**Control and Cybernetics**

vol. **31** (2002) No. 3

# Examples and counterexamples in Lipschitz analysis

by

**Peter Fusek**[1], **Diethard Klatte**[2], **Bernd Kummer**[3]

[1] Institut für Operations Research, Universität Zürich,
Moussonstrasse 15, CH-8044 Zürich, Switzerland
e-mail: fusek@ior.unizh.ch

[2] Institut für Operations Research, Universität Zürich,
Moussonstrasse 15, CH-8044 Zürich, Switzerland
e-mail: klatte@ior.unizh.ch

[3] Institut für Mathematik, Humboldt–Universität zu Berlin,
Unter den Linden 6, D-10099 Berlin, Germany
e-mail: kummer@mathematik.hu-berlin.de

**Abstract:** In the analysis of functions and multi-valued mappings of Lipschitzian type, there are many different notions of Lipschitz behavior, regularity and generalized derivatives. We collect relevant examples illustrating the interrelations between various concepts, the differences with the smooth case, and the importance of certain assumptions and special classes of Lipschitz mappings in applications.
**Keywords:** Lipschitz (multi-)functions, regularity concepts, generalized derivatives, application to stationary points, pathological Lipschitz functions.

## 1. Introduction

This paper is concerned with typical examples and counterexamples in the analysis of Lipschitz functions and multifunctions. Our purpose is to contribute to a better understanding of interrelations between different concepts of Lipschitz behavior, regularity and derivatives, and of their role in selected applications.

First we recall some notations and the classical definition of a locally Lipschitz (single-valued) function. To do this, let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces. Given $X^0 \subset X$, the usual point-to-set distance of $x \in X$ to $X^0$ is defined by $\operatorname{dist}(x, X^0) = \inf_{x' \in X^0} d_X(x, x')$, where $\operatorname{dist}(x, \emptyset) = +\infty$. We write $X^0 + \varepsilon B_X := \{x \mid \operatorname{dist}(x, X^0) \leq \varepsilon\}$, i.e., the specialization to $X^0 = \{x^0\}$ means for $y^0 \in Y$, $Y^0 \subset Y$ are defined. Given $y^0 \in Y$, a function $s : Y \to X$ is said to be

(L0) *Lipschitz near* $y^0$ if there are positive $L$ and $\varepsilon$ such that

$$d_X(s(y), s(y')) \leq L d_Y(y, y') \quad \forall y, y' \in y^0 + \varepsilon B_Y. \tag{1.1}$$

The function $s$ is called *locally Lipschitz on* $Y$ if $s$ is Lipschitz near $y$ for each $y \in Y$. We will also use the (standard) notation $s \in C^{0,1}$, in this situation.

## 2. Lipschitzian concepts for multifunctions

Given again metric spaces $(X, d_X)$ and $(Y, d_Y)$, let now $S : Y \rightrightarrows X$ be a multivalued mapping (briefly called *multifunction*). Let $\operatorname{gph} S := \{(y, x) \in Y \times X \mid x \in S(y)\}$ denote the *graph* of $S$, let $\operatorname{dom} S = \{y \in Y \mid S(y) \neq \emptyset\}$ denote the *domain* of $S$, and let $S(A) = \cup_{a \in A} S(a)$ be the *image* of $A \subset Y$. If $\operatorname{gph} S$ is closed in the product space $Y \times X$, then $S$ is said to be *closed*. Further, denote by $S^{-1}$ the *inverse* of $S$, the multifunction defined by $S^{-1}(x) := \{y \in Y \mid (y, x) \in \operatorname{gph} S\}$.

Let $(y^0, x^0)$ be an element of $\operatorname{gph} S$, and let $\emptyset \neq X^0 \subset S(y^0)$. The multifunction $S$ is said to be

(L1) *locally upper Lipschitz at* $(y^0, X^0)$ if there are positive $L$ and $\varepsilon$ such that

$$S(y) \cap (X^0 + \varepsilon B_X) \subset X^0 + L d_Y(y, y^0) B_X \quad \forall y \in y^0 + \varepsilon B_Y; \tag{2.1}$$

(L2) *pseudo-Lipschitz at* $(y^0, x^0)$ if there are positive $L$ and $\varepsilon$ such that

$$S(y) \cap (x^0 + \varepsilon B_X) \subset S(y') + L d_Y(y, y') B_X \; \forall y, y' \in y^0 + \varepsilon B_Y; \tag{2.2}$$

(L3) *calm at* $(y^0, x^0)$ if there are positive $L$ and $\varepsilon$ such that

$$S(y) \cap (x^0 + \varepsilon B_X) \subset S(y^0) + L d_Y(y, y^0) B_X \quad \forall y \in y^0 + \varepsilon B_Y; \tag{2.3}$$

(L4) *Lipschitz l.s.c. at* $(y^0, x^0)$ if there are positive $L$ and $\varepsilon$ such that

$$\operatorname{dist}(x^0, S(y)) \leq L d_Y(y, y^0) \quad \forall y \in y^0 + \varepsilon B_Y. \tag{2.4}$$

Here "l.s.c." means "lower semicontinuous", while "u.s.c." will refer to "upper semicontinuous". Sometimes we use "L." and "u.L." to abbreviate "Lipschitz" and "upper Lipschitz", respectively. The multifunction $S$ is said to be

(L5) *Lipschitz l.s.c. at* $y^0$ if there are positive $L$ and $\varepsilon$ such that

$$S(y^0) \subset S(y) + L d_Y(y, y^0) B_X \quad \forall y \in y^0 + \varepsilon B_Y; \tag{2.5}$$

(L6) *Lipschitz u.s.c. at* $y^0$ if there are positive $L$ and $\varepsilon$ such that

If (L4) is weakened by only requiring that $\operatorname{dist}(x^0, S(y)) \to 0$ hold for each sequence $y \to y^0$, then $S$ is called *l.s.c. at* $(y^0, x^0)$. If (L5) is replaced by the weaker condition $\operatorname{dist}(x^0, S(y)) \to 0$ for each sequence $y \to y^0$ and each $x^0 \in S(y^0)$, one says that $S$ is (Hausdorff-) *l.s.c. at* $y^0$, while $\operatorname{dist}(x, S(y^0)) \to 0$ for each sequence $y \to y^0$ and each $x \in S(y)$ leads to the notion of $S$ being (Hausdorff-) *u.s.c. at* $y^0$. If $S$ is a function, the pseudo-Lipschitz property (L2) is nothing else than (L0), and property (L6) is also called *pointwise Lipschitz continuity at* $y^0$. In all cases considered above, one says that $L$ is a *modulus* of the related Lipschitz property.

The introduced names of Lipschitz properties are used in conformity with the book of Klatte and Kummer (2002). The notion of (L6) was defined by Robinson (1981) in the context of polyhedral multifunctions. (L1) was introduced for $X^0 = \{x^0\}$ and called *locally upper Lipschitz at* $(y^0, x^0)$ in Dontchev (1995), our form of (L1) is an extension of this notion. The inclusion (2.1) particularly yields that $S(y^0) \cap (X^0 + \varepsilon B_X) = X^0$, i.e., the set $X^0$ is necessarily an isolated component of the set $S(y^0)$. The property (L2)—which is also called *Aubin property* in the literature (see Rockafellar and Wets, 1998)—is a basic stability condition in Aubin and Ekeland (1984), and calmness (L3) has been applied and investigated e.g. in Clarke (1983), for deriving optimality conditions. With this respect, calmness can be used in a similar way as the upper Lipschitz property (L1), see Klatte and Kummer (2002). An interesting sufficient condition for calmness of multifunctions can be found in Henrion and Outrata (2001). It uses so-called semismoothness, Mifflin (1977), and may be applied to the models in Outrata (2000) and many models in Luo, Pang and Ralph (1996). The concepts (L5) and (L6) are mainly studied in the context of polyhedral multifunctions and convex-set-valued maps.

