Blind Source Separation using Independent Component Analysis

Harshal Priyadarshi
harshal@cs.utexas.edu

February 19, 2016

1 Overview

Data is never clean. Usually it is corrupted by either random noise, or by background interference due to certain environment parameters that have not been quantified yet or that is known but unwanted. The problem is to extract the desired signal from the mixed signals that our sensors record. In a more general setting, the job is of reconstructing the original signals from the mixed signals provided. This is also called the Blind Source Separation Problem [1]. Fortunately, there is an algorithm which given the mixed signal can generate the original signals with very good accuracy, by minimizing the log-likelihood probability estimate of the observed mixed signals. Though a perfect reconstruction can't be expected on larger signals, it gives fairly accurate reconstructions, and if taken on smaller signals, the accuracy is very close to the exact reconstruction. This algorithm is called Independent Component Analysis [2]

2 Notation

- s : number of original signals used for mixing
- m : number of mixed signals created
- t : number of samples for each signal
- $U \in I\!R^{(s,t)}$: the original (unknown) signal used for mixing
- $A \in I\!R^{(m,s)}$: the mixing matrix
- $X \in I\!R^{(m,t)}$: the mixed signals to be separated
- $W \in I\!R^{(s,m)}$: the recovery matrix
- $U_{rec} \in I\!R^{(s,t)}$: the recovered signals using ICA on X

3 Concepts

So the biggest problem with blind-source separation is that we don't have the following quantities:

1. The original signals($= U$)

2. The mixing matrix($= A$)

What we have is the mixed signals which were obtained using some (A, U) pair, using the following equation:

$$X = AU \tag{1}$$

Now the task is basically to obtain the recovery matrix, so that we can obtain the best reconstruction of the original signals using the following equation.

$$U_{rec} = WX \tag{2}$$

The performance of the algorithm that generates the recovery matrix depends on how close are the original signals and the reconstructed signals.
ICA algorithm is one of the good candidates that efficiently does this. However, there are a lot of ambiguities associated with ICA, which must be constrained to ensure sufficient convergence. For certain signals, and for certain types of desired separation ICA will fail badly.
The ambiguities associated with ICA algorithm that may hinder successful separation are:

1. If X is mere permutation of U, then there is no way to exactly recover U in order
   The algorithm does not know if the X was obtained using identity transform from the original signals that were permutation of U or was obtained by using the permutation transform from the original signals U. Thus this algorithm might give the recovered signals that were in different order than the ordinary signals.
   This is not the problem though, as permutation can be easily handled using cross-correlation matric.

2. No way to recover the correct scaling of W and U
   These two are equally likely constructions

$$X = (A)(U) \tag{3}$$

$$X = (2A)(\frac{U}{2}) \tag{4}$$

   Now as the original signal is not known, thus ICA does not know if the original signal was U or $\frac{U}{2}$. Thus, it does not know if W should be closae to $(2A)^{-1}$ or $(A)^{-1}$. Thus it will not recover the correct scaling of the W and U_{rec}.
   However this is not a problem too, as we can easily normalize the signal.

3. ICA fails to separate gaussian distributed signals in U
   As we know that the product of gaussians is also a gaussian distribution. So X will also be a gaussian. Now gaussian distribution is invariant to rotation, as the covariance of the output signal,X, remains the same even with rotated input signal. Thus the algorithm does not know if the output signal is obtained from a rotated version of original signal or the signal is itself original. Now rotation applied to original signal changes the signal altogether. Thus it is impossible to reconstruct the original signal

