~~Introduction to o-minimality for~~ tame optim.

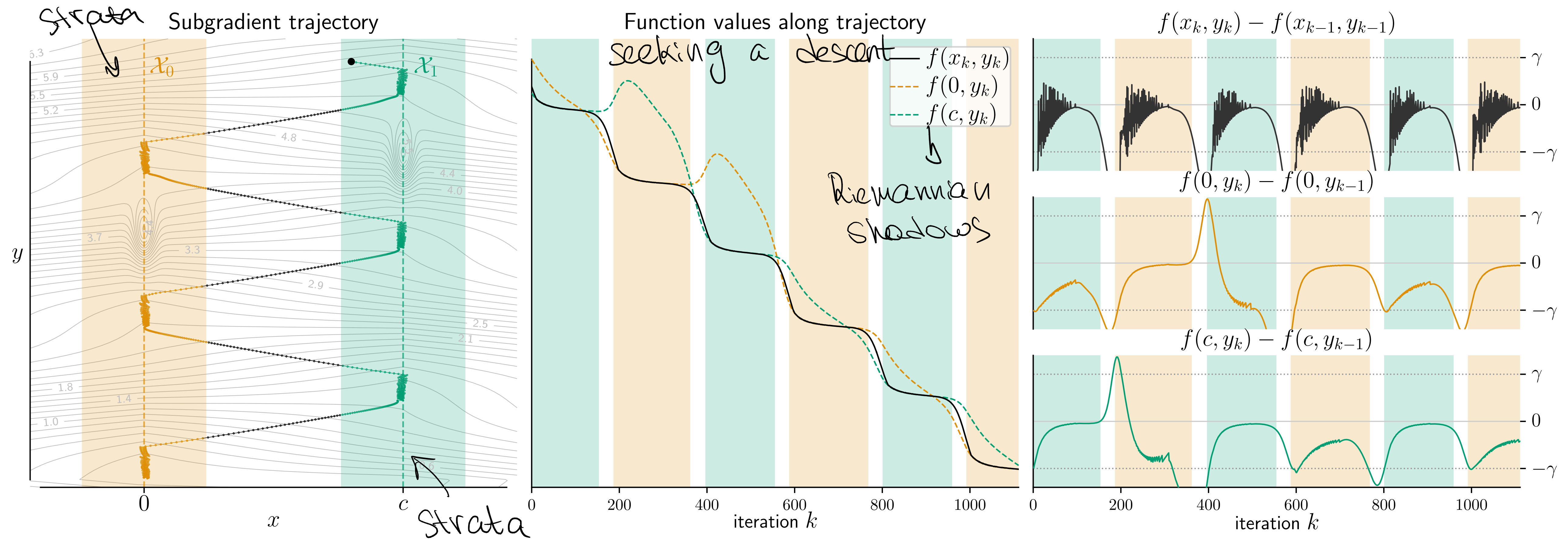
1) Good optim theory in convex case
   (smooth or not)

2) Good optim theory in smooth case

3) Only asymptotic theory in "non-convex"
   "non-smooth" case

non-convex + non-smooth is essentially everything, but in most applications there is structure (albeit unusual)

$\boxed{\text{o-minimality is a good candidate for "non-convex + non-smooth"}}$



Strata
Subgradient trajectory
x_0
x_1
y
x
0
c
Strata
Function values along trajectory
seeking a descent
$f(x_k, y_k)$
$f(0, y_k)$
$f(c, y_k)$
Riemannian shadows
iteration k
$f(x_k, y_k) - f(x_{k-1}, y_{k-1})$
$f(0, y_k) - f(0, y_{k-1})$
$f(c, y_k) - f(c, y_{k-1})$
γ
$-\gamma$
iteration k

Resources

1) An introduction to o-minimal geometry, M. Coste, '99

2) Clarke subgradients of stratifiable functions
Bolte, Daniilidis, Lewis, Shiota, 2007

3) For our results: C. & Schechtman
to be out on Monday

Stratification

$p \geq 1$

$A \subseteq \mathbb{R}^d$, a $\underbrace{C^p}_{\text{finite}}$ stratification of A is a finite collection $(X_i)_{i \in I}$ s.t.