In the following proposition, we compile several elementary interrelations between these Lipschitzian concepts.

PROPOSITION 2.1 *Let $X$ and $Y$ be metric spaces, $S : Y \rightrightarrows X$ and $z^0 = (y^0, x^0) \in \operatorname{gph} S$. Then the following properties hold:*

*a. $S$ is locally upper Lipschitz at $z^0 \Rightarrow S$ is calm at $z^0$.*

*b. $S$ is pseudo-Lipschitz at $z^0 \Rightarrow S$ is Lipschitz l.s.c. and calm at $z^0$.*

*c. In the case $S(y^0) = \{x^0\}$, $S$ is locally upper Lipschitz at $z^0 \Leftrightarrow S$ is calm at $z^0$.*

*d. $S$ is Lipschitz u.s.c. at $y^0 \Rightarrow S$ is calm at $(y^0, x^0)$ $\forall x^0 \in S(y^0)$.*

*e. $S$ is Lipschitz l.s.c. at $y^0 \Rightarrow S$ is Lipschitz l.s.c. at $(y^0, x^0)$ $\forall x^0 \in S(y^0)$.*

*Proof.* Immediate by the definitions. ■

The definitions (L2), (L4) and (L5) imply that for $y$ near $y^0$, the sets $S(y)$ or $S(y) \cap (x^0 + \varepsilon B_X)$ are nonempty, while (L1), (L3) and (L6) do not imply this. For this reason, none of the properties (L1), (L3) and (L6) implies any of the remaining ones, consider the trivial example $S(y) = \{x \in \mathbb{R} \mid |x| = y\}$.

missing directions of the implications a, b, d and e fail to hold. The following counterexamples refer to this situation.

EXAMPLE 2.2 (pseudo-L., calm, L.l.s.c. and L.u.s.c., but not locally u.L.). Consider $S(y) \equiv \mathbb{R}$, $y \in \mathbb{R}$, and let $B = [-1, 1]$. Trivially, $S$ is constant, hence it is Lipschitz l.s.c. and Lipschitz u.s.c. at each $y^0$, and for any $(y^0, x^0) \in \mathbb{R}^2$, $S$ is pseudo-Lipschitz and calm at $(y^0, x^0)$. However, for each $\varepsilon > 0$ and each $L > 0$, the point $y = y(\varepsilon, L) := y^0 + \varepsilon/(2L)$ satisfies

$$\begin{aligned} S(y) \cap (x^0 + \varepsilon B) &= [x^0 - \varepsilon, x^0 + \varepsilon] \not\subset [x^0 - \varepsilon/2\ ,\ x^0 + \varepsilon/2] \\ &= x^0 + L|y - y^0|B, \end{aligned}$$

i.e., for each $(y^0, x^0) \in \mathbb{R}^2$, $S$ is not locally upper Lipschitz at $(y^0, x^0)$.

EXAMPLE 2.3 (pseudo-L., calm and L.l.s.c., but not locally u.L. and not L.u.s.c.). Let $s(y) = 1 + \sqrt{|y|}$ and $S(y)$ be the interval $[-s(y), s(y)]$ for real $y$.

Then, if $x^0 \in S(0)$, the mapping $S$ is not locally upper Lipschitz at $(0, x^0)$, since for each $\varepsilon > 0$ and each $L > 0$, one finds points $x(y) \in S(y) \cap (x^0 + \varepsilon B)$ such that $|x(y) - x^0| > L|y|$ and $|y| < 1/L$ . Further, $\operatorname{dist}(s(y), S(0)) = \sqrt{|y|}$ for $y \neq 0$, i.e., $S$ is not Lipschitz u.s.c. at $y^0 = 0$.

On the other hand, $S$ is pseudo-Lipschitz (hence also calm) at each point $(0, x^0)$, $x^0 \in \operatorname{int} S(0)$. Note that $S$ is not calm at $(0, 1)$. Further, $S(0) \subset S(y)$ for $y \neq 0$ implies that $S$ is Lipschitz l.s.c. at $y^0 = 0$.

EXAMPLE 2.4 (L.u.s.c. at $y^0$ and L.l.s.c. at $(y^0, x^0)$, but not l.s.c. at $y^0$ and not pseudo-L. at $(y^0, x^0)$). Assign to each $x \in \mathbb{R}^n$ the line segment $F(x) = [0, x]$ (i.e., the convex hull of 0 and $x$), then the inverse $S(y) := F^{-1}(y)$ becomes

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

Obviously, $F^{-1}$ is Lipschitz u.s.c. with each $L$ at $y^0 = 0$ as well as Lipschitz l.s.c. with $L = 1$ at the origin $(y^0, x^0) = (0, 0)$.

However, $F^{-1}$ is not pseudo-Lipschitz at $(y^0, x^0) = (0, 0)$ and not l.s.c. at $y^0 = 0$. This can be seen as follows. Given $\varepsilon > 0$, let $\|x\| = \varepsilon$. Then $x \in S(0) \cap (0 + \varepsilon B_X)$, and for all $y$ of the form $y = -\lambda x$, $\lambda \geq 0$, it follows that $\operatorname{dist}(x, S(y)) = \varepsilon + \lambda\varepsilon \geq \varepsilon$. Hence, neither (L2) nor the l.s.c. condition at $y^0 = 0$ can be satisfied.

Note that a slight modification of the mapping $F$ leads to a regular situation. Now assign to each $x \in \mathbb{R}^n$, the line segment $F(x) = [\frac{1}{2}x, x] \subset \mathbb{R}^n$. The inverse multifunction is now

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

and $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is pseudo-Lipschitz at $(0, 0)$ as well as Lipschitz l.s.c. and

The same situation as in Example 2.4 can be found in the context of the constraint set mapping of a standard nonlinear optimization problem, see the next example, where the so-called *Mangasarian–Fromovitz condition* (MFCQ), Mangasarian and Fromovitz (1967), is violated. Recall that a finite-dimensional constraint map

$$S(y,z) = \{x \mid g(x) \geq y,\ \ h(x) = z\}, \tag{2.7}$$

for $(g,h) \in C^1(\mathbb{R}^n, \mathbb{R}^{m+k})$, is pseudo-Lipschitz at $(y^0, z^0, x^0)$ if and only if

$$\text{(MFCQ)} \qquad \begin{array}{l} Dh(x^0) \text{ has full rank, and there is some } u \text{ such that} \\ Dh(x^0)u = 0 \text{ and } g(x^0) + Dg(x^0)u > y^0 \end{array}$$

is satisfied, which was shown for the first time in Robinson (1976).

Example 2.5 (L.u.s.c. at $y^0$ and L.l.s.c. at $(y^0, x^0)$, but MFCQ violated). Consider the map

$$S(y) := \{x \in \mathbb{R}^2 \mid x_2(x_2 - x_1^2) \geq 0,\ x_2 = y\}, \quad y \in \mathbb{R}.$$

Obviously,

$$S(y) = \{(x_1, y) \mid -\sqrt{y} \leq x_1 \leq \sqrt{y}\}\ \ \forall y > 0$$

and

$$S(y) = \{(x_1, x_2) \mid x_2 = y\}\ \ \forall y \leq 0,$$

in particular, $\operatorname{dist}(x, S(0)) = |y|$ for all $y \neq 0$ and all $x \in S(y)$. So, $S$ is Lipschitz u.s.c. at $y^0 = 0$ and hence calm at each $(0, x^0)$, $x^0 \in S(0)$. Further, $\operatorname{dist}(0, S(y)) = |y|$ for all $y \neq 0$, hence, $S$ is also Lipschitz l.s.c. at $(0,0)$.

