TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 342, Number 1, March 1994

SECOND ORDER DIFFERENTIABILITY OF CONVEX FUNCTIONS IN BANACH SPACES

JONATHAN M. BORWEIN AND DOMINIKUS NOLL

ABSTRACT. We present a second order differentiability theory for convex functions on Banach spaces.

1. INTRODUCTION

The classical theorem of Alexandrov states that a convex function on $\mathbb{R}^n$ is almost everywhere second order differentiable. This was first proved by Busemann and Feller [12] for functions on $\mathbb{R}^2$ and later was extended by Alexandrov [2] to $\mathbb{R}^n$. More recent proofs were obtained by Mignot [26], Bangert [6], and Rockafellar [36].

Around 1975, Aronszajn [3] and Christensen [13] among others proved versions of Rademacher's theorem on almost everywhere differentiability of Lipschitz operators which apply in separable infinite dimensional Banach spaces. While these results typically do not extend beyond the separable case, it was only in 1990 when Preiss [35] proved the remarkable fact that every real valued Lipschitz function on a not necessarily separable Banach space which is an Asplund space is still at least densely Fréchet differentiable.

Motivated by these infinite dimensional versions of Rademacher's theorem, the present work is to attack Alexandrov's theorem in infinite dimensions. As it turns out, the situation here is less promising than it is for Rademacher's theorem. For instance, Alexandrov's theorem fails in the spaces l_p, L_p, $1 \leq p < 2$, and much to our surprise, even in nonseparable Hilbert spaces. This leads us to focus on the case of separable Hilbert spaces. Here in fact, a positive solution seems possible. As one of our central results here, we in fact obtain a partial positive answer by proving a version of Alexandrov's theorem for convex integral functionals.

Seemingly, the third of the classical results of measure theoretic geometry, the theorem of Sard, allows extensions to infinite dimensions only under comparatively strong hypotheses (see [1, 10]). In the light of our present investigation, this is explained to some extent by the fact that there is a strong link between Alexandrov's theorem and a version of Sard's theorem for monotone operators

Received by the editors June 20, 1991 and, in revised form, October 29, 1991 and December 9, 1991.

1991 *Mathematics Subject Classification.* Primary 46B22; Secondary 26B25.

The first author was supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

The second author was supported by Deutsche Forschungsgemeinschaft (DFG) and NSERC.

0002-9947/94 $1.00 + $.25 per page

on a Hilbert space. Roughly speaking, this tells that these theorems are equally difficult to establish in infinite dimensions.

It turns out that an appropriate understanding of Alexandrov's theorem in Hilbert space hinges on using the theory of *generalized second derivatives*. In finite dimensions, this has been exploited over recent years by Rockafellar [36, 37, 39]. Here we discuss the infinite dimensional theory, which bears some very interesting features of its own. Similar to the case of finite dimensions, however, its merits lie mainly in the fact that it allows for dualizing second order notions, permitting finally to express them in first order terms for associated monotone operators. In contrast with Hiriart-Urruty's theory of approximate *directional* second derivatives (cf. [21, 22–24]), generalized second differentiability in the present sense is designed to reflect the *global* second order behaviour of a convex function at a given point. This turns out to be extremally useful when approaching Alexandrov's theorem.

The structure of our paper is as follows. In §§2 and 3 we discuss different notions of second order differentiability arising in infinite dimensions. We single out the very one which allows for a reasonable second order theory. Section 4 presents a measure theoretic version of Alexandrov's theorem. Section 5 presents some basic information on the Yosida approximates of a convex function, providing a $C^{1,1}$ type approximation for continuous convex functions in Hilbert space. In §6 we discuss generalized second derivatives for convex functions on a Hilbert space. Our discussion is of interest also for functions on $\mathbb{R}^n$. In particular, we present an interesting class of examples of convex functions on $\mathbb{R}^2$ showing that the behaviour of the generalized second derivative may be strikingly different from the pointwise second derivative. Moreover, generalized second derivatives are shown to be useful for convex integral functionals.

Section 7 presents a way of dualising Lipschitz smoothness, which in tandem with the duality results in §6 finally leads to the dual description of second order differentiability. The latter then permits us to present the mentioned version of Alexandrov's theorem for convex integral functionals. Section 8 gives the link between Alexandrov's theorem and Sard's theorem.

The terminology we use is covered by any textbook on convex functions. See, for instance, Phelps' lecture notes [34].

2. Second order differentiability

What does it mean for a convex function f to be second order differentiable at a point x? Even in finite dimensions it is not clear a priori how this concept should be introduced, in particular when the function f is not necessarily first order differentiable at all points in a neighbourhood of x. Several notions of second order differentiability have been considered, e.g., by Busemann and Feller [12], Alexandrov [2], Mignot [26], Bangert [6], and Rockafellar [36]. As we shall indicate here, these all coincide in finite dimensions, but lead to nonequivalent notions when extended to infinite dimensions. The main task is then to decide which notions might be appropriate to obtain a reasonable infinite dimensional second order differentiability theory. One way of doing this is by testing whether for example a version of Alexandrov's theorem is possible.

Perhaps the most natural way of introducing second order differentiability is to consider the possibility of a second order Taylor's expansion. This was discussed by Alexandrov [2] and Bangert [6]. We adopt this point of view here.

Definition 2.1. Let f be a continuous convex function defined on a normed space E. Then f is said to be *second order differentiable* (resp. *strongly second order differentiable*) at $x \in E$ if there exist a $y^* \in \partial f(x)$ and a bounded linear operator $T: E \to E^*$ such that f has a representation of the form

$$f(x+th) = f(x) + t\langle y^*, h\rangle + (t^2/2)\langle Th, h\rangle + o(t^2) \qquad (t \to 0), \tag{2.1}$$

for every $h \in E$, respectively

$$f(x+h) = f(x) + \langle y^*, h\rangle + \tfrac{1}{2}\langle Th, h\rangle + o(\|h\|^2) \qquad (\|h\| \to 0). \tag{2.2}$$

Remarks. (1) It follows from (2.1) that $f(x+th) = f(x)+t\langle y^*, h\rangle+o(t)$; hence second differentiability of f at x implies first order Gâteaux differentiability of f at x, i.e., $y^* = \nabla^G f(x)$. (See [34] for the notion of the Gâteaux differential.) Clearly, the quadratic term $\frac{1}{2}\langle Th, h\rangle$ as well is uniquely determined by the Taylor expansions (2.1), resp. (2.2). Due to the convexity of f, this quadratic form is positive semidefinite, and therefore defines a *purely quadratic* continuous convex function

$$q(h) = \tfrac{1}{2}\langle Th, h\rangle. \tag{2.3}$$

Replacing the operator T by its symmetrization $\frac{1}{2}(T+T^*)$, where $T^*: E \to E^*$ is the restriction of the adjoint of T, defined by the identity $\langle Th, k\rangle = \langle T^*k, h\rangle$, $h, k \in E$, we may assume that, in (2.3), T is a nonnegative and symmetric (i.e., $\langle Th, h\rangle \geq 0$ and $T = T^*$) bounded linear operator. We use the notation $T = \nabla^2 f(x)$, calling $\nabla^2 f(x)$ the *Hessian* of f at x.

(2) A convex function $q: E \to \mathbb{R}$ is called *purely quadratic* if it admits a representation of the form (2.3) with a bounded linear operator T. Notice that in a Banach space E this is equivalent to saying that the graph ∂q of the subdifferential of q is a closed linear subspace of $E \times E^*$. Indeed, the representation (2.3) with symmetric T implies $\nabla^G q = T$, proving the linearity of ∂q. Conversely, assuming ∂q closed and linear, we first observe that q is *quadratic* in the sense that $q(\lambda x) = \lambda^2 q(x)$ for all $\lambda \in \mathbb{R}$, $x \in E$ and, on the other hand, that q is everywhere Gâteaux differentiable, with $\nabla^G q = T$ being a closed graph (and therefore bounded) linear operator. But now the identity

$$\langle Tx, x\rangle = \lim_{t\to 0} \frac{q(x+tx) - q(x)}{t} = \lim_{t\to 0} \frac{(1+t)^2 q(x) - q(x)}{t} = 2q(x)$$

gives the purely quadratic representation (2.3) of q. We will see in §6 that, in a Hilbert space, it is possible to obtain a similar characterization even for partially defined purely quadratic convex functions (similar to the one obtained in $\mathbb{R}^n$ by Rockafellar [36]).

Let us now consider an equivalent way of describing second order differentiability. Let f be a continuous convex function defined on a normed space E. We define the *second difference quotient* of f at $x \in E$ relative to $y^* \in \partial f(x)$ as

$$\Delta_{f,x,y^*,t}(h) = \frac{f(x+th) - f(x) - t\langle y^*, h\rangle}{t^2},$$

$t \neq 0$, $h \in E$. Notice that for any $t \neq 0$, $\Delta_{f,x,y^*,t}$ is a continuous convex and nonnegative function of h defined on E. Now second order differentiability has the following equivalent description in terms of $\Delta_{f,x,y^*,t}$.

Proposition 2.1. *The continuous convex function f is second order differentiable at $x \in E$ if and only if, for some $y^* \in \partial f(x)$, the second difference quotient $\Delta_{f,x,y^*,t}$ converges pointwise everywhere (as $t \to 0$) to a continuous convex purely quadratic function q of the form* (2.3). *The case of strong second order differentiability of f at x corresponds to $\Delta_{f,x,y^*,t}$ converging to q uniformly on bounded sets.*

Remarks. (1) Again here it is implicit in the statement that pointwise convergence of $\Delta_{f,x,y^*,t}$ $(t \to 0)$ for some $y^* \in \partial f(x)$ implies $y^* = \nabla^G f(x)$.

(2) It is well known that in the case of a Banach space E, pointwise convergence of a sequence of continuous convex functions forces this sequence to be uniformly convergent on compact sets (Ascoli). As bounded sets are relatively compact in finite dimensions, this immediately shows that both notions of second order differentiability discussed so far coincide in that case. On the other hand, this forcefully suggests that these concepts should be nonequivalent in infinite dimensions. A corresponding example will be given at the end of §3.

Suppose the second difference quotient $\Delta_t = \Delta_{f,x,y^*,t}$ converges pointwise everywhere to a limit function Δ, with $\operatorname{dom}(\Delta) = E$. We deduce immediately that Δ is nonnegative, convex, and due to the formula

$$\Delta_t(\lambda h) = \lambda^2 \Delta_{\lambda t}(h) \qquad (\lambda \neq 0,\ t \neq 0) \tag{2.4}$$

is *quadratic* (i.e., $\Delta(\lambda h) = \lambda^2 \Delta(h)$). Assuming, in addition, that E is a Banach space, we even know that Δ is continuous. But Δ need not be a purely quadratic function of the form (2.3) as occurring in the definition of second order differentiability. An example of this type may be obtained by considering the convex (!) function (cf. [41, p. 152])

$$f(x, y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} + 13(x^2 + y^2), & (x, y) \neq (0, 0), \\ 0, & x = y = 0, \end{cases}$$

which is directionally second order differentiable at $(0, 0)$, but whose mixed derivatives differ at $(0, 0)$. This means that the corresponding limit function Δ is not purely quadratic.

Yet, pointwise convergence of $\Delta_{f,x,y^*,t}$ to a limit function Δ is by itself still meaningful. Namely, it tells us that f is directionally second order differentiable at x, i.e., $f|_L$ is second order differentiable along any line L passing through x. Indeed, we obtain a directional Taylor expansion of the form

$$f(x + th) = f(x) + t\langle \nabla^G f(x), h\rangle + t^2 \Delta(h) + o(t^2) \qquad (t \to 0)$$

for any $h \in E$.

It seems at first glance that pointwise convergence of $\Delta_t := \Delta_{f,x,y^*,t}$ to a quadratic Δ should yield a rather weak second order differentiability concept. Yet it has a striking consequence on the first order behaviour of f at x. Assuming in the sequel that E is a Banach space, we may invoke the following "uniform boundedness principle for convex functions" to deduce that $(\Delta_t)_{0<|t|\leq 1}$ is uniformly bounded above on a neighbourhood of 0 in E. Namely, observing that, due to the convexity of f and the convergence of $\Delta_t(h) \to \Delta(h)$ (as $t \to 0$), the family $(\Delta_t(h))_{0<|t|\leq 1}$ is bounded for any fixed $h \in E$, we may define a lower semicontinuous convex and fully defined function ϕ by setting

$$\phi(h) = \sup_{0<|t|\leq 1} \Delta_t(h).$$

Completeness of E then renders ϕ to be continuous, hence bounded above on a neighbourhood of 0. So there exist $\delta > 0$ and $c > 0$ with $\Delta_t(h) \leq c$ being met for all $0 < |t| \leq 1$ and $\|h\| \leq \delta$. By the definition of Δ_t, this implies

$$f(x+th) - f(x) - t\langle y^*, h\rangle \leq ct^2 \tag{2.5}$$

for all $|t| \leq 1$ and $\|h\| \leq \delta$. This inequality is now recognized as the defining condition for what is called *Lipschitz smoothness* of the function f at x, as introduced by Fabian [16] (see also [9]). In particular, see [16], (2.5) implies that f is Fréchet differentiable at x. Due to the importance of this observation, we state this as

Proposition 2.2. *Let f be a continuous convex function defined on a Banach space E. Then every point of (directional) second order differentiability of f is a Lipschitz-smooth point. In particular, f is Fréchet differentiable at such x.*

An immediate consequence of Proposition 2.2 is that Gâteaux differentiability points which fail to be Fréchet differentiability points—and such are common in infinite dimensions—cannot be points of second order differentiability. Even more, for such a point x there must a priori exist a line L through x such that $f|_L$ does not have a finite second derivative at x along L.

Proposition 2.2 also produces interesting counterexamples. For instance, the norm $\|\cdot\|_1$ on l_1 is nowhere Fréchet differentiable (though Gâteaux differentiable at points $x = (x_n)$ with $x_n \neq 0$ for all n), so $f = \|\cdot\|_1$ is certainly nowhere second order differentiable in l_1. In other terms, Alexandrov's theorem fails in l_1. More generally, this is the case in any Banach space E which is not a LS-space in the sense of [16]. In particular, the validity of Alexandrov's theorem requires an Asplund space.

Notation. We write D_f^2 for the set of points of second order differentiability, and L_f for the set of Lipschitz-smooth points of a function f. So $D_f^2 \subset L_f$ by Proposition 2.2.

3. First order differentiability of ∂f

A quite natural way of studying second order differentiability properties of a convex function f is to look at the first order behaviour of its subdifferential operator ∂f. As ∂f is generally not a single-valued operator, this requires differentiability notions designed for set-valued operators. This approach has been exploited in finite dimensions by Mignot [26] and independently by Bangert [6]. Again, there are several nonequivalent ways of extending these concepts to infinite dimensions.

Definition 3.1. Let f be a continuous convex function defined on a normed space E. Then ∂f is said to be

(1) *weak-$*$ Gâteaux differentiable* (resp. *weak-$*$ Fréchet differentiable*) at $x \in E$ if there exists a bounded linear operator $T: E \to E^*$ such that

$$\lim_{t\to 0} t^{-1}(y_t^* - y^*) = Th$$

in the weak-$*$ sense for any fixed $h \in E$ and all $y_t^* \in \partial f(x+th)$, $y^* \in \partial f(x)$ respectively [uniformly over all $h \in E$, $\|h\| \leq 1$ and all $y_t^* \in \partial f(x+th)$, $y^* \in \partial f(x)$];

(2) *norm Gâteaux differentiable* (resp. *norm Fréchet differentiable*) at x if (1) holds with the respective limits to be understood in the sense of the dual norm.

