Tight Lower Bounds under Asymmetric High-Order Hölder Smoothness and Uniform Convexity

Cedar Site Bai
bai123@purdue.edu

Brian Bullins
bbullins@purdue.edu

Oracle Complexity / Convergence Rate

Oracle Complexity

- Objective: $\min_x f(x)$
- # times an algorithm accesses an oracle to reach an ϵ-approximate solution

e.g., $f(x_T) - f(x^*) \leq \epsilon$

Oracle Complexity / Convergence Rate

Oracle Complexity

- Objective: $\min_x f(x)$
- # times an algorithm accesses an oracle to reach an ϵ-approximate solution

  e.g., $f(x_T) - f(x^*) \leq \epsilon$
- First-order Oracle: gradient $\nabla f(x)$
- First-order oracle complexity: # times an algorithm computes gradient

ICLR International Conference On Learning Representations

Oracle Complexity / Convergence Rate

Oracle Complexity

- Objective: $\min_x f(x)$
- # times an algorithm accesses an oracle to reach an ϵ-approximate solution

  e.g., $f(x_T) - f(x^*) \leq \epsilon$
- First-order Oracle: gradient $\nabla f(x)$
- First-order oracle complexity: # times an algorithm computes gradient
- Oracle complexity of gradient descent for convex and smooth functions: $T \in \Theta(\epsilon^{-1})$

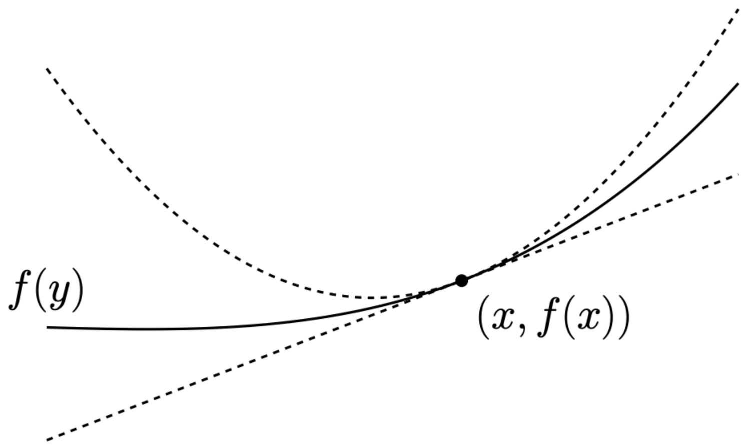
Convergence Rate

- How fast the approximation error decays with # iterations
- Convergence rate of gradient descent (convex & smooth): $f(x_T) - f(x^*) \leq \epsilon \in \Theta\left(\frac{1}{T}\right)$

Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le L \|\mathbf{x} - \mathbf{y}\|$$

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \le \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2$$



