Sheet IV: Matrix Decompositions

S. Wein and E. W. Lang
Computational Intelligence and Machine Learning Group,
Biophysics, University of Regensburg, 93040 Regensburg, Germany

February 10, 2021

1 Cocktail party for two

First plot a few timepoints of the mixed signals in *toydata.txt* and *audiodata.txt*, which you find in the attached data folder on the GRIPS server. Also plot the joint 2-dimensional distributions of the mixed signals.

- Now use principal component analysis (PCA) to project the two datasets onto basis systems that maximize the variance of the projections. Respectively use independent component analysis (ICA) to project the data on basis systems that maximize for statistical independence.
- Then plot the projected signals, as well as their joint distributions. Compare the projections between PCA and ICA and comment on their differences.

2 Ghostfaces

Download the cropped version of the *labeled faces in the wild dataset* (conradsanderson.id.au/lfwcrop/lfwcrop_grey.zip), and use a subset of 1000 faces for this exercise.

- First plot 4×5 original faces, and plot also the mean face over 1000 images. Reshape the image data into a matrix $X \in \mathbb{R}^{N \times P}$, with number of images $N = 1000$ and number of pixels $P = 64^2 = 4096$. Now use non negative matrix factorization (NMF) to approximate the data with two factor matrices $X \approx W \cdot H$, with $W \in \mathbb{R}^{N \times K}$ and $H \in \mathbb{R}^{K \times P}$. Then set $K = 20$, reshape the components in H back into $K = 20$ images, and plot them.
- Apply the same procedure employing PCA and ICA, in order to obtain images with maximal variance and maximal statistical independence. Compare the images to those generated with NMF and comment on your results.

3 Empirical Mode Decomposition

Now we consider a case where we observe 2 signals, which consist of a mixture of 3 sine waves with different frequencies. Plot the first 500 points of the two signals, which you can find in *ex3_signals.txt*.

- First apply ICA and estimate 2 source components. Plot the 2 extracted independent components (ICs).
- Next use empirical mode decomposition (EMD) and decompose one of the 2 signals into intrinsic mode functions (IMFs). Plot the IMFs and compare them with the estimated ICs.
- Now add small Gaussian noise with 0 mean and a standard deviation/scale of 0.1 to the first signal. Apply EMD to the noisy signal and plot the obtained IMFs. Then apply an ensemble EMD (EEMD) and try to vary the noise width used in the EEMD algorithm in order to compensate the noise on the signal. Comment on your results.

Note:

Here are some useful tools:

- https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html#sklearn.decomposition.PCA
- https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html#sklearn.decomposition.FastICA
- If you are using IPython, you can also listen the the audiodata with:
  $\textbf{from} \quad IPython.display \quad \textbf{import} \quad Audio$
  $Audio(yoursignal,\ rate = 9000)$
- https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html#sklearn.decomposition.NMF
- https://pyemd.readthedocs.io/en/latest/eemd.html
- https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html

Please send your report, as well as those from the other exercises, to either simon.wein@ur.de or elmar.lang@ur.de (ideally until the end of March).