Course: CS391L Machine Learning
UTEID: my3852

Use ICA to Recover Mixed Audio Signals

Anna Mengjie Yu[1,2]*

1. PhD Candidate in Bioinformatics, Department of Integrative Biology, The University of Texas at Austin, Austin, USA
2. MS Candidate in Statistics, Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, USA

* annayu2010@gmail.com

Abstract

Independent component analysis (ICA) is one of the linear transformation methods that finds a linear representation of non-Gaussian data capturing the essential structure of the original data. In this project, ICA is applied to a mixture of sound sources, with the goal to recover the original sound sources. Different iteration numbers and learning rates are experimented and their effects on the recovered sound result will be discussed.

Keywords: ICA, gradient descent method, learning rate

Introduction

Independent component analysis is one of the many different methods of blind signal separation, without aid of the information on the signal mixing process. Applications of ICA can be found in many areas, such as audio processing, image processing, stock market price prediction, mobile phone communications, *etc*. ICA separation of mixed signals are based on two assumptions: 1) The source signals are independent from each other; 2) The values in each source signal is non-Gaussian.

In this project, selected sound sources were mixed using a mixing matrix to simulate "cocktail party problem", the gradient descent algorithm was used to separate out the original speakers' speech signal. Different combination of iteration numbers and learning rates were experimented to achieve the optimal recovered sound signal.

Method

The data used in this project sounds.mat were downloaded from Machine Learning Course Canvas website. The sounds matrix consists of 5 different sound sources, with length of 44000 in each source. Three sound signals (sound1, sound3 and sound4) were selected from the original five sounds sources for the experiment, based on the fact they sound more distinct from each other (Figure 1.) (load_sound.m).

The gradient descent method with maximum number of iteration (ranges from 10,000 to 1,000,000,000) and set epsilon value (1e-6) as stopping criteria was implemented. Step 1: mix original sound using X = AU, where A (m by n) is the mixing matrix, U (n by t) is the original sound source signals. Step 2: Initialize W (n by m) with small random values. Step 3: Calculate Y = WX, where Y is current estimate of the source signals. Step 4: Calculate Z where $Z_{i,j} = g(Y_{i,j})$, where g $g(y_{i,j})$ a signmoid function, is a reasonable default that seems to work well for many problems. Step 5: Find Δ_W = η (I + (1-2Z)Y^T)W, where η (0.01,0.001 and 0.1 tested here) is the learning rate. Step 6: Update W = W + Δ_W, and repeat from step 3 till convergence or maximum number of iterations. Finally, we can recover the sources by computing Y = W. The ICA algorithm is implemented in ICA.m, and the sound mix and recover experiment were conducted in mixAndRecover_data.m.

The original sound sources, mixed sound sources and recovered sound sources were visualized in matlab using sound_plot.m.

