モノラル音響信号に対する音源分離のための無限半正定値テンソル分解

PSDTF

吉井 和佳[1)]　富岡 亮太[2)]
持橋 大地[3)]　後藤 真孝[1)]

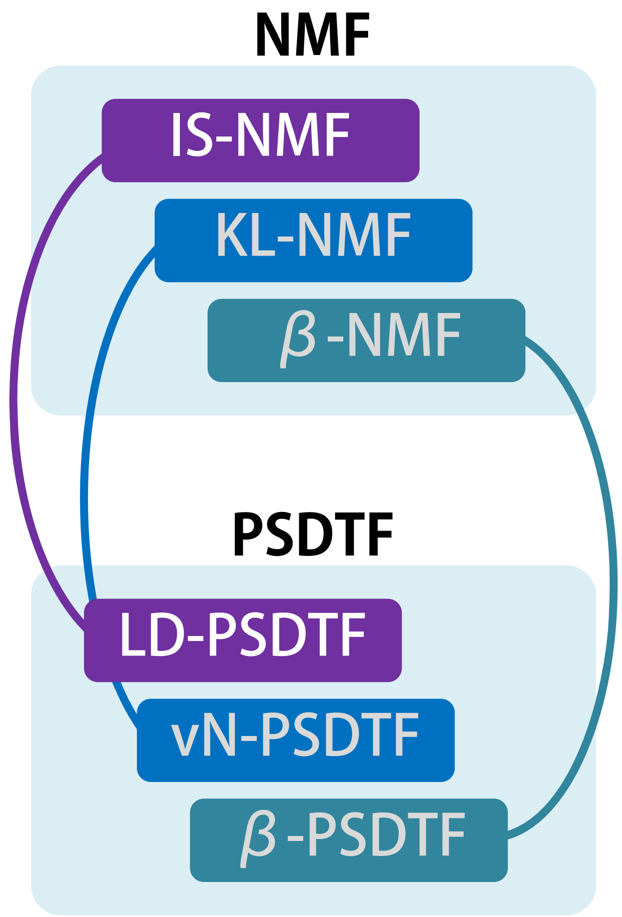
1) 産業技術総合研究所 (AIST)
2) 東京大学
3) 統計数理研究所 (ISM)

NMFの”美しい”数学的拡張

PSDTFを発見した！

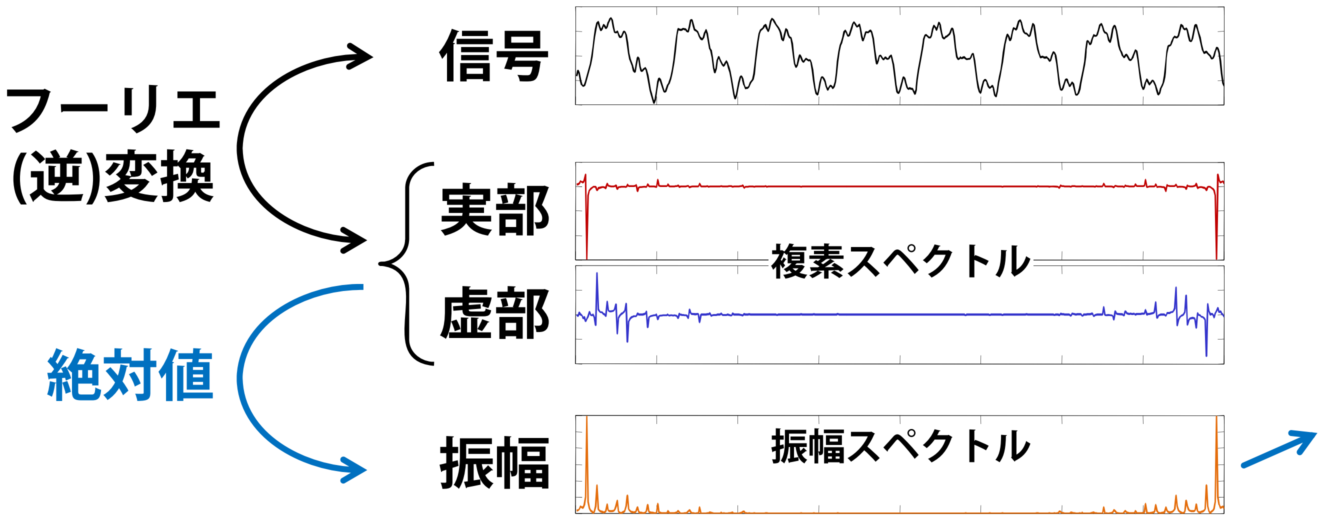
覚えて帰ってほしいこと

- 半正定値テンソル分解 (PSDTF)
  - 非負値行列分解 (NMF) の数学的に自然な拡張
  - 基底数の自動決定が可能
    - ノンパラメトリックベイズモデル
- Log-Determinant PSDTF (LD-PSDTF)
  - Itakura-Saito NMF (IS-NMF) の自然な拡張
  - 補助関数法に基づく反復最適化
    - 最尤推定：乗法更新則
    - ベイズ推定：変分ベイズ法
- モノラル音源分離への応用
  - NMFより優れた分離品質を達成
  - 計算量の削減が課題



研究の背景

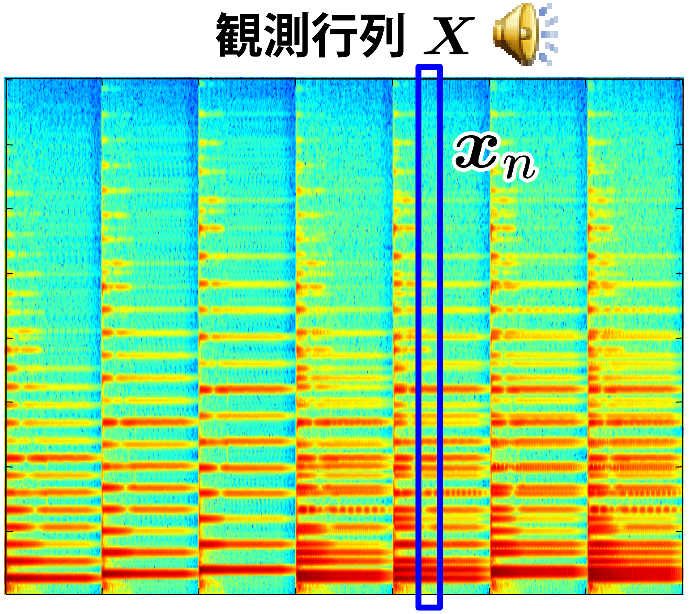
- モノラル音響信号の音源分離に関する研究がさかん
  - 応用例：音楽信号の高度な加工
    - 歌声と伴奏の分離 [Rafii2011, Huang2012]
    - 楽器音イコライザ [吉井2006, 宮本2008, 糸山2008]
- スペクトル領域で音源分離を行うのが一般的
  - 位相を無視すると音の特徴がとらえやすくなる
    - 調波構造・スパース性などに着目



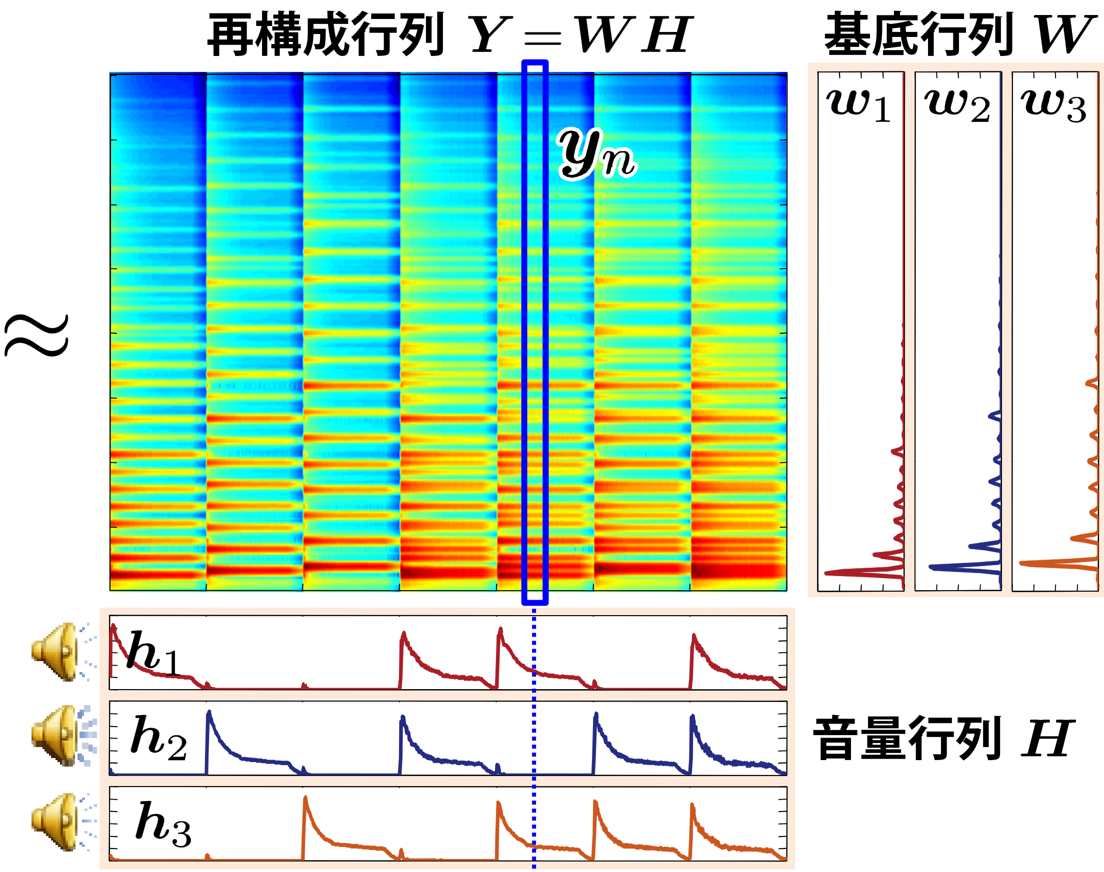
非負値行列分解 (NMF)
に基づく手法が席巻
- 直感に合う
- 実装が簡単
- 高速に動作