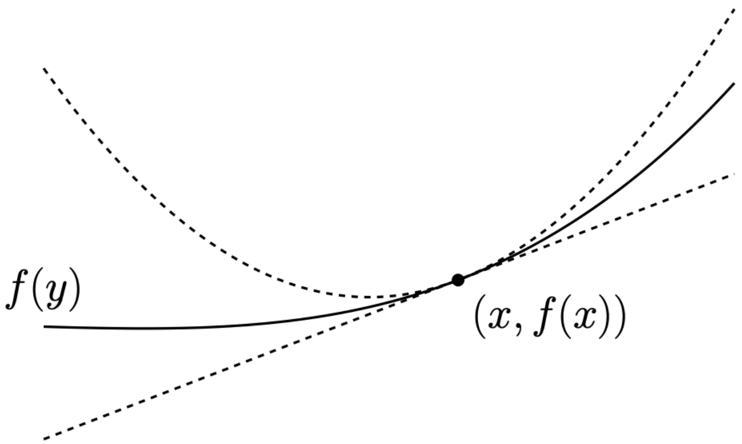
Problem Setting

Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|$$

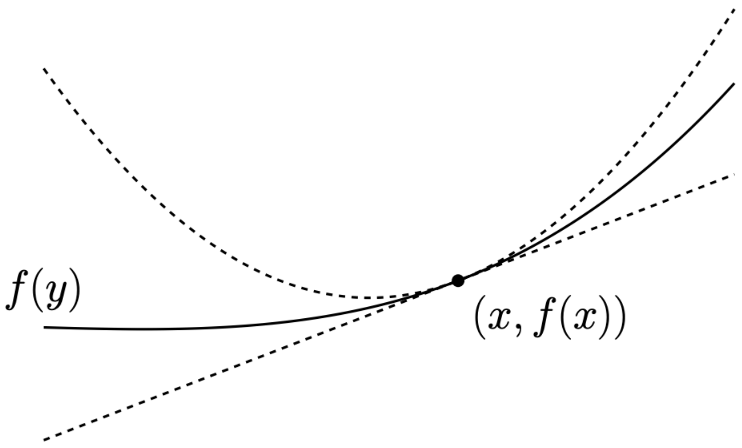
Hölder Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, H > 0, \nu \in (0, 1], \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq H \|\mathbf{x} - \mathbf{y}\|^{\nu}$$



Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\,\|\mathbf{x} - \mathbf{y}\|$$

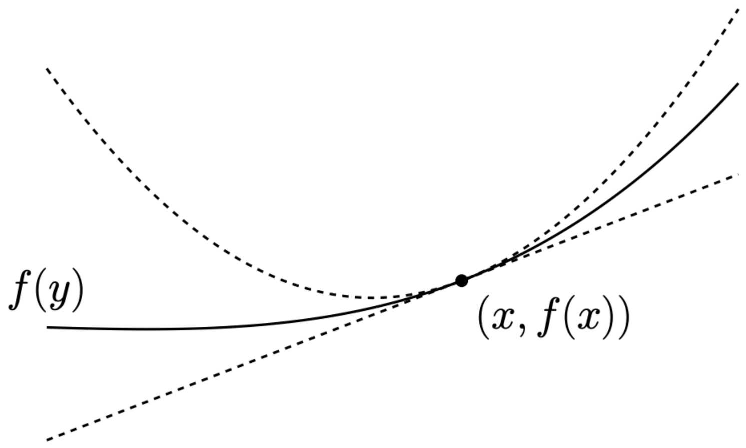


Hölder Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, H > 0, \nu \in (0,1], \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq H\|\mathbf{x} - \mathbf{y}\|^\nu$$

High-order Smoothness

$$p \in \mathbb{Z}^+, \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\| \leq L\,\|\mathbf{x} - \mathbf{y}\|$$



Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\, \|\mathbf{x} - \mathbf{y}\|$$

Hölder Smoothness

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, H > 0, \nu \in (0,1], \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq H \|\mathbf{x} - \mathbf{y}\|^{\nu}$$

High-order Smoothness

$$p \in \mathbb{Z}^+, \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, L > 0, \|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\| \leq L\, \|\mathbf{x} - \mathbf{y}\|$$

High-order Hölder Smoothness

$$p \in \mathbb{Z}^+, \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, H > 0, \nu \in (0,1], \|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\| \leq H \|\mathbf{x} - \mathbf{y}\|^{\nu}$$

ICLR International Conference On Learning Representations

Convexity

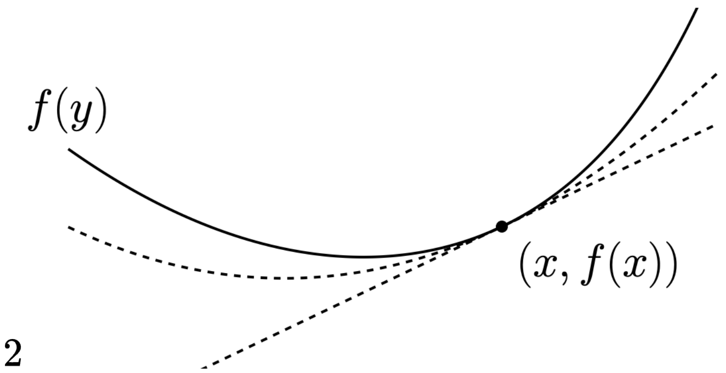
$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$$

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$$

ICLR International Conference On Learning Representations

Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$$



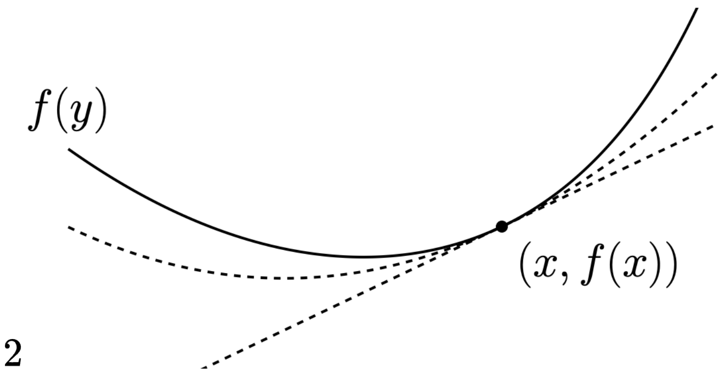
Strong Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \sigma > 0, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^2$$

