Introduction to Blind Source Separation and Non-negative Matrix Factorization

Tatsuya Yokota

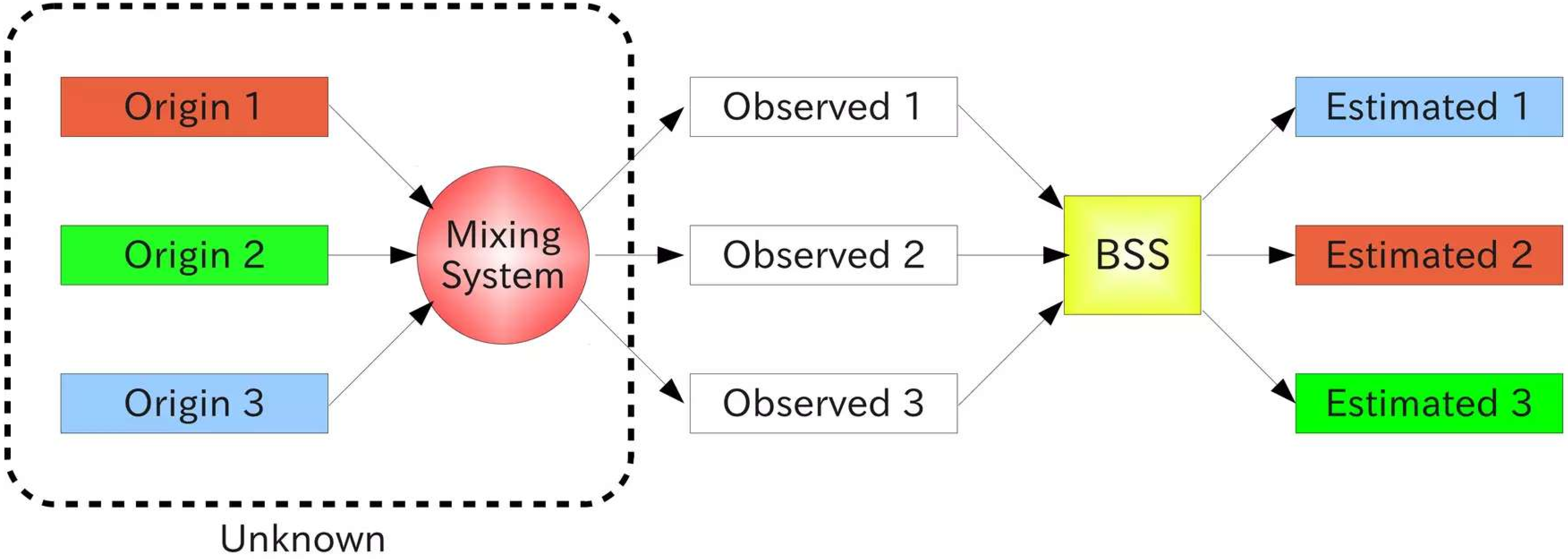
Tokyo Institute of Technology

November 29, 2011

Outline

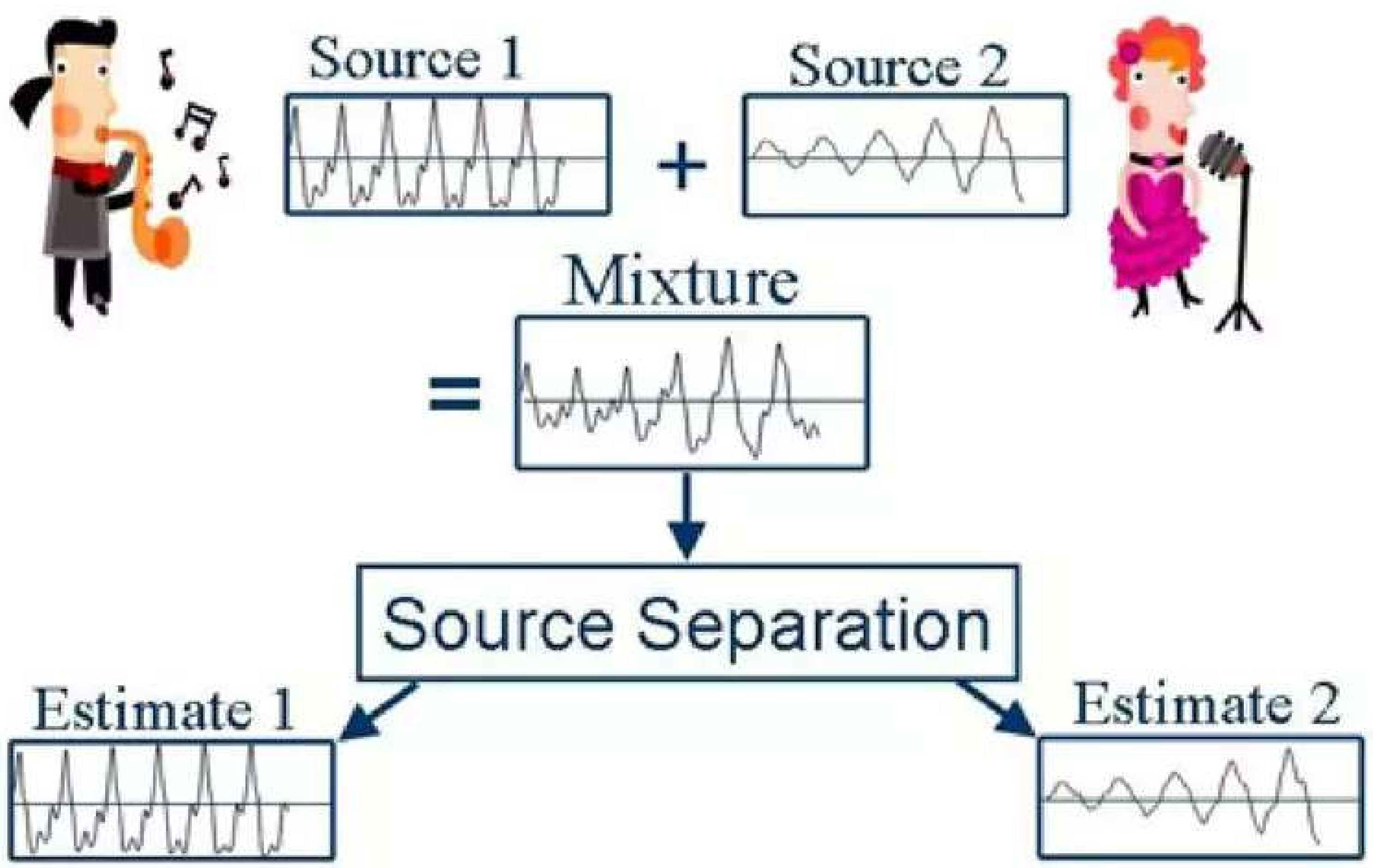
What's a Blind Source Separation

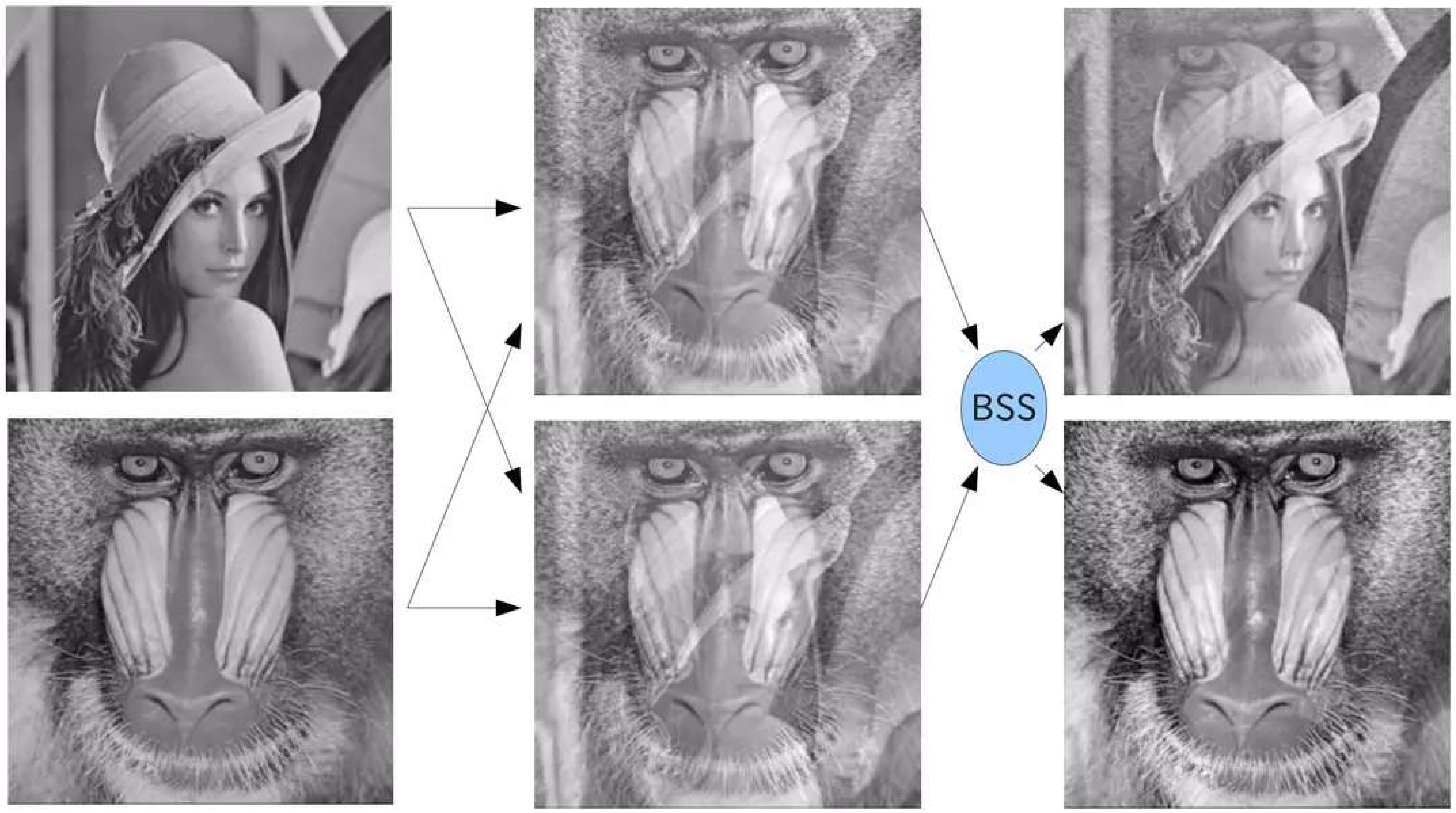
