APPLIED MACHINE LEARNING

Independent Component Analysis (ICA)

Principle

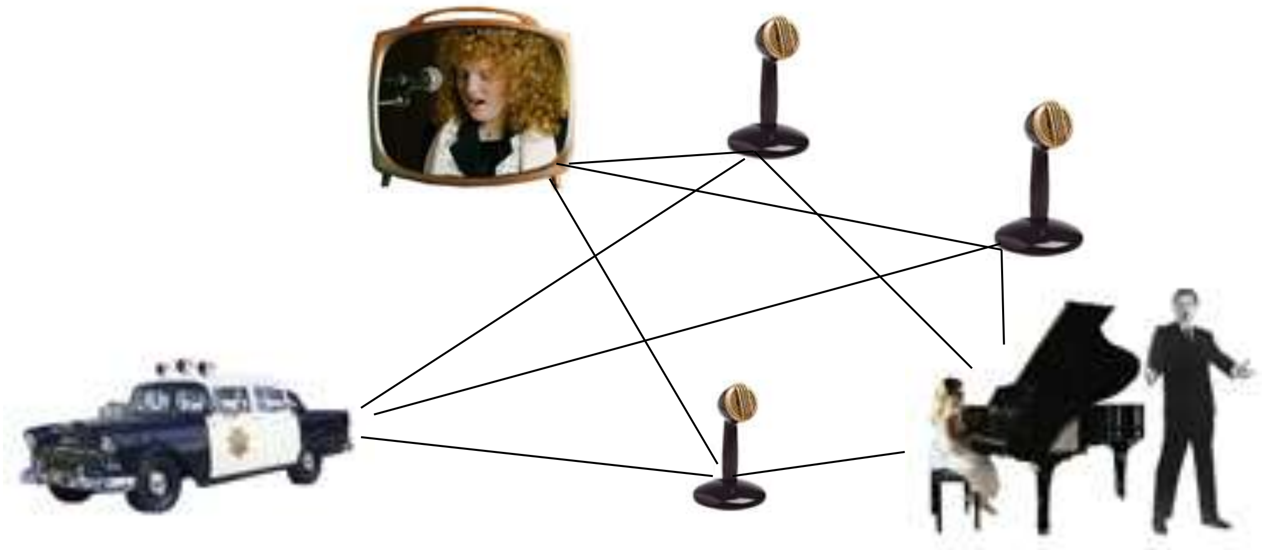
Cocktail Party Problem:

- You are in a room where two people are speaking simultaneously.
- How can you tell what each group of people say?

Cabrera Research

Blind Source analysis:

- **N** different sound sources
- **N** microphones located in three different locations in the room
- Each microphone records a sound **mixing** the three sound sources



Goal: Unmix the recorded signals and estimate the original sound signals

Original Signals - Sources (independent components)

