ICASSP 2018, Calgary, Canada
Tutorial T1
(April 16th, 2018)

Blind Audio Source Separation on Tensor Representation

NTT
TMU
THE UNIVERSITY OF TOKYO

Hiroshi Sawada (NTT Corporation)
Nobutaka Ono (Tokyo Metropolitan University)
Hirokazu Kameoka (NTT Corporation)
Daichi Kitamura (The University of Tokyo)

1

2

Tutorial structure

1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

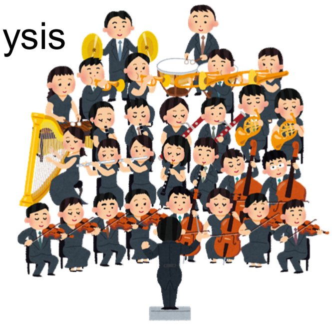
3

Separation of audio/speech signal

cocktail party effect

speech recognition
in noisy environment

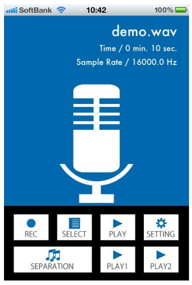
music analysis



4

Live demonstration

- Separate 2 speeches with 2 microphones
- iPhone app



- Script

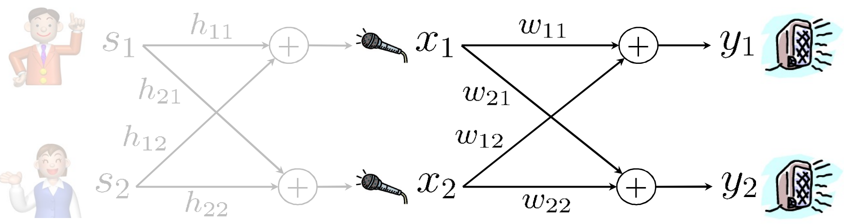
The ICASSP meeting is the world's largest and most comprehensive technical conference focused on signal processing and its applications.

We are demonstrating the Blind Source Separation for convolutive mixtures of speech in a real room with real talkers.

5

BSS: Blind Source Separation

- Separate the mixtures at microphones x_1, x_2
  - into the original sources s_1, s_2
    - assuming M=N=2 for simplicity
  - source activity and mixing system **H** is unknown (blind)

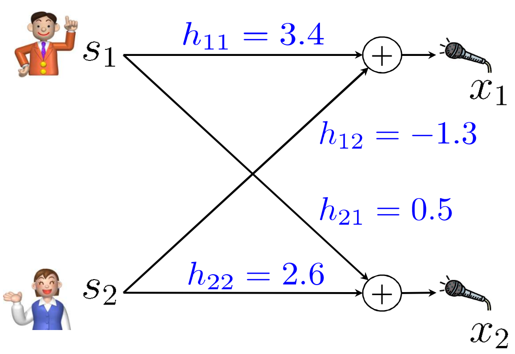


6

Instantaneous BSS

- Mixing system **H** is described by scalars
- Sources are multiplied by scalars and then mixed

$$x_m(t) = \sum_{n=1}^{N} h_{mn} s_n(t)$$

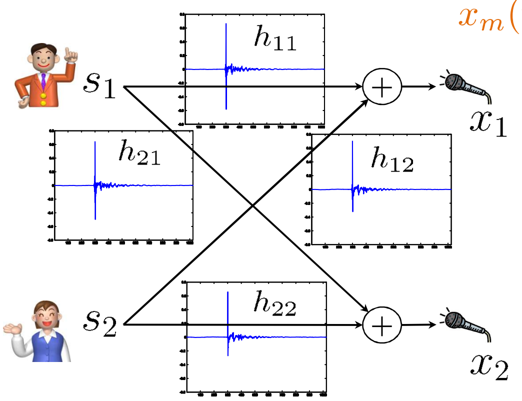


Simple and basic
BSS model

Convolutive BSS

- Delay and reverberations in a real room situation
  - mixing system **H** is described by impulse responses
- Sources are convolutively mixed

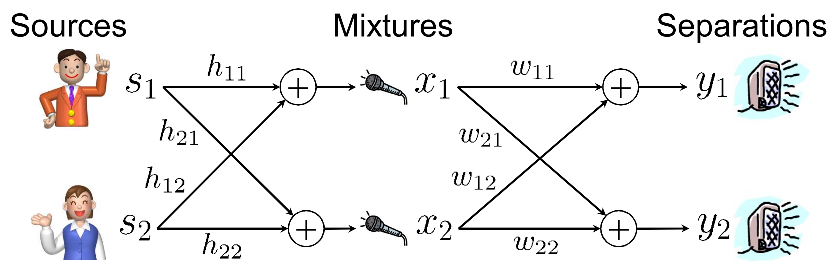
$$x_m(t) = \sum_{n=1}^{N} \sum_{\tau=0}^{L-1} h_{mn}(\tau) s_n(t-\tau)$$



Convolutive BSS is a much harder problem than instantaneous BSS

$$x_m(t) = \sum_{n=1}^{N} h_{mn} s_n(t)$$

The whole system



	Mixing system	**Separation system**
Convolution filter (time domain)	$x_m(t) = \sum_{n=1}^{N} \sum_{\tau=0}^{L-1} h_{mn}(\tau) s_n(t-\tau)$	$y_n(t) = \sum_{m=1}^{M} \sum_{\tau=0}^{L-1} w_{nm}(\tau) x_m(t-\tau)$
Frequency response (frequency domain)	$x_m(i,j) = \sum_{n=1}^{N} h_{mn}(i) s_n(i,j)$	$y_n(i,j) = \sum_{m=1}^{M} w_{nm}(i) x_m(i,j)$

STFT — STFT: short-time Fourier transform

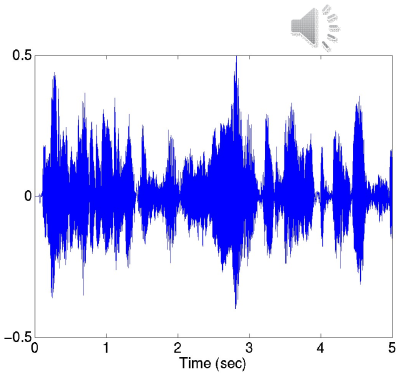
convolution → multiplication

9

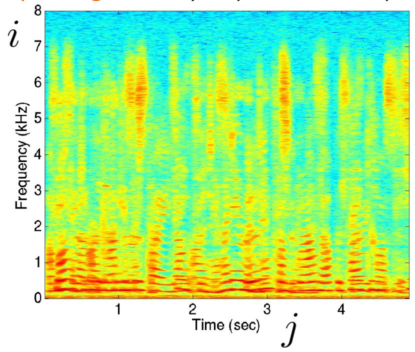
STFT: short-time Fourier transform

- From a time-domain real-valued signal
- To a time-frequency-domain complex-valued signal

$x(t) \in \mathbb{R}$ STFT $x(i, j) \in \mathbb{C}$



Spectrogram: only amplitudes are displayed



10

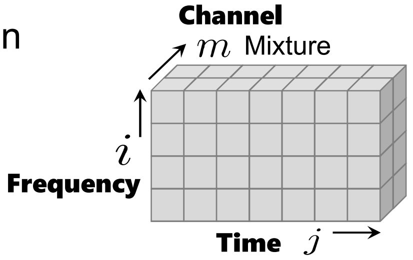
Three key axes

Frequency
- Frequency-domain processing is effective
  - source characteristic
  - convolution → multiplication

Time
- Source activity, Onset and offset

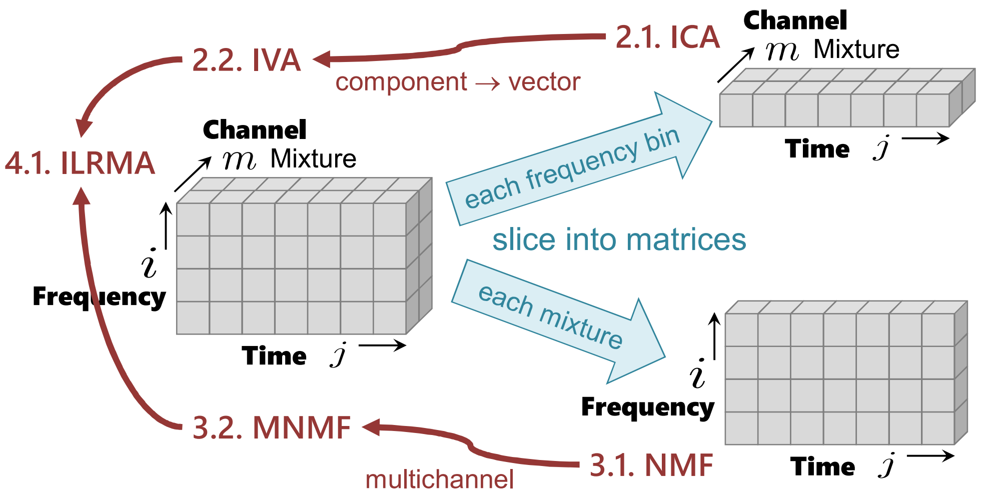
Channel
- Source, Mixture, Separation

Tensor representation



11

Tensor and sliced matrices



12

Notations

s: sources
x: mixtures/observations
y: separations

H: mixing system
W: separation system

i: frequency bin index
j: time frame index
m: microphone index
n: source/separation index

I: number of frequency bins
J: number of time frames
M: number of microphones
N: number of sources

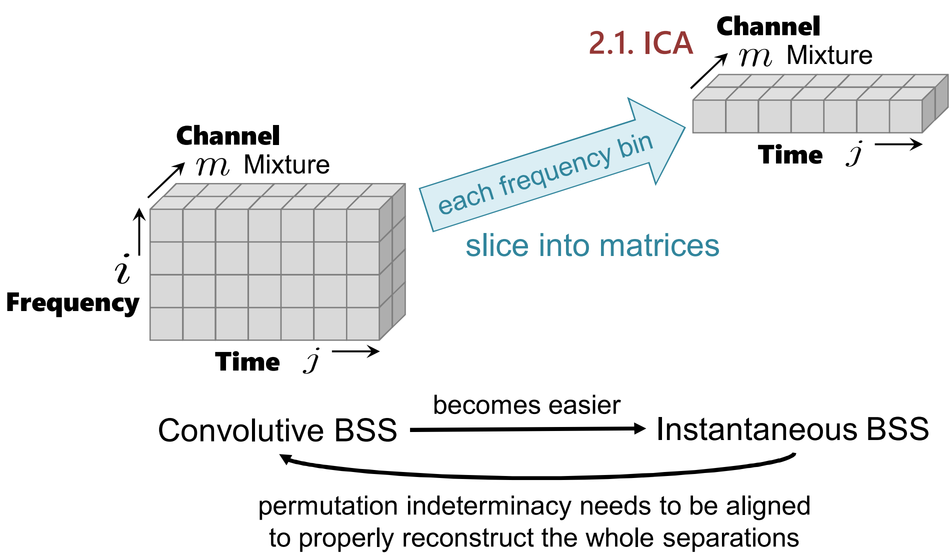
13

Tutorial structure

1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

14

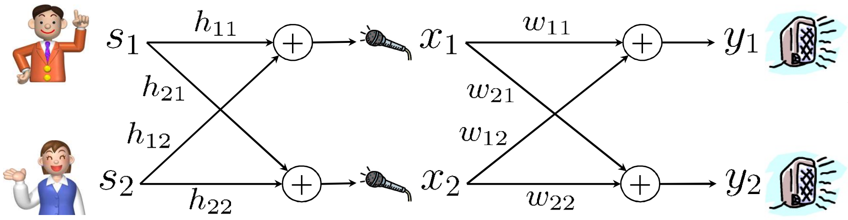
Tensor and sliced matrices



15

ICA: Independent Component Analysis

- Extract the original sources from the mixtures $x_1,\ x_2$
- Mixing matrix **H** cannot be obtained

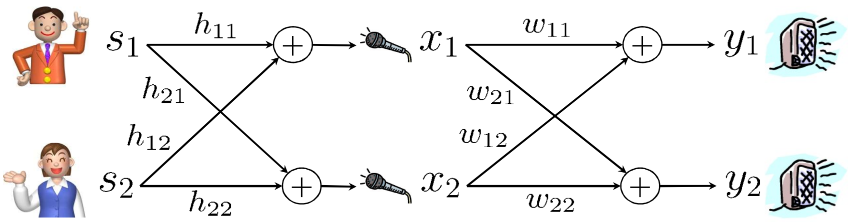


Sources are assumed to be independent → Makes y_1, y_2 independent to each other

16

ICA: Independent Component Analysis

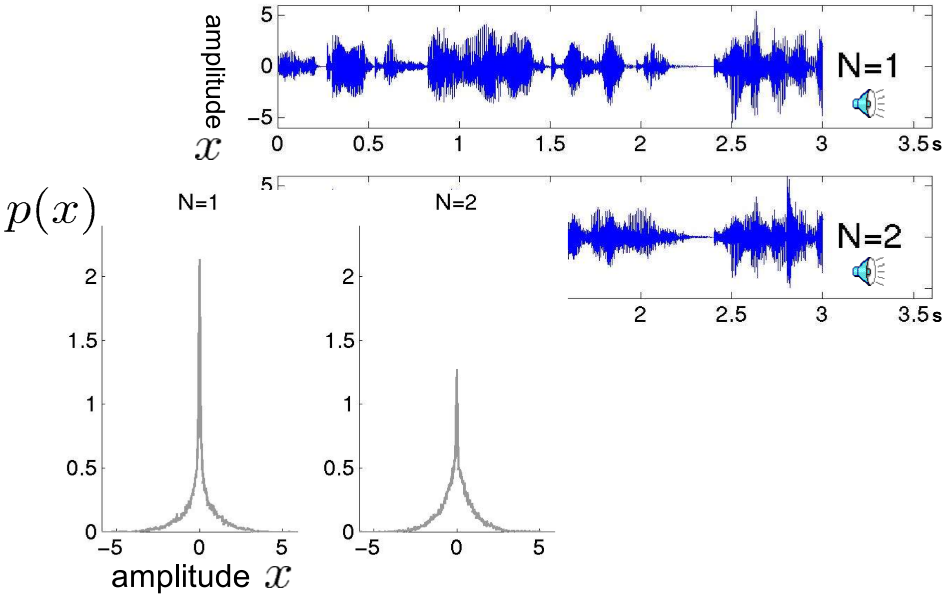
- In addition to independence, need to assume
  - source distributions are different from Gaussian



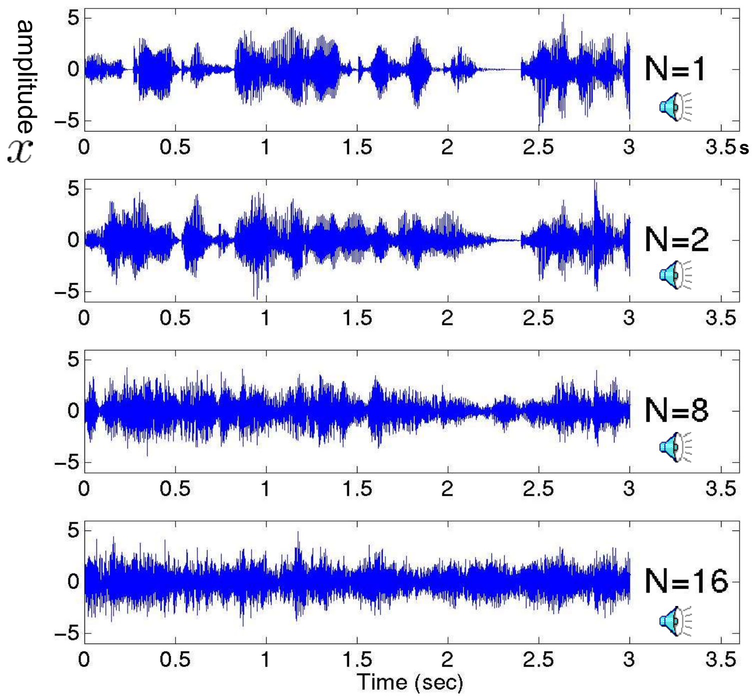
Distributions are different from Gaussian → Makes the distribution far from Gaussian

17

Speech mixtures

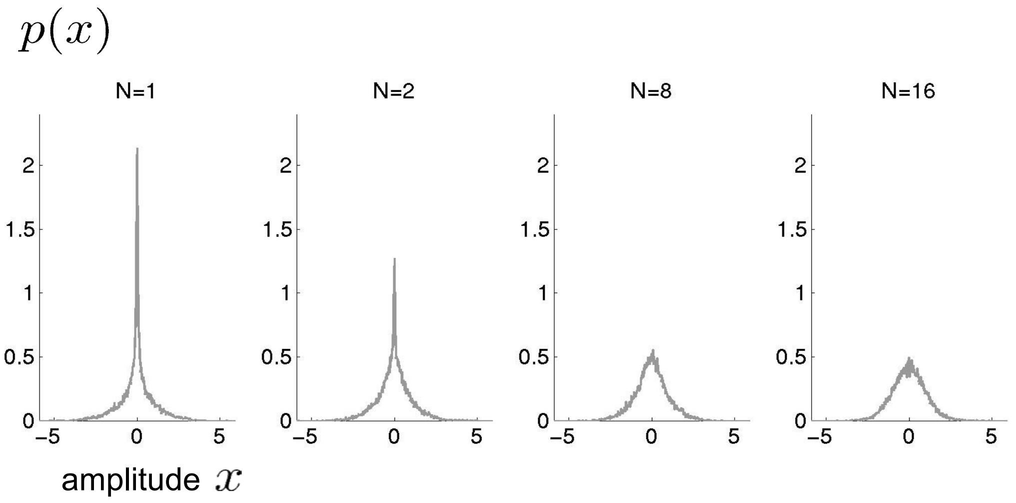


18



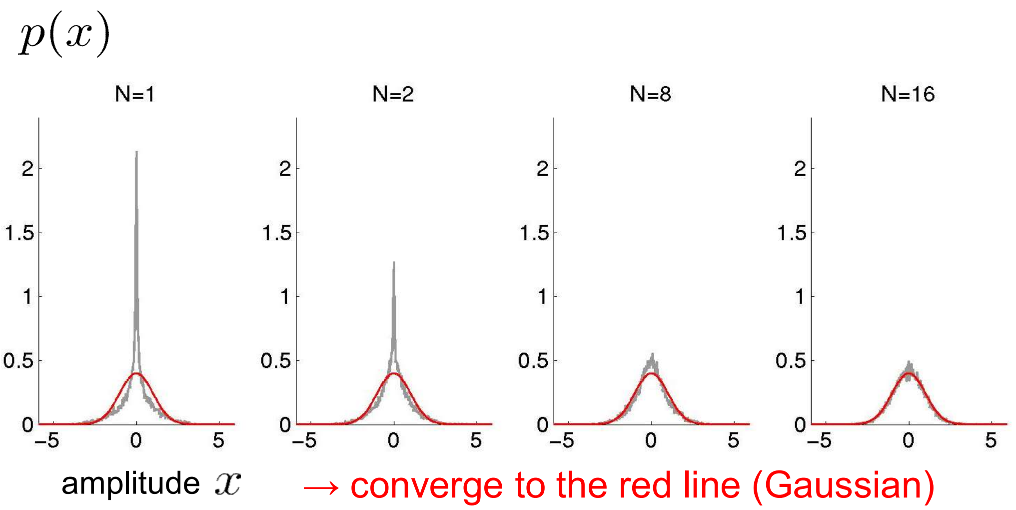
19

Statistics of the mixtures



20

Statistics of the mixtures



→ converge to the red line (Gaussian)

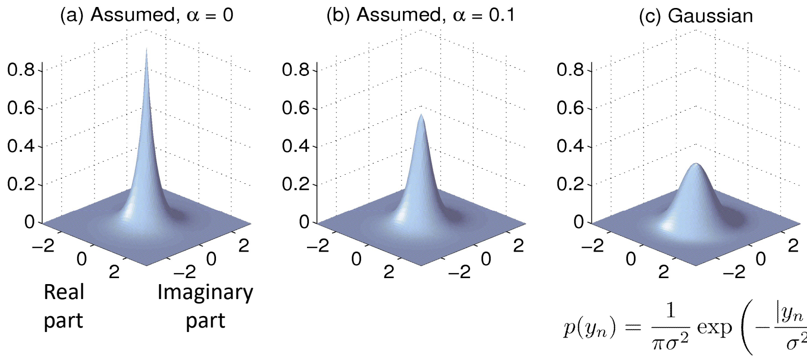
Central Limit Theorem

A complex-valued source model

- Super-Gaussian distribution

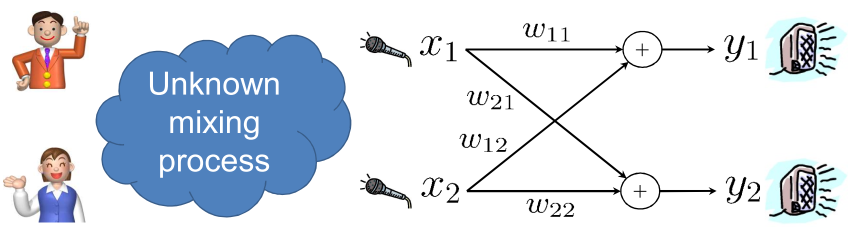
$$p(y_n) \propto \exp\left(-\frac{\sqrt{|y_n|^2 + \alpha}}{b}\right) \qquad y_n \in \mathbb{C}$$

- Sharper peak at the origin than Gaussian



$$p(y_n) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|y_n|^2}{\sigma^2}\right)$$

Independent component analysis



- Linear operation $\quad \mathbf{y}(j) = \mathbf{W}\mathbf{x}(j)$
- Output independence $\quad p(\mathbf{y}) = \prod_{n=1}^{N} p(y_n)$
- Non-gaussianity $\quad p(y_n) \neq \frac{1}{\pi\sigma^2}\exp\left(-\frac{|y_n|^2}{\sigma^2}\right)$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

23

Likelihood of separation matrix **W**

- Likelihood of **W** for the whole observations

$$p(\mathcal{X}|\mathbf{W}) = \prod_{j=1}^{J} p(\mathbf{x}(j)|\mathbf{W}) \qquad \mathcal{X} = \{\mathbf{x}(1), \ldots, \mathbf{x}(J)\}$$

- Probability density function, linear transformation

$$p(\mathbf{x}|\mathbf{W}) = |\det \mathbf{W}|^2 \, p(\mathbf{y}) \qquad \leftarrow \quad \mathbf{y}(j) = \mathbf{W}\mathbf{x}(j)$$

- Output independence

$$p(\mathbf{y}) = \prod_{n=1}^{N} p(y_n)$$

Log-likelihood function

$$\log p(\mathcal{X}|\mathbf{W}) = 2J \cdot \log|\det \mathbf{W}| + \sum_{j=1}^{J} \sum_{n=1}^{N} \log p(y_n(j))$$

24

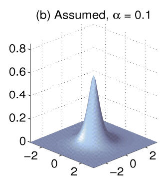
Objective function to be minimized

Maximum likelihood estimation = Minimize the negative log-likelihood

$$\mathcal{J}(\mathbf{W}) = J\left[\sum_{n=1}^{N} \mathrm{E}\{G(y_n)\} - 2\log|\det \mathbf{W}|\right]$$

Contrast function

$$p(y_n) \propto \exp\left(-\frac{\sqrt{|y_n|^2+\alpha}}{b}\right)$$



$$G(y_n) = -\log p(y_n) \qquad G(y_n) = \sqrt{|y_n|^2+\alpha}$$

1st order derivative

$$g(y_n) = \frac{\partial G(y_n)}{\partial y_n^*} = -\frac{\partial \log p(y_n)}{\partial y_n^*} \qquad g(y_n) = \frac{y_n}{2\sqrt{|y_n|^2+\alpha}}$$

2nd order derivative

$$g'(y_n) = \frac{\partial^2 G}{\partial y_n^* \partial y_n} = \frac{\partial g}{\partial y_n} \qquad g'(y_n) = \frac{1}{2\sqrt{|y_n|^2+\alpha}}\left[1 - \frac{1}{2}\frac{|y_n|^2}{|y_n|^2+\alpha}\right]$$

Optimization methods

- Gradient descent

$$\frac{\partial 2\log|\det \mathbf{W}|}{\partial \mathbf{W}^*} = (\mathbf{W}^{\mathsf{H}})^{-1}$$

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \cdot \frac{1}{J}\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}$$ η :step size

$$\frac{1}{J}\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*} = \mathrm{E}\{\mathbf{g}(\mathbf{y})\mathbf{x}^{\mathsf{H}}\} - (\mathbf{W}^{\mathsf{H}})^{-1}$$

$$\mathbf{g}(\mathbf{y}) = \begin{bmatrix} g(y_1) \\ \vdots \\ g(y_N) \end{bmatrix}$$

  - Expensive matrix inversion
  - Slow convergence

$$g(y_n) = \frac{y_n}{2\sqrt{|y_n|^2 + \alpha}}$$

- Three practical ways
  - **Natural gradient**
  - **Pre-whitening + FastICA**
  - **Auxiliary function-based optimization**

Natural gradient

[Amari et al., 1996]

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \cdot \frac{1}{J}\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}\mathbf{W}^{\mathsf{H}}\mathbf{W}$$

$$\frac{1}{J}\frac{\partial \mathcal{J}}{\partial \mathbf{W}^*}\mathbf{W}^{\mathsf{H}}\mathbf{W} = \left[\mathrm{E}\{\mathbf{g}(\mathbf{y})\mathbf{y}^{\mathsf{H}}\} - \mathbf{I}\right] \cdot \mathbf{W}$$

$$g(y_n) = \frac{y_n}{2\sqrt{|y_n|^2 + \alpha}}$$

- No matrix inversion
  - Efficient computation
- Equivariance property [Cardoso and Souloumiac, 1996]
  - Free from the characteristics of mixing matrix (e.g. close to singular)

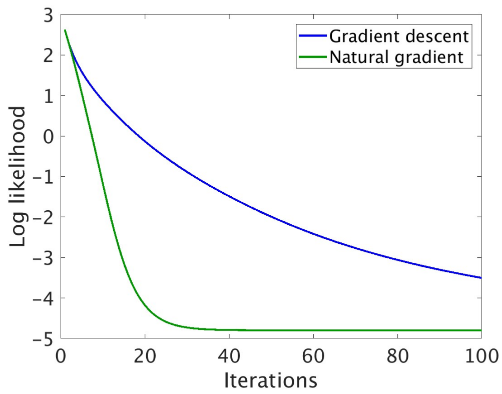
27

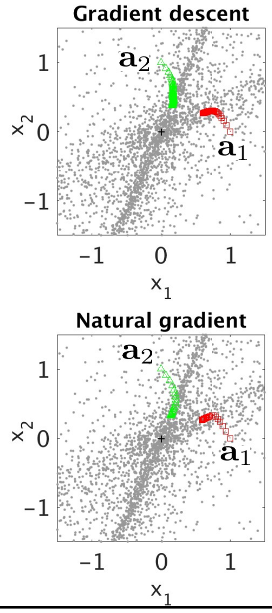
Convergence example

Estimated mixing matrix $\mathbf{A} = \mathbf{W}^{-1}$ ➡

Starting from $\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Step size: $\eta = 0.1$

