

信号処理論特論 第2回 (10/04)

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

- 9/27: 第1回 統計的音声音響信号処理概論
- 10/04: 第2回 非負値行列因子分解
- 10/11: 第3回 線形因子分析(ICA, IVA, ILRMA)
- 10/18: 第4回 エンハンスメント・バイノーラル再生
- 10/25: 第5回 音場再現・スパース最適化
- 11/01: 第6回 高次統計量解析とその応用
- 11/15: 第7回 【レポート課題1】
- 11/22: 第8回 話者認識
- 11/29: 休講
- 12/06: 第9回 音声合成・変換1
- 12/13: 休講
- 12/20: 第10回 音声合成・変換2
- 01/10: 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しておきます)

■ 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

■ 非負値行列因子分解

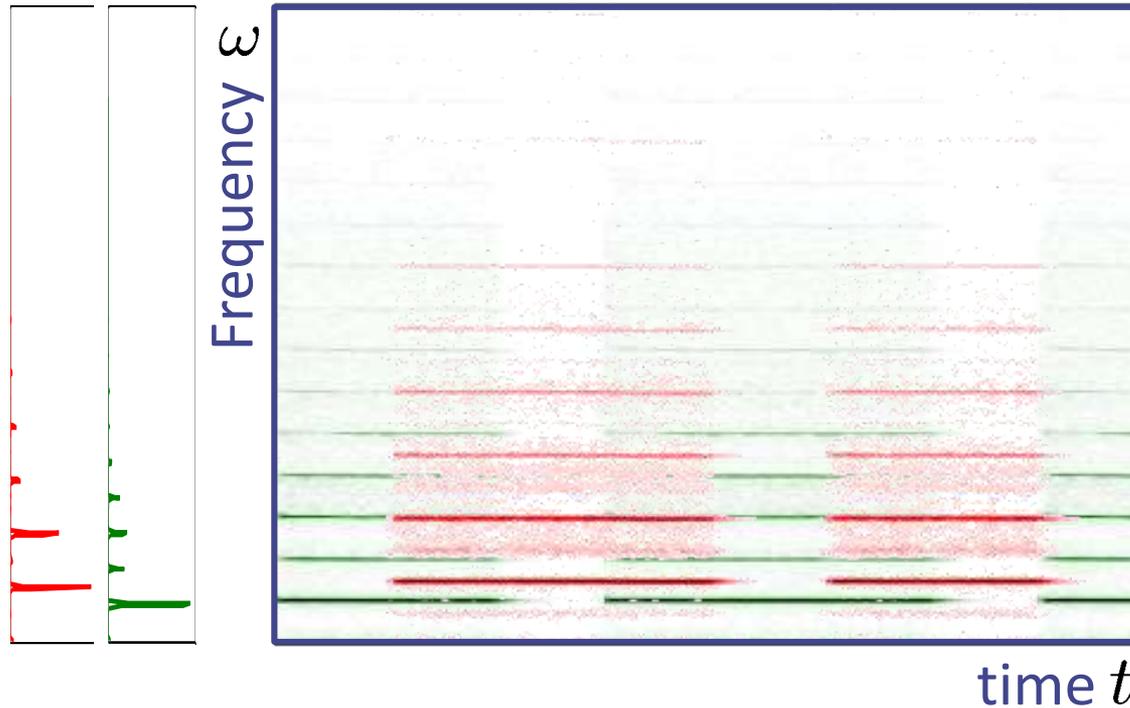
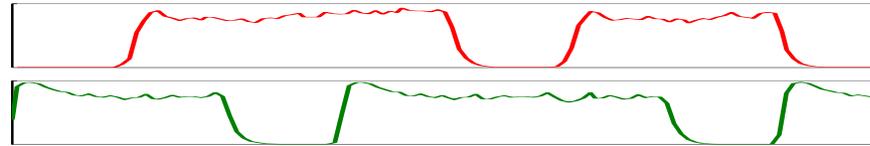
- データ行列を基底行列と係数行列の積に分解する
多変量解析手法の一つ
 - ◆ 非負制約を置く所がポイント
- 画像処理分野で登場した技術
 - ◆ 顔画像データから目・鼻・口などの特徴的パーツを抽出することが目的
- 近年, モノラル音源分離の手法としても注目
 - ◆ スペクトログラムをデータ行列と見なして, 行列の積に分解

■ 参考資料

- A. Cichocki, R. Zdunek, A. Huy Phan, S. Amari, *Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley & Sons, 2009.
- 亀岡弘和, "非負値行列因子分解," 計測と制御, Vol. 51, No. 9, pp. 835-844, Sep. 2012.

「行列の積」としてのスペクトログラム

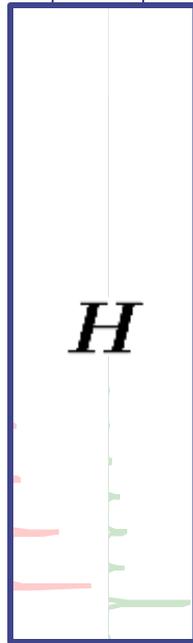
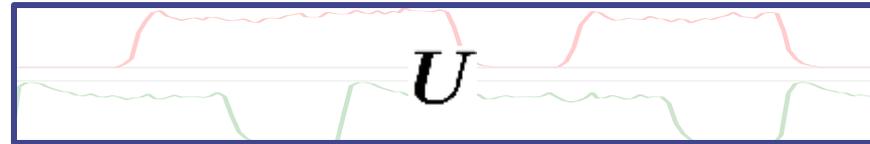
- 指定されたランクを持つスペクトログラムの生成



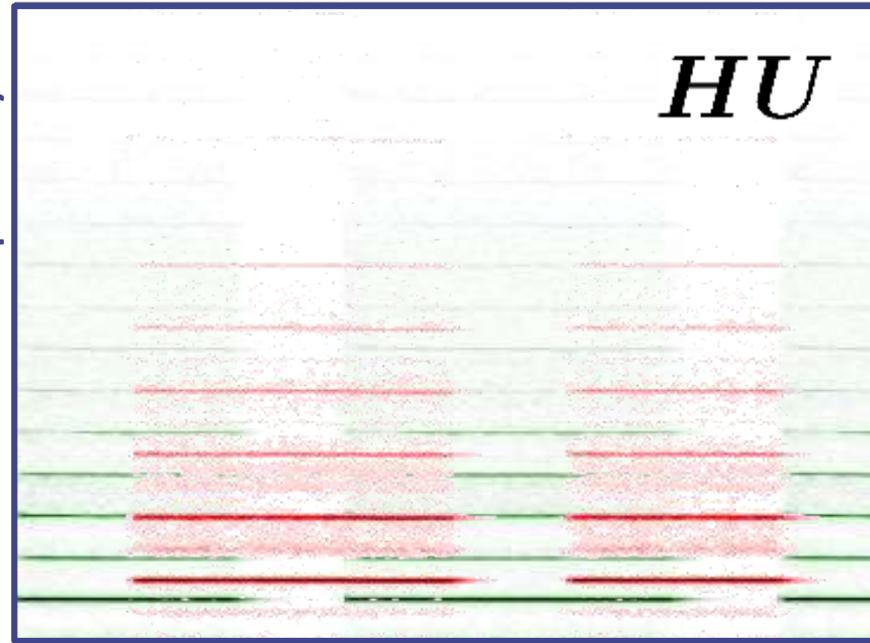
「行列の積」としてのスペクトログラム

各基底の「アクティベーション」

「基底スペクトル」



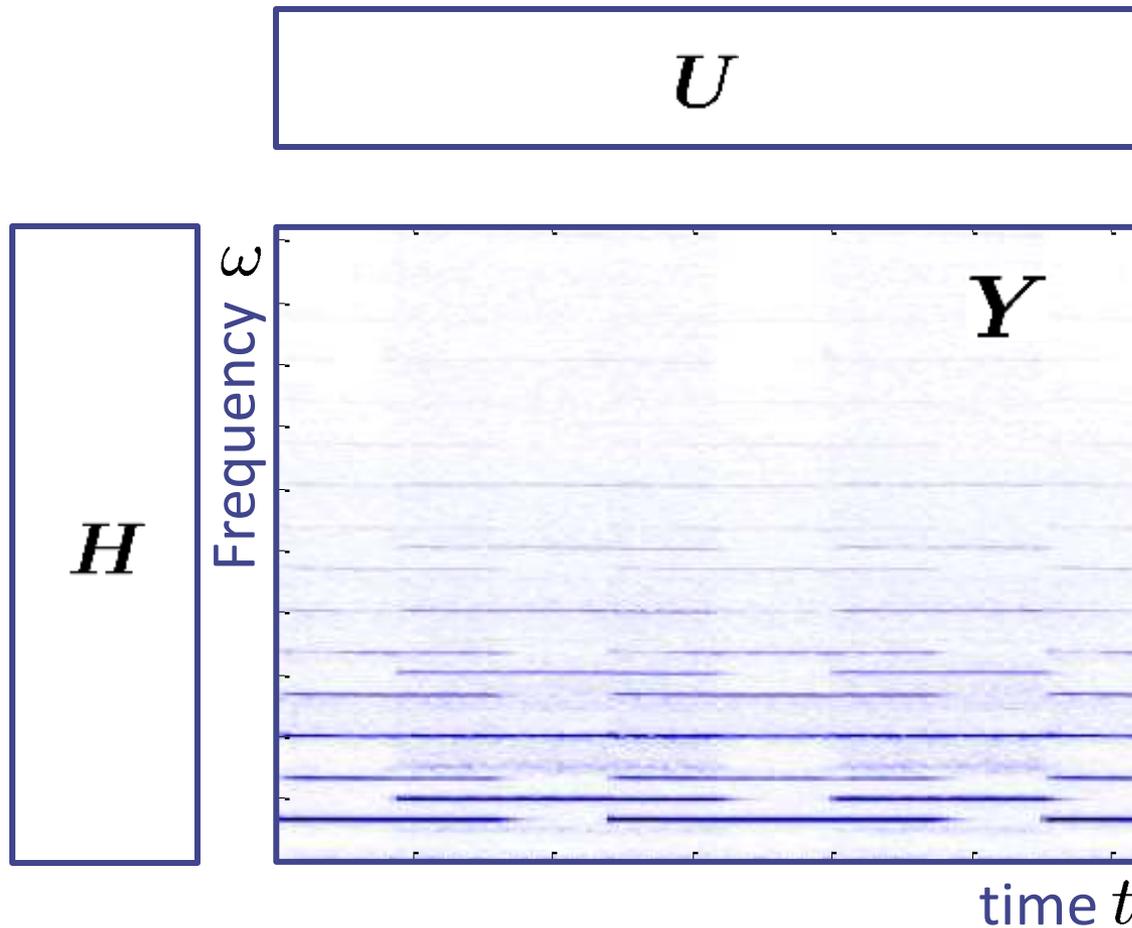
Frequency ϵ



time t

「行列の積」としてのスペクトログラム

- 観測スペクトログラムを低ランク行列積へ分解



本日の話題

- 非負値行列因子分解(NMF)の基礎
 - 何に使えるのか(音響信号処理を題材として)
 - どのような性質があるのか
 - どのように求めるのか
 - 統計モデルとしての解釈

- 様々なNMF及びその応用例
 - PLSAとの関係
 - スパースNMF、ベイジアンNMF、複素NMF etc.
 - 音楽信号分析例

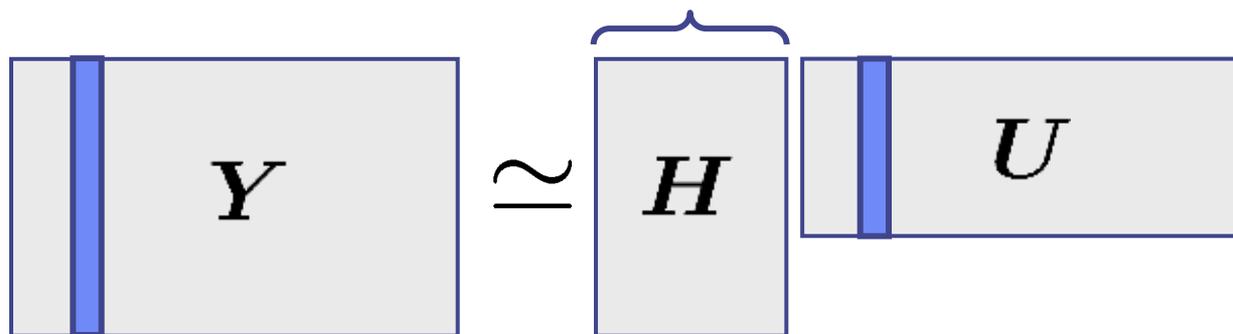
本日の話題

- 非負値行列因子分解(NMF)の基礎
 - 何に使えるのか(音響信号処理を題材として)
 - どのような性質があるのか
 - どのように求めるのか
 - 統計モデルとしての解釈

- 様々なNMF及びその応用例
 - PLSAとの関係
 - スパースNMF、ベイジアンNMF、複素NMF etc.
 - 音楽信号分析例

非負値行列因子分解 (Nonnegative Matrix Factorization) とは

- 非負値行列を2つの非負値行列の積で表現



- 行列因子分解の応用場面

- ブラインド信号分離(Blind Signal Separation)

H : 混合行列 U の各行: 源信号

- 教師有り音源分離(Informed Source Separation)

H : 基底スペクトル(一部はgiven) U : 生起時刻

- 次元圧縮

H の各列: 基底ベクトル U の各列: 結合係数

NMFが生まれた背景

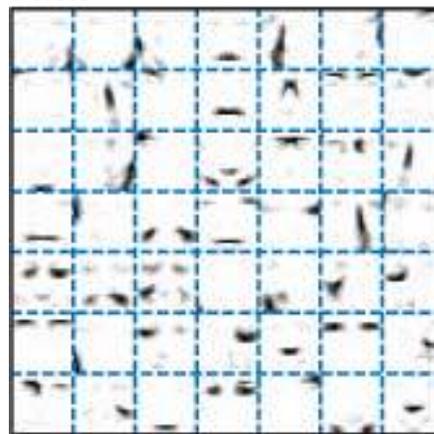
- ◆ 画像処理分野で生まれた技術 [Lee1999]
 - 顔画像から顔パーツを抽出するのが目的
 - 概念自体は90年代前半に登場 [Paatero1994]

- ◆ 音のスペクトルを画像と見なして適用(後述)
→ 音声分離・自動採譜等
[Smaragdis2003]以降極めて多数...

- ◆ 効率的な反復アルゴリズム [Lee2000]

$$H_{\omega,k} \leftarrow H_{\omega,k} \frac{[YU^T]_{\omega,k}}{[HUU^T]_{\omega,k}}$$
$$U_{k,t} \leftarrow U_{k,t} \frac{[H^T Y]_{k,t}}{[H^T HU]_{k,t}}$$

NMF



Original



x



=



VQ



||



x

PCA



||



x

なぜ非負値なのか？その意図は？

■ データ行列の非負性

- 実世界には非負値データが多い
(例) パワースペクトル, 画素値, 度数,

■ 基底行列の非負性

- 「非負値データの構成要素もまた非負値データであるべき(でないとは物理的に意味をなさない!)」
という考え方

(例) 負の値をもったパワースペクトルなんて解釈のしようがない

■ 係数行列の非負性

- 構成要素の混ざり方は「足し算」のみ
- 係数行列をスパースに誘導 → 基底の情報量をアップ

係数行列の非負性について

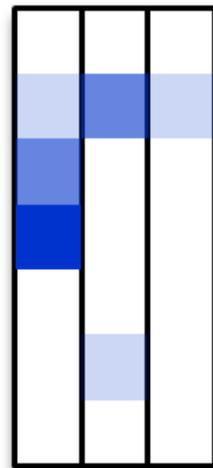
■ 係数が負の値を取っても良い場合

Observed vector

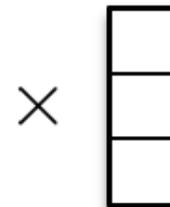


Basis vectors

(a)

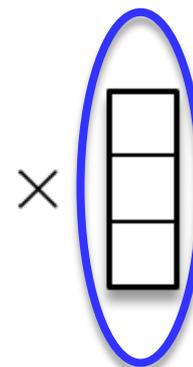
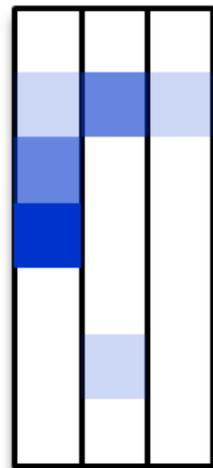


Coefficients



■ 係数が負の値を取ってはならない場合

(b)



