Implementable Tensor Methods in Unconstrained Convex Optimization

Yurii Nesterov, HSE (Moscow)

and CORE (UCL, Belgium)

April 14, 2018

Optimization@Work (MPTI, Moscow)

Taylor Approximation

Taylor Approximation

Let $x \in \operatorname{int}(\operatorname{dom} f)$.

Taylor Approximation

Let $x \in \operatorname{int}(\operatorname{dom} f)$. Then

$$f(x+h) \quad = \quad \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \operatorname{dom} f,$$

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \quad = \quad \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$,

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \quad = \quad \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \dots, h_p]$$

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \quad = \quad \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \ldots, h_p]$$

is the directional derivative of f at x along directions $h_i \in \mathbb{E}$, $i = 1, \ldots, p$.

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \;\; = \;\; \Phi_{x,p}(h) + o(\|h\|^p), \quad x + h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \dots, h_p]$$

is the directional derivative of f at x along directions $h_i \in \mathbb{E}$, $i = 1, \dots, p$.

Note:

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \;=\; \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \dots, h_p]$$

is the directional derivative of f at x along directions $h_i \in \mathbb{E}$, $i = 1, \dots, p$.

Note:

1. $D^p f(x)[\cdot]$ is a *symmetric p-linear form*.

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \quad = \quad \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \dots, h_p]$$

is the directional derivative of f at x along directions $h_i \in \mathbb{E}$, $i = 1, \dots, p$.

Note:

1. $D^p f(x)[\cdot]$ is a *symmetric p-linear form*.
2. If $h_1 = \cdots = h_p$,

Taylor Approximation

Let $x \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then

$$f(x+h) \;\; = \;\; \Phi_{x,p}(h) + o(\|h\|^p), \quad x+h \in \mathrm{dom}\, f,$$

where $\Phi_{x,p}(y) \stackrel{\mathrm{def}}{=} f(x) + \sum\limits_{i=1}^{p} \frac{1}{i!} D^i f(x)[y-x]^i$, $y \in \mathbb{E}$, and

$$D^p f(x)[h_1, \dots, h_p]$$

is the directional derivative of f at x along directions $h_i \in \mathbb{E}$, $i = 1, \dots, p$.

Note:

1. $D^p f(x)[\cdot]$ is a *symmetric p-linear form*.
2. If $h_1 = \dots = h_p$, we use notation $D^p f(x)[h]^p$.

Measuring the quality of approximations

Measuring the quality of approximations

Let us fix a norm $\|\cdot\|$ in $\mathbb{E}$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \| x - y \|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \| y - x \|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \| \nabla f(y) - \nabla \Phi_{x,p}(y) \|_* \leq \frac{L_p}{p!} \| y - x \|^p.$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \| x - y \|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}.$

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \le 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \le L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \le \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \le \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \le \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What?

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$,

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex*

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate*

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate polynomial*.

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\|D^p f(x)\| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate polynomial*.

2. Up to now, Algebraic Geometry cannot provide us with efficient tools for computing its stationary points

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \| x - y \|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}.$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate polynomial*.

2. Up to now, Algebraic Geometry cannot provide us with efficient tools for computing its stationary points (not speaking about global minimum).

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \; \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \| x - y \|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \| y - x \|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \| \nabla f(y) - \nabla \Phi_{x,p}(y) \|_* \leq \frac{L_p}{p!} \| y - x \|^p.$

C. *Hessian:* $\quad \| \nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y) \| \leq \frac{L_p}{(p-1)!} \| y - x \|^{p-1}. \qquad$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate polynomial*.

2. Up to now, Algebraic Geometry cannot provide us with efficient tools for computing its stationary points (not speaking about global minimum).

