Differentiability of convex functions outside small sets

Small sets. We are going to present some results of the following type: under a suitable assumption on a normed space X, each continuous convex function, defined on an open convex set $A \subset X$, is (Gâteaux or Fréchet) differentiable outside a set which is in some sense small.

It is natural to ask that a nonempty family $\mathcal{S} \subset 2^X$ whose elements are considered "small sets" satisfy the following conditions:

(a) $A \in \mathcal{S},\ B \subset A \ \Rightarrow\ B \in \mathcal{S}$;
(b) $A_n \in \mathcal{S}$ for each $n \in \mathbb{N} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{S}$;
(c) $A \in \mathcal{S} \Rightarrow A + v \in \mathcal{S}$ for each $v \in X$ (that is, $\mathcal{S}$ is translation invariant);
(d) $\mathcal{S}$ contains no nonempty open set.

Notice that (a) and the assumption that $\mathcal{S}$ is nonempty imply that $\emptyset \in \mathcal{S}$. Each family satisfying (a) and (b) is called a *σ-ideal.* Thus a *family of small sets* has to be a nonempty translation invariant σ-ideal that cointains no open ball (of positive radius).

Let us list a few important families of small sets.

At most countable sets. The σ-ideal

$$\mathcal{C} = \{E \subset X :\ E \text{ is at most countable}\}$$

satisfies (a)–(d).

Lebesgue null sets. Let m_d denote the Lebesgue measure on $\mathbb{R}^d$. For $X = \mathbb{R}^d$, the σ-ideal

$$\mathcal{N} = \{E \subset \mathbb{R}^d : m_d(E) = 0\}$$

satisfies (a)–(d). (The same holds for an arbitrary translation invariant complete measure on $\mathbb{R}^d$ which is positive on each open ball.)

Meager sets. Recall that a set $E \subset X$ is called *meager* or a *set of the first Baire category* if it is contained in a countable union of closed sets having empty interiors. Then the family

$$\mathcal{M} = \{E \subset X :\ E \text{ is meager}\}$$

satisfies (a)–(c). If X is a Banach space, then $\mathcal{M}$ satisfies also (d) by the Baire Category Theorem.

The meager sets are considered small from the topological point of view, while the null sets (in $\mathbb{R}^d$) are small from the measure point of view. However there is no inclusion relation between these two kinds of smallness. Indeed, it is well-known that (for each $d \in \mathbb{N}$) $\mathbb{R}^d$ contains a Borel set A such that $A \in \mathcal{N}$ and $\mathbb{R}^d \setminus A \in \mathcal{M}$.

Lipschitz-small sets. Let us start with the following observation. Let $H \subset X$ be a closed subspace of codimension 1, and $v_0 \in X \setminus H$. Then $X = H \oplus \mathbb{R}v_0$ in the algebraic sense, that is, each $x \in X$ has a unique expression as the sum of an element of H and a multiple of v_0. In other words, the mapping

$$(1) \qquad H \times \mathbb{R} \to X\,, \quad (u,t) \mapsto u + tv_0\,,$$

is an algebraic isomorphism of $H \times \mathbb{R}$ onto X. Let us show that it is also a topological isomorphism, that is, an isomorphism between normed spaces $H \times \mathbb{R}$ and X.

(Let us recall that the product topology of $H \times \mathbb{R}$ is generated by any of the equivalent norms $\|(u,t)\|_p = |(\|u\|, |t|)|_p$ where $|\cdot|_p$ is the standard ℓ_p-norm on $\mathbb{R}^2$, $1 \le p \le \infty$. For example, we can consider $H \times \mathbb{R}$ equipped with the norm $\|(u,t)\|_1 = \|u\| + |t|$.)

The mapping (1) is obviously continuous: $\|u + tv_0\| \le \|u\| + |t|\|v_0\| \le C\|(u,t)\|_1$ where $C = \max\{1, \|v_0\|\}$. Let us show that also its inverse is continuous. The subspace H is the kernel of some $x^* \in X^*$. By taking an appropriate multiple of x^*, we can suppose that $x^*(v_0) = 1$. Then we can write any $x \in X$ in the form

$$x = \big(x - x^*(x)v_0\big) + x^*(x)v_0$$

where $u_x := x - x^*(x)v_0 \in H$ and $t_x := x^*(x) \in \mathbb{R}$ are such that x is the image of (u_x, t_x) by the mapping (1). Obviously, $x \mapsto u_x$ and $x \mapsto t_x$ are continuous, and hence the inverse of (1) is continuous.

