Curiosities and counterexamples in smooth convex optimization

EDOUARD PAUWELS (IRIT, TOULOUSE 3)
collaboration with JÉRÔME BOLTE (TSE, TOULOUSE 1)

Learning and Optimization in Luminy

October 2022, CIRM, Lumniy

Forewords:

We work in $\mathbb{R}^2$ (negative results).

$k \geq 2$ is an arbitrary smoothness index.

Functions on $\mathbb{R}^2$: convex, compact sublevel sets, C^k.

Exact line search gradient descent

$f : \mathbb{R}^p \to \mathbb{R}$ convex, L-Lipschitz gradient, a minimum x^*. Set $x_0 \in \mathbb{R}^p$ and for $k \in \mathbb{N}$,

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k).$$

$(x_k)_{k \in \mathbb{N}}$ converges toward a solution (Féjer monotonicity: for any solution x^*, $\|x_k - x^*\|$ is monotone).

Exact line search gradient descent

$f : \mathbb{R}^p \to \mathbb{R}$ convex, L-Lipschitz gradient, a minimum x^*. Set $x_0 \in \mathbb{R}^p$ and for $k \in \mathbb{N}$,

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k).$$

$(x_k)_{k \in \mathbb{N}}$ converges toward a solution (Féjer monotonicity: for any solution x^*, $\|x_k - x^*\|$ is monotone).

Exact line search

$$x_{k+1} \in \arg\min_z f(z), \qquad s.t. \quad z \in x_k - \mathbb{R}_+ \nabla f(x_k).$$

Exact line search gradient descent

$f : \mathbb{R}^p \to \mathbb{R}$ convex, L-Lipschitz gradient, a minimum x^*. Set $x_0 \in \mathbb{R}^p$ and for $k \in \mathbb{N}$,

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k).$$

$(x_k)_{k\in\mathbb{N}}$ converges toward a solution (Féjer monotonicity: for any solution x^*, $\|x_k - x^*\|$ is monotone).

Exact line search

$$x_{k+1} \in \arg\min_z f(z), \qquad s.t. \quad z \in x_k - \mathbb{R}_+ \nabla f(x_k).$$

Contribution: does not converge in general,

- dimension 2
- f convex, C^k, coercive
- unique minimum along the trajectory (selection in the argmin).

Constructive proof

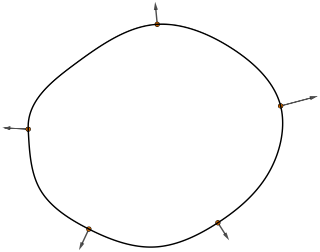
Construct f, C^k, with oscillating gradients around its minimum.

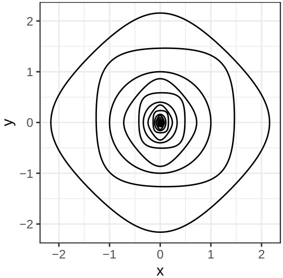
Constructive proof

Construct f, C^k, with oscillating gradients around its minimum.

Main ideas:

- Gradient orthogonal to level sets.
- Interpolate a sequence of level sets.





Additional counterexamples

C^k convex counter examples on $\mathbb{R}^2$

Additional counterexamples

C^k convex counter examples on $\mathbb{R}^2$

- **Thom's Gradient conjecture:** convergence of secants of gradient flow (Kurdyka-Mostowski-Parusinski 2000, Daniilidis-Haddou-Ley 2020)
- **Kurdyka-Łojasiewicz:** convex function with no error bound (Bolte, Daniilidis, Ley, Mazet 2009).
- **Tikhonov regularization:** infinite length regularization path (Torralba 1996).
- **Newton flow:** non convergence ($\nabla^2 f$ PD outside $\arg\min f$)

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

Additional counterexamples

C^k convex counter examples on $\mathbb{R}^2$

- **Thom's Gradient conjecture:** convergence of secants of gradient flow (Kurdyka-Mostowski-Parusinski 2000, Daniilidis-Haddou-Ley 2020)
- **Kurdyka-Łojasiewicz:** convex function with no error bound (Bolte, Daniilidis, Ley, Mazet 2009).
- **Tikhonov regularization:** infinite length regularization path (Torralba 1996).
- **Newton flow:** non convergence ($\nabla^2 f$ PD outside $\arg\min f$)

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

Algorithmic counter examples:

Additional counterexamples

C^k convex counter examples on $\mathbb{R}^2$

- **Thom's Gradient conjecture:** convergence of secants of gradient flow (Kurdyka-Mostowski-Parusinski 2000, Daniilidis-Haddou-Ley 2020)
- **Kurdyka-Łojasiewicz:** convex function with no error bound (Bolte, Daniilidis, Ley, Mazet 2009).
- **Tikhonov regularization:** infinite length regularization path (Torralba 1996).
- **Newton flow:** non convergence ($\nabla^2 f$ PD outside $\arg\min f$)

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

Algorithmic counter examples:

- **Bregman gradient / mirror descent:** non convergence (h Bregman, $c \in \mathbb{R}^2$).

$$x_{k+1} = \nabla h^*(\nabla h(x_k) + c).$$

Additional counterexamples

C^k convex counter examples on $\mathbb{R}^2$

- **Thom's Gradient conjecture:** convergence of secants of gradient flow (Kurdyka-Mostowski-Parusinski 2000, Daniilidis-Haddou-Ley 2020)
- **Kurdyka-Łojasiewicz:** convex function with no error bound (Bolte, Daniilidis, Ley, Mazet 2009).
- **Tikhonov regularization:** infinite length regularization path (Torralba 1996).
- **Newton flow:** non convergence ($\nabla^2 f$ PD outside $\arg\min f$)

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

Algorithmic counter examples:

- **Bregman gradient / mirror descent:** non convergence (h Bregman, $c \in \mathbb{R}^2$).

