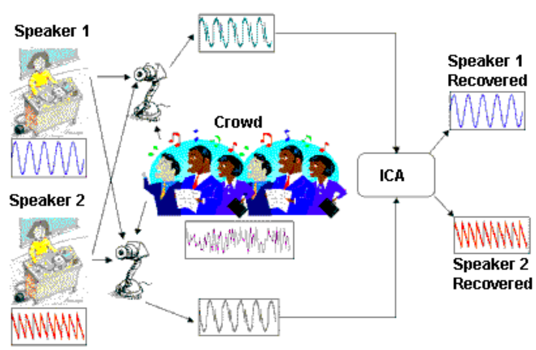
# Independent component analysis

111024503 王羿勳
111024509 陳冠霖
111024519 宇晉賢
108081132 江昀晏

2023/6/13

## 1 Introduction

### 1.1 Cocktail Party Problem




### 1.2 Assumption

- The source signals are independent of each other.
- The values in each source signal have non-Gaussian distributions.

### 1.3 Definition

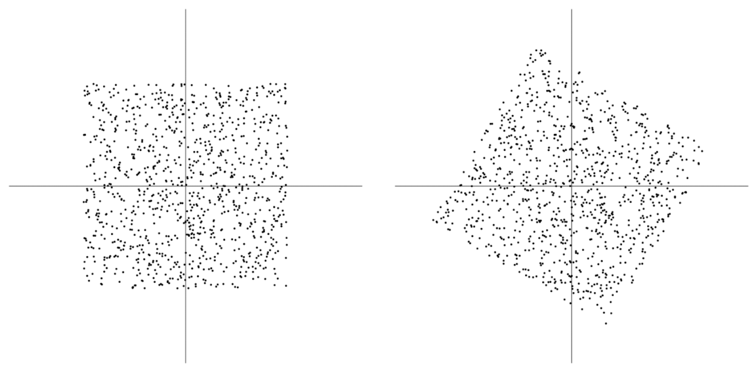
- d speakers $(s)$, d microphones $(x)$.
- Source $(s)$ : $s^{(i)}$ is a d-dimensional vector, $s_j^{(i)}$ is the sound that speaker $j$ was uttering at time $i$.
- Observation $(x)$ : $x^{(i)}$ is a d-dimensional vector, $x_j^{(i)}$ is the acoustic reading recorded by microphone $j$ at time $i$.

## 1.4 Definition (cont.)

- Mixing matrix $(A)$ : Satisfies $x = As$, where $A$ is an unknown square matrix.
- Unmixing matrix $(W)$ : $W = A^{-1} \Rightarrow s = Wx$.
- Estimates of unmixing matrix $(\hat{W})$ : $y = \hat{W}x$, and $y = s$ if all $y_i$ are independent.

# 2 Assumption 1 Independent sources of signal

## 2.1 Assumption 1: Independent sources of signal



## 2.2 Assumption 1: Independent sources of sound

**Principle1 : Nonlinear decorrelation**

Find the matrix $W$ so that for any $i \neq j$, the components $y_i$ and $y_j$ are uncorrelated, and the transformed components $g(y_i)$ and $h(y_i)$ are uncorrelated, where $g$ and $h$ are some suitable nonlinear functions.

# 3 Assumption 2 Non-Gaussion

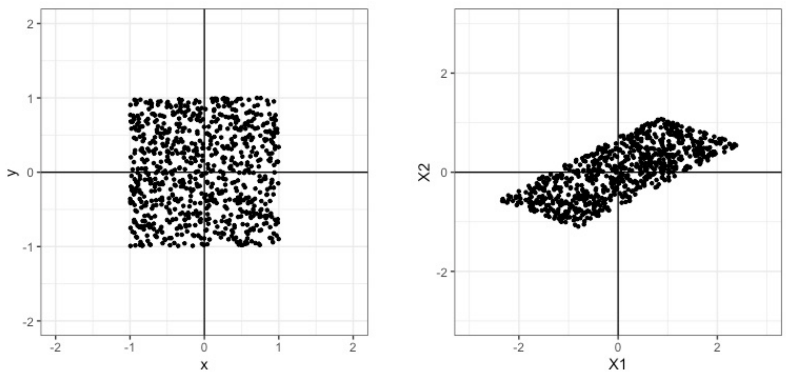
## 3.1 Independent components are the maximally nongaussian components

- The idea is that according to the central limit theorem, sums of nongaussian random variables are closer to gaussian that the original ones. Therefore, if we take a linear combination $y = \sum_i b_i x_i$ of the observed mixture variables (which, because of the linear mixing model, is a linear combination of the independent components as well),this will be maximally nongaussian if it equals one of the independent components.

