Local Polynomial Modeling and Variable Bandwidth Selection for Time-Varying Linear Systems

S. C. Chan, *Member, IEEE*, and Z. G. Zhang, *Member, IEEE*

***Abstract*—This paper proposes a local polynomial modeling (LPM) approach and variable bandwidth selection (VBS) algorithm for identifying time-varying linear systems (TVLSs). The proposed method models the time-varying coefficients of a TVLS locally by polynomials, which can be estimated by least squares estimation with a kernel having a certain bandwidth. The asymptotic behavior of the proposed LPM estimator is studied, and the existence of an optimal local bandwidth which minimizes the local mean-square error is established. A new data-driven VBS algorithm is then proposed to estimate this optimal variable bandwidth adaptively and locally. An individual bandwidth is assigned for each coefficient instead of the whole coefficient vector so as to improve the accuracy in fast-varying systems encountered in fault detection and other applications. Important practical issues such as online implementation are also discussed. Simulation results show that the LPM-VBS method outperforms conventional TVLS identification methods, such as the recursive least squares algorithm and generalized random walk Kalman filter/smoother, in a wide variety of testing conditions, in particular, at moderate to high signal-to-noise ratio. Using local linearization, the LPM method is further extended to identify time-varying systems with mild nonlinearities. Simulation results show that the proposed LPM-VBS method can achieve a satisfactory performance for mildly nonlinear systems based on appropriate linearization. Finally, the proposed method is applied to a practical problem of voltage-flicker-tracking problem in power systems. The usefulness of the proposed approach is demonstrated by its improved performance over other conventional methods.**

***Index Terms*—Bandwidth selection, least squares (LS), local polynomial modeling (LPM), maximum-likelihood estimation, system identification, time-varying linear systems (TVLSs).**

I. Introduction

DYNAMIC systems with time-varying behaviors, such as time-varying impedances in coaxial resonators [1], time-varying channels in wireless-communication systems [2], nonstationary mechanisms of physiological systems [3], fluctuations in power distribution systems [4], etc., are frequently encountered in various engineering applications. Given a dynamic system with known input and output measurements, it is crucial to accurately identify the underlying dynamics of the system for predicting future measurements and detecting system variations, for example, for fault detection, etc. The discrete-time time-varying linear system (TVLS) is a simple yet efficient model to characterize such a dynamic system [2]–[6]. Let $\boldsymbol{x}(t) = [x(1,t), x(2,t), \ldots, x(L,t)]^T$ be the known input vector to the TVLS. The measured output $y(t)$ of the system can be written as

$$y(t) = \sum_{k=1}^{L} a(k,t)x(k,t) + \sigma(t)\varepsilon(t) \tag{1}$$

or, in the matrix form

$$y(t) = \boldsymbol{a}^T(t)\boldsymbol{x}(t) + \sigma(t)\varepsilon(t) \tag{2}$$

where $\boldsymbol{a}(t) = [a(1,t), a(2,t), \ldots, a(L,t)]^T$ is the time-varying coefficient vector of the system, which is assumed to be a function of discrete-time instant t, L is the order of the system, $\sigma^2(t_0)$ is the conditional variance of additive noise at $t = t_0$, and $\varepsilon(t)$ is an additive noise, which is frequently modeled as a zero-mean white Gaussian process with unit variance. The input vector $\boldsymbol{x}(t)$ may contain a regression on $y(t)$, and it can also contain exogenous signals. For instance, if the input is $x(k,t) = y(t-k)$, the TVLS will be reduced to an autoregressive (AR) model. Similarly, if the input contains both past samples of $y(t)$ and exogenous signals, the AR with exogenous input model can also be regarded as a TVLS. For an AR moving average model (ARMA) or an ARMA with exogenous input (ARMAX) model, the system is driven by a zero-mean white Gaussian process. If the excitation is estimated at each time instant, then the past approximated excitation can be included in the input vector $\boldsymbol{x}(t)$, and both ARMA and ARMAX models can also be rewritten in the form of a TVLS. However, the estimated coefficients may not always converge to the true coefficients, and such application of the TVLS to ARMA and ARMAX models still requires a more detailed analysis.

Numerous methods have been proposed to identify or estimate the time-varying coefficients $\boldsymbol{a}(t)$ of the TVLS in (1). They can be broadly classified into three categories [5]: adaptive filtering/Kalman filtering (KF), basis expansion modeling (BEM), and weighted least square (LS) (WLS) methods. Adaptive-filtering methods, such as the least mean squares and recursive least squares (RLS) methods, estimate the coefficients recursively from the input and measured output. They usually offer efficient implementation with different performance/complexity tradeoffs. Most adaptive-filtering methods make use of past measurements for estimation, and the convergence speed

Manuscript received April 12, 2010; revised June 12, 2010; accepted July 19, 2010. This work was supported in part by the Grant from the Research Grants Council of Hong Kong and in part by The University of Hong Kong CRCG Small Project Funding. The Associate Editor coordinating the review process for this paper was Dr. Tadeusz Dobrowiecki.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong (e-mail: scchan@eee.hku.hk; zgzhang@eee.hku.hk).

Digital Object Identifier 10.1109/TIM.2010.2064850

is therefore limited [8], [9]. On the other hand, the KF method employs a state-space model to describe the coefficient variations where the current coefficients are treated as the system state and are obtained by a linear transformation of the previous coefficients or state plus an excitation variable through the state-space model. Given prior information of the coefficient variations in the form of the state-space model and the covariance matrices of the Gaussian distributed excitation and measurement noises, the KF is an optimal recursive state estimator in the minimum mean-square-error (MSE) sense. However, such prior knowledge is often vague in real-world applications. To address this, a generalized random walk Kalman filter (GRWKF), which uses a generalized random walk model to describe the variation of the system state, is widely used in practice for its simplicity and efficiency [5], while other KF variants aim to adaptively estimate the model recursively [10], [11]. In the BEM method, an explicit deterministic model of the coefficient variations is assumed, and the time-varying coefficients are approximated by a linear combination of known basis functions of time [12], [13]. The performance of the BEM method is greatly dependent on the basis functions used, and its optimal selection is not always accessible. The WLS method is similar to the conventional LS method, except that kernels or windows are employed to assign larger weights to local data and smaller weights to remote data. The time-varying coefficients are then estimated by minimizing a weighted sum-of-square estimation errors. The selection of the window size or kernel bandwidth is critical to the performance of the WLS method so that an appropriate compromise between estimation accuracy (variance) and modeling error (bias) can be achieved [5]. Unfortunately, automatic data-driven kernel-bandwidth selection for WLS is a difficult and open problem, which hinders considerably its practical implementation.

In this paper, a new local polynomial modeling (LPM) approach is proposed to estimate the time-varying coefficients of TVLSs. The LPM technique is originally a flexible and efficient nonparametric approach in statistics [14]–[16] and has been widely applied in data smoothing, derivative estimation, density estimation, etc. [17]–[20]. To deal with the TVLSs, the proposed LPM method models each element of the time-varying coefficient vector locally by a set of polynomials with a kernel having a certain bandwidth. Consequently, the estimation of time-varying coefficients is reduced to the estimation of the local polynomial coefficients, which can be easily performed using the LS method.

To establish the asymptotic behaviors of the proposed LPM estimator for TVLS, new asymptotic expressions for the estimation bias and variance are derived. The asymptotic expressions show that both the bias and variance are functions of the kernel bandwidth, and there exists an optimal local bandwidth which minimizes the MSE of each element in the coefficient vector at each time instant. While the analytical formulas derived are useful for theoretical work, it involves quantities which may not be easily estimated in practice. Therefore, a "data-driven" variable bandwidth selection (VBS) scheme is proposed for selecting the local bandwidth in the LPM method. The basic idea of the data-driven or empirical scheme is to approximate the bias and variance of the LPM for a given bandwidth. Hence, an estimate of the optimal bandwidth can be obtained by minimizing the approximate MSE over a bandwidth set, which consists of a finite set of candidate bandwidth values. Following the classical approach in [16], a novel algorithm to approximate the MSE is proposed. It employs a "pilot" LPM of the system with a slightly higher polynomial order to estimate the bias and variance required in the approximated MSE from which the optimal candidate bandwidth from the bandwidth set can be determined. Since this "pilot fit" also requires a pilot bandwidth, we propose to adopt the intersection of confidence interval (ICI) method [18]–[21] for choosing this pilot bandwidth because of its good performance and simple implementation. The ICI method is an empirical adaptive bandwidth-selection method proposed by Goldenshluger and Nemirovski [21]. It has been successfully employed in various areas, including local polynomial regression, image processing, and time–frequency analysis [18]–[20]. Interested readers are referred to [18]–[20] for more details. To further improve the estimation accuracy of the proposed LPM method with VBS (LPM-VBS), each coefficient in the coefficient vector of the LPM model is assigned a separate bandwidth, instead of using a global bandwidth for the whole coefficient vector. This facilitates the estimation and change detection of individual coefficient in the coefficient vector, which is very useful in fault detection and other applications. The performance of the proposed LPM-VBS method was evaluated using various types of simulated TVLSs, and the results show that the proposed method yields more accurate estimates than conventional TVLS identification methods such as the RLS, GRWKF, and generalized random walker Kalman smoother (GRWKS) algorithms in testing scenarios with moderate to high signal-to-noise ratio (SNR). At low coefficient variation, the GRWKS with smoothness-prior-constrained parameter estimation may perform slightly better than the proposed LPM method. At fast coefficient variation and low SNR, GRWKS with recursive parameter estimation may perform slightly better than LPM. Therefore, their good performance is considerably dependent on the appropriate selection or estimation of the model parameters, which is somewhat difficult to achieve.

Furthermore, we extend the proposed LPM method to handle time-varying systems with mild nonlinearity by means of linearization. A time-varying quadratic system was tested to demonstrate the usefulness of the LPM method for mildly nonlinear systems. Lastly, we applied the LPM-VBS method to track the instantaneous voltage flicker in power distribution systems. Voltage flickers may cause serious power-quality problems, and they need to be estimated accurately. Simulation results show that the proposed LPM-VBS method has better performance than conventional methods in tracking the envelope of voltage flickers. These results suggest that the proposed LBM-VBS method can be a valuable tool in modeling time-varying systems for various engineering and industrial applications.

It should be noted that the TVLS considered in the engineering field is closely related to the varying-coefficient model (VCM) [22] in the statistical community. As LPM is a powerful nonparametric modeling technique, it has received much attention and success in various branches of statistics [16]. Recently, local-polynomial-based estimators for VCM have been studied

in [23]–[26] by researchers in the statistical community. They were mainly concerned with the algorithmic aspects of the method, and a theoretically supported bandwidth selector has not been fully developed. Moreover, their performance for identifying TVLSs has not been tested and compared with other identification methods in the engineering literature. This paper is the first attempt to tackle the VBS problem for LPM with the following contributions: 1) An asymptotic analysis of the LPM estimator is developed, and the existence of the optimal bandwidth for each coefficient at each time instant is established; 2) a fully data-driven bandwidth selector is proposed, and related issues in practical implementation, particularly online, are addressed; 3) each coefficient in the coefficient vector of the LPM model is assigned a separate bandwidth, instead of using a global bandwidth for the whole coefficient vector, to improve the estimation accuracy, for example, for fault detection and location, etc.; 4) a detailed performance evaluation of the proposed method and other conventional methods for system identification in a variety of testing scenarios is carried out; 5) the LPM method is extended to the identification of time-varying systems with known and mild nonlinearity; and 6) it was applied to the tracking problem of voltage flickers in power delivery systems with improved performance over conventional methods.

The rest of this paper is organized as follows. In Section II, the LPM for TVLSs is introduced. Section III is devoted to the asymptotic analysis of the LPM estimator. The adaptive VBS method for LPM is developed in Section IV. The LPM-VBS method is further extended to time-varying and mildly nonlinear systems in Section V. Simulation results and comparisons to conventional methods are presented in Section VI. Finally, conclusions are drawn in Section VII.

II. LPM OF TVLS

In the proposed LPM method, the kth coefficient $a(k,t)$ of the TVLS in (1) is modeled locally at time $t = t_0$ as a pth-order polynomial [14], [15]

$$a(k,t) \approx \sum_{j=0}^{p} \frac{1}{j!}\alpha^{(j)}(k,t_0)(t-t_0)^j \tag{3}$$

where $\alpha^{(j)}(k,t_0)$ are the associated polynomial coefficients. These polynomial coefficients can be estimated locally by maximum-likelihood estimation. Since the additive noise is zero mean and white Gaussian distributed, maximizing the likelihood is equivalent to minimizing a locally weighted LS criterion between the observations and the desired local polynomials as follows:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{L}\sum_{j=0}^{p} \beta^{(j)}(k,t_0)(t_i-t_0)^j x(k,t_i) \right]^2 K_h(t_i-t_0) \tag{4}$$

where $y_i = y(t_i)$, n is the data length, $\beta^{(j)}(k,t_0) = \alpha^{(j)}(k,t_0)/j!$, and $K_h(t_i-t_0) = (1/h)K((1/h)(t_i-t_0))$ is a weighting function which controls the bandwidth h and, hence, the number of neighboring measurements around t_0 used to estimate $\beta^{(j)}(k,t_0)$. It can be seen that the weight function or kernel $K_h(\cdot)$ is obtained by scaling a basis kernel function $K(\cdot)$ in time by a factor of h.

