



Belief Propagation for Probabilistic Slow Feature Analysis

Omori, Toshiaki
Sekiguchi, Tomoki
Okada, Masato

(Citation)

Journal of the Physical Society of Japan, 86(8):084802-084802

(Issue Date)

2017-08-15

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

©2017 The Physical Society of Japan.

This article is available under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the article, journal citation, and DOI.

(URL)

<https://hdl.handle.net/20.500.14094/90004846>



Belief Propagation for Probabilistic Slow Feature Analysis

Toshiaki Omori^{1*}, Tomoki Sekiguchi², and Masato Okada^{2,3}

¹Graduate School of Engineering, Kobe University, Kobe 657-8501, Japan

²Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan

³RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan

(Received March 29, 2016; accepted May 26, 2017; published online July 11, 2017)

Slow feature analysis (SFA) is a time-series analysis method for extracting slowly-varying latent features from multi-dimensional data. A recent study proposed a probabilistic framework of SFA using the Bayesian statistical framework. However, the conventional probabilistic framework of SFA can not accurately extract the slow feature in noisy environments since its marginal likelihood function was approximately derived under the assumption that there exists no observation noise. In this paper, we propose a probabilistic framework of SFA with rigorously derived marginal likelihood function. Here, we rigorously derive the marginal likelihood function of the probabilistic framework of SFA by using belief propagation. We show using numerical data that the proposed probabilistic framework of SFA can accurately extract the slow feature and underlying parameters for the latent dynamics simultaneously even under noisy environments.

1. Introduction

Slow feature analysis (SFA) is a time-series analysis method for extracting slowly varying features from multi-dimensional data.¹⁾ In recent years, the SFA has attracted much attention in theoretical neuroscience and has been used to establish receptive field models for complex cells in the visual system, place cells in the hippocampus, and grid cells in the entorhinal cortex.^{2–4)} These theoretical studies suggest that the extraction of slowly varying features plays an important role in the information processings in neural systems. Recent machine learning studies have also used the SFA to improve pattern recognition and feature extraction algorithms from multi-dimensional data.^{5–11)}

A probabilistic version of SFA has been proposed using the framework of the state-space model,¹²⁾ a probabilistic perspective has been useful for many machine learning algorithms such as those for principal component analysis¹³⁾ and independent component analysis.¹⁴⁾ Probabilistic models have several merits, such as many powerful methods developed for learning and inference and so on. In the conventional probabilistic framework of SFA,¹²⁾ the marginal likelihood function is approximately derived by assuming that there exists no observation noise, although observed data would include noise in general. Therefore, it had been unclear whether the slow feature estimated by the conventional probabilistic framework of SFA is accurate for noisy data. From this point of view, a recent study showed that the conventional probabilistic framework of SFA cannot extract slow features accurately under noisy environments.¹⁵⁾

In this study, we propose a probabilistic version of SFA considering the effect of observation noise in order to extract the slow feature and its underlying parameters accurately. Here, we rigorously derive the marginal likelihood function of the probabilistic framework of SFA by using belief propagation.^{16–19)} The belief propagation for one-dimensional graphical models is also known as the *transfer-matrix method* in statistical physics,^{16,17,19–21)} and it enables the rigorous derivation of the marginal distribution for graphical models with no loops, where we find that the graphical model of the probabilistic framework of SFA has no loop structure. We show using numerical data that our proposed method can

accurately extract the slow feature even from noisy time-series data.

2. Theory

In this section, we first review the conventional SFAs: the deterministic SFA and the conventional probabilistic framework of SFA. Next, we propose a novel probabilistic framework of SFA with a rigorously derived marginal likelihood function. We rigorously derive the marginal likelihood function by means of belief propagation in order to realize the accurate and robust estimation of the slow feature and parameters even under noisy environments.

2.1 Deterministic SFA

The original SFA is a deterministic algorithm for extracting the most slowly varying feature (called the “slow feature”) from multi-dimensional time series data.¹⁾ A schematic diagram of the deterministic SFA is shown in Fig. 1(a). For multi-dimensional input time-series data $\mathbf{x}(t) \in \mathbb{R}^M$, the output of SFA $y_j(t)$ ($j = 1, \dots, N$) is obtained using the transformation $y_j(t) = g_j(\mathbf{x}(t))$. This transformation $g_j(\mathbf{x})$ is determined to minimize the following expression:

$$\Delta(y_j) = \langle \dot{y}_j^2 \rangle_t, \quad (1)$$

where $\Delta(y_j)$ is called a Δ -value and $\langle \cdot \rangle_t$ denotes an average with respect to time. Namely, in deterministic SFA, we perform the transformation $g_j(\mathbf{x})$, which minimizes the averaged square of the derivative of output $\mathbf{y}(t)$ with respect to time t . Within the outputs of deterministic SFA $\{y_j(t)\}$, the output $y_j(t)$ with the minimum Δ -value corresponds to a slow feature. The following three constraint conditions,

$$\langle y_j \rangle_t = 0 \quad (\text{zero mean}), \quad (2)$$

$$\langle y_j^2 \rangle_t = 1 \quad (\text{unit variance}), \quad (3)$$

$$\langle y_i y_j \rangle_t = 0 \quad (\text{decorrelation}), \quad (4)$$

are used to normalize the output signals, avoid trivial results, and guarantee that different output elements have different information.

