^1$. First note that (10.7.5) yields, due to $\lambda g^m(x) \le t$,

$$\frac{g_i(x+tu)-g_i(x)}{t} \le \frac{g^m(x)-g_i(x)}{\lambda\, g^m(x)} - \lambda \;\forall i \quad \text{and} \quad \lambda g^m(x) \le t \le \frac{1}{\lambda} g^m(x). \tag{10.7.9}$$

Due to the uniform convergence

$$\sup_{i\in I,\ \|u\|=1} \Big| \frac{g_i(x+tu)-g_i(x)}{t} - Dg_i(\bar{x})u \Big| \to 0 \quad \text{as } x \to \bar{x}, t \downarrow 0, \tag{10.7.10}$$

now (10.7.9) implies (10.7.6) (with possibly smaller λ). Hence (10.7.5) implies (10.7.6). Conversely, having (10.7.6) it suffices to put $t = \lambda g^m(x)$ in order to obtain (10.7.5) (possibly with $\lambda' < \lambda$), too. This completes the proof. □

Notes (modifying Thm. 10.7.3):

(i) Instead of considering all $x \in \Omega$ with $g^m(x) > 0$, it suffices to regard only

$$x \in \Omega \;\text{ with }\; 0 < g^m(x) < \lambda \|x - \bar{x}\| \tag{10.7.11}$$

since, for $g^m(x) \ge \lambda\|x-\bar{x}\|$, it holds the trivial calmness estimate

$$\operatorname{dist}(x, \Sigma(0)) \le \|x-\bar{x}\| \le \frac{1}{\lambda} g^m(x) \tag{10.7.12}$$

and one may put $u = \frac{\bar{x}-x}{\|\bar{x}-x\|},\ t = \|\bar{x}-x\|$ in the theorem. Since λ may be arbitrarily small, so *calmness depends only on sequences $x \to \bar{x}$ satisfying $g^m(x) = o(x-\bar{x}) > 0$.*

(ii) Trivially, (10.7.5) is equivalent to

$$g^m(x+tu) \le\; g^m(x) - \lambda t \;\text{ and }\; \lambda^2 g^m(x) \le \lambda t \le\; g^m(x). \tag{10.7.13}$$

(iii) For $g \in C^1$ condition (10.7.6) can be replaced (possibly with smaller Ω and λ) by

$$Dg_i(x)u \;\le\; \frac{s_i(x)}{\lambda} - \lambda \quad \forall i \tag{10.7.14}$$

The torus-condition (10.7.5):

Generally, because the stepsize t in (10.7.5) is restricted to a compact interval in the positive half-line, the left-hand side in (10.7.5), $\frac{g_i(x+tu)-g_i(x)}{t}$ compares points the difference tu of which belongs to a torus. Thus, without additional assumptions, the assigned quotients [*and in consequence, also calmness*] cannot be described by known generalized derivatives since such derivatives consider always preimage points $x, x+tu$ which may be *arbitrarily close* to each other. In particular, this holds for Tf, Cf, D^*f, discussed below though there are several papers where Cf and D^*f have been applied for establishing calmness conditions.

Notice: *It follows that every sufficient calmness condition in terms of these derivatives cannot be necessary in the class of all loc. Lipsch. functions.*

Remark 10.7.4 (*Infinitely many constraints.*) As in usual semi-infinite programs, one can consider Σ (10.7.3) with a compact topological space I, $\|p\| = \sup_i |p_i|$, and a continuous map $(i,x) \mapsto g_i(x)$ which is uniformly (in view of $i \in I$) locally Lipschitz w.r. to x near $\bar{x}$. Further, write $g \in C^1$ if all $Dg_i(x)$ w.r. to x exist and are continuous on $I \times X$. Then, due to (10.7.10),

Thm. 10.7.3 and the related notes remain true without changing the proof. The same holds for the next statements of this subsection, in particular for Thm. 10.7.6. ◇

10.7.2 The relative slack for solving C^1 inequality systems

In the C^1 case, the above calmness condition for Σ (10.7.3) becomes stronger after adding $\varepsilon\|x-\bar{x}\|^2$ to all g_i: Indeed, the set of all $x \in \Omega$ with $g^m(x) + \varepsilon\|x-\bar{x}\|^2 > 0$ is not smaller than before and the relative slack s_i is now smaller. Hence, the original system is calm whenever so is the perturbed one.

In order to *solve* the inequality system $\Sigma(0)$ of (10.7.3), recall that the minimizing sequence of Corollary 10.7.1 can be obtained by the successive assignment $x \mapsto x' = x+tu$, cf. (10.5.8). Finding u may be a hard task in general. However, if $g \in C^1$, we may replace (10.7.6) by condition (10.7.14) and put $t = \lambda g^m(x)$. This yields both an algorithm that finds some $x(\pi) \in \Sigma(0)$ and a calmness criterion as well.

ALG2: Given $x_k \in X$ and $\lambda_k > 0$, solve (if $g^m(x_k) > 0$) the system

$$Dg_i(x_k)u \le \frac{s_i(x_k)}{\lambda_k} - \lambda_k \ \forall i, \quad \|u\| = 1. \tag{10.7.15}$$

Having a solution u, put	$x_{k+1} = x_k + \lambda_k g^m(x_k)u,$	$\lambda_{k+1} = \lambda_k,$
otherwise put	$x_{k+1} = x_k,$	$\lambda_{k+1} = \frac{1}{2}\lambda_k.$

Corollary 10.7.5 (ALG2). *Let $g \in C^1$. Then Σ is calm at $(0,\bar{x}) \Leftrightarrow$ There is some $\alpha > 0$ such that, for $\|x_1 - \bar{x}\|$ small enough and $\lambda_1 = 1$, it follows $\lambda_k \ge \alpha \ \forall k$. In this case, the sequence x_k converges to some $x(\pi) \in \Sigma(0)$, and it holds (for proper steps $x_{k+1} \ne x_k$)*

$$g^m(x_{k+1}) \le (1-\beta^2)g^m(x_k) \text{ whenever } 0 < \beta < \alpha \text{ and } g^m(x_k) > 0. \quad \diamond \tag{10.7.16}$$

Proof. The first statements follow from Corollary 10.7.1 and Thm. 10.7.3. The estimate is ensured by formula (10.7.13) and $t = \lambda g^m(x)$. □

We used condition $\|u\| = 1$ in (10.7.15) to obtain the simple estimates (10.7.16). If one requires $\|u\| \le 1$ instead (*in order to define a more convenient convex auxiliary system*), then Corollary 10.7.5 is still true, only formula (10.7.16) becomes more complicated.

ALG3: Given $x_k \in X$ and $\lambda_k > 0$, solve (if $g^m(x_k) > 0$) the (convex) system

$$Dg_i(x_k)u \le \frac{s_i(x_k)}{\lambda_k} - \lambda_k \ \forall i, \quad \|u\| \le 1. \tag{10.7.17}$$

Having a solution u, put	$x_{k+1} = x_k + \lambda_k g^m(x_k)u,$	$\lambda_{k+1} = \lambda_k,$
otherwise put	$x_{k+1} = x_k,$	$\lambda_{k+1} = \frac{1}{2}\lambda_k.$

Theorem 10.7.6 (ALG3). *The statements of corollary 10.7.5 remain true with one difference only: for proper steps $x_{k+1} \ne x_k$ it holds*

$$g^m(x_{k+1}) \le (1-\beta^2)g^m(x_k) \text{ whenever } 0 < \beta < \frac{\alpha^2}{C} \text{ and } g^m(x_k) > 0 \tag{10.7.18}$$

with $C = 1 + \sup_i \|Dg_i(\bar{x})\|$. ◇

Proof. We verify the first statement, the estimate then follows from the proof. In view of Corollary 10.7.5, we have only to show that $\lambda_k \ge \alpha > 0$ for ALG3 implies $\inf \lambda_k > 0$ for ALG2. Hence let $\lambda_k \ge \alpha > 0$ hold, with x_1 near $\bar{x}$, for ALG3. We obtain $\|u\| > 0$ from (10.7.17) since there is always some $i = i(k)$ with $s_i(x_k) = 0$. For x_k near $\bar{x}$, we have $\|Dg_{i(k)}(x_k)\| \le C$ and obtain even $\|u\| \ge \lambda_k/C$. Setting now

$$u' = \frac{u}{\|u\|} \text{ and } \lambda_k' = \lambda_k\|u\| \tag{10.7.19}$$

we generate the same points

$$x_{k+1} = x_k + \lambda_k g^m(x_k)u = x_k + \lambda'_k g^m(x_k)u', \tag{10.7.20}$$

and $\lambda_k \geq \alpha$ implies $\lambda'_k = \lambda_k\|u\| \geq \lambda_k^2/C \geq \alpha' := \alpha^2/C$. Finally, it holds for all i, as required in (10.7.15),

$$Dg_i(x_k)u' = \frac{Dg_i(x_k)u}{\|u\|} \leq \frac{s_i(x_k)}{\lambda_k\|u\|} - \frac{\lambda_k}{\|u\|} = \frac{s_i(x_k)}{\lambda'_k} - \frac{\lambda'_k}{\|u\|^2} \leq \frac{s_i(x_k)}{\lambda'_k} - \lambda'_k.$$

This tells us that, up to getting new constants, we may claim $\|u\| \leq 1$ in ALG2. Estimate (10.7.16) implies, due to (10.7.19) and (10.7.20),

$$g^m(x_{k+1}) \leq (1-\beta^2)g^m(x_k) \quad \text{whenever } 0 < \beta < \alpha' \text{ and } g^m(x_k) > 0.$$

This is exactly (10.7.18). □

10.7.3 Calmness and crucial linear inequality systems

We continue in considering the mapping $S = \Sigma$ (10.7.3) in order to clarify that certain inequality systems of the kind $Dg_j(\bar{x})u < 0\ \forall j \in J$ are crucial for calmness, and to indicate the sets J which play the essential role.

