Change-Point Detection in Time-Series Data based on Subspace Identification

Yoshinobu Kawahara, Takehisa Yairi, and Kazuo Machida
The University of Tokyo
Komaba 4-6-1, Meguro-ku, Tokyo, 153-8904 Japan
{kawahara, yairi, machida}@space.rcast.u-tokyo.ac.jp

Abstract

In this paper, we propose series of algorithms for detecting change points in time-series data based on subspace identification, meaning a geometric approach for estimating linear state-space models behind time-series data. Our algorithms are derived from the principle that the subspace spanned by the columns of an observability matrix and the one spanned by the subsequences of time-series data are approximately equivalent. In this paper, we derive an batch-type algorithm applicable to ordinary time-series data, i.e. consisting of only output series, and then introduce the online version of the algorithm and the extension to be available with input-output time-series data. We illustrate the effectiveness of our algorithms with comparative experiments using some artificial and real datasets.

1. Introduction

Abrupt changes of the properties of time-series often contain critically important information from various perspectives, and hence, the problem of discovering time points where changes occur, called *change points*, has received much attention in statistics and data mining [2, 4]. The application of this issue includes fraud detection, network-instusion detection, irregular-motion detection in vision, fault detection in controlled systems, and others.

For change-point detection in time series, an approach possible to explicitly handle the interdependence over time need to be adopted. Thus, although a number of useful methods for novelty detection in multivariate data, i.e. not time-series data, have been explored [10, 11], it is inappropriate to directly apply those convensional approaches for this problem. Recently, Yamanishi and Takeuchi proposed an framework in which an autoregressive (AR) model is learned recursively and hence the nonstationarity in time-series can be dealt with explicitly [17, 16]. Also, change-point detection algorithms based on the singular-spectrum analysis (SSA), which is a classical nonparametric methodology for time-series analysis handled mainly in signal processing, are proposed [14, 5]. However, in the SSA-based detection algorithm, a state-space model (SSM) with no system noises is implicitly assumed as the dynamic model behind time-series data.

In this paper, we propose well-understood nonparametric algorithms for change-point detection in time-series data, based on the principle that the subspace spanned by the columns of an extended observability matrix is approximately equivalent to the one spanned by the subsequences of time-series data. In our framework, change-point detection is performed by estimating, on the basis of subspace identification, the column space of the extended observability matrix of the SSM behind time-series data, and the evaluating the subsequence of new-arrived data based on this subspace. Thus, our method can handle more abundant type of time-series data in pricise than conventional approaches because of implicitly utilizing generic SSMs, instead of AR models or constrained SSMs, as the model behind time-series data. Moreover, in this paper, we describe the online version of the algorithm and the extention to be available with input-output time-series data.

The remainder of this paper is organized as follows. Section 2 gives some background materials on subspace identification for stochastic dynamic systems. In Section 3, we derive an offline- and an online-algorithm for detecting change points based on subspace identification. Then, we extend these to be available with input-output time-series data in Section 4. Finally, empirical comparative results are given in Section 5, and we conclude this paper in Section 6.

2. Subspace Methods for Dynamic Systems

Consider a discrete-time wide-sense stationary vector process $\{\boldsymbol{y}(t) \in \mathbb{R}^p, t = 1, 2, \cdots\}$ which models the signal of the unknown stochastic system as a discret-time linear state-space system:

$$\begin{aligned} \boldsymbol{x}(t+1) &= A\boldsymbol{x}(t) + \boldsymbol{v}(t), \\ \boldsymbol{y}(t) &= C\boldsymbol{x}(t) + \boldsymbol{w}(t), \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $\boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{w} \in \mathbb{R}^p$ are the system and observation noises, respectively, and $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{p\times n}$ are the system matrices. Throughout this section, we shall assume that the process $\{\boldsymbol{y}(t)\}$ is a stationary and purely nondeterministic full rank

1550-4786/07 $25.00 © 2007 IEEE
DOI 10.1109/ICDM.2007.78

process [7] [1] . One of the promising approaches for solving this type of problem is subspace identification methods [15, 6], which was proposed in the early 1980s and has been one of the topics receiving most attention in system identification. Note that the case in which a target system includes no inputs, as equation (1), is called stochastic identification, and the origin goes back further [1, 3].