スパース

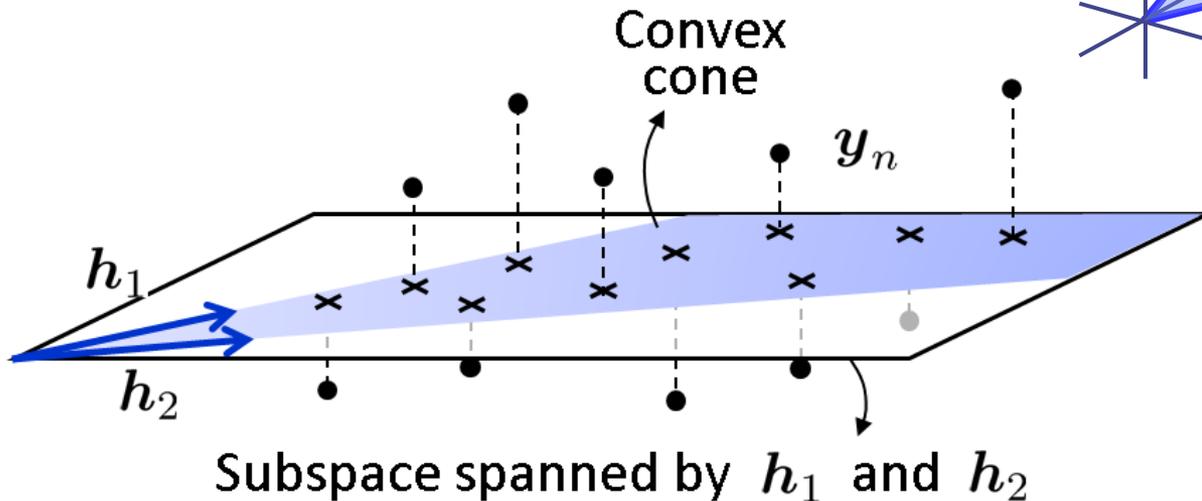
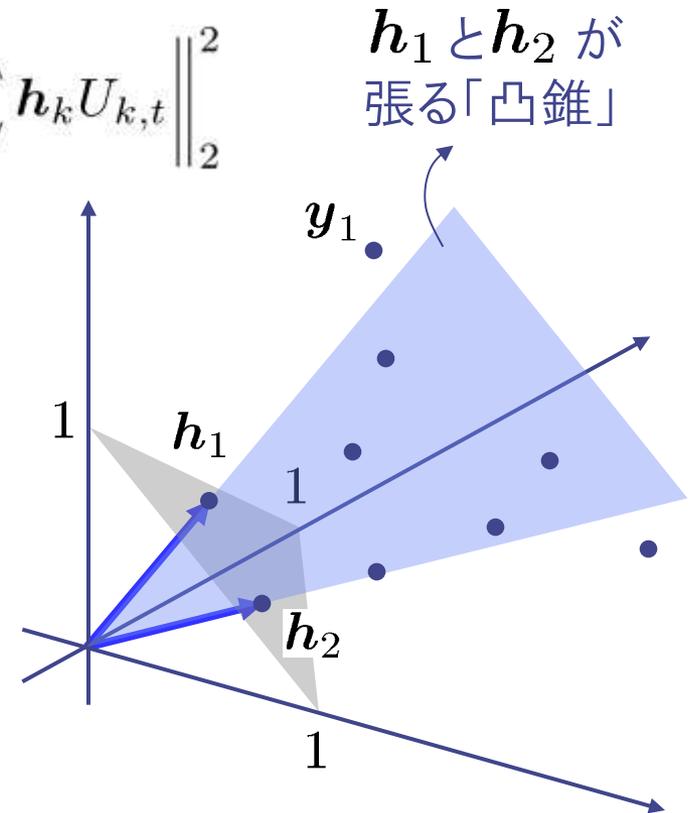
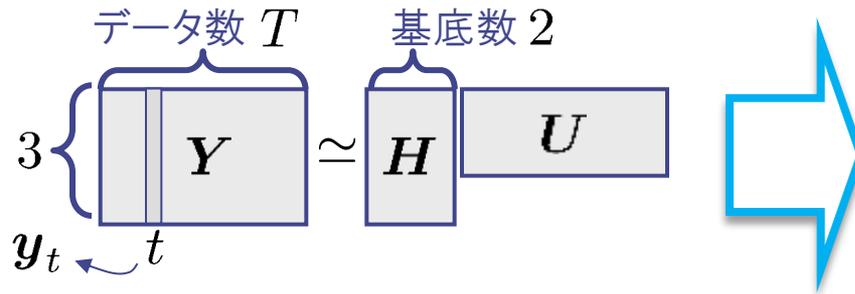
スパース性 \approx 統計的独立性

NMFの幾何学的イメージ

■ 二乗誤差規準のNMF

$$\text{minimize } \|Y - HU\|_F^2 = \sum_t \left\| \mathbf{y}_t - \sum_k \mathbf{h}_k U_{k,t} \right\|_2^2$$

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



共起成分のグルーピング効果

- 低ランク近似問題:
少ない基底でデータを良く表現したい

$$Y \approx H U$$

- H をどうした方が「節約的」?

$$Y = H U$$

Diagram illustrating the grouping effect of H . The matrix H is shown as a sparse matrix with columns labeled a , b , and c . The matrix U is shown as a matrix of coefficients. The diagram shows that the product of H and U results in a matrix with a block-like structure, where the non-zero elements are grouped together.

NMFで音声スペクトログラムを分解してみる

◆ 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)

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

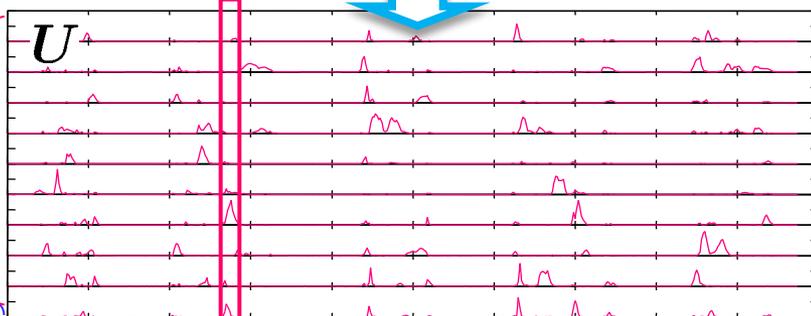
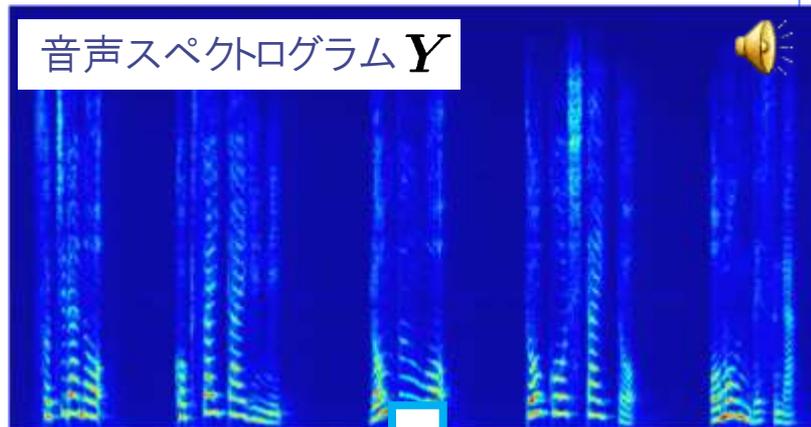
絶対値をとる

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

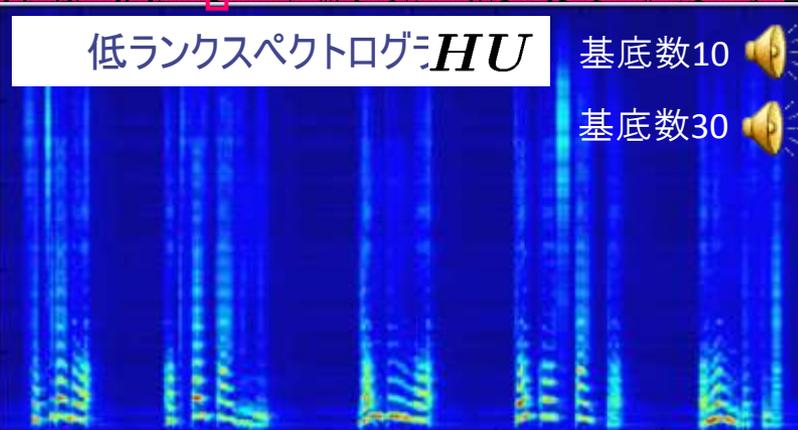
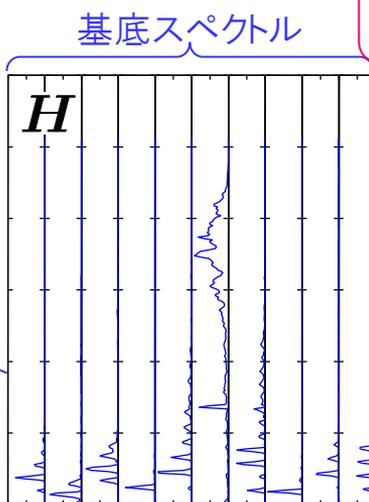
$$Y \simeq HU \quad (H \succeq 0, U \succeq 0)$$

各基底の
アクティビティ

周波数 \rightarrow



周波数 \rightarrow



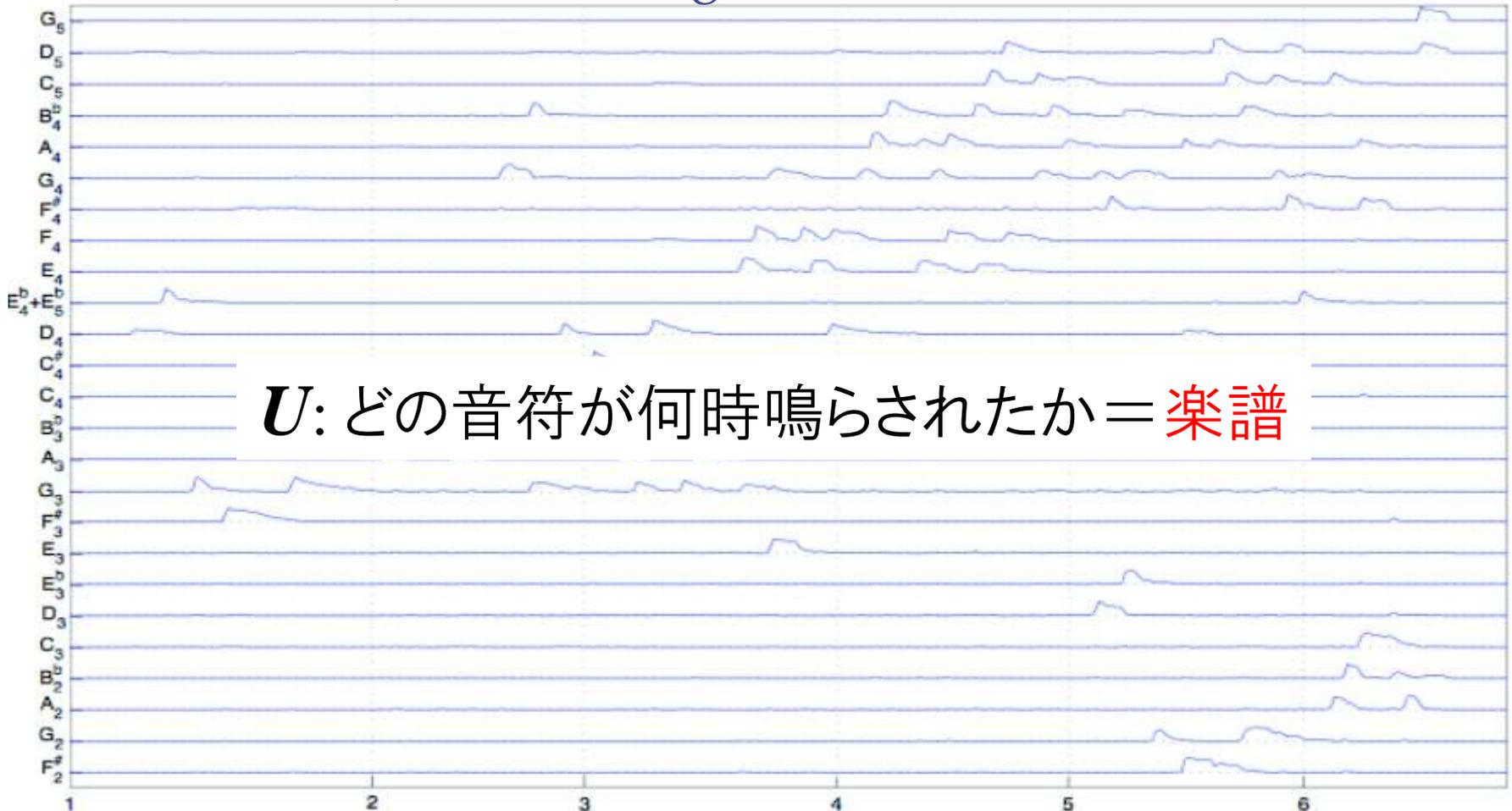
頻出するスペクトル
パーツが獲得される

時刻 \rightarrow

何に使えるのか？(1/6)

■ 自動採譜 [P. Smaragdis et al., 2003]

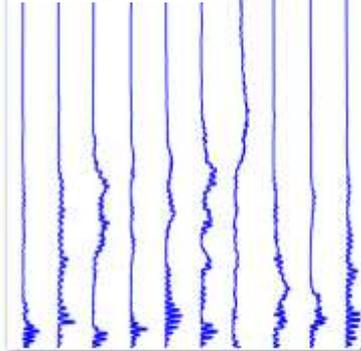
J.S. Bach: Fuge #16 in G minor



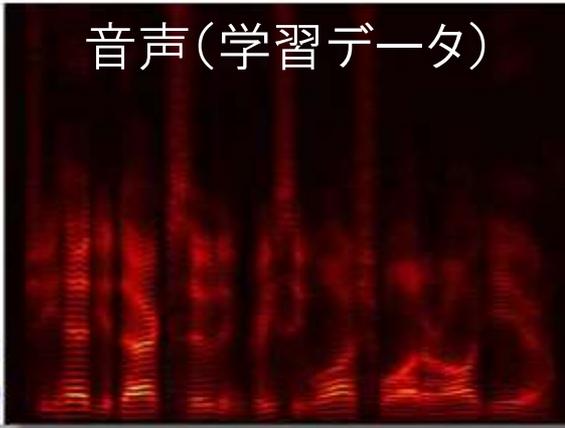
何に使えるのか？(2/6)

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

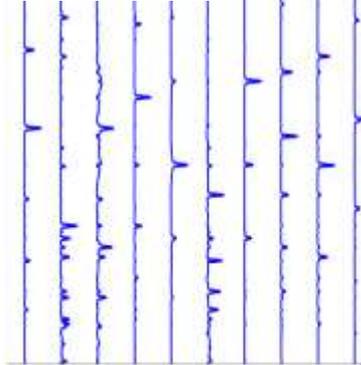
H_{speech}



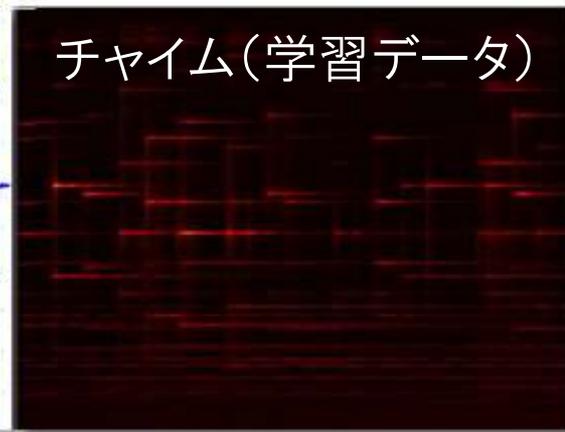
音声(学習データ)



H_{chime}

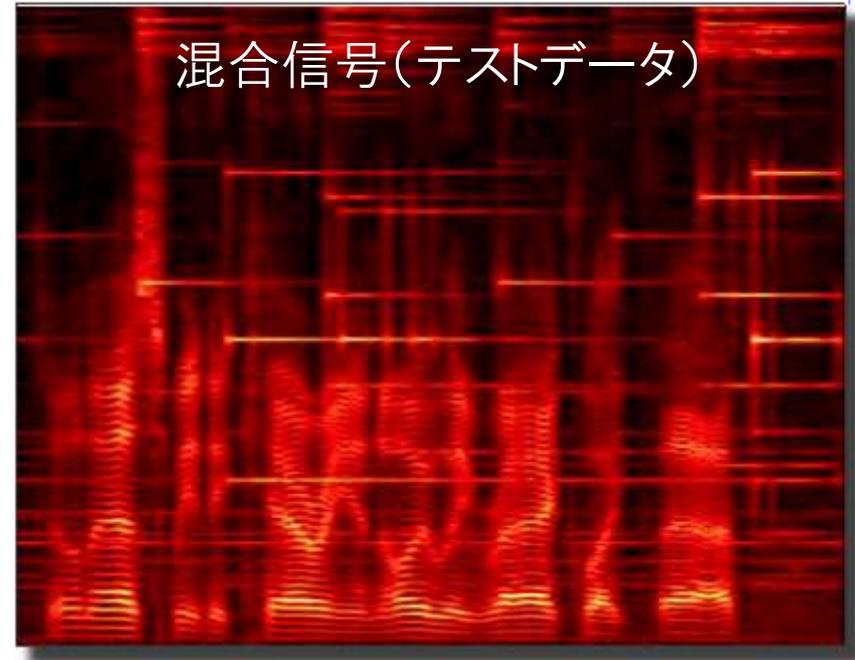


チャイム(学習データ)



Y

混合信号(テストデータ)



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

何に使えるのか？(3/6)

