Independent Component Analysis (ICA)

Pradeep Ravikumar

Co-instructor: Manuela Veloso

Machine Learning 10-701

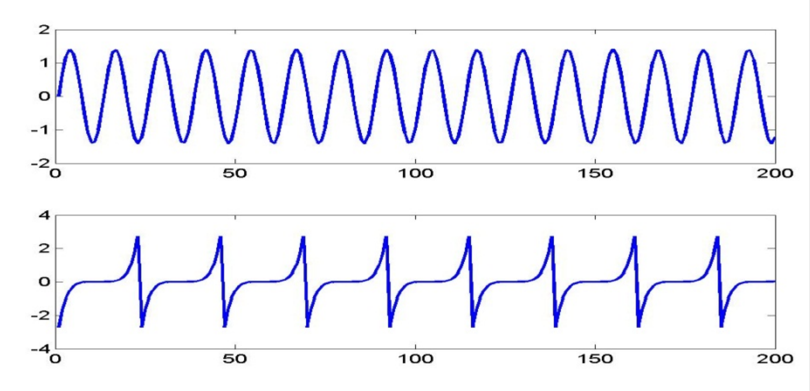
Slides courtesy of Barnabas Poczos

Independent Component Analysis

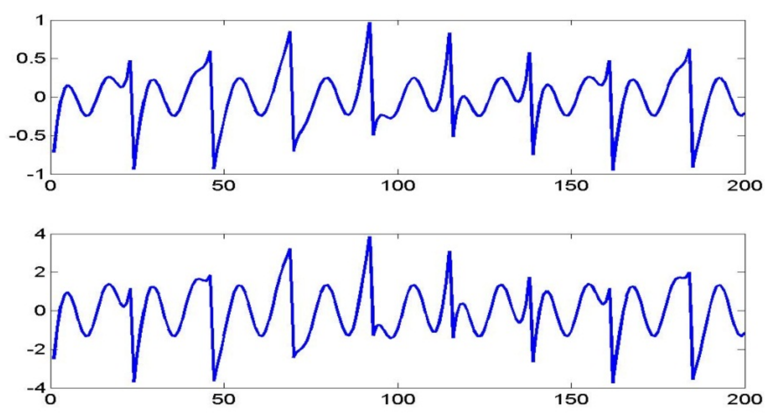
Independent Component Analysis

$$
\begin{aligned}
x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\
x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t)
\end{aligned}
$$

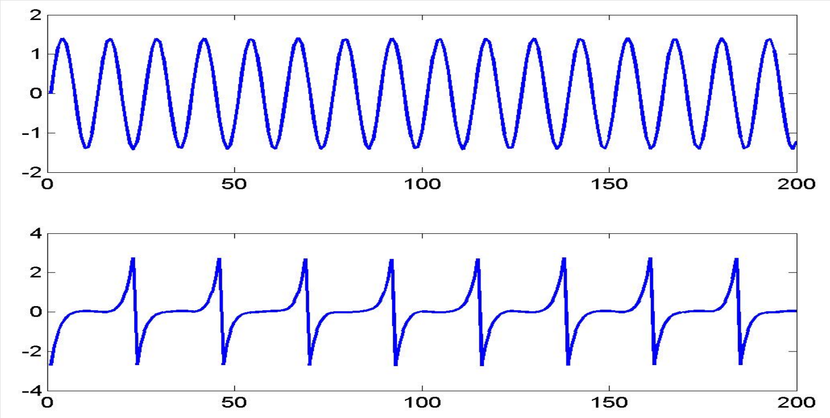
Model



original signals



Observations (Mixtures)



ICA estimated signals

Independent Component Analysys

Model

$$
\begin{aligned}
x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\
x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t)
\end{aligned}
$$

We observe

$$
\begin{pmatrix} x_1(1) \\ x_2(1) \end{pmatrix}, \begin{pmatrix} x_1(2) \\ x_2(2) \end{pmatrix}, \ldots, \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}
$$

