非ガウス性モノラル音響信号に対する音源分離のための非負値行列分解と半正定値テンソル分解

吉井 和佳　糸山 克寿 (京大)
後藤 真孝 (産総研)

第108回情報処理学会音楽情報科学研究会 (SIGMUS 108)

非ガウス性モノラル音響信号に対する

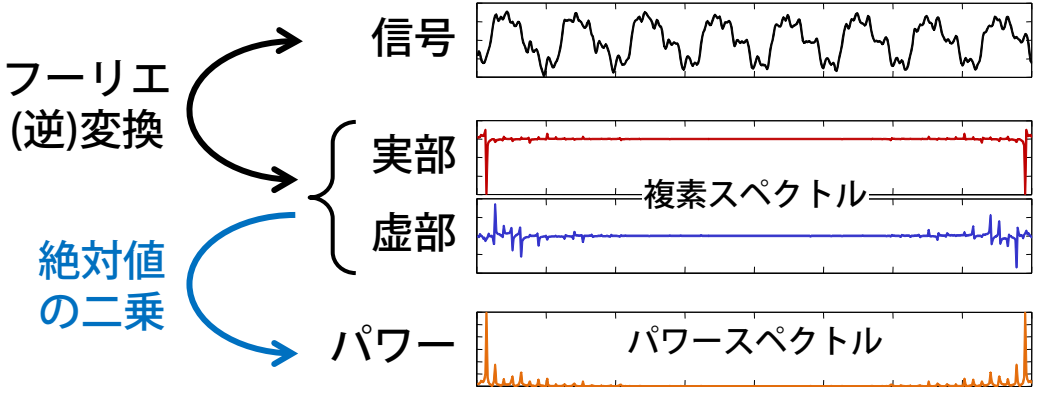
音源分離のための
非負値行列分解と半正定値テンソル分解

吉井 和佳　糸山 克寿 (京大)
後藤 真孝 (産総研)

第108回情報処理学会音楽情報科学研究会 (SIGMUS 108)

研究の背景

- モノラル音響信号の音源分離に関する研究がさかん
  - 応用例：音楽信号の高度な加工
    - 歌声と伴奏の分離 [Rafii2011, Huang2012, Yang2014, Ikemiya2015]
    - 楽器音イコライザ [吉井2006, 宮本2008, 糸山2008]
- スペクトル領域で音源分離を行うのが一般的
  - 位相を無視すると音の特徴 (調波構造・スパース性) がとらえやすくなる

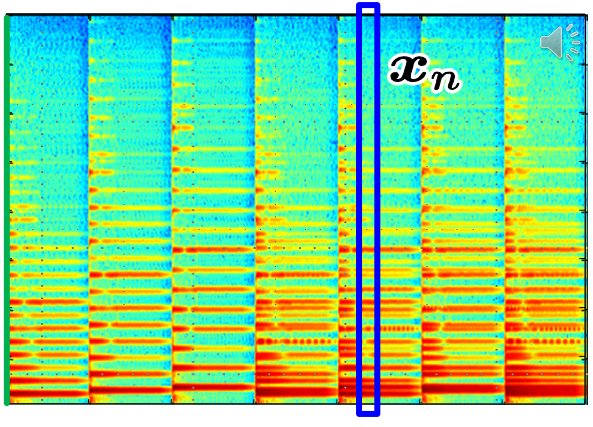


非負値行列分解 (NMF)
に基づく手法が席巻
- 直感に合う
- 実装が簡単
- 高速に動作

非負値行列分解 (NMF) [Lee 1999 / Févotte 2009]

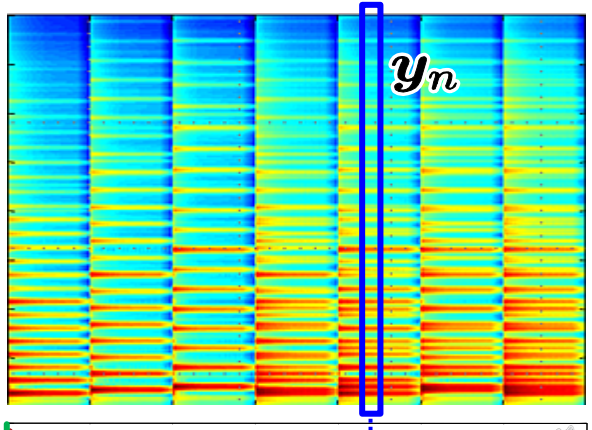
- 「非負値ベクトル」を少数の「非負値ベクトル」の錐結合で表現

観測行列 X

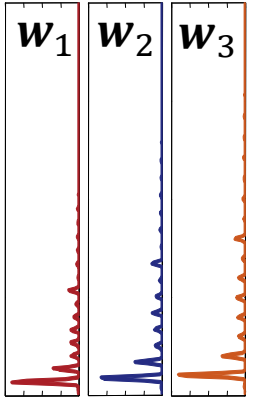


$\approx$

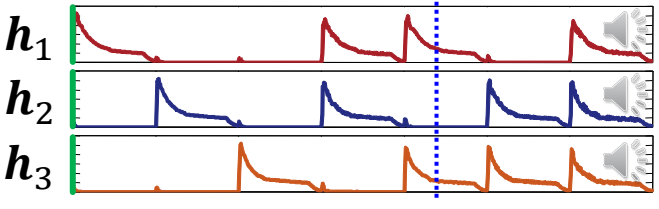
再構成行列 $Y = WH$



基底ベクトル群 W



$$x_n \approx y_n = \sum_{k=1}^{K} h_{kn} w_k$$



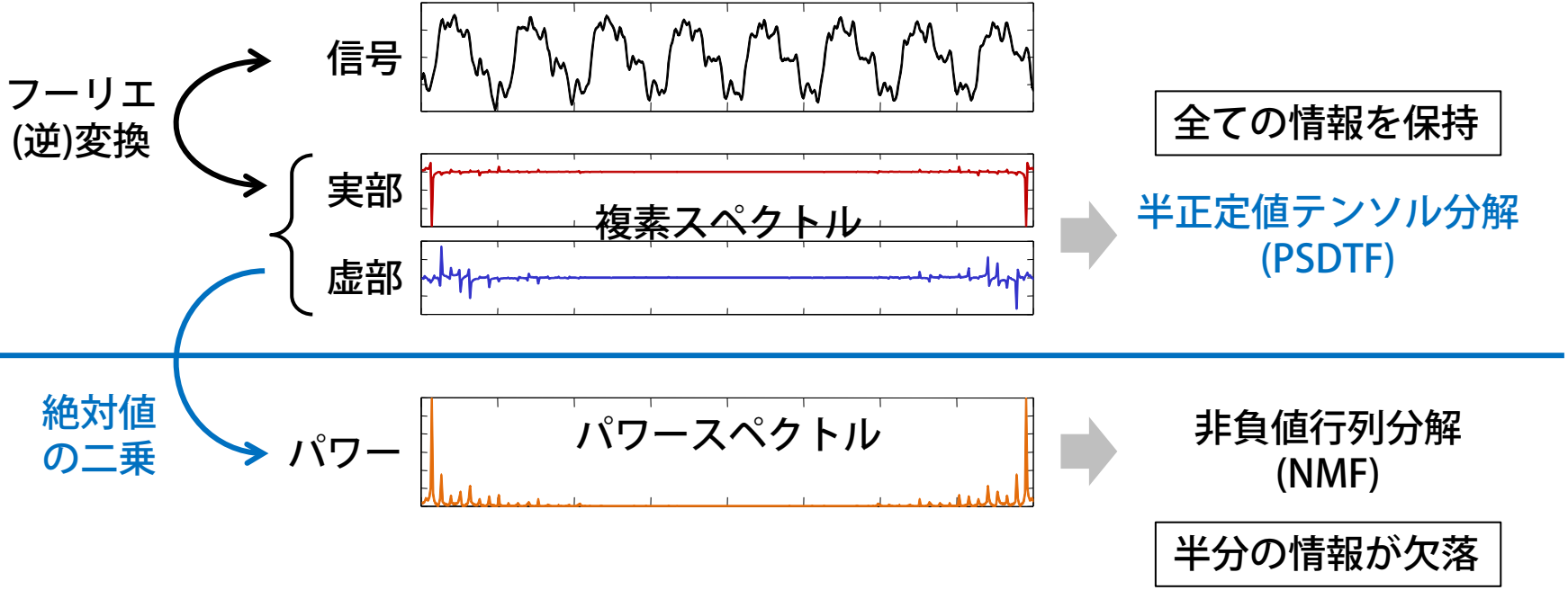
音量ベクトル群 H

音源分離に適したコスト関数：板倉・斎藤ダイバージェンス

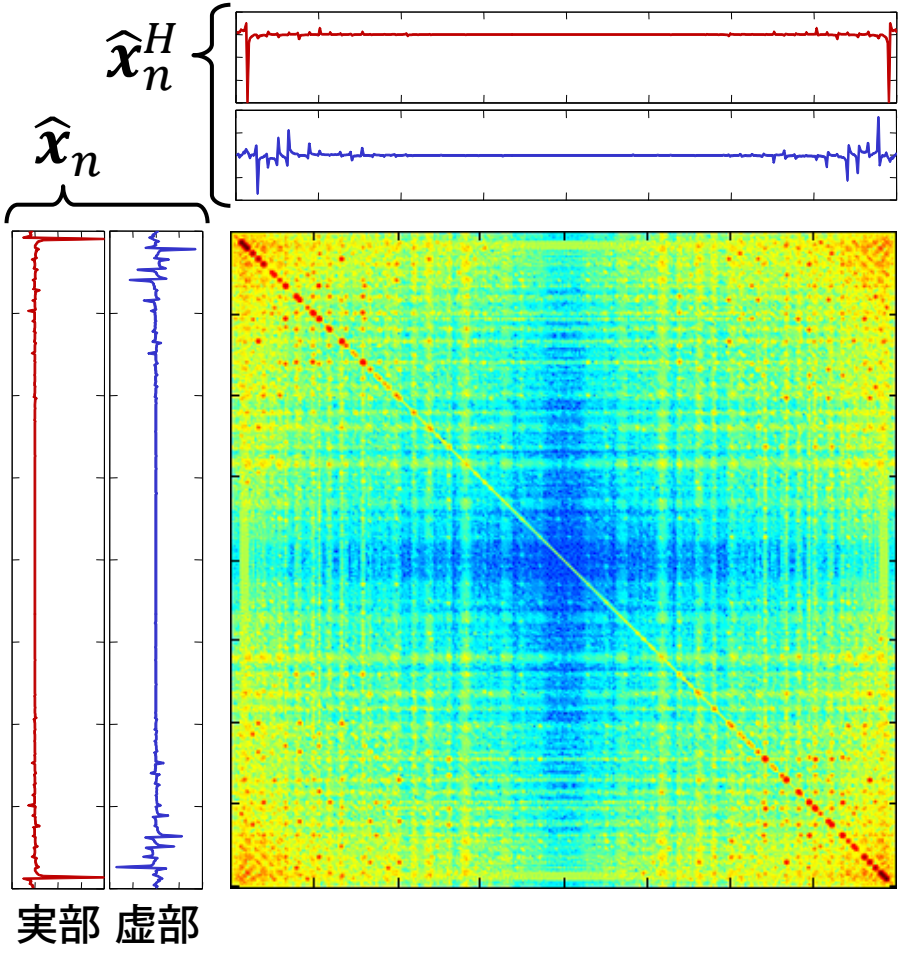
$$D_{IS}(x_n|y_n) = \sum_m (-\log x_{nm} y_{nm}^{-1} + x_{nm} y_{nm}^{-1} - 1)$$

