Homework 2: Independent Component Analysis

Jeremy Hintz

February 2016

1 Introduction

Independent Component Analysis (ICA) is a technique used to break down multivariate signals into independent non-Gaussian signals. It falls under the umbrella of Blind Source Separation and has been shown to be very effective for many applications. For this project, we used ICA to separate signals from short sound clips.

2 Mathematical Background

We start off with an $n \times t$ matrix U, which is the original source data. We then mix this data by operating on U with a matrix A. From there our task is to find a matrix W that recovers the source data from the mixed observations.

In this project, we used the batch technique described in the project description and lecture.

$$\Delta W = \eta(tI + (1 - 2Z)Y^T)W \tag{1}$$

One thing to note is that there is another method which was brought up in class that uses the inverse transpose of W:

$$\Delta W = \eta((1 - 2Z)x^T + (W^T)^{-1}) \tag{2}$$

where x represents an observation at a given point in time. In order to make 3 more comparable to 1, we can alter it to reflect the batch processing paradigm:

$$\Delta W = \eta((1-2Z)X^T + t(W^T)^{-1}) \tag{3}$$

The thing we must note with regards to the different forms of our equations is that they are only equivalent under the assumption that W is orthogonal. In this case, $W^T W = I$. Clearly, this assumption is not always valid. Thus, Equation 1 is not purely accurate. However, for the purposes of this report, since Equation 1 does not contain an inverse operation, we will tend to prefer it for its relative computational efficiency. In the following results, Equation 1, found in `hw2ICA.m` will be used, while Equation 3 can be found and experimented with in `hw2ICAv2.m`,

3 Model Evaluation

3.1 Running the Algorithm on the Test Set

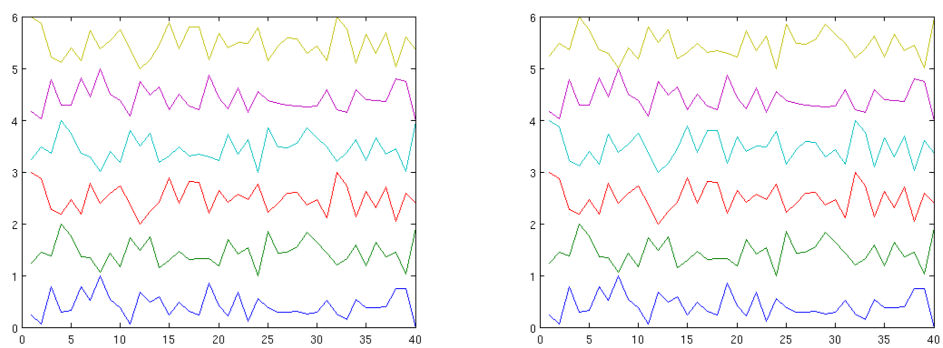

Figure 1: Two experiments showing original (bottom) and reconstructed (top) signals from `icaTest.mat` with $\eta = 0.01$ and $iterations = 1000$.

	Recon. 1	Recon. 2	Recon. 3
Source 1	-0.480	**0.992**	-0.443
Source 2	**0.994**	-0.386	-0.555
Source 3	-0.483	-0.501	**0.994**

Table 1: Correlations for source and reconstructed signals. Trial 1.

	Recon. 1	Recon. 2	Recon. 3
Source 1	-0.423	**0.990**	-0.495
Source 2	-0.544	-0.397	**0.993**
Source 3	**0.992**	-0.503	-0.481

Table 2: Correlations for source and reconstructed signals. Trial 2.

Along with `sounds.mat`, we are given a smaller set `icaTest.mat`. It will be useful for us to begin our analysis by examining the results of operating on this test set with the parameters $\eta = 0.01$ and $iterations = 1000$. The results are in shown in Figure 1.

I ran this trial two times and normalized the results on the interval $[0, 1]$. Because of this, scaling is possible and the signal could theoretically be flipped, although this never happened while I was running the experiments. We are looking for high correlations between source data and the corresponding reconstructed signals, which is the case for results close to either 1 or -1. When observing our results shown in Table 1 and 2, we observe that our algorithm generates signals with a correlation of more than 0.99.

3.2 Running the Algorithm on Sound Data

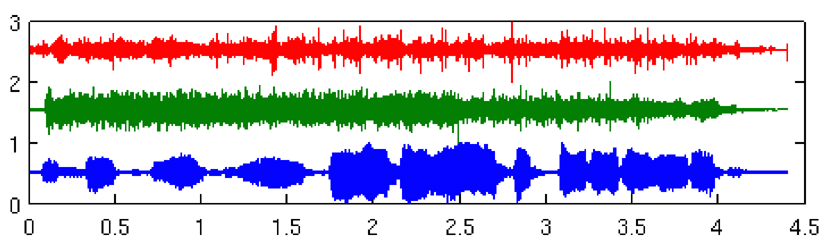

Figure 2: Original signals from `sounds.mat`

Next, for our primary experiment, we wish to attempt Blind Source Separation on the data found in `sounds.mat`. We follow a methodology similar to that found in the project description and in lecture, generating a Matrix A from random numbers from 0 to 1. Our parameters are $\eta = 10^{-6}$ and $iterations = 200$. The scaling is the same as in the small data set. The results shown in Table 3 again show a very high correlation. In order to visualize the results, see the plots in Figure 2 and 3.

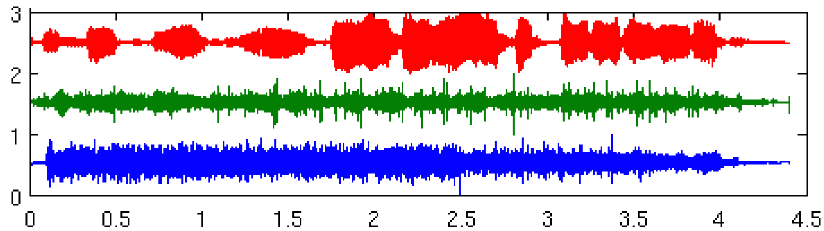

Figure 3: Reconstructed signals with $\eta = 10^{-6}$ $iterations = 200$.

	Recon. 1	Recon. 2	Recon. 3	Recon. 4	Recon. 5
Source 1	-0.003	0.003	-0.007	-0.003	**1.000**
Source 2	-0.001	-0.001	**0.999**	-0.013	0.002
Source 3	-0.003	-0.010	0.012	**0.999**	0.001
Source 4	0.002	**0.999**	0.008	-0.010	0.004
Source 5	**0.999**	0.005	0.002	0.013	-0.002

Table 3: Correlation between source and reconstructed signals. Sound data.

4 Conclusion

In this paper, we got a glimpse into the potential of using Independent Component Analysis for Blind Source Separation by running experiments on 44,000 short sound clips. The correlations we observed were very high and even with a large number of samples, processing them was achievable with relative efficiency.