

Advanced Data Analysis: Blind Source Separation

Masashi Sugiyama (Computer Science)

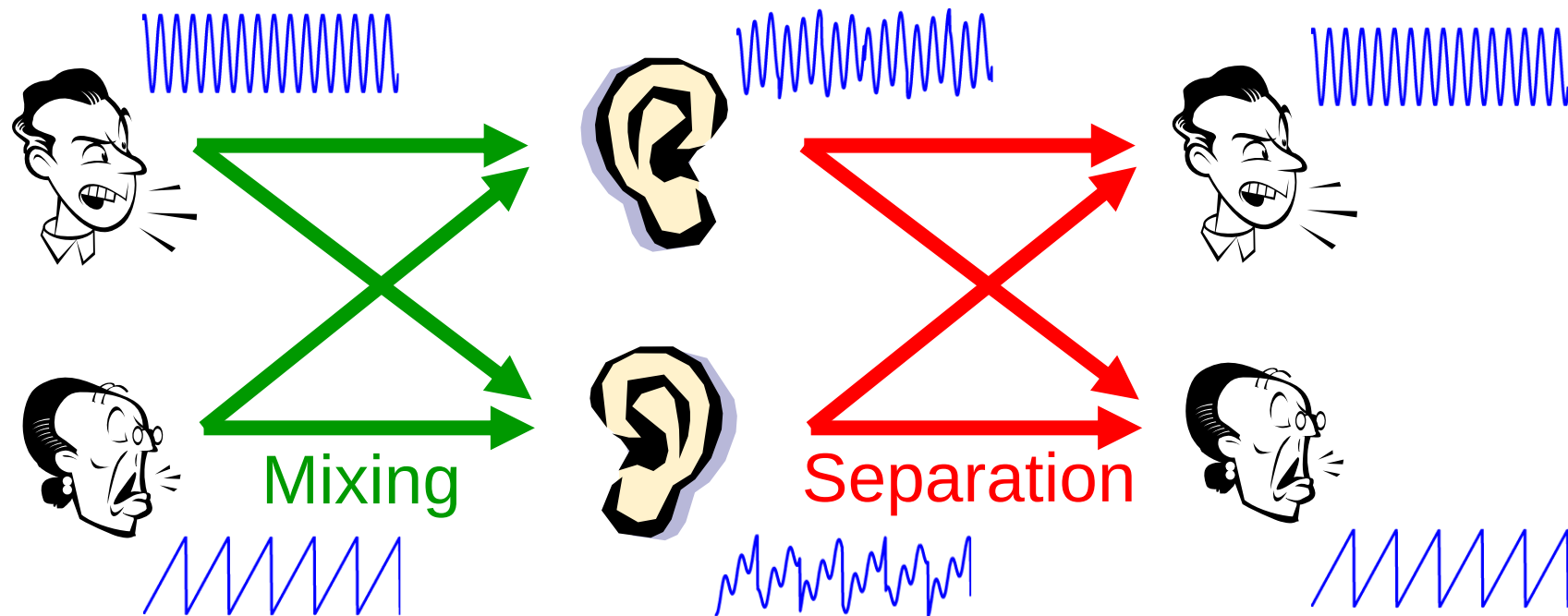
W8E-505, sugi@cs.titech.ac.jp

<http://sugiyama-www.cs.titech.ac.jp/~sugi>

Blind Source Separation

242

■ Cocktail-party problem:



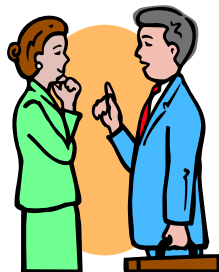
- We want to separate mixed signals into original ones.

