A Bestiary of Counterexamples in Smooth Convex Optimization

Jérôme Bolte,
with Edouard Pauwels (Toulouse)

Toulouse School of Economics
Université Toulouse Capitole & ANITI, France

Nesterov's 65th birthday

Sochi, Russia
July 12$^{\text{th}}$, 2021

What can go wrong with smooth convex functions?

Framework:
Very small scale very smooth convex coercive problems!

- f convex in $C^k(\mathbb{R}^n,\mathbb{R})$ with k arbitrarily large, and eventually $n = 2$
- $C \subset \mathbb{R}^n$ closed convex, most of the time $C = \mathbb{R}^n$, solve $\min_C f$.

Many things work:

Complexity FOM, acceleration, tensor's methods...

Yet many open questions:

Convergence of some basic methods?
Directional convergence?
Rigidity à la Lojasiewicz?
Length of generalized central paths?...

Open questions...

(i) **Gauss-Seidel method - Block coordinate descent:** (1823, Gauss)

$$u_{i+1} = \operatorname*{argmin}_{u \in \mathbb{R}^p} f(u, v_i)$$

$$v_{i+1} = \operatorname*{argmin}_{v \in \mathbb{R}^q} f(u_{i+1}, v)$$

$f(u_i, v_i)$ converges to $\min f$ **but** what about $(u_i, v_i)_{i \in \mathbb{N}}$ if it is uniquely defined?

(ii) **Gradient descent with exact line search:** (1944, Curry)

$$x_{i+1} = \operatorname*{argmin}_{t \geq 0} f(x_i - t\nabla f(x_i))$$

$f(x_i)$ converges to $\min f$ **but** what about $(x_i)_{i \in \mathbb{N}}$?

(iii) **Bregman or mirror descent method** (1983, Nemirovskii-Yudin)

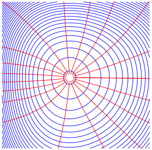
$$x_{i+1} = [\nabla h]^{-1} \left(\nabla h(x_i) - \frac{1}{L} \nabla f(x_i) \right)$$

with $Lh - f$ convex (relative smoothness).
$f(x_i)$ converges to $\min_{\overline{\operatorname{dom} h}} f$ **but** what about $(x_i)_{i \in \mathbb{N}}$?

Open questions: continuous time ODE...

(iv) **Directional convergence for gradient curves:**

$$x'(t) = -\nabla f(x(t)), \ t \geq 0. \qquad \text{(1847, Cauchy's descent)}$$



Theorem (Bruck 75)

f lower semicontinuous convex $\Rightarrow$ $x(t)$ converges whenever $\operatorname{argmin} f \neq \emptyset$

Assume f has positive definite Hessian on $\mathbb{R}^2 \setminus \{x^*\}$ where x^* is the unique minimizer.

Does the direction $\dfrac{x(t)-x^*}{\|x(t)-x^*\|}$ converges?

[Thom's conjecture 1989 when f is analytic. Solved by Kurdyka-Mostovsky-Parusinsky]

Intuitions around the Gauss-Seidel method in the plane

Consider $\min_{(u,v)\in\mathbb{R}^2} f(u,v)$ and a *uniquely* defined GS sequence

$$u_{i+1} = \operatorname*{argmin}_{u\in\mathbb{R}} f(u, v_i)$$
$$v_{i+1} = \operatorname*{argmin}_{v\in\mathbb{R}} f(u_{i+1}, v)$$

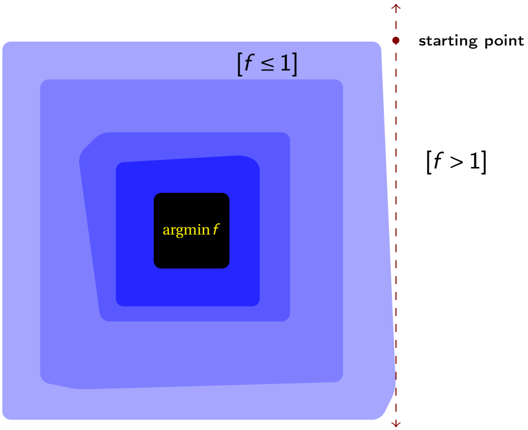
Writing the optimality conditions

- $\frac{\partial}{\partial u} f(u_{i+1}, v_i) = 0$ thus $\nabla f(u_{i+1}, v_i)$ parallel to the y-axis
- $\nabla f(u_{i+1}, v_{i+1})$ parallel to the x-axis