非負値行列分解 (NMF)

- 「非負のベクトル」を少数の「非負のベクトル」の凸結合で表現



$\approx$



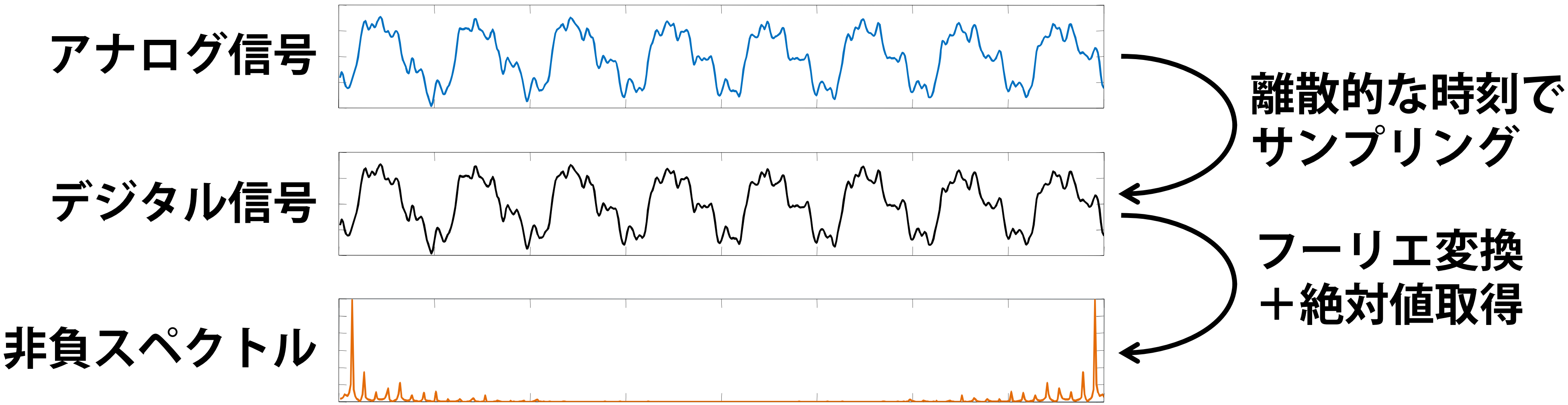
$$\boldsymbol{x}_n \approx \boldsymbol{y}_n = \sum_{k=1}^{K} h_{kn} \boldsymbol{w}_k$$

よく利用されるコスト関数：Bregmanダイバージェンス

$$\mathcal{D}_\phi(\boldsymbol{x}_n | \boldsymbol{y}_n) = \phi(\boldsymbol{x}_n) - \phi(\boldsymbol{y}_n) - \phi'(\boldsymbol{y}_n)^T (\boldsymbol{x}_n - \boldsymbol{y}_n)$$

研究の動機

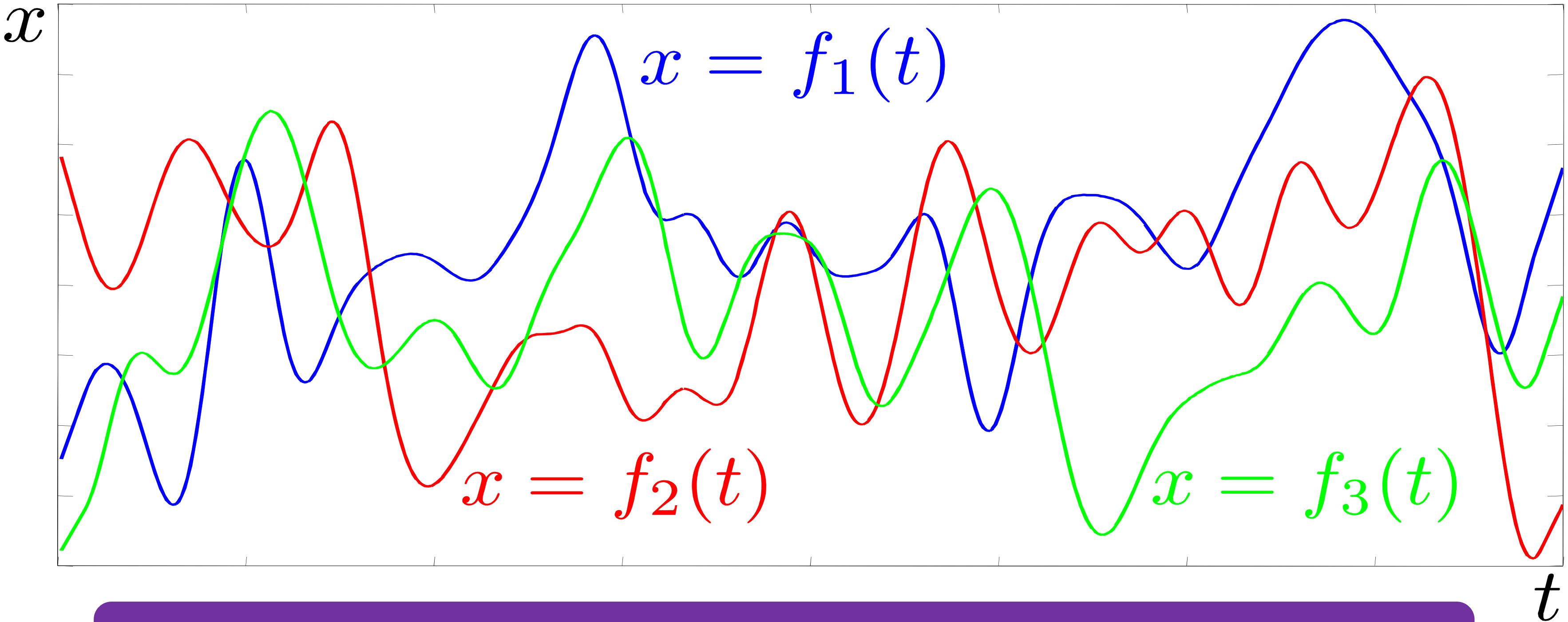
- 音響信号とは何か、原点にたちかえって考えよう！
  - 非負スペクトログラムを出発点 → 不適切
    - 位相情報は重要
  - デジタル信号を出発点 → 不適切
    - 空気振動 (物理現象) は連続時間上に存在



アナログ信号を直接モデル化する方法はあるだろうか？

ガウス過程 (GP)