The key problem dealt with by subspace identification, in which a state-space model (1) is estimated in the form of an innovation model [2] , is the consistent estimation of the column space of the extended observability matrix

$$\mathcal{O}_k := \left[\begin{array}{cccc} C^T & (CA)^T & \cdots & (CA^{k-1})^T \end{array} \right], \qquad (3)$$

from the measurements. Once the extended observability matrix is obtained, it is straightforward to calculate the system matrices and the Kalman gain. Although many types of the approaches for stochastic identification have been proposed, we proceed to a discussion in the basis of the balanced stochastic realization [3]. At the first step in this method, with the data for identification $\{\boldsymbol{y}(\bar{t}), \bar{t} = 0, \cdots, N+2k-2\}$, two Hankel matrices, defined as [3]

$$Y_p := Y_{k,N}(0) \ \text{ and } \ Y_f := Y_{k,N}(k), \qquad (4)$$

are constructed, where the suffix p means *past* and f means *future*. Then, for the computation of the extended observability matrix, we need to carry out the following SVD of the normalized conditional covariance

$$\Sigma_{ff}^{-1/2}\Sigma_{fp}\Sigma_{pp}^{-T/2} = USV^T \approx U_1 S_1 V_1^T, \qquad (5)$$

where the approximation $\approx$ is performed by selecting the elements compatible with the n significant singular values, and the covariance matrices Σ_{fp}, Σ_{ff} and Σ_{pp} are calculated using the matrices obtained by the LQ factorization

$$\frac{1}{\sqrt{N}} \left[\begin{array}{c} Y_p \\ Y_f \end{array} \right] = \left[\begin{array}{cc} L_{11} & 0 \\ L_{21} & L_{22} \end{array} \right] \left[\begin{array}{c} Q_1^T \\ Q_2^T \end{array} \right], \qquad (6)$$

[1] We will remove the assumption on stationarity implicitly in Section 3.2, in which the recursive algorithm is discussed.

[2] There are always infinitely many linear representations of the form in equation (1) equivalent to (conditional) second-order statistics for an identical time series. Therefore, in most subspace identification methods, a realization concerning the so-called innovation model

$$\begin{aligned} \boldsymbol{x}(t+1) &= A\boldsymbol{x}(t) + K\boldsymbol{e}(t), \\ \boldsymbol{y}(t) &= C\boldsymbol{x}(t) + \boldsymbol{e}(t), \end{aligned} \qquad (2)$$

is calculated, where $\boldsymbol{e}(t) \in \mathbb{R}^p (:= \boldsymbol{y}(t) - \hat{E}\{\boldsymbol{y}(t)|\mathcal{Y}_t^-\})$ is an innovation process (the error process of the model) and $K \in \mathbb{R}^{n\times p}$ is the stationary Kalman gain, where $\mathcal{Y}_t^-$ is the linear space spanned by the past subsequences and $\hat{E}\{\boldsymbol{y}(t)|\mathcal{Y}_t^-\}$ is the orthogonal projection of $\boldsymbol{y}$ onto $\mathcal{Y}_t^-$.

[3] In this paper, let $\{\boldsymbol{z}(t) \in \mathbb{R}^d, t = 1, \cdots, T\}$ be a time-series, then $\boldsymbol{z}_k(t) \in \mathbb{R}^{dk}$ will be referred to as the subsequence vector

$$\boldsymbol{z}_k(t) = \left[\begin{array}{cccc} \boldsymbol{z}(t)^T & \boldsymbol{z}(t+1)^T & \cdots & \boldsymbol{z}(t+k-1)^T \end{array} \right]^T,$$

where $t + k < T$ and $\bullet^T$ denotes the transpose of a matrix (or a vector) $\bullet$. Moreover, $Z_{k,N}(t) \in \mathbb{R}^{dk\times N}$ will be referred to as the Hankel matrix constructed by aligning the subsequence vectors as

$$Y_{k,N}(t) = \left[\begin{array}{cccc} \boldsymbol{y}_k(t) & \boldsymbol{y}_k(t+1) & \cdots & \boldsymbol{y}_k(t+N-1) \end{array} \right].$$

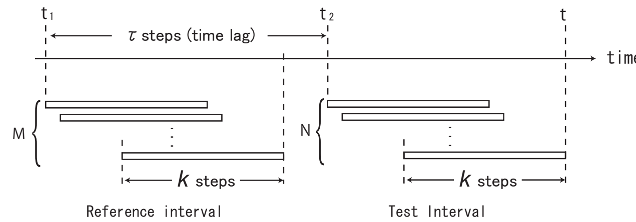

Figure 1. Scheme of change-point detection.

respectively [4] . And, using the obtained singular vectors, the extended observability matrix $\mathcal{O}_k$ is calculated as

$$\mathcal{O}_k = \Sigma_{ff}^{1/2} U_1 S_1^{1/2}. \qquad (8)$$

As is evident from the above descriptions, the estimation of state-space models by subspace identification can be implemented by fast and reliable numerical schemes. And it is well known that, in general, subspace methods require no apriori choice of identifiable parameterizations and avoid local minima, which are the serious problems in other approaches such as maximum likelihood estimate approaches.