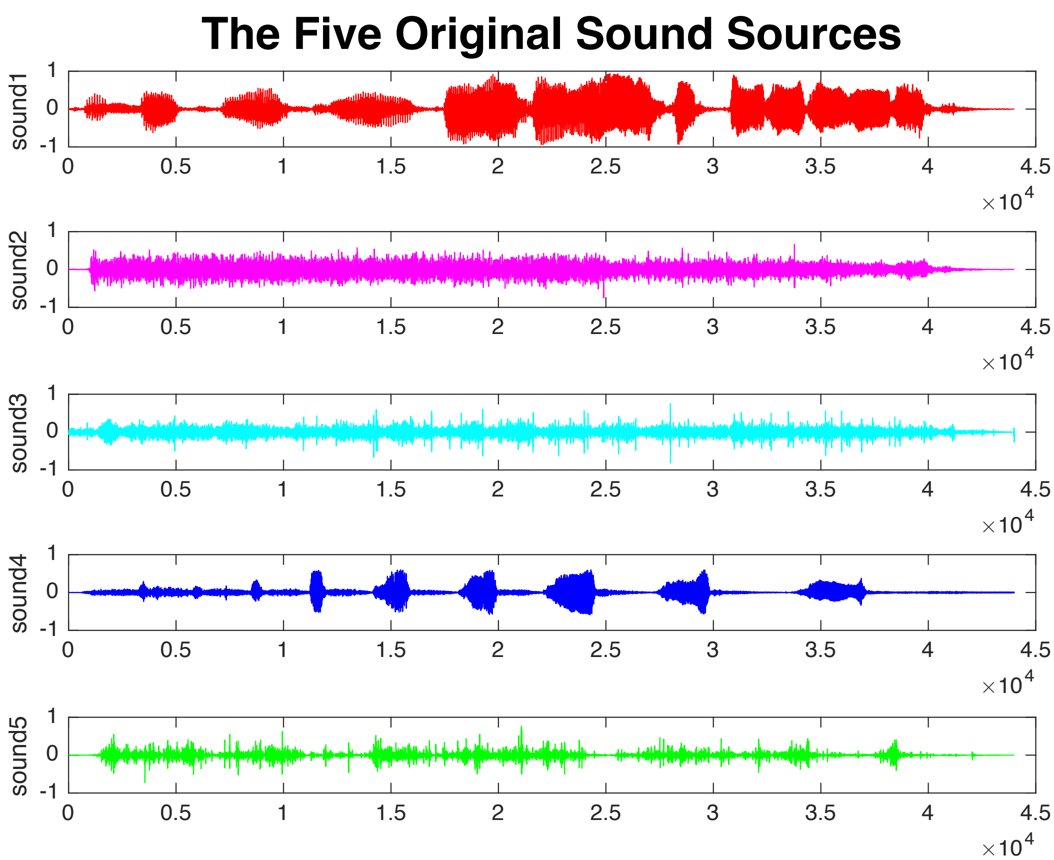

Figure 1. The Five Original Sound Sources
sound1: Old man talking the law of thermodynamics;
sound2: car sound;
sound3: applause sound;
sound4: man laughing;
sound5: clapping hands sound.

