THE UNIVERSITY OF RHODE ISLAND

NeuralPC Lab

Brain Signal processing & Applications

BME 473/ELE 573

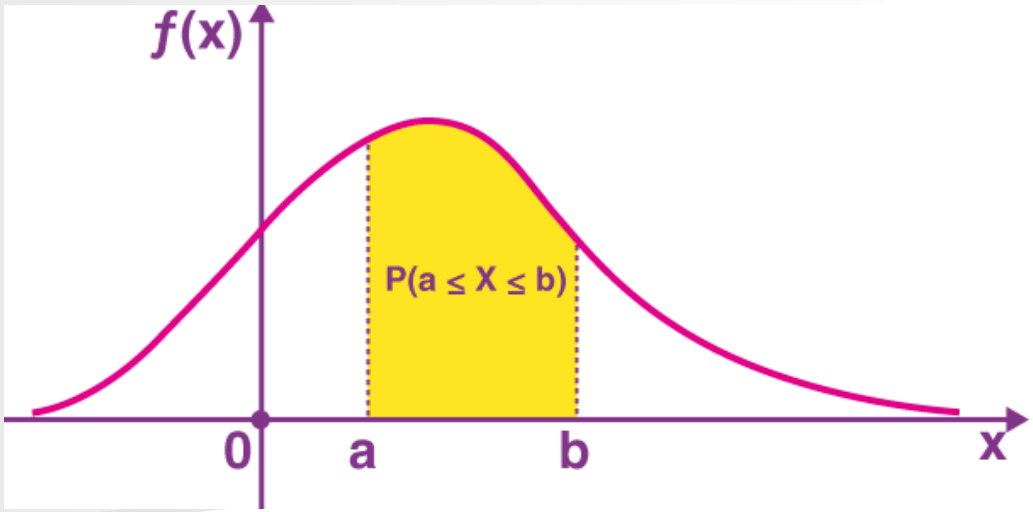
Lecture 4

Dr. Yalda Shahriari

Mathematical Preliminaries

- ✓ In probability theory, a probability density function (PDF) defines the random variable's probability coming within a distinct range of values, as opposed to taking on any one value.
  - ✓ In another word, the probability density function is defined as an integral of the density of the variable over a given range.

$$P(a<X<b) = \int_a^b p_x(x)dx$$



- ✓ The PDF ensures that the probability density function, remains non-negative for all possible values of x, i.e., f(x) ≥ 0, for every x.
- ✓ The area between the density curve and horizontal X-axis is equal to 1, i.e.

Mathematical Preliminaries

- Cumulative distribution function (cdf), F_x of a random variable x at point $x = x_0$ is defined as the probability that $x \leq x_0$

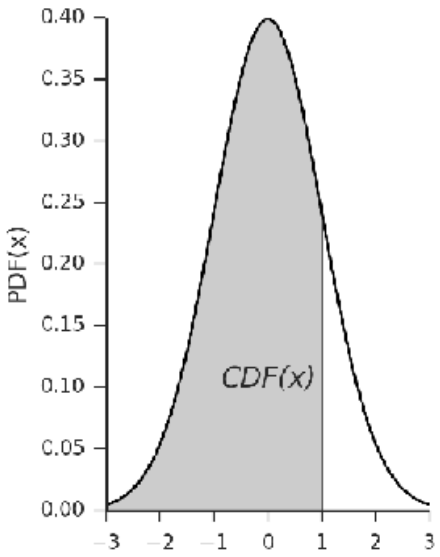
$$F_x(x_0) = P(x \leq x_0)$$

$$F_x(x_0) = \int_{-\infty}^{x_0} p_x(\xi) d\xi$$

Where P_x is the probability density function (pdf), and x_0 changes from -∞ to +∞ where $F_x(-\infty)=0$ and $F_x(\infty)=1$.

- The corresponding probability density function (pdf), P_x can be obtained as below:

$$p_x(x_0) = \left. \frac{dF_x(x)}{dx} \right|_{x=x_0}$$



http://work.thaslwanter.at/Stats/html/statsDistributions.html

Mathematical Preliminaries

- **Joint and marginal distributions:**

The joint probability density function (joint pdf) for two random variables x and y, is a function used to characterize the probability distribution of the two continuous random variables, which together form a continuous random vector.

$$p_{x,y}(x0, y0) = P(x = x0 \; and \; y = y0)$$

Let us assume we have two random variables x and y. The vectors x and y can be concatenated to a supervector $z^T = (x^T, y^T)$. The cdf of this supervector is called joint distribution function which is defined as below:

$$F_{\mathbf{x},\mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) = \int_{-\infty}^{\mathbf{x}_0} \int_{-\infty}^{\mathbf{y}_0} p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta}) d\boldsymbol{\eta} d\boldsymbol{\xi}$$

Where $p_{x,y}(x, y)$ is the joint density function and $p_x(x\,)$ and $p_y(y\,)$ are the marginal densities.

$$p_{\mathbf{x}}(\mathbf{x}) = \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \boldsymbol{\eta}) d\boldsymbol{\eta}$$

$$p_{\mathbf{y}}(\mathbf{y}) = \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \mathbf{y}) d\boldsymbol{\xi}$$

Mathematical Preliminaries

Expectation:

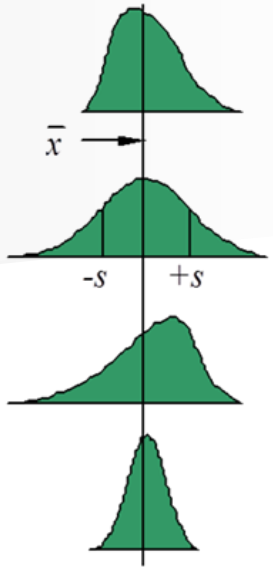
In probability theory, the **expected value** (also called **expectation**, **mean**, **average**, or **first moment**) is a generalization of the weighted average.

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Moments:

Moments of a random variable x are typical expectations used to characterize it.

1- Mean (measure of location)
2- Variance (measure of spread)
3- Skewness (measure of symmetry)
4- Kurtosis (measure of peakness)



Mathematical Preliminaries

- **First moment (Mean):**

$$\mathbf{m_x} = \mathrm{E}\{\mathbf{x}\} = \int_{-\infty}^{\infty} \mathbf{x} p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}$$

- **Second moment**
  - ✓ **Variance:**

$$\mathbf{C_x} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\}$$

  - ✓ **Covariance:**

$$\mathbf{C_{xy}} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\} = E(xy^T) - E(x)E(y^T) = E(xy^T) - m_x m_y$$

  - ✓ **Correlation:**

$$R_{xy} = \frac{C_{xy}}{\sqrt{Var(x)Var(y)}}$$

Mathematical Preliminaries

- **Third moment (Skewness):**

$$skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-1)s^3}$$

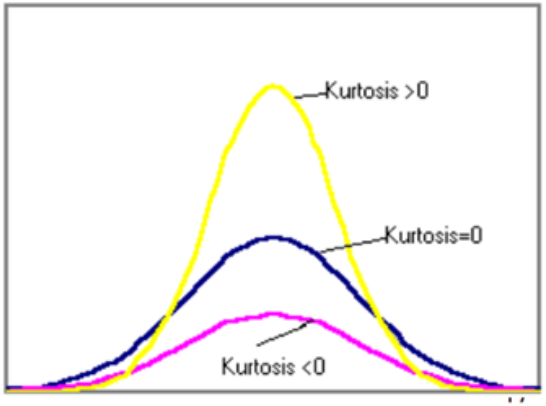
- **Fourth moment (Kurtosis):**

$$Kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)s^4}$$

Gaussian Dis. Kurtosis = 0

Super-Gaussian: kurtosis > 0

Sub-Gaussian: kurtosis < 0



$$C_{x,y} = \begin{bmatrix} \sigma_x^2 & C_{xy} \\ C_{yx} & \sigma_y^2 \end{bmatrix}$$

- ✓ $C_{x,y}$ is positive-semidefinite and symmetric (all eigenvalues are non-negative).
- ✓ A positive definite matrix is defined as a symmetric matrix whose every eigenvalue is positive.

The following definitions all involve the term $\mathbf{x}^* M\mathbf{x}$. Notice that this is always a real number for any Hermitian square matrix M.

An $n \times n$ Hermitian complex matrix M is said to be **positive-definite** if $\mathbf{x}^* M\mathbf{x} > 0$ for all non-zero $\mathbf{x}$ in $\mathbb{C}^n$. Formally,