3. Change-Point Detection Algorithms

3.1. Algorithm description

Using an extended observability matrix $\mathcal{O}_k$ and an innovation model (2), a subsequence can be expressed as

$$\boldsymbol{y}_k(t) = \mathcal{O}_k\boldsymbol{x}(t) + \Psi_k\boldsymbol{e}_k(t), \qquad (9)$$

where Ψ_k is the constant matrix defined with the system matrices and the Kalman gain, as in the footnote [5] . Moreover, by aligning equations (9) according to the structure of a Hankel matrix, $Y_{k,N}(t)$ can be described as

$$Y_{k,N}(t) = \mathcal{O}_k X_0 + \Psi_k E_{k,N}(t), \qquad (10)$$

where $X_0 := [\boldsymbol{x}(t), \cdots, \boldsymbol{x}(t+N-1)] \in \mathbb{R}^{n\times N}$. Hence, the subspace spanned by the column vectors of $Y_{k,N}(t)$ is equivalent to the one spanned by the columns of $\mathcal{O}_k$ plus Ψ_k. Because $\Psi_k E_{k,N}$ in Equation (10) is caused by noises and minutely small, the approximate relation

$$\mathcal{L}\{Y_{k,N}(t)\} \approx \mathcal{L}\{\mathcal{O}_k\} \qquad (11)$$

is satisfied, where $\mathcal{L}\{A\} := \{A\boldsymbol{\alpha} \in \mathbb{R}^n\}$ is the subspace spanned by the column vectors of a matrix $A \in \mathbb{R}^{m\times n}$.

Consider two intervals with a time lag τ; the test interval which consists of the latest $N+k-1$ steps and the reference

[4] In concrete terms, the covariances are calculated as

$$\Sigma_{fp} = L_{21}L_{11}^T, \ \Sigma_{ff} = L_{21}L_{21}^T + L_{22}L_{22}^T, \ \Sigma_{pp} = L_{11}L_{11}^T. \qquad (7)$$

[5]

$$\Psi_k := \left[\begin{array}{cccc} I_{p\times p} & 0 & \cdots & 0 \\ CK & I_{p\times p} & \cdots & 0 \\ \vdots & CK & \cdots & \vdots \\ CA^{k-2}K & CA^{k-3}K & \cdots & I_{p\times p} \end{array} \right]$$

Select k, M, N, τ and n.
At each time t
    Construct Y_p and Y_f correspoinding to the reference interval (cf. (4)).
    Calculate the covs. Σ_{pp}, Σ_{pf}, Σ_{ff} and the SVD $\Sigma_{ff}^{-1/2}\Sigma_{fp}\Sigma_{pp}^{-T/2} \approx U_1 S_1 V_1^T$ (cf.(5)), and then estimate the extended obsevability matrix $\mathcal{O}_k$.
    Construct the Hankel matrix $Y_{k,N}(t_2)$ of the test interval, and then evaluate the distance $\mathcal{D}$ (12).

Algorithm 1: Pseudo-code for the change-point detection based on subspace methods (offline).

interval of the past $M + k - 1$ steps, as shown in figure 1. Note that, although there is no temporal overlap between these intervals in the figure, it does not matter whether or not there is such overlap. Then, if the properties of time series do not change, from Equation (11), we can expect that there is not a large gap between the subspace of the extended observability matrix estimated in the reference interval and the one of the Hankel matrix consisting of the subsequences in the test interval. Hence, the following distance, which evaluates the gap between subspaces (the normalized type of this is equivalent to the *cepstral distance* [12]), can be utilized as the measure of the change in time-series:

$$\begin{aligned}\mathcal{D} := &\ Y_{k,M}(t_2)^T Y_{k,M}(t_2) \\ &- Y_{k,M}(t_2)^T U_1^{(1)} (U_1^{(1)})^T Y_{k,M}(t_2),\end{aligned} \tag{12}$$

where $U^{(1)}$ is calculated by the SVD of the extended observability matrix $\mathcal{O}_k$ estimated by subspace identification using the data in the reference interval as

$$\mathcal{O}_k^{(1)} = U^{(1)} S^{(1)} (V^{(1)})^T. \tag{13}$$

Note that other measures which evaluate the gap between subspaces can be available as the criteria of changes. The procedure for change-point detection can be outlined as shown in Algorithm 1. The choices of the parameters k and τ are not sensitive, as the SSA-based algorithms [14, 5]. The order n is determined from the SVD (5) and the length of the intervals M and N need to be selected according to the properties of objective data.

3.2. Recursive implementation

Change-point detection in time series need to be carried out in real time. One effectual alternative for turning down the computational cost in the execution is to utilize the results calculated at previous time, namely to calculate recursively. Recursive computation is also valuable in the case that the properties of time series changes smoothly as time advances. Hence, in this section, we describe the recursive computation of the algorithm derived above.

The main procedure of the algorithm described above is the estimation of the subspace spanned by the column vectors of the extended observability matrix. This can be performed recursively by *propagator methods* [9, 13], mainly

Select k, M, N, τ and n.
Initialize P, Σ_{y_1}, Σ_{y_2} and M.
At each time t
    Update P, Σ_{y_1}, Σ_{y_2} and M by (18)-(26), and estimate the observability subspace.
    Construct the Hankel matrix $Y_{k,N}(t_2)$ of the test interval, and then evaluate the distance $\mathcal{D}$ (12).

