Independent Component Analysis

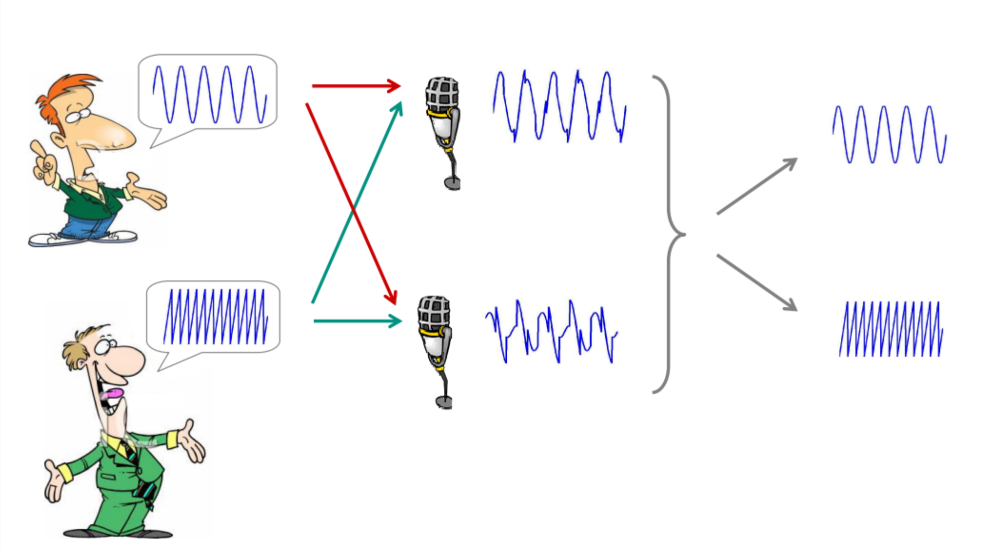
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Kunal Nema - MA20BTECH11007

Supriya Rawat - MA21MSCST11004

May 8, 2023

The Cocktail Party Problem



Blind Source Separation (BSS)

- An inductive inference problem.
- We **separate** and **estimate** the original source waveforms from the sensor array, without knowing the transmission channel characteristics and the source.
- Estimations are done upto certain indeterminacies as follows:
- Arbitrary scaling
- Permutation of resulting signals
- Delay of estimated source signals
- **Aim**-To process these observations so that the original source signals are extracted by the **adaptive system**. The whole process is done through the methods of **Linear Algebra** and **Statistics** at it's core.

Why to go for BSS techniques inspite of the indeterminacies?

1) The most relevant information about the source signals is contained in the **temporal waveforms** or time-frequency patterns of the source signals,
2) The amplitudes or the order of arrangement of the resulting waveforms in the output do not significantly enhance the result.

However, for some applications especially **biomedical signal models** such as **sEMG signals**, there is no guarantee that the estimated or extracted signals have exactly the same waveforms as the source signals.

Introduction to ICA

- One of the most widely used BSS techniques for revealing hidden factors that underlie sets of **random variables**, **measurements**, or **signals**.
- A method for extracting individual signals from mixture of signals

Core idea - Different physical processes generate unrelated signals.

ICA has a wide range of applications which are as follows: Image processing, Blind source separation, Financial forecasting, Medical imaging,etc.

ICA Model

Taking cocktail party problem as a reference,assuming we have sampled a mixture of spoken voices, the observations, we wish to separate each voice into a separate speaker channel. The common problems encoutered would be the time delay between microphones, echo, amplitude difference, voice order in speaker and underdetermined mixture signal.

An efficient separation could be done by reducing redundancy between signals,this observation lead to **Independent Component Analysis**.

ICA Model

Given N statistically independent signals, $s_i(t), i = 1, \dots, N$ to be **estimated**.

We obtain N **observation signals** $x_i(t), i = 1, \dots, N$.

The observation signals are obtained using sensors kept spatially well.

The mixing process with matrix multiplication as follows:

$$x(t) = A * s(t)$$

where $\boldsymbol{A}$ is an unknown matrix called the mixing matrix.