Demonstration

243

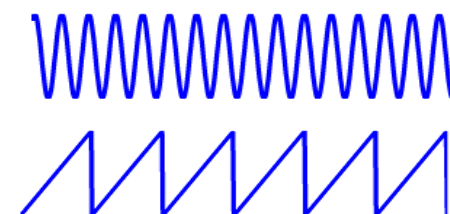
	Mixed signal	Separated signal 1	Separated signal 2
Conversation + Conversation			
Conversation + Instrument			

From <http://www.brain.kyutech.ac.jp/~shiro/research/blindsep.html>



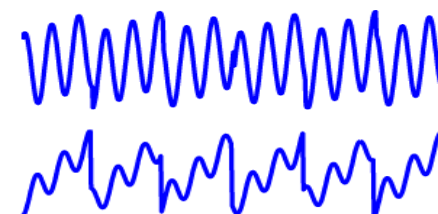
■ Source signals:

- Speaker 1: $s_1^{(1)}, s_2^{(1)}, \dots, s_n^{(1)}$
- Speaker 2: $s_1^{(2)}, s_2^{(2)}, \dots, s_n^{(2)}$



■ Mixed signals:

- Left ear: $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$
- Right ear: $x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}$



$$x_i^{(1)} = m_{11}s_i^{(1)} + m_{12}s_i^{(2)}$$
$$x_i^{(2)} = m_{21}s_i^{(1)} + m_{22}s_i^{(2)}$$

Formulation (cont.)

245

■ In matrix form:

$$\mathbf{x}_i = \mathbf{M} \mathbf{s}_i$$

$$\mathbf{x}_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \end{pmatrix}$$

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \quad \mathbf{s}_i = \begin{pmatrix} s_i^{(1)} \\ s_i^{(2)} \end{pmatrix}$$

■ More generally

- $\mathbf{x}_i, \mathbf{s}_i$: d -dimensional vectors
- $\mathbf{M}$: d -dimensional matrix.

Problem

246

$$\mathbf{x}_i = \mathbf{M} \mathbf{s}_i$$

- We want to estimate $\{\mathbf{s}_i\}_{i=1}^n$ from $\{\mathbf{x}_i\}_{i=1}^n$.
- Approach: Estimate $\mathbf{M}$, and use its inverse for obtaining $\{\hat{\mathbf{s}}_i\}_{i=1}^n$.

$$\hat{\mathbf{s}}_i = \hat{\mathbf{M}}^{-1} \mathbf{x}_i$$

- In BSS, the followings may not be important:
 - **Permutation** of separated signals
 - **Scaling** of separated signals
- Therefore, we estimate $\hat{\mathbf{M}}^{-1}$ up to permutation and scaling of rows.

Assumptions

247

- $\{s_i\}_{i=1}^n$ are i.i.d. random variables with mean zero and covariance identity:

$$\frac{1}{n} \sum_{i=1}^n s_i = 0$$

$$\frac{1}{n} \sum_{i=1}^n s_i s_i^\top = \mathbf{I}_d$$

- $\{s^{(j)}\}_{j=1}^d$ are mutually independent:

$$P(s^{(1)}, s^{(2)}, \dots, s^{(d)}) = P(s^{(1)})P(s^{(2)}) \dots P(s^{(d)})$$

- $\{s^{(j)}\}_{j=1}^d$ are **non-Gaussian**.

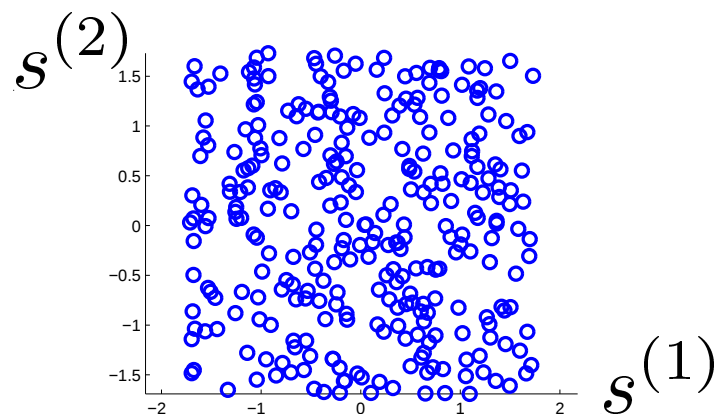
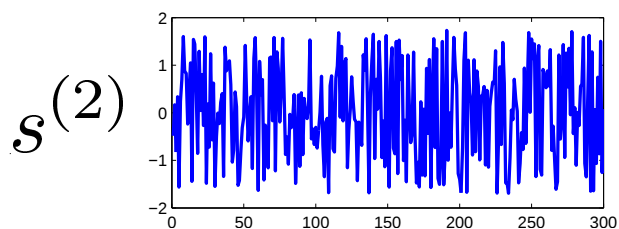
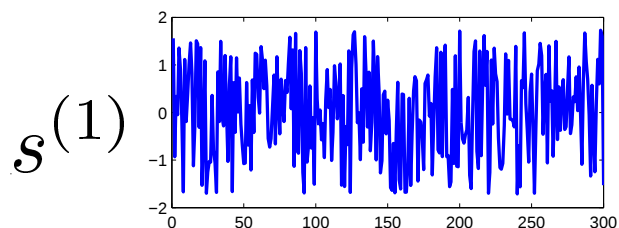
- **M** is invertible.

