Dalhousie Distributed Research Institute and Virtual Environment

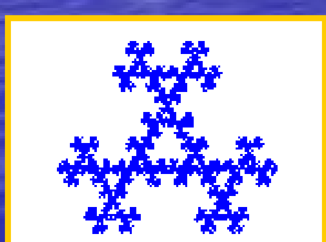
Spring School on Variational Analysis
(Paseky April 23-29, 2006)

Jonathan Borwein, FRSC **www.cs.dal.ca/~jborwein**

Canada Research Chair in Collaborative Technology

"Top mathematicians are becoming a new global elite. It's a force of barely 5,000, by some guesstimates, but every bit as powerful as the armies of Harvard University MBAs who shook up corner suites a generation ago."

Business Week Cover Story January 23, 2006



www.aarms.math.ca

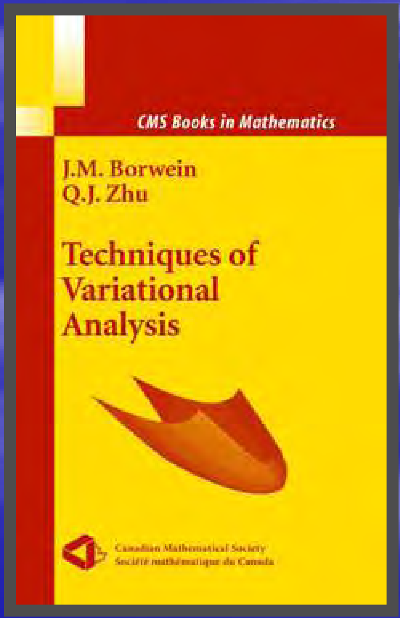
Revised 27/04/2006

DALHOUSIE UNIVERSITY *Inspiring Minds*

Four or Five Lectures on

Variational Principles and Convex Applications

based largely on



J M Borwein and Qiji Zhu

Techniques of Variational Analysis

CMS/Springer, 2005

http://users.cs.dal.ca/~jborwein/ToVA/

Atlantic Association for Research in the Mathematical Sciences

AARMS 2006 Summer School

1. Mark Bauer, University of Calgary: Elliptic Curve Cryptography.
2. Anthony Bonato, Wilfred Laurier University: Massive Networks and Internet Mathematics.
3. Rick Miranda, Colorado State University: Introduction to Algebraic Geometry.
4. Anita Tabacco, Polytecnico di Turino: Introduction to Wavelet Theory and Numerical applications.

July-August 2006 www.aarms.math.ca

And one presentation on

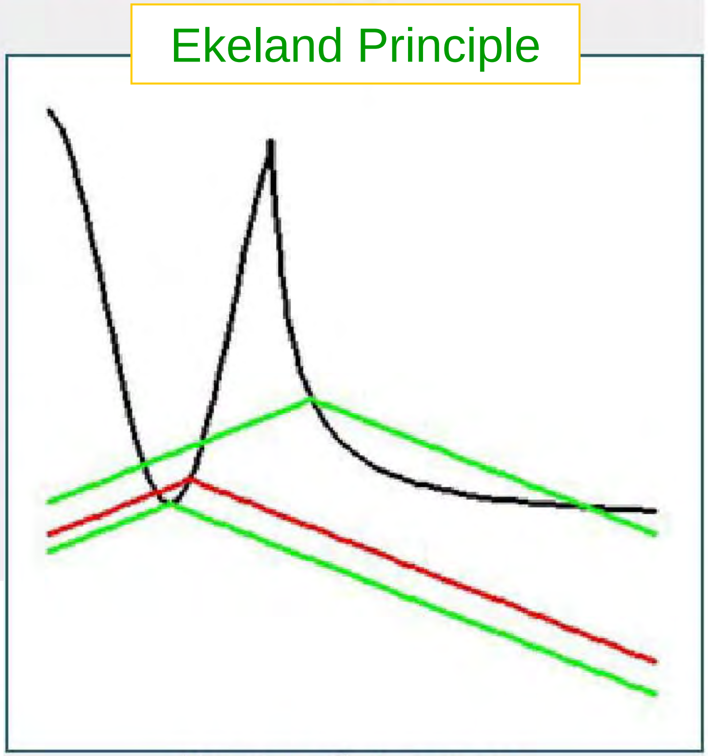
www.experimentalmath.info

http://users.cs.dal.ca/~jborwein/ToVA/

Techniques of Variational Analysis

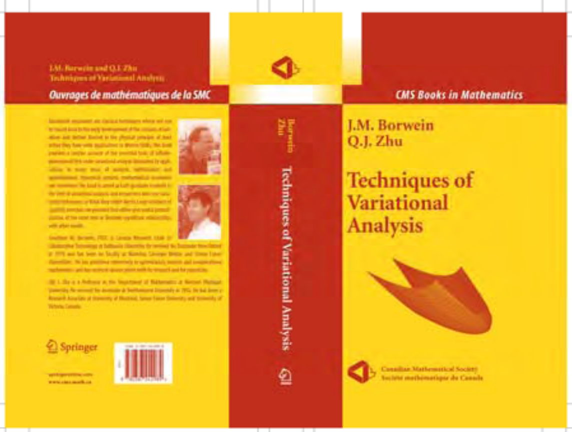
CMS | Sample | Order | Reviews | Addenda | Links | Contact | Springer

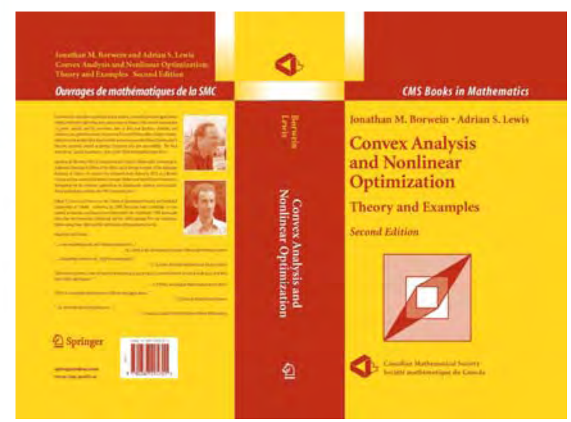
Jonathan Borwein
Dalhousie University



Now on a website near you.
A new CMS/Springer Book on
Techniques of Variational Analysis.
May 2005: intended for researchers, practioners, students in optimization, analysis and elsewhere.

Qiji (Jim) Zhu
Western Michigan University





Abstract I

- **Variational arguments** connote classical techniques whose use can be traced back to the early development of the calculus of variations and further. Rooted in the physical **principle of least action** they have wide applications in diverse fields.
- The discovery of modern variational principles and nonsmooth analysis further expands the range of applications of these techniques.
  - I anticipate a working knowledge of undergraduate analysis and the basic principles of functional analysis The recent monograph "Variational Analysis" by Rockafellar and Wets provides an authoritative account of variational analysis in finite dimensions
  - "Variational Analysis and Generalized Differentiation: I & II" by Boris Mordukhovich, is a comprehensive complement to the present text

Abstract II

- We shall start with an overview of "theory" in Lecture 1-2 and shall continue with concrete "applications" in Lectures 3-4 and 5.
  - the distinction is blurred
- As we proceed we shall see fewer broad results and more detailed proofs
  - full details of almost all results are in **ToVA** and **CaNo**

Rationale

- To talk about things I somewhat understand
- To complement my colleagues' lectures

- To revisit some hard old problems
- To show some very recent results
- To pose some open problems

Why Overheads ?

- I have pictures
- I can offer complete notes
- To complement my colleagues' lectures
- I have lousy blackboard style
- Since 2003 I work in a Computer Science Faculty

Atlantic Association for Research in the Mathematical Sciences

LECTURES I and II

Bumps, Cusps and Slices:

Functional-analytic Underpinnings of Variational Analysis

- a general tour

Atlantic Assocation for Research in the Mathematical Sciences

LECTURES II and III

The Fitzpatrick Function:

Monotone Operators as Convex Objects

- a detailed case study

Atlantic Assocation for Research in the Mathematical Sciences

LECTURE IV

Best Approximation

and

Chebysev Sets

- deep down and dirty

Slices, Bumps and Cusps:

Underpinnings of Nonsmooth Analysis

For Simon Fitzpatrick (1953—2004)

Dalhousie Distributed Research Institute and Virtual Environment

Jonathan M. Borwein, FRSC

Research Chair in IT
Dalhousie University

Halifax, Nova Scotia, Canada

Revised for Paseky, April 23, 2006

"I feel so strongly about the wrongness of reading a lecture that my language may seem immoderate. ··· The spoken word and the written word are quite different arts. ···
I feel that to collect an audience and then read one's material is like inviting a friend to go for a walk and asking him not to mind if you go alongside him in your car."

(Sir Lawrence Bragg)

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

Un sujet, un/deux langues, deux cultures

France

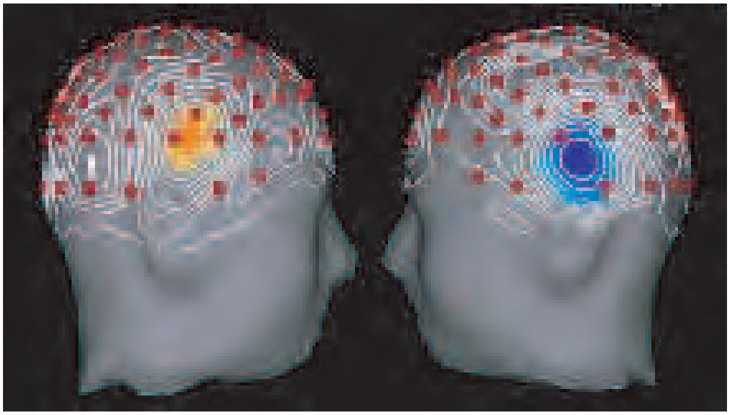
America

AS SMART AS HE WAS, ALBERT EINSTEIN COULD NOT FIGURE OUT HOW TO HANDLE THOSE TRICKY BOUNCES AT THIRD BASE.

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

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

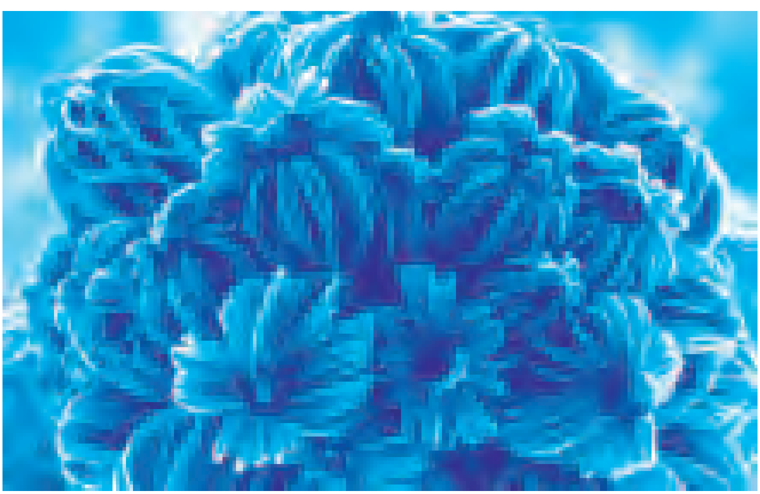
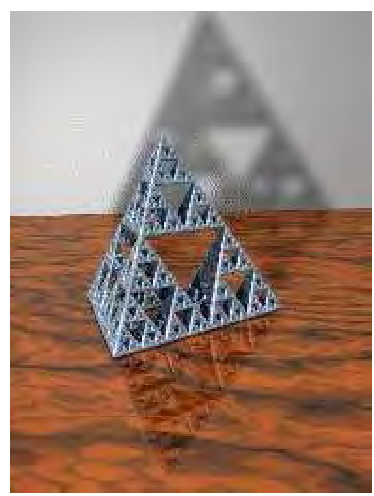
5. **Convex** functions and the Banach spaces they populate

• Full references will be found in

J.M. Borwein and Qiji (Jim) Zhu, *Techniques of Variational Analysis* CMS-Springer Books 2005.

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

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.

The Simpsons

Permission refused by Fox

TO: DAVID BAILEY
FROM: JACQUELINE ATKINS
DATE: 10/9/92
NUMBER OF PAGES: 1

FAX (310) 203-3852
PHONE (310) 203-3959

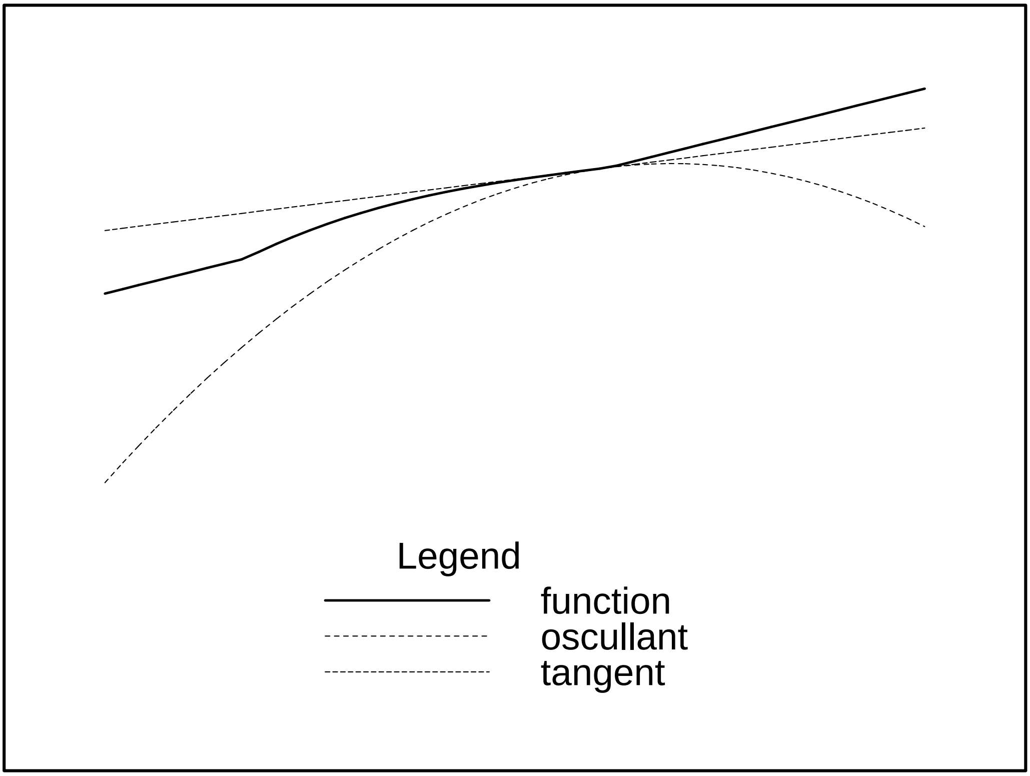
A Professor at UCLA told me that you might be able to give me the answer to: What is the 40,000th digit of Pi?

We would like to use the answer in our show. Can you help?

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"*

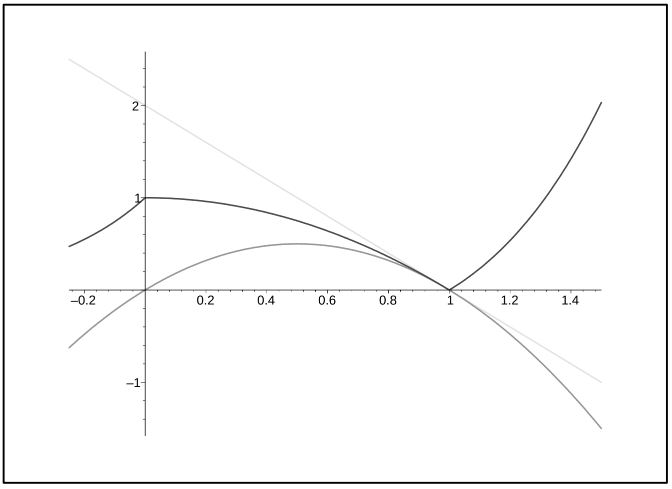
Viscosity is Fundamental

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,

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

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

FRANK & ERNEST

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*

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


Do not drop cigarette ends on the floor, as they burn the hands and knees of customers as they leave.
NOTICE-PUBLIC BAR
OUR PUBLIC BAR IS PRESENTLY NOT OPEN BECAUSE IT IS CLOSED. MANAGER

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.

GENERICALLY, TILT PERTURBATIONS ARE ATTAINED.

- 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*

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

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 $\operatorname{dom}\varphi$ *is relatively compact, the conclusion is true for any proper lsc f.*

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

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

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

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

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

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

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.

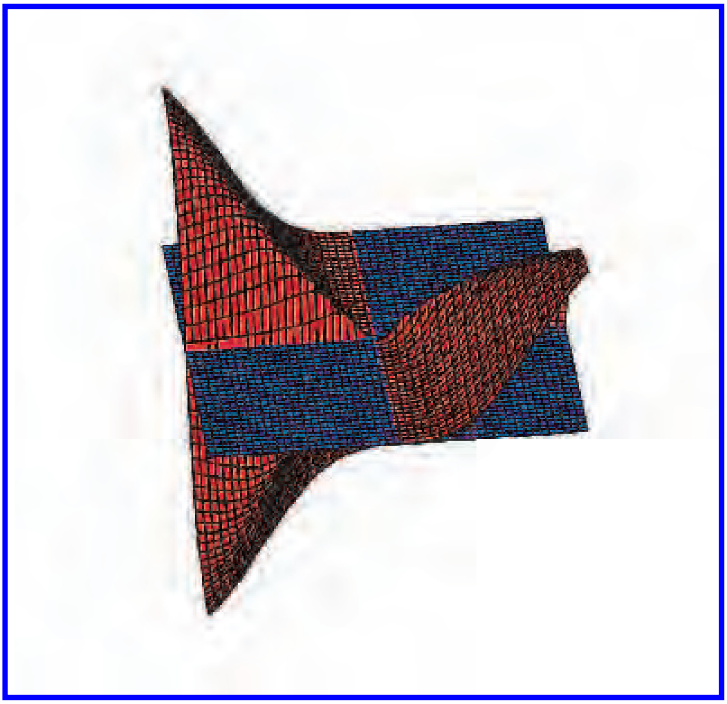
Two Open Questions

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 × Sep $\overset{?}{\subset}$ WASP.

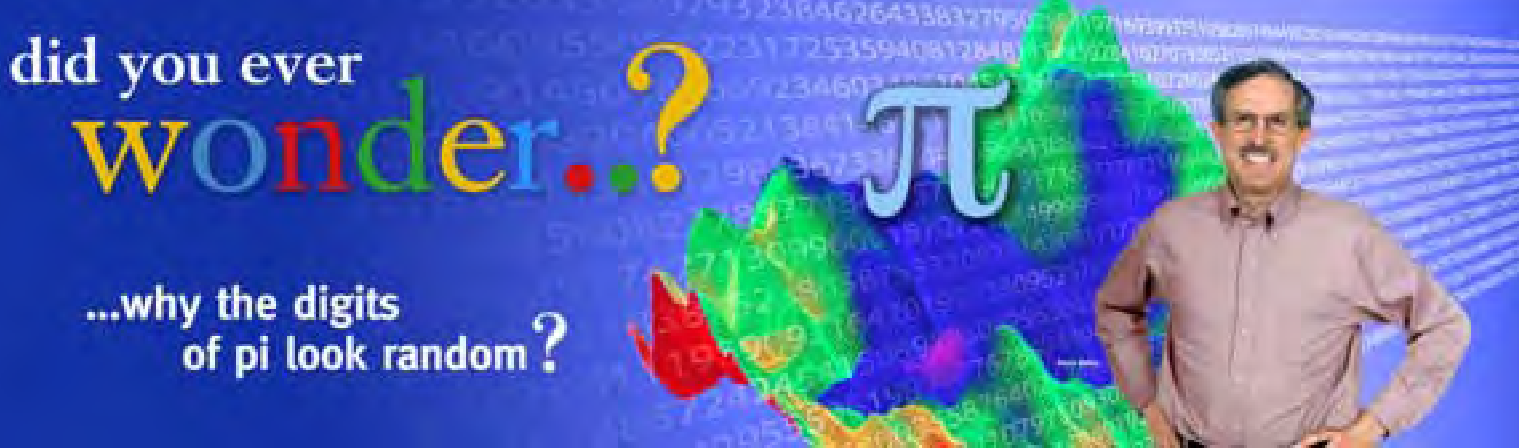
Star Trek

Kirk asks:

“ **Aren’t there some mathematical problems that simply can’t be solved**?”

And Spock ‘fries the brains’ of a rogue computer by telling it:

“ **Compute to the last digit the value of Pi**.”


did you ever
wonder...?
π
...why the digits
of pi look random?

Frank and Ernest

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.*

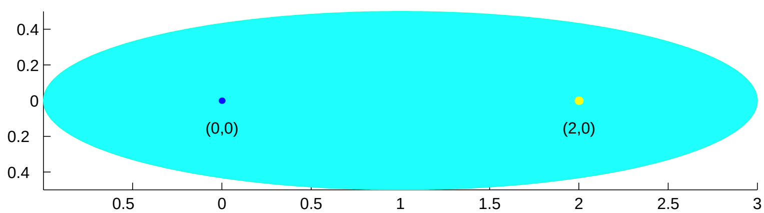
Step 1. The **support function** of a strictly convex body

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

leads to a bump

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

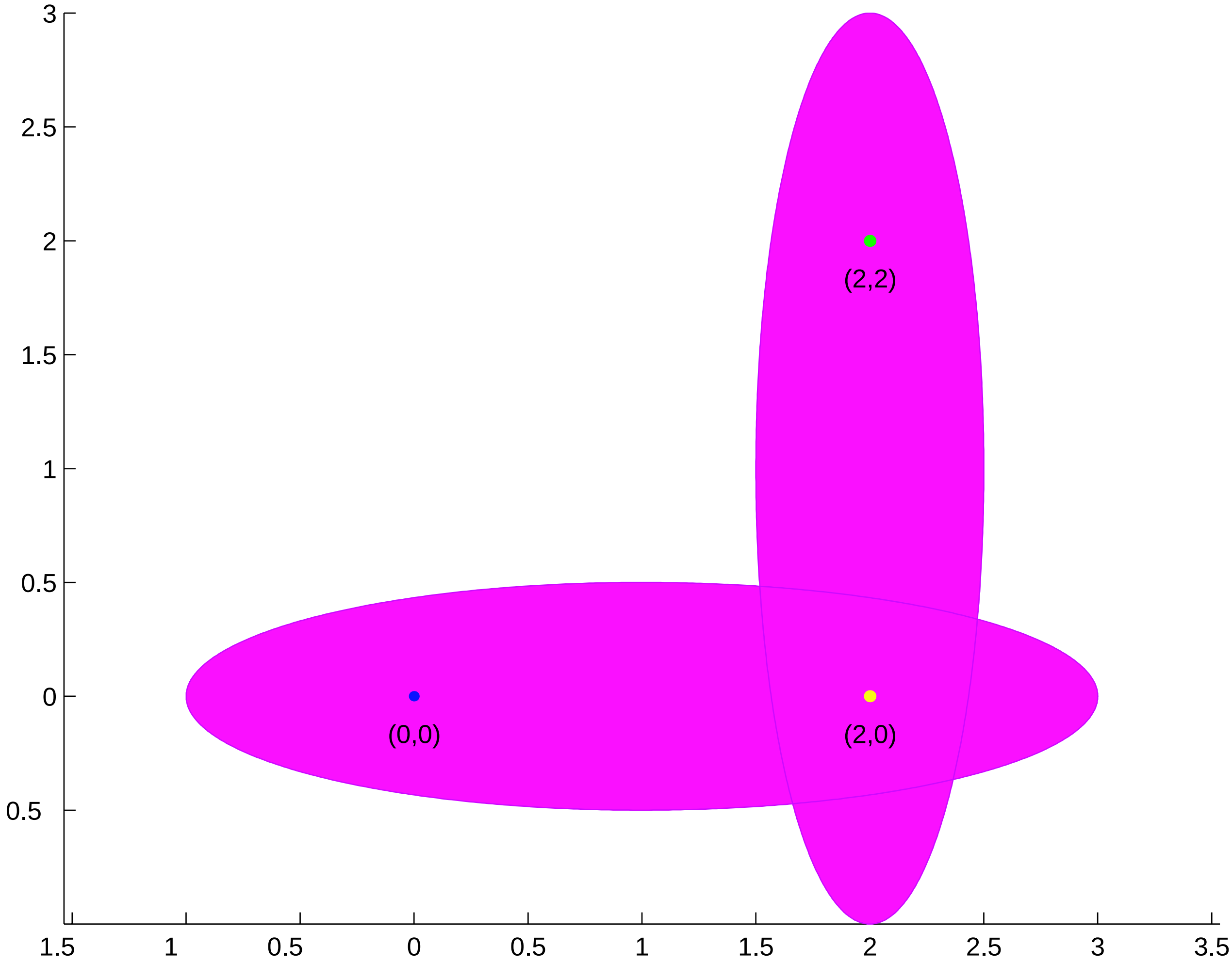
with range exactly C.



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

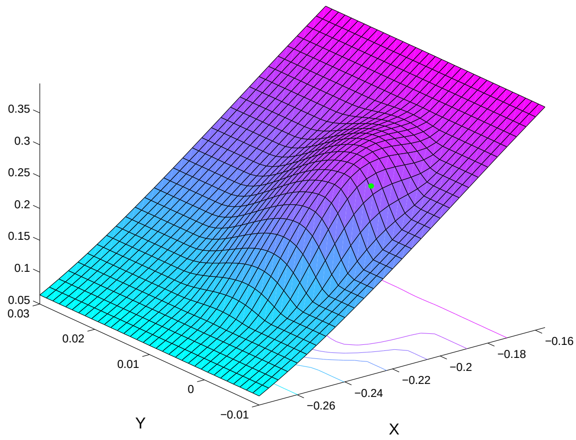
$$\sigma_E(y) = \langle Ax, x \rangle^{1/2}$$

Step 2. A disjoint sum then leads to

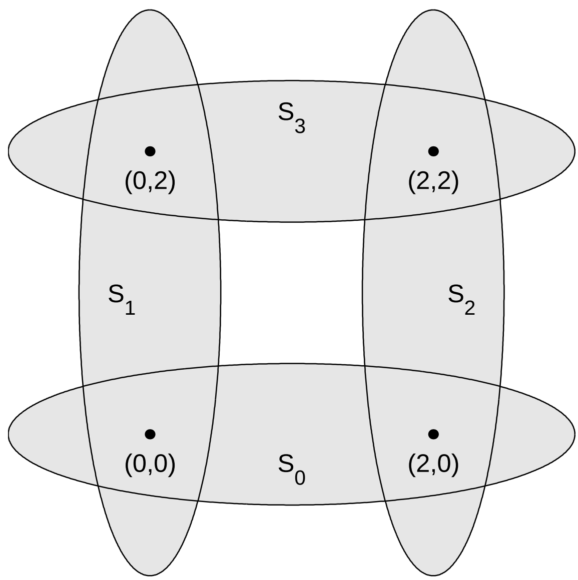


A Non-convex Gradient Range ∇b_C

Step 3. Build a flat patch on a bump range



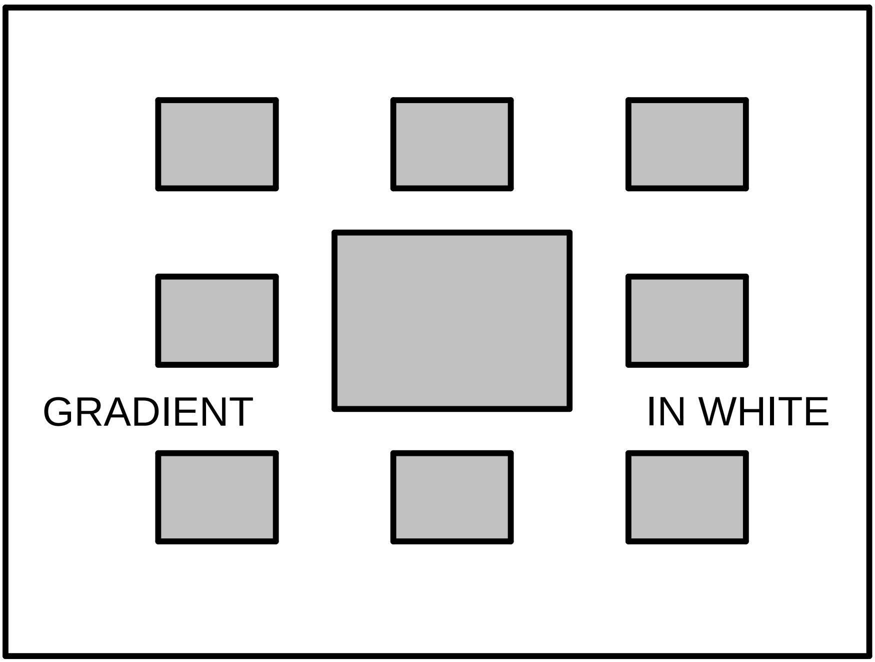
Step 4. Superposing a bump on a flat patch of another leads to



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).

