A One Perturbation Variational Principle and Applications

Jonathan M. Borwein, FRSC

Research Chair in IT
Dalhousie University

Halifax, Nova Scotia, Canada

Dalhousie Distributed Research Institute and Virtual Environment

Large Scale Nonlinear and Semidefinite Programming

Waterloo, 14th of May 2004

www.dal.cs.ca/~jborwein

www.cecm.sfu.ca/preprints.html

Revised: May 12th, 2004

A Large Scale Computation

Let

$$\zeta_N(\pm) := \sum_{n_1>n_2>\cdots>n_{2N}} \frac{(\pm 1)^{n_1}}{n_1^2 n_2} \cdots \frac{(\pm 1)^{n_{2N-1}}}{n_{2N-1}^2 n_{2N}}.$$

Conjecture. $\boxed{8^N \zeta_N(-) = \zeta_N(+)}$ for all N.

That is,

$$8 \sum_{n>m>0} \frac{(-1)^n}{n^2 m} = \sum_{n>m>0} \frac{1}{n^2 m}$$

$$64 \sum_{n>m>p>q>0} \frac{(-1)^n}{n^2 m} \frac{(-1)^p}{p^2 q} = \sum_{n>m>p>q>0} \frac{1}{n^2 m} \frac{1}{p^2 q}$$

$$\ddots$$

$$\ddots$$

- Found using **Lattice basis reduction** in 1995.

- Such sums are important in number theory, knot theory, QFT and elsewhere.

- They converge very slowly indeed—needed an 'FFT'.

- Checked to 1,000 digits for $N < 85$ (40 HP hrs) and $N = \mathbf{163}$ (10 hrs).

- Nothing to do with this conference.

Abstract

We study a variational principle with a *common* perturbation function φ for *all* proper lower semicontinuous extended real-valued functions f on a metric space X.

- Necessary and sufficient conditions are given for the perturbed $f+\varphi$ to attain its minimum.

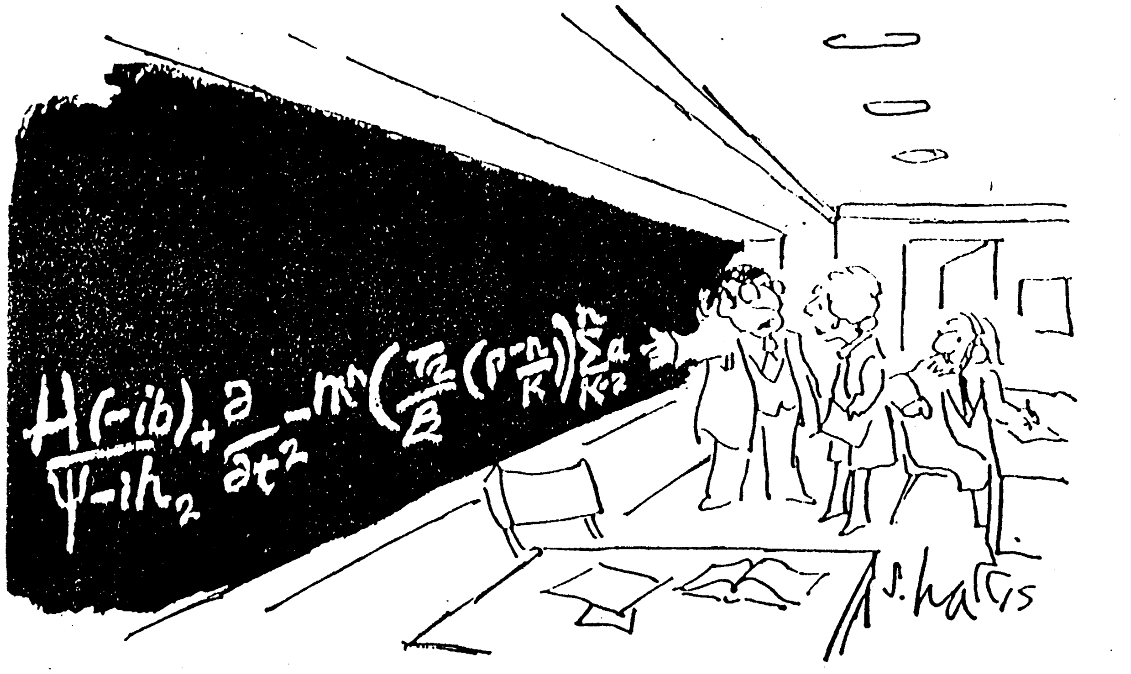
- For separable Banach space we may use a perturbation function that is also convex and *Hadamard-like* differentiable.

