Segmentation of Piecewise ARX Processes by Exploiting Sparsity in Tight-Dimensional Spaces

Hiroki Kuroda, Masao Yamagishi, and Isao Yamada
Department of Information and Communications Engineering, Tokyo Institute of Technology
Email: {kuroda,myamagi,isao}@sp.ce.titech.ac.jp

***Abstract*—Segmentation of piecewise Auto-Regressive eXogenous (ARX) processes has been a major challenge in time-series segmentation and change detection. In this paper, for piecewise ARX process segmentation, we exploit hidden sparsity in tight-dimensional representation spaces. More precisely, we strategically design a tight-dimensional linear transformation which reveals sparsity hidden in samples following piecewise ARX processes. Experiments on synthetic and real-world data demonstrate the effectiveness of the proposed method.**

***Index Terms*—Time-series segmentation, change detection, piecewise ARX process, sparse representation.**

I. Introduction

Segmentation of piecewise Auto-Regressive eXogenous (ARX) processes has been a challenging task in, e.g., segmentation of piecewise stationary time-series such as speech [1]–[3], video [4], [5] and EEG signals [6], [7], and change detection in production plants [8], [9]. More precisely, we consider that $y_n \in \mathbb{R}$ the observed signal at time instant $n \in \{1, \ldots, N\}$ is obtained from a piecewise ARX process:

$$y_n := \boldsymbol{\xi}_n^\top \boldsymbol{\vartheta}_\ell^\star + \varepsilon_n \quad (n_{\ell-1}^\star < n \le n_\ell^\star), \tag{1}$$

for $\ell = 1, \ldots, L+1$ where $n_1^\star < \cdots < n_L^\star$ are unknown *change points*,[1] $\boldsymbol{\vartheta}_1^\star, \ldots, \boldsymbol{\vartheta}_{L+1}^\star$ consist of unknown coefficients in each time-segment, $\boldsymbol{\xi}_n := (y_{n-1}, \ldots, y_{n-q_1}, x_{n-1}, \ldots, x_{n-q_2})^\top \in \mathbb{R}^K$ $(K := q_1 + q_2)$ consists of $y_{n-1}, \ldots, y_{n-q_1}$ the past observations and $x_{n-1}, \ldots, x_{n-q_2}$ the known input signals, and ε_n includes the observation noise and the modeling error. A major goal of piecewise ARX process segmentation is to find $n_1^\star, \ldots, n_L^\star \in \{1, \ldots, N\}$ from $(\boldsymbol{\xi}_n, y_n)_{n=1}^N$.[2]

The difficulty in estimation of $n_1^\star, \ldots, n_L^\star$ is clearly understood, e.g., from the fact that searching optimal estimates results in a certain combinatorial problem. Although dynamic programming based approaches [10], [11] are developed specially for the model in (1), these approaches are computationally intractable in practical situations as stated in, e.g., [3].

To avoid such difficulty, the approaches in [3], [5], [12], [13] cleverly reduce estimation of $n_1^\star, \ldots, n_L^\star$ to sparse optimizations after expressing discrete samples in (1) in KN-dimensional redundant representation spaces. Meanwhile, for recovery of certain piecewise continuous signals, similar redundant representations are also introduced, e.g., in [14]–[16]. Namely, to use the sparsity, the approaches in [14]–[16] as well as [3], [5], [12], [13] rely on redundant representation spaces whose dimensions are several times larger than the number of samples.

In our recent work [17], for piecewise continuous signal recovery, we reveal hidden sparsity in a *tight-dimensional representation space* which stands for the representation space whose dimension does not exceed the number of samples. It is demonstrated in [17] that the sparsity revealed in the tight-dimensional representation space yields more accurate estimation than that in the redundant representation space.

The first contribution of this paper is to clarify that the strategy in [17] proposed for piecewise continuous signal recovery is in fact applicable for piecewise ARX process segmentation thanks to similarity of problem settings.

The second contribution is to demonstrate the effectiveness of the proposed method for piecewise ARX process segmentation through experiments on synthetic and real-world data. These experiments are presented to show advantages of the proposed method against the method based on the redundant representation introduced in [3], [5], [12], [13]. First, by using a synthetic example exactly following (1), we show that there exists a case where the redundant representation-based method fails to detect change points even for small enough noise. Meanwhile, the proposed method correctly detects change points for small to moderate levels of noise. Next, for speech segmentation, we demonstrate that the proposed method yields significant performance improvements. Namely, the proposed method detects several change points which seem to reasonably reflect phoneme boundaries of speech, while the redundant representation-based method results in providing inadequate information for speech segmentation.

