Future Challenges for Variational Analysis

Jonathan M. Borwein

Abstract Modern non-smooth analysis is now roughly thirty-five years old. In this paper I shall attempt to analyse (briefly): where the subject stands today, where it should be going, and what it will take to get there? In summary, the conclusion is that the first order theory is rather impressive, as are many applications. The second order theory is by comparison somewhat underdeveloped and wanting of further advance.[1]

> It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never-satisfied man is so strange if he has completed a structure, then it is not in order to dwell in it peacefully,but in order to begin another. I imagine the world conqueror must feel thus, who, after one kingdom is scarcely conquered, stretches out his arms for others.—Carl Friedrich Gauss (1777-1855).[2]

1 Preliminaries and Precursors

I intend to first discuss *First-Order Theory*, and then *Higher-Order Theory*—mainly second-order—and only mention passingly higher-order theory which really devolves to second-order theory. I'll finish by touching on *Applications of Variational Analysis* or VA both inside and outside Mathematics, mentioning both successes and limitations or failures. Each topic leads to open questions even in the convex case which I'll refer to as CA. Some issues are technical and specialized, others are

Jonathan M. Borwein
CARMA, University of Newcastle NSW, e-mail: jonathan.borwein@newcastle.edu.au

[1] This paper is dedicated to Boris Mordukhovich on the occasion of his sixtieth birthday. It is based on a talk presented at the *International Symposium on Variational Analysis and Optimization* (ISVAO), Department of Applied Mathematics, Sun Yat-Sen University November 28-30, 2008.

[2] From an 1808 letter to his friend Farkas Bolyai (the father of Janos Bolyai)

some broader and more general. In nearly every case Boris Mordukhovich has made prominent or seminal contributions; many of which are elaborated in [24] and [25].

To work fruitfully in VA it is really important to understand both CA and *smooth analysis* (SA); they are the motivating foundations and very often provide the key technical tools and insights. For example, Figure 1 shows how an essentially strictly convex [8, 11] function defined on the orthant can fail to be strictly convex.

$$(x,y) \mapsto \max\{(x-2)^2+y^2-1, -(xy)^{1/4}\}$$

Understanding this sort of boundary behaviour is clearly prerequisite to more delicate variational analysis of lower semicontinuous functions as are studied in [8, 24, 13, 28].

In this note our terminology is for the most-part consistent with those references and since I wish to discuss patterns, not proofs, I will not worry too much about exact conditions. That said, f will at least be a proper and lower semicontinuous extended-real valued function on a Banach space X.

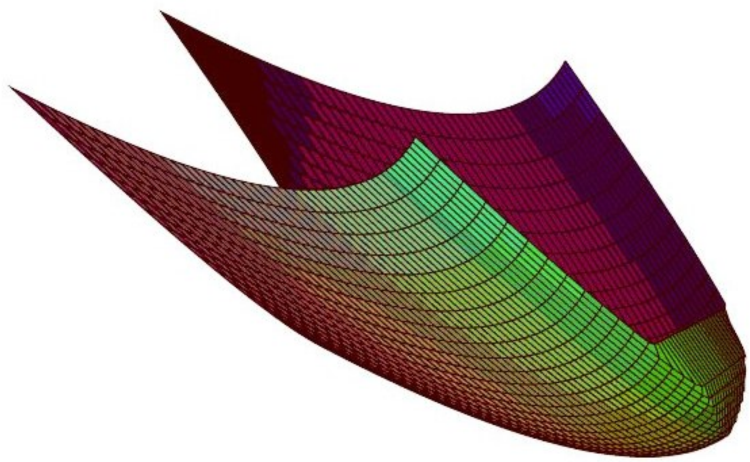

Fig. 1 A function that is essentially strictly but not strictly convex with nonconvex subgradient domain.

Let us first recall the two main starting points:

1.1 A Descriptive Approach

By 1968 Pshenichnii, as described in his book [27], had started a study of the large class of *quasi-differentiable* locally Lipschitz functions for which

$$f'(x;h) := \limsup_{t\to 0^+} \frac{f(x+th)-f(x)}{t}$$

is *required* to exist and be convex as a function of h. We define $\partial' f(x) := \partial_2 f'(x;0)$, where we take the classical convex subdifferential with respect to the second variable.

