Slices, Bumps and Cusps:

Underpinnings of Nonsmooth Analysis

Jonathan M. Borwein, FRSC

Research Chair in IT
Dalhousie University

Halifax, Nova Scotia, Canada

Dalhousie Distributed Research Institute and Virtual Environment

First Franco-Canadian Meeting

Toulouse, July 15th 2004

If mathematics describes an objective world just like physics, there is no reason why inductive methods should not be applied in mathematics just the same as in physics.

Kurt Gödel (1951)

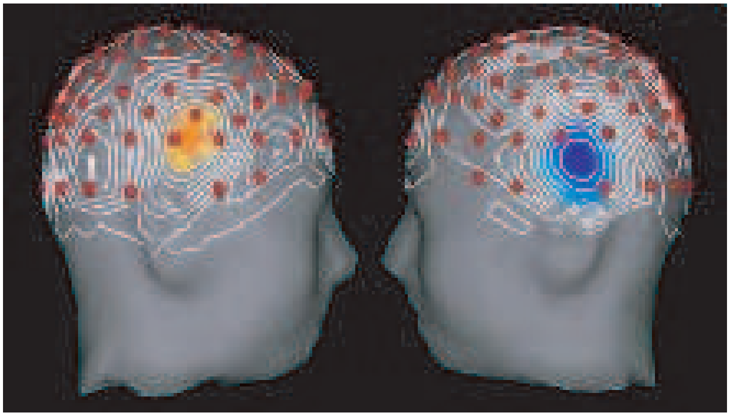
URL: `www.cs.dal.ca/~jborwein`

ToVA: CMS-Springer 04 **Talk revised**: 11-07-04

MY INTENTIONS IN THIS TALK

Most significant results or constructions in non-smooth analysis rely on exposing and really understanding underlying objects.

Insight taking place



Usually these objects are

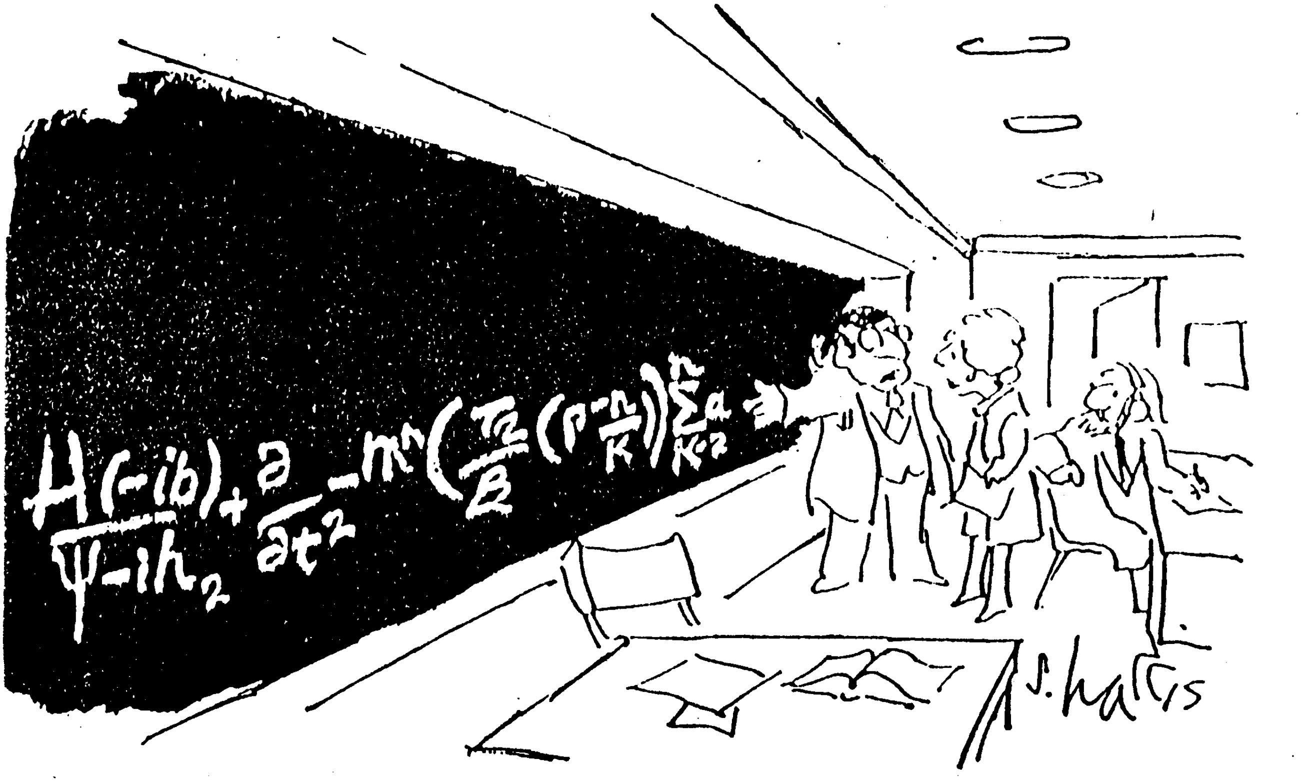
- **convex** or
- **differentiable**

or both

✓ As an illustration, in $\mathbb{R}^n$

Theorem 1 (BFKL, 2001) **Every "reasonable" connected set with zero interior to its domain is exactly the range of the gradient of a continuously differentiable bump function, i.e., with compact support.***

*Online slides are a superset of this talk



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

After a topological detour, I shall *illustrate* this in **five** ways:

1. Smooth variational principles and **bumps**

2. **Bumps** and generalized gradients

3. **Derivatives** and best approximations to sets

4. Non-**differentiable** mean value theorems and **convex** sandwich theorems

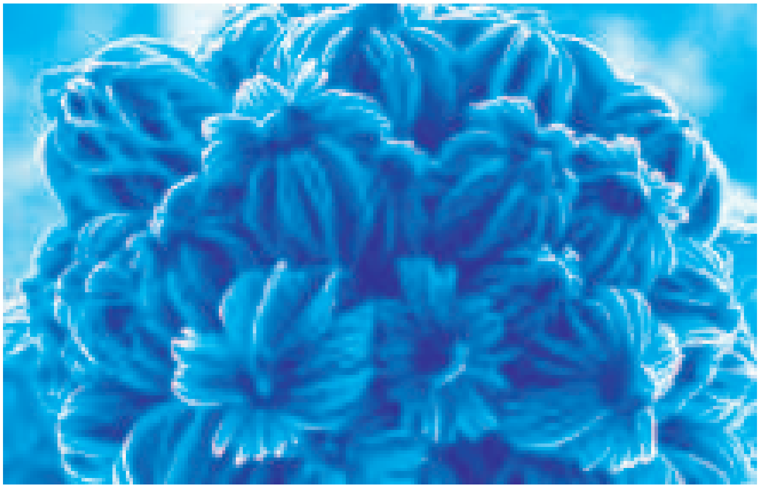
5. **Convex** functions and the Banach spaces they populate

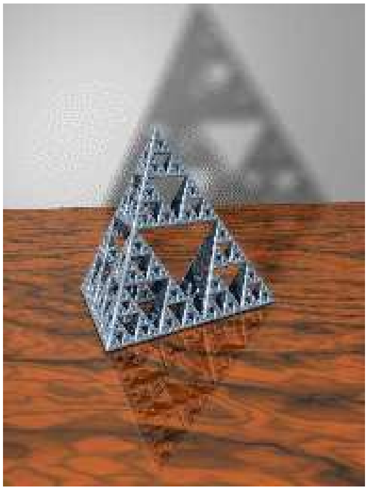
- Full references will be found in

  J.M. Borwein and Qiji (Jim) Zhu, *Techniques of Variational Analysis* CMS-Springer Books, in Press, 2004.