Definition 0.1. A set $L \subset X$ is a *Lipschitz hypersurface* if there exist a closed subspace $H \subset X$ of codimension 1, a vector $v_0 \in X \setminus H$, and a Lipschitz function $\varphi\colon H \to \mathbb{R}$ such that

$$L = \{u + \varphi(u)v_0 \colon u \in H\}\,.$$

Remark 0.2. Let $L \subset X$ be a Lipschitz hypersurface given by H, v_0, φ (in the sense of Definition 0.1).

(a) In the isomorphism (1), the hypersurface L corresponds to the graph of the Lipschitz function φ in $H \times \mathbb{R}$.
(b) L is a closed set without interior points (this follows e.g. from (a)).
(c) If $X = \mathbb{R}^d$, L is of Lebesgue measure zero by Fubini's theorem (since any line parallel to v_0 intersects L in exactly one point).

Definition 0.3. A set $E \subset X$ will be called *Lipschitz-small* if it is contained in a countable union of Lipschitz hypersurfaces.

The family

$$\mathcal{L} = \{E \subset X :\ E \text{ is Lipschitz small}\}$$

satisfies (a)–(c). If X is a Banach space, then $\mathcal{L}$ satisfies also (d) since its elements are meager by Remark 0.2(b).

Observe that Lipschitz small sets in $X = \mathbb{R}$ coincide with at most countable sets. Since the standard Cantor set is meager and uncountable, the family of Lipschitz small sets in $\mathbb{R}$ is strictly contained in the family of meager sets. (This example can be generalized to any nontrivial normed space X.)

For $X = \mathbb{R}^d$, $\mathcal{L}$ is contained in $\mathcal{N} \cap \mathcal{M}$. Moreover, the same example of the Cantor set shows that the inclusion is strict for $d = 1$ (and it can be generalized to any finite dimension d).

Zajíček's theorem. The following theorem by L. Zajíček is formulated for monotone mappings since monotonicity is the only property of the subdifferential which is used in its (quite simple) proof. Recall that a multivalued mapping $T\colon X \to 2^{X^*}$ is *monotone* if

$$(x^* - y^*)(x - y) \geq 0 \qquad \text{whenever } x, y \in X,\ x^* \in T(x),\ y^* \in T(y).$$

(Notice that we do not require that all values of T are nonempty.)

Let us start with a well-known extension lemma.

Lemma 0.4 (Extension of Lipschitz functions)**.** *Let X be a metric space, $E \subset X$ a nonempty set, and $\varphi\colon E \to \mathbb{R}$ an r-Lipschitz function. Then the formula*

$$\hat{\varphi}(x) = \inf\{\varphi(y) + r d(x, y) : y \in E\} \qquad (x \in X) \tag{2}$$

defines an r-Lipschitz extension $\hat{\varphi}\colon X \to \mathbb{R}$ of φ.
If, moreover, X is a normed space and φ is convex, then $\hat{\varphi}$ is convex, too.

Proof. $\hat{\varphi}$ is r-Lipschitz since it is the infimum of a family of r-Lipschitz functions. For each $x \in E$, we have

$$\hat{\varphi}(x) \leq \varphi(x) + r d(x, x) = \varphi(x) \leq \varphi(y) + r d(x, y) \quad (y \in E).$$

Passing to the infimum over $y \in E$, we get $\hat{\varphi}(x) = \varphi(x)$. Thus $\hat{\varphi}$ is an extension of φ.

Let us show the second part. Let $x_1, x_2 \in X$, $t \in (0, 1)$, and $\varepsilon > 0$. There exist $y_i \in E$ $(i = 1, 2)$ such that $\varphi(y_i) + r\|x - y_i\| \leq \hat{\varphi}(x_i) + \varepsilon$. Then

$$\begin{aligned}
\hat{\varphi}((1-t)x_1 + tx_2) &\leq \varphi((1-t)y_1 + ty_2) + r\|(1-t)x_1 + tx_2 - (1-t)y_1 - ty_2\| \\
&\leq (1-t)\varphi(y_1) + t\varphi(y_2) + (1-t)r\|x_1 - y_1\| + tr\|x_2 - y_2\| \\
&\leq (1-t)\hat{\varphi}(x_1) + t\hat{\varphi}(x_2) + \varepsilon\,.
\end{aligned}$$