Rotating bumps

Imagine that we have a smooth convex function

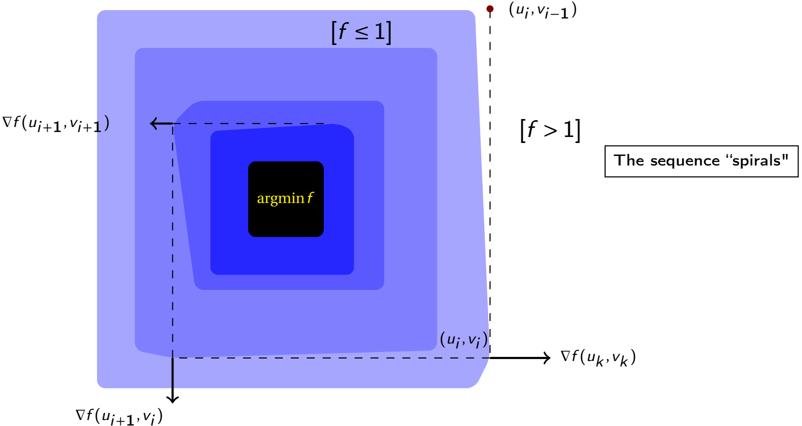
- having the bluish rounded squares as sublevel sets
- minimal on the black square



Rotating bumps yield a spiraling sequence

Imagine that we have a smooth convex function

- having the bluish rounded squares as sublevel sets
- minimal on the black square

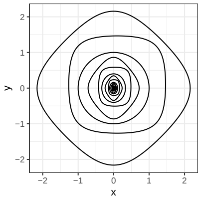


The interpolation problem: continuous case

♣ **“Strategy: Guess and draw a pathological sequence of convex sets and turn it into a counterexample"**

♣ Decreasing sequence $(T_i)_{i\in\mathbb{N}}$ of convex compact with $T_{i+1} \subset \mathrm{int}(T_i) \neq \emptyset$.

C^0 interpolation pb: **Find f convex such that the T_i are sublevel sets of f.**



Questions: de Finetti, Fenchel (50’s).

Kannai, Torralba (77, 96): f exists

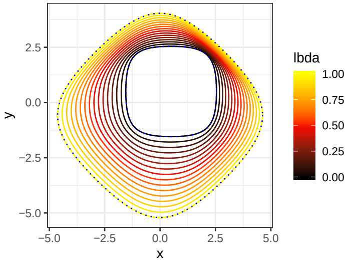
Interpolation between 2 sublevels: convex combination à la Minkowski

- Start with T_0, T_1: interpolate in between
- Set $\lambda \in [0,1]$ and define the (convex) set

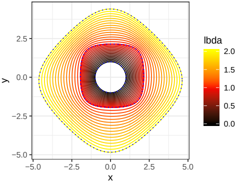
$$T_\lambda = (1-\lambda)\, T_0 + \lambda\, T_1.$$

Keep in mind the definition of T_λ

- Basic idea: "set $f(\partial T_\lambda) = 1 - \lambda$"



Interpolation with 3 or more sublevels



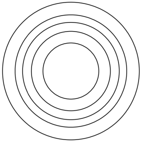
- Let us build f quasi-convex interpolating the T_i.
- **Choose** $\lambda_i \downarrow 0$.
  1. Assign λ_i to T_i.
  2. In between $[f \le \lambda] := \left(\frac{\lambda - \lambda_{i+1}}{\lambda_i - \lambda_{i+1}}\right) T_i + \left(\frac{\lambda_i - \lambda}{\lambda_i - \lambda_{i+1}}\right) T_{i+1}$.
- In addition get

$$\operatorname{argmin} f = \bigcap_{i \in \mathbb{N}} T_i, \text{ with } \min f = 0$$

Interpolating a sequence of concentric disks

But how to assign adequate values to enforce convexity?