- 連続時間上の確率過程の一種
  - 例：白色雑音・ブラウン運動など
  - 連続関数 f に対する確率分布 $p(f)$

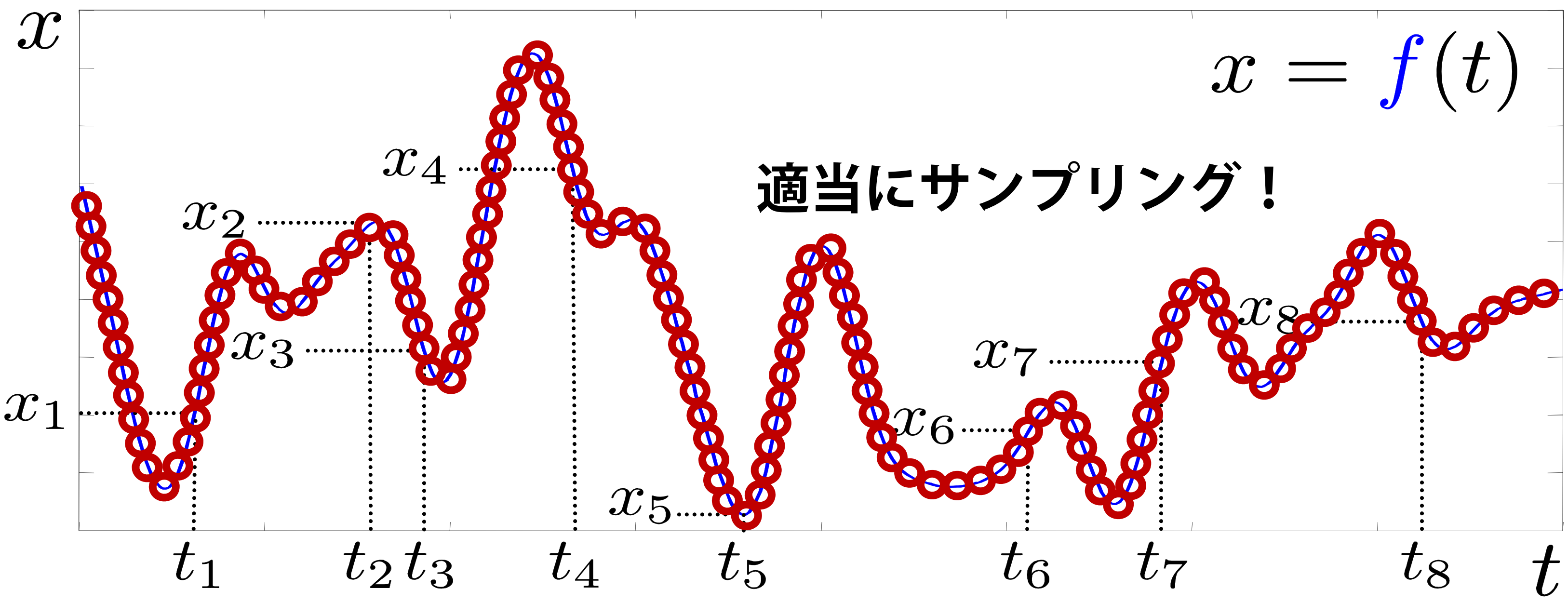


あらゆる信号波形の「もっともらしさ」を評価可能！

$p(f_1)$ $p(f_2)$ $p(f_3)$

ガウス過程はこわくない

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



  - 任意の M 点の周辺分布はガウス分布！

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

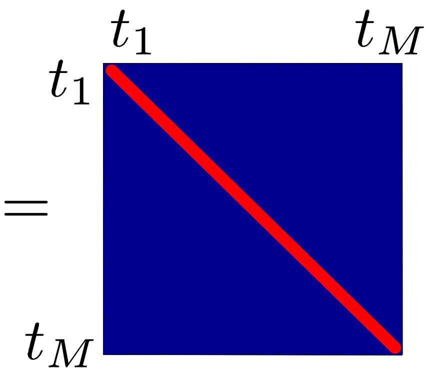
$$\boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T \qquad \boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K})$$

カーネル行列
(半正定値)

