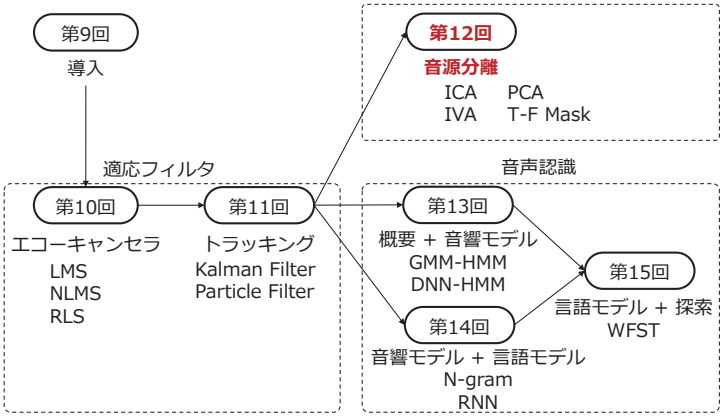
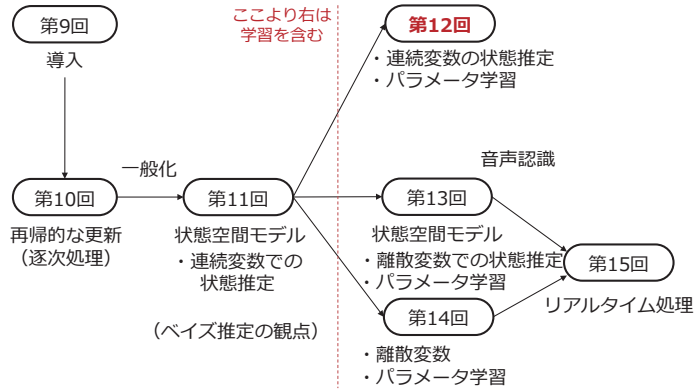
知的情報処理論 第12回

2023年7月4日（火）
武田

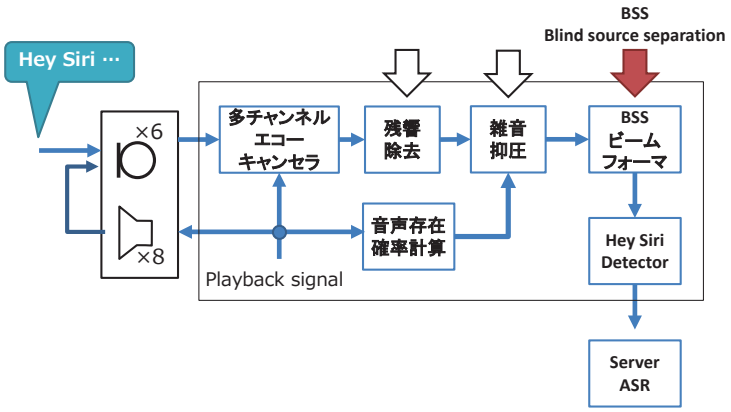
各回の内容（予定）



各回の内容（予定）



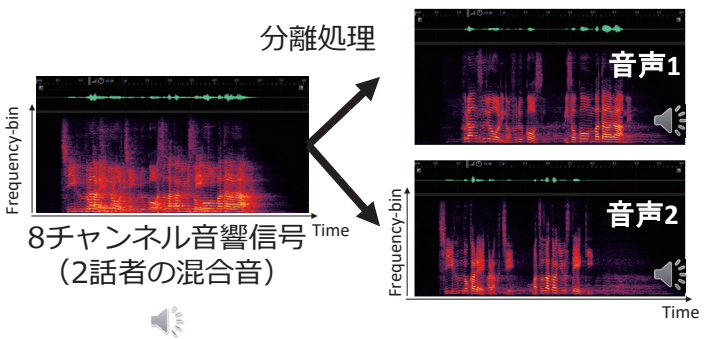
Apple Home Pod の構成 (SLT2018より)



音源分離

混ざった音を分離/特定方向の音を強調

ブラインド音源分離・残響除去



例: 正面方向以外の音を抑圧（マスクベースNN）

頭の片隅に置いとく方が良い大まかな要素

コスト関数
- 「良さ」を測る関数
  - 教師あり学習: 二乗誤差, CrossEntropy, etc…
  - 教師なし学習: 尤度, 再構成誤差, etc…
  - 認識・推定: 期待値, 最大事後確率(MAP), etc…
- ロス/目的関数ともいう

モデル
- 入出力の関係や拘束条件を記述
- 方程式や確率モデルで表現
  （コスト関数と一体化 or 切り離せない場合も）

ざっくりとした認識/学習/探索
- コスト関数を{最大化|最小化|良く}する値（集合）を求める手続き（**最適化等の分野**）
- モデル構造や補助変数・関数を利用した手続きも存在

「初期値依存性」 「局所的最適解（局所解）」 「大域的最適解」

本日の内容: 音源分離

0. 前置き・補足

前回: 状態空間モデル

1. 音源分離の問題設定とモデル

時間軸で独立
非ガウス

2. ブラインド音源分離（線形フィルタ）
   - 独立成分分析

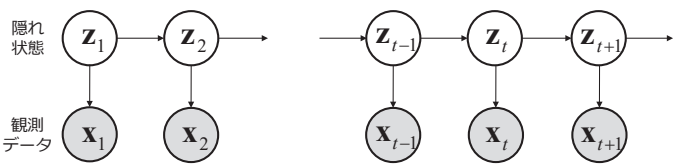
生成モデル的アプローチ
識別モデル的アプローチ

3. モノラル音源分離（非線形フィルタ）

前置き

前回の話題

状態空間モデル: state space model



- **「状態」と「観測」の2つの概念**
- **状態に基づいてデータが生成される過程をモデル化（系列データの確率的生成モデルの一つ）**
- **モデルパラメータが既知の元で，隠れ状態を推定**

（状態空間モデルの）トップダウン的な位置づけ 一般的な問題設定

z — 知りたい情報

⟺ 間をつなぐモデル
Θ: パラメータ

x — 使える情報

使える情報: 入力 x

知りたい情報: 出力 z （**直接観測できない**）
（推定対象のひとつ）

設定の例

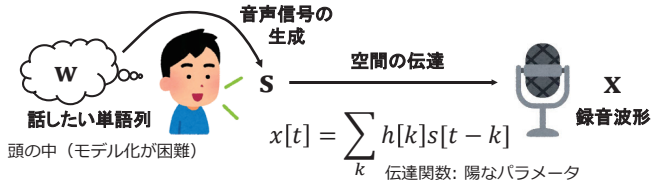
タスク	入力	出力
適応フィルタ	観測信号系列	現在時刻のフィルタ
音声認識	音声特徴量系列	単語列
話者識別	音声特徴量系列	話者ラベル

トップダウン的な位置づけ 整理: 前半で少しあった話

識別モデル（逆モデル）
– 事後確率 $p(\mathbf{z}|\mathbf{x})$ を直接モデル化

(確率的)生成モデル（順モデル）
– ベイズの定理: $p(\mathbf{z}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$

トップダウン的な位置づけ 具体的な例



生成モデル （この例: 音声の生成 と 伝達過程 の組み合わせ）
観測データ $\mathbf{x}$ に合うような $\mathbf{s}$ や $\mathbf{w}$, パラメータ を探す
$p(\mathbf{w},\mathbf{s}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{w},\mathbf{s};\boldsymbol{h})p(\mathbf{s}|\boldsymbol{w})p(\boldsymbol{w})$

識別モデル
観測データ $\mathbf{x}$ から直接 $\mathbf{w}$ を推定/ $\mathbf{w}$ の分布パラメータを推定

トップダウン的な位置づけ 整理: 前半で少しあった話

識別モデル（逆モデル）
– 事後確率 $p(\mathbf{z}|\mathbf{x})$ を直接モデル化
– 正解があれば強力なモデルを作りやすい

回帰 $p(\mathbf{z}|\mathbf{x},\boldsymbol{\Theta}) = \mathcal{N}(\mathbf{z};\mathbf{f}(\mathbf{x};\boldsymbol{\Theta}),\boldsymbol{\Sigma}(\mathbf{x};\boldsymbol{\Theta}))$ 〔$\mathbf{z} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$　$\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\boldsymbol{\Sigma}(\mathbf{x}))$〕
分類 $p(\mathbf{z}|\mathbf{x},\boldsymbol{\Theta}) = \mathrm{softmax}(\mathbf{f}(\mathbf{x};\boldsymbol{\Theta}))$　　e.g. $\mathbf{f}, \boldsymbol{\Sigma}$ をNNでモデル化

