Deep Learning, Autumn 2019

Homework 1. Matrix differentiation

Deadline (strict, no delay is allowed): 15 Sep. 2019, 23:59

Upload format: pdf file prepared using $\TeX$ formulas or scan of hand-written papers (in one pdf file with proper ordering of pages).

Hereinafter we use the following notation:

- $\langle x, y\rangle$ – Euclidean inner product;
- $\|x\| = \langle x, x\rangle^{1/2} = (x^T x)^{1/2}$ – Euclidean vector norm;
- $\|A\|_F = \langle A, A\rangle^{1/2} = \mathrm{tr}(A^T A)^{1/2}$ – Frobenius matrix norm;
- $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0 \ \forall i\}$, $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x_i > 0 \ \forall i\}$;
- $\mathbb{S}^n = \{A \in \mathbb{R}^{n\times n} \mid A = A^T\}$;
- $\mathbb{S}^n_+ = \{A \in \mathbb{S}^n \mid A \text{ is non-negative definite}\}$, $\mathbb{S}^n_{++} = \{A \in \mathbb{S}^n \mid A \text{ is positive definite}\}$.

1. For each of the following functions f find first and second derivatives:

   (a) $f : E \to \mathbb{R}$, $f(t) = \det(A - tI_n)$, where $A \in \mathbb{R}^{n\times n}$, $E = \{t \in \mathbb{R} : \det(A - tI) \neq 0\}$.

   (b) $f : \mathbb{R}^n_{++} \to \mathbb{R}$, $f(t) = \|(A + tI_n)^{-1} b\|$, where $A \in \mathbb{S}^n_+$, $b \in \mathbb{R}^n$.

2. For each of the following functions f find gradient ∇f and Hessian $\nabla^2 f$:

   (a) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2}\|xx^T - A\|_F^2$, where $A \in \mathbb{S}^n$;

   (b) $f : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$, $f(x) = \langle x, x\rangle^{\langle x, x\rangle}$;

   (c) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|Ax - b\|^p$, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $p \geq 2$.

3. For each of the following functions f find all stationary points and indicate possible parameter values when they exist:

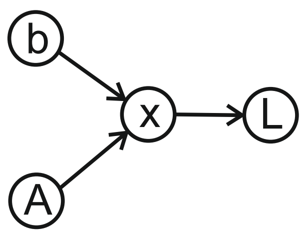
   (a) $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \langle c, x\rangle + \frac{\sigma}{3}\|x\|^3$, where $c \in \mathbb{R}^n$, $c \neq 0$, $\sigma > 0$;

   (b) $f : E \to \mathbb{R}$, $f(x) = \langle a, x\rangle - \ln(1 - \langle b, x\rangle)$, where $a, b \in \mathbb{R}^n$, $a, b \neq 0$, $E = \{x \in \mathbb{R}^n \mid \langle b, x\rangle < 1\}$;

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

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

4. Suppose we have the following calculation graph:



   Here $A \in \mathbb{R}^{n\times n}$ is some non-degenerate matrix, $b \in \mathbb{R}^n$ and x is a unique solution of the linear system $Ax = b$. L is some scalar loss function that somehow depends on x. The task is to backpropagate the gradient through this calculation graph, i.e. given $\nabla_x L$ find $\nabla_b L$ and $\nabla_A L$.