We have no information on the **mixing matrix**, or even on the sources themselves.

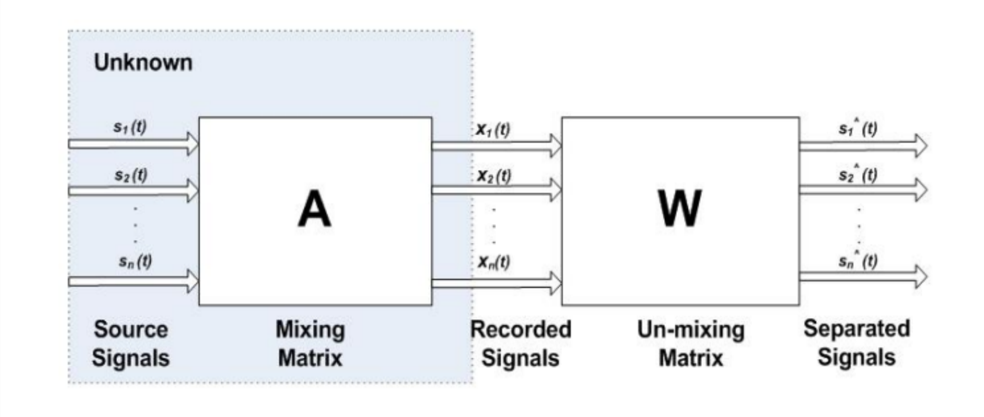
> The objective is to recover the original signals, $s_i(t)$, from only the observed vector $x_i(t)$.

We obtain estimates for the sources by first obtaining an **Unmixing Matrix**, **W**, where, $W = A^{-1}$.

This enables an estimate, $\hat{s}(t)$, of the independent sources to be obtained:

$$\hat{s}(t) = W * x(t)$$

ICA process



Unknown
$s_1(t)$
$s_2(t)$
$s_n(t)$
A
$x_1(t)$
$x_2(t)$
$x_n(t)$
W
$s_1^\wedge(t)$
$s_2^\wedge(t)$
$s_n^\wedge(t)$
Source Signals
Mixing Matrix
Recorded Signals
Un-mixing Matrix
Separated Signals

ICA assumptions

The sources being considered are **statistically independent**.

The independent components have **non-Gaussian** distribution.

Considering a Gaussian signal to be a linear combination of many independent signals, separation of independent signals from their mixtures can be accomplished by making the linear signal transformation as non-Gaussian as possible.

The mixing matrix **A** is invertible.

ICA ambiguities

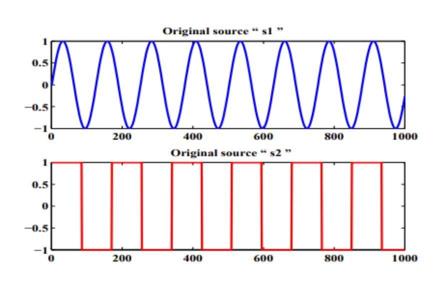
Magnitude and scaling ambiguity

We assume that the statistically independent sources have unit variance,so the inefficiency present due to difference in scales of data is dealt initially.

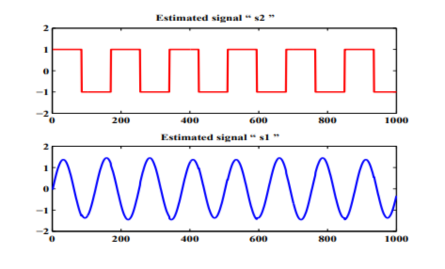
Permutation ambiguity

The order of the estimated independent components is unspecified.