- We give three applications to differentiability of convex functions on separable and more general Banach spaces.

- We pose various open questions: on viscosity, genericity and stability.

Credits

- This is NATO sponsored* joint work: J. Borwein, L. Cheng, M. Fabian and J. Revalski, "A one perturbation variational principle with applications," *Set-Valued Analysis*, **12** (2004), 49–60. [CECM Preprint 2003:205]†



"But this *is* the simplified version for the general public."

*Also funded by NSERC, CFI, CRC, etc. !!!

†Available from http://www.cecm.sfu.ca/preprints.

Especially as we go Retro-Digital

"I'll be glad if I have succeeded in impressing the idea that it is not only pleasant to read at times the works of the old mathematical authors, but this may occasionally be of use for the actual advancement of science."

Constantin Carathéodory (MAA, 1936).

Outline

1. One Theorem

2. Two Applications

3. Three Questions

4. Four References

1. One Theorem

Theorem 1 *Let X be a Hausdorff topological space which admits a proper lsc function*

$$\varphi : X \to \mathbb{R} \cup \{+\infty\}$$

whose level sets are all compact. Then for any proper lsc and bounded from below function $f : X \to \mathbb{R} \cup \{+\infty\}$ the function $f + \varphi$ attains its minimum. In particular, if $\operatorname{dom}\varphi$ *is relatively compact, the conclusion is true for any proper lsc function f.*

Key application. In separable Banach space, a *nice* convex choice is:

$$\varphi(x) := \begin{cases} \tan\left(\|S^{-1}x\|_H^2\right), & \text{if } \|S^{-1}x\|_H^2 < \frac{\pi}{2}, \\ +\infty, & \text{otherwise.} \end{cases}$$

for an appropriate compact, linear and injective mapping $S\colon H \to X$ ($H := \ell_2$). Also φ is *almost Hadamard* smooth:

$$\lim_{t \searrow 0} \sup_{h \in \operatorname{dom}\varphi} \frac{\varphi(x+th) + \varphi(x-th) - 2\varphi(x)}{t} = 0,$$

Remark 2 If $(X, \|\cdot\|)$ is *normed* and φ is *convex*, the result above holds for every proper lsc convex f, provided only that the level sets of φ are *weakly* compact, or that $\operatorname{dom}\varphi$ is.

Remark 3 In a normed space $(X, \|\cdot\|)$, by allowing translations of φ, we get a *localization* of the minimum of the perturbation (as in Bishop-Phelps, Ekeland, Borwein-Preiss [B-P], etc.).

With the same proof:

> *Suppose X admits a function φ as above. For any proper lsc (bounded below) function $f : X \to \mathbb{R} \cup \{+\infty\}$, for any $\bar{x} \in \operatorname{dom} f$ and each $\lambda > 0$, the function*
>
> $$f + \varphi((\cdot - \bar{x})/\mu)$$
>
> *(for some $\mu > 0$), attains its minimum at a u with $\|u - \bar{x}\| \le \lambda$.*

- Observe that in this case, formally, the perturbation function is *now* varying.

- The main requirement of Theorem 1 is also necessary.

Namely, we have:

Theorem 4 *Let* $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ *be a proper function on a metric space* X *with the property that* for every bounded continuous function $f : X \to \mathbb{R}$, the function $f + \varphi$ attains its minimum.

Then φ *is (i) a lower semicontinuous function, (ii) bounded from below, (iii) whose level sets are all compact.*

- This proof is significantly more subtle.

- The **punch-line** so far—a little geometry goes a long way.