ICLR International Conference On Learning Representations

Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$$



Strong Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \sigma > 0, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^2$$

Uniform Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \sigma > 0, q \geq 2, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^q$$

ICLR International Conference On Learning Representations

Motivating Example

Subproblem of Cubic-regularized Newton Method

Let M be a positive parameter. Define a modified Newton step using the following *cubic regularization* of quadratic approximation of function $f(x)$:

$$T_M(x) \in \operatorname{Arg\,min}_y \left[\langle f'(x), y-x \rangle + \frac{1}{2} \langle f''(x)(y-x), y-x \rangle + \frac{M}{6} \| y-x \|^3 \right], \quad (2.4)$$

Cubic regularization of Newton method
Initialization: Choose $x_0 \in R^d$.
Iteration k**,** $(k \geq 0)$**:** **1.** Find $M_k \in [L_0, 2L]$ such that $f(T_{M_k}(x_k)) \leq \bar{f}_{M_k}(x_k)$. **2.** Set $x_{k+1} = T_{M_k}(x_k)$.

ICLR International Conference On Learning Representations

Motivating Example

Subproblem of Cubic-regularized Newton Method

$$F(y) := \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2}\langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \frac{M}{6}\|y - x_k\|^3$$

$$x_{k+1} \in \arg\min_y F(y)$$

Cubic regularization of Newton method
Initialization: Choose $x_0 \in R^d$.
Iteration k**,** $(k \geq 0)$**:** **1.** Find $M_k \in [L_0, 2L]$ such that $f(T_{M_k}(x_k)) \leq \bar{f}_{M_k}(x_k)$. **2.** Set $x_{k+1} = T_{M_k}(x_k)$.

Motivating Example

Subproblem of Cubic-regularized Newton Method

$$F(y) := \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2}\langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \frac{M}{6}\|y - x_k\|^3$$

- Uniformly convex with degree q=3 and parameter σ:

$$F(y) - F(x) - \langle \nabla F(x), y - x \rangle \geq \frac{\sigma}{q}\|y - x\|^q$$

- Second-order smooth / Lipschitz Hessian:

$$\left\|\nabla^2 F(y) - \nabla^2 F(x)\right\| \leq M\,\|y - x\|$$

Motivating Example

Subproblem of Cubic-regularized Newton Method

$$F(y) := \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \frac{M}{6} \|y - x_k\|^3$$

- Uniformly convex with degree q=3 and parameter σ:

$$F(y) - F(x) - \langle \nabla F(x), y - x \rangle \geq \frac{\sigma}{q} \|y - x\|^q$$

- Second-order smooth / Lipschitz Hessian:

$$\left\| \nabla^2 F(y) - \nabla^2 F(x) \right\| \leq M \|y - x\|$$

- Ideally, solve with optimal second-order methods [Gasnikov et. al., 2019]

$$p = 2, q = 3, \nu = 1, q = p + \nu, \implies \mathcal{O}\left(\left(\frac{M}{\sigma}\right)^{\frac{2}{7}} \log\left(\frac{1}{\epsilon}\right)\right)$$

A. Gasnikov, P. Dvurechensky, E. Gorbunov, E. Vorontsova, D. Selikhanovych, CA. Uribe. "Optimal tensor methods in smooth convex and uniformly convex optimization." COLT, 2019.

Motivating Example

Subproblem of Cubic-regularized Newton Method

$$F(y) := \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2}\langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \frac{M}{6}\|y - x_k\|^3$$

- Ideally, solve with optimal second-order methods [Gasnikov et. al., 2019]

$$p = 2, q = 3, \nu = 1, q = p + \nu \implies \mathcal{O}\left(\left(\frac{M}{\sigma}\right)^{\frac{2}{7}} \log\left(\frac{1}{\epsilon}\right)\right)$$

- Computing Hessian is expensive: $\mathcal{O}(d^2)$
- Use first-order methods, e.g., (accelerated) gradient descent, instead: $\mathcal{O}(d)$

A. Gasnikov, P. Dvurechensky, E. Gorbunov, E. Vorontsova, D. Selikhanovych, CA. Uribe. "Optimal tensor methods in smooth convex and uniformly convex optimization." COLT, 2019.