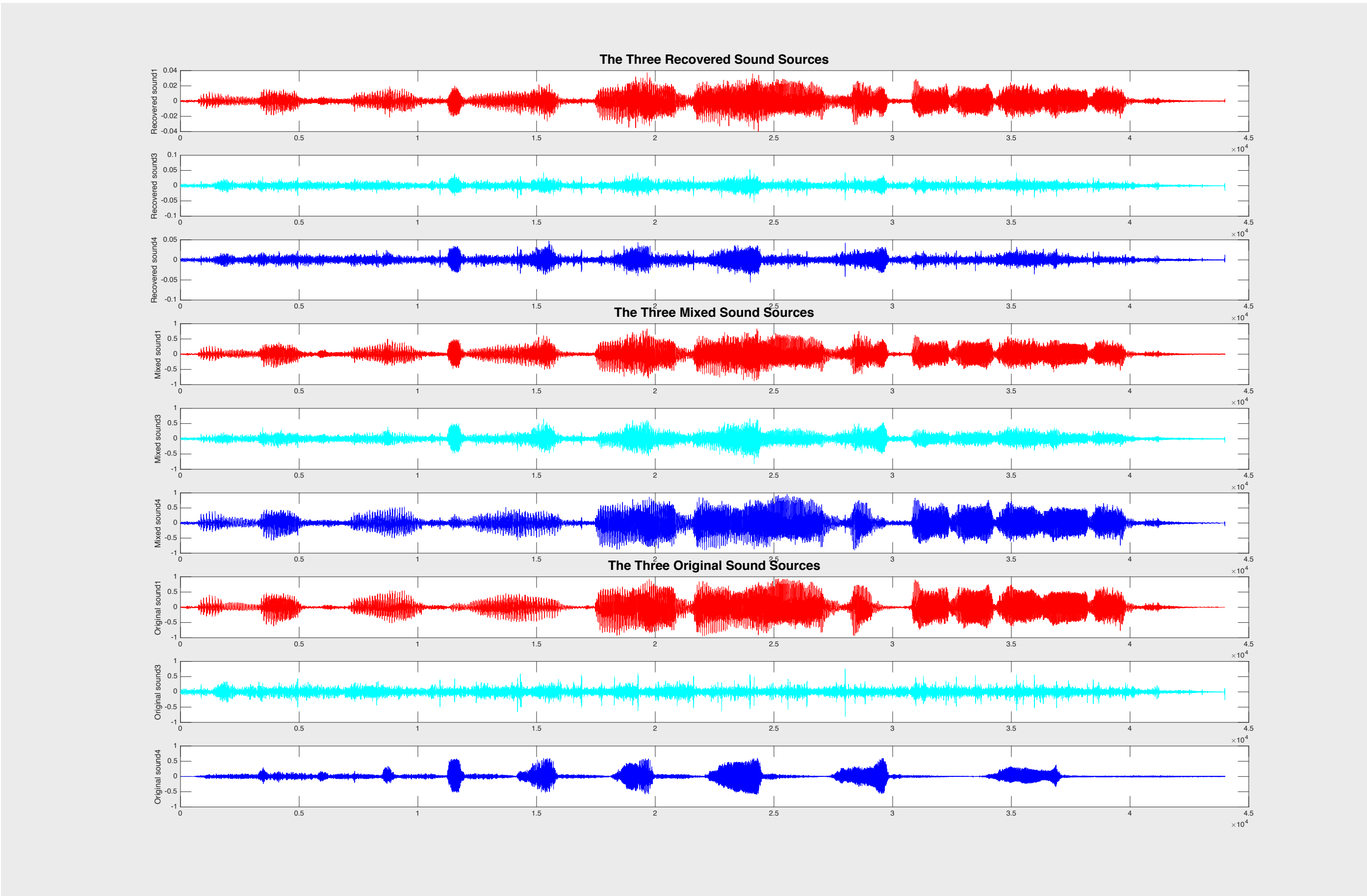

Figure 2. The Recovered, Mixed and Original Sounds with number of iteration = 10,000 and η=0.01

Result

1. Different Iteration Numbers

Two different iteration numbers (10,000 and 100,000) were tried when learning rate η=0.01. When iteration number was set to 10,000, we can see sound1(red) gets good recovery, however sound3(cyan) and sound4(blue) were not well recovered (Figure 2, Supplementary Table 1). When iteration number increased to 100,000, we can see all of the sounds were much better recovered, the sound patterns were more similar to the original sound patterns (Figure 3, Supplementary Table 1) (Note the scaling of the recovered sound might be different).