- BSS under source independence is called **independent component analysis**.

Example

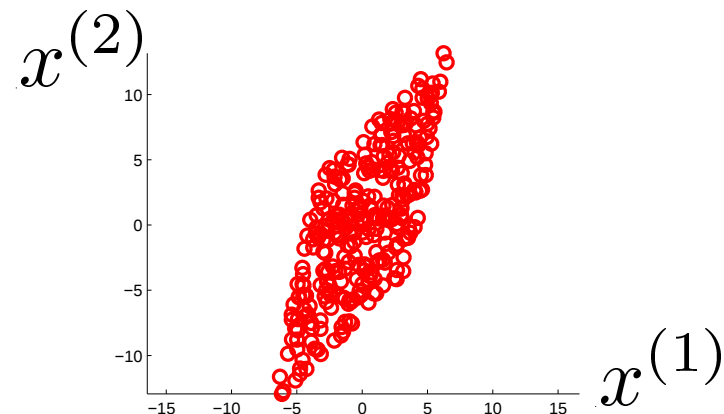
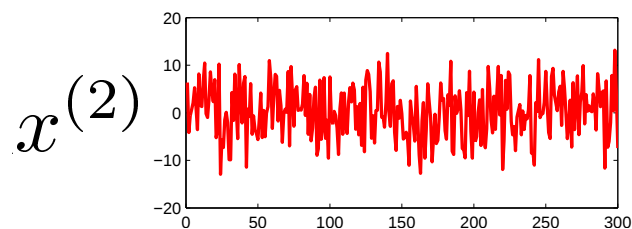
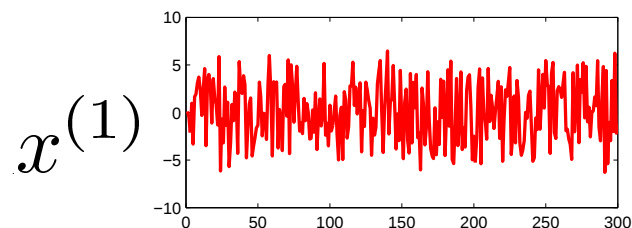
248

Source signals
(uniform)



Mixed signals

$$M = \begin{pmatrix} 1 & 3 \\ 5 & 1 \end{pmatrix}$$



■ Sphering (or pre-whitening):

$$\tilde{\mathbf{x}}_i = \mathbf{C}^{-\frac{1}{2}} \mathbf{x}_i \quad \mathbf{C} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \mathbf{x}_j^\top$$

■ Then

$$\tilde{\mathbf{x}}_i = \widetilde{\mathbf{M}} \mathbf{s}_i \quad \widetilde{\mathbf{M}} = \mathbf{C}^{-\frac{1}{2}} \mathbf{M}$$

