UML2015
Exercise set 4
Solutions to be presented in the 17.4.2015 session

Exercise 1:

1.1 Assume that $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are two real random variables with zero mean, and that $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^K$ are sample vectors whose elements are independent observations of the random variables $\mathcal{Y}_1$ and $\mathcal{Y}_2$. Explain how the covariance between $\mathcal{Y}_1$ and $\mathcal{Y}_2$ is related to possible orthogonality of vectors $\mathbf{y}_1$ and $\mathbf{y}_2$.

1.2 Assume that X is an $N \times K$ data matrix, seen as a collection of K sample vectors. Assume that U is an $N \times M$ matrix, whose columns are normalized eigenvectors of $\frac{1}{K}XX^T$ corresponding to the most significant eigenvalues. We define $\overline{X} = U^T X$. Prove that the rows of $\overline{X}$ are orthogonal to each other.

Advice: By definition, the covariance deals with the quantity $\mathbb{E}(\mathcal{Y}_1\mathcal{Y}_2)$, while the orthogonality deals with the quantity $\mathbf{y}_1 \cdot \mathbf{y}_2$.

Exercise 2:

2.1 Assume that A is some $N \times M$ matrix, $M < N$, and that the columns of A are linearly independent. Prove that AA^T has M positive eigenvalues, and that zero appears as the eigenvalue $N - M$ times.

2.2 Assume that A is some $N \times N$ diagonal matrix, and that its diagonal values are all positive. If a matrix B has a property $B^2 = A$, we say that B is a square root of A. Solve how many different diagonal matrices B there are with this property.

2.3 Assume that A is some $N \times M$ matrix and $M \leq N$. Also assume that we have found a square root $(A^TA)^{\frac{1}{2}}$ which is symmetric and invertible. Prove that if we define $\overline{A}$ by

$$\overline{A} = A(A^TA)^{-\frac{1}{2}}, \tag{1}$$

the columns of $\overline{A}$ will be orthogonal.

2.4 Assume that A is some $N \times M$ matrix and $N \leq M$. This time assume that we have found a square root $(AA^T)^{\frac{1}{2}}$ which is symmetric and invertible. Prove that if we define $\overline{A}$ by

$$\overline{A} = (AA^T)^{-\frac{1}{2}}A, \tag{2}$$

the rows of $\overline{A}$ will be orthogonal.

Advice: 2.1 deals with basics of linear algebra, so there is perhaps no one right solution. One possibility is to notice that the claim is almost equivalent to $\dim \ker(AA^T) = N - M$, and aim for this. One should examine $\ker(A)$ and $\ker(A^T)$ separately first. At some point it will be reasonable to resort to some well known results. For example, see: http://en.wikipedia.org/wiki/Rank-nullity_theorem and

http://en.wikipedia.org/wiki/Orthogonal_complement
If 2.1 is too difficult, do 2.2, 2.3 and 2.4 anyway.

Exercise 3:

3.1 Assume that the covariance matrix of a random vector $\mathbf{x} \in \mathbb{R}^2$ with a zero mean has the form

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{3}$$

Solve the eigenvalues and vectors of the covariance as functions of ρ.

3.2 Assume that a random vector $\mathbf{y} \in \mathbb{R}^2$ has been defined by the formula

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix} \tag{4}$$

where $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent from other random variables, and have means and variances $\mathbb{E}(\mathcal{E}_i) = 0$ and $\mathbb{E}(\mathcal{E}_i^2) = \sigma_i^2$ for $i = 1, 2$, where σ_1^2 and σ_2^2 are some positive constants. Solve the covariance matrix for $\mathbf{y}$, and solve its eigenvalues.

3.3 Approximate the eigenvalues under the assumption that σ_1^2 and σ_2^2 are small, by using the first order Taylor series with respect to σ_1^2 and σ_2^2.

3.4 Under the approximation that σ_1^2 and σ_2^2 are small, approximate the eigenvectors of the covariance of $\mathbf{y}$.

Exercise 4:

The objective function for a single factor is

$$J_{ls}(\mathbf{a}) = \|C - \mathbf{a}\mathbf{a}^T\|^2, \tag{5}$$

where C is some covariance matrix, and the squared norm $\|A\|^2$ of a matrix A is defined as $\|A\|^2 = \sum_{ij} A_{ij}^2$.

4.1 Write down the details of how to get from Equation (5) to

$$J_{ls} = \|\mathbf{a}\|^4 - 2\mathbf{a}^T C\mathbf{a} + \mathrm{Tr}(CC^T) \tag{6}$$

4.2 Calculate the gradient of J_{ls}, as defined in Equation (6), with respect to $\mathbf{a}$.

4.3 Show that if the gradient is zero for some vector $\mathbf{v}$, then $\mathbf{v}$ is an eigenvector. Let $\mathbf{e}$ be an eigenvector of C with unit norm. Find all the scalars α such that the vector $\mathbf{a}^* = \alpha\mathbf{e}$ makes the gradient of J_{ls} zero.

4.4 Calculate the value of J_{ls} for $\mathbf{a}^*$. Conclude that the eigenvector $\mathbf{e}$ that has the eigenvalue with the largest norm minimizes J_{ls}.

Exercise 5:

The optimization problem for quartimax is to maximize

$$J(U) = \sum_{ij} G((AU)_{ij}) \tag{7}$$

under the constraint of orthogonality of U (see Section 5.5).

5.1 Derive an iterative update rule for the matrix U which solves the optimization problem for $G(y) = y^4$.

5.2 Show that $J(U)$ is constant for all orthogonal U if $G(y) = y^2$.