

信号処理論特論

第3&4回 (5/8, 5/15)

情報理工学系研究科システム情報学専攻
猿渡 洋

hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

講義予定

- 04/10: 第1回 統計的音声音響信号処理概論
- 05/01: 第2回 非負値行列因子分解
- 05/08: 第3回 ブラインド音源分離その1
- 05/15: 第4回 ブラインド音源分離その2
- 05/22: 第5回 エンハンスメント・高次統計量解析とその応用
- 05/29: 第6回 【レポート課題1】
- 06/05: 第7回 音声合成・変換その1
- 06/12: 第8回 音声合成・変換その2
- 06/19: 第9回 音場再現の基礎
- 06/26: 第10回 学外講師・未定
- 07/03: 第11回 【レポート課題2】

講義資料と成績評価

■ 講義資料

- <http://www.sp.ipc.i.u-tokyo.ac.jp/~saruwatari/>

(システム情報第一研究室からたどれるようになってます)

■ 成績評価

- 出席点
- レポート点(2回の提出が必須)

本日の話題

■ ブラインド音源分離

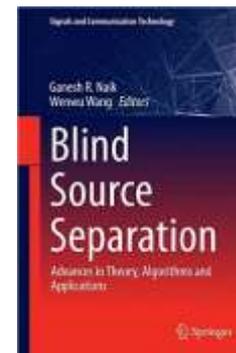
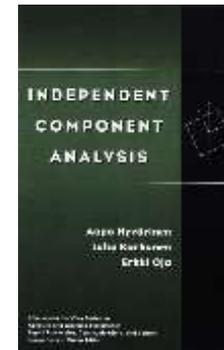
- 複数のマイクロホンで取得した観測信号を基に、同時に鳴っている複数の音源信号を「教師情報無しで」分離獲得する技術

■ 独立成分分析、独立ベクトル分析、独立低ランク行列分析

- ブラインド音源分離を行うための統計的手法

■ 参考資料

- A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- 村田昇, 入門 独立成分分析, 東京電機大学出版局, 2004.
- G. Naik and W. Wang, *Blind Source Separation: Advances in Theory, Algorithms and Applications*, Springer, 2013.



実音響環境の
独立成分分析に
ついて猿渡が執筆

本日の話題

■ ブラインド音源分離

- 複数のマイクロホンで取得した観測信号を基に、同時に鳴っている複数の音源信号を「教師情報無しで」分離獲得する技術

■ 独立成分分析、独立ベクトル分析、独立低ランク行列分析

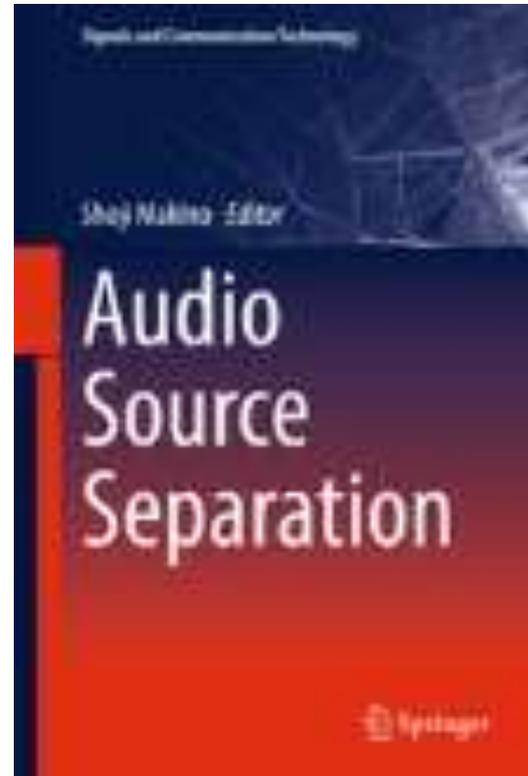
- ブラインド音源分離を行うための統計的手法

■ 参考資料

- S. Makino, Ed.,
Audio Source Separation,
Springer, 2018.

NEW!

独立低ランク行列分析と
ミュージカルノイズフリーBSS
の2章を執筆



目次

- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

目次

- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

音の分離

■カクテルパーティー効果

- 人の聞き分け能力の模擬
- 補聴器への応用



■音声認識

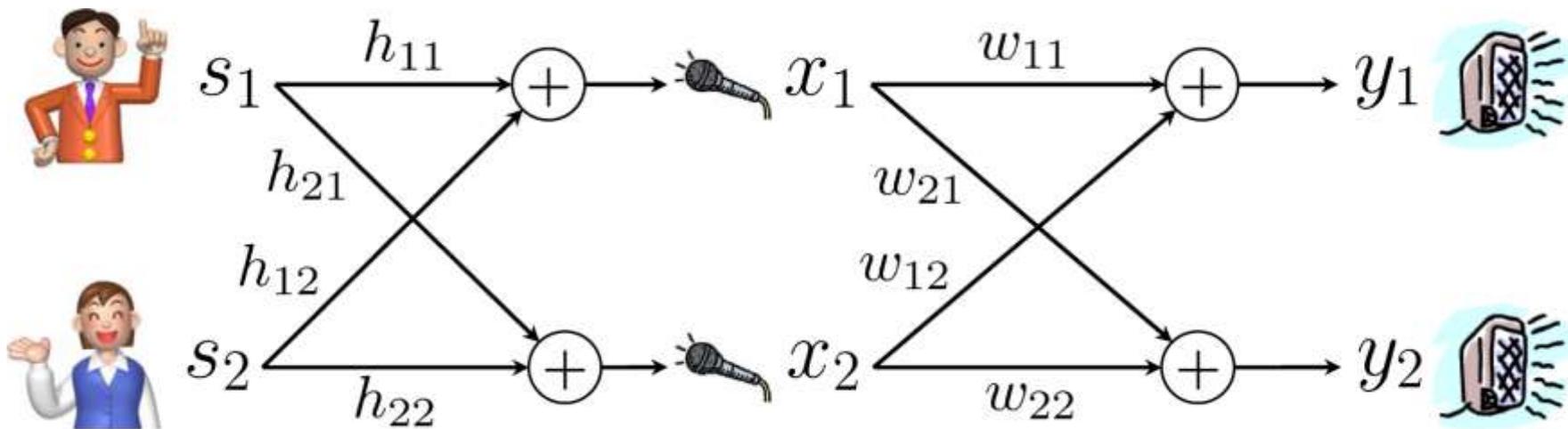
- マイクロホンと口の間の距離が大きくなるにつれて増大してくる妨害音を抑圧・除去
- 音声対話インターフェイスや対話ロボットへの応用
- 環境音認識・音監視システム

■音楽／楽器音分析

- ミックスダウンされてしまった録音データの解析・エディット
- 自動採譜、ユーザオリエンテッドな音楽情報処理

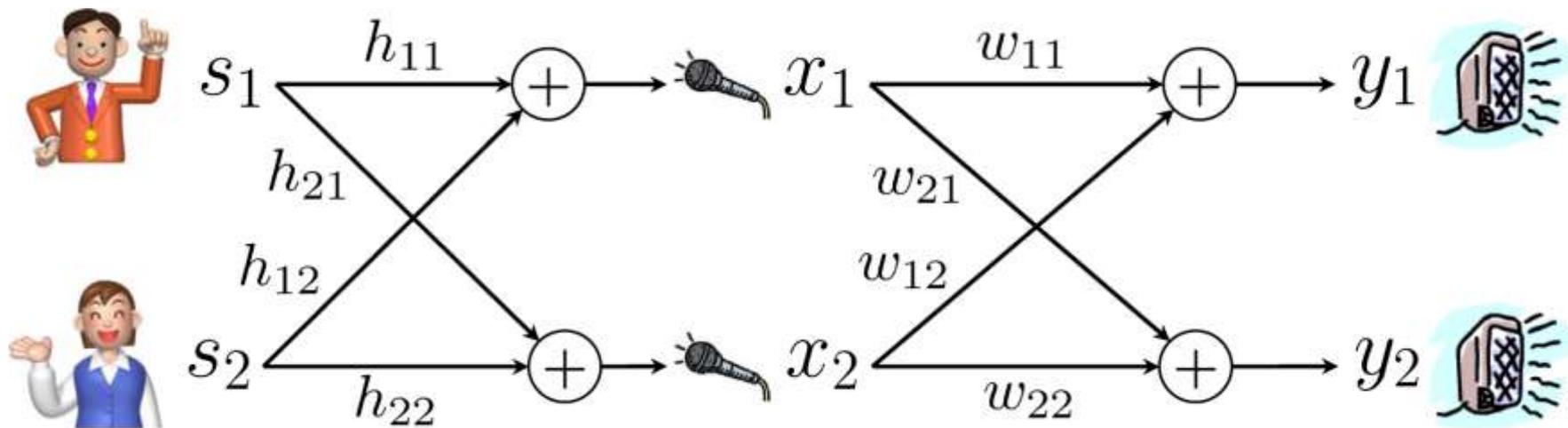
ブラインド音源分離 (Blind Source Separation)

- 混ぜり合った信号 x_1, x_2 から元の信号を取り出す
- どのように混ぜたかに関する情報 H は利用できない



ブラインド音源分離 (Blind Source Separation)

- 混ざり合った信号 x_1, x_2 から元の信号を取り出す
- どのように混ざったかに関する情報 H は利用できない



実は上記は**2つのことを同時に推定**している

- [空間] 統計的に独立な音源の分類問題 (分離行列 W の推定)
- [信号] 各音源が属する確率分布 $p(y)$ の推定問題

上記を閉形式で解く方法は存在せず凸問題でもない⇒**大変困難!**

音の分離

■何を手がかりにするか？

■音源の性質

- ◆音声の場合: 声質, ハーモニクス構造、話し方の特徴
- ◆楽器の場合: 音の高さ／特徴, リズム, 繰り返しパターン
- ◆それらが相互に異なるという統計的な「独立性」

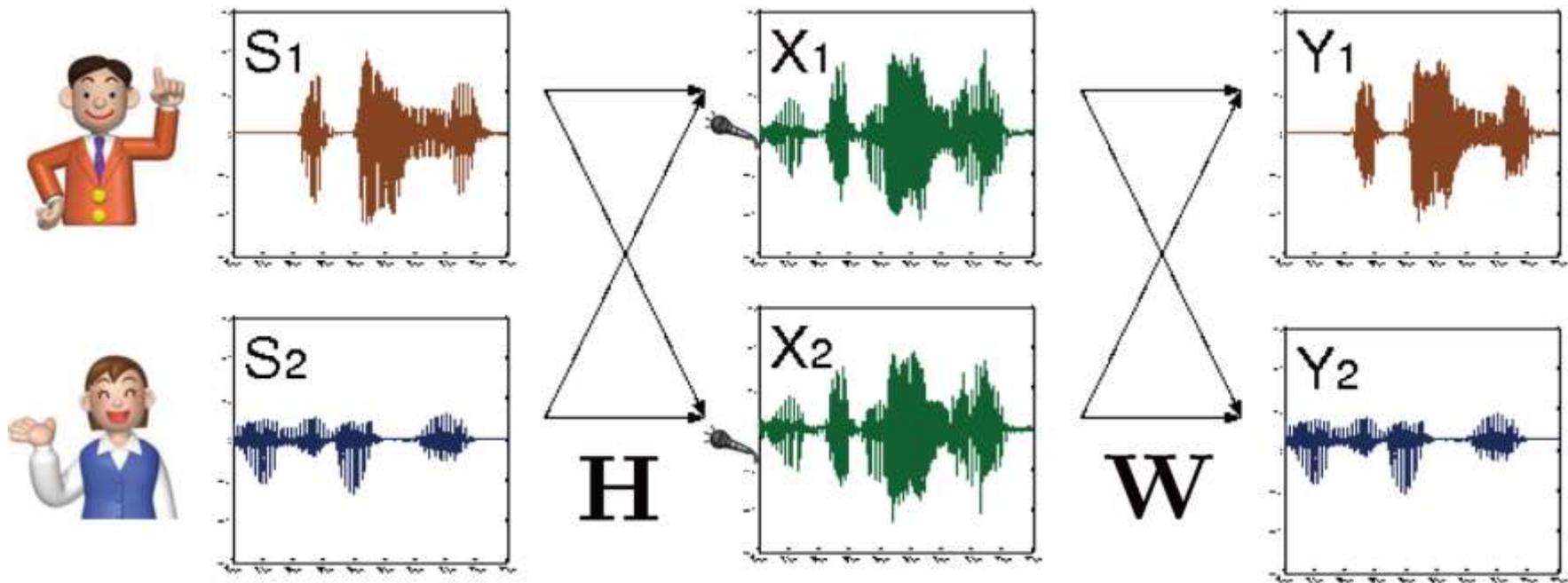
■空間情報

- ◆音源の方向、距離、壁や床の反射状況

■分離方法

- 線形フィルタ: 独立成分分析～独立低ランク行列分析
- 非線形処理: 時間周波数マスキング、ポストフィルタ

独立成分分析 (Independent Component Analysis)



元の信号は独立

- 一方から他方が推測できない

依存関係がある

- 同じ成分が双方に入っている
- 一方から他方が推測できる

なるべく独立に

- 具体的にはどうやる？

ICAによるブラインド音源分離

- I 個の音源信号 \mathbf{s} が混合行列 \mathbf{H} により混ざり合い, J 個の観測信号 \mathbf{x} が T 個得られたとする。

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) \quad t = 1, \dots, T$$
$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_I \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_I \end{bmatrix}$$

- 分離行列 \mathbf{W} により分離信号 \mathbf{y} を生成する。

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_I \end{bmatrix}$$

- \mathbf{W} の計算は観測信号 \mathbf{x} のみから行う。
 I 個の分離信号 y_1, \dots, y_I が互いに独立になるようにする。

ICAとその発展

■ 歴史：1990年代から世界的な広がり

- ◆ 1989 Cardoso
- ◆ 1990 Jutten (高次無相関化)
- ◆ 1994 Comon (ICAという言葉进行定義)
- ◆ 1995 Bell (infomaxによる定式化)
- ◆ 1998 Smaragdis, Ikeda-Murata, Saruwatari ... (音響信号へICAを導入)
- ◆ 2003 Saruwatari, Araki-Makino, Parra, etc. (空間逆フィルタとICAの融合)
- ◆ 2004 Sawada (ロバストパーミュテーションソルバの開発)
- ◆ 2006 Kim, Hiroe (独立ベクトル分析の提唱)
- ◆ 2009 Takahashi-Saruwatari (雑音推定ICA＋非線形BSS; 世界初の商用化)
- ◆ 2010 Ono (補助関数法による独立ベクトル分析)
- ◆ 2010 Ozerov, Sawada (多チャネルNMFとしての定式化)
- ◆ 2015 Kitamura-Ono-Sawada-Kameoka-Saruwatari (独立低ランク行列分析)

国際会議ICA(現在はLVA/ICA)を1999年より連続して開催
(第4回ICA2003は日本で開催された)

ICAとその発展

■ 歴史：1990年代から世界的な広がり

- ◆ 1989 Cardoso
- ◆ 1990 Jutten (高次無相関化)
- ◆ 1994 Comon (ICAという言葉を定義)
- ◆ 1995 Bell (infomaxによる定式化)
- ◆ 1998 Smaragdis, Ikeda-Murata, Saruwatari ... (音響信号へICAを導入)
- ◆ 2003 Saruwatari, Araki-Makino, Parra, etc. (空間逆フィルタとICAの融合)
- ◆ 2004 Sawada (ロバストパーミュテーションソルバの開発)
- ◆ 2006 Kim, Hiroe (独立ベクトル分析の提唱)
- ◆ 2009 Takahashi-Saruwatari (雑音推定ICA＋非線形BSS; 世界初の商用化)
- ◆ 2010 Ono (補助関数法による独立ベクトル分析)
- ◆ 2010 Ozerov, Sawada (多チャネルNMFとしての定式化)
- ◆ 2015 Kitamura-Ono-Sawada-Kameoka-Saruwatari (独立低ランク行列分析)

補助関数
独立ベクトル
分析

東大 NII

IS基準
多チャネル
NMF

NTT

IS-NMF、
複素NMF

東大 NTT

音響信号
BSS、周波
数領域ICA

NAIST 東大

Independent Factor Analysis Family

ICA
= Independent **Component** Analysis

= Scalar

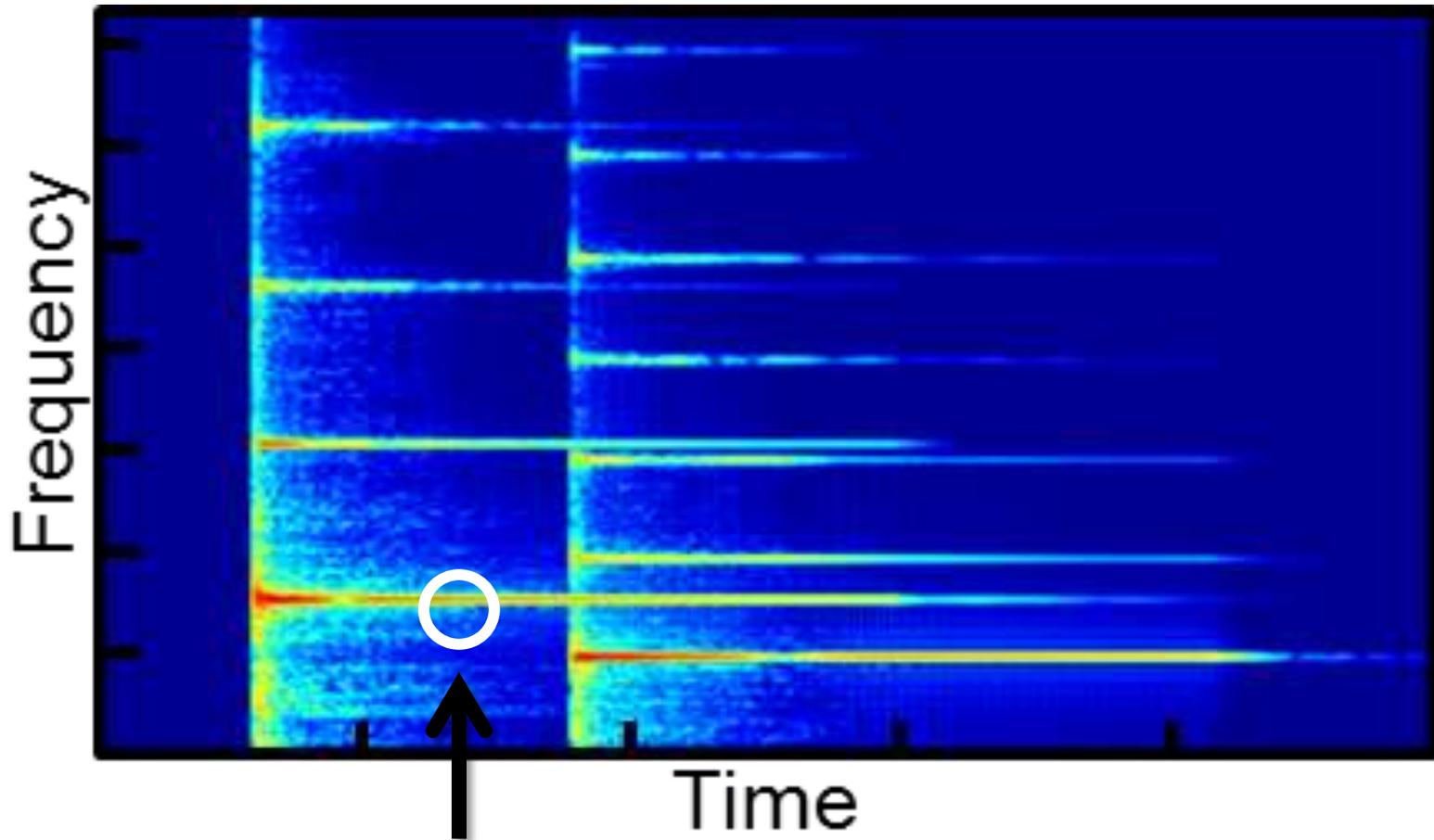
IVA
= Independent **Vector** Analysis

ILRMA
= Independent Low-Rank **Matrix** Analysis

Independent Factor Analysis family

Spectrogram:

Time-Frequency Representation via Short-Time FFT

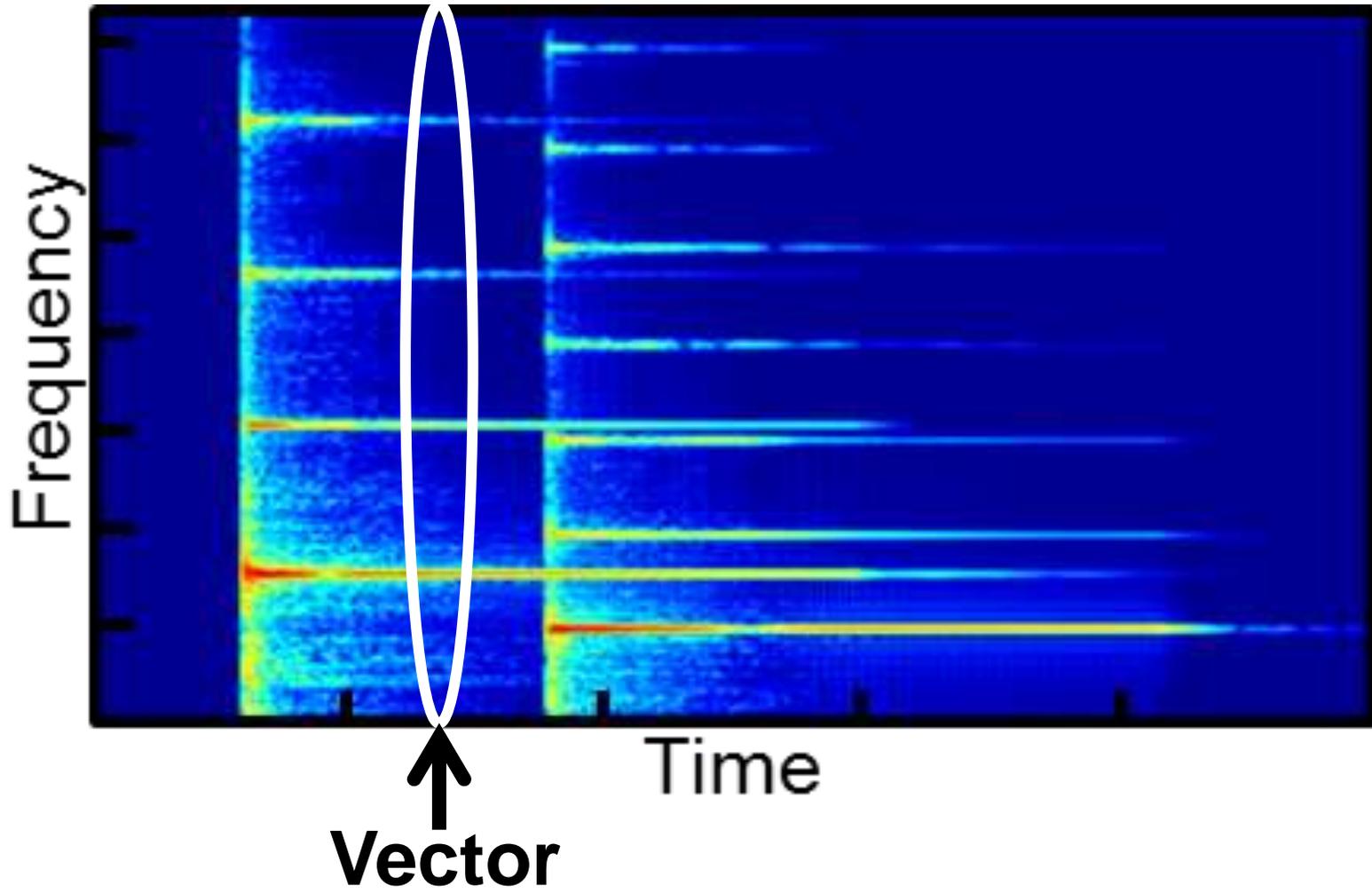


Component (Scalar)

Independent Factor Analysis family

Spectrogram:

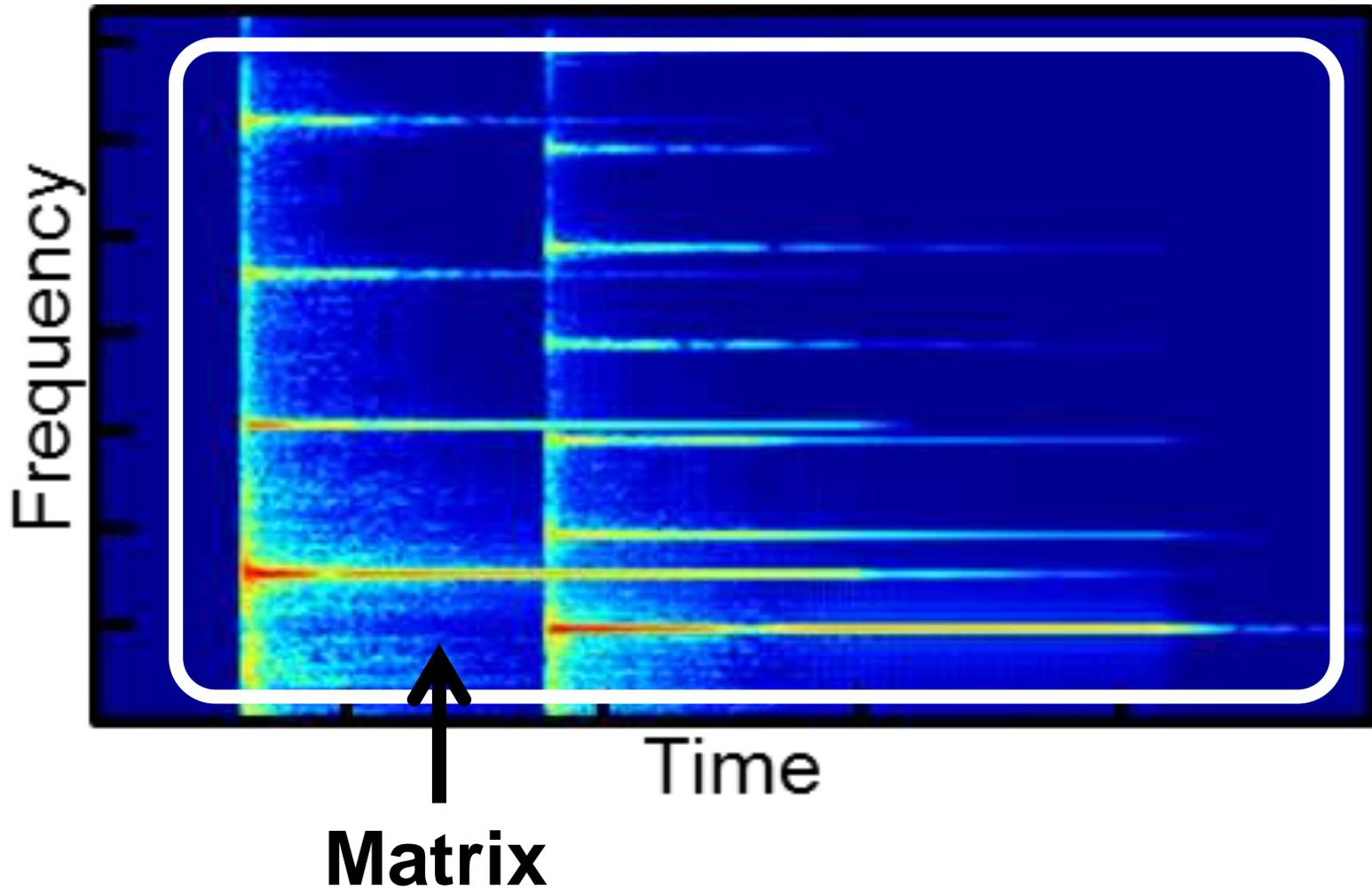
Time-Frequency Representation via Short-Time FFT



Independent Factor Analysis family

Spectrogram:

Time-Frequency Representation via Short-Time FFT

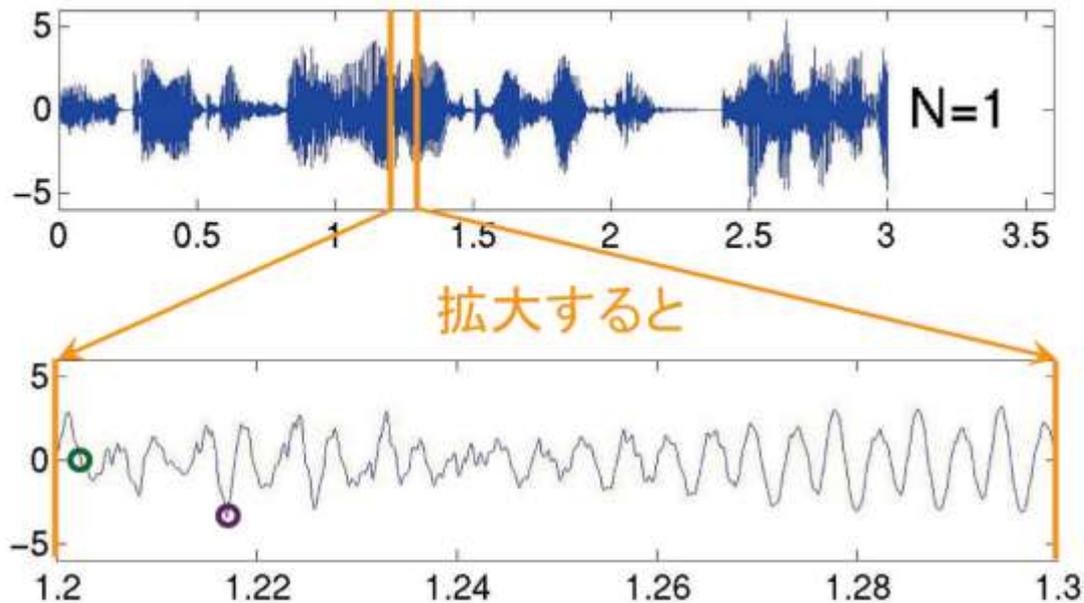


目次

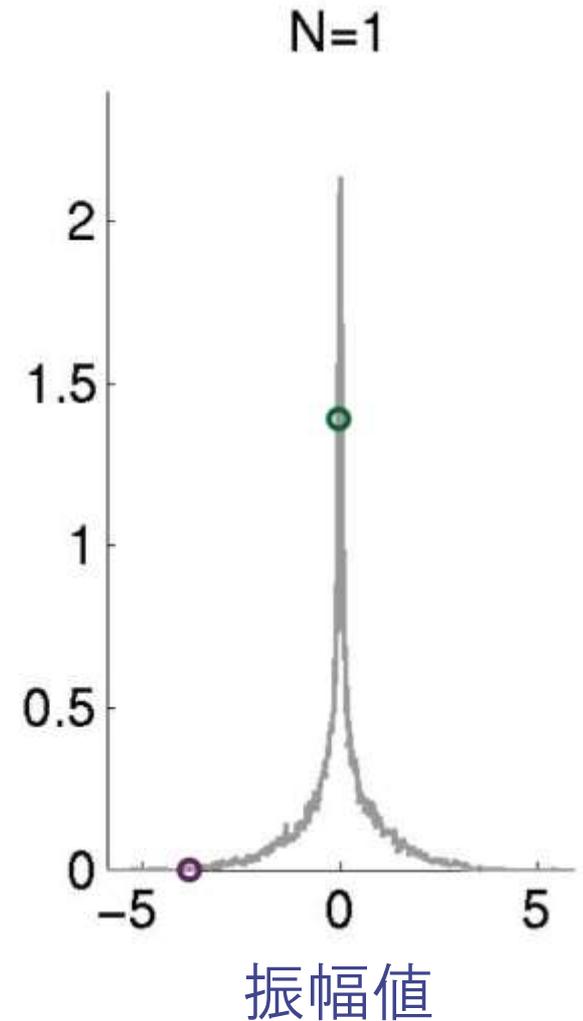
- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

音の統計的性質

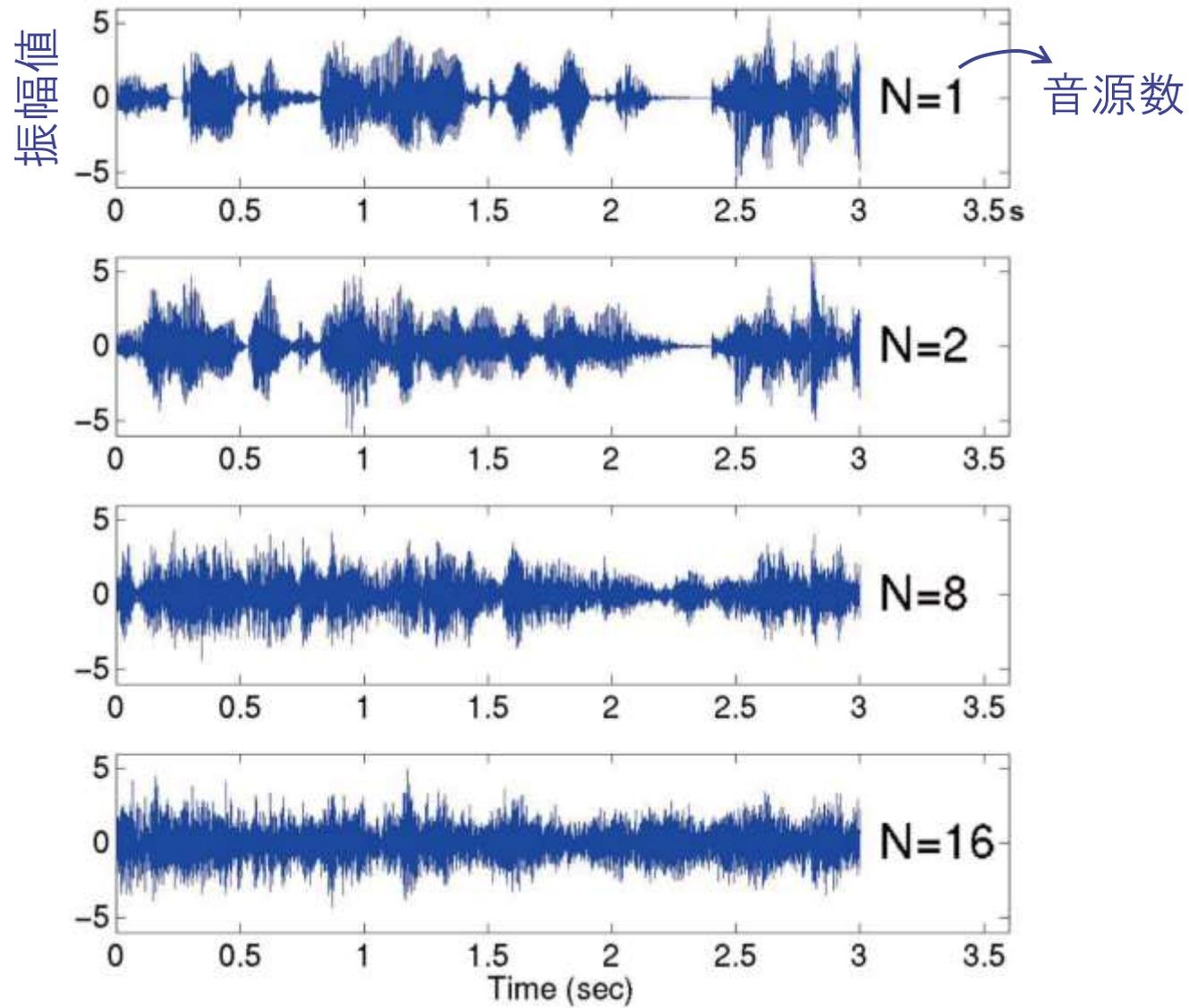
音声の波形



振幅値のヒストグラム

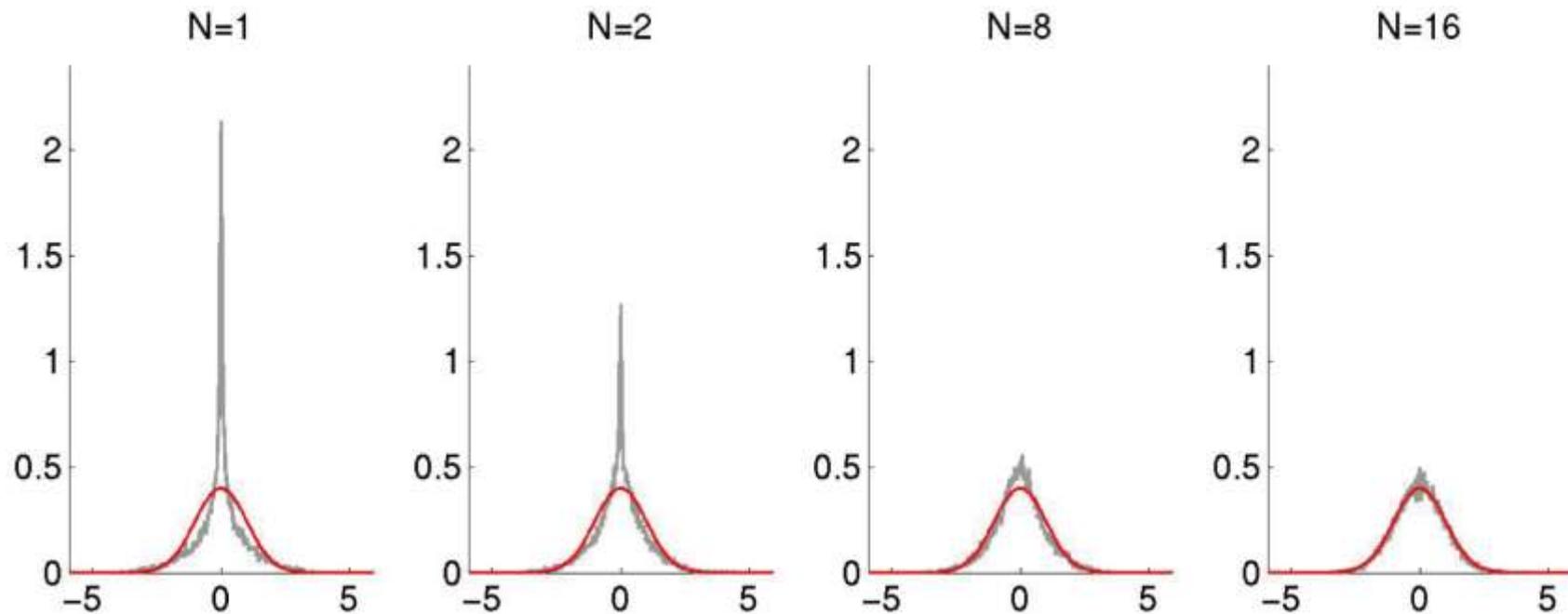


音を混ぜてみる



混ぜた音の統計的性質

■ 振幅値のヒストグラム

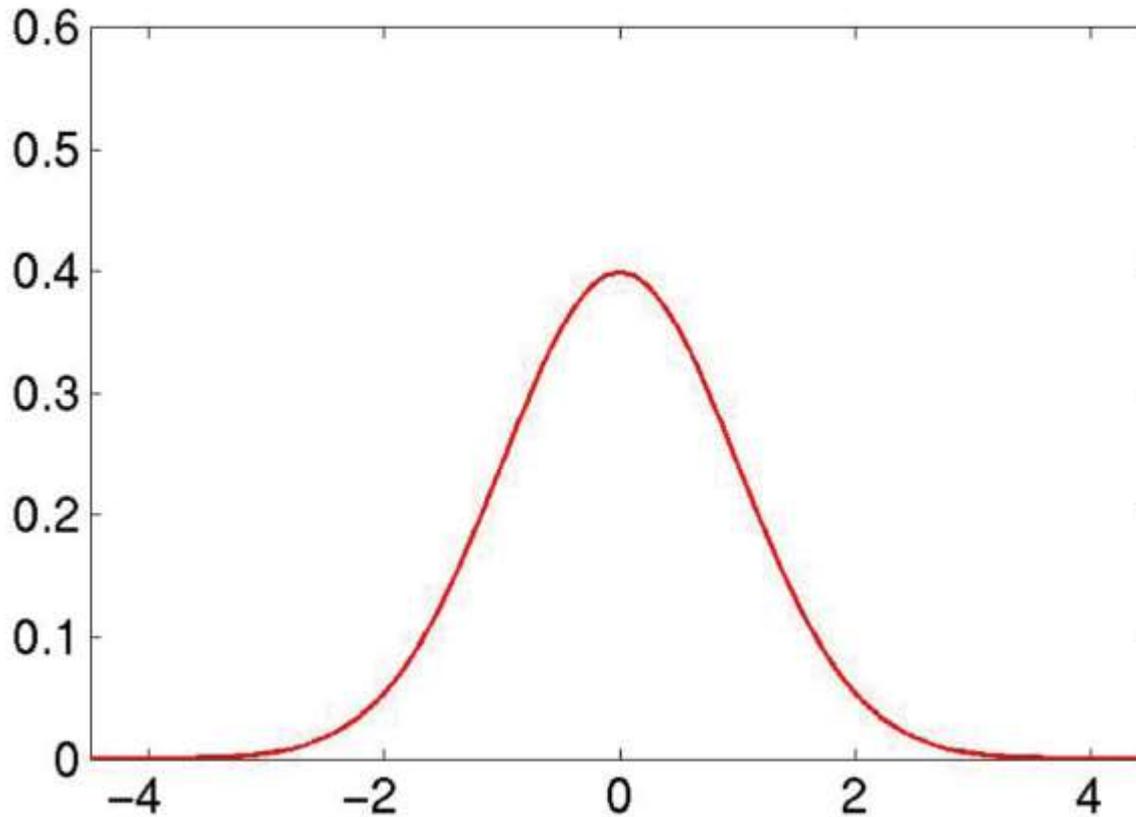


振幅値

→ 赤い線に近づいてくる

中心極限定理 (Central Limit Theorem)

