A New Parametric Adaptive Nonstationarity Detector and Application

Y. J. Chu and C. M. Mak[*]

Abstract—**Techniques for hypothesis testing can be used to solve a broad class of nonstationarity detection problems, which is a key issue in a variety of applications. To achieve lower complexity and to deal with real-time detection in practical applications, we develop a new adaptive nonstationarity detector by exploiting a parametric model. A weighted maximum *a posteriori* (MAP) estimator is developed to estimate the parameters associated with the parametric model. We then derive a regularized Wald test from the weighted MAP estimate, which is adaptively implemented by a regularized recursive least squares (RLS) algorithm. Several important issues are discussed, including model order selection, forgetting factor (FF) and regularization parameter selection for RLS, and numerically stable implementation using QR decomposition (QRD), which are intrinsic parts of the proposed parametric adaptive detector. Simulation results are presented to illustrate the efficiency of the proposed nonstationarity detector, with adaptive estimation and automatic model selection, especially for "slowly varying" type of nonstationarity such as time-varying spectrums and speeches.**

Index Terms—**Adaptive nonstationarity detection, Wald test, weighted maximum *a posteriori*, RLS, and adaptive model order selection.**

I. INTRODUCTION

FOR many statistical signal processing methods, e.g. time-series analysis [1], we often need to assume that signals under study are wide sense stationary (WSS) Gaussian random sequences. In many applications, however, data record usually exhibits a nonstationarity and hence results in biased estimate. Related areas include spectrum analysis [2][3], noise reduction [4], speech analysis [5], and biomedical signal processing [6]. It would therefore be important to determine whether the data record is stationary and suitable for further processing.

Much effort has been spent on developing nonstationarity detection methods [7]. These detection techniques can be mainly categorized as the batch- and sequential-based methods, depending on how the data is dealt with. Most of the spectrum-based algorithms belong to the first category [8]–[10]. Since they are based on the Fourier transform [11] that is only true asymptotically, they may not be viable for short data records [12]. Model-based approaches, on the other hand, are usually employed to extract high-resolution estimates in applications where only short data records are available [13]. For example, a time-varying (TV) autoregressive (TVAR) process [14] is often used for modeling signals of TV narrow-peak (or line) spectra [11][15], the high frequency resolution of which is also possible [16]. Then, the model parameters may be tested by using either a Rao test or a generalized likelihood ratio test (GLRT) as proposed, respectively, in [12] and [5]. The Rao test requires the maximum likelihood estimate (MLE) under the null hypothesis, which is usually easier to compute; whereas the GLRT requires MLEs under both null and alternative hypotheses, which is of higher arithmetic complexity and may lead to ill-conditioned problems [17] due to insufficient data sample. Another commonly used technique to determine between alternatives in a binary hypothesis testing problem is the Wald test [18]. It is asymptotically equivalent to GLRT, but only requires the estimate under the alternative hypothesis. This test depends on the parameter estimation method used, which affects its performance in nonstationary environments [7]. For example, insufficient excitation usually increases estimation variance [19]. An insightful investigation about the invariance and coincidence characteristic of GLRT, Rao and Wald tests and their corresponding decision statistics is carried out in [18]. These testing methods are also in a batch form, which are computationally consuming when overlapping is significant. More importantly, this type of nonstationarity detection is retrospective, i.e. to determine whether nonstationarity occurs at a particular point in the sequence requires all the data record available, including those after the change point, and hence is not quite amenable for online applications [3].

To cope with online problems, a class of adaptive detectors are devised, which make decisions at presence of each new sample, and not use a whole batch of processed data samples [20]–[23]. The direct application of these algorithms is the radar system. The practicality of such detectors, however, are quite limited since they usually involve the inversion of a correspondingly large matrix and the possibility of simplification has not been fully addressed in the current literature. Recently, a recursive least squares (RLS) algorithm is applied to nonstationarity detection that solves the matrix inversion recursively [24].

While there has been considerable progress in adaptive detection, it is still highly desirable to develop efficient methods for finding online solutions to the nonstationarity detection problems. In this paper, we propose a parametric adaptive nonstationarity detector, which uses a TV linear model for the unknown system. The model parameters are estimated using the developed weighted maximum *a posteriori* (MAP) estimator. In particular, an exponentially weighted window is

Y. J. Chu and C. M. Mak[*](corresponding author) are with the Acoustic and Noise Control Group, The Hong Kong Polytechnic University, Hong Kong (e-mail: yijing.chu@polyu.edu.hk; cheuk-ming.mak@polyu.edu.hk).

first employed in the likelihood function (LF), namely the weighted LF, which is augmented with an appropriate prior probability of the parameter. The resulting MAP estimator can be implemented adaptively using the so called state-regularized QR decomposition (QRD)-based RLS (SR-QRRLS) algorithm, which is numerically more stable than conventional RLS algorithms due to the QRD structure [25][26]. Then, the Wald statistic is employed, which results in the parametric adaptive nonstationarity detector, namely the regularized recursive Wald test (RWT). In addition, as an intrinsic part of the proposed parametric adaptive detector, we develop an adaptive method, called the state-regularized recursive Bayesian information criterion (SR-RBIC), for model order selection, which applies the weighted posterior probability to the conventional Bayesian model selection criterion [27]. Since both RWT and SR-RBIC use the estimation results from the SR-QRRLS algorithm, which is a critical step for detection, we derive a variable forgetting factor (FF) and regularization parameter for a better estimation through the analytical mean-square error (MSE) of the SR-RLS algorithm.

In summary, the advantages of the proposed detection method are as follows: 1) the adaptive detector can make a decision when each observation is received; 2) the detector has an arithmetic complexity of $O(L^2)$ with L is the model order; 3) the model order can be automatically selected during detection; 4) the estimation and detection performance has been improved at a low excitation; and 5) the FF and regularization parameter can be automatically selected without time-consuming try-and-error procedure.

The rest of the paper is organized as follows. In Section II, the system model and weighted posterior probability are presented. Section III is devoted to the derivation of the adaptive estimator, RWT and SR-RBIC methods. In Section IV, a detailed MSE analysis and implementation of the SR-QRRLS algorithm with variable FF and regularization parameter are discussed. The performance of the proposed estimation and detection methods is evaluated by simulations under different situations in Section V. Conclusions are drawn in Section VI.

II. SYSTEM MODEL AND STATISTICS

A. System Model and Problem Statement

Consider an L-order linear TV finite impulse response (FIR) system with coefficient vector $\boldsymbol{h}(n) = [h_1(n),...,h_L(n)]^T$. The unknown system is excited by an input $\{x(n)\}$. The observed data is $\{d(n)\}$, which is assumed to be corrupted by a zero-mean additive Gaussian noise $\{\eta(n)\}$

$$d(n) = \boldsymbol{x}^T(n)\boldsymbol{h}(n) + \eta(n) \qquad (1)$$

where $\boldsymbol{x}(n) = [x(n),...,x(n-L+1)]^T$ for $n \geq L$. Note the Gaussian noise assumption may need to be modified in case the estimator fails in fitting the model.

In the TV system, the model parameters may constantly change their values. We hence aim to find an online solution to dividing the observed data into pieces of WSS Gaussian sequences [12] via the parameter (hypothesis) test [7], which we call the parametric adaptive nonstationarity detection. Our

framework includes 1) an adaptive algorithm for parameter estimation; 2) a nonstationarity detector derived from the Wald test statistics as well as the developed adaptive algorithm; and 3) a model selector that automatically selects the candidate model during the test procedure. Based on the framework, we could develop an adaptive detection method with automatic model order selection.

In the following, a weighted posterior probability is introduced, based on which the adaptive estimation, detection and model selection algorithms are developed in Section III.

B. Weighted Posterior Probability

To estimate the system parameter in (1), we introduce the following Bayesian linear model, given the data set $\{d(n)\}$ and $\{x(n)\}$ [19]

$$\boldsymbol{d}(n) = \boldsymbol{X}(n)\boldsymbol{w}(n) + \boldsymbol{q}(n) \qquad (2)$$

where $\boldsymbol{d}(n) = [d(n), d(n-1),...,d(L)]^T$ is the observation vector, $\boldsymbol{X}(n) = [\boldsymbol{x}(n)\,\boldsymbol{x}(n-1)\,...\,\boldsymbol{x}(L)]^T$ is the variable matrix, $\boldsymbol{w}(n) = [w_1(n),...,w_L(n)]^T$ is a random vector with a prior probability density function (pdf) to be specified latter in (3), and $\boldsymbol{q}(n)$ is the measurement noise vector of appropriate size with a normal distribution of zero mean and covariance matrix $\sigma_q^2\boldsymbol{I}$, i.e. $\boldsymbol{q}(n) \sim N(\boldsymbol{0}, \sigma_q^2\boldsymbol{I})$. $\boldsymbol{I}$ denotes the identity matrix of appropriate size.

This model differs from the general linear model in that $\boldsymbol{w}(n)$ is modeled as a random variable with a prior pdf. A general assumption for the estimate is the state transition model as in the well-known Kaman filter. To make a trade-off between complexity and performance, a random walk prior on $\boldsymbol{w}(n)$ is used, i.e.

$$\boldsymbol{w}(n) = \boldsymbol{w}(n-1) + \boldsymbol{\varepsilon}(n) \qquad (3)$$

where the random variable $\boldsymbol{\varepsilon}(n)$ has a normal distribution of $N(\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{I})$.

Using Bayes' theorem, the conditional probability reads

$$p(\boldsymbol{w}(n)\,|\,\boldsymbol{d}(n)) = \frac{p(\boldsymbol{d}(n)\,|\,\boldsymbol{w}(n))\,p(\boldsymbol{w}(n))}{p(\boldsymbol{d}(n))}. \qquad (4)$$

Since $\boldsymbol{q}(n) \sim N(\boldsymbol{0}, \sigma_q^2\boldsymbol{I})$, we can derive $p(\boldsymbol{d}(n)\,|\,\boldsymbol{w}(n))$ from (2) via the transformation $q(i) = d(i) - \boldsymbol{w}^T(n)\boldsymbol{x}(i)$, i.e.

$$p(\boldsymbol{d}(n)\,|\,\boldsymbol{w}(n)) = A_1 \exp\left(-\sum_{i=L}^{n}\frac{(d(i) - \boldsymbol{w}^T(n)\boldsymbol{x}(i))^2}{2\sigma_q^2}\right) \qquad (5)$$

where $A_1 = (2\pi\sigma_q^2)^{-N_d/2}$ with $N_d = n - L + 1$.