1.2 A Prescriptive Approach

By contrast, Clarke in his 1972 thesis (described in his now classic book [15]) considered *all* locally Lipschitz functions for which

$$f^\circ(x;h) := \limsup_{t\to 0^+\ y\to x} \frac{f(y+th)-f(y)}{t}$$

is *constructed* to be convex. In convex terms we may now define a generalized subdifferential by $\partial^o f(x) := \partial_2 f^0(x;0)$. (Here the later is again the convex subdifferential with respect to the h variable.)

Both ideas capture the smooth and the convex case, both are closed under $+$ and $\vee$, and both satisfy a reasonable calculus; so we are off to the races. Of course we now wish to do as well as we can with more general lsc functions.

2 First-Order Theory of Variational Analysis

The key players are as I shall outline below. We start with:

1. **The (Fréchet) Subgradient** $\partial_F f(x)$, which denotes a one-sided lower Fréchet subgradient (i.e., the appropriate limit is taken uniformly on bounded sets) and which can (for some purposes) be replaced by a Gâteaux (uniform on finite sets), Hadamard (uniform on norm-compact sets) or weak Hadamard (uniform on weakly-compact sets) object. These are denoted by $\partial_G f(x), \partial_H f(x)$, and $\partial_{WH} f(x)$ respectively.
   That is $\phi \in \partial_F f(x)$, exactly when

   $$\phi(h) \le \liminf_{t\to 0^+\ \|h\|=1} \frac{f(x+th)-f(x)}{t}.$$

   A formally smaller and more precise object is a derivative bundle of F, G, H or WH-smooth (local) minorants:
2. **The Viscosity Subgradient**

   $$\partial_F^v f(x) := \{\phi : \phi = \nabla_F g(x), f(y)-g(y) \ge f(x)-g(x) \text{ for } y \text{ near } x\}$$

as illustrated in Figure 2. By its very definition $0 \in \partial_F^v f(x)$ when x is a local minimizer of f. In nice spaces, say those with Fréchet-smooth renorms as have reflexive spaces, these two subgradient notions coincide [13]. In this case we have access to a good generalization of the sum rule from convex calculus [11]:

3. **(Fuzzy) Sum Rule.** For each $\varepsilon > 0$

$$\partial_F(f+g)(x) \subseteq \partial_F f(x_1) + \partial_F g(x_2) + \varepsilon B_{X^*}$$

for points x_1, x_2 each within ε of x. In Euclidean space and even in Banach space—under quite stringent compactness conditions except in the Lipschitz case—with the addition of *asymptotic subgradients* one can pass to the limit and recapture *approximate* subdifferentials [13, 24, 25, 28].
For now we let ∂f denote any of a number of subgradients and have the appropriate tools to define a workable normal cone.

4. **Normal cones.** We define

$$N_{\mathrm{epi} f} := \partial \iota_{\mathrm{epi} f}.$$

Here ι_C denotes the convex *indicator function* of a set C.
Key to establishing the fuzzy sum rule and its many equivalences [13, 24] are:

5. **Smooth Variational Principles (SVP)** which establish the existence of many points, x, and locally smooth (with respect to an appropriate topology) minorants g such that

$$f(y) - g(y) \geq f(x) - g(x)$$

for y near x.
We can now establish the existence and structure of:

6. **Limiting Subdifferentials** such as

$$\partial^a f(x) := \limsup_{y \to_f x} \partial_F f(x),$$

for appropriate topological limits superior, and of:

7. **Coderivatives of Multifunctions.** As in [24] one may write

$$D^*\Omega(x,y)(y^*) = \{x^* \colon (x^*, -y^*) \in N_{\mathrm{gph}(\Omega)}(x,y)\}.$$

The fuzzy sum rule and its related calculus also leads to fine results about the notion of:

8. **Metric regularity.**
Indeed, we can provide very practicable conditions on a multifunction Ω, see [12, 13, 24, 17], so that locally around $y_0 \in \Omega(x_0)$ one has

$$K d(\Omega(x), y) \geq d(x, \Omega^{-1}(y)). \tag{1}$$

Estimate (1) allows one to show many things easily. For example, it allows one straight forwardly to produce C^k-implicit function theorems under second-order sufficiency conditions [3, 13]. Estimate (1) is also really useful in the very con-

crete setting of alternating projections on two closed convex sets C and D where one uses $\Omega(x) := x - D$ for $x \in C$ and $\Omega(x) := \emptyset$ otherwise [13].
The very recent book by Dontchev and Rockafellar [17] gives a comprehensive treatment of implicit function theory for Euclidean multifunctions (and much more).

