Cocktail Party Problem: retrieve original speeches from microphone recs

Microphone recordings *x*

Original Speeches *s*

At each mic we have a composition of orginal speeches (i.e. sources)

$$x_1 = a_{11}s_1 + a_{12}s_2 + \cdots + a_{15}s_5$$

Cocktail Party Problem: retrieve original speeches from microphone recs

Given all the recordings, we have that

$$x_1 = a_{11}s_1 + a_{12}s_2 + \cdots + a_{15}s_5$$
$$x_2 = a_{21}s_1 + a_{22}s_2 + \cdots + a_{25}s_5$$
$$\vdots$$
$$x_5 = a_{51}s_1 + a_{52}s_2 + \cdots + a_{55}s_5$$

A matrix notation can be used

$$x = As$$

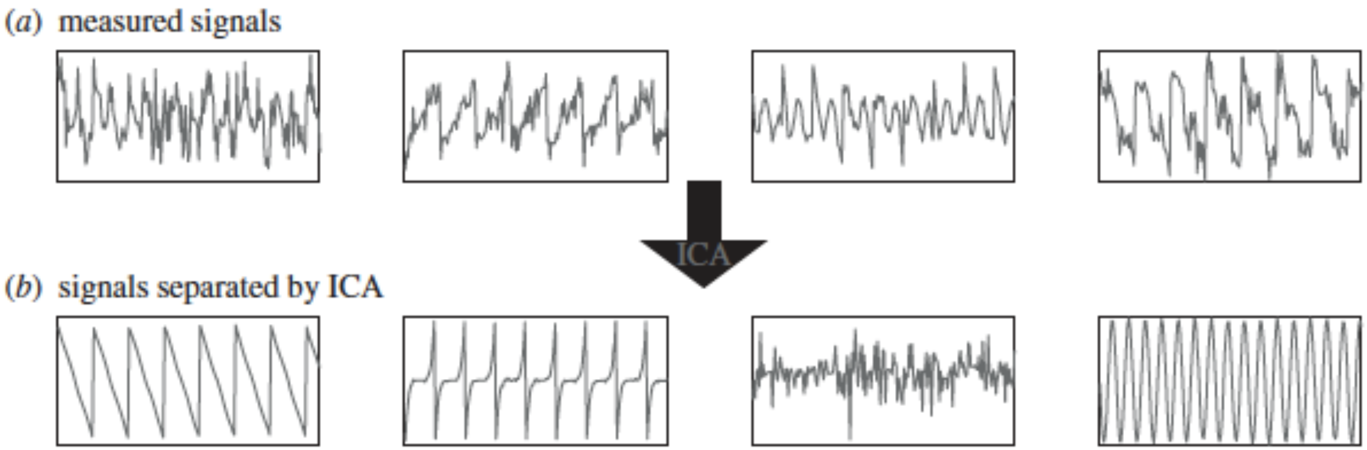
