Ark. Mat., **34** (1996), 383–406

On regularization in Banach spaces

Thomas Strömberg

1. Introduction

In the present paper we propose a regularization procedure for functions defined on Banach spaces admitting equivalent locally uniformly rotund norms the dual norm of which are also locally uniformly rotund. We demonstrate that with any bounded below lower semi-continuous (l.s.c.) proper function f defined on such a Banach space X can be associated a family of C^1 functions approximating f from below and enjoying favorable properties from the viewpoint of minimization. Our method reduces in the case where X is a Hilbert space to the one that was introduced and investigated by J. M. Lasry and P. L. Lions in their joint paper [10]. Their approach has subsequently been further explored by other authors, notably in [3] and [5], but never before outside the Hilbert space setting.

The Lasry–Lions method is based upon Moreau–Yosida approximation. Given f, an extended-real-valued function defined on a Hilbert space X, the *Moreau–Yosida approximates* of f are the functions f_t, $t>0$, that carry each $x\in X$ to

$$f_t(x)=\inf_{y\in X}\left(f(y)+\frac{1}{2t}\|x-y\|^2\right). \tag{1}$$

In the case where f is convex, l.s.c., and proper, the envelope functions f_t possess Lipschitz continuous Fréchet differentials (in symbols: $f_t\in C^{1,1}$) and $f_t\to f$, at least pointwise, as $t\downarrow 0$. There is even convergence, in certain senses, of the differentials df_t to the subdifferential ∂f. Furthermore, the infimal value of f together with the set of minimizers, as well as the stationary points and values, are preserved. The convexity hypothesis can actually be weakened; it suffices to assume that $f+(2T)^{-1}\|\cdot\|^2$ is convex for some $T>0$ in which case $f_t\in C^{1,1}$ etc. when $t\in(0,T)$.

In order to extend at least some of these results to non-convex functions, Lasry and Lions introduced a two-parameter family of approximates by putting $f_{t,s}=-(-f_t)_s$, $0<s<t$. Let us, for simplicity, assume that f is bounded from below which guarantees that the functions $f_{t,s}$ are all real-valued. It was proved in [10],

[3], without any convexity hypothesis on f, that $f_{t,s}$ enjoys $C^{1,1}$ smoothness when $0<s<t$, that $f_{t,s}\to f$ pointwise as $0<s<t\downarrow 0$ if f is l.s.c., and that the convergence is uniform if f is uniformly continuous. However, the techniques used in the seminal paper [10] and also in subsequent papers seem to depend on the specific properties of the Hilbertian norm and therefore to fail to carry over to a richer class of Banach spaces. The objective of the work reported here is to show that these problems can be overcome by using an approach by means of one of the chief tools in convex analysis, namely the Legendre–Fenchel transformation. The result is a regularization method having advantageous variational properties.

It deserves to be mentioned that one can view the Moreau–Yosida process as a regularization by way of the Cauchy problem

$$\begin{aligned} \partial u(x,t)/\partial t+\|d_x u(x,t)\|^2/2=0, \qquad & x\in X,\ t>0,\\ u(x,0)=f(x), \quad & x\in X. \end{aligned}$$

In fact, when $X=\mathbf{R}^n$ and f is l.s.c. and bounded from below, the viscosity solution to this initial-value problem is given by the Lax–Oleinik formula

$$u(x,t)=f_t(x)=\inf_{y\in X}\left(f(y)+\frac{1}{2t}\|x-y\|^2\right).$$

We refer to [11, Proposition 13.1] for a proof. Furthermore, we prove below (Proposition 3) that when X is an arbitrary Hilbert space and $f+(2T)^{-1}\|\cdot\|^2$ is convex, where T is some positive real number, then, as indicated by a formal application of the method of characteristics, the function $u(x,t)=f_t(x)$ satisfies the above Hamilton–Jacobi equation at each point (x,t) in $X\times(0,T)$. Also, if we define $S(t)f=f_t$, $t>0$, and $S(0)f=f$, for l.s.c. bounded below functions f, the family $(S(t))_{t\geq 0}$ forms a semigroup of operators (on the cone of all bounded below l.s.c. proper functions on X) with $S(t)f\to f$ pointwise and with respect to the epi-distance topology as $t\downarrow 0$.

The Lasry–Lions approximates can be written $f_{t,s}=\Sigma(s)S(t)f$ where $(\Sigma(s))_{s\geq 0}$ is the semigroup defined via

$$\Sigma(s)g\colon x\mapsto \sup_{y\in X}\left(g(y)-\frac{1}{2s}\|x-y\|^2\right).$$

The inequality $\Sigma(s)S(t)f\leq S(t-s)f$ is always true as long as $0<s<t$, while the equation $\Sigma(s)S(t)f=S(t-s)f$ ("time-reversal") holds when $0<s<t\leq T$ if and only if $f+(2T)^{-1}\|\cdot\|^2$ is convex; see Proposition 2.

At this stage we would like to remark that our method, outside the Hilbertian case, is not directly connected to Hamilton–Jacobi equations.

The paper is organized in the following way. Section 2 contains background material mainly on convex functions. The definition of the approximates and the statement of Theorem 1, the corner-stone of the paper, are given in Section 3 while the proof of the theorem can be found in Section 4. Theorem 1 summarizes the main results; several of the properties of the Lasry–Lions approximation process are shown to extend to the broader framework of functions defined on Banach spaces satisfying the above mentioned rotundity hypotheses. Also the epi-distance topology, which has received attention over the last years, is considered. In Section 5 approximation by twice Gâteaux differentiable $C^{1,1}$ functions in separable Banach spaces is examined. A combination of our Theorem 1 with techniques used in the papers [8], [13] yields a variant of certain results obtained in [13]. These authors have focused on convex functions which are bounded on bounded sets, whereas we relinquish convexity and consider functions which are uniformly continuous on bounded sets. Last on the agenda, the Lasry–Lions approximates of functions f (defined on Hilbert space) which are locally convex up to a positive multiple of the square of the norm are investigated in Section 6. We prove, firstly, that the derivatives of the approximates $df_{t,s}$ converge, in certain senses, to the Clarke subdifferential ∂f and, secondly, that the stationary points and values of f are preserved.

2. Preliminaries

We shall adopt some terminology and notation which are of a common use in the field of convex analysis. First of all, the *conjugate function* or *Legendre–Fenchel transform* of an arbitrary function $f\colon X\to[-\infty,+\infty]$ is the extended-real-valued function f^* on X^*, the topological dual space of X, that assigns to each $\xi\in X^*$ the number

$$f^*(\xi)=\sup_{x\in X}(\langle x,\xi\rangle-f(x)).$$

Symmetrically, for φ defined on X^*, $\varphi^*\colon X\to[-\infty,+\infty]$ sends each $x\in X$ to

$$\varphi^*(x)=\sup_{\xi\in X^*}(\langle x,\xi\rangle-\varphi(\xi)).$$

The biconjugate $f^{**}=(f^*)^*$ is equal to $\overline{\mathrm{co}}\,f$, the greatest convex l.s.c. minorant of f, provided f admits a continuous affine minorant.

The notation $\mathbf{\Gamma}(X)$ signifies the set of all convex l.s.c. proper functions on X.

Recall that f is termed *proper* provided it is somewhere finite and nowhere $-\infty$. The set of points at which a proper function f is finite is called the *effective domain* of f and is denoted by $\operatorname{dom} f$. We write $\inf f$ for $\inf\{f(x);x\in X\}$ and $\arg\min f$

for the possibly empty set of minimizers $\{x\in X; f(x)=\inf f\}$. Moreover, we term f *demi-convex* if $f+(\alpha/2)\|\cdot\|^2$ is convex for some $\alpha\geq 0$.

