Independent Component Analysis (ICA)

Bhaskar D Rao
University of California, San Diego
Email: brao@ucsd.edu

References

1. Hyvarinen, A., Karhunen, J., & Oja, E. (2004). Independent component analysis (Vol. 46). John Wiley & Sons.
2. Hyvrinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." Neural networks 13.4-5 (2000): 411-430.
3. Adali, Tulay, Matthew Anderson, and Geng-Shen Fu. "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging." IEEE Signal Processing Magazine 31.3 (2014): 18-33.
4. Novey, Michael, and Tulay Adali. "Complex ICA by negentropy maximization." IEEE Transactions on Neural Networks 19.4 (2008): 596-609.

ICA

Instantaneous Mixing Model: Assume N sensors, D sources with $D < N$, that are linearly mixed

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{Z}[n] or \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_N[n] \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1D} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ {}_{N1} & a_{N2} & \dots & a_{ND} \end{bmatrix} \begin{bmatrix} s_1[n] \\ s_2[n] \\ \vdots \\ s_D[n] \end{bmatrix} + \mathbf{Z}[n]$$

Problem: Both mixing matrix $\mathbf{A}$ and sources $\mathbf{s}[n]$ are unknown. From measurements of $\mathbf{x}[n]$, $n = 0, 1, ..., L-1$, recover $\mathbf{A}$ and $\mathbf{s}[n]$.
Assumptions:

- Sources are zero mean, i.e. $E(s_l[n] = 0, l = 1, .., D$.
- Sources are uncorrelated, i.e. $E(\mathbf{s}[n]\mathbf{s}^T[n]) = \text{diag}(p_l)$.
- Sources are independent, i.e. $p(s_1, s_2, .., s_D) = \Pi_{l=1}^{D} p_l(s_l)$.
- Sources are non-Gaussian distributed except for possibly one.

ICA relies on independence and non-Gaussian source assumption.

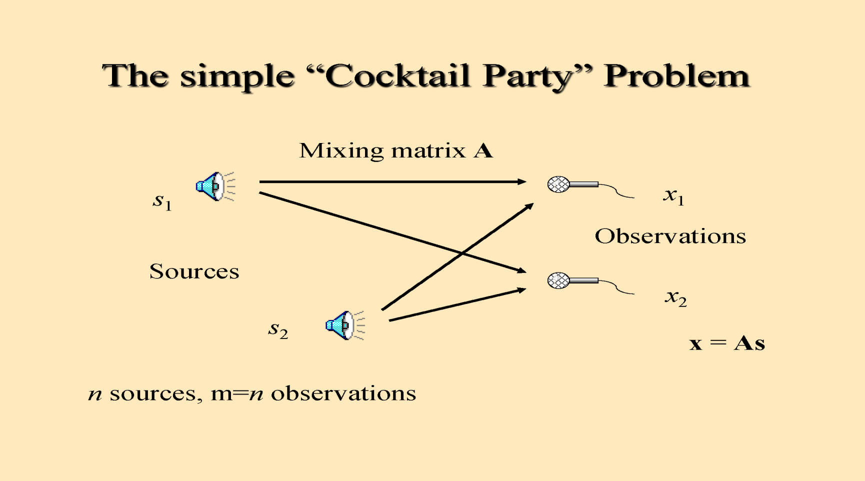
ICA and Applications

More general model is Convolutive Mixing: $x_l[n] = \sum_{k=1}^{D} h_{lk}[n] * s_k[n]$

ICA has several practical applications:

- The "cocktail party problem": separation of voices or music or sounds
- Sensor array processing, e.g. radar
- Biomedical signal processing with multiple sensors: EEG, ECG, MEG, fMRI
- Telecommunications: e.g. multiuser detection in CDMA
- Financial and other time series
- Noise removal from signals and images
- Feature extraction for images and signals
- Brain modelling

Cocktail Party Problem



ICA and EEG Processing[1]

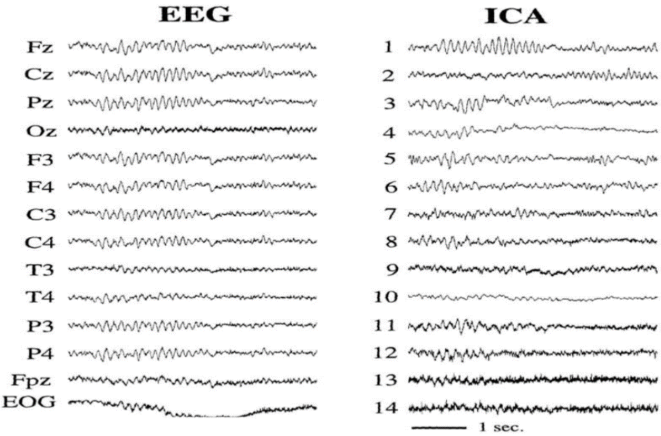


Figure 1: Left: 4.5 seconds of 14-channel EEG data. Right: an ICA transform of the same data, using weights trained on 6.5 minutes of similar data from the same session.

[1]Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In Advances in neural information processing systems (pp. 145-151).

