Counterexamples in Optimization and non-smooth Analysis

Vorlesung April-July 2011, Bernd Kummer

July, 27, 2011

Contents

1 Help

Um Gegenbeispiele in der Optimierung und der (eng verwandten sogenannten) Nonsmooth Analysis zu verstehen, braucht man sicher zunächst einige konstruktive, wesentliche Grundaussagen; z.B. über Dualität, notwendige Optimalitätsbedingungen, Verhalten von Lösungen bei Parameteränderungen u.s.w.

Andererseits versteht man diese nur halb, wenn man die Notwendigkeit der involvierten Voraussetzungen nicht kennt und die Möglichkeiten und Grenzen vieler Zugänge nicht bewerten kann. Beides geht nur konkret anhand von Beispielen, die oft den Charakter von Gegenbeispielen besitzen.

Zugänge in der "Nonsmooth Analysis" gruppieren sich in der Literatur zumeist um unterschiedliche Typen sogenannter verallgemeinerter Ableitungen für nicht-C^1-Funktionen bzw. für mehrwertige Abbildungen. Diese verallgemeinern sämtlich Fréchet-Ableitungen (oder deren adjungierte Operatoren), sind aber wegen fehlender Glattheit längst nicht so universell einsetzbar. Insbesondere benötigt man für klassische Anwendungen wie Optimalitätsbedingungen, implizite Funktionen und Newton-Verfahren jetzt jeweils unterschiedliche Verallgemeinerungen, und das Verhalten impliziter bzw. inverser Funktionen muss detaillierter als im klassischen Fall gefasst werden.

Daraus resultieren unterschiedliche Typen von Stabilität (etwa einer Gleichung) aber auch unterschiedliche, (nur) für gewisse Fragestellungen mehr oder weniger sinnvolle Definitionen verallgemeinerte Ableitungen. Um zu erkennen, inwieweit diese einsetzbar und berechenbar (also nützlich) sind, braucht es wieder konkrete Beispiele.

Im Folgenden balanzieren wir also zwischen konstruktiven und destruktiven Aussagen herum. Trotzdem hoffe ich, dass eine Linie erkennbar wird, auch wenn noch nicht alle aufgelisteten Literaturstellen (die natürlich wie immer nicht vollständig sind) eingeordnet wurden.

Einige Standardbezeichnungen in Kurzform:

$$r^+ = \max\{0, r\}, \quad r^- = \min\{0, r\} \text{ wenn } r \in \mathbb{R}, \qquad \text{conv}: \text{ konvexe Hülle}$$

$$A + B = \{a + b \mid a \in A,\ b \in B\} \text{ wenn } a + b \text{ erklärt ist}$$

$$f \in C^{0,1}: \ f \text{ ist locally Lipschitz}, \quad f \in C^{1,1}: Df \text{ exist. und ist locally Lipschitz}$$

$$\inf_{x \in \emptyset} f(x) = \infty, \qquad \langle x^*, x\rangle: \text{ Bilinearform}, \qquad \text{dist}(x, M) = \inf_{y \in M} d(x, y)$$

$$Mu = \{Au \mid A \in M\} \text{ wenn } M \text{ eine Menge von Operatoren ist, anwendbar auf } u.$$

$$B(x, r) \text{ abgeschl. Kugel mit Radius } r \text{ um } x.$$

Some statement holds *near* $\bar{x}$ if it holds for all x in some neighborhood of $\bar{x}$.

Lebesgue-measure: Eine *offene* beschränkte Menge $M \subset \mathbb{R}^n$ kann als abzählbare Vereinigung von Würfeln W_k mit Volumen v_k geschrieben werden, die sich nur in Randpunkten schneiden (Verfeinerung eines Gitters, das M überdeckt, und Auswahl derjenigen Würfel, die ganz in M liegen). Man definiert dann $\mu(M) = \sum v_k$. Eine beschränkte Menge $M \subset \mathbb{R}^n$ heisst messbar, wenn es zu jedem $\varepsilon > 0$ offene Mengen G und U gibt mit

$$M \subset G, \quad G \setminus M \subset U \qquad \text{and} \quad \mu(U) < \varepsilon.$$

Man definiert dann $\mu(M) = \inf\ \mu(G)$ bzgl. aller obigen Paare G, U. Ist M unbeschänkt, setzt man $\mu(M) = \lim_{r\to\infty} \mu(\ M \cap B(0, r)\ )$, wenn der Limes existiert.

1.1 The lectures

1. Vorl. 15. 4. Introduction and Examples: 2.1, 2.2, 2.3, 2.9, 2.10

2. Vorl. 29. 4. Duality:
Brauchen u.s.c. und l.s.c. für Abbildung M=M(b), Dualitätslücke mit $\sqrt{x^2 + y^2} - x$

3. Vorl. 6. 5. Duality:
Konstruktive Aussagen für schwach-analytische konvexe Funktionen und Michael's selection Satz, Notwendigkeit konvexer Bildmengen, Belousov Gegenbeispiel zur oberen Halbstetigkeit von M=M(b), Cantor Menge.

4. Vorl. 13. 5. Weiter mit Cantor Menge und -Funktion; begin Blitzfunktion

5. Vorl. 20. 5.
Weiter mit Blitzfunktion und konstruktiven Aussagen zum Clarke - Konzept, einschliesslich Generalized Jacobians und Beziehung zu Clarke's Subdifferential.
Begin Def. metrically und strongly regular für multifunctions $F : X \rightrightarrows Y$ at $(\bar{x}, \bar{y})$. Beispiele: Äquivalenz für $F : \mathbb{R}^n \to \mathbb{R}^n$, $F \in C^1$; Nicht- Äquiv. für Lipsch. functions F, example 7.10.

6. Vorl. 27. 5.
Wiederhol. Def. pseudo-Lipschitz und strongly regular für multifunctions $F : X \rightrightarrows Y$. Dazu:

locally upper Lipschitz, lower Lipschitz, calmness. Spezielle Fälle: F als C^1-Funktion. Example 7.10; $F = \Psi$ mit MFCQ Ex. 8.1 (nicht loc. upper Lipsch)

7. Vorl. 10. 6.
Weiter mit verallg. Ableitungen und Äquivalenz und Nicht- Äquivalenz von ps.Lip. Dazu: Die hinreich. Beding. Propos. 7.2 und 7.7 sowie example 6.2 für $Cf = \emptyset$. Example 7.9 dafür dass Clarke's Bed. im Inv.Satz 7.7 nur hinr. ist

8. Vorl. 17. 6.
Inv. Satz 7.8 zu Tf und Rechnen mit Tf speziell im KKT-System (Produktregel, <simple>). Ableit.-probleme, wenn $f(x) = g(h(x))$, weil i.a. nur $Tf(x)(u) \subset Tg(h(x))(Th(x)(u))$ gilt. Punktw. Surjekt. von CF mit example 7.3. Äquivalenz/Nicht-Äquival. von metrisch bzw. streng regulär für KKT-Systeme (C^2 oder $C^{1,1}$ -problem): Donch/Rockaf. Satz 7.14 und stückweise quadrat. example 7.15., Fusek-Satz 7.11 zu Isoliertheit.

9. Vorl. 24. 6.
Coderivative (Mordukh.) und die hinr. Bedingungen Propos. 7.2, 7.5 für metr. Regular. von F (ps. Lipsch.). Anwendung auf level sets im Hilbert Raum; example 7.6.

10. Vorl. 1. 8.
Partial derivatives and product rule for Tf, Cf and simple/non-simple Lipsch. functions. Stability conditions in original data: KKT via Kojima's funct. for loc. upper Lipsch. and strongly regular. Weak-strong stationary points under MFCQ in C^2-optim.problems.

11. Vorl. 8. 8.
Analysis around Propos. 8.3: $\exists$ two convex, polynomial optimiz. problems with the same first k-th derivat. for all functions at the critical point such that stationary points are strongly/not strongly regular. The same for metric regularity; expl. 8.7, 8.8.
$\exists$ two C^∞-functions with identical derivat. at the critical point and calm/not calm inverse. example 5.2, Dirichlet Funkt. 5.1 and calmness.

12. Vorl. 15. 8.
Newton-example 9.2. The space V of all non-empty, convex, compact sets in $\mathbb{R}^n$, and convex sets $K \neq \emptyset$ with algrelint $K = \emptyset$.

Also nicht alles Aufgeschrieben wurde wirklich geschafft.
Schliesslich sei bemerkt, dass Gegenbeispiele nicht nur die Theorie betreffen, sondern gelegentlich auch Lösungsverfahren ... , cf. [86].

1.2 Some basic definitions

Throughout, X and Y are - at least - metric spaces.

1.2.1 The usual convex subdifferential and argmin

Let X be a (real) Banach space with dual X^* and $f : X \to \mathbb{R} \cup \{\infty\}$.
f is convex if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall x, y \in X, \lambda \in (0,1)$.
The usual subdifferential $\partial f(x)$ of f (convex or not) at x with $f(x) < \infty$ consists of all $x^* \in X^*$ such that

$$f(y) \geq f(x) + \langle x^*, y - x\rangle \quad \forall y \in X \tag{1.1}$$

where $\langle x^*, x\rangle$ stands for $x^*(x)$ and x^* is called *subgradient*. Thus

$$x^* \in \partial f(x) \Leftrightarrow x \in \underset{X}{\operatorname{argmin}}\ f(.) - \langle x^*,\ .\rangle\ . \tag{1.2}$$

Remark 1.1. If $f : \mathbb{R}^n \to \mathbb{R}$ it convex then f is continuous and $\partial f(x) \neq \emptyset\ \forall x$, see e.g. [81]. If $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex, this fails to hold: Put $f = 0$ on $\operatorname{int} B$ (B = Euclidean unit-ball), $f = \infty$ on $\mathbb{R}^n \setminus B$ and $f(x) \in [1, 2]\ \forall x \in \operatorname{bd} B$ (the boundary). Now f is convex, discont. and $\partial f(x) = \emptyset\ \forall x \in \operatorname{bd} B$.

1.2.2 Generalized Jacobian

Definition 1.1. (locally Lipschitz) A function $f : X \to Y$ is called locally Lipschitz (in short $f \in C^{0,1}(X, Y)$) if $\forall x \in X\ \exists \varepsilon > 0 :$ f is Lipschitz on $B(x, \varepsilon) := \{x' \mid d(x', x) \leq \varepsilon\}$. $\diamond$

Functions $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ are almost everywhere differentiable (a poweful Theorem of Rademacher, cf. [17]). This permits to define,

Definition 1.2. Generalized Jacobian $\partial^{gJac} f(x)$ [8, 9]: Let $M_f(x)$ be the set of all limits A of Jacobians $Df(x_k)$ such that $Df(x_k)$ exists and $x_k \to x$. Then $\partial^{gJac} f(x)$ is the convex hull of $M_f(x)$. $\diamond$

Since $\|Df(x_k)\|$ is bounded by a local Lipsch. constant of f, both $M_f(x)$ and $\partial^{gJac} f(x)$ are non-empty and bounded. In addition, the mappings $M_f(.)$ and $\partial^{gJac} f(.)$ are closed.

Proposition 1.2.
If $f : \mathbb{R}^n \to \mathbb{R}$ *is convex then* $f \in C^{0,1}$ *and* $\partial^{gJac} f(x) = \partial f(x)$ *[9].*
For convex $f, g : \mathbb{R}^n \to \mathbb{R}$*, it holds* $\partial(f + g)(x) = \partial f(x) + \partial g(x)$ *[81] (Moreau/Rockafellar).*
But, for $f = |x|,\ g = -f$*:* $\partial^{gJac}(f + g)(x) = \{0\}$, $\partial^{gJac} f(x) + \partial^{gJac} g(x) = [-2, 2]$. $\diamond$

For insiders:
The mapping $M_f(.)$ is sometimes called B-derivative of f. However, B is also associated with the "Bouligand" or contingent derivative (see below) and there is no coincidence:
For the function $\phi = (y^+, y^-)$, $y \in \mathbb{R}$ in section 7.6.2, we have at the origin:
$M_\phi(0) = \{(1, 0), (0, 1)\}$. The contingent derivative consists of the directional derivative of ϕ, $\phi'(0; u) = (u, 0)$ if $u \geq 0$, $\phi'(0; u) = (0, u)$ if $u < 0$. Thus $\phi'(0; u) \neq M_\phi(0)u$.

1.2.3 L.s.c. and u.s.c. multifunctions

Let $F : X \rightrightarrows Y$ be multivalued - i.e., $F(x) \subset Y$.
The set $\operatorname{gph} F = \{(x, y) \mid y \in F(x),\ x \in X\ \}$ is called the graph of F.
The set $\operatorname{dom} F = \{x \mid F(x) \neq \emptyset\ \}$ is called the domain of F.
One says F *is closed* if $\operatorname{gph} F$ is closed in $X \times Y$.
The *inverse multifunction* $F^{-1} : Y \rightrightarrows X$ is defined by $F^{-1}(y) = \{x \in X \mid y \in F(x)\ \}$.
For Banach spaces X, Y, we call F *injective* if $0 \notin F(x) \forall x \neq 0$.
This will be applied to *positively homogeneous mappings* ($y \in F(x) \Rightarrow \lambda y \in F(\lambda x) \forall \lambda \geq 0$) which will play the role of certain derivatives.
If $F(x) = \{f(x)\}$ is single-valued, we identify $F(x)$ and $f(x)$.
If $f : X \to \mathbb{R}$ and $F(x) = \{y \in \mathbb{R} \mid f(x) \leq y\}$ then F^{-1} is the *level-set map* of f.

Definition 1.3. (continuity) At $\bar{x} \in X$, F is called

upper semi-continuous (u.s.c.) if $\forall y(x) \in F(x)$ and $x \to \bar{x}$: $\operatorname{dist}(y(x),\ F(\bar{x})) \to 0$.
lower semi-continuous (l.s.c.) if $\forall y \in F(\bar{x})$ and $x \to \bar{x}$: $\operatorname{dist}(y,\ F(x)) \to 0$.

Hence $F(\bar{x})$ has to be sufficiently big and, in the second case, sufficiently small. The multifunction $F : [0,1] \rightrightarrows \mathbb{R}$ as

$$F(x) = \begin{cases} \{\frac{1}{k}\} & \text{if } x = \frac{1}{k}, \quad k = 1, 2, ... \\ [0, x] & \text{otherwise} \end{cases} \tag{1.3}$$

is everywhere l.s.c., but not u.s.c. at $\bar{x} = \frac{1}{k}$.

2 Basic examples to Continuity and Differentiability

We begin with some classical examples.

2.1 General examples

Example 2.1. [25] A real function, continuous at irrational points and with jumps at rational points.

$$f = \begin{cases} 0 & \text{if } x \quad \textit{irrational} \\ 1/n & \text{if } x = m/n \ \textit{rational} \end{cases} \tag{2.1}$$

where integers $m, n \neq 0$ are prime to each other. $\diamond$

Example 2.2. [25] A C^1- function f such that f' is both positive and negative in intervals $I_\varepsilon = (-\varepsilon, 0)$ and $I^\varepsilon = (\varepsilon, 0)$ near the minimum 0: The function

$$f = \begin{cases} x^4 \ (\ 2 + \sin(1/x) \ ) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{2.2}$$

has the global minimum at 0. The continuous derivative

$$f' = \begin{cases} x^2 \ [\ 4x \ (\ 2 + \sin(1/x) \ ) - \cos(1/x) \ ] & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

attains posit. and neg. values on I_ε and I^ε (at max / min of $\cos(1/x)$). Thus f has stationary points ($f'(x) = 0$) arbitrarily close to 0. $\diamond$

Example 2.3. [25] Polynomials and minima on lines. Let

$$f = (y - x^2)(y - 3x^2). \tag{2.3}$$

Then $f(0, t) = t^2 > 0, \quad f(t, 2t^2) = -t^4 < 0$. Hence the origin is not a local minimum.

Lines: If $y = 0$ then $f = 3x^4 \geq 0$. If $x = 0$ then $f = y^2 \geq 0$. On every other line through the origin $y = \lambda x$ with $\lambda \neq 0$, we obtain

$$f = g(x) := (\lambda x - x^2)(\lambda x - 3x^2) = \lambda^2 x^2 - 4\lambda x^3 + 3x^4, \quad g'(0) = 0, \ g''(0) = 2\lambda^2 > 0.$$

Thus the origin is a (proper) local minimizer for f on each line through the origin. $\diamond$

Example 2.4. [45] *A real convex function, non-differentiable on a dense set.*
Consider all rational arguments $y = \dfrac{p}{q} \in (0, 1]$ such that $p, q \neq 0 \in I\!N$ are prime to each other, and put $h(y) = \frac{1}{q!}$. For fixed q, the sum $S(q)$ over all feasible $h(y)$ fulfills $S(q) \leq \frac{q}{q!}$ and $\sum_q S(q) = c < \infty$. Now define

$$g_1(x) = \begin{cases} 0 & \text{if } x = 0 \\ \sum_{y \leq x} h(y) & \text{if } x \in (0, 1]. \end{cases}$$

Then g_1 is increasing, bounded by c and has jumps of size $(q!)^{-1}$ at $x = y$. We extend g_1 on $\mathbb{R}_+$ by setting

$$g(x) = kg_1(1) + g_1(x-k) \;\text{ if } x \in [k, k+1), \quad k = 1, 2, ...$$

and put $g(x) = -g(-x)$ for $x < 0$. Since g is increasing, the function

$$f(t) = \int_0^t g(x)\, dx \quad \text{as Lebesgue integral}$$

is convex and for $t \downarrow y$ and $t \uparrow y$ (t irrational, y rational) one obtains different limits for the difference quotients of f. Thus f is not differentiable at y. $\diamond$

Linear functions on normed spaces

Example 2.5. Discontinuous lin. function on normed spaces:

$$X: \; x = (x_1, x_2, ...); \quad \text{only finitely many } x_k \neq 0. \tag{2.4}$$

Non-equivalent norms:

$$\|x\|_1 = \sum_k |x_k|, \qquad \|x\|_\infty = \max_k |x_k|.$$

$$L(x) = \sum_k k\; x_k \quad \text{is not bounded on the unit balls for both norms.}$$

$$L(x) = \sum_k \; x_k \quad \text{is bounded on the unit ball for } \|x\|_1 \text{ but not for } \|x\|_\infty. \; \diamond$$

2.2 Cantor's set and Cantor's function

The existence of a subset $C \subset [0,1]$ with the cardinality of $[0,1]$ and L-measure zero has many consequences. Next we follow [25]. Delete from $[0,1]$ the open middle segment of length $\frac{1}{3}$, i.e., $(\frac{1}{3}, \frac{2}{3})$ to obtain the union

$$A_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$$

of two intervals. Next delete the open middle segments of length $\frac{1}{9}$ of these intervals to obtain the union

$$A_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{3}{9}] \cup [\frac{6}{9}, \frac{7}{9}] \cup [\frac{8}{9}, 1].$$

Continuing in the same manner one obtains a sequence of closed sets $A_k \subset A_{k-1}$. The Cantor set is the intersection

$$C = \cap_k \; A_k \; .$$

Let C' be the set of all reals $x \in [0,1]$ which can be written in the triadic system as

$$x \equiv \; 0, c_1 c_2 c_3 \; ...; \quad x = \sum_i 3^{-i} c_i \quad \text{where } c_i \in \{0, 2\}$$

We show that $C' = C$. First notice that $x \equiv 0,1$ and $x' \equiv 0,02222....$ coincide. Given $x < y$ in C', we have (for some n)

$$x \equiv \; 0, c_1 ... c_n \; 0 \; p_{n+2} \; p_{n+3} \; ... \quad \text{and} \qquad y \equiv \; 0, c_1 ... c_n \; 2 \; q_{n+2} \; q_{n+3} \; ...$$

where $y - x$ is minimal for given $n \Leftrightarrow \; p_{n+2} = p_{n+3} ... = 2$ and $q_{n+2} = q_{n+3} ... = 0$. This yields