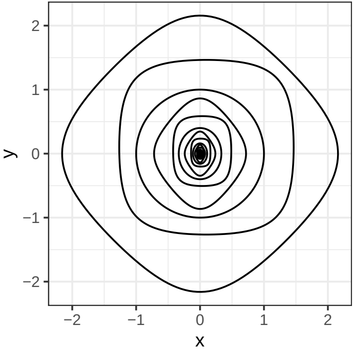
$$x_{k+1} = \nabla h^*(\nabla h(x_k) + c).$$

- **Frank-Wolfe algorithm:** non convergence (with C. Combettes).

$$x_{k+1} = (1 - \gamma_k) x_k + \gamma_k v_k, \qquad\qquad v_k \in \arg\min_{v \in \mathcal{C}} \langle \nabla f(x_k), v \rangle$$

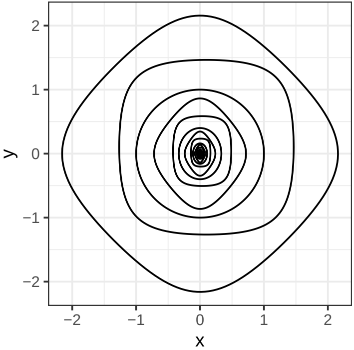
Plan

Convex interpolation problem: Let $(T_i)_{i\in-\mathbb{N}}$ be convex compact with $T_{i-1} \subset \mathrm{int}(T_i) \neq \emptyset$ for $i \in -\mathbb{N}$. Construct f convex with T_i as a sublevel for all $i \in -\mathbb{N}$.



Description of the problem

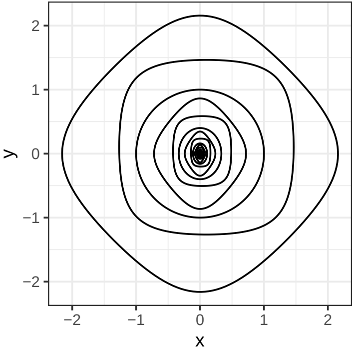
Convex interpolation problem: Let $(T_i)_{i\in-\mathbb{N}}$ be convex compact with $T_{i-1} \subset \mathrm{int}(T_i) \neq \emptyset$ for $i \in -\mathbb{N}$. Construct f convex with T_i as a sublevel for all $i \in -\mathbb{N}$.



Early questionning: de Finetti, Fenchel (50's).

Description of the problem

Convex interpolation problem: Let $(T_i)_{i\in -\mathbb{N}}$ be convex compact with $T_{i-1} \subset \mathrm{int}(T_i) \neq \emptyset$ for $i \in -\mathbb{N}$. Construct f convex with T_i as a sublevel for all $i \in -\mathbb{N}$.



Early questionning: de Finetti, Fenchel (50's).

Kannai-Torralba (77, 96): f exists (convex continuous).

Affine sublevel set interpolation with Minkowski sums

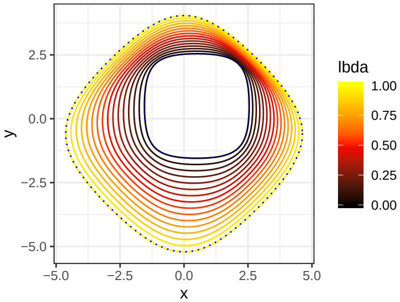
- T_0, sublevel set of value 0, T_1 sublevel set of value $f_1 > 0$.

Affine sublevel set interpolation with Minkowski sums

- T_0, sublevel set of value 0, T_1 sublevel set of value $f_1 > 0$.
- For $\lambda \in [0, f_1]$, set $T_\lambda = (1 - \lambda/f_1)\,T_0 + \lambda/f_1\,T_1$ the λ sublevel set.

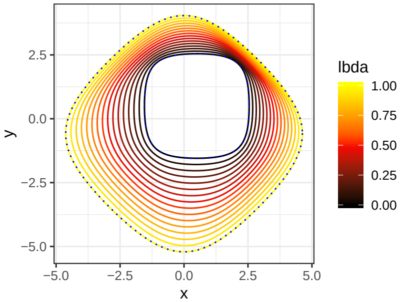
Affine sublevel set interpolation with Minkowski sums

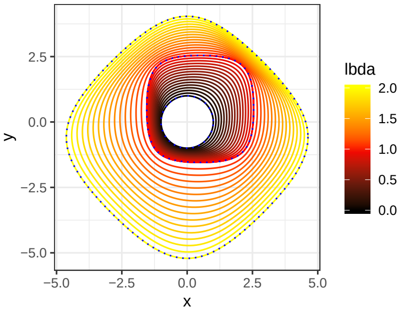
- T_0, sublevel set of value 0, T_1 sublevel set of value $f_1 > 0$.
- For $\lambda \in [0, f_1]$, set $T_\lambda = (1 - \lambda/f_1)T_0 + \lambda/f_1 T_1$ the λ sublevel set.



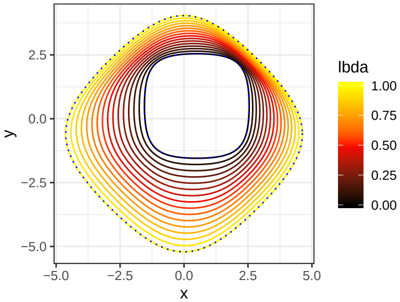
Affine sublevel set interpolation with Minkowski sums

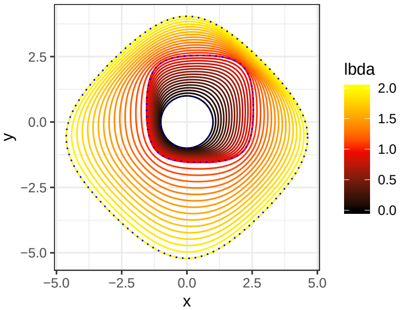
- T_0, sublevel set of value 0, T_1 sublevel set of value $f_1 > 0$.
- For $\lambda \in [0, f_1]$, set $T_\lambda = (1 - \lambda/f_1)\, T_0 + \lambda/f_1\, T_1$ the λ sublevel set.