We want

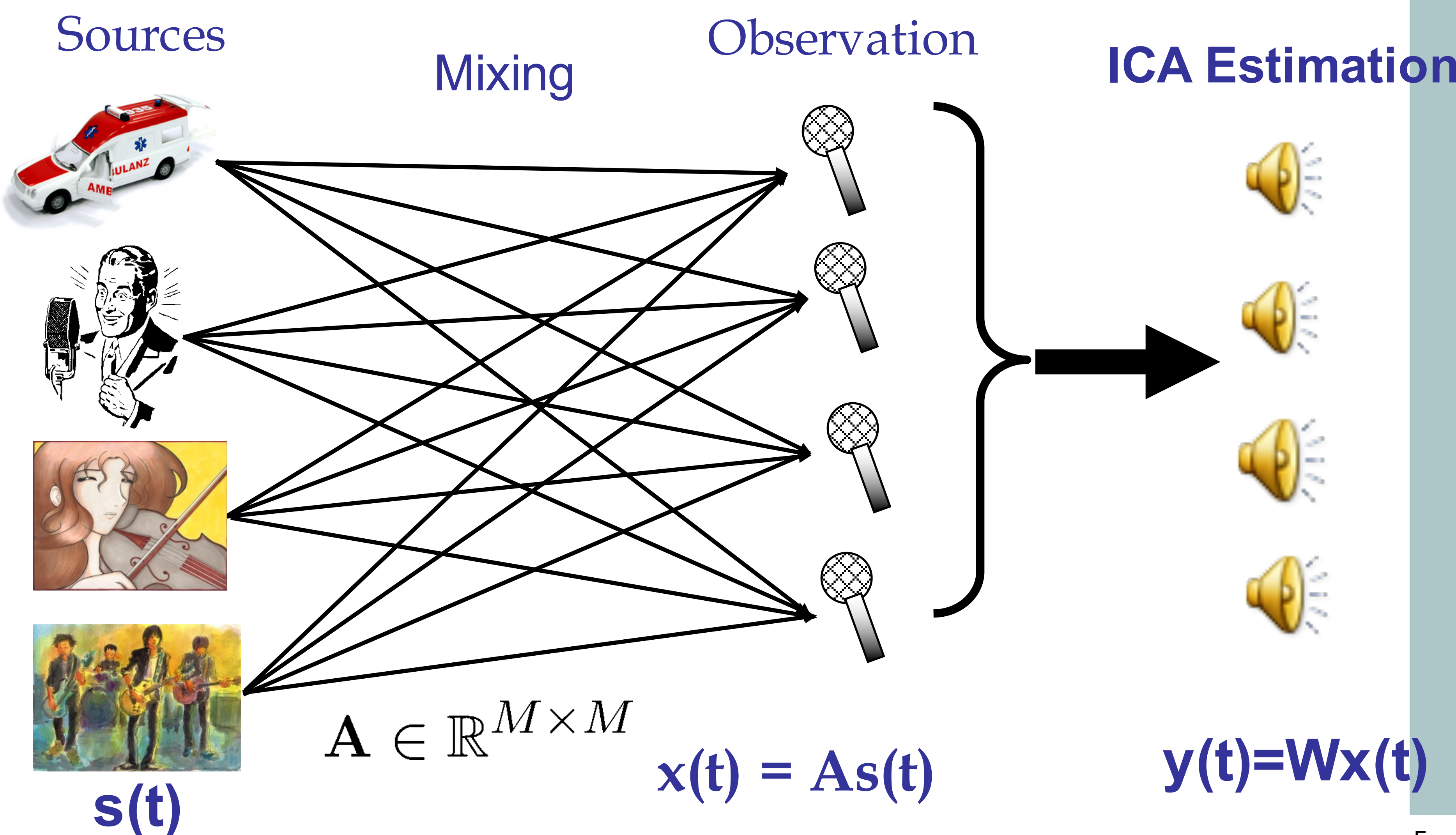
$$
\begin{pmatrix} s_1(1) \\ s_2(1) \end{pmatrix}, \begin{pmatrix} s_1(2) \\ s_2(2) \end{pmatrix}, \ldots, \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix}
$$

But we don't know $\{a_{ij}\}$, nor $\{s_i(t)\}$

Goal: Estimate $\{s_i(t)\}$, (and also $\{a_{ij}\}$)

