ICA for Blind Source Seperation

Lihang Liu
EID:ll32632
lihangliu@utexas.edu

1 Introduction

In this report, we will introduce how we apply ICA to the problem of Blind Source Separation. The iteration process is identical to that in the instrutions. In the experiment, we test how different factors affect the recovery result.

2 Methods

The algorithm is identical to that stated in the instructions. We cite it here for the purpose of a more complete report.

Assume $X = AU$.

Assume $X = AU$.

Initialize the (n by m) matrix W with small random values.

Calculate $Y = WX$. Y is our current estimate of the source signals.

Calculate Z where $z_{i,j} = g(y_{i,j}) = 1/(1 + e^{-y_{i,j}})$ for $i \in 1..n$ and $j \in 1..t$ (where t is the length of the signals). This helps us traverse the gradient of maximum information separation.

Find $\Delta W = \eta(I + (1 - 2Z)Y^T)W$ where η is a small learning rate.

Update $W = W + \Delta W$ and repeat from step 3 until convergence or R_max iterations (you get bored and decide it is done).

3 Results

In this section, we show the experiments on Blind Source Separation by the ICA algorithm. In the dataset, there are 5 original signals with the dimension of 44,000. In the experiments, we compare the influence of the number of dimension t, the number of recorders m and the distribution of original signals.

In all the experiments, the elements of A are randomly assigned between 0 and 1. The elements of W are randomly initialized between 0 and 0.1. And the gradient step η is set to be 0.01.

3.1 First Experiment

As the first experiment, we use all the 5 signals and utilize a randomly generated A to mix it into X. The iterations run for 500,000 times and the signal dimensions t is set to 1000. The result is shown in Fig. fig:overall, where only the first 200 dimension of signals are shown. From the figure, we can find that the extent that a signal is recovered differs. In this figure, the 1st, 2nd and 3rd signal are better separated than the 4nd and 5th signal. In the following section, we will discuss how different factors will affect the separation results.

