Independent Component Analysis

19 January 2021

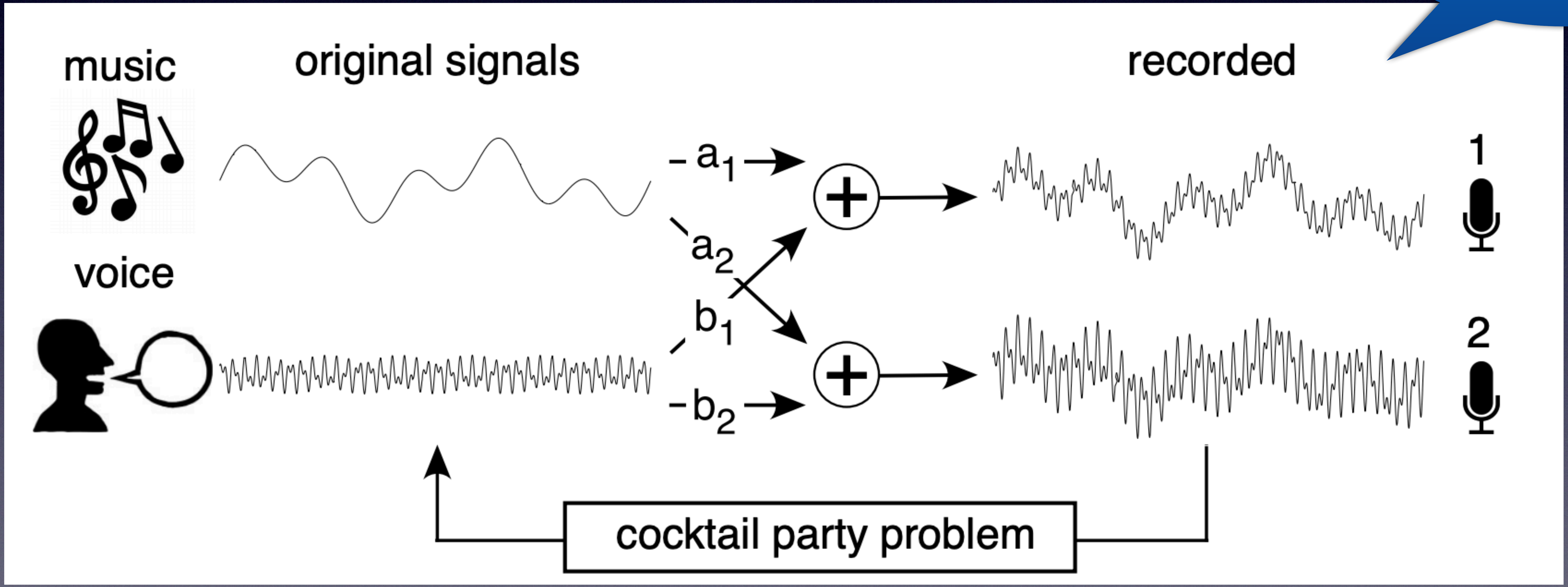
Rajesh R (S21005), MS Research Scholar

Cocktail Party Problem

Ref: https://jundurraga.github.io/COGS703/?print-pdf#/

Blind Source Separation

1



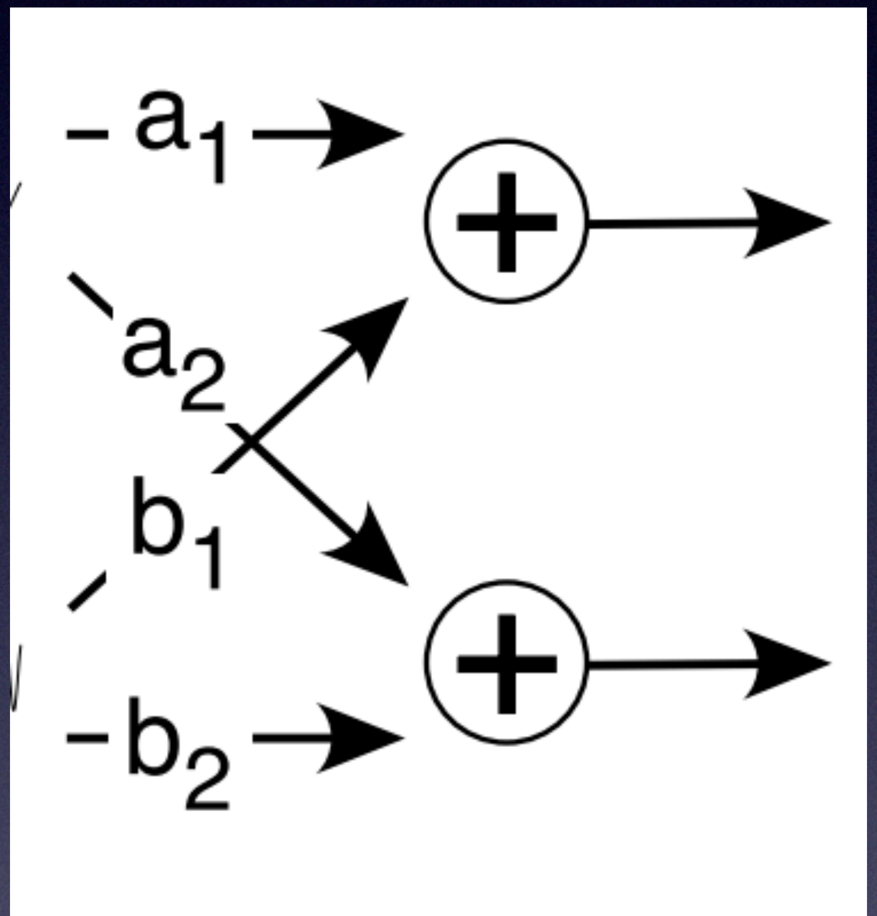
No Information or Little Information

Independent Component Analysis (ICA)

Mathematical Formation

Source 1

Source 2



Observed 1

Observed 2

Linear Combinations

Intuition ?

$$x = As$$

Observed Signals (Column Vector)

Mixing Matrix

Source Signals (Column Vector)

GOAL

$$x = As$$

Mixing Matrix, $A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$

Linear Combinations of Source 1 and Source 2

$$\hat{s} = Wx$$

Observed signal

Independent signals

W is an approximation of A^{-1} so that $\hat{s} \approx s$

Un-mixing Matrix

Impossible Problem?

Under constrained problem because number of unknown exceed the number of observations

$$\underbrace{x}_{\text{Known}} = \underbrace{A}_{\text{Unknown}}\underbrace{s}_{\text{Unknown}}$$

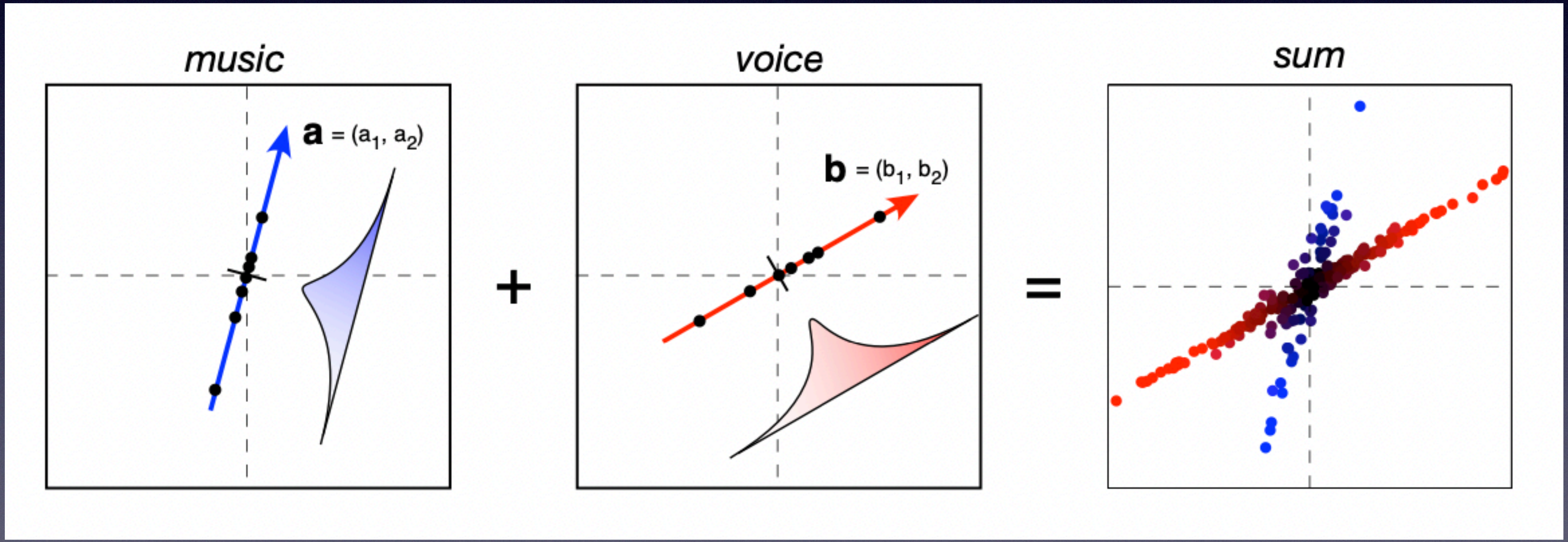
Assumptions

1. Sources are statistically Independent $p(x, y) = p(x)p(y)$
2. Independent components have Non Gaussian distribution
3. A is square and invertible

Why?