… under the assumption that signals are independent

The Cocktail Party Problem
SOLVING WITH ICA



ICA vs PCA

- PCA: $\mathbf{X=US, U^TU=I}$
- ICA: $\mathbf{X=AS}$, **A is invertible**

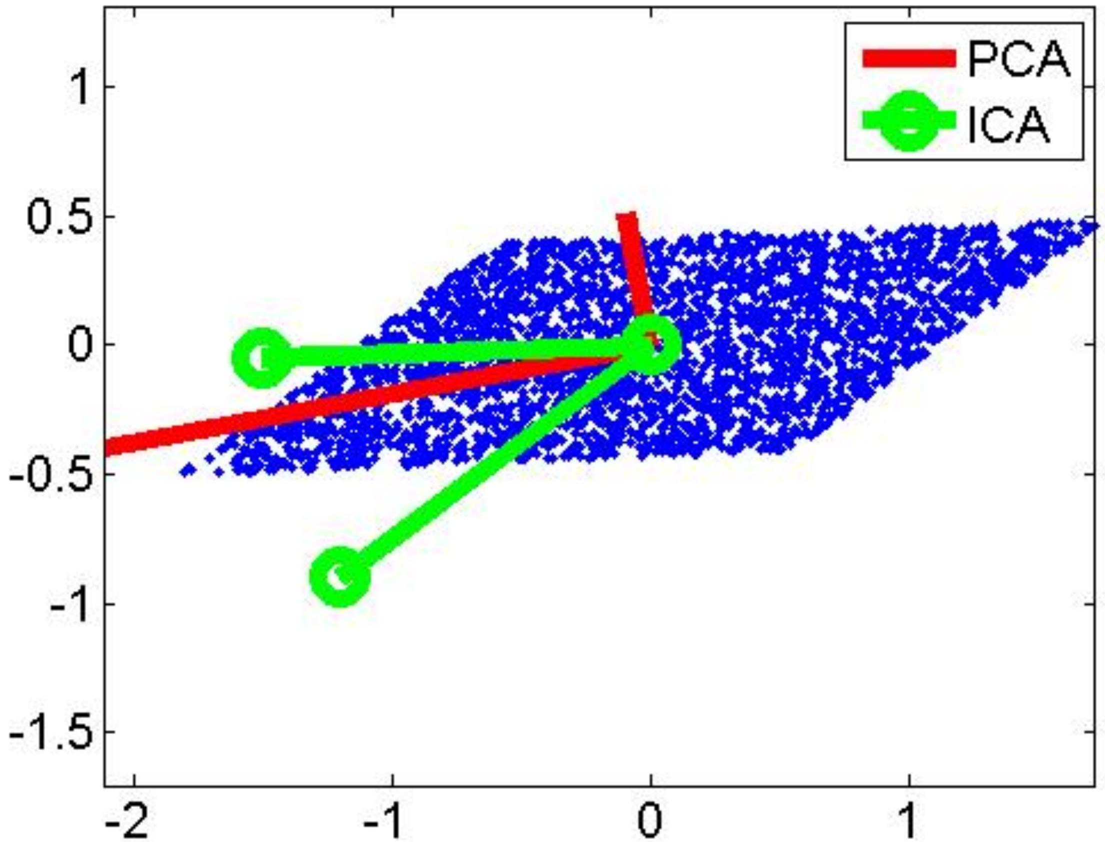
- PCA **does** compression
  - M<N

- ICA does **not** do compression
  - same # of features (M=N)

- PCA just removes correlations, **not** higher order dependence
- ICA removes correlations, **and** higher order dependence

- PCA: some components are **more important** than others (based on eigenvalues)
- ICA: components are **equally important**