On the other hand, for any $y > 0$, one has $(\sqrt{y}, y) \in S(y)$ and

$$\operatorname{dist}((\sqrt{y}, y), S(y/4)) = \|(\sqrt{y}, y) - (\sqrt{y}/2, y/4)\| = \sqrt{y/4 + 9y^2/16},$$

hence, (L2) is violated in $(0,0)$. It is easy to see that MFCQ does not hold at this point.

It is known that for a calm equality constraint $h(x) = 0$ of a nonlinear program (this means that $S = h^{-1}$ is calm at $(x^0, 0)$), a local minimizer $x^0$ of $f$ with respect to this constraint is necessarily a local minimizer of an unconstrained penalty-type function $F(x) = f(x) + \alpha\|h(x)\|$ for suitable $\alpha$, see, e.g., Clarke (1983). Similar results are true if the constraints satisfy a local upper Lipschitz or a pseudo-Lipschitz condition, see Klatte and Kummer (2002). The next example indicates that this (standard) penalization may yield a terrible unconstrained auxiliary function $F$, though the given equation satisfies "nice"

EXAMPLE 2.6 (the inverse of Dirichlet's function). For the real function

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

the inverse $h^{-1}$ is calm at $(y^0, x^0) = (0,0)$ and locally upper Lipschitz at $(0, h^{-1}(0))$ since $h^{-1}(y) = \emptyset$ for $y \neq 0$, $y$ near 0. These useful properties do not prevent that the terrible behavior of the constraint function $h$ being carried over to the auxiliary function $F(\cdot) = f(\cdot) + \alpha|h(\cdot)|$.

Note that the mapping $S(y) = \{x \mid h(x) \geq y\}$ is even pseudo-Lipschitz at $(0,0)$ since $h(x) \geq y$ holds for all irrational $x$ and all $y$ near 0.

## 3. Characterizations of regularity

Throughout this section, let $X$ and $Y$ be normed spaces (if not specified otherwise), though the regularity concepts make sense also for more general spaces. Let $S = F^{-1}$ be the inverse of a given multifunction $F : X \rightrightarrows Y$. In the following, we recall several regularity notions for $F$, where in general we will speak of *regularity* (of $F$) whenever $F^{-1}(y)$ satisfies a certain Lipschitz property. The *type* of regularity (*strong, pseudo, upper*) differs by the related Lipschitz properties of $F^{-1}$, where in the case of upper regularity we additionally suppose that $F^{-1}(y)$ is non-empty for $y$ near $y^0 \in F(x^0)$.

### Regularity Notions

(R1) If $S$ is pseudo-Lipschitz at $(y^0, x^0)$, then $F$ is called *pseudo-regular* at $(x^0, y^0)$.

(R2) If, additionally, neighborhoods $U$ and $V$ of $x^0$ and $y^0$, respectively, exist in such a way that $U \cap F^{-1}(y)$ is single-valued for $y \in V$, then we call $F$ *strongly regular* at $(x^0, y^0)$.

(R3) If $S$ is locally upper Lipschitz at $(y^0, x^0)$ *and* $S(y') \cap U$ is non-empty for all $y' \in V$ (for certain neighborhoods $U$ and $V$ of $x^0$ and $y^0$, respectively), then $F$ is said to be *upper regular* at $(x^0, y^0)$.

First we mention some typical examples for the defined regularity notions.

EXAMPLE 3.1 (regularity for $C^1$ functions). If $F : X = \mathbb{R}^n \to Y = \mathbb{R}^n$ is a continuously differentiable function, then all these regularity definitions coincide - due to usual implicit function theorem - with the requirement det $DF(x^0) \neq 0$.

EXAMPLE 3.2 Kummer (1997) (pseudo-regular, but not strongly regular). The complex function $F(z) = z^2/|z|$ for $z \neq 0$, $F(0) = 0$, is a Lipschitz function which is pseudo-regular and upper regular without being strongly regular at the origin.

EXAMPLE 3.3 (strong regularity for continuous functions). For a *continuous* function $F : \mathbb{R}^n \to \mathbb{R}^m$, strong regularity at $(x^0, y^0)$ induces that $F$ is a homeonecessarily true due to Brouwer's famous invariance of domain theorem. This is an essential fact being valid for functions, but not for (Lipschitz) continuous multifunctions (take $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{n+1}$ as $F(x) = \{(x, y) \mid y \in \mathbb{R}\}$).

EXAMPLE 3.4 (pseudo-regularity for linear operators). Let $F : X \to Y$ be a *linear operator* onto $Y$, where $X$ and $Y$ are normed spaces. Pseudo-regularity now requires that, given $y', x$ and $y = F(x)$, there is some $x'$ such that $F(x') = y'$ and $\|x' - x\| \leq L\|y' - y\|$. In other words, $F^{-1}$ is bounded as a mapping in the factor space $X/F^{-1}(0)$. Conversely, one may say that pseudo-regularity is just a nonlinear, local version of this property.

EXAMPLE 3.5 (subdifferential of the Euclidean norm). An interesting and relevant example of a *multifunction* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ being strongly regular at $(0, 0)$ is the following one. Consider the subdifferential (in the sense of convex analysis) $F(x) = \partial f(x)$ of the Euclidean norm $f(x) = \|x\|$: Then,

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

**Some generalized derivatives of multifunctions**

For normed spaces $X$ and $Y$ and $(x, y) \in \operatorname{gph} F$, the above regularity concepts are related to certain concepts of *generalized (directional) derivatives.* We associate with $F$ the following multifunctions:

(D1) $CF(x, y) : X \rightrightarrows Y$, defined by $v \in CF(x, y)(u)$ if there are certain (discrete) $t = t_k \downarrow 0$ and assigned elements $(u_t, v_t) \to (u, v)$ such that $y + tv_t \in F(x + tu_t)$.

(D2) $TF(x, y) : X \rightrightarrows Y$, defined by $v \in TF(x, y)(u)$ if there are certain (discrete) $t = t_k \downarrow 0$, assigned points $(x_t, y_t) \in \operatorname{gph} F$ with $(x_t, y_t) \to (x, y)$ and elements $(u_t, v_t) \to (u, v)$ such that $y_t + tv_t \in F(x_t + tu_t)$.

(D3) $D^*F(x, y) : Y^* \rightrightarrows X^*$, defined by $u^* \in D^*F(x, y)(v^*)$ if there are certain (discrete) $t = t_k \downarrow 0$, $r_t > 0$, assigned points $(x_t, y_t) \to (x, y)$ in $\operatorname{gph} F$ and dual elements $(u_t^*, v_t^*) \rightharpoonup^* (u^*, v^*)$ in $X^* \times Y^*$ such that $\langle u_t^*, \xi\rangle + \langle v_t^*, \eta\rangle \leq t\|\xi\|_X + t\|\eta\|_Y$ if $\|\xi\|_X + \|\eta\|_Y < r_t$ and $(x_t + \xi, y_t + \eta) \in \operatorname{gph} F$,

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

The mapping $CF(x, y)$ is the *contingent derivative*, Aubin and Ekeland (1984), also called *graphical derivative* or *Bouligand derivative*, while $D^*F(x, y)$ is (up to a sign) the *coderivative* in the sense of Mordukhovich (1993). $TF(x, y)$ was defined in Rockafellar and Wets (1998) and was called *strict graphical derivative* there. To be consistent with the terminology of the book by Klatte and Kummer (2002), we use the name *Thibault's limit set* (or *Thibault derivative*) for $TF(x, y)$. Note that this derivative has been first considered (however, For each of these generalized derivatives, the symmetric definitions induce that the inverse of the derivative is just the derivative of the inverse at corresponding points. As usually, we will say that a derivative is *injective* if the origin belongs only to the image of $u = 0$ or $v^* = 0$, respectively.