Algorithm 2: Pseudo-code for the change-point detection based on subspace methods (online).

used in signal processing and subspace methods. In propagator methods, we first need to partition an extended observability matrix as

$$\bar{\mathcal{O}}_k = \begin{bmatrix} \mathcal{O}_1 \\ \mathcal{O}_2 \end{bmatrix} \begin{matrix} \} \in \mathbb{R}^{n \times n} \\ \} \in \mathbb{R}^{(kp-n) \times n} \end{matrix}, \tag{14}$$

where $\mathcal{O}_1$ and $\mathcal{O}_2$ have the relation $\mathcal{O}_2 = P^T \mathcal{O}_1$, with a linear operator, called a *propagator*, $P \in \mathbb{R}^{n \times (kp-n)}$. That is, $\mathcal{O}_1$ is the matrix constructed from the n rows of the extended observability matrix that are linearly independent. Because the rank of an extended observability matrix is full, such rows can be always selected and unique P is determined. Hence, the estimation of the column subspace of $\mathcal{O}_k$ is solved by the following least square problem:

$$\underset{P}{\text{minimize}} \sum_{i=1}^{t} \lambda^{t-i} \| \boldsymbol{y}_2(i) - P^T \boldsymbol{y}_1(i) \|^2, \tag{15}$$

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are the vectors obtained by aligning corresponding to Equation (14), and λ is a suitable forgetting factor [6] . Thus, the propagator P can be calculated recursively by applying a recursive least square (RLS) method to this problem. However, the estimator has bias if noises have correlations with the subsequences, which occur in the case that system noises exist, and should be corrected. One of the useful methods that resolve this issue is the instrumental variable method, which calculates its consistent estimator using a new vector correlated with regression vectors ($\boldsymbol{y}_1$ in this case) and noncorrelated with noises, called instrumental variables [7] . In the case of (15), we can use past subsequences as instrumental variables. As a result, we obtain an recursive type of the algorithm for change-point detection,

[6] In the case that objective time-series is nonstationary, it is often useful to weight the latest data using a forgetting factor. Note that when such weighting is not necessary, set $\lambda = 1$.

[7] In the instrumental variable method for the calculation of the following least square problem, given $\{\boldsymbol{z}_1(t), \boldsymbol{z}_2(t), t = 0, \cdots, \tau\}$:

$$\underset{P}{\text{minimize}} \sum_{t=0}^{\tau} \lambda^{\tau-t} \| \boldsymbol{z}_2(t) - P^T \boldsymbol{z}_1(t) \|^2, \tag{16}$$

instrumental variables (IVs), which are an arbitary vector highly correlated with regression vectors ($\boldsymbol{z}_1$ in this case) and with noncorrelated with noises, are introduced. In the case of (16), if we denote IVs by ξ, we can replace the minimization problem to the following:

$$\underset{P}{\text{minimize}} \sum_{t=0}^{\tau} \lambda^{\tau-t} \| \boldsymbol{z}_2(t)\boldsymbol{\xi}(t)^T - P^T \boldsymbol{z}_1(t)\boldsymbol{\xi}(t)^T \|^2. \tag{17}$$

Then, the updates for this minimization problem are carried out as in which subsequences of the test intervals are evaluated using the column subspace of the extended observability matrix updated recursively, outlined in algorithm 2.

4. Input-Output Time-Series Data

In this section, we focus on the way that the algorithms derived in the previous section applies to input-output time-series data $\{\boldsymbol{u}(t) \in \mathbb{R}^q, \boldsymbol{y}(t) \in \mathbb{R}^p, t = 1, 2, \cdots\}$. Such a extension makes an intrinsic approach to time-series data which, in nature, should be devided into inputs and outputs, which occur frequently in the application of change-point detection, such as fault detection in engineering systems.

4.1. Modification in the offline algorithm

For identification with the input-output time-series data $\{\boldsymbol{u}(\bar{t}), \boldsymbol{y}(\bar{t}), \bar{t} = 0, \cdots, N + 2k - 2\}$, we first need to construct the Hankel matrices and the Lipchitz matrix [8];

$$Y_f := Y_{k,N}(k), \quad U_f := U_{k,N}(k), \quad \check{W}_p := \check{W}_{k,N}(0), \tag{27}$$

where $\boldsymbol{w}(t)^T := [\boldsymbol{y}(t)^T, \boldsymbol{u}(t)^T]^T$ means the joint input-output process. Using these matrices, the extended observability matrix can be calculated by, instead of (5), the SVD

$$\Sigma_{ff|u}^{-1/2} \Sigma_{fp|u} \Sigma_{pp|u}^{-T/2} = USV^T \approx U_1 S_1 V_1^T, \tag{28}$$