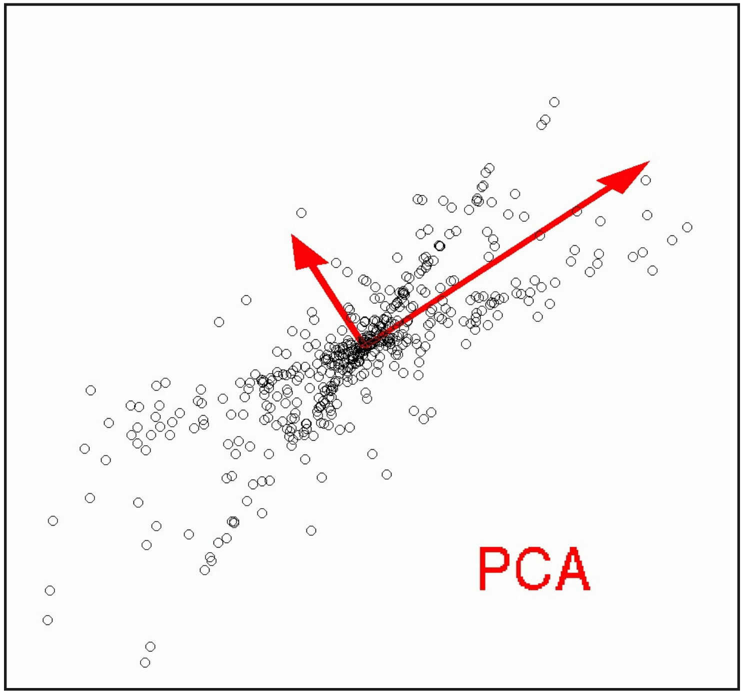
ICA vs PCA

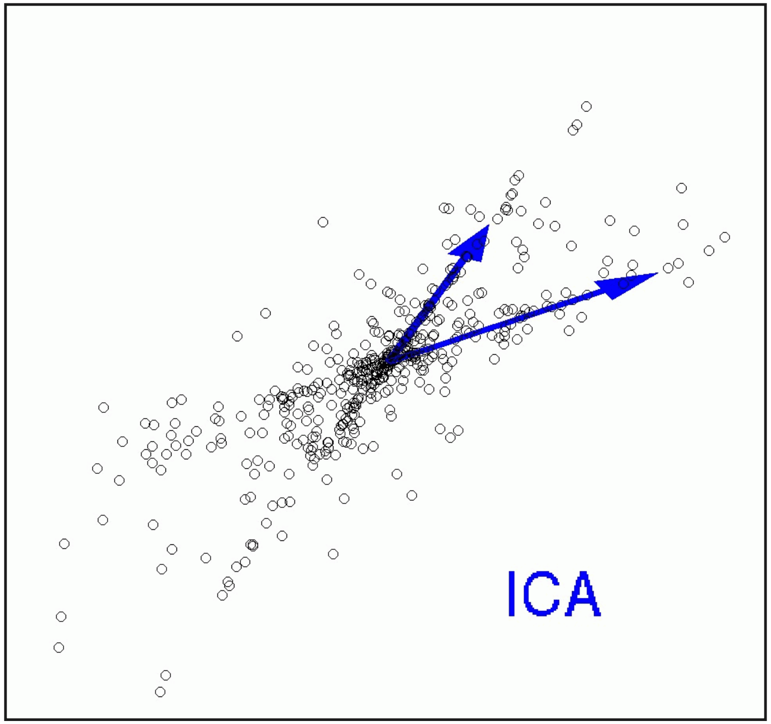


Note

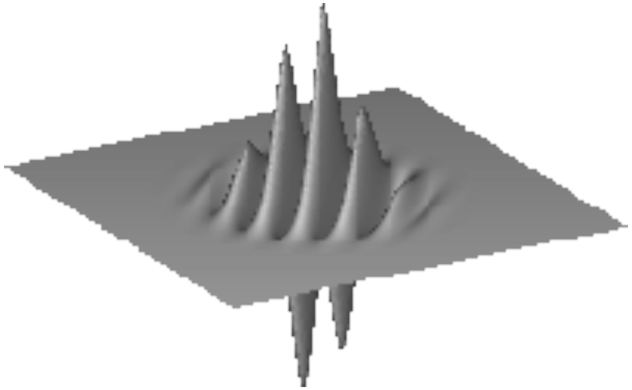
- PCA vectors are orthogonal
- ICA vectors are **not** orthogonal