Assume $(T_i)_{i \in \mathbb{N}}$ is a sequence of concentric disks:



$T_0, T_1, T_2, \ldots$

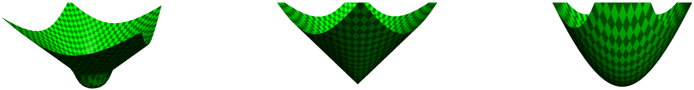


Figure: 3 interpolations: convexity (and smoothness) can easily be missed

Value assignation for convex interpolation

Translate the slope monotonicity characterization of convex functions in terms of sublevel sets.

- For $S \subset \mathbb{R}^p$ set $\sigma_S(x) = \sup\{\langle x, z \rangle, z \in S\}$ support function.

Theorem (de Finetti-Fenchel-Crouzeix)

$f : \mathbb{R}^p \to \mathbb{R}$ quasi-convex, T_λ the λ sublevel of f.
f is convex $\iff$ $F_v : \lambda \mapsto \sigma_{T_\lambda}(v)$ is concave for all fixed v is concave

- Choose $\lambda_i \to 0$ with a well adapted decrease rate.
  The ratio $(\lambda_i - \lambda_{i+1})/(\lambda_{i-1} - \lambda_i)$ must be lower than

$$K_i = \max_{||x^*||=1} \frac{\sigma_{T_{i-1}}(x^*) - \sigma_{T_i}(x^*)}{\sigma_{T_i}(x^*) - \sigma_{T_{i+1}}(x^*)}$$

The continuous interpolation result

Theorem (Kannai-Torralba)

$(T_i)_{i\in\mathbb{N}}$ convex compact in $\mathbb{R}^n$ such that $T_{i+1}\subset \operatorname{int} T_i$, $\forall i\geq 0$.

Then there is a continuous convex function f such that

$$T_i = [f\leq \lambda_i], \quad \text{for every } i\in\mathbb{N}$$

$$\operatorname{argmin} f = \bigcap_{i\in\mathbb{N}} T_i$$

Obstructions to the smooth case

We pertain to the plane $\mathbb{R}^2$ (...). Smoothness degree: $k \geq 2$

$(T_i)_{i\in\mathbb{N}}$ convex compact with C^k boundary, s.t. $T_{i+1} \subset \mathrm{int}(T_i) \neq \emptyset$.

Build $f : \mathbb{R}^2 \to \mathbb{R}$ convex C^k such that each T_i is a sublevel of f, and

$$\mathrm{argmin}\, f = \bigcap_{i\in\mathbb{N}} T_i$$

(T_i uniformly convex –> Positive Hessian out of the argmin?)

Issue I: Building C^k sublevels?

If $A \subset \mathbb{R}^2$ and $B \subset \mathbb{R}^2$ are C^k, $A + B$ is automatically C^k, uniquely when $k = 1, 2, 3, 4...$ (Kiselman 1987, Boman 1990).

Obstructions to the smooth case

We pertain to the plane $\mathbb{R}^2$ (...). Smoothness degree: $k \geq 2$

$(T_i)_{i\in\mathbb{N}}$ convex compact with C^k boundary, s.t. $T_{i+1} \subset \mathrm{int}(T_i) \neq \emptyset$.

Build $f : \mathbb{R}^2 \to \mathbb{R}$ convex C^k such that each T_i is a sublevel of f, and

$$\mathrm{argmin}\, f = \bigcap_{i\in\mathbb{N}} T_i$$

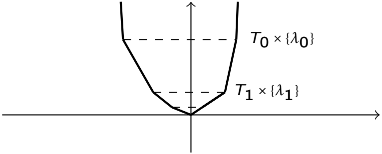
(T_i uniformly convex –> Positive Hessian out of the argmin?)

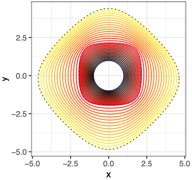
Issue I: Building C^k sublevels?

If $A \subset \mathbb{R}^2$ and $B \subset \mathbb{R}^2$ are C^k, $A + B$ is automatically C^k, uniquely when $k = 1, 2, 3, 4...$ (Kiselman 1987, Boman 1990). False for $k \geq 5$!

Obstructions to the smooth case II – III

Issue II: How to deal with more than 2 sets: “smooth the junctions"?

