Variational Analysis and Tame Optimization

Aris Daniilidis
http://mat.uab.cat/~arisd

Universitat Autònoma de Barcelona

April 15–17, 2010

PLAN OF THE TALK

Nonsmooth analysis

Genericity of
pathological situations

Interaction with geometry

Good structure in practice

A theoretical result:
Nonsmooth Sard Theorem

Application: (Thom First Integral Conjecture)
Genericity of non-existence of first integrals

Example (against common intuition)

There exists a continuous, nowhere differentiable function

(Weierstrass, 1872) For $a \in (0,1)$, $b > 0$ (odd integer) and $ab > 1 + \frac{3\pi}{2}$ set

$$f(x) = \sum_{n=1}^{+\infty} a^n \cos(b^n \pi t).$$

Then f is continuous but nowhere differentiable.

Fact (Pathology is ″generic”)

Generically all continuous functions defined on a compact set are nowhere differentiable (c.f. Baire)

Things cannot be better in variational analysis...

One of the basic objects of nonsmooth analysis:

Definition (Clarke subdifferential)

For any Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ the Clarke subdifferential at every $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) = \text{co} \left\{ \lim_{x_n \to x} Df(x_n) \,:\, x_n \in \text{dom}\,(Df) \diagdown \mathcal{N} \right\}.$$

Recall:

Fact

There exists a measurable set $A \subset \mathbb{R}$ such that for any nontrivial interval I we have $0 < m(A \cap I) < m(I)$.

Example (against (?) common intuition)

For the Lipschitz function

$$f(x) = \int_0^x [\chi_A(t) - \chi_{A^c}(t)]\, dt$$

we have

$$\partial f(x) = [-1, 1] \quad \text{for all} \;\; x \in \mathbb{R}.$$

But pathology is again generic !

Theorem (X. Wang, 1998)

Generically, all k-Lipschitz functions (on a compact subset S of $\mathbb{R}^n$) satisfy

$$\partial f(x) = k\mathbb{B} \quad \textit{for all } x \in S.$$

Classical Analysis

Smooth functions

Variational analysis

Well-structured functions

"GIVE UP FULL GENERALITY IN FAVOR OF A STRUCTURED THEORY"

(Grothendieck's *Sketch of a Programme*)

convex functions	(Fenchel, Moreau)
ϕ-convex functions	(Marino-Tosques, '80)
semiconvex	(Clarke, Ambrosio, '90)
amenable, lower-C^2	(Federer, Rockafellar, Penot)
prox-regular	(Thibault)
"primer-lower-nice"	(Poliquin, '90)
partly smooth	(Lewis, Hare, 2002)
semismooth	(Mifflin, Qi, Sun)

...

OBJECTIVE: AVOID PATHOLOGIES

Example (minimization of the sum of norms)

Consider the (location) problem

$$\min\ f(x) = \sum_{i \in \{1,\ldots,m\}} \|A_i\, x - b_i\|$$

- (active indices) $I(\bar{x}) = \{i : A_i\, \bar{x} = b_i\}$
- (active manifold) $\mathcal{M}_{\bar{x}} = \{x : I(x) = I(\bar{x})\}$

Fact

$\mathcal{M}_{\bar{x}}$ is an affine manifold and $f\ |_{\mathcal{M}_{\bar{x}}}$ is smooth.

Remark. The linear operators A_i's can be replaced by any smooth functions F_i.

Example (SDP Optimization)

Let $g = \Lambda_{\max} \circ F$ where $F : \mathbb{R}^n \to S^m$ is smooth and $\Lambda_{\max} : S^m \to \mathbb{R}$ is the maximum eigenvalue function, defined for each symmeric matric A as follows:

$$\Lambda_{\max}(A) = \max_{\|x\|=1} \langle Ax, x \rangle \qquad \text{(convex, continuous)}$$

Then

$$\Lambda_{\max}(A) = \lambda_1(A) \geq \ldots \geq \lambda_m(A)$$

Set for $1 \leq r \leq m$

$$\mathcal{M}_r = \{A \in S^m : \lambda_1(A) = \ldots = \lambda_r(A) > \lambda_{r+1}(A) \geq \lambda_m(A)\}$$

($\mathcal{M}_r$ is a Nash manifold)

