Lower Complexity Bounds for Minimizing Regularized Quadratic Functions

Nikita Doikov

Cornell University, ORIE

Joint work with Daniel Berg Thomsen (Inria and École Polytechnique, Paris)

UCLouvain

June 18, 2026

Optimization Problem

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

where f is a convex quadratic function with regularization, for $p > 2$:

$$f(x) \;=\; \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \frac{s}{p}\|x\|^p$$

- $A = A^\top \succeq 0$ is the main component, denote $L := \lambda_{\max}(A)$
- $b \in \mathbb{R}^d$
- Regularization parameter $s > 0$
- Euclidean norm: $\|x\| := \langle x, x\rangle^{1/2}$

Motivational Example: Regularized Newton's Method

A convex problem:

$$\min_{y \in \mathbb{R}^d} F(y)$$

- Taylor's approximation: $F(y_t + h) \approx F(y_t) + \langle \nabla F(y_t), h \rangle + \frac{1}{2} \langle \nabla^2 F(y_t) h, h \rangle$
- Step of the Newton method with regularization, for some $M > 0$:

$$\begin{aligned} y_{t+1} \quad &:= \quad y_t + \arg\min_{h \in \mathbb{R}^n} \Big[\langle \nabla F(y_t), h \rangle + \frac{1}{2} \langle \nabla^2 F(y_t) h, h \rangle + \frac{M}{p} \|h\|^p \Big] \\ &\equiv \quad y_t + \arg\min_{h \in \mathbb{R}^n} \Big[\frac{1}{2} \langle Ah, h \rangle - \langle b, x \rangle + \frac{s}{p} \|h\|^p \Big] \end{aligned}$$

where $A := \nabla^2 F(y_t) \succeq 0$ and $b := -\nabla f(y_t)$. We set $s := M$

Interesting cases:

- $p = 3$: Cubic Regularization of Newton's Method [Nesterov-Polyak, 2006]
- $2 \leq p \leq 3$: Second-order methods for Hölder Hessian [Grapiglia-Nesterov, 2017]
- $p = 4$: Quartic Regularization [Nesterov, 2021]
- $\ldots$

Motivational Example: Regularized Regression

- $X \in \mathbb{R}^{n \times d}$ is a given data matrix
- $y \in \mathbb{R}^{n}$ is a set of observations
- Linear Regression Problem:

$$\min_{w \in \mathbb{R}^d} \|Xw - y\|^2 + \lambda \|w\|^p$$

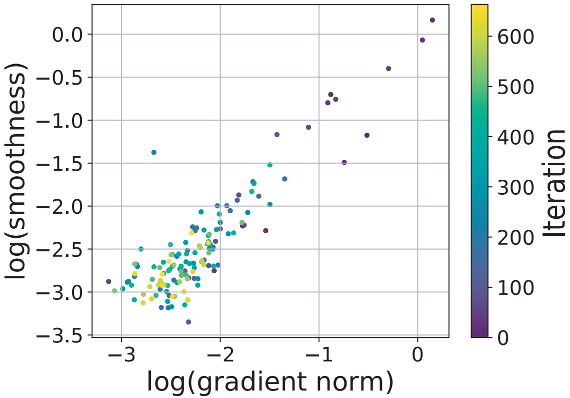
Fits our framework since

$$\frac{1}{2}\|Xw - y\|^2 \quad = \quad \frac{1}{2}\|Xw\|^2 + \frac{1}{2}\|y\|^2 - \langle Xw, y \rangle$$

- Set $A := X^\top X$
- Set $b := X^\top y$

Example: A Neural Network (LSTM)

- Classic assumptions: f is convex and smooth: $\forall x: \|\nabla^2 f(x)\| \leq L_1$ might not hold in practice



[Zhang-He-Sra-Jadbabaie, 2020]

Recent Trend in ML and Optimization: Non-Standard Smoothness

1. **Relative Smoothness:** [Bauschke-Bolte-Teboulle, 2016; Lu-Freund-Nesterov, 2018]

$$\nabla^2 f(x) \quad \leq \quad \beta \nabla^2 d(x)$$

2. (L_0, L_1)-**Functions:** [Zhang-He-Sra-Jadbabaie, 2020]

$$\|\nabla^2 f(x)\| \quad \leq \quad L_0 + L_1 \|\nabla f(x)\|$$

3. **Functions with bounded Global Curvature** [Nesterov, 2025]
4. ...