Consequence:

Measuring the quality of approximations

Let us fix a norm $\| \cdot \|$ in $\mathbb{E}$ and define the norm

$$\| D^p f(x) \| = \max_h \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}.$$

Then we can introduce Lipschitz constants:

$$\| D^p f(x) - D^p f(y) \| \leq L_p \|x - y\|, \quad x, y \in \mathrm{dom}\, f.$$

These constants ensure the high-quality of local approximations:

A. *Function*: $\quad |f(y) - \Phi_{x,p}(y)| \leq \frac{L_p}{(p+1)!} \|y - x\|^{p+1}, \quad x, y \in \mathrm{dom}\, f.$

B. *Gradient:* $\quad \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{L_p}{p!} \|y - x\|^p.$

C. *Hessian:* $\quad \|\nabla^2 f(y) - \nabla^2 \Phi_{x,p}(y)\| \leq \frac{L_p}{(p-1)!} \|y - x\|^{p-1}.$ ETC.

And What? **Note that:**

1. For $p \geq 3$, $\Phi_{x,p}(y)$ is a *non-convex multivariate polynomial*.

2. Up to now, Algebraic Geometry cannot provide us with efficient tools for computing its stationary points (not speaking about global minimum).

Consequence: Practical Optimization goes up to the 2nd-order methods.

Let us look ...

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1} g \rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function*

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \; = \; \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2} Bx,$$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \; = \; \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B$$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1} g \rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p} \|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2} Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2} B + (p-2) \|x\|^{p-4} Bxx^* B \succeq \|x\|^{p-2} B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!} \|y - x\|^{p+1}$.

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \; = \; \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y-x\|^{p+1}$.

NB:

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1} g \rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2} Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2} B + (p-2)\|x\|^{p-4} Bxx^* B \succeq \|x\|^{p-2} B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y - x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $\quad f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y-x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $\quad f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \ t \geq f(x)\}$ is a *convex set*.

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y - x\|^{p+1}.$

NB: 1. If $M \geq L_p$, then $\quad f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \ t \geq f(x)\}$ is a *convex set*.

Question:

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y-x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $\quad f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \; t \geq f(x)\}$ is a *convex set*.

Question: Is it easy to put a *nonconvex object*

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y-x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $\quad f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \; t \geq f(x)\}$ is a *convex set*.

Question: Is it easy to put a *nonconvex object* into the convex one?

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* \ = \ \langle g, B^{-1}g\rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $\quad d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2}Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2}B + (p-2)\|x\|^{p-4}Bxx^*B \succeq \|x\|^{p-2}B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y-x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $\quad f(y) \stackrel{(A)}{\leq} \Omega_{x,p,M}(y) \quad$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \ t \geq f(x)\}$ is a *convex set*.

Question: Is it easy to put a *nonconvex object* into the convex one?

Answer:

Let us look ...

Let us fix $B = B^* \succ 0 : \mathbb{E} \to \mathbb{E}^*$ and define the norms

$$\|x\| = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|g\|_* = \langle g, B^{-1} g \rangle^{1/2}, \quad g \in E^*.$$

Let us introduce the *power function* $d_p(x) = \frac{1}{p}\|x\|^p, \quad p \geq 2.$

Note that

$$\nabla d_p(x) = \|x\|^{p-2} Bx,$$

$$\nabla^2 d_p(x) = \|x\|^{p-2} B + (p-2)\|x\|^{p-4} Bxx^* B \succeq \|x\|^{p-2} B.$$

Define $\Omega_{x,p,M}(y) = \Phi_{x,p}(y) + \frac{M}{(p+1)!}\|y - x\|^{p+1}$.

NB: 1. If $M \geq L_p$, then $f(y) \overset{(A)}{\leq} \Omega_{x,p,M}(y)$ for all $y \in \mathbb{E}$.

2. The epigraph $\{(x,t) : \ t \geq f(x)\}$ is a *convex set*.

Question: Is it easy to put a *nonconvex object* into the convex one?

Answer: **NO!**

Main Theorem

Main Theorem

Let $M \geq pL_p$.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y)$$

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \frac{L_p}{(p-1)!}\|y - x\|^{p-1} B$$

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{L_p}{(p-1)!}\|y - x\|^{p-1} B$$

$$\preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{M}{p!}\|y - x\|^{p-1} B$$

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{L_p}{(p-1)!}\|y-x\|^{p-1}B$$

$$\preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{M}{p!}\|y-x\|^{p-1}B \ \preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{L_p}{(p-1)!}\|y-x\|^{p-1}B$$
$$\preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{M}{p!}\|y-x\|^{p-1}B \ \preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Consequences:

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{L_p}{(p-1)!}\|y-x\|^{p-1}B$$
$$\preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{M}{p!}\|y-x\|^{p-1}B \ \preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Consequences:

1. For $M > pL_p$ the point $T_{p,M}(x) = \arg\min\limits_{y\in\mathbb{E}} \Omega_{x,p,M}(y)$ is well defined.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{L_p}{(p-1)!} \|y - x\|^{p-1} B$$

$$\preceq \nabla^2 \Phi_{x,p}(y) + \tfrac{M}{p!} \|y - x\|^{p-1} B \ \preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Consequences:

1. For $M > pL_p$ the point $T_{p,M}(x) = \arg\min\limits_{y \in \mathbb{E}} \Omega_{x,p,M}(y)$ is well defined.
2. It can be computed by the techniques of *Convex Optimization*.