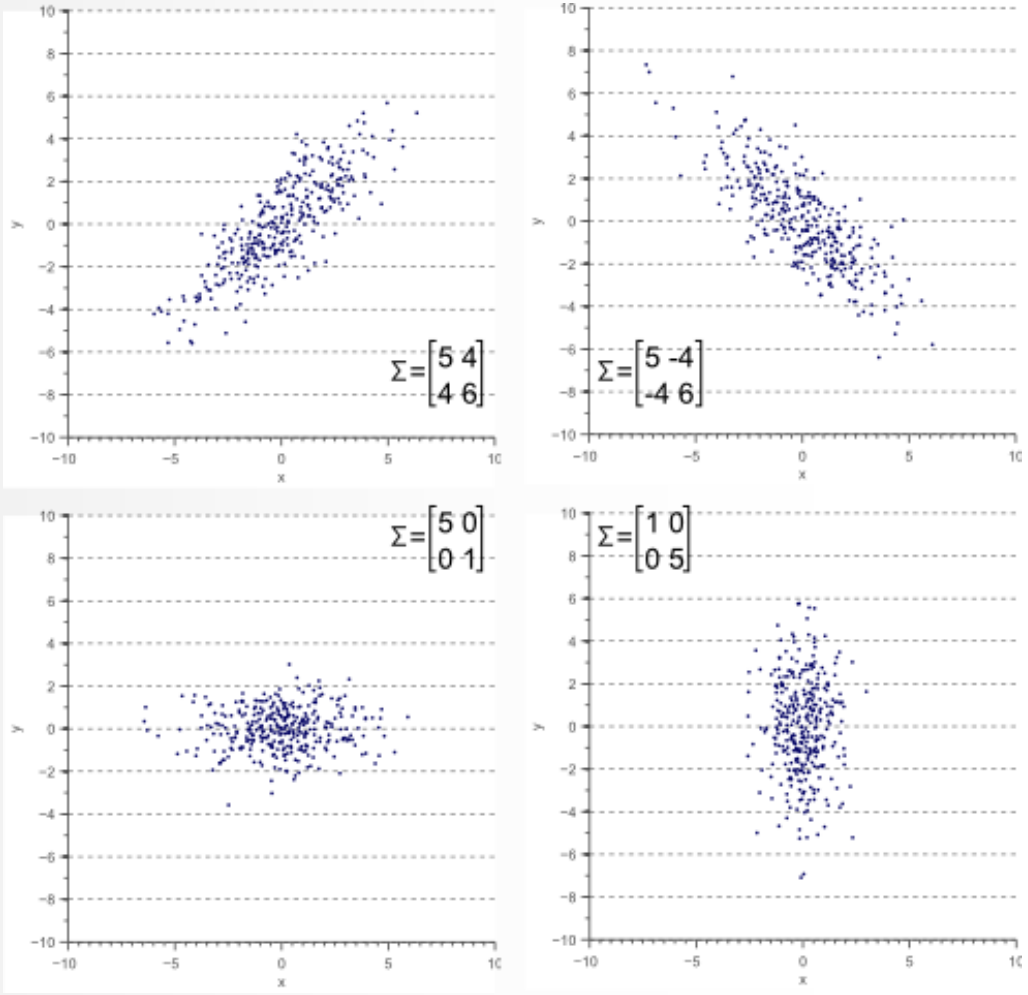
$$M \text{ positive-definite} \iff \mathbf{x}^* M\mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$$

An $n \times n$ Hermitian complex matrix M is said to be **positive semi-definite** or **non-negative-definite** if $x^* Mx \geq 0$ for all x in $\mathbb{C}^n$. Formally,

$$M \text{ positive semi-definite} \iff \mathbf{x}^* M\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{C}^n$$

$$C_{x,y} = \begin{bmatrix} \sigma_x^2 & C_{xy} \\ C_{yx} & \sigma_y^2 \end{bmatrix}$$

- ✓ The covariance matrix defines the shape of the data. Diagonal spread is captured by the covariance, while axis-aligned is captured by the variance.



➢ Uncorrelateness:

Two random vectors x and y are uncorrelated if their covariance matrix C_{xy} is zero:

$$\mathbf{C_{xy}} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\} = \mathbf{0}$$

Mathematical Preliminaries

Statistical Independence:

Two random variables x and y are said to be statistically independent if and only if:

$$p_{x,y}(x, y) = p_x(x)p_y(y)$$

Equivalently, independent random variables satisfy the basic property of:

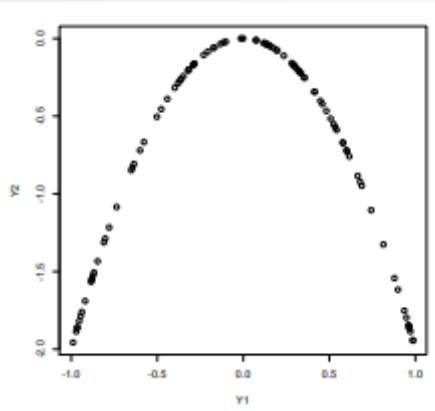
$$\mathrm{E}\{g(x)h(y)\} = \mathrm{E}\{g(x)\}\mathrm{E}\{h(y)\}$$

$$\begin{aligned}\mathrm{E}\{g(x)h(y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)p_{x,y}(x, y)\,dy\,dx \\ &= \int_{-\infty}^{\infty} g(x)p_x(x)\,dx \int_{-\infty}^{\infty} h(y)p_y(y)\,dy = \mathrm{E}\{g(x)\}\mathrm{E}\{h(y)\}\end{aligned}$$

Mathematical Preliminaries

- **Conceptual differences between independence and uncorrelation:**

- Uncorrelation means that there is no linear dependence between the two random variables, while independence means that no types of dependence exist between the two random variables. For example, in the figure below Y1 and Y2 are uncorrelated (no linear relationship) but not independent.
  - Independent random variables are always uncorrelated, but uncorrelated random variables may not be independent.
  - In other words, independence is a stronger statement than uncorrelation.



Mathematical Preliminaries

➢ Orthogonality:

- ✓ In mathematics, orthogonality is the generalization of the geometric notion of perpendicularity.
- ✓ Orthogonality is also used with various meanings that are often weakly related or not related at all with the mathematical meanings.

❑ Two variables X and Y are orthogonal if: $E(xy) = 0$

- ✓ If they are orthogonal, they may or may not be uncorrelated.
- ✓ If they are uncorrelated, then they are orthogonal only if E(X) = 0 or E(Y) = 0.

Mathematical Preliminaries

Remember: $Cov(x, y) = E(xy) - E(x)E(y) = E(xy) - m_x m_y$

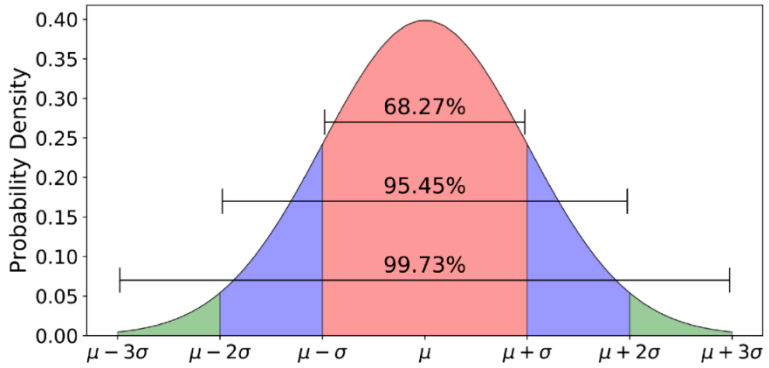
The relationship between orthogonality and correlation

	$E(XY) = 0$	$E(XY) \neq 0$
$\{E(X) \neq 0\} \cap \{E(Y) \neq 0\}$ **(both means non-zero)**	**orthogonal, correlated**	**non-orthogonal, maybe correlated, maybe not**
$\{E(X) = 0\} \cup \{E(Y) = 0\}$ **(at least one mean is zero)**	**orthogonal, uncorrelated**	**non-orthogonal, correlated**

Mathematical Preliminaries

Gaussian Distribution:

- ✓ The Gaussian distribution, also known as the normal distribution, is a continuous probability distribution characterized by symmetrical sides around its central point.
- ✓ In this distribution, the mean, median, and mode all coincide.
  - ✓ Essentially, it can help in simplying the model.
- ✓ Its graphical representation shows a bell shape with data points clustered around the mean, tapering off towards the tails.
- ✓ Gaussian distribution is the most important probability distribution in statistics because it fits many natural phenomena.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $f(x)$ = probability
- x = value of the variable
- μ = mean
- σ = standard deviation
- σ^2 = variance

Mathematical Preliminaries

➢ Multivariate Gaussian Density:

The general form of probability density function of a gaussian vector x:

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \mathbf{C}_{\mathbf{x}})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_{\mathbf{x}})^T \mathbf{C}_{\mathbf{x}}^{-1}(\mathbf{x}-\mathbf{m}_{\mathbf{x}})\right)$$

Recall that n is the dimension of $\mathbf{x}$, $\mathbf{m}_{\mathbf{x}}$ its mean, and $\mathbf{C}_{\mathbf{x}}$ the covariance matrix of $\mathbf{x}$.

➢ Properties of the gaussian density:

- ✓ Only first and second order statistics are needed
- ✓ Linear transformations are also gaussian:
  - ✓ If y=Ax is a linear transformation of gaussian vector x, then y is also gaussian with mean $m_y = Am_x$ and covariance $C_y = AC_xA^T$
- ✓ Uncorrelated gaussian random variables are also independent.