The *infimal convolute* $f\Box g$ of two proper extended-real-valued functions f and g on X is defined by

$$f\Box g(x)=\inf_{y\in X}(f(y)+g(x-y)) \quad \text{for all } x\in X.$$

Frequently, $(X,\|\cdot\|)$ stands for a Banach space having the property that $\|\cdot\|$ as well as the canonical dual norm $\|\cdot\|_*$ are locally uniformly rotund. Let us make precise the notion of (local) uniform rotundity. The norm $\|\cdot\|$ is called *locally uniformly rotund* (LUR) provided $\|\cdot\|^2$ is a locally uniformly convex function, that is, to each point $x\in X$ there exists a non-decreasing function Δ_x on $[0,+\infty)$ with $\Delta_x(u)>0$ if $u>0$ such that

$$\|(x+y)/2\|^2 \leq \tfrac{1}{2}\|x\|^2+\tfrac{1}{2}\|y\|^2-\Delta_x(\|x-y\|) \quad \text{for all } y\in X. \tag{2}$$

It is well known that local uniform rotundity of $\|\cdot\|_*$ forces $\|\cdot\|$ to be Fréchet differentiable on $X\setminus\{0\}$. This is actually a particular case of Lemma 4 below which we shall employ later on.

The norm $\|\cdot\|$ is *uniformly rotund* (UR) if there exists a non-decreasing function Δ defined on $[0,+\infty)$ with $\Delta(u)>0$ when $u>0$ such that

$$\|(x+y)/2\|^2 \leq \tfrac{1}{2}\|x\|^2+\tfrac{1}{2}\|y\|^2-\Delta(\|x-y\|) \quad \text{when } x,y\in X,\ \|x\|\leq 1. \tag{3}$$

The functions Δ_x and Δ are sometimes tacitly used in the sequel.

In part of Theorem 1 the l.s.c. proper functions on X will be endowed with the "epi-distance" topology which we proceed to define. Let A and B be two nonempty subsets of a Banach space Y. The *excess* of A over B is the number

$$E(A,B)=\sup\{d(a,B)\,;a\in A\}$$

where $d(a,B)=\inf_{b\in B}\|a-b\|$. For $\varrho>0$ we write

$$H_\varrho(A,B)=\max\{E(\varrho\mathbf{U}\cap A,B),E(\varrho\mathbf{U}\cap B,A)\}$$

where $\mathbf{U}$ stands for the closed unit ball in Y. The quantity $H_\varrho(A,B)$ is called the *ϱ-Hausdorff distance* between A and B. A closed nonempty subset A of Y is the limit, with respect to these distances, of a net of closed nonempty subsets $(A_\lambda)_{\lambda\in\Lambda}$ if $H_\varrho(A,A_\lambda)\to 0$ for each $\varrho>0$.

We shall in this paper only consider convergence of epigraphs of proper l.s.c. functions $X \to (-\infty, +\infty]$. If f and g are such functions, then we write, for notational simplicity, $H_\varrho(f,g)$ instead of $H_\varrho(\operatorname{epi} f, \operatorname{epi} g)$ where the latter is defined with the convention that $Y = X \oplus \mathbf{R}$ be equipped with the box norm $\|(x,\alpha)\| = \max\{\|x\|, |\alpha|\}$. The resulting function space topology is called the *epi-distance topology*. One reason for the interest in this topology is the fact that with respect to it the Legendre–Fenchel transformation is a homeomorphism of $\mathbf{\Gamma}(X)$ onto $\mathbf{\Gamma}^*(X^*)$. Another reason is that the convergence in the epi-distance topology implies the convergence of epigraphs in the sense of Kuratowski–Painlevé, and the latter plays a prominent role in variational analysis. For a recent exposition consult [4].

For $A \subseteq X$, $\mathfrak{I}_A$ stands for the *indicator function* of A:

$$\mathfrak{I}_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{if } x \in \complement A. \end{cases}$$

If $f: X \to \mathbf{R}$, the sum $f + \mathfrak{I}_A$ is the "restriction" of f to the subset A.

Finally, $B(x,r)$ denotes the closed ball centered at x and of radius r.

3. The approximates and main results

We proceed to our choice of approximation.

Definition. Suppose f is an extended-real-valued function on a real Banach space X. For positive real numbers $s < t$, $f_{t,s}$ is defined by

$$f_{t,s}(x) = \left[\left(f + \frac{1}{2t}\|\cdot\|^2 \right)^* (\cdot/t) + \left(\frac{1}{s} - \frac{1}{t} \right) \frac{1}{2} \|\cdot\|_*^2 \right]^* (x/s) - \frac{1}{2s}\|x\|^2, \quad x \in X.$$

A dual formulation in terms of infimal convolution is provided by the following proposition.

Proposition 1. *Let f be a proper bounded below function on X, $0 < s < t$, $x \in X$. Then*

$$f_{t,s}(x) = \inf_{y \in X} \left[\overline{\operatorname{co}} \left(f + \frac{1}{2t}\|\cdot\|^2 \right)(y) + \frac{1}{1/s - 1/t} \frac{1}{2} \left\| \frac{x}{s} - \frac{y}{t} \right\|^2 \right] - \frac{1}{2s}\|x\|^2. \tag{4}$$

Proof. Put

$$g = \overline{\operatorname{co}} \left(f + \frac{1}{2t}\|\cdot\|^2 \right)(t\cdot) \quad \text{and} \quad h = \frac{1}{1/s - 1/t} \frac{1}{2} \|\cdot\|^2;$$

with conjugates

$$g^* = \left(f + \frac{1}{2t}\|\cdot\|^2\right)^*(\cdot/t) \quad \text{and} \quad h^* = \left(\frac{1}{s} - \frac{1}{t}\right)\frac{1}{2}\|\cdot\|_*^2.$$

Using the convex duality formula $(g^*+h^*)^* = g\Box h$, which indeed is available since $g, h \in \mathbf{\Gamma}(X)$ and h is real-valued and strongly coercive, we find that

$$\begin{aligned} f_{t,s}(x) + \frac{1}{2s}\|x\|^2 &= (g^*+h^*)^*(x/s) = g\Box h(x/s) \\ &= \inf_{w\in X}\left[\overline{\mathrm{co}}\left(f + \frac{1}{2t}\|\cdot\|^2\right)(tw) + \frac{1}{1/s - 1/t}\frac{1}{2}\left\|\frac{x}{s} - w\right\|^2\right]. \end{aligned}$$

The asserted formula (4) clearly follows on substituting $tw = y$. $\square$

Without any further preparations we turn to the formulation of our main theorem.

Theorem 1. *Suppose X is a Banach space whose norm and dual norm are both* LUR. *Let $f: X \to (-\infty, +\infty]$ be proper, l.s.c., and minorized. If $0 < s < t$ then assertions* (i)–(iv) *concerning $f_{t,s}$ hold*:

(i) *$f_{t,s}$ is continuously differentiable and Lipschitz continuous on bounded sets*;

(ii) *$df_{t,s}$ is uniformly continuous on bounded sets* (*respectively, globally Lipschitz continuous*) *if $\|\cdot\|_*$ is* UR (*respectively, $J = d\|\cdot\|^2/2$ is globally Lipschitz continuous*);

(iii) $f_{t,s} \le f$, $\inf f_{t,s} = \inf f$, *and* $\arg\min f_{t,s} = \arg\min f$;

(iv) *$f_{t,s} + (2s)^{-1}\|\cdot\|^2$ is convex.*

Moreover, when $0 < s < t \downarrow 0$,

(v) *$f_{t,s} \to f$ pointwise, and provided $\|\cdot\|$ is* UR *in the epi-distance sense too*;

(vi) *$f_{t,s} \to f$ uniformly on bounded sets when $\|\cdot\|$ is* UR *and f is uniformly continuous on bounded sets.*

We emphasize that we impose no convexity assumptions whatsoever upon f.

The assertion (iii) expresses that the approximation is from below and, which is more important, that the method has the pleasant feature that it preserves the infimum of f and the associated set of minimizers.

We conclude this section with a list of corollaries to Theorem 1. It shows in particular that certain known results on approximation of uniformly continuous functions and on partitions of unity can rather easily be derived from Theorem 1.

Even specializing to indicator functions in Theorem 1 yields something nontrivial. This illustrates the advantage of the admissibility of the function value $+\infty$.

Corollary 1. *Let X be a Banach space whose norm and dual norm are both* LUR. *To any closed nonempty subset A of X there corresponds a demi-convex C^1 function ψ such that $\psi(x)=0$ when $x\in A$ and $\psi(x)>0$ when $x\in\complement A$.*

If the dual norm is UR *rather than merely* LUR, *ψ may be chosen so as to have a differential $d\psi$ that is uniformly continuous on bounded sets.*

Proof. To say that A is closed and nonempty means the same as to say that $\mathfrak{I}_A$ is l.s.c. and proper. Fix $0<s<t$ and let ψ be the approximate $(\mathfrak{I}_A)_{t,s}$. According to Theorem 1, $\psi\in C^1(X)$, $\inf\psi=\inf\mathfrak{I}_A=0$, and $\arg\min\psi=\arg\min\mathfrak{I}_A=A$. In addition, $\psi+(2s)^{-1}\|\cdot\|^2$ is convex.

