OSAKA KYOIKU UNIVERSITY

Image separation by wavelet analysis

Ryuichi Ashino

Joint work with Takeshi Mandai and Akira Morimoto

Workshop on Microlocal Methods in Medical Imaging
Fields Institute, August 16, 2012

Outline

Blind signal source separation

Time-scale information matrix defined by multiwavelets

Image separation by our previous method

Image separation by source reduction method (new method)

Blind signal source separation

The cocktail party problem:

Study the human listening ability to focus one's listening attention on a single talker among a mixture of conversations and background noise.

Make a machine to solve the cocktail party problem

in a satisfactory manner.

In the signal processing community, this problem is called

Blind source separation problem

Inverse problem

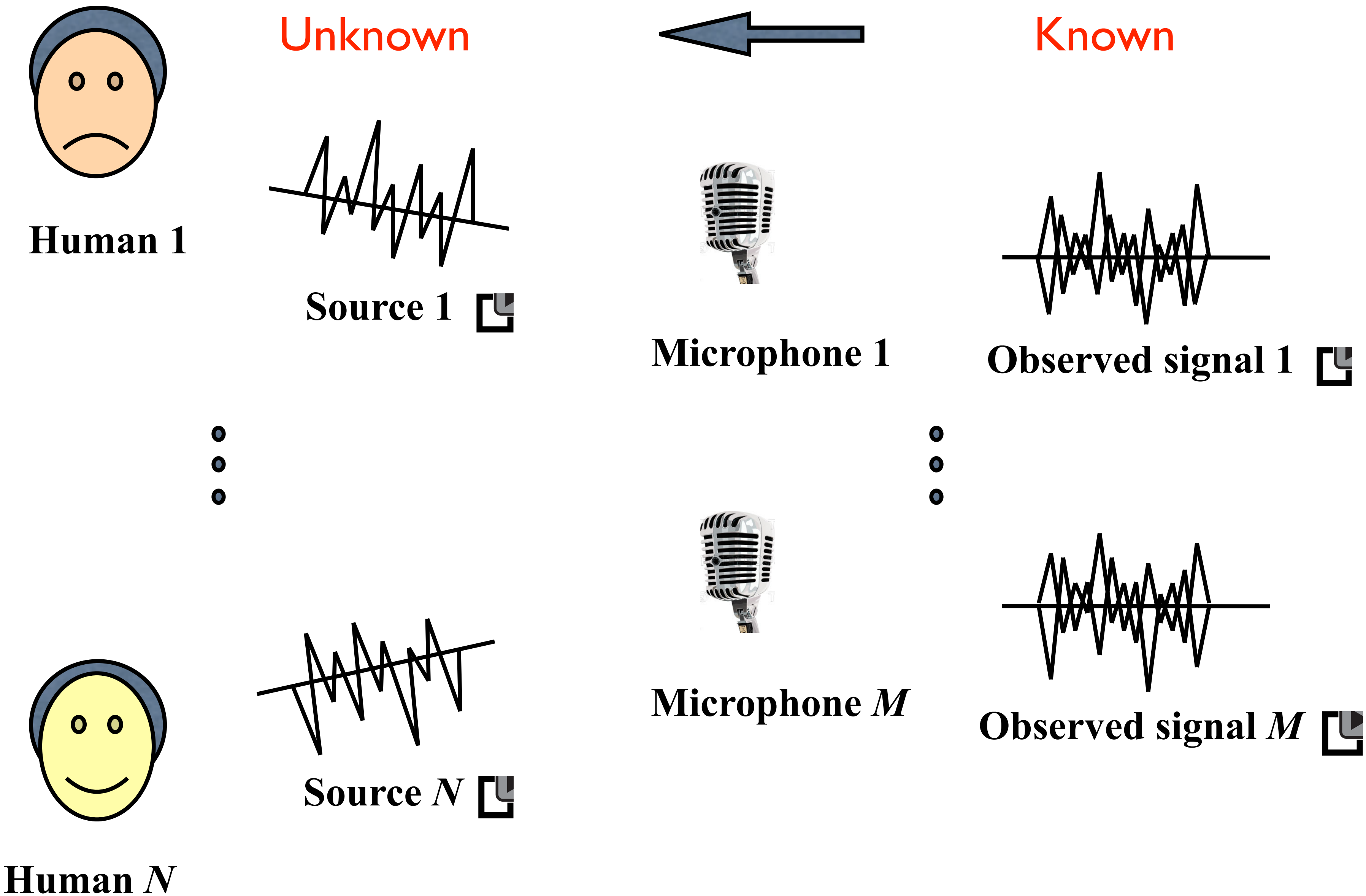
Independent Component Analysis

Independent Component Analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals.

To separate observed signals, we need to assume some properties which the sources should have.

ICA assumes statistical independency of sources.

We assume independency of time-frequency information of sources.


Unknown
Known
Human 1
Source 1
Microphone 1
Observed signal 1
Human N
Source N
Microphone M
Observed signal M

Mixing model

$$s(t) = (s_1(t), \ldots, s_N(t))^{\mathrm{T}} : \; N \text{ source signals}$$

$$x(t) = (x_1(t), \ldots, x_M(t))^{\mathrm{T}} : \; M \text{ observed signals}$$

$$t \in \mathbb{R}^1 : \text{ sound signal}, \quad t \in \mathbb{R}^2 : \text{ image}$$

Mixing model

$$x_j(t) = \sum_{k=1}^{N} a_{j,k}\, s_k(t), \qquad j = 1, \ldots, M$$

$$x(t) = As(t)$$

$$A = (a_{j,k}) \in \mathbb{R}^{M \times N} : \text{ mixing matrix}$$

Assume that $M \geq N$ and A is of full-rank.

Blind source separation

$$x(t) = (x_1(t), \dots, x_M(t))^{\mathsf{T}} : \ M \text{ observed signals}$$

+ Mixing model (assumption)

Unknown: N and $A = (a_{j,k}) \in \mathbb{R}^{M \times N}$

Estimate

$$N : \text{ the number of source signals}$$

$$A : \text{ mixing matrix}$$

$$s(t) = (s_1(t), \dots, s_N(t))^{\mathsf{T}} : \ N \text{ source signals}$$

Ambiguities:

Impossible to determine the energies of the column vectors of A.
(We will normalize the column vectors of A with length one.)

Impossible to determine the order of the sources.

The space-time mixture model (one dimensional case)

$$x_j(t) = \sum_{k=1}^{N} a_{j,k}\, s_k(t - c_{j,k}), \qquad j = 1, \ldots, M$$

$$A = (a_{j,k}) \in \mathbb{R}^{M\times N}: \text{ mixing matrix}$$

$$C = (c_{j,k}), \in \mathbb{R}^{M\times N}: \text{ delay matrix}$$

$$c_{j,k} > 0$$

Space-time mixture problem:

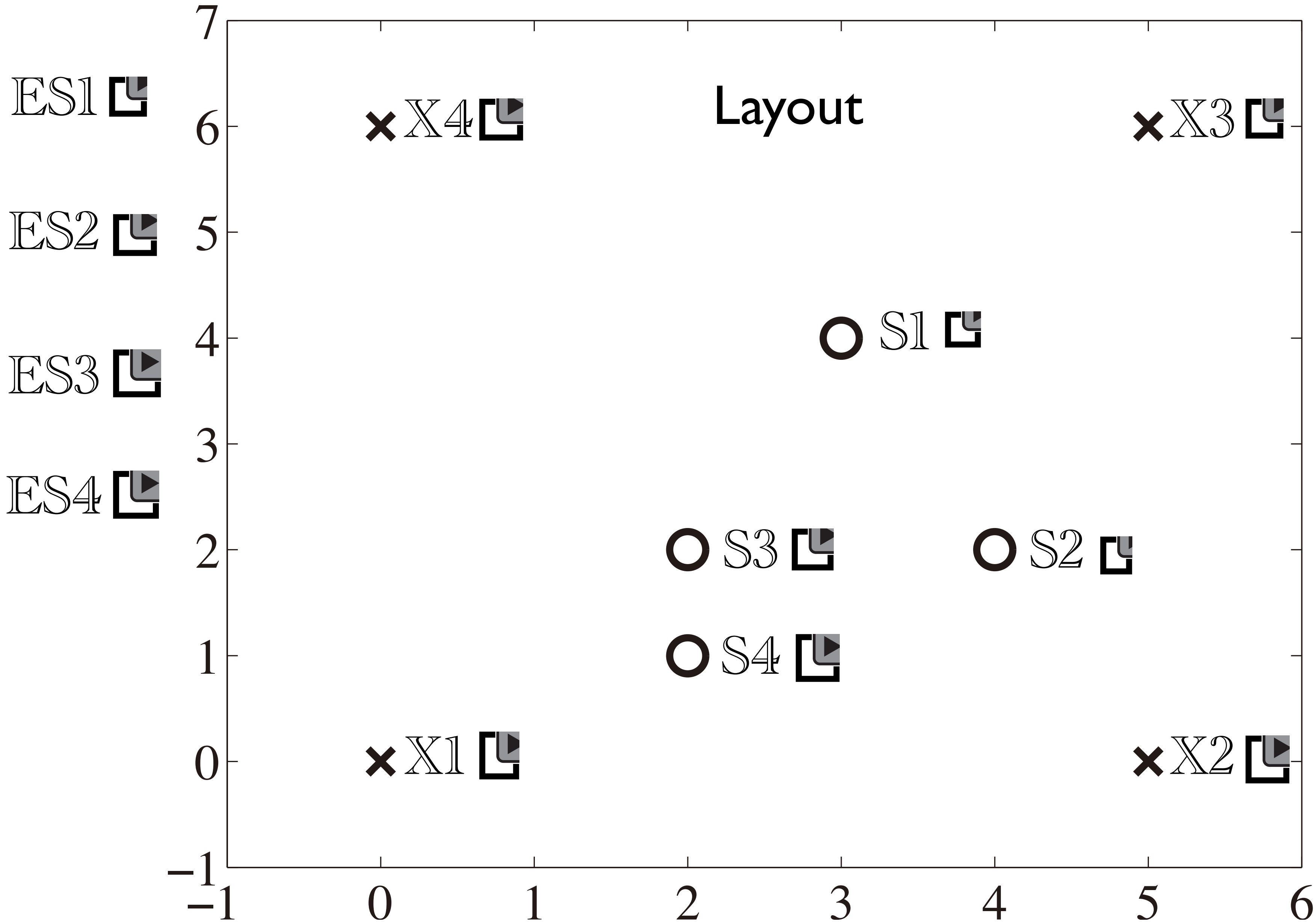
Wavelet is much simpler.

The wavelet transform of the model:

$$X_j(t,\omega) = \sum_{k=1}^{n} a_{j,k} S_k(t - c_{j,k}, \omega)$$

The windowed Fourier transform of the model:

$$\widetilde{X}_j(t,\omega) = \sum_{k=1}^{n} \boxed{e^{-ic_{j,k}\omega}} a_{j,k}\, \widetilde{S}_k(t - c_{j,k}, \omega)$$



Layout
ES1
ES2
ES3
ES4
X1
X2
X3
X4
S1
S2
S3
S4
7
6
5
4
3
2
1
0
−1
−1
0
1
2
3
4
5
6

Key idea

x_1

x_2

x_3

Find t_k such that
$$s_k(t_k) \neq 0$$
$$s_\ell(t_k) = 0, \quad \ell \neq k.$$

Only k^{th} source signal is active at $t = t_k$.

Mixing model: $$x_j(t) = \sum_{k=1}^{N} a_{j,k}\, s_k(t), \qquad j = 1, \ldots, M$$

$$x_j(t_k) = a_{j,k}\, s_k(t_k)$$

$$x(t_k) = \boxed{(a_{1,k}, \ldots, a_{M,k})^{\mathrm{T}}}\, s_k(t_k)$$

We can estimate *k*-th column of the mixing matrix *A* up to constant multiple.

The Fourier transform of the mixing model:

$$\widehat{x}_j(\xi) = \sum_{k=1}^{N} a_{j,k}\,\widehat{s}_k(\xi), \qquad j = 1, \ldots, M$$

$$\widehat{x}(\xi) = A\widehat{s}(\xi)$$

Find ξ_k such that $\widehat{s}_k(\xi_k) \neq 0$

$$\widehat{s}_\ell(\xi_k) = 0, \quad \ell \neq k$$

$$\widehat{x}(\xi_k) = (a_{1,k}, \ldots, a_{M,k})^{\mathtt{T}}\widehat{s}_k(\xi_k)$$

Observation:

Any integral transform can be applied to the mixing model.

Estimation of the mixing matrix A can be reduced to find points where only one transformed source signal is active.

Problem:

Since the number N of sources is unknown,
how many points do we need to find?

How to find points where only one transformed source signal is active?

Choice of the integral transform

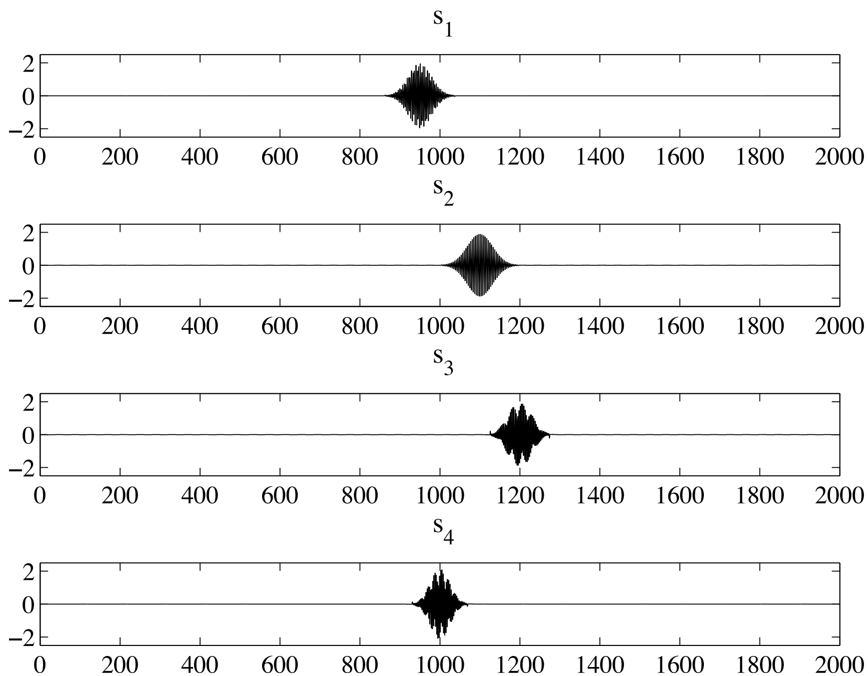
Human can focus on a single talker
if only one talker is speaking in a certain interval
or talkers' voices are different.

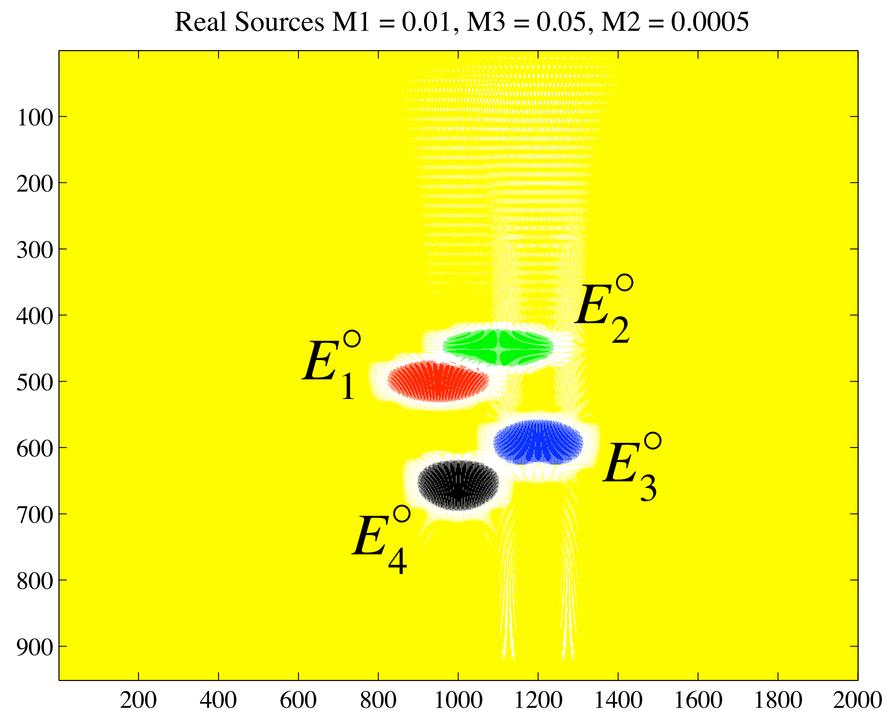
Integral transforms used in time-frequency analysis,
such as the wavelet transform, could be fine.