Visualisation of linear mixture

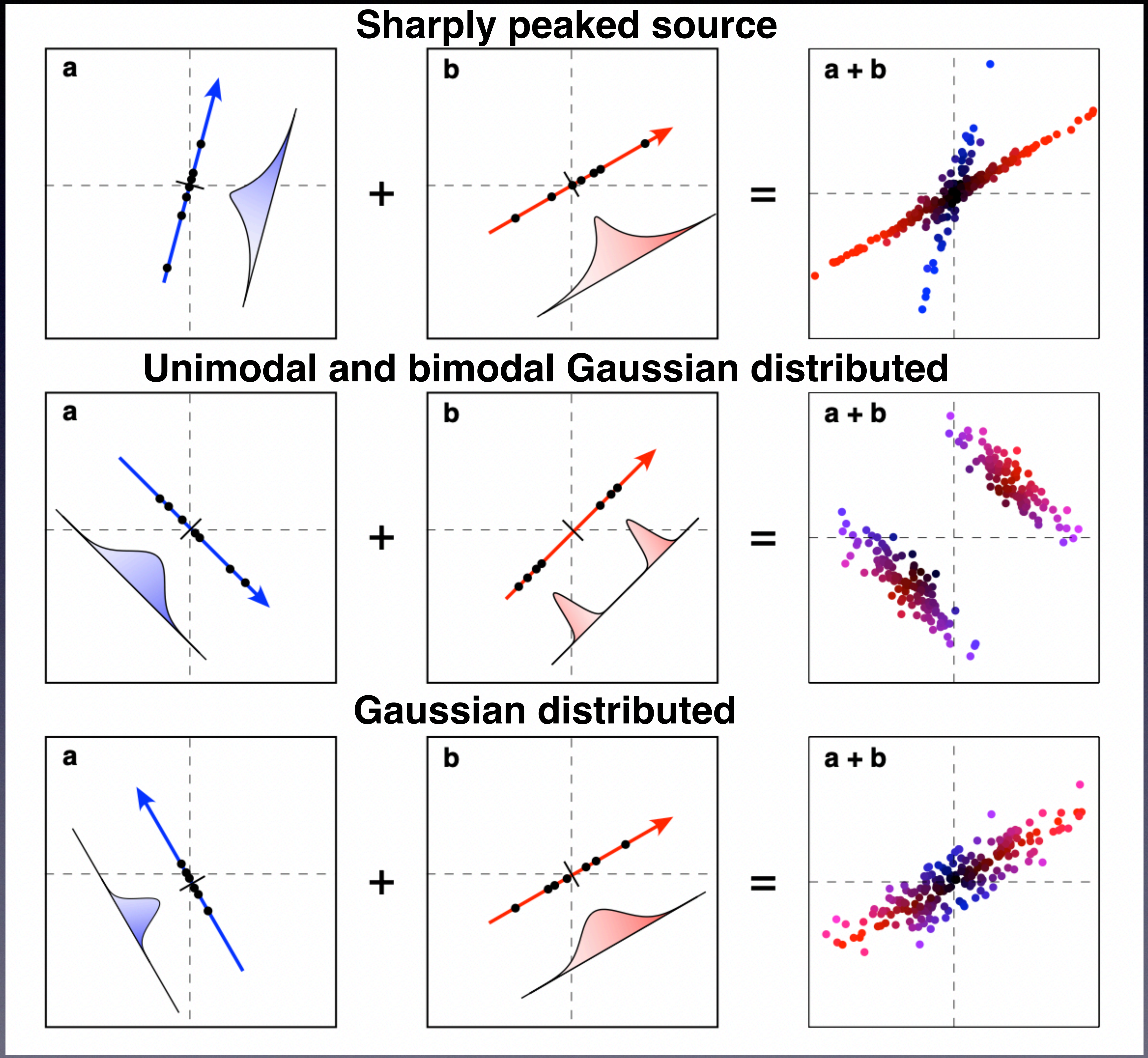
2

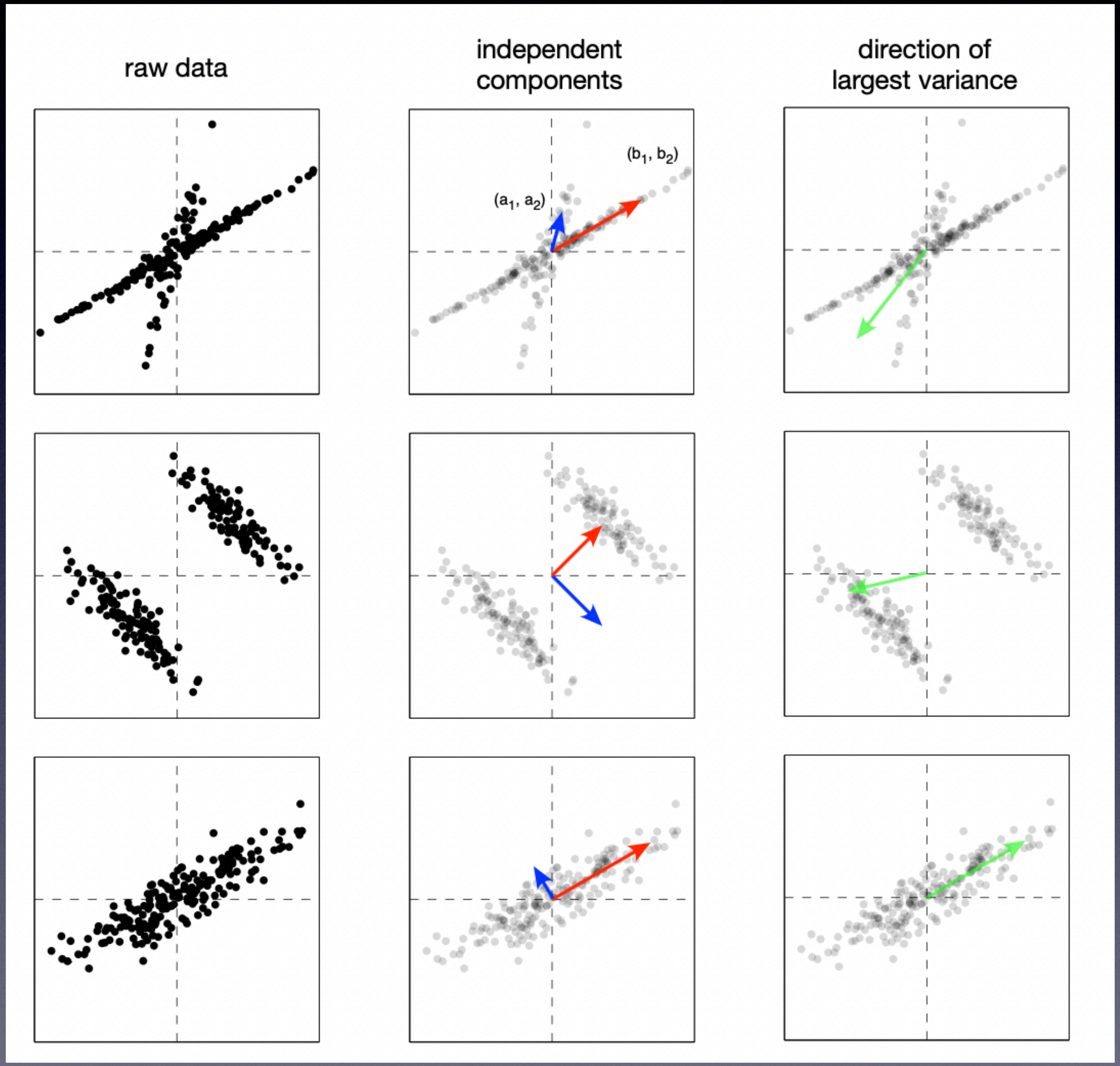


TYPES OF LINEAR MIXTURE

ANALYSIS OF LINEAR MIXTURE

3





Mathematical Preliminaries

Singular Value Decomposition

#1

$$A = U\Sigma V^{\dagger}$$

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} =$$

- Dimensionality Reduction
- Direction of maximum Stretching
- Rotation-Scaling-Rotation