生成モデル（順モデル） $p(\mathbf{z}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$
– 条件付き分布 $p(\mathbf{x}|\mathbf{z})$, 事前分布 $p(\mathbf{z})$ をモデル化
– 物理モデル導入，教師なし学習・適応，欠損値対応も可
（初期値依存性あり）

階層化したモデルやハイブリッドモデル、VAEなども存在
解きたい問題や対象の性質を捉えるように設計することが重要

トップダウン的な位置づけ 確率的生成モデル

ポイント: ※ **系列データに限った話ではない**

1. 問題設定の整理: 問題を適切にとらえる
   - 何が given で，何を推定したいのか, 制約は何か
   - Smoothing/filter/prediction，事前学習の有無
2. (データの生成過程を想像し) 潜在変数を設定する（設計者の仮定）
   - 中間的に設定する変数 ：パタンや分布に特徴（偏り）があるもの ：「もし」知っていたらデータを説明しやすいもの
3. 同時分布の形で書いてみる
4. 問題にあう依存関係で分解（設計者の仮定）
   - グラフィカルモデルなど
5. データや問題にあう分布を考える（設計者の仮定）

6. 推定の手続き（推論，学習）を導出
   コスト関数: 尤度，事後確率，それらを用いた何かしらの期待値, etc…

トップダウン的な位置づけ 状態空間モデルの場合

1. 問題設定
   - 観測データ $\mathbf{x}_{1:t}$ から状態 $\mathbf{z}_t$ を推定したい
2. 生成過程
   - $\mathbf{z}_{1:t-1}$ も $\mathbf{z}_t$ の生成に絡むし，推定対象でもある
3. 同時分布 $p(\mathbf{x}_{1:t}, \mathbf{z}_{1:t})$
4. 依存関係による分解: 尤度＋1次マルコフ
   - $p(\mathbf{x}_{1:t}, \mathbf{z}_{1:t}) = p(\mathbf{x}_{1:t}|\mathbf{z}_{1:t})p(\mathbf{z}_{1:t})$
   - $p(\mathbf{x}_{1:t}|\mathbf{z}_{1:t}) = \prod_t p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \prod_t p(\mathbf{x}_t|\mathbf{z}_t)$
   - $p(\mathbf{z}_{1:t}) = p(\mathbf{z}_t|\mathbf{z}_{1:t-1})p(\mathbf{z}_{1:t-1}) = \prod_t p(\mathbf{z}_t|\mathbf{z}_{t-1})$

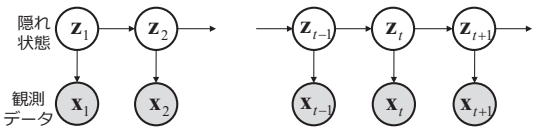
cf. 識別モデル: $p(\mathbf{z}_t|\mathbf{x}_{1:t})$ や $p(\mathbf{z}_{1:t}|\mathbf{x}_{1:t})$ をモデル化

トップダウン的な位置づけ 状態空間モデルの場合

4. 依存関係による分解: 尤度＋1次マルコフ
   - $p(\mathbf{x}_{1:t}|\mathbf{z}_{1:t}) = \prod_t p(\mathbf{x}_t|\mathbf{z}_t)$
   - $p(\mathbf{z}_{1:t}) = p(\mathbf{z}_1)\prod_t p(\mathbf{z}_t|\mathbf{z}_{t-1})$

この時点では
どんな分布を仮定するか決めてない
（抽象度が一段高いモデル）

グラフィカルモデル: 確率変数間の依存関係の図

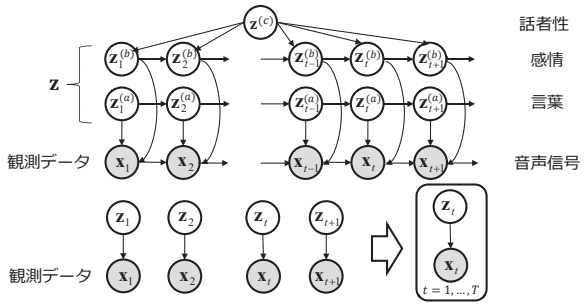


5. データや問題にあう分布を考える
   - 連続状態+線形・ガウス → カルマンフィルタのモデル
   - 離散状態 → (状態遷移は)隠れマルコフモデル

トップダウン的な位置づけ 状態空間モデルの場合

グラフィカルモデル・分布の仮定

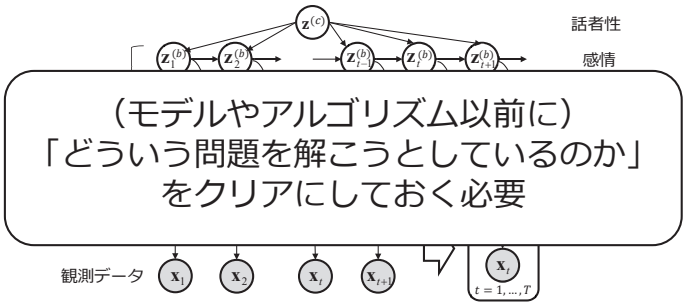
– 典型的・有名なモデルはすでにある
– 問題に応じて選択 or 詳細を設計 or 独自に提案



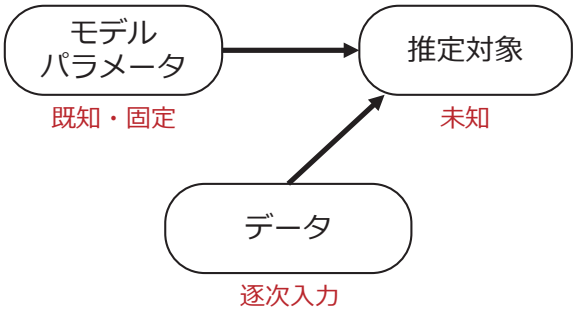
トップダウン的な位置づけ 状態空間モデルの場合

グラフィカルモデル・分布の仮定

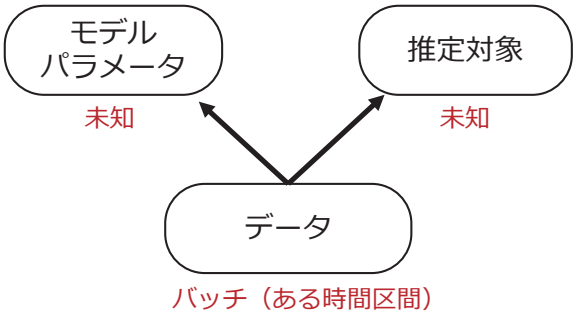
– 典型的・有名なモデルはすでにある
– 問題に応じて選択 or 詳細を設計 or 独自に提案



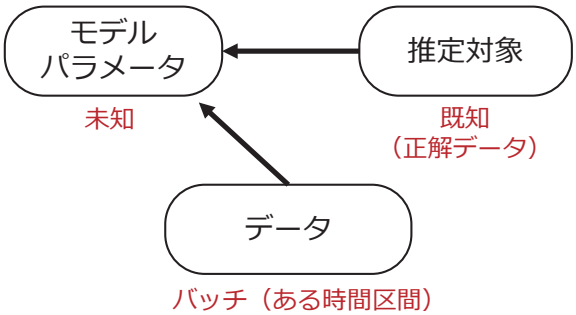
前回の話題 推論/予測/認識（生成モデルベース）



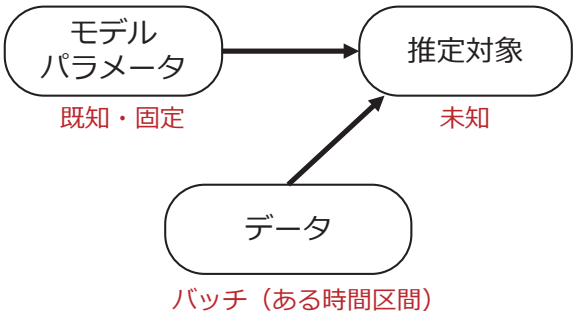
今回の話題: ブラインド音源分離 （生成モデルベース, 教師なし学習）



今回の話題: モノラル分離 教師あり学習時（識別モデルベース）



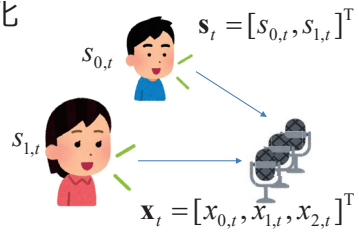
今回の話題: モノラル分離 分離実行時（識別モデルベース）



問題設定とモデル

具体的な例: ブラインド音源分離

目的: 入力信号を個々の音に分離
入力: 多チャンネルマイク入力 $\mathbf{x}_t$
出力: 音源信号 $\mathbf{s}_t$
前提: マイク入力のみで分離
– その他は個別に具体化