Remarks. (1) Notice that in any of these four notions it is an implicit consequence that $\partial f(x)$ must consist of a single element y^*, i.e., $y^* = \nabla^G f(x)$.

(2) We use the temporary notation $T = \nabla\partial f(x)$ for the operator T. We will see from Theorem 3.1 that $\nabla\partial f(x)$ is nonnegative and symmetric. For a direct proof of this see [6, Lemma 4.1].

(3) There are four immediate implications among these differentiability concepts, with norm Fréchet differentiability being the strongest and weak-$*$ Gâteaux differentiability being the weakest. As a consequence of Theorem 3.1 and Proposition 2.1, these all coincide in finite dimensions, but fail to do in infinite dimensions.

Theorem 3.1. *Let f be continuous convex and defined on a Banach space E. Then*

(1) *Second order differentiability of f at x is equivalent to weak-$*$ Gâteaux differentiability of ∂f at x.*
(2) *Strong second order differentiability of f at x is equivalent to norm Fréchet differentiability of ∂f at x.*

Moreover, in both cases, we have $\nabla^2 f(x) = \nabla\partial f(x)$.

Proof. First consider statement (1). Assume f is second order differentiable at x. As already noted, this implies Lipschitz smoothness of f at x, for E is a Banach space. Due to [16, Proposition 2.1(iii)] there exist $c > 0$ and $t_0 > 0$ such that

$$\|t^{-1}(y_t^* - y^*)\| \le c \tag{3.1}$$

for all $\|h\| \le 1$, $0 < |t| \le t_0$ and all $y_t^* \in \partial f(x + th)$, $y^* = \nabla^F f(x)$. So the difference quotient of ∂f is bounded. We have to show that it converges (as $t \to 0$) in the weak-$*$ sense.

Let $v \in E$ be fixed. Using (3.1) let $t_n \to 0$ be any null sequence such that the limit

$$\lim_{n\to\infty} \langle t_n^{-1}(y_{t_n}^* - y^*), v\rangle =: \alpha \tag{3.2}$$

exists. We show that there is only one possible such α. As

$$t_n^{-1}(y_{t_n}^* - y^*) \in t_n^{-1}(\partial f(x + t_n h) - \partial f(x)) = \partial\Delta_{t_n}(h),$$

the subgradient inequality implies

$$\langle t_n^{-1}(y_{t_n}^* - y^*), \rho v\rangle \le \Delta_{t_n}(h + \rho v) - \Delta_{t_n}(h) \cdot$$

for any $\rho \in \mathbb{R}$. Passing to the limit on both sides therefore gives

$$\rho\alpha \le q(h + \rho v) - q(h). \tag{3.3}$$

Using the fact that the limit function $q = \frac{1}{2}\langle T\cdot, \cdot\rangle$ with symmetric $T = \nabla^2 f(x)$ is purely quadratic and hence differentiable, dividing (3.3) by $\rho > 0$ and passing to the limit $\rho \searrow 0$, we obtain the inequality $\alpha \le q'(h, v)$. Using a negative ρ then gives equality

$$\alpha = q'(h, v) = \langle Th, v\rangle. \tag{3.4}$$

So fixing h and v, there is only one possible limit α in (3.2) for any sequence $t_n \to 0$ and any choice of $y^*_{t_n} \in \partial f(x + t_n h)$. This proves weak-$*$ convergence of the difference quotient. Moreover, (3.4) shows that

$$\langle Th, v\rangle = \alpha = \lim_{t\to 0}\langle t^{-1}(\partial f(x+th) - \partial f(x)), v\rangle,$$

so $T = \nabla\partial f(x)$, with the latter therefore being symmetric. This shows, in particular, the equality $\nabla^2 f(x) = \nabla\partial f(x)$.

Let us now assume conversely that the difference quotient of ∂f at x converges pointwise everywhere in the weak-$*$ sense to the limit $T = \nabla\partial f(x)$. We have to show that the second difference quotient $\Delta_{f,x,y^*,t}$ converges pointwise everywhere to a purely quadratic limit q. Moreover, we will find the representation $q(h) = \frac{1}{2}\langle Th, h\rangle$. This part of the proof is covered by the work of Bangert [6]. Let us fix $h \in E$. Then by [6, Lemma 4.8], f has a second order Taylor expansion at x along the line $x + \mathbb{R}h$, i.e.,

$$f(x+th) = f(x) + t\langle y^*, h\rangle + (t^2/2)f''(x; h) + o(t^2) \qquad (t \to 0)$$

where $f''(x; h)$ denotes the second derivative of $t \to f(x+th)$ at $t = 0$. The quoted reference also shows that

$$f''(x; h) = \langle Th, h\rangle. \tag{3.5}$$

This proves that, for any fixed h, the second difference quotient $\Delta_t(h)$ converges to the purely quadratic $q(h) = \frac{1}{2}\langle Th, h\rangle$ $(t \to 0)$. But then we may invoke (3.4) to deduce $q' = T$. This proves the final link, showing that T is symmetric, thus completing the proof of part (1).

Let us now consider statement (2). First assume $\Delta_t \to q$ uniformly on bounded sets, with a purely quadratic q. Then part (1) of the proof shows that the first difference quotient of ∂f converges to $T = \nabla^2 f(x) = \nabla\partial f(x)$ in the weak-$*$ sense. We have to show that it actually converges in norm with convergence being uniform over $\|h\| \le 1$. By assumption, given any sequence $t_n \to 0$, there exist $\varepsilon_n \to 0^+$ such that

$$|\Delta_{t_n}(k'_n) - \Delta_{t_n}(h'_n) - (q(k'_n) - q(h'_n))| \le \varepsilon_n \tag{3.6}$$

for all n and all $\|k'_n\| \le 1$, $\|h'_n\| \le 1$. Suppose now the statement is not satisfied. Then there exist $\varepsilon > 0$ and a sequence $t_n \to 0$ such that, for certain $\|h_n\| \le \frac{1}{2}$, say, and $y^*_{t_n} \in \partial f(x + t_n h_n)$, we have

$$\left\| \frac{1}{t_n}(y^*_{t_n} - y^*) - Th_n \right\| > \varepsilon \quad (\text{where } y^* = \nabla^F f(x)), \quad n = 1, 2, \ldots.$$

Find vectors $\|v_n\| \le 1$ such that

$$\left\langle \frac{1}{t_n}(y^*_{t_n} - y^*) - Th_n, v_n \right\rangle > \varepsilon, \tag{3.7}$$

for $n = 1, 2, \ldots$. Now choose ε_n according to (3.6), and find $\rho_n > 0$ such that $\rho_n \to 0$ and $\varepsilon_n/\rho_n \to 0$. Then we have $\|h_n + \rho_n v_n\| \le 1$ for n sufficiently large, and hence (3.6) gives

$$|\Delta_{t_n}(h_n + \rho_n v_n) - \Delta_{t_n}(h_n) - (q(h_n + \rho_n v_n) - q(h_n))| \le \varepsilon_n, \tag{3.8}$$

for $n = 1, 2, \ldots$. Using (3.7), the subgradient inequality, and (3.8) we obtain

$$\rho_n \varepsilon + \langle Th_n, \rho_n v_n \rangle \leq \left\langle \frac{1}{t_n}(y^*_{t_n} - y^*), \rho_n v_n \right\rangle \leq \Delta_{t_n}(h_n + \rho_n v_n) - \Delta_{t_n}(h_n)$$
$$\leq q(h_n + \rho_n v_n) - q(h_n) + \varepsilon_n.$$

Divided by $\rho_n > 0$, letting $n \to \infty$, and using $\varepsilon_n/\rho_n \to 0$, we obtain

$$\varepsilon + \limsup\langle Th_n, v_n \rangle \leq \limsup \frac{q(h_n + \rho_n v_n) - q(h_n)}{\rho_n}$$
$$= \limsup \left(\langle Th_n, v_n \rangle + \frac{\rho_n}{2} \langle Tv_n, v_n \rangle \right),$$

where the latter uses the purely quadratic representation $q(h) = \frac{1}{2}\langle Th, h \rangle$ of q. As $\rho_n \to 0$, we deduce $\varepsilon + \limsup\langle Th_n, v_n \rangle \leq \limsup\langle Th_n, v_n \rangle$, a contradiction. This proves one half of the argument.

The second half of the argument is again covered by Bangert's Lemma 4.8. We only have to observe that his argument maintains the uniformity of the convergence of the difference quotient of ∂f, translating it into the uniform o-condition (2.2) in the Taylor expansion. □

Remarks. (1) For convex functions f on the real line, Theorem 3.1 is just the classical result of Jessen (see [11, §2]).

(2) As a consequence of Theorem 3.1, when combined with Proposition 2.1, we obtain that all second order differentiability notions discussed so far coincide in finite dimensions. Surprisingly, this fact has not been observed in [6, 26, 36].

(3) Theorem 3.1 indicates that basing second order differentiability on the possibility of a second order Taylor expansion is justified for convex functions. This strongly relies on convexity, as the function $f(t) = t^3 \cos(1/t)$ shows.

We now provide examples showing that strong second order differentiability and second order differentiability are nonequivalent in infinite dimensions.

Example 1. Let C be a closed convex set in Hilbert space H, and let $P_C : H \to C$ be the metric projection onto C, i.e., the nearest point mapping. Then P_C is known to be the Fréchet derivative of a continuous convex function f on H, i.e., $P_C = \nabla^F f$, where

$$f(x) = \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}\|x - P_C x\|^2 \tag{3.10}$$

(see [18] for details). As P_C is a Lipschitz operator, it is almost everywhere Gâteaux differentiable in the sense of Aronszajn [3] (see §4) when H is assumed separable. Due to Theorem 3.1(1), this means that, on a separable H, f is almost everywhere second order differentiable. However, even in a separable Hilbert space, the set C may be chosen so that P_C is nowhere (norm) Fréchet differentiable. By Theorem 3.1(2), f is then nowhere strongly second order differentiable. We take $H = L^2[0, 1]$, and let $C = \{f \in H : |f| \leq 1 \text{ a.e.}\}$. Then, according to [18, §5], P_C is nowhere Fréchet differentiable. This shows that Alexandrov's theorem fails even in separable Hilbert space when based on strong second order differentiability. A similar example would be obtained by taking $H = l_2$, C the positive cone in H (see [18, §5]).

Example 2 (Example 1 continued). The situation is even worse in nonseparable Hilbert space. Here the set C may be chosen so that P_C is nowhere Gâteaux

differentiable. So here, by Theorem 3.1(1), f is nowhere second order differentiable, i.e., $D_f^2 = \varnothing$. Take $H = l_2(\Gamma)$ with $|\Gamma| > \aleph_0$, and let C be the positive cone in H. Then P_C is nowhere Gâteaux differentiable. This shows that there is no chance for a version of Alexandrov's theorem in nonseparable Hilbert space.

Example 3. A different type of counterexample is obtained by considering convex functions f on l_2 of the form

$$f(x) = \sum_{n=1}^{\infty} f_n(x_n), \qquad x = (x_n),$$

with appropriate convex functions f_n defined on the real line. Here f is Gâteaux differentiable at $x = (x_n)$ if and only if $f_n'(x_n)$ exists for every n. A necessary condition for $x \in D_f^2$ is the following: $f_n''(x_n)$ exists for every n and the sequence is bounded. However, this is not sufficient to guarantee $x \in D_f^2$, as shown in Example 2 in §6 by specifying the function f. Now one may find f such that $\nabla^F f = T : l_2 \to l_2$ is even a Lipschitz operator having no Fréchet differentiability point at all, while, by Aronszajn's result [3], T is almost everywhere Gâteaux differentiable. An explicit example of such T is [3, §3, Example I], with the corresponding convex f being easily supplemented.

So far we found that among the four differentiability notions for the subdifferential operator ∂f presented in Definition 3.1, weak-$*$ Gâteaux differentiability and norm Fréchet differentiability have counterparts stated in terms of the second order difference quotient of f, or what is the same, in terms of a second order Taylor expansion. This does not seem to be the case for the remaining two notions. There is an important case, however, in which second order differentiability of f at x may be upgraded to yield the stronger norm Gâteaux differentiability of ∂f at x. This was already observed by Aronszajn [3, Lemma 2.1] in the case of a Lipschitz operator. We include the argument here for the sake of completeness.

Proposition 3.2. *Let E be a separable reflexive Banach space, and let f be a continuous convex function on E. Suppose f is second order differentiable at $x \in E$ and ∂f is single-valued and directionally absolutely continuous in a neighbourhood of x along any line passing through x. Then ∂f is norm Gâteaux differentiable at x.*

Proof. Due to the fact that E^* is a space with the Radon Nikodým property, the result is essentially contained in [15, pp. 106, 107]. We sketch the idea for the sake of completeness.

By the assumption, $T = \partial f$ is absolutely continuous in a neighbourhood of x along any line $L = x + \mathbb{R}h$. So $\nabla_L T(x + \tau h)$ exists for almost all τ, when ∇_L denotes the differential along the line L. Therefore, for any fixed $v \in E$, we have the identity

$$\langle T(x+th)-T(x), v\rangle = \int_0^t \frac{d}{d\tau}\langle T(x+\tau h), v\rangle\, d\tau = \int_0^t \nabla_L T(x+\tau h)\, d\tau, \tag{3.11}$$

$t \in \mathbb{R}$. E being reflexive, this shows that

$$T(x+th) - T(x) = \int_0^t \nabla_L T(x+\tau h)\, d\tau \tag{3.12}$$

holds in the sense of the Pettis integral for $t \in \mathbb{R}$. But E being separable, (3.12) is actually true in the sense of a Bochner integral. This proves norm convergence of $1/t(T(x+th)-T(x))$ (as $t \to 0$) for any fixed $h \in E$. □

Example 4. Consider again the situation in Example 1, with $H = l_2$, $C = l_2^+$ the positive cone in H. Then P_C is nowhere Fréchet differentiable, but is Gâteaux differentiable at points x with $x_n \neq 0$ for all n, with convergence of the difference quotient being understood in the sense of the norm by Proposition 3.2. It is easy to see that $1/t(P_C(x+th)-P_C(x))$ also converges uniformly over all $\|h\| \leq 1$ in the weak sense. This shows that $\partial f = P_C$ may be both norm Gâteaux and weak-$*$ Fréchet differentiable, without being norm Fréchet differentiable.

We conclude this section with the following mild extension of a result by Fabian [16, Proposition 3.1].

Proposition 3.3. *Let E be a reflexive Banach space. Suppose there exists a norm $\|\cdot\|$ on E which is Lipschitz-smooth at all points $x \neq 0$. Then either*
(i) *the dual norm $\|\cdot\|^*$ on E^* is nowhere second order differentiable, or*
(ii) *E is isomorphic with a Hilbert space.*

Proof. Suppose the dual norm $\|\cdot\|^*$ is second order differentiable at some x^* satisfying $\|x^*\|^* = 1$. Let $x = \nabla^F \|\cdot\|^*(x^*) \in E$. By assumption, $\|\cdot\|$ is Lipschitz smooth at x with $\nabla^F \|\cdot\|(x) = x^*$. By [16, Proposition 2.2], the dual unit ball B^* is therefore Lipschitz exposed at x^* by x. But now we are prepared to follow almost verbally the argument presented in the proof of [16, Proposition 3.1]. Notice that the concept of second order Gâteaux differentiability, as used in [16], may with ease be replaced by our more general notion here. In fact, the existence of a first derivative at all points in a certain neighbourhood of x is not needed to give the argument. □

Remarks. (1) As a consequence of Proposition 3.3, we see that Alexandrov's theorem fails in any l_p $(1 < p < 2)$. Indeed, the dual $l_p^* = l_q$ has by $q > 2$ an everywhere Lipschitz smooth norm $\|\cdot\|_q$.