Permutation ambiguity



Observation signal



Output signal

Pre-processing involved in ICA

Step 1: Centering data

We "**center**" the observation vector **x** by subtracting its mean vector **m** = E[x].

$$x_c = x - E[x]$$

It results in zero mean and via unmixing matrix, we can obtain the actual estimates of the independent components as follows: $\hat{s}(t)=\mathrm{A}^{-1}(x_c + m)$

Step 2: Whitening the data

Whitening involves linearly transforming the observation vector such that its components are uncorrelated and have unit variance.

$$E[xx'] = I$$

It is done using the **eigen-value decomposition** of the observation vector.

$$E\left\{xx^T\right\} = VDV^T$$

$$x_w = VD^{-1/2}V^T x$$

$$D^{-1/2} = D^{-1/2} = \text{diag}\left\{\lambda_1^{-1/2}, \lambda_2^{-1/2}, \dots, \lambda_n^{-1/2}\right\}.$$

$$x_w = VD^{-1/2}V^T As = A_w s$$

$$
\begin{aligned}
E\left\{x_w x_w^T\right\} &= A_w E\left\{ss^T\right\} A_w^T \\
&= A_w A_w^T \\
&= I
\end{aligned}
$$

Since the new obtained matrix A_w is orthogonal, Instead of estimating n^2 elements of the original matrix A, we only need to estimate the half of n (n-1) elements of the matrix A_w.

A statistical representation of ICA process

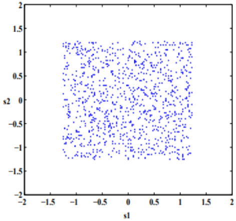


Figure 5: Original sources

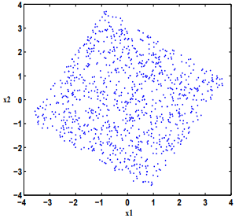


Figure 7: Joint density of whitened signals obtained fro
whitening the mixed sources

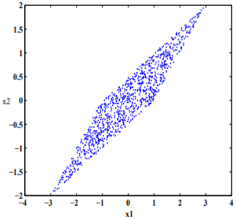


Figure 6: Mixed sources

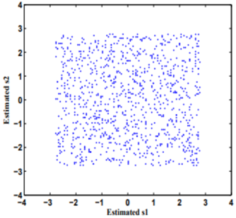


Figure 8: ICA solution (Estimated sources)

ICA algorithms: FastICA

- It is a **fixed-point** algorithm that uses higher order statistics to recover the independent components.
- FastICA can estimate ICs one by one (**deflation approach**) or simultaneously (**symmetric approach**) using the principle of **maximum entropy** on the mutual information.

Mutual information is a measure of independence between sources:

$$I(s) = \int f_s(s) \log \frac{f_s(s)}{\prod f_{s_i}(s_i)} ds$$

General layout of FastICA

The working of ICA ultimately depends on finding the **unmixing** matrix $\mathbf{W}$ to obtain the **estimates** of the original sources. FastICA is an efficient algorithm which converges quickly to the **unmixing** matrix.

Steps :

The cost function is given by:

$$\mathcal{J}_G = \sum_{i=1}^{n} \mathrm{E}\left\{G\left(\mathbf{w}_i^T \mathbf{z}\right)\right\}$$

where G is a nonlinear function which is usually assumed even and symmetrical. Here,the cost function is usually a measure of **non-gaussianity**, hence,it is to be maximized.

A widely used cost function is the fourth-order cumulant or **kurtosis**, defined for any random variable τ as

$$\mathrm{kurt}(\tau) = \mathrm{E}\left\{\tau^4\right\} - 3\left(\mathrm{E}\left\{\tau^2\right\}\right)^2$$

We restrict $\mathbf{w}_i^T\mathbf{z} = y_i$ (rows of $\mathbf{W}$) to have unit variance, and thus its kurtosis is $\mathrm{E}\left\{y_i^4\right\} - 3$.The second term being constant can be dropped, and the criterion becomes

$$\mathcal{J}_G^{\mathrm{kurt}} = \sum_{i=1}^{n} \mathrm{E}\left\{\left(\mathbf{w}_i^T\mathbf{z}\right)^4\right\}$$

The algorithm updates the basis vectors w_i in each iteration using a fixed point iteration rule, until the convergence criteria is met.