- $\mathcal{M}_1$ open dense – $\Lambda_{\max}$ differentiable on $\mathcal{M}_1$

$$\Lambda_{\max}\,|_{\mathcal{M}_r}\,(A) = \frac{\lambda_1(A) + \ldots + \lambda_r(A)}{r}$$

Fact

The restriction $\Lambda_{\max}\,|_{\mathcal{M}_r}$ of $\Lambda_{\max}$ onto $\mathcal{M}_r$ is smooth

- $\mathcal{M}_{r+1} \subset \mathrm{cl}(\mathcal{M}_r) \diagdown \mathcal{M}_r$ for all $r \in \{1, ..., m-1\}$
- $\bigcup\limits_{y} \mathcal{M}_r = S^m$

The manifolds $\mathcal{M}_r$ form a stratification of the space S^m.

Definition (Stratification)

$\mathcal{R} = \{\mathcal{S}_i\}$ is called a stratification of $X \subset \mathbb{R}^n$ if
(i) $\mathcal{R}$ is (locally) finite and $X = \bigcup_i \mathcal{S}_i$;
(ii) $\mathcal{S}_i \cap \mathrm{cl}(\mathcal{S}_j) \neq \emptyset \Longrightarrow \mathcal{S}_i \subset \mathrm{cl}(\mathcal{S}_j) \setminus \mathcal{S}_j$;
(iii) $\mathcal{S}_i$ C^k-manifold (for all i).

Definition

A function f (in $\mathbb{R}^n$) is called stratifiable, if Graph(f) admits a stratification (in $\mathbb{R}^{n+1}$).

Fact

Any set definable in some o-minimal structure admits a stratification.

Illustrative example

$f(x, y) = x^2 + |y|$

Stratification of the « graph »

$$\begin{cases} S_1 = \{ (x, y, f(x, y)) : y > 0 \} \\ S_2 = \{ (x, y, f(x, y)) : y < 0 \} \\ S_3 = \{ (x, y, f(x, y)) : y = 0 \} \end{cases}$$

Projection onto $\mathbb{R}^2$

$$f|_{X_i} \in C^\infty(X_i) \iff \begin{cases} X_1 = \{ (x, y) : y > 0 \} \\ X_2 = \{ (x, y) : y < 0 \} \\ X_3 = \{ (x, y) : y = 0 \} \end{cases}$$

"Frechet subdifferential and Riemann Gradient"

Definition

We say that p is a Fréchet subgradient of f at x if

$$\liminf_{\|u\|\to 0} \frac{f(x+u) - f(x) - \langle p, u\rangle}{\|u\|} \geq 0.$$

The set $\hat{\partial} f(x)$ of all subgadients is called Fréchet subdifferential.

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ continuous (nonsmooth), $x \in \operatorname{dom} f$

- **(Assumption)** Graph (f) admits a stratification $\{\mathcal{S}_i\}$ in $\mathbb{R}^{n+1}$
- Projecting $\mathcal{S}_j$ onto $\mathbb{R}^n$ we obtain a stratification $\{\mathcal{X}_i\}_j$ of $\operatorname{dom} f$
- Assume $x \in X_i$. Then (Basic projection formula´´)

$$\operatorname{Proj}_{T_{\mathcal{X}_i}(x)} \left[\hat{\partial} f(x)\right] \subset \{\nabla_R f(x)\}$$

Theorem (Projection formula)

Graph (f) *admits a* C^p*-Whitney stratification* $\{\mathcal{S}_i\}_{i\in I}$. *Then*★

$$\mathrm{Proj}_{T_{\mathcal{X}_i}(x)}\left[\partial f(x)\right] \subset \{\nabla_R f(x)\} \qquad \text{for all } x \in \mathrm{dom}\, f$$

($\mathcal{X}_i$ *projection of* S_i *onto* $\mathbb{R}^n$ *where* $(x, f(x)) \in S_i$*.)*

★ J. Bolte, A. Daniilidis, A. S. Lewis, M. Shiota, Clarke subgradients of stratifiable functions, *SIAM J. Optim.* **18** (2007), 556–572.

Corollary (Nonsmooth Sard theorem)

If $p \geq n$, *then the set of Clarke critical values of* f *has Lebesgue measure 0.*

(*Proof.* Apply classical Sard theorem to each stratum.)