1) $\bigcup_{i \in I} X_i = A$

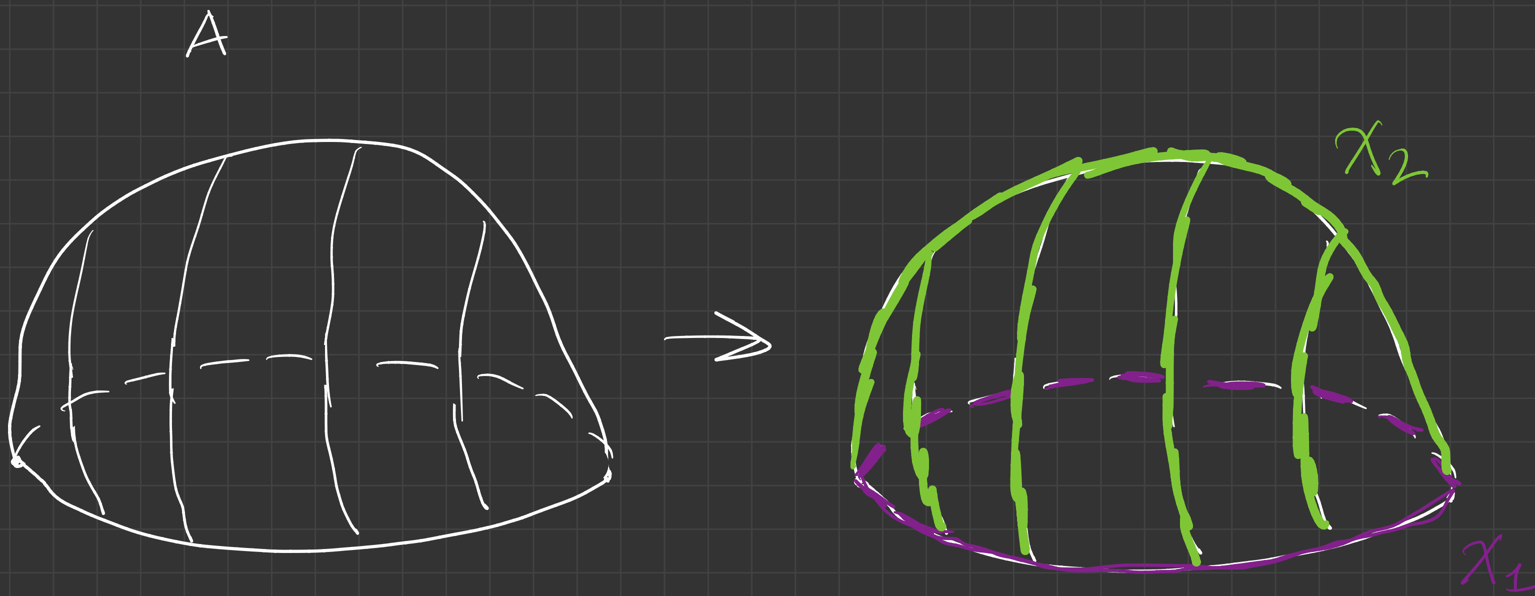
2) $X_i \cap X_j = \emptyset$

3) X_i is a C^p submanifold of $\mathbb{R}^d$

4) Boundary conditions : $\overline{X_i} \cap X_j \neq \emptyset \Rightarrow X_j \subseteq \partial X_i$

closure

topological frontier



lower dimensional manifolds live in frontiers of higher dimensional manifolds

One typically adds extra regularity conditions to glue nicely the boundaries

I will talk only about Whitney (a) condition

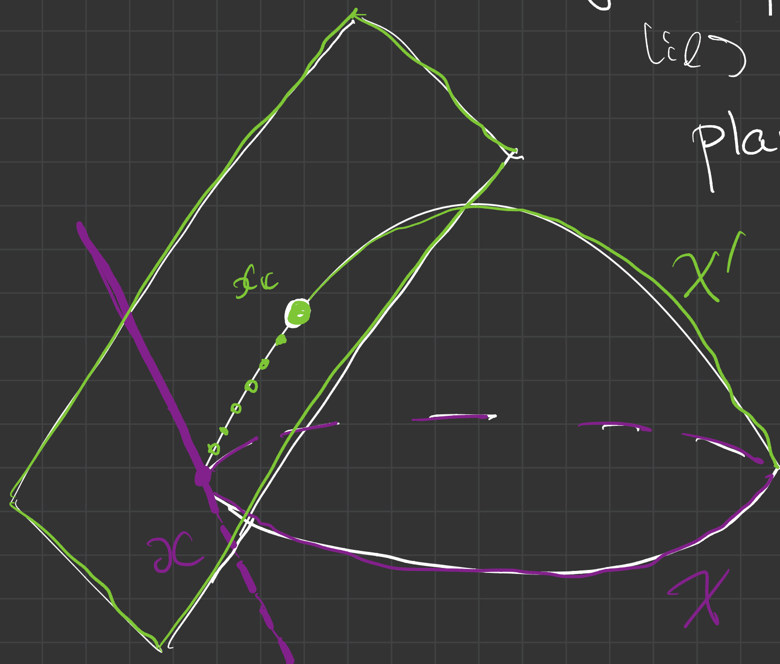
<u>Def</u> $A \subseteq \mathbb{R}^d$, $\mathbb{X} = (X_i)_{i \in I}$ a C^p-stratification

