# 信号処理論特論 第3&4回 (4/28, 5/12)

# 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

■ 04/14: 第1回 統計的音声音響信号処理概論 ■04/21: 第2回 非負值行列因子分解 ■04/28: 第3回 ブラインド音源分離その1 ■ 05/12: 第4回 ブラインド音源分離その2 ■ 05/19: 第5回 エンハンスメント・高次統計量解析とその応用 ■ 05/26: 第6回 【レポート課題1】 ■ 06/02: 第7回 音声合成 変換その1 ■ 06/09: 第8回 音声合成·変換その2 ■ 06/16: 第9回 音場の計測と制御 ■06/23: 第10回 音楽情報処理 ■ 06/30: 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

■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.



### 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

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

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





■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

■信号を混ぜる一 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



### ■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

■信号を混ぜる一 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





- ■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(<u>BlindSourceSeparation</u>)

■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



### ブラインド音源分離(<u>BlindSourceSeparation</u>)

■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(<u>I</u>ndependent<u>ComponentAnalysis</u>)



元の信号は独立

依存関係がある

なるべく独立に

- 一方から他方が
   推測できない
- 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

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 s が混合行列 H により混ざり合い, J 個の 観測信号 x が T 個得られたとする。



 $|_{\mathcal{U}_1}$ 

■分離行列 Wにより分離信号yを生成する。

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t)$$
  $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_1 \end{bmatrix}$ 

■ W の計算は観測信号 x のみから行う。 Ⅰ個の分離信号 y1,...,y1 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)

### 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)



### Independent Factor Analysis Family



### 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





### ■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

■信号を混ぜる一 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



### 中心極限定理 (Central Limit Theorem)

一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。



### 中心極限定理 (Central Limit Theorem)

■確率変数  $x_1, x_2, \ldots, x_N$  が独立に同一な確率密度関数  $p_x(x)$ に従うならば,  $z = (x_1 + x_2 + \cdots + x_N)/\sqrt{N}$ の 確率密度関数は  $N \to \infty$  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) $\mathbf{Z}$ 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 $\Theta_y(ju)$ $\Theta_{y}(ju) = \Theta_{x}(ju)^{N}$ (: 重畳積分定理) $z = y/\sqrt{N}$ の確率密度関数 $p_z(z)$ $p_z(z) = \sqrt{N} p_u(\sqrt{N}z)$ $z = y/\sqrt{N}$ 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  $\log \Theta_x(ju) = \sum_{n=0}^{\infty} \frac{k_n}{\sqrt{n!}} \frac{(ju)^n}{n!}$ n次キュムラント  $\log \Theta_z(ju)$  のマクローリン展開  $\log \Theta_z(ju) = N \log \Theta_x(ju/\sqrt{N})$  $=\sum_{n=0}^{\infty} \left(\frac{Nk_n}{N^{n/2}}\right) \frac{(ju)^n}{n!}$ 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

(こ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  $\mu$  と分散  $\sigma^2$  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  

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

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

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 $\mathcal{L}[p] = -\int p(x)\log p(x)\mathrm{d}x$ $+\lambda_0 \Big( \int p(x) dx - 1 \Big) + \lambda_1 \Big( \int x p(x) dx \Big) + \lambda_2 \Big( \int x^2 p(x) dx - 1 \Big)$ $\frac{\partial \mathcal{L}}{\partial n} = -1 - \log p(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2 = 0$ $\Rightarrow p(x) = \exp(1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2)$ 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





■信号を混ぜる一中心極限定理 ■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

■NMFと融合した多チャネルブラインド音源分離

FastICA



### 音源信号と混合信号の散布図



### 白色化(sphering) + FastICA


白色化

#### 相関行列が単位行列になる ように x(t) を変換



■無相関化 E{z<sub>i</sub>z<sub>j</sub>} = 0

■分散の正規化 E{z<sub>i</sub>z<sub>i</sub>} = 1



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

$$oldsymbol{y}(t) = oldsymbol{U} oldsymbol{z}(t)$$
  
 $oldsymbol{U} = egin{bmatrix} \cos heta & -\sin heta \ \sin heta & \cos heta \end{bmatrix}$ 



3



分離信号



 $G(y_i) = -\log p(y_i)$  $p(y_i) \propto \exp(-|y_i|)$ 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<sub>i</sub>) = -log p(y<sub>i</sub>) の平均を最小化
 求めたいのは U = [u<sub>1</sub>…u<sub>i</sub>]<sup>T</sup>
 解空間をユニタリ行列に限定

$$y_i \leftarrow u_i^{\mathrm{T}} z$$
 分離信号の計算  
 $u_i \leftarrow \mathbb{E}\{G''(y_i)\}u_i - \mathbb{E}\{G'(y_i)z\}$  Newton法による降下  
 $u_i \leftarrow u_i - \sum_{k=1}^{i-1} (u_k^{\mathrm{T}} u_i)u_k$  Gram-Schmidtの直交化  
 $u_i \leftarrow \frac{u_i}{\|u_i\|_2}$  ノルムの正規化

#### FastICAアルゴリズムによる推定の様子

■赤(□)