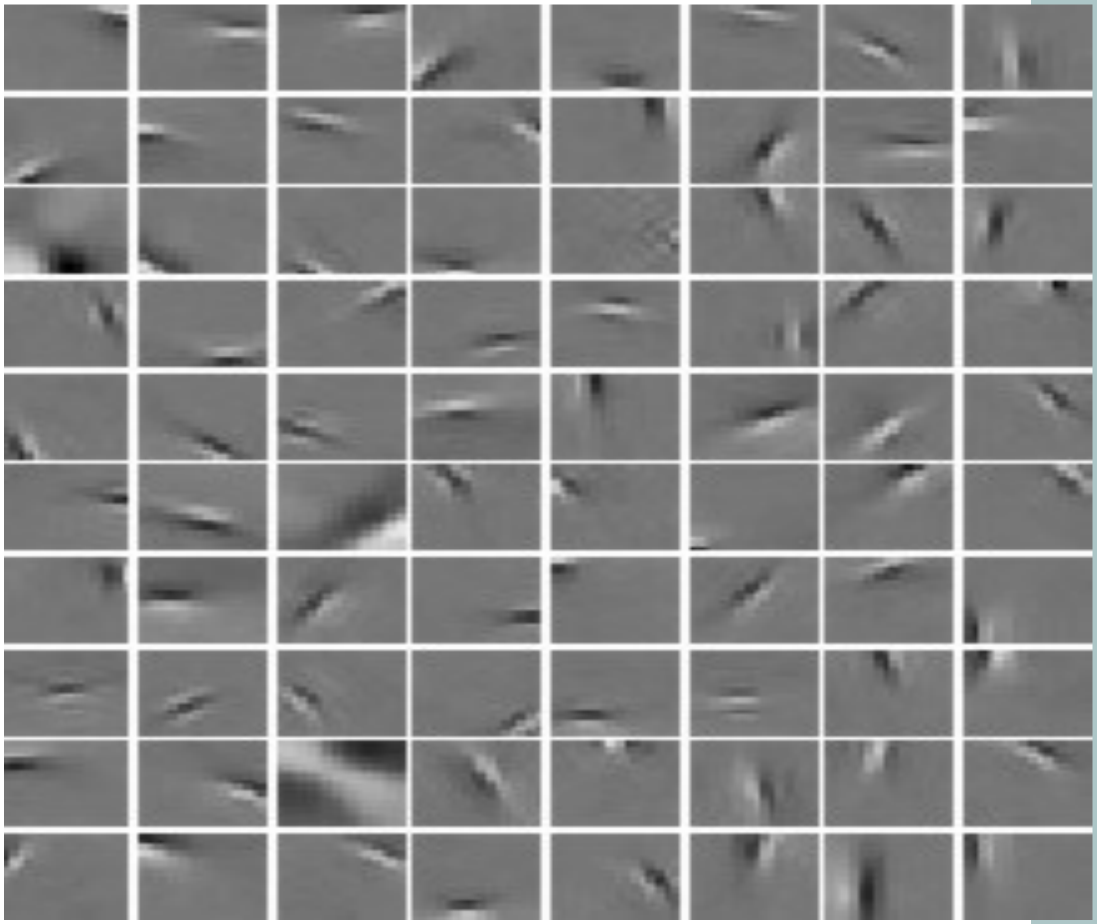
ICA vs PCA





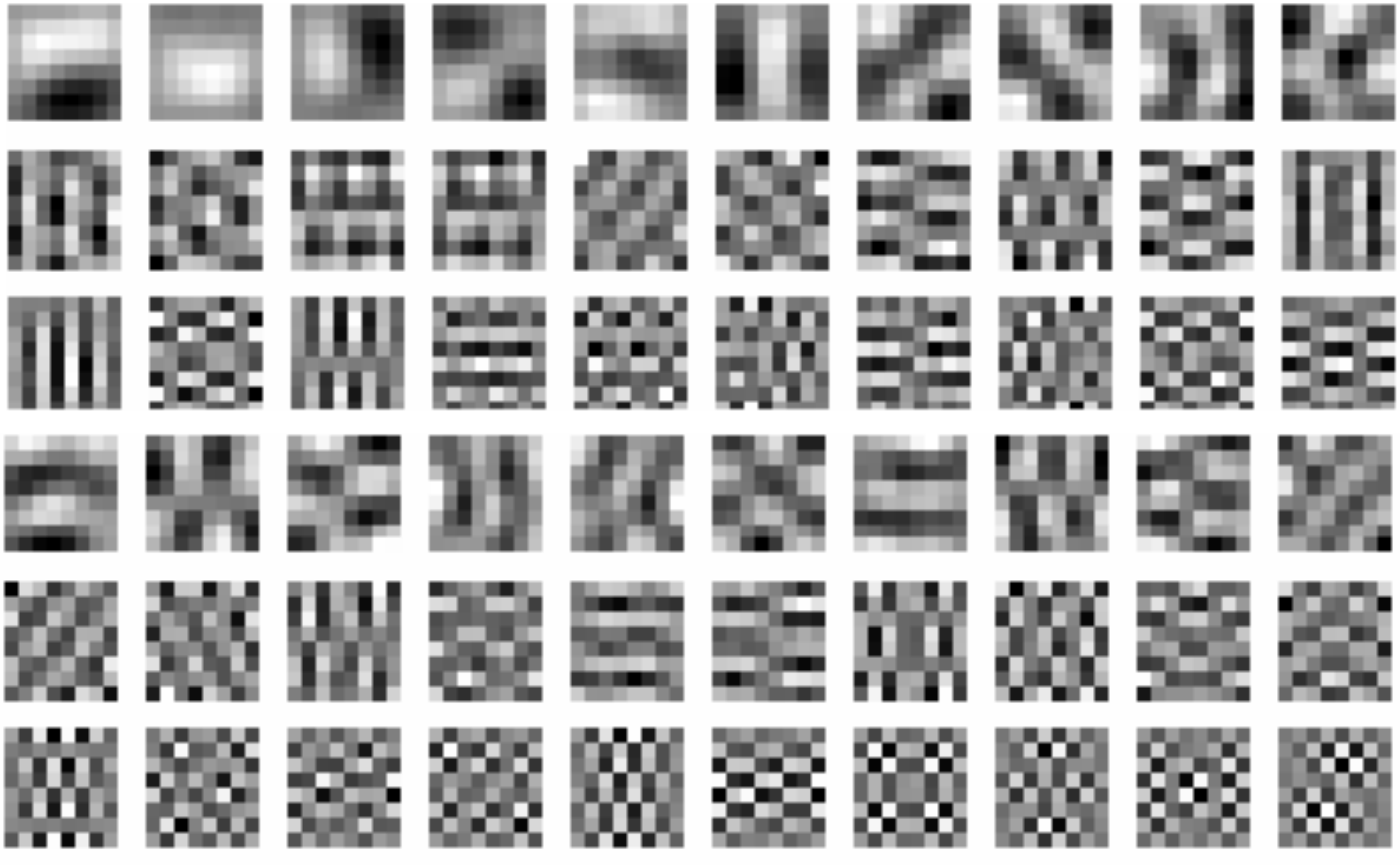
ICA basis vectors extracted from natural images





