PL condition and friends

Lectures by:
Dima Drusvyatskiy

I. A little history:

1. Boris Polyak ('63)
2. Stanisław Łojasiewicz ('63)
3. Krzystof Kurdyka ('98)

II. What does it really mean?

1. Descent Principle
2. Stability of sublevel sets

III. Nonsmooth extension & consequences

I A little history:

1. Boris Polyak ('63)
2. Stanisław Łojasiewicz ('63)
3. Krzystof Kurdyka ('98)

Boris Polyak

Let $f: \mathbb{R}^d \to \mathbb{R}$ be smooth & $\nabla f(\bar{x}) = 0$

1) Polyak Inequality (Polyak '63)

$$\boxed{\|\nabla f(x)\|^2 \geq \mu \left(f(x) - f(\bar{x}) \right) \quad \forall_{x \in B_r(\bar{x})}}$$

Polyak called this gradient dominance

Motivation: Grad. Descent converges linearly!

Proof sketch: Choose $\alpha < \frac{1}{\mathrm{lip}(\nabla f)}$

Set $x_{t+1} = x_t - \alpha \nabla f(x_t)$

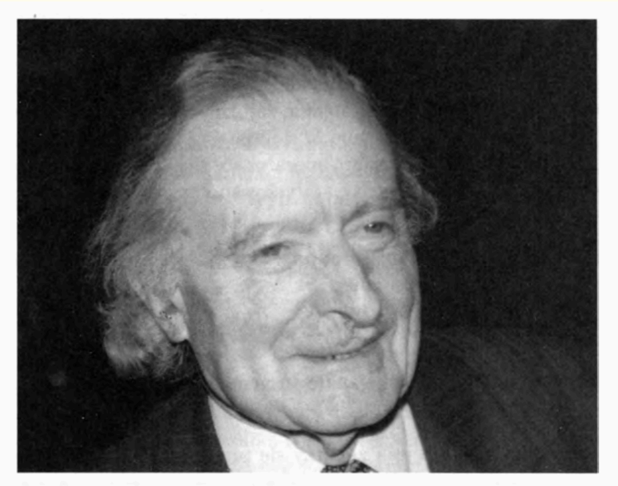
Then

$$f(x_t) - f(x_{t+1}) \geq \frac{\alpha}{2} \|\nabla f(x_t)\|^2 \geq \frac{\mu\alpha}{2} \left(f(x_t) - f(\bar{x}) \right)$$

Rearrange $\Rightarrow$ $f(x_{t+1}) - f(\bar{x}) \leq \left(1 - \frac{\mu\alpha}{2}\right) \left(f(x_t) - f(\bar{x}) \right)$

Then argue α small $\Rightarrow$ $x_i \in B_r(\bar{x})$ $\blacksquare$

Stanisław Łojasiewicz



2) Łojasiewicz Inequality (Ł '63)

$$\boxed{\|\nabla f(x)\|^{2/\theta} \geq \mu\, |f(x) - f(\bar{x})| \quad \forall x \in B_r(\bar{x})}$$

where $\theta \in (0,1]$

Łojasiewicz called it the grad. inequality

Thm (Łojasiewicz '63)
If f is analytic, then Ł-inequality holds at $\bar{x}$ with some μ, θ, r.

Motivation: gradient flow

Thm (Łojasiewicz '63)
If f satisfies Ł-inequality at every point, then every solution $\gamma(\cdot)$ of

$$\dot{\gamma}(t) = -\nabla f(\gamma(t))$$

has finite length and $\nabla f(x_\infty) = 0$

Surprisingly, not true for C^∞-functions (Palis - de Melo '82)