Source estimates

$$\hat{s} = Wx$$

... good when

$$W \cong A^{-1}$$

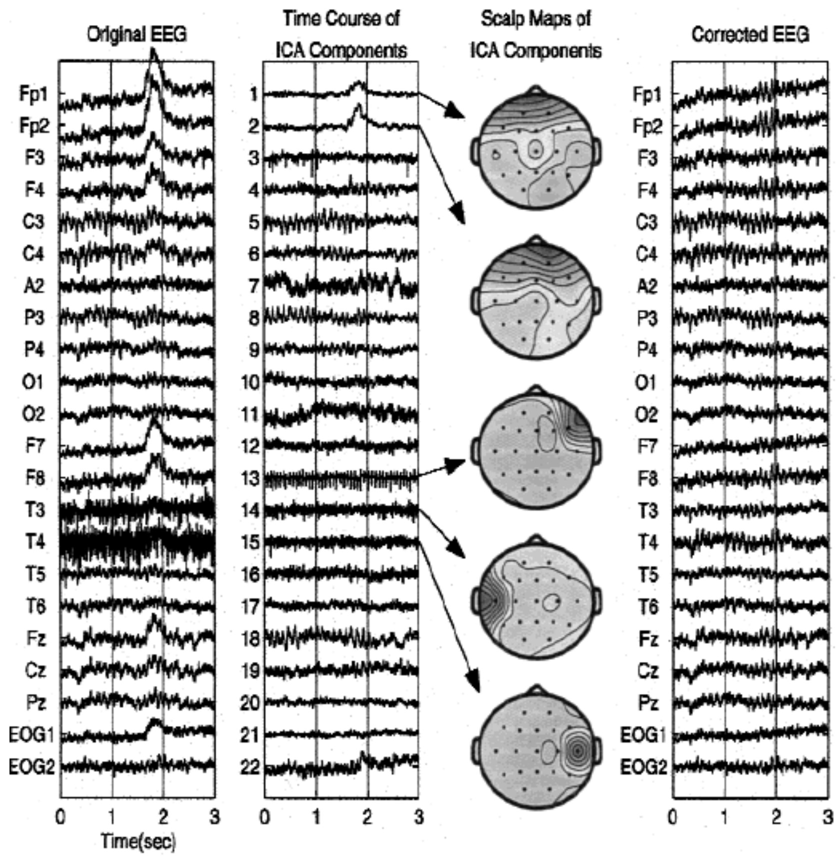
Independent Component Analysis



Hyvarinen A. 2013 Independent component analysis: recent advances. Phil Trans R Soc A 371: 20110534

Independent Component Analysis:
EEG Artifacts Reduction

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M.J., Iragui, V., Sejnowski, T.J. Removing electroencephalographic artifacts by blind source separation (2000) Psychophysiology, 37 (2), pp. 163-178. cited 1541 times........



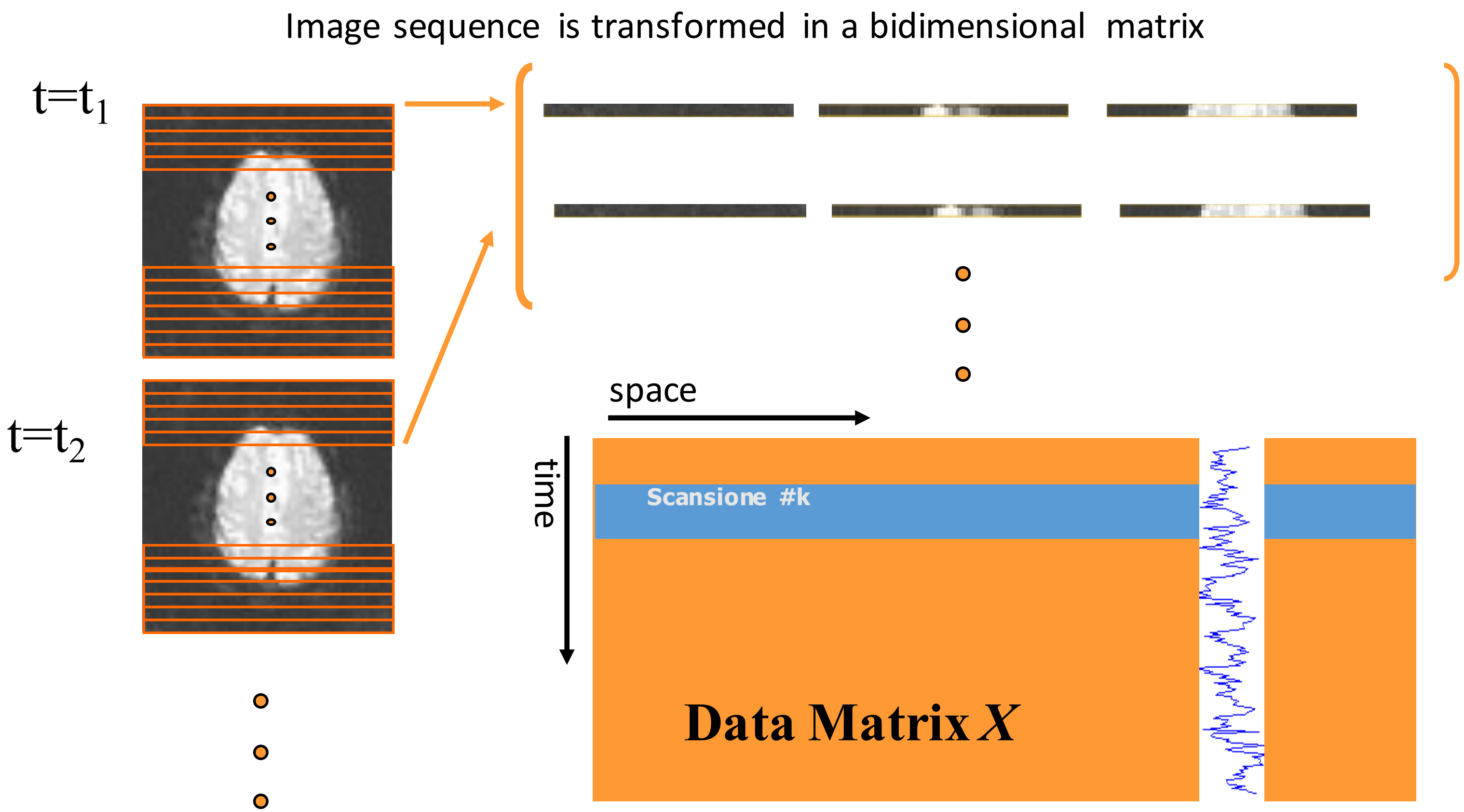
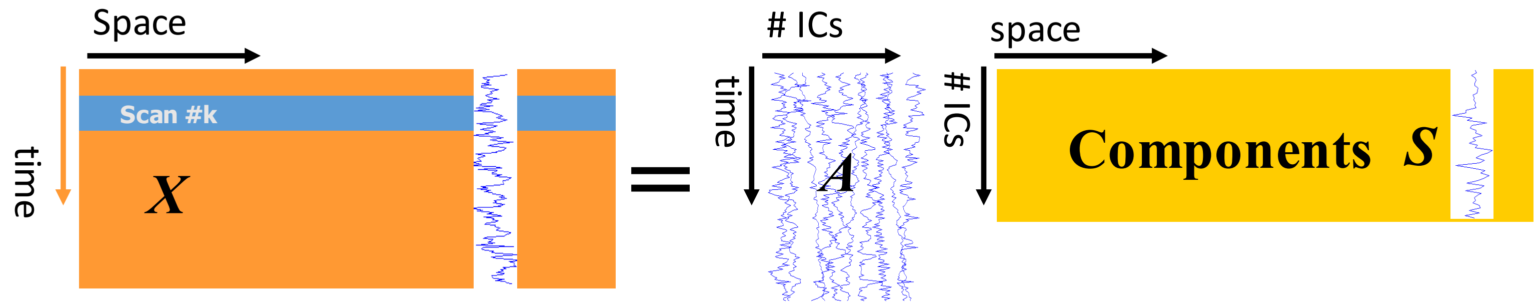


