Nonsmooth critical values and Sard type results

ARIS DANIILIDIS

variational Analysis, optimization and quantitative finance

limoges (may 18-22, 2015)

The (classical) Sard theorem

CRITICAL SET (SMOOTH CASE)

- $F : \mathbb{R}^n \to \mathbb{R}^m$ differentiable function

Definition (Critical point)

A point x_0 is called *critical*, if $dF(x_0)$ is not surjective.

- S := set of critical points
- $F(S)$:= set of critical values

Theorem (Sard theorem)

For every C^k-smooth function $F : \mathbb{R}^n \to \mathbb{R}^m$ we have $m(F(S)) = 0$, provided $k \geq n - m + 1$.

APPLICATIONS (DIFFERENTIAL GEOMETRY)

- Immersed submanifolds $\mathcal{N} \subset \mathcal{M}$ of positive co-dimension have zero measure in $\mathcal{M}$.
- (*Whitney embedding*)
  C^∞ mnfds of dimension d embed into $\mathbb{R}^{2d+1}$.

APPLICATIONS (OPTIMIZATION)

- Constraint qualification condition (*genericity result*)

$$\begin{cases} \min \quad f(x) \\ h_i(x) = r_i \\ i \in \{1, \ldots, m\} \end{cases} \quad \longleftrightarrow \quad \begin{array}{c} H : \mathbb{R}^n \to \mathbb{R}^m \\ H = (h_1, \ldots, h_m) \end{array} \qquad (\mathcal{P}_r)$$

The set of data $r = (r_1, \ldots, r_m) \in \mathbb{R}^m$ for which the constraints do not satisfy the *qualification condition* at the solution x_* is of measure zero.

APPLICATIONS (OPTIMIZATION)

- Constraint qualification condition (*genericity result*)

$$\begin{cases} \min \quad f(x) \\ h_i(x) = r_i \\ i \in \{1, \ldots, m\} \end{cases} \longleftrightarrow \begin{array}{c} H : \mathbb{R}^n \to \mathbb{R}^m \\ H = (h_1, \ldots, h_m) \end{array} \qquad (\mathcal{P}_r)$$

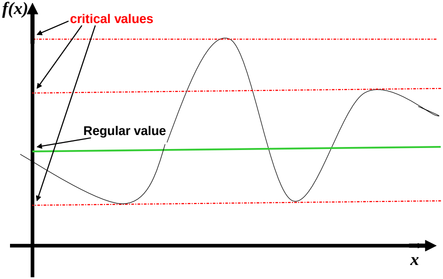
The set of data $r = (r_1, \ldots, r_m) \in \mathbb{R}^m$ for which the constraints do not satisfy the *qualification condition* at the solution x_* is of measure zero.

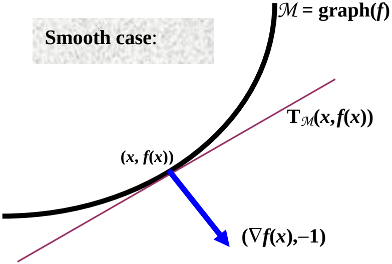
$$x_* \text{ solution of } (\mathcal{P}_r) \quad \Longrightarrow \quad \nabla f(x_*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x_*)$$

REAL-VALUED CASE

Theorem (Morse theorem)

For every C^k function $f : \mathbb{R}^n \to \mathbb{R}$ we have $m(f(S)) = 0$, provided $k \geq n$.

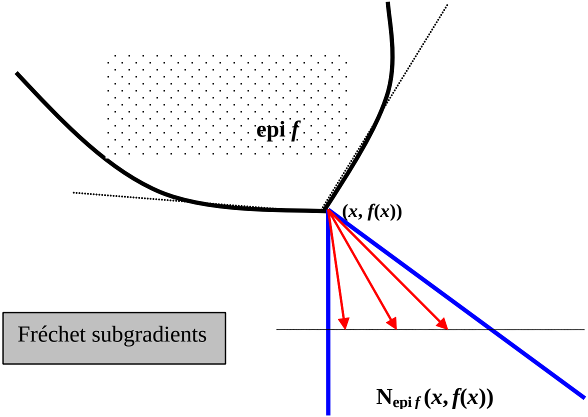




tangent space (to the graph) $T_{\text{graph}(f)}(u)$ → tangent cone (to the epigraph) $T_{\text{epi}(f)}(u)$

normal space (to the graph) $N_{\text{graph}(f)}(u)$ → normal cone (to the epigraph) $N_{\text{epi}(f)}(u)$