For *functions* $F$, we have $y = F(x)$ and may write $CF(x)$, $TF(x)$ and $D^*F(x)$. Nevertheless, the images of the derivatives as well as the pre-images $F^{-1}(y)$ may be multi-valued or empty. If the (one-sided) limit $\lim_{t\downarrow 0} t^{-1}(F(x+tu) - F(x))$ exists uniquely for a function $F$ and all sequences $t \downarrow 0$, then it is called the *directional derivative* of $F$ at $x$ in direction $u$, and denoted by $F'(x;u)$. Further, for $F : X \to \mathbb{R}$, *Clarke's directional derivative* of $F$ at $x^0$ in direction $u \in X$ is defined by the usual limes superior $F^c(x^0;u) = \limsup_{t\downarrow 0,\, x\to x^0} t^{-1}(F(x+tu) - F(x))$ which is obviously finite for locally Lipschitz functions.

If $f$ is a locally Lipschitz function from $\mathbb{R}^n$ to $\mathbb{R}^m$, then, by Rademacher's theorem, the set

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

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

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

has non-empty images. In addition, one easily sees that $\partial_\circ f$ is closed and locally bounded. The same properties are induced for the map $\partial f$, defined by

(D4) Clarke's (1983) *generalized Jacobian* $\partial f(x) = \operatorname{conv} \partial_\circ f(x)$ of $f$ at $x$, where $\operatorname{conv} Z$ means the convex hull of a set $Z$.

Note that

$$Tf(x)(u) \subset \partial f(x)u \tag{3.1}$$

and

$$\partial f(x)u = \operatorname{conv}(Tf(x)(u)) \tag{3.2}$$

hold, see Kummer (1991). The inclusion (3.1) may be strict even for piecewise linear functions $f : \mathbb{R}^2 \to \mathbb{R}^2$, see Example 3.9 below.

## Regularity characterizations

For the purposes of the present paper, we essentially restrict the following summary of interrelations between regularity notions and suitable properties of generalized derivatives to the case $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.

PROPOSITION 3.6 (regularity of multifunctions, summary). *Let* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be closed and* $z^0 = (x^0, y^0) \in \operatorname{gph} F$. *Then:*

$$\begin{aligned} &F \text{ is upper regular at } z^0 \\ &\Leftrightarrow\ CF(z^0) \text{ is injective and } F^{-1} \text{ is Lipschitz l.s.c. at } (y^0, x^0). \end{aligned} \tag{3.3}$$

$$\begin{aligned} &F \text{ is strongly regular at } z^0 \\ &\Leftrightarrow\ TF(z^0) \text{ is injective and } F^{-1} \text{ is Lipschitz l.s.c. at } (y^0, x^0). \end{aligned} \tag{3.4}$$

$$\begin{aligned} &F \text{ is pseudo-regular at } z^0 \\ &\Leftrightarrow\ \exists \varepsilon > 0 : \varepsilon B \subset CF(z)(B) \text{ for all } z \in \operatorname{gph} F \cap (z^0 + \varepsilon B) \\ &\Leftrightarrow\ D^* F(z^0) \text{ is injective.} \end{aligned} \tag{3.5}$$

$$\begin{aligned} &\text{If } F^{-1} \text{ is Lipschitz l.s.c. at } (y^0, x^0) \\ &\text{then there exists } r > 0 \text{ such that } B \subset CF(z^0)(rB). \end{aligned} \tag{3.6}$$

*If* $X$ *is a normed space, the conditions (3.3) and (3.4) are still necessary for the related regularity.*

The preceding proposition can be found as Theorem 5.1 in Klatte and Kummer (2002) and summarizes several results proved in Chapter 3 of that book. The characterization (3.3) was given in King and Rockafellar (1992). One of the referees pointed out that a prototype of condition (3.3) appeared first (without naming the property in question) in Rockafellar (1989). Condition (3.4) appears in Rockafellar and Wets (1998), Chapter 9 (in a different terminology). The "if"-direction of the first characterization in (3.5) goes back to Aubin and Ekeland (1984), for the "only if"-direction and for the implication (3.6) see, e.g., Kummer (2000), while the second characterization of (3.5) was shown in Mordukhovich (1993).

In the following, we give some examples which illustrate that crucial assumptions of the previous proposition may not be omitted.

The first example is taken from Kummer (2000) and shows that the conditions (given in (3.5)) for pseudo-regularity of $F$ in terms of the contingent derivative $CF$ and coderivative $D^*F$, respectively, are *not necessary* if $X$ is a Hilbert space, see Example 3.7.

The second example shows that the l.s.c. condition under (3.3) and (3.4) is, in general, not ensured by the already imposed injectivity of $CF$ and $TF$, respectively, see Example 3.8.

Further, Example 3.9 (given in Kummer, 1991) concerns strong regularity of a locally Lipschitzian function $F$ from $\mathbb{R}^n$ to $\mathbb{R}^n$: For such functions, without supposing l.s.c. in (3.4), the injectivity of $TF$ is a sufficient and necessary condition for strong regularity while the injectivity of the map $u \to \partial F(x^0)(u)$ (i.e., all matrices in $\partial F(x^0)$ are regular) is only a sufficient one, see Clarke (1976), Kummer (1991).

Finally, Example 3.10 will demonstrate that the (pointwise) condition (3.6)

EXAMPLE 3.7 (pseudo-regular, but the conditions (3.5) fail to hold). We give a function $f$ which is one of the simplest nonsmooth, nonconvex functions on a Hilbert space such that the following is true:

> Pseudo-regularity of the map $F(x) = \{y \in \mathbb{R} \mid f(x) \leq y\}$ holds. However, the conditions (3.5) in terms of contingent derivatives and coderivatives will not be satisfied.

Let

$$X = l^2, \; x = (x_1, x_2, \dots) \text{ and } f(x) = \inf_k x_k.$$

Now $F^{-1}(y) = \{x \in X \mid f(x) \leq y\}$ is the level set map of a globally Lipschitz functional. Since $f$ is concave the directional derivatives $f'(x; u)$ exist everywhere and are the only elements of $Cf(x, u)$. Further, $f$ is monotone with respect to the natural vector ordering, and $f$ is nowhere positive.

The mapping $F$ is (globally) pseudo-regular, e.g., with modulus $L = 2$.

Indeed, if $f(x) \leq y$ and $y' < y$, there is some $k$ such that $x_k < y + \frac{1}{2}|y' - y|$. Next, put $x' = x - 2|y' - y|e^k$ where $e^k$ is the $k$-th unit vector in $l^2$. Then, pseudo-regularity follows from $\|x' - x\| \leq 2|y' - y|$ and $x' \in F^{-1}(y')$ since $f(x') \leq x'_k \leq y - \frac{3}{2}|y' - y| \leq y'$.

In order to see that $0 \in D^*F(0,0)(-1)$, we refer to Kummer (2000) or Klatte and Kummer (2002), Example BE.2, for details. Here, we only note that the consideration of the points $(u^*, v^*) = (e^m, -1) \rightharpoonup^* (0, -1)$ and the assigned elements $(x^m, y^m) = (-e^m, -1)/m \in \operatorname{gph} F$ leads to the desired result as $m \to \infty$.