(2) Proposition 3.3 tells that, in a sense, a Banach space E and its dual E^* may not at the same time be expected to be appropriate for second order differentiability, unless E is isomorphic with some Hilbert space. Moreover, in the case of a Hilbert space, we know from Example 2 that separability is needed to give a reasonable theory.

4. Second order differentiability and measure theory

In his 1975 paper [3], Aronszajn proved an infinite dimensional version of the classical Rademacher theorem on almost everywhere differentiability of Lipschitz operators $T: \mathbb{R}^n \to \mathbb{R}^m$. To this end he defines a class of exceptional sets—henceforth referred to as the *Aronszajn class of exceptional sets* or *Aronszajn null sets*—which plays the role of the Lebesgue null sets in the infinite dimensional setting. Introducing such a class of exceptional sets becomes necessary due to the well-known fact that no Lebesgue measure exists on an infinite dimensional Banach space. The Aronszajn exceptional class $\mathbf{A}$, which may be defined on any separable Banach space E, has the following properties:

(1) It is invariant under translations, i.e., $A + x \in \mathbf{A}$ whenever $A \in \mathbf{A}$.

(2) Countable unions of sets of $\mathbf{A}$ are again in $\mathbf{A}$.
(3) No nonempty open set belongs to $\mathbf{A}$.
(4) $\mathbf{A}$ coincides with the Lebesgue null sets when E is finite dimensional.

Independently, Christensen [13] proved an infinite dimensional version of Rademacher's theorem, which is based on a larger class $\mathbf{H}$ of exceptional sets satisfying (1)–(4), the *Haar null sets*. This class can be defined on any Banach space. A third class $\mathbf{G}$ of exceptional sets satisfying (1)–(4), the *Gaussian null sets*, was discussed by Phelps [33]. Notice that $\mathbf{A} \subset \mathbf{G} \subset \mathbf{H}$.

In the light of these results, it is natural to ask whether, similarly, an infinite dimensional version of Alexandrov's theorem might be proved with the Lebesgue null sets replaced by any of these exceptional classes. As it turns out, this is not the case. Indeed, due to the fact that, by Proposition 2.2, second order differentiability implies (first order) Fréchet differentiability, any such result would automatically produce an almost everywhere type result for (first order) Fréchet differentiability of convex functions. And even these weaker types of results are known to fail. Let us consider some examples.

Example 1 (see Preiss [35, Remark 6.1]). Let E be a separable Banach space, and let μ be a finite Borel measure on E. As μ is regular, there exists an increasing sequence C_n, $n = 1, 2, \dots$, of compact convex sets such that $\mu(E \setminus \bigcup_{n=1}^{\infty} C_n) = 0$. Now let $f_n = d(\cdot, C_n)$ be the distance to the convex C_n, which is a continuous convex function. Then $f = \sum_{n=1}^{\infty} 2^{-n} f_n$ is continuous convex and fails to be Fréchet differentiable at the points of $\bigcup_{n=1}^{\infty} C_n$. This follows with some calculation from the fact that each f_n is Fréchet differentiable outside C_n but, due to $\operatorname{int} C_n = \varnothing$, fails to be Fréchet differentiable at points $x \in C_n$. So the set D_f^2 of points of second order differentiability of f is a μ-null set. Choosing for μ any *Gaussian measure* (see [33] for this notion), we derive that $E \setminus D_f^2$ is not a *Gaussian null set*, for this would require $E \setminus D_f^2$ to be null with respect to all Gaussian measures on E. As a consequence of Phelps [33], $E \setminus D_f^2$ therefore also fails to be an Aronszajn null set. It follows that an almost everywhere version of Alexandrov's theorem based on the exceptional classes of Aronszajn or Gaussian null sets is not valid in any separable Banach space.

Notice that this construction does not tell us, whether still such a result might be possible for the larger exceptional class of Haar null sets. Let us therefore consider the following.

Example 2. Consider the space c_0 of null sequences. Let C be the positive cone in c_0, and let $f = d(\cdot, C)$ be the distance from C. Then f is continuous and convex. As C has no interior points, f fails to be Fréchet differentiable at points $x \in C$. In particular, no $x \in C$ may be a point of second order differentiability of f. But notice that C is not a Haar null set in c_0. Indeed, suppose it were. Then, by definition (cf. [13]), there exists a finite, hence regular, Borel measure $\mu \neq 0$ on c_0 such that $C + x$ is a μ-null set for any translate $C + x$ of C. But notice that any compact set K in c_0 is contained in some translate $C + x$ of C, so consequently, any compact K has to be μ-null. Indeed, recall that K is contained in a set $\{y \in c_0 : |y_n| \leq x_n,\ n = 1, 2, \dots\}$ for a certain $x \in c_0$, so $K \subseteq C - x$. But $\mu(K) = 0$ for any compact K means $\mu \equiv 0$. This contradiction shows that, in c_0, no measure theoretic version

of Alexandrov's theorem based on the exceptional class of Haar null sets is possible.

Choosing for C any closed convex set with empty interior and not a Haar null set would work equally well here. Unfortunately, we do not have an explicit example of a set C of this type in separable Hilbert space l_2. In particular, the positive cone in l_2 does not work, for it is Haar null. Yet, it is quite unconceivable that no set C of this type should exist in l_2, which makes it at least very unlikely that in l_2 a measure theoretic version of Alexandrov's theorem based on the Haar null sets might be possible.

Example 3. We show that in contrast with the situation in c_0, the positive cone $C = l_p^+$ is Haar null in l_p for all $1 \leq p < \infty$. Fix an independent sequence of integrable functions $f_n : [0, 1] \to [0, 1]$ such that

$$m\{f_n = 0\} = 1 - 1/n, \qquad m\{f_n = 1/n^{1/p}\} = 1/n$$

($m =$ Lebesgue measure). Define a Borel measure μ on l_p by

$$\mu(A) = m\{t \in [0, 1] : (f_n(t))_{n=1}^{\infty} \in A\}.$$

We check that μ is a probability measure on l_p by showing that $(f_n(t)) \in l_p$ for almost all t. Indeed, by Levi's theorem we have

$$\int_0^1 \sum_{n=1}^{\infty} |f_n(t)|^p \, dt = \sum_{n=1}^{\infty} \int_0^1 |f_n(t)|^p \, dt = \sum_{n=1}^{\infty} \left(\frac{1}{n^{1/p}}\right)^p \frac{1}{n} = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

proving $\sum |f_n|^p < \infty$ a.e. Hence $\mu(l_p) = 1$.

Let us show that $\mu(x - C) = 0$ for every translate $x - C$ of $-C$. Fixing $x = (x_n) \in C$, we have

$$\mu(x - C) = m\{t : f_n(t) \leq x_n \text{ for all } n\} = \prod_{n=1}^{\infty} m\{t : f_n(t) \leq x_n\}$$

$$= \prod_{x_n^p < 1/n} \left(1 - \frac{1}{n}\right) =: \theta.$$

Clearly $\theta = 0$ if $\rho = \sum_{x_n^p < 1/n} 1/n = \infty$. But $\rho < \infty$ implies

$$\infty = \sum_{x_n^p \geq 1/n} \frac{1}{n} \leq \sum_{x_n^p \geq 1/n} x_n^p \leq \sum_{n=1}^{\infty} x_n^p < \infty,$$

a contradiction. This proves $\mu(x-C) = 0$ for $x \in C$. For $x \notin C$, $\mu(x-C) = 0$ is clear. Hence the negative cone $-C$ is Haar null in l_p, and consequently so is C. Notice that C is not Gaussian null, and hence not Aronszajn null, for C contains a Hilbert cube (cf. [33, 3]).

The discussion so far suggests that there is little hope for measure theoretic arguments, in the spirit of Aronszjan's approach, being helpful in second order theory. Yet, the situation is not so disastrous after all, at least not when we ask for the mere existence of points of second order differentiability.

Theorem 4.1. *Let E be a separable Banach space, and let f be a continuous convex function on E. Suppose the set L_f of Lipschitz smooth points of f*

is not an Aronszajn null set. Then also the set D_f^2 of points of second order differentiability of f is not Aronszajn null. In particular, $D_f^2 \neq \varnothing$ in this case.

Proof. By the definition of the Aronszajn exceptional class **A** (cf. [3]), there exists a dense sequence $\{e_n\}$ in E such that the set L_f of Lipschitz smooth points of f is not in the class $\mathbf{A}\{e_n\}$. According to [3, Corollary I, 1], we may assume that $\{e_n\}$ is closed under taking rational linear combinations.

Let $k, m, n \in \mathbb{N}$, and let $A_{k,m,n}$ be the set of all $x \in E$ such that either (i) one of the limits

$$\begin{aligned} f''(x; e_n, e_k) &= \lim_{t\to 0} t^{-1}\langle \partial f(x+te_n) - \partial f(x), e_k\rangle, \\ f''(x; e_m, e_k) &= \lim_{t\to 0} t^{-1}\langle \partial f(x+te_m) - \partial f(x), e_k\rangle, \\ f''(x; e_n+e_m, e_k) &= \lim_{t\to 0} t^{-1}\langle \partial f(x+t(e_n+e_m)) - \partial f(x), e_k\rangle \end{aligned} \tag{4.1}$$

fails to exist, or (ii) they all exist, but

$$f''(x; e_n, e_k) + f''(x; e_m, e_k) \neq f''(x; e_n+e_m, e_k). \tag{4.2}$$

We show that $A_{k,m,n}$ is in the exceptional class $\mathbf{A}\{e_k, e_m, e_n\}$. By definition, this means that we have to show that for any $x \in E$,

$$A_{k,m,n} \cap (x+S) \tag{4.3}$$

is a Lebesgue null set in the three-dimensional linear manifold $x+S$, where $S = \operatorname{lin}\{e_k, e_m, e_n\}$. But observe that Alexandrov's theorem is true in $S+x$, so the set of points $y \in S+x$ such that $f|S+x$ fails to be second order differentiable at y (in $S+x$) is Lebesgue null in $S+x$. Now notice that any $y \in A_{k,m,n} \cap (x+S)$ certainly is a point where $f|S+x$ fails to be second order differentiable in $S+x$. Here, in (4.1), we make use of the fact that, for any $e, e' \in S$, we have $\langle \partial f(x), e\rangle = \langle \partial_S f(x), e\rangle$, $\langle \partial f(x+te'), e\rangle = \langle \partial_S f(x+te'), e\rangle$, with ∂_S standing for the subdifferential operator in $S+x$. So the set (4.3) is Lebesgue null in $S+x$, proving $A_{k,m,n} \in \mathbf{A}\{e_k, e_m, e_n\}$.

Letting $A = \bigcup_{k,m,n} A_{k,m,n}$ therefore defines a set A in the exceptional class $\mathbf{A}\{e_n\}$ (cf. [3, Proposition I, 1]). Since $L_f \notin \mathbf{A}\{e_n\}$, we must have $B = L_f \backslash A \notin \mathbf{A}\{e_n\}$. In particular, this set is nonempty. We now show that f is second order differentiable at the points $x \in B$. This establishes the statement of the theorem.

By the definition of the sets $A_{k,m,n}$, the difference quotients

$$t^{-1}\langle \partial f(x+te_n) - \partial f(x), e_n\rangle, \qquad n = 1, 2, \ldots, \tag{4.4}$$

are convergent (as $t \to 0$). Invoking Bangert's result [3, Lemma 4.8], we deduce that, for any n, the second order difference quotient $\Delta_t(e_n) = \Delta_{f,x,y^*,t}(e_n)$ of f at x converges in direction e_n (as $t \to 0$), with $y^* = \nabla^F f(x)$. In other terms the family Δ_t is pointwise convergent on the dense set $\{e_n\}$. On the other hand, x is known to be a point of Lipschitz smoothness of f, so according to the argument leading to Proposition 2.2, (Δ_t) is uniformly bounded above $(0 < |t| \leq 1)$ on some neighbourhood of 0, i.e.,

$$\Delta_t(h) \leq c \tag{4.5}$$

for some $c > 0$, some $\delta > 0$, all $0 < |t| \leq 1$ and all $\|h\| \leq \delta$.

We deduce from (4.5) that the family Δ_t is equi-Lipschitz in a neighbourhood of 0, i.e., we have an estimate of the form

$$|\Delta_t(h) - \Delta_t(k)| \leq C\|h - k\| \tag{4.6}$$

for some $C > 0$, all $0 < |t| \leq 1$, and all $\|h\| \leq \delta/2$, $\|k\| \leq \delta/2$. Indeed, it follows from (4.5) and the subgradient inequality that, for any $\|h\| \leq \delta/2$, $\|v\| \leq \delta/2$, $0 < |t| \leq 1$, and $y_t^* \in \partial\Delta_t(h)$, we have the estimate

$$\langle y_t^*, v\rangle \leq \Delta_t(h+v) - \Delta_t(h) \leq c, \tag{4.7}$$

proving the uniform estimate $\|\partial\Delta_t(h)\| \leq 2c/\delta =: C$. But then, for any $\|h\| \leq \delta/2$, $\|k\| \leq \delta/2$, $0 < |t| \leq 1$, the mean value theorem provides a vector h_t on the segment $[h, k]$ such that

$$\Delta_t(h) - \Delta_t(k) = \langle \partial\Delta_t(h_t), h - k\rangle \leq C\|h - k\|,$$

proving (4.6).

Now pointwise convergence of Δ_t on the dense set $\{e_n\}$ combined with the equi-Lipschitz condition (4.6) readily implies pointwise convergence everywhere. Indeed, let $e \in E$, $\|e\| \leq \delta/2$ be fixed. Then, for $\|e_n\| \leq \delta/2$, we have

$$\begin{aligned} |\Delta_t(e) - \Delta_s(e)| &\leq |\Delta_t(e) - \Delta_t(e_n)| + |\Delta_t(e_n) - \Delta_s(e_n)| + |\Delta_s(e_n) - \Delta_s(e)| \\ &\leq C\|e - e_n\| + |\Delta_t(e_n) - \Delta_s(e_n)| + C\|e - e_n\|. \end{aligned} \tag{4.8}$$

Fixing $\varepsilon > 0$, we first find $\|e_n\| \leq \delta/2$ such that $2C\|e - e_n\| < \varepsilon/2$. Then we choose $t_0 > 0$ in such a way that

$$|\Delta_t(e_n) - \Delta_s(e_n)| < \varepsilon/2$$

for all $0 < |t|, |s| \leq t_0$. This proves pointwise convergence of (Δ_t) on $\|e\| \leq \delta/2$, and by formula (2.4), implies pointwise convergence everywhere. Let us denote the corresponding limit by q. It remains to check that q is purely quadratic. This is done by looking again at the first order difference quotient of ∂f.

