Independent Component Analysis

Pandian Raju (pandian@cs.utexas.edu, UT ID: pr22695)

February 2017

1 Introduction

The project aims at performing independent component analysis (a type of Blind Source Separation) using stochastic gradient descent algorithm to unmix a set of signals and recover the original signals. A set of source signals are given (each signal is of dimension 1*44000). A random mixing matrix is chosen and different sets of signals are mixed, the ICA algorithm is applied on the mixed signals and the orignal signals are recovered. Formally, if s denotes a set of source signals and A represents a mixing matrix, then

$$x = As$$

where, x is the set of mixed signals. To recover the original signals s from the mixed signals x, we need to find a matrix W (ideally equivalent to the A^{-1}) such that

$$s = Wx$$

Since we do not know the mixing matrix, we can only approximately find W that will recover the set of original signals and hence this is called Blind Source Separation. Analysis is performed to see which set of signals can be separated easier and which ones are difficult to separate and the reasons for the same are discussed. Also, some of the parameters used in the algorithm like the learning rate and the number of iterations are varied and the similarity between the recovered signals and original signals are compared and plotted.

Note: A primary assumption for ICA algorithm is that the input signals are not correlated as well as they are not Gaussian. In fact, signals not being Gaussian is a necessity for ICA algorithm, primarily because any orthogonal transformation of the joint Gaussian will again be a Gaussian and hence no information can be retrieved which can distinguish one signal from the other.

2 Methods

As far as the experiments are concerned, each experiment is repeated for 5 to 10 times (taking different random W matrices) and the average of the results is chosen to be the final value (for the quantitative measures). This ensures that the results are not biased due to the randomness of the initial weight matrix (W) chosen.

2.1 ICA algorithm

- A set of source signals is chosen and a mixing matrix A is chosen in random. The mixed signals are obtained by $x = As$.
- An initial W matrix is chosen in random with small values as an approximation for A^{-1}. As the algorithm proceeds, eventually W matrix gets closer and closer to A^{-1}.
- During every iteration of the algorithm, the gradient of W is calculated and is added to W. This, in turn, is calculated by maximizing the maximum likelihood (log-likelihood) of the function.
- Sigmoid function is chosen initially as the probability density function since it ranges from 0 and 1 and is differentiable.

- The gradient calculation and addition to W is repeated for a certain number of iterations (100000, say) and at the end of the iterations, W matrix is claimed to be the approximate matrix which recovers the original signals by $s = Wx$.
- The signals recovered using the W matrix are then scaled to the range -1 to 1 as the original signals and plotted.

2.2 Matching recovered signals with original signals

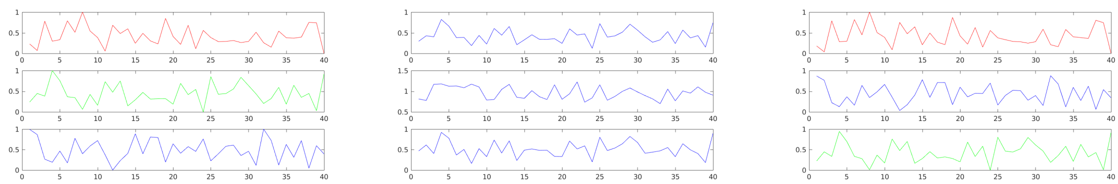
- Once the signals are recovered, a correlation matrix is constructed which calculates the correlation constant for each pair of recovered and original signals (using $corr2$ in MATLAB).
- For each recovered signal, the original signal with highest absolute correlation is chosen to be the matching signal and the recovered signals are color coded accordingly.

2.3 Quantitative measure of accuracy of the algorithm

- One way to examine how well the algorithm works is to observe the original, mixed and recovered signals and the similarities between them. But this doesn't provide a method to quantify the accuracy of the algorithm.
- Since the recovered signals can be in different order than the source signals, the rows of W matrix are permuted according to the matching. In an ideal case where the recovered signals exactly match the source signals, the permuted W will be equal to A^{-1}. Hence the $norm(W - A^{-1})$ is calculated and is used to quantify the similarity between original and recovered signals. The lower this error norm is, the better the algorithm works.
- Another measure calculated is the root mean square (RMS) of the correlation constants of the corresponding matching signals. Since a higher correlation constant means more similarity, higher the RMS value, better the algorithm works.