**Principle2 : Maximum nongaussianity**

Find the local maxima of nongaussianity of a linear combination $y = \sum_i b_i x_i$ under the constraint that the variance of $y$ is constant. Each local maximum gives one independent component.

### 3.2 Non-Gaussian




### 3.3 Gaussian

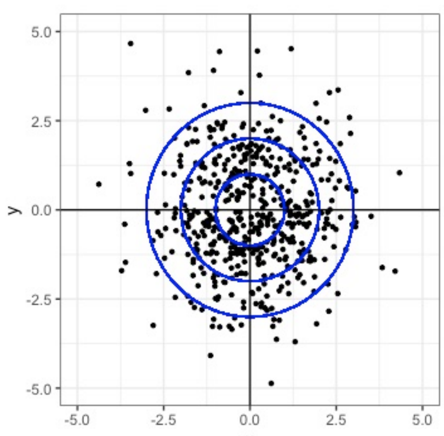



Figure 1: observe signal with Gaussian distribution

If we choose a Gaussian distribution for ICA, which is a rotational symmetry distribution, we may not find a unique solution for the decomposition of ICA.

### 3.4 The principle when finding independent component

- The idea is that according to the central limit theorem, sums of nongaussian random variables are closer to gaussian that the original ones. Therefore, if we take a linear combination $y = \sum_i b_i x_i$ of the observed mixture variables (which, because of the linear mixing model, is a linear combination of the independent components as well),this will be maximally nongaussian if it equals one of the independent components.

**Principle3 : Maximum nongaussianity**

Find the local maxima of nongaussianity of a linear combination $y = \sum_i b_i x_i$ under the constraint that the variance of $y$ is constant. Each local maximum gives one independent component.

### 3.5 Check ICA feasibility under assumption

**Darmois–Skitovich theorem**

let $S_j$ , $j = 1, 2, ..., n$ , $n \geq 2$ be independent random variables.
Let $\alpha_j$,$\beta_j$ be nonzero constants($\alpha_j \beta_j \neq 0$).
$if \mathrm{X}_1 = \alpha_1 S_1 + \cdots + \alpha_n S_n$ and $X_2 = \beta_1 S_1 + \cdots + \beta_n S_n$ are independent then all random variables $S_j$ have normal distributions.

By this Theorem we can conclude that if $S_j$ are non-gaussian and independent to each others,then $\alpha_j \beta_j = 0$
With this fact,we can find the unique $S_j$

### 3.6 Example

- Assume n=2,and $S$ is non gaussian

$$X_1 = \alpha_1 S_1 + \alpha_2 S_2$$
$$X_2 = \beta_1 S_1 + \beta_2 S_2$$

then we let $Z_1 = \boldsymbol{w}_1^T X_1$ , $Z_2 = \boldsymbol{w}_2^T X_2$
after the transform W

$$Z_1 = \alpha_1' S_1 + \alpha_2' S_2$$
$$Z_2 = \beta_1' S_1 + \beta_2' S_2$$

- By Darmois–Skitovich theorem,if we know $S$ is non gaussian and $Z_1$,$Z_2$ are independent,then $\alpha'_1\beta'_1 = \alpha'_2\beta'_2 = 0$.

$$Z_1 = \alpha'_1 S_1 + 0S_2$$
$$Z_2 = 0S_1 + \beta'_2 S_2$$

$$Z_1 = 0S_1 + \alpha'_2 S_2$$
$$Z_2 = \beta'_1 S_1 + 0S_2$$

# 4 ICA algorithm

## 4.1 Joint distribution

- We suppose that the distribution of each source $s_j$ is given by a density $p_s$, and that the joint distribution of the sources s is given by

$$p(s) = \prod_{j=1}^{d} p_s(s_j)$$

- By definition as previous $s = Wx$
  then we can rewrite the joint density as below

$$p(x) = \prod_{j=1}^{d} p_s(w_j^T x) \cdot |W|$$

## 4.2 Log likelihood

- Follow our previous discussion we can not choose Gaussian CDF to construct the likelihood function. we'll choose a reasonable CDF that slowly increases from 0 to 1,that is sigmoid function.

$$g(s) = \frac{1}{1 + e^{-s}}$$
$$\text{Hence, } p_s(s) = g'(s)$$

- The square matrix $W$ is the parameter in our model. Given a training set $\{x^{(i)}; i = 1, ..., n\}$,then we can get log likelihood.