2. Two Applications

1. We recover *Mazur's theorem* (and various *bornological* extensions).

Theorem 5 *Suppose X is a separable Banach space. Then every continuous convex $f : X \to \mathbb{R}$ is Gâteaux differentiable at the points of a generic (that is a dense G_δ) subset of X.*

Proof. First, we show f is Gâteaux differentiable at the points of a dense subset of X. After translation, it suffices to show every non empty open set Ω of X with $0 \in \Omega$, contains a point at which f is Gâteaux differentiable.

(**Step 1.**) Fix such an Ω and let φ be the function given by Theorem 1. We may suppose $\operatorname{dom}\varphi \subset \Omega$. Then there is an $x \in \operatorname{dom}\varphi \subset \Omega$ at which $-f+\varphi$ attains its minimum.

In particular, for any $h \in \operatorname{dom}\varphi$ and $t > 0$

$$-f(x \pm th) + \varphi(x \pm th) \geq -f(x) + \varphi(x).$$

Using this and convexity of f we obtain

$$\begin{aligned} 0 &\leq f(x+th) + f(x-th) - 2f(x) \\ &\leq \varphi(x+th) + \varphi(x-th) - 2\varphi(x) \end{aligned}$$

which together with the differentiability property (3.1) of φ shows that

$$\lim_{t \searrow 0} \frac{f(x+th) + f(x-th) - 2f(x)}{t} = 0,$$

for every $h \in \operatorname{dom}\varphi$. Since f is locally Lipschitz and $\operatorname{dom}\varphi$ is linearly dense, in fact, the latter limit is 0 for any $h \in X$.

(**Step 2.**) Finally, the fact that f is convex yields its (linear) Gâteaux differentiability at x.

To show the points of Gâteaux differentiability of f is *exactly* a G_δ-subset of X, let us observe that (3.1) yields a stronger conclusion: that X possesses a dense subset in which every x obeys the following stronger condition that as $t \searrow 0$,

$$\sup_{h \in \operatorname{dom} \varphi} \Big\{ f(x+th) + f(x-th) - 2f(x) \Big\} = o(t). \tag{3.4}$$

On the other hand, the set of all $x \in X$ satisfying (3.4) is always G_δ (possibly empty).

Therefore, f is Hadamard-like, as well as Gâteaux, differentiable on a dense G_δ-subset of X. ©

2. More generally, we recover [C-F, 2001],

$$\text{Sep} \times \text{GDS} \subset \text{GDS}.$$

Proof. Suppose Y is the GDS factor. Let $f : Y \times X \to \mathbb{R}$ be convex continuous, and $\Omega \subset Y \times X$ be a non empty open set. Assume, for ease, that $2B_Y \times 2B_X \subset \Omega$ and f is bounded on Ω.

(**Step 1.**) Let $\varphi : X \to [0, +\infty]$ be the function provided by Theorem 1 with domain in B_X, and define $g : Y \to (-\infty, +\infty]$ by

$$g(y) := \begin{cases} \inf\{-f(y,x) + \varphi(x);\ x \in X\}, & \text{if } y \in 2B_Y \\ +\infty, & \text{else .} \end{cases}$$

Then g is concave and continuous on $2B_Y$.

As Y is a Gâteaux differentiability space, the function g is Gâteaux differentiable at some y in B_Y.

(**Step 2.**) By Theorem 1, there is $x \in B_X$ so that

$$g(y) = -f(y,x) + \varphi(x).$$

Thus, for every $k \in Y$ and every $h \in \operatorname{dom} \varphi$ we have, for all $t > 0$ sufficiently small,

$$\begin{aligned} & f(y+tk, x+th) + f(y-t\,k, x-t\,h) - 2f(y,x) \\ \leq\ & -g(y+t\,k) + \varphi(x+t\,h) \\ -\ & g(y-t\,k) + \varphi(x-t\,h) + 2g(y) - 2\varphi(x) \\ =\ & o(t) + o(t). \end{aligned}$$

Finally, local Lipschitzness of f and linear density of $\operatorname{dom} \varphi$ in X imply

$$f(y+t\,k, x+t\,h) + f(y-t\,k, x-t\,h) - 2f(y,x) = o(t)$$

as $t \searrow 0$, for every $k \in Y$ and every $h \in X$.