Image sequence is transformed in a bidimensional matrix
t=t1
t=t2
space
time
Scansione #k
Data Matrix X

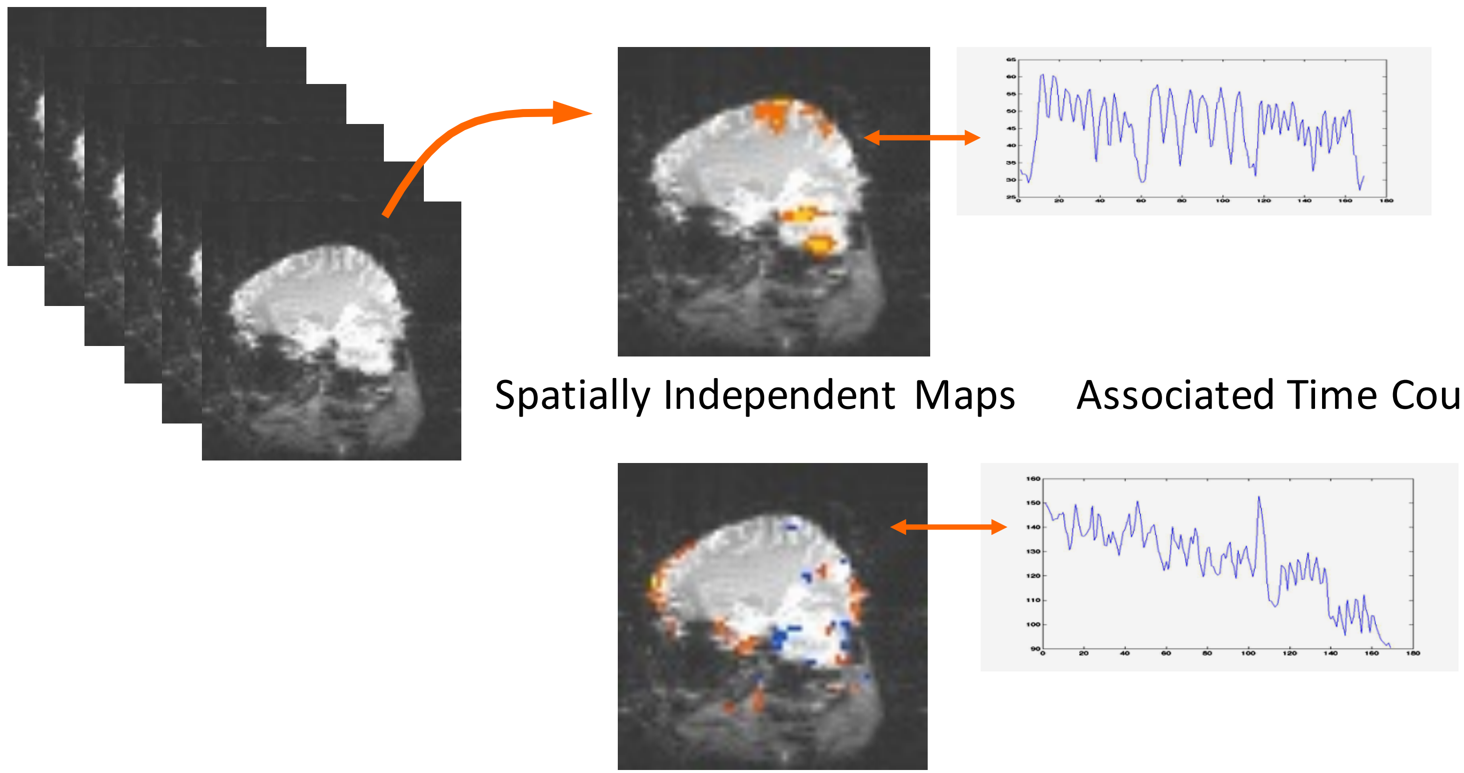
Independent Component Analysis



X Data Matrix
A Mixing Matrix
S Independent Component Matrix

$$X=AS$$

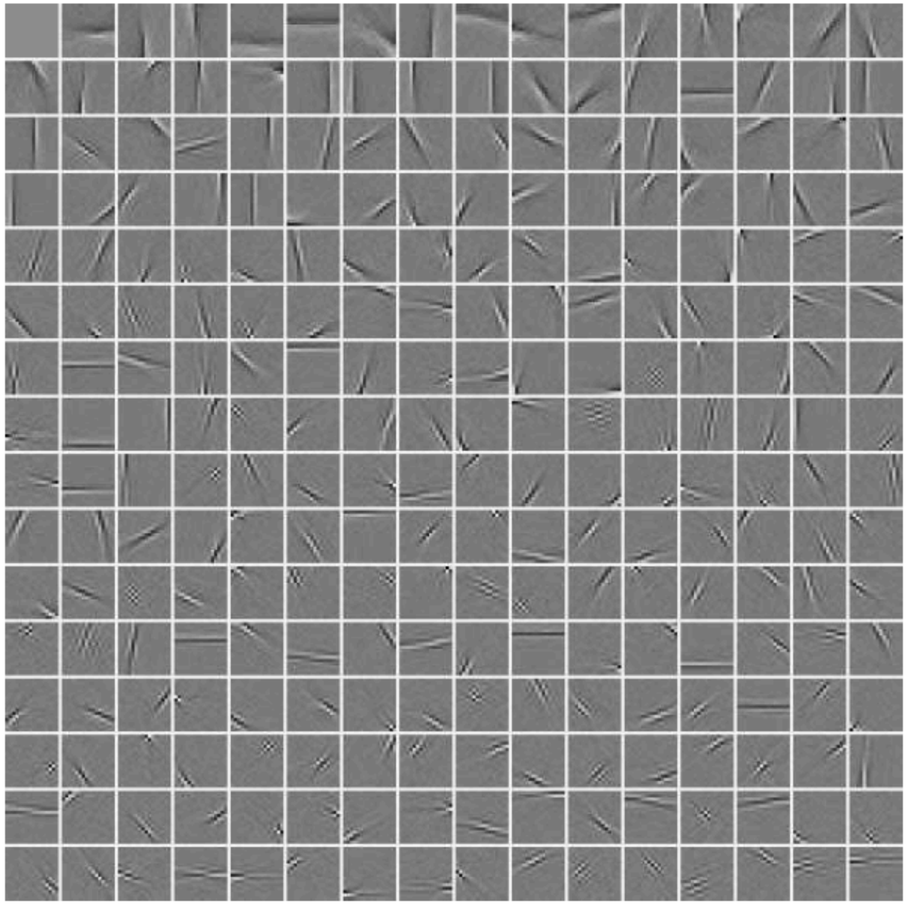
Independent Component Analysis



Is statistical independence relevant?

Independent components estimated from the analysis of natural images

Do they look like receptive fields in V1?



A. Hyvarinen, P. Hoyer and M. Inki, "Topographic ICA as a model of V1 receptive fields," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000.* Como, 2000, pp. 83-88 vol.4.

…. we suggest that the multifocal brain areas activated by performance of a psychomotor task should be unrelated to the brain areas whose signals are affected by artifacts, such as physiological pulsations, subtle head movements, and machine noise which may dominate fMRI experiments.

