Speech Separation

Name: Kunal Lad
UT Eid: KL28697
Email: kunal.lad@utexas.edu

Introduction

The goal of this assignment is to use Independent Component Analysis technique for Blind Source Separation (segregating a mixture of speech signals into individual signals). This technique makes some assumptions about the data and hence only works well if these assumptions are true. It assumes the following:

- Signals are independent
- Signals are non gaussian signals
- Observed signals are linear combination of original signals

Even after these assumptions it useful in a number of applications. Most of the applications where ICA is used are generally highly underdetermined (i.e. number of observed samples << number of original signals) [1]

In this assignment we will be use this segregating sounds. We will be combining the signals ourselves to produce the mixed signals. However in real world applications the captured mixed signals are more complex since there are time delays and echoes. But generally these factors are ignored.

Methodology

Description

We start with an initial guess of unmixing matrix W and compute initial estimate of signals Y = W X. Then we calculate the sigmoid matrix Z from Y. Next we find the error matrix ΔW

using our current guess W, learning rate **η**, Y and Z. Finally we update W to $W + \Delta W$
We continue this process until the error converges.

Algorithm

1. Assume $\mathbf{X} = \mathbf{AU}$.
2. Initialize the ($\mathbf{n}$ by $\mathbf{m}$) matrix $\mathbf{W}$ with small random values.
3. Calculate $\mathbf{Y} = \mathbf{WX}$.
   $\mathbf{Y}$ is our current estimate of the source signals.
4. Calculate $\mathbf{Z}$ where $\mathbf{z_{i,j}} = g(\mathbf{y_{i,j}}) = \mathbf{1/(1+e^{-y_{i,j}})}$ for $\mathbf{i} \in [\mathbf{1..n}]$ and $\mathbf{j} \in [\mathbf{1..t}]$ (where $\mathbf{t}$ is the length of the signals).
   This helps us traverse the gradient of maximum information separation.
5. Find $\mathbf{\Delta W} = \boldsymbol{\eta}(\mathbf{I} + (\mathbf{1\text{-}2Z})\mathbf{Y^T})\mathbf{W}$ where $\boldsymbol{\eta}$ is a small learning rate.
6. Update $\mathbf{W} = \mathbf{W} + \mathbf{\Delta W}$ and repeat from step 3 until convergence or R_max iterations (you get bored and decide it is done).

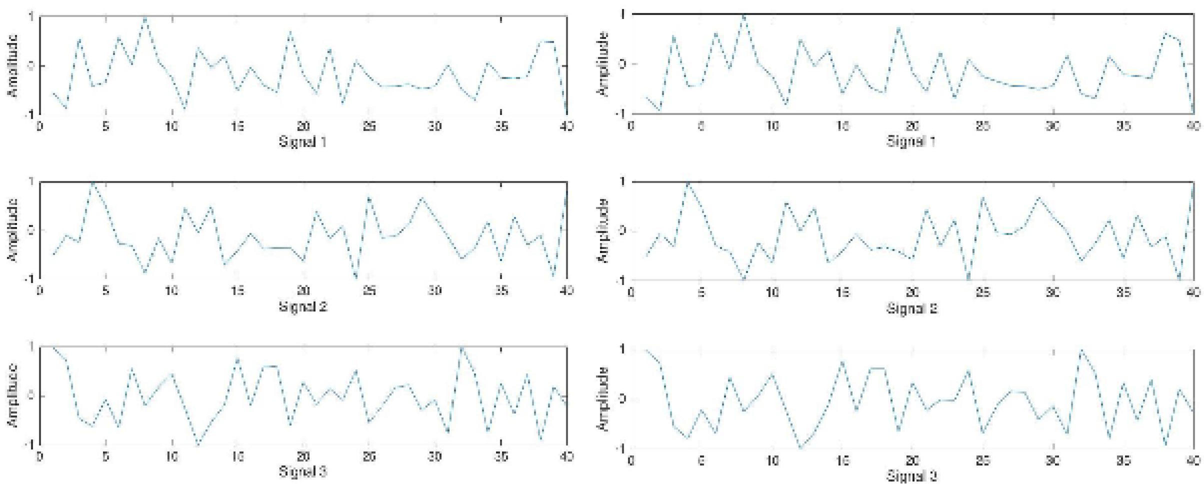
Experiments

In this section we briefly present the experiments performed and give some analysis of obtained results. For all the experiments mentioned in this section we use the following unless specified otherwise:

- Use number of samples equal to number of signals
- We use learning rate of 0.01 and 10000 iterations
- We use a random matrix with small values in the range [0, 0.1] for combining the signals to get the observed data.
- Rescale signals to lie in [0, 1] range for displaying output plots.