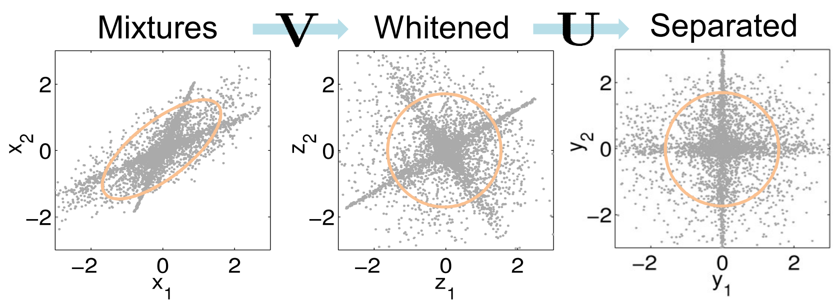




28

Pre-whitening ＋ FastICA

Separation matrix of the form: $\mathbf{W} = \mathbf{U}\mathbf{V}$ [Hyvarinen et al., 2001]



Pre-whitening

$\mathbf{z}(j) = \mathbf{V}\mathbf{x}(j)$ s. t. $\mathrm{E}\{\mathbf{z}\mathbf{z}^{\mathsf{H}}\} = \mathbf{I}$

via eigenvalue decomposition of $\mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathsf{H}}\}$

FastICA

$\mathbf{y}(j) = \mathbf{U}\mathbf{z}(j)$

Maximize the log-likelihood w.r.t. a unitary matrix **U**

FastICA

- Objective function w.r.t a unitary matrix **U**

$$\mathcal{J}(\mathbf{U}) = J\left[\sum_{n=1}^{N} \mathrm{E}\{G(y_n)\} - 2\log|\det \mathbf{U}|\right]$$
$$= J\sum_{n=1}^{N} \mathrm{E}\{G(y_n)\} \quad \Leftarrow \det\mathbf{U} = 1$$

- Minimize $\mathrm{E}\{G(y_n)\}$ for $y_1, y_2, \cdots, y_N$

$$G(y_n) = \sqrt{|y_n|^2 + \alpha}$$

  - with unitary constraint

$$\mathbf{u}_n^{\mathsf{H}}\mathbf{u}_k = \begin{cases} 1 & \text{if } n = k \\ 0 & \text{if } n \neq k \end{cases} \qquad \mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_N]^{\mathsf{H}}$$

FastICA algorithm

[Hyvarinen et al., 2001]

- For $n = 1, \ldots, N$ (sequentially)
  - Iterate the followings until convergence

Separated signal calculation

$$y_n(j) = \mathbf{u}_n^{\mathsf{H}}\mathbf{z}(j)$$

Optimization of G by Newton's method

$$\mathbf{u}_n \leftarrow \mathrm{E}\{g(y_n)\mathbf{z}\} - \mathrm{E}\{g'(y_n)\}\mathbf{u}_n$$

$$g(y_n) = \frac{y_n}{2\sqrt{|y_n|^2 + \alpha}} \qquad g'(y_n) = \frac{1}{2\sqrt{|y_n|^2 + \alpha}}\left[1 - \frac{1}{2}\frac{|y_n|^2}{|y_n|^2 + \alpha}\right]$$

Gram-schmidt orthogonalization

$$\mathbf{u}_n \leftarrow \mathbf{u}_n - \sum_{k=1}^{n-1}(\mathbf{u}_k^{\mathsf{H}}\mathbf{u}_n)\mathbf{u}_k$$

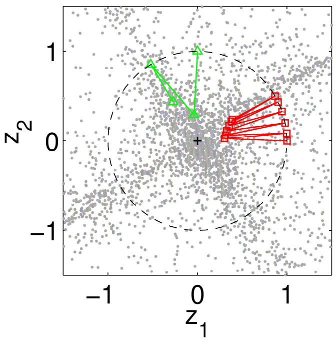
Unit-norm normalization

$$\mathbf{u}_n \leftarrow \frac{\mathbf{u}_n}{||\mathbf{u}_n||}$$

31

FastICA convergence example

- Red (□)
  - Starting from $\mathbf{u}_1 = [1 \;\; 0]^\mathrm{T}$
  - Likelihood maximization: points close to the origin
  - Unit-norm normalization: points on the unit sphere
  - Good solution only by 5 iterations
- Green (△)
  - Starting from $\mathbf{u}_2 = [0 \;\; 1]^\mathrm{T}$
  - One-step solution by orthogonalization



32

Auxiliary function

[Ono and Miyabe, 2010]
[Ono 2011]

- Objective function to be minimized

$$\mathcal{J}(\mathbf{W}) = J\left[\sum_{n=1}^{N} \mathrm{E}\{G(|y_n|)\} - 2\log|\det \mathbf{W}|\right]$$

- If $G(y_n) = G_R(|y_n|)$ and

  $\frac{G'_R(|y_n|)}{|y_n|}$ is monotonically decreasing

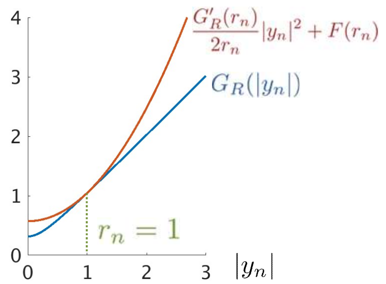
$$G(y_n) = \sqrt{|y_n|^2 + \alpha}$$
$$G_R(|y_n|) = \sqrt{|y_n|^2 + \alpha}$$
$$G'_R(|y_n|) = \frac{\partial G_R}{\partial |y_n|} = \frac{|y_n|}{\sqrt{|y_n|^2 + \alpha}}$$
$$\frac{G'_R(|y_n|)}{|y_n|} = \frac{1}{\sqrt{|y_n|^2 + \alpha}}$$

- Auxiliary function

$$G(y_n) \leq \frac{G'_R(r_n)}{2r_n}|y_n|^2 + F(r_n)$$

  with auxiliary variable r_n

  Equal when $r_n = |y_n|$



Auxiliary function-based optimization

- Iterate the followings until convergence

[Ono and Miyabe, 2010]
[Ono 2011]

Separated signal calculation

$$\mathbf{y} = \mathbf{W}\mathbf{x} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} \mathbf{w}_1^{\mathsf{H}} \\ \vdots \\ \mathbf{w}_N^{\mathsf{H}} \end{bmatrix}$$

Weighted covariance matrices for each separation

$$\mathbf{V}_n = \mathrm{E}\left\{ \frac{G'_R(|y_n|)}{2|y_n|} \mathbf{x}\mathbf{x}^{\mathsf{H}} \right\} \qquad \frac{G'_R(|y_n|)}{|y_n|} = \frac{1}{\sqrt{|y_n|^2 + \alpha}}$$

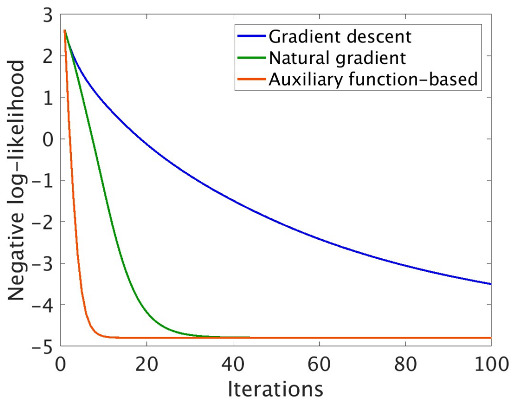
Solve the HEAD problem for $\mathbf{V}_1, \ldots, \mathbf{V}_N$

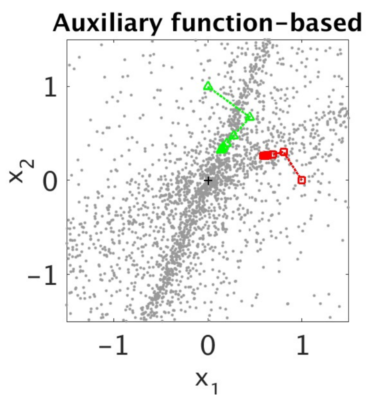
$$\mathbf{w}_m^{\mathsf{H}} \mathbf{V}_n \mathbf{w}_n = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}$$

Auxiliary ICA convergence example

Estimated mixing matrix $\mathbf{A} = \mathbf{W}^{-1}$

Starting from $\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$





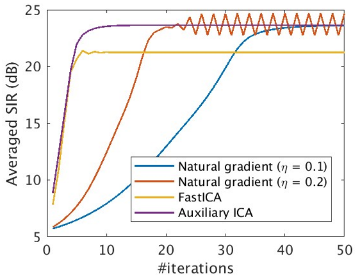
able to take a big step at an early stage

35

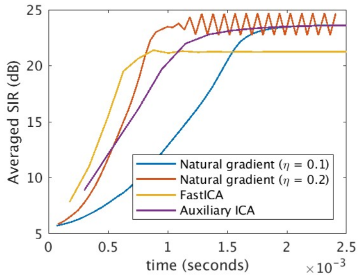
Comparisons of three methods

- Separation of 6-second 3 speeches with 3 microphones
  - measured by signal-to-interference ratio (SIR)
    - at frequency 3586 Hz, 201 samples, pre-whitening applied

convergence behavior



computational cost considered



✓ Natural gradient: sensitive to step-size η

✓ FastICA: fast convergence, slightly limited SIR (unitary constraint)

✓ Auxiliary ICA: fast convergence

36

Permutation and Scaling problem

- Ambiguities of ICA solutions

If $\mathbf{y}(j) = \mathbf{W}\,\mathbf{x}(j)$ is a solution, then

$\mathbf{y}(j) \leftarrow \boldsymbol{\Lambda}\,\mathbf{P}\,\mathbf{y}(j)$ is also a solution

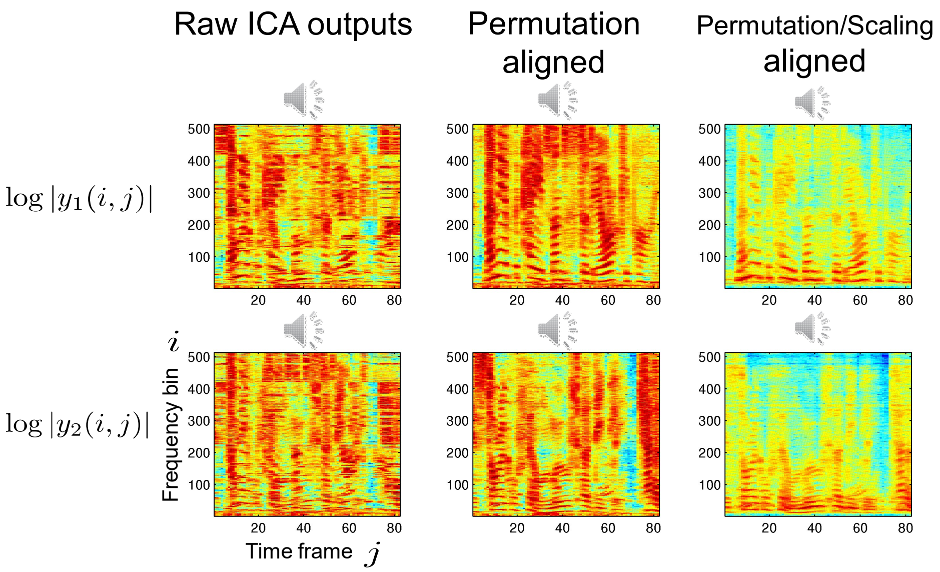
for any diagonal $\boldsymbol{\Lambda}$ and permutation $\mathbf{P}$ matrix

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \leftarrow \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Independence of y_1, y_2, y_3 does not change

37

Permutation and Scaling problem



38

Solving Permutation and Scaling Problems

- **Permutation**
  - Post-processing
    - see e.g. [Sawada et al., 2004], [Sawada et al., 2011]
  - Tensor methods (IVA, ILRMA)
    - will be explained in later sections
- **Scaling**
  - Refer to a microphone observation
    - So-called "projection back"
    - via mixing system estimation

[Cardoso 1998]
[Murata et al., 2001]
[Matsuoka and Nakashima 2001]
[Takatani et al., 2004]

39

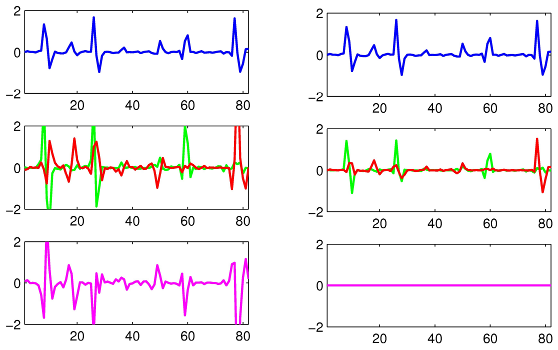
Scaling alignment example

Untouched | **Aligned** $y_n(j) \leftarrow a_{1n} y_n(j)$ | reference microphone $m = 1$

x_1

y_1, y_2

$x_1 - (y_1 + y_2)$



a_{1n} estimated mixing scalar

40

Mixing system estimation

Estimated mixing situation

$$\mathbf{x} = \sum_{n=1}^{N} \mathbf{a}_n y_n = \mathbf{A}\mathbf{y}$$

ICA result

$$\mathbf{y} = \mathbf{W}\,\mathbf{x}$$

$$\mathbf{a}_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{Mn} \end{bmatrix} \quad \mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}$$

- How to calculate matrix $\mathbf{A}$
  - If $\mathbf{W}$ has an inverse
    $\mathbf{A} = \mathbf{W}^{-1}$
  - Otherwise ($N < M$)
    $\mathbf{A} = \mathrm{E}\{\mathbf{x}\mathbf{y}^H\}(\mathrm{E}\{\mathbf{y}\mathbf{y}^H\})^{-1}$
    $\mathbf{A} = \mathbf{W}^{+}$

• Least-mean-square estimator that minimizes $\mathrm{E}\{||\mathbf{x} - \mathbf{A}\mathbf{y}||^2\}$

• Moore-Penrose pseudo inverse

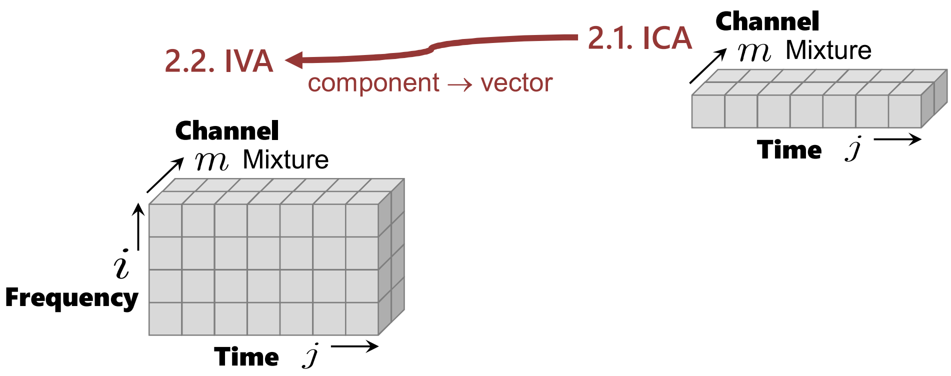
41

Tutorial structure

1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

42

Tensor and sliced matrices



43

Two papers at ICA 2006

Independent Vector Analysis: An Extension of ICA to Multivariate Components

Taesu Kim[1,2], Torbjørn Eltoft[3], and Te-Won Lee[1]

[1] Institute for Neural Computation, UCSD, USA
{taesu, tewon}@ucsd.edu
[2] Department of BioSystems, KAIST, Korea
[3] Department of Physics, University of Tromsø, Norway
torbjorn.eltoft@phys.uit.no

Abstract. In this paper, we solve an ICA problem where both source and observation signals are multivariate, thus, vectorized signals. To derive the algorithm, we define dependence between vectors as Kullback-Leibler divergence between joint probability and the product of marginal probabilities, and propose a vector density model that has a variance dependency within a source vector. The example shows that the algorithm successfully recovers the sources and it does not cause any permutation ambiguities within the sources. Finally, we propose the frequency domain blind source separation (BSS) for convolutive mixtures as an application of IVA, which separates 6 speeches with 6 microphones in a reverberant room environment.

Solution of Permutation Problem in Frequency Domain ICA, Using Multivariate Probability Density Functions

Atsuo Hiroe

Intelligent Systems Research Laboratory, Information Technologies Laboratories, Sony Corporation, 6-7-35 Kitashinagawa, Shinagawa-ku, Tokyo 141-0001, Japan
Atsuo.Hiroe@jp.sony.com

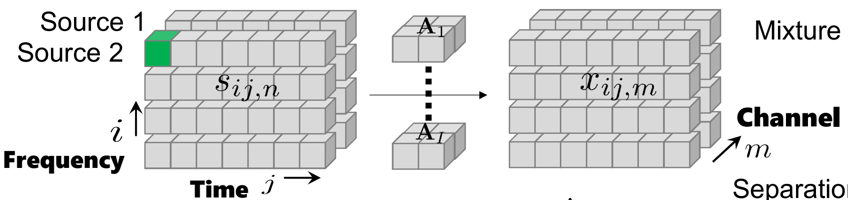
Abstract. Conventional Independent Component Analysis (ICA) in frequency domain inherently causes the permutation problem. To solve the problem fundamentally, we propose a new framework for separation of the whole spectrograms instead of the conventional binwise separation. Under our framework, a measure of independence is calculated from the whole spectrograms, not individual frequency bins. For the calculation, we introduce some multivariate probability density functions (PDFs) which take a spectrum as arguments. To seek the unmixing matrix that makes spectrograms independent, we demonstrate a gradient-based algorithm using multivariate activation functions derived from the PDFs. Through experiments using real sound data, we have confirmed that our framework is effective to generate permutation-free unmixed results.

Concept of "multivariate source model" was presented in two papers at the same conference independently

44

Independent vector analysis

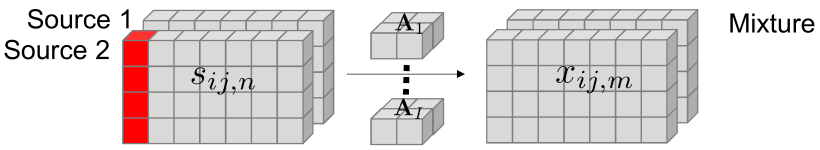
- ICA: Sources generate stochastic scalar variables



Mixing process: $\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij}$

Separation process: $\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij}$

- IVA: Sources generate stochastic vector variables



Same mixing / separation process as ICA, but multivariate source model

45

Likelihood of separation matrices **W**

- Likelihood of all **W**s for the whole observations

$$p(\mathcal{X}|\mathcal{W}) = \prod_{j=1}^{J} p(\mathbf{x}_{j,m}|\mathcal{W}) \quad \mathcal{X} = \{\mathbf{x}_{j,m}|j=1,\cdots,J, m=1,\cdots,N\} \quad \mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_I\}$$

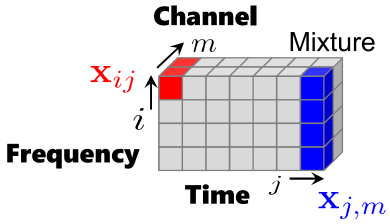
- Probability density function, linear transformation

$$p(\mathbf{x}_{j,1},\cdots,\mathbf{x}_{j,N}|\mathcal{W}) = \prod_{i=1}^{I} |\det \mathbf{W}_i|^2 p(\mathbf{y}_{j,1},\cdots,\mathbf{y}_{j,N})$$

$$\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij}$$

- Output independence

$$p(\mathbf{y}_{j,1},\cdots,\mathbf{y}_{j,N}) = \prod_{n=1}^{N} p(\mathbf{y}_{j,n})$$



Log-likelihood function

$$\mathcal{L} = \log p(\mathcal{X}|\mathbf{W}) = 2J\sum_{i=1}^{I} \log|\det \mathbf{W}_i| + \sum_{j=1}^{J}\sum_{n=1}^{N} \log p(\mathbf{y}_{j,n})$$

46

Objective function of IVA

- A set of demixing matrices to be estimated

$$\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_I\}$$

- Objective function of IVA

$$\mathcal{J}(\mathcal{W}) = -\mathcal{L}(\mathcal{W}) = \sum_{j=1}^{J}\sum_{n=1}^{N} G(\mathbf{y}_{j,n}) - 2J\sum_{i=1}^{I} \log|\det \mathbf{W}_i|$$

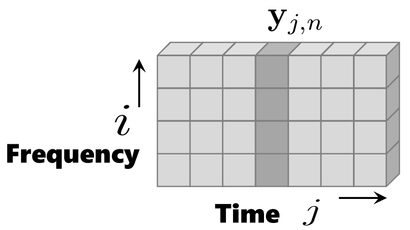
Contrast function

$$G(\mathbf{y}_{j,n}) = -\log p(\mathbf{y}_{j,n})$$

Multivariate p.d.f.

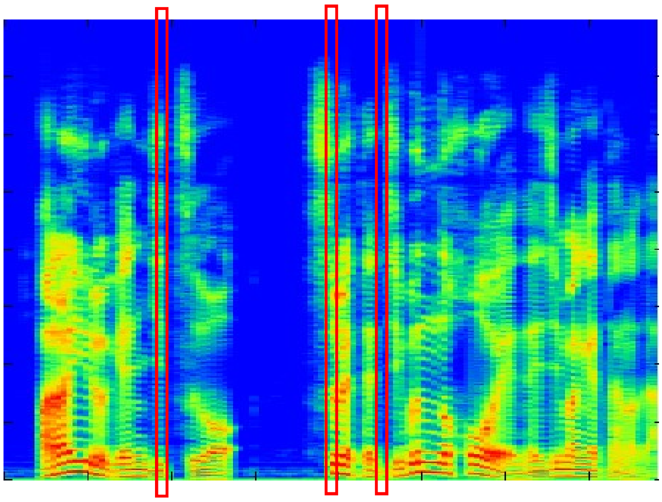
y is a function of w

$$\mathbf{y}_{j,n} = \begin{pmatrix} y_{1j,n} \\ y_{2j,n} \\ \vdots \\ y_{Ij,n} \end{pmatrix} \quad \begin{matrix} y_{1j,n} = \mathbf{w}_{1,n}^{\mathsf{H}} \mathbf{x}_{1j} \\ y_{2j,n} = \mathbf{w}_{2,n}^{\mathsf{H}} \mathbf{x}_{2j} \\ \\ y_{Ij,n} = \mathbf{w}_{I,n}^{\mathsf{H}} \mathbf{x}_{Ij} \end{matrix}$$



47

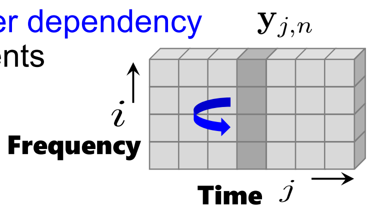
What p.d.f. is appropriate for a spectrum vector?



48

Choice of multivariate p.d.f.

- Necessary properties
  - Non-Gaussian (like ICA)
  - Representing higher-order dependency between vector components



- Well-used multivariate p.d.f.
  - Spherical super-Gaussian [Hiroe 2006],[Kim 2006]

$$p(\mathbf{y}_{j,n}) = C \exp(-||\mathbf{y}_{j,n}||_2)$$

  - Time-varying Gaussian [Ono+ 2012]

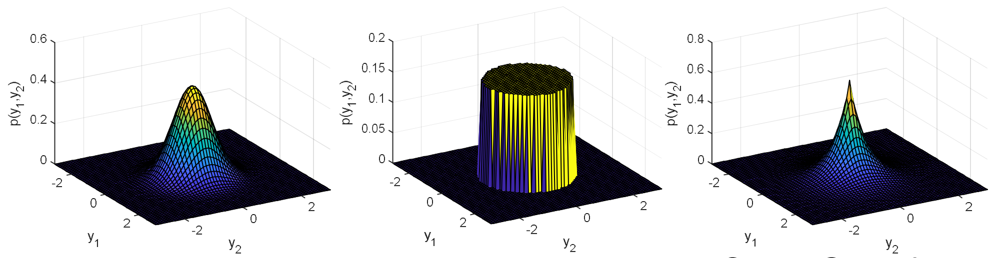
$$p(\mathbf{y}_{j,n}) = \frac{C}{\sigma_{j,n}^2} \exp\left(-\frac{||\mathbf{y}_{j,n}||_2^2}{\sigma_{j,n}^2}\right)$$

Variance is time-varying.
Totally it is super-Gaussian.
(Show later)

49

Why spherical super-Gaussian?

- Let $p(\mathbf{y}) = p(y_1, y_2) = q(r)$ where $r = \sqrt{y_1^2 + y_2^2}$
- When $q(r)$ is Gaussian, sub-Gaussian, and super-Gaussian, what dependency between y_1 and y_2 is represented?

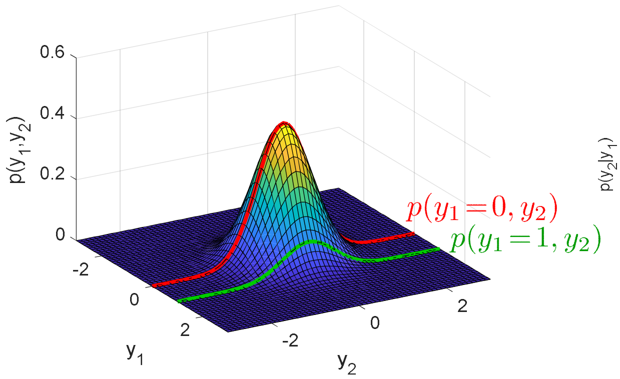


Gaussian   Sub-Gaussian   Super-Gaussian

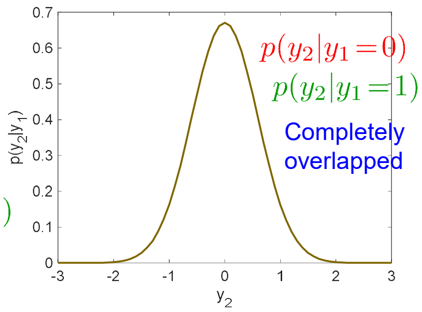
50

Spherical Gaussian

Joint p.d.f.



Conditional p.d.f.