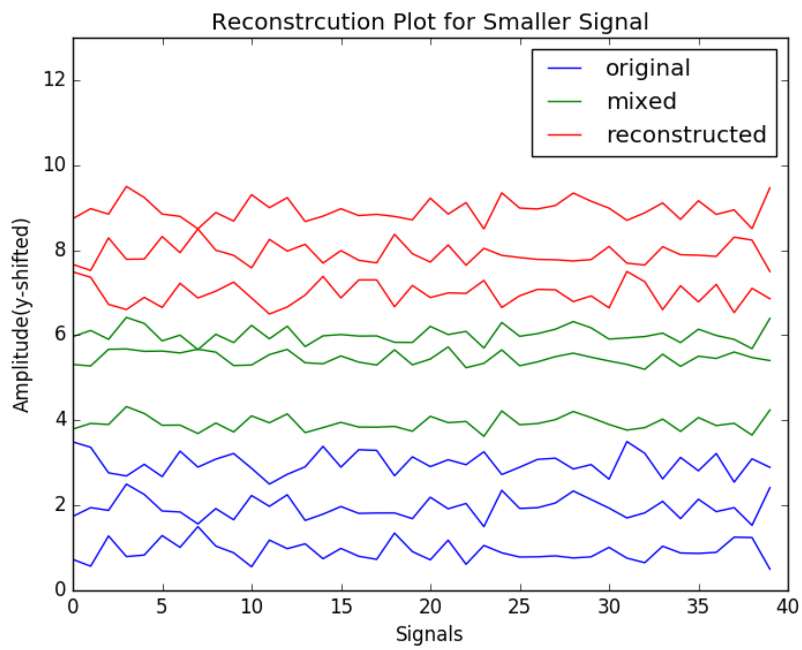


Figure 1

	1	2	3
1	-0.4282	**0.9908**	-0.4863
2	-0.5422	-0.4010	**0.9919**
3	**0.9925**	-0.4865	-0.5023

Table 1: Correlation Matrix (U_original,U_reconstructed) for smaller signal

Update Rule

Assuming that the true signals are independent of each other, the joint distribution of the signal is,

$$p(s) = \prod_{i=1}^{n} p_s(s_i) \tag{5}$$

Now we know that $s = w_i^T x$, thus transforming the probability density to the space of x, we get this

$$p(x) = \prod_{i=1}^{n} p_s(w_i^T x).|W| \tag{6}$$

Now we need to have a prior in the distribution, other than gaussian. If we take the cumulative distribution $g(.)$ as the sigmoid, and use the fact that probability distribution is the derivative of cumulative distribution, we get the following equation for the log likelihood

$$l(W) = \sum_{i=1}^{m} (\sum_{j=1}^{n} \log g'(w_j^T x^{(i)}) + \log |W|) \tag{7}$$

Now, for maximizing log-likelihood[3] w.r.t W, we use the gradient descent, for which the gradient is $\frac{\partial l(W)}{\partial W}$ is the same as the update rule given in the algorithm. Later when we are tweaking the parameter β for the sigmoid[4], we will also get the update rule by finding the derivative of $l(W)$ w.r.t sigmoid parameter,β

Implicit Learning Rate
The basic algorithm suggested below is a bit different from the actual ICA algorithm. The update ΔW is multiplied by a term $W^T W$. This is essentially a reduction in the learning rate, when you are close to the minima of the evaluation space. Since when W is reducing, $W^T W$ is also decreasing. Thus effectively the learning rate $\eta \leftarrow \eta W^T W$, is reduced. So we don't require explicitly reducing the learning rate (using methods like simulated annealing etc.) to ensure convergence.

4 Algorithm

Algorithm 1 Independent Component Analysis algorithm

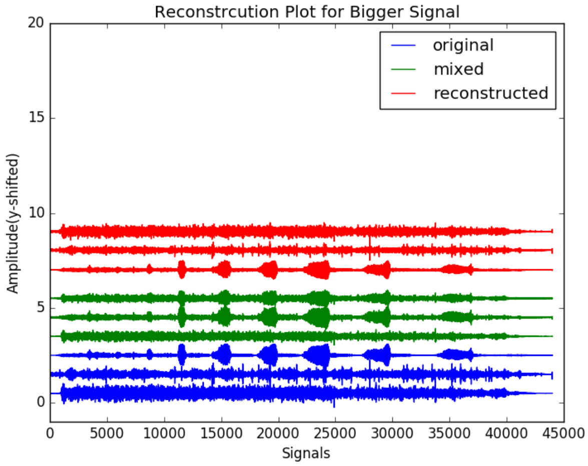
procedure ICA–RECONSTRUCTION
  Let the training mixed signals be $X = AU$
  Initialize W with uniform random values between [0,0.1]: $W \leftarrow rand(s, w)/10$
  Set learning rate η
  Set the maximum number of iterations k
  for $i = 1 : k$ **do**
    Estimate the present best reconstruction, $Y = WX$
    Compute the sigmoid CDF, Z for the signals: $Z = g(Y) = \frac{1}{(1+\exp^{-Y})}$
    $\Delta W \leftarrow \eta(I + (1 - 2Z)Y^T) * W$
    Update $W \leftarrow W + \Delta W$
  end for
end procedure