$\{\text{Limit points of } \gamma(\cdot)\} = \text{circle}$

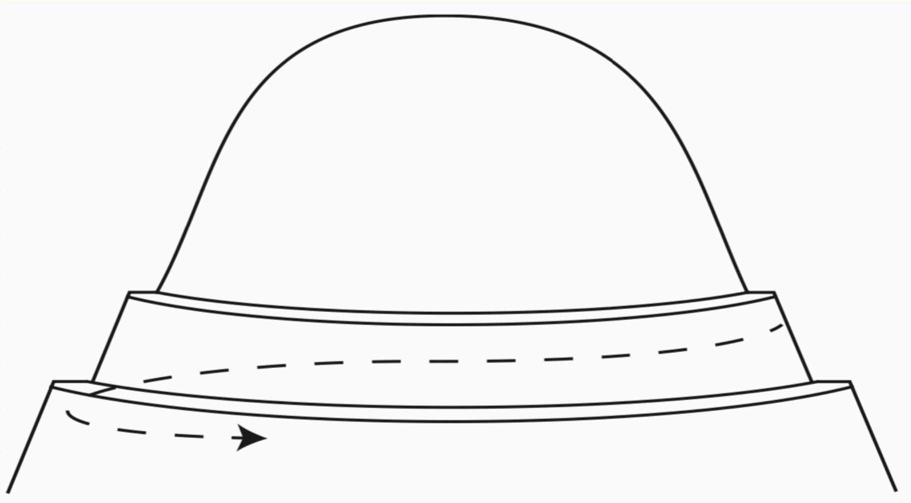
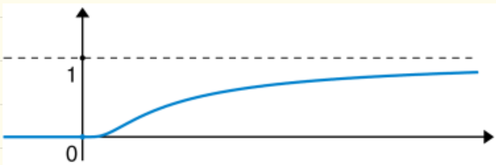


Fig from (Colding, Minicozzi II '15)

Example of C^∞ <u>but</u> not analytic

$$f(x) = \begin{cases} e^{-1/x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$



Krzystof Kurdyka

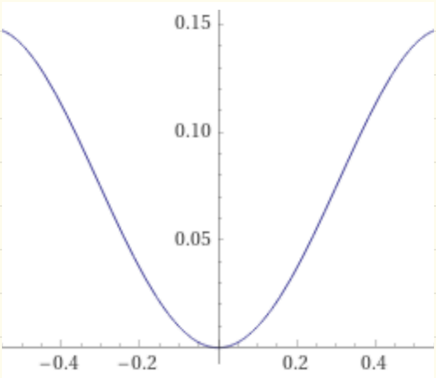
3) Kurdyka Inequality (Kurdyka '98)

Key idea: wlog assume $f(\bar{x}) = 0$.

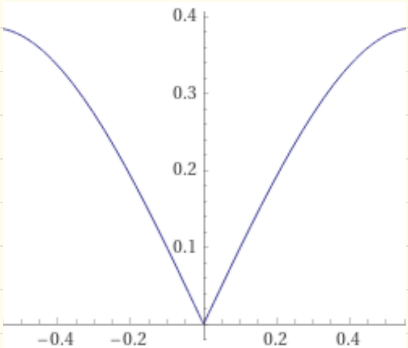
Polyak $\iff \|\nabla f(x)\|^2 \geq \mu \cdot f(x) \quad \forall x \in B_r(\bar{x})$

$\iff \|\nabla \sqrt{f(x)}\| \geq \frac{\sqrt{\mu}}{2} \quad \forall x \in B_r(\bar{x})$ with $f(x) \neq 0$

[Compute $\nabla \sqrt{f(x)} = \frac{\nabla f(x)}{2\sqrt{f(x)}}$]



$t \mapsto \sqrt{t}$



$f(x) = x^2(1-x^2)^2$ $\qquad \sqrt{f(x)} = |x|\,|1-x^2|$

$\varphi = \sqrt{\cdot}$ desingularizes f !

[Łojasiewicz $\iff \|\nabla f^{1-\frac{\theta}{2}}(x)\| \geq (1-\frac{\theta}{2})\mu^{\theta/2}$]

<u>Def</u> (Kurdyka '98)

A function $f: \mathbb{R}^n \to \mathbb{R}$ satisfies KŁ-inequality at $\bar{x}$ with $f(\bar{x})=0$ if $\exists_{\varepsilon, r>0}$ and $\varphi: [0,\varepsilon) \to [0,\infty)$ s.t.