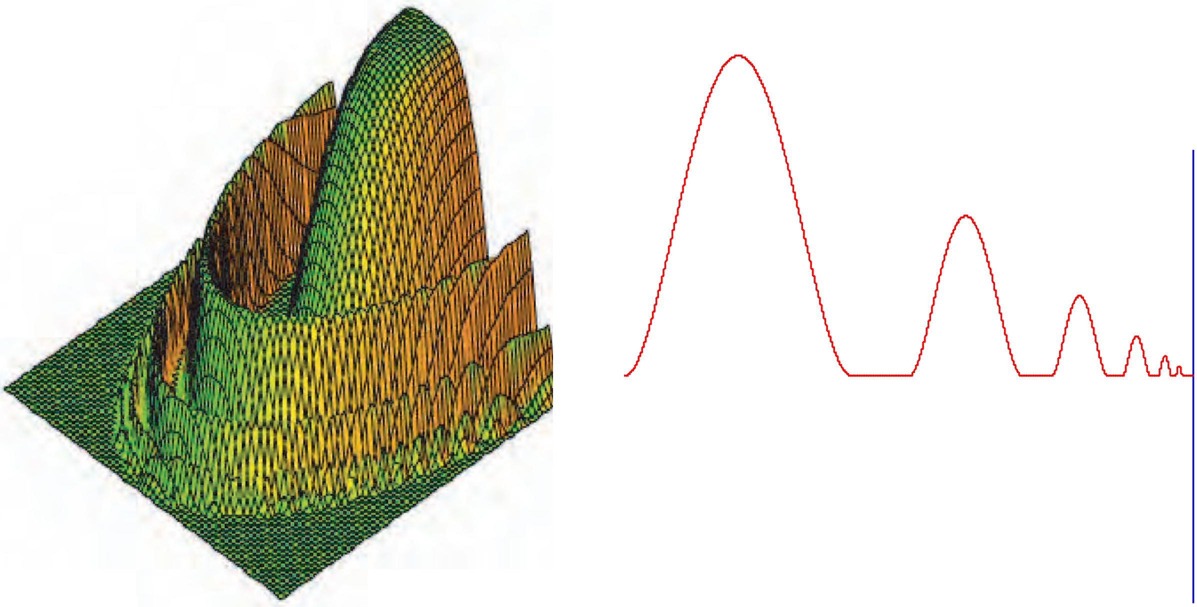


A Multiply Connected Gradient Range

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 $\partial_a f(x) \equiv B_2$?

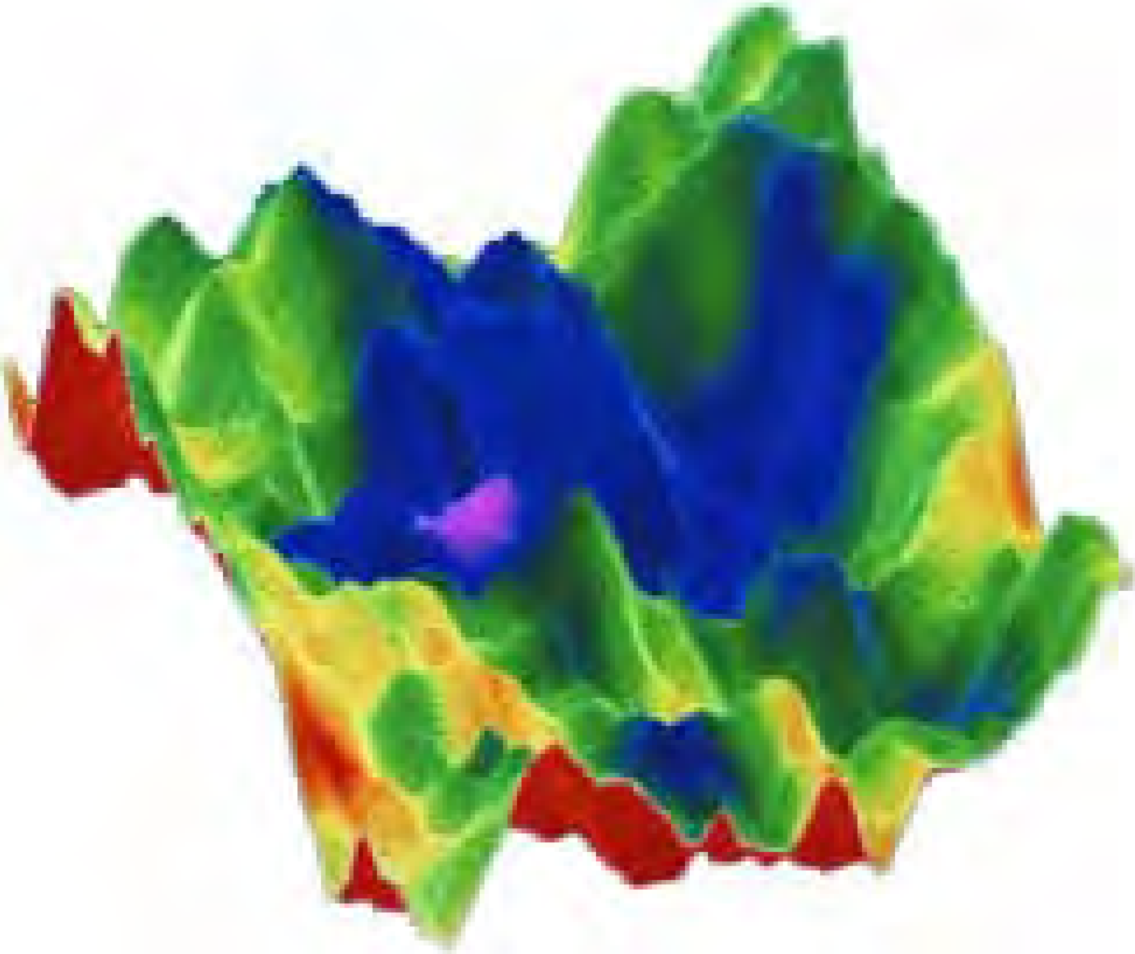
- Is it always true in $\mathbb{R}^N$ that the range of a C^1 bump's gradient is semi-closed:

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

  – 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.

The First Million Digits of π



- Pi as a random walk.

DERIVATIVES I: PROXIMALITY

- 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*

$$\boxed{\partial_F d_C(x) \subset \partial_F d_C(P_C(x))}$$

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.*

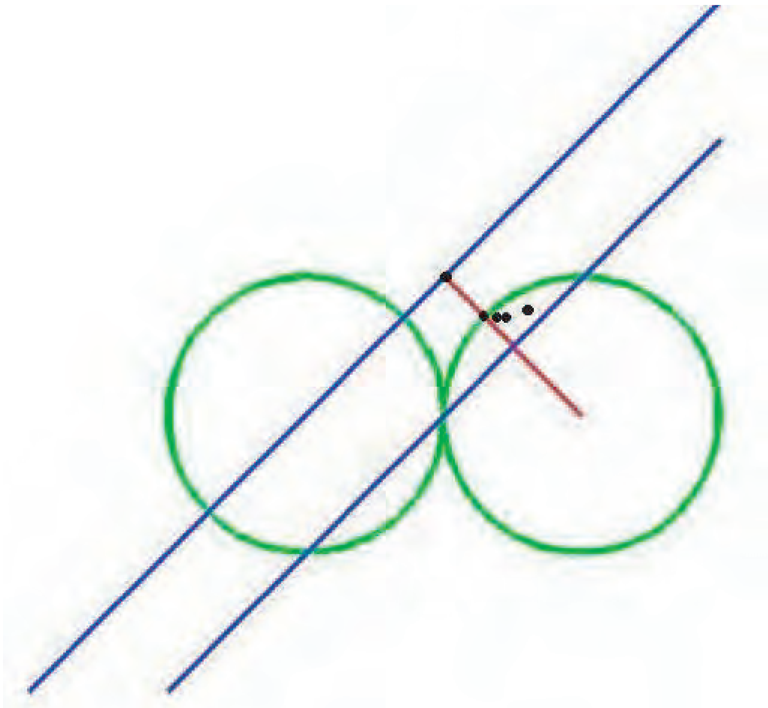
Proof. (a) We may assume $x_n \to_w p$ and at any of the dense set of points with

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

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

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

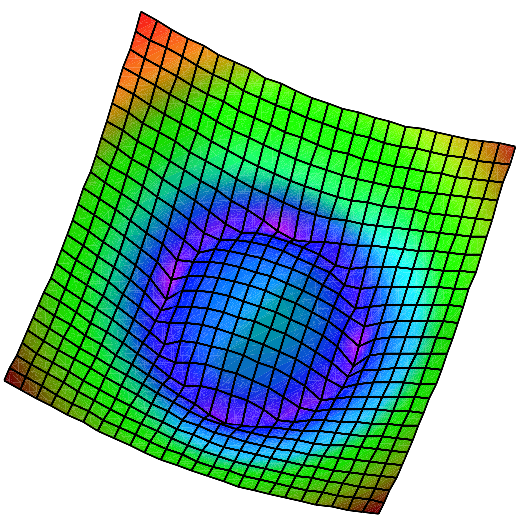
where ∂_π denotes the set of *proximal* subgradients.

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.]

Viete's formula

or **Vieta's formula,** *n.* the formula for π, derived from the infinite product for 2/π, namely

$$\sqrt{\frac{1}{2}} \times \sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}} \times \sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}} \times \cdots$$

published in 1593, and generally regarded as the first use of an infinite product. (Named after the French algebraist and geometer, *François Viete* or *Franciscus Vieta* (1540 - 1603), who introduced the use of literals to algebra, but rejected the existence of negative numbers. He made original contributions to trigonometry and the theory of equations, and decoded a complex code used by Philip II of Spain in his war against the French, being accused of witchcraft for his pains.)

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.*

DERIVATIVES II and CONVEXITY I

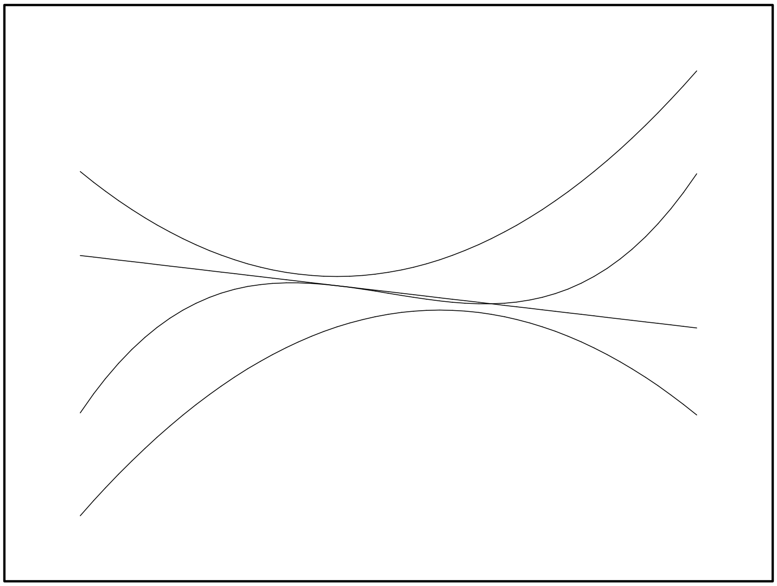
Duality Inequalities

- 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* $\mathrm{dom}\,(f) \cup \mathrm{dom}\,(h) \subset C$.

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* $\operatorname{dom}(f) \subset C$. *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 pleasant specializations follow.

Corollary 5 *Let* $C \subset \mathbb{R}^n$ *be compact convex and* f *proper convex lower semicontinuous with* $\operatorname{dom}(f) \subset C$. *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 proper convex lower semicontinuous 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*

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

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

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

✓ 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*

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

◇ 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*

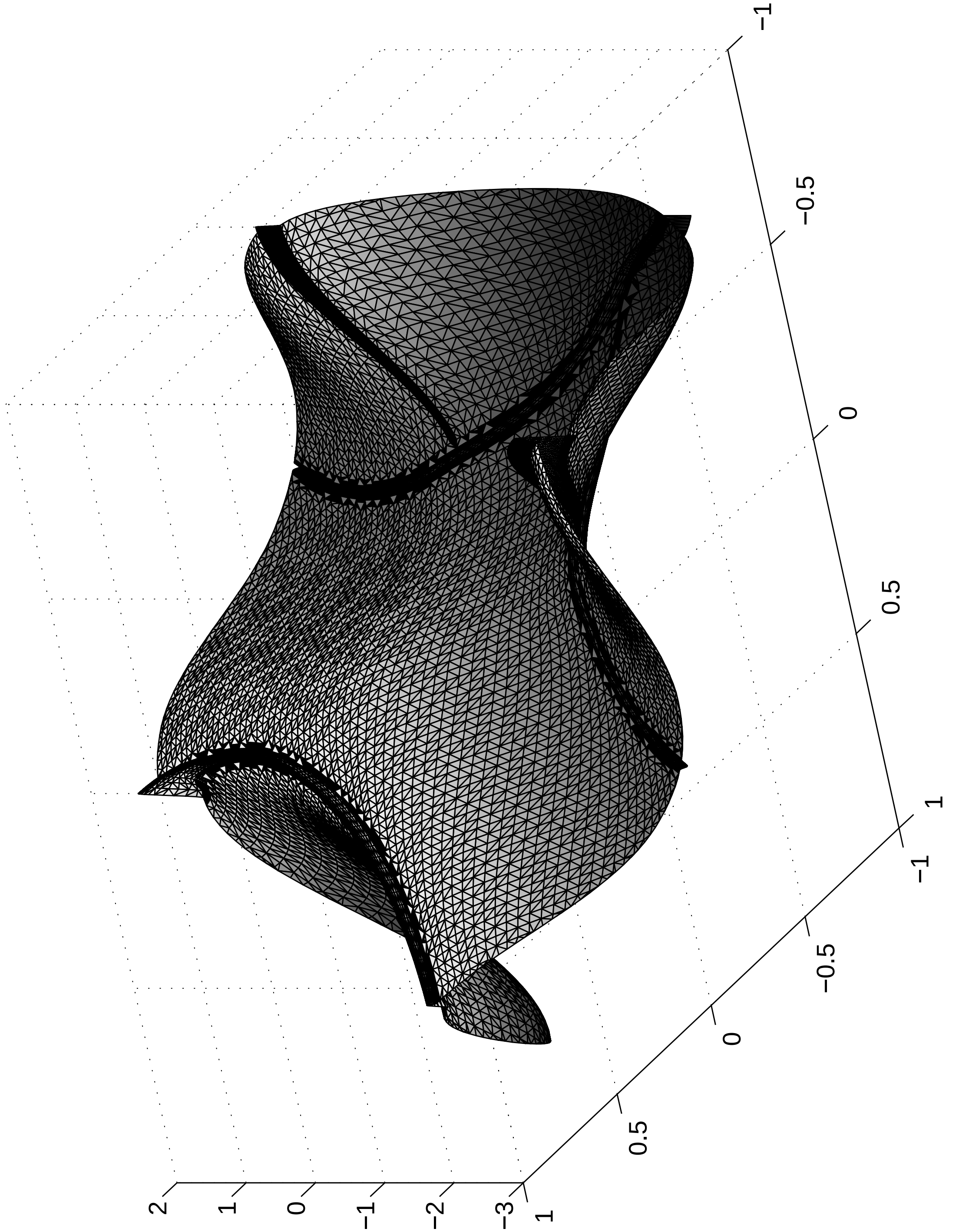
$$\|x^*\|_* \le \max_{a \in B} \frac{g(a) - g(-a)}{2}.$$

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

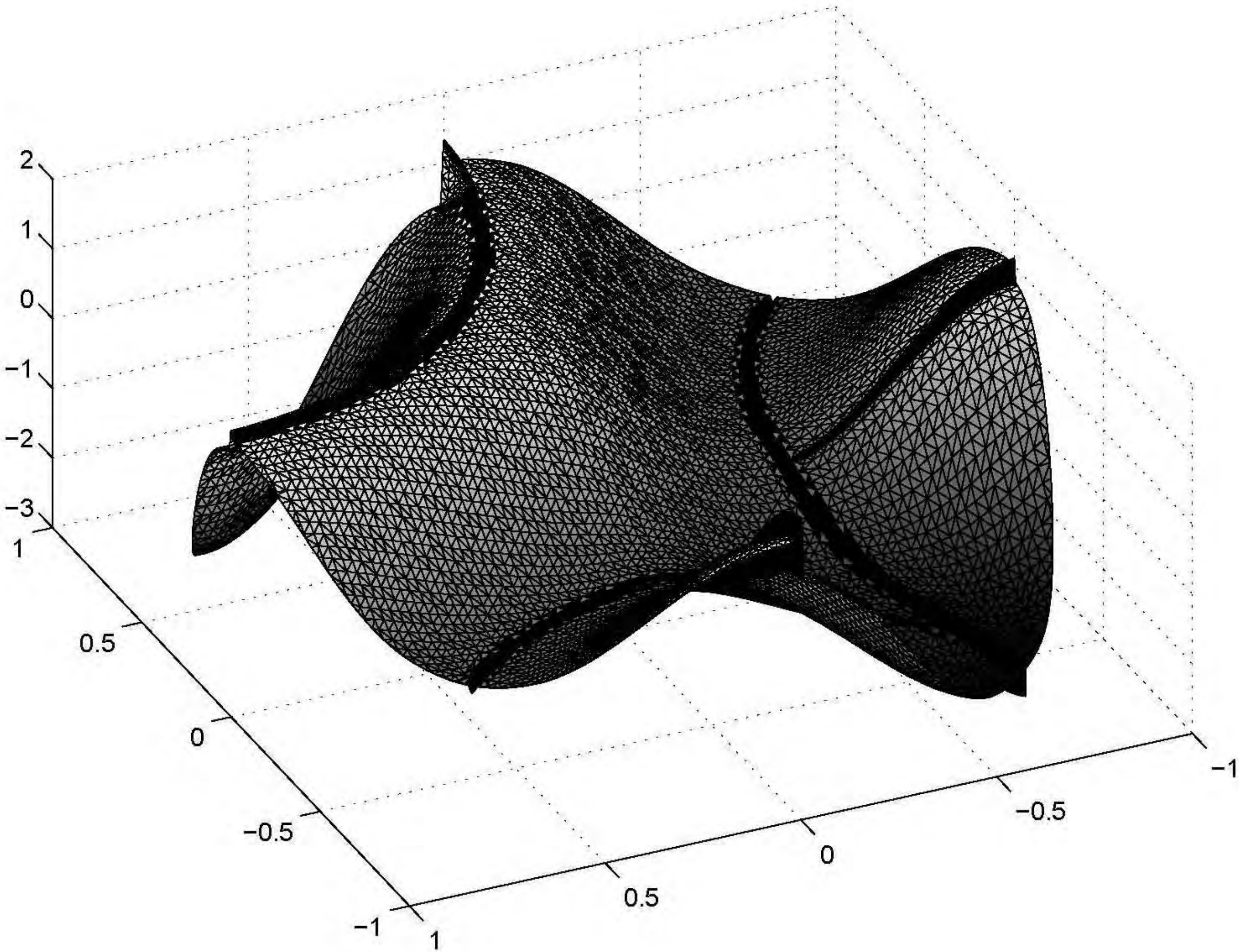
Remark 2 *Corollary 8 fails if B is replaced by the unit sphere S. Indeed, there is a C^1 mapping $f : B \subset \mathbb{R}^2 \to \mathbb{R}$ such that*

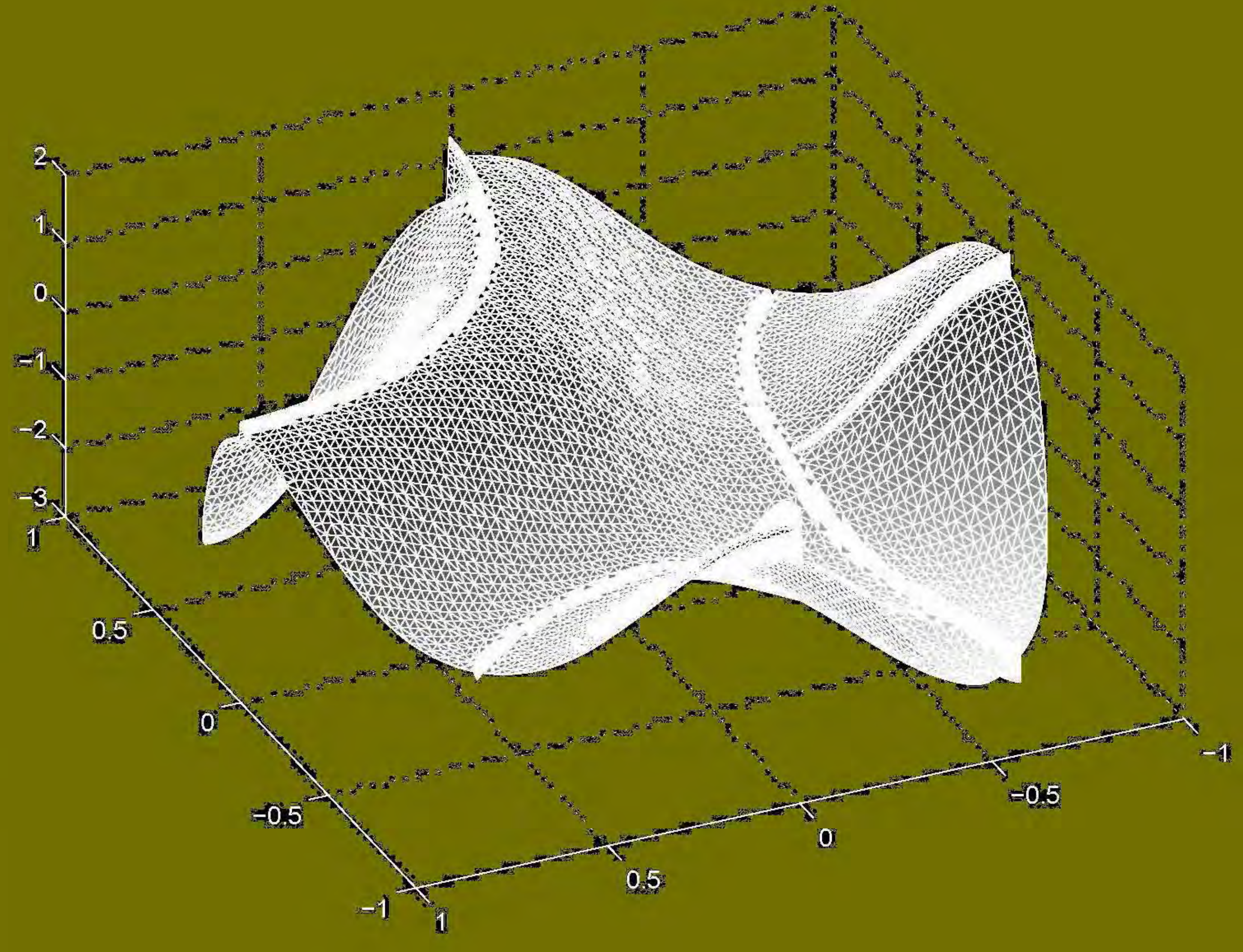
(i) $f|S$ *is even; but*

(ii) f *has no critical point in* B*.*

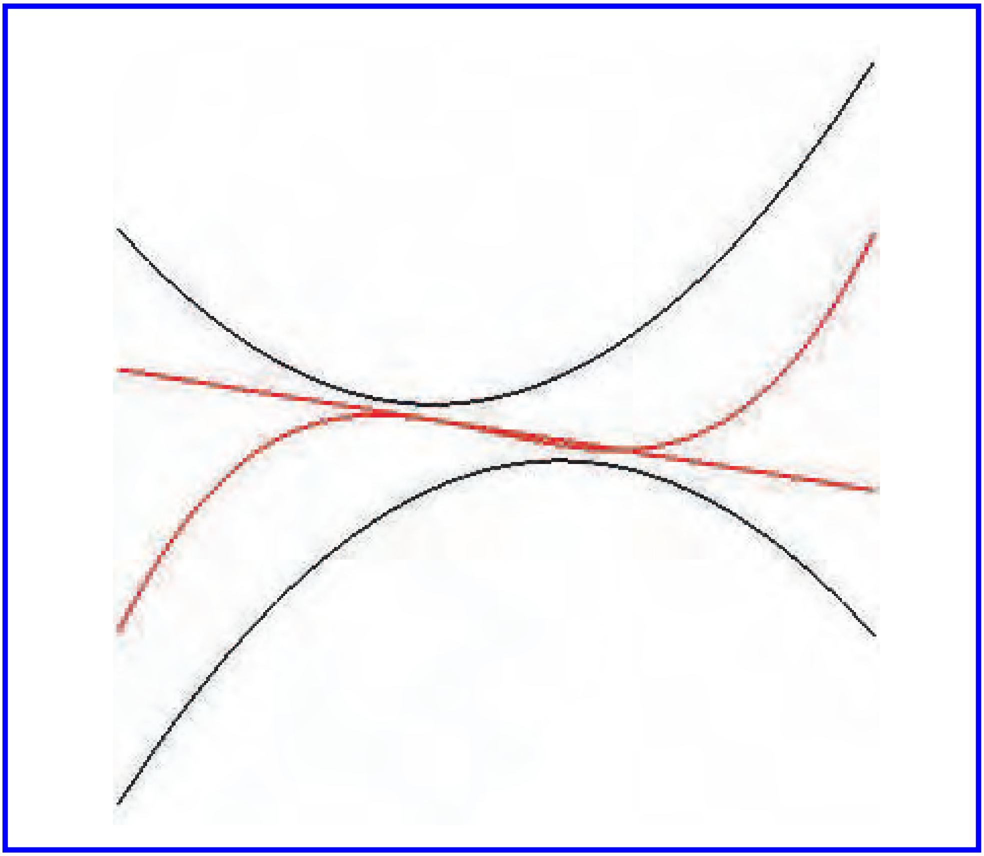


A Function Symmetric on S
With no Critical Point in B


2
1
0
−1
−2
−3
1
0.5
0
−0.5
−1
1
0.5
0
−0.5
−1


2
1
0
−1
−2
−3
1
0.5
0
−0.5
−1
1
0.5
0
−0.5
−1

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?


There's plenty of room for all God's creatures.
Right next to the mashed potatoes.
SASKATOON
STEAKS • FISH • WILD GAME
FAIRWAY

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

• many other similar results for reflexivity, Schur spaces, etc

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

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

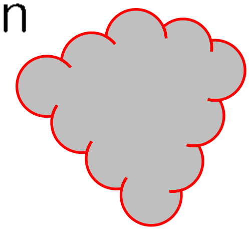
where $\psi \geq 0$ is convex, continuous with $\psi(1) = 1$ and $\psi([0, 1/2]) = 0$.

Then f is continuous since the sum is locally finite, and unbounded on B_X since $f(x_n) \geq 2^{n-1}$ for some $x_n \in B_X$

[(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

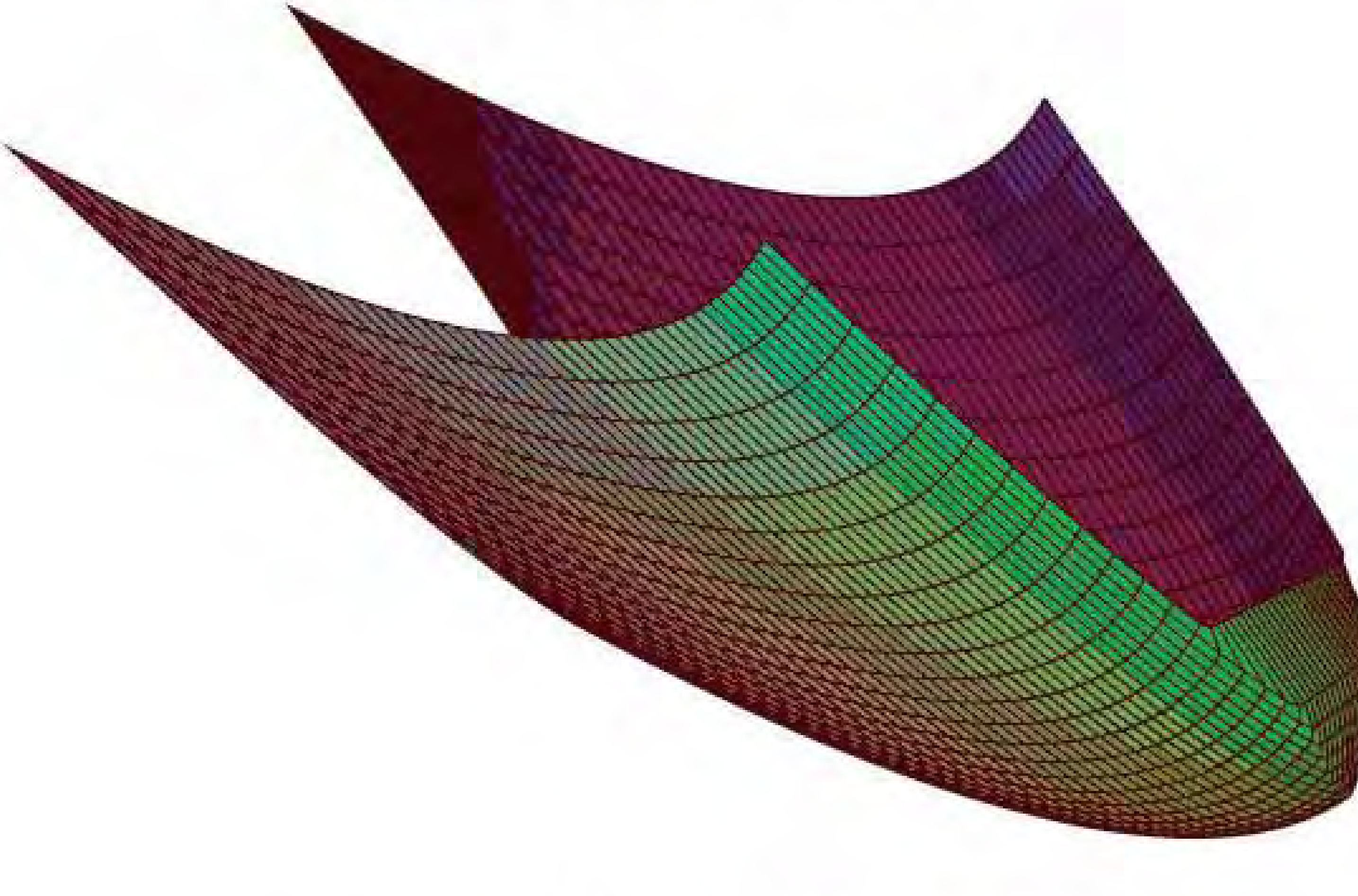
$$\langle y_n, x\rangle \leq \langle y_n, x_n\rangle + \frac{f(x) - f(x_n)}{\|z_n\|} \to 0.$$

Since the RHS < 1+ for all x in X.

