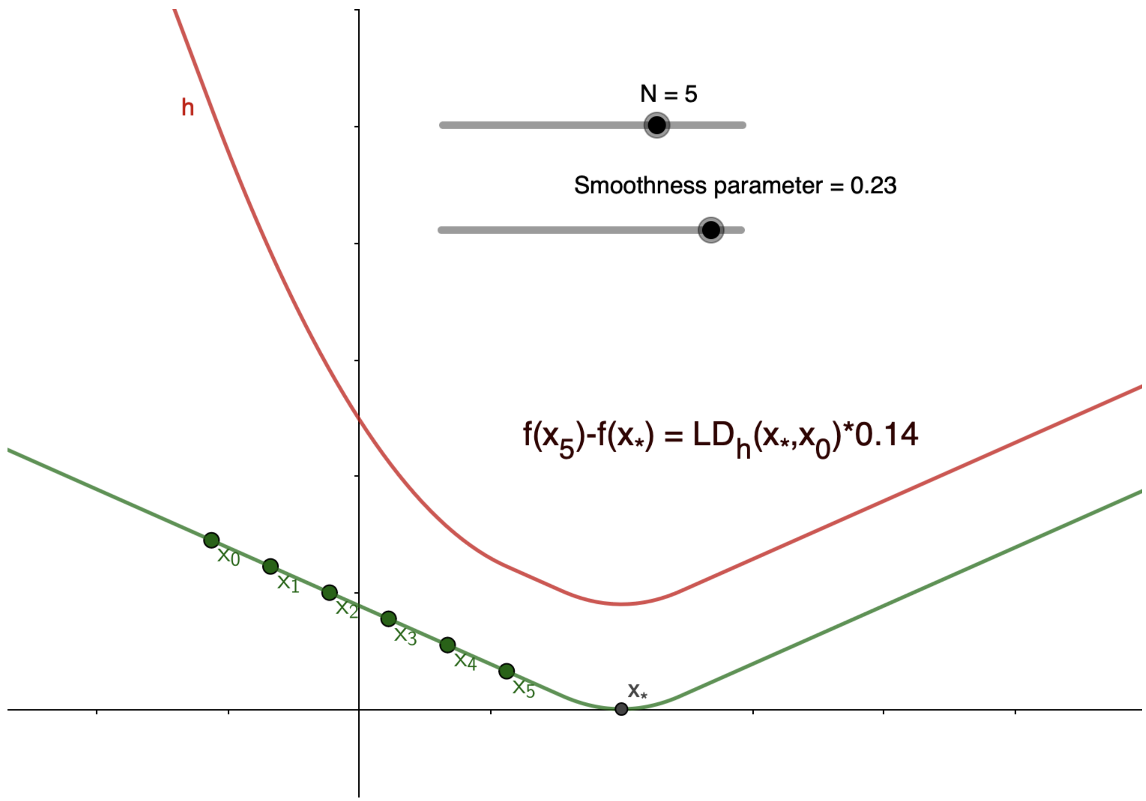
Computer-aided analyses of Bregman methods



Radu-Alexandru Dragomir,

Post-doctoral researcher, EPFL

joint work with Adrien Taylor, Alexandre d'Aspremont, Jérôme Bolte

PEP talks, Louvain-la-Neuve, 13/02/2023

Relatively-smooth convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f convex, differentiable.

Relatively-smooth convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f convex, differentiable.

Choose a **reference function** $h : \mathbb{R}^d \to \mathbb{R}$ convex and differentiable. Define

Relatively-smooth convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f convex, differentiable.

Choose a **reference function** $h : \mathbb{R}^d \to \mathbb{R}$ convex and differentiable. Define

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle \geq 0$$

the **Bregman divergence** of h.

Relatively-smooth convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f convex, differentiable.

Choose a **reference function** $h : \mathbb{R}^d \to \mathbb{R}$ convex and differentiable. Define

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle \geq 0$$

the **Bregman divergence** of h.

We say that f is **L-smooth relative to h** if

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + L D_h(x, y)$$

(Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)

Relatively-smooth convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

with f convex, differentiable.