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

学習プロセス

$$Y_{\text{target}} = F Q$$

目的の楽器の教師音のみを用いて学習した基底

分離プロセス

$$Y = F G + H U$$

教師基底 F を固定して G, H, U を最適化

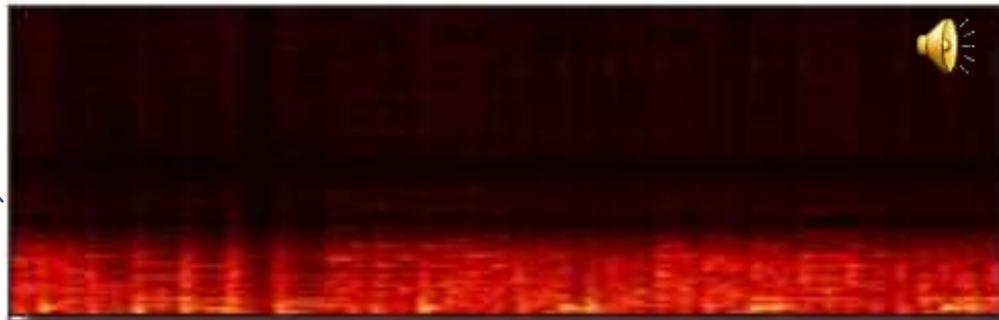
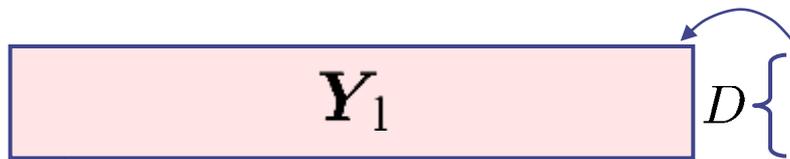
分離出力

H は F となるべく
無相関となるように求める

何に使えるのか？(4/6)

■音の「超解像」

[P. Smaragdis & B. Raj, 2007]

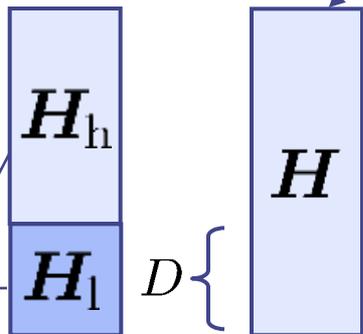


高サンプリングレート信号の学習データ



$$Y_1 \simeq H_1 U$$

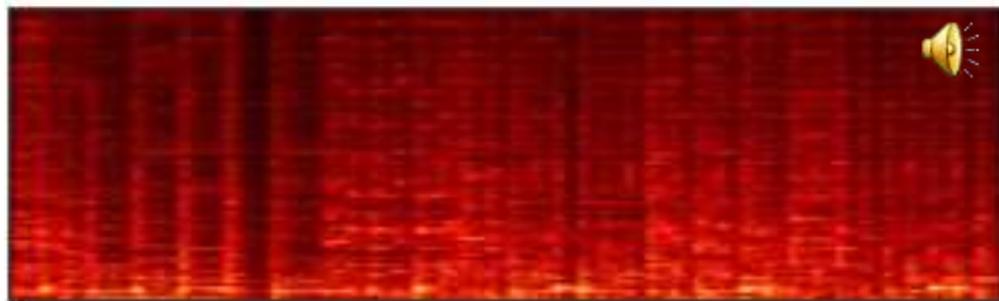
(固定)



$$H$$

$$H_h U = Y_h$$

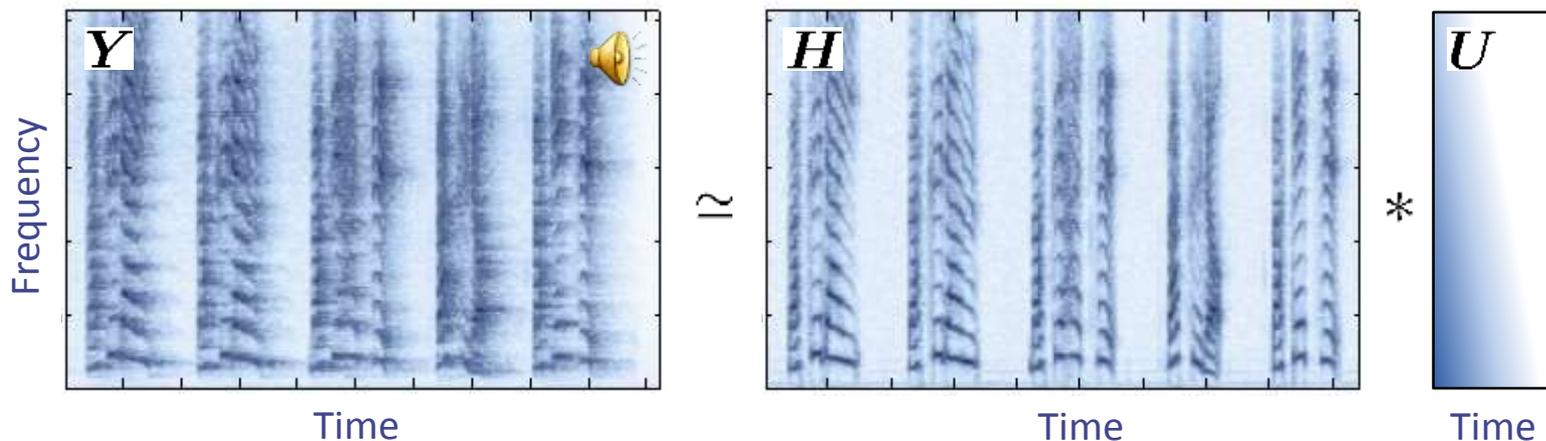
高周波帯域が復元された信号



何に使えるのか？(5/6)

■ ブラインド残響除去 [H. Kameoka et al., 2008]

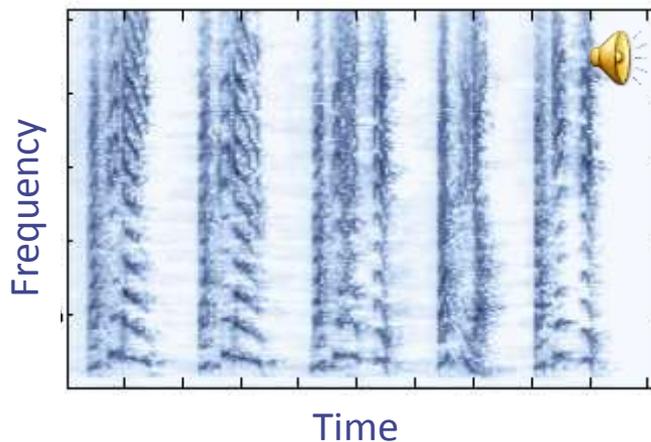
非負値行列「逆畳み込み」



$$Y_{\omega,t} \quad \text{is} \quad \sum_{\tau} H_{\omega,\tau} U_{\omega,t-\tau}$$

■ 残響除去音声

非負値行列逆畳み込み
によって求めた実際の H



何に使えるのか？(6/6)

- エコーキャンセラ

[戸上&川口 2009]

- 音響符号化

[Nikunen & Virtanen 2010]

- 教師有り音源分離(モノラル音分離)

[Smaragdis et al. 2007]

- ブラインド音源分離 (NMFの多チャンネル拡張)

[Ozerov et al. 2010] (多チャンネルNMFの先駆け)

[Sawada et al. 2011] (IS距離規範多チャンネルNMF)

[kitamura, Ono, Sawada, Kameoka, Saruwatari 2015] (独立低ランク行列分析)

- フォルマントラッキング

[Durrieu et al. 2011]

- 音素特徴量抽出

[Hurmalainen et al. 2011]

線形混合系の因子分析と関連
(来週の講義で扱う)

NMFの基本問題

- ◆ T 個の観測データ(非負値ベクトル) $\mathbf{y}_1, \dots, \mathbf{y}_T$
- ◆ K 個の非負値基底ベクトル $\mathbf{h}_1, \dots, \mathbf{h}_K$ 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

$$\mathbf{y}_t \simeq \sum_{k=1}^K \mathbf{h}_k U_{k,t}, \quad U_{k,t} \geq 0 \quad (t = 1, \dots, T)$$

$$\left. \begin{array}{l} \mathbf{Y} := (\mathbf{y}_1, \dots, \mathbf{y}_T) \\ \mathbf{H} := (\mathbf{h}_1, \dots, \mathbf{h}_K) \\ \mathbf{U} := (U_{k,t})_{K \times T} \end{array} \right\} \mathbf{Y} \simeq \mathbf{H}\mathbf{U} \quad H_{\omega,k} \geq 0, U_{k,t} \geq 0$$

\mathbf{Y} と $\mathbf{H}\mathbf{U}$ の乖離度を表す規準

$$\begin{array}{ll} \text{minimize} & \mathcal{D}(\mathbf{Y}|\mathbf{H}\mathbf{U}) \\ \text{subject to} & H_{\omega,k} \geq 0, U_{k,t} \geq 0 \end{array}$$

何をどうやって求めるのか

- 何を最小化するのか: NMFにおける代表的な最適化規準
 - 二乗誤差
 - lダイバージェンス(一般化KLダイバージェンス) etc.
- どうやって最小化するのか: 押さえておくべき基本原理
 - 補助関数法
 - 凸不等式(Jensenの不等式)

どのような最適化基準を選ぶのか？

- 様々なコスト関数が提案されている

- 二乗ユークリッド距離

$$D_{\text{EUC}}(a||b) = (a - b)^2$$

- 一般化カルバック・ライブラ(KL)ダイバージェンス

$$D_{\text{KL}}(a||b) = a \log \frac{a}{b} + b - a$$

- 板倉斎藤(IS)擬距離

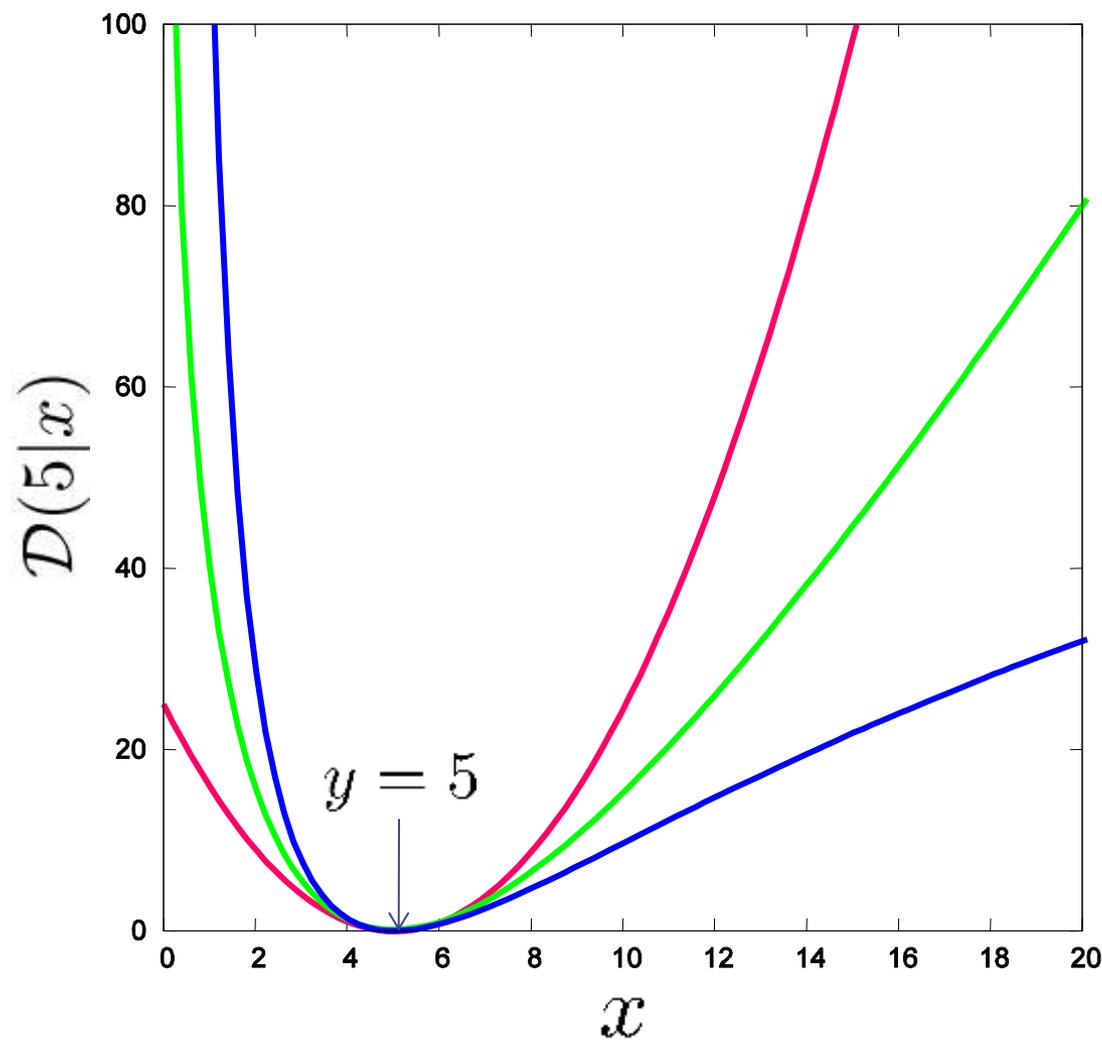
$$D_{\text{IS}}(a||b) = \frac{a}{b} - \log \frac{a}{b} - 1$$

但し、全要素の距離値の総和をコスト関数とするのが一般的

- NMF分解においてどのような特質を重視するかを左右
 - 例: スペクトルのピークの一致度を重要視した分解(KL, IS)
 - 例: 距離値がスケールに対して不変な分解(IS)
- より一般的な距離関数の β -divergenceというものもある
 - $\beta=0$ がIS, $\beta=1$ がKL, $\beta=2$ がEUCに対応する

二乗誤差、I ダイバージェンス、板倉斎藤擬距離の差

• $D(y|x)$: x の y からの近さの度合い



- 二乗誤差
 $D(x|y) = (y - x)^2$
- I ダイバージェンス
 $D(x|y) = y \log \frac{y}{x} - (y - x)$
- 板倉斎藤擬距離
 $D(x|y) = \frac{y}{x} - \log \frac{y}{x} - 1$

NMFにおける代表的な最適化規準

■ 二乗誤差

$$\begin{aligned} \text{minimize} \quad & \|Y - HU\|_F^2 = \sum_{\omega,t} \left| Y_{\omega,t} - \sum_k H_{\omega,k} U_{k,t} \right|^2 \\ \text{subject to} \quad & H_{\omega,k} \geq 0, U_{k,t} \geq 0 \end{aligned}$$

なんとかしたい部分

■ 1ダイバージェンス(一般化KLダイバージェンス)

$$\begin{aligned} \text{minimize} \quad & \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{\sum_k H_{\omega,k} U_{k,t}} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right) \\ \text{subject to} \quad & H_{\omega,k} \geq 0, U_{k,t} \geq 0 \end{aligned}$$

なんとかしたい部分

いずれも $Y = HU$ のとき 0 になる

押さえておくべき基本原理(1/2)

◆補助関数法

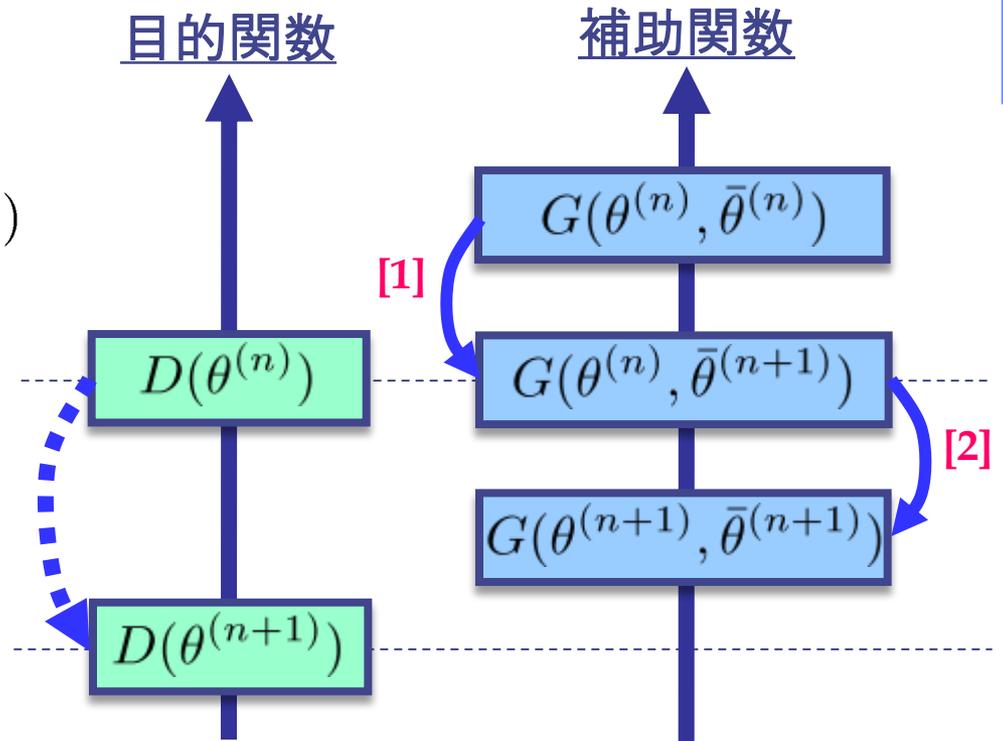
- $D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta})$ を満たす $G(\theta, \bar{\theta})$ を補助関数と定義
- 反復アルゴリズム

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

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

■ 収束性

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



目的関数を直接最小化するのが難しいなら、
とりあえずその上限関数を作ってみよう！

押さえておくべき基本原理(2/2)