$$l(W) = \sum_{i=1}^{n} (\sum_{j=1}^{d} \log g'(w_j^T x^{(i)}) + \log |W|)$$

### 4.3 Stochastic gradient ascent

- We would like to maximize log likelihood function in terms of $W$ ,then by taking derivatives and using the fact $\nabla_W|W| = |W|(W^T)^{-1}$ ,we derive a stochastic gradient ascent learning rule.

$$W := W + \alpha \left( \begin{bmatrix} 1 - 2g(w_1^T x^{(i)}) \\ 1 - 2g(w_2^T x^{(i)}) \\ . \\ . \\ . \\ 1 - 2g(w_d^T x^{(i)}) \end{bmatrix} x^{(i)T} + (W^T)^{-1} \right)$$

After the algorithm converges,
we then compute $s^{(i)} = Wx^{(i)}$ to recover the original sources.

## 5 FastICA

### 5.1 FastICA

- This algorithm uses the approximate neg-entropy to measure the nongaussianity.
To measure $y = w^T z$, the approximate neg-entropy is
$J(w) \propto [E\{G(w^T z)\} - E\{G(v)]^2$, where $v$ is the gaussion variable.

- By the Lagrange multiplier, under the constraint $E\{(w^T z)^2\} = \|w\|^2 = 1$, the maximum of $w$ must satisfies
$E\{zg(w_p^T z)\} + \beta w = 0$, where $g$ is the derivative of $G$.

- By Newton's method and some approximation methods, we have the iteration shown below :

$$\begin{aligned} w_p &\leftarrow E\{zg(w_p^T z)\} - E\{g^{'}(w_p^T z)\}w_p \\ w_p &\leftarrow \frac{w_p}{\|w_p\|} \end{aligned}$$

### 5.2 Choice of G

- $G_1(u) = \frac{1}{a} log\ cosh(au)$, $1 \leq a \leq 2$

- $G_2(u) = -exp(-\frac{u^2}{2})$

- $G_3(u) = \frac{1}{4}u^4$

### 5.3 FastICA algorithm

**Require:**

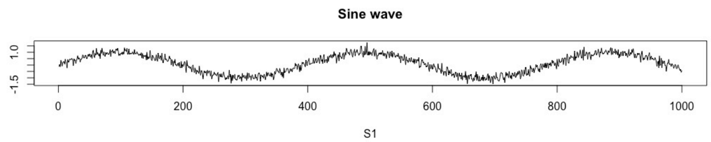
$x \leftarrow$ The observed variables
$s \leftarrow$ The independent components
$n \leftarrow$ The number of the observed variables
$m \leftarrow$ The number of the independent components

**procedure** FastICA$(x, s, n, m)$
  $x \leftarrow x - E(x)$ ▷ Centering
  $z \leftarrow Vx = VAs = \tilde{A}s$ ▷ Whitening
  $p \leftarrow 1$
  $t \leftarrow$ The iteration times
  **while** $p < m$ **do**
    **while** $w_p^{(t)} \neq w_p^{(t-1)}$ **do** ▷ Check convergence
      $w_p \leftarrow$ arbitary $(n \times 1)$vector
      $w_p \leftarrow E\{zg(w_p^T z)\} - E\{g^{`}(w_p^T z)\}w_P$
      $w_p \leftarrow \frac{w_p}{\|w_p\|}$
      $w_p \leftarrow w_p - \sum_{j=1}^{p-1}(w_j w_j^T)w_p$
      $w_p \leftarrow \frac{w_p}{\|w_p\|}$
    **end while**
    $p \leftarrow p + 1$
  **end while**
**end procedure**

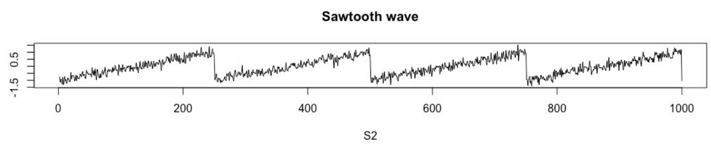
# 6 Simulation

## 6.1 Example 1: Blind Source Separation

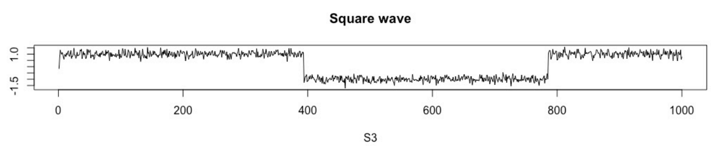
- Sine Wave : $sin\ (2t)$




- Sawtooth Wave : $sawtooth\ (f = 1, width = 1)$




- Square Wave : $sgn\ (sin\ (t))$




- Mixing Matrix :

$$A = \begin{bmatrix} 2, & 0.5, & 1 \\ 0.25, & 8, & 0.5 \\ 1.5, & 0.33, & 2 \end{bmatrix}$$