When $\|\cdot\|_*$ is UR, $d\psi$ is uniformly continuous on each bounded subset of X. □

We say that X admits C^1 *partitions of unity* if for any open covering $\mathcal{O}$ of X there exists a C^1 partition of unity which is subordinated to $\mathcal{O}$. It was proved in [17, Theorem 2.1] that X, subject to the hypotheses of Corollary 1, admits C^1 partitions of unity. We present next, in particular, a proof based upon Corollary 1.

Corollary 2. *Suppose X has an* LUR *norm whose dual is also* LUR.

(i) *X admits C^1 partitions of unity;*

(ii) *To any locally finite open covering $(O_\lambda)_{\lambda\in\Lambda}$ of X there exists a family of non-negative C^1 functions $(\varphi_\lambda)_{\lambda\in\Lambda}$ such that $\varphi_\lambda^{-1}(0,+\infty)=O_\lambda$ for each $\lambda\in\Lambda$ and $\sum_{\lambda\in\Lambda}\varphi_\lambda(x)=1$ for all $x\in X$.*

Proof. Let $\mathcal{O}^1(X)$ be $\{\psi^{-1}(0,+\infty);\psi\in C^1(X),\ \psi\geq 0\}$. Statement (i), concerning an arbitrary Banach space X, is equivalent to

(i$'$) If $A\subseteq B\subseteq X$, where A is closed and B is open, there exists $O\in\mathcal{O}^1(X)$ such that $A\subseteq O\subseteq B$.

Consult for instance [6, Chapter VIII, Lemma 3.6] for a proof. In our case, (i$'$) is certainly met since, by Corollary 1, $\mathcal{O}^1(X)$ consists of all open subsets of X.

(ii) Choose C^1 functions $\psi_\lambda\geq 0$ such that $\psi_\lambda(x)>0$ if and only if $x\in O_\lambda$, and define

$$\varphi_\lambda(x)=\psi_\lambda(x)\,/\textstyle\sum_{\mu\in\Lambda}\psi_\mu(x).$$

These functions will serve the purpose. □

Remark. Let k be a positive integer or ∞. The following conditions relative to a Banach space X are known to be equivalent (see [15, Proposition 2.1] and [6, Chapter VIII, Theorem 3.2]):

(a) *Every real-valued Lipschitz continuous function on X can be locally uniformly approximated by C^k functions;*

(b) *To every continuous $S\colon X\to Y$, where Y is a Banach space, and every continuous $p\colon X\to(0,+\infty)$ there exists a $T\in C^k(X,Y)$ such that*

$$\|S(x)-T(x)\|<p(x)\quad \textit{for all } x\in X;$$

(c) *X admits C^k partitions of unity.*

Recent contributions to the area of smooth partitions of unity appear in [9] and [12].

We show next that pointwise approximation of arbitrary l.s.c. proper functions is always possible in our setting.

Corollary 3. *Suppose X is a Banach space with an* LUR *norm the dual norm of which is* LUR. *Then every l.s.c. proper extended-real-valued function f on X can be approximated pointwise by continuously differentiable functions.*

Proof. Let $F(x)=\exp f(x)$ if $x\in\operatorname{dom} f$; $F(x)=+\infty$ otherwise. The function F maps X into $(0,+\infty]$ and is lower semicontinuous. The approximates $F_{t,s}$, $0<s<t$, are C^1 and >0. Indeed, if $\inf F_{t,s}=0$, then $\arg\min F_{t,s}=\arg\min F=\emptyset$ since F is nowhere 0 [Theorem 1(iii)]. Therefore, $\log F_{t,s}$ is C^1 and clearly $\log F_{t,s}(x)\to f(x)$ when $0<s<t\downarrow 0$ for each $x\in X$ [Theorem 1(v)]. □

By a similar argument we find the following variant of [14, Theorem 1].

Corollary 4. *If the norm of X and its dual norm are both* UR, *every real-valued function f on X which is uniformly continuous on bounded sets can be approximated uniformly on bounded sets by differentiable functions the differentials of which are uniformly continuous on bounded sets.*

4. Proof of Theorem 1

In the proof of Theorem 1 we shall use some auxiliary notation and definitions.

Notation 1. Unless otherwise stated, X will denote an LUR Banach space whose dual is also LUR. We introduce $K\colon X^2\to\mathbf{R}$ by putting

$$K(x,y)=\tfrac{1}{2}\|x\|^2+\tfrac{1}{2}\|y\|^2-\langle y,J(x)\rangle \quad \text{for all } (x,y)\in X^2,$$

where J denotes the *duality map* $J=d\|\cdot\|^2/2$ mapping X into X^*.

If X is a Hilbert space, J is simply the identity operator in X and $K(x,y)$ is nothing else but $\|x-y\|^2/2$.

Lemma 1. *The following hold:*
(i) *$K(x,x)=0$ for all $x\in X$;*
(ii) *For any $x\in X$,*

$$K(x,y)\geq\Delta_x(\|x-y\|) \quad \textit{for all } y\in X,$$

where Δ_x *is the function in* (2);

(iii) *If* $\|\cdot\|$ *is* UR *rather than just* LUR,

$$K(x,y)\ge \varrho^2\Delta(\|x-y\|/\varrho) \quad \textit{if } \varrho>0,\ x\in B(0,\varrho), \textit{ and } y\in X;$$

for the meaning of Δ *study* (3).

Proof. Part (i) translates to the familiar identity: $\langle x, J(x)\rangle=\|x\|^2$.

(iii) It suffices to consider the case in which $\varrho=1$ since K is homogeneous of second degree. Bearing inequality (3) in mind,

$$\begin{aligned}K(x,y)&=\tfrac12\|x\|^2+\tfrac12\|y\|^2-\langle y, J(x)\rangle\\ &\ge \|(x+y)/2\|^2+\Delta(\|x-y\|)-2\langle (x+y)/2, J(x)\rangle+\langle x, J(x)\rangle\\ &\ge \|(x+y)/2\|^2+\Delta(\|x-y\|)-2\big(\tfrac12\|(x+y)/2\|^2+\tfrac12\|x\|^2\big)+\|x\|^2\\ &=\Delta(\|x-y\|)\end{aligned}$$

provided $x\in B(0,1)$ and $y\in X$.

Assertion (ii) admits an analogous proof. □

We turn to yet another auxiliary definition.

Notation 2. Let f and g be extended-real-valued functions on X, s and t be positive real numbers. We define f_t and g^s by the formulas

$$\begin{aligned}f_t(z)&=\inf_{y\in X}\left(f(y)+\frac1t K(z,y)\right), \quad z\in X,\\ g^s(x)&=\sup_{z\in X}\left(g(z)-\frac1s K(z,x)\right), \quad x\in X.\end{aligned}$$

Obviously, $f_t\le f$ and $g^s\ge g$. Also, it is clear that if X is a Hilbert space, then f_t is the Moreau–Yosida approximate of order t while g^s equals $-(-g)_s$.

We observe that

$$f_t(z)=\frac1{2t}\|z\|^2-\left(f+\frac1{2t}\|\cdot\|^2\right)^*(J(z)/t).$$

Our interest in these constructions stems from the fact that they provide a decomposition of $f_{t,s}$ which will prove useful in the demonstration of Theorem 1. They also show that our approximates coincide with those of Lasry and Lions when X is a Hilbert space; namely,

Lemma 2. *$f_{t,s}=(f_t)^s$ when $0<s<t$.*

Proof. The sum

$$(f_t)^s(x)+\frac{1}{2s}\|x\|^2$$

equals

$$\begin{aligned}\sup_{z\in X}&\left[\langle x/s,J(z)\rangle-\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(J(z)/t)-\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|z\|^2\right]\\ &=\sup_{z\in X}\left[\langle x/s,J(z)\rangle-\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(J(z)/t)-\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|J(z)\|_*^2\right]\\ &=\sup_{\zeta\in X^*}\left[\langle x/s,\zeta\rangle-\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(\zeta/t)-\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|\zeta\|_*^2\right],\end{aligned}$$

and the last expression is $f_{t,s}(x)+(2s)^{-1}\|x\|^2$. In this string of equalities we have used the identity $\|z\|=\|J(z)\|_*$ and the density of the range of J (the Bishop–Phelps theorem); the latter in conjunction with the continuity of the concave function

$$X^*\ni\zeta\mapsto\langle x/s,\zeta\rangle-\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(\zeta/t)-\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|\zeta\|_*^2.$$

The proof is complete. □

Lemma 3. *Let f be a proper extended-real-valued l.s.c. bounded below function on X. For any $t>0$ it holds that*

(i) *f_t is a real-valued minorant of f;*

(ii) $\inf f_t=\inf f$;

(iii) $\arg\min f_t=\arg\min f$.