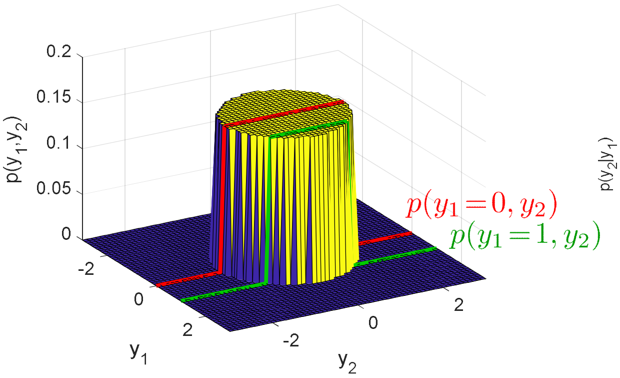


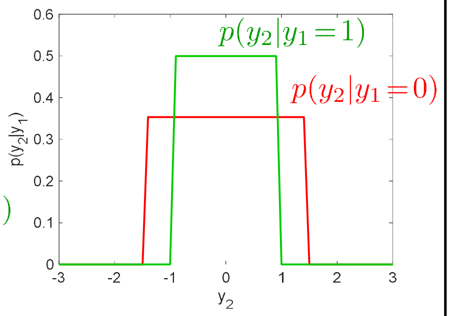
In spherical Gaussian case, the value of y_1 does not change the p.d.f. of y_2. It means y_1 and y_2 do not have any dependencies.

51

Spherical Sub-Gaussian

Joint p.d.f. | Conditional p.d.f.



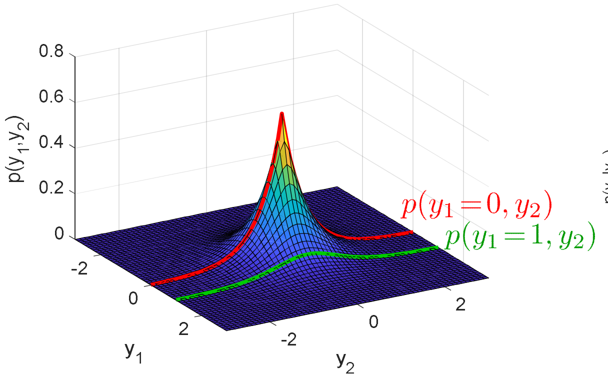


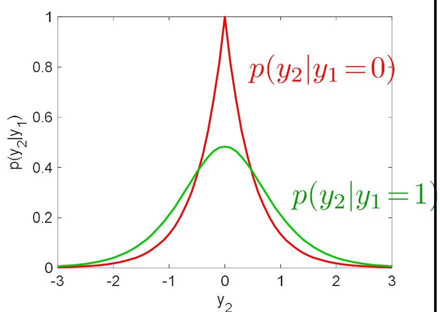
In spherical sub-Gaussian case, when $|y_1|$ is larger, $|y_2|$ tends to be smaller oppositely.

52

Spherical Super-Gaussian

Joint p.d.f. | Conditional p.d.f.





In spherical super-Gaussian case, when $|y_1|$ is larger, $|y_2|$ tends to be also larger. Therefore, it represents co-occurrence among components.

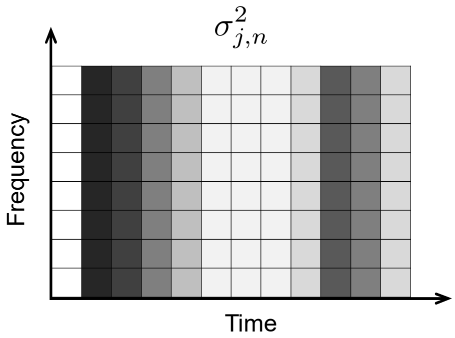
53

Time-Varying Gaussian

- P.d.f. of time-varying Gaussian

$$p(\mathbf{y}_{j,n}) = \frac{C}{\sigma_{j,n}^2} \exp\left(-\frac{||\mathbf{y}_{j,n}||_2^2}{\sigma_{j,n}^2} \right)$$

- Co-occurrence among frequency components are explicitly represented by $\sigma_{j,n}^2$.
  - Shared among all frequency *i*.
  - Can be changed at each time frame *j*.



54

Solutions for IVA

- The objective function of IVA is also nonlinear.

$$\mathcal{J}(\mathcal{W}) = \sum_{j=1}^{J} \sum_{n=1}^{N} G(\mathbf{y}_{j,n}) - 2J \sum_{i=1}^{I} \log |\det \mathbf{W}_i|$$

- Similarly as ICA, three typical methods
  - Natural gradient [Kim+ 2006,Hiroe 2006]
  - Pre-whitening + Fixed-point iteration (FastIVA) [Lee+ 2007]
    - Dr. Taesu Kim (an inventor of IVA)'s code is available https://github.com/teradepth/iva
  - Auxiliary function-based optimization (AuxIVA) [Ono2011, Ono2012b]

Auxiliary Function Approach

- Optimization Problem

  $J(\boldsymbol{\Theta}) \rightarrow \min$

- Auxiliary Function

  $Q(\boldsymbol{\Theta}, \tilde{\boldsymbol{\Theta}}) \geq J(\boldsymbol{\Theta})$

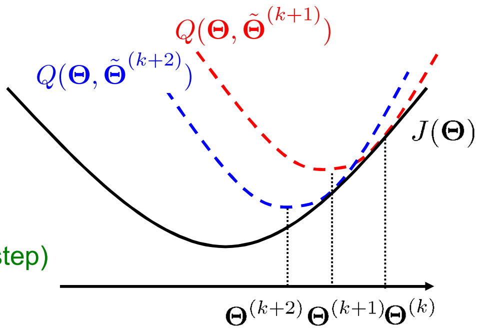
- Alternative Update Rules

  Auxiliary variable update (like E-step)

  $\tilde{\boldsymbol{\Theta}}^{(k+1)} = \mathrm{argmin}_{\tilde{\boldsymbol{\Theta}}}\ Q(\boldsymbol{\Theta}^{(k)}, \tilde{\boldsymbol{\Theta}})$

  Parameter update (like M-step)

  $\boldsymbol{\Theta}^{(k+1)} = \mathrm{argmin}_{\boldsymbol{\Theta}}\ Q(\boldsymbol{\Theta}, \tilde{\boldsymbol{\Theta}}^{(k+1)})$



- Advantages
  - Stable: Convergence is guaranteed
  - Simple: No tuning parameters such as step size

  But how to find useful auxiliary function is problem-dependent

More details will be explained later

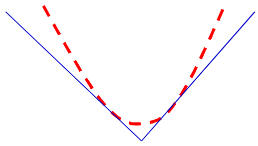
Theorem for quadratic auxiliary function

- If $G(\mathbf{y}) = G_R(||\mathbf{y}||_2)$ and $G'_R(r)/r$ is monotonically decreasing in $r \geq 0$,

  Spherical — Super-Gaussian

$$G(\mathbf{y}) \leq \frac{G'_R(r)}{2r}||\mathbf{y}||_2^2 + F(r)$$

Nonlinear contrast function — Quadratic auxiliary function



holds. The equality sign is valid iff $r = ||\mathbf{y}||_2$.

Examples

Multivariate p.d.f. $p(\mathbf{y})$	Contrast function $G(\mathbf{y}) = -\log p(\mathbf{y})$	Weight function $G'_R(r)/2r$
$Ce^{-\|\|\mathbf{y}\|\|_2^2}$	$\|\|\mathbf{y}\|\|_2^2 = r^2$	1
$Ce^{-\|\|\mathbf{y}\|\|_2}$	$\|\|\mathbf{y}\|\|_2 = r$	$1/2r$
$Ce^{-\log\cosh \|\|\mathbf{y}\|\|_2}$	$\cosh \|\|\mathbf{y}\|\|_2 = \cosh(r)$	$\tanh(r)/2r$

Auxiliary Function for IVA

- Objective Function of IVA

$$\mathcal{J}(\mathcal{W}) = \sum_{j=1}^{J}\sum_{n=1}^{N} G(\mathbf{y}_{j,n}) - 2J\sum_{i=1}^{I}\log|\det \mathbf{W}_i|$$

Nonlinear function of w

If G is spherical and derived from super-Gaussian

- Auxiliary Function for IVA

$$Q(\mathcal{W},\mathbf{r}) = 2J\left[\frac{1}{2}\sum_{i=1}^{I}\sum_{n=1}^{N}\mathbf{w}_{i,n}^{\mathsf{H}}\mathbf{V}_{i,n}\mathbf{w}_{i,n} - \sum_{i=1}^{I}\log|\det \mathbf{W}_i|\right] + F(\mathbf{r})$$

Quadratic function of w

$$\mathbf{r} = \{r_{j,n}\}$$

Source activity (auxiliary variable)

$$r_{j,n} = ||\mathbf{y}_{j,n}||_2 = \sqrt{\sum_{i=1}^{I}|\mathbf{w}_{i,n}^{\mathsf{H}}\mathbf{x}_{ij}|^2}$$

Weighted covariance matrix:

$$\mathbf{V}_{i,n} = \frac{1}{J}\sum_{j=1}^{J}\frac{G'_R(r_{j,n})}{2r_{j,n}}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathsf{H}}$$

Minimizing auxiliary function

- The demixing matrix should be updated such that auxiliary function is minimized.

$$Q(\mathcal{W},\mathbf{r}) = 2J\left[\frac{1}{2}\sum_{i=1}^{I}\sum_{n=1}^{N}\mathbf{w}_{i,n}^{\mathsf{H}}\mathbf{V}_{i,n}\mathbf{w}_{i,n} - \sum_{i=1}^{I}\log|\det \mathbf{W}_i|\right] + F(\mathbf{r})$$

$$\frac{\partial Q(\mathcal{W},\mathbf{r})}{\partial \mathbf{w}_{i,n}} = 0 \Rightarrow \mathbf{w}_{i,m}^{\mathsf{H}}\mathbf{V}_{i,n}\mathbf{w}_{i,n} = \delta_{mn} \quad (m = 1,\cdots,N)$$

- From $\frac{\partial Q(\mathcal{W},\mathbf{r})}{\partial \mathbf{w}_{i,n}} = 0$ for all $n = 1,\cdots,N$,

  HEAD (Hybrid Exact-Approximate Joint Diagonalization) problem [Yeredor 2009] is derived.

59

HEAD Problem (1/2)

For simplicity, a frequency index *i* is dropped in this slide.

Given N positive definite matrices $\mathbf{V}_1,\ \mathbf{V}_2, \cdots, \mathbf{V}_N$,
find an N × N matrix $\mathbf{W} = (\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_N)^{\mathsf{H}}$ such that

$$\mathbf{w}_m^{\mathsf{H}} \mathbf{V}_n \mathbf{w}_n = \delta_{mn} \quad (m, n = 1, 2, \cdots, N)$$

[Yeredor 2009]

ex. N=3 case

$$\begin{array}{lll} \mathbf{w}_1^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 1 & \mathbf{w}_1^{\mathsf{H}} \mathbf{V}_2 \mathbf{w}_2 = 0 & \mathbf{w}_1^{\mathsf{H}} \mathbf{V}_3 \mathbf{w}_3 = 0 \\ \mathbf{w}_2^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 0 & \mathbf{w}_2^{\mathsf{H}} \mathbf{V}_2 \mathbf{w}_2 = 1 & \mathbf{w}_2^{\mathsf{H}} \mathbf{V}_3 \mathbf{w}_3 = 0 \\ \mathbf{w}_3^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 0 & \mathbf{w}_3^{\mathsf{H}} \mathbf{V}_2 \mathbf{w}_2 = 0 & \mathbf{w}_3^{\mathsf{H}} \mathbf{V}_3 \mathbf{w}_3 = 1 \end{array}$$

60

HEAD Problem (2/2)

For simplicity, a frequency index *i* is dropped in this slide.

Given N positive definite matrices $\mathbf{V}_1,\ \mathbf{V}_2, \cdots, \mathbf{V}_N$,
find an N × N matrix $\mathbf{W} = (\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_N)^{\mathsf{H}}$ such that

$$\mathbf{w}_m^{\mathsf{H}} \mathbf{V}_n \mathbf{w}_n = \delta_{mn} \quad (m, n = 1, 2, \cdots, N)$$

[Yeredor 2009]

- Remarks
  - Number of equations = number of variables $= N^2$
  - When $N = 2$, it is equivalent to generalized eigenvalue problem, which can be solved in a closed form. (Show later)
  - When $N \geq 3$, a closed-form solution has never been found.

61

Derivation of sequential update rule

For simplicity, a frequency index *i* is dropped in this slide.

Let's consider to solve $\frac{\partial Q(\mathcal{W}, \mathbf{r})}{\partial \mathbf{w}_1} = 0$ in with fixing $\mathbf{w}_2, \mathbf{w}_3$.

$$\mathbf{w}_1^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 1$$
$$\mathbf{w}_2^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 0$$
$$\mathbf{w}_3^{\mathsf{H}} \mathbf{V}_1 \mathbf{w}_1 = 0$$

Some vector linearly independent of w_2 and w_3

$$\begin{bmatrix} \mathbf{b}^{\mathsf{H}} \\ \mathbf{w}_2^{\mathsf{H}} \\ \mathbf{w}_3^{\mathsf{H}} \end{bmatrix} \mathbf{V}_1 \mathbf{w}_1 = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

$\mathbf{B}$ $\quad c\mathbf{e}_1$

$$\mathbf{B}\mathbf{V}_1 \mathbf{w}_1 = c\mathbf{e}_1$$
$$\leftrightarrow \mathbf{w}_1 = c(\mathbf{B}\mathbf{V}_1)^{-1}\mathbf{e}_1$$

Scale can be fixed by

$$\mathbf{w}_1 \leftarrow \mathbf{w}_1 / \sqrt{\mathbf{w}_1^{\mathsf{H}} V_1 \mathbf{w}_1}$$

If we choose $\mathbf{w}_1$ obtained at the previous iteration as $\mathbf{b}$, $\mathbf{B}$ can be rewritten by $\mathbf{W}$.

$$\mathbf{w}_1 \leftarrow (\mathbf{W}\mathbf{V}_1)^{-1}\mathbf{e}_1$$
$$\mathbf{w}_1 \leftarrow \mathbf{w}_1 / \sqrt{\mathbf{w}_1^{\mathsf{H}} V_1 \mathbf{w}_1}$$

We can update $\mathbf{w}_2, \mathbf{w}_3$ sequentially in the same way.

62

Algorithm of AuxIVA

Iterate until convergence [Ono 2011]

For $n = 1 : N$ (every source)

$$y_{ij,n} = \mathbf{w}_{i,n}^{\mathsf{H}} \mathbf{x}_{ij}$$ Update of separation

$$r_{j,n} = ||\mathbf{y}_{j,n}||_2 = \sqrt{\sum_{i=1}^{I} |y_{ij,n}|^2}$$ Update of source activity (shared in all frequency)

For $i = 1 : I$ (every frequency)

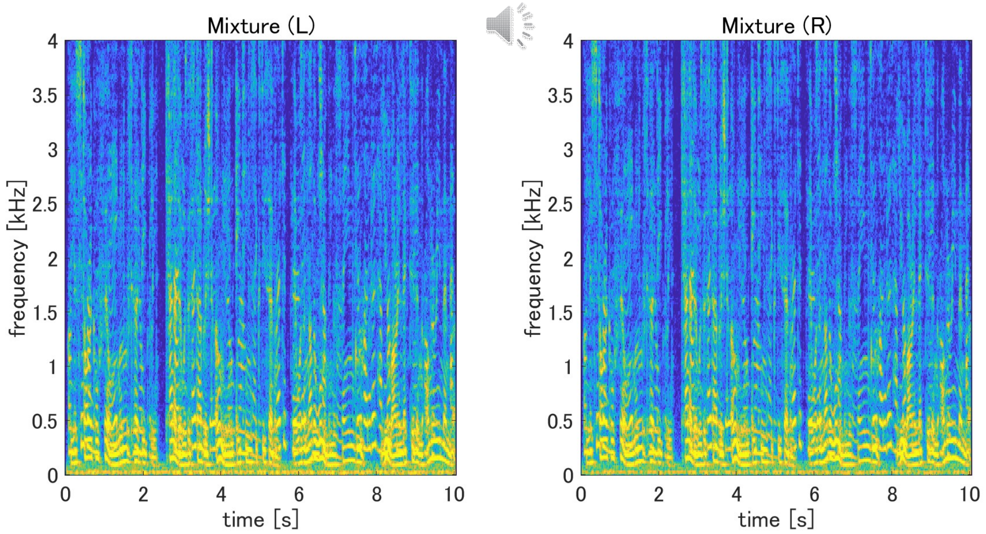
$$\mathbf{V}_{i,n} = \frac{1}{J}\sum_{j=1}^{J} \frac{G'_R(r_{j,n})}{2r_{j,n}} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathsf{H}}$$ Update of weighted covariance matrix

$$\mathbf{w}_{i,n} \leftarrow (\mathbf{W}_i \mathbf{V}_{i,n})^{-1}\mathbf{e}_n$$
$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n} / \sqrt{\mathbf{w}_{i,n}^{\mathsf{H}} \mathbf{V}_{i,n} \mathbf{w}_{i,n}}$$ Update of demixing matrix

Unit vector with the *n*th element unity

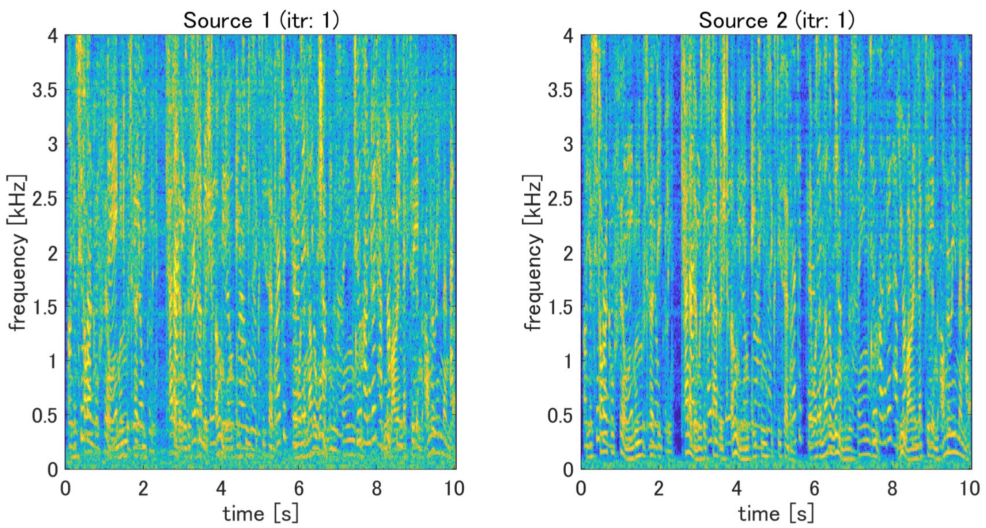
63

Separation Example (1/6)

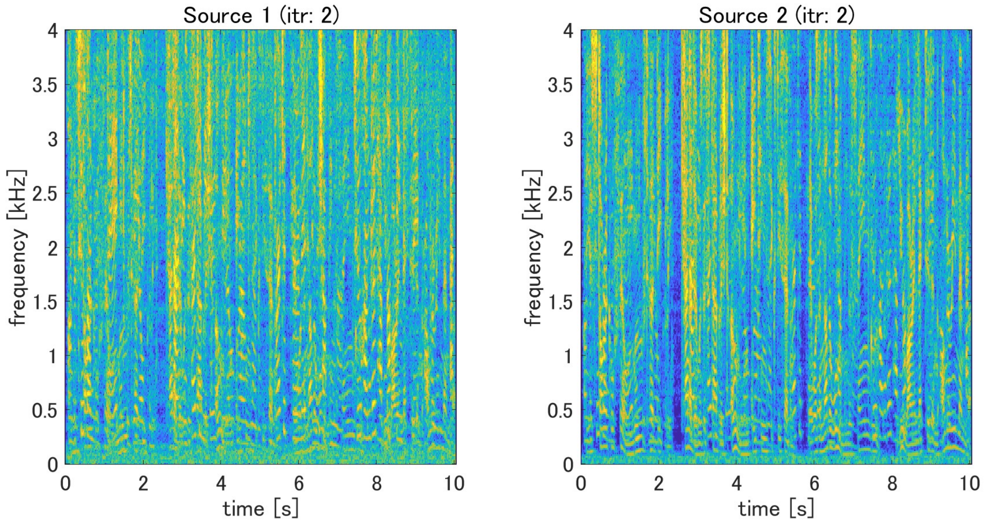


64

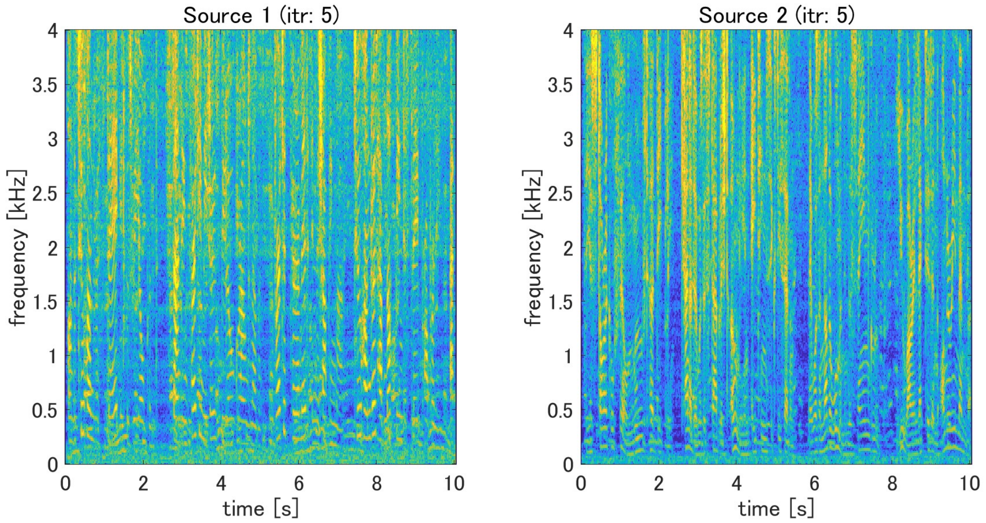
Separation Example (2/6)



Separation Example (3/6)

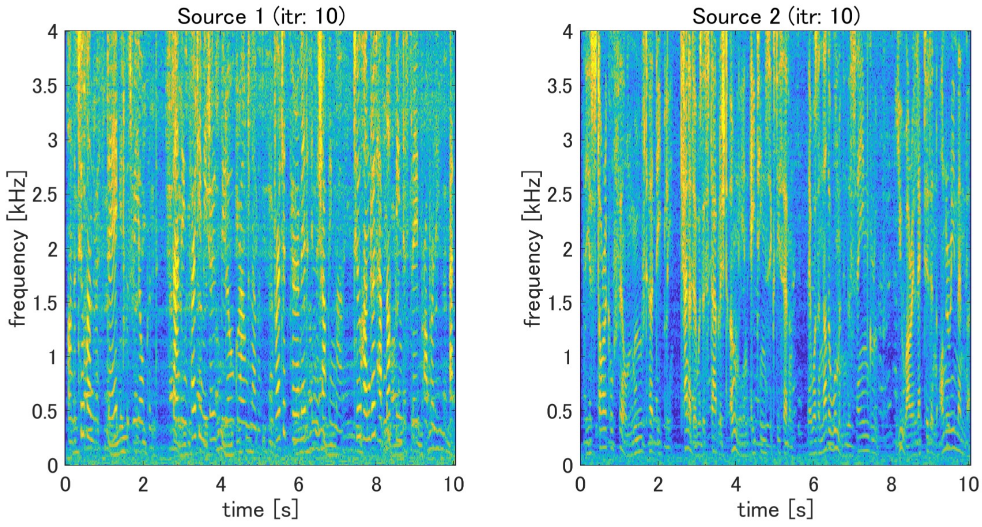


Separation Example (4/6)



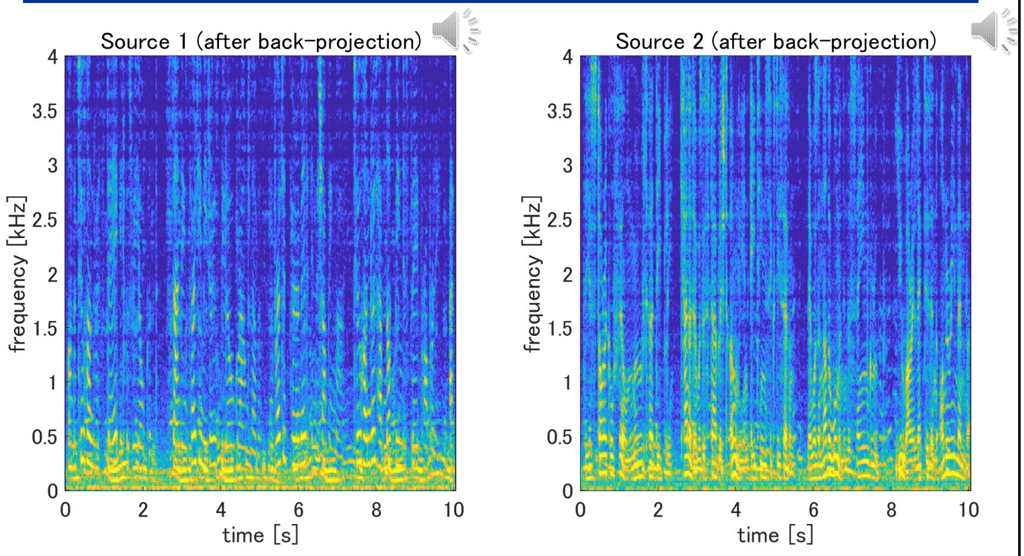
67

Separation Example (5/6)



68

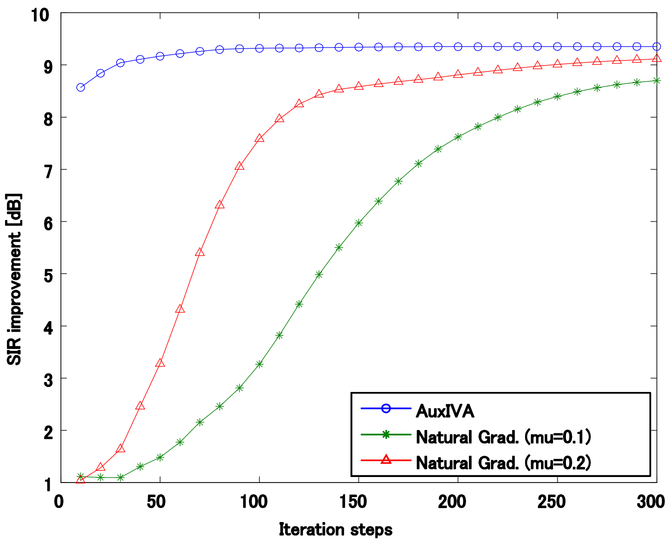
Separation Example (6/6)