Mathematical Preliminaries

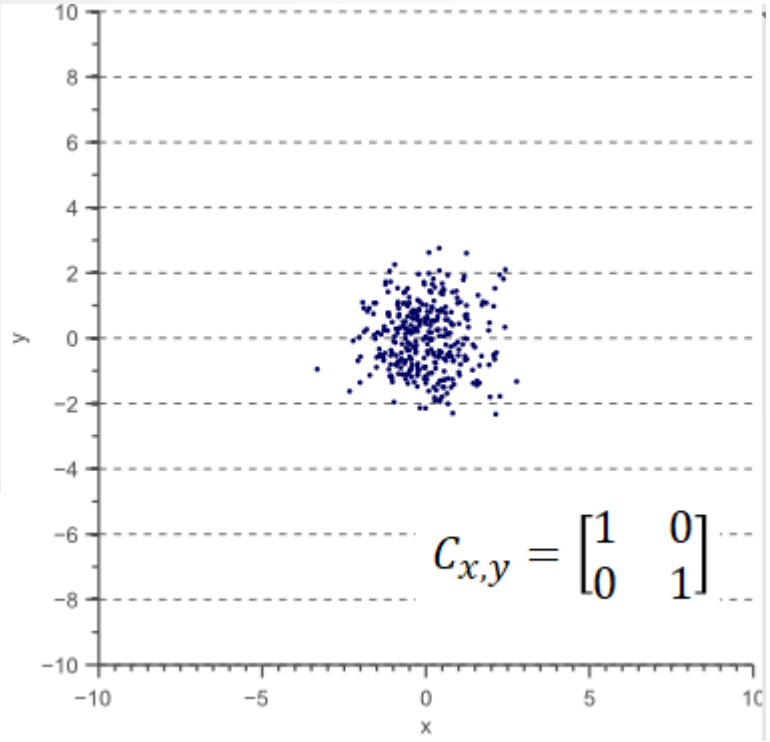
- **Whiteness:**

Random vectors having zero mean and unit covariance (and therefore zero correlation), are said to be white which satisfy the following conditions:

$$\mathbf{m_x} = \mathbf{0}, \quad \mathbf{R_x} = \mathbf{C_x} = \mathbf{I}$$

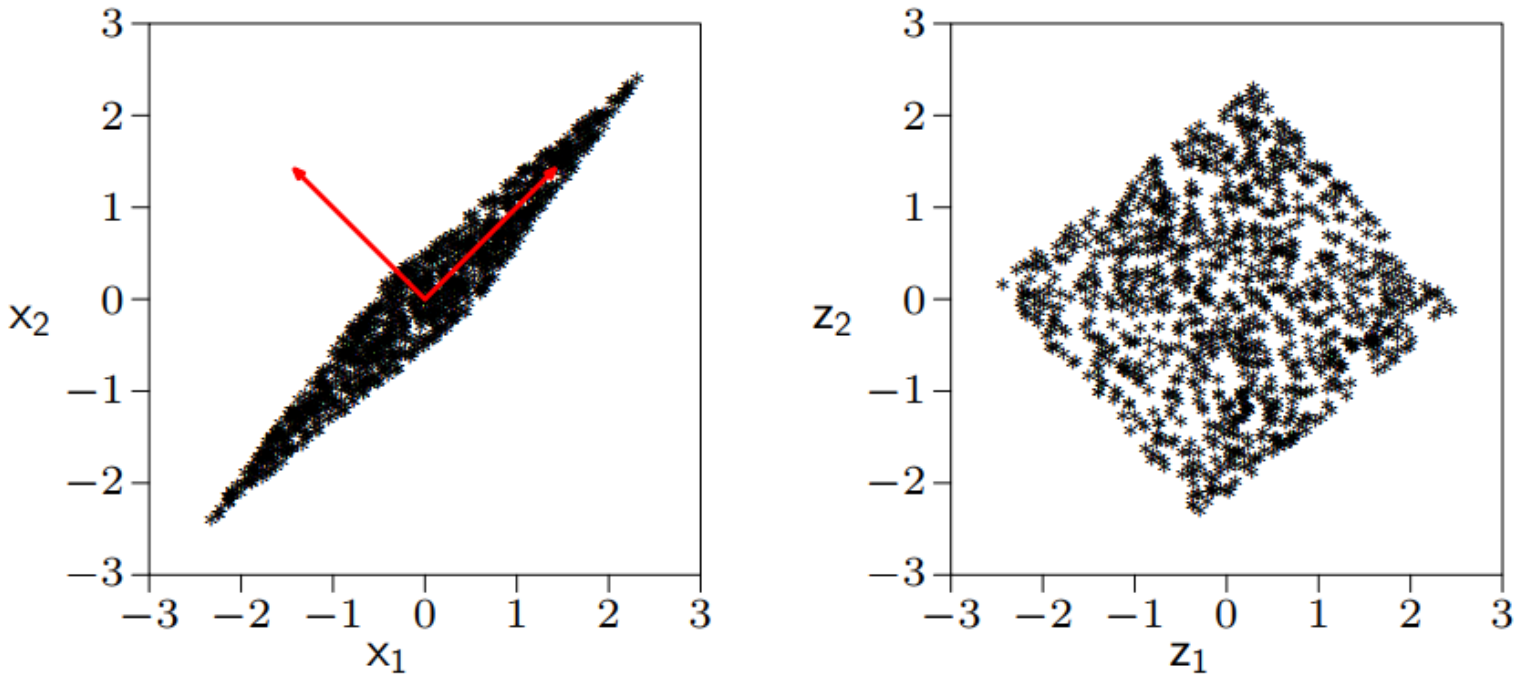
Where I is the identity matrix.

Data with unit covariance matrix (or uncorrelate and unit variance) is called white!



Mathematical Preliminaries

Whitening is a linear transformation that decorrelates and unifies the variance.



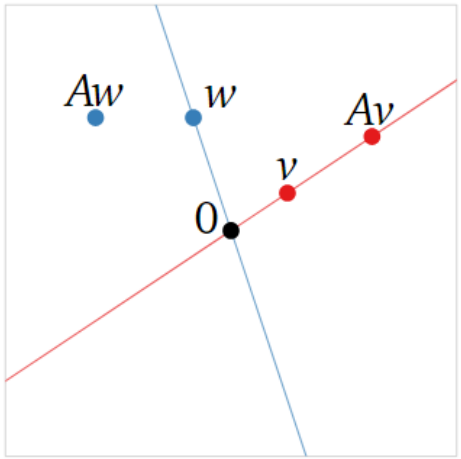
Mathematical Preliminaries

- **Eigendecomposition:**
  - ✓ The eigenvector of any matrix Σ (e.g., covariance matrix Σ) is the vector which satisfies the following condition:

$$\sum \vec{v} = \lambda \vec{v}$$

Where λ is a scalar value called "eigenvalue" and $\vec{v}$ is the eigenvector of Σ.

- ✓ Any linear transformation (e.g. matrix Σ here) of eigenvector, does not change the direction of the eigenvector and only stretches the length (depending on the eigenvalue).



v is an eigenvector

w is not an eigenvector

Mathematical Preliminaries

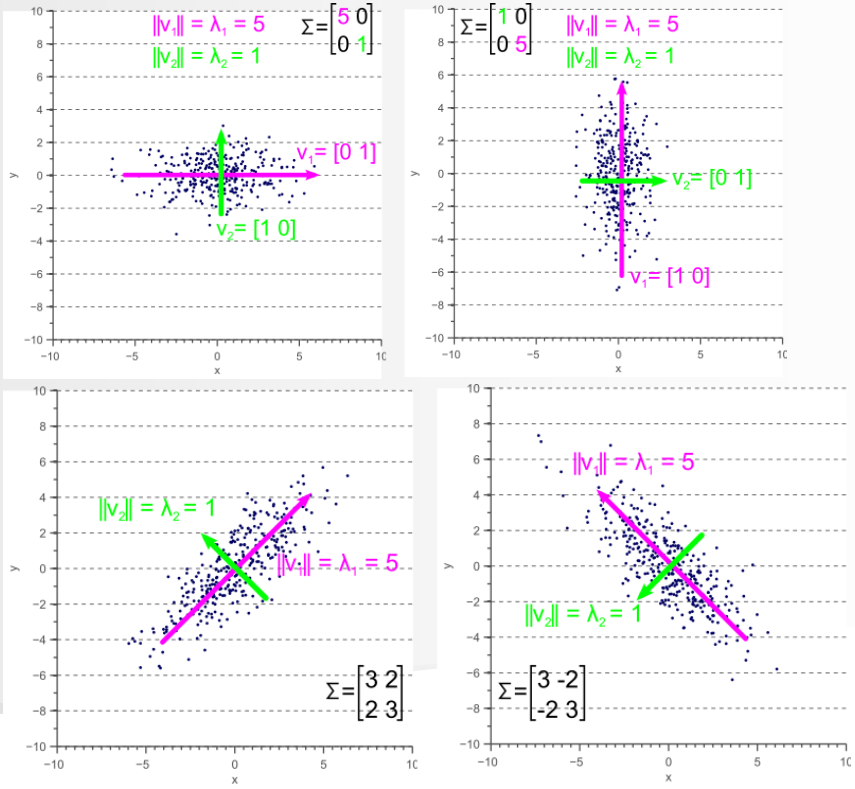
- **Properties of eigenvectors of covariance matrix:**

  - ✓ The eigenvectors of the covariance matrix ∑ are the directions in which the data varies the most.
  - ✓ All the eigenvectors are orthogonal to each other.
  - ✓ The first eigenvector is the direction of the largest variance.
  - ✓ The second eigenvector (which is orthogonal to the first eigenvector) is the direction in which data varies the most among those directions orthogonal to the first eigenvector, etc.

Mathematical Preliminaries

- ✓ If the covariance matrix of our data is diagonal such that the covariances are zero (the direction of the variances are axis-aligned), eigenvalues are equivalent to the variances of the data.