Choose a **reference function** $h : \mathbb{R}^d \to \mathbb{R}$ convex and differentiable. Define

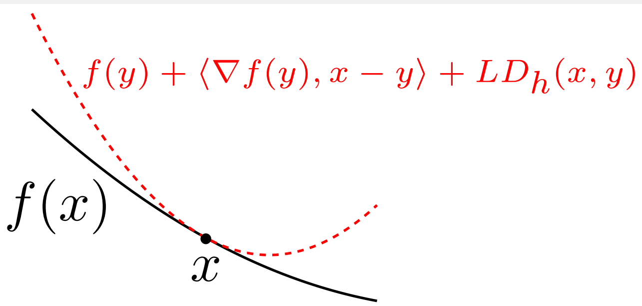
$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle \geq 0$$

the **Bregman divergence** of h.

We say that f is **L-smooth relative to h** if

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + L D_h(x, y)$$

(Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)



Equivalent to $Lh - f$ **convex**, or

$$\nabla^2 f(x) \preceq L \nabla^2 h(x)$$

Bregman gradient descent

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k) \qquad \text{(BGD)}$$

Also called *mirror descent* or *NoLips*.

Bregman gradient descent

$$x_{k+1} \in \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k) \tag{BGD}$$

Also called *mirror descent* or *NoLips*.

Convergence rate **(Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)**

If f is **convex**, **L-smooth relative to** h, and $\lambda \in (0, \frac{1}{L}]$, then

$$f(x_N) - f(x_*) \leq \frac{D_h(x_*, x_0)}{\lambda N},$$

for x_* that minimizes f.

Bregman gradient descent

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k) \qquad \text{(BGD)}$$

Also called *mirror descent* or *NoLips*.

Convergence rate (Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)

If f is **convex**, **L-smooth relative to h**, and $\lambda \in (0, \frac{1}{L}]$, then

$$f(x_N) - f(x_*) \leq \frac{D_h(x_*, x_0)}{\lambda N},$$

for x_* that minimizes f.

Is this **tight**? Can we design an **accelerated** version?

Bregman gradient descent

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k) \qquad \text{(BGD)}$$

Also called *mirror descent* or *NoLips*.

Convergence rate (Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)

If f is **convex**, **L-smooth relative to** h, and $\lambda \in (0, \frac{1}{L}]$, then

$$f(x_N) - f(x_*) \leq \frac{D_h(x_*, x_0)}{\lambda N},$$

for x_* that minimizes f.

Is this **tight**? Can we design an **accelerated** version?

... can we **PEP it?**

Bregman gradient descent

$$x_{k+1} \in \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k) \qquad \text{(BGD)}$$

Also called *mirror descent* or *NoLips*.

Convergence rate (Bauschke, Bolte, Teboulle, 2017) (Lu, Freund, Nesterov 2018)

If f is **convex**, **L-smooth relative to h**, and $\lambda \in (0, \frac{1}{L}]$, then

$$f(x_N) - f(x_*) \leq \frac{D_h(x_*, x_0)}{\lambda N},$$

for x_* that minimizes f.

Is this **tight**? Can we design an **accelerated** version?

... can we **PEP it?**

Yes, if we look for worst case over **both f and h!**

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k),$$

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k),$$

- f is convex and differentiable,
- h is convex and differentiable,

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k),$$

- f is convex and differentiable,
- h is convex and differentiable,
- f is L-smooth relative to h: $Lh - f$ is convex,

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$x_{k+1} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{\lambda} D_h(u, x_k),$$

- f is convex and differentiable,
- h is convex and differentiable,
- f is L-smooth relative to h: $Lh - f$ is convex,
- $D_h(x_*, x_0) \leq 1$.

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}_{i \in I}$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$\nabla h(x_{k+1}) = \nabla h(x_k) - \lambda \nabla f(x_k),$$

- f is convex and differentiable,
- h is convex and differentiable,
- $d = Lh - f$ is convex,
- $D_h(x_*, x_0) \leq 1$.

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}, u$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$\nabla h(x_{k+1}) = \nabla h(x_k) - \lambda \nabla f(x_k),$$

- f is convex and differentiable,
- $h = (f + d)/L$ with d convex and differentiable,
- $D_h(x_*, x_0) \leq 1$.