Motivating Example

Subproblem of Cubic-regularized Newton Method

$$F(y) := \langle \nabla f(x_k), y - x_k \rangle + \frac{1}{2}\langle \nabla^2 f(x_k)(y - x_k), y - x_k \rangle + \frac{M}{6}\|y - x_k\|^3$$

- Ideally, solve with optimal second-order methods [Gasnikov et. al., 2019]

$$p = 2, q = 3, \nu = 1, \boxed{q = p + \nu} \implies \mathcal{O}\left(\left(\frac{M}{\sigma}\right)^{\frac{2}{7}} \log\left(\frac{1}{\epsilon}\right)\right)$$

- Computing Hessian is expensive: $\mathcal{O}(d^2)$
- Use first-order methods, e.g., (accelerated) gradient descent, instead: $\mathcal{O}(d)$
- Order of smoothness limited by order of oracle, e.g., H-smooth (within some domain)

$$p = 1, q = 3, \nu = 1, \boxed{q > p + \nu} \implies \mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{1}{2}} \left(\frac{\sigma}{\epsilon}\right)^{\frac{1}{6}}\right)$$

[Roulet & d'Aspremont, 2017; Song et. al. 2021]

Roulet & d'Aspremont, "Sharpness, restart and acceleration." NeurIPS, 2017.
Song et. al., "Unified acceleration of high-order algorithms under general holder continuity." SIOPT, 2021.

Uniform Convexity and High-Order Hölder Smoothness

- Uniformly Convexity

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \sigma > 0, q \geq 2, \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^q$$

- High-order Holder-smoothness

$$p \in \mathbb{Z}^+, \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, H > 0, \nu \in (0, 1], \|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\| \leq H \|\mathbf{x} - \mathbf{y}\|^\nu$$

- **Question:**

What's the p^{th}-order oracle complexity for different combinations of p, q, ν?

Related Work

Upper Bounds

Upper Bounds [Song et. al. 2021]	
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$

C. Song, Y. Jiang, and Y. Ma, "Unified acceleration of high-order algorithms under general holder continuity." SIAM Journal on Optimization 31.3 (2021): 1797-1826.

Upper Bounds

Upper Bounds [Song et. al. 2021]

$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}+\log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]

Roulet & d'Aspremont, "Sharpness, restart and acceleration." NeurIPS, 2017.
Arjevani et. al., "Oracle complexity of second-order methods for smooth convex optimization." Mathematical Programming, 2019.

Lower Bounds

	Upper Bounds [Song et. al. 2021]	**Lower Bounds [Our Work]**
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$

Lower Bounds

	Upper Bounds [Song et. al. 2021]	Lower Bounds [Our Work]
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Thomsen & Doikov, 2024]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019] $\nu = 1$ [Kornowski & Shamir 2020]

Thomsen & Doikov, "Complexity of minimizing regularized convex quadratic functions." 2024.
Kornowski & Shamir, "High-order oracle complexity of smooth and strongly convex optimization." 2020.

Hard Function Construction for Case 1: q > p + ν

- Start with a sequence of non-smooth functions

$$g_t(\mathbf{x}) = \max_{1\le k\le t} r_k(\mathbf{x}) \quad \textit{where} \quad r_k(\mathbf{x}) = \xi_k\langle \mathbf{e}_{\alpha(k)}, \mathbf{x}\rangle - (k-1)\delta, \forall\, k \in [T]$$

$$\xi_k \in \{-1, 1\},\ \mathbf{e} : \text{standard basis},\ \alpha : \text{permutation of } [T],\ \delta > 0.$$

Hard Function Construction for Case 1: q > p + ν

- Start with a sequence of non-smooth functions

$$g_t(\mathbf{x}) = \max_{1\le k\le t} r_k(\mathbf{x}) \quad \textit{where} \quad r_k(\mathbf{x}) = \xi_k\langle \mathbf{e}_{\alpha(k)}, \mathbf{x}\rangle - (k-1)\delta, \forall\, k \in [T]$$

$$\xi_k \in \{-1, 1\},\ \mathbf{e} : \text{standard basis},\ \alpha : \text{permutation of } [T],\ \delta > 0.$$

➢ Smooth the functions with **a smoothing operator**, making it p^{th}-order H-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad \textit{where} \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable},\ S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times},\ \rho > 0.$$

