FUTURE CHALLENGES FOR VARIATIONAL ANALYSIS

JONATHAN M. BORWEIN

Abstract. I will also discuss open problems and current challenges for the subject Boris Mordukhovich has played a key role in the development of modern *Variational Analysis (VA) and its Applications.* 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? Summary: the first order theory is rather impressive, as are some applications. The second order theory is by comparison somewhat underdeveloped and wanting.

"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)[1]

1. Preliminaries

I intend to 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* both inside and outside Mathematics, mentioning both successes and limitations or failures. Each topic leads to open questions even in the convex case (CA). Some are

Date: May 23, 2009.

University of Newcastle, Australia.
Email: jonathan.borwein@newcastle.edu.au. 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.

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

technical and specialized, others are some broader and more general. In nearly every case Boris Mordukhovich has made prominent or seminal contributions; many elaborated in [20] and [21].

To work fruitfully in VA it is really important to understand both CA and *smooth analysis* (SA); they are the motivating foundations and often provide the key technical tools. For example, Figure 1 shows how an essentially strictly convex [6, 9] 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 studied in [6, 20, 11, 23]. In this note our terminology is for the most-part consistent with those references and since we wish to discuss patterns not proofs we will not worry too much about exact conditions. That said, f will be a proper and lower semicontinuous extended-real valued function on a Banach space X.

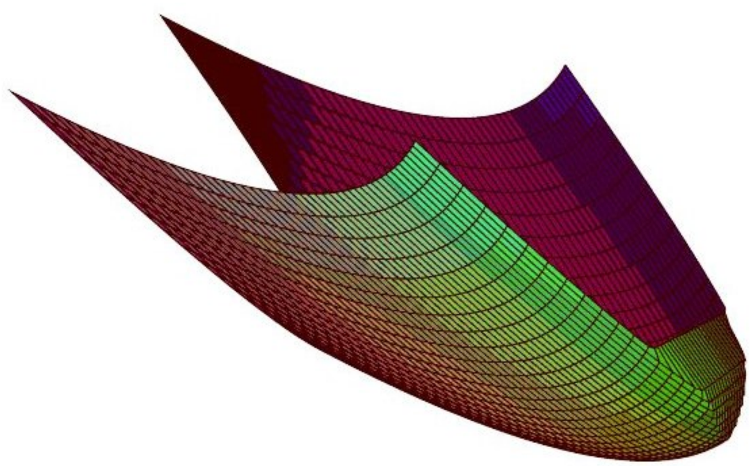

FIGURE 1. A function that is essentially strictly but not strictly convex with nonconvex subgradient domain.

Let us first recall the two main starting points:

1.1. **A Descriptive Approach.** Pshenichnii (1968) considered 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.** Clarke (1972) 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. We define $\partial^o f(x) := \partial_2 f^0(x;0)$.

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

2. First-Order Theory of Variational Analysis

The key players are as listed below. We start with:

- **The (Fréchet) Subgradient**

  $$\partial_F f(x)$$

  which denotes a one-sided lower Fréchet subgradient (i.e., uniform on bounded sets) and can 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 $\partial_G f(x), \partial_H f(x)$, and $\partial_{WH} f(x)$ respectively.

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

  $$\begin{aligned} \partial_F^v f(x) \;\; &:= \;\; \{\phi\colon \phi = \nabla_F g(x), \\ &\qquad f(y)-g(y) \geq f(x)-g(x) \text{ for } y \text{ near } x\} \end{aligned}$$

  as illustrated in Figure 2. In nice spaces, say those with Fréchet-smooth renorms, these coincide. In this case we have access to a good generalization of the sum rule from convex calculus [9]:
- **(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 some x_1, x_2 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 [11, 20, 21, 23].

For now we let ∂f denote any of a number of subgradients and can define a workable normal cone.

- **Normal cone**

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

Key to establishing the fuzzy sum rule and its many equivalences [11, 20] are:

- **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 of:

- **Limiting Subdifferentials**

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

and of:

- **Coderivative of a Multifunctions**

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

The fuzzy sum rule also leads to fine results about:

- **Metric regularity** Indeed, we can provide very functional conditions on a multifunction Ω, see [10, 11, 20], so that locally around $y_0 \in \Omega(x_0)$ one has

$$K d(\Omega(x), y) \geq d(x, \Omega^{-1}(y)). \tag{2.1}$$

Estimate 2.1 allows one to show many things easily. For example, to produce C^k-implicit function theorems under second-order sufficiency conditions.

