

スパース表現に基づく 音響信号のモデル化と分析

亀岡弘和

日本電信電話株式会社
NTTコミュニケーション科学基礎研究所
kameoka@cs.brl.ntt.co.jp

講演者紹介：2002年東京大学工学部計数工学科卒業
2004年同大学大学院情報理工学系研究科修士課程修了
2007年同大学大学院情報理工学系研究科博士課程修了
同年日本電信電話株式会社に入社
現在NTTコミュニケーション科学基礎研究所
音声や音楽を対象とした音響信号処理の研究に従事
日本音響学会, 情報処理学会, 電子情報通信学会, IEEE各会員

目次

1. スペース表現による音響信号処理の基本コンセプト

2. 非負値行列因子分解(NMF)の応用例と基礎

3. 複素NMF [H. Kameoka et al., 2008]

振幅スペクトルの基底関数によって構成される複素スペクトルモデル

4. 複合自己回帰系 [H. Kameoka & K. Kashino, 2009]

音声生成モデル(“ソースフィルタモデル”)におけるソースとフィルタがそれぞれ少数の基底関数により構成される音声信号の統計モデル

講演概要

実世界の音源から生成される音響信号は一見すると多様な波形として観測されるが、例えば、人の声には/a/や/i/等といった音素単位の要素があり、ピアノの演奏には音階単位の要素があるように、ある空間や量に着目すると限られた数の基本要素だけで構成されていると見ることができる。本講演では、事前には未知な信号の基本要素を自律的に学習するアプローチにより、このような仮定をもとに分析を行う音響信号処理の枠組を解説する。その上で、複素NMFと複合自己回帰分析法と呼ぶ上記の考え方をベースとした2つの信号解析法の原理と応用例を紹介する。

スパース表現

基底関数
Dictionary

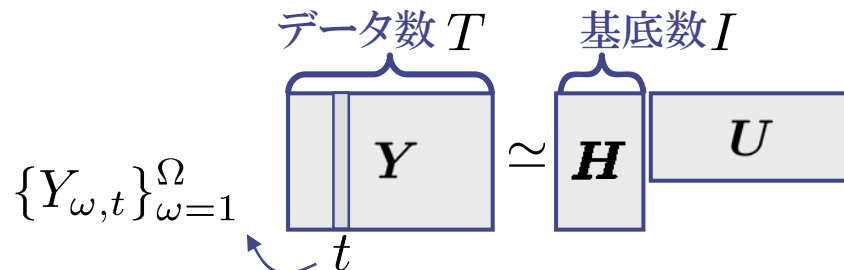
1. データ Y_ω を $\sum_{i=1}^I \underline{H_{\omega,i}} U_i$ の形のモデルで表現
2. $U_1, \dots, U_I$ がスパース (少数を除きほとんどが0)

$\{H_{\omega,2}\} \{H_{\omega,2}\}$
 $\dots \{H_{\omega,I}\}$

$\{Y_\omega\}$

◆ 低ランクモデル ($\Omega > I$)

- 複雑なデータセットを少ない基底で表現



◆ 基底のスパース正則化学習

- データに混在する独立な情報を抽出

スパース表現による音響信号処理

◆実世界音響信号処理

- 観測信号から現象を説明することが目的
- いかに実世界音響信号の構成音をうまくモデル化できるか

◆メッセージ伝達媒体としての音

- 離散的なシンボル情報を波形で表現し伝達
 - ◆音声 → 音素単位 (/a/, /i/, /u/, ...)
 - ◆音楽 → 音階単位 (ドレミファ, ...)
- 各シンボルがどういう波形で表現されるかは音源に固有(すなわち未知)

◆音源の仮定とモデル化

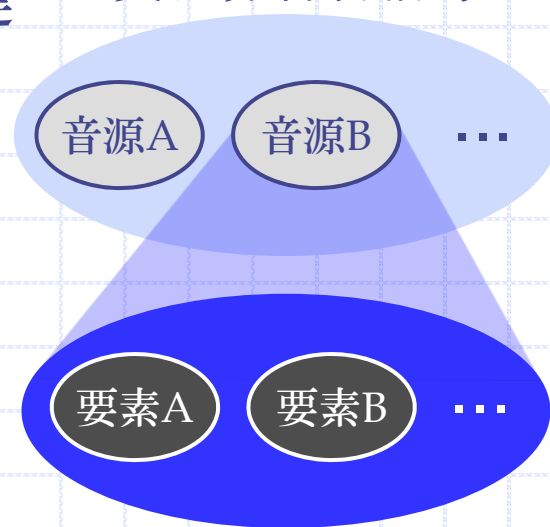
(仮定)

音源信号は限られた種類のシンボル単位に相当する未知の独立成分から成る



少ない基底関数の
スパースな重畳

実世界音響信号



背景

◆ 非負値行列因子分解 (Nonnegative Matrix Factorization; NMF)

$$Y \simeq HU \quad (H \succeq 0, U \succeq 0)$$

■ 画像処理分野で生まれた技術

[D.D. Lee & H.S. Seung, 1999]

■ 音のスペクトログラムを 画像と見なして適用 → 音声分離・自動採譜等

[P. Smaragdis et al., 2003]

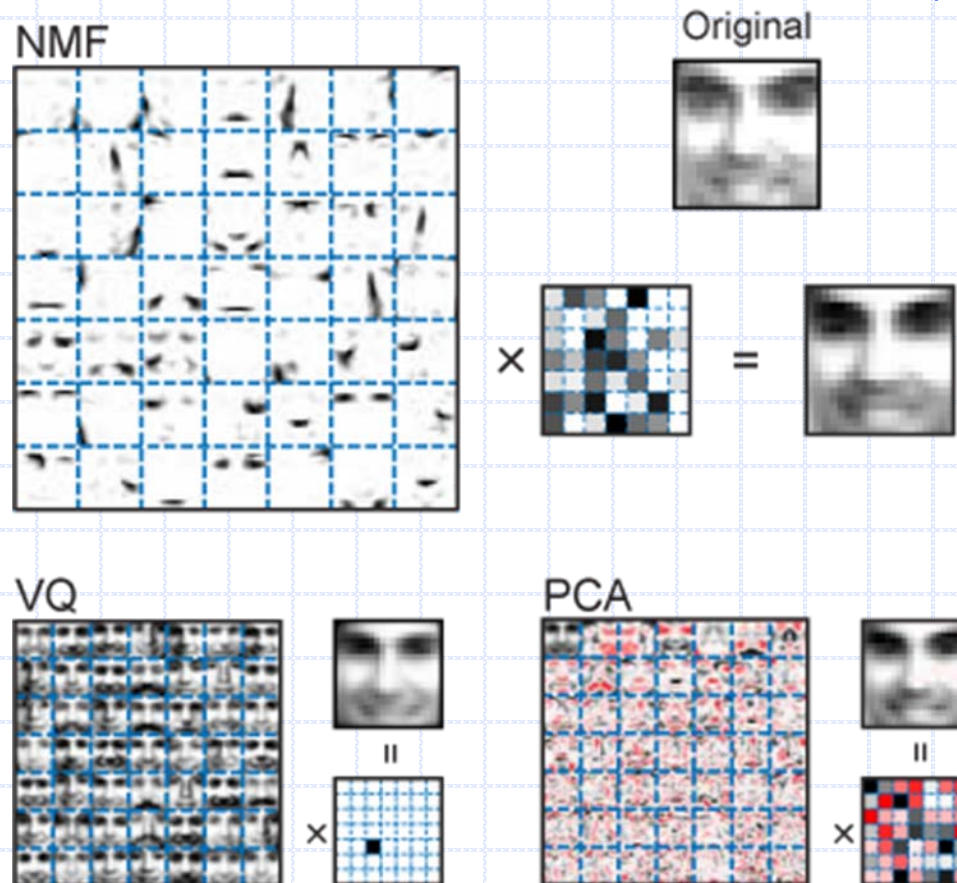
以降極めて多数...

■ 効率的な反復アルゴリズム

[D.D. Lee & H.S. Seung, 2000]

$$H_{k,i} \leftarrow H_{k,i} \frac{[YU^T]_{k,i}}{[HUU^T]_{k,i}}$$

$$U_{i,j} \leftarrow U_{i,j} \frac{[H^T Y]_{i,j}}{[H^T H U]_{i,j}}$$



NMFによる振幅スペクトログラムの分解表現

◆ 音響信号 $y(u) \in \mathbb{R}$

短時間フーリエ変換
(時間周波数分解)

$$Y_{\omega,t} = \int W(u)y(u+t)e^{j\omega u}du \in \mathbb{C}$$

絶対値をとる

$|Y_{\omega,t}|$: 時刻 t に周波数 ω の成分が
どれほど含まれているか

$$Y \simeq HU \quad (H \succeq 0, U \succeq 0)$$

非負値行列因子分解 (NMF)

$$H_{\omega}^i \leftarrow H_{\omega}^i \frac{[YU^T]_{k,i}}{[HUU^T]_{k,i}}$$

$$U_t^i \leftarrow U_t^i \frac{[H^TY]_{i,t}}{[H^THU]_{i,t}}$$

繰り返し生起する
振幅スペクトルパターンが出

周波数↑

振幅スペクトログラム Y

規則性が!

U

スパース(疎)になる!