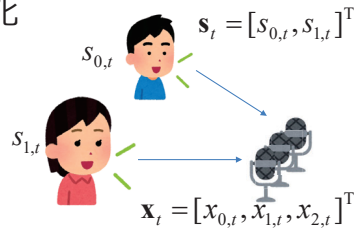
具体的な例: ブラインド音源分離

目的: 入力信号を個々の音に分離
入力: 多チャンネルマイク入力 $\mathbf{x}_t$
出力: 音源信号 $\mathbf{s}_t$
前提: マイク入力のみで分離
– その他は個別に具体化

例えば: 音の伝達過程のモデル

$$\mathbf{x}_t = \sum_d \mathbf{H}_d \mathbf{s}_{t-d} + \mathbf{w}_t$$

畳み込み混合　雑音・誤差



具体的な例: ブラインド音源分離

目的: 入力信号を個々の音に分離
入力: 多チャンネルマイク入力 $\mathbf{x}_t$
出力: 音源信号 $\mathbf{s}_t$
前提: マイク入力のみで分離
– その他は個別に具体化

例えば: 音の伝達過程のモデル

$$\mathbf{x}_t = \sum_d \mathbf{H}_d \mathbf{s}_{t-d} + \mathbf{w}_t$$

畳み込み混合　雑音・誤差

$\mathbf{H}_d$　$\mathbf{s}_t$
混合係数(時不変)とソースを観測データから推定
基本は"音源方向"で分離
音の到来時間差

よく使われるモデル: 短時間周波数領域 (STFT: short-time Fourier transform) でのモデル化

STFT領域 → 畳み込みを積で近似
– 膨大なタップ係数 → 各周波数ビン毎に積
– 時間領域モデルに比べて実用的

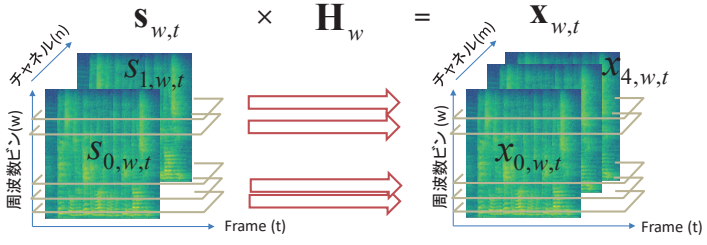
$$\mathbf{x}_t = \sum_d \mathbf{H}_d \mathbf{s}_{t-d} + \mathbf{w}_t \quad \Leftrightarrow \quad \mathbf{x}_{w,t} = \mathbf{H}_w \mathbf{s}_{w,t} + \mathbf{w}_t$$

t: 時間インデックス

t: フレームインデックス
w: 周波数ビンインデックス
変数はすべて複素数

よく使われるモデル: 短時間周波数領域 (STFT: short-time Fourier transform) でのモデル化

STFT領域 → 畳み込みを積で近似
– 膨大なタップ係数 → 各周波数ビン毎に積
– 時間領域モデルに比べて実用的
• STFT 領域での畳み込みモデルなどもある



2 つのモデル

音源モデル $\mathbf{s}_{w,t}$
- 音源パターンや「その音らしさ」を表現
- 方程式や確率モデルで拘束条件を与える
  e.g., 優ガウス分布, local Gaussian, NMF, 深層学習, etc..

空間モデル: 音源・マイク間の伝達関数 $\mathbf{H}_w$
もしくは, 分離フィルタ $\mathbf{W}_w$
- 伝達過程（の複雑度）を表現
- パラメータ扱い/ 確率モデルで拘束条件を与える

ブラインド分離では制約を適切に与えないと
【元の信号】を上手く抽出できない
そうでない場合，何らかの信号は出てくるが，真の信号とは限らない

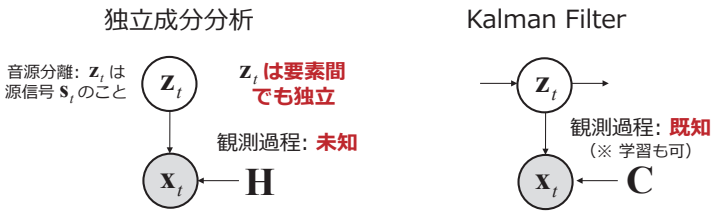
基本的な問題設定と色んなアプローチ

多チャンネル分離: 線形フィルタ
- ビームフォーマ: 音源方向が既知の下で推定
- ブラインド分離: 音源と空間モデルの両方を混合信号から（適当な区間毎に）推定
  ☺ マイクや音源位置の情報が不要
  ☹ ある程度のデータ量や反復推定が必要なことが多い

モノラル分離: 非線形フィルタ
- 混合音からターゲット成分を(NNで)直接予測
  ターゲットの音源種は仮定: 音声など
  ☹ 事前に大量のデータからモデル（分類/回帰）を構築
  ☺ 推論時は forward 計算だけで済む

独立成分分析とカルマンフィルタの仮定の対比（グラフィカルモデル）

時刻 t に関して潜在変数 $\mathbf{z}_t$ は独立



得られた観測値の系列 $\mathbf{x}_{1:T} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T]$ **から**
状態 $\mathbf{z}_{1:T}$ **と** $\mathbf{H}$ **(or 分離行列** $\mathbf{W}$**) を推定する問題**
後述
今回はバッチ処理

注意: 別の表現

ベクトルを時間軸でまとめて行列表現に

$$\mathbf{x}_t = \mathbf{H}\mathbf{s}_t \Rightarrow [\mathbf{x}_1 \quad \dots \quad \mathbf{x}_T] = \mathbf{H}[\mathbf{s}_1 \quad \dots \quad \mathbf{s}_T]$$

$$\mathbf{X} = \mathbf{H}\mathbf{S} \qquad \boxed{\mathbf{X}} = \boxed{\mathbf{H}} \times \boxed{\mathbf{S}}$$

- 行列分解の特殊な形ともとれる $\|\mathbf{X} - \mathbf{H}\mathbf{S}\|$
- 各問題における**対象の特性**や**モデルの仮定・制約条件, パラメータ推定方法**に注意

e.g. 変数間の具体的な関係, サンプル間の独立性, 次元数，各変数の特性: 非負値・離散，変数の事前分布，etc…

注意: 別の表現

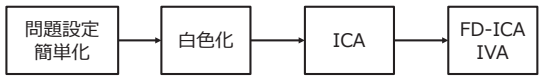
計画行列表記: 最小二乗法でも使われる

$$\mathbf{X} = [\mathbf{x}_1 \quad \dots \quad \mathbf{x}_T]^{\mathrm{T}}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = [\mathbf{x}_1 \quad \dots \quad \mathbf{x}_T][\mathbf{x}_1 \quad \dots \quad \mathbf{x}_T]^{\mathrm{T}}$$

$$= [\mathbf{x}_1 \quad \dots \quad \mathbf{x}_T] \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_T^{\mathrm{T}} \end{bmatrix}$$

$$= \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\mathrm{T}}$$



ブラインド音源分離

問題の簡単化: 基本モデルを扱う

STFT 領域: 畳み込み → 積

– 時間遅れのない瞬時混合モデルに絞る

$$\mathbf{x}_{w,t} = \mathbf{H}_w \mathbf{s}_{w,t} + \mathbf{w}_t \qquad \mathbf{H}_w \text{ 時間変化しない}$$