- T_0, sublevel set of value 0, T_1 sublevel set of value $f_1 > 0$.
- For $\lambda \in [0, f_1]$, set $T_\lambda = (1 - \lambda/f_1)T_0 + \lambda/f_1 T_1$ the λ sublevel set.





Kannai-Torralba (77, 96):

- Connect Minkowski interpolation.
- Choose function values to enforce convexity.

Add smoothness

Smooth convex interpolation problem:
Let $(T_i)_{i \in -\mathbb{N}}$ be convex compact with C^k boundary, such that $T_{i-1} \subset \mathrm{int}(T_i) \neq \emptyset$ for $i \in -\mathbb{N}$. Construct f convex C^k, such that, for all $i \in -\mathbb{N}$, T_i is a sublevel set of f.

Smooth convex interpolation problem:
Let $(T_i)_{i\in-\mathbb{N}}$ be convex compact with C^k boundary, such that $T_{i-1} \subset \mathrm{int}(T_i) \neq \emptyset$ for $i \in -\mathbb{N}$. Construct f convex C^k, such that, for all $i \in -\mathbb{N}$, T_i is a sublevel set of f.

Minkowski addition and smoothness?

- There are $A, B \subset \mathbb{R}^p$ convex with C^∞ boundary, such that $A + B$ does not have C^2 boundary (Kiselman 1986)

Plan

Positive curvature

Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Positive curvature

Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Gauss map: $A \subset \mathbb{R}^p$ compact, convex, nonempty interior, C^1 boundary.
$n_A \colon \partial A \to S_{p-1}$ unit outer normal on ∂A.

Positive curvature

Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Gauss map: $A \subset \mathbb{R}^p$ compact, convex, nonempty interior, C^1 boundary.
$n_A \colon \partial A \to S_{p-1}$ unit outer normal on ∂A.
Diffeomorphism if and only if, ∂A is C^2 with positive curvature, denoted C^2_+.

Positive curvature

Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Gauss map: $A \subset \mathbb{R}^p$ compact, convex, nonempty interior, C^1 boundary.
$n_A \colon \partial A \to S_{p-1}$ unit outer normal on ∂A.
Diffeomorphism if and only if, ∂A is C^2 with positive curvature, denoted C^2_+.

Normal parametrization: inverses of n_A, $(A \in C^2_+)$. $c_A \colon S_{p-1} \to \partial A$.

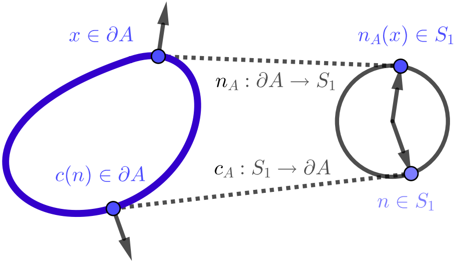
$$c_A \colon n \mapsto \mathrm{argmax}_{x \in A} \langle n, x \rangle .$$

Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Gauss map: $A \subset \mathbb{R}^p$ compact, convex, nonempty interior, C^1 boundary.
$n_A \colon \partial A \to S_{p-1}$ unit outer normal on ∂A.
Diffeomorphism if and only if, ∂A is C^2 with positive curvature, denoted C^2_+.

Normal parametrization: inverses of n_A, ($A \in C^2_+$). $c_A \colon S_{p-1} \to \partial A$.

$$c_A \colon n \mapsto \operatorname{argmax}_{x \in A} \langle n, x \rangle .$$

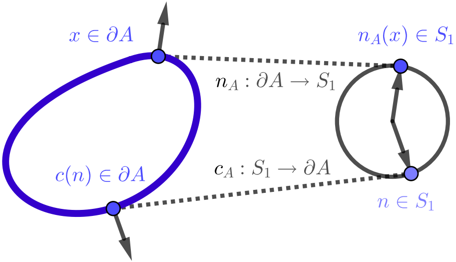


Schneider: Convex Bodies: The Brunn-Minkowski Theory. Section 2.5.

Gauss map: $A \subset \mathbb{R}^p$ compact, convex, nonempty interior, C^1 boundary.
$n_A \colon \partial A \to S_{p-1}$ unit outer normal on ∂A.
Diffeomorphism if and only if, ∂A is C^2 with positive curvature, denoted C^2_+.

Normal parametrization: inverses of n_A, $(A \in C^2_+)$. $c_A \colon S_{p-1} \to \partial A$.

$$c_A \colon n \mapsto \operatorname{argmax}_{x \in A} \langle n, x \rangle \,.$$

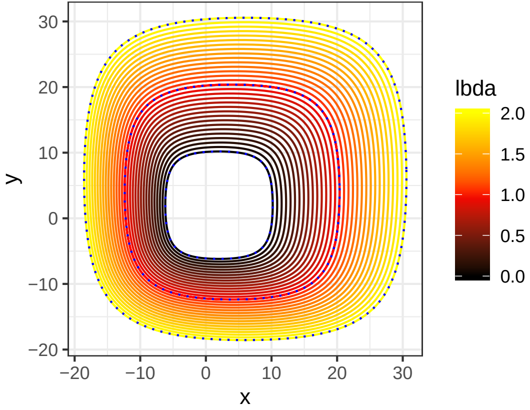


Get smoothness back: $A, B \in C^2_+$, C^k boundary, then $A + B \in C^2_+$ with C^k boundary

$$c_{A+B} = c_A + c_B$$

Smooth interpolation an easy example

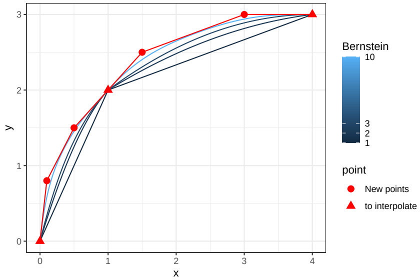
Homothetic sublevel sets.



Idea: Enforce a similar behavior for orders $0, 1, \ldots, k$.