Theorem 10.7.7 *[45]. Let $g^m(\bar{x}) = \max_i g_i(\bar{x}) = 0$. Then, S is calm at $(0,\bar{x})$ ⇔*

$$\textit{Each system } Dg_i(\bar{x})u < 0 \quad \forall i \in J \textit{ is solvable,} \tag{10.7.21}$$

whenever J fulfills $J = \{i \mid \lim_{k\to\infty} s_i(x_k) = 0\}$ for certain $x_k \to \bar{x}, g^m(x_k) > 0$. ◇

Comments:

(i) The set J collects the active ($g_i = g^m$) and "almost active" functions g_i for the given sequence of $x_k \notin S(0)$. It holds $J \subset I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$, and $J = \emptyset$ is possible (e.g. if $g(x) \equiv 0$). For $J = \emptyset$, system (10.7.21) is solvable by definition.

(ii) Well-known duality statements yield: (10.7.21) is unsolvable ⇔ $0 \in \operatorname{conv}\{Dg_i(\bar{x}) \mid i \in J\}$ ⇔ $u = 0$ minimizes $\max_{i\in J} Dg_i(\bar{x})u$.

For affine g_i, this holds for derivatives at x_k, too. Because of $g_i(x_k) > 0\ \forall i \in J$ we would obtain $\max_{i\in J} g_i(x_k + u) > 0$, a contradiction for $u = \bar{x} - x_k$. (Of course, this is not the simplest proof of Hoffman's Lemma).

(iii) Solvability of (10.7.21) means that $S^J(p) = \{x \in X \mid g_i(x) \leq p_i\ \forall i \in J\}$ obeys the Aubin property at $(0,\bar{x})$.

(iv) There is a formally similar statement of R. Henrion and J. Outrata [47], Thm. 3. *For $X = \mathbb{R}^n$, S is calm at $(0,\bar{x}) \in \operatorname{gph} S$ if, at $\bar{x}$, the Abadie CQ and (10.7.21) hold true whenever J fulfills $g_i(\xi_k) = 0\ \forall i \in J$ for certain $\xi_k \to \bar{x},\ \xi_k \in \operatorname{bd} S(0) \setminus \{\bar{x}\}$.*

But this condition, derived via co-derivatives, is much too strong: It is not necessary even for calm linear systems ($x \in \mathbb{R}^2,\ x_1 \leq p_1,\ -x_1 \leq p_2$) as noted in [73].

10.8 Modified successive approximation for $-h(x) \in F(x)$

Modified successive approximation is the typical method for analyzing solutions of $p \in h(x) + F(x)$ where h stands for a small variation of a closed mapping $F : X \rightrightarrows P$ (B-spaces). Notice that

$$x \in (h+F)^{-1}(p) \ \Leftrightarrow p \in (h+F)(x) \ \Leftrightarrow p \in h(x) + F(x).$$

The key observation for studying such inclusions consists in the fact that, if $T : X \rightrightarrows X$ obeys the Aubin property with rank $0 < \theta < 1$ at

$$(x_1, x_2) \in \operatorname{gph} T$$

and $d(x_1, x_2)$ is sufficiently small [compared with $(L =)\ \theta$ and ε, δ in Def. 10.1.1], there are elements x_{k+1} $(k > 1)$ such that

$$x_{k+1} \in T(x_k) \text{ and } d(x_{k+1}, x_k) \ \leq \ \theta \ d(x_k, x_{k-1}).$$

The sequence then behaves as in Banach's fixed-point theorem:

$$\begin{array}{l} d(x_{n+1}, x_1) \ \leq \ d(x_{n+1}, x^n) + \ \ldots \ + d(x_2, x_1) \\ \leq (\theta^{n-1} + \ \ldots + \theta^0)\ d(x_2, x_1) \leq \frac{1}{1-\theta}\ d(x_2, x_1). \end{array}$$

Hence the (Cauchy-) sequence converges and fulfills $x_k \to \hat{x} \in T(\hat{x})$ due to closeness. Our condition (10.8.1) below is (7.5.5) in section 7.5.

Theorem 10.8.1 *Let $S = F^{-1}$ obey the Aubin property with (any) rank L at $\bar{z} = (\bar{p}, \bar{x})$ and let $h : X \to P$ be a function with*

$$\|h(x') - h(x)\| \leq \alpha\, d(x', x) \text{ for all } x', x \text{ near } \bar{x} \ \text{ and } \ \|h(\bar{x})\| \leq \beta. \tag{10.8.1}$$

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

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

Proof. Outline of the proof: It holds $\pi \in (h+F)(x) \Leftrightarrow \pi - h(x) \in F(x) \Leftrightarrow x \in S(\pi - h(x))$. The mapping $T_\pi(x) = S(\pi - h(x))$ has Lipschitz rank $\theta := L\alpha$ which is arbitrarily small for small α. If also $\|(\pi - h(\bar{x})) - \bar{p}\|$ is small (here, small β is hidden), one finds a point $x_1 \in S(\pi - h(\bar{x}))$ (close to $\bar{x}$) which may be used as initial point of the above procedure. The detailed estimates are pointed out in [73]. □

Theorem 10.8.2 *(Lyusternik/Graves Theorem) Let $g : X \to P$ (B-spaces) be a C^1 function such that $Dg(\bar{x})X = P$. Then g^{-1} obeys the Aubin property at $(g(\bar{x}), \bar{x})$.* $\diamond$

Proof. Since $Dg(\bar{x})^{-1} : P \to X$ is pseudo-Lipschitz by Banach's inverse mapping theorem [after factorization of X w.r. to $\ker Dg(\bar{x})$], we may put

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

Then the suppositions of Thm. 10.8.1 hold for sufficiently small positive α and β (cf. Remark 7.5.7). Due to $\pi = h(x) + F(x) \Leftrightarrow \pi = g(x)$, this proves local solvability of the latter equation with related Lipschitz estimates which is the assertion. □

Other algorithmic approaches: For more special maps, methods like (10.5.1) have been already studied in [22, 87, 70] (generalized Newton methods and successive approximation). A similar approach and its relations to proximal point methods can be found in [70], too. To verify calmness for certain intersection maps, an algorithmic approach based on Newton's method for semismooth functions has been used in [46]. To characterize the Aubin property for intersections by MFCQ-like conditions in B-spaces, an algorithmic approach has been applied in [88], too.

Chapter 11

Ekelands Principle, Stability and gen. Derivatives

In various papers dealing with non-smooth and multivalued analysis (also called variational analysis), conditions of stability or optimality are given in terms of generalized derivatives. The latter are sets of certain limits and have a more complicated structure than Fréchet-derivatives since, at least, double limits are involved. As already mentioned, this has restrictive consequences for computing them and for all calculus rules. On the other hand, these derivatives indicate those local properties which are essential for the stability in question, and they show the differences which occur in comparison with the known case of C^1 functions. The most of the next statements can be found in several (quite distributed) papers. Our approach, based on Thm. 11.1.4, is self-contained, straightforward and establishes the bridge to the above-mentioned conditions. For other approaches under weaker assumptions and for more general stability properties we refer to [52] and [90].

11.1 Ekeland's principle and Aubin property

Basically, the mentioned bridge is Ekeland's principle which deals with ε-optimal points x_ε of a function, $f(x_\varepsilon) \le \varepsilon + \inf_X f(x)$. It tells us that, given $\delta > 0$, there is a second ε-optimal point z which minimizes the function $x \mapsto f(x) + \delta\ d(x,z)$. In other words, $g(x) := f(z) - \delta\ d(x,z)$ is majorized by f, $g \le f$ (compare with subgradients where g is affine). The distance $d(z, x_\varepsilon)$ can be estimated. Hence, even if $\min f$ does not exist, there exist minimizers after certain arbitrary small variations of f by the distance function.

Theorem 11.1.1 *[26] Let X be a complete metric space, $f : X \to \mathbb{R} \cup \{\infty\}$ be lower semi-continuous and $v := \inf_X f \in \mathbb{R}$. Then, given $\hat{x}$ with $f(\hat{x}) \le v + \varepsilon$ and $\alpha > 0$, there exists some $z \in X$ satisfying*

$$f(x) + \frac{\varepsilon}{\alpha}\ d(x,z) \ge f(z)\ \forall x \in X, \quad f(z) \le f(\hat{x}) \quad \textit{and} \quad d(z,\hat{x}) \le \alpha. \qquad \diamond$$

Proof. Put $\delta = \frac{\varepsilon}{\alpha}$ and $h(z) = \inf_{\xi \in X}\ [f(\xi) + \delta\ d(\xi, z)]$. For arbitrary ξ, z and z' in X, we observe

$$f(\xi) + \delta\ d(\xi,z) \le f(\xi) + \delta\ [d(\xi,z') + d(z',z)].$$

Taking the infimum over all $\xi \in X$ on both sides, we obtain $h(z) \le h(z') + \delta\ d(z',z)$. Therefore (by symmetry), h is a Lipschitz function. To construct a sequence z_k let $z_0 = \hat{x}$.

If $h(z_0) \geq f(z_0)$ then $z = z_0$ fulfills all assertions. Thus, beginning with $k = 0$, assume that $h(z_k) < f(z_k)$. Then one finds some z_{k+1} such that

$$f(z_{k+1}) + \delta\ d(z_{k+1}, z_k)\ <\ f(z_k) \tag{11.1.1}$$

and, in addition,

$$f(z_{k+1}) + \delta\ d(z_{k+1}, z_k)\ <\ h(z_k) + 2^{-k}. \tag{11.1.2}$$

From (11.1.1) and $f(z_0) \leq v + \varepsilon$, we obtain for each k,

$$\delta \sum_{s\leq k} d(z_{s+1}, z_s) < \sum_{s\leq k}(f(z_s) - f(z_{s+1})) = f(z_0) - f(z_{k+1}) \leq \varepsilon.$$

This yields particularly

$$d(z_{k+1}, z_0) \leq \sum_{s\leq k} d(z_{s+1}, z_s) \leq \frac{\alpha}{\varepsilon}\ [f(z_0) - f(z_{k+1}] \leq \alpha. \tag{11.1.3}$$

By (11.1.1) and (11.1.3), $z = z_k$ is again the point in question *whenever* $f(z_k) \leq h(z_k)$. Otherwise (11.1.3) shows that the Cauchy sequence z_k has a limit z^* in the complete space X. Since f is l.s.c., it holds $f(z^*) \leq \liminf f(z_k)$.

