

Switching Probabilistic Slow Feature Analysis for Time Series Data

Kazuki Tsujimoto and Toshiaki Omori

Abstract—Slow feature analysis (SFA) is a machine learning method for extracting slowly time-varying feature from multi-dimensional time series data. Recently, probabilistic SFA (PSFA) that extends SFA to a probabilistic framework has been proposed. The PSFA can be applied to stationary time series data with noise and missing values. In order to deal with non-stationary time series data including change points, we propose a switching probabilistic slow feature analysis (switching PSFA) in this paper. By introducing a switching state space model, it is possible to extract slowly varying information even when system parameters change with time. Using the proposed method, we show that slowly time-varying components can be extracted more accurately from time-series data with non-stationarity.

Index Terms—Slow feature analysis, switching state space model, Bayesian time series analysis, statistical machine learning.

I. INTRODUCTION

A large amount of multi-dimensional time-series data such as image data and sensor data has been handled, with the development of information technology and observation technology. In recent years, it has been required to make use of such large amount of data in order to discover useful knowledge [1]-[3].

Various machine learning studies have been conducted on feature extraction methods for high-dimensional time-series data, and slow feature analysis (SFA) has recently attracted attention as one of them, which was originally proposed from a viewpoint of human recognition system [4]. When a human acquires various visual information such as the position and shape of an object, the visual information is acquired from a large amount of retinal cells. Individual retinal signals are sensitive to slight environmental changes. However, higher sensations such as the position and shape of objects change very slowly compared to them. In this way, it is known in neuroscience that slowly changing components in visual information have an important role in the recognition of

objects and space [4]-[6]. The SFA is an unsupervised feature extraction method developed based on the knowledge of visual information, and is an algorithm for extracting slowly time-varying feature from an input multi-dimensional data. SFA is a model proposed in the field of neuroscience, but it is also applied in the field of machine learning such as pattern recognition and information extraction [7]-[15]. Moreover, Turner and Sahani proposed a probabilistic SFA by considering constraints to the state space model [6]. While deterministic SFA is difficult to estimate for data that includes noise, PSFA has the advantage that observation noise can be considered. There are also discussions on the effects of observation noise on PSFA [16], [17].

Conventionally, PSFA has been applied to multi-dimensional time-series data that does not include change points. However, in time series data, the behavior of observed values may change suddenly from a certain time, and it is important to detect changes behind the time series [18]. For example, in sensor data related to body movements, it is required to detect actions from time-series data in which multiple actions are continuously mixed [19].

In this study, we propose a switching PSFA in order to extract hidden slow feature from non-stationary multi-dimensional time series data. We formulate switching PSFA by using a switching state space model, and find a change point for non-steady multi-dimensional time series data so that PSFA can be performed. The rest of this paper is organized as follows. In Section II, we describe conventional SFA. In Section III, we propose switching probabilistic slow feature analysis. We formulate probabilistic model of switching PSFA, and derive variational learning framework. In Section IV, the proposed method is evaluated using simulated data. Concluding remarks are given in Section V.

II. EXISTING METHODS

In this section, we first explain conventional frameworks of slow feature analysis (SFA): deterministic SFA and probabilistic SFA. Next, we describe parameter estimation method for probabilistic SFA using EM algorithm.

A. Deterministic SFA

Deterministic SFA is an unsupervised algorithm for extracting slowly time-varying features from multi-dimensional time series data [4], [20]. Schematic of SFA is shown in Fig. 1. Let $\mathbf{x}(t)$ be the given input observation multi-dimensional time-series data. The output $\mathbf{y}(t)$ extracted by the SFA consists of $y_j(t) = g_j(\mathbf{x}(t))$ using the transformation $g_j(\cdot)$. The transformation $g_j(\mathbf{x})$ minimizes the square of the time derivative

Manuscript received January 21, 2020; revised July 14, 2020. 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 Grant No. JP25120010] and for Scientific Research [JSPS KAKENHI Grant No. JP16K00330], and a Fund for the Promotion of Joint International Research (Fostering Joint International Research) [JSPS KAKENHI Grant No. JP15KK0010] from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and Core Research for Evolutional Science and Technology (CREST) [Grant No. JPMJCR1914], Japan Science and Technology Agency, Japan.