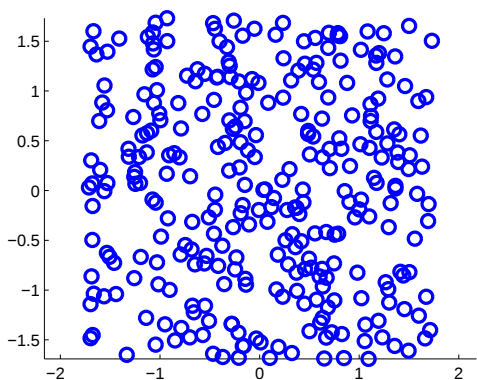
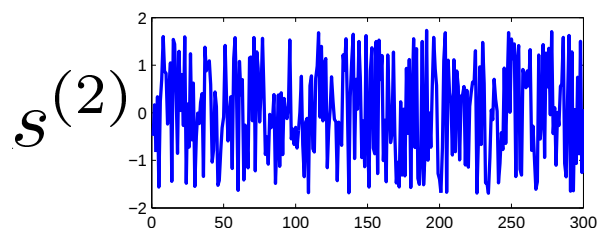
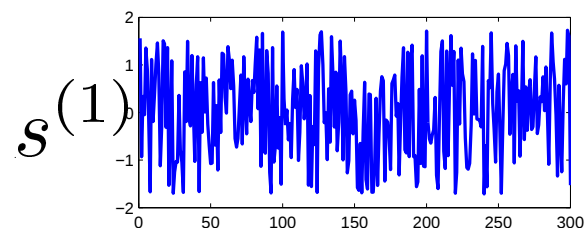
■ Now we want to estimate $\widetilde{\mathbf{M}}$ from $\{\tilde{\mathbf{x}}_i\}_{i=1}^n$, and obtain $\{\hat{\mathbf{s}}_i\}_{i=1}^n$ by

$$\hat{\mathbf{s}}_i = \mathbf{W} \tilde{\mathbf{x}}_i \quad \mathbf{W} = \widetilde{\mathbf{M}}^{-1}$$

Example

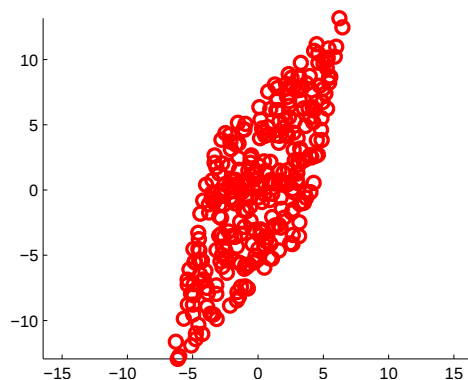
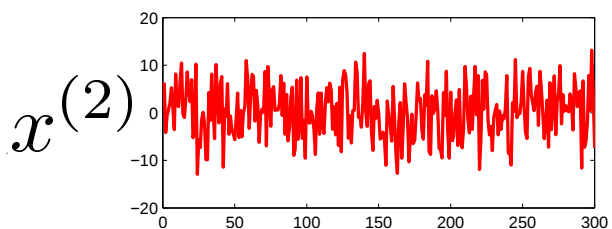
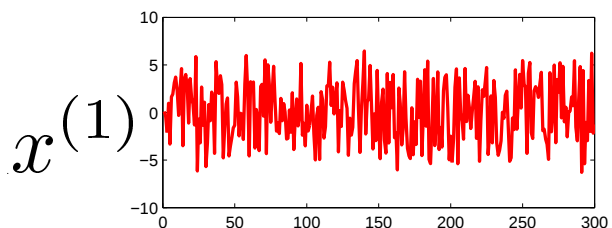
250

Source signals
(uniform)