3 Results

3.1 Smaller test signals

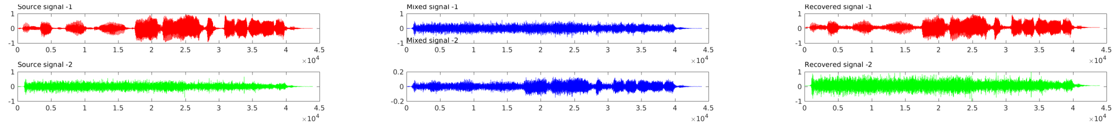


(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 1: Recovering test signals using ICA

- Figure 1 shows the original, mixed and recovered test signals which are smaller signals. We can see the mixed signals do not correspond to any single original signal but a linear combination of the original signals (depending on the mixing matrix A), and the individual signals are separared and recovered very well.
- The order of recovered signals is different than the order of original signals since separating the individual components out is the primary goal of ICA.
- A learning rate of 0.01 and initial W matrix with weights between 0 and 0.1 were chosen and the algorithm was performed for 100000 iterations. Since the original signals do not seem to have any correlation between them, they are easily separated and recovered.

3.2 Mixing two signals

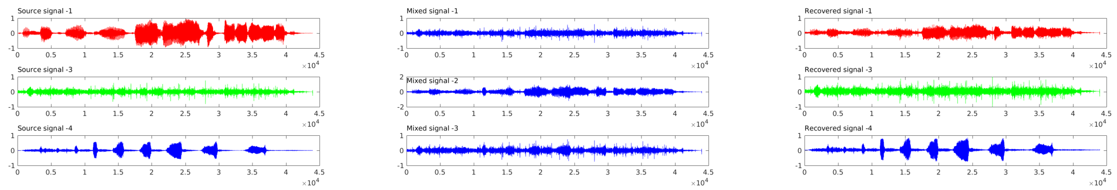


(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 2: Recovering signals 1 and 2 using ICA (Good)

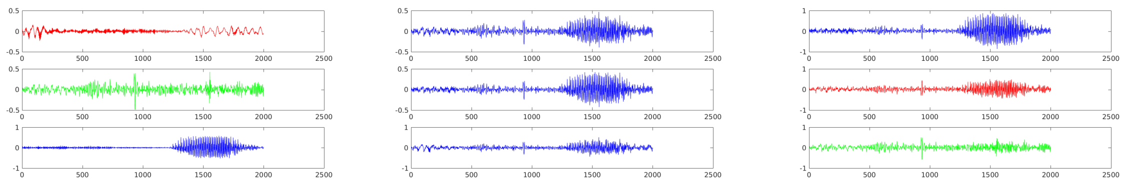
- Figure 2 shows the original, mixed and recovered signals when signals 1 and 2 are mixed (learning rate 0.01 and 100000 iterations).
- We can observe that the signals were able to be recovered pretty good. One reason is that the two signals chosen do not have similarity between them and hence have high entropy. Also, since it's just a mixture of two signals, the overall diffusion of one signal into the other signals is low (ie, if we had mixed 5 signals, each mixture will be very much diffused with all 5 signals making it difficult to separate).

3.3 Mixing three signals



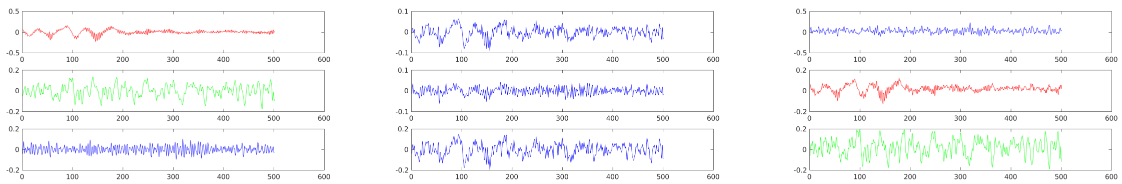
(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 3: Recovering signals 1, 3 and 4 using ICA (Good)



(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 4: Columns 10000 to 12000



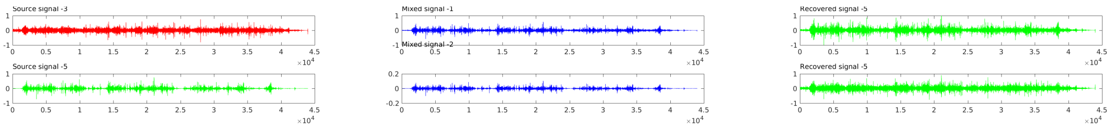
(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 5: Columns 10000 to 10500

- Figure 3 shows the original, mixed and recovered signals of the signals 1, 3 and 4 from the given data (learning rate 0.01 and 100000 iterations). Figures 4 and 5 correspond to the same experiment, but only display a chunk of the signal (from 10000 to 12000 and 10000 to 10500 respectively), just a zoomed in version of the Figure 3. It is clearer to observe the similarities between original and recovered signals in these images.