For one of the rows $\mathbf{w}_i^T$, the FastICA algorithm for kurtosis maximization makes the basic updating step

$$\mathbf{w}_{\text{new}} = \mathbf{w} - \eta\left(\mathbf{x}\left(-4(\mathbf{w}^T\mathbf{x})^3 + 6(\mathbf{w}^T\mathbf{x})\right)\right) \tag{1}$$

where $\mathbf{w}$ is the current estimate of the unmixing matrix, $\mathbf{x}$ is the preprocessed mixed data, η is the learning rate, and $\mathbf{w}_{\text{new}}$ is the updated estimate of the unmixing matrix.

It is followed by normalization of vector $\overline{\mathbf{w}}_i$ to unit norm. This updation is done until the matrix $\mathbf{w}$ converges.

Relevance of FastICA

- FastICA is widely used in signal processing, image processing, and machine learning.
- It can be used for speech separation, blind image deconvolution, image denoising, and feature extraction.
- FastICA has advantages over other blind source separation algorithms, such as PCA and ICA with Gaussian assumption, because it does not assume a specific distribution of the sources.
- FastICA can handle both underdetermined and overdetermined mixtures, and can separate more sources than the number of observations.

Infomax

The Infomax (Information Maximization) principle is a principle in information theory based on the maximization of the mutual information between its inputs and outputs.It has several application in for **blind source separation, feature extraction**, and **representation learning**.

Steps involved in Infomax:

- We first center and whiten the data.
- To estimate the **unmixing** matrix,we use **gradient ascent** method.
- We start by assuming a random unmixing matrix **W** and continuously iterate it to achieve the optimal solution.

- Here, W is the random matrix to be estimated $Y=W*S$.
- $I(Y;S) = H(Y) - H(Y|S)$ is the mutual information to be maximized, where, $H(Y)$ and $H(S)$ are the entropies of Y and S respectively, and $H(Y|S)$ is the conditional entropy of Y given S.
- We use gradient ascent algorithm to similarly achieve the maxima of this function.

ICA for different conditions

One of the important assumptions for ICA is number of sensors is equal to the number of sources. So when this constraint is not satisfied, we have

- Overcomplete ICA
- Undercomplete ICA

Overcomplete ICA

No. of sources (n) > No. of sensors (m)

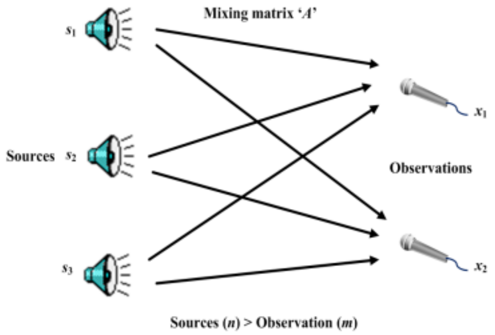


Figure 9: Illustration of "overcomplete ICA"

Sources: $s_1(t) \quad s_2(t) \quad s_3(t)$

Recordings: $x_1(t) \quad x_2(t)$

The $a_{ij} \in A$ are constant coefficients that give the mixing weights.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$$

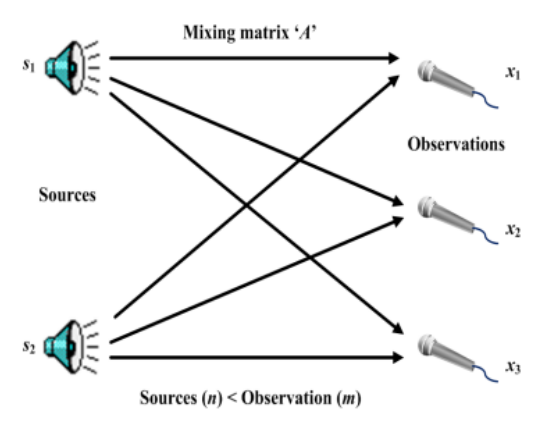
The unmixing process and estimation of sources can be written as

$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

In this example matrix A of size 2×3 matrix and unmixing matrix W is of size 3×2. Hence in overcomplete ICA it always results in pseudoinverse. Hence computation of sources in overcomplete ICA requires some estimation processes.