研究の背景

- 複素スペクトログラムを直接分解できる方法論が必要
  - 入力音響信号のもつ情報 (パワー・位相) を全て適切に取り扱いたい
  - パワースペクトルにした時点で位相情報が失われる



複素スペクトルの自己共分散

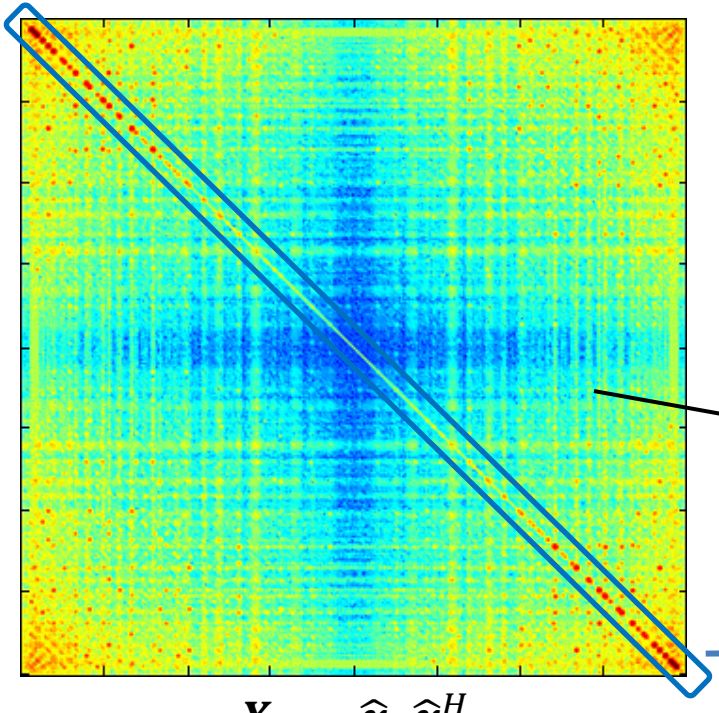


$$\boldsymbol{X}_n = \hat{\boldsymbol{x}}_n \hat{\boldsymbol{x}}_n^H$$

※複素行列は表示できないので
絶対値の大きさでカラー表示

観測データの与え方

- 各フレームの複素スペクトルの自己共分散を計算 → 半正定値行列
  - 各フレームの複素スペクトルの絶対値の二乗を計算 → 非負値ベクトル



$\boldsymbol{X}_n = \hat{\boldsymbol{x}}_n \hat{\boldsymbol{x}}_n^H$

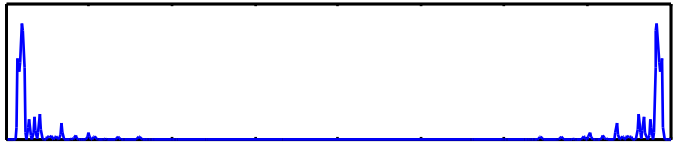
ある行列$\boldsymbol{X}$が半正定値行列であるとは
- すべての固有値が非負
- $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{A}^H$となる行列$\boldsymbol{A}$が存在

} 同値

半正定値性は非負値性の拡張概念

行列要素は任意のスカラ (負値や複素数もOK)

対角成分は常に非負 (パワースペクトル)



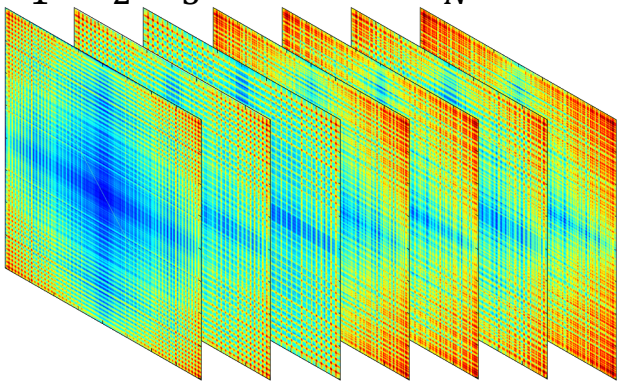
$\boldsymbol{x}_n = \hat{\boldsymbol{x}}_n \odot \hat{\boldsymbol{x}}_n$

半正定値テンソル分解 (PSDTF) [Yoshii 2013]

- 「半正定値行列」を少数の「半正定値行列」の錐結合で表現

観測テンソル $\mathcal{X}$

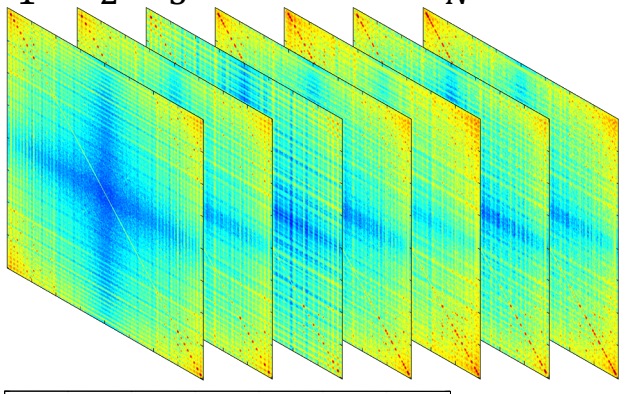
$\boldsymbol{X}_1$ $\boldsymbol{X}_2$ $\boldsymbol{X}_3$ $\cdots$ $\boldsymbol{X}_N$



$\approx$

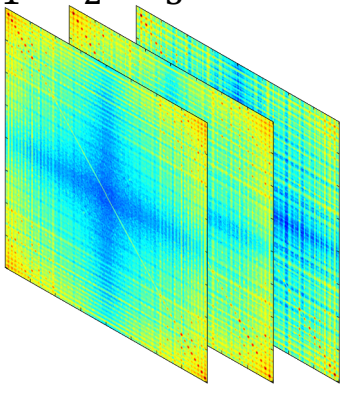
再構成テンソル $\mathcal{Y}$

$\boldsymbol{Y}_1$ $\boldsymbol{Y}_2$ $\boldsymbol{Y}_3$ $\cdots$ $\boldsymbol{Y}_N$

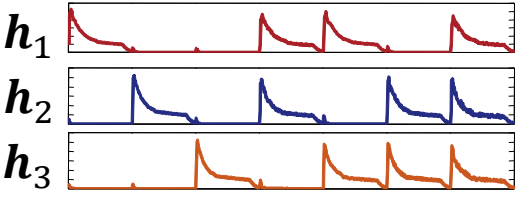


基底行列群 $\mathcal{W}$

$\boldsymbol{W}_1$ $\boldsymbol{W}_2$ $\boldsymbol{W}_3$



$$\boldsymbol{X}_n \approx \boldsymbol{Y}_n = \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k$$

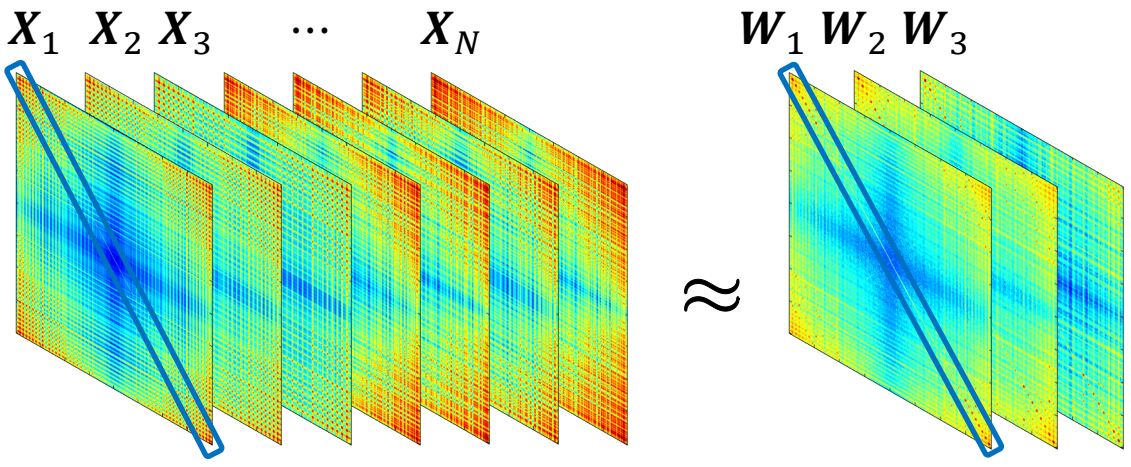


音量ベクトル群 $\boldsymbol{H}$

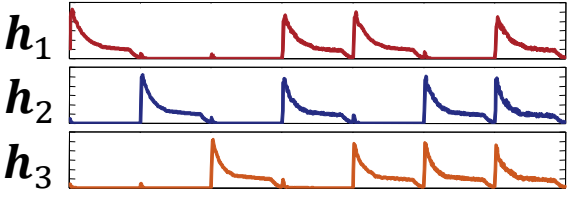
音源分離に適したコスト関数：Log-Determinantダイバージェンス

$$D_{LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = -\log|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}| + \boldsymbol{X}_n\boldsymbol{Y}_n^{-1} - M$$

PSDTF v.s. NMF

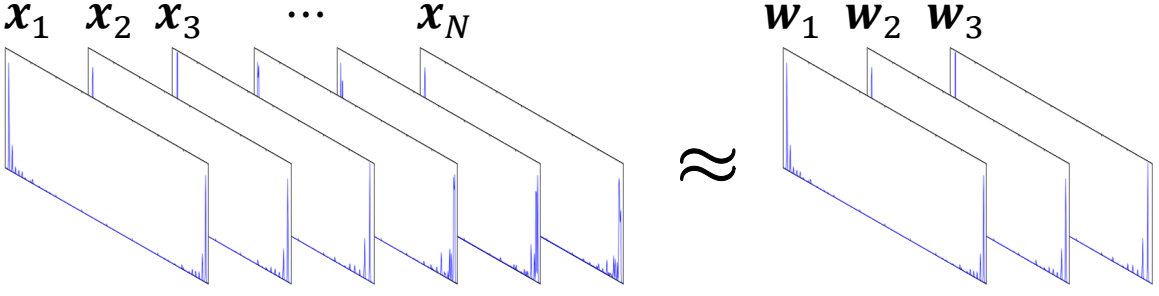


半正定値行列 $\boldsymbol{X}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k$ 半正定値行列

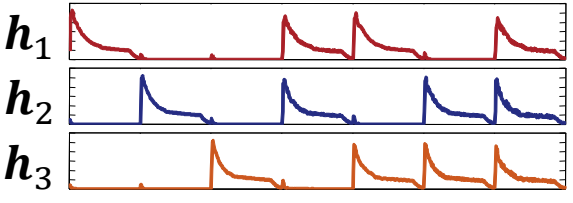


音源分離精度：SDR 26.7 dB

対角成分に限定 (周波数ビン間の相関を無視)