gradient $\nabla f(x)$ of f at x → subgradient $x^* \in \partial f(x)$ of f at x

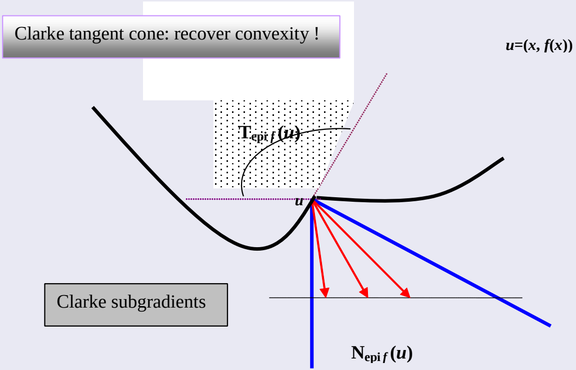


Unilateral definition:

$\partial f(x)$ may be empty (if f presents an "inward" corner).

Definition (Clarke subgradient - Lipschitz case)

$$\partial^C f(x) = \{p \in \mathbb{R}^n : (p, -1) \in N_{\mathrm{epi}\, f}(x, f(x))\}.$$



D_f := differentiability points (dense by Rademacher thm)

Clarke subdifferential

$$\partial f(x_0) = \text{conv}\left\{ \lim_{x_n \to x_0} \nabla f(x_n) : \{x_n\}_n \subset D_f \right\}$$

CRITICAL SET (NONSMOOTH CASE)

- $f : \mathbb{R}^n \to \mathbb{R}$ (nonsmooth) Lipschitz

Definition (Critical point)

A point x_0 is called *critical*, if $0 \in \partial f(x_0)$.

- If f is C^1 then (it is locally Lipschitz and)

$$\partial f(x) = \{\nabla f(x)\}, \quad \text{for all } x \in \mathbb{R}^n$$

- No hope to treat the general nonsmooth case if $n > 1$!

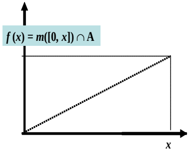
Theorem (Recall – smooth case)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^k-smooth function. Then $m(f(S)) = 0$ provided $k \geq n$ (or $k = n - 1$ and $D^{n-1}f$ is Lipschitz).

STRONG FAILURE OF SARD

$A \subset \mathbb{R}_+$ measurable: $\forall$ (nontrivial) interval I

$$0 < m(A \cap I) < m(I)$$



$$f(x) = m(A \cap [0,x]) = \int_0^x \chi_A(t)dt$$

$\partial f(x) = [0,1], \forall x \implies S = [0,1]$ and f is $\nearrow$ (!)

GENERIC PATHOLOGY

Theorem (X. Wang, 1998)

All points of a generic Lipschitz function are critical.

SEEKING A GOOD STRUCTURE

Recompense *nonsmoothness* assuming structure:

- Convex paradigm (convexity, generalizations)
- Semialgebraic paradigm (semialgebricity, tameness)

The convex paradigm does not lead far...

Fact (trivial)

Convex functions satisfy Morse-Sard theorem.

Fact (Failure for *DC* functions)

Every C^2 function can be written as a difference of a C^2 convex and a convex quadratic function (on compact sets)

Corollary (Failure in 3-dim)

The Morse-Sard theorem fails for semiconvex (and thus DC) functions in $\mathbb{R}^3$.

Semialgebraic paradigm: strong conclusions

- Semialgebricity (tameness) $\Longrightarrow$ C^p-Whitney stratifiability

Semialgebraic paradigm: strong conclusions

- Semialgebricity (tameness) $\Longrightarrow$ C^p-Whitney stratifiability
- The set of critical values is semialgebraic (Quantifier elimination principle)

Corollary (Tame case)

(Locally) finite nonsmooth critical values.

Semialgebraic paradigm: strong conclusions

- Semialgebricity (tameness) $\Longrightarrow$ C^p-Whitney stratifiability
- The set of critical values is semialgebraic
  Quantifier elimination principle ($\bigstar$)

Corollary (Tame case)

(Locally) finite nonsmooth critical values.