Maximizing (5) gives rise to the conventional MLE of $\boldsymbol{h}(n)$. Since MLE puts equal weights on data at different time indices, it leads to biased estimates for TV systems. To solve this problem, we maximize the weighted LS (WLS) estimate

$$p_W(\boldsymbol{d}(n)\,|\,\boldsymbol{w}(n)) = A_1 \exp\left(-\sum_{i=L}^{n}\lambda_{n-i}(n)\frac{(d(i) - \boldsymbol{w}^T(n)\boldsymbol{x}(i))^2}{2\sigma_q^2}\right) \qquad (6)$$

where $\lambda_{n-i}(n)$ is weight of the square error. $\lambda_{n-i}(n)$ decreases exponentially towards past data, which is calculated recursively

by using a FF $\lambda(n)$ satisfying $0 << \lambda(n) < 1$, i.e. $\lambda_{n-i}(n) = \lambda(n)\lambda_{n-i-1}(n-1)$ with $\lambda_0(n) = 1$. From (6), namely the weighted LF, we can recursively estimate $h(n)$, which leads to the conventional variable FF (VFF) RLS algorithm. Compared to MLE, the algorithm derived from (6) has a much better tracking capability [28]. The selection of the FF is based on the performance analysis, which will be discussed in Section IV.

Next, we can get the probability $p(w(n))$ from (3)

$$p(w(n)) = A_2 \exp\left(\frac{-1}{2\sigma_\varepsilon^2}(w(n) - w(n-1))^T(w(n) - w(n-1))\right) \quad (7)$$

where $A_2 = (2\pi\sigma_\varepsilon^2)^{-L/2}$.

Combining (4)–(7) and substituting the weighted LF (6) for $p(d(n) | w(n))$, the weighted posterior probability finds to be proportional to prior times weighted LF

$$p_W(w(n) | d(n)) \propto A_1 A_2 \exp\left(-\sum_{i=L}^n \frac{\lambda_{n-i}(n)}{2\sigma_q^2}(d(i) - w^T(n)x(i))^2 \right.$$
$$\left. - \frac{1}{2\sigma_\varepsilon^2}(w(n) - w(n-1))^T(w(n) - w(n-1))\right) \quad (8)$$

III. THE PROPOSED RWT

A. Parameter Estimation for RWT

To derive the proposed RLS-like algorithm that maximizes (8), the Fisher score vector (FSV) is obtained, which finds to be

$$s_f(w(n)) = \frac{\partial \ln p_W(w(n) | d(n))}{\partial w(n)} \quad (9)$$
$$= \sigma_q^{-2} \sum_{i=L}^n \lambda_{n-i}(n)q(i)x(i) - \sigma_\varepsilon^{-2}(w(n) - w(n-1)).$$

By setting (9) to zero [19], we find a weighted MAP estimator that recursively estimates the unknown system

$$(R_{XX}(n) + \kappa(n)I)w(n) = p_X(n) + \kappa(n)w(n-1) \quad (10)$$

where $R_{XX}(n) = \sum_{i=L}^n \lambda_{n-i}(n)x(i)x^T(i)$ is the covariance matrix, $p_X(n) = \sum_{i=L}^n \lambda_{n-i}(n)d(i)x(i)$ the cross-correlation vector, and $\kappa(n) = \sigma_q^2/\sigma_\varepsilon^2$ the regularization parameter, which can be made variable in practice. Eq. (10) is identical to the normal equation in [26] except that the FF and regularization parameter may be selected from different principles. The implementation of (10) using a QRD structure leads to the SR-QRRLS algorithm [26]. The procedure is shown in Appendix A for reference. In this paper, we have presented a new and rigorous derivation of the SR-QRRLS algorithm, which was previously designed from an intuitive perspective in one of our papers [26].

When $w(n)$ converges and hence equals to $w(n-1)$ asymptotically, the optimal solution to (10) reduces to the conventional RLS solution [25]

$$w_{opt}(n) = R_{XX}^{-1}(n)p_X(n). \quad (11)$$

It can be seen that the estimator (10) is asymptotically unbiased. A detailed MSE performance analysis of the SR-QRRLS algorithm is carried out in Section IV, from which the selection formulas of the user parameters are derived.

B. The Proposed RWT

We now formulate the nonstationarity detection problem as choosing between the null hypothesis H_0 and the alternative

hypothesis H_1. The pdf $p(d(n) | \theta)$ under H_0 and H_1, conditioned on unknown parameter θ, is the same except that the value of θ is different [7]. According to the linear system (1), the hypothesis test becomes the following parameter test

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1 \neq \theta_0 \quad (12)$$

where θ_0 is a constant parameter vector of order L; while θ_1 is a vector of order L that is different from θ_0. For the adaptive implementation, θ_0 can be assigned to an estimate of $h(n)$ at the last detected change point, say $w(n_0)$ ($n_0 < n$), and θ_1 the estimate at the current time index. Then, we can test the current estimate against $w(n_0)$.

To determine between the alternatives in (12), the Wald test is employed, the statistic of which reads

$$T_{WT} = (\theta_1 - \theta_0)^T B^{-1}(\theta_1)(\theta_1 - \theta_0) \quad (13)$$

where $B(\theta_1)$ is the Cramer-Rao bound (CRB) at θ_1 [7].

Based on the Wald statistic (13) and the weighted MAP, we can derive a testing method. Since this method employs results from the recursive estimate, we call it the recursive Wald test.

First, we can use (10) to estimate the parameter under H_1. Then, we need to determine CRB. For unbiased estimators, CRB equals to the inverse of Fisher information matrix (FIM). Using (8), the (l,m)-th entry of FIM is

$$[I_f(w(n))]_{lm} = -E\left[\frac{\partial^2 \ln p_W(w(n) | d(n))}{\partial w_l \partial w_m}\right]$$
$$= E\left[\sigma_q^{-2} \sum_{i=L}^n \lambda_{n-i}(n)x(i-l)x(i-m) + \sigma_\varepsilon^{-2}[I]_{lm}\right] \quad (14)$$
$$= \sigma_q^{-2} r_x(l-m)\sum_{i=L}^n \lambda_{n-i}(n) + \sigma_\varepsilon^{-2}[I]_{lm}$$

where $r_x(k)$ is the autocorrelation of $\{x(n)\}$ at lag k. Consequently, we have $I_f(w(n)) = \sigma_q^{-2}R_{XX}(n) + \sigma_\varepsilon^{-2}I$.

In the following, we discuss how to calculate the Wald test (13) recursively. To this end, we rewrite FSV at $w(n_0)$ as

$$s_f(w(n_0)) = \frac{1}{\sigma_q^2}\sum_{i=L}^n \lambda_{n-i}(n)q_0(i)x(i) - \frac{1}{\sigma_\varepsilon^2}(w(n_0) - w(n-1))$$
$$= \sigma_q^{-2}[p_X(n) + \kappa(n)w(n-1) - (R_{XX}(n) + \kappa(n)I)w(n_0)]$$
$$= \sigma_q^{-2}(R_{XX}(n) + \kappa(n)I)(w(n) - w(n_0)) \quad (15)$$
$$= I_f(w(n))(w(n) - w(n_0))$$

where $q_0(i) = d(i) - w^T(n_0)x(i)$, and (10) has been used to simplify the expression.

Replacing θ_1 and θ_0, respectively, with $w(n)$ and $w(n_0)$, and substituting (14)(15) into (13) gives the RWT formula

$$T_{WT} = (w(n) - w(n_0))^T s(w(n_0)) \quad (16)$$

where the weighted MAP estimate $w(n)$ is updated at each time index, FSV at $w(n_0)$ is updated recursively as follows

$$s_f(w(n_0)) = \sigma_q^{-2}\sum_{i=L}^n \lambda_{n-i}(n)q_0(i)x(i) - \kappa(n)(w(n_0) - w(n-1)) \quad,$$

and the noise variance σ_q^2 can be known "*a priori*" or approximated from the residue of the algorithm using a large FF [29]. $w(n)$ is then tested against the estimate at the last detected change point $w(n_0)$. We should reject the hypothesis

of stationarity if $T_{WT} > \gamma$, where γ is chosen to maintain a constant false alarm rate (CFAR) [7].

Under suitable technical conditions, such as MLE and MAP, likelihood ratio statistics take on a chi-squared distribution as the sample size grows large. Therefore, the developed Wald detector (16) is asymptotically Chi-squared distributed under H_0 with Lth degree of freedom [7]:

$$T_{WT} \sim \chi_L^2(0). \tag{17}$$

Under H_1, it has a noncentral Chi-squared distribution or

$$T_{WT} \sim \chi_L^2(\phi) \tag{18}$$

where the non-centrality parameter is given by [18]

$$\phi \approx \big(\boldsymbol{w}(n) - \boldsymbol{w}(n_0)\big)^T \boldsymbol{I}_f(\boldsymbol{w}(n))\big(\boldsymbol{w}(n) - \boldsymbol{w}(n_0)\big). \tag{19}$$

The statistical property of the detector indicates that the principle behind the hypothesis test is to assess whether the distribution of the detector follows the known distribution [7].

C. The SR-RBIC Method

One issue that has not been addressed is the question of what order to use in the adaptive detector. For example, if the order of the estimator is different (larger or smaller) from the true model, a stationary sequence could be detected as nonstationary sequence due to the increased estimation residue (variance or bias). To deal with this problem, a recursive order selection method employing Bayesian model selection criterion and the SR-QRRLS algorithm, namely SR-RBIC, is proposed. Using MAP rules in model selection was not newly proposed [30]. However, the work here focuses on adaptation procedures only, which can be extended to other state-of-the-art criteria [31] for the corresponding adaptive versions.

