Differentiability of conjugate functions and perturbed minimization principles

Jonathan Borwein[*] and Jon Vanderwerff[†]

ABSTRACT — We survey the tight connection between differentiability of conjugate functions and perturbed optimization principles.

Key words: Convex Function, conjugate, Fréchet differentability *2000 Mathematics Subject Classification*: Primary 52A41; Secondary 46G05, 46N10, 49J50, 90C25.

1 Introduction

The seminal paper of Asplund and Rockafellar [1] established duality results between Fréchet differentiability and well-posedness for convex functions. In this note we focus on what can be derived when the conjugate is differentiable, but the original function in not necessarily assumed to be convex. Many of the results we present are known; however, our proofs use a theorem on differentiability that is motivated by Šmulian's work [10], and it is our hope that this approach makes transparent the tight connection between differentiability of conjugate functions and certain minimization principles.

We work in real Banach spaces X whose closed unit ball is denoted by B_X. By a *proper function* $f : X \to (-\infty, +\infty]$ we mean a function that is somewhere finite valued. We use the notation $\partial f(x)$ for the *subdifferential* of f at x in the domain of f, and for $\varepsilon > 0$ we denote the ε*-subdifferential* of f at x in the domain of f by $\partial_\varepsilon f(x)$, that is,

$$\partial_\varepsilon f(x) = \{\phi \in X^* : \phi(y) - \phi(x) \leq f(y) - f(x) + \varepsilon,\ x \in X\};$$

when $\varepsilon = 0$, this is the definition of $\partial f(x)$. The *conjugate function* of $f : X \to (-\infty, +\infty]$ is defined for $x^* \in X^*$ by $f^*(x^*) = \sup_{x\in X}\langle x^*, x\rangle - f(x)$. Our main tool will be the following theorem and its variant for conjugate functions.

Theorem 1.1. *Suppose the convex function f is continuous at x_0. Then f is Fréchet differentiable at x_0 if and only if $\phi_n \to \phi$ whenever $\phi_n \in \partial_{\varepsilon_n} f(x_0)$, $\phi \in \partial f(x_0)$ and $\varepsilon_n \to 0^+$, and necessarily ϕ is the Fréchet derivative at f at x_0.*

A proof of this theorem can be found in Zalinescu's book [12, Theorem 3.3.2]. The following shows for conjugate functions that one need only consider the analogous epsilon subgradients in X.

[*]Computer Science Faculty, 325, Dalhousie University, Halifax, NS, Canada, B3H 1W5. Email: `jborwein@cs.dal.ca`.
Research supported by NSERC and the Canada Research Chair Program.
[†]Department of Mathematics, La Sierra University, Riverside, CA. Email: `jvanderw@lasierra.edu`.

Theorem 1.2. *Suppose a conjugate function $f^* : X^* \to (-\infty, +\infty]$ is continuous at ϕ_0. Then f^* is Fréchet differentiable at ϕ_0 if and only if $x_n \to \Phi$ whenever $x_n \in \partial_{\varepsilon_n} f^*(\phi_0)$, $\Phi \in \partial f^*(\phi_0)$ and $\varepsilon_n \to 0^+$. In particular, $\nabla f^*(\phi_0) \in X$ when f^* is Fréchet differentiable at ϕ_0.*

Proof. The "only if"' implication follows from the previous theorem. For the converse, suppose f^* is not Fréchet differentiable at ϕ_0. Then there exist $h_n \in S_{X^*}$, $t_n \to 0^+$ and $\varepsilon > 0$ such that

$$f^*(\phi_0 + t_n h_n) - f^*(\phi_0) - \Phi(t_n h_n) \geq \varepsilon t_n.$$

Choose $x_n \in \partial_{\varepsilon_n} f^*(\phi_0 + t_n h_n)$ where $\varepsilon_n = t_n \varepsilon/2$ (note that the definition of conjugate functions ensures ε-subdifferentials meet X). Then

$$(x_n - \Phi)(t_n h_n) \geq f^*(\phi_0 + t_n h_n) - f^*(\phi_0) - \Phi(t_n h_n) - t_n \frac{\varepsilon}{2} \geq t_n \frac{\varepsilon}{2}.$$

Consequently, $x_n \not\to \Phi$. □

2 Perturbed Minimization Principles

We begin with a basic fact that we include for completeness.