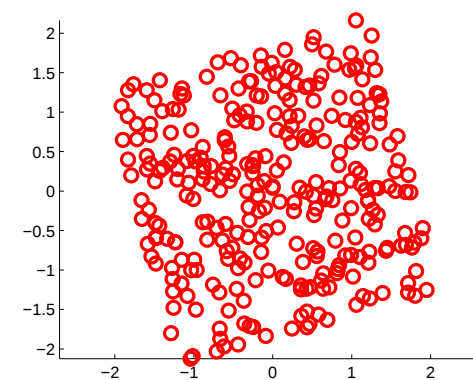
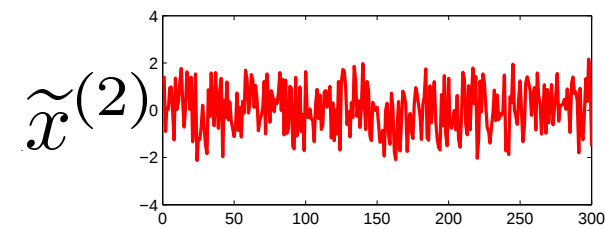
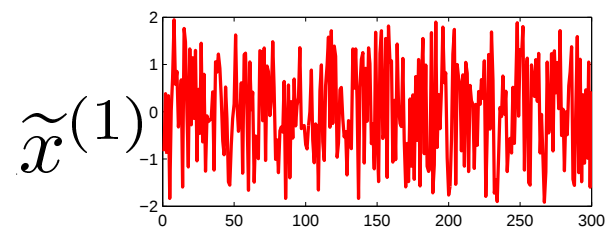
Mixed signals

$$\mathbf{M} = \begin{pmatrix} 1 & 3 \\ 5 & 1 \end{pmatrix}$$



Sphered signals

$$\tilde{x}_i = \mathbf{C}^{-\frac{1}{2}} \mathbf{x}_i$$



Orthogonal Matrix

251

■ $\widetilde{M}$ is an orthogonal matrix since

$$\widetilde{C} = \frac{1}{n} \sum_{i=1}^n \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^\top = \mathbf{I}_d$$

$$\widetilde{C} = \widetilde{M} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{s}_i \mathbf{s}_i^\top \right) \widetilde{M}^\top = \widetilde{M} \widetilde{M}^\top$$

■ Therefore,

$$\widehat{\mathbf{s}}_i = \mathbf{W} \widetilde{\mathbf{x}}$$

$$\mathbf{W} = \widetilde{M}^{-1} = \widetilde{M}^\top \equiv (\mathbf{w}^{(1)} | \mathbf{w}^{(2)} | \dots | \mathbf{w}^{(d)})^\top$$

$\{\mathbf{w}^{(j)}\}_{j=1}^d$: Orthonormal basis

$$\widehat{s}_i^{(j)} = \langle \mathbf{w}^{(j)}, \widetilde{\mathbf{x}}_i \rangle$$