Gabor wavelets,
edge detection,
receptive fields of V1 cells..., deep neural networks

PCA basis vectors extracted from natural images



Some ICA Applications

STATIC

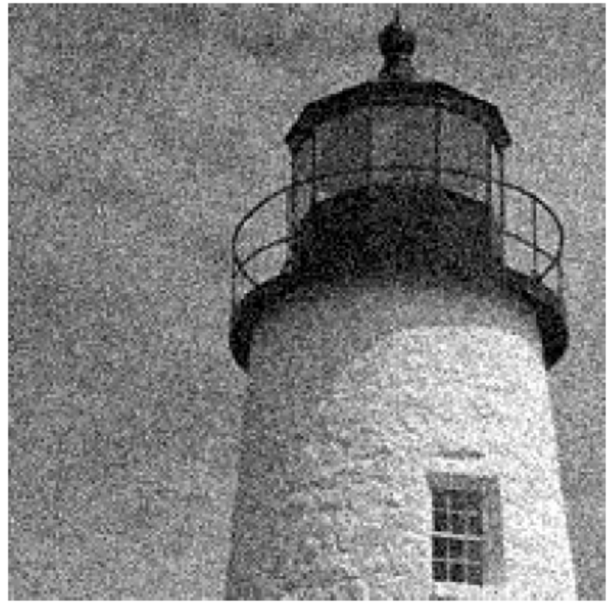
- Image denoising
- Microarray data processing
- Decomposing the spectra of galaxies
- Face recognition
- Facial expression recognition
- Feature extraction
- Clustering
- Classification
- Deep Neural Networks

TEMPORAL

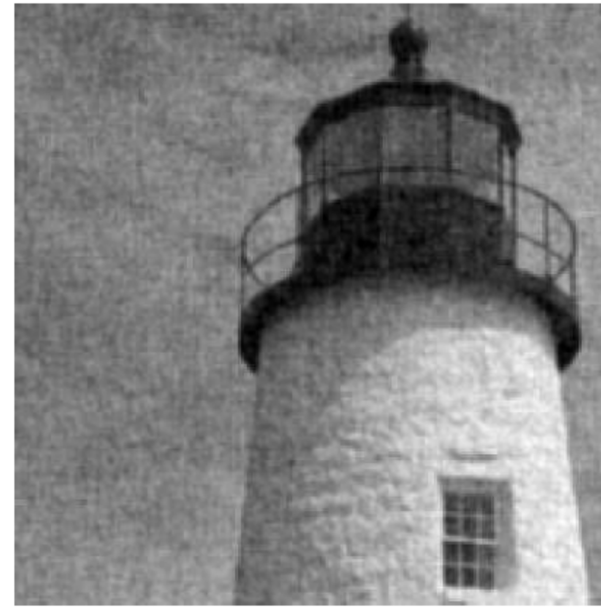
- Medical signal processing – fMRI, ECG, EEG
- Brain Computer Interfaces
- Modeling of the hippocampus, place cells
- Modeling of the visual cortex
- Time series analysis
- Financial applications
- Blind deconvolution