The authors are with the Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501 Japan (e-mail: omori@eedept.kobe-u.ac.jp).

$$\Delta(y_j) := \langle y_j^2 \rangle_t = \left\langle \left(\frac{d}{dt} g_j(\mathbf{x}(t)) \right)^2 \right\rangle_t \quad (1)$$

where $\langle \cdot \rangle_t$ represents the time average. An element of output $y_j(t) = g_j(\mathbf{x}(t))$ that minimizes $\Delta(\mathbf{y})$ is called slow feature. Besides, the following constraint conditions are added to the formula,

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

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

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

Here, (2) and (3) are used to normalize all outputs, and to avoid obvious solutions such as the output is zero. (4) is used to extract different information by making the outputs uncorrelated. In the following, we assume that the transformation $\mathbf{g}(\cdot)$ is a linear transformation and is expressed as $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$ using the matrix \mathbf{W} .

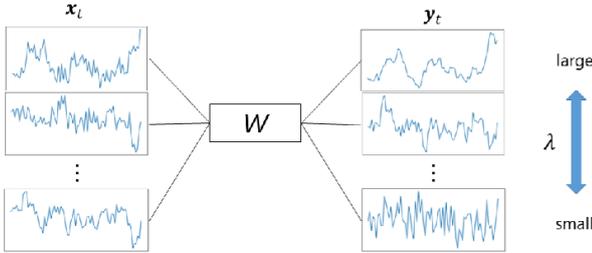


Fig. 1. Schematic of SFA.

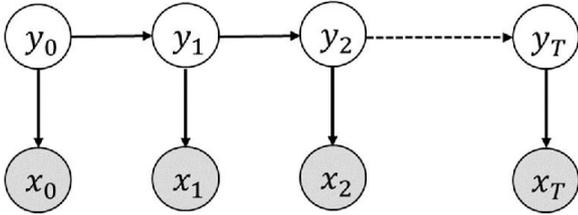


Fig. 2. Graphical model of probabilistic SFA (conventional method).

B. Probabilistic SFA

In recent studies, SFA with probabilistic framework was proposed by adding constraints to the system model of linear Gaussian state space model [6]. We assume that $\mathbf{y}(t)$ is the N -dimensional hidden variables including the slow feature and $\mathbf{x}(t)$ is the D -dimensional observable variable of the state space model. Fig. 2 shows the graphical model of probabilistic SFA. The state space model of PSFA is expressed by the following system model and observation model respectively,

$$\mathbf{y}_t = \boldsymbol{\lambda} \mathbf{y}_{t-1} + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (4)$$

$$\mathbf{x}_t = \mathbf{W}^{-1} \mathbf{y}_t + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma_x^2 \mathbf{I}). \quad (5)$$

In the system model as (5), the latent variable \mathbf{y}_t depends on \mathbf{y}_{t-1} , and $\boldsymbol{\lambda}$ is a parameter matrix that determines the degree of dependence. $\boldsymbol{\lambda}$ is an $N \times N$ diagonal matrix

consisting of values λ_n for each $y_{n,t}$. For system noise, the covariance matrix $\boldsymbol{\Sigma}$ is $N \times N$ diagonal matrix with diagonal elements σ_n^2 . In the observation model as (6), \mathbf{W}^{-1} is $D \times N$ matrix and \mathbf{I} is $D \times D$ unit matrix.

Here, the following restrictions are assumed in order to incorporate the properties of SFA,

$$\lambda_n^2 + \sigma_n^2 = 1. \quad (7)$$

If λ_n is large, corresponding latent variable $y_{n,t}$ depends strongly on the value of the previous time, and noise $\eta_{n,t}$ becomes small, thus $y_{n,t}$ changes slowly. On the other hand, if λ_n is small, it becomes less dependent on the value of the previous time, and noise $\eta_{n,t}$ becomes large, therefore $y_{n,t}$ changes quickly.

From (5) and (6), the state space model can be expressed by using the probability density function as follows:

$$p(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{y}_t | \boldsymbol{\lambda} \mathbf{y}_{t-1}, \boldsymbol{\Sigma}), \quad (8)$$

$$p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{W}, \sigma_x^2) = \mathcal{N}(\mathbf{x}_t | \mathbf{W}^{-1} \mathbf{y}_t, \sigma_x^2 \mathbf{I}). \quad (9)$$