2.2 Probabilistic SFA

A probabilistic version of SFA has been proposed recently

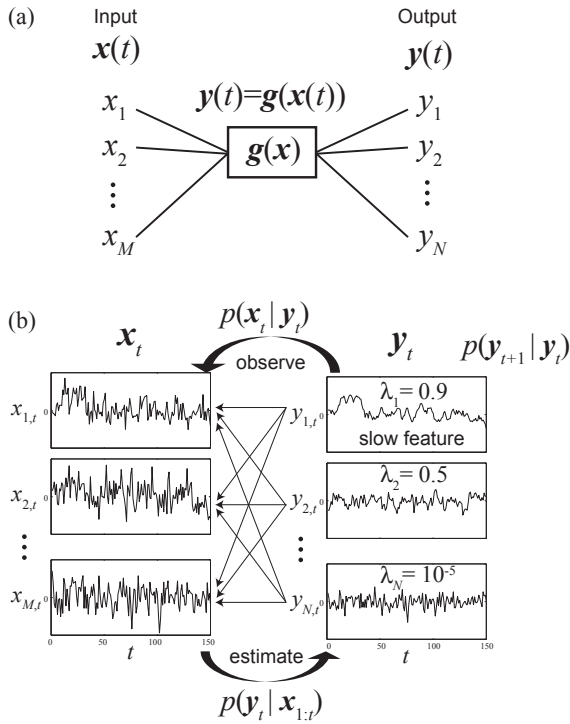


Fig. 1. Schematic diagrams of slow feature analysis (SFA). (a) In deterministic SFA, multi-dimensional time series data $x(t)$ are transformed via scalar functions $g_j(x)$ into outputs $y_j(t)$. An output element with minimal Δ -value corresponds to slow feature. (b) In probabilistic framework of SFA, latent variables y_t are estimated from observed variable x_t by using Bayesian statistical framework. The latent variable $y_{i,t}$ with the largest λ_i corresponds to the slow feature.

based on Bayesian statistical framework.¹²⁾ However, in that version of the SFA, the marginal likelihood function used to estimate parameters of the probabilistic model is approximately derived by assuming no observation noise.¹²⁾ A more recent theoretical study has shown that the conventional probabilistic framework of SFA cannot estimate the slow feature accurately under noisy conditions.¹⁵⁾

In this study, we propose a novel probabilistic framework of SFA with rigorous derivation of the marginal likelihood function. We consider the extraction of N -dimensional latent variables y_t from M -dimensional observed variables x_t for discrete-time steps $t = 1, \dots, T$ [Fig. 1(b)]. As shown in Fig. 1(b), observed multi-dimensional time series data are transformed into multiple latent time series with different degrees of slowness.

The probabilistic framework of SFA¹²⁾ is based on the state-space model, consisting of a system model and an observation model, and is expressed by the graphical model shown in Fig. 2. As shown by the gray circles in Fig. 2, we assume that the latent variables $y_t = [y_{1,t}, \dots, y_{N,t}]^T$ depend on those at the preceding time y_{t-1} . The dynamics of latent variables y_t are described by the following system model:

$$y_t = \lambda y_{t-1} + \eta_t. \quad (5)$$

Here, λ is a diagonal matrix with elements λ_n ($n \in \{1, \dots, N\}$). Each element λ_n determines the dynamics of the n th latent variable $y_{n,t}$. Namely, λ_n is an indicator of the slowness of each latent variable. Note that the latent variable with the largest value of λ_n corresponds to the slow feature to be extracted in the probabilistic framework of SFA. The

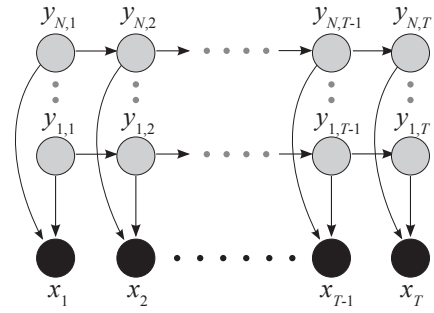


Fig. 2. A graphical structure of the probabilistic SFA. Each observed variable x_t at time t depends on a latent variable y_t at the same time t , whereas the latent variable y_t at time t depends on the latent variable y_{t-1} at the preceding time $t-1$. The graphical structure of the probabilistic SFA has a straight structure and no loops. Based on this graphical structure, we employ belief propagation in order to derive the marginal likelihood function rigorously.