Blind Source Separation is a method to estimate original signals from observed signals which consist of mixed original signals and noise.



Example of BSS

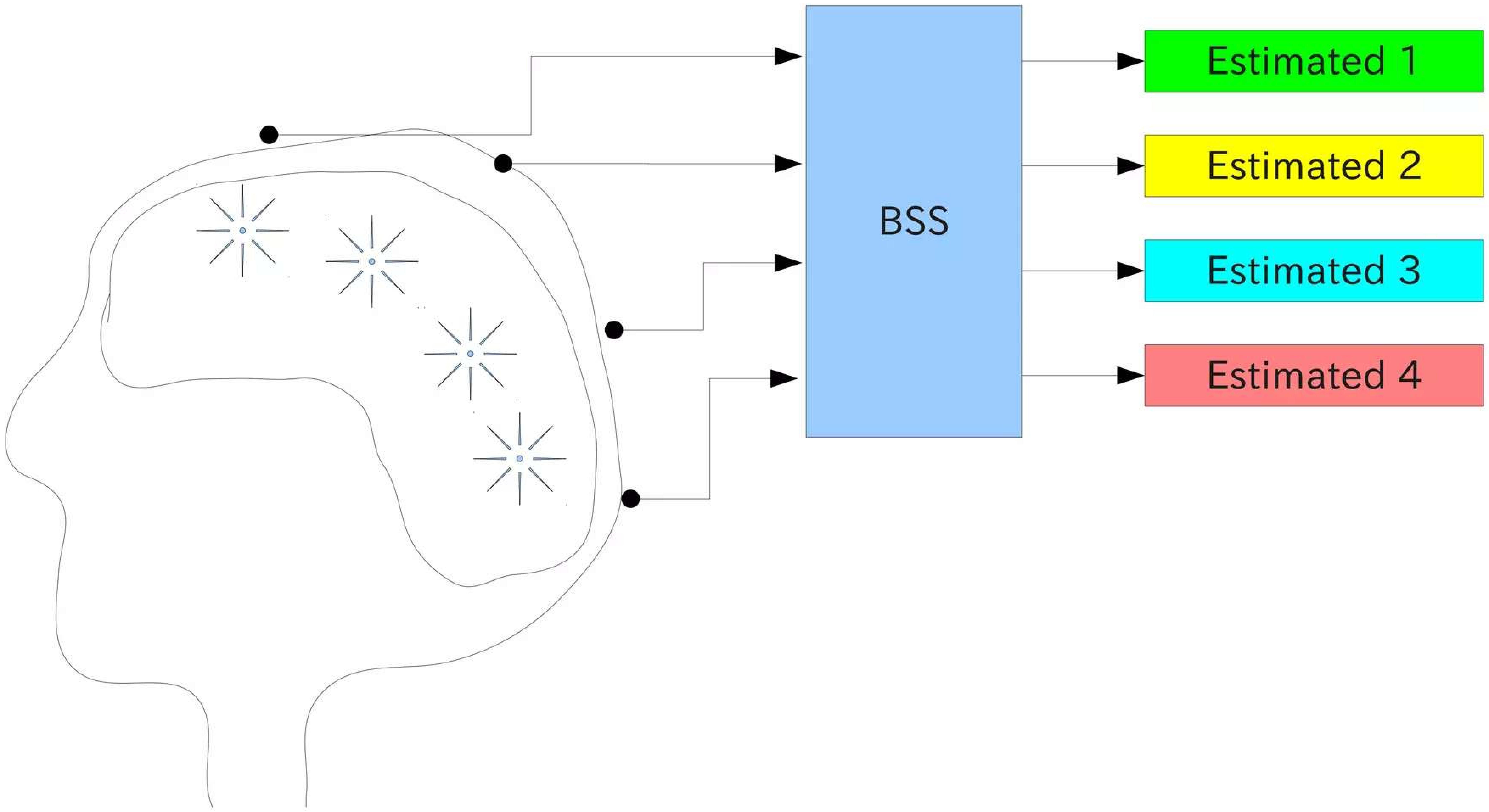
BSS is often used for Speech analysis and Image analysis.





Example of BSS (cont'd)

BSS is also very important for brain signal analysis.