Examples of Whitney stratifiable functions

f semialgebraic ⟺ Finite stratification (semialgebraic strata)

f subanalytic ⟺ (Locally) finite stratification (subanalytic strata)

f o-minimal ⟺ Finite stratification (o-minimal strata)

Corollary

(Locally) finite nonsmooth critical values.

Definition

– A subset of $\mathbb{R}^n$ is *semialgebraic* if it is a finite union of sets, each defined by finitely-many polynomial inequalities.

– A function (or set-valued mapping) is *semialgebraic* if its graph is semi-algebraic.

Properties:

- Stability (Quantifier elimination / Tarski-Seidenberg)
- Stratification: every semi-algebraic set can be written as a finite disjoint union of manifolds ("strata") that fit together in a regular way.
- Semialgebraic subsets of $\mathbb{R}$ are exactly the finite unions of points and intervals (o-minimal property).

How to recognize semialgebricity ?

- Basic semialgebraic set:

$$\{x \in \mathbb{R}^n : p(x) = 0, q_j(x) > 0, j = 1, \ldots, k\}$$

- Semialgebraic set: every set that is obtained via finite Boolean operations of basic semialgebraic sets.★

Quantifier elimination: In the language of ordered fields, we can eliminate quantifiers.

Fact

First ordered formulae in the language of ordered fields $(R, +, \cdot, >, 0, 1)$ *define semialgebraic sets.*

★ *that is*, sets defined by 0-order formulas.

Illustration. The set

$$\{(a,b,c) : (\forall x)(ax^2 + bx + c > 0)\}$$

can be equivalently rewritten as

$$\{(a,b,c) : a = 0, b = 0, c > 0\} \quad \bigcup \quad \{(a,b,c) : a > 0, b^2 - 4ac < 0\},$$

thus it is semialgebraic.

Examples

(i) If $f : \mathbb{R}^n \to \mathbb{R}$ is semialgebraic, then so is its derivative (resp. subdifferential), its level sets, the set of critical points, the set of critical values ...
(ii) If F is semialgebraic, then so is the normal cone mapping $N_F(\cdot)$, its indicator function $i_F(\cdot)$, its support function $\sigma_F(\cdot)$ ...

$$f(x) = \begin{cases} (1-\|x\|)^{1/2}, & \text{si } \|x\| \leq 1 \\ -\infty, & \text{si } \|x\| > 1 \end{cases}$$

$$h(t) = \begin{cases} |t|^p (1+|t|^q)^{-1}, & \text{si } t \neq 0 \\ 0, & \text{si } t = 0 \end{cases}$$

(p, q rationals)

Theorem

Every semialgebraic Lipschitz function is semismooth★.

★ J. BOLTE, A. DANIILIDIS, A. S. LEWIS, Tame functions are semismooth, *Math. Program.* **116** (2009), 115–127.

Example :

$$f(x,y) = \begin{cases} (y^2 - x^2)^{1/2}, & \text{si } y > 2|x| \\ 3^{1/2}|x|, & \text{si } y \leq 2|x| \end{cases}$$

f is semismooth (!)

(Globally) subanalytic function

$$g(x,y) = \begin{cases} y\,(\sin x)^{-1/2} & \text{if } \ 0 < x < \pi \\ 0\,, & \text{if } \ y = 0 \\ (\operatorname{sign} y)\,\infty\,, & \text{else} \end{cases}$$

O-minimal function (log-exp structure)

$$f(x,y,z) = \begin{cases} |x|^{p}\, 2^{\exp(x) - y^{-2}} \log(|\cos z|), & \text{if } \ x \neq 0,\, y \neq 0,\, |z| \leq N \\ 0\,, & \text{if } \ x = 0 \ \text{ or } \ y = 0 \\ +\infty\,, & \text{else} \end{cases}$$

$(N > 0,\ p \in \mathbb{R})$

Application (dynamical systems)

First Integral Conjecture (R. Thom)

"Generically", vector fields on compact connected smooth manifolds $\mathcal{M}$ cannot admit nontrivial continuous first integrals.

Thom's proposed scheme of a proof

Closing lemma ($\Longrightarrow$ genericity of structurally stable vector fields).
Sard theorem ($\Longrightarrow$ triviality of first integral).