Lemma 2.1. *Suppose $f : X \to (-\infty, +\infty]$ is a proper function such that f^* is Fréchet differentiable at ϕ and $\nabla f^*(\phi) = x_0 \in X$. If f is lower semicontinuous at x_0, then $f^{**}(x_0) = f(x_0)$.*

Proof. Because $f^{**}|_X \leq f$, it suffices to show $f^{**}(x_0) \geq f(x_0)$. Now choose $x_n \in X$ such that

$$\phi(x_n) - f(x_n) \geq f^*(\phi) - \varepsilon_n \quad \text{where } \varepsilon_n \to 0^+.$$

Then $\phi(x_n) - f^{**}(x_n) \geq f^*(\phi) - \varepsilon_n$ and it follows that $x_n \in \partial_{\varepsilon_n} f^*(\phi)$ for all n. According to Theorem 1.1, $x_n \to x_0$. In particular, $\phi(x_n) \to \phi(x_0)$. Therefore,

$$f^{**}(x_0) = \phi(x_0) - f^*(\phi) = \lim_{n\to\infty} \phi(x) - [\phi(x_n) - f(x_n)] = \lim_{n\to\infty} f(x_n).$$

Now f is lower semicontinuous at x_0, and so $\liminf_{n\to\infty} f(x_n) \geq f(x)$. Therefore, $f^{**}(x_0) \geq f(x_0)$ as desired. □

We will say a function f attains its *strong minimum* at $x_0 \in X$ if $\|x_n - x_0\| \to 0$ whenever $f(x_n) \to f(x_0)$ and $f(x_0) = \inf_X f$. We next present a simple minimization principle which can also be found in [12, Theorem 3.9.1] with a slightly different approach.

Proposition 2.2. *Suppose that X is a Banach space and $f : X \to (-\infty, +\infty]$ is a proper lower semicontinuous function such that f^* is Fréchet differentiable at $\phi \in X^*$, then*

(a) $\nabla(f^*)(\phi) = x_0$ where $x_0 \in \operatorname{dom} f$, and

(b) $(f - \phi)$ attains its strong minimum at x_0.

Proof. First, Theorem 1.2 shows $\nabla(f^*)(\phi) = x_0 \in X$ and then Lemma 2.1 shows $f^{**}(x_0) = f(x_0)$, and the Fenchel-Young equality ensures that $f^{**}(x_0) < +\infty$. This shows (a), and moreover implies that $f^*(\phi) = \phi(x_0) - f(x_0)$. Now suppose $(f - \phi)(x_n) \leq (f - \phi)(x_0) + \varepsilon_n$ where $\varepsilon_n \to 0^+$. This implies $x_n \in \partial_{\varepsilon_n} f^*(\phi)$. Because f^* is Fréchet differentiable at ϕ, Theorem 1.1 implies $\|x_n - x_0\| \to 0$ as desired. □

Using differentiability properites of conjugate functions we obtain:

Corollary 2.3 (Fabian, see e.g. [9]). *Suppose that X is a Banach space with the RNP and that $f : X \to (-\infty, +\infty]$ is a lower semicontinuous function for which there exist $a > 0$ and $b \in \mathbb{R}$ such that $f(x) \geq a\|x\| + b$ for all $x \in X$. Then the set $\{x^* \in aB_{X^*} : f - x^*$ attains its strong minimum on $X\}$ is residual in aB_{X^*}.*

Proof. The growth condition implies that $f^*(\phi) \leq -b$ whenever $\|\phi\| \leq a$. Therefore, f^* is continuous on the interior of aB_{X^*}; see e.g. [9, Proposition 3.3]. According to Collier's theorem [4], f^* is Fréchet differentiable on a dense G_δ subset G of aB_{X^*}. By Proposition 2.2, $f - x^*$ attains its strong minimum at $\nabla(f^*)(x^*) \in \operatorname{dom} f$ for each $x^* \in G$. $\square$