where the conditional covariances [9] $\Sigma_{fp|u}$, $\Sigma_{ff|u}$ and $\Sigma_{pp|u}$ are calculated using the matrices by the LQ factorization [10]. Then, the extended observability matrix can be estimated as

$$P^T(t) = P^T(t-1) + \Big(\boldsymbol{g}(t) - P^T(t-1)\Phi(t)\Big) K(t), \tag{18}$$

$$\boldsymbol{g}(t) = \left[\begin{array}{cc} \Sigma_{z_2\xi}(t-1)\boldsymbol{\xi} & \boldsymbol{z}_2 \end{array}\right], \tag{19}$$

$$\Lambda(t) = \left[\begin{array}{cc} -\boldsymbol{\xi}(t)^T\boldsymbol{\xi}(t) & \lambda \\ \lambda & 0 \end{array}\right], \tag{20}$$

$$\boldsymbol{q}(t) = \Sigma_{z_1\xi}(t-1)\boldsymbol{\xi}(t), \tag{21}$$

$$\Phi(t) = \left[\begin{array}{cc} \boldsymbol{q}(t) & \boldsymbol{z}_1(t) \end{array}\right], \tag{22}$$

$$K(t) = \Big(\Lambda(t) + \Phi^T(t)M(t-1)\Phi(t)\Big)^{-1} \Phi^T(t)M(t-1), \tag{23}$$

$$\Sigma_{z_1\xi}(t) = \lambda\Sigma_{z_1\xi}(t-1) + \boldsymbol{z}_1(t)\boldsymbol{\xi}(t)^T, \tag{24}$$

$$\Sigma_{z_2\xi}(t) = \lambda\Sigma_{z_2\xi}(t-1) + \boldsymbol{z}_2(t)\boldsymbol{\xi}(t)^T, \tag{25}$$

$$M(t) = \frac{1}{\lambda^2}\left(M(t-1) - M(t-1)\Phi(t)K(t)\right), \tag{26}$$

where $M(t) := (\Sigma_{z_1\xi}(t)\Sigma_{z_1\xi}(t)^T)^{-1}$.

[8] In this case, the block Lipchitz matrix corresponding to $\boldsymbol{w}(t)$ is defined as

$$\check{W}_{k,N}(0) := \left[\begin{array}{cccc} \boldsymbol{w}(k-1) & \boldsymbol{w}(k) & \cdots & \boldsymbol{w}(k+N-2) \\ \boldsymbol{w}(k-2) & \boldsymbol{w}(k-1) & \cdots & \boldsymbol{w}(k+N-3) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{w}(0) & \boldsymbol{w}(1) & \cdots & \boldsymbol{w}(N-1) \end{array}\right].$$

[9] Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ be random vectors, then the conditional covariances matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$ conditioned on $\boldsymbol{z}$ is defined as

$$\Sigma_{xy|z} := E\left[\hat{E}\{\boldsymbol{x}|\boldsymbol{z}^\perp\}\hat{E}\{\boldsymbol{y}|\boldsymbol{z}^\perp\}\right] = \Sigma_{xy} - \Sigma_{xu}\Sigma_{uu}^{-1}\Sigma_{by}^T.$$

[10] In concrete terms, these conditional covariances are calculated as

$$\Sigma_{fp|u} = L_{32}L_{22}^T, \Sigma_{ff|u} = L_{32}L_{32}^T + L_{33}L_{33}^T, \Sigma_{pp|u} = L_{22}L_{22}^T, \tag{29}$$

where L_{22} etc. are obtained by the following LQ factorization:

$$\frac{1}{\sqrt{N}}\left[\begin{array}{c} U_f \\ \check{W}_p \\ Y_f \end{array}\right] = \left[\begin{array}{ccc} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{array}\right]\left[\begin{array}{c} Q_1^T \\ Q_2^T \\ Q_3^T \end{array}\right]. \tag{30}$$

```
Select k, M, N, τ and n.
At each time t
    Construct U_f, Y_f and W̌_p corresponding to the reference interval (cf. (27)).
    Calculate Σ_{pp|u}, Σ_{pf|u}, Σ_{ff|u} and the SVD Σ_{ff|u}^{-1/2} Σ_{fp|u} Σ_{pp|u}^{-T/2} ≈ U_1 S_1 V_1^T (cf.(28)), and then estimate the extended obsevability matrix O_k.
    Construct the Hankel matrix Y_{k,N}(t_2) of the test interval and calculate Ê{Y_{k,N}(t)|U^⊥} (cf.(34)).
    Evaluate the distance D (12).
```

Algorithm 3: Pseudo-code for the change-point detection based on subspace methods (offline, in-out).