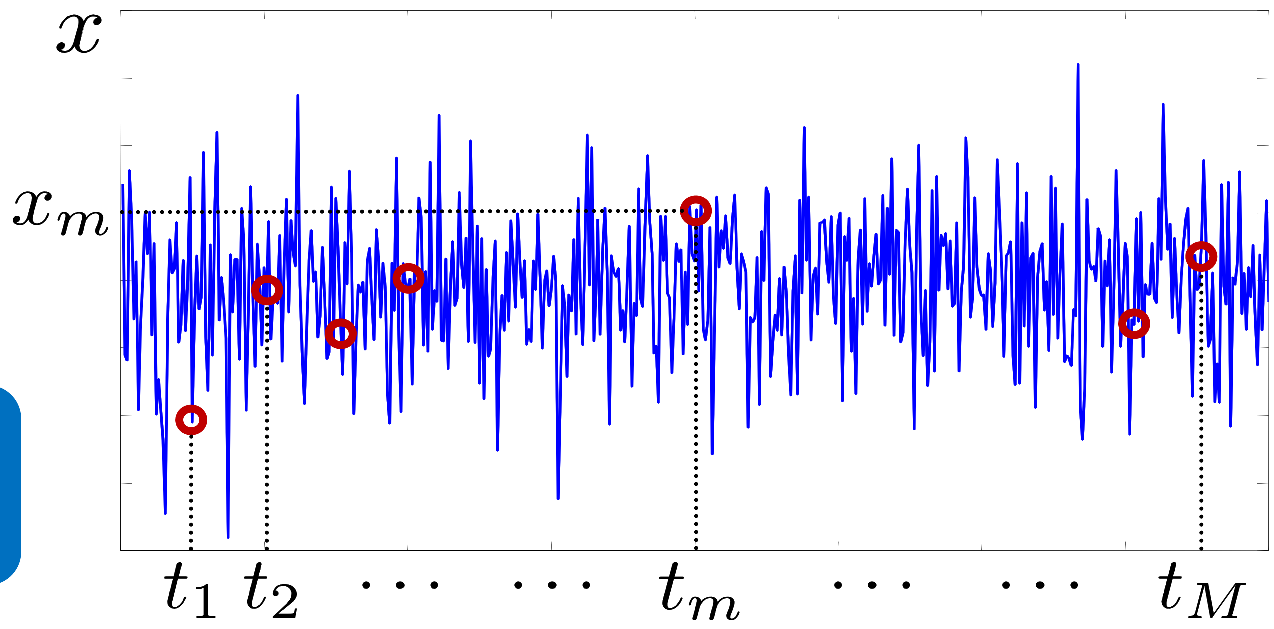
単位行列カーネル

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

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

$\boldsymbol{K} =$ 

単位行列であれば
各 x_m は独立

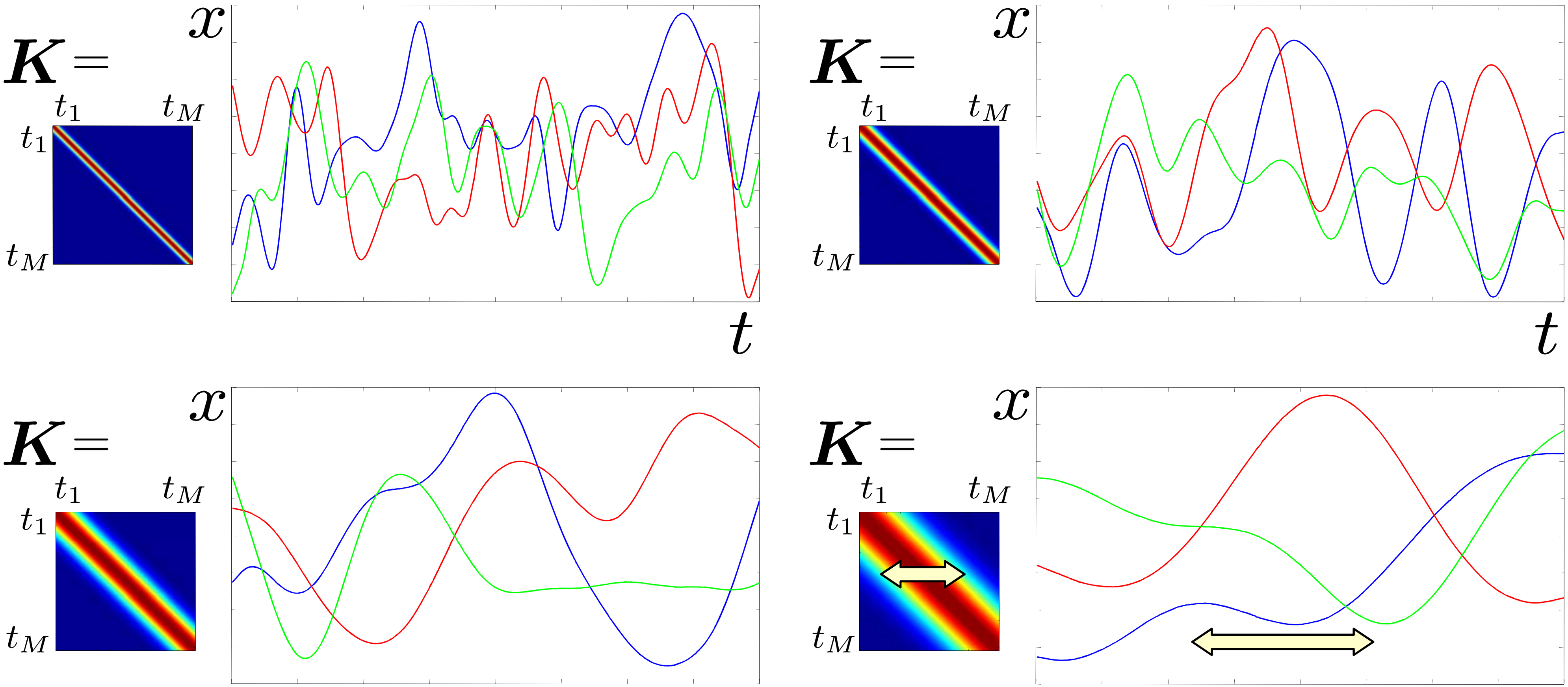


等間隔の離散的なサンプリング時刻を考える

ガウス関数型カーネル

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

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

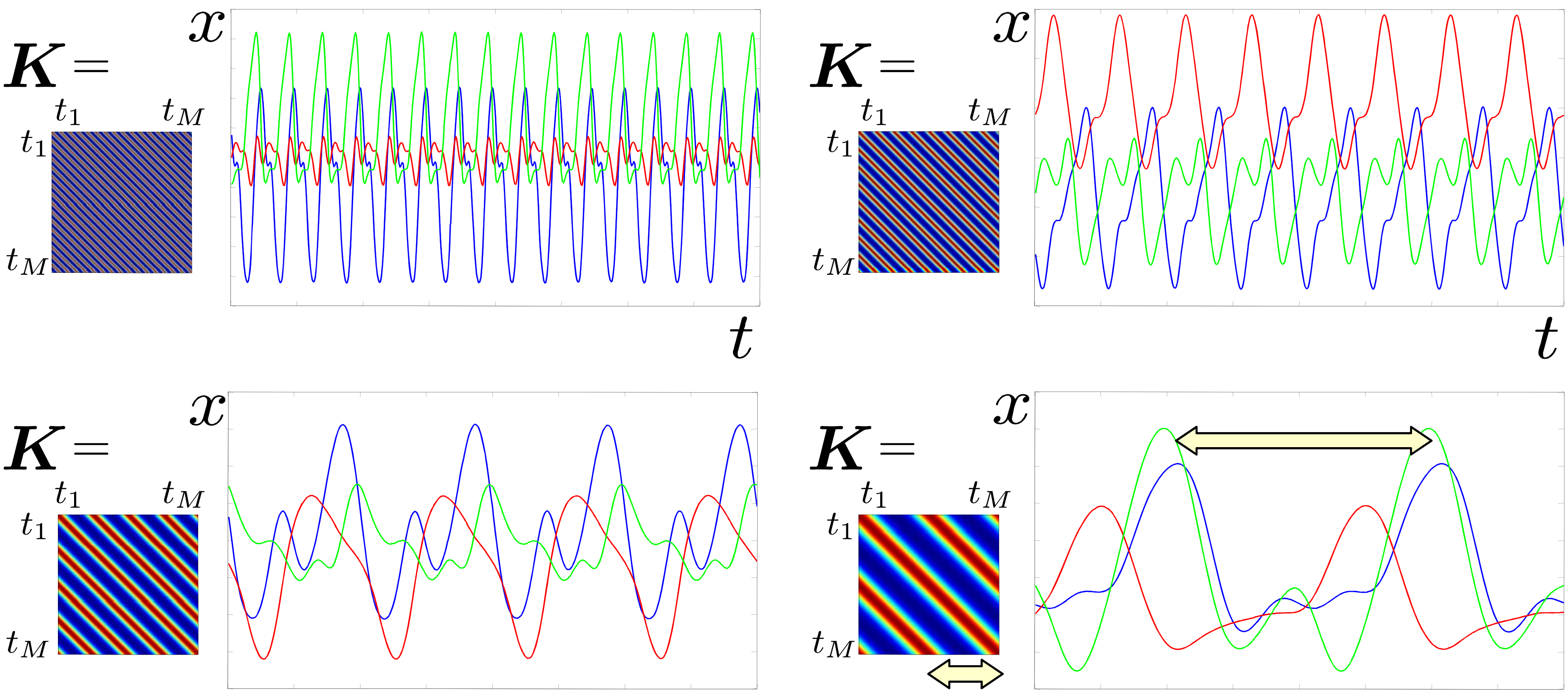


近傍間は相関が強い
→ 滑らかな変化

周期カーネル

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

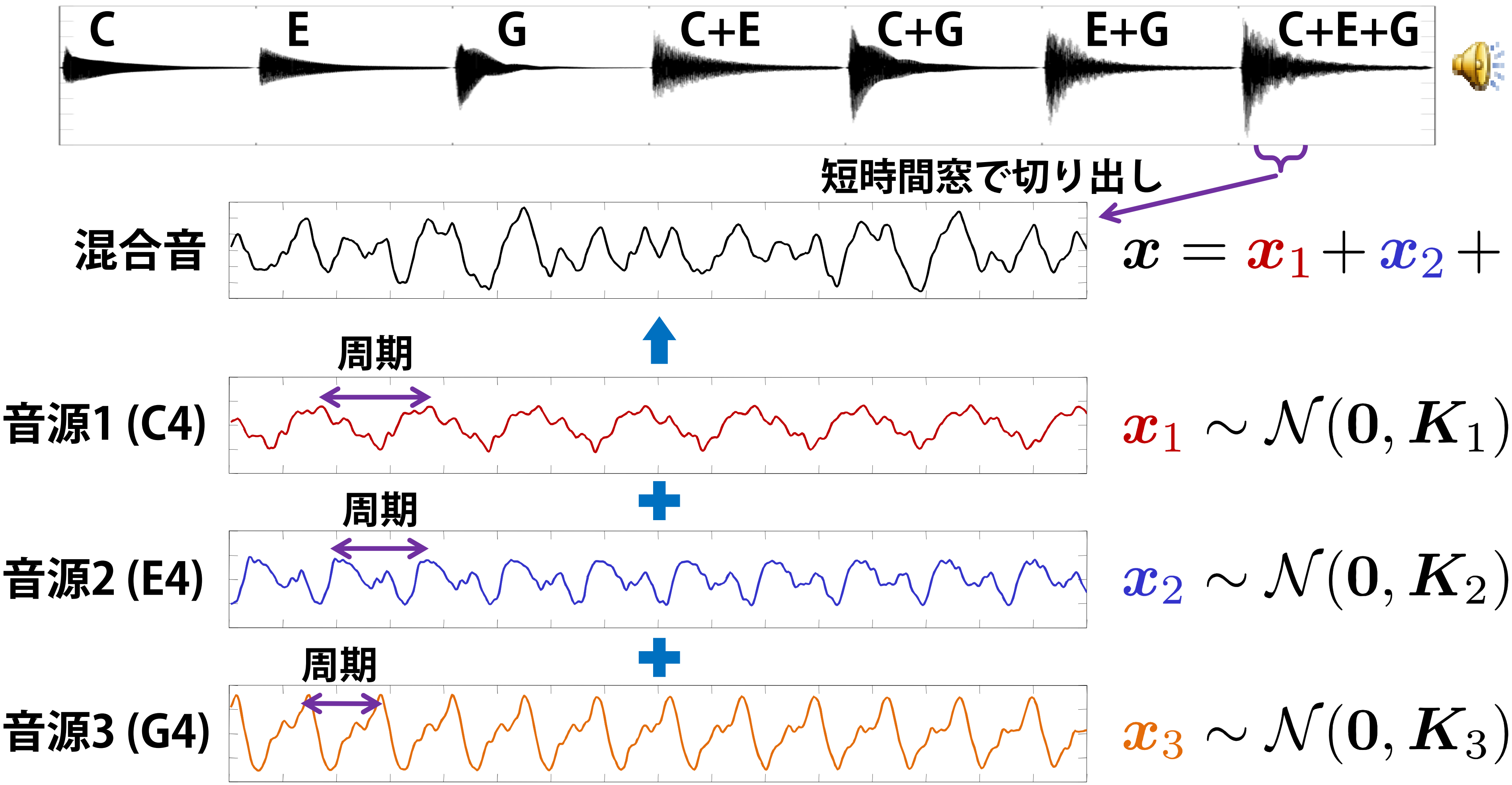
$$\boldsymbol{t} = [t_1, t_2, \cdots, t_M]^T$$
$$\downarrow f$$
$$\boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T$$



一定間隔で相関が強い
→ 周期的な変化

音響信号に適用できるか？

- 混合音はガウス過程の重ね合わせ
  - 各音源信号は周期カーネルをもつガウス過程？



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

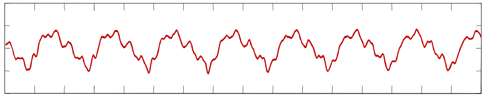
$$\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K}_1)$$

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

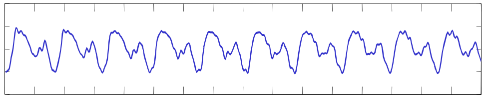
$$\boldsymbol{x}_3 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K}_3)$$

音響信号の生成と分離

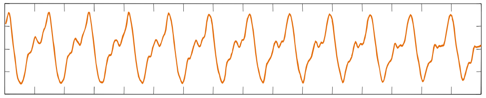
- **順問題：ガウス変数の和 → ガウス変数**



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



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



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

?