Notice that, due to the Lipschitz smoothness of f at x, we have an estimate of the form

$$|t^{-1}\langle \partial f(x + te_n) - \partial f(x), e_m\rangle| \leq C\|e_n\|\,\|e_m\| \tag{4.9}$$

for some $C > 0$, $t_0 > 0$, all $n, m \in \mathbb{N}$ and all $0 < |t| \leq t_0$ (cf. [16, Proposition 2.1(iii)]). Therefore the formula

$$\langle Te_n, e_m\rangle = \lim_{t\to 0}\langle t^{-1}(\partial f(x + te_n) - \partial f(x)), e_m\rangle \tag{4.10}$$

defines a bounded linear operator $T : E \to E^*$. Indeed, (4.10) first defines a linear $T\colon [\{e_n\}] \to [\{e_n\}]^* \cong E^*$, which by (4.9) has norm $\|T\| \leq C$ on $[\{e_n\}]$, and so by continuity extends to a bounded linear operator on E. Notice that, in (4.10), linearity in the e_n-coordinate follows since $x \notin A$, whence the phenomenon (4.2) is not allowed to occur. But now it remains to observe that, for $n \in \mathbb{N}$,

$$\Delta_t(e_n) \to q(e_n) = \tfrac{1}{2}\langle Te_n, e_n\rangle, \qquad t \to 0.$$

The $\{e_n\}$ being dense, this proves the purely quadratic representation $q(h) = \frac{1}{2}\langle Th, h\rangle$, so the proof is complete. □

Remarks. (1) It might be somewhat surprising that the finite dimensional version of Alexandrov's theorem is needed to obtain this result. However, the same is true in Aronszajn's approach, where the classical Rademacher theorem has to be invoked to prove the infinite dimensional version. Notice that in our case the one-dimensional version of Alexandrov's theorem would be sufficient if we wished to prove only convergence of the second difference quotient. Proving the limit q to be purely quadratic requires a more sophisticated reasoning.

(2) Notice that Theorem 4.1 does not guarantee the existence of points of second order differentiability. Even in separable Hilbert space we do not know whether $L_f \notin \mathbf{A}$ in general. Clearly, if we had $L_f \in \mathbf{A}$, Theorem 4.1 would not provide *any* information.

(3) On the other hand, we do not know of any example where the set L_f of Lipschitz smooth points *is* Haar null. Observe that the construction in Example 1 giving $\mu(L_f) = 0$ does *not* mean that L_f is Haar null, for this would require $x + L_f$ to be a μ-null set for every translate $x + L_f$ of L_f. We conjecture that the following is true in separable Hilbert space: $L_f \notin \mathbf{A}$. In §7 we will substantiate this for convex integral functionals f of the form

$$f(x) = \int_\Omega \phi(x(\omega), \omega)\mu(d\omega),$$

$x \in L^2_{\mathbb{R}^n}(\Omega)$, with $\phi : \mathbb{R}^n \times \Omega \to \mathbb{R}$ measurable and convex in the first coordinate.

We end this section with the following way of stating Theorem 4.1.

Corollary 4.2. *For any continuous convex function f on a separable Banach space E, the set $L_f \backslash D^2_f$ of Lipschitz smooth points which fail to be points of second order differentiability is an Aronszajn null set.*

This way of stating the result is interesting insofar as it badly fails in nonseparable Hilbert space. Indeed, let us consider

Example 4. Let Γ be uncountable, $H = l_2(\Gamma)$, and let C be the positive cone in H. Let P_C be the projection onto C, with f denoting the continuous convex function (3.10) having $\nabla^F f = P_C$. Then f is nowhere second order differentiable, yet is everywhere Lipschitz smooth. So here we have $L_f \backslash D^2_f = H$, which strongly contrasts Corollary 4.2. Notice that Lipschitz smoothness of f follows via [16, Proposition 2.1(iii)] when we recall that P_C is nonexpansive.

Remark. Observe that there is no chance for a category analogue of Alexandrov's theorem. Even for a convex function f on the real line, the set L_f may be of the first category. For an example see [16, §2, Example].

5. Smooth approximation of convex functions

It is well known that a convex function in $\mathbb{R}^n$ may be approximated uniformly on bounded sets by convex C^∞-functions. Recently, Vanderwerff [42] proved that a corresponding result does not hold in infinite dimensions, by constructing a norm on separable Hilbert space which cannot be approximated uniformly on bounded sets by even nonconvex C^2-functions having uniformly continuous second derivatives. It seems to be an open question whether a uniform C^2 approximation for convex functions on bounded sets in Hilbert space is possible, but the quoted result strongly suggests that uniform approximation by convex $C^{1,1}$-functions is the best we can hope for (see also [29]). Here

the Hessian $\nabla^2 f(x)$ of f at x is nondegenerate, it is known that the conjugate function f^* is second order differentiable at $y = \nabla f(x)$, its Hessian $\nabla^2 f^*(y)$ at y being the inverse of $\nabla^2 f(x)$, i.e.,

$$\nabla^2 f(x) = (\nabla^2 f^*(y))^{-1}. \tag{6.2}$$

Naturally, to prove this, we consider the second difference quotients of f and f^*. These are again conjugates, i.e., we have

$$(\Delta_{f,x,y,t})^* = \Delta_{f^*,y,x,t}, \tag{6.3}$$

where $y \in \partial f(x)$, $x \in \partial f^*(y)$. Now $\Delta_{f,x,y,t}$ converges pointwise everywhere to the purely quadratic limit function $q(h) = \frac{1}{2}\langle \nabla^2 f(x)h, h\rangle$. We may then deduce that the conjugate $\Delta^*_{f,x,y,t}$ converges pointwise everywhere to the limit q^* (see Proposition 6.3), hence, by (6.3), f^* is second order differentiable at y, with the identity (6.2) showing up. As it turns out, however, this argument relies strongly on the fact that $\nabla^2 f(x)$, or rather q, is nondegenerate. Yet, having the duality relation (6.3) at our disposal, we would like to maintain some duality of type (6.2) even when $\nabla^2 f(x)$ is degenerate. The function q then being degenerate, we would still expect convergence of the conjugate second order difference quotient $\Delta_{f^*,y,x,t}$ to the conjugate limit q^*, which in this case would be a purely quadratic function defined on a lower dimensional subspace of $\mathbb{R}^n$, the corresponding generalized Hessian then being a pseudo-inverse of $\nabla^2 f(x)$. Unfortunately, this generalized duality (6.2) is not valid in all cases (Example 1). The reason for the failure is that the Young-Fenchel conjugation $\phi \to \phi^*$ is not continuous with respect to pointwise convergence, i.e., pointwise convergence of a sequence (ϕ_n) does not necessarily imply pointwise convergence of the conjugate sequence (ϕ_n^*). This led Wijsman [43] to introduce a notion of convergence for convex functions which has the property of forcing the Young-Fenchel conjugation $\phi \to \phi^*$ to be continuous. This is widely known as *epiconvergence*. It was extended to the case of infinite dimensions by Mosco [28], where it is known as *Mosco convergence*. In order to maintain a duality-type relation (6.2) for second derivatives, we are therefore led to study second order differentiability concepts based on Mosco convergence rather than pointwise convergence of the second difference quotients. In finite dimensions, a programme of building a generalized second derivative based on epiconvergence has been initiated by Rockafellar [36, 37, 39]. Here we focus mainly on the case of infinite dimensions.

A function $q \in \Gamma_0(H)$, where $\Gamma_0(H)$ denotes the cone of proper lower semicontinuous convex functions on the Hilbert space H, will be called *purely quadratic* if there exists a closed symmetric and positive linear operator T, with domain $D(T) = D(q)$, such that q has a representation of the form

$$q(h) = \tfrac{1}{2}\langle Th, h\rangle, \qquad h \in D(q). \tag{6.4}$$

As in finite dimensions (cf. [36]), it can be shown that a function $q \in \Gamma_0(H)$ is purely quadratic if and only if the graph ∂q of its subdifferential operator is a closed linear subspace of $H \times H$. Indeed, the necessity of this condition being clear from (6.4), let us assume that ∂q is a closed linear subspace of $H \times H$. Then $\partial q(0)$ is a closed linear subspace of H. Hence, for every $x \in D(q)$ we find a unique vector $Tx \in \partial q(0)^\perp$ such that

$$\partial q(x) = Tx + \partial q(0).$$

Hence T is linear on $D(q)$ with values in $\partial q(0)^{\perp}$. Closedness of ∂q implies closedness of T. To check the identity (6.4), we use the fact that q is quadratic on $D(q)$. As $Th \in \partial q(h)$, we have

$$\langle Th, ph\rangle \leq q(h+ph) - q(h) = (2\rho + \rho^2)q(h)$$

for every $\rho \in \mathbb{R}$. Dividing by positive (resp. negative) ρ, and then passing to the limit $\rho \to 0$, we obtain the desired representation (6.4). In particular, T is positive. Finally, to check the symmetry of T, we use (6.4) to write

$$\langle Th, v\rangle \leq t^{-1}(q(h+tv) - q(h)) = \tfrac{1}{2}(\langle Th, v\rangle + \langle Tv, h\rangle + t\langle Tv, v\rangle)$$

for all $h, v \in D(q)$. As $t \to 0$, this proves the symmetry of T, and hence shows T to be uniquely determined by q.

Let us now briefly discuss the notion of Mosco convergence for functions in $\Gamma_0(H)$. Instead of stating the definition as presented in [4, §3; 5 or 40], we prefer to give the following equivalent, but for our purposes more workable, description. For a proof of the equivalence see [4, 3.0].

A sequence (ϕ_n) of functions in $\Gamma_0(H)$ is Mosco convergent to the limit $\phi \in \Gamma_0(H)$ if and only if the following two conditions are satisfied:

(α) *For any $x \in D(\phi)$ there exists $x_n \in D(\phi_n)$, $x_n \to x$ (norm) such that $\phi(x) \geq \overline{\lim}_{n\to\infty} \phi_n(x_n)$.*

(β) *For any strictly increasing sequence n_k of indices and any sequence (x_k) with $x_k \rightharpoonup x$ (weakly) we have $\phi(x) \leq \underline{\lim}_{k\to\infty} \phi_{n_k}(x_k)$.*

It is in the nature of Mosco convergence that it has an equivalent description in terms of convergence of the graphs $\partial\phi_n$ of the subdifferential operators, see, e.g., [4].

Definition 6.1. Let f be a continuous convex function on a Hilbert space H. Let $x \in H$. Then f is said to have a *generalized second derivative* or to be *second order differentiable in the generalized sense* at x relative to $y \in \partial f(x)$ if there exists a purely quadratic function $q \in \Gamma_0(H)$ such that the second order difference quotient $\Delta_{f,x,y,t}$ converges to q (as $t \to 0$) in the Mosco sense. The closed symmetric and positive linear operator T corresponding with q in the representation (6.4) is called the *generalized Hessian* of f at x relative to $y \in \partial f(x)$.

Remarks. (1) Take $f : \mathbb{R}^2 \to \mathbb{R}$, $f(\xi, \eta) = |\xi|$. Then f is not differentiable at $(0,0)$, $\partial f(0,0) = \{(\alpha, 0): |\alpha| \leq 1\}$. Now for $y = (\alpha, 0)$, $|\alpha| < 1$, the second difference quotient $\Delta_{f,0,y,t}$ converges both pointwise and in the Mosco sense to the limit q with $q(0,\eta) = 0$, $q(\xi,\eta) = \infty$ for $\xi \neq 0$. So the existence of a generalized second derivative q does not necessarily imply first order differentiability. However, in general, it is easy to see that the existence of q implies that all subgradients agree on $D(q)$, i.e., $\langle y, h\rangle = \langle y_1, h\rangle$ for all $h \in D(q)$ and all $y_1 \in \partial f(x)$. In particular, if the domain $D(q)$ of q is dense, this implies Gâteaux differentiability of f at x.

(2) Our notion of a generalized second derivative differs slightly from Rockafellar's [37, 39] since we consider two-sided limits and require the limit function q to be purely quadratic. More significantly, Rockafellar also considers partially defined functions f, while we restrict out attention to the fully defined

case. Clearly, new phenomena might be expected at boundary points x of the effective domain $\operatorname{dom} f$. It should be mentioned, however, that the second order behaviour of f at such points may be reduced to the fully defined case, by using a penalty-type approximation of f by fully defined functions $f_{[n]}$, where

$$f_{[n]} = f \square n\|\cdot\|, \qquad n \in \mathbb{N}. \tag{6.5}$$

The approximation (6.5) has been used by various authors for similar purposes, see [30, 31, 8, 19, 14, or 7].

(3) It is known that Mosco convergence and pointwise convergence are generally unrelated (see [4, 40] for examples). Here we show that the situation is similar for pointwise and Mosco convergence of second difference quotients. See also [32] for related examples.

Example 1. We construct a convex function f on $\mathbb{R}^2$ with $f(0,0) = 0$ which is differentiable at $(0,0)$ with $\nabla f(0,0) = (0,0)$ such that the following are true: The second difference quotient at $(0,0)$ converges both pointwise, with limit p, and in the Mosco sense, with limit q, but

$$q(0,1) < p(0,1), \qquad D(p) = D(q) = \{0\} \times \mathbb{R}. \tag{6.6}$$

Let $t_n = 2^{-n}$, $s_n = t_n^\alpha = 2^{-n\alpha}$ (for $1 < \alpha < 2$ fixed). We define f such that

$$f(0, t_n) = t_n^2 = 4^{-n}, \qquad f(s_n, t_n) = \tfrac{1}{2}t_n^2 = \tfrac{1}{2}4^{-n}. \tag{6.7}$$

For $n \geq 1$ let ε_n be the plane spanned by the points $(0, t_n, t_n^2)$, $(s_n, t_n, \frac{1}{2}t_n^2)$, $(0, t_{n+1}, t_{n+1}^2)$, and ε_{-n} the reflection of ε_n with the xz-plane. Also let δ_n be the plane spanned by the points $(s_n, t_n, \frac{1}{2}t_n^2)$, $(s_{n+1}, t_{n+1}, \frac{1}{2}t_{n+1}^2)$, and $(s_n, t_{n+1}, 3t_{n+1}^2)$, and let δ_{-n} be the reflection of δ_n with the xz-plane. Let f be defined by taking as its graph the upper envelope of all the planes $\varepsilon_{\pm n}$ and $\delta_{\pm n}$. It is easy to check that the slopes of these planes are uniformly bounded, so f is a globally defined convex function.

To check (6.6), notice that $p(0,1) = \lim_{n\to\infty} f(t_n(0,1))/t_n^2 = 1$, while, by condition (β) for Mosco convergence, $q(0,1) \leq \lim_{n\to\infty} f(s_n, t_n)/(s_n^2 + t_n^2) = \frac{1}{2}$. It follows from Proposition 6.1(4) that $D(q) \neq \mathbb{R}^2$; hence $D(q) = D(p) = \{0\} \times \mathbb{R}$.

Example 2. Let $f : l_2 \to \mathbb{R}$ be a convex function of the form

$$f(x) = \sum_{n=1}^{\infty} f_n(x_n), \qquad x = (x_n) \in l_2, \tag{6.8}$$

where the f_n are convex functions on the line. Suppose that $x = (x_n) \in l_2$ is a point such that each f_n is second order differentiable at x_n. Then f has generalized second derivative q at x given by

$$q(h) = \frac{1}{2}\sum_{n=1}^{\infty} f_n''(x_n)h_n^2, \tag{6.9}$$

with $D(q)$ the set of all $h \in l_2$ for which (6.9) converges. This may be checked either directly or using Proposition 6.4. Notice that $D(q) = l_2$ if and only if $f_n''(x_n)$ is bounded. But even in this case, q need not be a limit in the pointwise sense. To see this, let us specify (6.8) by taking $f_n(\xi) = n^{-\alpha}|\xi|$ $(\frac{1}{2} < \alpha < 1)$.

