Recursive Adaptive Algorithms for Fast and Rapidly Time-Varying Systems

Yuanjin Zheng, *Member, IEEE,* and Zhiping Lin, *Senior Member, IEEE*

***Abstract*—In this paper, some new schemes are developed to improve the tracking performance for fast and rapidly time-varying systems. A generalized recursive least-squares (RLS) algorithm called the trend RLS (T-RLS) algorithm is derived which takes into account the effect of local and global trend variations of system parameters. A bank of adaptive filters implemented with T-RLS algorithms are then used for tracking an arbitrarily fast varying system without knowing *a priori* the changing rates of system parameters. The optimal tracking performance is attained by Bayesian *a posteriori* combination of the multiple filter outputs, and the optimal number of parallel filters needed is determined by extended Akaike's Information Criterion and Minimum Description Length information criteria. An RLS algorithm with modification of the system estimation covariance matrix is employed to track a time-varying system with rare but abrupt (jump) changes. A new online wavelet detector is designed for accurately identifying the changing locations and the branches of changing parameters. The optimal increments of the covariance matrix at the detected changing locations are also estimated. Thus, for a general time-varying system, the proposed methods can optimally track its slowly, fast and rapidly changing components simultaneously.**

***Index Terms*—Dyadic wavelet transform (DWT), recursive wavelet change detector, system identification, time-varying system, trend recursive least-squares (T-RLS) algorithm.**

I. Introduction

IN MANY applications such as speech recognition, communication channel equalization, process control, and biomedical signal processing, the underlying time-varying systems are subject to fast and/or rapidly changing environments [1], [4]. To track a fast varying system, several variable-step-size adaptive algorithms have been proposed to improve the tracking performance [1], [3], [4], [8]. However, if the variations of system parameters show obvious deterministic local or global trends, stochastically perturbed difference equation constraints should be used as smooth priors for system parameters [5]. Based on these prior trend models, a multistep algorithm and a Kalman filtering algorithm have been used to track fast varying systems in [6], [7]. Further, when the varying trends of system parameters are unknown, one way to overcome this problem is to assume that the variations of system parameters satisfy a first-order Markov chain model and a multiple adaptive Kalman filtering (MAKF) algorithm is developed for parameter tracking (see, e.g., [9]). Another way is to employ the vector space adaptive filtering and tracking algorithm [10]. However, the computational loads of all the above methods are heavy and the exact statistical characteristics of system and measurement noises are required.

In this paper, a new trend reciursive least-squares (T-RLS) algorithm is derived for tracking a fast varying system with deterministic and known trends. One of the advantages of this T-RLS algorithm is that it does not require the exact information on system and measurement noise variances, and state space equation coefficient matrices. Moreover, extended Akaike's Information Criterion (AIC) and Minimum Description Length (MDL) criteria are proposed to determine the optimal order of a time-varying system online. For tracking a general fast varying system with unknown order trends, a multiple T-RLS algorithm is developed which attains the optimal posterior estimation and is computationally simpler than the MAKF algorithm.

When a time-varying system is subject to rare but abrupt (jumping) changes, the estimated parameters by conventional adaptive algorithms cannot track the variations of true system parameters in the vicinity of these jumping locations, resulting in the so called "lag" estimation. Three methods can be used to mitigate the effect of "lag" estimation. The first method is to use variable forgetting factor RLS algorithms [3]. The second is to increase the system estimation covariance matrix at the jumping locations [11], [12]. The third includes various Bayesian Kalman filtering algorithms [13], [14]. In this paper, the second method will be adopted to track the abrupt changes of system parameters. One difficulty of this method is how to identify the unknown locations and amplitudes of the abrupt changes online. Some approaches have been developed toward this task [15]–[18]. The obvious tradeoff between detection sensitivity and robustness exists in these methods. It has been shown that the idea of modification of covariance matrix only with respect to the detected jumping parameter branch(es) would improve the identification accuracy [12], [19]. Hence, it is desirable to have of a simple and yet efficient detection and modification algorithm.

To identify the rapidly changing points effectively, a new online detection algorithm based on a multiscale product sequence in wavelet domain is proposed in this paper. The new wavelet detector can efficiently suppress background noise and enhance

Manuscript received March 25, 2002; revised January 22, 2003. This work was supported by the Academic Research Fund, Ministry of Education, Republic of Singapore (RG 42/99). Please note that the review process for this paper was handled exclusively by the TCAS-I editorial office and that the paper has been published in TCAS-II for logistical reasons. This paper was recommended by Associate Editor N. Ling.

Y. Zheng was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798. He is now with the ICS Laboratory, Institute of Microelectronics, Singapore 117685 (e-mail: yuanjin@ime.a-star.edu.sg).

Z. Lin is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: EZPLin@ntu.edu.sg).

Digital Object Identifier 10.1109/TCSII.2003.816906

the abruptly changing components so that it is very robust to interferences (with low false alarm probability) and sensitive to jumping changes (with high detection probability) compared with the conventional detectors. A new algorithm for selectively modifying the elements of the covariance matrix is proposed. Moreover, the optimal increments of the covariance matrix are determined.

The rest of the paper is organized as follows. In Section II, a T-RLS algorithm and a multiple T-RLS algorithm are developed which can track an arbitrarily fast varying system. In Section III, a new wavelet detector is proposed for identifying the abrupt changes and a scheme for selectively modifying the estimation covariance matrix is presented for tracking rapidly changing systems. In Section IV, simulation results are provided which verify the superior performance of the proposed algorithms. Concluding remarks are given in Section V.

II. Recursive Adaptive Algorithm for Fast Changing Systems

A. T-RLS Algorithm

A time-varying system commonly can be represented by a linear regression equation and the changes of system parameters can be modeled with an order one (first-order) random walk model [10], [11], [25], [26]

$$\theta_{t+1} = \theta_t + w_t \tag{1}$$

$$y_t = \varphi_t^T \theta_t + e_t. \tag{2}$$

Here, θ_t is the true system parameter vector of size $N \times 1$, y_t is the scalar observation (output) signal, φ_t is the system (input) regressor vector of size $N \times 1$, w_t is the system noise vector of size $N \times 1$ with covariance matrix $\mathrm{Cov}(w_t) = Q$, and e_t is the measurement noise signal with variance $\mathrm{Var}(e_t) = \sigma^2$. When the variations of system parameters are slow enough, an RLS algorithm can be used to track the time-varying system [25], [26]

$$\hat{\theta}_t = \hat{\theta}_{t-1} + G_t \varepsilon_t \tag{3}$$

$$\varepsilon_t = y_t - \varphi_t^T \hat{\theta}_{t-1} \tag{4}$$

$$G_t = P_t \varphi_t = \frac{P_{t-1}\varphi_t}{\lambda_t + \varphi_t^T P_{t-1} \varphi_t} \tag{5}$$

$$P_t = \frac{1}{\lambda_t}\left(P_{t-1} - \frac{P_{t-1}\varphi_t \varphi_t^T P_{t-1}}{\lambda_t + \varphi_t^T P_{t-1}\varphi_t}\right) \tag{6}$$

where ε_t is the *a priori* prediction error, G_t is the filtering gain, P_t is the estimation covariance matrix, and λ_t is the forgetting factor.

However, if the time-varying parameters change fast, the first-order random walk model is not sufficient to describe the variations of the system parameters [5], [6]. To accurately model the time-dependent parameters and increase the tracking ability for a fast varying system, a general random walk model can be adopted which sufficiently includes the prior information such as the qth order deterministic trend, stochastic trend, and seasonal components of a nonstationary process [5]. By this model, Kalman filtering can be employed for system parameter estimation and tracking. However, the explicit statistical characteristics of system and measurement noises are required before filtering can be performed [6]. In this subsection, we derive a T-RLS algorithm which can adaptively track fast varying parameters without knowing the explicit statistical characteristics of (possibly nonstationary) system and measurement noises. It is well known that a conventional varying forgetting factor RLS algorithm has to compromise its tracking performance with estimation variance. Similar to the trend Kalman filter algorithm presented in [5], [6], the proposed T-RLS algorithm can achieve fast tracking and small estimation variance simultaneously.