Pepping Bregman

$$\text{maximize } f(x_N) - f(x_*)$$

in $f, h, \{x_i\}, u$, subject to

- $x_1, \ldots x_N$ are generated by BGD: for $k = 0 \ldots N-1$,

$$\nabla h(x_{k+1}) = \nabla h(x_k) - \lambda \nabla f(x_k),$$

- f is convex and differentiable,
- $h = (f + d)/L$ with d convex and differentiable,
- $h(x_*) - h(x_0) - \langle \nabla h(x_0), x_* - x_0 \rangle \leq 1.$

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f_i'\}_{i\in I}, \{h_i, h_i'\}_{i\in I}, \{x_i\}_{i\in I}, u,$ subject to

- for $k = 0 \dots N-1$,
$$h'_{k+1} = h'_k - \lambda f'_k,$$
- the set $\{x_i, f_i, f_i'\}_{i\in I}$ is **interpolable** by a differentiable convex function,
- the set $\{x_i, Lh_i - f_i, Lh_i' - f_i'\}_{i\in I}$ is **interpolable** by a differentiable convex function,
- $h_* - h_0 - \langle h_0', x_* - x_0\rangle \leq 1.$

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f_i'\}_{i\in I}, \{h_i, h_i'\}_{i\in I}, \{x_i\}_{i\in I}, u$, subject to

- for $k = 0 \dots N-1$,
$$h'_{k+1} = h'_k - \lambda f'_k,$$
- the set $\{x_i, f_i, f_i'\}_{i\in I}$ is **interpolable** by a differentiable convex function:
$$\begin{cases} f_i - f_j - \langle f_j', x_i - x_j\rangle \geq 0, & \\ f_i - f_j - \langle f_j', x_i - x_j\rangle > 0 & \text{if } f_i' \neq f_j'. \end{cases}$$
- the set $\{x_i, Lh_i - f_i, Lh_i' - f_i'\}_{i\in I}$ is **interpolable** by a differentiable convex function,
- $h_u - h_0 - \langle h_0', u - x_0\rangle \leq 1$.

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f'_i\}_{i\in I}, \{h_i, h'_i\}_{i\in I}, \{x_i\}_{i\in I}, u$, subject to

- for $k = 0 \ldots N-1$,
$$h'_{k+1} = h'_k - \lambda f'_k,$$
- the set $\{x_i, f_i, f'_i\}_{i\in I}$ is **interpolable** by a differentiable convex function:
$$\begin{cases} f_i - f_j - \langle f'_j, x_i - x_j\rangle \geq 0, & \\ f_i - f_j - \langle f'_j, x_i - x_j\rangle > 0 & \text{if } f'_i \neq f'_j. \end{cases}$$
- the set $\{x_i, Lh_i - f_i, Lh'_i - f'_i\}_{i\in I}$ is **interpolable** by a differentiable convex function:
$$[\ldots]$$
- $h_u - h_0 - \langle h'_0, u - x_0\rangle \leq 1.$

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f_i'\}_{i\in I}, \{h_i, h_i'\}_{i\in I}, \{x_i\}_{i\in I}, u$, subject to

- for $k = 0 \ldots N-1$,
$$h_{k+1}' = h_k' - \lambda f_k',$$
- the set $\{x_i, f_i, f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$f_i - f_j - \langle f_j', x_i - x_j\rangle \geq 0.$$
- the set $\{x_i, Lh_i - f_i, Lh_i' - f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$[\ldots]$$
- $h_u - h_0 - \langle h_0', u - x_0\rangle \leq 1.$

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f_i'\}_{i\in I}, \{h_i, h_i'\}_{i\in I}, \{x_i\}_{i\in I}, u$, subject to

- for $k = 0 \dots N-1$,
$$h_{k+1}' = h_k' - \lambda f_k',$$
- the set $\{x_i, f_i, f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$f_i - f_j - \langle f_j', x_i - x_j\rangle \geq 0.$$
- the set $\{x_i, Lh_i - f_i, Lh_i' - f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$[\dots]$$
- $h_u - h_0 - \langle h_0', u - x_0\rangle \leq 1.$

Problem is **linear** in f_i, h_i, and in the dot products

$$\langle h_i', x_j\rangle, \langle f_i', x_j\rangle, \quad i, j \in I,$$

Pepping Bregman

$$\text{maximize } f_N - f_u$$

in $\{f_i, f_i'\}_{i\in I}, \{h_i, h_i'\}_{i\in I}, \{x_i\}_{i\in I}, u$, subject to

- for $k = 0 \ldots N-1$,
$$h_{k+1}' = h_k' - \lambda f_k',$$
- the set $\{x_i, f_i, f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$f_i - f_j - \langle f_j', x_i - x_j\rangle \geq 0.$$
- the set $\{x_i, Lh_i - f_i, Lh_i' - f_i'\}_{i\in I}$ belongs to the closure of sets interpolable by differentiable convex functions:
$$[\ldots]$$
- $h_u - h_0 - \langle h_0', u - x_0\rangle \leq 1.$

Problem is **linear** in f_i, h_i, and in the dot products

$$\langle h_i', x_j\rangle, \langle f_i', x_j\rangle, \quad i, j \in I,$$

$\rightarrow$ can be solved for large d by ~~**SDP**~~ **LP** relaxation.

