UML2012
Exercise set 5
Solutions to be presented in the 20.4.2012 session

**Exercise 1:**

Let's examine algorithms, which approximate a number $x = \sqrt{y}$, once a parameter $y > 0$ is given as input. The following functions provide some possible ways to achieve the goal:

$$f_1(x,y) = (x^2 - y)^2, \qquad f_2(x,y) = x^2 - y \tag{1}$$

**1.1** Explain why 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.
**1.2** Explain why 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: No need for rigor asymptotics or studies of domains of convergence. The claimed results can be justified reasonably with some approximations. You can write $x_n = \sqrt{y} + \varepsilon_n$, and study approximative recursion formulas for the sequence $\varepsilon_1, \varepsilon_2, \ldots$.

**Exercise 2:**

Suppose we want to maximize the function

$$f(\mathbf{w}) = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{w}^T \mathbf{x}(t))^4 \tag{4}$$

with a constraint $\|\mathbf{w}\| = 1$, by using the recursive formulas

$$\begin{aligned} \tilde{\mathbf{w}}_{n+1} &= \mathbf{w}_n + \mu \nabla f(\mathbf{w}_n) \\ \mathbf{w}_{n+1} &= \frac{\tilde{\mathbf{w}}_{n+1}}{\|\tilde{\mathbf{w}}_{n+1}\|} \end{aligned} \tag{5}$$

Let's assume that we can write $\mathbf{w}_n = \mathbf{w}_{\max} + \boldsymbol{\varepsilon}_n$, where $\boldsymbol{\varepsilon}_n$ are some vectors with small norms ($\|\boldsymbol{\varepsilon}_n\| \approx 0$), and that $\mathbf{w}_{\max}$ is a vector that maximizes the $f$ with the constraint $\|\mathbf{w}_{\max}\| = 1$.

**2.1** Prove that formula

$$\tilde{\mathbf{w}}_{n+1} = (1+\alpha)\mathbf{w}_{\max} + (\mathrm{id} + \mu Q)\boldsymbol{\varepsilon}_n + O(\|\boldsymbol{\varepsilon}_n\|^2) \tag{6}$$

holds where $\alpha = \mu\|\nabla f(\mathbf{w}_{\max})\|$ is a real coefficient, and $Q \in \mathbb{R}^{N\times N}$ is some matrix.

**2.2** Prove the formula

$$Q = \frac{12}{T}\sum_{t=1}^{T}\left(x(t)x(t)^T w_{\max} w_{\max}^T x(t)x(t)^T\right) \tag{7}$$

**2.3** Prove the formula

$$\|\tilde{\mathbf{w}}_{n+1}\| = (1+\alpha) + \mathbf{w}_{\max}^T(\mathrm{id} + \mu Q)\boldsymbol{\varepsilon}_n + O(\|\boldsymbol{\varepsilon}_n\|^2) \tag{8}$$

**2.4** Prove the formula

$$\mathbf{w}_{n+1} = \mathbf{w}_{\max} + P\frac{\mathrm{id} + \mu Q}{1+\alpha}\boldsymbol{\varepsilon}_n + O(\|\boldsymbol{\varepsilon}_n\|^2), \tag{9}$$

where $P$ is the projection matrix to the space $\langle \mathbf{w}_{\max}\rangle^{\perp}$.

**Exercise 3:**

**3.1** Assume that $X$ and $Y$ are independent random real variables such that $E(X) = 0$ and $E(Y) = 0$, and define a function

$$f_\alpha(U) = E(U^4) + \alpha(E(U^2))^2 \tag{10}$$

with arbitrary real coefficient $\alpha$. Solve a formula for the quantity

$$f_\alpha(X+Y) - f_\alpha(X) - f_\alpha(Y). \tag{11}$$

How does it depend on the variable $\alpha$?

**3.2** Assume that $S_1$ and $S_2$ are independent random real variables with $E(S_1) = 0$, $E(S_2) = 0$, $\mathrm{kurt}(S_1) > 0$ and $\mathrm{kurt}(S_2) > 0$, and that $S$ is a two component random variable

$$S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} \tag{12}$$

Define a matrix $W$ as a function of real parameter $\theta$ by

$$W(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{13}$$

and a function $f(\theta)$ by the formula

$$f(\theta) = \mathrm{kurt}((WS)_1) + \mathrm{kurt}((WS)_2). \tag{14}$$

Solve and explicit formula for $f$, that shows how it depends on $\theta$.

**Exercise 4:**

**4.1** The Hermite polynomials are special functions which look like this:

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ &\vdots \end{aligned} \tag{15}$$

They satisfy an orthogonality relation

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} H_n(x) H_m(x) dx = \sqrt{2\pi}\, n! \delta_{nm} \tag{16}$$

Assume that a function $f$ can be written as

$$f(x) = e^{-\frac{1}{2}x^2} \sum_{n=0}^{\infty} a_n H_n(x) \tag{17}$$

and solve a formula for each coefficient $a_n$ in terms of the function $f$. (Advice: Multiply $f(x)$ with $H_m(x)$, integrate over $x$, change the order of integral and sum.)

**4.2** The Hermite polynomials also satisfy a formal completeness relation

$$\sum_{n=0}^{\infty} \frac{1}{n!} H_n(x) H_n(x') = \sqrt{2\pi} \delta(x - x') e^{\frac{1}{2}x^2} \tag{18}$$

Prove, under the assumption that the formal completeness formula works, that any $f$ can be written as series like in (17). (Advice: Substitute your formula for coefficients $a_n$ into the series (17), and change the order of integral and sum.)

**4.3** Assume that we can approximate a probability distribution $p(x)$ of some random variable by formula

$$p(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Big(1 + \sum_{n=3}^{N} a_n H_n(x)\Big) \tag{19}$$

where the coefficients $a_n$ are small in some sense. Prove the approximation

$$-\int_{-\infty}^{\infty} p(x) \log\big(p(x)\big) dx \approx \log(\sqrt{2\pi}) + \frac{1}{2} - \sum_{n=3}^{N} n!\, a_n^2 \tag{20}$$

The left side is the definition of entropy. (Advice: Use approximation $\log(1+t) \approx t$. Also notice $x^2 = H_2(x) + H_0(x)$, and use orthogonality properties.)