Notations: $\mathbb{N}$ and $\mathbb{R}$ denote the sets of all nonnegative integers and all real numbers respectively. For matrices or vectors, we denote the transpose by $(\cdot)^\top$. For $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{X} \in \mathbb{R}^{N \times M}$, $[\boldsymbol{x}]_n$ and $[\boldsymbol{X}]_{n,m}$ respectively denote the n-th component of $\boldsymbol{x}$ and the (n, m) entry of $\boldsymbol{X}$. We define the support of $\boldsymbol{x} \in \mathbb{R}^N$ by $\mathrm{supp}(\boldsymbol{x}) := \{n \in \{1, \ldots, N\} \mid [\boldsymbol{x}]_n \neq 0\}$. The Euclidean norm and the ℓ_1-norm of $\boldsymbol{x} \in \mathbb{R}^N$ are respectively denoted by $\|\boldsymbol{x}\| := \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$ and $\|\boldsymbol{x}\|_1 := \sum_{n=1}^N \left|[\boldsymbol{x}]_n\right|$. We define the range and the null spaces of $\boldsymbol{X} \in \mathbb{R}^{N \times M}$ respectively by $\mathcal{R}(\boldsymbol{X}) := \{\boldsymbol{X}\boldsymbol{u} \in \mathbb{R}^N \mid \boldsymbol{u} \in \mathbb{R}^M\}$ and $\mathcal{N}(\boldsymbol{X}) := \{\boldsymbol{u} \in \mathbb{R}^M \mid \boldsymbol{X}\boldsymbol{u} = \boldsymbol{0}\}$.

This work was supported in part by JSPS Grants-in-Aid 29-6921, 26730128, 18K18122 and B-15H02757. Hiroki Kuroda is now with the National Institute of Advanced Industrial Science and Technology.

[1] For convenience, we let $n_0^\star := 0$ and $n_{L+1}^\star := N$.

[2] Once $n_1^\star, \ldots, n_L^\star$ are estimated, the estimation of $\boldsymbol{\vartheta}_1^\star, \ldots, \boldsymbol{\vartheta}_{L+1}^\star$ reduces to estimation of ARX process in each time-segment.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

II. Revealing Sparsity for Piecewise ARX Processes

We show a strategic design of the matrix $\boldsymbol{W} \in \mathbb{R}^{(N-K)\times N}$ which makes $\boldsymbol{W}\boldsymbol{s}^\star$ sparse for *noiseless* samples

$$\boldsymbol{s}^\star := (y_1 - \varepsilon_1, \ldots, y_N - \varepsilon_N)^\top \in \mathbb{R}^N$$

in the range space $\mathcal{R}(\boldsymbol{W})$. Since the dimension of $\mathcal{R}(\boldsymbol{W})$ is smaller than N the number of samples, we call $\mathcal{R}(\boldsymbol{W})$ the tight-dimensional space. We exploit the fact that the condition

$$(\exists \ell \in \{1, \ldots, L+1\}) \quad \{n, \ldots, n+K\} \subset (n^\star_{\ell-1}, n^\star_\ell] \tag{2}$$

holds true except for consecutive n's around $n^\star_1, \ldots, n^\star_L$. The condition (2) implies the equation

$$([\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K})^\top = \boldsymbol{\Phi}_n \boldsymbol{\vartheta}^\star_\ell, \tag{3}$$

where

$$\boldsymbol{\Phi}_n := \begin{pmatrix} \boldsymbol{\xi}_n^\top \\ \vdots \\ \boldsymbol{\xi}_{n+K}^\top \end{pmatrix} \in \mathbb{R}^{(K+1)\times K}. \tag{4}$$

Since the equation (3) is similar to that utilized in [17] to reveal sparsity for piecewise continuous signals (see Appendix), we can construct $\boldsymbol{W}$ as shown in the following theorem.

Theorem 1 (Support of $\boldsymbol{W}\boldsymbol{s}^\star$)**.** *Let $\boldsymbol{w}_n \in \mathcal{N}(\boldsymbol{\Phi}_n^\top)\backslash\{\boldsymbol{0}\}$ ($n = 1, \ldots, N-K$) for $\boldsymbol{\Phi}_n$ in (4). For each $n = 1, \ldots, N-K$, construct n-th row of the matrix $\boldsymbol{W} \in \mathbb{R}^{(N-K)\times N}$ by*

$$([\boldsymbol{W}]_{n,1}, \ldots, [\boldsymbol{W}]_{n,N}) := (\underbrace{0, \ldots, 0}_{n-1}, \boldsymbol{w}_n^\top, \underbrace{0, \ldots, 0}_{N-n-K}). \tag{5}$$

Then, the support of $\boldsymbol{W}\boldsymbol{s}^\star$ is covered as

$$\operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star) \subset \{1, \ldots, N-K\} \cap \bigcup_{\ell=1}^{L} \{n^\star_\ell - K + 1, \ldots, n^\star_\ell\}. \tag{6}$$

Remark 1 (Sparsity of $\boldsymbol{W}\boldsymbol{s}^\star$)**.** The inclusion (6) implies that $\boldsymbol{W}\boldsymbol{s}^\star \in \mathbb{R}^{N-K}$ has at most KL nonzero components. Thus, under the condition $KL \ll N$ (i.e. the number of samples is sufficiently many), $\boldsymbol{W}\boldsymbol{s}^\star$ is sparse.