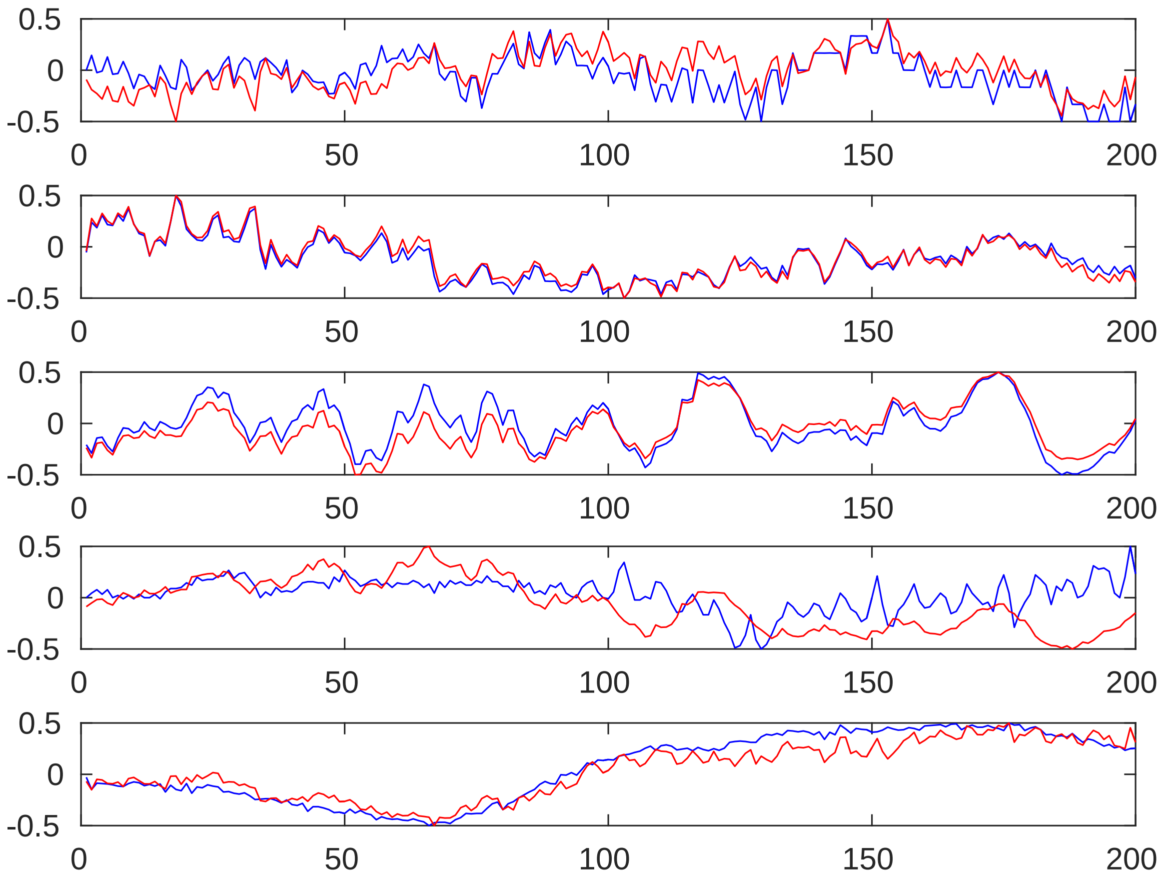


Figure 1: t = 10.

3.2 Signal Dimension t

The dimension of the signals affects both the distribution of the signal and information contained in the signal. In order to evaluate t, we compare them by different values $\{10, 50, 100, 200\}$. And we use the first 3 signals and the maximal iteration number is set to 50,000. The results are shown in Fig. 2,3,4,5. From the figures, we can find that with more dimension t, we can get a better result. Intuitively, with bigger t, we tend to have more data about the original signals, which also make the distribution of signals more stable. Thus they boost a better result.

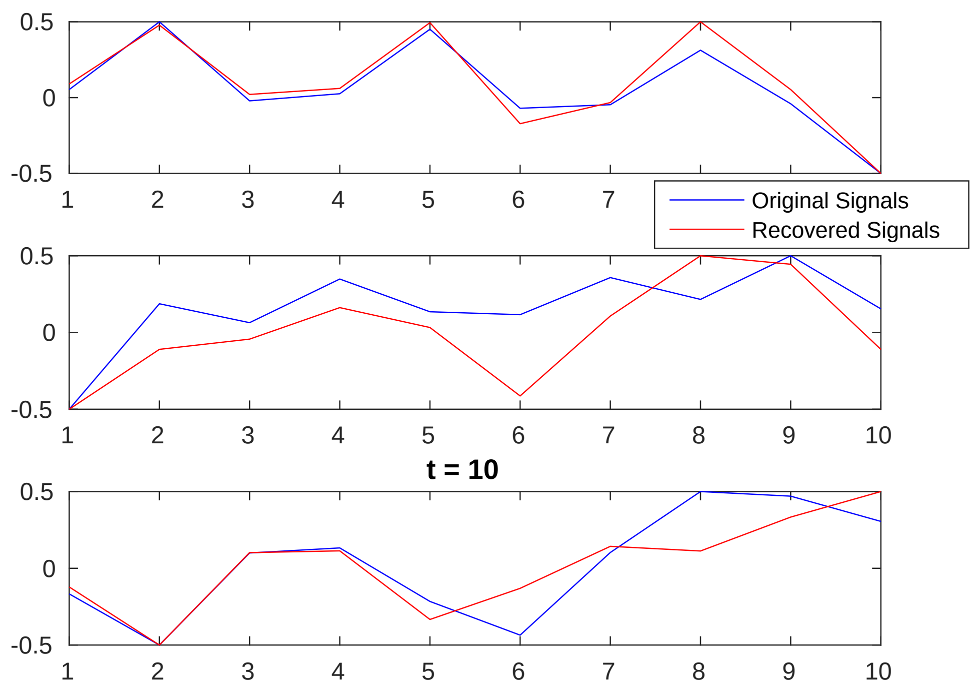


Figure 2: t = 10.