- **u**<sub>1</sub> = [1 0]<sup>T</sup> 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(△)

 ■ u<sub>2</sub> = [0 1]<sup>T</sup> 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] う離行列 W を直接推定 観測信号 $X = \{x(t)\}_{1 \le t \le T}$ に対する W の尤度関数 $\mathcal{L}(W) = \prod_{t=1}^{T} p(x(t)|W) \rightarrow \text{maximize}$

■線形変換と確率密度関数

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

-1

■ 音源信号の独立性と非Gauss性を仮定  
$$p(\mathbf{y}(t)) = \prod_{i=1}^{I} p(y_i(t)) \qquad p(y_i(t)): Laplace分布など$$

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

目的関数:  
$$\mathcal{J}(\boldsymbol{W}) = \log |\det \boldsymbol{W}| + \frac{1}{T} \sum_{t} \sum_{i} \log p(y_i(t))$$

■勾配法により W を反復的に更新

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

$$\begin{aligned} \frac{\partial \mathcal{J}(\boldsymbol{W})}{\partial \boldsymbol{W}} &= (\boldsymbol{W}^{\mathrm{T}})^{-1} - \frac{1}{T} \sum_{t} \boldsymbol{\Phi}(\boldsymbol{y}(t)) \boldsymbol{x}(t)^{\mathrm{T}} \\ \mathbf{\Phi}(\boldsymbol{y}(t)) &= \begin{bmatrix} \phi(y_{1}(t)) \\ \vdots \\ \phi(y_{I}(t)) \end{bmatrix} & \phi(y) \ \mathcal{O} \\ \end{bmatrix} \mathbf{\Phi}(y) \ \mathcal{O} \\ \mathbf{E}(y) \ \mathcal{O} \\ \mathbf{E}(y)$$

高次統計から眺めたICAにおける様々なコスト関数



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

- ■2次統計量 ■信号間相関を最小化(複数時間区間利用)
- ■高次統計量1■高次相関をも最小化

$$\mathrm{E}\left\{\mathbf{y}^{3}(t)\mathbf{y}^{\mathrm{T}}(t)\right\} \rightarrow \mathrm{diag}$$

■高次統計量2 ■源信号確率密度関数を仮定  $E\left\{ \Phi(y(t))y^{T}(t) \right\} \rightarrow diag$ 

 $\boldsymbol{\Phi}(\bullet)$ : 符号関数やtanh関数など

#### 非線型関数2の導出



■<u>独立⇒Kullback Leibler Divergenceの最小化問題</u>

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

$$KL(v,z) = \int p(z) \log \frac{p(z)}{p(v)} dz$$
  
上式において...

 $p(z) = p(y_1, \dots, y_K)$ 分離信号y(t)の同時分布密度関数  $p(v) = \prod_{k=1}^{K} p(y_k)$ 周辺分布密度関数の積

とおき、これらのKLを分離行列WIC関して最小化すれば独立

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

非線型関数2の導出(cont'd)



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

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

$$1. \frac{1}{2} \frac{1}$$

 $(:: p(\mathbf{x})d\mathbf{x} = p(\mathbf{y})d\mathbf{y})$ 

非線型関数2の導出(cont'd)



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

 $= -H(Y;W) + \sum_{k=1}^{K} H(Y_k;W)$ 1. 結合エントロピー 2. 周辺エントロピー和





#### 非線型関数2の導出(cont'd)

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

補足

$$\Delta \mathbf{W} \propto -\frac{\partial KL(\mathbf{W})}{\partial \mathbf{W}} = \left( (\mathbf{W}^{\mathrm{T}})^{-1} - \int p(x)\phi(\mathbf{y})\mathbf{x}^{\mathrm{T}}d\mathbf{x} \right)$$
$$= \left( (\mathbf{W}^{\mathrm{T}})^{-1} - \mathbf{E}_{x} \left[ \phi(\mathbf{y})\mathbf{x}^{\mathrm{T}} \right] \right)$$
$$= \left( \mathbf{I} - \mathbf{E}_{y} \left[ \phi(\mathbf{y})\mathbf{y}^{\mathrm{T}} \right] \right) \left( \mathbf{W}^{\mathrm{T}} \right)^{-1}$$
$$\phi(\mathbf{y}) = \left[ \frac{\partial \log p(y_{1})}{\partial y_{1}}, \dots, \frac{\partial \log p(y_{K})}{\partial y_{K}} \right]^{\mathrm{T}}$$
$$\frac{\partial \log p(y_{K})}{\partial y_{K}} = \left( \mathbf{E}_{y} \left[ \phi(\mathbf{y})\mathbf{y}^{\mathrm{T}} \right] \right) \left( \mathbf{W}^{\mathrm{T}} \right)^{-1}$$
$$\frac{\partial \log p(y_{1})}{\partial y_{1}} = \left( \mathbf{E}_{y} \left[ \phi(\mathbf{y})\mathbf{y}^{\mathrm{T}} \right] \right) \left( \mathbf{W}^{\mathrm{T}} \right)^{-1}$$