Corollary 2.4 (Stegall [11]). *Suppose $C \subset X$ is a nonempty closed bounded convex set with the RNP, and suppose that $f : C \to \mathbb{R}$ is a lower semicontinuous function on C that is bounded below. Then the set $S = \{x^* \in X^* : f - x^*$ attains its strong minimum on $X\}$ is residual in X^*.*

Proof. According to a locallization of Collier's theorem (see [3]), f^* is Fréchet differentiable on a dense G_δ subset of X^*. Hence, like the previous corollary, the result follows from Proposition 2.2. $\square$

The approach to derive variants of Stegall's variational principle using differentiability was used in [6] and then refined in [8] which used a perturbed function in the dual space rather than the conjugate function. We next show the equivalence of Stegall's variational principle with the locallized version of Collier's theorem just used. The key step is:

Proposition 2.5. *Suppose $f : X \to (-\infty, +\infty]$ is a proper lower semicontinuous function with bounded domain. Then $f - \phi_0$ attains its strong minimum at $x_0 \in \operatorname{dom} f$ where $\phi_0 \in X^*$ if and only if f^* is Fréchet differentiable at ϕ_0 with $\nabla f^*(\phi_0) = x_0$.*

Proof. The previous variational principle showed the "if" implication. For the "only if" implication let $M > 0$ be such that $M \geq \operatorname{diam} \operatorname{dom} f$. Now let $0 < r \leq M$ be given; because $f - \phi_0$ attains its strong minimum at x_0, we choose $\varepsilon > 0$ so that $(f - \phi_0)(x_0 + h) \geq \varepsilon$ if $\|h\| \geq r/2$. Define $g(\cdot) = \frac{\varepsilon}{M} d_C(\cdot) + (f - \phi_0)(x_0)$ where $C = \{x : \|x - x_0\| \leq r/2\}$. Then g is a continuous convex function such that $g \leq f - \phi_0$, and $g(x_0) = (f - \phi_0)(x_0)$ and $f^{**} - \phi_0 \geq g$; moreover $g(x) \geq r\varepsilon/(2M)$ whenever $\|x - x_0\| \geq r$, and so

$$(f^{**} - \phi_0)(x) \geq (f^{**} - \phi_0)(x_0) + \frac{r\varepsilon}{2M} \quad \text{if } \|x - x_0\| \geq r.$$

Because $0 < r \leq M$ was arbitrary, this shows $(f^{**} - \phi_0)|_X$ attains its strong minimum at x_0.

Now, $x_0 \in \partial f^*(\phi_0)$ and we suppose $x_n \in \partial_{\varepsilon_n} f^*(\phi_0)$ where $\varepsilon \to 0^+$. Then $(f^{**} - \phi_0)(x_n) \to (f^{**} - \phi_0)(x_0)$ and consequently $\|x_n - x_0\| \to 0$. According to Theorem 1.2, f^* is Fréchet differentiable at ϕ_0 with $\nabla f^*(\phi_0) = x_0$. $\square$

We are not aware that the preceding proposition has been noted in the literature, however, in the case when f is a proper *convex* lower semicontinuous function with no restriction on its domain, it is a well-known result of Asplund and Rockafellar [1]. Moreover, Propositon 2.5 may fail without a boundedness condition on the domain of a lower semicontinuous function. Indeed, let $f(x) = \min\{|x|, 1\}$, and $\phi_0 = 0$. Then $f - \phi_0$ attains it strong minimum at 0, but its conjugate is the indicator function of $\{0\}$ which is not Fréchet differentiable at 0.

Corollary 2.6 (Characterization of perturbed minimization principles)**.** *Let X be a Banach space, and let $C \subset X$ be a closed bounded convex set. Then the following are equivalent.*

(a) Every weak*-lower semicontinuous convex function $f : X^* \to \mathbb{R}$ such that $f \leq \sigma_C$ is Fréchet differentiable on a dense G_δ subset of X^*.

(b) Given any proper lower semicontinuous bounded below function $f : C \to (-\infty, +\infty]$ and $\varepsilon > 0$, there exist $\phi \in \varepsilon B_{X^*}$, $x_0 \in C$ and $k > 0$ such that $f - \phi$ attains its strong minimum at x_0.