Eigenvalues are equivalent to variance in a diagonal covariance matrix

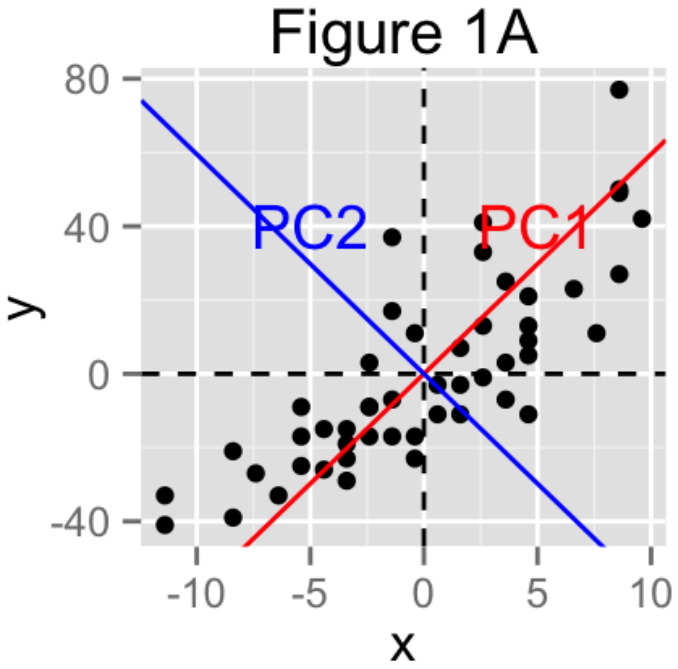


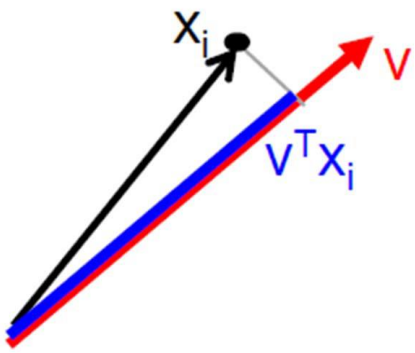
Diagonal covariance matrix

Non-diagonal covariance matrix

Principle Component Analysis (PCA)

- **A statistical data analysis, feature extraction, and data reduction technique**
- Orthogonal transform that projects (transforms) the data into new subspaces where the transformed data (the projected data) are uncorrelated.
- The first Principal component (PC) is a direction where the projection of the data into that direction has the maximum variance, the second PC is a direction orthogonal to the first and has the second maximum variance of the projected data, the third PC is a direction orthogonal to the first and second and has the third maximum variance of the projected data, so on…





Example: Projection of X_i onto the first PC is $V^T X_i$

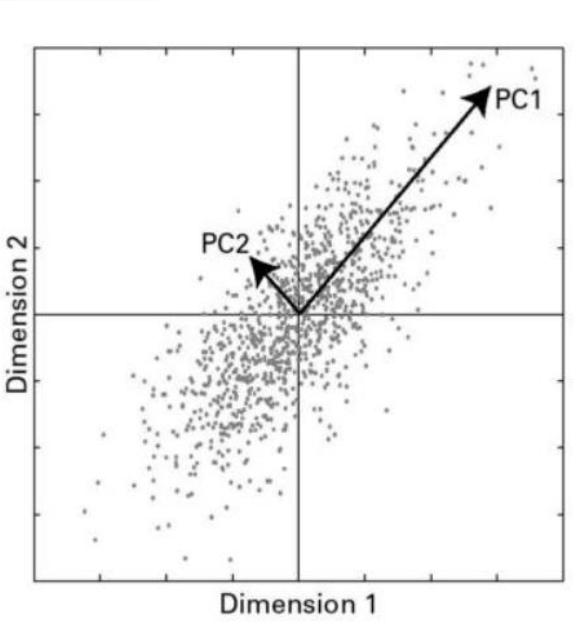
Principle Component Analysis (PCA)

How to find the PCs?!

- ✓ Consider vector X has elements $x_1, x_2, \ldots, x_n$ where the vector X is centered by subtracting the mean $\mathrm{X} \rightarrow X - E\{X\}$
- ✓ Consider a linear combination of X

$$y_1 = \sum_{k=1}^{n} w_{k1} x_k = \mathbf{w}_1^T \mathbf{x}$$

- ✓ $w_{11}, \ldots, x_{n1}$ are the scalar coefficients or weights.
- ✓ We know that y_1 is called the first principal component of x if it has the highest variance



Principle Component Analysis (PCA)

- Maximize the variance of the first PC assuming the data is centered:

$$J_1^{PCA}(\mathbf{w}_1) = \mathrm{E}\{y_1^2\} = \mathrm{E}\{(\mathbf{w}_1^T\mathbf{x})^2\} = \mathbf{w}_1^T\mathrm{E}\{\mathbf{x}\mathbf{x}^T\}\mathbf{w}_1 = \mathbf{w}_1^T\mathbf{C}_\mathbf{x}\mathbf{w}_1$$

$$\text{so that } \|\mathbf{w}_1\| = 1$$

Objective function

Euclidian norm (magnitude)

C_x is the covariance

Constraint: since variance depends on both norm and orientation of the weights (W1) and grows without limits as the norm grows, thus we impose a constraint that the norm of W1 is constant, in practice equal to 1.

Recall: $\|\mathbf{w}_1\| = (\mathbf{w}_1^T\mathbf{w}_1)^{1/2} = [\sum_{k=1} w_{k1}^2]^{1/2}$

$$max_{W1}(W_1^T C_x W_1) \quad s.t. \quad W_1^T W_1 = 1$$

Principle Component Analysis (PCA)

$$max_v\left(W_1^T C_x W_1\right) \quad s.t. \quad W_1^T W_1 = 1$$

- **Recall: Lagrange multiplier**

- One of the strategies for optimization where we want to find a local maxima and minima of a function subject to equality constraint.

maximize $f(x, y)$
subject to $g(x, y) = 0$.

- Assuming that both f and g have continues first partial derivatives, we assume λ, new variable called Lagrange multiplier, and then continue based on the Lagranigian function.

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot g(x, y),$$

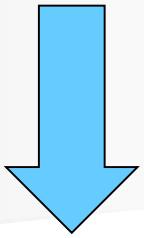
Principle Component Analysis (PCA)

$$max_{W1}\left(W_1^T C_x W_1\right) \quad s.t. \quad W_1^T W_1 = 1$$

- ✓ Lagranian function:

$$max_{W1}(W_1^T C_x W_1 - \lambda\, W_1^T W_1)$$

$$\partial/\partial W1 = 0 \qquad (C_x - \lambda \mathbf{I})\, W_1 = 0$$



$$C_x W_1 = \lambda W_1$$

This is a general form of eigenvalue decomposition!

- ✓ Thus, W_1 is the eigenvector of covariance matrix of X (i.e., C_x).

PCA in EEG

- EEG data over time can be conceptualized as a line traveling through high-dimensional space (electrodes)

- **Goal of PCA:**
- Constructs weights (PCs) based on the covariance of some correlated variables (in EEG the correlated variables are the electrodes).
- These components explain all the variance of the data such that the components are
  - (1) uncorrelated with each other
  - (2) the first PC has the most variance, the second PC has the most residual variance and orthogonal with the first PC, so on…

- **Interpretations:**
  - PCA creates a set of spatial filters in which weights for each electrodes are defined by patterns of interelectrode temporal covariance
  - Highlights specific features of the data
  - Data reduction techniques, useful for high dimensional data (assumes that the PCs for high variance are signal and the PCs with low variance are noise).

PCA in EEG

- ✓ **PCA vs. Laplacian filter:**

1- Laplacian filters are defined based on the distance (physical locations).

vs.

PCA weights are defined based on statistics of the data (no physical location consideration).

2- Laplacian removes low spatial frequencies and is focused on high spatial frequencies (localized activities)

vs.

PCA focuses on global activities (lower spatial frequencies). Thus, local activities are not likely to appear at first components.

PCA in EEG

- **Two main terms in PCA on EEG**: 1-observations, 2-variables.
  - PCA finds the patterns of covariance among the variables.
- Depends on the application we define the variables (e.g find the covariance patterns across the channels, thus, channels are the variables and observations are the time series).

- **How to compute PCA on EEG?**
- Example:
  - Assume x is the data (electrodes x time points)
  - Find the covariance of X.
  - Note: X needs to be centered (meaning that the mean should be subtracted from the data) and the first dimension is electrodes and the second is the time points (n).

$$covariance = (n-1)^{-1}\left(X-\bar{X}\right)\left(X-\bar{X}\right)^{T}$$

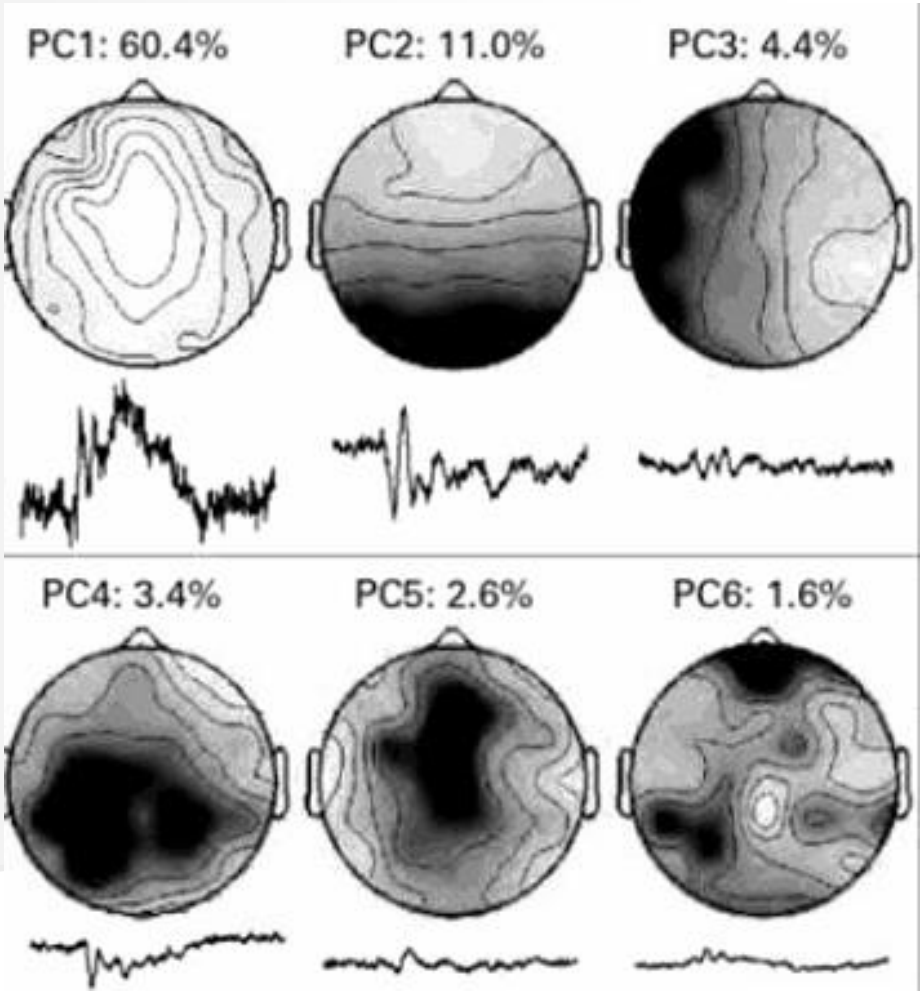
  - Perform the eigendecomposition of the "covariance".