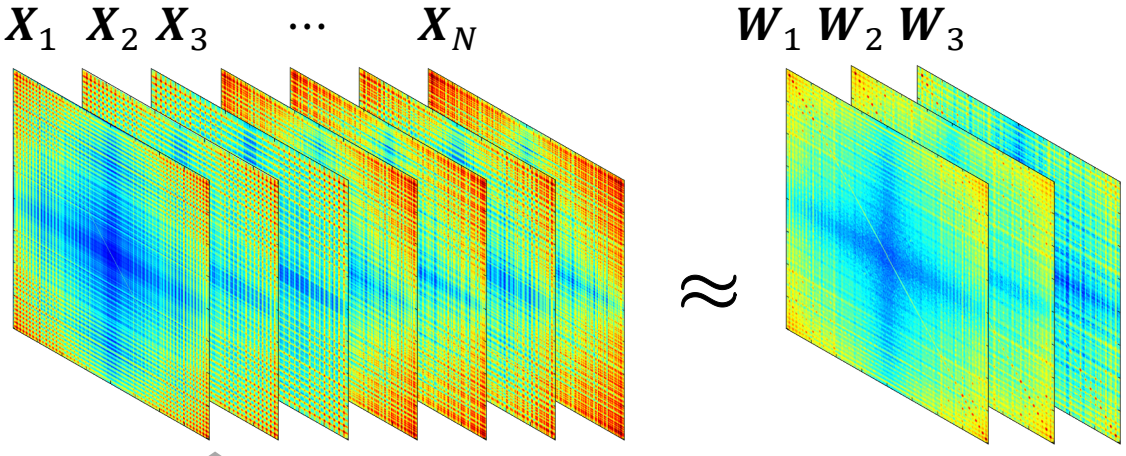


非負値ベクトル $\boldsymbol{x}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{w}_k$ 非負値ベクトル

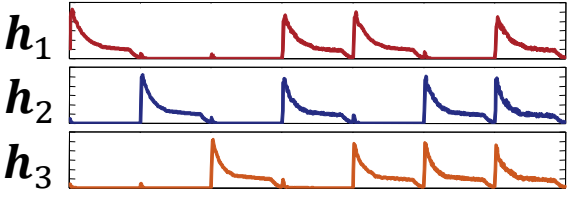


音源分離精度：SDR 18.9 dB

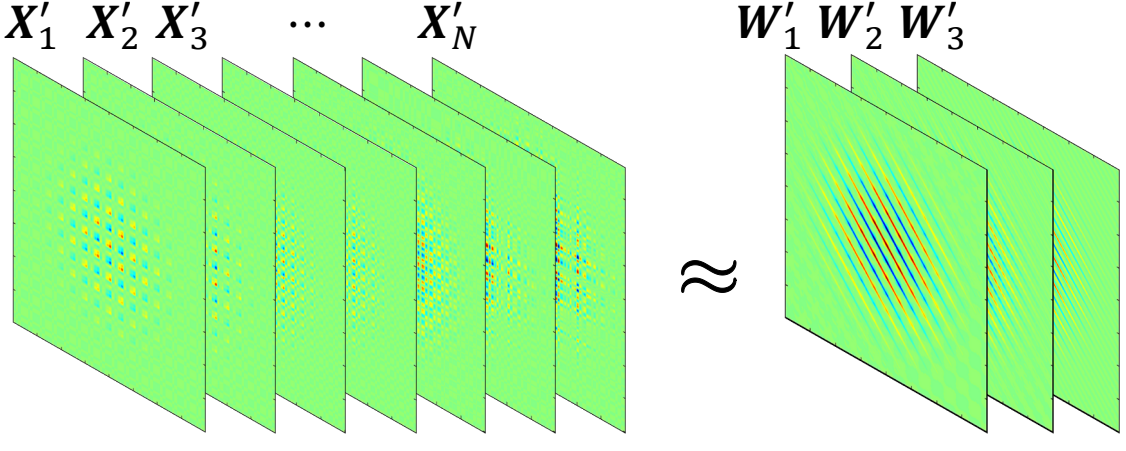
周波数領域PSDTF v.s. 時間領域PSDTF



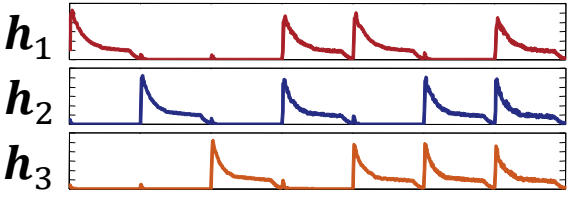
半正定値行列 $(FX'_nF^H) \approx \sum_{k=1}^{K} h_{kn}(FW'_kF^H)$ 半正定値行列



離散フーリエ変換行列Fで線形変換



半正定値行列 $X'_n \approx \sum_{k=1}^{K} h_{kn}W'_k$ 半正定値行列



非ガウス性モノラル音響信号に対する音源分離のための非負値行列分解と半正定値テンソル分解

吉井 和佳　糸山 克寿 (京大)
後藤 真孝 (産総研)

第108回情報処理学会音楽情報科学研究会 (SIGMUS 108)

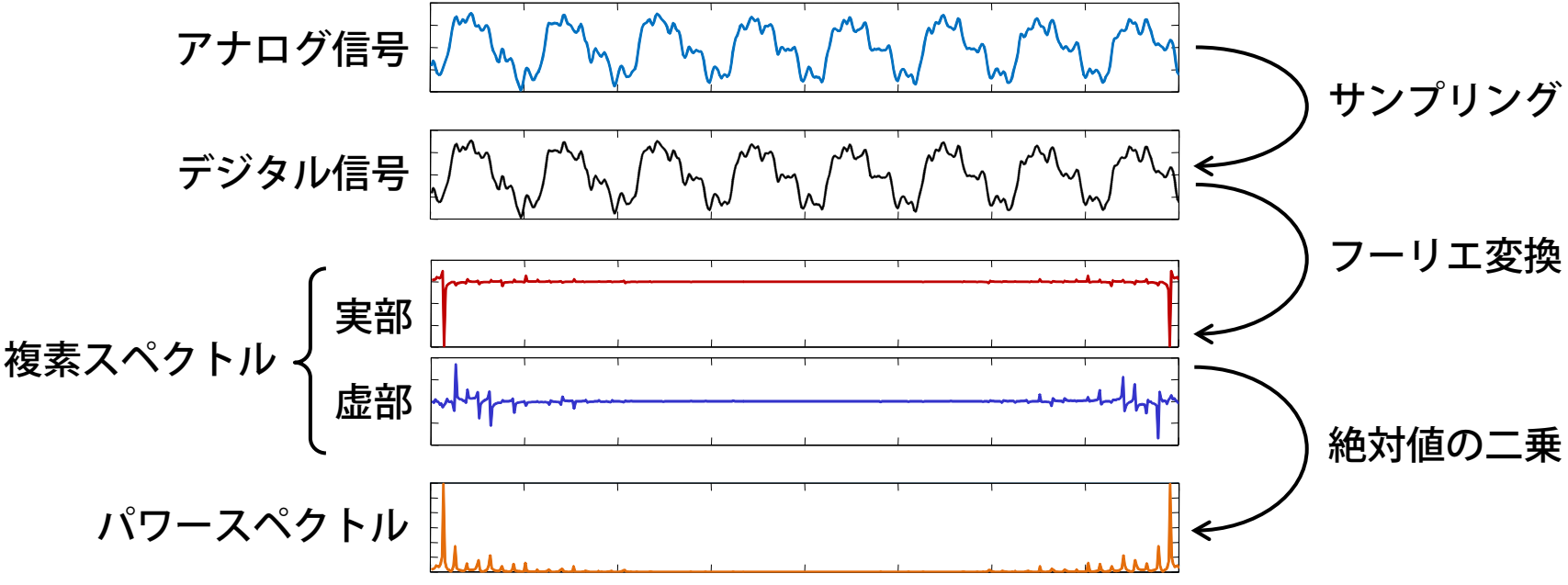
非ガウス性モノラル音響信号に対する音源分離のための非負値行列分解と半正定値テンソル分解

吉井 和佳　糸山 克寿 (京大)
後藤 真孝 (産総研)

第108回情報処理学会音楽情報科学研究会 (SIGMUS 108)