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \frac{L_p}{(p-1)!}\|y-x\|^{p-1}B$$

$$\preceq \nabla^2 \Phi_{x,p}(y) + \frac{M}{p!}\|y-x\|^{p-1}B \ \preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Consequences:

1. For $M > pL_p$ the point $T_{p,M}(x) = \arg\min\limits_{y \in \mathbb{E}} \Omega_{x,p,M}(y)$ is well defined.
2. It can be computed by the techniques of *Convex Optimization*.
3. It can be used for solving the problem $\quad f_* = \min\limits_{x \in \mathbb{E}} f(x)$

Main Theorem

Let $M \geq pL_p$. Then function $\Omega_{x,p,M}(\cdot)$ is convex.

Proof.

1. $\nabla^2 \Phi_{x,p,M}(\cdot)$ is a Taylor approximation of $\nabla^2 f(\cdot)$.
2. Therefore, for any $y \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(y) \preceq \nabla^2 \Phi_{x,p}(y) + \frac{L_p}{(p-1)!}\|y-x\|^{p-1}B$$

$$\preceq \nabla^2 \Phi_{x,p}(y) + \frac{M}{p!}\|y-x\|^{p-1}B \;\preceq \nabla^2 \Phi_{x,p,M}(y).$$

□

Consequences:

1. For $M > pL_p$ the point $T_{p,M}(x) = \arg\min\limits_{y \in \mathbb{E}} \Omega_{x,p,M}(y)$ is well defined.
2. It can be computed by the techniques of *Convex Optimization*.
3. It can be used for solving the problem $\quad f_* = \min\limits_{x \in \mathbb{E}} f(x)$

in the case $L_p(f) < +\infty$.

Local Method

Local Method

For $M \geq pL_p$, consider the process

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \;=\; T_{p,M}(x_t), \;\; t \geq 0.}$$

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \; = \; T_{p,M}(x_t), \;\; t \geq 0.}$$

Theorem 2.

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \;=\; T_{p,M}(x_t), \;\; t \geq 0.}$$

Theorem 2. For all $t \geq 0$ we have $f(x_{t+1}) \leq f(x_t)$.

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \ = \ T_{p,M}(x_t), \ \ t \geq 0.}$$

Theorem 2. For all $t \geq 0$ we have $f(x_{t+1}) \leq f(x_t)$.

At the same time,

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \;=\; T_{p,M}(x_t), \;\; t \geq 0.}$$

Theorem 2. For all $t \geq 0$ we have $f(x_{t+1}) \leq f(x_t)$.

At the same time,

$$f(x_t) - f_* \leq \frac{(M+L_p)D^{p+1}}{p!} \left(\frac{p+1}{t}\right)^p, \quad t \geq 1,$$

Local Method

For $M \geq pL_p$, consider the process

$$\boxed{x_{t+1} \;=\; T_{p,M}(x_t), \;\; t \geq 0.}$$

Theorem 2. For all $t \geq 0$ we have $f(x_{t+1}) \leq f(x_t)$.

At the same time,

$$f(x_t) - f_* \leq \frac{(M+L_p)D^{p+1}}{p!} \left(\frac{p+1}{t}\right)^p, \quad t \geq 1,$$

where

$$D = \max_{x \in E} \{\|x - x^*\| : \; f(x) \leq f(x_0)\}.$$

Accelerated Tensor Method

Accelerated Tensor Method

NB:

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$:

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \frac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

For these objects, we are going to maintain the following relations:

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

For these objects, we are going to maintain the following relations:

$$\mathcal{R}_k^1 : \; A_k f(x_k) \leq \psi_k^* \; \equiv \; \min_{x \in \mathbb{E}} \psi_k(x),$$

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

For these objects, we are going to maintain the following relations:

$$\mathcal{R}_k^1: \; A_k f(x_k) \leq \psi_k^* \; \equiv \; \min_{x \in \mathbb{E}} \psi_k(x),$$

$$\mathcal{R}_k^2: \; \psi_k(x) \leq A_k f(x) + \tfrac{M + L_p + C}{p!} d_{p+1}(x - x_0), \; \forall x \in \mathbb{E}, \; k \geq 1.$$

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \overset{\text{def}}{=} A_k + a_k$, $k \geq 1$.