$x \equiv\ 0, c_1...c_n\ 1$, $\ y \equiv\ 0, c_1...c_n\ 2$. Thus the reals of the open interval (x, y) form just one of the above deleted intervals and are in $[0,1] \setminus C$.
Beginning with $n = 0$, so $[0,1] \setminus C$ contains the union of the 2^n open intervals (assigned to the feasible combinations of $c_1...c_n \in \{0,2\}$)

$$\Omega_{n,c} = \ (\sum_{i=1}^{n} 3^{-i}c_i + 1 * 3^{-(n+1)}\ ,\ \sum_{i=1}^{n} 3^{-i}c_i + 2 * 3^{-(n+1)}\ )\ \text{ with length } 3^{-(n+1)}. \tag{2.5}$$

They are just used in order to define the gaps in A_{k+1}. Hence $C' = C$. Given $n < 3$, these 2^n intervals are

$$n = 0: \quad (0 + \frac{1}{3},\ 0 + \frac{2}{3}),$$

$$n = 1: \quad (\frac{0}{3} + \frac{1}{9},\ \frac{0}{3} + \frac{2}{9}) = (\frac{1}{9}, \frac{2}{9}), \qquad (\frac{2}{3} + \frac{1}{9},\ \frac{2}{3} + \frac{2}{9}) = (\frac{7}{9}, \frac{8}{9})$$

$$n = 2: \quad (\frac{0}{3} + \frac{1}{27},\ \frac{0}{3} + \frac{2}{27}) = (\frac{1}{27}, \frac{2}{27}), \qquad (\frac{2}{9} + \frac{1}{27},\ \frac{2}{9} + \frac{2}{27}) = (\frac{7}{27}, \frac{8}{27})$$

$$(\frac{2}{3} + \frac{1}{27},\ \frac{2}{3} + \frac{2}{27}) = (\frac{19}{27}, \frac{20}{27}), \qquad ([\frac{2}{3} + \frac{2}{9}] + \frac{1}{27},\ [\frac{2}{3} + \frac{2}{9}] + \frac{2}{27}) = (\frac{25}{27}, \frac{26}{27}),$$

and we obtain

$$C = [0,1] \setminus \cup_{n,c}\ \Omega_{n,c}.$$

C is not countable: To each $x \in C$, assign the *dual-number* $y = D(x)$ with the 0-1- digits

$$y \equiv 0, d_1\ d_2\ d_3\ ... \qquad \text{where } d_i = \frac{c_i}{2} \in \{0,1\}.$$

Then D maps C *onto* $[0,1]$ since all $y \in [0,1]$ have a preimage: $x \equiv 0, c_1 c_2 c_3\ ...\ ;\ c_i = 2d_i$. Thus C is not countable, $C \cong [0,1]$ (gleichmächtig). Notice: D is monotone and the pre-image is not unique for

$$y \quad \equiv \quad 0, d_1, ..., d_n 011... \qquad (\equiv \quad 0, d_1, ..., d_n 1\ )$$

since

$$D(x) = y = D(x')$$

holds for

$$x \equiv 0,(2d_1)...(2d_n)0222... \quad \text{and} \quad x' \equiv 0,(2d_1)...(2d_n)200... \quad \text{where } \ x \neq x'.$$

These are the boundary points of the (successively deleted) open intervals $\Omega_{c,n}$ (2.5).

The function D is not locally Lipschitz:
Consider $x_k = 3^{-k} \equiv 0,0...0222... \in C$ (k zeros after 0,). Then $D(x_k) \equiv 0,0...0111...$ (k zeros after 0, and dual) is $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}...$; hence $D(x_k) = \frac{1}{2^k}$ and $\frac{D(x_k)-D(0)}{x_k-0} = \frac{3^k}{2^k} \to \infty$.

The Cantor function: Define

$$g(x) = D(x)\ \ \forall x \in C \ \text{ and } \ g(y) = D(x) = D(x')\ \forall y \in (x, x') \ \text{ some deleted interval.}$$

This function is monotone (increasing) and constant on each set $\Omega_{c,n}$. It maps $[0,1]$ monotonically onto $[0,1]$. Hence g cannot have any jump.

In consequence, both g and D turn out to be continuous.
Moreover, on $\Omega_{c,n}$, the function g has derivative zero. Thus
$g' = 0$ *exists on the open set* $\Omega = \cup_{c,n}\ \Omega_{c,n}$ *with full measure* $\mu(\Omega) = 1$. Indeed,

$$\mu(\Omega) = \sum_{n=0}^{\infty} \mu(\Omega_{n,c}) = \sum_{n=0}^{\infty} 2^n\ 3^{-(n+1)} = \tfrac{1}{2} \sum_{n=0}^{\infty} (\frac{2}{3})^{n+1} = \tfrac{1}{2} \frac{\frac{2}{3}}{1 - \frac{2}{3}} = 1.$$

$$\text{Hence} \qquad \mu(C) = \mu([0,1]) - \mu(\Omega) = 0.$$

So it follows for the Lebesgue-Integral and Stieltjes-Radon-Integral, respectively:

$$\int_0^1 g'(x)dx = 0, \qquad \int_0^1 dg(x) = 1.$$

The Riemann-Integral $G(x) = \int_0^x g(t)dt$ of the continuous, monotone function g exists. Thus G is convex, increasing, and $G'(x) = g(x) \forall x \in (0,1)$. The second derivative $G''(x) = g'(x)$ exists almost everywhere, with value zero, but G is not linear.

Homeomorphism and measure: The function h as $h(x) = x + g(x)$ maps $[0,1]$ continuously onto $[0,2]$ and is strongly increasing. Thus the inverse $h^{-1} : [0,2] \to [0,1]$ exists and is again continuous. In consequence, h is a homeomorphism between $[0,1]$ and $[0,2]$.

The h- image of each interval $\Omega_{c,n} = (x, x')$ is an interval of the same length $x' - x$ (since $g(x) = g(x')$).

Thus $\mu(h(\Omega_{c,n})) = \mu(\Omega_{c,n})$ and $\mu(h(\Omega)) = \mu(\Omega) = 1$. It follows

$$\mu(h(C)) = 2 - \mu(\Omega) = 1 > 0 \quad \text{in spite of } \mu(C) = 0.$$

Let $M \subset h(C)$ be *not measurable* (such M exists for every measurable set with positive measure !). Then (by monotonicity) the set $h^{-1}(M)$ is a subset of C and has - in consequence- measure zero. Conversely, one obtains
The homeomorphism h maps certain sets of measure zero onto non-measurable sets.

2.3 The distance $x \mapsto \operatorname{dist}(y, F(x))$ for closed mappings

Let $F : X \rightrightarrows Y$ be closed and $d_y(x) = \operatorname{dist}(y, F(x))$ for fixed $y \in Y$. For many constructions, the following simple statements are useful.

Proposition 2.6. *If $Y = \mathbb{R}^m$ then*
(i) d_y is l.s.c. and,
(ii) whenever $F(x) \neq \emptyset$, some $y(x) \in F(x)$ realizes the distance. $\diamond$

Proof. Since F is closed so are the image-sets $F(x)$. Hence, if $F(x) \neq \emptyset$, some $y(x) \in F(x) \subset \mathbb{R}^m$ with $d(y, y(x)) = d_y(x)$ exists. Let $x_k \to \bar{x}$ and $d_y(x_k) \to \alpha$ $(< \infty)$. Select $y(x_k) \in F(x_k)$ such that $\|y - y(x_k)\| = \operatorname{dist}(y, F(x_k))$. The bounded sequence $\{y(x_k)\}$ in $\mathbb{R}^m$ has an accumulation point $\bar{y}$, and $\|y - \bar{y}\| = \alpha$. Since $\operatorname{gph} F$ is closed, it follows $\bar{y} \in F(\bar{x})$, and consequently $\operatorname{dist}(y, F(\bar{x})) \le \|y - \bar{y}\| = \alpha$. $\square$

In infinite dimensions, both statements of Propos. 2.6 may fail.

Example 2.7.
(i) Let $X = \mathbb{R}$, $Y = l^2$ and - with the unit-vectors $e^k \in l^2$ -

$$F(x) = \begin{cases} e^k, & \text{if } x = \frac{1}{k}, \quad k = 1, 2, \dots, \\ 2e^1, & \text{if } x = 0, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Then $\operatorname{gph} F$ is closed, $\operatorname{dist}(0, F(\frac{1}{k})) = d(0, e^k) = 1$, while $\operatorname{dist}(0, F(0)) = 2 > 1$.
(ii) Using example 7.6, we have only to change the role of x and y. This means: With $f(x) = \inf_k x_k$, $x \in l^2$, put $G(y) = \{x \in l^2 \mid f(x) \le y\}$. Now, for $\xi \in l^2$ with $\xi_k > 0 \forall k$, the distance $1 = \operatorname{dist}(\xi, G(-1))$ is not attained. $\diamond$

Example 2.8. Even for $X = Y = \mathbb{R}$, the distance d_y is not necessarily continuous. Take

$$F(x) = \begin{cases} \{1/x\} & \text{if } x \neq 0 \\ \{0\} & \text{if } x = 0 \end{cases}$$

as a closed, but neither u.s.c. nor l.s.c. multifunction at the origin. $\diamond$

2.4 Perfectly unstable quadratic one-parametric optimization

If an optimization problem (or any other problem) depends on some parameter p, the solution-behavior as a (multi-)function of p is of interest. Assume the problem should be solved for $p = 0$, but (due to some error) we solve it for p near 0. Evidently, we hope the error vanishes as $p \to 0$ if p is continuously involved. Nevertheless, the reverse situation may occur even for simple convex quadratic problems, i.e.,

Better approximation of involved functions can imply worser approximation of solutions.

Example 2.9. [2], [54]. Let us minimize, with real parameters $p \geq 0$, the convex function

$$g(x,p) = p^2x^2 - 2p(1-p)x \quad \text{under the constraint} \quad x \geq 0. \tag{2.6}$$

The problem has, for $p \in (0,1)$, the unique minimizer $x(p) = \frac{1-p}{p}$, and the extreme-value $\alpha(.)$ satisfies

$$\alpha(p) = -(1-p)^2 > -1. \tag{2.7}$$

For $p = 0$ and $p \geq 1$, $\alpha(p) = 0$ is obvious. In consequence,
1. The error of extreme-values

$$\mid \alpha(p) - \alpha(0) \mid$$

increases as $p \downarrow 0$, and vanishes if the error of arguments $|p - 0|$ is sufficiently large ≥ 1.
Due to (2.7),
2. g is a two-dimensional polynomial of (minimal) degree 4 which, on a polyhedron (the non-negative orthant), is bounded below (infimum $= -1$) without having a minimum. ◇

Example 2.10. [2] In the above example, the solution set at the critical parameter $p = 0$ was unbounded ($\mathbb{R}^+$). Now it is nonempty and compact. Minimize, again with $p \geq 0$,

$$\begin{array}{c} f_p(x,y) = xy - y + \ 2\ g(x,p) \quad \text{under the constraints} \quad 0 \leq y \leq 1,\ x \geq 0 \\ \text{with } g \text{ from (2.6)} \end{array} \tag{2.8}$$

Let $\phi(p)$ be the assigned infimum and $\Psi(p)$ the set of (global) minimizers. We write $z = (x,y)$.
Now it holds:
3. ϕ is not l.s.c., Ψ neither u.s.c. nor l.s.c. at 0, all sets $\Psi(p)$ are nonempty and compact.
Proofs:
Obviously, $\Psi(0) = \{(0,1)\}$ and $\phi(0) = -1$. Since $xy - y \geq -1$ and $\alpha(p) \geq -1$, it follows $\phi(p) \geq -3$.
The sets $\Psi(p)$ $(p > 0)$ are nonempty and compact since $f_p(z) \to \infty$ if $||z|| \to \infty$ and z is feasible.
For $p \downarrow 0$ and $z(p) = (\frac{1-p}{p}, 0)$, it holds $f_p(\ z(p)\ ) \to -2$. Hence

$$\liminf_{p \downarrow 0} \phi(p) \leq -2 < -1 = \phi(0), \tag{2.9}$$

which tells us that ϕ is not l.s.c. at 0. Next assume Ψ to be u.s.c. at 0. Then we have by Def. 1.3 : For any $z^*(p) \in \Psi(p)$ it holds $\operatorname{dist}(z^*(p), \Psi(0)) \to 0$ as $p \to 0$. Thus

$$z^*(p) \to (0,1) \ \text{ if } p \downarrow 0.$$

The same follows, for certain $z^*(p) \in \Psi(p)$, if Ψ is l.s.c. at 0. Continuity of f then implies

$$\phi(p) = f_p(\ z^*(p)\ ) \to f_0(0,1) = \phi(0) = -1 \ \text{ which contradicts (2.9).}$$

So Ψ is neither u.s.c. nor l.s.c. at 0. ◇

3 Duality and RHS-perturbations in convex problems

3.1 Lagrange-Duality for (classical) convex problems and duality gaps

Consider the problem

$$\text{(P)} \qquad \min \{f(x) \mid g_i(x) \le 0 \; \forall i = 1,...,m;\; x \in X = \mathbb{R}^n\} \qquad (f, g_i \text{ convex on } X) \tag{3.1}$$

and the Lagrangian

$$L(x,\lambda) = f(x) + \sum_i \lambda_i \; g_i(x) = f(x) + \langle \lambda \; , \; g(x)\rangle, \qquad \lambda \ge 0.$$

Let (P) have a finite infimum v_P (attained or not), and put

$$\gamma(\lambda) = \inf_{x \in X} L(x,\lambda).$$

The dual problem (D) consists in maximizing γ w.r. to $\lambda \ge 0$. The sup-inf relation

$$v_D := \quad \sup_{\lambda \ge 0} \gamma(\lambda) = \sup_{\lambda \ge 0} \inf_{x \in X} L(x,\lambda) \;\le \inf_{x \in X} \sup_{\lambda \ge 0} L(x,\lambda) \quad = v_P \tag{3.2}$$

is always true (even for arbitrary functions f, g).

Definition 3.1. We say that *weak* duality holds if $v_D = v_P$; and that *strong* duality holds if, in addition, (D) is solvable, i.e., $v_D = \max_{\lambda \ge 0} \gamma(\lambda)$.

Note: ! In some papers, already our weak duality is called (strong) duality and the trivial relation $v_D \le v_P$ is called weak duality !
If $v_D < v_P$, a *duality gap* occurs.

RHS perturbation function: Duality is closely connected with the (monotone) right-hand-side perturbation function

$$\phi(b) = \inf_{x \in M(b)} f(x) \text{ where } M(b) = \{x \mid g(x) \le b\},\; (b \in \mathbb{R}^m) \text{ and } \phi(b) = \infty \text{ if } M(b) = \emptyset. \tag{3.3}$$

For $b > 0$, it holds $M(0) \subset M(b)$ and $\phi(b) \le \phi(0) = v_P$, hence

$$\liminf_{b \to 0} \phi(b) \le \phi(0) \qquad \text{is always true.}$$

Proposition 3.1.
(i) strong duality is equivalent to the existence of a subgradient λ^* for ϕ at the origin (i.e., $\partial\phi(0) \ne \emptyset$), and is ensured if a Slater point x^S exists ($g_i(x^S) < 0 \forall i$).
(ii) weak duality holds true if ϕ is l.s.c. at 0.

We briefly show statement (ii), i.e: If $\liminf_{b \to 0} \phi(b) = \phi(0)$ (i.e. ϕ l.s.c. at 0) then weak duality holds true.

Proof. Since $M(0) \ne \emptyset$, all $x \in M(0)$ are Slater points in $M(b)$, $b > 0$, which yields - by (i) - strong duality for the related problem with Lagrangian $L_b(x,\lambda) = f(x) + \sum_i \lambda_i \; (g_i(x) - b_i)$. In (3.2), we thus obtain $v_P(b) = \phi(b)$ and

$$v_D(b) = \max_{\lambda \ge 0} \gamma_b(\lambda) = \max_{\lambda \ge 0} \inf_{x \in X} (\; L(x,\lambda) - \langle \lambda, b\rangle \;) \;= \inf_{x \in X} \sup_{\lambda \ge 0} L(x,\lambda) - \langle \lambda, b\rangle \quad = v_P(b),$$

and with any dual solutions $\lambda(b) \ge 0$,

$$v_D(b) = \inf_{x \in X} (\; L(x,\lambda(b)) - \langle \lambda(b), b\rangle \;) \;= \phi(b). \tag{3.4}$$

Since $\langle\lambda(b), b\rangle \geq 0$, it follows $\inf_{x\in X} L(x,\lambda(b)) \geq v_D(b) = \phi(b)$ and

$$v_D = \sup_{\lambda\geq 0}\inf_{x\in X} L(x,\lambda) \geq \inf_{x\in X} L(x,\lambda(b)) \geq \phi(b). \tag{3.5}$$

Hence $v_D \geq \liminf_{b\to 0}\phi(b) = \phi(0) = v_P$. □

Remark 3.2. ϕ is l.s.c. at 0 if $M = M(b)$ is u.s.c. at 0 and $M(0) \neq \emptyset$ is compact.

Indeed, take $x(b) \in M(b)$ with $f(x(b)) < \phi(b)+\varepsilon(b)$, $\varepsilon(b) \downarrow 0, b \to 0$. By $\text{dist}(x(b), M(0)) \to 0$ and compactness there is a cluster point $\bar{x} \in M(0)$, thus (since convex $f : \mathbb{R}^n \to \mathbb{R}$ are continuous)

$$\phi(0) \leq f(\bar{x}) \leq \liminf_{b\to 0}\phi(b).$$

Remark 3.3. Considering only $b \in \text{dom}\, M(.)$, it holds: ϕ is u.s.c. at 0 if M is l.s.c. at 0.

Indeed, take $x \in M(0)$ with $f(x) < \phi(0) + \varepsilon$, $\varepsilon > 0$. Since $\text{dist}(x, M(b)) \to 0$ there exist $x(b) \in M(b)$ with $x(b) \to x$. Hence $b \to 0$ yields

$$\phi(b) \leq f(x(b)) \to f(x) < \phi(0) + \varepsilon \quad \text{and} \quad \limsup_{b\to 0}\phi(b) \leq \phi(0) + \varepsilon.$$

Lack of strong duality; Study the real problem $\min\{x \mid x^2 \leq 0\}$ (exercise).
Lack of weak duality is more complicated. It needs $x \in \mathbb{R}^2$ and skillful constraints.

3.2 Convex problems with bad constraints and duality gaps

Now we specify the problem

Example 3.4. RHS perturbations of convex inequalities; systems $M(b) = \{x \mid g(x) \leq b\}$. Put

$$H(\alpha) = \{(x,y) \mid h(x,y) \leq \alpha\}; \quad h = \sqrt{x^2+y^2} - x. \tag{3.6}$$

Then

$$(x,y) \in H(0) \quad \Leftrightarrow \quad y = 0,\ x \geq 0.$$

To construct a duality gap, function h was already used in [27].

Note. Setting $\mu(x,y) = h(x,y) - y$, μ is a so-called NCP function (the zero-set coincides with the non-negative half-axes), cf. [18].

