Akanksha Bansal (ab45563)	**Assignment 2**
bansal@cs.utexas.edu	**Feb 21, 2015**

Basis Of the Assignment As part of this assignment we try and separate a mixture of non-gaussian signals using Independent Component Analysis(ICA) performed using gradient decent. We have been given the source signals, which are mixed using a randomly generate A matrix.

$X = A * U$

The signals can be mixed in any linear combination and can be used to generate any no of output signals (m). In order to recover the source signals we need to determine W. Such that :

$Y = W * X$

The first step in order to recover the signals is to make a guess of W. And then we perform gradient decent towards the value of W which maximizes information separation. Then we calculate the δW which is the estimated change in W which would help us maximize this information separation. And we repeat this until we have achieved desired accuracy.

Once convergence is achieved, we calculate the recovered signals Y. It is not necessary for the recovered rows to be ordered in the same manner as in the original signal. Therefore in order to measure the accuracy, we need to measure the correlation between the source signals and the recovered signals and we assign the highest co-related signals to be the same. One thing which needs to be taken care of is that, even negative correlation is possible between the signals(ICA is independent of sign, scale). Accuracy is then measured as the avg. of the magnitude of maximum correlation between the signals.

How to determine Convergence ? The step size used for gradient decent can effect the accuracy and the time taken to reach that accuracy a lot. If the step size if large, we might not reach convergence. If the step size is small, we get stuck in local minimum. Or we For example if the value of eta is 0.001 an accuracy of 95 is reached in around 200,000 iterations, which is achieved in 15,000 iterations with eta of 0.01. If the step size is large, then we might not even reach convergence in all cases, example eta of 0.7.

The first way with which I started to ensure convergence is by using annealing. This is used in order to control the step size used for gradient decent. Initially we start with a larger step size, and with increase in the no of iterations we would expect the step size η to reduce. This increased the chance of convergence.

Another way can be to make use of magnitude of ΔW indicates the change being made in the matrix in order to recover the source signals. Once we reach a maxima we would not expect this value to change much. Thus convergence can be said to have reached when the change in ΔW is less than a threshold value. This threshold value is determined from various tests run and determining the change in ΔW when accuracy is within expected values.