Mixed signals picked up by 3 microphones

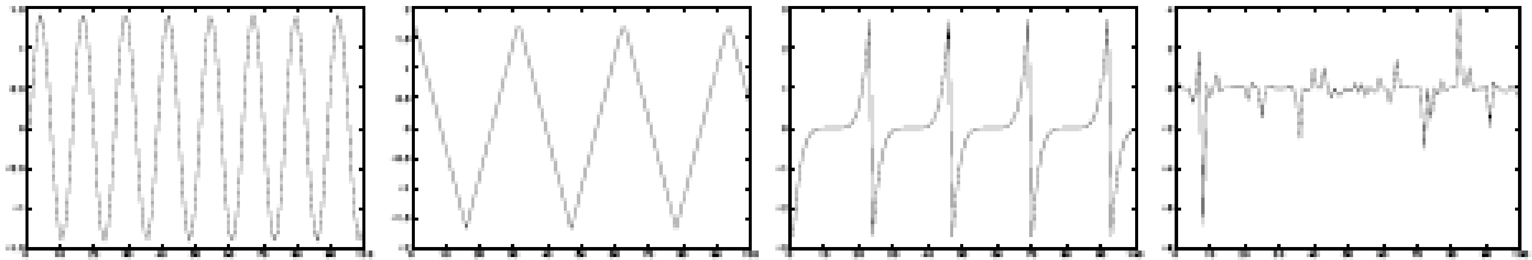
Reconstructed signals by ICA

Blind Source Analysis for Images

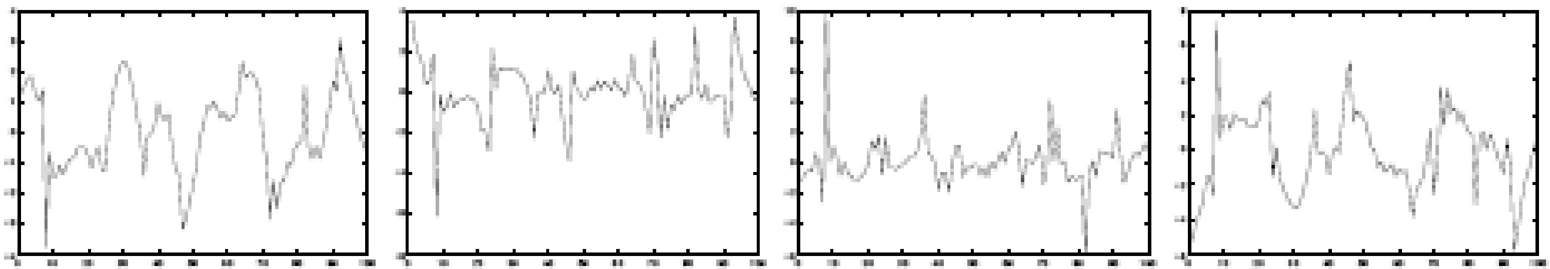
A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, "On separation of semitransparent dynamic images from static background", ***Proc. Intl. Conf. on Independent Component Analysis and Blind Signal Separation*****, pp. 934-940, 2006.**