In Bayesian framework, given data $\{d(n)\}$, the posterior probability $p(M_i \mid \boldsymbol{d}(n))$ of a model M_i is

$$p(M_i \mid \boldsymbol{d}(n)) = \frac{p(\boldsymbol{d}(n) \mid M_i) p(M_i)}{p(\boldsymbol{d}(n))} \tag{20}$$

where $p(M_i)$ is a prior probability for model M_i ($i = 1, 2, \ldots$). The probability of data conditioned on model M_i can be computed by the following integral

$$p(\boldsymbol{d} \mid M_i) = \int p(\boldsymbol{d} \mid \boldsymbol{\theta}, M_i) p(\boldsymbol{\theta} \mid M_i) d\boldsymbol{\theta} \tag{21}$$

where we have omitted the time index for simplicity. Since a posterior is proportional to prior multiplied by likelihood, we have $p(\boldsymbol{\theta} \mid \boldsymbol{d}, M_i) \propto p(\boldsymbol{d} \mid \boldsymbol{\theta}, M_i) p(\boldsymbol{\theta} \mid M_i)$ and hence get

$$p(\boldsymbol{d} \mid M_i) = \int p(\boldsymbol{\theta} \mid \boldsymbol{d}, M_i) d\boldsymbol{\theta}. \tag{22}$$

It is assumed that the posterior has a very sharp pick at the MAP estimate $\hat{\boldsymbol{\theta}}$ such that the quadratic Taylor expansion is sufficient. By letting $p(\boldsymbol{\theta} \mid \boldsymbol{d}, M_i) = \exp(g(\boldsymbol{\theta}))$, we get $g(\boldsymbol{\theta}) = g(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{I}_f^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. Then, the expansion finds

$$p(\boldsymbol{\theta} \mid \boldsymbol{d}, M_i) = p(\hat{\boldsymbol{\theta}} \mid \boldsymbol{d}, M_i) \exp\!\big(-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{I}_f^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\big). \tag{23}$$

Consequently, the marginal likelihood is found as

$$p(\boldsymbol{d} \mid M_i) \cong p(\hat{\boldsymbol{\theta}} \mid \boldsymbol{d}, M_i) \int \exp\!\big(-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{I}_f^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\big) d\boldsymbol{\theta} \tag{24}$$

$$= p(\hat{\boldsymbol{\theta}} \mid \boldsymbol{d}, M_i)(2\pi)^{L/2} \big|\boldsymbol{I}_f(\hat{\boldsymbol{\theta}})\big|^{-1/2}.$$

Substituting $p_W(\boldsymbol{w}(n) \mid \boldsymbol{d}(n))$ in (8) for $p(\hat{\boldsymbol{\theta}} \mid \boldsymbol{d}, M_i)$ and $\boldsymbol{w}(n)$ for $\hat{\boldsymbol{\theta}}$, we get the weighted posterior probability

$$p_W(\boldsymbol{d} \mid M_i) = A \exp\!\Big(- \sum_{i=L}^{n} \tfrac{\lambda_{n-i}(n)}{2\sigma_q^2} q^2(i) - \tfrac{1}{2\sigma_\varepsilon^2}[\boldsymbol{w}(n) - \boldsymbol{w}(n-1)]^T[\boldsymbol{w}(n) - \boldsymbol{w}(n-1)]\Big) \cdot (1 - \lambda(n))^{L/2} \big|\sigma_q^{-2}(\boldsymbol{R}_{xx}(n) + \xi(n)\boldsymbol{I})\big|^{-1/2} \tag{25}$$

where $A = A_1 A_2 (2\pi)^{L/2}$, and we define $\xi(n) = (1 - \lambda(n))\kappa(n)$ by using the relationship $\boldsymbol{R}_{xx}(n) \overset{n \to \infty}{=} \frac{1}{1-\lambda(n)}\boldsymbol{R}_{xx}$ with $\boldsymbol{R}_{xx} = E[\boldsymbol{x}(n)\boldsymbol{x}^T(n)]$ the input covariance matrix. Note, the above relationship is based on the fact that the FF does not change its value frequently [28]. To proceed further, the non-informative prior $p(M_i) = 1$ is used. Hence, we choose the optimal model M_i by minimizing the following SR-RBIC selector

$$
\begin{aligned}
T_M &= -2\ln(p_W(\boldsymbol{d} \mid M_i)) \\
&\overset{n \to \infty}{=} \sum_{i=L}^{n} \tfrac{1}{\sigma_q^2}\lambda_{n-i}(n)q^2(i) - L\ln(1 - \lambda(n)) + \ln\!\big(\sigma_q^{-2}\big|\hat{\boldsymbol{R}}_{xx}\big|\big) \\
&\quad + \kappa(n)\sigma_q^{-2}(\boldsymbol{w}(n) - \boldsymbol{w}(n-1))^T(\boldsymbol{w}(n) - \boldsymbol{w}(n-1)) + C
\end{aligned} \tag{26}
$$

where $\hat{\boldsymbol{R}}_{xx} = \boldsymbol{R}_{xx} + \xi(n)\boldsymbol{I}$ and C is a constant. It can be seen that the first two terms contribute as the adaptive version of the conventional BIC, where the number of sample changes with the window size. The third term can be viewed as the signal-to-noise ratio (SNR) while the last term comes from the regularization which would reject models with large variations. Compared to the conventional Akaike information criterion [32] and Bayesian information criterion [27], Eq. (26) considers the contribution of noises [30][31].

IV. USER PARAMETER SELECTION

Although developing advanced RLS algorithms is not the focus of this paper, adaptive estimation is an important and necessary step for the proposed detection method. Therefore, in this section, we analyze MSE of the SR-QRRLS algorithm based on a TV linear system (1), the coefficient of which is modeled by a local polynomial. From this analysis, we derive the locally optimal FF and regularization parameter.

Recall the normal equation (10). After some simple algebras, it can be written in a matrix notation for concise presentation

$$\tilde{\boldsymbol{R}}_{XX}(n)\boldsymbol{w}(n) = \boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{d}(n) + \kappa(n)\boldsymbol{w}(n-1) \tag{27}$$

where $\boldsymbol{\Lambda}(n) = diag([1, \lambda_1(n), \ldots, \lambda_{n-L}(n)])$ is a diagonal matrix, $\tilde{\boldsymbol{R}}_{XX}(n) = \boldsymbol{R}_{XX}(n) + \kappa(n)\boldsymbol{I}$ is the regularized covariance matrix with $\boldsymbol{R}_{XX}(n) = \boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{X}(n)$.

We assume that the impulse response vector of the channel is continuous and it admits a first order polynomial expansion at time t_n as described in [28] for the classical RLS algorithm:

$$\boldsymbol{h}(t_m) = \boldsymbol{h}(t_n) + \tfrac{1}{1!}\boldsymbol{h}^{(1)}(t_n)(t_m - t_n) + \boldsymbol{r}(t_m - t_n) \tag{28}$$

where t_m belongs to an appropriately close neighborhood of

t_n, and $\boldsymbol{r}(t_m - t_n)$ is the remainder of higher order terms. Both $\boldsymbol{h}^{(1)}(t_n)$ and $\boldsymbol{r}(t_m - t_n)$ can be considered as random vectors of zero mean and the latter is of order $o(t_m - t_n)$.

Using (28) and (1), and letting $t_n = nT_s$, we have

$$\boldsymbol{d}(n) = \boldsymbol{X}(n)\boldsymbol{h}(n) + \boldsymbol{D}_\tau(n)\boldsymbol{X}(n)\boldsymbol{h}^{(1)}(n) + \boldsymbol{\eta}(n) + \boldsymbol{\upsilon}(n) \quad (29)$$

where T_s is the sampling period, $\boldsymbol{D}_\tau(n) = diag(\boldsymbol{\tau}(n))$ with $\boldsymbol{\tau}(n) = [0,-1,...,-(n-L)]^T$, $\boldsymbol{\eta}(n) = [\eta(n),\eta(n-1),...\eta(L)]^T$ and $\boldsymbol{\upsilon}(n) = [\upsilon(n),\upsilon(n-1),...,\upsilon(L)]^T$ are, respectively, the noise and residue vectors with $\upsilon(n-m) = \boldsymbol{x}^T(n-m)\boldsymbol{r}(-m)$.

Substituting (29) into (27), we can obtain the RLS solution

$$\begin{aligned}
\boldsymbol{w}(n) = {} & \boldsymbol{h}(n) + \boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{R}_\tau(n)\boldsymbol{h}^{(1)}(n) \\
& + \boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\big(\boldsymbol{\eta}(n) + \boldsymbol{\upsilon}(n)\big) \\
& - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n)\Big[\boldsymbol{h}(n) - \boldsymbol{w}(n-1) + \boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{R}_\tau(n)\boldsymbol{h}^{(1)}(n) \\
& + \boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\big(\boldsymbol{\eta}(n) + \boldsymbol{\upsilon}(n)\big)\Big]
\end{aligned} \quad (30)$$

where $\boldsymbol{R}_\tau(n) = \boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\boldsymbol{D}_\tau(n)\boldsymbol{X}(n) = -\frac{1}{(1-\lambda(n))^2}\boldsymbol{R}_{xx}$ according to Appendix B and we used the Woodbury formula [33] to get $(\boldsymbol{R}_{XX}(n) + \kappa(n)\boldsymbol{I})^{-1} = \boldsymbol{R}_{XX}^{-1}(n) - \boldsymbol{R}_{XX}^{-1}(n)\big(\frac{1}{\kappa(n)}\boldsymbol{I} + \boldsymbol{R}_{XX}^{-1}(n)\big)^{-1}\boldsymbol{R}_{XX}^{-1}(n)$.

