Exercise 1:
Kurtosis of a zero mean random variable $x$ is defined as

$$\text{kurt}(x) = \mathbb{E}(x^4) - 3(\mathbb{E}(x^2))^2$$  \hspace{1cm} (1)$$

Kurtosis is a measure of the "peakedness" of the probability distribution of $x$.
Calculate the kurtosis for the

1.1 Uniform distribution $p(x)$,

$$p(x) = \begin{cases} 
\frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\
0 & \text{else.}
\end{cases}$$  \hspace{1cm} (2)$$

1.2 Laplacian distribution $p(x)$,

$$p(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|).$$  \hspace{1cm} (3)$$

1.3 Gaussian distribution $p(x)$ with mean zero and variance $\sigma^2$,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$  \hspace{1cm} (4)$$

1.4 Calculate the kurtosis for the following mixture of Gaussians (called a
Gaussian scale mixture)

$$p(x) = \frac{1}{2}(p_1(x) + p_2(x)),$$  \hspace{1cm} (5)$$

where

$$p_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{x^2}{2\sigma_i^2}\right).$$  \hspace{1cm} (6)$$

Show that the kurtosis is always $> 0$ if $\sigma_1 \neq \sigma_2$.

1.5 Consider now the following mixture of Gaussians of the same variance but
different means:

$$p(y) = \frac{1}{3}(p_{\mu}(y) + p_0(y) + p_{-\mu}(y))$$  \hspace{1cm} (7)$$

where

$$p_\alpha(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y - \alpha)^2}{2}\right).$$  \hspace{1cm} (8)$$
Calculate the kurtosis and show that it is always negative for nonzero mean. You can use the fact that the normal distribution has skewness of zero, i.e. \( E(u^3) = 0 \).

(Hint: You might want to use that \( E(u^2) = V(u) + E(u)^2 \). Also: \( (a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4 \).)

1.6 Let \( u = x + \alpha y \), where \( x \) follows the distribution in Equation (5), and \( y \) has the distribution in Equation (7). Furthermore, \( x \) and \( y \) are independent. How can you choose \( \alpha \in \mathbb{R} \) so that \( \text{kurt}(u) = 0 \)?

Exercise 2:
For a zero-mean random variable, skewness of a distribution is defined to be its third moment, i.e.

\[
\text{skew}(x) = E(x^3). \tag{9}
\]

It measures the asymmetry of a distribution. If the independent variables \( s \) have a highly asymmetric distribution, skewness can be used to perform ICA.

Suppose \( Z \) is \( N \times K \) data matrix, and denote as \( z_k \) the columns of \( Z \) with each fixed \( 1 \leq k \leq K \). We would like to maximize

\[
J(w) = \frac{1}{K} \sum_{k=1}^{K} (w \cdot z_k)^3 \tag{10}
\]

under the constraint that \( \|w\| = 1 \).

2.1 Find the gradient \( \nabla J(w) \).

2.2 What is the gradient-ascent optimization iteration, considering the constraint \( \|w\|? \)

2.3 Take the limit of large stepsizes, i.e. \( \mu \to \infty \). What is the optimization iteration now?

Exercise 3:
Assume the data \( z_1, \ldots, z_K \) is iid and follows the model \( z = As \), where \( z \in \mathbb{R}^N \) is white random vector, \( A \) is orthonormal, i.e. \( A^T A = I \), and the \( s_n \) are independent random variables.

3.1 Write down the log-likelihood \( \ell(A|z_1, \ldots, z_K) \) of \( A \) in terms of the distribution \( p_s(s) \) which may be arbitrary.

3.2 Show that the log-likelihood does not depend anymore on the matrix \( A \) if the distribution of \( s_n \) are Gaussian.

Exercise 4:
In the maximum likelihood estimation of the ICA model, we may not know the densities of the independent variables \( s \). Therefore, they must be approximated in one way or the other.
We have seen in the lecture that as long the approximation \( \tilde{p}_i(s_i) \) fulfills

\[
E(s_i g_i(s_i) - g'_i(s_i)) > 0 \quad (11)
\]

for all \( i \), where \( g_i = \tilde{p}_i'/\tilde{p}_i \), maximization of the likelihood will lead to the right solution for the mixing matrix \( B \) (see Theorem 1 on page 66).

**4.1** Assume that \( s_i \) is Gaussian (zero mean, unit variance). Is the condition in Eq (11) fulfilled? (Advise: Examine the quantities \( E(s_i g_i(s_i)) \) and \( E(g'_i(s_i)) \) with integration by parts. Assume that \( g_i \) grow slower than \( \exp(s_i^2/2) \).)

**4.2** Suppose you make the choice \( g_i(s_i) = s_i^3 \). To what does the condition in Eq (11) correspond to? (Assume that \( s_i \) is zero mean and normalized to unit variance.)

**4.3** Show that making the choice \( g_i(s_i) = -s_i \) corresponds to \( \tilde{p}_i \) being a Gaussian distribution.