Principal Component Analysis(PCA)

PCA is a data representation/ compression technique: Finds an efficient low-dimensional orthonormal basis to represent a random vector in a mean squared sense.

$$\mathbf{R}_{xx} = E(\mathbf{x}[n]\mathbf{x}^H[n]) = APA^H + \sigma_z^2\mathbf{I}, \text{ where } P = \text{diag}(p_l)$$

Eigen-decomposition should allow us to find the number of sources D and the noise variance.
Let us assume $N = D$ and $\sigma_z^2 = 0$.
Problem: Find an orthonormal basis W for a p-dimensional subspace ($p \leq N$), i.e $W \in C^{N\times p}$ and $W^H W = \mathbf{I}_{p\times p}$, such that $E(|W^H\mathbf{x}[n]|^2)$ is maximized.

$$\max_{W\in C^{N\times p}, W^H W = \mathbf{I}_{p\times p}} W^H \mathbf{R}_{xx} W$$

Solution (PCA): $W = [\mathbf{q}_1, .., \mathbf{q}_p]$, where $\mathbf{q}_l$ is an eigenvector of $\mathbf{R}_{xx}$ corresponding to the lth largest eigenvalue. The approximation for $\mathbf{x}[n]$ is obtained as

$$\hat{\mathbf{x}}[n] = WW^H\mathbf{x}[n]$$

PCA and Whitening

If $\mathbf{E} = [\mathbf{q}_1, .., \mathbf{q}_N]$, then $\mathbf{R}_{xx} = \mathbf{E}\Lambda\mathbf{E}^H$, is an eigen-decomposition where $\Lambda = \text{diag}(\lambda_l)$ contains the ordered eigenvalues, i.e. $\lambda_1 \geq \lambda_2 ... \geq \lambda_N \geq 0$.
Then $B = \Lambda^{-\frac{1}{2}}\mathbf{E}^H$ is a whitening matrix, i.e. $B\mathbf{R}_{xx}B^H = \mathbf{I}_{N\times N}$ and $\mathbf{y}[n] = B\mathbf{x}[n]$ has uncorrelated components.
If B is a whitening matrix, QB is also a whitening matrix where Q is an orthonormal matrix and $\mathbf{y}_Q[n] = QB\mathbf{x}[n]$ also has uncorrelated components.
Implications

- Gaussian sources are not separable: If the sources $(\mathbf{s}[n])$ are Gaussian, $\mathbf{y}[n]$ and $\mathbf{y}_Q[n]$ represent independent Gaussian sources. Non-unique
- Uncorrelated property is not enough to separate sources

Need Independent and Non-Gaussian sources: The goal is to find Q, the remaining unknown.

Ambiguities/Indeterminacies in ICA problem $\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n]$

1. We cannot determine the variances (energies) of the independent components (Scaling).
   - The reason is that we can always write $\mathbf{x}[n] = \mathbf{A}DD^{-1}\mathbf{s}[n]$, where D is a diagonal matrix
   - Therefore, the magnitudes of the $s_i[n]$ can be freely normalized, typically unit variance.
   - This still leaves the sign of $s_i[n]$ undetermined. Fortunately, often the waveform $s_i[n]$ is of interest.
2. We cannot determine the order of the independent components. (Permutation)
   - This is because we can always change the order in the sum and call any one of the IC's the first one.
   - Mathematically this means instead of the original $\mathbf{s}[n]$, we can only obtain $DP\mathbf{s}[n]$ where D is any nonsingular diagonal scaling matrix and P is any permutation matrix
   - Note that the elements of vector $DP\mathbf{s}[n]$ are independent if and only if the elements of $\mathbf{s}[n]$ are independent.

ICA Methods

Assumption: All random variables are zero mean, number of sensors equal to sources, i.e.$N = D$, and $\mathbf{A}$ is square and invertible.

Maximum Likelihood Estimate: Assume a certain non-Gaussian prior on the sources and compute a ML estimate of the mixing matrix $\mathbf{A}$. Once you have $\mathbf{A}$, can demix the sources.

Independence criteria

- Random variable Y_1 and Y_2 are independent implies $p_{Y_1 Y_2}(y_1, y_2) = p_{Y_1}(y_1) p_{Y_2}(y_2)$. Also $E(G(Y_1)H(Y_2)) = E(G(Y_1))E(H(Y_2))$, for arbitrary $G(.)$ and $H(.)$. Non-linear decorrelation.
- Kullback-Liebler distance $KL(p|q) = \int p(y) \ln \frac{p(y)}{q(y)} dy \geq 0$: Use $KL(p(s_1, .., s_D)|p(s_1)p(s_2)..p(s_D))$ to measure independence

Non-Gaussianity: Maximize distance from Gaussianity

- Kurtosis : $\text{kurt}(X) = E(X^4) - 3(E(X^2))^2$
- Neg-Entropy: $J(X) = H(X_{Gauss}) - H(X) \geq 0$, where $H(X)$ is the differential entropy of the random vector X and X_{Gauss} is a Gaussian random vector with same covariance matrix as X.