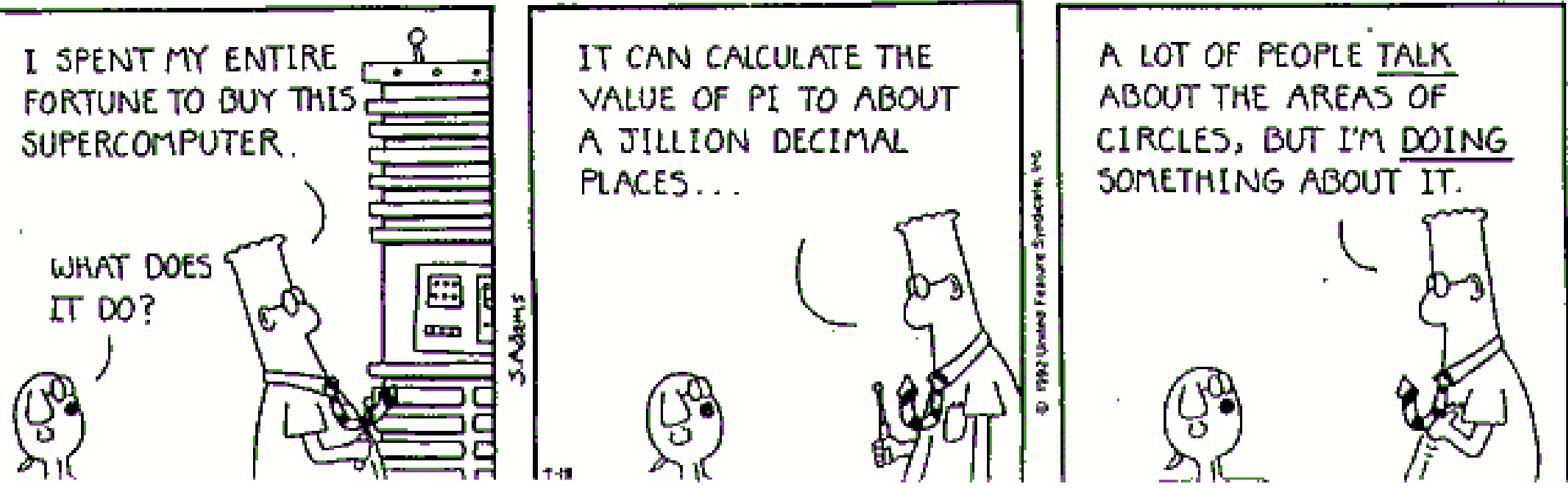


©

AN ESSENTIALLY STRICTLY CONVEX FUNCTION WITH
NONCONVEX SUBGRADIENT DOMAIN
AND WHICH IS NOT STRICTLY CONVEX



max{(x-2)^2+y^2-1,-(x*y)^(1/4)}


I SPENT MY ENTIRE FORTUNE TO BUY THIS SUPERCOMPUTER.
WHAT DOES IT DO?
S. Adams
IT CAN CALCULATE THE VALUE OF PI TO ABOUT A JILLION DECIMAL PLACES...
© 1992 United Feature Syndicate, Inc
A LOT OF PEOPLE TALK ABOUT THE AREAS OF CIRCLES, BUT I'M DOING SOMETHING ABOUT IT.

Two Open Questions

- Any two real valued Lipschitz functions on Hilbert space are *simultaneously densely Fréchet differentiable*. (L & P)

  ♢ 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.

TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 342, Number 1, March 1994

SECOND ORDER DIFFERENTIABILITY OF CONVEX FUNCTIONS IN BANACH SPACES

JONATHAN M. BORWEIN AND DOMINIKUS NOLL

ABSTRACT. We present a second order differentiability theory for convex functions on Banach spaces.

1. INTRODUCTION

The classical theorem of Alexandrov states that a convex function on $\mathbb{R}^n$ is almost everywhere second order differentiable. This was first proved by Busemann and Feller [12] for functions on $\mathbb{R}^2$ and later was extended by Alexandrov [2] to $\mathbb{R}^n$. More recent proofs were obtained by Mignot [26], Bangert [6], and Rockafellar [36].

Motivated by these infinite dimensional versions of Rademacher's theorem, the present work is to attack Alexandrov's theorem in infinite dimensions. As it turns out, the situation here is less promising than it is for Rademacher's theorem. For instance, Alexandrov's theorem fails in the spaces l_p, L_p, $1 \leq p < 2$, and much to our surprise, even in nonseparable Hilbert spaces. This leads us to focus on the case of separable Hilbert spaces. Here in fact, a positive solution seems possible. As one of our central results here, we in fact obtain a partial positive answer by proving a version of Alexandrov's theorem for convex integral functionals.

Seemingly, the third of the classical results of measure theoretic geometry, the theorem of Sard, allows extensions to infinite dimensions only under comparatively strong hypotheses (see [1, 10]). In the light of our present investigation, this is explained to some extent by the fact that there is a strong link between Alexandrov's theorem and a version of Sard's theorem for monotone operators

We now provide examples showing that strong second order differentiability and second order differentiability are nonequivalent in infinite dimensions.

Example 1. Let C be a closed convex set in Hilbert space H, and let $P_C : H \to C$ be the metric projection onto C, i.e., the nearest point mapping. Then P_C is known to be the Fréchet derivative of a continuous convex function f on H, i.e., $P_C = \nabla^F f$, where

$$f(x) = \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}\|x - P_C x\|^2 \tag{3.10}$$

(see [18] for details). As P_C is a Lipschitz operator, it is almost everywhere Gâteaux differentiable in the sense of Aronszajn [3] (see §4) when H is assumed separable. Due to Theorem 3.1(1), this means that, on a separable H, f is almost everywhere second order differentiable. However, even in a separable Hilbert space, the set C may be chosen so that P_C is nowhere (norm) Fréchet differentiable. By Theorem 3.1(2), f is then nowhere strongly second order differentiable. We take $H = L^2[0, 1]$, and let $C = \{f \in H : |f| \leq 1 \text{ a.e.}\}$. Then, according to [18, §5], P_C is nowhere Fréchet differentiable. This shows that Alexandrov's theorem fails even in separable Hilbert space when based on strong second order differentiability. A similar example would be obtained by taking $H = l_2$, C the positive cone in H (see [18, §5]).

Example 2 (Example 1 continued). The situation is even worse in nonseparable Hilbert space. Here the set C may be chosen so that P_C is nowhere Gâteaux differentiable. So here, by Theorem 3.1(1), f is nowhere second order differentiable, i.e., $D_f^2 = \varnothing$. Take $H = l_2(\Gamma)$ with $|\Gamma| > \aleph_0$, and let C be the positive cone in H. Then P_C is nowhere Gâteaux differentiable. This shows that there is no chance for a version of Alexandrov's theorem in nonseparable Hilbert space.

Example 3. A different type of counterexample is obtained by considering convex functions f on l_2 of the form

$$f(x) = \sum_{n=1}^{\infty} f_n(x_n), \qquad x = (x_n),$$

with appropriate convex functions f_n defined on the real line. Here f is Gâteaux differentiable at $x = (x_n)$ if and only if $f_n'(x_n)$ exists for every n. A necessary condition for $x \in D_f^2$ is the following: $f_n''(x_n)$ exists for every n and the sequence is bounded. However, this is not sufficient to guarantee $x \in D_f^2$, as shown in Example 2 in §6 by specifying the function f. Now one may find f such that $\nabla^F f = T : l_2 \to l_2$ is even a Lipschitz operator having no Fréchet differentiability point at all, while, by Aronszajn's result [3], T is almost everywhere Gâteaux differentiable. An explicit example of such T is [3, §3, Example I], with the corresponding convex f being easily supplemented.


Attention
Dog Guardians
Pick up after your
dogs. Thank you.
Attention Dogs
Grrrrr, bark, woof.
Good dog.
District of North Vancouver.
Bylaw 5981-11(i)

MONTY


SO, UH, WHY IS YOUR NAME "MR. PI"? DO YOU LIKE PIES OR SOMETHING?
THE PRIVY COUNCIL ON RIGEL-9 GAVE ME THIS NAME BEFORE EXILING ME FROM THE PLANET...
www.comics.com
MOST RIGELIANS HAVE RATIONAL NUMBERS FOR NAMES LIKE "14" OR "286733", BUT BECAUSE I'M HALF HUMAN AND TAINTED WITH HUMAN BLOOD THEY CHOSE TO LABEL ME WITH THE IRRATIONAL NUMBER "PI".
6/8
OK. WAIT. WAIT. WAIT.
©2004 Jim Meddick /Dist. by NEA, Inc.
"PI" IS A NUMBER?
I THOUGHT HUMANS WERE REQUIRED TO LEARN MATHEMATICS IN SCHOOL.
SOME DO. SOME BUY CALCULATORS.

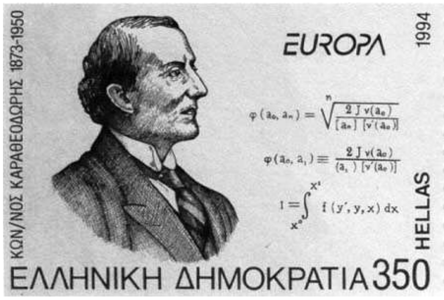
Monotone Operators as Convex Objects

Jonathan M. Borwein, FRSC
Research Chair in IT
Dalhousie University
Halifax, Nova Scotia, Canada

Fitzpatrick Memorial Workshop

Perth, September 25–26, 2005

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. (MAA 1936)

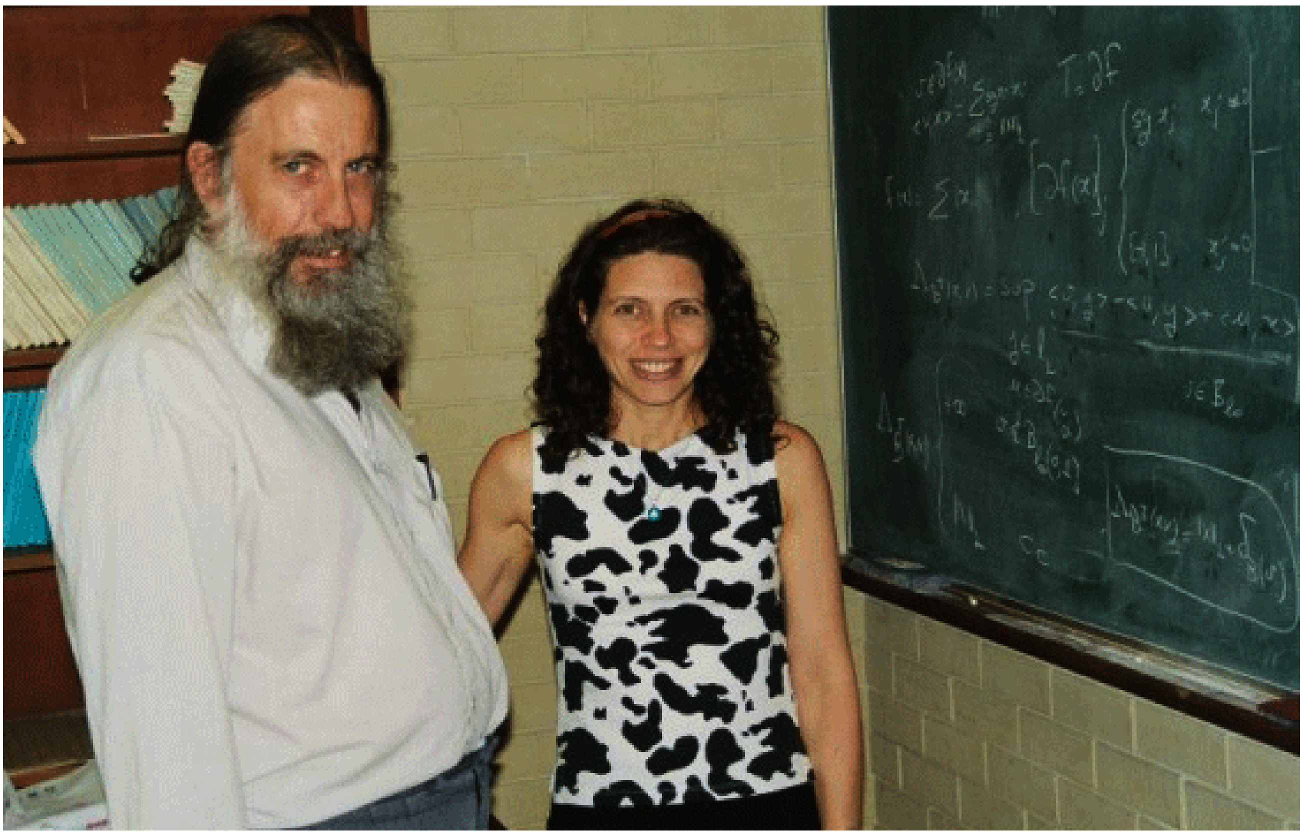


Constantin Carathéodory

In Memoriam

In his '23' "*Mathematische Probleme*" lecture to the Paris ICM in 1900*, David Hilbert wrote

> "*Besides it is an error to believe that rigor in the proof is the enemy of simplicity.*"



Simon Fitzpatrick† (1953–2004).

*See Ben Yandell's fine account of the *Hilbert Problems* and their solvers in *The Honors Class*, AK Peters, 2002. (He also died young in 2004.)

†At his blackboard with Regina Burachik

MOTIVATION and GOALS

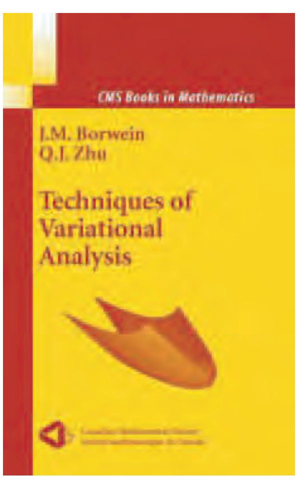
To reduce as much of monotone operator theory as possible to (elementary) convex analysis

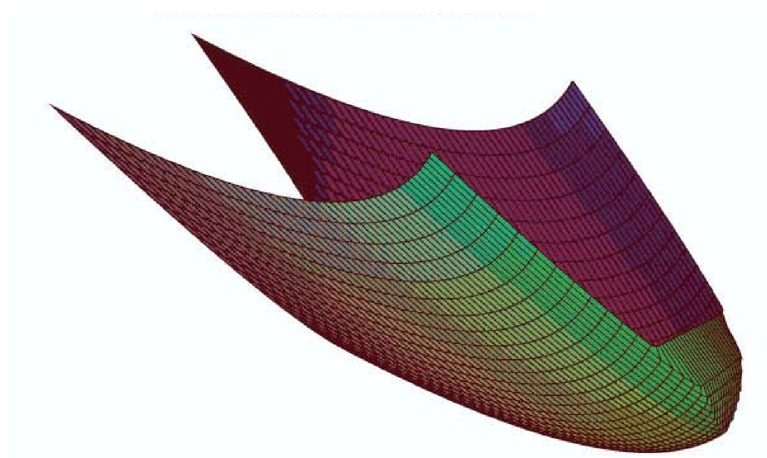
To thereby illustrate (some of) Simon Fitzpatrick's many fine contributions

To shed new light on the remaining open questions (in non-reflexive space)

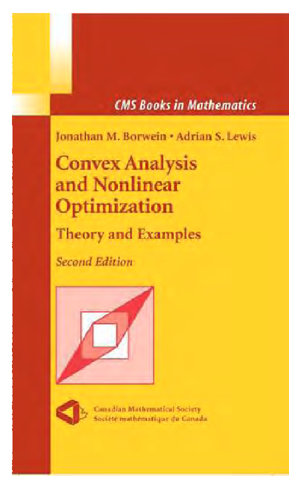
★ "Even convex objects are hard ..." ★

An essentially strictly convex function with non-convex subgradient domain and not strictly convex:





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



◀ JMB & J Zhu (Springer, 2005) JMB & A Lewis ▶

Most details will appear in: J.M. Borwein
Maximal Monotonicity via Convex Analysis
Fitzpatrick Memorial, *JCA*, **13–14**, 2006.

- http://users.cs.dal.ca/∼jborwein/mon-jca2.pdf

Coxeter's favourite 4-D polytope
(with 120 dodecahedronal faces)


CAUTION
THIS SIGN HAS
SHARP EDGES
DO NOT TOUCH THE EDGES OF THIS SIGN
ALSO, THE BRIDGE IS OUT AHEAD

1. Preliminaries

Throughout X is a real Banach space. The *domain* of an extended valued convex function, dom (f), is the set of values less than $+\infty$. A point s is in the *core* of a set S ($s \in$ core S) when $X = \bigcup_{\lambda>0} \lambda(S-s)$.

Now $x^* \in X^*$ is a *subgradient* of $f : X \to (-\infty, +\infty]$ at $x \in$ dom f provided that

$$f(y) - f(x) \geq \langle x^*, y - x \rangle$$

for all y in Y. The set of all subgradients of f at x is the *subdifferential* of f at x, denoted $\partial f(x)$.

We need the *indicator function* $\iota_C(x)$ which is zero for x in C and $+\infty$ otherwise, the *Fenchel conjugate* $f^*(x^*) := \sup_x\{\langle x, x^*\rangle - f(x)\}$ and the *infimal convolution*

$$f^* \Box \frac{1}{2}\|\cdot\|_*^2(x^*) := \inf\left\{f^*(y^*) + \frac{1}{2}\|z^*\|_*^2 : x^* = y^* + z^*\right\}$$

When f is convex and closed

$$x^* \in \partial f(x) \text{ exactly when } f(x) + f^*(x^*) = \langle x, x^*\rangle.$$

Finally, the *distance function* associated with a closed set C, given by $d_C(x) := \inf_{c \in C}\|x - c\|$, is convex if and only if C is. Moreover, $d_C = \iota_C \Box \|\cdot\|$.

We say $T : X \mapsto 2^{X^*}$ is *monotone* provided that for any $x, y \in X$, and $x^* \in T(x), y^* \in T(y)$,

$$\boxed{\langle y - x, y^* - x^* \rangle \geq 0,}$$

and that T is *maximal monotone* if its graph is not properly included in any other monotone graph.

- The *convex subdifferential* in Banach space* and a *skew linear matrix* are the canonical examples of maximal monotone multifunctions

We save the notation $J = J_X$ for the *duality map*

$$J_X(x) = \frac{1}{2}\partial\|x\|^2 = \left\{x^* \in X^* : \|x\|^2 = \|x^*\|^2 = \langle x, x^* \rangle\right\}$$

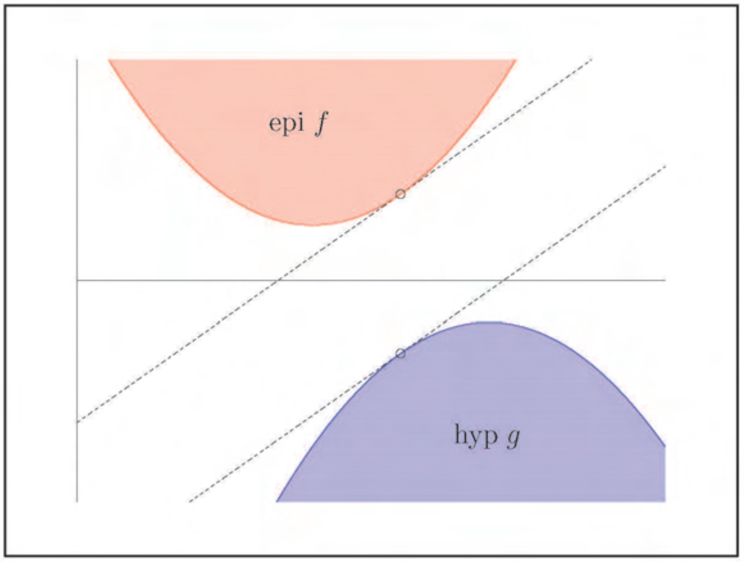
- It is not an exaggeration to say *the geometry of Banach space devolves to a deep study of J*

- The other foundational example is that of a second order nonlinear *elliptic PDE*

*There are several nice variational proofs. One based on the Mean value theorem follows.

Outline

Our goal is to derive *all* key results about maximal monotone operators *entirely from the existence of subgradients* and *Sandwich theorem* shown below



Section 2 considers general Banach spaces

Section 3 looks at (a-)cyclic operators

Section 4 presents our central result on maximality of the sum in reflexive space

Section 5 looks at more applications of the technique of Section 4

Section 6 provides limiting counter-examples.

on any convex su[...]

The existence of Subgradients on Three Slides

The following [...]e fundamental significance of subgrad[...]

Proposition 3.1.5 (Subgradients at optimality) *For any proper function $f : \mathbf{E} \to (\infty, +\infty]$, the point $\bar{x}$ is a (global) minimizer of f if and only if the condition $0 \in \partial f(\bar{x})$ holds.*

Alternatively put, minimizers of f correspond exactly to "zeroes" of ∂f.

The derivative is a local property whereas the subgradient definition (3.1.4) describes a global property. The main result of this section shows that the set of subgradients of a convex function is usually *nonempty*, and that we can describe it locally in terms of the directional derivative. We begin with another simple exercise.

Proposition 3.1.6 (Subgradients and directional derivatives) *If the function $f : \mathbf{E} \to (\infty, +\infty]$ is convex and the point $\bar{x}$ lies in* $\operatorname{dom} f$, *then an element ϕ of $\mathbf{E}$ is a subgradient of f at $\bar{x}$ if and only if it satisfies $\langle \phi, \cdot \rangle \leq f'(\bar{x}; \cdot)$.*

The idea behind the construction of a subgradient for a function f that we present here is rather simple. We recursively construct a decreasing sequence of sublinear functions which, after translation, minorize f. At each step we guarantee one extra direction of linearity. The basic step is summarized in the following exercise.

Lemma 3.1.7 *Suppose that the function $p : \mathbf{E} \to (\infty, +\infty]$ is sublinear and that the point $\bar{x}$ lies in* $\operatorname{core}(\operatorname{dom} p)$. *Then the function $q(\cdot) = p'(\bar{x}; \cdot)$ satisfies the conditions* (*i*) $q(\lambda\bar{x}) = \lambda p(\bar{x})$ *for all real* λ,

(*ii*) $q \le p$, *and*

(*iii*) $\operatorname{lin} q \supset \operatorname{lin} p + \operatorname{span}\{\bar{x}\}$.

Recall that
p is sublinear
and q(h)=p'(x;h)

With this tool we are now ready for the main result, which gives conditions guaranteeing the existence of a subgradient. Proposition 3.1.6 showed how to identify subgradients from directional derivatives; this next result shows how to move in the reverse direction.

Theorem 3.1.8 (Max formula) *If the function* $f : \mathbf{E} \to (\infty, +\infty]$ *is convex then any point* $\bar{x}$ *in* $\operatorname{core}(\operatorname{dom} f)$ *and any direction* d *in* $\mathbf{E}$ *satisfy*

$$f'(\bar{x}; d) = \max\{\langle \phi, d\rangle \mid \phi \in \partial f(\bar{x})\}. \tag{3.1.9}$$

In particular, the subdifferential $\partial f(\bar{x})$ *is nonempty.*

Proof. In view of Proposition 3.1.6, we simply have to show that for any fixed d in $\mathbf{E}$ there is a subgradient ϕ satisfying $\langle \phi, d\rangle = f'(\bar{x}; d)$. Choose a basis $\{e_1, e_2, \ldots, e_n\}$ for $\mathbf{E}$ with $e_1 = d$ if d is nonzero. Now define a sequence of functions $p_0, p_1, \ldots, p_n$ recursively by $p_0(\cdot) = f'(\bar{x}; \cdot)$, and $p_k(\cdot) = p'_{k-1}(e_k; \cdot)$ for $k = 1, 2, \ldots, n$. We essentially show that $p_n(\cdot)$ is the required subgradient.

First note that, by Proposition 3.1.2, each p_k is everywhere finite and sublinear. By part (iii) of Lemma 3.1.7 we know

$$\operatorname{lin} p_k \supset \operatorname{lin} p_{k-1} + \operatorname{span}\{e_k\} \quad \text{for } k = 1, 2, \ldots, n,$$

so p_n is linear. Thus there is an element ϕ of $\mathbf{E}$ satisfying $\langle \phi, \cdot \rangle = p_n(\cdot)$.

Part (ii) of Lemma 3.1.7 implies $p_n \le p_{n-1} \le \ldots \le p_0$, so certainly, by Proposition 3.1.6, any point x in $\mathbf{E}$ satisfies

$$p_n(x - \bar{x}) \le p_0(x - \bar{x}) = f'(\bar{x}; x - \bar{x}) \le f(x) - f(\bar{x}).$$

Thus ϕ is a subgradient. If d is zero then we have $p_n(0) = 0 = f'(\bar{x}; 0)$. Finally, if d is nonzero then by part (i) of Lemma 3.1.7 we see

$$\begin{aligned} p_n(d) \le p_0(d) = p_0(e_1) = -p_0'(e_1; -e_1) = \\ -p_1(-e_1) = -p_1(-d) \le -p_n(-d) = p_n(d), \end{aligned}$$

whence $p_n(d) = p_0(d) = f'(\bar{x}; d)$. □

Corollary 3.1.10 (Differentiability of convex functions) *Suppose the function $f : \mathbf{E} \to (\infty, +\infty]$ is convex and the point $\bar{x}$ lies in* core (dom f)*. Then f is Gâteaux differentiable at $\bar{x}$ exactly when f has a unique subgradient at $\bar{x}$ (in which case this subgradient is the derivative).*

2. Maximality in General Banach Space

For a monotone mapping T, we associate the *Fitzpatrick function* introduced in 1988 by Fitzpatrick. It is

$$\mathcal{F}_T(x, x^*) := \sup\{\langle x, y^*\rangle + \langle x^*, y\rangle - \langle y, y^*\rangle : y^* \in T(y)\}$$

which is clearly *lower semicontinuous and convex* as an affine supremum. Moreover,

Proposition 1 (Fitzpatrick) *For every maximal monotone operator T one has*

$$\mathcal{F}_T(x, x^*) \geq \langle x, x^*\rangle$$

with equality if and only if $x^* \in T(x)$.

- The equality $\mathcal{F}_T(x, x^*) = \langle x, x^*\rangle$ for $x^* \in T(x)$ requires only monotonicity not maximality.

- The idea of associating a convex function to a monotone operator and exploiting the relationship was neglected for many years after its introduction until revisited by Penot, Simons, Simons and Zālinescu, Burachik and Svaiter etc.

Proposition 2 *A proper lsc convex function on a Banach space (i) is continuous throughout the core of its domain; and (ii) has a non-empty subgradient throughout the core of its domain.*

These two basic facts lead to:

Theorem 1 (**Hahn-Banach sandwich**) *Suppose $f, -g$ are lsc convex on a Banach space X and* $f(x) \geq g(x)$, *for all x in X. Assume (CQ) holds:*

$$0 \in \mathrm{core}\,(\mathrm{dom}\,(f) - \mathrm{dom}\,(-g))\,. \tag{1}$$

Then there is an affine continuous function a such that

$$\boxed{f(x) \geq a(x) \geq g(x)}$$

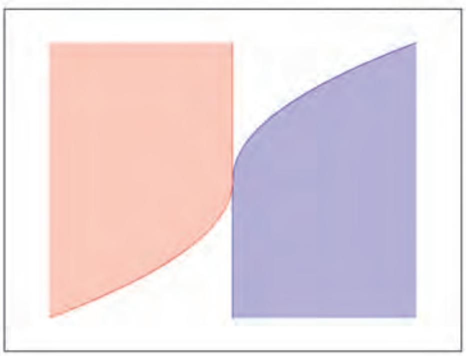
for all x in X.

Proof. The perturbation or *value function*

$$h(u) := \inf_{x \in X} f(x) - g(x-u)$$

is convex and (CQ) implies continuity at 0.* Hence there is $\lambda \in \partial h(0)$, which is the linear part of the affine separator. As needed, we have

$$f(x) - g(u-x) \geq h(u) - h(0) \geq \lambda(u). \qquad \blacksquare$$



$-\sqrt{-x} \geq \sqrt{x}$

- We refer to *constraint qualifications* like (1) as *transversality conditions*

◁ **CQ failure**

- It is easy to deduce complete *Fenchel duality theorem* from Thm 1

Proposition 3 *For a closed convex function* f *and* $f_J := f + \frac{1}{2}\|\cdot\|^2$ *we have that*

$$\left(f + \frac{1}{2}\|\cdot\|^2\right)^* = f^* \Box \frac{1}{2}\|\cdot\|_*^2$$

is everywhere continuous. Also

$$v^* \in \partial f(v) + J(v) \Leftrightarrow f_J^*(v^*) + f_J(v) - \langle v, v^* \rangle \leq 0.$$

$^*B_\varepsilon \subset \{f \leq M\} - \{g \leq M\} \Rightarrow h|B_\varepsilon \leq 2M.$

2a. Representative Functions

A convex function $\mathcal{H}_T$ is a representative function for a monotone T on $X \times X^*$ if (i) $\mathcal{H}_T(x, x^*) \geq \langle x, x^* \rangle$ for all x, x^*; (ii) $\mathcal{H}_T(x, x^*) = \langle x, x^* \rangle$ **if** $x^* \in T(x)$.