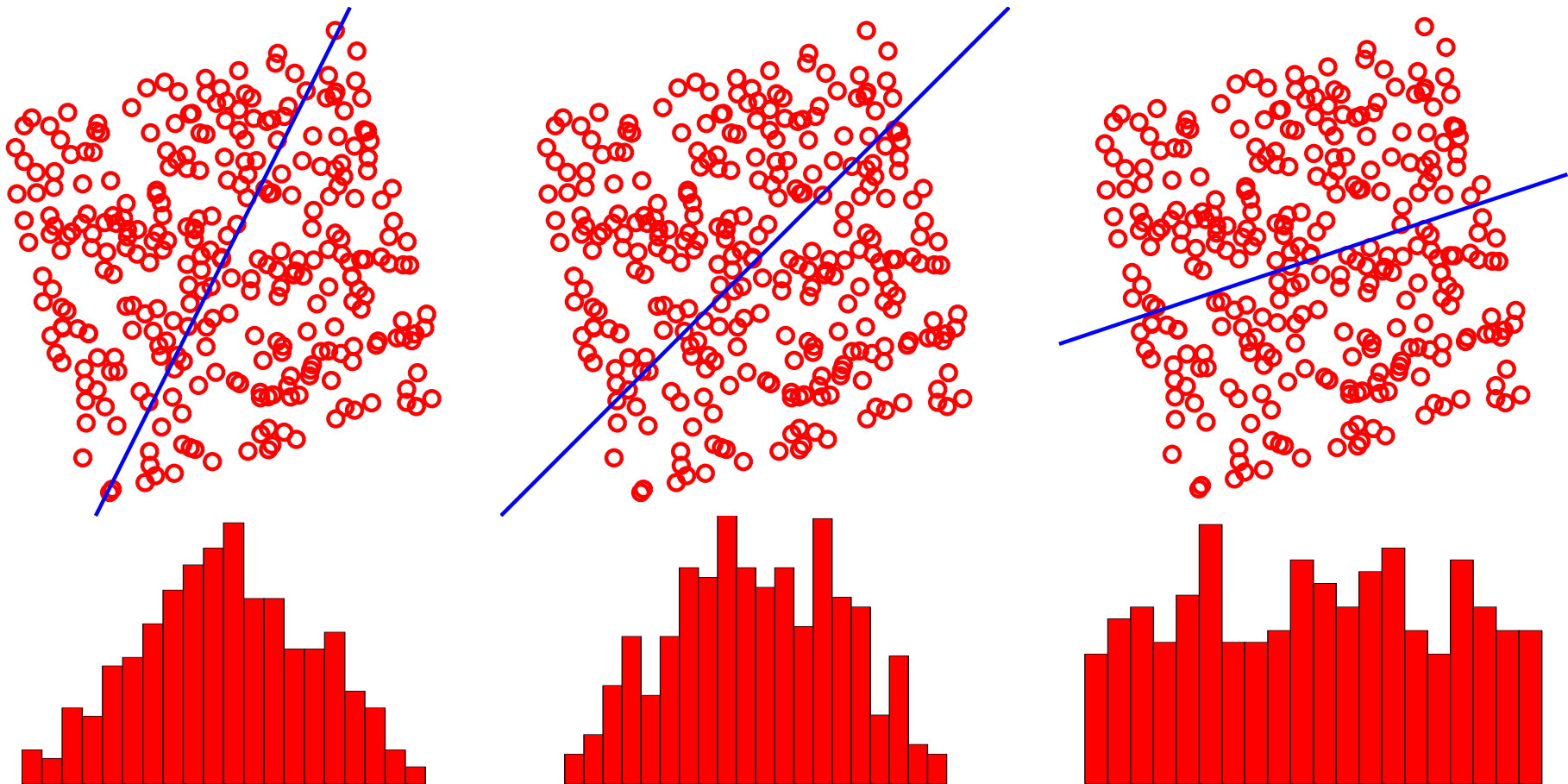
Non-Gaussian Is Independent²⁵²

- Now we want to find an ONB $\{\boldsymbol{w}^{(j)}\}_{j=1}^d$ such that $\{\hat{s}^{(j)}\}_{j=1}^d$ are independent.
- **Central limit theorem**: Sum of independent variables tends to be Gaussian.
- Conversely, **non-Gaussian variables are independent**.
- We find **non-Gaussian directions** in $\{\tilde{\boldsymbol{x}}_i\}_{i=1}^n$.

Example (cont.)

253

- Non-Gaussian direction is independent.



ICA by Projection Pursuit

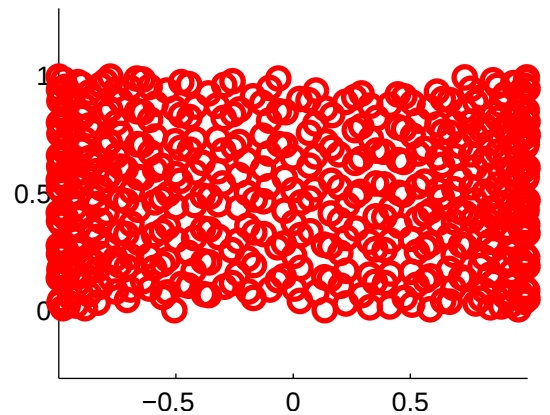
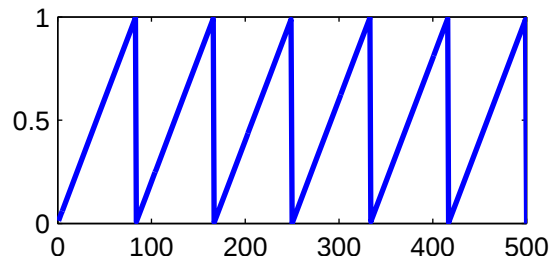
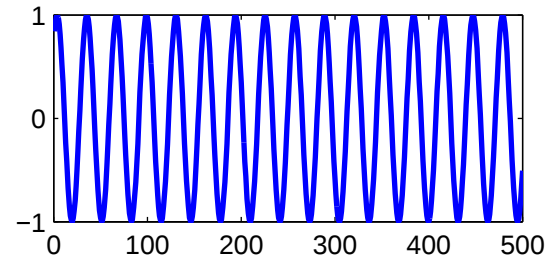
254

- Finding non-Gaussian directions can be achieved by **projection pursuit algorithms**!
 - Center and sphere the data.
 - Find non-Gaussian directions by PP.
- We may use an approximate Newton-based PP method, which is called **FastICA**.

