Labs
Optimization for Machine Learning
Spring 2024

Saarland University
CISPA Helmholtz Center for Information Security
Sebastian Stich
TAs: Yuan Gao & Xiaowen Jiang
https://cms.cispa.saarland/optml24/

Problem Set 7, June 4, 2024 (Local SGD)

1 Local SGD on Heterogeneous Functions

Consider the (generalized) example from the lecture, with $f\colon \mathbb{R} \to \mathbb{R}$ defined as:

$$f_1(x) = \frac{1}{2}x^2 \qquad f_2(x) = a(x-1)^2 \qquad f(x) = \frac{1}{2}\left(f_1(x) + f_2(x)\right)\,,$$

for $a \geq 0$. Verify that the optimal solution $x^\star := \operatorname{argmin} f(x)$ is given as $x^\star = \frac{2a}{1+2a}$.

1.1 The optimal solution is not a fix point of Local SGD

Consider local SGD with stepsize $\gamma > 0$, and $\tau = 2$ local steps. Prove that when we start local SGD at $x_0 = x^\star$ we end up at

$$x_2 = x^\star + \frac{(a-2a^2)\gamma^2}{1+2a}$$

after the first averaging round.

1.2 Similarity

Based the previous observation, can you derive conditions under which $x_2 = x^\star$, i.e. the optimal solution is a fixed point? Do these conditions also hold for $\tau > 2$ local steps?

2 Verify the proof of the Local SGD Theorem (general case):

In the lecture, we left out two steps:

- Plugging the (Difference) lemma into the (Decrease) lemma, and rearranging the terms to obtain

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left\|\nabla f(\bar{\mathbf{x}}_t)\right\|^2 = \mathcal{O}\left(\frac{\Delta}{\gamma T} + \gamma L \sigma^2 + \gamma^2 L^2 (\tau^2\zeta^2 + \tau\sigma^2)\right)$$

- And the tuning of the stepsize (with respect to the constraint $\gamma \leq \frac{1}{10L\tau}$).