Hard Function Construction for Case 1: q > p + ν

- Start with a sequence of non-smooth functions

$$g_t(\mathbf{x}) = \max_{1\le k\le t} r_k(\mathbf{x}) \quad where \quad r_k(\mathbf{x}) = \xi_k\langle \mathbf{e}_{\alpha(k)}, \mathbf{x}\rangle - (k-1)\delta, \forall\, k \in [T]$$

$$\xi_k \in \{-1, 1\},\ \mathbf{e}: \text{standard basis},\ \alpha: \text{permutation of } [T],\ \delta > 0.$$

- Smooth the functions with a smoothing operator, making it pth-order H-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable},\ S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times},\ \rho > 0.$$

- Add regularization, making it qth-order uniformly convex

$$F_t(\mathbf{x}) = \beta G_t(\mathbf{x}) + \frac{\sigma}{q}\,||\,\mathbf{x}\,||^{\,q}, \quad \mathbf{x} \in \mathcal{Q},\ \beta > 0 \qquad F(\mathbf{x}) = F_T(\mathbf{x}).$$

Hard Function Construction for Case 1: $q > p + \nu$

- Start with a sequence of non-smooth functions

$$g_t(\mathbf{x}) = \max_{1\le k\le t} r_k(\mathbf{x}) \quad \textit{where} \quad r_k(\mathbf{x}) = \xi_k\langle \mathbf{e}_{\alpha(k)}, \mathbf{x}\rangle - (k-1)\delta, \forall\, k \in [T]$$

$$\xi_k \in \{-1,1\},\ \mathbf{e}: \text{standard basis},\ \alpha: \text{permutation of } [T],\ \delta > 0.$$

➢ Smooth the functions with a smoothing operator, making it p^th-order H-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad \textit{where} \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x}+\rho V)],$$

$$V\text{: random variable},\ S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times},\ \rho > 0.$$

- Add regularization, making it q^th-order uniformly convex

$$F_t(\mathbf{x}) = \beta G_t(\mathbf{x}) + \frac{\sigma}{q}\|\mathbf{x}\|^q, \quad \mathbf{x}\in\mathcal{Q},\ \beta > 0 \qquad F(\mathbf{x}) = F_T(\mathbf{x}).$$

➢ Generate $\mathbf{x}_t$'s with F_t's and show oracle access of F_t is identical to F in the neighborhood of $\mathbf{x}_t$

$$\textit{Algorithm } \mathcal{A}: \ \mathbf{x}_{t+1} = \mathcal{A}(\mathcal{I}_t(\mathbf{x}_1),\cdots,\mathcal{I}_t(\mathbf{x}_t)) \quad \textit{for} \quad \mathcal{I}_t(\mathbf{x}) = \{F_t, \nabla F_t, \cdots, \nabla^p F_t\}.$$

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad \textit{for} \quad \mathbf{x}\in\mathcal{N}_\delta(\mathbf{x}_t), \forall\, t\in[T]$$

Hard Function Construction for Case 1: q > p + ν

- Start with a sequence of non-smooth functions

$$g_t(\mathbf{x}) = \max_{1 \le k \le t} r_k(\mathbf{x}) \quad \textit{where} \quad r_k(\mathbf{x}) = \xi_k \langle \mathbf{e}_{\alpha(k)}, \mathbf{x} \rangle - (k-1)\delta, \forall\, k \in [T]$$

$$\xi_k \in \{-1, 1\},\ \mathbf{e}: \text{standard basis},\ \alpha: \text{permutation of } [T],\ \delta > 0.$$

➢ Smooth the functions with a smoothing operator, making it pth-order H-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad \textit{where} \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable},\ S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times},\ \rho > 0.$$

- Add regularization, making it qth-order uniformly convex

$$F_t(\mathbf{x}) = \beta G_t(\mathbf{x}) + \frac{\sigma}{q} \|\mathbf{x}\|^q, \quad \mathbf{x} \in \mathcal{Q},\ \beta > 0 \qquad F(\mathbf{x}) = F_T(\mathbf{x}).$$

➢ Generate $\mathbf{x}_t$'s with F_t's and show oracle access of F_t is identical to F in the neighborhood of $\mathbf{x}_t$

$$\textit{Algorithm } \mathcal{A}: \ \mathbf{x}_{t+1} = \mathcal{A}(\mathcal{I}_t(\mathbf{x}_1), \cdots, \mathcal{I}_t(\mathbf{x}_t)) \quad \textit{for} \quad \mathcal{I}_t(\mathbf{x}) = \{F_t, \nabla F_t, \cdots, \nabla^p F_t\}.$$