Therefore, f is Gâteaux differentiable at the point $(y,x) \in \Omega$. ©

Example 6 (Moors and Somasundaram, 2003)

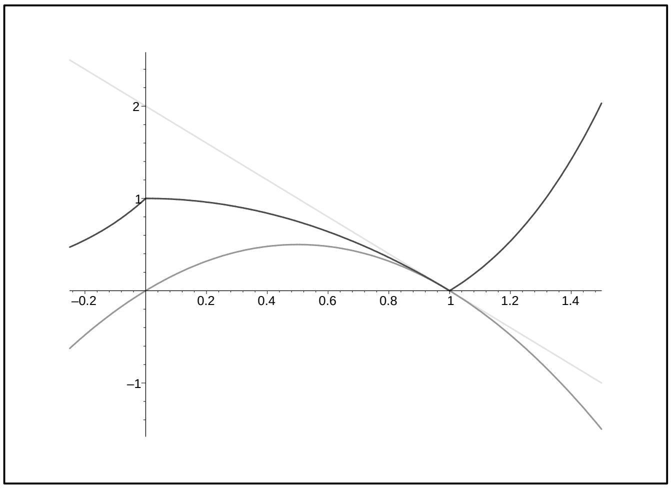
$$\boxed{\text{WASP} \underset{\neq}{\subset} \text{GDS}}$$

Viscosity is Fundamental

Definition [B-Z, 1996] f is β-*viscosity subdifferentiable* with subderivative x^* at x if there is a *locally Lipschitz* g, β-smooth at x, with

$$\nabla^\beta g(x) = x^*$$

and $f - g$ taking a local minimum at x. Denote all β-viscosity subderivatives by $\partial^V_\beta f(x)$.
All variational principles rely *implicitly* or *explicitly* on viscosity subdifferentials.



All Fréchet subdifferentials are viscosity subdifferentials

✓ We know many facts such as ...

- Bornology H = F in Euclidean space.

- Bornology F = WH in reflexive space.

- For locally Lipschitz f

$$\partial_G^V f = \partial_H^V f.$$

- Unless $\ell^1 \subset X$

$$\partial_{WH}^V f = \partial_F^V f$$

for locally Lipschitz *concave* f.

- When X has a Fréchet renorm

$$\partial_F^V f = \partial_F f$$

(e.g., reflexive or WCG Asplund spaces).

Example 7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ ($n > 1$) be continuous and Gateaux but not Fréchet differentiable at 0. Explicitly in $\mathbb{R}^2$, take*

$$\boxed{f(x,y) := \frac{xy^3}{x^2 + y^4}}$$

when $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Let

$$\boxed{g(h) := -|f(h) - f(0) - \nabla_G f(0)h|}$$

Then g is locally uniformly continuous and

(1) Uniquely, $\partial_G g(0) = \{0\}$.

(2) But $\partial_G^V g(0)$ is empty.

Proof. We check that $\nabla_G\, g(0) = 0$, so $\partial_G\, g(0) = \{0\}$. As always

$$\partial_G^V\, g(0) \subset \partial_G g(0).$$

Thus, if (2) fails, $\partial_G^V\, g(0) = \{0\}$, and there is a locally Lipschitz Gateaux-differentiable (hence Fréchet-differentiable) function k such that

$$k(0) = g(0) = 0, \quad \nabla^G\, k(0) = \nabla^G\, g(0) = 0$$

and

$$k \le g$$

in a neighbourhood of 0.

Thus, for small h,

$$\begin{aligned} \frac{|f(0+h) - f(0) - \nabla_G\, f(0)h|}{\|h\|} &\le -\frac{k(h) - k(0)}{\|h\|} \\ &\le \frac{|k(h) - k(0)|}{\|h\|} \end{aligned}$$