III. Piecewise ARX Process Segmentation by Exploiting Revealed Sparsity

The proposed segmentation method first obtains $\hat{\boldsymbol{s}} \approx \boldsymbol{s}^\star$, and then $\hat{n}_1, \ldots, \hat{n}_{\hat{L}} \approx n^\star_1, \ldots, n^\star_L$. This method is derived based on the algorithm in our previous work [17] for piecewise continuous signal recovery.

A. Support Estimation by Promoting Sparsity

To estimate $\boldsymbol{s}^\star$ or $\operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star)$, motivated by extensive studies on sparse vector estimation in, e.g., [18]–[23], we here present a simple ℓ_1 regularized formulation:

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}\in\mathbb{R}^N} \sum_{n=1}^{N} (y_n - [\boldsymbol{s}]_n)^2 + \lambda \|\boldsymbol{W}\boldsymbol{s}\|_1 \tag{7}$$

where $\|\boldsymbol{W}\boldsymbol{s}\|_1$ promotes the sparsity of $\boldsymbol{W}\boldsymbol{s}$, the standard square error $\sum_{n=1}^{N}(y_n - [\boldsymbol{s}]_n)^2$ is chosen as a data-fidelity term for simplicity, and $\lambda > 0$ is the regularization parameter.

B. Estimation of Change Points

By leveraging Theorem 1 which shows that the locations of change points $n^\star_1, \ldots, n^\star_L$ affects $\operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star)$, we obtain the proposed estimates $\hat{n}_1, \ldots, \hat{n}_{\hat{L}}$ via the following problem.

Problem 1 (Estimation of change points $n^\star_1, \ldots, n^\star_L$)**.** *For $\hat{\boldsymbol{s}} \in \mathbb{R}^N$, find $\hat{L} \in \mathbb{N}$ smallest as possible and $(\hat{n}_1, \ldots \hat{n}_{\hat{L}}) \in \{1, \ldots, N-K\}^{\hat{L}}$ which satisfy*

$$\left.\begin{aligned} \operatorname{supp}(\boldsymbol{W}\hat{\boldsymbol{s}}) \subset \bigcup_{\ell=1}^{\hat{L}} \{\hat{n}_\ell - K + 1, \ldots, \hat{n}_\ell\}, \\ \hat{n}_1 < \hat{n}_2 < \cdots < \hat{n}_{\hat{L}}. \end{aligned}\right\} \tag{8}$$

Since $\boldsymbol{W}\hat{\boldsymbol{s}}$ is given at this step, it is easy to compute a solution of Problem 1, e.g., by the following greedy algorithm. Note that the solution of Problem 1 is unique under the same conditions assumed in Theorem 2 below.

Algorithm 1 (Computation of a Solution of Problem 1)**.** *A solution of Problem 1 can be computed by*

$$\begin{aligned} \hat{n}_{\hat{L}} &= \max\{n \in \{1, \ldots, N-K\} \mid [\boldsymbol{W}\hat{\boldsymbol{s}}]_n \neq 0\}, \\ \hat{n}_{\hat{L}-1} &= \max\{n \in \{1, \ldots, \hat{n}_{\hat{L}} - K\} \mid [\boldsymbol{W}\hat{\boldsymbol{s}}]_n \neq 0\}, \\ &\vdots \\ \hat{n}_1 &= \max\{n \in \{1, \ldots, \hat{n}_2 - K\} \mid [\boldsymbol{W}\hat{\boldsymbol{s}}]_n \neq 0\}, \end{aligned}$$

where $\hat{L}$ is defined so that $\{n \in \{1, \ldots, \hat{n}_1 - K\} \mid [\boldsymbol{W}\hat{\boldsymbol{s}}]_n \neq 0\} = \varnothing$.

We show the estimation accuracy of $\hat{n}_1, \ldots, \hat{n}_{\hat{L}}$ for the case $\operatorname{supp}(\boldsymbol{W}\hat{\boldsymbol{s}}) = \operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star)$. Since the estimation accuracy depends on the solution of Problem 1 which is not always unique, for simplicity, we assume conditions for its uniqueness:

i) At least K sample indexes are contained in every time-segment: $(n^\star_{\ell-1}, n^\star_\ell]$ ($\ell = 2, \ldots, L$), $(0, n^\star_1]$ and $(n^\star_L, N]$.
ii) $[\boldsymbol{W}\boldsymbol{s}^\star]_{n^\star_\ell} \neq 0$ and $[\boldsymbol{W}\boldsymbol{s}^\star]_{n^\star_\ell - K + 1} \neq 0$ ($\ell = 1, \ldots, L$).