Next, we rewrite (4) more compactly in a matrix form as

$$\min_{\boldsymbol{\beta}} \big(\boldsymbol{y} - \boldsymbol{X}(t_0)\boldsymbol{B}(t_0)\big)^T \boldsymbol{W}(t_0)\big(\boldsymbol{y} - \boldsymbol{X}(t_0)\boldsymbol{B}(t_0)\big) \tag{5}$$

where the definitions are shown at the bottom of the page, with $\boldsymbol{\beta}^{(j)}(t_0) = [\beta^{(j)}(1,t_0),\ldots,\beta^{(j)}(L,t_0)]^T$. It can be seen that $\boldsymbol{B}(t_0)$, $\boldsymbol{X}(t_0)$, and $\boldsymbol{W}(t_0)$ are all functions of t_0, but for notation simplicity, we have dropped this dependence in the subsequent text.

The LS solution to (5) is given by

$$\hat{\boldsymbol{B}} = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} \tag{6}$$

and the kth coefficient at time instant t_0 is obtained as $\hat{a}(k,t_0) = \hat{\alpha}^{(0)}(k,t_0) = \hat{\boldsymbol{B}}(k,t_0)$. By estimating $\hat{a}(k,t_0)$ at each time instant, we obtain a smooth function of the time-varying coefficients from the input x and the noisy output y.

The LPM method can be viewed as a combination of WLS and BEM with polynomial basis functions. It extends the BEM method by employing a kernel function to put more emphasis on local information, and it extends the WLS method by employing a high-order polynomial expansion in the regression. When $p = 0$, LPM is reduced to the conventional WLS method. The Taylor series expansion, which grounds the LPM method, is the most fundamental and commonly used technique to

---

$$\begin{aligned}
\boldsymbol{y} &= [y_1, y_2, \ldots, y_n]^T \in \boldsymbol{R}^n \\
\boldsymbol{W}(t_0) &= \operatorname{diag}\{K_h(t_i-t_0); i = 1,\ldots,n\} \in \boldsymbol{R}^{n\times n} \\
\boldsymbol{X}(t_0) &= \begin{pmatrix} \boldsymbol{x}^T(t_1) & (t_1-t_0)\boldsymbol{x}^T(t_1) & \cdots & (t_1-t_0)^p\boldsymbol{x}^T(t_1) \\ \boldsymbol{x}^T(t_2) & (t_2-t_0)\boldsymbol{x}^T(t_2) & \cdots & (t_2-t_0)^p\boldsymbol{x}^T(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}^T(t_n) & (t_n-t_0)\boldsymbol{x}^T(t_n) & \cdots & (t_n-t_0)^p\boldsymbol{x}^T(t_n) \end{pmatrix} \in \boldsymbol{R}^{n\times(p+1)L} \\
\boldsymbol{B}(t_0) &= \left\{ \left[\boldsymbol{\beta}^{(0)}(t_0)\right]^T, \ldots, \left[\boldsymbol{\beta}^{(p)}(t_0)\right]^T \right\}^T \in \boldsymbol{R}^{(p+1)L}
\end{aligned}$$

approximate a smooth function when the underlying structure of the function is unknown. An important advantage of the proposed LPM method is that the bias and variance can be analytically derived, which paves the way to the solution of the key problem of automatic data-driven VBS. For the KF and BEM methods, the performance will heavily rely on a specified stochastic or deterministic model of the system. When an explicit model of the system is unavailable, the generalized random walk model in the GRWKF aims to stabilize the state update by restricting the variation by the simple state equation. This can be viewed as a regularization term and helps to reduce the variance of the state estimates. On the other hand, the LPM method aims to build a nonparametric model using local polynomial representation from the available data so that all available measurements may be taken into account during coefficient estimation.

Moreover, due to the analytical simplicity of local polynomial, both asymptotic and practical methods for estimating the bandwidth can be obtained, as illustrated in this paper. In short, the BEM and KF methods rely more on "explicit" models, while the LPM estimator relies more on "data-dependent" models. Therefore, the LPM method is a good alternative to conventional approaches for identifying TVLSs, particularly when the underlying coefficient model is unavailable and the measurement is reliable, i.e., the SNR is large. Finally, it is noted that it is possible to develop a novel state-space model which employs the LPM-estimated priors at each time instant as the state equation while having a variable number of measurement equations. This will be useful at low SNR because of the additional prior state information. However, since the number of measurements is variable at each time instant and the complexity to estimate the model parameters such as noise covariance will increase, we have not pursued such direction in this paper. Simulation results show that the performance of the proposed method is very satisfactory at moderate to high SNR.

In the next section, we shall derive new expressions for the asymptotic bias and variance of the LS estimator in (6), aiming to validate the existence of the optimal bandwidth parameters and to establish its asymptotic expressions. The data-driven or empirical method for approximating this optimal bandwidth parameter h will then be discussed in Section IV.

III. Asymptotic Analysis of LPM Estimator

The conditional bias based on $\Xi = \{\boldsymbol{X}, \boldsymbol{y}\}$ can be obtained by taking the expectation of (6) as

$$\begin{aligned} E(\hat{\boldsymbol{B}}|\Xi) &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{m} \\ &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{m}-\boldsymbol{X}\boldsymbol{B}) + \boldsymbol{B} \\ &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{r} + \boldsymbol{B} \end{aligned} \tag{7}$$

where $\boldsymbol{m} = E(\boldsymbol{y}|\Xi) = [m(t_1), \ldots, m(t_n)]^T$, $\boldsymbol{B}$ is the true parameter, and $\boldsymbol{r} = \boldsymbol{m} - \boldsymbol{X}\boldsymbol{B}$ is the residual vector of the local polynomial approximation. Thus, the conditional bias is

$$\text{Bias}(\hat{\boldsymbol{B}}|\Xi) = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{r}. \tag{8}$$

From definition, the conditional covariance of (6) $\text{var}(\hat{\boldsymbol{B}}|\Xi)$ is given by

$$\begin{aligned} &E\left(\left\{\hat{\boldsymbol{B}} - E(\hat{\boldsymbol{B}})\right\}\left\{\hat{\boldsymbol{B}} - E(\hat{\boldsymbol{B}})\right\}^T \middle| \Xi\right) \\ &\quad = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W} \\ &\qquad \times E\left((\boldsymbol{y}-\boldsymbol{m})(\boldsymbol{y}-\boldsymbol{m})^T|\Xi\right)\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} \\ &\quad = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\Sigma\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} \end{aligned} \tag{9}$$

where $\Sigma = \boldsymbol{W}E((\boldsymbol{y}-\boldsymbol{m})(\boldsymbol{y}-\boldsymbol{m})^T|\Xi)\boldsymbol{W} = \text{diag}\{K_h^2(t_i - t_0)\sigma^2(t_i),\ i = 1, \ldots, n\}$. We are interested in the asymptotic expressions of the bias in (8) and variance in (9) as functions of the bandwidth h when the number of measurements is large. To this end, we allow h to tend to zero so that we can employ the Taylor series expansion to reveal their order of dependence as $h \to 0$. However, n is still assumed to be large so that the central limit theorem (CLT) [27] can be applied to derive the asymptotic results. Therefore, we assume that $nh \to \infty$.

First of all, we define below two key quantities found in (8) and (9), namely, $\boldsymbol{S}_n$ and $\boldsymbol{S}_n^*$.

$$\boldsymbol{S}_n = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} \tag{10}$$

where the $(jL+q, lL+m)$th element in the matrix $\boldsymbol{S}_n \in \boldsymbol{R}^{(p+1)L\times(p+1)L}$, $(\boldsymbol{S}_n)_{jL+q,lL+m}$, is given by $s_{n,j+l,q,m}$ ($0 \le j, l \le p$ and $1 \le q, m \le L$) with

$$s_{n,\varsigma,q,m} = \sum_{i=1}^{n} x(q,t_i)x(m,t_i)K_h(t_i-t_0)(t_i-t_0)^{\varsigma} \quad \text{with } \varsigma = j+l. \tag{11}$$

$$\boldsymbol{S}_n^* = \boldsymbol{X}^T\Sigma\boldsymbol{X} \tag{12}$$

where the $(jL+q, lL+m)$th element in the matrix $\boldsymbol{S}_n^* \in \boldsymbol{R}^{(p+1)L\times(p+1)L}$, $(\boldsymbol{S}_n^*)_{jL+q,lL+m}$, ($0 \le j,\ l \le p$ and $1 \le q$, $m \le L$) is given by

$$s_{n,\varsigma,q,m}^* = \sum_{i=1}^{n} x(q,t_i)x(m,t_i)K_h^2(t_i-t_0)(t_i-t_0)^{\varsigma}\sigma^2(t_0), \quad \text{with } \varsigma = j+l. \tag{13}$$

The conditional variance is thus given by $\boldsymbol{S}_n^{-1}\boldsymbol{S}_n^*\boldsymbol{S}_n^{-1}$. Next, we shall evaluate the asymptotic expressions of $\boldsymbol{S}_n$ and $\boldsymbol{S}_n^*$. Then, we shall make use of them to evaluate the asymptotic bias and variance of $\hat{\boldsymbol{B}}$. We shall also make the following assumptions for derivation of the asymptotic results: 1) The input vector $\boldsymbol{x}(t)$ is a random vector with finite second-order statistics; 2) the system is persistently excited (i.e., the correlation matrix of $\boldsymbol{x}(t)$ is invertible); 3) the sampling density $f(t_0 + h\tau)$ and the correlation coefficient $E[x(q, t_0 + h\tau)x(m, t_0 + h\tau)]$ are continuous functions of h as $h \to 0$; and 4) the additional noise is white Gaussian distributed with zero mean and is independent of $\boldsymbol{x}(t)$ and $\boldsymbol{a}(t)$.

Assuming that the observations are independent and identically distributed, by CLT, we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x(q,t_i)x(m,t_i)K_h(t_i-t_0)(t_i-t_0)^{\varsigma}-\mu_S\right) \xrightarrow{D} \aleph\left(0,\sigma_S^2\right) \quad (14)$$

where $\mu_S = E[x(q,t_i)x(m,t_i)K_h(t_i-t_0)(t_i-t_0)^{\varsigma}]$ and $\sigma_S^2 = \mathrm{var}[x(q,t_i)x(m,t_i)K_h(t_i-t_0)(t_i-t_0)^{\varsigma}]$ are, respectively, the sample mean and variance and $\aleph(0,\sigma_S^2)$ is the normal distribution with zero mean and variance σ_S^2. By the definition of μ_S, the independence of $x(q,t)$ and the density of the samples, one gets

$$\mu_S = h^{-1}\int E\left[x(q,t)x(m,t)\right]K\left(\frac{1}{h}(t-t_0)\right)(t-t_0)^{\varsigma}f(t)dt \quad (15)$$

where $f(t)$ is the sampling density function at t. Using the substitution $t-t_0 = h\tau$, (15) becomes

$$\mu_S = h^{\varsigma}\int E\left[x(q,t_0+h\tau)x(m,t_0+h\tau)\right] \cdot K(\tau)\tau^{\varsigma}f(t_0+h\tau)d\tau. \quad (16)$$

As mentioned earlier, we are interested in a Taylor series expansion in terms of the bandwidth parameter h. Hence, we assume that $h\to 0$, while $nh\to\infty$, i.e., the number of measurements is still very large that CLT is applicable. To proceed further, we assume that $E[x(q,t_0+h\tau)x(m,t_0+h\tau)]$ and $f(t_0+h\tau)$ are continuous in h such that

$$E[x(q,t_0+h\tau)x(m,t_0+h\tau)] = E[x(q,t_0)x(m,t_0)]+O(h)\tau$$
$$f(t_0+h\tau) = f(t_0)+O(h)\tau. \quad (17)$$

This is equivalent to saying that $x(q,t)$ is nonstationary but is still smooth in the time domain. Accordingly, (16) becomes

$$\begin{aligned}\mu_S &= h^{\varsigma}\int\left\{E\left[x(q,t_0)x(m,t_0)\right]+O(h)\right\} \\ &\quad\times K(\tau)\tau^{\varsigma}\left(f(t_0)+O(h)\tau\right)d\tau \\ &= h^{\varsigma}f(t_0)r_{x,q,m}(t_0)\mu_s\left\{1+o(1)\right\} \end{aligned} \quad (18)$$

where $r_{x,q,m}(t_0) = E[x(q,t_0)x(m,t_0)]$ is the correlation of the inputs $x(q,t_0)$ and $x(m,t_0)$, and $\mu_{\varsigma} = \int\tau^{\varsigma}K(\tau)d\tau$. The notation $\chi(\zeta) = O(\delta(\zeta))$ means that a real-valued $\chi(\zeta)$ is less than some constant multiple of $|\delta(\zeta)|$, and the notation $\chi(\zeta) = o(\delta(\zeta))$ means that the quantity $\chi(\zeta)/\delta(\zeta)$ tends to zero as $\zeta\to\zeta_0$, where ζ_0 is a real-valued constant or infinity [27]. Thus, $o(1)$ in (18) denotes one real-valued quantity tending to zero as $h\to 0$.