Our objective: $\boxed{f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \frac{s}{p}\|x\|^p}$

- The gradient:

$$\nabla f(x) \quad = \quad Ax - b + s\|x\|^{p-2}x$$

- The Hessian:

$$\nabla^2 f(x) \quad = \quad A + s\|x\|^{p-2} I + s(p-2)\|x\|^{p-4} xx^\top \quad \preceq \quad (L + s(p-1)\|x\|^{p-2}) I$$

- **The simplest non-trivial example of a function with unbounded Hessian**

How We Can Solve It?

$$f(x) \quad = \quad \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \frac{s}{p}\|x\|^p$$

- "Set gradient to zero": $\nabla f(x^\star) = 0$.
- Nonlinear equation:

$$\boxed{Ax^\star + s\|x^\star\|^{p-2}x^\star \quad = \quad b}$$

1. First, we need to find $r = \|x^\star\|$ by solving the univariate equation

$$r \quad = \quad \left\|\left(A + sr^{p-1}I\right)^{-1} b\right\|$$

2. Then, set

$$x^\star \quad = \quad \left(A + sr^{p-2}I\right)^{-1} b$$

- Large-scale alternative: **gradient methods**! E.g., $\boxed{x_{k+1} = x_k - \eta\nabla f(x_k)}$

Uniform Convexity

We say that f is uniformly convex if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\sigma}{p}\|y - x\|^p$$

- For our objective: $\sigma := \frac{1}{2^{p-2}} s$

Bounds on the functional residual:

- $f(x) - f^\star \geq \frac{\sigma}{p}\|x - x^\star\|^p$ ($\Rightarrow$ solution $x^\star$ is unique)
- $f(x) - f^\star \leq \frac{p-1}{p}\left(\frac{1}{p}\right)^{\frac{1}{p}}\|\nabla f(x)\|^{\frac{p}{p-1}}$

Gradient Methods

$$f(x) \;=\; \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \frac{s}{p}\|x\|^p \;\to\; \min_{x \in \mathbb{R}^d}$$

- Composite gradient method. Represent $f(x) = q(x) + \psi(x)$, where

$$q(x) \;=\; \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle, \qquad \psi(x) \;=\; \frac{s}{p}\|x\|^p$$

Iterate, for $k \geq 0$:

$$x_{k+1} \;=\; \arg\min_{x \in \mathbb{R}^d}\Big[\, \langle \nabla q(x_k), x - x_k\rangle + \frac{L}{2}\|x - x_k\|^2 + \psi(x) \,\Big]$$

$$= \;\frac{L}{L + s\|x_{k+1}\|^{p-2}}\, x_k \;-\; \frac{1}{L + s\|x_{k+1}\|^{p-2}} \nabla q(x_k)$$

- Basic gradient method:

$$x_{k+1} \;=\; x_k - \eta \nabla f(x_k), \qquad \eta \;:=\; \frac{1}{L + s(p-1)2^{p-2}\|x^\star\|^{p-2}}$$

The Rates for Gradient Methods

- Basic rate:

$$f(x_k) - f(x_{k+1}) \;\geq\; \tfrac{\eta}{2}\|\nabla f(x_k)\|^2 \;\geq\; c \cdot \left(f(x_k) - f^\star\right)^{\frac{2(p-1)}{p}}.$$

Denoting $F_k := f(x_k) - f^\star$, we get

$$F_k - F_{k+1} \;\geq\; c \cdot F_k^{\frac{2(p-1)}{p}},$$

This gives the rate:

$$F_k \;=\; O\left(\left[\tfrac{1}{k}\right]^{\frac{p}{p-2}}\right)$$

- Acceleration (Composite Fast Gradient Method):

$$F_k \;=\; O\left(\left[\tfrac{1}{k}\right]^{\frac{2p}{p-2}}\right)$$

[Roulet-d'Aspremont, 2017; Nesterov, 2022]

Definition of Optimization Algorithm

Complexity theory of optimization [Nemirovski-Yudin, 1983]

- First-order black-box oracle: $\mathcal{I}_f(x) := \{f(x), \nabla f(x)\}$
- Any first-order method:

$$x_{k+1} \quad := \quad M_k\big(\mathcal{I}_f(x_0),\, \mathcal{I}_f(x_1),\, \ldots,\, \mathcal{I}_f(x_k)\big), \qquad k \geq 0$$

Optimization algorithm M $\quad\Leftrightarrow\quad$ sequence of mappings $(M_0, M_1, M_2 \ldots)$

- *What is the best possible rate of first-order methods on this problem?*

Main Result: Lower Complexity Bound

- Let $p > 2$, $s > 2$, and $L > 0$ be fixed.