In the following, we consider the mean square deviation (MSD) of $\boldsymbol{w}(n)$ from $\boldsymbol{h}(n)$. First, the difference equation for the bias vector $\boldsymbol{b}(n) = E[\boldsymbol{w}(n)] - \boldsymbol{h}(n)$ can be obtained from (30) as

$$\begin{aligned}
\boldsymbol{b}(n) = {} & \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n)\boldsymbol{b}(n-1) - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n)\overline{\boldsymbol{h}^{(1)}}(n) \\
& + (\boldsymbol{I} - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n))\boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{R}_\tau(n)\overline{\boldsymbol{h}^{(1)}}(n)
\end{aligned} \quad (31)$$

where $\overline{\boldsymbol{h}^{(1)}}(n) = E[\boldsymbol{h}^{(1)}(n)] = E[\boldsymbol{h}(n) - \boldsymbol{h}(n-1)]$. The remainder $\boldsymbol{r}(m-n)$ is independent of the input $\boldsymbol{X}(n)$ and the correlation between $\boldsymbol{X}(n)$ and $\boldsymbol{\upsilon}(n)$ is negligible [28][34]. From Appendix C, we can obtain the bias of the estimate

$$J_b = Tr(\boldsymbol{\Xi}_b(\infty)) \overset{n\to\infty}{\approx} \frac{\overline{\sigma}_h^2(n)}{(1-\lambda(n))^2}Tr(\boldsymbol{I}_\kappa^2) + \xi^2(n)\overline{\sigma}_h^2(n)Tr(\widetilde{\boldsymbol{R}}_{xx}^{-2}) \quad (32)$$

where $\boldsymbol{\Xi}_b(n) = E[\boldsymbol{b}(n)\boldsymbol{b}^T(n)]$, the matrix $\overline{\boldsymbol{R}}_h = \overline{\boldsymbol{h}^{(1)}}(n)[\overline{\boldsymbol{h}^{(1)}}(n)]^T$ can reduce to $\overline{\sigma}_h^2(n)\boldsymbol{I}$, and $\boldsymbol{I}_\kappa = \boldsymbol{I} - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n) = \boldsymbol{I} - \xi(n)\widetilde{\boldsymbol{R}}_{xx}^{-1}$ with $\widetilde{\boldsymbol{R}}_{xx} = \boldsymbol{R}_{xx} + \xi(n)\boldsymbol{I}$.

Next, the difference equation for the variance vector $\boldsymbol{v}(n) = \boldsymbol{w}(n) - E[\boldsymbol{w}(n)]$ can be obtained from (30) as

$$\begin{aligned}
\boldsymbol{v}(n) = {} & \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n)\boldsymbol{v}(n-1) \\
& + (\boldsymbol{I} - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n))\boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{R}_\tau(n)\delta\boldsymbol{h}^{(1)}(n) \\
& + (\boldsymbol{I} - \kappa(n)\widetilde{\boldsymbol{R}}_{XX}^{-1}(n))\boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{X}^T(n)\boldsymbol{\Lambda}(n)\big(\boldsymbol{\eta}(n) + \boldsymbol{\upsilon}(n)\big)
\end{aligned} \quad (33)$$

where $\delta\boldsymbol{h}^{(1)}(n) = \boldsymbol{h}^{(1)}(n) - \overline{\boldsymbol{h}^{(1)}}(n)$. From Appendix C, we have

$$J_v = Tr(\boldsymbol{\Xi}_v(\infty)) \overset{n\to\infty}{\approx} \frac{1-\lambda(n)}{1+\lambda(n)}\sigma_\Sigma^2 Tr\big(\boldsymbol{I}_\kappa^2\boldsymbol{R}_{xx}^{-1}\big) + \frac{\sigma_{\delta h}^2(n)}{(1-\lambda(n))^2}Tr(\boldsymbol{I}_\kappa^2) \quad (34)$$

where $\boldsymbol{\Xi}_v(n) = E[\boldsymbol{v}(n)\boldsymbol{v}^T(n)]$, $\sigma_\Sigma^2 = \sigma_\eta^2 + \sigma_\upsilon^2$, $\sigma_\eta^2 = E[\eta^2(n)]$, $\sigma_\upsilon^2 = E[\upsilon^2(n)]$, and $\boldsymbol{R}_{\delta h}(n) = \delta\boldsymbol{h}^{(1)}(n)[\delta\boldsymbol{h}^{(1)}(n)]^T = \sigma_{\delta h}^2(n)\boldsymbol{I}$.

Consequently, the mean square deviation (MSD) can be obtained from (32) and (34) as

$$\begin{aligned}
J_{MSD}(n) = {} & J_b + J_v \\
= {} & \frac{\sigma_h^2(n)}{(1-\lambda(n))^2}Tr(\boldsymbol{I}_\kappa^2) + \frac{1-\lambda(n)}{1+\lambda(n)}\sigma_\Sigma^2 Tr(\boldsymbol{I}_\kappa^2\boldsymbol{R}_{xx}^{-1}) \\
& + \xi^2(n)\overline{\sigma}_h^2(n)Tr(\widetilde{\boldsymbol{R}}_{xx}^{-2})
\end{aligned} \quad (35)$$

where $\sigma_h^2(n) = \overline{\sigma}_h^2(n) + \sigma_{\delta h}^2(n)$.

We now derive the locally optimal FF. To see the influence of $\lambda(n)$ on $J_{MSD}(n)$, the following assumptions are made. In the bias term, $\xi^2(n)\overline{\sigma}_h^2(n)Tr(\widetilde{\boldsymbol{R}}_{xx}^{-2}) << \frac{1}{(1-\lambda(n))^2}\overline{\sigma}_h^2 Tr(\boldsymbol{I}_\kappa^2)$ and hence the former can be ignored. Meanwhile, $\boldsymbol{I}_\kappa$ is very close to the identity matrix for small to medium regularization. Using these two approximations and taking the derivative of (35) with respect to the FF gives

$$\frac{\partial J_{MSD}(n)}{\partial \lambda(n)} \approx \frac{2\sigma_h^2 L}{(1-\lambda(n))^3} - \frac{2\sigma_\Sigma^2}{(1+\lambda(n))^2}Tr(\boldsymbol{R}_{xx}^{-1}). \quad (36)$$

Letting (36) be zero, we have

$$\sigma_\Sigma^2 Tr(\boldsymbol{R}_{xx}^{-1})/(1+\lambda(n))^2 = \sigma_h^2(n)L/(1-\lambda(n))^3$$

To proceed further, we let $\mu = (1+\lambda(n))/(1-\lambda(n))$ so as to have $\mu^2(\mu+1) = 2\sigma_\Sigma^2 Tr(\boldsymbol{R}_{xx}^{-1})/(L\sigma_h^2(n))$. Since $\mu >> 1$ [28], we have $\mu^2(\mu+1) \approx \mu^3$ and hence $\mu = [2\sigma_\Sigma^2 Tr(\boldsymbol{R}_{xx}^{-1})/(L\sigma_h^2(n))]^{\frac{1}{3}}$. Then, the locally optimal FF can be obtained

$$\lambda_{opt}(n) = (\mu-1)/(\mu+1), \text{ if } \lambda_{opt}(n) > 0 \quad (37)$$

This result is identical to that in [28], where no regularization technique has been employed. In practice, the noise variances can be obtained from prior information or calculated from the estimation error $q(n)$. The details can be found in [28].

To derive the optimal regularization parameter, we take the derivative of (35) with respect to ξ (the time index of which has been omitted to have a concise expression) and have

$$\begin{aligned}
\frac{\partial J_{MSD}(n)}{\partial \xi} = {} & \frac{\sigma_h^2(n)}{(1-\lambda(n))^2}\sum_{i=1}^L \frac{-2\sigma_{xi}^4}{(\sigma_{xi}^2+\xi)^3} + \frac{1-\lambda(n)}{1+\lambda(n)}\sigma_\Sigma^2\sum_{i=1}^L \frac{-2\sigma_{xi}^2}{(\sigma_{xi}^2+\xi)^3} \\
& + \overline{\sigma}_h^2(n)\sum_{i=1}^L\left[\frac{2\xi}{(\sigma_{xi}^2+\xi)^2} - \frac{2\xi^2}{(\sigma_{xi}^2+\xi)^3}\right]
\end{aligned} \quad (38)$$

where σ_{xi}^2 is the ith eigenvalue of $\boldsymbol{R}_{xx}$ so that the ith eigenvalue of $\boldsymbol{I}_\kappa$ reads $\sigma_{xi}^2/(\sigma_{xi}^2+\xi)$. To derive a formula for practical use, we assume that the input is white Gaussian distributed with variance σ_x^2 and let (38) equal to 0:

$$\overline{\sigma}_h^2(n)\xi = \frac{\sigma_h^2(n)}{(1-\lambda(n))^2}\frac{\sigma_x^4}{\sigma_x^2+\xi} + \frac{(1-\lambda(n))\sigma_\Sigma^2}{1+\lambda(n)}\frac{\sigma_x^2}{\sigma_x^2+\xi} + \frac{\xi^2\overline{\sigma}_h^2(n)}{\sigma_x^2+\xi}.$$

Since $\xi << \sigma_x^2$ and the last term is far smaller than the other terms, the following formula provides a good approximation for the selection of regularization parameter

$$\xi_{opt}(n) = \frac{\sigma_x^2}{\overline{\sigma}_h^2(n)}\left(\frac{1-\lambda(n)}{1+\lambda(n)}\frac{\sigma_\Sigma^2}{\sigma_x^2} + \frac{\sigma_h^2(n)}{(1-\lambda(n))^2}\right). \quad (39)$$

The resulting RLS algorithm using (37) and (39), respectively, as the locally optimal variable FF (LVFF) and regularization parameter is called the SR-LVFF-QRRLS algorithm.

V. SIMULATION RESULTS

First, we examine the performance of the estimator in system identification problems. The effect of FF and regularization parameter on estimation rate and accuracy is studied for both random walk and sudden change models. Next, the proposed RWT is applied to TVAR model for sequential detection. Then, SR-RBIC is tested. Finally, a sequential detection method with automatic order selection is evaluated. Unless specified, all results are averaged over 100 Monte-Carlo simulations.

