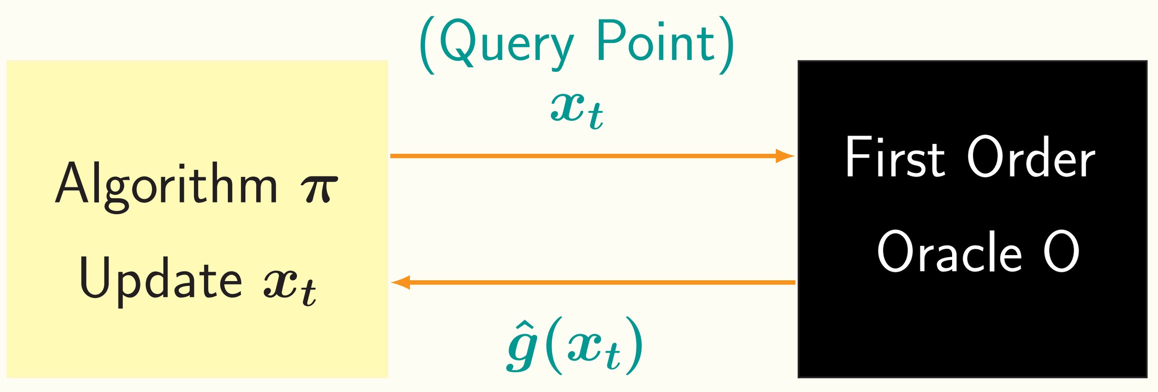
Information-Constrained Optimization: Can Adaptive Processing of Gradients help?

Jayadev Acharya, Cornell University,
Clément Canonne, University of Sydney
Prathamesh Mayekar, Indian Institute of Science
Himanshu Tyagi, Indian Institute of Science

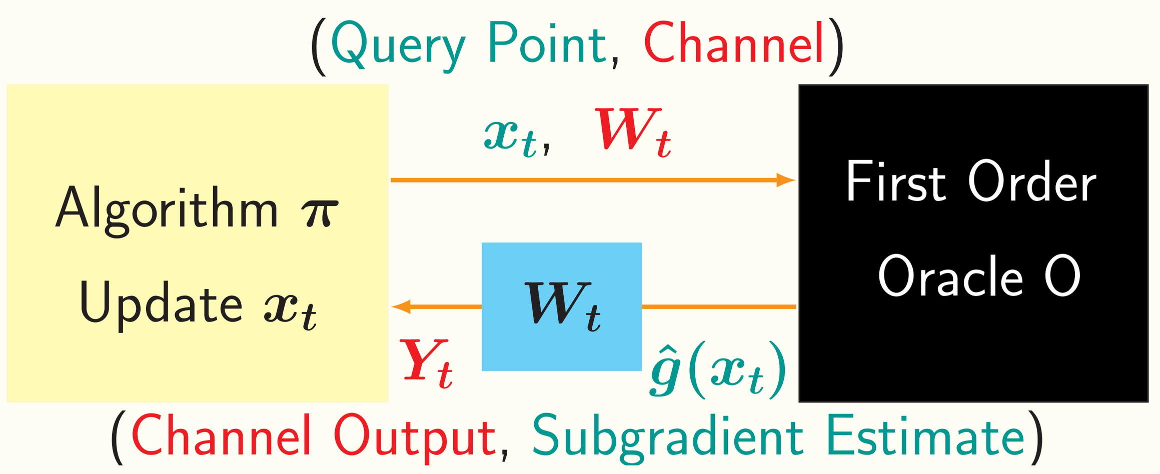
NEURAL INFORMATION PROCESSING SYSTEMS

Classical Setup: The Query Complexity framework [1]



Algorithm π: Minimize unknown function f using an oracle O.
Oracle O: Output noisy sub-gradient estimate $\hat{g}(x_t)$ for query x_t.

Our Setup: Information-Constrained Optimization



- $\hat{g}(x_t)$ can be sent using a channel $W_t \in \mathcal{W}$.
- Only the output Y_t available to the algorithm.
- $\mathcal{W}$ models the information constraints.

Does adaptive gradient coding help?