Assume the system parameters of a fast varying system can be modeled with a general random walk model [7]

$$\theta_{t+1} = A_t \theta_t + w_t \tag{7}$$

$$y_t = \varphi_t^T \theta_t + e_t \tag{8}$$

where the various prior trend information can be included in a nonsingular matrix A_t with size $N_A \times N_A$, and the sizes of φ_t and θ_t are both of $N_A \times 1$. Obviously, (7) is a more precise model for fast varying system parameters. In the following, assume that there are $n (1 \le t \le n)$ available observation samples $y_1, \ldots, y_n$ and the variables at time $t = 0$ represent the initial values in the following. To derive a T-RLS algorithm, rewriting (7) as $\theta_{n+1} = A_n \theta_n + w_n$ and then reversely iterating it $(n-j)$ times for $j = n, \ldots, 0$, we obtain

$$\theta_j = \Psi_{jn}\theta_n - \sum_{i=j+1}^{n} \Psi_{ji} w_{i-1}, \qquad j = 0, \ldots, n \tag{9}$$

where the inverse transition matrix Ψ_{jn} is defined as

$$\Psi_{jn} = \begin{cases} (A_{n-1} \cdots A_j)^{-1}, & \text{if } j < n \\ I, & \text{if } j = n. \end{cases} \tag{10}$$

Substituting (9) into (8) for $t = j = 0, \ldots, n$ gives

$$y_j = \varphi_j^T \Psi_{jn}\theta_n + e_j - \varphi_j^T \sum_{i=j+1}^{n} \Psi_{ji} w_{i-1}, \qquad j = 0, \ldots, n. \tag{11}$$

Define an $(n+1) \times N_A$ matrix

$$H_n = \left[\varphi_n, (\varphi_{n-1}^T \Psi_{(n-1)n})^T, \ldots, (\varphi_j^T \Psi_{jn})^T, \ldots, (\varphi_0^T \Psi_{0n})^T\right]^T \tag{12}$$

and two $(n+1) \times 1$ vectors

$$Y_n = [y_n, y_{n-1}, \ldots, y_j, \ldots, y_0]^T \tag{13}$$

$$V_n = \left[e_n, e_{n-1} - \varphi_{n-1}^T w_{n-1}, \ldots, e_j - \varphi_j^T \sum_{i=j+1}^{n} \Psi_{ji} w_{i-1}, \ldots, e_0 - \varphi_0^T \sum_{i=1}^{n} \Psi_{ji} w_0\right]^T. \tag{14}$$

Stack (11) in column for $0 \le j \le n$ and use notations (12)(14) to arrive a vector-matrix equation

$$Y_n = H_n \theta_n + V_n. \tag{15}$$

From (15), the optimal estimation of system parameters θ_n based on observation data Y_n (in the sense of weighted least-squares solution) can be obtained as [2]

$$\hat{\theta}_n = (H_n^T \Upsilon H_n)^{-1}(H_n^T \Upsilon Y_n) \tag{16}$$

where Υ denotes a weighting matrix. If Υ takes the inverse of the covariance matrix of the equivalent noise vector V_n, i.e.,

$$\Upsilon = (\mathrm{Var}(V_n))^{-1} = (E(V_n V_n^T))^{-1}. \tag{17}$$

Kalman filtering can be readily derived to achieve an unbiased minimal variance estimation [2]. For deriving the T-RLS algorithm, a reasonable choice is to introduce the following weighting matrix with a set of discounting factors on the diagonal elements [25]:

$$\Upsilon = \mathrm{diag}\,\{\alpha(n,\,n),\,\alpha(n,\,n-1),\,\alpha(n,\,n-1),\,\ldots,\,\alpha(n,\,j),\,\ldots,\,\alpha(n,\,0)\} \tag{18}$$

where $\alpha(n,\,n)=1$, $\{\alpha(n,\,j)\}_{0\leq j\leq n-1}$ are discounting factors defined as

$$\alpha(n,\,j) = \lambda_n \alpha(n-1,\,j) = \Pi_{k=j+1}^{n}\lambda_k \qquad j = n-1,\,n-2,\,\ldots,\,0 \tag{19}$$

and λ_k ($1 \leq k \leq n$) is named as the kth forgetting factor. Typically, $0 < \lambda_k \leq 1$ [25], [26]. To derive the recursive estimation equations, define

$$P_t = (H_t^T \Upsilon H_t)^{-1} = \left(\sum_{j=0}^{t} \alpha(t,\,j)\Psi_{jt}^T \varphi_j \varphi_j^T \Psi_{jt}\right)^{-1}. \tag{20}$$

Thus

$$\begin{aligned}
P_t^{-1} &= \sum_{j=0}^{t} \alpha(t,\,j)\Psi_{jt}^T \varphi_j \varphi_j^T \Psi_{jt} \\
&= \lambda_t \sum_{j=0}^{t-1} \alpha(t-1,\,j)(\Psi_{j(t-1)}A_{t-1}^{-1})^T \varphi_j \varphi_j^T \\
&\quad \cdot (\Psi_{j(t-1)}A_{t-1}^{-1}) + \varphi_t \varphi_t^T \\
&= \lambda_t (A_{t-1}^T)^{-1} \left(\sum_{j=0}^{t-1} \alpha(t-1,\,j)\Psi_{j(t-1)}^T \varphi_j \varphi_j^T \Psi_{j(t-1)}\right) \\
&\quad \cdot (A_{t-1}^{-1}) + \varphi_t \varphi_t^T \\
&= \lambda_t (A_{t-1}^T)^{-1} P_{t-1} A_{t-1}^{-1} + \varphi_t \varphi_t^T \\
&= \lambda_t (A_{t-1} P_{t-1}^{-1} A_{t-1}^T)^{-1} + \varphi_t \varphi_t^T.
\end{aligned} \tag{21}$$

From (21), it is easy to show

$$\lambda_t P_t (A_{t-1}^T)^{-1} P_{t-1}^{-1} = (I - P_t \varphi_t \varphi_t^T) A_{t-1}. \tag{22}$$

Substituting (12), (13), (19). and (20) into (16) and using (22) gives

$$\begin{aligned}
\hat{\theta}_t &= P_t (H_t^T \Upsilon Y_t) \\
&= P_t \sum_{j=0}^{t} \alpha(t,\,j)\Psi_{jt}^T \varphi_j y_j \\
&= P_t \left(\lambda_t (A_{t-1}^T)^{-1} \sum_{j=0}^{t-1} \alpha(t-1,\,j)\Psi_{j(t-1)}^T \varphi_j y_j + \varphi_t y_t\right) \\
&= P_t \left(\lambda_t (A_{t-1}^T)^{-1} P_{t-1}^{-1} \right. \\
&\qquad \left. \cdot \left(P_{t-1} \sum_{j=0}^{t-1} \alpha(t-1,\,j)\Psi_{j(t-1)}^T \varphi_j y_j\right) + \varphi_t y_t\right) \\
&= \lambda_t P_t (A_{t-1}^T)^{-1} P_{t-1}^{-1} \hat{\theta}_{t-1} + P_t \varphi_t y_t \\
&= (I - P_t \varphi_t \varphi_t^T) A_{t-1} \hat{\theta}_{t-1} + P_t \varphi_t y_t.
\end{aligned} \tag{23}$$

The final equality in (23) is just

$$\hat{\theta}_t = A_{t-1}\hat{\theta}_{t-1} + P_t \varphi_t \left(y_t - \varphi_t^T (A_{t-1}\hat{\theta}_{t-1})\right). \tag{24}$$

By applying the matrix inverse lemma [25], (21) can be rewritten in another equivalent form

$$P_t = \frac{1}{\lambda_t}\left(A_{t-1}P_{t-1}A_{t-1}^T - \frac{A_{t-1}P_{t-1}A_{t-1}^T \varphi_t \varphi_t^T A_{t-1} P_{t-1} A_{t-1}^T}{\lambda_t + \varphi_t^T A_{t-1} P_{t-1} A_{t-1}^T \varphi_t}\right). \tag{25}$$

In summary, for a fast varying system modeled with a general random walk model (7) and (8), the T-RLS algorithm can be realized by applying (24) and (25) alternatively. Notice that when $A_t = \mathbf{I}_{N_A} = \mathbf{I}_N$, (7) and (8) reduce to (1) and (2) and the T-RLS algorithm reduces to the conventional RLS algorithm. Commonly, we choose $\lambda_1 = \cdots = \lambda_n = \lambda$ with $0 < \lambda \leq 1$ to simplify the computation. If the order of a general random walk model is properly chosen, λ commonly takes a value less than but close to 1.

B. Multiple T-RLS Algorithm for Tracking Arbitrarily Fast Varying Systems

If the prior trend information of a fast time-varying system is exactly known, i.e., the matrix A_t is deterministic and known, the T-RLS algorithm can be used for accurate parameter estimation. However, A_t commonly is not explicitly known or even completely unknown. In this case, a bank of T-RLS filters with a spreading of assumed A_t matrices can be performed in parallel for parameter estimation separately and at the same time a Bayesian posterior combination is employed to attain the optimal parameter estimation. We name this scheme as the Multiple T-RLS (MT-RLS) algorithm. The MT-RLS algorithm does not estimate the unknown A_t of a fast varying system directly. Instead, it obtains the optimal parameter estimation through weighted averaging the multiple adaptive filter outputs. The optimal weighting coefficients can be determined by Bayesian probability inference. It is shown in the following that the MT-RLS algorithm can be realized with quite simple computations and yet has good performance.

For an unknown system (7) and (8) (where matrix A_t, variance of e_t and covariance of w_t all are unknown), M T-RLS adaptive filters can be adopted for parallel adaptive filtering. Let the ith $(1 \leq i \leq M)$ filter corresponds to the following hypothetical values of the design parameters $\aleph^i = \{A_t = A_t^i\}$. The recursive parameter estimation by the ith filter is performed as

$$\hat{\theta}_t^i = A_{t-1}^i \hat{\theta}_{t-1}^i + P_t^i \varphi_t \varepsilon_t^i \tag{26}$$

$$\varepsilon_t^i = y_t - \varphi_t^T (A_{t-1}^i \hat{\theta}_{t-1}^i) \tag{27}$$

$$P_t^i = \frac{1}{\lambda} \left(A_{t-1}^i P_{t-1}^i (A_{t-1}^i)^T - \frac{A_{t-1}^i P_{t-1}^i (A_{t-1}^i)^T \varphi_t \varphi_t^T A_{t-1}^i P_{t-1}^i (A_{t-1}^i)^T}{\lambda + \varphi_t^T A_{t-1}^i P_{t-1}^i (A_{t-1}^i)^T \varphi_t} \right) \tag{28}$$

where A_t^i is the assumed trend matrix for the ith filter. Without loss of generality, the ith priori prediction error ε_t^i is assumed an i.i.d. Gaussian distributed sequence with zero mean and variance $\mathrm{Var}(\varepsilon_t^i) = \sigma_t^2$ [12]. The optimal parameter estimation based on the completed observation sequence $Y^t = [y_t, y_{t-1}, \ldots, y_1]$ can be obtained as

$$\hat{\theta}_t = E\left(\theta_t | Y^t\right) = \sum_{i=1}^{M} \hat{\theta}_t^i \beta_t^i \tag{29}$$

where β_t^i is the posterior probability of the hypothetical model $\aleph^i$ given the data set Y^t, i.e., $\beta_t^i = \tilde{\beta}_t^i = p(i|Y^t)$.