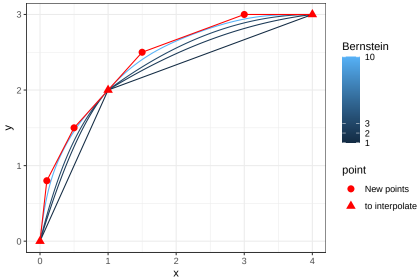
Nonlinear interpolation: idea from Bézier curves and Bernstein interpolation

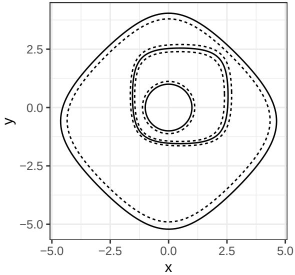
Preserves monotonicity, concavity, control derivatives up to order k at endpoints



Nonlinear interpolation: idea from Bézier curves and Bernstein interpolation

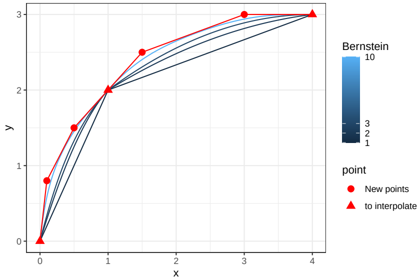
Preserves monotonicity, concavity, control derivatives up to order k at endpoints

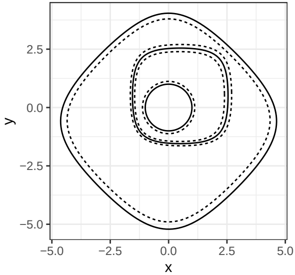




Nonlinear interpolation: idea from Bézier curves and Bernstein interpolation

Preserves monotonicity, concavity, control derivatives up to order k at endpoints



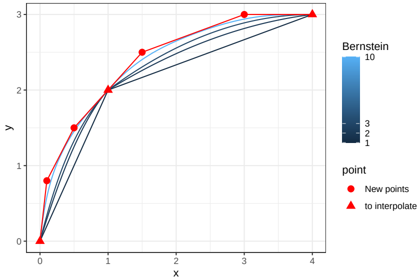


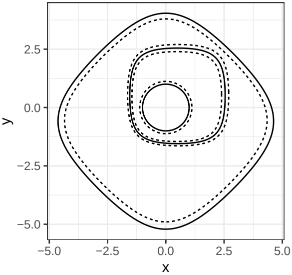
Conic interpolation:

$$T_\lambda = a(\lambda)\, T_0 + b(\lambda)\, T_1$$

a et b: Bernstein approximation of a piecewise affine, positive function.

Preserves monotonicity, concavity, control derivatives up to order k at endpoints





Conic interpolation:

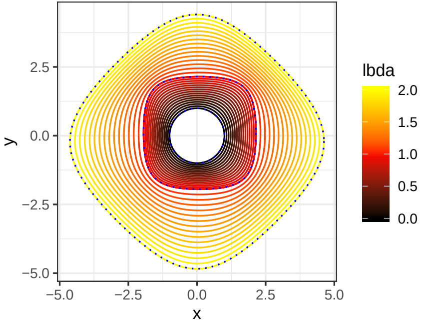
$$T_\lambda = a(\lambda)\,T_0 + b(\lambda)\,T_1$$

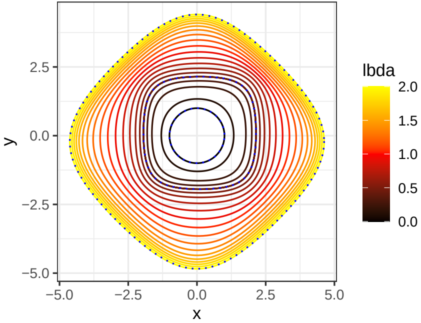
a et b: Bernstein approximation of a piecewise affine, positive function.

Explicit derivatives, C^k junction around boundaries:

- "normal" direction: C^1 control, variations of order $2, \dots, k$ vanish.
- "tangent" direction: C^k control.

Illustration





C^k convex interpolation

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f.

C^k convex interpolation

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f. $\nabla^2 f$ positive definite outside of: $\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i$.

C^k convex interpolation

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f. $\nabla^2 f$ positive definite outside of: $\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i$.

Original goal: construct f with oscillating gradients.

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f. $\nabla^2 f$ positive definite outside of: $\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i$.

Original goal: construct f with oscillating gradients.

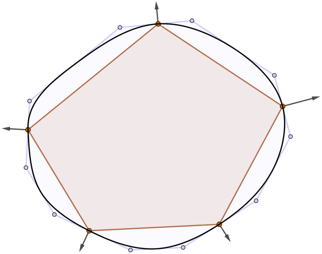
Construct C^2_+ sets in $\mathbb{R}^2$ with controled normals?

C^k convex interpolation

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f. $\nabla^2 f$ positive definite outside of: $\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i$.

Original goal: construct f with oscillating gradients.

Construct C^2_+ sets in $\mathbb{R}^2$ with controled normals? C^2_+ and C^k outer approximation of polygons with prescribed normals

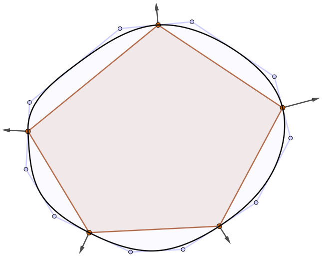


C^k convex interpolation

Theorem:
Let $(T_i)_{i\in\mathbb{Z}}$ be a sequence of C^2_+ and C^k subsets in $\mathbb{R}^2$, such that $T_i \subset \operatorname{int} T_{i+1} \neq \emptyset$ for all $i \in \mathbb{Z}$. Then there is C^k convex function f, such that each T_i is a sublevel set of f. $\nabla^2 f$ positive definite outside of: $\operatorname{argmin} f = \bigcap_{i\in\mathbb{Z}} T_i$.

Original goal: construct f with oscillating gradients.

Construct C^2_+ sets in $\mathbb{R}^2$ with controled normals? C^2_+ and C^k outer approximation of polygons with prescribed normals



"Practical" counterexamples: polygonal skeleton sequence + normals at vertices.
$\rightarrow$ C^k convex interpolation: gradient proportional to normals at vertices.