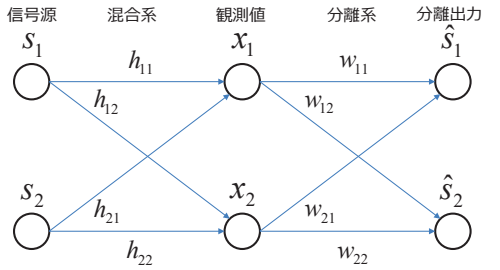
– 表記: 周波数ビンindexや雑音もとりあえず無視
時間表記も消去（時刻間で独立に生成されると仮定）

$$\mathbf{x} = \mathbf{H}\mathbf{s}$$

– マイク数と音源数も同じ, 変数は実数値と仮定
$\mathbf{x}$ と $\mathbf{s}$ は同じ次元数 N

– $\mathbf{s}$ の平均 0 を仮定 → $\mathbf{x}$ も平均 0

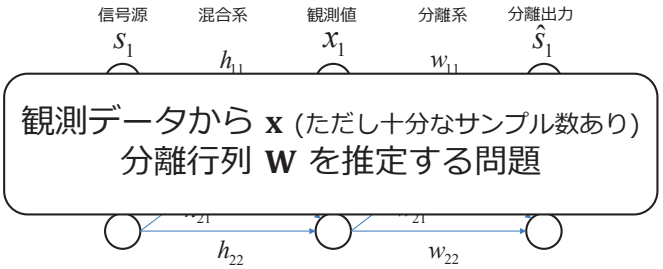
問題の簡単化



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \qquad \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\mathbf{x} = \mathbf{H}\mathbf{s}$ 混合行列　　$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ 分離行列

問題の簡単化



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \qquad \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\mathbf{x} = \mathbf{H}\mathbf{s}$ 混合行列　　$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ 分離行列

なお，伝達関数が既知の場合

状況としては・・

a) 音源（発生源）の座標が既知

b) 音源-マイク間のインパルス応答が既知
→ $\mathbf{H}$ がわかっている

– $\mathbf{H}$ の逆行列や疑似逆行列で

$$\mathbf{x} = \mathbf{H}\mathbf{s} \quad \Rightarrow \quad \hat{\mathbf{s}} = \mathbf{H}^{-1}\mathbf{x}$$

伝達関数既知は現実的にはなかなかつらい設定

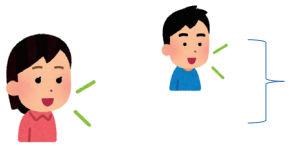
☹ 場所によっても H は異なる上，残響も存在=直接音の到来時間差のみ既知ではつらい
☹ そもそもマイク自体の周波数特性のキャリブレーションも必要 = 手間

直観的な方法

伝達関数の情報は使わない

– 一方，データ点はある程度数があると想定

– 統計的な性質の活用 →
もともとの音源信号は互いに無関係のはず

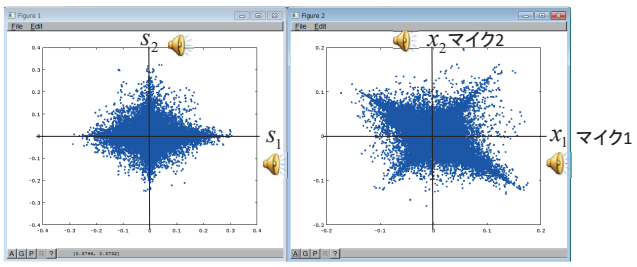


互いに独立な情報源？

出力が互いに無相関になるように $\mathbf{W}$ を決める？

実際の音声信号の同時分布（散布図のプロット）

混合前 $\mathbf{s}$　　混合後 $\mathbf{x}$



無相関ではありそう

相関を持っている
（マイク間で似た成分がある）

無相関化・白色化

白色

– N 個のデータから構成されるベクトル

$$\mathbf{z} = [z_1, ..., z_N]^{\mathrm{T}}$$

– 要素が互いに無相関かつ相互相関関数（行列）が次式を満たす

$$E[\mathbf{z}\mathbf{z}^{\mathrm{T}}] = \mathbf{I} \qquad \mathbf{I} \text{ 単位行列}$$

このような状態にすることを無相関化，白色化，あるいは球面化と呼ぶ

白色化

白色化を行うフィルタ $\mathbf{V}$

– 線形の過程

$$\mathbf{z} = \mathbf{V}\mathbf{x}$$

– 白色となる条件

$$E[\mathbf{z}\mathbf{z}^{\mathrm{T}}] = \mathbf{V}E[\mathbf{x}\mathbf{x}^{\mathrm{T}}]\mathbf{V}^{\mathrm{T}} = \mathbf{V}\mathbf{R}_z\mathbf{V}^{\mathrm{T}} = \mathbf{I}$$

$$\mathbf{R}_z = E[\mathbf{x}\mathbf{x}^{\mathrm{T}}] \qquad \text{※ 実際はサンプル平均で計算}$$

これを満たす何かしらの $\mathbf{V}$ は構築可能

白色化

白色化を行うフィルタ $\mathbf{V}$

– $\mathbf{R}_z$ の固有ベクトル $\{\mathbf{e}_i\}$ と固有値 $\{\lambda_i\}$ を用いて下記の行列を定義

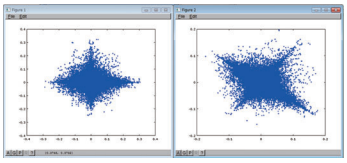
$$\mathbf{E} = [\mathbf{e}_1, ..., \mathbf{e}_N] \qquad \mathbf{\Lambda}^{-1/2} = \mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}}, ..., \frac{1}{\sqrt{\lambda_N}}\right)$$

N: $\mathbf{x}$ の次元数　　diag は要素を対角に並べた行列

– $\mathbf{V} = \mathbf{\Lambda}^{-1/2}\mathbf{E}$ は $\mathbf{x}$ を白色化する

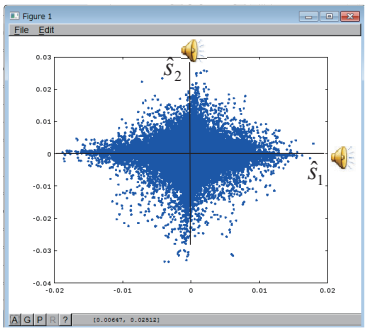
– $\mathbf{E}$ は無相関化，$\mathbf{\Lambda}^{-1/2}$ は分散を正規化

白色化フィルタをかけると



混合前　　混合後

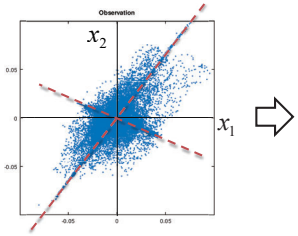
白色化後

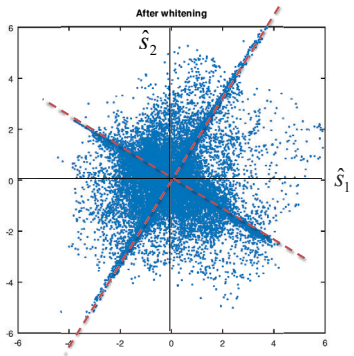


分離でき…ちゃった……？

先の例は偶然

無相関だが分離できてない例





ただし，軸のようなものは
直角に補正されたように見える

白色化行列

回転に関する不定性

– 任意のユニタリ行列 $\mathbf{U}$　　$\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$