A more general assumption is to consider a nonstationary system where the measurement noise itself possibly changes with time slowly. In this case, since σ_t and β_t^i are both slowly varying, only L most recent observation data samples are used to reliably estimate β_t^i so that σ_t is assumed invariant in this time interval (i.e., $\sigma_{t-L+1} = \cdots = \sigma_t$). Denote $Y_{t-L+1}^t = [y_t, y_{t-1}, \ldots, y_{t-L+1}]$, and redefine β_t^i as

$$\beta_t^i = p(i|Y_{t-L+1}^t). \tag{30}$$

To estimate β_t^i, Bayes rule can be applied

$$p(i|Y_{t-L+1}^t) = \frac{p(Y_{t-L+1}^t, i)}{\sum_{i=1}^{M} p(Y_{t-L+1}^t, i)} = \frac{\int_0^\infty p(Y_{t-L+1}^t | \sigma_t^2, i) p(\sigma_t^2 | i) p(i) d\sigma_t^2}{\sum_{i=1}^{M} p(Y_{t-L+1}^t, i)}. \tag{31}$$

The denominator of the above equation is a constant which is not relevant to i. Under i.i.d. Gaussian assumption and according to (27), the conditional likelihood function of y_t based on σ_t^2 and i is $p(y_t|\sigma_t^2, i) = (2\pi\sigma_t^2)^{-1/2} \exp\left(-(\varepsilon_t^i)^2 / 2\sigma_t^2\right)$. Thus the joint conditional likelihood function of recent L samples can be represented as

$$p(Y_{t-L+1}^t | \sigma_t^2, i) = \prod_{j=t-L+1}^{t} p(y_j | \sigma_t^2, i) = (2\pi\sigma_t^2)^{-L/2} \exp\left(- \sum_{j=t-L+1}^{t} \frac{(\varepsilon_j^i)^2}{2\sigma_t^2} \right). \tag{32}$$

As in [20], we assign noninformative prior distributions to i and σ_t^2 as

$$p(i) = \frac{1}{M} \tag{33}$$

$$p(\sigma_t^2 | i) = p(\sigma_t^2) = \frac{1}{\sigma_t^2}. \tag{34}$$

Substituting (32)–(34) into (31), it is easy to derive

$$p(i|Y_{t-L+1}^t) = \tilde{c}_1 \int_0^\infty p(Y_{t-L+1}^t | \sigma_t^2, i) p(\sigma_t^2|i) p(i) d\sigma_t^2 \tag{35}$$

$$= \tilde{c}_2 \left(\sum_{j=t-L+1}^{t} (\varepsilon_j^i)^2 \right)^{-L/2}. \tag{36}$$

Here, $\tilde{c}_1$ and $\tilde{c}_2$ are two different constants. Define

$$\varpi_t^i = \left(\sum_{j=t-L+1}^{t} (\varepsilon_j^i)^2 \right)^{-L/2} \tag{37}$$

which can be estimated recursively as

$$\varpi_t^i = \begin{cases} \left(\left(1 - \frac{1}{t}\right) (\varpi_{t-1}^i)^{-2/L} + \frac{L}{t} (\varepsilon_t^i)^2 \right)^{-L/2}, & \text{if } 1 \leq t \leq L, \\ \left((\varpi_{t-1}^i)^{-2/L} + (\varepsilon_t^i)^2 - (\varepsilon_{t-L}^i)^2 \right)^{-L/2}, & \text{if } t > L. \end{cases} \tag{38}$$

Since $\sum_{i=1}^{M} \beta_t^i = 1$, β_t^i can be recursively estimated based on ϖ_t^i as

$$\beta_t^i = \frac{\varpi_t^i}{\sum_{i=1}^{M} \varpi_t^i}. \tag{39}$$

In summary, the MT-RLS algorithm can be recursively implemented by (26)–(29), (38), and (39). Like the T-RLS algorithm, the MT-RLS does not require the information of system and measurement noises and, thus, is computationally simpler than the MAKF algorithm. At this point, several parameters L, M and $A_t^i (1 \leq i \leq M)$ in the MT-RLS algorithm remain to be selected. In the following section, we will discuss the problem of optimal parameter selection.

C. Parameter Selection for the MT-RLS Algorithm

The filtering length L should be chosen to ensure the signal in this interval is approximately stationary, so that σ_t^2 can be considered as unchanged and can be estimated efficiently by the observed data set in this interval. Recall that the memory of the RLS algorithm is approximately $(-1/\ln\lambda) \approx (1/(1-\lambda))$ if λ is less than but close to 1 [25], [26]. Thus, L can take

$$L = \left[\frac{1}{1-\lambda}\right]_+ \tag{40}$$

where $[.]_+$ means rounding toward positive infinity.

The forms and values of A_t^i may be chosen by referring to the varying trend of the system parameters to be studied. On the one hand, A_t^i should be selected general enough to encompass the true type and order of the underlying system. Commonly used trend models include local polynomial trend model, stochastic trend model and local polynomial seasonal component model etc. [5]. On the other hand, taking small number of parallel adaptive filters can reduce the algorithm complexity and improve the computational efficiency. Therefore, we should find a method to determine the optimal filter number $M = \hat{M}$ (acting as the bandwidth of the adaptive algorithm), which best balances the algorithm complexity and performance.

For tracking a time-varying system with deterministic but unknown trend, the assumed A_t in (7) can take different (integer) order fixed matrices. For example, if the system parameters are assumed to follow a qth order polynomial trend, the qth order stochastically perturbed difference equation can be used to represent the dynamics of the system parameters. Hence A_t will be a fixed matrix with size $qN \times qN$ ($N_A = qN$) as (41), shown at the bottom of the page, where $\otimes$ represents Kronecker product and $C_q^k = (q!/k!(q-k)!)$. Similarly, if the system parameters are assumed to follow a qth order autoregressive (AR) stochastic trend, A_t is also a $qN \times qN$ fixed matrix but with different elements (see [5] and [6] for details). Taking large orders can improve the tracking ability but also increase the model complexity, and taking small orders is computationally simple but will decrease the tracking ability. Obviously, an optimal trend model order can be chosen for an unknown system once the forms and values of $\{A_t^i\}_{i=1,\ldots,M}$ are selected. This order can be determined by the AIC or MDL information criteria, etc. [6] which optimally balance model representation ability and model complexity. Assume M T-RLS adaptive filters are performed in parallel and the ith $(1 \le i \le M)$ filter takes trend model order $\varrho(i)$. To determine the optimal trend model order, an extended AIC or MDL sequence AIC_t^i and MDL_t^i for the trend model of the ith filter at time t can be derived as follows:

$$\begin{aligned} AIC_t^i(\varrho(i)) &= \log\left(\frac{\sum_{j=t-L+1}^{t}(\varepsilon_j^i)^2}{L}\right) + \frac{2\varrho(i)}{L} \\ &= \log\left(\frac{(\varpi_t^i)^{-2/L}}{L}\right) + \frac{2\varrho(i)}{L} \end{aligned} \tag{42}$$

$$\begin{aligned} MDL_t^i(\varrho(i)) &= \log\left(\frac{\sum_{j=t-L+1}^{t}(\varepsilon_j^i)^2}{L}\right) + \frac{\varrho(i)\log(L)}{L} \\ &= \log\left(\frac{(\varpi_t^i)^{-2/L}}{L}\right) + \frac{\varrho(i)\log(L)}{L} \end{aligned} \tag{43}$$

where $\varrho(i)$ corresponds to the assumed A_t^i. For a system with all of the three trends,

$$\varrho(i) = M_1^i + 2M_2^i + M_3^i + \delta(M_1^i) + \delta(M_2^i) + \delta(M_3^i) \tag{44}$$

where M_1^i, M_2^i, and M_3^i are the model orders of the local polynomial trend, the AR trend, and the seasonal trend of the ith filter, respectively. $\delta(M_j^i) = 1$, if $M_j^i \neq 0$ and $\delta(M_j^i) = 0$, if $M_j^i = 0$ for each $1 \le i \le M$, $1 \le j \le 3$. See [5] for more details. We select the Ith filter as the most matched filter which leads to the minimal AIC_t^i or MDL_t^i, i.e.,

$$\begin{aligned} AIC_t^I \le AIC_t^i \quad &\text{or} \quad MDL_t^I \le MDL_t^i, \\ 1 \le i \le M; \quad &i \neq I \end{aligned} \tag{45}$$

and $\varrho(I)$ as the optimal integer trend model order. Without loss of generality, we sort $\{\varrho(i)\}_{i=1,\ldots,M}$ as

$$\varrho(1) \ge \cdots \ge \varrho(I) \le \cdots \le \varrho(M). \tag{46}$$

To reduce the number of parallel filters while keeping good performance, we can design an MT-RLS algorithm with *constant* filter number (bandwidth) which can optimally balance the algorithm performance and the computational complexity. Explicitly, we take $\hat{M}$ as a constant odd number and the corresponding $\hat{M}$ filter orders as $\varrho(I - ((\hat{M}-1)/2)), \ldots, \varrho(I), \ldots, \varrho(I + ((\hat{M}-1)/2))$. Here, the filter with tend model order $\varrho(I)$ corresponds to the most matched filter, the filter with trend model order $\varrho(I - (\hat{M}-1)/2))$ corresponds to the least underdetermined filter, the filter with trend model order $\varrho(I + ((\hat{M}-1)/2))$

$$A_t = A = \begin{bmatrix} (-1)^{2q}C_q^1, \ldots, (-1)^{2q-k+1}C_q^k, \ldots, (-1)^{q+2}C_q^{q-1}, & \vdots & (-1)^{q+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\vdots\cdots & \cdots\cdots\cdots \\ \mathbf{I}_{q-1}, & \vdots & \mathbf{0}_{(q-1)\times 1} \end{bmatrix} \otimes \mathbf{I}_N \tag{41}$$

corresponds to the most overdetermined filter in the chosen constant bandwidth ($\hat{M}$) adaptive algorithm.