The condition (i) likely holds if N is sufficiently many. The condition (ii) also likely holds because $[\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K}$ sampled from multiple time-segments mostly violates the condition (10) shown in Appendix.

Theorem 2 (Accuracy of change point estimation)**.** *Suppose $\operatorname{supp}(\boldsymbol{W}\hat{\boldsymbol{s}}) = \operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star)$. Then, under the conditions (i) and (ii) above, the solution of Problem 1 is unique and satisfies*

$$\begin{aligned} \hat{L} &= L, \\ \hat{n}_\ell &= n^\star_\ell \quad (\ell = 1, \ldots, L). \end{aligned}$$

In particular, Theorem 2 suggests that the estimation of $\operatorname{supp}(\boldsymbol{W}\boldsymbol{s}^\star)$ is the key step for estimation of $n^\star_1, \ldots, n^\star_L$.

Remark 2 (Computational costs of proposed method)**.** The proposed method can be divided into the following three steps:

1. Construct $\boldsymbol{W}$ as shown in Theorem 1.
2. Obtain $\hat{\boldsymbol{s}}$ by the convex optimization shown in (7).
3. Obtain $\hat{n}_1, \ldots, \hat{n}_{\hat{L}}$ from $\hat{\boldsymbol{s}}$ by Algorithm 1.

Computation needed for step 1 is to find $\boldsymbol{w}_n \in \mathcal{N}(\boldsymbol{\Phi}_n^\top)\backslash\{\boldsymbol{0}\}$ for each $n = 1, \ldots, N-K$. This computational cost is low because $\mathbf{\Phi}_n^\top$ is a K by $K+1$ matrix, and K is at most tens in practice (see, e.g., [1]–[9]). Step 2 requires to solve a nonsmooth convex optimization problem, but many efficient solvers are available, e.g., [24]–[27]. For instance, with the first order methods [24], [25], we can implement an iterative solver for (7) which needs only $\mathcal{O}(KN)$ arithmetic operations per iteration because $\boldsymbol{W}$ has at most $K+1$ nonzero entries in each row (see (5)). Finally, computational cost for step 3 is very low because this step simply searches nonzero components of $\boldsymbol{W}\hat{\boldsymbol{s}}$.

IV. Comparison with Segmentation Methods Using Sparsity in Redundant Representation Spaces

The approaches in [3], [5], [12], [13] introduce redundant representations $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_N \in \mathbb{R}^K$ which are trained so that $[\boldsymbol{s}^\star]_n \approx \boldsymbol{\xi}_n^\top \hat{\boldsymbol{\beta}}_n$ $(n = 1, \ldots, N)$. In this redundant representation, it is expected that $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_{n+1}$ can be same for reproducing $[\boldsymbol{s}^\star]_n$ and $[\boldsymbol{s}^\star]_{n+1}$. Based on this observation, $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_N$ are obtained, e.g., by solving

$$\min_{(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N)\in\mathbb{R}^{K\times N}} \sum_{n=1}^{N} (y_n - \boldsymbol{\xi}_n^\top \boldsymbol{\beta}_n)^2 + \kappa \sum_{n=1}^{N-1} \|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n+1}\|, \quad (9)$$

which we call modified-Total-Variation (TV) formulation, where $\kappa > 0$ is the regularization parameter. Modified-TV aims to promote sparsity of $(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n+1})_{n=1}^{N-1}$ the $K(N-1)$-dimensional vector. From $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_N$ the solution of (9), change points are estimated by finding $n \in \{1, \ldots, N-1\}$ such that $\|\hat{\boldsymbol{\beta}}_n - \hat{\boldsymbol{\beta}}_{n+1}\| > 0$.

Meanwhile, the proposed approach reveals sparsity hidden in $\boldsymbol{s}^\star$ in the tight-dimensional range space $\mathcal{R}(\boldsymbol{W})$ whose dimension does not exceed N. Compared with modified-TV, the tight-dimensionality of the proposed approach is a great advantage for computation (see Remark 3 below). In addition, it is hypothesized that the superior performances of the proposed method for numerical examples shown in Sect. V is thanks to the tight-dimensionality.

Remark 3 (Comparison of computational costs)**.** In both the proposed method and modified-TV, computational costs are dominated by convex optimization problems respectively shown in (7) and (9). The tight-dimensionality of the proposed approach yields that the number of variables to be optimized in the proposed formulation (7) is K times less than that of (9). Thus, although computation time depends on a employed solver, the proposed formulation (7) is expected to be solved faster than (9).

