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 5 — Solutions (Newton and Adaptive gradient methods)

Newton's Method

Exercise (Almost constant Hessians).

Solution:

We use that for any two matrices, $\|AB\| \leq \|A\| \|B\|$. Indeed,

$$\|AB\| = \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|AB\mathbf{v}\|}{\|\mathbf{v}\|} \leq \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|A\| \|B\mathbf{v}\|}{\|\mathbf{v}\|} = \|A\| \|B\| .$$

Hence,

$$1 = \left\|\nabla^2 f(\mathbf{x}^\star)\nabla^2 f(\mathbf{x}^\star)^{-1}\right\| \leq \left\|\nabla^2 f(\mathbf{x}^\star)\right\| \left\|\nabla^2 f(\mathbf{x}^\star)^{-1}\right\| \leq \left\|\nabla^2 f(\mathbf{x}^\star)\right\| \frac{1}{\mu},$$

so, $\left\|\nabla^2 f(\mathbf{x}^\star)\right\| \geq \mu$.

Next, we use the triangle inequality $\|A+B\| \leq \|A\| + \|B\|$. Indeed, for some vector $\mathbf{v}^\star \neq \mathbf{0}$,

$$\begin{aligned} \|A+B\| &= \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|(A+B)\mathbf{v}\|}{\|\mathbf{v}\|} \leq \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|A\mathbf{v}\| + \|B\mathbf{v}\|}{\|\mathbf{v}\|} = \frac{\|A\mathbf{v}^\star\| + \|B\mathbf{v}^\star\|}{\|\mathbf{v}^\star\|} \\ &= \frac{\|A\mathbf{v}^\star\|}{\|\mathbf{v}^\star\|} + \frac{\|B\mathbf{v}^\star\|}{\|\mathbf{v}^\star\|} \leq \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} + \max_{\mathbf{v}\neq\mathbf{0}} \frac{\|B\mathbf{v}\|}{\|\mathbf{v}\|} = \|A\| + \|B\| . \end{aligned}$$

Now, by the Lipschitz assumption and Corollary 6.5,

$$\left\|\nabla^2 f(\mathbf{x}_T) - \nabla^2 a f(\mathbf{x}^\star)\right\| \leq B \left\|\mathbf{x}_T - \mathbf{x}^\star\right\| \leq \mu \left(\frac{1}{2}\right)^{2^T-1} .$$

Together with $\left\|\nabla^2 f(\mathbf{x}^\star)\right\| \geq \mu$, the statement follows.

Exercise (Prove Young's inequality).

Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ be arbitrary vectors. Prove that for any $\gamma > 0$:

$$\mathbf{a}^\top \mathbf{b} \leq \frac{\gamma^2}{2} \|\mathbf{a}\|^2 + \frac{1}{2\gamma^2} \|\mathbf{b}\|^2 .$$

Solution:

Proof. Note that $\mathbf{a}^\top\mathbf{b} = (\gamma\mathbf{a})^\top(\gamma^{-1}\mathbf{b})$ and hence

$$\mathbf{a}^\top\mathbf{b} = (\gamma\mathbf{a})^\top(\gamma^{-1}\mathbf{b}) = \frac{1}{2}\|\gamma\mathbf{a}\|^2 + \frac{1}{2}\left\|\gamma^{-1}\mathbf{b}\right\|^2 - \frac{1}{2}\left\|\gamma\mathbf{a} - \gamma^{-1}\mathbf{b}\right\|^2 \leq \frac{\gamma^2}{2}\|\mathbf{a}\|^2 + \frac{1}{2\gamma^2}\|\mathbf{b}\|^2 .$$

□

Exercise (Prove Cauchy-Schwarz for random variables).

Suppose $A, B \in \mathbb{R}$ are random variables. Then

$$\mathbb{E}[AB] \le \sqrt{\mathbb{E}[A^2]\mathbb{E}[B^2]}\,.$$

Solution:

Proof. By Young's inequality, we have

$$\mathbb{E}[AB] \le \frac{\gamma^2\mathbb{E}[A^2]}{2} + \frac{E[B^2]}{2\gamma^2}\,.$$

Now set $\gamma^2 = \frac{\sqrt{\mathbb{E}[B^2]}}{\sqrt{\mathbb{E}[A^2]}}$. □