Consider

$$M(\alpha,\beta) = H(\alpha) \cap \{(x,y) \mid\ y = \beta\}. \tag{3.7}$$

Fixed x:
Put $x = t > 0$ fixed. Then $\sqrt{t^2+\beta^2} - t = \sigma(\beta) \downarrow 0$ if $\beta \to 0$. Now choose $\alpha = \sigma(\beta)$. Then

$$(t,\beta) \in M(\alpha,\beta) \ \text{ and } \ (s,\beta) \notin M(\alpha,\beta)\ \forall s < t \text{ since } s \mapsto \sqrt{s^2+\beta^2} - s \text{ is decreasing.}$$

1.) It follows that M is not lower semi-continuous at the origin
i.e., the condition of Def. 1.3

$$\lim_{(\alpha,\beta)\to 0,\ M(\alpha,\beta)\neq\emptyset} \text{dist}((x,y), M(\alpha,\beta)) = 0\ \forall (x,y) \in M(0,0) \tag{3.8}$$

is violated. Indeed, we may put $(x,y) = (0,0)$ and $\alpha = \sigma(\beta)$ (the limit is $\geq t$).

2.) the mapping $\alpha \mapsto H(\alpha)$ is not u.s.c.
To see this, fix $y = \beta \neq 0$. Feasibility now means $\sqrt{x^2+\beta^2} - x \le \alpha$. Due to

$$\sqrt{x^2+\beta^2} - x = \sqrt{x^2+\beta^2} - \sqrt{x^2} = \frac{\beta^2}{\sqrt{x^2+\beta^2} + \sqrt{x^2}} \to 0 \quad (\text{ if } x \to \infty),$$

points $(x, \beta) \in H(\alpha)$ exist for all $\alpha > 0$. This implies, since $\text{dist}(\ (x, \beta), H(0)\ ) \ge |\beta|$, that the mapping $\alpha \mapsto H(\alpha)$ is not u.s.c., cf. Def. 1.3.

3.) Perturbed infima finite, but ϕ is not l.s.c.
For the convex parametric optimization problem

$$\min\ \{y \mid (x,y) \in H(\alpha),\ y \ge -1\} \tag{3.9}$$

and its (finite) extreme-values $\phi = \phi(\alpha) \quad (\alpha \ge 0)$, one obtains:

$$\phi \text{ is not l.s.c. due to} \quad \phi(\alpha) = \begin{cases} 0 & \text{if } \alpha = 0 \\ -1 & \text{if } \alpha > 0. \end{cases} \tag{3.10}$$

4.) Duality gap
The convex problem

$$\min\ \{y \mid h(x,y) \le 0\} \tag{3.11}$$

with the solution set $H(0)$ and optimal-value $v = 0$ has a duality gap. This means, the Lagrangian

$$L(x,y,\lambda) = y + \lambda\ h(x,y), \qquad \lambda \ge 0$$

as well as

$$\gamma(\lambda) := \inf_{(x,y)\in \mathbb{R}^2} L(x,y,\lambda),$$

satisfy $\sup_{\lambda \ge 0} \gamma(\lambda) < v$. In other words, it holds

$$v_D := \sup_{\lambda\ge 0} \inf_{(x,y)\in\mathbb{R}^2} L(x,y,\lambda) \ < \ \inf_{(x,y)\in\mathbb{R}^2} \sup_{\lambda \ge 0} L(x,y,\lambda) = v.$$

Indeed, for any λ and any $y < 0$, we find (big) x such that

$$|\ \lambda\ h(x,y)\ | = |\ \lambda\ (\ \sqrt{x^2+y^2} - x\ )\ | < 1.$$

Hence $L(x,y,\lambda) < y + 1$. With $y \to -\infty$, this yields $\gamma(\lambda) = -\infty$ and $v_D = -\infty$. $\diamond$

Example 3.5. In [2], the properties of example 3.4 have been derived by using a (more complicated ?) convex function h with obvious behavior for fixed y,

$$h = \begin{cases} |y|\ e^{-x/|y|} & \text{if } x \ge 0,\ y \neq 0 \\ 0 & \text{if } x \ge 0,\ y = 0 \\ |y| - x & \text{if } x \le 0. \end{cases} \tag{3.12}$$

$\diamond$

3.3 Nice properties for weakly analytic convex functions f, g_i

We call f *weakly analytic* provided the following is true: If f is constant on a segment $[x, y] \subset \mathbb{R}^n$, $x \neq y$ then it is constant on the whole line which includes the segment.
In particular, this holds for all analytic functions and polynomials on $\mathbb{R}^n$.

Proposition 3.6. *Provided that all convex g_i in (3.1) are weakly analytic, then M in (3.3) is l.s.c. Moreover, ϕ is even continuous, if b is restricted to* $\text{dom}\, M := \{b \mid M(b) \neq \emptyset\}$.

A proof can be found in [2]. However, M is not necessarily u.s.c. There is a computer- generated counterexample with convex polynomials of degree 16, Belousov/Schironin (Moskau) ca 1982. Hence Remark 3.2 cannot be used for showing that ϕ is l.s.c.

Proposition 3.7. *If also f is convex and weakly analytic then there is even a continuous function $\psi : \operatorname{dom}\Psi \to \mathbb{R}^n$ such that $\psi(b) \in \Psi(b)$ (the solution set) for all $b \in \operatorname{dom}\Psi$.*

The proof uses E. **Michael's selection theorem** for l.s.c. multifunctions [64], simplified: *For every l.s.c. multifunction F with non-empty, convex images $F(x)$ there is a continuous function f satisfying $f(x) \in F(x) \forall x$.*

Here, convexity is essential since even u.s.c. *and* l.s.c. mappings F do not necessarily have a continuous selection.

Example 3.8. On the closed Euclidean unit ball B of $\mathbb{R}^2$, define

$$F(x) = \{y \in B \mid \|y - x\|_2 \geq \tfrac{1}{2}\} \qquad (x \in B). \tag{3.13}$$

It is easy to see that F is u.s.c. and l.s.c. everywhere. A continuous selection $f \in F$ cannot exist on B since $f : B \to B$ would have a fixed-point ξ (Brouwer's fixed-point theorem). Then $\xi = f(\xi) \in F(\xi)$ contradicts to the definition of F. $\diamond$

Exercise: Analyze the continuity properties for F with the Euclidean norm and with polyhedral norms, respectively, and with ">" instead of "≥".

4 The lightning function and constant Clarke subdifferential

4.1 Some preparations (Clarke's subdifferential)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function.
1. Given $x, u \in \mathbb{R}^n$, the limsup

$$f'_{Cl}(x, u) := \limsup_{x_k \to x,\ t_k \downarrow 0} \ t_k^{-1}(f(x_k + t_k u) - f(x_k)); \qquad x, u \in \mathbb{R}^n$$

is Clarke's directional derivative of f at x in direction u. If x is a local minimizer of f then (put $x_k = x$) it follows

$$f'_{Cl}(x, u) \geq 0 \ \forall u \quad \text{(necessary, weak condition for a local minimizer).} \tag{4.1}$$

This condition is weak since it also holds for $f = -|x|$ at $\bar{x} = 0$ (take a first sequence with $x_k < 0$ and a second one with $x_k > 0$).

2. For any $x^* \in \mathbb{R}^n$, consider $f - x^*$ with $(f - x^*)(x) = f(x) - \langle x^*, x\rangle$ and define a set $\partial^{Cl} f(x)$ by saying

$$x^* \in \partial^{Cl} f(x) \Leftrightarrow (f - x^*)'_{Cl}(x, u) \geq 0 \ \forall u. \tag{4.2}$$

3. By the definition only, one obtains: $\partial^{Cl} f(x) = \partial g(0)$ for the sublinear function $g(.) = f'_{Cl}(x, .)$. Hence, the generalized subdifferential $\partial^{Cl} f(x)$ - meanwhile called Clarke's subdifferential of f at x [9] - turns out to be a non-empty, convex compact set.

3. Clearly, $(f - x^*)'_{Cl}(x, u) = f'_{Cl}(x, u) - \langle x^*, u\rangle$. In consequence, condition

$$0 \in \partial^{Cl} f(x) \tag{4.3}$$

coincides with (4.1) and replaces the necessary condition $Df(x) = 0$ for $f \in C^1$.

Now we present a special real Lipschitz function G such that Clarke's subdifferential fulfills

$$\partial^{Cl} G(x) = [-1, 1] \ \forall x.$$

The existence of such functions has been clarified in [4]. Our construction gives a complete impression of such functions. We shall also see that the following sets are dense in $\mathbb{R}$:

the set $D_N = \{x \mid \text{G is not directionally differentiable (in the usual sense) at } x\}$,
the set of local minimizers and the set of local maximizers.
In addition, $h(x) = \frac{1}{2}(x + G(x))$ has further strange properties.

4.2 Construction

[45] To begin with, let $U : [a, b] \to \mathbb{R}$ be any affine-linear function with Lipschitz rank $L(U) < 1$, and let $c = \frac{1}{2}(a + b)$. As the key of the following construction, we define a linear function V by

$$V(x) = \begin{cases} U(c) - a_k(x - c) & \text{if } U \text{ is increasing,} \\ U(c) + a_k(x - c) & \text{otherwise.} \end{cases}$$

Here, we put

$$a_k := \frac{k}{k+1}, \tag{4.4}$$

and k denotes the step of the (later) construction. Given any $\varepsilon \in (0,\ \frac{1}{2}(b - a))$ consider the following 4 points in $\mathbb{R}^2$:

$$p_1 = (a, U(a)),\ \ p_2 = (c - \varepsilon,\ V(c - \varepsilon)),\ \ p_3 = (c + \varepsilon, V(c + \varepsilon)),\ \ p_4 = (b, U(b)).$$

By connecting these points in natural order, a piecewise affine function

$$w(\varepsilon, U, V) : [a, b] \to \mathbb{R}$$

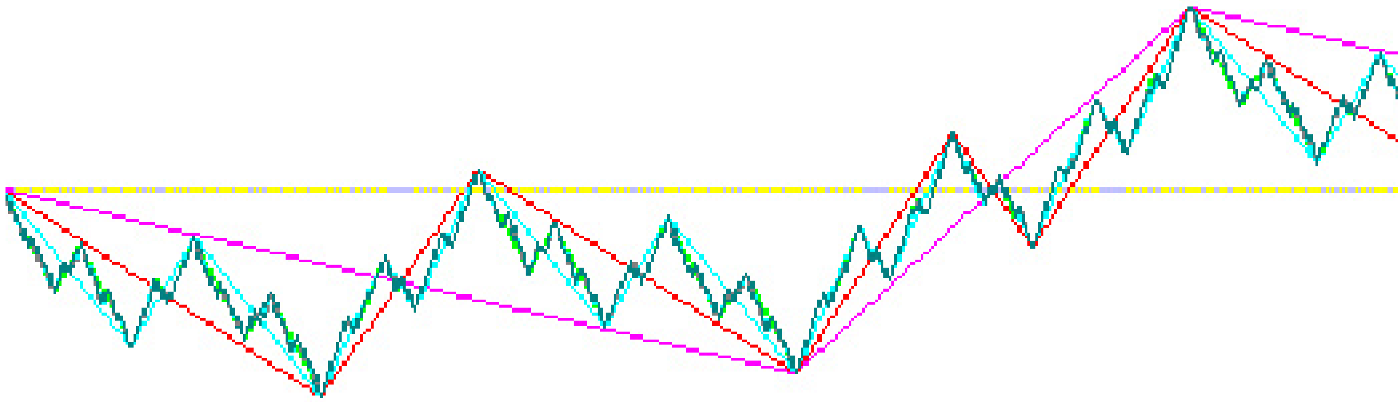
(the lightning) is defined. It consists of 3 affine pieces on the intervals

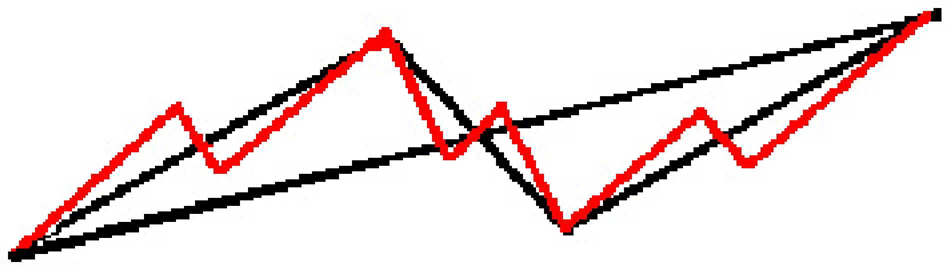
$$[a, c - \varepsilon], \quad [c - \varepsilon, c + \varepsilon], \quad [c + \varepsilon, b].$$

By the construction of V and p_1,...,p_4, it holds

$$Lip\ (w(\varepsilon, U, V)) < 1 \ \text{ if } \varepsilon > 0 \text{ is small.}$$

After taking ε in this way (it may depend on the interval and the step k of our construction), we repeat our construction (like defining Cantor's set) with each of the related 3 pieces and larger k.





Now, start this procedure on the interval [0, 1] with the initial function

$$U(x) = 0 \quad \text{and} \quad k = 1.$$

In the next step $k = 2$ we apply the construction to the 3 pieces just obtained, then with $k = 3$ to the now existing 9 pieces and so on. The concrete choice of the (feasible) $\varepsilon = \varepsilon(k) > 0$ is not important in this context. In any case, we obtain a sequence of piecewise affine functions

$$g_k \quad \text{on} \quad [0, 1]$$

with Lipschitz rank < 1. This sequence has a cluster point g in the space $C[0, 1]$ of continuous functions, and g has Lipschitz rank $L = 1$ due to (4.4). Let

$$N_k = \{y \in (0, 1) \mid g_k \text{ has a kink at } y\} \quad \text{and} \quad N \text{ be the union of all } N_k.$$

If $y \in N_k$, then the values $g_i(y)$ will not change during all forthcoming steps $i > k$. Hence $g(y) = g_k(y)$. The set N is dense in $[0, 1]$. Thus $g_k \to g$ in C.

Let c be a center point of some subinterval $I(k)$ used during the construction (Obviously, these c form a dense subset of the interval). Then c is again a centre point of some subinterval $I(k+i)$ for all $i > 0$. Thus, also $g(c) = g_{k+i}(c)$ holds true for all $i \geq 0$. Let $c_k^+ > c$ and $c_k^- < c$ be the nearest kink-points of g_k right and left from c. Then we have

$$d_k := \frac{g(c) - g(c_k^-)}{c - c_k^-} = \frac{g(c_k^+) - g(c)}{c_k^+ - c} = \pm\frac{k}{k+1} \tag{4.5}$$

where the sign alternates. Via $k \to \infty$ this shows that usual (not Clarke's) directional derivatives $g'(c, \pm 1)$ cannot exist. Thus g is not differentiable at c.

Assume $d_k > 0$. Then (since the orientation of the middle part changes with k) it holds

$$\frac{g(c) - g(c^-_{k+1})}{c - c^-_{k+1}} = \frac{k+1}{k+2} \qquad \text{and}$$

$$g(c) < \min\{g(c^+_k),\ g(c^-_{k+1})\}. \tag{4.6}$$

The inequality tells us that the function g has a local minimizer ξ in $\Omega_k := (c^-_{k+1}, c^+_k)$. If $|x^*| < 1$ and k is large enough then inequality (4.6) holds - due to (4.5) - for the function $g - x^*$, too. Hence also $g - x^*$ has a local minimizer $\xi(x^*)$ in Ω_k, and the sets of local minimizers for g and $g - x^*$, respectively, are dense. By definition, it holds

$$x^* \in \partial^{Cl} g(\xi(x^*)).$$

Since each x is limit of a sequence of minimizers to $g - x^*$, one easily obtains $x^* \in \partial^{Cl} g(x)$. Taking into account that $x \mapsto \partial^{Cl} g(x)$ is closed it follows

$$[-1, 1] \subset \partial^{Cl} g(x)\ \forall x.$$

Since g has Lipschitz rank 1, the equation has to hold.

$$[-1, 1] = \partial^{Cl} g(x)\ \forall x.$$

Starting with large k such that $d_k < 0$, we obtain that the local maximizers form also a dense set. Finally, by a mean-value theorem for Lipschitz functions [9], one obtains

$$\partial^{Cl} g(x) = [-1, 1] = \partial^{gJac} g(x) \quad \forall x \in (0, 1).$$

This tells us, for each $\varepsilon > 0$ and $x \in (0, 1)$: There are sequences $x_n, y_n \to x$ such that $Dg(x_k)$ and $Dg(y_k)$ exist and satisfy $Dg(x_n) \to 1$ and $Dg(y_n) \to -1$.

To extend g on $\mathbb{R}$ one may put $G(x) = g(x - \text{ integer}(x))$ where $\text{integer}(x)$ denotes the integer part of x.

G is nowhere semismooth (semismooth is a useful property for Newton's method; see below).

Derived functions: Let

$$h(x) = \frac{1}{2}(x + G(x)), \quad \text{ then } \quad \partial^{Cl} h(x) = [0, 1]\ \forall x.$$

The Lipschitz function h is strictly increasing and has a continuous inverse h^{-1} which is nowhere locally Lipschitz.

h is not directionally differentiable (in the usual sense) on a dense subset of $\mathbb{R}$.

In the negative direction -1, h is strictly decreasing, but Clarke's directional derivative $h'_{Cl}(x, -1)$ is identically zero.

The integral $F(t) = \int_0^t h(x)\, dx$ is a convex function with strictly increasing derivative h, such that (for generalized derivative-sets defined below),

$$0 \in Th(t)(1) = [0, 1]\ \forall t \quad \text{ and } \quad 0 \in Ch(t)(1) \text{ for all } t \text{ in a dense set.}$$

5 Lipschitzian stability / invertibility

5.1 Stability- Definitions for (Multi-) Functions

5.1.1 Metric and strong regularity

Let $F : X \rightrightarrows Y$ (metric spaces) be a multifunction. In many situations, then the behavior of "solution sets"

$$F^{-1}(y) = \{x \in X \mid y \in F(x)\}$$

is of interest. Multifunctions come into the play, even in the context of functions, if

$$F^{-1}(y) = \{x \in X \mid f(x) \leq y\}, \qquad F(x) = \{y \in \mathbb{R} \mid y \geq f(x)\}$$

for real-valued f and similarly for systems of equations and inequalities. Often F^{-1} describes solution sets (or stationary points) of optimization problems which depend on *parameter* y. Then, the following properties of F or F^{-1} reflect certain Lipschitz-stability of related solutions (being of interest, e.g., if such solutions are involved in other "multilevel" problems [11]). Let $\bar{y} \in F(\bar{x})$.

Definition 5.1. We call F^{-1} pseudo-Lipschitz at $(\bar{x}, \bar{y})$ if there are positive L, ε, δ such that

$$\begin{array}{l} \forall (x, y) : \ [\ x \in F^{-1}(y)\ ,\ y \in B(\bar{y}, \delta), x \in B(\bar{x}, \varepsilon)\ ]\ \ \forall y' \in B(\bar{y}, \delta) \\ \qquad\qquad \exists\ x' \in F^{-1}(y') \quad \text{such that } d(x', x) \leq L d(y', y). \end{array} \tag{5.1}$$

Definition 5.2. If, in addition, x' is unique, then F is called **strongly regular**. ◇

The latter means that - locally near $(\bar{x}, \bar{y})$ - the *inverse* F^{-1} is single-valued and a *Lipschitz function* with rank L. Notice that both properties describe the behavior of F^{-1} and remain valid if we exchange $(\bar{x}, \bar{y})$ by some $(\hat{x}, \hat{y}) \in \operatorname{gph} F$ sufficiently close to $(\bar{x}, \bar{y})$.

The pseudo-Lipschitz property of F^{-1} appears in the literature also under several other notions:

- sometimes F is called pseudo-Lipschitz and often F is called **metrically regular**
- or one says that F^{-1} obeys the Aubin-property.

In any case, one should look at the current definition.