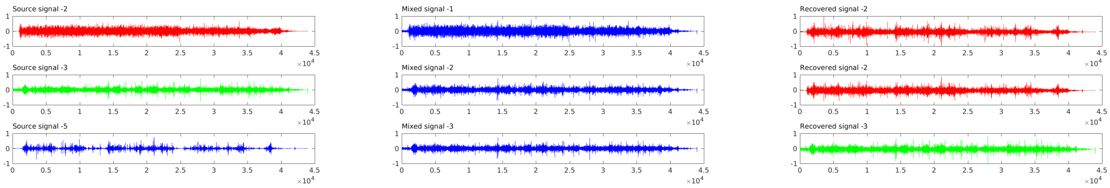
- It can be seen that the signals are properly separated and are matched with the corresponding original signals (as coded by the respective colors).
- Another observation to make is that the recovered signals maintain the same shape as the original signals but the magnitude need not be same. This is because the ICA algorithm can produce scaled versions of the original signals in any order. Irrespective of this, one can easily identify each of the recovered signal with the original signal.

3.4 Signals difficult to separate



(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 6: Recovering signals 3 and 5 using ICA (Bad)

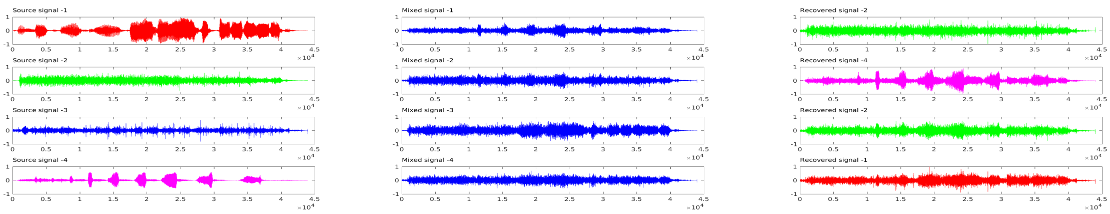


(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 7: Recovering signals 2, 3 and 5 using ICA (Bad)

- Figures 6 and 7 show the original, mixed and recovered signals for the input signals (3 and 5) and (2, 3 and 5) respectively (learning rate 0.01 and 100000 iterations).
- It can be seen that the the signals are not recovered properly and two of the recovered signals look similar to a single original signal and hence get matched (and color coded) incorrectly.
- The reason behind this is that all the input signals (or at least two) in this case seem to be corrleated, meaning they have similar shapes. So our basic assumption that the input signals are uncorrelated goes wrong here and hence the algorithm is not as efficient in separating these signals as the other signals. Another way to interpret is that the entropy between the signals in the mixture is very low and hence is difficult to distinguish the individual components.

3.5 Mixing 4 signals

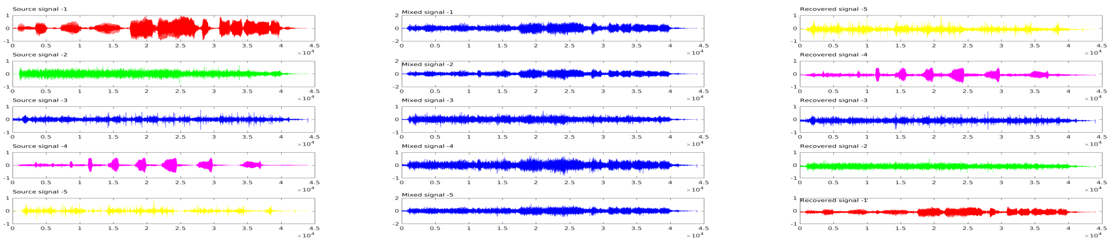


(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 8: Recovering signals 1, 2, 3 and 4 using ICA