For the sequential detection methods, the testing samples are summations of sinusoids or real speech signals, which are represented by a TVAR model and exhibit a "slowly varying" type of nonstationarity as opposed to a sudden change for which many efficient tests exist [12]. Such signal is of great concern in practical applications.

A. Evaluation of the Estimator

1) Convergence performance comparison

In this experiment, a TV linear system (1) is considered. The system to be identified is a random walk model of length $L = 5$: $\mathbf{h}(n+1) = \mathbf{h}(n) + \boldsymbol{\delta}(n)$, where $\boldsymbol{\delta}(n)$ is the white Gaussian random vector with a covariance matrix of $5 \times 10^{-5} \mathbf{I}$. The initial value of the channel is set to $\mathbf{h}_0 = [-1,1,1,1,-1]$, which changes to $\mathbf{h}_1 = [-1,-1,1,1,-1]$ at the 3000th sample. The length of the adaptive filter is set to $L = 5$. SNR is set to 10 and 20 dB, which represent a common environment in real applications. The input signal is generated from a first-order AR process: $x(n+1) = 0.9x(n) + g(n)$, where $g(n)$ is Gaussian process of a zero-mean, and the input power is normalized.

The performance of the conventional RLS algorithm, the GVFF-RLS algorithm [29], the LVFF-QRRLS algorithm [28] and the proposed SR-LVFF-QRRLS algorithm are tested. The FF for RLS is set to 0.999. Parameters for GVFF-RLS are set to $\alpha = 0.3$, $\beta = 0.99$, $\mu = 0.04$, as suggested in [29], and ω^* is chosen as 0.999 to achieve a similar MSD with the RLS algorithm. The simulation results are shown in Fig. 1. It can be seen that the LVFF-based algorithms have a much faster tracking speed in each case. The SR-LVFF-QRRLS algorithm could converge to a much lower steady-state MSD at a comparable or even faster speed, and the improvement is more significant when the noise level is higher due to the unbiased properties of the state regularization.

2) Effect of FF and regularization parameter

In this experiment, the effect of the FF and the regularization parameter on the performance of the estimator is examined. The system to be identified is identical to that in the previous experiment except it does not have a sudden change. The same inputs as in the previous experiment are used.

First, we evaluate the performance of SR-QRRLS algorithm at different FFs, given the regularization parameter calculated from (39). Different FFs $\lambda = \lambda_S$, λ_{opt} and λ_L are used, where $\lambda_S = 1 - 1/(0.1 L_{opt})$, $\lambda_L = 1 - 1/(10 L_{opt})$, and $L_{opt} = 1/(1 - \lambda_{opt})$ is the optimal window size. Applying these parameters to SR-QRRLS, the corresponding MSD curves at SNR = 0 and 20 dB are shown in Fig. 2. It can be seen that, SR-QRRLS with λ_{opt} generally converges faster to a lower steady-state MSD.
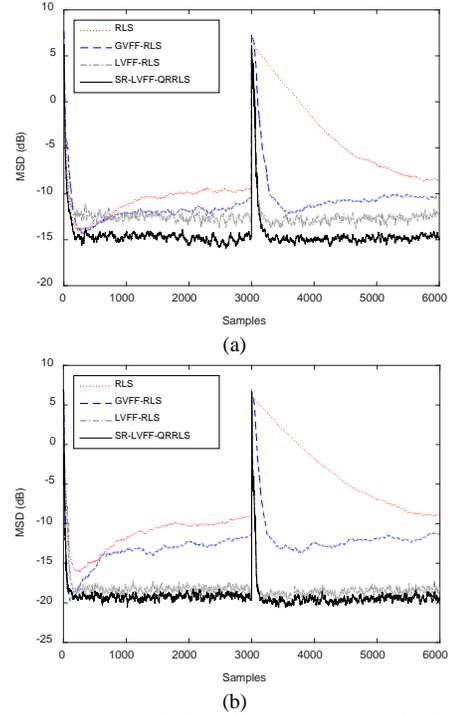

Fig. 1. Learning curves of MSD for sudden-change channels with the 1st order AR input at (a) SNR = 10 dB (b) SNR = 20 dB. $\sigma_h^2 = 0.00005$, $L = 5$.
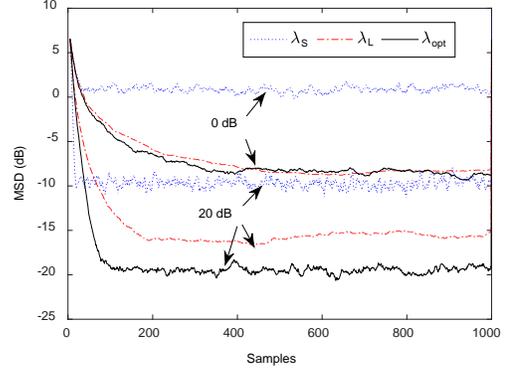

Fig. 2. Learning curves of MSD for the SR-QRRLS algorithm with the 1st order AR input at SNR = 0dB and 20dB. $L = 5$. $\xi = \xi_{opt}$. $\lambda = \lambda_S$, λ_{opt} and λ_L.

The algorithm with λ_S may have a slightly faster convergence initially, but it achieves much larger MSD due to the increased variance. The MSD curves at SNR = 10 dB are somewhere in between and the curve with λ_{opt} also has best performance. For presentation use, only curves at SNR = 0 and 20 dB are shown. It can be seen from the results that the accuracy of the proposed selection formula is within one order from its true value, which provides a good reference for practical use.

Secondly, we study the effect of the regularization parameter on the performance of SR-QRRLS. The theoretical and simulated steady-state MSD vs. regularization parameter curves are shown in Fig. 3. It can be seen that the simulated and theoretical results agree well with each other. It also illustrates that the optimal regularization parameter formulated in (39) gradually decreases as the SNR increases from 0 dB to 20 dB, which is mainly because the variance term in J_{MSD} decreases as the noise becomes smaller so that less regularization is needed to combat the measurement error. It also shows that ξ_{opt} is slightly overestimated for lower SNR. However, it still
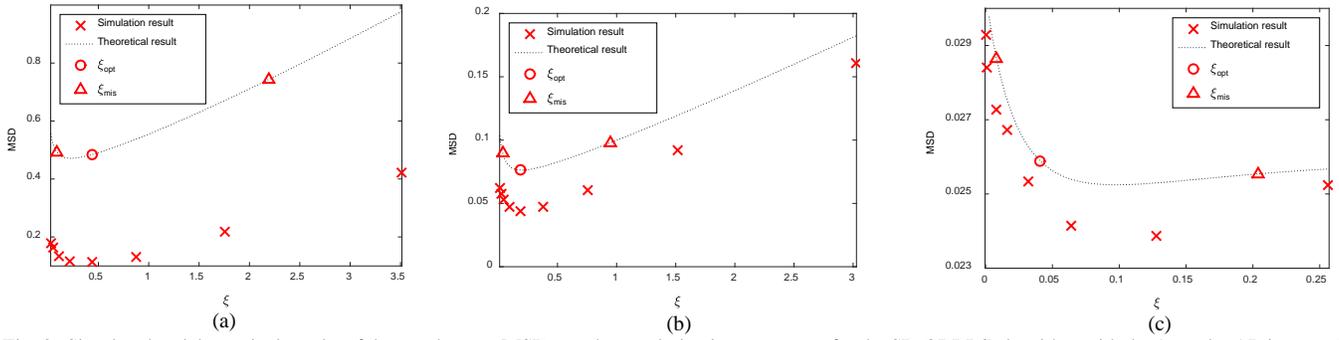
Fig. 3. Simulated and theoretical results of the steady-state MSDs vs. the regularization parameter for the SR-QRRLS algorithm with the 1st order AR input at SNR = (a) 0 dB (b) 10 dB, and (c) 20 dB. L = 5. λ=λ_{opt}. ξ = 0.1ξ_{opt}, ξ_{opt} and 5ξ_{opt}.

provides a good reference for selection of the regularization parameter in practice. Next, we examine in more detail the performance of (39) with mismatches, i.e. 0.1 ξ_{opt} and 5 ξ_{opt} are examined. The resulting steady-state MSDs are marked by 'Δ' in Fig. 3. It suggests that formula (39) is tolerant to mismatches if the prior information, such as noise variances, is not exactly known in real applications.

B. Evaluation of the Detector

1) Evaluation of the false alarm (FA) rate

In this example, the signal under study is a single sinusoid. A TVAR model is employed to represent the process, which reads

$$x_a(n) = \sum_{i=1}^{L} a_i(n)x_a(n-i) + \eta_a(n) = \boldsymbol{a}^T(n)\boldsymbol{x}_a(n) + \eta_a(n) \quad (40)$$

(a)

(b)
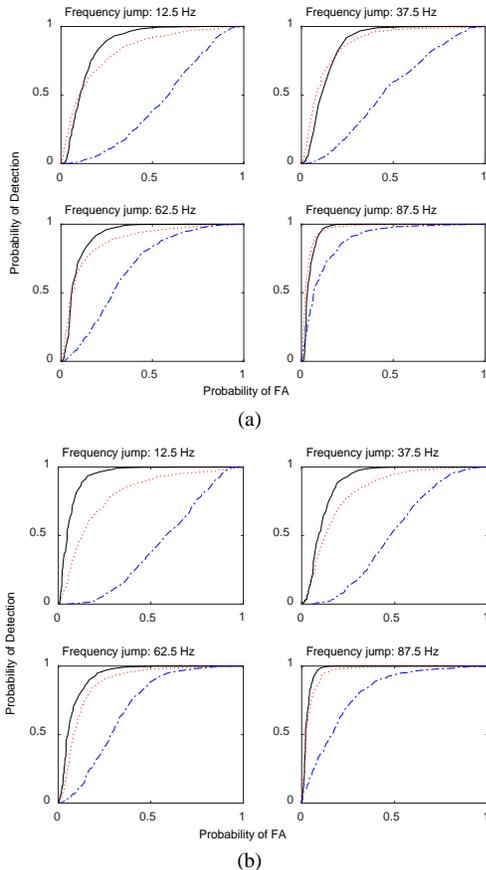
Fig. 4. Detection performance of the proposed RWT (solid line) and GLRT with q = 4 (dot line) and q = 2 (dot dash line). Fs = 4 kHz. (a) Time duration is 4 ms; (b) Time duration is 2 ms. TVAR(2) is used.

where $\boldsymbol{a}(n) = [a_1(n),...,a_L(n)]^T$ is the TV parameter vector of order L, $\boldsymbol{x}_a(n) = [x_a(n-1), x_a(n-2),..., x_a(n-L)]^T$, and $\eta_a(n)$ is a white Gaussian process for excitation. Such a model is simply denoted as TVAR(L) in the sequel. The developed algorithms can be easily extended to the TVAR model by replacing $d(n)$ and $\boldsymbol{x}(n)$ in (3), respectively, by $x_a(n)$ and $\boldsymbol{x}_a(n)$. The resulting RLS-based algorithm for the prediction of TV parameters is closely related to the problem of linear prediction in a variety of applications [35]–[37], which is popular with good computational efficiency and nearly optimum estimation accuracy [15].