vector η_t describes a system noise obeying white Gaussian noise with average $\mathbf{0}$ and covariance Σ , where Σ is a diagonal matrix with elements $\sigma_{1:N}^2$. Here, the relation between λ and Σ is assumed as $\lambda_n^2 + \sigma_n^2 = 1$, and each element λ_n takes a value between 0 and 1. As shown in Fig. 1(b), the dynamics of $y_{n,t}$ is slow for large values of λ_n and is fast for small values of λ_n . Note that the probabilistic framework of the SFA proposed by Turner and Sahani¹²⁾ is based on a system model with first-order time-differences for each latent dynamics, and an observation model. Hereafter, we express the probabilistic framework of SFA based on the first-order difference as probabilistic SFA for simplicity.

The observed variables x_t are assumed to be expressed using the following observation model:

$$x_t = W^{-1} y_t + \epsilon_t. \quad (6)$$

Namely, as shown in black circles in Fig. 2, the observed variables x_t are assumed to be generated from the latent variables y_t , which are converted by an $M \times N$ matrix W^{-1} under observation noise ϵ_t . The observation noise ϵ_t is assumed to be white Gaussian noise with average $\mathbf{0}$ and covariance $\sigma_x^2 \mathbf{I}$, where \mathbf{I} is the identity matrix.

From Eqs. (5) and (6), the state space model of probabilistic SFA can be expressed by using probability density functions as follows:

$$p(y_t | y_{t-1}, \lambda, \Sigma) = \frac{1}{\sqrt{(2\pi)^N \det \Sigma}} \exp \left[-\frac{1}{2} (y_t - \lambda y_{t-1})^T \Sigma^{-1} (y_t - \lambda y_{t-1}) \right], \quad (7)$$

$$p(x_t | y_t, W, \sigma_x^2 \mathbf{I}) = \frac{1}{\sqrt{(2\pi \sigma_x^2)^M}} \exp \left[-\frac{1}{2 \sigma_x^2} (x_t - W^{-1} y_t)^T (x_t - W^{-1} y_t) \right]. \quad (8)$$

The condition $\lambda_n < 1$ assumed in the probabilistic version of the SFA would seem to indicate that the latent variables $y_{n,t}$ simply decay to zero with time. However, the framework of the probabilistic SFA is formulated not to suffer from this problem.¹²⁾ The mean of the hidden variable $\langle y_{n,t} | y_{n,t-1} \rangle$ with respect to the conditional distribution given the hidden variable $y_{n,t-1}$ at the preceding time $p(y_{n,t} | y_{n,t-1})$,

$$\langle y_{n,t} | y_{n,t-1} \rangle = \int y_{n,t} p(y_{n,t} | y_{n,t-1}) dy_{n,t} = \lambda_n y_{n,t-1},$$

is smaller than the value at the preceding time $y_{n,t-1}$. On the other hand, the second moment of the hidden variable $\langle y_{n,t}^2 | y_{n,t-1} \rangle$ can be larger than that at the preceding time $y_{n,t-1}^2$ as follows:

$$\langle y_{n,t}^2 | y_{n,t-1} \rangle = \int y_{n,t}^2 p(y_{n,t} | y_{n,t-1}) dy_{n,t} = \lambda_n^2 y_{n,t-1}^2 + \sigma_n^2,$$

if $y_{n,t-1}^2 < \frac{\sigma_n^2}{1-\lambda_n^2}$. This condition $y_{n,t-1}^2 < \frac{\sigma_n^2}{1-\lambda_n^2}$ is satisfied when the hidden variable takes a small value ($y_{n,t-1}^2 < 1$) by setting $\sigma_n^2 + \lambda_n^2 = 1$. This condition $\sigma_n^2 + \lambda_n^2 = 1$ is an expression assumed in probabilistic SFA. Thus, the system model [Eqs. (5) and (7)] does not simply show that hidden variables $y_{n,t}$ decay to zero but, as shown by the numerical simulations in the present study (Fig. 1) and in a previous study by Turner and Sahani,¹²⁾ also describes fluctuating responses including slowly varying features.

2.3 Rigorous derivation of marginal likelihood function by belief propagation

To estimate the latent variables \mathbf{y}_t , we need to estimate parameters in the state space model: $\boldsymbol{\theta} = \{\mathbf{W}^{-1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}, \sigma_x^2\}$. For this purpose, we rigorously derive the marginal likelihood function of probabilistic SFA by means of the belief propagation.^{16–18,21,22)}