For image analysis,
the wavelet transform performs well.

Different wavelets give different time-scale information.
By using several proper wavelets (multiwavelet),
we can have better performance.

Time-scale independency

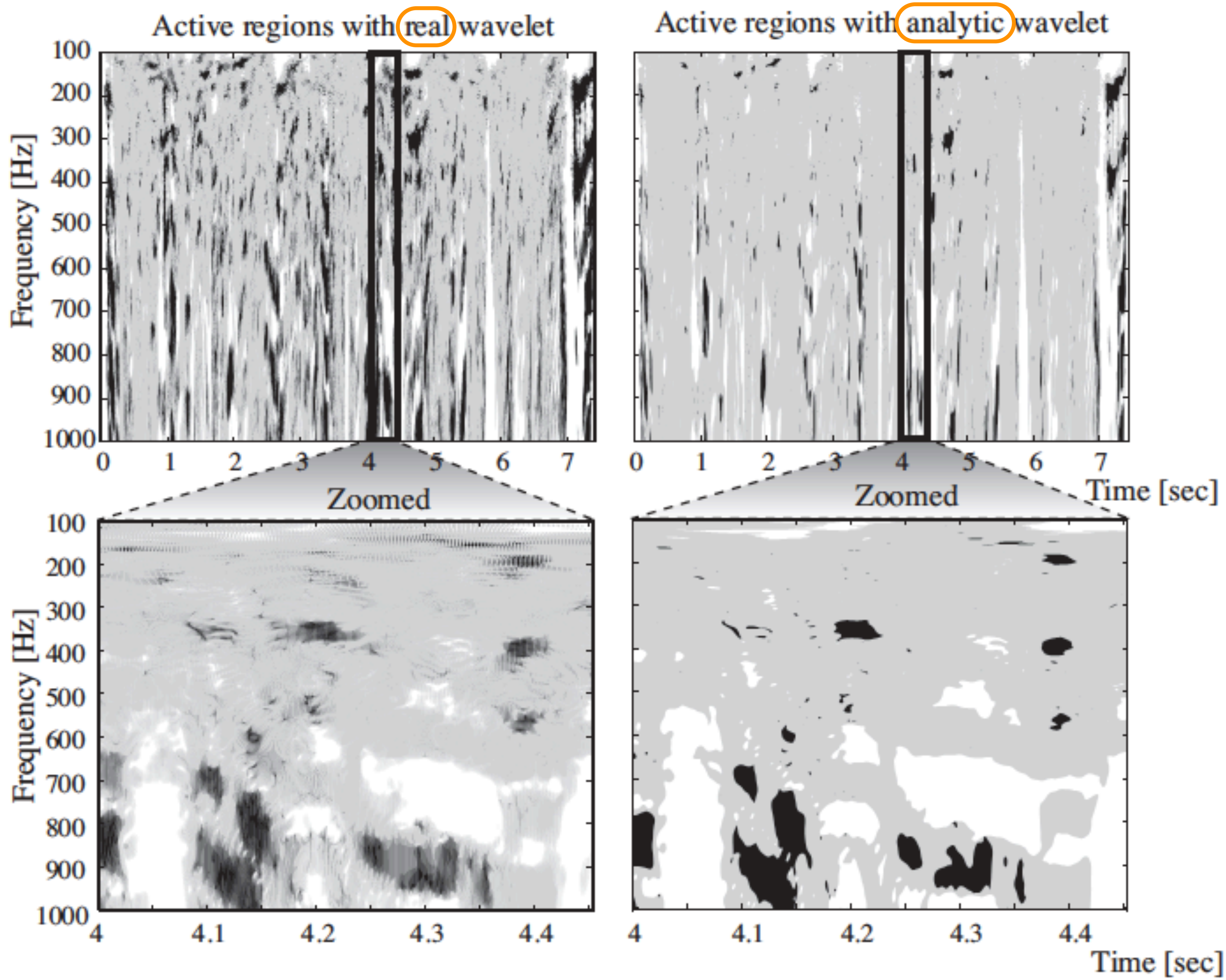




Red, **Green**, **Blue**, **Black**: only one signal is active.

White: several signals are active.

Yellow: no signal is active.


Active regions with real wavelet
Active regions with analytic wavelet
Frequency [Hz]
100
200
300
400
500
600
700
800
900
1000
0
1
2
3
4
5
6
7
Time [sec]
Zoomed
Zoomed
4
4.1
4.2
4.3
4.4
Time [sec]

Continuous wavelet transform

Wavelet transform of signal $f(t)$ with respect to wavelet $\psi(t)$

$$(W_\psi f)(b,\alpha) = \int_{\mathbb{R}^n} f(t)\,\alpha^{-n/2}\overline{\psi\left(\frac{t-b}{\alpha}\right)}\,dt, \quad \alpha \in \mathbb{R}_+,\ b \in \mathbb{R}^n$$

We use several wavelet functions.

$\psi^1(t), \ldots, \psi^P(t)$: real-valued wavelets

The same centers of time-domain and frequency-domain

Continuous multiwavelet transform

The continuous multiwavelet transform of $f \in L^2(\mathbb{R}^n)$

with respect to a multiwavelet function $\Psi = (\psi^p)_{p=1}^P \in L^2(\mathbb{R}^n)^P$

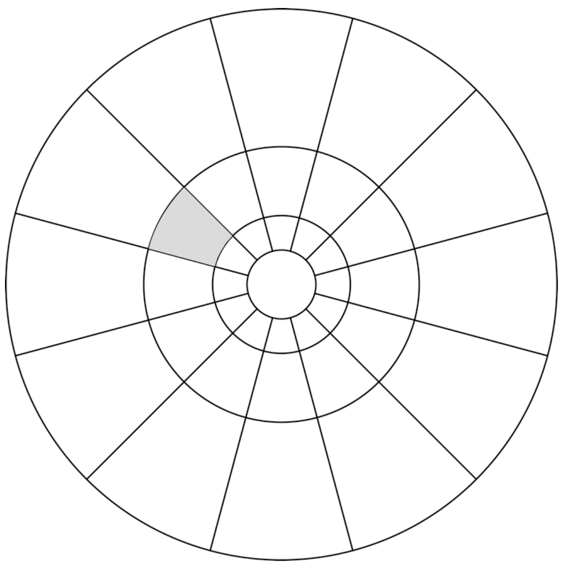
is defined by

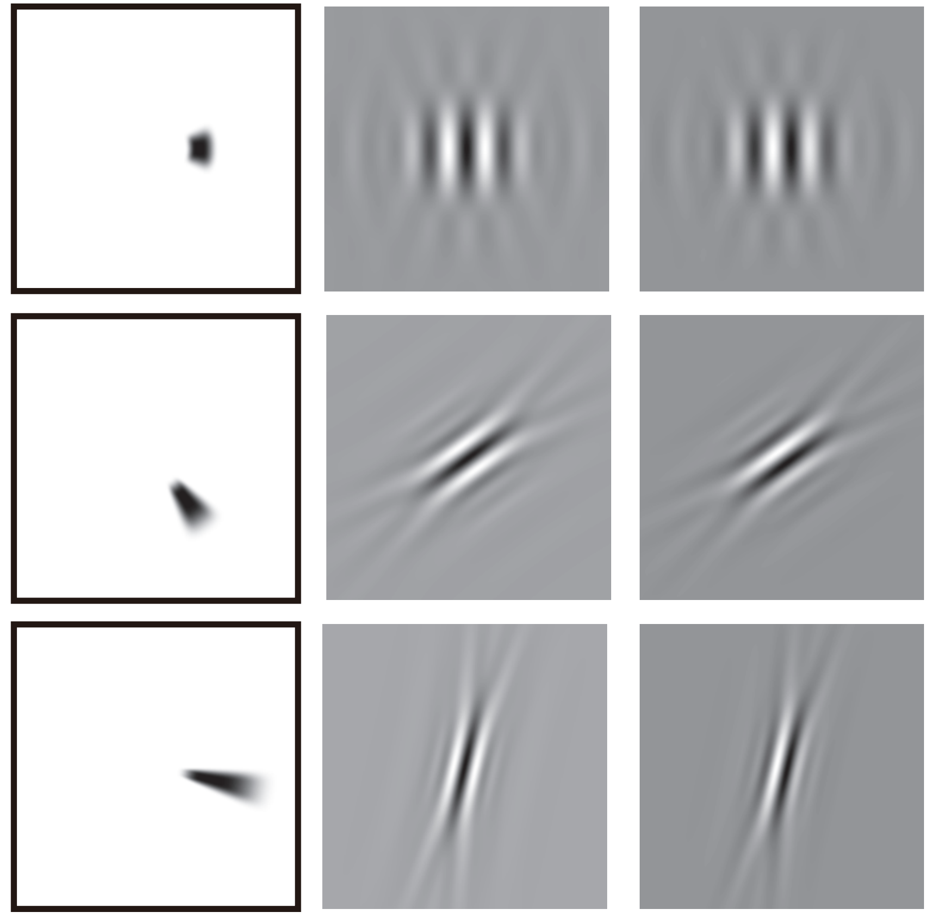
$$(W_\Psi f)(b,\alpha) := \Big((W_{\psi^p} f)(b,\alpha) \Big)_{p=1}^{P}, \quad \alpha \in \mathbb{R}_+,\; b \in \mathbb{R}^n$$

We omit the inversion formula, because it is not used here.