69

Comparison with Natural Gradient

Three sources case



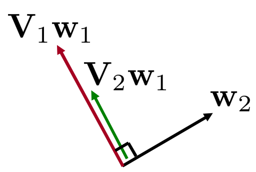
One iteration (Matlab)

AuxIVA : 0.34s
GradIVA: 0.16s

70

HEAD Problem in Two-Sources Case

$$\mathbf{w}_1^{\mathsf{H}}\mathbf{V}_1\mathbf{w}_1 = 1 \qquad \mathbf{w}_1^{\mathsf{H}}\mathbf{V}_2\mathbf{w}_2 = 0$$
$$\mathbf{w}_2^{\mathsf{H}}\mathbf{V}_1\mathbf{w}_1 = 0 \qquad \mathbf{w}_2^{\mathsf{H}}\mathbf{V}_2\mathbf{w}_2 = 1$$



$$\mathbf{V}_1\mathbf{w}_1 \perp \mathbf{w}_2$$
$$\mathbf{V}_2\mathbf{w}_1 \perp \mathbf{w}_2$$
$$\Rightarrow \mathbf{V}_1\mathbf{w}_1 \parallel \mathbf{V}_2\mathbf{w}_1 \Rightarrow \mathbf{V}_1\mathbf{w}_1 = \lambda\mathbf{V}_2\mathbf{w}_1$$

HEAD problem is deformed to generalized eigenvalue problem, which can be solved in a closed form [Yoshioka 2008,Ono 2010]

71

Algorithm of Stereo AuxIVA

Iterate until convergence [Ono 2012b]

$$y_{ij,n} = \mathbf{w}_{i,n}^{\mathsf{H}} \mathbf{x}_{ij}$$

$$r_{j,n} = \|\mathbf{y}_{j,n}\|_2 = \sqrt{\sum_{i=1}^{I} |y_{ij,n}|^2} \quad (n = 1, 2)$$

Update of separation

Update of source activity (shared in all frequency)

$i = 1 : I$

$$\mathbf{V}_{i,n} = \frac{1}{J} \sum_{j=1}^{J} \frac{G'_R(r_{j,n})}{2 r_{j,n}} \mathbf{x}_{ij} \mathbf{x}_{ij}^{\mathsf{H}} \quad (n = 1, 2)$$

Update of weighted covariance matrix

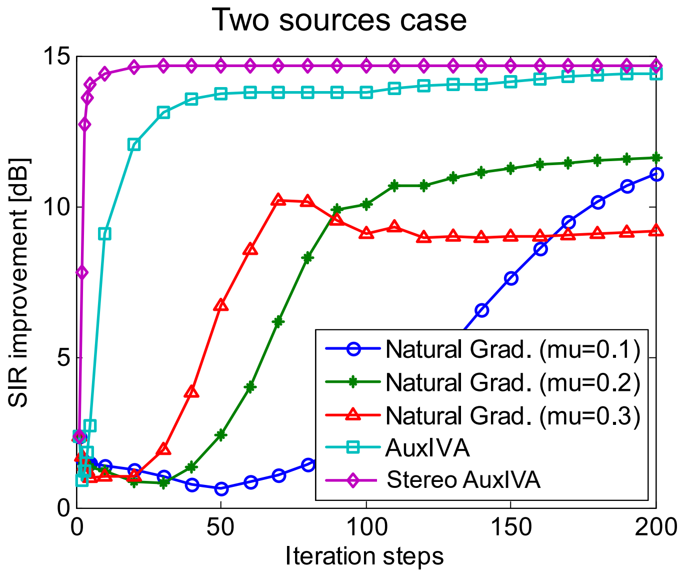
Find $\mathbf{u}_{i,1}, \ \mathbf{u}_{i,2}$ by solving $\mathbf{V}_{i,2} \mathbf{u}_{i,n} = \lambda_{i,n} \mathbf{V}_{i,1} \mathbf{u}_{i,n} \ (\lambda_{i,1} \geq \lambda_{i,2})$

$$\mathbf{w}_{i,n} = \mathbf{u}_{i,n} / \sqrt{\mathbf{u}_{i,k}^{H} \mathbf{V}_{i,n} \mathbf{u}_{i,n}} \quad (n = 1, 2)$$

Update of demixing matrix

72

Comparison with Natural Gradient (2)



73

Explicit Representation of Eigenvectors of 2 × 2 Matrix

For simplicity, a frequency index *i* is dropped in this slide.

- $\mathbf{V}_2\mathbf{u}_n = \lambda_n \mathbf{V}_1 \mathbf{u}_n \longleftrightarrow \mathbf{H}\mathbf{u}_k = \lambda_n \mathbf{u}_n \ (\mathbf{H} = \mathbf{V}_1^{-1}\mathbf{V}_2)$
- Two eigenvectors of 2 × 2 matrix can be explicitly given by

$$\lambda_1 = \frac{\mathrm{tr}(\mathbf{H}) + \sqrt{\mathrm{tr}(\mathbf{H})^2 - 4\det(\mathbf{H})}}{2}$$

$$\lambda_2 = \frac{\mathrm{tr}(\mathbf{H}) - \sqrt{\mathrm{tr}(\mathbf{H})^2 - 4\det(\mathbf{H})}}{2}$$

$$\mathbf{u}_1 = \begin{pmatrix} H_{22} - \lambda_1 \\ -H_{21} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -H_{12} \\ H_{11} - \lambda_2 \end{pmatrix}$$

$\lambda_1 \geq \lambda_2$ always holds

- See [Ono 2012b] about fast implementation using vector operation.

74

Implementation on iPhone

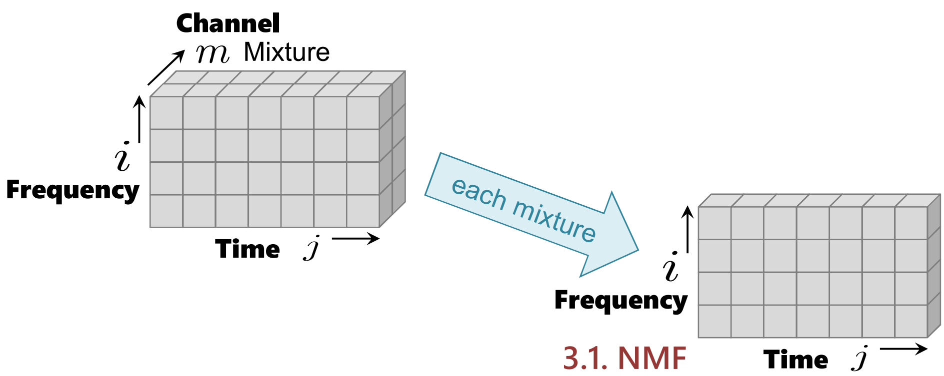
- Stereo AuxIVA was implemented on iPhone [Ono 2012b]
- Calculation time is almost linear to input signal length (RTF≒1/5 @ 16kHz, 10itr. on iPhone4)

- Demo on youtube (https://www.youtube.com/watch?v=ILMbflDMMeE)

Tutorial structure

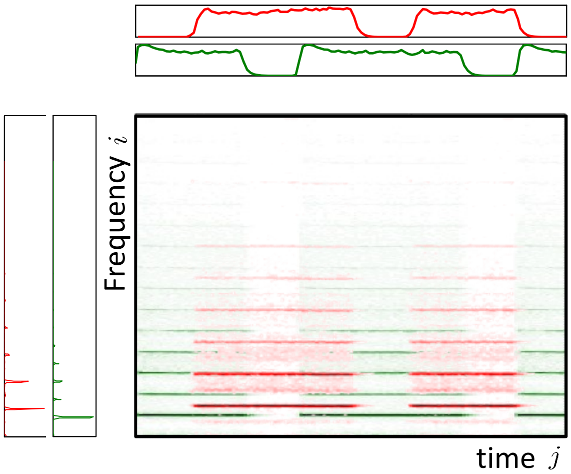
1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

Tensor and sliced matrices



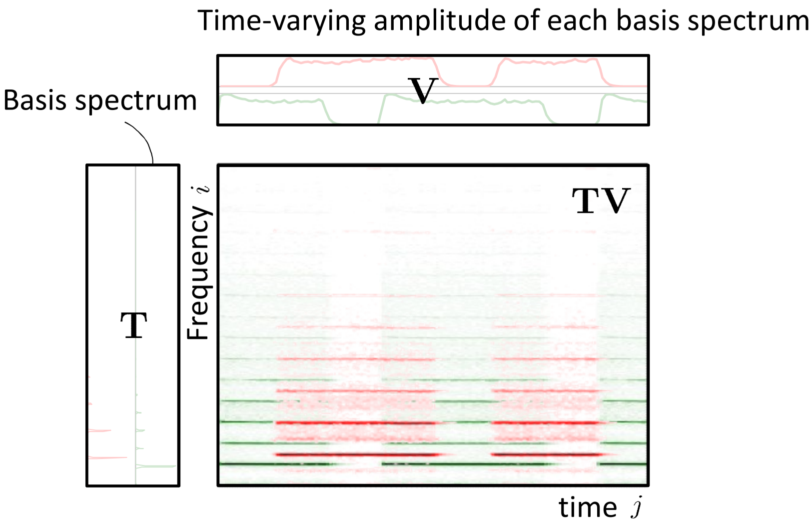
77

What is NMF? From “generative model” perspective



78

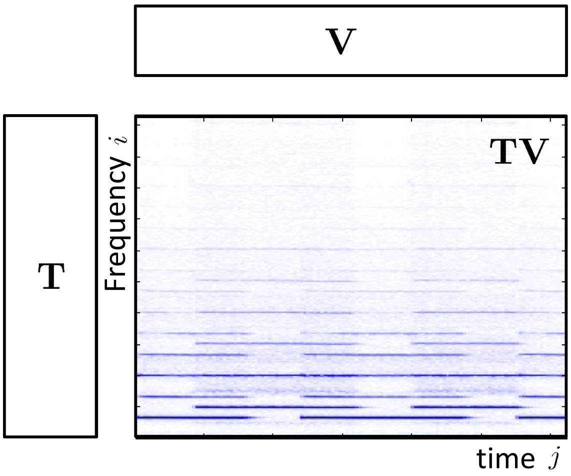
What is NMF? From “generative model” perspective



79

What is NMF? From “generative model” perspective

Source separation = *inverse problem* of estimating $\mathbf{T}$ and $\mathbf{V}$ from $\mathbf{X}$



80

NMF as spectrogram model fitting

- Model a mixture spectrum as the sum of basis spectra scaled by time-varying magnitudes

$$x_{ij} \simeq \sum_k t_{jk} v_{kj}$$

Time-varying magnitude

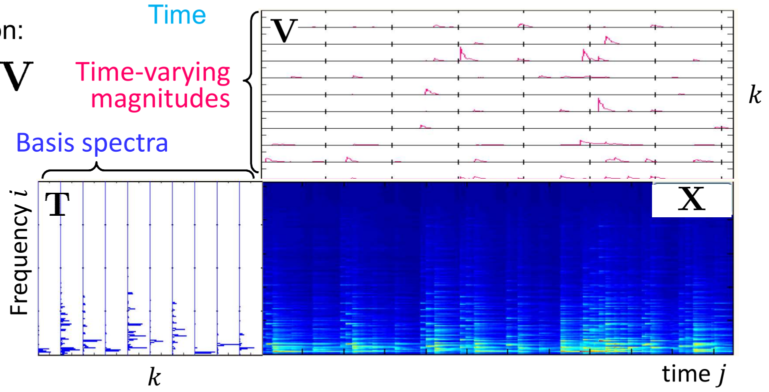
Basis spectrum

Frequency

Time

Matrix notation:

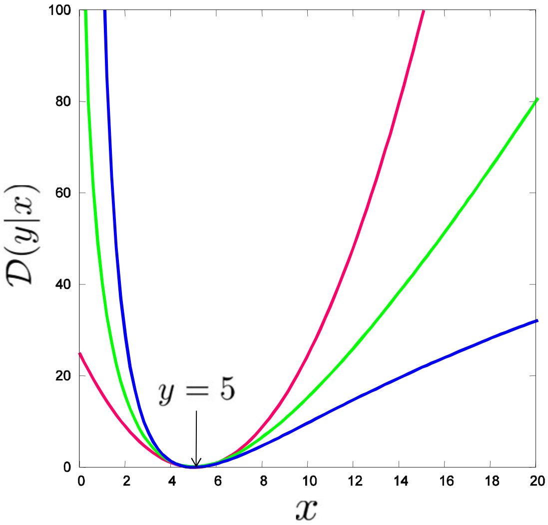
$$\mathbf{X} \simeq \mathbf{TV}$$



81

Popular divergence measures

- Measure of difference between x and y



Euclidean distance

$\mathcal{D}(y|x) = (y - x)^2$

Kullback-Leibler (KL) divergence

$\mathcal{D}(y|x) = y \log \frac{y}{x} - (y - x)$

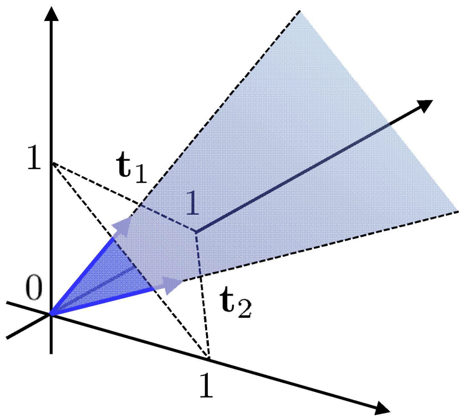
Itakura-Saito (IS) divegence

$\mathcal{D}(y|x) = \frac{y}{x} - \log \frac{y}{x} - 1$

82

Geometrical understanding of NMF

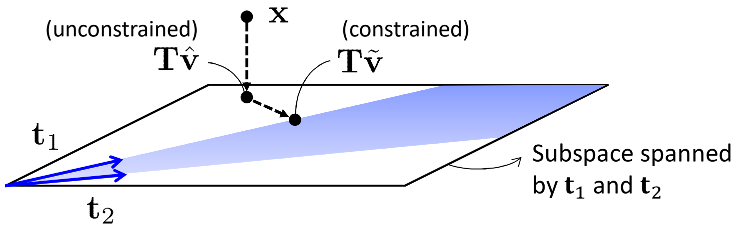
- Because of the non-negativity of $\mathbf{T}$, all basis vectors lie in the first quadrant.
- Because of the non-negativity of $\mathbf{V}$, $\mathbf{T}\mathbf{v}_j$ can only cover the area (a convex cone) enclosed by the extended lines of all the basis vectors.
- NMF attempts to find a convex cone that is closest to all the observed vectors.



83

Sparsity-inducing effect of NMF

- NMF naturally produces sparse representations
  - Let $\hat{\mathbf{v}}$ be the solution to an unconstrained optimization problem:
    $$\hat{\mathbf{v}} = \underset{\mathbf{v}}{\operatorname{argmin}}\, \mathcal{D}(\mathbf{x}|\mathbf{T}\mathbf{v})$$
  - $\mathbf{T}\hat{\mathbf{v}}$ corresponds to the closest point from $\mathbf{y}$ in the subspace spanned by $\mathbf{t}_1, \ldots, \mathbf{t}_K$.
  - Except for the case where $\hat{\mathbf{v}}$ is non-negative, the solution to the constrained optimization problem, $\mathbf{T}\tilde{\mathbf{v}}$, will be the closest point to $\mathbf{T}\hat{\mathbf{v}}$ in the area enclosed by the extended lines of all the basis vectors.
  - This means at least one of the elements of $\tilde{\mathbf{v}}$ becomes 0.



84

Itakura-Saito divergence NMF [Févotte+2009]

- Model mixture signal as the sum of Gaussian-distributed random signals with rank-1 power spectrograms

Complex spectrogram of component k

$$c_{ijk} \sim \mathcal{N}_{\mathbb{C}}(c_{ijk}|0, t_{ik}v_{kj})$$

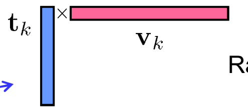
$$x_{ij} = \sum_k c_{ijk}$$

If c_{ijk} and $c_{ijk'}$ are independent

$$x_{ij} \sim \mathcal{N}_{\mathbb{C}}\left(x_{ij}\middle|0, \sum_k t_{ik}v_{kj}\right)$$

→ Likelihood function of $\boldsymbol{T}$ and $\boldsymbol{V}$ $p(\mathbf{X}|\mathbf{T},\mathbf{V})$

Power spectrogram of component k



Rank-1 spectrogram

- Maximum likelihood of $\mathbf{T}$ and $\mathbf{V}$ amounts to NMF using Itakura-Saito divergence

$$\underset{\mathbf{T},\mathbf{V}}{\operatorname{argmax}} \log p(\mathbf{X}|\mathbf{T},\mathbf{V})$$
$$= \underset{\mathbf{T},\mathbf{V}}{\operatorname{argmin}} \sum_{i,j} \left(\frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} - \log \frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} - 1 \right)$$

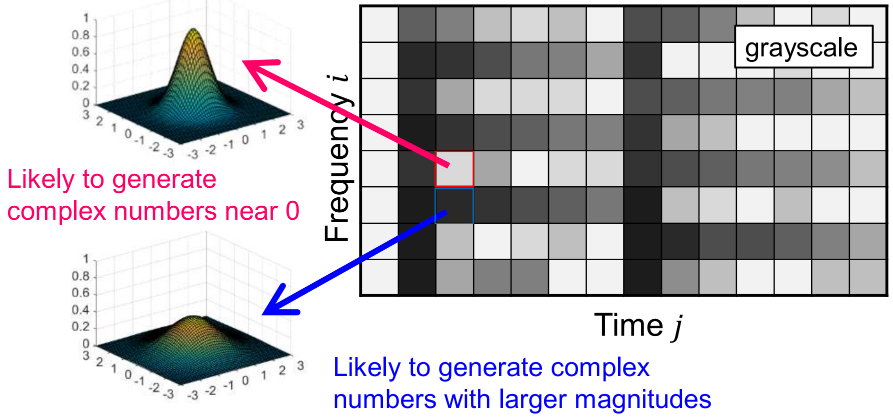
Itakura-Saito divergence

85

Local Gaussian model (LGM)

- Generative model assuming each element of a complex spectrogram to independently follow a zero-mean complex Gaussian distribution with a different variance (power)

$$c_{ijk} \sim \mathcal{N}_{\mathbb{C}}(c_{ijk}|0, \sigma_{ijk}^2)$$ variance



86

Itakura-Saito divergence NMF (IS-NMF)

- Optimization problem:
  Minimize

$$\mathcal{D}_{\mathrm{IS}}(\mathbf{T}, \mathbf{V}) = \sum_{i,j} \left(\frac{|x_{ij}|^2}{\sum_k t_{ik} v_{kj}} - \log \frac{|x_{ij}|^2}{\sum_k t_{ik} v_{kj}} - 1 \right)$$
$$= \sum_{i,j} \left(\frac{|x_{ij}|^2}{\sum_k t_{ik} v_{kj}} + \log \sum_k t_{ik} v_{kj} - \cdots \right)$$

  subject to $\forall i, j, k, \ t_{ik} \geq 0, \ v_{kj} \geq 0$
- How can we solve this?
  - EM algorithm [Févotte+2009]
  - Auxiliary function approach
    - –Majorization-Maximization (MM) algorithm [Kameoka+2006], [Nakano+2010], [Févotte+2011]
    - –Majorization-Equalization (ME) algorithm [Févotte+2011]

Expectation-Maximization (EM) algorithm

- When a certain likelihood function can be written as $p(\mathbf{x}|\theta) = \int p(\mathbf{x}, \mathbf{c}|\theta)\mathrm{d}\mathbf{c}$ where $\mathbf{c}$ is a set of latent variables, a stationary point of $\log p(\mathbf{x}|\boldsymbol{\theta})$ can be found by iteratively performing the following steps:
  - Expectation-step $q(\mathbf{c}|\mathbf{x}) \leftarrow \dfrac{p(\mathbf{x}, \mathbf{c}|\theta)}{\int p(\mathbf{x}, \mathbf{c}'|\theta)\mathrm{d}\mathbf{c}'} = p(\mathbf{c}|\mathbf{x}, \theta)$
  - Maximization-step $\theta \leftarrow \underset{\theta}{\operatorname{argmax}} \underbrace{\int q(\mathbf{c}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{c}|\theta)\mathrm{d}\mathbf{c}}$

$$Q(\theta, \theta') = \mathbb{E}_{\mathbf{c}\sim p(\mathbf{c}|\mathbf{x},\theta')}[\log p(\mathbf{x}, \mathbf{c}|\theta)]$$

Q function

IS-NMF optimization with EM algorithm

[Févotte+2009]

- Likelihood function for IS-NMF:

$$p(\mathbf{x}|\theta) = \int p(\mathbf{x}, \mathbf{c}|\theta)\mathrm{d}\mathbf{c} = \int \underline{p(\mathbf{x}|\mathbf{c})}\,\underline{p(\mathbf{c}|\theta)}\mathrm{d}\mathbf{c} \quad \text{where } \theta = \{\mathbf{T}, \mathbf{V}\}$$

$p(\mathbf{x}|\mathbf{c}) = \prod_{i,j} \delta\left(x_{ij} - \sum_k c_{ijk}\right)$ (independent of θ)

$p(\mathbf{c}|\theta) = \prod_k \prod_{i,j} \mathcal{N}_{\mathbb{C}}(c_{ijk}|0, t_{ik}v_{kj})$

- Q function

$$Q(\theta, \theta') = \sum_k \sum_{i,j} \mathbb{E}_{c_{ijk}\sim p(c_{ijk}|x_{ij},\theta')}\left[\underbrace{\log \mathcal{N}_{\mathbb{C}}(c_{ijk}|0, t_{ik}v_{kj})}_{-\log t_{ik}v_{kj} - \frac{|c_{ijk}|^2}{t_{ik}v_{kj}}}\right]$$

- E-step

$$r_{ijk} \leftarrow \mathbb{E}_{c_{ijk}\sim p(c_{ijk}|x_{ij},\theta')}[|c_{ijk}|^2]$$
$$= t_{ik}v_{kj} - \frac{t_{ik}^2 v_{kj}^2}{\phi_{ij}} + \frac{t_{ik}^2 v_{kj}^2 |x_{ij}|^2}{\phi_{ij}^2} \quad \text{where } \phi_{ij} = \sum_k t_{ik}v_{kj}$$

- M-step

$$\theta \leftarrow \underset{\theta}{\operatorname{argmax}} \sum_k \sum_{i,j} \left(-\log t_{ik}v_{kj} - \frac{r_{ikj}}{t_{ik}v_{kj}} \right)$$

89

Auxiliary function approach

- Techniques to find a stationary point of an objective function $D(\theta)$ using an auxiliary function $G(\theta, \alpha)$ that satisfies

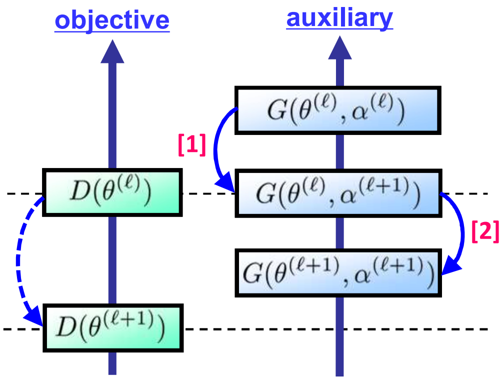
$$D(\theta) = \min_{\alpha} G(\theta, \alpha)$$

- Majorization-minimization
[Hunter & Lange 2004]

[1] $\alpha^{(\ell+1)} = \underset{\bar{\theta}}{\operatorname{argmin}}\, G(\theta^{(\ell)}, \alpha)$

[2] $\theta^{(\ell+1)} = \underset{\theta}{\operatorname{argmin}}\, G(\theta, \alpha^{(\ell+1)})$

$$\begin{aligned} D(\theta^{(\ell)}) &= G(\theta^{(\ell)}, \alpha^{(\ell+1)}) \\ &\geq G(\theta^{(\ell+1)}, \alpha^{(\ell+1)}) \\ &\geq D(\theta^{(\ell+1)}) \end{aligned}$$



- Majorization-equalization
[Févotte & Idier 2010]

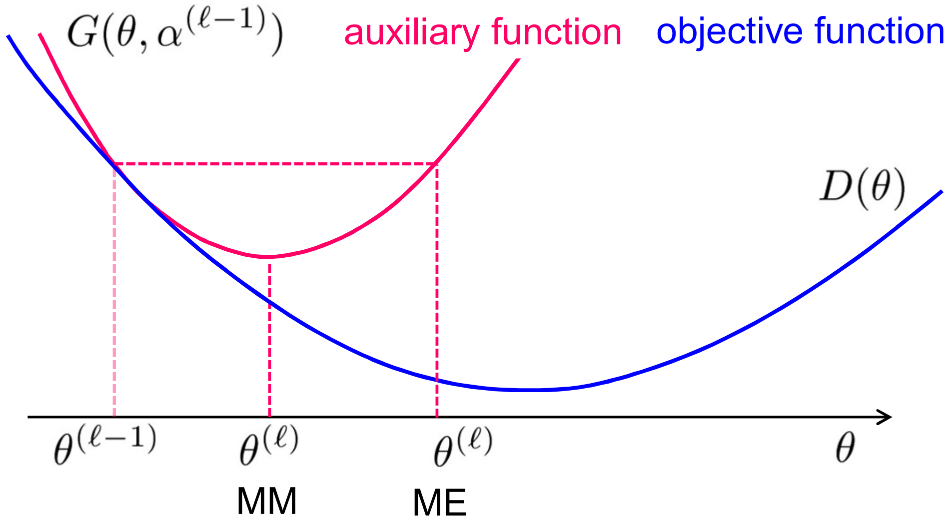
[1] $\alpha^{(\ell+1)} = \underset{\bar{\theta}}{\operatorname{argmin}}\, G(\theta^{(\ell)}, \alpha)$

[2] $\theta^{(\ell+1)} \leftarrow \theta \text{ such that } G(\theta, \alpha^{(\ell+1)}) = G(\theta^{(\ell)}, \alpha^{(\ell+1)}),\ \theta \neq \theta^{(\ell)}$