$$\mathcal{O}_k = \Sigma_{ff|u}^{1/2} U_1 S_1^{1/2}. \tag{31}$$

In addition, the Hankel matrix by aligning the subsequences of observation series is expressed by

$$Y_{k,N}(t) = \mathcal{O}_k X_0 + G_k U_{k,N}(t) + \Psi_k E_{k,N}(t), \tag{32}$$

where $G_k \in \mathbb{R}^{kp \times kq}$ is the constant matrix defined with the system matrices as well as Ψ_k (cf.(11)). Hence, as in Section 3.1, the following relation is approximately holded;

$$\mathcal{L}\{\hat{E}\{Y_{k,N}(t)|\mathscr{U}^\perp\}\} \approx \mathcal{L}\{\mathcal{O}_k\}, \tag{33}$$

where $\mathscr{U}^\perp$ is the linear space orthogonal to the column subspace of the Hankel matrix $U_{k,N}(t)$, and the projection in the left-hand side is calculated as

$$\hat{E}\{Y_{k,N}(t)|\mathscr{U}^\perp\} = L_{22}Q_2^T, \tag{34}$$

using the matrices obtained by the LQ factorization

$$\left[\begin{array}{c} U_{k,N}(t) \\ Y_{k,N}(t) \end{array}\right] = \left[\begin{array}{cc} L_{11} & 0 \\ L_{21} & L_{22} \end{array}\right]\left[\begin{array}{c} Q_1^T \\ Q_2^T \end{array}\right]. \tag{35}$$

As a result, the algorithm for input-output time-series data is outlined as in Algorithm 3.

4.2. Recursive implementation

The algorithm for input-output time-series data can be also modified to be performed recursively using the instrumental variable method. However, for input-output time-series data, after projecting outputs to the space orthogonal to input space, the method described in Section 3.2 should be applied, from the difference between (11) and (33).

While, in the offline case, the projection of outputs to the orthogonal space of inputs is calculated by the LQ factorization (34), the online computation can be carried out as follows. First, when new data $(\boldsymbol{u}(\tau), \boldsymbol{y}(\tau))$ arrives, by

Select k, M, N, τ and n.
Initialize P, Σ_{y_1}, Σ_{y_2}, M and L (in eq.(34)).
At each time t
    Calculate $\boldsymbol{z}(\tau)$ and update L by the succesive Givens rotations as in equations (36) and (37).
    Update P, Σ_{y_1}, Σ_{y_2} and M by (18)-(26), and estimate the observability subspace.
    Construct the Hankel matrix $Y_{k,N}(t_2)$ of the test interval and calculate $\hat{E}\{Y_{k,N}(t)|\mathscr{U}^{\perp}\}$ (cf.(34)).
    Evaluate the distance $\mathcal{D}$ (12).

Algorithm 4: Pseudo-code for the change-point detection based on subspace methods (online, in-out).

adding the new data to the rightmost column of (35), the following equation is satisfied:

$$\begin{bmatrix} U_{k,N+1}(t) \\ Y_{k,N+1}(t) \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \boldsymbol{u}_k(\tau) \\ L_{21} & L_{22} & \boldsymbol{y}_k(\tau) \end{bmatrix} \begin{bmatrix} Q_1^T & 0 \\ Q_2^T & 0 \\ 0 & 1 \end{bmatrix}. \tag{36}$$

A sequence of Givens rotation $G(\tau)$ can then be used to zero out the vector $\boldsymbol{u}(\tau)$ and bring back the L factor to block lower triangular form as

$$\begin{bmatrix} L_{11} & 0 & \boldsymbol{u}_k(\tau) \\ L_{21} & L_{22} & \boldsymbol{y}_k(\tau) \end{bmatrix} G(\tau) = \begin{bmatrix} \hat{L}_{11} & 0 & 0 \\ \hat{L}_{21} & \hat{L}_{22} & \boldsymbol{z}(\tau) \end{bmatrix}. \tag{37}$$

As a result, $\boldsymbol{z}(\tau)$ is the vector obtained by making the new output project to the orthogonal space of the inputs, and $\hat{L}_{11}$, etc. are the updated L factors in equation (35). Using the obtained vector $\boldsymbol{z}(\tau)$, we can estimate the column subspace of the extended observability matrix by applying the instrumental variable method to the following least square problem, as in the case of Section 3.2:

$$\underset{P}{\text{minimize}} \sum_{i=0}^{t} \lambda^{t-i} \|\boldsymbol{z}_2 - P^T \boldsymbol{z}_1\|^2. \tag{38}$$

Thus, change-point detection for input-output time-series data can be performed recursively by evaluating the subsequences after the projection in Equation (34) in the test interval with the subspace of the estimated observability matrix. The overall procedure is shown in Algorithm 4.