- 元々の信号がどのような統計的性質をもったものであれ, 多くの信号を足し合わせると, その振幅値の分布は**正規分布**に近づく。



平均0
分散1の
正規分布

中心極限定理 (Central Limit Theorem)

- 確率変数 x_1, x_2, \dots, x_N が独立に同一な確率密度関数 $p_x(x)$ に従うならば, $z = (x_1 + x_2 + \dots + x_N) / \sqrt{N}$ の確率密度関数は $N \rightarrow \infty$ でガウス分布に近づく
- ただし, x の平均は0とする

中心極限定理の略証 (1/2)

■ x の特性関数を $\Theta_x(ju)$ とする

■ $y = x_1 + x_2 + \cdots + x_N$ の特性関数 $\Theta_y(ju)$

$$\Theta_y(ju) = \Theta_x(ju)^N \quad (\because \text{重畳積分定理})$$

■ $z = y/\sqrt{N}$ の確率密度関数 $p_z(z)$

$$p_z(z) = \sqrt{N}p_y(\sqrt{N}z)$$

■ $z = y/\sqrt{N}$ の特性関数 $\Theta_z(ju)$

$$\Theta_z(ju)$$

$$= \int_{-\infty}^{\infty} p_z(z)e^{jzu} dz = \int_{-\infty}^{\infty} \sqrt{N}p_y(\sqrt{N}z)e^{jzu} dz$$

$$= \int_{-\infty}^{\infty} p_y(y)e^{jy(u/\sqrt{N})} dy = \Theta_y(ju/\sqrt{N}) = \Theta_x(ju/\sqrt{N})^N$$

中心極限定理の略証 (2/2)

- $\log \Theta_x(ju)$ のマクローリン展開

$$\log \Theta_x(ju) = \sum_{n=0}^{\infty} \frac{k_n}{n!} (ju)^n$$

n次キュムラント

- $\log \Theta_z(ju)$ のマクローリン展開

$$\begin{aligned} \log \Theta_z(ju) &= N \log \Theta_x(ju/\sqrt{N}) \\ &= \sum_{n=0}^{\infty} \left(\frac{N k_n}{N^{n/2}} \right) \frac{(ju)^n}{n!} \end{aligned}$$

zのキュムラントを
xのキュムラントで表したもの

- 3次以上のキュムラントは $N \rightarrow \infty$ で0に収束
⇒ 分布がガウス分布に収束

(∵ ガウス分布に従う確率変数の3次以上のキュムラントはすべて0)

正規分布(Gauss分布)

■ 確率密度関数

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

■ 統計学における最も重要な分布の一つ

■ さまざまな性質がある

- 平均 μ と分散 σ^2 によって確率密度関数が一意に決まる
- 最も「ランダム」な分布
 - ◆ 同じ分散をもつ分布の中でエントロピーが最大
- 中心極限定理

エントロピー(平均情報量)

■ある事象の情報量

$$I(x) = \log \frac{1}{p(x)} = -\log p(x)$$

めったに起こらないこと
ほど情報量は大きい

■エントロピー

$$H(x) = \mathbb{E}\{-\log p(x)\}$$

- 各事象の情報量の平均値
- 事象のランダムさを表す指標
- 正規分布の場合

$$H(x) = \log \sqrt{2\pi\sigma^2} + \frac{\mathbb{E}\{(x - \mu)^2\}}{2\sigma^2} = \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2}$$

エントロピーが最大になる確率密度関数

■問題:

■ Maximize $H(x) = - \int p(x) \log p(x) dx$ with respect to $p(x)$

Subject to $\int p(x) dx = 1, \int xp(x) dx = 0, \int x^2 p(x) dx = 1$

■ラグランジュ未定乗数法

■ ラグランジアンをpに関して偏微分して0と置く

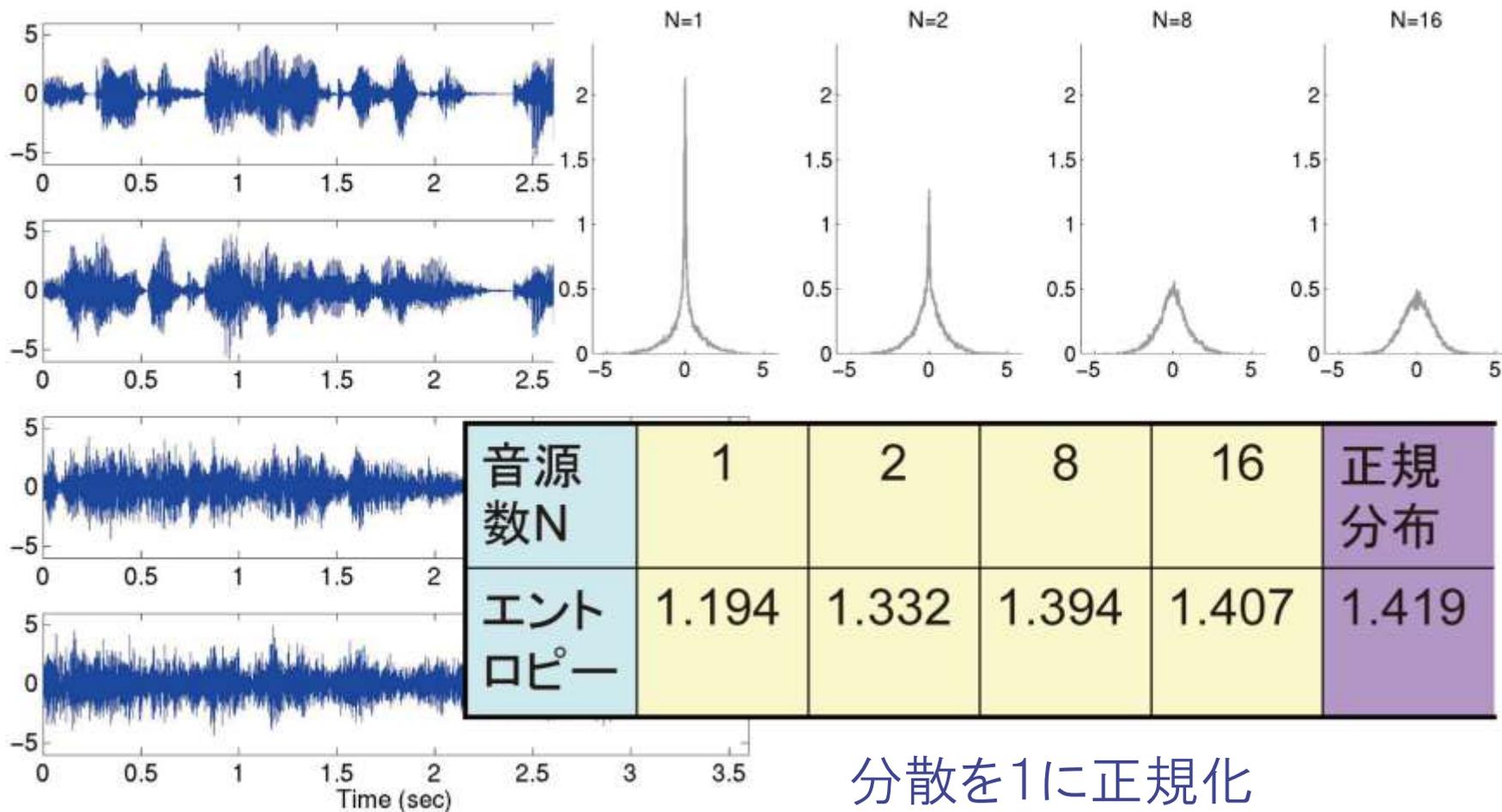
$$\begin{aligned} \mathcal{L}[p] = & - \int p(x) \log p(x) dx \\ & + \lambda_0 \left(\int p(x) dx - 1 \right) + \lambda_1 \left(\int xp(x) dx \right) + \lambda_2 \left(\int x^2 p(x) dx - 1 \right) \end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial p} = -1 - \log p(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2 = 0$$

$$\Rightarrow p(x) = \exp(1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2)$$

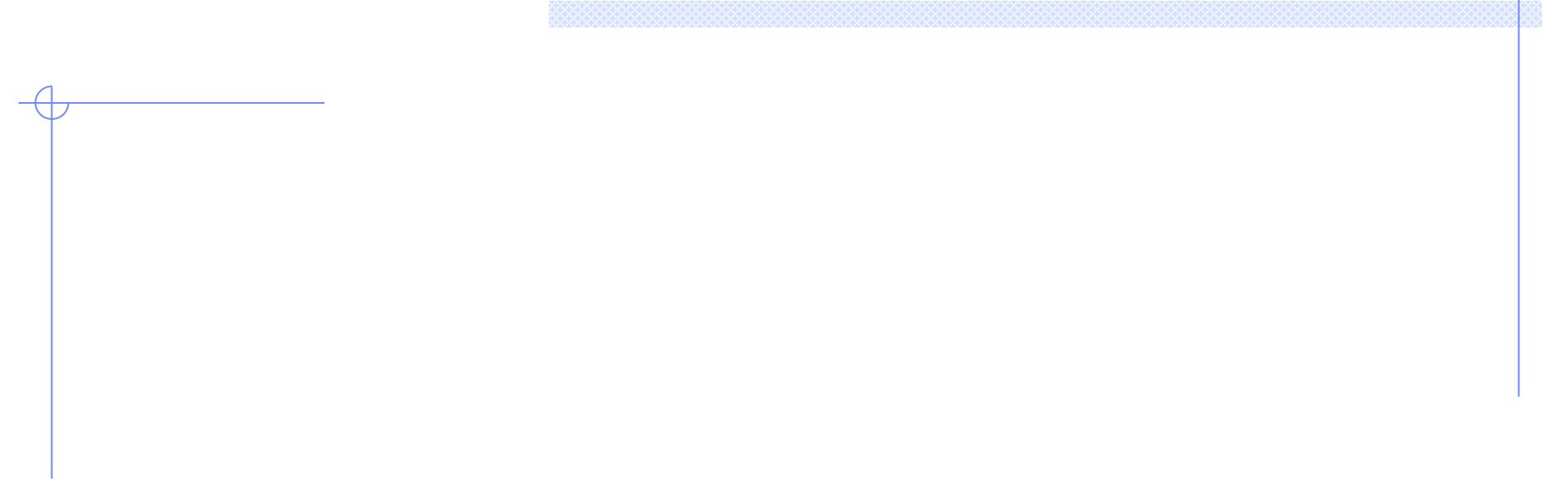
たしかに正規分布型になっている！

混ぜた音のエントロピー



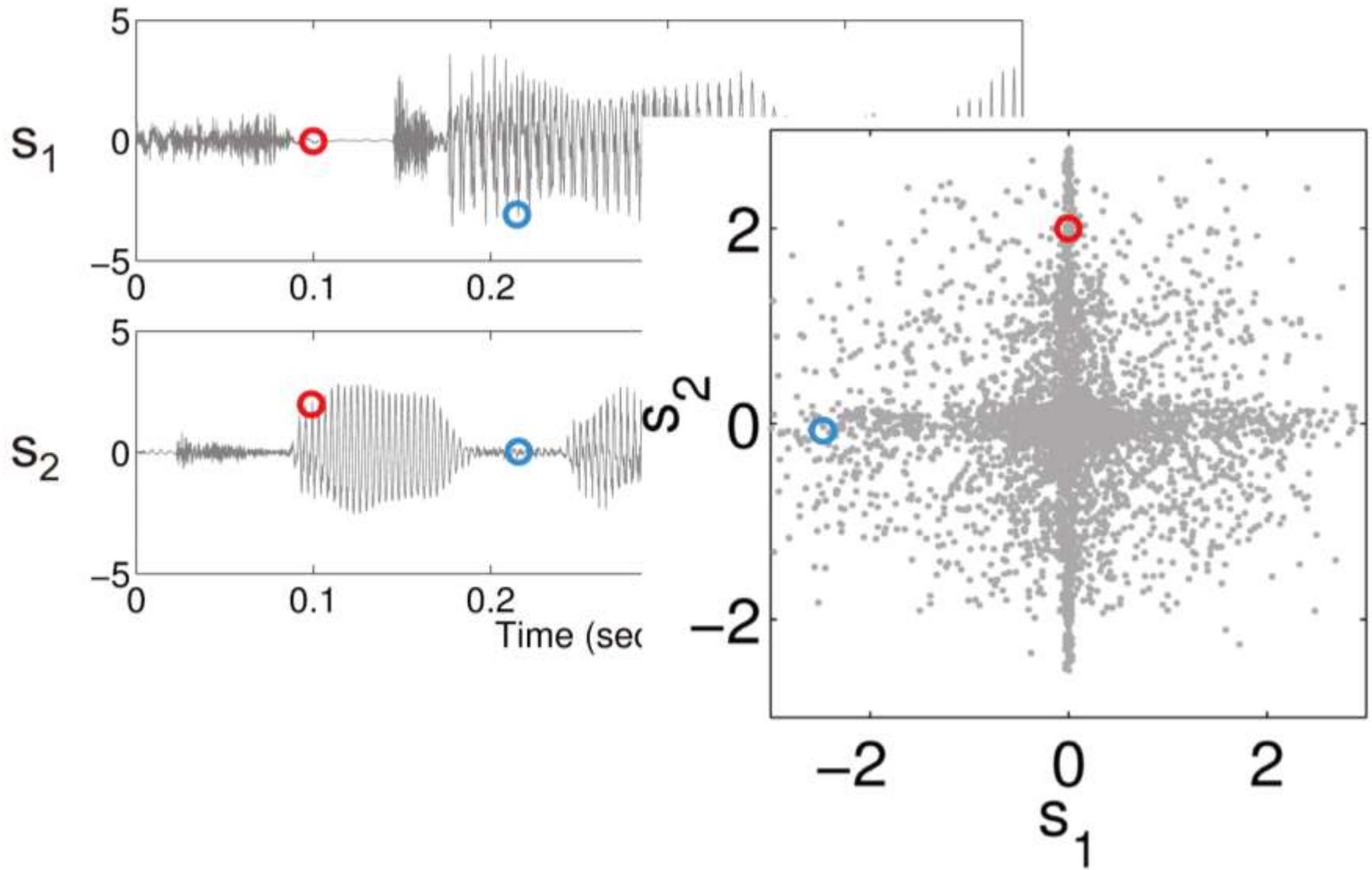
目次

- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

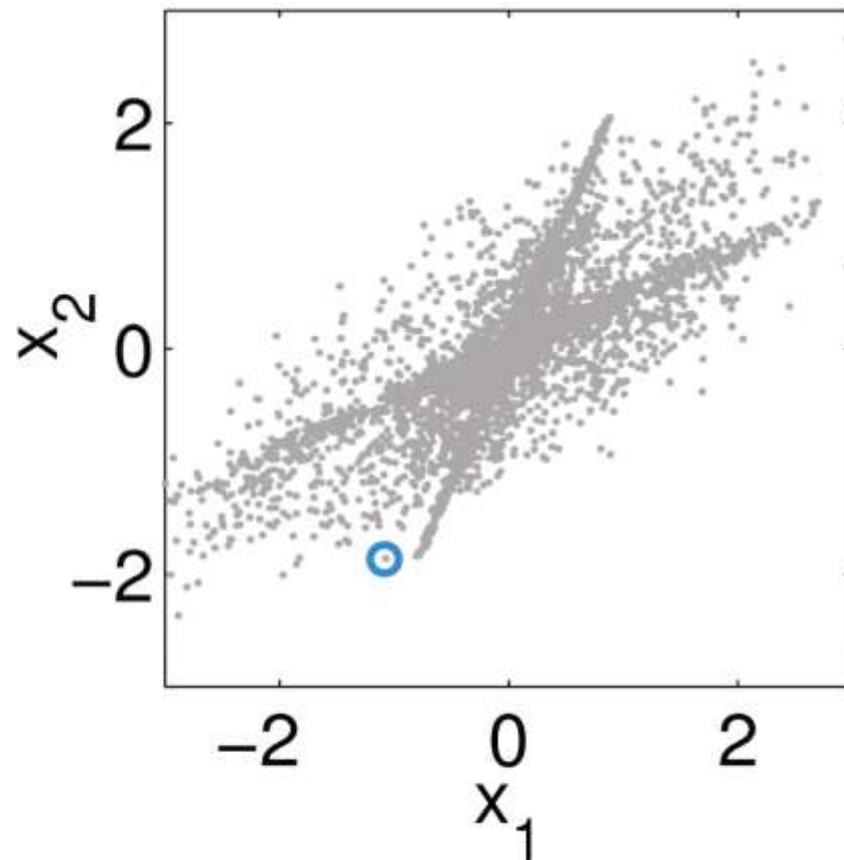
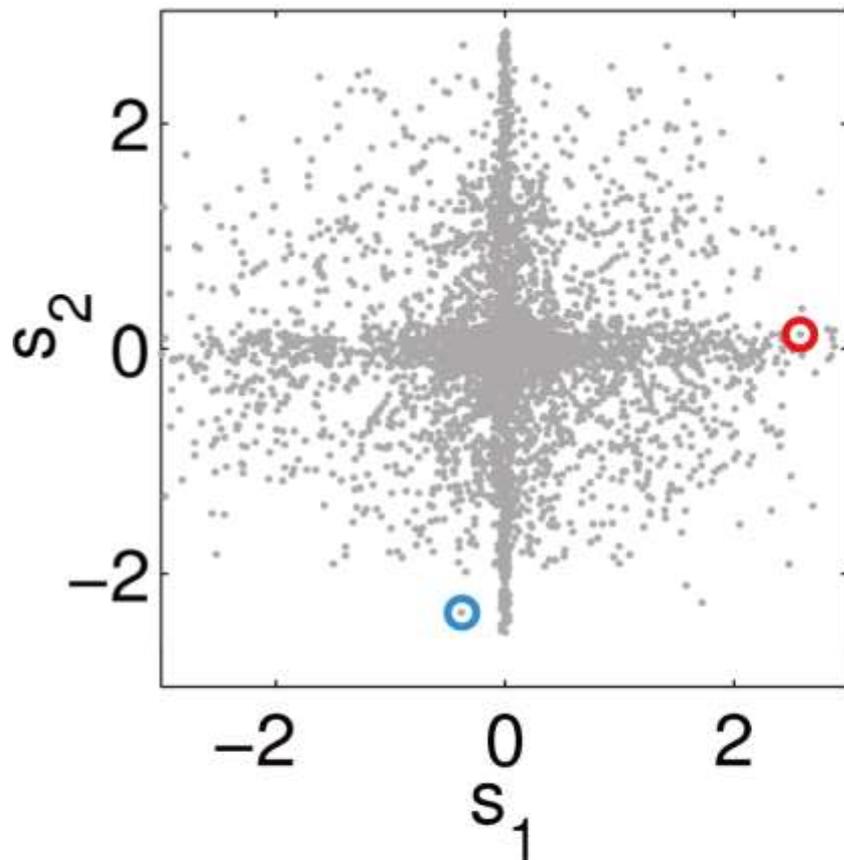


FastICA

混合信号の性質：音源信号の散布図



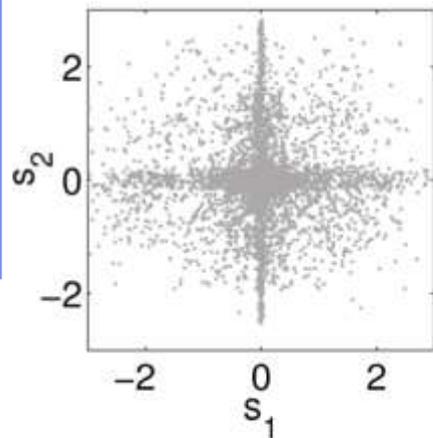
音源信号と混合信号の散布図



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.92 & 0.31 \\ 0.41 & 0.73 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

白色化(sphering) + FastICA

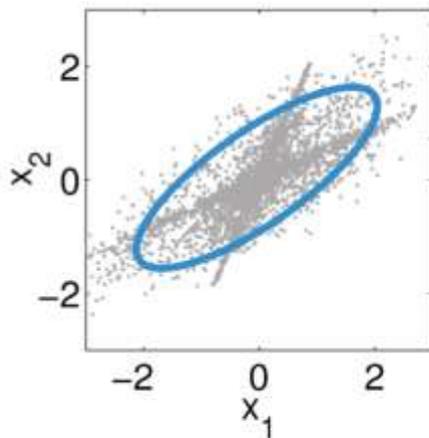
音源信号



H

混合行列
(未知)

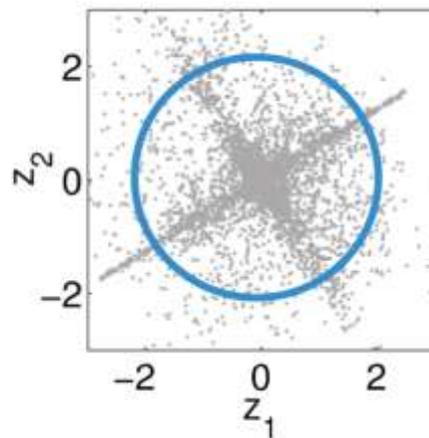
混合信号



V

白色化行列
(楕円を円に)

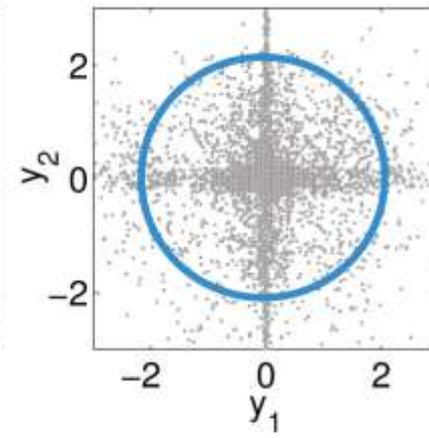
白色化信号



U

ユニタリ行列
(回転)

分離信号



FastICA

「白色化」という言い方は統計処理で主に用いられるが、信号処理では「スペクトルのフラット化」の意味で使われることがあったので「sphering」という語が推奨される。

白色化

- 相関行列が単位行列になるように $\mathbf{x}(t)$ を変換

$$\mathbf{z}(t) = \mathbf{V} \mathbf{x}(t)$$

$$\mathbf{R}_z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

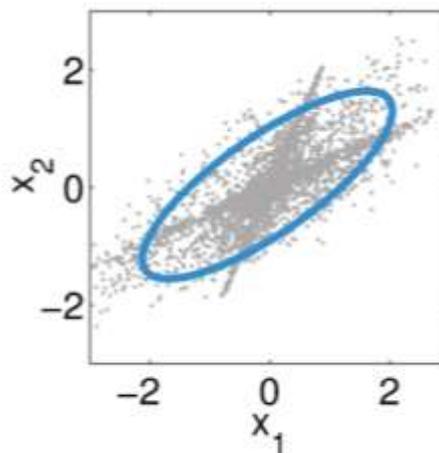
- 無相関化

$$\mathbb{E}\{z_i z_j\} = 0$$

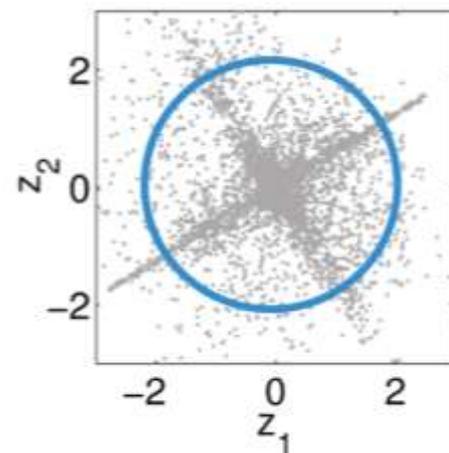
- 分散の正規化

$$\mathbb{E}\{z_i z_i\} = 1$$

混合信号



白色化信号



\mathbf{V}

白色化行列
(楕円を円に)

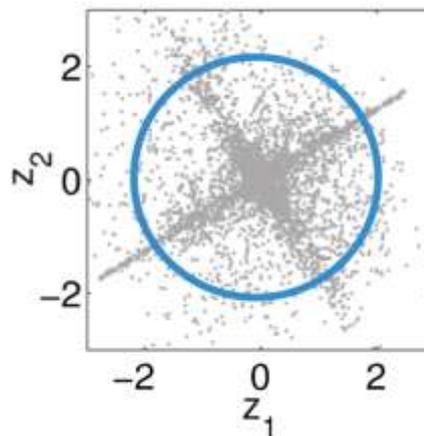
ユニタリ変換(回転)

- 相関行列が単位行列になるように $z(t)$ を変換

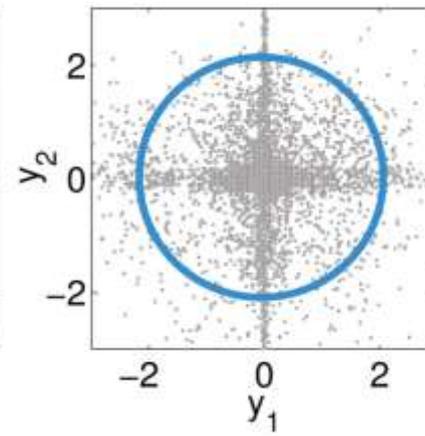
$$\mathbf{y}(t) = \mathbf{U} \mathbf{z}(t)$$

$$\mathbf{U} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

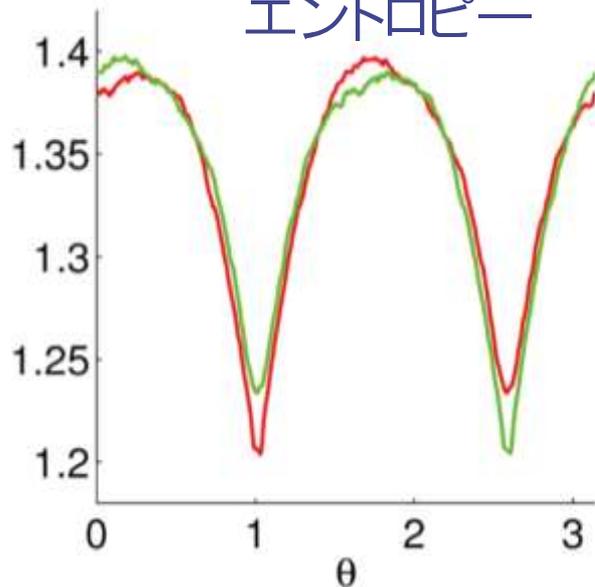
白色化信号



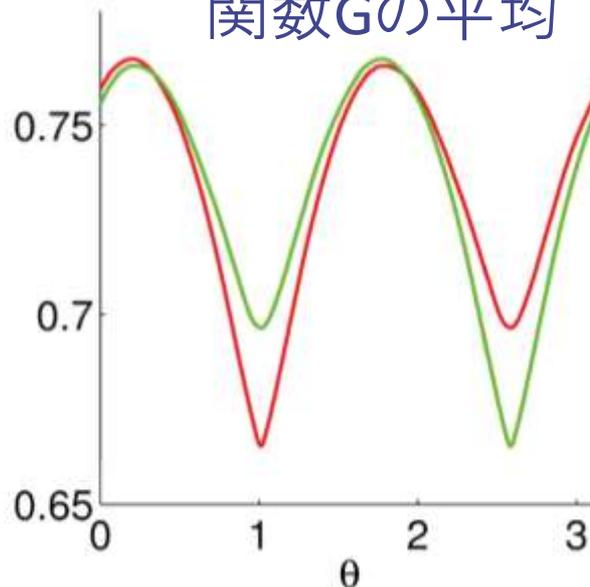
分離信号



エントロピー



関数Gの平均



$$G(y_i) = -\log p(y_i)$$

$$p(y_i) \propto \exp(-|y_i|)$$

Laplace分布による
情報量の近似

FastICAアルゴリズム

- 非線形関数 $G(y_i) = -\log p(y_i)$ の平均を最小化
 - 求めたいのは $U = [\mathbf{u}_1 \cdots \mathbf{u}_i]^T$
 - 解空間をユニタリ行列に限定

各 i について収束するまで以下を繰り返す

$$y_i \leftarrow \mathbf{u}_i^T \mathbf{z} \quad \text{分離信号の計算}$$

$$\mathbf{u}_i \leftarrow \mathbb{E}\{G''(y_i)\}\mathbf{u}_i - \mathbb{E}\{G'(y_i)\mathbf{z}\} \quad \text{Newton法による降下}$$

$$\mathbf{u}_i \leftarrow \mathbf{u}_i - \sum_{k=1}^{i-1} (\mathbf{u}_k^T \mathbf{u}_i) \mathbf{u}_k \quad \text{Gram-Schmidtの直交化}$$

$$\mathbf{u}_i \leftarrow \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|_2} \quad \text{ノルムの正規化}$$

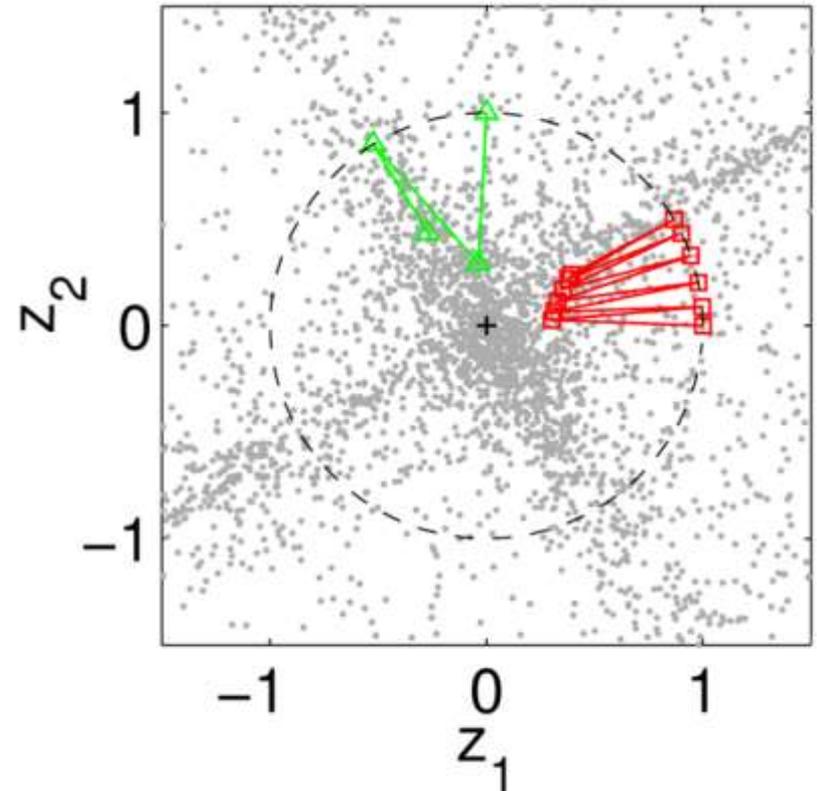
FastICAアルゴリズムによる推定の様子

■ 赤(□)

- $\mathbf{u}_1 = [1 \ 0]^T$ を初期値に設定
- Newton法による更新: 原点に向かっている
- ノルム1に正規化: 単位円上に引き戻されている
- 5回の繰り返しで良好な解へ

■ 緑(△)

- $\mathbf{u}_2 = [0 \ 1]^T$ を初期値に設定
- 直交化により, 1回だけで解に到達



最尤推定法

最尤推定法 (Infomax) [Bell et al., 1995]

- 分離行列 \mathbf{W} を直接推定
- 観測信号 $X = \{\mathbf{x}(t)\}_{1 \leq t \leq T}$ に対する \mathbf{W} の尤度関数

$$\mathcal{L}(\mathbf{W}) = \prod_{t=1}^T p(\mathbf{x}(t) | \mathbf{W}) \rightarrow \text{maximize}$$

- 線形変換と確率密度関数

$$\mathbf{y}(t) = \mathbf{W} \mathbf{x}(t) \longrightarrow \underline{p(\mathbf{y}(t))} = \frac{1}{|\det \mathbf{W}|} p(\mathbf{x}(t))$$

- 音源信号の独立性と非Gauss性を仮定

$$p(\mathbf{y}(t)) = \prod_{i=1}^I p(y_i(t)) \quad p(y_i(t)) : \text{Laplace分布など}$$

 $\log \mathcal{L}(\mathbf{W}) = T \log |\det \mathbf{W}| + \sum_{t=1}^T \sum_{i=1}^I \log p(y_i(t))$

最尤推定法

- 目的関数:

$$\mathcal{J}(\mathbf{W}) = \log |\det \mathbf{W}| + \frac{1}{T} \sum_t \sum_i \log p(y_i(t))$$

- 勾配法により \mathbf{W} を反復的に更新

$$\mathbf{W} \leftarrow \mathbf{W} + \eta \frac{\partial \mathcal{J}(\mathbf{W})}{\partial \mathbf{W}} \quad (\eta \text{ はステップサイズ})$$

$$\frac{\partial \mathcal{J}(\mathbf{W})}{\partial \mathbf{W}} = (\mathbf{W}^T)^{-1} - \frac{1}{T} \sum_t \Phi(\mathbf{y}(t)) \mathbf{x}(t)^T$$

$$\Phi(\mathbf{y}(t)) = \begin{bmatrix} \phi(y_1(t)) \\ \vdots \\ \phi(y_I(t)) \end{bmatrix}$$

$$\phi(y) = -\frac{\partial \log p(y)}{\partial y}$$

$\phi(y)$ の具体形

Laplace分布の場合: $\phi(y) = \text{sign}(y)$

先の近似分布の場合: $\phi(y) = \frac{y}{\sqrt{y^2 + \alpha}}$

分離信号ベクトル: $\mathbf{y}(t) = [y_1(t), \dots, y_2(t)]^T$

■ 2次統計量

$$\mathbf{E}\{\mathbf{y}(t)\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 信号間相関を最小化(複数時間区間利用)

■ 高次統計量1

$$\mathbf{E}\{\mathbf{y}^3(t)\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 高次相関をも最小化

■ 高次統計量2

$$\mathbf{E}\{\Phi(\mathbf{y}(t))\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 源信号確率密度関数を仮定

$\Phi(\bullet)$: 符号関数やtanh関数など

非線型関数2の導出

補足

■ 独立 ⇒ Kullback Leibler Divergenceの最小化問題

- 一般にKullback Leibler Divergenceとは2分布間の距離

$$KL(\mathbf{v}, \mathbf{z}) = \int p(\mathbf{z}) \log \frac{p(\mathbf{z})}{p(\mathbf{v})} d\mathbf{z}$$

上式において...