Plan

1. Overview of the convex interpolation problem

2. Positive curvature and smooth convex interpolation

3. Construction of algorithmic counter examples

4. Conclusion

Exact line search: f convex C^k.

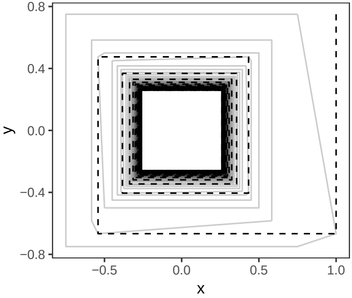
$$x_{k+1} \in \arg\min_x f(x), \qquad s.t. \quad x \in x_k - \mathbb{R}_+ \nabla f(x_k).$$

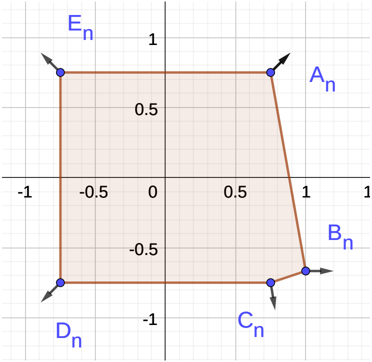
Optimality condition: $\langle x_{k+1} - x_k, \nabla f(x_{k+1}) \rangle = 0$.

Exact line search: f convex C^k.

$$x_{k+1} \in \arg\min_x f(x), \qquad s.t. \quad x \in x_k - \mathbb{R}_+ \nabla f(x_k).$$

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

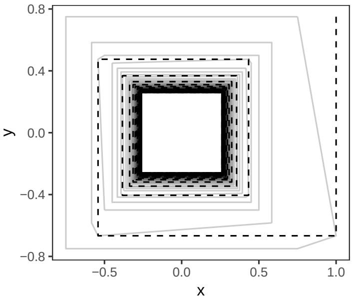


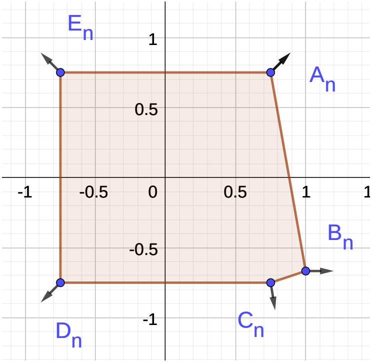


Exact line search: f convex C^k.

$$x_{k+1} \in \arg\min_x f(x), \qquad s.t. \quad x \in x_k - \mathbb{R}_+ \nabla f(x_k).$$

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





Alternating minimization / Gauss-Seidel: same counterexample.

$$b_{k+1} \in \arg\min_b f(a_k, b), \qquad\qquad a_{k+1} \in \arg\min_a f(a, b_{k+1})$$

Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Mirror descent / Bregman gradient (Nemirovski, Yudin)

Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\mathrm{int}\ \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Mirror descent / Bregman gradient (Nemirovski, Yudin)

Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\operatorname{int} \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Rockafellar: $\nabla h \colon \operatorname{int} \mathcal{C} \to \mathbb{R}^2$, homeomorphism, $(\nabla h)^{-1} = \nabla h^*$, h^* Legendre.

Mirror descent / Bregman gradient (Nemirovski, Yudin)

Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\operatorname{int} \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Rockafellar: $\nabla h \colon \operatorname{int} \mathcal{C} \to \mathbb{R}^2$, homeomorphism, $(\nabla h)^{-1} = \nabla h^*$, h^* Legendre.

Mirror descent: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$

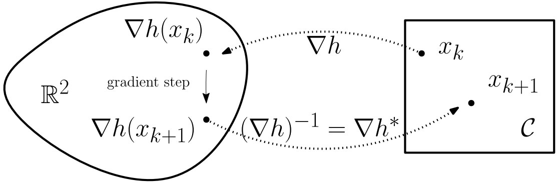
Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\operatorname{int} \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Rockafellar: $\nabla h \colon \operatorname{int} \mathcal{C} \to \mathbb{R}^2$, homeomorphism, $(\nabla h)^{-1} = \nabla h^*$, h^* Legendre.

Mirror descent: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$



Mirror descent / Bregman gradient (Nemirovski, Yudin)

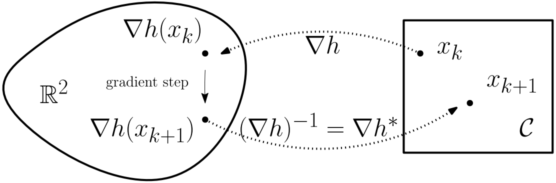
Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\mathrm{int}\, \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Rockafellar: $\nabla h \colon \mathrm{int}\, \mathcal{C} \to \mathbb{R}^2$, homeomorphism, $(\nabla h)^{-1} = \nabla h^*$, h^* Legendre.

Mirror descent: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$



- $O(1/k)$ optimality gap (Bauschke et. al. 17)
- sequential convergence?

Mirror descent / Bregman gradient (Nemirovski, Yudin)

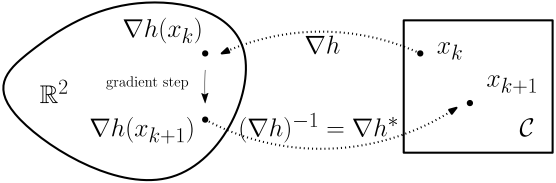
Linear program: $\mathcal{C}$ convex compact, $c \in \mathbb{R}^2$, $\max_{x \in \mathcal{C}} \langle x, c \rangle$

Legendre function h**:** on $\mathrm{int}\, \mathcal{C}$,

- essentially smooth (differentiable, ∇h explodes on the boundary)
- strictly convex $h(y) - h(x) - \langle \nabla h(x), y - x \rangle > 0$ for $x \neq y$.

Rockafellar: $\nabla h \colon \mathrm{int}\, \mathcal{C} \to \mathbb{R}^2$, homeomorphism, $(\nabla h)^{-1} = \nabla h^*$, h^* Legendre.