A a forthcoming book by Dontchev-Rockafellar gives a comprehensive treatment of implicit functions for Euclidean multifunctions. Estimate 2.1 is really useful in the very concrete 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 [11].

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) and extensions using $\partial_H f$ are limited and technically complicated. Correspondingly the coderivative is very beautiful theoretically but is hard

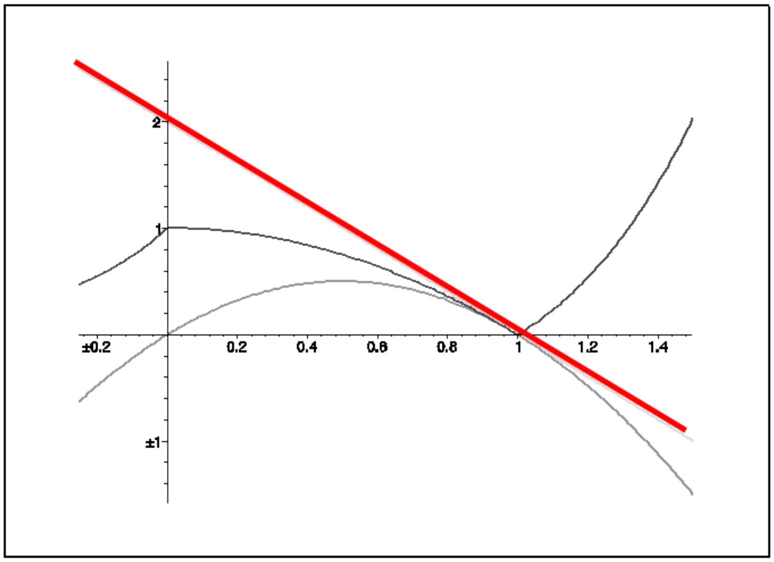

FIGURE 2. A function and its smooth minorant and a viscosity subdifferential (in red).

to compute even for ‘nice’ functions. Moreover the compactness restrictions (e.g., *sequential normal compactness* as described in [20]) are fundamental not technical.

Better results rely on restricting classes of functions (and spaces) such as considering, *prox-regular*, *lower* C^2, or *essentially smooth* functions [11].

Moreover, the limits of a prescriptive approach are highlighted by the fact that one can prove results like generically non-expansive functions have

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

in all (separable) Banach spaces [11, 8]. Similarly, one can show that nonconvex equilibrium results will frequently contain little or no information [11].

3. Higher-Order Theory of Variational Analysis

Recall that for closed *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

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

(See [23, 9].) For any $t > 0$,

$$\Delta_t f(x)$$

is closed, proper, convex and nonnegative.

Quite beautifully

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

This relates to a wonderful result:

Theorem 3.1 (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 [23, 9]. 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.** Some lovely patterns and fine theorems are available in Euclidean space [23, 20, 9] but no definitive corpus of results exists nor even canonical definitions outside of the convex case. There is interesting work by Jeyakumar-Luc [17], by Duta, and others much of which is surveyed in [14].

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 [9] applies.

More interesting are the fundamental results by Ioffe-Penot [16] on limiting 2-subjets and 2-coderivatives, with a more refined calculus of 'efficient' sub-Hessians given in Eberhard-Wenczel [15]. Ioffe and Penot [16] exploit Alexandrov-like theory, again starting with the subtle analysis in [13], 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. 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. Interesting refinements are given by Eberhard and Wenczel in [15].

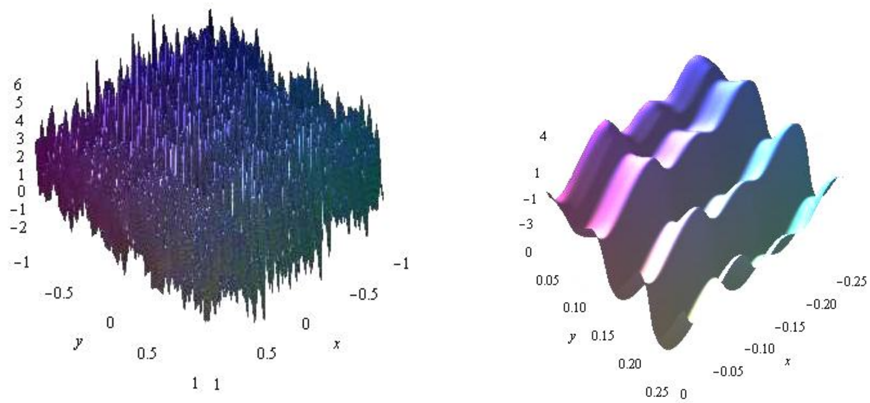

Figure 3. Nick Trefethen's digit-challenge function (4.1).