Also, when $t\downarrow 0$,

(iv) *$f_t\to f$ pointwise;*

(v) *$f_t\to f$ with respect to the epi-distance topology provided $\|\cdot\|$ is* UR;

(vi) *$f_t\to f$ uniformly on bounded sets when f is uniformly continuous on bounded sets and $\|\cdot\|$ is* UR.

Proof. Parts (i) and (ii) are immediate.

(iii) The assertions (i) and (ii) jointly imply $\arg\min f\subseteq\arg\min f_t$. To establish the reverse inclusion, suppose x minimizes f_t and let (y_j) be a minimizing sequence:

$$f(y_j)+\frac{1}{t}K(x,y_j)\to f_t(x)=\inf f,\quad j\to\infty.$$

Evidently $K(x,y_j)\to 0$ and hence $y_j\to x$ as $j\to\infty$ [Lemma 1(ii)]. By lower semicontinuity we find that

$$f(x)+\frac{1}{t}K(x,x)\leq \inf f,$$

which says $f(x)\leq \inf f$.

(iv) Choose $x\in X$. Of course,

$$\lim_{t\downarrow 0} f_t(x)=\sup_{t>0} f_t(x)\leq f(x).$$

The reverse inequality $\sup_{t>0} f_t(x)\geq f(x)$ is immediate if $\sup_{t>0} f_t(x)=+\infty$. Suppose instead that $S:=\sup_{t>0} f_t(x)\in\mathbf{R}$. Let $0<\varepsilon_t\downarrow 0$ as $t\downarrow 0$ and select y_t such that

$$f(y_t)+\frac{1}{t}K(x,y_t)\leq f_t(x)+\varepsilon_t\leq S+\varepsilon_t. \tag{5}$$

Then $y_t\to x$ as $t\downarrow 0$; otherwise there would exist a $\delta>0$ and a sequence $t_j\downarrow 0$ such that $\|x-y_{t_j}\|\geq\delta$ which would imply, again applying Lemma 1(ii),

$$f(y_{t_j})+\frac{1}{t_j}K(x,y_{t_j})\geq \inf f+\frac{1}{t_j}\Delta_x(\delta)\to+\infty,\quad j\to\infty,$$

contradicting (5). Thanks to the inequality $f(y_t)\leq f_t(x)+\varepsilon_t$ and the lower semicontinuity of f,

$$f(x)\leq \liminf_{t\downarrow 0}(f_t(x)+\varepsilon_t)=\lim_{t\downarrow 0} f_t(x).$$

(v) Fix ϱ, an arbitrary positive real number. Let $\mathbf{U}$ represent the unit ball in $X\oplus\mathbf{R}$ with respect to the box norm: $\mathbf{U}=\{(x,\alpha)\in X\oplus\mathbf{R};\|x\|\leq 1,\ |\alpha|\leq 1\}$. Since f_t minorizes f we have $E(\operatorname{epi} f\cap\varrho\mathbf{U},\operatorname{epi} f_t)=0$. Therefore,

$$\begin{aligned} H_\varrho(f,f_t)&=E(\operatorname{epi} f_t\cap\varrho\mathbf{U},\operatorname{epi} f)\\ &=\sup\{d((x,\alpha),\operatorname{epi} f)\,;\,(x,\alpha)\in\operatorname{epi} f_t,\ \|x\|\leq\varrho,\ \text{and } |\alpha|\leq\varrho\}.\end{aligned} \tag{6}$$

We claim that there exists a positive function $t\mapsto r(t)$ with $r(t)\to 0$ as $t\downarrow 0$ such that

$$f_t(x)=\inf_{y\in B(x,r(t))}\left(f(y)+\frac{1}{t}K(x,y)\right) \tag{7}$$

whenever $x\in B(0,\varrho)$ and $f_t(x)\leq\varrho$. In fact,

$$r(t)=\varrho\Delta^{-1}(t(\varrho-\inf f+1)/\varrho^2)$$

will do. (We may and do assume that Δ has a continuous inverse function.) Indeed, if $\|x\|\le\varrho$ and $\|x-y\|>r(t)$, by Lemma 1(iii),

$$f(y)+\frac{1}{t}K(x,y)\ge\inf f+\frac{1}{t}\varrho^2\Delta(r(t)/\varrho)=\varrho+1$$

implying (7) provided $f_t(x)\le\varrho$, and $r(t)$ tends to zero as $t\downarrow 0$, establishing the claim.

To estimate $H_\varrho(f,f_t)$ [see (6)] let $(x,\alpha)\in\operatorname{epi} f_t$, $\|x\|\le\varrho$, and $|\alpha|\le\varrho$. In particular, $f_t(x)\le\varrho$ so that formula (7) for $f_t(x)$ is valid. Fix $t>0$ temporarily, choose $\varepsilon>0$, and pick $y\in B(x,r(t))$ such that

$$f(y)+\frac{1}{t}K(x,y)\le f_t(x)+\varepsilon.$$

Then $f(y)\le f_t(x)+\varepsilon\le\alpha+\varepsilon$ and thus $(y,\alpha+\varepsilon)\in\operatorname{epi} f$. Consequently, $d((x,\alpha),\operatorname{epi} f)$ is at most the norm of $(x,\alpha)-(y,\alpha+\varepsilon)$:

$$d((x,\alpha),\operatorname{epi} f)\le\max\{\|x-y\|,\varepsilon\}\le\max\{r(t),\varepsilon\}.$$

Therefore, ε being arbitrary, $d((x,\alpha),\operatorname{epi} f)\le r(t)$ from which it follows that

$$H_\varrho(f,f_t)\le r(t).$$

We conclude that $f_t\to f$ for the epi-distance topology since $r(t)\to 0$ as $t\downarrow 0$.

(vi) Let us verify that

$$\sup_{x\in B(0,\varrho)}|f(x)-f_t(x)|=\sup_{x\in B(0,\varrho)}(f(x)-f_t(x))$$

approaches zero as $t\downarrow 0$. We can find $R>\varrho$ such that

$$f_t(x)=\inf_{y\in B(0,R)}\left(f(y)+\frac{1}{t}K(x,y)\right)\quad\text{when } t\in(0,1],\ x\in B(0,\varrho). \tag{8}$$

To see this notice that f is bounded on $B(0,\varrho)$ by say M, and that $K(x,y)\ge(\|y\|-\|x\|)^2/2$. Applying these observations we find that for arbitrary $x\in B(0,\varrho)$, $t\in(0,1]$,

$$f(y)+\frac{1}{t}K(x,y)\ge\inf f+\frac{1}{2}(\|y\|-\|x\|)^2\ge M+1$$

provided $\|y\|>\varrho+[2(M+1-\inf f)]^{1/2}=:R$; whence (8) as claimed.

With m being a modulus of continuity for $f|_{B(0,R)}$; in other words,

$$|f(x)-f(y)|\le m(\|x-y\|)\quad\text{for all } x,y\in B(0,R),$$

where m is continuous, non-decreasing, subadditive, with $m(0)=0$, we have

$$\begin{aligned}\sup\{f(x)-f_t(x)\,;x\in B(0,\varrho)\}&\\ &=\sup\{f(x)-f(y)-t^{-1}K(x,y)\,;x\in B(0,\varrho),\ y\in B(0,R)\}\\ &\le\sup\{m(\|x-y\|)-t^{-1}\varrho^2\Delta(\|x-y\|/\varrho)\,;x\in B(0,\varrho),\ y\in B(0,R)\}\\ &=\sup\{m(u)-t^{-1}\varrho^2\Delta(u/\varrho)\,;u\in[0,R+\varrho]\}\\ &\to 0\quad\text{as } t\downarrow 0.\end{aligned}$$

A proof of the elementary statement "$\to 0$" is omitted. □

The following lemma, which exploits the duality between differentiability and uniform convexity (for convex functions), will be utilized in the proof of smoothness of $f_{t,s}$.