By (11.1.1), the sequence of $f(z_k)$ is decreasing, hence $f(z^*) \leq \liminf f(z_k) \leq f(z_0)$, and (11.1.3) entails $d(z^*, z_0) \leq \alpha$.

Finally, since h is continuous, we infer due to (11.1.2), the key relation

$$f(z^*) \leq \liminf\ [f(z_{k+1}) + \delta\ d(z_{k+1}, z_k)] \leq \limsup\ [h(z_k) + 2^{-k}] \leq h(z^*).$$

The latter proves the theorem. □

Corollary 11.1.2 *If, in Thm. 11.1.1, X is a Banach space and $f \in C^1(X, \mathbb{R})$ then $\|Df(z)\| \leq \delta$.* ◇

Proof. Otherwise, choose $u \in B_X$ with $Df(z)u < -\delta$ and apply $f(z + tu) = f(z) + tDf(z)u + o(t)$. □

For many applications, one may put $\delta = \alpha = \sqrt{\varepsilon}$, but not for the next one.

11.1.1 Application to the Aubin property

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

$$\begin{array}{l} \exists\ (p_k, x_k) \to \bar{z} \text{ and } \pi_k \to \bar{p} \text{ such that } (p_k, x_k) \in \operatorname{gph} S \\ \text{and } \operatorname{dist}(x_k, S(\pi_k)) > L\|\pi_k - p_k\| > 0\ (\forall k > 0, k \to \infty). \end{array} \tag{11.1.4}$$

The inequality in (11.1.4) allows $S(\pi_k) = \emptyset$ and involves interesting particular cases.
Case 1: For $(p_k, x_k) \equiv (\bar{p}, \bar{x})$, (11.1.4) is the negation of S being lower Lipsch. with rank L.
Case 2: For $\pi_k \equiv \bar{p}$, (11.1.4) is the negation of S being calm with rank L.

Below, we shall need the function: $\phi_F(x) = \operatorname{dist}(\pi, F(x))$ for given $\pi \in P$.

Definition 11.1.1 We call S strongly closed if ϕ_F is l.s.c. and some $p \in F(x)$ realizes $\operatorname{dist}(\pi, F(x))$ whenever $F(x) \neq \emptyset$ and $\pi \in P$.

Being strongly closed:
Since $\operatorname{gph} F$ is closed, also $F(x)$ is closed. Hence S is strongly closed if F is locally compact [i.e., $\exists \varepsilon > 0 : \cup_{d(x,\bar{x}<\varepsilon} F(x) \subset C$ holds for a compact set C] or $\dim P < \infty$. By remark 3.1.3, S is strongly closed if P is a Hilbert space and all $F(x)$ are closed and convex. For multifunctions in general B-spaces P, our requirement is strong. But for continuous functions $F : X \to P$ the map S is trivially strongly closed.

For strongly closed S, all (p_k, x_k) in (11.1.4) can be replaced by "better" pairs (p_E^k, x_E^k). This was a basic tool in [5] and in various other papers dealing with the Aubin property. With the additional properties of (p_E^k, x_E^k), one obtains stability conditions in terms of generalized derivatives. The next Lemma was used in [70, 88] and is our key for deriving all subsequent conditions in an intrinsic manner. The proof is added since it demonstrates a typical application of Ekeland's principle.

Lemma 11.1.3 *Let S be strongly closed.*
(i) Condition (11.1.4) implies, with $\lambda = L$ and new points $(p_k, x_k) = (p_E^k, x_E^k)$:

$$\begin{array}{l} \exists\ (p_k, x_k) \to \bar{z} \text{ in } \operatorname{gph} S \text{ and } \pi_k \to \bar{p} \text{ such that } p_k \neq \pi_k\ (\forall k > 0) \\ \text{and } (p_k, x_k) \text{ minimizes } H_\lambda(p, x) := \|p - \pi_k\| + \frac{1}{\lambda} d(x, x_k) \text{ on } \operatorname{gph} S. \end{array} \tag{11.1.5}$$

(ii) Condition (11.1.5) implies (11.1.4) for each $L \in (0, \lambda)$.
(iii) In the same way, only with $\pi_k \equiv \bar{p}\ \forall k > 0$, calmness can be characterized.
(iv) For all $t \in (0, 1]$ and $p(t) = p_E^k + t(\pi_k - p_E^k)$, it holds $\operatorname{dist}(x_E^k, S(p(t)) \geq \lambda \|p(t) - p_E^k\|$.
◇

Notes: By (iv), even parameter changes on a fixed line show that every lower Lipschitz rank L of S at (p_E^k, x_E^k) fulfills $L \geq \lambda$. With $v^k = p_k - \pi_k\ (\neq 0)$, the minimum condition in (11.1.5) means

$$\|v^k + \eta\| - \|v^k\| \geq -\frac{1}{\lambda}\|\xi\| \quad \text{if } (p_k + \eta, x_k + \xi) \in \operatorname{gph} S. \tag{11.1.6}$$

Proof. of the Lemma:
(i) Let (11.1.4) be true. The current part of the proof is the same for $\pi_k \equiv \bar{p}$ and $\pi_k \to \bar{p}$, respectively. It remains also valid for $(p_k, x_k) \equiv (\bar{p}, \bar{x})$ in (11.1.4), which corresponds to the negation of being lower Lipschitz with rank L.
For fixed k, define the l.s.c. function

$$\phi(x) = \operatorname{dist}(\pi_k, F(x)), \quad put \quad \varepsilon_k = \|\pi_k - p_k\|$$

and note that $p_k \in F(x_k)$ yields

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

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

$$\phi(x) + \frac{1}{L} d(x, x_E^k) \geq \phi(x_E^k)\ \forall x \in X, \quad d(x_E^k, x_k) \leq \alpha_k \ \text{ and } \ \phi(x_E^k) \leq \phi(x_k). \tag{11.1.7}$$

Explicitly, the first inequality says

$$\operatorname{dist}(\pi_k, F(x)) + \frac{1}{L} d(x, x_E^k) \geq \operatorname{dist}(\pi_k, F(x_E^k))\ \forall x \in X. \tag{11.1.8}$$

Since $F(x_E^k) \neq \emptyset$ and S is strongly closed, some $p_E^k \in F(x_E^k)$ fulfills $\operatorname{dist}(\pi_k, F(x_E^k)) = \|\pi_k - p_E^k\|$. Obviously, $(p_E^k, x_E^k) \to \bar{z}$ as $k \to \infty$.
If $p_E^k = \pi_k$ then a contradiction follows from $x_E^k \in S(p_E^k)$ and (11.1.4):

$$\alpha_k \geq d(x_k, x_E^k) \geq \operatorname{dist}(x_k, S(p_E^k)) = \operatorname{dist}(x_k, S(\pi_k)) \ > \ L\|\pi_k - p_k\| = L\varepsilon_k = \alpha_k.$$

Hence $p_E^k \neq \pi_k$.
So (11.1.8) yields (11.1.5 with the new points $(p_k, x_k) = (p_E^k, x_E^k)$.
(ii) Conversely, (11.1.5) implies with $p = \pi_k$,

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

hence dist $(x_k, S(\pi_k)) \geq \lambda \|p_k - \pi_k\| > 0$. So (11.1.5) yields (11.1.4) for each $L < \lambda$.

(iv) For $x \in S(p(t))$, it follows $\|p(t) - \pi_k\| + \frac{1}{\lambda} d(x, x_E^k) \geq \|p_E^k - \pi_k\|$. Due to the choice of $p(t)$ on the segment $[p_E^k, \pi_k]$, this is $\frac{1}{\lambda} d(x, x_E^k) \geq \|p_E^k - \pi_k\| - \|p(t) - \pi_k\| = \|p(t) - p_E^k\|$. □

Condition (11.1.5) permits characterizations of calmness and the Aubin property by the simpler lower Lipschitz property. It also permits a formulation of Lemma 11.1.3 in terms of (10.1.2), where now (p, x) takes the place of (p_E^k, x_E^k).

Theorem 11.1.4 *For strongly closed S, the following statements are equivalent:*
(i) S is pseudo-Lipschitz at $\bar{z} = (\bar{p}, \bar{x})$
(ii) $\exists\ L, \varepsilon > 0$ such that

$$\begin{array}{l} \text{for all } x \in S_\varepsilon(p) \text{ and } \pi, p \in \bar{p} + \varepsilon B, \quad \pi \neq p, \text{ it holds} \\ \exists p' \in P: \quad L\|p' - \pi\| + \text{dist}\,(x, S(p')) < L\|p - \pi\|. \end{array} \tag{11.1.9}$$

(iii) S is lower Lipschitz at all $(p, x) \in \text{gph}\, S$ near $\bar{z}$ with uniform rank λ.
In addition, S is calm at $\bar{z}$ $\Leftrightarrow$ (11.1.9) holds for $\pi \equiv \bar{p}$. ◇

Proof. (i) $\Leftrightarrow$ (ii) Clearly, S is pseudo-Lipschitz iff (11.1.5) cannot hold for some (large) λ. The latter is (by formal negation) equivalent to condition (11.1.9). With respect to calmness, the same remains true after setting $\pi = \bar{p}$.
(iii) $\Rightarrow$ (i) We prove that (ii) is valid if (iii) holds with some $\lambda < L$. Indeed, setting $p' = p + t(\pi - p)$, there exists, for small $t > 0$, some $x' \in S(p')$ satisfying $d(x', x) \leq \lambda \|p' - p\| = \lambda t(\pi - p)$. So we obtain

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

(i) $\Rightarrow$ (iii) This is evident, so nothing remains to prove. □

Differently from Remark 10.2.2 (v), now the size of δ, included in the lower Lipschitz definition, may depend on (p, x), and the original definitions 10.1.1 claim (11.1.9) stronger with $p' = \pi$ and $p' = \bar{p}$, respectively (up to small changes of the rank L).