Proof. (a) $\Rightarrow$ (b): Suppose $f : C \to \mathbb{R}$ is bounded below on C. Then there exists $a \in \mathbb{R}$ so that $f + a \geq \delta_C$ where δ_C is the indicator function of C. Consequently, $f^* - a = (f + a)^* \leq \delta_C^* \leq \sigma_C$. Given $\varepsilon > 0$, there exists $\phi \in \varepsilon B_{X^*}$ so that $f^* - a$ and hence f^* is Fréchet differentiable at ϕ. According to Proposition 2.2, $f - \phi$ attains its strong minimum at x_0.

(b) $\Rightarrow$ (a): Take any weak*-lower semicontinuous convex $g \leq \sigma_C$ where $\sigma_C(\phi) = \sup_C \phi$ for $\phi \in X^*$. Let $f = g^*|_X$. Then $f \geq \delta_C$, and $f^* = g$. Now let $\Lambda \in X^*$ be arbitrary, then $f + \Lambda$ is bounded below on C, so $f + \Lambda$ is strongly exposed by some $\phi \in \varepsilon B_{X^*}$. This implies $(f + \Lambda)^*$ is Fréchet differentiable at ϕ. But $(f + \Lambda)^*(\cdot) = g^*(\cdot - \Lambda)$, and so g^* is Fréchet differentiable at $\Lambda + \phi$. Consequently, the set of points of differentiability of f is a dense (automatically) G_δ-set. $\square$

Concluding Remarks. We should mention that one can analogously study *Hölder smooth* or *Lipschitz smooth points* as studied by Fabian in [5] as a dual condition to minimization principles (naturally, these lead to a quantitative estimate in the convergence rate). A development of this will appear in the authors' forthcoming book [2]. Let us also mention that Lemma 2.1 can be used to show that a proper lower semicontinuous function f is convex when f^* is Fréchet differentiable at all $x^* \in \mathrm{dom}(\partial f^*)$. Then, one can efficiently recapture the result that a weakly closed subset of a Hilbert space is a *Chebyshev set* if and only if it is convex; see, for example, [12, Section 3.9]. Additionally, it is not difficult to formulate *bornological* versions of many of the results given herein; see for example [12, Section 3.9] and [2]. Finally, we should note that the paper [7] provides a unified approach to several variational principles using the notion of fragmentability.

References

[1] E. Asplund and R.T. Rockafellar. Gradients of convex functions. *Trans. Amer. Math. Soc.*, 139:443–467, 1969.

[2] J.M. Borwein and J. Vanderwerff. Convex functions: Constructions, Characterizations and Counterexamples. In preparation.

[3] L. Cheng and Y. Teng. A strong optimization theorem in locally convex spaces. *Chin. Ann. Math.*, 24B:395–402, 2003.

[4] J.B. Collier. The dual of a space with the Radon Nikodym property. *Pacific J. Math.*, 64:103–106, 1976.

[5] M. Fabian. Lipschitz smooth points of convex functions and isomorphic characterizations of Hilbert spaces. *Proc. London Math. Soc.*, 51:113–126, 1985.

[6] M. Fabian and V. Zizler. An elementary approach to some questions in higher order smoothness in Banach spaces. *Extracta Math.*, 14:295–327, 1999.

[7] M. Lassonde and J. Revalski. Fragmentability of sequences of set-valued mappings with applications to variational principles. *Proc. Amer. Math. Soc.*, 133:2637–2646, 2005.

[8] P. Loewen and X. Wang. A generalized variational principle. *Canad. J. Math*, 53:1174–1193, 2001.

[9] R.R. Phelps. *Convex Functions, Monotone Operators and Differentiability*, volume 1364 of *Lecture Notes in Mathematics.* Springer-Verlag, New York–Heildelberg–Berlin, 2nd edition, 1993.

[10] V.L. Smulian. Sur la structure de la sphere unitair dans l'espace de Banach. *Mat. Sb.*, 51:545–561, 1941.

[11] C. Stegall. Optimization of functions on certain subsets of Banach spaces. *Math. Ann.*, 236:171–176, 1978.

[12] C. Zălinescu. *Convex Analysis in General Vector Spaces.* World Scientific, New Jersey-London-Singapore-Hong Kong, 2002.