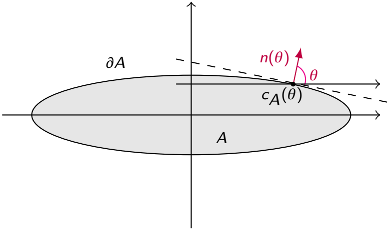




Issue III: Smooth near the argmin

$$\operatorname{argmin} f = \bigcap_{i=1}^{+\infty} T_i$$

Solve issue I: Positive curvature & smoothness of Minkowski sum



A smooth convex set, ∂A is C^2

$$n(\theta) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \in S^1,$$

Choose $c_A(\theta) \in \operatorname{argmax}\{\langle n(\theta), u\rangle : u \in A\}$

We have a **normal parametrization**, if $c_A : S^1 \to \partial A$ is uniquely-defined and is a diffeomorphism.
Then A is said to have *positive curvature.*

Lemma (Parametrization of a sum)

Let A, B with positive curvature

- *Then $c_{A+B} = c_A + c_B$ and $A + B$ has positive curvature*
- *A,B have C^k boundary then $A + B$ has C^k boundary*

Parametrization of rings: $T_i \setminus \text{int}\, T_i$

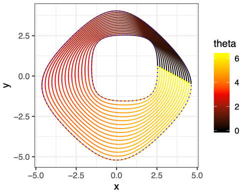
- $T_1 \subset \text{int}(T_0)$, compact C^k convex with positive curvature, then

$$T_\lambda = (1-\lambda)\,T_0 + \lambda\,T_1 \quad \text{is } C^k \qquad \forall \lambda \in [0,1]$$

- Gives a family of normal parametrizations of T_λ through

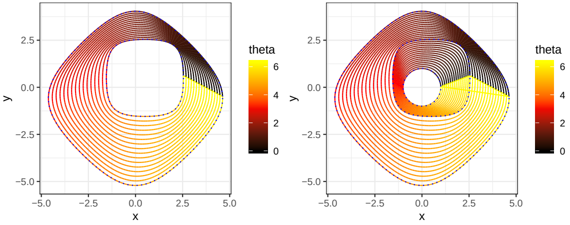
$$c_\lambda = (1-\lambda)c_0 + \lambda c_1.$$

- $c : (\lambda, \theta) \to c_\lambda(\theta)$ is a parametrization of the ring $T_0 \setminus \text{int}\, T_1$
  1. **Iso-value**: $\theta \to c_\lambda(\theta)$ is the normal parametrization of ∂T_λ
  2. **Iso-angle**: $\lambda \to c_\lambda(\theta)$ are monochromatic segments



Issue II: beyond 2 sets? The normal vector gluing issue

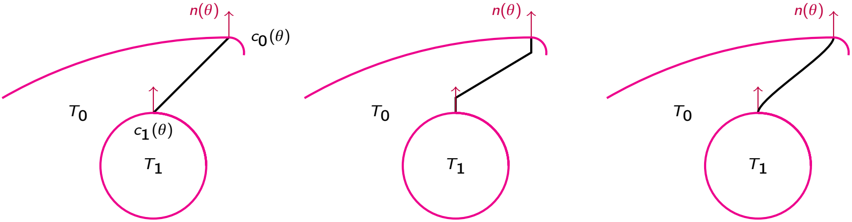
Iso-angle figures:



- On the yellow-black iso-angle line $n(\theta) = (1,0)$ and $\theta = 0$.
- ∇f colinear to $n(\theta)$ $\Rightarrow$ non differentiability at the junction.

Gluing normal and preserving convexity

Bend iso-angles $\lambda \to c_\lambda(\theta)$ to match their derivatives with the normals at endpoints



$c_\lambda(\theta) = (1-\lambda)c_0(\theta) + \lambda c_1(\theta)$ — Broken line — Bézier curve

$(\lambda,\theta) \to G(\lambda,\theta)$

⚠ **Need to preserve convexity of the sets: fundamental properties of Bézier curves and Bernstein polynomial**

Smoothing: angles and values

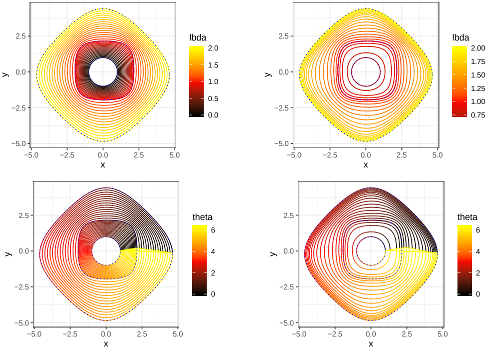


Figure: Left: Raw Minkowski sum
Right: "Smoothed pictures"

The C^k convex interpolation theorem

Let $k \geq 2$.