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 る配法では W の逆行列計算が厄介  $\frac{\partial \mathcal{J}(W)}{\partial W} = (W^{\mathsf{T}})^{-1} - \frac{1}{T} \sum_{t} \Phi(y(t)) x(t)^{\mathsf{T}}$ 自然勾配(Natural Gradient)  $\frac{\partial \mathcal{J}(W)}{\partial W} W^{\mathsf{T}} W = \left(I - \frac{1}{T} \sum_{t} \Phi(y(t)) x(t)^{\mathsf{T}}\right) W$ 

- ■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

$$\begin{aligned} \boldsymbol{y}(t) &\leftarrow \boldsymbol{W} \boldsymbol{x}(t) \\ \boldsymbol{W} &\leftarrow \boldsymbol{W} + \eta \Big( \boldsymbol{I} - \frac{1}{T} \sum_{t} \boldsymbol{\Phi}(\boldsymbol{y}(t)) \boldsymbol{y}^{\mathsf{T}} \Big) \boldsymbol{W} \end{aligned}$$

#### 自然勾配の導出 (1/2)

#### ■最急降下方向

- WをW + ϵdW に変化させたときに最もJ(W + ϵdW) 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 ≠ 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

 $\boldsymbol{W} + \epsilon \mathrm{d} \boldsymbol{W} = (\boldsymbol{I} + \epsilon \mathrm{d} \boldsymbol{W} \boldsymbol{W}^{-1}) \boldsymbol{W}$ 

と変形し、実質的な変化分を $dWW^{-1}$ と捉える

■ dWW<sup>-1</sup>のFrobeniusノルム

 $\|\mathbf{d} \boldsymbol{W} \boldsymbol{W}^{-1}\|_{F}^{2} = \operatorname{tr}(\boldsymbol{W}^{-\mathsf{T}} \mathbf{d} \boldsymbol{W}^{\mathsf{T}} \mathbf{d} \boldsymbol{W} \boldsymbol{W}^{-1}) = \operatorname{tr}(\boldsymbol{W}^{-1} \boldsymbol{W}^{-\mathsf{T}} \mathbf{d} \boldsymbol{W}^{\mathsf{T}} \mathbf{d} \boldsymbol{W})$ を  $\mathbf{d} \boldsymbol{W}$  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 •  $tr(W^{-1}W^{-T}dW^{T}dW) = 1$  という制約の下で  $\mathcal{J}(\boldsymbol{W} + \epsilon \mathrm{d} \boldsymbol{W}) - \mathcal{J}(\boldsymbol{W}) \simeq \epsilon \mathrm{tr}(\nabla \mathcal{J}(\boldsymbol{W})^{\mathsf{T}} \mathrm{d} \boldsymbol{W})$ を最小化する  $\mathrm{d} \boldsymbol{W}$ を 求める最適化問題として定式化 ■ Lagrangeの未定乗数法  $\frac{\partial}{\partial(\mathrm{d}\boldsymbol{W})} \{ \epsilon(\nabla \mathcal{J}(\boldsymbol{W})^{\mathsf{T}}\mathrm{d}\boldsymbol{W}) - \underline{\lambda}(1 - \mathrm{tr}(\boldsymbol{W}^{-1}\boldsymbol{W}^{-\mathsf{T}}\mathrm{d}\boldsymbol{W}^{\mathsf{T}}\mathrm{d}\boldsymbol{W})) \} = 0$  $\Rightarrow \epsilon \nabla \mathcal{J}(\boldsymbol{W})^{\mathsf{T}} + 2\lambda \boldsymbol{W}^{-1} \boldsymbol{W}^{-\mathsf{T}} \mathrm{d} \boldsymbol{W}^{\mathsf{T}} = 0$  $\Rightarrow \mathrm{d} \boldsymbol{W} \propto -\nabla \mathcal{J}(\boldsymbol{W}) \boldsymbol{W}^{\mathsf{T}} \boldsymbol{W}$ ▶逆行列をキャンセルしてくれる

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

自然勾配法による推定の様子

■赤(□): w<sub>1</sub> = [1 0]<sup>T</sup> を初期値に設定 ■緑(△): w<sub>2</sub> = [0 1]<sup>T</sup> 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

- <u> 周波数領域ICA(frequency-domain ICA: FDICA)</u>
  - [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

- <u> 周波数領域ICA(frequency-domain ICA: FDICA)</u>
  - [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<u>因子の順番</u>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<u>因子の大きさ</u>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<sup>-1</sup>diag{0,...,1,...0}の乗算

#### 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列不定性)



空間音響の視点からパーミュテーションを解く

- <u>FDICA+DOAクラスタリング</u> [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



但し, 分離フィルタの タップ長はフーリエ変換 の窓長と同じ

- 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
- FDICAはブラインドな音源分離手法

逆問 題的 観点

信号

処理 観点

- 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

## 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



## 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

コスト L(W) =  $\Sigma_f$  FDICA(W(f)) +  $\lambda$ ・BeamFormer(DOA) 介 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

 $\exists \mathsf{A} \vdash L_f(W) = \lambda_1 \operatorname{FDICA}(W(f)) + \lambda_2 \operatorname{BeamFormer}(\operatorname{DOA})$ 