Passing to limit for $\varepsilon \to 0^+$, we get the needed inequality $\hat{\varphi}((1-t)x_1 + tx_2) \leq (1-t)\hat{\varphi}(x_1) + t\hat{\varphi}(x_2)$. $\square$

Theorem 0.5 (Zajíček)**.** *Let X be a separable normed space, $T\colon X \to 2^{X^*}$ a monotone multivalued mapping. Then the set*

$$M(T) = \{x \in X :\ \operatorname{card}[T(x)] > 1\}$$

is Lipschitz-small.

Proof. Let D be a countable dense subset of S_X (the boundary of the unit ball). Given an arbitrary $x \in M(T)$, choose $a_x^*, b_x^* \in T(x)$ such that $a_x^* \neq b_x^*$. There exists $v_x \in D$ such that $a_x^*(v_x) < b_x^*(v_x)$. Choose rational numbers α_x and β_x such that

$$a_x^*(v_x) < \alpha_x < \beta_x < b_x^*(v_x)\,,$$

and a natural number m_x such that $\|a_x^*\| \leq m_x$ and $\|b_x^*\| \leq m_x$.

Now, we can write

$$M(T) = \bigcup \{E(v, \alpha, \beta, m) :\ v \in D,\ \alpha, \beta \in \mathbb{Q},\ \alpha < \beta,\ m \in \mathbb{N}\}$$

where

$$E(v, \alpha, \beta, m) = \{x \in M(T) :\ v_x = v,\ \alpha_x = \alpha,\ \beta_x = \beta,\ m_x = m\}\,.$$

Since the above union is countable, it remains to show that each $E(v, \alpha, \beta, m)$ is contained in a Lipschitz hypersurface.

Fix v, α, β, m. Chose $v^* \in X^*$ such that $v^*(v) \neq 0$, and put $H = \ker(v^*)$. Then X is the direct sum of H and $\mathbb{R}v$. Thus, each point $x \in E(v, \alpha, \beta, m)$ admits a unique expression of the form

$$x = u_x + t_x v \qquad \text{where } u_x \in H \text{ and } t_x \in \mathbb{R}.$$

By monotonicity, we have for each $x, y \in E(v, \alpha, \beta, m)$

$$0 \le (a_x^* - b_y^*)(x - y) = (a_x^* - b_y^*)(u_x - u_y) + (t_x - t_y)[a_x^*(v) - b_y^*(v)]\,.$$

It follows that

$$(t_x - t_y)[b_y^*(v) - a_x^*(v)] \le (a_x^* - b_y^*)(u_x - u_y) \le 2m\|u_x - u_y\|\,.$$

Notice that $b_y^*(v) - a_x^*(v) \ge \beta - \alpha > 0$. Thus we can devide by the expression in square brackets to obtain

$$t_x - t_y \le \frac{2m}{b_y^*(v) - a_x^*(v)}\,\|u_x - u_y\| \le \frac{2m}{\beta - \alpha}\,\|u_x - u_y\|\,.$$

Interchanging the roles of x and y, we easily conclude that

$$|t_x - t_y| \le \frac{2m}{\beta - \alpha}\,\|u_x - u_y\| \qquad \text{whenever } x, y \in E(v, \alpha, \beta, m). \tag{3}$$

In particular, if $x, y \in E(v, \alpha, \beta, m)$ and $u_x = u_y$, then $t_x = t_y$. In other words, for each $x \in E(v, \alpha, \beta, m)$ there exists a unique $t \in \mathbb{R}$ such that $u_x + tv \in E(v, \alpha, \beta, m)$ (namely, $t = t_x$). Define

$$A = \{u_x : x \in E(v, \alpha, \beta, m)\}\,, \quad \varphi\colon A \to \mathbb{R}\,,\ \varphi(u_x) = t_x\,.$$

Then $A \subset H$ and, by(3), φ is Lipschitz. By Lemma 0.4, there exists a Lipschitz extension $\hat{\varphi}\colon H \to \mathbb{R}$ of φ. Then

$$E(v, \alpha, \beta, m) = \{u + \varphi(u)v : u \in A\} \subset \{u + \hat{\varphi}(u)v : u \in H\}$$

where the last set is a Lipschitz hypersurface. This completes the proof. $\square$

Corollary 0.6. *Let X be a separable Banach space, $A \subset X$ an open convex set, and $f\colon A \to \mathbb{R}$ a continuous convex function. Then the set*

$$NG(f) = \{x \in A :\ f \text{ is not Gâteaux differentiable at } x\}$$

is Lipschitz-small, and hence meager.