Theorem. For any first-order method running for $K \leq \frac{d-1}{2}$ iterations, there is a function f of the form

$$f(x) \quad = \quad \tfrac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \tfrac{s}{p}\|x\|^p$$

with $\lambda_{\max}(A) = L$ such that

$$f(x_K) - f^\star \quad \geq \quad \Omega\left(\left[\frac{L}{s^{2/p}K^2}\right]^{\frac{p}{p-2}}\right)$$

- FGM is optimal on this problem
- Proof technique is based on

"*Information-Based Complexity of Convex Programming.*" Arkadi S. Nemirovski. **(Lecture Notes)**, 1995

Proof Technique: Krylov Subspaces

Define:

- $A_\star := A + s\|x^\star\|^{p-2} I$
- Then $x^\star = A_\star^{-1} b$
- Krylov subspaces, for every $k \geq 1$:

$$\mathcal{E}_k(f) \;=\; \mathrm{span}\left\{b,\ A_\star b,\ \ldots\ ,\ A_\star^{k-1} b\right\} \;=\; \mathrm{span}\left\{b,\ Ab,\ \ldots\ ,\ A^{k-1} b\right\}$$

Facts:

1. $\mathcal{E}_0(f) \subseteq \mathcal{E}_1(f) \subseteq \ldots \subseteq \mathcal{E}_k(f)$
2. If $\mathcal{E}_k(f) = \mathcal{E}_{k+1}(f)$ then $\mathcal{E}_{k+1}(f) = \mathcal{E}_{k+2}(f) = \ldots$
3. Therefore:

$$\mathcal{E}_0(f) \;\subset\; \mathcal{E}_1(f) \;\subset\; \ldots \;\subset\; \mathcal{E}_\ell(f) \;=\; \mathcal{E}_{\ell+1}(f) \;=\; \ldots$$

4. Cayley-Hamilton theorem: $x^\star = A_\star^{-1} b = c_0 b + c_1 A_\star b + \ldots + c_{d-1} A_\star^{d-1} b$ for some $c_0, c_1, \ldots, c_d$
5. Hence: $\boxed{x^\star \in \mathcal{E}_\ell(f)}$

Problem Parametrization

Let

$$A = U \Lambda U^\top,$$

where $UU^\top = I$ is orthogonal b-invariant matrix: $Ub = b$

Fix:

- $r > 0$
- $\pi \in \Delta_d := \left\{ x \in \mathbb{R}^d_+ \; : \; \sum_{i=1}^{d} x^{(i)} = 1 \right\}$

Define:

$$\boxed{b := r\Big(\Lambda + s r^{p-2} I\Big)\sqrt{\pi} \in \mathbb{R}^d}$$

Lemma. $x^\star = rU\sqrt{\pi}$. Consequently, $\|x^\star\| = r$.

Corollary. For any $r > 0$ there exists a problem such that $\|x^\star\| = r$.

Main Lemma

Fixed parameters:

- $p \geq 2$ and $s > 0$
- $L \geq \lambda_d > \ldots > \lambda_1 \geq 0$
- $r > 0$ and $\pi \in \Delta_d$
- We have, $\ell = \#$nonzeros in vector π $\quad \Rightarrow \quad$ we can assume $\boxed{\ell = d}$

Lemma. For any method performing $K \leq \frac{\ell}{2}$ iterations, there exists an orthogonal b-invariant matrix U such that

$$x_k \in \mathcal{E}_{2k}(f), \qquad 0 \leq k \leq K.$$

Corollary.

$$x_k = c_0 b + c_1 A_\star b + c_2 A_\star^2 b + \ldots + c_{2k} A_\star^{2k} b = q(A_\star) b, \qquad q \in \mathcal{P}_{2k}$$

Bound for the Distance to the Solution

We obtain:

$$\begin{aligned}
\|x_K - x^\star\|^2 &\geq \min_{q\in\mathcal{P}_{2K}} \|q(A_\star)b - x_\star\|^2 \\
&= \min_{q\in\mathcal{P}_{2K}} \|\left[q(A_\star)A_\star - I\right]x^\star\|^2 \\
&= \min_{q\in\mathcal{P}_{2K}} \|\left[q(A_\star)A_\star - I\right]Ur\sqrt{\pi}\|^2 \\
&= r^2 \cdot \min_{q\in\mathcal{P}_{2K}} \sum_{i=1}^{d} \pi_i \left[1 - \lambda_i^\star q(\lambda_i^\star)\right]^2
\end{aligned}$$

where $\lambda_i^\star = \lambda_i + sr^{p-2} \in [\mu_\star, L_\star]$,

$$\begin{aligned}
\mu_\star &= sr^{p-2}, \\
L_\star &= L + sr^{p-2}.
\end{aligned}$$

Chebyshev Polynomials Bound

- The worse instance: **there exist λ_i and π_i such that**

$$\|x_K - x^\star\| \geq \min_{q\in\mathcal{P}_{2K}} \max_{\mu_\star\leq t\leq L_\star} |1 - tq(t)| = \min_{\substack{u\in\mathcal{P}_{2K+1} \\ \text{s.t.}\, u(0)=1}} \max_{\mu_\star\leq t\leq L_\star} |u(t)| = \Theta(Q_\star)$$