Results for BGD

- The numerical value of (PEP) is **exactly** L/N, i.e., the bound

$$f(x_N) - f(x_*) \leq \frac{L D_h(x_*, x_0)}{N}$$

is **tight** in the worst case for BGD.

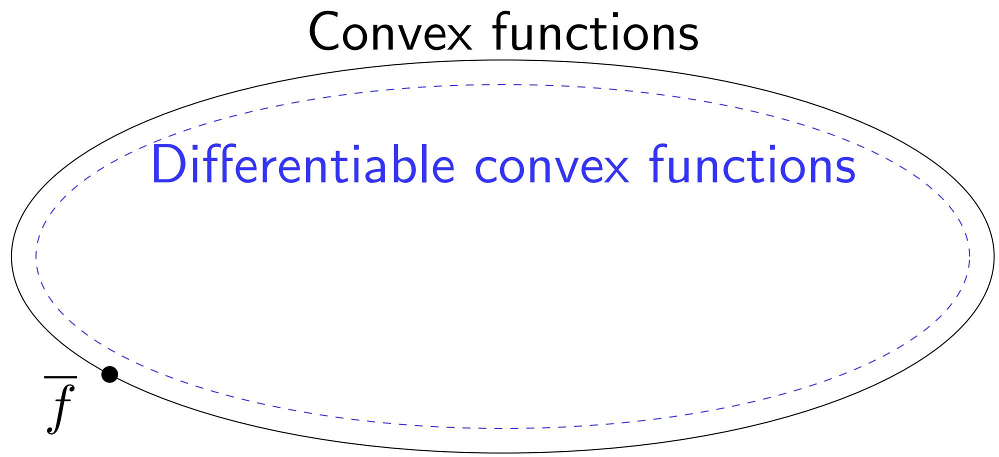
Results for BGD

- The numerical value of (PEP) is **exactly** L/N, i.e., the bound

$$f(x_N) - f(x_*) \leq \frac{LD_h(x_*, x_0)}{N}$$

  is **tight** in the worst case for BGD.

- **Limiting nonsmooth worst-case:** the original feasible set is not closed; the supremum is reached as (f, h) approach some nonsmooth functions $(\overline{f}, \overline{h})$.



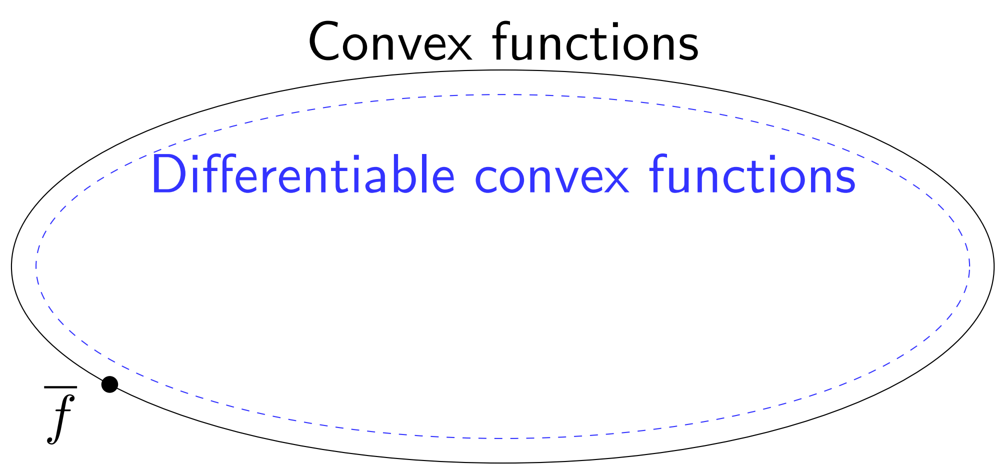
Results for BGD

- The numerical value of (PEP) is **exactly** L/N, i.e., the bound

$$f(x_N) - f(x_*) \le \frac{L D_h(x_*, x_0)}{N}$$

  is **tight** in the worst case for BGD.

- **Limiting nonsmooth worst-case:** the original feasible set is not closed; the supremum is reached as (f, h) approach some nonsmooth functions $(\bar{f}, \bar{h})$.



Worst-case **sequence** of functions:
`https://www.geogebra.org/classic/re5c2phw`

Lower bound for general Bregman methods

Can we find a lower bound that holds for all Bregman-type methods?