The likelihood function for estimating the parameters $\boldsymbol{\theta} = \{\boldsymbol{\lambda}, \mathbf{W}^{-1}, \boldsymbol{\Sigma}, \sigma_x^2\}$ of PSFA is expressed as follows:

$$p(\mathbf{x}_{1:T} | \boldsymbol{\theta}) = \int d\mathbf{y}_{1:T} \left(\prod_{t=1}^T p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{W}, \sigma_x^2) \right) \times \prod_{n=1}^N \prod_{t=1}^T p(y_{n,t} | y_{n,t-1}, \lambda_n, \sigma_n^2) p(y_{n,1} | \sigma_{n,1}^2). \quad (10)$$

C. Parameter Estimation Using EM Algorithm

In the PSFA proposed by Turner and Sahani [6], parameters are estimated assuming that the observation noise is zero. However, recent research has shown that observation noise affects the accuracy of slow feature estimation. In this section, we describe a PSFA proposed by Takeuchi and Omori [21]. It introduces an EM algorithm so that all parameters including observation noise can be estimated. The EM algorithm is a maximum likelihood estimation algorithm devised by Dempster *et al.* [22]. When it is difficult to apply the maximum likelihood estimation, the estimated maximum likelihood is calculated for incomplete data using the likelihood assuming that complete data is obtained [23]. The EM algorithm guarantees that a local optimal solution can be obtained by alternately repeating the E (Expectation) step and the M (Maximization) step. In the E step, the expected value of the log likelihood of the complete data $p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \boldsymbol{\theta})$ is calculated based on the distribution of latent variables currently estimated using the Kalman smoother,

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)}) = E_{\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}^{(k)}} [\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \boldsymbol{\theta})]. \quad (11)$$

Here, Q is the expected value of the log likelihood of the complete data in the PSFA model, k is the number of steps in the EM algorithm, and the log likelihood is represented as

$$\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T} | \boldsymbol{\theta}) = \sum_{t=1}^T \log p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{W}, \sigma_x^2) + \log p(\mathbf{y}_1 | \boldsymbol{\Sigma}_1) + \sum_{t=2}^T \log p(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\lambda}). \quad (12)$$

In the M step, the expected value of the log likelihood

obtained in the E step is maximized for the parameter θ ,

$$\theta^{(k+1)} = \operatorname{argmax}_{\theta} Q(\theta | \theta^{(k)}). \quad (13)$$

The EM algorithm repeats the E and M steps up to a predetermined number of times or until the log likelihood converges. However, since the parameter estimated by the EM algorithm is a local optimal solution, it may differ from the true value depending on the initial parameter $\theta^{(1)}$.

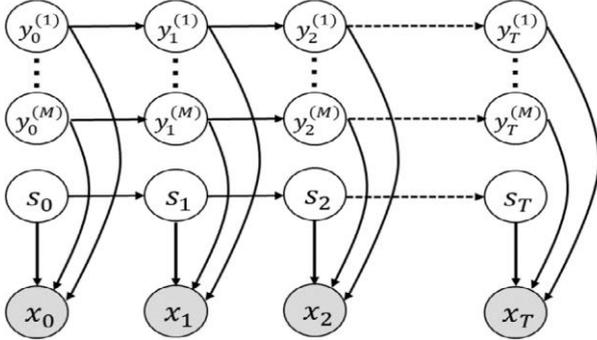


Fig. 3. Graphical model of the switching probabilistic slow feature analysis (proposed method).

III. PROPOSED METHOD: SWITCHING PROBABILISTIC SLOW FEATURE ANALYSIS

In this section, we formulate switching PSFA by applying the concept of switching state space model to the PSFA. In order to extract slowly time-varying features even for time series data including change points, we estimate latent variables and parameters using the variational Bayesian method. We derive variational Bayesian learning framework for switching PSFA.

A. Switching Framework for Probabilistic Slow Feature Analysis

The conventional PSFA assumes one system model, since it does not assume non-steady systems. In the proposed method, by introducing a switching state space model [24]-[27] into PSFA, it was possible to analyze time series data including change points. As shown in Fig. 3, switching state space model (SSSM) prepares multiple state variables $\mathbf{y}_t^{(m)}$ ($1 \leq m \leq M$) and switch variables $\mathbf{s}_t \in \{0,1\}^M$, and the observation value \mathbf{x}_t is generated from the latent variables of the model m selected by the switch variables \mathbf{s}_t . \mathbf{s}_t is an M -dimensional vector having 1 in one element and 0 in the other elements. The latent variable $\mathbf{y}_t^{(m)}$ and the switch variable \mathbf{s}_t have a Markov property that depends on the previous value ($\mathbf{y}_{t-1}^{(m)}, \mathbf{s}_{t-1}$), and SSSM is a combination of a hidden Markov model (HMM) and a state space model (SSM). The system model and observation model of SSSM are expressed as follows:

$$\mathbf{y}_t^{(m)} = \lambda^{(m)} \mathbf{y}_{t-1}^{(m)} + \eta_t^{(m)} \quad (m = 1, \dots, M), \quad (14)$$

$$\mathbf{x}_t = \mathbf{W}^{-1(m)} \mathbf{y}_t^{(m)} + \epsilon_t^{(m)} \quad (m = 1, \dots, M). \quad (15)$$

For (15), if the observation \mathbf{x}_t is output from the $\mathbf{y}_t^{(m)}$, \mathbf{s}_t is a one-hot vector with unity in the m -th element. The model

for switch variables is as follows:

$$p(\mathbf{s}_t^{(j)} = 1 | \mathbf{s}_{t-1}^{(i)} = 1) = \phi_{j,i}. \quad (16)$$

This is the probability of transition from state $\mathbf{s}_{t-1}^{(i)} = 1$ at time $(t-1)$ to state $\mathbf{s}_t^{(j)} = 1$ at time t . This is each element of the transition matrix of $M \times M$.

The joint probability of the proposed switching PSFA is given by

$$\begin{aligned} & p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}^{(1)}, \dots, \mathbf{y}_{1:T}^{(M)}, \mathbf{x}_{1:T}) \\ &= p(\mathbf{s}_1) \prod_{t=2}^T p(\mathbf{s}_t | \mathbf{s}_{t-1}) \prod_{m=1}^M p(\mathbf{y}_1^{(m)}) \prod_{t=2}^T p(\mathbf{y}_t^{(m)} | \mathbf{y}_{t-1}^{(m)}) \\ & \times \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{y}_t^{(1)}, \dots, \mathbf{y}_t^{(M)}, \mathbf{s}_t), \end{aligned} \quad (17)$$

where $p(\mathbf{y}_t^{(m)} | \mathbf{y}_{t-1}^{(m)})$ is derived from (14). The log-likelihood is expressed by

$$\begin{aligned} & \log p(\mathbf{x}_t | \mathbf{y}_t^{(1)}, \dots, \mathbf{y}_t^{(M)}, \mathbf{s}_t^{(m)} = 1) \\ &= -\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{R}^{(m)}| \\ & - \frac{1}{2} (\mathbf{x}_t - \mathbf{W}^{-1(m)} \mathbf{y}_t^{(m)})^T \mathbf{R}^{(m)-1} (\mathbf{x}_t - \mathbf{W}^{-1(m)} \mathbf{y}_t^{(m)}). \end{aligned} \quad (18)$$

Here, D represents the number of dimensions of the observation and $\mathbf{R}^{(m)}$ is the covariance matrix of the observation noise. As shown in (18), when the switch variable \mathbf{s}_t has 1 in the m -th element ($\mathbf{s}_t^{(m)} = 1$), the likelihood is expressed by an observation model for m -th SSM.

B. Variational Inference

In this study, we derive learning algorithm for parameters of switching probabilistic slow feature analysis by generalizing the EM algorithm [22]. The log-likelihood of the observation data to be maximized is expressed by the following equation:

$$\begin{aligned} \log p(\mathbf{x}_{1:T} | \theta) &= \log \sum_{\mathbf{s}_{1:T}} \int p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}, \mathbf{x}_{1:T} | \theta) d\mathbf{y}_{1:T} \\ &= \log \sum_{\mathbf{s}_{1:T}} \int Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}) \left[\frac{p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}, \mathbf{x}_{1:T} | \theta)}{Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})} \right] d\mathbf{y}_{1:T} \\ &\geq \sum_{\mathbf{s}_{1:T}} \int Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}) \log \left[\frac{p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}, \mathbf{x}_{1:T} | \theta)}{Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})} \right] d\mathbf{y}_{1:T}, \end{aligned} \quad (19)$$

where we used Jensen's inequality. The purpose here is to maximize the lower bound of (19). In the E step, the posterior probability of $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}) = p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_{1:T}, \theta)$ is obtained using the current parameter estimates. However, it is quite difficult to find the exact posterior probability in SSSM. Therefore, we employ an estimation method using the variational Bayes method.