$p(\mathbf{z}) = p(y_1, \Lambda, y_K)$ 分離信号 $\mathbf{y}(t)$ の同時分布密度関数

$p(\mathbf{v}) = \prod_{k=1}^K p(y_k)$ 周辺分布密度関数の積

とおき, これらのKLを分離行列 W に関して最小化すれば独立

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{k=1}^K p(y_k)} d\mathbf{y} \quad \Rightarrow \quad \text{最小化}$$

非線型関数2の導出(cont'd)

補足

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{k=1}^K p(y_k)} d\mathbf{y}$$

$$= -H(\mathbf{Y}; W) + \sum_{k=1}^K H(Y_k; W)$$

1. 結合エントロピー

2. 周辺エントロピー和

$$\begin{aligned} H(\mathbf{Y}; W) &= -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} \\ &= -\int p(\mathbf{x}) (\log p(\mathbf{x}) - \log |\mathbf{W}|) d\mathbf{x} \\ &= H(\mathbf{X}) + \log |\mathbf{W}| \quad (\ominus p(\mathbf{y}) = p(\mathbf{x}) / |\mathbf{W}|) \end{aligned}$$

$$\begin{aligned} H(Y_k; W) &= -\int p(\mathbf{y}) \log p(y_k) d\mathbf{y} \\ &= -\int p(\mathbf{x}) \log p(y_k) d\mathbf{x} \\ &\quad (\ominus p(\mathbf{x}) d\mathbf{x} = p(\mathbf{y}) d\mathbf{y}) \end{aligned}$$

非線型関数2の導出(cont'd)

補足

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{k=1}^K p(y_k)} d\mathbf{y}$$

$$= -H(\mathbf{Y}; W) + \sum_{k=1}^K H(Y_k; W)$$

1. 結合エントロピー

2. 周辺エントロピー和

$$H(\mathbf{Y}; W) =$$

=

=

この値を最大化

⇒ 音源間の関連を無くす

$$H(Y_k; W) =$$

=

この値を最小化

⇒ 個々の音を非ガウス化

非線型関数2の導出(cont'd)

補足

分離信号の同時確率密度と周辺確率密度積のKL擬距離 $KL(W)$ の W に関する勾配を求め、その逆方向に W を更新学習

$$\Delta \mathbf{W} \propto -\frac{\partial KL(\mathbf{W})}{\partial \mathbf{W}} = \left((\mathbf{W}^T)^{-1} - \int p(x) \phi(\mathbf{y}) \mathbf{x}^T d\mathbf{x} \right)$$

$$= \left((\mathbf{W}^T)^{-1} - E_x [\phi(\mathbf{y}) \mathbf{x}^T] \right)$$

$$= \left(\mathbf{I} - E_y [\phi(\mathbf{y}) \mathbf{y}^T] \right) (\mathbf{W}^T)^{-1}$$

$$\phi(\mathbf{y}) \equiv \left[\frac{\partial \log p(y_1)}{\partial y_1}, \dots, \frac{\partial \log p(y_K)}{\partial y_K} \right]^T$$

音声の場合はsign関数やsigmoidで近似可能

様々なバリエーション

- EMアルゴリズムによる $p(y)$ の同時推定
- 二次統計量による $p(y)$ の推定 + 高次統計量ICA [Saruwatari, ICASSP2009]
- ICAと音響信号処理の類似点を明らかにし、相補性を生かした高速収束アルゴリズム [Saruwatari, IEEE Trans. SAP 2003 & 2006]

自然勾配(Natural Gradient)法

- 勾配法では W の逆行列計算が厄介

$$\frac{\partial \mathcal{J}(W)}{\partial W} = (W^\top)^{-1} - \frac{1}{T} \sum_t \Phi(y(t))x(t)^\top$$

- 自然勾配(Natural Gradient)

$$\frac{\partial \mathcal{J}(W)}{\partial W} W^\top W = \left(I - \frac{1}{T} \sum_t \Phi(y(t))x(t)^\top \right) W$$

- 逆行列計算が不要
- Equivariance Property
混合行列 H の影響 (特異点に近くて不安定, etc.) を受けない
- アルゴリズムは以下のとおり

$$y(t) \leftarrow W x(t)$$

$$W \leftarrow W + \eta \left(I - \frac{1}{T} \sum_t \Phi(y(t))y^\top \right) W$$

自然勾配の導出 (1/2)

■ 最急降下方向

- W を $W + \epsilon dW$ に変化させたときに最も $\mathcal{J}(W + \epsilon dW) - \mathcal{J}(W)$ を小さくする dW が最急降下方向
- モデルの接空間がユークリッド的でない場合は「偏微分方向 \neq 最急降下方向」

■ dW のノルムの定義

- W における変化を単位行列 I に移し変えて

$$W + \epsilon dW = (I + \epsilon dW W^{-1})W$$

と変形し, 実質的な変化分を $dW W^{-1}$ と捉える

- $dW W^{-1}$ のFrobeniusノルム

$$\|dW W^{-1}\|_F^2 = \text{tr}(W^{-T} dW^T dW W^{-1}) = \text{tr}(W^{-1} W^{-T} dW^T dW)$$

を dW のノルムと定義

自然勾配の導出 (2/2)

■ 最急降下方向の導出

- $\text{tr}(\mathbf{W}^{-1}\mathbf{W}^{-\top}\text{d}\mathbf{W}^{\top}\text{d}\mathbf{W}) = 1$ という制約の下で $\mathcal{J}(\mathbf{W} + \epsilon\text{d}\mathbf{W}) - \mathcal{J}(\mathbf{W}) \simeq \epsilon\text{tr}(\nabla\mathcal{J}(\mathbf{W})^{\top}\text{d}\mathbf{W})$ を最小化する $\text{d}\mathbf{W}$ を求める最適化問題として定式化

- Lagrangeの未定乗数法

$$\frac{\partial}{\partial(\text{d}\mathbf{W})} \{ \epsilon(\nabla\mathcal{J}(\mathbf{W})^{\top}\text{d}\mathbf{W}) - \lambda(1 - \text{tr}(\mathbf{W}^{-1}\mathbf{W}^{-\top}\text{d}\mathbf{W}^{\top}\text{d}\mathbf{W})) \} = 0$$

→ Lagrangeの未定乗数

$$\Rightarrow \epsilon\nabla\mathcal{J}(\mathbf{W})^{\top} + 2\lambda\mathbf{W}^{-1}\mathbf{W}^{-\top}\text{d}\mathbf{W}^{\top} = 0$$

$$\Rightarrow \text{d}\mathbf{W} \propto -\nabla\mathcal{J}(\mathbf{W})\mathbf{W}^{\top}\mathbf{W}$$

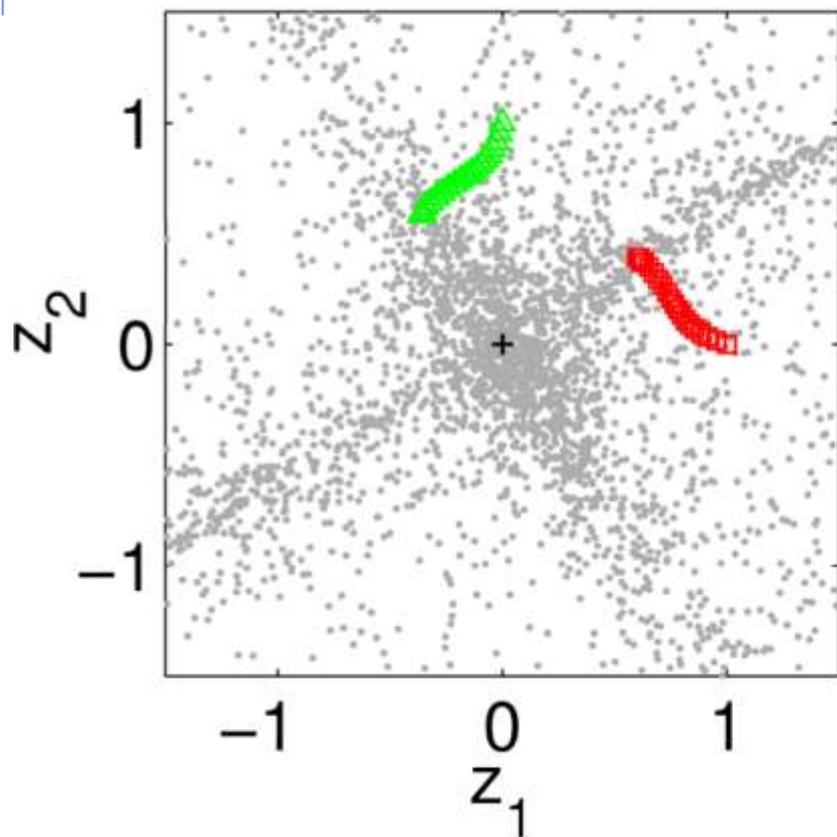
→ 逆行列をキャンセルしてくれる

\mathbf{W} の変化を単位行列に移し変えて考えているため、
 \mathbf{W} に依らず一定の収束特性を示す
 \Rightarrow 等価性 (Equivalence)

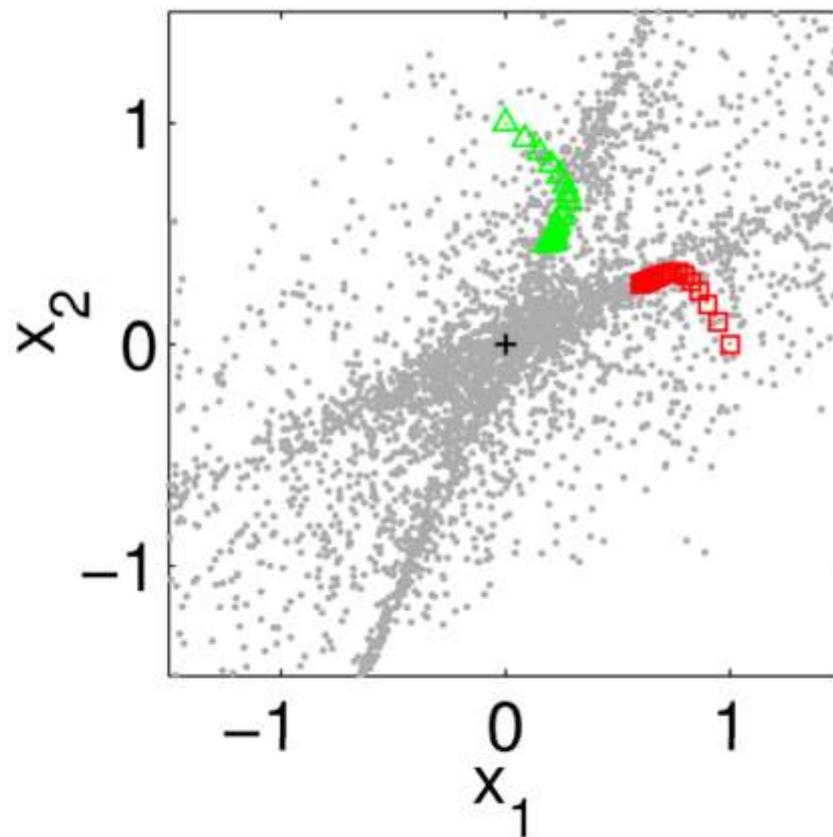
自然勾配法による推定の様子

- 赤(□): $w_1 = [1 \ 0]^T$ を初期値に設定
- 緑(△): $w_2 = [0 \ 1]^T$ を初期値に設定
- 観測信号の白色化をせずとも良好な解に収束

Whitened mixtures



Original mixtures



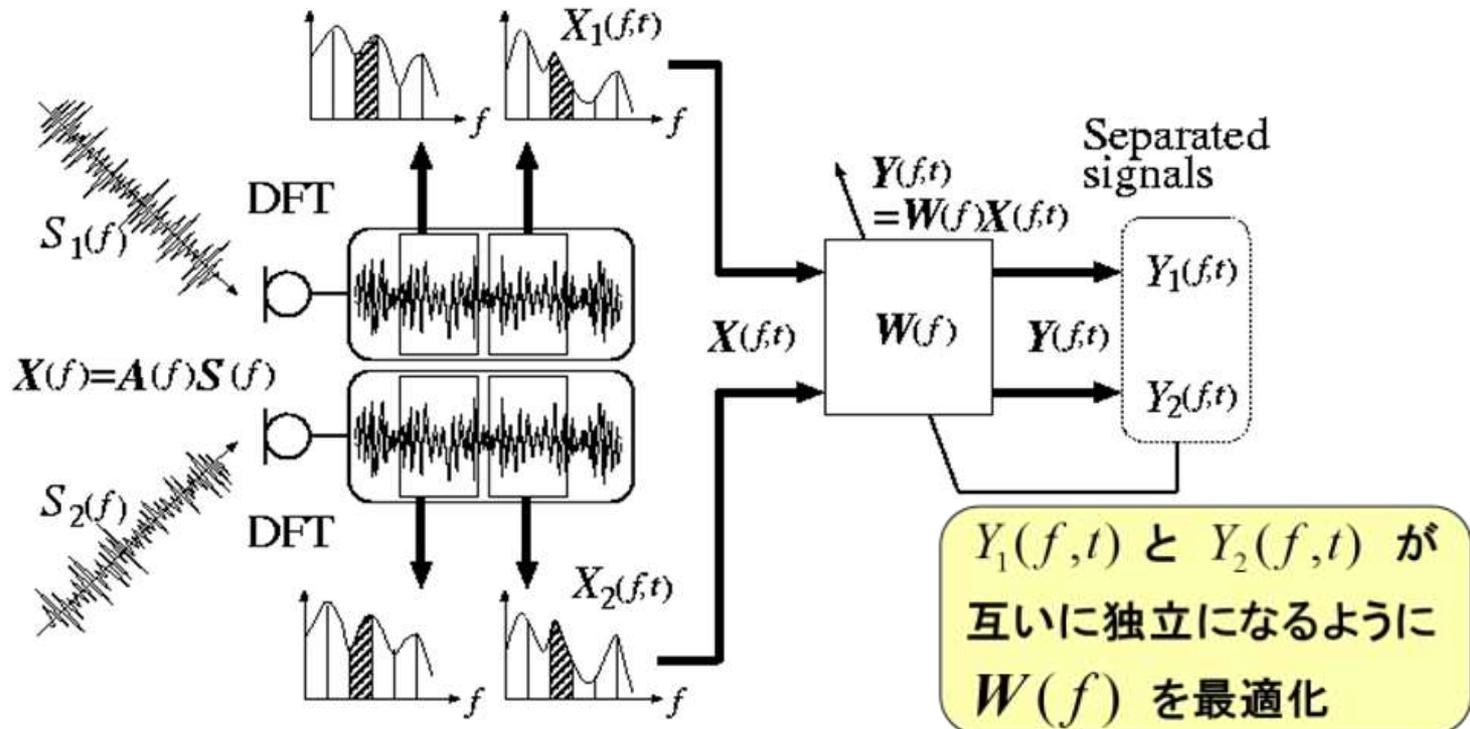
周波数領域ICA

周波数領域ICAによる音響信号BSS

• 周波数領域ICA (frequency-domain ICA: FDICA)

[Smaragdis, 1998], [Ikeda-Murata, 1998], [Saruwatari, 2000], etc.

- 音響信号は空間伝達関数(時間遅れ・畳み込み)を経て混合
- 時間領域では逆畳み込みの問題になってしまう
- 上記を解決するため、短時間フーリエ変換によって時間-周波数領域へ変換し、そのドメインで**周波数別**に分離行列 W をかける



周波数領域ICAによる音響信号BSS

• 周波数領域ICA (frequency-domain ICA: FDICA)

[Smaragdis, 1998], [Ikeda-Murata, 1998], [Saruwatari, 2000], etc.

- 音響信号は空間伝達関数(時間遅れ・畳み込み)を経て混合
- 時間領域では逆畳み込みの問題になってしまう
- 上記を解決するため、短時間フーリエ変換によって時間-周波数領域へ変換し、そのドメインで**周波数別**に分離行列 W をかける

周波数帯域別にICAを行うと...

ICAでは因子の順番は不問

⇒ 周波数帯域毎に分離信号が入れ替わってしまう

パーミュテーション問題

- 何らかの事前仮説が必要

「**プロジェクションバック**」法という

ICAでは因子の大きさは不問

⇒ 周波数帯域毎に分離信号の利得がバラバラに...

スケーリング問題

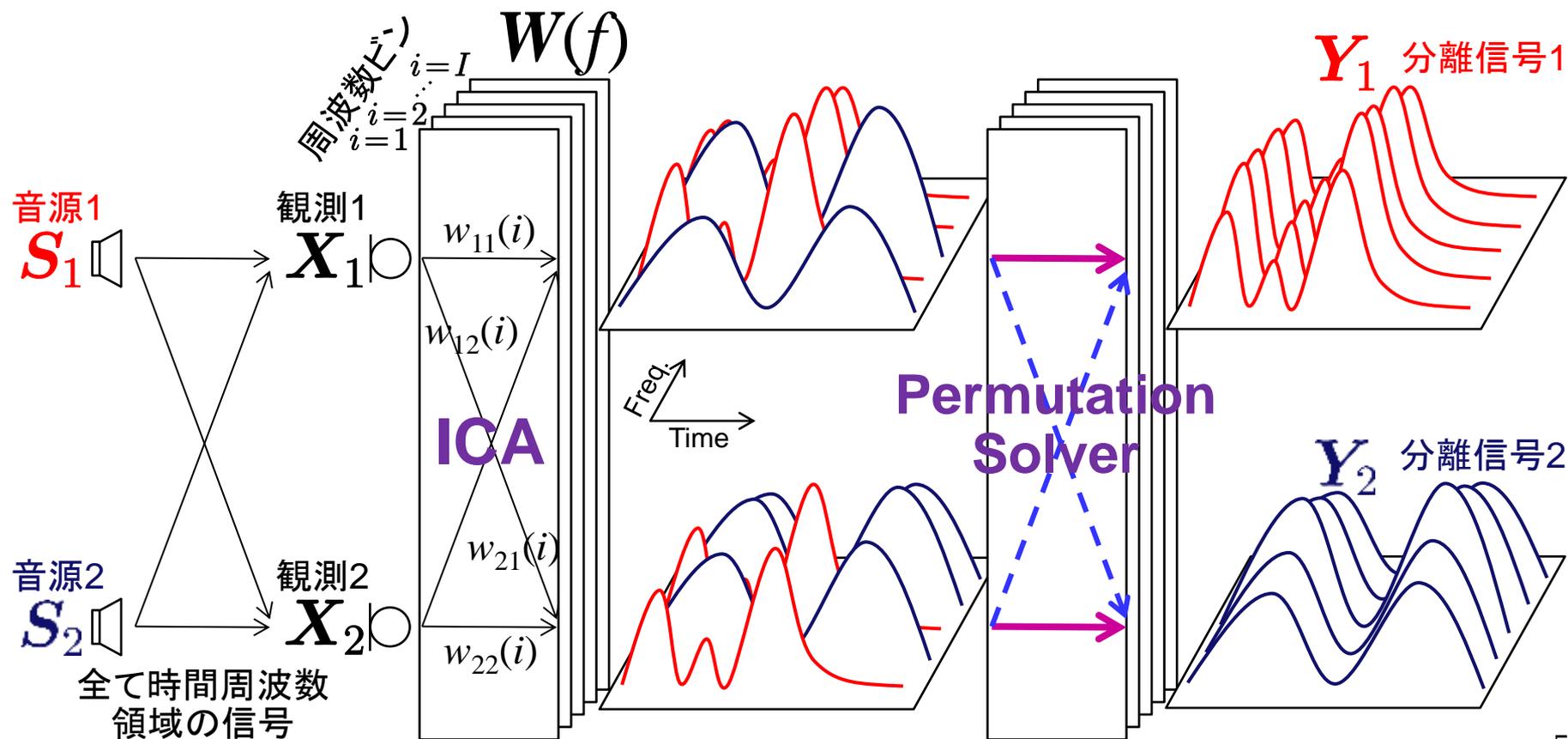
- 基準センサへの射影で解決
- $W^{-1} \text{diag}\{0, \dots, 1, \dots, 0\}$ の乗算

目的音出力番号

FDICAと音源パーミュテーション問題

FDICAに付随する問題

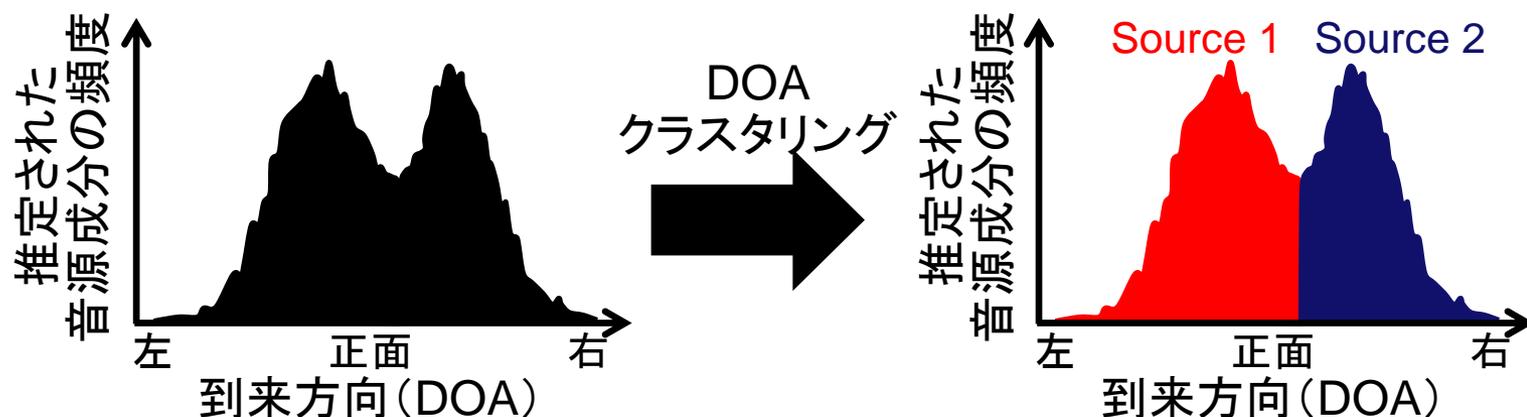
- スペクトログラムの周波数ビン毎に独立なICAを適用
- 音源及び空間モデルはパーミュテーション解決法に依存
(各周波数における音源順序不定性)



空間音響の視点からパーミュテーションを解く

• FDICA+DOAクラスタリング [Saruwatari 2000]

- 分離行列 $W(f)$ からステアリングベクトルを逆算し音源の到来方向 (DOA) でクラスタリングすることでパーミュテーションを解く



— 音源モデル

- 具体的なモデルスペクトログラムは無し (ICAの非ガウス性制約のみ)

— 空間モデル (「音源位置は物理的に点」仮定)

- **点DOAを中心に広がる混合分布**を成す ⇒ **クラスタリング**で解決

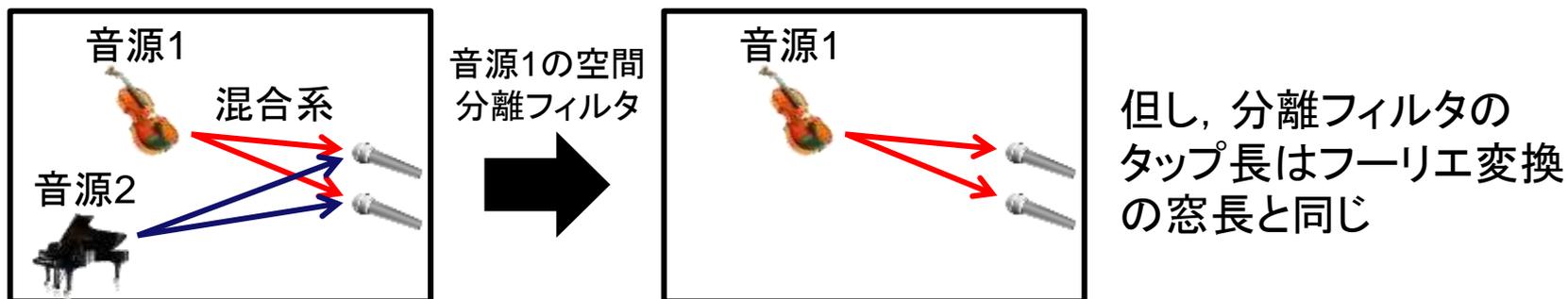
※事前情報の活用 [Parra, IEEE Trans. 2003] [Saruwatari, IEEE Trans. SAP 2006]

- **物理量DOAは点推定可能** ⇒ **DOA事前知識や正則化を導入可能**

ICAによる信号源分離の音響学的なメカニズム

• FDICAで推定される分離行列(フィルタ)とは？

- 周波数領域での瞬時混合を仮定
- **音源毎の空間特徴の違い**を用いた線形の空間分離フィルタ



– 適応ビームフォーマ (ABF) と本質的に等価 [Araki-Saruwatari, 2003]

- ABF: 妨害音のみがアクティブな時間の出力二乗誤差最小化
- 妨害音に対してヌル(感度0)を打つような空間分離
- ABFは音源位置とマイク位置が既知で音源アクティビティ検出が必要

信号
処理
観点

– FDICAはブラインドな音源分離手法

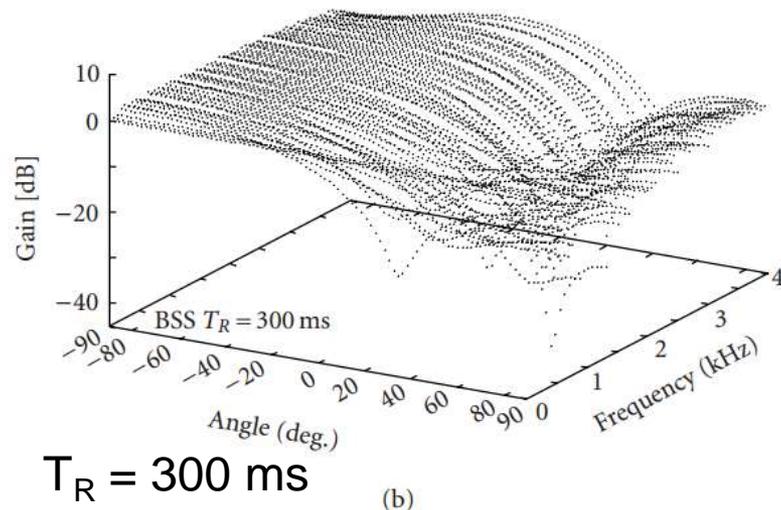
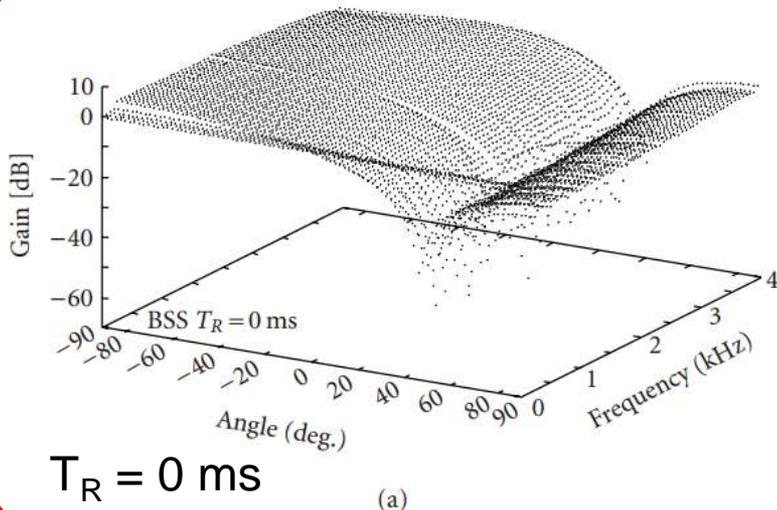
- 混合系未知, アクティビティ検出不要
- 厳密な音源位置とマイク位置が既知の場合のビームフォーマがFDICAの上限性能といえる

逆問
題的
観点

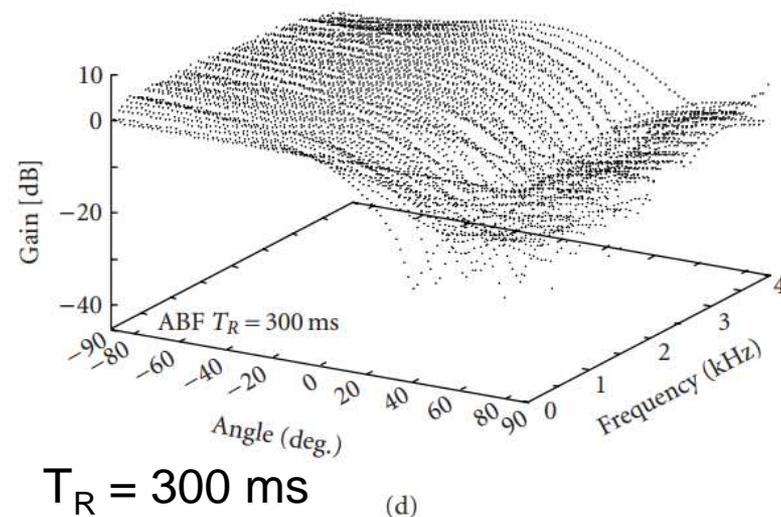
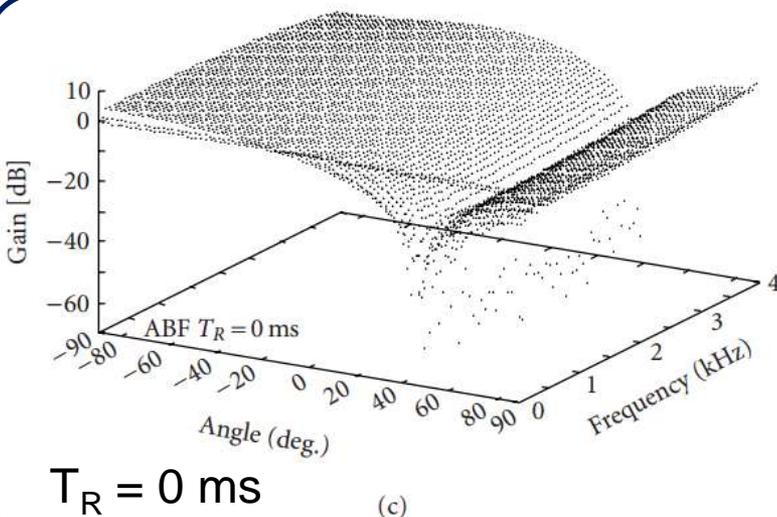
FDICAの分離フィルタとABFの分離フィルタ

- 図は [Araki-Makino-Saruwatari, 2003] より引用

BSSの
空間分離
フィルタ



ABFの
空間分離
フィルタ



勾配法の収束改善: 空間モデルによる正則化

- Geometric Source Separation (GSS) [Parra IEEE Trans. 2003]
 - 事前点DOA情報からの乖離を正則化(空間線スペクトルモデル制約)
 - 正則化重み(ハイパーパラメータ)は固定値

$$\text{コスト } L(W) = \sum_f \text{FDICA}(W(f)) + \lambda \cdot \text{BeamFormer}(\text{DOA})$$

↑
事前情報からimpose

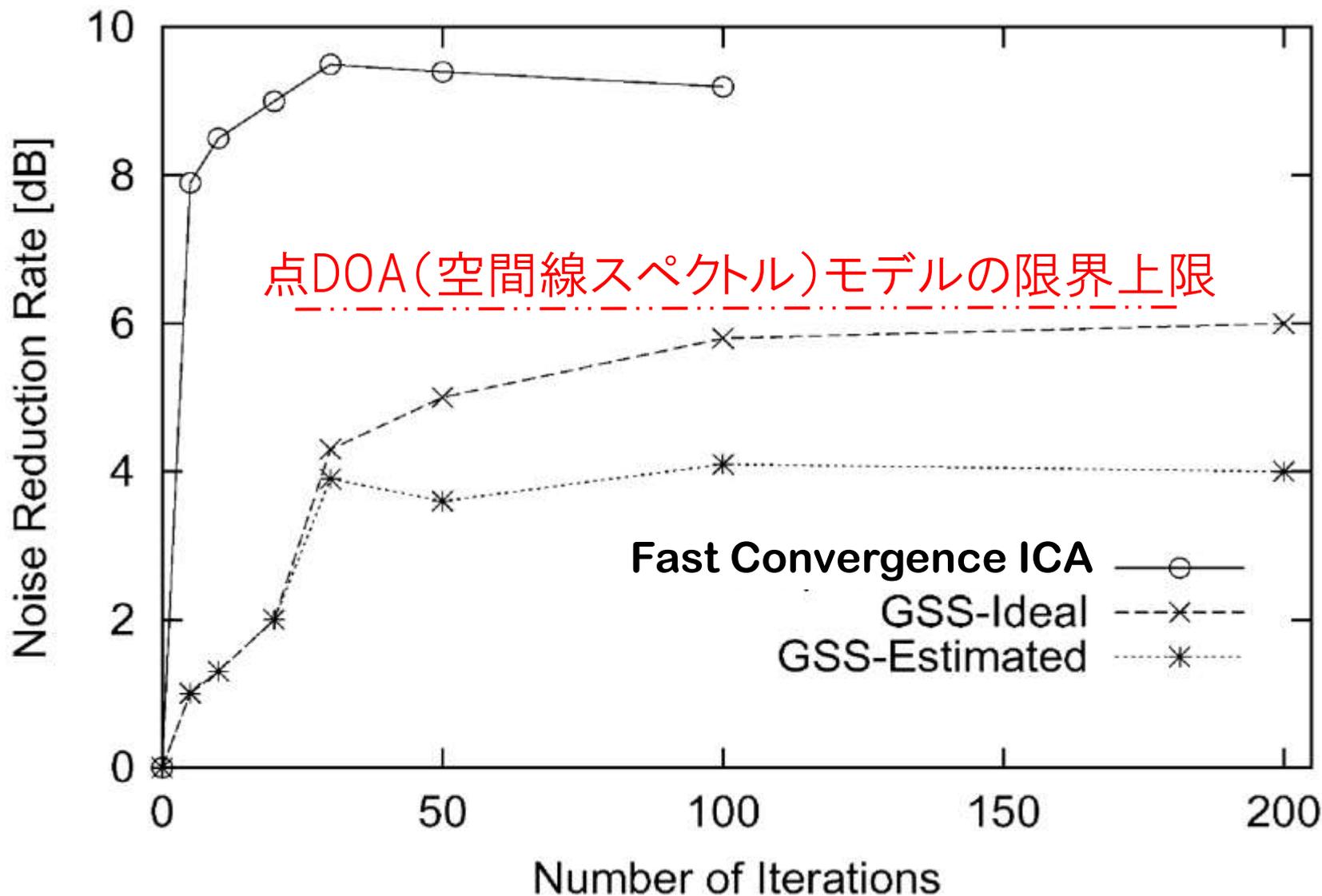
- Fast Convergence ICA [Saruwatari Eurospeech2001, IEEE Trans. 2006]
 - 空間モデルDOA期待値(平均DOA)をICAからempirical推定
 - 正則化重みは二次統計量から反復毎に自動選択される
 - 反復前半ではBF重視だが後半ではICA重視⇒DOA分散考慮

$$\text{コスト } L_f(W) = \lambda_1 \text{FDICA}(W(f)) + \lambda_2 \text{BeamFormer}(\text{DOA})$$

↑
空間モデル期待値 $\text{DOA} = E_f [\text{DOA}(W(f))]$

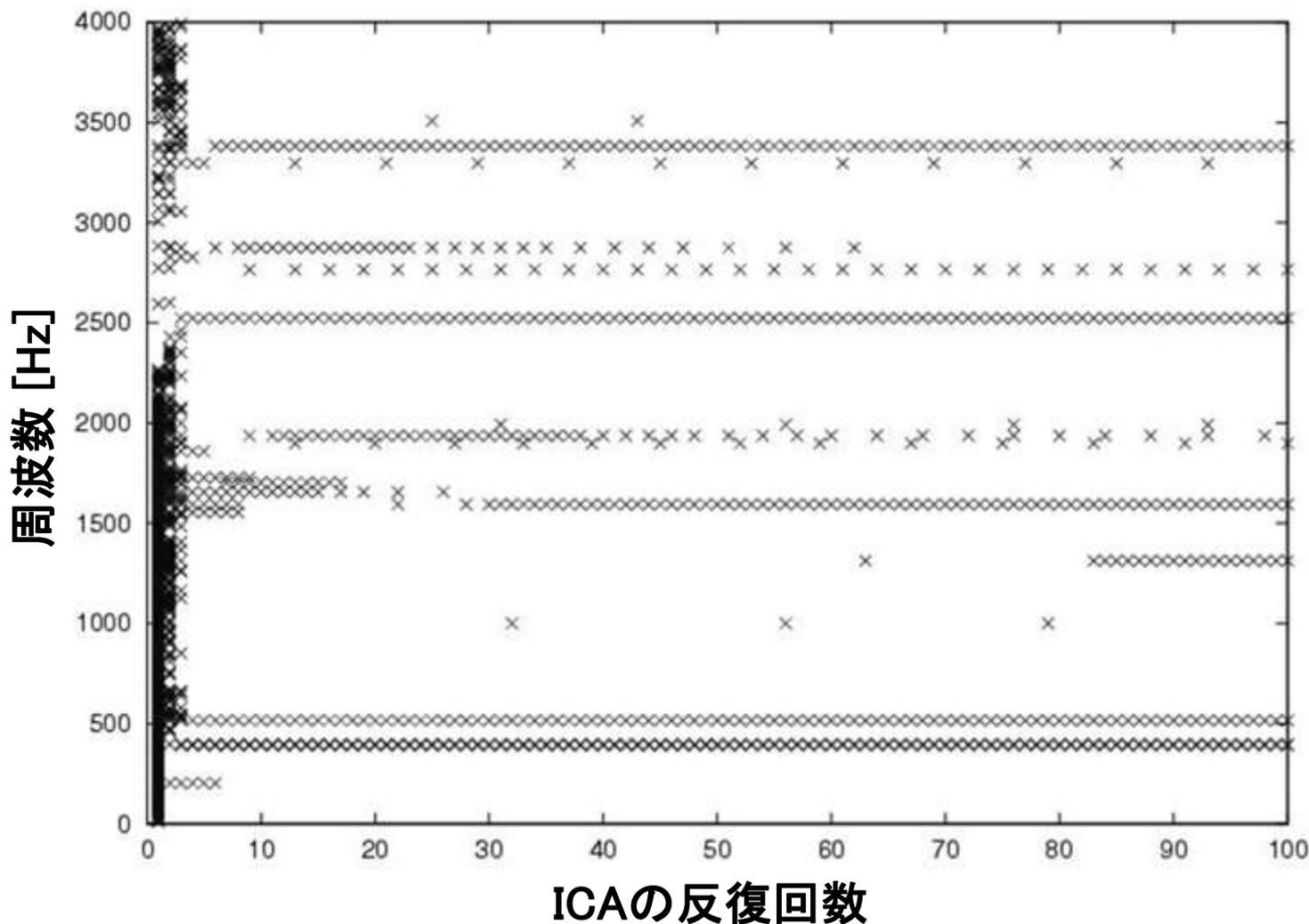
勾配法の収束改善：空間モデルによる正則化

300 ms残響音場における2音源分離の例：GSSの事前DOAは正解値 or 外部推定



勾配法の収束改善: 空間モデルによる正則化

$$\text{コスト } L_f(\mathbf{W}) = \lambda_1 \text{FDICA}(\mathbf{W}(f)) + \lambda_2 \text{BeamFormer}(\text{DOA})$$



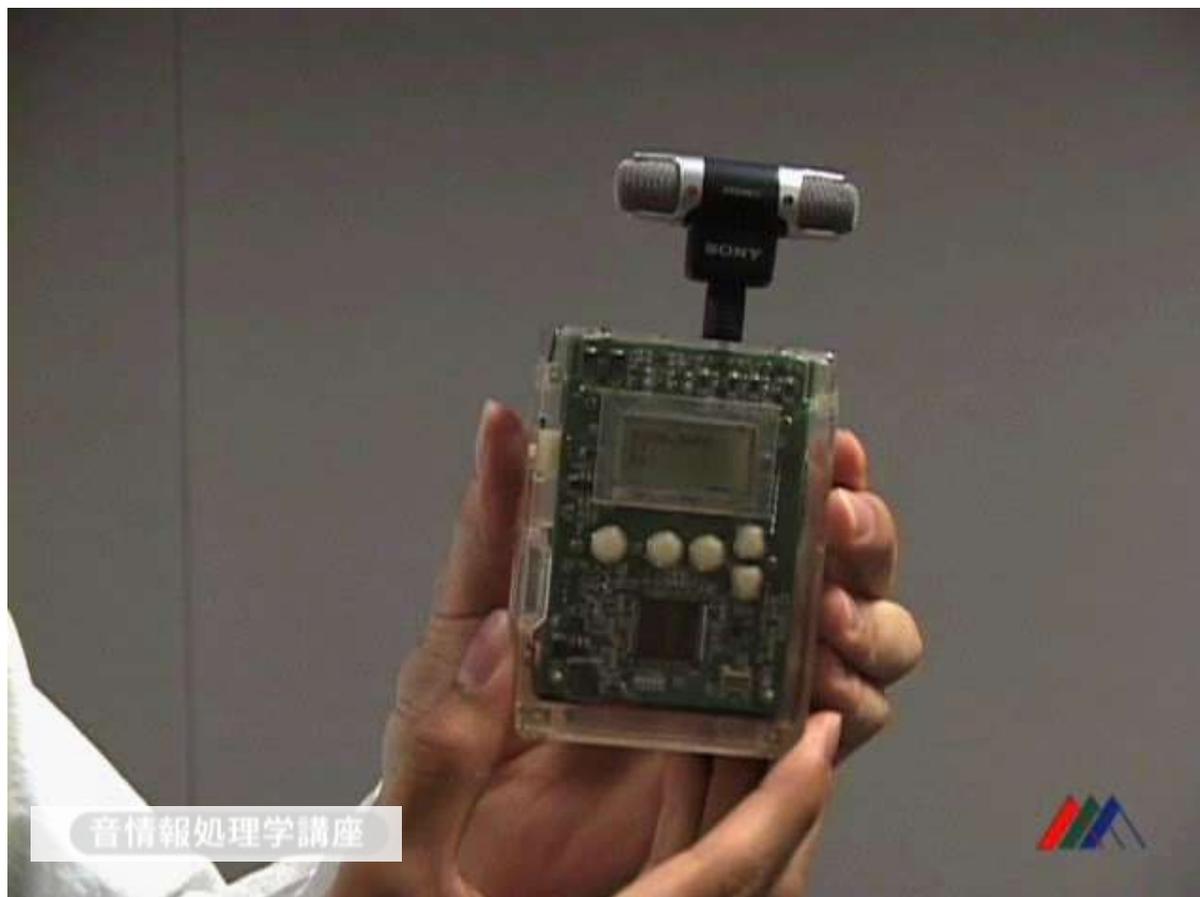
×印
 $\lambda_1=0, \lambda_2=1$
それ以外
 $\lambda_1=1, \lambda_2=0$

周波数領域ICAのまとめ

- GSS と Fast Convergence ICA の違いは、空間モデルに関する「実データに関する経験的知識(DOA分散の考慮)」の差であった(Parraは機械学習、Saruwatariは統計的音声信号処理出身)。
- [Araki-Makino-Saruwatari2003]の論文を皮切りにして、音響信号ICAの研究は様々な空間モデルの導入へと発展し、レガシーな音響信号処理(ビームフォーマ)との融合も盛んに行われた。
- 一方、学習理論研究者の興味は「音源モデル」の方にもあった。時系列としてのモデリングは多々拡張されたが、スペクトログラム(2次元データ)としてのモデリング拡張はまだ未開発であった(2006年のIVA登場まで)。

リアルタイムICAの実用化 [Saruwatari, et al. 2009]

- 4chマイクとDSP (TI社製C67) から構成される独立成分分析BSS専用モジュールを2005年に開発
- 2009年に世界初の商用化 (警察備品に採用)



ロボット対話 デモンストレーション

Demonstration for
spoken-oriented hands-free
interactive robot.