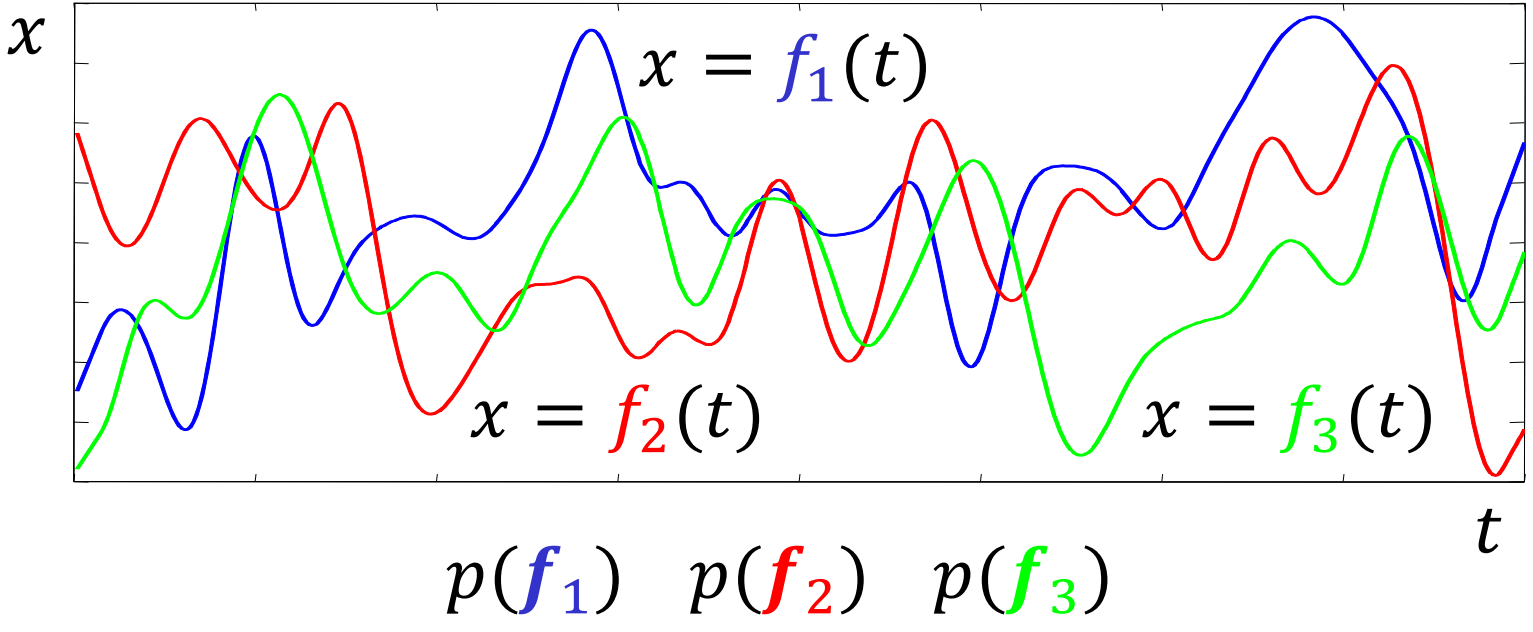
研究の動機

- 音響信号は連続信号 → 連続信号に対する確率モデルが必要
  - 空気振動は連続時間上の物理現象
  - 「サンプリング」によって離散信号が得られる



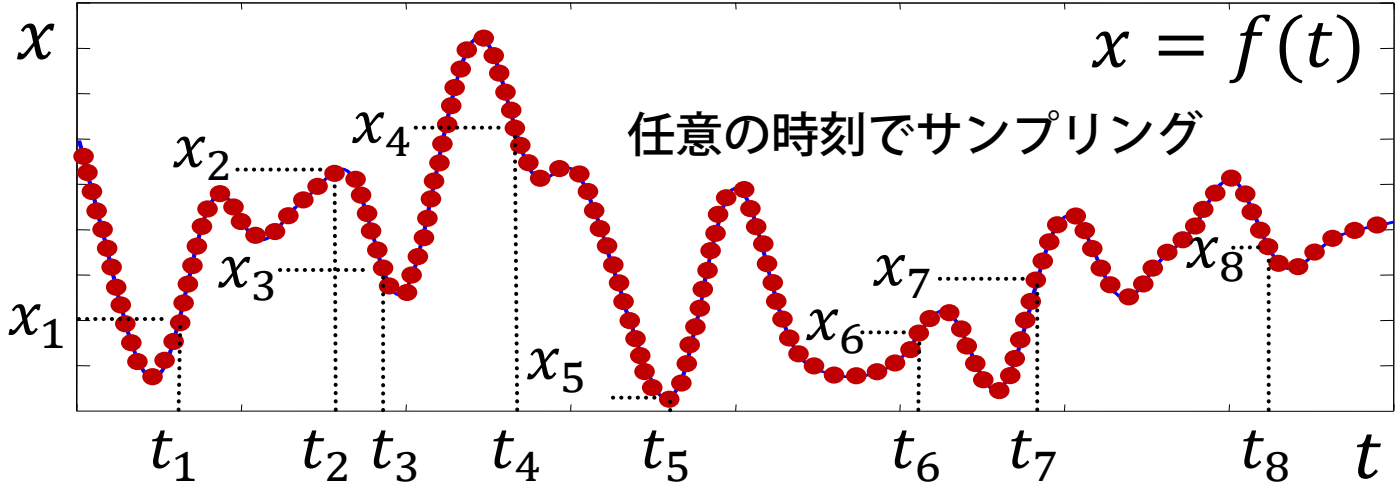
ガウス過程 (GP)

- 連続時間上の確率過程 (無限次元の確率分布) のひとつ
  - 連続関数 $\boldsymbol{f}$ (無限次元のベクトル) に対する確率分布 $p(\boldsymbol{f})$
  - 任意の音響信号のもっともらしさを計算可能



ガウス過程とガウス分布

- ガウス過程 ⇔ 無限次元のガウス分布
  - 連続関数 f ⇔ 無限個の出力値 $\boldsymbol{x} = [x_1, x_2, \cdots, x_\infty]^T$



関数 f

$$\boldsymbol{t} = [t_1, t_2, \cdots, t_\infty]^T$$

$$\boldsymbol{x} = [x_1, x_2, \cdots, x_\infty]^T$$

無限次元のガウス分布

$$\boldsymbol{x} \sim \mathrm{GP}(0, \boldsymbol{K})$$

任意の周辺分布はガウス分布

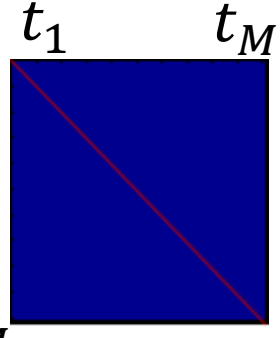
Deltaカーネル

- 定常な白色雑音を生成
  - 周辺分布 $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{K})$

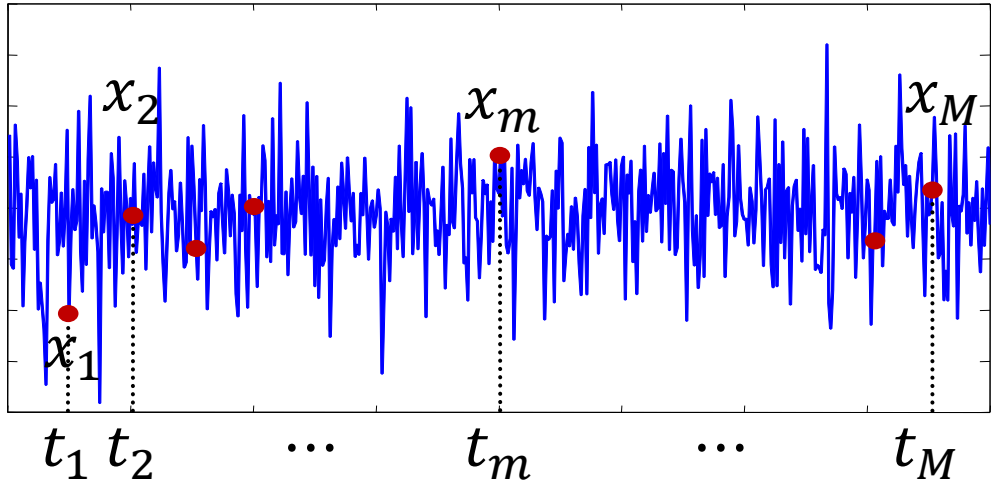
関数 f

$$\boldsymbol{t} = [t_1, t_2, \cdots, t_M]^T$$
$$\boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T$$

$\boldsymbol{K} =$



各x_mは独立な
ガウス分布に従う

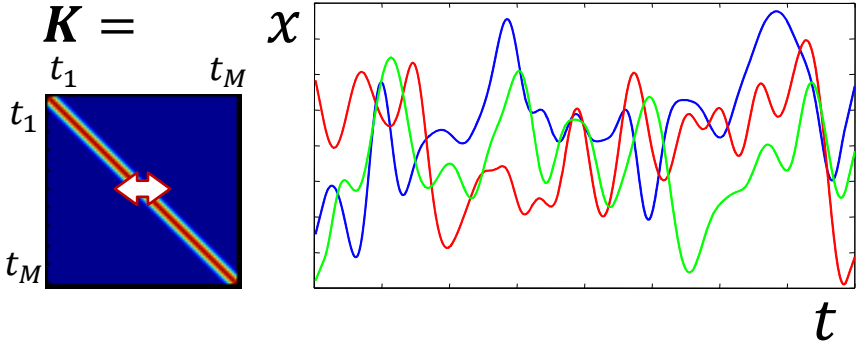


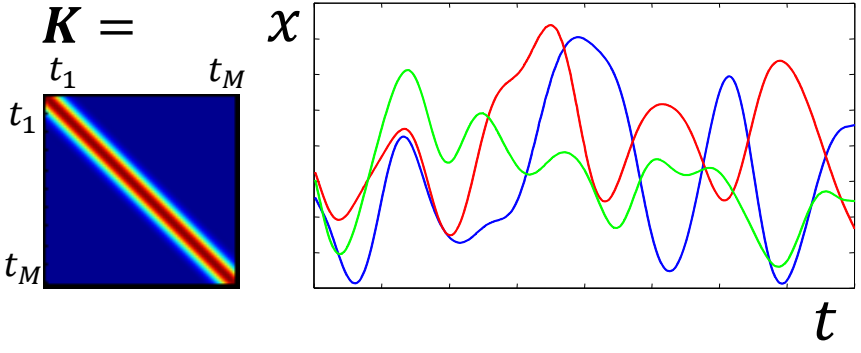
Squared Exponentialカーネル

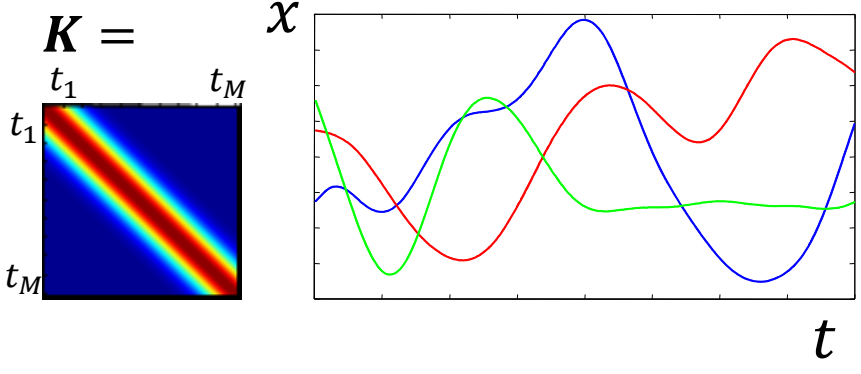
- なめらかな連続関数を生成
  - 周辺分布 $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{K})$

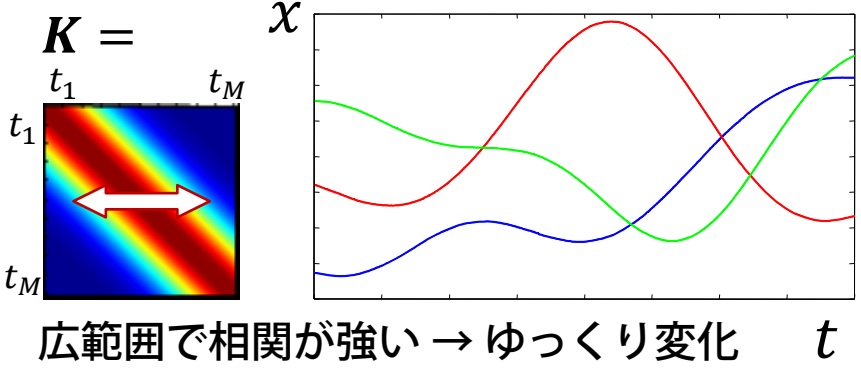
関数 f $\quad \boldsymbol{t} = [t_1, t_2, \cdots, t_M]^T$

$\quad \boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T$





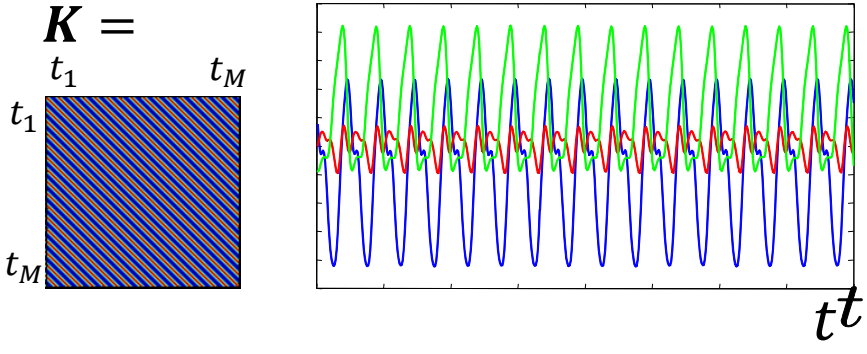


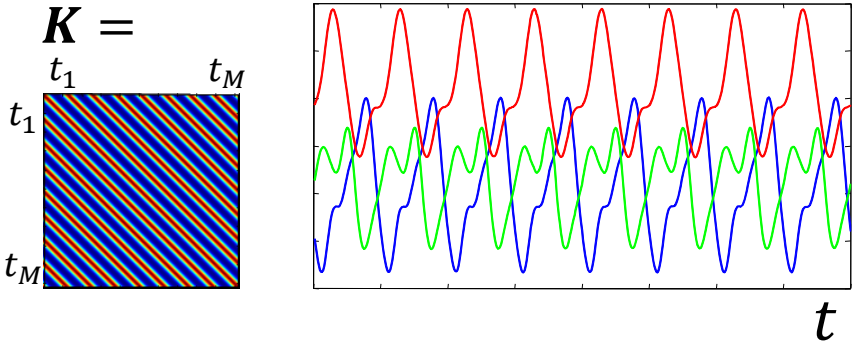


周期カーネル

- 周期的な連続関数を生成
  - 周辺分布 $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{K})$