1) $\varphi(0)=0$; φ is continuous, increasing, and differentiable on $(0,\varepsilon)$

2) Inequality $\|\nabla(\varphi \circ f)(x)\| \geq 1$ holds for all $x \in B_r(\bar{x}) \cap [0<f<\varepsilon]$.

Thm: (Kurdyka '98)
Every tame C^1 function satisfies KL.

Eg: x^r, $x^{1/p}$, $\log(x)$, $\exp(x)$, $\max\{\cdot,\cdot\}$,
their products, sums, compositions

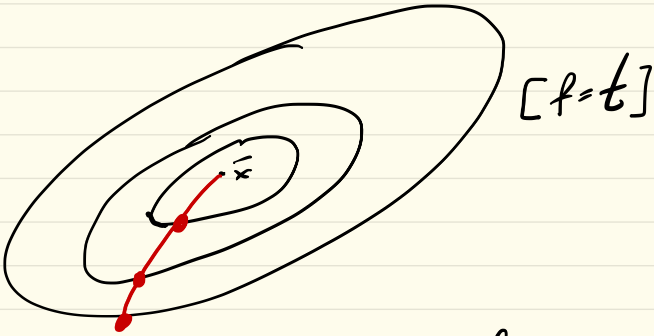
A little History:

"the foundations of topology are inadequate is manifest from the very beginning, in the form of 'false problems' (at least from the point of view of the topological intuition of shape)." These false problems include the existence of wild phenomena (space-filling curves, etc.) that add complications which are not essential. He states that a new field of topology is needed, one which should be adapted to a theory of "dévissage" (unscrewing) of stratified structures, a device which he was led to

"Sketch of a Program"
Grothendieck '84

van den Dries '98 answered this challenge(!)

pf sketch when f is coercive:



Define: For $t \in (0, \varepsilon]$ define

$$\omega(t) = \min_{x \in [f=t]} \|\nabla f(x)\|$$

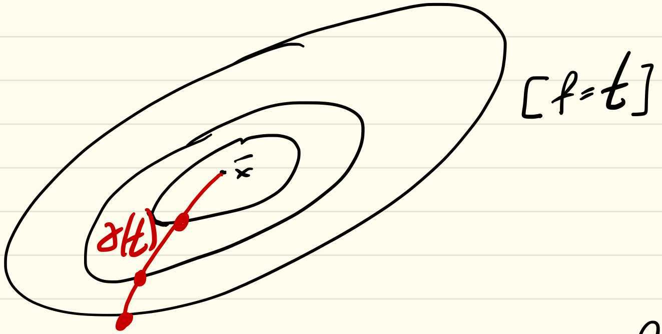
Define $\boxed{\varphi(t) \triangleq \int_0^t \frac{1}{\omega(s)}\, ds}$

$$\Rightarrow \|\nabla(\varphi \circ f)(x)\| = \varphi'(f(x)) \|\nabla f(x)\|$$

$$= \frac{1}{\omega(f(x))} \|\nabla f(x)\|$$

$$\geq 1$$

Key missing step: why is $\varphi(t) < \infty$?



Define: $\gamma(t) \in \operatorname*{argmin}_{x \in [f=t]} \|\nabla f(x)\|$

Claim: $\varphi(t) \leq \text{length}(\gamma) \overset{\boxed{\text{tame}}}{<} \infty$