[Takahashi, Saruwatari+, IEEE Trans. ASLP 2009]

人工知能研究会優秀賞・平成23年度RIEC Award を受賞

目次

- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

独立ベクトル分析及びその補助関数による解法に関して、発明者であるNII小野順貴先生より資料のご提供を受けました。ここに深く感謝いたします。

IVAの登場

パーミュテーションフリー なICAを目指して

周波数成分間の高次相関を用いたBSS(2006~)

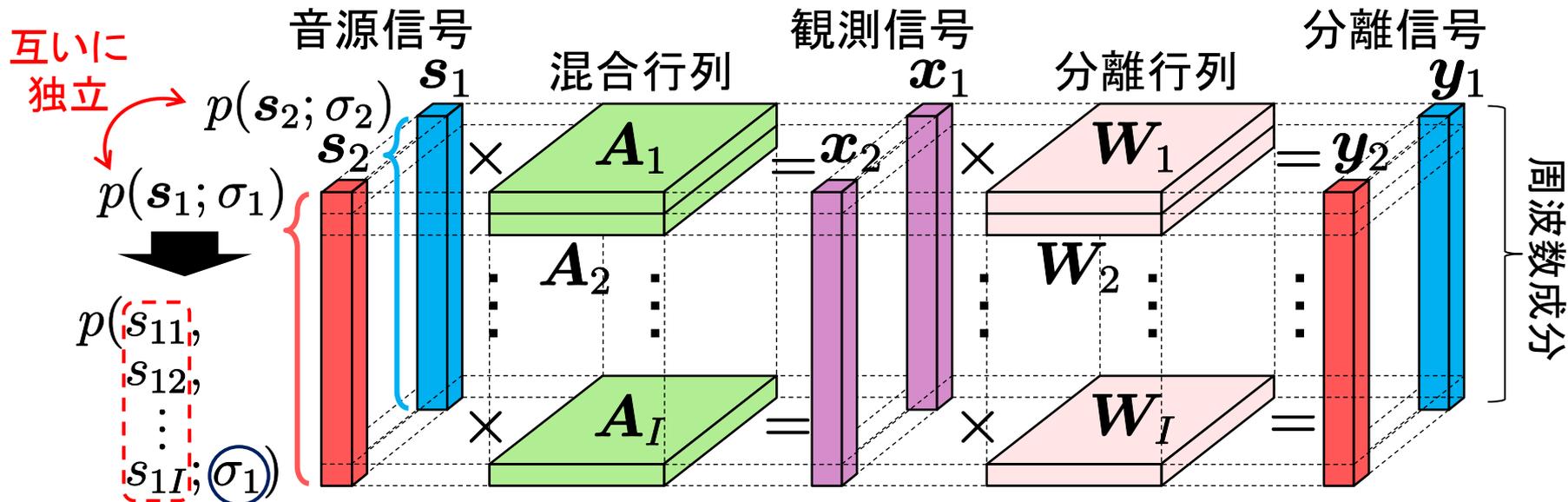
独立ベクトル分析 (independent vector analysis: IVA)

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

– FDICAの多変量拡張手法, パーミュテーション問題を回避

– 球対称な多変量(多次元)分布モデルの最尤推定

- ICAにおける音源の事前分布 $p(s_n)$ を多次元分布 $p(s_{n1}, \dots, s_{nI}; \sigma_n)$ に拡張
- 球対称性: 無相関な周波数成分間で同じ分散を持つ多変量分布



周波数成分間
は無相関だが
依存性をもつ

共通の分散
を持つ

IVAの仮定する周波数成分間の依存性

- FDICAとIVAの違いはコントラスト関数のみ
 - コントラスト関数: パラメタの対数尤度のgradient

$$\varphi_i(s_{n1}, \dots, s_{nI}) = -\frac{\partial p(s_{n1}, \dots, s_{nI}; \sigma_n)}{\partial s_{ni}}$$

- 「音源の事前分布が一次元か多次元か」のみ
- IVAの仮定する音源の事前分布 $p(\mathbf{s}_n)$
 - ラプラス分布の例 (音声信号のモデルとして一般的)

周波数毎に独立な事前分布

$$p(\mathbf{s}_n) = \prod_{i=1}^I p(s_{ni}) = \alpha \prod_{i=1}^I \exp\left(-\frac{|s_{in} - \mu_{in}|}{\sigma_{ni}}\right)$$

周波数間で高次相関をもつ事前分布

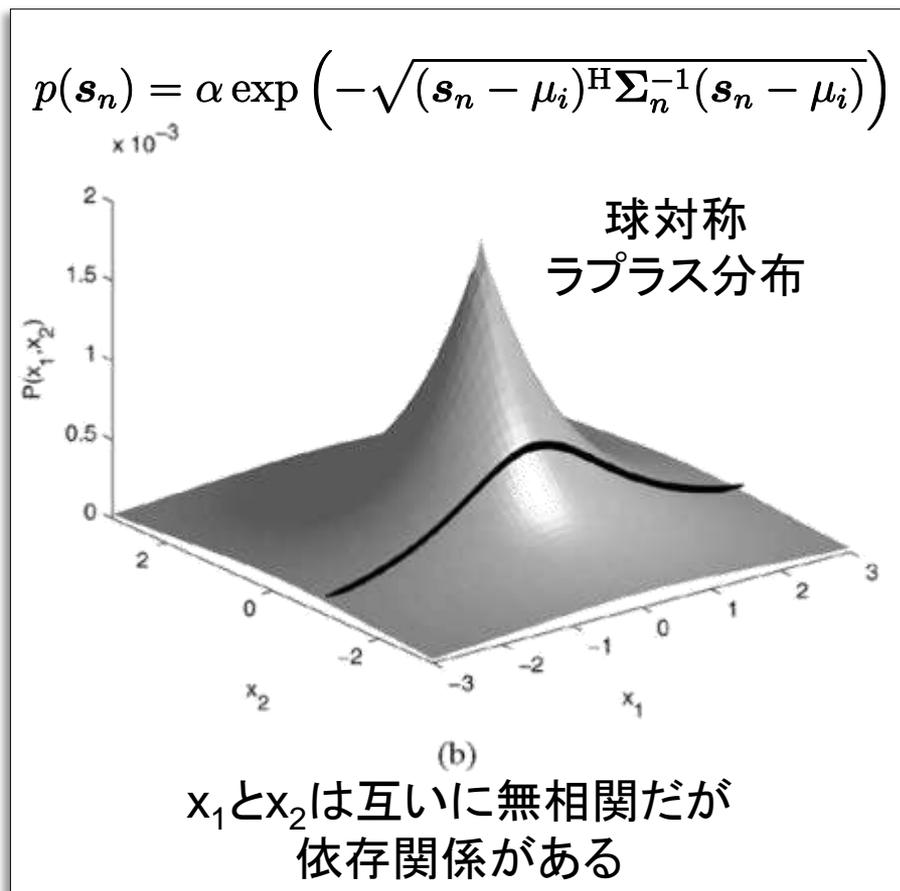
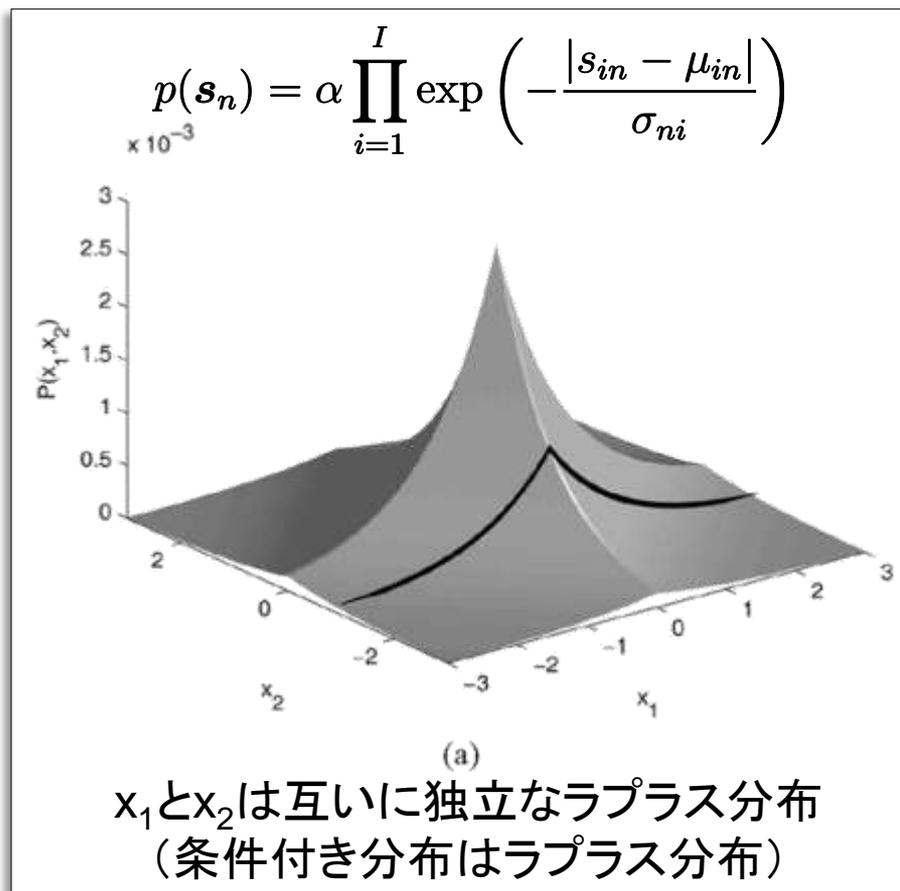
$$p(\mathbf{s}_n) = \alpha \exp\left(-\sqrt{(\mathbf{s}_n - \boldsymbol{\mu}_i)^H \boldsymbol{\Sigma}_n^{-1} (\mathbf{s}_n - \boldsymbol{\mu}_i)}\right)$$

分散共分散行列

- 後者は $\boldsymbol{\Sigma}_n = \mathbf{I}$ (互いに無相関) の場合でも, s_{ni} が互いに依存

IVAの仮定する周波数成分間の依存性

- 図は [Kim, 2007] より引用



Higher-order correlation
Higher-order dependency

IVAの仮定する周波数成分間の依存性

- IVAの音源事前分布の分散に関して
 - 分散はその周波数ビンの信号のスケールに対応
 - 残念ながら独立性基準ではスケール自体の推定はできない
 - 分散を全周波数で1にしても問題ない \blacksquare 共通分散(球対称)
 - どのみち「プロジェクションバック」法で復元できるので問題にはならない

$$p(\mathbf{s}_n) = \alpha \exp \left(-\sqrt{(\mathbf{s}_n - \boldsymbol{\mu}_i)^H \boldsymbol{\Sigma}_n^{-1} (\mathbf{s}_n - \boldsymbol{\mu}_i)} \right)$$



零平均, 零共分散, 共通分散とすると

$$p(\mathbf{s}_n) = \alpha \exp \left(-\sqrt{\sum_{i=1}^I \left| \frac{s_{ni}}{\sigma_n} \right|} \right)$$

- 球対称事前分布の定性的な効果
 - 周波数間で同じアクティベーションを持つ成分を一つの音源としてまとめる傾向にある \blacksquare パーミュテーション問題の自動回避
 - スペクトログラム(二次元)音源モデルの先駆けでもあった

補助関数法の導入

[小野順貴先生2013年5月音声研究会講演資料より抜粋]

補助関数型IVA [Ono 2010]: 補助関数法とは？

■ 解くべき最適化問題

$$J(\Theta) \rightarrow \min$$

■ 補助関数

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

■ 反復更新則

補助変数の更新 (E-step)

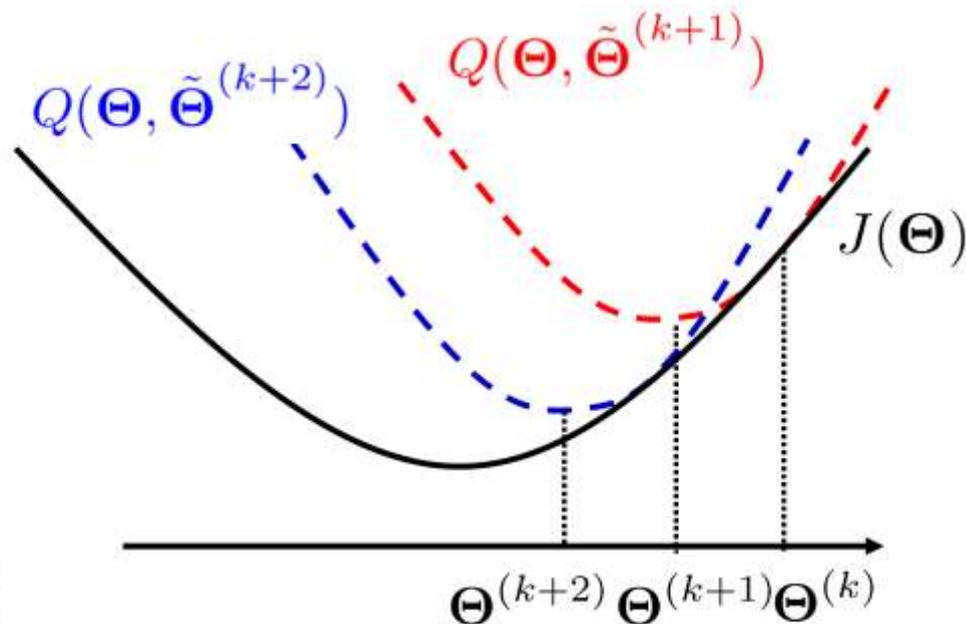
$$\tilde{\Theta}^{(k+1)} = \operatorname{argmin}_{\tilde{\Theta}} Q(\Theta^{(k)}, \tilde{\Theta})$$

パラメータの更新 (M-step)

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

■ 補助関数の例(絶対値関数の最小化)

$$|x| \leq \frac{1}{2x_0} |x|^2 + \frac{x_0}{2} \quad \Rightarrow \quad \begin{array}{l} \text{二次関数の和も二次関数なので、} \\ \text{絶対値関数の和も容易に最小化できる} \end{array}$$



利点: 目的関数の収束を保証しつつ比較的容易に反復最小化可能

補助関数型IVA: 補助関数の定義

- IVAのコントラスト関数として以下の関数を考えてみる

$$G(\mathbf{z}) = G_R(\|\mathbf{z}\|_2) \quad \begin{array}{l} \mathbf{z} \text{ は多次元複素変数} \\ \text{球対称(ノルムだけの)分布} \end{array}$$

IVAのコスト関数 J 及びその補助関数 Q を以下のように定める

$$J(\mathbf{W}) = \frac{1}{N_\tau} \sum_{\tau=1}^{N_\tau} \sum_{k=1}^K \underline{G(\mathbf{y}_k(\tau))} - \sum_{\omega=1}^{N_\omega} \log |\det W(\omega)|$$



$$Q(\mathbf{W}, \mathbf{V}) = \sum_{\omega=1}^{N_\omega} \left[\frac{1}{N_\tau} \sum_{\tau=1}^{N_\tau} \sum_{k=1}^K \underline{\mathbf{w}_k^h(\omega) V_k(\omega) \mathbf{w}_k(\omega)} - \log |\det W(\omega)| \right]$$

補助変数

$$V_k(\omega) = \frac{1}{N_\tau} \sum_{\tau=1}^{N_\tau} \left[\frac{G'(r_k(\tau))}{r_k(\tau)} \mathbf{x}(\omega, \tau) \mathbf{x}^h(\omega, \tau) \right] \quad r_k(\tau) = \sqrt{\sum_{\omega=1}^{N_\omega} |\mathbf{w}_k^h(\omega) \mathbf{x}(\omega, \tau)|^2}$$

補助関数型IVA: パラメータ更新

- 重み付き共分散行列(補助変数)の更新

$$V_k(\omega) = \frac{1}{N_\tau} \sum_{\tau=1}^{N_\tau} \left[\frac{G'(r_k(\tau))}{r_k(\tau)} \mathbf{x}(\omega, \tau) \mathbf{x}^h(\omega, \tau) \right] \quad r_k(\tau) = \sqrt{\sum_{\omega=1}^{N_\omega} |\mathbf{w}_k^h(\omega) \mathbf{x}(\omega, \tau)|^2}$$

- 補助関数 Q が小さくなるように W の更新

$$\frac{\partial Q}{\partial w_1} = 0$$

$$\frac{\partial Q}{\partial w_2} = 0$$

$$\frac{\partial Q}{\partial w_3} = 0$$

$$\mathbf{w}_1^h V_1 \mathbf{w}_1 = 1$$

$$\mathbf{w}_2^h V_1 \mathbf{w}_1 = 0$$

$$\mathbf{w}_3^h V_1 \mathbf{w}_1 = 0$$

$$\mathbf{w}_1^h V_2 \mathbf{w}_2 = 0$$

$$\mathbf{w}_2^h V_2 \mathbf{w}_2 = 1$$

$$\mathbf{w}_3^h V_2 \mathbf{w}_2 = 0$$

$$\mathbf{w}_1^h V_3 \mathbf{w}_3 = 0$$

$$\mathbf{w}_2^h V_3 \mathbf{w}_3 = 0$$

$$\mathbf{w}_3^h V_3 \mathbf{w}_3 = 1$$

どう
解く
のか？

$K=2$ ならば一般化固有値問題(閉解法あり)、 $K=3$ 以上は閉形式解が無い。
分離行列を行ベクトル毎に更新することは閉形式で可能(補助関数は減少する)。

補助関数型IVA: パラメータ更新

- 重み付き共分散行列(補助変数)の更新

$$V_k(\omega) = \frac{1}{N_\tau} \sum_{\tau=1}^{N_\tau} \left[\frac{G'(r_k(\tau))}{r_k(\tau)} \mathbf{x}(\omega, \tau) \mathbf{x}^h(\omega, \tau) \right] \quad r_k(\tau) = \sqrt{\sum_{\omega=1}^{N_\omega} |\mathbf{w}_k^h(\omega) \mathbf{x}(\omega, \tau)|^2}$$

- 補助関数 Q が小さくなるように W の更新

【IP(反復射影)アルゴリズム】

for $k = 1 : K$

$$\mathbf{w}_k(\omega) \leftarrow (W(\omega)V_k(\omega))^{-1} \mathbf{i}_k$$

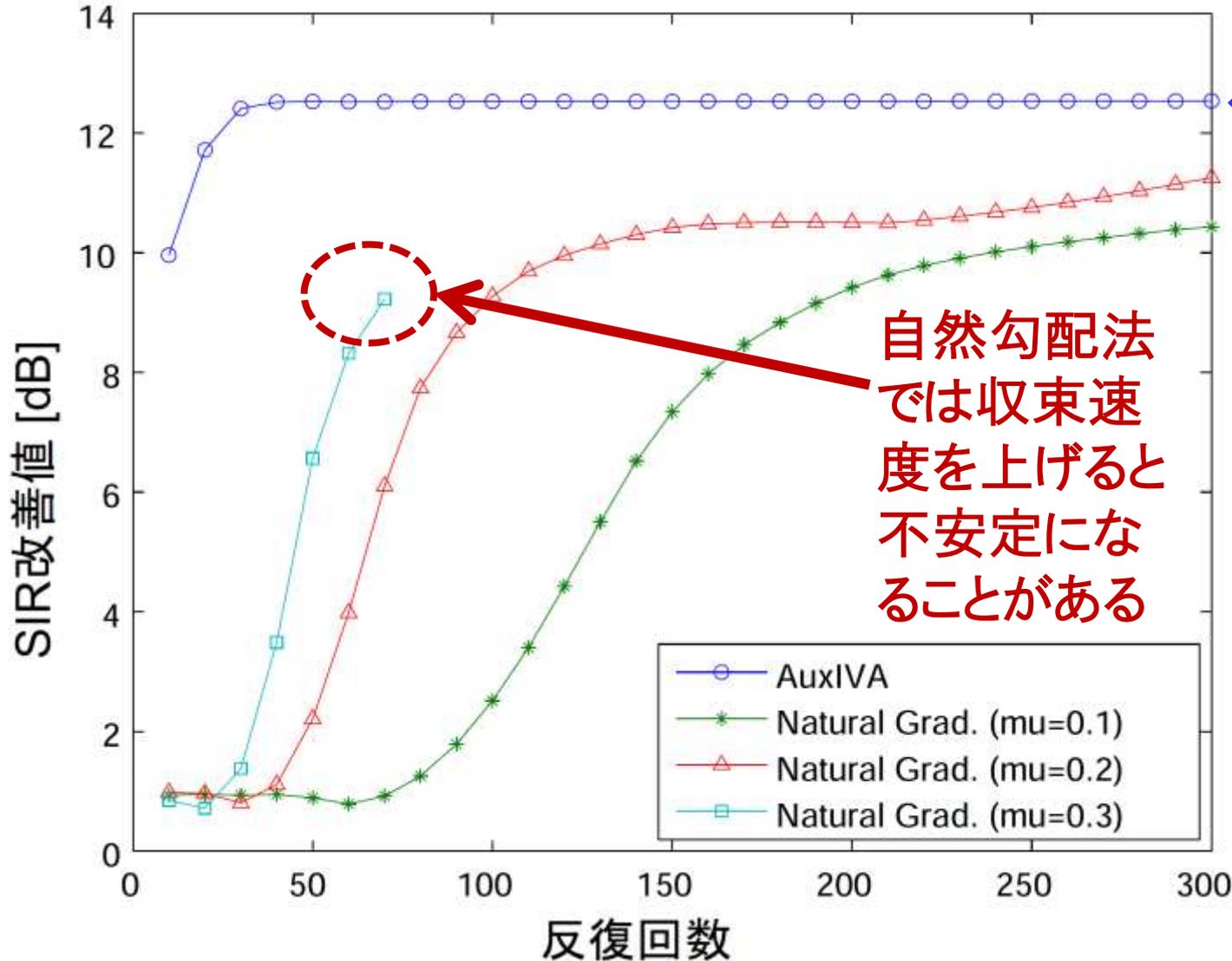
$$\mathbf{w}_k(\omega) \leftarrow \mathbf{w}_k(\omega) / \sqrt{\mathbf{w}_k^h(\omega) V_k(\omega) \mathbf{w}_k(\omega)}$$

end

(\mathbf{i}_k は第 k 要素のみ
が1の単位ベクトル)

$K=2$ ならば一般化固有値問題(閉解法あり)、 $K=3$ 以上は閉形式解が無い。
分離行列を行ベクトル毎に更新することは閉形式で可能(補助関数は減少する)。

■ 2音源分離の場合: 自然勾配法によるIVAと比較



自然勾配法では収束速度を上げると不安定になることがある

補助関数型IVAでは非常に高速かつ高精度な分離が達成されている

目次

- ブラインド音源分離について
 - 定式化, 歴史, 応用
- 信号の統計的性質
 - 信号を混ぜるー 中心極限定理
- 独立成分分析のアルゴリズム
 - 白色化 + FastICA
 - 最尤推定法 by Natural Gradient
 - 周波数領域における独立成分分析
- 独立ベクトル分析
 - パーミュテーションフリー
 - 補助関数法の導入
- 独立低ランク行列分析
 - NMFと融合した多チャネルブラインド音源分離

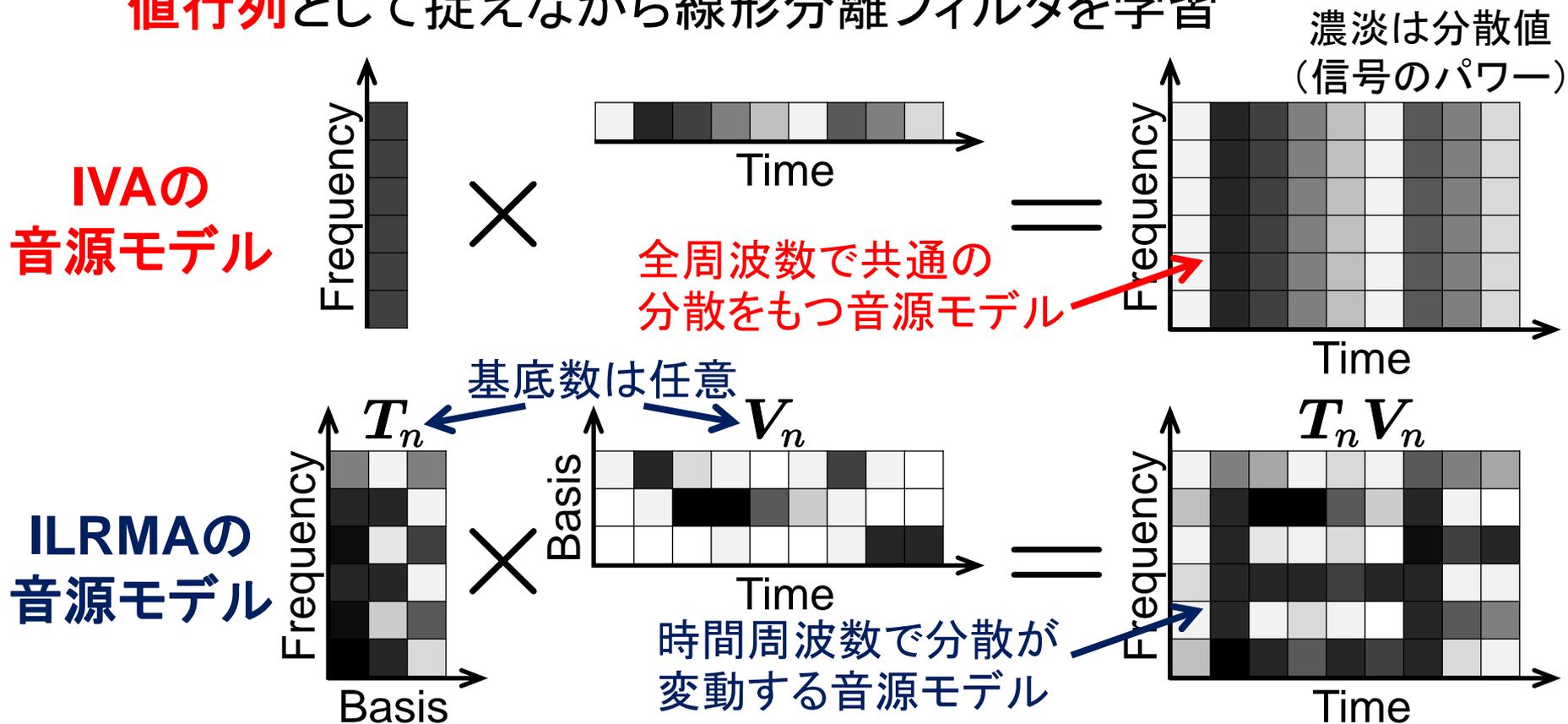
独立低ランク行列分析

NMFと融合した多チャネルBSS

IVAとNMFを融合した新しいBSS(2016~)

- 独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA)
[Kitamura-Ono-Sawada-Kameoka-Saruwatari, ICASSP2015, IEEE-Trans.2016]

- 独立音源の詳細な時間周波数構造(のパワー)を**低ランク非負値行列**として捉えながら線形分離フィルタを学習



- 板倉斎藤擬距離基準NMFの事前分布を活用

板倉斎藤擬距離基準NMFにおける生成モデル

• 従来のNMF分解の問題点

- データ行列(非負実数)は1本の基底と1本のアクティベーションからなる**ランク1行列の線形結合**として表現

$$\boxed{X} = \boxed{t_1 v_1} + \boxed{t_2 v_2} + \boxed{t_3 v_3} + \dots$$

- X は振幅スペクトログラムなのか？あるいはパワーなのか？
- いずれにしても線形結合(加法性)は成り立たない
 - 理論的には複素スペクトログラムの加法モデルが正しい
- 位相スペクトログラムはどうするのか？

• 板倉斎藤擬距離基準NMFでは下記のように解決される

- 複素スペクトログラムに対する生成モデルが与えられる
- **複素数成分の線形結合なので理論的に正しいモデル**
- 位相は無情報な形(一様分布)で保持される

板倉斎藤擬距離基準NMFにおける生成モデル

- 板倉斎藤擬距離基準NMF (Itakura-Saito NMF: ISNMF)

[Févotte, 2009]

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

最小化は等価  複素球対称ガウス分布 (零平均)

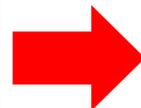
観測の複素数値 ← 複素ガウスの分散

$$\begin{aligned} \mathcal{L} &= - \sum_{i,j} \log \mathcal{N}_c(x_{ij} | 0, \sum_k t_{ik} v_{kj}) \\ &= IJ \log \pi + \sum_{i,j} \left[-\log \left(\sum_k t_{ik} v_{kj} \right) + \frac{|x_{ij}|^2}{\sum_k t_{ik} v_{kj}} \right] \end{aligned}$$

- この生成モデルはガウス分布の再生性を用いて分解可

– $x_{ijk} = \sum_{k=1}^K c_{ijk}$ とおくと

$$x_{ij} \sim \mathcal{N}_c(x_{ij} | 0, \sum_k t_{ik} v_{kj})$$



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

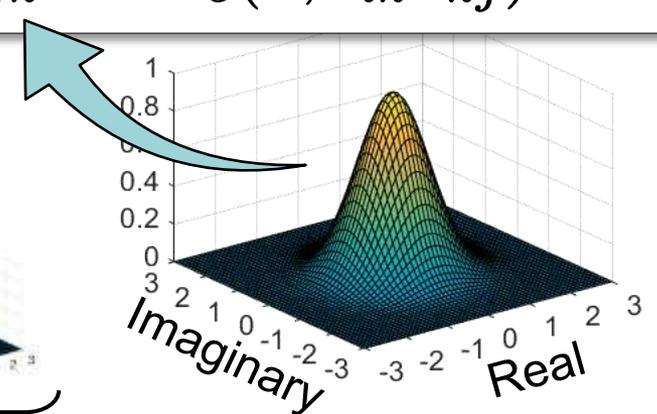
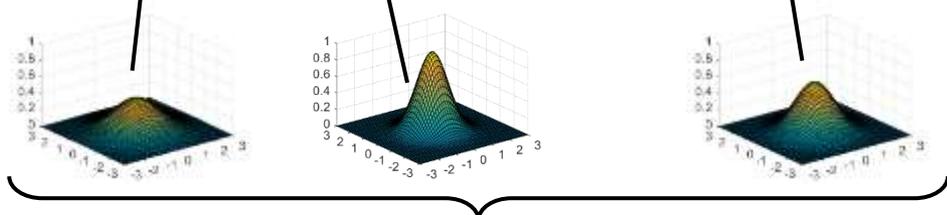
板倉斎藤擬距離基準NMFにおける生成モデル

- X を複素スペクトログラムとする(STFTしたそのもの)
 - 各時間周波数要素は複素要素 c_{ijk} を K 個足し合わせたもの

とある時間周波数要素 $x_{ij} = \sum_{k=1}^K c_{ijk}$

平均0, 分散 $t_{ik}v_{kj}$ の複素球対称ガウス分布
 $c_{ijk} \sim \mathcal{N}_c(0, t_{ik}v_{kj})$

$$x_{ij} = c_{ij1} + c_{ij2} + \dots + c_{ijK}$$



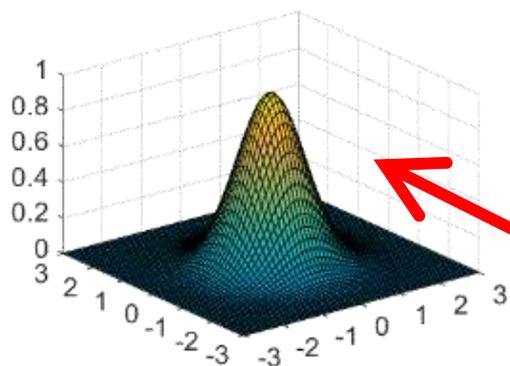
これらの複素ガウス分布は互いに独立 (i.i.d.)