The variance σ_S^2 can be derived similarly as

$$\begin{aligned}\sigma_S^2 &= E\left[x^2(q,t)x^2(m,t)K_h^2(t-t_0)(t-t_0)^{2\varsigma}\right] \\ &= h^{2\varsigma-1}\int r^*_{x,q,m}(t_0)K^2(\tau)\tau^{2\varsigma}\left\{f(t_0)\left[1+o(1)\right]\right\}d\tau \\ &= h^{2\varsigma-1}f(t_0)r^*_{x,q,m}(t_0)\nu_{\varsigma}\left\{1+o(1)\right\} \end{aligned} \quad (19)$$

where $r^*_{x,q,m}(t_0) = E[x^2(q,t_0)x^2(m,t_0)]$ and $\nu_{\varsigma} = \int\tau^{\varsigma}K^2(\tau)d\tau$. Combining (14), (18), and (19) gives

$$\begin{aligned} & s_{n,\varsigma,q,m} \\ &= n\mu_S+\sqrt{n}O_P\left(\sqrt{\sigma_S^2}\right) \\ &= nh^{\varsigma}\Big\{f(t_0)r_{x,q,m}(t_0)\mu_{\varsigma}\left[1+o(1)\right] \\ &\qquad + O_P\left(\sqrt{(nh)^{-1}\nu_{\varsigma}f(t_0)r^*_{x,q,m}(t_0)\left[1+o(1)\right]}\right)\Big\} \\ &= nh^{\varsigma}\Big\{f(t_0)r_{x,q,m}(t_0)\mu_{\varsigma}\left[1+o(1)\right]+O_P\Big(\sqrt{(nh)^{-1}}\Big)\Big\} \\ &= nh^{\varsigma}f(t_0)r_{x,q,m}(t_0)\mu_{\varsigma}\left\{1+o_P(1)\right\}. \end{aligned} \quad (20)$$

Here, the stochastic order symbols $O_P(\cdot)$ and $o_P(\cdot)$ are used since $s_{n,\varsigma,q,m}$ is a random variable. The notation $\chi(\zeta) = O_p(\delta(\zeta))$ means that the random variable $\chi(\zeta)$ is stochastically bounded by some constant multiple of $|\delta(\zeta)|$, and the notation $\chi(\zeta) = o_p(\delta(\zeta))$ means that $\chi(\zeta)/\delta(\zeta)$ converges to zero in probability [27].

Using (20) and (11), one gets

$$\begin{aligned}(\boldsymbol{S}_n)_{jL+q,lL+m} &= s_{n,j+l,q,m} \\ &= nf(t_0)h^{(j+l)}r_{x,q,m}(t_0)\mu_{j+l}\left\{1+o_P(1)\right\} \end{aligned} \quad (21)$$

$$\boldsymbol{S}_n = nf(t_0)\boldsymbol{Q}\otimes\boldsymbol{R}_{\boldsymbol{X}}\left\{1+o_P(1)\right\} \quad (22)$$

where $\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{U}\boldsymbol{H}$, $\boldsymbol{H} = \mathrm{diag}(1,h,\ldots,h^p)$, $(\boldsymbol{U})_{0\le j,l\le p} = \mu_{j+l}$, and $(\boldsymbol{R}_{\boldsymbol{X}})_{1\le q,m\le L} = r_{x,q,m}(t_0)$. Note that $\boldsymbol{U}\in\boldsymbol{R}^{(p+1)\times(p+1)}$, $\boldsymbol{H}\in\boldsymbol{R}^{(p+1)\times(p+1)}$, $\boldsymbol{Q}\in\boldsymbol{R}^{(p+1)\times(p+1)}$, and $\boldsymbol{R}_{\boldsymbol{X}}\in\boldsymbol{R}^{L\times L}$.

Using similar arguments, we have

$$\boldsymbol{S}_n^* = nf(t_0)h^{-1}\sigma^2(t_0)\boldsymbol{Q}^*\otimes\boldsymbol{R}_{\boldsymbol{X}}\left\{1+o_P(1)\right\} \quad (23)$$

where $\boldsymbol{Q}^* = \boldsymbol{H}\boldsymbol{U}^*\boldsymbol{H}$ and $(\boldsymbol{U}^*)_{0\le j,l\le p} = \nu_{j+l}$. Likewise, $\boldsymbol{U}^*\in\boldsymbol{R}^{(p+1)\times(p+1)}$, and $\boldsymbol{Q}^*\in\boldsymbol{R}^{(p+1)\times(p+1)}$. Equations (22) and (23) can be substituted back to (8) and (9) to obtain the asymptotic expressions for the bias and variance, respectively.

More precisely, from (22) and (23) and using the Slutsky's theorem, we get the variance of $\hat{\boldsymbol{B}}$ in (9) as

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{B}}|\Xi) &= \boldsymbol{S}_n^{-1}\boldsymbol{S}_n^*\boldsymbol{S}_n^{-1} \\ &= \frac{\sigma^2(t_0)}{nf(t_0)h}(\boldsymbol{Q}\otimes\boldsymbol{R}_{\boldsymbol{X}})^{-1}(\boldsymbol{Q}^*\otimes\boldsymbol{R}_{\boldsymbol{X}}) \\ &\quad\times(\boldsymbol{Q}\otimes\boldsymbol{R}_{\boldsymbol{X}})^{-1}\left\{1+o_P(1)\right\} \\ &= \frac{\sigma^2(t_0)}{nf(t_0)h}(\boldsymbol{Q}^{-1}\boldsymbol{Q}^*\boldsymbol{Q}^{-1}) \\ &\quad\otimes\left(\boldsymbol{R}_{\boldsymbol{X}}^{-1}\boldsymbol{R}_{\boldsymbol{X}}\boldsymbol{R}_{\boldsymbol{X}}^{-1}\right)\left\{1+o_P(1)\right\} \\ &= \frac{\sigma^2(t_0)}{nf(t_0)h}(\boldsymbol{H}^{-1}\boldsymbol{U}^{-1}\boldsymbol{U}^*\boldsymbol{U}^{-1}\boldsymbol{H}^{-1}) \\ &\quad\otimes\boldsymbol{R}_{\boldsymbol{X}}^{-1}\left\{1+o_P(1)\right\}. \end{aligned} \quad (24)$$

For the asymptotic bias in (8), we first employ the Taylor series expansion of the residual $\boldsymbol{r} = \boldsymbol{m} - \boldsymbol{X}\boldsymbol{B}$ at $\hat{\boldsymbol{B}}$ to obtain

$$\begin{aligned}\text{Bias}(\hat{\boldsymbol{B}}|\Xi) &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{r} = \boldsymbol{S}_n^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{r} \\ &= \boldsymbol{S}_n^{-1}\boldsymbol{X}^T\boldsymbol{W}\left\{\boldsymbol{X}_{p+1}\boldsymbol{\beta}^{(p+1)} + o_P(\boldsymbol{X}_{p+1})\right\} \\ &= \boldsymbol{S}_n^{-1}\left\{\mathbf{c}_n\boldsymbol{\beta}^{(p+1)} + o_P(nh^{p+1})\right\} \end{aligned} \tag{25}$$

where $\boldsymbol{x}_{p+1}=[(t_1-t_0)^{p+1}\boldsymbol{x}(t_1),(t_2-t_0)^{p+1}\boldsymbol{x}(t_2),\ldots,(t_n-t_0)^{p+1}\boldsymbol{x}(t_n)]^T \in \boldsymbol{R}^{n\times L}$, $\boldsymbol{c}_n \in \boldsymbol{R}^{(p+1)L\times L} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{x}_{p+1} = nf(t_0)\boldsymbol{q}\otimes\boldsymbol{R}_{\boldsymbol{X}}\{1+o_P(1)\}$ with $\boldsymbol{q} = [h^{p+1}\mu_{p+1}, h^{p+2}\mu_{p+2}, \ldots, h^{2p+1}\mu_{2p+1}]^T \in \boldsymbol{R}^{p+1}$, and $\boldsymbol{\beta}^{(p+1)} = [\beta^{(p+1)}(1,t_0), \ldots, \beta^{(p+1)}(L,t_0)]^T \in \boldsymbol{R}^L$ is the $(p+1)$th derivative of β.

Using the asymptotic expressions of $\boldsymbol{S}_n$ and $\boldsymbol{S}_n^*$ in (21) and (22), the asymptotic conditional bias of $\hat{\boldsymbol{B}}$ in (25) becomes

$$\begin{aligned}\text{Bias}(\hat{\boldsymbol{B}}|\Xi) &= (\boldsymbol{Q}\otimes\boldsymbol{R}_{\boldsymbol{X}})^{-1}(\boldsymbol{Q}\otimes\boldsymbol{R}_{\boldsymbol{X}})\boldsymbol{\beta}^{(p+1)}\{1+o_P(1)\} \\ &= \left[(\boldsymbol{H}^{-1}\boldsymbol{U}^{-1}\boldsymbol{U})\otimes\boldsymbol{I}_L\right]\boldsymbol{\beta}^{(p+1)}h^{p+1}\{1+o_P(1)\}\end{aligned} \tag{26}$$

where $\boldsymbol{u} = (\mu_{p+1},\ldots,\mu_{2p+1})^T \in \boldsymbol{R}^{p+1}$.

Since the polynomial coefficients and the regression coefficients are related by $\alpha^{(j)}(k,t_0) = j!\beta^{(j)}(k,t_0) = j!\boldsymbol{B}(jL+k,t_0)$, the asymptotic conditional bias and variance of the local polynomial coefficients $\hat{\alpha}^{(j)}(k,t_0)$ are, respectively

$$\begin{aligned}&\text{Bias}\left(\hat{\alpha}^{(j)}(k,t_0)|\Xi\right) \\ &= \frac{\left\{e_v^T\left[\boldsymbol{H}^{-1}\boldsymbol{U}^{-1}\boldsymbol{u}\otimes\boldsymbol{I}_L\right]\boldsymbol{\alpha}^{(p+1)}(t_0)\right\}j!h^{p+1}}{(p+1)!}\{1+o_P(1)\} \\ &= \frac{\left\{e_v^T\left[\boldsymbol{U}^{-1}\boldsymbol{u}\otimes\boldsymbol{I}_L\right]\boldsymbol{\alpha}^{(p+1)}(t_0)\right\}j!h^{p+1-j}}{(p+1)!} + o_P(h^{p+1-j})\end{aligned} \tag{27}$$

$$\begin{aligned}&\text{Var}\left(\hat{\alpha}^{(j)}(k,t_0)|\Xi\right) \\ &= \frac{\left\{e_v^T\left[\boldsymbol{H}^{-1}\boldsymbol{U}^{-1}\boldsymbol{U}^*\boldsymbol{U}^{-1}\boldsymbol{H}^{-1}\otimes\boldsymbol{R}_{\boldsymbol{X}}^{-1}(t_0)\right]e_v\right\}j!^2\sigma^2(t_0)}{nf(t_0)h} \\ &\quad + \{1+o_P(1)\} \\ &= \frac{\left\{e_v^T\left[\boldsymbol{U}^{-1}\boldsymbol{U}^*\boldsymbol{U}^{-1}\otimes\boldsymbol{R}_{\boldsymbol{X}}^{-1}(t_0)\right]e_v\right\}j!^2\sigma^2(t_0)}{nf(t_0)h^{2j+1}} \\ &\quad + o_P\left(\frac{1}{nh^{2j+1}}\right)\end{aligned} \tag{28}$$

where $v = jL+k$ and $e_v = (0,\ldots,0,1,0,\ldots,0)^T$ with a value one on the vth position and zero elsewhere.

It can be seen from (27) and (28) that as h increases, the squared bias will increase while the variance will decrease. Hence, there exists a locally optimal bandwidth $h^{\text{opt}}(k,t_0)$ for estimating $\hat{\alpha}^{(j)}(k,t_0)$, which minimizes the MSE as follows:

$$\begin{aligned}\text{MSE}\left(\hat{\alpha}^{(j)}(k,t_0)|\Xi\right) &= \left[\text{Bias}\left(\hat{\alpha}^{(j)}(k,t_0)|\Xi\right)\right]^2 \\ &\quad + \text{Var}\left(\hat{\alpha}^{(j)}(k,t_0)|\Xi\right).\end{aligned} \tag{29}$$

By setting the derivative of (29) with respect to h to zero, the following asymptotic optimal bandwidth is obtained:

$$\begin{aligned}h_{\text{opt}}(k,t_0) = \Bigg\{&\frac{e_v^T\left[\boldsymbol{U}^{-1}\boldsymbol{U}^*\boldsymbol{U}^{-1}\otimes\boldsymbol{R}_{\boldsymbol{X}}^{-1}(t_0)\right]e_v}{\left\{e_v^T[\boldsymbol{U}^{-1}\boldsymbol{u}\otimes\boldsymbol{I}_L]\boldsymbol{\alpha}^{(p+1)}(t_0)\right\}^2} \\ &\times\frac{\{(p+1)!\}^2(2j+1)}{2(p+1-j)}\frac{\sigma^2(t_0)}{nf(t_0)}\Bigg\}^{1/(2p+3)}.\end{aligned} \tag{30}$$

Now, we have already established the asymptotic expressions of the bias and variance of the proposed LPM approach and the existence of an optimal bandwidth. These results, although useful in analytical work, cannot be used directly in practice. It is because some quantities in (30), mainly $\boldsymbol{\alpha}^{(p+1)}(t_0)$, are difficult to be calculated, which makes the optimal bandwidth difficult to be estimated accurately. Instead of computing an optimal bandwidth in an analytical form, we adopt the "data-driven" approach and propose an empirical method to select the optimal bandwidth from a finite set of possible bandwidths.