Mirror descent: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$



- $O(1/k)$ optimality gap (Bauschke et. al. 17)
- sequential convergence? (yes, under additional assumptions, no in general).

Absence of convergence for Bregman gradient algorithm

Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$

$$\nabla h(x_k) = \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc$$

Absence of convergence for Bregman gradient algorithm

Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\begin{aligned} \nabla h(x_k) &= \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc \\ x_k &= \nabla h^*(kc). \end{aligned}$$

Absence of convergence for Bregman gradient algorithm

Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\begin{aligned}\nabla h(x_k) &= \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc \\ x_k &= \nabla h^*(kc).\end{aligned}$$

Skeleton for h^*: h Legendre bounded domain,

Absence of convergence for Bregman gradient algorithm

Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\begin{aligned} \nabla h(x_k) &= \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc \\ x_k &= \nabla h^*(kc). \end{aligned}$$

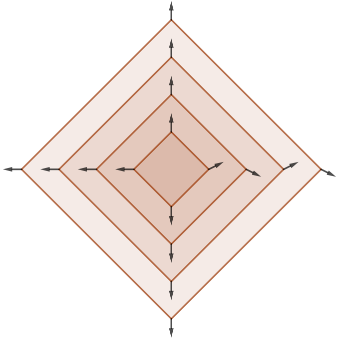
Skeleton for h^*: h Legendre bounded domain,
h^*, C^k, strictly convex, *Lipschitz*.

Absence of convergence for Bregman gradient algorithm

Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\nabla h(x_k) = \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc$$
$$x_k = \nabla h^*(kc).$$

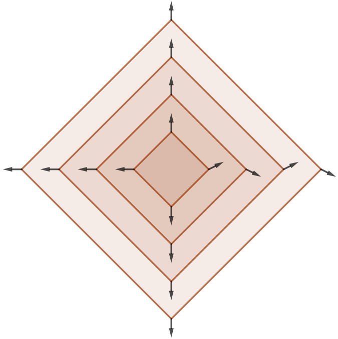
Skeleton for h^*: h Legendre bounded domain, h^*, C^k, strictly convex, *Lipschitz*.
polygons, oscilating normals.



Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\begin{aligned} \nabla h(x_k) &= \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc \\ x_k &= \nabla h^*(kc). \end{aligned}$$

Skeleton for h^*: h Legendre bounded domain,
h^*, C^k, strictly convex, *Lipschitz*.
polygons, oscilating normals.
$\rightarrow$ h^*, $\nabla^2 h^*$ P.D outside of argmin (singleton)



Absence of convergence for Bregman gradient algorithm

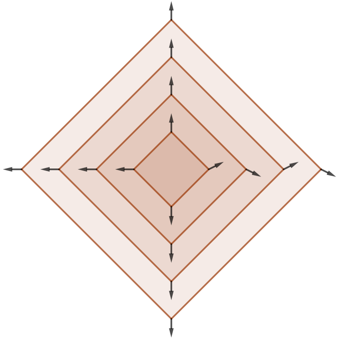
Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

$$\nabla h(x_k) = \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc$$
$$x_k = \nabla h^*(kc).$$

Skeleton for h^*: h Legendre bounded domain,
h^*, C^k, strictly convex, *Lipschitz*.
homothetic polygons, oscilating normals.
$\rightarrow h^*$, $\nabla^2 h^*$ P.D outside of argmin (singleton)

Additional properties:

- h^* Lipschitz.
- h Legendre compact domain (unit square).



Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

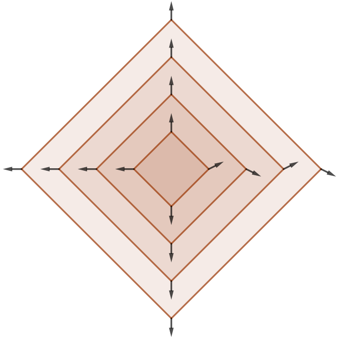
$$\nabla h(x_k) = \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc$$
$$x_k = \nabla h^*(kc).$$

Skeleton for h^*: h Legendre bounded domain,
h^*, C^k, strictly convex, *Lipschitz*.
homothetic polygons, oscilating normals.
$\rightarrow h^*$, $\nabla^2 h^*$ P.D outside of argmin (singleton)

Additional properties:

- h^* Lipschitz.
- h Legendre compact domain (unit square).

$\rightarrow (x_k)_{k \in \mathbb{N}}$ does not converge.



Algorithm: $x_{k+1} = \nabla h^*(\nabla h(x_k) + c)$, set $x_0 = \arg\min_x h(x)$

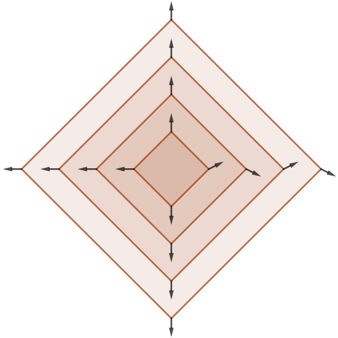
$$\nabla h(x_k) = \nabla h(x_{k-1}) + c = \nabla h(x_0) + kc = kc$$
$$x_k = \nabla h^*(kc).$$

Skeleton for h^*: h Legendre bounded domain,
h^*, C^k, strictly convex, *Lipschitz*.
homothetic polygons, oscilating normals.
$\rightarrow h^*$, $\nabla^2 h^*$ P.D outside of argmin (singleton)

Additional properties:

- h^* Lipschitz.
- h Legendre compact domain (unit square).

$\rightarrow (x_k)_{k\in\mathbb{N}}$ does not converge.



Remark: Possibly related phenomenon,
$\rightarrow$ no acceleration *à la* Nesterov (Dragomir et. al. 20)

Conditional gradient / Frank-Wolfe algorithm (with C. Combettes)

Frank-Wolfe algorithm: $\min_{x \in \mathcal{C}} f(x)$, $\mathcal{C}$ compact convex, f convex, Lipschitz gradient.

Conditional gradient / Frank-Wolfe algorithm (with C. Combettes)

Frank-Wolfe algorithm: $\min_{x\in\mathcal{C}} f(x)$, $\mathcal{C}$ compact convex, f convex, Lipschitz gradient.