- 複素ガウス分布の線形結合なので x_{ij} も複素ガウス分布
 - ガウス分布の再生性
 - x_{ij} の複素ガウス分布の分散は $\sum_{k=1}^K t_{ik}v_{kj}$
- 分散が時間周波数で変動する複素ガウス分布が生成モデル

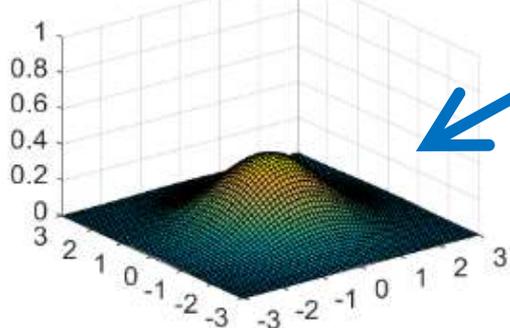
板倉斎藤擬距離基準NMFにおける生成モデル

- パワースペクトログラムは複素ガウスの分散に対応

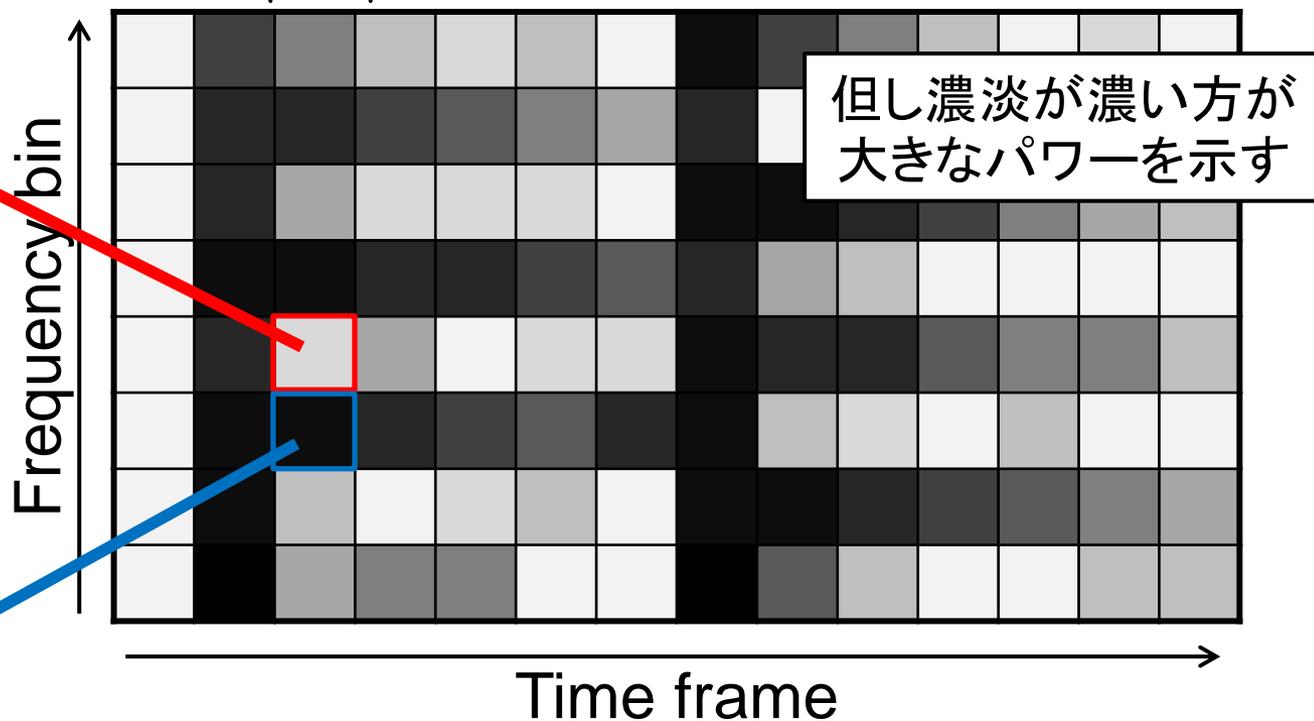
$|X|^2$: パワースペクトログラム



パワーが小＝分散が小
殆ど0付近の複素数しか
生成しない



パワーが大＝分散が大
大きな振幅の複素数も
生成する



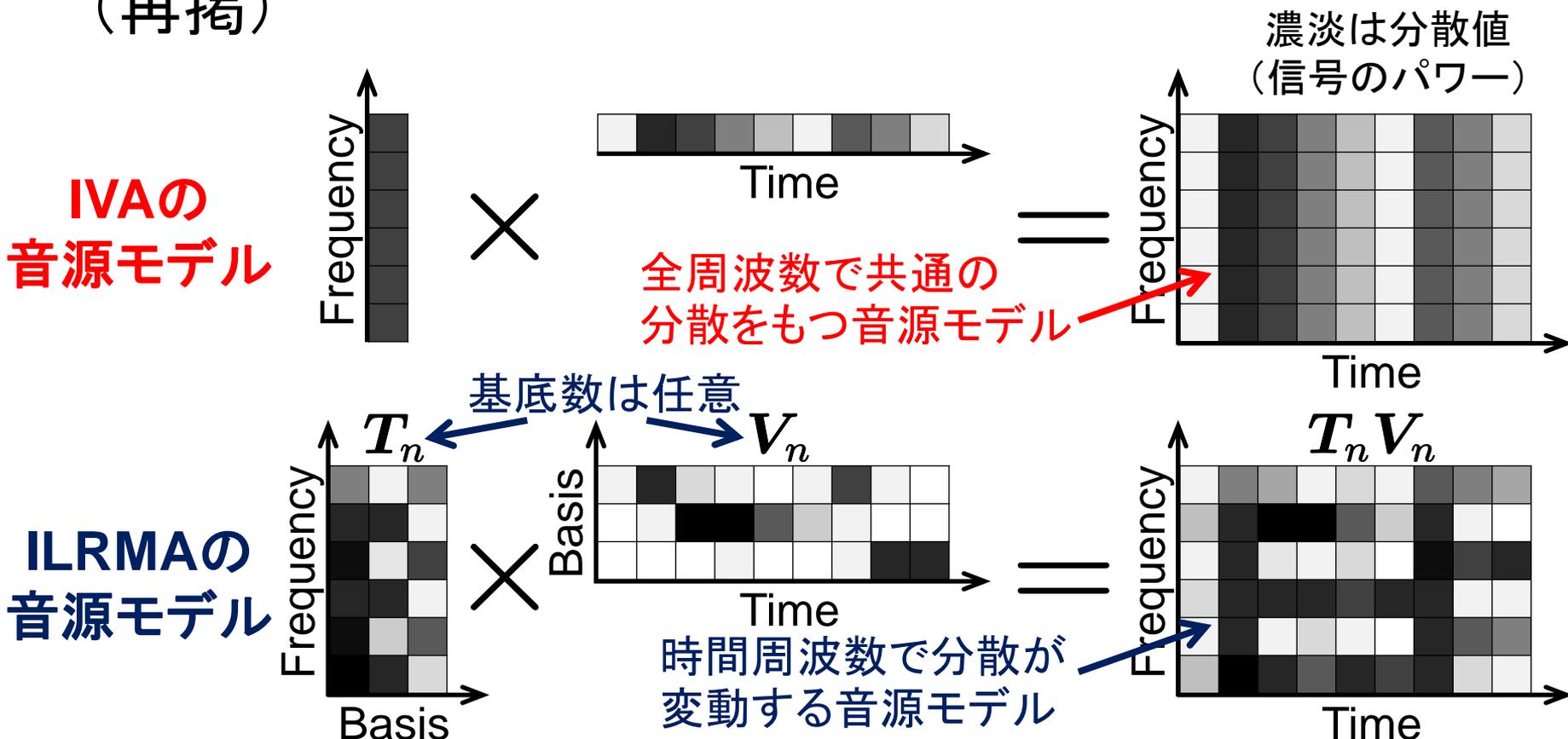
各時間周波数で分散が変動する複素ガウス分布

巨視的(マクロ)に考えると分散が逐一変動する為、スペクトログラム全体の密度分布はスーパーガウシアン(カートシスがガウス分布より大)な分布になっている

IVAとNMFを融合した新しいBSS(2016~)

- 独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA)

(再掲)



- 但し, IVAは共通分散だがスケールは周波数毎に不定なので, フラットなスペクトルだけではなく任意の形状を表現可能

ILRMAのコスト関数と更新則

- ILRMAのコスト(対数尤度)関数

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

主なパラメータ

空間モデル

\mathbf{W}_i : 分離フィルタ行列

$y_{ij,m} = \mathbf{w}_{i,m}^H \mathbf{x}_{ij}$: 分離信号

音源モデル

t_{ik} : 基底行列要素

(TV : K ランクNMFモデル)

v_{kj} : アクティベーション行列要素

z_{mk} : 分割関数(基底の共有を許す重み)

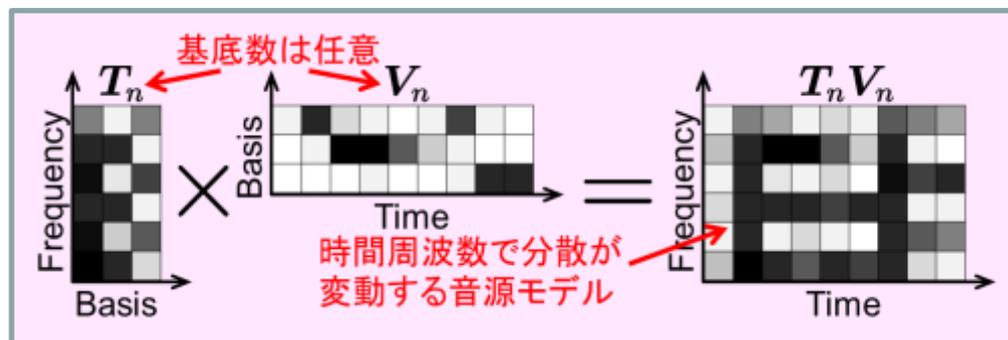
ILRMAのコスト関数と更新則

- ILRMAのコスト(対数尤度)関数→これを最小化

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

音源の低ランク性コスト関数
(ISNMFモデルの推定に寄与)

音源の独立性コスト関数
(空間モデル W の推定に寄与)



$$p(\mathbf{y}) = p(y_1) p(y_2) \dots p(y_m)$$

となる W を推定

両者を交互にMajorization-Minimization(補助関数)アルゴリズムで反復最小化

- ✓ コスト値の単調減少性を保証(勾配法には無い特徴)
- ✓ 高速かつ安定な求解法を実現(従来の多入力NMFと比較して2ケタ速い)

ILRMAのコスト関数と更新則

- ILRMAコスト関数はIVAコストとNMFコストの複合
 - 全変数の最適化は**全て補助関数法に基づく更新式**
 - 分離フィルタは**IP更新**、音源モデルは**乗法更新**

分離フィルタと分離信号の更新

$$V_{i,m} = \frac{1}{J} \sum_j \frac{1}{r_{ij,m}} \mathbf{x}_{ij} \mathbf{x}_{ij}^H$$

$$\mathbf{w}_{i,m} \leftarrow (\mathbf{W}_i V_{i,m})^{-1} \mathbf{e}_m$$

$$\mathbf{w}_{i,m} \leftarrow \mathbf{w}_{i,m} (\mathbf{w}_{i,m}^H V_{i,m} \mathbf{w}_{i,m})^{-\frac{1}{2}}$$

$$y_{ij,m} \leftarrow \mathbf{w}_{i,m}^H \mathbf{x}_{ij}$$

但し, $\mathbf{W}_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,M})^H$,
 \mathbf{e}_m は m 番目の要素のみ1で
 他が0の縦ベクトル

音源モデルと推定分散の更新

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

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

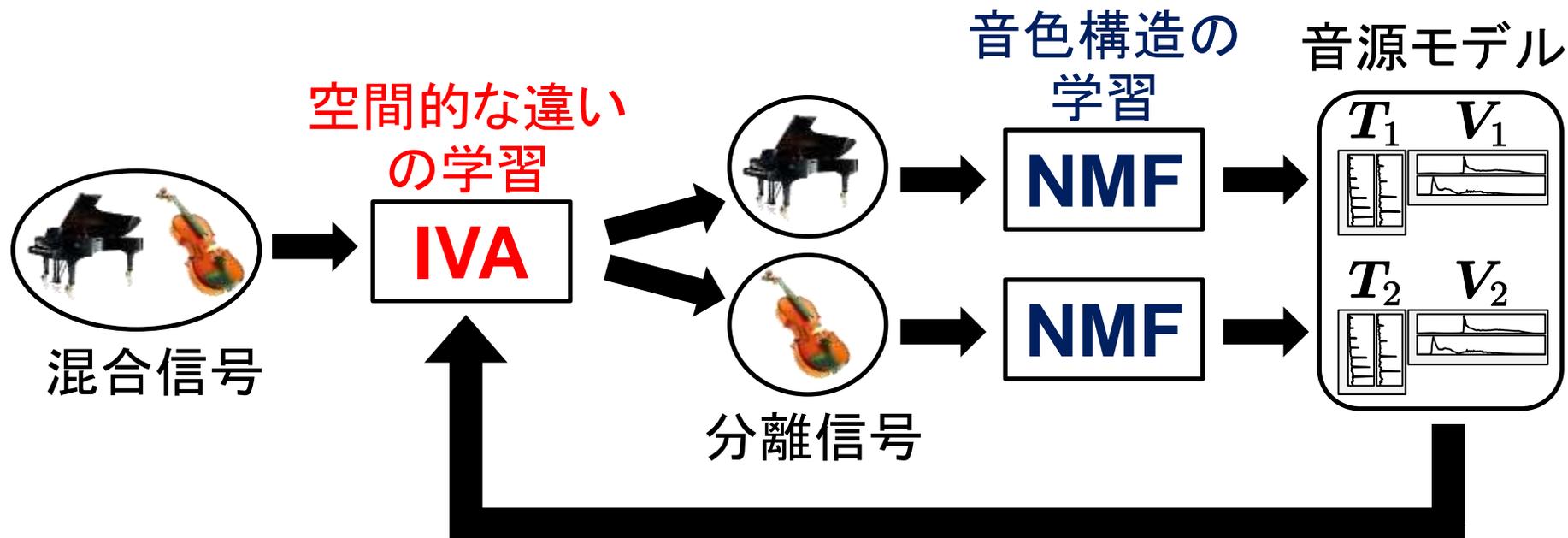
$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,m} |y_{ij,m}|^2 z_{mk} t_{ik} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-2}}{\sum_{i,m} z_{mk} t_{ik} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-1}}}$$

$$r_{ij,m} = \sum_k z_{mk} t_{ik} v_{kj} : \text{音源毎の推定分散}$$

- 複雑なスペクトrogram音源モデルを内包しつつ、反復更新で尤度が単調増加することを保証するアルゴリズム (cf. 自然勾配)

ILRMAのコスト関数と更新則

- 音源毎の空間的な違い(空間モデル)と各音源の音色構造(音源モデル)を交互に学習



- 音源毎の時間周波数構造を正確に捉えることで、独立性基準での線形時不変空間分離の性能向上が期待できる
- 従来のICAと異なり、全パラメータを補助関数法で更新することにより収束性を保証する

アルゴリズム概観

「生成モデル」 vs. 「分離モデル」

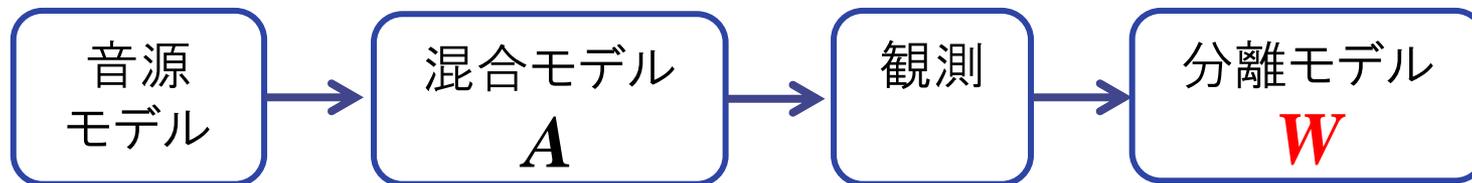
音源分離問題における双対性

生成モデル型



- 音源と混合モデル A のパラメータを同時に推定
- 混合系が劣決定の場合でも適用可能
- 多チャンネルNMFなどに利用されるモデル

分離モデル型



- 音源と分離モデル $W=A^{-1}$ のパラメータを同時に推定
- 混合系がランク1空間混合の場合に適用可能
- ICA, IVA, ILRMA等利用されるモデル

何が違うのか？どちらが良いのか？

FDICAにおける最適化: 自然勾配法

- 勾配法では W の逆行列計算が厄介

$$\frac{\partial \mathcal{J}(W)}{\partial W} = (W^T)^{-1} - \frac{1}{T} \sum_t \Phi(y(t)) x(t)^T$$

- 自然勾配 (Natural Gradient) [Amari, 1997]

$$\frac{\partial \mathcal{J}(W)}{\partial W} W^T W = \left(I - \frac{1}{T} \sum_t \Phi(y(t)) x(t)^T \right) W$$

分離フィルタ
のみの更新

- 逆行列計算が不要 (音響系では「逆フィルタ」計算が不要！)
- 混合系の推定よりも安定
- ビームフォーマとの類似性
- アルゴリズムは以下のとおり

$$y(t) \leftarrow W x(t)$$

$$W \leftarrow W + \eta \left(I - \frac{1}{T} \sum_t \Phi(y(t)) y^T \right) W$$

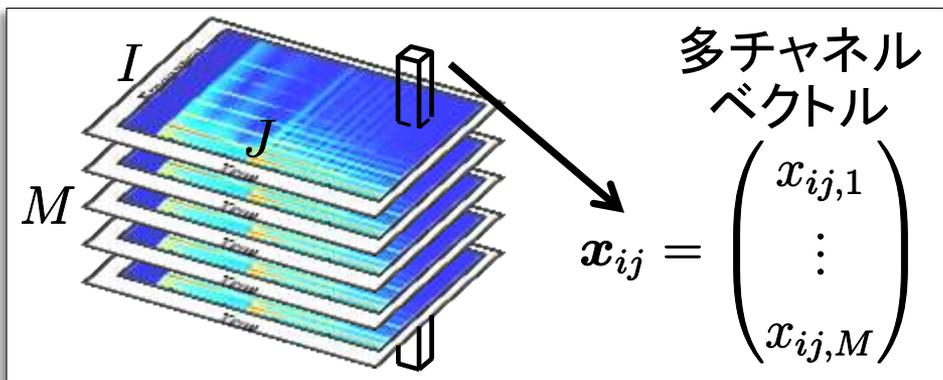
生成モデル型の例：多チャンネルNMF

- 全音源の時間周波数構造をNMFでモデル化し音源毎の空間伝達特性(空間相関行列 R)を用いてクラスタリング
 - 「NMFモデル」と「音源毎の空間相関行列」を同時に推定

提案者と年代	空間的な混合系と音源モデル	各音源の空間相関行列	音源のスペクトログラム	各変数の最適化手法
Ozerov and Fevotte, 2010	$R^{(x)} = R^{(s)} \circ TV + R^{(b)}$	$R^{(s)}$ はランク1 $R^{(b)}$ はフルランク	NMF	EMで $R^{(s)}$, $R^{(b)}$, NMF変数を推定
Arberet et al., 2010	$R^{(x)} = R^{(s)} \circ TV + R^{(b)}$	$R^{(s)}$ も $R^{(b)}$ もフルランク	NMF	EMで $R^{(s)}$, $R^{(b)}$, NMF変数を推定
Ozerov et al., 2011	$R^{(x)} = (R^{(s)} Z) \circ TV + R^{(b)}$	$R^{(s)}$ はランク1 $R^{(b)}$ はフルランク	分割関数付きのNMF	EMで $R^{(s)}$, $R^{(b)}$, NMF変数を推定
Sawada et al., 2013	$R^{(x)} = (R^{(s)} Z) \circ TV$	フルランクの $R^{(s)}$	分割関数付きのNMF	乗算更新式で $R^{(s)}$, NMF変数を推定
Kitamura et al., 2015	$R^{(x)} = (R^{(s)} Z) \circ TV$	ランク1の $R^{(s)}$	分割関数付きのNMF	反復射影(IP)で分離行列 $W = A^{-1}$, 乗算更新式で NMF変数を推定

NMFの多チャンネル信号への拡張

- 最も一般化された多チャンネルNMF [Sawada, 2013]

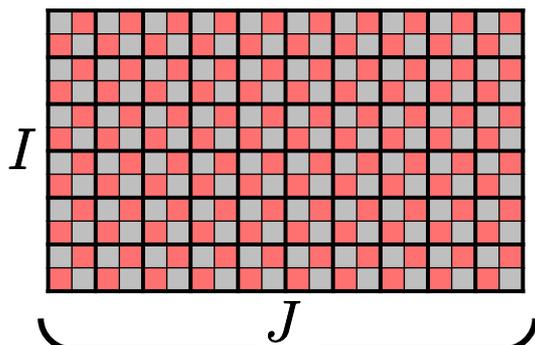


瞬時空間(チャンネル間)相関行列

$$\begin{aligned} X_{ij} &= \mathbf{x}_{ij} \mathbf{x}_{ij}^H \\ &= \begin{pmatrix} |x_{ij,1}|^2 & \cdots & x_{ij,1} x_{ij,M}^* \\ \vdots & \ddots & \vdots \\ x_{ij,M} x_{ij,1}^* & \cdots & |x_{ij,M}|^2 \end{pmatrix} \end{aligned}$$

時間周波数毎の観測チャンネル間相関

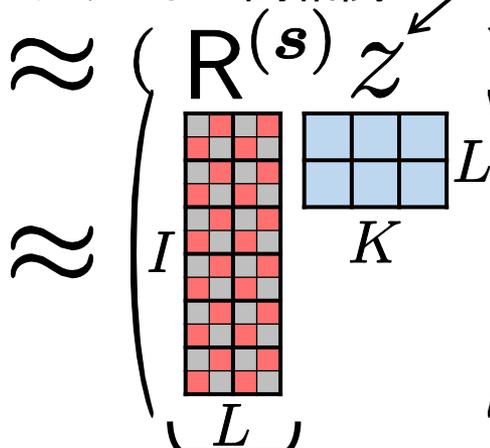
\mathbf{X}



多チャンネル観測信号

音源周波数毎のチャンネル間相関

\approx

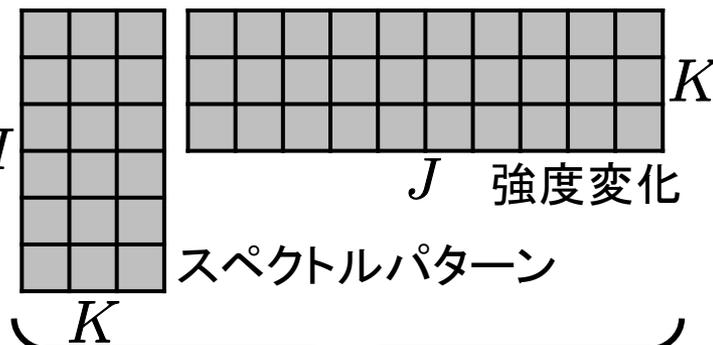


空間モデル
音源毎の空間的な違い

基底行列 アクティベーション行列

\mathbf{T}

\mathbf{V}



音源モデル
全音源の音色構造

空間相関行列とは

- 空間相関行列 又は 空間共分散行列 [Duong, 2010]
 - Duong modelとも呼ばれる
 - 音源とマイク間の伝達系と音響的拡散度合を含む特徴量
 - ステアリングベクトルの拡張
 - 観測信号 \mathbf{x}_{ij} 中の n 番目の音源成分のみを $\mathbf{x}_{ij}^{(n)}$ と表すとき

音源毎の
空間共分散

$$\mathbb{E} \left[\mathbf{x}_{ij}^{(n)} \mathbf{x}_{ij}^{(n)\text{H}} \right] = r_{ij,n} \mathbf{R}_i^{(s_n)}$$

Source image
マイクロホンへの伝達系
に寄与する時不変な成分
(空間相関行列)

時変な音源の分散(パワースペクトログラム)

観測の
空間共分散

$$\mathbf{R}_{ij}^{(x)} = \mathbb{E} \left[\mathbf{x}_{ij} \mathbf{x}_{ij}^{\text{H}} \right] = \sum_{n=1}^N r_{ij,n} \mathbf{R}_i^{(s_n)}$$

時間周波数で分散共分散が
変動する多変量ガウス分布

観測の
生成モデル

$$p(\mathbf{x}_{ij}; r_{ij}, \mathbf{R}_{ij}^{(x)}) = \frac{1}{\det(\pi \mathbf{R}_{ij}^{(x)})} \exp \left(-\mathbf{x}_{ij}^{\text{H}} \mathbf{R}_{ij}^{(x)-1} \mathbf{x}_{ij} \right)$$

多チャンネル
Wiener filter

$$\hat{\mathbf{x}}_{ij}^{(n)} = r_{ij,n} \mathbf{R}_i^{(s_n)} \mathbf{R}_{ij}^{(x)-1} \mathbf{x}_{ij}$$

➡ 時変分散と音源毎の空間共分散
から音源分離が可能(劣決定も可)

空間相関行列のランク

- 空間相関行列は瞬時空間相関の期待値

音源毎の
空間共分散

$$E \left[\mathbf{x}_{ij}^{(n)} \mathbf{x}_{ij}^{(n)H} \right] = r_{ij,n} \mathbf{R}_i^{(s_n)}$$

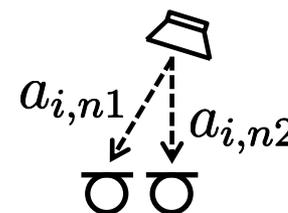
- 「瞬時相関の期待値」のランクが1

➡ 伝達系が時不変な一つの空間基底でモデル化できる

- 時不変な空間基底: ステアリングベクトル $\mathbf{a}_{i,n}$

ランク1
空間モデル

$$\mathbf{R}_i^{(s_n)} = \mathbf{a}_{i,n} \mathbf{a}_{i,n}^H$$



- 「瞬時相関の期待値」のランクが1より大きい(フルランク)

➡ 拡散性音源, 音響放射特性の変動, 過度な残響

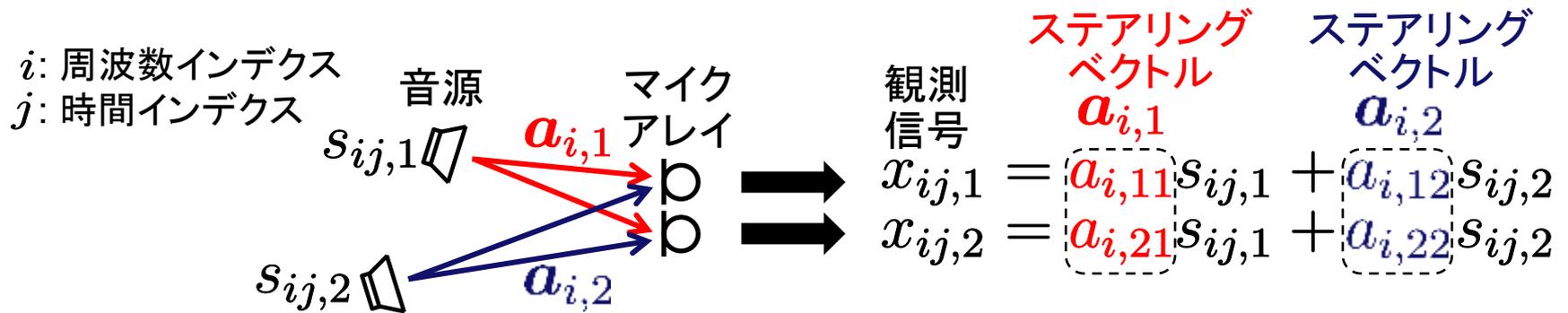
- ステアリングベクトルのような空間基底ではもはや表現不可

- 複数本の空間基底になる(空間基底の数=空間相関行列のランク)

- 周波数領域での瞬時混合仮定が成り立たない

ランク1空間モデル

- 音源毎の空間相関行列のランクが1
 - 時間周波数領域において1つの音源の伝達系が**1本の時不変なステアリングベクトル**で表現可能



- 時間周波数領域における**時不変複素瞬時混合モデル**

時不変混合行列 $\mathbf{A}_i = (\mathbf{a}_{i,1} \ \mathbf{a}_{i,2})$

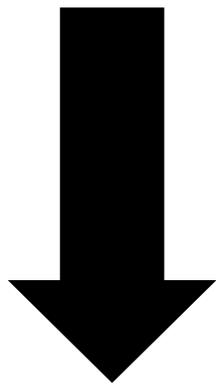
$$\begin{pmatrix} x_{ij,1} \\ x_{ij,2} \end{pmatrix} = \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i,22} \end{pmatrix} \begin{pmatrix} s_{ij,1} \\ s_{ij,2} \end{pmatrix} \quad \Rightarrow \quad \mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij}$$

- 「音源・マイク位置が時不変」かつ「残響時間がフーリエ変換の窓長より短い」場合に成立する仮定 (ICA, IVAでは前提条件)

多チャンネルNMFとILRMAのミッシングリンク

- 多チャンネルNMFの目的関数にランク1空間モデルを導入

$$\mathcal{J} = \sum_{i,j} \left[\text{tr} \left(X_{ij} \left(\sum_k \left(\sum_m \mathbf{R}_i^{(s_m)} z_{mk} \right) t_{ik} v_{kj} \right)^{-1} \right) + \log \det \sum_k \left(\sum_m \mathbf{R}_i^{(s_m)} z_{mk} \right) t_{ik} v_{kj} \right]$$



1. ランク1チャンネル間相関を導入 ($\mathbf{R}_i^{(s_m)} = \mathbf{a}_{i,m} \mathbf{a}_{i,m}^H$)
2. 混合行列 \mathbf{A}_i を用いて表現しなおす
3. 分離行列 $\mathbf{W}_i = \mathbf{A}_i^{-1}$ と分離信号 $\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij}$ に変数変換
ここで逆行列ドメインでの最適化にしたのがポイント!

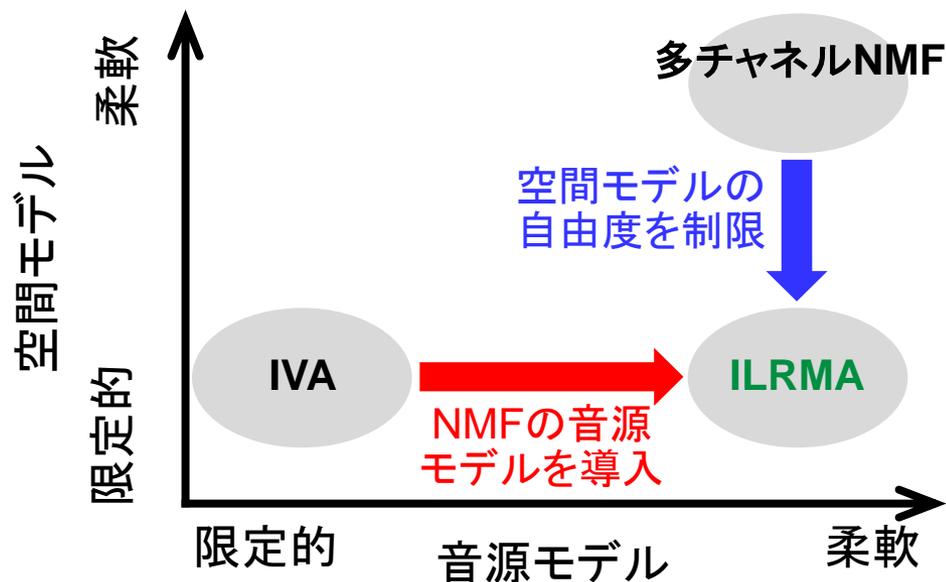
ILRMAのコスト関数

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

- ランク1空間モデル制約付き多チャンネルNMFとILRMAは双対であるが、ILRMAは「分離モデル型」であり高速な最適化が可能 98

IVAと多チャンネルNMFの両理論をつなぐILRMA

- 実は独立に誕生したIVAと多チャンネルNMFは深く関連
 - IVAの音源モデルの基底数拡張
 - 音源モデルの推定にNMFによる分解表現を導入
 - 多チャンネルNMFの空間モデルの自由度の制約
 - フルランクで推定される音源毎の空間相関にランク1制約を導入
 - 「独立性に基づく高速な空間モデルの最適化」と「NMFに基づく柔軟な音源モデル」の両立を実現



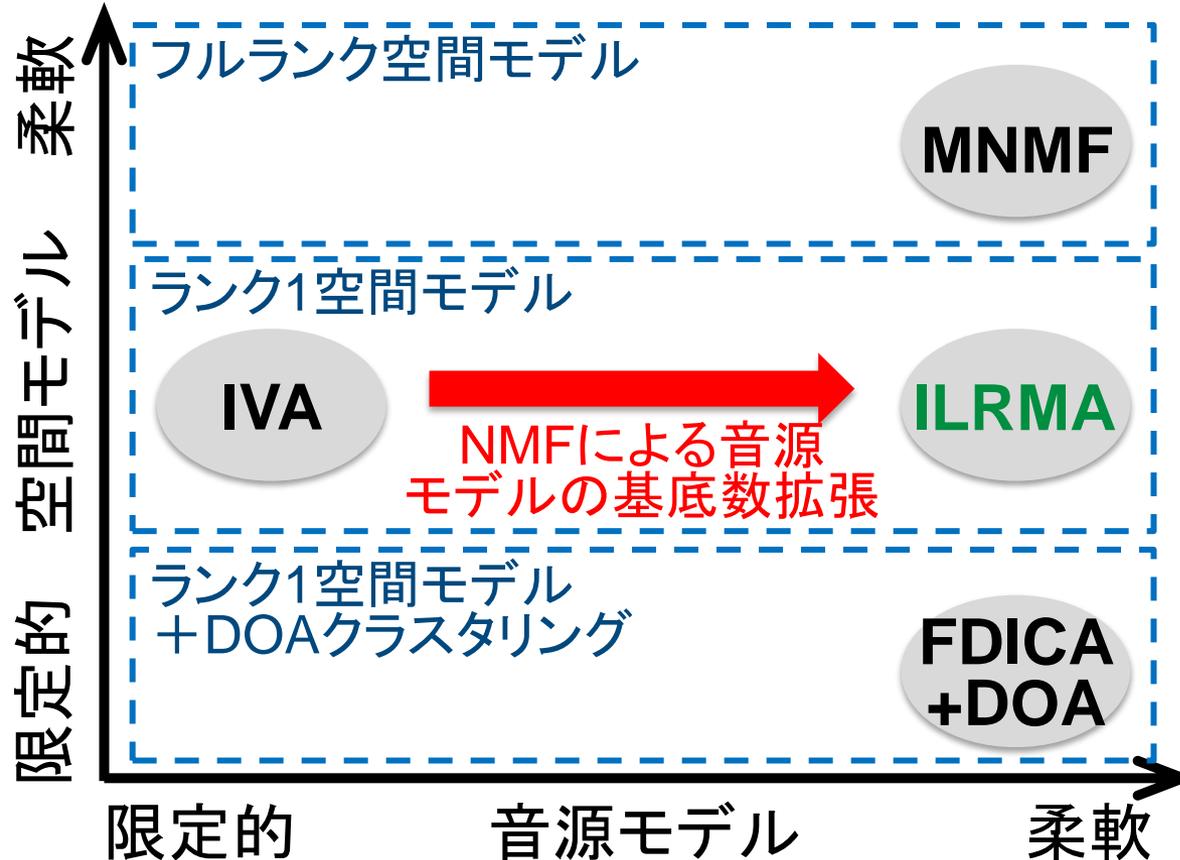
多チャネルNMFや 他の独立因子分析法 との比較

他の独立因子分析法との実証比較実験

手法	空間モデル	音源モデル	推定対象
FDICA	ランク1空間モデル 点DOAを中心に広がる混合分布という制約	なし (時系列の非ガウス性制約のみ)	周波数毎に独立な分離行列
IVA	ランク1空間モデル	基底1本での表現 球対称多変量ラプラス分布	分離行列 音源毎の周波数共通分散
多チャンネルNMF	フルランク空間モデル (周波数領域の瞬時混合仮定は不要)	任意基底数のNMF 時間周波数分散変動型複素ガウス分布	音源毎の空間相関行列 全音源のNMF表現
ILRMA	ランク1空間モデル	任意基底数のNMF 時間周波数分散変動型複素ガウス分布	分離行列 全音源のNMF表現

実験的な実証方法

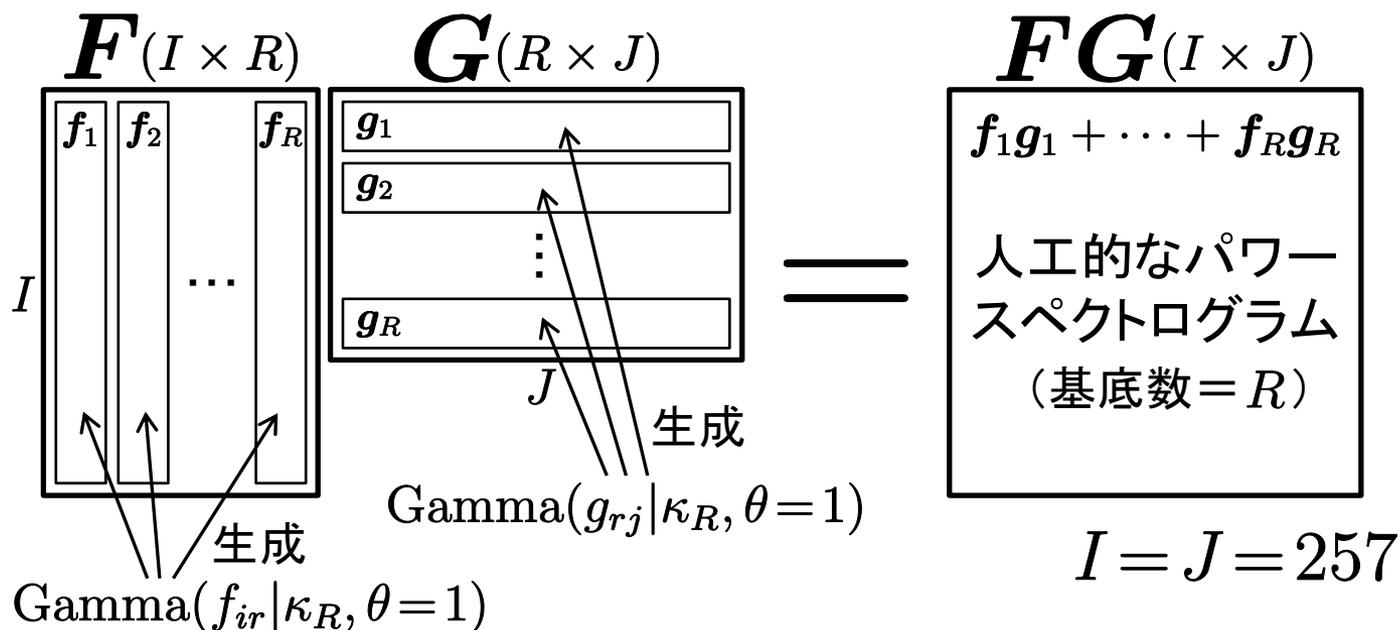
- 音源スペクトログラムの基底数の違い
 - 音源の混合系の違い
- 分離精度に与える影響
- 実測信号ではその他の要因が多すぎる
 - 完全に人工的な音源と混合系を用いて実証する



実験条件

- 実験に用いる信号

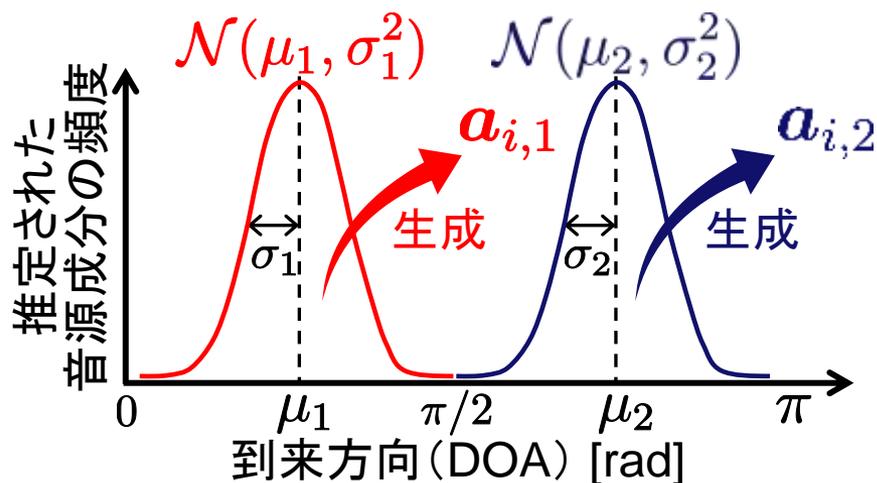
- ガンマ分布に従う変数の積とその線形結合に基づくパワースペクトログラムを2音源分生成
- いかなる基底数 R においても FG が等カートシス(kurt = 50)となるように形状母数 κ_R を設定(導出は省略)
 - カートシスの違いによる音源分離性能の変化を除外するため
- $0 \sim 2\pi$ の一様分布に従う位相を付与



実験条件

• 実験に用いる混合系

- DOA上で2音源分のガウス分布を定義
- 各ガウス分布からステアリングベクトルを生成
- 先の人工音源に対して人工ステアリングベクトルを乗算



人工混合行列

$$\mathbf{A}_i = (\mathbf{a}_{i,1} \ \mathbf{a}_{i,2})$$

人工観測信号

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

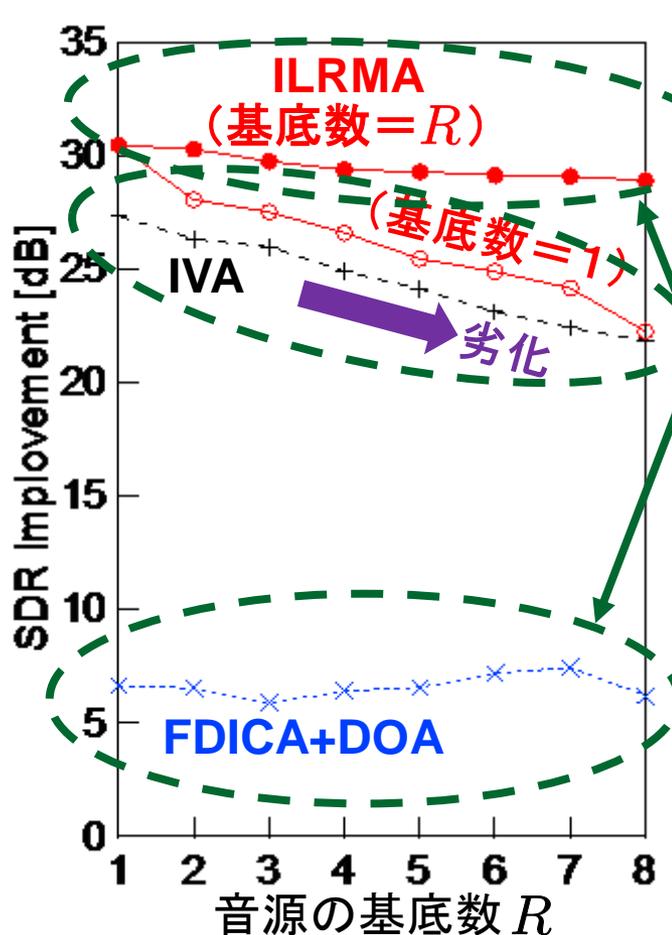
• その他の実験条件

最適化アルゴリズム	FDICA: 自然勾配法, DOAクラスタリング: k-means法, IVA: 補助関数法, ILRMA: 補助関数法
反復回数	200回 (FDICAの反復ステップサイズは実験的に調整済)
評価値	SDR改善量 (総合分離性能)