For these objects, we are going to maintain the following relations:

$$\mathcal{R}_k^1 : \; A_k f(x_k) \leq \psi_k^* \; \equiv \; \min_{x \in \mathbb{E}} \psi_k(x),$$

$$\mathcal{R}_k^2 : \; \psi_k(x) \leq A_k f(x) + \tfrac{M + L_p + C}{p!} d_{p+1}(x - x_0), \; \forall x \in \mathbb{E}, \; k \geq 1.$$

(In nonconvex setting,

Accelerated Tensor Method

NB: We apply the standard technique of *estimating sequences*

(see the theory of 1st and 2nd-order methods).

We choose $M \geq pL_p$ and recursively update the following sequences.

1. Sequence of estimating functions

$$\psi_k(x) = \ell_k(x) + \tfrac{C}{p!} d_{p+1}(x - x_0), \quad k \geq 1,$$

where $\ell_k(\cdot)$ are linear functions in $x \in \mathbb{E}$, and $C > 0$.

2. Minimizing sequence $\{x_k\}_{k=1}^{\infty}$.

3. Sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

For these objects, we are going to maintain the following relations:

$$\mathcal{R}_k^1: \; A_k f(x_k) \leq \psi_k^* \; \equiv \; \min_{x \in \mathbb{E}} \psi_k(x),$$

$$\mathcal{R}_k^2: \; \psi_k(x) \leq A_k f(x) + \tfrac{M + L_p + C}{p!} d_{p+1}(x - x_0), \; \forall x \in \mathbb{E}, \; k \geq 1.$$

(In nonconvex setting, a similar approach was analyzed in an unpublished report by M. Baes (ETHZ, 2009).)

Define $A_k = \left[\frac{(p-1)(M^2-p^2L_p^2)}{4(p+1)M^2}\right]^{\frac{p}{2}} \left(\frac{k}{p+1}\right)^{p+1}$, Define $A_k = \left[\frac{(p-1)(M^2-p^2L_p^2)}{4(p+1)M^2}\right]^{\frac{p}{2}} \left(\frac{k}{p+1}\right)^{p+1}$, and $a_k = A_{k+1} - A_k$, $k \geq 0$.

Define $A_k = \left[\frac{(p-1)(M^2-p^2L_p^2)}{4(p+1)M^2}\right]^{\frac{p}{2}} \left(\frac{k}{p+1}\right)^{p+1}$, and $a_k \;=\; A_{k+1} - A_k,\; k \geq 0.$

Initialization: Choose $x_0 \in \mathbb{E}$ and $M > pL_p$. Compute $x_1 = T_{p,M}(x_0)$.

Define $C = \frac{p}{2}\sqrt{\frac{(p+1)}{(p-1)}(M^2 - p^2 L_p^2)}$ and $\psi_1(x) = f(x_1) + \frac{C}{p!} d_{p+1}(x - x_0)$.

Iteration k, ($k \geq 1$):

1. Compute $v_k = \arg\min_{x \in \mathbb{E}} \psi_k(x)$ and choose $y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_k}{A_{k+1}} v_k$.

2. Compute $x_{k+1} = T_{p,M}(y_k)$ and update

$$\psi_{k+1}(x) = \psi_k(x) + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle].$$

Define $A_k = \left[\frac{(p-1)(M^2-p^2L_p^2)}{4(p+1)M^2}\right]^{\frac{p}{2}} \left(\frac{k}{p+1}\right)^{p+1}$, and $a_k = A_{k+1} - A_k$, $k \geq 0$.

Initialization: Choose $x_0 \in \mathbb{E}$ and $M > pL_p$. Compute $x_1 = T_{p,M}(x_0)$.

Define $C = \frac{p}{2}\sqrt{\frac{(p+1)}{(p-1)}(M^2 - p^2 L_p^2)}$ and $\psi_1(x) = f(x_1) + \frac{C}{p!} d_{p+1}(x - x_0)$.