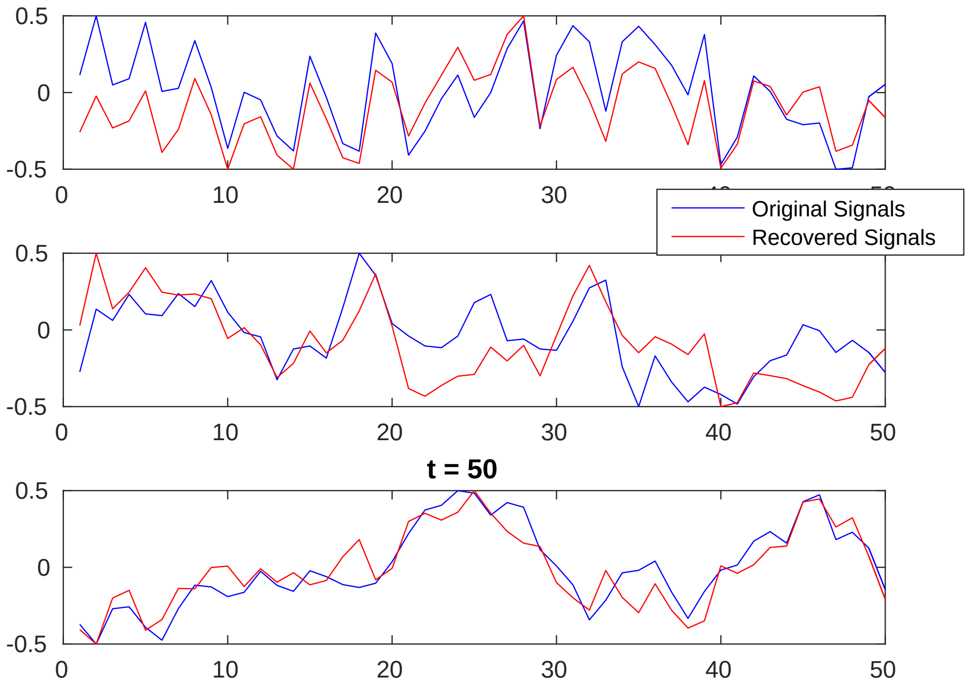


Figure 3: t = 50.