Multiwavelet design

Annual sectors
in Fourier space





$\widehat{\psi}$ $\Re\psi$ $\Im\psi$

We use $\Re\psi$ and $\Im\psi$.

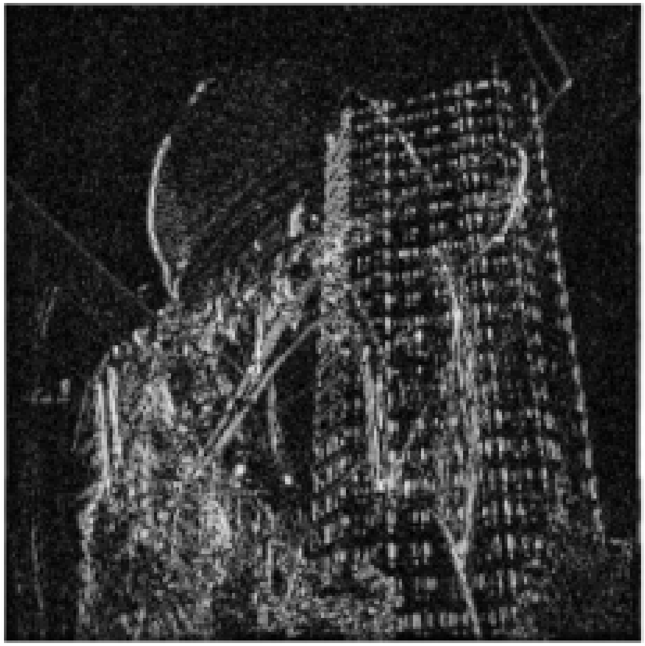
Previous method

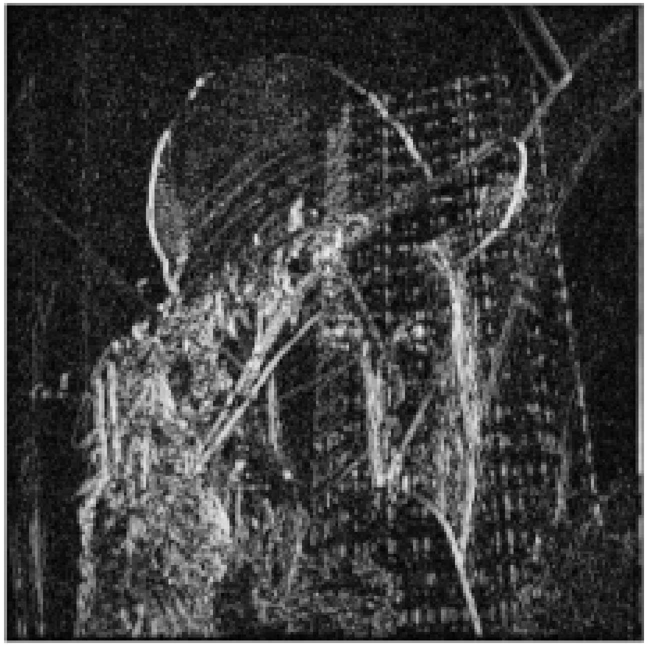
[1] R. Ashino, K. Fujita, T. Mandai, A. Morimoto, K. Nishihara, Blind source separation using time-frequency information matrix given by several wavelet transforms, *Information*, **10** (5), 555–568, 2007.

[2] R. Ashino, S. Kataoka, T. Mandai, A. Morimoto, Blind image source separations by wavelet analysis, *Appli. Anal.*, **91** (4), 617–644, 2012.

Most edges are curve segments.
Intersections of curve segments are points generally.

Points on edges are candidates for the points
where only one image is active.





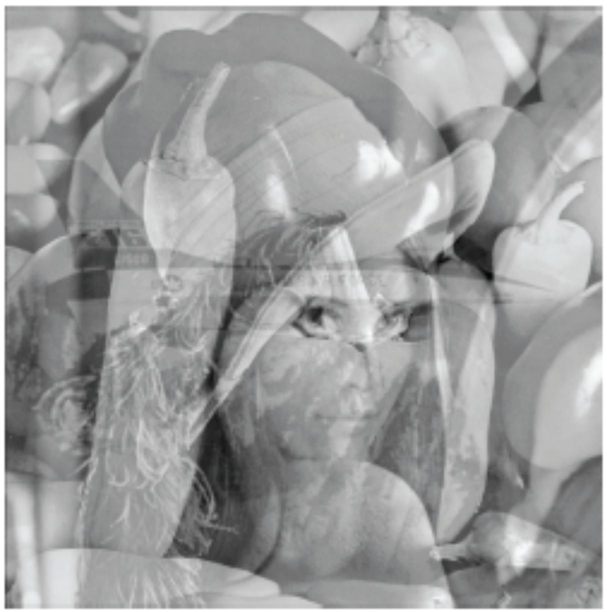
Color image separation

Separation of RGB images can be reduced to the separation of grey scale images.

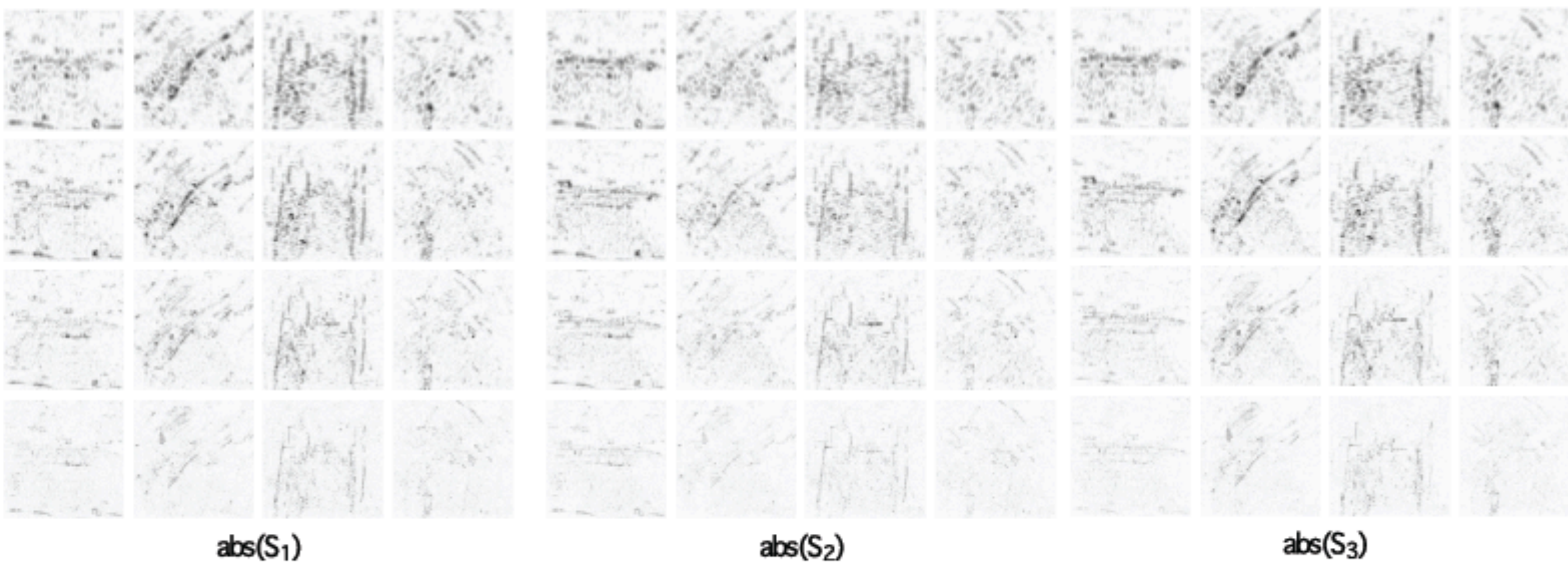
Original RGB images

Mixed images

Convert RGB images to grey scale images
and apply the proposed method



Edges analysis (We omit the detail.)