IV. Data-Driven VBS for LPM

In the empirical bandwidth-selection methods to be introduced, an approximate optimal bandwidth can be obtained by minimizing an approximated MSE over a set of possible bandwidths. Toward this end, we need the following approximation methods to determine the bias, variance, and MSE.

A. Approximated Bias and Variance

Although the bias and variance cannot be directly computed because of the unknown quantities, good finite sample approximations of the bias and variance can still be derived, as in the local polynomial regression [16]. For instance, the conditional bias, which contains the unknown residual $\boldsymbol{r} = \boldsymbol{m} - \boldsymbol{X}\boldsymbol{B}$, can be estimated using a Taylor series expansion with an order $p+p_{ex}$

$$\boldsymbol{b}\left(\hat{\boldsymbol{B}}(t_0)\right) = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\tau} \tag{31}$$

where $\boldsymbol{\tau}$ is an $n\times 1$ vector with $\sum_{\upsilon=1}^{p_{ex}}\sum_{k=1}^{L}\beta^{(p+\upsilon)}(k,t_0)(t_i-t_0)^{p+\upsilon}x(k,t_i)$ as its ith $(i=1,2,\ldots,n)$ element and the quantities $\beta^{(p+\upsilon)}(k,t_0)$ can be estimated by fitting a polynomial of degree $p+p_{ex}$. The basic idea is that the observations can be better fitted by a higher order polynomial, and hence, the model provided by this higher order polynomial can be used to estimate the bias for a local polynomial model with lower order. The excess or extra order p_{ex} is generally chosen as $p_{ex}=2$ because this selection would reduce the computational costs and lead to a bandwidth selector which is not far from being $\sqrt{n}$-consistent [16]. Since the higher order model is still based on LPM, we still need a bandwidth h^* in this $(p+p_{ex})$th-order LPM. This bandwidth h^* is usually referred to as the "pilot bandwidth," which will be discussed later in Section IV-B.

Next, suppose local homoscedasticity; the conditional variance can be estimated as

$$\boldsymbol{V}\left(\hat{\boldsymbol{B}}(t_0)\right) = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}^2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\sigma^2(t_0). \tag{32}$$

The noise variance $\sigma^2(t_0)$ is estimated as the normalized weighted residual sum of squares [16]

$$\hat{\sigma}^2(t_0) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 K_{h*}(t_i - t_0)}{tr\left\{\boldsymbol{W}^* - \boldsymbol{W}^*\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{W}^*\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}^*\right\}} \tag{33}$$

where $\boldsymbol{X}^*$ and $\boldsymbol{W}^*$ are, respectively, the design matrix and weighting matrix in the $(p + p_{ex})$th-order LPM using the pilot bandwidth h^*.

The MSE of the vth entry of $\hat{\boldsymbol{B}}(t_0)$ at a given point t_0 is given by

$$\text{MSE}_v\left(\hat{\boldsymbol{B}}(t_0)\right) = b_v^2\left(\hat{\boldsymbol{B}}(t_0)\right) + V_v\left(\hat{\boldsymbol{B}}(t_0)\right) \tag{34}$$

where $\hat{b}_v(\hat{\boldsymbol{B}}(t_0))$ is the vth element of $\boldsymbol{b}(\hat{\boldsymbol{B}}(t_0))$ and $V_v(\hat{\boldsymbol{B}}(t_0))$ is the vth diagonal element of $\boldsymbol{V}(\hat{\boldsymbol{B}}(t_0))$. Note that we only focus on the estimation of the coefficients $\hat{a}(k, t_0) = \hat{\alpha}^{(0)}(k, t_0) = \boldsymbol{B}(k, t_0)$, and thus, it is only required to approximate $\text{MSE}(\hat{a}(k, t_0)) = \text{MSE}_k(\hat{\boldsymbol{B}}(t_0))$ by the approximated bias and variance from (31)–(33). Their derivatives can also be estimated in a similar manner. We now propose an ICI-method-based algorithm for finding the pilot bandwidth h^* and, hence, estimate the quantities $\boldsymbol{X}^*$, $\boldsymbol{W}^*$, and the higher derivatives $\beta^{(p+\upsilon)}(k, t_0)$, as required by (33).

B. Pilot Bandwidth Selection (the ICI Method)

The ICI method is an empirical adaptive bandwidth-selection method proposed by Goldenshluger and Nemirovski [16], and it has been successfully applied to various areas, including local polynomial regression, image processing, and time–frequency analysis, for selecting the variable bandwidth [18]–[20]. The theoretical background of the ICI method is omitted to save space, and more details can be found in [18]–[20]. We now briefly review the basic concept and the algorithm of the ICI method.

Given a set of bandwidth parameters in an ascending order, generally, in a form of an exponential series

$$\widetilde{\boldsymbol{H}} = \{h_m | h_m = h_0(h_a)^m, \quad m = 0, \ldots, M-1\} \tag{35}$$

where $h_a > 1$ is a step factor, $h_0 > 0$ is the base bandwidth, and M is the number of possible bandwidths, the ICI method determines the optimal bandwidth by comparing the confidence intervals of the estimates with different bandwidths in the bandwidth set.

Consider a series of confidence intervals $D_m = [L_m, U_m]$ obtained from the estimated $\hat{a}(k, t_0)$ with different values of bandwidth h_m in the set $\widetilde{\boldsymbol{H}}$

$$U_m = \hat{a}(k, t_0 h_m) + \kappa \cdot SD\left(\hat{a}(k, t_0 h_m)\right) \tag{36}$$

$$L_m = \hat{a}(k, t_0 h_m) - \kappa \cdot SD\left(\hat{a}(k, t_0 h_m)\right) \tag{37}$$

where the standard deviation $SD(\hat{a}(k, t_0; h_m))$ is the square root of $V_k(\hat{\boldsymbol{B}}(t_0; h_m))$; κ is a threshold parameter used to adjust the width of the confidence interval, and it can be chosen as the one that minimizes the generalized cross-validation criterion [18].

For a small value of h, we expect that the bias is small and the confidence interval will gradually decrease with increasing value of h, while the center of the interval remains more or less fixed. When h is increased to a point where the observations cannot be satisfactorily modeled, a large bias will be observed, and the center of the interval will change significantly with respect to the others, while the length of the interval will still be small. As a result, the confidence interval will no longer intersect those with smaller values of h.

Motivated by this observation, the ICI bandwidth-selection method computes and examines the following quantities from the confidence intervals in order to detect this sudden change:

$$\begin{aligned} \overline{L}_m &= \max[\overline{L}_{m-1}, L_m], \quad \text{for} \quad m = 1, 2, \ldots, M-1 \\ \underline{U}_m &= \min[\underline{U}_{m-1}, U_m], \quad \text{for} \quad m = 1, 2, \ldots, M-1 \\ \overline{L}_0 &= \underline{U}_0 = 0. \end{aligned} \tag{38}$$

It can be seen that $\overline{L}_m$ is the largest upper bound of the confidence interval for bandwidth evaluated up to h_m, while $\underline{U}_m$ is the corresponding lower bound. The largest value of these m's for which $\underline{U}_m \geq \overline{L}_m$ gives the desirable bandwidth $h_{\text{ICI}}(k, t_0)$ because the confidence intervals no longer intersect with each other above the bandwidth $h_{\text{ICI}}(k, t_0)$.

In this paper, we employ the ICI method to determine the pilot bandwidth and then generate a relatively smooth bandwidth using the method described in Section IV-A. The latter is based on the Fan and Gijbels' method for local polynomial regression [16]. Note that the ICI method is an effective adaptive VBS method, and therefore, it can also be used to find the data-driven bandwidth of our LPM model. It has the advantages of lower arithmetic complexity and better performance in systems with jump discontinuities than the Fan and Gijbels' method [20]. On the other hand, the Fan and Gijbels' method usually gives a smoother bandwidth estimate and yields better performance for smooth time-varying signals and at low SNR. For a more detailed comparison, the readers are referred to [20]. It should be noted that variations of these basic schemes exist, and the proposed method is chosen for its good performance.

C. Data-Driven VBS

Suppose that we have used the ICI method to obtain the variable bandwidths $h_{ICI}(k, t_0)$ for the kth coefficient at each time instant t_0. Then, $h_{ICI}(k, t_0)$ will be used as pilot bandwidths $h^*(k, t_0)$ to compute $\boldsymbol{X}^*$, $\boldsymbol{W}^*$, and $\hat{\beta}^{(p+\upsilon)}(k, t_0)$ in (31)–(33). Next, with the pilot estimates, we can calculate a series of bias, variance, and MSE values of the LPM estimators using each bandwidth in the set $\widetilde{\boldsymbol{H}}$. Finally, the optimal data-driven bandwidth is the bandwidth having the minimum estimated MSE.

As observed in [16], the estimated MSE usually exhibit considerable variation. Hence, it is better to estimate an optimal

bandwidth for a small subinterval instead of at each time instant t_0. In other words, a constant bandwidth will be used within the subinterval. To this end, the whole measurement is first split up into a set of nonoverlapping subintervals, for example, I_l. The bandwidth function $\widetilde{h}_{\mathrm{MMSE}}(k,t)$ in the lth subinterval I_l can be determined by minimizing the integrated MSE

$$\widetilde{h}_{\mathrm{MMSE}}(k,t|t\in I_l)=\arg\min_{h}\int_{I_l}\mathrm{MSE}\left(\hat{a}(k,th)\right)dt. \quad (39)$$

The main difference in using (39) with (34) is that, the estimated MSE for each sample inside the subinterval I_l are summed up together to form the integrated MSE: $\int_{I_l}\mathrm{MSE}(\hat{a}(k,t;h))dt$. The proposed bandwidth-selection method can then be invoked to determine the optimal data-driven bandwidth $\widetilde{h}_{\mathrm{MMSE}}(k,t|t\in I_l)$ for this subinterval. The averaging operation reduces the sensitivities of the algorithm to additive noise.

Since a rapidly fluctuating bandwidth is generally less desirable, $\widetilde{h}_{\mathrm{MMSE}}(k,t)$ is usually smoothed (averaged) with a length of I_l to yield the final local bandwidth $h_{\mathrm{MMSE}}(k,t)$, i.e., $h_{\mathrm{MMSE}}(k,t)=\int_{t-I_l/2}^{t+I_l/2}\widetilde{h}_{\mathrm{MMSE}}(k,\tau)d\tau$. Lastly, the LPM-based WLS proposed in Section II with the computed local bandwidth $h_{\mathrm{MMSE}}(k,t)$ is used to compute the final estimate $\hat{a}(k,t;h_{\mathrm{MMSE}}(k,t))$. As suggested in [16], the length of the subinterval I_l can be selected around $10\log_{10}(n)$, where n is the size of the data. For online application, it is usually chosen as the interval where the system will remain somewhat stationary.

We now summarize the proposed LPM-VBS method as follows.

- Step 1) For each bandwidth h in $\widetilde{\boldsymbol{H}}$ of (35), an estimate of the coefficient $\hat{a}(k,t;h)$ is calculated by a pth-order LPM with the constant bandwidth h at every time instant using (6).
- Step 2) A pilot bandwidth $h_{\mathrm{ICI}}(k,t)$ is estimated using the ICI method (36)–(38) based on the estimates $\hat{a}(k,t;h)$.
- Step 3) For each bandwidth h in $\widetilde{\boldsymbol{H}}$, a $(p+p_{ex})$th-order LPM with the pilot bandwidth $h_{\mathrm{ICI}}(k,t)$ is carried out to approximate the MSE $\mathrm{MSE}(\hat{a}(k,t;h))$ using (31)–(34).
- Step 4) In the lth subinterval, the optimal bandwidth $\widetilde{h}_{\mathrm{MMSE}}(k,t|t\in I_l)$ is approximated as the bandwidth having the minimum integrated MSE (39), and the desired variable bandwidth $h_{\mathrm{MMSE}}(k,t)$ is obtained by smoothing $\widetilde{h}_{\mathrm{MMSE}}(k,t)$ with the length of subinterval I_l.
- Step 5) A pth-order LPM with $h_{\mathrm{MMSE}}(k,t)$ is finally performed to obtain the final estimates of the time-varying coefficient $\hat{a}(k,t;h_{\mathrm{MMSE}}(k,t))$, $k=1,\ldots,L$, using (6).