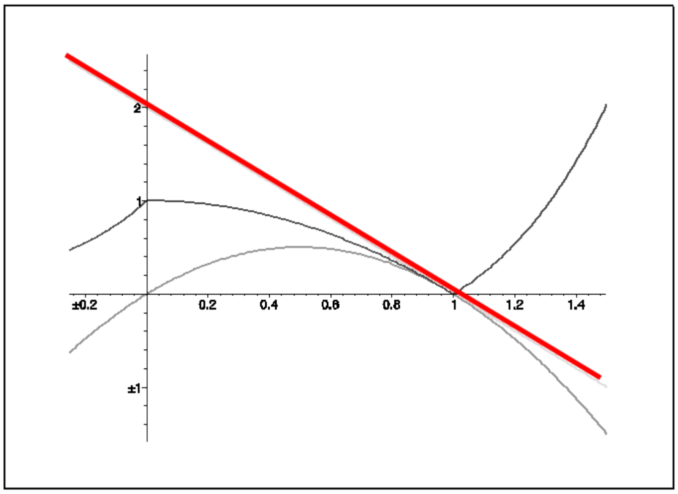

Fig. 2 A function and its smooth minorant and a viscosity subdifferential (in red).

2.1 Achievements and limitations

Variational principles meshed with viscosity subdifferentials provide a fine first-order theory. Sadly, $\partial^a f(x)$ is inapplicable outside of *Asplund space* (such as reflexive space or spaces with separable duals) and extensions using $\partial_H f$ are limited and technically complicated. Correspondingly, the coderivative is very beautiful theoretically but is hard to compute even for 'nice' functions. Moreover the compactness restrictions (e.g., *sequential normal compactness* as described in [24]) are fundamental, not technical. Better results rely on restricting classes of functions (and spaces) such as considering, *prox-regular* [28], *lower* C^2 [28], or *essentially smooth* functions [13].

Moreover, the limits of a prescriptive approach are highlighted by the fact that one can prove results showing that in all (separable) Banach spaces X a *generic* non-expansive function has no information in its generalized derivative:

$$\partial^a f(x) = \partial^o f(x) \equiv B_{X^*}$$

for all points $x \in X$ [13, 10]. Similarly, one can show that nonconvex equilibrium results will frequently contain little or no non-trivial information [13].

3 Higher-Order Theory of Variational Analysis

Recall that for closed proper *convex functions* the *difference quotient* of f is given by

$$\Delta_t f(x) : h \mapsto \frac{f(x+th) - f(x)}{t};$$

and the *second-order difference quotient* of f by

$$\Delta_t^2 f(x) : h \mapsto \frac{f(x+th) - f(x) - t\langle \nabla f(x), h\rangle}{\frac{1}{2}t^2}.$$

Analogously let

$$\Delta_t[\partial f](x) : h \mapsto \frac{\partial f(x+th) - \nabla f(x)}{t}.$$

For any $t > 0$, $\Delta_t f(x)$ is closed, proper, convex and nonnegative [28, 11]. Quite beautifully, as Rockafellar [28, 11] discovered,

$$\partial \left[\frac{1}{2}\Delta_t^2 f(x)\right] = \Delta_t[\partial f](x).$$

Hence, we reconnect the two most natural ways of building a second-order convex approximation.

This relates to a wonderful result [1, 11]:

Theorem 1 (Alexandrov (1939)). *In Euclidean space a real-valued continuous convex function admits a second-order Taylor expansion at almost all points (with respect to Lebesgue measure).*

My favourite proof is a specialization of Mignot's 1976 extension of Alexandrov's theorem for monotone operators [28, 11]. The theorem relies on many happy coincidences in Euclidean space. This convex result is quite subtle and so the paucity of definitive non-convex results is no surprise.