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

- **逆問題：ガウス変数の分解 → ガウス変数**

?

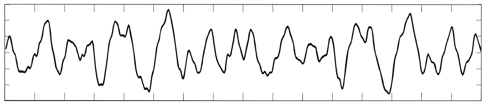
$\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K}_1)$

?

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

?

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



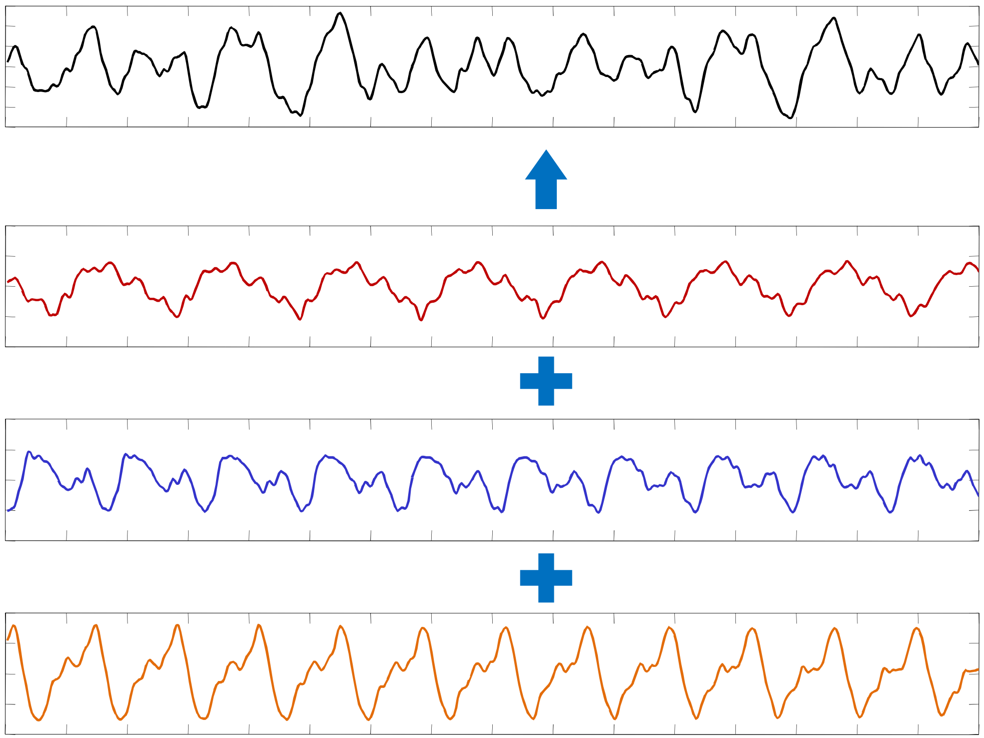
$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3$$
$$\sim \mathcal{N}(\mathbf{0}, \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3)$$

$$\boldsymbol{x}_1 | \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{K}_1 \boldsymbol{K}^{-1} \boldsymbol{x},\ \boldsymbol{K}_1 - \boldsymbol{K}_1 \boldsymbol{K}^{-1} \boldsymbol{K}_1)$$
$$\boldsymbol{x}_2 | \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{K}_2 \boldsymbol{K}^{-1} \boldsymbol{x},\ \boldsymbol{K}_2 - \boldsymbol{K}_2 \boldsymbol{K}^{-1} \boldsymbol{K}_2)$$
$$\boldsymbol{x}_3 | \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{K}_3 \boldsymbol{K}^{-1} \boldsymbol{x},\ \boldsymbol{K}_3 - \boldsymbol{K}_3 \boldsymbol{K}^{-1} \boldsymbol{K}_3)$$

カーネルさえ求まれば
ウィナーフィルタで
音源分離が可能！

ここで問題

- **観測信号だけからどうやってカーネルを求めるか？**
  - **カーネルの種類すら不明**



観測信号

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

音源信号1

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

音源信号2

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

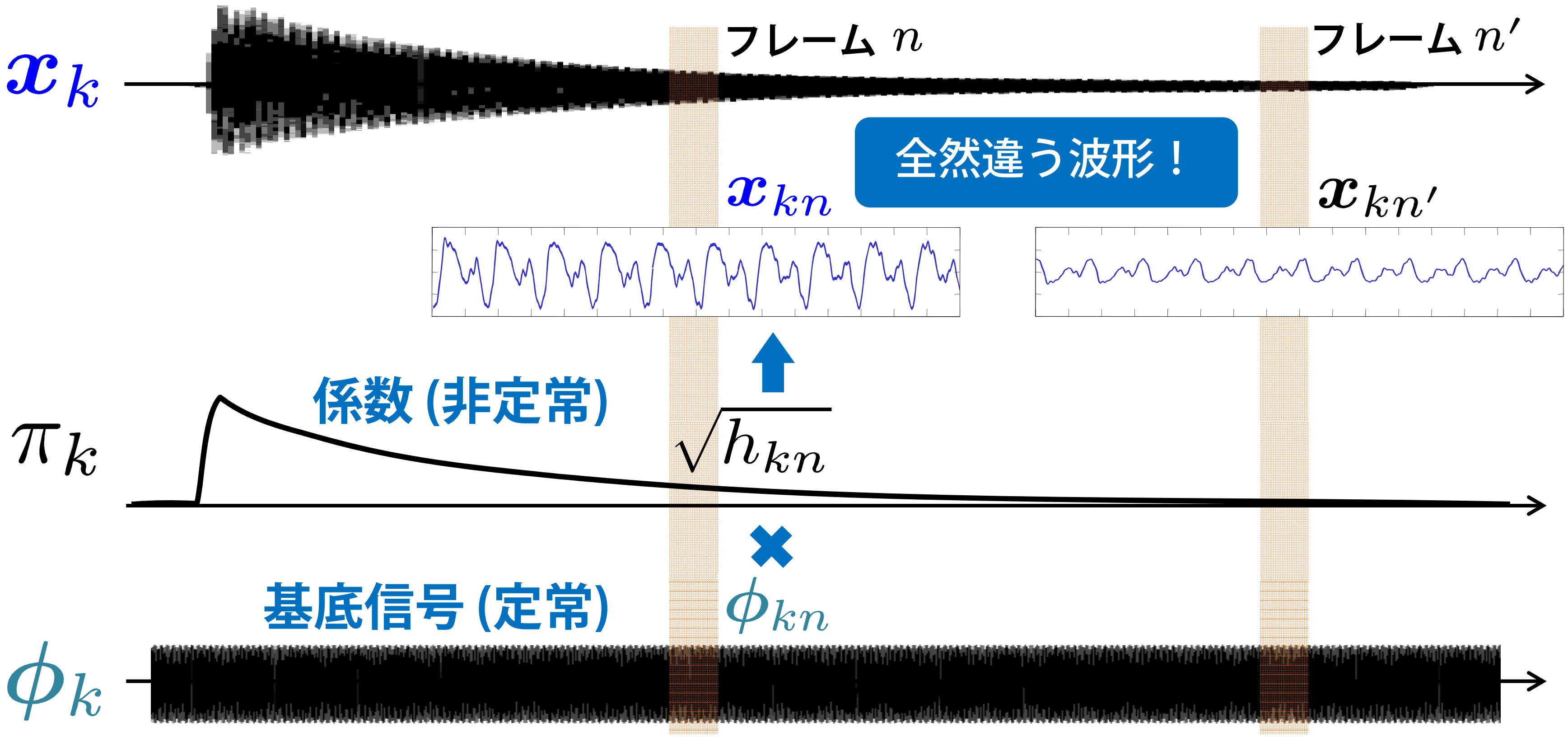
音源信号3

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

音楽信号の「非定常性」に着目する！

非定常な楽器音のモデル

- 非定常 = 時刻によって各種統計量 (例えば分散) が変化



$$\phi_{kn} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{V}_k) \longrightarrow \boldsymbol{x}_{kn} \sim \mathcal{N}(\mathbf{0}, h_{kn}\boldsymbol{V}_k)$$