The sufficient condition in terms of $CF$ does not hold because of the property $f'(x, u) \geq 0 \; \forall u \in l^2$, which is valid at all $x$ satisfying

$$f(x) < x_k \; \forall k. \tag{3.7}$$

In fact, having such $x$, it follows $f(x) = 0$ immediately. Now assume that some of these directional derivatives are negative, i.e., let the inequality $f(x + tu) < f(x) - t\delta = -t\delta$ hold for some fixed $\delta > 0$ and sufficiently small $t > 0$. Then, the infimum $f(x + tu)$ must be attained at some component $k = k(t)$:

$$f(x + tu) = x_{k(t)} + tu_{k(t)} < -t\delta. \tag{3.8}$$

If $k(t) < K_0$ remains bounded (as $t \to 0$), then we have $x_{k(t)} \geq \mu$ with some positive $\mu$, and (3.8) implies the contradiction $u_{k(t)} \to -\infty$. Otherwise, the inequality $u_{k(t)} < -\delta$ follows for an infinite number of components of $u$ whereafter $u$ cannot belong to $l^2$. Since points $x$ having property (3.7) exist with arbitrarily small norm, the condition

$$\varepsilon B \subset CF(z)(B) \text{ for all } z \in \operatorname{gph} F \cap (z^0 + \varepsilon B)$$

Example 3.8 ($TF$ or $CF$ injective, but $F^{-1}$ not l.s.c.). First consider the real function $f(x) = |x|$. One has $Cf(0)(u) = |u|\ \forall u$, i.e., $Cf(0)$ is injective. On the other hand, $f^{-1}(y) := \{x \mid |x| = y\} = \emptyset$ if $y < 0$, i.e., $f^{-1}$ is not l.s.c. at $(0,0)$.

Next consider $F : \mathbb{R} \to \mathbb{R}^2$ defined by $F(x) := (x,x)$ for all $x \in \mathbb{R}$. Thus

$$TF(0)(u) = CF(0)(u) = \{DF(0)u\} = \left\{ \begin{pmatrix} u \\ u \end{pmatrix} \right\}.$$

Hence, $TF(0)$ and $CF(0)$ are injective, but

$$F^{-1}(y_1, y_2) := \{x \in \mathbb{R} | (y_1, y_2) = (x,x)\} = \emptyset \text{ if } y_1 \neq y_2,$$

i.e., $F^{-1}$ is not l.s.c. at $((0,0),0)$.

Example 3.9 (piecewise linear bijection of $\mathbb{R}^2$ with $0 \in \partial f(0)$). On the sphere of $\mathbb{R}^2$, let vectors $a^k$ and $b^k$ ($k = 1, 2, \ldots, 6$) be arranged as follows (we put $a^7 = a^1, b^7 = b^1$ in order to simplify the notation):

(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 pointed (the angle between the vectors is smaller than $\pi$).

Let $L_i : \mathbb{R}^2 \to \mathbb{R}^2$ be the unique linear function satisfying $L_i(a^i) = b^i$ and $L_i(a^{i+1}) = b^{i+1}$. By setting $f(x) = L_i(x)$ if $x \in K_i$ we define a piecewise linear function which maps $K_i$ onto $P_i$. By the construction, $f$ is surjective and has a well-defined inverse; hence it is a (piecewise linear) Lipschitzian homeomorphism of $\mathbb{R}^2$. Moreover, $f = id$ (:= identity) on $\operatorname{int} K_1$ and $f = -id$ on $\operatorname{int} K_4$.

Thus, $\partial f(0)$ contains the unit-matrix $E$ as well as $-E$ and, by convexity, the zero-matrix, too.

Example 3.10 (counterexample ($n = m = 2$) showing that the pointwise condition (3.6) is not sufficient for the Lipschitz l.s.c. of $F^{-1}$). We construct $f : \mathbb{R}^2 \to \mathbb{R}^2$ continuous with

$$f'(0; u) = u \quad \forall u \in \mathbb{R}^2 \quad \text{and} \quad 0 \notin \operatorname{int} f(\mathbb{R}^2).$$

Let

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

and $G = \operatorname{conv} M$. For $(x,y) \in M$ , let $f(x,y) = (x,y)$. For $(x,y) \in G \backslash M$ and $y \geq 0$ put $f(x,y) = (x, x^2)$.

In order to define $f$ at $(x,y) \in G \backslash M$ with $0 > y > -x^2$ , let $D$ be the triangle given by the points

Then $t \in (0,1)$. We shift the point $(x, t(-x^2) + (1-t)x^2)$ to the left boundary of $D$ and define $f$ to be the related point:

$$f(x,y) = \begin{cases} (2t-1)(x, -x^2) & \text{for } t \geq \frac{1}{2}, \\ (1-2t)(x, x^2) & \text{for } t \leq \frac{1}{2}. \end{cases}$$

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

**Pseudo- and strong regularity of stationary points**

Given a stationary point $x^0$ of a function $f \in C^2(\mathbb{R}^n, \mathbb{R})$, the above characterizations for different types of regularity immediately imply that pseudo-regularity and strong regularity of the function $Df$ at $(x^0, 0)$ coincide, i.e., if the stationary point map $S = [Df]^{-1}$ is pseudo-Lipschitz at $(0, x^0)$, then it is locally single-valued. The following example shows that this property does not carry over to $C^1$ functions, which are piecewise $C^2$.

It is worth noting that the equivalence of strong regularity and pseudo-regularity still holds in the context of the constrained $C^2$ optimization problems; for results of this type see, e.g., Dontchev and Rockafellar (1996), Klatte and Kummer (2002), Kummer (1997).

EXAMPLE 3.11 (a piecewise quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ having pseudo-Lipschitzian stationary points being not unique). We put $z = (x, y) \in \mathbb{R}^2$ in polar-coordinates,

$$z = r(\cos\phi + i \sin\phi),$$

and describe $f$ as well as the partial derivatives $D_x f, D_y f$ over eight cones

$$C(k) = \left\{ z \mid \phi \in \left[\frac{1}{4}(k-1)\pi, \frac{1}{4}k\pi\right]\right\}, \ (1 \leq k \leq 8),$$

by

| *cone* | $f$ | $D_x f$ | $D_y f$ |
|---|---|---|---|
| $C(1)$ | $y(y-x)$ | $-y$ | $2y-x$ |
| $C(2)$ | $x(y-x)$ | $-2x+y$ | $x$ |
| $C(3)$ | $x(y+x)$ | $+2x+y$ | $x$ |
| $C(4)$ | $-y(y+x)$ | $-y$ | $-2y-x$ |

and on the remaining cones $C(k+4)$, $(1 \leq k \leq 4)$, $f$ being defined as in $C(k)$.

Upon 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

REMARK 3.12 (pseudo-regularity and isolated zeros of $F \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$). The function $F = Df : \mathbb{R}^2 \to \mathbb{R}^2$ of Example 3.11 had the same topological properties as $F$ in Example 3.2 (pseudo-regular, upper regular, but not strongly regular at $(0,0)$). It is remarkable that, under pseudo-regularity at $(0,0)$, the related upper regularity of a directionally differentiable function $F \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is an immediate consequence, because the origin is necessarily isolated in $F^{-1}(0)$, see Fusek (2001). However, for $F \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ without directional derivatives at the origin, nothing can be said up to now concerning this implication.