The sinusoids under study have a time duration of T and changes its frequency by δ Hz at T/2. The sampling frequency is assumed to be 4 kHz. The proposed RWT and GLRT method [5] are used to detect this change. For both algorithms under testing, the TVAR(2) model is applied since two TVAR parameters correspond to one frequency component. For the TV linear prediction algorithm in GLRT, q = 2, and 4 Legendre polynomials are used. The signal lengths considered are 2 ms and 4 ms. The signal changes its frequency by δ = 12.5Hz, 37.5Hz, 62.5Hz and 87.5Hz. The SNR is set to 20 dB. To calculate the probability of detection and FA, 1000 trial simulations have been performed for each setting. As can be seen from Fig. 4, the detection performance for both methods improves when δ is increased while T is fixed. If the data records are sufficient (as shown in Fig. 4(a)), the GLRT method with appropriate power order (q = 4) has a comparable detection performance with the proposed RWT method at the cost of increased computational complexity. At the presence of insufficient data, however, the proposed detector usually outperforms the GLRT due to the regularization technique employed as shown in Fig. 4(b).

2) The sequential nonstationarity detection method

In this section, we give two examples showing how to detect nonstationarity using the proposed RWT. Different from the retrospective property of the batch-based methods, the method here can process the observations sequentially over time. So we call it a sequential nonstationarity detection method, which is summarized as follows:

Method I: Sequential Nonstationarity Detection
1) Initialization:
• set γ via (17) given a certain CFAR;   • set n = 1, n_0 = 0, and $\boldsymbol{w}_0$ = $\boldsymbol{0}$;

2) Detector calculation
- estimate the model parameter at the current time instant n, $\boldsymbol{w}$, using SR-LVFF-QRRLS;
- calculate the corresponding T_{WT} via (16);
3) Detection
- if $T_{WT} < \gamma$ (no nonstationarity is detected), go to step 4);
- else (nonstationarity detected), mark this time instant as $n_0 = n$ and set $\boldsymbol{w}_0 = \boldsymbol{w}$;
4) Set $n = n + 1$ and return to 2) for the next data sample.

In this method, $\boldsymbol{w}$ is the parameter estimate at the current time instant n whereas $\boldsymbol{w}_0$ is the estimate at the last detected change point, say at n_0, and (16) will be used to test nonstationarity using the current estimate. If a change is detected, n_0 will be set to the current time instant n and we can repeat the above process, otherwise, it will proceed to the next time instant. The proposed method is compared with the conventional GLRT method [5]
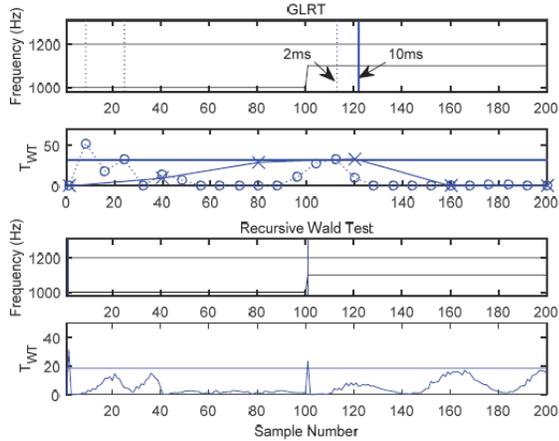
Fig. 5. Detection performance of the proposed RWT-base method (Method I) and the GLRT-based sequential detection (Algorithm 1) in [5] for synthetic data. Fs = 4 kHz. For GLRT, the batch sizes used are 2 ms (dotted lines) and 10 ms (solid lines).
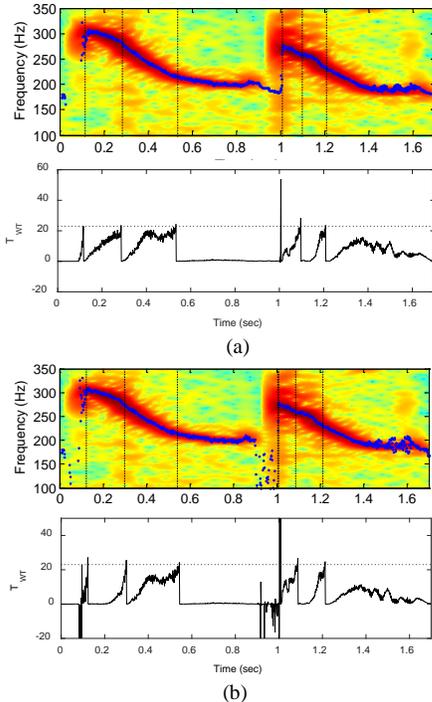

Fig. 6. Detection performance of the proposed RWT method (Method I) using (a) SR-LVFF-QRRLS, and (b) LVFF-QRRLS for the real speech.

(Algorithm 1) without overlapping so that the complexity of the two methods is comparable. In this experiment and the rest of the paper, all results are not averaged.

In the first experiment, the signal under study is shown in Fig. 5, which contains two sinusoids, one of which changes its frequency at the 100th sample. A TVAR(4) model is used for both algorithms. Other settings are identical to those in the previous experiment. For the GLRT-based method ($q = 4$), two different rectangular windows are used, i.e. $M = 8$ (2 ms) and $M = 40$ (10 ms). The CFAR used is 1%. We observe in Fig. 5 that for the GLRT-based method, a shorter batch size (dotted lines in the first two subplots) leads to a more prompt response to signal changes, but may have more FAs due to insufficient data samples (see Fig. 4). On the other hand, a longer batch size (solid lines in the first two subplots) can suppress large variances during detection, but leads to longer detection latency. The proposed method can provide a detection result with much less detection latency as well as less FAs due to its recursive property and the state regularization employed.

Next, the detection method is applied to a segment of real speech signal. The real data used are the waveform of a vowel [a] (as in "father") followed by [ai] (as in "life") [38]. It was downsampled to 1 kHz in order to focus on lower frequency formants [5]. According to the rule-of-thumb that "2 coefficients per kHz", a TVAR(2) model is applied. The TV center frequencies are extracted from $\boldsymbol{a}(n)$ using the roots z_i of the equation $z^L - a_1(n)z^{L-1} \dots - a_L(n) = 0$. We choose the frequency estimate as the angles of these roots. For real-valued signals, only roots in the upper half of the complex plane are selected. Both LVFF-QRRLS and SR-LVFF-QRRLS algorithms are used in the detection method and the results are shown in Fig. 6. Since the center frequency mainly ranges from 100 Hz to 350 Hz, a spectrum covering this range is presented. From the T_{WT} curves and estimated frequencies, it can be seen that the regularization helps to stablize the estimation and detection process.

C. Evaluation of the Model Selector

In this section, we illustrate the performance of the proposed SR-RBIC and examine the effect of background noises on it.

In the first example, the true linear regression model is $y(n) = 1 + 0.5x(n) + \eta(n)$, where the input $\{x(n)\}$ and additive noise $\{\eta(n)\}$ are both random Gaussian sequences. The SNR is set to 20 dB. We fit the following models to the values of $\{y(n)\}$ by using SR-LVFF-QRRLS: (1) the constant model $y_1(n) = \beta_0$; (2) the straight line model $y_2(n) = \beta_0 + \beta_1 x(n)$; and (3) the second order model $y_3(n) = \beta_0 + \beta_1 x(n) + \beta_2 x(n-1)$.

The 1002th–1016th samples fitted by the three models are shown in Fig. 7(a). It seems that the straight line model y_2 and the second order model y_3 are almost overlapping and they are quite close to the true values, while the constant model y_1 has significant deviation. These 15 samples are randomly chosen. The results for other samples show similar performance. Then, using the SR-RBIC selector (26), we proceed with the model selection from the three candidates. From Fig. 7(b), it can be seen that y_2 should be selected as the best model since its T_M curve is stable and lower than others. The second best model is y_1, which may also be correct since regularization could help to suppress the variance caused by modeling error (longer filter
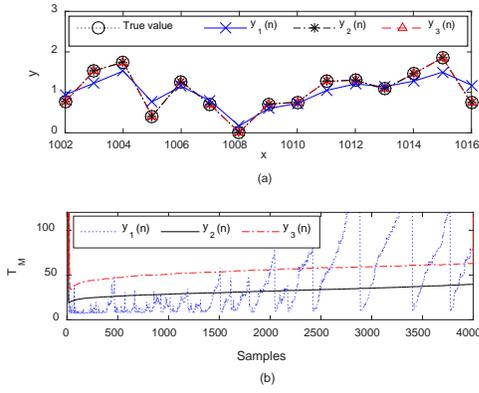
Fig. 7. Comparison of (a) fitting between three different models and (b) the corresponding values of T_M for SR-RBIC.
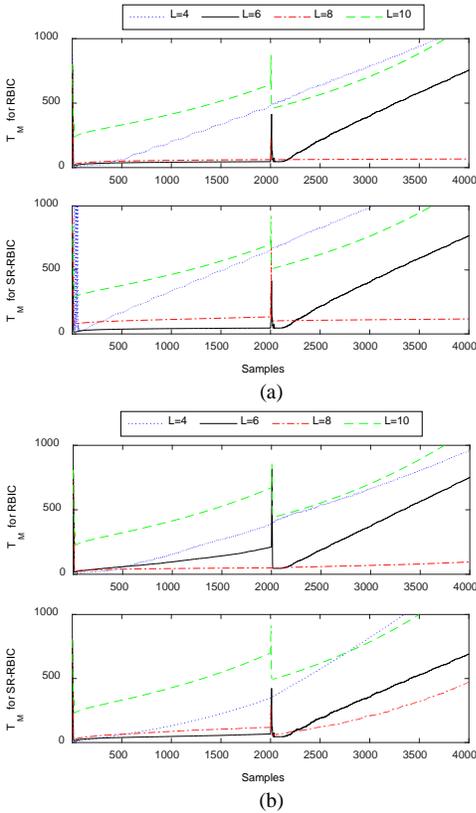

Fig. 8. Performance comparison of the conventional RBIC and the proposed SR-RBIC at: (a) SNR = 100 dB and (b) SNR = 20 dB.

length). The constant model should be rejected due to the significant fluctuation of the corresponding T_M curve.

