Optimization in Machine Learning HWS 2024

Solution Sheet 5

For the exercise class on the 21.11.2024.

Hand in your solutions by 10:15 in the lecture on Tuesday 19.11.2024.

Exercise 1 (Conditional Expectation).

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, \mathcal{F} a subalgebra of \mathcal{A} and X, Y random vectors. Prove for \mathcal{F} -measurable $X \in \mathbb{R}^d$ that we have

$$\mathbb{E}[\langle X, Y \rangle \,|\, \mathcal{F}] = \langle X, \mathbb{E}[Y \,|\, \mathcal{F}] \rangle$$

Solution. We simply use linearity of the conditional expectation

$$\mathbb{E}[\langle X, Y \rangle \,|\, \mathcal{F}] = \mathbb{E}\left[\sum_{i=1}^{d} X_i Y_i \,|\, \mathcal{F}\right] = \sum_{i=1}^{d} \mathbb{E}[X_i Y_i \,|\, \mathcal{F}] = \sum_{i=1}^{d} X_i \mathbb{E}[Y_i \,|\, \mathcal{F}] = \langle X, \mathbb{E}[Y \,|\, \mathcal{F}] \rangle. \quad \Box$$

Exercise 2 (Convexity and Expectation).

Let Z be a random variable. Let f(x) := f(x, Z) be a random function $(f(x, \omega) = f(x, Z(\omega)))$ if you want) and its expectation $F(x) = \mathbb{E}[f(x)]$

Is
$$f$$
 almost surely convex if and only if F is convex? Prove or disprove both directions.

Solution. Let us first assume f was almost surely convex. Then due to monotonicity of expectation we have

$$F(\lambda x + (1 - \lambda)y) = \mathbb{E}[f(\lambda x + (1 - \lambda)y)]$$

$$\leq \mathbb{E}[\lambda f(x) + (1 - \lambda)f(y)]$$

$$= \lambda \mathbb{E}[f(x)] + (1 - \lambda)\mathbb{E}[f(y)] = \lambda F(x) + (1 - \lambda)F(y)$$

so F is convex. The other direction is false. For this let $\mathbb{P}(Z = -1) = (1 - p)$ and $\mathbb{P}(Z = 1) = p$ with p > 0.5, and $f(x, z) = zx^2$. Then

$$F(x) = \mathbb{E}[f(x,Z)] = \mathbb{E}[Z]x^2 = (2p-1)x^2$$

is convex, but f is not convex, i.e. $f(x) = -x^2$ with probability (1 - p).

Exercise 3 (Convergence of SGD on Strongly Convex Functions). (4 Pe In the lecture we proved for *L*-smooth functions *F* and X_n generated by Algorithm 6 (SGD)

$$\|\nabla F(X_n)\|^2 \to 0$$
 a.s.

1

If we additionally have strong convexity of F, prove $||X_n - x_*|| \to 0$ almost surely.

(4 Points)

(4 Points)

(4 Points)

Solution. On sheet 3 we proved the PL inequality for *L*-smooth, strongly convex functions. This together with strong convexity implies

$$\frac{\mu}{2} \|X_n - x_*\|^2 \le F(X_n) - F(x_*) - \langle \underbrace{\nabla F(x_*)}_{=0}, X_n - x_* \rangle \stackrel{\mathsf{PL}}{\le} \frac{L}{2\mu} \|\nabla F(X_n)\|^2 \to 0.$$

From part in the middle we get $F(X_n) \to F(x_*)$ for free.

Exercise 4 (Swap Integration with Differentiation).

(i) What formal requirements on $f: V \times \Omega \to \mathbb{R}$ with $V \subseteq \mathbb{R}$ and measure μ on Ω are needed, for the following argument using the fundamental theorem of calculus (FTC) to work?

$$\begin{split} \frac{\partial}{\partial t} \int_{\Omega} f(t_0, \omega) d\mu(\omega) &\stackrel{\text{linear}}{=} \lim_{\epsilon \to 0} \int \frac{f(t_0 + \epsilon, \omega) - f(t_0, \omega)}{\epsilon} d\mu(\omega) \\ & \underset{\epsilon \to 0}{\text{FTC II}} \lim_{\epsilon \to 0} \int \frac{1}{\epsilon} \int_{t_0}^{t_0 + \epsilon} \frac{\partial}{\partial t} f(t, \omega) dt d\mu(\omega) \\ & \underset{\epsilon \to 0}{\text{Fubini}} \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{t_0}^{t_0 + \epsilon} \int \frac{\partial}{\partial t} f(t, \omega) d\mu(\omega) dt \\ & \underset{\epsilon \to 0}{\text{def.+lin.}} \frac{d}{dy} \int_{t_0}^y \int \frac{\partial}{\partial t} f(t, \omega) d\mu(\omega) dt \bigg|_{y=t_0} \\ & \underset{\epsilon \to 0}{\text{FTC I}} \int \frac{\partial}{\partial t} f(t_0, \omega) d\mu(\omega). \end{split}$$

Formulate the corresponding theorem.