## 4. Generalized derivatives of Lipschitz functions

In this section, we consider locally Lipschitz functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. If $f$ is such a function with modulus $L$ near some $x$, then the Thibault derivative $Tf$ and the contingent derivative $Cf$ take on the particular forms

$$Tf(x)(u) = \left\{ v \;\middle|\; \begin{array}{l} v = \lim t_k^{-1}[f(x^k + t_k u) - f(x^k)] \\ \text{for certain } t_k \downarrow 0 \text{ and } x^k \to x \end{array} \right\}, \tag{4.1}$$

$$Cf(x)(u) = \left\{ v \;\middle|\; \begin{array}{l} v = \lim t_k^{-1}[f(x + t_k u) - f(x)] \\ \text{for certain } t_k \downarrow 0 \end{array} \right\}. \tag{4.2}$$

These sets are non-empty, closed and bounded ($\subset L\|u\|B$). $Tf(x)(u)$ and $Cf(x)(u)$ are connected sets, and both mappings are Lipschitz in $u$. For $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, there is $Cf = Tf = \{Df\}$. For the absolute value $f(x) = |x|$ we observe that $Cf(0)(u) = \{f'(0; u)\}$ (the usual directional derivative), and $Tf(0)(u) = [-|u|, |u|]$ (a closed interval). So $Cf$ and $Tf$ are different even for elementary functions.

The following example gives a Lipschitz function $f$ having images in an infinite-dimensional space with empty contingent derivatives for nontrivial directions (and without directional derivatives).

EXAMPLE 4.1 (a Lipschitz function $f : [0, \frac{1}{2}) \to C[0,1]$ such that directional derivatives $f'$ nowhere exist, neither as strong nor weak (pointwise) limits; and contingent derivatives are empty). For $x \in [0, \frac{1}{2})$ define a continuous function $h_x : [0,1] \to \mathbb{R}$ by

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

The mapping $f(x) := h_x$ is a Lipschitz function from the interval $[0, \frac{1}{2})$ into $C[0,1]$. For small $|\lambda| > 0$, consider the function

$$g(x, \lambda) = (f(x + \lambda) - f(x))/\lambda.$$

If $\lambda > 0$, then

Hence, the limit $\lim g(x,\lambda)$ (as $\lambda \downarrow 0$) cannot exist in $C[0,1]$ (neither in a strong nor in a weak sense). If $\lambda < 0$, then we obtain for $x > 0$ that

$$g(x,\lambda)(2x) \geq 0 \text{ and } g(x,\lambda)(2x+2\lambda) = -1.$$

Thus $\lim g(x,\lambda)$ (as $\lambda \uparrow 0$) cannot exist, too.

**Chain rules and "simple" Lipschitz functions**

The following chain rules are the key for many applications of Lipschitz analysis, for example, in the study of regularity of the Karush–Kuhn–Tucker system of a nonlinear program. We consider

$$f(x,y) = h(x,g(y)); \quad h: \mathbb{R}^{n+q} \to \mathbb{R}^p,\ g: \mathbb{R}^m \to \mathbb{R}^q,\ f: \mathbb{R}^{n+m} \to \mathbb{R}^p.$$

Here, $h$ and $g$ are supposed to be locally Lipschitz. We are interested in the formula

$$Tf(x,y)(u,v) = T_x h(x,g(y))(u) + T_g h(x,g(y))(Tg(y)(v)), \tag{4.3}$$

where $T_x h$ and $T_g h$ denote the *partial $T$-derivatives*, defined - as usually - by fixing the remaining arguments. In general, (4.3) is not true, we need a special property of $g$. According to Kummer (1991), we say that a locally Lipschitz function $g: \mathbb{R}^m \to \mathbb{R}^q$ is *simple* at $y$ if, for all $v \in \mathbb{R}^m$, $w \in Tg(y)(v)$ and each sequence $t_k \downarrow 0$, there is a sequence $y^k \to y$ such that

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

A similar requirement for double limits (but in the context of contingent derivatives for multifunctions) is involved in the definition of proto-derivatives, see Levy and Rockafellar (1996).

Note that all $g \in C^{0,1}(\mathbb{R}^m, \mathbb{R})$ are simple, further simple functions are $y \mapsto y^+$ and $y \mapsto (y^+, y^-)$ (which are of particular interest for the Kojima's form of the Karush–Kuhn–Tucker conditions, Klatte and Kummer, 1999, Kojima, 1980), for the proofs see Klatte and Kummer (2002), Kummer (1991). The following proposition shows the importance of simple functions in the chain rule under consideration.

PROPOSITION 4.2 (partial derivatives for $Tf$). *Let $g$ and $h$ be locally Lipschitz, $f = h(x,g(y))$, and let $D_g h(\cdot,\cdot)$ exist and be locally Lipschitz, too. Then*

$$Tf(x,y)(u,v) \subset T_x h(x,g(y))(u) + T_g h(x,g(y))(Tg(y)(v)).$$

*Let, additionally, $g$ be simple at $y$. Then the equation (4.3) holds true.*

NOTE. Clearly, $T_g h = \{D_g h\}$.

For the proof, we refer to Kummer (1991) or Klatte and Kummer (2002), Theorem 6.8. It is remarkable that neither all functions in $C^{0,1}(\mathbb{R}, \mathbb{R}^2)$ nor all

EXAMPLE 4.3 ($f \in C^{0,1}(\mathbb{R}, \mathbb{R}^2)$ not simple). Put $a_0 = 1$ and consider for $k \in \mathbb{N}$ the points

$$a_k = 2^{-k}, \quad b_k = \frac{9}{8}\,2^{-k}, \quad c_k = \frac{1}{2}(a_k + b_k) = \frac{17}{16}\,2^{-k},$$
$$d_k = \frac{15}{8}\,2^{-k}, \quad e_k = \frac{1}{2}(d_k + a_{k-1}) = \frac{31}{16}\,2^{-k}.$$

Let the function $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f = (f_1, f_2)^T$,

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

and

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

where $k \in \mathbb{N}$. The function $f$ is locally Lipschitz everywhere with the modulus $L = 1$. Considering the direction $u = 1$ and the sequences $x^k = c_k$, $t_k = e_k - c_k$, we obtain

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

Now let the sequence $r_k \downarrow 0$ be given. Our goal is to find points $y^k$, $y^k \to 0$ 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 $k$ sufficiently large 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 limit $(-\frac{1}{14}, \frac{1}{14})^T$, only the subsequences with $\ell(k) = n(k)$ are suitable. As a consequence we have $r_k \geq d_{n(k)} - b_{n(k)} = \frac{3}{4}\,2^{-n(k)}$ and $r_k \leq a_{n(k)-1} - a_{n(k)} = 2^{-n(k)}$.

In other words, for every sequence $\{r_k\}$ with $2^{-(k+1)} < r_k < \frac{3}{4}\,2^{-k}$, $k \in \mathbb{N}$ it is impossible to find a suitable sequence of indices $\{n(k)\}$. Hence, $f$ is not

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

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

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 $I_k(g) \cap I_\nu(f) = \emptyset \ \forall k, \nu$. Now, with the definition

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

the point $(v(f), v(g))$ cannot belong to $Th(0,0)(1,1)$, and the chain rule (4.3) fails to hold even for a sum of functions.