Here we derive a method for estimating parameters in the switching PSFA by means of variational framework. In SSSM, it is difficult to obtain the posterior probability distribution $p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_{1:T})$ precisely. Therefore, we approximate the posterior distribution by variational distribution $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})$ based on the mean field

approximation. In E-step, we estimate $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})$ that minimizes the Kullback-Leibler (KL) divergence:

$$\text{KL}(Q||P) = \sum_{\mathbf{s}_{1:T}} \int Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}) \log \left[\frac{Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})}{p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_{1:T})} \right] d\mathbf{y}_{1:T} \quad (20)$$

Here, (20) is a metric of the distance between true distribution $p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_{1:T})$ and approximated distribution $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})$, and we aim to minimize (20) in E-step. The distribution $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})$ that approximates the posterior distribution is separated as follows:

$$Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T}) = \frac{1}{Z_Q} \left[\psi(\mathbf{s}_1) \prod_{t=2}^T \psi(\mathbf{s}_{t-1}, \mathbf{s}_t) \right] \prod_{m=1}^M \left[\psi(\mathbf{y}_1^{(m)}) \prod_{t=2}^T \psi(\mathbf{y}_{t-1}^{(m)}, \mathbf{y}_t^{(m)}) \right], \quad (21)$$

where ψ is defined as

$$\psi(\mathbf{s}_1^{(m)} = 1) = p(\mathbf{s}_1^{(m)} = 1) q_1^{(m)}, \quad (22)$$

$$\psi(\mathbf{s}_{t-1}, \mathbf{s}_t^{(m)} = 1) = p(\mathbf{s}_t^{(m)} = 1 | \mathbf{s}_{t-1}) q_1^{(m)}, \quad (23)$$

$$\psi(\mathbf{y}_1^{(m)}) = p(\mathbf{y}_1^{(m)}) [p(\mathbf{x}_1 | \mathbf{y}_1^{(m)}, \mathbf{s}_1^{(m)} = 1)]^{h_1^{(m)}}, \quad (24)$$

$$\psi(\mathbf{y}_{t-1}^{(m)}, \mathbf{y}_t^{(m)}) = p(\mathbf{y}_t^{(m)} | \mathbf{y}_{t-1}^{(m)}) [p(\mathbf{x}_t | \mathbf{y}_t^{(m)}, \mathbf{s}_t^{(m)} = 1)]^{h_t^{(m)}}. \quad (25)$$

Here, Z_Q is a normalization factor. The KL divergence of the posterior distribution $p(\mathbf{s}_{1:T}, \mathbf{y}_{1:T} | \mathbf{x}_{1:T})$ and its approximate $Q(\mathbf{s}_{1:T}, \mathbf{y}_{1:T})$ is minimized by repeatedly updating the following equations for $h_t^{(m)}$ and $q_t^{(m)}$:

$$h_t^{(m)} = \langle s_t^{(m)} \rangle, \quad (26)$$

$$\log q_t^{(m)} = -\frac{1}{2} \left\langle \left(\mathbf{x}_t - \mathbf{W}^{-1(m)} \mathbf{y}_t^{(m)} \right)^T \mathbf{R}^{(m)-1} \left(\mathbf{x}_t - \mathbf{W}^{-1(m)} \mathbf{y}_t^{(m)} \right) \right\rangle, \quad (27)$$

where brackets $\langle \cdot \rangle$ represent an expectation. $\langle s_t^{(m)} \rangle$ is the probability that the m -th element of the switch variable \mathbf{s}_t is 1 and is calculated by the forward-backward algorithm [28] using $q_t^{(m)}$. Here, $q_t^{(m)}$ is obtained from the value obtained by using Kalman smoother [29]-[31] on the data weighted by $h_t^{(m)}$. In this way, the E-step aims to minimize the KL divergence.