5 Experiment & Insights

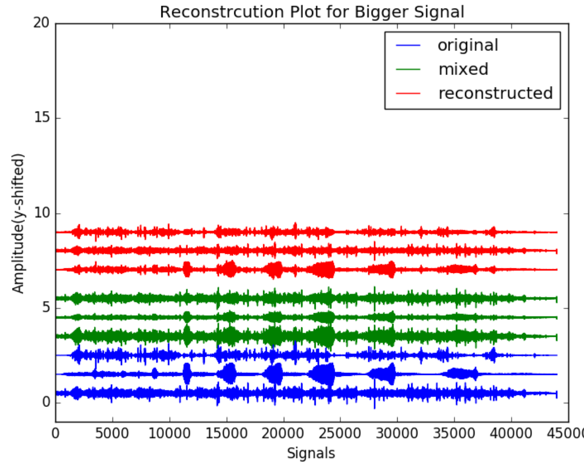
The experiment was conducted first for the smaller dataset for the proof of concept. The parameters chosen for the ICA algorithm for this smaller dataset were, $\eta = 0.01$, $k = 10000$. The **Figure 1**, shows the original, mixed and recovered signal in the bottom up way. The original and the reconstructed signals are normalized to lie between $[0, 1]$.

The exact correspondance (original $\rightarrow$ reconstruction), upon visualization is *top* $\rightarrow$ *bottom*, *middle* $\rightarrow$ *top*, **and** *bottom* $\rightarrow$ *middle*

As it can be visually examined, the reconstruction though not perfect, is surprisingly close. This shows, that given the data is sufficiently non-gaussian, we can get state-of-art reconstruction, for small signals.
Now, when ICA was evaluated on the larger signal dataset for $\eta = 0.01$ and $k = 10000$, for all 3 of the signals, we get the following reconstruction, as shown in the Figure 2(a) and 2(b). The reconstruction though not as accurate as on the smaller data, is still very descent. Listening to reconstructed sound one can hear:



(a) Signal Chosen: [2,3,4]



(b) Signal Chosen:[3,4,5]

Figure 2: Reconstruction plot for bigger signal ($f_s = 11025$)

	1	2	3
1	-0.1765	0.4394	**0.8807**
2	0.1155	**0.8984**	-0.4236
3	**0.9740**	-0.0397	0.2229

Table 2: Correlation Matrix (U_original,U_reconstructed) for [2,3,4]

	1	2	3
1	-0.1762	**0.9610**	0.2130
2	**0.9076**	0.0559	0.4158
3	-0.3845	-0.2586	**0.8861**

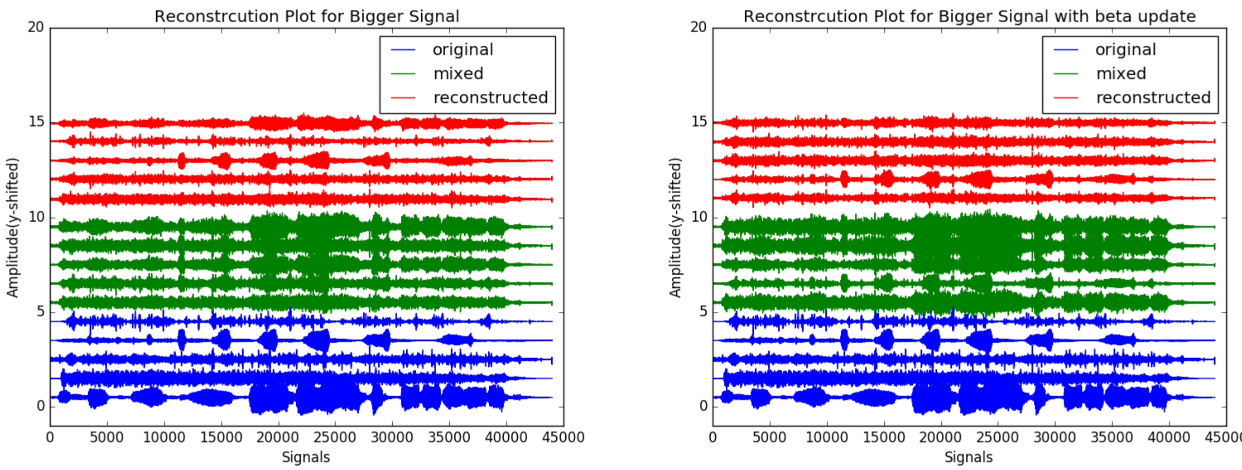
Table 3: Correlation Matrix (U_original,U_reconstructed) for [3,4,5]