3.1 The state of higher-order theory

Various lovely patterns and fine theorems are available in Euclidean space [28, 24, 11] but no definitive corpus of results exists, nor even canonical definitions, outside of the convex case. There is interesting work by Jeyakumar-Luc [21], by Dutta, and others, much of which is surveyed in [18].

Starting with Clarke, many have noted that

$$\partial^2 f(x) := \partial \nabla_G f(x)$$

is a fine object when the function f is Lipschitz smooth in a separable Banach space—so that the Banach space version of Rademacher's Theorem [11] applies.

More interesting are the quite fundamental results by Ioffe and Penot [20] on limiting 2-subjets and 2-coderivatives in Euclidean space, with a more refined calculus of 'efficient' sub-Hessians given by Eberhard and Wenczel [19]. Ioffe and Penot [20] exploit Alexandrov-like theory, again starting with the subtle analysis in [16], to carefully study a *subjet* of a reasonable function f at x, the subjet $\partial_-^2 f(x)$ being defined as the collection of second-order expansions of all C^2 local minorants g with $g(x) = f(x)$. The (non-empty) *limiting 2-subjet* is then defined by

$$\overline{\partial}^2 f(x) := \limsup_{y \to_f x} \partial_-^2 f(x).$$

Various distinguished subsets and limits are also considered in their paper. They provide a calculus, based on a sum rule for limiting 2-subjets (that holds for all lower-C^2 functions and so for all continuous convex functions) making note of both the similarities and differences from the first-order theory. As noted, interesting refinements have been given by Eberhard and Wenczel in [19].

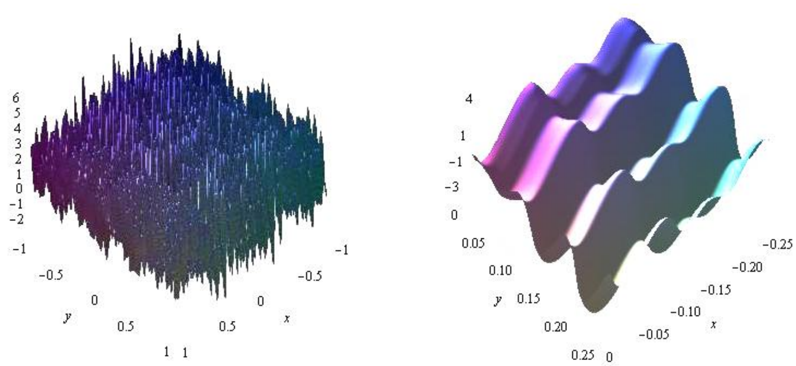

Fig. 3 Nick Trefethen's digit-challenge function (2).

There is little 'deep' work in infinite dimensions, that is, when reasonably obvious extensions fail even in Hilbert space. Outside separable Hilbert space general positive results are not to be expected [11]. So it seems clear to me that research should focus on structured classes of functions for which more can be obtained; such as *integral functionals* as in Moussaoui-Seeger [26], *semi-smooth* and *prox-regular functions* [8], or *composite convex functions* [28].

4 Some Reflections on Applications of Variational Analysis

The tools of variational analysis are now an established part of pure non-linear and functional analysis. This is a major accomplishment.

There are also more concrete successes:

- There is a convergence theory for “pattern search” derivative-free optimization algorithms (see [23] for an up to date accounting of such methods) based on the Clarke subdifferential.
- Eigenvalue and singular value optimization theory has been beautifully developed [8], thanks largely to Adrian Lewis. There is a quite delicate second-order theory due to Lewis and Sendov [22]. There are even some results for Hilbert-Schmidt operators [13, 11].
- We can also handle a wide variety of differential inclusions and optimal control problems well [25].
- There is a fine approximate Maximum Principle and a good accounting of Hamilton-Jacobi equations [24, 25, 13].
- Non-convex mathematical economics and *Mathematical Programs with Equilibrium Constraints* (MPECS) are much better understood than before [24, 25].
- Exact penalty and universal barrier methods are well developed, especially in finite dimensions [11].
- Counting convex optimization—as we certainly should—we have many more successes [14].