ICA for Image Denoising

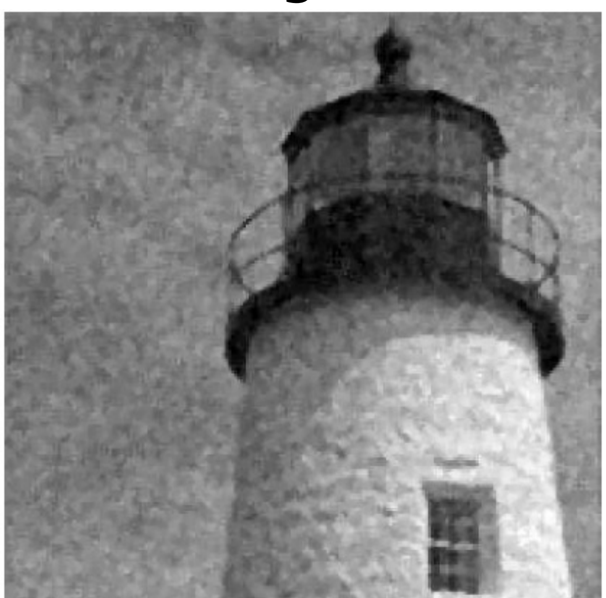
original



noisy



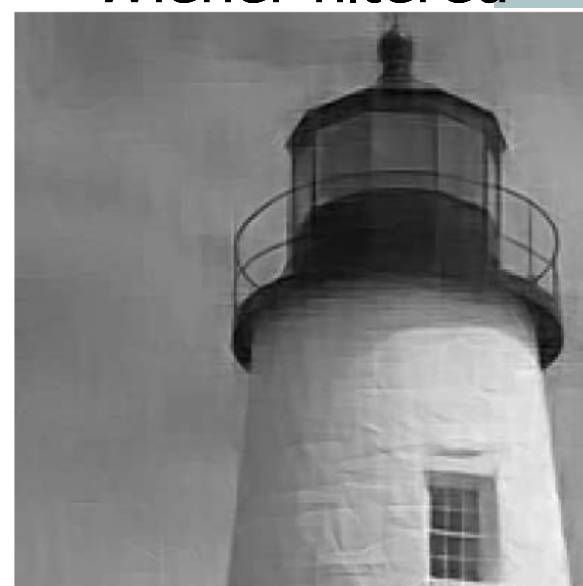
Wiener filtered



median filtered

ICA denoised

(Hoyer, Hyvarinen)



ICA Theory

Solving the ICA problem with i.i.d. sources

ICA model: $\mathrm{x} = \mathrm{A}\mathrm{s}$, $\mathrm{s} = [s_1; \ldots ; s_M]$ are jointly independent.

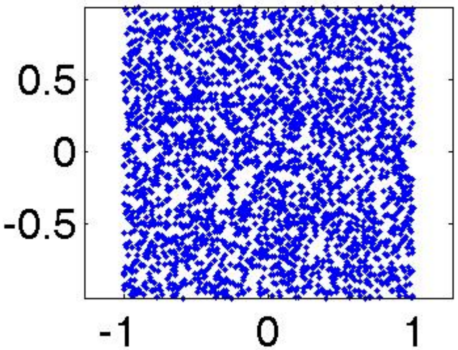
ICA task: Given **x**, find **A** and **s**

ICA solution: **y=Wz** **y** – surrogate for **s**, **z** – surrogate for **x**

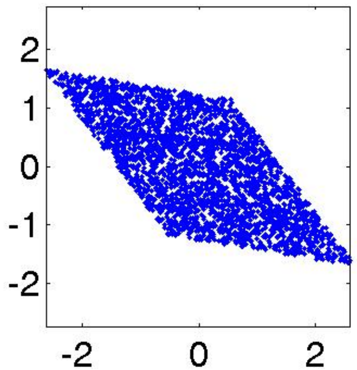
- Remove mean and whiten **x** to get **z** with $\mathbf{E[z]=0}$, $\mathbf{E[zz^T] = I}$

  $$\mathbf{z = D^{-1/2}U'\,(x - E[x])} \qquad \mathbf{UDU' = eig(\Sigma_x)}$$

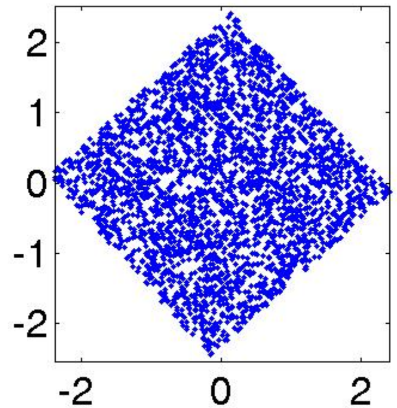
- Find an orthogonal **W** minimizing **dependence** between **y**