5.1.2 Weaker stability requirements

Setting $(x, y) = (\bar{x}, \bar{y})$**,** condition (5.1) requires

$$\forall y' \in B(\bar{y}, \delta)\ \exists x' \in F^{-1}(y') \text{ such that } \operatorname{dist}(x', \bar{x}) \leq L d(y', \bar{y}) \tag{5.2}$$

which means that F^{-1} is **lower Lipschitz** at $(\bar{x}, \bar{y})$ with rank L. In particular, this implies local solvability of $y' \in F(x)$ if $d(y', \bar{y}) < \delta$.
Setting $y' = \bar{y}$, condition (5.1) requires

$$\begin{array}{l} \forall (x, y) : \ [x \in F^{-1}(y),\ y \in B(\bar{y}, \delta),\ x \in B(\bar{x}, \varepsilon)\ ] \\ \exists x' \in F^{-1}(\bar{y}) \text{ such that } d(x', x) \leq L d(\bar{y}, y). \end{array} \tag{5.3}$$

This requirement defines so-called **calmness of** F^{-1} at $(\bar{y}, \bar{x})$.

Definition 5.3. We call F weak-strong regular at $(\bar{x}, \bar{y})$ if there are positive L, ε, δ such that

$$\begin{array}{c} \forall (x, y) \text{ with } \ y \in F(x),\ y \in B(\bar{y}, \delta),\ x \in B(\bar{x}, \varepsilon) \\ \forall y' \in B(\bar{y}, \delta) \text{ with } \ M := F^{-1}(y') \cap B(\bar{x}, \varepsilon) \neq \emptyset : \\ M \text{ is a singleton and } \ x' \in M \text{ fulfills } d(x', x) \leq L d(y', y). \end{array} \tag{5.4}$$

In other words, we consider F^{-1} on $Y_\varepsilon := \{y' \mid F^{-1}(y') \cap B(\bar{x}, \varepsilon) \neq \emptyset\}$ only. If $\bar{y} \in \operatorname{int} Y_\varepsilon$, we obtain strong regularity and vice versa. The linear function $f : l^2 \to l^2$ as $f(x_1, x_2, ...) = (0, x_1, x_2, ...)$ is weak-strong regular but neither strongly nor metrically regular.

Finally, F^{-1} is called **locally upper Lipschitz** with rank L at $(\bar{x}, \bar{y})$ if (as for $F = |x|$)

$$\forall y' \in B(\bar{y}, \delta): \ (\ F^{-1}(y') \cap B(\bar{x}, \varepsilon)\ ) \subset B(\bar{x},\ Ld(y', \bar{y})). \tag{5.5}$$

In this situation, F^{-1} is calm and $\bar{x}$ is isolated in $F^{-1}(\bar{y})$ (put $y' = \bar{y}$). The sets $F^{-1}(y')$ may be empty. Property (5.5) does not follow from metric regularity (put $F(x) = x_1 + x_2$).

Notice:

Strong regularity implies all other mentioned stability properties.

Calmness follows from all other mentioned stability properties excepted lower Lipschitz.

If $\bar{x}$ is a local minimizer of $f : X \to \mathbb{R}$ then the level set mapping $F^{-1}(y) = \{x \mid f(x) \leq y\}$ is never lower Lipschitz at $(f(\bar{x}), \bar{x})$.

5.1.3 The common question

All introduced stabilities involve a clear and classical analytical question for functions $f = F$: Given (x, y) near $(\bar{x}, \bar{y})$ such that $f(x) = y$ as well as y' near $\bar{y}$, we ask for certain x' satisfying $f(x') = y'$ with small (Lipschitzian) distance $d(x', x)$. The different stability types arise from additional hypotheses or requirements like $y' = \bar{y}$, uniqueness of x' and so on. For multifunctions, the same question concerns the inclusion $y \in F(x)$. Having the differentiable case in mind, many approaches are thinkable to this question.

(1) Try to find x' constructively by a solution method: of Newton-type, by a descent method if f maps into $\mathbb{R}$ and $y' < y$ or by another method [51], [58].

(2) Generalize implicit/inverse function theorems by allowing that certain non-differentiable situations (typical for the problem under consideration) occur [83], [76].

(3) Define new derivatives and show (if possible) how the well-known calculus around implicit functions can be adapted [1], [82], [70].

All these ideas appear in the framework of nonsmooth analysis and not any of them dominates the others. They have specific advantages and disadvantages which will be discussed now.

5.2 First stability-examples

Calmness of f^{-1} does not depend on differentiability.

Example 5.1. The inverse f^{-1} of Dirichlet's function

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases}$$

is *calm* at $(0, 0)$ since $f^{-1}(y) = \emptyset$ for $y \neq 0$ near 0. The mapping $S(y) = \{x \mid f(x) \geq y\}$ is even *pseudo-Lipschitz* at $(0, 0)$ since $f(x) \geq y$ holds for all irrational x and all y near 0. Clearly, f is not closed. $\diamond$

Though calmness may hold for very strange functions, note that:
Even for $f \in C^\infty(\mathbb{R}, \mathbb{R})$, calmness cannot be checked by considering derivatives only.

Example 5.2. Calm and not calm for functions with identical derivatives. Let $f \equiv 0$ and

$$g(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } \ x \ \neq 0 \\ 0 & \text{if } \ x \ = 0. \end{cases}$$

Then $g^{(n)}(0) = f^{(n)}(0) = 0$ for all $n \in I\!N$. f^{-1} is calm at $(0,0)$ since $f^{-1}(y) = \emptyset$ for $y \neq 0$. On the other hand, it holds with each fixed L: If $y > 0$ is small enough and $x \in g^{-1}(y)$ then $\text{dist}(x, g^{-1}(0)) = |x| > Ld(y,0)$. Moreover, given any $q \in (0,1]$, also $\text{dist}(x, g^{-1}(0)) > Ld(y,0)^q$ follows for small $y > 0$. The latter means that g^{-1} is even not *Hoelder-calm* at the origin. $\diamond$

Strong regularity of multifunctions is possible.

Example 5.3. A strongly regular $\mathbb{R}^n$- *multi* function. Let $f(x) = \|x\|_2$ on $\mathbb{R}^n$. Then the usual subdifferential (sect. 1.2.1) has the form

$$\partial f(x) = \begin{cases} B & \text{if } x = 0 \\ \{\frac{x}{\|x\|}\} & \text{if } x \neq 0 \end{cases} \qquad \text{(B = Euclidean unit ball)}$$

$\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is strongly regular at $(0,0)$ since $(\partial f)^{-1}$ is locally constant and single-valued,

$$(\partial f)^{-1}(x^*) = \{x \mid x \text{ minimizes } \ \xi \mapsto f(\xi) - \langle x^*, \xi\rangle \ \} \ = \ \{0\} \qquad \text{if } \|x^*\| < 1. \qquad \diamond$$

6 Basic generalized derivatives

6.1 CF, TF and D^*F

Below, we shall use certain "directional limits" of a function $f : X \to Y$ (normed spaces) at x in direction $u \in X$. They collect certain limits v of difference quotients, namely

Definition 6.1.

$$\begin{aligned} Cf(x;u) &= \{\ v \mid \exists u_k \to u, \qquad\qquad\quad t_k \downarrow 0 : \ \ v = \lim t_k^{-1}\ [f(\ x + t_k u_k) - f(\ x)\ ]\ \ \}, \\ Tf(x;u) &= \{\ v \mid \exists (x_k, u_k) \to (x,u), \ \ t_k \downarrow 0 : \ \ v = \lim t_k^{-1}\ [f(x_k + t_k u_k) - f(x_k)\ ]\ \}. \end{aligned}$$

The mapping Cf is said to be the *contingent derivative (also Bouligand-) derivative of f*. Alternatively, one can define Cf by using the contingent (also Bouligand-) cone to gph f, see below. The limits of Tf were introduced by Thibault in [84, 85] (to define other objects) and called *limit sets*. They appeared in [45, 56] (to study inverse Lipschitz functions) as $\Delta-$ or T- *derivatives.*
Evidently, $Cf(x;u) \subset Tf(x;u)$ is always true. Other useful properties are, for $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$,

$$conv\ Tf(x;u) = \partial^{gJac} f(x)u := \{Au \mid A \in \partial^{gJac} f(x)\}, \tag{6.1}$$

$$Tf(x;u+v) \subset Tf(x;u) + Tf(x;v) \text{ (element-wise sum; by definitions only).} \tag{6.2}$$

$$Tf(x;ru) \ = \ r\ Tf(x;u)\ \forall r \in \mathbb{R} \ \text{ (element-wise multipl.; by definitions only).} \tag{6.3}$$

If $f \in C^1$ then $Cf(x;u) = Tf(x;u) = \{Df(x)u\}$.
If $f(x) = |x|$ then $Cf(0;1) = \{1\}$ and $Tf(0;1) = [-1, \ 1] = Tf(0;-1)$.
In what follows, we also write

$$Cf(x;u) = Cf(x)(u) \ \text{ and } \ Tf(x;u) = Tf(x)(u).$$

Now let $F : X \rightrightarrows Y$ (normed spaces) be multivalued.

Definition 6.2. Given $y \in F(x)$, define CF as: $v \in CF(x,y)(u)$ if
$\exists (u_k, v_k) \to (u,v)$ and $t_k \downarrow 0$ such that $(x + t_k u_k, y + t_k v_k) \in \text{gph}\, F$. $\diamond$

This means that (u,v) is some (Bouligand-) tangent direction to $\mathrm{gph}\, F$ at $z := (x,y)$,

$$v \in CF(z)(u) \Leftrightarrow (u,v) \in C_{\mathrm{gph}\, F}(z) \Leftrightarrow \ \exists t_k \downarrow 0: \ t_k^{-1}\, \mathrm{dist}(z + t_k(u,v), \ \mathrm{gph}\, F) \to 0. \quad (6.4)$$

Evidently, CF corresponds to Cf for functions where $y + t_k v_k = f(x + t_k u_k)$.

If $\mathrm{gph}\, F \subset \mathbb{R}^n \times \mathbb{R}^m$ is a finite union of polyhedral sets [79] then $CF(x,y)$ can be easily determined via classical feasible directions for $\mathrm{gph}\, F$

$$v \in CF(x,y)(u) \Leftrightarrow \exists \varepsilon > 0 : (x,y) + t(u,v) \in \mathrm{gph}\, F \ \forall t \in [0,\varepsilon], \quad (6.5)$$

which leads to linear inequality systems for characterizing (u,v).

Definition 6.3. Similarly, but with more limits: $v \in TF(x,y)(u)$ if
$\exists (u_k, v_k, x_k, y_k) \to (u,v,x,y)$ and $t_k \downarrow 0$: $(x_k, y_k) \in \mathrm{gph}\, F$ and $(x_k + t_k u_k, y_k + t_k v_k) \in \mathrm{gph}\, F$. $\diamond$

This defines a (bigger) set called *strict graphical derivative* in [82]. TF has been applied (up to now) only to such F which can be linearly transformed into $C^{0,1}$ functions, [60, 61, 62], and is hard to compute even for polyhedral F as in example 7.9. We shall see that

- CF plays a role for metric regularity and for being locally upper Lipschitz,
- Tf is crucial for strong and weak-strong regularity.

Remark 6.1. For $f \in C^{0,1}(X, \mathbb{R}^n)$, $Cf(x;u)$ and $Tf(x;u)$ are nonempty and compact, and one may put $u_k \equiv u$ in Def. 6.1, without changing these sets. $\diamond$

Example 6.2. [45] A $C^{0,1}$-function $f : [0, \frac{1}{2}) \to C$ such that directional derivatives f' nowhere exist and $Cf(x;u) = \emptyset$. For $x \in [0, \frac{1}{2})$ define a contin. function $h_x : [0,1] \to \mathbb{R}$ by

$$h_x(t) = \begin{cases} 0 & \text{if } \ 0 \ \le t < x \\ t - x & \text{if } \ x \ \le t < 2x \\ x & \text{if } \ 2x \le t \le 1 \ . \end{cases}$$

Now $f(x) := h_x$ defines a $C^{0,1}$ function $f : [0, \frac{1}{2}) \to C$. Consider the difference quotients

$$g(x,\lambda) = \frac{f(x+\lambda) - f(x)}{\lambda} \qquad \text{and notice that } g(x,\lambda) \in C[0,1].$$

If $\lambda > 0$, then $g(x,\lambda)(2x) \le 0$ and $g(x,\lambda)(2x + 2\lambda) = 1$. Hence $\lim_{\lambda \downarrow 0} g(x,\lambda)$ cannot exist in $C[0,1]$. If $\lambda < 0$, we obtain for $x > 0$ that $g(x,\lambda)(2x) \ge 0$ and $g(x,\lambda)(2x + 2\lambda) = -1$. Thus $\lim_{\lambda \uparrow 0} g(x,\lambda)$ cannot exist, too. In consequence, f has no directional derivative and $Cf(x;u)$ is empty for all directions $u \in \mathbb{R} \setminus \{0\}$. $\diamond$

Again let X and Y be Banach spaces.

Definition 6.4. Mordukhovich's co-derivative [69, 70] $D^*F(x,y) : Y^* \rightrightarrows X^*$. Write $x^* \in D^*F(x,y)(y^*)$ if there exist sequences
$\varepsilon_k, \delta_k \downarrow 0$, $(x_k^*, y_k^*) \to (x^*, y^*)$ (weak*) and $(x_k, y_k) \to (x,y)$ in $\mathrm{gph}\, F$ (strong) such that

$$\begin{array}{c} \langle \ y_k^*, \ v \ \rangle \ \ge \ \langle x_k^*, \ u \ \rangle \ \ - \varepsilon_k \ \|(u,v)\|_{X \times Y} \\ \text{if } (x_k + u, y_k + v) \in \mathrm{gph}\, F \ \text{ and } \ \|(u,v)\|_{X \times Y} < \delta_k. \end{array} \quad (6.6)$$

Having (6.6), $(x_k^*, -y_k^*)$ is said to be an ε_k–normal to $\mathrm{gph}\, F$ at (x_k, y_k) while $(x^*, -y^*)$ is called limiting ε-normal. $\diamond$

Specializations:
case 0. $F = f$ is a $C^{0,1}$ function: Now $y = f(x)$ is unique and
$x^* \in D^*f(x)(y^*) \;\Leftrightarrow \exists\, \varepsilon_k, \delta_k \downarrow 0,\; (x_k^*, y_k^*) \to (x^*, y^*)$ (weak*) and $x_k \to x$ such that

$$\langle\, y_k^*,\; f(x_k + u) - f(x_k)\,\rangle \;\geq\; \langle x_k^*,\; u\rangle \;-\varepsilon_k\, \|u\|_X \quad \text{if } \|u\| < \delta_k. \tag{6.7}$$

If $\dim Y < \infty$, we may obviously put $y_k^* = y^*$ without changing the Definition. If $y^* = e^1$ we simply consider the first component f_1. If $\dim X < \infty$, we may put $x_k^* = x^*$.
case 1. If $F = L : X \to Y$ is linear and continuous then $D^*L(x) = L^*$ coincides with the adjoint operator (direct proof by the definition).
case 2. Level sets. Let $F(\xi) = \{\eta \in \mathbb{R} \mid f(\xi) \leq \eta\}$ $(\xi \in X)$, $f \in C^{0,1}(X, \mathbb{R})$ and $f(x) = y$.
Then: $x^* \in D^*F(x, y)(1) \Leftrightarrow$

$$\begin{array}{c} \exists\, \varepsilon_k, \delta_k \downarrow 0,\; x_k^* \to x^* \text{ (weak*)} \quad \text{and} \quad x_k \to x \text{ such that} \\ f(x_k + u) - f(x_k) \geq \;\langle x_k^*, u\rangle - \varepsilon_k\, \|u\| \quad \text{if} \;\; \|u\| \leq \delta_k. \end{array} \tag{6.8}$$

Any x^* satisfying (6.8) is a so-called limiting Fréchet-subgradient of f at x. Other subdifferentials and their nice and bad properties: see the script to Optimization and Variational Inequalities.
case 3. For $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$, there is a direct relation to Cf (by the definitions only)

$$\begin{array}{c} x^* \in D^*f(x)(y^*) \quad \Leftrightarrow \quad \exists\, x_k \to x,\; \varepsilon_k \downarrow 0 \text{ such that} \\ \langle y^*, v\rangle + \varepsilon_k \geq \langle x^*, u\rangle \;\; \forall (u, v) :\; \|(u, v)\| \leq 1 \;\; \text{and} \;\; v \in Cf(x_k)(u). \end{array} \tag{6.9}$$

case 4. For $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, this yields $v = Df(x_k)u$ and $D^*f(x) = [Df(x)]^* = Df(x)^T$.

6.2 Chain rules and simple Lipsch. functions

By the symmetric/asymmetric definitions, it holds

$$v \in TF(x, y)(u) \Leftrightarrow u \in T(F^{-1})(y, x)(v) \qquad \text{the same for } CF$$

$$-x^* \in D^*F(x, y)(-y^*) \Leftrightarrow y^* \in C(F^{-1})(y, x)(x^*) \qquad \text{multiply with -1 in (6.6).}$$

Computing CF, TF or D^*F may be a hard job not only for multifunctions, but also for Lipschitz functions f in finite dimension. In the standard situation

$$f(x) = g(h(x)) \text{ for } g, h \in C^{0,1} \qquad \text{(appropriate finite dimension)}$$

the inclusion

$$Tf(x)(u) \subset Tg(h(x))[Th(x)(u)] =^{Def.} \{a \mid a \in Tg(h(x))(b) \text{ for some } b \in Th(x)(u)\} \tag{6.10}$$

holds true. If $g \in C^1$, the equation holds. Both statements are direct consequences of the definitions. If $h \in C^1$ the equation may fail.

Example 6.3. *(chain rule).* Let $x \in \mathbb{R}$, $h(x) = (x, 0)$ and $f(x) = g(h(x))$ where

$$g(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 \leq 0; \\ y_1 & \text{if } 0 \leq y_1 \leq |y_2| \\ |y_2| & \text{otherwise.} \end{cases}$$

Then $h(0) = (0, 0)$, $Dh(0) = (1, 0)$, $g(y_1, 0) \equiv 0$, $f \equiv 0$. It follows $Tf(0)(1) = \{0\}$, but $1 \in Tg(h(0))(1, 0)$ (Take $y_{1,k} = 0$, $y_{2,k} = \frac{1}{k} = t_k$ in the definition of T). ◇

"Simple" Lipschitz Functions

For several chain rules, the following property plays a key role.

Definition 6.5. (a private Def. of Fusek, Klatte, Kummer) A function $g \in C^{0,1}[\mathbb{R}^m, \mathbb{R}^q]$ is said to be *simple* at y if, for all $v \in \mathbb{R}^m$, all $w \in Tg(y)(v)$ and each sequence $t_k \downarrow 0$, there is a sequence $y_k \to y$ such that

$$\begin{array}{l} w = \lim t_k^{-1}[g(y_k + t_k v) - g(y_k)] \text{ holds} \\ \text{at least for some subsequence of } k \to \infty. \end{array}$$

All $g \in C^{0,1}(\mathbb{R}^m, \mathbb{R})$ are simple [56]. Further simple functions are $y \mapsto y^+$ and $y \mapsto (y^+, y^-)$, [45], see also Exercise 15.

However, neither all $g \in C^{0,1}(\mathbb{R}, \mathbb{R}^2)$ nor all PC^1-functions g into $\mathbb{R}^2$ are simple. Roughly speaking, then the following situation occurs: To obtain a given limit w_1 for the first component, it may happen that certain special sequences t_k must be taken, which are inappropriate to obtain the limit w_2 for the second component.

Detailed investigations of simple functions and relations to the following chain rule can be found in [21] and [24].