The marginal likelihood function of the probabilistic SFA is expressed as follows:

$$\begin{aligned} p(\mathbf{x}_{1:T} | \boldsymbol{\theta}) &= \int d\mathbf{y}_{1:T} p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \boldsymbol{\theta}) \\ &= \int d\mathbf{y}_{1:T} \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{W}, \sigma_x^2 \mathbf{I}) p(\mathbf{y}_1) \\ &\quad \times \prod_{t=2}^T p(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}), \end{aligned} \quad (9)$$

where $\mathbf{x}_{1:T}$ and $\mathbf{y}_{1:T}$ denote sets of vectors $\mathbf{x}_{1:T} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$ and $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$. $p(\mathbf{y}_1)$ expresses an initial condition obeying a Gaussian distribution. To evaluate this marginal likelihood function, the previous study by the Turner and Sahani¹²⁾ employed an approximation by assuming that the observation noise σ_x^2 is zero. In their approximation, the probabilistic model of the observation model $p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{W}, \sigma_x^2 \mathbf{I})$ is assumed to be Dirac's delta function $\delta(\mathbf{x}_t - \mathbf{W}^{-1} \mathbf{y}_t)$ instead of Eq. (8) in order to simplify the integration in the marginal likelihood function. However, this approximation assumes that there exists no observation noise. A recent study showed that the estimation accuracy in the conventional probabilistic SFA is lowered by noisy data.¹⁵⁾

In this study, we derive the marginal likelihood function rigorously. Equation (9) includes the high-dimensional integration of the joint probability distribution with respect to \mathbf{y}_t . As shown in Fig. 2, the graphical model of probabilistic SFA has a straight structure and no loops. This structure enables us to overcome the difficulty of high-dimensional integration by using the belief propagation. Thus, we can perform integration with respect to each latent variable \mathbf{y}_t consecutively from time $t = 1$. Since \mathbf{y}_t depends on \mathbf{y}_{t-1} , integrations after time $t = 2$ can be performed by using the integration for the preceding time. By the belief propagation,

a marginal distribution $\alpha(\mathbf{y}_t)$ can be propagated as a *message* from time $t = 1$ to time $t = T$ using the following recursion relation:

$$c_t \alpha(\mathbf{y}_t) = p(\mathbf{x}_t | \mathbf{y}_t) \int d\mathbf{y}_{t-1} \alpha(\mathbf{y}_{t-1}) p(\mathbf{y}_t | \mathbf{y}_{t-1}), \quad (10)$$

where the marginal distribution $\alpha(\mathbf{y}_t)$ obeys a Gaussian distribution:

$$\alpha(\mathbf{y}_t) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{V}_t}} \exp \left[-\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\mu}_t)^T \mathbf{V}_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t) \right]. \quad (11)$$

Detailed derivation of Eq. (10) is given in the Appendix. The coefficient c_t here is the conditional distribution of the observation model as follows:

$$c_t = p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}), \quad (12)$$

where $p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1})$ obeys a Gaussian distribution.

Substituting the system model [Eq. (7)] into the integral in the right-hand side in Eq. (10), we obtain the following equation:

$$\begin{aligned} &\int d\mathbf{y}_{t-1} \alpha(\mathbf{y}_{t-1}) p(\mathbf{y}_t | \mathbf{y}_{t-1}) \\ &= \frac{1}{\sqrt{(2\pi)^{2N} \det \mathbf{V}_{t-1} \det \boldsymbol{\Sigma}}} \int d\mathbf{y}_{t-1} \\ &\quad \times \exp \left[-\frac{1}{2} (\mathbf{y}_{t-1} - \boldsymbol{\mu}_{t-1})^T \mathbf{V}_{t-1}^{-1} (\mathbf{y}_{t-1} - \boldsymbol{\mu}_{t-1}) \right] \\ &\quad \times \exp \left[-\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\lambda} \mathbf{y}_{t-1})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\lambda} \mathbf{y}_{t-1}) \right] \\ &= \frac{1}{\sqrt{(2\pi)^N \det \mathbf{P}_{t-1}}} \\ &\quad \times \exp \left[-\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1})^T \mathbf{P}_{t-1}^{-1} (\mathbf{y}_t - \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1}) \right], \end{aligned} \quad (13)$$

where we put $\mathbf{P}_{t-1} = \boldsymbol{\Sigma} + \boldsymbol{\lambda} \mathbf{V}_{t-1} \boldsymbol{\lambda}$. By substituting this expression into Eq. (10), we obtain the following recursion relations:

$$\boldsymbol{\mu}_t = \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1} + \mathbf{K}_t (\mathbf{x}_t - \mathbf{W}^{-1} \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1}), \quad (14)$$

$$\mathbf{V}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{W}^{-1}) \mathbf{P}_{t-1}, \quad (15)$$

$$\begin{aligned} c_t &= \frac{1}{\sqrt{(2\pi)^M \det \mathbf{Z}_{t-1}}} \\ &\quad \times \exp \left[-\frac{1}{2} (\mathbf{x}_t - \mathbf{W}^{-1} \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1})^T \mathbf{Z}_{t-1}^{-1} (\mathbf{x}_t - \mathbf{W}^{-1} \boldsymbol{\lambda} \boldsymbol{\mu}_{t-1}) \right], \end{aligned} \quad (16)$$