D. Practical Issues

1) Computational Complexity: First, we consider the arithmetic complexity of the LPM estimator in (6): $\hat{\boldsymbol{B}}=(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$, where $\boldsymbol{X}$ is an $n\times(p+1)L$ matrix, $\boldsymbol{y}$ is an $n\times 1$ vector, and $\boldsymbol{W}$ is an $n\times n$ diagonal matrix. The complexities of computing $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$ and $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$ are $O\{n(p+1)^2L^2\}$ and $O\{n(p+1)L\}$, respectively, and the matrix inversion needs a complexity of $O\{[(p+1)L]^3\}$. If $n\gg(p+1)L$, the complexity of the LPM solution at each time point can be regarded as $O(n)$. Since the approximations of bias and variance require a $(p+p_{ex})$th fitting and the ICI bandwidth selector needs to calculate LPM solutions using a series of bandwidth parameters, the actual complexity of the LPM estimator is higher than the adaptive filtering and BEM methods. The RLS and KF methods have a complexity of $O\{[(p+1)L]^3\}$ at each time instant, while the complexity of BEM is around $O(n)$ for all time instants if $n\gg(p_{\mathrm{BEM}}+1)L$, where p_{BEM} is the numbers of basis functions used in BEM.

The computational complexity of the basic LPM algorithm can be further simplified. If the kernel has a limited support, such as the Epanechnikov kernel used in this paper, then the number of actual time points included in the kernel is finite. Let us denote it as n_K. Hence, the corresponding complexity of the LPM solution will decrease considerably to $O(n_K)$ if $n_K\gg(p+1)L$. Since n_K increases with the bandwidth parameter h, a large bandwidth will increase the computational complexity.

2) Selection of Kernel: As mentioned earlier, a kernel with compact support is desirable because it can reduce the complexity significantly. In this paper, following the recommendation of literature [16] and [24], the following basis Epanechnikov kernel is employed:

$$K(u)=\begin{cases}\frac{3}{4}\left(1-|u|^2\right), & |u|<1\\ 0, & |u|\geq 1.\end{cases} \quad (40)$$

Consequently, for an Epanechnikov kernel with bandwidth h, $K_h(u)=(1/h)K(u/h)$, only the time points included in interval (t_0-h,t_0+h) are required for local estimation, and thus, the effective length of the kernel is $2h$. Other types of kernels, such as Gaussian kernels and biweight kernels, can also be used without much performance difference [16].

3) Selection of Bandwidth Set: The next problem is how to select the bandwidth set $\widetilde{\boldsymbol{H}}$ used in the empirical bandwidth-selection method. The bandwidth parameter h_0 should be selected as the minimum bandwidth that makes the LPM estimators solvable. More precisely, the number of measurements n_K included in the interval (t_0-h_0,t_0+h_0) should be equal to or larger than $(p+p_{ex}+1)L$ because a $(p+p_{ex})$-order LPM will be employed to approximate the MSE. If the data are uniformly distributed at a sampling frequency of f_s, we get the following condition for the minimum bandwidth: $h_0\geq(p+p_{ex}+1)L/(2f_s)$. On the other hand, the largest element in the bandwidth set $h_0h_a^{M-1}$ should be large enough so that the corresponding local kernel can include all data points. However, as mentioned before, too large a kernel will result in high computational complexity. As a compromise, the largest bandwidth used in this paper enables the largest kernel to cover at least 1/8 of the whole data set.

The step factor h_a determines the distances between two adjacent bandwidth parameters in the bandwidth set and, thus, the total number of bandwidth parameters in the set. A small step factor h_a implies more bandwidth parameters in $\widetilde{\boldsymbol{H}}$, which may

lead to more refined estimation results but will also increase the complexity. To achieve a good tradeoff between performance and complexity, the parameters h_a is chosen to make sure that the number of bandwidths M is larger than or equal to four. From simulation results, it was found that the proposed bandwidth setting given earlier gave satisfactory results in practice. In real-time and online applications, the bandwidth set can be determined experimentally from the variation of the system responses, for example, through a small and fixed bandwidth.

4) Selection of Polynomial Order: As seen from the asymptotic expression of the bias in (27), the estimation bias is a function of polynomial order p, and a larger p can lead to a smaller bias, in theory. However, it does not imply that the order p can be arbitrarily large due to two reasons. First, $(p+1)L$ should be smaller than the number of measurements n_K in the smallest kernel to ensure that the LPM solution in (6) is well posed. Second, a higher order increases the variability of the LPM estimator because it implies more parameters (i.e., with the dimension of $(p+1)L$) to be estimated. Although the asymptotic expression of variance in (28) does not include the polynomial order p, the constant term $\boldsymbol{S}'^{-1}\boldsymbol{S}'^{*}\boldsymbol{S}'^{-1}$ will generally increase with p. More details about the influence of p on the variance and the selection of p can be found in the comprehensive book by Fan and Gijbels [16]. They also recommended a local linear estimator $(p=1)$ because, compared with local constant estimator $(p=0)$, the linear estimator will not suffer from any loss in variance and can reduce the bias effectively. Therefore, in this paper, the polynomial order is chosen as $p=1$.

5) Online Implementation: Adaptive filtering and KF make use of past measurements for current estimates, and thus, they are suitable for tracking purposes. The proposed LPM method can be easily extended for the purpose of online tracking by employing a one-sided kernel. For example, the one-sided Epanechnikov kernel is given by

$$K(u)=\begin{cases}\frac{3}{4}(1-u^2), & -1<u\le 0\\ 0, & u\le -1 \text{ or } u>0.\end{cases} \tag{41}$$

Because the one-sided kernel only has a support of (t_0-h,t_0), its effective length is h. As a result, the minimum bandwidth in the bandwidth set $\widetilde{\boldsymbol{H}}$ should satisfy $h_0\ge (p+p_{ex}+1)L/f_s$ when the observations are uniformly distributed at a sampling frequency of f_s. In addition, to obtain a smooth bandwidth function, the optimal bandwidth $h(k,t)$ can be smoothed recursively as

$$h(k,t)=\lambda_h h(k,t) + (1-\lambda_h)\text{median}\left(\bar{h}(k,t-1),\ldots,\bar{h}(k,t-I_l)\right) \tag{42}$$

where $\bar{h}(k,t)$ is the raw bandwidth estimated by the ICI method or the minimum MSE criterion, λ_h is the forgetting factor, which should be close to one to ensure the smoothness of the resultant bandwidth function, and I_l is the length of the smoothing window. The use of the median operation avoids possible large variations in $\bar{h}(k,t)$ from affecting significantly the smoothed bandwidth $h(k,t)$. No other modification is necessary when the proposed VBS scheme is extended for LPM using one-sided kernel.

V. LPM for Time-Varying Nonlinear Systems

Let us now consider the extension of the proposed LPM method to identify a weakly nonlinear time-varying system with a known structure as follows:

$$y(t)=\phi\left(\boldsymbol{x}^T(t)\boldsymbol{a}(t)\right)+\sigma(t)\varepsilon(t) \tag{43}$$

where $\phi(u)$ is a nonlinear scalar-valued function of $u=\boldsymbol{x}^T(t)\boldsymbol{a}(t)$ and the other variables follow the definitions in (2). Generally, the time-varying parameters $\boldsymbol{a}(t)$ can be estimated by nonlinear estimation methods, such as the extended KF (EKF), which linearizes the nonlinear function at the current state. The LPM approach has the advantage of using a variable bandwidth to cater to nonstationary parameter vector. To extend the LPM approach to (43), the nonlinear function $\phi(\boldsymbol{x}^T\boldsymbol{a})$ can be approximated as a first-order Taylor series expansion at a given point $\boldsymbol{a}_0$

$$\phi(\boldsymbol{x}^T\boldsymbol{a})=\phi(\boldsymbol{x}^T\boldsymbol{a}_0)+d\phi/du|_{u=\boldsymbol{x}^T\boldsymbol{a}_0}\cdot\boldsymbol{x}^T(\boldsymbol{a}-\boldsymbol{a}_0) \tag{44}$$

where $d\phi/du$ is the derivative of ϕ evaluated at $\boldsymbol{x}^T\boldsymbol{a}_0$. Provided that the nonlinearity of ϕ is mild and the expansion points are appropriately chosen, the nonlinear function around the expansion points can be well approximated by a first-order Taylor series expansion of (44), which is a linear function of $\boldsymbol{a}$. Using the expansion of (44), the nonlinear system (43) becomes

$$y(t)\approx\phi\left(\boldsymbol{x}^T(t)\boldsymbol{a}_0(t)\right)+d\phi/du|_{u=\boldsymbol{x}^T(t)\boldsymbol{a}_0(t)} \cdot\boldsymbol{x}^T(t)\left(\boldsymbol{a}(t)-\boldsymbol{a}_0(t)\right)+\sigma(t)\varepsilon(t). \tag{45}$$

For convenience, we can absorb the linearization error into the noise term and replace the approximate sign by the equality sign. Consequently, (45) can be written as a linear system of $\boldsymbol{a}(t)$ as follows:

$$y_\phi(t)=d\phi/du|_{u=\boldsymbol{x}^T(t)\boldsymbol{a}_0(t)}\boldsymbol{x}^T(t)\boldsymbol{a}(t)+\sigma(t)\varepsilon(t) \tag{46}$$

where

$$y_\phi(t)=y(t)-\phi(\boldsymbol{x}^T(t)\boldsymbol{a}_0(t)) + d\phi/du|_{u=\boldsymbol{x}^T(t)\boldsymbol{a}_0(t)}\boldsymbol{x}^T(t)\boldsymbol{a}_0(t).$$

We can see that, through the linearization process, the time-varying nonlinear system can be approximated by a TVLS, which can be solved by the proposed LPM method. Although bias may be introduced by the linearization process, its contribution is small if the expansion point is close to the true value of the coefficient, and the nonlinearity is mild. To this end, the expansion points are selected by the following two procedures: 1) In the ICI method, expansion points can be assigned recursively as the coefficient estimates at previous time instant, i.e., $\boldsymbol{a}_0(t)=\hat{\boldsymbol{a}}(t-1)$, and 2) during the approximation of MSE, the expansion points are selected as the coefficients estimated by the ICI method at the same time point. Simulation results showed that the proposed scheme for selecting expansion points can achieve satisfactory results.

Before presenting the simulation results, let us clarify the two Taylor expansions used in the LPM method for time-varying nonlinear system. The Taylor series expansion mentioned in this section is employed to linearize the nonlinear system, and its order is one, i.e., linear. On the other hand, the Taylor series

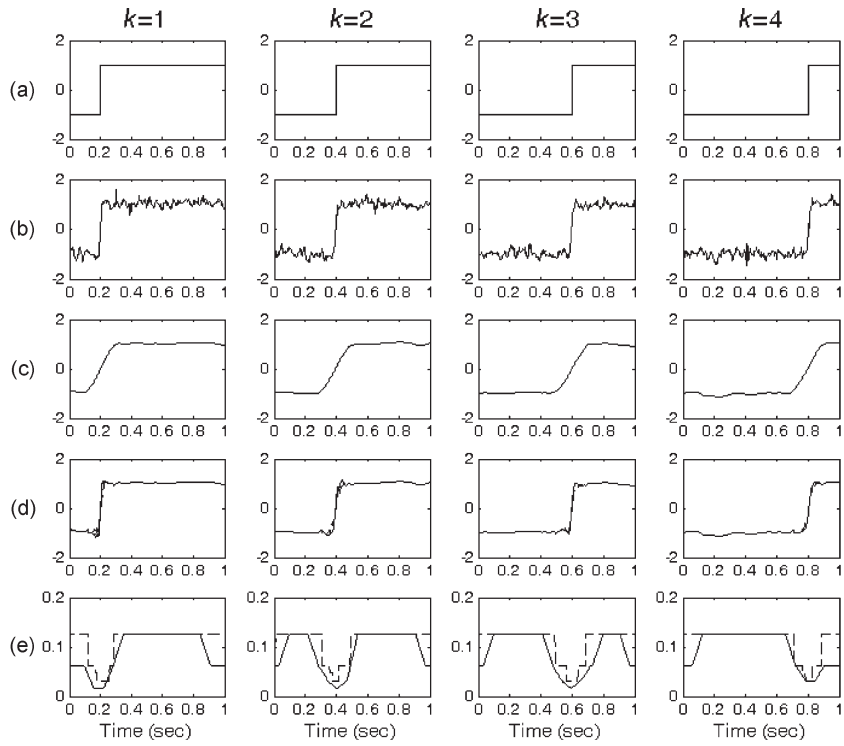

Fig. 1. LPM for identification of a jumping-coefficient linear system: (a) True coefficients. (b) LPM estimates with a constant small kernel ($h = 1/64$). (c) LPM estimates with a constant large kernel ($h = 1/8$). (d) LPM-VBS estimates (dashed lines: estimates with h_{ICI}; solid lines: estimates with h_{MMSE}). (e) Adaptive variable bandwidths (dashed lines: h_{ICI}; solid lines: h_{MMSE}).

expansion discussed in Section II is used to approximate the variations of the coefficient vector, and the expansion order p can be any possible integers including one. The latter reduces to a local linear approximation $(p = 1)$ and is used in the simulation in this paper. The estimation bias considered in this paper results from the finite expansion order of the latter Taylor series expansion. However, the linearization error of the nonlinearity is somewhat difficult to characterize.