Model Formalization

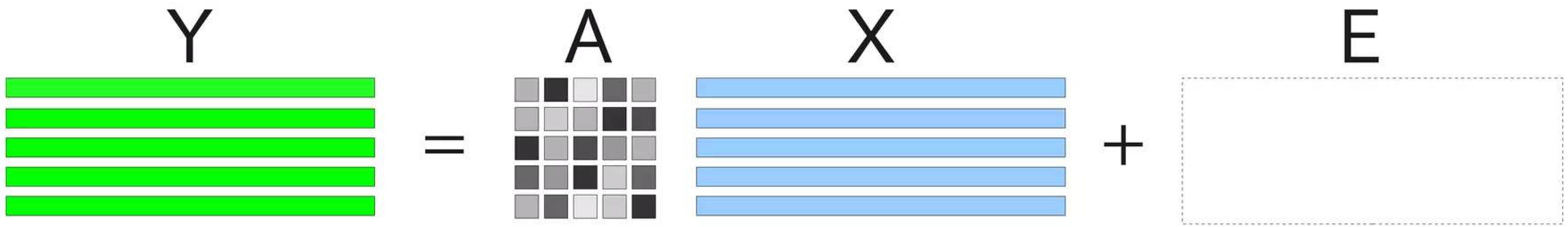
The problem of BSS is formalized as follow:
The matrix

$$X \in \mathbb{R}^{m \times d} \tag{1}$$

denotes original signals, where m is number of original signals, and d is dimension of one signal.
We consider that the observed signals $Y \in \mathbb{R}^{n \times d}$ are given by linear mixing system as

$$Y = AX + E, \tag{2}$$

where $A \in \mathbb{R}^{n \times m}$ is the unknown mixing matrix and $E \in \mathbb{R}^{n \times d}$ denotes a noise.
Basically, $n \geq m$.
The goal of BSS is to estimate $\hat{A}$ and $\hat{X}$ so that $\hat{X}$ provides unknown original signal as possible.



Kinds of BSS Methods

Actually, degree of freedom of BSS model is very high to estimate A and X.
Bcause there are a huge number of combinations (A, X) which satisfy $Y = AX + E$.
Therefore, we need some constraint to solve the BSS problem such as:

- PCA : orthogonal constraint
- SCA : sparsity constraint
- NMF : non-negativity constraint
- ICA : independency constraint

In this way, there are many methods to solve the BSS problem depending on the constraints. What we use is depend on subject matter.
I will introduce about **Non-negative Matrix Factrization(NMF)** based technique for BSS.

Non-negativity

Non-negativity means that value is 0 or positive (i.e. at least 0). Non-negative matrix is a matrix which its all elements are at least 0. For example

$$\begin{pmatrix} 1 & 10 \\ -10 & 1 \end{pmatrix} \notin \mathbb{R}_+^{2\times 2}, \quad \begin{pmatrix} 1 & 10 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}_+^{2\times 2}. \tag{3}$$

Why nonnegativity constraints?

This is the key-point of this method.
We can say that because many real-world data are nonnegative and the corresponding hidden components have a physical meaning only when nonnegative.
Example of nonnegative data,

- image data (luminance)
- power spectol of signal (speech, EEG etc.)

In general, observed data can be considered as a linear combination of such nonnegative data with noise.(i.e. $Y = AX + E$)

Standard NMF Model

Consider the standard NMF model, given by:

$$Y = AX + E, \tag{4}$$
$$A \geq 0, X \geq 0. \tag{5}$$

For simply, we denote $A \geq 0$ as nonnegativity of matrix A.
The NMF problem is given by:

NMF problem

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}||Y - AX||_F^2 && (6) \\ &\text{subject to} \quad A \geq 0,\ X \geq 0 && (7) \end{aligned}$$

I will introduce **the alternating least square(ALS) algorithm** to solve this problem.

ALS algorithm

Our goal is to estimate $(\hat{A}, \hat{X}) = \min_{A,X} \frac{1}{2}||Y - AX||_F^2$, s.t. $A \geq 0, X \geq 0$.
However, it is very difficult to solve such a multivariate problem.
The ALS strategy is to solve following sub-problems alternately.

sub-problem for A

$$\min \frac{1}{2}||Y - AX||_F^2 \tag{8}$$

$$\text{s.t. } A \geq 0 \tag{9}$$

sub-problem for X

$$\min \frac{1}{2}||Y - AX||_F^2 \tag{10}$$

$$\text{s.t. } X \geq 0 \tag{11}$$

How to Solve the sub-problem

Solving procedure of sub-problem for A is only two steps as follow:

sub-procedure for A

1. $A \leftarrow \operatorname{argmin}_A \frac{1}{2}||Y - AX||_F^2$
2. $A \leftarrow [A]_+$

where $[a]_+ = \max(\varepsilon, a)$ rectify to enforce nonnegativity.
In a similar way, we can solve sub-problem for X as follow:

sub-prodecure for X

1. $X \leftarrow \operatorname{argmin}_X \frac{1}{2}||Y - AX||_F^2$
2. $X \leftarrow [X]_+$

Solution of Least Squares

Here, I show basic calculation of least squares. First, objective function can be transformed as

$$\frac{1}{2}||Y - AX||_F^2 \tag{12}$$

$$= \frac{1}{2}\mathrm{tr}(Y - AX)(Y - AX)^T \tag{13}$$

$$= \frac{1}{2}\mathrm{tr}(YY^T - 2AXY^T + AXX^TA^T) \tag{14}$$

The stationary points $\hat{A}$ can be found by equating the gradient components to 0:

$$\frac{\partial}{\partial A^T}\frac{1}{2}\mathrm{tr}(YY^T - 2AXY^T + AXX^TA^T) \tag{15}$$

$$= -XY^T + XX^TA^T = 0 \tag{16}$$

Therefore,

$$\hat{A}^T = (XX^T)^{-1}XY^T. \tag{17}$$

Solution of Least Squares (cont'd)

In similar way, we can also obtain a solution of X. The objective function can be transformed as

$$
\begin{aligned}
&\frac{1}{2}||Y - AX||_F^2 && (18)\\
&= \frac{1}{2}\mathrm{tr}(Y - AX)^T(Y - AX) && (19)\\
&= \frac{1}{2}\mathrm{tr}(Y^TY - 2X^TA^TY + X^TA^TAX) && (20)
\end{aligned}
$$

The stationary points $\hat{X}$ can be found by equating the gradient components to 0:

$$
\begin{aligned}
&\frac{\partial}{\partial X}\frac{1}{2}\mathrm{tr}(Y^TY - 2X^TA^TY + X^TA^TAX) && (21)\\
&= -A^TY + A^TAX = 0 && (22)
\end{aligned}
$$