各基底の
アクティビティ

振幅スペクトル基底

H

周波数↑

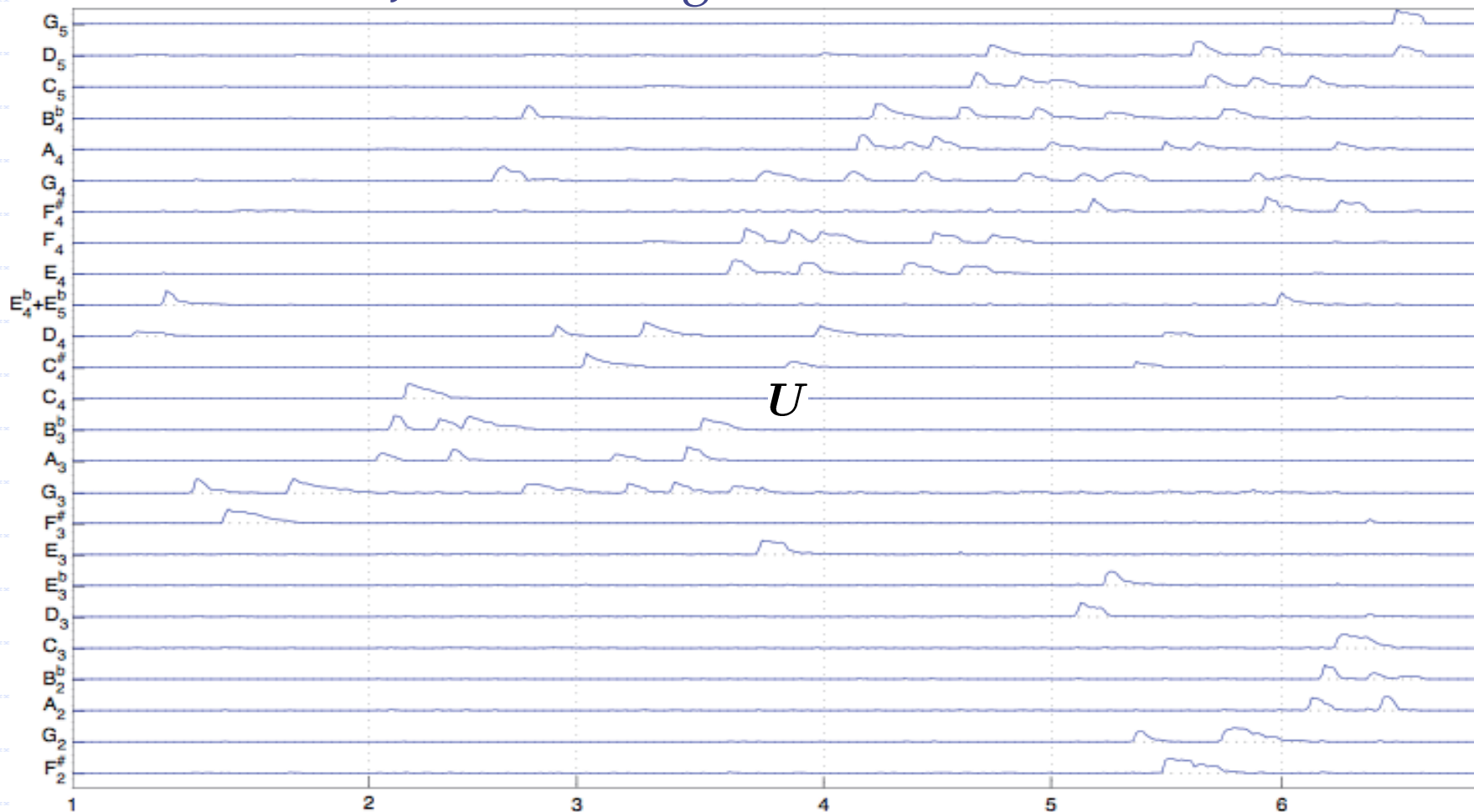
低ランクスペクトログラム HU

時刻→

NMFの音響信号処理応用事例1

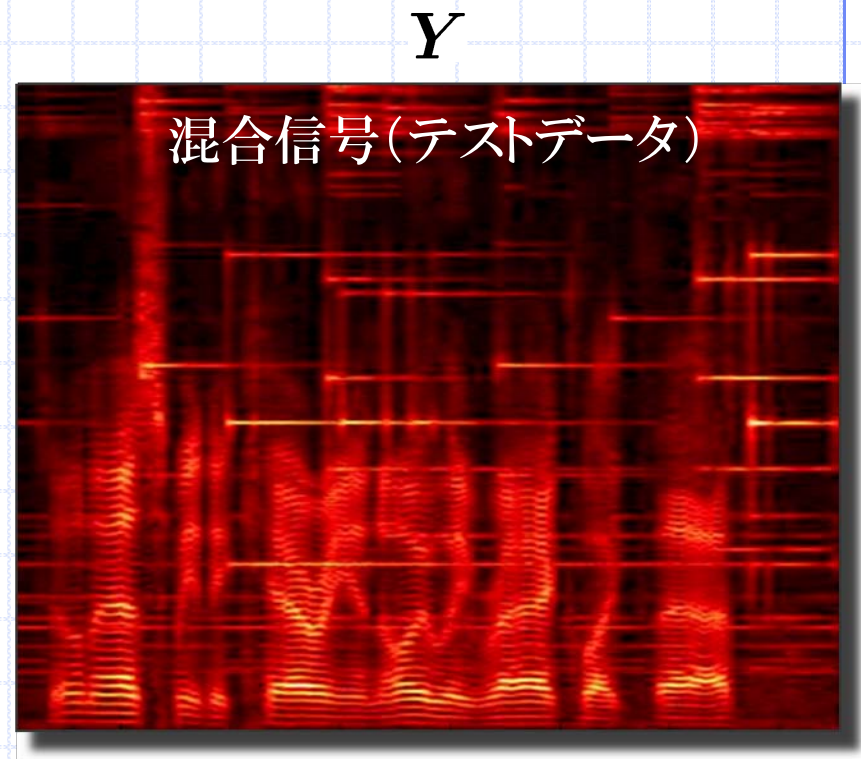
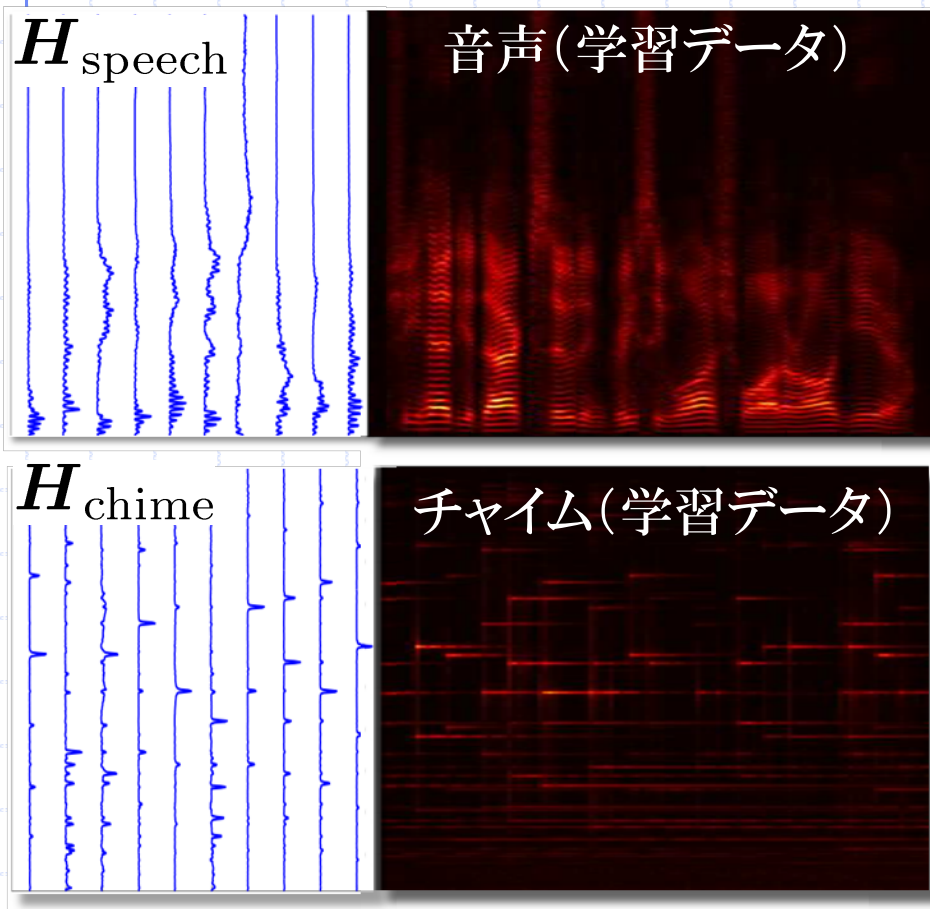
◆ 自動採譜 [P. Smaragdis et al., 2003]

J.S. Bach: Fuge #16 in G minor



NMFの音響信号処理応用事例2

◆教師ありモノラル音源分離 [P. Smaragdis et al., 2007]

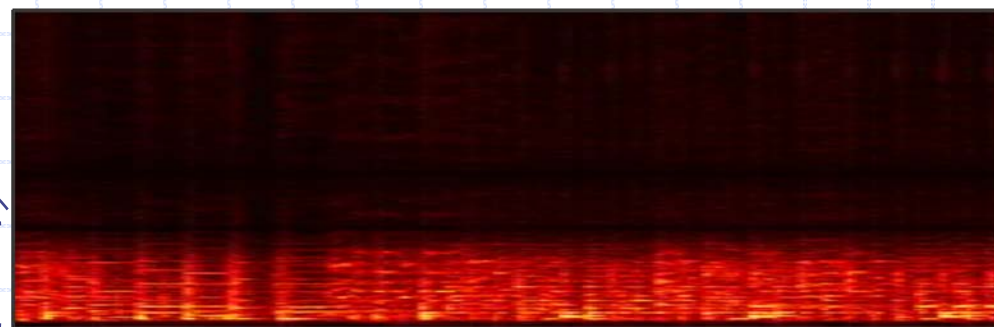
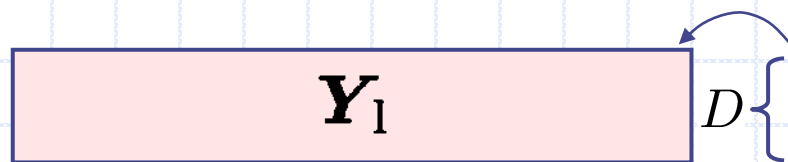


$$Y \simeq \underbrace{\begin{bmatrix} H_{\text{speech}} & H_{\text{chime}} \end{bmatrix}}_{\text{固定}} \begin{bmatrix} U_{\text{speech}} \\ U_{\text{chime}} \end{bmatrix}$$

NMFの音響信号処理応用事例3

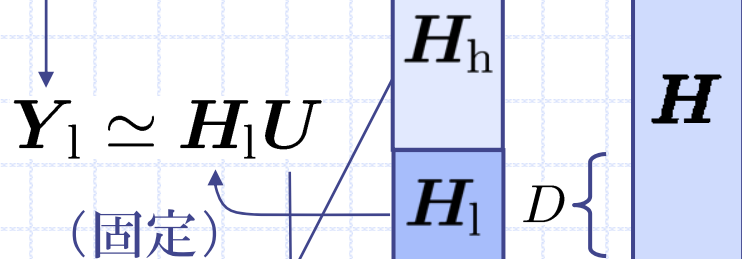
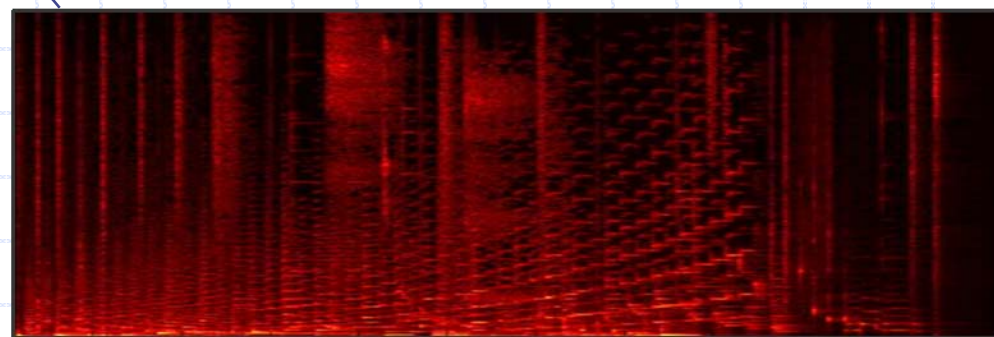
◆音の「超解像」

[P. Smaragdis & B. Raj, 2007]

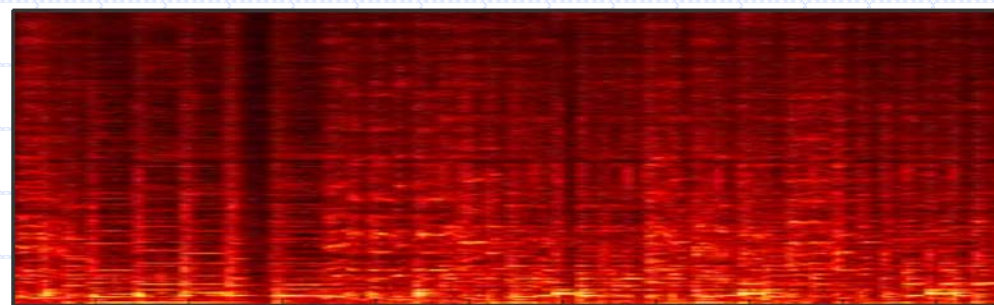


低サンプリングレートの信号

高サンプリングレート信号の学習データ



$$H_h U = Y_h$$



高周波帯域が復元された信号

NMF [D.D. Lee & H.S. Seung, 1999] の基礎

- ◆ T 個の観測データ(非負値ベクトル) $y_1, \dots, y_T$
- ◆ I 個の非負値基底ベクトル $h_1, \dots, h_I$ の非負結合でどの観測データも良く表現できる基底セットを求めたい

$$y_t \simeq \sum_{i=1}^I h_i u_{i,t}, \quad u_{i,t} \geq 0 \quad (t = 1, \dots, T)$$

- $I \leq \Omega$ のとき, 各基底ベクトルはいくつかの観測データで共起するパターン(パーツと呼ぶ)を表現するよう誘導される

$$\left. \begin{array}{l} Y := (y_1, \dots, y_T) \\ H := (h_1, \dots, h_I) \\ U := (u_{i,t})_{I \times T} \end{array} \right\} Y \simeq HU \quad H_{\omega,i} \geq 0, U_{i,t} \geq 0$$

minimize $D(Y, HU)$

subject to $H_{\omega,i} \geq 0, U_{i,t} \geq 0$

- ◆ NMFが想定する状況

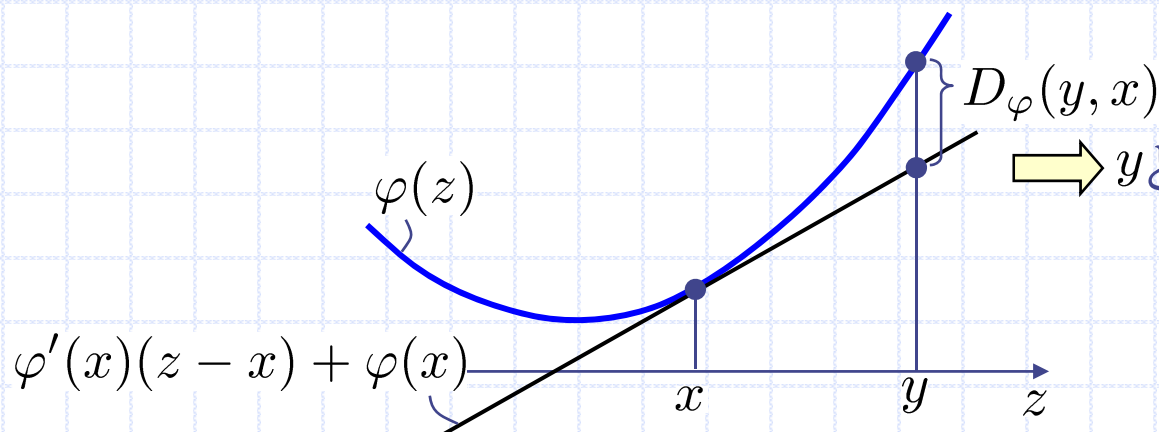
- データはすべて非負値
- 高々 I 種類のパーツだけですべての観測データは構成されている
- パーツの和で観測データをモデル化

NMFアルゴリズム：近さを測る

◆準備1. Bregmanダイバージェンス [L.M. Bregman, 1967]

$$D_{\varphi}(y, x) := \varphi(y) - \varphi(x) - \varphi'(x)(y - x)$$

$\varphi : \mathbb{R} \rightarrow \mathbb{R}$ は任意の微分可能な凸関数



⇒ y と x が近いほど小さい

ただし距離の公理
は満たさない(対称
性が成り立たない)

例) 二乗誤差, Iダイバージェンス, 板倉齋藤距離

- ◆ $\varphi(z) = z^2 \longrightarrow D_{\varphi}(y, x) = y^2 - x^2 - 2x(y - x) = (y - x)^2$
- ◆ $\varphi(z) = z \log z \longrightarrow D_{\varphi}(y, x) = y \log y - x \log x - (1 + \log x)(y - x) = y \log \frac{y}{x} - (y - x)$
- ◆ $\varphi(z) = -\log z \longrightarrow D_{\varphi}(y, x) = -\log y + \log x + \frac{1}{x}(y - x) = \frac{y}{x} - \log \frac{y}{x} - 1$