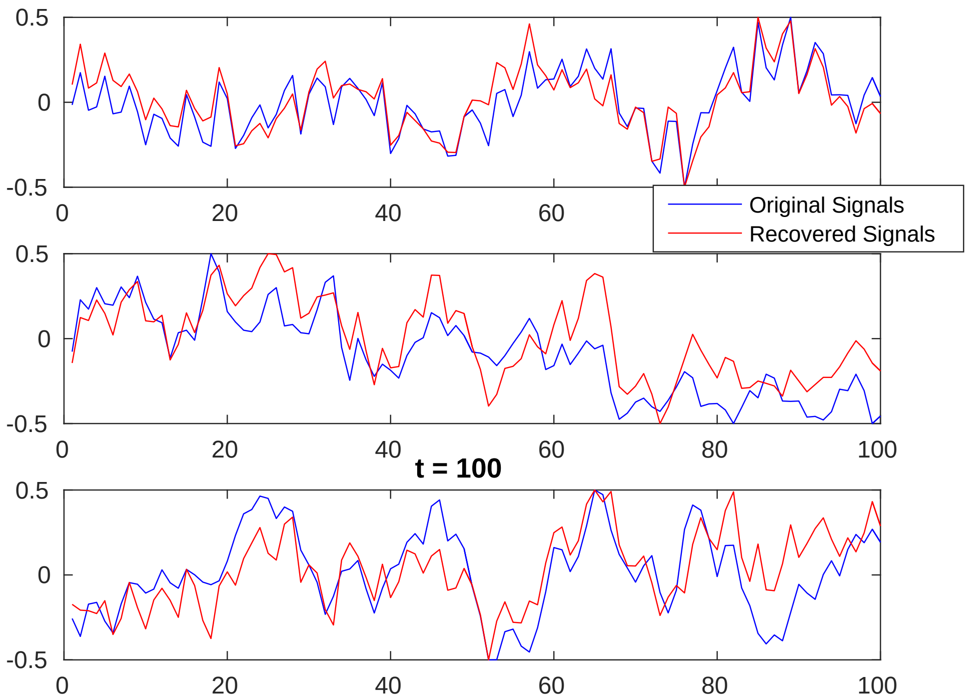


Figure 4: t = 100.

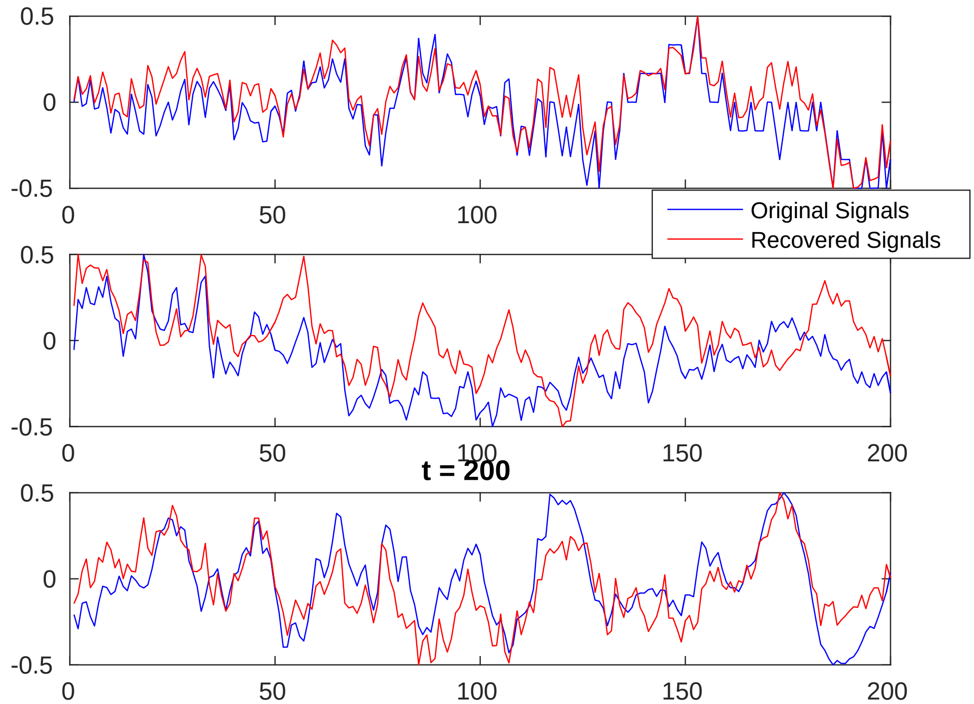


Figure 5: t = 200.

3.3 m of the A

An interpretation for the m is the number of recorders in the case of Blind Source Separation. With bigger m, we tends to get more data about the original signals. To evaluate this, we test m on different values $\{1, 3, 5\}$ while using the first 3 orignal signals. The maximal iteration number is set to 50,000 and the signal dimensions t is set to 50. The results are shown in Fig. 6,7,8. From the figures, we can find that when $m = 1$, the result is worse than the case of $m = 3$, which can be explained as a lack of information. However, when set m to be 5, it doesn't have obvious gains compared to $m = 3$. A guess would be $m = 3$ is enough to recover 3 orignal signals. And as for the part that the setting $m = 3$ isn't doing well, which may due to gaussian factors in the orignal signals, the setting $m = 5$ can't also compensate as the ambiguity is not solved by introducing larger m.