where \mathbf{K}_t is a Kalman gain matrix as follows:

$$\mathbf{K}_t = \mathbf{P}_{t-1} \mathbf{W}^{-1T} (\mathbf{W}^{-1} \mathbf{P}_{t-1} \mathbf{W}^{-1T} + \sigma_x^2 \mathbf{I})^{-1}, \quad (17)$$

and \mathbf{Z}_t is defined as

$$\mathbf{Z}_t = \mathbf{W}^{-1} \mathbf{P}_t \mathbf{W}^{-1T} + \sigma_x^2 \mathbf{I}. \quad (18)$$

By conducting analytical integrations consecutively using the belief propagation, we obtain the rigorously derived expression for the marginal likelihood function of the probabilistic SFA as follows:

$$p(\mathbf{x}_{1:T} | \boldsymbol{\theta}) = \prod_{t=1}^T \frac{1}{\sqrt{(2\pi)^M \det \mathbf{Z}_{t-1}}}$$

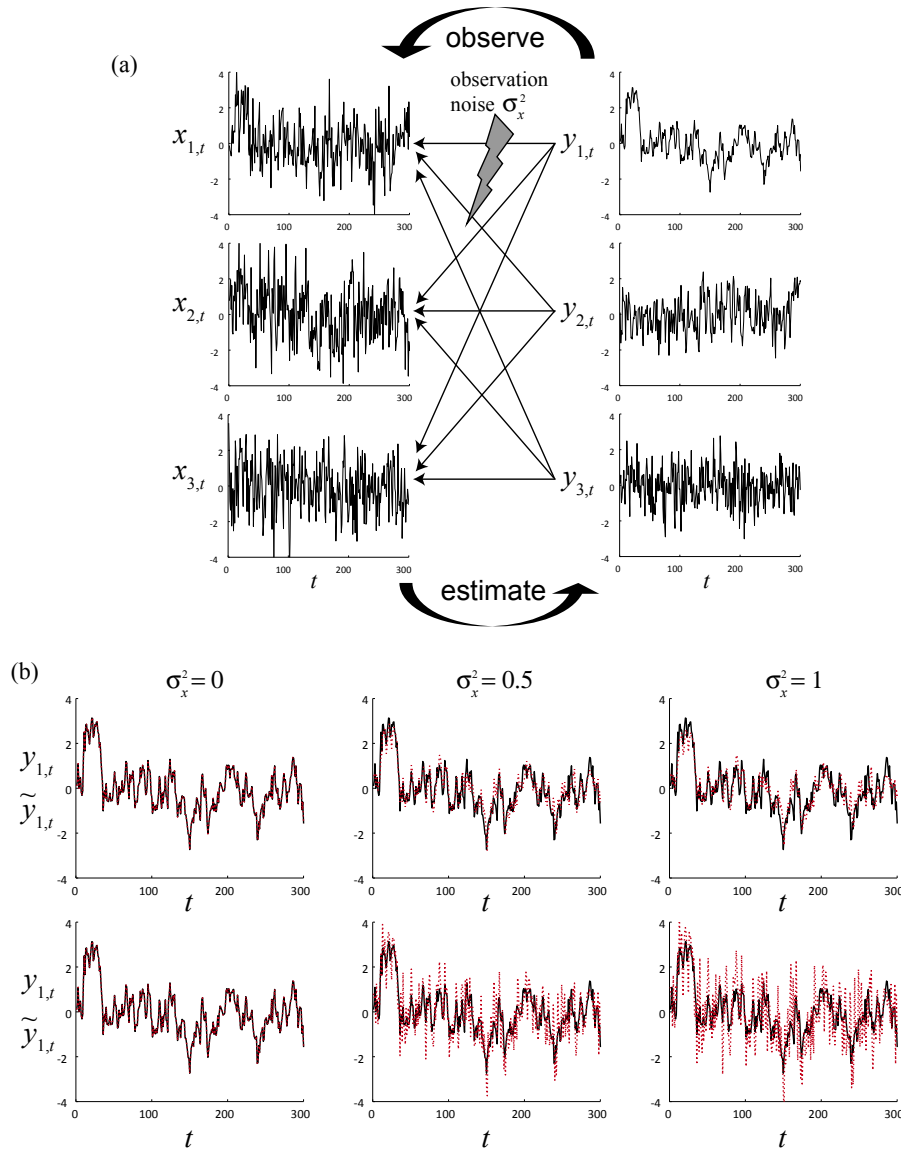


Fig. 3. (Color online) Estimation of slow feature $\tilde{y}_{1,t}$. (a) True latent variables $y_{i,t}$ ($i = 1, 2, 3$) and observed variables $x_{i,t}$ ($i = 1, 2, 3$) under noisy condition ($\sigma_x^2 = 1$). (b) Estimated slow feature by using the proposed method (top figures) and the conventional method (bottom figures). The estimated slow feature $\tilde{y}_{1,t}$ is indicated by a dotted line, whereas the true slow feature $y_{1,t}$ is indicated by a solid line. Results for different levels of observation noise σ_x^2 are shown.