Michael Faraday

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.





- So I offer nano-flowers and nourishing tubers

Franciscus Vieta

 (1540-1603)

Arithmetic is absolutely as much science as geometry [is]. Rational magnitudes are conveniently designated by rational numbers, and irrational magnitudes by irrational [numbers]. If someone measures magnitudes with numbers and by his calculation get them different from what they really are, it is not the reckoning's fault but the reckoner's.

Rather, says Proclus, **ARITHMETIC IS MORE EXACT THAN GEOMETRY**. *To an accurate calculator, if the diameter is set to one unit, the circumference of the inscribed dodecagon will be the side of the binomial [i.e. square root of the difference]* $72-\sqrt{3888}$. *Whosoever declares any other result, will be mistaken, either the geometer in his measurements or the calculator in his numbers.*

SOME TOPOLOGY

- The acronym *usco (cusco)* denotes a (convex-valued) upper semicontinuous non-empty compact-valued multifunction (set-valued function).

- These are fundamental because they describe common features of maximal monotone operators, convex subdifferentials and Clarke generalized gradients.

◇ Cuscos are the most natural extensions of continuous (single-valued) functions.

- The Clarke gradient is usually much too large (generically "maximal", see below).

◇ By contrast convex subdifferentials and maximal monotone operators are always "minimal" (interior to their domains), as are the Clarke subdifferentials of a.e. strictly differentiable functions (BM).

- An usco (cusco) mapping Φ from a topological space T to subsets of a (linear) topological space X is a *minimal usco (cusco)* if its graph does not strictly contain the graph of any other usco (cusco) on T.

- A Banach space is of *class* (S) (Stegall) provided every weak* usco from a Baire space into X^* has a selection which is generically weak* continuous. Every smooth Banach space is class (S).

- A Banach space is (*weak*) *Asplund* if convex functions on the space are generically Fréchet (Gateaux) differentiable. Equivalently, every separable subspace has a separable dual (e.g., reflexive spaces).

In our setting a fundamental result is:

- A Banach space X is Asplund if and only if every locally bounded minimal weak* cusco from a Baire space into X^* is generically singleton and norm-continuous. A fortiori, Asplund spaces are class (S).

We show the power of minimality by easily proving a generic (partial) differentiability result:

Theorem 2 *Suppose that f is locally Lipschitz on an open subset A of a Banach space X and possesses a minimal subgradient on A.*

(a) *When Y is a class (S) subspace of X then f is generically $Y-$Hadamard smooth throughout A.*

(b) *When Y is an Asplund subspace of X then f is generically $Y-$Fréchet smooth throughout A.*

Proof. Let Ω_Y be the restriction of elements of ∂f to Y.

As the composition of the 'restriction' linear operator

$$R : x^* \to x^*|Y$$

and the minimal cusco ∂f, Ω_Y is a minimal cusco from $A \subset X$ to Y^*.

(a) Consider first the class (S) case.

Then Ω_Y is generically single-valued on the open (Baire) set A. An easy application of Lebourg's mean-value theorem establishes that at each such point f is (strictly) Y-Hadamard smooth.

(b) The Asplund case follows similarly. ©

◇ Note how Y and X^* have been 'detached'!

- An immediate consequence is that in *any* Banach space, continuous convex functions are generically Fréchet (respectively Gateaux) differentiable with respect to any fixed Asplund (respectively class (S)) subspace.

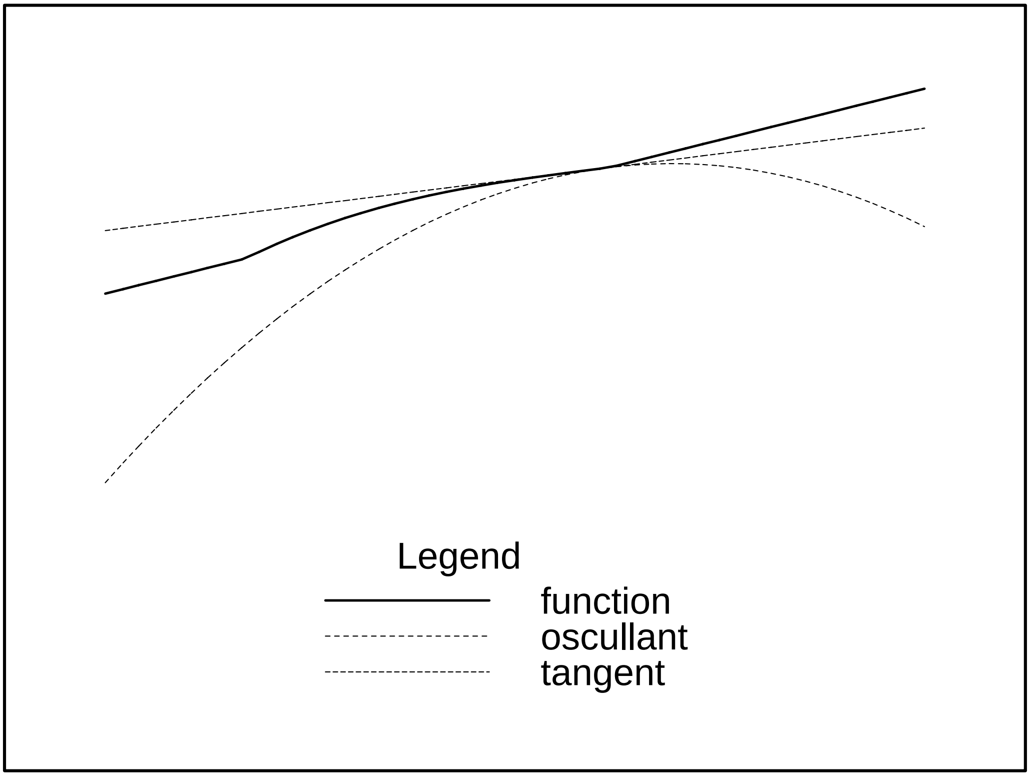
Remark 1 *Fabian, Zajíček and Zizler give a category version of Asplund's result that if a Banach space and its dual have rotund renorms one can find a rotund renorm whose dual norm is rotund simultaneously.*

- Their technique allows us to show that if Y is a subspace of X such that both X and X^* admit 'Y-rotund' renorms (appropriately defined), then X can be renormed to be simultaneously Y-smooth and Y-rotund.

BUMPS I: VARIATIONAL PRINCIPLES

- All variational principles devolve from Ekeland's powerful (1974) reworking of the Bishop-Phelps theorem* (1961).

- More powerful recent ones exploit smoothness of the underlying space—by partially capturing the smoothness of an **osculating** norm or bump function



***All Banach spaces are "sub-reflexive"*

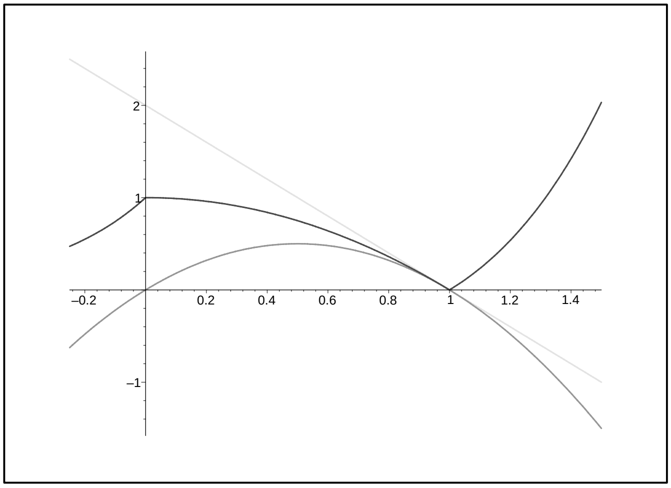
Viscosity is Fundamental

Definition [BZ, 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 $\partial_\beta^v f(x)$.