V. Numerical Experiments

To show the effectiveness of the proposed method, we present experiments on segmentation of synthetic and real-world data. The proposed method first estimates $\boldsymbol{s}^\star$ by (7), and then estimates $\hat{n}_1, \ldots, \hat{n}_{\hat{L}}$ by Algorithm 1. Note that, for construction of $\boldsymbol{W}$ shown in Theorem 1, there exists an arbitrariness of scalar multiplication in $\boldsymbol{w}_n \in \mathcal{N}(\mathbf{\Phi}_n^\top)\backslash\{\mathbf{0}\}$. As a typical example, we here use $\boldsymbol{w}_n$ of unit Euclidean norm $\|\boldsymbol{w}_n\| = 1$ $(n = 1, \ldots, N-K)$. We compare the proposed method against modified-TV formulation, shown in (9), promoting the sparsity in the redundant representation space introduced in [3], [5], [12], [13].

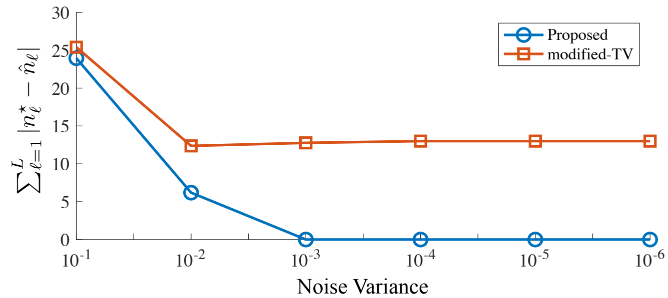


Fig. 1: Comparison of the proposed method and modified-TV for segmentation of synthetic data where results are averaged over 100 realizations of white Gaussian noise for each noise variance.

A. Synthetic Example

In (1), we set $\boldsymbol{\xi}_n = (y_{n-1}, \ldots, y_{n-4}, x_{n-1})^\top$, $n_1^\star = 40$, $n_2^\star = 70$, $\boldsymbol{\vartheta}_1^\star = (3.0797, -4.2766, 3.0012, -0.9475, 0.1)^\top$, $\boldsymbol{\vartheta}_2^\star = (2.6916, -3.6977, 2.6235, -0.9477, 0.1)^\top$, $\boldsymbol{\vartheta}_3^\star = (2.8945, -3.9908, 2.8210, -0.9476, 0.1)^\top$ with $N = 100$. The results are averaged for 100 realizations of the white Gaussian noise $(\varepsilon_n)_{n=1}^N$, while the input signals $(x_n)_{n=0}^{N-1}$ are fixed to values drawn from i.i.d. Gaussian distribution $\mathcal{N}(0, 1)$.

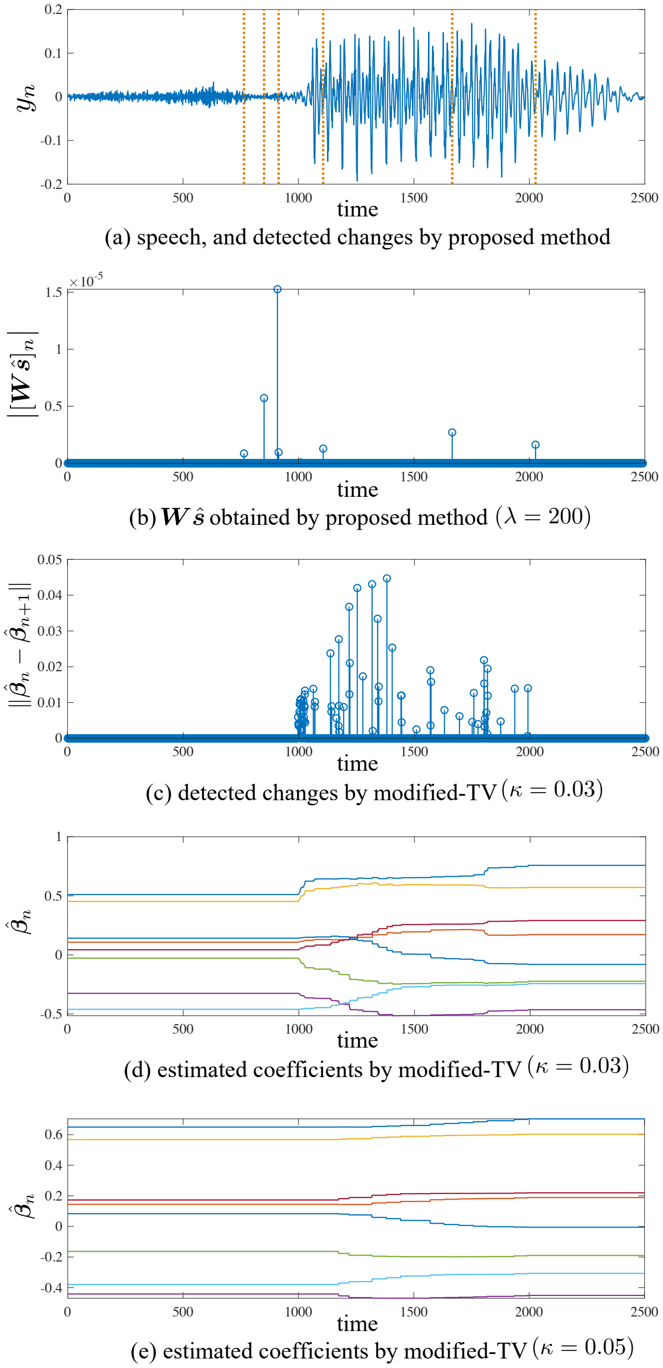
To simply measure the difference between $\hat{n}_1, \ldots \hat{n}_{\hat{L}}$ and $n_1^\star, \ldots, n_L^\star$, it is convenient to make $\hat{L} = L$. Thus, we here modify Problem 1 by approximating the relation (8) with $(\hat{n}_1, \ldots \hat{n}_L) \in \{1, \ldots, N-K\}^L$ chosen by