Consider $x = (n^{-2}) \in l_2$. Then (6.9) gives $q \equiv 0$ as the generalized second derivative at x. However, the second difference quotient fails to converge pointwise to q. In fact, if $\Delta_t \to 0$ pointwise, then f had to be Lipschitz smooth at x by Proposition 2.2. We show that this is not the case.

Let $h \in l_2$, then

$$f(x+h) - f(x) - \langle \nabla^F f(x), h\rangle = \sum_{n=1}^{\infty} n^{-\alpha}(|x_n + h_n| - x_n - h_n) \tag{6.10}$$
$$= \sum_{n^{-2}+h_n<0} (-2n^{-\alpha-2} - 2n^{-\alpha}h_n).$$

Assuming that f Lipschitz smooth at $x = (n^{-2})$, we would find $C > 0$ and $\delta > 0$ such that (6.10) is majorized by $C\|h\|^2$ for all h having $\|h\| \le \delta$.

Let $\delta > 0$ be fixed. Define $h^\delta \in l_2$ by $h_n^\delta = 0$ for $n \le \delta^{-2}$, $h_n^\delta = -n^{-1}$ for $n > \delta^{-2}$. Then $\|h^\delta\|^2 \doteq \delta^2$. Inserting h^δ in (6.10), and observing that $n^{-2} + h_n^\delta < 0$ for $n > \delta^{-2}$, we obtain the term

$$\sum_{n>\delta^{-2}} (-2n^{-\alpha-2} + 2n^{-\alpha-1}) \doteq \delta^{2\alpha} - \delta^{2\alpha+2},$$

which by $\alpha < 1$ is certainly not of the form $O(\delta^2)$, $\delta \to 0$. So f is not Lipschitz smooth at x.

Notation. In the following we will use the notation $x \in GD_f^2$ if f has a generalized second derivative q at x satisfying $D(q) = H$. Also we recall the notations D_f^2 for the set of points of second order differentiability of f (Definition 2.1), and L_f for the set of Lipschitz smooth points of f (see §2).

Our first result gives conditions under which Mosco convergence of the second difference quotient may be improved to pointwise convergence.

Proposition 6.1. *Let f be a continuous convex function on a Hilbert space H. Let $x \in H$, $y \in \partial f(x)$, and suppose the second difference quotient $\Delta_{f,x,y,t}$ converges to the limit $q \in \Gamma_0(H)$ in the Mosco sense. Then*

(1) $\Delta_{f,x,y,t}(h) \to \infty = q(h)$ *for* $h \notin D(q)$.

(2) *Suppose for some $h \in D(q)$ there exist $y_t \in \partial f(x+th)$ such that*

$$\|t^{-1}(y_t - y)\| \le C \tag{6.11}$$

for some $C > 0$ and small t. Then $\Delta_{f,x,y,t}(h) \to q(h)$ $(t \to 0)$.

(3) *If f is Lipschitz smooth at x, then $\Delta_{f,x,y,t}$ converges pointwise everywhere to q. In other terms, $L_f \cap GD_f^2 \subset D_f^2$.*

(4) *In finite dimensions, $H = \mathbb{R}^n$, if $D(q) = \mathbb{R}^n$, then $\Delta_{f,x,y,t}$ converges pointwise everywhere. In other words, $D_f^2 = GD_f^2$ in $\mathbb{R}^n$.*

Proof. Statement (1) is clear from condition (β) of Mosco convergence, when applied to the constant sequence h. Statement (3) follows from (2) using [16, Proposition 2.1(iii)], which tells that for a Lipschitz smooth point x of f, (6.11) is met uniformly over all $\|h\| \le 1$. Statement (4) is the second half of [37, Proposition 2.3]. It might be obtained directly from (2), using an Ascoli-type argument as in [4, Remarque 1.12, $\Rightarrow$].

Let us prove (2). Let $h \in D(q)$ satisfying (6.11) be fixed. By condition (β) we have $q(h) \le \underline{\lim}_{t\to 0} \Delta_t(h)$, where $\Delta_t = \Delta_{f,x,y,t}$. So it remains to prove the reverse estimate $\overline{\lim}_{t\to 0} \Delta_t(h) \le q(h)$.

Using condition (α) of Mosco convergence, we find $h_t \to h$ (norm) such that $\overline{\lim}\,\Delta_t(h_t) \le q(h)$. Choosing $y_t \in \partial f(x+th)$ as in (6.11), we have

$$t^{-1}(y_t - y) \in t^{-1}(\partial f(x+th) - y) = \partial\Delta_t(h);$$

hence the subgradient inequality gives

$$\langle t^{-1}(y_t - y),\, h_t - h\rangle \le \Delta_t(h_t) - \Delta_t(h). \tag{6.12}$$

By (6.11), the left-hand side of (6.12) tends to 0; hence $\overline{\lim}\,\Delta_t(h) \le \overline{\lim}\,\Delta_t(h_t) \le q(h)$. This proves statement (2). □

Notice that Example 2 shows that (4) above fails in infinite dimensions.

Let us now ask for conditions under which conversely pointwise convergence of Δ_t entails Mosco convergence.

Proposition 6.2. *Let f be a continuous convex function on a separable Hilbert space. Let $x \in H$, $y \in \partial f(x)$, and suppose $\Delta_{f,x,y,t}$ converges pointwise everywhere to a limit $p \in \Gamma_0(H)$.*

(1) If f is Lipschitz smooth at x, then also $\Delta_{f,x,y,t} \to p$ in the Mosco sense.

(2) Any point of second order differentiability is also a point of generalized second order differentiability. In particular, $D_f^2 = L_f \cap GD_f^2$.

Proof. Statement (2) is immediate from (1), since second order differentiability implies Lipschitz smoothness (cf. §2).

To prove statement (1), it suffices to show that $\Delta_t = \Delta_{f,x,y,t}$ *has* a Mosco limit, say q. For then Proposition 6.1(3) shows $p = q$. We prove that every sequence $t_n \to 0$ has a subsequence t_{n_k} such that $\Delta_{t_{n_k}}$ converges to some limit q in the Mosco sense. Then $p = q$ by Proposition 6.1(3). Let us fix $t_n \to 0$.

According to [4, Proposition 6.1], the cone $\Gamma_0(H)$ is a complete metric space in its Mosco topology when endowed with the metric

$$d(\phi, \psi) = |\phi_1(0) - \psi_1(0)| + \sum_{n=0}^{\infty} 2^{-n} \min[1\,,\, \|J_\phi(h_n) - J_\psi(h_n)\|]. \tag{6.13}$$

Here $\{h_n\}$ is any fixed dense sequence in H, and $\phi_1 = \phi \,\square\, \frac{1}{2}\|\cdot\|^2$, $\psi_1 = \psi \,\square\, \frac{1}{2}\|\cdot\|^2$, while J_ϕ, J_ψ are the resolvents (5.3) of parameter $\lambda = 1$, i.e., $J_\phi = (\mathrm{id} + \partial\phi)^{-1}$, $J_\psi = (\mathrm{id} + \partial\psi)^{-1}$. By the definition of the second difference quotient we have $(\Delta_t)_1(0) = (\Delta_t \,\square\, \frac{1}{2}\|\cdot\|^2)(0) = 0$, so the first term in (6.13) cancels for distances $d(\Delta_{t_n}, \Delta_{t_m})$. Next observe that since each $J_t := J_{\Delta_t}$ is nonexpansive, we have (using (5.5))

$$\|J_{t_n}(h_j)\| = \|J_{t_n}(h_j) - J_{t_n}(0)\| \le \|h_j\|, \tag{6.14}$$

which proves that, for any j, the sequence $(J_{t_n}(h_j))_{n=1}^{\infty}$ is bounded. Using a diagonal procedure, we may therefore extract a subsequence t_{n_k} such that $J_{t_{n_k}}(h_j)$ converges $(k \to \infty)$ for $j = 1, 2, \ldots$. By (6.13), this means $d(\Delta_{t_{n_k}}, \Delta_{t_{n_l}}) \to 0$ as $k, l \to \infty$. So $\{\Delta_{t_{n_k}}\}$ is Mosco-Cauchy, and hence has a Mosco limit q. This completes our argument. □

Let f be continuous convex on a separable Hilbert space H. As we are interested in the second order behaviour of f, we lose no information on perturbing by a C^2-function. Indeed, consider the perturbed function

$$g = f + \tfrac{1}{2}\|\cdot\|^2. \tag{6.15}$$

In the case of second order differentiability (Definition 2.1), it is immediate that the properties of f and g are equally good. Concerning generalized second order differentiability, the same is true, but we have to be more careful. Observe that the second difference quotients of f and g are related by the identity

$$\Delta_{g,x,x+y,t} = \Delta_{f,x,y,t} + \tfrac{1}{2}\|\cdot\|^2, \tag{6.16}$$

where $y \in \partial f(x)$, $x+y \in \partial g(x)$. Based on conditions (α), (β), we may now check using (6.16) that $\Delta_{f,x,y,t} \to q$ in the Mosco sense if and only if $\Delta_{g,x,x+y,t} \to q_0 = q + \frac{1}{2}\|\cdot\|^2$ in the Mosco sense, which proves our claim.

The perturbation (6.15) makes duality much easier. Indeed, $g^* = f^* \square \frac{1}{2}\|\cdot\|^2$ is now C^1 and even Lipschitz smooth according to (5.1), (5.2), (5.4). This gives the following dual description of generalized second order differentiability.

Proposition 6.3. *Let f be a continuous convex function on a separable Hilbert space. Let $y \in \partial f(x)$. Then the following are equivalent:*

(1) *f has generalized second derivative q at x with respect to y.*
(2) *$g = f + \frac{1}{2}\|\cdot\|^2$ has generalized second derivative $q_0 = q + \frac{1}{2}\|\cdot\|^2$ at x with respect to $x+y \in \partial g(x)$.*
(3) *$g^* = f^* \square \frac{1}{2}\|\cdot\|^2$ is second order differentiable at $x+y$ (with $\nabla^F g^*(x+y) = x$).*
(4) *$\nabla^F g^*$ is (first order) norm Gâteaux differentiable at $x+y$.*
(5) *$J_{f^*} = (\mathrm{id} + \partial f^*)^{-1}$ is (first order) norm Gâteaux differentiable at $x+y$.*

Proof. The equivalence of (1) and (2) was observed above.The equivalence of (2) and (3) follows from the fact that Mosco convergence is invariant under Young-Fenchel conjugation (cf. [4, 40]), so that Mosco convergence of $\Delta_{g,x,x+y,t}$ is equivalent to Mosco convergence of $(\Delta_{g,x,x+y,t})^* = \Delta_{g^*,x+y,x,t}$. But g^* is Lipschitz smooth, so Propositions 6.1, 6.2 show that Mosco convergence and pointwise convergence are equivalent for $\Delta_{g^*,x+y,x,t}$.

The equivalence of (3) and (4) follows from the fact that second order differentiability is equivalent to first order Gâteaux differentiability of the derivative by Proposition 4.3. Equivalence of (4) and (5) is immediate since $\nabla g^* = \mathrm{id} - J_{f^*}$ (see (5.2)). □

As a consequence of Proposition 6.3, we may now check the representation (6.9) of the generalized second derivative of a function (6.8). This may be considered as a special case of the following.

Example 3 (Integral functionals). Let $(\Omega, \mathbf{A}, \mu)$ be a measure space, and let $H = L^2_{\mathbb{R}^n}(\Omega, \mathbf{A}, \mu)$. To apply our present theory, we need to assume that H is separable, which is satisfied, e.g., when $\mathbf{A}$ is countably generated and μ is σ-finite. Now let $\phi : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be a measurable function which is convex in the first argument. Then

$$f(x) = \int_\Omega \phi(x(\tau), \tau)\, d\mu(\tau) \tag{6.17}$$

defines a closed convex integral functional f on H. For convenience we assume that f is finite everywhere and hence continuous. This requires a growth condition on ϕ as for instance given in [38, §2]. See this reference for some

Moreover, in these cases, $q_n \to q_0 = q + \frac{1}{2}\|\cdot\|^2$ $(n \to \infty)$ *pointwise and in the Mosco sense, with convergence being uniform on all bounded sets on which* f *is bounded when* q *is fully defined.*

Proof. The equivalence of (1) and (2) was proved above. Mosco convergence of q_n to q_0 is immediate from the fact that, dually, $q_n^* = q_0^* + \|\cdot\|^2/2n \to q_0^*$ in the Mosco sense. As for pointwise convergence of q_n to q_0, observe that q_n is the Yosida approximate of q_0 with parameter n, which converges pointwise to q_0 by reasoning as in Lemma 5.1. Finally, in the case $D(q_0) = H$, the statement is just Theorem 5.2(1). □

Remark. Suppose we wish to calculate the generalized second derivative $\nabla^2 f(x)$ for a point $x \in GD_f^2$. We consider $g = f + \frac{1}{2}\|\cdot\|^2$ instead. We then approximate g by taking one of its Yosida approximates $g_n = g \,\square\, n\|\cdot\|^2/2$, and evaluate $\nabla^2 g_n(x_n)$, where $x_n = x + (x + \nabla f(x))/n$. This is justified, for $g_n \to g$ uniformly on bounded sets on which f is bounded, $\nabla g_n(x) \to \nabla g(x)$ at differentiability points of g, and even $\nabla^2 g_n(x_n) \to \nabla^2 g(x)$ for points $x \in GD_f^2$ (resp. $x \in GD_g^2$). We may proceed in this way even when $x \in GD_f^2$ is not clear. We then have $x \in GD_f^2$ if and only if $x_n \in D_{g_n}^2$, $n = 1, 2, \ldots$, with $\|\nabla^2 g_n(x_n)\| \leq C < \infty$. In finite dimensions, this is therefore a test for second order differentiability of f (resp. g) at x.

We end this section by giving a short proof of Alexandrov's theorem in finite dimensions.

Theorem 6.7 (Alexandrov's theorem). *Every convex function* f *on* $\mathbb{R}^n$ *is almost everywhere second order differentiable.*

Proof. Let $g = f + \frac{1}{2}\|\cdot\|^2$. By Rademacher's theorem there exists a null set A such that $T = \nabla g^*$ is differentiable on $\mathbb{R}^n \backslash A$. Then $T(A)$ is again a null set. By Sard's theorem (cf. [17, Corollary 3.2.5]) the set

$$M = \{T(z) : \nabla T(z) \text{ exists but is degenerate}\}$$

is null, hence so is $N = M \cup T(A)$. By Proposition 6.3, f is second order differentiable in the generalized sense at every $x \in \mathbb{R}^n \backslash N$. Let q be the generalized second derivative at x with respect to y, and $q_0 = q + \frac{1}{2}\|\cdot\|^2$. As $x+y \notin A$, g^* is second order differentiable at $x+y$ with Hessian $\nabla^2 g^*(x+y) = \nabla T(x+y)$. By definition of M, $\nabla T(x+y)$ is nondegenerate, so $q_0^*(h) = \frac{1}{2}\langle \nabla T(x+y)h, h\rangle$ is nondegenerate, which means that q_0, and hence q, is fully defined. By Proposition 6.1(4), f is therefore second order differentiable at x. □

7. Dual Lipschitz smoothness

The results obtained in §6 tell us two facts. First, a continuous convex function f on separable Hilbert space is second order differentiable at x if and only if it has a generalized second derivative and if, in addition, it is Lipschitz smooth at x. Second, the shift $g = f + \frac{1}{2}\|\cdot\|^2$ allows for a dual description of generalized second order differentiability in terms of the function g^* which, being $C^{1,1}$, is almost everywhere second order differentiable by the result of Aronszajn [3]. It is therefore natural to ask for a dual description of Lipschitz

smoothness, which then, in tandem with the above, allows us to express second order differentiability of f completely in terms of the dual data.