Lower bound for general Bregman methods

Can we find a lower bound that holds for all Bregman-type methods?

Bregman first-order method, informal definition

An algorithm $\mathcal{A}$ is a Bregman first-order method if it uses only the oracles ∇f, ∇h, ∇h^*, where

$$\nabla h^*(y) = \operatorname*{argmin}_{u \in \mathbb{R}^d} h(u) - \langle y, u \rangle,$$

and linear combinations.

Lower bound for general Bregman methods

Can we find a lower bound that holds for all Bregman-type methods?

Bregman first-order method, informal definition

An algorithm $\mathcal{A}$ is a Bregman first-order method if it uses only the oracles ∇f, ∇h, ∇h^*, where

$$\nabla h^*(y) = \operatorname*{argmin}_{u \in \mathbb{R}^d} h(u) - \langle y, u \rangle,$$

and linear combinations.

We look for a worst-case function that *hides information in high dimension* by adding to the PEP the constraints

$$\langle \nabla f(x_i), \nabla f(x_j) \rangle = 0, \quad \forall i \neq j.$$

Similar approach to "worst function in the world" (Nesterov, 2003).

Lower bound for general Bregman methods

Can we find a lower bound that holds for all Bregman-type methods?

Bregman first-order method, informal definition

An algorithm $\mathcal{A}$ is a Bregman first-order method if it uses only the oracles ∇f, ∇h, ∇h^*, where

$$\nabla h^*(y) = \operatorname*{argmin}_{u \in \mathbb{R}^d} h(u) - \langle y, u \rangle,$$

and linear combinations.

We look for a worst-case function that *hides information in high dimension* by adding to the PEP the constraints

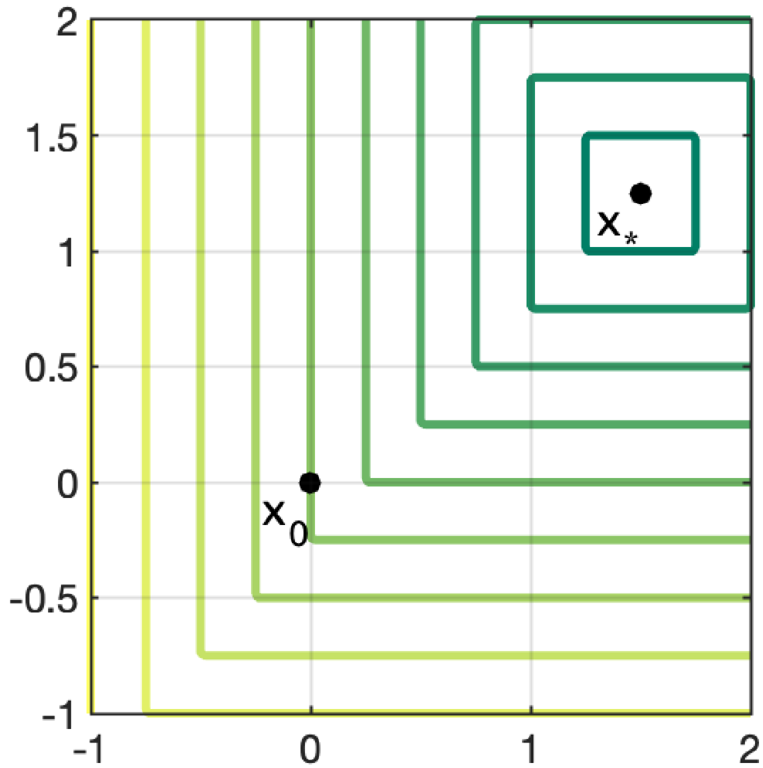
$$\langle \nabla f(x_i), \nabla f(x_j) \rangle = 0, \quad \forall i \neq j.$$

Similar approach to "worst function in the world" (Nesterov, 2003).