関数 f $\boldsymbol{t} = [t_1, t_2, \cdots, t_M]^T$ → $\boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T$

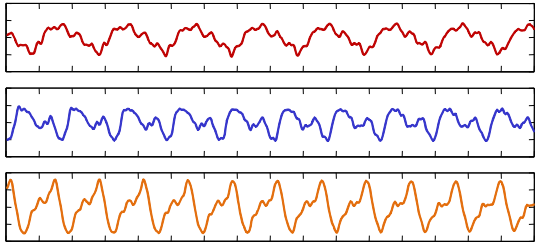




一定間隔で相関が強い → 周期的な変化

混合音の生成と分離

- 生成：ガウス変数の和 → ガウス変数　再生性をもつ



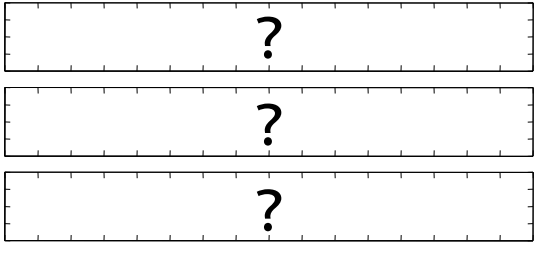
$\boldsymbol{x}_1 \sim N(\boldsymbol{0}, \boldsymbol{K}_1)$

$\boldsymbol{x}_2 \sim N(\boldsymbol{0}, \boldsymbol{K}_2)$

$\boldsymbol{x}_3 \sim N(\boldsymbol{0}, \boldsymbol{K}_3)$

$\underbrace{\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3}_{\boldsymbol{x}} \sim N(\boldsymbol{0}, \underbrace{\boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3}_{\boldsymbol{K}})$

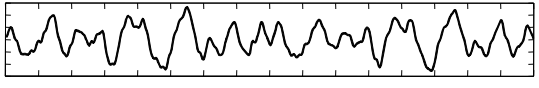
- 分離：ガウス変数の分解 → ガウス変数　条件付き期待値が求まる



$\boldsymbol{x}_1 \sim N(\boldsymbol{0}, \boldsymbol{K}_1)$

$\boldsymbol{x}_2 \sim N(\boldsymbol{0}, \boldsymbol{K}_2)$

$\boldsymbol{x}_3 \sim N(\boldsymbol{0}, \boldsymbol{K}_3)$



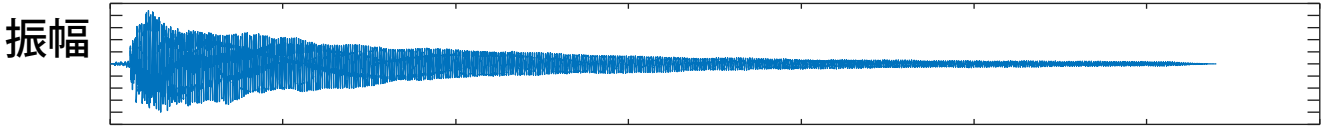
$\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 \sim N(\boldsymbol{0}, \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3)$

ウィナーフィルタによる音源分離

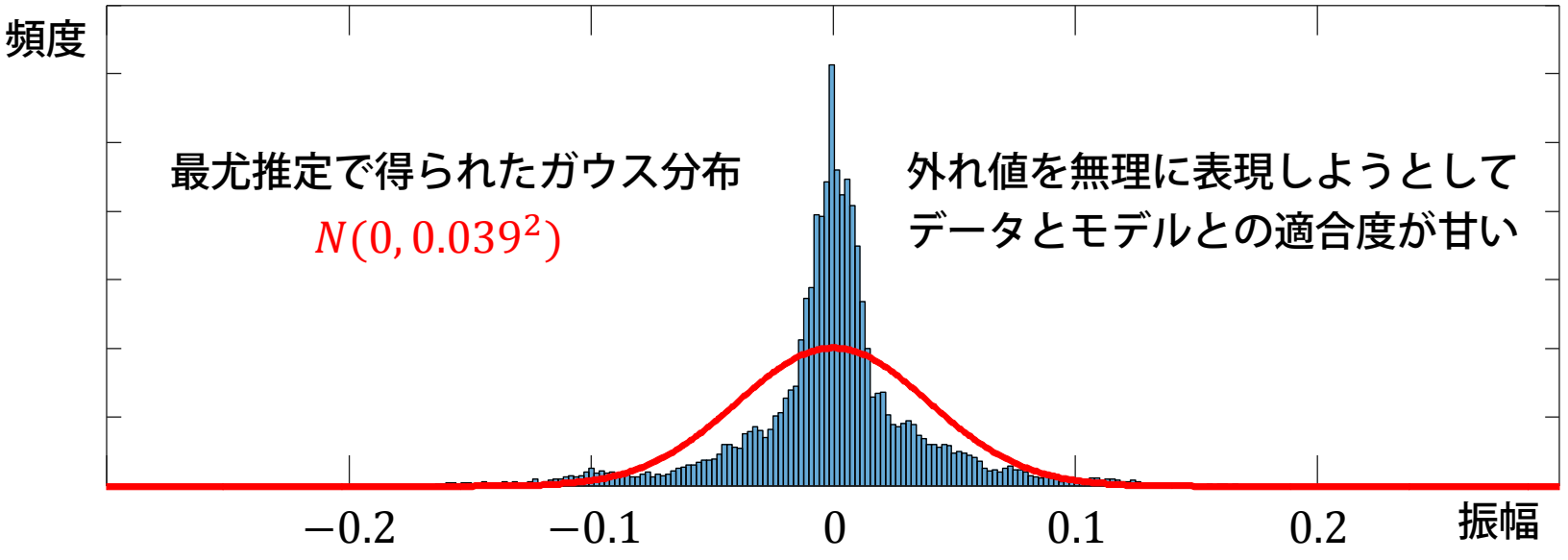
$E[\boldsymbol{x}_1|\boldsymbol{x}] \sim \boldsymbol{K}_1\boldsymbol{K}^{-1}\boldsymbol{x}$　$E[\boldsymbol{x}_2|\boldsymbol{x}] \sim \boldsymbol{K}_2\boldsymbol{K}^{-1}\boldsymbol{x}$　$E[\boldsymbol{x}_3|\boldsymbol{x}] \sim \boldsymbol{K}_3\boldsymbol{K}^{-1}\boldsymbol{x}$

ガウス分布の限界

- 実際の音源信号はガウス分布には従わないことが多い
  - ガウス分布では値の散らばり (はずれ値含む) を十分に表現できない

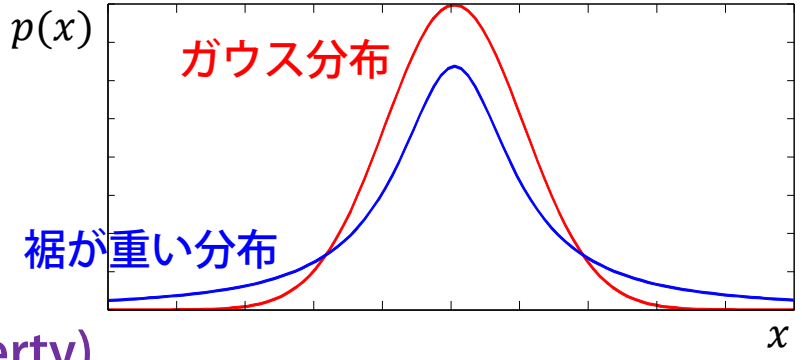


011PFNOM
(RWC-MDB-I-2001)

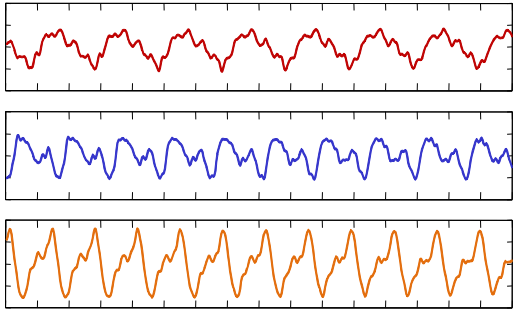


どのような確率分布を用いるべきか

- 裾が重い (heavy-tailed)
  - 裾が指数関数的には減衰しない
  - 外れ値に頑健なモデル化が可能



- 再生性をもつ (reproductive property)
  - 確率変数の和が同じクラスの確率分布に従う
  - 音源信号の加法性を成立させるためには不可欠



$\boldsymbol{x}_1 \sim N(\mathbf{0}, \boldsymbol{K}_1)$

$\boldsymbol{x}_2 \sim N(\mathbf{0}, \boldsymbol{K}_2)$

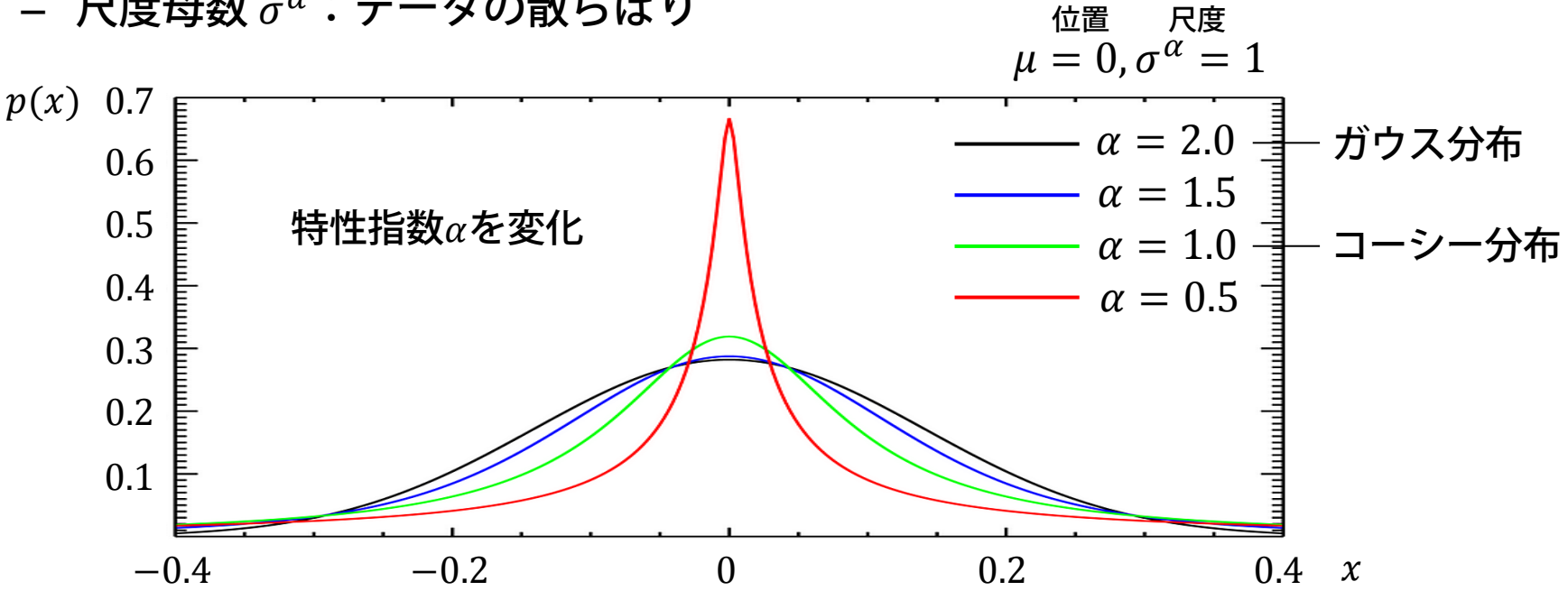
$\boldsymbol{x}_3 \sim N(\mathbf{0}, \boldsymbol{K}_3)$

?