For T maximal, Prop. 1 shows $\mathcal{F}_T$ is a representative function as is the (closed) convexification

$$\mathcal{P}_T(x, x^*) = \inf \sum_{i=1}^{N} \lambda_i \langle x_i, x_i^* \rangle$$
$$\text{s.t.} \sum_i \lambda_i (x_i, x_i^*, 1) = (x, x^*, 1), x_i^* \in T(x_i), \lambda_i \geq 0.$$

Proposition 4 (Penot) *For any monotone mapping T, $\overline{\mathcal{P}}_T$ is a representative convex function.*

Proof. By monotonicity we have

$$\mathcal{P}_T(x, x^*) \geq \langle x^*, y \rangle + \langle y^*, x \rangle - \langle y^*, y \rangle,$$

for $y^* \in T(y)$. Thus, for *all* points

$$\mathcal{P}_T(x, x^*) + \mathcal{P}_T(y, y^*) \geq \langle x^*, y \rangle + \langle y^*, x \rangle.$$

By definition $\mathcal{P}_T(x, x^*) \leq \langle x^*, x \rangle$ for $x^* \in T(x)$. Setting $x = y$ and $x^* = y^*$ shows $\mathcal{P}_T(x, x^*) = \langle x^*, x \rangle$ for $x^* \in T(x)$ while $\mathcal{P}_T(z, z^*) \geq \langle z^*, z \rangle$ for (z^*, z) in $\operatorname{conv} \operatorname{graph} T$: (also for $\overline{\mathcal{P}}_T$). ■

2b. Monotone Extension Theorems

A direct calculation shows $(\mathcal{P}_T)^* = \mathcal{F}_T$ for any monotone T. This convexification originates with Simons but was much refined by Penot.

We illustrate its flexibility by proving a central case of the Debrunner-Flor theorem *without* Brouwer's theorem.

Theorem 2 *Suppose T is monotone on X with range contained in αB_{X^*}, for some $\alpha > 0$. Then*

(a) *For every x_0 in X there is $x_0^* \in \overline{\text{conv}}^* R(T) \subset \alpha B_{X^*}$ such that (x_0, x_0^*) is monotonically related to* graph (T).

(b) *Hence, T has a bounded monotone extension $\overline{T}$ with* dom $(\overline{T}) = X$ *and $R(\overline{T}) \subset \overline{\text{conv}}^* R(T)$.*

(c) *Thence, a maximal monotone T with bounded range has* dom $(T) = X$.

Proof. (a) It is enough, after translation, to show $x_0 = 0 \in \text{dom}\,(T)$. Fix $\alpha > 0$ with $R(T) \subset C := \overline{\text{conv}}^*\, R(T) \subset \alpha\, B_{X^*}$.
Consider

$$\pi_T(x) := \inf\,\{\mathcal{P}_T(x, x^*) : x^* \in C\}\,.$$

Then π_T is convex since $\mathcal{P}_T$ is. Observe that

$$\mathcal{P}_T(x, x^*) \geq \langle x, x^* \rangle$$

and so $\pi_T(x) \geq \inf_{x^* \in C} \langle x, x^* \rangle \geq -\alpha\, \|x\|$ for all x in X. As $x \mapsto \inf_{x^* \in C} \langle x, x^* \rangle$ is concave and continuous the Sandwich Theorem 1 applies.

Thus, there exist w^* in X^* and γ in $\mathbf{R}$ with

$$\mathcal{P}_T(x, x^*) \geq \pi_T(x) \geq \langle x, w^* \rangle + \gamma \geq \inf_{x^* \in C} \langle x, x^* \rangle \geq -\alpha\, \|x\|$$

for all x in X and x^* in $C \subset \alpha\, B_{X^*}$.
Setting $x = 0$ shows $\gamma \geq 0$. Now, for any (y, y^*) in the graph of T we have $\mathcal{P}_T(y, y^*) = \langle y, y^* \rangle$. Thus,

$$\langle y - 0, y^* - w^* \rangle \geq \gamma \geq 0,$$

which shows that $(0, w^*)$ is monotonically related to the graph of T.

Finally, $\langle x, w^* \rangle + \gamma \geq \inf_{x^* \in C} \langle x, x^* \rangle \geq -\alpha\, \|x\|$ for all $x \in X$ involves three sublinear functions, and so implies that $w^* \in C \subset \alpha\, B_{X^*}$.

(b) Consider the set $\mathcal{E}$ of all monotone extensions of T with range in $C \subset \alpha\, B_{X^*}$, ordered by inclusion. By Zorn's lemma $\mathcal{E}$ admits a maximal member $\overline{T}$ and by (a) $\overline{T}$ has domain the whole space.
(c) follows immediately. ■

▶ $R(T) \subset M B_{X^*} \Rightarrow \pi_T := \inf_{X^*} \mathcal{P}_T(\cdot, x^*) \geq -M\|\cdot\|$

$$\boxed{x^* \in \partial \pi_T(x) \Leftrightarrow \pi_T(x) + \mathcal{F}_T(0, x^*) = \langle x, x^* \rangle}$$

- (a) holds on *any* w^*-closed convex set C in Hilbert space (Brezis). Our proof applies if

$$x_0 \in \text{core}\,(\text{dom}\,\pi_T + \text{dom}\,\sup_C).$$

The full Debrunner-Flor extension theorem is next:

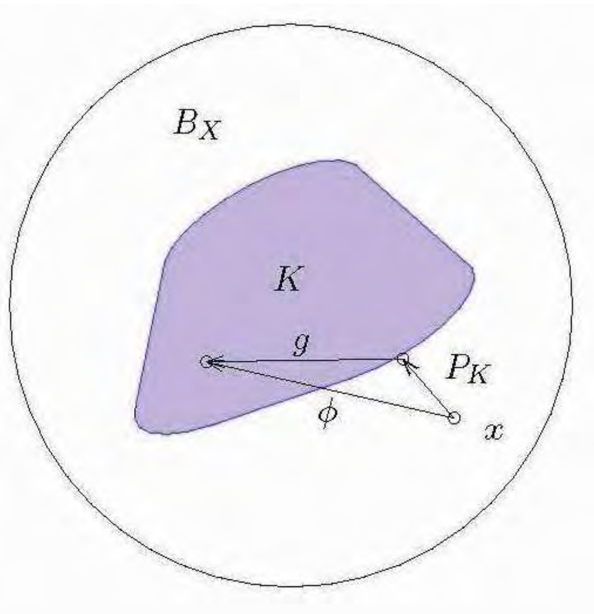
Theorem 3 (**Debrunner-Flor**) *Suppose T is a monotone operator on X with range $T \subset C$ for some weak-star compact and convex C. Suppose also $\varphi\colon C \mapsto X$ is weak-star to norm continuous. Then there is some $c^* \in C$ with*

$$\langle x - \varphi(c^*), x^* - c^* \rangle \geq 0$$

for all $x^* \in T(x)$.

Theorem 4 *The full Debrunner-Flor extension theorem is* equivalent *to Brouwer's theorem.*

Proof. Phelps derives Debrunner-Flor from Brouwer. Conversely, let g be a continuous self-map of a compact convex set $K \subset \operatorname{int} B_X$ in finite dimensions.

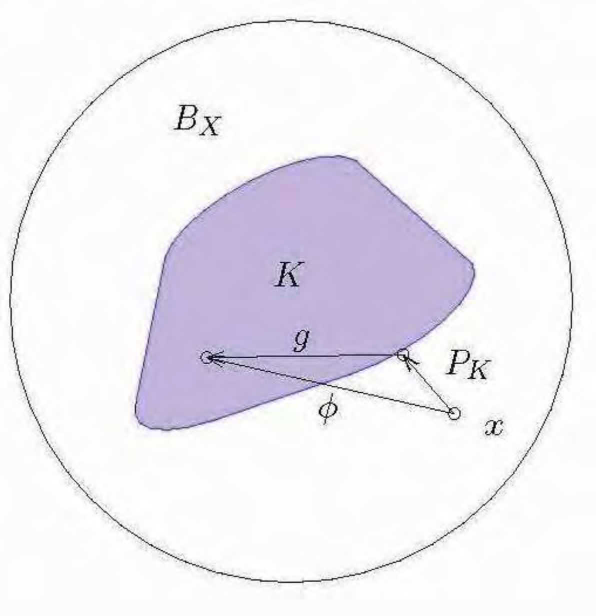


Apply Debrunner-Flor to the identity I on B_X and to $\varphi\colon B_X \mapsto X$ given by $\varphi(x) := g(P_K x)$, where P_K is the metric projection. We have $x_0^* \in B_X$, $x_0 := \varphi(x_0^*) = g(P_K x_0^*) \in K$,

$$\langle x - x_0, x - x_0^* \rangle \geq 0$$

for all $x \in B_X$.

Since $x_0 \in \operatorname{int} B_X$, for $h \in X$ and small $\epsilon > 0$ we have $x_0 + \epsilon h \in B_X$ and so $\langle h, x_0 - x_0^* \rangle \geq 0$ for all $h \in X$. Thus, $x_0 = x_0^*$ and so $P_K x_0^* = P_K x_0 = x_0 = g(P_K x_0^*)$, is a fixed point of the arbitrary self-map g. ■



Apply Debrunner-Flor to the identity I *on* B_X and to $\varphi\colon B_X \mapsto X$ given by $\varphi(x) := g(P_K\, x)$, where P_K is the metric projection. We have $x_0^* \in B_X$, $x_0 := \varphi(x_0^*) = g(P_K\, x_0^*) \in K$,

$$\langle x - x_0, x - x_0^* \rangle \geq 0$$

for all $x \in B_X$.

Since $x_0 \in \operatorname{int} B_X$, for $h \in X$ and small $\epsilon > 0$ we have $x_0 + \epsilon h \in B_X$ and so $\langle h, x_0 - x_0^* \rangle \geq 0$ for all $h \in X$. Thus, $x_0 = x_0^*$ and so $P_K\, x_0^* = P_K\, x_0 = x_0 = g(P_K\, x_0^*)$, is a fixed point of the arbitrary self-map g. ■

2c. Local Boundedness Results

Recall that an operator T is *locally bounded* around a point x if $T(B_\varepsilon(x))$ is bounded for some $\varepsilon > 0$.

Theorem 5 (**Simons, Veronas**) *Let $S, T\colon X \to 2^{X^*}$ be monotone operators. Suppose*

$$0 \in \operatorname{core}\,[\operatorname{conv}\operatorname{dom}(T) - \operatorname{conv}\operatorname{dom}(S)].$$

There exist $r, c > 0$ so that, for all x with $t^* \in T(x)$ and $s^* \in S(x)$,

$$\max(\|t^*\|, \|s^*\|) \le c\,(r + \|x\|)(r + \|t^* + s^*\|).$$

Proof. Consider the convex lsc function*

$$\sigma_T(x) := \sup_{z^* \in T(z)} \frac{\langle x - z, z^* \rangle}{1 + \lambda\|z\|}.$$

First, $\operatorname{conv}\operatorname{dom}(T) \subset \operatorname{dom}\sigma_T$, and $0 \in$ core

$$\bigcup_{i=1}^{\infty} [\{x : \sigma_S(x) \le i, \|x\| \le i\} - \{x\colon \sigma_T(x) \le i, \|x\| \le i\}],$$

and apply conventional Baire category techniques–with some care. ■

Corollary 1 *Let X be any Banach space. Suppose T is monotone and*

$$x_0 \in \operatorname{core} \operatorname{conv} \operatorname{dom}(T).$$

Then T is locally bounded around x_0.

Proof. Let $S = 0$ in Theorem 5 or directly apply Proposition 2 to σ_T. ■

We can also improve Theorem 2.

Corollary 2 *A monotone mapping T with bounded range admits an everywhere defined maximal monotone extension with bounded range contained in $\overline{\operatorname{conv}}^* R(T)$.*

Proof. Let $\widehat{T}$ denote the extension of Theorem 2 (b) Clearly it is everywhere locally bounded. The desired extension $\widetilde{T}(x)$ is the operator whose graph is the norm-weak-star closure of the graph of $x \mapsto \operatorname{conv} \widehat{T}(x)$, since this is both monotone and is a norm-w* cusco.
Explicitly,

$$\widetilde{T}(x) := \cap_{\varepsilon>0} \overline{\operatorname{conv}}^* \widehat{T}(B_\varepsilon(x))$$

(see ToVA). ■

A mapping is *locally maximal monotone*, or *type (FP)*, if $(\operatorname{graph} T^{-1}) \cap (V \times X)$ is maximal monotone in $V \times X$, for every convex open set V in X^* with $V \cap \operatorname{range} T \neq \emptyset$.

- Simons showed subgradients are (FP). So are maximal monotones on reflexive space (SF-P).

We may usefully apply Corollary 2 to

$$T_n(x) := T(x) \cap n\, B_{X^*}.$$

Often the extension, $\widehat{T_n}$ is unique:

Proposition 5 (Fitzpatrick-Phelps) *Suppose T is maximal and n is such that $R(T) \cap n \operatorname{int} B_{X^*} \neq \emptyset$.*
(a) There is a unique maximal monotone $\hat{T}_n$ with

$$T_n(x) \subset \widehat{T_n}(x) \subset nB_{X^*}$$

whenever $M_n(x) :=$

$$\{x^* \in nB^* : \langle x^* - z^*, x - z\rangle \geq 0, \forall z^* \in T(z) \cap n \operatorname{int} B_{X^*}\}$$

is monotone; in which case $M_n = \hat{T}_n$.
(b) This holds if T is type (FP) and B_{X^*} is strictly convex; so for any maximal monotone on a rotund dual reflexive norm, e.g. Hilbert space.

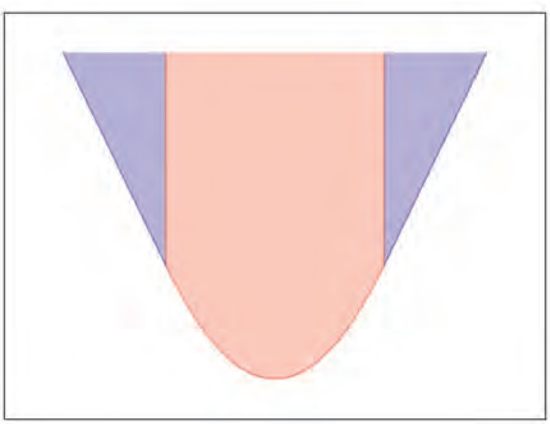
Proof. Since $\widehat{T_n}$ exists by Corollary 2 and since $\widehat{T_n}(x) \subset M_n(x)$, (a) follows. We refer to Fitzpatrick and Phelps for the fairly easy proof of (b). ■

★ $\{\widehat{T_n}\}_{n\in\mathbb{N}}$ is a non-reflexive generalization of the resolvent -based *Yosida approximate* or the *Hausdorff-Moreau Lipschitz regularization* of a convex function.

In the (FP) case one also easily shows (F-P) that:

(I) $\widehat{T_n}(x) = T(x) \cap n\,B_{X^*}$ if $T(x) \cap \operatorname{int} n\,B_{X^*} \neq \emptyset$

(II) $\widehat{T_n}(x) \setminus T(x) \subset n\,S_{X^*}$.



- cl $R(T)$ is convex if cl $R(\widehat{T_n})$ is for T type (II)

◀ function regularization

- For local properties (e.g. differentiability) one may replace T by $\widehat{T_n}$

2d. Maximality of Subgradients

Theorem 6 *Every closed convex function has a (locally) maximal monotone subgradient.*[*]

Proof. (Sketch) Without loss we may suppose

$$\langle 0 - x^*, 0 - x\rangle \geq 0 \text{ for all } x^* \in \partial f(x)$$

but $0 \notin \partial f(0)$; so $f(\overline{x}) - f(0) < 0$ for some $\overline{x}$.

The *Approximate mean value theorem* (see [ToVA, Thm. 3.4.6]) lets us find $x_n \xrightarrow{f} c \in (0, \overline{x}]$ and $x_n^* \in \partial f(x_n)$ with

$$\limsup_n \langle x_n^*, x_n - c\rangle \leq 0, \limsup_n \langle x_n^*, \overline{x}\rangle \leq f(\overline{x}) - f(0) < 0.$$

Now $c = \theta\,\overline{x}$ for some $\theta > 0$. Hence,

$$\limsup_n \langle x_n^*, x_n\rangle < 0,$$

a contradiction. The locally maximal case follows 'similarly' on exploiting that $f(x_n) \to f(c)$, and that ∂f is dense type. ■

[*]This fails in *all* incomplete normed spaces and in *some* Fréchet spaces

2e. Convexity of Range and Domain

Corollary 3 *Let X be any Banach space. Suppose that T is maximal monotone with* core conv $D(T)$ *nonempty. Then*

$$\operatorname{core}\operatorname{conv} D(T) = \operatorname{int}\operatorname{conv} D(T) \subset D(T). \quad (2)$$

In consequence dom (T) *has both a convex closure and a convex interior.*

Proof. We first prove the inclusion in (2). Fix $x + \varepsilon B_X \subset \operatorname{int}\operatorname{conv}\operatorname{dom}(T)$ and, via Cor. 1, select $M := M(x, \varepsilon) > 0$ so that $T(x + \varepsilon B_X) \subset M B_{X^*}$. For $N > M$ define w^*-closed nested sets

$$T_N(x) := \{x^* : \langle x - y, x^* - y^* \rangle \geq 0, \forall y^* \in T(y) \cap N B_{X^*}\}.$$

By Theorem 2 (b), the sets are non-empty, and by the next lemma, bounded, hence w^*-compact. By maximality of T, $T(x) = \cap_N T_N(x) \neq \emptyset$, as a nested intersection, and x is in dom (T) as asserted.

Then $\operatorname{int}\operatorname{conv}\operatorname{dom}(T) = \operatorname{int}\operatorname{dom}(T)$ and so the final conclusion follows. ■

Lemma 1 *For* $x \in \operatorname{int}\operatorname{conv}\operatorname{dom}(T)$ *and* N *sufficiently large,* $T_N(x)$ *is bounded.*

Proof. A Baire category argument shows for N large and $u \in 1/N\, B_X$ that $x + u \in \operatorname{cl}\operatorname{conv} D_N$ for

$$D_N := \{z \colon z \in D(T) \cap N\, B_X, T(z) \cap N\, B_{X^*} \neq \emptyset\}.$$

Now for each $x^* \in T_N(x)$, since $x + u$ lies in the closed convex hull of D_N, we have

$\langle u, x^* \rangle \leq \sup\{\langle z - x, z^* \rangle \colon z^* \in T(z) \cap N B_{X*}, z \in N B_X\}$ $\leq 2N^2$ and so $\|x^*\| \leq 2N^3$. ■

Another nice application is:

Corollary 4 (**Verona**) *Let* X *be Banach and let* $S, T : X \to 2^{X^*}$ *be maximal monotone. Suppose*

$$0 \in \operatorname{core}[\operatorname{conv}\operatorname{dom}(T) - \operatorname{conv}\operatorname{dom}(S)].$$

Then for any $x \in \operatorname{dom}(T) \cap \operatorname{dom}(S)$*,* $T(x) + S(x)$ *is a* w^**-closed subset of* X^**.*

Proof. Theorem 5 shows bounded w^*-convergent nets in $T(x) + S(x)$ have limits in $T(x) + S(x)$. We apply the *Krein-Smulian theorem*. ■

- Thus, we preserve some structure. It is still open if $T+S$ must actually be maximal.

We may neatly recover convexity of $\operatorname{int} D(T)$:

Theorem 7 (**Simons, 2005**) *Suppose T is maximal monotone and int $\operatorname{dom}(T)$ is nonempty. Then* $\operatorname{int}\operatorname{dom}(T) = \operatorname{int}\{x\colon (x,x^*) \in \operatorname{dom}\mathcal{F}_T\}$.

- Suppose T is *domain regularizable*: for $\varepsilon > 0$, there is a maximal T_ε with $H\left(D(T), D(T_\varepsilon)\right) \leq \varepsilon$ and $\operatorname{core} D(T_\varepsilon) \neq \emptyset$. In reflexive space we can use

$$T_\varepsilon := \left(T^{-1} + N_{\varepsilon B_X}^{-1}\right)^{-1}.$$

Then $\overline{\operatorname{dom}}\,(T)$ is convex.

3. Cyclic and Acyclic Monotone Operators

For $N = 2, 3, \ldots$, an operatorn T is *N-monotone* if

$$\sum_{k=1}^{N} \langle x_k^*, x_k - x_{k-1} \rangle \geq 0$$

whenever $x_k^* \in T(x_k)$ and $x_0 = x_N$.

T is *cyclically monotone* if T is N-monotone for all $N \in \mathbb{N}$, as holds for convex subgradients.

- Monotonicity = 2-monotonicity:
  $\langle x_1^*, x_1 - x_2 \rangle + \langle x_2^*, x_2 - x_1 \rangle \geq 0$

- $(N+1)$-monotone $\subsetneq$ N-monotone (Asplund):
  $\langle x_1^*, x_1 - x_3 \rangle + \langle x_2^*, x_2 - x_1 \rangle + \langle x_3^*, x_3 - x_2 \rangle \geq 0.$

- It is a classical result of Rockafellar that *every maximal cyclically monotone operator is the subgradient of a proper closed convex function* (and conversely).

We recast this result to make the parallel with the Debrunner-Flor Theorem 2 explicit.

Theorem 8 (**Rockafellar**) *Suppose C is cyclically monotone on a Banach space X.*

Then C has a maximal cyclically monotone extension $\overline{C}$, which is of the form $\boxed{\overline{C} = \partial f_C}$ for some proper closed convex function f_C.

Moreover $R(\overline{C}) \subset \overline{\text{conv}}^* R(C)$.

Proof. We fix $x_0 \in \text{dom}\, C, x_0^* \in C(x_0)$ and define

$$f_C(x) := \sup_{x_k^* \in C(x_k)} \{\langle x_n^*, x - x_n\rangle + \sum_{k=1}^{n-1} \langle x_{k-1}^*, x_k - x_{k-1}\rangle\}$$

where the 'sup' is over all $n \in \mathbb{N}$ and all such chains. The proof in Phelps' monograph shows that

$$C \subset \overline{C} := \partial f_C.$$

The range assertion follows because f_C is the supremum of affine functions whose linear parts all lie in range C. This is most easily seen by writing $f_C = g_C^*$ with

$$g_C(x^*) := \inf\{\sum_i t_i \alpha_i \colon \sum_i t_i x_i^* = x^*, \sum_i t_i = 1, t_i > 0\}$$

for appropriate $\alpha_i \in \mathbb{R}$. ∎

The relationship of $\mathcal{F}_{\partial f}$ and ∂f is complicated:

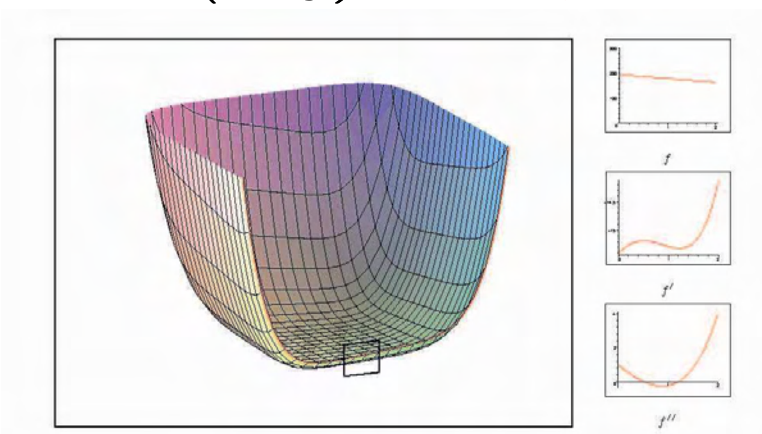
$$\begin{aligned}\langle x, x^* \rangle &\leq \mathcal{F}_{\partial f}(x, x^*) \leq f(x) + f^*(x^*) \leq \mathcal{F}^*_{\partial f}(x, x^*)\\ &\leq \langle x, x^* \rangle + \iota_{\partial f}(x, x^*),\end{aligned}$$

(see Bauschke et al.) Two central questions are:

Q1. When is a maximal monotone operator T the sum of a subgradient ∂f and a skew linear S? This is closely related to the behaviour of

$$\mathcal{F}\mathcal{L}_T(x) := \int_0^1 \sup_{x^*(t) \in T(tx)} \langle x, x^*(t) \rangle \, dt$$

when $0 \in \operatorname{core} \operatorname{dom} T$, then $\mathcal{F}\mathcal{L}_T = \mathcal{F}\mathcal{L}_{\partial f} = f$ and we call T *(fully) decomposable*.



Fitzpatrick's Last Function [*][†]

[*]The use of $\mathcal{F}\mathcal{L}_T$ originates in discussions I had with Fitzpatrick shortly before his death.

[†]T *'inherits the differentiability'* of $\mathcal{F}\mathcal{L}_T$.

Example 6. *Consider the mapping*

$$T(x,y) := \big(\sinh(x) - \alpha\, y^2/2, \sinh(y) - \alpha\, x^2/2)\big) .$$

Then

$$DT = \begin{pmatrix} \cosh(x) & -\alpha y \\ -\alpha x & \cosh(y) \end{pmatrix}$$

which is monotone iff

$$\alpha^2 \le \frac{\cosh(x)}{x} \frac{\cosh(y)}{y}$$

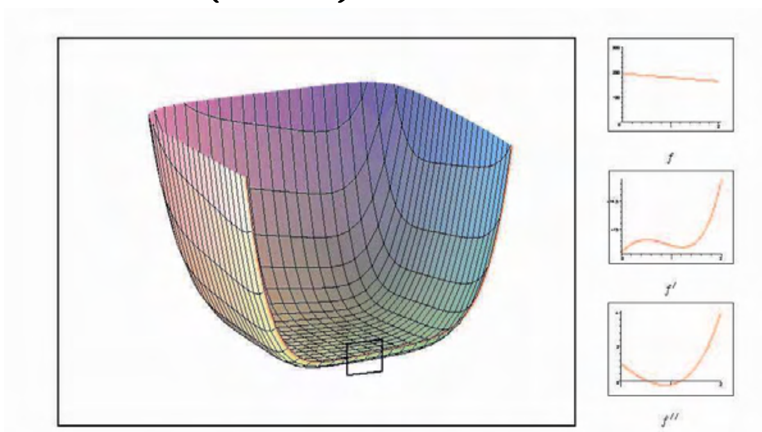
for all $x, y > 0$. *The right hand side is a separable convex function, and is minimized at* $x = y = x_0 = \coth(x_0) = 1.199678\ldots$. *So* T *is monotone iff* $\alpha^2 \le \sinh^2(x_0) = 2.276717\ldots$.

As before, the off-diagonal entries of DT *are nonconstant and unequal, so* T *is indecomposable.*

Q1. **When is a maximal monotone operator T the sum of a subgradient ∂f and a skew linear S?** This is closely related to the behaviour of

$$\mathcal{F}\mathcal{L}_T(x) := \int_0^1 \sup_{x^*(t)\in T(tx)} \langle x, x^*(t)\rangle \, dt$$

when $0 \in \operatorname{core} \operatorname{dom} T$, then $\mathcal{F}\mathcal{L}_T = \mathcal{F}\mathcal{L}_{\partial f} = f$ and we call T *(fully) decomposable*.



Fitzpatrick's Last Function *†

*The use of $\mathcal{F}\mathcal{L}_T$ originates in discussions I had with Fitzpatrick shortly before his death.

†T *'inherits the differentiability'* of $\mathcal{F}\mathcal{L}_T$.

A MONOTONE CONVERGENCE THEOREM FOR SEQUENCES OF NONLINEAR MAPPINGS

Edgar Asplund

In this paper we prove a theorem generalizing the elementary theorem on convergence of bounded, monotone sequences of real numbers, and also the theorem of Vigier and Nagy, cf. [**2**, Appendice II] on the convergence of certain sequences of symmetric linear operators on Hilbert space.

The paper consists of two sections. In the first we prove the main monotone convergence theorem (Theorem 1) and apply it to prove a decomposition for monotone operators which generalizes the decomposition of a linear operator into symmetric and antisymmetric parts. In the second section we apply Theorem 1

Q2. **How does one generalize the decomposition of a linear monotone operator** L into a symmetric (cyclic) and a skew (acyclic) part? Viz

$$L = \frac{1}{2}(L + L^*|_X) + \frac{1}{2}(L - L^*|_X).$$

3a. Asplund's approach to Q2

Every 3-monotone operator such that $0 \in T(0)$ has the local property that

$$\langle x, x^* \rangle + \langle y, y^* \rangle \geq \langle x, y^* \rangle \tag{3}$$

whenever $x^* \in T(x)$ and $y^* \in T(y)$. We call a monotone operator satisfying (3), *3^--monotone*, and write $T \geq_N S$ if $T = S + R$ with R being N-monotone ($T \geq_{\omega_0} S$ if R is cyclically monotone.)

Proposition 6 (**Dini Property**) *Let N be $3^-, 3, 4$, ..., or ω_0. Consider an increasing (infinite) net of monotone operators on a space X, satisfying*

$$0 \leq_N T_\alpha \leq_N T_\beta \leq_2 T$$

if $\alpha < \beta \in \mathcal{A}$. Suppose that $0 \in T_\alpha(0), 0 \in T(0)$ and that $0 \in \operatorname{core} \operatorname{dom} T$. Then

a) *There is a N-monotone $T_{\mathcal{A}}$ with $T_\alpha \leq_N T_{\mathcal{A}} \leq_2 T$, for all $\alpha \in \mathcal{A}$.*

b) *If $R(T) \subset MB_{X^*}$ for some $M > 0$ then one may suppose $R(T_{\mathcal{A}}) \subset MB_{X^*}$.*

Proof. a) **The single-valued case.** Since $0 \leq_2 T_\alpha \leq_2 T_\beta \leq_2 T$, while $T(0) = 0 = T_\alpha(0)$, we have

$$0 \leq \langle x, T_\alpha(x)\rangle \leq \langle x, T_\beta(x)\rangle \leq \langle x, T(x)\rangle,$$

for all x in $\operatorname{dom} T$. This shows $\langle x, T_\alpha(x)\rangle$ converges as α goes to ∞. Fix $\varepsilon > 0, M > 0$ with $T(\varepsilon B_X) \subset M B_{X^*}$. We write $T_{\beta\alpha} = T_\beta - T_\alpha$ for $\beta > \alpha$, so that $\langle T_{\beta\alpha} x, x\rangle \to 0$ for $x \in \operatorname{dom} T$ as $\alpha, \beta \to \infty$.
We appeal to (3) to obtain

$$\langle x, T_{\beta\alpha}(x)\rangle + \langle y, T_{\beta\alpha}(y)\rangle \geq \langle T_{\beta\alpha}(x), y\rangle, \tag{4}$$

for $x, y \in \operatorname{dom} T$. Also, $0 \leq \langle x, T_{\beta\alpha}(x)\rangle \leq \varepsilon$ for $\beta > \alpha > \gamma(x)$ for all $x \in \operatorname{dom} T$.

