Dr. Peter Arndt

Master for AI and Data Science
Universität Düsseldorf

Math for AI
Winter term 2023/24

Exercise sheet 5

Exercise 16 (10 points)

(a) (3 point) Compute the differential and the derivative (or its transpose) of the map $\mathbb{R}^{n\times m} \to \mathbb{R}$ $X \mapsto \mathrm{tr}(AX^TB)$ by the method given in the lecture (here A should also be an $n \times m$-matrix and B an $n \times k$-matrix).

(b) (4 points) Compute the differential and the derivative of the map $\mathbb{R}^{n\times n} \to \mathbb{R}$ $X \mapsto \mathrm{tr}(X^n)$ by the method given in the lecture.

[Remark: You can use without proof that $\mathrm{tr}(Y\epsilon Z\epsilon W)$ is $o(\|\epsilon\|)$ for any matrices Y, Z, W.]

(c) (3 points) The matrix exponential function is the function $\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, $X \mapsto e^X$ defined by $e^X := \sum_{k=0}^{\infty} \frac{1}{k!} X^k$. One can show that this sum of matrices converges to a matrix. Compute the differential of the map $\mathbb{R}^{n\times n} \to \mathbb{R}$ $X \mapsto \mathrm{tr}(e^X)$ by the method given in the lecture.
[Remark: You can use without proof that you can pull the above infinite sum out of the trace. Reason: trace is a linear map (can pull out sums), and therefore continuous (can pull out the limit).]

Exercise 17 (10 points)

A casino owner wants to offer a betting game with n possible outcomes, which give her revenues of $a_1, \dots, a_n$ with $a_i \in \mathbb{R}$. She can freely set up the game choosing the the probability p_i of outcome i. She wants to base her choice on various factors, e.g. the expected revenue should be high enough, it should also be reliable enough (she wants not too big variance), but also not too predictable (the variance should not be low). The following questions are part of checking which of the requirements can be achieved by solving *convex* optimization problems.

(a) (2 points) Consider the set of possible probability distributions $P = \{p = (p_1, \dots, p_n) \mid p_i \geq 0,\ \Sigma_i p_i = 1\}$. Show that this set is a convex subset of $\mathbb{R}^n$.

(b) (2 points) The expected revenue is $E(p) := \sum_{i=1}^n p_i a_i$. Is $E\colon P \to \mathbb{R}$ a convex function? Is it a concave function?

(c) (2 points, tricky) The variance of her revenue is $Var(p) := \sum_{i=1}^n p_i(a_i - E(p))^2$. Is $Var\colon P \to \mathbb{R}$ a convex function? Is it a concave function?

[Hint: You can use without proof that $Var(p) = \sum_{i=1}^n p_i a_i^2 - E(p)^2$ – this is the formula for random variables $Var(X) = E(X^2) - (EX)^2$ You might also want to prove and use that $x \mapsto x^2$ is a convex function and then apply Example 3.3.9.7 of the notes]

(d) (2 points) Let $\alpha \in \mathbb{R}$. Is $\{p \in P \mid Var(p) \geq \alpha\}$ a convex set?

(e) (2 points) Let $\alpha \in \mathbb{R}$. Is $\{p \in P \mid Var(p) \leq \alpha\}$ a convex set?

If the function or set in question is not convex (or concave), give concrete values for n, a_i, p_i where the necessary condition is violated.
[Hint: You can use without proof that if $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the “epigraph” $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq y\}$ is a convex set (you can at least convince yourself of this visually, but you can also prove this as a further exercise - it’s not hard).]

Exercise 18 (12 points)

(a) (2 points) Is the function $f\colon \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto \frac{x}{y}$ convex? Prove your answer.

(b) (2 points) A quadratic function $f\colon \mathbb{R}^n \to \mathbb{R}$ is a function of the form $x \mapsto x^T Ax + b^T x + c$, for some $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Give conditions on A, b and b which characterize when exactly such a function is convex.

(c) (4 points) Consider the set of probability distributions on an n-element set, $P := \{p = (p_1, \ldots, p_n) \mid p_i \geq 0,\ \Sigma_i p_i = 1\}$. Show that the following function is strictly convex on P:

$$f\colon P \to \mathbb{R}, \qquad f(p) := \sum_{i=1}^{n} p_i \log p_i$$

Here we adopt the convention that $0 \cdot \log 0 = 0$.

[Remark: The function f is called *negative entropy.* The *entropy* of a probability distribution (i.e. $-f(p)$) is a measure of how uncertain is the result of a random experiment with the probabilities given by p. It is widely used in Machine Learning, and we will cover it in the part on Information Theory. By (c) entropy is a concave function, and that is good, because its maximization is a standard task in Bayesian Statistics.]

(d) (4 points) The *Kullback-Leibler divergence* is a distance measure between probability distributions. It is the map $D_{KL}(-\|-)\colon P \to \mathbb{R}$ defined by $D_{KL}(p\|q) := \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right)$. It does not satisfy all the properties of a metric, but at least the following two of them:

Show that $D_{KL}(p\|q) \geq 0$ for all $p, q \in P$ and that $D_{KL}(p\|q) = 0$ if and only if $p = q$.

[Hint: Use the function f of part (c): Show that $D_{KL}(p\|q) = f(p) - f(q) - \nabla f(q)^T(p - q)$. Then use the first order characterization of strict convexity from Prop. 3.3.3(a) of the manuscript, applied to f: This says that a differentiable function f is convex if and only if its domain is convex and $f(y) \geq f(x) + (grad f)(x)(y - x)$ for all x, y.]

Exercise 19 (8 points)

(a) (5 points) Write a function `min_list()` that, given a list of points in $\mathbb{R}^2$, returns a minimal list of points that have the same convex hull as the given list.

Print the results for the following lists:

L_1`:=[(0.0,0.0), (7.0,0.0), (3.0,1.0), (5.0,2.0), (5.0,5.0), (3.0,3.0), (1.0,4.0), (9.0,6.0)]`
L_2`:=[(0.0,1.0), (1.0,0.0), (3.0,1.0), (2.5,2.0), (1.5,1.0), (1.0,3.0), (1.5,1.5)]`
L_3`:=[(0.5,0.5), (1.0,1.0), (0.0,0.5), (1.0,1.5), (0.5,2.0), (1.0,0.5)]`

[Remark: It is convenient for part (b), if your function returns the minimal set of points in the order that you would encounter them travelling around the border of your polygon. A possible algorithm is to start at the point l which is furthest to the left (and the lowest if there are several). Then look for a point n such that all other points are left of the line from l to n (to realize this concretely it may help to consider the halfspaces for the line on which the points lie). Replace l by n and repeat until you come back to the original l. You can also use any other correct algorithm.]

(b) (3 points) Write a function `is_in_convex_hull()` that, given a list of points in $\mathbb{R}^2$ and another single point $P \in \mathbb{R}^2$ (in that order), returns `true` if P lies in the convex hull of the other points, and `false` otherwise.

Print out a table showing the results for the lists from part (a) and the following points (in all combinations, i.e. also $P_3 \in conv(L_1)$ etc.):

$P_1 := (-1.0, 4.0)$, $P_2 := (0.5, 0.5)$, $P_3 := (5.0, 3.0)$

Deadline: Friday 17th of November, 10:00.
Upload your solution to this link.