Therefore,

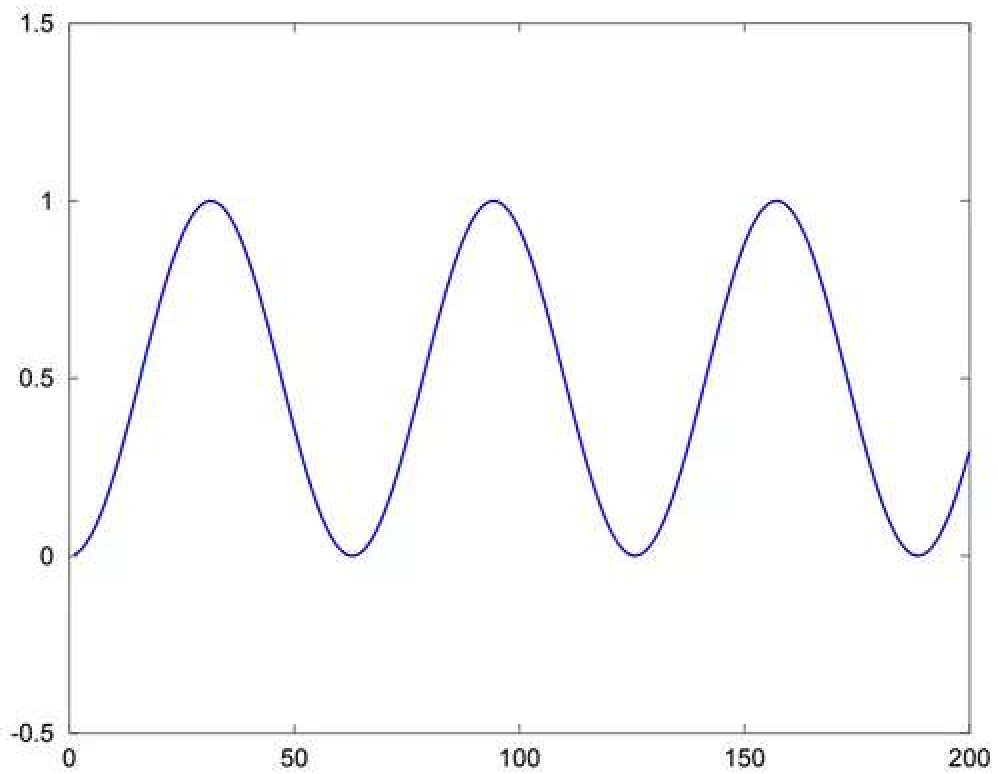
$$\hat{X} = (A^TA)^{-1}A^TY. \tag{23}$$

Implimentation

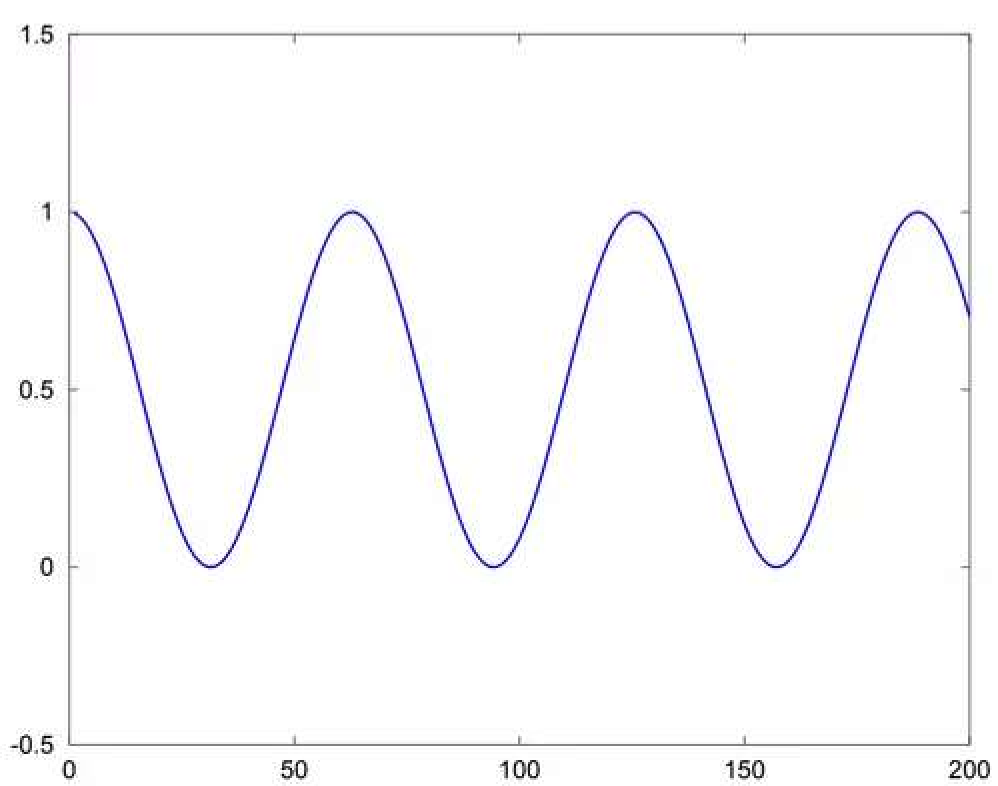
Implimentation in “octave” is very simple:

```
i=0;
while 1
  i++;
  printf("%d: %f \n",i,sqrt(sum(sum(E.^2))));

  if sqrt(sum(sum(E.^2))) < er || i > maxiter
    break;
  end

  At=(X*X'+ 1e-5*eye(J))\(X*Y');A=At';
  A(A < 0)=1e-16;
  X=(A'*A + 1e-5*eye(J))\(A'*Y);
  X(X < 0)=1e-16;

  E=Y-A*X;
end
```