$$A = U\Sigma V^T$$

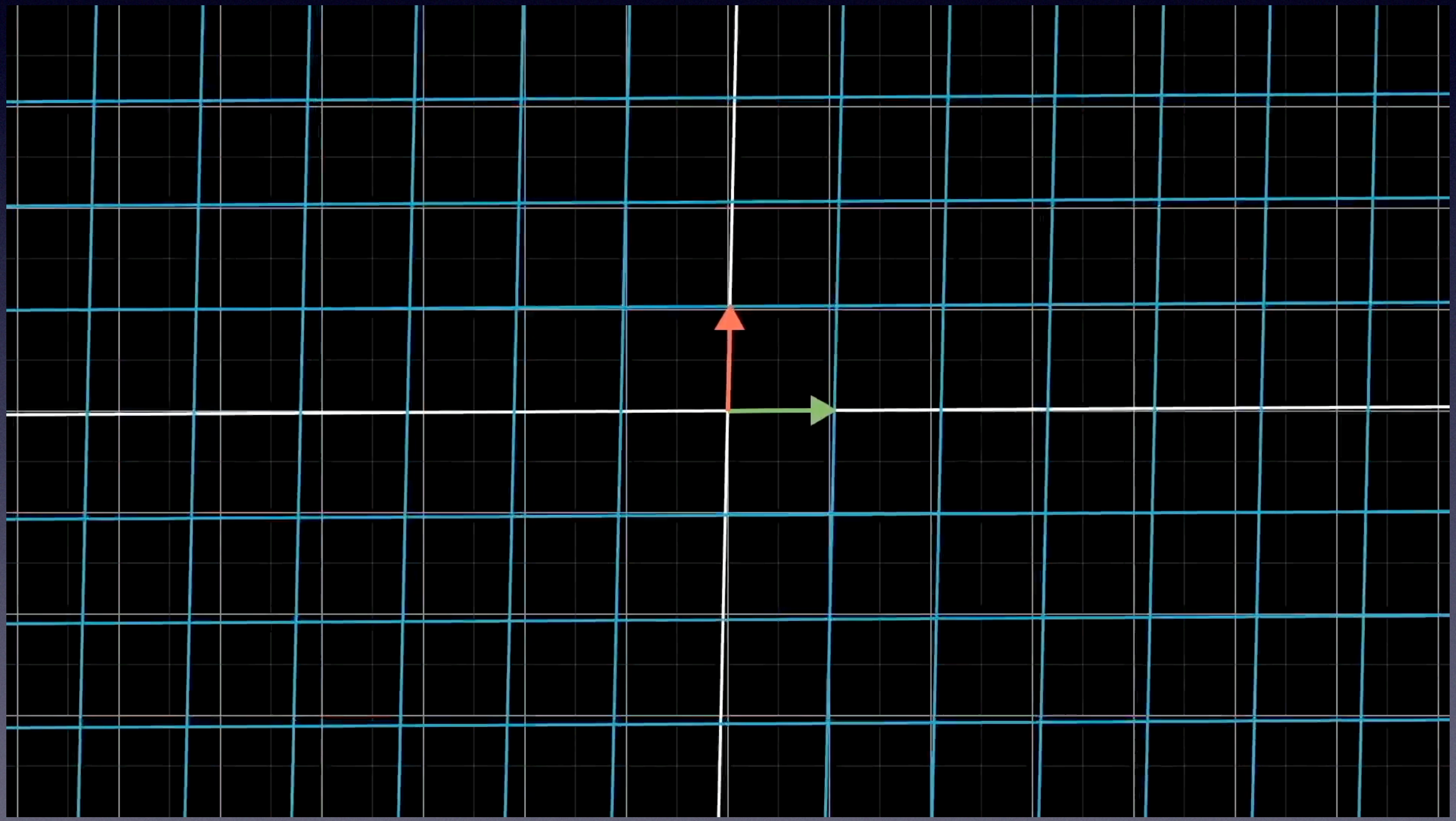
$U =$ Eigen vectors of AA^T

$V =$ Eigen vectors of A^TA

Mathematical Preliminaries

Covariance Matrix and PCA

#2



- Transformation that rotate any vector into the direction of the greatest variance
- PCA uses greatest variance direction as principal component and does dimensionality reduction

Strategy!

Divide and Conquer

$$x = As$$

$$A \qquad S$$

$$\downarrow$$

$$U\Sigma V^T$$

$$\hat{s} = Wx$$

$$W = A^{-1} = (U\Sigma V^T)^{-1}$$

$$W = (V^T)^{-1}\Sigma^{-1}U^{-1}$$

Solution

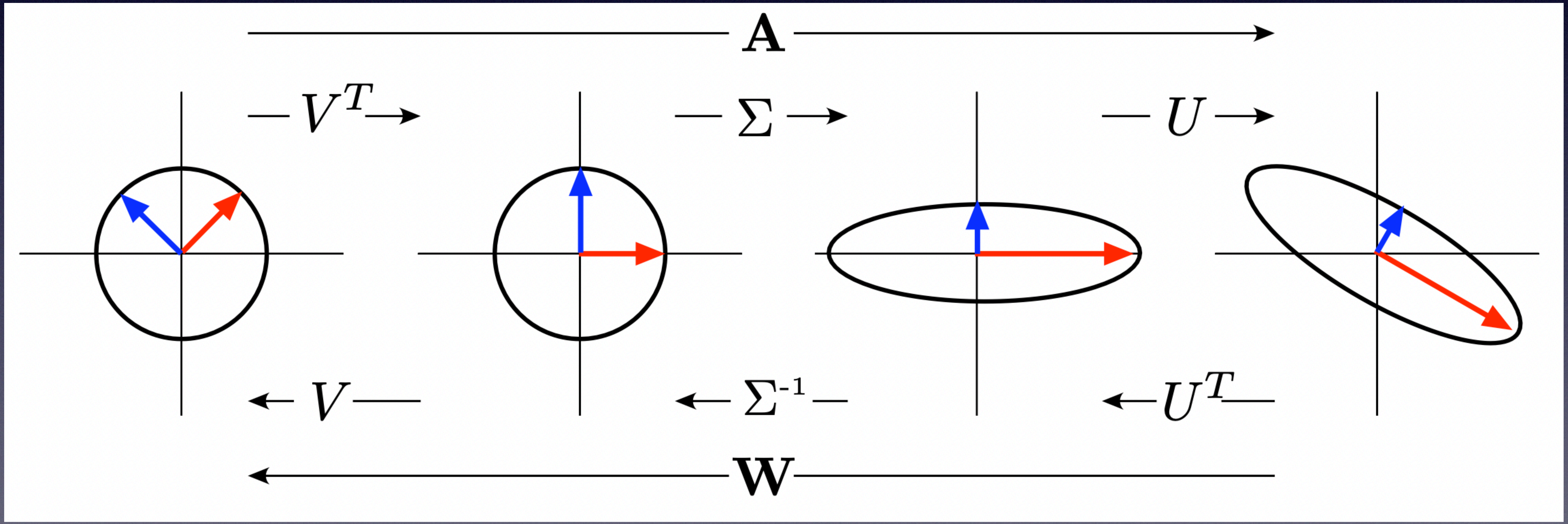
$$W = V\Sigma^{-1}U^T$$

$$W = V\Sigma^{-1}U^T$$

Assumption of independence of s to solve for V

Examine the covariance of the data x inorder to calculate U and Σ

4



Covariance of data

Covariance = Expected value of the outer product of individual data points

$$E[xx^T] = \langle xx^T \rangle = E\left[[x_1 x_2 \ldots x_n][x_1 x_2 \ldots x_n]^T\right]$$

Assumption!

- Zero mean
- Covariance of the source is whitened

$$\langle ss^T \rangle = I$$

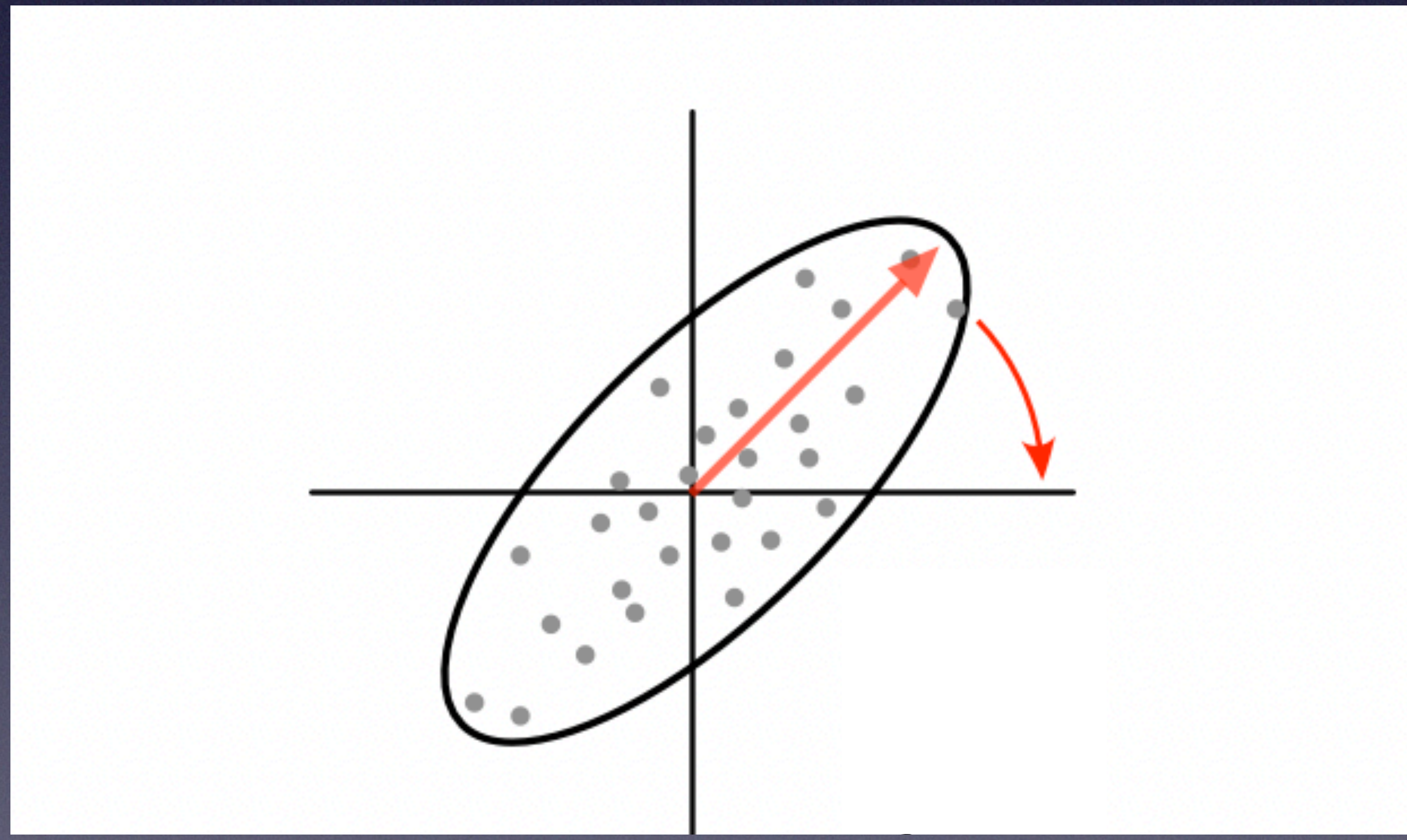
Identity Matrix

Whitening Intuition!