5. Experiments

5.1. Change detection in patient monitoring

We first show comparative results with the SSA-based algorithm [14] using the respiration datasets [11] in the UCR Time Series Data Mining Archive. This dataset records patients' respiration measured by thorax extension, as they wake up, and manually segmented by a medical expert, Dr. J. Rittweger of the Institute for Physiology, Free University of Berlin [8]. Figures 2 and 3 show the distances (12) calculated by the proposed algorithm (Algorithm 1) and the SSA-based method, in which the parameters are selected as

[11] Available from the website www.cs.ucr.edu/~eamonn/discords/

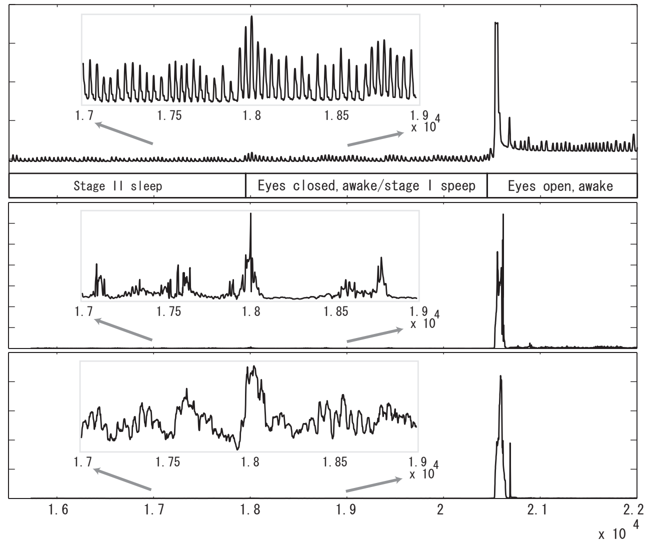

Figure 2. The raw nprs44 data (top), and the degrees of change (the distances $\mathcal{D}$ (12)) calculated by the proposed algorithm (middle) and the SSA-based algorithm (bottom).

$\tau = 150$, $k = 8$, $N = 150$, $N = 80$ and $n = 2$. As can be seen from these results, our algorithm outperform the SSA-based method regarding the correctness and the sharpness of detecting change points, meaning time points switching from *sleep* to *awake* or from *awake* to *sleep*.

5.2. Fault detection in engineering systems

As the second example, we illustrate the application of our algorithm to fault detection, using an artificial data simulating a malfunction in the attitude controller of a spacecraft [12]. In the simulation, the attitude of the spacecraft is controlled by feedback using three reaction wheels (RWs), which are instruments for attitude control based on gyroscopic procession, and kept near its target attitude. The data consists of inputs ($\boldsymbol{r}_1 \in \mathbb{R}^6, \boldsymbol{r}_2 \in \mathbb{R}^3$) and outputs ($\boldsymbol{u} \in \mathbb{R}^3, \boldsymbol{y} \in \mathbb{R}^9$) (see figure 4), and its frequency and length are respectively 0.25 [s] and 5,500 [s] (i.e. the data consists of 22,000 samples). First, the spacecraft has been controlled from time 0 [s] to 1,500 [s] so that it is kept constant posture (all attitude angles and angular rates are zeros). Then, at time $1,500$[s], the command of the attitude change (the change of the precession angle to 22.5 [deg]) is sent and the spacecraft begin to carry out the command using the RWs. After that, the spacecraft continue to control for keeping the attitude but, from 3,500 [s], an abnormality occurs in one of the RWs (the RW for the control of the roll axis), in which the power of the RW descends gradually to the half of its original power. Note that detecting the fault

[12] This simulation was carried out using MATLAB software Spacecraft Control Toolbox (www.psatellite.com/products/html/sct.php) produced by Princeton Satellite Systems, Inc..

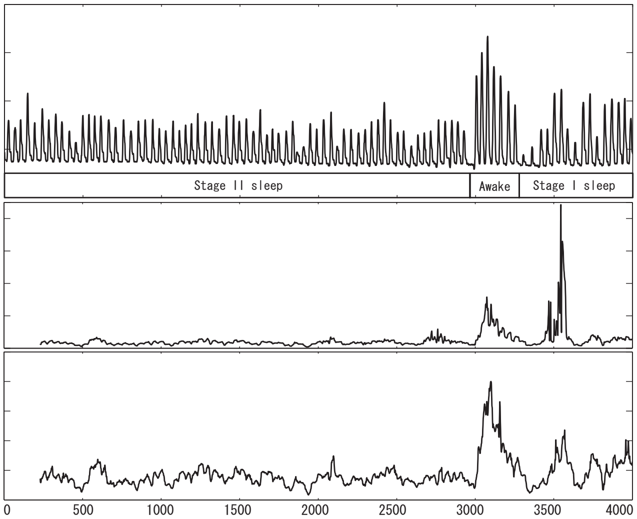

Figure 3. The raw nprs43 data (top), and the degrees of change (the distances $\mathcal{D}$ (12)) calculated by the proposed algorithm (middle) and the SSA-based algorithm (bottom).

fraught with gradual changes, as in this example, at an early stage is a relatively difficult problem compared with abrupt changes, but this is important tasks in practice.