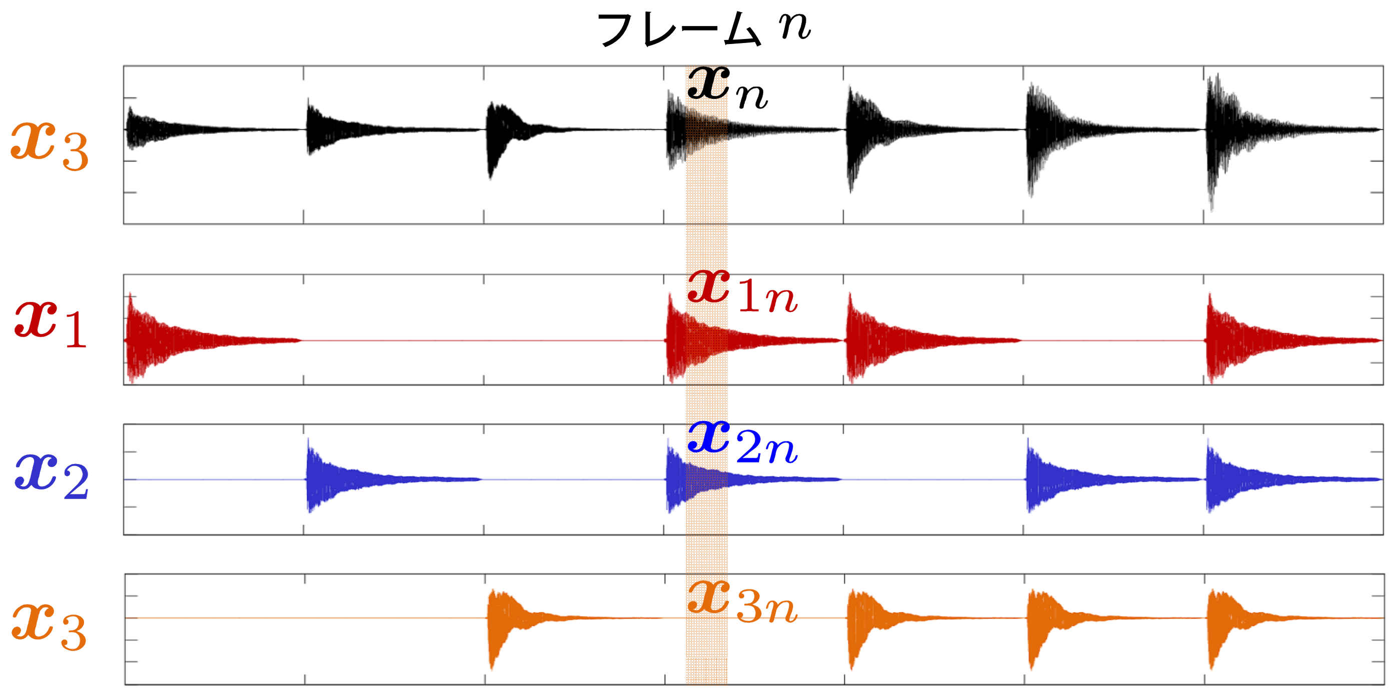
定常 → nに依存しない

≥ 0 非負　$\succeq 0$ 半正定

混合音の確率モデル

- **ガウス変数 (局所信号) の足し合わせはガウス変数**

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



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

時変スケール×カーネル

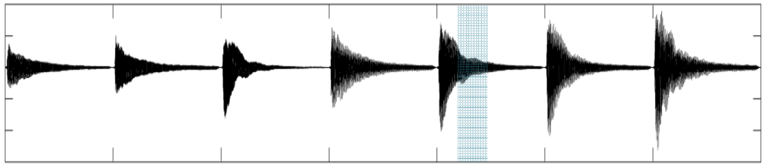
半正定 $\succeq 0$

非負 ≥ 0

最尤推定

- 対数尤度を最大化するパラメータH,Vを求めたい

混合音 $\boldsymbol{x}_n$



$$\boldsymbol{x}_n \sim \mathcal{N}\left(\mathbf{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{V}_k\right) \longrightarrow \text{最大化}$$

半正定値行列　半正定値行列

$$\boldsymbol{X}_n = \boldsymbol{x}_n\boldsymbol{x}_n^T \qquad \boldsymbol{Y}_n = \sum_{k=1}^{K} h_{kn}\boldsymbol{V}_k$$ と定義

ガウス変数の対数尤度

$$\log p(\boldsymbol{X}_n|\boldsymbol{Y}_n) \stackrel{c}{=} -\frac{1}{2}\log|\boldsymbol{Y}_n| - \frac{1}{2}\mathrm{tr}(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}) \longrightarrow \text{最大化}$$

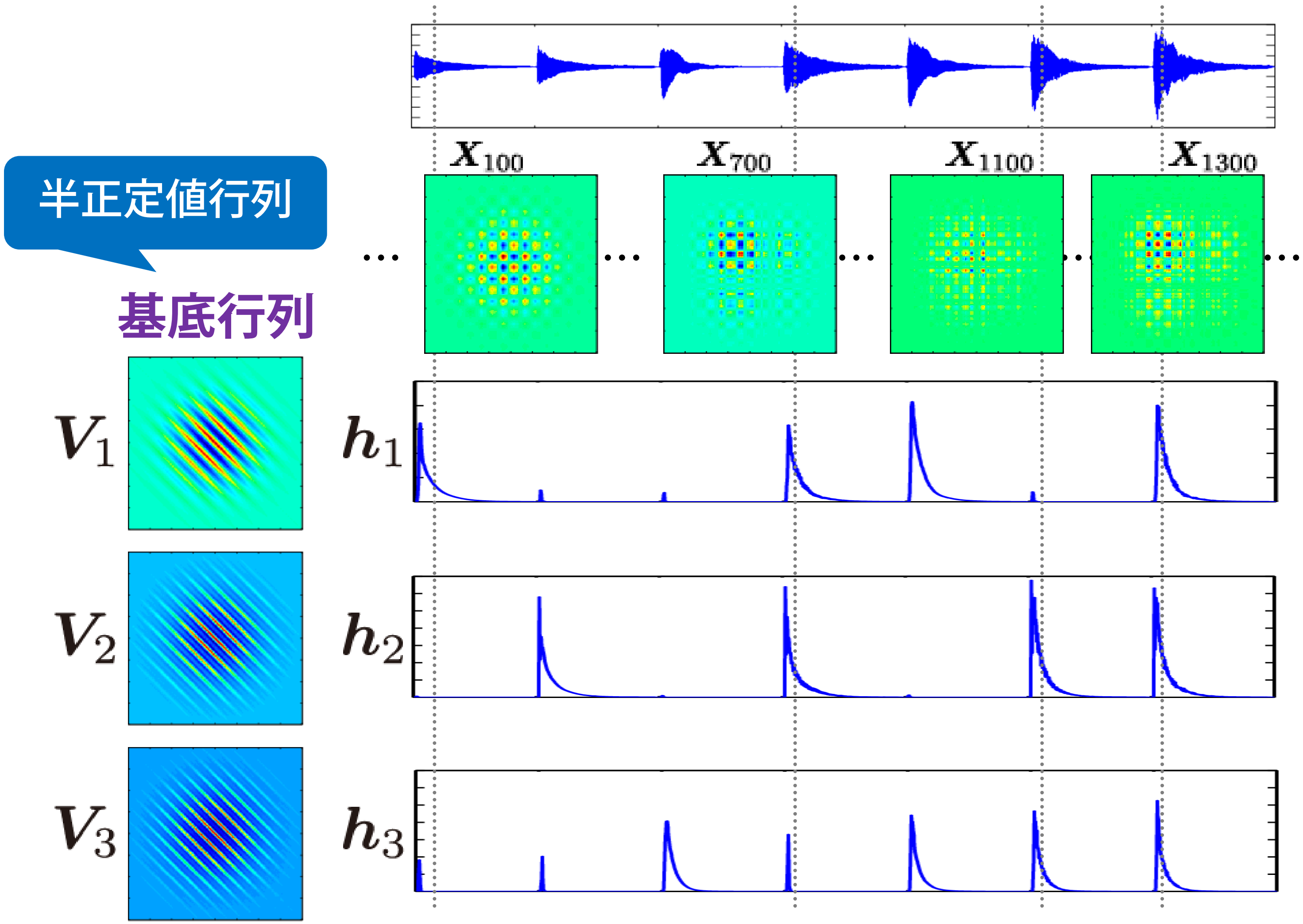
等価！

Log-Detダイバージェンス

$$\mathcal{D}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = -\log\left|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right| + \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) - M \longrightarrow \text{最小化}$$