That said, there has been only limited numerical success even in the convex case—excluding somewhat spectral optimization, semidefinite programming code, and bundle methods.

For example, consider the following two-variable well-structured very smooth function taken from [4] in which only the first two rather innocuous terms couple the variables

$$
\begin{aligned}
(x,y) \mapsto &+ (x^2+y^2)/4 - \sin(10(x+y)) + \exp(\sin(50x)) \\
&+ \sin(\sin(80y)) + \sin(70\sin x) + \sin(60e^y). \qquad (2)
\end{aligned}
$$

This function is quite hard to minimize. Actually, the global minimum occurs at $(x^*,y^*) \approx (-0.024627\ldots, 0.211789\ldots)$ with minimal value of $\approx -3.30687\ldots$.

The pictures in Figure 3, plotted using 10^6 grid points on $[0,1]\times[0,1]$ and also—after ‘zooming in’—on $[-00.25,0]\times[0,0.25]$, shows that we really can not robustly distinguish the function from a nonsmooth function. Hence, it makes little sense to look at practicable nonsmooth algorithms without specifying a realistic subclass of functions on which they should operate.

Perhaps we should look more towards projects like Robert Vanderbei’s SDP/Convex package *LOQO/LOCO*[3] and Janos Pinter’s Global Optimization *LGO*[4] package,

[3] http://www.princeton.edu/∼rvdb/

[4] http://myweb.dal.ca/jdpinter/index.html

while working with composite convex functions and smoothing techniques, and adopting the “disciplined convex programming”[5] approach advocated by Steve Boyd.

5 Open Questions and Concluding Remarks

I pose six problems below which should either have variational solutions or instructive counter-examples. Details can be found in the specified references.

5.1 Alexandrov Theorem in Infinite Dimensions

For me, the most intriguing open question about convex functions is:

> *Does every continuous convex function on separable Hilbert space admit a second order Gâteaux expansion at at least one point (or perhaps on a dense set of points)?* ([9, 7, 13])

This fails in non-separable Hilbert space and in every separable $\ell_p(\mathbf{N}), 1 \leq p < \infty, p \neq 2$. It also fails in the Fréchet sense even in $\ell_2(\mathbf{N})$.

The following example from [9] provides a continuous convex function d on any nonseparable Hilbert space which is nowhere second-order differentiable: *Let A be uncountable and let C the positive cone of $\ell_2(A)$. Denote by d the distance function to C and let $P := \nabla d$. Then d is nowhere second-order differentiable and P is nowhere differentiable (in the sense of Mignot [28]).*

Proof. Clearly, $P(a) = a^+$ for all $a \in \ell_2(A)$, where $a^+ = (a^+_\alpha)_{\alpha\in A}$ and $a^+_\alpha = \max\{0, a_\alpha\}$. Pick $x \in \ell_2(A)$ and $\alpha \in A$, then P is differentiable in the direction e_α if and only if $x_\alpha \neq 0$. Here e_α stands for an element of the canonical basis. Since each $x \in \ell_2(A)$ has only countably many nonzero coordinates, d is nowhere second-order differentiable. Likewise the maximal monotone operator P is nowhere differentiable. ♢

So I suggest to look for a counter-example. I might add that, despite the wonderful results of Preiss [11] and others on differentiability of Lipschitz functions, it is still also unknown whether two arbitrary real-valued Lipschitz functions on a separable Hilbert space must share a point of Fréchet differentiability.

5.2 Subjets in Hilbert space

I turn to a question about nonsmooth second-order behaviour:

---

[5] http://www.stanford.edu/∼boyd/cvx/

> *Are there sizeable classes of functions for which* subjets *or other useful second order expansions can be built in separable Hilbert space?* ([20, 19, 11])

I have no precise idea what "useful" means and even in convex case this is a tough request; if one could handle the convex case then one might be able to use Lasry-Lions regularization or other such tools more generally. A potentially tractable case is that of continuous integral functionals for which positive Alexandrov like results are known in the convex case [9].