Undercomplete ICA

No. of sources (n) $<$ No. of sensors (m)



Sources: $s_1(t) \quad s_2(t)$

Recordings: $x_1(t) \quad x_2(t) \quad x_3(t)$

The $a_{ij} \in$ A are constant coefficients that give the mixing weights.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{23} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

- We have redundancy in the observation vector
- Dimension Reduction using PCA:

$$(z)_{2x1} = (V)_{2x3}(x)_{3x1}$$

- Now, standard ICA can be applied:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

- The unmixing process and estimation of sources can written as

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Sub band decomposition ICA

Problem: A signal may have different spectral characteristics in different frequency bands.

Introduction to sub band decomposition:

- Technique that divides a signal into a number of frequency sub-bands, each representing a different frequency range.
- The sub-bands can be processed separately, providing a more efficient way to analyze and manipulate signals.
- The sub-bands can be processed separately, providing a more efficient way to analyze and manipulate signals.

Implementation:

- The mixed signal is first decomposed into sub-bands using a filter bank.
- Independent component analysis (ICA) is then applied to each sub-band to separate the sources.
- The separated sources from each sub-band are then combined to obtain the original signals.

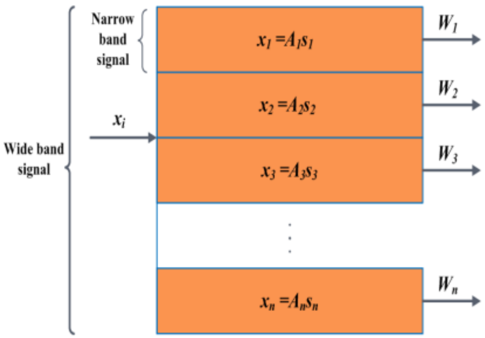


Figure 11: Sub band ICA block diagram.

Sub-band Decomposition

Limitations of ICA

- **Linear assumptions**- ICA assumes that the sources are linearly mixed and that the mixing process is governed by multiplying with a matrix,hence the resulting function is always linear,which may not be true in real-world situations.
- **Computationally expensive**: ICA can be computationally expensive, especially when dealing with high-dimensional data. The **FastICA** algorithm helps in alleviating this problem, but it can still be slow for very large datasets.
- **Model selection**: ICA requires the user to specify the number of sources or components to be estimated, which can be challenging in practice. Choosing an incorrect number of sources can lead to poor separation results,hence the results in **undercomplete** and **overcomplete** can be affected.

- **Non-Gaussianity assumption**: ICA assumes that the sources are non-Gaussian, which may not be true in all cases. If the sources are Gaussian, ICA may not be the best method for source separation.
- **Identifiability**: ICA suffers from the problem of identifiability, meaning that the estimated sources are only unique up to a permutation and scaling. In other words, ICA cannot determine the order of the sources, and the amplitudes of the sources can only be estimated up to a scaling factor.

Reconstructive-ICA

- Reconstructive ICA is a technique used to solve the ambiguities involved in the ICA problem, such as scaling, permutation, and sign ambiguities.
- It involves reconstructing the original sources from the estimated mixing matrix and the estimated independent components.
- Reconstructive ICA is useful in scenarios where the sign or scaling of the sources is important, such as in signal processing applications where the sign of the signal indicates the direction or polarity of the signal.

Applications of ICA

Image Enhancement: ICA can be used to enhance images by decomposing them into independent components and manipulating them individually. For example, we can use ICA to remove noise from an image by identifying and removing components that correspond to noise.

Biomedical Applications: Separation of EEG (electroencephalogram) signals from different brain regions.It is also useful in separation of fMRI (functional magnetic resonance imaging) signals for brain mapping.

Other Applications

- Audio signal processing
- Telecommunications
- Speech separation and recognition
- Seismic data analysis

References

Dinesh Kumar Ganesh Naik.
Overview of ICA its Applications:.
Researchgate Publication, 2011.