$$\times \exp \left[-\frac{1}{2} (\mathbf{x}_t - \mathbf{W}^{-1} \lambda \boldsymbol{\mu}_{t-1})^T \mathbf{Z}_{t-1}^{-1} (\mathbf{x}_t - \mathbf{W}^{-1} \lambda \boldsymbol{\mu}_{t-1}) \right]. \quad (19)$$

The rigorously derived marginal likelihood function $p(\mathbf{x}_{1:T}|\boldsymbol{\theta})$ is a product of c_t . The average and covariance of each c_t can be obtained from the matrices and vectors of the preceding time $t - 1$ by means of belief propagation.

Using the rigorously derived marginal likelihood function and the state space model of the probabilistic SFA, the proposed method extracts the slow feature and its underlying parameters simultaneously. Note that the approximated marginal likelihood function used in the conventional method¹²⁾ can be obtained in the limit of $\sigma_x^2 \rightarrow 0$.

3. Results

In this section, we evaluate the validity of proposed probabilistic SFA. Both latent and observed variables are generated numerically based on the probabilistic SFA. Namely, we assume that latent variables y_t at time t are generated from those at the preceding time, and that the

observed data \mathbf{x}_t is generated through noisy observation under linear transformation of the latent variables \mathbf{y}_t .

We estimate the latent variables $\mathbf{y}_{1:T}$ and the parameters $\boldsymbol{\theta} = \{\mathbf{W}^{-1}, \lambda, \boldsymbol{\Sigma}, \sigma_x^2\}$ simultaneously from the observed data $\mathbf{x}_{1:T}$. The latent variables $\mathbf{y}_{1:T}$ are estimated using the Kalman filter, whereas the parameters $\boldsymbol{\theta}$ are estimated using the maximum likelihood method for the rigorously derived marginal likelihood function: $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}_{1:T}|\boldsymbol{\theta})$. For simplicity, the dimension of the observed variables \mathbf{x}_t is set to be the same as that of the latent variables \mathbf{y}_t ($M = N = 3$), and the total number of time steps is set to be $T = 10000$. The parameters of probabilistic SFA are set to be $\lambda_1 = 0.9$, $\lambda_2 = 0.5$, $\lambda_3 = 10^{-5}$, $\sigma_1^2 = 0.19$, $\sigma_2^2 = 0.75$, and $\sigma_3^2 = 0.99999$.

3.1 Estimation of slow feature from multi-dimensional data

Here, we extract the slow feature from noisy observed data \mathbf{x}_t by using the proposed method. We use the observed variables \mathbf{x}_t generated under noisy observation ($\sigma_x^2 = 0.5, 1$) or noiseless observation ($\sigma_x^2 = 0$). Figure 3(a) shows the

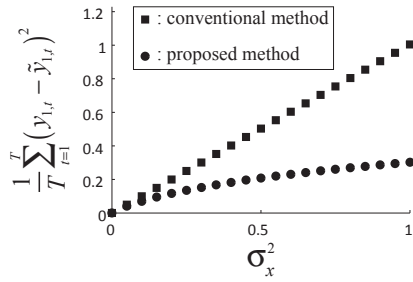


Fig. 4. Discrepancy between the true slow feature $y_{1,t}$ and the estimated slow feature $\tilde{y}_{1,t}$ evaluated for different levels of observation noise σ_x^2 . The discrepancies obtained from the proposed method and the conventional method¹²⁾ are indicated by circles and squares, respectively.

numerical data for $\sigma_x^2 = 1$. We find that each of the observed variables \mathbf{x}_t includes both slowly-varying and fast-varying elements, whereas each of the latent variables \mathbf{y}_t shows dynamics with different specific slowness. Note that the latent variable $y_{1,t}$ with the largest λ_i corresponds to the *slow feature*.

The upper graphs in Fig. 3(b) show the slow feature $\tilde{y}_{t,1}$ estimated by the proposed method. We find that, even for noisy data, the estimated slow feature $\tilde{y}_{t,1}$ (red dotted line) exhibits dynamical behaviors similar to those of the true slow feature $y_{t,1}$ (black solid line). In contrast, the slow feature estimated from noisy data by the conventional method is less similar to the true one [bottom graphs in Fig. 3(b)]. These results show that the proposed probabilistic SFA with a rigorously derived marginal likelihood function gives better performance than the previous probabilistic SFA.¹²⁾

3.2 Effect of observation noise on performance

To evaluate the effect of observation noise on estimation performance quantitatively, we perform estimation for different levels of observation noise. Figure 4 shows how the discrepancy between the estimated and the true slow feature changes with the observation noise. We find that, for finite values of observation noise, the discrepancy between the estimated and the true slow feature in the proposed method is smaller than that in the conventional method.¹²⁾ This result shows that our proposed method gives better performance than the previous method; even though the results of the two methods are similar when there exists no observation noise, the estimation errors for the proposed method are much smaller than those for the previous method when there exists observation noise. From these results, we find that our proposed method using probabilistic SFA with the rigorously derived marginal likelihood function extracts the slow feature more accurately.