Next, we evaluate the effect of noises on the performance of model selection criteria in a TV system. T_M curves of the conventional RBIC (the first two terms in (26)) and SR-RBIC (26) are computed for the candidate models. We suppose the input signal $\{x(n)\}$ is a summation of sinusoids containing three frequency components (500 Hz, 760 Hz, and 1600 Hz) from the 1st to the 2000th sample and an extra sinusoid of 1250 Hz appears after that. We now apply TVAR(4), TVAR(6), TVAR(8) and TVAR(10), respectively, to $\{x(n)\}$ and calculate T_M values at each sample of both RBIC and SR-RBIC. The results have been shown in Fig. 8. It can be seen that when SNR = 100 dB, both criteria select the same and also the true order (i.e. $L = 6$ for the first 2000 samples and $L = 8$ for the rest) after the initial iterations. However, when SNR decreases to 20 dB as

shown in Fig. 8(b), only the proposed SR-RBIC selects the best model while RBIC selects $L = 8$ for all samples.

It should be mentioned that although there are many newly proposed model order selectors, the adaptive versions are still not available in the current literature and we therefore do not provide comparisons with other model order selectors.

D. The Sequential Nonstationarity Detection Method with Automatic Model Order Selection

In this section, we introduce a sequential detection method with automatic model order selection. The method has been summarized as follows:

Method II: Sequential Nonstationarity Detection with Automatic Model Order Selection

---

1) Initialization:
   - set γ via (17) given a certain CFAR;
   - set the model order L by an initial guess;
   - set $n = 1$, $n_0 = 0$, and $w_0 = 0$;
2) Detector calculation
   - estimate the model parameter at the current time instant n, w, using SR-VFF-QRRLS;
   - calculate the corresponding T_{WT} via (16) (if the length of w_0 and w is different, the shorter will be added zeros to the end so that they have the same length);
   - calculate the corresponding T_M via (26) for model order $M_i = \max(1, L-1), L, L+1$;
3) Detection
   - if $T_{WT} < \gamma$ (no nonstationarity is detected), go to step 5);
   - else (nonstationarity detected), mark this time instant as $n_0 = n$ and set $w_0 = w$;
4) Model order selection iff $T_{WT} \geq \gamma$
   - select the model order from $[\max(1, L-1), L, L+1]$ based on T_M at time instant n;
   - set the selected model order;
5) Set $n = n + 1$ and return to 2) for the next data sample.
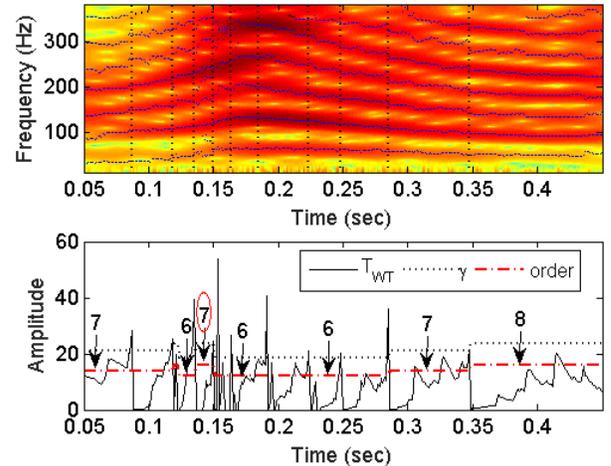
Fig. 9. Results of the Sequential Nonstationarity Detection with Automatic Model Order Selection method (Method II) using the speech signal (the word "jazz"). The selected model orders have been denoted by arrows.

According to Method II, the sequential detection is first executed using an initial guess for the model. The model selection will then be carried out once a change is detected since it is where model order is most likely to change. It should be noted that we assume the model order varies slowly and select

from (L–1, L, L+1), where L is the previously selected model order. In this way, much computational complexity can be saved. The above method is applied to a segment of the real speech "jazz" [38]. The sampling frequency is 1 kHz.

The results are shown in Fig. 9. It can be seen that the segmentation around 0.15 sec is comparatively short and the order "7" (marked in red circle) is overestimated, although it is also correct (i.e., the parameter corresponding to the redundant order can be made 0). This may be caused by the variance of the significant changes of the frequency components. Generally, the sequential detection and order selection system provides satisfactory performance. As a byproduct, the center frequencies are well estimated as shown in the spectrum.

VI. CONCLUSION

In this paper, we derive a parametric adaptive nonstationarity detector with model order selection from a weighted LF and the Bayesian framework. A sequential technique could be obtained that separates the data record into stationary segments, which allows users to determine nonstationarity of data for statistical analysis before a biased estimation occurs, and is very important in online applications such as speech and economics. The byproduct is the adaptive estimator using the SR-LVFF-QRRLS algorithm and the adaptive model order selector. This paper has provided a framework for employing adaptation procedures and developing sequential methods in detection and model selection. It can be extended to other advanced detection and information criteria in order to derive more efficient methods. A problem that is very common in practice and has not been addressed is the high level of noises. In this case, techniques for noise reduction can be used as pretreatment of data such that the proposed detection methods can provide satisfactory results. It is a comparatively separate problem, we hence have omitted any discussion and will explore it in more detail in the future.

Appendix A: Table I

Appendix B

In this Appendix, we derive the term $R_\tau(n)$ as $n \to \infty$. Since $R_\tau(n)$ is symmetric, it can be written as

$$R_\tau = \begin{pmatrix} s_0 & s_1 & \cdots & s_{L-1} \\ s_1 & s_0 & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{L-1} & & \cdots & s_0 \end{pmatrix},$$

where the time index at ∞ has been omitted.

By using some simple algebras, s_k can be written as

$$
\begin{aligned}
s_k &= -\sum_{i=0}^{n-L+1} i\lambda^i(\infty)x(n-i)x(n-i-k) \\
&\overset{n\to\infty}{\approx} -\sum_{i=0}^{n-L+1} i\lambda^i(\infty)r_{xx}(k) \\
&\overset{n\to\infty}{=} -\left(\frac{1-\lambda^{n-L+1}(\infty)}{(1-\lambda(\infty))^2} - \frac{(n-L+1)\lambda^{n-L+1}(\infty)}{1-\lambda(\infty)} \right)r_{xx}(k) \\
&\overset{n\to\infty}{\approx} -1/[(1-\lambda(\infty))^2]r_{xx}(k)
\end{aligned}
$$

A.1

TABLE I THE SR-QRRLS ALGORITHM

Initialization:
$R(0) = \sqrt{\delta}I$, δ is a small positive constant;
$u(0) = 0$, $w(0) = 0$ are null vectors.

Recursion:
Given $R(n-1)$, $u(n-1)$, $w(n-1)$, $x(n)$ and $d(n)$, compute at time n:
(i). The first update:
$$\begin{bmatrix} R^{(1)}(n) & u^{(1)}(n) \\ 0^T & c^{(1)}(n) \end{bmatrix} = Q^{(1)}(n)\begin{bmatrix} \sqrt{\lambda(n)}R(n-1) & \sqrt{\lambda(n)}u(n-1) \\ x^T(n) & d(n) \end{bmatrix}$$
The second update:
$$\begin{bmatrix} R(n) & u(n) \\ 0^T & c(n) \end{bmatrix} = Q(n)\begin{bmatrix} R^{(1)}(n) & u^{(1)}(n) \\ \sqrt{\mu(n)}d_l & \sqrt{\mu(n)}w_l(n-1) \end{bmatrix}$$
where $Q^{(1)}(n)$ and $Q(n)$ are calculated by Givens rotation to obtain the left hand side of each equation above, $\mu(n)=\xi(n)L$, and d_l is the l-th row of the identity matrix.
(ii). $w(n) = R^{-1}(n)u(n)$ (back-substitution).

where $r_{xx}(k)$ is the correlation between $x(n)$ and $x(n-k)$.