11.1.2 Weakly stationary points

The Ekeland principle can be canonically used to define stationary and weakly stationary points for non-differentiable functions. Let $f : X \to \mathbb{R} \cup \{\infty\}$.

Definition 11.1.2 We say that $z \in X$ is a (local) $\varepsilon-$ Ekeland point of f if $f(z)$ is finite and $f(x) + \varepsilon\ d(x, z) \geq f(z) \quad \forall x \in X \quad (\forall x \text{ near } z)$. ◇

We call x *stationary* if $\exists \varepsilon_k \downarrow 0$ such that x is a local ε_k-Ekeland point of f. For $f \in C^1$, this means $Df(x) = 0$.
We call x *weakly stationary* if $\exists \varepsilon_k \downarrow 0, x_k \to x:\ x_k$ is a local ε_k-Ekeland point of f.
Example: Once more the origin for $f = \min\{x, -x^2\}$.

Theorem 11.1.5 *Let* $\dim X + \dim P < \infty$ *and* $F \in C^{0,1}(X,P)$. *Then* $S = F^{-1}$ *is pseudo Lipschitz at* $(\bar p, \bar x) \Leftrightarrow$ *there is no* $p^* \neq 0$ *such that* $\bar x$ *is weakly stationary for* $x \mapsto f(x) = \langle p^*, F(x)\rangle$. $\diamond$

Note: After a regular linear transformation in P, it simply holds $f(x) = F_1(x)$.

Proof. ($\Leftarrow$) (by contradiction) Under the assumption that S is not pseudo Lipschitz at $(\bar p, \bar x)$, we first verify the subsequent formula (11.1.11). For later use we permit, in this part, that F is a closed multifunction: We already know that, with every fixed λ and $v^k = p_k - \pi_k$, condition (11.1.6) holds true. With normalized $p_k^* = \frac{v^k}{\|v^k\|}$ and Euclidean norms, (11.1.6) implies

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

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

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

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

$$\langle p_k^*,\ \eta\rangle + \frac{\alpha_k}{\lambda}\|\eta\| \geq -\frac{1}{\lambda}\|\xi\| \ \text{ if } \|(\xi,\eta)\| < \delta_k \text{ and } (x_k + \xi, p_k + \eta) \in \operatorname{gph} F. \qquad (11.1.10)$$

Since S is not pseudo-Lipschitz, (11.1.10) hold for $\lambda = \lambda_\nu \to \infty$. We select, to $\nu > 1$, sufficiently large index $k = k(\nu) > k(\nu-1)$ and the related points in (11.1.10). Then $p^*_{k(\nu)} \to p^* \neq 0$ may be assumed (otherwise pass to a related subsequence), and (11.1.10) tells us that, with vanishing $\varepsilon_\nu = \frac{1}{\lambda_\nu}$, $\beta_\nu = \|p^*_{k(\nu)} - p^*\| + \alpha_{k(\nu)}/\lambda_\nu$ and $\delta'_\nu = \delta_{k(\nu)}$,

$$\langle p^*,\ \eta\rangle + \ \beta_\nu\|\eta\| \geq -\varepsilon_\nu\|\xi\| \ \text{ if } \|(\xi,\eta)\| < \delta'_\nu \text{ and } (x_{k(\nu)} + \xi, p_{k(\nu)} + \eta) \in \operatorname{gph} F. \quad (11.1.11)$$

Now, since F is loc. Lipschitz with some rank L_F near $\bar x$, we also have

$$p_{k(\nu)} = F(x_{k(\nu)}), \quad p_{k(\nu)} + \eta = F(x_{k(\nu)} + \xi) \quad \text{ and } \ \|\eta\| \leq L_F\|\xi\|.$$

Thus each $x_{k(\nu)}$ is a local $(\varepsilon_\nu + \ \beta_\nu L_F)-$ Ekeland point for $\langle p^*, F(x)\rangle$ with $p^* \neq 0$. This proves ($\Leftarrow$).

($\Rightarrow$) If, in contrary, $\bar x$ is weakly stationary for $p^* \neq 0$ and $f(x) = \langle p^*, F(x)\rangle$ then, considering the equation $F(x_k + \xi) = F(x_k) - tp^*$ for small $t > 0$ at the ε_k- Ekeland points x_k of f, it follows that S cannot be pseudo-Lipschitz. $\square$

11.2 Stability in terms of generalized derivatives

As before, we study closed maps $S : P \rightrightarrows X$ (Euclidean spaces) at $\bar z = (\bar p, \bar x) \in \operatorname{gph} S$, put $F = S^{-1}$. There is a basic device for describing the desired Lipschitz properties by generalized derivatives: use limiting negation.

11.2.1 Strongly Lipschitz

By definition, the map S is not strongly Lipschitz iff

$$\begin{array}{l}\exists x_k \in S(p_k),\ \xi_k \in S(\pi_k) \text{ with } x_k, \xi_k \to \bar x \text{ and } p_k, \pi_k \to \bar p \\ \text{such that } x_k \neq \xi_k \text{ and } \ \|\pi_k - p_k\|/\|\xi_k - x_k\| \to 0 \ (k \to \infty).\end{array} \qquad (11.2.1)$$

Writing, in situation (11.2.1), $\xi_k = x_k + t_k u_k$, where $\|u_k\| = 1$ and $t_k > 0$, and selecting a subsequence such that $u_k \to u$, one obtains $\pi_k = p_k + t_k v_k$ with $v_k \to 0$ and

$$\text{some } u \neq 0 \text{ belongs to } TS(\bar{z})(0), \tag{11.2.2}$$

and vice versa. Hence, (11.2.1) and (11.2.2) coincide. In terms of F, the negation of (11.2.2) (i.e., the strong Lipschitz property of S) is just injectivity of $TF(\bar{x}, \bar{y})$:

$$0 \in TF(\bar{x}, \bar{y})(u) \Rightarrow u = 0. \tag{11.2.3}$$

11.2.2 Upper Lipschitz

The negation of the upper Lipschitz property is just

$$\begin{array}{l} \exists x_k \in S(p_k) \text{ with } x_k \to \bar{x} \text{ and } p_k \to \bar{p} \text{ such that} \\ x_k \neq \bar{x} \text{ and } \ \|p_k - \bar{p}\|/\|x_k - \bar{x}\| \to 0 \ (k \to \infty). \end{array} \tag{11.2.4}$$

Writing $u_k = \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}$, and selecting a subsequence such that $u_k \to u$, one sees that

$$\text{some } u \neq 0 \text{ belongs to } CS(\bar{z})(0), \tag{11.2.5}$$

and vice versa. Hence, (11.2.4) and (11.2.5) coincide, too. In terms of F, the negation of (11.2.5) (= the upper Lipschitz property of S) requires exactly injectivity of $CF(\bar{x}, \bar{y})$:

$$0 \in CF(\bar{x}, \bar{y})(u) \Rightarrow u = 0. \tag{11.2.6}$$

This interrelation has been mentioned (perhaps first) in [62].

11.2.3 Lower Lipschitz

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

$$CS(\bar{z})(v) \cap LB \neq \emptyset \ \ \forall v \in B \tag{11.2.7}$$

is *necessary* for S to be lower Lipschitz at $\bar{z}$. In terms of F, (11.2.7) means *surjectivity* (with linear rate) of $CF(\bar{x}, \bar{y})$, i.e.,

$$B \subset CF(\bar{x}, \bar{y})(LB). \tag{11.2.8}$$

Though (11.2.8) is not sufficient for being lower Lipschitz [cf. Ex.9, [70], with $F \in C(\mathbb{R}^2, \mathbb{R}^2)$], (11.2.8) plays a crucial role if it holds for all $(x, p) \in \operatorname{gph} F$ *near* $(\bar{x}, \bar{p})$.

11.2.4 Aubin property

We shall apply Lemma 11.1.3 for deriving two equivalent characterizations of the Aubin property of $S = F^{-1}$ at $\bar{z}$, namely:

$$\exists \lambda > 0 : \ B \subset CF(x, p)(\lambda B) \ \ \forall \ (p, x) \in \operatorname{gph} S \text{ near } \bar{z} \tag{11.2.9}$$

$$\text{and} \qquad 0 \in D^* F(\bar{x}, \bar{y})(p^*) \Rightarrow p^* = 0. \tag{11.2.10}$$

Remark 11.2.1 Our assumption $\dim X + \dim P < \infty$ is important in this context. In [70], example BE.2, $F = f_{\leq}$ is given by a Lipschitz function, $f : X = l^2 \to \mathbb{R}$ and both conditions are not necessary (for $\dim X = \infty$, Def. 7.3.2 must be modified). $\diamond$

Necessity:
Condition (11.2.9) is necessary since the Aubin property implies that S is lower Lipschitz for $z \in \operatorname{gph} S$ near $\bar{z}$ with the same rank, cf. (11.2.8).
Next we consider (11.2.10), assume $x^* \in D^*F(\bar{x},\bar{y})(p^*)$, $p^* \neq 0$ and verify:
If $x^* \neq 0$ then the rank L of the Aubin property fulfills $L \geq \|p^*\|\,\|x^*\|^{-1}$.
If $x^* = 0$ then S does not obey the Aubin property.
Let $x^* \in D^*F(\bar{x},\bar{y})(p^*)$ hold with sequences ε_k, δ_k and $(p_k, x_k) \to \bar{z}$ from the definition. Then (7.3.2) ensures due to $|\langle x^*, \xi\rangle| \leq \|x^*\|\,\|\xi\|$,

$$\varepsilon_k\|(\xi,\eta^k)\| + \|x^*\|\,\|\xi\| \geq \|p^*\|\,\|\eta^k\| \quad \text{if } \|(\xi,\eta^k)\| < \delta_k \text{ and } p_k + \eta^k \in F(x_k+\xi). \tag{11.2.11}$$