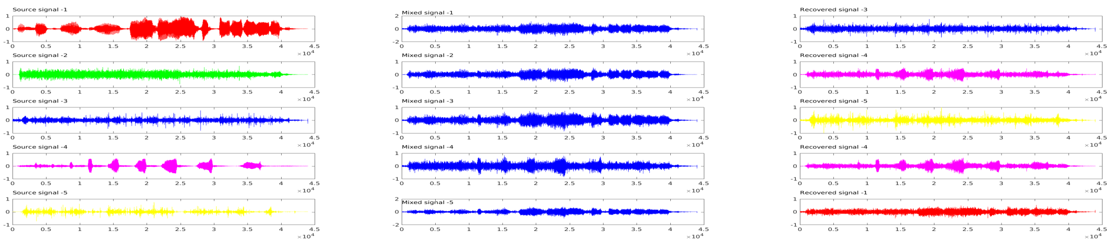
- Figure 8 shows the original, mixed and recovered signals when the input signals 1, 2, 3 and 4 are mixed (learning rate 0.01 and 100000 iterations).
- It can be seen that signal 1 and signal 4 are recovered or at least identified to be matching with appropriate original signals while signals 2 and 3 are not separated properly. The recovered signal 1 is also very much distorted.
- The reason again is that signals 2 and 3 are kind of correlated (they have similar shape and distribution) and hence the algorithm fails to separate them. Also, signals 1 and 4 have higher entropy (as seen by spikes and accumulation of signals at discrete points) and hence are separated easier than the other signals.

3.6 Mixing all 5 signals



(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 9: Recovering signals 1, 2, 3, 4 and 5 using ICA (Good)



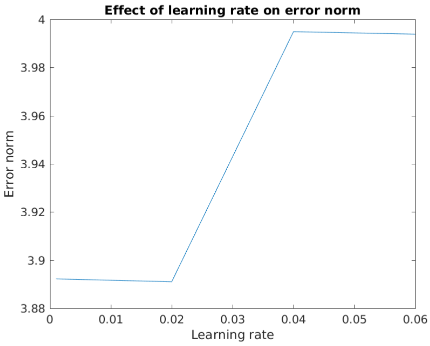
(a) Original signals (b) Mixed signals (c) Recovered signals

Figure 10: Recovering signals 1, 2, 3, 4 and 5 using ICA (Bad)

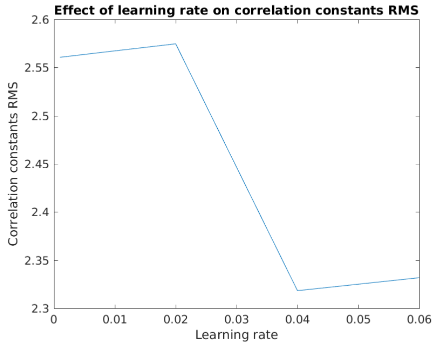
- Figures 9 and 10 show the original, mixed and recovered signals when all 5 sources are mixed (learning rate 0.01 and 100000 iterations).
- We can see that in the first experiment, the signals are separated properly and can be identified to be the appropriate original signals (although the recovered signals are scaled, they retain the shape of the original signals).
- At the same time, in the second experiment, the signals cannot be recovered properly (signals 1 and 4 cannot be separated and other signals also look very similar and do not retain the exact shape of the original signals).
- The reason for these two observations is that the ICA algorithm depends on the parameters used for the experiment - like the initial W matrix, learning rate, the number of iterations etc. It might so be the case that the initial W matrix chosen for the first experiment was better than the one chosen for the second for convergence. Also, the first experiment was run for 1000000 iterations while the second one was run for 100000 iterations. So, it might also be the case that the second experiment finished before meeting the criteria of convergence and hence the variation.

4 Discussion

4.1 Effect of varying learning rate



(a) Learning rate vs error norm

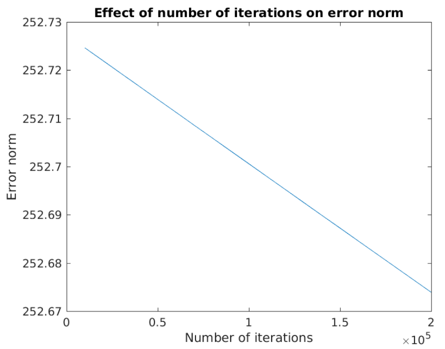


(b) Learning rate vs correlation RMS

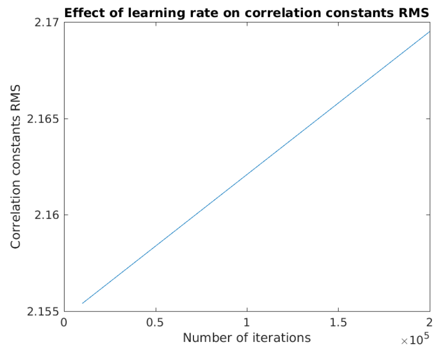
Figure 11: Effect of varying learning rate on the algorithm accuracy