- The solution is achieved for

$$u^\star_{2K+1}(t) = \Theta(Q_\star) \cdot T_{2K+1}\left(\frac{Q_\star + 1 - \frac{2}{\mu_\star} t}{c - 1}\right),$$

where $Q_\star := \frac{L_\star}{\mu_\star}$ and $T_{2K+1}(\cdot)$ is the Chebyshev polynomial (of the first kind)

- **Main fact:**

$$\|x_K - x^\star\| \geq r \exp\left(-\frac{8K}{\sqrt{Q_\star - 1}}\right)$$

Lower Bound

- Assume $Q_\star = \frac{L_\star}{\mu_\star} = \frac{L + sr^{p-2}}{sr^{p-2}} \geq 2$. Then,

$$\|x_K - x^\star\| \geq r\exp\left(-\frac{8K}{\sqrt{Q_\star - 1}}\right) \geq r\exp\left(-16K\sqrt{\frac{sr^{p-2}}{L}}\right)$$

- Using uniform convexity, we have

$$f(x_K) - f^\star \geq \frac{s}{p2^{p-2}}\|x_K - x^\star\|^p \geq \frac{s}{p2^{p-2}} r^p \exp\left(-16pK\sqrt{\frac{sr^{p-2}}{L}}\right)$$

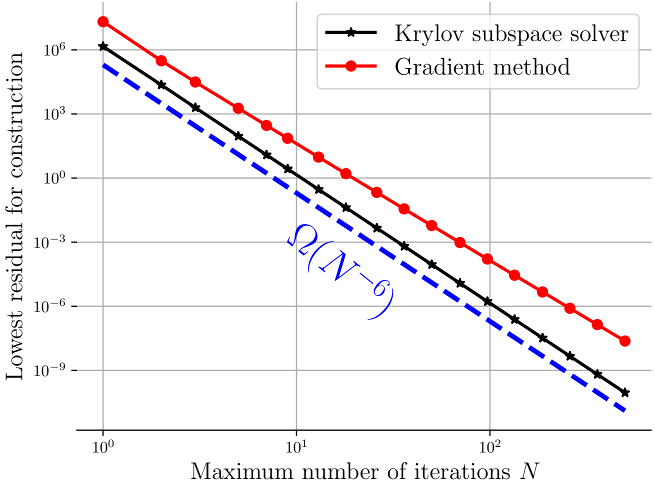
- r is arbitrary! Set

$$\boxed{r := \left(\frac{L}{3sK^2}\right)^{\frac{1}{p-2}}}$$

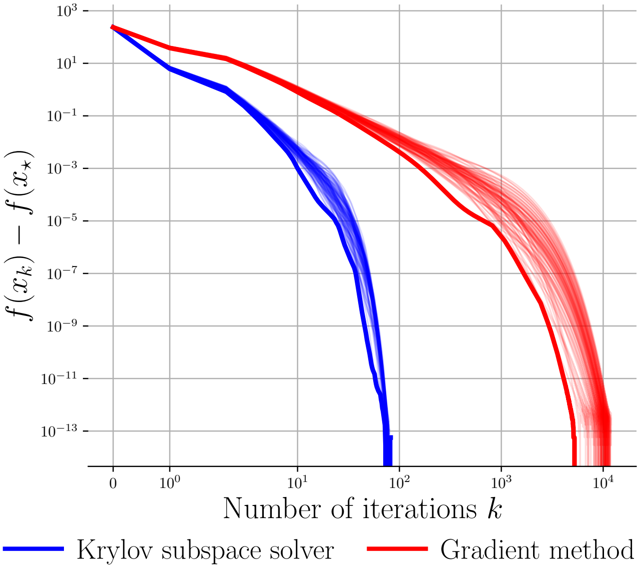
which gives the desired lower bound:

$$f(x_K) - f^\star = \Omega\left(\left[\frac{L}{s^{2/p}K^2}\right]^{\frac{p}{p-2}}\right)$$

Practical Run



Random Instances



Conclusions

- $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + \frac{s}{p}\|x\|^p$ is a natural step beyond classic problem classes
- Interpolates between convex ($s = 0$) and strongly convex ($p = 2, \sigma > 0$) functions
- Acceleration due to uniform convexity $p > 2$: the **optimal rate is**

$$f(x_K) - f^\star \quad = \quad O\Big(\Big[\frac{L}{s^{2/p}K^2}\Big]^{\frac{p}{p-2}}\Big)$$

E.g. for $p = 3$ this gives $O(1/K^6)$ and for $p = 4$ this is $O(1/K^4)$

1. "*Lower Complexity Bounds for Minimizing Regularized Functions.*" Nikita Doikov. **Optimization Letters**, 2025
2. "*Complexity of Minimizing Regularized Convex Quadratic Functions.*" Daniel Berg Thomsen, Nikita Doikov. Appear in **SIOPT**, 2026

Thank you for your attention!