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

$$\partial^v_G f = \partial^v_H f \qquad \partial_G f = \partial_H f$$

- When $\ell^1 \not\subseteq X$

$$\partial^v_{WH} f = \partial^v_F f$$

  for locally Lipschitz *concave* f

- When X has a Fréchet renorm

$$\partial^v_F f = \partial_F f$$

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

Example 1 *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.*

✓ The proof is easy but instructive ...

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

$$\partial^v_G\, g(0) \subset \partial_G g(0).$$

Thus, if (2) fails, $\partial^v_G\, g(0) = \{0\}$, and yet there is a locally Lipschitz Gateaux (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 zero.

Thus, for small h,

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

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

The Smooth Variational Principle

Theorem 3 (Borwein-Preiss, 1987) *Let X be Banach and let $f : X \to (-\infty, \infty]$ be lsc, let $\lambda > 0$ and let $p \geq 1$. Suppose $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \varepsilon.$$

Then there exist y and a sequence $\{x_i\} \subset X$ with $x_1 = z$ and a continuous convex *function $\varphi_p : X \to \mathbb{R}$ of the form*

$$\varphi_p(x) := \sum_{i=1}^{\infty} \mu_i \|x - x_i\|^p,$$

where $\mu_i > 0$ and $\sum_{i=1}^{\infty} \mu_i = 1$ such that

(i) $\|x_i - y\| \leq \lambda, n = 1, 2, \ldots,$

(ii) $f(y) + (\varepsilon/\lambda^p)\varphi_p(y) \leq f(z)$, *and*

(iii) $f(x) + \frac{\varepsilon}{\lambda^p}\varphi_p(x) > f(y) + \frac{\varepsilon}{\lambda^p}\varphi_p(y)$ *for $x \neq y$*

Corollary 1 *All extended real-valued lsc (resp. convex) functions on a smoothable (Gateaux, Fréchet, ...) space are densely* subdifferentiable *(resp. differentiable) in the same sense.*

- $f : X \to (\infty, \infty]$ attains a *strong minimum* at $x \in X$ if $f(x) = \inf_X f$ and whenever $x_i \in X$ and $f(x_i) \to f(x)$, we have $\|x_i \to x\|$ (The problem is *well posed*.)

- also we set $\|g\|_\infty := \sup\{|g(x)| : x \in X\}$.

Theorem 4 (Deville-Godefroy-Zizler, 1992) *Let X be Banach and let Y be a Banach space of continuous bounded functions on X such that*

(i) $\|g\|_\infty \le \|g\|_Y$ *for all $g \in Y$.*

(ii) *For $g \in Y$ and $z \in X$, $x \mapsto g_z(x) = g(x+z)$ is in Y and $\|g_z\|_Y = \|g\|_Y$.*

(iii) *For $g \in Y$ and $a \in \mathbb{R}$, $x \mapsto g(ax)$ is in Y.*

(iv) *There exists a bump function in Y.*

Then, whenever $f : X \to (\infty, \infty]$ is lsc and bounded below, the set G of $g \in Y$ such that $f + g$ attains a strong minimum on X is residual (in fact a dense G_δ set).

- Picking Y appropriately leads to:

Theorem 5 *Let X be Banach with a Fréchet smooth bump and let f be lsc. There is $a > 0$ ($a = a(X)$) such that for $\varepsilon \in (0, 1)$ and $y \in X$ satisfying*

$$f(y) < \inf_X f + a\varepsilon^2,$$

there is a Lipschitz Fréchet differentiable g and $x \in X$ such that

(i) *$f + g$ has a strong minimum at x,*

(ii) *$\|g\|_\infty < \varepsilon$ and $\|g'\|_\infty < \varepsilon$,*

(iii) *$\|x - y\| < \varepsilon$.*

Corollary 2 *For any C^1 bump function b on a finite dimensional space*

$$0 \in \operatorname{int} R(\nabla b)$$

The Stegall Variational Principle

As we add more geometry we may often refine the variational principle:

- Again, $x \in S$ is a *strong minimum* of f on S if $f(x) = \inf_S f$ and $f(x_i) \to f(x)$ implies $\|x - x_i\| \to 0$.

- A *slice* for f bounded above on S is:

$$S(f, S, \alpha) := \{x \in S : \ f(x) > \sup_S \ f - \alpha\}.$$

- A necessary and sufficient condition for a f to attain a strong minimum on a closed set S is diam $S(-f, S, \alpha) \to 0$ as $\alpha \to 0+$.

Theorem 6 (Stegall, (1978)) *Let X be Banach and let $C \subset X$ be a closed bounded convex set with the Radon-Nikodym property, Let f be lsc on C and bounded from below.*

For any $\varepsilon > 0$ there exists $x^* \in X^*$ such that $\|x^*\| < \varepsilon$ and $f + x^*$ attains a strong minimum on C.

- The following variant due to Fabian (1983) is often convenient in applications

Corollary 3 *Let X be Banach with the Radon-Nikodym property (e.g., reflexive) and let f be lsc. Suppose there exists $a > 0$ and $b \in \mathbb{R}$ such that*

$$f(x) > a\|x\| + b, \quad x \in X.$$

Then for any $\varepsilon > 0$ there exists $x^* \in X^*$ such that $\|x^*\| < \varepsilon$ and $f + x^*$ attains a strong minimum on X.

✓ In separable space we may set the perturbation in advance:

A One-perturbation Variational Principle

Theorem 7 *Let X be a Hausdorff space which admits a proper lsc function*

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

with compact level sets. For any proper lsc bounded below function $f : X \to \mathbb{R} \cup \{+\infty\}$ the function $f + \varphi$ attains its minimum.

In particular, if dom φ *is relatively compact, the conclusion is true for any proper lsc 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 $S\colon H \to X$ $(H := \ell_2)$.

- φ is *almost Hadamard* smooth: $x \in \operatorname{dom}\varphi$

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

- We recover a recent result (CF, 2001) open for 25 years:

Corollary 4 GDS $\times$ Sep $\subset$ GDS.

Proof Sketch. Suppose Y is the Gateaux differentiability space factor. Let $f : Y \times X \to \mathbb{R}$ be convex continuous, and $\Omega \subset Y \times X$ non empty open. Without loss, $2B_Y \times 2B_X \subset \Omega$ and f is bounded on Ω.

Let $\varphi : X \to [0, +\infty]$ be as in Theorem 7 with domain in B_X, and define

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

Then g is concave and continuous on $2B_Y$. As Y is a GDS, the function g is Gâteaux differentiable at some y in B_Y.

Moreover

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

and (y,x) is a point of joint differentiability ... ©

- This is particularly interesting because we cannot show the corresponding generic result:

$$\text{WASP} \times \text{Sep} \overset{?}{\subset} \text{WASP},$$

  while recently Moors and Somasundaram (2003) showed—unconditionally—that

Example 2

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

answering another long open question with delicate set-theoretic topological tools.