$$\min_{\hat{n}_1 < \cdots < \hat{n}_L} \sum_{n \notin \bigcup_{\ell=1}^{L} \{\hat{n}_\ell - K + 1, \ldots, \hat{n}_\ell\}} \left|[\boldsymbol{W}\hat{\boldsymbol{s}}]_n\right|.$$

Note that, since $\hat{\boldsymbol{s}}$ is given at this step, this minimization can be easily done in a greedy way. Then, we measure the error by $\sum_{\ell=1}^{L} |n_\ell^\star - \hat{n}_\ell|$. Similar modification is also performed for $(\|\hat{\boldsymbol{\beta}}_n - \hat{\boldsymbol{\beta}}_{n+1}\|)_{n=1}^{N-1}$ obtained by modified-TV (9). Note that regularization parameters respectively in the proposed formulation and modified-TV are chosen so that the performances become best. Fig. 1 shows that there exists a case where modified-TV fails to detect change points even if noise is small enough. Meanwhile, the proposed method precisely estimates change point locations for small to moderate levels of noise.

B. Speech Segmentation

As an instance of real-world data, we show experiments on phonetic segmentation of English male and female speech respectively pronouncing /seɪm/ and /klaʊd/. Data are taken from [28] with the 8kHz sampling rate. According to the setting in [3], we here adopt the piecewise AR model of order 8, i.e., $\boldsymbol{\xi}_n = (y_{n-1}, \ldots, y_{n-8})^\top$. As shown in Fig. 2(c), modified-TV detects too many change point locations (Note: nonzero components in Fig. 2(c) indicate the change point locations), i.e., provides inadequate information for speech segmentation. This is caused by the situation that $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_N$ obtained by modified-TV gradually change (see Fig. 2(d)),



(a) speech, and detected changes by proposed method

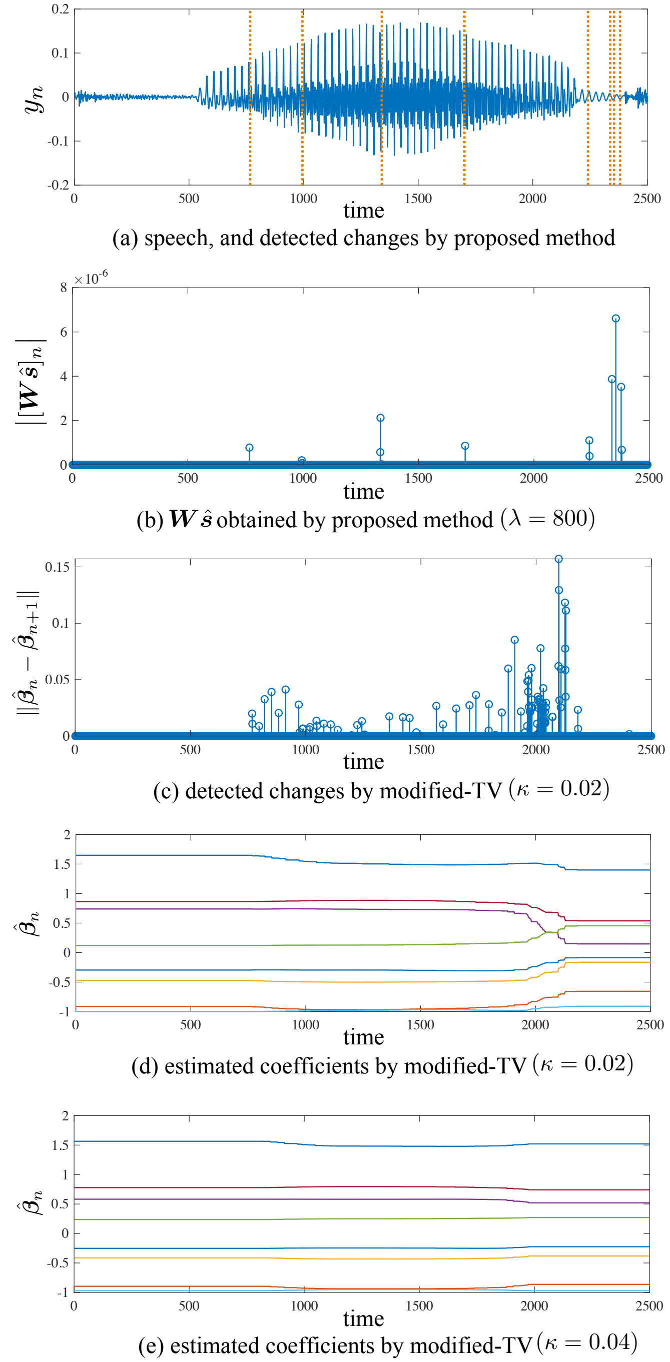
(b) $\boldsymbol{W}\hat{\boldsymbol{s}}$ obtained by proposed method ($\lambda = 200$)

