UML2012
Exercise set 1
Solutions to be presented in the 23.3.2012 session

Exercise 1:
1.1 Given two vertical vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in $\mathbb{R}^n$, show that

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{a}_1 \\ \mathbf{u}_2 &= \mathbf{a}_2 - \frac{\mathbf{u}_1^T \mathbf{a}_2}{\|\mathbf{u}_1\|_2^2}\mathbf{u}_1 \end{aligned} \tag{1}$$

are orthogonal to each other. Furthermore, show that any linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ can be written in terms of $\mathbf{u}_1$ and $\mathbf{u}_2$.

1.2 Let's assume that $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$ are linearly independent vectors (so that $k \leq n$), and define vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ by formula

$$\mathbf{u}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} \frac{\mathbf{u}_j^T \mathbf{a}_i}{\|\mathbf{u}_j\|_2^2}\mathbf{u}_j \tag{2}$$

Prove that $\mathbf{u}_i$ are well defined, by showing that $\mathbf{u}_j \neq \mathbf{0}$ for all $j = 1, 2, \ldots, k$.

1.3 Prove that vectors $\mathbf{u}_i$ are orthogonal to each other (meaning $\mathbf{u}_i^T \mathbf{u}_j = \|\mathbf{u}_i\|_2^2 \delta_{ij}$). Also prove that any linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_k$ can be written in terms of $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

Exercise 2:
2.1 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 use exercise 1.1 to prove that the area S of the parallelogram satisfies

$$S^2 = \|\mathbf{a}_1\|_2^2 \|\mathbf{a}_2\|_2^2 - (\mathbf{a}_2^T \mathbf{a}_1)^2 \tag{3}$$

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

$$S^2 = \det(A)^2 \tag{4}$$

2.3 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?

2.4 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

$$\int_{AU_v} f(\mathbf{y})dy = \int_{U_v} f(A\mathbf{x})|\det(A)|dx \tag{5}$$

Exercise 3:

3.1 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 $\alpha_1, \alpha_2 \in \mathbb{R}$.

3.2 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.

3.3 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.

$$\begin{aligned} A &= U\Lambda V^T \\ A &= \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^T \\ A^{-1} &= U\Lambda^{-1}V^T \\ A^{-1} &= \sum_{i=1}^{n} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T \end{aligned} \tag{6}$$

Exercise 4:

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

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2 \Big) \tag{7}$$

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 $L(\mu, \sigma^2 | x_1, \ldots, x_N)$.

4.2 Calculate the log-likelihood $\ell(\mu, \sigma^2 | x_1, \ldots, x_N) = \log(L(\mu, \sigma^2 | x_1, \ldots, x_N))$.

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

$$\begin{aligned} \hat{\mu} &= \frac{1}{N} \sum_{n=1}^{N} x_N \\ \hat{\sigma}^2 &= \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{\mu})^2 \end{aligned} \tag{8}$$

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 $\frac{\partial}{\partial \sigma^2}\ell \ldots$ is allowed.

Exercise 5:
The gradient of a function $J : \mathbb{R}^n \to \mathbb{R}$ at point $\mathbf{w}$ is usually defined as

$$\nabla J(\mathbf{w}) = \begin{pmatrix} \frac{\partial J(\mathbf{w})}{\partial w_1} \\ \vdots \\ \frac{\partial J(\mathbf{w})}{\partial w_n} \end{pmatrix} \tag{9}$$

Alternatively, it can be defined as a vector with the property

$$J(\mathbf{w} + \varepsilon\mathbf{h}) = J(\mathbf{w}) + \varepsilon(\nabla J(\mathbf{w}))^T\mathbf{h} + o(\varepsilon) \qquad (\text{as } \varepsilon \to 0) \tag{10}$$

for all $\mathbf{h} \in \mathbb{R}^n$. Sometimes the second definition is more convenient because you can avoid multiplying out scalar products. Use either of the two definitions to find $\nabla J(\mathbf{w})$ in the following cases.

$$\begin{aligned} J(\mathbf{w}) &= \mathbf{a}^T\mathbf{w} \\ J(\mathbf{w}) &= \mathbf{w}^T A\mathbf{w} \\ J(\mathbf{w}) &= \mathbf{w}^T\mathbf{w} \\ J(\mathbf{w}) &= \|\mathbf{w}\| \\ J(\mathbf{w}) &= f(\|\mathbf{w}\|) \\ J(\mathbf{w}) &= f(\mathbf{a}^T\mathbf{w}) \end{aligned} \tag{11}$$

Here $\mathbf{a} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, and $f : \mathbb{R} \to \mathbb{R}$ is some differentiable function.