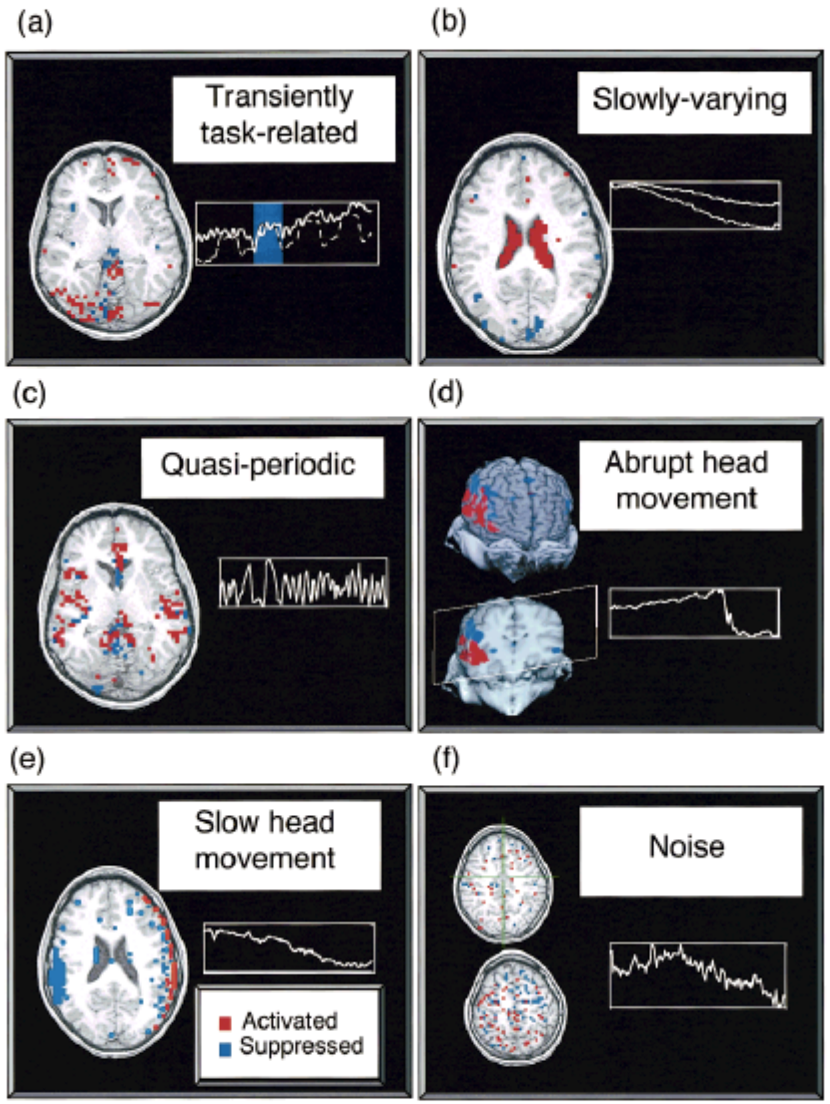


Figure 15.

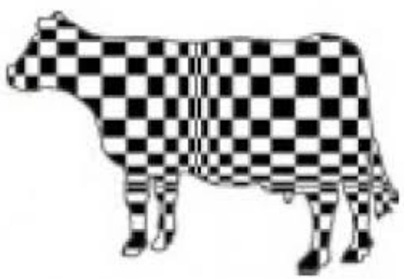
McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.-P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J. Analysis of fMRI data by blind separation into independent spatial components (1998) Human Brain Mapping, 6 (3), pp. 160-188

How Estimation Can Be Performed

Then the direction where the mutual information between each pair of estimate is minimal, is found

$$I\left(s_i, s_j\right) \ \forall i, j$$

$$\hat{s}_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{15}x_5$$
$$\hat{s}_2 = w_{21}x_1 + w_{22}x_2 + \cdots + w_{25}x_5$$
$$\vdots$$
$$\hat{s}_5 = w_{51}x_1 + w_{52}x_2 + \cdots + w_{55}x_5$$



Harald Stögbauer, Alexander Kraskov, Sergey A. Astakhov, and Peter Grassberger, "Least Dependent Component Analysis Based on Mutual Information", Phys. Rev. E 70 (6) 066123, 2004

Mutual information of a random vector:

Do we have some advantage describing the variables all together?

$$\mathrm{I}(y_1, y_2, \dots, y_n) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{H}(y_i) - \mathrm{H}(\boldsymbol{y})$$

Entropy is an information related concept:

$$\mathrm{H}(\mathrm{x}) = \int_{-\infty}^{+\infty} \mathrm{f_x}(\mathrm{x}) \log\big(\mathrm{f_x}(\mathrm{x})\big) \mathrm{dx}$$

For instance is related to coding length
(how many bits we have to use the possible states of the brain)

Saxe GN, Calderone D, Morales LJ Brain entropy and human intelligence: A resting-state fMRI study. PLoS One. 2018 Feb 12;13(2):e0191582.