90

Graphical illustration of MM and ME algorithms

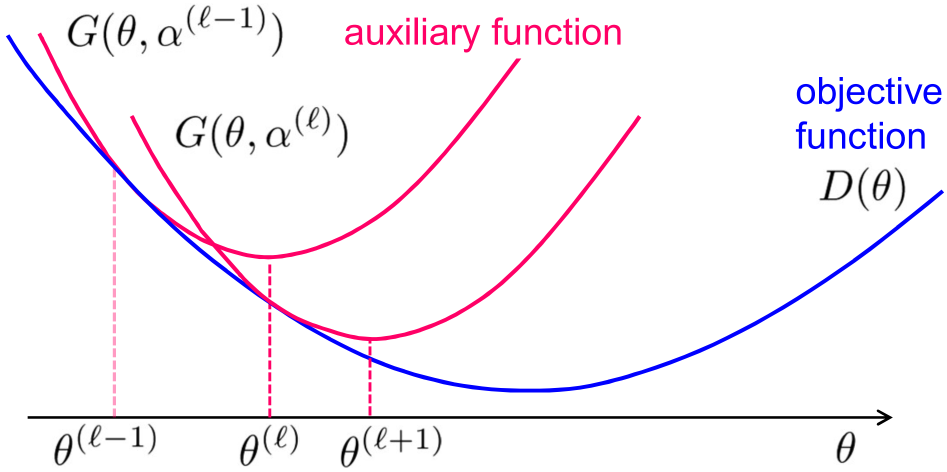
- Comparison between MM and ME updates



91

Graphical illustration of MM and ME algorithms

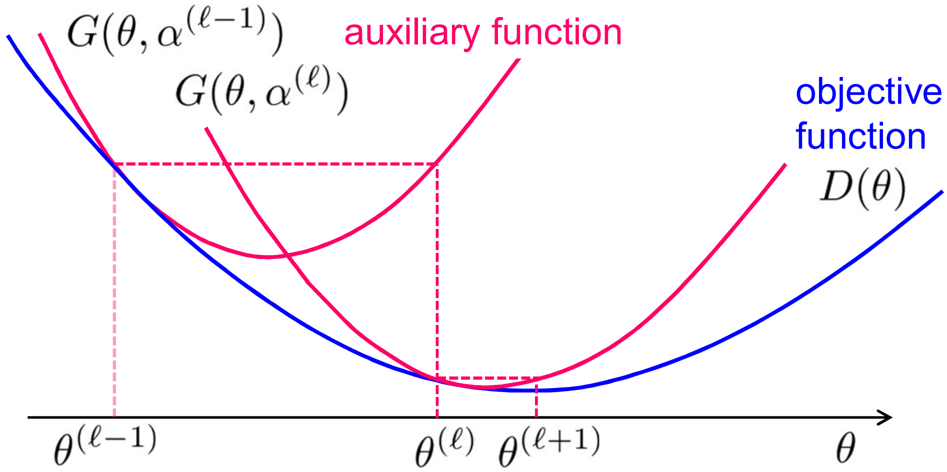
- MM algorithm



92

Graphical illustration of MM and ME algorithms

- ME algorithm



Relation to EM algorithm

- Objective: maximize $p(\mathbf{x}|\theta)$ w.r.t. θ

$$\log p(\mathbf{x}|\theta) = \log \int p(\mathbf{x}, \mathbf{c}|\theta) \mathrm{d}\mathbf{c}$$

$\mathbf{c}$: latent variable

$$= \log \int q(\mathbf{c}|\mathbf{x}) \frac{p(\mathbf{x}, \mathbf{c}|\theta)}{q(\mathbf{c}|\mathbf{x})} \mathrm{d}\mathbf{c}$$

$$\geq \int q(\mathbf{c}|\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{c}|\theta)}{q(\mathbf{c}|\mathbf{x})} \mathrm{d}\mathbf{c}$$

← Jensen's inequality
Auxiliary function

E step:
Maximize auxiliary function w.r.t. q

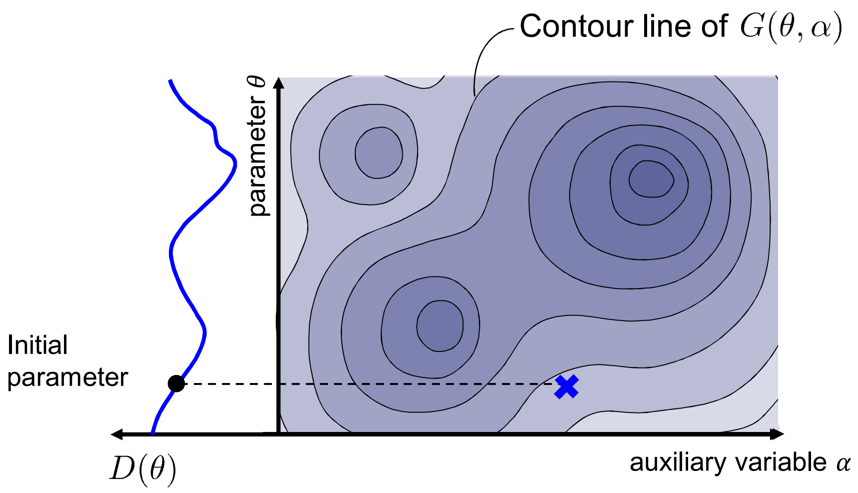
$$q(\mathbf{c}|\mathbf{x}) \leftarrow \frac{p(\mathbf{x}, \mathbf{c}|\theta)}{\int p(\mathbf{x}, \mathbf{c}'|\theta) \mathrm{d}\mathbf{c}'} = p(\mathbf{c}|\mathbf{x}, \theta)$$

M step:
Maximize auxiliary function w.r.t. θ

$$\theta \leftarrow \underset{\theta}{\operatorname{argmax}} \int q(\mathbf{c}|\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{c}|\theta)}{q(\mathbf{c}|\mathbf{x})} \mathrm{d}\mathbf{c}$$

Another look at MM algorithm

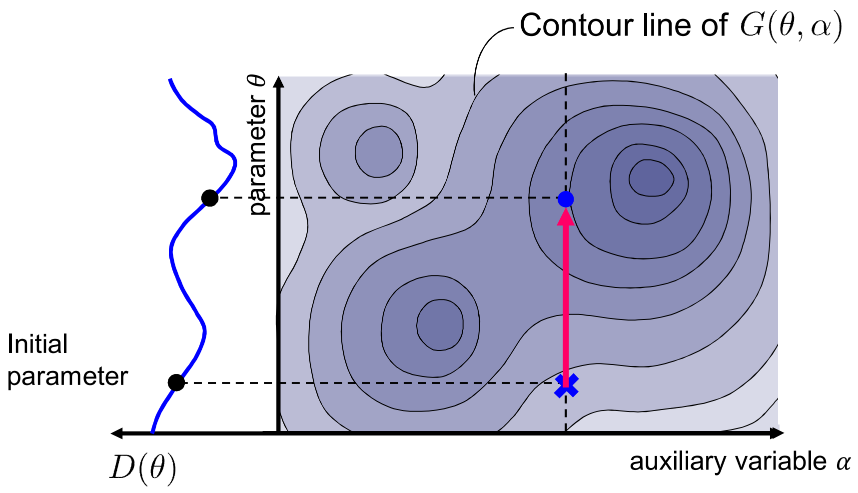
- Coordinate descent of $G(\theta, \alpha)$



95

Another look at MM algorithm

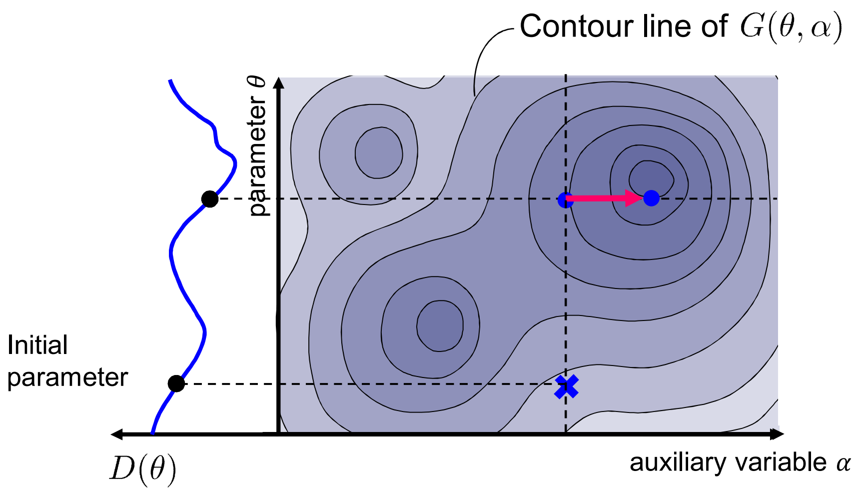
- Coordinate descent of $G(\theta, \alpha)$



96

Another look at MM algorithm

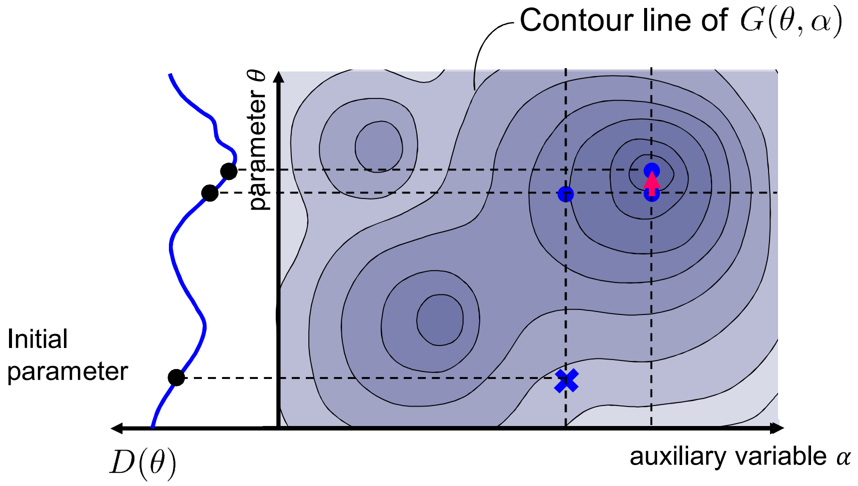
- Coordinate descent of $G(\theta, \alpha)$



97

Another look at MM algorithm

- Coordinate descent of $G(\theta, \alpha)$



98

Motivations for auxiliary function approach

- Auxiliary function can be useful and effective when one wants to
  - handle a non-convex objective function with multiple local optima and stationary points,
  - handle an objective function that has discontinuous/non-differentiable points,
  - handle equality/inequality constraints, and
  - avoid matrix inversions.

99

Useful inequalities for auxiliary function design (1/7)

- Jensen's inequality for non-negative arguments

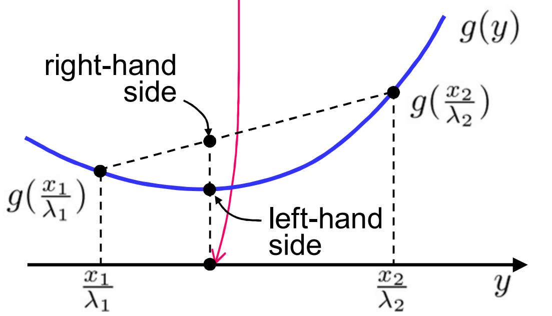
g : convex function
$x_k \geq 0,$
$\lambda_k \geq 0, \quad \sum_k \lambda_k = 1$

$$g\left(\sum_k x_k\right) \leq \sum_k \lambda_k g\left(\frac{x_k}{\lambda_k}\right)$$

Equality holds when $\lambda_k = \dfrac{x_k}{\sum_{k'} x_{k'}}$

E.g.) when $g(y) = -\log y$:

$$-\log\left(\sum_k \lambda_k \frac{x_k}{\lambda_k}\right) \leq -\sum_k \lambda_k \log \frac{x_k}{\lambda_k}$$



100

Useful inequalities for auxiliary function design (2/7)

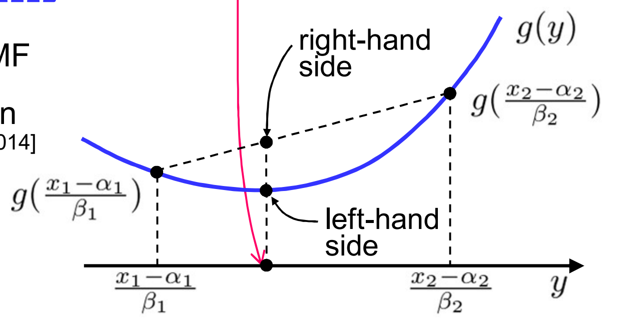
- Jensen's inequality for real number arguments

g : convex function
$\sum_k \alpha_k = 0,$
$\beta_k \geq 0, \quad \sum_k \beta_k = 1$

$$g\left(\sum_k x_k\right) \leq \sum_k \beta_k g\left(\frac{x_k - \alpha_k}{\beta_k}\right)$$

Equality holds when $\alpha_k = x_k - \beta_k\left(\sum_k x_k\right)$

- Used for complex NMF
[Kameoka+2009]
and RBM optimization
[Kameoka+2014][Takamune+2014]



Useful inequalities for auxiliary function design (3/7)

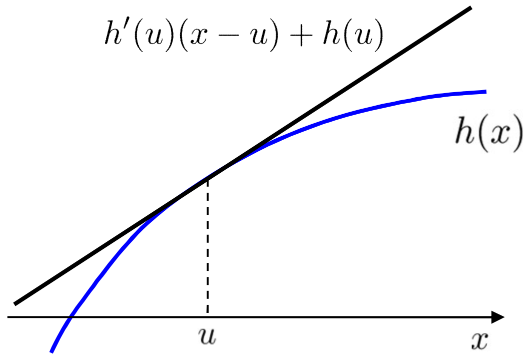
- 1st order Taylor expansion of convex/concave functions

h: concave function → $h(x) \leq h'(u)(x-u) + h(u)$

Equality holds when $u = x$

E.g.) when $h(x) = -\log x$:

$$\log x \leq \frac{1}{u}(x-u) + \log u$$



Useful inequalities for auxiliary function design (4/7)

- 1st order Taylor expansion of logarithmic function

Scalar case:

$$\log x \leq \frac{x}{u} + \log u - 1$$

Equality holds when $u = x$

Extension to matrix case:

$$\log\det \mathbf{X} \leq \mathrm{tr}(\mathbf{U}^{-1}\mathbf{X}) + \log\det \mathbf{U} - M$$

Equality holds when $\mathbf{U} = \mathbf{X}$

→ Equals to the sum of the logarithms of the eigenvalues of $\mathbf{X}$

Used in multichannel NMF frameworks [Sawada+2012, Higuchi+2014] and Positive Semidefinite Tensor Factorization [Yoshii+2013]

103

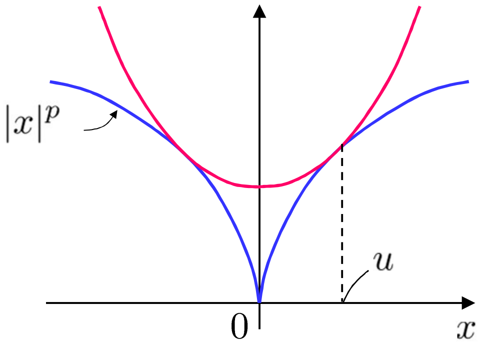
Useful inequalities for auxiliary function design (5/7)

- Quadratic function tangent to power functions

When $0 < p \leq 2$:

$$|x|^p \leq \frac{p}{2}|u|^{p-2}x^2 + |u|^p - \frac{p}{2}|u|^p$$

Equality holds when $u = x$



- Used for sparse regularization for complex NMF [Kameoka+2009]

104

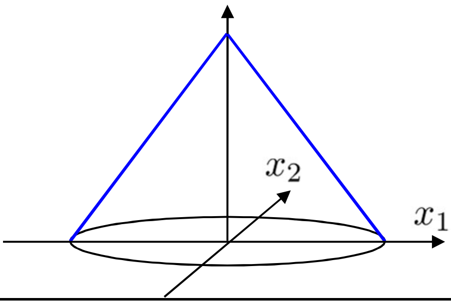
Useful inequalities for auxiliary function design (6/7)

- 1st order Taylor expansion of L2 norm

When $\|\mathbf{w}\|_2 = 1$ → $-\|\mathbf{x}\|_2 \leq -\mathbf{w}^\mathsf{T}\mathbf{x}$ Equality holds when $\mathbf{w} = \mathbf{x}/\|\mathbf{x}\|_2$

For complex number:

When $|\omega| = 1$ → $-|z| \leq -\mathrm{Re}(\omega^* z)$ Equality holds when $\omega = z/|z|$



- Used for sound source localization [Ono+2009][Ono+2010] and Time-domain Spectrogram Factorization (TSF) [Kameoka2015]

105

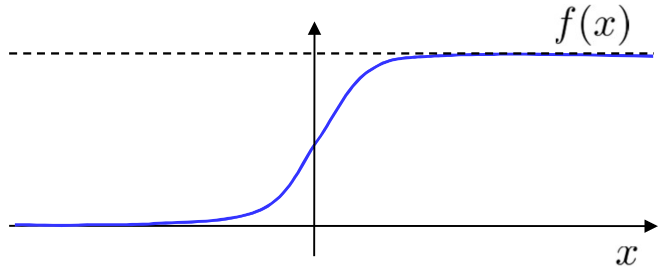
Useful inequalities for auxiliary function design (7/7)

- Logistic function

$$f(x) = \frac{1}{1+\exp(-x)}$$
$$\geq f(\eta)\exp\left(\frac{x-\eta}{2} - \frac{\tanh(\eta/2)}{4\eta}(x^2-\eta^2)\right)$$

Gaussian distribution function

Equality holds when $\eta = x$



106

IS-NMF optimization with MM algorithm

[Kameoka+2006][Nakano+2010][Févotte+2011]

- Objective function to be minimized:

$$\mathcal{D}_{\mathrm{IS}}(\mathbf{T},\mathbf{V}) = \sum_{i,j}\left(\frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} + \log\sum_k t_{ik}v_{kj} - \cdots\right)$$

- Reciprocal function is convex in positive domain

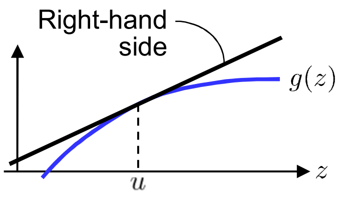
$$\frac{1}{\sum_k z_k} = \frac{1}{\sum_k \lambda_k \frac{z_k}{\lambda_k}} \leq \sum_k \lambda_k \frac{1}{\frac{z_k}{\lambda_k}} = \sum_k \frac{\lambda_k^2}{z_k}$$

Jensen's inequality ⇒ $\frac{1}{\sum_k t_{ik}v_{kj}} \leq \sum_k \frac{\lambda_{ijk}^2}{t_{ik}v_{kj}}$

- Logarithmic function is concave

$$g(u) \leq g'(u)(z-u) + g(u)$$

⇒ $\log\sum_k t_{ik}v_{kj} \leq$
$$\frac{1}{u_{ij}}\left(\sum_k t_{ik}v_{kj} - u_{ij}\right) + \log u_{ij}$$



IS-NMF optimization with MM algorithm

[Kameoka+2006][Nakano+2010][Févotte+2011]

- Majorizer

$$\mathcal{G}_{\mathrm{IS}}(\mathbf{T},\mathbf{V},\boldsymbol{\lambda},\mathbf{U})=\sum_{i,j}\left(\sum_k\frac{\lambda_{ijk}^2|x_{ij}|^2}{t_{ik}v_{kj}}+\frac{1}{u_{ij}}\left(\sum_k t_{ik}v_{kj}-u_{ij}\right)+\log u_{ij}-\cdots\right)$$

- Update rules for $\boldsymbol{\lambda}$ and $\mathbf{U}$

$$\lambda_{ijk}\leftarrow\frac{t_{ik}v_{kj}}{\sum_{k'}t_{ik'}v_{k'j}}\qquad u_{ij}\leftarrow\sum_k t_{ik}v_{kj}$$

- Update rules for $\mathbf{T}$ and $\mathbf{V}$

$$\frac{\partial\mathcal{G}_{\mathrm{IS}}}{\partial t_{ik}}=\sum_j\left(-\frac{\lambda_{ijk}^2|x_{ij}|^2}{t_{ik}^2v_{kj}}+\frac{v_{kj}}{u_{ij}}\right)=0\ \Leftrightarrow t_{ik}=\sqrt{\frac{\sum_j\lambda_{ijk}^2|x_{ij}|^2/v_{kj}}{\sum_j v_{kj}/u_{ij}}}$$

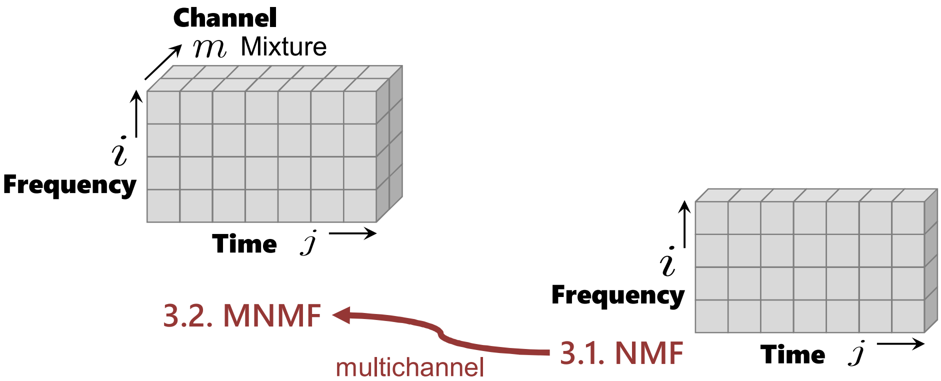
$$\frac{\partial\mathcal{G}_{\mathrm{IS}}}{\partial v_{kj}}=\sum_i\left(-\frac{\lambda_{ijk}^2|x_{ij}|^2}{t_{ik}v_{kj}^2}+\frac{t_{ik}}{u_{ij}}\right)=0\ \Leftrightarrow v_{kj}=\sqrt{\frac{\sum_i\lambda_{ijk}^2|x_{ij}|^2/t_{ik}}{\sum_i t_{ik}/u_{ij}}}$$

* Update rules of ME algorithm can be obtained by solving $\mathbf{T}$ and $\mathbf{V}$ that satisfies $\mathcal{G}_{\mathrm{IS}}(\mathbf{T},\mathbf{V},\boldsymbol{\lambda},\mathbf{U})=\mathcal{G}_{\mathrm{IS}}(\mathbf{T}',\mathbf{V}',\boldsymbol{\lambda},\mathbf{U})$

Tutorial structure

1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

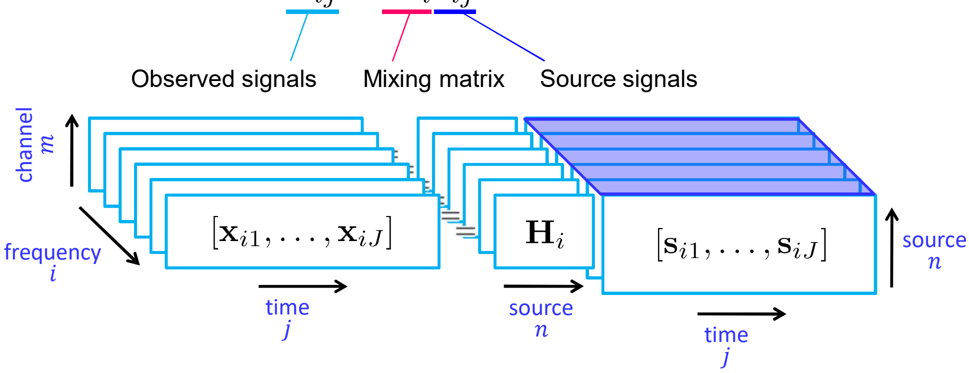
Tensor and sliced matrices



Extension to multichannel input

- Frequency-domain instantaneous mixture:

$$\mathbf{x}_{ij} = \mathbf{H}_i \mathbf{s}_{ij}$$



- Assume local Gaussian model (LGM) with source power spectrograms expressed using NMF (low-rank) model

111

Local Gaussian model

- Source model assuming each element of a complex spectrogram to independently follow a zero-mean complex Gaussian distribution with a different variance (power)

$$s_{ij,n} \sim \mathcal{N}_{\mathbb{C}}(s_{ij,n}|0, \sigma_{ij,n}^2)$$

variance

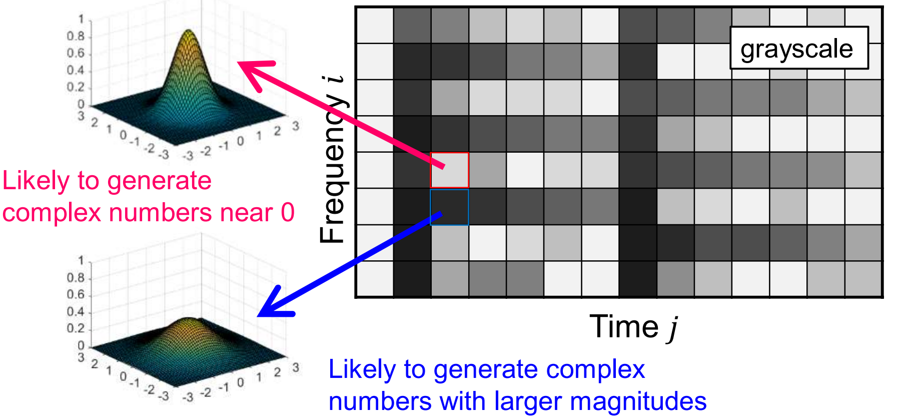
112

Local Gaussian model

- Source model assuming each element of a complex spectrogram to independently follow a zero-mean complex Gaussian distribution with a different variance (power)

$$s_{ij,n} \sim \mathcal{N}_{\mathbb{C}}(s_{ij,n}|0, \sigma_{ij,n}^2)$$

variance



113

Local Gaussian model

- Source model assuming each element of a complex spectrogram to independently follow a zero-mean complex Gaussian distribution with a different variance (power)

$$s_{ij,n} \sim \mathcal{N}_{\mathbb{C}}(s_{ij,n}|0, \sigma^2_{ij,n})$$ variance

  - Allows us to incorporate power spectrogram model in $\sigma^2_{ij,n}$
  - $\sigma^2_{ij,n} = \sum_k t_{ik,n} v_{kj,n}$ corresponds to NMF model

114

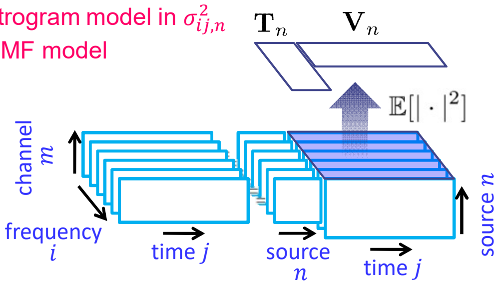
Local Gaussian model

- Source model assuming each element of a complex spectrogram to independently follow a zero-mean complex Gaussian distribution with a different variance (power)

$$s_{ij,n} \sim \mathcal{N}_{\mathbb{C}}(s_{ij,n}|0, \sigma^2_{ij,n})$$ variance

  - Allows us to incorporate power spectrogram model in $\sigma^2_{ij,n}$
  - $\sigma^2_{ij,n} = \sum_k t_{ik,n} v_{kj,n}$ corresponds to NMF model

$$\mathbf{x}_{ij} = \mathbf{H}_i \mathbf{s}_{ij} \qquad \mathbf{s}_{ij} = \begin{bmatrix} s_{ij,1} \\ \vdots \\ s_{ij,N} \end{bmatrix}$$



Likelihood

$$\mathbf{x}_{ij} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{x}_{ij} \middle| \mathbf{0}, \mathbf{H}_i \begin{bmatrix} \sigma^2_{ij,1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{ij,N} \end{bmatrix} \mathbf{H}_i^{\mathsf{H}}\right)$$

115

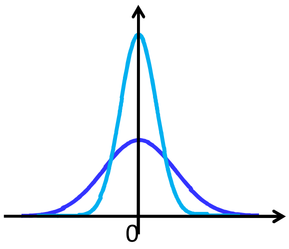
Super-Gaussianity of local Gaussian model

- Theorem

  The time average of zero-mean Gaussian distributions with time-varying variances is super-Gaussian.

  $$p(x) = \sum_n w_n \mathcal{N}(x; 0, v_n) \qquad \left(\sum_n w_n = 1\right)$$

- Proof omitted



Kurtosis becomes 0 if and only if all the variances are equal.