Theorem (B-Pauwels, 2020)

$(T_i)_{i\in\mathbb{Z}}$ a sequence of C^k convex compact subsets of $\mathbb{R}^2$ with positive curvature, with

$$T_{i+1} \subset \operatorname{int} T_i \neq \emptyset \text{ for all } i.$$

Then there exists a C^k convex function f having the T_i as sublevel sets.

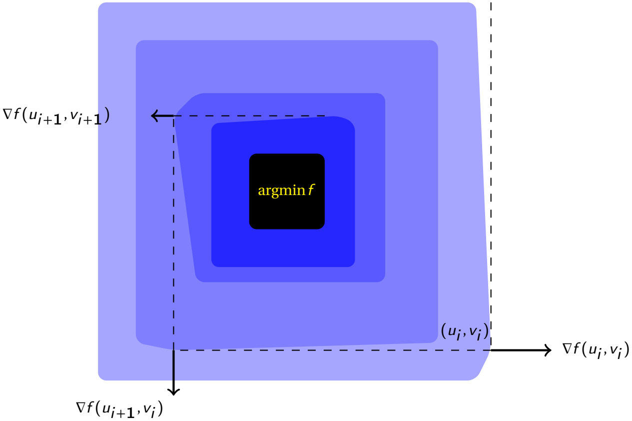
In addition the Hessian of f is positive definite out of

$$\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i.$$

Gauss-Seidel method does not converge

Choose the rotating bumps sequence $(T_i)_{i\in\mathbb{N}}$ and **add slight curvature...**

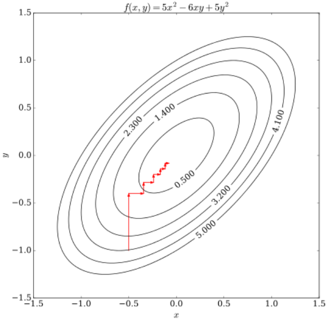
$$\bigcap_{i=1}^{\infty} T_i = T_\infty = \text{black rounded square}$$



Exact line search does not converge

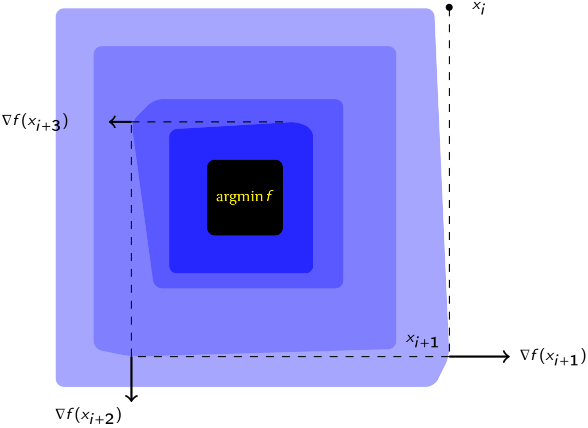
$$x_{i+1} = \operatorname{argmin}\left\{f(x_i - t\nabla f(x_i)) : t \geq 0\right\}$$

Optimality condition: $\langle x_{i+1} - x_i, \nabla f(x_{i+1})\rangle = 0.$



Exact line search does not converge

Same sequence $(T_i)_{i \in \mathbb{N}}$ as in Gauss-Seidel! Same starting point and same sequence of points!

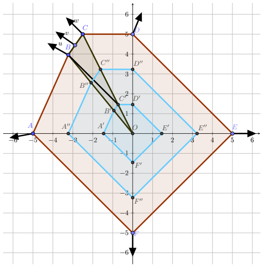


Secant convergence for convex potential

f convex C^k with compact sublevel sets, positive Hessian out of $\operatorname{argmin} f = \{0\}$

$$x'(t) = -\nabla f(x(t)), \qquad \text{does the secant } \frac{x(t)}{\|x(t)\|} \text{ converge?}$$

No! **"Build a swirling-decreasing sequence of repulsing-triangles".**



Triangles exist at all scale near 0 and along "many" directions.