In M step, the expected value of log likelihood is partially differentiated for each parameter and the parameter is updated. The update formula derived for parameters in the proposed switching PSFA is as follows:

$$\mathbf{W}^{-1(m)} = \frac{\sum_{t=1}^T \langle s_t^{(m)} \rangle \mathbf{x}_t \langle \mathbf{y}_t^{(m)T} \rangle}{\sum_{t=1}^T \langle s_t^{(m)} \rangle \langle \mathbf{y}_t^{(m)} \mathbf{y}_t^{(m)T} \rangle}, \quad (28)$$

$$\mathbf{R} = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T \langle s_t^{(m)} \rangle \left(\langle \mathbf{y}_t^{(m)} \mathbf{y}_t^{(m)T} \rangle - \mathbf{W}^{-1(m)} \mathbf{x}_t \langle \mathbf{y}_t^{(m)T} \rangle \right), \quad (29)$$

$$\begin{aligned} \mathcal{T}(\lambda_n^{(m)})^3 - \sum_t \langle \mathbf{y}_{n,t}^{(m)} \mathbf{y}_{n,t-1}^{(m)} \rangle (\lambda_n^{(m)})^2 \\ + \sum_t \left(\langle \mathbf{y}_{n,t}^{(m)2} \rangle + \langle \mathbf{y}_{n,t-1}^{(m)2} \rangle \right) \\ - 1) \lambda_n^{(m)} - \sum_t \langle \mathbf{y}_{n,t}^{(m)} \mathbf{y}_{n,t-1}^{(m)} \rangle = 0. \end{aligned} \quad (30)$$

As for (30), it satisfies the condition $0 < \lambda_n^{(m)} < 1$ in the solution of the cubic equation. Note that covariance matrix of the observation noise for each m can also be derived by differentiation of the expected log-likelihood. Algorithm 1 summarizes the algorithm of the proposed method.

Algorithm 1: Switching probabilistic SFA

Input: Observation data \mathbf{x} , Iteration limit of EM algorithm K , Initial values of parameters for each model $\theta^{(m)} = \{\lambda^{(m)}, \mathbf{W}^{-1(m)}, \mathbf{R}^{(m)}\}$, and Initial value of transition probability ϕ
Output: Estimated parameters $\theta^{(m)}, \phi$

- 1: **while** Repeat until the lower log likelihood converges.
 - 2: (E-step)
 - 3: **while** Repeat until KL divergence converges.
 - 4: Calculate the expected value $E[\mathbf{s}_t]$ using forward-backward algorithm, and update $h_t^{(m)}$.
 - 5: Calculate $E[\mathbf{y}_t^{(m)}]$, $E[\mathbf{y}_t^{(m)} \mathbf{y}_t^{(m)T}]$, $E[\mathbf{y}_t^{(m)} \mathbf{y}_{t-1}^{(m)T}]$ using Kalman smoother, and update $q_t^{(m)}$.
 - 6: **end while**
 - 7: (M-step)
 - 8: Update parameters $\theta^{(m)}$
 - 9: **end while**
-

IV. EXPERIMENT

A. Experimental Settings

In this section, in order to show the effectiveness of the proposed switching PSFA, we estimate parameters using simulated data and evaluate latent variables compared to conventional PSFA. The observation data including the change points were synthesized from three different PSFA models. The number of data points is $T = 1500$; we assume that the entire data consists of 500 data points generated from each model ($m = 1, 2$ and 3). The latent variable was set to 3 dimensions, and the observation data was set to 6 dimensions. The parameter $\lambda^{(m)}$ of switching PSFA has a different value in each dimension, $\mathbf{W}^{(m)}$ was set to different values for different m . Covariance matrix of the observation noise \mathbf{R} is common. We use switching PSFA to estimate latent variable \mathbf{y}_t and switch variable \mathbf{s}_t with Kalman smoother and forward-backward algorithm in E step. We estimate the parameters such as $\theta^{(m)} = \{\lambda^{(m)}, \mathbf{W}^{-1(m)}, \mathbf{R}\}$ and state transition rates with M-step. State transition rates are estimated as conventional switching state space model framework [24]. We repeat E-step and M-step alternately until log likelihood has converged.

TABLE I: MEAN SQUARE ERROR BETWEEN TRUE AND ESTIMATED VALUES OF LATENT VARIABLES

Proposed method	Conventional method
0.03407	0.13226

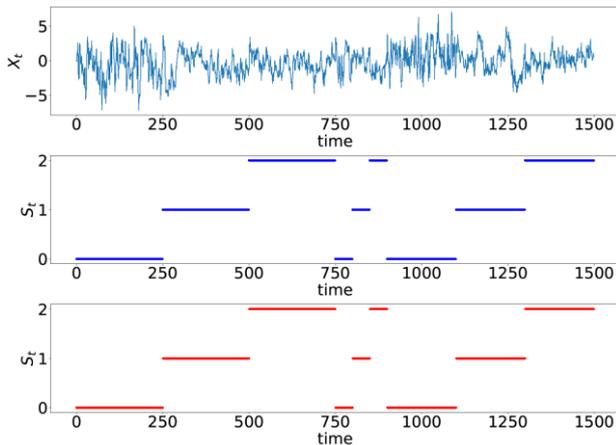


Fig. 4. Estimation of switch variables. Upper: one dimension $x_{n,t}$ of the six-dimensional observation data x_t . Middle: true switch variable s_t . Lower: estimated switch variable \hat{s}_t .

