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 — Solutions (CD and SGD)

## Coordinate descent

**Exercise (Efficient Implementation of Coordinate Descent).**

**Solution:**

a) $\nabla f(\mathbf{x}) = A^\top (A\mathbf{x} - \mathbf{b})$. Computing the gradient requires two matrix-vector products of complexity $\mathcal{O}(dn)$ and one addition of complexity $\mathcal{O}(n)$. Therefore the time-complexity is $\mathcal{O}(dn)$.

b) $\nabla_i f(\mathbf{x}) = \mathbf{e}_i \cdot \mathbf{e}_i^\top (A^\top (A\mathbf{x} - \mathbf{b}))$. Note that the time-complexity remains $\mathcal{O}(dn)$.

c) Observe that the property $\mathbf{y}_t = A\mathbf{x}_t$ holds for all $t \geq 0$ and compare with part b).
Note that computing $\mathbf{e}_{i_t}^\top A^\top$ requires only $\mathcal{O}(n)$ time, as it is just required to access the $i$-th column of $A$. Therefore, updating $\mathbf{x}_{t+1}$ requires one addition of complexity $\mathcal{O}(n)$, one scalar product of complexity $\mathcal{O}(n)$, and one addition of complexity $\mathcal{O}(1)$. Updating $\mathbf{y}_{t+1}$ requires one addition of complexity $\mathcal{O}(n)$.

## SGD

**Exercise (WGC).**

**Solution:**

- By SGC: $\mathbb{E}\|\nabla f(\mathbf{x}, \xi)\| \leq c\, \|\nabla F(\mathbf{x})\|^2$. By smoothness, we have: $\|\nabla F(\mathbf{x})\|^2 \leq 2L[F(\mathbf{x}) - F(\mathbf{x}^\star)]$
- Note that $\mu$-strong convexity implies PL condition with $\mu$, i.e. $\|\nabla F(\mathbf{x})\| \geq 2\mu[F(\mathbf{x}) - F(\mathbf{x}^\star)]$. Hence, $\mathbb{E}\|\nabla f(\mathbf{x}, \xi)\| \leq 2cL[F(\mathbf{x}) - F(\mathbf{x}^\star)] \leq \frac{2cL}{\mu}[F(\mathbf{x}) - F(\mathbf{x}^\star)]$

.

## Practical Implementation

Solutions to the Coordinate Descent exercise