空間モデル期待値  $DOA = E_f [DOA(W(f))]$ 

## 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





×印  $\lambda_1 = 0, \lambda_2 = 1$ 

それ以外 λ<sub>1</sub>=1, λ<sub>2</sub>=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



■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

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。

■NMFと融合した多チャネルブラインド音源分離



## 周波数成分間の高次相関を用いたBSS(2006~)

<u>独立ベクトル分析(independent vector analysis: IVA)</u>

[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<sub>n</sub>)を多次元分布p(s<sub>n1</sub>,...,s<sub>nI</sub>;σ<sub>n</sub>)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



### 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

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

- 「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*(*s<sub>n</sub>*)
  - 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 
$$p(\mathbf{s}_n) = \alpha \exp\left(-\sqrt{(\mathbf{s}_n - \mu_i)^{\mathrm{H}} \Sigma_n^{-1} (\mathbf{s}_n - \mu_i)}\right)$$

- 後者は $\Sigma_n = I$ (互いに無相関)の場合でも, $S_{ni}$ が互いに依存

分散共分散行列

## 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 → 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

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

零平均,零共分散,共通分散とすると

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

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 → 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(Θ) → min
- 補助関数  $Q(\Theta, \tilde{\Theta}) \ge J(\Theta)$
- 反復更新則 補助変数の更新 (E-step)  $\tilde{\Theta}^{(k+1)} = \operatorname{argmin}_{\bar{\Theta}} Q(\Theta^{(k)}, \tilde{\Theta})$ パラメータの更新 (M-step)  $\Theta^{(k+1)} = \operatorname{argmin}_{\bar{\Theta}} Q(\Theta, \tilde{\Theta}^{(k+1)})$
- 補助関数の例(絶対値関数の最小化)

 $|x| \leq \frac{1}{2x_0} |x|^2 + \frac{x_0}{2} \implies$ 二次関数の和も二次関数なので、 絶対値関数の和も容易に最小化できる 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<sub>73</sub>



#### 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

 $G(z) = G_R(||z||_2)$  求 は多次元複素変数 球対称(ノルムだけの)分布

IVAのコスト関数」及びその補助関数Qを以下のように定める

$$J(\boldsymbol{W}) = \frac{1}{N_{\tau}} \sum_{\tau=1}^{N_{\tau}} \sum_{k=1}^{K} \underbrace{G(\boldsymbol{y}_{k}(\tau))}_{\boldsymbol{\omega}} - \sum_{\omega=1}^{N_{\omega}} \log |\det W(\omega)|$$
$$Q(\boldsymbol{W}, \boldsymbol{V}) = \sum_{\omega=1}^{N_{\omega}} \left[ \frac{1}{N_{\tau}} \sum_{\tau=1}^{N_{\tau}} \sum_{k=1}^{K} \underbrace{\boldsymbol{w}_{k}^{h}(\omega) V_{k}(\omega) \boldsymbol{w}_{k}(\omega)}_{\boldsymbol{\omega}} - \log |\det W(\omega)| \right]$$

補助変数

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

#### 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

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

■補助関数2が小さくなるように₩の更新



#### 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

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

補助関数Qが小さくなるようにWの更新

【P(反復射影)アルゴリズム】 for k = 1: K  $w_k(\omega) \leftarrow (W(\omega)V_k(\omega))^{-1}i_k$   $w_k(\omega) \leftarrow w_k(\omega)/\sqrt{w_k^h(\omega)V_k(\omega)w_k(\omega)}$ end

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。 分離行列を行べクトル毎に更新することは閉形式で可能(<u>補助関数は減少する</u>)。

76

#### 補助関数型IVA:実験結果例

[小野順貴先生2013年5月音 声研究会講演資料より抜粋]

■ 2音源分離の場合:自然勾配法によるIVAと比較





■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~)



## 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

$$X = t_1v_1 + t_2v_2 + t_3v_3 + \cdots$$

- 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]

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

最小化は等価 複素球対称ガウス分布(零平均) <u>観測の複素数値</u>  $\mathcal{L} = -\sum_{i,j} \log \mathcal{N}_c (x_{ij}|0, \sum_k t_{ik} v_{kj}) \leftarrow 複素ガウスの分散$   $= IJ \log \pi + \sum_{i,j} \left[ -\log \left( \sum_k t_{ik} v_{kj} \right) + \frac{|x_{ij}|^2}{\sum_k t_{ik} v_{kj}} \right]$ 

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

## 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



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



# IVAとNMFを融合した新しいBSS(2016~)

・ <u>独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA)</u>



- 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(対数尤度)関数

$$\mathcal{J} = \sum_{i,j} \left[ \sum_{m} \log \sum_{k} z_{mk} t_{ik} v_{kj} + \sum_{m} \frac{|y_{ij,m}|^2}{\sum_{k} z_{mk} t_{ik} v_{kj}} - 2 \log |\det W_i| \right]$$
  