Proposition 6.4. (partial derivatives for Tf). *Let g and h be locally Lipschitz, $f = h(x, g(y))$, and let $D_g h(\cdot,\cdot)$ exist and be locally Lipschitz, too. Then*

$$Tf(x,y)(u,v) \subset T_x h(x, g(y))(u) + D_g h(x, g(y))(Tg(y)(v)). \tag{6.11}$$

If, additionally, g is simple at y then (6.11) holds as equation. ◇

This implies for the product $f(x,y) = A(x)B(y)$ of two $C^{0,1}$- matrix- functions with appropriate dimensions:

Proposition 6.5. *(Product rule) If A or B is simple then*

$$Tf(\bar{x}, \bar{y})(u,v) = [TA(\bar{x})(u)]\ B(\bar{y}) + A(\bar{x})\ [TB(\bar{y})(v)].$$

Both statements also hold for the contingent derivative Cf where <simple> becomes <directional differentiable>. Applied to multifunctions and CF, <simple> leads to the so-called proto-derivative [82], [75].

6.3 A non-simple Lipschitz function $f : \mathbb{R} \to \mathbb{R}^2$

Here, some detailed calculations are needed.

Example 6.6. [24] Put $a_0 = 1$ and consider for $k \in \mathbb{N}$ the points

$$a_k = 2^{-k}, \qquad b_k = \frac{9}{8}\, 2^{-k}, \quad c_k = \frac{1}{2}(a_k + b_k) = \frac{17}{16}\, 2^{-k},$$

$$d_k = \frac{15}{8}\, 2^{-k}, \qquad\qquad e_k = \frac{1}{2}(d_k + a_{k-1}) = \frac{31}{16}\, 2^{-k}.$$

Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by the components

$$f_1(x) = \begin{cases} x - a_k & \text{if} \quad x \in [a_k, c_k], \\ b_k - x & \text{if} \quad x \in [c_k, b_k], \\ 0 & \text{else} \end{cases}$$

$$f_2(x) = \begin{cases} x - d_k & \text{if} \quad x \in [d_k, e_k], \\ a_{k-1} - x & \text{if} \quad x \in [e_k, a_{k-1}], \\ 0 & \text{else,} \end{cases}$$

The function f is globally Lipschitz with modulus $L = 1$. Considering the direction $u = 1$ and the sequences $x_k = c_k,\ \ t_k = e_k - c_k$ we obtain

$$\frac{1}{t_k}\,[f(x_k + t_k u) - f(x_k)] = \frac{16}{14}\,2^k \left[\begin{pmatrix} 0 \\ \frac{1}{16}\,2^{-k} \end{pmatrix} - \begin{pmatrix} \frac{1}{16}\,2^{-k} \\ 0 \end{pmatrix}\right] = \begin{pmatrix} -\frac{1}{14} \\ \frac{1}{14} \end{pmatrix} \in Tf(0)(u).$$

Now let a sequence $\{r_k\}$ with

$$2^{-(k+1)} \ < \ r_k \ < \ \frac{3}{4}\,2^{-k}, \qquad k \in \mathbb{N} \tag{6.12}$$

be given. Our goal (to show <simple>) is to find reals $y_k \to 0$ such that

$$v^k := r_k^{-1}[f(y_k + r_k) - f(y_k)] \to (-\frac{1}{14}, \frac{1}{14})^T$$

at least for some subsequence. This implies that, for large k, the first (second) component of v^k has to be negative (positive), respectively. Hence there are indices $n(k)$, $\ell(k)$, $n(k) \geq \ell(k)$ with $y_k \in [a_{n(k)}, b_{n(k)}]$ and $y_k + r_k \in [d_{\ell(k)}, a_{\ell(k)-1}]$, and we have $r_k \geq d_{\ell(k)} - b_{n(k)}$.

For $\ell(k) \leq n(k) - 1$ we would get $r_k \geq d_{n(k)-1} - b_{n(k)} = \frac{21}{8}\,2^{-n(k)}$ and

$$|v_1^k| \ \leq \ r_k^{-1}\,\frac{1}{16}\,2^{-n(k)} \ \leq \ \frac{1}{42} \ < \ \frac{1}{14}.$$

Thus, in order to obtain the limes $(-\frac{1}{14}, \frac{1}{14})^T$ only subsequences with $\ell(k) = n(k)$ are suitable. As a consequence, it follows

$$r_k \geq d_{n(k)} - b_{n(k)} = \frac{3}{4}\,2^{-n(k)} \ \text{ and } \ r_k \leq a_{n(k)-1} - a_{n(k)} = 2^{-n(k)}.$$

In other words, for the sequence $\{r_k\}$ (6.12), one cannot find a suitable sequence of indices $\{n(k)\}$. Hence f is not simple at 0. $\diamond$

Violation of formula

$$Tf(x, y)(u, v) = T_x h(x, g(y))(u) + T_g h(x, g(y))(Tg(y)(v)). \tag{6.13}$$

For the function f of this example, there were pair–wise disjoint intervals $I_k(f)$ and some $v(f) \in Tf(0)(1)$, such that the equation

$$v(f) = \lim r_k^{-1}[f(y_k + r_k) - f(y_k)] \ \text{ with } y_k \to 0$$

can only hold if $r_k \in I_k(f)$ (for some infinite subsequence). Let the same situation occur with respect to a second function $g : \mathbb{R} \to \mathbb{R}^2$ and intervals $I_k(g)$ such that $I_k(g) \cap I_\nu(f) = \emptyset\ \forall k, \nu$. Defining now

$$h(x, y) = (f(x), 0) + (0, g(y)) \in \mathbb{R}^4, \quad x, y \in \mathbb{R},$$

the point $(v(f), v(g))$ cannot belong to $Th(0, 0)(1, 1)$, and chain rule (6.13) fails to hold even for a sum of functions.

7 Sufficient stability conditions

We connect now stability with properties of generalized derivatives. Of course, this makes only sense when the latter can be determined.

7.1 Main motivations for defining CF and TF

First of all, let us note that - directly by the definitions - injectivity of $CF(\bar{x},\bar{y})$ and of $TF(\bar{x},\bar{y})$ is *equivalent* to certain stability properties, provided $\dim X + \dim Y < \infty$.

Proposition 7.1. *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be closed and $\bar{y} \in F(\bar{x})$.*

$$\begin{array}{llll} (a) & F^{-1} & \text{is loc. upper Lipsch. at } (\bar{y},\bar{x}) & \Leftrightarrow \; CF(\bar{x},\bar{y})(.) \text{ is injective.} \\ (b) & F & \text{is weak-strong regular at } (\bar{x},\bar{y}) & \Leftrightarrow \; TF(\bar{x},\bar{y})(.) \text{ is injective.} \end{array} \tag{7.1}$$

Statement (a) was shown in [40], statement(b) in [60] (by negation of the stab. requirements).

7.2 Metric regularity

For characterizing $F^{-1} : Y \rightrightarrows X$ (now Banach spaces) to be pseudo-Lipschitz, one can again apply the contingent derivative CF. Let us claim:

$$\exists L > 0 : \forall (x,y) \in \operatorname{gph} F \text{ near } (\bar{x},\bar{y}) : \; B_Y(0,1) \subset CF(x,y)(B_X(0,L)). \tag{7.2}$$

This condition requires *uniform surjectivity of the multifunctions $CF(x,y)$ near $(\bar{x},\bar{y})$ with a linear rate.* Such mappings are also called *uniformly open.* Conditions like (7.2) are often equivalent to certain stability, cf. [19], [20], [73], [87], but checking them is highly non-trivial.

Proposition 7.2. *(Aubin/Ekeland [1])*
If F is closed and (7.2) holds true then F is metrically regular at $(\bar{x},\bar{y})$ with constant L.

In finite dimension, the point-wise inclusion $B_Y(0,1) \subset CF(\bar{x},\bar{y})(B_X(0,L))$ is necessary for F^{-1} to be lower Lipschitz at $(\bar{y},\bar{x})$. However, even for continuous functions f, it is not sufficient to ensure (weaker) $f(\bar{x}) \in \operatorname{int} f(X)$.

Example 7.3. [45]. *The pointwise condition (7.2).* We construct $f : \mathbb{R}^2 \to \mathbb{R}^2$ (continuous) with $f(0) = 0$, $f'(0;u) = u \;\forall u \in \mathbb{R}^2$ and $0 \notin \operatorname{int} f(\mathbb{R}^2)$. Let

$$M = \{(x,y) \in \mathbb{R}^2 \mid \; |y| \geq x^2 \text{ if } x \geq 0, \;\; x^2 + y^2 \leq 1, \;\; x \leq \tfrac{1}{2}\}$$

and $G = \operatorname{conv} M$. For $(x,y) \in M$, let $f(x,y) = (x,y)$. For $(x,y) \in G \backslash M$ with $y \geq 0$ put $f(x,y) = (x,x^2)$. In order to define f at the remaining points $(x,y) \in G \backslash M$ with

$$-x^2 < y < 0$$

let D be the nonlinear triangle given by upper/lower parables and the points

$$P^1 = (x,-x^2),\; P^2 = (0,0),\; P^3 = (x,x^2) \quad \text{and let } t = t(x,y) = -\frac{y}{x^2}.$$

Then $t \in (0,1)$. We assign, to (x,y), the point $h(x,y) = (x,\; t(-x^2) + (1-t)x^2)$ between the parables. Then h is continuous. Next we shift the point $h(x,y)$ to the left boundary of D and call this (continuous) horizontal projection $p(x,y)$. Finally, define f by

$$f(x,y) = p(x,y).$$

So f becomes a continuous function of the type $G \to M$. Setting $g(z) = f(\pi(z))$ where $\pi(z)$ is the projection of $z = (x,y)$ onto G, f can be continuously extended to the whole space. We identify f and g. Clearly, $f'(0;u) = u$ holds for all $u \in \mathbb{R}^n$, and $0 \notin \operatorname{int} f(\mathbb{R}^2)$. $\diamond$

Propos. 7.2 generalizes the classical Graves-Lyusternik-Theorem,

Proposition 7.4. *[28], [67].*
If $f \in C^1(X,Y)$ and $Df(\bar{x}) : X \to Y$ is surjective then f is metrically regular at $(\bar{x}, f(\bar{x}))$.

If $X = Y$ are Hilbert spaces, then the image $f(B(\bar{x}, \varepsilon))$ is even convex for small $\varepsilon > 0$, cf. [74]. Along with Propos. 7.11, this is one of only few statements which tell us something about the structure of gph f in case of metric regularity.

Provided that X and Y are Asplund spaces (like $\mathbb{R}^n$ and L^p, $1 < p < \infty$) also injectivity of D^*F (in place of the uniform surjectivity in Propos. 7.2) ensures a sufficient condition.

Proposition 7.5. *(Mordukhovich [70])*
If F is closed, X, Y are Asplund and $0 \notin D^*F(\bar{x}, \bar{y})(y^*)$ whenever $y^* \neq 0$ then F is metrically regular at $(\bar{x}, \bar{y})$.

For $\dim X + \dim Y < \infty$, these 3 sufficient conditions are even necessary. But already in Hilbert spaces, the sufficient conditions Propos. 7.2 and Propos. 7.5 are very strong, far from being necessary.

7.3 The sufficient conditions of Mordukhovich and Aubin/Ekeland in l^2

We consider the level-set map for one of the simplest nonsmooth, nonconvex functions on a Hilbert space. f is monotone in all components, is concave, globally Lipschitz and nowhere positive.

Example 7.6. [45]. Let $X = l^2$, $x = (x_1, x_2, ...)$ and $f(x) = \inf_k x_k$. Put $F(x) = \{y \in \mathbb{R} \mid f(x) \leq y\}$ such that $F^{-1}(y) = \{x \in X \mid f(x) \leq y\}$ is a level set map. Since f is concave the usual directional derivatives $f'(x; u)$ exist and (due to the Lipsch. property) $Cf(x; u) = \{f'(x; u)\}$. Recalling $f \leq 0$, it holds $f'(x; u) \leq 0 \forall u$ if $f(x) = 0$ (In particular for $x = \xi$ in (7.3)). Now we summarize the main properties of f and F^{-1}.

(i) F^{-1} is (globally) pseudo-Lipschitz, e.g., with rank $L = 2$. Indeed, if $f(x) \leq y$ and $y' < y$, there is some k such that $x_k < y + \frac{1}{2}|y' - y|$.

Put $x' = x - 2|y' - y|e^k$ where e^k is k-th unit vector in l^2. Then, $\|x' - x\| \leq 2|y' - y|$ is trivial, and $x' \in F^{-1}(y')$ follows from $f(x') \leq x'_k \leq y - \frac{3}{2}|y' - y| \leq y'$.

(ii) At each $\xi \in l^2$ with $\xi_k > f(\xi)\ \forall k$, it holds

$$f'(\xi; u) \geq 0 \quad \forall u \in X. \tag{7.3}$$

In consequence, condition (7.2) is violated. We show even more for ξ from (ii):
If $f(\xi + tu) \leq f(\xi) - t$ holds for certain $t \downarrow 0$ and bounded u, say for $\|u\| \leq C$, then $u = u(t)$ necessarily depends on t, and there is no (strong) accumulation point of $u(t)$.

Proof: By assumption, we have

$$\xi_k > f(\xi) = \inf_n \xi_n = 0\ \forall k \quad \text{and} \quad \xi_k + tu_k < -\frac{1}{2}t \text{ for some } k.$$

Due to $|u_k| \leq C$ and $\xi_k > 0$, the second inequality cannot hold for $t \downarrow 0$ if k is fixed. Similarly, it cannot hold if $k = k(t) \leq m$ is bounded since $\min_{k \leq m} \xi_k > 0$. Thus $k(t)$ diverges. So one obtains from $\xi_k > 0$ by division that $u_{k(t)} < -\frac{1}{2}$ holds for an infinite number of components. If u is fixed, this yields the contradiction $u \notin l^2$.

Hence u depends on t. Assuming $u(t) \to u^0$ for certain $t \downarrow 0$, we obtain again a contradiction, namely $\liminf_{t \downarrow 0} u(t)_{k(t)} \leq -\frac{1}{2}$ for certain $k(t) \to \infty$, though $u^0 \in l^2$ yields necessarily $\lim_{k \to \infty} u^0_k = 0$. $\square$

(iii) Mordukhovich's injectivity condition is violated since $0 \in D^*F(0,0)(1)$. To see this, let $x^k = -\frac{e^k}{k}$ and $x^{*k} = e^k$. Then $x^{*k} \to 0$ (weak*). We show according to (6.8) that $\exists\ \varepsilon_k, \delta_k \downarrow 0$ such that

$$f(x^k + u) - f(x^k) \geq \langle x^{*k}, u \rangle - \varepsilon_k \|u\| \quad \text{if } \|u\| \leq \delta_k. \tag{7.4}$$

Obviously, we have $f(x^k) = -\frac{1}{k}$, $\langle x^{*k}, u\rangle = u_k$ and $f(x^k+u) = \inf\left\{-\frac{1}{k} + u_k,\ \inf_{\nu\neq k} u_\nu\right\}$. With $\|u\| < \delta_k := \frac{1}{2k}$, then $f(x^k+u) = -\frac{1}{k} + u_k$ follows and (7.4) holds true since

$$(-\frac{1}{k} + u_k) + \frac{1}{k} \ \geq \ u_k \ - \varepsilon_k \ \|u\|. \qquad \diamond$$

7.4 Strong regularity for $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ via Tf and $\partial^{gJac} f$

Notice that the mapping $u \mapsto \partial^{gJac} f(\bar{x})u$ is injective iff all $A \in \partial^{gJac} f(\bar{x})$ are regular matrices.

Proposition 7.7. *[8]*
Any $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(\bar{x}, f(\bar{x}))$ if all $A \in \partial^{gJac} f(\bar{x})$ are regular. $\diamond$

Proposition 7.8. *[56]*
Any $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(\bar{x}, f(\bar{x})) \ \Leftrightarrow \ Tf(\bar{x}, .)$ is injective. $\diamond$

These conditions do not coincide, see below.

7.5 Strong regularity with singular generalized Jacobians

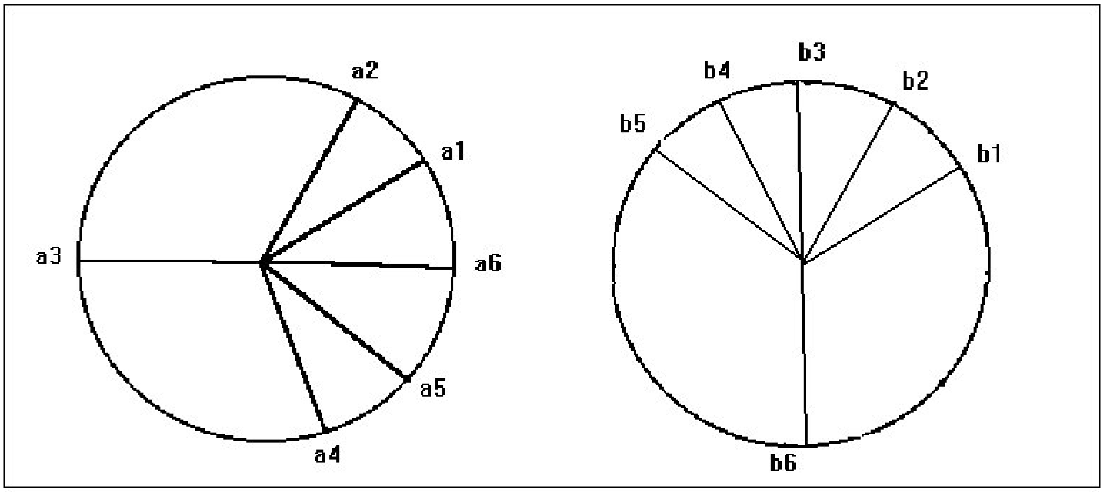
Example 7.9. *A piecewise linear bijection of $\mathbb{R}^2$ with $0 \in \partial^{gJac} f(0)$.* [56], [45].
On the sphere of $\mathbb{R}^2$, let vectors a^k and b^k $(k = 1, 2, ..., 6)$ be arranged as follows:
Put $a^7 = a^1, b^7 = b^1$ and ensure the following properties, see the picture below:

(i) $a^1 = b^1, a^2 = b^2;\ a^4 = -b^4, a^5 = -b^5$.

(ii) The vectors a^k and b^k turn around the sphere in the same order.

(iii) The cones K_i generated by a^i and a^{i+1}, and P_i generated by b^i and b^{i+1}, are proper.

Let $L_i : \mathbb{R}^2 \to \mathbb{R}^2$ be the unique linear function satisfying $L_i(a^i) = b^i$ and $L_i(a^{i+1}) = b^{i+1}$. Setting $f(x) = L_i(x)$ if $x \in K_i$ we define a piecewise linear, continuos function which maps K_i onto P_i. By construction, f is surjective and has a well-defined piecewise linear, continuous inverse (given by L_i^{-1} on P_i); hence f is a strongly regular piecewise linear homeomorphism of $\mathbb{R}^2$. Moreover, $f = id$ on $\operatorname{int} K_1$ and $f = -id$ on $\operatorname{int} K_4$. Thus, $\partial^{gJac} f(0)$ contains E and $-E$ and, by convexity, the zero-matrix, too. $\diamond$



7.6 General relations between strong and metric regularity

7.6.1 Loc. Lipschitz functions

To begin with let $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ and $\bar{y} = f(\bar{x})$.
If $m = n$, then the usual implicit function theorem ensures
metrically regular $\Leftrightarrow$ strongly regular at $(\bar{x}, \bar{y})$ $\Leftrightarrow$ $\det Df(\bar{x}) \neq 0$.
If $\operatorname{rank} Df(\bar{x}) = m < n$, one obtains metric regularity (again by the implicit function theorem) but never strong regularity. If $\operatorname{rank} Df(\bar{x}) < m$, metric regularity fails. Hence, for C^1 functions in finite dimension, the characterization of strong/metric regularity is evident.
We study now locally Lipschitz functions for $m = n$.