For sublinear functions $\phi = \sup\langle C, \cdot\rangle$, Fabian [16, Proposition 2.2] presents a dual version of Lipschitz smoothness in terms of a geometric condition. Namely, ϕ is Lipschitz smooth at a point x $(\phi(x) = 1,\ y = \nabla^F\phi(x))$, if and only if the convex set C is *Lipschitz exposed* at $y \in C$ by its tangent hyperplane τ, which is to say that the surface ∂B of a Hilbert norm ball B may be fitted in between ∂C and its tangent hyperplane τ at y. In other terms, $C \subseteq B$, with B and C having common tangent hyperplane τ at y. Here our program is to obtain a dual description of Lipschitz smoothness of a general convex function f, using g^*, ∇g^*, J_{f^*}.

Proposition 7.1. *Let f be a globally Lipschitz convex function on a separable Hilbert space H. Then the following are equivalent:*

(1) *f is Lipschitz smooth at x (with $y = \nabla^F f(x)$).*

(2) *There exists $c > 0$ such that $\Delta_{f,x,y,t}(h) \le c\|h\|^2$ for all $h \in H$ and $t \neq 0$.*

(3) *There exists $\alpha > 0$ such that $\Delta_{f^*,y,x,t}(k) \ge \alpha\|k\|^2$ for all $k \in H$ and $t \neq 0$.*

(4) *The dual second difference quotient $\Delta_{f^*,t} = \Delta_{f^*,y,x,t}$ of f^* is equicoercive, i.e.,*

$$\lim_{\|k\|\to\infty} \sup_{t\neq 0} \frac{\Delta_{f^*,t}(k)}{\|k\|} = \infty. \tag{7.1}$$

Proof. (1) $\Rightarrow$ (2) By the definition of Lipschitz smoothness, we find $\tilde{c} > 0$ and $\delta > 0$ such that (with $\Delta_{f,t} := \Delta_{f,x,y,t}$) $\Delta_{f,t}(h) \le \tilde{c}$ for all $\|h\| \le 1$ and $0 < |t| \le \delta$. For $|t| \ge \delta$, $\|h\| \le 1$ we have

$$\Delta_{f,t}(h) = \frac{f(x+th) - f(x) - t\langle y, h\rangle}{t^2} \le \frac{2L}{|t|} \le \frac{2L}{\delta},$$

with $L > 0$ a global Lipschitz constant of f. So $\Delta_{f,t}(h) \le c$ for all $t \neq 0$ and $\|h\| \le 1$ with $c = \max\{\tilde{c}, 2L/\delta\}$. Clearly this gives the desired estimate (2).

(2) $\Rightarrow$ (3) Using (6.3), we have (with $\Delta_{f^*,t} = \Delta_{f^*,y,x,t}$)

$$\Delta_{f^*,t}(k) = \sup_h\{\langle k, h\rangle - \Delta_{f,t}(h)\} \ge \sup_h\{\langle k, h\rangle - c\|h\|^2\} = \frac{1}{4c}\|k\|^2, \tag{7.2}$$

which is (3) with $\alpha = 1/4c$.

The implication (3) $\Rightarrow$ (4) being obvious, let us prove (4) $\Rightarrow$ (1). By [16, Proposition 2.1(iii)], we have to show that, for any fixed $h \in H$,

$$\sup_{t\neq 0} \left\| \frac{\partial f(x+th) - \partial f(x)}{t} \right\| < \infty. \tag{7.3}$$

Let $y_t \in \partial f(x+th)$, $y = \nabla^F f(x)$, and suppose $\|t^{-1}(y_t - y)\| \to \infty$ $(t \to 0)$. Now (6.3) gives

$$\Delta_{f,t}(h) + \Delta_{f^*,t}((y_t - y)/t) = \langle (y_t - y)/t, h\rangle. \tag{7.4}$$

Dividing (7.4) by $\|t^{-1}(y_t - y)\|$ produces a bounded right-hand side, while by equicoercivity the second term on the left-hand side tends to ∞. As both terms on the left-hand side are positive, this is a contradiction. □

Our next step is to express Lipschitz smoothness of f in terms of the Lipschitz operator ∇g^*, resp. using (5.2), in terms of the proximity mapping J_{f^*}.

Proposition 7.2. *Suppose f is convex and globally Lipschitz on the separable Hilbert space H. Then the following are equivalent:*

(1) *f is Lipschitz smooth at x (with $y = \nabla^F f(x)$).*

(2) *There exists $0 < \beta < 1$ such that*

$$\langle J_{f^*}(x+y+tk) - J_{f^*}(x+y), k\rangle/t \le \beta\|k\|^2 \tag{7.5}$$

for all $t \neq 0$ and $k \in H$.

Proof. First assume that f is Lipschitz smooth at x. By Proposition 7.1(2) there exists $\tilde{c} > 0$ such that $\Delta_{f,x,y,t} \le \tilde{c}\|\cdot\|^2$ for all $t \neq 0$. Letting $g = f + \frac{1}{2}\|\cdot\|^2$, (6.16) implies $\Delta_{g,x,x+y,t} \le c\|\cdot\|^2$ for all $t \neq 0$, where $c = \tilde{c} + \frac{1}{2}$. Now the implication (2) $\to$ (3) of Proposition 7.1, which does not use the global Lipschitz condition, implies the dual estimate

$$\Delta_{g^*,t}(k) \ge \alpha\|k\|^2, \qquad k \in H, \tag{7.6}$$

for some $\alpha > 0$, where $\Delta_{g^*,t} = \Delta_{g^*,x+y,x,t}$. But observe that, on the other hand, the identity $\Delta_{g^*,t} = \Delta_{f^*,t} \square \frac{1}{2}\|\cdot\|^2$ shows that $\Delta_{g^*,t} \le \frac{1}{2}\|\cdot\|^2$, hence $0 < \alpha \le \frac{1}{2}$ in (7.6) (with equality $\alpha = \frac{1}{2}$ occurring precisely when $\Delta_{f,x,y,t} \equiv 0$, which is to say that f is affine).

Let $z = x + y$, and let $k \in H$. Applying the mean value theorem to g^* along the ray $z + \mathbb{R}_+ k$ provides $\tau = \tau(t, k) \in (0, t)$ such that

$$\alpha\langle k, k\rangle \le \Delta_{g^*,t}(k) = \frac{\langle \nabla g^*(z+\tau k) - \nabla g^*(z), k\rangle}{\tau}\frac{\tau}{t}. \tag{7.7}$$

Using (5.2), this may be recast as

$$(1 - t\alpha/\tau)\langle k, k\rangle \ge \langle J_{f^*}(z+\tau k) - J_{f^*}(z), k\rangle/\tau, \tag{7.8}$$

$t > 0$, $\tau = \tau(t, k)$. Let $\beta = 1 - \alpha$. As $t > \tau$, (7.8) is the desired estimate (7.5), with the proviso that we still have to check that $\tau = \tau(t, k)$ ranges over all positive reals when t is allowed to vary over $\mathbb{R}_+$.

To do this let $k \neq 0$ be fixed. We consider the real convex function $\gamma : t \to g^*(z + tk)$, which is C^1, for so is g^*. There are two cases to be discussed.

First assume that γ is not affine on any interval $[0, t_0]$ with $t_0 > 0$. Then the mapping $t \to \tau(t, k)$ is continuous and monotonically increasing on $\mathbb{R}_+$. Indeed, first we have to observe that $t \to \tau(t, k)$ is a well-defined function. This follows from our temporary assumption and the fact that γ is C^1. The latter also implies continuity of $\tau(\cdot, k)$. As for monotonicity, we observe that, in the defining identity

$$\gamma'(\tau(t, k)) = (\gamma(t) - \gamma(0))/t \tag{7.9}$$

for $\tau(\cdot, k)$, the difference quotient on the right-hand side is monotone as a function of t. Monotonicity of γ' then shows the monotonicity of $\tau(\cdot, k)$. Notice that, a priori, the case $s^{-1}(\gamma(s) - \gamma(0)) = t^{-1}(\gamma(t) - \gamma(0))$ for $0 < s < t$ seems possible. However, due to convexity, this forces γ to be affine on $[0, t]$, the case which was excluded.

As $\tau(\cdot, k)$ is continuous and monotone, we deduce that, in (7.8), τ ranges over all values close to 0. As for large values τ, suppose there was an upper

bound $\tau(t, k) \leq \bar{\tau} < \infty$, as $t \to \infty$. By (7.9), this would mean that γ had an asymptote with slope $\gamma'(\bar{\tau})$ as $t \to \infty$. But γ' is increasing. So we must have $\gamma'(\tau) = \gamma'(\bar{\tau})$ for $\tau \geq \bar{\tau}$, which means that γ is affine on $[\bar{\tau}, \infty)$. Hence we have

$$\gamma(t) = g^*(z + tk) = mt + b, \qquad t \geq \bar{\tau}, \tag{7.10}$$

where $m = \gamma'(\bar{\tau})$, $b = g^*(z + \bar{\tau}k) - m\bar{\tau}$. We show that this is impossible. Indeed, assume $m > 0$. Then we have

$$\begin{aligned} g\left(\frac{2m}{\|k\|^2}k\right) &\geq \sup_{t \geq \bar{\tau}} \left[\left\langle \frac{2m}{\|k\|^2}k, z + tk \right\rangle - g^*(z + tk)\right] \\ &= \sup_{t \geq \bar{\tau}} \left[\left\langle \frac{2m}{\|k\|^2}k, z \right\rangle + \frac{2m}{\|k\|^2 t}\langle k, k\rangle - mt - b\right] \\ &= \text{constant} + \sup_{t \geq \bar{\tau}} mt = \infty, \end{aligned} \tag{7.11}$$

a contradiction. For $m \leq 0$ take $g(k)$ to produce a similar contradiction. This completes the argument in the first case.

Now consider the case where γ is affine on some interval $[0, t_0]$, $t_0 > 0$. As $\tau(t, k)$ may then be any value in $(0, t)$, there is no problem for the $t \in [0, t_0]$. For convenience, we may fix $\tau(t, k)$ to equal t on $[0, t_0]$, with the previous argument then applying on $[t_0, \infty)$. This completes the proof of statement (2).

Conversely, suppose now that (7.5) is satisfied. Fix $k \in H$, $\|k\| = 1$. Writing again $\gamma(t) = g^*(z + tk)$, $z = x + y$, the assumption reads as

$$1 - (\gamma'(t) - \gamma'(0))/t \leq \beta, \qquad t \neq 0, \tag{7.12}$$

where $\gamma'(t) = \langle \nabla g^*(z + tk), k\rangle$, and where we use (5.2). Let $\alpha = 1 - \beta$. Then (7.12) is $\gamma'(t) \geq \gamma'(0) + \alpha t$. Integrating gives

$$\gamma(\tau) - \gamma(0) = \int_0^\tau \gamma'(t)\, dt \geq \gamma'(0)\tau + \frac{\alpha}{2}\tau^2,$$

for all $\tau \geq 0$, so $\Delta_{g^*, \tau}(k) \geq \alpha/2$ for all $\tau \geq 0$. Hence condition (3) in Proposition 7.1 is met, which proves that g (and so f) is Lipschitz smooth at x. Notice that the implications (3) $\Rightarrow$ (4) $\Rightarrow$ (1) in 7.1 do not use the global Lipschitz assumption, so they apply to g. □

Proposition 7.3. *Suppose f is convex and globally Lipschitz on the separable Hilbert space H. Then the following are equivalent:*

(1) *f is Lipschitz smooth at x (with $y = \nabla^F f(x)$).*

(2) *There exists $0 < \beta < 1$ and $\delta > 0$ such that*

$$\langle J_{f^*}(x + y + k) - J_{f^*}(x + y), k\rangle \leq \beta\|k\|^2 \quad \textit{for all } \|k\| \leq \delta. \tag{7.13}$$

Proof. By Proposition 7.2, Lipschitz smoothness of f at x implies that (7.13) is met globally for all $k \in H$. Conversely, we have to show that it is possible to obtain the global estimate (7.5) by starting with the local version (7.13).

Let $k \in H$ having $\|k\| = 1$ be fixed. With the notation

$$v(t) = \langle J_{f^*}(z + tk), k\rangle, \qquad z = x + y,$$

statement (2) means $v(t) - v(0) \leq \beta t$ for $|t| \leq \delta$. Let $\delta \leq t \leq 3L$, where L is a global Lipschitz constant for f. Then

$$\begin{aligned} \frac{v(t)-v(0)}{t} &= \frac{v(t)-v(\delta)}{t-\delta}\frac{t-\delta}{t} + \frac{v(\delta)-v(0)}{\delta}\frac{\delta}{t} \\ &\leq \frac{t-\delta}{t} + \beta\frac{\delta}{t} = 1 - \frac{\delta-\beta\delta}{t} \leq 1 - \frac{\delta(1-\beta)}{3L}, \end{aligned} \tag{7.14}$$

using (2) and the fact that v is nonexpansive. Finally, for $t > 3L$, we have

$$\frac{v(t)-v(0)}{t} \leq \frac{2L}{3L} = \frac{2}{3},$$

using that v is globally bounded by L. This shows that (7.5) is now satisfied for all $t \neq 0$ with β replaced by $\tilde{\beta} = \max\{\beta, 1 - \delta(1-\beta)/3L, 2/3\} < 1$, since the above estimates are met uniformly over all $\|k\| = 1$. □

As a consequence of Proposition 7.3, we obtained a localized condition (7.13) which dually characterizes Lipschitz smoothness. This suggests asking for the *best* constant $\beta > 0$ occurring in (7.13).

Definition 7.1. Let $T : H \to H$ be a monotone locally Lipschitz operator on a Hilbert space H. Then

$$b(T, z) = \lim_{\delta \searrow 0} \sup_{0<\|k\|\leq\delta} \frac{\langle T(z+k) - T(z), k\rangle}{\|k\|^2} \tag{7.15}$$

is called the *one-sided Lipschitz constant* of T at $z \in H$.

Notice that we have $b(T, z) \leq l(T, z)$, where $l(T, z)$ denotes the *Lipschitz constant* of T at z, i.e., the infimum over all Lipschitz constants of T on small neighbourhoods of z.

Definition 7.2. Let $\phi : H \to \mathbb{R}$ be continuous and convex. Then

$$c(\phi, z) = \lim_{\delta \searrow 0} \sup_{0<\|k\|\leq\delta} \frac{\phi(z+k) - \phi(z) - \langle \nabla^F \phi(z), k\rangle}{\|k\|^2} \tag{7.16}$$

is called the *Lipschitz-smooth constant* of ϕ at z. If ϕ fails to be Lipschitz smooth at z, we write $c(\phi, z) = \infty$.

In the case of a convex $C^{1,1}$ function ϕ, we may ask for the relation between $c(\phi, z)$ and $b(\nabla^F \phi, z)$. This is covered by the following

Lemma 7.4. *Let ϕ be convex and $C^{1,1}$. Then*

(1) $c(\phi, z) \leq \frac{1}{2} b(\nabla^F \phi, z)$.

(2) *If ϕ is second order differentiable at z, then*

$$\tfrac{1}{2}\|\nabla^2 \phi(z)\| \leq c(\phi, z) \leq \tfrac{1}{2} b(\nabla^F \phi, z). \tag{7.17}$$

(3) *If ϕ is strongly second order differentiable at z, then equality holds in* (7.17).