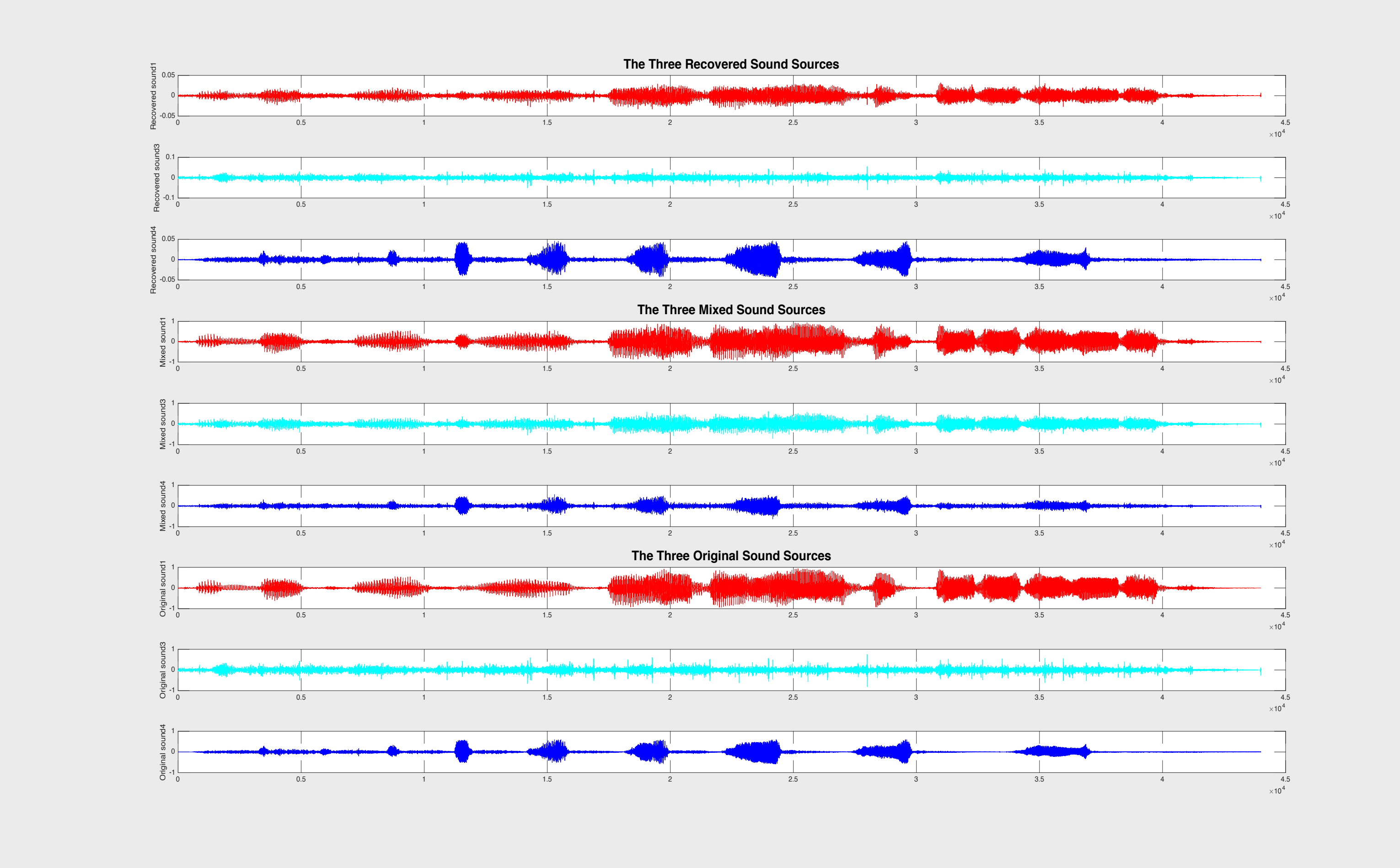

Figure 3. The Recovered, Mixed and Original Sounds with number of iteration = 100,000 and η=0.01

2. Different learning rates η

Two different learning rates (0.01 and 0.1) were experimented at iteration number = 100,000. When η = 0.01, we can see all three sounds got good recovery (Figure 3, Supplementary Table 1). However, when η increases 0.1, even with 100,000 numbers of iterations, none of the sounds got well recovered, and the recovered sounds exhibit similar patterns, making them hard to distinguish (Figure 4, Supplementary Table 1).

3. Large iteration number and small learning rate

One extremely large iteration number 1,000,000,000 was tried using very small η = 0.001. We can see all the three sound sources got good recovery, showing very similar pattern to the original sound sources, *esp.* sound4 (blue) (Figure 5).