5.3 Chebyshev Sets

The *Chebyshev problem* as posed by Klee (1961) asks:

> *Given a non-empty set C in a Hilbert space H such that every point in H has a unique nearest (also called proximal) point in C must C convex?* ([5, 8, 11])

Such sets are called *Chebyshev sets*. Clearly convex closed sets in Hilbert space are Chebyshev sets. The answer is 'yes' in finite dimensions. This is the *Motzkin-Bunt theorem* of which four proofs are given in Euclidean space in [8] and [5]. In [5, 11] a history of the problem, which fails in some incomplete normed spaces, is given.

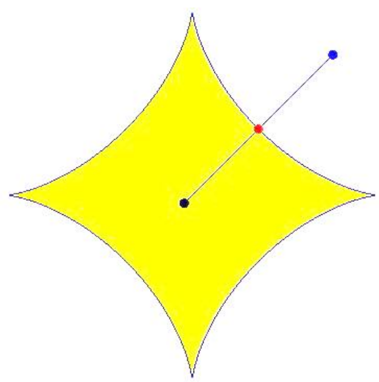

Fig. 4 A proximal point on the boundary of the $(2/3)$-ball.

5.4 Proximality

The most striking open question I know regarding proximality is:

(a) *Let C be a closed subset of a Hilbert space H. Fix an arbitrary equivalent renorming of H. Must some (many) points in H have a nearest point in C in the given renorming?*

(b) *More generally, is it possible that in every reflexive Banach space, the* proximal points *on the boundary of C (see Figure 4) are dense in the boundary of C?* ([6, 13])

The answer is 'yes' in if the set is bounded or the norm is *Kadec-Klee* and hence if the space is finite dimensional or if it is locally uniformly rotund [6, 13, 11].

So any counter-example must be a wild set in a weird equivalent norm on H.

5.5 Legendre Functions in Reflexive Space

Recall that a convex function is of *Legendre-type* if it is both essentially smooth and essentially strictly convex. In the reflexive setting, the property is preserved under Fenchel conjugacy.

Find a generalization of the notion of a Legendre function for convex functions on a reflexive space that applies when the functions have no points of continuity such as is the case of the (negative) Shannon entropy. ([2, 11])

When f has a point of continuity, a quite useful theory is available but it does not apply to entropy functions like $x \mapsto \int_0^1 x(t)\log x(t)\,\mu(dt)$ or $x \mapsto -\int_0^1 \log x(t)\,\mu(dt)$, whose domains are subsets of the non-negative cone when viewed as operators on $L_2(T,\mu)$. More properly to cover these two examples, the theory should really apply to integral functionals on non-reflexive spaces such as $L_1(T,\mu)$.

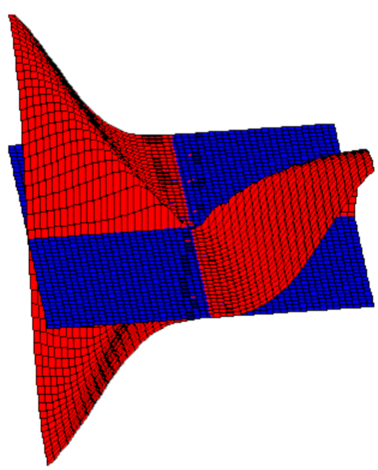

Fig. 5 A function and non-viscosity *subderivative* of 0.

5.6 Viscosity Subdifferentials in Hilbert Space

A more technical but fundamental question is:

Is there a real-valued locally Lipschitz function f on $\ell_2(\mathbf{N})$ such that properly

$$\partial_H^v f(x) \subset \partial_H f(x)$$

for some $x \in \ell_2(\mathbf{N})$? ([12, 13])

As shown in Figure 5, the following continuous but non-Lipschitz function

$$(x,y) \mapsto \frac{xy^3}{x^2+y^4}$$

with value zero at the origin has $0 \in \partial_H f(0)$ but $0 \notin \partial_H^v f(0)$ [12, 13].

For a Lipschitz function in Euclidean space the answer is 'no' since $\partial_F f = \partial_H f$ in this setting. And as we have noted $\partial_F f = \partial_F^v f$ in reflexive space. A counter-example would be very instructive, while a positive result would allow for many results to be extended from the Fréchet case to the Gateaux case: as $\partial_G f = \partial_H f$ for all locally Lipschitz f.