Notice: $s = f(\gamma(s))$

$$\Rightarrow \quad 1 = (f \circ \gamma)'(s) = \langle \nabla f(\gamma(s)), \dot{\gamma}(s) \rangle$$

$$\leq \omega(s) \cdot \|\dot{\gamma}(s)\|$$

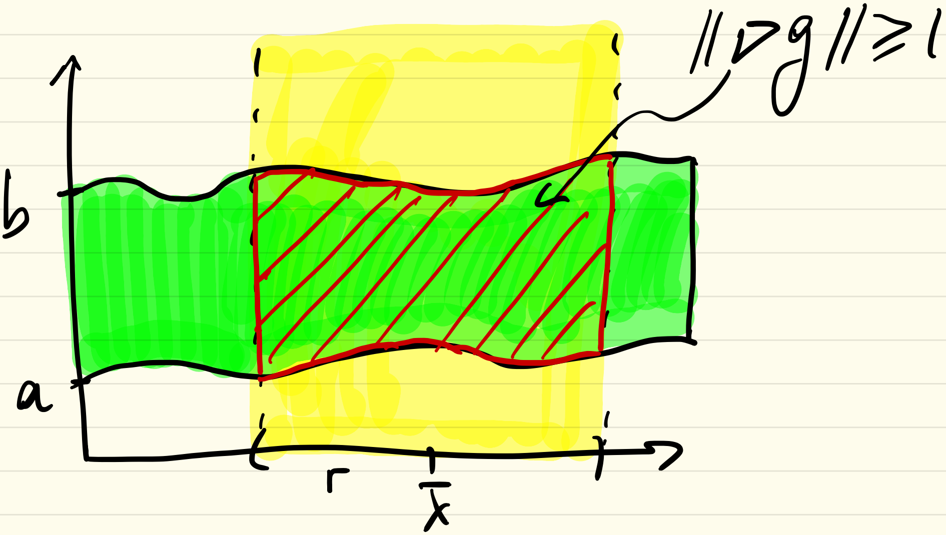
So $\frac{1}{\omega(s)} \leq \|\dot{\gamma}(s)\|$ $\blacksquare$

Summary:

(P), (t), (K)-inequalities reduce to:

$$\|\nabla g(x)\| \geq 1$$

for all $x \in B_r(\bar{x}) \cap [a < g < b]$



(II) What does it really mean for g?

(1) Descent Principle

(2) Stability of sublevel sets

...smoothness is irrelevant

Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ where $(\mathcal{X}, d)$ is a complete metric space

<u>Slope:</u> (De Giorgi et al. '80)

$$|\nabla f|(x) \triangleq \limsup_{y \to x} \frac{(f(x) - f(y))^+}{d(x,y)}$$

<u>Remark:</u>

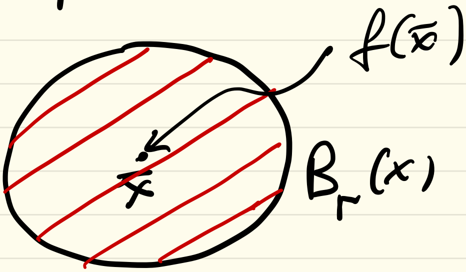
- f smooth $\Rightarrow$ $|\nabla f|(x) = \|\nabla f(x)\|$
- f convex $\Rightarrow$ $|\nabla f|(x) = \operatorname{dist}(0, \partial f(x))$

<u>Remark:</u> We'll assume f is closed.

① Decrease Principle:

Question: When is the following true?

$$f(\bar{x}) - \inf_{x \in B_r(\bar{x})} f(x) \geq \underbrace{\sigma}_{\text{constant}} \cdot r$$



Thm: Fix $\bar{x} \in \mathcal{X}$ and $r > 0$.

Set $$\sigma \triangleq \inf_{x \in B_r(\bar{x})} |\nabla f|(x)$$

Then

$$\boxed{f(\bar{x}) - \inf_{x \in B_r(\bar{x})} f(x) \geq \sigma r}$$

Follows from the more general thm.