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad \textit{for} \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x}_t), \forall\, t \in [T]$$

- Lower bound optimality gap $F(\mathbf{x}_T) - F(\mathbf{x}^*) \ge -\beta(T-1)\delta - \frac{5}{4} p\beta\rho\sqrt{d} + \frac{q-1}{q}\left(\frac{\beta^q}{\sigma T^{\frac{q}{2}}}\right)^{\frac{1}{q-1}}$

Smoothing

- Smooth the functions with **a smoothing operator**, making it pth-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable, } S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times, } \rho > 0.$$

Smoothing

- Smooth the functions with a smoothing operator, making it p^th-order $H = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x}+\rho V)],$$

V: random variable, $S_\rho^p[\cdot]$: $S_\rho[\cdot]$ for p times, $\rho > 0$.

  - Uniform smoothing over a unit ℓ_2 ball: $H = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate [Agarwal & Hazan, 2018]

$$S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x}+\rho V)], \quad \mathbb{P}[V=\mathbf{v}] = \frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{\frac{d}{2}}} \mathbb{I}_{[\|\mathbf{v}\|_2 \le 1]}$$

ICLR International Conference On Learning Representations

Smoothing

- Smooth the functions with a smoothing operator, making it pth-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

V: random variable, $S_\rho^p[\cdot]$: $S_\rho[\cdot]$ for p times, $\rho > 0$.

- Uniform smoothing over a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate [Agarwal & Hazan, 2018]
- Moreau smoothing: $\mathrm{H} = \mathcal{O}(1)$ [Doikov, 2022], but only 1st-order smooth

$$S_\rho\left[g_t\right](\mathbf{x}) = \min_{\mathbf{y}}\left\{g_t(\mathbf{y}) + \frac{\rho}{2}\|\mathbf{y} - \mathbf{x}\|^2\right\}$$

Agarwal & Hazan, "Lower bounds for higher-order convex optimization." COLT, 2018
Doikov, “Lower complexity bounds for minimizing regularized functions.” 2022.

Smoothing

- Smooth the functions with a smoothing operator, making it p-th-order $H = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad \textit{where} \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x}+\rho V)],$$

$$V\text{: random variable, } S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times, } \rho > 0.$$

  - Uniform smoothing over a unit ℓ_2 ball: $H = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate [Agarwal & Hazan, 2018]
  - Moreau smoothing: $H = \mathcal{O}(1)$ [Doikov, 2022], but only 1^{st}-order smooth
  - Softmax smoothing: $H = \mathcal{O}(1)$, p^{th}-order smooth [Bullins, 2020]

$$\mathrm{smax}_\rho\,[g_t](\mathbf{x}) = \rho \log\left(\sum_{i=1}^{d} e^{\frac{x_i}{\rho}}\right)$$

  - Gaussian smoothing: $H = \mathcal{O}(1)$, p^{th}-order smooth after p times [Duchi et. al., 2012]

$$S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x}+\rho V)], \quad \mathbb{P}\left[V = \mathbf{v}\right] = \frac{1}{(2\pi)^{\frac{d}{2}}}\exp\left\{-\frac{\mathbf{v}^\top \mathbf{v}}{2}\right\}$$

Bullins, “Highly smooth minimization of non-smooth problems.” COLT, 2020
Duchi et. al., “andomized smoothing for stochastic optimization." SIOPT, 2012.

ICLR International Conference On Learning Representations

Smoothing

- Smooth the functions with a smoothing operator, making it p[th]-order $H = \mathcal{O}(1)$-smooth

  $$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

  V: random variable, $S_\rho^p[\cdot]$: $S_\rho[\cdot]$ for p times, $\rho > 0$.

  - Uniform smoothing over a unit ℓ_2 ball: $H = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate
  - Moreau smoothing: $H = \mathcal{O}(1)$, but only 1[st]-order smooth
  - Softmax smoothing: $H = \mathcal{O}(1)$, p[th]-order smooth
  - Gaussian smoothing: $H = \mathcal{O}(1)$, p[th]-order smooth after p times

- Generate x_t's with F_t's and show oracle access of F_t is identical to F in the neighborhood of x_t

  $$F_t(\mathbf{x}) = F(\mathbf{x}) \quad for \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x_t}), \forall\, t \in [T]$$

  - The smoothing operator needs to be **local**, accessing information only **within the neighborhood**

ICLR International Conference On Learning Representations

Smoothing

- Smooth the functions with a smoothing operator, making it pth-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

V: random variable, $S_\rho^p[\cdot]$: $S_\rho[\cdot]$ for p times, $\rho > 0$.