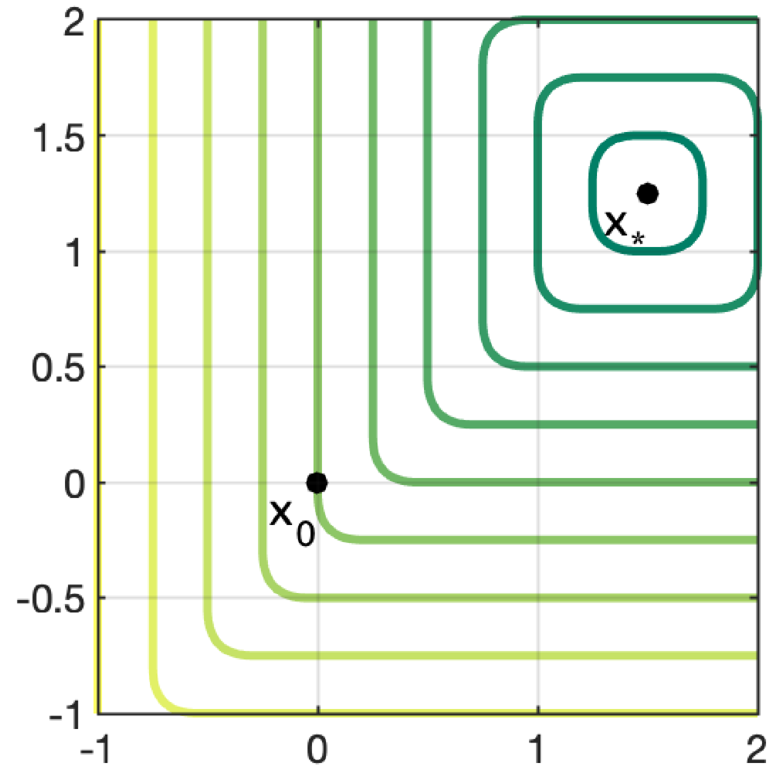
No acceleration theorem (D., Taylor, d'Aspremont, Bolte, 2021)

The rate $1/N$ is **optimal** for Bregman first-order methods on relatively-smooth convex problems for **general reference functions** h.

High-dimensional worst case function



(a) Limiting function $\overline{f}$

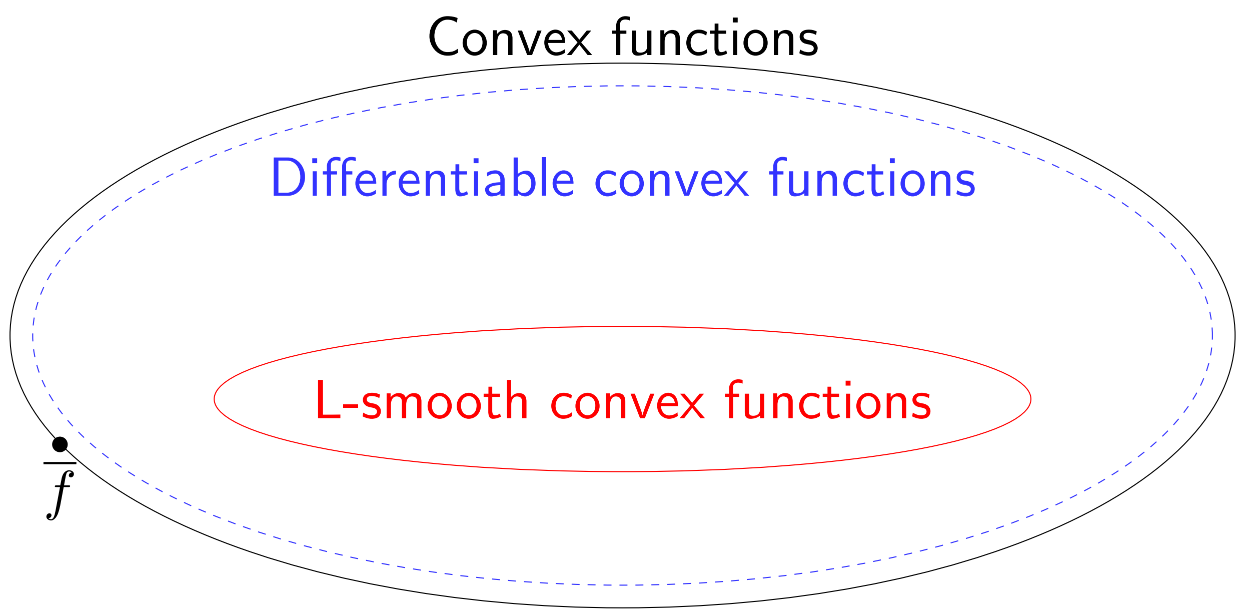


(b) Feasible f approaching $\overline{f}$

Conclusion

Besides helping to prove the result, PEPs allowed us to understand

- the **structure of the class of functions** characterized by a set of inequalities,
- the many possible ways **things can go wrong in the worst case...**



R-A. Dragomir, A. B. Taylor, A. d'Aspremont, J. Bolte. *Optimal Complexity and Certification of Bregman First-Order Methods.* Mathematical Programming, 2021.