Proof. First consider (1). Let $\beta > b(\nabla^F \phi, z)$. We prove that ϕ is Lipschitz smooth at z with constant $\beta/2$. Let $\delta > 0$ be such that

$$\langle \nabla^F \phi(z + tk) - \nabla^F \phi(z), k\rangle \leq \beta t \tag{7.18}$$

for all $\|k\| = 1$ and $|t| \leq \delta$. Fixing $\|k\| = 1$, we let $\gamma(t) = \phi(z + tk)$. Then (7.18) becomes $\gamma'(t) \leq \gamma'(0) + \beta t$, $|t| \leq \delta$. Integrating implies $\gamma(t) - \gamma(0) \leq t\gamma'(0) + \beta t^2/2$, proving $c(\phi, z) \leq \beta/2$.

Concerning (2), observe that by the definition of the second difference quotient, we obtain

$$\tfrac{1}{2}\langle \nabla^2\phi(z)k, k\rangle \leq c(\phi, z) \tag{7.19}$$

for all $\|k\| = 1$. But $\nabla^2\phi(z)$ is a positive and symmetric operator, so (7.19) implies (7.17).

Finally, concerning (3), let $T = \nabla^F\phi$, and suppose $\nabla^2\phi(z) = \nabla^F T(z)$ is a Fréchet derivative. Suppose we had $\|\nabla^2\phi(z)\| + \varepsilon < b(T, z) =: b$ for some $\varepsilon > 0$. By (7.15), there exist $k_n \to 0$ such that

$$(b - \varepsilon)\|k_n\|^2 < \langle T(z + k_n) - T(z), k_n\rangle \leq \|T(z + k_n) - T(z)\|\,\|k_n\|, \tag{7.20}$$

for $n = 1, 2, \ldots$. On the other hand, by the definition of the Fréchet derivative, we have

$$\|(T(z + k_n) - T(z))/\|k_n\| - \nabla^F T(z)k_n/\|k_n\|\| \to 0,$$

which, with (7.20), gives $\|\nabla^2\phi(z)\| < b - \varepsilon \leq \|\nabla^2\phi(z)\|$, a contradiction. □

Using the constant b, we may give the following dual formula for the set L_f of Lipschitz smooth points of the function f:

$$L_f = \{\nabla g^*(z) : b(\mathrm{id} - \nabla g^*, z) < 1\} \qquad (g = f + \tfrac{1}{2}\|\cdot\|^2). \tag{7.21}$$

Using both b and c, we now obtain the following dual characterization of second order differentiability, which is one of our main steps towards Alexandrov's theorem.

Theorem 7.5. *Let f be a continuous convex function on a separable Hilbert space H. Then the following are equivalent:*

(1) *f is second order differentiable at x (with $y = \nabla^F f(x)$).*

(2) *$f_1 = f \,\square\, \frac{1}{2}\|\cdot\|^2$ is second order differentiable at $x+y$ and $c(f_1, x+y) < \frac{1}{2}$.*

(3) *J_{f^*} is Gâteaux differentiable at $x + y$ and $b(J_{f^*}, x + y) < 1$.*

Proof. By Propositions 7.3 and 6.3(5), statements (1) and (3) are equivalent for a globally Lipschitz function f, as (7.13) translates into $b(J_{f^*}, x+y) < 1$. We now argue that the global Lipschitz assumption may be dropped. Invoking a standard shift here, we may assume without loss that $x = 0$ and $y = \nabla^F f(0) = 0$.

Let f be Lipschitz on some neighbourhood $B(0, \delta)$ of 0 with constant $n \in \mathbb{N}$, say. Then, according to [20], f and its penalty approximate $f_{[n]} = f \,\square\, n\|\cdot\|$ of order n coincide on $B(0, \delta)$, with ∂f and $\partial f_{[n]}$ being equal on $B(0, \delta)$, too. As the data in (1) and (3) above are clearly local, all we have to check is that $J_{f^*} = J_{(f_{[n]})^*}$ on a neighbourhood of $x + y = 0$. Using the definition (5.3) of the resolvent parameter and the fact that J_{f^*} is nonexpansive, it can be shown that the desired neighbourhood is $B(0, \delta/2)$. This proves that (1) and (3) are equivalent.

Now observe that, according to [27], J_{f^*} is the Fréchet derivative of f_1. Thus $J_{f^*} = J_{(f_{[n]})^*}$ on a neighbourhood of $x + y = 0$ implies that, up to a

constant, $f_1 = (f_{[n]})_1$ on the same neighbourhood. Therefore, by the reasoning above, it suffices to prove the equivalence of (2) and (3) for a globally Lipschitz function f, which we proceed to do.

Recall another fact from [27], namely $g^* + f_1 = \frac{1}{2}\|\cdot\|^2$, where as usual, $g = f + \frac{1}{2}\|\cdot\|^2$. This implies

$$\Delta_{g^*,x+y,x,t} + \Delta_{f_1,x+y,y,t} = \tfrac{1}{2}\|\cdot\|^2 \tag{7.22}$$

for the second difference quotients. Now Propositions 7.1 and 7.2 show that $b(J_{f^*}, x+y) < 1$ is equivalent to $\Delta_{g^*,t} \geq \alpha\|\cdot\|^2$ for some $0 < \alpha < \frac{1}{2}$. By (7.22), the latter is equivalent to $c(f_1, x+y) < \frac{1}{2}$. This completes the proof of the theorem. □

Remark. As a consequence of Theorem 7.5, we now obtain the following sufficient condition for the existence of points $x \in D_f^2$ in a separable Hilbert space:

$$l(J_{f^*}, z_0) < 1 \quad \text{for some } z_0 \in H. \tag{7.23}$$

Indeed, if (7.23) is satisfied, then $b(J_{f^*}, z) < 1$ in a neighbourhood of the point z_0. By Aronszajn's result [3], we may pick a Gâteaux differentiability point of J_{f^*} herein.

Example 1. Let C be a bounded closed convex set in the Hilbert space H. Then $f = \sup\langle C, \cdot\rangle$ is a sublinear function with $f^* = \delta(C|\cdot)$, the support function of C. Now $g^* = (f + \frac{1}{2}\|\cdot\|^2)^* = f^* \square \frac{1}{2}\|\cdot\|^2$ is of the form $g^*(x) = \frac{1}{2}\|x\|^2 - \frac{1}{2}\|x - P_C x\|^2$, which proves $J_{f^*} = P_C$ in this case. So every projection onto a bounded closed convex set C occurs as a mapping J_{f^*}. It is known that even in separable H, such P_C may fail to have Fréchet differentiability points (cf. [18, §5]); hence it is not clear whether equality holds in (7.17). We show that this need not be the case, by specifying the set C.

Example 2. Let $H = l_2$, $C = \prod_{n=1}^{\infty}[\alpha_n, \beta_n]$, where $\alpha_n < \beta_n$, $(\alpha_n), (\beta_n) \in l_2$. Then P_C is Gâteaux differentiable at points $z = (z_n)$ having $z_n \neq \alpha_n, \beta_n$. Notice that either $\|\nabla^G P_C(z)\| = 1$, namely if $\alpha_n < z_n < \beta_n$ for some n, or $\nabla^G P_C(z) = 0$, if $z_n \notin [\alpha_n, \beta_n]$ for all n. On the other hand, we can obtain a formula for $b(P_C, z)$ at the Gâteaux differentiability points z of P_C. Clearly $b(P_C, z) = 1$ if $\alpha_n < z_n < \beta_n$ for some n. So let us assume $z_n \notin [\alpha_n, \beta_n]$ for all n. Then we have

$$b(P_C, z) = \max\left(\varlimsup_{\substack{n\to\infty \\ z_n > \beta_n}} \frac{\beta_n - \alpha_n}{z_n - \alpha_n}, \varlimsup_{\substack{n\to\infty \\ z_n < \alpha_n}} \frac{\beta_n - \alpha_n}{\beta_n - z_n}\right). \tag{7.24}$$

In particular, this shows that $b(P_C, z) = 1$ is possible while $\nabla^G P_C(z) = 0$, so we see that equality in (7.17) may fail even with the largest possible gap of 1.

We next use Theorem 7.5(3) to prove the following infinite dimensional version of Alexandrov's theorem. For a related result see [29a].

Theorem 7.6. *Every continuous convex function $f : l_2 \to \mathbb{R}$ of the form $f(x) = \sum_{n=1}^{\infty} f_n(x_n)$ is densely second order differentiable. Even more, D_f^2 is not an Aronszajn null set.*

Proof. We assume that $f(0) = 0$, and that f is Lipschitz smooth at 0 with $\nabla^F f(0) = 0$. Moreover, it suffices to prove the result under the additional assumption that f is globally Lipschitz.

Because of the special form of f, the operator J_{f^*} is of the form $z \to (J_{f_n^*}(z_n))$, with each $J_n := J_{f_n^*} = (\mathrm{id} + \partial f_n^*)^{-1}$ being a monotone and nonexpansive function on the real line. Observe that, due to our special arrangements, the J_n are nondecreasing.

Based on Theorem 7.5(3), it will be sufficient to find differentiability points z of J_{f^*} satisfying $b(J_{f^*}, z) < 1$. Now, for any n and $v \in \mathbb{R}$, let $\sigma_n(v)$ be the *largest slope* of J_n at v, i.e., $\sigma_n(v) = \max\{\sigma_n^+(v), \sigma_n^-(v)\}$, where

$$\sigma_n^+(v) = \sup_{z>v} \frac{J_n(z) - J_n(v)}{z - v}, \qquad \sigma_n^-(v) = \sup_{y<v} \frac{J_n(v) - J_n(y)}{v - y}. \tag{7.25}$$

Let us fix a positive sequence $(\alpha_n) \in l_2$, $\alpha_n \leq 1$. By Lemma 7.7, there exists some $\delta > 0$ such that, for every $n \in \mathbb{N}$, there exists a subset B_n of $[0, \alpha_n]$ of positive Lebesgue measure such that

$$\sigma_n(v) \leq 1 - \delta < 1 \quad \text{for all } v \in B_n, \tag{7.26}$$

$n - 1, 2, \ldots$. Let $B = \prod_{n=1}^{\infty} B_n \subseteq l_2$. It follows from (7.26) that

$$b(J_{f^*}, z) \leq 1 - \delta \tag{7.27}$$

for all $z \in B$. Now we define Borel measures μ_n on the line by

$$\mu_n(X) = m(X \cap B_n)/m(B_n) \qquad (m = \text{Lebesgue measure})$$

and let μ be the product measure of the μ_n. As $(\alpha_n) \in l_2$, μ is a Borel probability measure on l_2 having $\mu(B) = 1$. But observe that $\mu_n \ll m$ for every n. This proves that $\mu \ll \nu$ for every Gaussian measure ν on l_2 constructed on the canonical basis of l_2 (cf. [33] for this). In particular, $\mu(A) = 0$ for every Gaussian null set A, and hence, by the result in [33], for every Aronszajn null set A. By Aronszajn's result [3], the operator J_{f^*} is Gâteaux differentiable outside an Aronszajn null set A. As $\mu(A) = 0$, we must have $\mu(B \backslash A) > 0$, hence $B \backslash A \neq \varnothing$. Choosing $z \in B \backslash A$ provides a differentiability point of J_{f^*} satisfying (7.27). This completes the proof. □

So it remains to find $\delta > 0$ and the sets B_n. This is done in the following:

Lemma 7.7. *There exists $\delta > 0$ such that for every $n \in \mathbb{N}$, the set $B_n = \{v \in [0, \alpha_n] : \sigma_n(v) \leq 1 - \delta < 1\}$ has positive Lebesgue measure.*

Proof. Notice that by assumption, f is Lipschitz smooth at 0. As $J_{f^*}(0) = 0$, Proposition 7.3 shows $b(J_{f^*}, 0) < 1$. Hence there exists $0 < c < 1$ so that $\langle J_{f^*}(k), k\rangle \leq c\|k\|^2$ for all $\|k\| \leq 1$, say. This implies

$$J_n(w) \leq c \cdot w, \qquad |w| \leq 1, \quad n = 1, 2, \ldots. \tag{7.28}$$

Choose $\delta > 0$ with $\delta < \frac{1}{2}(1 - c)$. We show that δ is as desired. Assume the contrary. Then, for some n,

$$\sigma_n(v) > 1 - \delta \quad \text{for almost all } v \in [0, \alpha_n]. \tag{7.29}$$

Observe that by the global Lipschitz assumption on f, each J_n has bounded range, hence its difference quotient decays at infinity. Therefore, the suprema in (7.25) are attained.

Let v be a differentiability point of J_n. Suppose first that $\sigma_n(v) = J_n'(v) > 1 - \delta$. Then we select $x(v) > v$ such that

$$J_n(w) > (1 - \delta)(w - v) + J_n(v), \qquad w \in [v, x(v)]. \tag{7.30}$$

In this case let us define $I(v) = [v, x(v)]$ and call $I(v)$ an *interval of the first kind*. Next consider the case where $\sigma_n(v) > J_n'(v)$. There are two possibilities.

First assume $\sigma_n(v) = \sigma_n^+(v) > 1 - \delta$. Then let $z(v)$ be the largest possible value $> v$ such that the supremum $\sigma_n^+(v)$ is attained at $z(v)$. Here we define $I(v) = [v, z(v)]$, calling $I(v)$ an *interval of the second kind*. Finally, in the case $\sigma_n(v) = \sigma_n^-(v) > 1 - \delta$, $\sigma_n^-(v) > \sigma_n^+(v)$, let $y(v)$ the smallest possible value $< v$ such that the supremum $\sigma_n^-(v)$ is attained at $y(v)$. Here set $I(v) = [y(v), v]$ and call $I(v)$ an *interval of the third kind*.

Observe that the intervals $I(v)$ cover $[0, \alpha_n]$ up to a null set because of (7.29) and the fact that J_n is a.e. differentiable.

Let $\varepsilon = \frac{1}{2}(1 - c - 2\delta) \cdot \alpha_n > 0$. There exist finitely many intervals $I(v_1), \ldots, I(v_r)$ covering $[0, \alpha_n]$ up to a set of measure $< \varepsilon$. As these intervals may have some overlap, we need to collect some facts about the $I(v)$.

First observe that two intervals $[v, z(v)]$ and $[w, z(w)]$ of the second kind cannot overlap. For suppose $v < w < z(v)$. Then according to the definition of $z(v)$, the value $J_n(w)$ cannot lie strictly above the line joining $J_n(v)$ and $J_n(z(v))$. Suppose $J_n(w)$ lies on the line. Then necessarily $\sigma_n(w) = \sigma_n(v)$, and hence $z(w) = z(v)$. If $J_n(w)$ lies strictly below the line joining $J_n(v)$ and $J_n(z(v))$, we see that

$$(J_n(z(v)) - J_n(w))/(z(v) - w) > (J_n(x) - J_n(w))/(x - w)$$

for all $x > z(v)$. In other terms, these x do not contribute to the supremum $\sigma_n^+(w)$, hence $z(w) \leq z(v)$. This proves that the intervals of the second kind do not overlap. A similar reasoning shows that neither may the intervals of the third kind have an overlap.

Let us arrange the $I(v_j)$ in blocks $I(v_{j_{i-1}+1}), \ldots, I(v_{j_i})$, $i = 1, 2, \ldots, k$, of pairwise overlapping intervals, but with a gap between $I(v_{j_i})$ and $I(v_{j_i+1})$. Let

$$[u^{(i)}, v^{(i)}] = I(v_{j_{i-1}+1}) \cup \cdots \cup I(v_{j_i})$$

be the interval covered by the ith block. Then the gaps have length $\varepsilon_i = u^{(i+1)} - v^{(i)}$ with $\sum_{i=1}^{k} \varepsilon_i < \varepsilon$ by assumption.