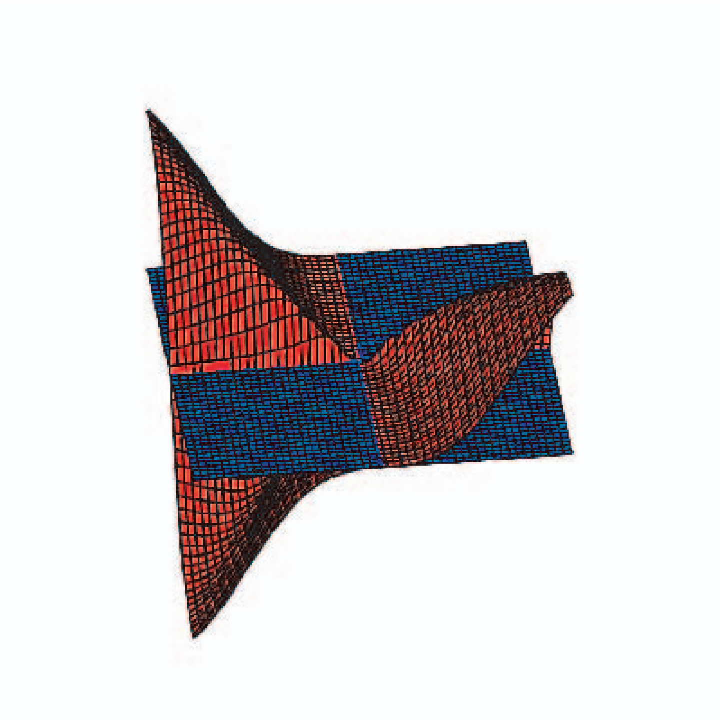
- Lassonde and Revalski (2004) have extended the single perturbation principle to ensure generic strong minimality.

Two 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 was:



A non-viscosity subdifferential

2. **Genericity.** WASP $\times$ Sep $\stackrel{?}{\subset}$ WASP.

BUMPS II: SUBDIFFERENTIALS

Maximality and Genericity

- These powerful positive results are complemented by the following negative ones:

Below B_{X^*} is the dual ball, $(\mathcal{X}_{B_{X^*}}, \rho)$ is the space of real-valued non-expansive mappings

$$|f(x) - f(y)| \leq \|x - y\|$$

in the uniform metric, while ∂_0 and ∂_a denote the *Clarke and approximate subdifferentials*

$$\partial_a f(x) := \{x^* \colon \; x^* \overset{w*}{\leftarrow} x_n^* \in \partial_H f(x_n), x_n \to x\}$$

and

$$\partial_0 f(x) = \overline{co}^* \partial_a f(x).$$

- In reasonable (reflexive or separable) spaces, $\partial_0 f(x)$ is the limit of nearby gradients.

Theorem 8 *(Maximal Subdifferentials) Let A be open in a Banach space X.*
(i) Then

$$\{g \in \mathcal{X}_{B_{X^*}} : \partial_0 g(x) = B_{X^*} \text{ for all } x \in A\}$$

is residual in $(\mathcal{X}_{B_{X^*}}, \rho)$.

(ii) If X is smooth

$$\{g \in \mathcal{X}_{B_{X^*}} : \partial_a g(x) = B_{X^*} \text{ for all } x \in A\}$$

is residual in $(\mathcal{X}_{B_{X^*}}, \rho)$.

- ◇ Thus usually (generically) even the limiting subdifferential is everywhere maximal (and convex, agreeing with the Clarke subdifferential).

- $T(x) := \nabla f(x) + B_{X^*}$ is also a subgradient. Much more is true (BMW).

- Despite this, the limiting subdifferential of a Lipschitz function can be non-convex a.e. (BBW)—save on $\mathbb{R}$ where it differs from the Clarke subdifferential at most countably.

Moreover,

Theorem 9 *Let* $0 \in A$ *be an open connected and bounded subset of* $\mathbb{R}^N$ *and let* $\varepsilon > 0$.

There is a locally Lipschitz function $f : \mathbb{R}^N \to \mathbb{R}$ *such that*

$$R(\partial_a f) \subset \overline{A}$$

and

$$\mu\{x : \partial_a f(x) \neq \overline{A}\} < \varepsilon.$$

The proof relies on two facts:

Fact 1 *By Theorem 1, such connected A can be realized as the range of the gradient of a continuously differentiable bump (bounded support) function b_A.*

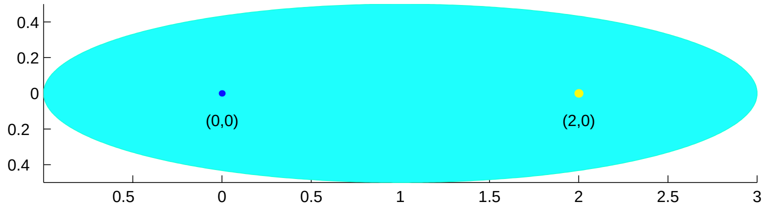
Step 1. The **support function** of a strictly convex body

$$\sigma_C(x) := \sup_{y \in C} \langle y, x \rangle$$

leads to a bump

$$b_C(x) := \frac{3\sqrt{3}}{8} \left(\max \left\{ 1 - \sigma_C(-x)^2, 0 \right\} \right)^2$$

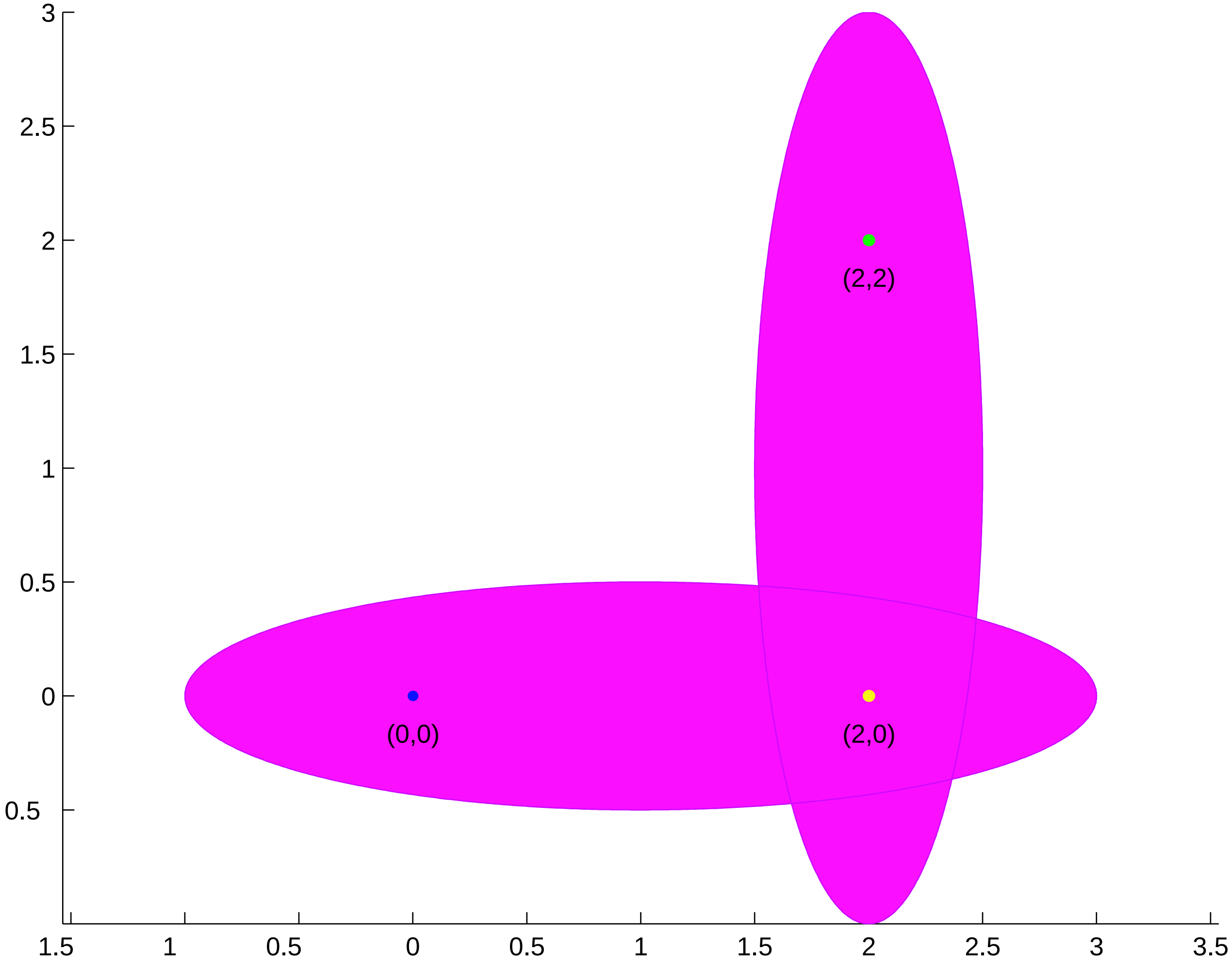
with range exactly C.