Iteration k, ($k \geq 1$):

1. Compute $v_k = \arg\min_{x \in \mathbb{E}} \psi_k(x)$ and choose $y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_k}{A_{k+1}} v_k$.

2. Compute $x_{k+1} = T_{p,M}(y_k)$ and update

$$\psi_{k+1}(x) = \psi_k(x) + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle].$$

Convergence:

Define $A_k = \left[\frac{(p-1)(M^2-p^2L_p^2)}{4(p+1)M^2}\right]^{\frac{p}{2}} \left(\frac{k}{p+1}\right)^{p+1}$, and $a_k = A_{k+1} - A_k,\ k \geq 0$.

Initialization: Choose $x_0 \in \mathbb{E}$ and $M > pL_p$. Compute $x_1 = T_{p,M}(x_0)$.

Define $C = \frac{p}{2}\sqrt{\frac{(p+1)}{(p-1)}(M^2 - p^2 L_p^2)}$ and $\psi_1(x) = f(x_1) + \frac{C}{p!} d_{p+1}(x - x_0)$.

Iteration k, ($k \geq 1$):

1. Compute $v_k = \arg\min\limits_{x \in \mathbb{E}} \psi_k(x)$ and choose $y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_k}{A_{k+1}} v_k$.

2. Compute $x_{k+1} = T_{p,M}(y_k)$ and update

$$\psi_{k+1}(x) = \psi_k(x) + a_k[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle].$$

Convergence:

$$f(x_k) - f(x^*) \leq \frac{M+L_p+C}{(p+1)!} \left[\frac{4(p+1)M^2}{(p-1)(M^2-p^2L_p^2)}\right]^{\frac{p}{2}} \left(\frac{p+1}{k}\right)^{p+1} \|x_0 - x^*\|^{p+1}.$$

Lower Complexity Bounds

Lower Complexity Bounds

For $p \geq 2$,

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

$$\text{Let } U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k},$$

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function:

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Lower Complexity Bounds

For $p \geq 2$, define $\eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3.

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$

Lower Complexity Bounds

For $p \geq 2$, define $\eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n$.

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)!\ \kappa(t)}, \ t \geq 1.$$

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n$.

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)!\ \kappa(t)},\ t \geq 1.$$

Then for all $t :\ 2t + 1 \leq n$ we have

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

$$\text{Let } U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}, \text{ and } A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}.$$

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)! \; \kappa(t)}, \; t \geq 1.$$

Then for all $t: \; 2t + 1 \leq n$ we have $\kappa(t) \leq \frac{1}{3p} 2^{p+1} (2t+2)^{\frac{3p+1}{2}}$.

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n$.

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)! \; \kappa(t)}, \; t \geq 1.$$

Then for all $t: \; 2t + 1 \leq n$ we have $\kappa(t) \leq \frac{1}{3p} 2^{p+1} (2t+2)^{\frac{3p+1}{2}}$.

NB:

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)! \; \kappa(t)}, \; t \geq 1.$$

Then for all $t: \; 2t + 1 \leq n$ we have $\kappa(t) \leq \frac{1}{3p} 2^{p+1} (2t+2)^{\frac{3p+1}{2}}$.

NB: 1. Our proof is simpler,

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)!\ \kappa(t)},\ t \geq 1.$$

Then for all $t:\ 2t + 1 \leq n$ we have $\kappa(t) \leq \frac{1}{3p} 2^{p+1} (2t+2)^{\frac{3p+1}{2}}$.

NB: 1. Our proof is simpler, but we need some natural assumptions on the rules of $\mathcal{M}$.

Lower Complexity Bounds

For $p \geq 2$, define $\quad \eta_{p+1}(x) = \frac{1}{p+1} \sum\limits_{i=1}^{n} |x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$

Let $U_k = \begin{pmatrix} 1 & -1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$, and $A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$.

Difficult function: $\quad f_k(x) = \eta_{p+1}(A_k x) - x^{(1)}$, $2 \leq k \leq p$.