半正定値テンソル分解

- 「半正定値行列」を少数の「半正定値行列」の凸結合で表現



半正定値行列

観測行列

$$X_n = x_n x_n^T$$

Log-Det
ダイバージェンス
$\mathcal{D}_{\mathrm{LD}}(X_n | Y_n)$
を最小化！

$$Y_n = \sum_{k=1}^{K} h_{kn} V_k$$

再構成行列

NMF と PSDTF

- 非負値行列分解 (NMF)
  - 非負ベクトルを非負ベクトルの和に分解 $\boldsymbol{x}_n \approx \boldsymbol{y}_n = \sum_{k=1}^{K} h_{kn} \boldsymbol{w}_k$
  - Bregmanダイバージェンス
    - カルバック・ライブラ (KL) ダイバージェンス
      $\mathcal{D}_{\mathrm{KL}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_m (x_{mn} \log x_{mn} y_{mn}^{-1} - x_{mn} + y_{mn})$
    - 板倉・齋藤 (IS) ダイバージェンス
      $\mathcal{D}_{\mathrm{IS}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_m (-\log x_{mn} y_{mn}^{-1} + x_{mn} y_{mn}^{-1} - 1)$
- 半正定値テンソル分解 (PSDTF)
  - 半正定値行列を半正定値行列の和に分解 $\boldsymbol{X}_n \approx \boldsymbol{Y}_n = \sum_{k=1}^{K} h_{kn} \boldsymbol{V}_k$
  - Bregman"行列"ダイバージェンス
    - フォン・ノイマン (vN) ダイバージェンス
      $\mathcal{D}_{\mathrm{vN}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = \mathrm{tr}\left(\boldsymbol{X}_n \log \boldsymbol{X}_n \boldsymbol{Y}_n^{-1} - \boldsymbol{X}_n + \boldsymbol{Y}_n\right)$
    - Log-Det (LD) ダイバージェンス
      $\mathcal{D}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = -\log\left|\boldsymbol{X}_n \boldsymbol{Y}_n^{-1}\right| + \mathrm{tr}\left(\boldsymbol{X}_n \boldsymbol{Y}_n^{-1}\right) - M$

半正定性

- 実対称行列 $A = A^T$ が半正定
  ⇔ すべての固有値が非負
  ⇔ $A = XX^T$ をみたす実行列 X が存在
- 複素エルミート行列 $A = A^H$ が半正定
  ⇔ すべての固有値が非負
  ⇔ $A = XX^H$ をみたす複素行列 X が存在

任意の実数の二乗は常に非負

共役な複素数の積は常に非負

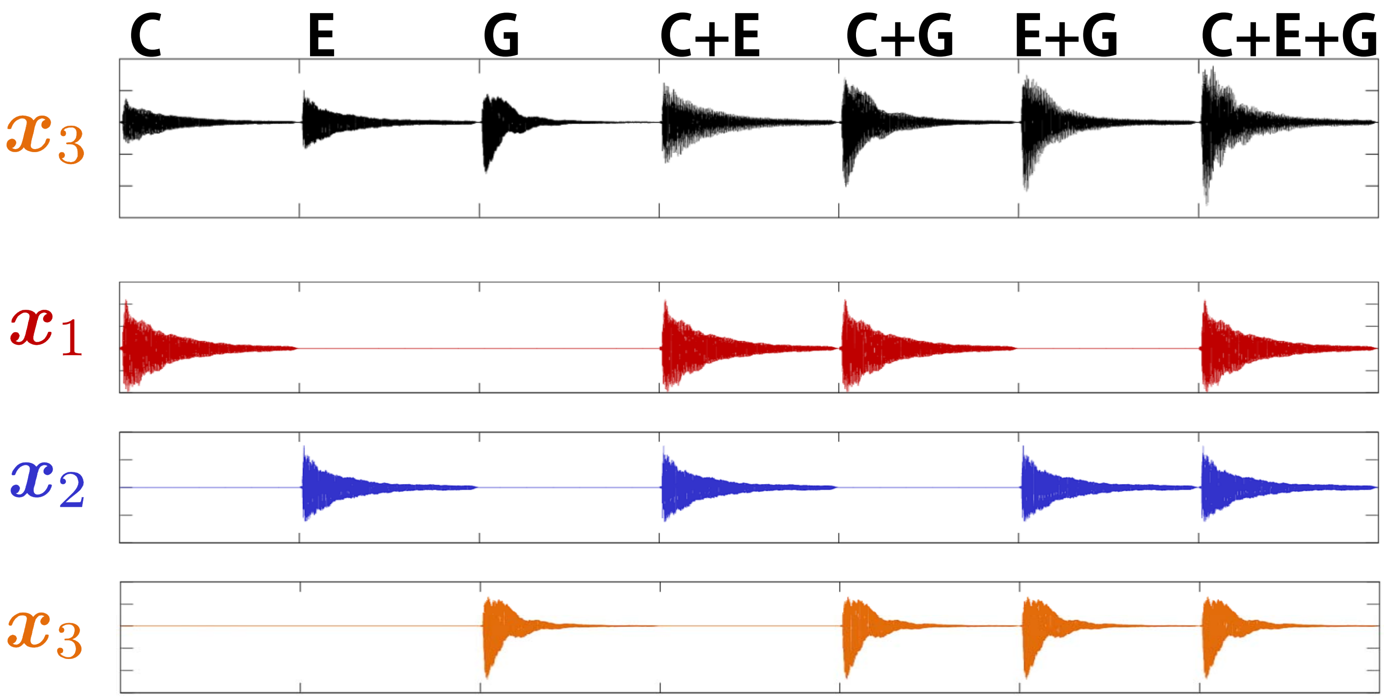
「行列の半正定性」は「スカラ/ベクトルの非負性」の拡張概念！

非負値	＋	非負値	＝	非負値
半正定値行列	＋	半正定値行列	＝	半正定値行列
非負値	−	非負値	≠	非負値
半正定値行列	−	半正定値行列	≠	半正定値行列

評価実験

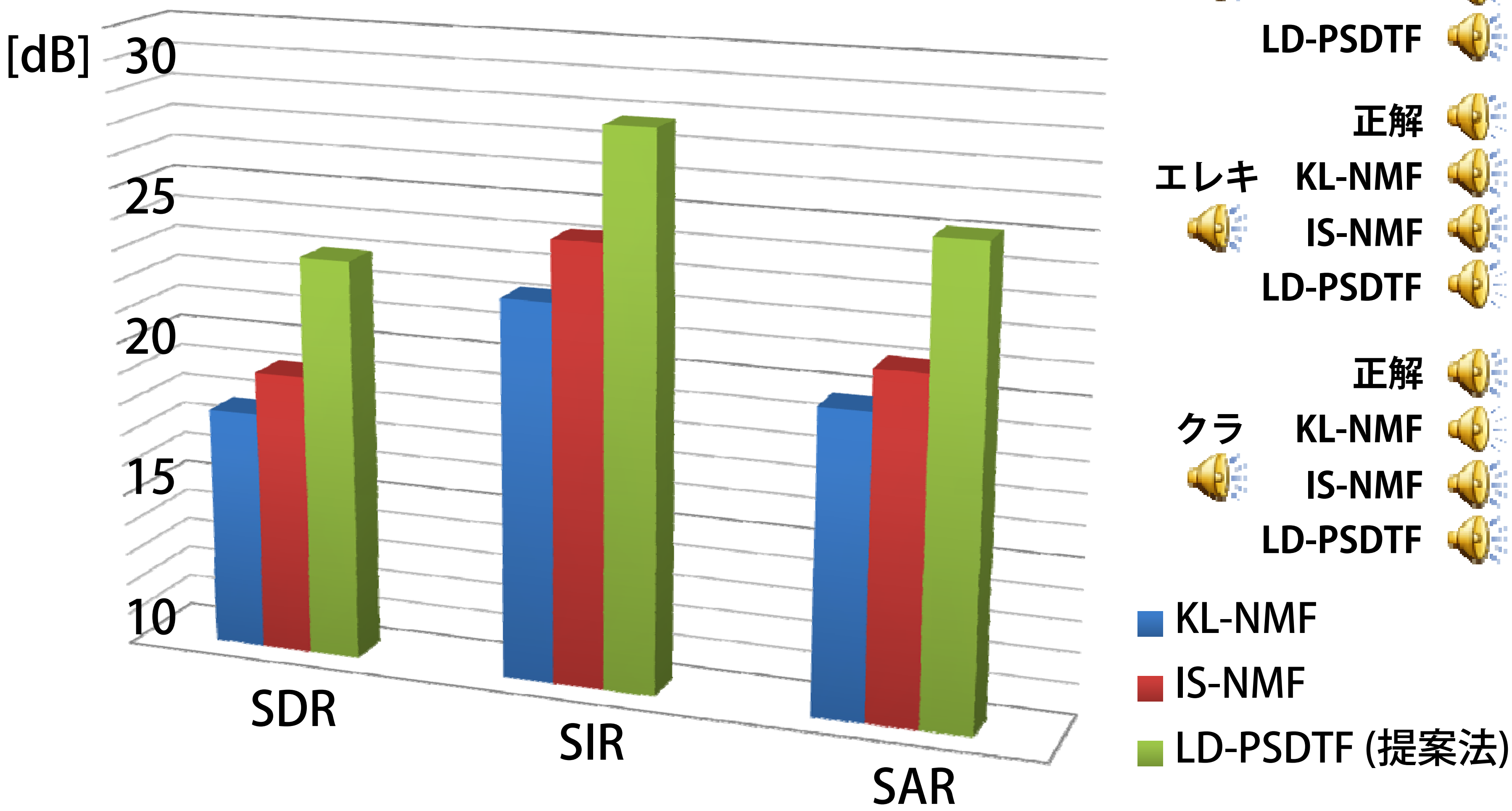
- 人工的に混合音を作成して基本的な音源分離性能を評価
  - 基底数はK=3 (C4, E4, G4に対応)
  - RWC研究用音楽データベースから3種類の楽器の単独音を使用
    - ピアノ・エレキギター (減衰音) / クラリネット (持続音)
  - BSS Eval Toolbox [Vincent2006] を使用して評価