$\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 \sim N(\mathbf{0}, \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3)$

対称α安定分布

- 3個のパラメータを持つ左右対称な確率分布 $S_\alpha(x|\mu,\sigma^\alpha)$
  - 特性指数 α：SαS分布を特徴づける最も重要な量 $(0<\alpha\leq 2)$
  - 位置母数 μ：本研究では0 (音響信号は0を中心に振動)
  - 尺度母数 σ^α：データの散らばり



$S\alpha S$分布の性質

- 理論的には$S\alpha S$分布が音源分離に最も適している [Liutkus 2015]
  - 裾が重い → 音源信号に含まれる外れ値に対応
  - 再生性をもつ → 音源信号の加法性が成立

非ガウス信号分離のための
一般化ウィナーフィルタを提案

「安定」分布と呼ばれる由来

$$X_1, X_2 \sim S_\alpha(0, \sigma^\alpha) \Rightarrow AX_1 + BX_2 \sim S_\alpha(0, (A^\alpha + B^\alpha)\sigma^\alpha)$$

$(A, B > 0)$　　やはり安定分布！

確率変数の加法性

$$\begin{cases} X_1 \sim S_\alpha(0, \sigma_1^\alpha) \\ X_2 \sim S_\alpha(0, \sigma_2^\alpha) \end{cases}$$

中心極限定理の例外！

$$\Rightarrow AX_1 + BX_2 \sim S_\alpha(0, A^\alpha\sigma_1^\alpha + B^\alpha\sigma_2^\alpha)$$

$\alpha = 1$ (コーシー分布) なら振幅領域で
$\alpha = 2$ (ガウス分布) ならパワー領域で
加法性が成立

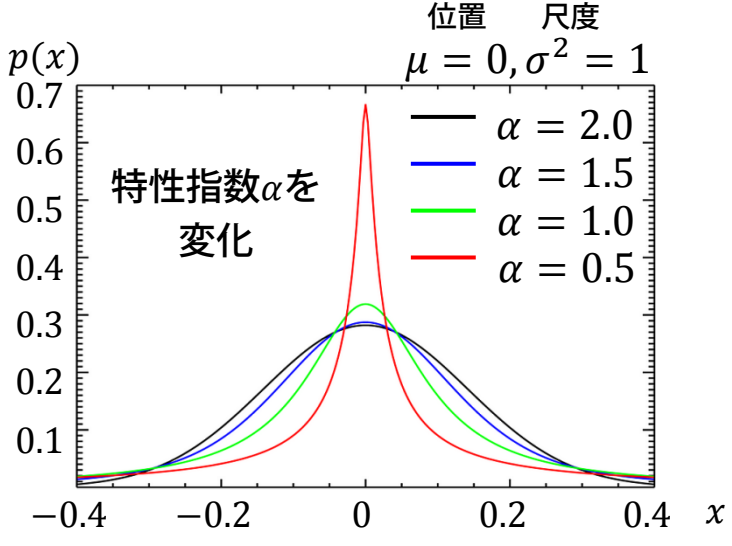
$S\alpha S$分布の性質

- 一般の$0 < \alpha \leq 2$に対する確率密度関数$p(x)$は解析的に計算不可能
  - ガウス分布 $(\alpha = 2)$ とコーシー分布 $(\alpha = 1)$ のみ解析的に表現可能
  - ただし、特性関数$\psi(t)$は簡潔に表現可能

$$\psi(t) = E_x\left[e^{itx}\right] = \int_{-\infty}^{\infty} e^{itx} p(x)\, dx = \exp(i\mu t - \sigma^{\alpha}|t|^{\alpha})$$

$p(x)$のフーリエ変換

次善の策として用いるべき確率分布の要件
- 裾が重い
- ガウス/コーシー分布を特殊形に含む
- 確率密度関数が解析的に表現可能



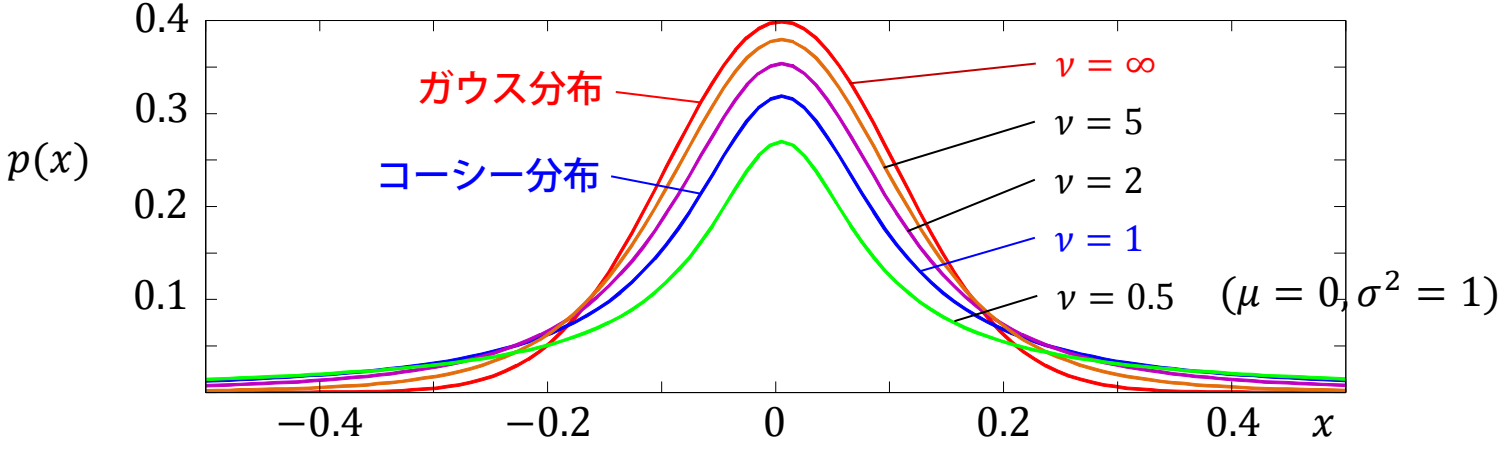
t分布

- ガウススケール混合分布の一種 (Gaussian scale mixture: GSM)
  - 裾が重い・コーシー分布・ガウス分布を特殊形に含む

$$p(x|\mu,\sigma^2,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\pi\nu\sigma^2}}\left(1+\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}} \Rightarrow N(x|\mu,\sigma^2)=\frac{1}{\sqrt{\pi\sigma^2}}\exp\left(\frac{(x-\mu)^2}{\sigma^2}\right)$$