Once we know the number of source images and the mixing matrix *A*, we can apply the inverse of *A* to the observed RGB images.

Separated images

Previous Strategies:

Define the time-scale information matrix X.

Find a candidate of a column vector of the mixing matrix A
by applying SVD to X and using Kohonen's self-organizing map.

Time-scale information matrix

$$X_j^p(t,\alpha) = W_{\psi^p} x_j(t,\alpha) \ : \ \text{time-scale information}$$

$$X(t,\alpha) = \begin{pmatrix} X_1^1(t,\alpha) & \dots & X_M^1(t,\alpha) \\ \vdots & \ddots & \vdots \\ X_1^P(t,\alpha) & \dots & X_M^P(t,\alpha) \end{pmatrix} \in \mathbb{R}^{P\times M}$$

$$S_k^p(t,\alpha) = W_{\psi^p} s_k(t,\alpha) \ : \ \text{time-scale information}$$

$$S(t,\alpha) = \left(S_k^p(t,\alpha)\right)_{p=1,\dots,P;k=1,\dots,N} \in \mathbb{R}^{P\times N}$$

Mixing model

$$x(t) = As(t)$$

Linearity of CWT

Time-scale version

$$X(t,\alpha) = S(t,\alpha)\,A^{\mathrm{T}}$$

For $\mu_1 \gg \mu_2 > 0$,

$$\boxed{\begin{aligned} \mathfrak{E}_k =& \{(t,\alpha) \mid |S_k^p(t,\alpha)| > \mu_1 \text{ for some } p, \\ & |S_m^p(t,\alpha)| < \mu_2 \; (m \neq k,\ p = 1,\ldots,P)\} \end{aligned}}$$

The time-scale region $\mathfrak{E}_k$ where only $s_k(t)$ is active.

For $(t,\alpha) \in \mathfrak{E}_k$, we have

$$\boxed{X(t,\alpha) \approx \left(S_k^1(t,\alpha),\ldots,S_k^P(t,\alpha)\right)^{\mathsf{T}} (a_{1,k},\ldots,a_{M,k})\,.}$$

$\boxed{\text{The rank of } X(t,\alpha) \text{ is one.}}$

Our previous method

$\boxed{\text{Search } (t,\alpha) \text{ such that } \operatorname{rank} X(t,\alpha) = 1}$

Take the singular value decomposition of $X(t,\alpha)$:

$$\boxed{\begin{aligned} X(t,\alpha) &= U(t,\alpha)\Sigma(t,\alpha)V(t,\alpha)^{\mathrm{T}} \\ &= \sum_{p=1}^{P} \sigma_p(t,\alpha)u_p(t,\alpha)v_p(t,\alpha)^{\mathrm{T}}, \end{aligned}}$$

where $\sigma_1(t,\alpha) \geq \sigma_2(t,\alpha) \geq \cdots \geq \sigma_P(t,\alpha) \geq 0$, and

$$U(t,\alpha) = (u_1(t,\alpha),\ldots,u_P(t,\alpha)) \in \mathcal{O}_P,$$

$$\Sigma(t,\alpha) = \operatorname{diag}(\sigma_1(t,\alpha),\ldots,\sigma_P(t,\alpha)) \in \mathbb{R}^{P\times M},$$

$$V(t,\alpha) = (v_1(t,\alpha),\ldots,v_M(t,\alpha)) \in \mathcal{O}_M.$$

$\mathcal{O}_n$: the group of orthogonal matrices of degree n.

$\boxed{\text{Search } (t, \alpha) \text{ such that } \operatorname{rank} X(t, \alpha) = 1}$

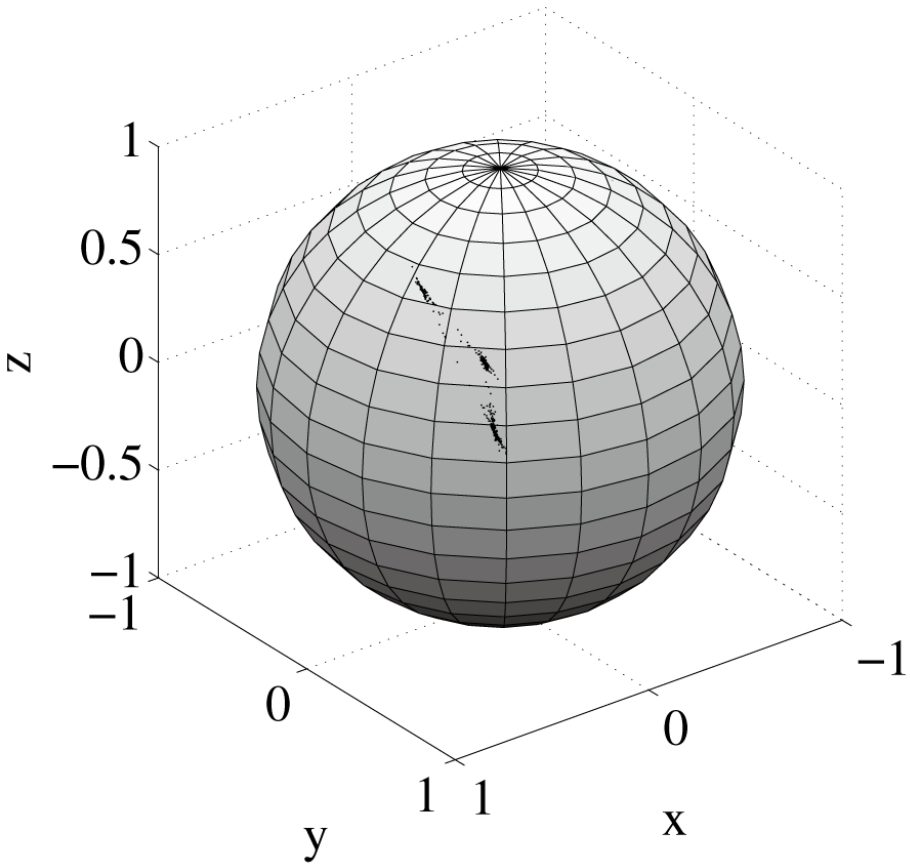
Criterion:

If $\sigma_1(t, \alpha) \gg \sigma_2(t, \alpha)$, that is, $\operatorname{rank} X(t, \alpha) = 1$, then record $v_1(t, \alpha)$.

$\boxed{v_1(t, \alpha) \text{ is a candidate of some } (a_{1,k}, \ldots . a_{M,k})^{\mathsf{T}}.}$

Kohonen's self-organizing maps

Make a clustering of $v_1(t,\alpha)$
on the $(M-1)$-dimensional unit sphere $\mathbb{S}^{M-1}$.



Three clusters

The number of sources $\widetilde{N} = 3$.

$$\widetilde{A} = \begin{pmatrix} 0.56088 & 0.60055 & 0.61022 \\ 0.62537 & 0.74897 & 0.27266 \\ 0.54245 & 0.27955 & 0.74355 \end{pmatrix}$$

Representatives of 1^{st}, 2^{nd}, 3^{rd} clusters

Observation:

The number of clusters coincides with the number of sources.

The representative of a cluster coincides with a column vector of the mixing matrix A up to constant multiple.

Merit : we can estimate mixing matrix A at once.

Demerit : 1. If the number of observed signals $M \geq 4$,
we cannot see $\mathbb{S}^{M-1}$.
2. Some clusters are very crowded than the others.
3. For 512x512 image separation,
we can separate up to $(M, N) = (5, 4)$.