Information-Constraints

- Local Differential Privacy: The family $\mathcal{W}_{\text{priv},\varepsilon}$ comprising W s.t.

$$\ln \frac{W(y \mid x)}{W(y \mid x')} \leq \varepsilon \quad \forall x, x' \in \mathcal{X}, y \in \mathcal{Y}.$$

- Communication Constraints: the family $\mathcal{W}_{\text{com},r}$ comprising W s.t.

$$\text{(the output range) } |\mathcal{Y}| \leq 2^r.$$

- Random Coordinate Descent: the family $\mathcal{W}_{\text{obl}}$ comprising W s.t. for any input $g \in \mathbb{R}^d$,
W outputs $\{g(I), I\}$ where I is a randomly chosen coordinate.

Our Goal

Characterize

$$\mathcal{E}(T, \mathcal{W}) := \inf_{\pi \in \Pi_T} \inf_{\{W_t\}_{t\in[T]} \in \mathcal{W}} \sup_{\{f,O\} \in \mathcal{O}} \mathbb{E} f(x(\pi, W)) - f^*$$

Function, Oracle class $\mathcal{O}$ consists of tuples of $\{f, O\}$ such that

- $f : \mathcal{X} \to \mathbb{R}$ is convex.
- $\mathcal{X}$ has Euclidean diameter atmost D.
- *Unbiased Oracle:* $\mathbb{E}[\hat{g}(x)|x] \in \partial f(x)$.
- *Almost surely norm-bounded*: $\|\hat{g}(x)\|_2 \leq B$.

Classical Result (no channel constraints): $\mathcal{E}(T) = \Theta\left(\frac{DB}{\sqrt{T}}\right)$.

Main Result: Information-Constrained Opt. Lower Bounds

For T large enough,
(Private Optimization) and $\varepsilon \in [0, 1]$,

$$\mathcal{E}(T, \mathcal{W}_{\text{priv},\varepsilon}) \geq \frac{DB}{\sqrt{T}} \cdot \sqrt{\frac{d}{\varepsilon^2}}.$$

(Communication-Constrained Optimization) and $r \in [d]$,

$$\mathcal{E}(T, \mathcal{W}_{\text{com},r}) \geq \frac{DB}{\sqrt{T}} \cdot \sqrt{\frac{d}{r}}.$$

(Random Coordinate descent)

$$\mathcal{E}(T, \mathcal{W}_{\text{obl}}) \geq \frac{DB}{\sqrt{T}} \cdot \sqrt{d}.$$

These LBs are tight and achieved by nonadaptive gradient coding.

Proof: Key Ideas

- Similar to [1], [2], [3].
- Lower bound the error due to each coordinate as opposed to the total error and use a result from [3].
- Our techniques extend to the strongly convex function class.

A structured optimization problem where adaptivity helps

For $\mathcal{X} = [-1, 1]^d$, consider

$$\min_{\mathcal{X}} \|x - v\|_2^2,$$

where $v \in [-1, 1]^d$ is s-block sparse. That is,

- Only one of the block $\{is, \ldots, i(s+1)\}$ of coordinates can have non-zero values.
- All the coordinates in a block have the same absolute value.

Oracle: Outputs $2(x_t - Z_t)$,
where $\{Z_t\}_{t=1}^{\infty}$ is i.i.d., $Z_1 = \{-1, 1\}^d$, and $\mathbb{E}Z_1 = v$.
Channel Constraint: Algorithm can only see one coordinate of the gradient estimate. (RCD channel family)

The gap between adaptive and nonadaptive protocols

Lower bound for nonadaptive protocols:
For any nonadaptive protocol we can find a block-sparse v s.t.

$$\mathbb{E}\|x_T - v\|_2^2 \geq \frac{ds}{T}.$$

Upper bound for a adaptive protocol:
There exist an adaptive protocol such that

$$\mathbb{E}\|x_T - v\|_2^2 \lessapprox \frac{d + s^2}{T}.$$

References

1. Nemirovsky, A. S., and Yudin, D. B. (1983). Problem complexity and method efficiency in optimization.
2. Agarwal, A., Bartlett, P. L., Ravikumar, P., & Wainwright, M. J. (2012). Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. IEEE Transactions on Information Theory, 58(5), 3235-3249.
3. Mayekar, P., & Tyagi, H. (2020, June). RATQ: A universal fixed-length quantizer for stochastic optimization. In International Conference on Artificial Intelligence and Statistics (pp. 1399-1409). PMLR.
4. Acharya, J., Canonne, C. L., Sun, Z., & Tyagi, H. (2020). Unified lower bounds for interactive high-dimensional estimation under information constraints. arXiv preprint arXiv:2010.06562.