Independent Component Analysis report

DIYUAN DAI, EID: DD33653,

Department of Computer Science, The University of Texas at Austin

1 1. INTRODUCTION

ICA, also known as Blind source separation (BSS), assumes that the observed random signal X obeys a model in which S is an unknown source signal whose components are independent of each other and A is an unknown mixing matrix. The purpose of ICA is to estimate the mixing matrix A and the source signal S by only observing x.

ICA considers that a signal can be decomposed into linear combinations of several statistically independent components, with the latter carrying more information. We can show that this decomposition is unique as long as the source signal is non-Gaussian.

Uncorrelation is a kind of weak linear independence and generally has to be used with other properties to produce relatively unique results. In PCA, we combine the uncorrelation and maximum variance of data. In general, linear independence is much stronger than uncorrelation, which is strong enough to recover data sources under certain conditions.

This project is a handwritten data separator using Independent Component Analysis (ICA). It separates the chosen sound source from randomly mixed sound data.

2 METHOD

2.1 Principles of ICA algorithm

- Firstly, a set of recordings (each length of 44000) from sound samples is chosen and transformed into a 2D matrix of shape (n * 44000). Each row of the matrix represents one recording of sound.
- Then, apply a linear mixing to the selected samples by initializing A matrix randomly and do a multiplication to the source data.
- Starting the algorithm with a initialization of a weight matrix W. By calculating Calculate $Y = WX$, where Y is our current estimate of the source signals. We have obtained an m by t matrix X of m mixed signals ($m >= n$) of length t that consist of different linear mixtures of U, where t is the length of source, 44000.
- To help us traverse the gradient of maximum information separation, we need to calculate the sigmoid value of Y, matrix Z. Sigmoid function:

  $$Z = 1/(1 + e^{-y})$$

  Then we implement the descending gradient strategy to find the convergence of estimate:

  $$\Delta W = \alpha(I + (1 - 2Z)Y^T)W$$

  where α is the learning rate, I is the identity matrix of size n.
- Finally, we update W by adding ΔW to it until convergence.

2.2 Measure of convergence

- According to the theory of descending gradient strategy, ΔW will be continually decreasing until a point.

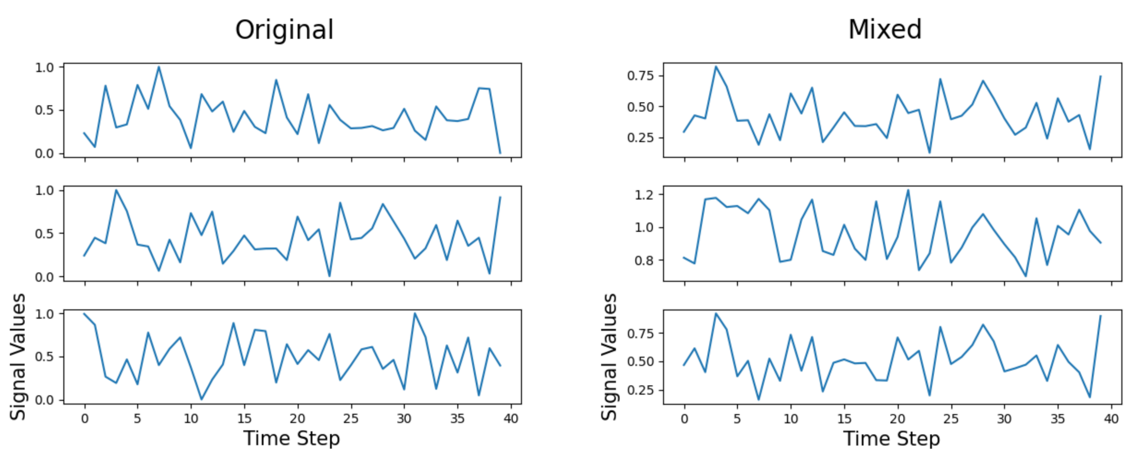
- We can track on the norm of W and pay attention on the change of norm after each iteration, in this implication, a result is considered convergent once the norm of ΔW is less than 1/1000 of last W.