Example 7.10. metrically regular $\neq$ strongly regular for a function $f \in C^{0,1}(\mathbb{R}^2, \mathbb{R}^2)$. Take the complex function

$$f(z) = \begin{cases} \frac{z^2}{|z|} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0 \end{cases}$$

(as a $\mathbb{R}^2$ function) and study the equation $f(z) = \zeta$ with two solutions for $\zeta \neq 0$. $\diamond$

Example 7.10 is typical for a general property of loc. Lipschitz functions.

Proposition 7.11. *(Fusek, [23]) Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be metrically regular at $(\bar{x}, f(\bar{x}))$ and directionally differentiable at $\bar{x}$. Then $\bar{x}$ is isolated in $f^{-1}(f(\bar{x}))$ and $f'(\bar{x}; .)$ is injective.* $\diamond$

Nevertheless, the equations $f(x) = y$ may have solutions $x_1(y) \neq x_2(y)$, both converging to $\bar{x}$ as $y \to \bar{y} = f(\bar{x})$. If f is not directionally differentiable, there is neither a proof nor a counterexample for $\bar{x}$ being isolated in $f^{-1}(\bar{y})$ as yet.

7.6.2 KKT-mapping and Kojima's function with/without C^2- functions

We are now going to consider particular $C^{0,1}$ functions $\Phi : \mathbb{R}^\mu \to \mathbb{R}^\mu$ which are closely related to stationary points in optimization problems.
For parametric optimization problems $P(p)$ with parameter $p = (a, b, c) \in \mathbb{R}^{n+m+m_h}$

$$\min \{f(x) - \langle a, x\rangle \mid g_i(x) \le b_i,\ h_j(x) = c_j;\ i = 1, ..., m,\ j = 1, ..., m_h\} \quad f, g, h \in C^1 \tag{7.5}$$

the set $KKT(p)$ of Karush-Kuhn-Tucker- points $(x, y, z) \in \mathbb{R}^{n+m+m_h}$ is given by

$$\begin{array}{lcl} Df(x) + \sum y_i Dg_i(x) + \sum z_j Dh_j(x) & = & a \\ g(x) \le b,\ h(x) = c; \quad y \ge 0, \quad y_i(g_i(x) - b_i) & = & 0 \ \ \forall i. \end{array} \tag{7.6}$$

This is the usual Lagrange condition if inequalities are deleted.

Proposition 7.12. *Under some regularity of the constraints, e.g.*
- *calmness of the constraint map* $M(b, c) = \{x \in \mathbb{R}^n \mid g(x) \le b,\ h(x) = c\ \}$ *at* $(0, 0, \bar{x})$,
- *or the stronger condition MFCQ at* $\bar{x}$
($\text{rank}\, Dh(\bar{x}) = m_h$ *and* $\exists u :\ Dh(\bar{x})u = 0$ *and* $Dg_i(\bar{x})u < 0\ \forall i$ *with* $g_i(\bar{x}) = 0$*),*
it holds:
If $\bar{x}$ *solves (locally) problem (7.5) at* $p = 0$ *then* $\exists y, z$ *such that* $(\bar{x}, y, z) \in KKT(0)$. $\diamond$

As well-known, MFCQ is equivalent to the pseudo-Lipschitz property of $M(.)$ at $(0, 0, \bar{x})$. (Once more a consequence of the implicit function theorem).

Kojima's function: The KKT-System for $p = 0$ can be written in terms of Kojima's [52] function $\Phi : \mathbb{R}^\mu \to \mathbb{R}^\mu$ which has the components

$$\begin{array}{rclcccl} \Phi_1 & = & Df(x) & + & \sum_i y_i^+ Dg_i(x) & + \sum_\nu z_\nu Dh_\nu(x), & y_i^+ = \max\{0, y_i\}, \\ \Phi_{2i} & = & g_i(x) & - & y_i^-, & & y_i^- = \min\{0, y_i\}, \\ \Phi_3 & = & h(x). & & & & \end{array} \tag{7.7}$$

The zeros of Φ are related to KKT- points via the (loc. Lipschitzian) transformations

$$\begin{array}{lll} (x, y, z) \in \Phi^{-1}(0) & \Rightarrow\ (x,\ y^+, z) & \text{is KKT-point} \\ (x, y, z) \text{ a KKT-point} & \Rightarrow\ (x,\ y + g(x),\ z) \in \Phi^{-1}(0) & \end{array} \tag{7.8}$$

and Φ is, for $f, g, h \in C^2$, one of the simplest nonsmooth functions.

The product form: Moreover, Φ can be written as a (separable) product

$$\Phi(x, y, z) = \mathcal{M}(x)\ N(y, z) \tag{7.9}$$

$$\text{where} \qquad N = (1,\ \ y_1^+, ..., y_m^+,\ \ y_1^-, ..., y_m^-,\ z)^T \in \mathbb{R}^{1+2m+m_h} \tag{7.10}$$

and

$$\mathcal{M}(x) = \begin{pmatrix} Df(x) & Dg_1(x)... & Dg_m(x) & 0... & 0... & 0 & Dh_1(x)... & Dh_{m_h}(x) \\ g_i(x) & 0\ \ ... & 0 & 0... & -1... & 0 & 0\ \ \ ... & 0 \\ h(x) & 0\ \ ... & 0 & 0... & 0... & 0 & 0\ \ \ ... & 0 \end{pmatrix} \tag{7.11}$$

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

$$\Phi(x, y, z) = (a, b, c)^T \tag{7.12}$$

describes by (7.8) the KKT-points $KKT(p)$ of problem (7.5).

Replacing Df by another function of corresponding dimension and smoothness, the system describes solutions of variational inequalities over $M(b,c)$.

Due to the structure of Φ and since $N(.)$ is <simple> , the derivatives $T\Phi$ and $C\Phi$ (Def. 6.1) can be exactly determined for $f,g,h \in C^{1,1}$ (derivatives loc. Lipsch.) by the product rule Propos. 6.5 (provided $T\mathcal{M}$ or $C\mathcal{M}$ is available). After that, questions on stability of solutions (locally upper Lipsch., strong regularity) can be reduced to injectivity of $C\Phi$ and $T\Phi$), respectively.

All other known concepts for strong/metric regularity require $f,g,h \in C^2$ due to the used technique. 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 [11], [71].

For $f,g,h \in C^2$, non-smoothness is *only* implied by the components of N:

$$\phi(y_i) = (y_i^+,\ y_i^-) = (y_i^+,\ y_i - y_i^+) = \tfrac{1}{2}\ (y_i + |y_i|,\ y_i - |y_i|). \tag{7.13}$$

So, Φ is a PC^1 function (useful for Newton's method, sect. 9.2), and we need generalized derivatives of the *absolute value* at the origin only. In addition, the equation

$$TN(\bar y)(v) = \partial^{gJac} N(\bar y)(v) := \{Av \mid A \in \partial^{gJac} N(\bar y)\}$$

is obvious. This implies, since $\mathcal{M}(.)$ is C^1 (for more explicit formulas see [45]),

$$\partial^{gJac}\Phi(\bar x,\bar y)(u,v) = T\Phi(\bar x,\bar y)(u,v) = [D\mathcal{M}(\bar x)u]N(\bar y) + \mathcal{M}(\bar x)\partial^{gJac} N(\bar y)(v).$$

7.6.3 Stability of KKT points

The final results follow by computing $T\Phi$ or $C\Phi$ in terms of the given functions. Once more, this is possible by the product rule since N is <simple>.

Assume $f,g \in C^2$ and delete equations (only for a more compact description). Again, let $KKT(a,b) = KKT(p)$ be the set of KKT points. We shall see:

(i) The local upper Lipschitz property
of KKT at $(0,(\bar x,\bar y))$ can be checked by studying the linear system

$$\begin{array}{lcll} D^2L_x(\bar x,\bar y^+)u & + & Dg(\bar x)^T\,\alpha & = 0,\\ Dg(\bar x)\ u & - & \beta & = 0,\\ \alpha_i = 0\ \text{ if }\ g_i(\bar x) < 0, & & \beta_i = 0 \quad \text{if}\ \ \bar y_i > 0, & \end{array} \tag{7.14}$$

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

$$\alpha_i\beta_i = \ 0, \quad \alpha_i \ \ge\ 0\ \ge\ \beta_i \quad \text{if}\ \ \bar y_i = g_i(\bar x) = 0. \tag{7.15}$$

(ii) The strong regularity of KKT^{-1} (or of Kojima's function Φ)
at $(0,(\bar x,\bar y))$ can be checked by studying system (7.14) where (α,β) has, instead of (7.15), to satisfy the weaker condition

$$\alpha_i\beta_i \ge 0 \quad \text{if}\ \ \bar y_i = g_i(\bar x) = 0. \tag{7.16}$$

These systems have the trivial solution $(u,\alpha,\beta) = 0 \in \mathbb{R}^{n+2m}$. They do not change after replacing the original problem (7.5) at $p = 0$ by its *quadratic approximation* at $(\bar x,\bar y)$:

$$\min\ \{Df(\bar x)(x-\bar x) + \tfrac{1}{2}(x-\bar x)^T D^2L_x(\bar x,\bar y^+)(x-\bar x) \mid g_i(\bar x) + Dg_i(\bar x)(x-\bar x) \le 0\}. \tag{7.17}$$

Proposition 7.13. *In both cases,*
the related Lipschitz property for KKT just means (equivalently), that the corresponding systems (7.14, 7.15) and (7.14, 7.16), respectively, are only trivially solvable. $\diamond$

For $f, g \in C^{1,1}$, proofs and history of these statements we refer to [45]. By considering solutions with $u = 0$, both stabilities imply the constraint qualification LICQ at $\bar{x}$ (the gradients of active constraints are linearly independent) which makes Lagrange multipliers unique.

7.6.4 The Dontchev-Rockafellar Theorem for Lipschitzian gradients ?

Again we study the problem (7.5) and use the notations above. Recall that $KKT(.)$ is pseudo-Lipschitz (by definition) iff Φ is metrically regular.

Proposition 7.14. *(Dontchev/Rockafellar [15]). Let all involved functions f, g, h be C^2. Then, if Φ is metrically regular at $(\bar{x}, y, z, 0)$, Φ is even strongly regular at this point.* $\diamond$

This statement (formulated for variational inequalities) fails to hold for $C^{1,1}$-functions under (7.5), even without constraints.

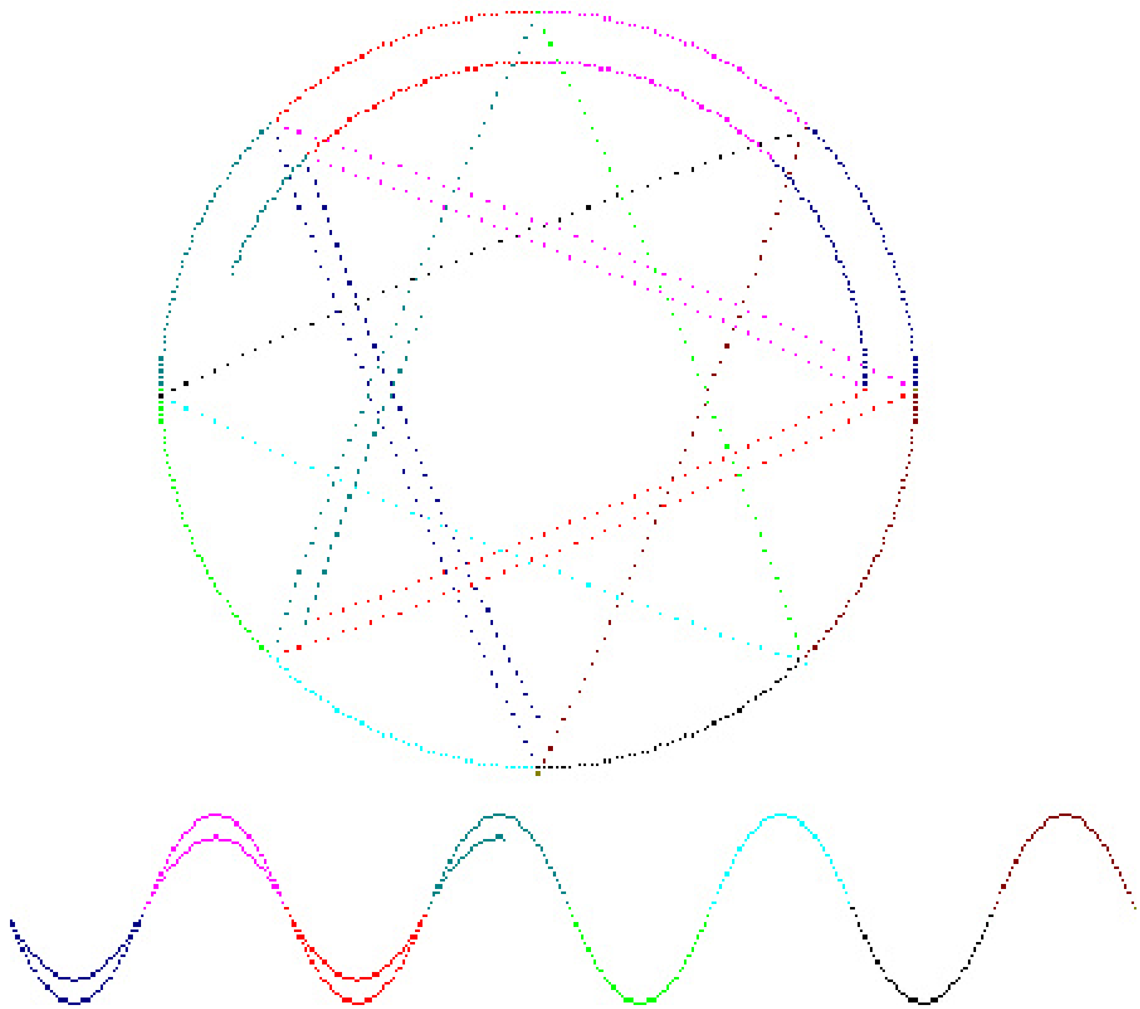
Example 7.15. [45] *A piecewise quadratic function $f \in C^{1,1}(\mathbb{R}^2, \mathbb{R})$ having pseudo-Lipsch. stationary points (solutions of $Df(x, y) = a \in \mathbb{R}^2$) which are -locally- not unique (hence also not strongly regular).*
We write $(x, y) \in \mathbb{R}^2$ in polar-coordinates, $r(\cos\phi, \sin\phi)$, and describe f as well as the partial derivatives $D_x f$, $D_y f$ over 8 cones (of size $\pi/4$)

$$C(k) = \{\ (x, y) \mid \ \phi \in [\frac{k-1}{4}\pi,\ \frac{k}{4}\pi]\ \},\ \ (1 \le k \le 8), \qquad \text{by}$$

cone	f	$D_x f$	$D_y f$
$C(1)$	$y\,(y-x)$	$-y$	$+2y-x$
$C(2)$	$x\,(y-x)$	$-2x+y$	x
$C(3)$	$x\,(y+x)$	$+2x+y$	x
$C(4)$	$-y(y+x)$	$-y$	$-2y-x.$

On the remaining cones $C(k+4)$, $(1 \le k \le 4)$, f is defined as in $C(k)$.

Studying the Df-image of the sphere, it is not difficult to see (but needs some effort) that Df is continuous and $(Df)^{-1}$ is pseudo-Lipschitz at the origin. For $a \in \mathbb{R}^2 \backslash \{0\}$, there are exactly 3 solutions of $Df(x, y) = a$. Our picture shows Df and f if (x, y) turns around the sphere. $\diamond$



8 Explicite stability conditions for *stationary* points

Now let S denote the map of *stationary points* for (7.5). We assume $f, g \in C^2$ and delete equations (only for a more compact description), i.e.,

$$S(a,b) = \{x \mid \exists y: \ (x,y) \text{ is a KKT point for } P(a,b) \ \}, \quad p = (a,b). \tag{8.1}$$

Obviously, $S(p)$ is a projection of $KKT(p)$. Let $\bar{x} \in S(0)$ be the crucial point and suppose throughout MFCQ at $\bar{x}$ for $p = 0$ (without MFCQ, nearly nothing is known for stability under nonlinear constraints). Even with MFCQ, the behavior of S is not Lipschitz for simple examples.

Example 8.1. Consider the "classical" problem (Bernd Schwartz ca 1970) for $x \in \mathbb{R}^2$,

$$\min \ x_2 \quad \text{such that} \ \ g_1(x) = -x_2 \le b_1, \ \ g_2(x) = x_1^2 - x_2 \le b_2.$$

At the origin, MFCQ holds true with $u = (0,1)$. Setting $a \equiv 0$, $b_2 = 0$; $b_1 = -\varepsilon$ we obtain $S(0,b) = \{(x_1, \varepsilon) \mid |x_1| \le \sqrt{\varepsilon}\}$. Hence S is neither calm nor loc. upper Lipschitz at 0. $\diamond$

8.1 Necessary and sufficient conditions

8.1.1 Locally upper Lipschitz

Proposition 8.2. *(upperLip) S is locally upper Lipschitz at $(0, \bar{x})$ $\Leftrightarrow$ each solution of system (7.14), (7.15) (for each Lagr. multiplier $\bar{y}$ to $\bar{x}$) satisfies $u = 0$.*
If $\bar{x}$ was a local minimizer for $p = 0$, the condition even implies $S(p) \neq \emptyset$ for small $\|p\|$. $\diamond$

For a proof see Thm. 8.36 [45]. The proof of the first statement uses the fact that MFCQ ensures - with the Kojima function Φ

$$u \in CS(0, \bar{x})(\alpha, \beta) \;\Leftrightarrow\; (\alpha, \beta) \in \cup_{\bar{y} \in Y(0,\bar{x}),\; v \in \mathbb{R}^m} C\Phi(\bar{x}, \bar{y})(u, v). \tag{8.2}$$

Thus the local 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 (7.14), (7.15). In consequence, for two problems with the same first and second derivatives of f, g at $\bar{x}$, the stationary point mappings are either both locally upper Lipschitz or both not.

The same remains true (only the formulas change) for $S = S(a)$ with fixed constraints $[b \equiv 0]$, though this situation is surprisingly more involved, cf. [60, 61, 62].

8.1.2 Weak-strong regularity

Similar statements, beginning with formula (8.2) for TS, are not known for metric and strong regularity. In contrary, we shall see (sect. 8.2) that a comparable simple answer does not exist - even in the subclass of convex, polynomial problems.

Without loss of generality (since inactive constraints can be removed), we suppose $g(\bar{x}) = 0$. We also put $A_i = Dg_i(\bar{x})$.

Proposition 8.3. *[46]. (strLip) The mapping S^{-1} is not weak-strong regular at $(0, \bar{x})$ $\Leftrightarrow$*

$$\begin{array}{l} \textit{There exist } u \in \mathbb{R}^n \setminus \{0\} \textit{ and a Lagrange vector } y \textit{ to } (0,\bar{x}) \textit{ such that} \\ y_i\; A_i u \;=\; 0\; \forall i, \textit{ and with certain } x_k \to \bar{x} \textit{ and } \alpha^k \in \mathbb{R}^m, \textit{ one has} \\ \alpha_i^k\; A_i u \geq 0\; \forall i \quad \textit{and} \quad \lim_{k\to\infty} \sum_i \alpha_i^k\; Dg_i(x_k) = -D_x^2 L(\bar{x}, y) u. \end{array} \tag{8.3}$$
$\diamond$

If all constraints are linear (disregarding only one quadratic constraint) the limit condition (where $\|\alpha^k\| \to \infty$ is possible) can be simplified into a non-limit form. Generally, (8.3) cannot be replaced by a condition in terms of derivatives (for f, g at $\bar{x}$) until a fixed order.

Next put again $p = (a, b)$ and let $Y(p, x)$ be the set of Lagr. multipliers for p and x.