What we need to know about a random variable?

$$\mathrm{H(x)} = \int_{-\infty}^{+\infty} \mathrm{f_x(x)\log\big(f_x(x)\big)dx}$$

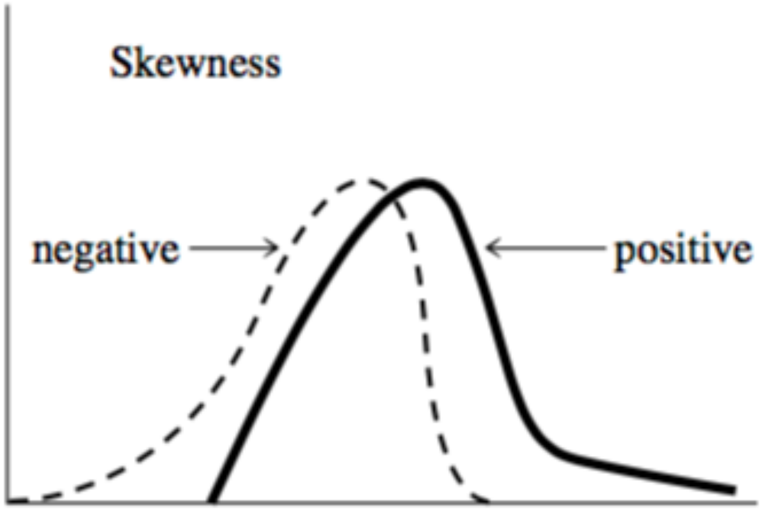
$f_x(x)$ Probability density function.... i.e. everything

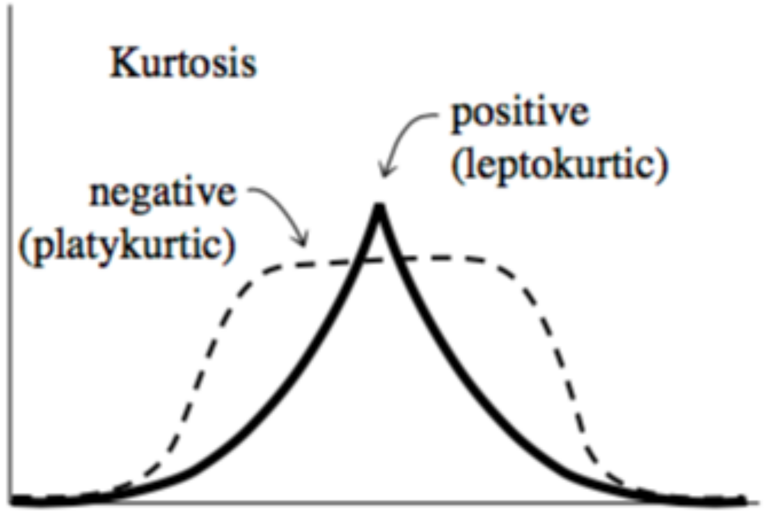
If we don't know we have to estimate it from the data...... How?

$f_x(x)$

We can estimate

Mean value : position index
Standard deviation: dispersion index
Skewness: tilt of the distribution, asymmetry
Kurtosis: peaked or smooth





Images from :Numerical Recipes in C: The Art of Scientific Computing

Kurtosis and skewness are the fourth and third order cumulants, respectively

Cumulants are the coefficients of the Taylor expansion of the second generating functions of the PDF (the logarithm of the Fourier Transform of the PDF):
They can be used to describe the statistical properties of a variable.

They have very nice properties they are preferred with respect to central moments

$$k_1 = E\{x\}$$
$$k_2 = E\{x^2\} - [E\{x\}]^2$$
$$k_3 = E\{x^3\} - 3E\{x^2\}E\{x\} + 2[E\{x\}]^3$$
$$k_4 = E\{x^4\} - 3[E\{x^2\}]^2 - 4E\{x^3\}E\{x\} + 12E\{x^2\}[E\{x\}]^2 - 6[E\{x\}]^4$$

Decorrelation: it is possible to diagonalize the covariance matrix. Covariances are zero
And only diagonal elements (variances) are different from zero

Eigen-value decomposition of the covariance matrix can be used

To estimate independent components decorrelation is not enough
It can be possible to estimate higher order cumulants of the observations and transform such that they become zero