EXAMPLE 4.4 ($f \in PC^1$ not simple). Put for $k \in \mathbb{N}$

$$a_k = \left(2^k\pi + \frac{\pi}{2}\right)^{-1}, \qquad b_k = \left(2^k\pi - \frac{\pi}{2}\right)^{-1}.$$

We define a piecewise differentiable function $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f = (f_1, f_2)^T$,

$$f_1(x,y) = \begin{cases} g_1(x,y) & \text{if } x \geq 0,\ g_1(x,y) \geq 0,\ y \in [a_k, b_k], \\ g_2(x,y) & \text{if } x \leq 0,\ g_2(x,y) \geq 0,\ y \in [a_k, b_k], \\ 0 & \text{otherwise,} \end{cases}$$

where

$$g_1(x,y) = y^3 \cos\frac{1}{y} - x, \qquad g_2(x,y) = y^3 \cos\frac{1}{y} + x,$$

$$f_2(x,y) = \begin{cases} g_3(x,y) & \text{if } x \geq 8y^3,\ g_3(x,y) \geq 0,\ y \in [a_k, b_k], \\ g_4(x,y) & \text{if } x \leq 8y^3,\ g_4(x,y) \geq 0,\ y \in [a_k, b_k], \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_3(x,y) = g_1(x,y) + 8y^3, \qquad g_4(x,y) = g_2(x,y) - 8y^3.$$

By setting $u = (1,0)$ and considering the sequence $\xi^k = (x^k, y^k) = (0, (2^k\pi)^{-1}) \to (0,0)$, with $t_k = 8(y^k)^3$ we obtain

$$\frac{1}{t_k}\left[f(\xi^k + t_k u) - f(\xi^k)\right] = \frac{1}{t_k}\left[\begin{pmatrix} 0 \\ (y^k)^3 \end{pmatrix} - \begin{pmatrix} (y^k)^3 \\ 0 \end{pmatrix}\right]$$
$$= \begin{pmatrix} -\frac{1}{8} \\ \frac{1}{8} \end{pmatrix} \in Tf(0,0)(1,0).$$

Let the sequence $r_k \downarrow 0$ be given. In order to show that $f$ is simple at $(0,0)$ we have to find a sequence $\zeta^k = (z^k, w^k) \to (0,0)$ with $v^k = r_k^{-1}[f(\zeta^k + r_k u) - f(\zeta^k)] \to (-\frac{1}{8}, \frac{1}{8})^T$ at least for some subsequence. Necessarily, both components of $v^k$ have to be nonzero for $k$ sufficiently large. This is only possible if $f_1(\zeta^k) \neq$

Without loss of generality let $w^k \geq 0$. Because of $u = (1, 0)$ we obtain $f_1(\zeta^k) = f_1(z^k, w^k) \neq 0$ and $f_2(\zeta^k + r_k u) = f_2(z^k + r_k, w^k) \neq 0$. This yields that for $k$ sufficiently large there exists an index $n(k)$ with

$$w^k \in [a_{n(k)}, b_{n(k)}] \tag{4.4}$$

and $|z^k| \leq (w^k)^3$, $|z^k + r_k - 8(w^k)^3| \leq (w^k)^3$. It follows $z^k + r_k \in [7(w^k)^3, 9(w^k)^3]$ and $r_k \in [6(w^k)^3, 10(w^k)^3]$. Together with (4.4) we obtain for $k$ sufficiently large

$$r_k \in [6a^3_{n(k)}, 10b^3_{n(k)}]. \tag{4.5}$$

On the other hand one can easily see that $\forall k \geq 2$ there is $10b_k^3 < 6a_{k-1}^3$. This means that for every sequence $\{r_k\}$ with $10b_k^3 < r_k < 6a_{k-1}^3$ $\forall k \geq 2$ the condition (4.5) cannot be satisfied. Thus, $f$ is not simple at $(0, 0)$.

## 5. Pathological Lipschitz functions

In this final section we give examples of Lipschitz functions which are pathological with respect to properties of (generalized) derivatives.

In the basic Example 5.1 (see Klatte and Kummer, 2002), we construct a special real Lipschitz function $G$ such that the Clarke subdifferential satisfies $\partial G(x) \equiv [-1, 1]$.

Further, Example 5.2 is taken from Kummer (1988) and presents a real Lipschitz function $f$ such that, for almost all initial points, the standard Newton method provides alternating Newton sequences, though $f$ is differentiable at all iteration points. It illustrates why one has to utilize suitable local approximations in the analysis of Newton-type methods for locally Lipschitz functions (see, e.g., Kummer 1988, 1992, 2000, Pang, 1990, Qi and Sun, 1993, Robinson, 1994).

Finally, Example 5.3 (compare Klatte and Kummer, 2002) presents a convex real function which is non-differentiable on a dense set.

EXAMPLE 5.1 (a pathological real Lipschitz function: lightning function). We present a simple construction of a special real Lipschitz function $G$ such that F.H. Clarke's subdifferential fulfills $\partial G(x) \equiv [-1, 1]$. The existence of such functions has been clarified in Borwein, Moors and Xianfy (1994).

It will be seen that the following sets are dense in $\mathbb{R}$:

the set $D_N = \{x |\ G$ is not directionally differentiable at $x\}$,

the set of local minimizers, and the set of local maximizers.

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

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

Here,

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

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

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

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

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

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 $p_1, \ldots, p_4$, we have

$$\operatorname{Lip}(w(\varepsilon, U, V)) < 1 \text{ provided that } \varepsilon \text{ is small.}$$

After taking $\varepsilon$ in this way, we may repeat our construction (like defining Cantor's set) with each of the related three pieces and larger $k$.

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

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

In the next step $k = 2$ we apply the construction to the three pieces just obtained, then with $k = 3$ to the now existing nine pieces, and so on.

The concrete choice of the (feasible) $\varepsilon = \varepsilon(k) > 0$ is not important in this context. We obtain a sequence of piecewise affine functions $g_k$ on $[0, 1]$ with Lipschitz modulus $< 1$. This sequence has a cluster point $g$ in the space $C[0, 1]$ of continuous functions, and $g$ has the Lipschitz modulus $L = 1$. Let

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

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

Connecting arbitrary three neighboring kink-points of $g_k$ and taking into account that these points belong to the graph of $g$, one sees that $g$ has a dense set of local minimizers (and maximizers).

Further, let $D$ be the dense set of all centre points $c$ belonging to some subinterval used during the construction. Then, each $y \in D$ is again a centre point of some subinterval $I(k)$ for each step with sufficiently large $k$. Thus, $g(y) = g_k(y)$ is again true. Moreover, for arbitrary $\delta \in (0, 1)$, one finds points

$$y', y'' \in (y, y+\delta) \text{ such that } y', y'' \in N$$
$$\text{and } g(y') - g(y) > (1-\delta)(y'-y) \text{ as well as}$$

namely the nearest kinks of $g_k$ on the righthand side of $y$ where $k$ is (large and) odd or even, respectively. This shows that directional derivatives $g'(y;1)$ cannot exist for $y \in D$. In addition, by the mean-value theorem for Lipschitz functions, Clarke (1983), one obtains $\partial g(x) = [-1,1]\,\forall x \in (0,1)$.

To finish the construction, define $G$ on $\mathbb{R}$ by setting $G(x) = g(x-\text{integer}\,(x))$, where integer $(x)$ denotes the integer part of $x$. It is worth noting that $G$ is nowhere semismooth in the sense of Mifflin (1977).

*Derived functions:* Let $h(x) = \frac{1}{2}(x + G(x))$. Then $\partial h(x) = [0,1]$ for all $x$, $h$ is strictly increasing, has a continuous inverse $h^{-1}$ which is nowhere locally Lipschitz, and $h$ is not directionally differentiable on a dense subset of $\mathbb{R}$. In the negative direction $-1$, $h$ is strictly decreasing, but Clarke's directional derivative $h^c(x;-1)$ is identically zero. The integral

$$F(t) = \int_0^t h(x)\,dx$$

is a convex $C^{0,1}$ function with strictly increasing derivative $h$, such that

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

holds true.

EXAMPLE 5.2 (alternating Newton sequences for real, Lipschitzian $f$ with almost all initial points).