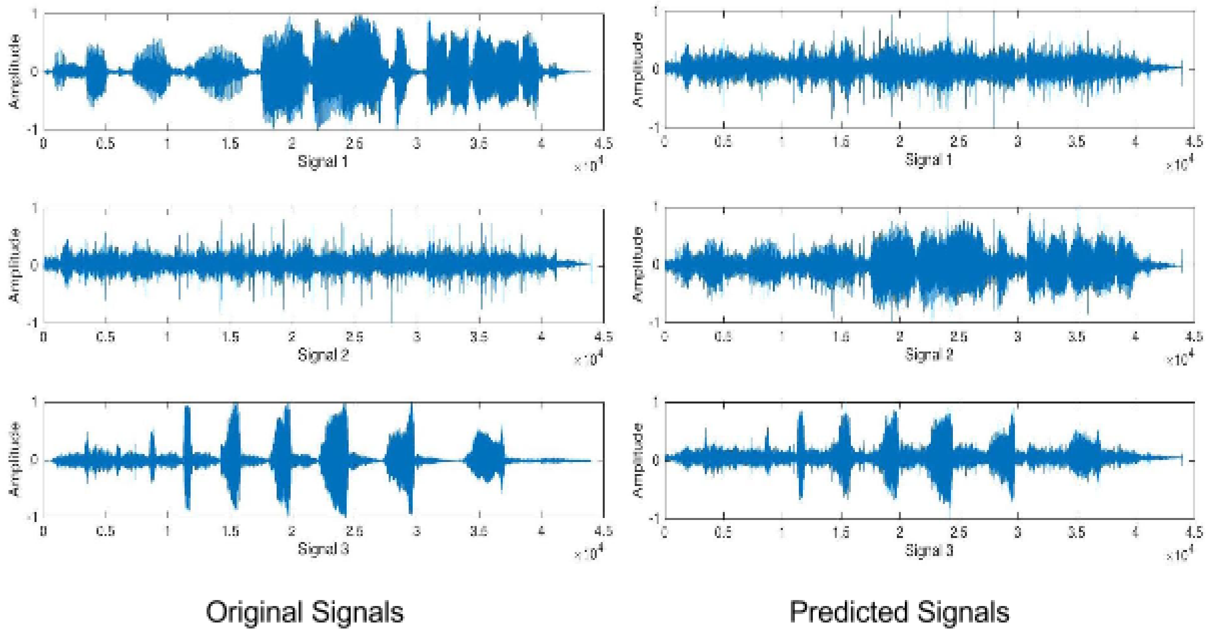
NOTE : Signals are not plotted in the order of match.

