ICA2: Noise & Kurtosis

The first exercise discuss some properties of Infomax-ICA. Then in the second problem the moments of some popular distributions are to be calculated. The third exercise illustrates why (approximately) independent components can be obtained by maximizing *"non-Gaussianity"*. The task requires to compute the *kurtosis* for different distributions of toy data.

6.1 Natural Gradient (3 points)

(a) Extend your code from the previous problem sheet to get an ICA-learning scheme based on the natural gradient with a learning rate ε that decays slowly to 0 (e.g. $\varepsilon_{t+1} = \lambda\varepsilon_t$ with $\lambda \approx 1, \lambda < 1$). Note that depending on λ you have to iterate over the (shuffled) data more than once for proper convergence.

(b) Use the two sound signals from the last problem sheet and add (as third source s_3) an additional "noise" source (normally distributed random numbers with a standard deviation similar to the two signals). Mix the signals using a mixing matrix of your choice and apply your ICA-algorithm. Plot the Mixed Sounds and recovered Sources

(c) Do the same analysis but adding a different "noise"-source (e.g. Laplace distributed) instead of the normal one.

6.2 Moments of univariate distributions (3 points)

Calculate the first 4 moments of the different random variables depending on the respective parameters. In addition to providing the derivation (e.g. by using the characteristic function) fill the following table:

	Laplace (μ, b)	Gauß (μ, σ)	Uniform (a, b)
mean: first moment $\langle X \rangle$			
variance: second centered moment $\langle X^2 \rangle_c$			
skewness: third standardized moment $\langle X^3 \rangle_s$			
kurtosis: fourth standardized moment $\langle X^4 \rangle_s$			

The i-th centered moment is defined by $\langle X^i \rangle_c = \left\langle (X - \langle X \rangle)^i \right\rangle$ and the standardized one by $\langle X^i \rangle_s = \dfrac{\langle X^i \rangle_c}{\langle X^2 \rangle_c^{i/2}}$.

6.3 Kurtosis of Toy Data (4 points)

The file `distrib.mat` contains three toy datasets (`uniform`, `normal`, `laplacian`), each 10000 samples of 2 sources. Do the following for each dataset (which can be read for example using Python with `loadmat` from `scipy.io`):

(a) Apply the following mixing matrix $\mathbf{A}$ to the original data $\mathbf{s}$:

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} \\ \mathbf{x} &= \mathbf{A}\mathbf{s}. \end{aligned}$$

(b) Center the mixed data to zero mean.

(c) Decorrelate the data by applying principal component analysis (PCA) and project them onto the principal components (PCs).

(d) Scale the data to unit variance in each PC direction (now the data is *whitened* or *sphered*).

(e) Rotate the data by different angles θ

$$\begin{aligned} \mathbf{x}_\theta &= \mathbf{R}_\theta \mathbf{x} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \mathbf{x} \\ \theta &= 0, \frac{\pi}{50}, \ldots, 2\pi, \end{aligned}$$

and calculate the kurtosis[1] empirically for each dimension:

$$\operatorname{kurt}(x_\theta) = \langle x_\theta^4 \rangle - 3\underbrace{\langle x_\theta^2 \rangle}_{=1}^2.$$

(f) Find the minimum and maximum kurtosis value for the first dimension and rotate the data accordingly.

- Plot the original dataset (sources) and the mixed dataset after the steps (a), (b), (c), (d), and (f) as a scatter plot and display the respective marginal histograms. For step (e) plot the kurtosis value as a function of angle for each dimension.
- Compare the histograms after rotation by θ_{min} and θ_{max} for the different distributions.

[1] In this exercise and in the script a different notion of Kurtosis is used in comparison with the previous problem. Here and in the lecture notes the so-called *excess* Kurtosis is used which yields a value of 0 for normally distributed random variables. Additionally in this definition no normalization by the standard deviation is applied but this is at least for whitened data not of any relevance.