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 6, May 21, 2024 (Mini-batch and Async)

1 Tuning the Stepsize

Let $A, B, C \geq 0$ and $D > 0$ be given parameters. Consider the expression

$$\Psi(T, \gamma) := \frac{A}{\gamma T} + B\gamma + C\gamma^2$$

depending on T and γ. Show that for any $T \geq 1$

$$\min_{\gamma \leq \frac{1}{D}} \Psi(T, \gamma) \leq 2\left(\frac{AB}{T}\right)^{1/2} + 2C^{1/3}\left(\frac{A}{T}\right)^{2/3} + \frac{AD}{T}\,.$$

Hint: Prove the result first for the special case $C = 0$.

2 Bias-Variance Decmposition

Let $f \colon \mathbb{R}^d \to \mathbb{R}$ be a differentiable function and $\mathbf{g}(\mathbf{x})$ a gradient oracle $\mathbf{g} \colon \mathbb{R}^d \to \mathbb{R}^d$ with $\mathbb{E}[\mathbf{g}(\mathbf{x})] = \nabla f(\mathbf{x})$, $\mathbb{E}\left\|\mathbf{g}(\mathbf{x}) - \nabla f(\mathbf{x})\right\|^2 \leq M \left\|\nabla f(\mathbf{x})\right\|^2 + \sigma^2$, $\forall \mathbf{x} \in \mathbb{R}^d$. Show that

$$\mathbb{E}\left\|\mathbf{g}(\mathbf{x})\right\|^2 \leq (M+1)\left\|\nabla f(\mathbf{x})\right\|^2 + \sigma^2$$

3 Hogwild!

Consider the Howild! algorithm from the lecture. We want to prove its convergence under atomic coordinate-writes (in contrasts to atomic vector-writes as studied in the lecture).

3.1 Notation

Suppose we want to express the iterates of the algorithm as

$$\mathbf{x}_t = \mathbf{x}_0 - \gamma \sum_{k=0}^{t-1} \mathbf{J}_k^t \mathbf{g}_k$$

for matrices $\mathbf{J}_k^t \in \mathbb{R}^{d \times d}$, $k < t$. Define $\mathbf{J}_k^t$.

Hint: Considering diagonal matrices suffices.

3.2 "Difference" Lemma

Prove that the difference Lemma still holds (under the same assumptions on f and γ_{crit} as in the lecture):

$$\mathbb{E}\left\|\mathbf{x}_t - \tilde{\mathbf{x}}_t\right\|^2 \leq \frac{1}{50L^2\tau} \sum_{k=(t-\tau)_+}^{t-1} \mathbb{E}\left\|\nabla f(\mathbf{x}_k)\right\|^2 + \frac{\gamma}{5L}\sigma^2\,.$$