To construct $f : \mathbb{R} \to \mathbb{R}$, consider intervals $I(k) = [k^{-1}, (k-1)^{-1}] \subset \mathbb{R}$ for integers $k \geq 2$, and put

$$c(k) = \frac{1}{2}[k^{-1} + (k-1)^{-1}] \quad \text{(the center of } I(k))$$
$$c(2k) = \frac{1}{2}[(2k)^{-1} + (2k-1)^{-1}] \quad \text{(the center of } I(2k)).$$

In the $(x,y)$-plane, define

$g_k = g_k(x)$ to be the linear function through
the points $((k-1)^{-1}, (k-1)^{-1})$ and $(-c(k), 0)$,

i.e.,

$$g_k(x) = a_k(x + c(k)), \text{ where } a_k = (k-1)^{-1}/[(k-1)^{-1} + c(k)].$$

Similarly, let

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

i.e.,

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

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

We finish the construction by setting $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0$.

The related properties can be seen as follows:

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

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

On the lefthand-side of the interval $I(k), f$ coincides with $h_k$, on the right with $g_k$. Since $g_k(c(k)) < h_k(c(k)), f$ coincides with $g_k$ on a small neighborhood of the center point $c(k)$.

Now, let us start Newton's method at some $x^0 \in \Theta^1$, where $\Theta^1$ is the set of $C^1$ points of $f$. Then the next iterate $x^1$ is some point $\pm c(k) \in \Theta^1$. There, $Df = Dg_k$ (or $Df = -Dg_k$ for negative arguments) holds. Hence, the method generates the alternating sequence $x^0, x^1, x^2 = -x^1, x^3 = x^1, \ldots$

EXAMPLE 5.3 (a convex function $f : \mathbb{R} \to \mathbb{R}$, non-differentiable on a dense set). Consider all rational arguments $y = \dfrac{p}{q} \in (0, 1]$ such that $p$, $q$ are positive integers, prime to each other, and put

$$h(y) = \frac{1}{q!}.$$

For fixed $q$, the sum $S(q)$ over all feasible $h(y)$ is bounded by

$$S(q) \leq \frac{q}{q!} \quad \text{and} \quad \sum_q S(q) = c < \infty.$$

Now define $g_1$ by

$$g_1(0) = 0 \text{ and } g_1(x) = \textstyle\sum_{y \leq x} h(y) \text{ for } x \in (0, 1].$$

Then $g_1$ is increasing, bounded by $c$ and has jumps of size $(q!)^{-1}$ at $x = y$.

Next extend $g_1$ on $\mathbb{R}_+$ by setting $g(0) = 0$ and

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

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

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

is convex. For $t \downarrow y$ and $t \uparrow y$ ($t$ irrational, $y$ rational) one obtains different limits of $Df(t)$. Thus $f$ is not differentiable at $y$.

**Acknowledgement.** The authors gratefully acknowledge the comments and

## References

AUBIN, J.-P. and EKELAND, I. (1984) *Applied Nonlinear Analysis.* Wiley, New York.

BORWEIN, J.M. MOORS, W.B. and XIANFY, W. (1994) Lipschitz functions with prescribed derivatives and subderivatives. CECM Information document 94-026, Simon Fraser Univ., Burnaby.

CLARKE, F.H. (1976) On the inverse function theorem. *Pacific Journal of Mathematics*, **64**, 97–102.

CLARKE, F.H. (1983) *Optimization and Nonsmooth Analysis.* Wiley, New York.

DONTCHEV, A.(1995) Characterizations of Lipschitz stability in optimization. In: R. Lucchetti and J. Revalski, eds., *Recent Developments in Well-Posed Variational Problems*, Kluwer, 95–116.

DONTCHEV, A. and ROCKAFELLAR, R.T. (1996) Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization*, **6**, 1087–1105.

FUSEK, P. (2001) Isolated zeros of Lipschitzian metrically regular $R^n$ functions. *Optimization*, **49**, 425–446.

HENRION, R. and OUTRATA, J. (2001) A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications*, **258**, 110–130.

KING, A. and ROCKAFELLAR, R.T. (1992) Sensitivity analysis for nonsmooth generalized equations. *Mathematical Programming*, **55**, 341–364.

KLATTE, D. and KUMMER, B. (1999) Generalized Kojima functions and Lipschitz stability of critical points. *Computational Optimization and Applications*, **13**, 61–85.

KLATTE, D. and KUMMER, B. (2002) Constrained minima and Lipschitzian penalties in metric spaces. *SIAM Journal on Optimization*, to appear.

KLATTE, D. and KUMMER, B. (2002) *Nonsmooth Equations in Optimization—Regularity, Calculus, Methods and Applications.* Nonconvex Optimization and Its Applications. Kluwer Academic Publ., Dordrecht-Boston-London.

KOJIMA, M. (1980) Strongly stable stationary solutions in nonlinear programs. In: S.M. Robinson, ed., *Analysis and Computation of Fixed Points*, Academic Press, New York, 93–138.

KUMMER, B. (1988) Newton's method for non-differentiable functions. In: J. Guddat et al., eds., *Advances in Math. Optimization*, Akademie Verlag Berlin (Ser. Math. Res. 45), Berlin, 114–125.

KUMMER, B. (199) Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ optimization. *Journal of Optimization Theory and Applications*, **70**, 559–580.

KUMMER, B. (1992) Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In: W. Oettli and D. Pallaschke, KUMMER, B. (1997) Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear programming. In: A.V. Fiacco, ed., *Mathematical Programming with Data Perturbations*, Marcel Dekker, New York, 201–222.

KUMMER, B. (2000a) Generalized Newton and NCP-methods: Convergence, regularity, actions. *Discussiones Mathematicae - Differential Inclusions*, **20**, 209–244.

KUMMER, B. (2000b) Inverse functions of pseudo regular mappings and regularity conditions. *Mathematical Programming, Series B*, **88**, 313–339.

LEVY, A.B. and ROCKAFELLAR, R.T. (1996) Variational conditions and the proto-differentiation of partial subgradient mappings. *Nonlinear Analysis: Theory, Methods & Applications*, **26**, 1951–1964.

LUO, Z.-Q. PANG, J.-S. and RALPH, D. (1996) *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge.

MANGASARIAN, O.L. and FROMOVITZ, S. (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, **17**, 37–47.

MIFFLIN, R. (1977) Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, **15**, 957–972.

MORDUKHOVICH, B.S. (1993) Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions. *Transactions of the American Mathematical Society*, **340**, 1–35.

OUTRATA, J. (2000) A generalized mathematical program with equilibrium constraints. *SIAM Journal on Control and Optimization*, **38**, 1623–1638.

PANG, J.-S. (1990) Newton's method for B-differentiable equations. *Mathematics of Operations Research*, **15**, 311–341.

QI, L. and SUN, J. (1993) A nonsmooth version of Newton's method. *Mathematical Programming*, **58**, 353–367.

ROBINSON, S.M. (1976) Regularity and stability for convex multivalued functions. *Mathematics of Operations Research*, **1**, 130–143.

ROBINSON, S.M. (1981) Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, **14**, 206–214.

ROBINSON, S.M. (1994) Newton's method for a class of nonsmooth functions. *Set-Valued Analysis*, **2**, 291–305.

ROCKAFELLAR, R.T. (1989) Proto-differentiability of set-valued mappings and its applications in optimization. In: H. Attouch, J.-P. Aubin, F. Clarke, and I. Ekeland, eds., *Analyse Non Linéaire*, Gauthier-Villars, Paris, 449–482.

ROCKAFELLAR, R.T. and WETS, R.J.-B. (1998) *Variational Analysis*. Springer, Berlin.

THIBAULT, L. (1980) Subdifferentials of compactly Lipschitz vector-valued functions. *Annali di Matematica Pura ed Applicata*, **4**, 157–192.

THIBAULT, L. (1982) On generalized differentials and subdifferentials of Lipschitz vector-valued functions. *Nonlinear Analysis: Theory, Methods &*