Thus if $\frac{x(t)}{\|x(t)\|}$ converges, x should stay forever in some triangle...

Secant convergence for convex potential

A recent construction from Daniilidis-Haddou-Ley following our work has an appealing form.

“Rotating ellipses yield spiraling curves"

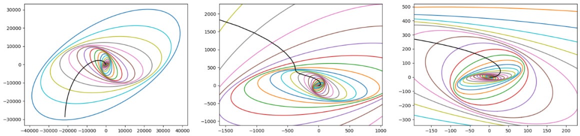


Figure: The sublevel sets and zoomed images from Daniilidis-Haddou-Ley

Mirror descent aka Bregman minimization: $\min_C f$.

- h a Legendre function on C
  - h is a convex function $\operatorname{int} C \subset \operatorname{dom} h \subset C$
  - h smooth on the interior
  - ∇h is a diffeomorphism from $\operatorname{int} C$ to its image
  - Blow-up: when $z_i \in \operatorname{int} C$ is such that $\operatorname{dist}(z_i, \partial C) \to 0$ then

$$\lim \|\nabla h(z_i)\| = +\infty.$$

- Examples: $x \log x, -\log x, -\sqrt{x}$ for $x \in C = \mathbb{R}_+$
- The problem
  - Let f convex with $Lh - f$ convex ("relative smoothness")
  - Goal: minimize f over $C = \overline{\operatorname{dom} h}$
  - Run $x_{i+1} = (\nabla h)^{-1}(\nabla h(x_i) - \text{step} \nabla f(x_i))$ with step $< 1/L$.
- Then

$$\lim_{i \to +\infty} f(x_i) = \min_C f$$

But convergence?

Mirror descent aka Bregman minimization: $\min_C f$.

- h a Legendre function on C
  - h is a convex function $\operatorname{int} C \subset \operatorname{dom} h \subset C$
  - h smooth on the interior
  - ∇h is a diffeomorphism from $\operatorname{int} C$ to its image
  - Blow-up: when $z_i \in \operatorname{int} C$ is such that $\operatorname{dist}(z_i, \partial C) \to 0$ then

$$\lim \|\nabla h(z_i)\| = +\infty.$$

- Examples: $x \log x, -\log x, -\sqrt{x}$ for $x \in C = \mathbb{R}_+$
- The problem
  - Let f convex with $Lh - f$ convex ("relative smoothness")
  - Goal: minimize f over $C = \overline{\operatorname{dom} h}$
  - Run $x_{i+1} = (\nabla h)^{-1}(\nabla h(x_i) - \text{step}\nabla f(x_i))$ with step $< 1/L$.
- Then

$$\lim_{i \to +\infty} f(x_i) = \min_C f$$

But convergence?

Mirror descent aka Bregman minimization: $\min_C f$.

- h a Legendre function on C
  - h is a convex function $\operatorname{int} C \subset \operatorname{dom} h \subset C$
  - h smooth on the interior
  - ∇h is a diffeomorphism from $\operatorname{int} C$ to its image
  - Blow-up: when $z_i \in \operatorname{int} C$ is such that $\operatorname{dist}(z_i, \partial C) \to 0$ then

$$\lim \|\nabla h(z_i)\| = +\infty.$$

- Examples: $x \log x, -\log x, -\sqrt{x}$ for $x \in C = \mathbb{R}_+$
- The problem
  - Let f convex with $Lh - f$ convex ("relative smoothness")
  - Goal: minimize f over $C = \overline{\operatorname{dom} h}$
  - Run $x_{i+1} = (\nabla h)^{-1}(\nabla h(x_i) - \text{step} \nabla f(x_i))$ with step $< 1/L$.
- Then

$$\lim_{i \to +\infty} f(x_i) = \min_C f$$

But convergence?

Mirror descent aka Bregman minimization: $\min_C f$.