主なパラメータ
  
空間モデル
  
 $W_i$ :分離フィルタ行列
  
 $y_{ij,m} = w_{i,m}^{\mathrm{H}} x_{ij}$ :分離信号
  
 $t_{ik}$ :基底行列要素
  
 $v_{kj}$ :アクティベーション行列要素
  
 $TV$ : KランクNMFモデル)

*Z<sub>mk</sub>*:分割関数(基底の共有を許す重み)

ILRMAのコスト(対数尤度)関数→これを最小化



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) 87

- <u>ILRMAコスト関数はIVAコストとNMFコストの複合</u>
  - 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  

$$V_{i,m} = \frac{1}{J} \sum_{j} \frac{1}{r_{ij,m}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{H}}$$
  
 $\boldsymbol{w}_{i,m} \leftarrow (\boldsymbol{W}_i \boldsymbol{V}_{i,m})^{-1} \boldsymbol{e}_m$   
 $\boldsymbol{w}_{i,m} \leftarrow \boldsymbol{w}_{i,m} \left( \boldsymbol{w}_{i,m}^{\mathrm{H}} \boldsymbol{V}_{i,m} \boldsymbol{w}_{i,m} \right)^{-\frac{1}{2}}$   
 $y_{ij,m} \leftarrow \boldsymbol{w}_{i,m}^{\mathrm{H}} \boldsymbol{x}_{ij}$ 

但し、 $W_i = (w_{i,1} \cdots w_{i,M})^{\mathrm{H}}$ ,  $e_m$ 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  

$$z_{mk} \leftarrow z_{mk} \sqrt{\frac{\sum_{i,j} |y_{ij,m}|^2 t_{ik} v_{kj} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-2}}{\sum_{i,j} t_{ik} v_{kj} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-1}}}$$
  
 $t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{j,m} |y_{ij,m}|^2 z_{mk} v_{kj} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-2}}{\sum_{j,m} z_{mk} v_{kj} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-1}}}}$   
 $v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,m} |y_{ij,m}|^2 z_{mk} t_{ik} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-2}}{\sum_{i,m} z_{mk} t_{ik} (\sum_{k'} z_{mk'} t_{ik'} v_{k'j})^{-1}}}}$   
 $r_{ij,m} = \sum_{k} z_{mk} t_{ik} v_{kj}$ : 音源毎の推定分散

- 複雑なスペクトログラム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) 88

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





#### 何が違うのか?どちらが良いのか?

FDICALこおける最適化:自然勾配法  
■ 勾配法では 
$$W$$
 の逆行列計算が厄介  
 $\frac{\partial \mathcal{J}(W)}{\partial W} = (W^{\mathsf{T}})^{-1} - \frac{1}{T} \sum_{t} \Phi(y(t))x(t)^{\mathsf{T}}$   
■ 自然勾配(Natural Gradient) [Amari, 1997]  $\mathcal{O}_{\mathfrak{M}}^{\mathfrak{m}} \mathcal{O}_{\mathfrak{M}} \mathcal{O}_{\mathfrak{M}}^{\mathsf{T}} \mathcal{W} = \left(I - \frac{1}{T} \sum_{t} \Phi(y(t))x(t)^{\mathsf{T}}\right) \mathcal{W}$   
■ 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  
 $y(t) \leftarrow Wx(t)$   
 $W \leftarrow W + \eta \left(I - \frac{1}{T} \sum_{t} \Phi(y(t))y^{\mathsf{T}}\right) \mathcal{W}$ 

-

### 生成モデル型の例:多チャネル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	$R^{(\boldsymbol{x})} = R^{(\boldsymbol{s})} \circ \boldsymbol{T} \boldsymbol{V} + R^{(\boldsymbol{b})}$	R <sup>(s)</sup> はランク1 R <sup>(b)</sup> はフルランク	NMF	EMで R <sup>(s)</sup> , R <sup>(b)</sup> , NMF変数を推定
Arberet et al., 2010	$R^{(\boldsymbol{x})} = R^{(\boldsymbol{s})} \circ \boldsymbol{T} \boldsymbol{V} + R^{(\boldsymbol{b})}$	R <sup>(s)</sup> もR <sup>(b)</sup> もフル ランク	NMF	EMで R <sup>(s)</sup> , R <sup>(b)</sup> , NMF変数を推定
Ozerov et al., 2011	$R^{(\boldsymbol{x})} = \left(R^{(\boldsymbol{s})}\boldsymbol{Z}\right) \circ \boldsymbol{T}\boldsymbol{V} + R^{(\boldsymbol{b})}$	R <sup>(s)</sup> はランク1 R <sup>(b)</sup> はフルランク	分割関数付き のNMF	EMで R <sup>(s)</sup> , R <sup>(b)</sup> , NMF変数を推定
Sawada et al., 2013	$R^{(m{x})} = \left(R^{(m{s})}m{Z} ight) \circ m{T}m{V}$	フルランクの R <sup>(s)</sup>	分割関数付き のNMF	乗算更新式でR <sup>(®)</sup> , NMF変数を推定
Kitamura et al., 2015	$R^{(x)} = \left(R^{(s)}Z ight) \circ TV$	ランク1の R <sup>(s)</sup>	分割関数付き のNMF	反復射影(IP)で分 離行列 <i>W=A<sup>-1</sup></i> , 乗算更新式で NMF変数を推定