Experiments: Artificial Data 1

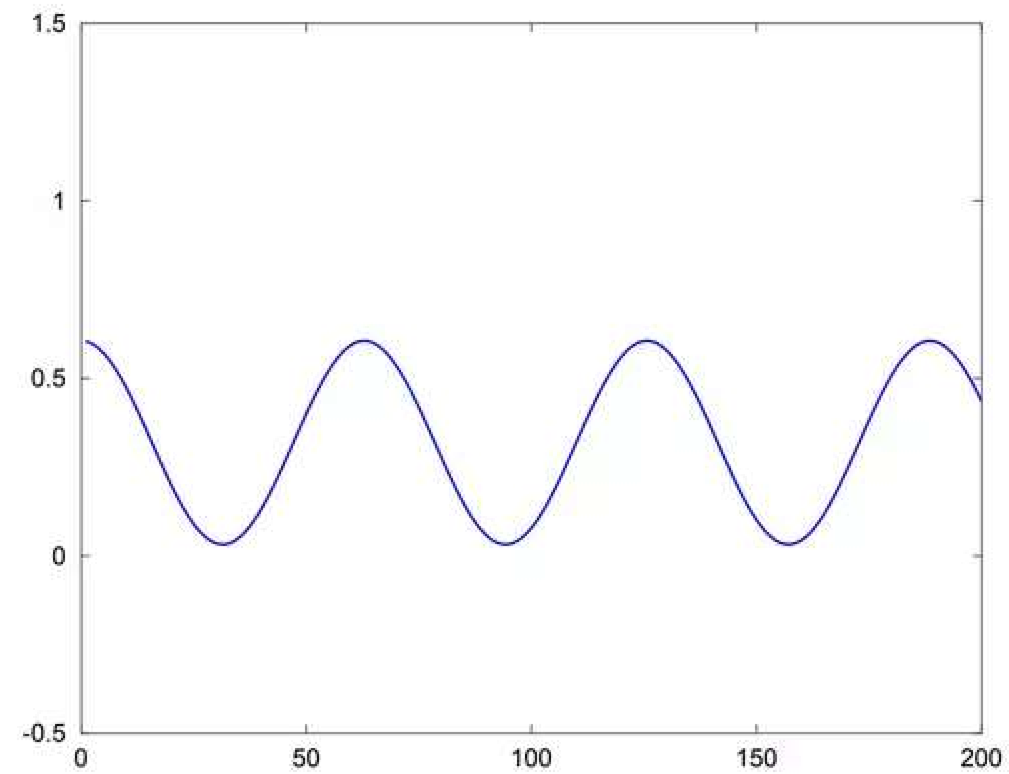


(a) $\sin^2(t)$

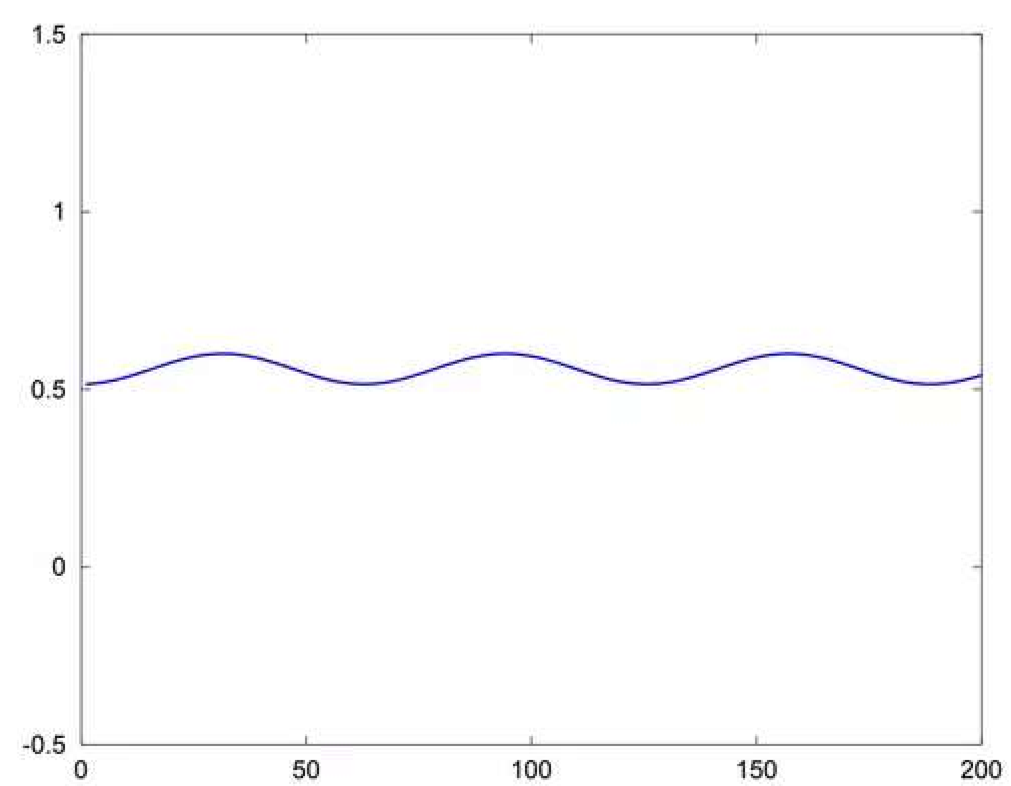


(b) $\cos^2(t)$

Figure: Original Signals

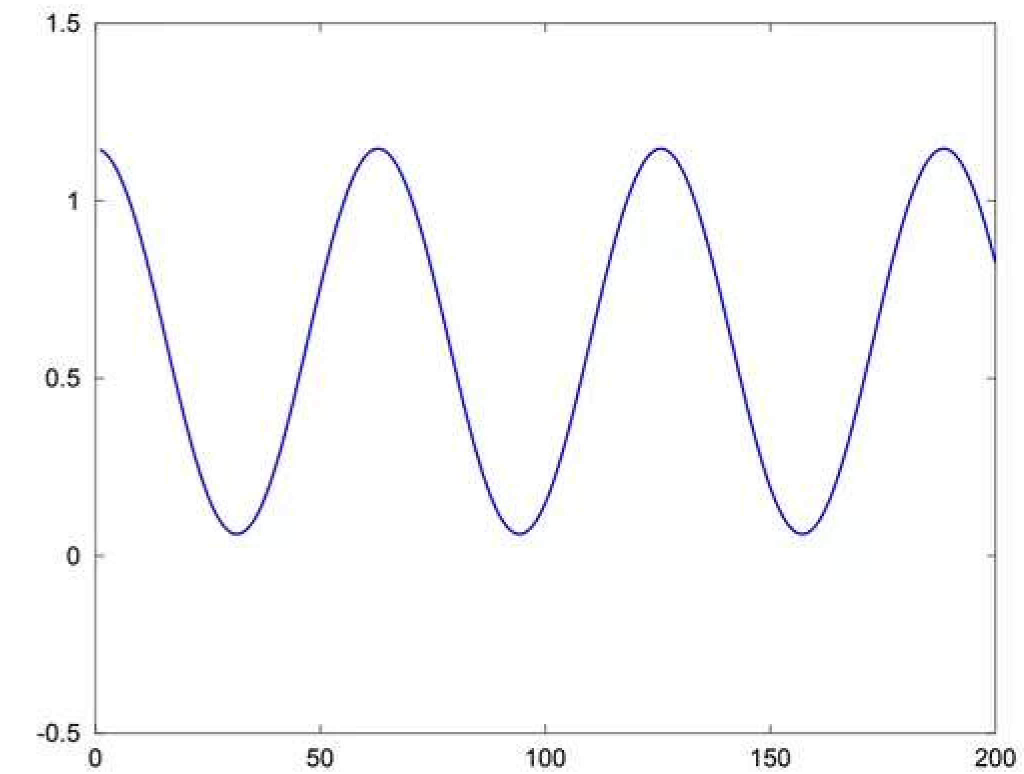


(a) mixed (observed) signal 1

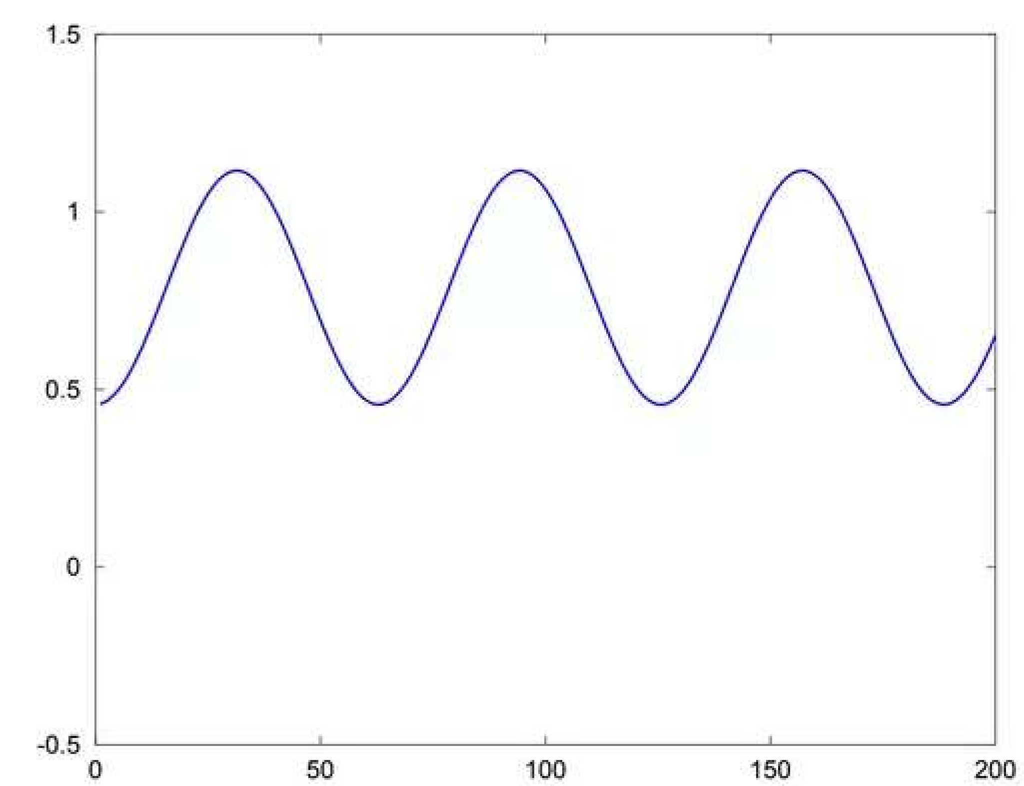