There is little 'deep' work in infinite dimensions, that is, if obvious extensions fail even in Hilbert space. Outside separable Hilbert space general positive results are not to be expected [9]. So it seems clear that research should focus on structured classes such as *integral functionals* as in Moussaoui-Seeger [22], *semi-smooth functions* [6], or *composite convex functions.*

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.

There are also concrete successes:

- There is a convergence theory for "pattern search" derivative-free optimization algorithms (see [19] for an up to date accounting of such methods) based on the Clarke subdifferential.

- Eigenvalue and singular value optimization theory has been beautifully developed [6], thanks largely to Adrian Lewis. There is a quite delicate second-order theory due to Lewis and Sendov [18]. There are even some results for Hilbert-Schmidt operators [11, 9].
- We can also handle differential inclusions and optimal control problems well [21].
- There is a fine approximate Maximum Principle and a good accounting of Hamilton-Jacobi equations [20, 21, 11].
- Non-convex mathematical economics and *Mathematical Programs with Equilibrium Constraints* (MPECS) are much better understood than before [20, 21].
- Exact penalty and universal barrier methods are well developed, especially in finite dimensions [9].
- Counting convex optimization—as we certainly should—we have many more successes [12].

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

For instance, the following two-variable well-structured very smooth function (only the first two innocuous terms couple the variables) is very hard to minimize [2]:

$$\begin{aligned}(x, y) \mapsto \quad &+ \quad (x^2 + y^2)/4 - \sin(10(x + y)) + \exp(\sin(50x)) \\ &+ \quad \sin(\sin(80y)) + \sin(70 \sin x) + \sin(60e^y). \end{aligned} \tag{4.1}$$

The 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 distinguish it from a nonsmooth function and, hence, it makes little sense to look at practicable nonsmooth algorithms without specifying a realistic subclass of functions.

Perhaps we should look more towards Robert Vanderbei's SDP/Convex *LOQO/LOCO*[2] and Janos Pinter's Global Optimization *LGO*[3] package, working with composite convex functions and smoothing techniques, and adopting the "disciplined convex programming"[4] approach advocated by Steve Boyd.

[2] `http://www.princeton.edu/~rvdb/`
[3] `http://myweb.dal.ca/jdpinter/index.html`
[4] http://www.stanford.edu/~boyd/cvx/

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. ([5, 11])

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

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

5.2. Subjets in Hilbert space. ([16, 15, 9])

> *Are there sizeable classes of functions for which* subjets *or other useful second order expansions can be built in separable Hilbert space?*

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.

5.3. Chebyshev Sets. ([3, 6, 9])

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?*

The answer is 'yes' in finite dimensions. This is the *Motzkin-Bunt theorem* of which four proofs are given in Euclidean space in [6] and [3]. In [3, 9] a history of the problem, which fails in some incomplete normed spaces, is given.

5.4. Proximality. ([4, 11])

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

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 [4, 11, 9].

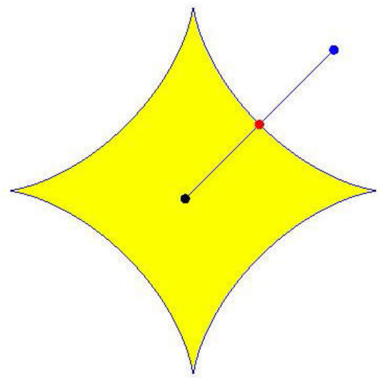

FIGURE 4. A proximal point on the boundary of the $\frac{2}{3}$-ball.

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

5.5. **Legendre Functions in Reflexive Space.** ([1, 9])

Recall that a convex function is of *Legendre-type* if it both essentially smooth and essentially strictly convex. 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 have no points of continuity such (negative) Shannon entropy.*

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. More properly to cover these two examples, the theory should really apply to integral functionals on non-reflexive spaces such as $L_1(T, \mu)$.

5.6. **Viscosity Subdifferentials in Hilbert Space.** ([10, 11])

> *Is there a real-valued locally Lipschitz function f on $\ell_2(\mathbf{N})$ such that*
> $$\partial_H^v f(x) \subsetneqq \partial_H f(x)$$
> *for some $x \in H$?*

As shown in Figure 5,

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

has $0 \in \partial_H f(0)$ but $0 \notin \partial_H^v f(0)$ [10, 11].