  - Uniform smoothing over a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate
  - Moreau smoothing: $\mathrm{H} = \mathcal{O}(1)$, but only 1st-order smooth
  - Softmax smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth
  - Gaussian smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth after p times

- Generate x_t's with F_t's and show oracle access of F_t is identical to F in the neighborhood of x_t

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad for \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x_t}), \forall\, t \in [T]$$

  - The smoothing operator needs to be **local**, accessing information only **within the neighborhood**
  - Gaussian smoothing and softmax smoothing are **global on $\mathbb{R}^d$**

$$S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$\mathbb{P}[V = \mathbf{v}] = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left\{-\frac{\mathbf{v}^\top \mathbf{v}}{2}\right\}$$

$$\mathrm{smax}_\rho[g_t](\mathbf{x}) = \rho \log\left(\sum_{i=1}^d e^{\frac{x_i}{\rho}}\right)$$

Smoothing

- **Smooth the functions** with a smoothing operator, making it p^th^-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable, } S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times, } \rho > 0.$$

  - Uniform smoothing over a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate
  - Moreau smoothing: $\mathrm{H} = \mathcal{O}(1)$, but only 1st-order smooth
  - Softmax smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth
  - Gaussian smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth after p times

- Generate x_t's with F_t's and show oracle access of F_t is identical to F in the neighborhood of x_t

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad for \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x_t}), \forall\, t \in [T]$$

  - The smoothing operator needs to be local, accessing information only within the neighborhood
  - Gaussian smoothing and softmax smoothing are global on $\mathbb{R}^d$

- **Truncated Gaussian Smoothing**

ICLR International Conference On Learning Representations

Smoothing

- **Smooth the functions** with a smoothing operator, making it pth-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad \textit{where} \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

$$V\text{: random variable, } S_\rho^p[\cdot]\text{: } S_\rho[\cdot] \text{ for } p \text{ times, } \rho > 0.$$

  - Uniform smoothing over a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate
  - Moreau smoothing: $\mathrm{H} = \mathcal{O}(1)$, but only 1st-order smooth
  - Softmax smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth
  - Gaussian smoothing: $\mathrm{H} = \mathcal{O}(1)$, pth-order smooth after p times

- Generate x_t's with F_t's and show oracle access of F_t is identical to F in the neighborhood of x_t

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad \textit{for} \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x_t}), \forall\, t \in [T]$$

  - The smoothing operator needs to be local, accessing information only within the neighborhood
  - Gaussian smoothing and softmax smoothing are global on $\mathbb{R}^d$

- **Truncated Gaussian Smoothing**
  - Truncated in a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(poly(d)) \Longrightarrow$ suboptimal rate

$$\mathbb{P}[V = \mathbf{v}] = \frac{1}{Z(d)(2\pi)^{\frac{d}{2}}} \exp\left\{-\frac{\mathbf{v}^\top \mathbf{v}}{2}\right\} \mathbb{I}_{[\|\mathbf{v}\|_2 \le 1]}$$

Smoothing

- Smooth the functions with a smoothing operator, making it p^th-order $\mathrm{H} = \mathcal{O}(1)$-smooth

$$G_t(\mathbf{x}) = S_\rho^p[g_t](\mathbf{x}) \quad where \quad S_\rho[g_t](\mathbf{x}) = \mathbb{E}_V[g_t(\mathbf{x} + \rho V)],$$

V: random variable, $S_\rho^p[\cdot]$: $S_\rho[\cdot]$ for p times, $\rho > 0$.

