Open Problem: The Oracle Complexity of Smooth Convex Optimization in Nonstandard Settings

Cristóbal Guzmán

Universidad de Chile (DII)

cguzman@gatech.edu

COLT 2015
Paris, 5-07-2015

Oracle-based Algorithms

- Instance family: (κ, L)-smooth cvx. functions, $\kappa \in (1, 2]$

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L \|x - y\|^{\kappa-1} \qquad \mathcal{F}_{\|\cdot\|}(\kappa, L)$$

- Domain $X \subseteq \mathbb{R}^n$: symmetric convex body
- *Oracle* $\mathcal{O}$: e.g., $x \longmapsto (f(x), \nabla f(x))$
- Algorithm A after T (adaptive) queries outputs $x^T(A)$
- Worst-case oracle complexity:

$$\text{Risk}(T) = \inf_A \sup_{f \in \mathcal{F}_{\|\cdot\|}(\kappa, L)} \left[f(x^T(A)) - \text{Opt}(f) \right]$$

Oracle-based Algorithms

- Instance family: (κ, L)-smooth cvx. functions, $\kappa \in (1, 2]$

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L\,\|x - y\|^{\kappa-1} \qquad \mathcal{F}_{\|\cdot\|}(\kappa, L)$$

- Domain $X \subseteq \mathbb{R}^n$: symmetric convex body
- *Oracle* $\mathcal{O}$: e.g., $x \longmapsto (f(x), \nabla f(x))$
- Algorithm A after T (adaptive) queries outputs $x^T(A)$
- Worst-case oracle complexity:

$$\text{Risk}(T) = \inf_A \sup_{f \in \mathcal{F}_{\|\cdot\|}(\kappa, L)} \left[f(x^T(A)) - \text{Opt}(f) \right]$$

Oracle-based Algorithms

- Instance family: (κ, L)-smooth cvx. functions, $\kappa \in (1, 2]$

$$\|\nabla f(x) - \nabla f(y)\|_* \le L \, \|x - y\|^{\kappa - 1} \qquad \mathcal{F}_{\|\cdot\|}(\kappa, L)$$

- Domain $X \subseteq \mathbb{R}^n$: symmetric convex body
- *Oracle* $\mathcal{O}$: e.g., $x \longmapsto (f(x), \nabla f(x))$
- Algorithm A after T (adaptive) queries outputs $x^T(A)$
- Worst-case oracle complexity:

$$\text{Risk}(T) = \inf_A \sup_{f \in \mathcal{F}_{\|\cdot\|}(\kappa, L)} \left[f(x^T(A)) - \text{Opt}(f) \right]$$

Standard setting: $X = B_{\|\cdot\|}$

Non-Standard Settings

Example: Linear prediction

- Predictor $x \in X = B_p^n$
- Random examples: $(1/q + 1/q_* = 1)$

$$(a_1, b_1), \ldots, (a_m, b_m) \in B_{q_*}^n \times [-1, 1]$$

- Linear regression model

$$\text{min}\left\{ \underbrace{\frac{1}{m}\sum_{j=1}^{m}(a_j^T x - b_j)^2}_{\mathcal{F}_{\|\cdot\|_q}(2,L)} : \underbrace{\|x\|_p \leq 1}_{B_p^n} \right\}$$

ℓ_p/ℓ_q-Lower Bounds

Range q	Range p	Rate LB	UB/LB
$1 \leq q \leq 2$	$p < q$	$\tilde{\Omega}\left(\frac{1}{T^{\kappa[\frac{3}{2}+\frac{1}{p}-\frac{1}{q}]-1}}\right)$	$\tilde{O}\left(T^{\kappa[\frac{1}{p}-\frac{1}{q}]}\right)$
	$p \geq q$	$\tilde{\Omega}\left(\frac{n^{\kappa[\frac{1}{q}-\frac{1}{p}]}}{T^{\frac{3\kappa}{2}-1}}\right)$	$\tilde{O}(1)$
$2 < q \leq \infty$	$p < q$	$\tilde{\Omega}\left(\frac{1}{T^{\kappa[1+\frac{1}{p}]-1}}\right)$	$\tilde{O}\left(T^{\kappa[\frac{1}{p}-\frac{1}{q}]}\right)$
	$p \geq q$	$\tilde{\Omega}\left(\frac{n^{\kappa[\frac{1}{q}-\frac{1}{p}]}}{T^{\kappa[1+\frac{1}{q}]-1}}\right)$	$\tilde{O}(1)$

Analogous LBs for matrix Schatten norms

Non-Standard Settings

Application: Compressed Sensing ($p = 1$, $q = 2$)

$$\min\{\|Ax - b\|_2^2 : \ \|x\|_1 \leq 1\}$$

$$\text{UB} = O\left(\frac{1}{T^2}\right) \quad \text{vs} \quad \text{LB} = \Omega\left(\frac{1}{T^3}\right)$$

Conjecture

In the $\ell_1 - \ell_2$ setup above, oracle complexity is $\ll \frac{1}{T^2}$

When it would be useful?

- E.g., compressed sensing

  ($p = 1$, $q = 2$, and $f(x) = \|Ax - b\|_2^2$)
- Smoothness constants

$$\|A\|_{1\to 1} \le \|A\|_{1\to 2} \le \sqrt{m}\,\|A\|_{1\to 1}$$

When it would be useful?

- E.g., compressed sensing
  ($p = 1$, $q = 2$, and $f(x) = \|Ax - b\|_2^2$)
- Smoothness constants

$$\|A\|_{1\to1} \le \|A\|_{1\to2} \le \sqrt{m}\,\|A\|_{1\to1}$$

- Complexities *(if conjecture were true)*

$$O\left(\frac{\|A\|_{1\to1}^2}{T^2}\right) \quad \text{vs.} \quad O\left(\frac{\|A\|_{1\to2}^2}{T^3}\right)$$

- Substantial acceleration when $\|A\|_{1\to1} \approx \|A\|_{1\to2}$