E1: Performance variation with size of data set



Original Signals Predicted Signals

ICAtest Data

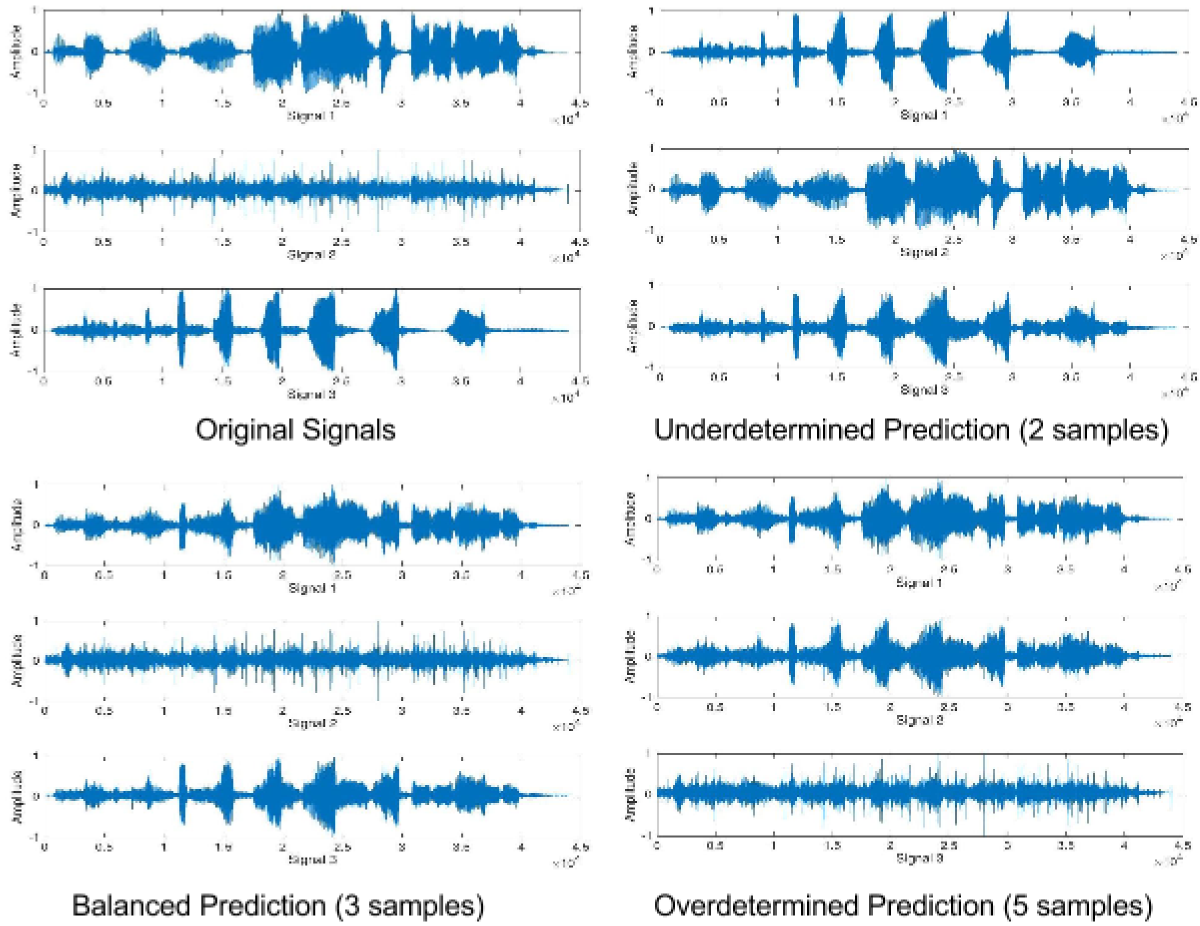


Sounds Test Data

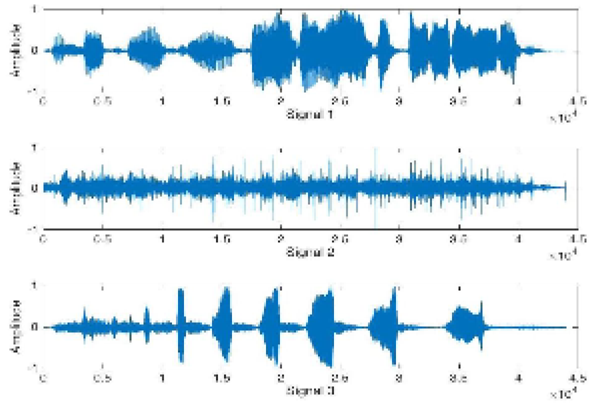
We observe that algorithm performs extremely well for smaller dataset (almost retrieves back the original signals) and reasonably well for larger data sets.

E2: Performance comparison for underdetermined, overdetermined and balanced system.

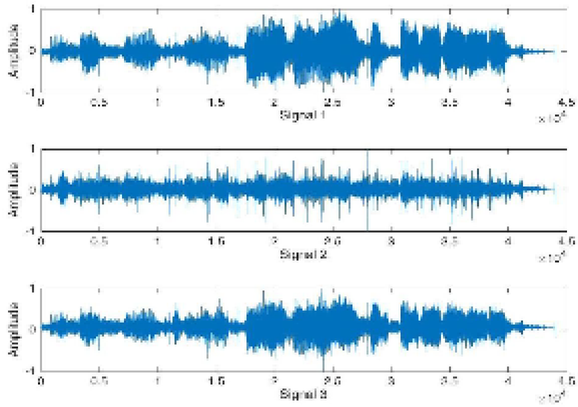
In this experiment we try estimating a 3 original signals from a data having 2 samples (underdetermined), 3 samples (balanced) and 5 samples (overdetermined). We observe that there is not a huge difference in quality of results for these 3 cases. Results are plotted below:


Original Signals
Underdetermined Prediction (2 samples)
Balanced Prediction (3 samples)
Overdetermined Prediction (5 samples)

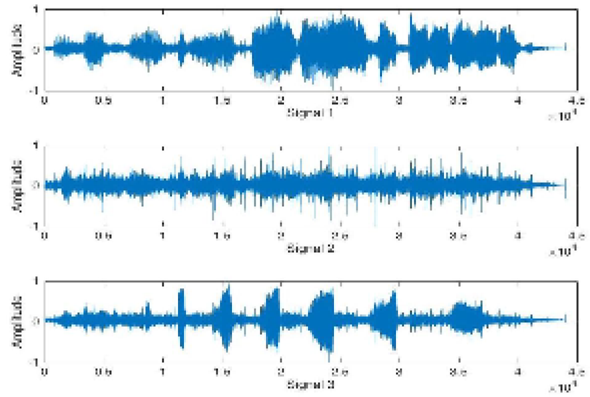
E3: Performance variation with number of iterations