$\mathbb{X}$ satisfies Whitney (a) if

$\forall \quad X \leq \partial X' \quad \forall \quad x \in X \quad \forall \quad (x_k)_{k \geq 1} \in X'$

$$\boxed{\begin{cases} x_k \to x \\ T_{x_k} X' \to \tau \end{cases} \implies \quad T_x X \subseteq \tau}$$

In words: take a sequence x_c that goes to a boundary point. Take corresponding sequence of tangent planes, it goes somewhere. We want that tangent plane of the boundary (i.e.) is in this limiting tangent plane



$$\boxed{X \subseteq \partial X'}$$

Why would we care?

(some details are hidden)

<u>Theorem</u> (Bolte, Danilidis, Lewis, Shiota 2007)

Let $f : \mathbb{R}^d \to \mathbb{R}$ be locally lips.

$$\text{Graph } f = \{ (x, f(x)) : x \in \mathbb{R}^d \}$$

Assume that $(S_i)_{i \in I}$ is a finite C^p-stratification of Graph f, satisfying Whitney (a) cond an such that

projection onto the first d coord.

1) $X_i \overset{\text{def}}{=} \Pi(S_i)$ is C^1 manifold

2) $f_{|X_i}$ is C^1 (As a function from C^1 manifold)

Then $\forall x \in \mathbb{R}^d \ \forall g \in \partial f(x)$

Riemannian

$$\boxed{\operatorname{Proj}_{T_x \mathcal{X}}(g) = \nabla_{\mathcal{X}} f(x)}$$

$\boxed{\text{In words}}$: Clarke subgradients are Riemannian gradients + something in normal plane at $x \in \mathcal{X}$!

Idea of the proof!

One needs to figure out tangent planes of $\text{Graph}_{|\mathcal{X}} f = \{(x, f(x)) : x \in \mathcal{X}\}$ at an arbitrary $x \in \mathcal{X}$

$$(x,y) \in \mathrm{Graph}_{\mathcal{X}} f \iff \boxed{x \in \mathcal{X}_i \qquad y = f(x)}$$

$$G(x,y) = y - f(x) \qquad G : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$$

$$\mathrm{Graph}_{\mathcal{X}_i}(f) = G^{-1}(0)$$

$$T_{(x,f(x))} = \ker dG(x,f(x))$$

$$= \{ (h, \langle \nabla_{\mathcal{X}} f(x), h\rangle) : h \in T_x \mathcal{X} \}$$

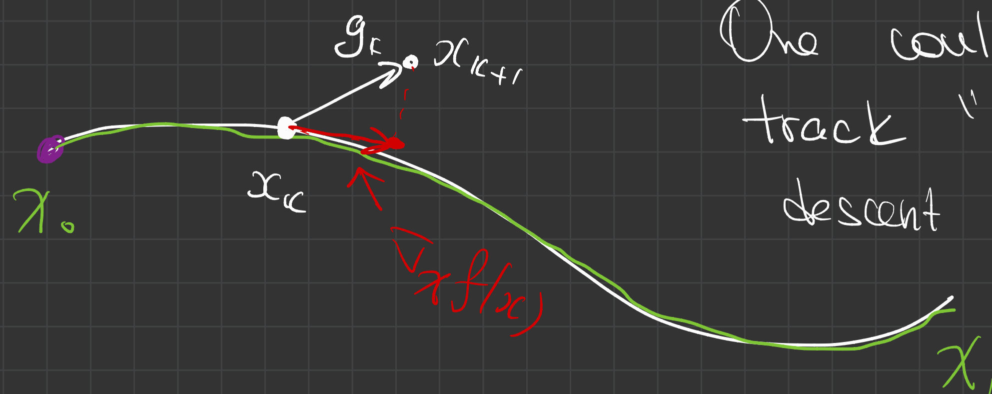
Now just write Whitney (a) for those tangent planes.

Ex

Moral

1) Conditions on stratification of Graph f give some behaviour of subgradients

2) $x_{k+1} = x_k - \gamma g_k \qquad g_x \in \partial f(x_k)$



One could try to track "Shadow Riemannian" descent along strata!