音源の基底数の違いによる分離精度比較

音源の基底数 R を変化した場合の分離精度比較

— 混合系は $\mu_1 = 5\pi/12, \mu_2 = 7\pi/12, \sigma_1^2 = \sigma_2^2 = 0.05$ に固定



音源1の結果グラフ

— FDICA+DOAは R によらない

- 具体的な音源モデルが無い

— IVAとILRMA(基底数=1)は音源の基底数の増加に伴って精度が劣化する

- 音源モデルが表現できなくなるため

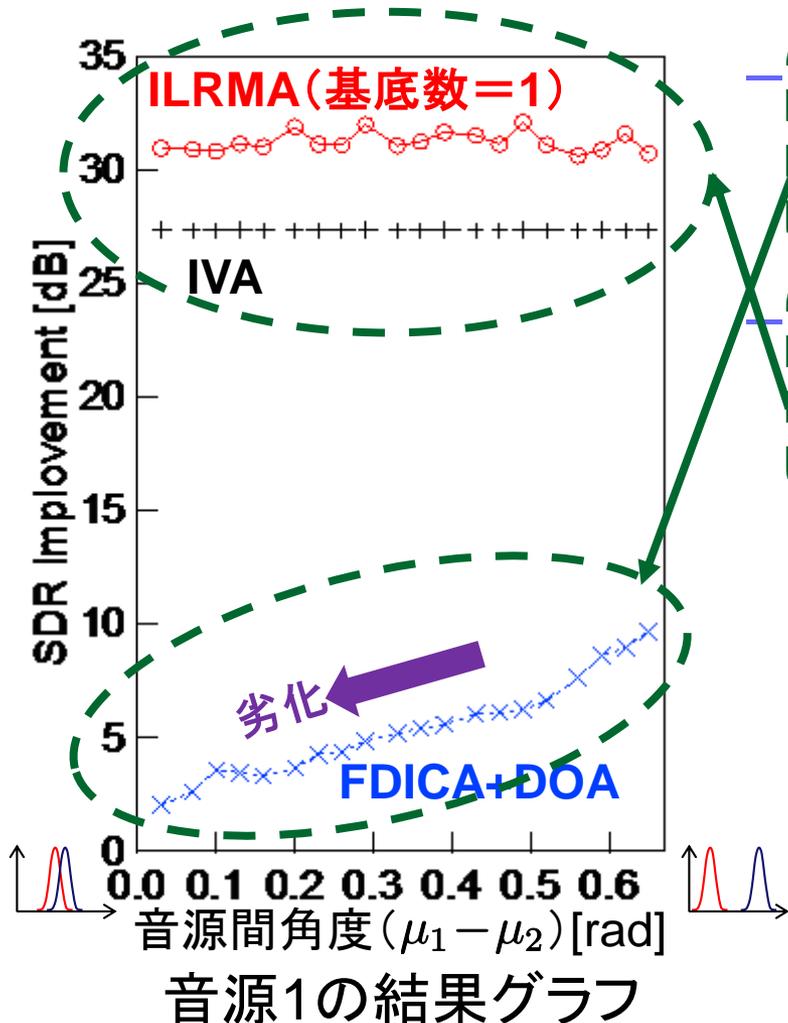
— ILRMA(基底数= R)は精度劣化しない

- 適切で具体的な音源モデルが仮定されているため

ILRMAの持つ音源モデル(NMF表現)が最も柔軟

音源の混合系の違いによる分離精度比較

- 音源の混合系 μ_1, μ_2 を変化した場合の分離精度比較
 - 音源の基底数は $R=1$, 分散は $\sigma_1^2 = \sigma_2^2 = 0.05$ に固定

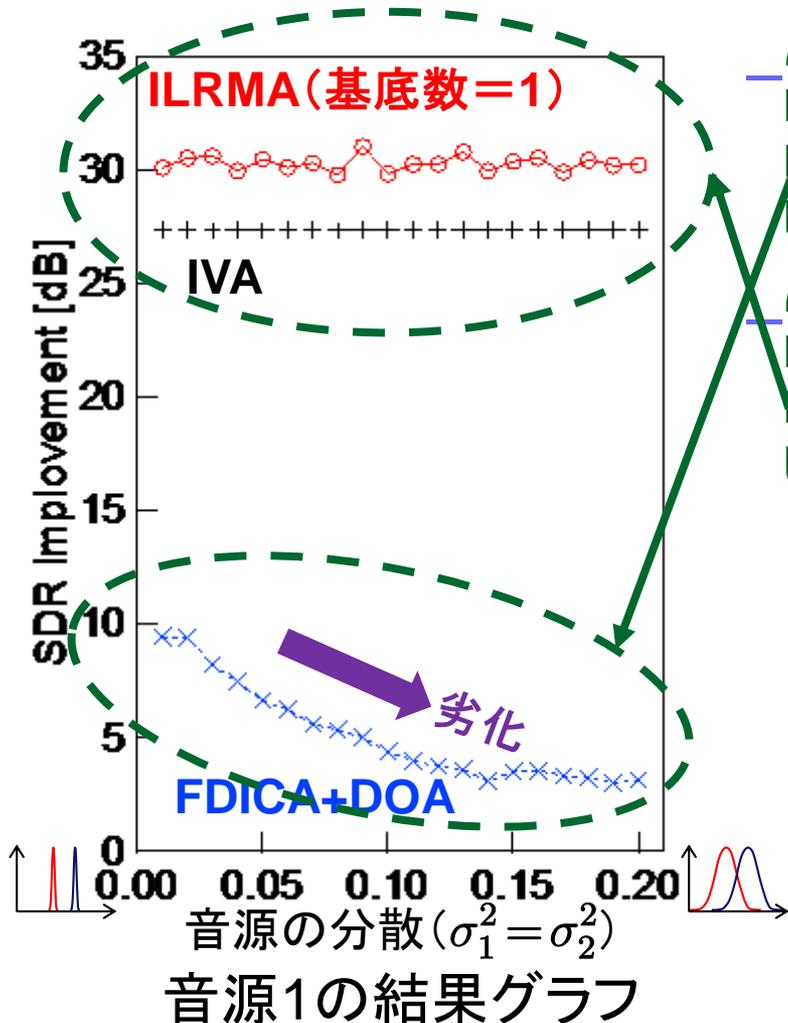


- FDICA+DOAは2音源が接近すると精度が劣化
 - DOAクラスタリングのエラーが増加する
- IVAとILRMA(基底数=1)は音源の混合系によらない
 - 空間モデルによる制約が無いため

IVAとILRMAの無制約な空間モデルが柔軟

音源の混合系の違いによる分離精度比較

- 音源の混合系 σ_1^2, σ_2^2 を変化させた場合の分離精度比較
 - 音源の基底数は $R=1$, 平均は $\mu_1 = 5\pi/12, \mu_2 = 7\pi/12$ に固定



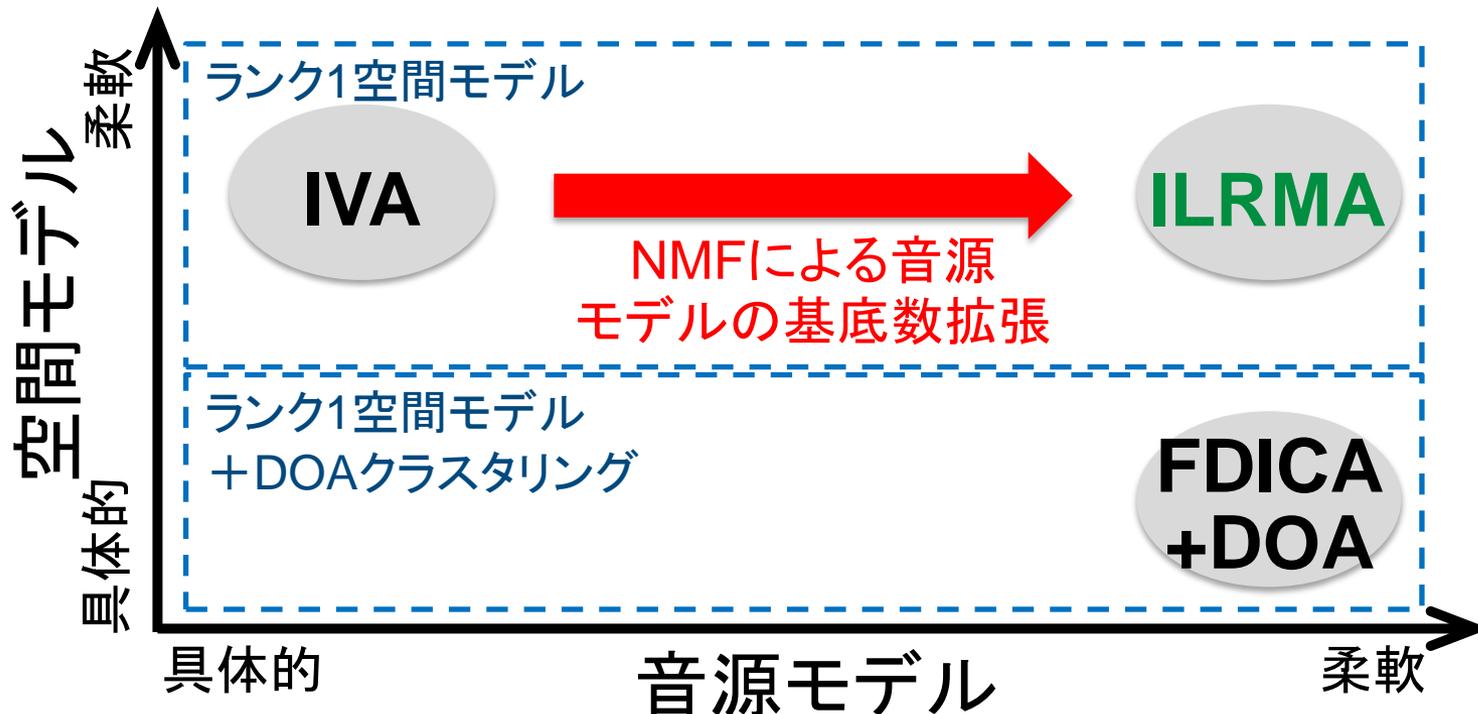
- FDICA+DOAは2音源のDOA幅が増加すると精度が劣化
 - DOAクラスタリングのエラーが増加する
- IVAとILRMAは音源の混合系によらない
 - 空間モデルによる制約が無いため

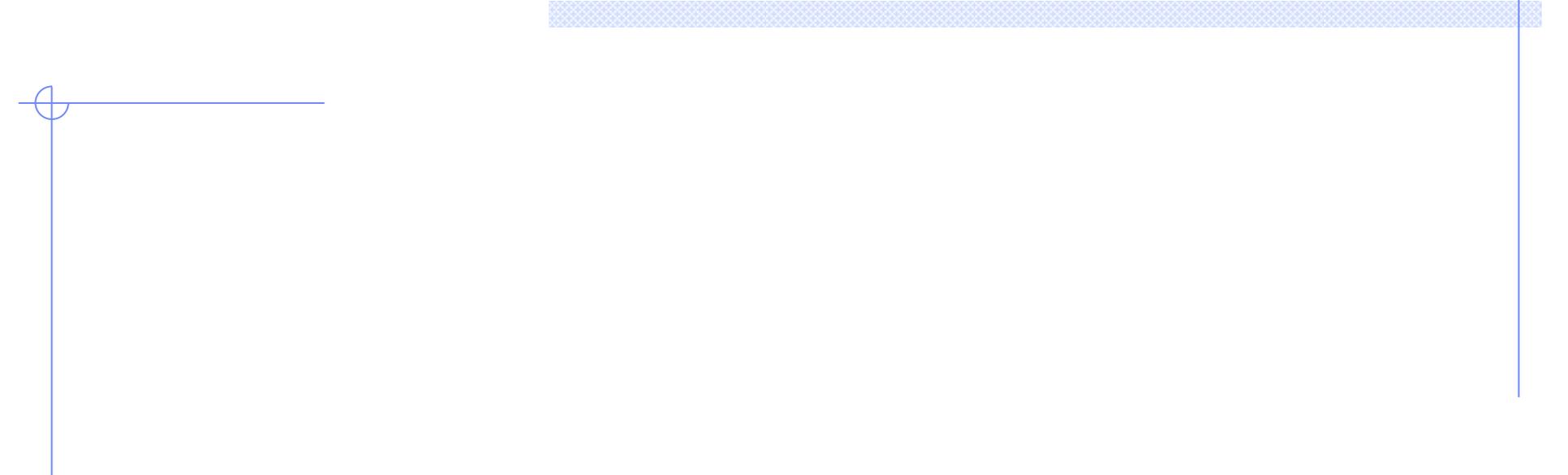
IVAとILRMAの無制約な空間モデルが柔軟

比較実験の結論

• ILRMA

- IVAの音源モデルを任意の基底数に拡張
- 空間モデルへの制限は「ランク1」のみ
- モデル複雑化に伴う最適化の不安定さもさほど問題にならない
- ランク1空間近似を用いる優決定条件BSSとして優秀な手法





ILRMAの実験的評価

従来手法とILRMAの比較による性能評価

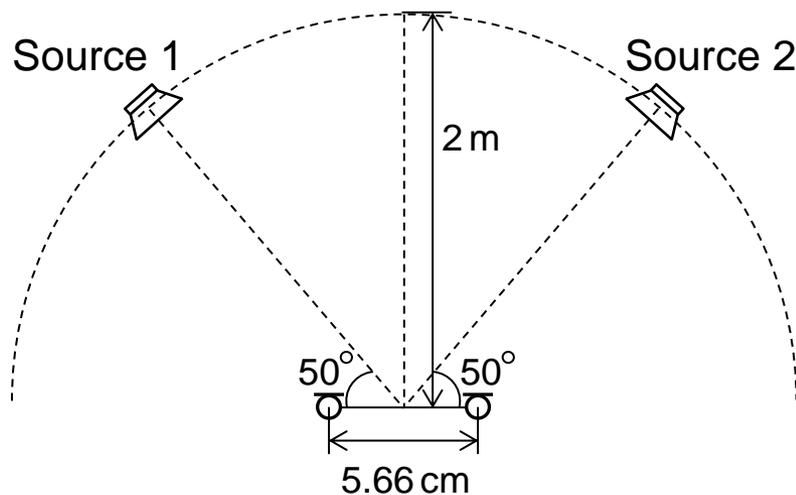
- ILRMAへの期待
 - NMF音源モデルの導入による性能向上(IVAと比較して)
 - ランク1空間モデルの導入による安定性の向上(多チャンネルNMFと比較して)
- ランク1空間モデルが成立する条件での実験
 - マイク位置時不変混合系(多チャンネルNMFを含む全手法での必須条件)
 - 残響時間が窓長より短い(ランク1空間モデル)
 - 特異な音響放射特性などがない(ランク1空間モデル)
- インパルス応答の畳み込みによるシミュレーション混合
 - 残響時間が短ければランク1空間モデルが完全に成立
- 実際のライブ録音による混合観測
 - より現実的な条件での実験

音楽音源分離実験の条件

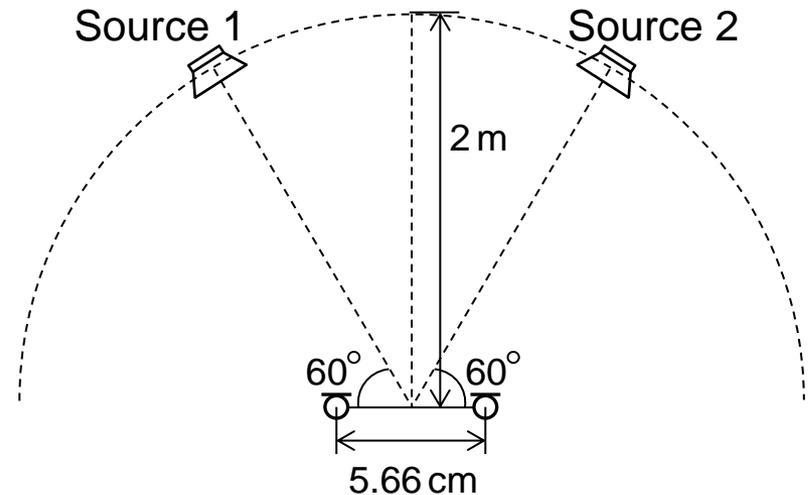
実験条件

音源信号	SiSECのプロ音楽信号に, RWCP収録のマイクアレイインパルス応答で畳み込んで作成, 2チャンネルで2音源の混合信号
窓長(FFT長)	512 ms, ハニング窓
シフト長	128 ms (1/4シフト)
基底数	1音源につき30本(ILRMA1), 全音源で60本(ILRMA2)
主観評価値	SDR改善値(音質と分離度合いを含む総合的な分離性能)

Impulse response E2A
(reverberation time: 300 ms)

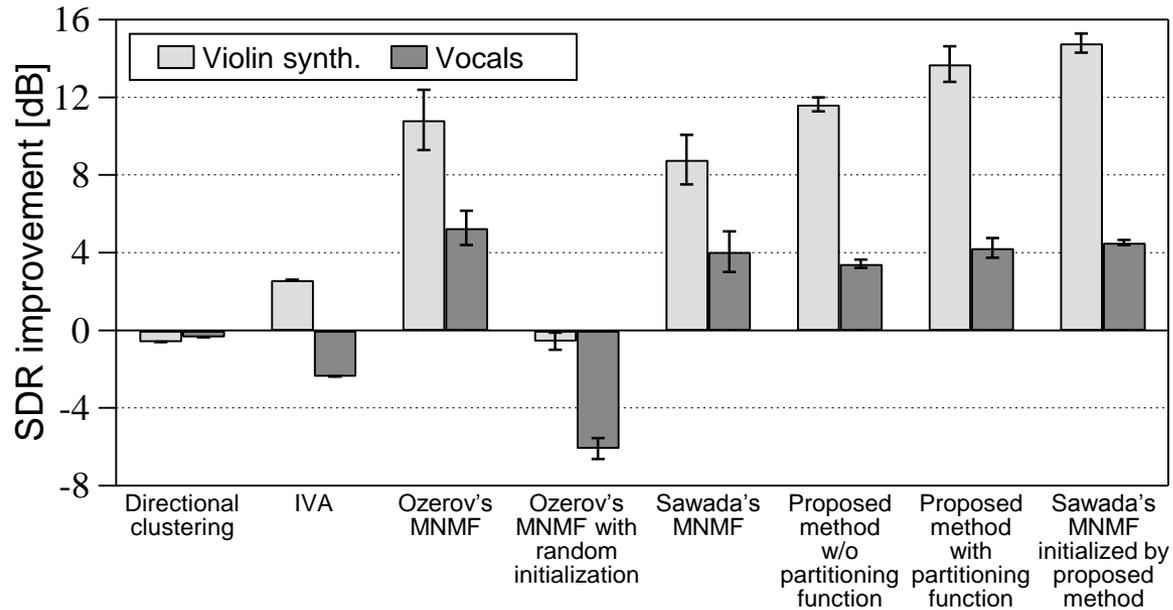


Impulse response JR2
(reverberation time: 470 ms)

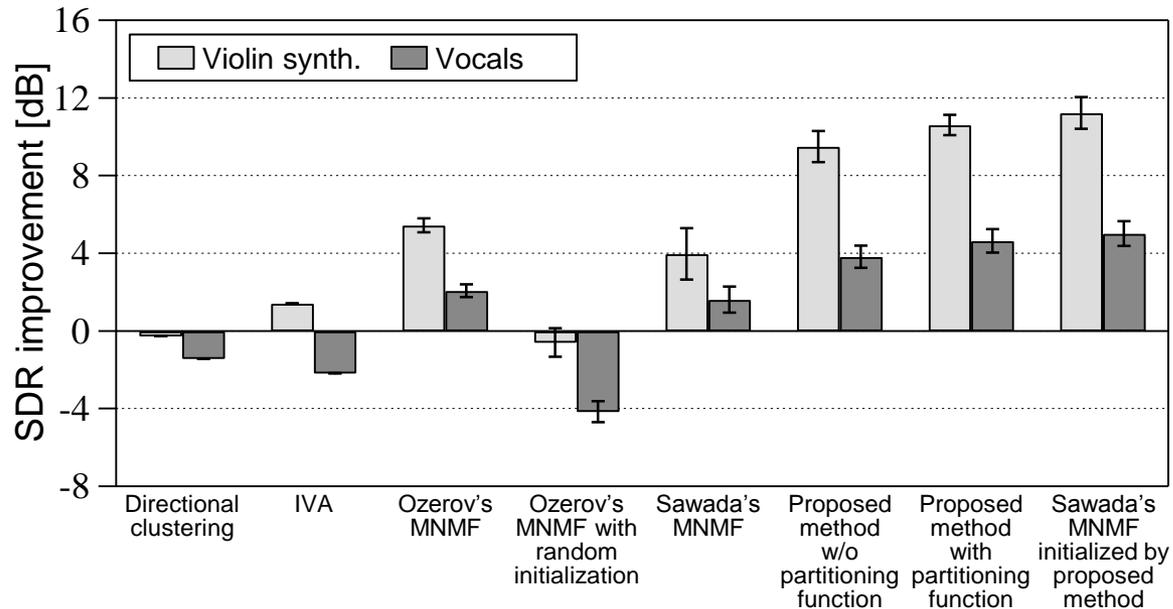


実験結果: fort_minor-remember_the_name

E2A
(300ms)

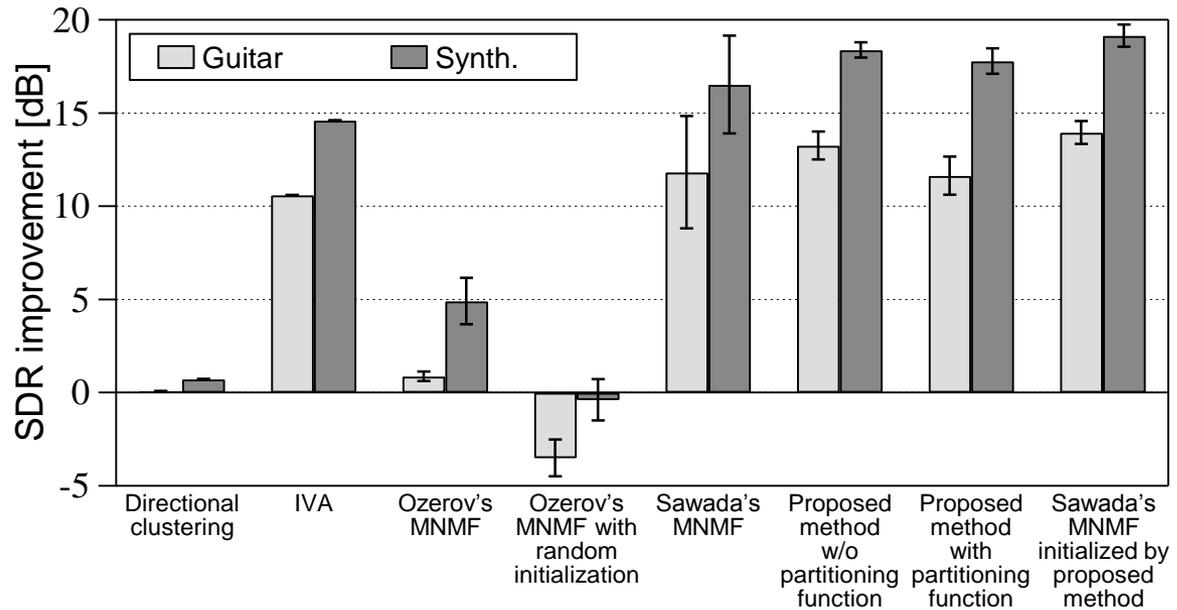


JR2
(470ms)

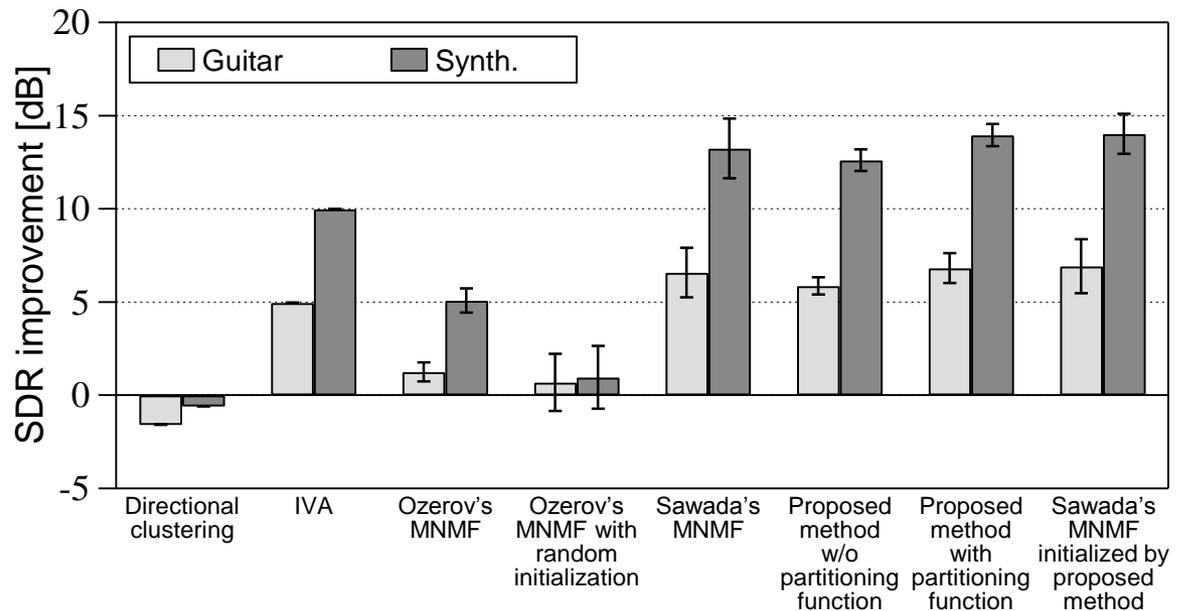


実験結果: ultimate_nz_tour

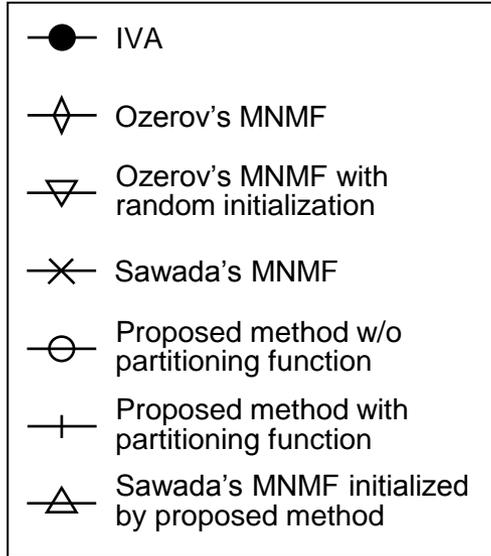
E2A
(300ms)



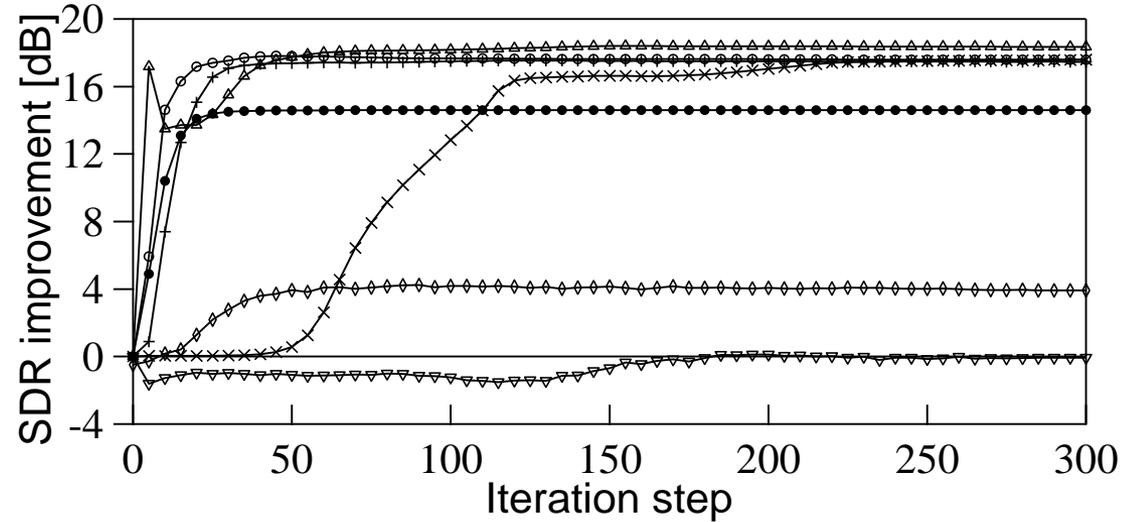
JR2
(470ms)



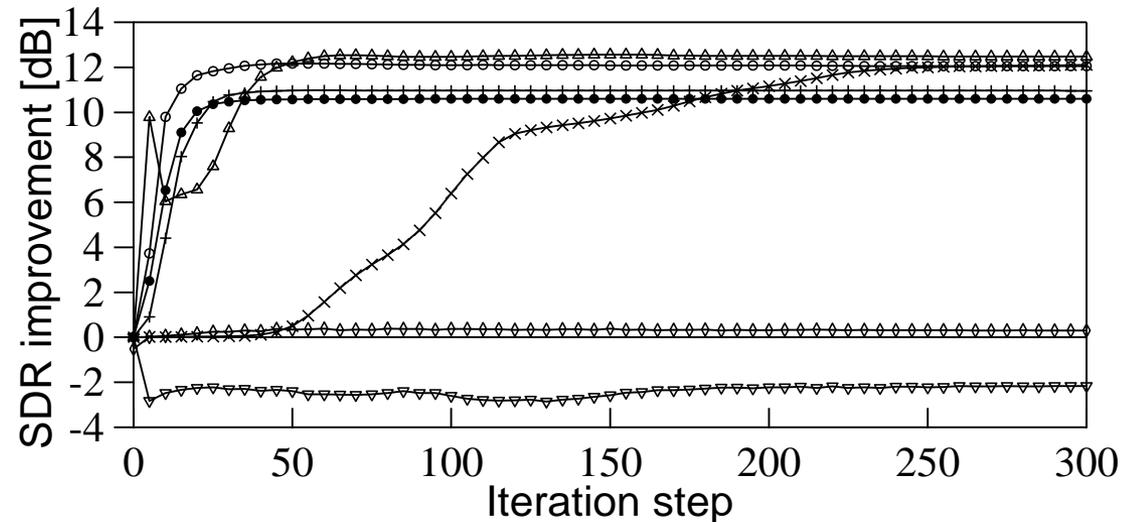
実験結果：ultimate_nz_tour(収束の様子)



Guitar



Synth.



高速, 頑健, 高精度
な音源分離を達成

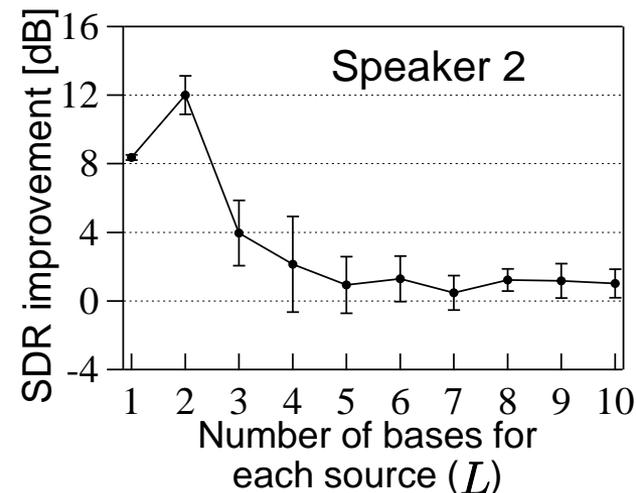
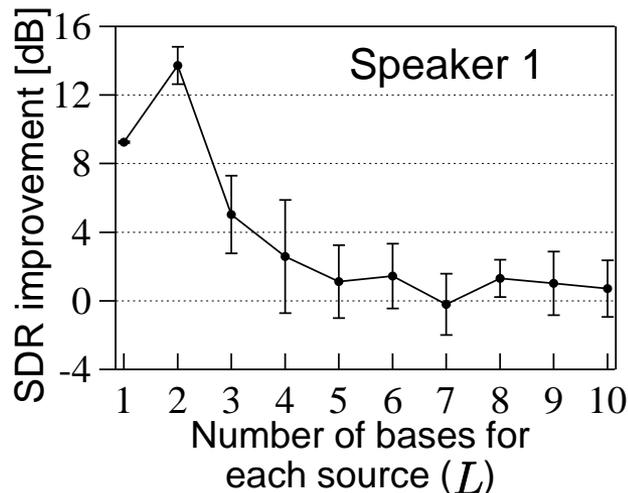
音声音源分離実験の条件

実験条件

音源信号	SiSECのライブ録音音声信号, 2チャンネルで2話者の混合信号
窓長(FFT長)	256 ms, ハニング窓
シフト長	128 ms (1/4シフト)
基底数	1音源につき2本(ILRMA1), 全音源で4本(ILRMA2)
主観評価値	SDR改善値(音質と分離度合いを含む総合的な分離性能)

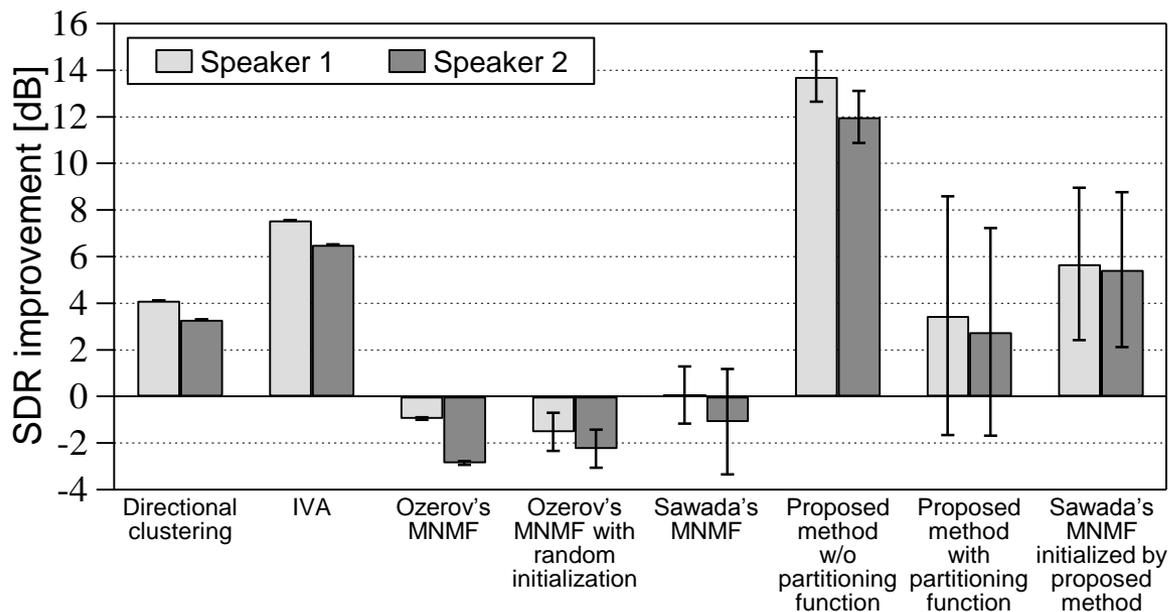
— 予備実験より, 音声信号に対しては基底数を大きくすると音源分離に失敗する事実を確認

