

Unified Inference for Variational Bayesian Linear Gaussian State-Space Models

David Barber and Silvia Chiappa

Advances in Neural Information Processing Systems 19

(担当者)

金崎 弘文

東京大学 先端科学技術研究センター

町田・矢入研究室

NIPS2006 を読む会

東京工業大学・大岡山キャンパス

2007/1/26 (fri.)

論文の概要

- 対象とする問題

 - システムのパラメータ推定（時間普遍）および状態推定（時間変化）

- 対象とするシステム（扱うモデル）

 - ベイジアン線型・ガウスノイズシステム

 - (Bayesian LGSSMs; Linear Gaussian State-Space Models)

- 論文の貢献

 - 変分ベイズ法による解法が提案されているBayesian LGSSMs を,
 - それと等価な LGSSMs に変換して,
 - Kalman Filtering/Smoothing アルゴリズムで解く方法を導出した.

- 研究特徴

 - パラメータの曖昧性（事前分布）を考慮したベイズ推定手法
 - 再帰計算可能なベイズ学習アルゴリズム,

 - * 変分ベイズ法（バッチ型, 反復計算）に基づくアルゴリズムをKalman Filter/ Smootherと対応する型（逐次計算型）に変形.

 - ベイズ学習アルゴリズムとKF/KSを統一的に理解 ("Unified") .
 - 変分ベイズ法を用いる従来手法よりエレガント

ON BAYES PROPAGATION. Our framework both simplifies and unifies the inference problem, so that future applications may be easily developed. We demonstrate the elegance of the approach on Bayesian temporal ICA, with an application to finding independent components in noisy EEG signals.

(Abstractより抜粋)

説明の流れ

- 扱うモデル（線型ガウスモデル, ベイジアン線型ガウスモデル）の説明
 - 1. Linear Gaussian State-Space Models
 - 2. Bayesian Linear Gaussian State-Space Models
- 変分ベイズ法（バッチ, 反復）の説明
 - 2. 節の "Variational Bayes"
- 問題の変換
 - 3. Unified Inference on $q(h)$
- 実験
 - 4. An Application to Bayesian ICA
- まとめ

1. Linear Gaussian State-Space Models

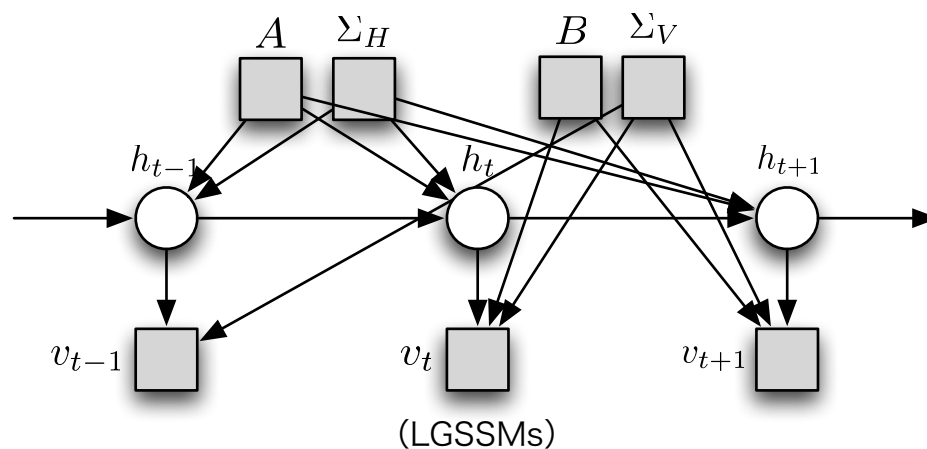
• LGSSMsとは？

- 観測方程式, 運動方程式が線型. $v_t = Bh_t + \eta_t^v, \eta_t^v \sim \mathcal{N}(\mathbf{0}_v, \Sigma_v)$ (観測方程式)
- ノイズがガウスノイズ. $h_t = Ah_{t-1} + \eta_t^h, \eta_t^h \sim \mathcal{N}(\mathbf{0}_H, \Sigma_H)$ (運動方程式)
- パラメータ (6個)
 $\Theta = \{A, B, \Sigma_H, \Sigma_V, \mu, \Sigma\}$

初期状態のパラメータ

(研究の動機)