($\bigstar$) The set of critical points is semialgebraic
($\rightarrow$ finite union of arc-connected parts)

Further extensions:

nonsmooth (and nontame!) case

- f is a *maximum* of smooth functions
  (scalar case)
- f is a (Lipschitz) *selection* of smooth functions
  (vector case)

Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ C^k-smooth, $k \geq n$.

$$f(x) = \max\{f_1(x), f_2(x)\},$$

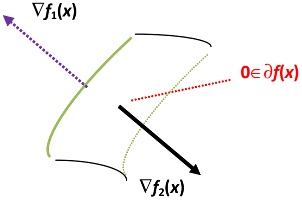
Clarke subdifferential:

$$\partial f(x) = \begin{cases} \nabla f_1(x) & \text{if } f_1(x) > f_2(x) \\ \nabla f_2(x) & \text{if } f_1(x) < f_2(x) \\ \mathrm{co}\{\nabla f_1(x), \nabla f_2(x)\} & \text{if } f_1(x) = f_2(x) \end{cases}$$

Clarke critical points:

$$S = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}$$

A SIMPLE OBSERVATION



$$\Phi = f_1 - f_2 \Longrightarrow \nabla\Phi(x) \neq 0.$$

Set

$$S_i := \{x \in \mathbb{R}^n : \nabla f_i(x) = 0\}$$

and

$$\mathcal{M} = \{x \in \mathbb{R}^n : f(x) = f_1(x) = f_2(x)\}$$

Fact

If $x \in \mathcal{M}$ is Clarke critical and $x \notin S_1 \cup S_2$ (i.e. $\nabla f_i(x) \neq 0 \ \forall i$) then

– $\mathcal{M}$ is a C^k smooth submanifold of $\mathbb{R}^n$ around x

– $f|_{\mathcal{M}}$ is C^k-smooth on $\mathcal{M}$ and $\nabla_{\mathcal{R}}(f|_{\mathcal{M}})(x) = 0$.

Deduce a nonsmooth result from the classical Morse-Sard theorem

Apply classical Morse-Sard theorem to f_1, f_2 and $f|_{\mathcal{M}}$

$$\Longrightarrow f_1(S_1), f_2(S_2) \text{ and } f|_{\mathcal{M}}(S) \text{ null.}$$

- Conclude:

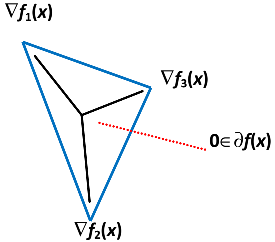
$$f(S) \subset f_1(S_1) \cup f_2(S_2) \cup f|_{\mathcal{M}}(S)$$

is null.

Fact

This argument extends to finite continuous selections

$$f(x) \in \{f_1(x), \ldots f_k(x)\}.$$



$$\mathcal{M} = \{x : f_1(x) = f_2(x) = f_3(x)\}$$

Assume

$$\bar{x} \in \mathcal{M}, \qquad 0 \in \mathrm{co}\{\nabla f_1(\bar{x}), \nabla f_2(\bar{x}), \nabla f_3(\bar{x})\}$$

and

$$0 \notin \mathrm{co}\{\nabla f_1(\bar{x}), \nabla f_2(\bar{x})\} \bigcup \mathrm{co}\{\nabla f_2(\bar{x}), \nabla f_3(\bar{x})\} \bigcup \mathrm{co}\{\nabla f_1(\bar{x}), \nabla f_3(\bar{x})\}$$

Then:

– $\mathcal{M}$ is a C^k submnfd of $\mathbb{R}^n$ around $\bar{x}$

– $f|_{\mathcal{M}}$ (is C^k-smooth and) $\nabla_{\mathcal{R}}(f|_{\mathcal{M}})(x) = 0$.

GENERAL NONSMOOTH MORSE-SARD RESULT

Consider $f : \mathbb{R}^n \to \mathbb{R}$ *continuous* selection:

$$f(x) \in \{F(x,t) : t \in T\}, \quad \text{for all } x \in \mathbb{R}^n$$

- T compact countable
- $x \mapsto F(x,t)$ C^k-smooth ($k \geq n$)
- $F, \nabla_x F$ jointly continuous.

Theorem (Barbet, Dambrine, Daniilidis (Adv. Math. 2013))

f is locally Lipschitz and satisfies the Morse-Sard theorem.

Corollary (lower-C^k functions over a *countable* set)

Assume :

– T countable compact,

– $F(\cdot,t)$ C^k-smooth ($k \geq n$) and

– F, $\nabla_x F$ (jointly) continuous.

Then the locally Lipschitz function

$$f(x) = \max_{t \in T} F(x,t), \quad x \in \mathbb{R}^n$$

satisfies the Morse–Sard theorem.