5.7 Final Comments

My view is that rather than looking for general prescriptive results based on universal constructions, we would do better to spend some real effort, or 'brain grease' as Einstein called it,[6] on descriptive results for problems such as the six above. Counter-examples or complete positive solutions would be spectacular, but even somewhat improving best current results will require sharpening the tools of variational analysis in interesting ways. That would also provide great advertising for our field.

References

1. Alexandrov A.D. (1939) Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. Leningrad State Univ. Ann. Math Ser. 6:3–35.
2. Bauschke HH, Borwein JM, and Combettes PL (2001) Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. Communications in Contemporary Mathematics 3:615–647
3. Borwein JM (1986) Stability and regular points of inequality systems. Journal of Optimization Theory and Applications 48:9–52

[6] "On quantum theory, I use up more brain grease than on relativity." (Albert Einstein in a letter to Otto Stern in 1911).

4. Borwein JM (2005) The SIAM 100 Digits Challenge. Extended review in Math. Intelligencer 27:40–48
5. Borwein JM (2007) Proximality and Chebyshev sets. Optimization Letters 1:21–32
6. Borwein JM and Fitzpatrick S (1989) Existence of nearest points in Banach spaces. Canadian Journal of Mathematics 61:702–720.
7. Borwein JM and Ioffe A (1996) Proximal analysis in smooth spaces. Set-Valued Analysis 4:1–24
8. Borwein JM and Lewis AS (2005) Convex Analysis and Nonlinear Optimization. Theory and Examples. Canadian Math Society-Springer Books, 2nd Ed.
9. Borwein JM and Noll D (1994) Second order differentiability of convex functions in Banach spaces. Trans. Amer. Math. Soc. 342:43–82
10. Borwein JM and Sciffer S (2010) An explicit non-expansive function whose subdifferential is the entire dual ball. Contemporary Mathematics. In press
11. Borwein JM and Vanderwerff JM (2010) Convex Functions: Constructions, Characterizations and Counterexamples. Cambridge University Press
12. Borwein JM and Qiji Zhu (1996) Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity. SIAM J. Optimization 34:1568–1591
13. Borwein JM and Qiji Zhu (2005) Techniques of Variational Analysis. Canadian Math Society-Springer Books .
14. Boyd S and Vandenberghe L (2004) Convex Optimization. Cambridge University Press
15. Clarke FH (1983) Optimization and nonsmooth analysis. Canadian Math Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc, New York
16. Crandall M, Ishii H, and Lions P-L (1992) User's guide to viscosity solutions of second-order partial differential equations. Bull. Amer. Math. Soc. 27:1–67
17. Dontchev AL and Rockafellar RT (2009) Implicit Functions and Solution Mappings: A View from Variational Analysis. Springer Monographs in Mathematics, New York
18. Dutta J (2005) Generalized derivatives and nonsmooth optimization, a finite dimensional tour. With discussions and a rejoinder by the author. TOP 11:85–314
19. Eberhard AC and Wenczel R (2007) On the calculus of limiting subhessians. Set-Valued Analysis 15:377–424
20. Ioffe AD, and Penot J-P (1997) Limiting subhessians, limiting subjets and their calculus. Trans. Amer. Math. Soc. 349:789–807
21. Jeyakumar V and Luc DT (2008) Nonsmooth Vector Functions and Continuous Optimization. Springer Optimization and Its Applications, vol 10
22. Lewis AS and Sendov HS (2005) Nonsmooth analysis of singular values, I & II. Set-Valued Analysis 13:213–264.
23. Macklem M (2009). Parallel Continuous Optimization. Doctoral Thesis, Dalhousie University
24. Mordukhovich BS (2006) Variational Analysis & Generalized Differentiation I. Basic Theory. Springer-Verlag Grundlehren, vol 330.
25. Mordukhovich BS (2006) Variational Analysis & Generalized Differentiation II. Applications. Springer-Verlag Grundlehren, vol 331
26. Moussaoui M and Seeger A.(1999) Second-order subgradients of convex integral functionals. Trans. Amer. Math. Soc. 351:3687–3711
27. Pshenichnii B (1971) Necessary Conditions for an Extremum. Marcel Dekker, New York
28. Rockafellar RT and Wets RJ-B (1998) Variational Analysis. Springer-Verlag Grundlehren, vol 317