Open Problem: Do Good Algorithms Necessarily Query Bad Points?

Rong Ge, Prateek Jain, Sham Kakade, Rahul Kidambi,
Dheeraj Nagaraj, Praneeth Netrapalli

Duke University
Microsoft Research India
University of Washington
MIT

Stochastic Approximation

Goal: Compute $w^* \in \mathrm{argmin}_w\, F(w) = \mathrm{E}_{\xi\sim D}[f(w;\xi)]$

- Approach: Use SGD [RM51]

$$w_{t+1} \leftarrow w_t - \eta_t \cdot \nabla f(w_t;\xi_t);$$
$$\mathrm{E}_{\xi_t}[\nabla f(w_t;\xi_t)|w_t, Z_{t-1}] = \nabla F(w_t)$$

- Iterate averaging [R88, PJ92, RSS11, JSB12, BM13]: anytime minimax optimal.

- SGD's final iterate (with fixed time horizon):
  - Non-smooth: [JNN19] optimal rates achievable.
  - Least Squares: [GKKN19] near optimal (up to a $\log T$ factor).

The Infinite Horizon Case

- **Goal:** Understand query point behavior of SGD style methods in the limit.
- SGD's final iterate behavior (in an anytime sense):
  - Non-smooth case: [SZ12, HLPR18] sub-optimal by a $\log T$ factor.
  - Strongly Convex Least Squares: [GKKN19] sub-optimal by a condition number factor.
- Does **any*** stochastic gradient procedure have to query sub-optimal points infinitely often?

* Consider the following **non-adaptive** procedure:

Suppose an algorithm can query any iterate which is expressed as a fixed (potentially non-stationary) linear combination of all previous stochastic gradients, which is defined in advance at the start of the algorithm. That is,

$$\boldsymbol{w_t} = \boldsymbol{\alpha_0 w_0} + \sum_{j<t} \boldsymbol{\alpha_j^{(t)}} \nabla \mathbf{f}(\mathbf{w}_j; \xi_j)\,, \boldsymbol{\alpha_j^{(t)}} \in \mathrm{R}\ \forall\ \boldsymbol{j}, \boldsymbol{t}.$$

Our Question

- Does a **non-adaptive procedure** query sub-optimal iterates (beyond constant factors of the minimax rate) infinitely often?
- Special cases:
  - Non-smooth case: Can we bridge the $\log T$ factor indicated by [SZ12,HLPR18]
  - Smooth + Strongly convex case: Can we bridge the condition number factor indicated by [GKKN19]
  - Gradient norm: Can we achieve similar results as [A18].