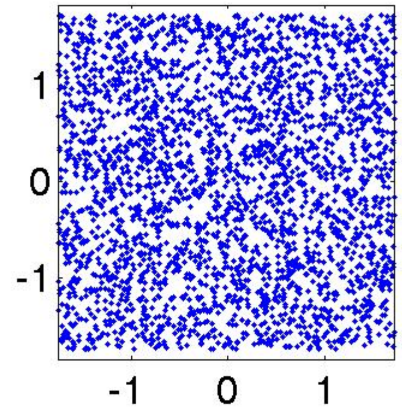
Original **s**



Mixed **x**



Whitened **z**



Demixed **y**

Statistical (in)dependence

Definition (Independence)

Y_1, Y_2 are independent $\Leftrightarrow p(y_1, y_2) = p(y_1)\,p(y_2)$

Definition (Shannon entropy)

$$H(\mathbf{Y}) \doteq H(Y_1, \ldots, Y_m) \doteq -\int p(y_1, \ldots, y_m) \log p(y_1, \ldots, y_m) d\mathbf{y}.$$

Definition (Mutual Information) between more than 2 variables

$$0 \leq I(Y_1, \ldots, Y_M) \doteq \int p(y_1, \ldots, y_M) \log \frac{p(y_1, \ldots, y_M)}{p(y_1) \ldots p(y_M)} d\mathbf{y}$$

Definition (KL divergence)

$$0 \leq KL(f \| g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

ICA Cost Functions

Let $\mathbf{y} \doteq \mathbf{W}\mathbf{x}$, $\mathbf{y} = [y_1; \ldots; y_M]$, and let us measure the dependence using Shannon's mututal information:

$$J_{ICA_1}(\mathbf{W}) \doteq I(y_1, \ldots, y_M) \doteq \int p(y_1, \ldots, y_M) \log \frac{p(y_1,\ldots,y_M)}{p(y_1)\ldots p(y_M)} d\mathbf{y},$$

Let $H(\mathbf{y}) \doteq H(y_1, \ldots, y_m) \doteq -\int p(y_1, \ldots, y_m) \log p(y_1, \ldots, y_m) d\mathbf{y}.$

Lemma

$$H(\mathbf{W}\mathbf{x}) = H(\mathbf{x}) + \log|\det \mathbf{W}|$$

Proof: Recitation

Therefore,

$$\begin{aligned} I(y_1, \ldots, y_M) &= \int p(y_1, \ldots, y_M) \log \frac{p(y_1, \ldots, y_M)}{p(y_1) \ldots p(y_M)} \\ &= -H(y_1, \ldots, y_M) + H(y_1) + \ldots + H(y_M) \\ &= -H(x_1, \ldots, x_M) - \log|\det \mathbf{W}| + H(y_1) + \ldots + H(y_M). \end{aligned}$$

ICA Cost Functions

$$\begin{aligned} I(y_1,\ldots,y_M) &= \int p(y_1,\ldots,y_M) \log \frac{p(y_1,\ldots,y_M)}{p(y_1)\ldots p(y_M)} \\ &= -H(y_1,\ldots,y_M) + H(y_1) + \ldots + H(y_M) \\ &= -H(x_1,\ldots,x_M) - \log|\det \mathbf{W}| + H(y_1) + \ldots + H(y_M). \end{aligned}$$

$H(x_1,\ldots,x_M)$ is constant

Does not depend on **W**

Suppose we select W so that:

$\log|\det \mathbf{W}| = 0.$

ICA Cost Functions

$$
\begin{aligned}
I(y_1, \ldots, y_M) &= \int p(y_1, \ldots, y_M) \log \frac{p(y_1, \ldots, y_M)}{p(y_1) \ldots p(y_M)} \\
&= -H(y_1, \ldots, y_M) + H(y_1) + \ldots + H(y_M) \\
&= -H(x_1, \ldots, x_M) - \log|\det \mathbf{W}| + H(y_1) + \ldots + H(y_M).
\end{aligned}
$$

$H(x_1, \ldots, x_M)$ is constant, $\log|\det \mathbf{W}| = 0.$

Therefore,

$$
J_{ICA_2}(\mathbf{W}) \doteq H(y_1) + \ldots + H(y_M)
$$

The covariance is fixed: Which distribution has the largest entropy?

➔ go away from normal distribution