(c) detected changes by modified-TV ($\kappa = 0.03$)

(d) estimated coefficients by modified-TV ($\kappa = 0.03$)

(e) estimated coefficients by modified-TV ($\kappa = 0.05$)

Fig. 2: Comparison of the proposed method and modified-TV for segmentation of English male speech where phoneme boundaries between /s/, /e/, /ɪ/, and /m/ would be around 1050, 1600, and 2000 samples.



(a) speech, and detected changes by proposed method

(b) $\boldsymbol{W}\hat{\boldsymbol{s}}$ obtained by proposed method ($\lambda = 800$)

(c) detected changes by modified-TV ($\kappa = 0.02$)

(d) estimated coefficients by modified-TV ($\kappa = 0.02$)

(e) estimated coefficients by modified-TV ($\kappa = 0.04$)

Fig. 3: Comparison of the proposed method and modified-TV for segmentation of English female speech where phoneme boundaries between /k/, /l/, /aʊ/, and /d/ would be around 550, 900, and 2400 samples.

contrary to the intension of modified-TV (see Sect. IV). This issue is not resolved by increasing κ the regularization parameter (see Fig. 2(e)). Similar tendencies are also observed in Fig. 3(c)(d)(e). Meanwhile, as shown in Fig. 2(a)(b) and Fig. 3(a)(b), the proposed method detects several change point locations. From inspection of the speech, detections by the proposed method seem to reasonably reflect phoneme boundaries. For the speech /seɪm/, detections by the proposed method adequately matches phoneme boundaries except that several changes are detected in /s/. We observe that the first and latter parts of /s/ have rather different acoustical properties, and thus detections by the proposed method would be reasonable in terms of acoustical changes in speech. For segmentation of the speech /klaʊd/ which is more challenging due to diphthong /aʊ/, the proposed method performs fairly well, though the detected boundary between /k/ and /l/ is slightly different from the true one, and several changes are detected in /aʊ/.

VI. Conclusion

For segmentation of the piecewise ARX process shown in (1), in the tight-dimensional representation spaces, we presented an idea to exploit sparsity hidden in $\boldsymbol{s}^\star$ consisting of samples following the piecewise ARX process. More precisely, by exploiting the fact that (2) is expected to hold for most of $n \in \{1, \ldots, N-K\}$, we design the linear transformation $\boldsymbol{W}$ which makes $\boldsymbol{W}\boldsymbol{s}^\star$ sparse in the tight-dimensional range space $\mathcal{R}(\boldsymbol{W})$ whose dimension is smaller than the number of samples. We also show how the revealed sparsity can be exploited for piecewise ARX process segmentation. Numerical examples demonstrate the effectiveness of the proposed method.

Appendix
Essence to Reveal Sparsity in Tight-Dimensional Representation Spaces

In our previous work [17], we reveal sparsity in tight-dimensional spaces for certain piecewise continuous signals where the equation similar to (3) is utilized. We explain the idea in [17] for the setting in this paper. First, the equation (3) implies the inclusion

$$([\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K})^\top \in \mathcal{R}(\boldsymbol{\Phi}_n) = \mathcal{N}(\boldsymbol{\Phi}_n^\top)^\perp, \tag{10}$$

where $\mathcal{N}(\boldsymbol{\Phi}_n^\top)^\perp$ is the orthogonal complement of $\mathcal{N}(\boldsymbol{\Phi}_n^\top)$. The inclusion (10) is equivalent to

$$\boldsymbol{u}^\top([\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K})^\top = 0 \quad (\forall \boldsymbol{u} \in \mathcal{N}(\boldsymbol{\Phi}_n^\top)).$$

Suppose the technical assumption:

Assumption 1. *For every $n \in \{1, \ldots, N-K\}$, the matrix $\boldsymbol{\Phi}_n$ in* (4) *is full-column rank, i.e., $\mathcal{N}(\boldsymbol{\Phi}_n^\top)$ is 1-dimensional.*

This assumption holds almost surely, roughly speaking, if input signals $(x_n)_{n=1-q_2}^{N-1}$ and noises $(\varepsilon_n)_{n=1}^{N}$ in (1) are independently drawn from continuous probability distributions. Then, we deduce that, for any $\boldsymbol{w}_n \in \mathcal{N}(\boldsymbol{\Phi}_n^\top)\backslash\{\boldsymbol{0}\}$,

$$\left.\begin{array}{r}([\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K}) \text{ satisfies Condition (10)} \\ \Leftrightarrow \boldsymbol{w}_n^\top([\boldsymbol{s}^\star]_n, \ldots, [\boldsymbol{s}^\star]_{n+K})^\top = 0.\end{array}\right\} \tag{11}$$

Note that "$\Rightarrow$" in (11) holds true without Assumption 1. From this relation and the simple observation shown in (2), for $\boldsymbol{W}$ constructed by (5) in Theorem 1, $[\boldsymbol{W}\boldsymbol{s}^\star]_n \neq 0$ happens only in consecutive n's around $n_1^\star, \ldots, n_L^\star$. This implies that sparsity is revealed in the range space $\mathcal{R}(\boldsymbol{W})$. This fact is more precisely expressed in Theorem 1 and Remark 1. We say that $\mathcal{R}(\boldsymbol{W})$ is the tight-dimensional space because the dimension of $\mathcal{R}(\boldsymbol{W})$ does not exceed N the number of samples.

References

[1] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Process.*, vol. 5, no. 3, pp. 267–285, 1983.

[2] R. Andre-Obrecht, "A new statistical approach for the automatic segmentation of continuous speech signals," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 36, no. 1, pp. 29–40, Jan 1988.

[3] D. Angelosante and G. B. Giannakis, "Group lassoing change-points in piecewise-constant AR processes," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, March 2012.

[4] R. Vidal, S. Soatto, and A. Chiuso, "Applications of hybrid system identification in computer vision," in *European Control Conference*, July 2007, pp. 4853–4860.

[5] N. Ozay, M. Sznaier, C. M. Lagoa, and O. I. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Trans. Automat. Ctrl.*, vol. 57, no. 3, pp. 634–648, March 2012.

[6] G. Bodenstein and H. M. Praetorius, "Feature extraction from the electroencephalogram by adaptive segmentation," *Proc. IEEE*, vol. 65, no. 5, pp. 642–652, May 1977.

[7] F. Gustafsson, *Adaptive filtering and change detection*. Wiley, 2000.

[8] B. Liu, H. Chen, A. Sharma, G. Jiang, and H. Xiong, "Modeling heterogeneous time series dynamics to profile big sensor data in complex physical systems," in *IEEE Int. Conf. Big Data*, Oct. 2013, pp. 631–638.

[9] Z. Han, H. Chen, T. Yan, and G. Jiang, "Time series segmentation to discover behavior switching in complex physical systems," in *IEEE Int. Conf. Data Mining*, Nov. 2015, pp. 161–170.

[10] R. Bellman and R. Roth, "Curve fitting by segmented straight lines," *J. Amer. Statist. Assoc.*, vol. 64, no. 327, pp. 1079–1084, 1969.

[11] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1365–1373, May 1998.

[12] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107–1111, 2010.

[13] B. Zhang, J. Geng, and L. Lai, "Multiple change-points estimation in linear regression models via sparse group lasso," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2209–2224, May 2015.

[14] T. Nir and A. M. Bruckstein, "On over-parameterized model based TV-denoising," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 1, July 2007, pp. 1–4.

[15] D. Angelosante, G. B. Giannakis, and N. D. Sidiropoulos, "Estimating multiple frequency-hopping signal parameters via sparse linear regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5044–5056, Oct. 2010.

[16] R. Giryes, M. Elad, and A. M. Bruckstein, "Sparsity based methods for overparameterized variational problems," *SIAM J. Imag. Sci.*, vol. 8, no. 3, pp. 2133–2159, 2015.

[17] H. Kuroda, M. Yamagishi, and I. Yamada, "Exploiting sparsity in tight-dimensional spaces for piecewise continuous signal recovery," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6363–6376, Dec. 2018.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[20] J. J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.

[21] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[22] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inv. Probl.*, vol. 23, no. 3, pp. 947–968, 2007.

[23] S. Vaiter, G. Peyre, C. Dossal, and J. Fadili, "Robust sparse analysis regularization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2001–2016, April 2013.

[24] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013.

[25] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[26] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[27] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 1, pp. 293–318, April 1992.

[28] P. Kabal, "TSP speech database," Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, Tech. Rep., 2002.