(b) mixed (observed) signal 2

Figure: Observed Signals



(a) estimated signal 1

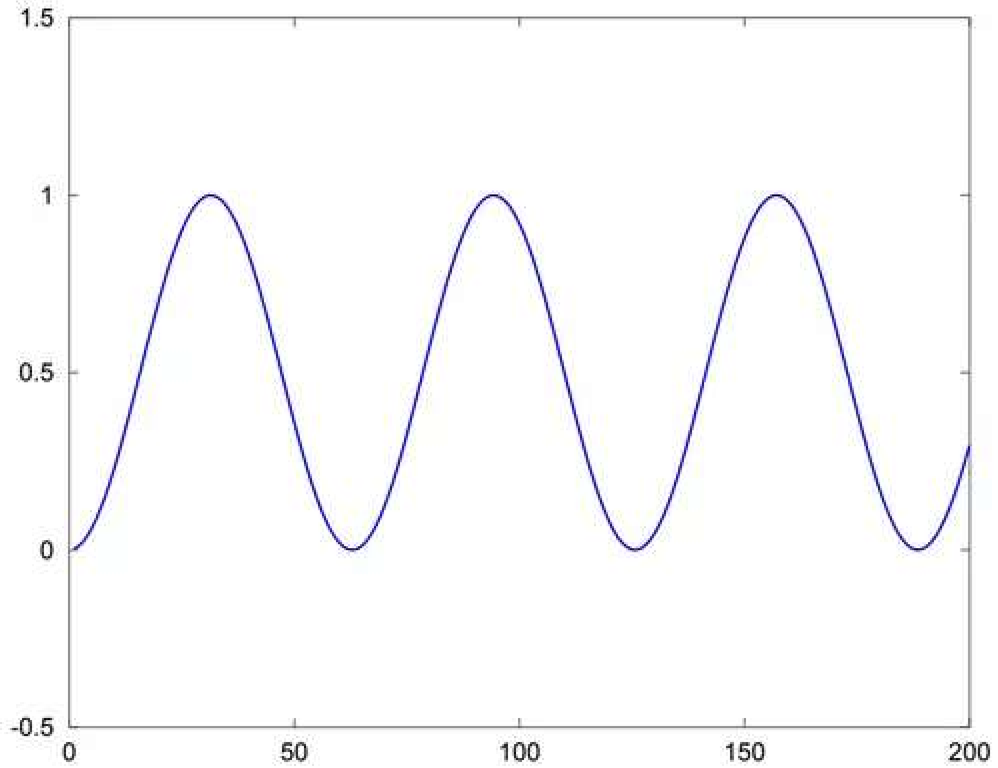


(b) estimated signal 2

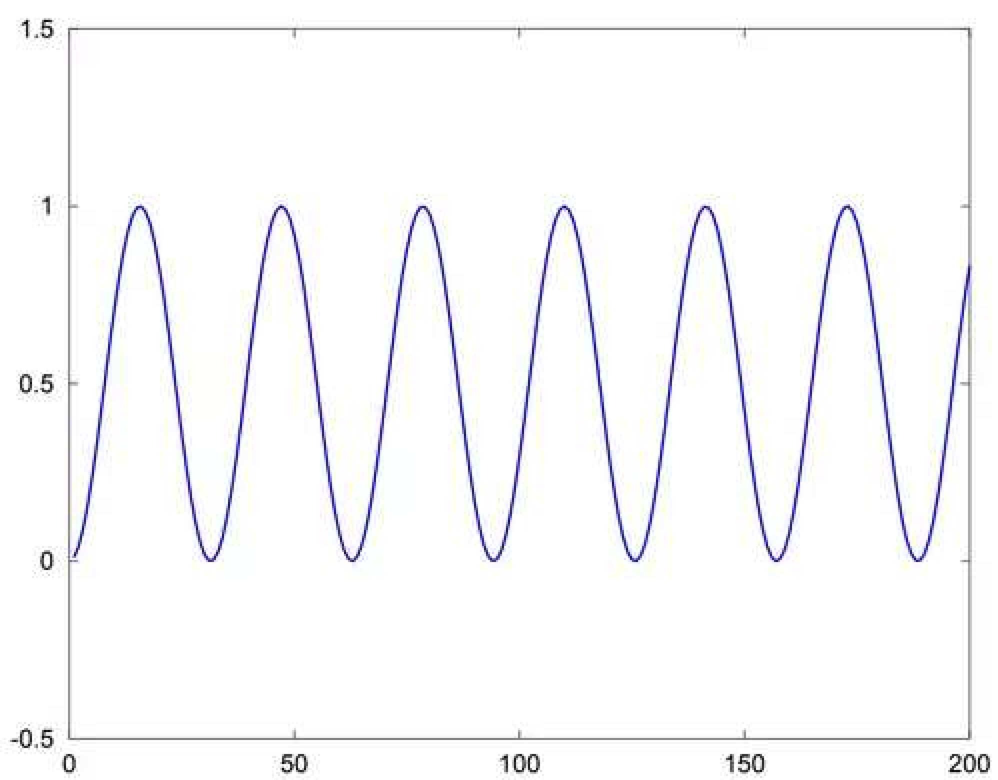
Figure: Estimated Signals

We can see that results are almost separated.