Specify $\eta^k := -t_k p^*$, $t_k = \|\delta_k\|^2$ $(\to 0)$ and assume that any Lipschitz estimate $\|\xi\| \leq L\|\eta^k\|$ holds for certain solutions $\xi = \xi(k)$ to $\pi_k := p_k + \eta^k \in F(x_k+\xi)$.
Since $\|(\xi,\eta^k)\| < \delta_k$ follows for large k, (11.2.11) may be applied:
If $x^* = 0$ this yields $\varepsilon_k\|(\xi,\eta^k)\| \geq \|p^*\|\,\|\eta^k\|$ which contradicts $p^* \neq 0$.
If $x^* \neq 0$, (11.2.11) yields, with every $\beta > 0$, that $\|\xi\| \geq \frac{\|p^*\|+\beta}{\|x^*\|}\,\|\eta^k\|$ for large k. This verifies $L \geq \|p^*\|\,\|x^*\|^{-1}$. □

Sufficiency:
Condition (11.2.9): As for Thm. 11.1.4, we verify (11.1.9) provided that (11.2.9) holds with $\lambda < L$: Given (p,x), choose some $u \in \lambda B$ such that $\frac{\pi-p}{\|\pi-p\|} \in CF(x,p)(u)$. Then $x' \in S(p')$ holds for certain elements $p' = p + t(\pi-p) + o_1(t)$ and $x' = x + t\|\pi-p\|u + o_2(t)$ where $t = t_k \downarrow 0$ and $o_i(t)/t \to 0$. So we obtain for small $t = t_k$,

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

which finishes the proof. □
The criterion (11.2.9) is known from [5].
Condition (11.2.10): If S is not pseudo-Lipschitz at $\bar{z}$, the proof of Thm. 11.1.5 shows that (11.1.11) holds for related sequences. This tells us by definition that

$$p^* \neq 0 \text{ and } 0 \in D^*F(\bar{x},\bar{y})(p^*).$$

Condition (11.2.10) excludes that such p^* exists, hence it implies the Aubin property. □
The criterion (11.2.10) is known from [104] and also from [79] where the equivalent property of openness with linear rate has been investigated.

11.2.5 Summary

We obtained the following stability conditions (for closed F in finite dimension).

Theorem 11.2.2 *It holds for $\bar{z} \in \operatorname{gph} S$ and $S = F^{-1}$:*
(i) The Aubin property of S is equivalent to each of the conditions (11.2.9), (11.2.10).
(ii) S is strongly Lipschitz iff $\quad 0 \in TF(\bar{x},\bar{y})(u) \Rightarrow u = 0$.
(iii) S is upper Lipschitz iff $\quad 0 \in CF(\bar{x},\bar{y})(u) \Rightarrow u = 0$.
(iv) For S to be lower Lipschitz, condition (11.2.8) is necessary. ◇

By (7.4.4), the (injectivity-) conditions in (ii) and (iii) can be also written as

$$\{0\} = TS(\bar{z})(0) \quad \text{and} \quad \{0\} = CS(\bar{z})(0),$$

respectively (but usually, F is the given mapping). We emphasize once again that these facts were observed in many papers, e.g., [12, 83, 62, 134, 91, 70, 79, 104], and the statements have been modified for more general spaces, e.g., in [5, 108, 52, 80, 70].
Notice however that (ii) and (iii), by our definitions, do not imply local solvability.

11.2.6 Minimizer and Taylor expansion

Let $f : \mathbb{R}^n \to \mathbb{R}$ be loc. Lipschitz and $\hat{x}$ be a local minimizer. Then:
(i) [13] $0 \in \partial^c f(\hat{x})$,
(ii) [104] $0 \in D^*F(\hat{x})(1)$ for $\operatorname{gph} F = \{(x, r) \mid r \geq f(x)\}$

The necessary condition (ii) for a local *minimizer* is stronger than (i).
Example: For $f = -|x|$, the origin $\hat{x} = 0$ satisfies (i), but not (ii). For $f = \min\{x, -x^2\}$ however, both conditions are satisfied though $f(\hat{x} - t) = f(\hat{x}) - t$ for small $t > 0$.

Let $f \in C^{1,1}(\mathbb{R}^n, \mathbb{R})$ (loc. Lip. deriv.). Then, it holds

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

[49] for some $\theta \in (0, 1)$ and $q \in \partial^c Df(x + \theta u)u$ as well as
[84] for some $\theta \in (0, 1)$ and $q \in TDf(x + \theta u)u$,

11.3 Persistence of solvability

11.3.1 The direct fixed point approach

Consider now two closed mapping $F, G : X \rightrightarrows P$ and $S = F^{-1}$ in finite dimension. By Thm. 11.2.2, the Aubin property of S at $\bar{z}$ along with the identity

$$D^*G(\bar{x}, \bar{y}) = D^*F(\bar{x}, \bar{y}) \tag{11.3.1}$$

entails local solvability of

$$p \ \in \ G(x). \tag{11.3.2}$$

We would like to show that upper and strong Lipschitz stability are similarly invariable if $CF = CG$ or $TF = TG$ coincide at $\bar{x}, \bar{y}$. But this is not true.

Example 11.3.1 Let F be the real function $F(x) = \{x\}$, and

$$G(x) = \begin{cases} \{x\} & \text{if } x = 0 \ \ or \ |x| = 1/k, \ k = 1, 2, ... \\ \emptyset & otherwise. \end{cases} \tag{11.3.3}$$

For F and G at $(0, 0)$, the $C-$ and $T-$derivative is just the identity, but $F^{-1} = F$ obeys the related stability property in contrast to $G^{-1} = G$. Co-derivatives are: $D^*F(0, 0)(p^*) = \{p^*\}$, $D^*G(0, 0)(p^*) = \mathbb{R}$. $\diamond$

Hence solvability of (11.3.2) does not only depend on CG or TG at $\bar{x}, \bar{y}$. It needs extra assumptions. In addition, solvability may disappear if these derivatives slightly differ at the reference point.
On the other hand, solvability can be handled by the help of Thm. 10.8.1 and via standard fixed-point techniques. The latter will be investigated now.

Suppose we vary F by a small loc. Lipschitz function h as in Thm. 10.8.1 such that

$$G = h + F$$

with small α and β. Then, we already know that the inclusion (11.3.2), $p \in h + F$, leads us to fixed-points of

$$T_p(x) := S(p - h(x)), \qquad \text{namely}$$

$$p \in h(x) + F(x) \quad \Leftrightarrow \quad x \in T_p(x). \tag{11.3.4}$$

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

$$\begin{array}{rl} T_{p,\,\varepsilon}(x) := (\bar{p} + \varepsilon B) \cap T_p(x) & \subset \ \bar{x} + \lambda \ (\ \|p - \bar{p}\| + \|h(x) - h(\bar{x})\| \ ) \ B \\ & \subset \ \bar{x} + \lambda \ (\ \|p - \bar{p}\| + \ \ \alpha\delta \ ) \ B \end{array}$$

ensures for p near $\bar{p}$ and small α, namely (e.g.) if

$$\|p - \bar{p}\| < \tfrac{1}{2}\lambda^{-1} \min\{\varepsilon, \delta\} \ \text{ and } \ \alpha\lambda\delta < \tfrac{1}{2} \min\{\varepsilon, \delta\}, \tag{11.3.5}$$

that $T_{p,\,\varepsilon}$ maps X_δ into itself.
Hence fixed-point Theorems can be applied to verify solvability. We mention here only those approaches which are closely related to the stability notions of this paper.

(i) Studying the fixed-points of (11.3.4) was the key idea in [124]: Apply Kakutani's Theorem to $T_{p,\,\varepsilon}$ if S is upper Lipschitz stable and has convex images.

(ii) If S is strongly Lipschitz stable, Banach's principle can be directly applied since $T_{p,\,\varepsilon}$ is contractive for $\alpha < \lambda^{-1}$. This was a crucial observation in [125].

(iii) Under the Aubin property of S at $\bar{z}$, solvability of (11.3.2) follows from Thm. 10.8.1

In [70], also perturbations by *multifunctions* h are allowed in view of (iii). In [14], [125], [87] and [70] the underlying spaces were Banach spaces. Recall that Thm. 10.8.1 then remains valid. In some of the mentioned papers, only the case of small C^1 functions h has been taken into account, but the proofs for small loc. Lipschitz functions use basically the same principles.

Persistence of upper Lipschitz stability for more general variations G in (11.3.1), has been shown in [81] (S upper Lipschitz stable, S and G closed and convex-valed, $\bar{x} \in \operatorname{dom} G$). So one may summarize.

Theorem 11.3.1 *The Aubin property of S at $\bar{z}$ and s.L.s. are persistent (at least) with respect to small perturbations h as in Thm. 10.8.1. The same is true, in finite dimension, for upper Lipschitz stability, if S is closed and convex-valued.* ◇

11.3.2 Invariance w.r. to first-order approximations

In Thm. 11.3.1, one can put

$$F(x) = \hat{f}(x) + \mathcal{N}(x) \quad \text{ and } \quad h(x) = \hat{f}_L(x) - \hat{f}(x)$$

where $\hat{f}_L(x) = \hat{f}(\bar{x}) + D\hat{f}(\bar{x})(x - \bar{x})$ is the linearization of a C^1 function $\hat{f}$ at a zero $\bar{x} \in F^{-1}(0)$. Then $p \in G(x) := h(x) + F(x)$ means

$$p \in \hat{f}_L(x) + \mathcal{N}(x) \tag{11.3.6}$$

and h becomes an arbitrarily small Lipschitz function in the sense of Thm. 10.8.1. This is a basic situation Thm. 11.3.1 can be applied to. It ensures

Remark 11.3.2 Strong Lipschitz stability, the Aubin property as well as upper Lipschitz stability (for convex-valued S) is invariant w. r. to replacing $\hat{f}$ by its linearization $\hat{f}_L$. ◇

Moreover, derivative formulas for solution mappings then follow from the fixed-point representation

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

and can be computed, by the chain rules in sect. 7.4, if (and only if) related derivatives for $\mathcal{N}$ are known.
As already mentioned, the validity of the equivalence between

$$p \in \hat{f}(x) + \mathcal{N}(x) \quad \text{and} \quad p \in \hat{f}_L(x) + \mathcal{N}(x)$$

in view of being s.L.s. was first shown by S. Robinson. Concerning the same principle for the other stability notions in Thm. 11.3.1, we refer to [124, 81, 24, 87, 134, 70].