■ Jensenの不等式

■ $\phi(\cdot)$: 凸関数

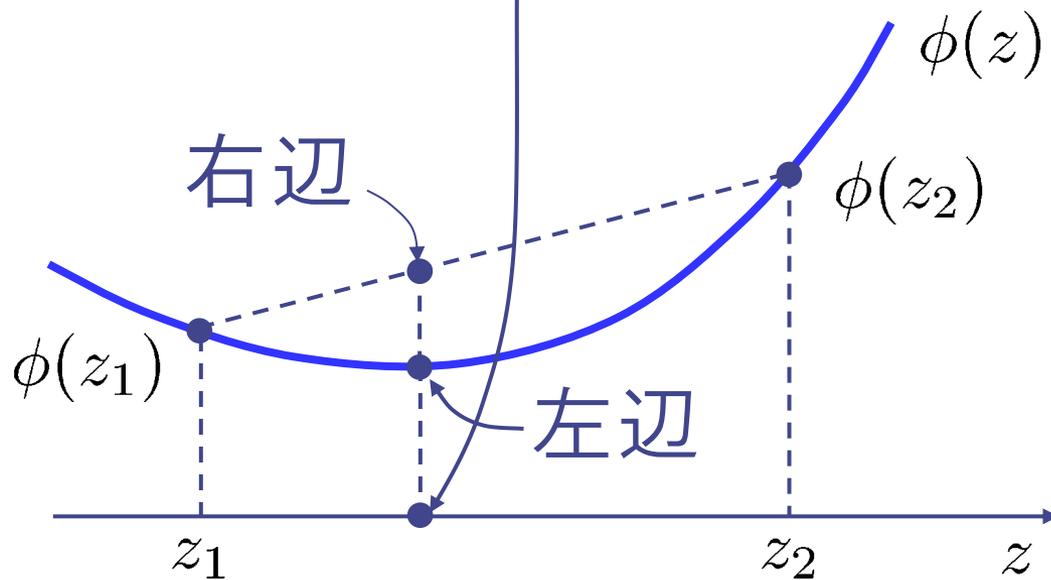
■ $\lambda_i \geq 0, \sum_i \lambda_i = 1$

$$\Rightarrow \phi\left(\sum_i \lambda_i z_i\right) \leq \sum_i \lambda_i \phi(z_i)$$

例えば,

$\phi(z) = -\log z$ の場合:

$$\begin{aligned} & -\log\left(\sum_i \lambda_i z_i\right) \\ & \leq -\sum_i \lambda_i \log z_i \end{aligned}$$



Frobeniusノルム規準のNMFアルゴリズム

- 目的関数

$$\begin{aligned}\|Y - HU\|_F^2 &= \sum_{\omega,t} \left| Y_{\omega,t} - \sum_i H_{\omega,i} U_{i,t} \right|^2 \\ &= \sum_{\omega,t} \left(|Y_{\omega,t}|^2 - 2Y_{\omega,t} \sum_k H_{\omega,k} U_{k,t} + \underbrace{\left| \sum_k H_{\omega,k} U_{k,t} \right|^2}_{\text{Jensen's inequality}} \right)\end{aligned}$$

- 下線部に対してJensenの不等式を立ててみる

$$\left(\sum_i z_i \right)^2 = \left(\sum_i \lambda_i \frac{z_i}{\lambda_i} \right)^2 \leq \sum_i \lambda_i \left(\frac{z_i}{\lambda_i} \right)^2 = \sum_i \frac{z_i^2}{\lambda_i}$$

適用 $\Rightarrow \left| \sum_k H_{\omega,k} U_{k,t} \right|^2 \leq \sum_k \frac{H_{\omega,k}^2 U_{k,t}^2}{\lambda_{k,\omega,t}}$

- 補助関数が完成

$H_{\omega,k}$ または $U_{k,t}$ ごとの二次関数の和になっている

$$\|Y - HU\|_F^2 \leq \sum_{\omega,t} \left(|Y_{\omega,t}|^2 - 2Y_{\omega,t} \sum_k H_{\omega,k} U_{k,t} + \sum_k \frac{H_{\omega,k}^2 U_{k,t}^2}{\lambda_{k,\omega,t}} \right)$$

Frobeniusノルム規準のNMFアルゴリズム

- 補助関数が完成したらあとはステップ1とステップ2を導出すればOK!

$$G := \sum_{\omega,t} \left(|Y_{\omega,t}|^2 - 2Y_{\omega,t} \sum_k H_{\omega,k} U_{k,t} + \sum_k \frac{H_{\omega,k}^2 U_{k,t}^2}{\lambda_{k,\omega,t}} \right)$$

[1] $\lambda_{k,\omega,t} \leftarrow \operatorname{argmin}_{\lambda_{k,\omega,t}} G = \frac{H_{\omega,k} U_{k,t}}{\sum_{k'} H_{\omega,k'} U_{k',t}}$ 代入

[2] $H_{\omega,k} \leftarrow \operatorname{argmin}_{H_{\omega,k}} G = \frac{\sum_t Y_{\omega,t} U_{k,t}}{\sum_t \frac{U_{k,t}^2}{\lambda_{k,\omega,t}}}$

$$H_{\omega,k} \leftarrow H_{\omega,k} \frac{\sum_t Y_{\omega,t} U_{k,t}}{\sum_t U_{k,t} \sum_{k'} H_{\omega,k'} U_{k',t}}$$

$$U_{k,t} \leftarrow \operatorname{argmin}_{U_{k,t}} G = \frac{\sum_{\omega} Y_{\omega,t} H_{\omega,k}}{\sum_{\omega} \frac{H_{\omega,k}^2}{\lambda_{k,\omega,t}}}$$

$$U_{k,t} \leftarrow U_{k,t} \frac{\sum_{\omega} Y_{\omega,t} H_{\omega,k}}{\sum_{\omega} H_{\omega,k} \sum_{k'} H_{\omega,k'} U_{k',t}}$$

ダイバージェンス規準のNMFアルゴリズム

- 目的関数

$$\begin{aligned}\mathcal{D} &:= \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{\sum_k H_{\omega,k} U_{k,t}} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right) \\ &= \sum_{\omega,t} \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \log \sum_k H_{\omega,k} U_{k,t} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right)\end{aligned}$$

- 下線部に対してJensenの不等式を立ててみる

$$-\log \left(\sum_i z_i \right) = -\log \left(\sum_i \lambda_i \frac{z_i}{\lambda_i} \right) \leq -\sum_i \lambda_i \log \left(\frac{z_i}{\lambda_i} \right)$$

適用 \rightarrow
$$-\log \sum_k H_{\omega,k} U_{k,t} \leq -\sum_k \lambda_{k,\omega,t} \log \frac{H_{\omega,k} U_{k,t}}{\lambda_{k,\omega,t}}$$

- 補助関数が完成

$$\mathcal{D} \leq \sum_{\omega,t} \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \sum_k \lambda_{k,\omega,t} \log \frac{H_{\omega,k} U_{k,t}}{\lambda_{k,\omega,t}} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right)$$

ダイバージェンス規準のNMFアルゴリズム

- 補助関数が完成したらあとはステップ1とステップ2を導出すればOK!

$$\mathcal{G} := \sum_{\omega,t} \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \sum_k \lambda_{k,\omega,t} \log \frac{H_{\omega,k} U_{k,t}}{\lambda_{k,\omega,t}} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right)$$

$$[1] \lambda_{k,\omega,t} \leftarrow \operatorname{argmin}_{\lambda_{k,\omega,t}} G = \frac{H_{\omega,k} U_{k,t}}{\sum_{k'} H_{\omega,k'} U_{k',t}} \quad \text{代入}$$

$$[2] H_{\omega,k} \leftarrow \operatorname{argmin}_{H_{\omega,k}} G = \frac{\sum_t Y_{\omega,t} \lambda_{k,\omega,t}}{\sum_t U_{k,t}} \quad H_{\omega,k} \leftarrow H_{\omega,k} \frac{\sum_t \frac{Y_{\omega,t} U_{k,t}}{\sum_{k'} H_{\omega,k'} U_{k',t}}}{\sum_t U_{k,t}}$$

$$U_{k,t} \leftarrow \operatorname{argmin}_{U_{k,t}} G = \frac{\sum_{\omega} Y_{\omega,t} \lambda_{k,\omega,t}}{\sum_{\omega} H_{\omega,k}} \quad U_{k,t} \leftarrow U_{k,t} \frac{\sum_{\omega} \frac{Y_{\omega,t} H_{\omega,k}}{\sum_{k'} H_{\omega,k'} U_{k',t}}}{\sum_{\omega} H_{\omega,k}}$$

統計モデルとしての解釈

■ NMFは以下を仮定した最尤推定問題と等価

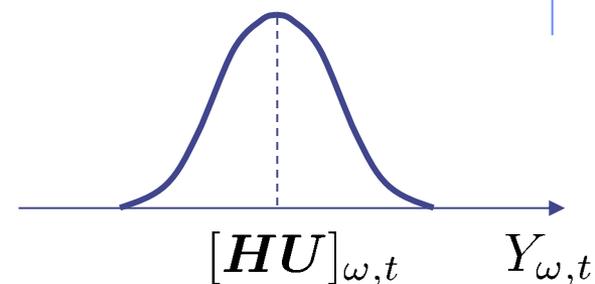
■ Frobeniusノルム規準

$$Y_{\omega,t} \sim \mathcal{N}([HU]_{\omega,t}, \sigma^2)$$

$$P(\mathbf{Y}|\mathbf{H}, \mathbf{U}) = \prod_{\omega,t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y_{\omega,t} - [HU]_{\omega,t})^2}$$

$$\rightarrow -\log P(\mathbf{Y}|\mathbf{H}\mathbf{U}) \stackrel{c}{=} \frac{1}{2\sigma^2} \sum_{\omega,t} (Y_{\omega,t} - [HU]_{\omega,t})^2$$

正規分布



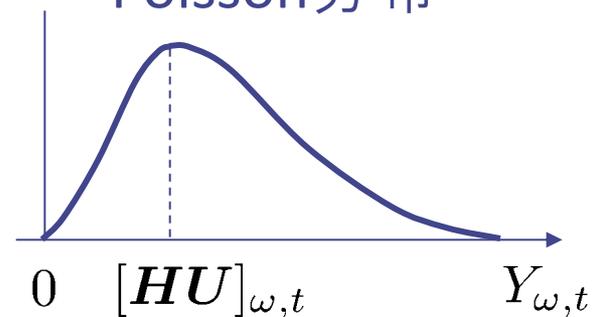
■ 1ダイバージェンス規準

$$Y_{\omega,t} \sim \text{Poisson}([HU]_{\omega,t})$$

$$P(\mathbf{Y}|\mathbf{H}, \mathbf{U}) = \prod_{\omega,t} \frac{[HU]_{\omega,t}^{Y_{\omega,t}} e^{-[HU]_{\omega,t}}}{Y_{\omega,t}!}$$

$$\rightarrow -\log P(\mathbf{Y}|\mathbf{H}\mathbf{U}) \stackrel{c}{=} \sum_{\omega,t} (-Y_{\omega,t} \log [HU]_{\omega,t} + [HU]_{\omega,t})$$

Poisson分布



板倉斎藤距離規準NMF [Févotte 2008]

- 観測信号をガウス性信号の重ね合わせとしてモデル化
- 各要素信号のパワースペクトログラムがRank-1構造をもつと仮定

$$C_{k,\omega,t} \sim \mathcal{N}_{\mathbb{C}}(0, H_{\omega,k}U_{k,t})$$

$$Y_{\omega,t} = \sum_k C_{k,\omega,t}$$

→ ここでの $Y_{\omega,t}$ と $C_{k,\omega,t}$ は複素スペクトル成分(複素数値)であることに注意。



$$Y_{\omega,t} \sim \mathcal{N}_{\mathbb{C}}\left(0, \sum_k H_{\omega,k}U_{k,t}\right)$$

→ $C_{k,\omega,t}$ と $C_{k',\omega,t}$ の独立性を仮定



$$\begin{aligned} \mathcal{J} &= \sum_{\omega,t} \mathcal{D}\left(|Y_{\omega,t}|^2 \middle| \sum_k H_{\omega,k}U_{k,t}\right) \\ &= \sum_{\omega,t} \left(\frac{|Y_{\omega,t}|^2}{\sum_k H_{\omega,k}U_{k,t}} - \log \frac{|Y_{\omega,t}|^2}{\sum_k H_{\omega,k}U_{k,t}} - 1 \right) \end{aligned}$$

板倉斎藤距離規準のNMFアルゴリズム [Kameoka 2006]

- 目的関数

$$\mathcal{D}_{\text{IS}}(\mathbf{H}, \mathbf{U}) = \sum_{\omega, t} \left(\frac{|Y_{\omega, t}|^2}{\sum_k H_{\omega, k} U_{k, t}} + \log \sum_k H_{\omega, k} U_{k, t} - \dots \right)$$

- 逆数関数は凸関数 \Rightarrow Jensenの不等式

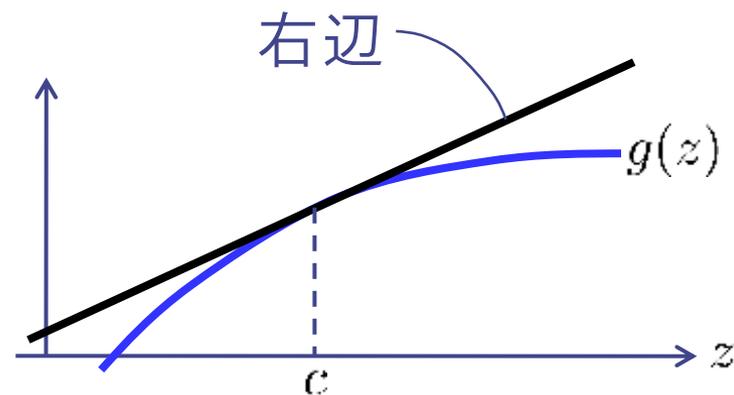
$$\frac{1}{\sum_i z_i} = \frac{1}{\sum_i \lambda_i \frac{z_i}{\lambda_i}} \leq \sum_i \lambda_i \frac{1}{\frac{z_i}{\lambda_i}} = \sum_i \frac{\lambda_i^2}{z_i}$$

適用 \Rightarrow
$$\frac{1}{\sum_k H_{\omega, k} U_{k, t}} \leq \sum_k \frac{\lambda_{k, \omega, t}^2}{H_{\omega, k} U_{k, t}}$$

- 対数関数は凹関数 \Rightarrow この場合はどうしたら??

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

適用 \Rightarrow
$$\log \sum_k H_{\omega, k} U_{k, t} \leq \frac{1}{c_{\omega, t}} \left(\sum_k H_{\omega, k} U_{k, t} - c_{\omega, t} \right) + \log c_{\omega, t}$$



NMFアルゴリズムの(できるだけ)一般化

◆ 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ダイバージェンスにおいて

$$\varphi(z) = \frac{1}{(\beta - 1)\beta} (z^{\beta} - \beta z + \beta - 1) \text{ と置いたもの}$$

例) 二乗誤差, 1ダイバージェンス, 板倉斎藤擬距離

- $\beta = 2$ $D_{\beta}(y, x) = \frac{y^2}{2} + \frac{x^2}{2} - yx = \frac{1}{2}(y - x)^2$
- $\beta \rightarrow 1$ $D_{\beta}(y, x) \rightarrow y \log \frac{y}{x} + x$ $\because \lim_{\beta \rightarrow 1} \frac{yx^{\beta-1}}{\beta - 1} = \lim_{\beta \rightarrow 1} yx^{\beta-1} \log x = y \log x$
- $\beta \rightarrow 0$ $D_{\beta}(y, x) \rightarrow \frac{y}{x} - \log \frac{y}{x}$ $\because \lim_{\beta \rightarrow 0} \frac{x^{\beta}}{\beta} = \lim_{\beta \rightarrow 0} x^{\beta} \log x = \log x$

最初の β ダイバージェンス規準NMF [Kompas2007]

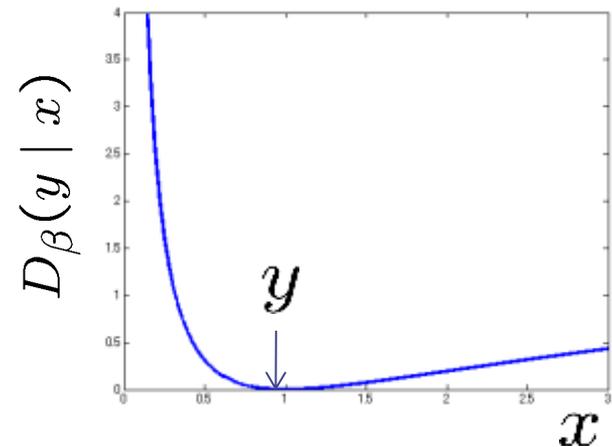
- β ダイバージェンス[Eguchi2001]を局所最小化するアルゴリズム

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

$\beta = 0$: 板倉齋藤距離

◀ $\beta = 1$: I ダイバージェンス

$\beta = 2$: 二乗誤差



$$H \leftarrow H \cdot \frac{(Y \cdot (HU)^{\beta-2})U^T}{(HU)^{\beta-1}U^T} \quad U \leftarrow U \cdot \frac{H^T(Y \cdot (HU)^{\beta-2})}{H^T(HU)^{\beta-1}}$$

$1 \leq \beta \leq 2$ の場合しか収束性が保証されない

補助関数版 β ダイバージェンス基準NMF

[M. Nakano et al., 2010]

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

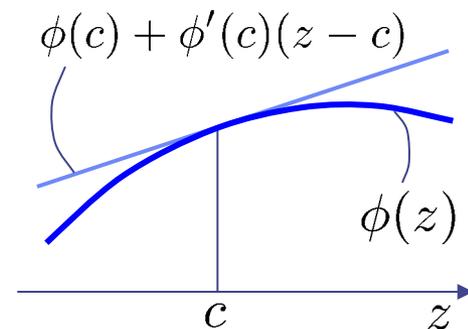
$$\text{subject to} \quad H_{\omega,i} \geq 0, U_{i,t} \geq 0 \quad (1 \leq \omega \leq \Omega, 1 \leq t \leq T)$$