➡ Using local Gaussian models implies assuming source signals to follow super-Gaussian distributions

116

Multichannel NMF

[Ozerov&Févotte2010][Kameoka+2010][Sawada+2012]

- Probability density function of microphone array observations

$$\mathbf{x}_{ij} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{x}_{ij} \middle| \mathbf{0}, \mathbf{H}_i \begin{bmatrix} \sigma_{ij,1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{ij,N}^2 \end{bmatrix} \mathbf{H}_i^{\mathsf{H}}\right) \longleftarrow \text{local Gaussian model}$$

$$\sum_k t_{ik,n} v_{kj,n} \longleftarrow \text{NMF model}$$

log-likelihood

$$\mathcal{L}(\mathbf{H}, \mathbf{T}, \mathbf{V}) = \sum_{i,j} \{-\log\det \mathbf{\Phi}_{ij} - \mathbf{x}_{ij}^{\mathsf{H}} \mathbf{\Phi}_{ij}^{-1} \mathbf{x}_{ij}\}$$

where $\mathbf{\Phi}_{ij} = \mathbf{H}_i \begin{bmatrix} \sigma_{ij,1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{ij,N}^2 \end{bmatrix} \mathbf{H}_i^{\mathsf{H}}$ and $\sigma_{ij,n}^2 = \sum_k t_{ik,n} v_{kj,n}$

Assumption on mixing matrix
- None [Ozerov+2010][Sawada+2012]
  → Applicable for underdetermined system
- Invertible [Kameoka+2010][Kitamura+2015]
  → Specialized for determined system

Optimization
- EM algorithm [Ozerov+2010][Kameoka+2010]
- Auxiliary function approach [Sawada+2012][Kitamura+2015]

117

Multivariate extension of IS divergence

- Different expression of $\boldsymbol{\Phi}_{ij}$

$$\mathbf{x}_{ij} = \mathbf{H}_i \mathbf{s}_{ij} = \sum_n \mathbf{h}_{i,n} \underline{s_{ij,n}}$$

$$s_{ij,n} \sim \mathcal{N}_\mathbb{C}(s_{ij,n} | 0, \sum_k t_{ik,n} v_{kj,n})$$

$$\sim \mathcal{N}_\mathbb{C}\left(\mathbf{x}_{ij} \middle| 0, \sum_n \underbrace{\mathbf{h}_{i,n}\mathbf{h}_{i,n}^\mathsf{H}}_{} \sum_k t_{ik,n} v_{kj,n}\right)$$

$\mathsf{H}_{i,n}$: spatial property of source n

$$\hat{\mathsf{X}}_{ij} (= \boldsymbol{\Phi}_{ij})$$

- Multivariate extension of IS divergence
  (a.k.a. log-determinant divergence)

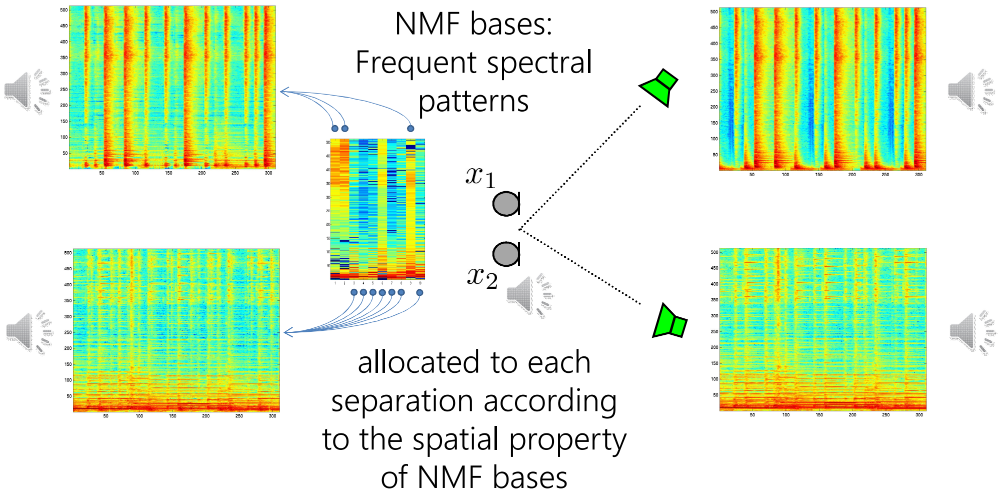
$$\mathcal{L}(\mathbf{H}, \mathbf{T}, \mathbf{V}) = \sum_{i,j} \{-\log \det \hat{\mathsf{X}}_{ij} - \mathbf{x}_{ij}^\mathsf{H} \hat{\mathsf{X}}_{ij}^{-1} \mathbf{x}_{ij}\}$$

$$= \sum_{i,j} \{-\log \det \underline{\hat{\mathsf{X}}_{ij}} - \mathrm{tr}(\underline{\hat{\mathsf{X}}_{ij}^{-1}} \underline{\mathbf{x}_{ij}\mathbf{x}_{ij}^\mathsf{H}})\}$$

118

Treating NMF bases as individual sources

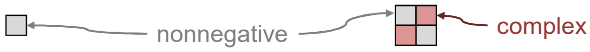
Special case where $\hat{\mathsf{X}}_{ij} = \sum_k \mathsf{H}_{ik} t_{ik} v_{kj}$



119

Nonnegativity: from scalar to vector

	Scalar	Vector
Observation	x	$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}$
Nonnegative representation	$\lvert x \rvert^2 = x x^*$	$\mathsf{X} = \mathbf{x}\mathbf{x}^{\mathsf{H}} = \begin{bmatrix} \lvert x_1 \rvert^2 & \dots & x_1 x_M^* \\ \vdots & \ddots & \vdots \\ x_M x_1^* & \dots & \lvert x_M \rvert^2 \end{bmatrix}$
Low-rank model	$\hat{x}_{ij} = \sum_{k=1}^{K} t_{ik} v_{kj}$	$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K} \mathsf{H}_{ik} t_{ik} v_{kj}$



- Hermitian positive-semidefinite matrix
  $\mathsf{H} = \mathsf{H}^H$ All eigenvalues are nonnegative
- Spatial property of the *k*-th NMF basis at frequency *i*

120

Optimization problem of multichannel NMF

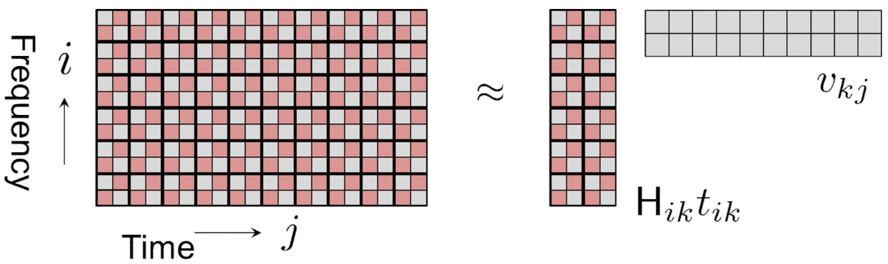
- Objective function to be minimized

$$\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \sum_{i,j} d_*(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij})$$

typically IS divergence

with $\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K} \mathsf{H}_{ik} t_{ik} v_{kj}$

$$d_{IS}(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) = \mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) - \log\det \mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1} - M$$



ij-element wise:

121

Auxiliary function design

[Sawada+2012, Higuchi+2014, Yoshii+2013]

Objective function

$$\mathcal{L}(\mathbf{T},\mathbf{V},\mathbf{H}) = \sum_{i,j}\left[\mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) + \log\det\hat{\mathsf{X}}_{ij}\right] \quad \text{with} \quad \hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K}\mathsf{H}_{ik}t_{ik}v_{kj}$$

$$\mathrm{tr}\left[\left(\sum_k \mathbf{X}_k\right)^{-1}\right] \le \sum_k \mathrm{tr}\left(\mathsf{R}_k\mathbf{X}_k^{-1}\mathsf{R}_k\right)$$

matrix extension ⬆

Jensen's inequality applied to reciprocal function

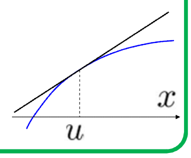
$$\frac{1}{\sum_k x_k} = \frac{1}{\sum_k \lambda_k \frac{x_k}{\lambda_k}} \le \sum_k \frac{\lambda_k^2}{x_k}$$

$$\log\det\hat{\mathsf{X}} \le \mathrm{tr}(\hat{\mathsf{X}}\mathsf{U}^{-1}) + \log\det\mathsf{U} - M$$

matrix extension ⬆

1st order Taylor expansion of logarithmic function

$$\log x \le \frac{x}{u} + \log u - 1$$



Auxiliary function

$$\mathcal{L}^{+}(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{R},\mathbf{U}) = \sum_{i,j}\left[\sum_k \frac{\mathrm{tr}(\mathsf{X}_{ij}\mathsf{R}_{ijk}\mathsf{H}_{ik}^{-1}\mathsf{R}_{ijk})}{t_{ik}v_{kj}} + \mathrm{tr}\left(\sum_k t_{ik}v_{kj}\mathsf{H}_{ik}\mathsf{U}_{ij}^{-1}\right) + \log\det\mathsf{U}_{ij} - M\right]$$

122

Multiplicative update rules

Multichannel IS-NMF

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}{\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}}$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}{\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}}$$

$\mathsf{H}_{ik}\mathsf{A}\mathsf{H}_{ik} = \mathsf{B}$ Algebraic Riccati equation

$$\mathsf{A} = \sum_j t_{ik}v_{kj}\hat{\mathsf{X}}_{ij}^{-1}$$

$$\mathsf{B} = \mathsf{H}_{ik}\left(\sum_j t_{ik}v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\right)\mathsf{H}_{ik}$$

IS-NMF

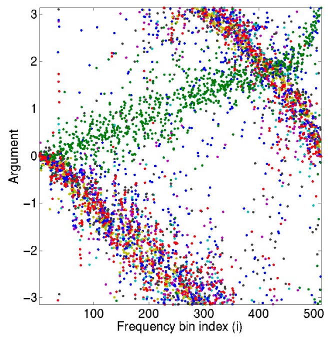
$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}\frac{x_{ij}}{\hat{x}_{ij}}}{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}}}$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}\frac{x_{ij}}{\hat{x}_{ij}}}{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}}}$$

123

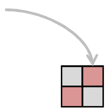
Learned spatial property example

Basis-wise

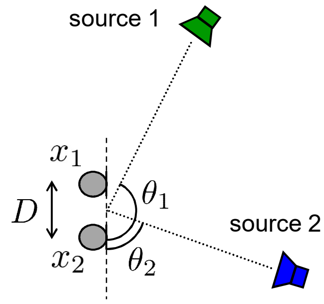


$\arg([\mathsf{H}_{ik}]_{12})$

$k = 1, \ldots, 10$



The 10 bases seem to form 2 clusters, each of which corresponds to each source.



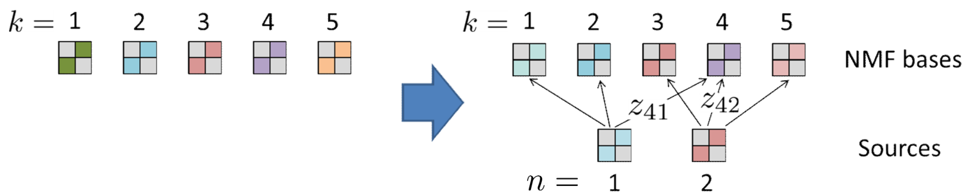
Inter-channel phase difference of source k becomes $\omega_i \frac{D \cos \theta_k}{C}$ where C is speed of sound

124

Clustering NMF bases for sources

- Modify the modeling of the spatial property

H_{ik} $\quad\quad$ $\sum_n z_{kn} \mathsf{H}_{in}$



- $\hat{\mathsf{X}}_{ij}$ with source-wise spatial property

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K} \sum_{n=1}^{N} z_{kn} \mathsf{H}_{in} t_{ik} v_{kj}$$

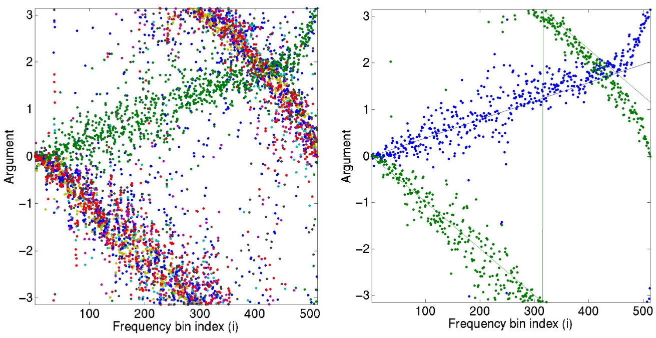
Similar multiplicative updates derived

125

Learned spatial property example

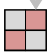
Basis-wise

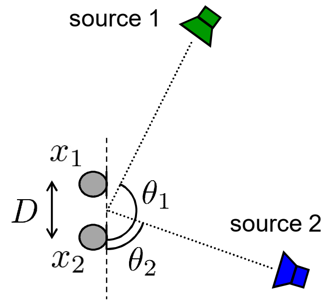
Source-wise



$\arg([\mathsf{H}_{ik}]_{12})$ $\arg([\mathsf{H}_{in}]_{12})$

$k = 1, \ldots, 10$ $n = 1, 2$



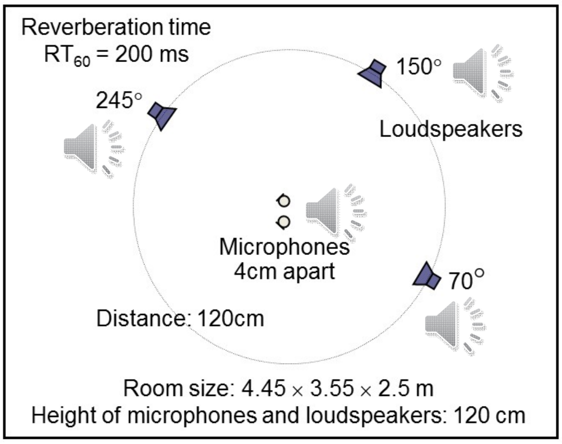


Inter-channel phase difference of source k becomes $\omega_i \frac{D \cos \theta_k}{C}$ where C is speed of sound

126

3 music parts separation example

- 3 sources and 2 microphones (underdetermined case)



The computational burden was heavy: it took 838.30 seconds for separating 24-second mixture.

4 examples in total can be found at
http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/mchnmf/

127

Determined multichannel NMF

[Kameoka+2010]

- Special case of multichannel NMF where mixing matrix $\mathbf{H}_i$ is invertible
- When $\mathbf{W}_i = \mathbf{H}_i^{-1}$, we can write $\mathbf{W}_i \mathbf{x}_{ij} = \mathbf{s}_{ij}$ and so

$$\mathcal{L}(\mathbf{H}, \mathbf{T}, \mathbf{V}) = \sum_{i,j} \{-\log\det \mathbf{\Phi}_{ij} - \mathbf{x}_{ij}^{\mathsf{H}} \mathbf{\Phi}_{ij}^{-1} \mathbf{x}_{ij}\}$$

$$= \sum_{i,j} \left(\log\det \mathbf{W}_i^{\mathsf{H}} \mathbf{\Sigma}_{ij}^{-1} \mathbf{W}_i - \mathbf{x}_{ij}^{\mathsf{H}} \mathbf{W}_i^{\mathsf{H}} \mathbf{\Sigma}_{ij}^{-1} \mathbf{W}_i \mathbf{x}_{ij} \right)$$

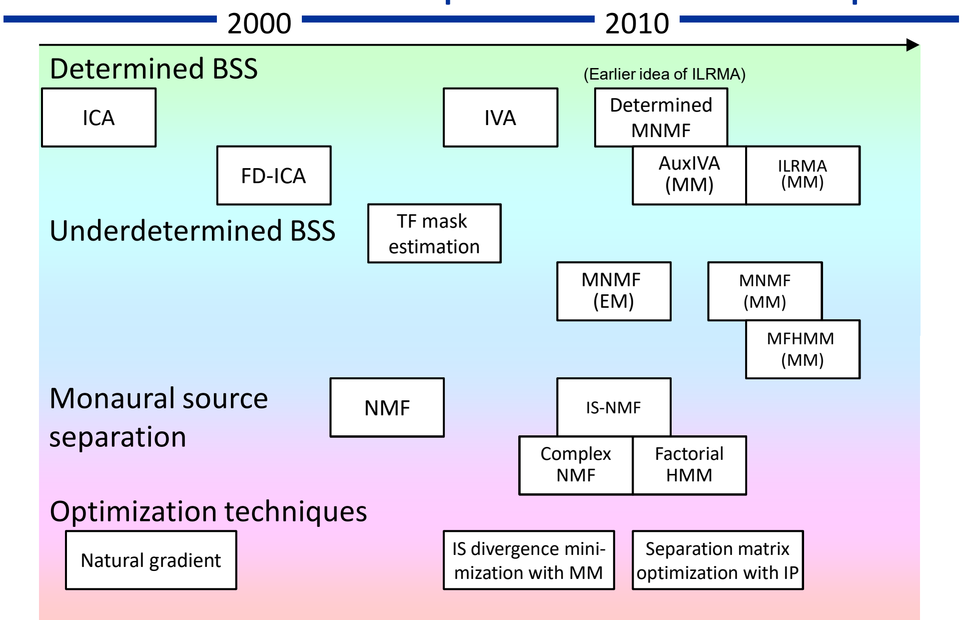
$$= \sum_{i,j} \left\{ 2\log\det \mathbf{W}_i - \sum_n \left(\log \sigma_{ij,n}^2 + \frac{|s_{ij,n}|^2}{\sigma_{ij,n}^2} \right) \right\}$$

where $\mathbf{\Sigma}_{ij} = \begin{bmatrix} \sigma_{ij,1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{ij,N}^2 \end{bmatrix}$ and $\sigma_{ij,n}^2 = \sum_k t_{ik,n} v_{kj,n}$

- Earlier idea of Independent Low-Rank Analysis (ILRMA)

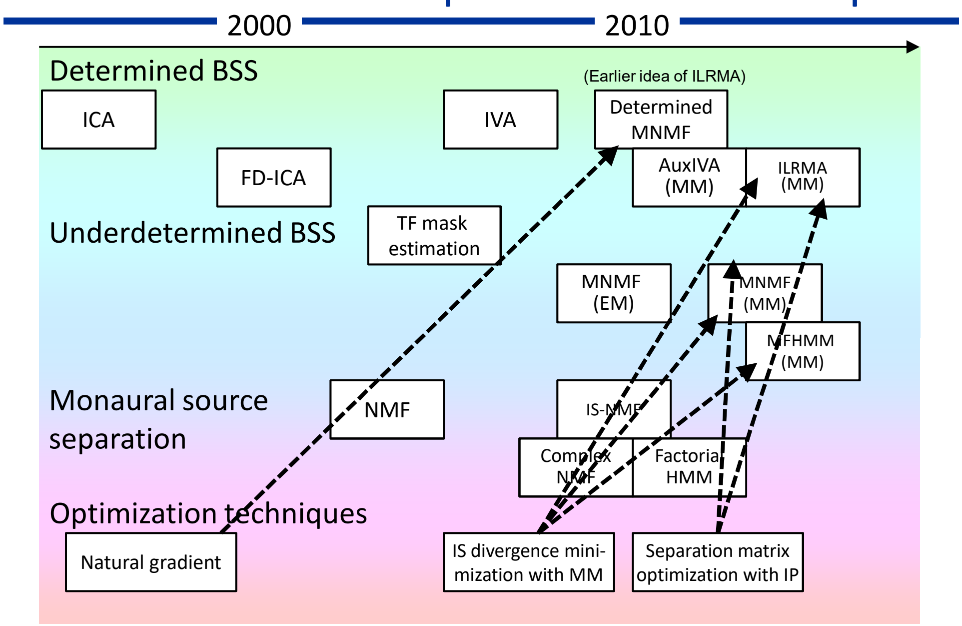
128

BSS methods and optimization techniques



129

BSS methods and optimization techniques



130

Categorization of LGM-based BSS methods

Generative model $\mathbf{x}_{i,j} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}_{i,j} | \boldsymbol{\mu}_{i,j}, \boldsymbol{\Phi}_{i,j})$

Method	$\boldsymbol{\mu}_{i,j}$	$\boldsymbol{\Phi}_{i,j}$
Attias (2003)	$\mathbf{0}$	$\sum_n \sum_{j'} \sigma^2_{i,j-j',n} \mathsf{H}_{i,j',n}$
Ozerov & Févotte (2010)	$\mathbf{0}$	$\sum_n \sigma^2_{i,j,n} \mathsf{H}_{i,n} + \mathsf{B}_i$
Duong et al. (2010)	$\mathbf{0}$	$\sum_n \sigma^2_{i,j,n} \mathsf{H}_{i,n}$
Kameoka et al. (2010)	$\mathbf{W}_i^{-1} \sum_{j'=1}^{J'-1} \mathbf{F}_{i,j'} \mathbf{x}_{i,j-j'}$	$\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_{i,j} \mathbf{W}_i^{-1}$
Yoshioka et al. (2011)	$\mathbf{W}_i^{-1} \sum_{j'=1}^{J'-1} \mathbf{F}_{i,j'} \mathbf{x}_{i,j-j'}$	$\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_{i,j} \mathbf{W}_i^{-1}$
Ono et al. (2012)	$\mathbf{0}$	$\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_{i,j} \mathbf{W}_i^{-1}$
Sawada et al. (2013)	$\mathbf{0}$	$\sum_n \sigma^2_{i,j,n} \mathsf{H}_{i,n}$
Higuchi et al. (2014)	$\mathbf{0}$	$\sum_n \sum_{j'} \sigma^2_{i,j-j',n} \mathsf{H}_{i,j',n}$
Kitamura et al. (2015)	$\mathbf{0}$	$\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_{i,j} \mathbf{W}_i^{-1}$
López et al. (2015)	$\mathbf{0}$	$\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_{i,j} \mathbf{W}_i^{-1}$
Adiloğlu & Vincent (2016)	$\mathbf{0}$	$\sum_n \sigma^2_{i,j,n} \mathsf{H}_{i,n}$
Kounades-Bastian et al. (2016)	$\mathbf{0}$	$\sum_n \sigma^2_{i,j,n} \mathsf{H}_{i,j,n}$

131

Categorization of LGM-based BSS methods

Constraints on $\sigma^2_{i,j,n}$

Method	$\sigma^2_{i,j,n}$
Attias (2003)	$t_{i,k_{j,n},n}$ where $k_{j,n} \sim \boldsymbol{\pi}$
Ozerov & Févotte (2010)	$\sum_k t_{i,k,n} v_{k,j,n}$
Duong et al. (2010)	$\sigma^2_{i,j,n}$
Kameoka et al. (2010)	$\sum_{k,l} t_{i,k,n} a_{i,l,n} v_{k,j,n}$
Yoshioka et al. (2011)	$v_{j,n} a_{i,j,n}$
Ono et al. (2012)	$v_{j,n}$
Sawada et al. (2013)	$\sum_k z_{k,n} t_{i,k} v_{k,j}$
Higuchi et al. (2014)	$t_{i,k_{j,n},n} v_{j,n}$ where $k_{j,n} \sim \boldsymbol{\pi}$
Kitamura et al. (2015)	$\sum_k z_{k,n} t_{i,k} v_{k,j}$
López et al. (2015)	$\sigma^2_{i,j,n}$
Adiloğlu & Vincent (2016)	$(\sum_k t^{\mathrm{e}}_{i,k,n} v^{\mathrm{e}}_{k,j,n})(\sum_{k'} t^{\mathrm{f}}_{i,k',n} v^{\mathrm{f}}_{k',j,n})$
Kounades-Bastian et al. (2016)	$\sum_k t_{i,k,n} v_{k,j,n}$

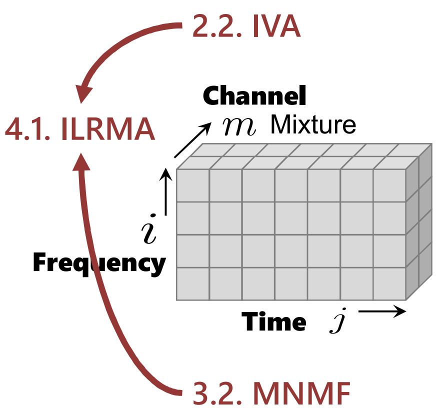
132

Tutorial structure

1. **Introduction**
   1. Separation of audio/speech signals
   2. Live demonstration
2. **ICA and IVA**
   1. ICA: Independent Component Analysis
   2. IVA: Independent Vector Analysis
3. **NMF**
   1. NMF: Nonnegative Matrix Factorization
   2. MNMF: Multichannel NMF
4. **ILRMA**
   1. ILRMA: Independent Low-Rank Matrix Analysis