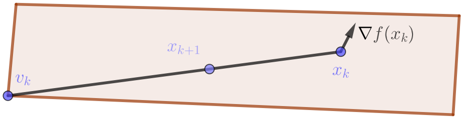
$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k v_k, \qquad v_k \in \arg\min_{v\in\mathcal{C}} \langle \nabla f(x_k), v\rangle$$

$$\gamma_k = \arg\min_{z=(1-\gamma)x_k+\gamma v_k} f(x_k) + \langle \nabla f(x_k), z - x_k\rangle + \frac{L}{2}\|x_k - z\|^2$$

Frank-Wolfe algorithm: $\min_{x\in\mathcal{C}} f(x)$, $\mathcal{C}$ compact convex, f convex, Lipschitz gradient.

$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k v_k, \qquad v_k \in \arg\min_{v\in\mathcal{C}} \langle \nabla f(x_k), v\rangle$$

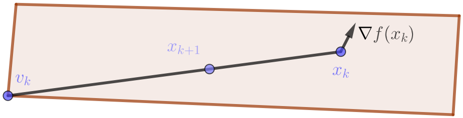
$$\gamma_k = \arg\min_{z=(1-\gamma)x_k+\gamma v_k} f(x_k) + \langle \nabla f(x_k), z - x_k\rangle + \frac{L}{2}\|x_k - z\|^2$$



Frank-Wolfe algorithm: $\min_{x \in \mathcal{C}} f(x)$, $\mathcal{C}$ compact convex, f convex, Lipschitz gradient.

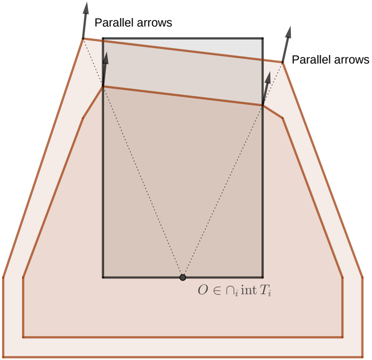
$$x_{k+1} = (1 - \gamma_k) x_k + \gamma_k v_k, \qquad v_k \in \arg\min_{v \in \mathcal{C}} \langle \nabla f(x_k), v \rangle$$

$$\gamma_k = \arg\min_{z = (1-\gamma)x_k + \gamma v_k} f(x_k) + \langle \nabla f(x_k), z - x_k \rangle + \frac{L}{2} \|x_k - z\|^2$$

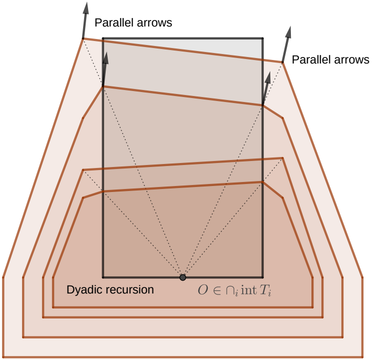


- Fixed points = minimizers.
- $O(1/k)$ optimality gap.
- Many step size variants.
- Sequential convergence?

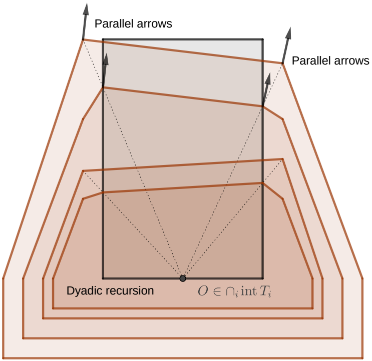
Controling gradients beyond polygonal skeleton vertices



Controling gradients beyond polygonal skeleton vertices

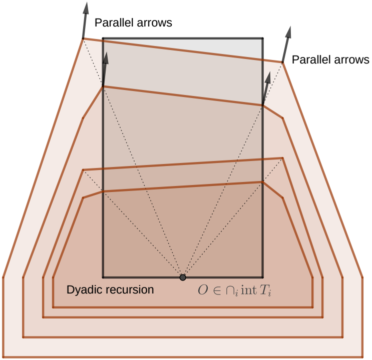


Controling gradients beyond polygonal skeleton vertices

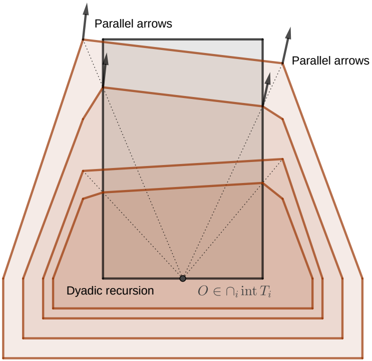


- **Conic interpolation:** alignment $\rightarrow$ control gradient directions on segments.

Controling gradients beyond polygonal skeleton vertices



- **Conic interpolation:** alignment $\to$ control gradient directions on segments.
- **Linear oracle:** controled on a significative portion of the trajectory.



- **Conic interpolation:** alignment $\rightarrow$ control gradient directions on segments.
- **Linear oracle:** controled on a significative portion of the trajectory.
- **No convergence:** if $\gamma_k \rightarrow 0$, $\gamma_k < 1$ for all k, not summable (*e.g.* L slight over estimation).

Plan

Conclusion

Smooth convex interpolation in $\mathbb{R}^2$:

- Sublevel sets with positive curvature.
- Specification using polygonal skeleton.

Not detailed in the presentation:

- Additional counter-examples.
- Subtleties and computational aspects of the construction.

Next steps:

- More counter examples.
- C^∞ interpolation.
- Relax positive curvature.

Conclusion

Smooth convex interpolation in $\mathbb{R}^2$:

- Sublevel sets with positive curvature.
- Specification using polygonal skeleton.