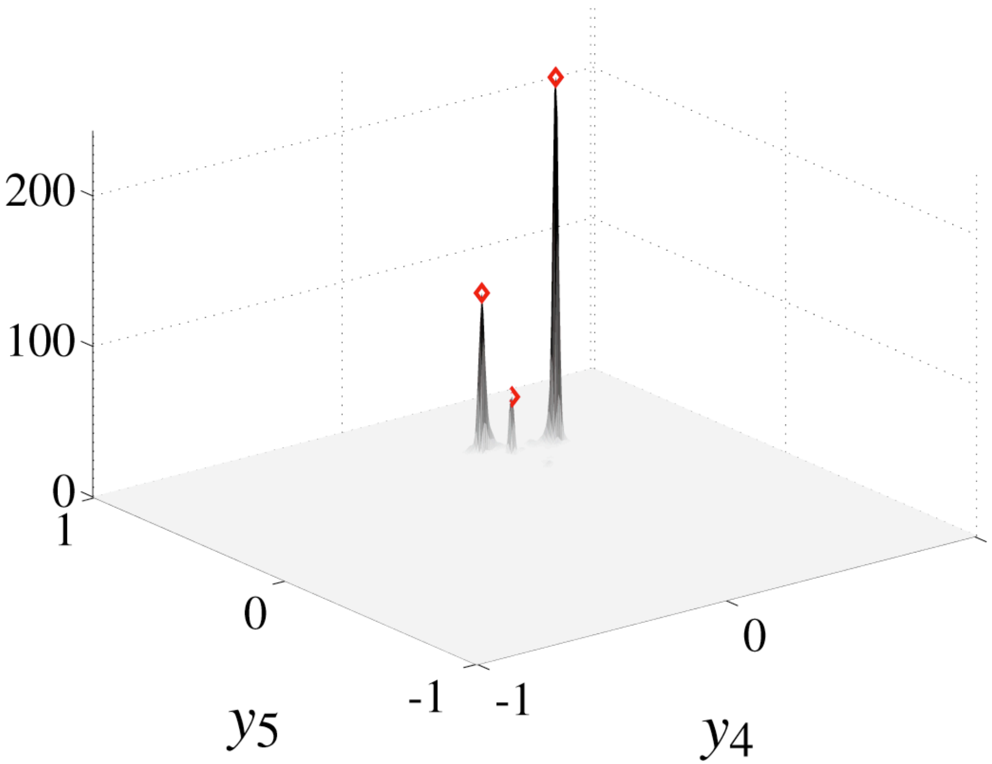
An example in the case of $(M, N) = (5, 5)$.

$v_1(t, \alpha) = (y_1, \ldots, y_5)^{\mathtt{T}}$.

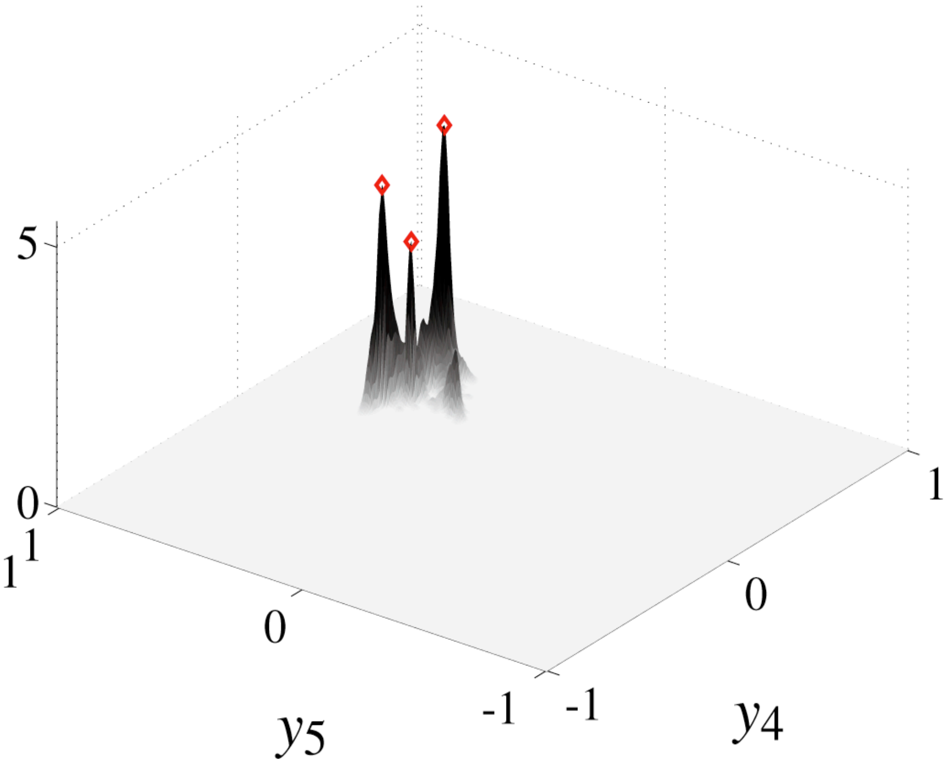
Take a two-dimensional projective histogram for (y_4, y_5).

We cannot select five clusters (peaks).

Best histogram No.10



Best histogram No.10 (log scale)



Observation and new strategies:

Reduction procedure:

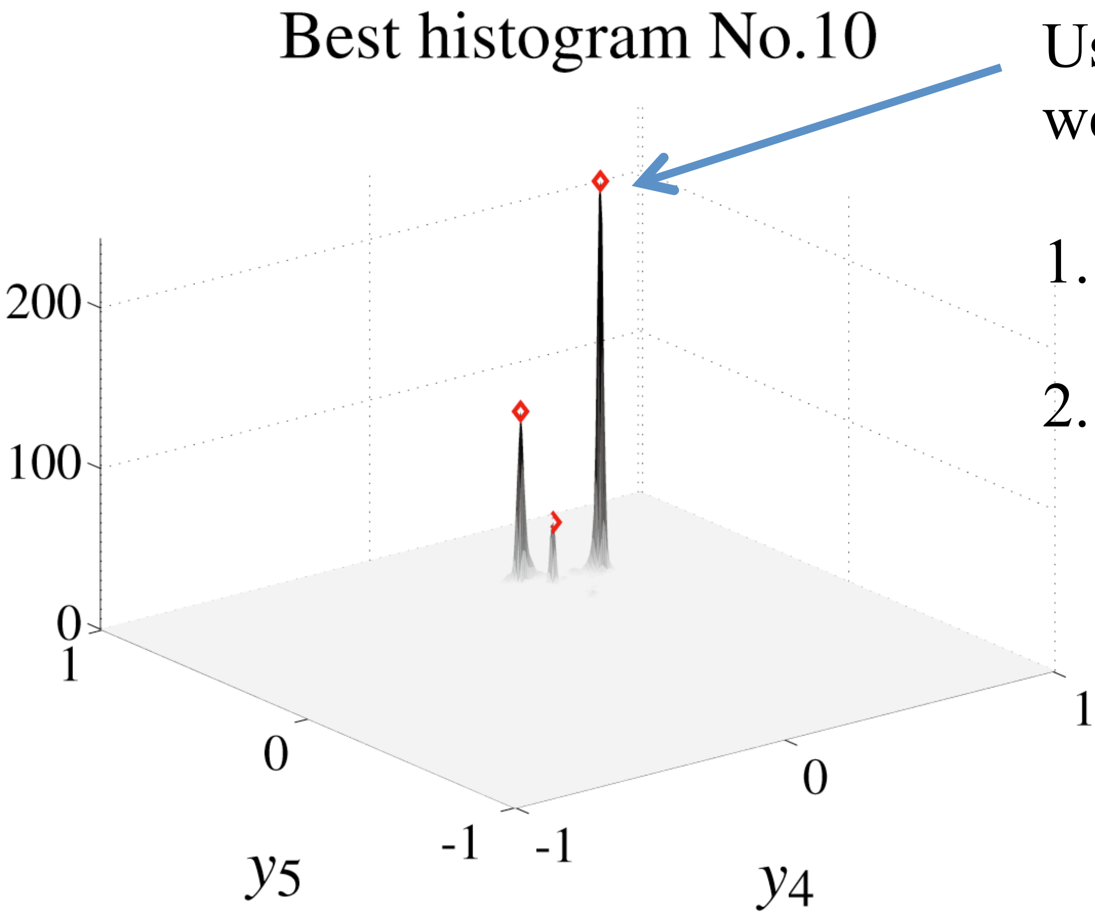
Our previous method can have at least the most crowded cluster, that is, a candidate of a column vector of the mixing matrix A.

