UML2015
Exercise set 2
Solutions to be presented in the 27.3.2014 session

Exercise 1:
Suppose we wanted to construct algorithms that approximate the number $x = \sqrt{y}$, once a parameter $y > 0$ is given as input. Consider the functions

$$f_1(x,y) = (x^2 - y)^2 \quad \text{and} \quad f_2(x,y) = x^2 - y. \tag{1}$$

These functions have the properties $f_1(\sqrt{y}, y) = 0$ and $f_2(\sqrt{y}, y) = 0$.
1.1 Show that iterations

$$x_{n+1} = x_n - \mu \frac{\partial f_1(x_n, y)}{\partial x} \tag{2}$$

can produce a sequence $x_1, x_2, \ldots$ that converges to $\sqrt{y}$ at an exponential rate. Here μ is some real constant.
1.2 Show that iterations

$$x_{n+1} = x_n - \frac{f_2(x_n, y)}{\frac{\partial f_2(x_n, y)}{\partial x}} \tag{3}$$

can produce a sequence $x_1, x_2, \ldots$ that converges to $\sqrt{y}$ at a rate that is faster than exponential.
Advice: First solve explicit formulas for the partial derivatives. Then write $x_n = \sqrt{y} + \varepsilon_n$, and study approximative recursion formulas for the sequences $\varepsilon_1, \varepsilon_2, \ldots$. It is not necessary to prove any rigor asymptotics for this exercise. The essential will become clear when the recursion formulas for $\varepsilon_1, \varepsilon_2, \ldots$ are studied. Compare to Exercise 5 of Set 1.

Exercise 2:
2.1 Assume that $J : \mathbb{R}^N \to \mathbb{R}$ is some differentiable function. Its partial derivatives are defined as

$$\frac{\partial J(\mathbf{x})}{\partial x_i} = \lim_{\varepsilon \to 0} \frac{J(\mathbf{x} + \varepsilon \mathbf{e}_i) - J(\mathbf{x})}{\varepsilon}. \tag{4}$$

Assume that $\mathbf{x}$ is fixed, and that some vector $\mathbf{v}$ has the property

$$J(\mathbf{x} + \varepsilon \mathbf{h}) = J(\mathbf{x}) + \varepsilon \mathbf{v} \cdot \mathbf{h} + o(\varepsilon) \qquad \text{when} \quad \varepsilon \to 0 \tag{5}$$

for all vectors $\mathbf{h}$. Prove that $\mathbf{v}$ cannot be anything else than the $\nabla f(\mathbf{x})$. In other words $v_i = \frac{\partial f(\mathbf{x})}{\partial x_i}$ for all i.

2.2 Assume that $J : \mathbb{R}^{N\times N} \to \mathbb{R}$ is some differentiable function. Its partial derivatives are defined as

$$\frac{\partial J(A)}{\partial A_{ij}} = \lim_{\varepsilon\to 0} \frac{J(A + \varepsilon \mathbf{e}_i \mathbf{e}_j^T) - J(A)}{\varepsilon}. \tag{6}$$

Assume that A is fixed, and that some matrix B has the property

$$J(A + \varepsilon \mathbf{h}\mathbf{w}^T) = J(A) + \varepsilon \mathbf{h}^T B \mathbf{w} + o(\varepsilon) \qquad \text{when} \quad \varepsilon \to 0 \tag{7}$$

for all vectors $\mathbf{h}, \mathbf{w}$. Prove that B cannot be anything else than the $\nabla f(A)$. In other words $B_{ij} = \frac{\partial J(A)}{\partial A_{ij}}$ for all i, j.

2.3 Assume that matrices $A, B \in \mathbb{R}^{N\times N}$ and vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N^2}$ have the forms

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & \cdots & B_{1N} \\ \vdots & & \vdots \\ B_{N1} & \cdots & B_{NN} \end{pmatrix} \tag{8}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{N^2} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{N^2} \end{pmatrix}, \tag{9}$$

and also assume that they are related by the relations

$$A = \begin{pmatrix} x_1 & \cdots & x_N \\ x_{N+1} & \cdots & x_{2N} \\ \vdots & & \vdots \\ x_{N^2-N} & \cdots & x_{N^2} \end{pmatrix}, \quad B = \begin{pmatrix} y_1 & \cdots & y_N \\ y_{N+1} & \cdots & y_{2N} \\ \vdots & & \vdots \\ y_{N^2-N} & \cdots & y_{N^2} \end{pmatrix} \tag{10}$$

Prove that $\mathbf{y}^T\mathbf{x} = \mathrm{Tr}(B^T A)$.

2.4 Again assume that $J : \mathbb{R}^{N\times N} \to \mathbb{R}$ is some differentiable function, and that $\nabla J(A)$ is its gradient (as $N \times N$ matrix) at location A. Prove that

$$J(A + \varepsilon W) = J(A) + \varepsilon \mathrm{Tr}\big(W^T \nabla J(A)\big) + o(\varepsilon) \qquad \text{when} \quad \varepsilon \to 0 \tag{11}$$

holds for all $N \times N$ matrices W.