- **パラメータに曖昧性**があるという条件下で, **パラメータおよび状態を推定**をしたい.
- 確率伝搬法 (BP; Belief Propagation), 変分ベイズ法 (VB; Variational Bayes)
 - LGSSMsに関して広く知られているカルマンフィルターとはどんな関係なのか？
 - VB法とKF/KSを組み合わせると, エレガントな形 (Predictor-Corrector, Raugh-Tung-Striebel recursion) にしたい！！



2. Bayesian Linear Gaussian State-Space Models (2)

変分ベイズ法の復習 (1) : 変分近似・平均場近似

$$\begin{aligned}
 \mathcal{L} &= \log p(v_{1:T} | \hat{\Theta}) && \text{ログエビデンス} \\
 & && \text{(エビデンス : 未知変数で周辺化された周辺尤度)} \\
 &= \log \sum_{h_{1:T}} \int p(v_{1:T}, h_{1:T}, \Theta | \hat{\Theta}) d\Theta \\
 & && \text{未知変数 (知りたいもの)} \\
 &= \log \sum_{h_{1:T}} \int \boxed{q(h_{1:T}, \Theta)} \frac{p(v_{1:T}, h_{1:T}, \Theta | \hat{\Theta})}{\boxed{q(h_{1:T}, \Theta)}} d\Theta && \text{テスト分布} \\
 & && \text{(未知変数の分布を近似的に表現)} \\
 &\geq \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \frac{p(v_{1:T}, h_{1:T}, \Theta | \hat{\Theta})}{q(h_{1:T}, \Theta)} d\Theta && \text{(変分近似 : Jensenの不等式を適用)} \\
 & && \text{エントロピーの定義そのまま} \\
 &= \boxed{- \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log q(h_{1:T}, \Theta) d\Theta} + \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \boxed{p(v_{1:T}, h_{1:T}, \Theta | \hat{\Theta})} d\Theta && \text{さらに, 因数分解する. (モデル構造)} \\
 &= H_q(\Theta, h_{1:T}) + \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log p(v_{1:T}, h_{1:T} | \Theta, \hat{\Theta}) d\Theta + \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log p(\Theta | \hat{\Theta}) d\Theta \\
 &= H_q(\Theta, h_{1:T}) + \langle E(h_{1:T}, \Theta | \hat{\Theta}) \rangle_{q(\Theta, h_{1:T})} + \langle \log p(\Theta | \hat{\Theta}) \rangle_{q(\Theta)} \\
 &\equiv \mathcal{F}(q)
 \end{aligned}$$

Jensenの不等式

$$\begin{aligned}
 E[\phi(x)] &\geq \phi(E[x]) \quad \because \phi(x) : \text{convex} \\
 E[\phi(x)] &\leq \phi(E[x]) \quad \because \phi(x) : \text{concave}
 \end{aligned}$$

例) $\phi(x) = -\log x$

$$\frac{\sum_{i=1}^n x_i}{n} \geq \left(\prod_{i=1}^n x_i \right)^{1/n}$$

2. Bayesian Linear Gaussian State-Space Models (2)

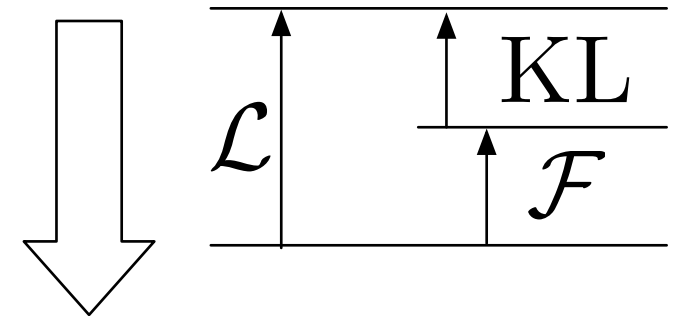
変分ベイズ法の復習 (2) : 変分問題を解く.