Experiments: Artificial Data 2

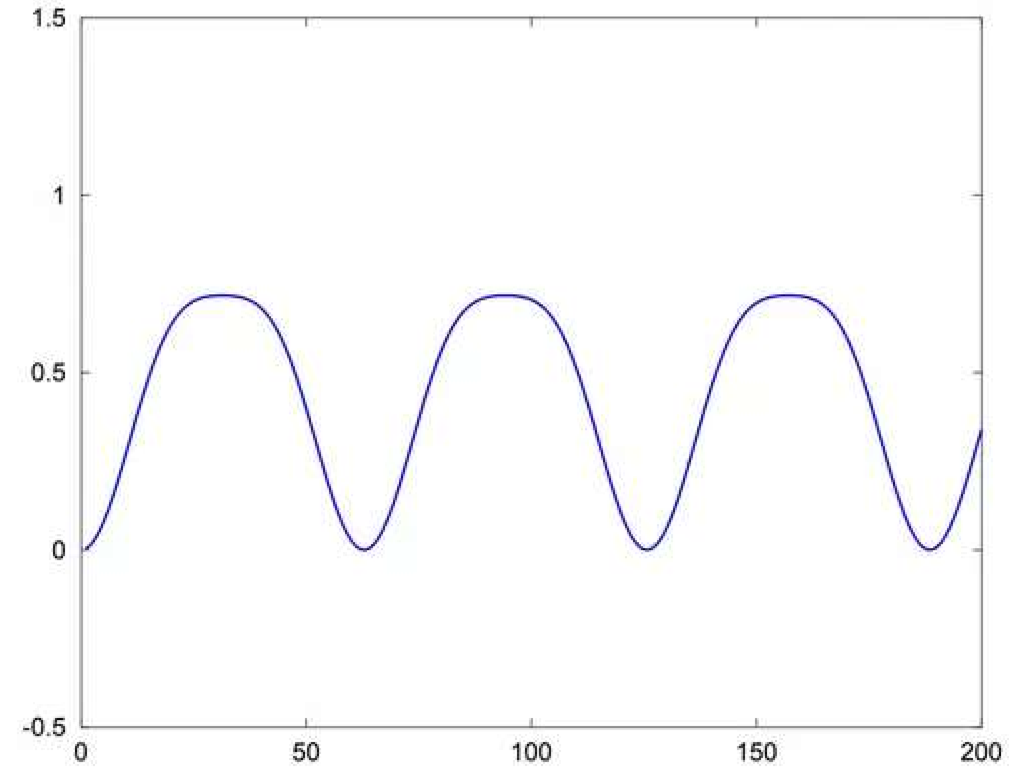


(a) $\sin^2(t)$

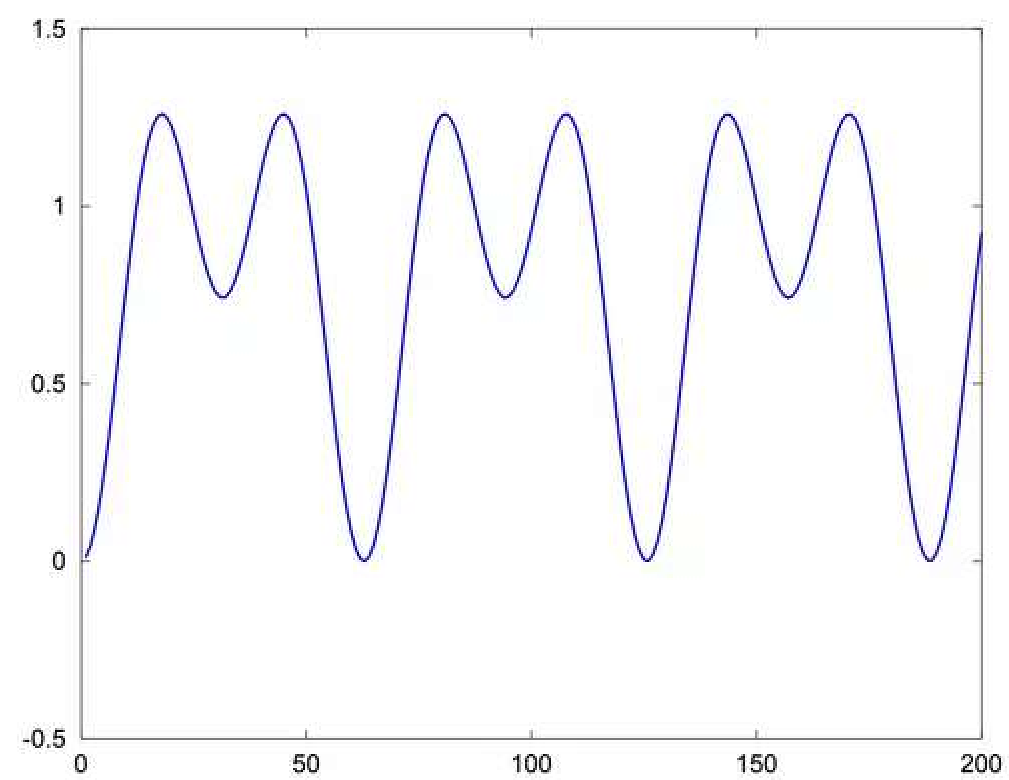


(b) $\sin^2(2t)$

Figure: Original Signals

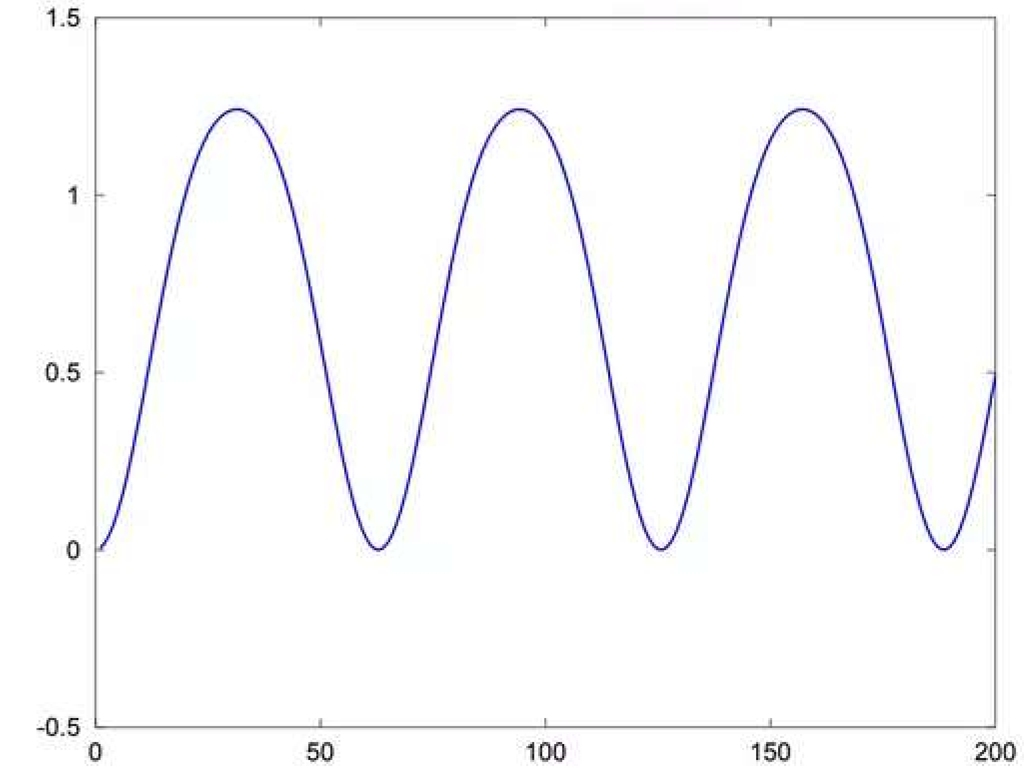


(a) mixed (observed) signal 1

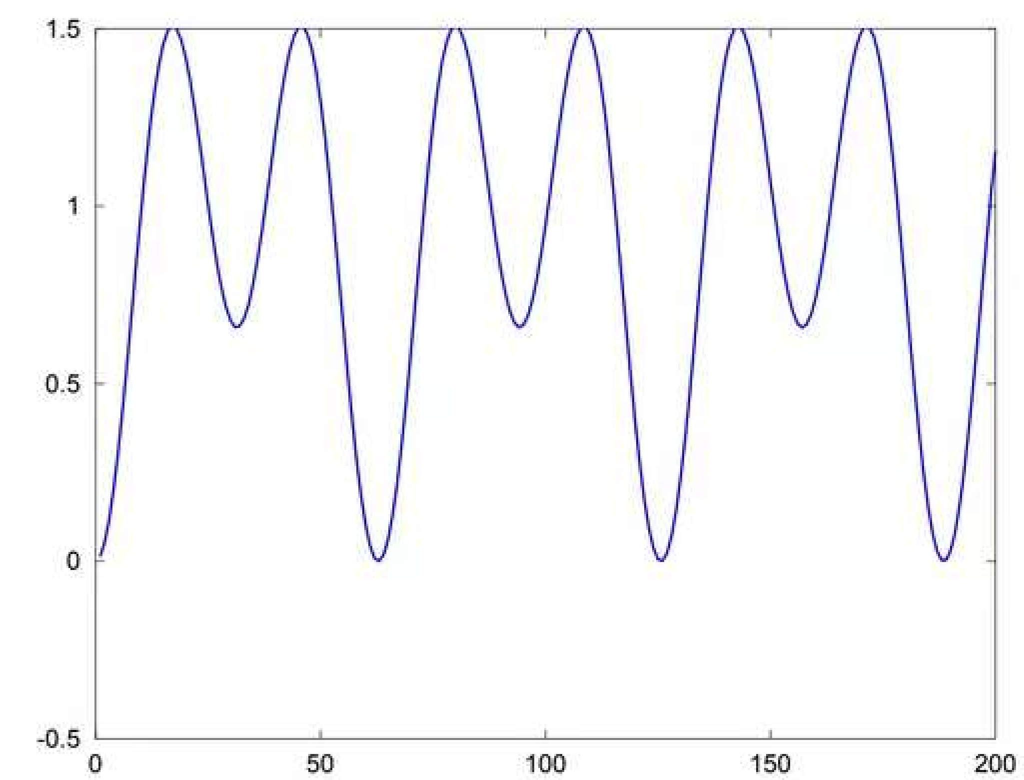


(b) mixed (observed) signal 2

Figure: Observed Signals



(a) estimated signal 1



(b) estimated signal 2

Figure: Estimated Signals

We can see that results are not so separated.

Experiments: Real Image 1

(a) newyork

(b) shanghai

Figure: Original Signals

(a) mixed (observed) signal 1

(b) mixed (observed) signal 2

Figure: Observed Signals

(a) estimated signal 1

(b) estimated signal 2

Figure: Estimated Signals

We can see that results are almost separated.

Experiments: Real Image 2

(a) rock

(b) pig

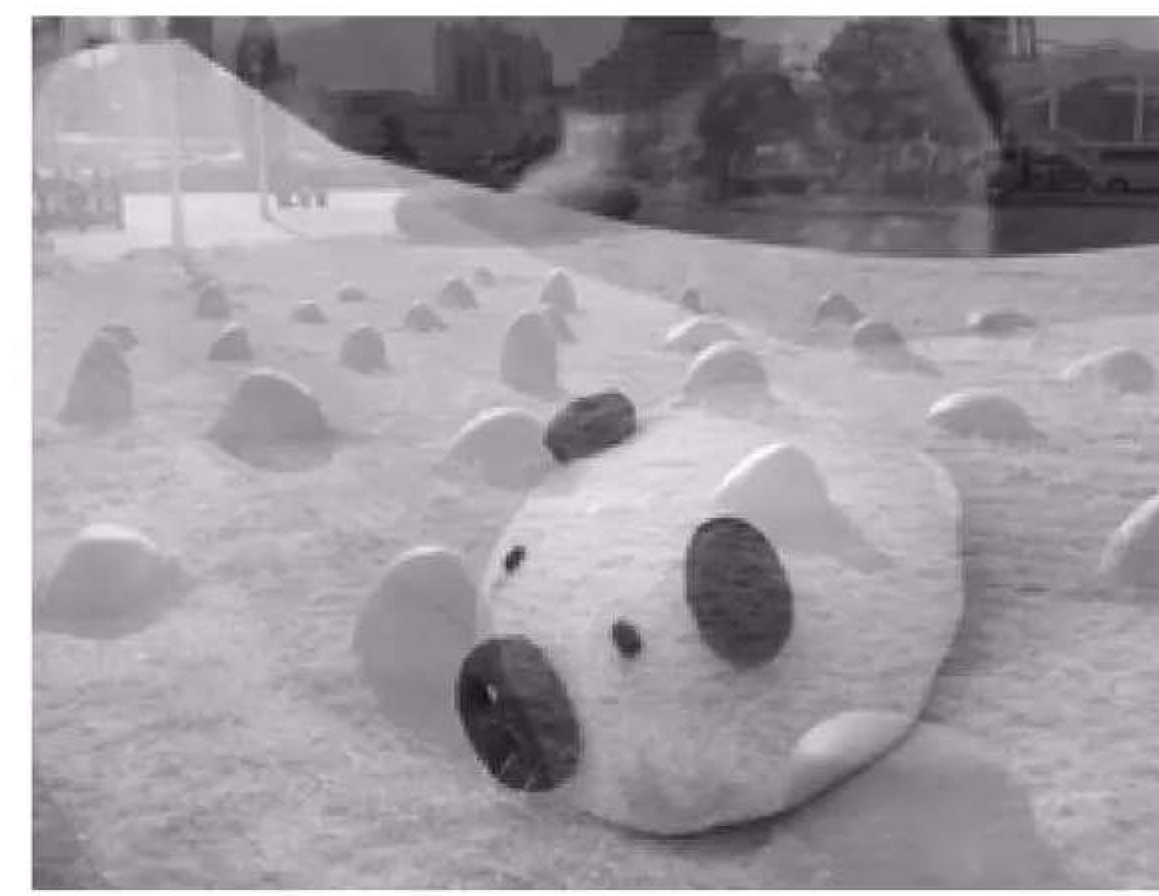
Figure: Original Signals

(a) mixed (observed) signal 1

(b) mixed (observed) signal 2

Figure: Observed Signals

(a) estimated signal 1



(b) estimated signal 2

Figure: Estimated Signals

We can see that estimate 2 is not so separated.

Experiments: Real Image 3


(a) nyc


(b) sha


(c) rock


(d) pig


(e) obs1


(f) obs2


(g) obs3


(h) obs4

Figure: Ori. & Obs.


(a) estimated signal 1


(b) estimated signal 2


(c) estimated signal 3


(d) estimated signal 4

Figure: Estimated Signals: Results are not good. Signals are still mixed.

Issues

There are some issues of this method as follow:

- Solution is non-unique (i.e. result is depend on starting point)
- It is not enough only by non-negativity constraint (shoud consider independency of each components)
- When number of estimate signals is large, features of components are overlapped.

And we should consider more general issues of BSS as follow:

- Number of original signals is unknown (How can we decide its number?)

About this research area

In this research area, many method for BSS are studied and proposed as follow:

- KL-divergence based NMF [Honkela et al., 2011]
- Alpha-Beta divergences based NMF [Cichocki et al., 2011]
- Non-Gaussianity based ICA [Hyvärinen et al., 2001]
  - Kurtosis based ICA
  - Negentropy based ICA
- Solving method for ICA
  - gradient method
  - fast fixed-point algorithm [Hyvärinen and Oja, 1997]
- MLE based ICA
- Mutual information based ICA
- Non-linear ICA
- Tensor ICA

In this way, this research area is very broad!!

Summary

- I introduced about BSS problem and basic NMF tequnique from [Cichocki et al., 2009].
- EEG classification is very difficult problem.

Future Work

- To study about extension of NMF and Independent Component Analysis (ICA).
- To apply the BSS problem, EEG analysis, and some pattern recognition problem.

Bibliography I

[Cichocki et al., 2011] Cichocki, A., Cruces, S., and Amari, S.-i. (2011).
Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization.
Entropy, 13(1):134–170.

[Cichocki et al., 2009] Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S. (2009).
Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis.
Wiley.

[Honkela et al., 2011] Honkela, T., Duch, W., Girolami, M., and Kaski, S., editors (2011).
Kullback-Leibler Divergence for Nonnegative Matrix Factorization, volume 6791 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.

[Hyvärinen et al., 2001] Hyvärinen, A., Karhunen, J., and Oja, E. (2001).
Independent Component Analysis.
Wiley.

Bibliography II

[Hyvärinen and Oja, 1997] Hyvärinen, A. and Oja, E. (1997).
A fast fixed-point algorithm for independent component analysis.
Neural Computation, 9:1483–1492.

Thank you for listening