– 白色化フィルタ $\mathbf{V}$ に縦続接続

$$\mathbf{y} = \mathbf{U}\mathbf{V}\mathbf{x}$$

$$E[\mathbf{y}\mathbf{y}^{\mathrm{T}}] = \mathbf{U}\mathbf{V}E[\mathbf{x}\mathbf{x}^{\mathrm{T}}]\mathbf{V}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}$$

$$= \mathbf{U}\mathbf{U}^{\mathrm{T}}$$

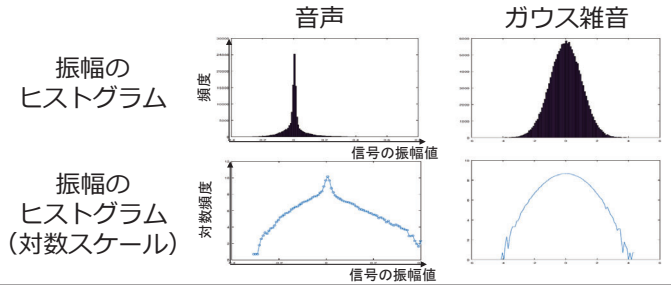
$$= \mathbf{I} \qquad \mathbf{U}\mathbf{V} \text{ も白色化行列}$$

白色化の条件では分離可能なフィルタは定まらない

音源の統計的性質をよく見る

非ガウス分布かつ互いに独立

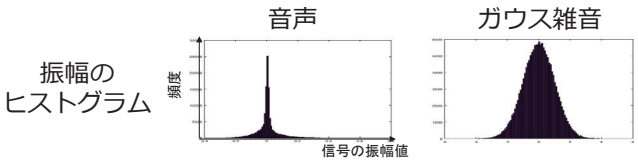
– 音声，楽器，大体の音信号: ガウス分布ではない
– 独立: 無相関より強い条件



音源の統計的性質をよく見る

非ガウス分布かつ互いに独立

– 音声，楽器，大体の音信号
  • ガウス分布ではない．音声などは定常でもない．
– 独立: 無相関より強い条件

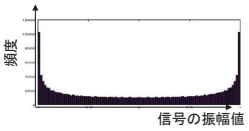


このような性質（非ガウス性 や 非定常性）を
キチンと捉える/利用しないと期待通りの分離はできない

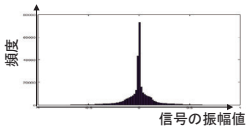
Mini Quiz #1

• どの信号がどの振幅ヒストグラムに対応するだろうか？

スズムシの鳴き声　A ○　a ○

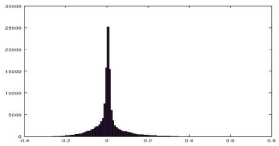


$x_n = \sin(2\pi \frac{5n}{16000})$　B ○　b ○



考察してみましょう

<u>0値付近の出現頻度が高い分布</u>のとき…



いわゆる “疎（スパース）”な分布

人の音声や生物の鳴き声などの音信号

1. ガウス分布と比較して，(独立な)複数の音信号が同時に0(付近)以外の値を取る確率(可能性)はどうなりそうか?
2. 分布を仮定しない場合と比較して，次のような非0値の関係を満たすような，組み合わせパタンそれぞれの<u>可能性(生起確率)</u>はどうなりそうか?

$5 = ax_1 + by_1$　$6 = cx_1 + dy_1$
$7 = ax_2 + by_2$　$8 = cx_2 + dy_2$
$-10 = ax_3 + by_3$　$-9 = cx_3 + dy_3$

a, b, c, d は定数
各 x_i, y_i は上の分布に従う
（さらに平均0, 分散1であるとする）

考察してみましょう

<u>白色化について</u>

– エコキャンのフィルタをLMSで推定する際，もし仮に参照信号 $\mathbf{u}_t$ を事前に白色化できると，何かうれしくなるだろうか？
  • 参照信号が元から白色だと意味なし

$$d_t = \boldsymbol{w}^{\mathrm{T}}\mathbf{u}_t \quad \Rightarrow \quad \overline{d_t} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v}_t$$

– 白色化した後の参照信号で推定したフィルタと，白色化しないで推定したフィルタの間に何か違いがあるだろうか

独立成分分析

独立成分分析
(ICA; Independent Component Analysis)

信号の統計的独立性・非ガウス性を利用

– 独立性: $p(x,y) = p(x)p(y)$
  • KL-divergence で測る, 生成モデルでの仮定
– 非ガウス性/スパース性
  1. 具体的な分布（信号が従うもの）を利用
  2. エントロピー,ネゲントロピー,尖度を利用: fastICA

⇩ そのこころ

混合後の信号はガウス分布に近づく(**中心極限定理**)
分離: 逆の過程=出力 $\hat{\mathbf{s}}$ をガウス分布から遠ざける

これらに基づいて分離行列 $\mathbf{W}$ を推定($\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$)

注意点: ICA の曖昧性

「混合行列と信号の両方が未知」で生じる

$$\mathbf{x} = \mathbf{H}\mathbf{s} = \sum_i \mathbf{a}_i s_i$$

$\mathbf{a}_i$は$\mathbf{H}$の i 番目の列ベクトル

1. 独立成分の分散（パワー）を決定することはできない: スケールの不定性（掛け算があるから）

$$\mathbf{x} = \sum_i \left(\frac{1}{\alpha_i}\mathbf{a}_i\right)\alpha_i s_i$$

結果的に
出力信号の分散が 1 と
仮定しても良い

2. 独立成分の出力順序を決めることはできない
   FD-ICA ではパーミュテーション問題となる

さっき出てきた白色化の役割

白色化: 前処理として有効

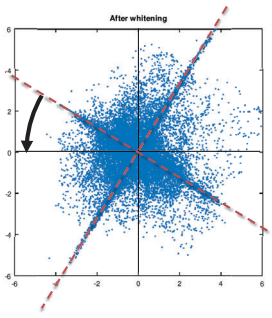
– 意味がないわけではない
– フィルタ学習における収束速度改善も期待

問題としては「半分解けている」

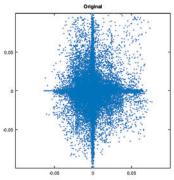
– 出力の分散は 1 に拘束
– 後はユニタリ行列（行ベクトルはノルム1）を求めるだけ → 解の候補に制約が掛かっている
– 高速に収束するアルゴリズム(fastICAなど)は白色化を前提

白色化の役割

白色化: 前処理として有効



実際，後は軸を
上手く回転させれば
分離は達成される
（元の散布図になる）
回転＝ユニタリ行列



KL-divergence による定式化

KL-divergence: 分布間の距離尺度の一つ

$$D_{KL}(p,q) = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \mathrm{d}\mathbf{x}$$

– $D_{KL} \geq 0$ かつ 等号は $p(\mathbf{x}) = q(\mathbf{x})$ の時のみ成立
– p, q に関しては非対称

分離出力の独立性を KLD で測れば，独立な成分になっているかどうかわかる

KLD を用いた分離行列推定をざっと見る

KL-divergence による定式化モデル

KLD に基づく独立性尺度

– 分離出力 $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$
– 同時分布 $p(\hat{\mathbf{s}}) = p(\hat{s}_1, \ldots, \hat{s}_N)$
– 周辺分布の積 $\prod_i p_i(\hat{s}_i)$

として，KLD に当てはめる

方程式とパラメータ $\mathbf{W}$

確率モデル（独立）

$$I(\hat{\mathbf{s}}) = \int p(\hat{\mathbf{s}}) \log \frac{p(\hat{\mathbf{s}})}{\prod_i p_i(\hat{s}_i)} \mathrm{d}\hat{\mathbf{s}}$$

コスト関数

これを最小化するように$\mathbf{W}$ を推定する

KL-divergence による定式化 コスト関数

– 変形するとエントロピー H の形で書ける

$$I(\hat{\mathbf{s}}) = \int p(\hat{\mathbf{s}}) \log \frac{p(\hat{\mathbf{s}})}{\prod_i p_i(\hat{s}_i)} \mathrm{d}\hat{\mathbf{s}}$$

$$= \int p(\hat{\mathbf{s}}) \log p(\hat{\mathbf{s}}) \mathrm{d}\hat{\mathbf{s}} - \sum_i \int p(\hat{s}_i) \log p_i(\hat{s}_i) \mathrm{d}\hat{s}_i$$

「$p(\hat{\mathbf{s}})$ が消えた？」と思った人は一回展開して積分を考えてみよ

$$= -H(\hat{\mathbf{s}}) + \sum_i H(\hat{s}_i)$$

こっちはなんとかなる

$$H(\hat{s}_i) = -\int p(\hat{s}_i) \log p(\hat{s}_i) \mathrm{d}\hat{s}_i = -E[\log p_i(\hat{s}_i)]$$

$$H(\hat{\mathbf{s}}) = -\int p(\hat{\mathbf{s}}) \log p(\hat{\mathbf{s}}) \mathrm{d}\hat{\mathbf{s}} = -E[\log p(\hat{\mathbf{s}})]$$