PCA in EEG

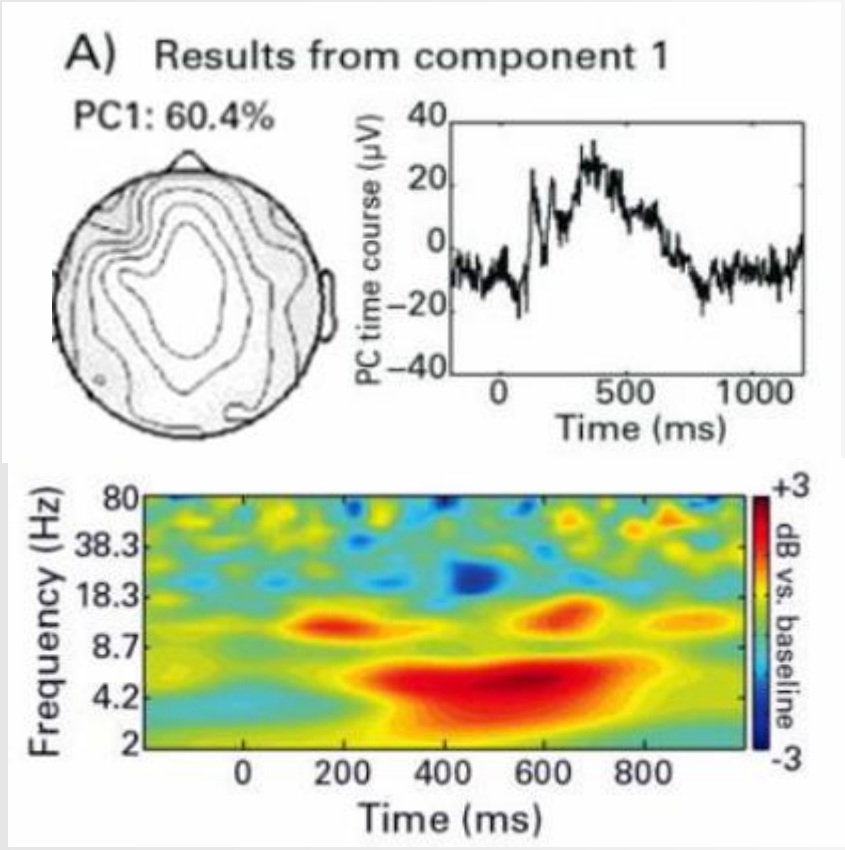
- **Eigenvectors are the PCs!**
- ✓ Note: Each column in the obtained eigendecomposition matrix (i.e. a square matrix) is PC and each row is the weights
- ✓ The importance of the PCs are related to the amount of their variance obtained by their eigenvalues.
- ✓ To understand each electrode weights for each PC, we can plot each PC over the entire head.
- ✓ Each PC can be projected back to the time domain which gives us the time course of each PC.
  - ✓ The time course of each PC is the weighted sum of the activity of all the electrodes.
- ✓ The first few components account for "a lot" of variance and the last components have little variance.
- ✓ Usually in PCA the first several components are considered as signal and the last ones are considered as noise. But this is not all the time true (we need to have visual inspection to make sure).

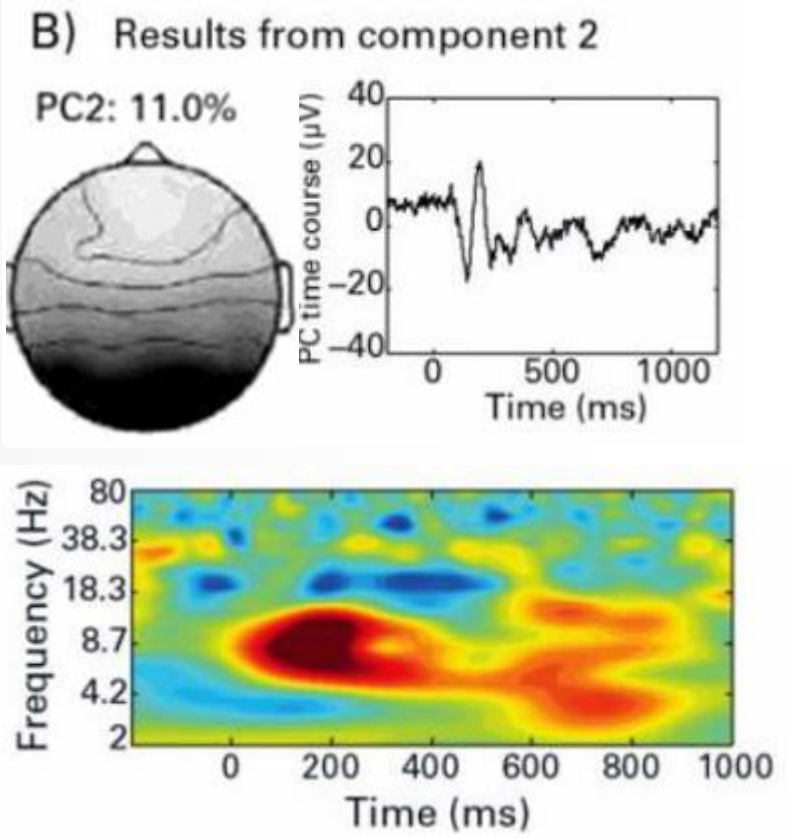
PCA in EEG

The first 6 principal components (PCs), their percentage of variance, and the associated time course



PCA in EEG



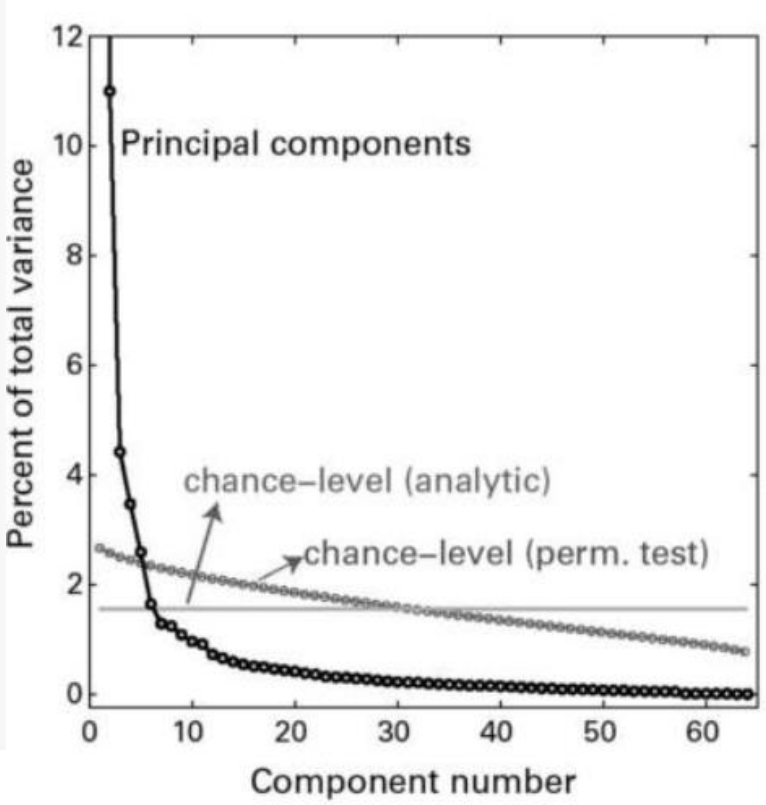


Dissociation between early sensory (PC2) processing and later cognitive processing (PC1) on the same data.

PCA in EEG

- **What PCs are significant?**
  - Define percentage variance threshold.
  - PCs with variance above the threshold can be considered as significant and below that as non-significant.

- **How to define a variance threshold?**
  - **Analytic approach:** 100/M considering M numbers of electrodes/components.
  - **Permutation testing:** randomly shuffle the data many times (e.g. 1000 times), obtain the PCA and thus amount of variance on the shuffled data, average over all the repetition is the threshold.