Now, $0 \leq \langle y, T_{\beta\alpha}(y)\rangle \leq \langle y, T(y)\rangle \leq \varepsilon M$ for $\|y\| \leq \varepsilon^2$. Thus, for $\|y\| \leq \varepsilon$ and $\beta > \alpha > \gamma(x)$ we have

$$\begin{aligned} \varepsilon(M+\varepsilon) &\geq \langle x, T_{\beta\alpha}(x)\rangle + \langle y, T(y)\rangle \\ &\geq \langle x, T_{\beta\alpha}(x)\rangle + \langle y, T_{\beta\alpha}(y)\rangle \\ &\geq \langle y, T_{\beta\alpha}(x)\rangle, \end{aligned} \tag{5}$$

from which we obtain $\|T_{\beta\alpha}(x)\| \leq M + \varepsilon$ for all $x \in \operatorname{dom} T$, while $\langle y, T_{\beta\alpha}(x)\rangle \to 0$ for all $y \in X$.

We conclude that $\{T_\alpha(x)\}_{\alpha \in \mathcal{A}}$ is a norm-bounded weak-star Cauchy net and so weak-star convergent to the desired N-monotone limit $T_{\mathcal{A}}(x)$.

<u>The set-valued case</u> uses (3) to deduce that $T_\beta = T_\alpha + T_{\beta\alpha}$ where (i) $T_{\beta\alpha} \subset (M+\varepsilon)B_{X^*}$ and (ii) for each $t^*_{\beta\alpha} \in T_{\beta\alpha}$ one has $t^*_{\beta\alpha} \rightharpoondown^* 0$ as $\alpha, \beta \to \infty$. The conclusion is as before but somewhat more technical.

b) Fix $x \in X$, and apply (3 to T_α to write

$$\langle Tx, x\rangle + \langle Ty, y\rangle \geq \langle T_\alpha x, x\rangle + \langle T_\alpha y, y\rangle \geq \langle T_\alpha x, y\rangle$$

for all $y \in D(T) = X$, by Theorem 2 (c). Hence

$$\langle Tx, x\rangle + M\|y\| \geq \|T_\alpha x\|\,\|y\|, \qquad \forall \|y\|$$

Let $\|y\| \to \infty$ to show $T_\alpha(x)$ lies in the M-ball, and since the ball is weak-star closed, so does $T_{\mathcal{A}}(x)$.

The set-valued case is analogous but *messier*. ■

- $0 \leq_2 (-ny, nx) \leq_2 (-y, x)$ for $n \in \mathbb{N}$, shows the need for (3) in the deduction that $T_{\beta\alpha}(x)$ are equi-norm bounded.

★ **(Daniel property)** If X is an *Asplund space*, the proof of Prop 6 can be adjusted to show

$$\boxed{T_{\mathcal{A}}(x) = \text{norm}-\lim_{\alpha\to\infty} T_\alpha(x)}$$

Definition 1 *We say a maximal monotone operator A is* acyclic *if whenever $A = \partial g + S$ with S maximal monotone and g closed and convex then g is necessarily linear.*

We provide a broad extension of Asplund's original idea:

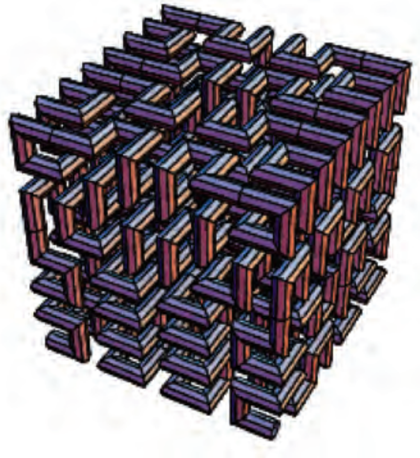
Theorem 9 (Asplund Decomposition) *Suppose T is maximal monotone with $\operatorname{core}\operatorname{dom} T \neq \emptyset$.*

a) *Then T may be decomposed as $T = \partial f + A$, where f is closed and convex while A is acyclic.*

b) *If the range of T lies in $M\,B_{X^*}$ then f may be assumed M-Lipschitz.*

♠ There is a like $N-$*cyclic* decomposition.

A Hilbert curve in 3D
is more constructive



Proof. a) We normalize so $0 \in T(0)$. Zorn's lemma applies to the cyclically monotone operators

$$\mathcal{C} := \{C \colon 0 \leq_{\omega_0} C \leq_2 T,\, 0 \in C(0)\}$$

in the cyclic order. By Prop. 6 every chain in $\mathcal{C}$ has a cyclically monotone upper-bound.

Fix a maximal $\overline{C}$ with $0 \leq_{\omega_0} \overline{C} \leq_2 T$. Hence $T = \overline{C} + A$ where by construction A is acyclic. Now, $T = \overline{C} + A \subset \partial f + A$, by Rockafellar's result. Since T is maximal the decomposition is as asserted.

b) We use the facts that (i) $0 \leq_{3^-} U \leq_2 T$ implies $\|U(x)\| \leq \|T(x)\|$ for all x and (ii) an M-bounded cyclically monotone operator extends to an M-Lipschitz subgradient—as Theorem 8 confirms. ■

By way of application we offer:

Corollary 5 *Let T be an arbitrary maximal monotone operator T. For $\mu > 0$ one may decompose*

$$T \cap \mu B_{X^*} \subset \widehat{T_\mu} = \partial f_\mu + A_\mu,$$

where f_μ is μ-Lipschitz and A_μ is acyclic (with bounded range).

Proof. Combining Theorem 9 with Proposition 5 we deduce that the composition is as claimed. ∎

- In Corollary 5, range A_μ is bounded. Thus, it is only skew and linear when T is cyclic—so a non-cyclic range bounded monotone operator is never fully decomposable in the sense of **Q1**.

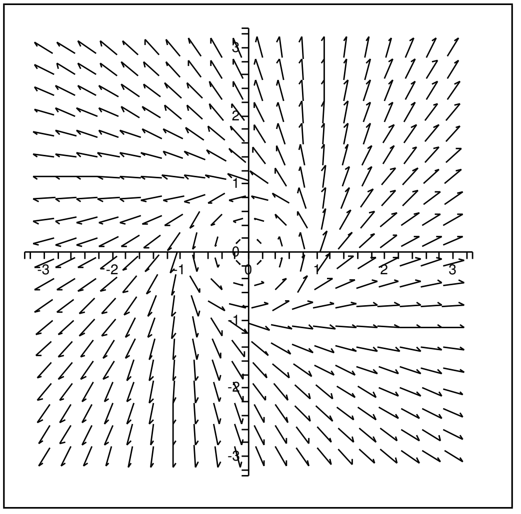
- Theorem 9 et al are entirely existential: can one prove Theorem 9 constructively in finite dimensions?

- How does one effectively diagnose acyclicity?

An Acyclic Monotone Operator

A concrete example in $\mathbb{R}^2$ is implicit in these observations (JMB-Wiersma).

- R_θ: rotation by $\theta < \pi/2$
- $\widehat{R_\theta}$: the range restriction to B_1 extended to be maximal with range in B_1.
- **CONJECTURE** $\widehat{R_\theta}$ is acyclic.



Theorem. Let

$$\alpha(x) := \sqrt{1 - 1 \wedge \frac{1}{\|x\|^2}}, \qquad \beta(x) := 1 \wedge \frac{1}{\|x\|}.$$

Then

$$\widehat{R}_{\pi/2}(x) = \alpha(x)\, R\left(\frac{x}{\|x\|}\right) + \beta(x)\, \frac{x}{\|x\|}$$

is acyclic.

► The proof is delicate and needs $T^2 = -I$.

3b. Fitzpatrick Functions of Order N

- The *Fitzpatrick function of order N* is:

$$\mathcal{F}_T^N(x,x^*) := \sup_{x_N = x} \left\{ \langle x_1, x^* \rangle + \sum_{k=1}^{N-1} \langle x_{k+1} - x_k, x_k^* \rangle \right\}$$

where $x_k^* \in T(x_k)$ for $1 \le k \le N-1$.

- The *Rockafellar function of order N* is:

$$\mathcal{R}_T^N(x, x_1, x_1^*) : =$$

$$\sup \langle x - x_{N-1}, x_{N-1}^* \rangle \; + \; \sum_{i=1}^{N-2} \langle x_{i+1} - x_i, x_i^* \rangle,$$

for $x_1^* \in T(x_1)$, $x \in X$ and $N \ge 3$, over all $x_k^* \in T(x_k)$ (for $2 \le k \le N-1$).

Then $\mathcal{F}_T^\infty := \left(\mathcal{P}_T^\infty\right)^* = \sup \mathcal{F}_T^N$, $\mathcal{P}_T^\infty := \inf \mathcal{P}_T^N$, and $\mathcal{R}_T :=\sup \mathcal{R}_T^N$. Moreover, for a maximal N-monotone T we have

$$\mathcal{F}_T^N(x,x^*) \ge \langle x, x^* \rangle$$

with equality if and only if $x^* \in T(x)$.

We recast Rockafellar's Theorem 8:

Theorem 10 *Suppose A is cyclically monotone. For $a_1^* \in A(a_1)$, $x \mapsto \mathcal{R}_A(x, a_1, a_1^*)$ is closed and convex and $\mathcal{R}_A(a_1, a_1, a_1^*) = 0$. Also for every $x \in X$, $A(x) \subset \partial\mathcal{R}_A(x, a_1, a_1^*)$. When A is maximal cyclically monotone one has $A = \partial\mathcal{R}_A$. Moreover, for every closed f satisfying $\partial f = A$, one has*

$$f(x) - f(a_1) = \mathcal{R}_A(x, a_1, a_1^*) \quad \text{for } x \in X.$$

We now connect the infinite Fitzpatrick function to the Rockafellar function.

Theorem 11 (**Bartz-Bauschke-Borwein-Reich - Wang**) *Let A be cyclically monotone. For each closed convex function f on X such that $A \subset \partial f$ one has*

$$\mathcal{F}_A^\infty(x, x^*) = f(x) + \sup_{a_1^* \in A(a_1)} \langle x^*, a_1 \rangle - f(a_1),$$

for $(x, x^*) \in X \times X^*$. If actually $\operatorname{dom} A = \operatorname{dom} \partial f$ *then*

$$\mathcal{F}_A^\infty(x, x^*) = (f \oplus f^*)(x, x^*) := f(x) + f^*(x^*),$$

for all $(x, x^*) \in X \times X^*$.

The Fitzpatrick Functions of a Rotation

Theorem 12 (BaBW) *Let $\theta \in [0, \pi/2]$ and*

$$A_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

1. $\theta = 0$. *then* $A_\theta = I = \nabla \frac{1}{2}\|\cdot\|^2$ *is cyclically monotone*, $F_I^\infty = \frac{1}{2}\|\cdot\|^2 \oplus \frac{1}{2}\|\cdot\|^2$, *and* $n \geq 2$

$$F_I^n \colon (x, u) \mapsto \frac{n-1}{2n}\Big(\|x\|^2 + \|u\|^2\Big) + \frac{1}{n}\langle x, u\rangle. \quad (6)$$

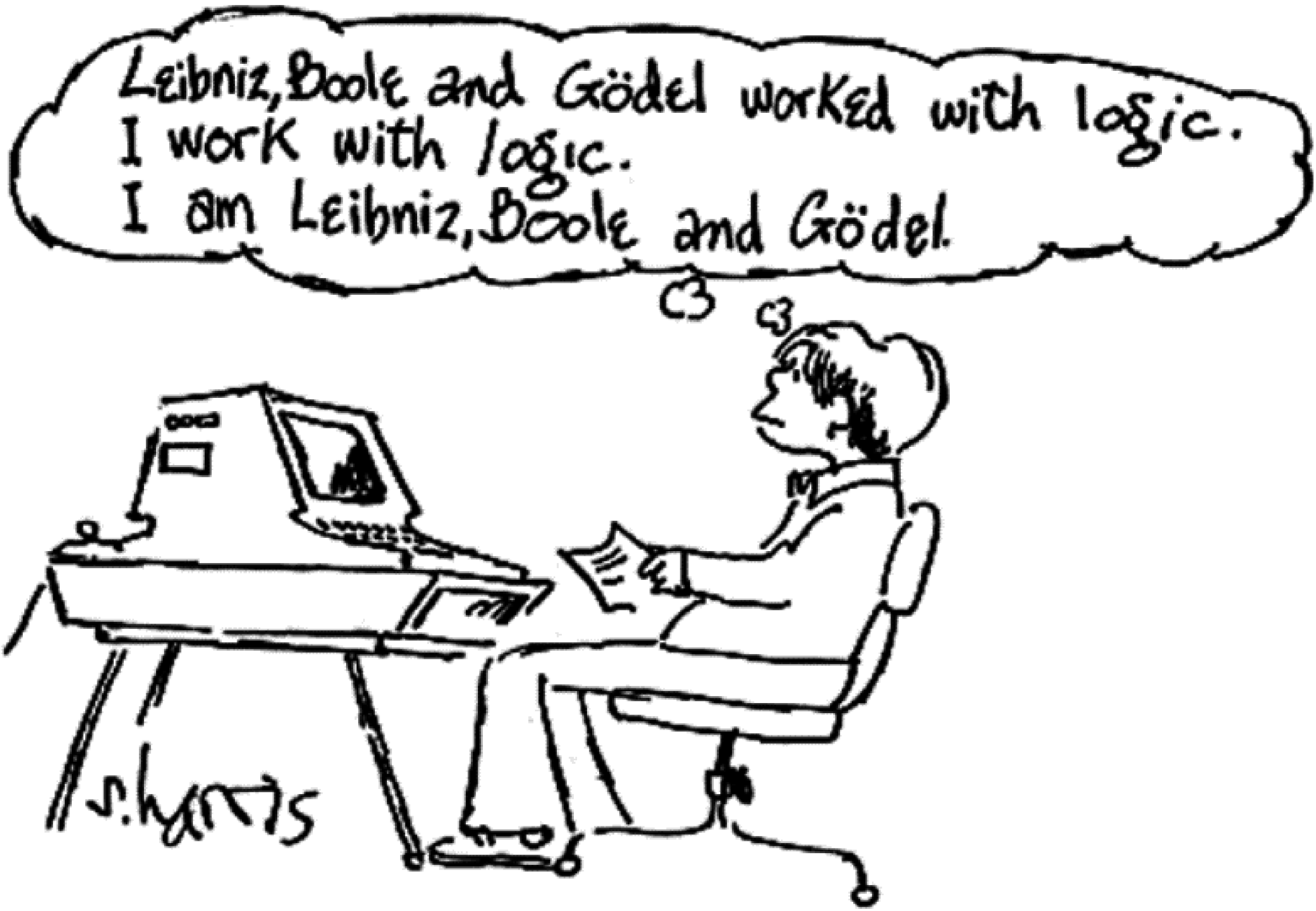
2. $\theta \in]0, \pi/2]$. *For* $n \geq 2$, *if* $n \in [2, \pi/\theta[$, *then* A_θ *is* n*-cyclically monotone and*

$$F_{A_\theta}^n \colon (x, u) \mapsto \frac{\sin(n-1)\theta}{2\sin n\theta}\Big(\|x\|^2 + \|u\|^2\Big) + \frac{\sin\theta}{\sin n\theta}\langle x, A_\theta^{n-1}u\rangle. \quad (7)$$

For $\pi/\theta \in \mathbb{N}$, A_θ *is* (π/θ)*-monotone and*

$$F_{A_\theta}^{\pi/\theta} = \iota_{\text{Graph}\, A_\theta} + \langle\cdot,\cdot\rangle. \quad (8)$$

If $n \in]\pi/\theta, +\inf[$, *then* A_θ *is not* n*-cyclically monotone since* $F_{A_\theta}^n \equiv +\infty$.


Leibniz, Boole and Gödel worked with logic.
I work with logic.
I am Leibniz, Boole and Gödel.
s.harris

4. Maximality in Reflexive Banach Space

We begin with:

Proposition 7 *A monotone operator T on a reflexive Banach space is maximal iff the mapping $T(\cdot + x) + J$ is surjective for all x in X.*

Moreover, when J and J^{-1} are both single valued, a monotone mapping T is maximal if and only if $T + J$ is surjective.

Proof. We prove the 'if'. The 'only if' is completed in Corollary 8. Assume (w, w^*) is monotonically related to the graph of T. By hypothesis, we may solve $w^* \in T(x + w) + J(x)$. Thus $w^* = t^* + j^*$ where $t^* \in T(x + w), j^* \in J(x)$. Hence,

$$\begin{aligned} 0 &\leq \langle w - (w + x), w^* - t^* \rangle \\ &= -\langle x, w^* - t^* \rangle = -\langle x, j^* \rangle = -\|x\|^2 \leq 0. \end{aligned}$$

Thus, $j^* = 0, x = 0$. So $w^* \in T(w)$, and we are done. ∎

We now prove our central result whose proof—originally hard and due to Rockafellar—has been revisited over many years culminating in recent results of Simons, Penot, Zālinescu among others:

Theorem 13 (Sum) *Let X be reflexive, let T be maximal monotone and f closed and convex. Suppose* $0 \in \operatorname{core}\{\operatorname{conv}\operatorname{dom}(T) - \operatorname{conv}\operatorname{dom}(\partial f)\}$. *Then*

(a) $\partial f + T + J$ is surjective.

(b) $\partial f + T$ is maximal monotone.

(c) ∂f is maximal monotone.

Proof. (a) We consider the Fitzpatrick function $\mathcal{F}_T(x, x^*)$ and $f_J(x) := f(x) + 1/2\|x\|^2$.

Let $G(x, x^*) := -f_J(x) - f_J^*(-x^*)$. Observe that

$$\mathcal{F}_T(x, x^*) \geq \langle x, x^* \rangle \geq G(x, x^*)$$

pointwise thanks to the *Fenchel-Young inequality*

$$f_J(x) + f_J^*(-x^*) \geq \langle x, -x^* \rangle,$$

for all $x \in X, x^* \in X^*$, along with Proposition 1. The (CQ) assures the *Sandwich theorem* applies to $\mathcal{F}_T \geq G$ since f_J^* is everywhere finite by Prop. 3.

Then there are $w \in X$ and $w^* \in X^*$ such that

$$\mathcal{F}_T(x, x^*) - G(z, z^*) \geq w(x^* - z^*) + w^*(x - z) \quad (9)$$

for all x, x^* and all z, z^*. In particular, for $x^* \in T(x)$ and for all z^*, z we have

$$\begin{aligned} \langle x - w, x^* - w^* \rangle \quad &+ \quad [f_J(z) + f_J^*(-z^*) + \langle z, z^* \rangle] \\ &\geq \quad \langle w - z, w^* - z^* \rangle. \end{aligned}$$

Now use the fact that $-w^* \in \text{dom}\,(\partial f_J^*)$, by Prop. 3, to deduce that $-w^* \in \partial f_J(v)$ for some v and so

$$\begin{aligned} \langle v - w, x^* - w^* \rangle \quad &+ \quad [f_J(v) + f_J^*(-w^*) + \langle v, w^* \rangle] \\ &\geq \quad \langle w - v, w^* - w^* \rangle = 0. \end{aligned}$$

The second term on the left is zero and so by maximality $w^* \in T(w)$. Substitution of $x = w$ and $x^* = w^*$ in (9), and rearranging yields

$$\begin{aligned} \langle w, w^* \rangle \quad &+ \quad \{\langle -z^*, w \rangle - f_J^*(-z^*)\} \\ &+ \quad \{\langle z, -w^* \rangle - f_J(z)\} \leq 0, \end{aligned}$$

for all z, z^*. Taking the supremum over z and z^* produces $\langle w, w^* \rangle + f_J(w) + f_J^*(-w^*) \leq 0$.

This shows $-w^* \in \partial f_J(w) = \partial f(w) + J(w)$ via the sum formula for subgradients, implicit in Prop. 3.

Thus, $0 \in (T + \partial f_J)(w)$. As all translations of $T + \partial f$ may be used, while (CQ) is undisturbed, we see that $(\partial f + T)(x + \cdot) + J$ is surjective which completes (a).

(b) $\partial f + T$ is maximal by Proposition 7.

(c) Setting $T \equiv 0$ we recover the reflexive case of the maximality for a lsc convex function. ■

Recall that the *normal cone* $N_C(x)$ to a closed convex set C at a point x in C is $N_C(x) = \partial \iota_C(x)$.

Corollary 6 *The sum of a maximal monotone operator T and a (necessarily maximal) normal cone N_C on a reflexive space is maximal monotone whenever the transversality condition*

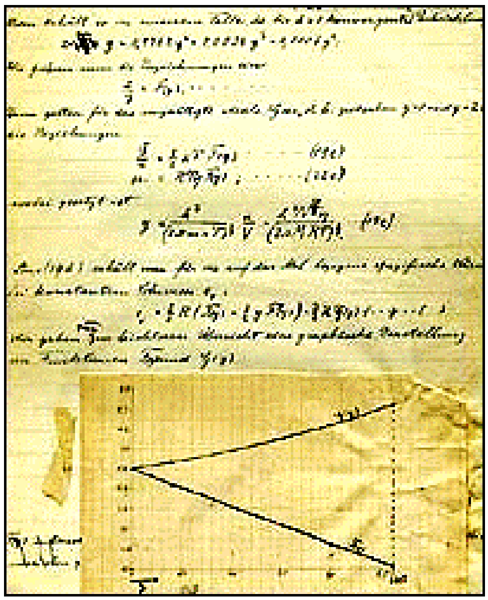
$$0 \in \text{core}\,[C - \text{conv}\,\text{dom}\,(T)]$$

holds.

- In particular, if T is monotone and

$$C := \mathrm{cl\,conv\,dom}\,(T)$$

has nonempty interior, then for any maximal extension $\overline{T}$ the sum $\overline{T} + N_C$ is a *'domain preserving'* maximal monotone extension of T.



Einstein, 1924

- "Quantentheorie des einatomigen idealen Gases"

- On Bose-Einstein condensates, in Paul Ehrenfest' papers in Leiden. Confirmed in 1995.

Corollary 7 (Rockafellar) *The sum of maximal monotone operators T_1 and T_2, on a reflexive space, is maximal when the transversality condition*
$0 \in \text{core}\,[\text{conv dom}\,(T_1) - \text{conv dom}\,(T_2)]$ *holds.*

Proof. Theorem 13 applies to the product $T(x,y) := (T_1(x), T_2(y))$ and the indicator function $f(x,y) := \iota_{\{x=y\}}$ of the diagonal in $X \otimes X$.

We check that the given transversality condition implies the needed (CQ), as in Theorem 13. Hence, $T + J_{X \otimes X} + \partial \iota_{\{x=y\}}$ is surjective. Thus, so is

$$T_1 + T_2 + 2\,J$$

and we are done. ■

- One may easily replace the core condition by a relativized version—wrt the closed affine hull.

We re-record that $\mathcal{F}_{\partial f}(x, x^*) \leq f(x) + f^*(x^*)$, and that we have exploited the beautiful inequality

$$\mathcal{F}_T(x, x^*) + f(x) + f^*(-x^*) \geq 0, \qquad (10)$$

for all $x \in X, x^* \in X^*$, valid for *any* maximal monotone T and *any* convex function f.

Ludolph's Rebuilt Tombstone in Leiden

Ludolph van Ceulen (1540-1610)

- Tombstone reconsecrated July 5, 2000.

4a. The Fitzpatrick Inequality

We have a *stronger* *Fitzpatrick inequality*

$$\mathcal{F}_{T_1}(x, x^*) + \mathcal{F}_{T_2}(x, -x^*) \geq 0 \qquad (11)$$

for all $x \in X, x^* \in X^*$, valid for *any* maximal monotone T_1, T_2. By Proposition 1

$$\begin{aligned} \mathcal{F}_T^*(x^*, x) &\geq \sup_{y^* \in T(y)} \langle x, y^* \rangle + \langle x^*, y \rangle - \mathcal{F}_T(y, y^*) \\ &= \mathcal{F}_T(x, x^*) \qquad (12) \end{aligned}$$

and we clearly have an extension of (11) in that

$$\mathcal{H}_T^1(x, x^*) + \mathcal{H}_S^2(x, -x^*) \geq 0,$$

for any representative functions $\mathcal{H}_T^1$ and $\mathcal{H}_S^2$. Letting $\widehat{\mathcal{F}_S}(x, x^*) := \mathcal{F}_S(x, -x^*)$, we may establish:

Theorem 14 (Sums) *Let S and T be maximal monotone on a reflexive space. Suppose that**

$0 \in \operatorname{core}\{\operatorname{dom}(\mathcal{F}_T) - \operatorname{dom}(\widehat{\mathcal{F}_S})\}$ *as happens if*
$0 \in \operatorname{core}\{\operatorname{conv}\operatorname{graph}(T) - \operatorname{conv}\operatorname{graph}(-S)\}$.

Then

$$0 \in \operatorname{range}(T + S).$$

*This works for any representative functions.

Proof. We use Fenchel duality or follow the steps of Theorem 13. We have $\mu \in X, \lambda \in X^*, \beta \in \mathbb{R}$ with

$$\begin{aligned}\mathcal{F}_T(x,x^*) &- \langle x,\lambda\rangle - \langle \mu,x^*\rangle + \langle \mu,\lambda\rangle \geq \beta \\ &\geq -\mathcal{F}_S(y,-y^*) + \langle y,\lambda\rangle - \langle \mu,y^*\rangle - \langle \mu,\lambda\rangle,\end{aligned}$$

for all variables x, y, x^*, y^*. Hence for $x^* \in T(x)$ and $-y^* \in S(y)$ we obtain

$$\langle x-\mu, x^*-\lambda\rangle \geq \beta \geq \langle y-\mu, y^*+\lambda\rangle.$$

If $\beta \leq 0$, we derive that $-\lambda^* \in S(\mu)$ and so $\beta = 0$; consequently, $\lambda \in T(\mu)$ and since $0 \in (T+S)(\mu)$ we are done. If $\beta \geq 0$ we argue first with T. ■

- A graph (CQ) is formally tougher than a domain (CQ) as conv graph (J_{ℓ^2}) is the diagonal in $\ell^2 \otimes \ell^2 = \operatorname{dom}(F_{J_{\ell^2}})$, while

$$\mathcal{F}_{J_{\ell^2}}(x,x^*) = \frac{1}{4}\|x+x^*\|^2,$$

yielding a simple proof in ℓ^2 of Cor. 8 below.

- Zalinescu has adapted this to extend results like those of Simons in the reflexive case: the sum has a *semi-convex graph*.

Corollary 8 (Rockafellar-Minty surjectivity theorem) *For a maximal monotone operator on a reflexive Banach space,* range $(T+J) = X^*$.

Proof. Let $f \equiv 0$ in Theorem 13. Alternatively, on noting that $\mathcal{F}_J(x,x^*) \leq \frac{\|x\|^2+\|x^*\|^2}{2}$, we may apply Theorem 14. ∎

4b. Extensions to Non-reflexive Space

Let $\overline{T}$ denote the *monotone closure* of T in $X^{**} \times X^*$. That is, $x^* \in \overline{T}(x^{**})$ when

$$\inf_{y^* \in T(y)} \langle x^* - y^*, x^{**} - y \rangle \geq 0.$$

Recall that T is *type (NI)* if

$$\inf_{y^* \in T(y)} \langle x^* - y^*, x^{**} - y \rangle \leq 0$$

for all $x^{**} \in X^{**}, x^* \in X^*$:

Corollary 9 (Gossez for (D)). *For T type (NI)*

$$R(\overline{T} + \partial f^{**} + J^{**}) = X^*.$$

Proof. Mimic the steps of Theorem 13. ∎

4c. A Non-reflexive Sum Rule

Theorem 15 *Suppose that A and B are maximal monotone in Banach space. If either*

a) $\operatorname{int} D(A) \cap \operatorname{int} D(B)$ *is nonempty;*

b) $\operatorname{int} D(A) \cap D(B) \neq \emptyset$ *while $D(B)$ is closed and convex; or*

c) (Voisei) Both $D(A), D(B)$ are closed and convex and

$$0 \in \operatorname{core} \operatorname{conv} \{D(A) - D(B)\}. \tag{13}$$

Then $A + B$ is maximal monotone.

Let

$$\Phi_{A,B}(x, x^*) := \inf_{\{u^*+v^*=x^*\}} \{\mathcal{F}_A(x, u^*) + \mathcal{F}_B(x, v^*)\}$$

$$\Psi_{A,B}(x, x^*) := \inf_{\{u^*+v^*=x^*\}} \{\mathcal{P}_A(x, u^*) + \mathcal{P}_B(x, v^*)\}.$$

Proof. Voisei (2005) shows, as in §5, that (13) implies the lower-semicontinuity and attainment of $\Phi_{A,B}$ as the conjugate of $\Psi_{A,B}$. Hence

$$\Phi_{A,B}(x, x^*) \geq \langle x, x^* \rangle$$

with equality if and only if $x^* \in (A + B)(x)$.