Let L denote the line with slope $1 - \delta$ emanating from the origin. We show that for all points $w \in [0, \alpha_n]$,

$$J_n(w) + \delta\alpha_n + \varepsilon \geq (1 - \delta)w. \tag{7.31}$$

In other terms, the graph of the function J_n is above the line L up to an error $\leq \delta\alpha_n + \varepsilon$. We prove (7.31) by proceeding in positive direction from interval to interval.

First consider the interval $[0, u^{(1)}]$, i.e., the first possible gap of length ε_0. Clearly, as $J_n \geq 0$, we have

$$J_n(w) + \varepsilon_0 \geq (1 - \delta)w \tag{7.32}$$

for all $w \in [0, u^{(1)}]$.

Now consider the first interval $I(v_1)$ of the first block. Let $I(v_1) = [e, e^1]$, $e = u^{(1)}$. By (7.32), the value $J_n(e)$ is above the line L up to the error ε_0. As the line of the first interval joining $J_n(e)$ and $J_n(e^1)$ has slope $> 1 - \delta$, the right endpoint e^1 of $I(v_1)$ satisfies $J_n(e^1) + \varepsilon_0 \geq (1 - \delta)e^1$. What about the points within $I(v_1)$? If $I(v_1)$ is an interval of the first kind, then by (7.30),

the points within $I(v_1)$ have the same behaviour as e and e^1. If $I(v_1)$ is an interval of the third kind, there is no problem either, for by the definition of $I(v_1)$, the graph of the function J_n on $I(v_1)$ must then be entirely above the line joining $J_n(e)$ and $J_n(e^1)$. So for $I(v_1)$ of the first or third kind, the relation (7.32) is valid for all $w \in I(v_1)$. In particular (7.32) holds for the left endpoint of the second interval $I(v_2)$ of the first block, so that the process may be continued. Finally, if $I(v_1)$ happens to be an interval of the second kind, $I(v_1) = [e, z(e)]$, then the graph of J_n is below the line joining $J_n(e)$ and $J_n(z(e))$. So here we have to estimate by what amount this may happen. Now using the facts that J_n is nonexpansive and that the line joining $J_n(e)$ and $J_n(z(e))$ has slope $> 1 - \delta$, we find that, for every $w \in [e, z(e)]$,

$$J_n(w) \geq (1-\delta)(w-e) + J_n(e) - \delta(z(e) - w), \tag{7.33}$$

which means that $J_n(w)$ lies below the line with maximum possible error $\delta(z(e) - e) = \delta|I(v_1)|$. In this case, by adding this error, we obtain

$$J_n(w) + \delta|I(v_1)| + \varepsilon_0 \geq (1-\delta)w \tag{7.34}$$

for all $w \in I(v_1)$, replacing (7.32). In particular, (7.34) pertains to the left-hand endpoint of the interval $I(v_2)$, so that the process may be continued.

Proceeding in this way, we see that we have to add an error $\delta|I(v_j)|$ whenever an interval $I(v_j)$ of the second kind occurs, and an error ε_i whenever a gap between blocks occurs. On the other hand, intervals of the first or third kind do not cause any trouble. Finally, we see that the total error is

$$\delta \sum_j |I(v_j)| + \sum_i \varepsilon_i \leq \delta\alpha_n + \varepsilon,$$

where the first summation is over the intervals of the second kind, and the second summation is over the gaps. Notice here that the intervals of the second kind are mutually disjoint and contained in $[0, \alpha_n]$. This establishes formula (7.31).

Combining (7.31), (7.28) and choosing $w = \alpha_n$, we obtain that $\varepsilon \geq (1 - c - 2\delta)\alpha_n$, which contradicts the choice of ε. Hence the proof of the lemma is complete. □

Using Lemma 7.7, the proof of Theorem 7.6 may now be modified to prove a version of Alexandrov's theorem for integral functionals.

Corollary 7.8. *Let f be an integral functional of the form* (6.17) *with $\phi : \mathbb{R}\times\Omega \to \mathbb{R}$, and such that $L^2(\Omega)$ is separable. Then D_f^2 is dense in $L^2(\Omega)$.*

Proof. Indeed, here J_{f^*} is of the form $z \to J_\cdot(z(\cdot))$, with

$$J_\tau := (\mathrm{id} + \partial\phi(\cdot, \tau)^*)^{-1}.$$

Fixing a strictly positive $\alpha \in L^2(\Omega)$ (μ is σ-finite), we apply the lemma to see that, for every $\tau \in \Omega$,

$$B_\tau = \{0 \leq v \leq \alpha(\tau) : J_\tau'(v) \text{ exists and } \sigma_\tau(v) \leq 1 - \delta\}$$

has positive measure, with $\delta > 0$ chosen as above. Observing that $\tau \to B_\tau$ is a measurable multifunction, we invoke the Kuratowski/Ryll-Narczewski selection theorem to fix a measurable selection $\tau \to z(\tau)$. This provides an element $z \in L^2(\Omega)$ having $b(J_{f^*}, z) \leq 1 - \delta < 1$. We have only to observe that J_{f^*} is Gâteaux differentiable at z. Now recall that $J_{f^*} = \nabla^F f_1$, where $f_1 = f \square \frac{1}{2}\|\cdot\|^2$. Hence f_1 is again an integral functional, namely

$$f_1(z) = \int_\Omega \phi_1(z(\tau), \tau)\, d\mu(\tau),$$

where $\nabla\phi_1(\cdot, \tau) = J_\tau$. But then $z \in GD^2_{f_1}$ by Proposition 6.3, and as f_1 is everywhere Lipschitz smooth, $z \in D^2_{f_1}$ by Proposition 6.1(3), which proves the result. □

Remarks. (1) With some effort it is possible to prove existence $D^2_f \neq \varnothing$ for multidimensional integral functionals. This requires an n dimensional version of Lemma 7.7. We do not give the details of the result here.

(2) Observe that the proof of Theorem 7.6 applies to any $f = \sum_n f_n$ defined on a space l_p, $p \geq 2$, while the result fails in l_q, $1 \leq q < 2$, by Proposition 3.3.

8. Sard's theorem

Let us explain one connection between Alexandrov's theorem and the theorem of Sard. Recall that a version of Sard's theorem tells that, for a Lipschitz operator $T : \mathbb{R}^n \to \mathbb{R}^n$, the set of singular values of T

$$\{T(z) : \nabla^G T(z) \text{ exists, but } \operatorname{rank} \nabla^G T(z) < n\} \tag{8.1}$$

is a Lebesgue null set (see [26; 17, 3.2.5]). There are several ways of introducing the notion of singular resp. regular points in infinite dimensions:

Let $T : H \to H$ be a (locally) Lipschitz operator. Then $z \in H$ is called a *regular point* (resp. a *generalized regular point*) of T if T is Gâteaux differentiable at z and

$$\inf_{0<\|k\|\leq\delta} \frac{|\langle T(z+k) - T(z), k\rangle|}{\|k\|^2} > 0 \quad \text{for some } \delta > 0; \tag{8.2}$$

respectively

$$\inf_{\|k\|=1} |\langle \nabla^G T(z)k, k\rangle| > 0. \tag{8.3}$$

Clearly, in finite dimensions, both conditions (8.2), (8.3) are equivalent to saying that rank $\nabla^G T(z)$ is maximal, so we get the classical notion of regular points.

Interpreting our results obtained in §7 for the operator $\nabla g^* = J_f$ (where $g = f + \frac{1}{2}\|\cdot\|^2$ as usual), we obtain the identities

(i) $L_f = \{J_f(z) : (8.2) \text{ holds for } J_f \text{ at } z\}$;

(ii) $D^2_f = \{J_f(z) : z \text{ is regular for } J_f\}$;

(iii) $GD^2_f = \{J_f(z) : z \text{ is generalized regular for } J_f\}$.

Indeed, (i) is just rewriting (7.21), (ii) is Theorem 7.5, while (iii) comes from Proposition 6.3. By identity (ii), the failure of Alexandrov's theorem,

i.e., $D_f^2 = \varnothing$ for some f, would imply the failure of Sard's theorem for the operator J_f in the sense that the singular values of J_f are all of H. A similar interpretation applies to $GD_f^2 = \varnothing$. Concerning (i), observe that by Fabian's result [16], L_f is always a dense set in H, so we at least have a weak version of Sard's theorem, saying that the set of values $J_f(z)$ such that (8.2) fails for z is small in the sense that its complement L_f is dense. We conjecture, however, that this set is even smaller in the sense that L_f (and therefore D_f^2) is not an Aronszajn null set.

We mention that, typically, infinite dimensional versions of Sard's theorem as for instance the Sard-Smale theorem, need strong assumptions which seem to be violated in natural situations (see [1, 10]).

ACKNOWLEDGMENT

We thank the referee for pointing out an inaccuracy in the original proof of Theorem 7.5.

BIBLIOGRAPHY

1. R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, tensor analysis, and applications*, Addison-Wesley, Reading, MA, 1983.
2. A. D. Alexandrov, *Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it*, Leningrad State Univ. Ann. [Uchenye Zapiski] Math. Ser. **6** (1939), 3–35. (Russian)
3. N. Aronszajn, *Differentiability of Lipschitzian mappings between Banach spaces*, Studia Math. **57** (1976), 147–190.
4. H. Attouch, *Familles d'opérateurs maximaux monotones et mesurabilité*, Ann. Mat. Pura Appl. (4) **120** (1979), 35–111.
5. H. Attouch and R. J. B. Wets, *Isometries for the Legendre-Fenchel transform*, Trans. Amer. Math. Soc. **296** (1986), 33–60.
6. V. Bangert, *Analytische Eigenschaften konvexer Funktionen auf Riemannschen Mannigfaltigkeiten*, J. Reine Angew. Math. **307** (1979), 309–324.
7. A. Ben-Tal and J. Zowe, *Directional derivatives in nonsmooth optimization*, J. Optim. Theory Appl. **47** (1985), 483–490.
8. J. M. Borwein, S. Fitzpatrick, and P. Kenderov, *Minimal convex uscos and monotone operators on small sets*, Canad. J. Math. **43** (1991), 461–476.
9. J. M. Borwein and D. Preiss, *A smooth variational principal with applications to subdifferentiability and to differentiability of convex functions*, Trans. Amer. Math. Soc. **303** (1987), 517–527.
10. K. Buchner, *Sard's theorem on Banach manifolds. Topics in differential geometry*. I, II, Colloq. Math. Soc. János Bolyai, vol. 46, North-Holland, Amsterdam, 1989.
11. H. Busemann, *Convex surfaces*, Interscience, New York, 1955.
12. H. Busemann and W. Feller, *Krümmungseigenschaften konvexer Flächen*, Acta Math. **66** (1936), 1–47.
13. J. P. R. Christensen, *Measure theoretic zero-sets in infinite dimensional spaces and applications to differentiability of Lipschitz mappings*, II, Colloq. Analyse Fonctionnel (Bordaux, 1973), pp. 29–39.
14. V. F. Demyanov, C. Lemaréchal, and J. Zowe, *Approximation to a set-valued mapping*, I: *A proposal*, Appl. Math. Optim. **14** (1986), 203–214.
15. J. Diestel and J. J. Uhl, *The theory of vector measures*, Math. Surveys Monographs, vol. 15, Amer. Math. Soc., Providence, RI, 1977.

16. M. Fabian, *Lipschitz smooth points of convex functions and isomorphic characterizations of Hilbert spaces*, Proc. London Math. Soc. (3) **51** (1985), 113–126.

17. H. Federer, *Geometric measure theory*, Springer-Verlag, Berlin and New York, 1969.

18. S. Fitzpatrick and R. R. Phelps, *Differentiability of the metric projection in Hilbert space*, Trans. Amer. Math. Soc. **270** (1982), 483–501.

19. ——, *Bounded approximants to monotone operators in Banach spaces* (to appear).

20. J. B. Hiriart-Urruty, *Lipschitz r-continuity of the approximate subdifferential of a convex function*, Math. Scand. **47** (1980), 124–134.

21. ——, *Calculus rules on the approximate second order directional derivative of a convex function*, SIAM J. Control Optim. **22** (1984), 381–404.

22. J. B. Hiriart-Urruty and A. Seeger, *Calculus rules on a new set-valued second order derivative for convex functions*, Nonlinear Anal. **13** (1989), 721–738.

23. ——, *The second order subdifferential and the Dupin indicatrices of a non-differentiable convex function*, Proc. London Math. Soc. **58** (1989), 351–365.

24. ——, *Complément de Schur et sous-différentiel du second ordre d'une fonction convexe* (to appear).

25. N. Kato, *On the second derivatives of convex functions in Hilbert space*, Proc. Amer. Math. Soc. **106** (1989).

26. F. Mignot, *Contrôle dans les inéquations variationelles elliptiques*, J. Funct. Anal. **22** (1976), 130–185.

27. J. J. Moreau, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France **93** (1965), 273–299.

28. U. Mosco, *Convergence of convex sets and solutions to variational inequalities*, Adv. Math. **3** (1969), 510–585.

29. A. S. Nemirovskii and S. M. Semenov, *On polynomial approximation of functions on Hilbert space*, Mat. Sb. **92(134)** (1973), 255–277.

29a. D. Noll, *Second order differentiability of integral functionals on Sobolev spaces and L^2-spaces*, J. Reine Angew. Math. **436** (1993), 1–17.

30. D. Noll, *Generic Gâteaux-differentiability of convex functions on small sets*, J. Math. Anal. Appl. **147** (1990), 531–544.

31. ——, *Generic Fréchet-differentiability of convex functions on small sets*, Arch. Math. **54** (1990), 487–492.

32. ——, *Generalized second fundamental form for Lipschitzian hypersurfaces by way of second epi derivatives*, Canad. Math. Bull. **35** (1992), 523–536.

33. R. R. Phelps, *Gaussian null sets and differentiability of Lipschitz maps on Banach spaces*, Pacific J. Math. **77** (1978), 523–531.

34. ——, *Convex functions, monotone operators and differentiability*, Lecture Notes in Math., vol. 1364, Springer-Verlag, Berlin and New York, 1989.

35. D. Preiss, *Differentiability of Lipschitz functions in Banach spaces*, J. Funct. Anal. **91** (1990).

36. R. T. Rockafellar, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré. Anal. Non Linéaire **2** (1985), 167–184.

37. ——, *Generalized second derivatives of convex functions and saddle functions*, Trans. Amer. Math. Soc. **322** (1990), 51–77.

38. ——, *Conjugate duality and optimization*, SIAM, Philadelphia, Pa., 1974.

39. ——, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res. **14** (462–484).

40. G. Salinetti and R. J.-B. Wets, *On the relation between two types of convergence for convex sets*, J. Math. Anal. Appl. **60** (1977), 211–226.

41. J. Stoer and A. Witzgall, *Convexity and optimization in finite dimensions.* I, Springer-Verlag, Berlin and New York, 1970.

42. J. Vanderwerff, *Smooth approximation in Banach spaces* (to appear).

43. R. A. Wijsman, *Convergence of sequences of convex sets, cones and functions.* II, Trans. Amer. Math. Soc. **123** (1966), 32–45.

Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, Canada B3H 3J5

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada

Current address, J. M. Borwein: Shrum Professor of Science, Simon Fraser Centre for Experimental and Constructive Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6 Canada

(Dominikus Noll) Universität Stuttgart, Mathematisches Institut B, Pfaffenwaldring 57, 70550 Stuttgart 80, Deutschland