(this is what we eventually do, but with different conditions)

The big question: $\boxed{\text{what's up with o-minimality?}}$

I just realized that most likely my time is over...

Theorem: let $A \subseteq \mathbb{R}^d$ be a set $\boxed{\text{definable}}$ in some $\boxed{\text{o-minimal structure}}$, then it admits a C^p, Whitney (a) stratification

[Whitney (b)
Kuo-Verdier
Lipschitz*]

(some details hidden as usual)

o-minimality buys you a lot of properties!

Definition: $\mathcal{O} = (\mathcal{O}_n)_{n\geq 1}$ is an o-minimal structure, where $\mathcal{O}_n \subseteq \mathcal{P}(\mathbb{R}^n)$ if

$\hookrightarrow$ power set

Structure

1*) $Q : \mathbb{R}^n \to \mathbb{R}$, a polinomial

$\Rightarrow \{x \in \mathbb{R}^n : Q(x) = 0\} \in \mathcal{O}_n$

2). $\forall n \geq 1$ $\mathcal{O}_n$ is boolean algebra

3). $A \subseteq \mathcal{O}_n$, $B \in \mathcal{O}_m \Rightarrow A \times B \in \mathcal{O}_{n+m}$

4) $A \in \mathcal{O}_{n+1} \Rightarrow \Pi(A) \in \mathcal{O}_n$

$\hookrightarrow$ projection onto n first coord

o-minimal

5) $A \in \mathcal{O}_1 \Rightarrow$ A is a finite union of point and intervals

Def:

1) $A \subseteq \mathbb{R}^n$ is definable in $\mathcal{O}$ if $A \in \mathcal{O}_n$

2) $f: \mathbb{R}^n \to \mathbb{R}^d$ is definable in $\mathcal{O}$ if

$$\text{Graph} f \in \mathcal{O}_{n+d}$$

Examples:

I Semialgebraic sets (Tarski - Seinderberg)
$\mathbb{R}_{alg}$

II Semialgebraic + $\{(x, \exp(x)) : x \in \mathbb{R}\}$
$\mathbb{R}_{exp}$ (Wilkie)

III Subanalytic (Gabrielov)
$\mathbb{R}_{an}$

IV $\mathbb{R}_{an,exp}$ (van den Dries, Miller)

More explicit examples

1) $f(x,y) = \sqrt{x^2 + y^8}$ $\quad (\mathbb{R}_{alg})$

2) $f(x,y) = \dfrac{y}{\sin x}$ $\quad x \in (0,\pi)$ $\quad (\mathbb{R}_{an})$

3) $f(x,y) = x^2 \exp\left(- \dfrac{(y-\mu)^2}{x^4}\right)$ $\quad (\mathbb{R}_{exp})$

4) $f(x,y) = x^{\sqrt{2}} \log(\sin(y))$ $\quad x > 0, y \in (0,\pi)$

$(\mathbb{R}_{an,exp})$

5) $\sin x$ for $x \in \mathbb{R}$ is not definable in any structure!

Proof of 5) Assume the opposite, then

$$\{x : \sin(x) = 0\} \in \mathcal{O}_1$$

but it cannot be written as finite union of points and intervals

↯

How to work with these objects?

Informal theorem: if you can define a set using first order formula that only involves sets definable in the same structure, the outcome will be definable in this structure

Example: fix some $\mathcal{O}$, let $A \in \mathcal{O}_n$, then

$\overline{A} \in \mathcal{O}_n$

↳ closure

Proof

$$\overline{A} = \{x \in \mathbb{R}^n : \forall \varepsilon > 0 \ \exists y \in A \text{ s.t. } \forall i = 1, \dots, n \ |x_i - y_i|^2 < \varepsilon\}$$

formula

definable

$\boxtimes$

Some facts that make you enjoy your math life

<u>Thm</u>: $f : \mathbb{R} \to \mathbb{R}$ definable $\Rightarrow$ f is (finite) piece-wise $\{$cont, monotone, $C^p\}$

<u>Thm</u>: $A \subseteq \mathbb{R}^n$ definable, manifold $\Rightarrow$

i) $\dim(\partial A) < \dim(A)$

ii) A is path connected $\Leftrightarrow$ connected ☺

(think about graph of $\sin \frac{1}{x}$)

Final remark: all known to me neural nets are definable in some o-minimal structure.

Conclusion: o-minimality gives a lot of structure, is it useful for you or not is a different question.