NMFアルゴリズム：Bregmanダイバージェンス最小化

[I.S. Dhillon & S. Sra, 2005]

$$\begin{aligned} &\text{minimize} && \sum_{\omega,t} D_{\varphi}(Y_{\omega,t}, \sum_i H_{\omega,i} U_{i,t}) \quad \text{or} \quad \sum_{\omega,t} D_{\varphi}(\sum_i H_{\omega,i} U_{i,t}, Y_{\omega,t}) \\ &\text{subject to} && H_{\omega,i} \geq 0, U_{i,t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T) \end{aligned}$$

◆ 補助関数法 [J. de Leeuw, 1994]

■ $D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta})$ を満たす $G(\theta, \bar{\theta})$ を補助関数と定義

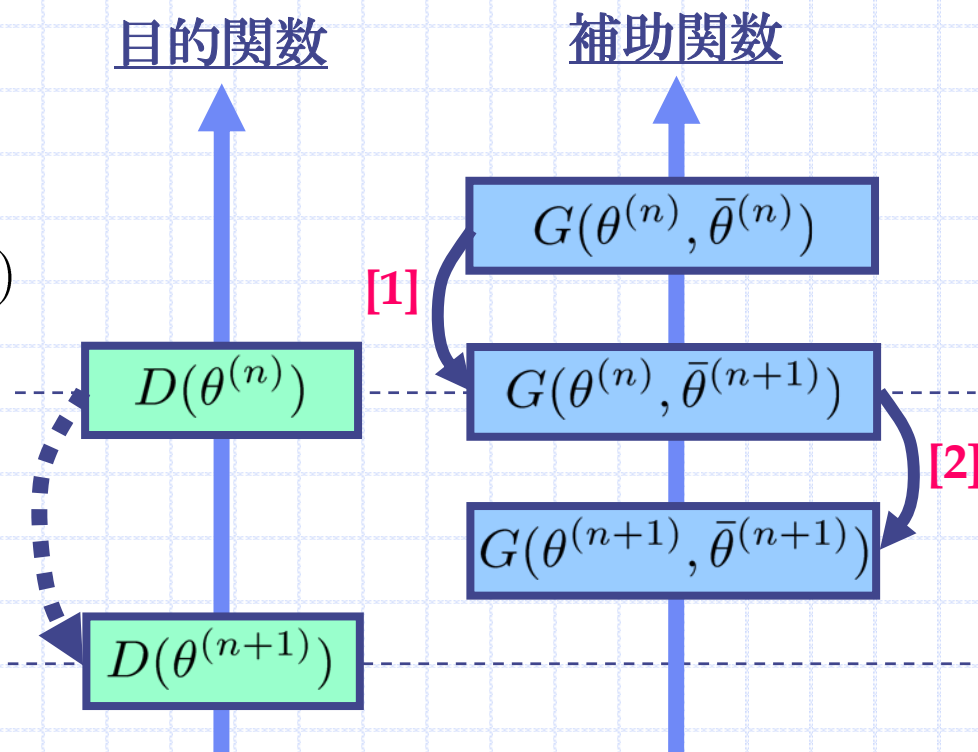
■ 反復アルゴリズム

[1] $\bar{\theta}^{(n+1)} = \operatorname{argmin}_{\bar{\theta}} G(\theta^{(n)}, \bar{\theta})$

[2] $\theta^{(n+1)} = \operatorname{argmin}_{\theta} G(\theta, \bar{\theta}^{(n+1)})$

■ 収束性

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



NMFアルゴリズム：Bregmanダイバージェンス最小化

[I.S. Dhillon & S. Sra, 2005]

$$\begin{aligned} &\text{minimize} \quad \sum_{\omega,t} D_{\varphi}(Y_{\omega,t}, \sum_i H_{\omega,i} U_{i,t}) \quad \text{or} \quad \sum_{\omega,t} D_{\varphi}(\sum_i H_{\omega,i} U_{i,t}, Y_{\omega,t}) \\ &\text{subject to} \quad H_{\omega,i} \geq 0, U_{i,t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T) \end{aligned}$$

タイプ1

◆ 目的関数 $\theta \equiv \{H, U\}$

φ : 凸関数

■ Bregmanダイバージェンス $D_{\varphi}(x, y) \equiv \varphi(x) - \varphi(y) - \varphi'(y)(x - y)$

$$D(\theta) = \sum_{\omega,t} \left[\varphi \left(\sum_i H_{\omega,i} U_{i,t} \right) - \varphi(Y_{\omega,t}) - \varphi'(Y_{\omega,t}) \left(\sum_i H_{\omega,i} U_{i,t} - Y_{\omega,t} \right) \right]$$

◆ 補助関数

■ Jensenの不等式 $\phi \left(\sum_i \lambda_i z_i \right) \leq \sum_i \lambda_i \phi(z_i) \quad \left\{ \begin{array}{l} \phi : \text{凸関数} \\ \sum_i \lambda_i = 1 \end{array} \right.$

$$\Rightarrow \varphi \left(\sum_i H_{\omega,i} U_{i,t} \right) = \varphi \left(\sum_i \lambda_{i,\omega,t} \frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right) \leq \sum_i \lambda_{i,\omega,t} \varphi \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right)$$

$$G(\theta, \bar{\theta}) = \sum_{\omega,t} \left[\sum_i \lambda_{i,\omega,t} \varphi \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right) - \varphi(Y_{\omega,t}) - \varphi'(Y_{\omega,t}) \left(\sum_i H_{\omega,i} U_{i,t} - Y_{\omega,t} \right) \right]$$

$\bar{\theta} \equiv \{\lambda\}$

NMFアルゴリズム: Bregmanダイバージェンス最小化

[I.S. Dhillon & S. Sra, 2005]

$$G(\theta, \bar{\theta}) = \sum_{\omega, t} \left[\sum_i \lambda_{i, \omega, t} \varphi \left(\frac{H_{\omega, i} U_{i, t}}{\lambda_{i, \omega, t}} \right) - \varphi(Y_{\omega, t}) - \varphi'(Y_{\omega, t}) \left(\sum_i H_{\omega, i} U_{i, t} - Y_{\omega, t} \right) \right]$$

◆Step1) $\bar{\theta}^{(n+1)} = \operatorname{argmin}_{\bar{\theta}} G(\theta^{(n)}, \bar{\theta})$

$$\lambda_{i, \omega, t}^{(n+1)} = \frac{H_{\omega, i}^{(n)} U_{i, t}^{(n)}}{\sum_i H_{\omega, i}^{(n)} U_{i, t}^{(n)}}$$

◆Step2) $\theta^{(n+1)} = \operatorname{argmin}_{\theta} G(\theta, \bar{\theta}^{(n+1)})$ $\begin{cases} \mathbf{H}^{(n+1)} = \operatorname{argmin}_{\mathbf{H}} G(\{\mathbf{H}, \mathbf{U}^{(n)}\}, \bar{\theta}^{(n+1)}) \\ \mathbf{U}^{(n+1)} = \operatorname{argmin}_{\mathbf{U}} G(\{\mathbf{H}^{(n)}, \mathbf{U}\}, \bar{\theta}^{(n+1)}) \end{cases}$

$$\frac{\partial G}{\partial H_{\omega, i}} = \sum_t \left[\lambda_{i, \omega, t}^{(n+1)} \varphi' \left(\frac{H_{\omega, i} U_{i, t}^{(n)}}{\lambda_{i, \omega, t}^{(n+1)}} \right) \frac{U_{i, t}^{(n)}}{\lambda_{i, \omega, t}^{(n+1)}} - \varphi'(Y_{\omega, t}) U_{i, t}^{(n)} \right] = 0$$

$$\frac{\partial G}{\partial U_{i, t}} = \sum_t \left[\lambda_{i, \omega, t}^{(n+1)} \varphi' \left(\frac{H_{\omega, i} U_{i, t}^{(n)}}{\lambda_{i, \omega, t}^{(n+1)}} \right) \frac{H_{\omega, i}^{(n)}}{\lambda_{i, \omega, t}^{(n+1)}} - \varphi'(Y_{\omega, t}) H_{\omega, i}^{(n)} \right] = 0$$

を解きたい ...が...一般の φ では解けない...

NMFアルゴリズム：Bregmanダイバージェンス最小化

[I.S. Dhillon & S. Sra, 2005]

$$\frac{\partial G}{\partial H_{\omega,i}} = \sum_t \left[\cancel{\lambda_{i,\omega,t}^{(n+1)}} \varphi' \left(\frac{H_{\omega,i} U_{i,t}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} \right) \frac{U_{i,t}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} - \varphi'(Y_{\omega,t}) U_{i,t}^{(n)} \right] = 0$$

$$\frac{\partial G}{\partial U_{i,t}} = \sum_t \left[\cancel{\lambda_{i,\omega,t}^{(n+1)}} \varphi' \left(\frac{H_{\omega,i} U_{i,t}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} \right) \frac{H_{\omega,i}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} - \varphi'(Y_{\omega,t}) H_{\omega,i}^{(n)} \right] = 0$$

◆ φ のクラスを制限 $\varphi'(xy) = \varphi'(x)\varphi'(y)/\varphi'(1)$ $\varphi(z) = z^2$ はこのタイプ

$$\frac{\partial G}{\partial H_{\omega,i}} = \sum_t \left[\varphi' \left(\frac{H_{\omega,i}}{H_{\omega,i}^{(n)}} \right) \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) U_{i,t}^{(n)} / \varphi'(1) - \varphi'(Y_{\omega,t}) U_{i,t}^{(n)} \right] = 0$$

$$\Leftrightarrow \varphi' \left(\frac{H_{\omega,i}}{H_{\omega,i}^{(n)}} \right) \sum_t \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) U_{i,t}^{(n)} = \varphi'(1) \sum_t \varphi'(Y_{\omega,t}) U_{i,t}^{(n)}$$

$$H_{\omega,i} = H_{\omega,i}^{(n)} \cdot \varphi'^{-1} \left(\frac{\varphi'(1) \sum_t \varphi'(Y_{\omega,t}) U_{i,t}^{(n)}}{\sum_t \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) U_{i,t}^{(n)}} \right)$$

$\varphi' : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$

ならば、非負値から
非負値へ更新

$$\lambda_{i,\omega,t}^{(n+1)} = \frac{H_{\omega,i}^{(n)} U_{i,t}^{(n)}}{\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)}}$$

具体的に代入してみると...

$$\left. \begin{array}{l} \varphi(z) = z^2 \\ \varphi'(z) = 2z \\ \varphi'^{-1}(z) = z/2 \end{array} \right\} H_{\omega,i} = H_{\omega,i}^{(n)} \cdot \left(\frac{\sum_t Y_{\omega,t} U_{i,t}^{(n)}}{\sum_t \sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} U_{i,t}^{(n)}} \right)$$

二乗誤差(Frobeniusノルム)
規準のNMF更新公式

NMFアルゴリズム：Bregmanダイバージェンス最小化

[I.S. Dhillon & S. Sra, 2005]

$$\frac{\partial G}{\partial H_{\omega,i}} = \sum_t \left[\cancel{\lambda_{i,\omega,t}^{(n+1)}} \varphi' \left(\frac{H_{\omega,i} U_{i,t}^{(n)}}{\lambda_{i,\omega,t}^{(n+1)}} \right) \frac{U_{i,t}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} - \varphi'(Y_{\omega,t}) U_{i,t}^{(n)} \right] = 0$$

$$\frac{\partial G}{\partial U_{i,t}} = \sum_t \left[\cancel{\lambda_{i,\omega,t}^{(n+1)}} \varphi' \left(\frac{H_{\omega,i} U_{i,t}^{(n)}}{\lambda_{i,\omega,t}^{(n+1)}} \right) \frac{H_{\omega,i}^{(n)}}{\cancel{\lambda_{i,\omega,t}^{(n+1)}}} - \varphi'(Y_{\omega,t}) H_{\omega,i}^{(n)} \right] = 0$$

◆ φ のクラスを制限

$\varphi'(xy) = \varphi'(x) + \varphi'(y) + \varphi'(1)$ $\varphi(z) = z \log z, -\log z$ はこのタイプ

$$\frac{\partial G}{\partial H_{\omega,i}} = \sum_t \left[\left\{ \varphi' \left(\frac{H_{\omega,i}}{H_{\omega,i}^{(n)}} \right) + \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) + \varphi'(1) \right\} U_{i,t}^{(n)} - \varphi'(Y_{\omega,t}) U_{i,t}^{(n)} \right] = 0$$

$$\Leftrightarrow \varphi' \left(\frac{H_{\omega,i}}{H_{\omega,i}^{(n)}} \right) \sum_t U_{i,t}^{(n)} = \sum_t \left[\varphi'(Y_{\omega,t}) - \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) - \varphi'(1) \right] U_{i,t}^{(n)}$$

$$H_{\omega,i} = H_{\omega,i}^{(n)} \cdot \varphi'^{-1} \left(\frac{\sum_t \left[\varphi'(Y_{\omega,t}) - \varphi' \left(\sum_i H_{\omega,i}^{(n)} U_{i,t}^{(n)} \right) - \varphi'(1) \right] U_{i,t}^{(n)}}{\sum_t U_{i,t}^{(n)}} \right)$$

NMFアルゴリズム：Bregmanダイバージェンス最小化

タイプ2

$$\text{minimize} \quad \sum_{\omega,t} D_{\varphi} (Y_{\omega,t}, \sum_i H_{\omega,i} U_{i,t}) \quad \text{or} \quad \sum_{\omega,t} D_{\varphi} (\sum_i H_{\omega,i} U_{i,t}, Y_{\omega,t})$$

$$\text{subject to} \quad H_{\omega,i} \geq 0, U_{i,t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T)$$