TOOLS OF THE PROOF

- Cantor-Bendinxon index
- Formula for the Clarke subgradients:

$$\partial f(x) \subset \text{conv}\,\{\nabla_x F(x,t) : t \in T(x)\}$$

  where $T(x)$ = active indices of f at x.
- Finite representation of critical points (*Caratheodory*)

Fact

The proof uses the Morse–Sard theorem, but it also recovers it.

VECTORIAL CASE (SARD THEOREM) ?

$f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ *continuous* selection:

$$f(x) \in \{F(x,t) : t \in T\}, \quad \text{for all } x \in \mathbb{R}^n$$

where $F : \mathbb{R}^n \times T \longrightarrow \mathbb{R}^p$

- T compact countable
- $x \mapsto F(x,t)$ C^k-smooth ($k \geq n$)
- $F, \nabla_x F$ jointly continuous.

VECTORIAL CASE (SARD THEOREM) ?

$f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ *continuous* selection:

$$f(x) \in \{F(x,t) : t \in T\}, \quad \text{for all } x \in \mathbb{R}^n$$

where $F : \mathbb{R}^n \times T \longrightarrow \mathbb{R}^p$

Question (How to proceed ?)

The previous approach (i.e. Riemann gradients to naturally arising submanifols) requires f to be scalar valued.

PREPARATORY SARD THEOREM

$$\begin{cases} \Psi : \operatorname{inn}\Delta^m \times \mathcal{M} \to \mathbb{R}^p \\ \Psi(\boldsymbol{\lambda}, x) := \sum_{i=0}^m \lambda_i \phi^i(x). \end{cases}$$

where

$$\dim \mathcal{M} = n, \qquad \phi^i : \mathcal{M} \to \mathbb{R}^p$$

$$\begin{cases} \Psi(\boldsymbol{\lambda}, x) := \sum_{i=0}^m \lambda_i \phi^i(x), \\ (\boldsymbol{\lambda}, x) \in \operatorname{inn}\Delta^m \times \mathcal{M}. \end{cases}$$

Define $\widehat{\mathrm{Crit}}\Psi$ *strongly critical points*:

$$(\lambda, x) \in \widehat{\mathrm{Crit}}\Psi \qquad \Longleftrightarrow \qquad \begin{cases} \phi^i(x) = \phi^0(x), & i \in \{0, \ldots, m\} \\ \mathrm{rank}\ \left(\sum_{i=0}^m \lambda_i D\phi^i(x)\right) < p. \end{cases}$$

Theorem ("Preparatory Sard theorem")

$\mathcal{M}$ is C^k mnfd, $\dim \mathcal{M} = n$
$\phi^i : \mathcal{M} \to \mathbb{R}^p$, $i \in \{0, \ldots, m\}$ of class C^k and

$$\begin{cases} \Psi : \mathrm{inn}\,\Delta^m \times \mathcal{M} \to \mathbb{R}^p \\ \Psi(\boldsymbol{\lambda}, x) := \sum_{i=0}^m \lambda_i \phi^i(x). \end{cases}$$

If $k \geq n - p + 1$, then $\Psi(\widehat{\mathrm{Crit}}\Psi)$ is null in $\mathbb{R}^p$.

Proof of Preparatory Sard theorem.

– Wnlog $\mathcal{M} = \mathbb{R}^n$ and $\operatorname{dom} f = \operatorname{dom} \phi^i = [0,1]^n$
– Divide $[0,1]^n$ in ℓ subcubes.

– Approximate ϕ^i by a polynomial P_j^i
(in the subcube $j \in \{1, \ldots, \ell\}$)

– Apply the *Yomdin technique* to estimate the ε-critical values. □

APPLICATION (NONSMOOTH SARD THEOREM - VECTOR CASE)

$$\begin{cases} f(x) \in \{F(x,t) : t \in T\} \\ F : \mathbb{R}^n \times T \longrightarrow \mathbb{R}^p \end{cases}$$

Assume $T = \{1, \ldots, \ell\}$.
Fix $m_i \in \{0, 1, \ldots, n\}$, $i \in \{1, \ldots, p\}$ and set:

$$\begin{cases} G : \operatorname{inn}\Delta^{m_1} \times \ldots \times \operatorname{inn}\Delta^{m_p} \times \mathbb{R}^n \longrightarrow \mathbb{R}^p \\ G(\lambda^1, \ldots, \lambda^p, x) = \left(\sum_{j_1=0}^{m_1} \lambda^1_{j_1} F_1(x, t^1_{j_1}), \ldots, \sum_{j_p=0}^{m_\ell} \lambda^p_{j_p} F_p(x, t^p_{j_\ell}) \right) \end{cases}$$

Fact (transferring criticality from f to some G)

$\bar{x} \in \operatorname{Crit} f \implies \exists \{m_1, \ldots, m_p\}$ *and* $\lambda^i \in \operatorname{inn}\Delta^{m_i}$ *(depending on* $\bar{x}$*):*

– $(\lambda^1, \ldots, \lambda p, \bar{x}) \in \widehat{\operatorname{Crit}}\, G$;

– $f(\bar{x}) = G(\lambda^1, \ldots, \lambda^p, \bar{x})$.