III. Recursive Adaptive Algorithm for Rapidly Changing (Jumping) Systems

A. Changing Points Detection for Tracking Rapidly Changing Systems

There are two methods to adjust the RLS algorithm (3)–(6) when it is used for tracking a rapidly changing system. One method is to adaptively adjust the forgetting factor λ_t at the rapidly changing points while keeping its nominal values at other locations [3]. The other is to increase the estimation covariance matrix P_t or P_{t-1} at the locations of jumping points so that the filtering gain can be increased significantly to track the rapidly changing components [11]. When using either method, the jumping points need to be known *a priori* and this commonly is unrealistic in practice. Therefore, a recursive parameter change detection algorithm is required to identify the locations of jumping points online. Once a jumping change is detected, the above RLS algorithm with changing the forgetting factor or the covariance matrix at the detected jumping point can be adopted.

Some recursive change detection algorithms have been developed in [15]–[18]. An attractive method is the one used by Trigg and Leach (T & L) [15]. In this method, two filtering signals gained from the prediction error signal ε_t are used

$$\varepsilon_t^o = (1-\gamma)\varepsilon_{t-1}^o + \gamma\varepsilon_t \tag{47}$$

$$\varepsilon_t^a = (1-\gamma)\varepsilon_{t-1}^a + \gamma|\varepsilon_t| \tag{48}$$

where $|\cdot|$ denotes absolute value and γ takes a very small positive value (commonly, $0.005 \le \gamma \le 0.05$). The T & L detection signal is defined as [15]

$$d_t = \frac{\varepsilon_t^o}{\varepsilon_t^a}. \tag{49}$$

According to the central limiting theorem, d_t is asymptotically Gaussian distributed. It is shown in [15] and [16] that, for small γ, d_t is a zero-mean signal with variance approximately as

$$\mathrm{Var}(d_t) = E(d_t^2) \approx \frac{\pi}{2}\frac{\gamma}{2-\gamma}. \tag{50}$$

Assume a detection threshold is r. When $|d_t| > r$ at time index t, a parameter change is considered to have happened [11]. Denotes f_p as the false alarm probability of detection. According to the Chebyshev's inequality $p(|d_t| \ge r) \le (E(d_t^2)/r^2)$, the detection threshold r can be chosen as

$$r \approx \sqrt{\frac{\pi}{2}\frac{\gamma}{2-\gamma}\frac{1}{f_p}}. \tag{51}$$

From the above, we see that there exists a tradeoff between the false alarm probability and the detection probability of the T & L detector. In the following subsection, we will develop a wavelet domain change detection algorithm which can achieve much higher detection probability when using the same false alarm probability as the one used by the T & L detector (i.e., using the same detection threshold).

B. Wavelet Jump Detector for Abrupt Change Detection

A dyadic wavelet transform (DWT) of function f_t at time t and scale j ($1 \le j \le J$ and J is the maximum decomposition scale) can be implemented via a set of discrete digital filters as follows [21], [22] (without loss of generality, the initial conditions are assumed as $f_0 = f_{-1} = \cdots = f_{(-2^j+2)} = 0$):

$$z_t(j) = W_{2^j} f_t = \sum_{k=-2^j+2}^{2^j-1} h_k(j) f_{t-k} \tag{52}$$

where $h_k(j)(-2^j+2 \le k \le 2^j-1)$ is the equivalent digital filter of DWT in the jth scale. It is odd symmetrical with respect to $1/2$ [i.e., $h_k(j) = -h_{-k+1}(j)$] and its region of support (the number of nonzero coefficients) at scale j is $\sum_{k=1}^{j} 2^k = 2^{j+1} - 2$ [21], [23]. Unfortunately, $h_k(j)$ for each scale j is a noncausal filter. Hence, (52) cannot be used for calculating DWT online causally. For detecting the changing points online, a *recursive* DWT algorithm is desired that can calculate the DWT coefficients in different scales once a new observation data sample is available. Assuming the available data sequence at time t is $Y_f^t = [f_1, \ldots, f_t]$, we consider an extended sequence $\overline{Y}_f^t = [\overline{f}_1, \ldots, \overline{f}_{t+2^j-2}]$ by an even-symmetric extension of Y_f^t as

$$\overline{f}_k = \begin{cases} f_k, & 1 \le k \le t \\ f_{2t-k}, & t+1 \le k \le t+2^j-2. \end{cases} \tag{53}$$

Denote the DWT of $\overline{f}_t$ at time t and scale j as $\overline{z}_t(j)$, i.e., $\overline{z}_t = W_{2^j}\overline{f}_t$. We will show in the following theorem that $\overline{z}_t(j)$ can be online calculated from Y_f^t using an equivalent causal filter. Obviously, if a jumping point of f_t occurring at time t can be detected from $z_t(j)$, it also can be detected from $\overline{z}_t(j)$ since the even-symmetric extension signal $\overline{f}_t$ does not alter the jumping point features of the original signal f_t. Since $\overline{z}_t(j)$ can be recursively calculated online, it should be used instead of $z_t(j)$ for online detecting the abrupt changes of signal f_t.

Theorem 1:: Let the even-symmetric extension sequence $\overline{Y}_f^t$ be given in (53), the DWT $\overline{z}_t(j)$ of $\overline{f}_t$ at time t and scale j can be calculated from Y_f^t as

$$\overline{z}_t(j) = W_{2^j}\overline{f}_t = \overline{W}_{2^j} f_t = \sum_{k=0}^{2^j} h_k^e(j) f_{t-k} \tag{54}$$

where $h_k^e(j)$, $k = 0, \ldots, 2^j$ is an equivalent causal filter for scale j defined as

$$h_k^e(j) = \begin{cases} h_0(j), & k = 0 \\ h_k(j) + h_{-k}(j), & 1 \le k \le 2^j - 2 \\ h_{2^j-1}(j), & k = 2^j - 1. \end{cases} \tag{55}$$

Furthermore, $\overline{z}_t(j)$ can be recursively calculated as

$$\Omega^{(t)}(j) = \tilde{\Omega}^{(t-1)}(j) + f_t H^e(j) \tag{56}$$

$$\overline{z}_t(j) = \Omega_1^{(t)}(j). \tag{57}$$

Here, $\Omega^{(t)}(j)$, $\tilde{\Omega}^{(t-1)}(j)$, and $H^e(j)$ are column vectors of size $2^j \times 1$ which are defined as follows:

$$\Omega^{(t)}(j) = \left[\overline{Z}_t^{(t)}(j), \ldots, \overline{Z}_{t+2^j-1}^{(t)}(j)\right]' \tag{58}$$

$$\tilde{\Omega}^{(t-1)}(j) = \left[\overline{Z}_{t+1}^{(t-1)}(j), \ldots, \overline{Z}_{t+2^j-1}^{(t-1)}(j), 0\right]' \tag{59}$$

$$H^e(j) = [h_0^e(j), \ldots, h_{2^j-1}^e(j)]'. \tag{60}$$

To save space, the proof is omitted here but can be found in [24, pp. 91–92].

We name (54) as a causal DWT of the original signal f_t. The above theorem shows that $\overline{z}_t(j)$ at time t can be recursively calculated online using (56) and (57) once a new data sample f_t arrives. The scale filters $\{h_k^e(j), j = 1, \ldots J\}$ can be calculated and stored in advance before performing the recursive causal DWT. In (52), $z_k(j) = h_k(j)$ when $f_k = \delta_k$. Thus, the filter coefficients of $\{h_k(j), j = 1, \ldots J\}$ can be calculated by applying the iterative DWT algorithm introduced in [21] to a δ_k input signal. Once $h_k(j)$ is obtained, $h_k^e(j)$ can be easily calculated by (55). The filter coefficients of $h_k^e(j)$ for scale j =1–4 can be found in [24].