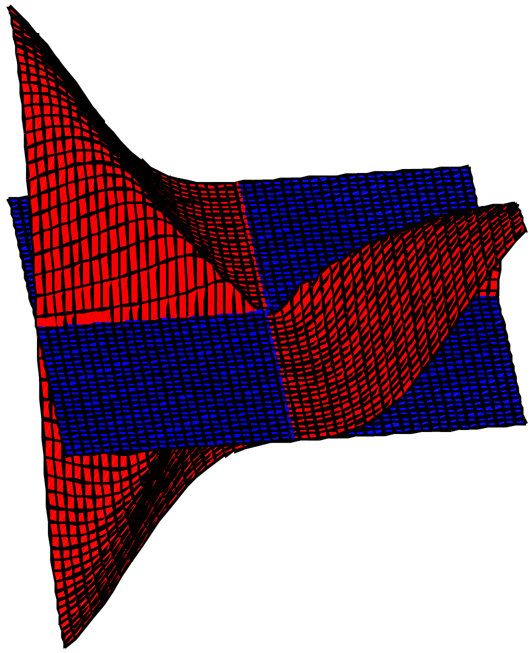
which implies that f is Fréchet-differentiable at 0, a contradiction. ©

3. Three Open Questions

1. *Viscosity.* In *Hilbert space* is

$$\partial_G^V f(x) \subsetneq \partial_G f(x)$$

possible for *Lipschitz* f? For continuous f we saw it is:



A non-viscosity 0-subdifferential

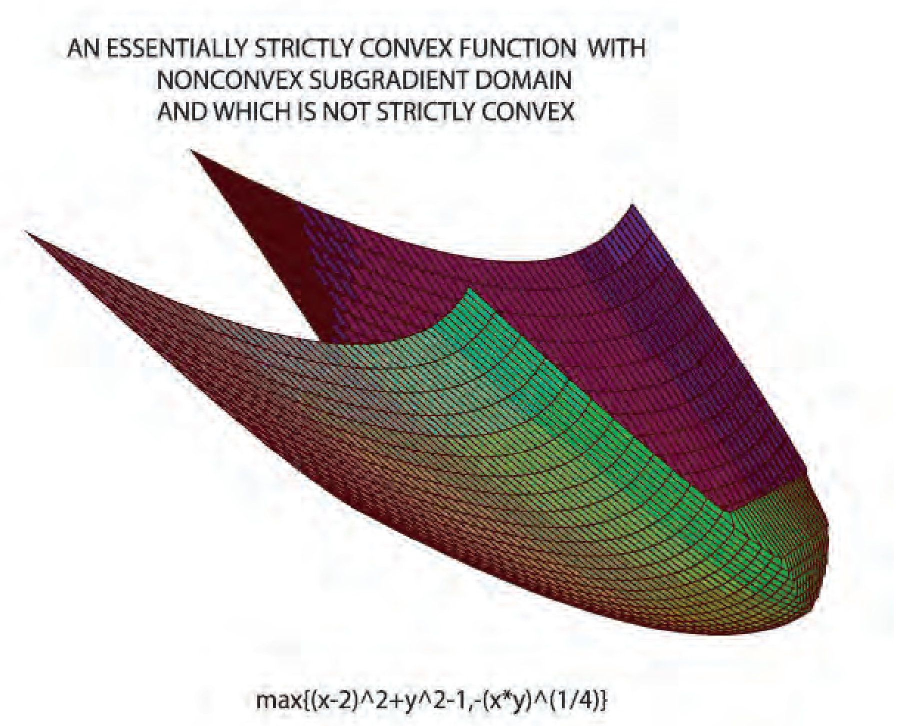
- Relatedly, show $\partial_H f(x) \neq \partial_H^V f(x)$ in (non-separable) *Hilbert space* for *continuous* f?

2. *Stability.* In light of the (unconditional ZFC) result of *Moors and Somasundaram*, is

$$\mathrm{Sep} \times \mathrm{WASP} \subset \mathrm{WASP}?$$

- Our methods seem at most to show that: $\mathrm{Sep} \times \mathrm{a}-\mathrm{WASP} \subset \mathrm{a}-\mathrm{WASP}$.

- For *Stegall's Class*: (G) $\subset$ (S) $\subset$ WASP, and (S) $\times$ (S) $\subset$ (S).



A Legendre-type function

- Drawing often helps . . .