Proof. The multivalued mapping $T\colon X \to 2^{X^*}$, given by

$$T(x) = \begin{cases} \partial f(x) & \text{if } x \in A, \\ \emptyset & \text{otherwise,} \end{cases}$$

is monotone and $NG(f) = M(T)$. Apply Theorem 0.5. $\square$

Corollary 0.7. *Let $A \subset \mathbb{R}^d$ be an open convex set, $f\colon A \to \mathbb{R}$ a convex function. Then the set*

$$NF(f) = \{x \in A : \ f \text{ is not Fréchet differentiable at } x\}$$

is Lipschitz-small, and hence meager and of Lebesgue measure zero. Moreover, the Fréchet derivative $f'(\cdot)$ is continuous on the set $A \setminus NF(f)$.

Proof. By a well-known fact about continuity of convex functions, f is locally Lipschitz on A. Thus $NF(f) = NG(f)$ by a general fact about Lipschitz functions in $\mathbb{R}^d$. By Corollary 0.6, $NF(f)$ is Lipschitz-small. The last assertion follows immediately from the fact that, for each $x \in A \setminus NF(f)$, ∂f is single-valued and continuous with $\partial f(x) = \{f'(x)\}$. $\square$

Notice that, for $d = 1$, Corollary 0.7 says nothing else that the set of nondifferentiability points of a convex function of one real variable is at most countable.

An infinite-dimensional result about generic Fréchet differentiability.

Lemma 0.8. *Let X be a normed space. If X^* is separable then also X is separable, but not vice-versa.*

Proof. If X^* is separable, there exists a sequence $\{v_n^*\} \subset S_{X^*}$ which is dense in S_{X^*}. For each n, choose $v_n \in S_X$ such that $v_n^*(v_n) > \frac{1}{2}$. The closed linear subspace

$$Y = \overline{\text{span}}\{v_n : n \in \mathbb{N}\}$$

of X is separable since the countable set of all finite linear combinations with rational coefficients of the vectors v_n ($n \in \mathbb{N}$) is dense in Y.

Assume that $Y \neq X$. By the Hahn-Banach Theorem, there exists $x^* \in S_{X^*}$ such that $Y \subset \ker(x^*)$. There exists $m \in \mathbb{N}$ such that $\|v_m - x^*\| < \frac{1}{2}$. Then we get

$$\tfrac{1}{2} < v_m^*(v_m) = x^*(v_m) + (v_m^* - x^*)(v_m) \leq 0 + \|v_m^* - x^*\| \, \|v_m\| < \tfrac{1}{2}\,,$$

a contradiction. Thus $X = Y$ is separable.

Finally, $X = \ell_1$ is an example of a separable Banach space such that $X^* = \ell_\infty$ is not separable. $\square$

Theorem 0.9 (Preiss–Zajíček)**.** *Let X be a Banach space with X^* separable. Then, for every continuous convex function f defined on an open convex set $A \subset X$, the set $NF(f)$ (see Corollary 0.7) is meager.*

One more theorem. Let us conclude with one more theorem which easily follows from theorems by Ekeland–Lebourg (1976) and Preiss–Phelps–Namioka (1990) about differentiability of convex functions under existence of nice equivalent norms on the space. Our theorem says that if a Banach space X contains *one* continuous convex function which is coercive and everywhere differentiable then *every* continuous convex function is differentiable everywhere except a small set. Recall that a function $f\colon X \to \mathbb{R}$ is called coercive if $\lim_{\|x\|\to+\infty} f(x) = +\infty$.

Theorem 0.10. *Let X be a Banach space. Assume that there exists a continuous convex function $g\colon X \to \mathbb{R}$ such that g is coercive and everywhere Gâteaux/Fréchet differentiable. Then each continuous convex function, defined on an open convex set $A \subset X$, is Gâteaux/Fréchet differentiable except a meager set.*