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



音源分離結果

- **KL-NMF や IS-NMF より優れた音源分離品質を達成**
  - **課題：計算量の削減・複雑な信号への適用**



ピアノ
- 正解
- KL-NMF
- IS-NMF
- LD-PSDTF

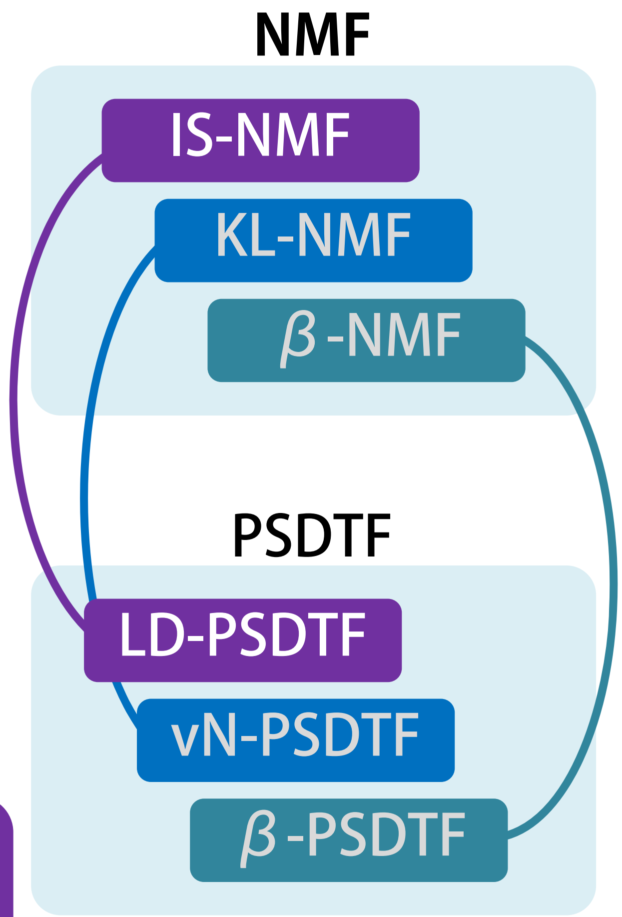
エレキ
- 正解
- KL-NMF
- IS-NMF
- LD-PSDTF

クラ
- 正解
- KL-NMF
- IS-NMF
- LD-PSDTF

NMFの”美しい”数学的拡張
PSDTFを発見した！

まとめと今後の課題

- 半正定値テンソル分解 (PSDTF)
  - 非負値行列分解 (NMF) の数学的に自然な拡張
  - 基底数の自動決定が可能
    - ノンパラメトリックベイズモデル
- Log-Determinant PSDTF (LD-PSDTF)
  - Itakura-Saito NMF (IS-NMF) の自然な拡張
  - 補助関数法に基づく反復最適化
- モノラル音源分離への応用
  - NMFより優れた分離品質を達成
  - 計算量の削減が課題

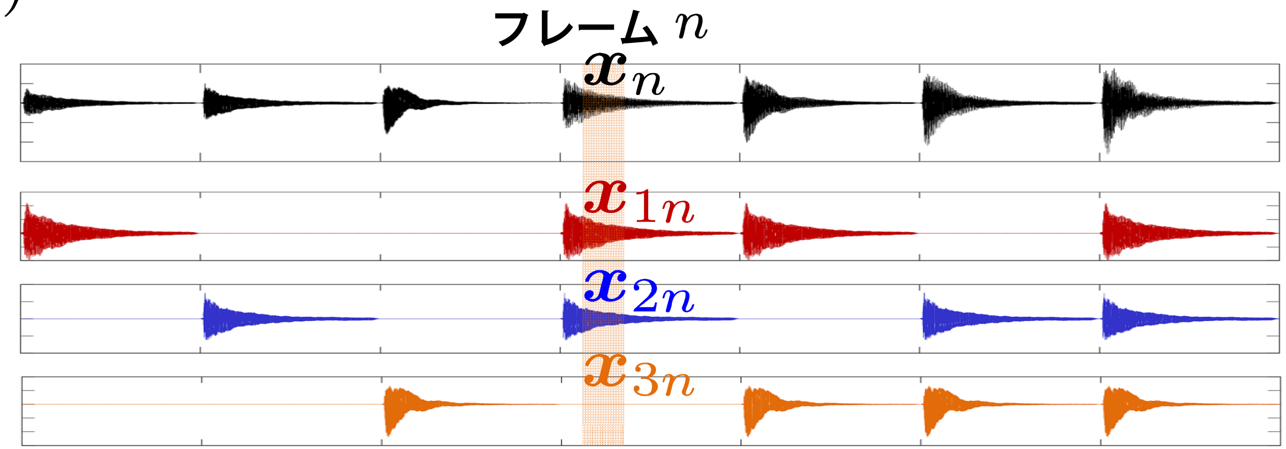


・NMFに基づくすべての手法が改善する可能性！
・他のコスト関数を用いたPSDTFの開発

PSDTFの計算量

- 窓幅 M フレーム数 N 基底数 K → 計算量 $O(M^3NK)$
  - NMFは $O(MNK)$

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



$\boldsymbol{x}_1$

$\boldsymbol{x}_2$

$\boldsymbol{x}_3$

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

時変スケール×カーネル

サイズ M の行列の基本演算に $O(M^3)$

計算量削減：低ランク近似すれば $O(R^2M)$

計算量削減&局所解回避：NMFである程度最適化してからPSDTFに移行

PSDTFの最適化技法

- 補助関数法に基づく乗法更新アルゴリズム
  - もとのパラメータと補助パラメータとを反復更新

Log-Detダイバージェンス

$$\mathcal{D}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = -\log\left|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right| + \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) - M \longrightarrow \text{最小化}$$

任意の半正定値行列 Xに対して

$$\log|\boldsymbol{X}| \leq \log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{X}) - M$$

任意の半正定値行列 X, Yに対して

$$\mathrm{tr}\left(\left(\textstyle\sum_{k=1}^{K}\boldsymbol{X}_k\right)^{-1}\boldsymbol{Y}\right) \leq \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{X}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{Y}\boldsymbol{\Phi}_k^T\right)$$

$$\begin{cases} \boldsymbol{X}_n = \boldsymbol{x}_n\boldsymbol{x}_n^T \\ \boldsymbol{Y}_n = \displaystyle\sum_{k=1}^{K} h_{kn}\boldsymbol{V}_k \end{cases}$$

$$\begin{aligned}\mathcal{D}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) &\overset{c}{=} \log|\boldsymbol{Y}_n| + \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) \\ &\leq \log|\boldsymbol{\Omega}_n| + \mathrm{tr}\left(\boldsymbol{Y}_n\boldsymbol{\Omega}_n^{-1}\right) - M + \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{Y}_{kn}^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right)\end{aligned}$$

$\longrightarrow$ 最小化

PSDTFの無限化

- ガンマ過程 (GaP) を用いると可算無限個の基底数を考慮可能
  - 「半正定値行列」を無限個の「半正定値行列」の凸結合に分解
  - 観測データに合わせて一部の基底のみが使われる

$$\boldsymbol{X}_n \approx \boldsymbol{Y}_n = \sum_{k=1}^{K\to\infty} \theta_k h_{kn} \boldsymbol{V}_k$$

重みパラメータを導入！

無限次元の
非負ベクトル

$$\boldsymbol{\theta} \sim \mathrm{GaP}(\alpha)$$

ガンマ過程事前分布