ログエビデンスとその変分近似 \mathcal{F} , およびテスト分布 q の関係

$$\begin{aligned} \mathcal{L} - \mathcal{F} &= \log p(v_{1:T}|\hat{\Theta}) - \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \frac{p(v_{1:T}, h_{1:T}, \Theta|\hat{\Theta})}{q(h_{1:T}, \Theta)} d\Theta \\ &= \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log p(v_{1:T}|\hat{\Theta}) d\hat{\Theta} - \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \frac{p(v_{1:T}, h_{1:T}, \Theta|\hat{\Theta})}{q(h_{1:T}, \Theta)} d\Theta \\ &= \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \frac{q(h_{1:T}, \Theta)p(v_{1:T}|\hat{\Theta})}{p(v_{1:T}, h_{1:T}, \Theta|\hat{\Theta})} d\Theta \\ &= \sum_{h_{1:T}} \int q(h_{1:T}, \Theta) \log \frac{q(h_{1:T}, \Theta)}{p(h_{1:T}, \Theta|v_{1:T}, \hat{\Theta})} d\Theta \\ &= \text{KL} \left(q(h_{1:T}, \Theta), p(h_{1:T}, \Theta|v_{1:T}, \hat{\Theta}) \right) \end{aligned}$$

テスト分布と未知変数の分布の非類似度最小化問題

テスト分布と未知変数の分布 (知りたいもの) との非類似度



汎関数 \mathcal{F} の変関数 q に関して最大化する変分問題

$$\mathcal{F}(q) \equiv H_q(\Theta, h_{1:T}) + \langle E(h_{1:T}, \Theta|\hat{\Theta}) \rangle_{q(\Theta, h_{1:t})} + \langle \log p(\Theta|\hat{\Theta}) \rangle_{q(\Theta)}$$

変分問題を解くと・・・**反復計算**により最適解を求める計算式が出てくる.

(EMアルゴリズムに類似, BPと等価.)

$$q(h_{1:T}) \propto \exp \left(\langle E(h_{1:T}, \Theta|\hat{\Theta}) \rangle_{q(\Theta)} \right)$$

$$q(\Theta) \propto p(\Theta) \exp \left(\langle E(h_{1:T}, \Theta|\hat{\Theta}) \rangle_{q(h_{1:T})} \right)$$

VB法に基づいて知りたいものを求める
計算法を導出した.

3. Unified Inference on $q(h)$ (2)

Mean + Fluctuation Decomposition

- 変分ベイズ法に基づいて、パラメータ・状態推定を行うための反復計算法を導出した。

$$q(h_{1:T}) \propto \exp \left(\langle E(h_{1:T}, \Theta | \hat{\Theta}) \rangle_{q(\Theta)} \right) \quad E(h_{1:T}, \Theta | \hat{\Theta}) \equiv \log p(v_{1:T}, h_{1:T} | \Theta, \hat{\Theta})$$

$$q(\Theta) \propto p(\Theta) \exp \left(\langle E(h_{1:T}, \Theta | \hat{\Theta}) \rangle_{q(h_{1:T})} \right) \quad = -\frac{1}{2} \sum_{t=1}^T [(v_t - Bh_t)^T \Sigma_V^{-1} (v_t - Bh_t) + (h_t - Ah_{t-1})^T \Sigma_H^{-1} (h_t - Ah_{t-1})]$$

$$\langle E(h_{1:T}, \Theta | \hat{\Theta}) \rangle_{q(\Theta)} = -\frac{1}{2} \sum_{t=1}^T [\langle (v_t - Bh_t)^T \Sigma_V^{-1} (v_t - Bh_t) \rangle_{q(B)} + \langle (h_t - Ah_{t-1})^T \Sigma_H^{-1} (h_t - Ah_{t-1}) \rangle_{q(A)}] \dots (2) \text{ 式}$$

- Kalman Filterのようなpredictor-corrector的な計算アルゴリズムにできないか？
 - exp の肩にかかる部分の計算を逐次的に行えないか？

- 上記更新式に含まれている $\langle E(\dots) \rangle$ に関する考察

A useful decomposition is to write

$$\langle (v_t - Bh_t)^T \Sigma_V^{-1} (v_t - Bh_t) \rangle_{q(B, \Sigma_V)} = \underbrace{(v_t - \langle B \rangle h_t)^T \langle \Sigma_V^{-1} \rangle (v_t - \langle B \rangle h_t)}_{\text{mean}} + \underbrace{h_t^T S_B h_t}_{\text{fluctuation}},$$

and similarly

$$\langle (h_t - Ah_{t-1})^T \Sigma_H^{-1} (h_t - Ah_{t-1}) \rangle_{q(A, \Sigma_H)} = \underbrace{(h_t - \langle A \rangle h_{t-1})^T \langle \Sigma_H^{-1} \rangle (h_t - \langle A \rangle h_{t-1})}_{\text{mean}} + \underbrace{h_{t-1}^T S_A h_{t-1}}_{\text{fluctuation}},$$

where the parameter covariances are $S_B = V H_B^{-1}$ and $S_A = H H_A^{-1}$ (see Appendix A).