Theorem 3. Let for any function $f \in \mathcal{F}_p$ with $L_p((f) < +\infty$ method $\mathcal{M}$ ensures the rate of convergence

$$\min_{0 \leq k \leq t} f(x_k) - f_* \leq \frac{L_p \|x_0 - x^*\|^{p+1}}{(p+1)!\ \kappa(t)},\ t \geq 1.$$

Then for all $t:\ 2t+1 \leq n$ we have $\kappa(t) \leq \frac{1}{3p} 2^{p+1} (2t+2)^{\frac{3p+1}{2}}$.

NB: 1. Our proof is simpler, but we need some natural assumptions on the rules of $\mathcal{M}$.

2. The most general construction was justified in a sequence of papers (see Y.Arjevani, O.Shamir, R.Shiff in arXiv (2017) for the last one).

Degree of Non-Optimality

Degree of Non-Optimality

Accelerated method:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^{0}, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}$

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p=1 \Rightarrow \left(\frac{1}{\epsilon}\right)^{0}$, $\quad p=2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}$, $\quad p=3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}$, $\quad 2^{20} \approx 10^{6}$.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}, \quad 2^{20} \approx 10^6.$

Conclusion:

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}, \quad 2^{20} \approx 10^6$.

Conclusion: "Optimal methods" with expensive line search

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}, \quad 2^{20} \approx 10^6$.

Conclusion: "Optimal methods" with expensive line search do not work in practice.

Degree of Non-Optimality

Accelerated method:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{p+1}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p+1}}\right)$.

Lower bound:

- Rate of convergence: $O\left(\left(\frac{1}{t}\right)^{\frac{3p+1}{2}}\right)$.
- Complexity bound: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{2}{3p+1}}\right)$.

Extra factor: $O\left(\left(\frac{1}{\epsilon}\right)^{\frac{p-1}{(p+1)(3p+1)}}\right)$.

NB: $p = 1 \Rightarrow \left(\frac{1}{\epsilon}\right)^0, \quad p = 2 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{21}}, \quad p = 3 \Rightarrow \left(\frac{1}{\epsilon}\right)^{\frac{1}{20}}, \quad 2^{20} \approx 10^6$.

Conclusion: "Optimal methods" with expensive line search do not work in practice. (E.g. Monteiro, Svaiter, *SIOPT* (2013), for $p = 2$.)

Third-order methods: implementation details

Third-order methods: implementation details

Taylor polynomial:

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem:

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24} \| h \|^4 \;\to\; \min\limits_{h \in \mathbb{E}}$.

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24}\|h\|^4 \;\to\; \min\limits_{h \in \mathbb{E}}$.

Main Theorem:

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24} \| h \|^4 \; \to \; \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\rm def}{=} \Phi_x(h) + \frac{M}{24} \| h \|^4 \ \rightarrow \ \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \| h \|^2 B.$$

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h\rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24}\|h\|^4 \ \rightarrow \ \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \|h\|^2 B.$$

Conclusion:

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24} \|h\|^4 \;\to\; \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \|h\|^2 B.$$

Conclusion: For any $h \in \mathbb{E}$,

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24} \|h\|^4 \;\to\; \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \|h\|^2 B.$$

Conclusion: For any $h \in \mathbb{E}$, the Hessian $\nabla^2 \Phi_x(h)$ is *similar* to the Hessian of the function

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24}\|h\|^4 \;\to\; \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \|h\|^2 B.$$

Conclusion: For any $h \in \mathbb{E}$, the Hessian $\nabla^2 \Phi_x(h)$ is *similar* to the Hessian of the function

$$\rho_x(h) = \frac{1}{2}(1 - \frac{1}{\tau})\langle \nabla^2 f(x)h, h \rangle + \frac{M - \tau L_3}{10}\|h\|^4$$

Third-order methods: implementation details

Taylor polynomial: $\Phi_x(h) = \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \frac{1}{6} D^3 f(x)[h]^3$.

Auxiliary Problem: $\Omega_{x,M}(h) \stackrel{\mathrm{def}}{=} \Phi_x(h) + \frac{M}{24} \|h\|^4 \ \rightarrow \ \min\limits_{h \in \mathbb{E}}$.

Main Theorem: for all $h \in \mathbb{E}$ we have

$$0 \preceq \nabla^2 f(x) + D^3 f(x)[h] + L_3 \|h\|^2 B.$$

Conclusion: For any $h \in \mathbb{E}$, the Hessian $\nabla^2 \Phi_x(h)$ is *similar* to the Hessian of the function

$$\rho_x(h) = \frac{1}{2}(1 - \frac{1}{\tau}) \langle \nabla^2 f(x) h, h \rangle + \frac{M - \tau L_3}{10} \|h\|^4$$

with some $\tau > 1$.