Now, consider multiscale product of the first K scale sequences in wavelet domain at time index t

$$\xi_t^K = \prod_{j=1}^{K} \overline{z}_t(j). \tag{61}$$

Since the wavelet used for DWT in this paper is chosen as the first-order derivative of a smooth function (a cubic spline function, see [21]), the DWT $\overline{z}_t(j)$ can be interpreted as the derivative of local smooth (average) of $\overline{f}_t$ at scale j [22]. Hence, if f_t (thus $\overline{f}_t$) has some singular points (especially jumping points), $\overline{z}_t(j)$ will appear as modulus maxima at these locations. More importantly, the amplitude of noise modulus maxima will decrease from small scales to large scales while the amplitude of signal modulus maxima will increase from small scales to large scales in the wavelet domain [21], [22]. Therefore, multiscale product sequence ξ_t^K sharpens and enhances the modulus maxima dominated by signal edges and at the same time suppresses the modulus maxima dominated by noises. It has been further shown in that the probability density function (PDF) of a multiscale product sequence is heavy tailed compared with that of a Gaussian distributed one with the same variance. Employing these characteristics, a DWT multiscale product sequence of an existing detection signal (for example, obtained from the T & L detector) can be used as a new detection signal. It will enhance the components representing possible abrupt changes in the original detection signal and thus a larger detection threshold can be used, which will lead to a smaller false alarm probability. At the same time, it will suppress the noise interference components in the original detection signal, which will decrease the miss alarm probability and thus increase the detection probability. Motivated by the above discussion, a new wavelet jump detector is now proposed for online change detection.

Denote $\tilde{z}_t(j)$ as the causal DWT of the T & L detection signal d_t (49) at time t and scale j. That is

$$\tilde{z}_t(j) = \overline{W}_{2^j}\, d_t \tag{62}$$

which can be recursively calculated as (56) and (57). The multiscale product signal of the first K scales can be calculated as

$$\tilde{\xi}_t^K = \prod_{j=1}^{K} \tilde{z}_t(j). \tag{63}$$

Define a new (multiscale product) detection signal ζ_t by filtering $\tilde{\xi}_t^K$ as follows:

$$\zeta_t = (1-\eta)\zeta_{t-1} + \eta\tilde{\xi}_t^K \tag{64}$$

where η is an exponential smoothing factor which commonly takes a value in the range [0.05, 0.13]. Although ξ_t^K is heavy-tailed non-Gaussian distributed, ζ_t obtained above is a Gaussian distributed signal according to the central limiting theorem. Now, a new wavelet detector can be formed as

$$\tilde{d}_t = \sqrt{\frac{\mathrm{Var}(d_t)}{\mathrm{Var}(\zeta_t)}}\, \zeta_t. \tag{65}$$

Obviously, if ζ_t is a Gaussian distributed signal, $\tilde{d}_t$ is also a Gaussian distributed signal whose variance is the same as that of d_t. However, if d_t has some local maxima (minima) corresponding to the abrupt changes of the original signal, these local maxima (minima) will be enlarged and sharpened in $\tilde{d}_t$. This characteristics can be employed to provide a more robust and accurate identification of the possible abrupt changes. Thus, if we choose the detection threshold $\tilde{r}$ of the wavelet detection signal $\tilde{d}_t$ equal to the threshold r of the T & L detection signal d_t, we can achieve much higher detection probability. To get the accurate estimation of the variance of ζ_t ($\mathrm{Var}(\zeta_t)$) in (65), a robust method will be proposed shortly based on an empirical equation. We first postulate two transformations as follows.

First, the log of the ratio of the variance of d_t to the variance of $\tilde{\xi}_t^K$ satisfies an order m polynomial function of γ

$$\log\left(\frac{\mathrm{Var}(d_t)}{\mathrm{Var}(\tilde{\xi}_t^K)}\right) = \sum_{i=0}^{m} \nu_i \gamma^{m-i} \tag{66}$$

where $\nu_0 = 1$. Taking log operation can compress the dynamic range of $\mathrm{Var}(d_t)/\mathrm{Var}(\tilde{\xi}_t^K)$ and hence an order m polynomial function is sufficient to approximate the log function in the left side of (66).

Second, the log of the ratio of the variance of ζ_t to the variance of $\tilde{\xi}_t^K$ satisfies a log function $\log(\eta/(2-\eta))$ plus a linear function $\kappa\eta + \tau$ of η

$$\log\left(\frac{\mathrm{Var}(\zeta_t)}{\mathrm{Var}(\tilde{\xi}_t^K)}\right) = \log\left(\frac{\eta}{2-\eta}\right) + (\kappa\eta + \tau). \tag{67}$$

If $\tilde{\xi}_t^K$ is a white-noise sequence, the variance ratio $\mathrm{Var}(\zeta_t)/\mathrm{Var}(\tilde{\xi}_t^K)$ is $\eta/(2-\eta)$ and thus $\log(\mathrm{Var}(\zeta_t)/\mathrm{Var}(\tilde{\xi}_t^K)) = \log(\eta/(2-\eta))$ [obtained by taking variance on both sides of (64)]. Although wavelet transform can decorrelate a signal to some degree, $\tilde{z}_t(j)$ in (62) is a correlated signal since d_t is highly correlated. Therefore, $\tilde{\xi}_t^K$ in (63) is a correlated signal. A linear function of η is added in (67) to account for the bias produced by the correlation of signal $\tilde{\xi}_t^K$.

Combining (66) and (67), the ratio of the variance of new wavelet multiscale product detection signal to the variance of the T & L detection signal (R–W–TL) is defined as

$$\begin{aligned}\mathrm{R{-}W{-}TL} &= \frac{\mathrm{Var}(\zeta_t)}{\mathrm{Var}(d_t)} = \frac{\mathrm{Var}(\zeta_t)/\mathrm{Var}(\tilde{\xi}_t^K)}{\mathrm{Var}(d_t)/\mathrm{Var}(\tilde{\xi}_t^K)} \\ &= \exp\left(\log\left(\frac{\eta}{2-\eta}\right) + \kappa\eta + \tau - \sum_{i=0}^{m}\nu_i\gamma^{m-i}\right) \quad (68)\end{aligned}$$

and the wavelet detector (65) using the empirical variance ratio estimation (68) can be represented as $\check{d}_t$

$$\check{d}_t = \frac{\zeta_t}{\sqrt{\mathrm{R{-}W{-}TL}}} \quad (69)$$

where the values of κ, τ, and $\{\nu_i\}_{i=1,\ldots,m}$ can be estimated by applying least-squares method to experimental data through Monte Carlo simulations. Recommendation values are $\kappa = -5.1043$, $\tau = 1.0026$, $m = 9$, and $\nu = \{1.272\times10^{11}, -5.9549\times10^{10}, 1.179\times10^{10}, -1.2877\times10^9, 8.4984\times10^7, -3.5005\times10^6, 9.0861\times10^4, -1.5622\times10^3, 2.5120\times10^1\}$. Extensive simulations have verified that the empirical equations (68) and (69) are effective and produce quite accurate results (see [11] for more details).

C. Selectively Tracking of Rapidly Changing Systems Using Wavelet Jump Detectors

For a time-varying system, different branches of system parameters are not always subject to abrupt changes simultaneously when a jump occurs. When modifying the matrix P_{t-1} or P_t [in (5) or (6)] with Δ_t, it is common to select Δ_t as a diagonal matrix where each diagonal element reflects the change of the corresponding parameter branch. When one or several branches have changed rapidly at a specific time t, the corresponding elements in Δ_t should be increased while the remaining elements should keep unchanged [11]. This requires that the jump detector cannot only identify the locations where the jumps have happened but also determine the branches producing these jumps.

It is well known that the *priori* prediction error signal can be used to construct the jump detector [15], [16]. However, this detector (named as *prediction detector*) only can determine where a jump happens for a time-varying system. To judge which branches this jump is produced by, a set of jump detectors can be constructed directly from the estimated filtering gains (named as *gain detectors*). Combining the *prediction detector* with *gain detectors*, a new *selective wavelet detector* is proposed in the following, which can determine not only the locations of jumping points but also the branches that have produced the jumps.

Assume a wavelet detector $\check{d}_t^{\varepsilon}$ (*prediction detector*) is obtained from the priori prediction error signal ε_t [see (4)]. Assume other N wavelet detectors $\check{d}_t^1, \ldots, \check{d}_t^N$ (*gain detectors*) are obtained from the estimated filtering gains $G_t(1), \ldots, G_t(N)$ [see (5)], respectively. Without loss of generality, we assume here a system jumping change at a specific time is produced by an abrupt change of only one parameter branch (the case of several parameter branches changing at the same time is a simple extension). The proposed *selective wavelet detector* uses both the *prediction detector* and the *gain detectors* for parameter change detection. More explicitly, an abrupt change is considered to be detected at the ith $(i = 1, \ldots, N)$ parameter branch at time t, if

$$|\check{d}_t^{\epsilon}| > \tilde{r} \quad \text{and} \quad |\check{d}_t^{i}| > \tilde{r} \quad (70)$$

where the detection threshold $\tilde{r}$ is set as $\tilde{r} = r$ and, thus, can be determined by (51) in advance. At this time t, we set $\Delta_t = [\mathbf{0}]_{N\times N} except \Delta_t(i, i) \neq 0$. To determine the value of $\Delta_t(i, i)$, consider the following equality (see [25, Appendix 3.D]):

$$\Xi_t = \lambda\Xi_{t-1} + \frac{\lambda}{\lambda + \varphi_t^T P'_{t-1}\varphi_t}\varepsilon_t^2 \quad (71)$$

$$= \lambda\Xi_{t-1} + (1 - \varphi_t^T P'_t\varphi_t)\varepsilon_t^2 \quad (72)$$

where Ξ_t is the sum of the *a priori* prediction mean-square errors at time t. When λ is close to 1, we can take $(1-\lambda)\Xi_t = \sigma^2$ and σ^2 is the measurement noise variance [25]. Modify the matrix P_{t-1} as (setting $\Delta_t = \Delta'_t$)

$$P'_{t-1} = P_{t-1} + \Delta'_t. \quad (73)$$

Substituting (73) into (71) gives

$$\frac{\sigma^2}{\varepsilon_t^2} = \frac{\lambda}{\lambda + \varphi_t^T P_{t-1}\varphi_t + \varphi_t^T\Delta'_t\varphi_t}. \quad (74)$$

Thus, $\Delta'_t(i, i)$ can be estimated as

$$\hat{\Delta}'_t(i, i) = \frac{\lambda\left(\frac{\sigma^2}{\varepsilon_t^2} - 1\right) - \varphi_t^T P_{t-1}\varphi_t}{\varphi_t^2(i)}. \quad (75)$$

Similarly, modify the matrix P_t as (setting $\Delta_t = \Delta''_t$)

$$P'_t = P_t + \Delta''_t. \quad (76)$$

Substituting (76) into (72), and $\Delta''_t(i, i)$ can be estimated as

$$\hat{\Delta}''_t(i, i) = \frac{\frac{\lambda}{\lambda+\varphi_t^T P_{t-1}\varphi_t} - \frac{\sigma^2}{\varepsilon_t^2}}{\varphi_t^2(i)}. \quad (77)$$

In a summary, we list the complete RLS algorithm using estimation covariance matrix modification and selective wavelet detector (abbreviated as RLS-MSWD) at time t as follows.