- This is clearest for the case of an ellipse $E := \{x \colon \langle Ax, x \rangle \leq 1\}$ where

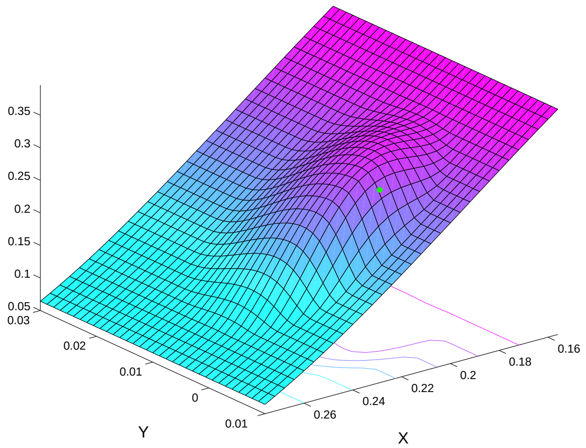
$$\sigma_E(y) = \langle Ax, x \rangle^{1/2}$$

Step 2. A disjoint sum then leads to

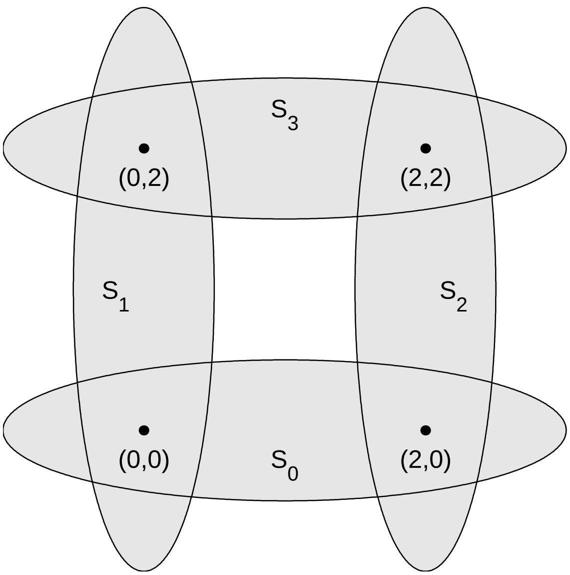


A Non-convex Gradient Range ∇b_C

Step 3. Build a flat patch on a bump range



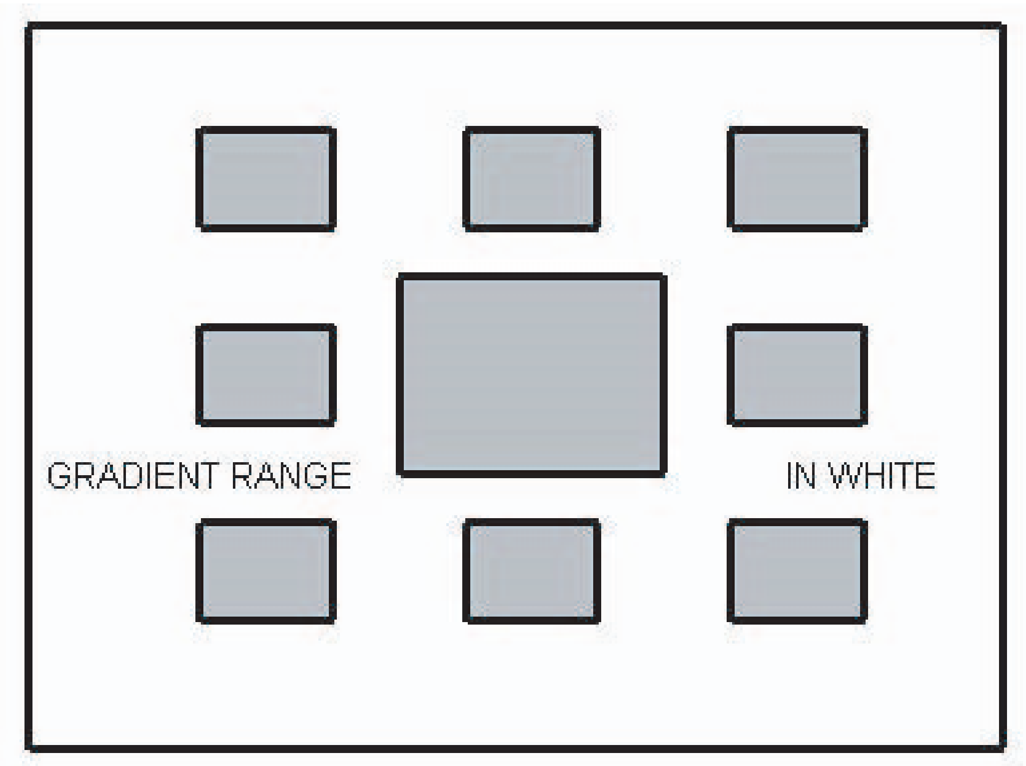
Step 4. Superposing a bump on a flat patch of another leads to



A Non-simply Connected Gradient Range $\nabla b_{C_1 \cup C_2}$

• **Step 5.** Careful analysis leads, in the limit, to the general result.

◇ Indeed, there is a C^1 bump $b : \mathbb{R}^2 \to \mathbb{R}$ such that $\nabla b(\mathbb{R}^2)$ is exactly the k-th approximation to the Sierpinski carpet (BFKL).