Relative Smoothness Condition (Lu, Freund, N., 2018)

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition:

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu\nabla^2\rho(x) \preceq \nabla^2 f(x) \preceq L\nabla^2\rho(x).$$

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu\nabla^2\rho(x) \preceq \nabla^2 f(x) \preceq L\nabla^2\rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu\nabla^2\rho(x) \preceq \nabla^2 f(x) \preceq L\nabla^2\rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle\nabla\rho(x), y - x\rangle$.

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018):

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}.$

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB:

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB: 1. For 3rd-order method with $\rho = \rho_x$,

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB: 1. For 3rd-order method with $\rho = \rho_x$, we have $\mu = 1$, $L = \frac{\tau+1}{\tau-1}$.

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB: 1. For 3rd-order method with $\rho = \rho_x$, we have $\mu = 1$, $L = \frac{\tau+1}{\tau-1}$.

2. Solution of problem (*) is simple:

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x,y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB: 1. For 3rd-order method with $\rho = \rho_x$, we have $\mu = 1$, $L = \frac{\tau+1}{\tau-1}$.

2. Solution of problem (*) is simple:

$$\min_{h \in \mathbb{E}} \{ \langle g, h \rangle + \tfrac{1}{2} \langle Gh, h \rangle + \gamma \|h\|^4 \},$$

Relative Smoothness Condition (Lu, Freund, N., 2018)

Definition: Function $f(\cdot)$ satisfies the strong relative smoothness condition with respect to $\rho(\cdot)$ if

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 f(x) \preceq L \nabla^2 \rho(x).$$

(Upper bounding was suggested by Bauschke, Bolte, Teboulle (2017).)

Define the Bregman distance $\beta_\rho(x,y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle$.

Consider the method:

$$x_{k+1} = \arg\min_{y \in \mathbb{E}} \{ \langle \nabla f(x_k), y - x_k \rangle + L \beta_\rho(x_k, y) \}. \quad (*).$$

Theorem 4 (LFN 2018): $\quad f(x_k) - f^* \leq \frac{\mu \beta_\rho(x_0, x^*)}{\left(\frac{L}{L-\mu}\right)^k - 1}$.

NB: 1. For 3rd-order method with $\rho = \rho_x$, we have $\mu = 1$, $L = \frac{\tau+1}{\tau-1}$.

2. Solution of problem (*) is simple:

$$\min_{h \in \mathbb{E}} \{ \langle g, h \rangle + \tfrac{1}{2} \langle Gh, h \rangle + \gamma \|h\|^4 \},$$

especially after an appropriate factorization of matrix $G \succeq 0$.

Conclusion

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple:

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization:

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x \rangle)$,

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x\rangle)$, functions with explicit structure (by fast backward differentiation),

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x\rangle)$, functions with explicit structure (by fast backward differentiation), etc.)

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$. (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x\rangle)$, functions with explicit structure (by fast backward differentiation), etc.)

3. Main open question:

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x\rangle)$, functions with explicit structure (by fast backward differentiation), etc.)

3. Main open question: The Optimal Methods.

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$. (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x \rangle)$, functions with explicit structure (by fast backward differentiation), etc.)

3. Main open question: The Optimal Methods.

4. What about 4th-order schemes?

Conclusion

1. We developed accelerated 3rd order schemes for minimizing smooth convex functions with the global rate of convergence $O(\frac{1}{k^4})$. This is the fastest sublinear rate known so far.

2. These methods are *implementable*. Complexity of each iteration is comparable with that of the 2nd-order methods:

- Linear convergence rate of auxiliary process depends only on absolute constant.
- Algorithmic complexity of one iteration is $O(n^2)$.
- The oracle is simple: we need to compute the vector $D^3 f(x)[h]^2$.

  (e.g. Separable Optimization: $\sum\limits_{i=1}^{N} f_i(\langle a_i, x \rangle)$, functions with explicit structure (by fast backward differentiation), etc.)

3. Main open question: The Optimal Methods.

4. What about 4th-order schemes?

THANK YOU FOR YOUR ATTENTION!