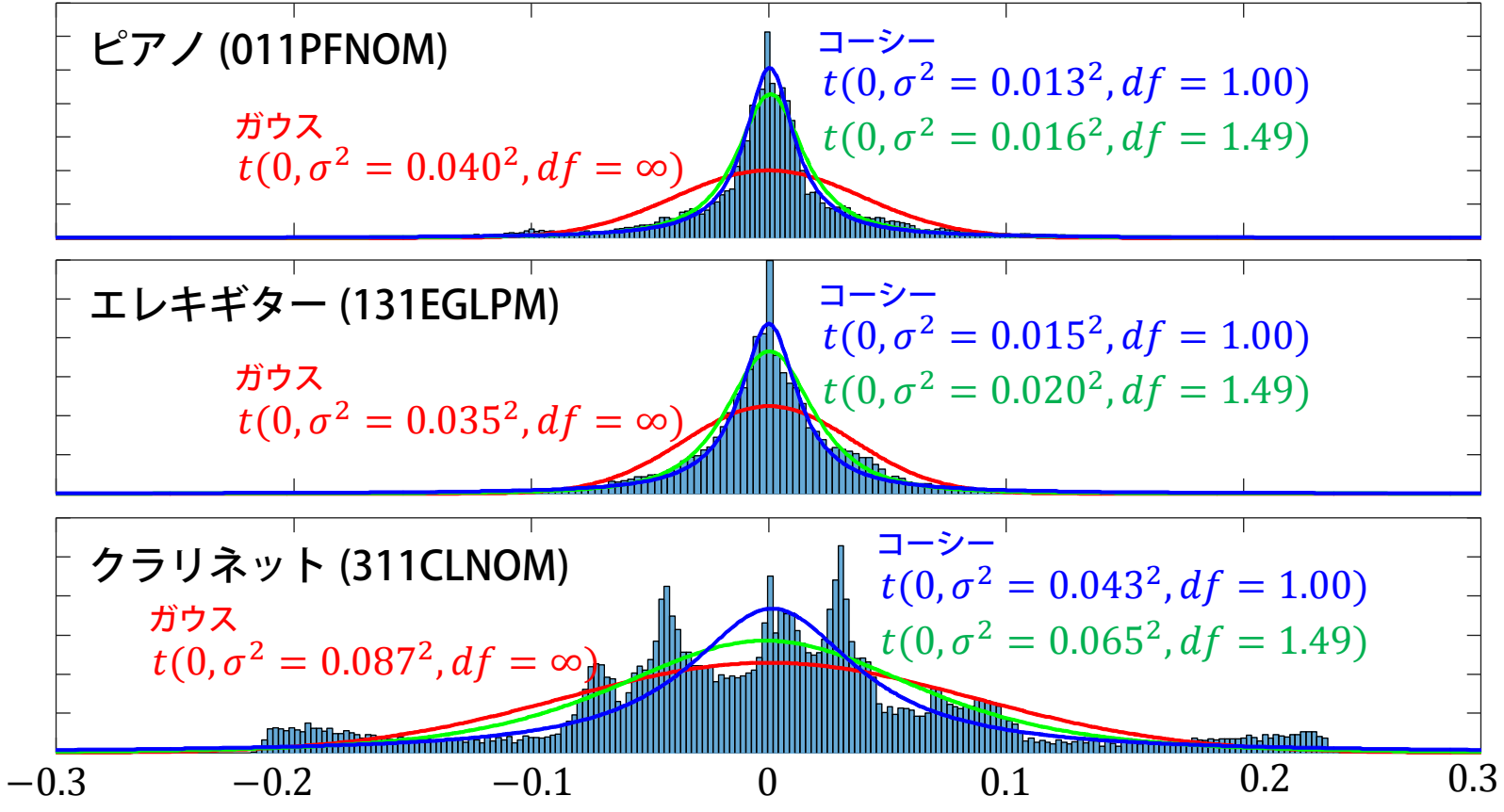
自由度

自由度$\nu\to\infty$でガウス分布に一致



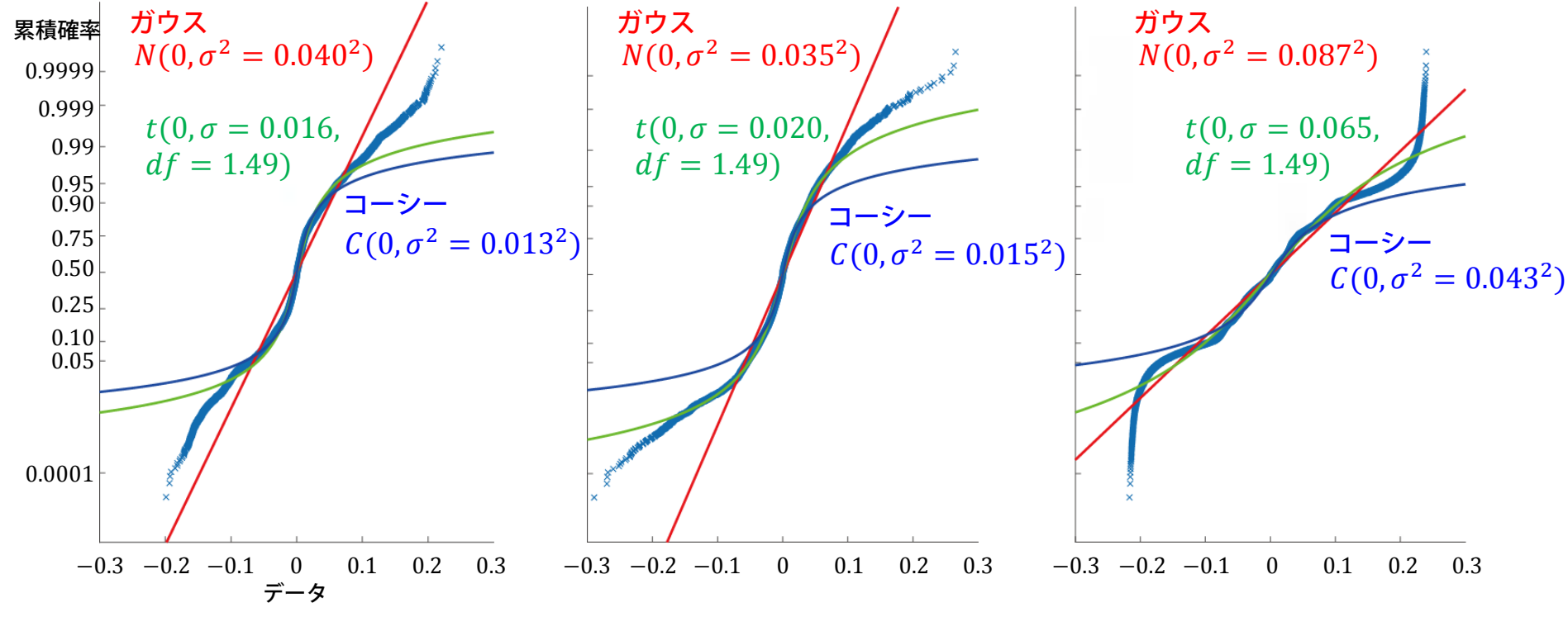
t分布による音響信号のモデル化

- t分布はガウス分布より当てはまりがよい



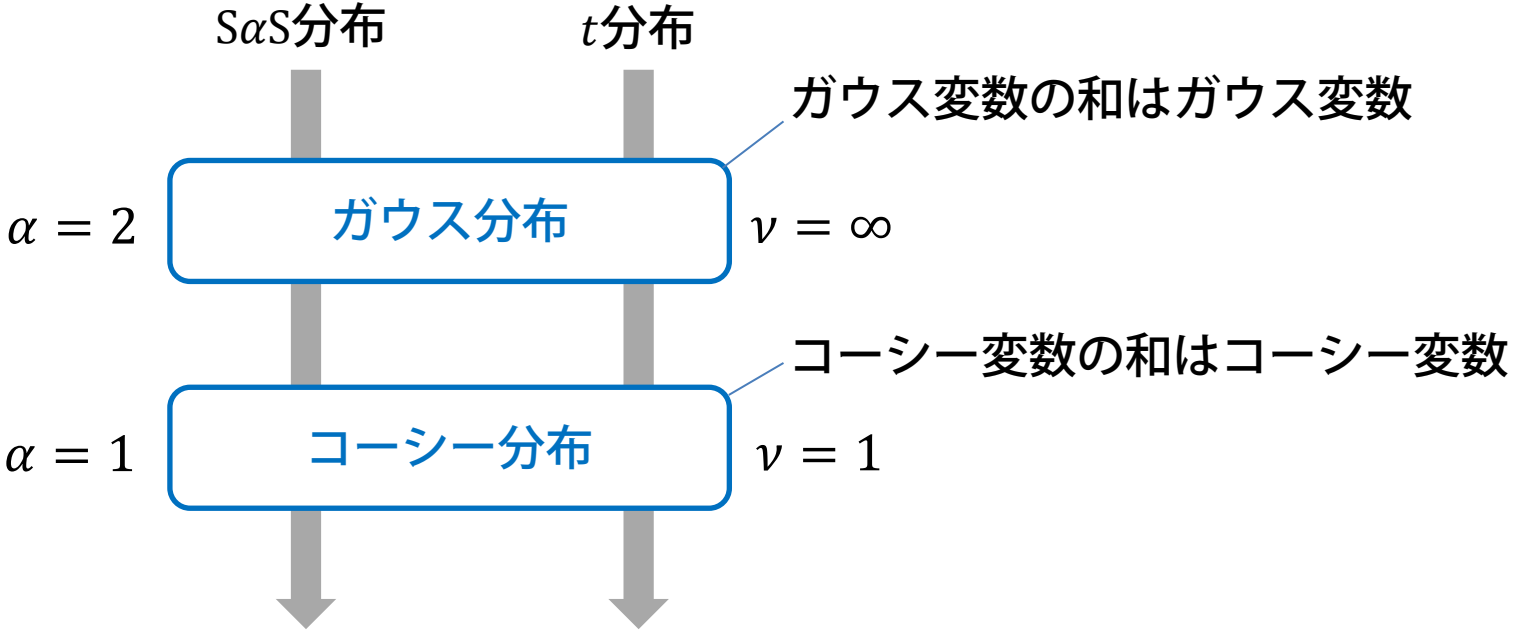
t分布による音響信号のモデル化

- t分布はガウス分布より当てはまりがよい
  - 中心部の当てはまりは良いが、裾の厚みを大きく見積もりすぎる傾向



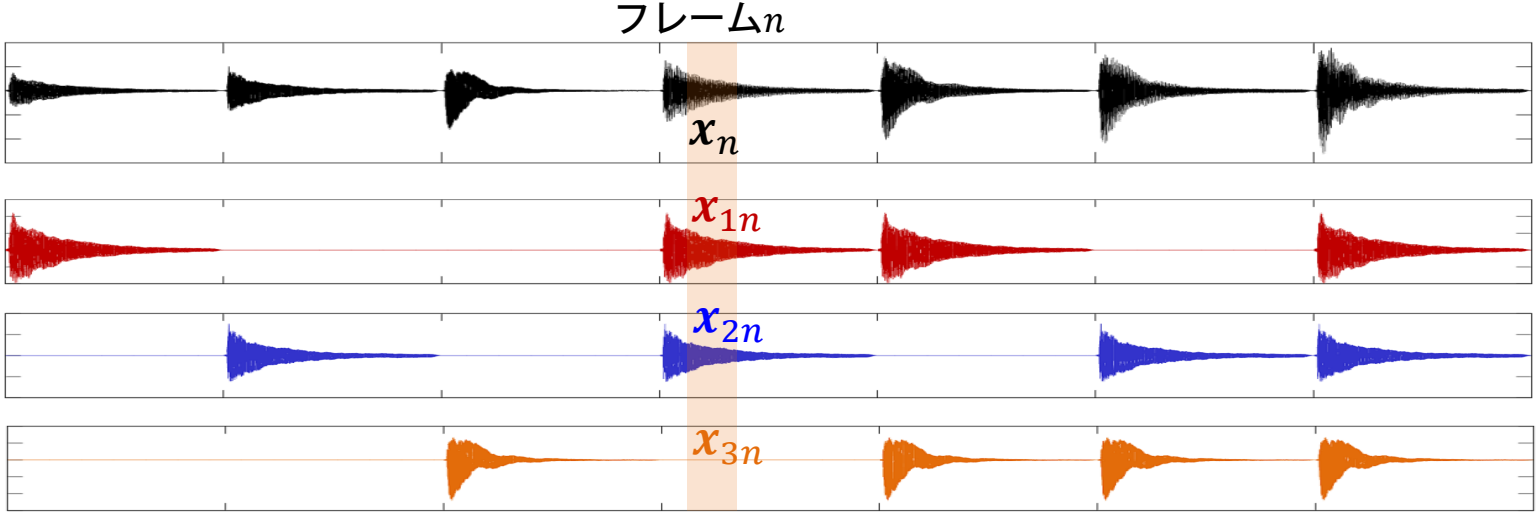
$S\alpha S$分布とt分布

- いずれもガウス分布とコーシー分布を特殊形として含む
  - 本来は$S\alpha S$分布が適切 [Liutkus 2015]
  - 確率密度関数が解析的に計算可能なt分布に着目