... (2') 式

$$\begin{aligned} & \int \sigma_v^{-1} (v_t - b h_t)^2 q(b) db \\ &= \int \sigma_v^{-1} (v_t - b h_t)^2 q(b) db \\ &= \sigma_v^{-1} \int \{v_t^2 - 2v_t b h_t + b^2 h_t^2 q(b)\} db \\ &= \sigma_v^{-1} \{v_t^2 - 2v_t \langle b \rangle h_t + \langle b^2 \rangle h_t^2\} \\ &= \sigma_v^{-1} [v_t^2 - 2v_t \langle b \rangle h_t + \{\langle b \rangle^2 + \sigma_b\} h_t^2] \\ &= \sigma_v^{-1} (v_t - 2\langle b \rangle h_t)^2 + \frac{\sigma_b}{\sigma_v} h_t^2 \end{aligned}$$

元のBayesian-LGSSMsと「等価」（再帰的な尤度計算が（2'）と一致する）LGSSMsは、ないだろうか？
 LGSSMsに変換できれば、KF/KSのアルゴリズムをそのまま使えるのではないか？

3. Unified Inference on $q(h)$ (3)

Appendix A

ここでやっていること： $q(A)$ および $q(B)$ を表現するガウシアンを、平均と分散を求める。

A.1 Determining $q(B|\Sigma_V)$

By examining \mathcal{F} , the contribution of $q(B|\Sigma_V)$ can be interpreted as the negative KL divergence between $q(B|\Sigma_V)$ and a Gaussian. Hence, optimally, $q(B|\Sigma_V)$ is a Gaussian. The covariance $[\Sigma_B]_{ij,kl} \equiv \langle (B_{ij} - \langle B_{ij} \rangle)(B_{kl} - \langle B_{kl} \rangle) \rangle$ (averages wrt $q(B|\Sigma_V)$) is given by:

観測行列の曖昧性

観測値のバラツキ

$$[\Sigma_B]_{ij,kl} = [H_B^{-1}]_{jl} [\Sigma_V]_{ik}, \quad \text{where } [H_B]_{jl} \equiv \sum_{t=1}^T \langle h_t^j h_t^l \rangle_{q(h_t)} + \beta_j \delta_{jl}.$$

係数

The mean is given by $\langle B \rangle = N_B H_B^{-1}$, where $[N_B]_{ij} \equiv \sum_t \langle h_t^j \rangle v_t^i + \beta_j \hat{B}_{ij}$.

Determining $q(A|\Sigma_H)$

Optimally, $q(A|\Sigma_H)$ is a Gaussian with covariance

遷移行列の曖昧性

状態のノイズ

$$[\Sigma_A]_{ij,kl} = [H_A^{-1}]_{jl} [\Sigma_H]_{ik}, \quad \text{where } [H_A]_{jl} \equiv \sum_{t=1}^{T-1} \langle h_t^j h_t^l \rangle_{q(h_t)} + \alpha_j \delta_{jl}.$$

係数

The mean is given by $\langle A \rangle = N_A H_A^{-1}$, where $[N_A]_{ij} \equiv \sum_{t=2}^T \langle h_{t-1}^j h_t^i \rangle + \alpha_j \hat{A}_{ij}$.

3. Unified Inference on $q(h)$ (3)

Bayesian LGSSMs \rightarrow LGSSMs

Bayesian LGSSMs	Augmented LGSSMs
v_t	$\tilde{v}_t = \begin{pmatrix} v_t \\ 0_H \\ 0_H \end{pmatrix}$
B	$\tilde{B} = \begin{pmatrix} \langle B \rangle \\ U_A \\ U_B \end{pmatrix}$
A	$\tilde{A} = \langle A \rangle$
Σ_H	$\tilde{\Sigma}_H = \langle \Sigma_H^{-1} \rangle^{-1}$
Σ_V	$\tilde{\Sigma}_V = \begin{pmatrix} \langle \Sigma_V^{-1} \rangle^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
μ	$\tilde{\mu} = \mu$
Σ	$\tilde{\Sigma} = \Sigma$
α	
β	

ただし,

$$U_A^T U_A = S_A = H H_A^{-1}$$

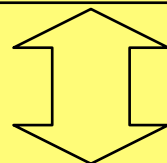
$$U_B^T U_B = S_B = H H_B^{-1}$$

- ・左のテーブルのようにパラメータを書き換えたLGSSMsを考える.
- ・(2') を直接計算可能.