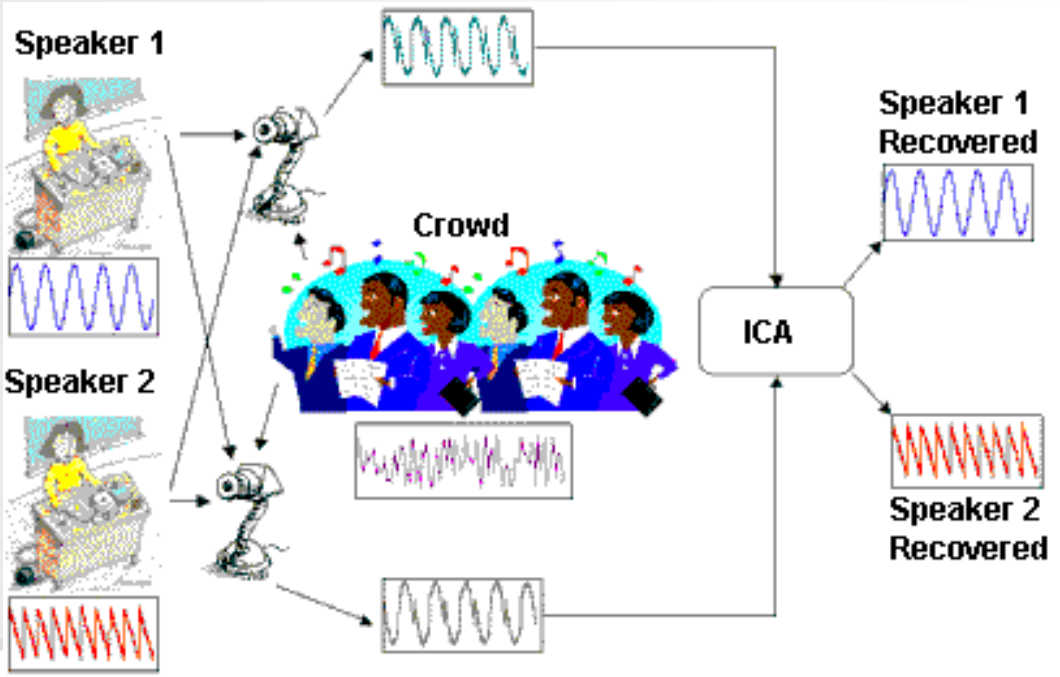
PCA in EEG

- **Now what we can do with the significant PCs?**
  - Use PCA for the dimensionality reduction by analyzing only the significant components instead of all of the channels
  - Treat PCA as a de-noising option
  - PCA can be used as a measure of complexity of a system. The more complex the system is the more independent component we have; thus, the more components are above the threshold (Note: for the simple systems the eigenvalue function decline more steeply).

- **Note: PCA mostly focuses on low spatial frequency activities and larger topographical covariances. Thus, local activities might be neglected in PCA even if they reflect meaningful patterns.**

Independent Component Analysis (ICA)

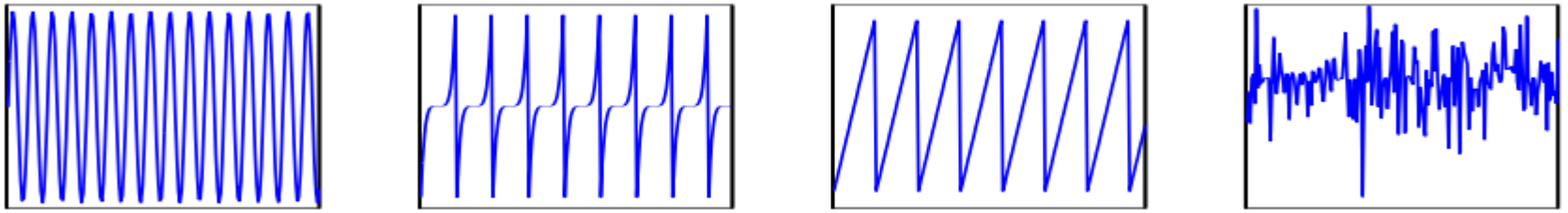
- ✓ The classical "cocktail party" problem.
- ✓ Separates the mixed signal into sources



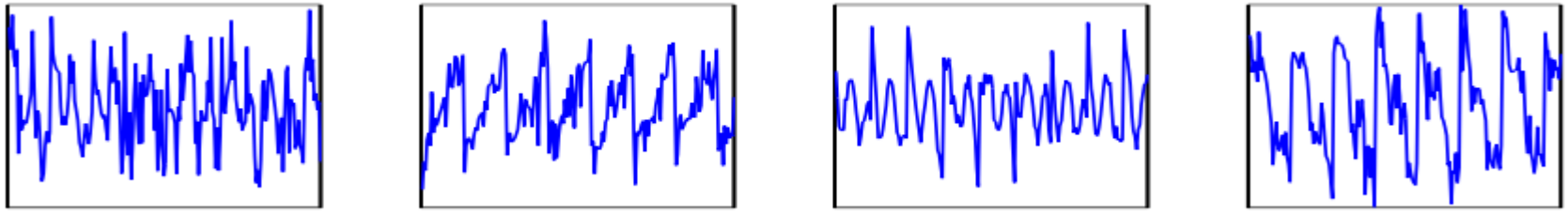
http://www.statsoft.com/Textbook/Independent-Components-Analysis

Independent Component Analysis (ICA)

Source signals



Observed signals (linear mixture of source signals)



Now we want to estimate the source signals

Independent Component Analysis (ICA)

Recall:

- **Basic idea in PCA:** Find directions that maximize the variance of the signal.

$$y_1 = \sum_{k=1}^{n} w_{k1} x_k = \mathbf{w}_1^T \mathbf{x}$$

- **Goals in PCA:**
- ✓ De-noisng
- ✓ Dimensionality reduction
- ✓ Vain hope: finds original underlying components!!

Independent Component Analysis (ICA)

Why PCA can not find the sources (original signals)?

- ✓ They only use covariances $cov(x_i, x_j)$
- ✓ Covariances are symmetric, thus, only $n^2/2$ equations would be available
- ✓ Estimating the mixing matrix needs estimating n^2 parameters. So, not enough information.

Independent Component Analysis (ICA)

- ✓ Given a set of observations of random variables: $(x_1(t), x_2(t), \ldots, x_n(t))$, where t is the time/sample index, assume that they are generated as a linear mixture of independent sources:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix}$$

Where A is the mixing matrix and s is source (hidden factors).

- ✓ Using ICA we would like to find linear transformation given by $(A^{-1})X$, so that random variables $S_i\{i = 1, \ldots, n\}$ are as independent as possible.
- ✓ To fully recover N sources we need to have at least N observations (e.g. microphones, EEG electrodes) available

Independent Component Analysis (ICA)

Problem: Estimate both A and S observing only X.

- **Assumptions in basic ICA:**
  - ✓ The sources are statistically independent from each other (i.e. the information from one variable does not provide any information from the other variable).

  Random variable $y_1, y_2, \ldots, y_n$ are independent if and only if the joint pdf can be defined by multiplication of the all marginal pdfs:

  $$P(y_1, y_2, \ldots, y_n) = p_1(y_1)\, p_2(y_2) \ldots p_n(y_n)$$

  - ✓ The values in each source signal have non-Gaussian distribution. Gaussian distributions are "too simple" and do not have higher order statistical moments which is necessary for ICA.
  - ✓ For simplicity we assume that unknown mixing matrix is square (i.e. the number of sources is equivalent to the number of observed mixtures).

Then both mixing matrix and components can be identified!!

Independent Component Analysis (ICA)

- **Two effects of mixing source signals:**
  - ✓ Independence: The source signals are independent but the signal mixtures are not (as they are sharing the same sources)
  - ✓ Normality: Based on Central Limit Theorem (CLT), the distribution of a sum of independent random variables with finite variance tends towards a Gaussian distribution.
    - ✓ So, roughly any mixture of the components will be more Gaussian than the components themselves.

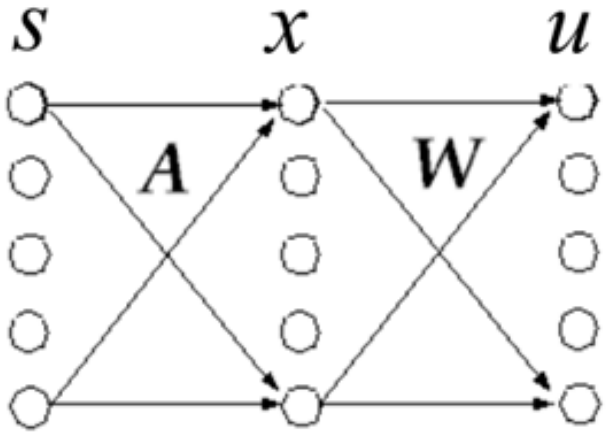
Independent Component Analysis (ICA)

The general problem in ICA:

We want to recover the original version of the sources (S) by multiplying the data (X) by an unmixing matrix (W)

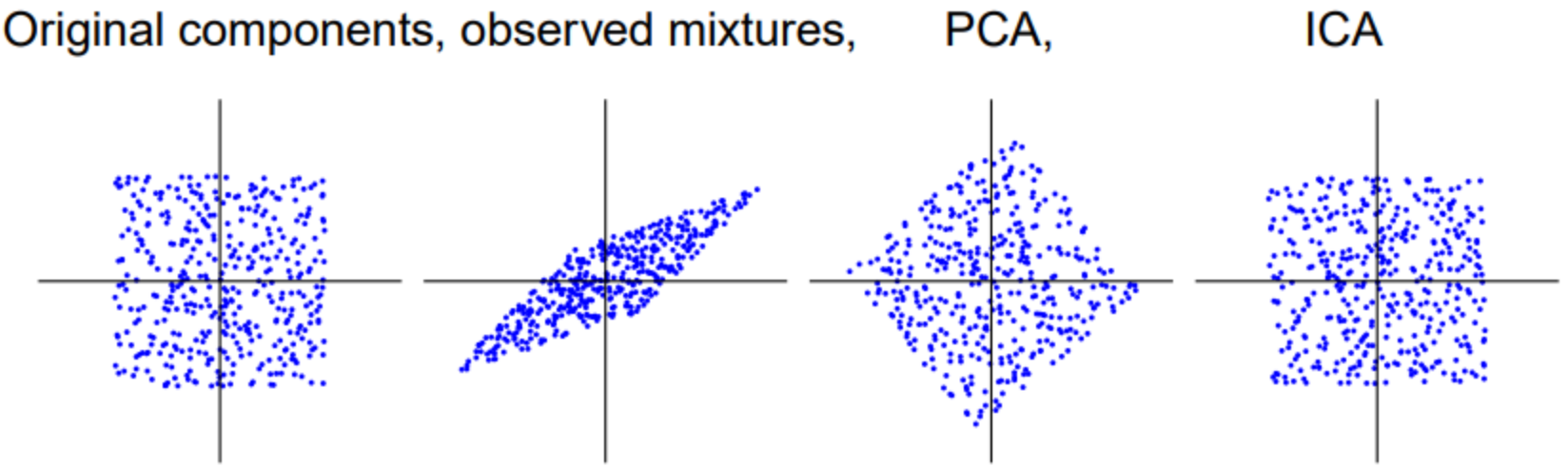
$$u = Wx,$$

$$x = As.$$



PCA vs. ICA

- PCA decorrolates, ICA demixes
- PCA is based on second order statistics, ICA is based on higher order statistics (e.g. kurtosis)
- ICA uses iterative methods based on minimizing the cost functions to adjust the weights, PCA does not use any iteration methods.
- PCA components are constrained to be orthogonal to each other. ICA does not assume sources are orthogonal.