ICA: Application to Signal Decomposition

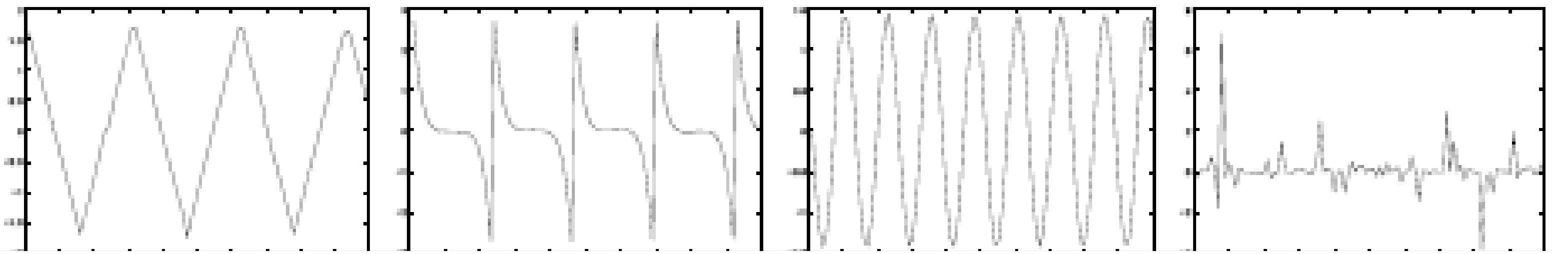
4 independent sources



4 measurements: sources are mixed

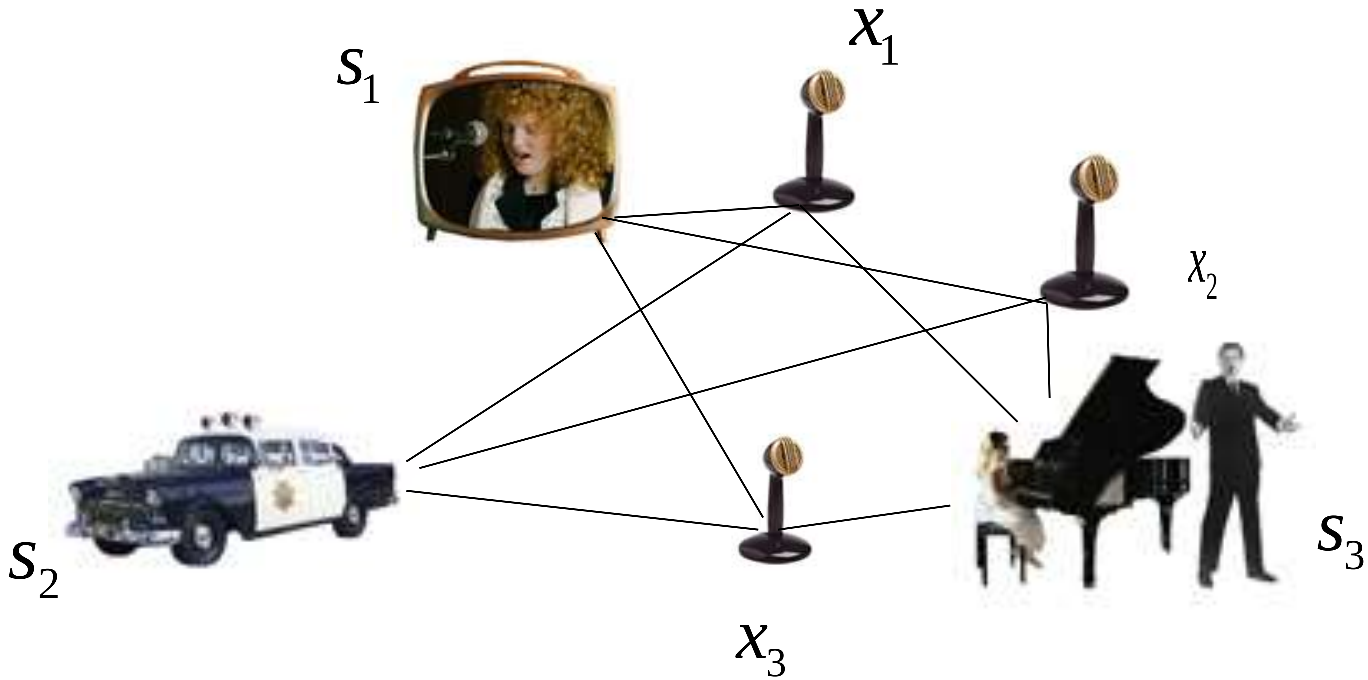


4 signals de-mixed after ICA



Two of the sources are inverted

ICA: Formalism



N-dimensional observation vector $x \in \mathbb{R}^N, \; N = 3.$

x was generated by a linear combination of N sources, $s \in \mathbb{R}^N$.

$x = As$, mixing matrix $A: N \times N$

ICA uncovers both A and s.

ICA: A projection pursuit method

$x = As$ is the equation of a projection.

The N sources are projected through A.

Each of the rows of A is a projection vector.

How to choose the projections

ICA is a **projection pursuit method.**

It searches for ***interesting*** *projections.*

Projections convey information on the sources

ICA assumes that

the underlying ***sources*** are:

- *Statistically Independent*
- *Non-Gaussian*

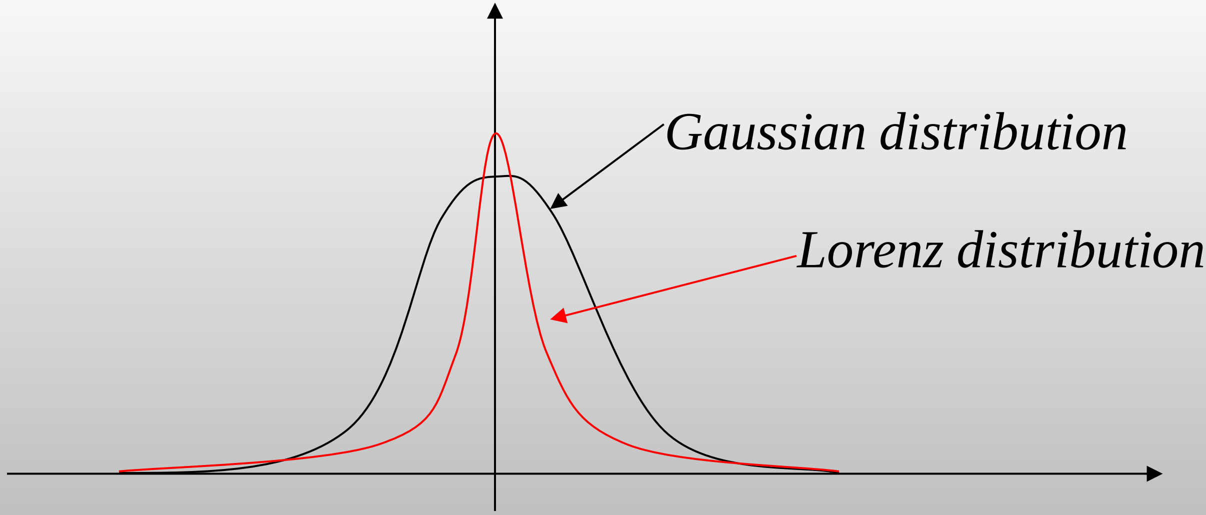
Maximizing Statistical Independence

Maximize non-Gaussianity

The Central Limit Theorem states that, ‘the sum of several independent random variables tends towards a Gaussian distribution’.

The distribution of the mixed components is closer to a Gauss distribution than the distribution of the source components

Non-gaussian distributions

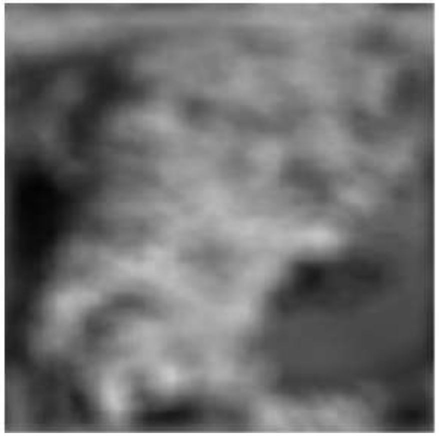


*Lorentz distribution is a n*on-Gaussian Distribution

It represents well variables that are more likely to have either very large or very small (close to zero) values

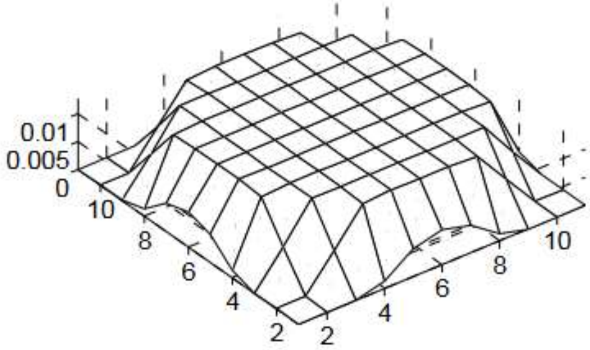
Non-gaussian distributions

Real images are subjected to all sort of noise. The resulting (mixed) image is a composite of the noise distribution and original image's distribution.



Noisy (out of focus)

Original image



Pdf of noisy image

A uniform blurring produces correlation between adjacent pixels, leading to a flatter distribution

Statistical Dependency

Mutual Information

One approach to measure statistical dependency is through mutual information.

Mutual Information for x, y:

$$I(x,y) = H(x) - H(x|y)$$

$$H(x) = -\int p(x)\log p(x)$$: entropy of x

Measuring non-Gaussianity: Negentropy

Neg-entropy measures how much a distribution departs from the "normal" (or Gauss) distribution.

Measuring non-Gaussianity: Negentropy

Neg-entropy measures how much a distribution departs from the "normal" (or Gauss) distribution.

It is computed as the difference between the entropy of the equivalent Gauss distribution and the entropy of the true distribution

$$\boxed{\text{Negentropy: } J(x) = H_{Gauss}\{x\} - H\{x\}}$$

$H_{Gauss}\{x\}$: Gauss distribution with same mean and variance as true distribution of x.

H is always **positive**. **H=1 for a Gauss function** *(for normalized dataset).*

A value of **entropy (H) < 1** means that the distribution is **non-Gaussian**.

The smaller H, the more non-Gaussian.