- 音声信号の時間周波数構造がNMF表現に不向き？

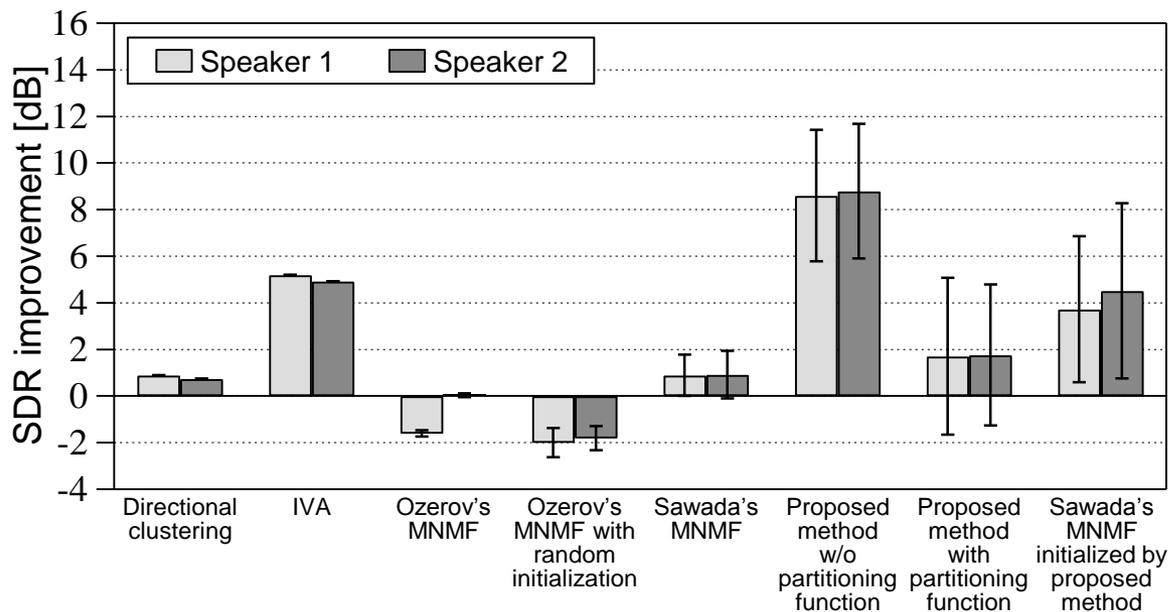


実験結果: female3_liverec_1m

130ms

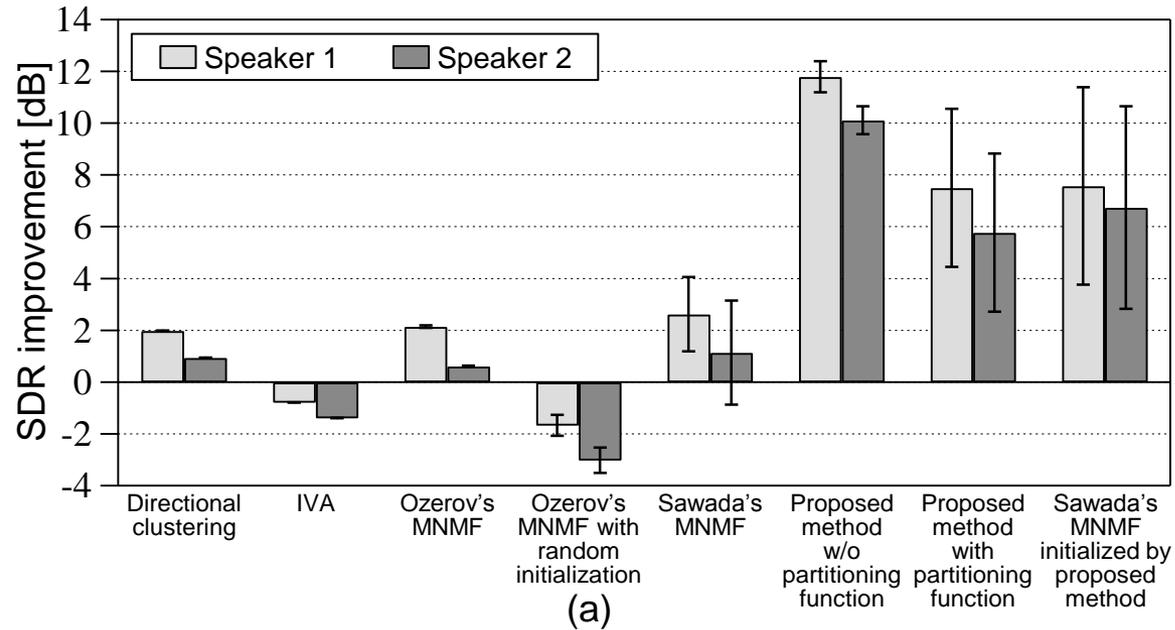


250ms

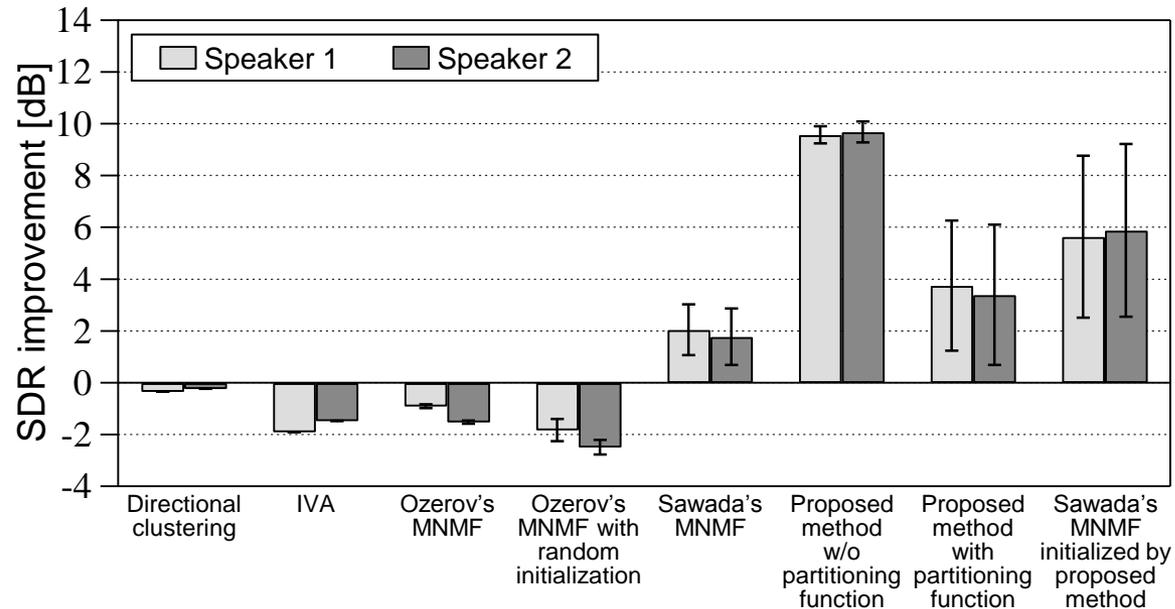


実験結果: male3_liverec_1m

130ms

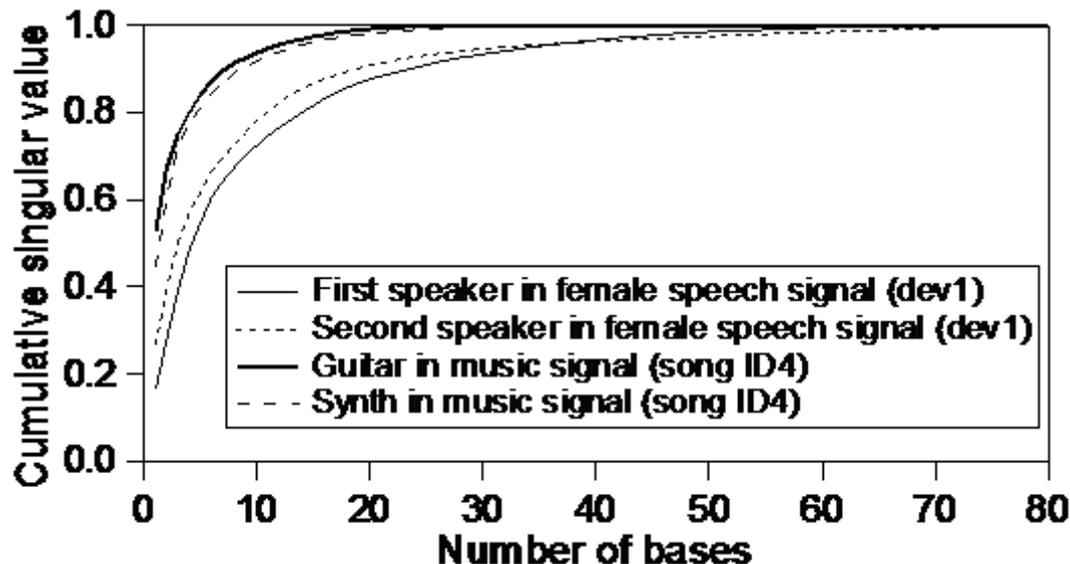


250ms



考察

- ほぼすべての場合で高速, 高精度, 安定な分離を達成
 - 多チャンネルNMFと比較するとモデルの自由度に優位性はない
 - 精度向上はランク1空間モデルの導入による空間モデル変数の最適化が容易になったことに起因
- 音声信号に対しては基底数を増加できない
 - 基底数が増加すると性能が不安定
 - 音声の時間周波数構造は音楽信号ほど低ランクではない



各種音源分離手法の比較(演算量)

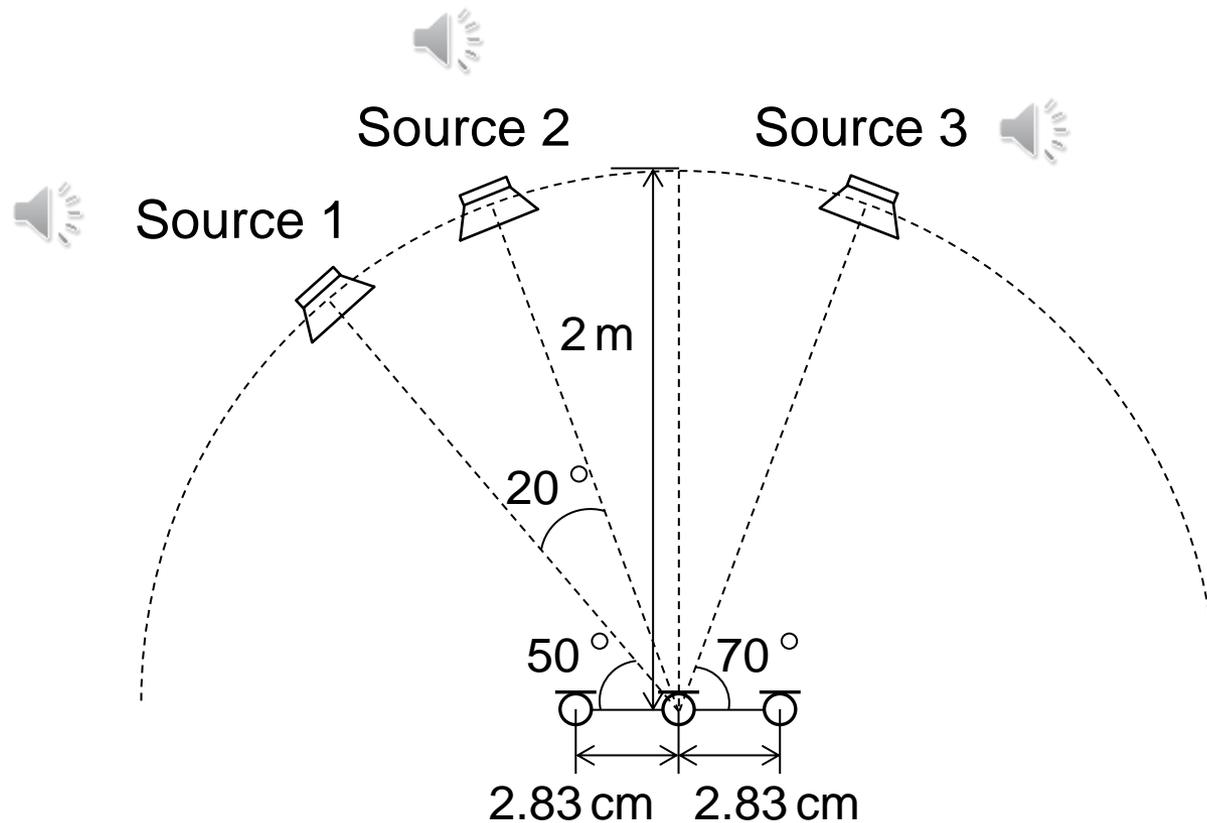
- 実際の音響信号 & 空間混合の分離実験
 - SiSECデータ(実録音・2音源混合)による分離実験結果
 - 演算時間はIVAを基準に正規化した

Algorithm	SDR improv.	Comp. time
Soft masking	-0.1	-
IVA	2.6	1.0
Ozerov's MNMF	1.2	-
Sawada's MNMF	5.0	49.1
ILRMA	8.7	1.3

ILRMAが高い分離精度と低演算時間を実現

ILRMAによるデモ

- ドラム、ストリング、音声からなる複合音の分離



内閣府ImPACT災害対応タフロボット [2016年6月プレスリリース]

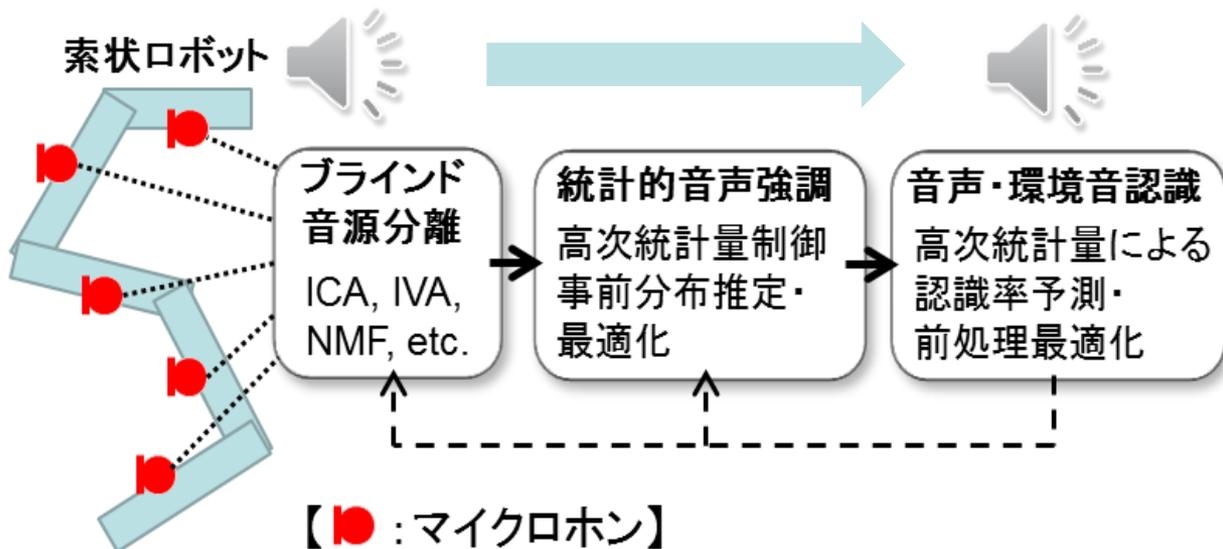
- 災害時の倒壊家屋に入り込んで被災者発見
- 環境音認識による状況把握・救助支援

[Bando, Saruwatari+, J. Robotics & Mechatronics 2017]



被災者はあるのか？

人の声を発見！



いかなる曲がりくねった形状においても位置不定マイク同士が協調して騒音の中から被災者の声を見つけ出す

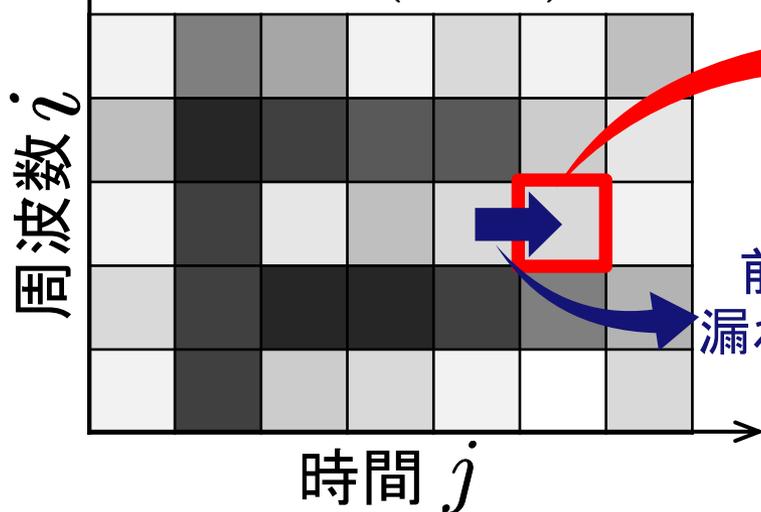
高残響下におけるILRMA の拡張

高残響下における応用手法

- 高残響下では短時間フーリエ変換の窓長よりも長い残響が生じる
 - 残響成分が次の時間フレームに漏れるため複素瞬時混合では表現できなくなる
 - チャンネル間相関がランク1で無くなる \Rightarrow 分離性能が劣化

観測パワースペクトログラム

$X (I \times J)$



音源信号

s_{ij}

観測信号

x_{ij}

$s_{ij,1}$

$s_{ij,2}$

A_i

$x_{ij,1}$

$x_{ij,2}$

前フレームから $n_{ij-1,1}$
漏れ出した残響成分 $n_{ij-1,2}$

$$x_{ij} = A_i s_{ij} + n_{ij-1}$$

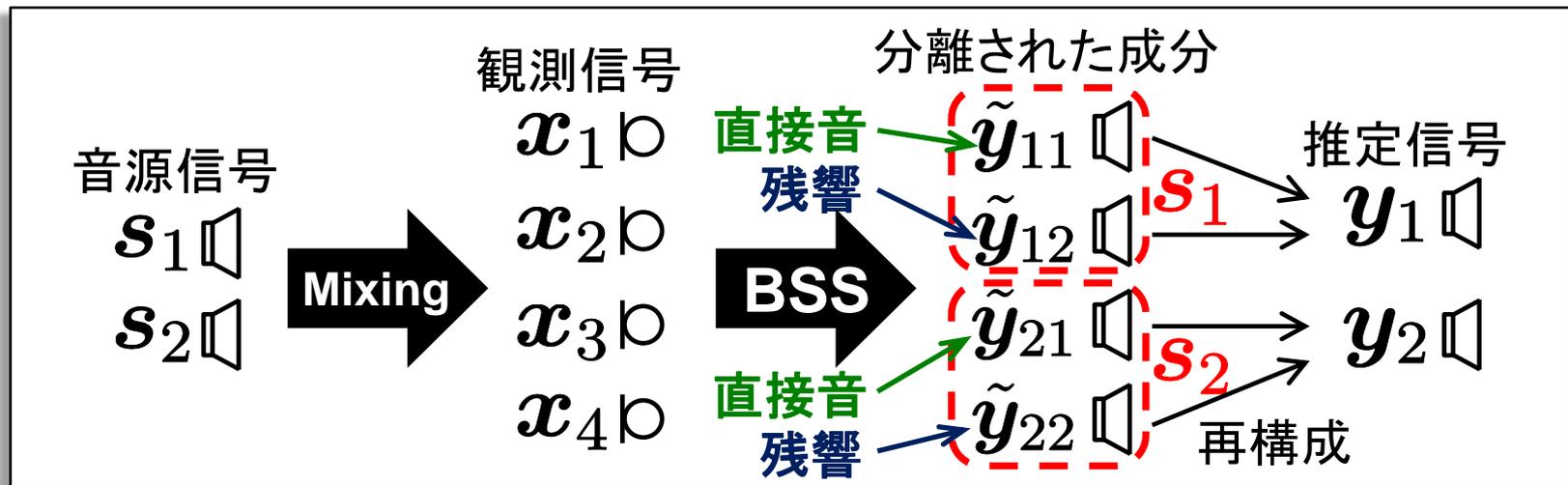
高残響下における応用手法

- 応用として余剰な観測チャンネルを用いることを提案

[Kitamura, 2015]

- 音源数 N の P 倍の観測チャンネルがある状況を仮定
 - 通常のBSSでは事前に主成分分析(PCA)を用いて次元圧縮
 - 提案手法ではPCAを用いずにそのままILRMAで分離
- 各音源の直接音成分及び残響成分を別の独立成分として分離
- 分離後に同じ音源に属する成分同士を足し合わせて復元

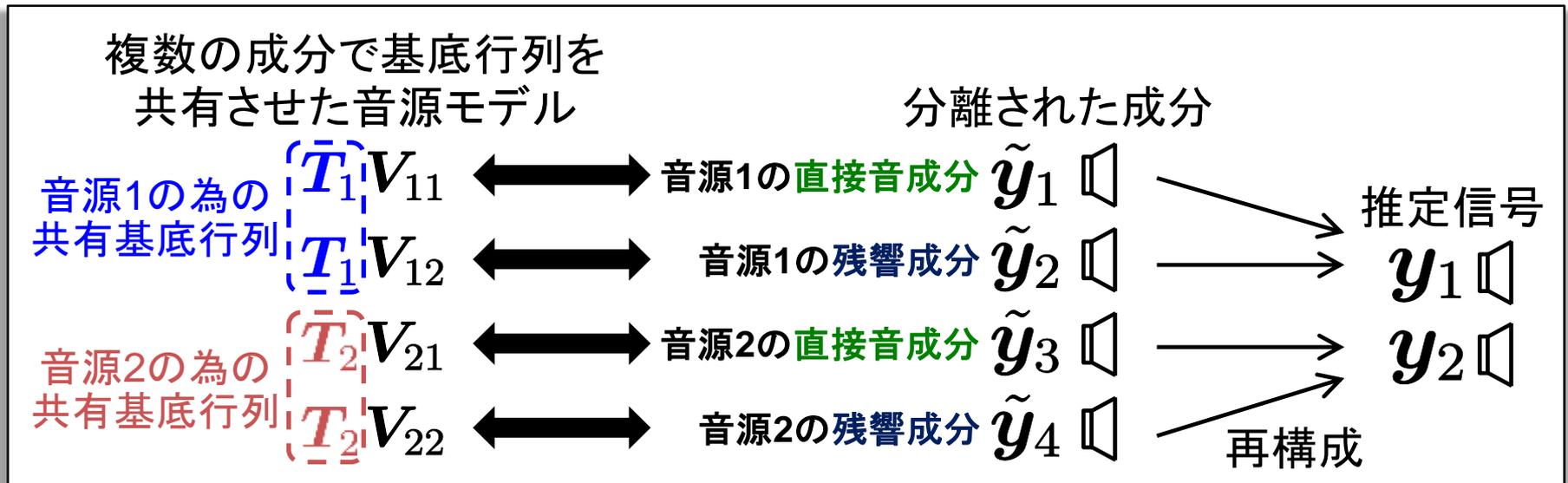
例: 音源数 $N=2$, 観測チャンネル数 $M=4$ ($P=2$)



高残響下における応用手法

- 分離された各成分のクラスタリング問題
 - 各成分が音源毎に自動的にまとまる手法として**基底共有型 ILRMA**を提案 [Kitamura et al., EUSIPCO2015]
 - 複数の分離成分においてNMFによる音源モデルの基底行列 (スペクトルパターン)を共有して推定する

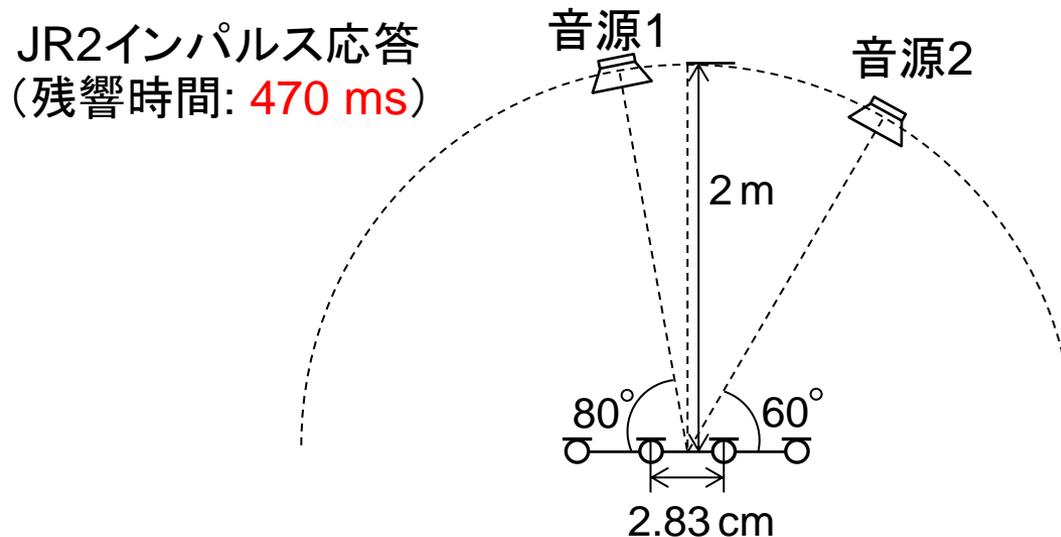
例: 音源数 $N=2$, 観測チャンネル数 $M=4$ ($P=2$)



高残響下における音源分離実験

実験条件

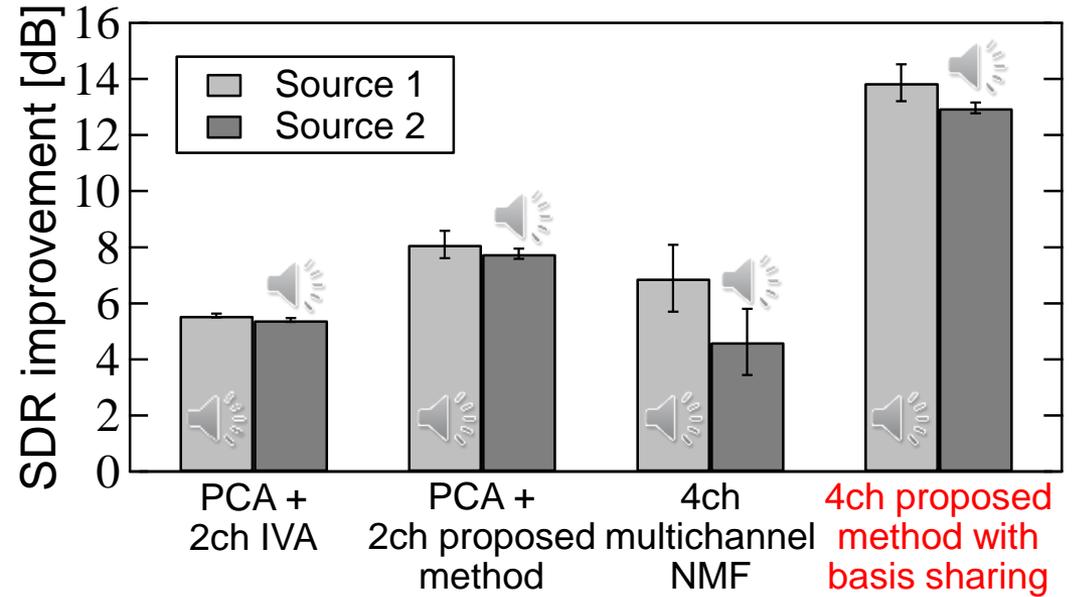
音源信号	SiSECのプロ音楽信号に, RWCP収録のマイクアレーインパルス応答で畳み込んで作成, 2チャンネルで2音源の混合信号
比較手法	PCAを適用してIVA, PCAを適用してILRMA, Sawada's多チャンネルNMF, 基底共有型ILRMA
窓長(FFT長)	128 ms, ハニング窓
シフト長	64 ms (1/2シフト)
基底数	1音源につき30本
主観評価値	SDR改善値(音質と分離度合いを含む総合的な分離性能)



高残響下における音源分離実験

● 実験結果 (曲名: ultimate nz tour, guitar and vocal)

- 初期値を変えて10回試行した際の平均と標準偏差を示したグラフ
- 各音源の残響成分を含んだ状態での音源分離を達成



- 実計算時間の比較においても効率的な最適化を保っていることが確認できる

PCA + 2ch IVA	PCA + 2ch proposed method	4ch multichannel NMF	4ch proposed method with basis sharing
53.8 s	67.6 s	8307.1 s	330.97 s

注: 全て200回反復した場合の計算時間

スパースな生成モデルの導入 ～スパース性と低ランク性の関係～

安定分布と Student's t 分布

• Symmetric α -stable ($S\alpha S$) 分布に基づく音源モデル

[A. Liutkus+, 2015], [U. Şimşekli+, 2015], [S. Leglaive+, 2017], [M. Fontaine+, 2017]

- これは複素変数の重ね合わせに関して安定分布

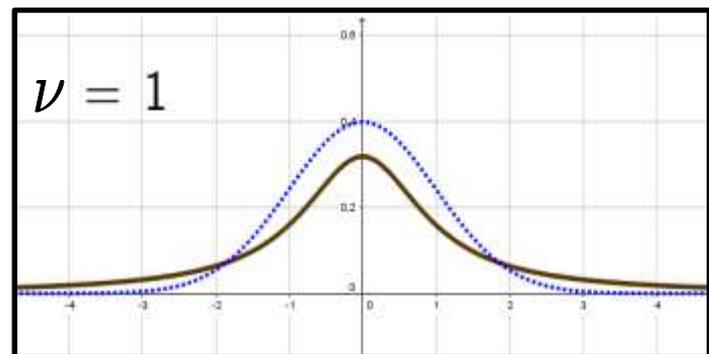
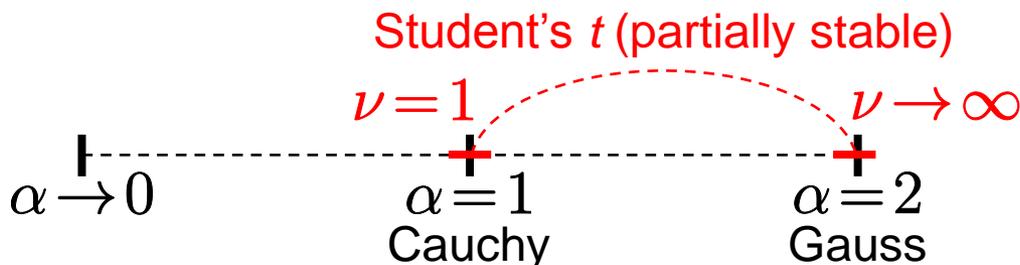
$$\begin{array}{l} z_1 \sim S\alpha S(z; 0, |z_1|^\alpha) \\ z_2 \sim S\alpha S(z; 0, |z_2|^\alpha) \end{array} \iff z_1 + z_2 \sim S\alpha S(z; 0, |z_1|^\alpha + |z_2|^\alpha)$$

- α を小さくしていくと裾の重いスパースな分布になる

• Student's t 分布

[C. Févotte+, 2006], [K. Yoshii+, 2016], [K. Kitamura+, 2016], [S. Leglaive+, 2017]

- これは 2つの $S\alpha S$ であるCauchy分布($\nu=1$)と Gaussian分布を表すことが出来る ($\nu \rightarrow \infty$)



t -ILRMA: 複素Student's t 分布生成モデル型ILRMA

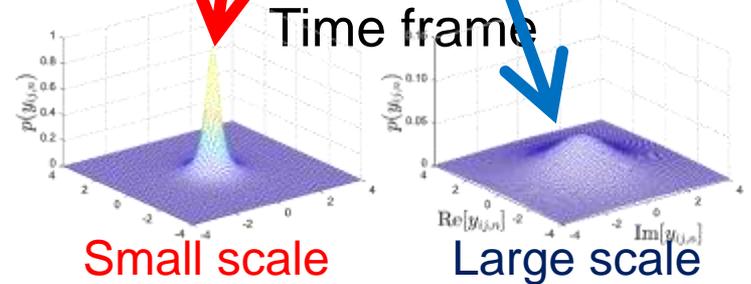
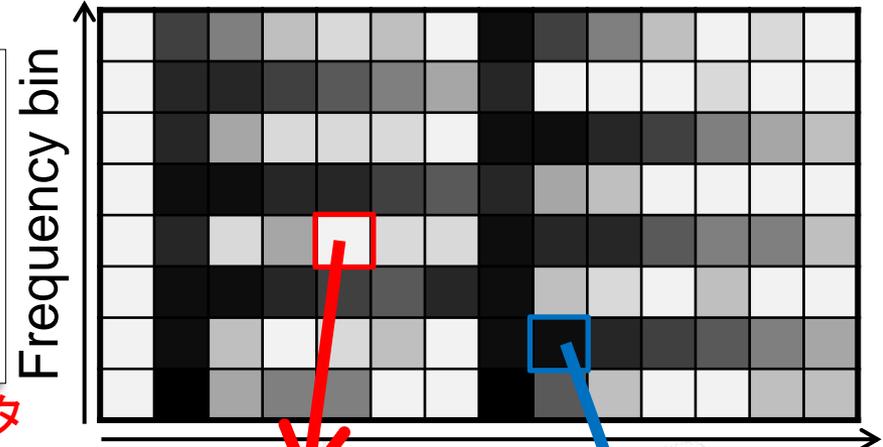
- 時間周波数領域における複素Student's t 分布

n 番目音源のスペクトログラム

$$\begin{aligned}
 p(\mathbf{Y}_n) &= \prod_{i,j} p(y_{ij,n}) \\
 &= \prod_{i,j} \frac{1}{\pi \sigma_{ij,n}^2} \left(1 + \frac{2}{\nu} \frac{|y_{ij,n}|^2}{\sigma_{ij,n}^2} \right)^{-\frac{2+\nu}{2}}
 \end{aligned}$$

時間周波数グリッドにおけるスケールパラメータ
 ↓ この時間周波数スケールが以下の低ランク構造を持つ

$$\sigma_{ij,n}^p = \sum_k t_{ik,n} v_{kj,n} \quad (\Sigma_n^p = \mathbf{T}_n \mathbf{V}_n)$$



- t -ILRMA [Mogami, Saruwatari+, MLSP2017]におけるコスト関数

$$\mathcal{L} = -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2} \right) \log \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{\left(\sum_k t_{ik,n} v_{kj,n} \right)^{\frac{2}{p}}} \right) + \log \left(\sum_k t_{ik,n} v_{kj,n} \right)^{\frac{2}{p}} \right]$$

分離行列の最適化: t -ILRMAにおけるIPの適用

- 分離行列 W の更新にはIPを使いたいのだが...
 - IPは “ $-\log |\det \mathbf{W}_i|$ ” と “ $w_{i,n}$ の二次形式” の和の形式にのみ適用が出来る
 - t -ILRMAにおいてはどうなっているのでしょうか？
- コスト関数の比較
 - 従来の(時変ガウス型)ILRMA (IP適用可能)

$$\mathcal{L} = -2J \sum_i \left\{ \log |\det \mathbf{W}_i| \right\} + \sum_{i,j,n} \left(\frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{\sum_k t_{ik,n} v_{kj,n}} + \log \sum_k t_{ik,n} v_{kj,n} \right)$$

- t -ILRMA (“log”がついているのでそのままではIP適用不可能)

$$\mathcal{L} = -2J \sum_i \left\{ \log |\det \mathbf{W}_i| \right\} + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2} \right) \underline{\log} \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{\left(\sum_k t_{ik,n} v_{kj,n} \right)^{\frac{2}{p}}} \right) + \log \left(\sum_k t_{ik,n} v_{kj,n} \right)^{\frac{2}{p}} \right]$$

対数関数を外すような工夫が必要!

t-ILRMAにおいてどうやってIPを適用するか？

- 補助関数によって上から抑えてやれば良い

— 例えば接線不等式 $\log(z) \leq \frac{1}{\lambda}(z - \lambda) + \log \lambda$

$$\begin{aligned} \mathcal{L} &= -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{(\sum_k t_{ik,n} v_{kj,n})^{\frac{2}{p}}}\right) + \log \left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right] \\ &\leq -2J \sum_i \left[\log |\det \mathbf{W}_i| \right] + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \frac{1}{\alpha_{ij,n}} \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{(\sum_k t_{ik,n} v_{kj,n})^{\frac{2}{p}}} - \alpha_{ij,n}\right) \right. \\ &\quad \left. + \left(1 + \frac{\nu}{2} \log \alpha_{ij,n}\right) + \log \left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right] \end{aligned}$$

t-ILRMAにおいてどうやってIPを適用するか？

- 補助関数によって上から抑えてやれば良い

— 例えば接線不等式 $\log(z) \leq \frac{1}{\lambda}(z - \lambda) + \log \lambda$

$$\begin{aligned} \mathcal{L} &= -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{(\sum_k t_{ik,n} v_{kj,n})^{\frac{2}{p}}}\right) + \log \left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right] \\ &\leq -2J \sum_i \left[\log |\det \mathbf{W}_i| \right] + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \frac{1}{\alpha_{ij,n}} \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{(\sum_k t_{ik,n} v_{kj,n})^{\frac{2}{p}}} - \alpha_{ij,n}\right) \right. \\ &\quad \left. + \left(1 + \frac{\nu}{2} \log \alpha_{ij,n}\right) + \log \left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right] \end{aligned}$$

補助関数の最小点を求めるためにIPを適用可能

- IPに基づく \mathbf{W}_i の最適化

$$\begin{aligned} \mathbf{U}_{i,n} &= \frac{1}{J} \sum_j \frac{\mathbf{x}_{ij} \mathbf{x}_{ij}^H}{\frac{\nu}{\nu+2} (\sum_k t_{ik,n} v_{kj,n})^{\frac{2}{p}} + \frac{\nu}{\nu+2} |y_{ij,n}|^2} \\ \mathbf{w}_{i,n} &\leftarrow (\mathbf{W}_i \mathbf{U}_{i,n})^{-1} \mathbf{e}_n \\ \mathbf{w}_{i,n} &\leftarrow \mathbf{w}_{i,n} (\mathbf{w}_{i,n}^H \mathbf{U}_{i,n} \mathbf{w}_{i,n})^{-\frac{1}{2}} \end{aligned}$$

\mathbf{e}_n : unit vector whose n th element is unity

Cf. MNMF [H. Sawada+, 2013] は一つの空間相関行列を更新するのに J 回の逆行列演算と2回の固有値分解が必要となる。

t-ILRMAにおけるスパース性と低ランク性の関係

- コスト関数における音源モデルパラメータ項

$$\mathcal{L} = -2J \sum_i \log |\det \mathbf{W}_i| + \sum_{i,j,n} \left[\left(1 + \frac{\nu}{2}\right) \underline{\log} \left(1 + \frac{2}{\nu} \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{\left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}}\right) + \log \left(\sum_k t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right]$$

$$\left(\sum_k t_{ik,n} v_{kj,n}\right)^{-\frac{2}{p}} \leq \sum_k \eta_{ij,nk} \left(\frac{t_{ik,n} v_{kj,n}}{\eta_{ij,nk}}\right)^{-\frac{2}{p}} \quad \log(z) \leq \frac{1}{\lambda} (z - \lambda) + \log \lambda$$

更新式

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\sum_j \frac{z_{ij,n}^p}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \right]^{\frac{p}{p+2}}$$

指数に関して一般化された
擬似観測 $z_{ij,n}$ との板倉斎藤
擬距離NMFと等価

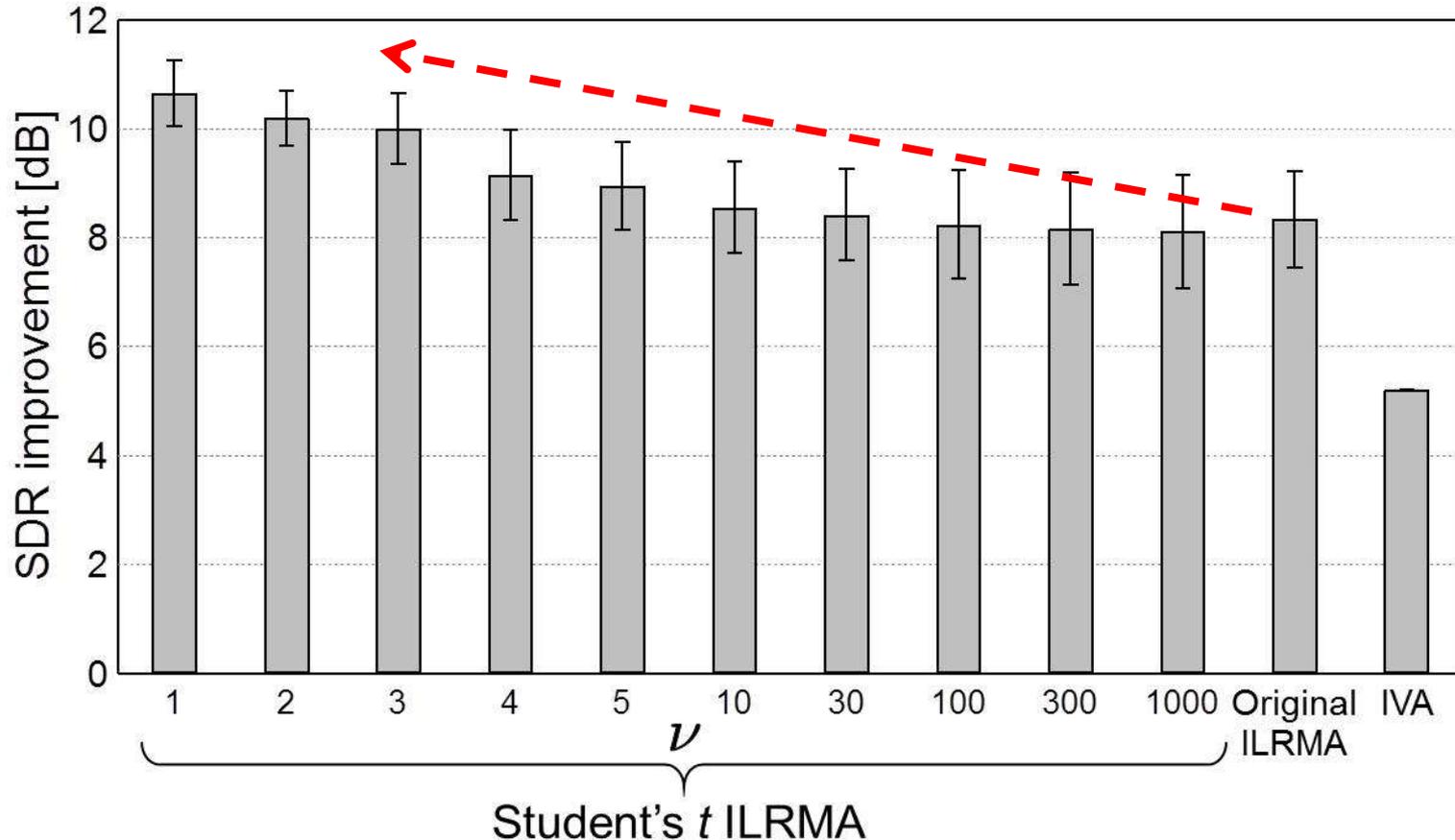
$$z_{ij,n} = \sigma_{ij,n}^{p-2} \left(\frac{\nu}{\nu+2} |y_{ij,n}|^{-2} + \frac{2}{\nu+2} \sigma_{ij,n}^{-2} \right)^{-1}$$

真の観測 y と低ランクモデル σ との $\nu : 2$ 調和平均

$\Rightarrow \nu$ を小さく (スパース音源に) すると低ランク性が強調される!

