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 10, July 9, 2024 (Proximal Methods & Compression)

1 Proximal Methods

1.1 Properties of Proximal Operator

1.1.1 Reformulation of proximal operator

Let g be proper closed convex, recall the following optimality condition:

$$\mathbf{x} = \operatorname*{argmin}_{\mathbf{y}} g(\mathbf{y}) \Leftrightarrow 0 \in \partial g(\mathbf{x})$$

Now show that for a proper closed convex f, the proximal operator can be formulated as the following:

$$\mathbf{u} = \operatorname{prox}_f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$$

1.1.2 Monotonicity of partial differential

Show that the subdifferential of a convex function $f(\mathbf{x})$ at $\mathbf{x} \in \mathbf{dom}(f)$ is a monotone operator, i.e.,

$$(\mathbf{u} - \mathbf{v})^\top (\mathbf{x} - \mathbf{y}) \geq 0, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f), \mathbf{u} \in \partial f(\mathbf{x}), \mathbf{v} \in \partial f(\mathbf{y}).$$

1.1.3 Firm nonexpansiveness

Let $f : \mathbb{R}^d \to \mathbb{R}$ be proper closed convex, write $P(\mathbf{x}) = \operatorname{prox}_f(\mathbf{x})$ and $Q(\mathbf{x}) = \mathbf{x} - P(\mathbf{x})$. Prove the following:

$$\|P(\mathbf{x}) - P(\mathbf{y})\|^2 + \|Q(\mathbf{x}) - Q(\mathbf{y})\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

1.2 LASSO

Consider the LASSO problem:

$$\min_{\mathbf{w}} \underbrace{\frac{1}{2} \|A\mathbf{w} - \mathbf{b}\|_2^2}_{f(\mathbf{w})} + \underbrace{\mu \|\mathbf{w}\|_1}_{\psi(\mathbf{w})} \tag{1}$$

Solving the LASSO problem with proximal gradient method requires the computation of the proximal operator of the following form (for some stepsize γ):

$$\operatorname{prox}_{\gamma\psi}(\mathbf{z}) = \operatorname*{argmin}_{\mathbf{y}} \left\{ \frac{1}{2\gamma} \|\mathbf{y} - \mathbf{z}\|^2 + \mu \|\mathbf{y}\|_1 \right\} \tag{2}$$

where $\mathbf{z} = \mathbf{w} - \gamma \nabla f(\mathbf{w}) = \mathbf{w} - \gamma A^T (A\mathbf{w} - \mathbf{b})$. Find the solution to problem (2).

2 Compression

Recall the following definitions:

Definition 1 (ω-quantizer)**.** *We say that a (possibly randomized) mapping $\mathcal{Q}_\omega : \mathbb{R}^d \to \mathbb{R}^d$ is a quantizer if for some $\omega \geq 0$ it holds*

$$\mathbb{E}_{\mathcal{Q}_\omega}\left[\mathcal{Q}_\omega(\mathbf{x})\right] = \mathbf{x}, \quad \mathbb{E}_{\mathcal{Q}_\omega}\left[\left\|\mathcal{Q}_\omega(\mathbf{x})\right\|^2\right] \leq (1+\omega)\left\|\mathbf{x}\right\|^2 \tag{3}$$

Definition 2 (δ-compressor)**.** *We say that a (possibly randomized) mapping $\mathcal{C} \colon \mathbb{R}^d \to \mathbb{R}^d$ is a contractive compression operator if for some constant $0 < \delta \leq 1$ it holds*

$$\mathbb{E}\left[\left\|\mathcal{C}(\mathbf{x}) - \mathbf{x}\right\|^2\right] \leq (1-\delta)\left\|\mathbf{x}\right\|^2 \qquad \forall \mathbf{x} \in \mathbb{R}^d. \tag{4}$$

2.1 Top-k

Show that the top-k operator is a δ-compressor. For which δ?

The $\mathrm{top}_k \colon \mathbb{R}^d \to \mathbb{R}^d$ operator is defined as

$$\left(\mathrm{top}_k(\mathbf{x})\right)_i = \begin{cases} \left(\mathbf{x}\right)_{\pi(i)} & \text{if } i \leq k \\ 0 & \text{otherwise} \end{cases}$$

where $k \in [d]$ is a parameter and π a permutation of the indices $\{1, \ldots, d\}$, such that $\left(|\mathbf{x}|\right)_{\pi(i)} \geq \left(|\mathbf{x}|\right)_{\pi(i+1)}$ for $i = 1, \ldots, d-1$. Here $\left(\mathbf{x}\right)_i$ denotes the i-th coordinate of the vector $\mathbf{x}$.

2.2 Rescaled Quantizer

Let $\mathcal{Q} \colon \mathbb{R}^d \to \mathbb{R}^d$ be an unbiased ω-quantizer. Show that $\frac{1}{1+\omega}\mathcal{Q}(\mathbf{x})$ is a δ-compressor. For which δ?