Then, Gaussian elimination can remove the the most crowded cluster and its corresponding source.

This will give another image separation problem with new source images (unknown mixture of original source images which are not removed) whose number is reduced by one.

Continue the reduction procedure until we have only one source image, which coincides with one of the original images up to constant multiple. The number of reduction steps coincides with the number of sources minus one.

Source reduction method (new method)



Using Gaussian elimination, we can

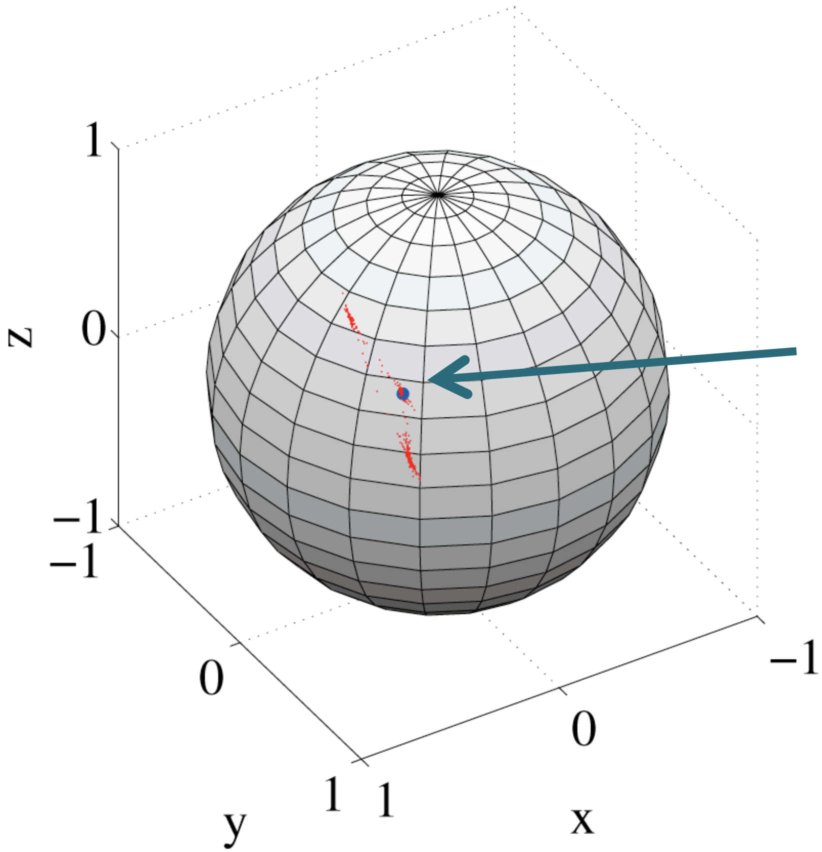
1. remove one source signal,

2. make a new set of observed signals which do not contain the source signal.

Separation procedure:

Backward substitution and solving new image separation problem at each step will solve the original separation problem.

Reduction procedure



In the case of $M = 3$.

Select the most crowded cluster.

Pick up the representative of this cluster.

$$b^{(0)} = (0.5609, \boxed{0.62537}, 0.5424)^{\mathrm{T}} \in \mathbb{S}^2$$

Then, our mixing model is

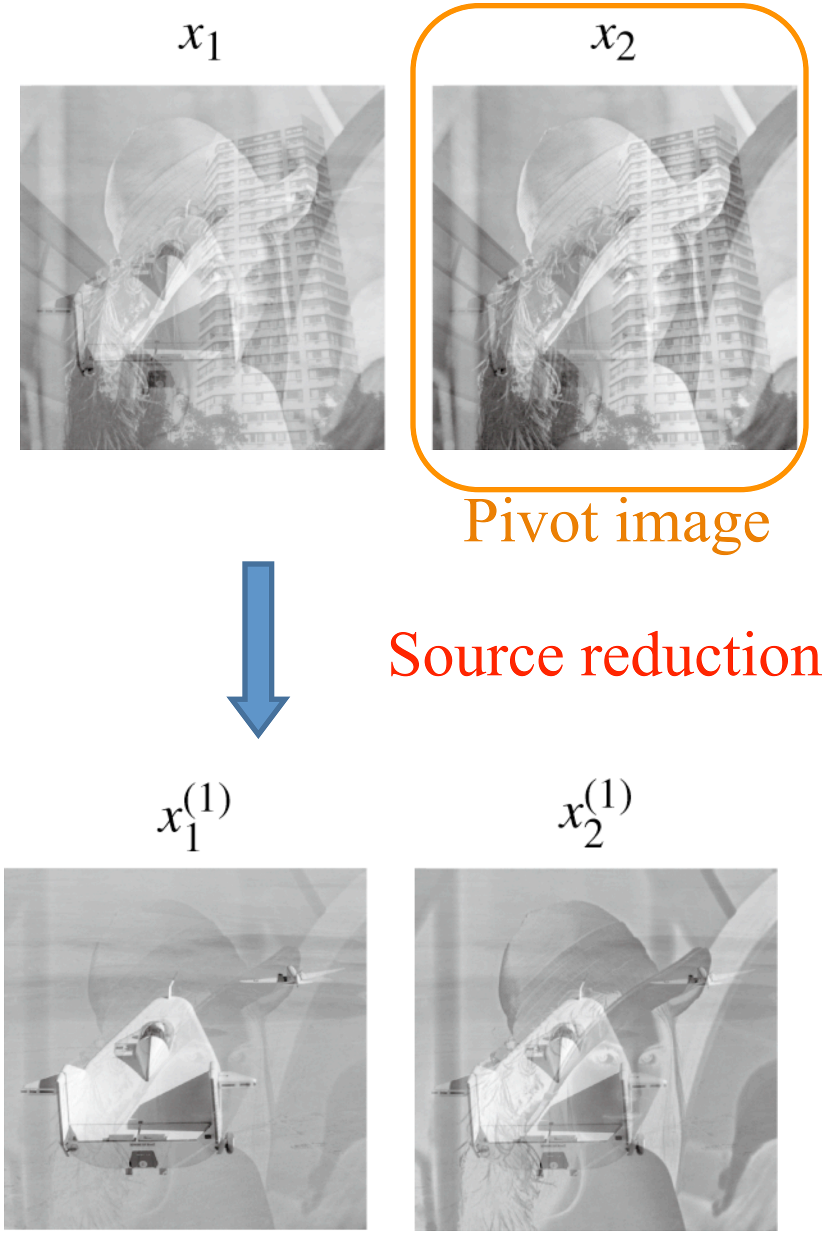
$$x(t) = (b^{(0)}, *, *)s(t)$$

Choose $b_2^{(0)}$ as the pivot.

Apply Gaussian elimination.

$$x(t) = (x_1(t), x_2(t), x_3(t))^{\mathrm{T}}$$

$$\boxed{\bar{x}_j(t) = x_j(t) - \frac{b_j^{(0)}}{b_2^{(0)}} x_2(t), \quad j \neq 2}$$

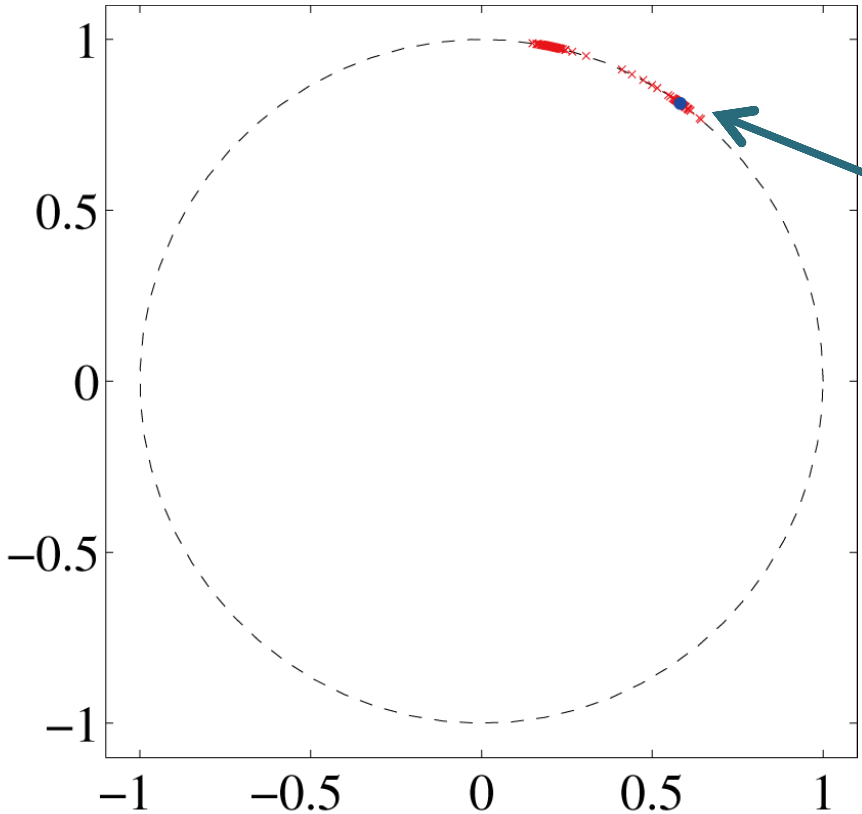


Renumber $\bar{x}_j$.