◆ 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]

補助関数版βダイバージェンス基準NMF

[M. Nakano et al., 2010]

◆補助関数

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

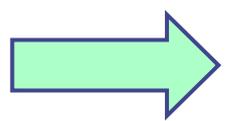
$$-\frac{1}{\beta-1} \left(\sum_i H_{\omega,i} U_{i,t} \right)^{\beta-1} \leq \begin{cases} \text{⏏} \left\{ -\frac{1}{\beta-1} \sum_i \lambda_{i,\omega,t} \left(\frac{H_{\omega,i} U_{i,t}}{\lambda_{i,\omega,t}} \right)^{\beta-1} \right. & (\beta \leq 2, \beta \neq 1) \\ \text{⏏} \left\{ -\frac{1}{\beta-1}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

(例) $0 < \beta < 1$ のケース:

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

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

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



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

補助関数版 β ダイバージェンス基準NMF

[M. Nakano et al., 2010]

◆補助関数 (例) $0 < \beta < 1$ のケース:

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

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

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

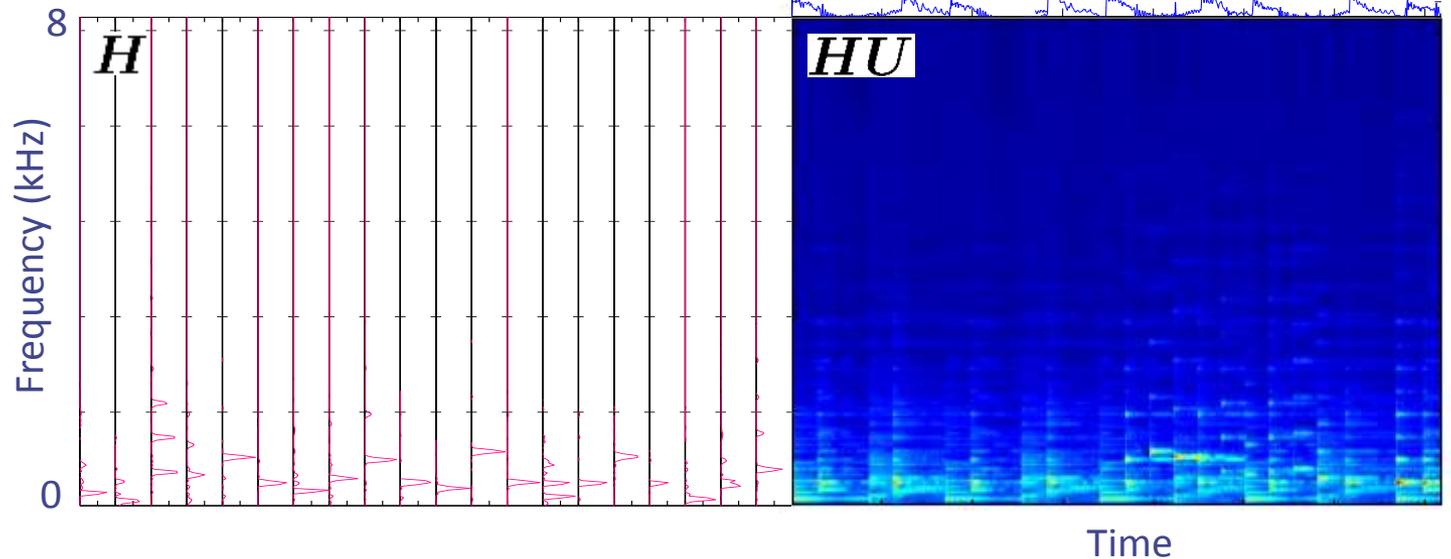
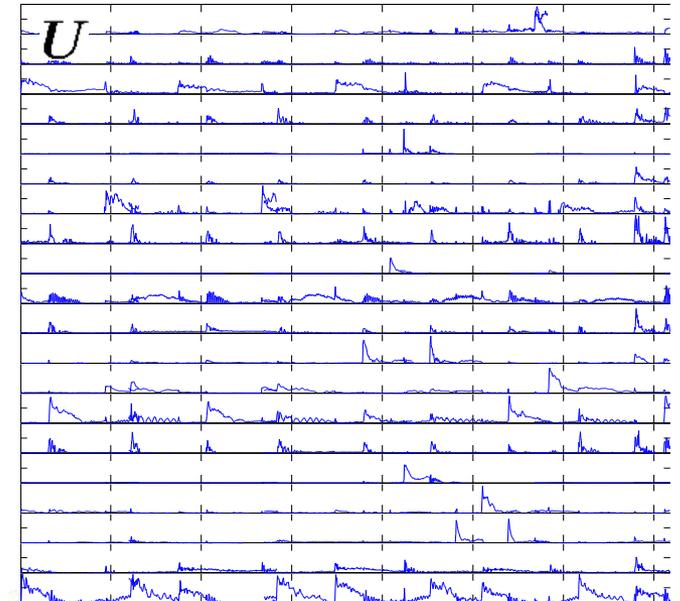
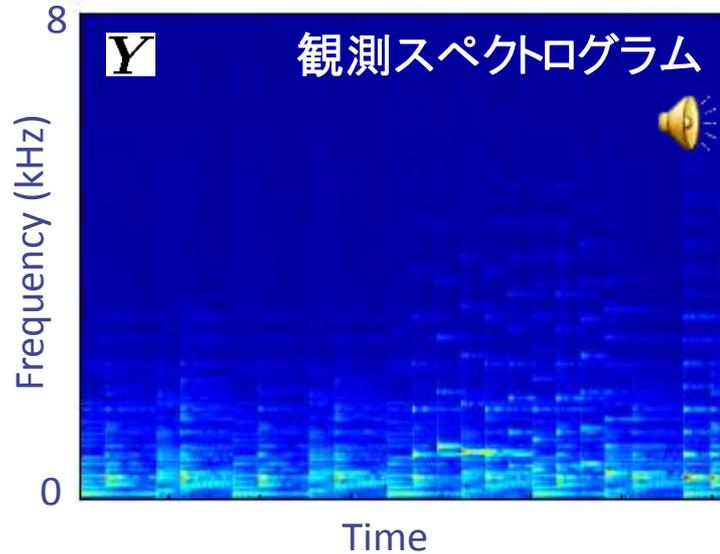
$$\left\{ \begin{aligned} \frac{\partial G(\theta, \bar{\theta})}{\partial H_{\omega, i}} &= \sum_t \left[c_{\omega, t}^{(n+1)\beta-1} U_{i, t}^{(n)} - \frac{Y_{\omega, t} U_{i, t}^{(n)\beta-1}}{\lambda_{i, \omega, t}^{(n+1)\beta-2}} H_{\omega, i}^{\beta-2} \right] = 0 \\ \frac{\partial G(\theta, \bar{\theta})}{\partial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

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

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

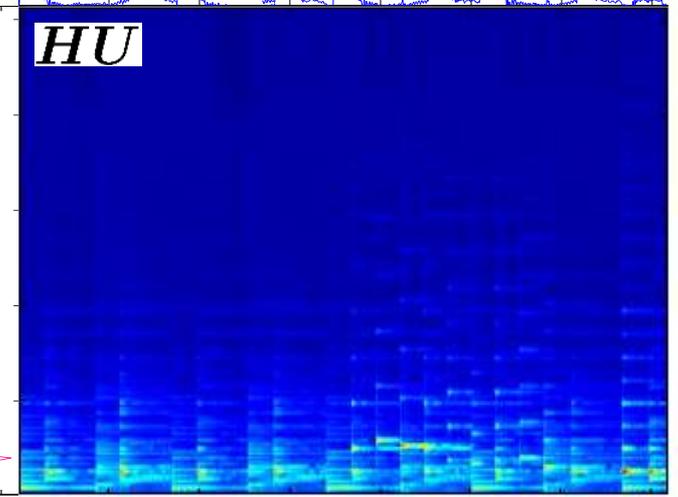
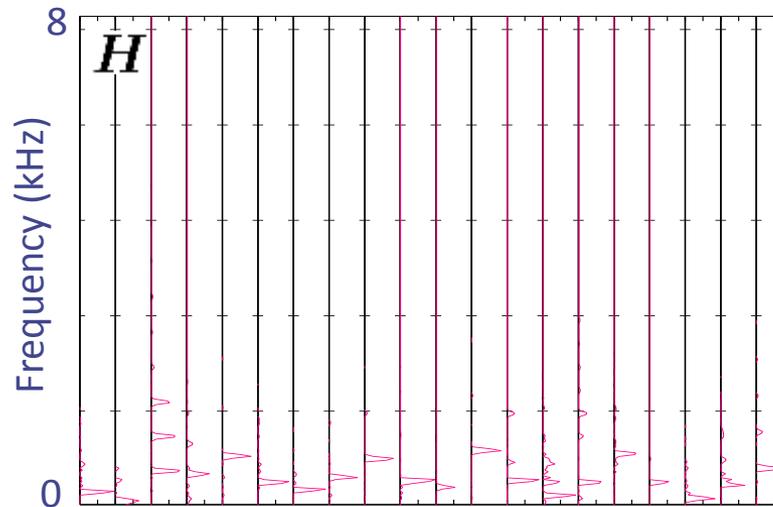
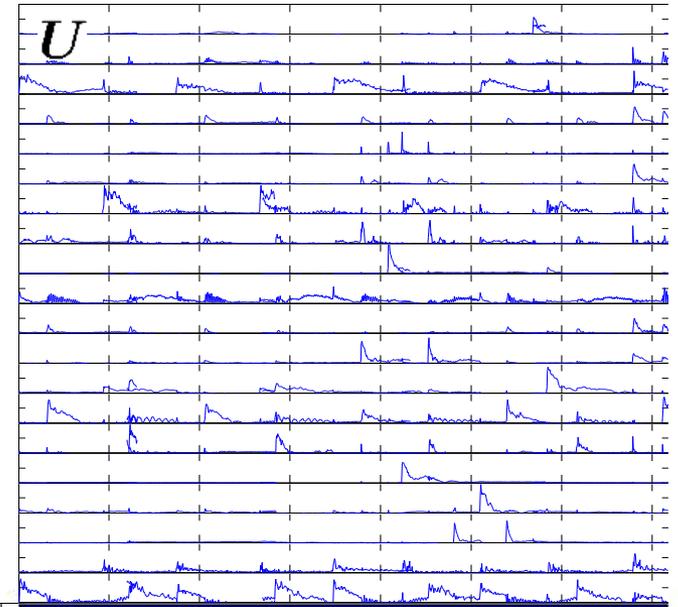
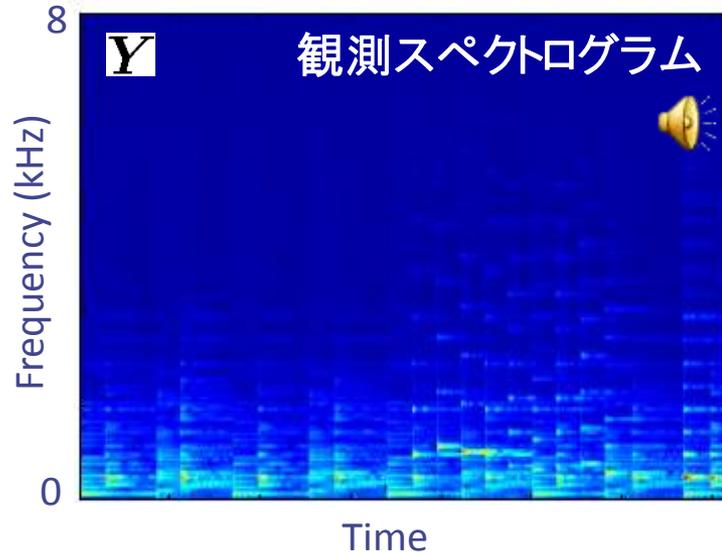
NMFの規準別の比較

■ 二乗誤差規準



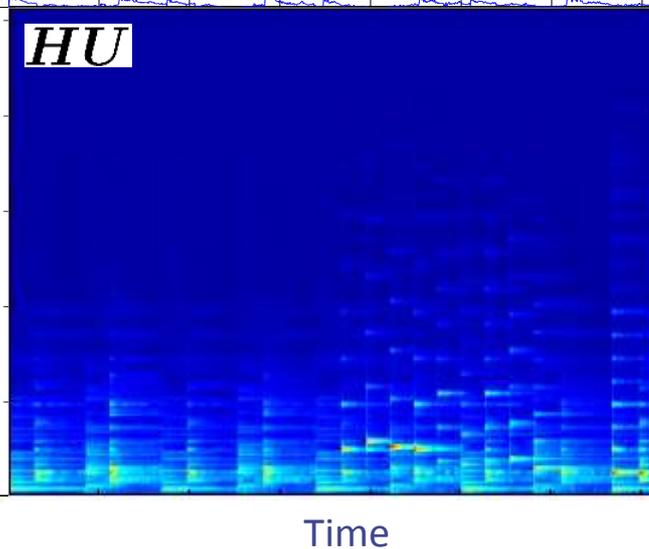
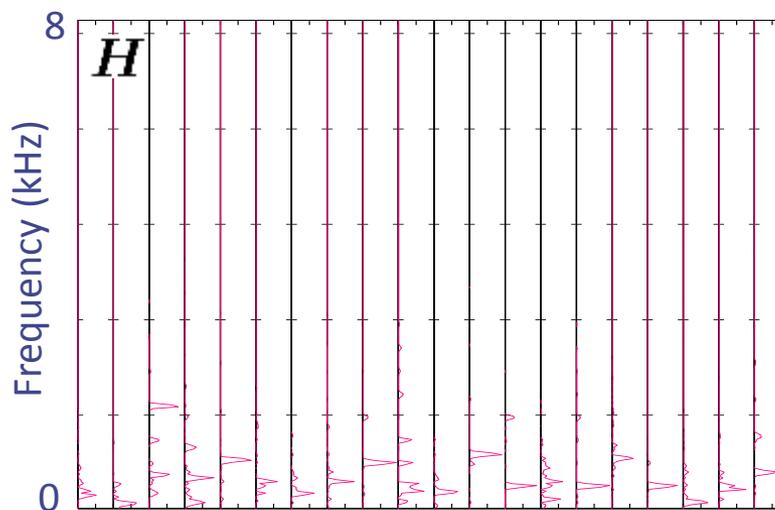
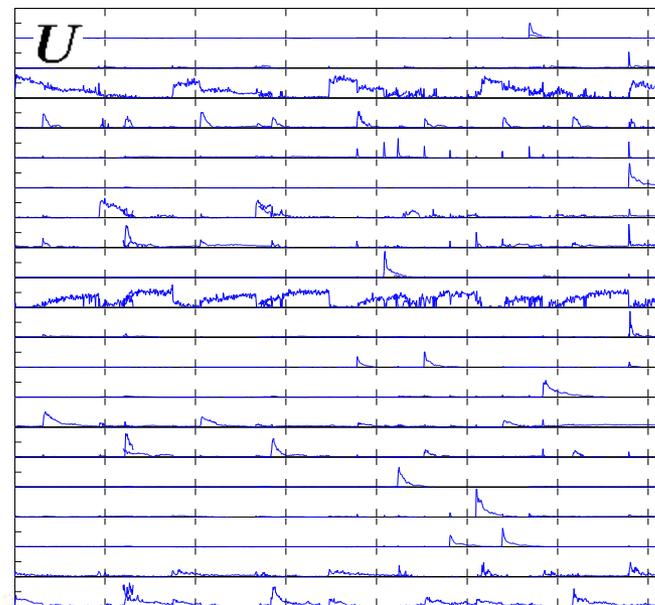
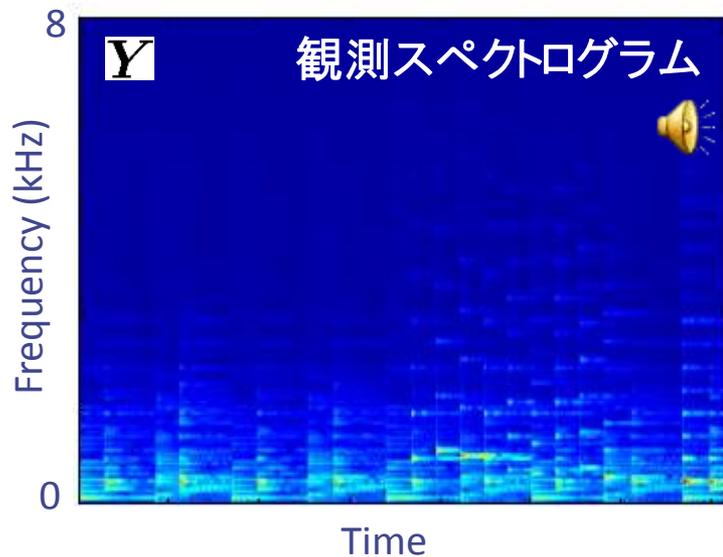
NMFの規準別の比較

■ I ダイバージェンス規準



NMFの規準別の比較

■ 板倉斎藤距離規準



NMFの規準別の比較(まとめ)

- データと変数のフィッティングを、**どの確率分布をもって測っていくか**を制御している
 - その上で期待値(変数)を求める最尤推定問題を解いている
 - 一般に、 β の小さな場合ほど、スパースな基底・アクティベーションが得られる傾向にあり
- 板倉斎藤擬距離基準NMFに関しては**複素数の観測データ**に対する別の生成モデル的な解釈が存在する
 - 複素観測データが**複素平面で球対称なガウス分布(複素ガウス分布)**から生成
 - 観測データが**複数(基底数)個の独立な複素ガウス分布に従う確率変数の線形結合**であることを仮定(ガウス分布の再生性)
 - IS-NMFによる基底分解の(期待値の意味での)妥当性を保証