2.3 Quantitative measure of accuracy of recovery

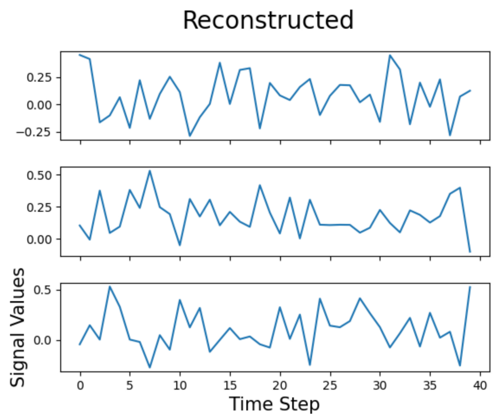
- Firstly, we can simply observe the shape of output signals and comparing them to the original. Although it is not a quantitative measure but it is simple and effective way to judge whether the ICA algorithm is implemented right at the very beginning of the environment.
- In this lab, Pearson product-moment correlation coefficients between reconstructed data and source are calculated to give out a quantitative presentation of how good the reconstructed sound is.

3 RESULTS

3.1 Small sample test



Figures shows the original signals from small sample and the mixed data after using A matrix provided in data set.

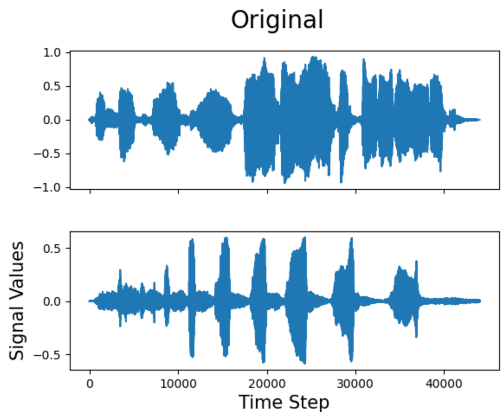


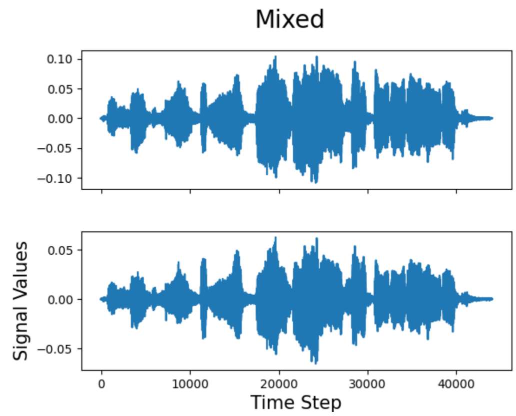
-0.6556813790940755	0.9808799743048849	-0.550103691124151
-0.311915549533343	-0.589701845861982	0.9843240858942878
0.9585140262753442	-0.26624178312631497	-0.4276528036415969

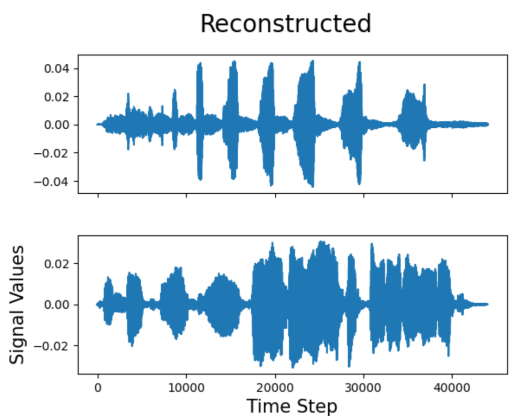
This figure shows the reconstructed signals and the correlation table, We can obviously find that the model can reconstruct recognizable signals. This result is given out under a learning rate 0.1, $m * n = 3 * 3$, 10000 epochs.

3.2 Big sample test, mixing two of source

In this session, we choose first source and fourth source to conduct the experiment. It took much longer time to obtain a convergence under such a 2 * 44000 data set. However, since there are only 2 sources, the reconstruction performs well.