Lemma 4. *Suppose $h\in\mathbf{\Gamma}(X)$ where X is a Banach space. Suppose moreover that $h^*(\xi)/\|\xi\|_*\to+\infty$ when $\|\xi\|_*\to+\infty$ and that h^* is locally uniformly convex. Then h is Lipschitz continuous on bounded sets and C^1. If h^* happens to be uniformly convex on bounded sets, then the differential dh is uniformly continuous on bounded sets.*

The assumption that h^* be locally uniformly convex means that to each $\xi\in\operatorname{dom}h^*$ there should exist a function p_ξ on $[0,+\infty)$ such that

$$h^*((\xi+\eta)/2)\le\tfrac12h^*(\xi)+\tfrac12h^*(\eta)-p_\xi(\|\xi-\eta\|_*)\quad\text{for all }\eta\in\operatorname{dom}h^*$$

and $p_\xi(u)>0$ if $u>0$. The concept of "uniform convexity on bounded subsets" is defined similarly.

The contents of the lemma are essentially classical, see for instance [1], [18], but for the sake of completeness we include a proof.

Proof. The strong coerciveness of h^* clearly implies that h is bounded and hence Lipschitz continuous on bounded subsets of X.

Let $x;x_0,x_1,\dots$ be elements of X such that $\|x-x_j\|\to0$ as $j\to\infty$. Choose $\xi\in\partial h(x)$ and $\xi_j\in\partial h(x_j)$, $j\in\mathbf{N}$. The differentiability follows if we establish that $\|\xi-\xi_j\|_*\to0$. To this end we make use of the identities

$$\langle x,\xi\rangle=h(x)+h^*(\xi),\quad\langle x_j,\xi_j\rangle=h(x_j)+h^*(\xi_j)\text{ for all }j\in\mathbf{N}.$$

These equations imply

$$\begin{aligned}2h^*((\xi+\xi_j)/2)-h^*(\xi)-h^*(\xi_j)&\\ &\ge2(\langle(x+x_j)/2,(\xi+\xi_j)/2\rangle-h((x+x_j)/2))-h^*(\xi)-h^*(\xi_j)\\ &\ge-\langle x-x_j,\xi-\xi_j\rangle/2\to0\quad\text{when } j\to\infty,\end{aligned}$$

from which it follows, thanks to the local uniform convexity of h^*, that $\|\xi-\xi_j\|_*\to 0$. Hence h is C^1.

Replacing x_j, ξ, and ξ_j by y, $dh(x)$, and $dh(y)$, respectively, we find that

$$(9)\qquad h^*((dh(x)+dh(y))/2)-\tfrac{1}{2}h^*(dh(x))-\tfrac{1}{2}h^*(dh(y))\geq \tfrac{1}{4}\langle x-y, dh(x)-dh(y)\rangle.$$

Assume h^* is uniformly convex on bounded subsets, choose $\varrho>0$, and let L be a Lipschitz constant for $h|_{B(0,\varrho)}$. By assumption there exists a non-decreasing function p such that

$$(10)\qquad h^*((\xi+\eta)/2)\leq \tfrac{1}{2}h^*(\xi)+\tfrac{1}{2}h^*(\eta)-p(\|\xi-\eta\|_*)\quad \text{if } \|\xi\|_*\leq L,\ \|\eta\|_*\leq L,$$

and $p(u)>0$ if $u>0$. The inequalities (9) and (10) imply

$$4p(\|dh(x)-dh(y)\|_*)\leq \|x-y\|\,\|dh(x)-dh(y)\|_* \quad \text{when } x,y\in B(0,\varrho),$$

since $\|dh(x)\|_*\leq L$ for all $x\in B(0,\varrho)$. The preceding inequality forces dh to be uniformly continuous on $B(0,\varrho)$. □

Now we are ready to prove Theorem 1.

Proof of Theorem 1. (i) To begin we observe that $\|\cdot\|_*^2$ is locally uniformly convex since $\|\cdot\|_*$ is locally uniformly rotund by hypothesis. The function

$$X^*\ni\xi\mapsto\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(\xi/t)+\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|\xi\|_*^2$$

is strongly coercive and locally uniformly convex for the sum of a convex function and a locally uniformly convex function is again locally uniformly convex. Invoking Lemma 4 it is then clear that

$$X\ni x\mapsto f_{t,s}(x)=\left[\left(f+\frac{1}{2t}\|\cdot\|^2\right)^*(\cdot/t)+\left(\frac{1}{s}-\frac{1}{t}\right)\frac{1}{2}\|\cdot\|_*^2\right]^*(x/s)-\frac{1}{2s}\|x\|^2$$

is C^1 and Lipschitz continuous on bounded sets.

(ii) Again by Lemma 4, $df_{t,s}$ is uniformly continuous on bounded sets if $\|\cdot\|_*$ is UR.

The infimal convolute $g\Box h$ is $C^{1,1}$ provided $g\in\Gamma(X)$, and h is convex, $C^{1,1}$, $h(x)/\|x\|\to\infty$ when $\|x\|\to\infty$; see for instance [13, Proposition 2.5]. As a particular case, $f_{t,s}$ is $C^{1,1}$ provided $\|\cdot\|^2$ is $C^{1,1}$ [see (4)].

(iii) With $y=x$ in the infimum (4), $f_{t,s}(x)\leq f(x)$. Evidently, $f_{t,s}=(f_t)^s\geq f_t$, see Lemma 2. The established inequalities $f_t\leq f_{t,s}\leq f$ imply, by virtue of the equality $\inf f_t=\inf f$ [Lemma 3(ii)], that

$$\inf f=\inf f_t\leq \inf f_{t,s}\leq \inf f.$$

On the other hand, taking into account the identity $\arg\min f_t = \arg\min f$, from Lemma 3(iii), it follows that

$$\arg\min f = \arg\min f_t \supseteq \arg\min f_{t,s} \supseteq \arg\min f.$$

Assertion (iv) is obvious.

(v), (vi) The inequalities $f_t \leq f_{t,s} \leq f$ and the pointwise convergence $f_t \to f$ [Lemma 3(iv)] together imply the pointwise convergence $f_{t,s} \to f$. Arguing similarly, with "pointwise convergence" replaced either by "convergence in epi-distance" or by "uniform convergence on bounded sets", the remaining assertions follow with the aid of Lemma 3(v), (vi). □

5. A remark on second-order smoothness

Unfortunately, the method in this paper need not provide approximates enjoying second-order differentiability, even if the underlying space X is finite-dimensional. In fact, the following elementary example shows that we cannot expect more regularity than $C^{1,1}$ smoothness and that it hence becomes necessary to combine the method with another one in order to achieve second-order differentiability.

Example 1. Let f be defined on the line by the assignments

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x^2/2 & \text{if } x > 0. \end{cases}$$

Then f is $C^{1,1}$ and convex but fails to be twice differentiable at the origin. A straightforward computation yields $f_{t,s} = (1+t-s)^{-1} f$, $0 < s < t$. Regularity is thus neither gained nor lost.

In this connection it should be noted, as pointed out in [14] and [13], respectively, that

(a) *There exists on l^2 a $C^{1,1}$ function f that fails to be uniformly approximable by functions with two uniformly continuous derivatives (although l^2 admits C^∞ partitions of unity);*

(b) *Certain separable Banach spaces have norms such that the associated duality maps J are Lipschitz continuous yet admit no C^2 bump functions.*

Nevertheless, if X is a Banach space with the properties in (6), then every convex $f\colon X \to \mathbf{R}$ which is Lipschitz continuous on bounded sets can be approximated uniformly on bounded sets by twice Gâteaux differentiable $C^{1,1}$ convex functions; study [13]. As we shall see in this section the convexity assumptions are superfluous which will be made clear by combining our Theorem 1 with results in [8], [13].