Advice: Notice that in 2.1 the claimed property holds for all $\mathbf{h}$, so we have freedom to choose all kinds of vectors for $\mathbf{h}$. When proving $v_i = \frac{\partial f(\mathbf{x})}{\partial x_i}$, first fix i, then choose suitable $\mathbf{h}$. The 2.2 must be done similarly as 2.1. First fix a pair (i, j), then choose $\mathbf{h}$ and $\mathbf{w}$ suitably. In 2.3 the trace is defined as $\mathrm{Tr}(C) = \sum_{n=1}^{N} C_{nn}$. When proving $\mathbf{y}^T\mathbf{x} = \mathrm{Tr}(B^T A)$, simply use the definition of trace, and then the definition of a matrix multiplication. It turns out that the way the vector elements are ordered into the matrix form doesn't matter, so don't be mislead by unnecessary details. In 2.4 transform the matrix function into a vector function, use vector calculus (which we can assume to be known), and transform the vector inner product into a matrix form. Finally, check that pieces fit together and stuff makes sense: Is the quantity $\mathbf{h}^T \nabla f(A)\mathbf{w}$ related to the quantity $\mathrm{Tr}\big(W^T \nabla f(A)\big)$?

Exercise 3:

3.1 Assume two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are in $\mathbb{R}^2$. Together they span a parallelogram. Assume it known, that the area of the parallelogram is the length of the base multiplied by the height, and prove that the area S of the parallelogram satisfies

$$S^2 = \|\mathbf{a}_1\|_2^2 \|\mathbf{a}_2\|_2^2 - (\mathbf{a}_2^T \mathbf{a}_1)^2 \tag{12}$$

3.2 Form the matrix $A = (\mathbf{a}_1, \mathbf{a}_2) \in \mathbb{R}^{2\times 2}$, and show that

$$S^2 = \det(A)^2 \tag{13}$$

3.3 Consider the linear transformation $\mathbf{y} = A\mathbf{x}$ where A is a 2×2 matrix. What kind of set is AU_v, where $U_v = [v_1, v_1 + \ell_1] \times [v_2, v_2 + \ell_2]$ is some rectangle? (AU_v means the image of U_v under the mapping A.) What is the area of AU_v?

3.4 Assume you had to integrate a function f over the set AU_v. Give an intuitive explanation why we have equality in the change of variables formula

$$\int_{AU_v} f(\mathbf{y}) dy = \int_{U_v} f(A\mathbf{x}) |\det(A)| dx \tag{14}$$

Advice: There is no one right way to do this exercise, but it is recommended that you draw some pictures. The integrals can be approximated as sums over some grid points.

Exercise 4:

4.1 Assume that A is $n \times n$ matrix, and that vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors of A with the same eigenvalue λ. (So that $A\mathbf{u}_i = \lambda \mathbf{u}_i$ for both $i = 1, 2$.) Prove that also $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2$ is an eigenvector of A for all $\alpha_1, \alpha_2 \in \mathbb{R}$.

4.2 Assume that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent eigenvectors of A, and form a matrix $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n) \in \mathbb{R}^{n\times n}$. Prove that $AU = U\Lambda$, where Λ is a diagonal matrix whose diagonal contains the eigenvalues of A.

4.3 Let's define a matrix $V = (U^{-1})^T$, and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ so that $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Prove the following formulas, and assume that the eigenvalues are non-zero, when needed.

$$
\begin{aligned}
A &= U\Lambda V^T \\
A &= \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^T \\
A^{-1} &= U\Lambda^{-1} V^T \\
A^{-1} &= \sum_{i=1}^{n} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T
\end{aligned}
\tag{15}
$$

Exercise 5:

5.1 Recall that a Gaussian random variable $X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2 has the density

$$
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2 \Big) \tag{16}
$$

Assume that $X_1, \ldots, X_N$ are iid (independent and indentically distributed) following a Gaussian distribution of mean μ and variance σ^2, and solve the likelihood $L(\mu, \sigma^2 | x_1, \ldots, x_N)$ and the log-likelihood $\ell(\mu, \sigma^2 | x_1, \ldots, x_N) = \log(L(\mu, \sigma^2 | x_1, \ldots, x_N))$.

5.2 Show that the estimates $\hat{\mu}$ and $\hat{\sigma}^2$ which maximize the likelihood are given by formulas

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{N} \sum_{n=1}^{N} x_N \\
\hat{\sigma}^2 &= \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{\mu})^2
\end{aligned}
\tag{17}
$$

Advice: Remember $p(\mathbf{x}) = p(x_1) \cdot \ldots \cdot p(x_N)$, and use zeros of derivatives for maxima. The study of second order derivatives can be omitted for simplicity, but you can examine them if you want. The parameter σ^2 is denoted in such way that it appears to be a square of some parameter σ, but you can consider σ^2 as a parameter of its own too, later defining $\sigma := \sqrt{\sigma^2}$. Calculating $\frac{\partial \ell}{\partial \sigma^2}$ is allowed.