Moreover,

$$\mathcal{F}_{A+B} \leq \Phi_{A,B} \leq \mathcal{P}_{A+B}.$$

Hence $A + B$ is maximal **iff**

$$\mathcal{F}_{A+B}(x, x^*) \geq \langle x, x^* \rangle, \tag{14}$$

for all x, x^*. Now all three conditions imply that

$$\overline{\text{conv}}\, D(A) \cap \overline{\text{conv}}\, D(B) \subset \overline{D(A+B)}^{alg},$$

since $\overline{D(A)}$ is convex when $D(A)$ has nonempty interior. This in turn implies (14). ■

Corollary 10 *Suppose that T is maximal monotone, C is closed and convex while $C \cap \text{int}\, D(T) \neq \emptyset$.*

Then $T + N_C$ is maximal monotone.

In particular, when $D(T)$ has nonempty interior, then T is of type (FPV).

4d. The Case of a Subgradient

We can significantly improve the result in this case:

Theorem 16 *Suppose T is maximal monotone and f is convex and closed. Suppose* $\operatorname{dom} f \cap \operatorname{int} D(T)$ *is nonempty. Then $T + \partial f$ is maximal.*

Proof. We use $\mathcal{F}_{T,f}(x, x^*) :=$

$$f(x) + \sup_{y^* \in T(y)} \{\langle x, y^* \rangle + \langle y, x^* \rangle - \langle y, y^* \rangle - f(y)\},$$

with conjugate $\mathcal{P}_{T,f}(x, x^*) :=$

$$f(x) + \overline{\text{conv}}_{y_i^* \in T(y_i)} \{\langle y_i, y_i^* \rangle - f(y_i)\}.$$

We define $\mathcal{V}_{T,f}(x, x^*) := f(x) + (\mathcal{F}_{T,f}(x, \cdot) \Box f^*)(x^*)$. Then

$$\mathcal{F}_{T,f}(x, x^*) \leq \mathcal{V}_{T,f}(x, x^*) \leq \mathcal{P}_{T,f}(x, x^*),$$

and (a) $\mathcal{F}_{T,f}(x, x^*) \leq \langle x, x^* \rangle$ for (x, x^*) monotonically related to Gr$(T + \partial f)$ while (b) $\mathcal{V}_{T,f}(x, x^*) \geq \langle x, x^* \rangle$ for all x, x^* with equality exactly for $x^* \in T(x) + \partial f(x)$. (c) The (CQ) ensures $\mathcal{F}_{T,f}$ represents $T + \partial f$. As before

$$\mathcal{F}_{T,f}(x, x^*) = \langle x, x^* \rangle \Rightarrow \mathcal{P}_{T,f}(x, x^*) = \langle x, x^* \rangle.$$

[Note: $\mathcal{F}_{T,0} = \mathcal{F}_T \quad \mathcal{P}_{T,0} = \overline{\mathcal{P}}_T$] ■

5. Further Reflexive Applications

Another very useful result is:

Theorem 16 (Composition) *Suppose X and Y are Banach spaces with X reflexive, that T is maximal monotone operator on Y, and that $A\colon X \mapsto Y$, is a bounded linear mapping.* Then

$$T_A := A^* \circ T \circ A$$

is maximal monotone on X whenever

$$0 \in \mathrm{core}\,(\mathrm{range}\,(A) + \mathrm{conv}\,\mathrm{dom}\,T)$$

Proof. Monotonicity is clear. To obtain maximality, use the Fitzpatrick inequality (11) to write

$$f(x, x^*) + g(x, x^*) \geq 0,$$

where

$$f(x, x^*) := \inf\{\mathcal{F}_T(Ax, y^*)\colon A^*y^* = x^*\}$$

and

$$g(x, x^*) := \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^*\|^2.$$

Apply Fenchel's duality theorem—or use the Sandwich theorem directly—to deduce the existence of $\overline{x} \in X, \overline{x}^* \in X^*$ with

$$f^*(\overline{x}^*, \overline{x}) + g^*(\overline{x}^*, \overline{x}) \leq 0. \tag{15}$$

Carefully using the standard formula for the conjugate of a convex composition —we have for some $\overline{y}^*$ with $A^*\overline{y}^* = \overline{x}^*$:

$$\begin{aligned} f^*(\overline{x}^*, \overline{x}) &= \inf\{\mathcal{F}_T^*(A\overline{x}, y^*)\colon A^*y^* = \overline{x}^*\} \\ &= \min\{\mathcal{F}_T^*(y^*, A\overline{x})\colon A^*y^* = \overline{x}^*\} \\ &= \mathcal{F}_T^*(\overline{y}^*, A\overline{x}) \geq \mathcal{F}_T(A\overline{x}, \overline{y}^*), \end{aligned}$$

the last inequality following from (12). Moreover,

$$g^*(\overline{x}^*, \overline{x}) = \frac{1}{2}\|\overline{x}\|^2 + \frac{1}{2}\|A^*\overline{y}^*\|^2.$$

Thus, (15) implies that

$$\begin{aligned} &\left\{\mathcal{F}_T(A\overline{x}, \overline{y}^*) - \langle \overline{y}^*, A\overline{x}\rangle\right\} \\ + &\left\{\frac{1}{2}\|\overline{x}\|^2 + \frac{1}{2}\|A^*\overline{y}^*\|^2 + \langle \overline{y}^*, A\overline{x}\rangle\right\} \leq 0. \end{aligned}$$

We see that $\overline{y}^* \in T(A\overline{x})$, $-\overline{x}^* := -A^*\overline{y}^* \in J_X(\overline{x})$, since both bracketed terms are non-negative. Hence,

$$0 \in J_X(\overline{x}) + T_A(\overline{x}).$$

In the same way if we start with

$$f(x,x^*) := \inf\{\mathcal{F}_T(Ax,y^*)\colon A^*y^* = x^* + x_0^*\},$$

$$g(x,x^*) := \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^*\|^2 - \langle x, x_0^*\rangle,$$

we deduce, $x_0^* \in J_X(\overline{x}) + T_A(\overline{x})$. This applies to all *domain* translations of T. As in Theorem 13, this is sufficient to conclude T_A is maximal. ■

- This recovers the reflexive case of the formula that $A^*\partial f(Ax) = \partial(fA)(x)$ with the same (CQ).

- A recent paper [Bot et al] relaxes the (CQ) to

$$\{(A^*y^*, Ax, r)\colon \mathcal{F}_T^*(Ax,y^*) \le r\} \qquad (16)$$

  is relatively closed in $X^* \times R(A) \times \mathbb{R}$.

- Application of Theorem 16 to

$$T(x,y) := (T_1(x), T_2(y)),$$

  and $A(x) := (x,x)$ yields $T_A(x) = T_1(x) + T_2(x)$ and recovers Theorem 13. With more effort one may equally embed Theorem 16 in Theorem 13.

Note only X need be reflexive. A key case of Theorem 16 is a *reflexive injection*.

Corollary 11 *Let T be maximal monotone on a Banach space Y. Let ι denote the injection of a reflexive subspace $Z \subset Y$ into Y.*
Then $T_Z := \iota^* \circ T \circ \iota$ is maximal monotone on Z if

$$0 \in \operatorname{core}(Z + \operatorname{conv}\operatorname{dom} T).$$

Hence, if $0 \in \operatorname{core}(\operatorname{conv}\operatorname{dom} T)$, then T_Z is maximal for each reflexive Z.

- In this case, (16) implies the result holds when

$$\{(y^*|_Z, z, r) \colon \mathcal{H}_T^*(z, y^*) \le r, z \in Z\}$$

  is relatively closed in $Z^* \times Z \times \mathbb{R}$
  *What happens generally?**

***Conjecture**: 'most' subspaces behave well $\Rightarrow T$ is (FPV) and so $\overline{D(T)}$ convex.

Conjectural Details

1. For a lsc representative $\mathcal{H}_T$ and dim $F < \infty$, if
$$\mathcal{H}_T^F(y, y^*) := \inf\{\mathcal{H}_T(y, x^*) \colon x^*|F = y^*\}$$
is lsc on $F \times F^*$ then T_F is maximal.

2. Equivalently, this holds if
$$\operatorname{epi}\mathcal{H} + \{0\} \times F^{\perp} \times \{0\} \qquad (17)$$
is closed.

3. Hence, if (17) holds for 'most' F meeting $\operatorname{dom} T$, we have a net of approximating 'nice' maximal monotone (e.g., FPV, FP) operators.

Example 1 *Consider* $T(x_1, x_2) := \partial f(x_1, x_2)$ *and* $\mathcal{H}_T(x_1, x_2, x_1^*, x_2^*) := f(x_1, x_2) + f^*(x_1^*, x_2^*)$ *where*
$$f(x_1, x_2) := \max\{|x_1|, 1 - \sqrt{x_2}\}, \qquad x_2 \geq 0$$
$$f^*(x_1^*, x_2^*) = \frac{\{(|x_1^*| - 1) \vee x_2^*\}^2}{4x_2} - (|x_1^*| - 1) \vee x_2^*,$$
and $|x_1^*| \leq 1, x_2^* < 0$. *Then (only)* $T_{R\times 0}$ *is not maximal and, necessarily,* $\mathcal{H}_T^{R\times 0}$ *is not lsc.*

A Dense Limiting Example

Example 2 *Let C be closed convex and bounded in an infinite dimensional Banach space X and fix $x_0 \neq 0$ in X. Define*

$$f_C(x) := \inf\{t \in \mathbb{R} \colon x + t\, x_0 \in C\}.$$

Set $c_x := x - f_C(x) x_0 \in C$. Then f_C is closed and convex and has no global minimum. Moreover, $\partial f_C(x) = \partial f_C(c_x)$. This implies that

$$\operatorname{dom} \partial f_C \subset \operatorname{supp} C + \mathbb{R} x_0$$

Now arrange that $0 \in C$, that

$$Y \bigcap \operatorname{span}\,(C \cup \{x_0\}) = 0$$

for a dense subspace Y, while span C is also dense. It follows that $(\partial f_C)_F$ fails to be maximal for every non-trivial finite dimensional subspace $F \subset Y$.

Explicitly, take the (norm-compact) Hilbert cube $K := \{x \in \ell_2 \colon |x_n| \leq 1/2^n, \forall n \in \mathbb{N}\}$ and $x_0 := (1/2^n)$ so that

$$f_K(x) := \sup_{n \in \mathbb{N}} |2^n\, x_n - 1|,$$

and take $Y \setminus \{0\}$ to contain only more slowly decreasing sequences.

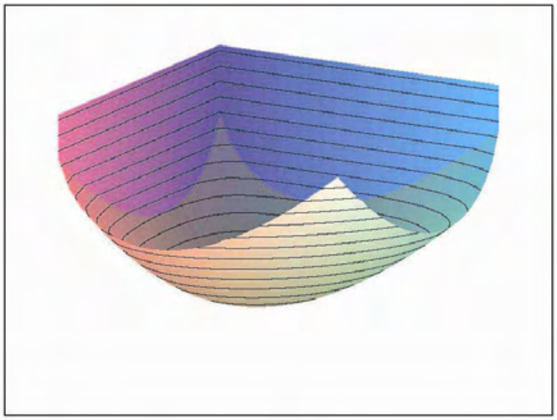
5a. Variational Inequalities

T is *coercive on C* if

$$\inf_{y^* \in T(y) + \partial\iota_C(y)} \langle y, y^* \rangle / \|y\| \to \infty$$

as $y \in C$ goes to infinity in norm.[a]

[a]This may be weakened significantly, especially if $0 \in C$.



A *variational inequality* **V(T,C)** requests a solution $y \in C$ and $y^* \in T(y)$ to

$$\langle y^*, x - y \rangle \geq 0 \qquad \forall x \in C.$$

Equivalently

$$0 \in T(y) + N_C(y)$$

or

$$0 \in T(y) + \partial\iota_C(y).$$

- This models the *necessary condition*

  $$\langle \nabla f(x), c - x \rangle \geq 0$$

  for all $c \in C$.

Corollary 12 *Suppose T is maximal monotone on a reflexive space and is coercive on the closed convex set C while $0 \in \mathrm{core}\,(C - \mathrm{conv}\,\mathrm{dom}\,(T))$. Then $V(T,C)$ has a solution.*

Proof. Let $f := \iota_C$, the indicator function. For $n = 1, 2, 3 \cdots$, let $T_n := T + J/n$. We solve

$$0 \in (T_n + \partial\iota_C)\,(y_n) = (T + \partial\iota_C) + \frac{1}{n}J(y_n) \quad (18)$$

and take limits as n goes to infinity.

More precisely, Theorem 13, yields y_n in C, and $y_n^* \in (T + \partial\iota_C)\,(y_n), j_n^* \in J(y_n)/n$ with $y_n^* = -j_n^*$. Then

$$\langle y_n^*, y_n\rangle = -\frac{1}{n}\langle j_n^*, y_n\rangle = -\frac{1}{n}\|y_n\|^2 \leq 0,$$

and so coercivity of $T + \partial\iota_C$ implies that $\|y_n\|$ remains bounded and so $j_n^* \to 0$. We may assume $y_n \rightharpoonup y$.

Since $T + \partial\iota_C$ is maximal monotone (again by Theorem 13), it is demi-closed. It follows that $0 \in (T + \partial\iota_C)(y)$, and y is as required. ■

Letting $C := X$ in Corollary 12 we deduce

Corollary 13 *Every coercive maximal monotone operator on a Banach space is surjective if (and only if) the space is reflexive.*

Proof. To complete the proof we recall that, by *James' theorem*, surjectivity of J is equivalent to reflexivity of the corresponding space. ■

We may improve Corollary 3 in the reflexive setting:

Theorem 17 *Suppose T is maximal monotone on a reflexive space. Then* dom (T) *and range (T) have convex closure (and interior).*

Proof. Without loss, we assume 0 is in the closure of conv dom (T). Fix $y \in$ dom (T), $y^* \in T(y)$. Corollary 8 applied to T/n solves $w_n^*/n + j_n^* = 0$ with $w_n^* \in T(w_n), j_n^* \in J(w_n)$, for integer $n > 0$. By monotonicity

$$\frac{1}{n}\langle y^*, y - w_n\rangle \geq \frac{1}{n}\langle w_n^*, y - w_n\rangle = \|w_n\|^2 - \langle j_n^*, y\rangle$$

where $\|w_n\|^2 = \|j_n^*\|^2 = \langle j_n^*, w_n\rangle$ and $w_n \in$ dom (T).

We deduce $\sup_n \|w_n\| < \infty$. Thus, (j_n^*) has a weak cluster point j^*. Thence, denoting $D := \operatorname{dom}(T)$

$$\begin{aligned} d_D^2(0) &\leq \liminf_{n\to\infty} \|w_n\|^2 \leq \inf_{y\in D} \langle j^*, y\rangle \\ &= \inf_{y\in \operatorname{conv} D} \langle j^*, y\rangle \leq \|j^*\|\, d_{\operatorname{conv} D}(0) = 0. \end{aligned}$$

We have shown that $\operatorname{cl}\operatorname{conv}\operatorname{dom}(T) \subset \operatorname{cl}\operatorname{dom}(T)$ and so $\operatorname{cl}\operatorname{dom}(T)$ is convex as required.
As $\operatorname{range}(T) = \operatorname{dom}(T^{-1})$ and X^* is reflexive we are done. ■

More generally:

Theorem 18 (**Fitzpatrick, Phelps**) *Every locally maximal monotone operator on a Banach space has* $\operatorname{cl}\operatorname{range} T$ *convex.*

Proof. We suppose not and then that there are $\pm x^*$ in $\operatorname{cl}\operatorname{range} T$ of unit-norm but with midpoint $0 \notin \operatorname{cl}\operatorname{range} T$.

Proof. We build the ball

$$B' := \text{conv}\,\{\pm 2x^*, \alpha\, B_X^*\}$$

where $0 < \alpha < 1/2$ is chosen with

$$(\text{range}\, T) \cap 2\alpha B_X^* = \emptyset.$$

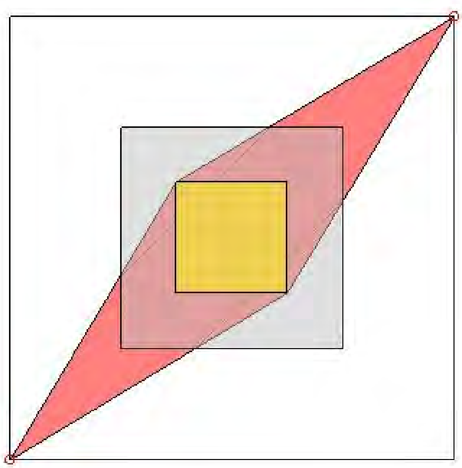
We extend $T \cap B'$ as in Prop. 5, so that

$R(\widehat{T}) \subset \text{cl conv}\,\{R(T) \cap B'\}$ and $R(\widehat{T}) \setminus R(T) \subset \text{bd}\, B'$.

It follows that

$\text{range}\,\widehat{T} \subset (R(T) \cap B') \bigcup (\text{cl conv}\,\{R(T) \cap B'\} \cap \text{bd}\, B')$.

Hence range $\widehat{T}$ is weak-star disconnected. As $\widehat{T}$ is a weak-star cusco it has a weak-star connected range which contradicts the construction. ■



B' (red), $\alpha\, B_{X^*}$ (yellow) and $2\alpha\, B_{X^*}$ (grey)

Corollary 14 *Suppose T is maximal monotone on a reflexive Banach space X and is locally bounded at each point of* cl dom(T). *Then* dom$(T) = X$.

Proof. Observe dom(T) must be closed and so convex. By the Bishop-Phelps theorem, there is some boundary point $\overline{x} \in$ dom(T) with a non-zero support functional $\overline{x}^*$.
Then $T(\overline{x}) + [0, \infty)\,\overline{x}^*$ is monotonically related to the graph of T. By maximality

$$T(\overline{x}) + [0, \infty)\,\overline{x}^* = T(\overline{x})$$

which is non-empty and (linearly) unbounded. ■

6. Limiting Examples and Constructions

- It is unknown outside reflexive space whether $\operatorname{cl}\operatorname{dom}(T)$ must always be convex for a maximal monotone operator

- Reflexivity in Theorem 17 may be relaxed to $R(T+J)$ is boundedly w^*-dense—as an examination of the proof will show

We do however have the following result:

Theorem 19 (JB-SF-Vanderwerff) *TFAE.*

(a) *A Banach space X is reflexive*

(b) $\operatorname{int}\operatorname{range}(\partial f)$ *is convex for each coercive lsc convex function f on X*

(c) $\operatorname{int}\operatorname{range}(T)$ *is convex for each coercive maximal monotone mapping T.*

Proof. Suppose X is nonreflexive and $p \in X$ with $\|p\| = 5$ and $p^* \in Jp$ where J is the duality map. Define

$$f(x) := \max\left\{\frac{1}{2}\|x\|^2, \|x \mp p\| - 12 \pm \langle p^*, x\rangle\right\}$$

for $x \in X$. By the max-formula, for $x \in B_X$,

$$\partial f(\pm p) = B_{X^*} \pm p^*, \partial f(x) = Jx \tag{19}$$

using inequalities like $\|p - p\| - 12 + \langle p^*, p\rangle = 13 > \frac{25}{2} = \frac{1}{2}\|p\|^2$.

Moreover, $f(0) = 0$ and $f(x) > \frac{1}{2}\|x\|$ for $\|x\| > 1$, thus $\|x^*\| > \frac{1}{2}$ if $x^* \in \partial f(x)$ and $\|x\| > 1$. Combining this with (19) shows

$$\text{range}\,(\partial f) \cap \frac{1}{2}B_{X^*} = \text{range}\,(J) \cap \frac{1}{2}B_{X^*}.$$

Let $U := U_{X^*}$ denote the open unit ball in X^*. Now James' theorem gives $x^* \in \frac{1}{2}U_{X^*} \setminus \text{range}\,(J)$, thus $U_{X^*} \setminus \text{range}\,(\partial f) \neq \emptyset$. However, from (19)

$$U \subset \text{conv}\,\{(p^* + U) \cup (-p^* + U)\} \subset \text{conv}\,\text{int}\,R(\partial f)$$

so range$\,(\partial f)$ has non-convex interior. This shows that (b) implies (a) while (c) implies (b) is clear.

Finally (a) $\Rightarrow$ (c) follows from Theorem 17. ■

- Every locally maximal operator T has $\operatorname{cl}\operatorname{range} T$ convex (Fitzpatrick-Phelps)

Observe the two roles of convexity in the proof of (a) $\Leftrightarrow$ (c). One often uses the same logic to establish a result of the form

> *"Property P holds for all maximal monotone operators if and only if X is a Banach space with property Q."*

Two other examples are:

- *"Every monotone operator T on a space X is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} T$ *iff X is finite dimensional."*

- *"Every monotone operator T on a space X is single valued and norm-continuous on a generic subset of* $\operatorname{int}\operatorname{dom} T$ *iff X is an Asplund space."*

Example 3 *Most explicitly Fitzpatrick and Phelps used c_0, the space of null sequences, and*

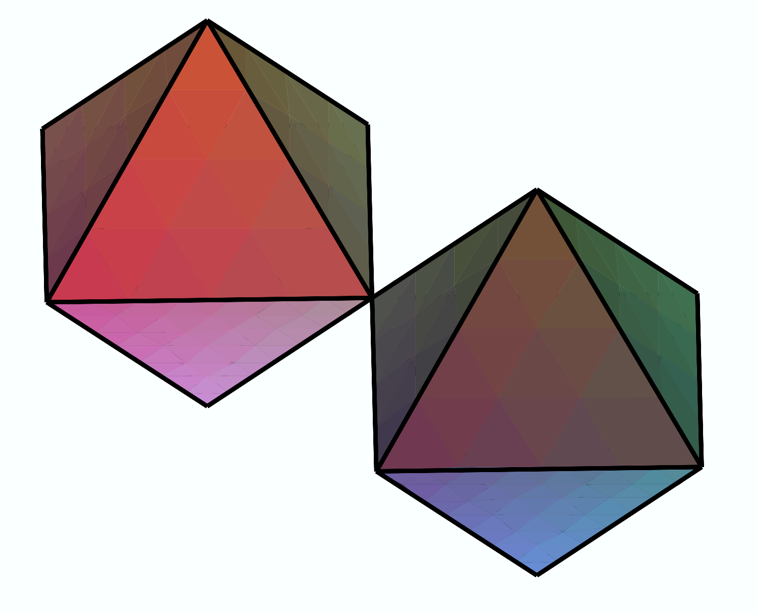
$$f(x) := \|x - e_1\|_\infty + \|x + e_1\|_\infty \tag{20}$$

where e_1 is first unit vector. Then int range ∂f is not convex (disconnected):

$$\text{int range}(\partial f) = \left\{U_{\ell_1} + e_1\right\} \cup \left\{U_{\ell_1} - e_1\right\}$$

$$\text{cl-int range}(\partial f) = \left\{B_{\ell_1} + e_1\right\} \cup \left\{B_{\ell_1} - e_1\right\}$$

both of which are far from convex. ■



The range of ∂f in ℓ^1

▼ It is instructive to compute cl-range(∂f)

Example 4 *Gossez gives a coercive maximal monotone operator T with full domain whose range has a non-convex closure.*

T is of the form $2^{-n} J_{\ell^1} + S$ for some $n > 0$ large with bounded linear $S : \ell_1 \to \ell_\infty$ given by

$$(Sx)_n := -\sum_{k<n} x_k + \sum_{k>n} x_k, \quad \forall x = (x_k) \in \ell_1, n \in N.$$

In fact, $\mp S : \ell_1 \mapsto \ell_\infty$ is skew bounded and S^* is not monotone but $-S^*$ is.

- Hence, $-S$ is both of dense type and locally maximal monotone (also called FP) while S is in neither class (Bauschke-JMB) ■

- Relatedly, let ι be the injection of ℓ^1 into ℓ^∞. For small $\epsilon > 0$

  $$S_\varepsilon := \varepsilon\iota + S$$

  is a coercive maximal monotone operator for which the closure $\overline{S_\varepsilon}$ fails to be coercive in X^{**}.

  One may use a smooth renorming of ℓ_1. This means $T + \lambda J$ is single-valued, demicontinuous.

Example 5 (Some further related results) *More abstractly, one can show that if the underlying space X is* rugged, *meaning* $\operatorname{cl}\operatorname{span}\operatorname{range}(J - J) = X^*$, *then the following are equivalent whenever T is bounded linear and maximal monotone:*

i) T *is of dense type.*

ii) $\operatorname{cl} - \operatorname{range}(T + \lambda J) = X^*$, $\forall \lambda > 0$.

iii) $\operatorname{cl} - \operatorname{range}(T + \lambda J)$ *is convex,* $\forall \lambda > 0$.

iv) $T + \lambda J$ *is locally maximal monotone,* $\forall \lambda > 0$.

- Equivalences i)–iv) hold for the following rugged spaces: c_0, c, ℓ_1, ℓ_∞, $L_1[0,1]$, $L_\infty[0,1]$, $C[0,1]$.

  In cases like c_0 or $C[0,1]$, which contain no complemented copy of ℓ_1, a maximal monotone bounded linear T is always of dense type.*

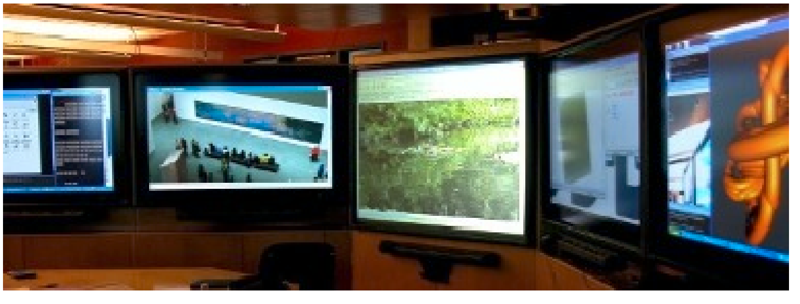
  In particular, S in Example 4 is necessarily not of dense type, etc.

*SF and JMB spent several weeks in 1994 looking for a counter-example in C[0,1].

7. Conclusion

Fitzpatrick's function was built to provide a transparent convex alternative to earlier saddle function constructions of Krauss. His interests were more in differentiation theory for Lipschitz functions.

Results relating when a maximal monotone T is single-valued to differentiability of $\mathcal{F}_T$ were not forthcoming, and he put the function aside.



D-Drive

- This is still the one area where to the best of my knowledge $\mathcal{F}_T$ has proved of little help—in part because generic properties of $\operatorname{dom} \mathcal{F}_T$ and of $\operatorname{dom}(T)$ seem poorly related.

- By contrast, Fitzpatrick's function and its relatives now provide the easiest access to a gamut of solvability and boundedness results.

The clarity of the constructions also offers hope for resolving some of the most persistent open questions about maximal monotone operators such as:

Q3. Must $\mathrm{cl}\,\mathrm{dom}\,(T)$ always be convex? Simons shows this is so for operators of *dual type (FPV)*.

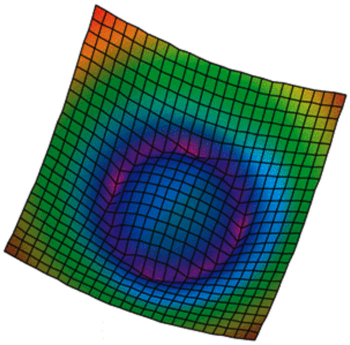
Q4. Can $T_1 + T_2$ fail be maximal when

$$0 \in \mathrm{core}\,\mathrm{conv}\,(\mathrm{dom}\,(T_1) - \mathrm{dom}\,(T_2))?$$

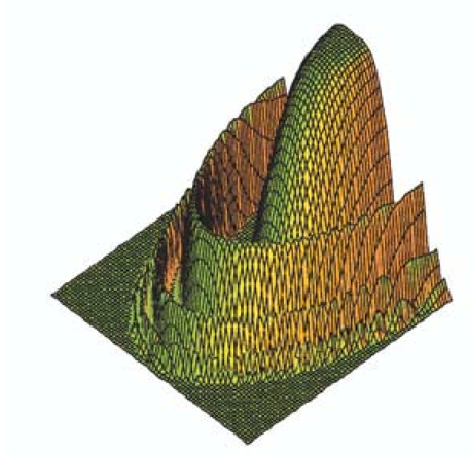
Q5. Given a maximal monotone T, can one associate a convex f_T with T in such fashion that $T(x)$ is singleton as soon as $\partial f_T(x)$ is?

Q6. Are there some nonreflexive spaces, such as c_0, for which such questions can be answered in the affirmative?*

*__Conjecture.__ On c_0 all maximal operators are type (NI).



Non-convex
functions are
hard too ...



References

1. Borwein J.M. and Fitzpatrick, S., "Local boundedness of monotone operators under minimal hypotheses," *Bull. Aust. Math. Soc.*, **39** (1988), 339–441.

2. Borwein J.M., Fitzpatrick S. and Vanderwerff J., "Examples of convex functions and classifications of normed spaces," *J. Convex Analysis*, **1** (1994), 61–73.

3. Fitzpatrick, S. "Representing monotone operators by convex functions," *Workshop on Functional Analysis and Optimization* (Canberra, 1988), 59–65.