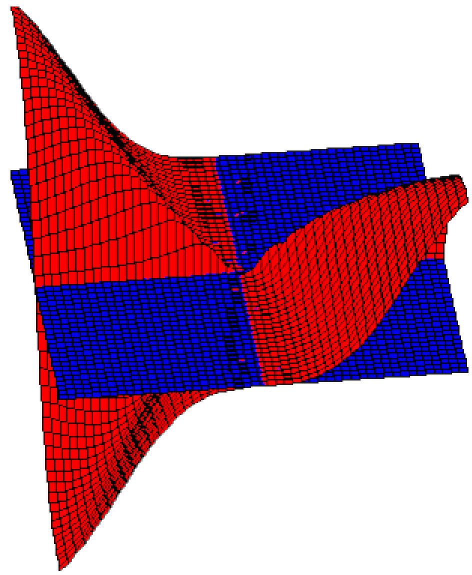

FIGURE 5. A function and non-viscosity *subderivative* of 0.

For a Lipschitz function in Euclidean space the answer is 'no' since $\partial_F f = \partial_H f$ in this setting. 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: $\partial_G f = \partial_H f$ 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,[5] on descriptive results 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] H.H. Bauschke, J.M. Borwein, and P.L. Combettes, "Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces," *Communications in Contemporary Mathematics*, **3** (2001), 615–647.

[2] J.M. Borwein, "The SIAM 100 Digits Challenge", Extended review in the *Math. Intelligencer*, **27** (2005) , 40–48. [D-drive preprint 285]

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

[3] J.M. Borwein, “Proximality and Chebyshev sets,” *Optimization Letters*, **1** (2007), 21–32.
[4] J.M. Borwein and S. Fitzpatrick, “Existence of nearest points in Banach spaces,” *Canadian Journal of Mathematics*, **61** (1989), 702–720.
[5] J.M. Borwein and A. Ioffe, “Proximal analysis in smooth spaces,” *Set-Valued Analysis,* **4** (1996), 1–24.
[6] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples,* Can. Math Society-Springer Books, 2nd Ed., 2005.
[7] J.M. Borwein and D. Noll, “Second order differentiability of convex functions in Banach spaces,” *Trans. Amer. Math. Soc.*, **342** (1994), 43-82.
[8] J.M. Borwein and S. Sciffer, “An explicit non-expansive function whose subdifferential is the entire dual ball,” *Contemporary Mathematics*, (2009). In press.
[9] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples,* Cambridge University Press, 2009.
[10] J.M. Borwein and Qiji Zhu, “Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity,” *SIAM J. Optimization*, **34** (1996), 1568–1591.
[11] J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis,* Can. Math Society-Springer Books, 2005.
[12] S. Boyd. and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, 2004.
[13] M. Crandall, H. Ishii, H., and P.-L. Lions, “User’s guide to viscosity solutions of second-order partial differential equations,” *Bull. Amer. Math. Soc.*, **27** (1992), 1–67.
[14] J. Dutta, “Generalized derivatives and nonsmooth optimization, a finite dimensional tour. With discussions and a rejoinder by the author.” *TOP* **13** (2005), 185–314.
[15] A.C. Eberhard and R. Wenczel, “On the calculus of limiting subhessians,” *Set-Valued Anal.*, **15**, (2007) 377–424.
[16] A.D. Ioffe, J.-P. Penot, “Limiting subhessians, limiting subjets and their calculus,” *Trans. Amer. Math. Soc*, **349**, (1997) 789–807.
[17] V. Jeyakumar, and D.T. Luc, *Nonsmooth Vector Functions and Continuous Optimization*. Springer Optimization and Its Applications, **10**, 2008.
[18] A.S. Lewis, and H.S. Sendov, “Nonsmooth analysis of singular values”, I & II. *Set-Valued Anal.* **13**, (2005) 213–264.
[19] M. Macklem, *Parallel Continuous Optimization*, Doctoral Thesis, Dalhousie University, 2009.

[20] B.S. Mordukhovich, *Variational Analysis & Generalized Differentiation I. Basic Theory,* Springer-Verlag Grundlehren, **330**, 2006.
[21] B.S. Mordukhovich, *Variational Analysis & Generalized Differentiation II. Applications,* Springer-Verlag Grundlehren, **331**, 2006.
[22] M. Moussaoui and A. Seeger, "Second-order subgradients of convex integral functionals," *Trans. Amer. Math. Soc.* **351** (1999), 3687–3711.
[23] R.T. Rockafellar and RJ-B. Wets, *Variational Analysis,* Springer-Verlag Grundlehren, **317**, 2006.