- h a Legendre function on C
  - h is a convex function $\operatorname{int} C \subset \operatorname{dom} h \subset C$
  - h smooth on the interior
  - ∇h is a diffeomorphism from $\operatorname{int} C$ to its image
  - Blow-up: when $z_i \in \operatorname{int} C$ is such that $\operatorname{dist}(z_i, \partial C) \to 0$ then

$$\lim \|\nabla h(z_i)\| = +\infty.$$

- Examples: $x \log x, -\log x, -\sqrt{x}$ for $x \in C = \mathbb{R}_+$
- The problem
  - Let f convex with $Lh - f$ convex ("relative smoothness")
  - Goal: minimize f over $C = \overline{\operatorname{dom} h}$
  - Run $x_{i+1} = (\nabla h)^{-1}(\nabla h(x_i) - \text{step}\nabla f(x_i))$ with $\text{step} < 1/L$.
- Then

$$\lim_{i \to +\infty} f(x_i) = \min_C f$$

But convergence?

Mirror descent aka Bregman minimization: $\min_C f$.

- $\min_{C:=[0,1]^2} f(x) := \langle e_1, x\rangle$ that is $\min_{[0,1]^2} x_1$

  f is relatively smooth with respect to *any kernel* and for all $L > 0$
- **Algorithm:**

$$x_{i+1} = (\nabla h)^{-1}(\nabla h(x_i) - e_1)$$
$$\nabla h(x_{i+1}) - \nabla h(x_i) = -e_1$$

  Thus by telescopic sum

$$\nabla h(x_{i+1}) - \nabla h(x_0) = -(i+1)e_1$$

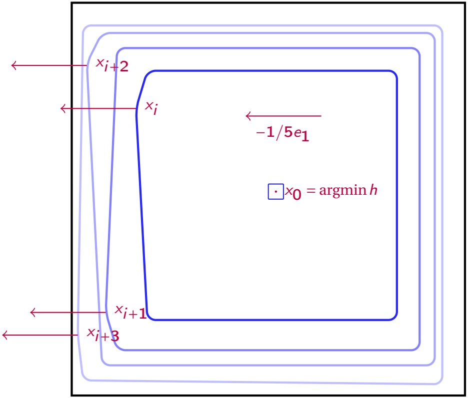
- Counterexample

Theorem (B-Pauwels 2020)

There is $h\colon [0,1]^2 \mapsto \mathbb{R}$ Legendre, continuous on $[0,1]^2$, C^k on $(0,1)^2$ such that the accumulation set of x_i is the entire left edge of the square $[0,1]^2$.

A Legendre function with sliding bumps

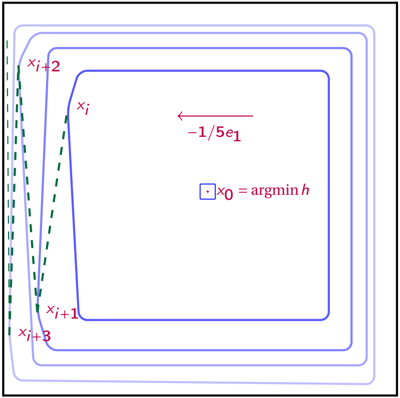
- Recall $\nabla h(x_{i+1}) - \nabla h(x_0) = -(i+1)e_1$.



- $C = [0,1]^2$
- In **blue** level lines of h
- $\nabla h(x_0) = 0 \Rightarrow \nabla h(x_{i+1}) = -(i+1)e_1$

A Legendre function with sliding bumps

- Recall $\nabla h(x_{i+1}) - \nabla h(x_0) = -(i+1)e_1$.



- $C = [0,1]^2$
- In **blue** level lines of h
- $\nabla h(x_0) = 0 \Rightarrow \nabla h(x_{i+1}) = -(i+1)e_1$

Other counterexamples

- a C^k convex coercive function failing to have the KL property (as in Bolte-Daniilidis-Ley-Mazet 2009 but C^k and not merely C^2).
- A non converging Newton's curve

$$x' = -\nabla^2 f(x)^{-1} \nabla f(x)$$

- a Tikhonov path of infinite length (à la Torralba 1996).
- a nonconverging central path
- nonconverging Hessian-Riemannian gradient curves

Conclusion

Today's results

- smooth convex interpolation in the plane
- counterexamples

What you did not see:

- Other counterexamples
- Subtle issues: hessian, global Lipschitz properties, Legendre functions
- Computational difficulties of the construction

What we are working on:

- More counterexamples
- Finite dimensional setting
- C^∞ interpolation?
- Removing the curvature assumption??