(this is to Tensorial methods like JADE)

The Central Limit Theorem

The pdf of a sum of random variables (unders some hypotheses)
Tends towards a gaussian distribution

If we start from independent and non gaussian sources s

The sum is more gaussian than the sources

$$x_1 = a_{11}s_1 + a_{12}s_2 + \cdots + a_{15}s_5$$
$$x_2 = a_{21}s_1 + a_{22}s_2 + \cdots + a_{25}s_5$$
$$\cdot$$
$$\cdot$$
$$x_5 = a_{51}s_1 + a_{52}s_2 + \cdots + a_{55}s_5$$

$$x = As$$

Mixing Process

$$\hat{s} = Wx$$

Source estimates

$$\hat{s}_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{15}x_5 =$$
$$= w_{11}\,(a_{11}s_1 + a_{12}s_2 + \cdots + a_{15}s_5) + w_{12}(a_{21}s_1 + a_{22}s_2 +$$

Each estimate is a combination of the original sources:

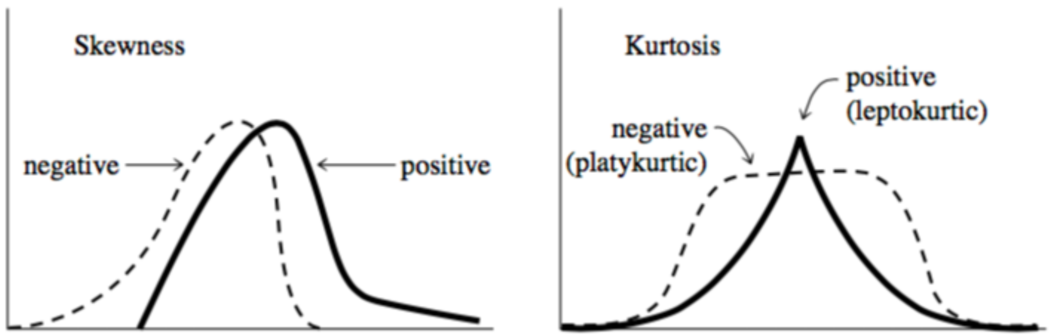
is least gaussian when the sum comprises only one source

$$\hat{s}_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{15}x_5 =$$
$$= w_{11}\,(a_{11}s_1 + a_{12}s_2 + \cdots + a_{15}s_5) + w_{12}(a_{21}s_1 + a_{22}s_2 + \cdots + a_{25}s_5) + \cdots =$$

$$\hat{s} = Wx$$

We change *W* till we maximize nongaussianity of the estimastes

Again we have to work with all those statistical measures that can distinguish Non gaussian variables from gaussian



Several algorithms work by optimizing non gaussianity…..

And even those that are supposed to behave differently are somehow similar

FastICA maximizes negentropy $\mathrm{J}(\mathrm{x}) = \mathrm{H}(\mathrm{x}_{\mathrm{gauss}}) - \mathrm{H}(\mathrm{x})$

.. But even in this case we have to estimate higher order moments or cumulants…… (kurtosis, skewness......)

It is possible to weight differently higher order moments

By using non linear transformations of the variables

$$J(\mathbf{x}) \approx \sum_{i=1}^{p} k_i \left[E\{G^i(\nu)\} - E\{G^i(x)\} \right]^2$$

$E\{x\}$ is the expectation operator

While $G^i(\nu)$ are non polynomial functions

Is is possible to think about a Taylor expansion of these ... So that we get the higher order moments properly weighted

$$G(x) = G(\bar{x}) + \left.\frac{\partial G(x)}{\partial x}\right|_{x=\bar{x}} (x-\bar{x}) + \left.\frac{\partial^2 G(x)}{\partial x^2}\right|_{x=\bar{x}} \frac{(x-\bar{x})^2}{2!} + \cdots$$

So if you take the $\mathrm{E}\{\mathrm{G}^{\mathrm{i}}(\nu)\}$ ......