計算例:

$$\begin{aligned} & \langle (v_t - B h_t)^T \Sigma_V^{-1} (v_t - B h_t) \rangle_{q(B)} \\ &= (v_t - \langle B \rangle h_t)^T \langle \Sigma_V^{-1} \rangle (v_t - \langle B \rangle h_t) + h_t^T S_B h_t \\ &= (\tilde{v}_t - \tilde{B} h_t)^T \tilde{\Sigma}_V (\tilde{v}_t - \tilde{B} h_t) \end{aligned}$$

元のBayesian-LGSSMsの状態推定問題



元のBayesian-LGSSMsと「等価」(再帰的な尤度計算が(2')と一致する)なAugmented LGSSMsに変換.

(Augmented-) LGSSMsの状態推定問題

3. Unified Inference on $q(h)$ (4)

Inference Using an Augmented LGSSMs

Algorithm 1 LGSSM: Forward and backward recursive updates. The smoothed posterior $p(h_t|v_{1:T})$ is returned in the mean \hat{h}_t^T and covariance P_t^T .

- Forwardアルゴリズム
 - 状態量推定 (フィルタリング)
 - a系列: 通常のカルマンフィルター
 - b系列: LGSSMsのパラメータに変換行列を加えた記述法.

- Backwardアルゴリズム
 - 通常のカルマン smoother

- 変換行列Dの意味・解釈

システム (A, B) の曖昧性がある分だけ, 通常のカFよりも予測値に内在する分散を割ります.

```

procedure FORWARD
    1a:  $P \leftarrow \Sigma$ 
    1b:  $P \leftarrow D\Sigma$ , where  $D \equiv I - \Sigma U_{AB} (I + U_{AB}^T \Sigma U_{AB})^{-1} U_{AB}^T$ 
    2a:  $\hat{h}_1^0 \leftarrow \mu$ 
    2b:  $\hat{h}_1^0 \leftarrow D\mu$ 
    3:  $K \leftarrow PB^T(BPB^T + \Sigma_V)^{-1}$ ,  $P_1^1 \leftarrow (I - KB)P$ ,  $\hat{h}_1^1 \leftarrow \hat{h}_1^0 + K(v_1 - B\hat{h}_1^0)$ 
    for  $t \leftarrow 2, T$  do
        4:  $P_t^{t-1} \leftarrow AP_{t-1}^{t-1}A^T + \Sigma_H$ 
        5a:  $P \leftarrow P_t^{t-1}$ 
        5b:  $P \leftarrow D_t P_t^{t-1}$ , where  $D_t \equiv I - P_t^{t-1} U_{AB} (I + U_{AB}^T P_t^{t-1} U_{AB})^{-1} U_{AB}^T$ 
        6a:  $\hat{h}_t^{t-1} \leftarrow A\hat{h}_{t-1}^{t-1}$ 
        6b:  $\hat{h}_t^{t-1} \leftarrow D_t A\hat{h}_{t-1}^{t-1}$ 
        7:  $K \leftarrow PB^T(BPB^T + \Sigma_V)^{-1}$ ,  $P_t^t \leftarrow (I - KB)P$ ,  $\hat{h}_t^t \leftarrow \hat{h}_t^{t-1} + K(v_t - B\hat{h}_t^{t-1})$ 
    end for
end procedure

procedure BACKWARD
    for  $t \leftarrow T - 1, 1$  do
         $\overleftarrow{A}_t \leftarrow P_t^t A^T (P_{t+1}^t)^{-1}$ 
         $P_t^T \leftarrow P_t^t + \overleftarrow{A}_t (P_{t+1}^T - P_{t+1}^t) \overleftarrow{A}_t^T$ 
         $\hat{h}_t^T \leftarrow \hat{h}_t^t + \overleftarrow{A}_t (\hat{h}_{t+1}^T - A\hat{h}_t^t)$ 
    end for
end procedure
    
```