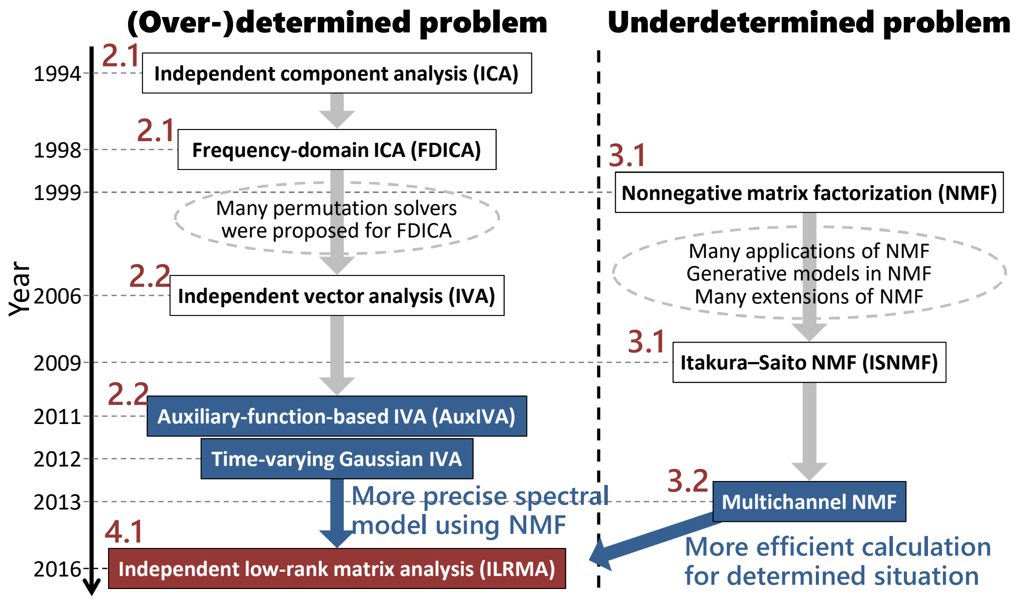
133

Tensor and sliced matrices



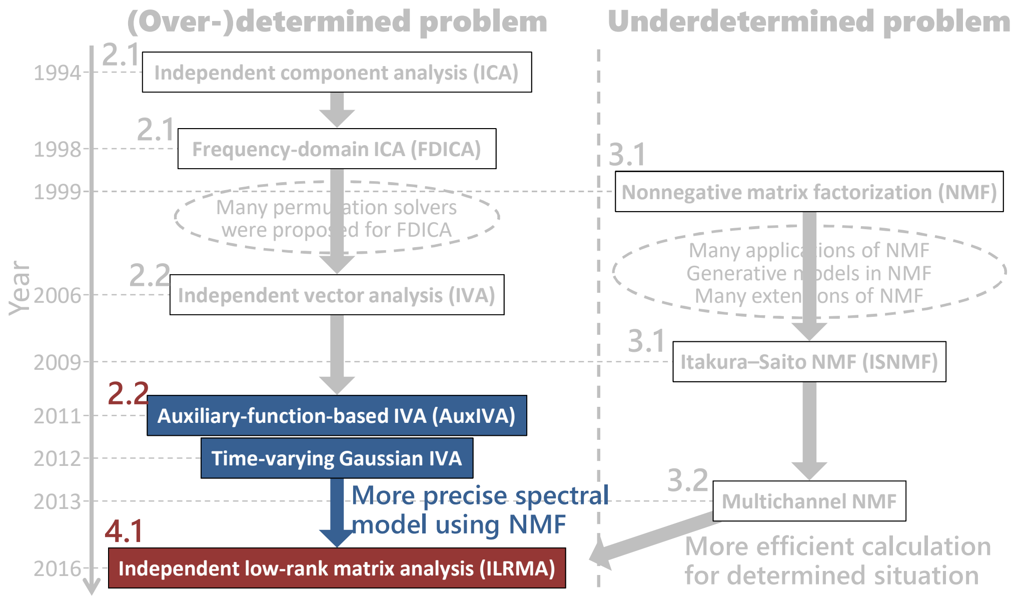
134

Historical development



135

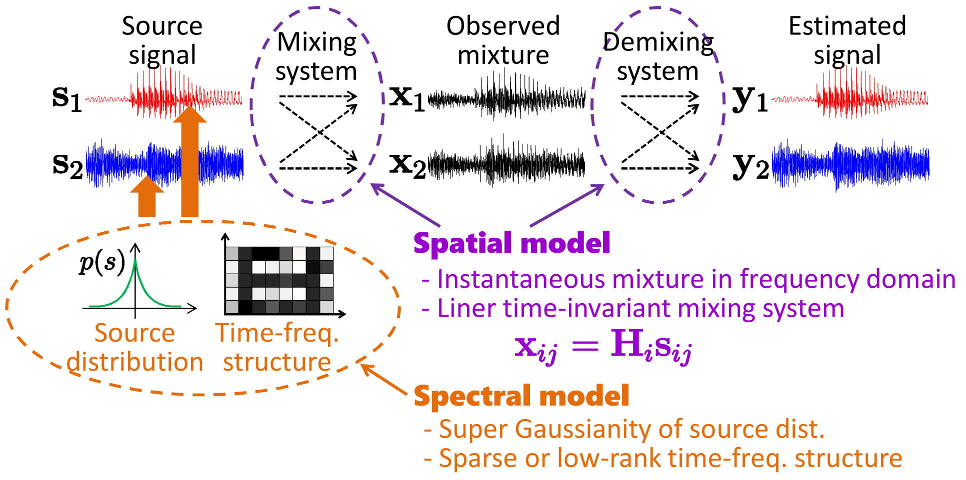
Historical development



136

Spatial and spectral models in BSS

- Spatial model: assumption of mixing/demixing system
- Spectral model: assumption for each source



137

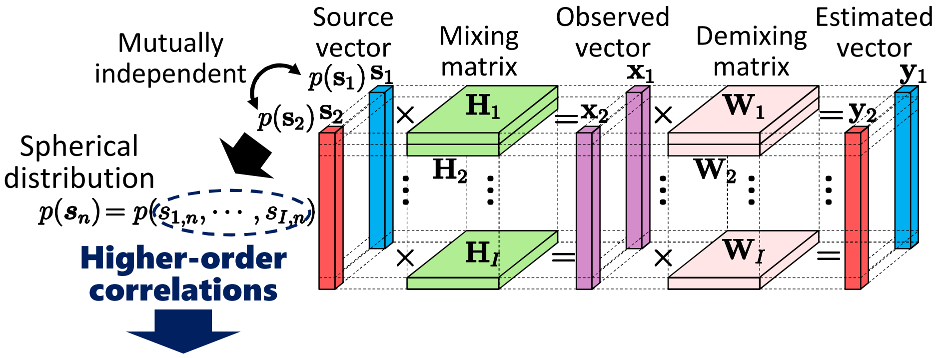
IVA revisited: model

- IVA extends ICA to multivariate probabilistic model

[Hiroe+, 2006], [Kim+, 2006], [Kim+, 2007]

Frequency-domain ICA: frequency **scalar** random variables

IVA: frequency **vector** random variables

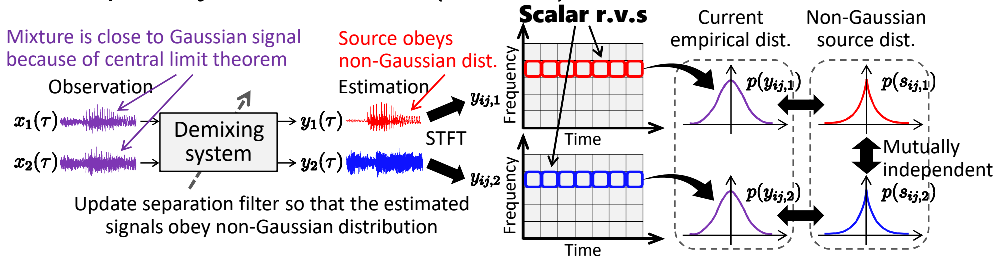


Permutation-problem-free estimation of $\mathbf{W}_i$

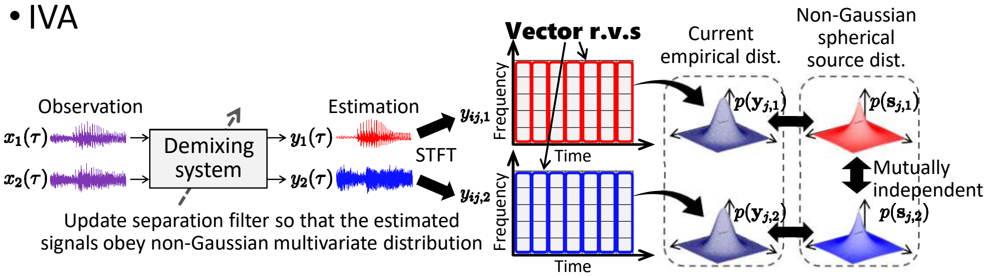
138

IVA revisited: compared to FDICA

- Frequency-domain ICA (FDICA)



- IVA



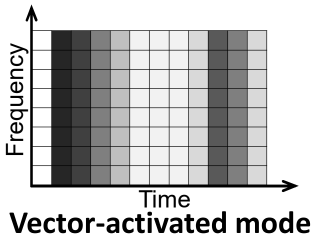
139

Extension of vector source model in IVA

- Frequency vector spectral model (IVA)

  Co-occurrence among **frequency bins** of each source

Extend vector model to low-rank matrix model



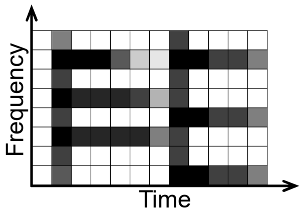
Vector-activated model

- NMF spectral model

  Co-occurrence among **time-frequency slots** of each source with a low-rank structure

More precise representation of time-frequency structure

Incrementation of frequency bases



Low-rank model

140

ISNMF revisited: low-rank spectral model

- Itakura–Saito NMF (ISNMF) [Févotte+, 2009]

$$\mathcal{D}_{\mathrm{IS}}\left(|\mathbf{X}|^{\cdot 2} \| \mathbf{TV}\right) = \sum_{i,j}\left[\frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} + \log\sum_k t_{ik}v_{kj}\right]$$

Minimization of the above is equivalent to a maximum likelihood (ML) estimation with a following generative model:

At each time-frequency slot, complex-valued component x_{ij} obeys

Variance

$$x_{ij} \sim \mathcal{N}_c\left(x_{ij}|0, \sum_k t_{ik}v_{kj}\right) = \frac{1}{\pi\sum_k t_{ik}v_{kj}}\exp\left(-\frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}}\right)$$

If x_{ij} can be decomposed as $x_{ij} = \sum_k c_{ijk}$, then

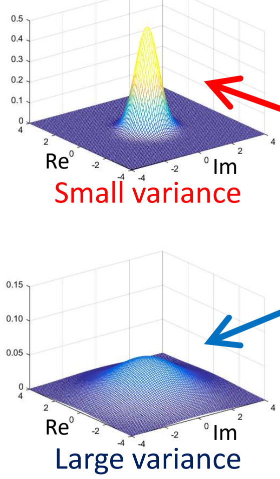
$$c_{ijk} \sim \mathcal{N}_c\left(c_{ijk}|0, t_{ik}v_{kj}\right)$$

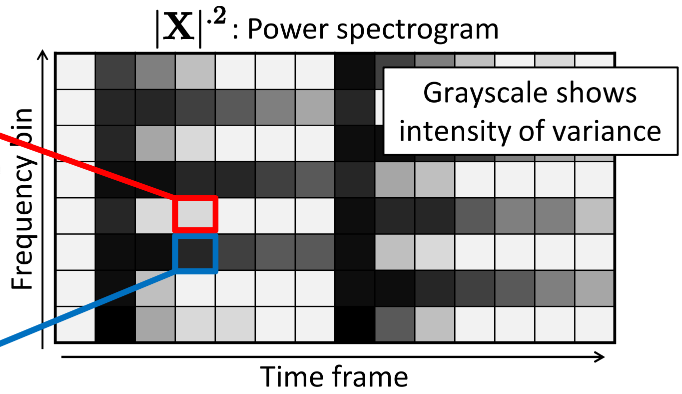
Parameters are also decomposed

141

ISNMF revisited: low-rank spectral model

- Itakura–Saito NMF (ISNMF) [Févotte+, 2009]





NOTE: $p(\mathbf{X}) = \prod_{i,j} p(x_{ij}; 0, \sum_k t_{ik} v_{kj})$ is **non-Gaussian** because the variance $\sum_k t_{ik} v_{kj}$ can fluctuate along time and frequency

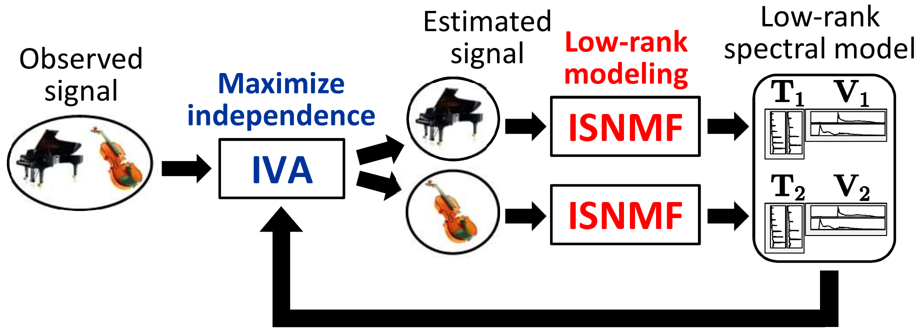
142

ILRMA: unified method of IVA and NMF

- Independent low-rank matrix analysis (ILRMA) [Kitamura+, 2016]

  Unification of

  1. Estimation of demixing matrix $\mathbf{W}_i$ (**IVA spatial model**)
  2. Low-rank approximation using $\mathbf{T}_n \mathbf{V}_n$ (**ISNMF spectral model**)



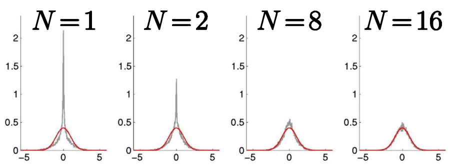
ILRMA: unified method of IVA and NMF

- Independent low-rank matrix analysis (ILRMA) [Kitamura+, 2016]

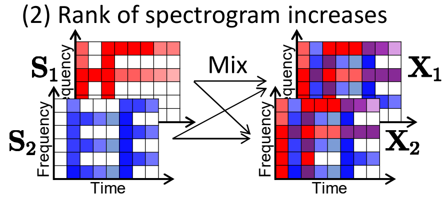
  Motivation of ILRMA?

When sources are mixed…

(1) Mixture dist. approaches to Gaussian



(2) Rank of spectrogram increases



To separate the sources…

Maximize non-Gaussianity
(using ICA or IVA theory)

Restrict rank of separated signal
(using NMF theory)

Low-rank assumption can avoid the permutation problem

Cost function in ILRMA

- Cost function in FDICA, IVA, or ILRMA

Source distribution (spectral model)

$$\mathcal{L} = -2J \sum_i \log |\det \mathbf{W}_i| - \sum_n \log p(\mathbf{Y}_n)$$

FDICA (Laplace)
e.g., [Sawada+, 2003]

$$p(\mathbf{Y}_n) = \prod_{i,j} \alpha \exp(-|y_{ij,n}|)$$

Separable for frequency

IVA (spherical Laplace)
[Hiroe+, 2006], [Kim+, 2006]

$$p(\mathbf{Y}_n) = \prod_j \alpha \exp(-\|\mathbf{y}_{j,n}\|_2)$$

$$\mathbf{y}_{j,n} = \begin{pmatrix} y_{1j,n} \\ \vdots \\ y_{Ij,n} \end{pmatrix}$$

Frequency vector

IVA (time-varying Gauss)
[Ono+, 2012]

$$p(\mathbf{Y}_n) = \prod_j \frac{\alpha}{\sigma_{j,n}^2} \exp\left(-\frac{\|\mathbf{y}_{j,n}\|_2^2}{\sigma_{j,n}^2}\right)$$

ILRMA (ISNMF model)
[Kitamura+, 2016]

$$p(\mathbf{Y}_n) = \prod_{i,j} \frac{1}{\pi \sum_k t_{ik,n} v_{kj,n}} \exp\left(-\frac{|y_{ij,n}|^2}{\sum_k t_{ik,n} v_{kj,n}}\right)$$

145

Cost function in ILRMA

- Cost function in ILRMA

Demixing matrix: $\mathbf{W}_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,N})^{\mathrm{H}}$ Estimated signal: $y_{ij,n} = \mathbf{w}_{i,n}^{\mathrm{H}}\mathbf{x}_{ij}$

$$\mathcal{L} = -2J\sum_i \log|\det \mathbf{W}_i| + \sum_{i,j,n}\left[\frac{|y_{ij,n}|^2}{\sum_k t_{ik,n}v_{kj,n}} + \log\sum_k t_{ik,n}v_{kj,n}\right]$$

Cost function in **time-varying Gaussian IVA** (estimates demixing matrix)

Cost function in **ISNMF** (estimates low-rank spectral structure)

Spatial model update: AuxIVA
Spectral model update: ISNMF
➡ All the variables are alternatively updated

146

Update rule of parameters in ILRMA

- Maximum-likelihood-based update rules

Alternatively update both models

Spatial model (demixing matrix)

Update demixing filter [Ono+, 2011]

$$\mathbf{V}_{i,n} = \frac{1}{J}\sum_j \frac{1}{r_{ij,n}}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{H}}$$

$$\mathbf{w}_{i,n} \leftarrow (\mathbf{W}_i\mathbf{V}_{i,n})^{-1}\mathbf{e}_n$$

$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n}\left(\mathbf{w}_{i,n}^{\mathrm{H}}\mathbf{V}_{i,n}\mathbf{w}_{i,n}\right)^{-\frac{1}{2}}$$

Update estimated signal

$$y_{ij,n} = \mathbf{w}_{i,n}^{\mathrm{H}}\mathbf{x}_{ij}$$

Spectral model (NMF variables)

Update NMF parameters [Nakano+, 2010]

$$t_{ik,n} \leftarrow t_{ik,n}\sqrt{\frac{\sum_j |y_{ij,n}|^2 v_{kj,n}\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{-2}}{\sum_j v_{kj,n}\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{-1}}}$$

$$v_{kj,n} \leftarrow v_{kj,n}\sqrt{\frac{\sum_j |y_{ij,n}|^2 t_{ik,n}\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{-2}}{\sum_j t_{ik,n}\left(\sum_{k'} t_{ik',n}v_{k'j,n}\right)^{-1}}}$$

Update estimated variance

$$r_{ij,n} = \sum_k t_{ik,n}v_{kj,n}$$

$\mathbf{e}_n$: one-hot vector with one at nth element

See also (pseudo code)
http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf

147

Clustering NMF bases for sources

- NMF bases should be automatically clustered into each source

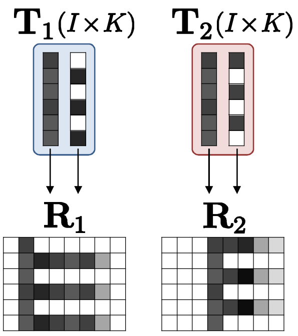
$$r_{ij,n} = \sum_k t_{ik,n} v_{kj,n}$$

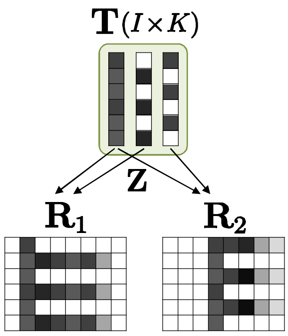
Fixed number of bases for each source

Introduce partitioning function

$$r_{ij,n} = \sum_k z_{kn} t_{ik} v_{kj}$$

Adaptive number of bases for each source





148

Update rule of parameters in ILRMA

- Maximum-likelihood-based update rules with $\mathbf{Z}$

Alternatively update both models

Spatial model
(demixing matrix)

Update demixing filter [Ono+, 2011]

$$\mathbf{V}_{i,n} = \frac{1}{J} \sum_j \frac{1}{r_{ij,n}} \mathbf{x}_{ij} \mathbf{x}_{ij}^{\mathrm{H}}$$

$$\mathbf{w}_{i,n} \leftarrow (\mathbf{W}_i \mathbf{V}_{i,n})^{-1} \mathbf{e}_n$$

$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n} \left(\mathbf{w}_{i,n}^{\mathrm{H}} \mathbf{V}_{i,n} \mathbf{w}_{i,n} \right)^{-\frac{1}{2}}$$

Update estimated signal

$$y_{ij,n} = \mathbf{w}_{i,n}^{\mathrm{H}} \mathbf{x}_{ij}$$

$\mathbf{e}_n$: one-hot vector with one at nth element

Spectral model
(NMF variables)

Update NMF parameters [Nakano+, 2010]

$$z_{kn} \leftarrow z_{kn} \sqrt{\frac{\sum_{i,j} |y_{ij,n}|^2 t_{ik} v_{kj} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-2}}{\sum_{i,j} t_{ik} v_{kj} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-1}}}$$

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{j,n} |y_{ij,n}|^2 z_{kn} v_{kj} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-2}}{\sum_{j,n} z_{kn} v_{kj} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-1}}}$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,n} |y_{ij,n}|^2 z_{kn} t_{ik} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-2}}{\sum_{i,n} z_{kn} t_{ik} \left(\sum_{k'} z_{k'n} t_{ik'} v_{k'j} \right)^{-1}}}$$

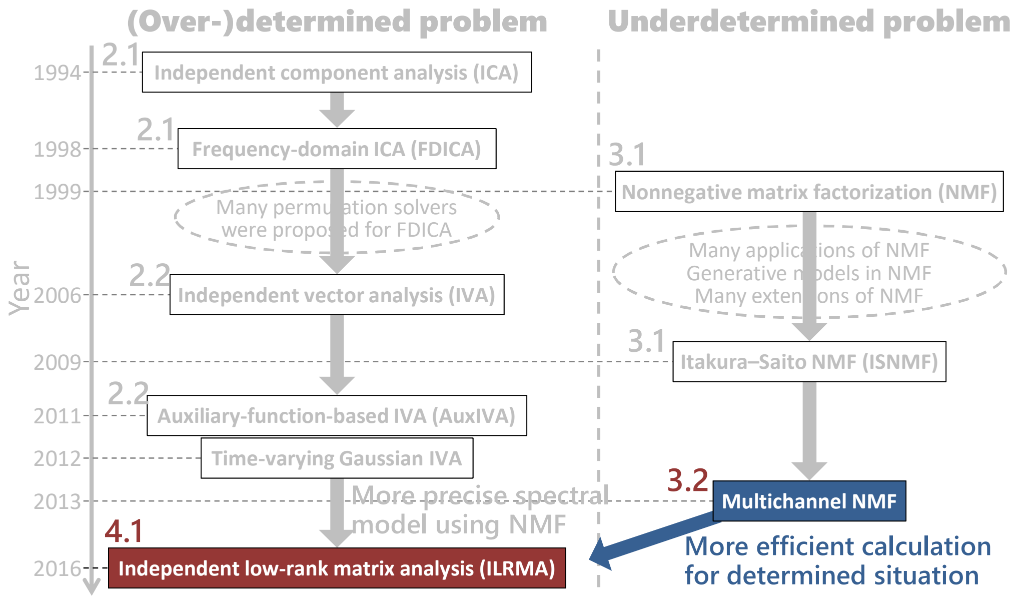
Update estimated variance

$$r_{ij,n} = \sum_k z_{kn} t_{ik} v_{kj}$$

See also (pseudo code)
http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf

149

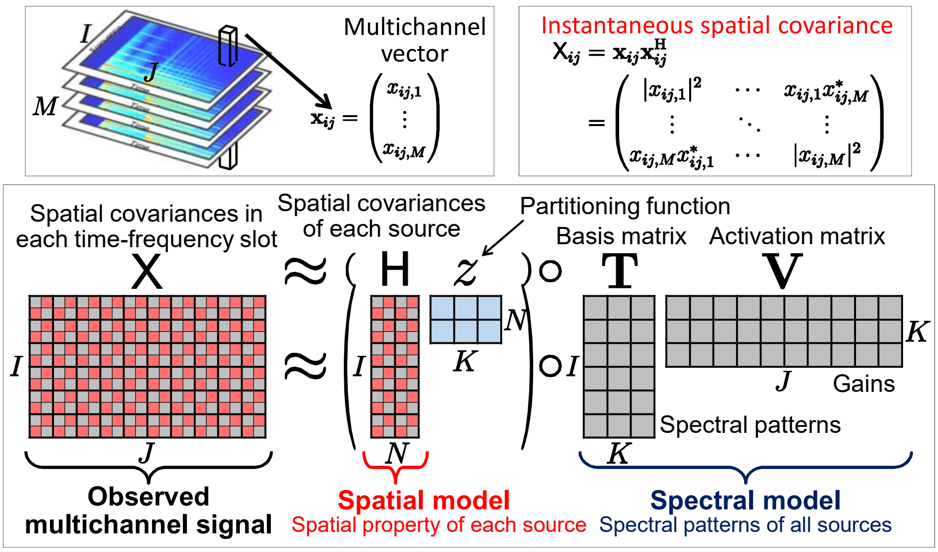
Historical development



150

Multichannel NMF revisited: model

- Multichannel NMF with full-rank covariance [Sawada+, 2013]



151

ILRMA and multichannel ISNMF

- Relationship b/w ILRMA and multichannel ISNMF?