Notice however, that replacing $\hat{f}$ by $\hat{f}_L$ does not work in view of calmness: With $\mathcal{N} = \{0\}$ and $\hat{f} = x^2$, calmness is violated at the origin though it holds true for $\hat{f}_L \equiv 0$. This is a consequence of the possible "discontinuous change" of $S(\bar{p})$ when passing from $\hat{f}$ to $\hat{f}_L$.

Remark 11.3.3 The invariance principle simplifies stability conditions only up to a certain level. If one cannot translate conditions including $\mathcal{N}$ or some "derivative" of $\mathcal{N}$ in original terms, nothing is known about stability of (11.3.6) as well. So the description and possible simplification of $\mathcal{N}$ become important. ◇

11.3.3 Invariance w.r. to second-order approximations

If $\mathcal{N}$ has a concrete structure, defined by some function ψ and related systems of equations and inequalities or their polar cones, a similar invariance principle can be observed.

In many cases, ψ may be replaced by its *quadratic approximation* at the reference point, cf. e.g., [126, 117, 23, 134, 24, 70, 130]. This means for the original problem:

The related stability property is invariant w.r. to quadratic approximation of all involved functions near the reference point.

The main example:
For a VI $\hat{f}(x) \in \mathcal{N}_M(x)$ with $M = \{x \mid g_i(x) \leq 0,\ i = 1, .., m\}$ and $g_i \in C^2, \hat{f} \in C^1$, it holds (under a CQ),

$$\mathcal{N}_M(x) = \{\ x^* \mid \exists y \in \mathbb{R}^m :\ x^* = \sum y_i^+ Dg_i(x),\ g(x) = y^-\ \}.$$

Let $\bar{x}$ solve the VI. The quadratic approximation we are speaking about is

$$\hat{f}(\bar{x}) + D\hat{f}(\bar{x})(x - \bar{x}) \in \mathcal{N}_M^q(x) \tag{11.3.7}$$

where

$$\mathcal{N}_M^q(x) = \{\ x^* \mid \exists y \in \mathbb{R}^m :\ x^* = \sum y_i^+ Dq_i(x),\ q(x) = y^-\ \}$$

with

$$q_i(x) = g_i(\bar{x}) + Dg_i(\bar{x})(x - \bar{x}) + \tfrac{1}{2}(x - \bar{x})^T D^2 g_i(\bar{x})(x - \bar{x}).$$

This principle works for the upper Lipschitz property of stationary points (for C^2 optimization), provided that now $p^T x$ perturbs the objective and M is fixed.

Theorem 11.3.4 *The mapping $S(p) = \{x \mid p + \hat{f}(x) \in \mathcal{N}_M(x)\}$ is loc. upper Lipschitz at $(0, \bar{x})$ iff so is the mapping $S^q(p) = \{x \mid p + \hat{f}(\bar{x}) + D\hat{f}(\bar{x})(x - \bar{x}) \in \mathcal{N}_M^q(x)\}$.* ◇

The theorem remains true for additional equations in M.

However, it fails if $\hat{f}$ or ψ are not sufficiently smooth (even if generalized second derivative are applied). Then also the results in [94, 95, 134, 91, 92] for solutions of optimization problems cannot be applied since proper second derivatives are decisive used.

Unfortunately, this complete quadratic approximation may fail for other stability notions (applied to the same model), even if all involved functions are convex polynomials. A typical example (strongly Lipschitz stationary points) will be presented now.

Chapter 12

Explicit analytical stability conditions

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

$$P(a,b): \quad \min f(x) - \langle a, x\rangle, \ s.t.\ x \in \mathbb{R}^n, \ \ g_i(x) \le b_i \ \ \forall\ i = 1, ..., m. \tag{12.0.1}$$

The following statements remain true (with the same proofs) for variational inequalities where Df is replaced by any function $\hat{f}$ of related dimension and smoothness, cf. the comments after Thm. 3.3.3.
Recall: S is upper Lipschitz at $(0, \bar{x})$ if $\bar{x}$ is isolated in $S(0)$ and S is calm at $(0, \bar{x})$. For

$$p = (a, b)$$

near 0, the sets $S(p) \cap B(\bar{x}, \varepsilon)$ may be multivalued or empty.

12.1 Stability of KKT points

Let $S_{KKT}(a,b) = S_{KKT}(p)$ be the set of KKT points to problem (12.0.1).

(i) The upper Lipschitz property
of S_{KKT} at $(0, (\bar{x}, \bar{y}))$ can be checked by studying the linear system

$$\begin{array}{llll} D^2L_x(\bar{x}, y^{0+})u & + & Dg(\bar{x})^T\,\alpha & = 0, \\ Dg(\bar{x})\ u & - & \beta & = 0, \\ \alpha_i = 0 \ \text{ if } \ g_i(\bar{x}) < 0, & & \beta_i = 0 \quad \text{if } \ \bar{y}_i > 0, & \end{array} \tag{12.1.1}$$

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

$$\alpha_i\beta_i = \ 0, \quad \alpha_i \ \ge\ 0 \ \ge\ \beta_i \quad \text{ if } \ \bar{y}_i = g_i(\bar{x}) = 0. \tag{12.1.2}$$

(ii) The strong Lipschitz stability
of S_{KKT} at $(0, (\bar{x}, \bar{y}))$ can be checked by studying system (12.1.1) where (α, β) has, instead of (12.1.2), to satisfy the weaker condition

$$\alpha_i\beta_i \ge 0 \quad \text{ if } \ \bar{y}_i = g_i(\bar{x}) = 0. \tag{12.1.3}$$

These systems have the trivial solution $(u, \alpha, \beta) = (0, 0, 0)$.

Theorem 12.1.1 *In both cases,*
the related Lipschitz property for S_{KKT} just means (equivalently), that the corresponding systems (12.1.1, 12.1.2) and (12.1.1, 12.1.3), respectively, are only trivially solvable. ◇

For proofs and the history of these statements we refer to [70]. Notice that both properties imply LICQ at $\bar{x}$. The next statement was the message of [23].

Theorem 12.1.2 *The Aubin property of S_{KKT} at $(0,(\bar{x},\bar{y}))$ and the strong Lipschitz stability coincide.* ◇

Proof. Otherwise, the Aubin property holds true and the strong Lipschitz stability (Lipschitzian invertibility) is violated. If strict complementarity holds true, i.e., the situation $\bar{y}_i = g_i(\bar{x}) = 0$ does not appear, the KKT system is locally equivalent to a C^1 equation $F(x,y) = 0, F : R^s \to R^s$. So both stability conditions are equivalent to regularity of the related Jacobian. Hence, suppose that $\bar{y}_i = g_i(\bar{x}) = 0$ holds for some i, say for $i = m$. In this situation, consider the reduced problem without constraint g_m and with the KKT point $(\bar{x}, \bar{y}_1, ..., \bar{y}_{m-1})$. By direct application of the definitions and a topological device one can show (see [87] or [70]) that the same situation (Aubin prop., but not strongly Lipschitz stable) remains valid for the reduced problem. Repeating this reduction as long as strict complementarity is violated, we obtain again the C^1 situation for F. This proves the theorem. □

Thm. 12.1.2 fails to hold for $f \in C^{1,1}$ (even without constraints) [87].

12.2 Stability of stationary points

Now let S denote the map of *stationary points* for (12.0.1), i.e.,

$$S(a,b) = \{x \mid \exists y : \ (x,y) \text{ is a KKT point for } P(a,b) \ \}, \quad p = (a,b)$$

and let $\bar{x} \in S(0)$ be the crucial point. Suppose MFCQ at $\bar{x}$ for $p = 0$. Then, given (p,x) near $(0,\bar{x})$, the polyhedron $Y(p,x)$ of Lagrange multipliers to (p,x) is bounded and

$$Y(p,x) \subset Y(0,\bar{x}) + L_1 \ \|(p,x) - (0,\bar{x})\| B_Y$$

holds with some constant L_1 (cf. [10], [64] or Coroll. 2.9 [70]). If S is upper Lipschitz, then also $\|x - \bar{x}\| \leq L_2\|p\|$ holds for $x \in S_\varepsilon(p) := S(p) \cap (\bar{x} + \varepsilon B)$ and small $\|p\|$. This implies for the KKT points

$$((\bar{x} + \varepsilon B) \times \mathbb{R}^m) \cap S_{KKT}(p) \ \subset \ S_\varepsilon(p) \times Y(p, S_\varepsilon(p)) \ \subset \ S_{KKT}(0) + L\|p\| \ B_{XY} \ \text{ and}$$

Remark 12.2.1 If MFCQ holds at $(\bar{x},0)$ and S is upper Lipschitz at $(0,\bar{x})$ then S_{KKT} is calm at $(0,(\bar{x},\bar{y}))$ for every $\bar{y} \in Y(0,\bar{x})$ (with uniform rank L). ◇

Remark 12.2.2 If LICQ holds at $(\bar{x},0)$ then, for $(p,x) \in \operatorname{gph} S$ near $(0,\bar{x})$, there is exactly one Lagrange multiplier $y(p,x)$, and $y(.)$ is loc. Lipschitz. So S and S_{KKT} are locally "lipeomorph" and Thm. 12.1.1 is valid for both mappings. ◇

Hence let only MFCQ be supposed (without MFCQ, nearly nothing is known for stability under nonlinear constraints).

Theorem 12.2.3 *(upperLip) S is upper Lipschitz at $(0,\bar{x})$ $\Leftrightarrow$ each solution of system (12.1.1), (12.1.2) (for each LM $\bar{y}$ to $\bar{x}$) satisfies $u=0$. If $\bar{x}$ was a local minimizer for $p=0$, then this condition even ensures $S(p)\neq\emptyset$ for small $\|p\|$ (u.L.s.).* ◇

For a proof see Thm. 8.36 [70]. □
The proof uses the fact that MFCQ ensures the formula

$$u\in CS(0,\bar{x})(\alpha,\beta) \;\Leftrightarrow\; (\alpha,\beta)\in \cup_{\bar{y}\in Y(0,\bar{x}),\ v\in\mathbb{R}^m} CF(\bar{x},\bar{y})(u,v).$$

Thus the upper Lipschitz property can be checked by solving a finite number of linear systems, defined by the first and second derivatives of f,g at $\bar{x}$ via (12.1.1), (12.1.2).