VI. Simulation Results

A. Jumping-Coefficient Linear Systems

We first test the proposed LPM-VBS method using a TVLS with both jumping and static coefficients, as shown in Fig. 1(a). These types of situations occur, for example, in fault detection of dynamic systems and other related applications. The input data x was generated from a Gaussian process with zero mean and unit variance. The order of the TVLS was $L = 4$. The sampling rate was 512 Hz, and the number of measurements was $n = 512$. The length of I_l was set to 32 to produce a total of 16 subintervals. A zero-mean white Gaussian noise with variance $\sigma^2 = 0.1$ was added to the test signal so that the SNR was around 10 dB. The polynomial order used in LPM was $p = 1$, and the excess order for approximating the bias was $p_{ex} = 2$. Epanechnikov kernels were employed, and the bandwidth set for ICI was chosen as $\widetilde{\boldsymbol{H}} = \{1/64, 1/32, 1/16, 1/8\}$, i.e., $h_0 = 1/64$, $h_a = 2$, and $M = 4$.

The LPM-VBS with variable local bandwidth (h_{ICI} and h_{MMSE}) was compared with LPM with a constant bandwidth, which used one bandwidth for the whole data. It can be seen clearly from Fig. 1 that, in LPM 1) A small bandwidth can detect fast change of coefficients, but it may lead to large variations for slow-varying coefficients, as shown in Fig. 1(b); 2) a large bandwidth can obtain smooth estimates when the coefficients varied slowly, but it cannot accurately estimate fast-varying coefficients, as shown in Fig. 1(c); and 3) local variable bandwidth can obtain satisfactory results for the whole data set by employing small bandwidths for jumping coefficients and large bandwidths for slowly varying coefficients, as shown in Fig. 1(d) and (e).

The estimation bias, variance, and MSE for each coefficient can be estimated by means of Monte Carlo simulations. Suppose that we perform Γ independent Monte Carlo realizations of the simulated TVLS and denote the average of the estimated coefficients as $\bar{a}(k,t) = (1/\Gamma)\sum_{\gamma=1}^{\Gamma} \hat{a}(k,t;\gamma)$, where $\hat{a}(k,t;\gamma)$ is the estimated coefficient of the γth realization. The estimation bias, variance, and MSE were computed analogously as

$$\mathrm{Bias}\left[\hat{a}(k,t)\right] = \bar{a}(k,t) - a(k,t) \tag{47}$$

$$\mathrm{Var}\left[\hat{a}(k,t)\right] = \frac{1}{\boldsymbol{\Gamma}} \sum_{\gamma=1}^{\Gamma} \left[\hat{a}(k,t\gamma) - \bar{a}(k,t)\right]^2 \tag{48}$$

$$\mathrm{MSE}\left[\hat{a}(k,t)\right] = \frac{1}{\boldsymbol{\Gamma}} \sum_{\gamma=1}^{\Gamma} \left[\hat{a}(k,t\gamma) - a(k,t)\right]^2. \tag{49}$$

The estimation bias, variance, and MSE of $a(1,t)$ in the jumping-coefficient system are shown in Fig. 2. It can be seen that the estimation bias for small h is much smaller than that for large h around the jump discontinuity, while at flat areas, the estimation bias is rather small. On the other hand, the

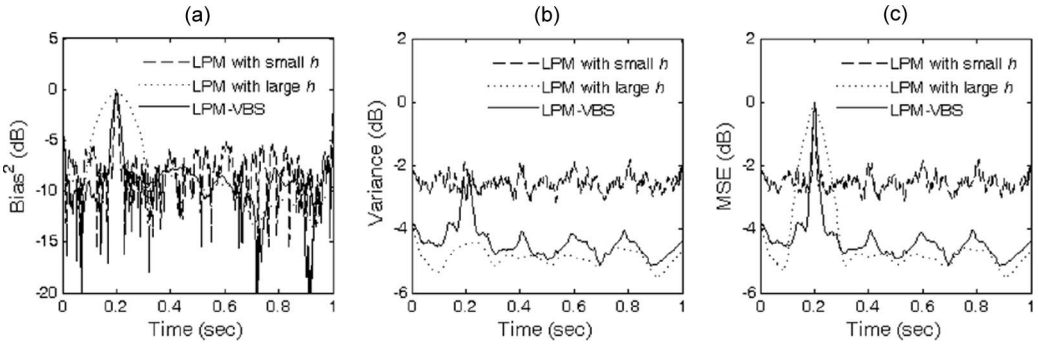

Fig. 2. Estimation bias, variance, and MSE of $a(1,t)$ in the jumping-coefficient linear system. (a) Squared bias. (b) Variance. (c) MSE. Dashed lines: LPM estimates with a constant small kernel ($h = 1/64$); dotted lines: LPM estimates with a constant large kernel ($h = 1/8$); solid lines: LPM-VBS estimates with h_{MMSE}.

TABLE I
EMSD COMPARISONS BETWEEN LPM WITH DIFFERENT BANDWIDTH-SELECTION SCHEMES FOR JUMPING-COEFFICIENT LINEAR SYSTEMS (UNIT: DECIBELS)

Methods	SNR=20	SNR=10	SNR=5	SNR=0
LPM with h=1/64	-15.87	-6.65	-1.73	3.13
LPM with h=1/32	-18.98	-10.75	-6.19	-1.66
LPM with h=1/16	-18.24	-12.06	-8.34	-4.62
LPM with h=1/8	-11.23	-9.17	-7.49	-5.39
LPM-VBS with h_{ICI}	-19.25	-13.13	-8.97	-5.41
LPM-VBS with h_{MMSE}	-20.82	-13.68	-9.09	-4.93

estimation variance for large h is obviously smaller than that for small h. As for the MSE, we can see that a small h achieves a smaller MSE around the jump discontinuity, while a large h has a smaller MSE at flat areas. The proposed local variable bandwidths, which are small around the jump discontinuity and large at flat areas, can obtain satisfactory results for the whole period.

Next, the ensemble mean-squared deviation (EMSD) from the true coefficients was calculated and used as the performance measure for the whole time period

$$\mathrm{EMSD} = 10\log_{10}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{L}\left[a(k,t_i)-\hat{a}(k,t_i)\right]^2\right\}. \quad (50)$$

Table I shows the averaged EMSD values over $\Gamma = 100$ independent runs. It can be seen that the LPM-VBS with variable bandwidth (h_{ICI} and h_{MMSE}) generally had better performances than those with constant bandwidths, and the LPM with h_{MMSE} slightly outperformed the LPM with h_{ICI}. At high noise level, for example, SNR $= 0$, a larger constant bandwidth may achieve a slightly better result than adaptive variable bandwidth. This may be due to the large variance of the estimated quantities which degrade slightly the performance of the VBS method.

B. Smooth-Coefficient Linear Systems

The aforementioned jumping-coefficient system contains two extreme types of coefficient variations so that it can effectively illustrate the results of VBS. In order to check whether the optimal variable bandwidths obtained by the proposed method are in high accordance with the asymptotic optimal variable bandwidths derived in (30), a smooth-coefficient linear system with time-varying coefficients $a(1,t) = \cos(3t)$ and $a(2,t) = 2\exp(-16t^2)$ was also tested, as the coefficients contain different extents of variations and their derivatives can be easily calculated. The previous jump-coefficient system, whose coefficient derivatives are mostly zero, is unsuitable because most of the asymptotic optimal bandwidth will be infinity. In total, $n = 200$ samples were uniformly distributed in the time interval $[-2, 2]$ so that the sampling density $f(t)$ is 1/4. The input vector $x(1,t)$ and $x(2,t)$ were independently generated from the Gaussian distribution with zero mean and unit variance. As a result, the correlation matrix of the input was given as $\boldsymbol{R}_x = \boldsymbol{I}_2$. A zero-mean white Gaussian noise was added with an SNR of 5 dB. The bandwidth set was chosen as $\widetilde{\boldsymbol{H}} = \{0.08, 0.16, 0.32, 0.64, 1.28\}$. The other parameters for the LPM were the same as those in previous simulations. Thus, all the quantities in the expression of asymptotic optimal variable bandwidth (30) can be computed, and the resultant asymptotic optimal variable bandwidth h_{asy} is shown as dashed lines in Fig. 3(c) and (d). The estimated optimal bandwidth in Fig. 3(c) and (d) were averaged from 100 independent

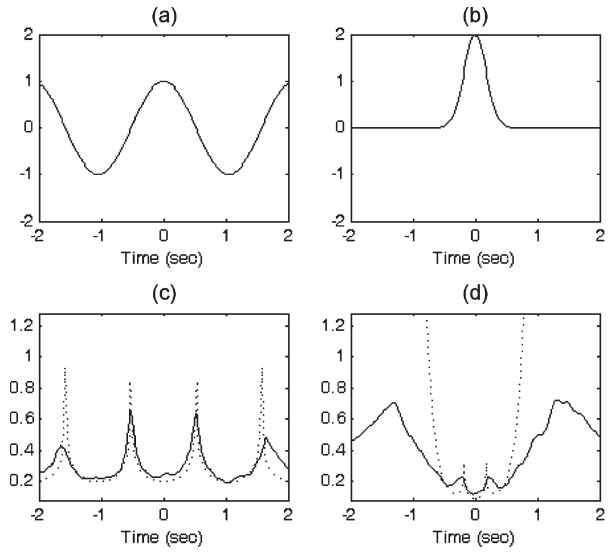

Fig. 3. LPM for identification of a smooth-coefficient linear system: (a) True coefficients $a(1,t)$. (b) True coefficients $a(2,t)$. (c) (Dashed line) Asymptotic optimal variable bandwidth $h_{\mathrm{asy}}(1,t)$ and (solid line) adaptive variable bandwidth $h_{\mathrm{MMSE}}(1,t)$. (d) (Dashed line: those too large values over 1.28 were not shown) Asymptotic optimal variable bandwidth $h_{\mathrm{asy}}(2,t)$ and (solid line) adaptive variable bandwidth $h_{\mathrm{MMSE}}(2,t)$.

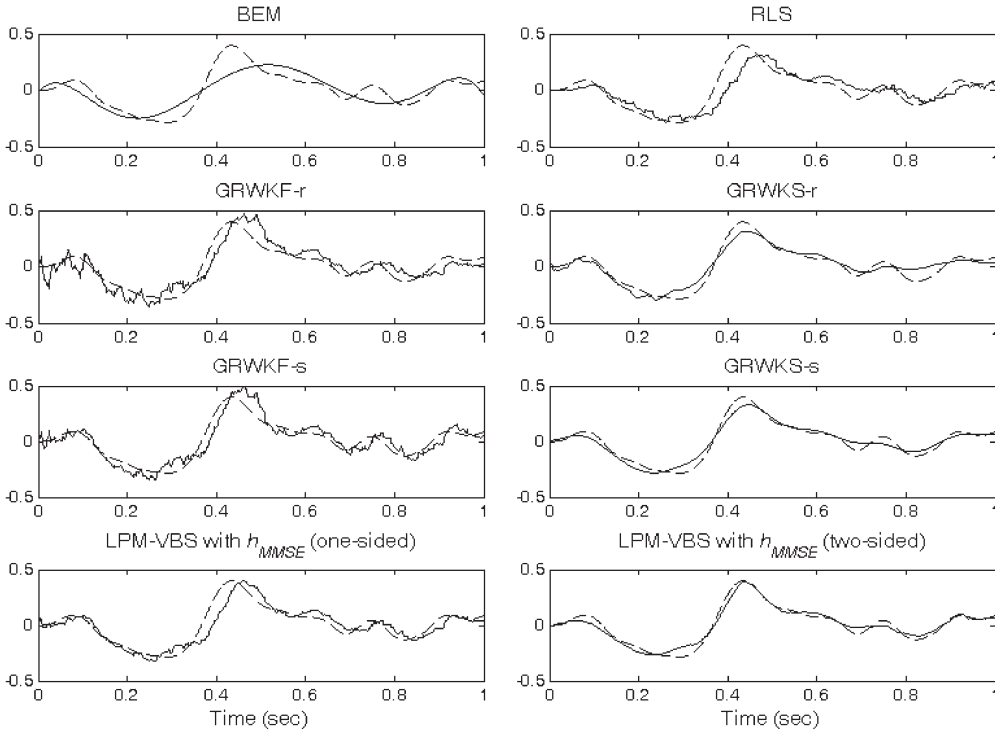

Fig. 4. Comparisons between various TVLS identification methods for a random-coefficient linear system ($f_c = 0.02,\ \text{SNR} = 5$ dB). Dashed lines denote the true coefficients, and solid lines are the estimates.

Monte Carlo runs. It can be seen that, overall, the consistency between the theoretically derived asymptotic optimal bandwidth and the estimated optimal bandwidth is good, particularly when the theoretical bandwidth is small (because one of the asymptotic assumptions is $h \to 0$). When the derivative is small, the estimated bandwidth h_{MMSE} is not as large as the theoretical asymptotic bandwidth h_{asy}. It is due to the fact that the asymptotic bandwidth h_{asy} is based on the assumption that $n \to \infty$, while the estimated bandwidth h_{MMSE} is obtained from a finite number of samples. In addition, the bandwidth-selection results were not as good at the beginning and the end of the finite-sample data because of the zero-padded points on the left and right sides of the samples, respectively.