PCA does not find the original coordinates but ICA does!

Independent Component Analysis (ICA)

PCA maximizes variance of the projected data (i.e., $\sum_i w_i x_i$), ICA maximizes the independence

- **How to maximize the independence?**

1- Minimize the mutual information

2- Maximize the non-gaussianity

Independent Component Analysis (ICA)

- **Three main ingredients in ICA:**

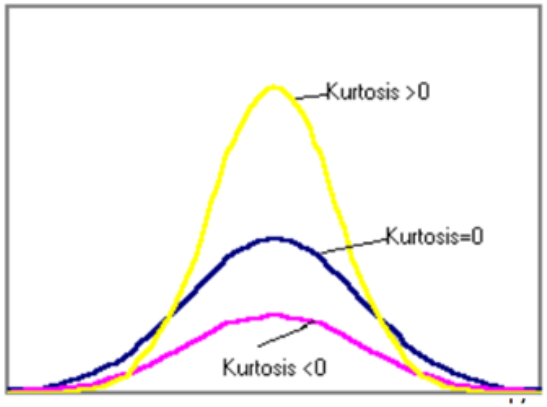
1- Nongaussianity measure

- ✓ **Kurtosis:** a classic approach but sensitive to outliers, and thus not robust (Its value may depend on only a few observations in the tails of the distribution which may be erroneous or)

Gaussian Dis. Kurtosis = 0

Super-Gaussian: kurtosis > 0

Sub-Gaussian: kurtosis < 0



Independent Component Analysis (ICA)

2- Mutual Information

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\, p(y)} \right)$$

Where P(x,y) is the joint PDF, P(x) and P(y) are the marginal PDF.

3- Optimization algorithm

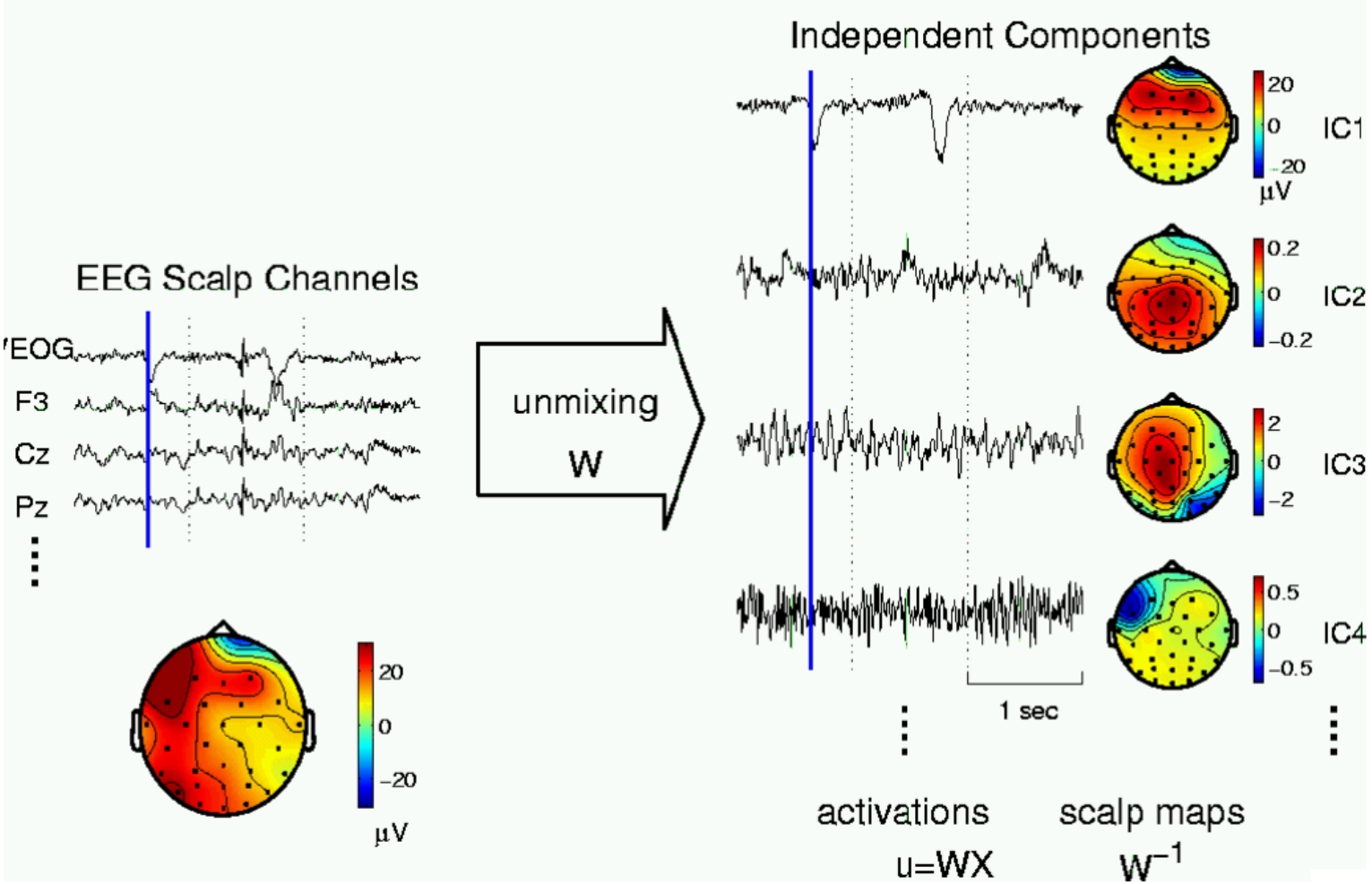
- ✓ Gradient based method: "infomax"
- ✓ Fixed-point algorithm: FastICA

ICA in EEG

ICA is a standard and widely used technique in EEG data analysis mostly to clean the data prior to analysis.

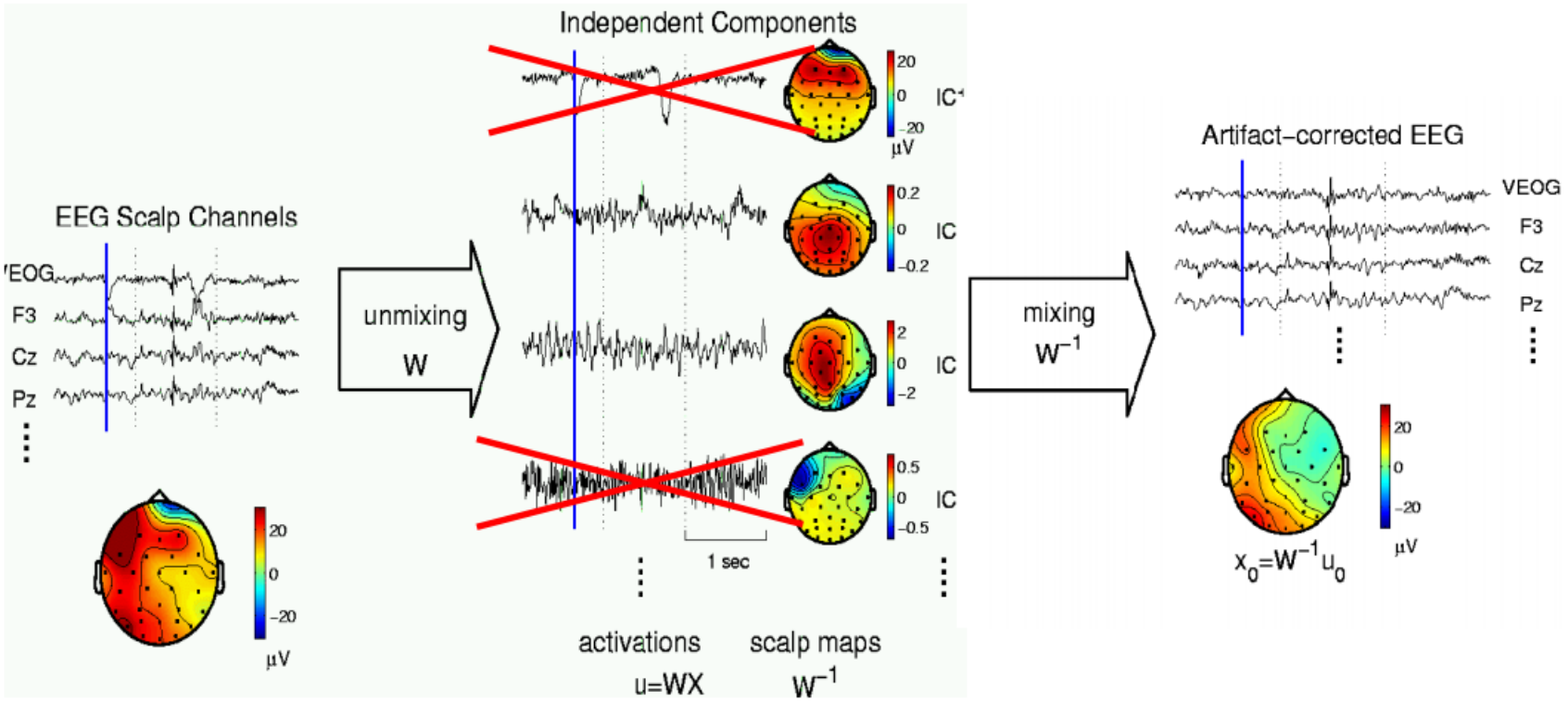
- ✓ Sources in EEG are contaminated by volume conduction and noises.
- ✓ We assume the mixing process is linear and the sources (component time courses) are independent.
- ✓ The maximum number of components is equivalent to the number of electrodes (number of components ≤ number of electrodes)
- ➢ ICA can not tell which component is noise which is signal; thus, needs careful inspection of signal components to avoided not to remove the components that have good amount of signal