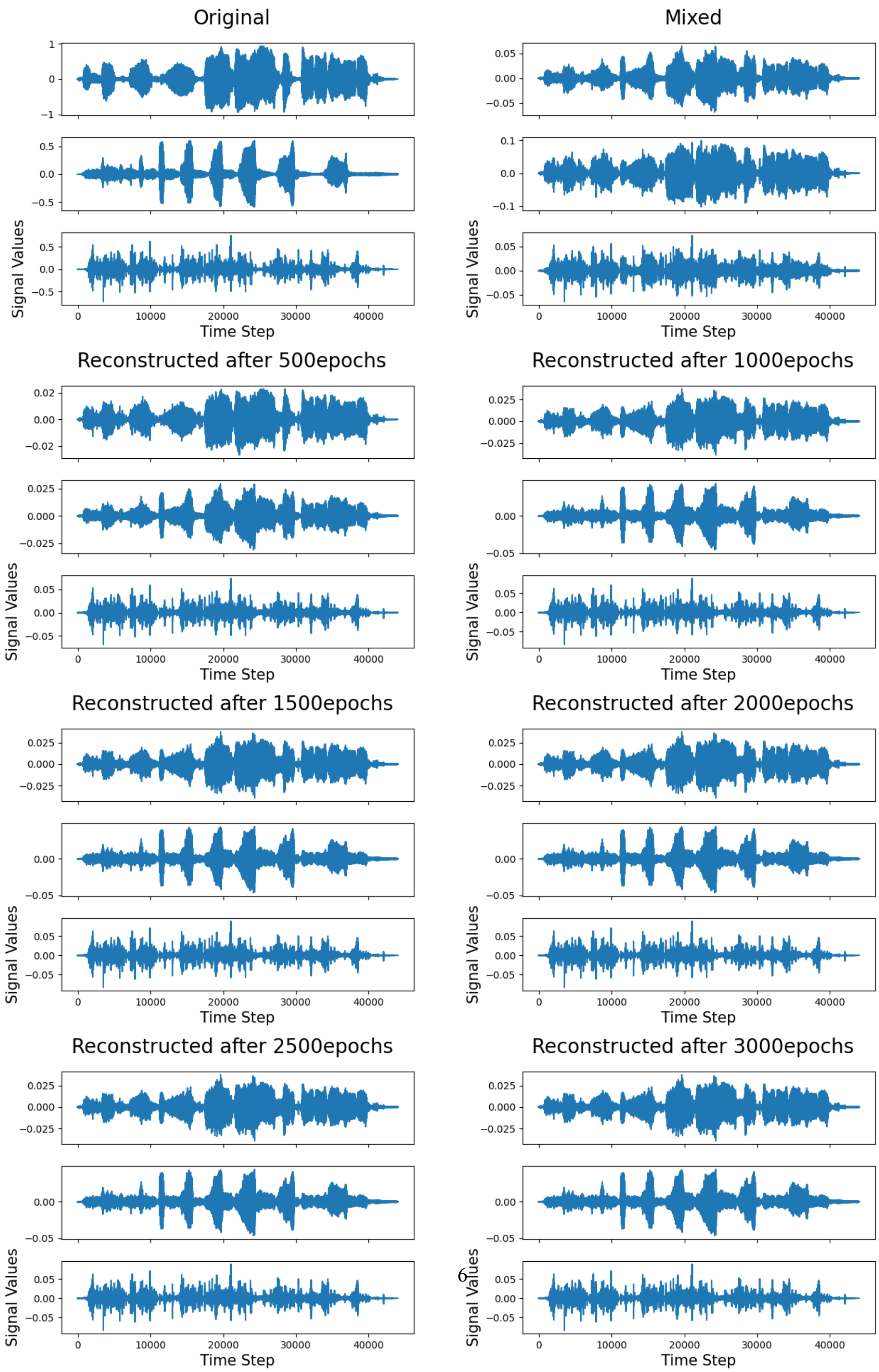




0.46518245	0.98752016
0.99077593	-0.4603333

3.3 Big sample test, mixing three of source

In this session, we choose first source, fourth and fifth source to conduct the experiment. And here are the shapes reconstructed signals against epoch. We can see clearly the sources are gradually separated and the tendency of changing become much smaller after a certain amount of epochs, when reach convergence.