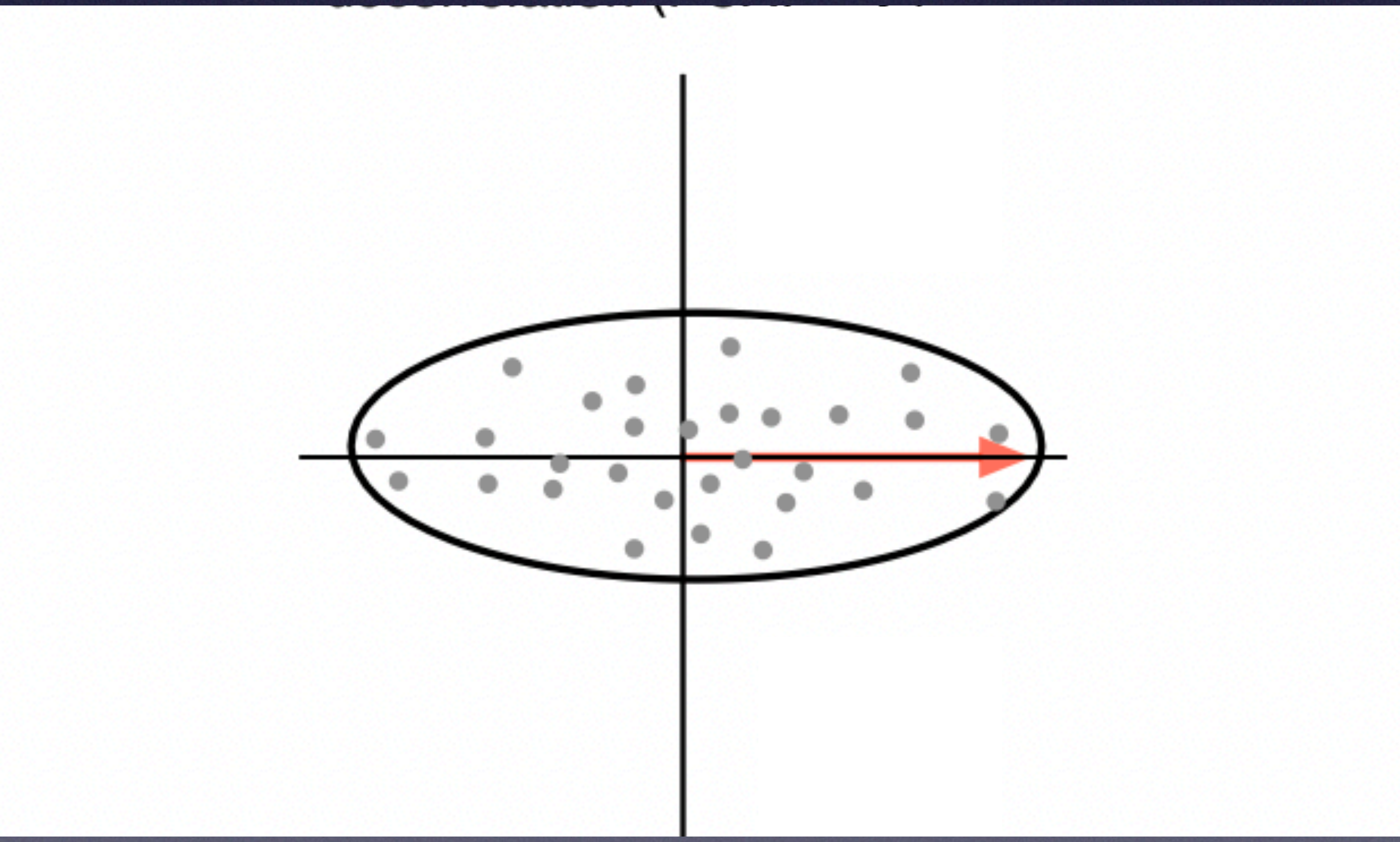
Operation that removes all linear dependencies in data set (second order correlations) & Normalise the variance among all dimensions

Decorrelation PCA **Normalisation**

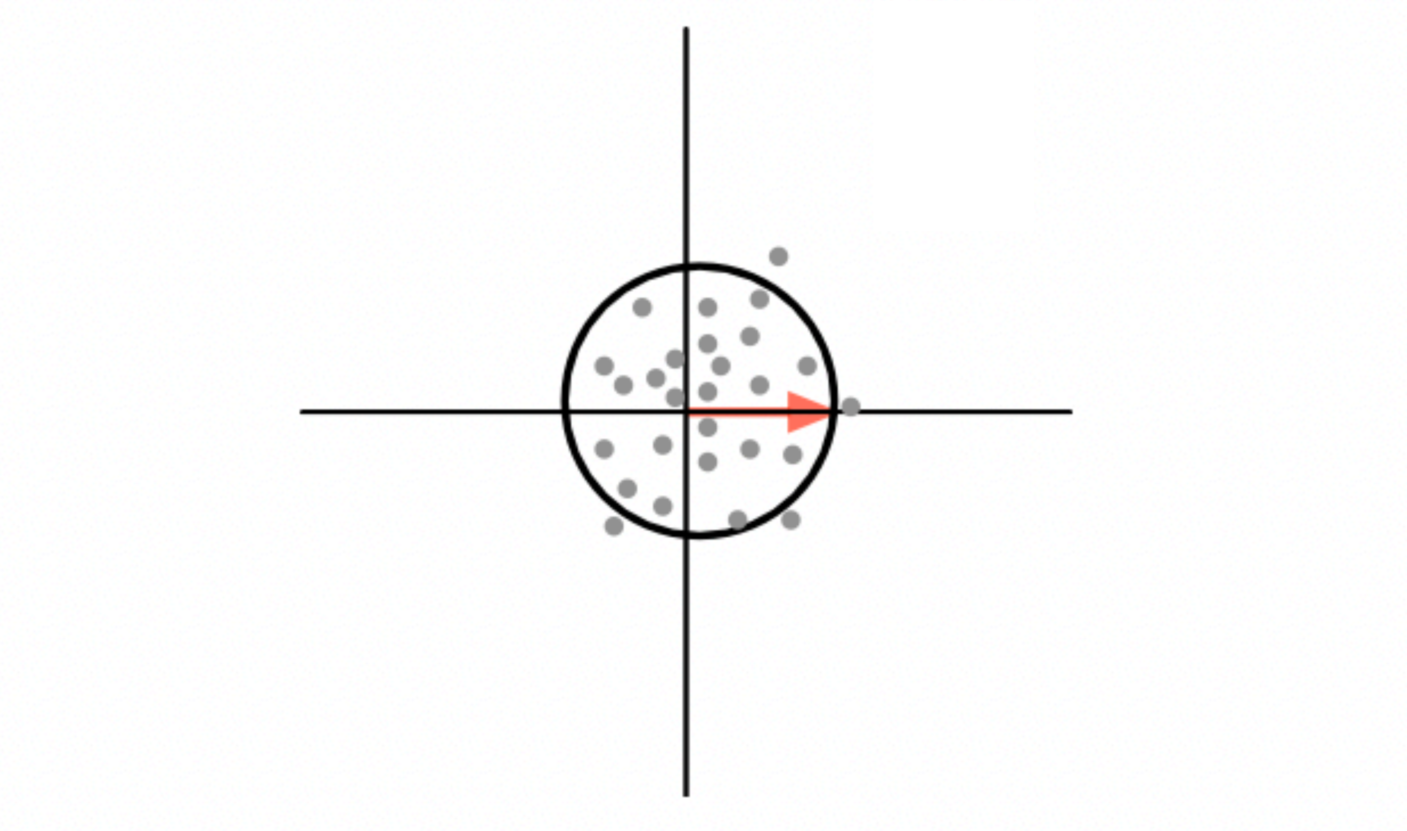
5



$$\xrightarrow{E^T}$$



$$\xrightarrow{D^{-\frac{1}{2}}}$$



Mathematically,

Decorrelation means the covariance of the transformed data is diagonalised

$$\langle xx^T \rangle = EDE^T$$

$$E^T \langle xx^T \rangle = E^T EDE^T$$

$$E^T \langle xx^T \rangle E = DE^T E$$

$$E^T \langle xx^T \rangle E = D$$

$$\langle (E^T x)(x^T E) \rangle = D$$

$$\langle (E^T x)(E^T x)^T \rangle = D$$

Covariance of data

$$\langle xx^T \rangle = \langle (As)(As)^T \rangle$$

$$= \langle (U\Sigma V^T s)(U\Sigma V^T s)^T \rangle$$

$$= \langle U\Sigma V^T s s^T V \Sigma^T U^T \rangle$$

$$= U\Sigma V^T \langle s s^T \rangle V \Sigma U^T$$

$$\langle xx^T \rangle = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T$$

Independent of sources S as well as V

Covariance of data

$$\langle xx^T \rangle = U\Sigma^2 U^T$$

Familiar form!