C. Random-Coefficient Linear Systems

We further test the performance of the proposed LPM-VBS method in a more general random-coefficient system under different noise situations. In the stimulated TVLSs, the time-varying coefficients were generated by filtering white Gaussian signals with zero mean and unit variance using low-pass filters. The resultant TVLS has a coefficient function varying considerably over time, and the extent of variations is determined by the cutoff frequency of the low-pass filter. Four cutoff frequencies f_c (normalized by the sampling rate $f_s = 512$ Hz): 0.01, 0.02, 0.05, and 0.1, were used to simulate different extents of coefficient variations. Zero-mean white Gaussian noises with different SNRs: 20, 10, 5, and 0 dB, were added to simulate different noise conditions. These examples can be used to demonstrate the effectiveness of the proposed method in detecting rapid system change, which is a problem extensively studied with various practical applications [5], [29]. Conventionally, the change-detection problem is addressed by various adaptive filtering and KF approaches. Here, we compare the proposed LPM-VBS method with other conventional TVLS identification or change-detection methods, including the RLS, the KF, and the BEM methods. We also considered the LPM-VBS with one-sided kernel so that the online performances of various identification methods can be fairly compared. In the BEM method, polynomial basis functions were employed, and a set of expansion orders $p_{\text{BEM}} = [1, 5, 10, 15, 20]$ was tested to determine the one with the best EMSD value for further comparison with other TVLS methods. For RLS, a set of forgetting factors $\lambda = [0.8, 0.85, 0.9, 0.95, 0.99]$ was tested, and the results with the best EMSD performance were selected for further comparison. As for the KF, since it is difficult and outside the scope of this paper to compare the proposed LPM method with all KF variants, we shall compare the proposed method with the conventional GRWKF and the fixed-interval GRWKS methods. The fixed-interval GRWKS is also tested because it can take advantage of future measurements to effectively avoid the tracking-lag problem of the GRWKF [10]. For a fair comparison with LPM using order $p = 1$, the order of the generalized random walk model in GRWKF and GRWKS is set as $p = 2$, which implies that coefficients can be presented as a first-order polynomial plus a white-noise process. Two parameter-estimation schemes for the GRWKF were considered, and they were the smoothness-priors-constrained approach in [6] and the recursive updating approach in [28]. Hereinafter, the GRWKF with smoothness-prior-constrained parameter estimation is denoted as GRWKF-s, while the GRWKF with recursive parameter estimation is denoted as GRWKF-r. For LPM using two-sided kernels, $\widetilde{\boldsymbol{H}} = \{1/64, 1/32, 1/16, 1/8\}$, while for LPM using one-sided kernels, $\widetilde{\boldsymbol{H}} = \{1/32, 1/16, 1/8, 1/4\}$.

TABLE II
EMSD COMPARISONS BETWEEN VARIOUS TVLS IDENTIFICATION METHODS FOR RANDOM-COEFFICIENT LINEAR SYSTEMS (UNIT: DECIBELS)

Methods	f_c=0.01 Hz			
	SNR=20	SNR=10	SNR=5	SNR=0
BEM	-25.79	-25.49	-25.06	-24.35
RLS	-33.69	-28.33	-25.44	-22.42
GRWKF-r	-29.78	-26.16	-22.96	-19.49
GRWKS-r	-36.38	-31.93	-28.75	-24.90
GRWKF-s	-34.23	-28.70	-24.31	-20.24
GRWKS-s	-43.49	-35.43	-30.84	-26.53
LPM-VBS $^{(I)}$ with h_{MMSE}	-35.99	-27.66	-23.88	-19.94
LPM-VBS $^{(II)}$ with h_{MMSE}	-42.53	-34.81	-30.88	-26.59
Methods	f_c=0.02 Hz			
	SNR=20	SNR=10	SNR=5	SNR=0
BEM	-15.62	-16.07	-15.45	-15.07
RLS	-25.03	-22.47	-19.53	-16.84
GRWKF-r	-21.41	-19.73	-17.94	-14.50
GRWKS-r	-26.23	-24.43	-22.81	-19.41
GRWKF-s	-20.29	-20.24	-18.66	-15.31
GRWKS-s	-24.40	-25.48	-24.17	-20.92
LPM-VBS $^{(I)}$ with h_{MMSE}	-28.64	-22.17	-18.32	-14.46
LPM-VBS $^{(II)}$ with h_{MMSE}	-34.88	-28.18	-24.31	-20.92
Methods	f_c=0.05 Hz			
	SNR=20	SNR=10	SNR=5	SNR=0
BEM	-9.06	-8.79	-9.04	-8.54
RLS	-14.68	-13.52	-12.36	-9.82
GRWKF-r	-11.84	-11.44	-10.60	-8.75
GRWKS-r	-14.48	-14.36	-13.72	-12.27
GRWKF-s	-9.52	-9.32	-9.09	-8.19
GRWKS-s	-10.62	-10.73	-10.55	-10.19
LPM-VBS $^{(I)}$ with h_{MMSE}	-17.59	-12.83	-10.64	-8.06
LPM-VBS $^{(II)}$ with h_{MMSE}	-23.79	-17.98	-14.99	-11.67
Methods	f_c=0.1 Hz			
	SNR=20	SNR=10	SNR=5	SNR=0
BEM	-5.41	-5.31	-5.25	-5.28
RLS	-8.19	-7.57	-6.78	-5.54
GRWKF-r	-6.48	-6.39	-5.77	-4.43
GRWKS-r	-7.97	-8.16	-7.79	-6.94
GRWKF-s	-5.23	-5.13	-4.90	-4.28
GRWKS-s	-5.81	-5.83	-5.78	-5.56
LPM-VBS $^{(I)}$ with h_{MMSE}	-7.15	-6.29	-5.51	-4.56
LPM-VBS $^{(II)}$ with h_{MMSE}	-10.41	-8.35	-7.03	-6.06

The other parameters for the LPM were the same as those in previous simulations.

Fig. 4 shows the estimation results of one coefficient of a sample four-order TVLS ($f_c = 0.02$ Hz and SNR $= 5$ dB), which have both slowly changing and rapid-changing coefficients. It can be seen clearly that BEM gives a very smooth estimation, but the rapid change around 0.4 s is smoothed out at the same time. The RLS, GRWKF, and GRWKS have better tracking abilities than BEM, but it is still difficult for them to track rapid changes accurately. Overall, the LPM-VBS methods are seen to offer a better adaptability than other methods tested for both rapid changes and slow-varying parts.

For quantitative comparison, Table II lists the EMSD values averaged over 200 independent runs, and these values are shown graphically in Fig. 5. We can conclude from Table II and Fig. 5 the following: 1) The proposed LPM-VBS method has a better performance than other methods tested for TVLSs having various degrees of coefficient variations and different noise levels; 2) the BEM method can provide a result comparable with the proposed LPM-VBS method if the system coefficients are very smooth (f_c is small); 3) the LPM using

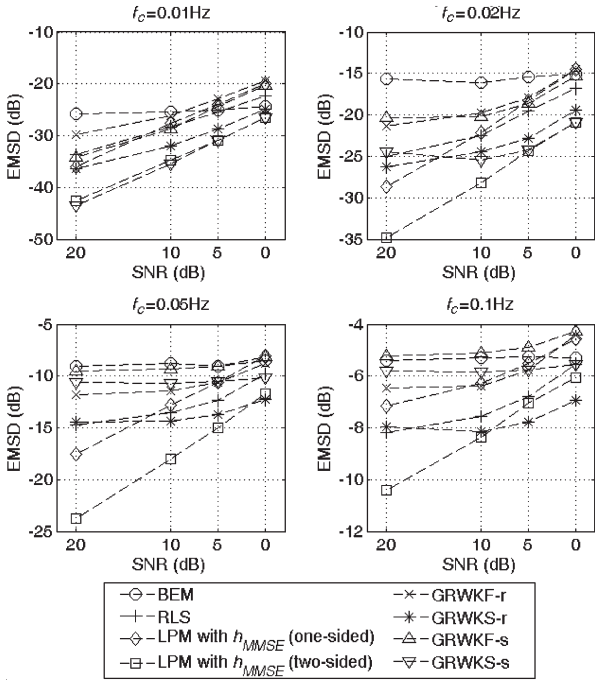

Fig. 5. EMSD comparisons between various TVLS identification methods for random-coefficient linear systems.

one-sided kernel method has relatively lower EMSD values than other tracking methods when the noise is small and the variability of the system is not large, but the performances of all tracking methods are comparable under heavy noise and large system variability; and 4) the GRWKF-s method has a slightly better performance than LPM-VBS and GRWKF-r when the coefficient variation is small ($f_c = 0.01$ Hz) (top left panel in Fig. 5), while the GRWKF-r method outperformed LPM-VBS and GRWKF-s if the coefficient variation is large ($f_c = 0.1$ Hz) with low SNR (lower right panel in Fig. 5). The performance difference between GRWKF-r and GRWKF-s is due to the facts that the smoothness-constrained parameter selection adapts better to smooth coefficient variations, while the recursive updating can respond better to fast coefficient variations. In most testing scenarios, the proposed LPM-VBS method (in particular, LPM-VBS with two-sided kernel supports) achieves better performance than the GRWKF/GRWKS methods except for the two situations mentioned previously. In conclusion, the LPM method, which is more "data-based," has an evident advantage over the BEM/GRWKF/GRWKS methods, which are more "model-based" when *a prior* model is not precisely given, and the measurements are trustable (i.e., SNR is high). When the amount of noise is large (i.e., the data are not so reliable) or the prior model knowledge is correctly given or estimated, it is possible for the model-based methods such as GRWKS to outperform the proposed LPM method.

D. Random-Coefficient Quadratic Systems

In this experiment, we evaluate the performances of the proposed LPM-VBS method for time-varying and mildly nonlinear systems. Such systems are frequently encountered in linear systems with output passing through memoryless nonlinearities such as the sigmoidal function for modeling amplitude

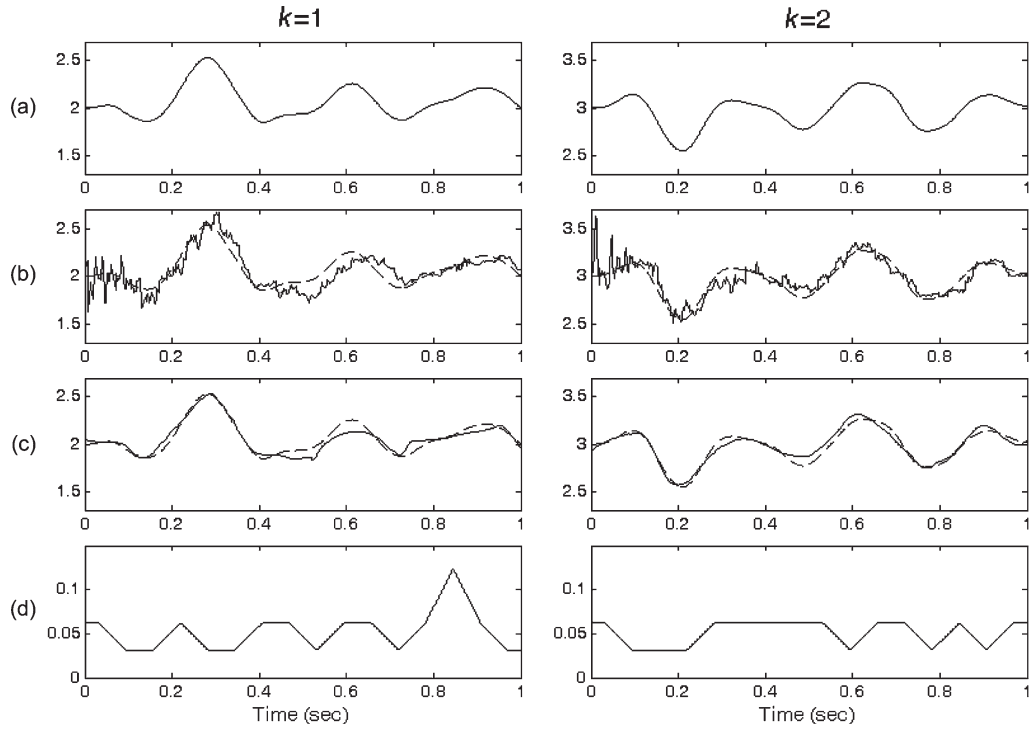

Fig. 6. Comparisons between various nonlinear identification methods for random-coefficient quadratic systems ($f_c = 0.02$ Hz, SNR $= 10$ dB): (a) True coefficients (b) EKF. (c) LPM-VBS estimates with variable bandwidths h_{MMSE}. (d) Variable bandwidths h_{MMSE}. Dashed lines in (b) and (c) denote the true coefficients, and solid lines are the estimates.

saturation [30]. For illustrative purposes, the following nonlinear quadratic system $y(t) = [\boldsymbol{x}^T(t)\boldsymbol{a}(t)]^2 + \sigma(t)\varepsilon(t)$ is considered in this paper. The order of the quadratic system is $L = 2$, and the two channels of coefficients $a(1,t)$ and $a(2,t)$ were, respectively, generated by filtering Gaussian processes $\aleph(2,1)$ and $\aleph(2,1)$ using low-pass filters. Similar to previous experiments, different cutoff frequencies f_c: 0.01, 0.02, 0.05, and 0.1 Hz, and different amount of noise with SNRs: 20, 10, 5, and 0 dB, were used to simulate a wide variety of testing conditions. The input data x was generated from a uniform distribution on [0, 1]. The sampling rate was 512 Hz, and the number of measurements used was $n = 512$. The length of I_l was set to 32. The other parameters were the same as those used in previous simulations. The EKF was tested for comparison with the proposed LPM-VBS (two-sided kernel, variable bandwidth h_{MMSE}) method.