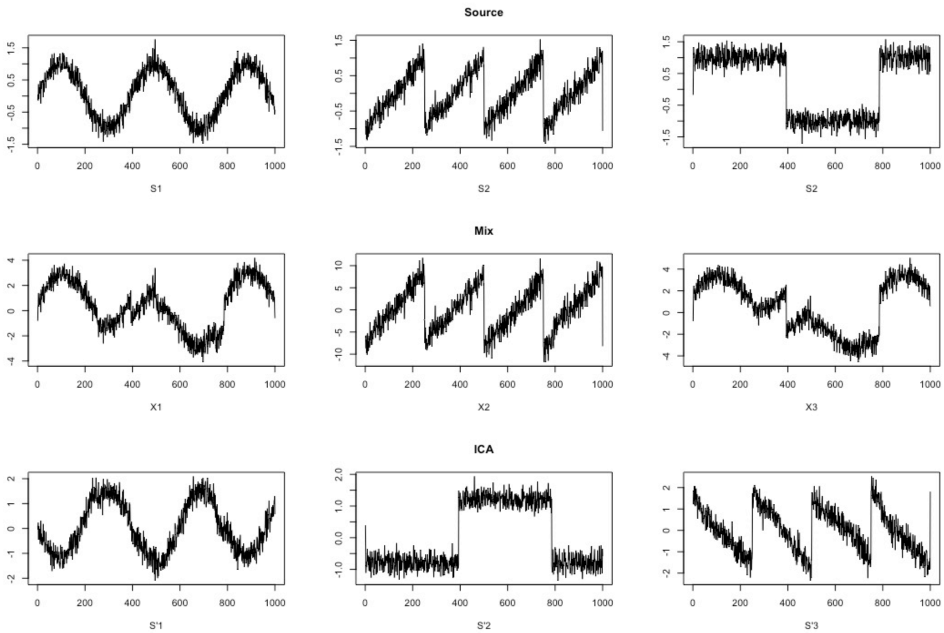
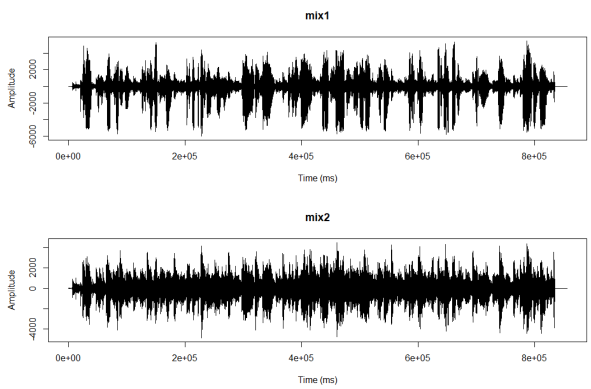
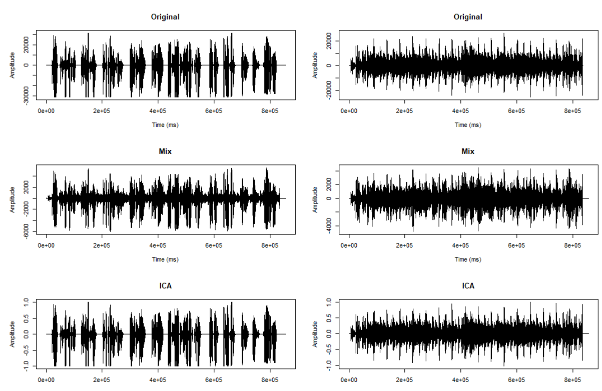



Figure 2: Blind Source Separation

Figure 3: Example 1 shiny.app

### 6.2 Example 2: Melodic Demo




# 7 Reference

1. Hyvärinen A, Karhunen J, Oja E. 2001 Independent component analysis. London, UK: Wiley Interscience.

2. Hyvärinen Aapo. 2013 Independent component analysis: recent advances. Phil. Trans. R. Soc. A.

3. 連憶如, " 頻域獨立成分分析法於語音訊號分離之研究", 交大碩士論文, 2003.

4. Stanford CS229: Machine Learning Lecture 16 - Independent Component Analysis & RL by Andrew Ng.