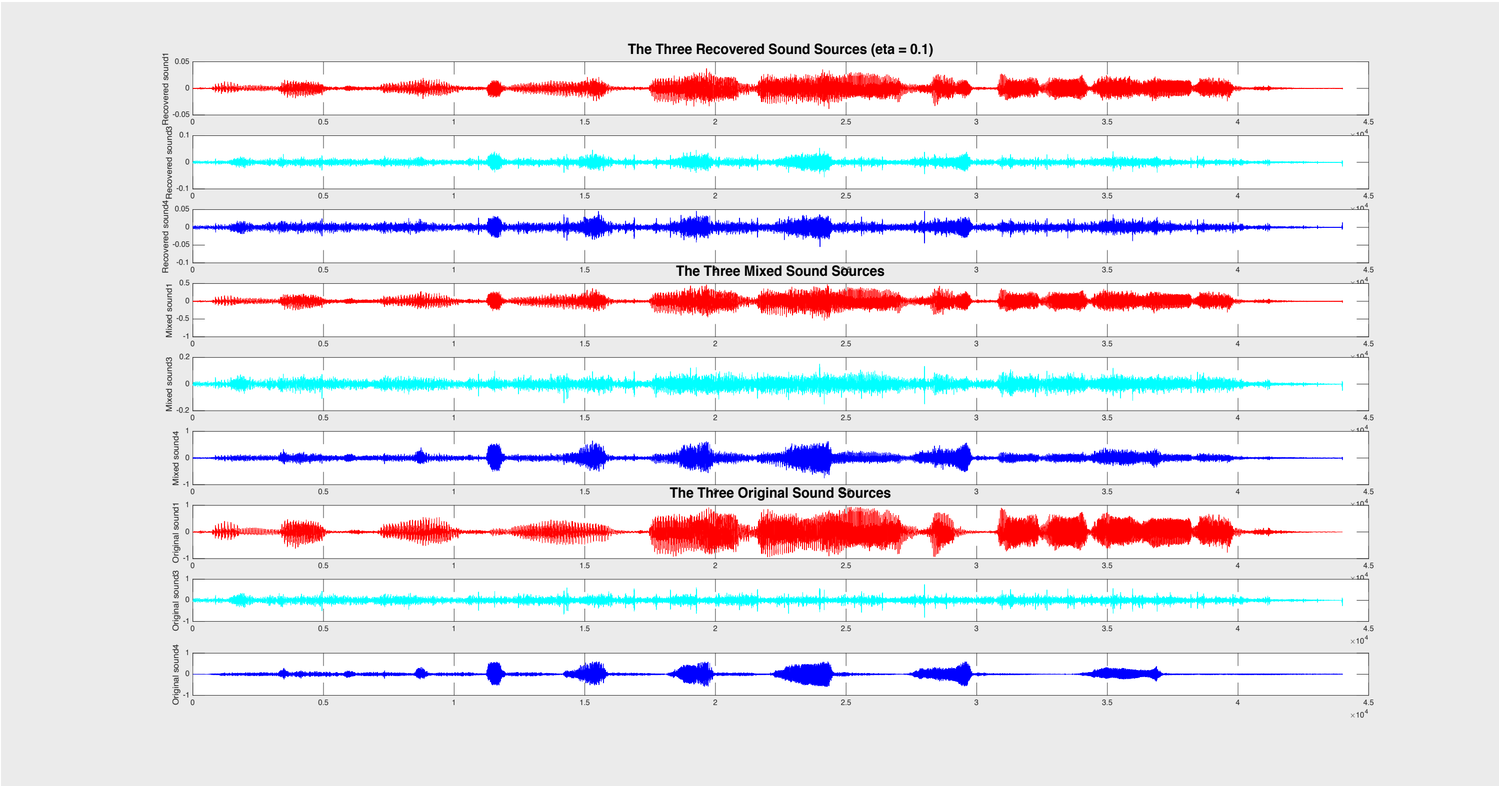

Figure 4. The Recovered, Mixed and Original Sounds with number of iteration = 100,000 and η=0.1

Discussion

ICA is a statistical technique which linearly transforms data into components that are maximally independent from each other. One thing to notice is that, ICA cannot identify the uniquely correct ordering of the source signals, nor the proper scaling (including the sign) of the source signals. However, scaling a speaker's speech signal by some positive factor α affects only the volume of that speaker's speech. Also, sign changes do not matter when played in a speaker.

The larger number of iterations, the better the recovered result. However, the larger number of iterations will make the computation more expensive. The choice of the learning rate also affects the recovery result. Learning rate η too small, the gradient descent can be slow, will take long time to converge; Learning rate η too big, gradient descent can overshoot the minimum, may fail to converge, or even diverge. A not too big and not too small η is essential for optimal recovery.

In real world, *e.g.* speech data and other time series data are hardly independent, which might affect the performance of the algorithm. FastICA, which seeks an orthogonal rotation of prewhitenend data through a fixed-point iteration scheme, and maximizes a measure of the non-Gaussianity of the rotated componenets, might be a future solution for large scale data analysis.