  - Uniform smoothing over a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(\sqrt{d}) \Longrightarrow$ suboptimal rate
  - Moreau smoothing: $\mathrm{H} = \mathcal{O}(1)$, but only 1st-order smooth
  - Softmax smoothing: $\mathrm{H} = \mathcal{O}(1)$, p^th-order smooth
  - Gaussian smoothing: $\mathrm{H} = \mathcal{O}(1)$, p^th-order smooth after p times

- Generate x_t's with F_t's and show oracle access of F_t is identical to F in the neighborhood of x_t

$$F_t(\mathbf{x}) = F(\mathbf{x}) \quad for \quad \mathbf{x} \in \mathcal{N}_\delta(\mathbf{x_t}), \forall\, t \in [T]$$

  - The smoothing operator needs to be local, accessing information only within the neighborhood
  - Gaussian smoothing and softmax smoothing are global on $\mathbb{R}^d$

- **Truncated Gaussian Smoothing**
  - Truncated in a unit ℓ_2 ball: $\mathrm{H} = \mathcal{O}(poly(d)) \Longrightarrow$ suboptimal rate
  - ✓ Truncated in a unit ℓ_∞ ball: $\mathrm{H} = \mathcal{O}(1)$, p^th-order smooth after p times [Ours, Def. 4 & Lemma 1]

$$\mathbb{P}[V = \mathbf{v}] = \frac{1}{[\Phi(1) - \Phi(-1)]^d (2\pi)^{\frac{d}{2}}} \exp\left\{-\frac{\mathbf{v}^\top \mathbf{v}}{2}\right\} \mathbb{I}_{[\|\mathbf{v}\|_\infty \le 1]}$$

ICLR International Conference On Learning Representations

Hard Function Construction for Case 2: q < p + ν

Orthogonal basis: $\mathbf{v}_i \perp \mathbf{x}_1, \cdots, \mathbf{x}_i$ *and* $\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \ \forall\, i \in [\tilde{T}]$

[Arjevani et. al., 2019] $q = 2, p = 2, \nu = 1$

$$f(\mathbf{x}) = \frac{H}{12}\left(\frac{1}{3}\sum_{i=1}^{\tilde{T}} |\langle \mathbf{v}_i, \mathbf{x}\rangle - \langle \mathbf{v}_{i+1}, \mathbf{x}\rangle|^3 - \gamma \langle \mathbf{v}_1, \mathbf{x}\rangle\right) + \frac{\sigma}{2} \|\mathbf{x}\|^2$$

[Our Work]

$$f(\mathbf{x}) = \frac{H}{2^{p+\nu+1}(p+\nu-1)!}\left(\frac{1}{p+\nu}\sum_{i=1}^{\tilde{T}} |\langle \mathbf{v}_i, \mathbf{x}\rangle - \langle \mathbf{v}_{i+1}, \mathbf{x}\rangle|^{p+\nu} - \gamma \langle \mathbf{v}_1, \mathbf{x}\rangle\right) + \frac{\sigma}{q} \|\mathbf{x}\|^q$$

Conclusion

	Upper Bounds [Song et. al. 2021]	**Tight Lower Bounds [Our Work]**
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Thomsen & Doikov, 2024]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019] $\nu = 1$ [Kornowski & Shamir 2020]

Conclusion

	Upper Bounds [Song et. al. 2021]	**Tight Lower Bounds [Our Work]**
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Thomsen & Doikov, 2024]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019] $\nu = 1$ [Kornowski & Shamir 2020]

Ad. #1

On job market, open to research in industry & academia, both ML theory & applications.

Conclusion

	Upper Bounds [Song et. al. 2021]	**Tight Lower Bounds [Our Work]**
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Thomsen & Doikov, 2024]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019] $\nu = 1$ [Kornowski & Shamir 2020]

Ad. #1

On job market, open to research in industry & academia, both ML theory & applications.

Ad. #2

Brian is super nice, passionate, knowledgeable. You are welcome to collaborate and join us!

Conclusion

	Upper Bounds [Song et. al. 2021]	**Tight Lower Bounds [Our Work]**
$q > p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Roulet & d'Aspremont, 2017]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\left(\frac{\sigma}{\epsilon}\right)^{\frac{2(q-p-\nu)}{q(3(p+\nu)-2)}}\right)$ $p = 1$ [Thomsen & Doikov, 2024]
$q = p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}}\log\left(\frac{1}{\epsilon}\right)\right)$ [Gasnikov et. al., 2019]	Future Work
$q < p + \nu$	$\mathcal{O}\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019]	$\Omega\left(\left(\frac{H}{\sigma}\right)^{\frac{2}{3(p+\nu)-2}} + \log\log\left(\left(\frac{\sigma^{p+\nu}}{H^q}\right)^{\frac{1}{p+\nu-q}}\frac{1}{\epsilon}\right)\right)$ $q = 2, p = 2, \nu = 1$ [Arjevani et al., 2019] $\nu = 1$ [Kornowski & Shamir 2020]

Thank you! Q&A

More @Poster#430 3PM Today