We also investigate the estimation accuracy of underlying parameters. The dependence of estimated parameters $\tilde{\lambda}_i$ ($i = 1, 2, 3$) on observation noise is shown in Fig. 5. We see that with the proposed method, the estimated parameters are similar to the true ones even for a noisy environment. With the conventional method, however, the discrepancy between the estimated and true parameters increases as the observation noise increases. These results show that the proposed probabilistic SFA with a rigorously derived marginal likelihood function accurately estimates the slow feature and its underlying parameters.

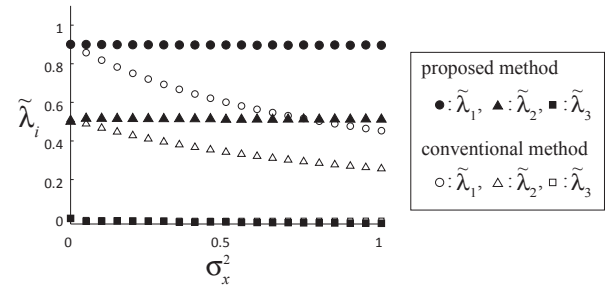


Fig. 5. Dependence of estimated parameters $\tilde{\lambda}_i$ ($i = 1, 2, 3$) on the variance of observation noise σ_x^2 . Estimated parameters in the proposed method are indicated by filled marks whereas those in the conventional method¹²⁾ are indicated by open marks. For both methods, estimated values of $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, and $\tilde{\lambda}_3$ are indicated by circles, triangles, and squares, respectively. Note that the true parameter values are $\lambda_1 = 0.9$, $\lambda_2 = 0.5$, $\lambda_3 = 10^{-5}$. We find that parameters are accurately estimated even in noisy case in the proposed method whereas the discrepancy between the true and estimated parameters increases with the observation noise in the conventional method.

4. Concluding Remarks

In this paper, we have proposed probabilistic SFA with a rigorously derived marginal likelihood function for extracting the latent slow feature from multi-dimensional data. The marginal likelihood function in the probabilistic SFA has been derived rigorously by means of belief propagation in the proposed method, whereas the marginal likelihood function was approximately evaluated in the conventional probabilistic SFA by assuming that there is no observation noise. Furthermore, we have shown using numerical data that our proposed method can estimate the slow feature and its underlying parameters for latent dynamics accurately even under noisy conditions.

The conventional and proposed probabilistic frameworks of the SFA assume a simple system model described by first-order autoregressive models with some constraints. This assumption may cause inaccurate estimation in the case of latent dynamics including complex oscillations. To overcome this limitation, it would be important to extend the probabilistic frameworks of the SFA to a framework with system model including an orthogonal matrix with off-diagonal elements and to consider more general dynamical systems including higher-order autoregressive models. We leave this extension as a future work.

Acknowledgments This work is partially supported by Grants-in-Aid for Scientific Research for Innovative Areas “Initiative for High-Dimensional Data-driven Science through Deepening of Sparse Modeling” [JSPS KAKENHI Grants No. JP25120010 (T.O.) and No. JP25120009 (M.O.)] and for Scientific Research [JSPS KAKENHI Grant No. JP16K00330 (T.O.)], and a Fund for the Promotion of Joint International Research (Fostering Joint International Research [JSPS KAKENHI Grant No. JP15KK0010 (T.O.)] from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Appendix

In this Appendix, we derive the recursion relation [Eq. (10)] in belief propagation.^{16,17,22)} A joint probability density function $p(\mathbf{x}_{1:t}, \mathbf{y}_t)$ can be expressed as

$$\begin{aligned} p(\mathbf{x}_{1:t}, \mathbf{y}_t) &= p(\mathbf{x}_{1:t} | \mathbf{y}_t) p(\mathbf{y}_t) \\ &= p(\mathbf{x}_{1:t-1} | \mathbf{y}_t) p(\mathbf{x}_t | \mathbf{y}_t) p(\mathbf{y}_t) \\ &= p(\mathbf{x}_{1:t-1}, \mathbf{y}_t) p(\mathbf{x}_t | \mathbf{y}_t). \end{aligned} \quad (\text{A} \cdot 1)$$