Proposition 8.4. *(AubStat) The pseudo-Lipschitz 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*

$$\begin{array}{ll} Dg_i(x_k)u^* = 0 & \textit{if } \; y_i > 0 \textit{ for some } y \in Y(p_k, x_k), \\ \alpha_i^* \leq 0 \;\; \textit{and} \;\; Dg_i(x_k)u^* \leq 0 & \textit{if } \; y_i = g_i(x_k) - b_i^k = 0 \textit{ for some } y \in Y(p_k, x_k), \\ \alpha_i^* = 0 & \textit{if } \; g_i(x_k) - b_i^k \; < 0 \end{array} \tag{8.4}$$

$$\textit{and} \qquad \|D_x^2 L(\bar{x}, y)u^* + Dg(\bar{x})^T \alpha^*\| < \varepsilon_k \downarrow 0 \quad \forall y \in Y(x_k, p_k).$$
$\diamond$

A proof and specializations of Propos. 8.4 can be found in [45], Thm. 8.42. By choosing an appropriate subsequence, the index sets in (8.4) can be fixed. But setting $(p_k, x_k) \equiv (0, \bar{x})$ violates again the equivalence for nonlinear g.

Remark 8.5. The conditions of Propos. 8.3 and 8.4 are equivalent to non-injectivity of TS^{-1} and D^*S^{-1}, respectively (at the point in question), cf. Propositions 7.1, 7.5. Hence verifying injectivity of these generalized derivatives (not to speak about computing them) requires to study the same limits.

8.2 Bad properties for strong and metric regularity of stationary points

Next we have $\bar{x} = 0 \in \mathbb{R}^2, \bar{p} = 0 \in \mathbb{R}^2$ and write $A_i = Dg_i(0)$.

We will show - by modifying example 8.1 as in [46] - that condition (8.3) cannot be simplified and that (weak-) strong regularity cannot be handled by looking at the first 123 derivatives of the involved functions at $\bar{x}$ alone.

Example 8.6. Consider the following problem for parameter $(a,b) = 0$ with some real constant r:

$$\min\ rx_1^2 + x_2 \ \text{ such that } \ g_1(x) = -x_2 \le 0,\ g_2(x) = x_1^2 - x_2 \le 0.$$

Then $Df = (2rx_1, 1), Dg_1 = (0,-1), Dg_2 = (2x_1, -1)$ and $\bar{x} = (0,0)$ is a stationary point with $Y^0 = \{y \ge 0 \mid \ y_1 + y_2 = 1\}$ and $A_1 = A_2 = (0,-1)$. With $\gamma = 2r + 2y_2$, we have

$$Q(y) := D_x^2 L(\bar{x}, y) = \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence $u\,Q(y) = (\gamma u_1, 0)$.

Since at least one y_i is positive for $y \in Y^0$, it follows $u \perp A_i\ \forall i$ from $y_i A_i u = 0 \forall i$. Hence all u of interest have the form $u = (u_1, 0),\ u_1 \ne 0$. Condition (8.3) now requires exactly that for some sequence of $(\alpha_1, \alpha_2) \in \mathbb{R}^2$ and of converging $x \to \bar{x}$, it holds

$$(\gamma u_1, 0) + \alpha_1(0,-1) + \alpha_2(2x_1, -1) \to 0.$$

This condition cannot be satisfied *with fixed* $x = \bar{x} = 0$ whenever $\gamma \ne 0$. Note that $\gamma \ne 0$ holds for all $y \in Y^0$ if $r \notin [-1,0]$, so convexity of the problem plays no role.

On the other hand, we can define the sequences $x = (\frac{1}{k},\ 0),\ \alpha_2 = -\frac{1}{2}\ k\ \gamma u_1,\ \alpha_1 = -\alpha_2$ in order to satisfy the singularity condition.

Thus, if $r \notin [-1,0]$, S^{-1} *is not weak-strong regular* (the same for $r \in [-1,0]$ by other arguments). Moreover, if $r > 0$ then - in spite of singularity - Kojima's condition [52]

$$\text{For each } y \in Y^0, \quad Q(y) \text{ is positive definite on } K(y) = \{u \mid u \perp A_i \ \text{ if } y_i > 0\}. \tag{8.5}$$

for his modified definition of *strong stability* is satisfied at $(0, \bar{x})$. ◇

Example 8.7. Change example 8.6, with some integer $q \ge 2$ and $r = 1$, as follows

$$\min\ x_1^2 + x_2 \ \text{ such that } \ g_1(x) = -x_2 \le 0,\ g_2(x) = x_1^{q+1} - x_2 \le 0. \tag{8.6}$$

We obtain again *singularity* at $(0,0)$, since for any $u = (u_1, 0) \ne 0$, it holds

$$(2u_1, 0) + \alpha_1(0,-1) + \alpha_2(\ (q+1)x_1^q, -1\ ) \to 0$$

for the sequences $x = (\frac{1}{k},\ 0),\ \alpha_2 = -\frac{2u_1}{(q+1)x_1^q}$ and $\alpha_1 = -\alpha_2$.

For odd q, we are still in the class of convex, polynomial problems with unique and continuous solutions $x(p)$ for all parameters $p = (a,b)$.

Nevertheless,
one cannot identify the singularity by using alone the first q derivatives of f and g at $\bar{x}$,

since these derivatives are the same for the next, strongly regular example with $r = 1$. ◇

Example 8.8. Change only the second constraint in example 8.6,

$$\min\ rx_1^2 + x_2 \ \text{ such that } \ g_1(x) = -x_2 \le 0,\ g_2(x) = -x_2 \le 0.$$

Now the mapping S^{-1} is *strongly regular* at $(0, \bar{x})$ for every $r \ne 0$ (Dg_i is constant). If $r < 0$, the stationary points are never minimizers. ◇

Finally, our problems had unique solutions for $r > 0$. So weak-strong regularity is strong regularity and, moreover, the same unpleasant situations occur in view of metric regularity.

9 The nonsmooth Newton method

9.1 Convergence

We summarize properties of f which are necessary and sufficient for solving an equation

$$f(x) = 0, \qquad f : \mathbb{R}^n \to \mathbb{R}^n \qquad \text{loc. Lipschitz}$$

by a Newton-type method. Such methods can be applied to KKT-systems (after any reformulation as an equation).

The crucial conditions

Newton's method for computing a zero $\bar{x}$ of $f : X \to Y$ (normed spaces) is determined by

$$x_{k+1} = x_k - A^{-1} f(x_k), \tag{9.1}$$

where $A = Df(x_k)$ is supposed to be invertible. The formula means that x_{k+1} solves

$$f(x_k) + A(x - x_k) = 0, \qquad A = Df(x_k). \tag{9.2}$$

Forgetting differentiability replace A by any invertible linear operator $A_k : X \to Y$, assigned to x_k (if $Df(x_k)$ exists, A_k could take the place of an approximation). To replace also the regularity condition of $Df(\bar{x})$ for the usual C^1-Newton method, suppose:

$$\exists\ K^+, K^- \text{ such that } \|A_k\| \le K^+ \text{ and } \|A_k^{-1}\| \le K^- \quad \forall A_k \text{ and small } \|x_k - \bar{x}\|. \tag{9.3}$$

The *locally superlinear convergence* of Newton's method means that, for some o-type function r and initial points x_0 near $\bar{x}$, we have

$$x_{k+1} - \bar{x} = z_k \quad \text{ with } \quad \|z_k\| \le r(x_k - \bar{x}). \tag{9.4}$$

Substituting x_{k+1} from (9.1) and applying A_k to both sides, this requires

$$f(x_k) = f(x_k) - f(\bar{x}) = A_k\ (x_k - x_{k+1}) = A_k\ [(x_k - \bar{x}) - z_k] \text{ with } \|z_k\| \le r(x_k - \bar{x}). \tag{9.5}$$

Condition (9.5) claims equivalently (with $A = A_k$)

$$A(x_k - \bar{x}) = f(x_k) - f(\bar{x}) + Az_k, \quad \|z_k\|_X \le r(x_k - \bar{x}) \tag{9.6}$$

and yields *necessarily*, with

$$o(u) = K^+ r(u)\ : \tag{9.7}$$

$$A_k(x_k - \bar{x}) = f(x_k) - f(\bar{x}) + v^k \quad \text{ for some } v^k \in B(0,\ o(x_k - \bar{x})) \subset Y. \tag{9.8}$$

Sufficiency: Conversely, having (9.8), it follows

$$x_k - \bar{x} = A_k^{-1}(f(x_k) - f(\bar{x})) + A_k^{-1} v^k \quad \text{ for some } v^k \in B(0,\ o(x_k - \bar{x})). \tag{9.9}$$

So the solutions of equation (9.1) fulfill

$$\begin{aligned} z_k := x_{k+1} - \bar{x} \ &= (x_{k+1} - x_k)\ +\ (x_k - \bar{x}) \\ &= -A_k^{-1} f(x_k)\ +\ A_k^{-1}(f(x_k) - f(\bar{x})) + A_k^{-1} v^k =\ A_k^{-1} v^k. \end{aligned}$$

Hence $\|z_k\| = \|A_k^{-1} v^k\| \le K^- o(x_k - \bar{x})$. This ensures the convergence (9.4) with

$$r(u) = K^- o(u) \tag{9.10}$$

for all initial points near $\bar{x}$. So we have shown

Proposition 9.1. *(Convergence) Under the regularity condition (9.3), method (9.1) fulfills the convergence-condition (9.4)* $\Leftrightarrow$ *the assignment* $x_k \mapsto A_k$ *satisfies, for* x_k *near* $\bar{x}$*, the approximation-condition (9.8).* $\diamond$

Hence we may use any $A = A(x) \in Lin(X, Y)$ whenever the conditions

$$(CI) \qquad \|A(x)\| \le K^+ \quad \text{and} \quad \|A(x)^{-1}\| \le K^- \ \ \text{(injectivity)} \ \ \text{as well as} \tag{9.11}$$

$$(CA) \qquad A(x)(x - \bar{x}) \in f(x) - f(\bar{x}) + o(x - \bar{x})B \ \ \text{(approximation)} \tag{9.12}$$

are satisfied for sufficiently small $\|x - \bar{x}\|$. The quantities $o(.)$ and $r(.)$ are directly connected by (9.7) and (9.10).
A multifunction $\mathcal{N}$ which assigns, to x near $\bar{x}$, a non-empty set $\mathcal{N}(x)$ of linear functions $A(x)$ satisfying (CA) with the same $o(.)$ is called a *Newton map* (at $\bar{x}$) in [45] .

In the current context, the function $f : X \to Y$ *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 $C^{0,1}$ functions we cannot suggest any reasonable definition for $A(x)$ since (9.11) and (9.12) already imply the (pointwise) Lipschitz estimate

$$\|f(x) - f(\bar{x})\| \le (1 + K^+)\|x - \bar{x}\|.$$

Because the zero $\bar{x}$ is usually unknown, this estimate should be required at least for all $\bar{x}$ *near* the zero of f. Similarly, it makes sense to require (CA) for points $\bar{x}$ *near* the solution, too. If this can be satisfied, f is called *slantly differentiable* in [34].

9.2 Semismoothness

Condition (9.12) appears in various versions in the literature. Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$. If all $A(x)$ in $\partial^{gJac} f(x)$ satisfy (9.12) with the same $o(.)$, then f is called *semismooth* at $\bar{x}$, [63]; sometimes – if $o(\cdot)$ is even quadratic – also *strongly semismooth*. In others papers, A is a mapping that approximates $\partial^{gJac} f$; and all f satisfying the related conditions (9.12) are called weakly semismooth. The simplest example is $f(x) = |x|$.

However, neither a relation between A and $\partial^{gJac} f$ nor the existence of directional derivatives is essential for the interplay of the conditions (9.4), (9.11) and (9.12) in Propos. 9.1. The main problem is the characterization of those functions f which allow us to *find practically relevant Newton functions* $A = A(x)$ satisfying (9.12). This class of functions f is not very big (for finite dimension, see [45], locally PC^1-functions).

A particular Newton map:
Let $f = PC^1(\alpha_1, ..., \alpha_k)$ be a piecewise C^1 function, i.e., f is continuous, $\alpha_j \in C^1(X, Y)$, and for each x there is some (*active*) $j = j(x)$ such that $f(x) = \alpha_j(x)$. In this case, all

$$A(x) \in \{D\alpha_j(x) \mid \alpha_j(x) = f(x)\}$$

fulfill (9.12); this is a simple exercise. So one may select any $A(x) = D\alpha_{j(x)}(x)$.

Also semi-smoothness of the Euclidean norm makes no problems. Near $\bar{x} \ne 0$, it is C^∞. For $\bar{x} = 0$ and $x \ne 0$, we have $Df(x)(x - \bar{x}) = \|x\| - \|\bar{x}\|$. For other approaches to such methods, more references and basic papers, cf. [18], [45], [30], [53].

9.3 Alternating Newton sequences everywhere for $f \in C^{0,1}(\mathbb{R}, \mathbb{R})$

Newton methods cannot be applied to all loc. Lipsch. functions, even if $n = 1$ (provided the steps have the usual form at C^1-points of f). Assumptions like strong regularity due not help, in the present context, since the subsequent function is everywhere strongly regular (with uniform L) and even differentiable at the unique zero.

Example 9.2. [55], [45]
Alternating sequences for $f \in C^{0,1}(\mathbb{R}, \mathbb{R})$ with almost all initial points.
To construct f, put $I(k) = [k^{-1}, (k-1)^{-1}] \subset \mathbb{R}$ for integers $k \geq 2$, and put

$$\begin{aligned} c(k) &= \tfrac{1}{2}\,[k^{-1} + (k-1)^{-1}] && \text{(the center of } I(k)) \\ c(2k) &= \tfrac{1}{2}\,[(2k)^{-1} + (2k-1)^{-1}] && \text{(the center of } I(2k)). \end{aligned}$$

In the (x, y)-plane, let $g_k = g_k(x)$ be the lin. function through the points $((k-1)^{-1}, (k-1)^{-1})$ and $(-c(k), 0)$,

$$\text{i.e.,} \qquad g_k(x) = a_k(x + c(k)), \quad \text{where} \quad a_k = \frac{(k-1)^{-1}}{(k-1)^{-1} + c(k)}.$$

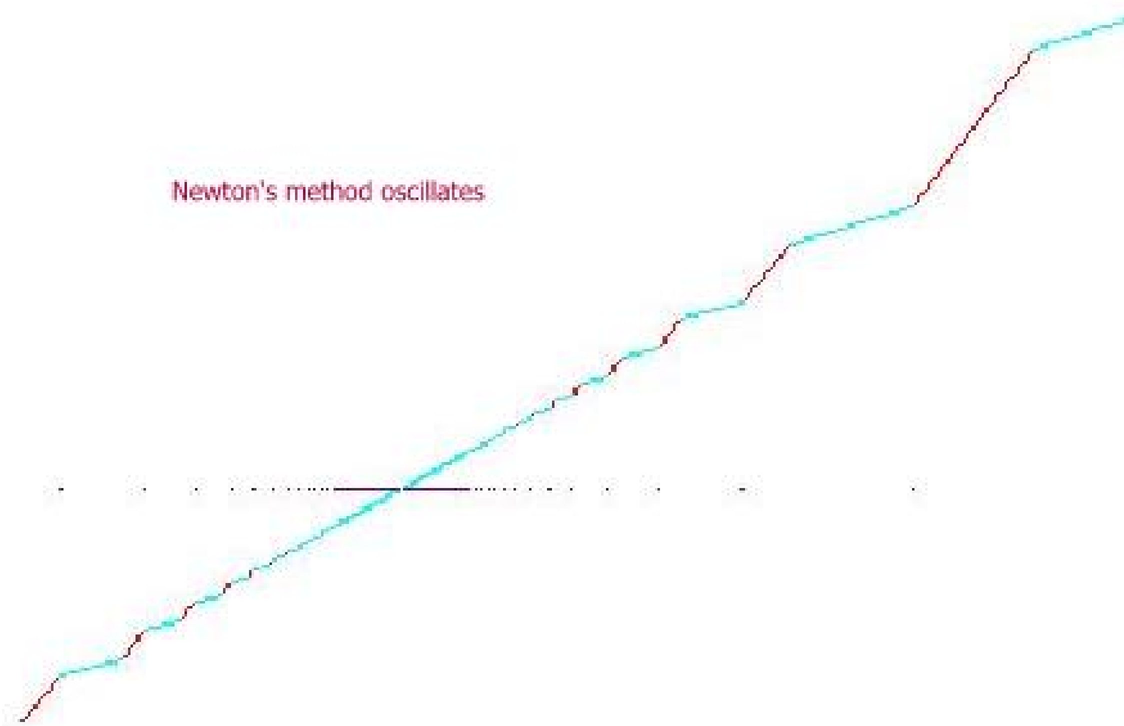
Similarly, let $h_k = h_k(x)$ be the lin. function through the points (k^{-1}, k^{-1}) and $(c(2k), 0)$

$$\text{i.e.,} \qquad h_k(x) = b_k(x - c(2k)) \ \text{ where } \ b_k = \frac{k^{-1}}{k^{-1} - c(2k)}.$$

Evidently, $g_k = 0$ at $x = -c(k)$, $\quad h_k = 0$ at $x = c(2k)$. Now define f for $x > 0$ as

$$f(x) = \min\,\{g_k(x),\ h_k(x)\} \ \text{ if } x \in I(k) \qquad \text{and} \qquad f(x) = g_2(x) \ \text{ if } x > 1.$$

Finally, let $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0$. $\diamond$



Properties of f: For $k \to \infty$, one obtains $\lim a_k = \frac{1}{2}$ and $\lim b_k = 2$. The assertion $Df(0) = 1$ can be directly checked. Again directly, one determines the global Lipschitz rank

$$L = \max b_k = b_2 = \frac{1}{2} / \left[\ \frac{1}{2} - \frac{1}{2}(\frac{1}{4} + \frac{1}{3})\ \right] = \frac{12}{5}.$$

On the left side of interval $I(k), f$ coincides with h_k, on the right with g_k. Because of

$$g_k(c(k)) < h_k(c(k)),$$

it holds $f = g_k$ at $c(k)$, and f is differentiable near the center point $c(k)$.

Now, start Newton's method at any $x^0 \neq 0$ where $Df(x^0)$ exists. Then the next iterate x^1 is some point $\pm c(k)$. There, it holds $Df = Dg_k$ (or $Df = -Dg_k$ for negative arguments). Hence, the next point x^2 is the center point on the opposite side. It follows that the method generates the alternating sequence $x^0, x^1, x^2 = -x^1, x^3 = x^1, ...$

Other counterexamples: Study also $f(x) = x^q,\ q \in (0,1)$, which shows the difficulties if f is everywhere C^1 except the origin, and if f is not locally Lipschitz.