Original
Mixed
Reconstructed after 500epochs
Reconstructed after 1000epochs
Reconstructed after 1500epochs
Reconstructed after 2000epochs
Reconstructed after 2500epochs
Reconstructed after 3000epochs
Signal Values
Time Step

Followed by the corresponding correlation tables and W after 500 epochs and 3000 epochs.

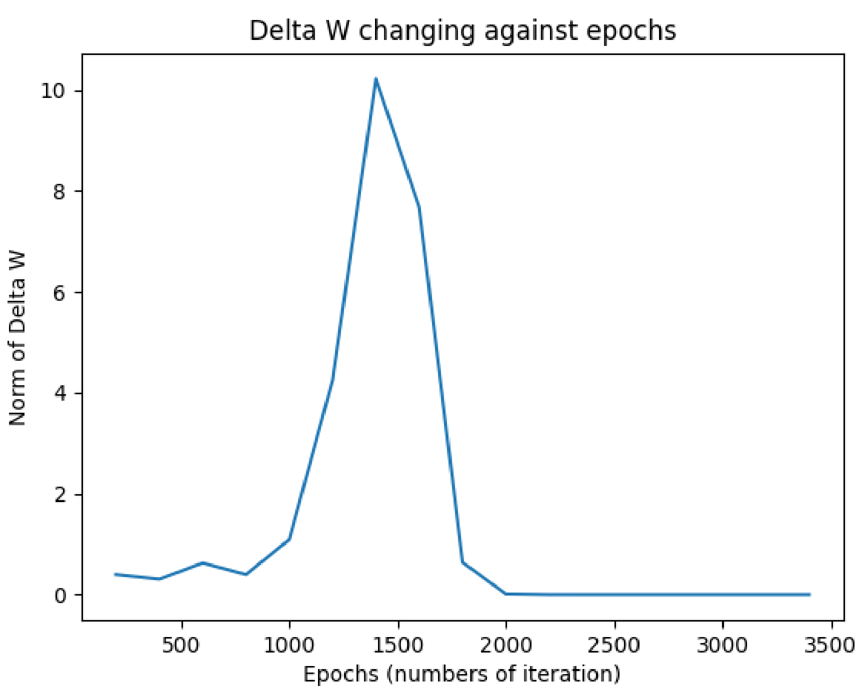
```
correlation after 500epochs
[[ 0.97763043  0.75023254  0.19929938]
 [-0.00752068  0.66424581  0.18282761]
 [-0.21374505  0.03758542  0.96307234]]
100%|██████████| 500/500 [00:01<00:00, 319.26it/s]


w =
[[-1.46746080e-04  9.10375875e-05 -5.73728694e-05]
 [ 1.72401797e-04 -1.06788456e-04  6.68624020e-05]
 [-5.54816025e-05  3.41116147e-05 -2.06585871e-05]]
```

```
correlation after 3000epochs
[[ 0.90559267  0.38745974  0.17249049]
 [-0.39244042  0.9193918   0.02693775]
 [-0.15364559 -0.08895145  0.98411796]]
100%|██████████| 500/500 [00:01<00:00, 323.28it/s]


w =
[[ 1.99392572e-08 -8.13696627e-09  5.25095815e-08]
 [-7.28686529e-09 -2.11153107e-08  6.60060531e-08]
 [-1.71737483e-08 -1.26363130e-08  2.42546855e-08]]
```

In this experiment, Norm of W is changing with time dramatically. We can see the convergence is reached after 2000 epochs.



4 SUMMARY

According to the above explanation of the mathematical principles of ICA, we can understand some of the capabilities and limitations of ICA.

From the familiar sample-feature perspective, we use ICA with the prerequisite that the sample data is considered to be generated by independent non-Gaussian distributed implied factors, with the number of implied factors equal to the number of features. It is more suitable for signal reduction (since signals are more regular and often not Gaussian distributed).

ICA is a powerful method in the field of blind signal analysis and a method for finding the implied factors of non-Gaussian distributed data.

I come up with a scenario where ICA come into real world application:
If we replace the microphone with an electrode that captures brain waves, the signal source S could represent the brain independent processes, heartbeat, blink, etc. By subtracting the signal x from the useless signals such as heartbeat, blink using ICA, we could get the Independent internal brain signals.