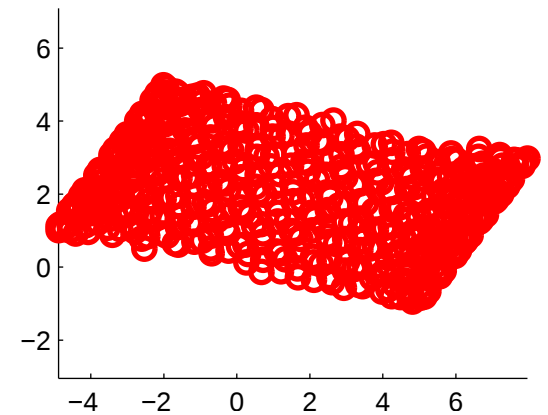
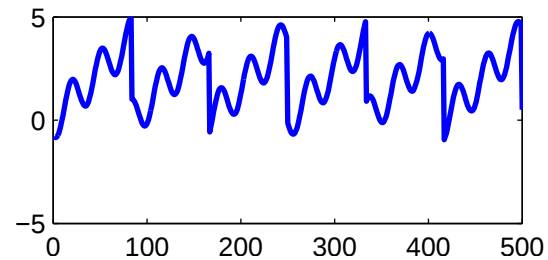
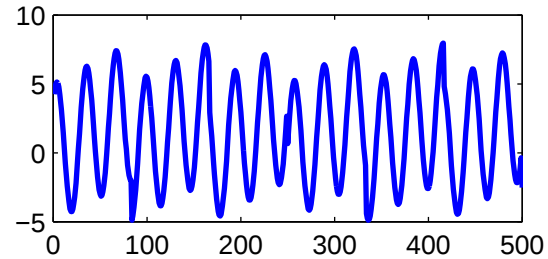
Example 2

255

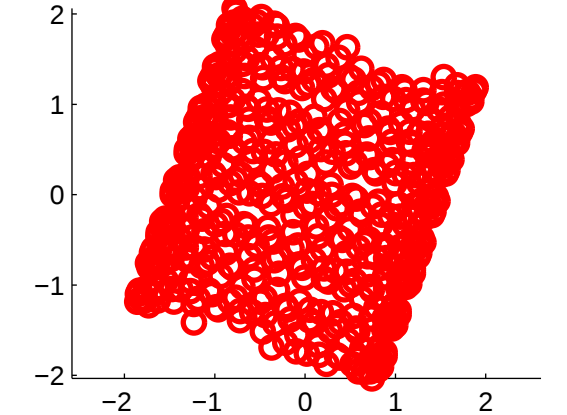
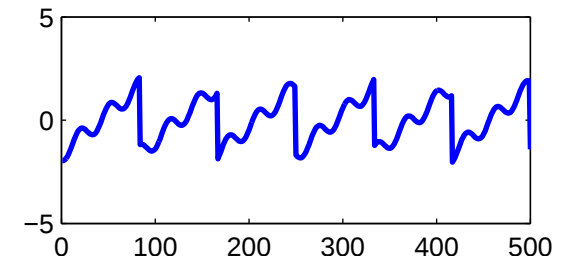
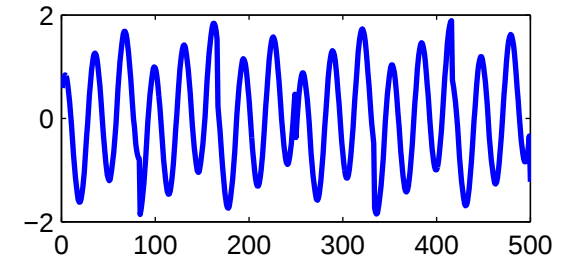
Source