- **(a) RLS algorithm**
  Using (3)–(6) to calculate $\hat{\theta}_t$, ε_t, G_t and P_t;
- **(b) Selective wavelet detector for change detection**
  — (b1) From ε_t, calculating (47)–(49), (62) [implemented with (56) and (57)], (63), (64), (68), and (69) to get the *predictive detector* $\check{d}_t^{\varepsilon}$,
  — (b2) For i = 1: N {Using $G_t(i)$ instead of ε_t in (47) and (48), calculating equations as in (b1) to get the ith *gain detector* $\check{d}_t^{i}$} End
  — (b3) Using (70) to detect if a jumping change has happened. If yes, determine which parameter branch pro-

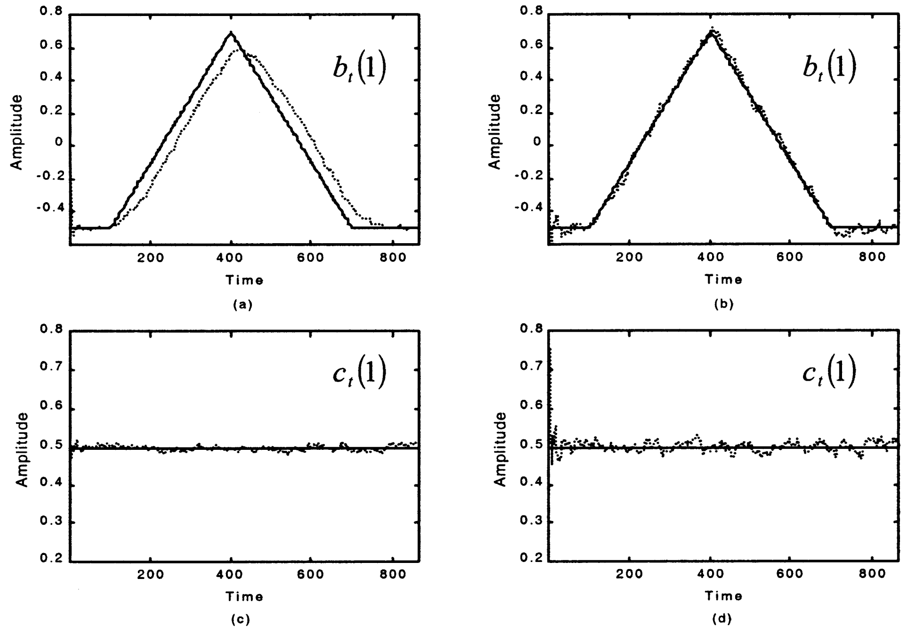

Fig. 1. Track a ramp time-varying system by (a), (c) the RLS algorithm and (b), (d) the T-RLS algorithm. (Solid lines represent true values and dotted lines represent estimation results.)

duces this change and go to (c); Otherwise $t = t + 1$ and go to (a);

- **(c) Estimation covariance matrix modification**
  Modify P_{t-1} or P_t as (73) and (75) or (76) and (77). $t = t + 1$ and go to (a).

The RLS-MSWD algorithm above is performed recursively to track a rapidly changing system. If we use a T-RLS or a MT-RLS algorithm instead of the RLS algorithm in step (a) and keep steps (b) and (c) unchanged, the extended algorithm can identify and track an arbitrary time-varying system with slowly, fast and rapidly varying components simultaneously. A typical example will be used in the next section to illustrate this promising tracking method.

IV. Simulation Results

The T-RLS and RLS algorithm are used for tracking a ramp function in Fig. 1. The system simulated is an $ARX(1, 1)$ model

$$y_t = b_t(1)y_{t-1} + c_t(1)u_{t-1} + u_t + e_t$$

where $b_t(1)$ is a ramp-like function, $c_t(1)$ is a constant and u_t is assumed a random ± 1 pulse input. The measurement noise variance of e_t takes 0.01. The RLS algorithm is used for the identification of the above $ARX(1, 1)$ system assuming that the system parameters are modeled with an order one random walk model. The T-RLS algorithm is used for the identification of the same system assuming that the system parameters are modeled with a second order deterministic trend model, where we take vector $\theta_t = [b_t(1), c_t(1), b_{t-1}(1), c_{t-1}(1)]'$, vector $\varphi_t = [y_{t-1}, u_{t-1}, 0, 0]'$, and matrix

$$A_t = A = \begin{bmatrix} 2\mathbf{I_2} & -\mathbf{I_2} \\ \mathbf{I_2} & \mathbf{0} \end{bmatrix}$$

in recursive equations (7) and (8) [see also (41)]. The estimation results by the RLS algorithm (with a forgetting factor 0.97) are shown in Fig. 1(a) and (c). Obviously, the estimated ramp function lags the true values. The estimation results by the T-RLS algorithm with the same forgetting factor are shown in Fig. 1(b) and (d), where the estimated ramp function can track the true values very well.

In Figs. 2 and 3, we show that the performance of a MT-RLS algorithm. An $ARX(1, 1)$ system with $b_t(1)$ as sin-like function and $c_t(1)$ as a constant is simulated. Here we assume that the true order of the trend model used for modeling system parameters is unknown. The MDL values for assumed deterministic trend model order from 1 to 4 are calculated and shown in Fig. 2. The forgetting factor is taken as 0.98 and thus $L = 50$. We can see that the third-order trend model is the best matched model since its MDL sequence values are minimal among the four MDL sequences of different order trend models after the algorithm converges. If we select the number of parallel filters $\hat{M} = 3$, the three filters should take the trend models with orders 2, 3, and 4 respectively. The optimal posterior estimation by the MT-RLS algorithm using these three filters is shown in Fig. 3(a), and the estimation results by the T-RLS algorithm using order 1–4 trend model are shown in Fig. 3(c) respectively. Fig. 3(b) and (d) are the zoom-in parts of Fig. 3(a) and (c), respectively. Obviously, the adaptive filter taking order 1 or 2 trend model underestimates the true system parameters while the adaptive filter taking order 3 or 4 trend model overestimates the true system parameters. Thus, we can infer that the true trend model order is between 2–3 but closer to 3. Although we cannot estimate the true trend model order accurately, the optimal estimation of system parameters can still be obtained by the MT-RLS algorithm.

In Fig. 4, the proposed wavelet detector is compared with the T & L detector. Fig. 4(a) shows a stationary white Gaussian noisy signal which has three abrupt changes at the vicinity of

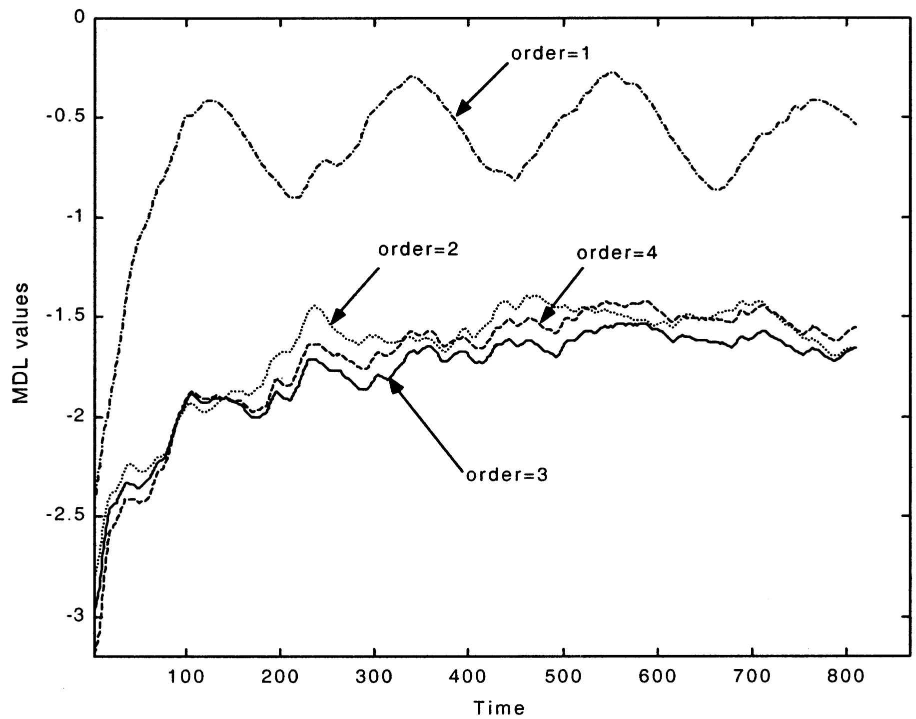

Fig. 2. MDL values of the T-RLS algorithm for different order trend models.