板倉斎藤擬距離NMF再考(1)

従来のNMF分解 (EUC, KL, etc.) 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

$$Y = F_1 G_1 + F_2 G_2 + F_3 G_3 + \dots$$

- Y 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を与える
- **複素生成モデル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NMF再考(2)

- では、パワードメインのIS-NMFがスペクトログラムの分解や音源分離にとって最良な結果を与えるのか？
 - 否！
 - 例えば音源分離タスクにおいては「振幅ドメインのKL-NMFが良い」といろいろな文献で報告されている
 - Ex. [D. FitzGerald, et al., 2009], [D. Kitamura, et al., 2014]
 - 場当たりの的に振幅ドメインでIS-NMFを適用することもしばしば
 - 理論的妥当性とはなんだったのか…
- 他に加法に関して「妥当」なものはあるのか？（例えば振幅ドメインで理論的に妥当なNMFとか）
 - Cauchy NMF [A. Liutkus, et al., 2015]
 - コーシー分布の再生性を用いて振幅ドメインでの加法性が（期待値の意味で）理論的に保証されたNMF

Cauchy NMF [A. Liutkus et al., 2015]

コーシーNMFの生成モデル

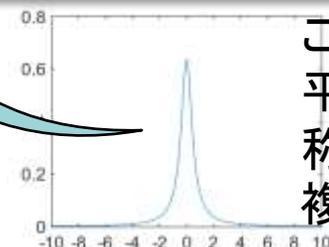
とある時間
周波数要素

$$x_{\omega,t} = \sum_{k=1}^K c_{\omega,t,k}$$

中心値0, スケール $f_{\omega,k}g_{k,t}$ の球対称コーシー
 $c_{\omega,t,k} \sim \mathcal{C}_c(0, f_{\omega,k}g_{k,t})$

$$x_{\omega,t} = \underbrace{c_{\omega,t,1} + c_{\omega,t,2} + \dots + c_{\omega,t,K}}_{\text{これらの複素コーシー分布は互いに独立}}$$

これらの複素コーシー分布は互いに独立



これを複素
平面で球対
称に回した
複素分布

– 複素コーシー分布の線形結合なので $x_{\omega,t}$ も複素コーシー分布

- コーシー分布の再生性

- $x_{\omega,t}$ の複素コーシー分布のスケールは $\sum_{k=1}^K f_{\omega,k}g_{k,t}$

– スケールが時間周波数で変動する複素コーシー分布生成系

最小化規範

$$\mathcal{D}(|\mathbf{X}| || \mathbf{F}\mathbf{G}) = \sum_{\omega,t} \left[\frac{3}{2} \log (|x_{\omega,t}|^2 + (\sum_k f_{\omega,k}g_{k,t})^2) - \log \sum_k f_{\omega,k}g_{k,t} \right]$$

Cauchy NMF [A. Liutkus et al., 2015]

- 音源分離タスクにおいてパワードメインのISNMFよりは良い性能をマーク

- しかし依然として振幅ドメインのKLには及ばず

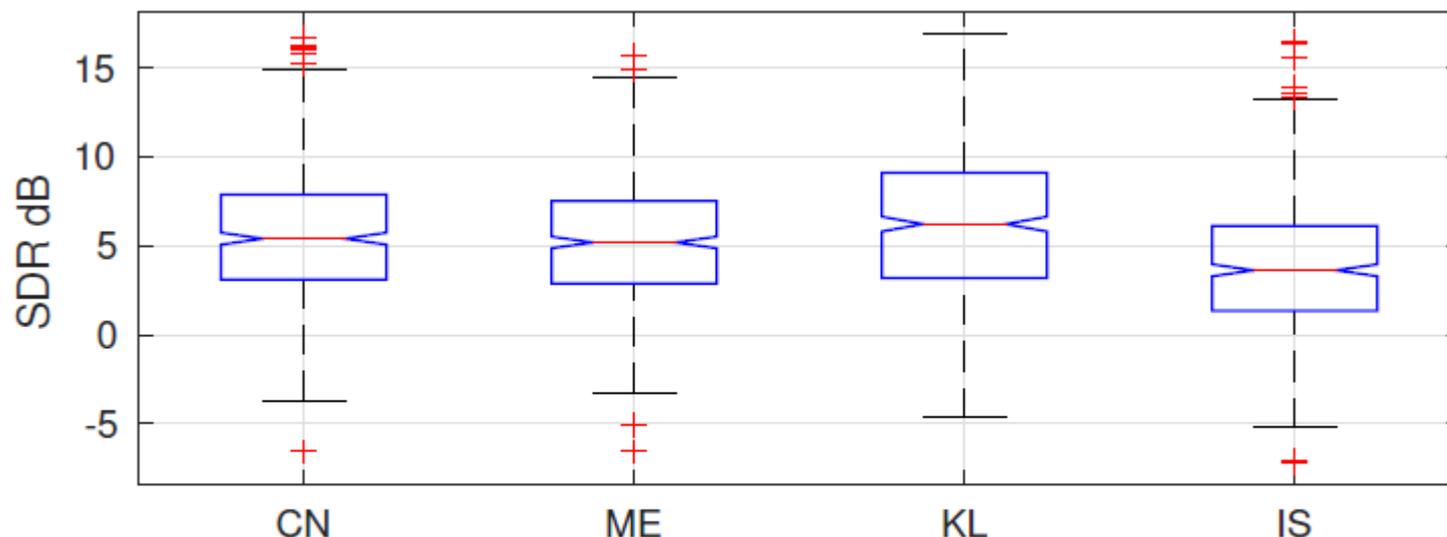


Figure 1. SDR for Cauchy Naive (CN), Cauchy ME (ME), KL and IS NMF algorithms (10 runs). Higher is better

- 依然として「妥当な加法生成モデルに基づくNMF」は色々試されている(複素NMFのところでも後述する)

「乗算型」更新式の意義

- 補助関数法によって得られた更新式を眺めると...

$$H_{\omega,k} \leftarrow H_{\omega,k} \frac{\sum_t Y_{\omega,t} U_{k,t}}{\sum_t U_{k,t} \sum_{k'} H_{\omega,k'} U_{k',t}} \quad U_{k,t} \leftarrow U_{k,t} \frac{\sum_{\omega} Y_{\omega,t} H_{\omega,k}}{\sum_{\omega} H_{\omega,k} \sum_{k'} H_{\omega,k'} U_{k',t}}$$

- 非負値の係数を「乗算」の形で更新する⇒更新後も非負！
- 一般の最急降下法(加法更新)ではこうはいかない。

(1) 加法型更新式:
$$x^{(i+1)} \leftarrow x^{(i)} - \eta \frac{\partial f(x^{(i)})}{\partial x}$$

(2) 乗法型更新式:
$$x^{(i+1)} \leftarrow x^{(i)} \cdot g(f(x^{(i)}))$$

- 事前知識(構造)の導入

- 乗法更新の特徴として「零値」の保存性がある(あるパラメータを一旦ゼロにすると、以後そのままゼロに)
- 事前知識や外部情報で得られる拘束条件を付与可能
(例)楽譜情報により「この時刻は絶対に音は無い」

User-Guided Source Separation [Smaragdakis 2009, Ozerov 2011]

本日の話題

- 非負値行列因子分解(NMF)の基礎
 - 何に使えるのか(音響信号処理を題材として)
 - どのような性質があるのか
 - どのように求めるのか
 - 統計モデルとしての解釈

- 様々なNMF及びその応用例
 - PLSAとの関係
 - スパースNMF、ベイジアンNMF、複素NMF etc.
 - 音楽信号分析例

NMF と pLSA [T. Hofmann, 1999] との関係

(probabilistic Latent Semantic Analysis)

- 各行列要素を確率値だと考えてみよう(非負値だから)

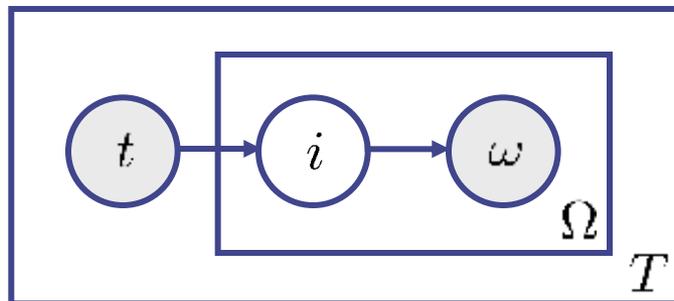
$$\mathbf{X} = \mathbf{H} \mathbf{U}$$

$P(\omega|t)$ 文書 t に単語 ω が現われる確率

$P(\omega|i)$ トピック i のとき単語 ω が生成される確率

$P(i|t)$ 文書 t がトピック i のものである確率

$$\rightarrow P(\omega|t) = \sum_i P(\omega|i)P(i|t)$$



NMF と pLSA [T. Hofmann, 1999] との関係

(probabilistic Latent Semantic Analysis)

- pLSAでは各文書中の単語数 $N_{\omega,t}$ から $P(\omega|i)$, $P(i|t)$ を推定

$$\log P(\omega|t)^{N_{\omega,t}} \rightarrow \text{maximize}$$

$$\begin{aligned} &\text{maximize} && \sum_{\omega,t} N_{\omega,t} \log P(\omega|t) \\ &\text{subject to} && P(\omega|i) \geq 0, P(i|t) \geq 0 \\ &&& \sum_{\omega} P(\omega|i) = 1, \sum_i P(i|t) = 1 \end{aligned}$$

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

スパースNMF [Hoyer2004]

■ アクティベーションを強制的にスパース化

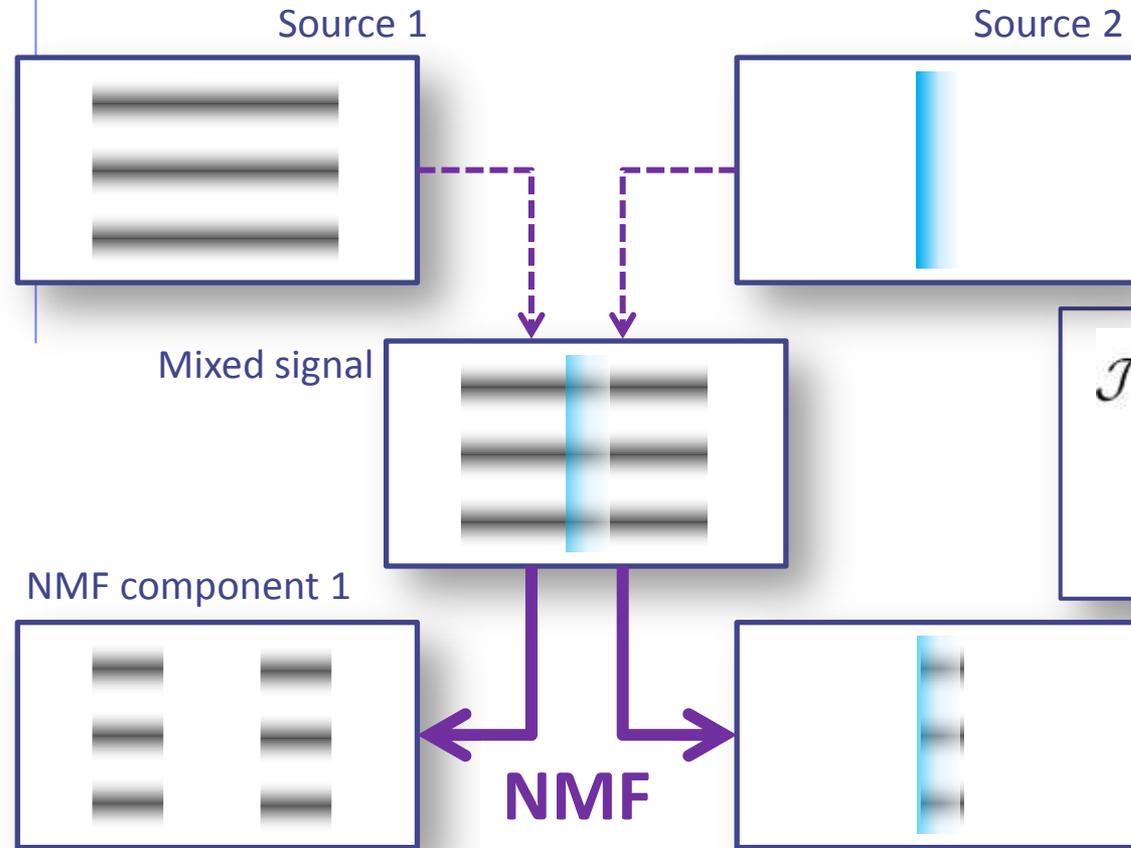
$$\begin{aligned} & \text{minimize} \quad \sum_{\omega,t} \mathcal{D}\left(Y_{\omega,t} \mid \sum_i H_{\omega,i} U_{i,t}\right) \\ & \text{subject to} \quad \frac{\sqrt{T} - \sum_{k,t} |U_{k,t}| / \sqrt{\sum_{k,t} U_{k,t}^2}}{\sqrt{T} - 1} = S_U \end{aligned}$$

スパース性尺度(1のときスパース, 0のとき非スパース)



アクティベーションの連続性規準入りNMF [Virtanen2007]

- 楽器音のパワーエンベロープは滑らかという仮定



そこで、アクティベーションの連続性規準を導入・・・

$$\mathcal{J} = \sum_{\omega,t} \mathcal{D}\left(Y_{\omega,t} \mid \sum_k H_{\omega,k} U_{k,t}\right) + \sum_{k,t} (U_{k,t} - U_{k,t-1})^2$$

- アクティベーションが滑らかなほど小さい値をとるコスト関数
- この項を含めて最適化を行うことで、左図のようなことが起こりにくなる

音源のストリームが分断されてしまっている！

別音源のストリームが混入してしまっている！

ベイジアンNMF [Cemgil2008]

- Iダイバージェンス規準NMFを生成モデルの観点から解釈

$$\begin{aligned}\mathcal{J} &= \sum_{\omega,t} \mathcal{D}\left(Y_{\omega,t} \mid \sum_k H_{\omega,k} U_{k,t}\right) \\ &= \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{\sum_k H_{\omega,k} U_{k,t}} - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right)\end{aligned}$$



$$Y_{\omega,t} \sim \text{Poisson}\left(\sum_k H_{\omega,k} U_{k,t}\right)$$

$$\text{Poisson}(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$



$$\begin{aligned}C_{\omega,t,k} &\sim \text{Poisson}(H_{\omega,k} U_{k,t}) \\ Y_{\omega,t} &= \sum_k C_{\omega,t,k}\end{aligned}$$

$$H_{\omega,k} \sim \text{Gamma}(a, b)$$

$$U_{k,t} \sim \text{Gamma}(c, d)$$

ベイズノンパラメトリックNMF [Hoffman2011, Nakano2011]

■ NMFにおける基底数をデータから推論できるようにしたい

■ Infinite NMF [Schmidt2010]

■ Gamma Process (GaP)-NMF [Hoffman2011]

◆ スペクトログラムの生成モデル

$$H_{\omega,k} \sim \text{Gamma}(a, a)$$

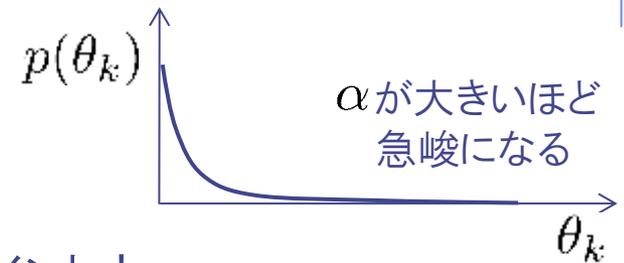
$$U_{k,t} \sim \text{Gamma}(b, b)$$

$$\theta_k \sim \text{Gamma}(\alpha/K, \alpha c) \longrightarrow \text{ここがポイント！}$$

$$Y_{\omega,t} \sim \text{Exponential}\left(\sum_k \theta_k H_{\omega,k} U_{k,t}\right) \rightarrow \text{FévotteのIS-NMF方式}$$

◆ $\theta_l \sim \text{Gamma}(\alpha/K, \alpha c)$ の意味

$\theta_1, \theta_2, \dots$ をスパースにする = 不要なコンポーネントはできるだけ使わない



■ 状態遷移NMFにおけるコンポーネント数と状態数をデータから推論できるようにしたい

■ Infinite Factorial infinite Hidden Markov Model (iFiHMM) [Nakano2011]

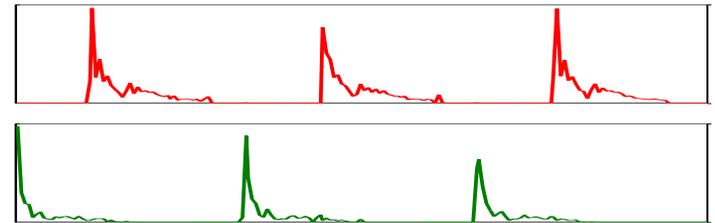
Nonnegative Matrix Factor Deconvolution [Smaragdis2004]