ICA in EEG

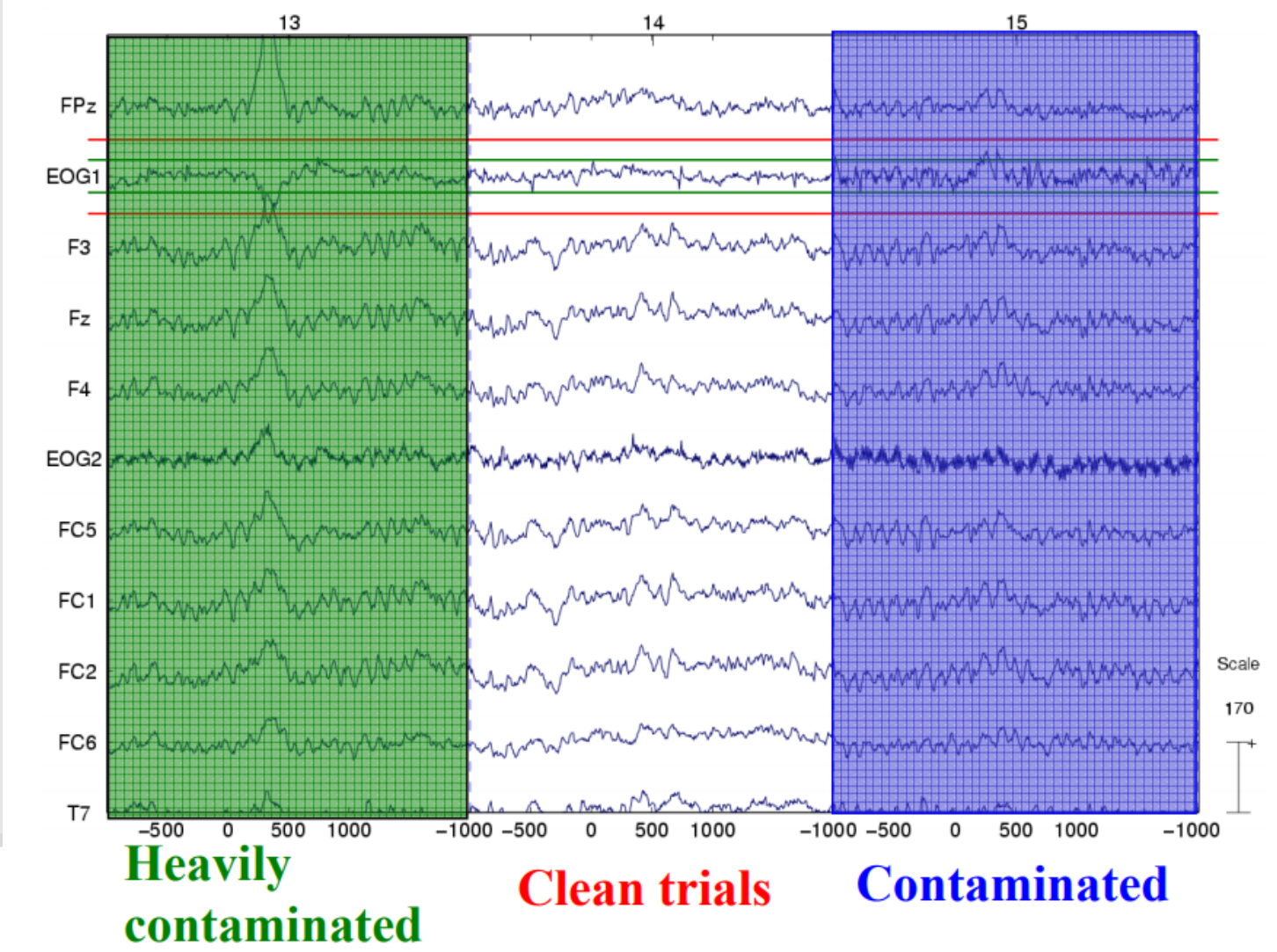


From Jung et al., *Clinical Neurophysiology*, 2000.

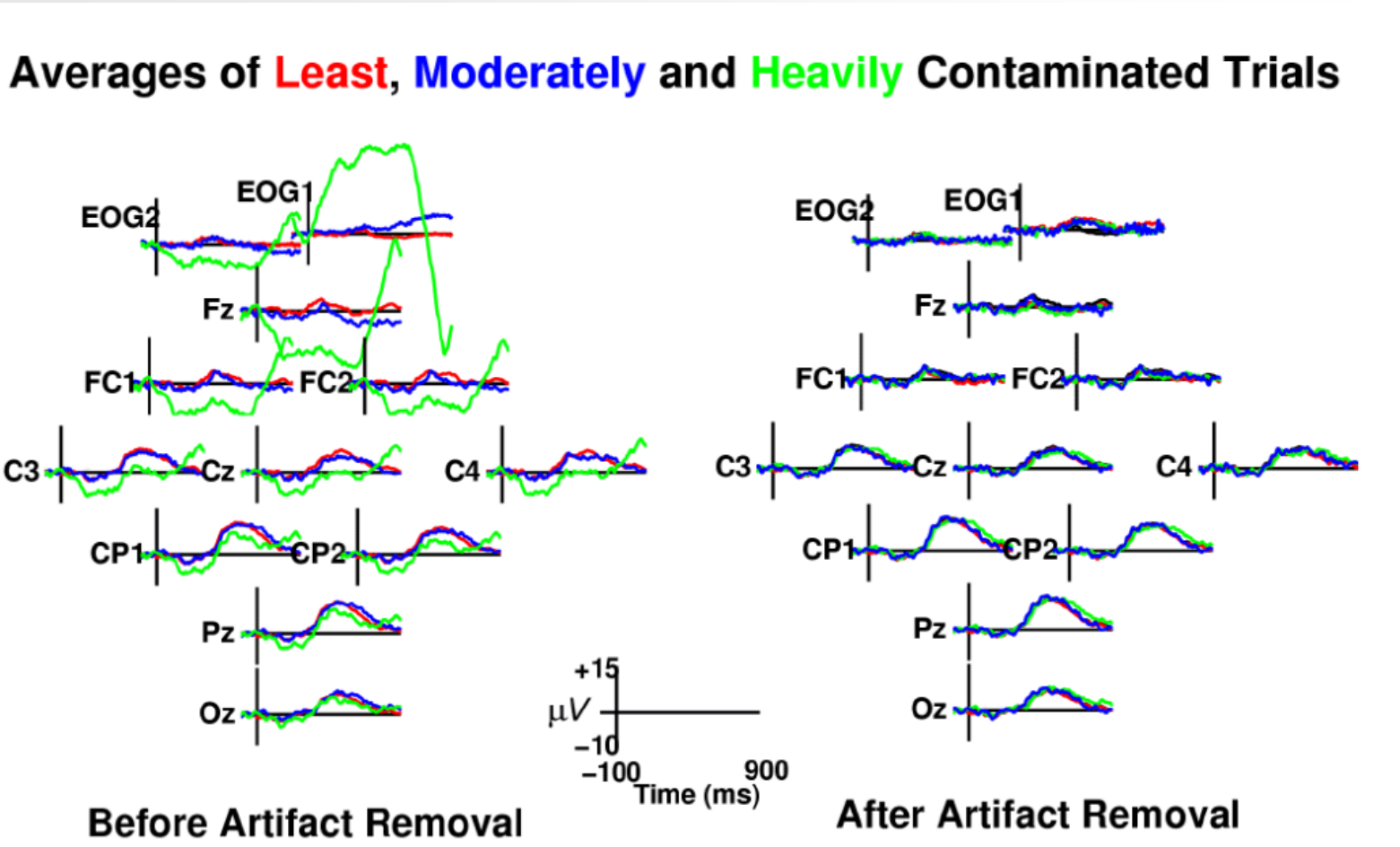
ICA in EEG


EEG Scalp Channels
VEOG
F3
Cz
Pz
20
0
−20
μV
unmixing
W
Independent Components
20
0
−20
μV
IC
0.2
0
−0.2
IC
2
0
−2
IC
0.5
0
−0.5
IC
1 sec
activations
u=WX
scalp maps
W−1
mixing
W−1
Artifact−corrected EEG
VEOG
F3
Cz
Pz
20
0
−20
μV
x0=W−1u0

ICA in EEG


13
14
15
FPz
EOG1
F3
Fz
F4
EOG2
FC5
FC1
FC2
FC6
T7
−500 0 500 1000 −1000 −500 0 500 1000 −1000 −500 0 500 1000 −1000
Scale
170
Heavily contaminated
Clean trials
Contaminated

ICA in EEG



From Jung et al., *Clinical Neurophysiology*, 2000.

ICA in EEG