Such systems are not known for the Aubin property and for strong stability. Recently, it has been shown that a comparable simple answer does not exist (even for convex, polynomial problems), cf. [71]. Notice that our current definition 10.1.1 of being strongly Lipschitz requires locally uniqueness and Lipschitz behavior of solutions only for parameters p with $S(p)\neq\emptyset$ in contrast to being strongly stable (then $S(p)\neq\emptyset$ for all p near $\bar{p}$ is required, too). Let (without loss of generality) $g(\bar{x})=0$.

Theorem 12.2.4 *(strLip) The mapping S is not strongly Lipschitz at $(0,\bar{x})$ $\Leftrightarrow$*

$$\begin{array}{l}\text{There exist } u\in\mathbb{R}^n\setminus\{0\} \text{ and a Lagrange vector } y \text{ to } (0,\bar{x}) \text{ such that}\\ y_i\ Dg_i(\bar{x})u=0\ \forall i, \text{ and with certain } x_k\to\bar{x} \text{ and } \alpha^k\in\mathbb{R}^m, \text{ one has}\\ \alpha_i^k\ Dg_i(\bar{x})u\geq 0\ \forall i \quad \text{and} \quad \lim_{k\to\infty}\ \sum_i\ \alpha_i^k\ Dg_i(x_k)=-D_x^2L(\bar{x},y)u. \ \diamond\end{array} \tag{12.2.1}$$

Examples demonstrate that the limit condition cannot be replaced by a condition in terms of derivatives (for f,g at $\bar{x}$) up to a fixed order. Next put again $p=(a,b)$ and let $Y(p,x)$ be the set of Lagr. multipliers for p and x.

Theorem 12.2.5 *(AubStat) The Aubin property is violated for S at $(0,\bar{x})$ $\Leftrightarrow$ there is some $(u^*,\alpha^*)\in\mathbb{R}^{n+m}\setminus\{0\}$ and a sequence $(p_k,x_k)\to(0,\bar{x})$ in $\operatorname{gph} S$, such that*

$$\begin{array}{ll} Dg_i(x_k)u^*=0 & \text{if } y_i>0 \text{ for some } y\in Y(p_k,x_k),\\ \alpha_i^*\leq 0 \text{ and } Dg_i(x_k)u^*\leq 0 & \text{if } y_i=g_i(x_k)-b_i^k=0 \text{ for some } y\in Y(p_k,x_k),\\ \alpha_i^*=0 & \text{if } g_i(x_k)-b_i^k<0\end{array} \tag{12.2.2}$$

$$\text{and} \qquad \|D_x^2L(\bar{x},y)u^*+Dg(\bar{x})^T\alpha^*\|<\varepsilon_k\downarrow 0 \quad \forall y\in Y(x_k,p_k). \qquad \diamond$$

A proof and specializations can be found in [70], Thm. 8.42. By choosing an appropriate subsequence, the index sets in (12.2.2) can be fixed. But setting $(p_k,x_k)\equiv(0,\bar{x})$ violates again the equivalence for nonlinear g.

Remark 12.2.6 The stability properties of Thm. 12.2.4 and 12.2.5 are equivalent to injectivity of the so-called strict graphical derivative and the coderivative of S, respectively (at the point in question), see e.g. [134]. Hence verifying injectivity of these generalized derivatives for S (not to speak about computing them) requires to study the same limits.

12.3 Strongly Lipschitz ⇒ local solvability ?

Recently this implication has been studied for mappings $S:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$, based on the condition of Thm. 11.2.2 only, i.e.,

$$\{0\}=TS(0,\bar{z})(0) \quad [\text{ equiv. to } T(S^{-1})(\bar{z},0) \text{ being injective }]. \tag{12.3.1}$$

We give a brief summary and direct arguments for the essentials:

(i) It is supposed (described by being *kernel inverting*) that, near $(0,\bar z)$, S satisfies

$$S(p) = U^{-1}\sigma^{-1}(Vp) + Lp \tag{12.3.2}$$

with loc. Lip. $\sigma : \mathbb{R}^n \to \mathbb{R}^n$, regular linear transformations U, V of the $\mathbb{R}^n$ and linear L.

Theorem 12.3.1 *(inverse mapping theorem [91, 92]) Having (i) then S is s.L.s. at $(0,\bar z)$ $\Leftrightarrow$ (12.3.1) holds true.* ◇

Proof. Since U, V are linear and regular, it follows by the chain rules
(12.3.1) $\Leftrightarrow$ $\{0\} = T(\sigma^{-1})(0, U\bar z)(0)$.
By Thm. 7.5.2, it holds
$\{0\} = T(\sigma^{-1})(0,U\bar z)(0)$ $\Leftrightarrow$ σ^{-1} is s.L.s. at $(0,U\bar z)$.
Finally, being a regular linear transformation of σ^{-1}, S is s.L.s. iff so is σ^{-1}. □

Similarly, one could use C^1- diffeomorphisms U, V of $\mathbb{R}^n$ and a C^1 function L in (12.3.2).

(ii) Property (12.3.2) holds for mappings $S = H$ and $S = H^{-1}$ if H is *max- hypomonotone*, which requires that

$$H + \mu\,\mathrm{id} \qquad (\mathrm{id} = \textit{identity})$$

is maximal-monotone for some $\mu > 0$, locally around the (by μ transformed) reference point, say the origin for sake of simplicity. To see this, one may apply arguments of [118], related to prox- regularization. However, after increasing μ, we may directly assume that $H + \mu$ id is closed and strongly monotone. Setting $\lambda = \mu + \delta$, $\delta > 0$, the (strongly monotone) inclusion

$$\lambda(\xi - x) \in H(x), \quad \text{i.e.} \quad \delta(\xi - x) \in \mu(x - \xi) + H(x)$$

has (by local maximal monotonicity) locally unique solutions $x = x(\xi)$ for small $\|\xi\|$, and the estimate $\lambda\,\|\xi'' - \xi'\| \geq \delta\,\|x(\xi'') - x(\xi')\|$ holds true (cf. 3.5.15 in sect. 3.5.3). Setting $\pi(\xi) = x(\xi)$, so π is locally well-defined and Lipschitz and

$$\pi^{-1}(x) = x + \lambda^{-1}H(x) \quad \text{yields, with} \quad \sigma = \pi,$$
$$H = \lambda\,(\pi^{-1} - id) = \lambda\,\sigma^{-1} - \lambda\,id \tag{12.3.3}$$

as required in (12.3.2). With $\sigma = id - \pi$, also H^{-1} satisfies (12.3.2) since

$$\xi \in \sigma^{-1}(x) \Leftrightarrow \xi - x = \pi(\xi) \Leftrightarrow \lambda(\xi - (\xi - x)) \in H(\xi - x) \Leftrightarrow \lambda x \in H(\xi - x) \Leftrightarrow \xi \in x + H^{-1}(\lambda x).$$

Hence

$$H^{-1} = \sigma^{-1}(\lambda^{-1}id) - \lambda^{-1}id. \tag{12.3.4}$$

(iii) Under MFCQ at $\bar x$ and for $g, h \in C^2$, the map of normals

$$H(x) := \mathcal{N}_M(x) = \{x^* \mid x^* = Dh(x)^T z + Dg(x)^{\mathsf{T}} y^+,\ \ g(x) = y^-\ \} \quad x \in M$$

is max-hypomonotone (basically shown in [118]) at $(\bar x, x^{0*}) \in \operatorname{gph} H$. To see this directly, consider $(x, x^*), (x', x'^*)$ in $\operatorname{gph} H$ near $(\bar x, x^{0*})$ and put $\varepsilon = \mu^{-1}$. The products in the monotonicity condition have the form $\langle x - x' + \varepsilon(x^* - x'^*),\ x - x'\rangle$. Using now Lemma 3.5.9, strong monotonicity follows for small ε. The local maximality follows from arguments of closeness and convexity.

Applying (i), (ii) and (iii) to H^{-1}, one obtains that non-singularity (12.3.1) of the stationary point mapping $S(.\,,0)$ (fixed constraints) implies local solvability, too.

However, one does not know whether $S(0,b)$ is locally empty or not and, more important, [91, 92] present no tool for checking (12.3.1) if S is different from the known case of a KKT map in sect. 12.1. If $\bar x$ is a local minimizer, local solvability already follows from the weaker condition of Thm. 12.2.3.

Chapter 13

Future research

(1) All given characterizations of stability do nothing say about the *topological properties* of the solution sets, in particular (and important due to possible characterizations via Kojima functions) if

$$F \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n).$$

Up to now, only the following is known:

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

If F^{-1} obeys the Aubin property at the origin and F has directional derivatives, then $x = 0$ is necessarily an isolated zero of F and $CF(0)$ is injective. This deep statement has been shown by P. Fusek [35]. After deleting directional differentiability, the same statement or counterexamples are unknown.

(2) For stability of B- space problems, the direct approach via algorithms seems to be most appropriate to decrease the gap between stability theory and practical aspects of applications (basically) by refinements of the algorithms for particular classes of problems. The most relevant (and simplest) classes are those where the graph of the mapping is a union of finitely many smooth manifolds. Nevertheless, algorithmic approaches can be used quite general, [88]. For mappings in separable Hilbert spaces, approximate projections can be constructed by the help of dense finite-dimensional subspaces.

(3) Already for $\mathbb{R}^n$ problems, it would be a big step ahead to characterize subclasses of problems which permit more convenient conditions in Thm. 12.2.4. Up to now, this has been done (without requiring $LICQ$) only for problems having linear constraints with at most one quadratic exception [71].

Bibliography

[1] P.S. Alexandroff and H. Hopf. Topologie. Springer, Berlin, 1935.

[2] W. Alt. *Lipschitzian perturbations of infinite optimization problems.* in: A.V. Fiacco (Ed.), Mathematical Programming with Data Perturbations, Dekker, New York, 1983, pp. 7-21.

