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 4, May 7, 2024 (CD and SGD)

Coordinate descent

Exercise 1. Efficient Implementation of Coordinate Descent *Consider the least squares objective*

$$f(\mathbf{x}) = \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|^2$$

for $A \in \mathbb{R}^{n\times d}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^d$.

a) Derive $\nabla f(\mathbf{x})$. *What is the time-complexity to compute the gradient vector* $\nabla f(\mathbf{x})$*?*

b) Given an index $i \in [d]$, *derive* $\nabla_i f(\mathbf{x})$. *What is the time-complexity to compute* $\nabla_i f(\mathbf{x})$*?*

c) Consider the following implementation of coordinate descent, where $\mathbf{y}_t$ *denotes a sequence of auxiliary variables,* $\mathbf{y}_0 = A\mathbf{x}_0$. *At iteration* t, *pick index* $i_t \in [d]$ *uniformly at random and update:*

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t - \gamma(\mathbf{e}_{i_t}^\top A^\top) \cdot (\mathbf{y}_t - \mathbf{b}) \cdot \mathbf{e}_{i_t}\,, \\ \mathbf{y}_{t+1} &= \mathbf{y}_t - \gamma(\mathbf{e}_{i_t}^\top A^\top) \cdot (\mathbf{y}_t - \mathbf{b}) \cdot (A\mathbf{e}_{i_t})\,, \end{aligned} \tag{1}$$

where $\mathbf{e}_i$ *denotes the* i*-th unit vector. Show that this is equivalent to the coordinate descent update* $\mathbf{x}_{t+1} = \mathbf{x}_t - \gamma \nabla_{i_t} f(\mathbf{x}_t)$. *What is the time-complexity of updating both sequences in* (1)*?*

SGD

Exercise 2 (Weak Growth Condition)**.** *Suppose* $F(\cdot)$ *is* L*-smooth and has a minima at* $x^\star$. *We say the stochastic gradient satisfies the* weak growth condition *with constant* c *if*

$$\mathbb{E}[\|\nabla f(\mathbf{x}, \boldsymbol{\xi})\|_2^2] \leq 2cL[F(\mathbf{x}) - F(\mathbf{x}^\star)].$$

Prove that

1. *For convex function* F, *strong growth condition implies weak growth condition.*
2. *For* μ*-strongly convex function* F, *weak growth condition implies strong growth condition.*

Practical Implementation of CD

Follow the Python notebook provided here:

colab.research.google.com/github/epfml/OptML_course/blob/master/labs/ex08/template/Lab_8.ipynb