#### NMFの多チャネル信号への拡張

<u>最も一般化された多チャネルNMF</u> [Sawada, 2013]



# 空間相関行列とは

- <u>空間相関行列 スは空間共分散行列</u> [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 Source image

 $oldsymbol{R}_{ij}^{(oldsymbol{x})} = \mathrm{E}\left[oldsymbol{x}_{ij}oldsymbol{x}_{ij}^{\mathrm{H}}
ight] = \sum_{n=1}^{N} r_{ij,n}oldsymbol{R}_{i}^{(oldsymbol{s}_{n})}$ 

 $\hat{x}_{ij}^{(n)} = r_{ij,n} R_i^{(s_n)} R_{ij}^{(x)^{-1}} x_{ij}$ 

- 観測信号 $x_{ij}$ 中のn番目の音源成分のみを $x_{ij}^{(n)}$ と表すとき

の 分散  $E\left[\boldsymbol{x}_{ij}^{(n)}\boldsymbol{x}_{ij}^{(n)H}\right] = r_{ij,n}\boldsymbol{R}_{i}^{(\boldsymbol{s}_{n})}$  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)

から音源分離が可能(劣決定も可)

時間周波数で分散共分散が 変動する多変量ガウス分布

測の  
成モデル
$$p(\boldsymbol{x}_{ij};r_{ij},\boldsymbol{R}_{ij}^{(\boldsymbol{x})}) = \frac{1}{\det(\pi \boldsymbol{R}_{ij}^{(\boldsymbol{x})})} \exp\left(-\boldsymbol{x}_{ij}^{\mathrm{H}} \boldsymbol{R}_{ij}^{(\boldsymbol{x})^{-1}} \boldsymbol{x}_{ij}\right)$$
  
チャネル

多チャネル Wiener filter

音源毎の

観測の

観

牛

空間共分散

空間共分散

95

### 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 
$$\mathrm{E}\left[\boldsymbol{x}_{ij}^{(n)}\boldsymbol{x}_{ij}^{(n)^{\mathrm{H}}}\right] = r_{ij,n}\boldsymbol{R}_{i}^{(\boldsymbol{s}_{n})}$$

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

ランク1  
空間モデル 
$$oldsymbol{R}_i^{(oldsymbol{s}_n)} = oldsymbol{a}_{i,n} a_{i,n}^{\mathrm{H}}$$
  $oldsymbol{a}_{i,n2}$ 

- 「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 <sub>時不変混合行列</sub>  $A_i = (a_{i,1} \ a_{i,2})$ 

$\langle \rangle$		/	\	$\langle \rangle$
$(x_{ij,1})$	_	$a_{i,11}$	$a_{i,12}$	$(s_{ij,1})$
$\left( x_{ij,2} \right)$		$\langle a_{i,21} \rangle$	$a_{i,22}$	$\left( s_{ij,2} \right)$



- 「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)

多チャネルNMFとILRMAのミッシングリンク

多チャネルNMFの目的関数にランク1空間モデルを導入

$$\mathcal{J} = \sum_{i,j} \left[ \operatorname{tr} \left( \mathsf{X}_{ij} \left( \sum_{k} \left( \sum_{m} \mathbf{R}_{i}^{(s_{m})} z_{mk} \right) t_{ik} v_{kj} \right)^{-1} \right) + \log \det \sum_{k} \left( \sum_{m} \mathbf{R}_{i}^{(s_{m})} z_{mk} \right) t_{ik} v_{kj} \right]$$

1. ランク1チャネル間相関を導入  $(R_i^{(s_m)} = a_{i,m}a_{i,m}^{H})$ 

2. 混合行列 $A_i$ を用いて表現しなおす

3. 分離行列  $W_i = A_i^{-1}$ と分離信号  $y_{ij} = W_i x_{ij}$  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!

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

- ランク1空間モデル制約付き多チャネル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



- 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 $\kappa_R$ 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 ~ 2π**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



・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)







# 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


#### LRMA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	
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)	



#### 実験結果: fort\_minor-remember\_the\_name



#### 実験結果: ultimate\_nz\_tour



#### 実験結果: ultimate\_nz\_tour(収束の様子)



#### 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



116

#### 実験結果: male3\_liverec\_1m



117

者察

- 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



#### 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] 巻いた状態 被災者はいるのか? 人の声を発見! 索状ロボット ブラインド 統計的音声強調 音声·環境音認識 音源分離 高次統計量制御 高次統計量による 伸ばした状態(先端部) ICA, IVA, 事前分布推定· 認識率予測· 最適化 前処理最適化 NMF, etc. 10011100 1 2-DOF 分布推進 【 | :マイクロホン】 屈曲 機構 いかなる曲がりくねった形状においても 小型IMU. 位置不定マイク同士が協調して騒音の マイクロフォン. ガスセンサ、など 先端に装着した 中から被災者の声を見つけ出す カメラとLED照明 内蔵可能

# 高残響下におけるLRMA 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 ➡ 分離性能が劣化



#### 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 其底共有型II RMA
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 (SαS) 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

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

- 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

 $z_1 \sim S\alpha S(z; 0, |z_1|^{\alpha})$ 

[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



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

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

### 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 **W**<sub>i</sub> |"と"**w**<sub>i,n</sub> 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あろうか?
- コスト関数の比較
  - 従来の(時変ガウス型)ILRMA (IP適用可能)

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

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

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

対数関数を外すような工夫が必要!