- ◆ スペクトログラム素片パーツを時間方向に畳み込んで
楽音スペクトログラムを表現

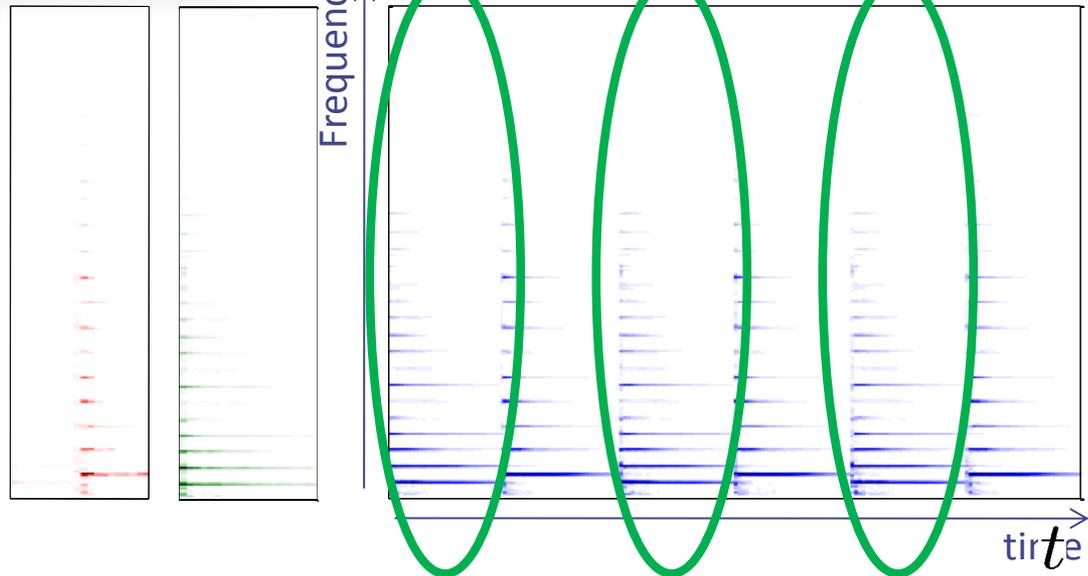
$$Y_{\omega,t} \simeq \sum_{k,\phi} H_{\omega,\phi,k} U_{k,t-\phi}$$

スペクトログラム素片 オンセット

オンセット



Frequency ξ

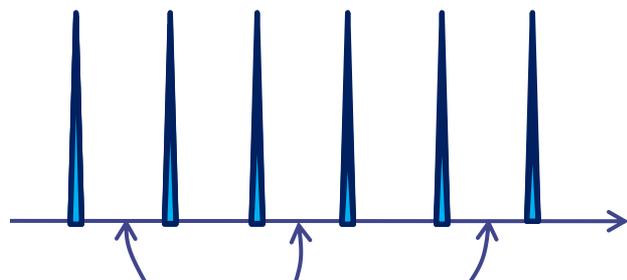
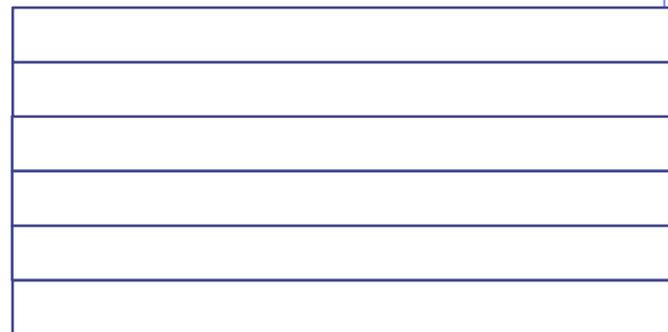


スペクトログラム
素片パーツ

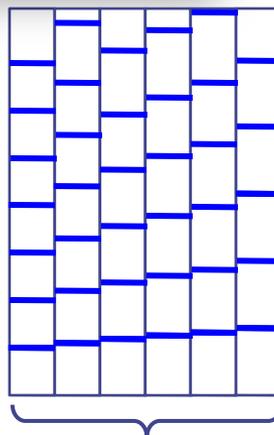
ハーモニックNMF [Raczynski2007]

- 調波構造形の基底スペクトル
- アクティベーションの無相関規準

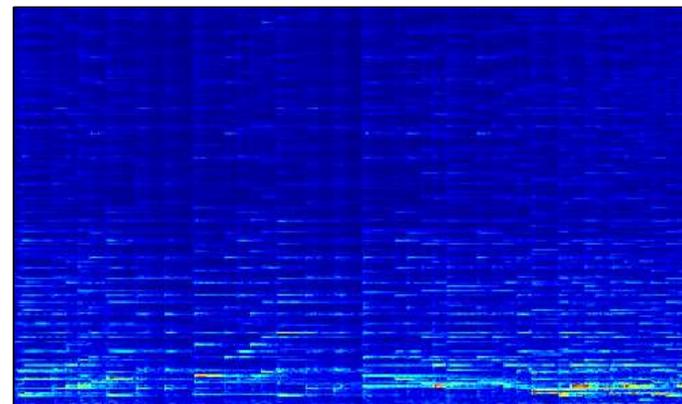
$$\mathcal{J} = \sum_{\omega,t} \mathcal{D} \left(Y_{\omega,t} \middle| \sum_k H_{\omega,k} U_{k,t} \right) + \sum_{k \neq j} \sum_t (U_{k,t} - U_{j,t})^2$$



基本周波数・高調波周波数
以外の成分を0に初期設定

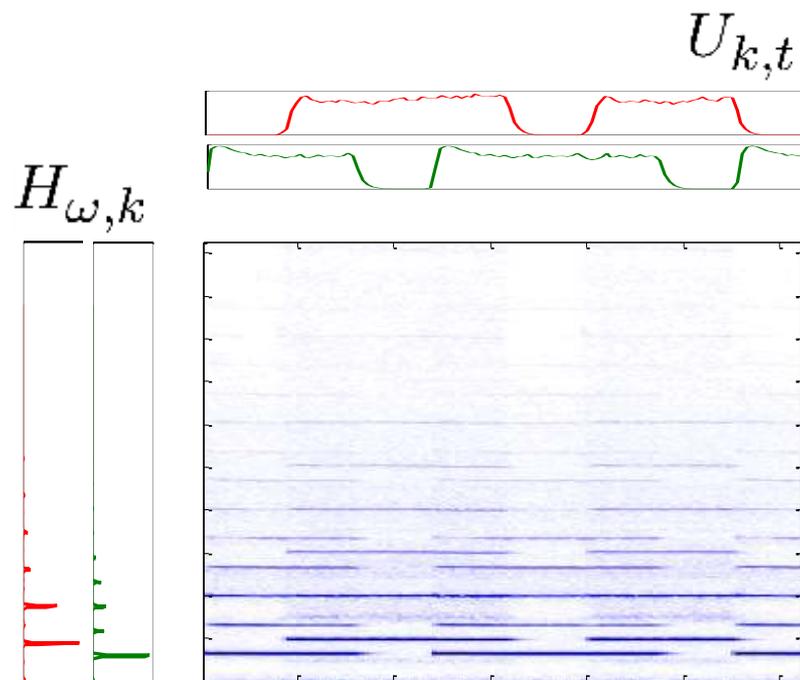
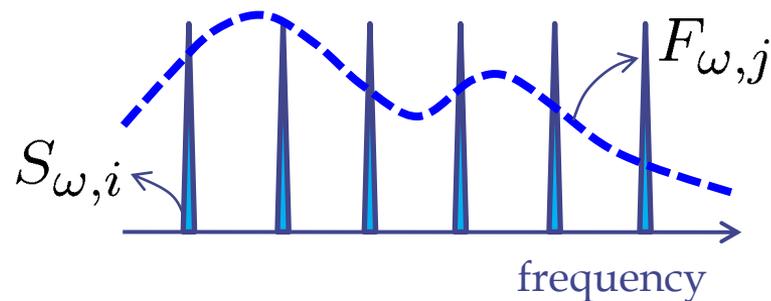


音名に対応する分だけ基底スペクトルを用意



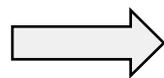
ソースフィルタモデルに基づくNMF [Kameoka 2009]

- $\{i, j\}$ 番目の基底スペクトルを i 番目のソーススペクトルと j 番目の全極フィルタの積で表現



従来NMF:

$$\sum_k H_{\omega, k} U_{k, t}$$



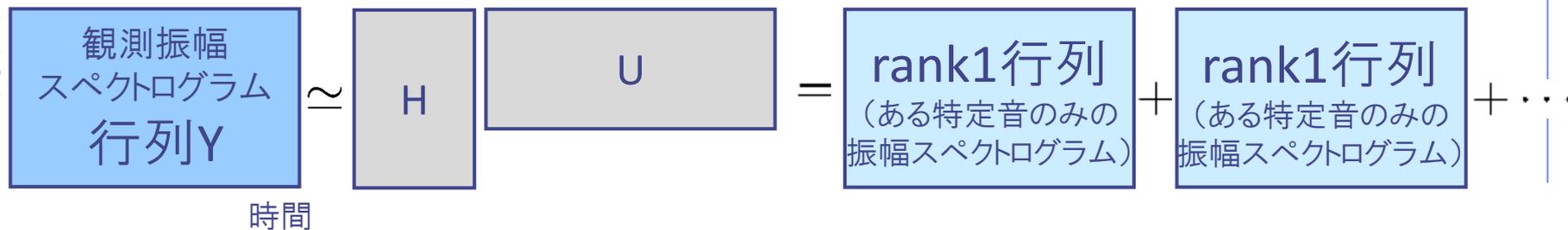
ソースフィルタNMF:

$$\sum_{i, j} \underline{S_{\omega, i} F_{\omega, j}} U_{i, j, t}$$

状態遷移NMF [Ozerov2009, Nakano2010, Mysore2010]

- NMF: スペクトログラムをrank1行列の和に分解していることに相当

周波数



- 楽音のスペクトログラムは実際にはrank1にならない
(ピアノ音にattack/decay/sustainなどの複数状態があるように,
通常, 楽音スペクトルの形は発音中に多様に時間変化する)
- 各楽音スペクトログラムをrank1行列でモデル化する代わりに
非負値を出力する隠れマルコフモデル(HMM)でモデル化
→ Factorial HMM [Ghahramani1997]と同形のモデルになる

A. Ozerov, C. Févotte, M. Charbit, "Factorial scaled hidden Markov model for polyphonic audio representation and source separation," in *Proc. WASPAA'09*, 2009.

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," 情報処理学会研究報告, 2010-MUS-84-10, Feb. 2010.

G. J. Mysore, P. Smaragdis, B. Raj, "Non-negative hidden Markov modeling of audio with application to source separation," in *Proc. LVA/ICA'2010*. pp.140-148, 2010.

複素NMF [Kameoka2008]

- パワースペクトルも振幅スペクトルも本来は非加法的
- 複素スペクトル領域モデルでNMFと同様のスペクトルパーツ獲得機能が実現できないか？

$$\begin{aligned} x(u) &= \sum_k c^{(k)}(u) \\ &\quad \downarrow \text{(STFT)} \\ X_{\omega,t} &= \sum_k C_{\omega,t,k} \\ &\quad \downarrow \text{(magnitude)} \\ |X_{\omega,t}| &= \left| \sum_k C_{\omega,t,k} \right| \\ &\neq \sum_k |C_{\omega,t,k}| \end{aligned}$$

NMFにおけるモデルが厳密には不適切であることを示している

要素信号の複素スペクトログラム $C_{\omega,t,k}$ の振幅成分が“rank1”構造をもつように混合信号をモデル化

$$\begin{aligned} X_{\omega,t} &= \sum_k |C_{\omega,t,k}| e^{j\phi_{\omega,t,k}} \\ &= \sum_k H_{\omega,k} U_{k,t} e^{j\phi_{\omega,t,k}} \end{aligned}$$

$$\sum_{\omega,t} |Y_{\omega,t} - X_{\omega,t}|^2 \rightarrow \text{minimum}$$

ここでの $Y_{\omega,t}$ と $C_{\omega,t,k}$ も複素スペクトル成分(複素数値)であることに注意。

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...線形な変換
(時間周波数分解)

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

↓ 絶対値をとる ...非線形な変換

$|Y_{\omega,t}|$: 時刻 t に周波数 ω の成分が
どれほど含まれているか \longrightarrow $|Y_{\omega,t}| = \left| \sum_i X_{\omega,t}^i \right|$

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

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

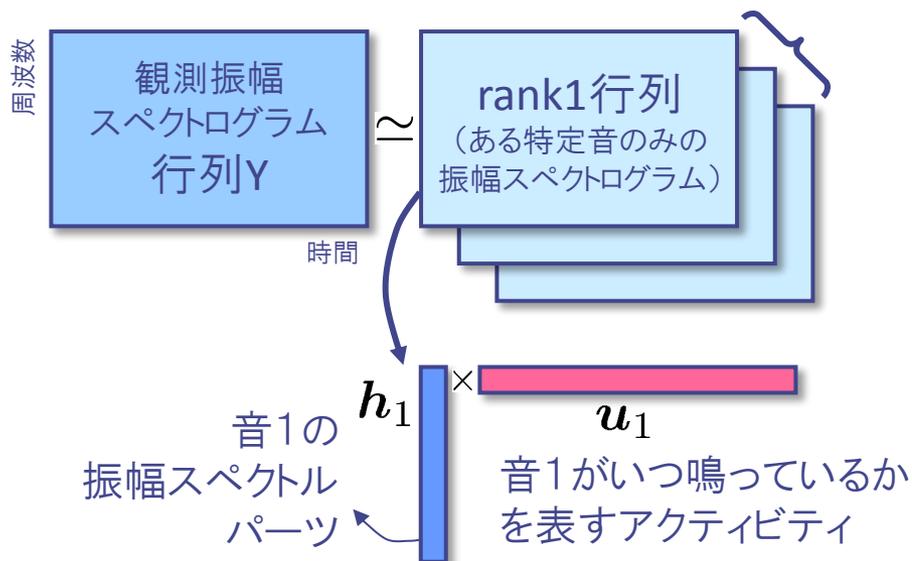
NMFモデル

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

複素NMFのモデル [H. Kameoka 2008]

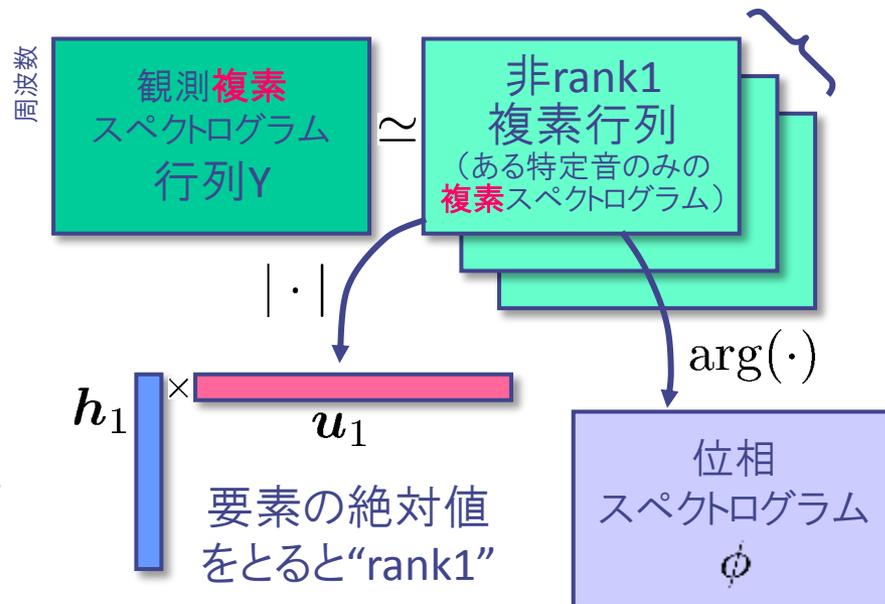
従来のNMFの捉え方

何種類かの音
からなると仮定



複素NMFの考え方

加法性が
成り立つ!