New observed signals

$$x^{(1)}(t) = (x_1^{(1)}(t), x_2^{(1)}(t))^{\mathsf{T}}$$

which do not contain $s_1(t)$.



For $x^{(1)}(t) = (x_1^{(1)}(t), x_2^{(1)}(t))^{\mathtt{T}}$,

Select the most crowded cluster.

Pick up the representative of this cluster.

$$b^{(1)} = (0.58203, \boxed{0.81308})^{\mathtt{T}} \in \mathbb{S}^1$$

Then, our new mixing model is

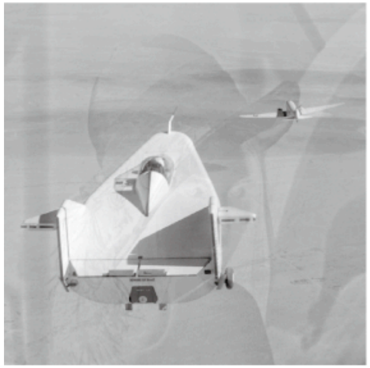
$$x^{(1)}(t) = (b^{(1)}, *)s^{(1)}(t)$$

Choose $b_2^{(1)}$ as the pivot.

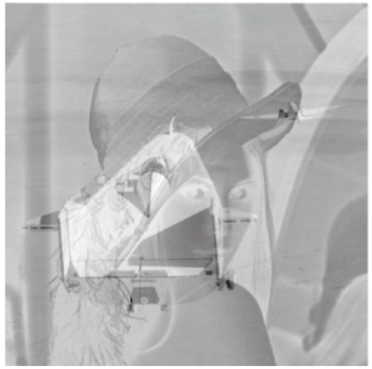
Apply Gaussian elimination.

$$\boxed{x_1^{(2)}(t) = x_1^{(1)}(t) - \frac{b_1^{(1)}}{b_2^{(1)}} x_2^{(1)}(t)}$$

$x_1^{(1)}$



$x_2^{(1)}$



Pivot image

Source reduction

$x_1^{(2)}$

New observed signal

$$x^{(2)}(t) = (x_1^{(2)}(t))^{\mathsf{T}}$$

which does not contain $s_2(t)$.

New observed image $x_1^{(2)}(t)$
seems the source image by human observation.

The number of sources $\widetilde{N}$ is estimated as $\widetilde{N} = 3$.

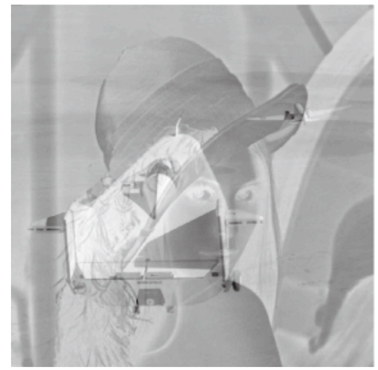
Separation procedure

$\boxed{\text{Set the estimated source image } \widetilde{s}_3(t) = x_1^{(2)}(t).}$

Consider the blind source separation problem : $\{x_2^{(1)}(t), \widetilde{s}_3(t)\}$.
This mixing model is

$$\begin{pmatrix} x_2^{(1)}(t) \\ \widetilde{s}_3(t) \end{pmatrix} = \begin{pmatrix} 1 & c_{1,2}^{(1)} \\ 0 & c_{2,2}^{(1)} \end{pmatrix} \begin{pmatrix} s_2(t) \\ \widetilde{s}_3(t)/c_{2,2}^{(1)} \end{pmatrix}.$$

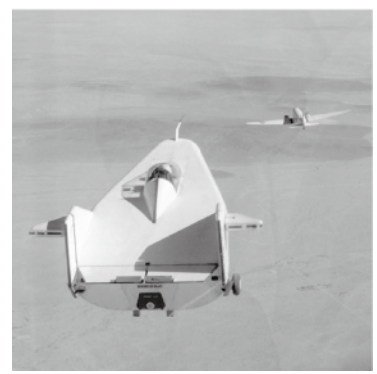
$x_2^{(1)}$



Pivot image

$\tilde{s}_3$

$\tilde{s}_2$

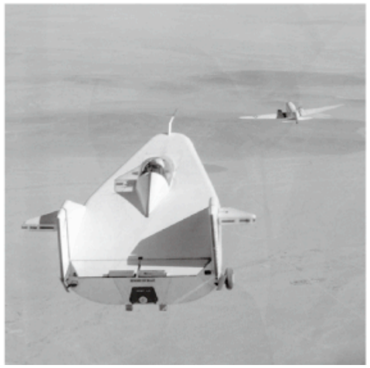


Since we know the form of mixing matrix, we can easily separate source images.

x_2

Pivot image

$\tilde{s}_2$



$\tilde{s}_3$

Consider the blind source separation problem : $\{x_2(t), \widetilde{s}_2(t), \widetilde{s}_3(t)\}$. This mixing model is

$$\begin{pmatrix} x_2(t) \\ \widetilde{s}_2(t) \\ \widetilde{s}_3(t) \end{pmatrix} = \begin{pmatrix} 1 & c_{1,2}^{(0)} & c_{1,3}^{(0)} \\ 0 & c_{2,2}^{(0)} & 0 \\ 0 & 0 & c_{3,3}^{(0)} \end{pmatrix} \begin{pmatrix} s_1(t) \\ \widetilde{s}_2(t)/c_{2,2}^{(0)} \\ \widetilde{s}_3(t)/c_{3,3}^{(0)} \end{pmatrix}.$$

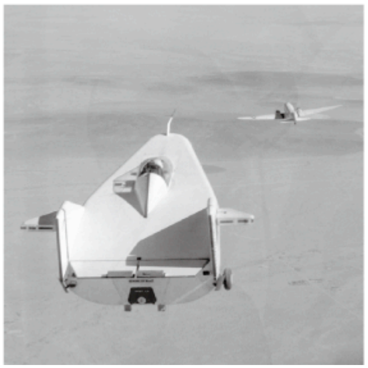
$\tilde{s}_1$

Error estimation

ES	S	max error (%)	ℓ^1-error (%)	ℓ^2-error (%)	SNR (dB)
$\widetilde{s}_3$	s_1	0.7410	0.6307	0.5931	44.54
$\widetilde{s}_2$	s_3	2.298	2.072	2.011	33.93
$\widetilde{s}_1$	s_2	0.6641	0.8365	0.6146	44.23

$\tilde{s}_3$

$\tilde{s}_2$



$\tilde{s}_1$

Results of numerical experiments

1. $(M, N) = (5, 5), (6, 5), (6, 6), (7, 6), (8, 7), (9, 8), (10, 9)$.
2. Two types of original sources $s(t)$:
   (a) 7.4 seconds speech signals with sampling rate 8820 Hz.
   (b) 512×512 gray scale standard images.

For each combination of (M, N) and $s(t)$, we generated twenty random mixing matrices A uniformly distributed in $[0.2, 0.8]$.

The results are the following:

1. $\boxed{\text{The proposed method performed well for all the cases listed above.}}$
2. Most of the above cases could not be separated by our previous separation method.

Conclusion

Our previous estimation methods had a difficulty when several peaks are quite large comparing to others.

We overcame this difficulty by a new source reduction method, which removes the largest peak step by step by applying Gaussian elimination.

The computation cost of our new method is expensive.

Various numerical experiments demonstrate that our new method performs well for both speech and image separations.

Future work

Find real applications whose mixing model is discussed here.

ICA is widely used in real applications especially in medical images. Compare our method to ICA in some of them.

Thank you for your attention.