$$\langle xx^T \rangle = EDE^T$$

$$W = V\Sigma^{-1}U^T$$

$$W = VD^{-\frac{1}{2}}E^T$$

Eigen value of correlation data

Eigen vector of correlation data

Only Unknown

Defining x_w

$$x_w = (D^{-\frac{1}{2}}E^T)x$$

Whitened version of observed data such that $\langle x_w x_w^T \rangle = I$

$$W = VD^{-\frac{1}{2}}E^T$$

$$\hat{s} = Wx$$

$$\hat{s} = Vx_w$$

Statistics of Independence

Statistical independence is the strongest measure of dependency between random variables

$$p(a, b) = p(a)p(b)$$

$$p(s) = \prod_i P(s_i)$$

Why? Coz all second order correlations are removed

Solution using Information Theory

Multi information: $$I(y) = \int p(y) log_2 \left(\frac{p(y)}{\prod_i p(y_i)} \right) dy$$

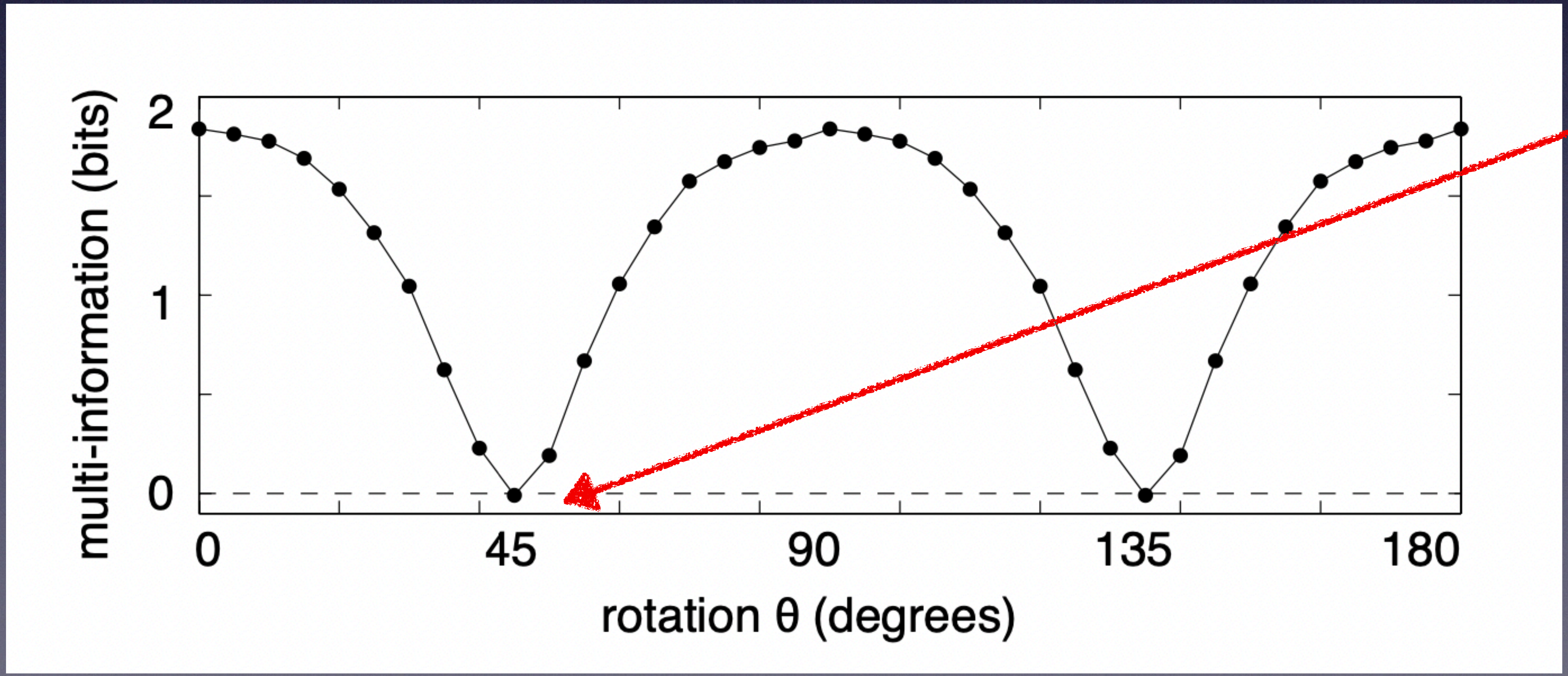
Non negative quantity that reaches minimum of zero only if all variables are statistically independent

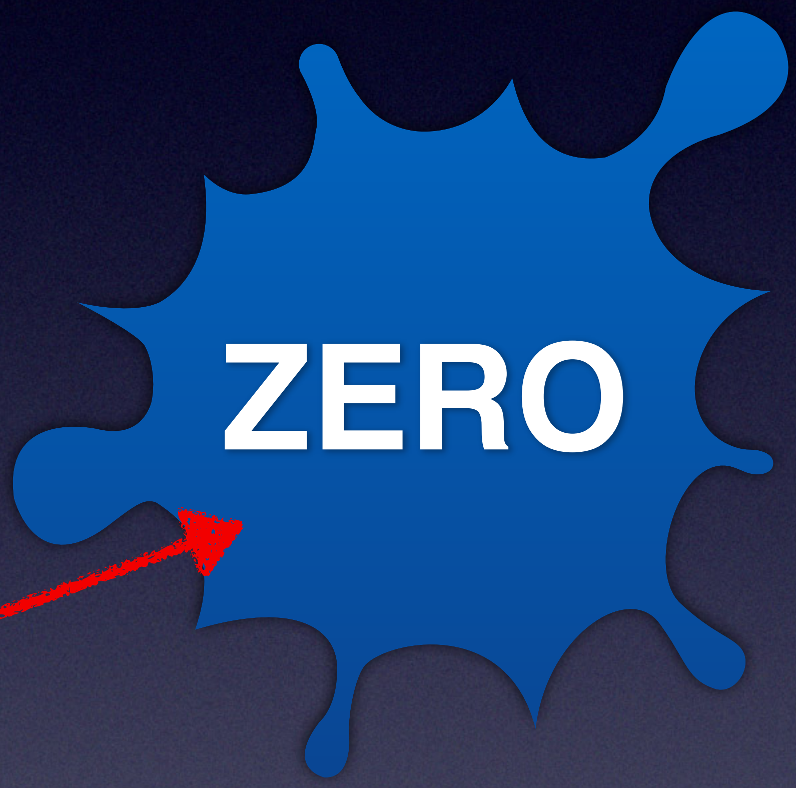
If $p(y) = \prod_i p(y_i)$, then $log_2(1) = 0$ so, $I(y) = 0$

Goal of ICA: Minimise the multi information until it reaches zero

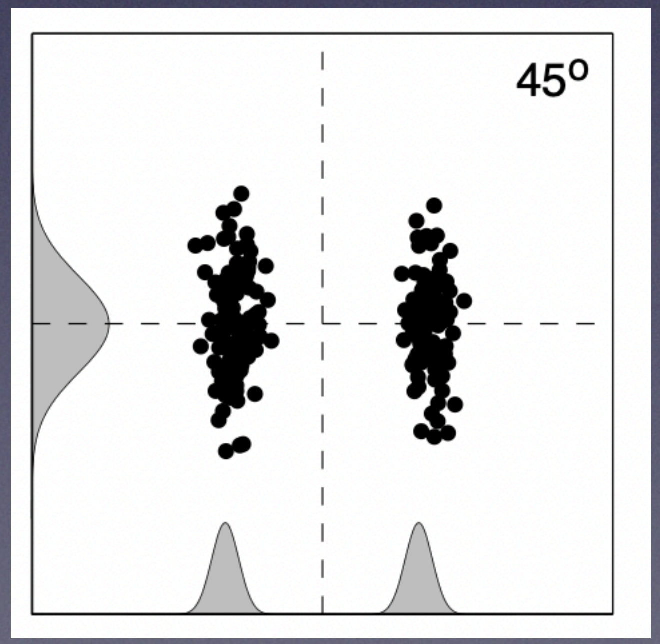
$$V = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

6





7



Entropy: Amount of uncertainty about a distribution $p(y)$

$$H(y) = -\int p(y) log_2 p(y) dy$$

Multi information is given by:

$$I(y) = \sum_i H(y_i) - H(y)$$

Entropies of marginal distribution

Entropies of joint distribution

Towards rotation V

$$I(\hat{s}) = \sum_i H[(Vx_w)_i] - H(Vx_w)$$

For linear transformation V and random vector x_w,

$$H(Vx_w) = H(x_w) + log_2 |V|$$

$$I(\hat{s}) = \sum_i H[(Vx_w)_i] - H(x_w)$$

Determinant of the rotation matrix is 1

$$I(\hat{s}) = \sum_i H[(Vx_w)_i] - H(x_w)$$

Independent of V

Finding rotation matrix that minimises the sum of the marginal entropies of $\hat{s}$

$$V = arg\ min \sum_i H(Vx_w)_i$$

Maximize the statistical independence of $\hat{s}$

Many ICA strategy focus on approximating this

⛸ Summary

1. Subtract off the mean of the data in each dimension.
2. Whiten the data by calculating the eigenvectors of the covariance of the data.
3. Identify final rotation matrix that optimises statistical independence

Why Non-Gaussian?

