FEDERAL STATE AUTONOMOUS
EDUCATIONAL INSTITUTION OF HIGHER EDUCATION
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY
(STATE UNIVERSITY)
PHYSTECH-SCHOOL OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

Homework.

# Deep learning

3-rd year student, group B05-003
**Kreinin Matvei**

**Moscow, 2022**

# Contents

# 1 Matrix calculus

## 1.1 Problem № 1

Find the gradient $\nabla f(x)$, if $f(X) = det\left(X^{-1} + A\right)$

**Solution:**

$$df(X) = det\left(X^{-1} + A\right)\langle (X^{-1}+A)^{-T}, d(X^{-1}+A)\rangle = det\left(X^{-1} + A\right)\langle X^{T}+A^{-T}, dX^{-1}\rangle =$$

$$= det\left(X^{-1} + A\right)\langle (X + A^{-1})^{T}, -X^{-1}dXX^{-1}\rangle$$

$$= -det\left(X^{-1} + A\right)\langle X^{-T}(X + A^{-1})^{T}X^{-T}, dX\rangle$$

**Answer:** $\nabla_X f(x) = -det\left(X^{-1} + A\right) X^{-T}(X + A^{-1})^{T}X^{-T}$

## 1.2 Problem № 2

Find $\nabla f(x)$ and $\nabla^2 f(x)$ for:

- $f : \mathbb{R}_{++} \to \mathbb{R}$, $f(t) = ||(A + tI_n)^{-1}b||$, where $A \in \mathbb{S}^n_+, b \in \mathbb{R}^n$
- $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2}||xx^T - A||^2_F$, where $A \in \mathbb{S}^n$

**Solution:**

$C = (A + tI_n)^{-1}$, $\frac{dC}{dt} = -CCdt$

$$df(t) = d(||Cb||) = \frac{1}{2||Cb||}d\langle Cb, Cb\rangle = \frac{-1}{||Cb||}\langle CCb, Cb\rangle dt$$

$$d^2f(t) = \frac{1}{||Cb||}\left(\langle CCb, CCb\rangle + 2\langle CCCb, Cb\rangle\right)dt^{'}dt - \frac{1}{2||Cb||^3}\langle CCb, Cb\rangle d\langle Cb, Cb\rangle dt^{'} =$$

$$= \frac{3}{||Cb||}\langle CCb, CCb\rangle dt^{'}dt - \frac{1}{||Cb||^3}(\langle CCb, Cb\rangle)^2 dt^{'}dt$$

Let's solve the next one:

$$df = d\langle xx^T - A, xx^T - A\rangle = \langle dxx^T + xdx^T, xx^T - A\rangle =$$

$$= tr(x^T(xx^T - A)^T dx) + tr(dx^T(xx^T - A)^T x) = 2\cdot\langle (xx^T - A)x, dx\rangle$$

$$d^2 f = 2\cdot\langle (xx^T - A)dx_2 + 2\cdot dx_2 xx^T, dx_1\rangle = \langle 6xx^T - 2A, dx_1 dx_2^T\rangle$$

**Answer:**

- $\nabla f(t) = -\frac{\langle CCb, Cb\rangle}{||Cb||}$, $\nabla^2 f(t) = \frac{3}{||Cb||}\langle CCb, CCb\rangle - \frac{1}{||Cb||^3}(\langle CCb, Cb\rangle)^2$, where $C = (A + tI_n)^{-1}$
- $\nabla f = 2(xx^T - A)x$, $\nabla^2 f = 6xx^T - 2A$

### 1.3 Problem № 3

Find $\nabla_b L$ and $\nabla_A L$, if we know $\nabla_x L$

**Solution:** $Ax = b$, $x = A^{-1}b$, $\frac{\partial x}{\partial b} = A^{-T}$

$$\nabla_b L = \frac{\partial L}{\partial x}\frac{\partial x}{\partial b} = A^{-T}\nabla_x L$$

$$dL = \frac{\partial L}{\partial x} - A^{-1}dAA^{-1}b = -(A^{-T}\frac{\partial L}{\partial x}^T b^T A{-}T)$$

$$\nabla_A L = \frac{\partial L}{\partial x}\frac{\partial x}{\partial A} = -A^{-T}\nabla_x L b^T A^{-T}$$

**Answer:** $\nabla_A L = -A^{-T}\nabla_x L b^T A^{-T}$, $\nabla_b L = A^{-T}\nabla_x L$

### 1.4 Problem № 4

For each of the following functions, find all the stationarity points and specify the parameter values at which they exist.