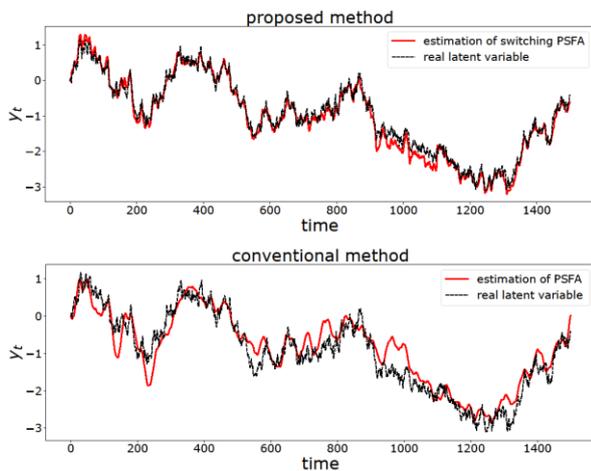


Fig. 5. A comparison of estimated latent variables y_t between proposed method and conventional method. True and estimated latent variables y_t are represented by black dotted line and red solid lines, respectively. Upper: Slow feature estimated by conventional PSFA. Lower: Slow feature estimated by switching PSFA.

B. Result

Here, we perform estimation of hidden variables y_t using the proposed switching PSFA. Fig. 4 shows the result for estimating the switch variable s_t by the forward-backward algorithms. It was shown that the switching of observation data model could be estimated by using switching PSFA from the state even when no information on change points or true parameters was given.

Fig. 5 shows a comparison between the latent value y_t obtained from the same observation data x_t using the conventional PSFA and switching PSFA and the true value. Here, latent variables y_t are calculated in E-step using a Kalman smoother. In Fig. 5, we find that the proposed method extracts the slow feature more accurately; the discrepancy between true and estimated slow feature is large in the existing PSFA compared with the proposed switching PSFA. Estimation accuracy is compared quantitatively between conventional PSFA and proposed switching PSFA by using mean squared error between true and estimated values of latent variables. Table I shows that our proposed switching PSFA extract latent variables more accurately compared with conventional PSFA. Since the conventional PSFA does not assume that the model is switched, it is impossible to estimate

latent variables from observation data including change points. However, switching PSFA was able to estimate the latent variables well. It was also confirmed that the estimated value converged to the true value as each parameter was updated.

V. CONCLUSION

In this paper, the framework of the switching state space model is introduced to the probabilistic SFA, and the parameter update formula is derived using the EM algorithm and variational Bayesian method. It was also shown that latent variables, switch variables, and parameters can be estimated for time-series data including change points.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

TO designed research; KT and TO performed research; KT and TO analyzed the data; KT and TO wrote the paper; all authors had approved the final version.