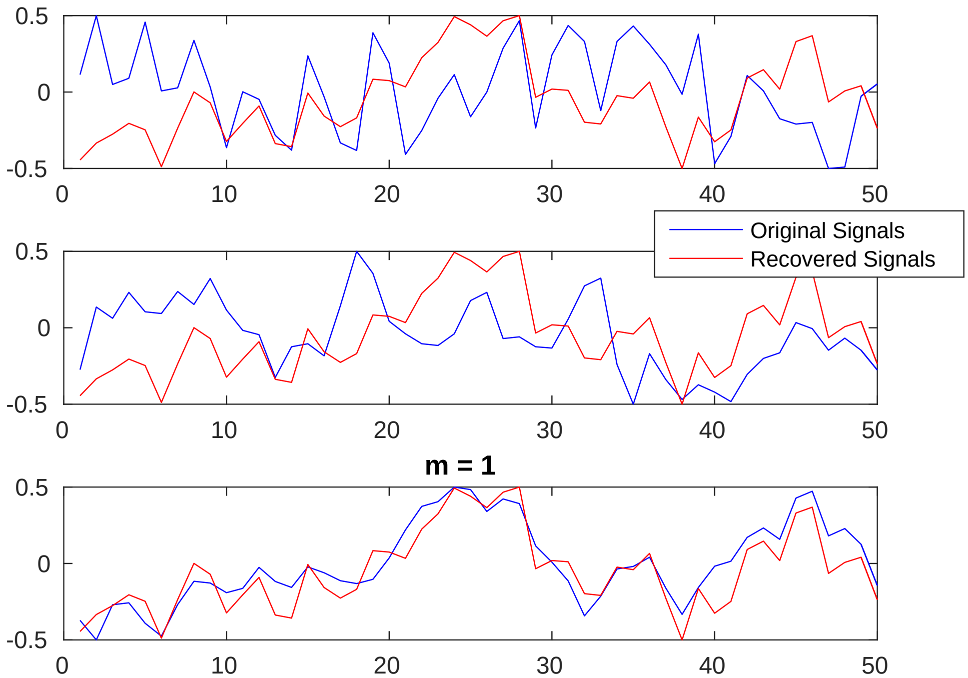


Figure 6: m = 1

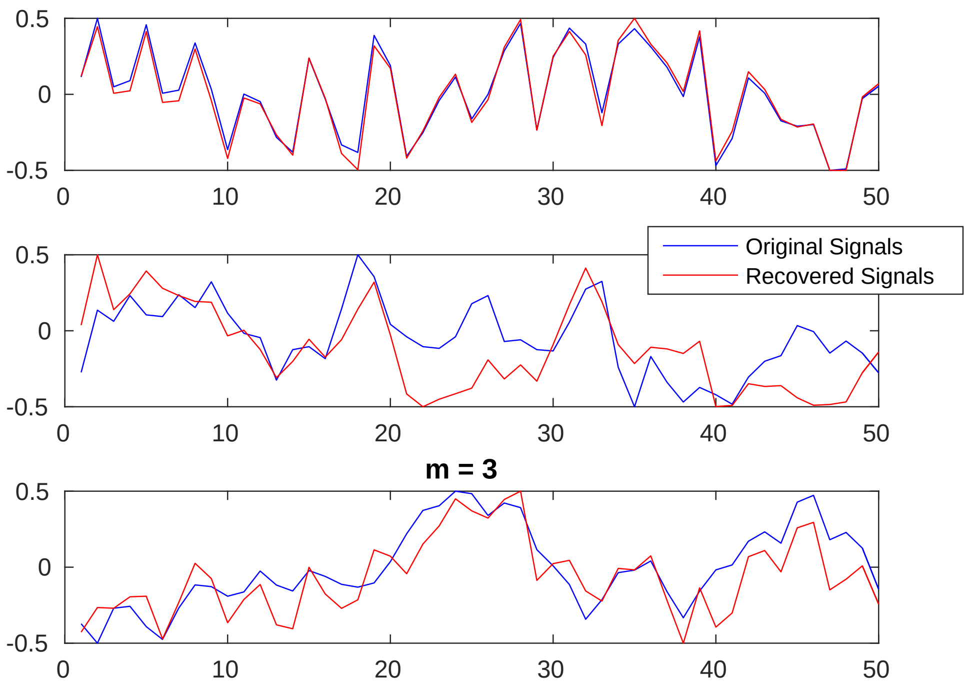


Figure 7: m = 3

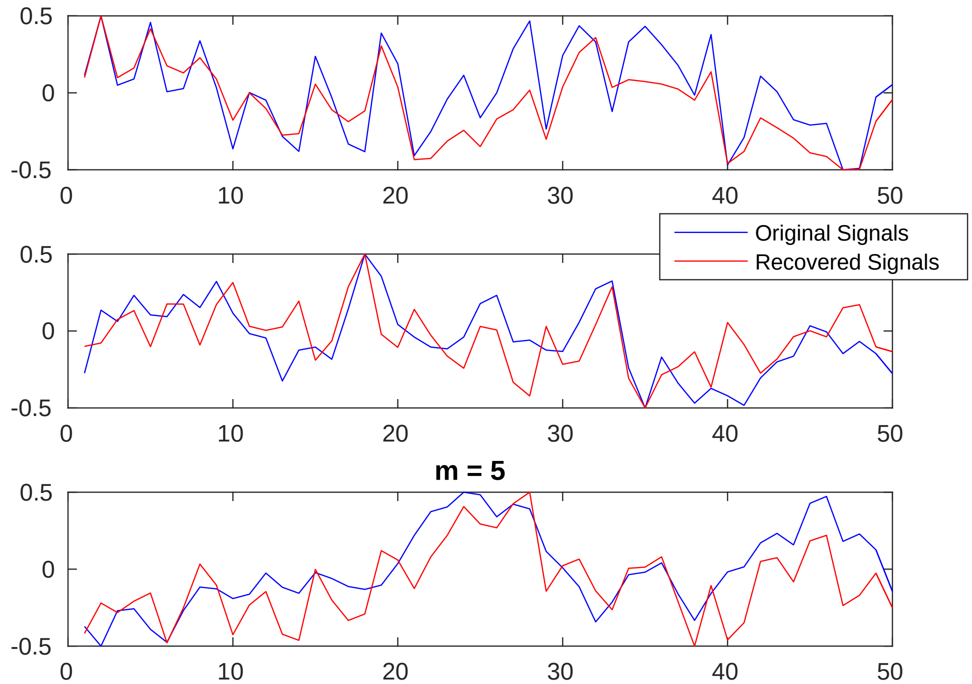


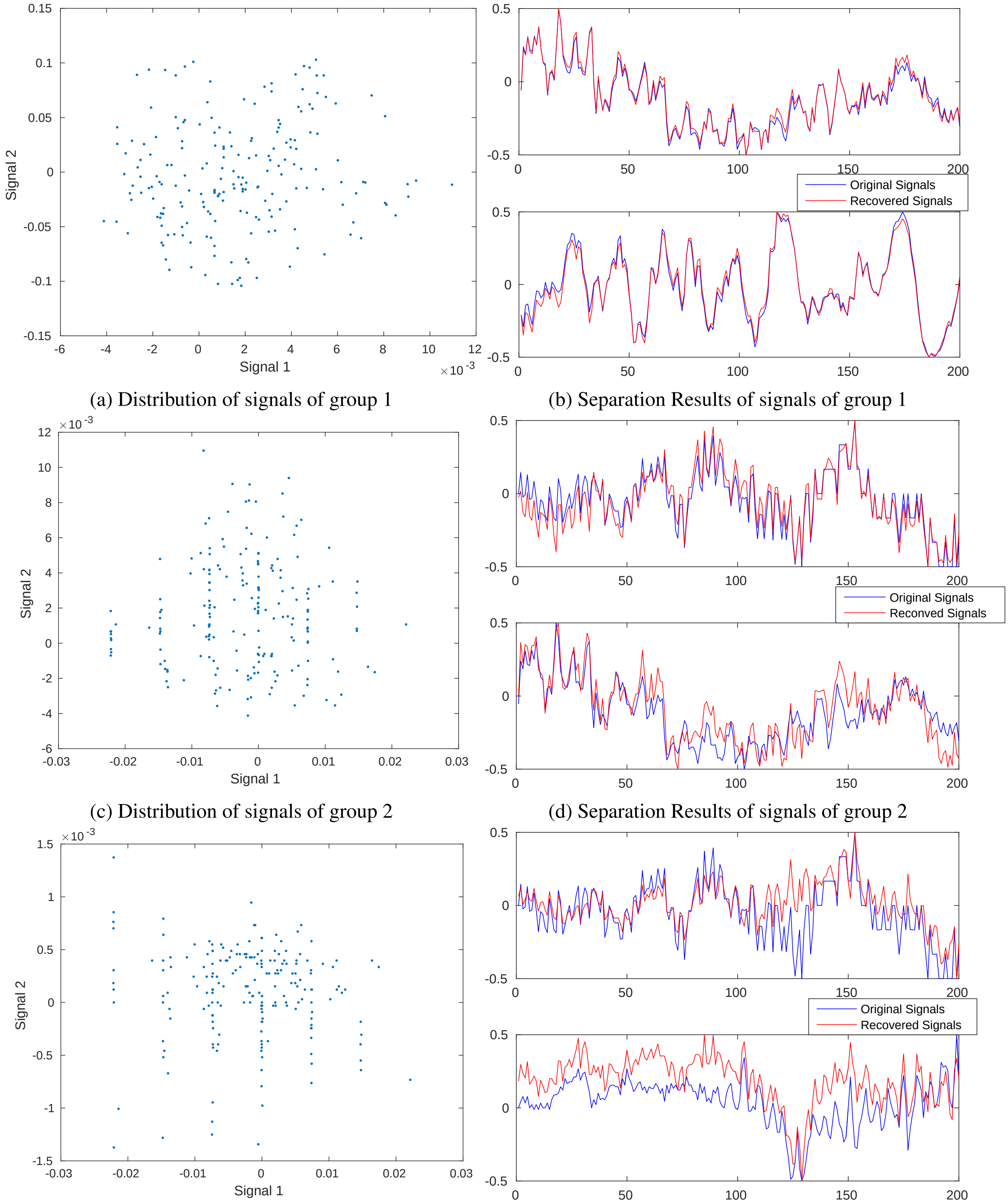
Figure 8: m = 5

3.4 Correlation of signals

In this section, we inspect how the correlation between signals affect the separation. Theoretically, ICA assumes that the original signals are dependent and non-gaussian. From this assumption, we guess that the correlation will lead to the failure of signal recovery. To simplify the problem, we only can two signals as a group when evaluating the correlation. We randomly select 3 group of signal pair out of 5 and test the recovery as shown in the Fig 9. The maximal iteration number is set to 50,000 and the signal dimensions t is set to 200. From the figures we can find that the distribution becomes less independent from group 1 to group 3. While the separation result also becomes worse from group 1 to group 3. In other words, we can guess that the correlation between the signals (indicated by the distribution) affects the result of signal separation.

4 Conclusion

In this report, we show how ICA can apply the problem of Blind Source Separation. Also, in the experiment, we test how different factors affect the separation result.



(a) Distribution of signals of group 1

(b) Separation Results of signals of group 1

(c) Distribution of signals of group 2

(d) Separation Results of signals of group 2

(e) Distribution of signals of group 3

(f) Separation Results of signals of group 3

Figure 9