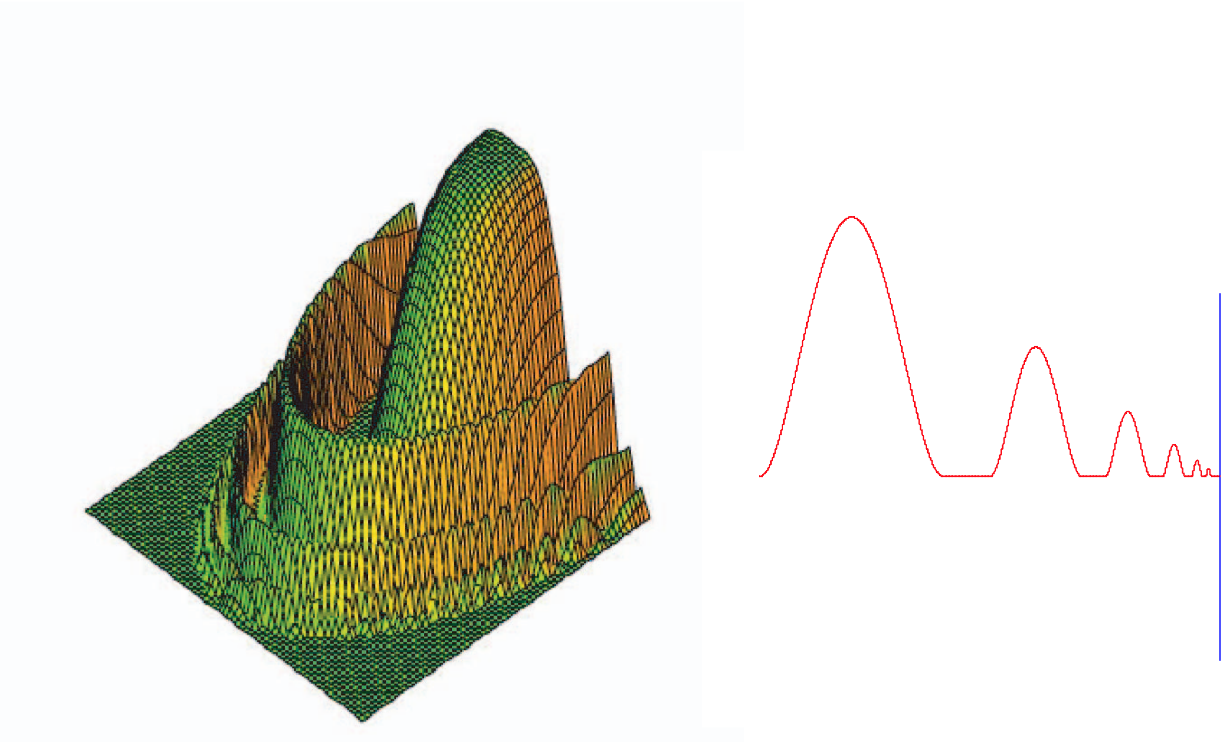


A Multiply Connected Gradient Range

Fact 2 *One can 'seed' an open dense set of small measure with dilated bumps of constant gradient range, A, forcing all limits to be A.*

Reason. As observed by Ioffe, dilation and translation do not effect the range. Consider

$$f_A(x) := \sum_{n=0}^{\infty} 2^{-n-1}\, b_A(a_n + 2^{n+1}x)$$



Scaled bumps in one and two dimensions
Limiting blue subdifferential at right

✓ Now, Facts 1 and 2 prove Theorem 9.

Two Open Questions

- Can one build an *explicit* example of a function on $\mathbb{R}^2$ with $\partial_a f(x) \equiv B_2$?

- Is it always true in $\mathbb{R}^N$ that the range of a C^1 bump's gradient is semi-closed:

$$\mathrm{R}\,(\nabla b) = \mathrm{cl} - \mathrm{int}\,\mathrm{R}\,(\nabla b)?$$

  – with enough smoothness this is true (C^{N+1}, Rifford, 2003).

- The situation is quite different in infinite dimensions (BFL, Deville-Hajek and others): the interior may be empty and one can achieve many strange sets.

- A norm is *Kadec-Klee* (sequentially) if the weak and norm topologies coincide (sequentially) on the boundary of the unit ball, as in Hilbert space.

Theorem 10 *Let C be a closed subset of a reflexive Banach space X with a* Kadec-Klee *norm.*

(a) *(Density) The set of points in X at which every minimizing sequence clusters to a best approximation is dense in X.*

(b) *(Projection) If in addition, the original norm is Fréchet then*

$$\partial_F d_C(x) \subset \partial_F d_C(P_C(x))$$

where $P_C(x)$ is the (set of) best approximations of x on C.

(c) *In particular, in any Fréchet LUR norm on a reflexive space, this holds for all sets in the Fréchet sense with a single-valued metric projection.*

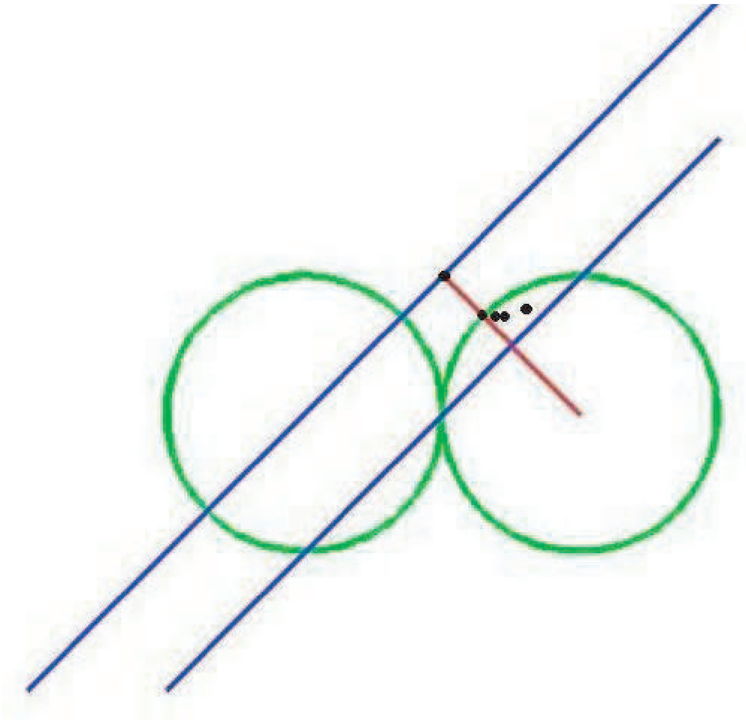
Proof. (a) We may assume $x_n \rightarrow_w p$ and at any of the dense set of points with

$$\phi \in \partial_F d_C(x) \neq \emptyset$$

all minimizing sequences actually converge in norm to p since

$$\phi(x_n - x) \to d_C(x) \Rightarrow \|x_n - x\| \to \|p - x\|,$$

and by Kadec-Klee $x_n \to p$, and $p = P_C(x)$.



The Fréchet slice forces the approximating sequence to line up

The corresponding subgradient is a proximal normal to C at p.

(b-c) Finally, when the norm is F-smooth, simple derivative estimates show that any member of $\partial_F d_C(x)$ must lie in

$$\partial_F d_C(P_C(x)).$$

©

✓ This used to be hard.

- (**Lau-Konjagin** (1976-86)) X is reflexive and Kadec-Klee iff best approximations always exist densely (or generically).

- Theorem 10 easily shows the *normal cone* defined in terms of *distance functions* is always contained in the normal cone defined in terms of *indicator functions*.

- In Hilbert space we may conclude

$$\partial_F d_C(x) \subset \partial_\pi d_C(P_C(x)),$$

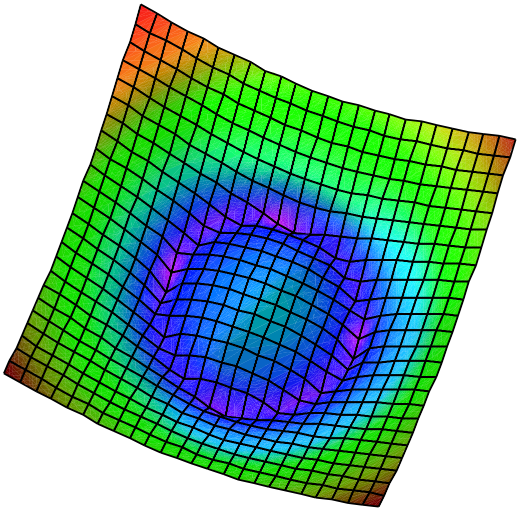
where ∂_π denotes the set of *proximal* subgradients.

Random Subgradients

- $\partial_0 d_C$ is a minimal cusco for all closed C iff the norm is uniformly Gateaux.

- While d_C is often too well behaved, $\sqrt{d_C(x)}$ is not Lipschitz and choosing C wisely provides many counter-examples:

$$\sqrt{d_S(x)} = \sqrt{|1 - \|x\||}$$



Burke
Lewis
Overton

How random gradients fail

Two Open Questions

- Every closed set in every reflexive space (every renorm of Hilbert space) admits at least *one best approximation*.

  (**Stronger variant.**) For every closed set of every reflexive space the *proximal normal points are norm dense* in the norm boundary.

  ✓ Any counter-example is necessarily unbounded (and fractal-like)

- Every norm closed set in a reflexive Banach space with unique best approximations for every point in A (a Chebyshev set) is convex.