- Transferring criticality from f to G

$$(\bar{x} \in \mathrm{Crit}\, f \Longrightarrow) \quad (\lambda^1, \ldots, \lambda^p, x) \in \widehat{\mathrm{Crit}}\, G \iff$$

$$\iff \begin{cases} F_i(x, t^i_{j_i}) = F_i(x, t^i_0), & \begin{cases} \text{for all } i \in \{1, \ldots, p\} \\ \text{for all } j_i \in \{0, \ldots, m_i\} \end{cases} \\ \text{rank } \left(\sum_{j_1=0}^{m_1} \lambda^1_{j_1} D_x F_1(x, t^1_{j_1}), \ldots, \sum_{j_p=0}^{m_\ell} \lambda^p_{j_p} D_x F_p(x, t^p_{j\ell}) \right) < p. \end{cases}$$

- Set:

$$m = \left(\prod_{i=1}^{p} (m_i + 1) \right) - 1 \qquad \text{and} \qquad \begin{cases} \vec{i} = (i_1, i_2, \ldots, i_\ell) \\ i_j \in \{0, \ldots, m_j\} \end{cases}$$

Describe $\operatorname{inn} \Delta^m$ by multi-indices, i.e. $\boldsymbol{\lambda} = (\lambda_{\vec{i}}) \in \operatorname{inn} \Delta^m$.

Set:

$$\begin{cases} \phi^{\vec{i}} : \mathbb{R}^n \to \mathbb{R}^p, \quad \vec{i} = (i_1, i_2, \ldots, i_p) \\ \phi^{\vec{i}}(x) = \left(F_1(x, t^1_{i_1}), \ldots, F_p(x, t^p_{i_p}) \right) \end{cases}$$

and

$$\Psi(\boldsymbol{\lambda}, x) := \sum_{i_1=0}^{m_1} \cdots \sum_{i_p=0}^{m_p} \lambda_{\vec{i}} \phi^{\vec{i}}(x), \quad (\boldsymbol{\lambda}, x) \in \operatorname{inn} \Delta^m \times \mathbb{R}^n.$$

- Transferring strong criticality from G to Ψ

$(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x) \in \widehat{\mathrm{Crit}}\, G \Longrightarrow$ for $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{\vec{i}}) \in \operatorname{inn} \Delta^m$ with

$$\boldsymbol{\lambda}_{(i_1,\ldots,i_p)} = \lambda^1_{i_1} \cdots \lambda^p_{i_p}$$

it holds

$$(\boldsymbol{\lambda}, x) \in \widehat{\mathrm{Crit}}\Psi \qquad \text{and} \qquad \Psi(\boldsymbol{\lambda}, x) = G(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x).$$

- By the Sard Preparatory theorem:

$$\Psi(\widehat{\mathrm{Crit}}\Psi) \text{ is null in } \mathbb{R}^p.$$

APPLICATION: SEMI-INFINITE PROGRAMMING

$$(\mathcal{P}_r) \qquad \min_{g_t(x) \le r} \quad u(x)$$

- Necessary optimality conditions for $(\mathcal{P}_r)$.

For *a.a.* $r \in \mathbb{R}$, and solution $\bar{x}$ of $(\mathcal{P}_r)$ there exist $\lambda_1, \ldots, \lambda_n \geq 0$ and $\{t_1, \ldots, t_n\} \subset T(\bar{x})$ such that

$$0 \in \partial u(\bar{x}) + \sum_{i=1}^{n} \lambda_i \nabla g_{t_i}(\bar{x})$$

REFERENCES

L. BARBET, A. DANIILIDIS, M. DAMBRINE,
The Morse-Sard Theorem for Clarke critical values,
Adv. Math. **242** (2013), 217–227.

L. BARBET, A. DANIILIDIS, M. DAMBRINE, L. RIFFORD
Sard theorems for Lipschitz functions
and applications in Optimization
Israel. Math. J. (to appear)