実験的比較例

- ν パラメータを変えて音声の分離を試みた



ν を小さく(スパース音源に)すると分離精度が向上する

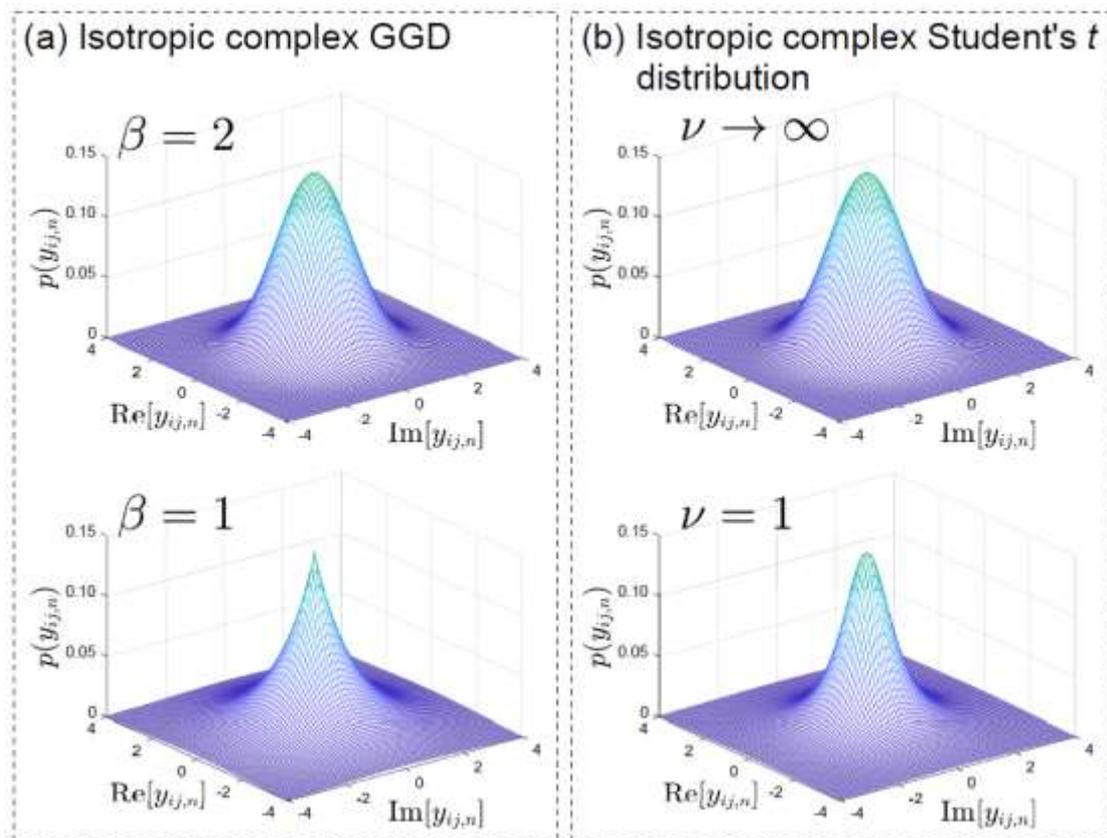
GGD-ILRMA: 複素一般化ガウス分布生成モデルILRMA

- 複素生成モデルを複素一般化ガウス分布 (GGD) に変更

[Kitamura, Saruwatari+, EURASIP JASP2018][Ikeshita+, ICASSP2018]

- GGDにも t 分布と同様にテイルの重さを変える形状母数 β があり、それを変更することによって分布を制御可能

- β を小さくするとスパースな分布になる
- 特に $\beta = 1$ は複素ラプラス分布となりIVAの自然な拡張に帰着
- $\beta \rightarrow 1$ につれて、**低ランク性が幾何平均の意味で強調される** (cf. t -ILRMAは調和平均)



GGD-ILRMA: スパース性と低ランク性の関係

- コスト関数における音源モデルパラメータ項

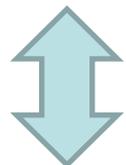
更新式

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\sum_j \frac{z_{ij,n}}{(\sum_{k'} t_{ik',n} v_{k',n})^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik',n} v_{k',n}} v_{kj,n}} \right]^{\frac{p}{\beta+p}}$$

指数に関して一般化された
擬似観測 $z_{ij,n}$ との板倉斎藤
擬距離NMFと等価

$$z_{ij,n} = \frac{\beta}{2} \left(|y_{ij,n}|^{\frac{\beta}{p}} \sigma_{ij,n}^{1-\frac{\beta}{p}} \right)^p$$

真の観測 y と低ランクモデル σ との $\beta/p : (1-\beta/p)$ 比の **幾何平均**
 $\Rightarrow \beta$ を小さく (**スパース音源に**) すると低ランク性が**強調**される！

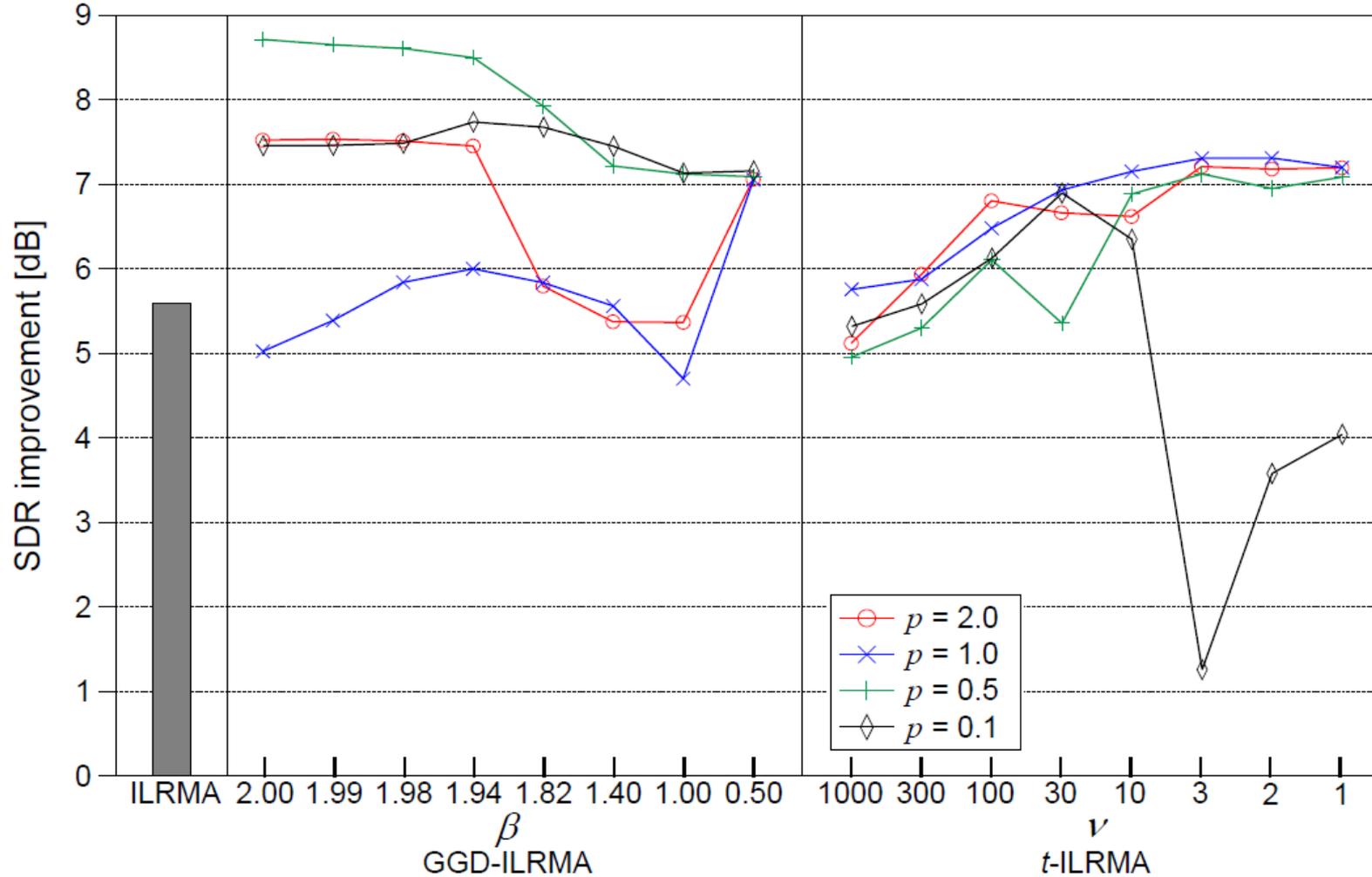


cf. Student's t分布の場合 (**調和平均**)

$$z_{ij,n} = \sigma_{ij,n}^{p-2} \left(\frac{\nu}{\nu+2} |y_{ij,n}|^{-2} + \frac{2}{\nu+2} \sigma_{ij,n}^{-2} \right)^{-1}$$

実験的比較例

- β パラメータを変えて音声の分離を試みた



音源依存ではあるが平均的にはGGD-ILRMA > t-ILRMA ?

事前分布・正則化の導入 ～Vectorwise座標降下法の提案～



■ 動機

- パラメータに事前分布を導入することにより音源分離精度を向上させる
- 例えば「だいたいあの方角からの音を抽出したいんだけど…」という場合
- ILRMAのメインコストに正則化項を加算することによって実現される

■ 低ランク音源パラメータ TV については...

- ISNMF項にスパース正則化を付与する（L1正則化等）だけで実装可能

■ 分離行列 W については...

- ICAの時代は最急降下法だったので、メインコストと正則化項の偏微分を単純に足せばよかった
- **IPを使った補助関数法の場合、偏微分だけではなく、「偏微分 = 0（停留点）」を解析的に解く必要があるのでそれほど簡単ではない**

正則化付き目的関数の変形



ILRMAの目的関数 (2次形式 + logdet)

$$\mathcal{J} = J \sum_i \left[\sum_n \mathbf{w}_{i,n}^H \mathbf{D}_{i,n} \mathbf{w}_{i,n} - \log |\det \mathbf{W}_i|^2 \right] + \mathcal{C}$$

ガウシアン事前分布による正則化 (2次形式 + **1次内積** + logdet)

$$\begin{aligned} \mathcal{J}_P &= J \sum_i \left[\sum_n \mathbf{w}_{i,n}^H \mathbf{D}_{i,n} \mathbf{w}_{i,n} - \log |\det \mathbf{W}_i|^2 \right] + \underbrace{J \sum_{i,n} \lambda_n \|\mathbf{w}_{i,n} - \hat{\mathbf{w}}_{i,n}\|^2}_{\text{Red dashed box}} + \mathcal{C} \\ &= J \sum_i \left[\sum_n \left(\mathbf{w}_{i,n}^H \hat{\mathbf{D}}_{i,n} \mathbf{w}_{i,n} - \lambda_n \hat{\mathbf{w}}_{i,n}^H \mathbf{w}_{i,n} - \lambda_n \mathbf{w}_{i,n}^H \hat{\mathbf{w}}_{i,n} \right) - \log |\det \mathbf{W}_i|^2 \right] + \mathcal{C}' \end{aligned}$$

$\hat{\mathbf{W}}_i = (\hat{\mathbf{w}}_{i,1}, \dots, \hat{\mathbf{w}}_{i,N})^H$: 事前情報から定まる目標分離行列

- 新たな目的関数は $\mathbf{w}_{i,n}$ の**1次項 (内積)** を含む → **IPが利用できない!**
- **1次項が加わった新たな目的関数の反復最適化法**が提案されている

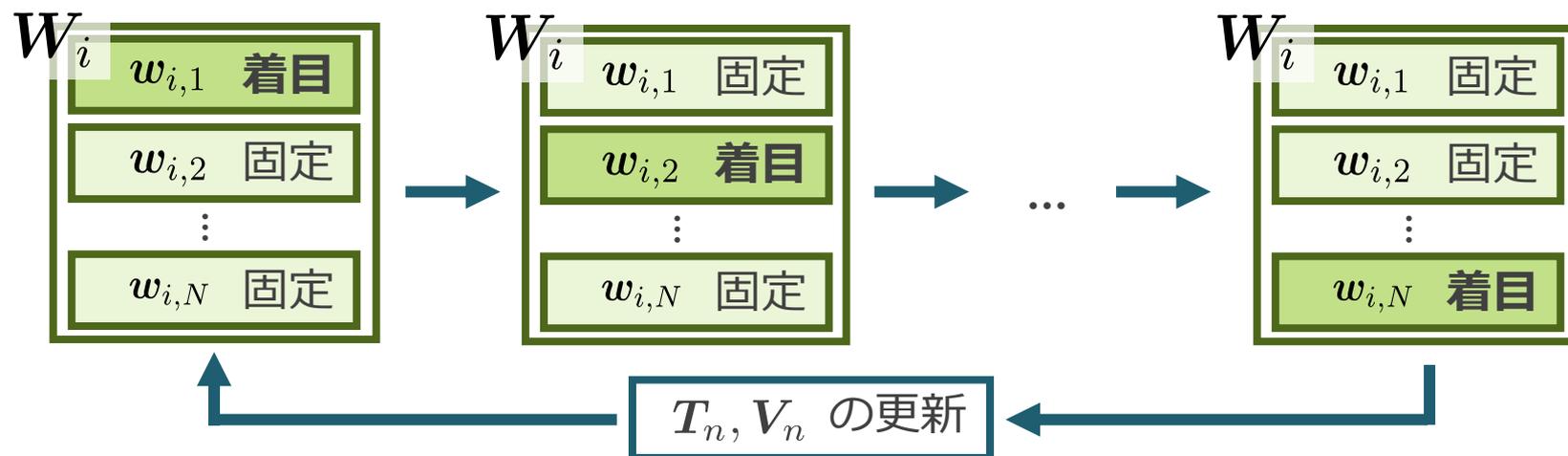
[Mitsui, Saruwatari+, ICASSP2018]

目的関数最適化のアプローチ

正則化付きILRMAの目的関数

$$\mathcal{J}_P = J \sum_i \left[\sum_n \left(\mathbf{w}_{i,n}^H \hat{\mathbf{D}}_{i,n} \mathbf{w}_{i,n} - \lambda_n \hat{\mathbf{w}}_{i,n}^H \mathbf{w}_{i,n} - \lambda_n \mathbf{w}_{i,n}^H \hat{\mathbf{w}}_{i,n} \right) - \log |\det \mathbf{W}_i|^2 \right] + \mathcal{C}'$$

- \mathbf{W}_i に関する最適解を解析的に求めるのは困難 → **反復最適化**が必要
- **ブロック座標降下法**を利用して反復最適化を行う
 - あるベクトル $\mathbf{w}_{i,n}$ のみに着目すれば, 目的関数の停留点が代数的に求まる
 - 行ベクトルごとに更新 → **Vectorwise coordinate descent (VCD)** と呼ぶ



停留点の導出：余因子展開アプローチ

正則化付きILRMAの目的関数

$$\mathcal{J}_P = J \sum_i \left[\sum_n \left(\mathbf{w}_{i,n}^H \hat{\mathbf{D}}_{i,n} \mathbf{w}_{i,n} - \lambda_n \hat{\mathbf{w}}_{i,n}^H \mathbf{w}_{i,n} - \lambda_n \mathbf{w}_{i,n}^H \hat{\mathbf{w}}_{i,n} \right) - \log |\det \mathbf{W}_i|^2 \right] + C'$$

目的関数 \mathcal{J}_P の停留点に $\mathbf{w}_{i,n}$ を更新 … $\log |\det \mathbf{W}_i|$ の偏微分が厄介

- 余因子行列を次のように定義： $\mathbf{B}_i = (\mathbf{b}_{i,1}, \dots, \mathbf{b}_{i,N}) \stackrel{\text{def}}{=} (\det \mathbf{W}_i) \mathbf{W}_i^{-1}$
- logdet項は、次のように変形できる： $\log |\det \mathbf{W}_i|^2 = \log |\mathbf{w}_{i,n}^H \mathbf{b}_{i,n}|^2$
- $\mathbf{b}_{i,n}$ は $\mathbf{w}_{i,n}$ と関係なく定まるため、 $\mathbf{w}_{i,n}^*$ によるlogdetの偏微分は以下：

$$\frac{\partial \log |\mathbf{w}_{i,n}^H \mathbf{b}_{i,n}|^2}{\partial \mathbf{w}_{i,n}^*} = \frac{\mathbf{b}_{i,n}}{\mathbf{w}_{i,n}^H \mathbf{b}_{i,n}}$$

→ $\frac{\partial \mathcal{J}_P}{\partial \mathbf{w}_{i,n}^*} = 0$ の解へと $\mathbf{w}_{i,n}$ を更新すればよい！

「(目的関数の偏微分) = 0」の代数的求解

「(目的関数の偏微分) = 0」は、次の通り

$$\hat{D}_{i,n} \mathbf{w}_{i,n} - \lambda_n \hat{\mathbf{w}}_{i,n} - \frac{1}{\mathbf{w}_{i,n}^H \mathbf{b}_{i,n}} \mathbf{b}_{i,n} = 0$$

$\beta_{i,n} = 1 / \mathbf{w}_{i,n}^H \mathbf{b}_{i,n}$ において整理すると、以下の方程式を満たす $\beta_{i,n}$ を求める問題へと帰着可能

$$\mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n} |\beta_{i,n}|^2 + \lambda_n \hat{\mathbf{w}}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n} \beta_{i,n} - 1 = 0$$

この方程式の解は、次のように代数的に求まる:

$$\beta_{i,n} = e^{j\phi_{i,n}} / \sqrt{\mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n}} \quad (\lambda_n \hat{\mathbf{w}}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n} = 0)$$

$$\beta_{i,n} = \frac{\lambda_n \mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \hat{\mathbf{w}}_{i,n}}{2 \mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n}} \left[-1 \pm \sqrt{1 + \frac{4 \mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n}}{\lambda_n^2 |\mathbf{b}_{i,n}^H \hat{D}_{i,n}^{-1} \hat{\mathbf{w}}_{i,n}|^2}} \right] \quad (\lambda_n \hat{\mathbf{w}}_{i,n}^H \hat{D}_{i,n}^{-1} \mathbf{b}_{i,n} \neq 0)$$



最終的な反復更新式

$$\mathbf{u}_{i,n} = \hat{\mathbf{D}}_{i,n}^{-1} \mathbf{W}_i^{-1} \mathbf{e}_n$$

$$\hat{\mathbf{u}}_{i,n} = \lambda_n \hat{\mathbf{D}}_{i,n}^{-1} \hat{\mathbf{w}}_{i,n}$$

$$h_{i,n} = \mathbf{u}_{i,n}^H \hat{\mathbf{D}}_{i,n} \mathbf{u}_{i,n}$$

$$\hat{h}_{i,n} = \mathbf{u}_{i,n}^H \hat{\mathbf{D}}_{i,n} \hat{\mathbf{u}}_{i,n}$$

$$\mathbf{w}_{i,n} \leftarrow \begin{cases} \frac{\mathbf{u}_{i,n}}{\sqrt{h_{i,n}}} + \hat{\mathbf{u}}_{i,n} & (\text{if } \hat{h}_{i,n} = 0) \\ \frac{\hat{h}_{i,n}}{2h_{i,n}} \left[-1 + \sqrt{1 + \frac{4h_{i,n}}{|\hat{h}_{i,n}|^2}} \right] \mathbf{u}_{i,n} + \hat{\mathbf{u}}_{i,n} & (\text{otherwise}) \end{cases}$$

ベクトル毎の座標降下法に基づき $\mathbf{w}_{i,n}$ を順々に更新する

音源分離例 : Average SDRi of various BSS methods [dB]

(θ_1, θ_2)	AuxIVA	MNMF	<i>t</i> -MNMF	ILRMA (Method 1)	Regularized ILRMA (Method 3)
$(-40^\circ, +40^\circ)$	3.97	3.84	4.80	7.35	12.62
$(-40^\circ, +20^\circ)$	4.15	3.80	4.46	3.51	11.10

独立深層学習行列分析

Independent Deeply Learned Matrix Analysis
(IDLMA: 発音はアイドルエムエー)

Independent Factor Analysis family

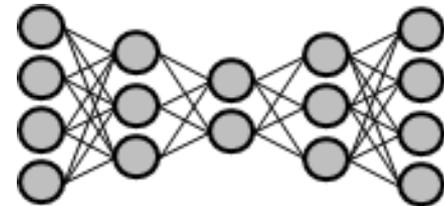
ILRMA

= Independent **Low-Rank** Matrix Analysis



IDLMA

= Independent **Deeply Learned** Matrix Analysis



{ Deeply = “*using DNN*”
 Learned = “*trained in advance*”

ILRMAにおける問題点：音源の低ランク性？

$$\mathcal{J}_{\text{ILRMA}} = \frac{1}{J} \sum_{i,j,n} \left[\log \sum_l t_{il,n} v_{lj,n} + \frac{|w_{i,n}^H x_{ij}|^2}{\sum_l t_{il,n} v_{lj,n}} \right] - \sum_i \log |\det \mathbf{W}_i|^2$$

音源モデル (低ランク性)

空間モデル (音源間が独立)

音源によっては低ランク性が
成り立たない場合がある

音源・マイク位置, 部屋の形状,
残響時間などの膨大な物理要因に依存

ならば！

事前に学習データを用いて音源モデル
の分散を推定する写像を作る

学習データの用意は非現実的
ブラインドに推定

ILRMAにおける問題点：音源の低ランク性？

$$\mathcal{J}_{\text{ILRMA}} = \frac{1}{J} \sum_{i,j,n} \left[\log \sum_l t_{il,n} v_{lj,n} + \frac{|w_{i,n}^H x_{ij}|^2}{\sum_l t_{il,n} v_{lj,n}} \right] - \sum_i \log |\det \mathbf{W}_i|^2$$

音源モデル (低ランク性)

空間モデル (音源間が独立)

- 深層学習 (DNN) による強力なモデリング能力を活用する！
- 今まで培ってきた「教師あり音源分離 (例：教師ありNMF)」の技術を昇華させる形で研究を発展できる。
- 急速に発展するDNN研究を我々ならではの視点で拡張する。

事前に学習データを用いて音源モデルの分散を推定する写像を作る

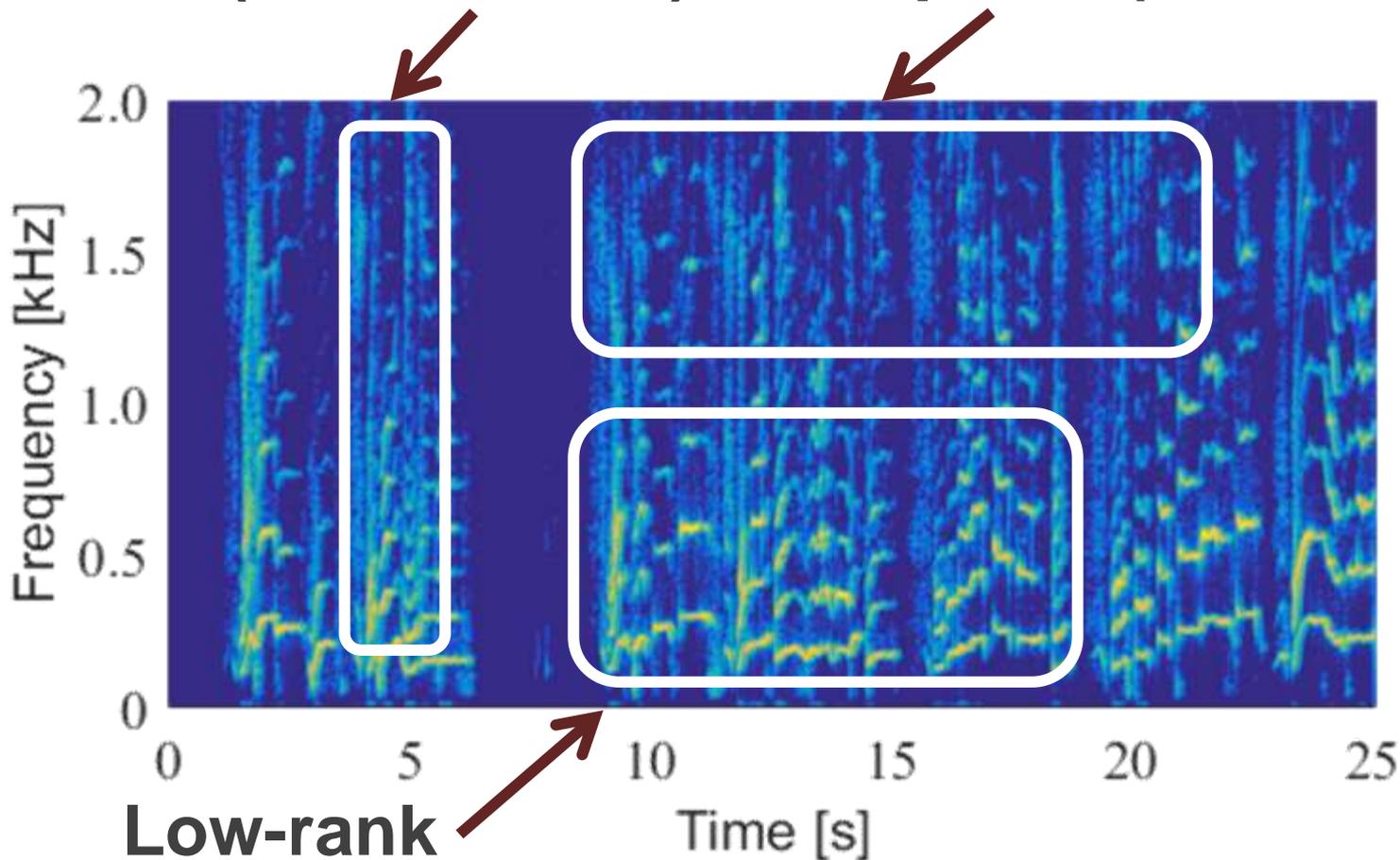
学習データの用意は非現実的
ブラインドに推定

音源の低ランク性？（例：音声信号）



Dense (not low-rank)

Sparse (not low-rank)



このような複雑な構造を持つ信号はDNNでモデリングする

提案手法: DNN音源モデルによる最尤推定

■ 独立深層学習行列分析 (IDLMA)

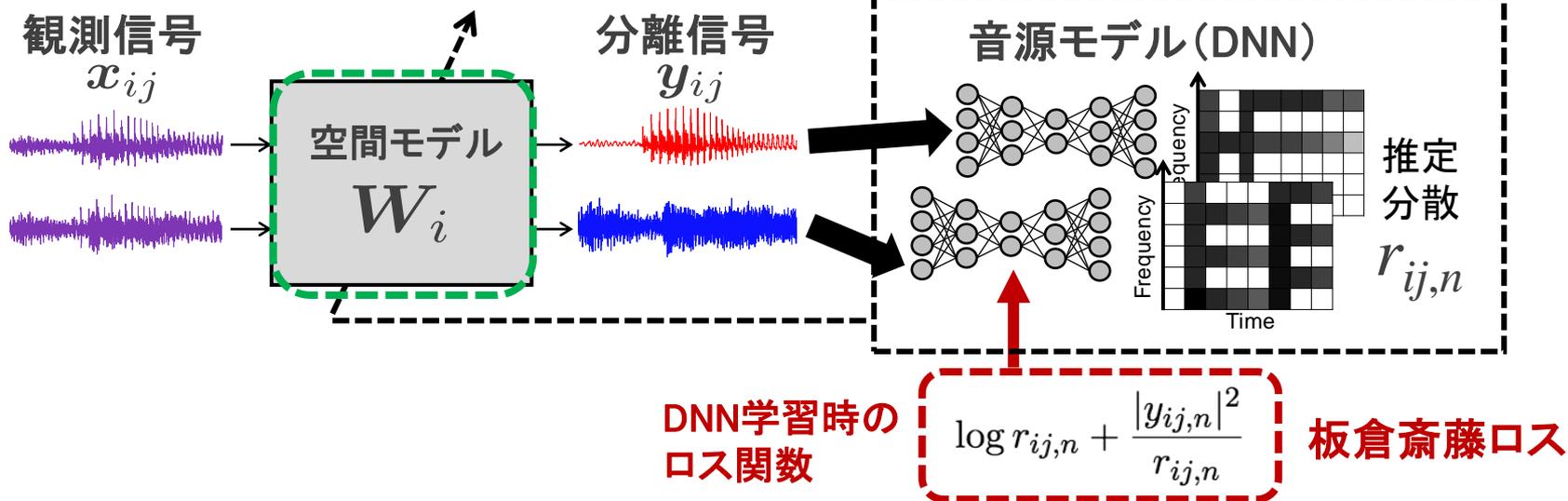
$$y_{ij,n} = \mathbf{w}_{i,n}^H \mathbf{x}_{ij}$$

$$\mathcal{J} = \frac{1}{J} \sum_{i,j,n} \left(\log r_{ij,n} + \frac{|\mathbf{w}_{i,n}^H \mathbf{x}_{ij}|^2}{r_{ij,n}} \right) - \sum_i \log |\det \mathbf{W}_i|^2$$

音源モデル (DNN)

交互に最適化

空間モデル (音源間が独立)



■ **空間モデル**: 各音源が統計的に独立となる分離行列を推定

■ **音源モデル**: J を最小化するような分散 $r_{ij,n}$ を推定するDNNを各音源ごとに構成

Update rules of ILRMA (again)

- ML-based iterative update rules
 - Update rule for \mathbf{W}_i is based on iterative projection
 - Update rules for NMF variables is based on MM algorithm

Spatial model (demixing matrix)

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

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

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

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

where $\mathbf{W}_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,M})^H$
and \mathbf{e}_n is a one-hot vector
that has 1 at n th element

Source model (NMF source model)

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

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

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



Alternating Parameter Update

Update rules of IDLMA (NEW)

- ML-based iterative update rules
 - Update rule for \mathbf{W}_i is based on iterative projection
 - Spectrogram (variance) is estimated via DNN's inference

Spatial model (demixing matrix)

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

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

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

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

where $\mathbf{W}_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,M})^H$
and \mathbf{e}_n is a one-hot vector
that has 1 at n th element

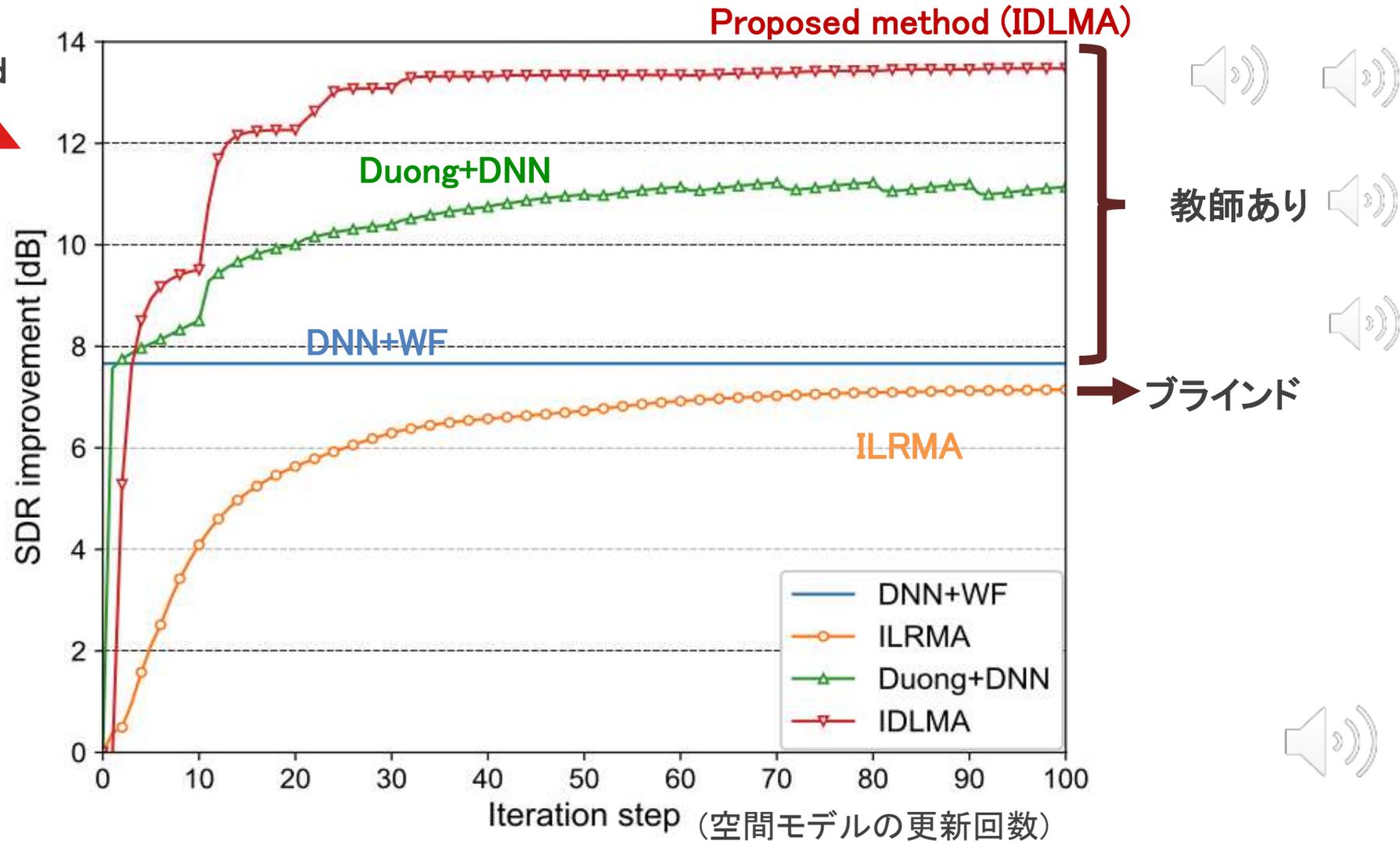
Source model (DNN inference)

$$\text{Train } \log r_{ij,n} + \frac{|y_{ij,n}|^2}{r_{ij,n}} \rightarrow \min$$

$$\text{Infer } r_{ij,n} = \text{DNN}(|y_{ij,n}|^2)$$

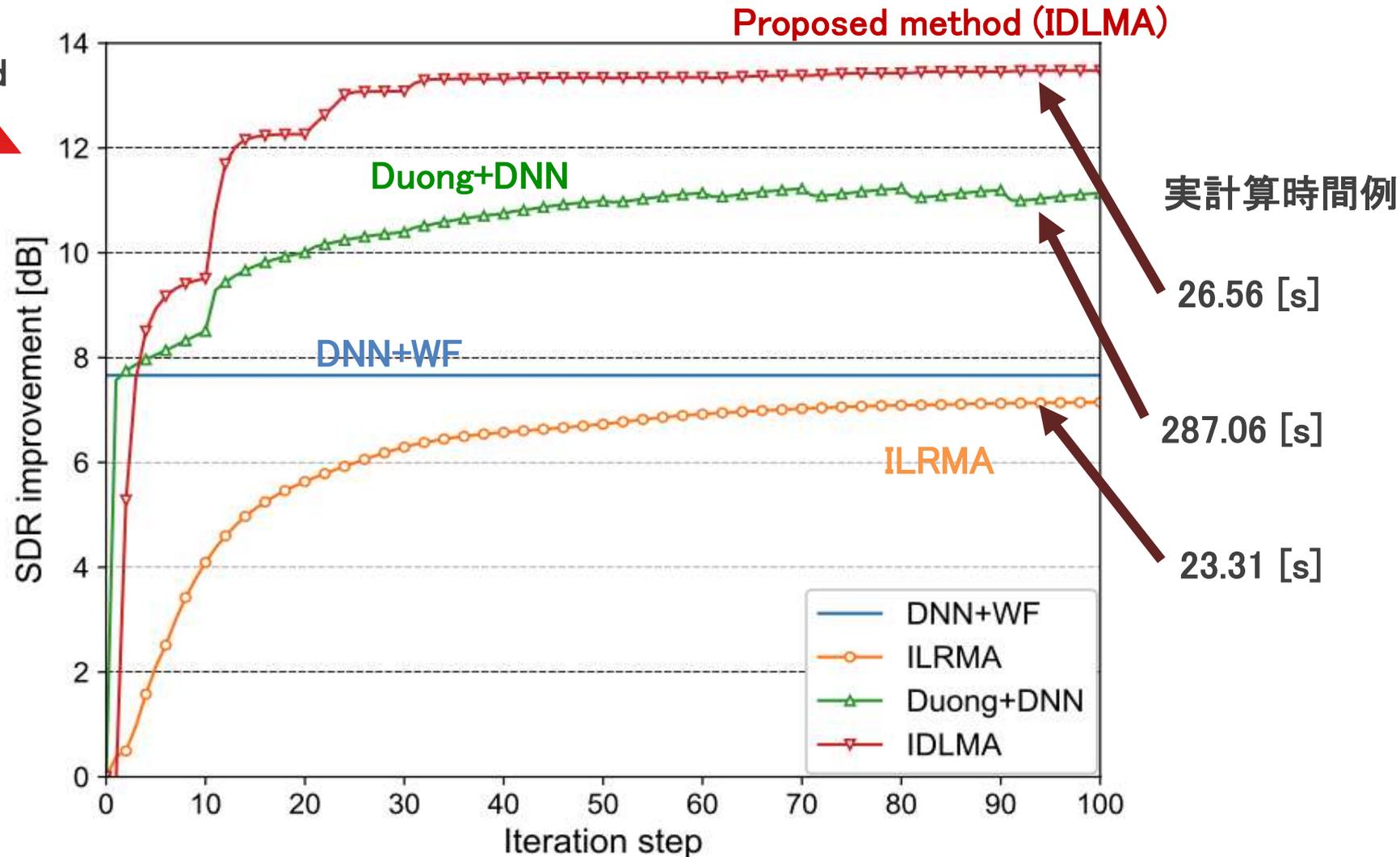
↔
Alternating Parameter Update

簡易実験結果 (角野十、2018年3月音響学会)



■ 10回に1回 DNNで分散行列を更新

簡易実験結果 (角野十、2018年3月音響学会)



■ 10回に1回 DNNで分散行列を更新