- (6 pts)
- Solution. (a) Linearity of the integral, requires measurability and either positivity $(f \ge 0)$ or μ -integrability of f with regard to ω at t_0 .
- (b) For Fubini, we need either $\frac{\partial}{\partial t}f(t,\omega)\geq 0$, or

$$\int_{t_0}^{t_0+\epsilon} \int \Big| \frac{\partial}{\partial t} f(t,\omega) \Big| d\omega dt < \infty.$$

Since we let $\epsilon \to 0$, there only needs to be a small environment around t_0 , where this is the case. I.e. we need "local-integrability" around t_0 of $\partial_t f$ with regard to t

$$\exists a, b \in \mathbb{R}: \quad t_0 \in [a, b], \ a < b, \ \int_a^b \int \Big| \frac{\partial}{\partial t} f(t, \omega) \Big| d\omega dt < \infty.$$

This also covers the second usage of linearity.

(c) For the second fundamental theorem of calculus (FTC II), we do not even need $f(\cdot, \omega)$ to be continuous. It is sufficient, if it is for μ -almost-all ω absolutely continuous (i.e. a density exists).

(12	Points)

(d) For the first fundamental theorem of calculus (FTC I), we need continuity. I.e.

$$t\mapsto \int \partial_t f(t,\omega) d\mu(\omega)$$

needs to be continuous.

So we get

Theorem (Swap Integration and Differentiation). Let $f : U \times \Omega \to \mathbb{R}$ be a measurable function for $U \subseteq \mathbb{R}$ which satisfies

- (a) $f \mu$ -integrable over $\omega \in \Omega$ at $t_0 \in U$,
- (b) for μ -almost all $\omega \in \Omega$ we have: $t \to f(t, \omega)$ is differentiable (or absolutely continuous) in t.
- (c) If $\frac{\partial}{\partial t}f$ is further "locally integrable in t_0 ", i.e. there exists $a < b \in \mathbb{R}$, such that $t_0 \in \mathbb{R}$ $[a, b] \subseteq U$ and

$$\exists a < b \in \mathbb{R} : \quad t_0 \in [a, b] \subseteq U, \quad \int_a^b \int_\Omega \left| \frac{d}{dt} f(t, \omega) \right| d\mu(\omega) dt < \infty,$$

or $\frac{\partial}{\partial t} f \ge 0$ in the neighborhood [a, b]. (d) In a similar local neighborhood [a, b] of t_0 assume that

$$t\mapsto \int \frac{\partial}{\partial t} f(t,\omega) d\mu(\omega)$$

is continuous.

then swapping derivative in t_0 and integration over ω is allowed

$$\frac{\partial}{\partial t} \int f(t_0, \omega) d\mu(\omega) = \int \frac{\partial}{\partial t} f(t_0, \omega) d\mu(\omega).$$

(ii) We want to find an example for a function, where you can not swap integration with differentiation. So for a function $f(t, \omega)$ we need some t_0 such that

$$\frac{\partial}{\partial t}\int_{\Omega}f(t_0,\omega)d\omega\neq\int_{\Omega}\frac{\partial}{\partial t}f(t_0,\omega)d\omega.$$

For this consider $f(t,\omega) = t^3 e^{-t^2 \omega}$. Prove the inequality at $t_0 = 0$ and $\Omega = [0,\infty)$. Why is this not a contradiction to (i)? (6 pts)

Solution. We have

$$\int_0^\infty f(t,\omega)d\omega = \int_0^\infty t^3 e^{-t^2\omega}d\omega = \begin{cases} -te^{-t^2\omega} \Big|_{\omega=0}^{\omega=\infty} & t \neq 0\\ 0 & t = 0 \end{cases}$$
$$= t.$$

So its derivative is constant

$$\frac{\partial}{\partial t} \int_0^\infty f(t,\omega) d\omega = 1.$$

In particular this is also the case at $t_0 = 0$. On the other hand we have

$$\frac{\partial}{\partial t}f(t,\omega) = 3t^2e^{-t^2\omega} - 2t^4\omega e^{-t^2\omega} = t^2e^{-t^2\omega}(3-2t^2\omega).$$

In particular $\frac{\partial}{\partial t}f(t_0,\omega)=0.$ Therefore

$$\int_0^\infty \frac{\partial}{\partial t} f(t_0, \omega) d\omega = 0 \neq 1 = \frac{\partial}{\partial t} \int_0^\infty f(t_0, \omega) d\omega$$

We also have for $t \neq 0$

$$\begin{split} \int_0^\infty \frac{\partial}{\partial t} f(t,\omega) d\omega &= -3e^{-t^2\omega} \Big|_{\omega=0}^{\omega=\infty} -2t^2 \int_0^\infty t^2 \omega e^{-t^2\omega} d\omega \\ &= 3 - 2t^2 \bigg[\omega e^{-t^2\omega} \Big|_{\omega=0}^{\omega=\infty} - \int_0^\infty \underbrace{\left(\frac{d}{d\omega}\omega\right)}_{=1} e^{-t^2\omega} d\omega \bigg] \\ &= 3 - 2 \int_0^\infty t^2 e^{-t^2\omega} d\omega \\ &= 1 \end{split}$$

So in total

$$\int_0^\infty \frac{\partial}{\partial t} f(t,\omega) d\omega = \begin{cases} 0 & t = 0\\ 1 & t \neq 0. \end{cases}$$

In particular

$$t\mapsto \int_0^\infty \frac{\partial}{\partial t} f(t,\omega) d\omega$$

is not continuous. But this was a requirement for (i) so this is not a contradiction.