  Source distribution: **same** $p(y_{ij,n}) = \mathcal{N}_c(y_{ij,n}; 0, \sum_k t_{ik,n} v_{kj,n})$

  Spatial model: **different**

ILRMA	**Multichannel ISNMF**
Instantaneous mixture in frequency domain	Mixture of full-rank covariances (and power spectrograms)
$\mathbf{x}_{ij} = \mathbf{H}_i \mathbf{s}_{ij}$	$\hat{\mathsf{X}}_{ij} = \sum_n \sum_k z_{kn} t_{ik} v_{kj} \mathsf{H}_{i,n}$

- Rank-1 spatial model [Duong+, 2010]

  Equivalent to instantaneous mixture

$$\mathsf{H}_{i,n} = \mathbf{h}_{i,n} \mathbf{h}_{i,n}^{\mathrm{H}}$$

$\mathbf{h}_{i,n}$: steering vector (column vector of mixing matrix)

$\mathbf{H}_i = (\mathbf{h}_{i,1} \ \cdots \ \mathbf{h}_{i,N})$

152

ILRMA and multichannel ISNMF

- Multichannel ISNMF with rank-1 spatial model

  Cost function of multichannel ISNMF

$$\mathcal{J} = \sum_{i,j} \left[\mathrm{tr}\left(\mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \right) + \log \det \hat{\mathsf{X}}_{ij} \right]$$

$$\hat{\mathsf{X}}_{ij} = \sum_k \sum_n \mathsf{H}_{i,n} z_{kn} t_{ik} v_{kj}$$

Substitute rank-1 spatial model $\mathsf{H}_{i,n} = \mathbf{h}_{i,n} \mathbf{h}_{i,n}^{\mathrm{H}}$ into $\hat{\mathsf{X}}_{ij}$

$$\begin{aligned} \hat{\mathsf{X}}_{ij} &= \sum_k \sum_n \mathbf{h}_{i,n} \mathbf{h}_{i,n}^{\mathrm{H}} z_{kn} t_{ik} v_{kj} \\ &= \sum_n \mathbf{h}_{i,n} \mathbf{h}_{i,n}^{\mathrm{H}} \sum_k z_{kn} t_{ik} v_{kj} \\ &= \mathbf{H}_i \mathbf{D}_{ij} \mathbf{H}_i^{\mathrm{H}} \end{aligned}$$

$$\mathbf{H}_i = (\mathbf{h}_{i,1} \ \cdots \ \mathbf{h}_{i,N}), \ \mathbf{D}_{ij} = \begin{pmatrix} \sum_k z_{k1} t_{ik} v_{kj} & 0 & \cdots & 0 \\ 0 & \sum_k z_{k2} t_{ik} v_{kj} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sum_k z_{kN} t_{ik} v_{kj} \end{pmatrix}$$

153

ILRMA and multichannel ISNMF

- Multichannel ISNMF with rank-1 spatial model

Substitute $\hat{\mathrm{X}}_{ij} = \mathbf{H}_i \mathbf{D}_{ij} \mathbf{H}_i^{\mathrm{H}}$ into the cost function of multichannel ISNMF

$$\mathcal{J} = \sum_{i,j} \left[\mathrm{tr}\left(\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{H}} (\mathbf{H}_i \mathbf{D}_{ij} \mathbf{H}_i^{\mathrm{H}})^{-1} \right) + \log\det \mathbf{H}_i \mathbf{D}_{ij} \mathbf{H}_i^{\mathrm{H}} \right]$$
$$= \sum_{i,j} \left[\mathrm{tr}\left(\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{H}} \mathbf{H}_i^{-\mathrm{H}} \mathbf{D}_{ij}^{-1} \mathbf{H}_i^{-1} \right) + \log(\det \mathbf{H}_i)(\det \mathbf{D}_{ij})(\det \mathbf{H}_i)^{\mathrm{H}} \right]$$

Transform the variables as $\mathbf{W}_i = \mathbf{H}^{-1}$ and $\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij}$

$$\mathcal{J} = \sum_{i,j} \left[\mathrm{tr}\left(\mathbf{W}_i^{-1} \mathbf{y}_{ij}\mathbf{y}_{ij}^{\mathrm{H}} \mathbf{W}_i^{-\mathrm{H}} \mathbf{W}_i^{\mathrm{H}} \mathbf{D}_{ij}^{-1} \mathbf{W}_i \right) + \log|\det \mathbf{H}_i|^2 + \log\det \mathbf{D}_{ij} \right]$$
$$= -2J \sum_i \log|\det \mathbf{W}_i| + \sum_{i,j} \left[\log \prod_n \sum_k z_{kn} t_{ik} v_{kj} + \mathrm{tr}\left(\mathbf{y}_{ij}\mathbf{y}_{ij}^{\mathrm{H}} \mathbf{D}_{ij}^{-1} \right) \right]$$
$$= -2J \sum_i \log|\det \mathbf{W}_i| + \sum_{i,j,n} \left[\sum_k z_{kn} t_{ik} v_{kj} + \frac{|y_{ij,n}|^2}{\sum_k z_{kn} t_{ik} v_{kj}} \right]$$

Cost in ILRMA

154

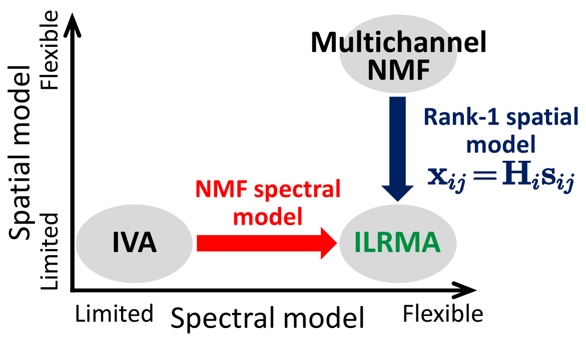
Summary of ILRMA

- From IVA side:

Introduce NMF spectral model (basis incrementation)

- From multichannel NMF side:

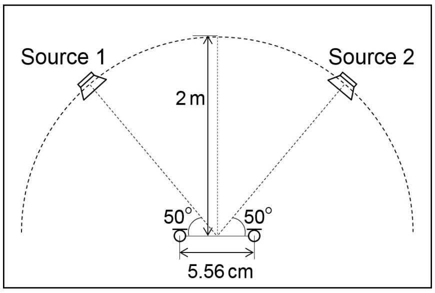
Introduce rank-1 spatial model (instantaneous mixture assumption)



Experiment

- Conditions

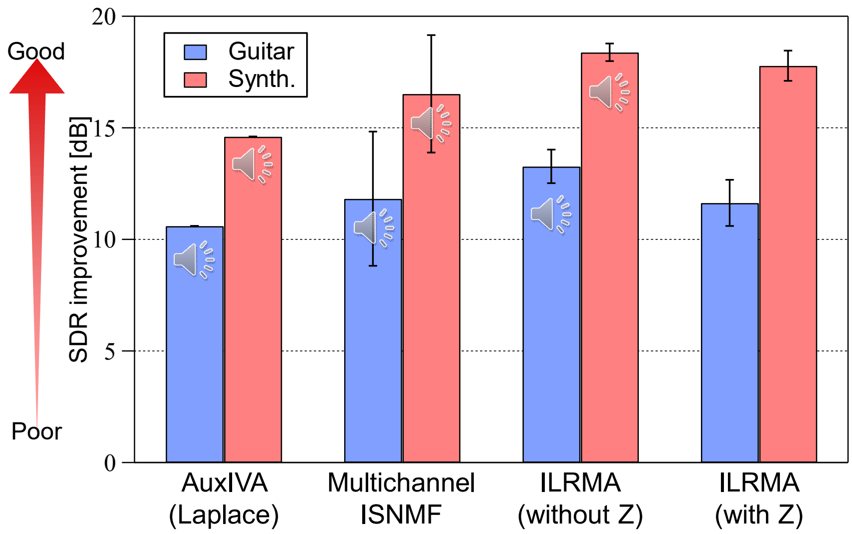
Source signals	Music signals obtained from SiSEC2011 Two microphones and two sources (determined)
Analysis window	512-ms-long Hamming window
Shift length	128 ms (1/4 shift)
Number of bases	30 per each source/60 for all sources
Evaluation score	Improvement ot signal-to-distortion ratio (SDR)



Impulse response E2A
(reverberation time: 300 ms)

Experiment

- Two source case (ultimate nz tour)

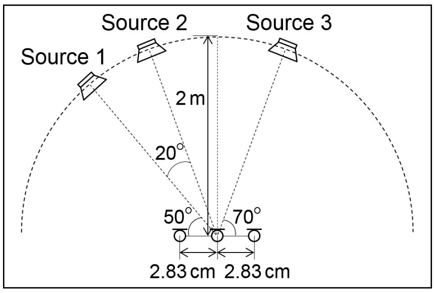


157

Experiment

- Conditions

Source signals	Music signals obtained from SiSEC2011 Three microphones and three sources (determined)
Analysis window	512-ms-long Hamming window
Shift length	128 ms (1/4 shift)
Number of bases	30 per each source/90 for all sources
Evaluation score	Improvement ot signal-to-distortion ratio (SDR)

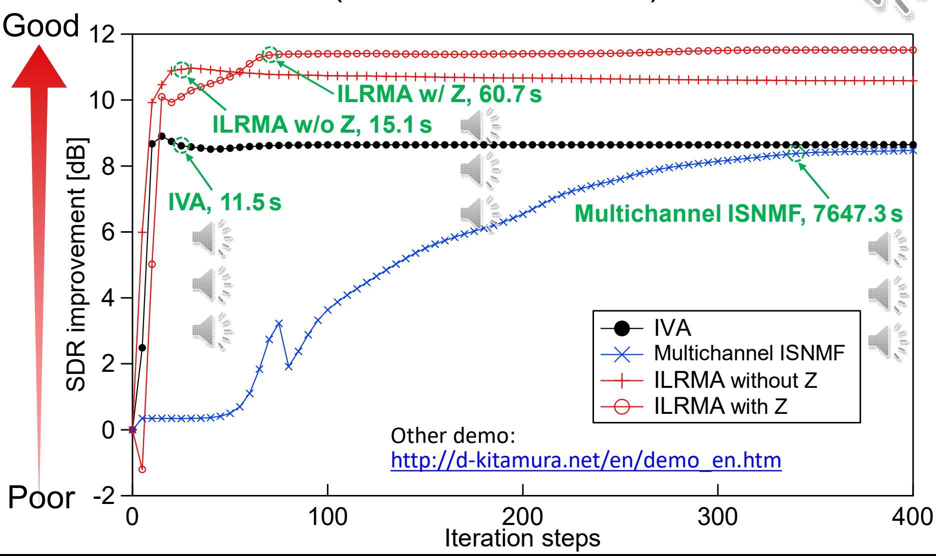


Impulse response E2A
(reverberation time: 300 ms)

158

Experiment

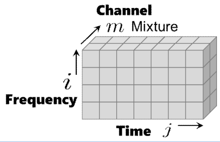
- Three source case (bearlin-roads, 14 s)



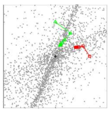
159

Conclusion

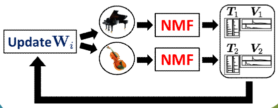
Complex-valued Tensor
- Multichannel audio mixture
- Frequency domain via STFT



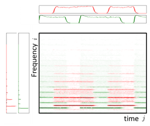
ICA, IVA
- Independence
- Super-Gaussian



ILRMA

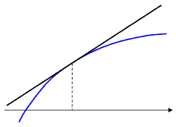


NMF, MNMF
- Spectral bases
- Low-rank



Component → Vector → Low-rank Matrix

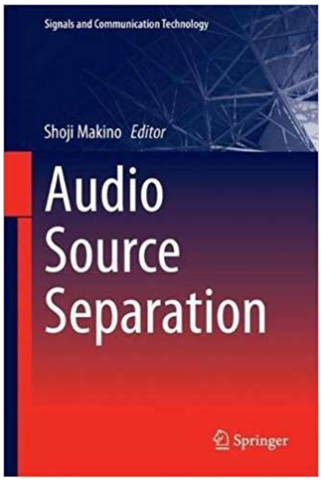
Auxiliary function-based optimization
- Fast convergence
- Problem dependent, but "cookbook" works



160

Advertisement: book chapters

- MNMF and ILRMA will be published from Springer in March, 2018!



Audio Source Separation (Signals and Communication Technology)
1st ed. 2018 by Ed. Shoji Makino

Ch. 5
General formulation of multichannel extensions of NMF variants;
Hirokazu Kameoka, Hiroshi Sawada, and Takuya Higuchi.

Ch. 6
Determined Blind Source Separation with Independent Low-Rank Matrix analysis;
Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari.

161

Thank you very much
for attending this tutorial !

Hiroshi Sawada

Nobutaka Ono

Hirokazu Kameoka

Daichi Kitamura

Selected references

Frequency-domain approach to convolutive BSS

- [Smaragdis, 1998], [Parra and Spence, 2000], [Schobben and Sommen, 2002], [Anemüller and Kollmeier, 2000], [Asano et al., 2003], [Saruwatari et al., 2003], [Saruwatari et al., 2006], [Yoshioka et al., 2008], [Duong et al., 2010]
- Permutation alignment [Sawada et al., 2004], [Sawada et al., 2007], [Sawada et al., 2011]
- Scaling adjustment to microphone observations [Cardoso, 1998], [Murata et al., 2001], [Matsuoka and Nakashima, 2001], [Takatani et al., 2004], [Mori et al., 2006]

ICA: Independent Component Analysis

- Information-maximization approach [Bell and Sejnowski, 1995]
- Maximum likelihood (ML) estimation [Cardoso, 1997]
- Natural gradient [Amari et al., 1996], [Cichocki and Amari, 2002]
- Equivariance property [Cardoso and Souloumiac, 1996]
- FastICA [Hyvärinen et al., 2001]
- Complex-valued ICA [Bingham and Hyvärinen, 2000], [Sawada et al., 2003]
- Auxiliary function based ICA [Ono and Miyabe, 2010]

IVA: Independent Vector Analysis

- Multivariate p.d.f. [Hiroe, 2006], [Kim et al., 2006], [Kim et al., 2007]
- FastIVA [Lee et al., 2006], [Lee et al., 2007]
- HEAD problem [Yeredor, 2009]
- Auxiliary function based IVA (AuxIVA) [Ono, 2011], [Ono, 2012b], [Ikeshita et al., 2017]
- Time-varying Gaussian p.d.f. [Ono, 2012a], [Ono et al., 2012]
- Supervised or model-based IVA [Ono et al., 2012], [Lopez et al., 2015], [Nesta and Koldovský, 2017]
- Online IVA [Kim, 2010], [Taniguchi et al., 2014], [Sunohara et al., 2017]

NMF: Nonnegative Matrix Factorization

- Auxiliary function based optimization for Euclidean distance NMF and generalized Kullback-Leibler divergence NMF [Lee and Seung, 1999], [Lee and Seung, 2001]
- EM-based optimization for Itakura-Saito divergence NMF [Févotte et al., 2009]
- Auxiliary function based optimization for Itakura-Saito divergence NMF [Kameoka et al., 2006], [Nakano et al., 2010], [Févotte and Idier, 2011]
- Auxiliary function based optimization for β divergence NMF [Nakano et al., 2010], [Févotte and Idier, 2011]
- Auxiliary function based optimization for sparse NMF [Kameoka et al., 2009]

Multi-channel NMF

- EM-based optimization [Ozerov and Févotte, 2010]
- Auxiliary function based optimization [Sawada et al., 2012], [Sawada et al., 2013], [Higuchi and Kameoka, 2014]

ILRMA: Independent Low-Rank Matrix Analysis

- Earlier idea (determined multi-channel NMF) [Kameoka et al., 2010]
- Multichannel NMF with rank-1 spatial model [Kitamura et al., 2015a]
- ILRMA [Kitamura et al., 2016], [Kitamura et al., 2018]
- Relaxation of Rank-1 spatial model [Kitamura et al., 2015b]
- Maximization-equalization algorithm [Mitsui et al., 2017b]
- Optimal window length [Kitamura et al., 2017]
- Based on Student's t-distribution [Mogami et al., 2017]
- With sparse regularization [Mitsui et al., 2017a]
- With spatial regularization [Mitsui et al., 2018]

Other references related to auxiliary function based optimization

- [Lange et al., 2000], [Hunter and Lange, 2000], [Hunter and Lange, 2004], [Ono et al., 2009], [Ono and Sagayama, 2010], [Févotte and Idier, 2011], [Yoshii et al., 2013], [Kameoka and Takamune, 2014], [Sun et al., 2017]

References

[Amari et al., 1996] Amari, S., Cichocki, A., and Yang, H. H. (1996). A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. The MIT Press.

[Anemüller and Kollmeier, 2000] Anemüller, J. and Kollmeier, B. (2000). Amplitude modulation decorrelation for convolutive blind source separation. In *Proc. ICA 2000*, pages 215–220.

[Asano et al., 2003] Asano, F., Ikeda, S., Ogawa, M., Asoh, H., and Kitawaki, N. (2003). Combined approach of array processing and independent component analysis for blind separation of acoustic signals. *IEEE Trans. Speech Audio Processing*, 11(3):204–215.

[Bell and Sejnowski, 1995] Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.

[Bingham and Hyvärinen, 2000] Bingham, E. and Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 10(1):1–8.

[Cardoso, 1997] Cardoso, J.-F. (1997). Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114.

[Cardoso, 1998] Cardoso, J.-F. (1998). Multidimensional independent component analysis. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 4, pages 1941–1944. IEEE.

[Cardoso and Souloumiac, 1996] Cardoso, J.-F. and Souloumiac, A. (1996). Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164.

[Cichocki and Amari, 2002] Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing*. John Wiley & Sons.

[Duong et al., 2010] Duong, N., Vincent, E., and Gribonval, R. (2010). Under-determined reverberant audio source separation using a full-rank spatial covariance model. *IEEE Trans. Audio, Speech, and Language Processing*, 18(7):1830–1840.

[Févotte et al., 2009] Févotte, C., Bertin, N., and Durrieu, J.-L. (2009). Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830.

[Févotte and Idier, 2011] Févotte, C. and Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation*, 23(9):2421–2456.

[Higuchi and Kameoka, 2014] Higuchi, T. and Kameoka, H. (2014). Joint audio source separation and dereverberation based on multichannel factorial hidden markov model. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pages 1–6. IEEE.

[Hiroe, 2006] Hiroe, A. (2006). Solution of permutation problem in frequency domain ICA using multivariate probability density functions. In *Proc. ICA 2006 (LNCS 3889)*, pages 601–608. Springer.

[Hunter and Lange, 2000] Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, 9(1):60–77.

[Hunter and Lange, 2004] Hunter, D. R. and Lange, K. (2004). A tutorial on mm algorithms. *The American Statistician*, 58(1):30–37.

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

[Ikeshita et al., 2017] Ikeshita, R., Kawaguchi, Y., Togami, M., Fujita, Y., and Nagamatsu, K. (2017). Independent vector analysis with frequency range division and prior switching. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 2329–2333. IEEE.

[Kameoka et al., 2006] Kameoka, H., Goto, M., and Sagayama, S. (2006). Selective amplifier of periodic and non-periodic components in concurrent audio signals with spectral control envelopes. *Proc. SIG Tech. Reports on Music and Computer of IPSJ*, pages 77–84.

[Kameoka et al., 2009] Kameoka, H., Ono, N., Kashino, K., and Sagayama, S. (2009). Complex NMF: A new sparse representation for acoustic signals. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3437–3440. IEEE.

[Kameoka and Takamune, 2014] Kameoka, H. and Takamune, N. (2014). Training restricted boltzmann machines with auxiliary function approach. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pages 1–6. IEEE.

[Kameoka et al., 2010] Kameoka, H., Yoshioka, T., Hamamura, M., Le Roux, J., and Kashino, K. (2010). Statistical model of speech signals based on composite autoregressive system with application to blind source separation. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 245–253. Springer.

[Kim, 2010] Kim, T. (2010). Real-time independent vector analysis for convolutive blind source separation. *IEEE Trans. Circuits and Systems I: Regular Papers*, 57(7):1431–1438.

[Kim et al., 2007] Kim, T., Attias, H. T., Lee, S.-Y., and Lee, T.-W. (2007). Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. Audio, Speech and Language Processing*, 15(1):70–79.

[Kim et al., 2006] Kim, T., Eltoft, T., and Lee, T.-W. (2006). Independent vector analysis: An extension of ica to multivariate components. In *International Conference on Independent Component Analysis and Signal Separation*, pages 165–172. Springer.

[Kitamura et al., 2017] Kitamura, D., Ono, N., and Saruwatari, H. (2017). Experimental analysis of optimal window length for independent low-rank matrix analysis. In *Proc. EUSIPCO 2017*, pages 1210–1214.

[Kitamura et al., 2015a] Kitamura, D., Ono, N., Sawada, H., Kameoka, H., and Saruwatari, H. (2015a). Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model. In *Proc. ICASSP 2015*, pages 276–280.

[Kitamura et al., 2015b] Kitamura, D., Ono, N., Sawada, H., Kameoka, H., and Saruwatari, H. (2015b). Relaxation of rank-1 spatial constraint in overdetermined blind source separation. In *Proc. EUSIPCO 2015*, pages 1271–1275.

[Kitamura et al., 2016] Kitamura, D., Ono, N., Sawada, H., Kameoka, H., and Saruwatari, H. (2016). Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9):1626–1641.

[Kitamura et al., 2018] Kitamura, D., Ono, N., Sawada, H., Kameoka, H., and Saruwatari, H. (2018). Determined blind source separation with independent low-rank matrix analysis. In Makino, S., editor, *Audio Source Separation*. Springer.

[Lange et al., 2000] Lange, K., Hunter, D. R., and Yang, I. (2000). Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, 9(1):1–20.

[Lee and Seung, 2001] Lee, D. and Seung, H. (2001). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, pages 556–562.

[Lee and Seung, 1999] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401:788–791.

[Lee et al., 2006] Lee, I., Kim, T., and Lee, T.-W. (2006). Complex FastIVA: A robust maximum likelihood approach of MICA for convolutive BSS. In *Proc. ICA 2006 (LNCS 3889)*, pages 625–632. Springer.

[Lee et al., 2007] Lee, I., Kim, T., and Lee, T.-W. (2007). Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. *Signal Processing*, 87(8):1859–1871.

[Lopez et al., 2015] Lopez, A. R., Ono, N., Remes, U., Palomaki, K., and Kurimo, M. (2015). Designing multichannel source separation based on single-channel source separation. In *Proc. ICASSP*, pages 469–473.

[Matsuoka and Nakashima, 2001] Matsuoka, K. and Nakashima, S. (2001). Minimal distortion principle for blind source separation. In *Proc. ICA 2001*, pages 722–727.

[Mitsui et al., 2017a] Mitsui, Y., Kitamura, D., Takamichi, S., Ono, N., and Saruwatari, H. (2017a). Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity. In *Proc. ICASSP 2017*, pages 21–25.

[Mitsui et al., 2017b] Mitsui, Y., Takamune, N., Kitamura, D., Saruwatari, H., Takahashi, Y., and Kondo, K. (2017b). Independent low-rank matrix analysis based on parametric majorization-equalization algorithm. In *Proc. CAMSAP 2017*, pages 98–102.

[Mitsui et al., 2018] Mitsui, Y., Takamune, N., Kitamura, D., Saruwatari, H., Takahashi, Y., and Kondo, K. (2018). Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis. In *Proc. ICASSP 2018*.

[Mogami et al., 2017] Mogami, S., Kitamura, D., Mitsui, Y., Takamune, N., Saruwatari, H., and Ono, N. (2017). Independent low-rank matrix analysis based on complex student's t-distribution for blind audio source separation. In *Proc. MLSP 2017*.

[Mori et al., 2006] Mori, Y., Saruwatari, H., Takatani, T., Ukai, S., Shikano, K., Hiekata, T., Ikeda, Y., Hashimoto, H., and Morita, T. (2006). Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking. *EURASIP Journal on Applied Signal Processing*, 2006:Article ID 34970, 17 pages.

[Murata et al., 2001] Murata, N., Ikeda, S., and Ziehe, A. (2001). An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41:1–24.

[Nakano et al., 2010] Nakano, M., Kameoka, H., Le Roux, J., Kitano, Y., Ono, N., and Sagayama, S. (2010). Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence. In *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, pages 283–288. IEEE.

[Nesta and Koldovský, 2017] Nesta, F. and Koldovský, Z. (2017). Supervised independent vector analysis through pilot dependent components. In *Proc. ICASSP*, pages 536–540.

[Ono, 2011] Ono, N. (2011). Stable and fast update rules for independent vector analysis based on auxiliary function technique. In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pages 189–192. IEEE.

[Ono, 2012a] Ono, N. (2012a). Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–4.

[Ono, 2012b] Ono, N. (2012b). Fast stereo independent vector analysis and its implementation on mobile phone. In *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*. IEEE.

[Ono et al., 2009] Ono, N., Kohno, H., Ito, N., and Sagayama, S. (2009). Blind alignment of asynchronously recorded signals for distributed microphone array. In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pages 161–164. IEEE.

[Ono and Miyabe, 2010] Ono, N. and Miyabe, S. (2010). Auxiliary-function-based independent component analysis for super-Gaussian sources. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 165–172. Springer.

[Ono and Sagayama, 2010] Ono, N. and Sagayama, S. (2010). R-means localization: A simple iterative algorithm for source localization based on time difference of arrival. In *Proc. ICASSP*, pages 2718–2721. IEEE.

[Ono et al., 2012] Ono, T., Ono, N., and Sagayama, S. (2012). User-guided independent vector analysis with source activity tuning. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 2417–2420. IEEE.

[Ozerov and Févotte, 2010] Ozerov, A. and Févotte, C. (2010). Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, 18(3):550–563.

[Parra and Spence, 2000] Parra, L. and Spence, C. (2000). Convolutive blind separation of non-stationary sources. *IEEE Trans. Speech Audio Processing*, 8(3):320–327.

[Saruwatari et al., 2006] Saruwatari, H., Kawamura, T., Nishikawa, T., Lee, A., and Shikano, K. (2006). Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. *IEEE Trans. Audio, Speech and Language Processing*, 14(2):666–678.

[Saruwatari et al., 2003] Saruwatari, H., Kurita, S., Takeda, K., Itakura, F., Nishikawa, T., and Shikano, K. (2003). Blind source separation combining independent component analysis and beamforming. *EURASIP Journal on Applied Signal Processing*, 2003(11):1135–1146.

[Sawada et al., 2007] Sawada, H., Araki, S., and Makino, S. (2007). Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS. In *Proc. ISCAS 2007*, pages 3247–3250.

[Sawada et al., 2011] Sawada, H., Araki, S., and Makino, S. (2011). Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. *IEEE Trans. Audio, Speech, and Language Processing*, 19(3):516–527.

[Sawada et al., 2012] Sawada, H., Kameoka, H., Araki, S., and Ueda, N. (2012). Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization. In *Proc. ICASSP 2012*, pages 261–264.

[Sawada et al., 2013] Sawada, H., Kameoka, H., Araki, S., and Ueda, N. (2013). Multichannel extensions of non-negative matrix factorization with complex-valued data. *IEEE Trans. Audio, Speech, and Language Processing*, 21(5):971–982.

[Sawada et al., 2003] Sawada, H., Mukai, R., Araki, S., and Makino, S. (2003). Polar coordinate based nonlinear function for frequency domain blind source separation. *IEICE Trans. Fundamentals*, E86-A(3):590–596.

[Sawada et al., 2004] Sawada, H., Mukai, R., Araki, S., and Makino, S. (2004). A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Trans. Speech Audio Processing*, 12(5):530–538.

[Schobben and Sommen, 2002] Schobben, L. and Sommen, W. (2002). A frequency domain blind signal separation method based on decorrelation. *IEEE Trans. Signal Processing*, 50(8):1855–1865.

[Smaragdis, 1998] Smaragdis, P. (1998). Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22:21–34.

[Sun et al., 2017] Sun, Y., Babu, P., and Palomar, D. P. (2017). Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65(3):794–816.

[Sunohara et al., 2017] Sunohara, M., Haruta, C., and Ono, N. (2017). Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components. In *Proc. ICASSP*, pages 216–220.

[Takatani et al., 2004] Takatani, T., Nishikawa, T., Saruwatari, H., and Shikano, K. (2004). High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis. *IEICE Trans. Fundamentals*, E87-A(8):2063–2072.

[Taniguchi et al., 2014] Taniguchi, T., Ono, N., Kawamura, A., and Sagayama, S. (2014). An auxiliary-function approach to online independent vector analysis for real-time blind source separation. In *Proc. HSCMA*, pages 107–111.

[Yeredor, 2009] Yeredor, A. (2009). On hybrid exact-approximate joint diagonalization. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2009 3rd IEEE International Workshop on*, pages 312–315. IEEE.

[Yoshii et al., 2013] Yoshii, K., Tomioka, R., Mochihashi, D., and Goto, M. (2013). Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *International Conference on Machine Learning*, pages 576–584.

[Yoshioka et al., 2008] Yoshioka, T., Nakatani, T., and Miyoshi, M. (2008). An integrated method for blind separation and dereverberation of convolutive audio mixtures. In *Proc. EUSIPCO*.