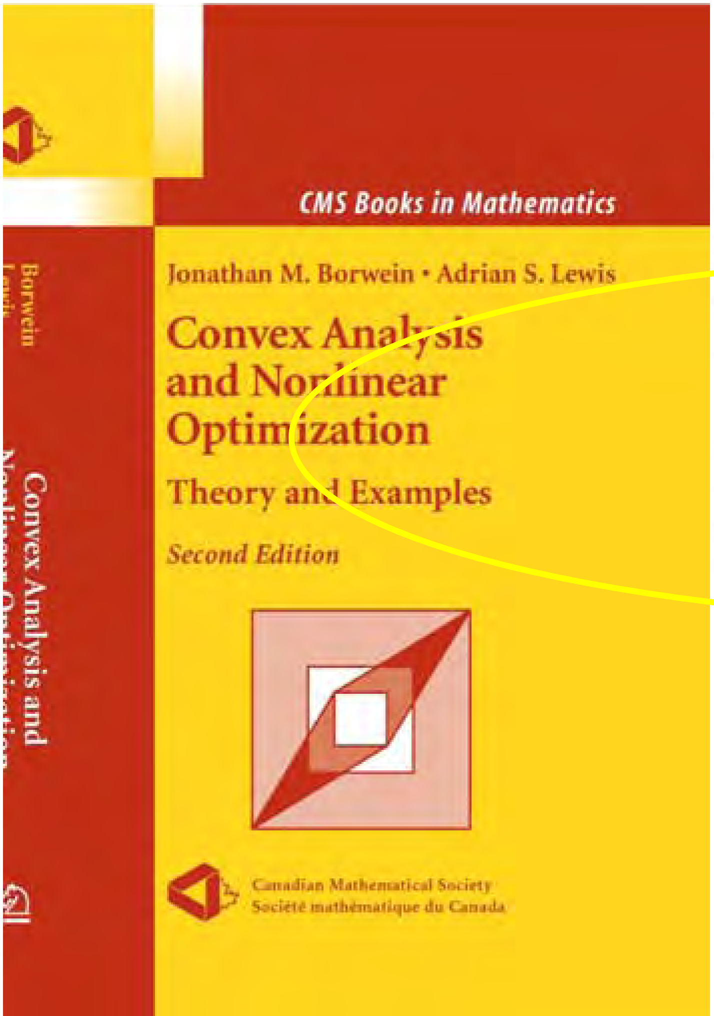
4. Fitzpatrick, S. and Phelps R.R., "Some properties of maximal monotone operators on nonreflexive Banach spaces," *Set-Valued Anal.* **3** (1995), 51–69.

5. Phelps, R. R. *Convex Functions, Monotone Operators and Differentiability,* Lecture Notes in Mathematics, **1364**, Springer-Verlag,Berlin, 1989.

6. Phelps R.R., "Lectures on maximal monotone operators," *arXiv:math.FA/9302209*, (posted 4 February 1993).

Convex Analysis and Nonlinear Optimization

Second Edition



New Material

8.3 now includes Fitzpatrick Function

9.1 Rademacher's Theorem

We mentioned Rademacher's fundamental theorem on the differentiability of Lipschitz functions in the context of the Intrinsic Clarke subdifferential formula (Theorem 6.2.5):

$$\partial_\circ f(x) = \operatorname{conv}\{\lim_r \nabla f(x^r) \mid x^r \to x,\ x^r \notin Q\}, \tag{9.1.1}$$

valid whenever the function $f : \mathbf{E} \to \mathbf{R}$ is locally Lipschitz around the point $x \in \mathbf{E}$ and the set $Q \subset \mathbf{E}$ has measure zero. We prove Rademacher's theorem in this section, taking a slight diversion into some basic measure theory.

Theorem 9.1.2 (Rademacher) *Any locally Lipschitz map between Euclidean spaces is Fréchet differentiable almost everywhere.*

Proof. Without loss of generality (Exercise 1), we can consider a locally Lipschitz function $f : \mathbf{R}^n \to \mathbf{R}$. In fact, we may as well further suppose that f has Lipschitz constant L throughout $\mathbf{R}^n$, by Exercise 2 in Section 7.1.

Fix a direction h in $\mathbf{R}^n$. For any $t \neq 0$, the function g_t defined on $\mathbf{R}^n$ by

$$g_t(x) = \frac{f(x+th) - f(x)}{t}$$

is continuous, and takes values in the interval $I = L\|h\|[-1,1]$, by the Lipschitz property. Hence, for $k = 1, 2, \ldots$, the function $p_k : \mathbf{R}^n \to I$

defined by

$$p_k(x) = \sup_{0<|t|<1/k} g_t(x)$$

is lower semicontinuous and therefore Borel measurable. Consequently, the upper Dini derivative $D_h^+ f : \mathbf{R}^n \to I$ defined by

$$D_h^+ f(x) = \limsup_{t\to 0} g_t(x) = \inf_{k\in\mathbf{N}} p_k(x)$$

is measurable, being the infimum of a sequence of measurable functions. Similarly, the lower Dini derivative $D_h^- f : \mathbf{R}^n \to I$ defined by

$$D_h^- f(x) = \liminf_{t\to 0} g_t(x)$$

is also measurable.

The subset of $\mathbf{R}^n$ where f is not differentiable along the direction h, namely

$$A_h = \{x \in \mathbf{R}^n \mid D_h^- f(x) < D_h^+ f(x)\},$$

is therefore also measurable. Given any point $x \in \mathbf{R}^n$, the function $t \mapsto f(x+th)$ is absolutely continuous (being Lipschitz), so the fundamental theorem of calculus implies this function is differentiable (or equivalently, $x + th \notin A_h$) almost everywhere on $\mathbf{R}$.

Consider the nonnegative measurable function $\phi : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$ defined by $\phi(x,t) = \delta_{A_h}(x+th)$. By our observation above, for any fixed $x \in \mathbf{R}^n$ we know $\int_{\mathbf{R}} \phi(x,t)\,dt = 0$. Denoting Lebesgue measure on $\mathbf{R}^n$ by μ, Fubini's theorem shows

$$0 = \int_{\mathbf{R}^n} \Big(\int_{\mathbf{R}} \phi(x,t)\,dt \Big)\,d\mu = \int_{\mathbf{R}} \Big(\int_{\mathbf{R}^n} \phi(x,t)\,d\mu \Big)\,dt = \int_{\mathbf{R}} \mu(A_h)\,dt$$

so the set A_h has measure zero. Consequently, we can define a measurable function $D_h f : \mathbf{R}^n \to \mathbf{R}$ having the property $D_h f = D_h^+ f = D_h^- f$ almost everywhere.

Denote the standard basis vectors in $\mathbf{R}^n$ by $e_1, e_2, \ldots, e_n$. The function $G : \mathbf{R}^n \to \mathbf{R}^n$ with components defined almost everywhere by

$$G_i = D_{e_i} f = \frac{\partial f}{\partial x_i} \qquad (9.1.3)$$

for each $i = 1, 2, \ldots, n$ is the only possible candidate for the derivative of f. Indeed, if f (or $-f$) is regular at x, then it is easy to check that $G(x)$ is the Fréchet derivative of f at x (Exercise 2). The general case needs a little more work.

Consider any continuously differentiable function $\psi : \mathbf{R}^n \to \mathbf{R}$ that is zero except on a bounded set. For our fixed direction h, if $t \neq 0$ we have

$$\int_{\mathbf{R}^n} g_t(x)\,\psi(x)\,d\mu = \int_{\mathbf{R}^n} f(x) \frac{\psi(x-th) - \psi(x)}{t}\,d\mu.$$

As $t \to 0$, the bounded convergence theorem applies, since both f and ψ are Lipschitz, so

$$\int_{\mathbf{R}^n} D_h f(x)\,\psi(x)\,d\mu = -\int_{\mathbf{R}^n} f(x)\,\langle \nabla\psi(x), h\rangle\,d\mu.$$

Setting $h = e_i$ in the above equation, multiplying by h_i, and adding over $i = 1, 2, \ldots, n$, yields

$$\int_{\mathbf{R}^n} \langle h, G(x)\rangle\,\psi(x)\,d\mu = -\int_{\mathbf{R}^n} f(x)\,\langle \nabla\psi(x), h\rangle\,d\mu = \int_{\mathbf{R}^n} D_h f(x)\,\psi(x)\,d\mu.$$

Since ψ was arbitrary, we deduce $D_h f = \langle h, G\rangle$ almost everywhere.

Now extend the basis $e_1, e_2, \ldots, e_n$ to a dense sequence of unit vectors $\{h_k\}$ in the unit sphere $S_{n-1} \subset \mathbf{R}^n$. Define the set $A \subset \mathbf{R}^n$ to consist of those points where each function $D_{h_k} f$ is defined and equals $\langle h_k, G\rangle$. Our argument above shows A^c has measure zero. We aim to show, at each point $x \in A$, that f has Fréchet derivative $G(x)$.

Fix any $\epsilon > 0$. For any $t \neq 0$, define a function $r_t : \mathbf{R}^n \to \mathbf{R}$ by

$$r_t(h) = \frac{f(x + th) - f(x)}{t} - \langle G(x), h\rangle.$$

It is easy to check that r_t has Lipschitz constant $2L$. Furthermore, for each $k = 1, 2, \ldots$, there exists $\delta_k > 0$ such that

$$|r_t(h_k)| < \frac{\epsilon}{2} \quad \text{whenever } 0 < |t| < \delta_k.$$

Since the sphere S_{n-1} is compact, there is an integer M such that

$$S_{n-1} \subset \bigcup_{k=1}^{M} \Big(h_k + \frac{\epsilon}{4L}B\Big).$$

If we define $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_M\} > 0$, we then have

$$|r_t(h_k)| < \frac{\epsilon}{2} \quad \text{whenever } 0 < |t| < \delta,\ k = 1, 2 \ldots, M.$$

Finally, consider any unit vector h. For some positive integer $k \leq M$ we know $\|h - h_k\| \leq \epsilon/4L$, so whenever $0 < |t| < \delta$ we have

$$|r_t(h)| \leq |r_t(h) - r_t(h_k)| + |r_t(h_k)| \leq 2L\frac{\epsilon}{4L} + \frac{\epsilon}{2} = \epsilon.$$

Hence $G(x)$ is the Fréchet derivative of f at x, as we claimed. □

1. Assuming Rademacher's theorem with range $\mathbf{R}$, prove the general version.

2. * **(Rademacher's theorem for regular functions)** Suppose the function $f : \mathbf{R}^n \to \mathbf{R}$ is locally Lipschitz around the point $x \in \mathbf{R}^n$. Suppose the vector $G(x)$ is well-defined by equation (9.1.3). By observing

$$0 = f^-(x; e_i) + f^-(x; -e_i) = f^\circ(x; e_i) + f^\circ(x; -e_i)$$

and using the sublinearity of $f^\circ(x; \cdot)$, deduce $G(x)$ is the Fréchet derivative of f at x.

3. ** **(Intrinsic Clarke subdifferential formula)** Derive formula (9.1.1) as follows.

(a) Using Rademacher's theorem (9.1.2), show we can assume that the function f is differentiable everywhere outside the set Q.

(b) Recall the one-sided inclusion following from the fact that the Clarke subdifferential is a closed multifunction (Exercise 12 in Section 6.2)

(c) For any vector $v \in \mathbf{E}$ and any point $z \in \mathbf{E}$, use Fubini's theorem to show that the set $\{t \in \mathbf{R} \mid z + tv \in Q\}$ has measure zero, and deduce

$$f(z+tv) - f(z) = \int_0^t \langle \nabla f(z+sv), v \rangle \, ds.$$

(d) If formula (9.1.1) fails, show there exists $v \in \mathbf{E}$ such that

$$f^\circ(x; v) > \limsup_{w \to x,\ w \notin Q} \langle \nabla f(w), v \rangle.$$

Use part (c) to deduce a contradiction.

4. ** **(Generalized Jacobian)** Consider a locally Lipschitz map between Euclidean spaces $h : \mathbf{E} \to \mathbf{Y}$ and a set $Q \subset \mathbf{E}$ of measure zero

outside of which h is everywhere Gâteaux differentiable. By analogy with formula (9.1.1) for the Clarke subdifferential, we call

$$\partial_Q h(x) = \operatorname{conv}\{\lim_r \nabla h(x^r) \mid x^r \to x,\ x^r \notin Q\},$$

the *Clarke generalized Jacobian* of h at the point $x \in \mathbf{E}$.

(a) Prove that the set $J_h(x) = \partial_Q h(x)$ is independent of the choice of Q.

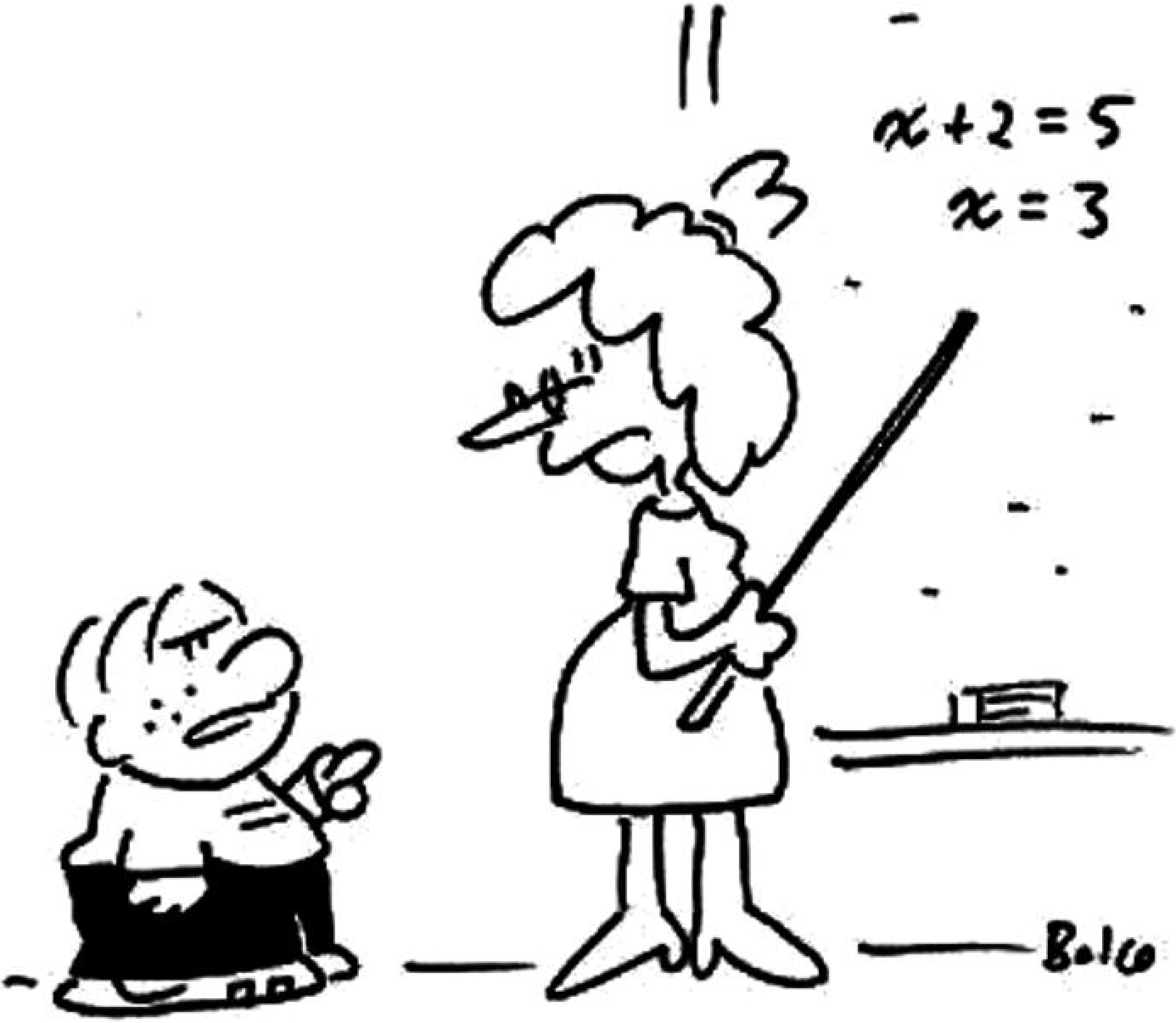
(b) **(Mean value theorem)** For any points $a, b \in \mathbf{E}$, prove

$$h(a) - h(b) \subset \operatorname{conv} J_h[a, b](a - b).$$

(c) **(Chain rule)** If the function $g : \mathbf{Y} \to \mathbf{R}$ is locally Lipschitz, prove the formula

$$\partial_\circ(g \circ h)(x) \subset J_h(x)^* \partial_\circ g(h(x)).$$

(d) Propose a definition for the generalized Hessian of a continuously differentiable function $f : \mathbf{E} \to \mathbf{R}$.



"Just a darn minute! — Yesterday you said that X equals two!"

9.2 Proximal Normals and Chebyshev Sets

We introduced the Clarke normal cone in Section 6.3 (Tangent Cones), via the Clarke subdifferential. An appealing alternative approach begins with a more geometric notion of a normal vector. We call a vector $y \in \mathbf{E}$ a *proximal normal* to a set $S \subset \mathbf{E}$ at a point $x \in S$ if, for some $t > 0$, the nearest point to $x + ty$ in S is x. The set of all such vectors is called the *proximal normal cone*, which we denote $N_S^p(x)$.

The proximal normal cone, which may not be convex, is contained in the Clarke normal cone (Exercise 3). The containment may be strict, but we can reconstruct the Clarke normal cone from proximal normals using the following result.

Theorem 9.2.1 (Proximal normal formula) *For any closed set $S \subset \mathbf{E}$ and any point $x \in S$, we have*

$$N_S(x) = \operatorname{conv}\Big\{ \lim_r y_r \mid y_r \in N_S^p(x_r),\ x_r \in S,\ x_r \to x \Big\}.$$

One route to this result uses Rademacher's theorem (Exercise 7). In this section we take a more direct approach.

The Clarke normal cone to a set $S \subset \mathbf{E}$ at a point $x \in S$ is

$$N_S(x) = \operatorname{cl}\,(\mathbf{R}_+ \partial_\circ d_S(x)),$$

by Theorem 6.3.8, where

$$d_S(x) = \inf_{z \in S} \|z - x\|$$

is the distance function. Notice the following elementary but important result that we use repeatedly in this section (Exercise 4(a) in Section 7.3).

The Clarke normal cone to a set $S \subset \mathbf{E}$ at a point $x \in S$ is

$$N_S(x) = \mathrm{cl}\,(\mathbf{R}_+ \partial_\circ d_S(x)),$$

by Theorem 6.3.8, where

$$d_S(x) = \inf_{z \in S} \|z - x\|$$

is the distance function. Notice the following elementary but important result that we use repeatedly in this section (Exercise 4(a) in Section 7.3).

Proposition 9.2.2 (Projections) *If $\bar{x}$ is a nearest point in the set $S \subset \mathbf{E}$ to the point $x \in \mathbf{E}$, then $\bar{x}$ is the unique nearest point in S to each point on the half-open line segment $[\bar{x}, x)$.*

To derive the proximal normal formula from the subdifferential formula (9.1.1), we can make use of some striking differentiability properties of distance functions, summarized in the next result.

Theorem 9.2.3 (Differentiability of distance functions) *Consider a nonempty closed set $S \subset \mathbf{E}$ and a point $x \notin S$. Then the following properties are equivalent:*

(i) *the Dini subdifferential $\partial_- d_S(x)$ is nonempty;*

(ii) *x has a unique nearest point $\bar{x}$ in S;*

(iii) *the distance function d_S is Fréchet differentiable at x.*

In this case,

$$\nabla d_S(x) = \frac{x - \bar{x}}{\|x - \bar{x}\|} \in N_S^p(\bar{x}) \subset N_S(\bar{x}).$$

The proof is outlined in Exercises 4 and 6.

For our alternate proof of the proximal normal formula without recourse to Rademacher's theorem, we return to an idea we introduced in Section 8.2. A cusco is a USC multifunction with nonempty compact convex images. In particular, the Clarke subdifferential of a locally Lipschitz function on an open set is a cusco (Exercise 5 in Section 8.2).

Suppose $U \subset \mathbf{E}$ is an open set, $\mathbf{Y}$ is a Euclidean space, and $\Phi : U \to \mathbf{Y}$ is a cusco. We call Φ *minimal* if its graph is minimal (with respect to set inclusion) among graphs of cuscos from U to Y. For example, the subdifferential of a continuous convex function is a minimal cusco (Exercise 8). We next use this fact to prove that Clarke subdifferentials of distance functions are also minimal cuscos.

Theorem 9.2.4 (Distance subdifferentials are minimal) *Outside a nonempty closed set $S \subset \mathbf{E}$, the distance function d_S can be expressed locally as the difference between a smooth convex function and a continuous convex function. Consequently, the Clarke subdifferential $\partial_\circ d_S : \mathbf{E} \to \mathbf{E}$ is a minimal cusco.*

Proof. Consider any closed ball T disjoint from S. For any point y in S, it is easy to check that the Fréchet derivative of the function $x \mapsto \|x - y\|$ is Lipschitz on T. Suppose the Lipschitz constant is $2L$. It follows that the function $x \mapsto L\|x\|^2 - \|x - y\|$ is convex on T (see Exercise 9). Since the function $h : T \to \mathbf{R}$ defined by

$$h(x) = L\|x\|^2 - d_S(x) = \sup_{y \in S}\{L\|x\|^2 - \|x - y\|\}$$

is convex, we obtain the desired expression $d_S = L\|\cdot\|^2 - h$.

To prove minimality, consider any cusco $\Phi : \mathbf{E} \to \mathbf{E}$ satisfying $\Phi(x) \subset \partial_\circ d_S(x)$ for all points x in $\mathbf{E}$. Notice that for any point $x \in \operatorname{int} T$ we have

$$\partial_\circ d_S(x) = -\partial_\circ(-d_S)(x) = \partial h(x) - Lx.$$

Since h is convex on $\operatorname{int} T$, the subdifferential ∂h is a minimal cusco on this set, and hence so is $\partial_\circ d_S$. Consequently, Φ must agree with $\partial_\circ d_S$ on $\operatorname{int} T$, and hence throughout S^c, since T was arbitrary.

On the set $\operatorname{int} S$, the function d_S is identically zero. Hence for all points x in $\operatorname{int} S$ we have $\partial_\circ d_S = \{0\}$ and therefore also $\Phi(x) = \{0\}$. We also deduce $0 \in \Phi(x)$ for all $x \in \operatorname{cl}(\operatorname{int} S)$.

Now consider a point $x \in \operatorname{bd} S$. The Mean value theorem (Exercise 9 in Section 6.1) shows

$$\begin{aligned}\partial_\circ d_S(x) &= \operatorname{conv}\Big\{0, \lim_r y^r \;\Big|\; y^r \in \partial_\circ d_S(x^r),\ x^r \to x,\ x^r \notin S\Big\} \\ &= \operatorname{conv}\Big\{0, \lim_r y^r \;\Big|\; y^r \in \Phi(x^r),\ x^r \to x,\ x^r \notin S\Big\},\end{aligned}$$

where 0 can be omitted from the convex hull unless $x \in \operatorname{cl}(\operatorname{int} S)$ (see Exercise 10). But the final set is contained in $\Phi(x)$, so the result now follows. □

The Proximal normal formula (Theorem 9.2.1), follows rather quickly from this result (and indeed can be strengthened), using the fact that Clarke subgradients of the distance function are proximal normals (Exercise 11).

We end this section with another elegant illustration of the geometry of nearest points. We call a set $S \subset \mathbf{E}$ a *Chebyshev set* if every point in E has a unique nearest point $P_S(x)$ in S. Any nonempty closed convex set is a Chebyshev set (Exercise 8 in Section 2.1). Much less obvious is the converse, stated in the following result.

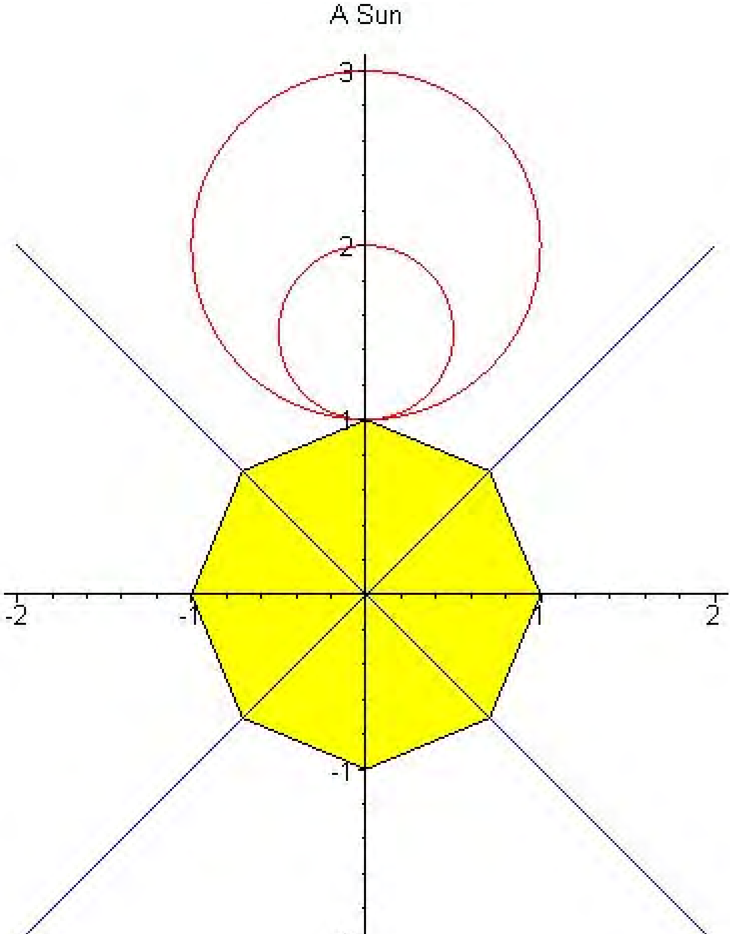
Theorem 9.2.5 (Convexity of Chebyshev sets) *A subset of a Euclidean space is a Chebyshev set if and only if it is nonempty, closed and convex.*

Proof. Consider a Chebyshev set $S \subset \mathbf{E}$. Clearly S is nonempty and closed, and it is easy to verify that the projection $P_S : \mathbf{E} \to \mathbf{E}$ is continuous. To prove S is convex, we first introduce another new notion. We call S a *sun* if, for each point $x \in \mathbf{E}$, every point on the ray $P_S(x) + \mathbf{R}_+(x - P_S(x))$ has nearest point $P_S(x)$. We begin by proving that the following properties are equivalent (see Exercise 13):

(i) S is convex;

(ii) S is a sun;

(iii) P_S is nonexpansive.


A Sun
3
2
1
-2
-1
1
2
-1

So, we need to show that S is a sun.

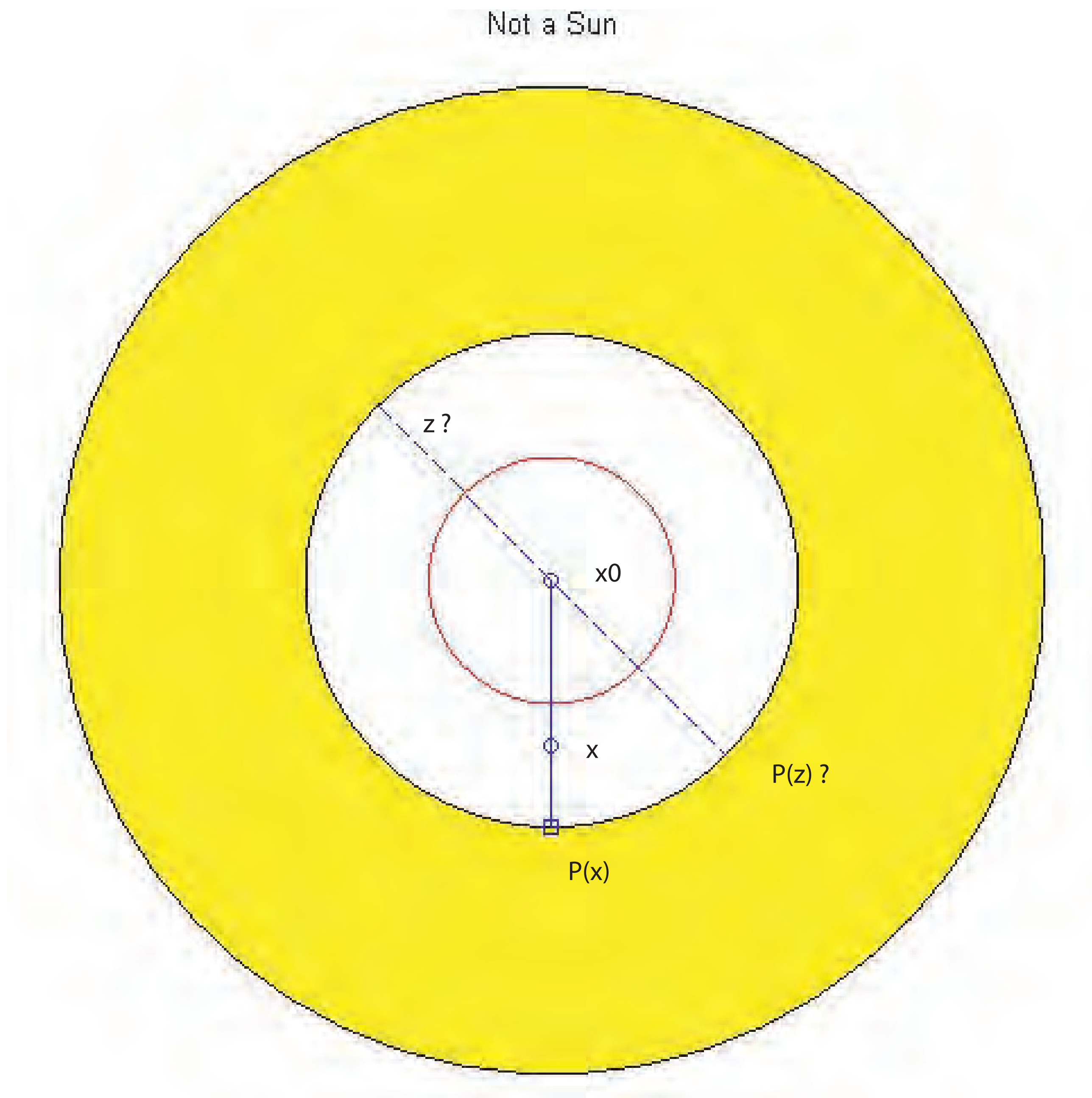
Suppose S is not a sun, so there is a point $x \notin S$ with nearest point $P_S(x) = \bar{x}$ such that the ray $L = \bar{x} + \mathbf{R}_+(x - \bar{x})$ strictly contains

$$\{z \in L \mid P_S(z) = \bar{x}\}.$$

Hence by Proposition 9.2.2 (Projections) and the continuity of P_S, the above set is nontrivial closed line segment $[\bar{x}, x_0]$ containing x.