- A person shouting about thermodynamics
- People clapping
- Some kind of printer sound
- Person laughing
- Crumbling of paper

Now the reconstruction for the entire set of 5 signals, when constructed using ICA was not that good, as can be seen in the Figure 3.a, for the same algorithm parameters of $\eta = 0.01$ and $k = 10000$.

Additional Experiment

Now I did an additional experiment of changing the β in the sigmoid function, $g(z) = \frac{1}{1+\exp -\beta z}$



(a) All signals without β update

(b) All signals with β update

Figure 3

	1	2	3	4	5
1	-0.2354	-0.1137	-0.2448	-0.2498	**-0.8995**
2	**-0.7397**	-0.5276	-0.0566	-0.2282	0.3450
3	0.5709	**-0.6662**	0.2546	-0.4062	-0.0181
4	0.2416	-0.2301	**-0.9114**	0.1820	0.1572
5	0.0589	0.4506	-0.2306	**-0.8303**	0.2252

Table 4: Correlation Matrix (U_original,U_reconstructed) for Fig 3.a

	1	2	3	4	5
1	**-0.5911**	-0.2623	-0.4630	-0.4452	0.4110
2	-0.3190	0.1965	**-0.6064**	0.5567	-0.4265
3	0.4931	-0.02030	-0.4695	**-0.5707**	-0.4584
4	0.1116	**0.8750**	-0.1770	-0.0996	0.4249
5	**0.5448**	-0.3514	-0.4044	0.3883	0.5149

Table 5: Correlation Matrix (U_original,U_reconstructed) for Fig 3.b

so that each signal has its own Cumulative distribution, which seems reasonable. The above algorithm was used with an outer β update given by,

$$\beta = \beta + \frac{\eta}{10} * column\_sum(Y * (1 - 2Z)) \tag{8}$$

where,

$$Y = WX \tag{9}$$
$$Z = g(Y) \tag{10}$$

The plot for the bigger data can be seen in Figure 3.b.

6 Observation & Justification

1. The reconstruction was close to original signal for smaller data. This can be clearly seen in the plots. The high boldened correlation factor in Table 1, shows that reconstruction was very accurate.

2. For the larger signal, the reconstruction was not as clear, and some mix remained in the output. This can be seen by a relatively less strong correlation factor in Table 2 and Table 3. However still it is 88 - 97 % correlation, which is pretty strong

3. The recovery of all 5 signals was less accurate than set of 3 signals, as can be seen in Table 4, where correlation lies in 66 - 91% range. This is reasonable, as more parameters (25 vs 9) are to be estimated

4. The β update did lead to a worse solution both visually and in correlation plot on Table 5. This was expected, as now convergence was desired in a 5D β space and 25D W space. The performance of gradient descent is known to deteriorate with increase in convergence space.

7 Conclusion

Thus we can easily conclude, based on experimental observations, that the Independent Component Analysis is a decent algorithm for approaching blind csource separation problem and β update is rather not required as it gives poor performance.

References

[1] Zarzoso, V., and A. K. Nandi. "Blind source separation." In Blind Estimation Using Higher-Order Statistics, pp. 167-252. Springer US, 1999.

[2] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. Independent component analysis. Vol. 46. John Wiley & Sons, 2004.

[3] Myung, In Jae. "Tutorial on maximum likelihood estimation." Journal of mathematical Psychology 47, no. 1 (2003): 90-100.

[4] Sigmoid Tutorial - http://www.computing.dcu.ie/ humphrys/Notes/Neural/sigmoid.html