◆一般の φ に対する公式はまだ知られていない

➔それぞれの φ についての更新公式は個別に提案されている：

$$\varphi(z) = z^2 \quad : [\text{D.D. Lee \& H.S. Seung, 2000}]$$

$$\varphi(z) = z \log z \quad : [\text{D.D. Lee \& H.S. Seung, 2000}]$$

$$\varphi(z) = -\log z \quad : [\text{H. Kameoka et al., 2006}][\text{C. Févotte et al., 2008}]$$

Bregmanダイバージェンスと対応する距離尺度

y : 観測値

$$\min \begin{cases} D_{\varphi}(y, x) \\ D_{\varphi}(x, y) \end{cases}$$

[I.S. Dhillon & S. Sra, 2005]

	$D_{\varphi}(y, x)$	$D_{\varphi}(x, y)$
$\varphi(z) = z^2$	二乗誤差 $(y - x)^2$	二乗誤差 $(x - y)^2$
$\varphi(z) = z \log z$	Iダイバージェンス $y \log \frac{y}{x} - (y - x)$	双対Iダイバージェンス $x \log \frac{x}{y} - (x - y)$
$\varphi(z) = -\log z$	板倉斎藤距離 $\frac{y}{x} - \log \frac{y}{x} - 1$	双対板倉斎藤距離 $\frac{x}{y} - \log \frac{x}{y} - 1$

Bregmanダイバージェンスと対応する尤度関数

y : 観測値

$$\min \begin{cases} D_{\varphi}(y, x) \\ D_{\varphi}(x, y) \end{cases} \Leftrightarrow \max P(y|x)$$

	$D_{\varphi}(y, x)$	$D_{\varphi}(x, y)$
$\varphi(z) = z^2$	Gauss分布 $y x \sim \mathcal{N}(x, \sigma^2)$	Gauss分布 $y x \sim \mathcal{N}(x, \sigma^2)$
$\varphi(z) = z \log z$	Poisson分布 $y x \sim \text{Pois}(x)$	Gamma分布 $y x \sim \text{Gamma}(1, x + 1)$
$\varphi(z) = -\log z$	指数分布 $y x \sim \text{Exp}(1/x)$	逆Gamma分布 $y x \sim \text{InvGamma}(0, x)$

NMFアルゴリズム：近さを測る

◆準備2. β ダイバージェンス [S. Eguchi & Y. Kano, 2001]

$$D_{\beta}(y, x) := \frac{y^{\beta}}{\beta(\beta-1)} + \frac{x^{\beta}}{\beta} - \frac{yx^{\beta-1}}{\beta-1}$$

Bregmanダイバージェンス
同様、距離の公理を満たさない
(対称性が成り立たない)

例) 二乗誤差, Iダイバージェンス, 板倉齋藤距離

$$\diamond \beta = 2 \quad D_{\beta}(y, x) = \frac{y^2}{2} + \frac{x^2}{2} - yx = \frac{1}{2}(y-x)^2$$

$$\diamond \beta \rightarrow 1 \quad D_{\beta}(y, x) \rightarrow y \log \frac{y}{x} + x \quad \because \lim_{\beta \rightarrow 1} \frac{yx^{\beta-1}}{\beta-1} = \lim_{\beta \rightarrow 1} yx^{\beta-1} \log x = y \log x$$

$$\diamond \beta \rightarrow 0 \quad \mathbf{D_{\beta}(y, x) \rightarrow \frac{y}{x} - \log \frac{y}{x}} \quad \because \lim_{\beta \rightarrow 0} \frac{x^{\beta}}{\beta} = \lim_{\beta \rightarrow 0} x^{\beta} \log x = \log x$$

NMFアルゴリズム： β ダイバージェンス最小化

[M. Nakano et al., 2010]

$$\begin{aligned} &\text{minimize} && \sum_{\omega,t} D_{\beta} \left(Y_{\omega,t}, \sum_i H_{\omega,i} U_{i,t} \right) \\ &\text{subject to} && H_{\omega,i} \geq 0, U_{i,t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T) \end{aligned}$$

◆ 補助関数法 [J. de Leeuw, 1994]

■ $D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta})$ を満たす $G(\theta, \bar{\theta})$ を補助関数と定義

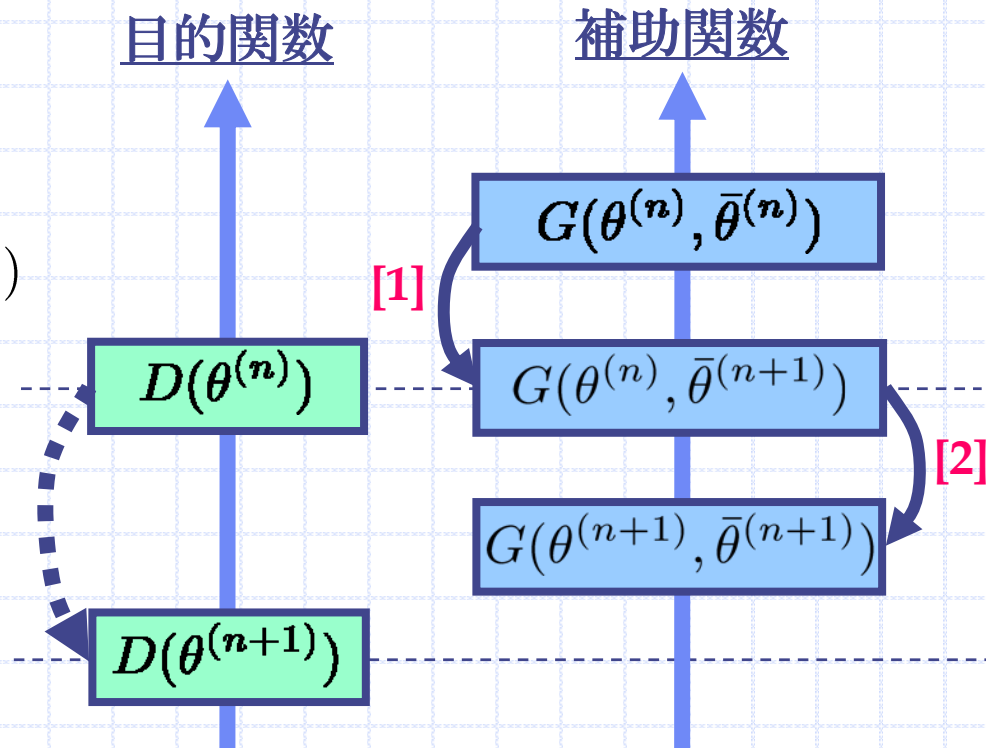
■ 反復アルゴリズム

[1] $\bar{\theta}^{(n+1)} = \operatorname{argmin}_{\bar{\theta}} G(\theta^{(n)}, \bar{\theta})$

[2] $\theta^{(n+1)} = \operatorname{argmin}_{\theta} G(\theta, \bar{\theta}^{(n+1)})$

■ 収束性

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



NMFアルゴリズム： β ダイバージェンス最小化

[M. Nakano et al., 2010]

$$\text{minimize} \quad \sum_{\omega, t} D_{\beta} \left(Y_{\omega, t}, \sum_i H_{\omega, i} U_{i, t} \right)$$

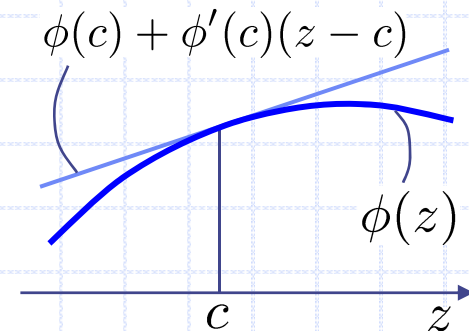
$$\text{subject to} \quad H_{\omega, i} \geq 0, U_{i, t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T)$$

◆ 目的関数 $\theta := \{H, U\}$

$$D(\theta) = \sum_{\omega, t} \left[\frac{Y_{\omega, t}^{\beta}}{\beta(\beta-1)} + \frac{1}{\beta} \underbrace{\left(\sum_i H_{\omega, i} U_{i, t} \right)^{\beta}}_{\text{凸/凹}} - \frac{Y_{\omega, t}}{\beta-1} \underbrace{\left(\sum_i H_{\omega, i} U_{i, t} \right)^{\beta-1}}_{\text{凸/凹}} \right]$$

◆ 補助関数の設計方針

- 凸不等式 $\phi \left(\sum_i \lambda_i z_i \right) \leq \sum_i \lambda_i \phi(z_i) \quad \left\{ \begin{array}{l} \phi : \text{凸関数} \\ \sum_i \lambda_i = 1 \end{array} \right.$
- 接線不等式 $\phi(z) \leq \phi(c) + \phi'(c)(z - c) \quad \phi : \text{凹関数}$



➡ 凸関数項の上限を凸不等式，凹関数項の上限を接線不等式を使って設計し，補助関数を構成 [H. Kameoka et al., 2006]

NMFアルゴリズム： β ダイバージェンス最小化

[M. Nakano et al., 2010]

◆補助関数

$$\frac{1}{\beta} \left(\underbrace{\sum_i H_{\omega,i} U_{i,t}}_{\geq 0} \right)^{\beta} \leq \begin{cases} \text{凸} & \frac{1}{\beta} \sum_i \lambda_{i,\omega,t} \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right)^{\beta} & (\beta \geq 1) \\ \text{凹} & \frac{1}{\beta} c_{\omega,t}^{\beta} + c_{\omega,t}^{\beta-1} \left(\sum_i H_{\omega,i} U_{i,t} - c_{\omega,t} \right) & (\beta \leq 1, \beta \neq 0) \end{cases}$$

$$-\frac{1}{\beta-1} \left(\sum_i H_{\omega,i} U_{i,t} \right)^{\beta-1} \leq \begin{cases} \text{凸} & -\frac{1}{\beta-1} \sum_i \lambda_{i,\omega,t} \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right)^{\beta-1} & (\beta \leq 2, \beta \neq 1) \\ \text{凹} & -\frac{1}{\beta-1} c_{\omega,t}^{\beta-1} - c_{\omega,t}^{\beta-2} \left(\sum_i H_{\omega,i} U_{i,t} - c_{\omega,t} \right) & (\beta \geq 2) \end{cases}$$

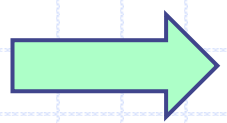
(例) $0 < \beta < 1$ のケース:

ただし, $\sum_i \lambda_{i,\omega,t} = 1$

$$G(\theta, \bar{\theta}) = \sum_{\omega,t} \left[\frac{Y_{\omega,t}^{\beta}}{\beta(\beta-1)} + \frac{1}{\beta} c_{\omega,t}^{\beta} + c_{\omega,t}^{\beta-1} \left(\sum_i H_{\omega,i} U_{i,t} - c_{\omega,t} \right) - \frac{Y_{\omega,t}}{\beta-1} \sum_i \lambda_{i,\omega,t} \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right)^{\beta-1} \right]$$

subject to $\sum_i \lambda_{i,\omega,t} = 1$

$\{c_{\omega,t}\}, \{\lambda_{i,\omega,t}\}$



$$D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta}) \quad \begin{cases} \lambda_{i,\omega,t}^{(n+1)} = \operatorname{argmin}_{\lambda_{i,\omega,t}} G(\theta^{(n)}, \bar{\theta}) = \frac{H_{\omega,i}^{(n)}}{\sum_{i'} H_{\omega,i'}^{(n)} U_{i',t}^{(n)}} \\ c_{\omega,t}^{(n+1)} = \operatorname{argmin}_{c_{\omega,t}} G(\theta^{(n)}, \bar{\theta}) = \sum_{i'} H_{\omega,i'}^{(n)} U_{i',t}^{(n)} \end{cases}$$

NMFアルゴリズム： β ダイバージェンス最小化

[M. Nakano et al., 2010]