これを扱わないといけない

KL-divergence による定式化 コスト関数

ここで，確率変数 $\mathbf{x}$ と $\mathbf{y}$ が可逆な線形変換 $\mathbf{y} = \mathbf{A}\mathbf{x}$ の関係にあるとき，下記が成立

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{|\det \mathbf{A}|} p_{\mathbf{x}}(\mathbf{A}^{-1}\mathbf{y})$$

伸縮・膨張分をキャンセル（密度は非負値，積分値＝1）

これを $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ に対して用いると

$$H(\hat{\mathbf{s}}) = -E[\log p_{\hat{\mathbf{s}}}(\hat{\mathbf{s}})] = -E[\log \frac{1}{|\det \mathbf{W}|} p_{\mathbf{x}}(\mathbf{x})]$$

$$= -E[\log p_{\mathbf{x}}(\mathbf{x})] + \log |\det \mathbf{W}|$$

$$= H(\mathbf{x}) + \log |\det \mathbf{W}|$$

KL-divergence による定式化 最適化（学習）

下記を最小化する $\mathbf{W}$ を勾配法で求める

$\hat{s}_i = \mathbf{w}_i^{\mathrm{T}}\mathbf{x}$　　$\mathbf{w}_i^{\mathrm{T}}$は $\mathbf{W}$ の i 番目の行ベクトル

$$I(\hat{\mathbf{s}}) = -H(\mathbf{x}) - \log |\det \mathbf{W}| - \sum_{i=1}^{N} E[\log p_i(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})]$$

分離行列と無関係（定数扱い）　白色化後だと定数

– **（一応）勾配を計算: 第2項目**

$$\frac{\partial}{\partial \mathbf{W}} \log |\det \mathbf{W}| = (\mathbf{W}^{-1})^{\mathrm{T}}$$ （余因子行列表現を利用）

$$\left(= \mathbf{W}^{-\mathrm{T}} \right)$$

KL-divergence による定式化 最適化（学習）

– **勾配を計算: 第3項目**

$$\frac{\partial}{\partial \mathbf{w}_i} \sum_{i=1}^{N} E[\log p_i(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})] = \frac{\partial}{\partial \mathbf{w}_i} E[\log p_i(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})]$$

$$= E[\frac{\partial}{\partial \hat{s}_i} \log p_i(\hat{s}_i) \frac{\partial}{\partial \mathbf{w}_i} \mathbf{w}_i^{\mathrm{T}}\mathbf{x}]$$

$$= E[\mathbf{x}\varphi_i(\hat{s}_i)] \quad \varphi_i(\hat{s}_i) = -\frac{\partial}{\partial \hat{s}_i} \log p_i(\hat{s}_i)$$ という

スコア関数を定義（具体的なものは後述）

行列形式で整理 $\frac{\partial}{\partial \mathbf{W}} = \left[\frac{\partial}{\partial \mathbf{w}_1} \quad \cdots \quad \frac{\partial}{\partial \mathbf{w}_N} \right]^{\mathrm{T}}$

$$\frac{\partial}{\partial \mathbf{W}} \sum_{i=1}^{N} E[\log p_i(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})] = E[\varphi(\hat{\mathbf{s}})\mathbf{x}^{\mathrm{T}}]$$

$$\varphi(\hat{\mathbf{s}}) = [\varphi_1(\hat{s}_1), \ldots, \varphi_N(\hat{s}_N)]^{\mathrm{T}}$$

KL-divergence による定式化 最適化（学習）

勾配法まとめ

$$\frac{\partial}{\partial \mathbf{W}} I(\hat{\mathbf{s}}) = -\mathbf{W}^{-T} + E[\varphi(\hat{\mathbf{s}})\mathbf{x}^{\mathrm{T}}]$$

Amari ら(1996) による自然勾配を用いると

$$\frac{\partial}{\partial \mathbf{W}} I(\hat{\mathbf{s}})\mathbf{W}^{\mathrm{T}}\mathbf{W} = -\mathbf{W} + E[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{\mathrm{T}}]\mathbf{W}$$

逆行列が消える

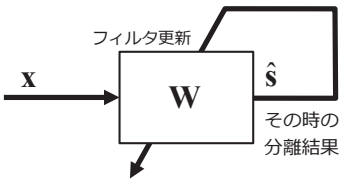
$$= -(\mathbf{I} - E[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{\mathrm{T}}])\mathbf{W}$$

中が 0 になると停止

更新式　$\mathbf{W} \leftarrow \mathbf{W} + \mu(\mathbf{I} - E[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{\mathrm{T}}])\mathbf{W}$

期待値はサンプル平均に置き換え

勾配法による手続き



誤差のフィードバック
$(\mathbf{I} - E[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{\mathrm{T}}])\mathbf{W}$

$E[\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{\mathrm{T}}]$　非線形な相関

正解の"値"との誤差計算がない＝教師なし

– 停止条件: 例 - スコア関数が線形の場合(ガウス分布)

$\mathbf{I} - \mathbb{E}[\hat{\mathbf{s}}\hat{\mathbf{s}}^{\mathrm{T}}] = \mathbf{0}$ → $\mathbb{E}[\hat{\mathbf{s}}\hat{\mathbf{s}}^{\mathrm{T}}] = \mathbf{I}$：無相関になった時

→ モデルが仮定する統計的性質を満たすと停止

☹ 実データとモデルや仮定の間に乖離があると悪影響: 過剰フィット

適当な拘束条件も重要: ICA 線形 (混合音 – 分離音)

– 初期値依存性: もちろんある

スコア関数

定義

$$\varphi(x) = -\frac{\partial}{\partial x}\log p(x)$$

源信号の確率密度から計算
- 必要に応じて，分散1に正規化したもの

確率密度が既知: 定義に従って計算

確率密度が未知: 近似したものを適用
- 大雑把でも動作するときはする

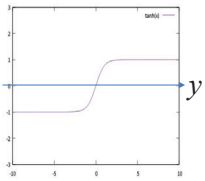
変数が複素数の場合（※ コスト関数は実数）
- 実部虚部をわける or ウィルティンガーの微分で計算

スコア関数（実数版）

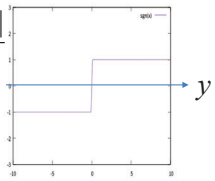
	$p(y)$	$\varphi(y)$
双曲線余弦	$\frac{1}{\pi\cosh(y/\sigma^2)}$	$\tanh\frac{y}{\sigma^2}$
ラプラス	$\frac{1}{2\sigma}\exp\left(-\frac{\lvert y\rvert}{\sigma}\right)$	$\frac{1}{\sigma}\frac{y}{\lvert y\rvert}$

$\varphi(y)$

$\tanh y$



$\frac{\lvert y\rvert}{y}$



参考: 最尤推定

生成モデルの観点

$$p(\mathbf{x}\mid\mathbf{W}) = \lvert\det\mathbf{W}\rvert\prod_i p(\hat{s}_i)$$

確率モデル & コスト関数

$$= \lvert\det\mathbf{W}\rvert\prod_i p(\mathbf{w}_i^{\mathrm{T}}\mathbf{x})$$

$\mathbf{w}_i^{\mathrm{T}}$ は $\mathbf{W}$ の i 番目の行ベクトル

– 全データに対する尤度の負の対数を取る
KLDと同じ形の式が出現 → 同様の計算

$$-\log p(\mathbf{x}_{1:T}\mid\mathbf{W}) = -\log\prod_t p(\mathbf{x}_t\mid\mathbf{W})$$