- $f : E \to \mathbb{R}$, $f(x) = \langle a, x\rangle - ln(1 - \langle b, x\rangle)$,
  where $a, b \in \mathbb{R}^n$, $a, b \neq 0$, $E = \{x \in \mathbb{R}^n | \langle b, x\rangle < 1\}$
- $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \langle c, x\rangle \exp(-\langle Ax, x\rangle)$, where $c \in \mathbb{R}^n$, $c \neq 0$, $A \in \mathbb{S}^n$
- $f : \mathbb{S}^n_{++} \to \mathbb{R}$, $f(X) = \langle X^{-1}, I_n\rangle - \langle A, X\rangle$, where $A \in \mathbb{S}^n$

**Solution:** $f : E \to \mathbb{R}$, $f(x) = \langle a, x\rangle - ln(1 - \langle b, x\rangle)$

$$\nabla f = a + \frac{b}{1 - \langle b, x\rangle} = 0$$

We can easily add two equations:

$$\langle a, a\rangle + \frac{\langle a, b\rangle}{1 - \langle b, x\rangle} = 0$$

$$\langle a, b\rangle + \frac{\langle b, b\rangle}{1 - \langle b, x\rangle} = 0$$

$$\langle b, x\rangle = \frac{\langle a, b\rangle}{\langle a, a\rangle} + 1$$

$$\langle b, x\rangle = \frac{\langle b, b\rangle}{\langle a, b\rangle} + 1$$

And here from that we get:

$$\frac{\langle b, b\rangle}{\langle a, b\rangle} = \frac{\langle a, b\rangle}{\langle a, a\rangle}$$

And from that we get conditions on a and b:

$$\langle a, b\rangle^2 = \langle a, a\rangle\langle b, b\rangle$$

$f : \mathbb{R}^n \to \mathbb{R},\ f(x) = \langle c, x\rangle \exp(-\langle Ax, x\rangle)$, where $c \in \mathbb{R}^n,\ c \neq 0,\ A \in \mathbb{S}^n$

$$\nabla f = c\exp(-\langle Ax, x\rangle) - 2\langle c, x\rangle \exp(-\langle Ax, x\rangle)x^T A = 0$$

$$c - 2\langle c, x\rangle x^T A =$$

Multiply it on c:

$$\langle cA^{-1}, c\rangle = 2\langle c, x\rangle^2$$

And from that we get:

$$x = \frac{cA^{-1}}{2\langle c, x\rangle} = \frac{\pm cA^{-1}}{\sqrt{2\langle cA^{-1}, c\rangle}}$$

$f : \mathbb{S}^n_{++} \to \mathbb{R},\ f(X) = \langle X^{-1}, I_n\rangle - \langle A, X\rangle$, where $A \in \mathbb{S}^n$

$$\nabla f(X) = \langle -X^{-1}dXX^{-1}, I\rangle - \langle A, dX\rangle = -tr(X^{-T}dX^TX^{-T}) - tr(A^TdX) =$$

$$= -\langle X^{-1}X^{-1} + A, dX\rangle$$

From that we get that: $XX = (-A)^{-1},\ X = \sqrt{(-A)^{-1}}$

In course of optimization I had the same problem and I have already found answer in the internet: $X = Q^T\Lambda^{-0.5}Q$, where $A = -Q^T\Lambda Q$, we could prove it by putting X into equation and show that it's true.

**Answer:**

- $\langle a, b\rangle^2 = \langle a, a\rangle\langle b, b, x\rangle$, it means that $a = b \cdot t$, where $t \in \mathbb{R}$ and x such that: $\langle b, x\rangle = \frac{\langle b,b\rangle}{\langle a,b\rangle} + 1$
- $x = \frac{\pm cA^{-1}}{\sqrt{2\langle cA^{-1},c\rangle}}$
- $X = Q^T\Lambda^{-0.5}Q$

## 1.5 Problem № 5

X – symmetric matrix and $X = Q^T\Lambda Q$, we know $\nabla_Q f,\ \nabla_\Lambda f$. Find $\nabla_X f = ?$

**Solution:**

$$df(X) = df(Q^T\Lambda Q) = \langle \nabla_Q, f d_Q(Q^T\Lambda Q)\rangle + \langle \nabla_\Lambda f, d_\Lambda(Q^T\Lambda Q)\rangle =$$

$$= \langle \nabla_Q f, Q^T\Lambda dQ\rangle + \langle \nabla_Q f, dQ^T\Lambda Q\rangle + \langle \nabla_\Lambda f, Q^T d\Lambda Q\rangle$$

$$= \langle \Lambda Q\nabla_Q f, dQ\rangle + \langle \Lambda Q\nabla_Q f^T, dQ\rangle + \langle Q\nabla_\Lambda f Q^T, d\Lambda\rangle = \langle \nabla_X f, dX\rangle =$$

$$= \langle \nabla_X f, dQ^T \Lambda Q + Q^T d\Lambda Q + Q^T \Lambda dQ \rangle =$$
$$= \langle Q \nabla_X f Q^T, d\Lambda \rangle + \langle \Lambda Q \nabla_X f^T, dQ \rangle + \langle \Lambda Q \nabla_X f, dQ \rangle$$

From that we get two equations:

$$Q \nabla_X f Q^T = Q \nabla_\Lambda f Q^T$$

and

$$\Lambda Q \nabla_Q f + \lambda Q \nabla_Q f^T = \Lambda Q \nabla_X f + \Lambda Q \nabla_X f^T$$

and it's trivial that: $\nabla_X f = \nabla_\Lambda f$

**<u>Answer:</u>** $\nabla_X f = \nabla_\Lambda f$