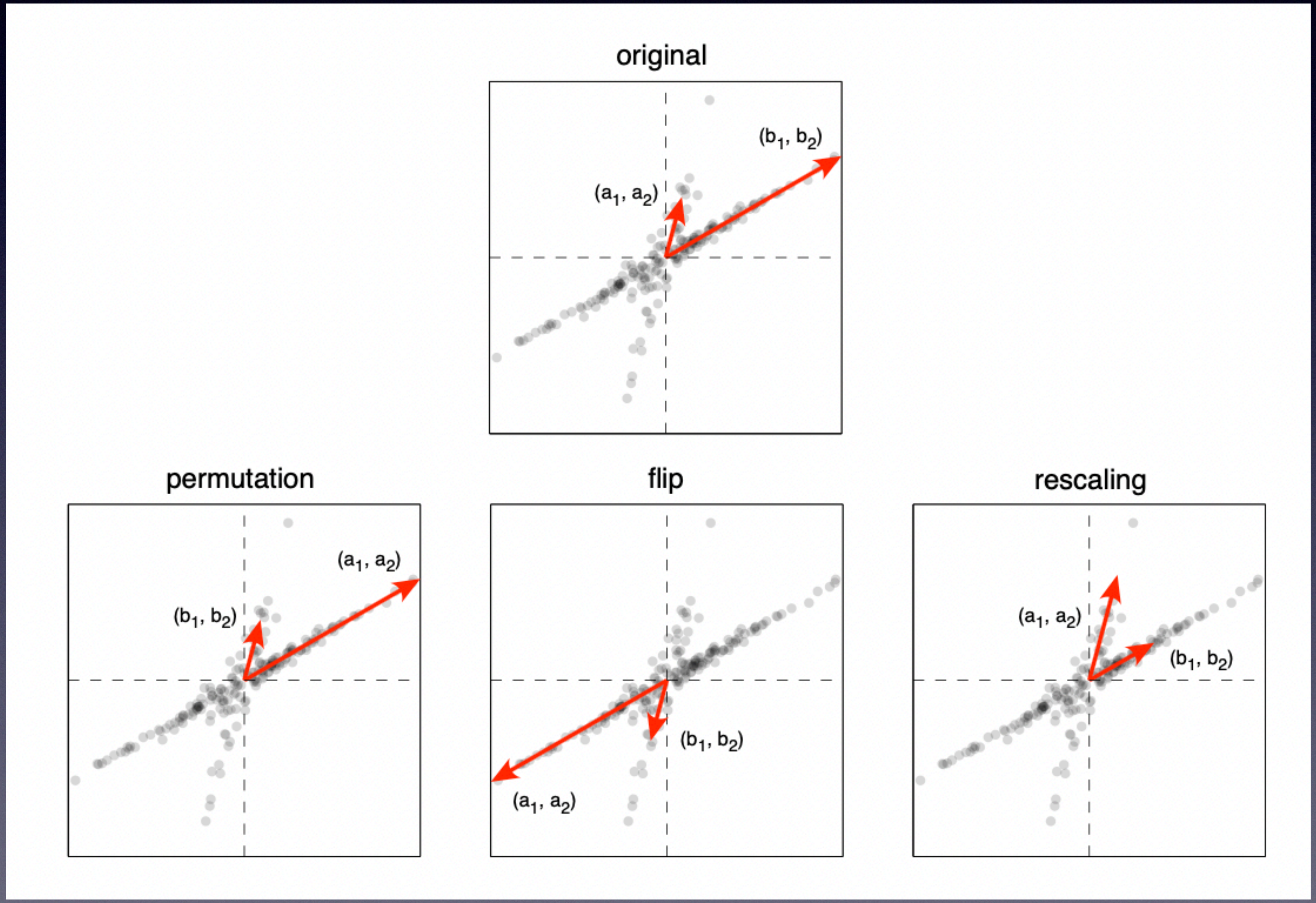
Let $x = As$ where $s \approx N(0,1)$, symmetric and rotation	Now assume some arbitrary rotation R, $RR^T = I$
$Cov(x) = E[(x-0)(x-0)^T]$ $= E[(As)(As)^T]$ $= E[Ass^TA^T]$ $= AE[ss^T]A^T$ $= ACov(s)A^T$ $= AA^T$	$Cov(x) = E[(x-0)(x-0)^T]$ $= E[(ARs)(ARs)^T]$ $= E[ARss^TR^TA^T]$ $= ARE[ss^T]R^TA^T$ $= ARR^TA^T$ $= AA^T$

Ambiguities in Solution

1. Optimisation of V

2. Each a degree of freedom that provides an additional solution

8



ICA vs PCA

ICA provides a fairly unrestrictive linear basis for representing a data set

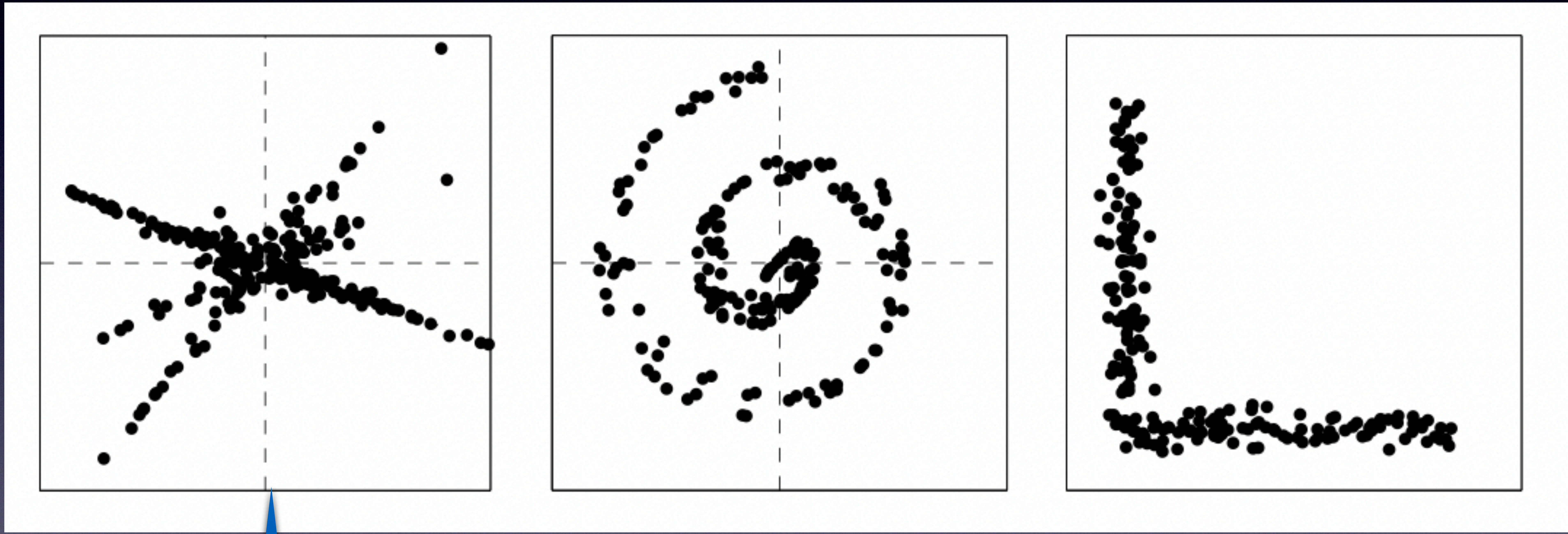
PCA identifies an orthogonal linear basis which maximises the variance of the data.

ICA is not restricted to an orthogonal basis because statistical independence makes no such requirement

PCA can identify the underlying sources only when data is distributed appropriately, such as a Gaussian distribution

Challenges

9



number of independent sources exceed the number of measurements

Nonlinear manifold

Occlusion

⌘ Extensions

- Over complete problem
- Non linearity!