Thm: (Ioffe 2000)
Fix $\bar{x}$ and $\alpha < f(\bar{x})$. Define

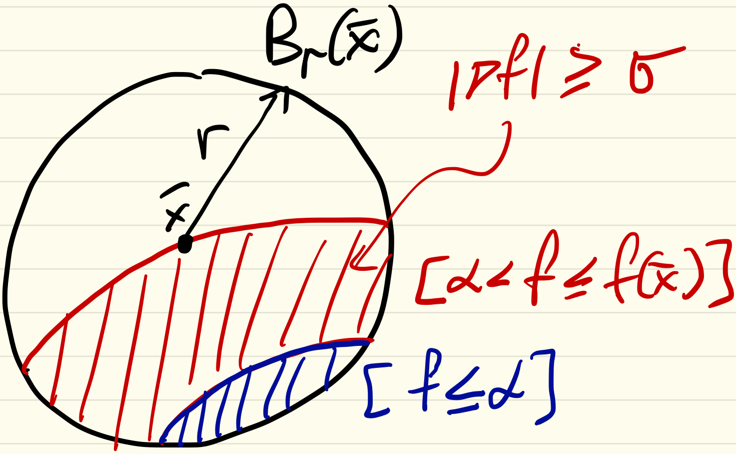
$$\sigma \triangleq \inf\{|\nabla f|(x) : x \in B_r(\bar{x}) \cap [\alpha < f \le f(\bar{x})]\}$$

Suppose $f(\bar{x}) - \alpha < r\sigma$.

Then $B_r(\bar{x}) \cap [f \le \alpha]$ is nonempty.

and

$$\boxed{\operatorname{dist}(\bar{x}, [f \le \alpha]) \le \sigma^{-1}(f(\bar{x}) - \alpha)}$$



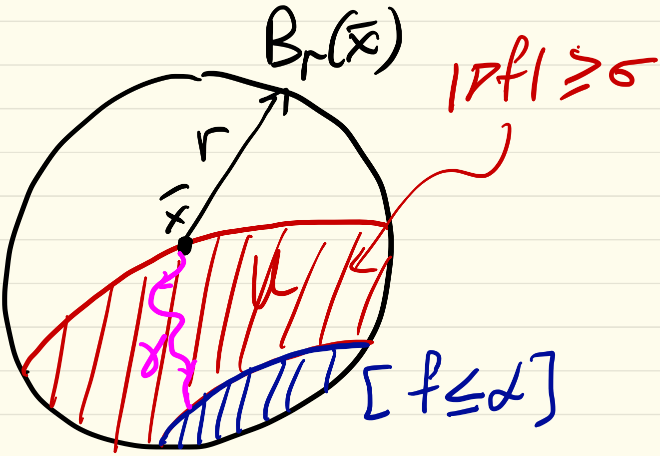
proof when $f:\mathbb{R}^d \to \mathbb{R}$ is smooth.

Set $\mathcal{U} = B_r(\bar{x}) \cap [0 < f \le f(\bar{x})]$

Let $\gamma : [0, \eta) \to \mathcal{U}$ such that

$$\dot{\gamma}(t) = -\nabla f(\gamma(t)) \quad \forall t > 0.$$

Extend γ to have maximal domain.



Compute:

$$f(\gamma(t)) - f(\bar{x}) = \int_0^t (f \circ \gamma)'(s) \, ds$$

$$= \int_0^t \langle \nabla f(\gamma(s)), \dot{\gamma}(s) \rangle \, ds$$

$$= \int_0^t - \|\nabla f(\gamma(s))\| \cdot \|\dot{\gamma}(s)\| \, ds$$

$$\leq -\sigma \int_0^s \|\nabla f(\gamma(s))\| \, ds$$

So

$$\text{length}(\gamma) \leq \sigma^{-1}(f(\bar{x}) - \alpha))$$

$\Rightarrow$ γ converges to some x^*

$\Rightarrow$ $\|\bar{x} - x^*\| \leq \sigma^{-1}(f(\bar{x}) - \alpha)) < r$

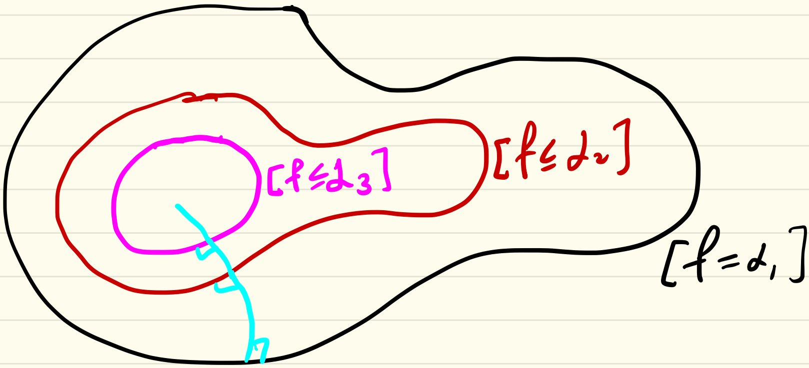
$\Rightarrow$ $x^* \in B_r(\bar{x}) \cap [f \leq \alpha]$

$\blacksquare$

We proved something more (!)

Fact: If φ is smooth, then

$$\left\{\begin{matrix}\text{grad flow} \\ \text{of } f\end{matrix}\right\} \overset{\text{as curves}}{=} \left\{\begin{matrix}\text{grad flow} \\ \text{of } \varphi\circ f\end{matrix}\right\}$$



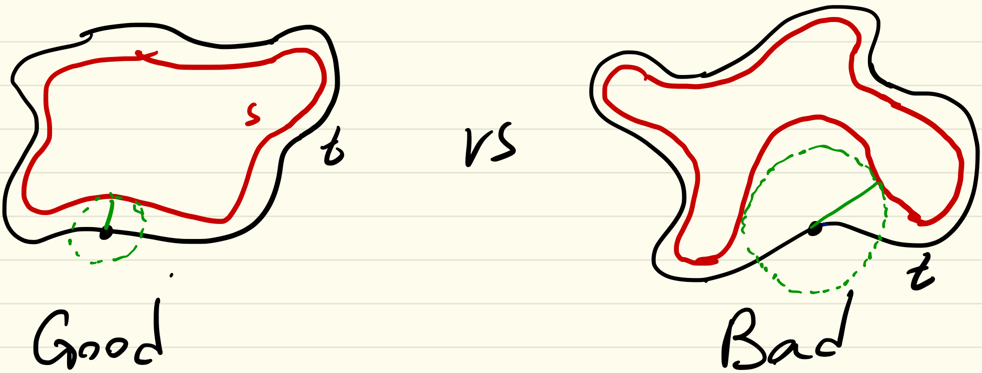
So we "essentially" proved that bounded grad flow of analytic/tame functions has finite length (Łojasiewicz '63) (Kurdyka '98)

② Stability of Sublevel Sets

<u>Question</u>: When is the following true?

$$d_H([f \le s], [f \le t]) \le \sigma^{-1} |s-t|$$

constant



Hausdorff Distance (in $\mathbb{R}^d$)

$$d_H(A,B) = \inf\left\{ \varepsilon > 0 : \begin{array}{l} A \subseteq B + \varepsilon \mathbb{B} \\ B \subseteq A + \varepsilon \mathbb{B} \end{array} \right\}$$

<u>Thm:</u> (Aze-Corvellac '04)

For any $a<b$ the following are equivalent:

① $|\nabla f|(x) \geq \sigma \quad \forall_{x \in [a<f<b]}$

② $d_H([f\leq t], [f\leq s]) \leq \sigma^{-1}|s-t|$ for all $s,t \in (a,b)$

III Nonsmooth extension & consequences

Beyond smoothness:

Let $f: \mathbb{R}^d \to \mathbb{R}$ be loc. Lipschitz.