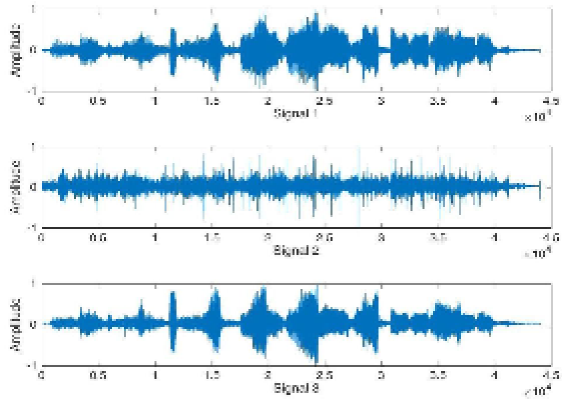
Original Signals



Predicted Signals after 100 iterations



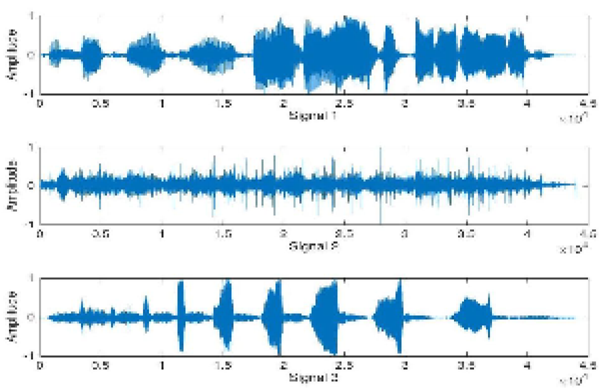
Predicted Signals after 1000 iterations



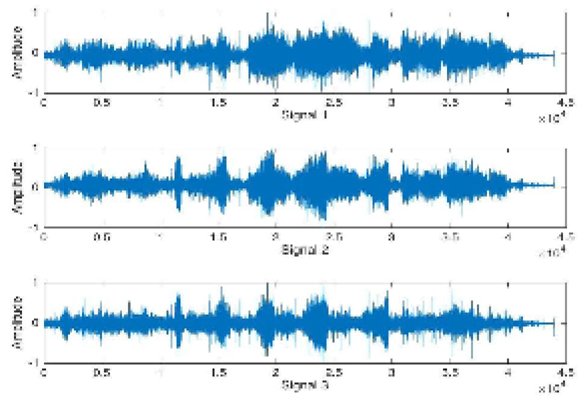
Predicted Signals after 10000 iterations

We can observe from the results that our algorithm converges very quickly and gives very good results even after 100 iterations.

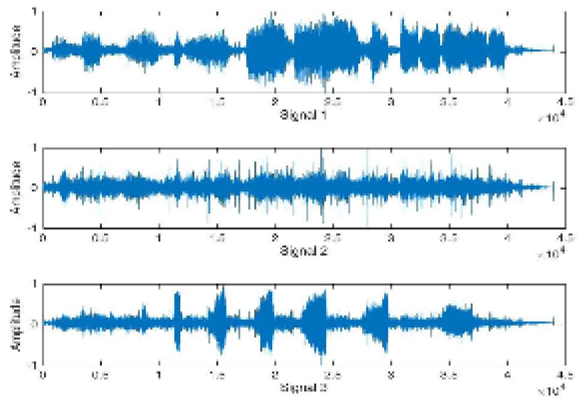
E4: Performance variation with learning rate



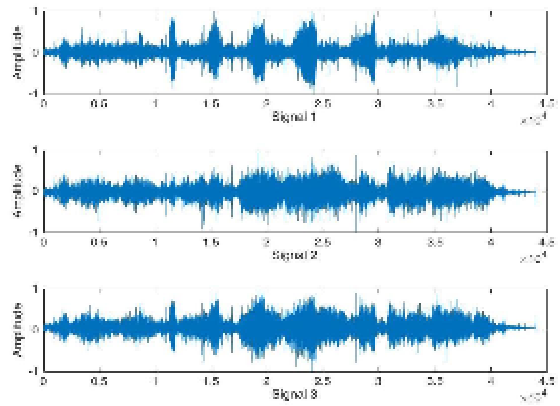
Original Signals



Learning Rate 0.001



Learning Rate 0.01



Learning Rate 0.1

We observe that learning rate of 0.01 gives the best results. 0.001 is too less and will take more iterations to converge, whereas 0.1 is too large.

Conclusions

Our experimental results suggest that performance depends on learning rate to some extent but is not affected a lot by other parameters like size of signals, number of iterations, over determined, balanced or under determined system etc. We also found that signal 4 in sounds.mat was harder to reconstruct than other signals.

References

[1] https://en.wikipedia.org/wiki/Independent_component_analysis