Accordingly, let us briefly describe the approach of these authors. Suppose that X is a separable Banach space and that $f\colon X\to\mathbf{R}$ is convex and $C^{1,1}$. To approximate f by twice Gâteaux differentiable convex functions whose first differentials remain Lipschitz continuous we may proceed as follows. Fix $e_1, e_2, \ldots$, a sequence which is dense in the unit sphere of X. For $0\le\psi\in C_0^\infty(\mathbf{R})$ with $\int\psi(t)\,dt=1$ put $\psi_j(t)=2^j\psi(2^j t)$ and

$$f^\psi(x)=\lim_{n\to\infty}\int_{\mathbf{R}^n} f\big(x-\textstyle\sum_{j=1}^n t_j e_j\big)\prod_{j=1}^n\psi_j(t_j)\,dt_1\,dt_2\ldots dt_n,\quad x\in X.$$

It turns out that the functions f^ψ have the desired regularity and that $f^\psi\to f$ uniformly on bounded subsets of X as $\operatorname{supp}\psi\to\{0\}$.

Recall that a function $g\colon X\to\mathbf{R}$ is said to be *twice Gâteaux differentiable* at $x\in X$ if g is Gâteaux differentiable in some neighborhood of x, and if

$$g''(x)(h,k)=\lim_{\lambda\to 0}\langle h, g'(x+\lambda k)-g'(x)\rangle/\lambda$$

exists for each $(h,k)\in X^2$ making $g''(x)$ a continuous symmetric bilinear form.

Theorem 2. *Let X be a separable Banach space. If X admits an equivalent smooth* UR *norm $\|\cdot\|$ such that $J=d\|\cdot\|^2/2$ is globally Lipschitz continuous, every bounded below $f\colon X\to\mathbf{R}$ which is uniformly continuous on bounded sets can be approximated uniformly on bounded sets by twice Gâteaux differentiable functions whose first differentials are globally Lipschitz continuous. Moreover, the approximates may be chosen as differences of convex functions.*

Proof. Theorem 1 implies that f can be approximated uniformly on bounded sets by $C^{1,1}$ functions $g-q$ with the functions g convex and the functions q positive multiples of $\|\cdot\|^2/2$. According to the preceding considerations (for details see [8, Theorem 3.1] and [13, Lemma 2.6]) such convex functions g and q can, in turn, be approximated uniformly on bounded sets by convex functions enjoying the asserted regularity. □

6. Approximation in Hilbert spaces

Henceforth X will be a Hilbert space with scalar product $(\cdot\,|\,\cdot)$. A function f on X will be called *locally demi-convex* if to each $x\in X$ there correspond c and r, positive reals, such that $f+(2c)^{-1}\|\cdot\|^2$ is convex on the ball $B(x,r)$. We shall demonstrate that the continuous locally demi-convex real-valued functions on X fit the regularization scheme like a glove. In the process, various properties of the Moreau–Yosida approximates will be discussed.

We remark that R. T. Rockafellar (see [16]) showed that a continuous function $f\colon \mathbf{R}^n \to \mathbf{R}$ is locally demi-convex if and only if it is *lower-*C^2 which means, somewhat loosely, that f can be expressed locally as a pointwise maximum of C^2 functions. To be precise, it is required that with each x_0 there may be associated a function $F\colon \mathcal{X}\times\mathcal{K}\to\mathbf{R}$, where $\mathcal{X}$ is an open neighborhood of x_0 while $\mathcal{K}$ is a compact space, which along with its first and second derivatives is jointly continuous, that is, $(x,y)\mapsto(F(x,y),F'_x(x,y),F''_{xx}(x,y))$ is continuous, and such that the representation

$$f(x)=\max_{y\in\mathcal{K}} F(x,y), \quad x\in\mathcal{X},$$

holds.

The lower-C^2 functions were singled out by Rockafellar for their nice properties with respect to subgradient optimization.

Notation 3. Let f be an extended-real-valued function on X and $t>0$. By $C(t)f$ we understand the function

$$C(t)f=\overline{\mathrm{co}}(f+(2t)^{-1}\|\cdot\|^2)-(2t)^{-1}\|\cdot\|^2.$$

The following proposition will prove useful in what follows. It is of independent interest too.

Proposition 2. *Let f be an l.s.c. extended-real-valued function on X.*

(i) $f_{t,s}=[C(t)f]_{t-s}$ *when* $0<s<t$.

(ii) *If* $T>0$ *and* f_T *is real-valued, the following conditions are equivalent*:

(a) $f_{t,s}=f_{t-s}$ *for all* $0<s<t\le T$;

(b) $f+(2T)^{-1}\|\cdot\|^2\in\mathbf{\Gamma}(X)$.

Proof. (i) Equation (4) may be recast in the desired form

$$f_{t,s}(x)=\inf_{y\in X}\left[\overline{\mathrm{co}}\left(f+\frac{1}{2t}\|\cdot\|^2\right)(y)-\frac{1}{2t}\|y\|^2+\frac{1}{2(t-s)}\|x-y\|^2\right]$$

by applying the identity

$$\frac{1}{2t}\|y\|^2+\frac{1}{1/s-1/t}\frac{1}{2}\left\|\frac{x}{s}-\frac{y}{t}\right\|^2-\frac{1}{2s}\|x\|^2=\frac{1}{2(t-s)}\|x-y\|^2,$$

which holds true when $\|\cdot\|$ is the canonical norm of a Hilbert space.

(ii) The implication (b) $\Rightarrow$ (a) follows at once from (i).

Assume (a) so that, in particular, $[C(T)f]_{T-s}=f_{T-s}$ when $0<s<T$. By letting $s\uparrow T$ we find that $C(T)f=f$, which is a reformulation of (b). $\square$

We investigate next the local behaviour of $C(t)f$ for small values of t and functions f that satisfy certain local convexity conditions.

Lemma 5. *Suppose that f is bounded from below, $x_0 \in X$, and that $f + (2c)^{-1}\|\cdot\|^2$ is convex on the ball $B(x_0, R)$, where $c, R > 0$. Assume moreover that f satisfies, on $B(x_0, R)$, a Lipschitz condition of rank L. Let $0 < r < R$ and*

$$T = \min\{(R-r)^2[2(f(x_0) - \inf f + RL)]^{-1}, c\}.$$

Then

$$C(t)f(x) = f(x) \quad \text{when } x \in B(x_0, r) \text{ and } t \in (0, T].$$

Proof. Let $x \in B(x_0, r)$ and $\xi \in \partial f(x)$. Let $t \in (0, T]$. It suffices to show that $\xi + t^{-1}x$ is the "slope" of a continuous affine minorant of $f + (2t)^{-1}\|\cdot\|^2$ which is exact at x.

If $y \in \complement B(x_0, R)$,

$$\begin{aligned}\left(f(y) + \frac{1}{2t}\|y\|^2\right) - \left(f(x) + \frac{1}{2t}\|x\|^2\right) - (y - x|\xi + t^{-1}x) \\ = f(y) - f(x) + \frac{1}{2t}\|y - x\|^2 - (y - x|\xi) \\ \geq \inf f - (Lr + f(x_0)) + \frac{1}{2t}\|y - x\|^2 - \|y - x\|L \\ \geq \inf f - (Lr + f(x_0)) + \frac{1}{2t}(R - r)^2 - (R - r)L \geq 0\end{aligned}$$

since $t \leq (R-r)^2[2(f(x_0) - \inf f + RL)]^{-1}$.

If $y \in B(x_0, R)$,

$$\left(f(y) + \frac{1}{2t}\|y\|^2\right) - \left(f(x) + \frac{1}{2t}\|x\|^2\right) - (y - x|\xi + t^{-1}x) \geq 0$$

by the assumed convexity. □

It is well known that the derivatives of the Moreau–Yosida approximates df_t converge to ∂f if $f \in \mathbf{\Gamma}(X)$. We aim to prove that $df_{t,s} \to \partial f$ if f is a locally demi-convex continuous function. To reach this goal we start with demi-convex functions.

Notation 4. If $\mathcal{C}$ is a nonempty convex closed subset of X, we denote by $[\mathcal{C}]_0$ the element of minimal norm in $\mathcal{C}$.