[3] W. Alt. Numerische Verfahren der konvexen, nichtglatten Optimierung, 2004.

[4] M. Asche. *On the Structure of Nash Equilibrium Sets in Partially Convex Games. J. of Convex Analysis*, Vol. 4, No. 2, 363–372, 1997.

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

[6] S. Banach. *Théorie des opérations linéaires.* Warschau, 1932.

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

[8] M.S. Bazaraa, H.D. Sherali, C.M. Shetty. Nonlinear Programming, Theory and Algorithms. Wiley, New York, 1993.

[9] E.G. Belousov and V.G. Andronov. Solvability and Stability for Problems of Polynomial Programming. Moscow University Publishers, Moscow, 1993 (in Russian).

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

[11] J.V. Burke. *Calmness and exact penalization. SIAM J. Control Optim.*, 29: 493–497, 1991.

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

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

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

[15] W.J. Cook, W.H. Cunningham, W.R. Pulleyblank and A. Schrijver. Combinatorial Optimization, John Wiley and Sons, 1998.

[16] Korte/Vygen. Combinatorial Optimization, Springer, 2003.

[17] A. Schrijver. Combinatorial Optimization, Springer, 2004.

[18] S. Dempe. *Foundations of Bilevel Programming.* Kluwer Academic Publ., Dordrecht-Boston-London, 2002.

[19] V.F. Demyanov, V.N. Malozemov. Introduction to Minimax. Wiley, 1974, New York.

[20] A.V. Dmitruk, A.A.Milyutin and N.P. Osmolovsky. Lyusterniks Theorem and the theory of the extremum. Uspekhy Mat. Nauk 35: 11-46, in Russian, MR 82c: 58010, 1980.

[21] S. Dolecki and S. Rolewicz. *Exact penalties for local minima. SIAM J. Control Optim.*, 17: 596–606, 1979.

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

[23] 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.

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

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

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

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

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

[29] H. Federer. Geometric Measure Theory. Springer, New York, Heidelberg, 1969.

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

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

[32] R. Fletcher. Practical Methods of Optimization, Volume 2, Constrained Optimization. John Wiley, New York, 1981.

[33] H. Frankowska. *An open mapping principle for set-valued maps J. of Math. Analysis and Appl.*, 127: 172–180, 1987.

[34] H. Frankowska. *High order inverse function theorems.* in Analyse Non Lineaire, Attouch, Aubin, Clarke, Ekeland Eds., Gauthier-Villars and C.R.M. Universite de Montreal, 283–304, 1989.

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

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

[37] D. Gale, H.W. Kuhn and A.W. Tucker. Linear programming and the theory of games. Activity analysis of production and allocation. Cowles Commission Monograph 13, Wiley, New York, 1951.

[38] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. Mathematical Programming, 12: 136–138, 1977.

[39] J. Gauvin. Theory of Nonconvex Programming. Les Publications CRM, Montreal, 1994.

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

[41] E.G. Golstein. Theory of Convex Programming. American Mathematical Society, Ser. Transactions of Mathematical Monographs 36, 1972, Providence, RI.

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

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

[44] C. Grossmann, D. Klatte, B. Kummer. Convergence of primal-dual solutions for the nonconvex log-barrier method without LICQ. *Kybernetika*, 40:571–584, 2004.

[45] J. Heerda and B. Kummer. *Characterization of Calmness for Banach space mappings.* Preprint Reihe, Instit. f. Math., HU-Berlin, 2006.

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

[47] R. Henrion and J. Outrata. *Calmness of constraint systems with applications. Mathematical Programming Ser. B*, 104: 437–464, 2005.

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

[49] 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.

[50] A.J. Hoffman. On approximate solutions of systems of linear inequalities. Journal of Research of the National Bureau of Standards. 49: 263–265, 1952.

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

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

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

[54] Jarre and Stoer. Optimierung, Springer, 2004.

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

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

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

[58] S. Kakutani. A generalization of Brouwer's fixed-point theorem. Duke Mathematical Journal, 8: 457–459, 1941.

[59] L.W. Kantorovich and G.P. Akilov. Funktionalanalysis in normierten Räumen. Akademie Verlag, Berlin 1964.

[60] A. Kaplan and R. Tichatschke. Stable Methods for ill-posed Variational Problems. Akademie-Verlag, Berlin, 1994.

[61] J.E. Kelley. The cutting plane method for solving convex programs. *Journ. Soc. Industr. Appl. Math.* 8 (4), 1960.

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

[63] D. Klatte. *On the stability of local and global optimal solutions in parametric problems of nonlinear programming.* Seminarbericht Nr. 80, Sektion Math. HU Berlin, 1985, pp. 1-21.

[64] D. Klatte. *Strong stability of stationary solutions and iterated local minimization.* in Parametric Optimization and Related Topics II, Akademie–Verlag, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička, Berlin, 119–136, 1991.

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

[66] D. Klatte. *On quantitative stability for non-isolated minima.* Control Cybernet. 23 (1994) 183-200, 1994

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

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

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

[70] D. Klatte and B. Kummer. *Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications*. Ser. Nonconvex Optimization and Its Applications, Vol. 60. Kluwer Academic Publ., Dordrecht-Boston-London, 2002.

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

[72] D. Klatte and B. Kummer. *Stability of inclusions: Characterizations via suitable Lipschitz functions and algorithms*. *Optimization*, 55: Issue 5,6: 627-660, Oct. 2006.

[73] D. Klatte and B. Kummer. *Optimization methods and stability of inclusions in Banach spaces*. Mathematical Programming, Series B; Vol. 117, 2009, 305-330.

[74] D. Klatte and B. Kummer. *Newton methods for stationary points: an elementary view of regularity conditions and solution schemes*. *Optimization*, 56: No. 4; 441-462, Aug. 2007.

[75] 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.

[76] R. Kluge, Nichtlineare Variationsungleichungen und Extremalaufgaben. Deutscher Verlag der Wissenschaften. 1979.

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

[78] 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.

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

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

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

[82] B. Kummer. Newton's method for non-differentiable functions. in Advances in Math. Optimization. eds. J. Guddat et al. Akademie Verlag Berlin, Ser. Math. Res. 45: 114–125, 1988.

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

[84] 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.

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

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

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

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

[89] B. Kummer. *How fast is fast fictitious play ?*. Preprint Reihe, Instit. f. Math., HU-Berlin, 2006.

[90] B. Kummer. Inclusions in general spaces: Hoelder stability, Solution schemes and Ekeland's principle *J. Math. Anal. Appl. 358:* 327-344, 2009.

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

[92] A.B. Levy. Lipschitzian Multifunctions and a Lipschitzian Inverse Mapping Theorem. Mathematics of Operations Research, 26: 105–118, 2001.

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

[94] 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*, 215–223. World Scientific Press, Singapore, 1995.

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

[96] D.G. Luenberger. Linear and Nonlinear Programming. 2nd Edition. Addison-Wesley Inc., Reading, Massachusetts, 1984.

[97] D.G. Luenberger. Optimization by Vector Space Methods. John Wiley and Sons, 2009.

[98] Z.-Q. Luo, J.-S. Pang and D. Ralph. Mathematical Programs with Equilibrium Constraints. Cambridge University Press. Cambridge, 1996.

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

[100] K. Malanowski. Second-order conditions and constraint qualifications in stability and sensitivity analysis to solutions to optimization problems in Hilbert spaces. Applied Mathematics and Optimization. 79: 51–79, 1992.

[101] O.L. Mangasarian. Nonlinear Programming. SIAM, Classics in Applied Mathematics, 1994, Philadelphia (republication of the work first published by McGraw-Hill Book Company, New York, 1969).

[102] 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.

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

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

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

[106] 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.

[107] B.S. Mordukhovich. *Variational Analysis and generalized differentiation. Vol. I: Basic Theory, Vol II: Applications. Springer, Berlin*, 2005.

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

[109] J.F. Nash. Equilibrium Point in n-Person Games. Proceeding of the National Academy of Sciences of the USA, 36: 48-49, 1950.

[110] J.F. Nash. Noncooperative Games. Annals of Mathematics, 54: 286-295, 1951.

[111] Y. Nesterov and A. Nemirovskii. Interior–Point Polynomial Algorithms in Convex Programming. SIAM Studies in Applied Mathematics. SIAM, Philadelphia, 1994.

[112] Y. Nesterov. Introductory Lectures on Convex Optimization. Kluwer, 2004.

[113] J. v. Neumann, O. Morgenstern. Theory of Games and Economic Behavior. Princeton, Univ. Press, 1944 (in deutsch u.a. Wuerzburg 1961).

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

[115] R.R. Phelps. Convex Functions, Monotone Operators and Differentiability. 2nd edition, Springer, Lecture Notes in Mathematics, Vol. 1364, New York, 1993.

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

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

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

[119] E. Polyak. Optimization: Algorithms and Consistent Approximations. Springer, Berlin, 1997.

[120] L.S. Pontrjagin, V.G. Boltjanski, R.V. Gamkrelidse, E.F. Mischtschenko. Mathematische Theorie Optimaler Prozesse. Nauka, Moskau, 1969.

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

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

[123] J. Robinson. An iterative method of solving a game. Annals of Mathematics, 54: 296–301, 1951.

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

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

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

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

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

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

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

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

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

[133] J. B. Rosen. *Existence and uniqueness of equilibrium points for concave n-person games. Econometrica*, 33, 520–534, 1965.

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

[135] H. Scarf. The approximation of fixed points of continuous mappings. SIAM J. Appl. Math. 15 (1967), 1328-1343

[136] H. Scarf. The core of a n-person game. J. Econ. 35 (1967), 50-69

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

[138] M. Slater. Lagrange multipliers revisited. Cowles Commission Discussing Paper, Math. 403, 1950.

[139] J. Stoer and C. Witzgall. Convexity and Optimization in Finite Dimensions I. Springer, Berlin, 1970.

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

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

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

[143] E. Zeidler. Vorlesungen über nichtlineare Funktionalanalysis. Teubner Verlag, Leipzig 1976.

Index