Mixed



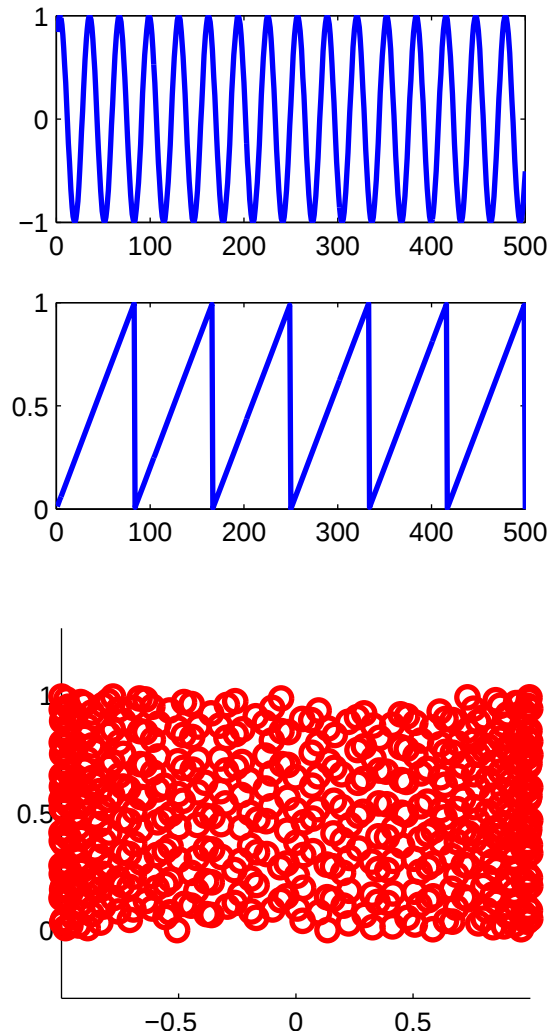
Sphered



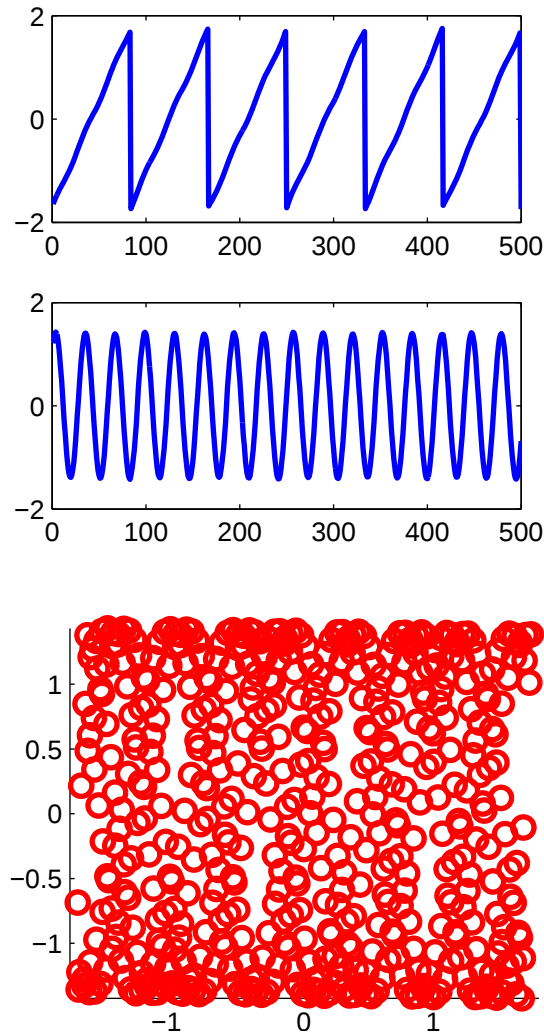
Example 2 (cont.)

256

Source



Separated



Original signals are recovered up to permutation and scaling.