Appendix C

To evaluate the mean square behavior of $b(n)$, we determine a difference equation for the weight error covariance matrix $\Xi_b(n) = E[b(n)b^T(n)]$. From (31) and assuming the bias vector $b(n)$ is uncorrelated with $\overline{h^{(1)}}(n) = E[h^{(1)}(n)]$, we have

$$
\begin{aligned}
\Xi_b(n) &= \kappa^2(n)\tilde{R}_{XX}^{-1}(n)\Xi_b(n-1)\tilde{R}_{XX}^{-1} + B_1(n) \\
&\quad + B_2(n) - (B_3(n) + B_3^T(n))
\end{aligned}
$$
A.2

where $B_1(n) = \kappa^2(n)\tilde{R}_{XX}^{-1}(n)\overline{h^{(1)}}(n)[\overline{h^{(1)}}(n)]^T\tilde{R}_{XX}^{-1}(n)$, $B_2(n) = I_\kappa R_{XX}^{-1}(n)R_\tau(n)\overline{h^{(1)}}(n)[\overline{h^{(1)}}(n)]^T R_\tau(n)R_{XX}^{-1}(n)I_\kappa$ with $I_\kappa = I - \kappa(n)\tilde{R}_{XX}^{-1}(n) = I - \xi(n)\tilde{R}_{xx}^{-1}$ and $\tilde{R}_{xx} = R_{xx} + \xi(n)I$, and $B_3(n) = \kappa(n)I_\kappa R_{XX}^{-1}(n)R_\tau(n)\overline{h^{(1)}}(n)[\overline{h^{(1)}}(n)]^T\tilde{R}_{XX}^{-1}(n)$. $B_3(n)$ is quite small compared to $B_2(n)$ and hence can be ignored. Then, using the eigenvalue decomposition $\tilde{U}^T\tilde{R}_{XX}(n)\tilde{U} = D_{XX}$ and expressing $\Xi_b(n)$ in the transformed coordinate: $\Xi_B(n) = \tilde{U}^T\Xi_b(n)\tilde{U}$, the ith diagonal element of $\Xi_B(n)$ converges to

$$\Xi_{B,i}(\infty) = \frac{1}{1-\kappa^2(\infty)/\lambda_i^2}([\tilde{U}^T B_1(\infty)\tilde{U}]_i + [\tilde{U}^T B_2(\infty)\tilde{U}]_i)$$
A.3

where λ_i is the ith diagonal element of D_{XX}. Since $\frac{1}{1-\kappa^2(\infty)/\lambda_i^2} \approx 1$, the bias can be obtained as

$$J_b = \sum_{i=1}^{L} \Xi_{B,i}(\infty) \overset{n\to\infty}{\cong} \frac{\overline{\sigma}_h^2 Tr(I_\kappa^2)}{(1-\lambda(n))^2} + \xi^2(n)\overline{\sigma}_h^2 Tr(\tilde{R}_{xx}^{-2}).$$
A.4

where the covariance matrix $\overline{R}_h = \overline{h^{(1)}}(n)[\overline{h^{(1)}}(n)]^T = \overline{\sigma}_h^2 I$.

The mean square behavior of $v(n)$ can be evaluated from its covariance matrix $\Xi_v(n) = E[v(n)v^T(n)]$. From (33), we have

$$\Xi_v(n) = \kappa^2(n)\tilde{R}_{XX}^{-1}(n)\Xi_v(n-1)\tilde{R}_{XX}^{-1} + V_1(n) + V_2(n)$$
A.5

where $V_1(n) = \sigma_\Sigma^2 I_\kappa R_{XX}^{-1}(n)\hat{R}_{XX}(n)R_{XX}^{-1}(n)I_k$ with $\hat{R}_{XX}(n) = X^T(n)\Lambda^2(n)X(n) \overset{n\to\infty}{\approx} (1-\lambda^2(n))^{-1}R_{xx}$ [28], $\sigma_\Sigma^2 = \sigma_\eta^2 + \sigma_\upsilon^2$,

$\sigma_\eta^2 = E[\eta^2(n)]$ and $\sigma_\upsilon^2 = E[\upsilon^2(n)]$, and $\boldsymbol{V}_2(n) = \boldsymbol{I}_\kappa \boldsymbol{R}_{XX}^{-1}(n) \cdot \boldsymbol{R}_\tau(n)\boldsymbol{R}_{\delta h}(n)\boldsymbol{R}_\tau(n)\boldsymbol{R}_{XX}^{-1}(n)\boldsymbol{I}_\kappa$ with $\boldsymbol{R}_{\delta h}(n) = \delta\boldsymbol{h}^{(1)}(n)[\delta\boldsymbol{h}^{(1)}(n)]^T = \sigma_{\delta h}^2 \boldsymbol{I}$. Expressing $\boldsymbol{\Xi}_v(n)$ in the transformed coordinate: $\boldsymbol{\Xi}_V(n) = \tilde{\boldsymbol{U}}^T\boldsymbol{\Xi}_v(n)\tilde{\boldsymbol{U}}$ and summarizing the diagonal elements

$$\boldsymbol{\Xi}_{V,i}(\infty) = \frac{1}{1-\kappa^2(\infty)/\lambda_i^2}([\tilde{\boldsymbol{U}}^T\boldsymbol{V}_1(\infty)\tilde{\boldsymbol{U}}]_i + [\tilde{\boldsymbol{U}}^T\boldsymbol{V}_2(\infty)\tilde{\boldsymbol{U}}]_i), \qquad \text{A.6}$$

we can obtain

$$J_v = \sum_{i=1}^{L} \boldsymbol{\Xi}_{V,i}(\infty) \overset{n\to\infty}{\approx} \frac{1-\lambda(n)}{1+\lambda(n)}\sigma_\Sigma^2 Tr(\boldsymbol{I}_\kappa^2\boldsymbol{R}_{xx}^{-1}) + \frac{\sigma_{\delta h}^2}{(1-\lambda(n))^2}Tr(\boldsymbol{I}_\kappa^2) \qquad \text{A.7}$$

where we have used the following relationship [28]: $\tilde{\boldsymbol{R}}_{XX}^{-1}(n) \overset{n\to\infty}{\approx} (1-\lambda(n))\tilde{\boldsymbol{R}}_{xx}^{-1}(n)$ and $\boldsymbol{R}_{xx}(n) \overset{n\to\infty}{\approx} (1-\lambda(n))\boldsymbol{R}_{XX}(n)$. If $\kappa(n) = 0$, the result for estimation variance coincides with that in [28].

Acknowledgment

References

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.

[2] J. I. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 32, pp. 561–582, April 1975.

[3] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. Norwell, MA: Artech House, Apr. 30, 2005.

[4] R. Tavares and R. Coelho, "Speech enhancement with nonstationary acoustic noise detection in time domain," *IEEE Trans. Signal Process. Lett.*, vol. 23, no. 1, pp. 2704–2715, Jan. 2016.

[5] D. Rudoy, T. Quatieri, and P. Wolfe, "Time-varying autoregressions in speech: detection theory and application," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 4, pp. 977–989, May 2011.

[6] Khalil and J. Duchêne, "Uterine EMG analysis: a dynamic approach for change detection and classification," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 6, pp. 748–756, Jun. 2000.

[7] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[8] M. B. Priestley and T. S. Rao, "A test for non-stationarity of time-series," *J. Royal Stat. Soc.*, vol. 31, pp. 1401–149, 1969.

[9] M. B. Priestley, *Non-linear and non-stationary time series analysis*. London, U.K.: Academic, 1988.

[10] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: a time-frequency approach," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3459–3470, Jul. 2010.

[11] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson Prentice-Hall, 2005.

[12] S. Kay, "A new nonstationarity detector," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1440–1451, Apr. 2008.

[13] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[14] L. A. Liporace, "Linear estimation of nonstationary signals," *J. Acoust. Soc. Amer.*, vol. 58, no. 6, pp. 1268–1295, Dec. 1975.

[15] Z. Zhou, H. C. So and M. G. Christensen, "Parametric modeling for damped sinusoids from multiple channels," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp.3895–3907, Aug. 2013.

[16] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-31, pp. 899–911, 1983.

[17] D. Rudoy and T. Georgiou, "Regularized parametric models of nonstationary processes," in *Proc. the 19th Int. Symp. Math. Theory Netw. Syst., (MTNS2010)*, Budapest, Hungary, pp. 5–9, July 2010.

[18] A. Maio, S. Kay, and A. Farina, "On the invariance, coincidence, and statistical equivalence of the GLRT, Rao test, and Wald test," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1967–1979, Apr. 2010.

[19] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[20] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 115–127, Mar. 1986.

[21] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[22] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, pp. 2538–2541, Sep. 1999.

[23] H. Li and J. H. Michels, "Parametric adaptive signal detection for hyperspectral imaging," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2704–2715, Jul. 2006.

[24] Y. J. Chu, S. C. Chan, Z. G. Zhang, and K. M. Tsui, "A new regularized TVAR-based algorithm for recursive detection of nonstationarity and its application to speech signals," in *Pro. Statistical Signal Process. (SSP 2012)*, pp. 361–364, Ann Arbor, USA, Aug. 2012.

[25] S. C. Chan, Y. J. Chu, Z. G. Zhang, and K. M. Tsui, "A new variable regularized QR decomposition-based recursive least M-estimate algorithm—performance analysis and acoustic applications," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 907–922, May 2013.

[26] S. C. Chan and Y. J. Chu, "A new state-regularized QRRLS algorithm with variable forgetting factor," *IEEE Trans. Circ. Syst. II*, vol. 59, no. 3, pp. 183–187, Mar. 2012.

[27] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6 no. 2, pp. 461–464, 1978.

[28] Y. J. Chu and S. C. Chan, "A new local polynomial modeling-based variable forgetting factor RLS algorithm and its acoustic applications," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 2059–2069, Nov. 2015.

[29] S. H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, Aug. 2005.

[30] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.* vol. 46, no.10, pp. 2726–2735, Oct. 1998.

[31] S, Kallummil and S. Kalyani, "High SNR consistent linear model order selection and subset selection," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4307–4312, Aug. 2016.

[32] H. Akaike, "Information theory and the maximum likelihood principle," *International Symposium on Information Theory*, 1973.

[33] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd ed.*. Baltimore, MD: Johns Hopkins, p. 51, 1996.

[34] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," in *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[35] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Process.*, vol. 5, pp. 267–285, 1983.

[36] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1298–1312, Apr. 2007.

[37] Z. G. Zhang, Y. S. Hung, and S. C. Chan, "Local polynomial modeling of time-varying autoregressive models with application to time-frequency analysis of event-related EEG," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 557–566, Mar. 2011.

[38] T. Quatieri, *Discrete-time Speech Signal Processing*, Prentice Hall, 2001.