Labs
Optimization for Machine Learning
Spring 2025

EPFL
School of Computer and Communication Sciences
Nicolas Flammarion
github.com/epfml/OptML_course

Problem Set 6 — Solutions (Non-Convex Optimization)

Theoretical Exercises

Exercise 38. *Let $f : \mathbb{R}^n \to \mathbb{R}$ twice differentiable, with $X \subseteq \mathbf{dom}(f)$ an open convex set, and suppose that f is smooth with parameter L over X. Prove that under these conditions, the largest eigenvalue of the Hessian $\lambda_{max}(\nabla^2 f(\mathbf{x})) \le L$ for all $\mathbf{x} \in X$.*

Solution: To prove that the largest eigenvalue of $\nabla^2 f(\mathbf{x})$ is $\le L$ for all $\mathbf{x} \in X$, we need to show that $\mathbf{y}^\top \nabla^2 f(\mathbf{x})\mathbf{y} \le L \|\mathbf{y}\|_2^2$ for all $\mathbf{x} \in X$ and $\mathbf{y} \in \mathbb{R}^d$.

Since X is an open set, for any $\mathbf{x} \in X$, there exists a $c > 0$ such that for all $\lambda \in [0, c]$ and $\mathbf{y} \in \mathbb{R}^d$, $\mathbf{x} + \lambda\mathbf{y} \in X$. For any such $\mathbf{x}, \mathbf{y}$, and λ, we can use Taylor's theorem and write

$$f(\mathbf{x} + \lambda\mathbf{y}) = f(\mathbf{x}) + \lambda\langle\nabla f(\mathbf{x}), \mathbf{y}\rangle + \frac{\lambda^2}{2}\mathbf{y}^\top\nabla^2 f(\mathbf{x})\,\mathbf{y} + o(\lambda^2)\,.$$

Using the smoothness of f over X, we also have that

$$f(\mathbf{x} + \lambda\mathbf{y}) \le f(\mathbf{x}) + \lambda\langle\nabla f(\mathbf{x}), \mathbf{y}\rangle + \frac{L\lambda^2}{2}\|\mathbf{y}\|_2^2\,.$$

These two expressions can be combined into the observation that

$$\frac{\lambda^2}{2}\mathbf{y}^\top\nabla^2 f(\mathbf{x})\,\mathbf{y} \le \frac{L\lambda^2}{2}\|\mathbf{y}\|_2^2 - o(\lambda^2)\,.$$

For any $\lambda > 0$, we can divide both sides by $\frac{\lambda^2}{2}$ and tend λ to zero to obtain:

$$\mathbf{y}^\top\nabla^2 f(\mathbf{x})\,\mathbf{y} \le \lim_{\lambda\to 0} L\|\mathbf{y}\|_2^2 - 2\frac{o(\lambda^2)}{\lambda^2} = L\|\mathbf{y}\|^2\,.$$

This proves the exercise.

Exercise 39. *Prove that the statement of Theorem 8.2 implies that*

$$\lim_{t\to\infty}\|\nabla f(\mathbf{x}_t)\|^2 = 0.$$

Solution: Let $\nabla_t = \|\nabla f(\mathbf{x}_t)\|^2$. The statement of Theorem 8.2 is that

$$S_T := \sum_{t=0}^{T-1}\nabla_t \le K := 2L\big(f(\mathbf{x}_0) - f(\mathbf{x}^\star)\big), \quad T > 0.$$

Hence, $(S_T)_{T>0}$ is a monotone increasing and bounded sequence and as such has a limit $K' \le K$. This means that for all $\varepsilon > 0$, there exists T_0 such that $S_T > K' - \varepsilon$ for all $T \ge T_0$. As $S_{T+1} \le K'$ for all T, we have

$$\nabla_T = S_{T+1} - S_T < \varepsilon$$

for all $T \ge T_0$. And this means that

$$\lim_{T\to\infty}\nabla_T = 0.$$