Estimating the entropy is difficult as it requires to estimate the distribution of the data.

$$\boxed{\text{Negentropy: } J(x) = H_{Gauss}\{x\} - H\{x\}}$$

$H_{Gauss}\{x\}$: Gauss distribution with same mean and variance as true distribution of x.

Measuring non-Gaussianity: Kurtosis

When data is preprocessed to be zero-mean and have unit variance, kurtosis is equal to the fourth moment of the data.

$$kurtosis(x) = E\left\{x^4\right\} - 3\left(E\left\{x^2\right\}\right)^2$$

(for zero-mean distribution)

$$kurtosis(x) = E\left\{x^4\right\}$$

(for unit variance distribution)

Measuring non-Gaussianity: Kurtosis

When data is preprocessed to be zero-mean and have unit variance, kurtosis is equal to the fourth moment of the data.

$$kurtosis(x) = E\{x^4\} - 3\left(E\{x^2\}\right)^2$$

(for zero-mean distribution)

$$kurtosis(x) = E\{x^4\}$$

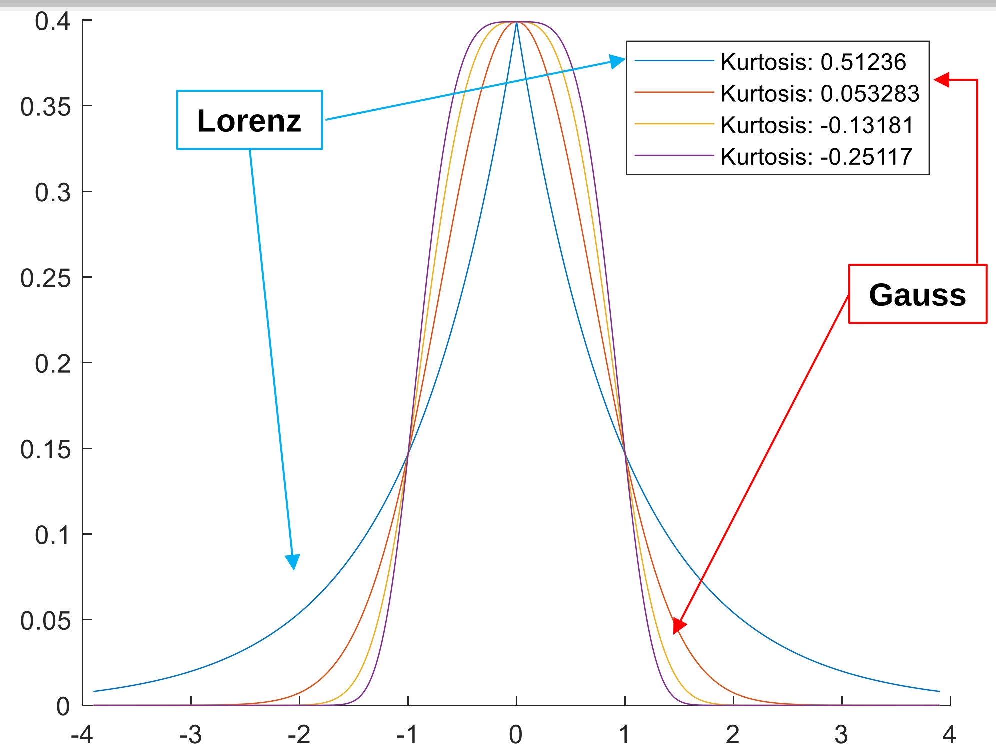
(for unit variance distribution)

kurtosis measures "spikiness"

$kurtosis(x) = 0$ (Gauss distribution)

$kurtosis(x) > 0$ (Spiky distribution)

$kurtosis(x) < 0$ (Flat distribution)



N.B.: While the $kurtosis(x) = 3$ for a Gauss distribution, we use standard notation and substraction 3 to set it to zero.

Key concepts of ICA

What is ICA?

Dimensionality reduction technique

for

Blind Source Analysis

ICA principle

Find **projections** in which the data distribution is **statistically independent**.

Non-gaussian is independent

→ maximize non-gaussiantity

Kurtosis or Negentropy