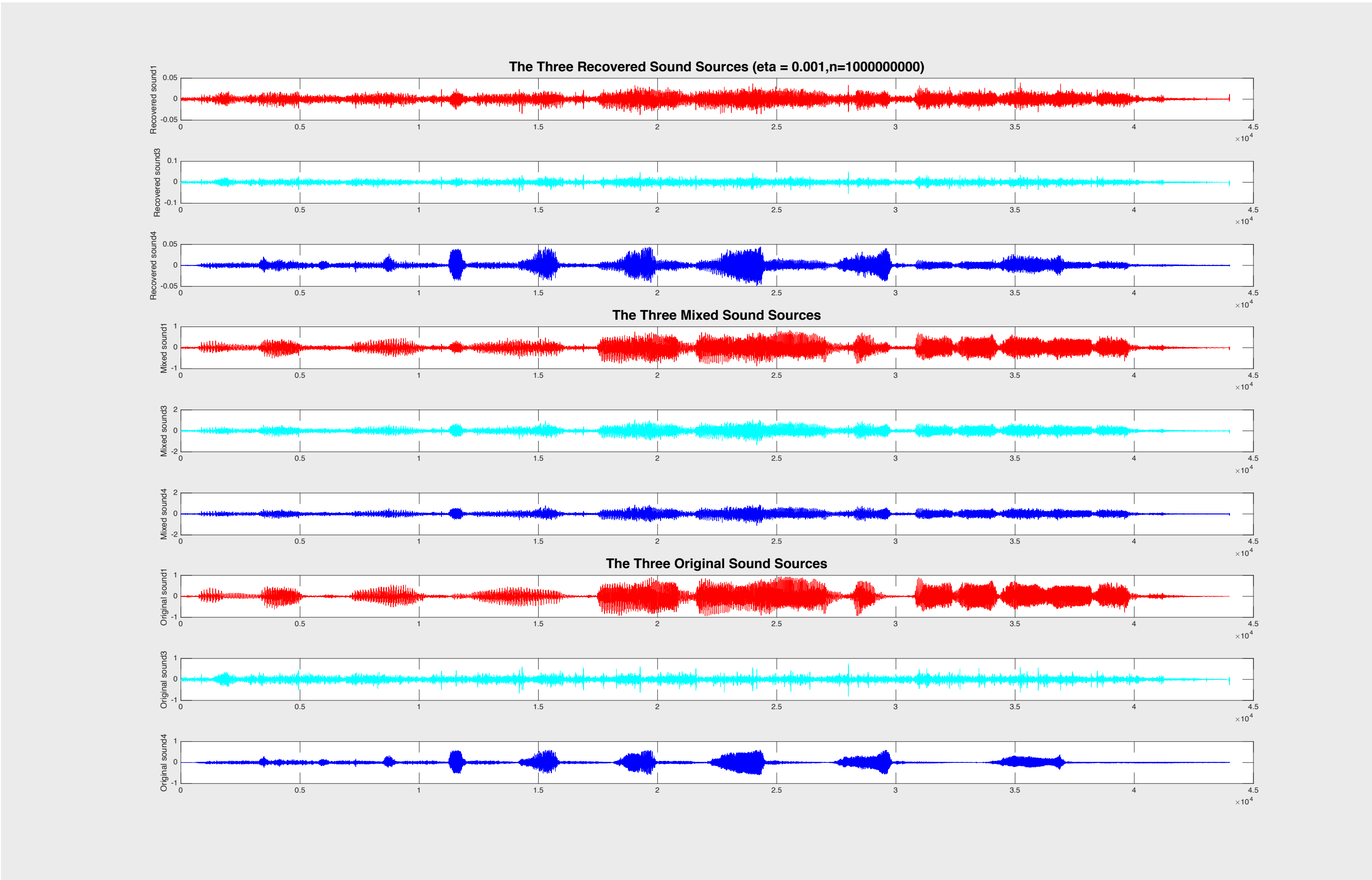

Figure 5. The Recovered, Mixed and Original Sounds with number of iteration = 1,000,000,000 and η=0.001

Supplementary

Iter=10,000 η=0.01	0.8516	0.5884	0.7421
Iter=100,000 η=0.01	0.8965	0.9842	0.9105
Iter=100,000 η=0.1	0.7115	0.7203	0.7641
Iter=1,000,000,000 η=0.001	0.8516	0.9297	0.9814

Supplementary Table 1: Correlation score of Recovered and Original Signal.
(calculated using corrcoef([U' Y'])