- Figure 11 shows the effect of varying the learning rate on the error norm and correlation RMS. Error norm is the 2-norm of $(W - A^{-1})$ and correlation RMS is the root mean square of the correlation constants (obtained using corr2) of the corresponding matching signals.
- The experiment is repeated for 10 trials and 10 random W matrices are chosen initially and are resued for each of the learning rate so that the randomness of W doesn't bias the results. The number of iterations was set to 100000 for these experiments.
- It can be seen that the error goes to a minimum value at learning rate of 0.02 and increasing the learning rate increases the error norm as well.
- The same observation can also be verified from the other graph wherein the correlation increases initially and reaches a peak at learning rate of 0.02 and then starts decreasing. This concludes that 0.02 is an optimal learning rate for this set of signals used (1, 3 and 4).

4.2 Effect of varying number of iterations

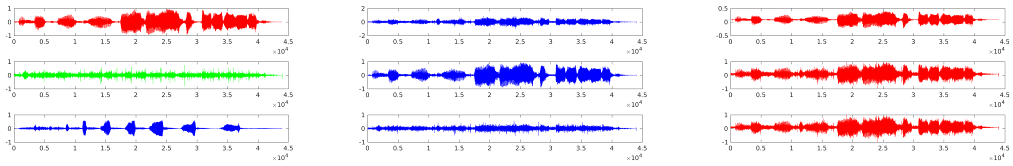


(a) Number of iterations vs error norm

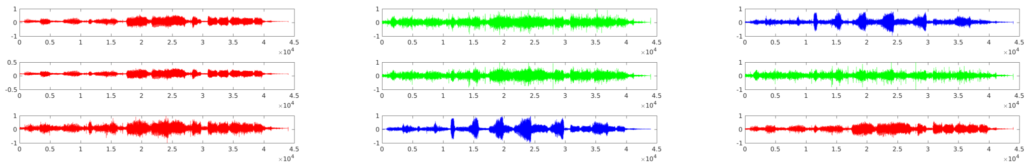


(b) Number of iterations vs correlation RMS

Figure 12: Effect of varying num-iterations on the algorithm accuracy



(a) Original signals (b) Mixed signals (c) Recovered (10 iterations)



(d) Recovered (100 iterations) (e) Recovered (1000 iterations) (f) Recovered (10000 iterations)

Figure 13: Recovering signals 1, 3 and 4 using ICA for different num-iterations

- Figure 12 shows the effect of varying the number of iterations on error and correlation RMS.
- It can be observed that as the number of iterations increases, the error norm decreases and the correlation increases. The reason is that for lower number of iterations, the algorithm doesn't go to convergence yet but for higher number of iterations, it does and hence we get better results.
- The experiments were conducted for 10 trials and the average was taken. Learning rate of 0.01 was used for these experiments.
- Figure 13 shows the recovered signals for different number of iterations. It can be observed that the signals were not recovered correctly for lower number of iterations while for higher number of iterations, they were recovered with good accuracy. This is again because of the algorithm not getting to convergence for lower number of iterations.

4.3 Performance of the algorithm

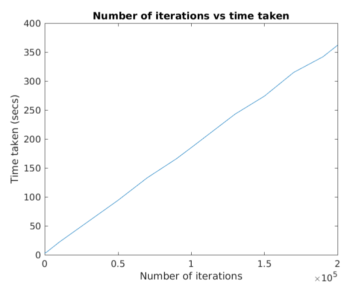


Figure 14: Number of iterations vs time taken

- Figure 14 shows the plot between the number of iterations and the time taken by the algorithm.
- It can be seen that as the number of iterations increases, the time taken also linearly increases. Hence, depending on the level of the accuracy and the performance needed to be guaranteed, the number of iterations can be chosen.

5 Conclusion

By this experiment, we performed ICA on a set of mixed signals and did a series of analysis, picking different sets of input signals and altering parameters like learning rate and the number of iterations. By training the algorithm with set of sample inputs, these parameters can be set to give better accuracy for the algorithm and can be used to effectively separate the independent components from a non-gaussian mixture.