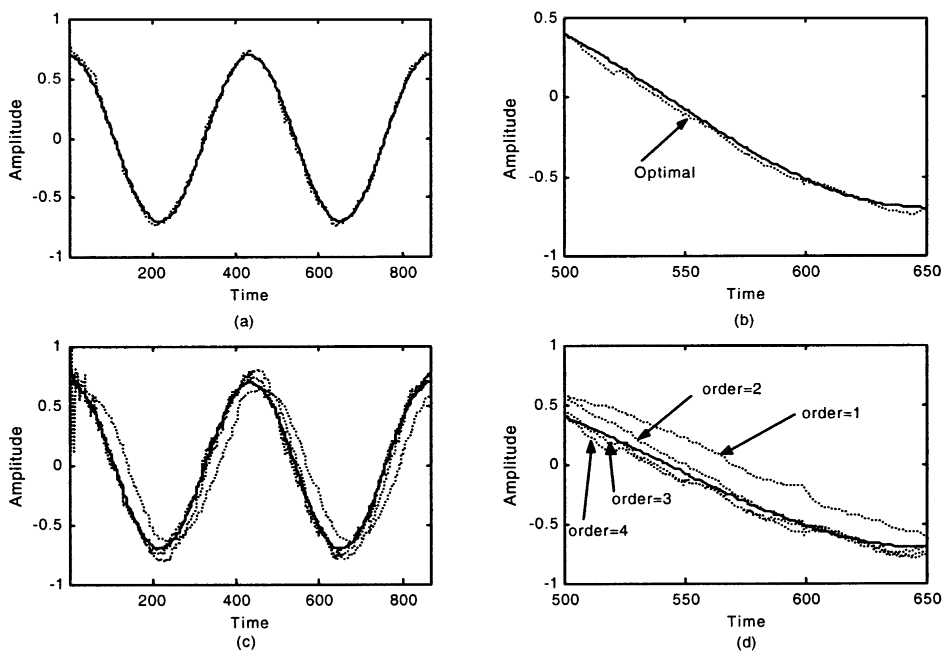

Fig. 3. Track a sin-like unknown order time-varying system by the MT-RLS algorithm. (a) Optimal tracking by the MT-RLS algorithm. (c) Tracking by the T-RLS algorithm with different order trend models. (b), (d) Zoom in of (a) and (c).

time locations 100, 700, and 1500, respectively. The amplitudes and shapes of these changes are shown in Fig. 4(b). In Fig. 4(c), the solid line represents the T & L detection signal and the dotted line represents the wavelet detection signal obtained using the theoretical R–W–TL. Fig. 4(d) shows the same trace as the one represented by the dotted line in Fig. 4(c), i.e., wavelet detection signal obtained using the theoretical R–W–TL (wavelet decomposition scale number $K = 3$). Comparing the wavelet detection signal with the T & L detection signal in Fig. 4(c), the former can provide sharper and more accurate indication of the abrupt changing points and this is very important for detecting small amplitude or/and concentrated abrupt changes.

Next, an ARX(2, 1) system

$$y_t = b_t(1)y_{t-1} + b_t(2)y_{t-2} + c_t(1)u_{t-1} + u_t + e_t$$

is used to verify the performance of the proposed abrupt change tracking algorithm. Here, the system parameters $b_t(1)$ and $b_t(2)$

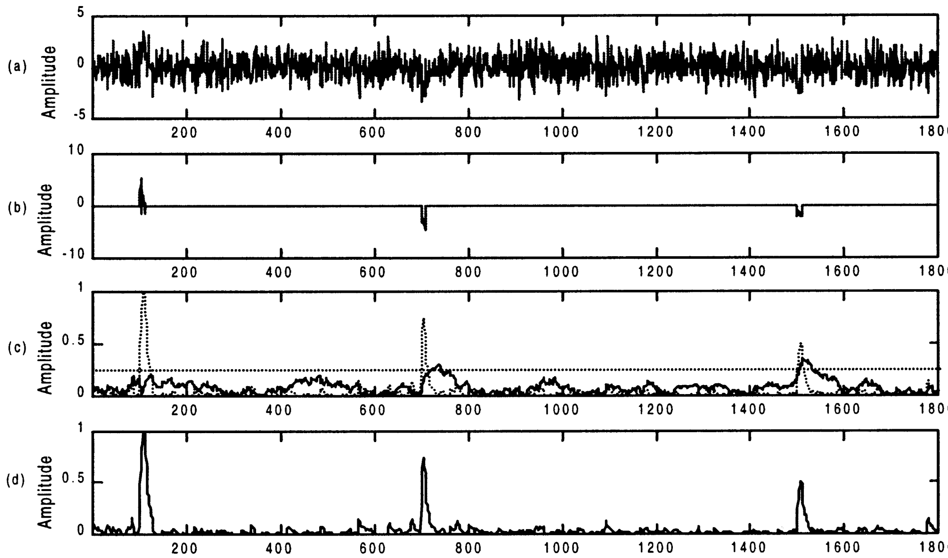

Fig. 4. Comparison of the wavelet detector with the T & L detector. (a) Original signal. (b) Abrupt change locations and shapes. (c) Solid line: T & L detection signal; dotted line: wavelet detection signal obtained from empirical equations. (d) Wavelet detection signal obtained from empirical equations.

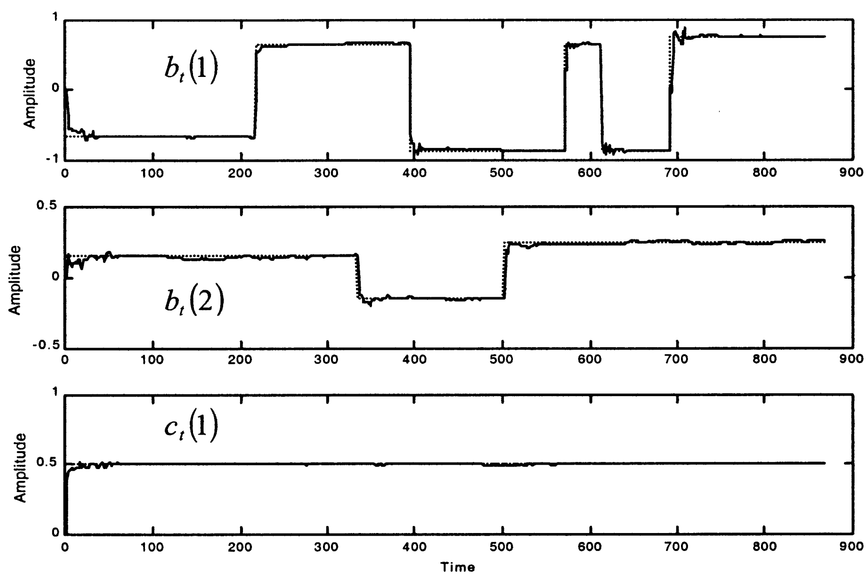

Fig. 5. An ARX(2, 1) abruptly changing system identification by the proposed RLS-MSWD algorithm. (Solid lines represent tracking results and dotted lines represent true values.)

are both with abrupt changes and $c_t(1)$ is constant as shown in Fig. 5. The identification results by the proposed RLS-MSWD are shown in Fig. 5, where $\gamma = 0.02$, $\eta = 0.10$, $K = 3$, and the empirical formulas (68) and (69) are used for producing the wavelet detectors. It can be seen that the estimation coincides with the true parameter values very well. For comparison, identification results by the RLS algorithm using T & L detector (abbreviated as RLS-MTLD) are shown in Fig. 6. Since the T & L detector is not so sensitive to the abrupt changes as the selective wavelet detector, the identification results by the RLS-MTLD method can not track abrupt changes with small amplitude [see $b_t(2)$ between time index 330 and 500] and concentrated abrupt changes [see $b_t(1)$ between time index 570 and 620] in Fig. 6. Moreover, from Fig. 5 we can see that the proposed RLS-MSWD method can selectively track the abrupt changes of different parameter branches; while the estimation of different parameter branches by the RLS-MTLD method in Fig. 6 are disturbed and affected by each other.