◆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

$$\begin{aligned}
 Y_{\omega,t} &= \sum_k X_{k,\omega,t} \quad |X_{k,\omega,t}| = H_{\omega,k} U_{k,t} \\
 &= \sum_i \underline{|X_{k,\omega,t}|} e^{j\phi_{k,\omega,t}} = \sum_k H_{\omega,k} U_{k,t} e^{j\phi_{k,\omega,t}}
 \end{aligned}$$

行列積の形にならない!

$$H_{\omega,k} \geq 0, U_{k,t} \geq 0$$

複素NMFアルゴリズム [H. Kameoka 2008]

■ 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 $\mathcal{H} \equiv \{H_{\omega,k}\}$
- 位相スペクトログラム $\phi \equiv \{\phi_{k,\omega,t}\}$
- ゲイン $\mathcal{U} \equiv \{U_{k,t}\}$

音響信号モデル

$$F_{\omega,t} = \sum_{k=1}^K H_{\omega,k} U_{k,t} e^{j\phi_{k,\omega,t}}$$

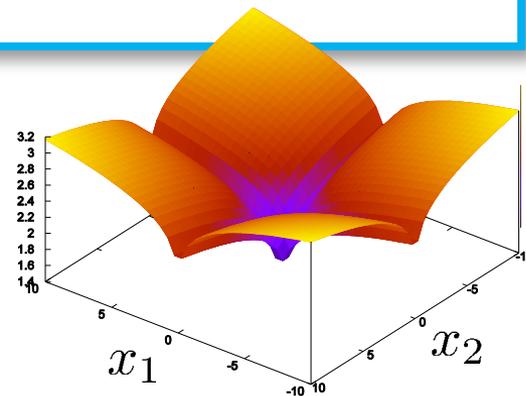
■ 最適化問題

$$\text{minimize } f(\mathcal{H}, \mathcal{U}, \phi) := \sum_{\omega,t} |Y_{\omega,t} - F_{\omega,t}|^2 + 2\lambda \sum_{i,t} |U_{k,t}|^p$$

$$\text{subject to } \sum_{\omega} H_{\omega,k} = 1 \quad (k = 1, \dots, K)$$

スパース正則化項

$$|x_1|^{0.2} + |x_2|^{0.2}$$



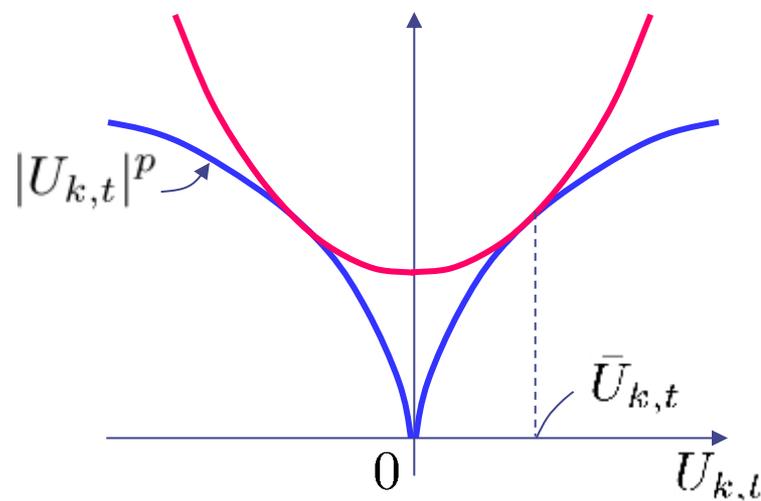
複素NMFアルゴリズム [H. Kameoka 2008]

■ 補助関数法

$$\begin{aligned}
 f(\mathcal{H}, \mathcal{U}, \phi) &= \sum_{\omega, t} |Y_{\omega, t} - F_{\omega, t}|^2 + 2\lambda \sum_{k, t} |U_{k, t}|^p \\
 &\leq \sum_{\omega, t} \sum_i \frac{|\bar{X}_{k, \omega, t} - H_{\omega, k} U_{k, t} e^{j\phi_{k, \omega, t}}|^2}{\beta_{k, \omega, t}} + \lambda \sum_{k, t} (p|\bar{U}_{k, t}|^{p-2} U_{k, t}^2 + (2-p)|\bar{U}_{k, t}|^p) \\
 &\quad \text{subject to } \sum_k \bar{X}_{k, \omega, t} = Y_{\omega, t}
 \end{aligned}$$

$0 < \beta_{k, \omega, t} < 1, \sum_k \beta_{k, \omega, t} = 1$
 を満たす任意の定数

$=: g(\mathcal{H}, \mathcal{U}, \phi, \bar{Y}, \bar{U})$



複素NMFアルゴリズム [H. Kameoka 2008]

■ 補助関数法

$$\begin{aligned}
 f(\mathcal{H}, \mathcal{U}, \phi) &= \sum_{\omega, t} |Y_{\omega, t} - F_{\omega, t}|^2 + 2\lambda \sum_{k, t} |U_{k, t}|^p \\
 &\leq \sum_{\omega, t} \sum_i \frac{|\bar{X}_{k, \omega, t} - H_{\omega, k} U_{k, t} e^{j\phi_{k, \omega, t}}|^2}{\beta_{k, \omega, t}} + \lambda \sum_{k, t} (p|\bar{U}_{k, t}|^{p-2} U_{k, t}^2 + (2-p)|\bar{U}_{k, t}|^p) \\
 &\quad \text{subject to } \sum_k \bar{X}_{k, \omega, t} = Y_{\omega, t}
 \end{aligned}$$

$0 < \beta_{k, \omega, t} < 1, \sum_k \beta_{k, \omega, t} = 1$
 を満たす任意の定数

$$=: g(\mathcal{H}, \mathcal{U}, \phi, \bar{Y}, \bar{U})$$

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

$$\bar{U}_{k, t} \leftarrow U_{k, t}$$

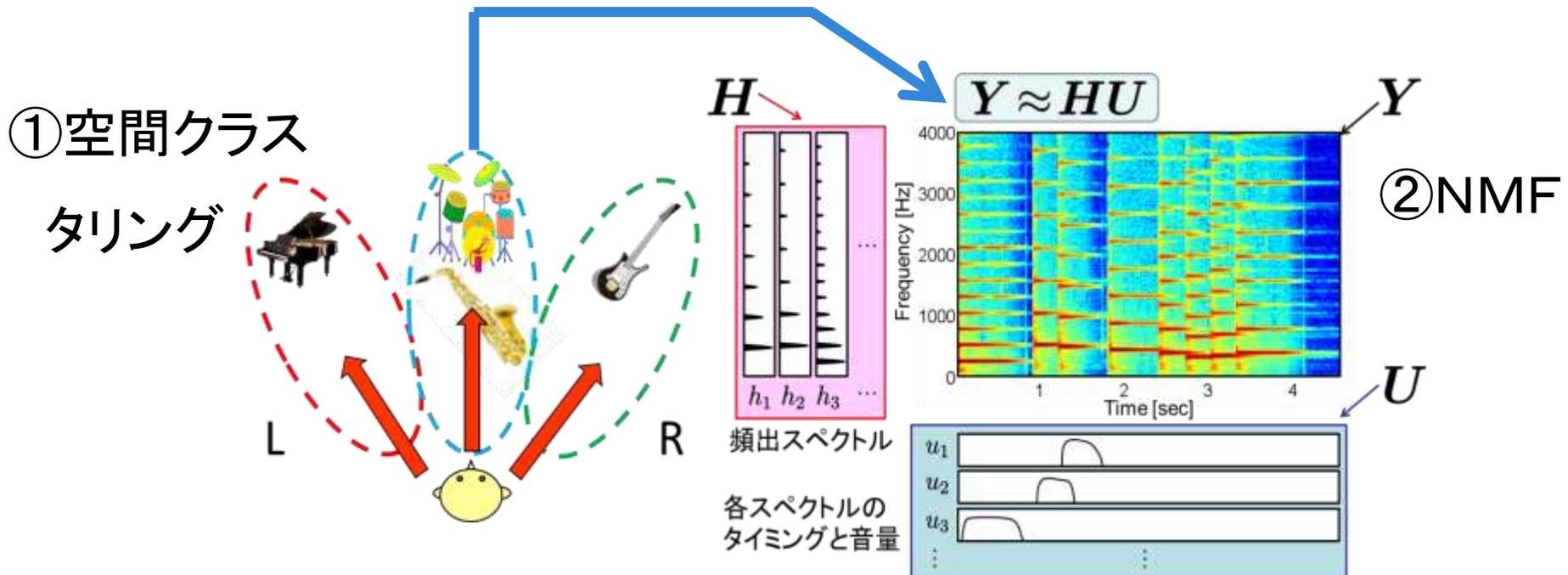
Step 2) $H_{\omega, k} \leftarrow \frac{\sum_t U_{k, t} \operatorname{Re}[\bar{X}_{k, \omega, t}^* e^{j\phi_{k, \omega, t}}] / \beta_{k, \omega, t}}{\sum_t U_{k, t}^2 / \beta_{k, \omega, t}}$

$$U_{k, t} \leftarrow \frac{\sum_t H_{\omega, k} \operatorname{Re}[\bar{X}_{k, \omega, t}^* e^{j\phi_{k, \omega, t}}] / \beta_{k, \omega, t}}{\sum_t U_{k, t}^2 / \beta_{k, \omega, t} + \lambda p |\bar{U}_{k, t}|^{p-2}}$$

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

音源分離例：半教師有り超解像NMF [Kitamura, Saruwatari et al. 2014]

- ステレオ音楽信号を効率よく分解するための、空間クラスタリングとスペクトル頻出パターン分解(NMF)を組み合わせた手法
- 本手法では、空間クラスタリングでの分類エラーを学習基底で補修する機能が備わっているため、**分離に必要なスパース性**と**補修に必要なアンチスパース性**との間でトレードオフが生じる。
- そこで、各音源に合ったスパース性を選択する機能を導入した。



半教師有り超解像NMF [Kitamura, Saruwatari et al. 2014]

- 空間・時間・周波数で表される確率的信号モデルを、「空間」及び「時間周波数」領域における2つの個別モデルに分割

$$p(\text{空間}, \text{時間}, \text{周波数}) \approx p(\text{空間}) \cdot p(\text{時間}, \text{周波数})$$

教師無し分類法による
方位クラスタリング

教師有りNMFによる
低ランク近似・分解

着眼点

- 方位クラスタリングが真の分布を表さない場合の誤差は？
- 教師有りNMFにおける「超解像」能力を利用し、上記の分類誤差を回復できないか？
- ではどのようなNMFを設計すればよいか？

半教師有り超解像NMF [Kitamura, Saruwatari et al. 2014]

分解モデル: $Y \approx FG + HU$ ただし Y の各要素は空間クラスタリングにより $i_{\omega,t} = \{0,1\}$ がかかる

教師スペクトル基底 (固定)

一般化されたコスト関数:

$$\mathcal{J} = \sum_{\omega,t} i_{\omega,t} \mathcal{D}_{\beta_{\text{NMF}}} (y_{\omega,t} \parallel \sum_k f_{\omega,k} g_{k,t} + \sum_l h_{\omega,l} u_{l,t})$$

音源分離項 (低ランク近似による分解)

$$+ \lambda \sum_{\omega,t} \overline{i_{\omega,t}} \mathcal{D}_{\beta_{\text{reg}}} (0 \parallel \sum_k f_{\omega,k} g_{k,t}) + \mu \|F^T H\|_{\text{Fr}}^2$$

正則化項 (非観測点における外挿)

$i_{\omega,t}, y_{\omega,t}, f_{\omega,k}, g_{k,t}, h_{\omega,l}, u_{l,t}$: それぞれ行列 I, Y, F, G, H, U の要素,
 $\overline{\cdot}$: 論理反転, λ, μ : 正則化項と罰則項の重み係数, $\|\cdot\|_{\text{Fr}}$: フロベニウスノルム

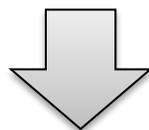
- NMFのメイン項と正則化項には一般化された距離関数である β -divergence を用いる ($\beta \rightarrow$ 小の時スパース化)

超解像NMFにおけるスパース性の制御

- この手法には2つのタスクがある

$$\begin{array}{c} \text{提案} \\ \text{教師ありNMF} \end{array} = \begin{array}{c} \text{目的音源} \\ \text{の分離} \end{array} + \begin{array}{c} \text{教師基底} \\ \text{の外挿} \end{array}$$

- 過去の研究では、音源分離に適切なNMFの距離規範が検討されてきた
 - KL-divergence規範 ($\beta_{\text{NMF}} = 1$) が精度よく分離できる
 - 結論として、「よりスパースな表現」が良い分離を与える？
- しかし、教師基底の外挿に適切な距離規範についての議論はされていない



そこで

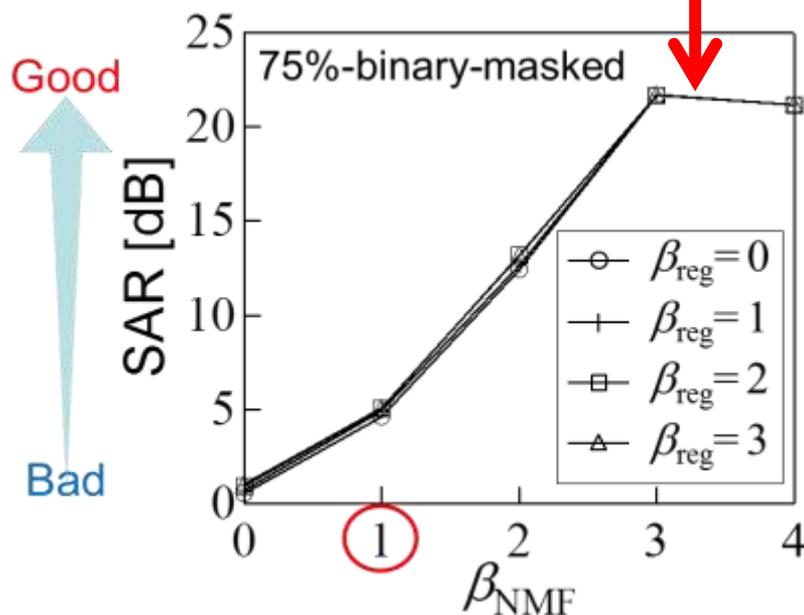
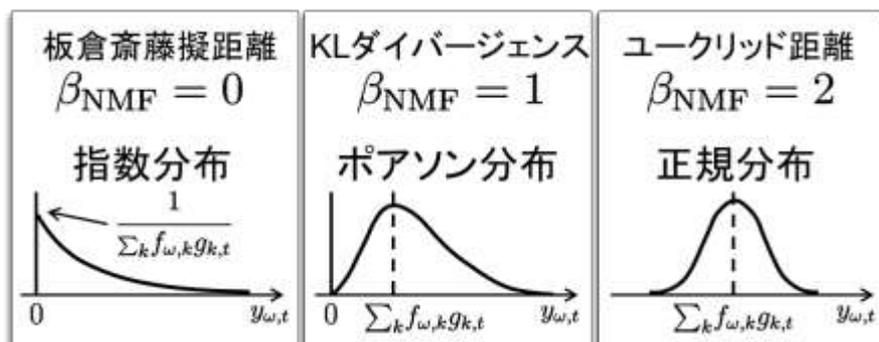
- 距離規範ごとの教師基底の外挿能力を、**NMFの仮定する生成モデル**の観点から議論するとどうなるか？

生成モデル的観点に基づく外挿能力

- NMFによる分解は観測データ Y の確率的生成モデルを仮定した最尤推定問題と等価
 - β -divergence最小化における β の値によって変化

$$\arg \min_{F, G} \mathcal{J} = \mathcal{D}_{\beta_{\text{NMF}}} (Y \| FG)$$

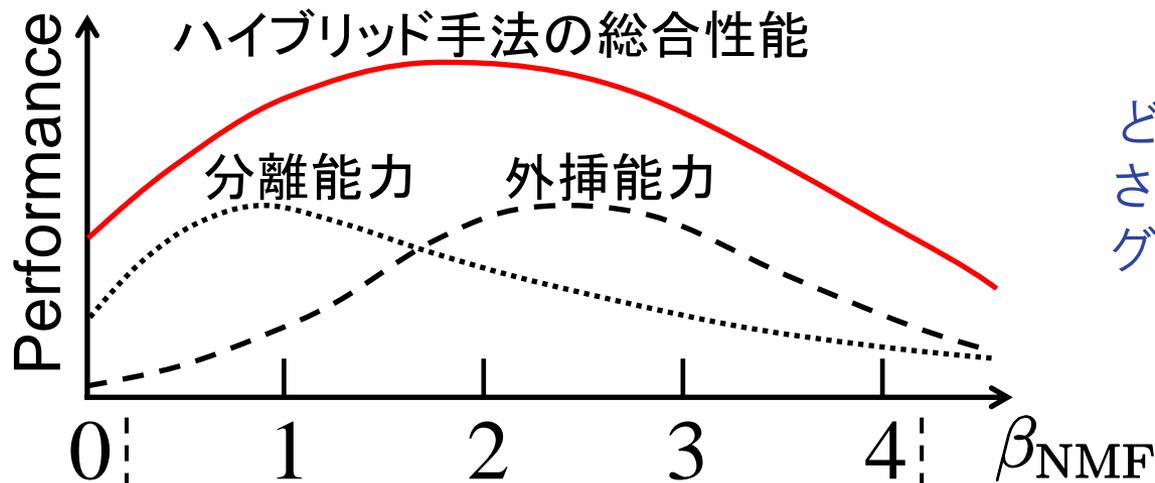
- その外挿能力はどのような傾向にあるか？ ⇒ アンチスパース



最適ダイバージェンスの導出 (Adaptive Divergence)

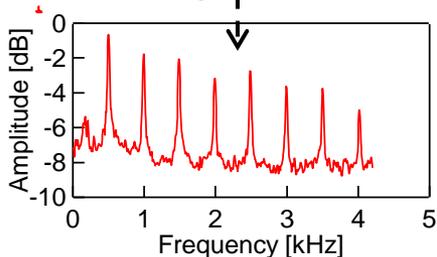
- 最適なダイバージェンスは音源分離能力 (スパース) と外挿能力 (アンチスパース) のトレードオフとなる
- β の値を観測データにあわせて最適化する機構を導入

[Kitamura, Saruwatari et al., IEEE Trans. Audio, Speech, Lang. Processing, 2014]

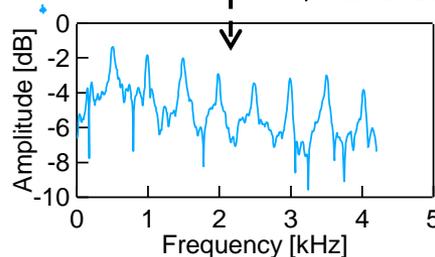


どのダイバージェンスが使用されるかは、空間クラスタリングの結果に応じて決定される

↓
時間的に変動するダイバージェンス



スパース性: 強



スパース性: 弱

ステレオ音楽信号分離デモ

■ 実際の演奏曲を半教師有りNMFで分解してみた。

原曲



教師1



分離音1



教師2



分離音2



ステレオ音楽信号分離デモ2

■ プロレコーディングに対応できる品質を目指して。

原曲（プロ演奏）



Saxのみを抜いた
伴奏部分



Copyright © 2014 Yamaha Corp.
All rights reserved.

サクソ奏者が
消えた!?