9.4 Difficulties for elementary $f \in C^{0,1}(l^2, l^2)$

can be seen by

Example 9.3. [29]. Let $f : l^2 \to l^2,\ \ f_i(x) = x_i^+,\ i = 1, 2, ...$, and define, as in $\mathbb{R}^n$,

$$A_i(x)u = \begin{cases} u_i & \text{if } x_i > 0 \\ 0 & \text{if } x_i \leq 0. \end{cases}$$

At $\bar{x}$ with $\bar{x}_k \neq 0\ \forall k$, we check condition (CA), i.e., $A(x)(x - \bar{x}) \in f(x) - f(\bar{x}) + o(x - \bar{x})B$. Put $x^k = \bar{x} - 2\bar{x}_k e^k \in l^2$. Then $x^k - \bar{x} = -2\bar{x}_k e^k \to 0\ (k \to \infty)$ (not possible in $\mathbb{R}^n$),

$$f_i(x^k) = \begin{cases} (-\bar{x}_i)^+ & \text{if } i = k \\ (+\bar{x}_i)^+ & \text{if } i \neq k \end{cases} \quad \text{and} \quad A_i(x^k)(x^k - \bar{x}) = \begin{cases} -2\bar{x}_k e^k & \text{if } x_i^k > 0 \text{ and } i = k \\ 0 & \text{otherwise.} \end{cases}$$

If $i \neq k$ this implies $f_i(x^k) - f_i(\bar{x}) = 0 = A_i(x^k)(x^k - \bar{x})$.
If $i = k$ this implies, due to $x_i^i = -\bar{x}_i$,

$$f_i(x^k) - f_i(\bar{x}) = (-\bar{x}_i)^+ - (\bar{x}_i)^+ \ \text{ and } \ A_i(x^k)(x^k - \bar{x}) = \begin{cases} -2\bar{x}_i & \text{if } x_i^i = -\bar{x}_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus we obtain for $i = k$,

$$\begin{array}{llll} f_i(x^k) - f_i(\bar{x}) = |\bar{x}_i|; & A_i(x^k)(x^k - \bar{x}) = 2|\bar{x}_i| & \text{if} & -\bar{x}_i > 0; \\ f_i(x^k) - f_i(\bar{x}) = -|\bar{x}_i|; & A_i(x^k)(x^k - \bar{x}) = 0 & \text{if} & -\bar{x}_i \leq 0. \end{array}$$

The difference is at least $|\bar{x}_i| = |\bar{x}_k|$. Since $\|x^k - \bar{x}\| = 2|\bar{x}_k|$, (CA) is violated. ◇

10 Convex sets with empty algebraic relative interior

For many statements of functional analysis, crucial sets have to possess interior points. Usually, this requires more than to have a nonempty algebraic interior or a relative algebraic interior. Given a convex set K in a linear space V one defines

$$\begin{array}{llll} x \in \text{algrelint } K & \text{if } \forall y \in K\ \exists r > 0: & x - r(y - x) \in K \\ x \in \text{algint } K & \text{if } \forall y \in V\ \exists r > 0: & x - r(y - x) \in K. \end{array} \tag{10.1}$$

For $V = \mathbb{R}^n$, these notions coincide with the related topological definitions.

10.1 The space of convex compact subsets of $\mathbb{R}^n$

Now we study a space of convex sets (see [72] for details and references). It has been investigated in order to extend the usual subdifferential to functions which are the difference of two convex functions (Demianov, Rubinov).

Let $\mathcal{K}$ be the set of all nonempty, convex and compact subsets of $\mathbb{R}^n$. By the addition

$$A + B = \{a + b \mid \ a \in A, b \in B\}$$

we obtain a commutative half group with zero $0 = \{0\}$. Using the separation theorem, one easily shows the cancellation law

$$A + C = B + C \ \Rightarrow \ A = B.$$

Thus embedding in a group is possible by considering all pairs $(A, B) \in \mathcal{K}\times\mathcal{K}$ with equivalence relation

$$(A, B) \sim (C, D) \text{ if } A + D = B + C$$

and new addition

$$(A, B) + (C, D) =^{Def} (A + C, B + D)$$

which is invariant with respect to the equivalence relation and gives

$$(A, 0) + (0, A) =^{Def} (A, A) \sim (0, 0).$$

Hence we may identify the pairs (A, B) and the equivalence classes $[A, B]$ like in the context of natural and integer numbers, respectively. To simplify we still write (A, B) and identify equivalent elements. Now define a multiplication with real r: First put $rA = \{ra \mid a \in A\}$ for $r \geq 0$ and next

$$r(A, B) = (rA, rB) \text{ if } r \geq 0, \qquad r(A, B) = (|r|B, |r|A) \text{ if } r < 0.$$

We obtain a vector space V (even a metric can be introduced) where $(-1)(A, B) = (B, A)$ is the inverse element w.r. to addition. The set $\mathcal{K}$ is now the convex cone $(\mathcal{K},\ 0) \subset V$ (the "non-negative orthant"). We show

Proposition 10.1. *algrelint* $(\mathcal{K}, 0) = \emptyset$ *for all* $n > 1$. ◇

Proof. Given $A, B \in \mathcal{K}$ we have to consider, for small $r > 0$, the ray

$$(A, 0) - r(B, A) = (A, 0) + r(A, B),$$

and to ask for convexity of $(A, 0) + r(A, B)$ which means by definition,

$$(A, 0) + r(A, B) \ \sim \ (C_r, 0) \quad \text{ for some } C_r \in \mathcal{K}.$$

The latter means $A + rA = C_r + rB$ and

$$A \in \text{algrelint } \mathcal{K} \ \Leftrightarrow \ \forall B \in \mathcal{K}\ \exists r > 0\ \exists C_{B,r} \in \mathcal{K}: \ (1 + r)A = C_{B,r} + rB. \tag{10.2}$$

If $n = 1$, the interval $A = [-1, 1]$ has the claimed property. Now assume that $A \in$ algrelint $\mathcal{K}$ exists for $n > 1$. For any $u \in \mathbb{R}^n$, consider

$$p_A(u) = \max_{a\in A}\langle u, a\rangle \quad \text{ and the set of maximizer } \ \Psi_A(u). \tag{10.3}$$

Choose $u \neq 0,\ v \perp u,\ v \neq 0$ and put $B = [-1, 1]v$ (to simplify, study $u = e^1$, $v = e^2$). We obtain

$$(1 + r)A = C + r[-1, 1]v, \qquad C = C_{B,r}. \tag{10.4}$$

Let $x \in \Psi_C(u)$ (it depends on v and r, too).

Then all $\xi \in x + r[-1, 1]v$ satisfy $\langle u, \xi\rangle = \langle u, x\rangle$ and belong to $(1 + r)A$. In addition, (10.4) tells us that all ξ maximize $\langle u, a\rangle$ on $(1 + r)A$, too. In consequence, all $\xi' = \frac{\xi}{1+r}$ maximize $\langle u, a\rangle$ on A. Thus $\Psi_A(u)$ is not a singleton.

However, then p_A cannot be differentiable at u. Since this holds for each $u \neq 0$ and p_A is convex (even sublinear = positively homogeneous and subadditive), we arrived at a contradiction. □

10.2 Spaces of sublinear and convex functions

- By the definitions (10.3), an isomorphism between $\mathcal{K}$ and the set Π of sublinear, positive homogeneous functions $p : \mathbb{R}^n \to \mathbb{R}$ is established:

$$A \mapsto p_A, \qquad p \mapsto A := \partial p(0)$$

  (use again the separation theorem to verify this fact, called Minkovski duality). Now, the space V corresponds with the space D of all functions $p - q,\ p, q \in \Pi$. Again, algrelint $(\Pi, 0)$ is empty in D.

- Let C be the set of all continuous real functions $x = x(t), t \in \mathbb{R}$ and K be the subset of all convex x. Given any $x \in K$ define $y \in K$ as $y(t) = e^{(t^2)} + e^{\max\{0,x(t)\}}$.

  Since $\lim_{t\to\pm\infty}[x(t) - r(y(t) - x(t))] = -\infty\ \forall r > 0$, the function $x - r(y - x)$ is not convex. Thus $x \notin$ algrelint K.

- Another example, related to Michael's selection theorem, can be found in [2], p. 31. There, $F : X = [0, 1] \rightrightarrows \mathbb{R}$ is the l.s.c. multifunction defined by (1.3) and K is the convex set of all continuous f such that $f(x) \in F(x)\ \forall x$.

11 Exercises

Exercise 7 [45] *Verify: If $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(\bar{x}, f(\bar{x}))$ and directionally differentiable near $\bar{x}$ then the local inverse f^{-1} is directionally differentiable near $f(\bar{x})$.*

Otherwise one finds images $y = f(x)$ for x near $\bar{x}$ and $v \in \mathbb{R}^n$ such that $Cf^{-1}(y)(v)$ contains at least two different elements p and q. Since f' exists and $p \in Cf^{-1}(y)(v)$ iff $v \in Cf(x)(p)$, one obtains $f'(x; p) = v = f'(x; q)$. For small $t > 0$, then the images

$$f(x + tp) - f(x + tq) = f(x + tp) - f(x) - (f(x + tq) - f(x))$$

differ by a quantity of type $o(t)$ while the pre-images differ by $t(p - q)$. Therefore, the local inverse f^{-1} cannot be Lipschitz near $(f(\bar{x}), \bar{x})$ for $p \neq q$.

Exercise 13 [45] *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be strongly regular at $(\bar{x}, 0)$. Show, e.g., by applying (6.1) and (6.2), that the local inverse f^{-1} is semismooth at 0 if so is f at $\bar{x}$.*

Otherwise, $\partial^{gJac}(f^{-1})$ is not a Newton map at 0. Then, due to

$$conv\ Tf^{-1} = \partial^{gJac}(f^{-1}),$$

also Tf^{-1} is not a Newton map at 0: There exist $c > 0$ and elements $u \in Tf^{-1}(y)(y - 0)$ such that

$$\|u - (f^{-1}(y) - f^{-1}(0))\| > c\|y\| \quad \text{where } u = u(y) \text{ and } y \to 0.$$

Setting $x = f^{-1}(y)$ and using that f and f^{-1} are locally Lipschitz, we obtain with some new constant $C > 0$:

$$\|u - (x - \bar{x})\| \geq C\|x - \bar{x}\|.$$

Since Tf is a Newton map at $\bar{x}$, we may write (with different $o-$ functions)

$$Tf(x)(x - \bar{x}) \subset f(x) - f(\bar{x}) + o(x - \bar{x})B = y + o(x - \bar{x})B.$$

Next apply $u \in Tf^{-1}(y)(y) \;\Leftrightarrow y \;\in Tf(x)(u)$. By *subadditivity* of the homogeneous map Tf, we then observe

$$\begin{array}{rcl} y \in Tf(x)(u) & \subset & Tf(x)(u+\bar{x}-x) + Tf(x)(x-\bar{x}) \\ & \subset & Tf(x)(u+\bar{x}-x) + y + o(x-\bar{x})B. \end{array}$$

Hence

$$0 \in Tf(x)(u+\bar{x}-x) + w \quad \text{ holds with certain } w \in o(x-\bar{x})B.$$

We read the latter as

$$u+\bar{x}-x \in Tf^{-1}(y)(-w)$$

which yields, with some Lipschitz rank L of f^{-1} near the origin,

$$C\|x-\bar{x}\| \le \|u-(x-\bar{x})\| \le L\|w\| \le Lo(x-\bar{x}).$$

This is impossible for o-type functions and proves the statement.

Exercise 15 [45] *Verify that positively homogeneous $g \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ are* <simple> *at the origin.*
Let $v \in Tg(0)(r)$ and $t_k \downarrow 0$ be given ($k = 1, 2, ...$). We know by the structure of $Tg(0)(r)$ that there exist q_k such that $v_k := g(q_k + r) - g(q_k) \to v$.

Given k select some $\nu > k$ such that $\|t_\nu q_k\| < 1/k$ and put $p_\nu = t_\nu q_k$. Then

$$v_k = t_\nu^{-1}[g(t_\nu q_k + t_\nu r) - g(t_\nu q_k)] = t_\nu^{-1}[g(p_\nu + t_\nu r) - g(p_\nu)].$$

Next select $k' > \nu$ and choose a related $v' > k'$ in the same way as above. Repeating this procedure, the subsequence of all $s \in \{t_\nu, t_{\nu'}, t_{\nu''}, ...\}$ then realizes, with the assigned $p(s) \in \{p_\nu, p_{\nu'}, p_{\nu''}, ...\}$, $\;v = \lim s^{-1}[g(p(s) + sr) - g(p(s))]$ and $p(s) \to 0$.

References

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

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

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

[4] J.M. Borwein, W.B. Moors, and W. Xianfy. *Lipschitz functions with prescribed derivatives and subderivatives.* CECM Information document 94-026, Simon Fraser Univ., Burnaby, 1994.

[5] J. M. Borwein and D. M. Zhuang. *Verifiable necessary and sufficient conditions for regularity of set-valued and single-valued maps. J. of Math. Analysis and Appl.*, 134: 441–459, 1988.

[6] J. V. Burke. *Calmness and exact penalization. SIAM J. Control Optim.*, 29: 493–497, 1991.

[7] J. V. Burke and S. Deng. *Weak sharp minima revisited, Part III: Error bounds for differentiable convex inclusions.* Manuscript 2006, to appear in *Mathematical Programming.*

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

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

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

[11] S. Dempe. Foundations of Bilevel Programming. Kluwer Academic Publ., Dordrecht-Boston-London, 2002.

[12] A. V. Dmitruk and A. Kruger. *Metric regularity and systems of generalized equations. J. of Math. Analysis and Appl.* 342: 864-873, 2008.

[13] S. Dolecki and S. Rolewicz. *Exact penalties for local minima. SIAM J. Control Optim.*, 17: 596–606, 1979.

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

[15] 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.

[16] I. Ekeland. *On the variational principle. J. of Math. Analysis and Appl.*, 47: 324–353, 1974.

[17] H. Federer. Geometric Measure Theory. Springer, New York, Heidelberg, 1969.

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

[19] H. Frankowska. *An open mapping principle for set-valued maps J. of Math. Analysis and Appl.*, 127: 172–180, 1987.

[20] 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.

[21] P. Fusek. *Ueber Kettenregeln in Gleichungsform fuer Ableitungen nichtglatter Funktionen.* Diplomarbeit, Fachbereich Mathematik. HU-Berlin, 1994

[22] P. Fusek. *Eigenschaften pseudo-regularer Funktionen und einige Anwendungen auf Optimierungsaufgaben.* Dissertation, Fachbereich Mathematik. HU-Berlin, Februar 1999.

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

[24] P. Fusek, D. Klatte, B. Kummer. *Examples and Counterexamples in Lipschitz Analysis. Control and Cybernetics* Vol. 31 (2002) No. 3, 471–492.

[25] B.R. Gelbaum , J.M.H. Olmsted. Counterexamples in Analysis. Holden-Day, San Francisco, London, Amsterdam, 1964.

[26] H. Gfrerer. *Hoelder continuity of solutions of perturbed optimization problems under Mangasarian-Fromovitz Constraint Qualification.* In Parametric Optimization and Related Topics, eds. J. Guddat, H.Th. Jongen, B. Kummer and F. Nožička, Akademie–Verlag, Berlin, 113–124, 1987.

[27] E.G. Golstein. Theory of Convex Programming. American Mathematical Society, Ser. Transactions of Mathematical Monographs 36, 1972, Providence, RI.

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

[29] R. Griesse, D.A. Lorenz. *A semismooth Newton method for Tikhonov functionals with sparsity constraints.* Inverse Problems, 24, 2008.

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

[31] J. Heerda and B. Kummer. *Characterization of Calmness for Banach space mappings. Preprint 06-26 of the Institut of Mathematics, Humboldt-University Berlin*, 2006 (http://www.mathematik.hu-berlin.de/publ/pre/2006/p-list-06.html)

[32] R. Henrion and J. Outrata. *A subdifferential condition for calmness of multifunctions. J. of Math. Analysis and Appl.*, 258: 110–130, 2001.

[33] R. Henrion and J. Outrata. *Calmness of constraint systems with applications. Mathematical Programming B*, 104: 437–464, 2005.

[34] M. Hintermueller, K. Ito, K. Kunisch. *The primal-dual active set strategy as a semi-smooth Newton method.* SIAM J. Optimiz., 13, 3: 865–888, 2003.

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

[36] A.D. Ioffe. *On regularity concepts in variational analysis Preprint: Department of Mathematics Technion, Haifa, Israel* 2006.

[37] A.D. Ioffe and J. Outrata *On metric and calmness qualification conditions in subdifferential calculus. Preprint: Department of Mathematics Technion, Haifa, Israel* 2006; J. Set-valued Anal. 16: 199-227, 2008.

[38] K. Jittorntrum. Solution point differentiability without strict complementarity in nonlinear programming. *Mathematical Programming Study*, 21:127–138, 1984.

[39] H.Th. Jongen, T. Möbert, and K. Tammer. On iterated minimization in nonconvex optimization. *Mathematics of Operations Research*, 11:679–691, 1986.

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

[41] D. Klatte. *On the stability of local and global optimal solutions in parametric problems of nonlinear programming.* Seminarbericht Nr. 80 der Sektion Mathematik, Humboldt-Universitaet Berlin., 1–21, 1985.

[42] 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.

[43] D. Klatte. *On quantitative stability for non-isolated minima.* Control and Cybernetics, 23, 183–200, 1994.

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

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

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

[47] D. Klatte and B. Kummer. *Stability of inclusions: Characterizations via suitable Lipschitz functions and algorithms. Optimization*, 55: Issue 5,6: 627-660, 2006.

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

[49] D. Klatte. *Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs. Mathematical Programming*, 88: 285–311, 2000.

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

[51] D. Klatte and B. Kummer. *Optimization methods and stability of inclusions in Banach spaces. Mathematical Programming, Series B*, Vol. 117: 305-330, 2009; online 2007 or as preprint HUB, 2006 under http://www.mathematik.hu-berlin.de/publ/pre/2006/p-list-06.html .

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

[53] 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.

[54] B. Kummer. *Globale Stabilitätsuntersuchungen für parametrische Optimierungsaufgaben.* Diss. B; HU-Berlin, 1977.

[55] 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.

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

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

[58] B. Kummer. *Inclusions in general spaces: Hoelder stability, Solution schemes and Ekeland's principle. J. of Math. Analysis and Appl.* 358: 327–344, 2009.

[59] D. N. Kutzarova and S. Rolewicz. *On drop property for convex sets. Archiv der Mathematik*, Vol 56, Nr. 5 : 501-511, 1991.

[60] A.B. Levy and R.T. Rockafellar. *Sensitivity analysis of solutions to generalized equations. Transactions of the American Mathematical Society*, 345: 661–671, 1994.

[61] A.B. Levy and R.T. Rockafellar. *Sensitivity analysis of solutions in nonlinear programming.* In D.-Z. Du, L. Qi and R.S. Womersley, editors, *Recent Advances in Nonsmooth Optimization*: 215–223. World Scientific Publ., 1995.

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

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

[64] E. Michael. *Continuous selections I. Annals of Math.*, 63:361–382, 1956.

[65] J.–S. Pang. *A degree–theoretic approach to parametric nonsmooth equations with multivalued perturbed solution sets. Mathematical Programming*, 62:359–383, 1993.

[66] W. Li. *Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. SIAM Journal on Optimization*, 7: 966-978, 1997.

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

[68] L. I. Minchenko. *Multivalued analysis and differential properties of multivalued mappings and marginal functions. J. of Mathem. Sciences*, 116, No. 3: 3266–3302, 2003.

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

[70] B.S. Mordukhovich. Variational Analysis and generalized differentiation. Vol. I: Basic Theory, Vol II: Applications *Springer, Berlin*, 2005.

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

[72] D. Pallaschke and R. Urbanski: Pairs of Compact Convex Sets, Kluwer Academic Publishers Dordrecht/Boston/London, 2002.

[73] J.-P. Penot. *Metric regularity, openness and Lipschitz behavior of multifunctions. Nonlinear Analysis*, 13: 629–643, 1989

[74] B. T. Polyak. *Convexity of nonlinear image of a small ball with applications to optimization* , *Set–Valued Analysis* 9 (1-2): 159–168, 2001.

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

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

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

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

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

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

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

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

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

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

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

[86] Philip B. Zwart. *Nonlinear programming: Counterexamples to two global optimization algorithms.* Operations Research. Vol. 21, No. 6, Nov. - Dec., 1973.

[87] N. D. Yen, J.-C. Yao and B. T. Kien. *Covering properties at positive order rates of multifunctions and some related topics. J. of Math. Analysis and Appl.* 338: 467–478, 2008.

Index