By using an integral representation of $p(\mathbf{x}_{1:t-1}, \mathbf{y}_t)$ in Eq. (A.1), the joint probability density function at time t , $p(\mathbf{x}_{1:t}, \mathbf{y}_t)$, can be expressed using that at time $t-1$, $p(\mathbf{x}_{1:t-1}, \mathbf{y}_{t-1})$, as follows:

$$\begin{aligned} p(\mathbf{x}_{1:t}, \mathbf{y}_t) &= p(\mathbf{x}_t | \mathbf{y}_t) \int d\mathbf{y}_{t-1} p(\mathbf{x}_{1:t-1}, \mathbf{y}_{t-1}, \mathbf{y}_t) \\ &= p(\mathbf{x}_t | \mathbf{y}_t) \int d\mathbf{y}_{t-1} p(\mathbf{x}_{1:t-1} | \mathbf{y}_{t-1}) p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}) \\ &= p(\mathbf{x}_t | \mathbf{y}_t) \int d\mathbf{y}_{t-1} p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{x}_{1:t-1}, \mathbf{y}_{t-1}). \quad (\text{A} \cdot 2) \end{aligned}$$

By substituting $\alpha(\mathbf{y}_t) = p(\mathbf{y}_t | \mathbf{x}_{1:t}) = p(\mathbf{x}_{1:t}, \mathbf{y}_t) / p(\mathbf{x}_{1:t})$ into Eq. (A.2), we obtain the following equation:

$$\begin{aligned} p(\mathbf{x}_{1:t}) \alpha(\mathbf{y}_t) &= p(\mathbf{x}_t | \mathbf{y}_t) p(\mathbf{x}_{1:t-1}) \int d\mathbf{y}_{t-1} p(\mathbf{y}_t | \mathbf{y}_{t-1}) \alpha(\mathbf{y}_{t-1}). \quad (\text{A} \cdot 3) \end{aligned}$$

By putting $c_t = p(\mathbf{x}_t | \mathbf{x}_{1:t-1})$, we obtain the following recursion relation shown in Eq. (10):

$$c_t \alpha(\mathbf{y}_t) = p(\mathbf{x}_t | \mathbf{y}_t) \int d\mathbf{y}_{t-1} p(\mathbf{y}_t | \mathbf{y}_{t-1}) \alpha(\mathbf{y}_{t-1}). \quad (\text{A} \cdot 4)$$

*omori@eedept.kobe-u.ac.jp

- 1) L. Wiskott and T. J. Sejnowski, *Neural Comput.* **14**, 715 (2002).
- 2) P. Berkes and L. Wiskott, *J. Vision* **5**, 9 (2005).
- 3) M. Franzius, H. Sprekeler, and L. Wiskott, *PLOS Comput. Biol.* **3**, e166 (2007).

- 4) H. Sprekeler and L. Wiskott, *Neural Comput.* **23**, 303 (2011).
- 5) P. Berkes, CogPrints 4104 (2005).
- 6) M. Franzius, N. Wilbert, and L. Wiskott, *Lect. Notes Comput. Sci.* **5163**, 961 (2008).
- 7) R. Legenstein, N. Wilbert, and L. Wiskott, *PLOS Comput. Biol.* **6**, e1000894 (2010).
- 8) Y. P. Huang, J. L. Zhao, Y. H. Liu, S. W. Luo, Q. Zou, and M. Tian, *Inf. Sci.* **181**, 3284 (2011).
- 9) H. Sprekeler, T. Zito, and L. Wiskott, *J. Mach. Learn. Res.* **15**, 921 (2014).
- 10) J. P. Cunningham and Z. Ghahramani, *J. Mach. Learn. Res.* **16**, 2859 (2015).
- 11) V. R. Kompella, M. Luciw, M. F. Stollenga, and J. Schmidhuber, *Neural Comput.* **28**, 1599 (2016).
- 12) R. Turner and M. Sahani, *Neural Comput.* **19**, 1022 (2007).
- 13) M. E. Tipping and C. M. Bishop, *J. R. Stat. Soc.* **61**, 611 (1999).
- 14) C. F. Beckmann and S. M. Smith, *IEEE Trans. Med. Imaging* **23**, 137 (2004).
- 15) T. Sekiguchi, T. Omori, and M. Okada, *IPSJ Trans. Math. Model. Appl.* **5**, 26 (2012) [in Japanese].
- 16) J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, Burlington, MA, 1988).
- 17) K. Tanaka, *J. Phys. A* **35**, R81 (2002).
- 18) K. Tanaka, M. Yasuda, and D. M. Titterton, *J. Phys. Soc. Jpn.* **81**, 114802 (2012).
- 19) J. S. Yedidia, W. T. Freeman, and Y. Weiss, in *Exploring Artificial Intelligence in the New Millenium*, ed. G. Lakemeyer and B. Nebel (Morgan Kaufmann, Burlington, MA, 2002) p. 239.
- 20) R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).
- 21) M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).
- 22) C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Heidelberg, 2006).