  [True in weak topology, and so in R^N.]

DERIVATIVES II and CONVEXITY I

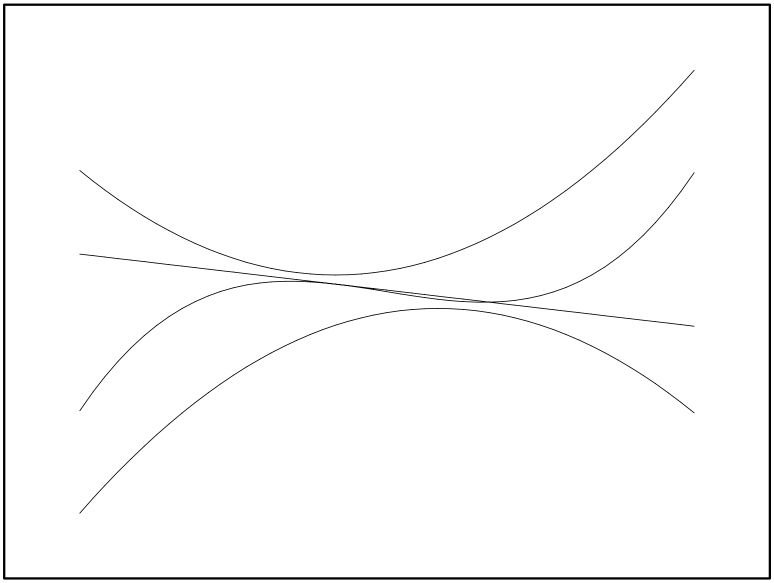
Duality Inequalities

- The following hybrid inequality is based on the two-set Mean Value theorem of Clarke and Ledyaev (94) and its Fenchel reworking by Lewis & Ralph (96).

Theorem 11 *(Three Functions) Let* $C \subset \mathbb{R}^n$ *be nonempty compact convex and let* f *and* h *be lsc functions with* $\operatorname{dom}(f) \cup \operatorname{dom}(h) \subset C$.

For any Lipschitz $g : C \to \mathbb{R}$ *there is* $z^* \in \partial_0 g(C)$ *(the* Clarke *subdifferential) such that*

$$\begin{aligned} (\min(f-g) \;&+\; \min(h+g)) \\ \leq -f^*(z^*) - h^*(-z^*) \;&\leq\; \min(f+h). \end{aligned}$$



A Three Function Sandwich

- The smooth case (BF) applies the classical Mean value theorem to $t \mapsto g(\overline{x}(t))$ for an arc, $\overline{x}$, on $[0,1]$ obtained via Schauder's fixed point theorem.

- The nonsmooth case follows by 'mollification'— the limits lie in the Clarke subdifferential.

- Fenchel Duality is 'recovered' from $g := f$. Recall, $f^*(t) = \sup_x y(x) - f(x)$.

Finding the arc. We may smoothify since $\left(f + \varepsilon\|\cdot\|^2\right)^*$ is differentiable.

Let $M := 2\sup\{\|c\| : c \in C\}$ and

$$W := \{x : [0,1] \to C : \mathrm{Lip}(x) \le M\}.$$

By Arzela-Ascoli, W is compact in the uniform norm topology.

For $x \in W$ define a continuous self map $T : W \to W$ by

$$Tx(t) := \int_0^t \nabla f^* \circ \nabla g \circ x + \int_t^1 \nabla h^* \circ (-\nabla g) \circ x.$$

Since W is compact and convex, the Schauder fixed point theorem shows there is $x \in W$ such that $\overline{x} = T\overline{x}$. That is,

$$\overline{x}(t) = \int_0^t \nabla f^* \circ \nabla g \circ \overline{x} + \int_t^1 \nabla h^* \circ (-\nabla g) \circ \overline{x}.$$

• A striking partner is:

Theorem 12 *(Two Functions) Let $C \subset \mathbb{R}^n$ be nonempty compact convex and f proper convex lower semicontinuous with* dom $(f) \subset C$*. If $\alpha \neq 1$ and $g : [C, \alpha C] \to \mathbb{R}$ is Lipschitz then there are $z^* \in \partial_0 g([C, \alpha C])$ and $a \in C$ such that*

$$[g(\alpha a) - g(a)]/(\alpha - 1) - f(a) \geq f^*(z^*).$$

⋄ Two fine specializations follow.

Corollary 5 *Let $C \subset \mathbb{R}^n$ be compact convex and f proper convex lower semicontinuous with* dom $(f) \subset C$*. If $g : [C, -C] \to \mathbb{R}$ is Lipschitz then there are $z^* \in \partial_0 g([C, -C])$ and $a \in C$ such that*

$$[g(a) - g(-a)]/2 - f(a) \geq f^*(z^*).$$

Hence

$$f^*(z^*) \leq 0$$

if f dominates the <u>odd part</u> of g on C.

- The comparison of f to the odd part of g reinforces the suggestion that fixed point theory is central to these results.

Corollary 6 *Let* $C \subset R^n$ *be nonempty, compact and convex and* f *lsc with* $\mathrm{dom}\,(f) \subset C$*. If* $g : [C, 0] \to \mathbb{R}$ *is Lipschitz then there are* $z^* \in \partial_0 g([C, 0])$ *and* $a \in C$ *such that*

$$f(a) + f^*(z^*) \leq g(a) - g(0).$$

Hence

$$f^*(z^*) \leq 0$$

whenever f *dominates* $g - g(0)$ *on* C*.*

- By contrast, this corollary can be obtained and strengthened by variational methods.

Theorem 13 *Let A be nonempty open bounded in a Banach space and let $g : \overline{A} \to \mathbb{R}$ be Lipschitz. If $x \in \operatorname{int} A$ and*

$$t := \inf\{\|z^*\| : z^* \in \partial_0 g(z), z \in A\} > 0$$

then

$$\sup_{u \in \partial \overline{A}} (g(u) - t\|u - x\|) \geq g(x).$$

✓ Specialized to the unit ball with $x := 0$ we obtain, a la Corvallec:

Corollary 7 *(Rolle Theorem) Let B be the closed unit ball in $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ a Lipschitz function. Then there is $x^* \in \partial_0 g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in \partial B} |g(a)|.$$

⋄ Contrastingly:

Corollary 8 *(Odd Rolle Theorem) Let B be the closed unit ball in $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ a Lipschitz function. Then there is $x^* \in \partial_0 g(B)$ such that*