3. *Stability.* Linear iterations (needed in work on an amazing continued fraction of *Ramanujan*). Consider

$$\boxed{z_{n+1} = z_n + c_n z_{n-1}, \qquad c_n \to c \in \mathbf{C}}$$

with $z_0 := a, z_1 := b$. *How does its behaviour relate to the case* $c_n \equiv c$*?*

- We were able to resolve the issue via a tail and **non-symmetric word analysis**:

  Theorem: in any matrix algebra if

  $$A_n A_{n-1} \cdots A_1 \to L$$

  with L *non-singular*, but perhaps only conditionally convergent, then

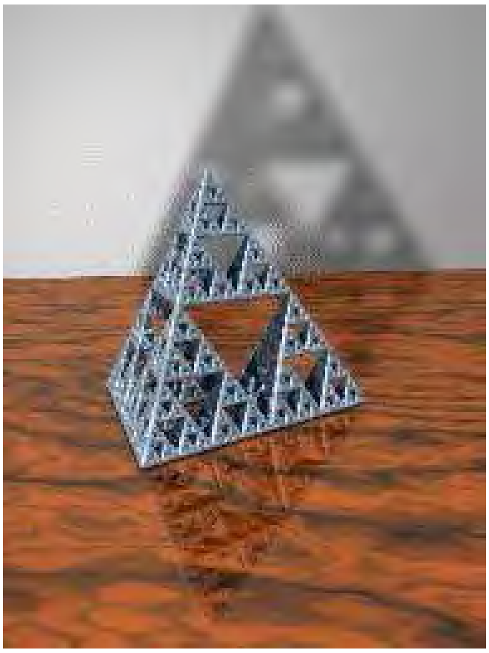
  $$(A_n + B_n)(A_{n-1} + B_{n-1}) \cdots (A_1 + B_1) \to M$$

  whenever

  $$\sum_n \|B_n\| < \infty.$$

- What can our field say to address this and other *linear algebra stability* issues?

"The most prominent requisite to a lecturer, though perhaps not really the most important, is a good delivery; for though to all true philosophers science and nature will have charms innumerably in every dress, yet I am sorry to say that the generality of mankind cannot accompany us one short hour unless the path is strewed with flowers."



Four References

B-P: J. Borwein and D. Preiss, "A smooth variational principle with applications to subdifferentiability and differentiability of convex functions," *TAMS*, **303**(1987), 517–527.

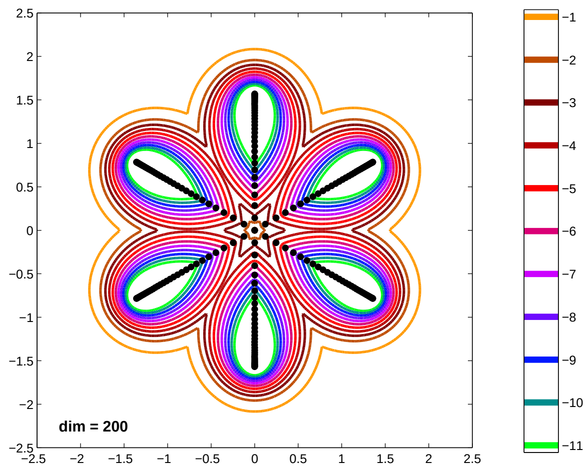
B-Z: J.M. Borwein and Qiji Zhu, "Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity," *SIAM J. Optimization*, **34**(1996), 1568–1591.

C-F: L. Cheng, M. Fabian, "The product of a Gâteaux differentiability space and a separable space is a Gâteaux differentiability space," *PAMS* **129**(2001), 3539–3541.

M-S: W. B. Moors and S. Somasundaram, "A Gâteaux differentiability space that is not weak Asplund," *PAMS*, in press.

Kuratowski-Ulam theorem. If X is Baire and Y has a countable π-base then a Borel subset B is residual in $X \times Y$ iff there exists a residual subset R of X s.t. $(\{x\} \times Y) \cap B$ is residual (in $\{x\} \times Y$) for all $x \in R$.

A π-base B is a family of non-empty open sets s.t. for each open set U there exists an element of B inside U



A tri-diagonal pseudospectrum