We applied our recursive algorithm 4 to this data, in which the first 2,000 samples (5,00[s]) were used for the initialization and the parameters were set as $\lambda = 0.98$, $\tau = 20$, $k = 25$, $N = 200$ and $n = 7$. Figure 5 illustrates the distance $\mathcal{D}$ calculated by Algorithm 4, in which two changes, i.e. the abrupt change that accompanies attitude change at time $t = 1,500$[s] and the gradual change caused by the fault from time $t = 3,500$[s], can be detected, increasing rapidly and gradually the distance (12). Note that the visible sympton of the failure manifests itself in the raw data right before going out of control (near at 4800 seconds).

6. Conclusions

In this paper, we proposed series of algorithms for detecting change points in time-series data based on subspace identification. These are based on the approximate equivalence between the subspaces spanned by the subsequences of time-series data and the columns of the extended observability matrix. On the basis of this, we introduced the offline- and the online-algorithms, and discussed those extension to be available with input-output data. The comparative experiments showed the high performance of our methods. In future work, we will develop the idea for nonlinear time-series based on reproducing kernels and so on.

References

[1] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13(1):162–173, 1975.

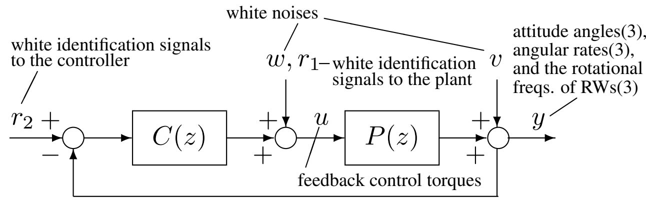

Figure 4. Experimental model.

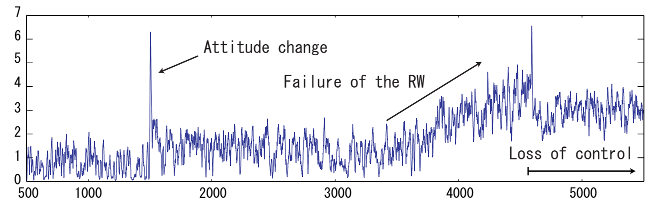

Figure 5. The degree of change (the distance $\mathcal{D}$ (12)) calculated by Algorithm 4.

[2] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Inc., Simon & Schuster Company, Englewood Cliff, 1993.
[3] U. B. Desai, D. Pal, and R. D. Kirkpatrick. A realization approach to stochastic model reduction. *Int. J. Control*, 42(4):821–838, 1985.
[4] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley & Sons Inc, 2000.
[5] T. Ide and H. Kashima. Eigenspace-based anomaly detection in computer systems. In *Proc. of the 10th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 440–449, 2004.
[6] T. Katayama. *Subspace Methods for System Identification: A Realization Approach*. Communications and Control Engineering. Springer Verlag, 2005.
[7] T. Katayama and G. Picci. Realization of stochastic systems with exogenous inputs and subspace identification. *Automatica*, 35(10):1635–1652, 1999.
[8] E. Keogh, J. Lin, and A. Fu. Hot sax: Finding the most unusual time series subsequence: Algorithms and applications. In *Proc. of the 5th IEEE Int. Conf. on Data Mining*, pages 440–449, 2004.
[9] S. Marcos, A. Marsal, and M. Benidir. The propagator method for source bearing estimation. *Signal Processing*, 42:121–138, 1995.
[10] M. Markou and S. Singh. Novelty detection: A review - part 1: Statistical approaches. *Signal Processing*, 83(12):2481–2497, 2003.
[11] M. Markou and S. Singh. Novelty detection: A review - part 2: Neural network based approaches. *Signal Processing*, 83(12):2499–2521, 2003.
[12] R. J. Martin. A metric for arma processes. *IEEE Trans. on Signal Processing*, 48(4):1164–1170, 2000.
[13] G. Mercère, S. Lecoeuche, and M. Lovera. Recursive subspace identification based on instrumental variable unconstrained quadratic optimization. *Int. J. Adaptive Control and Signal Processing*, 18:771–797, 2004.
[14] V. Moskvina and A. A. Zhigljavsky. An algorithm based on singular spectrum analysis for change-point detection. *Communication in Statistics: Simulation & Computation*, 32(2):319–352, 2003.
[15] P. V. Overschee and B. D. Moor. *Subspace Identification for Linear Systems: Theory, Implementation and Applications*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1996.
[16] J. Takeuchi and K. Yamanishi. A unifying framework for detecting outliers and change points from time series. *IEEE Trans. on Knowledge and Data Engineering*, 18(4):482–489, 2006.
[17] K. Yamanishi and J. Takeuchi. A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proc. of the 8the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 676–681, 2002.