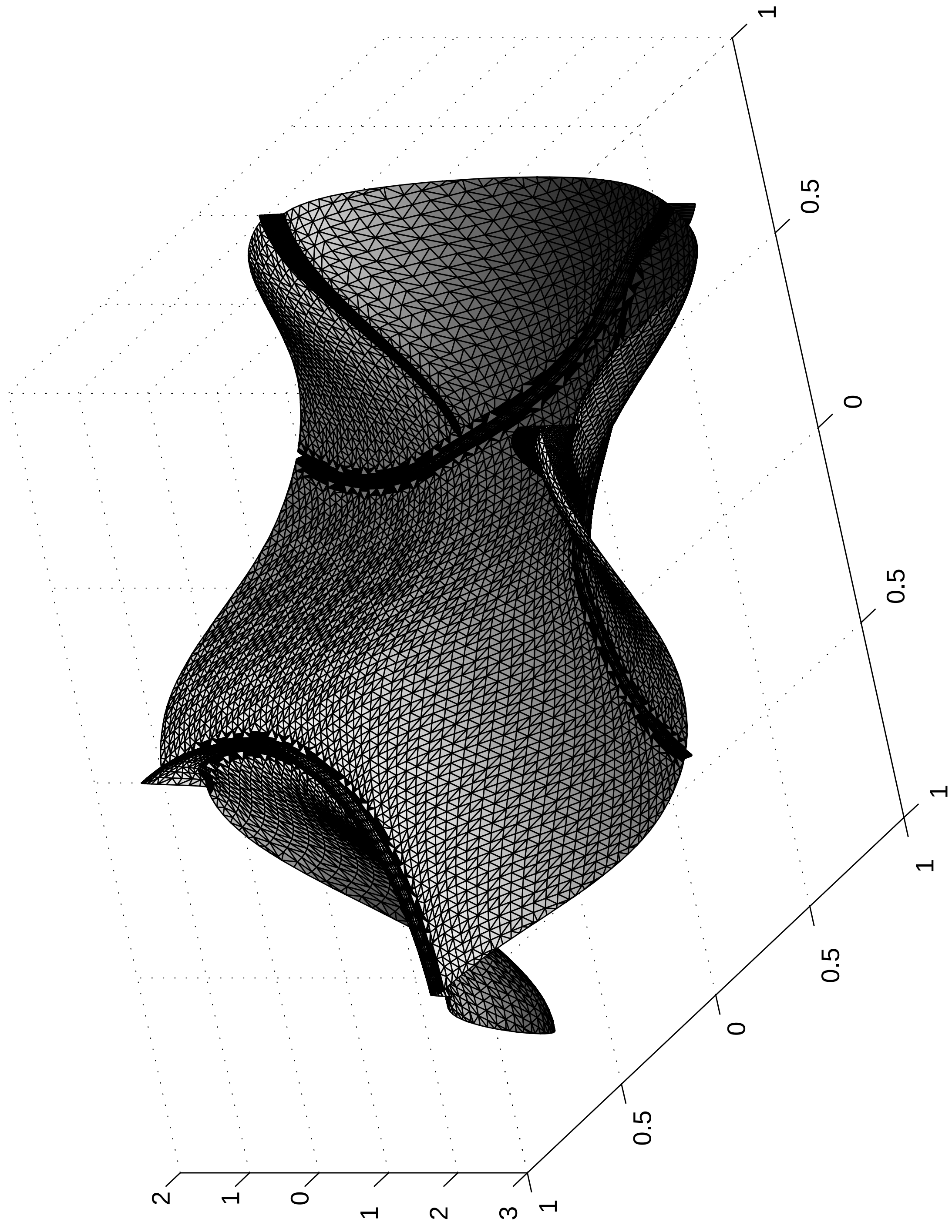
$$\|x^*\|_* \le \max_{a \in B} \frac{g(a) - g(-a)}{2}.$$

- That this last result is 'topological' is heightened by the following example (BKW):

Remark 2 *Corollary 8 fails if B is replaced by the unit sphere S. Indeed, there is a C^1 mapping $f : B \subset \mathbb{R}^2 \to \mathbb{R}$ such that*

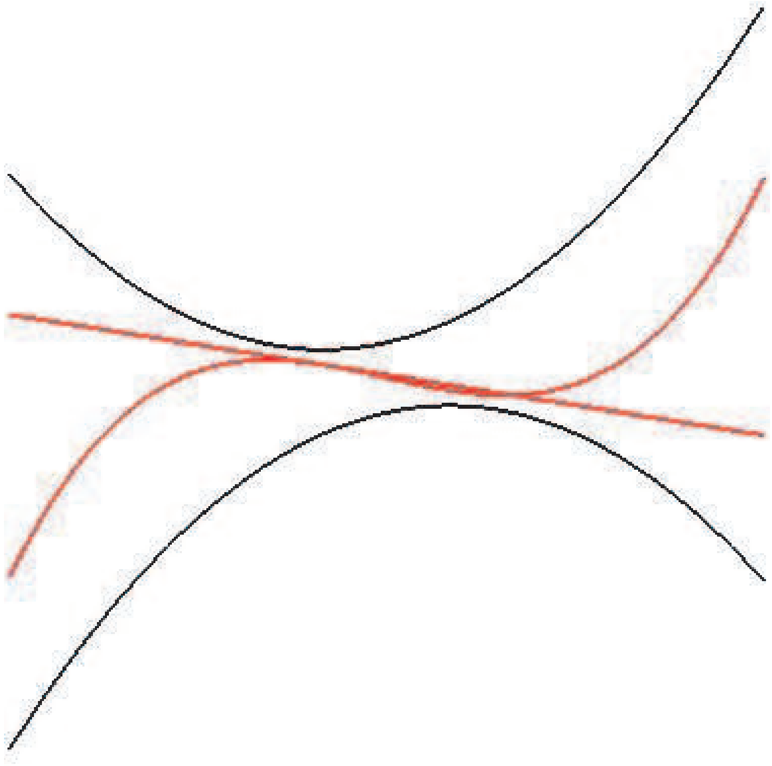
(i) $f|S$ is even; but

(ii) f has no critical point in B.



A Function Symmetric on S
With no Critical Point in B

Two Open Questions



- The picture suggests that in the sandwich theorem the slope is actually achieved by a tangent. Is this true?

- Can one avoid using Brouwer's fixed point theorem in the proof—a variational proof?

CONVEXITY II: BANACH SEQUENCES

Convex function properties are tightly coupled to the sequential properties of the spaces they may inhabit. We finish by illustrating this in three cases.

1. Finite dimensional spaces

2. Spaces containing ℓ_1

3. Grothendiek spaces.

Fact 3 (Josephson-Nissensweig) *A Banach space is infinite dimensional* **iff** *it contains a* **JN sequence***: that is, a norm-one but weak-star null sequence.*

- This is easy in separable space—e.g., the unit vectors in ℓ^2—but appears *hard* in general.

Theorem 14 *(a) Every continuous convex function finite throughout X is bounded on bounded sets* **iff** *(b) X is a* **JN space***: weak-star and norm convergence of sequences coincides iff (c) X is finite dimensional.*

Theorem 15 *Every continuous convex function finite on X has f^{**} finite on X^{**}* **iff** *X is a* **Grothendiek space***: weak-star and weak convergence of sequences coincides (e.g., in reflexive space or ℓ^∞).*

Theorem 16 *Gateaux and Fréchet differentiability agree for convex functions on X* **iff** *X is a* **JN-space***.*

Theorem 17 *Weak Hadamard and Fréchet differentiability agree for convex functions on X* **iff** *X is a* **sequentially reflexive space***: $\ell^1 \not\subseteq X$ iff norm and Mackey convergence of sequences coincides.*

Proof of Theorem 14

[(a) implies (b)] Suppose $\{y_n\}$ is JN. Define

$$f(x) := \sum 2^n \psi(y_n(x))$$

where $\psi \geq 0$ is convex, continuous with $\psi(1) = 1$ and $\psi([0, 1/2]) = 0$.

Then f is continuous since the sum is locally finite, and unbounded on B_X since $f(y_n) = 1$.

[(b) implies (a)] if $f \geq 0$ is unbounded on B_X, so by the MVT, is ∂f. Thus, there is $x_n \in B_X$, $z_n \in \partial f(x_n)$ and $\|z_n\| \to \infty$. Then $y_n := z_n/\|z_n\|$ is JN. Indeed

$$\langle y_n, x\rangle \leq \langle y_n, x_n\rangle + \frac{f(x) - f(x_n)}{\|z_n\|} \to 0.$$

©

♠ There are many other such results (e.g., characterizing Schur spaces, reflexive spaces, strong separability etc).

Two Open Questions

- Any two real valued Lipschitz functions on Hilbert space are *simultaneously densely Fréchet differentiable*.

  ◊ True in the separable Gateaux case.

- A convex continuous function on separable Hilbert space admits a *second-order Gateaux expansion* densely.

  ◊ True in finite dimensions.

  ◊ False for Fréchet or nonseparable ℓ^2.