Finally, an example of tracking a general system with slowly, fast(with ramp and quadratic trend), and abruptly changing components is illustrated in the following. The system to be simulated is an ARX(1, 1) system whose parameter $b_t(1)$ is shown by the dotted line in Fig. 7 and $c_t(1)$ is a constant. The MT-RLS algorithm using both estimation covariance matrix modification and selective wavelet detector is applied for system identification and parameter tracking. The algorithm parameters are selected as $\lambda = 0.98$, $\hat{M} = 3$, $\gamma = 0.01$, $\eta = 0.12$, and $K = 3$. The estimation result of $b_t(1)$ is shown in Fig. 7(a) which coincides with the true values very well. The MT-RLS alone ($\lambda = 0.98$, $\hat{M} = 3$) is used for system

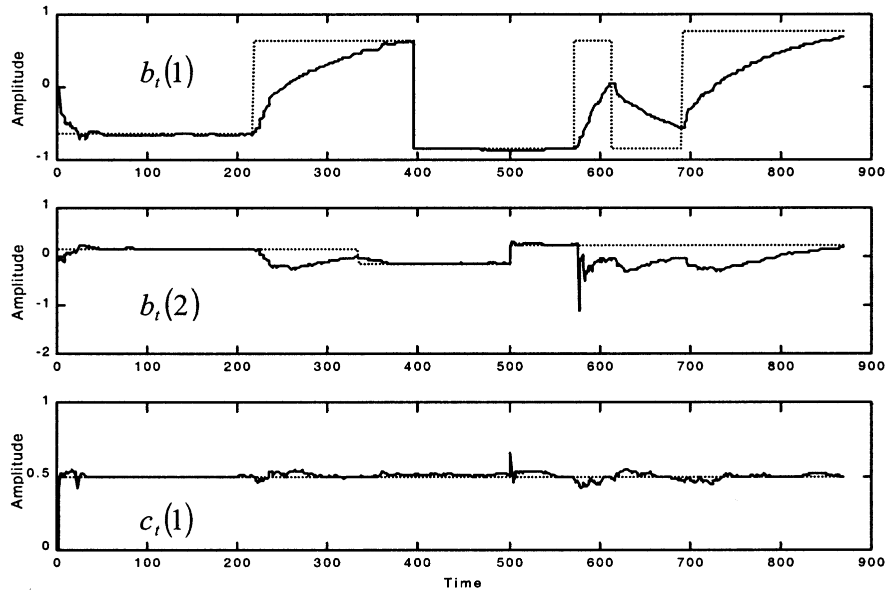

Fig. 6. An ARX(2, 1) abruptly changing system identification by the RLS-MTLD algorithm. (Solid lines represent tracking results and dotted lines represent true values.)

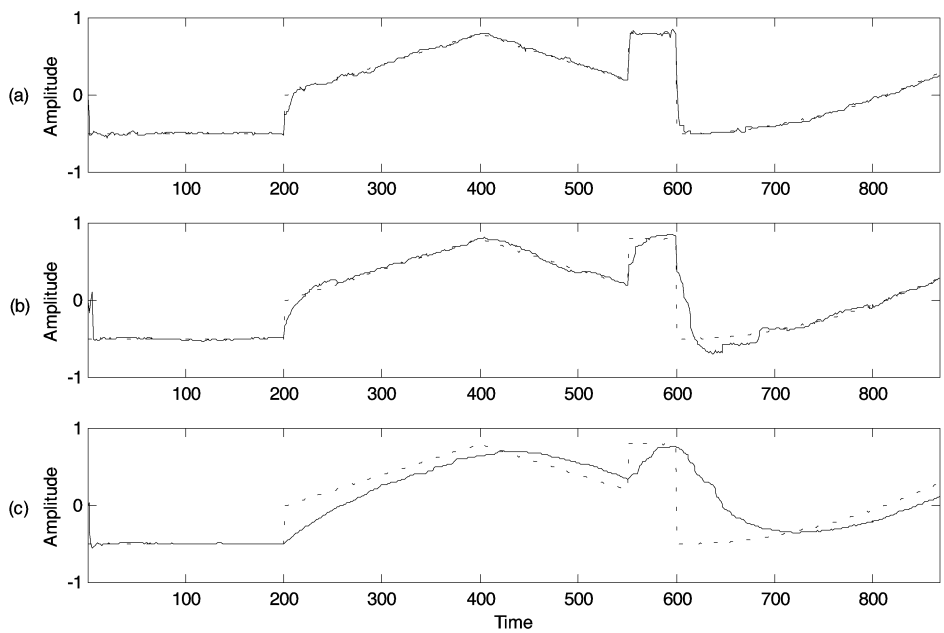

Fig. 7. Track a general system with slowly varying, fast varying, and rapidly changing components by the: (a) MT-RLS algorithm with covariance modification using selective wavelet detector; (b) MT-RLS algorithm; (c) RLS algorithm. (Solid lines represent tracking results and dotted lines represent true values.)

parameter estimation and the result is shown in Fig. 7(b). Although the trend changes of time-varying parameters are estimated well, this method cannot track the abrupt changes sufficiently. The estimation by the RLS method ($\lambda = 0.98$) is shown in Fig. 7(c) which fails to track both the fast and the abruptly changing components.

V. CONCLUSION

In this paper, the problem of tracking fast and abruptly changing systems has been tackled. A T-RLS algorithm has been proposed to track fast changing parameters with local and global trends. For an unknown fast varying system, a MT-RLS algorithm has been developed to optimally estimate the system parameters through Bayesian posterior combination of multiple adaptive filter outputs. To track an abruptly changing system, a new online wavelet detector has been proposed which is computationally simpler and can achieve much higher detection probability than commonly used abrupt detection methods. Selectively tracking the rapidly changing parameter branches via estimation covariance modification at the jumping points has been rigorously discussed. Both the jumping locations and increment values of covariance matrix for the detected parameter branches can be determined. Combining the proposed MT-RLS algorithm with the covariance modification method using wavelet detectors, slowly, fast, and abruptly changing components of a general time-varying system all can be tracked well.

REFERENCES

[1] P. E. Wellstead and M. B. Zarrop, *Self-Tuning Systems: Control and Signal Processing*. New York: Wiley, 1991.
[2] C. K. Chui and G. Chen, *Kalman Filtering With Real-Time Applications*. Berlin, Germany: Springer-Verlag, 1999.
[3] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self-tuning regulators with variable forgetting factor," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.
[4] D. G. Childers, J. C. Principle, and Y. T. Ting, "Adaptive WRLS-VFF for speech analysis," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 209–213, May 1995.
[5] G. Kitagawa and W. Gersch, "A smoothness priors-state space modeling of time-series with trend and seasonality," *J. Amer. Stat. Assoc.*, vol. 79, pp. 378–389, 1984.
[6] ——, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time-series," *IEEE Trans. Automat. Contr.*, vol. AC-30, no. 1, pp. 48–56, 1985.
[7] A. Benvensite, "Design of adaptive algorithms for the tracking of time-varying systems," *Int. J. Adapt. Control Signal Process.*, vol. 1, pp. 3–29, 1987.
[8] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[9] E. Emre and J. Seo, "A unifying approach to multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 520–528, July 1989.
[10] G. A. Williamson, "Tracking random walk systems with vector space adaptive filters," *IEEE Trans. Circuits Syst. II*, vol. 42, pp. 543–547, Aug. 1995.
[11] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–21, 1990.
[12] M. J. Chen and J. P. Norton, "Estimation technique for tracking rapid parameter changes," *Int. J. Control*, vol. 45, no. 4, pp. 1387–1398, 1987.
[13] P. Andersson, "Adaptive forgetting in recursive identification through multiple models," *Int. J. Control*, vol. 42, no. 5, pp. 1175–1193, 1985.
[14] M. Niedzwiecki, "Identification of time-varying systems with abrupt parameter changes," *Automatica*, vol. 30, no. 3, pp. 447–459, 1994.
[15] D. W. Trigg and A. G. Leach, "Exponential smoothing with an adaptive response rate," *Oper. Res.*, vol. 18, pp. 53–59, 1967.
[16] B. Carlsson, "Digital differentiating filters and model based fault detection," Ph.D. dissertation, Dept. Automat. Control, Uppsala Univ., Uppsala, Sweden, 1989.
[17] T. Hägglund, "New estimation techniques for adaptive control," Ph.D. dissertation, Dept. Automat. Control, Lund Inst. Technol., Lund, Sweden, 1983.
[18] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[19] J. Holst and N. K. Poulsen, "Self tuning control of plant with abrupt changes," in *Proc. IFAC 9th Triennial World Congr.*, Budapest, Hungary, 1984, pp. 923–928.
[20] M. Niedzwiecki, "Identification of Nonstationary Stochastic systems using parallel estimation schemes," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 329–334, Mar. 1990.
[21] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.
[22] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 617–643, Feb. 1992.
[23] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1043–1051, Mar. 1999.
[24] Y. J. Zheng, "Contributions to nonstationary adaptive signal processing and time-varying system identification," Ph.D. dissertation, School Elect. Electron. Eng., Nanyang Technol. Univ., Singapore, 2002.
[25] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[26] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1997.

Yuanjin Zheng (M'03) received the B.Eng. degree with first class honors and the M.Eng. degree with best graduate student honors from Xi'an Jiaotong University, Xi'an, China, in 1993 and 1996, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2001.

From July 1996 to April 1998, he was with the National Key Lab of Optical Communication Technology, University of Electronic Science and Technology of China, as a Research Assistant. In March 2001, he joined the Institute of Microelectronics, Singapore, as a Senior Research Engineer. His research is currently focused on communication system design and performance evaluation, analog and digital IC design, and DSP algorithm design and implementation. He has authored over 15 papers published in international journals and conference proceedings. He is the holder of four U.S. patents pending.

Zhiping Lin (M'86–SM'00) received the Ph.D. degree in information engineering from the University of Cambridge, Cambridge, U.K., in 1987.

He was with the University of Calgary, Calgary, AB, Canada, from 1987 to 1988, with Shantou University, China, from 1988 to 1993, and with DSO National Laboratories, Singapore, from 1993 to 1999. Since 1999, he has been an Associate Professor in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include multidimensional systems and signal processing, wavelets, and biomedical signal processing.

Dr. Lin is currently on the Editorial Board of *Multidimensional Systems and Signal Processing* and is an Associate Editor of *Circuits, Systems, and Signal Processing*.