◆補助関数 (例) $0 < \beta < 1$ のケース:

$$G(\theta, \bar{\theta}) = \sum_{\omega, t} \left[\frac{Y_{\omega, t}^{\beta}}{\beta(\beta - 1)} + \frac{1}{\beta} c_{\omega, t}^{\beta} + c_{\omega, t}^{\beta-1} \left(\sum_i H_{\omega, i} U_{i, t} - c_{\omega, t} \right) - \frac{Y_{\omega, t}}{\beta - 1} \sum_i \lambda_{i, \omega, t} \left(\frac{H_{\omega, i} U_{i, t}}{\lambda_{i, \omega, t}} \right)^{\beta-1} \right]$$

$$H_{\omega, i}^{(n+1)} = \operatorname{argmin}_{H_{\omega, i}} G(\theta, \bar{\theta}^{(n+1)})$$

$$U_{i, t}^{(n+1)} = \operatorname{argmin}_{U_{i, t}} G(\theta, \bar{\theta}^{(n+1)})$$

$$\left\{ \begin{array}{l} \frac{\partial G(\theta, \bar{\theta})}{\partial H_{\omega, i}} = \sum_t \left[c_{\omega, t}^{(n+1)\beta-1} U_{i, t}^{(n)} - \frac{Y_{\omega, t} U_{i, t}^{(n)\beta-1}}{\lambda_{i, \omega, t}^{(n+1)\beta-2}} H_{\omega, i}^{\beta-2} \right] = 0 \\ \frac{\partial G(\theta, \bar{\theta})}{\partial U_{i, t}} = \sum_{\omega} \left[c_{\omega, t}^{(n+1)\beta-1} H_{\omega, i}^{(n)} - \frac{Y_{\omega, t} H_{\omega, i}^{(n)\beta-1}}{\lambda_{i, \omega, t}^{(n+1)\beta-2}} U_{i, t}^{\beta-2} \right] = 0 \end{array} \right.$$

$$H_{\omega, i}^{(n+1)} = \left(\frac{\sum_t c_{\omega, t}^{(n+1)\beta-1} U_{i, t}^{(n)}}{\sum_t Y_{\omega, t} U_{i, t}^{(n)\beta-1} / \lambda_{i, \omega, t}^{(n+1)\beta-2}} \right)^{\beta-2} \xrightarrow{\beta \rightarrow 1} H_{\omega, i}^{(n+1)} = \frac{\sum_t Y_{\omega, t} \lambda_{i, \omega, t}^{(n+1)}}{\sum_t U_{i, t}^{(n)}}$$

$$U_{i, t}^{(n+1)} = \left(\frac{\sum_{\omega} c_{\omega, t}^{(n+1)\beta-1} H_{\omega, i}^{(n)}}{\sum_{\omega} Y_{\omega, t} H_{\omega, i}^{(n)\beta-1} / \lambda_{i, \omega, t}^{(n+1)\beta-2}} \right)^{\beta-2} \xrightarrow{\beta \rightarrow 1} U_{i, t}^{(n+1)} = \frac{\sum_{\omega} Y_{\omega, t} \lambda_{i, \omega, t}^{(n+1)}}{\sum_{\omega} H_{\omega, i}^{(n)}}$$

NMF と Sparse coding [B.A. Olshausen, 1996] との関係

- ◆ T 個の観測データ(実数値ベクトル) $y_1, \dots, y_T$
- ◆ すべての観測データが I 個の実数値基底ベクトル $a_1, \dots, a_I$ の”スパース”な線形結合で表される基底セットを求めたい

■ exactなreconstruction

$$\begin{array}{ll} \text{minimize} & S(U) \\ \text{subject to} & Y = AU \end{array}$$

NMFでは副次効果的に
スパースになっていたのに
対し, sparse codingで
はスパース性コストにより
強制的にスパース化

■ approximation

$$\begin{array}{ll} \text{minimize} & \|Y - AU\|_F^2 + \lambda S(U) \\ \text{subject to} & \|U\|_F^2 = 1 \end{array}$$

勾配法等

- ◆ スパース性コスト [B.A. Olshausen, 1996][G. Harpur, 2000]

■ L_p ノルム $S(U) = \sum_{i,t} |U_{i,t}|^p$

$$S(U) = \sum_{i,t} \log(1 + U_{i,t}^2) \quad \text{尖度} \quad S(U) = \sum_i \left(\frac{\langle (U_{i,t} - \mu_i)^4 \rangle_t}{\sigma_t^2} - 3 \right)$$

NMF と pLSA [T. Hofmann, 1999] との関係

(probabilistic Latent Semantic Analysis)

◆ Aspectモデル

$$P(\omega, t) = \sum_c P(c) P(\omega, t|c) = \sum_c P(c) P(\omega|c) P(t|c) = P(t) \sum_c P(\omega|c) P(c|t)$$

word document latent class

◆ ドキュメント t に含まれる単語 ω の総数: $N(\omega, t)$

➔ 最尤推定によるAspectモデルの”フィッティング”

■ 対数尤度 $\log \prod_{\omega, t} P(\omega, t)^{N(\omega, t)}$ を最大化する $P(t), P(\omega|c), P(c|t)$ 推定

$$\text{maximize} \quad \sum_{\omega, t} N(\omega, t) \log P(t) \sum_c P(\omega|c) P(c|t)$$

$$\text{subject to} \quad P(t) \geq 0, P(\omega|c) \geq 0, P(c|t) \geq 0$$

$$\sum_t P(t) = 1, \sum_{\omega} P(\omega|c) = 1, \sum_c P(c|t) = 1$$

$N(\omega, t)$ と $P(\omega, t)$
のIダイバージェンス
最小化と同じ意味

複素NMF

NMFによる振幅スペクトログラムの分解表現の問題点

観測振幅スペクトルを振幅スペクトルパーツの和でモデル化

😞 振幅スペクトルの加法性を仮定した不正確なモデル

- 振幅スペクトルは本当は非加法的（加法的なのは波形！）
- ➔ モデル自体に誤差を含んでいる

😞 位相情報の放棄

- 推定したモデルが波形に対応しない
- 波形領域で実現される多くの信号処理法と組み合わせられない

NMFモデルが
仮定される世界



周波数成分の
含有率の情報

波形の世界
(豊富な情報)
|||
複素スペクトルの世界

広い
応用展開

位相を捨てるため元に戻せない！

振幅スペクトルの非加法性について

◆ 音響信号 $y(u) \in \mathbb{R}$ $\longrightarrow$ $y(u) = \sum_i x^i(u)$ (波形同士は加法的)

↓ 短時間フーリエ変換
(時間周波数分解) $\cdots$ 線形な変換

$$Y_{\omega,t} = \int W(u)y(u+t)e^{j\omega u}du \in \mathbb{C} \longrightarrow Y_{\omega,t} = \sum_i X_{\omega,t}^i \quad (\text{もちろん加法的})$$

↓ 絶対値をとる $\cdots$ 非線形な変換

$$|Y_{\omega,t}| : \text{時刻 } t \text{ に周波数 } \omega \text{ の成分が} \longrightarrow |Y_{\omega,t}| = \left| \sum_i X_{\omega,t}^i \right|$$

どれほど含まれているか

$$\mathbf{Y} \simeq \mathbf{H}\mathbf{U} \quad (\mathbf{H} \succeq 0, \mathbf{U} \succeq 0)$$

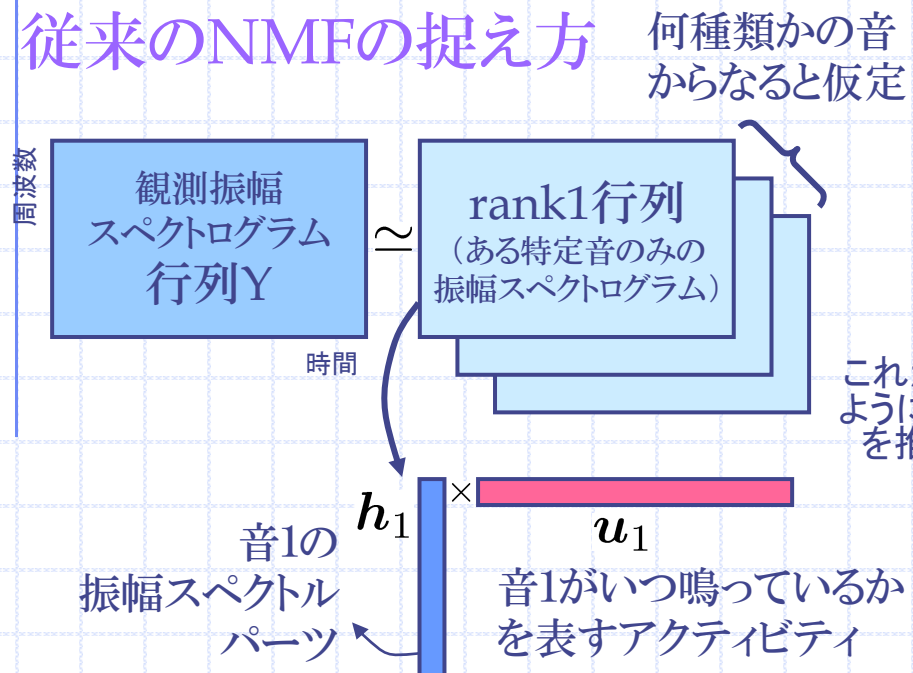
NMFモデル

$$\neq \sum_i |X_{\omega,t}^i| \quad (\text{振幅スペクトル} \\ \text{同士は非加法的})$$

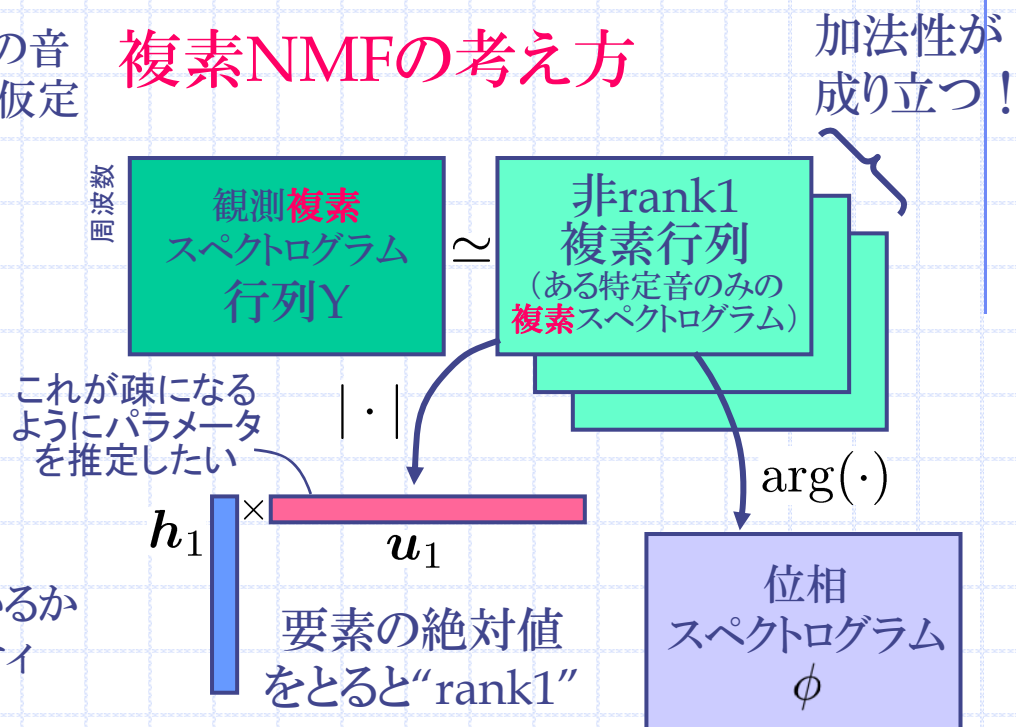
$$|Y_{\omega,t}| = \sum_i \underbrace{H_{\omega}^i U_t^i}_{\text{振幅スペクトル?}}$$

新モデル「複素NMF」の提案

従来のNMFの捉え方



複素NMFの考え方



◆複素NMFモデル

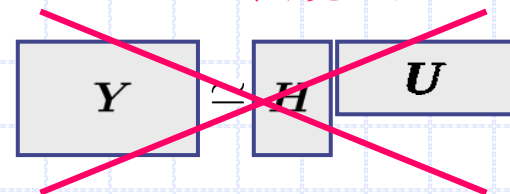
■複素スペクトログラム $Y_{\omega,t} \in \mathbb{C}$ をモデル化

$$Y_{\omega,t} = \sum_i X_{\omega,t}^i \quad |X_{\omega,t}^i| = H_{\omega}^i U_t^i$$

$$= \sum_i \underline{|X_{\omega,t}^i|} e^{j\phi_{\omega,t}^i} = \sum_i \underline{H_{\omega}^i U_t^i} e^{j\phi_{\omega,t}^i}$$

$$H_{\omega}^i \geq 0, U_t^i \geq 0$$

行列積の形にならない！
(新しいクラスのスパース表現モデル)



複素NMFアルゴリズム

◆定義

- 複素スペクトログラム $Y_{\omega,t} \in \mathbb{C}$
- 振幅スペクトル基底 $\mathcal{H} \equiv \{H_{\omega}^i\}$
- 位相スペクトログラム $\phi \equiv \{\phi_{\omega,t}^i\}$
- ゲイン $\mathcal{U} \equiv \{U_t^i\}$

音響信号モデル

$$F_{\omega,t} = \sum_{i=1}^I H_{\omega}^i U_t^i e^{j\phi_{\omega,t}^i}$$

◆最適化問題

$$Y_{\omega,t} = F_{\omega,t} + \epsilon_{\omega,t} \begin{cases} \epsilon_{\omega,t} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2) \\ U_t^i \sim \mathcal{GGD}(0, b, p) \quad (0 < p < 2) \end{cases}$$

↪ 一般化正規分布

$$\begin{aligned} &\text{minimize} && f(\mathcal{H}, \mathcal{U}, \phi) := \sum_{\omega,t} |Y_{\omega,t} - F_{\omega,t}|^2 + 2\lambda \sum_{i,t} |U_t^i|^p \\ &\text{subject to} && \sum_{\omega} H_{\omega}^i = 1 \quad (i = 1, \dots, I) \end{aligned}$$

スプース正則化項

複素NMFアルゴリズム

◆ 補助関数法

$$\begin{aligned}
 f(\mathcal{H}, \mathcal{U}, \phi) &= \sum_{\omega, t} |Y_{\omega, t} - F_{\omega, t}|^2 + 2\lambda \sum_{i, t} |U_t^i|^p \\
 &\leq \sum_{\omega, t} \sum_i \frac{|\bar{Y}_{\omega, t}^i - H_{\omega}^i U_t^i e^{j\phi_{\omega, t}^i}|^2}{\beta_{\omega, t}^i} + \lambda \sum_{i, t} (p|\bar{U}_t^i|^{p-2} U_t^{i2} + (2-p)|\bar{U}_t^i|^p) \\
 &\quad \text{subject to } \sum_i \bar{Y}_{\omega, t}^i = Y_{\omega, t} \\
 &\quad \quad \quad \sum_i \beta_{\omega, t}^i = 1 \\
 &\quad \quad \quad 0 < \beta_{\omega, t}^i < 1, \sum_i \beta_{\omega, t}^i = 1 \text{ を満たす任意の定数} \\
 &=: g(\mathcal{H}, \mathcal{U}, \phi, \bar{Y}, \bar{U})
 \end{aligned}$$