サンプル間は独立

$$= -\sum_t \log p(\mathbf{x}_t\mid\mathbf{W})$$

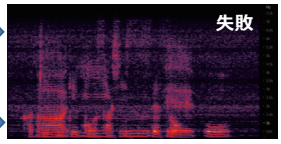
実用上の問題

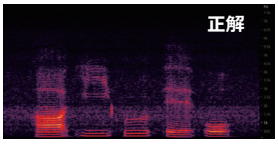
STFT 領域でのICAの適用: FD-ICA

– 周波数ビン毎に独立に 複素 ICA を適用
→ 分離自体は上手く動作

– 不定性の問題が致命的　復元できない
- スケールの不定性 → 分散1 のままだと白色信号
- パーミュテーションの不定性 → 帯域毎に音源成分が混ざってしまう

音源成分が入り乱れる





スケールの復元

Projection back

– 分離行列の逆行列は混合行列の近似

$$\hat{\mathbf{A}} = \mathbf{W}^{-1} \qquad \mathbf{x} = \sum_i \hat{\mathbf{a}}_i\hat{s}_i$$

– 不定性がキャンセルされる下記の成分を出力

$\hat{\mathbf{a}}_i\hat{s}_i$ ：“マイクで観測された信号”に復元

$\hat{s}_i$

ICAの出力

$\hat{\mathbf{a}}_i\hat{s}_i$

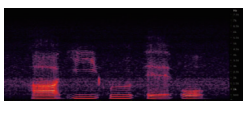
伝達関数がかかったもの

Projection back

パーミュテーションへの対応

クラスタリング

– パワースペクトルパターンを利用
- 音声の調波構造：シマシマパターン
- スペクトルの包絡
- 基本周波数



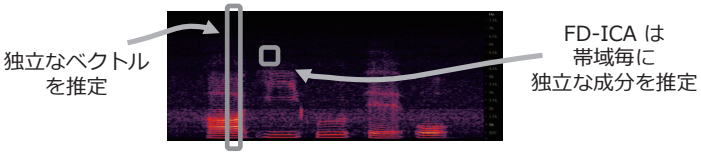
– 音の到来方向を利用
- 推定した伝達関数（ステアリングベクトル）$\hat{\mathbf{a}}_i$
- 観測スペクトル

などから方向情報（位相差，到来時間差）を抽出
= マイクアレーの特性や物理的な情報を利用

IVA (Independent Vector Analysis)

独立ベクトル分析 (Kim, Hiroe 2006)

- 独立な ”ベクトル成分” を抽出
- FD-ICA のパーミュテーション問題を ICA の枠組みを拡張して対応
  → “独立なフレームベクトルを出力”する



周波数軸上: パワースペクト上では関連づいている

IVA (Independent Vector Analysis)

混合過程・分離過程は同じ: 帯域毎

違い: 確率密度 $p(\mathbf{s}_{i,*,t})$・スコア関数

- 「ベクトル」単位で定義される
- 音源 i , フレーム t のベクトル

$$\mathbf{s}_{i,*,t} = [s_{i,1,t}, \ldots, s_{i,W,t}]^{\mathrm{T}}$$

周波数ビン $1,\ldots,W$

$s_{i,w,t}$ 音源 i, フレームt, 周波数ビン w の成分（複素数）

- 密度関数の例

$$p(\mathbf{s}_{i,*,t}) = \alpha \exp\left(-\sqrt{(\mathbf{s}_{i,*,t}-\boldsymbol{\mu}_i)^{\mathrm{H}} \boldsymbol{\Sigma}_t^{-1} (\mathbf{s}_{i,*,t}-\boldsymbol{\mu}_i)}\right)$$

IVA (Independent Vector Analysis)

違い: 確率密度 $p(\mathbf{s}_{i,*,t})$・スコア関数

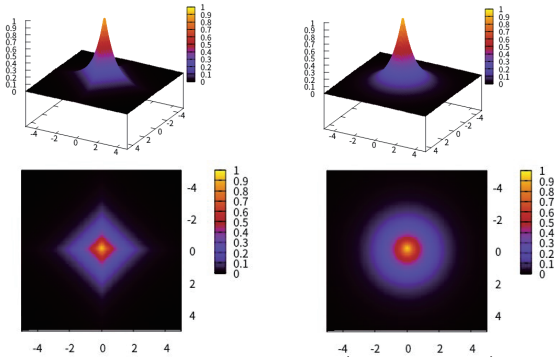
- $\boldsymbol{\Sigma}_t^{-1} = \mathbf{I}$ でも帯域間で独立にはならない
- 最も単純な形（$\boldsymbol{\mu}_i = \mathbf{0}$ ）とすると

$$p(\mathbf{s}_{i,*,t}) = \alpha \exp\left(-\sqrt{\sum_w |s_{i,w,t}|^2}\right)$$

$$-\frac{\partial}{\partial s_{i,w,t}} \log p(\mathbf{s}_{i,*,t}) = \frac{1}{2} \frac{s_{i,w,t}}{\sqrt{\sum_w |s_{i,w,t}|^2}} \quad \Leftrightarrow \quad \text{ICA} \quad \frac{1}{2} \frac{s_{i,w,t}}{|s_{i,w,t}|}$$

周波数軸上での正規化が
音源の「活性化(activation)」をまとめるような効果
(lasso → group lasso のようなもの)

密度関数の”形状” 比較



ラプラス分布 $\exp(-|x|-|y|)$
$= \exp(-|x|)\exp(-|y|)$

$\exp\left(-\sqrt{|x|^2+|y|^2}\right)$

その他のモデル

音源モデル

- 確率モデル: 時変分散ガウス, KF, HMM, …
- 音源パワースペクトル（時変分散） = 周波数 x フレーム → 非負値行列分解系 (NMF)
- 観測スペクトル = チャネル x フレーム x 周波数 → テンソル分解系, multichannel NMF

  → NNs を用いた教師あり/なし事前学習へシフト…

空間モデル: 空間相関行列の制約 $\mathbf{a}_i\mathbf{a}_i^{\mathrm{H}}$

- rank-1: 伝達関数が時不変の場合＝瞬時混合
- full-rank: 瞬時混合とみなせない場合のモデル

<u>いずれも方程式や確率モデルの形で制約を与える</u>

(時間があれば)Mini quiz #2

ICA ではデータの時間方向の構造
（変化パターン）は何も仮定していない

このメリットとデメリットに関して何か
あったりするか考えてみましょう

- 適用可能な信号
- 時間方向構造/パタン
  - 事前学習？ → 使うデータは？
  - その場でブラインド推定？
    - 合唱音声だと音声パターンは?

補足: 主成分分析

データについて

多くのデータ集合

- **観測データの次元数 >> 実効次元数**
- 例: 手書き数字データの1つから人工的に生成
  - もと: 64x64 画素のグレースケール画像
  - 位置と方向をランダムに変え 100x100画素の画像に配置: 100x100=10,000次元のデータ空間1点

https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/ より

- 実効次元数 = 3　　潜在変数に相当
  - 実際の自由度: 平行移動(垂直・水平) + 回転の3つ

主成分分析, PCA
Principal Component Analysis

用途

- 次元削減, 非可逆データ圧縮, 特徴抽出, データの可視化などに広く適用
  - 限られた次元数の特徴量で元のデータを再現
- Karhunen-Loève 変換と同じ

PCA を定義する一般的な方法

低次元の線形空間の上へのデータ点の直交射影

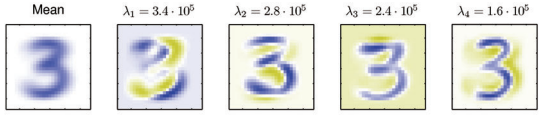
- 分散最大化
- 誤差最小化

} 等価（同じアルゴリズムを与える）

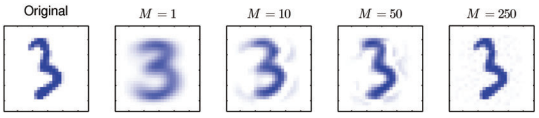
固有値問題などと関係（行列分解系）

ずらし数字データ集合での例

- 「3」の平均画像＋最初の4つの主成分分析の固有ベクトル・固有値



- 原画像と M 個の主成分を使って得られた主成分分析による再現画像



https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/ より

確率的主成分分析など

ICA: 生成モデルの側面あり

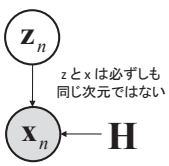
- 信号源: 非ガウス分布

確率的主成分分析

- 生成モデル（ベイズ的に扱う場合）

$$\mathbf{x}_n = \mathbf{H}\mathbf{z}_n + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_n \qquad \mathbf{z}_n \sim N(0, \mathbf{I}) \quad \boldsymbol{\varepsilon}_n \sim N(0, \sigma^2 \mathbf{I})$$