• 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  $\log(z) \leq \frac{1}{\lambda}(z-\lambda) + \log \lambda$ 

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

$$\leq -2J \sum_{i} \left[ \log |\det \mathbf{W}_{i}| + \sum_{i,j,n} \left[ \left( 1 + \frac{\nu}{2} \right) \frac{1}{\alpha_{ij,n}} \left( 1 + \frac{2}{\nu} \frac{\left[ \left[ \mathbf{w}_{i,n}^{\mathrm{H}} \mathbf{x}_{ij} \right]^{2} \right]}{\left(\sum_{k} t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}}} - \alpha_{ij,n} \right)$$

$$+ \left( 1 + \frac{\nu}{2} \log \alpha_{ij,n} \right) + \log \left( \sum_{k} t_{ik,n} v_{kj,n} \right)^{\frac{2}{p}} \right]$$

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  $\log(z) \leq \frac{1}{\lambda}(z-\lambda) + \log \lambda$ 

• IPに基づく
$$W_i$$
の最適化

$$egin{aligned} oldsymbol{U}_{i,n} &= rac{1}{J} \sum_{j} rac{oldsymbol{x}_{ij} oldsymbol{x}_{ij}}{rac{
u}{
u+2} \left(\sum_{k} t_{ik,n} v_{kj,n}
ight)^{rac{2}{p}} + rac{
u}{
u+2} |y_{ij,n}|^2} \ oldsymbol{w}_{i,n} &\leftarrow \left(oldsymbol{W}_{i}oldsymbol{U}_{i,n}
ight)^{-1} oldsymbol{e}_n \ oldsymbol{w}_{i,n} &\leftarrow oldsymbol{w}_{i,n} \left(oldsymbol{w}_{i,n}^{\mathrm{H}} oldsymbol{U}_{i,n} oldsymbol{w}_{i,n}
ight)^{-rac{1}{2}} \end{aligned}$$

 $e_n$ : unit vector whose *n*th element is unity

Cf. MNMF [H. Sawada+, 2013] は-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  

$$\mathcal{L} = -2J\sum_{i} \log |\det W_{i}| + \sum_{i,j,n} \left[ \left(1 + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \frac{|W_{i,n}^{H} x_{ij}|^{2}}{\left(\sum_{k} t_{ik,n} v_{kj,n}\right)^{2}}\right) + \log \left(\sum_{k} t_{ik,n} v_{kj,n}\right)^{\frac{2}{p}} \right]$$

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

$$\underbrace{\mathsf{Pfst}}_{ik,n} \leftarrow t_{ik,n} \left[ \frac{\sum_{j} \frac{z_{ij,n}^{p}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^{2}} v_{kj,n}}{\sum_{j} \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \right] \frac{\frac{p}{p+2}}{\text{fibble for the state state$$

#### 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



### 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
- (a) Isotropic complex GGD (b) Isotropic complex Student's t βを小さくするとスパ distribution = 2 $\rightarrow \infty$ B ースな分布になる  $\binom{u_{ijn}}{0}_{0}^{0}$  $p(y_{ij,n})$ ▶ 特に *B* =1 は 複素 ラプ ラス分布となりIVAの  $\operatorname{Re}[y_{ij,n}]$  -2  $\operatorname{Re}[y_{ij,n}] \sim$ 自然な拡張に帰着  $\operatorname{Im}[y_{ij,n}]$ Im yin 4  $\succ \beta \rightarrow 1 に つ れ て 、 低$ B = 1 $p_{(y_{ij,n})}^{(n)}$  $\binom{n_{ij,n}}{0}{}^{0}$ 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) Re yin  $\operatorname{Re}[y_{ij,n}]$  -2 4 4 Imy 4 4 Im yin

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  $t_{ik,n} \leftarrow t_{ik,n} \begin{bmatrix} \sum_{j} \frac{z_{ij,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} v_{kj,n} \\ \frac{1}{\sum_{j} \frac{1}{\sum_{k'} t_{ik',n} v_{k'j,n}} v_{kj,n}} \end{bmatrix}^{\frac{p}{\beta+p}}$ 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

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

真の観測yと低ランクモデル  $\sigma$  との  $\beta/p$ :(1- $\beta/p$ )比の幾何平均 ⇒  $\beta$  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



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~

### ILRMAにおける事前分布の導入



#### ■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、偏微分だけでは十分ではなく、「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)

$$\mathcal{J} = J \sum_{i} \left[ \sum_{n} \boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{D}_{i,n} \boldsymbol{w}_{i,n} - \log |\det \boldsymbol{W}_{i}|^{2} \right] + \mathcal{C}$$