ℓ 回目の反復計算後のパラメータ値: $\mathcal{H}^{(\ell)} = \{H_{\omega}^{i(\ell)}\}$, $\mathcal{U}^{(\ell)} = \{U_t^{i(\ell)}\}$, $\phi^{(\ell)} = \{\phi_{\omega, t}^{i(\ell)}\}$

$$\left. \begin{aligned}
 \text{Step 1)} \quad &\bar{Y}_{\omega, t}^{i(\ell+1)} \leftarrow H_{\omega}^{i(\ell)} U_t^{i(\ell)} e^{j\phi_{\omega, t}^{i(\ell)}} + \beta_{\omega, t}^i (Y_{\omega, t} - F_{\omega, t}^{(\ell)}) \\
 &\bar{U}_t^{i(\ell+1)} \leftarrow U_t^{i(\ell)} \\
 \text{Step 2)} \quad &H_{\omega}^{i(\ell+1)} \leftarrow \frac{\sum_t U_t^{i(\ell)} \text{Re}[\bar{Y}_{\omega, t}^{i(\ell+1)*} e^{j\phi_{\omega, t}^{i(\ell)}}] / \beta_{\omega, t}^i}{\sum_t U_t^{i(\ell)2} / \beta_{\omega, t}^i} \\
 &U_t^{i(\ell+1)} \leftarrow \frac{\sum_t H_{\omega}^{i(\ell)} \text{Re}[\bar{Y}_{\omega, t}^{i(\ell+1)*} e^{j\phi_{\omega, t}^{i(\ell)}}] / \beta_{\omega, t}^i}{\sum_t U_t^{i(\ell)2} / \beta_{\omega, t}^i + \lambda p |\bar{U}_t^{i(\ell+1)}|^{p-2}} \\
 &e^{j\phi_{\omega, t}^{i(\ell+1)}} \leftarrow \bar{Y}_{\omega, t}^{i(\ell+1)} / |\bar{Y}_{\omega, t}^{i(\ell+1)}|
 \end{aligned} \right\} \begin{aligned}
 &f(\mathcal{H}^{(\ell)}, \mathcal{U}^{(\ell)}, \phi^{(\ell)}) \\
 &= g(\mathcal{H}^{(\ell)}, \mathcal{U}^{(\ell)}, \phi^{(\ell)}, \bar{Y}^{(\ell+1)}, \bar{U}^{(\ell+1)}) \\
 &\geq g(\mathcal{H}^{(\ell+1)}, \mathcal{U}^{(\ell+1)}, \phi^{(\ell+1)}, \bar{Y}^{(\ell+1)}, \bar{U}^{(\ell+1)}) \\
 &\geq f(\mathcal{H}^{(\ell+1)}, \mathcal{U}^{(\ell+1)}, \phi^{(\ell+1)})
 \end{aligned}$$

f は増加しない!

NMFと等価となる条件 複素NMFはNMFを包含

Step 1) $\bar{Y}_{\omega,t}^i \leftarrow H_{\omega}^i U_t^i e^{j\phi_{\omega,t}^i} + \beta_{\omega,t}^i (Y_{\omega,t} - F_{\omega,t})$

$$\bar{U}_t^i \leftarrow U_t^i$$

Step 2) $H_{\omega}^i \leftarrow \frac{\sum_t U_t^i \operatorname{Re}[\bar{Y}_{\omega,t}^{i*} e^{j\phi_{\omega,t}^i}]/\beta_{\omega,t}^i}{\sum_t U_t^{i2}/\beta_{\omega,t}^i}$

$$U_t^i \leftarrow \frac{\sum_t H_{\omega}^i \operatorname{Re}[\bar{Y}_{\omega,t}^{i*} e^{j\phi_{\omega,t}^i}]/\beta_{\omega,t}^i}{\sum_t U_t^{i2}/\beta_{\omega,t}^i + \lambda p |\bar{U}_t^i|^{p-2}}$$

$$e^{j\phi_{\omega,t}^i} \leftarrow \bar{Y}_{\omega,t}^i / |\bar{Y}_{\omega,t}^i|$$

[条件2]

Step 3) $\beta_{\omega,t}^i \leftarrow \frac{H_{\omega}^i U_t^i}{\sum_k H_{\omega}^k U_t^k}$ を実行

[条件1]

$\phi_{\omega,t}^i$ を $\arg(Y_{\omega,t})$ に
初期設定する

$\phi_{\omega,t}^i = \arg(Y_{\omega,t})$ は
Step1, Step2に対して
不動点になっている

$\beta_{\omega,t}^i : 0 < \beta_{\omega,t}^i < 1, \sum_i \beta_{\omega,t}^i = 1$
を満たす任意の定数

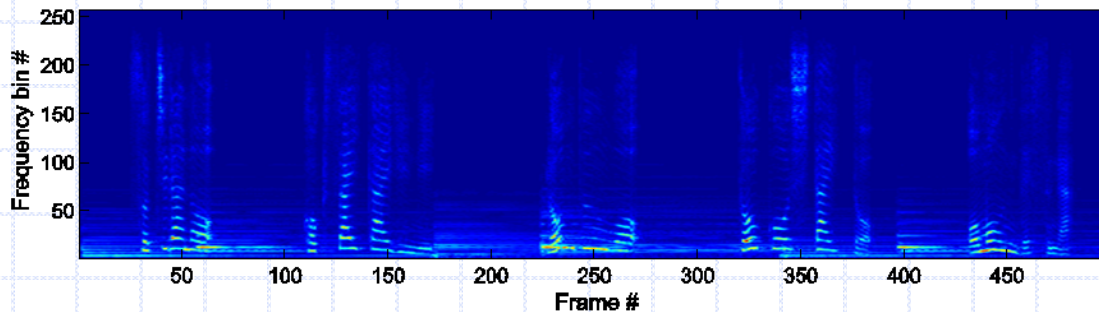
$$H_{\omega}^i \leftarrow H_{\omega}^i \frac{\sum_t U_t^i |Y_{\omega,t}|}{\sum_t U_t^i \sum_k H_{\omega}^k U_t^k}$$

$$U_t^i \leftarrow U_t^i \frac{\sum_{\omega} H_{\omega}^i |Y_{\omega,t}|}{\sum_{\omega} H_{\omega}^i \sum_k H_{\omega}^k U_t^k + \lambda p |U_t^i|^{p-2}}$$

Lee & Seungの二乗誤差規準のNMFアルゴリズム[Lee & Seung, 2000]と等価!

複素NMFによる音声と音楽の混合信号の分離

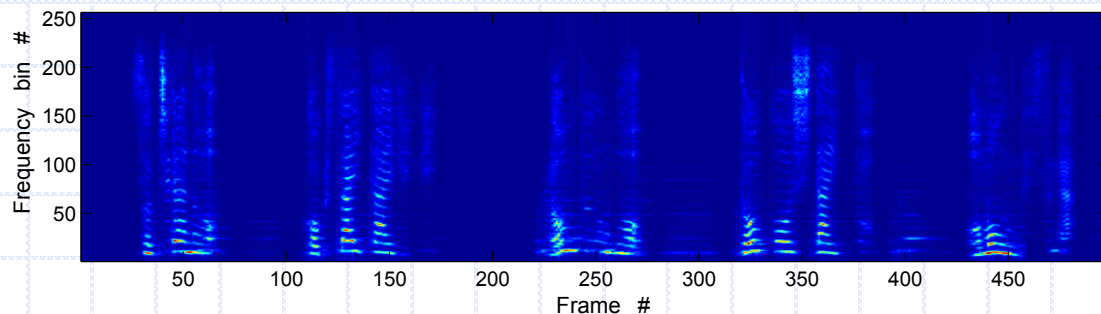
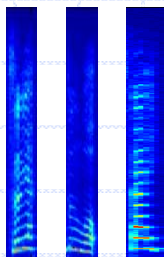
混合信号(音声+音楽)



|

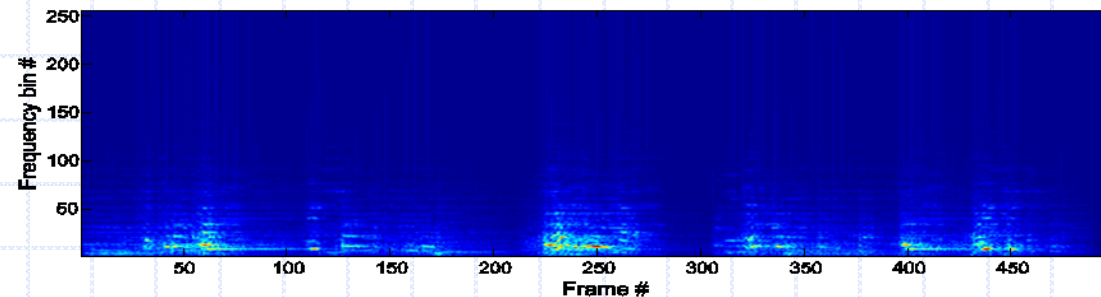
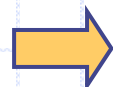
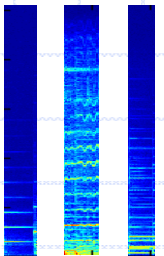
抽出音声信号

音声パート



+

音楽パート



Complex NMF

by H. Kameoka

Open	サンプリング	16000 Hz	フレーム長	64 ms
	トータル時間	23 s		
Factorize	スペクトルパーツ数	30	Save	
	エンベロープパーツ数	5		

表示するスペクトルパーツ

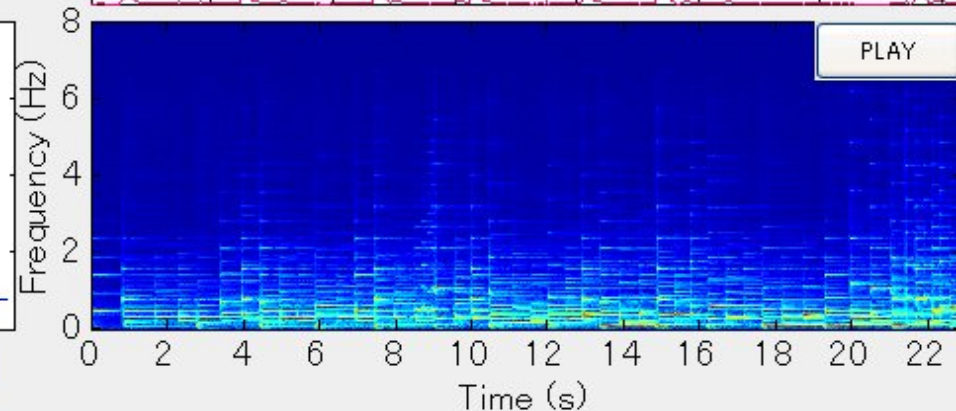
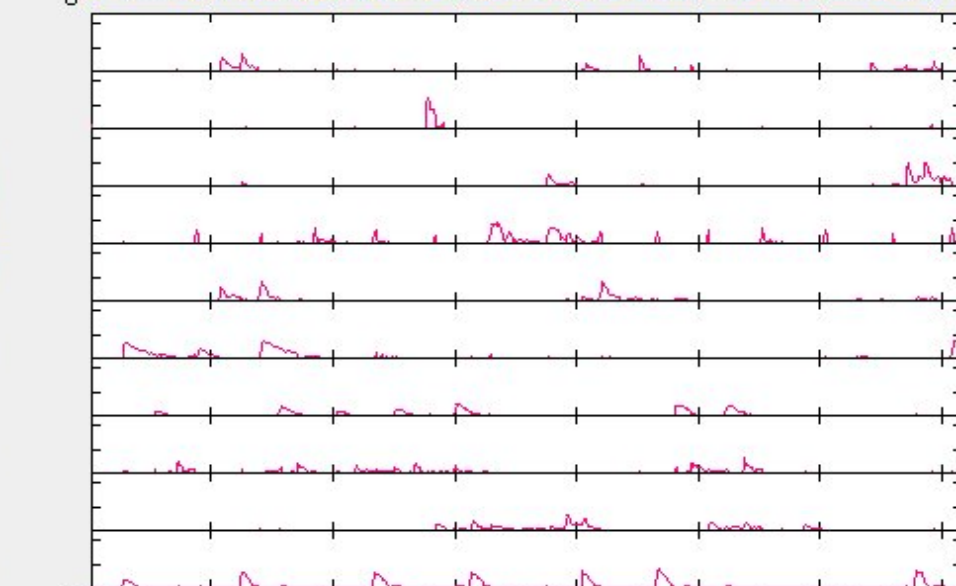
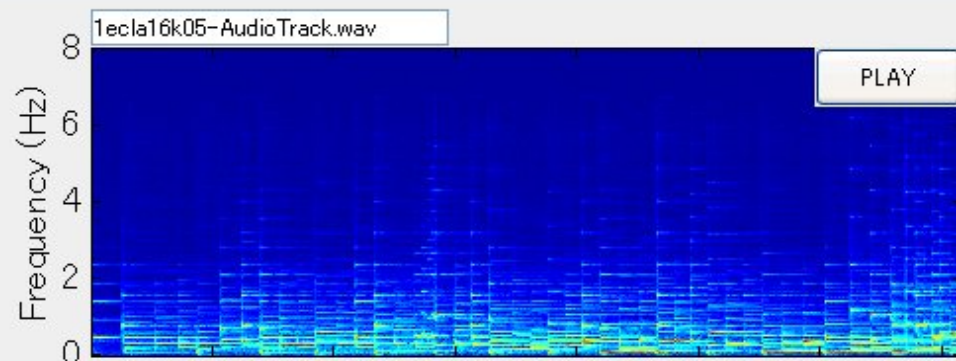
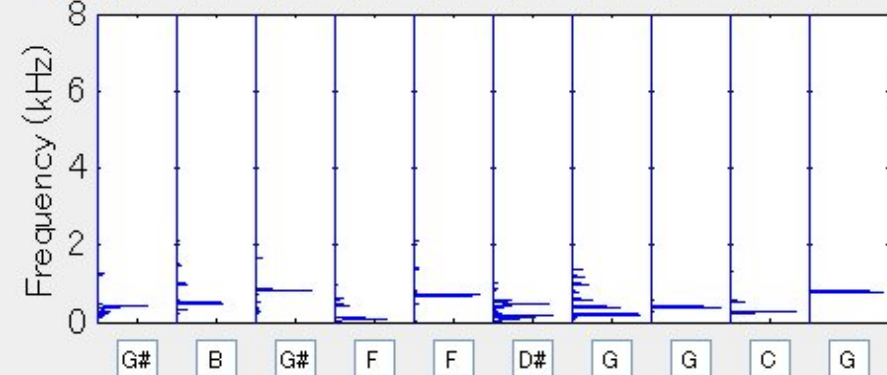
No. 1 - 10