Python Implementation

FastICA Algorithm

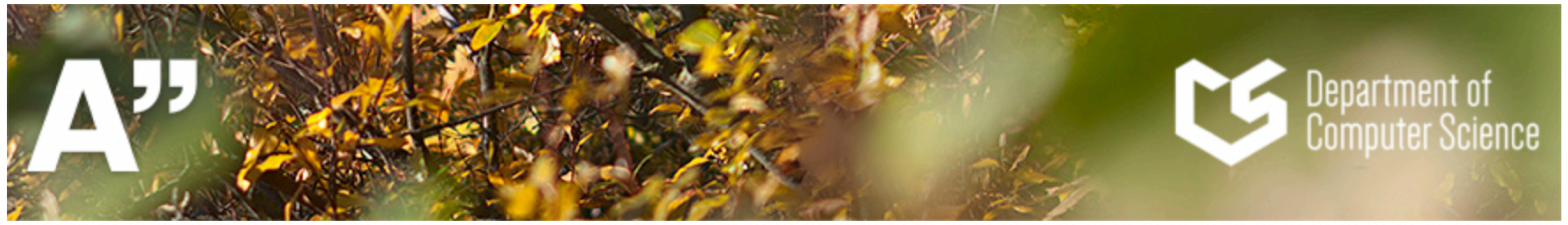
Invented by **Aapo Hyvärinen**
at Helsinki University of Technology

https://research.ics.aalto.fi/ica/software.shtml

sklearn.decomposition.FastICA

class sklearn.decomposition.FastICA(*n_components=None*, ***, *algorithm='parallel'*, *whiten=True*, *fun='logcosh'*, *fun_args=None*, *max_iter=200*, *tol=0.0001*, *w_init=None*, *random_state=None*) [source]

https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html#sklearn.decomposition.FastICA.fit



Independent Component Analysis (ICA) and Blind Source Separation (BSS)

Home | Research | Members | Demos | Software | Links

The FastICA package for MATLAB

The FastICA package is a free (GPL) MATLAB program that implements the fast fixed-point algorithm for independent component analysis and projection pursuit. It features an easy-to-use graphical user interface, and a computationally powerful algorithm.

Download software package

Related sofware

ICASSO: analysing and visualising the reliability of independent components

ISCTEST: principled statistical testing of independent components

FastICA for other environments

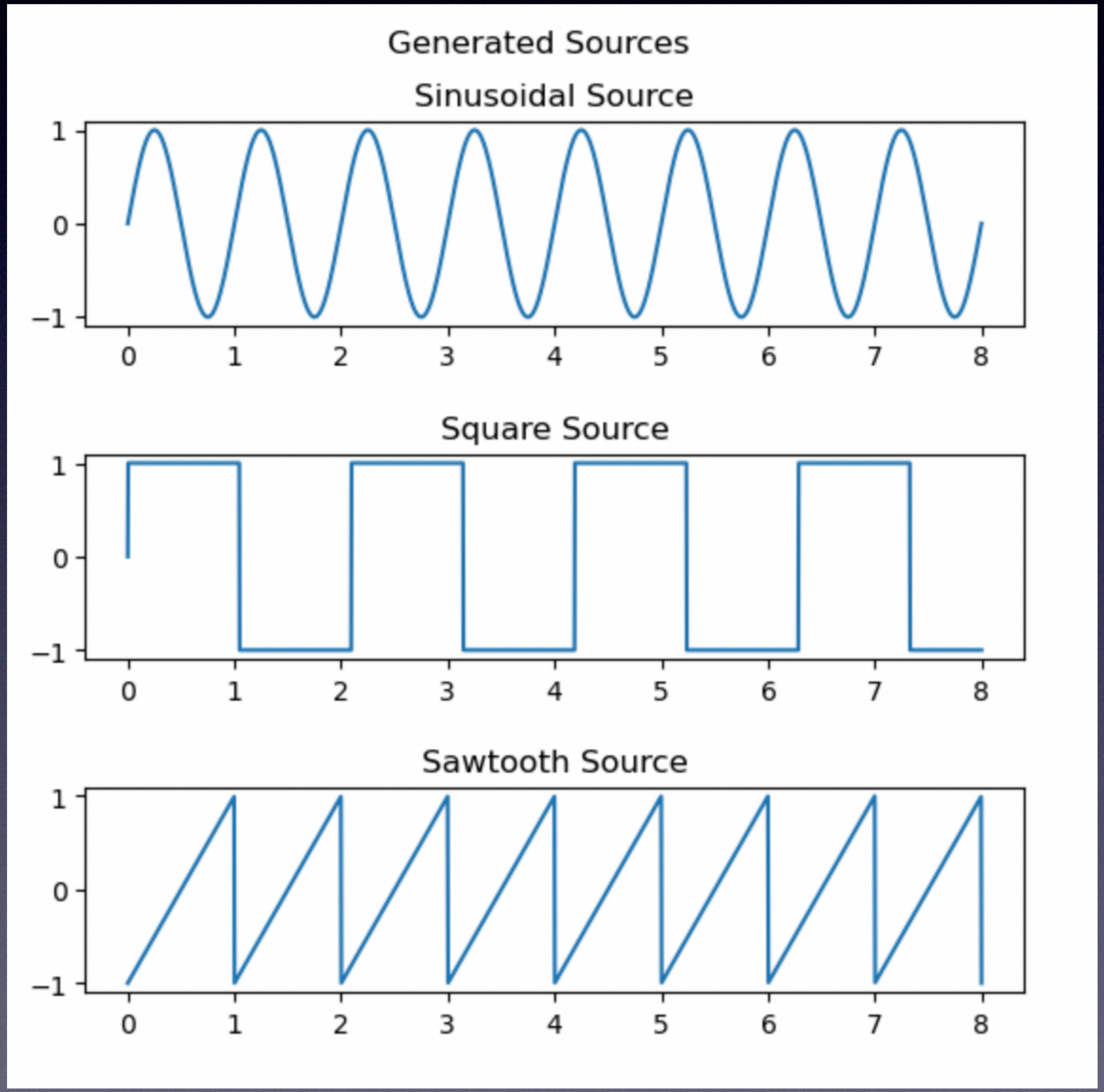
FastICA in R

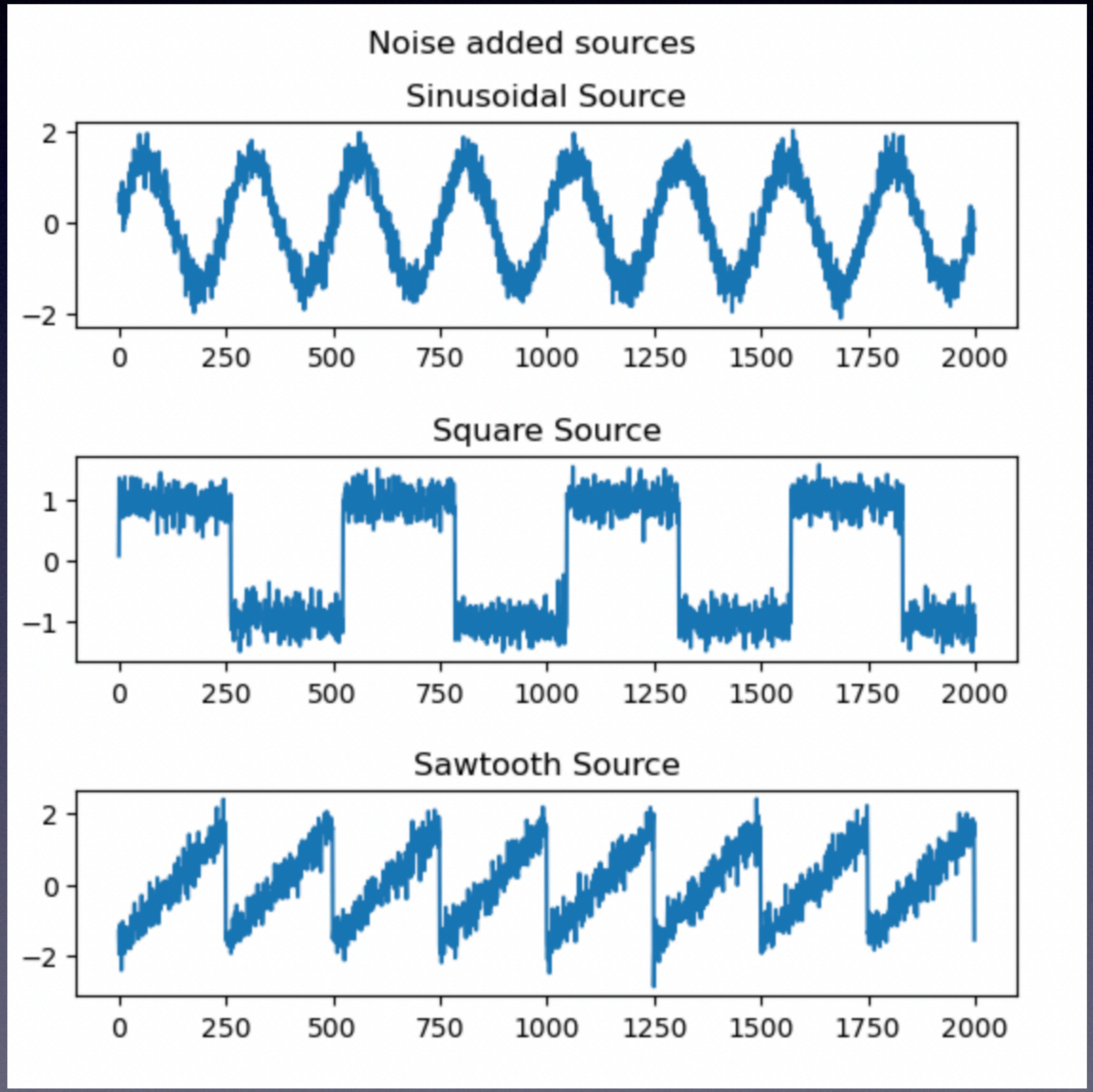
FastICA in C++ (part of IT++ package)