Proposition 3. *Suppose $f + (2T)^{-1}\|\cdot\|^2 \in \mathbf{\Gamma}(X)$ where $T > 0$.*

(i) *Assume $\partial f(x)$ is nonempty. Then $df_t(x) \to [\partial f(x)]_0$ when $t \downarrow 0$.*

(ii) *Assume $\partial f(x)$ is nonempty. Then $f_t(x) = f(x) - (t/2)\|[\partial f(x)]_0\|^2 + o(t)$ as $t \downarrow 0$.*

(iii) *At each* $(x,t)\in X\times(0,T)$,

$$\partial f_t(x)/\partial t+\|df_t(x)\|^2/2=0.$$

(iv) *For arbitrary* $0<t<T$, $x\in X$,

$$f_t(x)=f(x) \iff df_t(x)=0 \iff \partial f(x)\ni 0.$$

Proof. Denote by g the function $f+(2T)^{-1}\|\cdot\|^2$; thus $g\in\mathbf{\Gamma}(X)$.

(i), (ii) It is readily verified that

$$f_t(x)=g_{tT/(T-t)}\left(\frac{Tx}{T-t}\right)-\frac{1}{2(T-t)}\|x\|^2,\quad 0<t<T,\ x\in X. \tag{11}$$

In particular, f_t is $C^{1,1}$ when $0<t<T$. Using convex duality,

$$\begin{aligned} g_{tT/(T-t)}\left(\frac{Tx}{T-t}\right)&=\max_{\xi\in X}\left[\left(\frac{Tx}{T-t}\,\middle|\,\xi\right)-g^*(\xi)-\frac{tT}{2(T-t)}\|\xi\|^2\right]\\ &=\frac{t}{2T(T-t)}\|x\|^2-\min_{\xi\in X}\left[g^*(\xi)-(x|\xi)+\frac{tT}{2(T-t)}\left\|\xi-\frac{x}{T}\right\|^2\right]. \end{aligned}$$

These minimization problems have unique solutions denoted ξ_t which are equal to $dg_{tT/(T-t)}(Tx/(T-t))$, $0<t<T$, by classical convex analysis. Letting

$$\Phi(\xi)=g^*(\xi)-(x|\xi),\quad V(\xi)=\tfrac{1}{2}\|\xi-T^{-1}x\|^2,\quad \varepsilon(t)=\frac{tT}{T-t},$$

the minimization problems read

$$\min_{\xi\in X}(\Phi(\xi)+\varepsilon(t)V(\xi)).$$

Observe that $\arg\min\Phi=\partial g(x)$ is nonempty. Viscosity methods are now applicable: since $\Phi\in\mathbf{\Gamma}(X)$, V is convex, continuous, strongly coercive, non-negative, and such that $\eta_j\to\eta$ weakly and $V(\eta_j)\to V(\eta)$ implies $\|\eta_j-\eta\|\to 0$, we have, according to [2, Theorem 5.1],

(a) $\xi_t\to\xi_0$ where $\xi_0\in\arg\min\Phi$;
(b) $V(\xi_0)=\min\{V(\xi);\xi\in\arg\min\Phi\}$; and
(c) $\min(\Phi+\varepsilon(t)V)=\min\Phi+t\min\{V(\xi);\xi\in\arg\min\Phi\}+o(t)$.

Because $\arg\min\Phi = \partial g(x)$, (a) and (b) imply that $\xi_t \to \xi_0$, $\xi_0 \in \partial g(x)$, and

$$\|\xi_0 - T^{-1}x\| = \min\{\|\xi - T^{-1}x\|\,;\xi \in \partial g(x)\} = \min\{\|\eta\|\,;\eta \in \partial f(x)\}.$$

Differentiating (11) and passing to the limit,

$$df_t(x) = \frac{T}{T-t}(\xi_t - T^{-1}x) \to \xi_0 - T^{-1}x = [\partial f(x)]_0 \quad \text{when } t \downarrow 0,$$

as asserted in (i).

The asymptotic development (c) implies (ii) since a direct calculation reveals that

$$\min(\Phi + \varepsilon(t)V) - \min\Phi = f(x) - f_t(x).$$

(iii) Equation (11) implies that $f_t + [2(T-t)]^{-1}\|\cdot\|^2$ is convex when $0<t<T$, while $-f_t + (2t)^{-1}\|\cdot\|^2$ is convex for all $t>0$. On the one hand, by (ii) and the semigroup property $(f_t)_h = f_{t+h}$,

$$\lim_{h\downarrow 0}\frac{f_{t+h}(x) - f_t(x)}{h} = \lim_{h\downarrow 0}\frac{(f_t)_h(x) - f_t(x)}{h} = -\tfrac{1}{2}\|df_t(x)\|^2.$$

On the other hand, $-(-f_t)_h = f_{t-h}$ when $0<h<t<T$ [Proposition 2(ii)]; hence,

$$\lim_{h\downarrow 0}\frac{f_t(x) - f_{t-h}(x)}{h} = \lim_{h\downarrow 0}\frac{(-f_t)_h(x) - (-f_t)(x)}{h} = -\tfrac{1}{2}\|d(-f_t)(x)\|^2.$$

For a proof of assertion (iv) see [7] or [3]. □

Our final theorem gives results on the convergence of the differentials $df_{t,s}$ to ∂f, and on the preservation of stationary points and values. In this connection we accentuate the fact that stationary points, in contrast to minimizers, need not be preserved by the approximation process.

Example 2. We consider three functions neither of which is locally demi-convex (lower-C^2).

(a) Let $f\colon \mathbf{R}\to\mathbf{R}$ be the concave C^1 function defined via $f(x)=0$ if $x\le 0$, $f(x) = -x^{3/2}$ if $x>0$. An elementary calculation shows that $df_{t,s}(0)<0$ when $0<s<t$ whereas evidently $df(0)=0$.

(b) Let $g(x) = \min\{0, -x\}$ for all $x\in\mathbf{R}$. Then $dg_{t,s}(0) = -1$ if $0<s\le t/2$ and $\partial g(0) = [-1, 0] \ni 0$.

(c) Let n be an arbitrary positive integer. R. T. Rockafellar exhibited in [16] a Lipschitz continuous function h on $\mathbf{R}^n$ whose subdifferential $\partial h(x)$ is equal to $[-1,1]^n$ at every $x\in\mathbf{R}^n$. For this h every point $x\in\mathbf{R}^n$ is stationary in the sense that $\partial h(x)\ni 0$. But $dh_{t,s}$ is not identically equal to zero for any $0<s<t$. Otherwise some $h_{t,s}$ would be constant which in turn would force h to be constant, a contradiction. Indeed, $h_{t,s}=k$, k being a real constant, implies $\arg\min h = \arg\min h_{t,s} = \mathbf{R}^n$ and $\inf h = \inf h_{t,s} = k$ according to Theorem 1.

In Theorem 3 we use a localization argument to obtain the promised results.

Theorem 3. *Let X be a Hilbert space and $f\colon X\to\mathbf{R}$ be bounded from below, continuous, and locally demi-convex.*

(i) *To each $x_0\in X$ there exist $r>0$ and $T>0$ such that*

$$\begin{aligned} f_{t,s}(x) &= (f+\mathfrak{I}_{B(x_0,2r)})_{t-s}(x) \\ &= f(x)-(t-s)\|[\partial f(x)]_0\|^2/2+\gamma(x,t-s) \\ &\geq f(x)-(t-s)L^2/2 \end{aligned}$$

for all $x\in B(x_0,r)$ and $0<s<t\leq T$, where $\gamma(x,u)/u\to 0$ as $u\downarrow 0$ and L is a Lipschitz constant for $f|_{B(x_0,2r)}$;

(ii) *$df_{t,s}\to\partial f$ in the sense of Kuratowski–Painlevé convergence of graphs as $0<s<t\downarrow 0$;*

(iii) $\lim_{0<s<t\downarrow 0} df_{t,s}(x)=[\partial f(x)]_0$ *for every $x\in X$;*

(iv) (Preservation of stationary points and values.) *The following conditions relative to an arbitrary point $x\in X$ are mutually equivalent:*

(a) $\partial f(x)\ni 0$;

