Meeting with Stefan: June 18, 2020

Consider $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^p$ (zero-th order oracle) and algebraically available $h : \mathbb{R}^p \to \mathbb{R}$. We want to minimize

$$f(\mathbf{x}) = h \circ \mathbf{F}(\mathbf{x})$$

In the paper, we write

$$f(\mathbf{x}) = h(\mathbf{F}(\mathbf{x}); \mathbf{x}) = h \circ g(\mathbf{x}) \tag{1}$$

where $g(\mathbf{x}) = \begin{bmatrix} \mathbf{F}(\mathbf{x}) \\ \mathbf{x} \end{bmatrix}$

Therefore, we want to minimize f. To summarize:

- h algebraically available
- $\mathbf{F}$ initially deterministic
  - Randomness is from algorithm
- $\mathbf{F}, h$ smooth as needed.
- When possible, we would like to not require explicit knowledge of $\mathbf{F}$ (eg. Lipschitz constant)
- Some cases, consider when function h is convex in its primary arguments.

Let us write

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} F_1(\mathbf{x}) \\ \vdots \\ F_p(\mathbf{x}) \end{bmatrix} \tag{2}$$

We consider finite difference approximation to the directional derivative.

$$\delta_{F_i}(F_i; \mathbf{x}_k; \mathbf{v}; \Delta_k) := \frac{F_i(\mathbf{x}_k + \Delta_k \mathbf{v}) - F_i(\mathbf{x}_k)}{\Delta_k}, \quad i = 1, \ldots, p \tag{3}$$

where $\Delta_k > 0, \|\Delta_k\| \neq 0$. We assume a specific form of f, which is

$$f(\mathbf{x}) = \sum_{i=1}^{p} F_i(\mathbf{x})^2 \tag{4}$$

then we have

$$D_v F(x_k) = \nabla_x F(\mathbf{x}_k) v \Rightarrow D_v F(\mathbf{x}_k) = 2\mathbf{v}^\top \nabla_x F(\mathbf{x}) \mathbf{F}(\mathbf{x}) \tag{5}$$

Questions:

- Swipe v through n linearly independent directions? Too expensive. But swipe 1 direction and hope that is good enough. Though we might hit a saddle point.
- How do we update $\mathbf{x}_k$? Look aat some papers Stefan suggested.
- Algebraically available, deterministic, explicit knowledge.

- Bandit methods?
- Zeroth order Oracle
- Inner function, look at two (2) points to estimate directional derivative
- Need $n+1$ coordinates to evaluate if we want to estimate full Jacobian
- 1st order - local min/max and saddle
- Single derivative - 2 zeroth order calls.
- Gaussian smoothing, bandits method.
- If $F_i(\mathbf{x}) \approx 0$ equation

  $$\nabla^2_{x,x} f(x) = 2\sum_{i=1}^{p}(\nabla_x F_i(x)\nabla_x F_i(x)^\top + F_i(x)\nabla^2_x F_i(x)) \tag{6}$$

  then it's enough to look at when directional derivative is close to zero.
- batch sampling
- Full gradient estimate
- 1st order stationary point.

Look at ref 1 in journal.text
Nesterov and Spokoing
Alg. 6

Bandits method

Homework: Read these papers

Convex stochastic optimization

Non-composite setting

(3) Walk these ideas in the paper in error bound

- Summarized them in a pdf and put in latex (important parts in repo)

Gaussian smoothing in nested optimization looks like

- Successful project leads to publication advisor not on.... This is good which says you can collaborate without your advisor and him not giving you all the ideas.

Carve something our - next one (project) - can collaborate

Obstacle - involve in Argonne (which isn't a large obstacle)

Industry: Publish without advisor.

Remiss. Time to delay graduation. Talk through what I have read and summarize it with reference in github when I talk to Matt or Raghu.

This should be a growing document in journal.text

Slack, email, talk and stay in touch with my group members.