(**Pugh, 1967**)

– Validity of C^1-closing lemma

– General density theorem (GDT)

Formal proofs (under slight variations in the assumptions)

Peixoto (1967), Bewley (1971), Mañe (1973), Hurley (1986)

$\phi_t \colon \mathcal{M} \to \mathcal{M}$: group of diffeomorphisms generated by $\mathcal{X}$.

Definition

A point $p \in \mathcal{M}$ is called *nonwandering*, if given any neighborhood U of p, there are arbitrarily large values of t for which $U \cap \phi_t(U) \neq \emptyset$. [Ω = set of nonwandering points]

General density theorem (GDT): The set G_M of vector fields $\mathcal{X} \in \mathcal{X}_M$ satisfying **(G$_1$)–(G$_4$)** is residual in $\mathcal{X}_M$.

(G$_1$) $\mathcal{X}$ has only a finite number of singularities, all generic ;

(G$_2$) closed orbits of $\mathcal{X}$ are generic ;

(G$_3$) stable and unstable manifolds of singularities and closed orbits of $\mathcal{X}$ are transversal ;

(G$_4$) $\Omega = \bar{\Gamma}$ ($\Gamma \equiv$ union of singular points and closed orbits)

Common technical assumption

Common technical assumption in all proofs:
The first integral is assumed C^k, $k \geq \dim \mathcal{M}$ *(c.f. Sard theorem).*

- Relaxation of regularity: Lipschitz continuity

- Notion of criticality

$$0 \in \partial f(x) \qquad \text{(Clarke criticality)}$$

Tools

– Generic singularities and closed orbits of hyperbolic type are located at the Clarke-critical level sets (for each Lipschitz first integral).
– Nonsmooth Sard theorem

$\mathcal{M},\mathcal{N}$ C^1 manifolds

Definition (Approximation of a manifold)

(i) $(\mathcal{N},F)$ is a C^1-approximation of $\mathcal{M}$ (of precision ϵ), if $\mathcal{N} \subset U_\epsilon$ (ϵ-tubular neighborhood of $\mathcal{M}$) and F can be extended to a C^1-diffeomorphism $\tilde{F}\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$, which is isotopic to the identity id, satisfies $\tilde{F}|_{\mathbb{R}^n \setminus U_r} \equiv \mathrm{id}$ and

$$\max_{x\in\mathbb{R}^n} \left\{ ||\tilde{F}(x) - x|| + ||d\tilde{F}(x) - \mathrm{id}|| \right\} < \epsilon.$$

(ii) A C^1-approximation $(\mathcal{N},F)$ of $\mathcal{M}$ is called semialgebraic if $\mathcal{N}$ is a semialgebraic subset of $\mathbb{R}^n$

In the sequel, we shall need the following approximation result.

Lemma (Semialgebraic approximation)

Let $\mathcal{M}$ be a C^1 compact submanifold of $\mathbb{R}^n$. Then for every $\varepsilon > 0$, there exists a semialgebraic ε-approximation of $\mathcal{M}$.

Lemma

$\mathcal{M}$ C^1 compact manifold. If $\mathcal{X}$ is a C^1 vector field of GDT type, then it does not admit Lipschitz continuous definable first integral.

Theorem

For the C^1 topology, the set of vector fields $\mathcal{X}$ of $\mathcal{M}$ that do not admit essentially o-minimal★ Lipschitz continuous first integrals (with respect to a given definable approximation of $\mathcal{M}$) is generic★★.

★ $f : \mathcal{M} \to \mathbb{R}$ is called essentially o-minimal with respect to a definable approximation $(\mathcal{N}, F)$ of $\mathcal{M}$ if the mapping $f \circ F^{-1} : \mathcal{N} \to \mathbb{R}$ is o-minimal.

★★ J. Cresson, A. Daniilidis, M. Shiota, On the first integral conjecture of René Thom, *Bull. Sci. Math.* **132** (2008), 625–631.

Application (Set-valued analysis)

Theorem

Every semialgebraic multifunction is generically (outer and inner) continuous and differentiable (in a set-valued sense)★.

★ A. Daniilidis, J. Pang, Continuity of set-valued maps revisited in the light of tame geometry, preprint 2009 (in revision in *Proc. London Math. Soc.*)

announcement
(conference in Barcelona)

Advances in Optimization and Related Topics
November 29 to December 3, 2010.

Conference Web: http://www.crm.cat/RPOPT2010
Personal Page: http://mat.uab.cat/~arisd/