混合音の生成モデル

- ガウス変数 (局所信号) の足し合わせはガウス変数



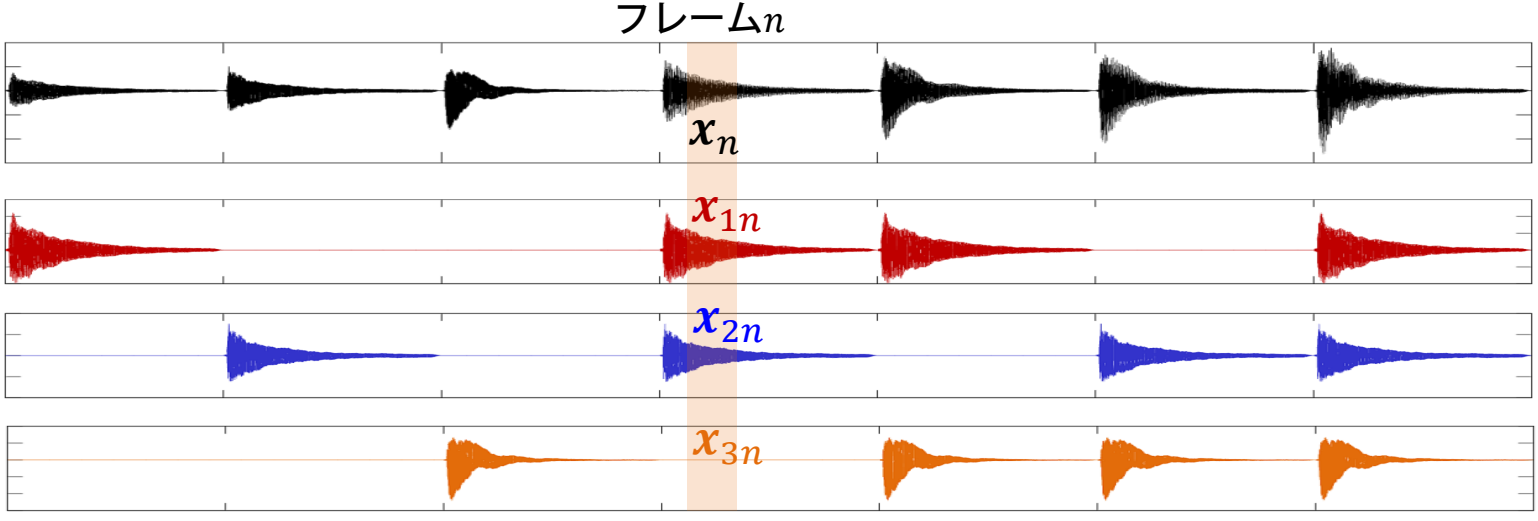
$$\left.\begin{aligned}\boldsymbol{x}_{1n} &\sim N(\mathbf{0}, h_{1n}\boldsymbol{W}_1)\\ \boldsymbol{x}_{2n} &\sim N(\mathbf{0}, h_{2n}\boldsymbol{W}_2)\\ \boldsymbol{x}_{3n} &\sim N(\mathbf{0}, h_{3n}\boldsymbol{W}_3)\end{aligned}\right\} \boldsymbol{x}_n \sim N\left(\mathbf{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k\right)$$

$\boldsymbol{x}_n\boldsymbol{x}_n^T \geqslant 0$が与えられた時に
$h_{kn} \geq 0$と$W_k \succcurlyeq 0$を求めたい

ガウス過程に基づくPSDTF (時間領域)

混合音の生成モデル

- コーシー変数 (局所信号) の足し合わせはコーシー変数



$$\left.\begin{aligned}\boldsymbol{x}_{1n} &\sim C(\mathbf{0}, h_{1n}\boldsymbol{W}_1)\\ \boldsymbol{x}_{2n} &\sim C(\mathbf{0}, h_{2n}\boldsymbol{W}_2)\\ \boldsymbol{x}_{3n} &\sim C(\mathbf{0}, h_{3n}\boldsymbol{W}_3)\end{aligned}\right\} \boldsymbol{x}_n \sim C\left(\mathbf{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k\right)$$

$\boldsymbol{x}_n\boldsymbol{x}_n^T \geqslant 0$が与えられた時に
$h_{kn} \geq 0$と$W_k \succcurlyeq 0$を求めたい

コーシー過程に基づくPSDTF (時間領域)

t-PSDTFの定式化

- 尤度関数としてガウス過程の代わりにt過程を利用
  - 加法性が成立する場合は限られる
    - $\nu = 1$：コーシー過程に基づくPSDTF
    - $\nu = \infty$：ガウス過程に基づくPSDTF
    - 他のνでも実用上うまくいく可能性はある
  - パラメータ$\boldsymbol{H}, \mathcal{W}$の更新則は補助関数を用いて導出
    - ガウス過程に基づくPSDTFの更新式を含む
  - 周波数領域での等価なモデル化も可能

$$\boldsymbol{x}_n \sim t\left(\boldsymbol{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k, \nu\right)$$

離散フーリエ変換行列$\boldsymbol{F}$による線形変換

$$\boldsymbol{F}\boldsymbol{x}_n = \hat{\boldsymbol{x}}_n \sim t_c\left(\boldsymbol{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{F}\boldsymbol{W}_k\boldsymbol{F}^H, \nu\right)$$

複素スペクトル

t-NMFの定式化

- 複素t分布に基づくt-NMF (周波数領域)

IS-NMF

$$\hat{x}_{nm} \sim N_c\left(0, \sum_{k=1}^{K} h_{kn} w_{km}, \nu\right)$$

複素スペクトルの値
(フレームn・ビンm)

$\Rightarrow$

t-NMF

$$\hat{x}_{nm} \sim t_c\left(0, \sum_{k=1}^{K} h_{kn} w_{km}, \nu\right)$$

更新式

$$w_{km} \leftarrow w_{km} \sqrt{\frac{\sum_n \frac{(\pi_n x_{nm}) h_{kn}}{y_{nm}^2}}{\sum_n \frac{h_{kn}}{y_{nm}}}} \qquad h_{kn} \leftarrow h_{kn} \sqrt{\frac{\sum_m \frac{(\pi_n x_{nm}) w_{km}}{y_{nm}^2}}{\sum_m \frac{w_{km}}{y_{nm}}}}$$

$\pi_n \to 1\ (\nu \to \infty)$でIS-NMFの更新式に帰着

$$\pi_n x_{nm} = \left(\frac{2}{2+\nu} y_{nm}^{-1} + \frac{\nu}{2+\nu} x_{nm}^{-1}\right)^{-1}$$

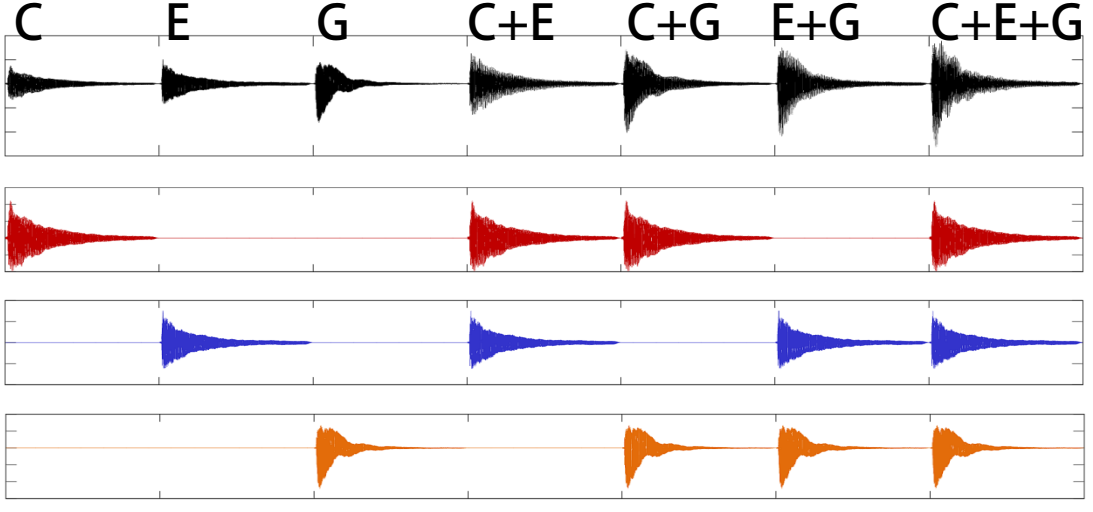
$$\to x_{nm}\ (\nu \to \infty)$$

観測データx_{nm}と再構成データy_{nm}を
ν: 2で調和平均をとったものを観測データと
みなしてIS-NMF (過学習の抑制効果？)

評価実験

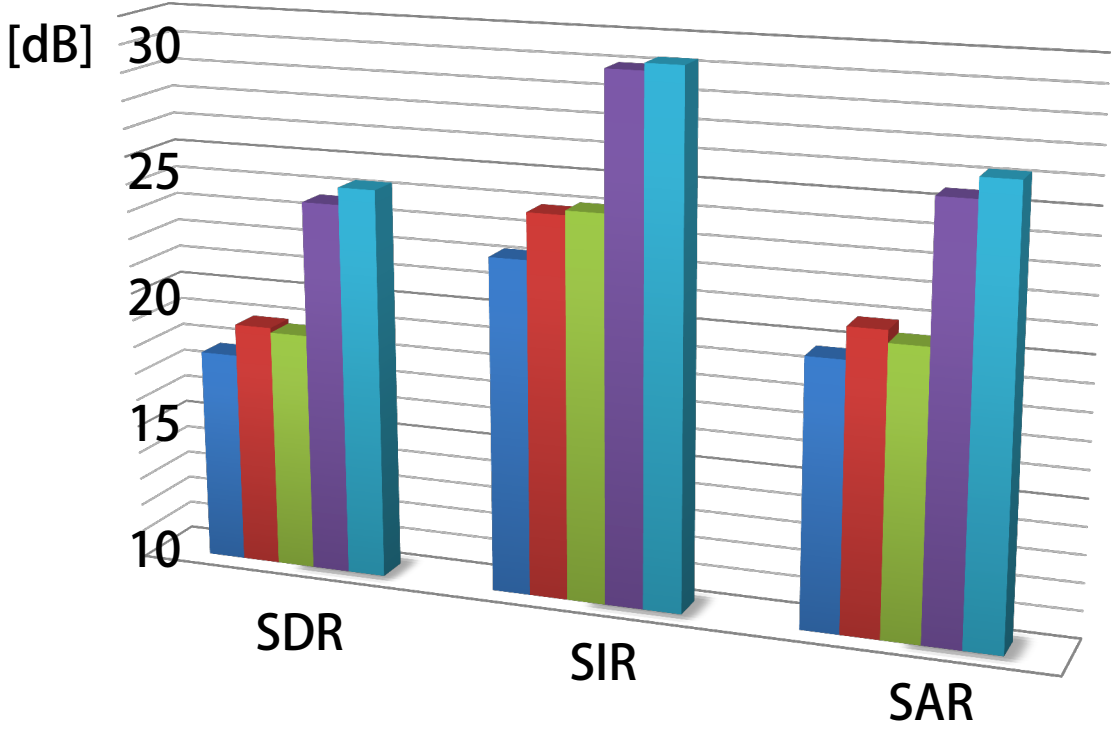
- 人工的に混合音を作成して基本的な音源分離性能を評価
  - 基底数はK=3 (C4, E4, G4に対応)
  - RWC研究用音楽データベースから3種類の楽器の単独音を使用
    - ピアノ・エレキギター (減衰音) / クラリネット (持続音)
  - BSS Eval Toolbox [Vincent2006] を使用して評価

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3$$



音源分離結果

- t-PSDTFは優れた音源分離品質を達成・さらなる調査が必要
  - IS-NMFよりt-NMFの方がわずかに精度が低いが局所解に陥りにくい傾向
  - LD-PSDTFよりt-PSDTFの方がわずかに精度が高い



まとめと今後の課題

- Done: t過程に基づく半正定値テンソル分解 (PSDTF)
  - ガウス過程($\nu = \infty$)とコーシー過程($\nu = 1$)を特殊形に含む
    - 再生性が成立するのはこれらのみ
  - 過学習抑制効果で局所解を回避しやすい傾向
    - t-NMFはIS-NMFの局所解問題の解決策？
    - νを徐々に上昇させるアニーリングが有効？
- Todo: SαS過程に基づく半正定値テンソル分解
  - ガウス過程とコーシー過程を特殊形に含む
    - 再生性が全てのαで成立
  - より音響信号にフィットしたモデル化が可能