トランスポーズ

0	+	+	+	+	+	+	+	+	+	+
m	-	-	-	-	-	-	-	-	-	-

ボリュームコントロール

0										
M										



複合自己回帰系

目的: ブラインド音声強調のための音声モデリング

◆音源分離/残響除去・・・室内伝達系と音声信号が未知

→ 音声信号らしさの規準(モデル)をうまく仮定することが重要

(例) 非ガウス性 → 独立成分分析に基づくブラインド音源分離

◆観測モデルの例 (マイクロホン数: M , 音源数: M) [T. Yoshioka et al., 2008]

観測信号の時間周波数成分
 $(Y_{\omega,t}^{(1)}, \dots, Y_{\omega,t}^{(M)})^T$

$\xrightarrow{\text{残響除去フィルタ}} \mathbf{Y}_{\omega,t} = \sum_{\tau=1}^{N_{\omega}} \mathbf{G}_{\omega,\tau} \mathbf{Y}_{\omega,t-\tau} + \hat{\mathbf{S}}_{\omega,t}$ (ω : 周波数, t : 時刻)

$\xleftarrow{\text{分離行列}} \mathbf{W}_{\omega} \hat{\mathbf{S}}_{\omega,t} = \mathbf{S}_{\omega,t}$ 音源成分 $(S_{\omega,t}^{(1)}, \dots, S_{\omega,t}^{(M)})^T$

瞬時混合信号

◆統計モデル $f_{S_{\omega,t}}(S_{\omega,t}; \theta)$ の定義 → 尤度関数 $f_Y(Y; \theta)$

■ 仮定: $S_{\omega,t}^{(m)}$ と $S_{\omega',t'}^{(m')}$ ($(\omega, t, m) \neq (\omega', t', m')$) は独立 $\rightarrow \{Y_{\omega,t}\}_{\forall(\omega,t)}$

$$f_Y(Y) = \prod_{\omega} |\det \mathbf{W}_{\omega}| \prod_{t,m} f_{S_{\omega,t}^{(m)}} \left(\left[\mathbf{W}_{\omega} \left(\mathbf{Y}_{\omega,t} - \sum_{\tau=1}^{N_{\omega}} \mathbf{G}_{\omega,\tau} \mathbf{Y}_{\omega,t-\tau} \right) \right]_m; \theta^{(m)} \right)$$

➡ 音源分離&残響除去: $\theta^{(m)}, \mathbf{W}_{\omega}, \mathbf{G}_{\omega,t}$ の最尤/MAP推定

音声生成モデル(ソースフィルタモデル)

◆自己回帰系による短時間フレーム t 内の信号モデル

$$\epsilon_t[n] \longrightarrow G_t(z) \longrightarrow s_t[n] = \sum_{p=1}^P a_{p,t} s_t[n-p] + \epsilon_t[n]$$

声帯による駆動源に対応

◆定常Gauss過程

$$(\epsilon_t[1], \dots, \epsilon_t[N])^T \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

◆白色性 $\Sigma = \sigma^2 \mathbf{I}$

声道特性(音素)に対応

◆ P 次の全極型モデル

$$\mathbf{G}_t(z) = 1/\mathbf{A}_t(z)$$

$$\mathbf{A}_t(z) = 1 - a_{1,t}z^{-1} \dots - a_{P,t}z^{-P}$$

$$E_{\omega,t} := \mathcal{F}\{\epsilon[n]\} \\ \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2 \mathbf{I})$$

$$S_{\omega,t} := \frac{E_{\omega}}{A_t(e^{j2\pi\omega/\Omega})} \\ \sim \mathcal{N}_{\mathbb{C}}\left(0, \underbrace{\frac{\sigma^2}{|A_t(e^{j2\pi\omega/\Omega})|^2}}_{\text{パワースペクトル密度(PSD)}}\right)$$

パワースペクトル密度(PSD)

framewise自己回帰系

複合自己回帰系

フィル
タ

フレームごとに別個
の全極モデル

入力

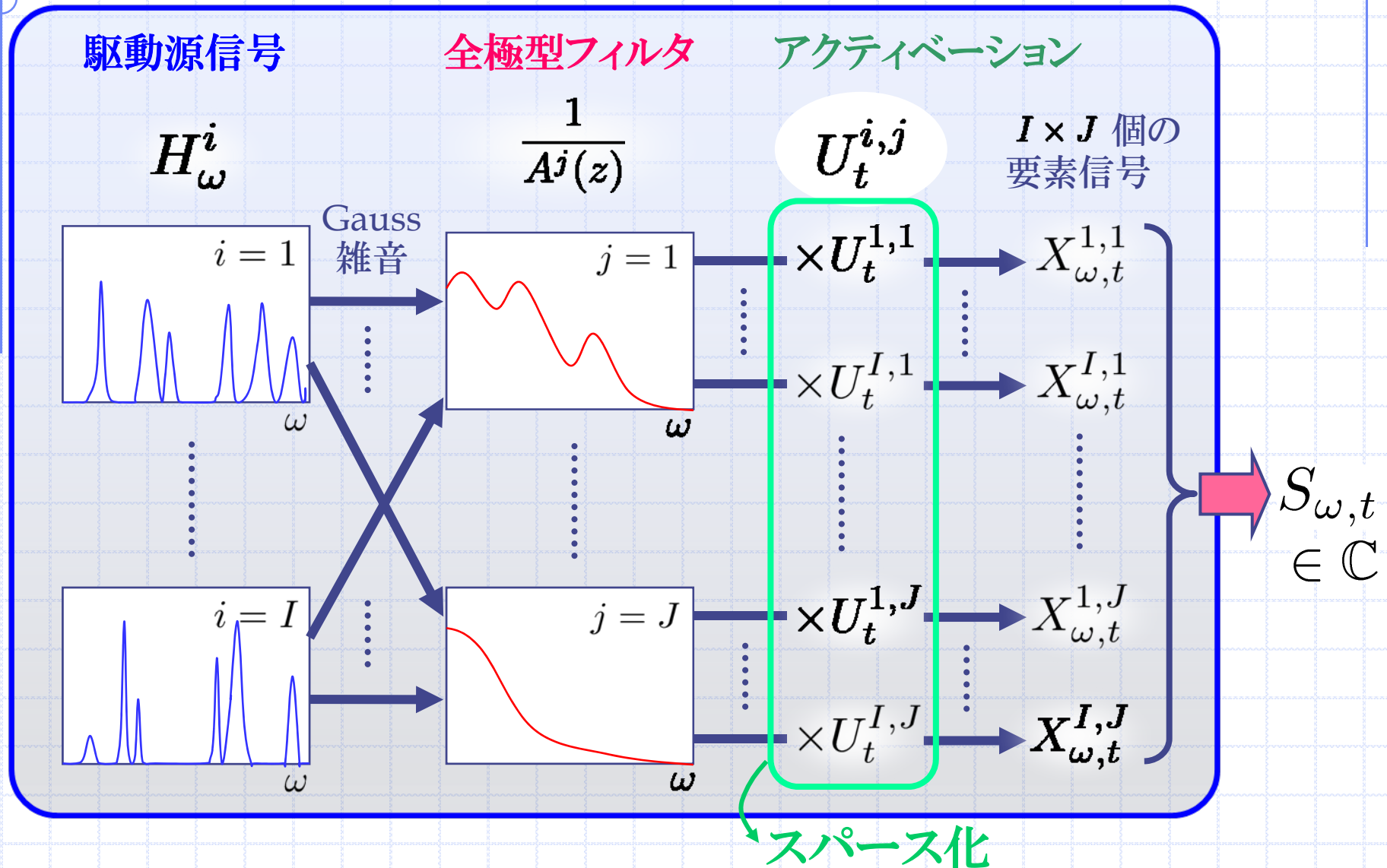
白色性を仮定
(パワースペクトル
密度が平坦)
実際は違う!

全フレームで高々 J 種類
の全極モデルを仮定

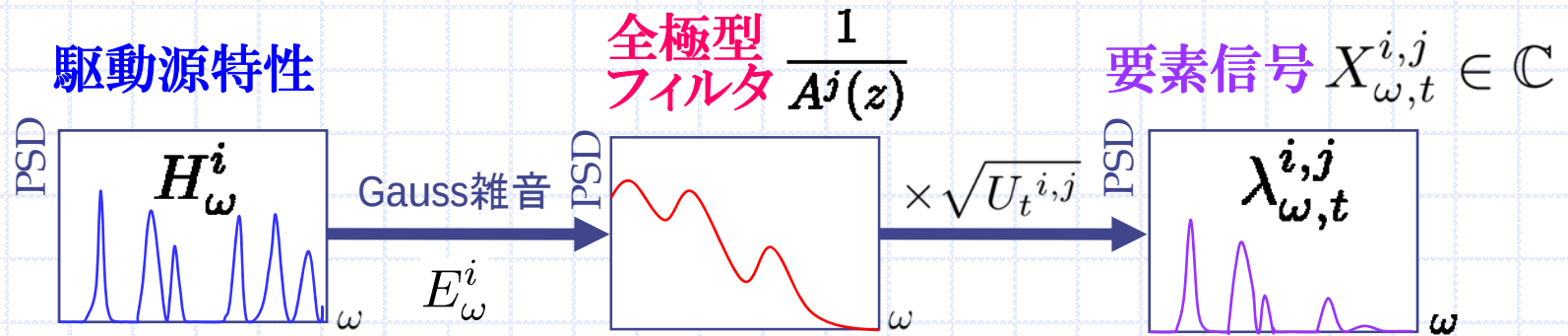
パワースペクトル密度
自体がパラメータ

全フレームで高々 I 種類
のパワースペクトル密度

複合自己回帰系による音声信号 $S_{\omega,t}$ のモデル化



要素信号スペクトルの確率密度関数



- 駆動信号スペクトル: $E_{\omega}^i \sim \mathcal{N}_{\mathbb{C}}(0, H_{\omega}^i)$
- 声道フィルタ通過後: $\frac{E_{\omega}^i}{A^j(e^{j2\pi\omega/\Omega})} \sim \mathcal{N}_{\mathbb{C}}\left(0, \frac{H_{\omega}^i}{|A^j(e^{j2\pi\omega/\Omega})|^2}\right)$
- アクティベート後: $\frac{E_{\omega}^i \sqrt{U_t^{i,j}}}{A^j(e^{j2\pi\omega/\Omega})} \sim \mathcal{N}_{\mathbb{C}}\left(0, \frac{H_{\omega}^i U_t^{i,j}}{|A^j(e^{j2\pi\omega/\Omega})|^2}\right)$

ここで, $X_{\omega,t}^{i,j} := \frac{E_{\omega}^i \sqrt{U_t^{i,j}}}{A^j(e^{j2\pi\omega/\Omega})}$, $\lambda_{\omega,t}^{i,j} := \frac{H_{\omega}^i U_t^{i,j}}{|A^j(e^{j2\pi\omega/\Omega})|^2}$ とすると

$$X_{\omega,t}^{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_{\omega,t}^{i,j})$$

音声信号スペクトルの確率密度関数

◆ 要素信号 $X_{\omega,t}^{i,j}$ の和を音声信号 $S_{\omega,t}$ と考える

$$X_{\omega,t}^{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_{\omega,t}^{i,j})$$

◆ $X_{\omega,t}^{i,j}$ と $X_{\omega,t}^{i',j'}$ は $(i,j) \neq (i',j')$ のとき独立する

$$\Rightarrow S_{\omega,t} = \sum_{i,j} X_{\omega,t}^{i,j} \\ \sim \mathcal{N}_{\mathbb{C}}\left(0, \sum_{i,j} \lambda_{\omega,t}^{i,j}\right)$$

logをとって少し式操作すると

$$|S_{\omega,t}|^2 \text{ と } \sum_i \sum_j \frac{H_{\omega}^i U_t^{i,j}}{|A^j(e^{j2\pi\omega/\Omega})|^2}$$