REFERENCES

- [1] W. Fan and A. Bifet, "Mining big data: Current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*, vol. 14, pp. 1-5, 2012.
- [2] T. Omori, T. Kuwatani, A. Okamoto, and K. Hukushima, "Bayesian inversion analysis of nonlinear dynamics in surface heterogeneous reactions," *Physical Review E*, vol. 94, pp. 033305:1-11, 2016.
- [3] S. Otsuka and T. Omori, "Estimation of neuronal dynamics based on sparse modeling," *Neural Networks*, vol. 109, pp. 137-146, 2019.
- [4] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Computation*, vol. 14, pp. 715-770, 2002.
- [5] G. E. Hinton, "Connectionist learning procedures," *Artificial Intelligence*, vol. 40, pp. 185-234, 1989.
- [6] R. Turner and M. Sahani, "A maximum-likelihood interpretation for slow feature analysis," *Neural Computation*, vol. 19, pp. 1022-1038, 2007.
- [7] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 436-450, 2012.
- [8] D. Jayaraman and K. Grauman, "Slow and steady feature analysis: higher order temporal coherence in video," in *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3852-3861.
- [9] P. Koch, W. Konen, and K. Hein, "Gesture recognition on few training data using slow feature analysis and parametric bootstrap," in *Proc. The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1-8.
- [10] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sensing of Environment*, vol. 199, pp. 241-255, 2017.
- [11] L. Zafeiriou, M. A. Nicolaou, S. Zafeiriou, S. Nikitidis, and M. Pantic, "Probabilistic slow features for behavior analysis," *IEEE Trans Neural Netw Learn Systems*, vol. 27, pp. 1034-1048, 2016.
- [12] H. Sakaino, "Spatio-temporal feature extraction/recognition in videos based on energy optimization," *IEEE Trans Image Process*, vol. 28, pp. 3395-3407, 2019.
- [13] S. Zhang, C. Zhao, and B. Huang, "Simultaneous static and dynamic analysis for fine-scale identification of process operation statuses," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 5320-5329, Sep. 2019.
- [14] X. Zou and C. Zhao, "Concurrent assessment of process operating performance with joint static and dynamic analysis," *IEEE Transactions on Industrial Informatics*, pp. 1-1, 2019.
- [15] T. Sekiguchi, T. Omori, and M. Okada, "The effect of observation noise in probabilistic slow feature analysis," *IPSI Transactions in Mathematical Modeling and Its Applications*, vol. 5, pp. 26-31, 2012.

- [16] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 9976-9992, 2019.
- [17] T. Omori, T. Sekiguchi, and M. Okada, "Belief propagation for probabilistic slow feature analysis," *Journal of the Physical Society of Japan*, vol. 86, pp. 084802:1-6, 2017.
- [18] S. Aminikhanghahi and D. J. Cook, "A Survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, pp. 339-367, 2017.
- [19] V. Pavlovic, J. M. Rehg, and J. MacCormick, "Learning switching linear models of human motion," *Advances in Neural Information Processing Systems*, vol. 13, pp. 981-987, 2001.
- [20] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *Journal of Vision*, vol. 5, pp. 49-64, 2005.
- [21] A. Takeuchi and T. Omori, "Estimation of hyperparameters in probabilistic slow feature analysis," in *Proc. Joint 7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems*, 2014, pp. 1606-1608.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1-38, 1977.
- [23] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, The MIT Press, 2012.
- [24] Z. Ghahramani and G. E. Hinton, "Variational learning for switching state space models," *Neural Computation*, vol. 12, pp. 831-864, 2000.
- [25] V. Pavlovic, J. M. Rehg, T. Cham, and K. P. Murphy, "A dynamic Bayesian network approach to figure tracking using learned dynamic models," in *Proc. the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 1, pp. 94-101.
- [26] K. P. Murphy, "Switching Kalman filter," *Tech. Rep.*, U. C. Berkeley 1998.
- [27] J. J. Dabrowski, J. P. de Villiers, and C. Beyers, "Naive Bayes switching linear dynamical system: A model for dynamic system modelling, classification, and information fusion," *Information Fusion*, vol.42, pp. 75-101, 2018.
- [28] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, pp. 4-16, 1986.
- [29] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Fluids Eng.*, vol. 82, pp. 35-45, 1960.
- [30] M. Athans, "The Importance of Kalman filtering methods for economic systems," *Annals of Economic and Social Measurement*, vol. 3, pp. 49-64, 1974.
- [31] J. Hu, K. Fennel, J. P. Mattern, and J. Wilkin, "Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the middle Atlantic bight," *Journal of Marine Systems*, vol. 94, pp. 145-156, 2012.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Kazuki Tsujimoto was born in 1996. He received his B.E. degree from Kobe University in 2018. His research interests include machine learning theory and its applications.



Toshiaki Omori received his B.S. degree in physics from University of Tsukuba in 1999, and his Ph.D. degree in information science from Tohoku University in 2004. He was a predoctoral research fellow of Japan Society for the Promotion of Science (JSPS) from 2003 to 2004, a postdoctoral researcher at Japan Science and Technology Agency (JST) from 2004 to 2006, and a postdoctoral research fellow of JSPS from 2006 to 2008. He was a visiting researcher at University of Arizona, U.S.A. in 2007. He became a research assistant professor and an assistant professor at the University of Tokyo in 2008. He is currently an associate professor at the Graduate School of Engineering, Kobe University. His research interests include machine learning theory and its applications, data-driven science, probabilistic information processing, and computational neuroscience.