初期値

分散の予測

状態予測

変換行列

カルマンゲイン

分散の更新

状態更新

$$D_t \equiv I - P_t^{t-1} U_{AB} (I + U_{AB}^T P_t^{t-1} U_{AB})^{-1} U_{AB}^T$$

4. An Application to Bayesian ICA (1)

- ・提案手法を**独立成分分析 (ICA; Independent Component Analysis)** の問題に適用

- ・ICAとは・・・

複数系列の入力信号から、信号を構成している独立な信号成分を抽出する。

例1) 複数話者の会話を複数マイク録音しておき、録音した音声信号から、各話者の会話を分離する。

例2) 複数の測定器で計測した脳波（元の脳波の混合した信号）から、意味のある信号を分離する。

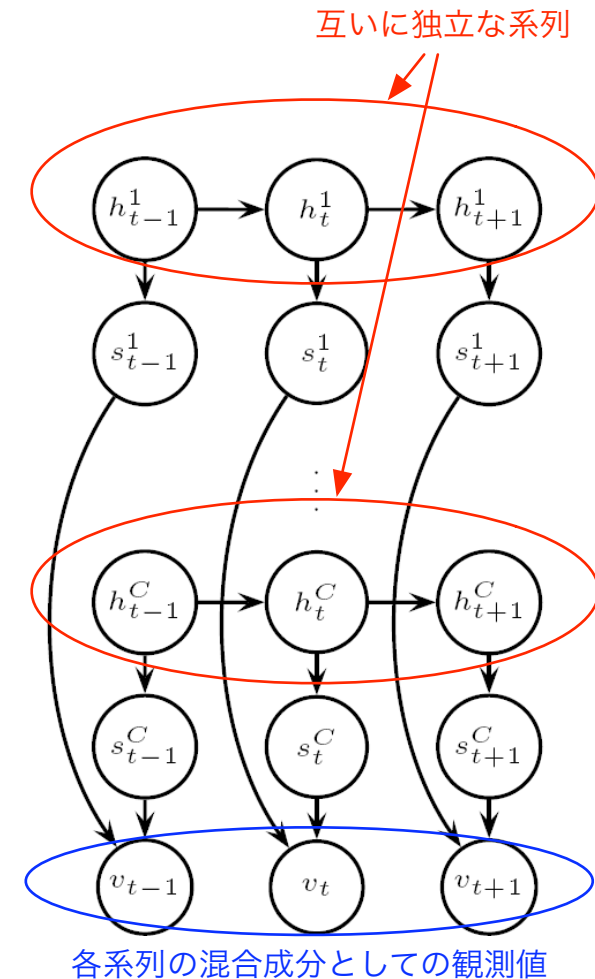


Figure 1: The structure of the LGSSM for ICA.

4. An Application to Bayesian ICA (2)

- ICAで扱う問題は、観測行列に曖昧性のある状態空間モデルとしてモデル化することができる。
- ICAでは状態推定というよりは、独立な各信号系列とその値を求めることが目的。
- 各系列は、単一のLGSSMs.
- 全体は、観測行列 B に曖昧性がある Bayesian LGSSMs.
- 観測行列 $B=WP$ の形でモデル化
 - 各系列の信号出力： $P \cdots \cdots$ 既知 (単位行列)
 - 混合： $W \cdots \cdots$ 未知