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

$$\mathcal{J}_{\mathrm{P}} = J \sum_{i} \left[ \sum_{n} \boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{D}_{i,n} \boldsymbol{w}_{i,n} - \log |\det \boldsymbol{W}_{i}|^{2} \right] + \left[ J \sum_{i,n} \lambda_{n} || \boldsymbol{w}_{i,n} - \hat{\boldsymbol{w}}_{i,n} ||^{2} \right] + \mathcal{C}$$
$$= J \sum_{i} \left[ \sum_{n} \left( \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n} \boldsymbol{w}_{i,n} - \lambda_{n} \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}} \boldsymbol{w}_{i,n} - \lambda_{n} \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{w}}_{i,n} \right) - \log |\det \boldsymbol{W}_{i}|^{2} \right] + \mathcal{C}'$$

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

 ■ 新たな目的関数は w<sub>i,n</sub> 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

$$\mathcal{J}_{\mathrm{P}} = J \sum_{i} \left[ \sum_{n} \left( \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n} \boldsymbol{w}_{i,n} - \lambda_{n} \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}} \boldsymbol{w}_{i,n} - \lambda_{n} \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{w}}_{i,n} \right) - \log |\det \boldsymbol{W}_{i}|^{2} \right] + \mathcal{C}'$$

- W<sub>i</sub>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 w<sub>i,n</sub> 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

$$\mathcal{J}_{\mathrm{P}} = J \sum_{i} \left[ \sum_{n} \left( \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n} \boldsymbol{w}_{i,n} - \lambda_{n} \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}} \boldsymbol{w}_{i,n} - \lambda_{n} \boldsymbol{w}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{w}}_{i,n} \right) - \log |\det \boldsymbol{W}_{i}|^{2} \right] + \mathcal{C}'$$

目的関数  $\mathcal{J}_{\mathrm{P}}$  の停留点に  $w_{i,n}$  を更新 … log|det $W_i$ |の偏微分が厄介 ■ 余因子行列を次のように定義:  $B_i = (b_{i,1}, \dots, b_{i,N}) \stackrel{\text{def}}{=} (\det W_i) W_i^{-1}$ ■ logdet項は、次のように変形できる: log | det  $W_i$ |<sup>2</sup> = log | $w_{i,n}^{\mathrm{H}} b_{i,n}$ |<sup>2</sup> ■  $b_{i,n}$  は  $w_{i,n}$  と関係なく定まるため、 $w_{i,n}^{*}$  によるlogdetの偏微分は以下:  $\frac{\partial \log |w_{i,n}^{\mathrm{H}} b_{i,n}|^2}{\partial w_{i,n}^{*}} = \frac{b_{i,n}}{w_{i,n}^{\mathrm{H}} b_{i,n}}$  $\rightarrow \frac{\partial \mathcal{J}_{\mathrm{P}}}{\partial w_{i,n}^{*}} = 0$ の解へと $w_{i,n}$ 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

$$\widehat{D}_{i,n}w_{i,n} - \lambda_n \widehat{w}_{i,n} - \frac{1}{w_{i,n}^{\mathsf{H}} b_{i,n}} b_{i,n} = 0$$
  
 $\beta_{i,n} = 1/w_{i,n}^{\mathsf{H}} b_{i,n}$  とおいて整理すると、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

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

この方程式の解は、次のように代数的に求まる:

$$\begin{split} \beta_{i,n} &= e^{j\phi_{i,n}} / \sqrt{\boldsymbol{b}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \boldsymbol{b}_{i,n}} \\ \beta_{i,n} &= \frac{\lambda_n \boldsymbol{b}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \widehat{\boldsymbol{w}}_{i,n}}{2\boldsymbol{b}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \boldsymbol{b}_{i,n}} \left[ -1 \pm \sqrt{1 + \frac{4\boldsymbol{b}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \boldsymbol{b}_{i,n}}{\lambda_n^2 |\boldsymbol{b}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \widehat{\boldsymbol{w}}_{i,n}|^2}} \right] \quad (\lambda_n \widehat{\boldsymbol{w}}_{i,n}^{\mathsf{H}} \widehat{\boldsymbol{D}}_{i,n}^{-1} \boldsymbol{b}_{i,n} \neq 0) \end{split}$$
### 反復更新式





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

$( heta_1, heta_2)$	AuxIVA	MNMF	t-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における問題点:音源の低ランク性?



### ILRMAにおける問題点:音源の低ランク性?



• 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。



#### 音源の低ランク性?(例:音声信号)





■ 音源モデル:Jを最小化するような分散 r<sub>iin</sub>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  $W_i$  is based on iterative projection
  - Update rules for NMF variables is based on MM algorithm



**Alternating Parameter Update** 

# Update rules of IDLMA (NEW)

- ML-based iterative update rules
  - Update rule for  $W_i$  is based on iterative projection
  - Spectrogram (variance) is estimated via DNN's inference



Alternating Parameter Update

# 簡易実験結果(角野+、2018年3月音響学会)



# 簡易実験結果(角野+、2018年3月音響学会)