One representative example when $f_c = 0.02$ Hz and SNR $=$ 10 dB was shown in Fig. 6. We can see that the LPM-VBS method has a reduced estimation variability than EKF. Moreover, the LPM-VBS method can accurately estimate the rapid changes in the system coefficients. The EMSD values listed in Table III were averages of 100 Monte Carlo runs, and they further substantiated the usefulness of the proposed LPM-VBS method quantitatively.

TABLE III
EMSD COMPARISONS BETWEEN VARIOUS NONLINEAR IDENTIFICATION METHODS FOR RANDOM-COEFFICIENT QUADRATIC SYSTEMS (UNIT: DECIBELS)

f_c (Hz)	Methods	SNR=20	SNR=10	SNR=5	SNR=0
0.01	EKF	-23.00	-15.19	-10.49	-6.74
	LPM-VBS	-28.27	-20.20	-15.20	-10.55
0.02	EKF	-20.96	-15.04	-10.22	-6.22
	LPM-VBS	-25.33	-18.95	-14.72	-9.48
0.05	EKF	-16.72	-12.63	-9.20	-5.09
	LPM-VBS	-18.48	-13.86	-11.68	-7.90
0.1	EKF	-12.06	-9.66	-7.51	-4.87
	LPM-VBS	-12.22	-10.08	-8.45	-6.27

E. Tracking of Voltage Flicker in Power Distribution Systems

In power distribution systems, voltage flicker or fluctuation is often produced by high electricity load alternations, such as arc-furnace operations and resistance wielding, and it will cause serious quality problems to the power systems and consumers [4]. Accurate envelope tracking of the voltage measurements is therefore important not only for evaluating the flicker level but also for compensating the flicker and regulating the voltage. Usually, the measured voltage $v(t)$ can be modeled as a sinusoidal waveform with time-varying amplitudes as follows [4]:

$$\begin{aligned} v(t) &= A(t)\sin(\omega t + \phi) + e_v(t) \\ &= A(t)[\sin\omega t\cos\phi + \cos\omega t\sin\phi] + e_v(t) \\ &= [\sin\omega t, \cos\omega t]\begin{bmatrix} A(t)\cos\phi \\ A(t)\sin\phi \end{bmatrix} + e_v(t) \\ &= \Omega^T(t)\Phi(t) + e_v(t) \end{aligned} \tag{51}$$

where $A(t)$ is the voltage amplitude, ω is the supply angular frequency, ϕ is the phase angle, and $e_v(t)$ is the additive measurement noise. The voltage envelope is obtained as the L_2 norm of the estimated $\Phi(t)$. We can see that the voltage model (51) is in the form of a TVLS with $v(t)$ as the measured output, $\Omega(t)$ as the known input, $\Phi(t)$ as the time-varying coefficient,

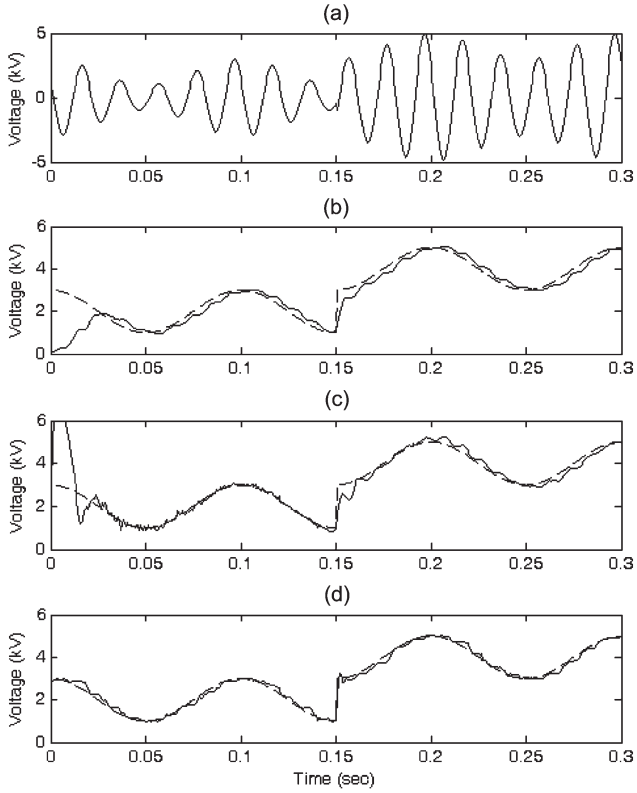

Fig. 7. Comparisons between various methods for tracking envelope of voltage flicker: (a) Measured voltage (b) RLS. (c) GRWKF. (d) LPM-VBS. Dashed lines in (b)–(d) denote the true envelope, and solid lines are the estimates.

and the system order L is equal to two. We now evaluate the performance of the proposed LPM-VBS method in tracking the envelope of the flicker and compare it with conventional RLS and GRWKF methods [4]. The measured voltage is assumed to be

$$v(t) = \begin{cases} (2+\cos 20\pi t)\cos(100\pi t + \pi/3), & 0 < t \le 0.15s \\ (4+\cos 20\pi t)\cos(100\pi t + \pi/3), & t > 0.15s. \end{cases} \tag{52}$$

The envelope of the voltage, which was modulated by a sinusoid, fluctuated periodically and mildly, while a rapid change occurred at time 0.15 s. The sampling rate was set at 1600 Hz (32 samples per cycle). A zero-mean white Gaussian noise component with SNR $= 30$ dB was added. The forgetting factor for the RLS algorithm was 0.7, as recommended in [4]. In the GRWKF method, the order of generalized random walk model is set as two, and the covariance matrices of state and observation noises were estimated recursively for online tracking as in [28]. For tracking purposes, LPM-VBS with one-sided kernel was tested, and the bandwidth set was $\widetilde{\boldsymbol{H}} = \{1/200, 1/100, 1/50, 1/25\}$. The other parameters for the LPM were the same as those in previous simulations.

We can see from Fig. 7 that the proposed LPM-VBS method achieved a good envelope-tracking performance for both rapid voltage change and slow voltage fluctuations. On the other hand, the RLS and GRWKF methods showed a slow response to rapid change and a slow convergence rate. By comparing the EMSD values (after 0.05 s when RLS and GRWKF converge) of the three methods (RLS: -11.93 dB, GRWKF: -14.16 dB, LPM-VBS: -17.67 dB), we can conclude that the LPM-VBS had a better envelope-tracking performance than the RLS and GRWKF methods.

VII. Conclusion

A novel LPM method for identification of TVLSs and its asymptotic performance analysis have been presented. A new data-driven VBS scheme was also developed to minimize the local MSE. Simulation results showed that the performance of the LPM-VBS method was better than the conventional RLS-, GRWKF-, and GRWKS-based TVLS identification methods in most testing scenarios, particularly when the SNR is moderate to high. Moreover, the LPM-VBS method was further extended for identification of time-varying systems with mild nonlinearity. An application of proposed LPM-VBS method for tracking of voltage flicker was presented with better performance over conventional RLS and GRWKF methods. The LPM-VBS method is expected to find various other applications like change detection and model identification in communications, biosignal processing, instrument testing, power systems and delivery, etc.

Acknowledgment

The authors would like to thank Dr. K. M. Tsui and the anonymous reviewers for their valuable suggestions, which greatly improve the presentation of this paper.

References

[1] J. Kantz, J. Waldmann, and F. M. Landstorfer, "Measuring system for time-variant impedances," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 1, pp. 258–263, Feb. 2005.

[2] A. B. Gershman, "Adaptive beamforming for multiantenna communications," in *Adaptive Signal Processing for Wireless Communications*, M. Ibnkahla, Ed. Boca Raton, FL: CRC Press, 2008, pp. 201–231.

[3] M. Aboy, O. W. Márquez, J. McNames, R. Hornero, T. Thong, and B. Goldstein, "Adaptive modelling and spectral estimation of nonstationary biomedical signals based on Kalman filtering," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1485–1489, Aug. 2005.

[4] M. I. Marei, E. F. El-Saadany, and M. M. A. Salama, "Envelope tracking techniques for flicker mitigation and voltage regulation," *IEEE Trans. Power Del.*, vol. 19, no. 4, pp. 1854–1861, Oct. 2004.

[5] M. Niedzwiecki, *Identification of Time-Varying Processes*. Chichester, U.K.: Wiley, 2000.

[6] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Autom. Control*, vol. AC-30, no. 1, pp. 48–56, Jan. 1985.

[7] Y. J. Zheng and Z. P. Lin, "Recursive adaptive algorithms for fast and rapidly time-varying systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 50, no. 9, pp. 602–614, Sep. 2003.

[8] D. G. Manolakis, V. K. Ingle, and S. M. Kogan, *Statistical and Adaptive Signal Processing*. New York: McGraw-Hill, 2000.

[9] J. Proakis, C. Rader, F. Ling, C. Nikias, M. Moonen, and I. Proudler, *Algorithms for Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[10] C. K. Chui and G. Chen, *Kalman Filtering: With Real-Time Applications*, 2nd ed. New York: Springer-Verlag, 1999.

[11] E. Kamen and J. Su, *Introduction to Optimal Estimation*. London, U.K.: Springer-Verlag, 1999.

[12] T. S. Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *J. R. Stat. Soc. Ser. B*, vol. 32, no. 2, pp. 312–322, 1970.

[13] G. B. Giannakis and C. Tepedelenliolu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.

[14] J. Fan and I. Gijbels, "Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation," *Stat. Sin.*, vol. 57, no. 2, pp. 371–394, 1995.

[15] J. Fan, I. Gijbels, T. C. Hu, and L. S. Huang, "A study of variable bandwidth selection for local polynomial regression," *Stat. Sin.*, vol. 6, no. 1, pp. 113–127, 1996.
[16] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall, 1996.
[17] H. Lanshammar, "On precision limits for derivatives numerically calculated from noisy data," *J. Biomech.*, vol. 15, no. 6, pp. 459–470, 1982.
[18] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567–2571, Sep. 1999.
[19] V. Katkovnik, K. Egiazarian, and J. Astola, *Local Approximation Techniques in Signal and Image Processing*. Bellingham, WA: SPIE Press, Sep. 2006.
[20] Z. G. Zhang, S. C. Chan, K. L. Ho, and K. C. Ho, "On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data," *J. Signal Process. Syst.*, vol. 52, no. 3, pp. 263–280, Sep. 2008.
[21] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Methods Statist.*, vol. 6, no. 2, pp. 135–170, 1997.
[22] T. J. Hastie and R. J. Tibshirani, "Varying-coefficient models," *J. R. Stat. Soc. Ser. B*, vol. 55, no. 4, pp. 757–796, 1993.
[23] D. R. Hoover, J. A. Rice, C. O. Wu, and L. P. Yang, "Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data," *Biometrika*, vol. 85, no. 4, pp. 809–822, 1998.
[24] J. Fan and W. Zhang, "Statistical estimation in varying coefficient models," *Ann. Statist.*, vol. 27, no. 5, pp. 1491–1518, 1999.
[25] J. Fan, Q. Yao, and Z. Cai, "Adaptive varying-coefficient linear models," *J. R. Stat. Soc. Ser. B*, vol. 65, no. 1, pp. 57–80, 2003.
[26] W. Zhang and S. Y. Lee, "Variable bandwidth selection in varying-coefficient models," *J. Multivar. Anal.*, vol. 74, no. 1, pp. 116–134, Jul. 2000.
[27] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.
[28] Z. G. Zhang, S. C. Chan, and K. M. Tsui, "A recursive frequency estimator using linear prediction and a Kalman filter-based iterative algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 6, pp. 576–580, Jun. 2008.
[29] F. Gustafsson, *Adaptive Filtering and Change Detection*. Chichester, U.K.: Wiley, 2000.
[30] S. C. Chan, Y. Zhou, and W. Y. Lau, "Approximate QR-based algorithms for recursive nonlinear least squares estimation," in *Proc. IEEE ISCAS*, Kobe, Japan, May 23–26, 2005, vol. 5, pp. 4333–4336.

S. C. Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from The University of Hong Kong, Hong Kong, in 1986 and 1992, respectively.

He was with the City Polytechnic of Hong Kong in 1990 as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he is currently a Professor. He was a Visiting Researcher at Microsoft Corporation, Redmond, U.S.; Microsoft, Beijing, China; University of Texas, Arlington; and Nanyang Technological University, Singapore. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed AD converter architecture, bioinformatics, and image-based rendering.

Z. G. Zhang (S'05–M'07) received the B.Sc. degree in electrical and electronic engineering from Tianjin University, Tianjin, China, in 2000, the M.Eng. degree in electrical and electronic engineering from the University of Science and Technology of China, Hefei, China, in 2003, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2008.

He is currently a Postdoctoral Fellow with The University of Hong Kong. His research interests are in general area of statistical signal processing and digital signal processing, and, in particular, in adaptive filtering, nonparametric regression, time–frequency analysis, and biomedical signal processing.