(b) *$df_{t,s}(x)=0$ for all sufficiently small $0<s<t$;*

(c) *$f_{t,s}(x)=f(x)$ for all sufficiently small $0<s<t$.*

Proof. We observe first that in the definition of $f_\lambda(x)$ [see (1)] it is sufficient to restrict the infimum to $y\in B(x,r)$, that is to say

$$f_\lambda(x)=\inf_{y\in B(x,r)}\left(f(y)+\frac{1}{2\lambda}\|x-y\|^2\right),$$

provided

$$[2\lambda(f(x)-\inf f)]^{1/2}<r. \tag{12}$$

We proceed by analyzing the "domain of dependence" of $f_{t,s}$ upon f. Choose positive real numbers r, c such that $f+(2c)^{-1}\|\cdot\|^2$ is convex and Lipschitz continuous on the ball $B(x_0,3r)$. By Lemma 5 there is $T\in(0,c]$ such that

$$C(t)f(x)=f(x)\quad\text{when } x\in B(x_0,2r) \text{ and } t\in(0,T]. \tag{13}$$

Moreover, by if necessary decreasing T we can assume that when $x\in B(x_0,r)$, $0<s<t\leq T$, we have

$$[C(t)f]_{t-s}(x)=\inf_{y\in B(x,r)}\left(C(t)f(y)+\frac{1}{2(t-s)}\|x-y\|^2\right)\quad\text{and} \tag{14}$$

$$f_{t-s}(x)=\inf_{y\in B(x,r)}\left(f(y)+\frac{1}{2(t-s)}\|x-y\|^2\right). \tag{15}$$

Here we have restricted the infima to $y \in B(x,r)$ which is possible for all sufficiently small $0<s<t$ since $C(t)f$ and f are both bounded from below on X and from above on $B(x_0,2r)$. (Indeed, the expression $[...]^{1/2}$ in (12) goes to zero uniformly over $x \in B(x_0,2r)$ as $\lambda \downarrow 0$.) Thus from (13)–(15) and Proposition 2(i) it follows that

$$(16) \qquad f_{t,s}(x) = f_{t-s}(x) = (f + \mathfrak{I}_{\mathcal{B}})_{t-s}(x) \quad \text{when } x \in B(x_0,r),\ 0<s<t \le T,$$

with $\mathcal{B} = B(x_0,2r)$.

By (16), for $0<s<t\le T$,

$$\sup_{x \in B(x_0,r)} (f(x) - f_{t,s}(x)) = \sup_{\substack{x \in B(x_0,r) \\ y \in B(x_0,2r)}} (f(x) - f(y) - 2^{-1}(t-s)^{-1}\|x-y\|^2)$$
$$\le \sup_{u \ge 0} (Lu - 2^{-1}(t-s)^{-1}u^2) = L^2(t-s)/2.$$

Also, by Proposition 3(ii) and (16),

$$f_{t,s}(x) = f(x) - (t-s)\|[\partial f(x)]_0\|^2/2 + \gamma(x,t-s)$$

where $\gamma(x,u)/u \to 0$ as $u \downarrow 0$ for every $x \in B(x_0,r)$, concluding the proof of (i).

(ii) It is known that the derivatives of the Moreau–Yosida approximates dg_λ converge to ∂g in the sense of Kuratowski–Painlevé convergence of graphs provided g is demi-convex and l.s.c., see [3, Theorem 3.4]. An application of this fact to the particular case $g = f + \mathfrak{I}_{\mathcal{B}}$ implies the assertion because of (16) and the local definition of the Clarke subdifferential.

(iii) By (16) it holds that

$$df_{t,s}(x_0) = d(f + \mathfrak{I}_{\mathcal{B}})_{t-s}(x_0), \quad 0<s<t\le T,$$

while it follows from Proposition 3(i) that

$$\lim_{0<s<t\downarrow 0} d(f + \mathfrak{I}_{\mathcal{B}})_{t-s}(x_0) = [\partial(f + \mathfrak{I}_{\mathcal{B}})(x_0)]_0 = [\partial f(x_0)]_0.$$

(iv) Use (16) and Proposition 3(iv). □

Note that the assumption that f be locally demi-convex is also necessary for (i). Simply observe that the formula $f_{T,s} = [C(T)f]_{T-s}$ together with (i) implies, by letting $s \uparrow T$, that $C(T)f(x) = f(x)$ for all $x \in B(x_0,r)$ so that $f + (2T)^{-1}\|\cdot\|^2$ is necessarily convex on $B(x_0,r)$.

We close the paper by remarking that under the hypotheses of Theorem 3, the Moreau–Yosida approximate f_t is $C^{1,1}$ near a given point x_0 for all sufficiently small t. Actually, there exist positive real numbers T and r such that f_t is smooth on $B(x_0,r)$ when $t \in (0,T)$ and

$$\partial f_t(x)/\partial t + \|df_t(x)\|^2/2 = 0 \quad \text{at each } (x,t) \in B(x_0,r) \times (0,T).$$

Also, $f_t \to f$ locally uniformly while $df_t \to \partial f$ in the sense of parts (ii) and (iii) in Theorem 3. Furthermore, obvious analogues of (i) and (iv) are true.

References

1. ASPLUND, E. and ROCKAFELLAR, R. T., Gradients of convex functions, *Trans. Amer. Math. Soc.* **139** (1969), 443–467.
2. ATTOUCH, H., Viscosity solutions of optimization problems. Epi-convergence and scaling, *Sém. Anal. Convexe* **22**:8 (1992), 1–48.
3. ATTOUCH, H. and AZÉ, D., Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry–Lions method, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **10** (1993), 289–312.
4. BEER, G., *Topologies on Closed and Closed Convex Sets,* Mathematics and its Applications **268**, Kluwer Academic Publishers Group, Dordrecht, 1993.
5. BENOIST, J., Convergence de la dérivée de la régularisée Lasry–Lions, *C. R. Acad. Sci. Paris Sér. I Math.* **315** (1992), 941–944.
6. DEVILLE, R., GODEFROY, G. and ZIZLER, V., *Smoothness and Renormings in Banach Spaces,* Pitman Monographs and Surveys in Pure and Applied Mathematics **64**, Pitman, New York, 1993.
7. EKELAND, I. and LASRY, J. M., On the number of periodic trajectories for a Hamiltonian flow on a convex energy surface, *Ann. of Math.* **112** (1980), 283–319.
8. FABIAN, M., WHITFIELD, J. H. M. and ZIZLER, V., Norms with locally Lipschitzian derivatives, *Israel J. Math.* **44** (1983), 262–276.
9. FRONTISI, J., Smooth partitions of unity in Banach spaces, *Rocky Mountain J. Math.* **25** (1995), 1295–1304.
10. LASRY, J. M. and LIONS, P. L., A remark on regularization in Hilbert spaces, *Israel J. Math.* **55** (1986), 257–266.
11. LIONS, P. L., *Generalized Solutions of Hamilton–Jacobi Equations,* Research Notes in Mathematics **69**, Pitman (Advanced Publishing Program), Boston–London–Melbourne, 1982.
12. MCLAUGHLIN, D., Smooth partitions of unity and approximating norms in Banach spaces, *Rocky Mountain J. Math.* **25** (1995), 1137–1148.
13. MCLAUGHLIN, D., POLIQUIN, R., VANDERWERFF, J. and ZIZLER, V., Second-order Gâteaux differentiable bump functions and approximations in Banach spaces, *Canad. J. Math.* **45** (1993), 612–625.
14. NEMIROVSKIĬ, A. S. and SEMENOV, S. M., On polynomial approximation of functions on Hilbert space, *Mat. Sb.* **92 (134)** (1973), 257–281, 344 (Russian). English transl.: *Math. USSR-Sb.* **21** (1973), 255–277.
15. POLIQUIN, R., VANDERWERFF, J. and ZIZLER, V., Renormings and convex composite representations of functions, *Preprint.*
16. ROCKAFELLAR, R. T., Favorable classes of Lipschitz-continuous functions in subgradient optimization, in *Progress in nondifferentiable optimization* (Numinski, E. A., ed.), pp. 125–143, Internat. Inst. Appl. Systems Anal., Laxenburg, 1982.
17. VANDERWERFF, J., Smooth approximations in Banach spaces, *Proc. Amer. Math. Soc.* **115** (1992), 113–120.

18. Zălinescu, C., On uniformly convex functions, *J. Math. Anal. Appl.* **95** (1983), 344–374.

Received January 29, 1996
in revised form May 19, 1996

Thomas Strömberg
Department of Mathematics
University of Lund
P.O. Box 118
S-221 00 Lund
Sweden
email: thomas@maths.lth.se