Clarke Subdifferential:

$$\partial f(x) = \text{conv}\left\{ \lim_{y \to x} \nabla f(y) : y \in \text{dom}(\nabla f) \right\}$$

Subgradient Flow:

Absolutely continuous $\gamma(\cdot)$ with

$$\dot{\gamma}(t) \in \partial f(\gamma(t)) \quad \text{for a.e. } t$$

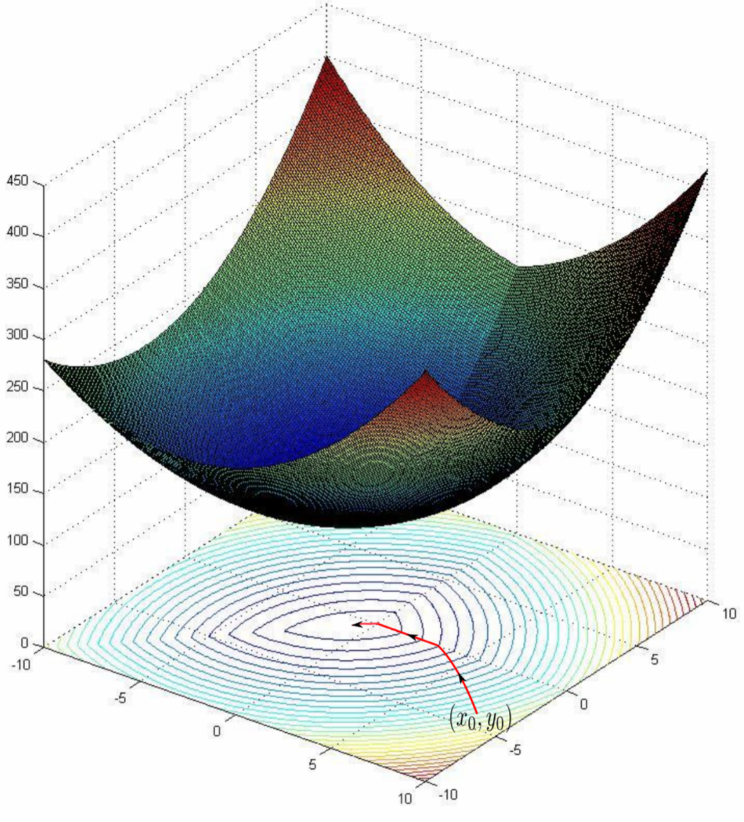


FIG. 3.1. $f(x,y) = \max\{x+y,\, |x-y|\} + x(x+1) + y(y+1) + 100$

All proofs discussed work relied on computing $(f\circ\gamma)'(t)$

Definition:

f is *path-differentiable* if for any absolutely continuous $\gamma(\cdot)$ and a.e. t it holds:

$$\underbrace{(f\circ\gamma)'(t)}_{\text{primal}} = \langle \underbrace{\partial f(\gamma(t))}_{\text{dual}}, \dot{\gamma}(t)\rangle,$$

Amazing fact

Includes virtually all functions you'll see in practice

Main Examples

1) convex & concave (Brezis '73)

2) tame

(Bolte et al. '06), (D-Ioffe-Lewis '15)

Some consequences of the assumption

① KŁ-inequality is valid

Subgradient curves have <u>finite length</u>

(Bolte, Daniilidis, Lewis, Shiota '06)

② Back-prop oracle for nonsmooth

(Bolte, Pauwels '19)

③ Proximal-type algos converge

(Attouch-Bolte-Svaiter '09)

(Bolte-Teboulle-Sabach '14)

Eg: Consider

$$\min_{x,y} \underbrace{g(x,y)}_{\text{smooth}} \quad \text{s.t. } x \in A,\ y \in B$$

Algo (PALM)

$$x_{t+1} = \mathrm{proj}_A\left(x_t - \frac{1}{L}\nabla_x g(x_t, y_t)\right)$$

$$y_{t+1} = \mathrm{proj}_B\left(y_t - \frac{1}{L}\nabla_y g(x_{t+1}, y_t)\right)$$

④ Limit points of SGD satisfy $0 \in \partial f(\bar{x})$

(Davis-D-Kakade-Lee '19)

⑤ Implicit regularization of SGD

(Du-Hu-Lee '18, Ji-Telgarsky '20)