FastICA in Python as part of MDP package

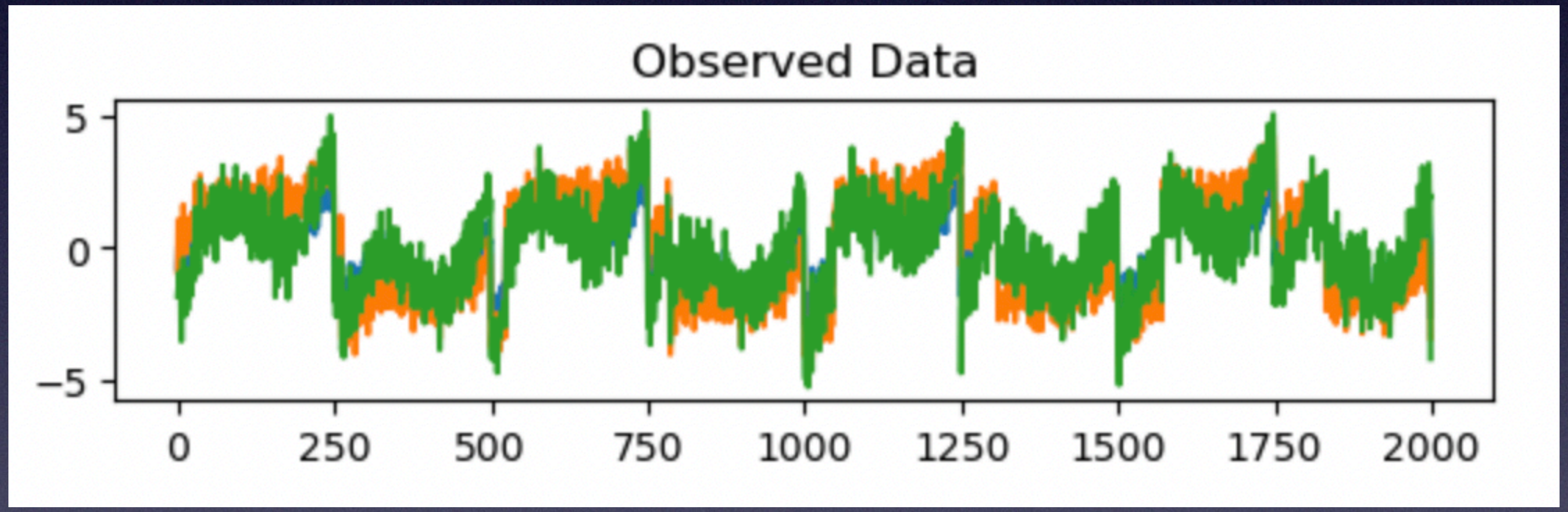
FastICA in Python as part of scikit-learn package

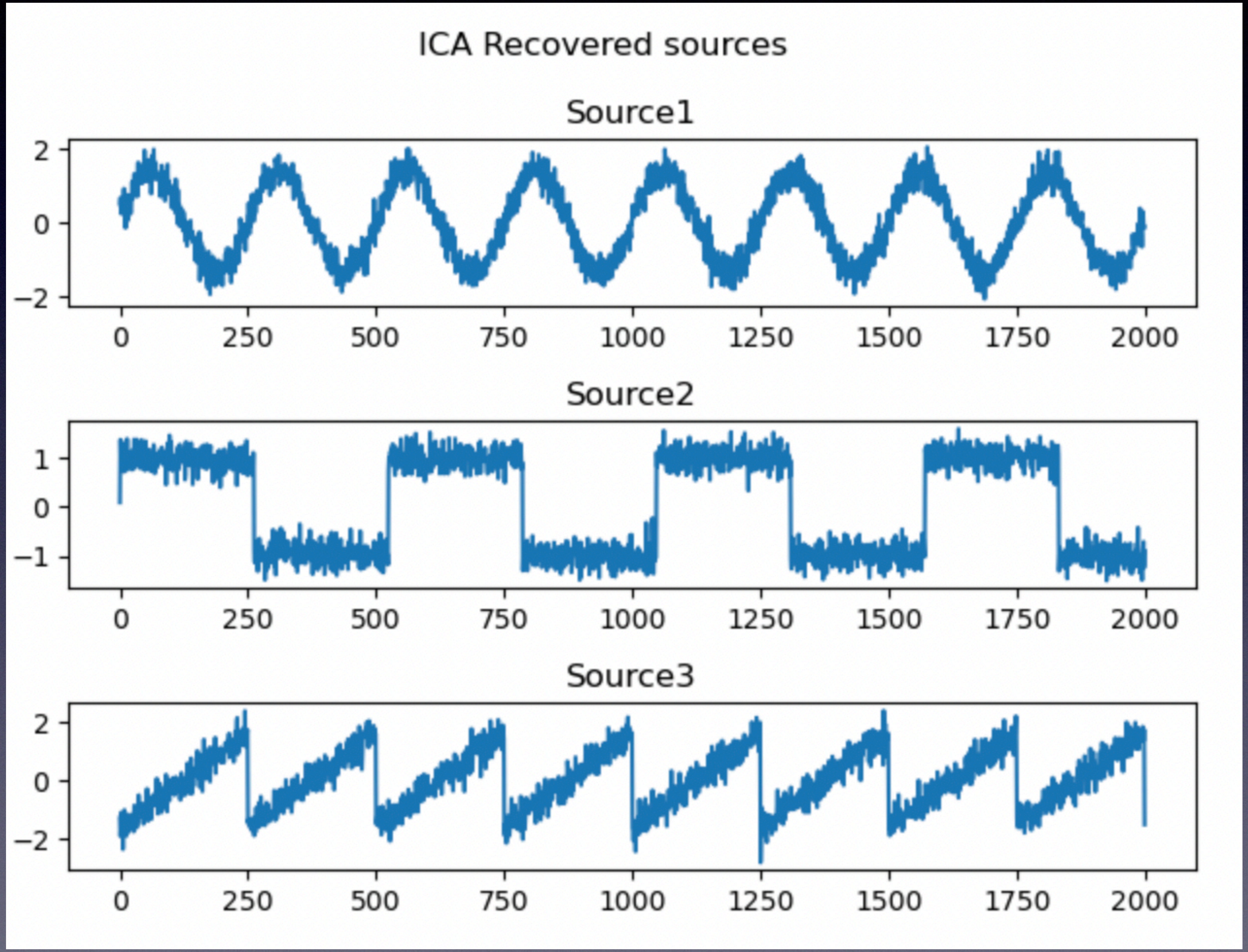
1. Basic signals



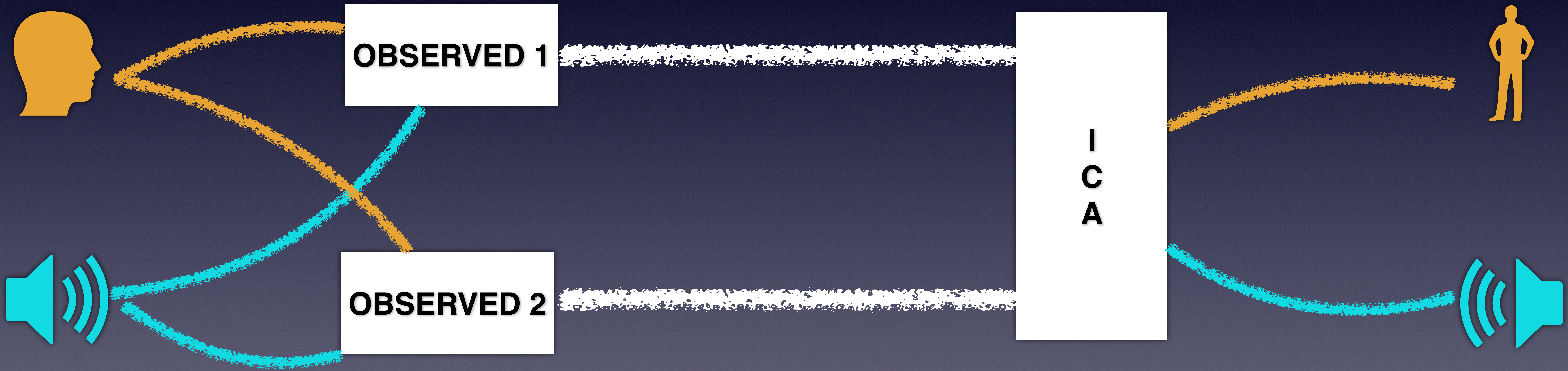


Gitlink: https://github.com/Rajesh-Smartino/Independent-Component-Analysis





2. Audios



References:

1. [1, 2, 3, 4, 5, 6, 7, 8, 9] A Tutorial on Independent Component Analysis **Jonathon Shlens** Google Research Mountain View, CA 94043
2. "*Independent Component Analysis" book by* **Aapo Hyvarinen**, Juha Karhunen, Erkki Oja.

Video credits:

1. [#2] Eigenvectors and eigenvalues | Chapter 14, Essence of linear algebra by **3Blue1Brown**. https://youtu.be/PFDu9oVAE-g
2. [#1] Singular Value Decomposition (SVD) and Image Compression by **Serrano.Academy.** https://youtu.be/DG7YTlGnCEo

Thank you