の板倉斎藤距離になる

音声信号 $S_{\omega,t}$ の統計モデル

$$f_{S_{\omega,t}}(S_{\omega,t}; \theta) = \frac{1}{\pi \phi_{\omega,t}} \exp\left(-\frac{|S_{\omega,t}|^2}{\phi_{\omega,t}}\right) \quad \text{但し, } \phi_{\omega,t} = \sum_{i,j} \lambda_{\omega,t}^{i,j}$$

アクティベーションのスパースネスを保障する事前確率

$$f_{\theta}(\theta) \propto \prod_{i,j,t} \frac{1}{U_t^{i,j} \alpha} \exp\left(-\frac{\beta}{U_t^{i,j}}\right) \quad (\text{逆ガンマ分布})$$

EMアルゴリズムによる最適化

$$L(\Theta) := \log f_{Y|\Theta}(Y|\Theta) + \log f_{\Theta}(\Theta) \rightarrow \text{maximize}$$

$$\Theta := \{ \{W_{\omega}\}, \{G_{\omega,\tau}\}, \{H_{\omega}^{(m)i}\}, \{a_p^{(m)j}\}, \{U_t^{(m)i,j}\} \}$$

$$L(\Theta) = \sum_{\omega,t} \left\{ \log \det(W_{\omega}^H \Phi_{\omega,t}^{-1} W_{\omega}) - \left(Y_{\omega,t} - \sum_{\tau} G_{\omega,\tau} Y_{\omega,t-\tau} \right)^H W_{\omega}^H \Phi_{\omega,t}^{-1} W_{\omega} \left(Y_{\omega,t} - \sum_{\tau} G_{\omega,\tau} Y_{\omega,t-\tau} \right) \right\} + \log f_{\Theta}(\Theta)$$

$$\Phi_{\omega,t} := \begin{bmatrix} \phi_{\omega,t}^{(1)} & & 0 \\ & \ddots & \\ 0 & & \phi_{\omega,t}^{(M)} \end{bmatrix}$$

M番目の話者の
音声パワースペクトル密度推定値

(Step 1: 音源分離) $\{\Phi_{\omega,t}\}, \{G_{\omega,\tau}\}$ を固定

$$\{W_{\omega}\} \leftarrow \operatorname{argmax}_{\{W_{\omega}\}} L(\Theta)$$

(Step 2: 残響除去) $\{\Phi_{\omega,t}\}, \{W_{\omega}\}$ を固定

$$\{G_{\omega,\tau}\} \leftarrow \operatorname{argmax}_{\{G_{\omega,\tau}\}} L(\Theta)$$

$$\bar{S}_{k,n}^{(m)} \leftarrow \left[W_k \left(Y_{k,n} - \sum_{\tau} G_{k,\tau} Y_{k,n-\tau} \right) \right]_m$$

(Step 3: 音声パラメータ推定) $\bar{S}_{\omega,t}^{(m)}$ を固定

$$\{ \{H_{\omega}^{(m)i}\}, \{a_p^{(m)j}\}, \{U_t^{(m)i,j}\} \} \leftarrow \operatorname{argmax} L(\Theta)$$

をEMアルゴリズムにより行う!

Step 3の詳細

Q関数

$$Q(\Theta, \Psi) := \sum_{m,i,j,\omega,t} \left[\log \frac{|A^{(m)j}(e^{j2\pi\omega/\Omega})|^2}{H_\omega^{(m)i} U_t^{(m)i,j}} - \frac{\Psi_{\omega,t}^{(m)i,j} |A^{(m)j}(e^{j2\pi\omega/\Omega})|^2}{H_\omega^{(m)i} U_t^{(m)i,j}} \right] + \sum_{m,i,j,t} \left[\alpha \log U_t^{(m)i,j} + \frac{\beta}{U_t^{(m)i,j}} \right]$$

E-step

$$\lambda_{\omega,t}^{(m)i,j} \leftarrow \frac{H_\omega^{(m)i} U_t^{(m)i,j}}{|A^{(m)j}(e^{j2\pi\omega/\Omega})|^2}$$

$$\phi_{\omega,t}^{(m)} \leftarrow \sum_{i,j} \lambda_{\omega,t}^{(m)i,j}$$

$$\Psi_{\omega,t}^{(m)i,j} \leftarrow \lambda_{\omega,t}^{(m)i,j} \left[1 + \frac{\lambda_{\omega,t}^{(m)i,j} (|\bar{S}_{\omega,t}^{(m)}|^2 - \phi_{\omega,t}^{(m)})}{\phi_{\omega,t}^{(m)2}} \right]$$

M-step

[駆動源特性]

$$H_\omega^{(m)i} \leftarrow \frac{1}{TJ} \sum_{t,j} \frac{\Psi_{\omega,t}^{(m)i,j} |A^{(m)j}(e^{j2\pi\omega/\Omega})|^2}{U_t^{(m)i,j}}$$

[ゲイン]

$$U_t^{(m)i,j} \leftarrow \frac{1}{\Omega + \alpha} \sum_{\omega} \left[\frac{\Psi_{\omega,t}^{(m)i,j} |A^{(m)j}(e^{j2\pi\omega/\Omega})|^2}{H_\omega^{(m)i}} + \beta \right]$$

$$r_p^{(m)j} \leftarrow \text{DFT}^{-1} \left\{ \sum_{i,n} \frac{\Psi_{\omega,t}^{(m)i,j}}{H_\omega^{(m)i} U_t^{(m)i,j}} \right\}$$

[自己回帰係数]

$$\begin{bmatrix} a_1^{(m)j} \\ \vdots \\ a_P^{(m)j} \end{bmatrix} \leftarrow \begin{bmatrix} r_0^{(m)j} & \dots & r_{1-P}^{(m)j} \\ \vdots & \ddots & \vdots \\ r_{P-1}^{(m)j} & \dots & r_0^{(m)j} \end{bmatrix}^{-1} \begin{bmatrix} -r_1^{(m)j} \\ \vdots \\ -r_P^{(m)j} \end{bmatrix}$$

板倉齋藤距離規準のNMFと等価となる条件

◆ Step3は音源 m ごとのパラメータ最適化ステップより以下では m は省略

◆ 板倉齋藤距離: $D_{\text{IS}}(Y||\phi) = \sum_{\omega,t} \left(\frac{|Y_{\omega,t}|^2}{\phi_{\omega,t}} - \log \frac{|Y_{\omega,t}|^2}{\phi_{\omega,t}} - 1 \right)$

$P = 0, J = 1$ の下では $\phi_{\omega,t} = \sum_i H_{\omega}^i U_t^{i,1}$ となり複合自己回帰系におけるパワースペクトル密度モデルはNMFモデルと等価

$$|Y_{\omega,t}|^2 \simeq \phi_{\omega,t} = \sum_i H_{\omega}^i U_t^{i,1} \Rightarrow Y \simeq HU$$

このとき、 H と U のEMアルゴリズムの更新則が板倉齋藤距離規準のNMF [C. Févotte et al., 2008] の更新則と等価となる。

$$H_k^i = \frac{1}{N} \sum_n \frac{\Psi_{k,n}^{i,1}}{U_n^{i,1}}, \quad U_n^{i,1} = \frac{1}{K} \sum_k \frac{\Psi_{k,n}^{i,1}}{F_k^i}$$

残響環境下ブラインド音源分離実験

◆実験結果

•混合条件

	Mic#1	Mic#2	Mic#3	Mic#4
Src#1	−0.59	−0.32	−0.14	+0.57
Src#2	+0.59	+0.32	+0.14	−0.57

単位: [dB]

•Src#1の分離性能 : SIR (Signal-to-Interference ratio)

	提案法	従来法
SIR	+18.6 dB	+17.2 dB

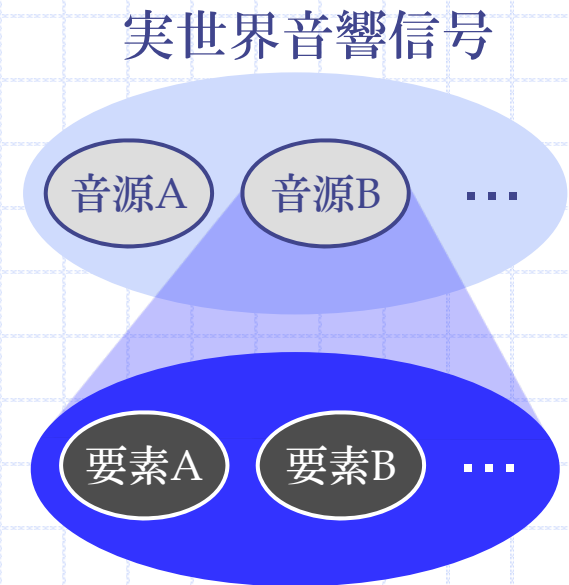
•Src#2の残響除去性能: DRR (Direct-to-Reverberate ratio)

	提案法	従来法
DRR	+13.5 dB	+12.6 dB

まとめ (1/2)

◆実世界音響信号処理へのアプローチ

- 実世界音響信号が何個かの統計的に独立な音源信号によって構成されると仮定するのと同様に、各音源信号もまた何らかの離散的なシンボル情報に対応した何個かの独立成分によって構成される、と仮定
- 実世界音響信号を階層的に独立な成分に分解したモデルで簡潔に記述し、現象をモデルパラメータ最適化の視点から推論



まとめ (2/2)

◆ スパース表現(NMF)の考え方をヒントにした新しい音響信号モデルを提案

通常のNMFモデル: $Y_{\omega,t} \simeq \sum_{i=1}^I H_{\omega}^i U_t^i$

低ランク表現

$$\boxed{Y} \simeq \boxed{H} \boxed{U}$$

1. 複素NMF

- ◆ NMFによる振幅スペクトログラム分解表現の特長を継承した複素スペクトログラムの分解表現

複素スペクトログラム

$$Y_{\omega,t} \simeq \sum_{i=1}^I H_{\omega}^i U_t^i e^{j\phi_{\omega,t}^i}$$

2. 複合自己回帰系

- ◆ 音声のソースフィルタモデルをバックボーンとした音声版のスパース表現モデル

パワースペクトログラム

$$Y_{\omega,t} \simeq \sum_{i=1}^I \sum_{j=1}^J \frac{H_{\omega}^i U_t^{i,j}}{|A^j(e^{j2\pi\omega/\Omega})|^2}$$

参考文献

- [1] 亀岡, 小野, 柏野, 嵯峨山, “複素NMF: 新しいスパース信号分解表現と基底系学習アルゴリズム,” 日本音響学会2008年秋季研究発表会講演論文集, 2-8-13, pp. 657-660, 2008.
- [2] 亀岡, 柏野, “複合自己回帰系に基づく音響信号の生成モデル,” 日本音響学会2009年春季研究発表会講演論文集, 2-9-8, pp. 671-674, 2009.
- [3] D. D. Lee and H. S. Seung, “Learning the parts of objects with nonnegative matrix factorization,” *Nature*, vol. 401, pp. 788–791, 1999.
- [4] P. Smaragdis and J.C. Brown, “Non-negative matrix factorization for music transcription,” *Proc. WASPAA 2003*, pp. 177–180, 2003.
- [5] D.D. Lee and H.S. Seung, “Algorithms for nonnegative matrix factorization,” *Proc. NIPS 2000*, pp. 556–562, 2000.
- [6] P. Smaragdis, B. Raj, M.V. Shashanka, “Supervised and semi-supervised separation of sounds from single-channel mixtures,” *Proc. 7th International Conference on Independent Component Analysis and Signal Separation*, 2007.
- [7] P. Smaragdis and B. Raj, “Example-driven bandwidth expansion,” *Proc. WASPAA 2007*, in CD-ROM, 2007.
- [8] L.M. Bregman, “The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming,” *USSR Computational Mathematics and Mathematical Physics* 7: pp. 200–217, 1967.
- [9] I.S. Dhillon and S. Sra, “Generalized nonnegative matrix approximations with Bregman divergences,” *Proc. NIPS 2005*, pp. 283-290, 2005.

参考文献

- [10] J. de Leeuw. Block-relaxation methods in statistics. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Information Systems and Data Analysis*, Berlin, 1994. Springer Verlag.
- [11] 亀岡, 後藤, 嵯峨山, “スペクトル制御エンベロープによる混合音中の周期および非周期成分の選択的イコライザ,” 情処研報, 2006-MUS-66-13, pp. 77-84, 2006.
- [12] C. Févotte, N. Bertin and J.-L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis,” Technical Report TELECOM ParisTech 2008D006, 2008.
- [13] S. Eguchi and Y. Kano, “Robustifying maximum likelihood estimation,” Technical report, Institute of Statistical Mathematics, Research Memo. 802, 2001.
- [14] 中野, 北野, ルルー, 亀岡, 小野, 嵯峨山, “可変基底NMFに基づく音楽音響信号の解析,” 情処研報, 2010-MUS-84-10, 2010.
- [15] B.A. Olshausen, and D.J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, pp. 607-609, 1996.
- [16] G. Harpur, R. Prager, “Experiments with low-entropy neural networks,” in R. Baddeley, P. Hancock, P. Földiák (eds.) *Information Theory and the Brain*, Cambridge University Press, pp. 84-100, 2000.
- [17] T. Hofmann, “Probabilistic latent semantic analysis,” *Proc. UAI*, pp. 289-296, 1999.
- [18] T. Yoshioka, T. Nakatani and M. Miyoshi, “An integrated method for blind separation and dereverberation of convolutive audio mixtures,” *Proc. EUSIPCO 2008*, in CD-ROM, 2008.