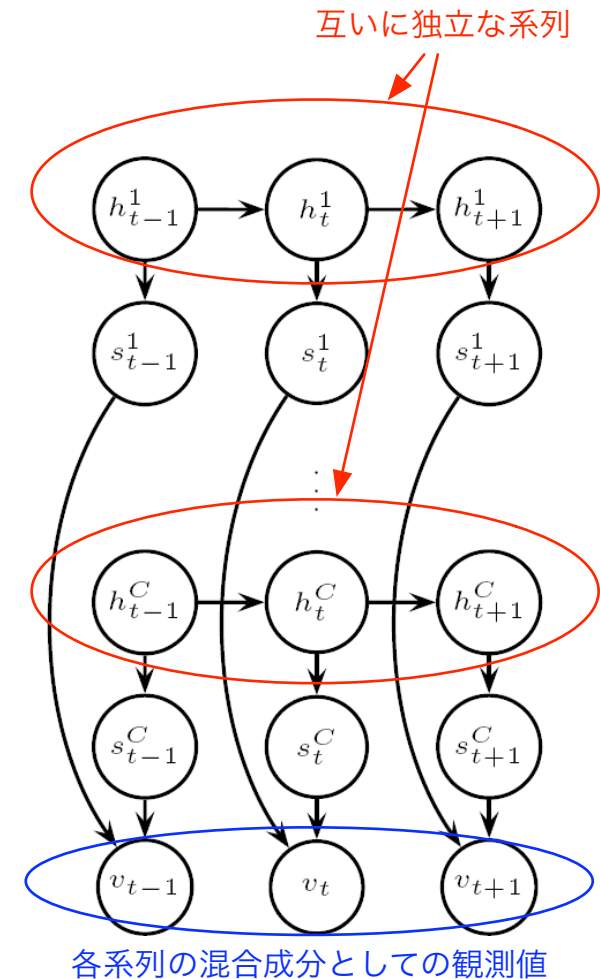


Figure 1: The structure of the LGSSM for ICA.

4.1 Demonstration (実験1)

(設定)

- 元信号：状態空間が5次元のLGSSMs x 3系列
- 学習に使うデータ：元信号が混合した信号 x 3系列
- 学習に使うモデル：各状態空間が7次元 x 5系列

(結果)

- 提案手法とMAP推定での性能比較

(MAP推定に用いる事後確率)

$$\log p(v_{1:t} | A, W, \Theta) + \log p(A | \alpha) + \log p(W | \beta)$$

- MAPでは余計な系列の存在を推定しているが、提案手法では3系列を特定.