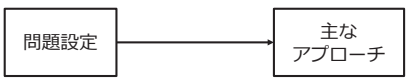
同じグラフィカルモデル



- 信号源がガウス分布（線形-ガウス）
  - 固有値分解など以外に EMアルゴリズムなどが適用可

他: ベイズ主成分分析

- $\mathbf{H}$ を確率変数化し, モデル（M の次元数）選択
- 事前分布&積分消去 $p(\mathbf{H}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \left(\frac{\alpha_i}{2\pi}\right)^{D/2} \exp\left(-\frac{1}{2}\alpha_i \mathbf{h}_i^{\mathrm{T}} \mathbf{h}_i\right)$



モノラル音源分離

モノラル信号（single channel）で音源分離

1. 使えるリソースによる制限
   – 録音データがそれしかない
2. 同じ方向に音源が存在
   – マイクアレーでは分離が困難

問題設定の例

入力: 音声信号 + 背景雑音
出力: 音声信号
音声かそれ以外かを区別する問題

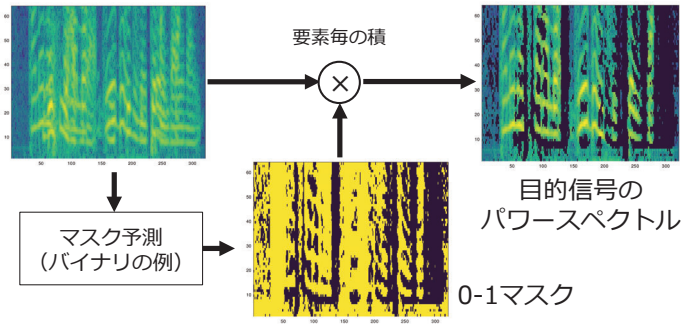
入力: 複数楽器の混合信号
出力: 楽器毎の信号
異なる性質の信号をそれぞれ分離する問題

入力: 複数の音声信号
出力: それぞれの音声信号
同種の信号をそれぞれ分離する問題

各問題設定で前提・仮定は通常異なる

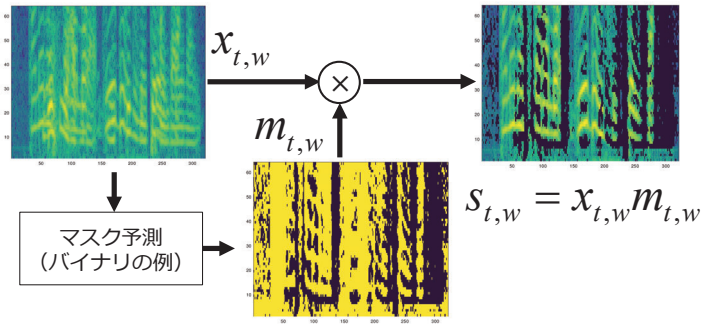
基本的なアプローチの一つ マスク生成

（パワー）スペクトルをマスク処理
– 時間周波数マスク（TF-mask）を推定



基本的なアプローチの一つ マスク生成

（パワー）スペクトルをマスク処理
– 時間周波数マスク（TF-mask）を推定



分類/セグメンテーション問題として定式化する場合の一例

各t,w成分の音源クラスを推定

例: 音声+雑音 → 音声成分抽出
– 音声（1） or not（0） の2値分類

$$\hat{\mathbf{M}} = \arg\max_{\mathbf{M}} p(\mathbf{M} \mid \mathbf{X})$$


$\mathbf{X}$ 入力のスペクトル: w 行 t 列成分 $x_{w,t}$


$\mathbf{M}$ クラスを表す変数: w 行 t 列成分 $m_{w,t}$

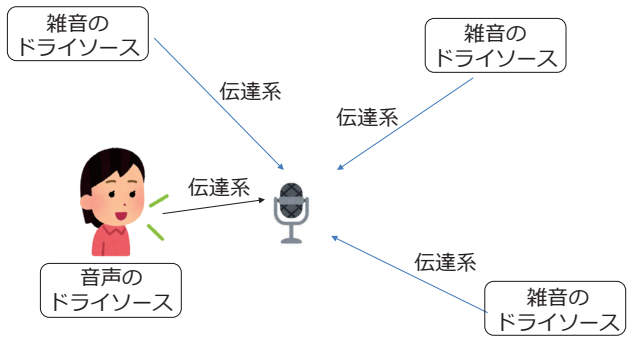
事後確率を NN でモデル化
（事前に教師あり学習しておく）

学習用のデータ・ラベルの生成例（シミュレートに基づく生成の一例）

必要物
– (a) 音声信号, (b) 雑音信号: それぞれ単体
– シミュレートに必要なインパルス応答など

手順
1. 伝達系をシミュレートした (c)音声信号, (d) 雑音信号 を生成
2. (e) 混合音 を (c) + (d) で生成
3. (e), (c), (d) などを用いて教師信号の生成（確率値，クラスラベル）

学習用のデータ・ラベルの生成例（シミュレートに基づく生成の一例）



背景雑音は色んなパターンを用意する必要

2クラスにおけるベーシックな教師信号

$p_{w,t,1}$：$x_{w,t}$ が音声である確率

Ideal Binary Mask

$$p_{w,t,1} = \begin{cases} 1 & \text{if } \mathrm{SNR}_{w,t} > C \\ 0 & \text{otherwise} \end{cases}$$

$x_{w,t}$ の信号対雑音比 $\mathrm{SNR}_{w,t}$

Ideal Ratio Mask

$$p_{w,t,1} = \left(\frac{|s_{w,t}|^2}{|s_{w,t}|^2 + |n_{w,t}|^2} \right)^{\beta}$$

大量データ生成に基づき教師あり学習を行う

最近の処理フロー

End-to-end 構成: 全部学習

エンコーダ: 波形 → 特徴量 変換
セパレータ/Masker: 特徴量領域でマスク推定
デコーダ: 特徴量 → 波形 変換

マスク処理は内部に隠蔽 ＋ 回帰問題 $\boldsymbol{s} = \boldsymbol{f}(\mathbf{X})$
– NN は 事前学習

- 教師あり: SI-SDR + PIT(Permutation invariant training)
- 教師なし: 最尤推定, Mixture invariant training

参考文献

- Aapo Hyvarinen 他, 根本・川勝 訳「詳解 独立成分分析 信号解析の新しい世界」
- C.M.ビショップ「パターン認識と機械学習 上下」
- 持橋大地，大羽成征「ガウス過程と機械学習」
- 浅野太「音のアレイ信号処理」

参考文献

T. Kim, T. Eltoft, TW. Lee, “Independent Vector Analysis: An Extension of ICA to Multivariate Components” in ICA, 2006.

A. Hiroe, “Solution of Permutation Problem in Frequency Domain ICA Using Multivariate Probability Density Functions” in ICA 2006.

N. Q. K. Duong, E. Vincent and R. Gribonval, "Under-Determined Reverberant Audio Source Separation Using a Full-Rank Spatial Covariance Model," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1830-1840, 2010.

H. Sawada, H. Kameoka, S. Araki and N. Ueda, “Multichannel Extensions of Non-Negative Matrix Factorization With Complex-Valued Data,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 5, pp. 971-982, 2013.

DeLiang Wang and Jitong Chen, “Supervised Speech Separation Based on Deep Learning: An Overview,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 26, No. 10, pp.1702—1726, 2018.

Scott Wisdom, Efthymios Tzinis, Hakan Erdogan, Ron Weiss, Kevin Wilson, John Hershey, “Unsupervised Sound Separation Using Mixture Invariant Training,” *Advances in Neural Information Processing Systems 33*, 2020.

M. Delcroix, J. B. Vázquez, T. Ochiai, K. Kinoshita, Y. Ohishi and S. Araki, "SoundBeam: Target Sound Extraction Conditioned on Sound-Class Labels and Enrollment Clues for Increased Performance and Continuous Learning," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 121-136, 2023.