1. Question: Derivatives of functions taking scalars as inputs *(elementary)*

1.1. Calculate the gradient of the following two functions

(i) $F : \mathbb{R} \longrightarrow \mathbb{R}^2$

$$F(x) = \begin{pmatrix} x^3 \\ 2e^x \end{pmatrix}.$$

(ii) $G : \mathbb{R} \longrightarrow \mathbb{R}^3$

$$G(x) = \begin{pmatrix} 0 \\ x^3 + 2x^2 \\ \cos(x) \end{pmatrix}.$$

1.2. Calculate the gradient of the following two functions

(i) $F : \mathbb{R} \longrightarrow \mathbb{R}^{2\times 3}$

$$F(x) = \begin{pmatrix} x^2 & 2e^x & 0 \\ 0 & x & \ln(x) \end{pmatrix}.$$

(ii) $G : \mathbb{R} \longrightarrow \mathbb{R}^{3\times 2}$

$$G(x) = \begin{pmatrix} 5x & \sin(x) \\ 2 & x^3 + 2x^2 \\ x^2 + 3x & 1 \end{pmatrix}.$$

1.3. Consider two functions $\boldsymbol{f} : \mathbb{R} \longrightarrow \mathbb{R}^n$ and $\boldsymbol{g} : \mathbb{R} \longrightarrow \mathbb{R}^n$. Verify the general **sum rule** and **product rule** for these two functions.

2. Question: Derivatives of functions taking vectors as inputs *(elementary)*

2.1. Calculate the Jacobian matrix of the following two functions

(i) $F : \mathbb{R}^2 \to \mathbb{R}^3$ where:

$$F(x, y) = \begin{bmatrix} x^2 + \sin(x) \\ x(y-2) \\ y^2 - 3xy \end{bmatrix}$$

(ii) $G : \mathbb{R}^3 \to \mathbb{R}^2$ where:

$$G(x, y, z) = \begin{bmatrix} x^2 - y^2 \\ 3xyz - 5 \end{bmatrix}$$

2.2. Determine the gradient $\frac{\mathrm{d}\boldsymbol{f}}{\mathrm{d}\boldsymbol{x}}$ of the following function, where $M, N \in \mathbb{N}_{>0}$

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^M, \quad \boldsymbol{A} \in \mathbb{R}^{M\times N}, \quad \boldsymbol{x} \in \mathbb{R}^N.$$

2.3. Consider the function $h : \mathbb{R} \to \mathbb{R}, h(t) = (f \circ g)(t)$ with

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R} \\ g &: \mathbb{R} \to \mathbb{R}^2 \\ f(\boldsymbol{x}) &= \exp\left(x_1 x_2^2\right), \\ \boldsymbol{x} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = g(t) = \begin{bmatrix} t\cos t \\ t\sin t \end{bmatrix} \end{aligned}$$

and compute the gradient of h with respect to t.

2.4. Use the chain rule, **both according to Proposition 7.1 and according to Remark 7.1**, to find the gradient of

$$F : \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad (x, y, z) \mapsto f \circ \varphi\,(x, y, z)$$

for

$$\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad (x, y, z) \mapsto (h(x), g(x, y), z)$$

and scalar-valued functions f, g, and h defined as $f(x, y, z) := x^2 + yz$, $g(x, y) := y^3 + xy$, and $h(x) := \sin x$.

3. Question: Derivatives of functions taking matrices as inputs (*elementary*)

Note that the Booklet only contains instructions on taking the derivative of scalar-valued functions taking matrices as inputs (matrix norms being a common case). If you are interested in the derivation of vector and matrix valued functions taking matrices as indices, see Examples 5.12 *and* 5.13 *of* Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for Machine Learning

3.1. For a matrix $A \in \mathbb{R}^{m\times n}$, the *Frobenius norm* is defined as $\|\boldsymbol{X}\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$.

Calculate the gradient of the squared Frobenius norm, i.e. the function

$$f : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}, \quad \boldsymbol{X} \mapsto \|\boldsymbol{X}\|_F^2.$$

3.2. Prove the following identities

(i) $\nabla_{A^T} f(A) = (\nabla_A f(A))^T$, for a differentiable function $f : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}$, $m, n \in \mathbb{N}_{>0}$.

(ii) $\nabla_A \operatorname{tr}(AB) = B^T$.

4. Question: Directional derivative (*a bit more advanced*)

Evaluating partial derivatives only gives us the slope of a function in the direction of one of the inputs, or, equivalently, the direction of the corresponding canonical vector. (A canonical vector is a vector each of whose components are all zero, except one that equals 1.)

If we are interested in the slope of a function in the direction of a non-canonical vector, i.e. when changing several inputs at once, we can use the ***directional derivative****. The directional derivative of function f at $\mathbf{x}$ along $\mathbf{u}$ is defined as*

$$D_{\mathbf{u}} f(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}.$$

For differentiable functions f and unit vector $\mathbf{u}$, i.e. $\|\mathbf{u}\| = 1$, the directional derivative is simply computed as $D_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$.

4.1. Evaluate the directional derivative $D_{\mathbf{u}} f(\mathbf{x})$ for the following

(i) $f(x, y) = e^x \cos(\pi y)$, $\mathbf{x} = (0, -1)^\top$ and $\mathbf{u} = \left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^\top$.

(ii) $f(x, y) = xy^2 + x^3 y$, $\mathbf{x} = (4, -2)^\top$ and $\mathbf{u} = \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)^\top$

4.2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called homogeneous of degree m if $f(tx) = t^m f(x)$ for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. If f is differentiable, show that for $x \in \mathbb{R}^n$,

$$\nabla f(x)x = mf(x), \ \text{that is,} \ \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i} = mf(x).$$

Show that maps multilinear in k variables, which are characterized by the following property

$$\begin{aligned}&L\left(x_1,\ldots,x_{i-1},\alpha u+\beta w,x_{i+1},\ldots,x_n\right)\\=&\alpha L\left(x_1,\ldots,x_{i-1},u,x_{i+1},\ldots,x_n\right)+\beta\left(x_1,\ldots,x_{i-1},w,x_{i+1},\ldots,x_n\right)\end{aligned}$$

give rise to homogeneous functions of degree k. Give other examples.

If you have any questions or feedback, please feel free to contact me via E-mail at hannah.kuempel@stat.uni-muenchen.de!!

Also, thank you to the authors of the books *Mathematics for Machine Learning* as well as Steven J. Miller and Anthony Varilly whose exercises this sheet was inspired by.