Choose a radius $\epsilon > 0$ so that the ball $x_0 + \epsilon B$ is disjoint from S. The continuous self map of this ball

$$z \mapsto x_0 + \epsilon \frac{x_0 - P_S(z)}{\|x_0 - P_S(z)\|}$$


Not a Sun
z ?
x0
x
P(z) ?
P(x)

has a fixed point by Brouwer's theorem (8.1.3). We then quickly derive a contradiction to the definition of the point x_0. □

Exercises and Commentary

Proximal normals provide an alternative comprehensive approach to nonsmooth analysis: a good reference is [56]. Our use of the minimality of distance subdifferentials here is modelled on [38]. Theorem 9.2.5 (Convexity of Chebyshev sets) is sometimes called the "Motzkin-Bunt theorem". Our discussion closely follows [62]. In the exercises, we outline three nonsmooth proofs. The first (Exercises 14, 15, 16) is a variational proof following [82]. The second (Exercises 17, 18, 19) follows [96], and uses Fenchel conjugacy. The third argument (Exercises 20 and 21) is due to Asplund [2]. It is the most purely geometric, first deriving an interesting dual result on furthest points, and then proceeding via inversion in the unit sphere. Asplund extended the argument to Hilbert space, where it remains unknown whether a norm-closed Chebyshev set must be convex. Asplund showed that, in seeking a nonconvex Chebyshev set, we can restrict attention to "Klee caverns": complements of closed bounded convex sets.

2. **(Projections)** Prove Proposition 9.2.2.

3. **(Proximal normals are normals)** Consider a set $S \subset \mathbf{E}$. Suppose the unit vector $y \in \mathbf{E}$ is a proximal normal to S at the point $x \in S$.

   (a) Use Proposition 9.2.2 (Projections) to prove $d'_S(x; y) = 1$.

   (b) Use the Lipschitz property of the distance function to prove $\partial_\circ d_S(x) \subset B$.

   (c) Deduce $y \in \partial_\circ d_S(x)$.

   (d) Deduce that any proximal normal lies in the Clarke normal cone.

4. * **(Unique nearest points)** Consider a closed set $S \subset \mathbf{E}$ and a point x outside S with unique nearest point $\bar{x}$ in S. Complete the following steps to prove

$$\frac{x - \bar{x}}{\|x - \bar{x}\|} \in \partial_- d_S(x).$$

(a) Assuming the result fails, prove there exists a direction $h \in \mathbf{E}$ such that

$$d_S^-(x; h) < \langle \|x - \bar{x}\|^{-1}(x - \bar{x}), h \rangle.$$

(b) Consider a sequence $t_r \downarrow 0$ such that

$$\frac{d_S(x + t_r h) - d_S(x)}{t_r} \to d_S^-(x; h)$$

and suppose each point $x + t_r h$ has a nearest point s_r in S. Prove $s_r \to \bar{x}$.

(c) Use the fact that the gradient of the norm at the point $x - s_r$ is a subgradient to deduce a contradiction.

5. **(Nearest points and Clarke subgradients)** Consider a closed set $S \subset \mathbf{E}$ and a point x outside S with a nearest point $\bar{x}$ in S. Use Exercise 4 to prove

$$\frac{x - \bar{x}}{\|x - \bar{x}\|} \in \partial_\circ d_S(x).$$

6. * **(Differentiability of distance functions)** Consider a nonempty closed set $S \subset \mathbf{E}$.

(a) For any points $x, z \in \mathbf{E}$, observe the identity

$$d_S^2(z) - d_S^2(x) = 2d_S(x)(d_S(z) - d_S(x)) + (d_S(z) - d_S(x))^2.$$

(b) Use the Lipschitz property of the distance function to deduce

$$2d_S(x)\partial_- d_S(x) \subset \partial_- d_S^2(x).$$

Now suppose $y \in \partial_- d_S(x)$.

(c) If $\bar{x}$ is any nearest point to x in S, use part (b) to prove $\bar{x} = x - d_S(x)y$, so $\bar{x}$ is in fact the unique nearest point.

(d) Prove $-2d_S(x)y \in \partial_-(-d_S^2)(x)$.

(e) Deduce d_S^2 is Fréchet differentiable at x.

Assume $x \notin S$.

(f) Deduce d_S is Fréchet differentiable at x.

(g) Use Exercises 3 and 4 to complete the proof of Theorem 9.2.3.

7. * **(Proximal normal formula via Rademacher)** Prove Theorem 9.2.1 using the subdifferential formula (9.1.1) and Theorem 9.2.3 (Differentiability of distance functions).

8. **(Minimality of convex subdifferentials)** If the open set $U \subset \mathbf{E}$ is convex and the function $f : U \to \mathbf{R}$ is convex, use the Max formula (Theorem 3.1.8) to prove that the subdifferential ∂f is a minimal cusco.

9. **(Smoothness and DC functions)** Suppose the set $C \subset \mathbf{E}$ is open and convex, and the Fréchet derivative of the function $g : C \to \mathbf{R}$ has Lipschitz constant $2L$ on C. Deduce that the function $L\|\cdot\|^2 - g$ is convex on C.

10. ** **(Subdifferentials at minimizers)** Consider a locally Lipschitz function $f : \mathbf{E} \to \mathbf{R}_+$, and a point x in $f^{-1}(0)$. Prove

$$\partial_\circ f(x) = \operatorname{conv}\Big\{0, \lim_r y^r \;\Big|\; y^r \in \partial_\circ f(x^r),\ x^r \to x,\ f(x^r) > 0\Big\},$$

where 0 can be omitted from the convex hull if $\operatorname{int} f^{-1}(0) = \emptyset$.

11. ** (**Proximal normals and the Clarke subdifferential**) Consider a closed set $S \subset \mathbf{E}$ and a point x in S Use Exercises 3 and 5 and the minimality of the subdifferential $\partial_\circ d_S : \mathbf{E} \to \mathbf{E}$ to prove

$$\partial_\circ d_S(x) = \operatorname{conv}\Big\{0, \lim_r y^r \,\Big|\, y^r \in N_S^p(x^r),\ \|y^r\| = 1,\ x^r \to x,\ x^r \in S\Big\}.$$

Deduce the Proximal normal formula (Theorem 9.2.1). Assuming $x \in \operatorname{bd} S$, prove the following stronger version. Consider any dense subset Q of S^c, and suppose $P : Q \to S$ maps each point in Q to a nearest point in S. Prove

$$\partial_\circ d_S(x) = \operatorname{conv}\Big\{0, \lim_r \frac{x^r - P(x^r)}{\|x^r - P(x^r)\|} \,\Big|\, x^r \to x,\ x^r \in Q\Big\},$$

and derive a stronger version of the Proximal normal formula.

12. **(Continuity of the projection)** Consider a Chebyshev set S. Prove directly from the definition that the projection P_S is continuous.

13. * **(Suns)** Complete the details in the proof of Theorem 9.2.5 (Convexity of Chebyshev sets) as follows.

    (a) Prove (iii) $\Rightarrow$ (i).

    (b) Prove (i) $\Rightarrow$ (ii).

    (c) Denoting the line segment between points $y, z \in \mathbf{E}$ by $[y, z]$, prove property (ii) implies

$$P_S(x) = P_{[z, P_S(x)]}(x) \quad \text{for all } x \in \mathbf{E},\ z \in S. \tag{9.2.6}$$

(d) Prove (9.2.6) $\Rightarrow$ (iii).

(e) Fill in the remaining details of the proof.

14. ** (**Basic Ekeland variational principle** [**43**]) Prove the following version of the Ekeland variation principle (Theorem 7.1.2). Suppose the function $f : \mathbf{E} \to (\infty, +\infty]$ is closed and the point $x \in \mathbf{E}$ satisfies $f(x) < \inf f + \epsilon$ for some real $\epsilon > 0$. Then for any real $\lambda > 0$ there is a point $v \in \mathbf{E}$ satisfying the conditions

(a) $\|x - v\| \leq \lambda$,

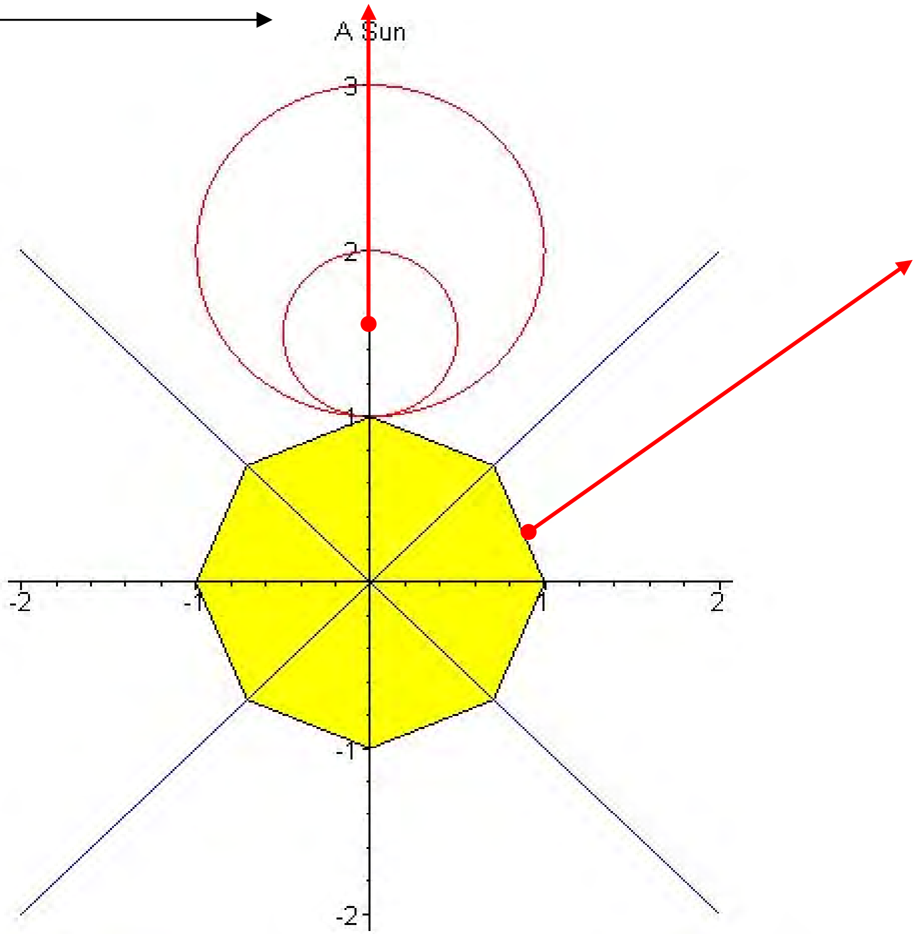
(b) $f(v) + (\epsilon/\lambda)\|x - v\| \leq f(x)$, and

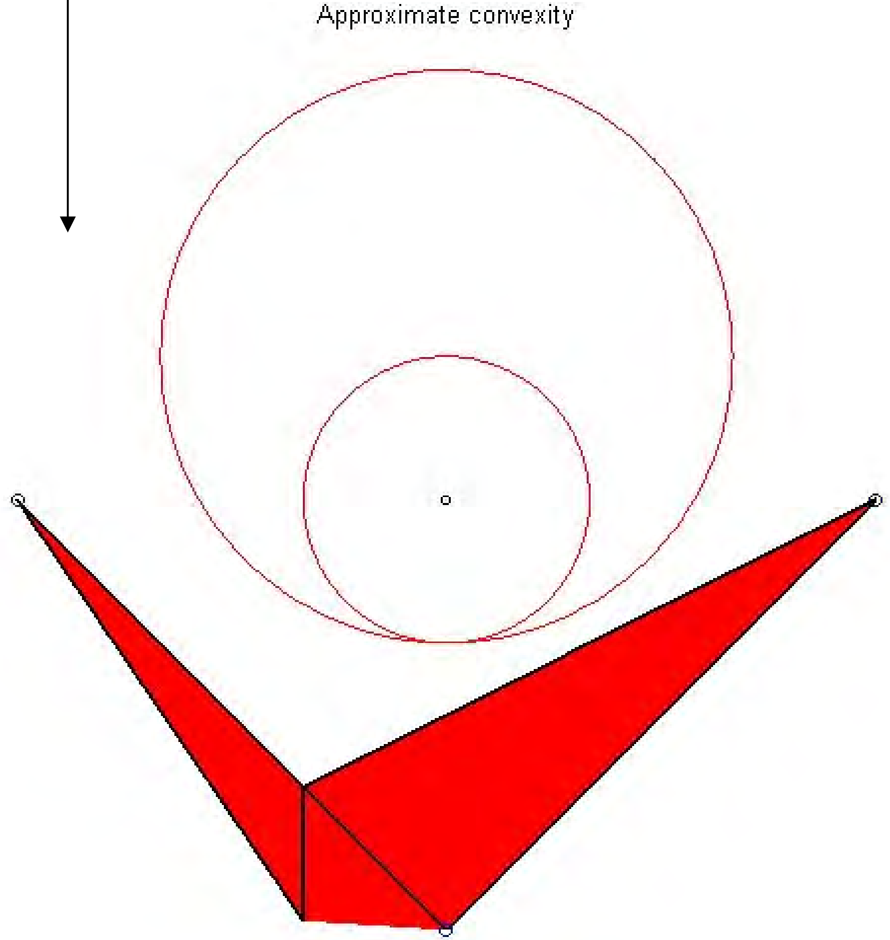
(c) v minimizes the function $f(\cdot) + (\epsilon/\lambda)\|\cdot - v\|$.

15. * (**Approximately convex sets**) Consider a closed set $C \subset \mathbf{E}$. We call C *approximately convex* if, for any closed ball $D \subset \mathbf{E}$ disjoint from C, there exists a closed ball $D' \supset D$ disjoint from C with arbitrarily large radius.

Ekeland replaces Brouwer

A Chebyshev Sun is Approximately Convex





- replaces exact Hilbert proximal analysis by ε-variations
- works for weakly closed sets in smooth rotund space

(a) If C is convex, prove it is approximately convex.

(b) Suppose C is approximately convex but not convex.

(i) Prove there exist points $a, b \in C$ and a closed ball D centered at the point $c = (a+b)/2$ and disjoint from C.

(ii) Prove there exists a sequence of points $x_1, x_2, \ldots \in \mathbf{E}$ such that the balls $B_r = x_r + rB$ are disjoint from C and satisfy $D \subset B_r \subset B_{r+1}$ for all $r = 1, 2, \ldots$.

(iii) Prove the set $H = \mathrm{cl} \cup_r B_r$ is closed and convex, and its interior is disjoint from C but contains c.

(iv) Suppose the unit vector u lies in the polar set H°. By considering the quantity $\langle u, \|x_r - x\|^{-1}(x_r - x)\rangle$ as $r \to \infty$, prove H° must be a ray.

(v) Deduce a contradiction.

(c) Conclude that a closed set is convex if and only if it is approximately convex.

16. ** (**Chebyshev sets and approximate convexity**) Consider a Chebyshev set $C \subset \mathbf{E}$, and a ball $x + \beta B$ disjoint from C.

(a) Use Theorem 9.2.3 (Differentiability of distance functions) to prove

$$\limsup_{v \to x} \frac{d_C(v) - d_C(x)}{\|v - x\|} = 1.$$

(b) Consider any real $\alpha > d_C(x)$. Fix reals $\sigma \in (0,1)$ and ρ satisfying

$$\frac{\alpha - d_C(x)}{\sigma} < \rho < \alpha - \beta.$$

By applying the Basic Ekeland variational principle (Exercise 14) to the function $-d_C + \delta_{x+\rho B}$, prove there exists a point $v \in \mathbf{E}$ satisfying the conditions

Approx convex implies convex iff norm is rotund

$$\begin{aligned} d_C(x) + \sigma\|x - v\| &\leq d_C(v) \\ d_C(z) - \sigma\|z - v\| &\leq d_C(v) \quad \text{for all } z \in x + \rho B. \end{aligned}$$

Use part (a) to deduce $\|x - v\| = \rho$, and hence $x + \beta B \subset v + \alpha B$.

(c) Conclude that C is approximately convex, and hence convex by Exercise 15.

(d) Extend this argument to an arbitrary norm on $\mathbf{E}$.

17. ** **(Smoothness and biconjugacy)** Consider a function $f : \mathbf{E} \to (\infty, +\infty]$ that is closed and bounded below and satisfies the condition

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = +\infty.$$

Consider also a point $x \in \operatorname{dom} f$.

(a) Using Carathéodory's theorem (Section 2.2, Exercise 5), prove there exist points $x_1, x_2, \ldots, x_m \in \mathbf{E}$ and real $\lambda_1, \lambda_2, \ldots, \lambda_m > 0$ satisfying

$$\sum_i \lambda_i = 1, \quad \sum_i \lambda_i x_i = x, \quad \sum_i \lambda_i f(x_i) = f^{**}(x).$$

(b) Use the Fenchel-Young inequality (Proposition 3.3.4) to prove

$$\partial(f^{**})(x) = \bigcap_i \partial f(x_i).$$

Suppose furthermore that the conjugate f^* is everywhere differentiable.

(c) If $x \in \operatorname{ri}(\operatorname{dom}(f^{**}))$, prove $x_i = x$ for each i.

(d) Deduce $\operatorname{ri}(\operatorname{epi}(f^{**})) \subset \operatorname{epi}(f)$.

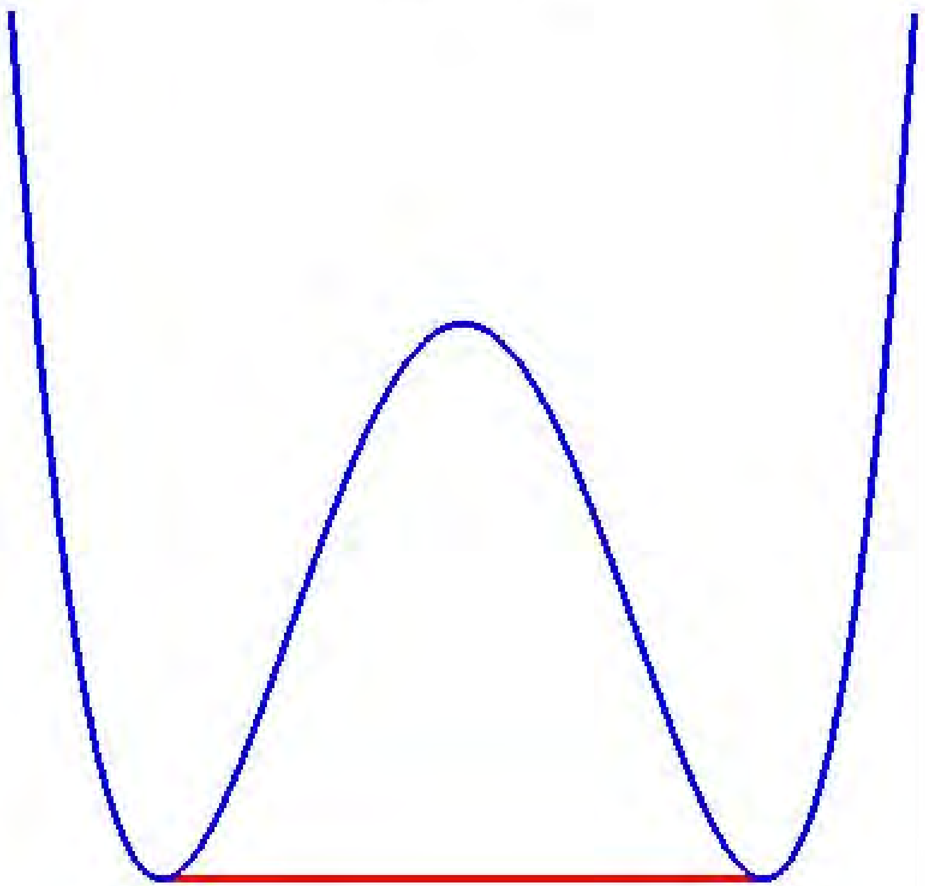
(e) Use the fact that f is closed to deduce $f = f^{**}$, so f is convex.

18. * **(Chebyshev sets and differentiability)** Use Theorem 9.2.3 (Differentiability of distance functions) to prove that a closed set $S \subset \mathbf{E}$ is a Chebyshev set if and only if the function d_S^2 is Fréchet differentiable throughout $\mathbf{E}$.

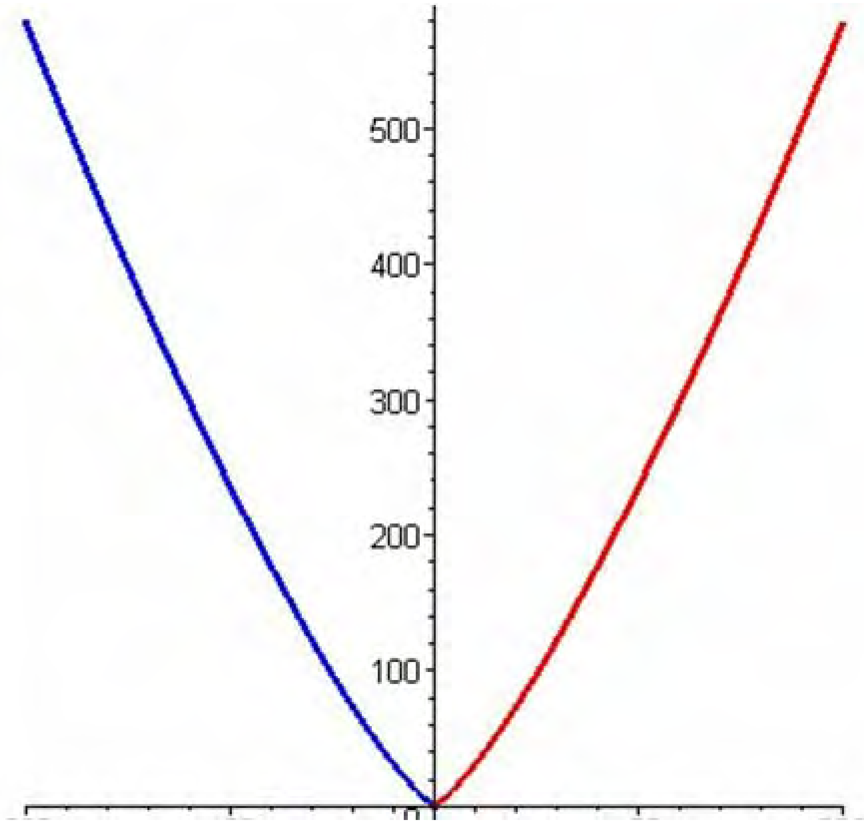
W is convex if and only if W* is Frechet

$$W(x) = \left(1 - x^2\right)^2$$

W and W**



W*



19. * (**Chebyshev convexity via conjugacy**) For any nonempty closed set $S \subset \mathbf{E}$, prove

$$\left(\frac{\|\cdot\|^2 + \delta_S}{2}\right)^* = \frac{\|\cdot\|^2 - d_S^2}{2}$$

Deduce, using Exercises 17 and 18, that Chebyshev sets are convex.

20. ** (**Unique furthest points**) Consider a set $S \subset \mathbf{E}$, and define a function $r_S : \mathbf{E} \to [-\infty, +\infty]$ by

$$r_S(x) = \sup_{y \in S} \|x - y\|.$$

Any point y attaining the above supremum is called a *furthest point* in S to the point $x \in \mathbf{E}$.

(a) Prove that the function $(r_S^2 - \|\cdot\|^2)/2$ is the conjugate of the function

$$g_S = \frac{\delta_{-S} - \|\cdot\|^2}{2}.$$

(b) Prove that the function r_S^2 is strictly convex on its domain.

Now suppose each point $x \in \mathbf{E}$ has a unique nearest point $q_S(x)$ in S.

(c) Prove that the function q_S is continuous.

We consider two alternative proofs that a set has the unique furthest point property if and only if it is a singleton.

(d) (i) Use Section 6.1 , Exercise 10 (Max-functions) to show that the function $r_S^2/2$ has Clarke subdifferential the singleton $\{x - q_S(x)\}$ at any point $x \in \mathbf{E}$, and hence is everywhere differentiable.

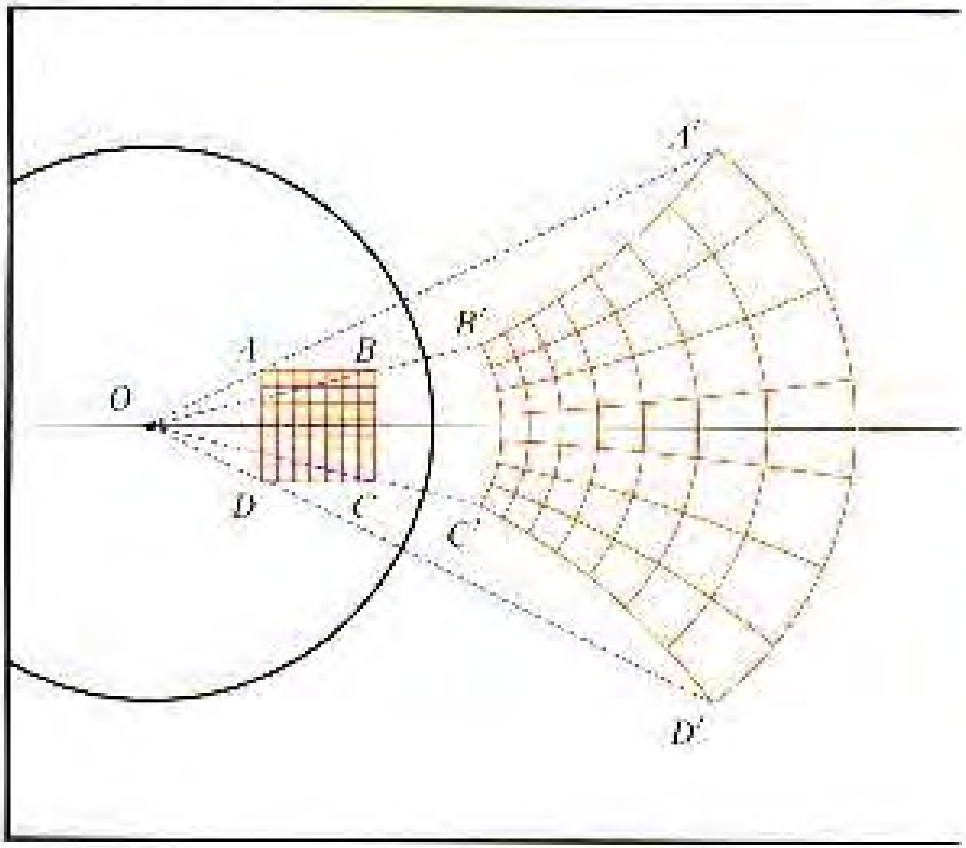
(ii) Use Exercise 17 (Smoothness and biconjugacy) to deduce that the function g_S is convex, and hence that S is a singleton.

(e) Alternatively, suppose S is not a singleton. Denote the unique minimizer of the function r_S by y. By investigating the continuity of the function q_S on the line segment $[y, q_S(y)]$, derive a contradiction without using part (d).

21. ** **(Chebyshev convexity via inversion)** The map $\iota : \mathbf{E} \setminus \{0\} \to \mathbf{E}$ defined by $\iota(x) = \|x\|^{-2}x$ is called the *inversion in the unit sphere*.

(a) If $D \subset \mathbf{E}$ is a ball with $0 \in \operatorname{bd} D$, prove $\iota(D \setminus \{0\})$ is a halfspace disjoint from 0.

Inverse Geometry for Hunters



Preserves circles (spheres, lines …)

(b) For any point $x \in \mathbf{E}$ and radius $\delta > \|x\|$, prove

$$\iota((x + \delta B) \setminus \{0\}) = \frac{1}{\delta^2 - \|x\|^2}\{y \in \mathbf{E} : \|y + x\| \geq \delta\}.$$

Prove that any Chebyshev set $C \subset \mathbf{E}$ must be convex as follows. Without loss of generality, suppose $0 \notin C$ but $0 \in \mathrm{cl}\,(\mathrm{conv}\, C)$. Consider any point $x \in \mathbf{E}$.

(c) Prove the quantity

$$\rho = \inf\{\delta > 0 \mid \iota C \subset x + \delta B\}$$

satisfies $\rho > \|x\|$.

(d) Let z denote the unique nearest point in C to the point

$$\frac{-x}{\rho^2 - \|x\|^2}.$$

Use part (b) to prove that ιz is the unique furthest point in ιC to x.

(e) Use Exercise 20 to derive a contradiction.

The Chebyshev Problem in Infinite Dimensions is OPEN

In any Banach space (JMB & JV, CUP in press):

Corollary 3.14.2. *Suppose* $f : X \to (-\infty, \infty]$ *is such that* f^{**} *is proper.*

(a) If f^* *is Fréchet differentiable at all* $x^* \in \mathrm{dom}(\partial f^*)$ *and* f *is lower semicontinuous, then* f *is convex.*

(a) If f^* *is Gâteaux differentiable at all* $x^* \in \mathrm{dom}(\partial f^*)$ *and* f *is sequentially weakly lower semicontinuous, then* f *is convex.*

Theorem 3.14.7. *Let* X *be a Hilbert space and suppose* C *is a nonempty weakly closed subset of* X. *Then the following are equivalent.*

(i) C *is convex.*

(ii) C *is a Chebyshev set.*

(iii) $d(\cdot, C)^2$ *is Fréchet differentiable.*

(iv) $d(\cdot, C)^2$ *is Gâteaux differentiable.*

"...and, as you go out into the world, I predict that you will, gradually and imperceptibly, forget all you ever learned at this university."