Not detailed in the presentation:

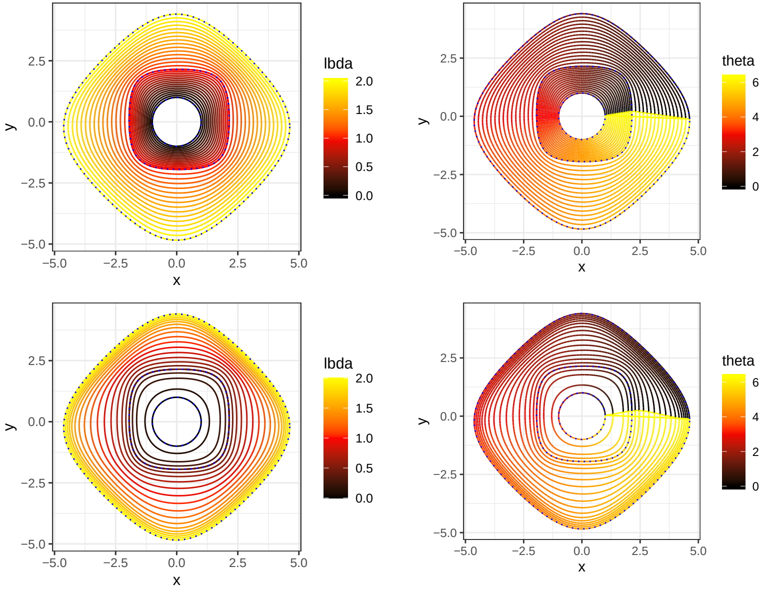
- Additional counter-examples.
- Subtleties and computational aspects of the construction.

Next steps:

- More counter examples.
- C^∞ interpolation.
- Relax positive curvature.

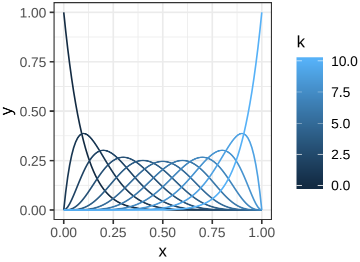
Thanks

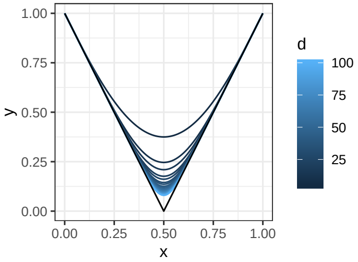
Illustration



Bernstein approximation

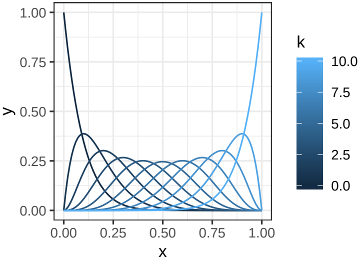
Bernstein polynomial and approximation of absolute value:

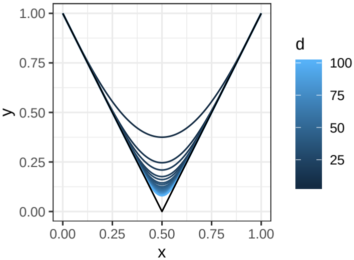




Bernstein approximation

Bernstein polynomial and approximation of absolute value:



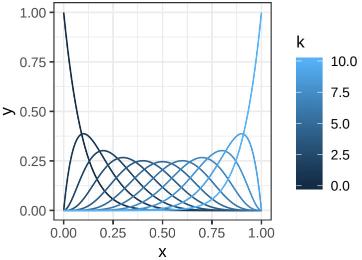


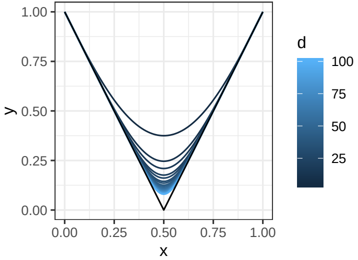
$f : [0,1] \mapsto \mathbb{R}$, $d \in \mathbb{N}^*$, $m \in \mathbb{N}^*$,

$$B_{d,f} : x \mapsto \sum_{k=0}^{d} f\left(\frac{k}{d}\right) \binom{d}{k} x^k (1-x)^{d-k}$$

Bernstein approximation

Bernstein polynomial and approximation of absolute value:





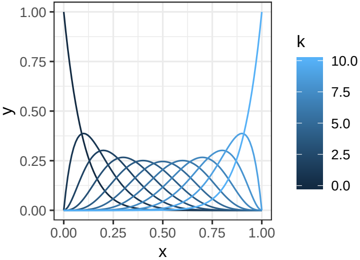
$f : [0,1] \mapsto \mathbb{R}$, $d \in \mathbb{N}^*$, $m \in \mathbb{N}^*$, $\Delta^1 f(x) = f(x + 1/d) - f(x)$, derivative.

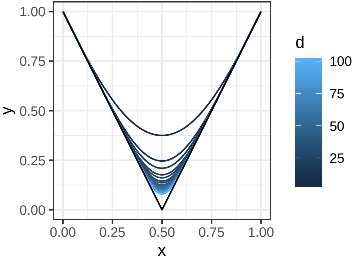
$$B_{d,f} : x \mapsto \sum_{k=0}^{d} f\left(\frac{k}{d}\right) \binom{d}{k} x^k (1-x)^{d-k}$$

$$B'_{d,f} : x \mapsto d \sum_{k=0}^{d-1} \Delta^1 f\left(\frac{k}{d}\right) \binom{d-1}{k} x^k (1-x)^{d-k-1}$$

Bernstein approximation

Bernstein polynomial and approximation of absolute value:





$f : [0,1] \mapsto \mathbb{R}$, $d \in \mathbb{N}^*$, $m \in \mathbb{N}^*$, $\Delta^1 f(x) = f(x + 1/d) - f(x)$, derivative.

$$B_{d,f} : x \mapsto \sum_{k=0}^{d} f\left(\frac{k}{d}\right)\binom{d}{k} x^k (1-x)^{d-k}$$

$$B'_{d,f} : x \mapsto d \sum_{k=0}^{d-1} \Delta^1 f\left(\frac{k}{d}\right)\binom{d-1}{k} x^k (1-x)^{d-k-1}$$

Linear in f, preserves monotonicity, convexity, control derivatives at endpoints.