You estimate the higher order moments weighted by some coefficients
That are related to the local properties of the nonlinear function

Several nonlinear functions can be chosen

Smooth nonpolinomial even functions are prefererred: do not have to increase too quickly

The infomax consists in maximizing the entropy of the output of some Neural networks

$$\mathrm{y}(x) = g_i\left(b_i^T \boldsymbol{x}\right) + n$$

g_i are nonlinear functions
b_i^T are the weights of the network

It can be shown that maximizing the output entropy

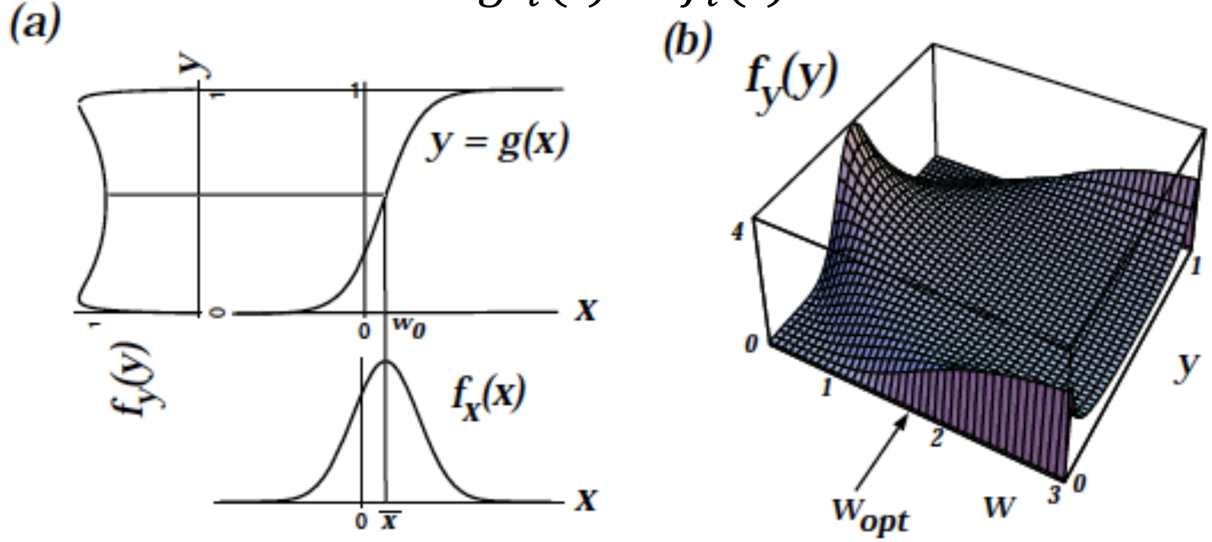
$$H(\boldsymbol{y}) = H\left(g_1(b_1^T\boldsymbol{x}), g_2(b_2^T\boldsymbol{x}), \dots . g_n(b_n^T\boldsymbol{x})\right)$$

Is equivalent to minimize the mutual information among the output variables

Bell, A.J., Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. (1995) Neural computation, 7 (6), pp. 1129-1159.

Noticeably,
The cumulative density function of the original sources should be matched with the nonlinearity

$$g'_i(.) = f_i(.)$$



The high sloping parts of g(x) are aligned with the regions where the pdf is higher

Extended Infomax

This approach can switch between
Two non linearities to optimize estimation of
Supergaussian (unimodal nonlinearity)
And subgaussian (bimodal nonlinearity)

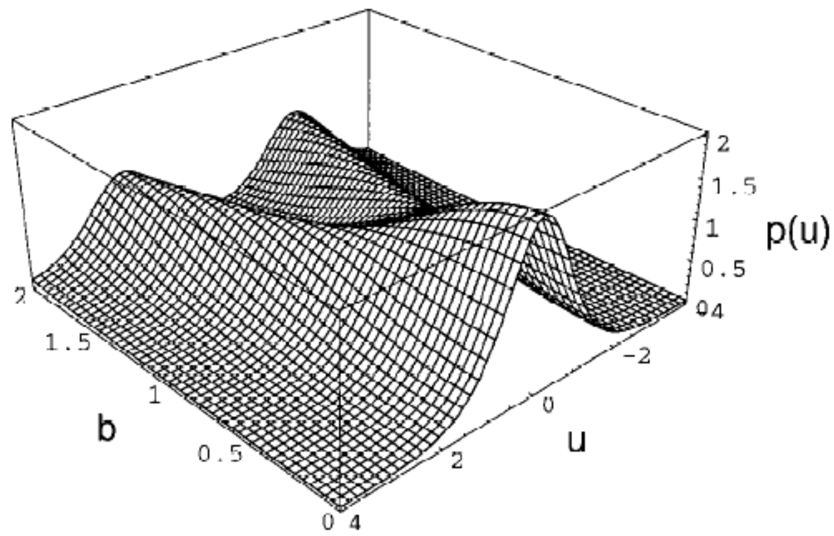


Figure 3: $p(u)$ as a function of b. For $b = 0$ the density estimate is suited to separate supergaussian sources. If, for example, $b = 2$ the density estimate is bimodal and therefore suited to separate subgaussian sources.

Lee et al., Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources, Neural Comput, 1999