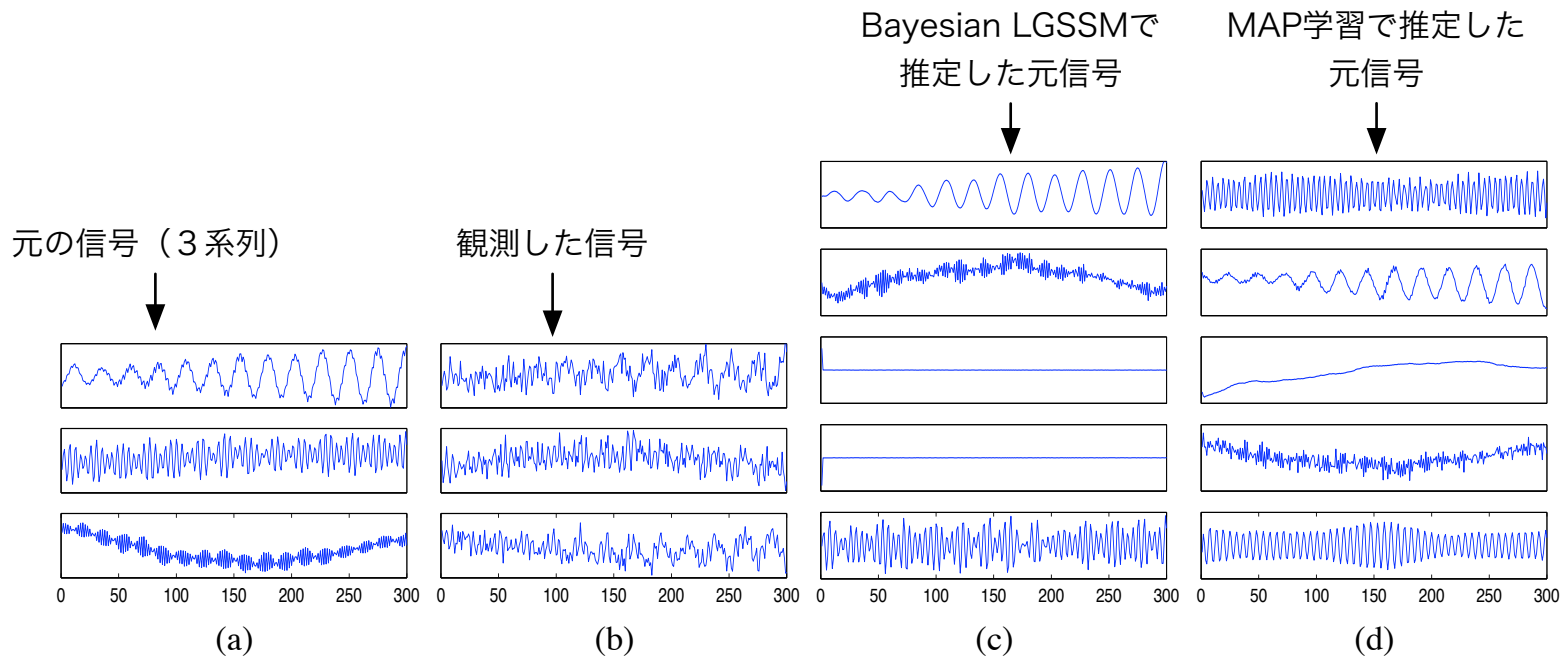


Figure 2: (a) Original sources s_t . (b) Observations resulting from mixing the original sources, $v_t = Ws_t + \eta_t^v$, $\eta_t^v \sim \mathcal{N}(0, I)$. (c) Recovered sources using the Bayesian LGSSM. (d) Sources found with MAP LGSSM.

4.2 Application to EEG Analysis (実験2)

(1) 概要

- EEG (ElectroEncephaloGraph) : 脳波
- 通常のICA (例 : FastICA) を用いる場合の問題点
ノイズの載った生データに対して, 帯域通過フィルタを適用するなどの前処理が必要.
- EEGの分析に求められる性能
観測されるEEGは多くの脳内信号の混合であるが, そこから独立成分を何個抽出するかを柔軟に設定できること.

4.2 Application to EEG Analysis (実験2)

(2) 中身

(設定)

- 学習に使うデータ：4系列の脳波
- 学習に使うモデル：遷移行列の4つの周波数帯ごとに各4つ、全16通り指定。
(見つけたい独立成分の周波数帯および、見つけたい独立成分数の最大値を指定。)

(結果)

- 0.5Hz帯：低周波数の変動成分
- 10-11Hz帯：3つだけ。
- 20-21Hz帯：1つだけ。
- 50Hz帯：50Hz帯のものが2つと、他の周波数帯のもの2つ。

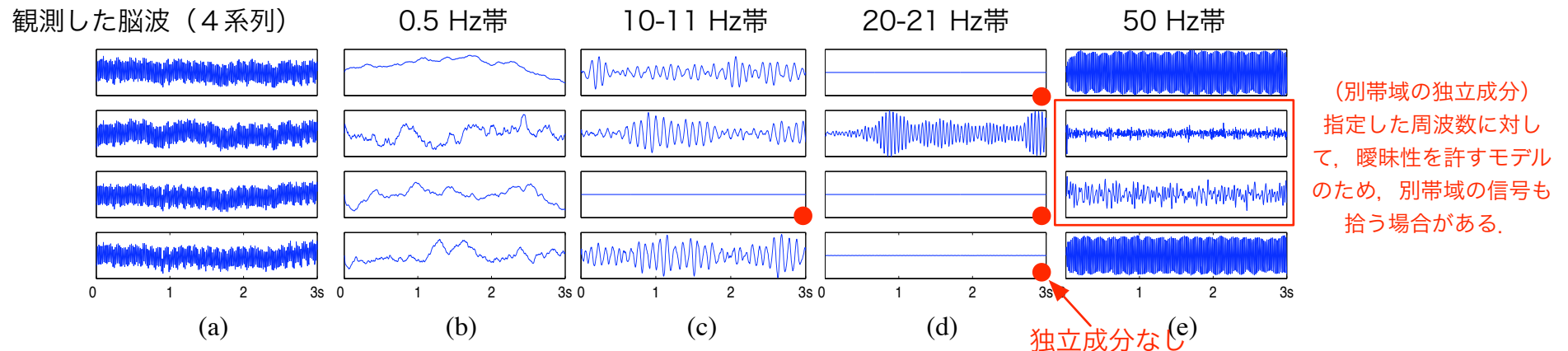


Figure 3: (a) Original raw EEG recordings from 4 channels. (b-e) 16 sources s_t estimated by the Bayesian LGSSM.

5. Conclusion

- ・ LGSSMsの状態量およびパラメータをベイズ推定する問題を扱った.
- ・ 従来の変分ベイズに基づく解法は、効率的で安定な既存の手法（Kalman Filter/ Smoother）との対応関係が不明であった.
- ・ Bayesian-LGSSMsをそれと等価なLGSSMsに変換した。これにより、KF/KSでベイズ推定が可能になった。 [main contribution]
- ・ ICA問題に適用して、提案手法が機能することを確認した.
- ・ 同様の考え方によって、他のベイズ推定問題を既存のアルゴリズムで解ける可能性がある.