Difficulties with Experimentation

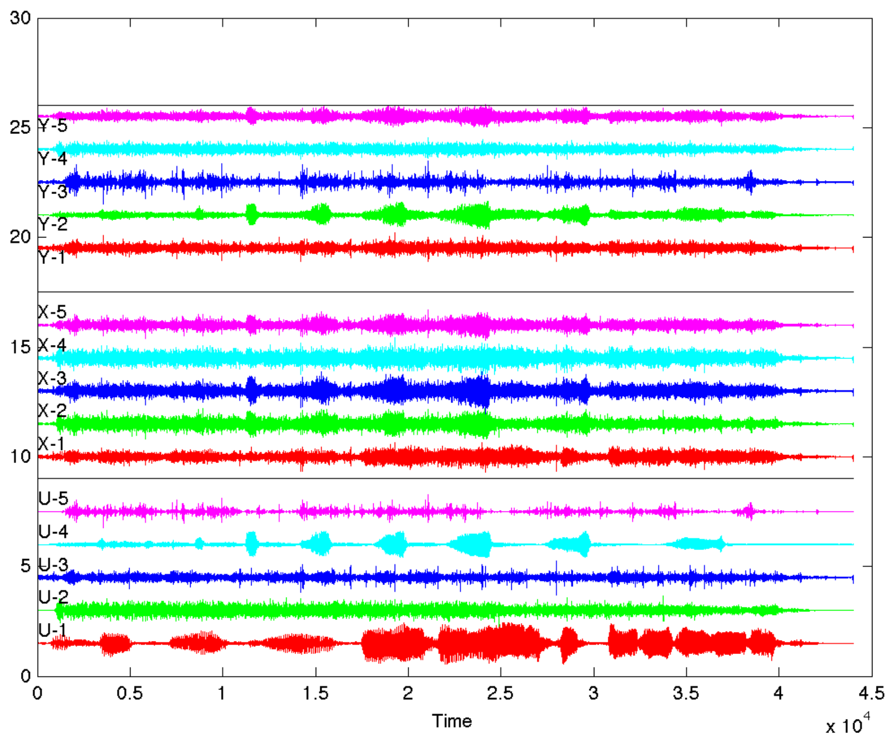
- Random Mixing Matrix

  Blind Source Separation aims to identify original signals given the linear mixtures. The mixture are generated from random linear combination of the sources. Let A be the mixing matrix, which produces the mixed signals X by applying a linear transformation on the sources. Now for different values of A convergence would be expected to reach in different value of η and different no of *iterations*. Therefore some tests purposes, I decided to chose a single randomly generated matrix A.

- Random Recover matrix

  The initial guess of the recover matrix W too seemed to effect the rate at which convergence is achieved. However, for these tests I didn't fix the W matrix because in a real life problem (where we are not given the source matrix) the initial start matrix W would most probably randomly generated. However I kept the matrix W small at the beginning because we would not want to start with a huge step size, we might skip the optimal minimum.

  Blind Source Separation aims to identify original signals given the linear mixtures. The mixture are generated from random linear combination of the sources. Let A be the mixing matrix, which produces the mixed signals X by applying a linear transformation on the sources. Now for different values of A convergence would be expected to reach in different value of η and different no of *iterations*. Therefore some tests purposes, I decided to chose a single randomly generated matrix A.



Experiments • Different combinations of Signals I tested with various no of sets of signals.

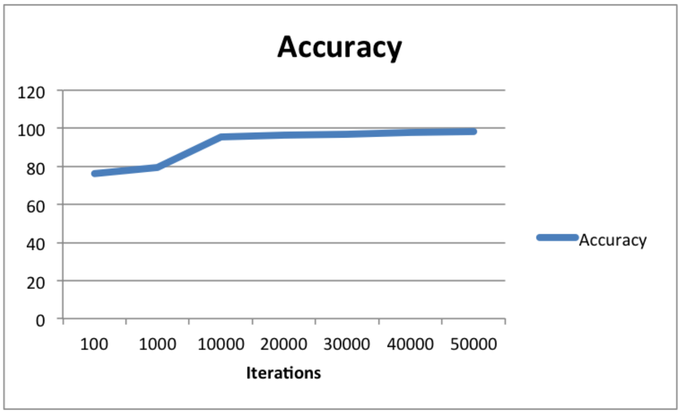
No of Signals being combined	Avg Accuracy
3 signals	85
4 signals	75
5 signals	67

The first three signals have the maximum diversity within the source signals, therefore best results are achieved if the first three signals are combined. Mixing the last 3 signals gave accuracy of 79. This indicates that separation is easier within signals if the signals themselves are different from one another.

Mixing the source signals in more linear combinations didn't have much effect on the accuracy.

- Different Iterations and change in Accuracy

  With a very basic approach of gradient decent, increasing the no of iteration increased the accuracy of the model. However, after certain point the increase in the accuracy wasn't much.



- Annealing

  It was extremely useful in leading the signals to converge on a good optima (atleast prevented converging on a poor local optima). Without annealing, the gradient descent algorithm usually converged to a local optimum which was not useful in separating the sources. Annealing also reduced the number of iterations required to converge to achieve a given accuracy.

No of Signals being combined	Avg Accuracy	Iterations
3 signals	95	10000
4 signals	81	10000
5 signals	72	10000

- Convergence based of ΔW IN order to understand the behavior of the model while making use of this convergence method, I changed the program to come out of converge when the change in ΔW would be $1 * 10^{-13}$

No of Signals being combined	Avg Accuracy	Iterations
3 signals	99.2	57152
4 signals	84.7	89294
5 signals	76.3	124728