Structure-Exploiting Zeroth-Order Optimization

By: Anthony Nguyen
Mentors: Raghu Bollapragada, Matt Menickelly and Stefan M. Wild

Argonne National Laboratory

August 27, 2020

Background

Optimization problem: $\min_{x \in \mathbb{R}^n} \{f(x) = h \circ F(x)\}$

where $h(x) = \|x\|^2, F(x) = (F_1(x), \ldots, F_p(x))$

Goal: Minimize f using zeroth order information (using $F(x)$ evaluations and $\nabla F(x)$ unknown).

Question: Can we exploit h using an efficient algorithm to solve the optimization problem?

Assumptions and Gaussian Smoothing

Structured case: $f = h \circ F, F : \mathbb{R}^n \to \mathbb{R}^p, f : \mathbb{R}^n \to \mathbb{R}$

Assumptions: F smooth, ∇F_i are lipschitz, f is convex, and ∇f lipschitz.

Gaussian Smoothing:

$$f_\mu(x) = \mathbb{E}_{u \sim \mathcal{N}(0, I_n)}[f(x + \mu u)]$$
$$\nabla f_\mu(x) = \mathbb{E}_{u \sim \mathcal{N}(0, I_n)}\left[\frac{f(x + \mu u) - f(x)}{\mu} u\right]$$

Estimators

Draw $u \sim \mathcal{N}(0, I_n)$.

Estimators $g_\mu(x)$ for $\nabla f_\mu(x)$:

1. $\frac{f(x+\mu u)-f(x)}{\mu} u$
2. $2F(x+\mu u)^\top \left\{ \frac{F(x+\mu u)-F(x)}{\mu} \right\} u$
3. $2F(x)^\top \left\{ \frac{F(x+\mu u)-F(x)}{\mu} \right\} u$

Scheme: $x_{k+1} = x_k - h g_\mu(x_k)$

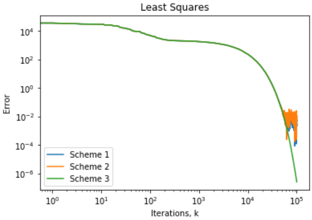
Remark: Estimator 1 is unbiased, while estimators 2 and 3 are biased for ∇f_μ.

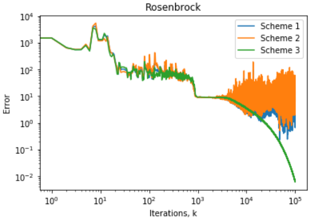
Structured Algorithm

Define: $\mathcal{U}_k = (u_0, \ldots, u_k)$, $\phi_0 = f(x_0), \phi_k = \mathbb{E}_{\mathcal{U}_{k-1}} f(x_k)$

Theorem: Suppose f is convex. If g_μ is an **unbiased estimator** for ∇f_μ, $h = \frac{1}{4\left(\frac{\mu^2}{2} L^2_{\nabla F}(n+6)^3 + (n+4) L_{\nabla f}\right)}$, and $f^* = 0$, then

$$\frac{1}{N+1} \sum_{k=0}^{N} (\phi_k - f^*) \leq \frac{4(n+4) L_{\nabla f} \|x_0 - x^*\|^2}{N+1} + \mu^2 n L_{\nabla f}$$





Future Work: Why biased estimators perform better