ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

Various Recursive Mixed L2-Linfty Algorithms for Linear-in-the-Parameters Models

Qi Zhu
Department of Computer Science, University of Houston – Victoria, TX, 77901

Abstract - In this paper, various Recursive Mixed L2-Linfty (RML) learning algorithms are developed by choosing different forgetting factor matrix function $\Lambda(\theta)$ for Linear-in-the-Parameters (LIP) models, including Projection, Recursive Mixed L2-Linfty, Recursive Mixed Mean L2-Linfty, λ-weighted Mixed L2-Linfty, instantaneous RML, and Batch RML algorithms. A few models are given to apply the proposed RML algorithms for system identification, and some simulations are carried out to show the algorithms' efficiency and effectiveness.

Index Terms: **Recursive Mixed L2-Linfty, System Identification, Parameter Estimation.**

I. INTRODUCTION

Least square algorithm is a standard approach to find the numerical values of a set of parameters to fit a function optimally or to characterize the statistical properties of estimates [2, 14]. Especially for digital signal processing applications, least square algorithm is widely used to establish the mathematical framework for the system [10], such as some areas of data communications [17, 18], signal control [7], radar systems [4], and seismology analysis [16]. Recursive parameter estimation methods are identification procedures that continually and recursively update the parameters of a process model on-line. Recursive least squares and on-line procedures are more useful when parameters are identified from recurring in time [5, 13], which can be used in a wide variety of real world applications when the model structure is well understood and input data becomes available at regular intervals of time, such as speech [15], vehicle mass estimation [], and structural damage assessment [3], and many more.

In [21], we have developed a High-order Mixed L2-Linfty Estimation for LIP models under noiseless and noisy data, based on that the data samples $\{(x_i, y_i)\}(i=1,2,...,N)$ are known before starting the learning process. This assumption is valid in many real identification situations where the whole set of sampling data can be collected before the learning can start. However, the sample data may only be obtained sequentially in real time in many identification fields and adaptive control situations. Therefore, we develop the recursive (on-line) version of the Mixed L2-Linfty learning algorithms for these situations.

In order to develop a recursive format, we assume that the number of data samples up to time t is $\rho(t)$. That means, up to time t, the sampling data is $\{(x_i, y_i)\}(i=1,2,...,\rho(t))$, where $\rho(t)$ is an increasing function on the non-negative integer set. If at time t, $\rho(t)=\rho(t-1)$, that means no new data coming in at time t; when $\rho(t)-\rho(t-1)>1$ that means more than one data samples coming in at time t. Under this assumption, we rearrange our positive semi-definite forgetting matrix $\Lambda(\theta)$ in [21] as $\Lambda(\theta)=diag\{\lambda_{t1},\lambda_{t2},...,\lambda_{t\rho(t)}\}$, the results obtained at time $t-1$ can be used in order to get the estimation at time t. In what follows, we begin to discuss the *Recursive Mixed L2-Linfty* (RML) learning algorithms.

In this paper, various RML learning algorithms are developed by choosing different kinds of time-variant symmetric positive semi-definite forgetting factor matrix function $\Lambda(\theta)$. When $\Lambda_t = diag\{0,...,0,1\}$, RML learning algorithm becomes the **projection learning algorithm** that minimizes the cost function $J_t=\frac{1}{2}e_t^2$; When $\Lambda_t=diag\{\lambda^{t-1},...,\lambda,1\}$, RML learning algorithm becomes the **λ-weighted RML learning algorithm** to minimize the commonly used cost function $J_t=\frac{1}{2}\sum_{i=1}^{t}\lambda^{t-i}e_i^2$ where $0<\lambda<1$; When $\Lambda_t = diag\{0,...,0,\underbrace{1}_{t-k-1},...,\underbrace{1}_{t}\}$, Mixed L2-Linfty learning algorithm becomes an **instantaneous k-order RML learning algorithm** that minimizes the adaptive cost function $J_t=\frac{1}{2}\sum_{i=t-k+1}^{t}e_i^2$; finally when $\Lambda_t = diag\{0,...,0,\underbrace{1}_{kt-k+1},...,\underbrace{1}_{kt}\}$, Mixed L2-Linfty learning algorithm becomes the **batch k-order RML learning algorithm** that minimizes the adaptive cost function $J_t=\frac{1}{2}\sum_{i=kt-k+1}^{kt}e_i^2$, and so forth.

This paper is organized as follows. In Section 2, the general form of recursive Mixed L2-Linfty (RML) learning algorithm is developed. Followed in Section 3, we derive the various specialized RML learning algorithms such as the projection learning algorithm, the Recursive Mixed L2-Linfty learning algorithm, the Recursive Mixed Mean L2-Linfty learning algorithm, the λ-weighted Recursive Mixed L2-Linfty learning algorithm, the instantaneous k-order RML learning algorithm, and the batch k-order RML learning algorithm. In Section 4, we discuss some applied models of RML

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

learning algorithm in the fields of identification and adaptive control. In Section 5, we supply simulations of several specialized RML learning algorithm in realistic industry applications. Finally, in Section 6, we conclude this paper and point out the future research directions.

II. GENERAL FORM OF RML LEARNING ALGORITHM

A. *Preliminaries*

The integral function $\rho(t)$ is defined as follows:

Definition 1 $\rho(t)$ is defined to be a function from the set of non-negative integers, it satisfies the following conditions:

(i) $0=\rho(0)\le\rho(1)\le\rho(2)\le\cdots\le\rho(t)\le\cdots$ (1)

(ii) The number of elements of the set $t \overset{\triangle}{=} \rho^{-1}(\rho(t))$ is uniformly bounded for all t. That is, there exists a positive integer N_ρ which

$$card(\rho^{-1}(\rho(t)))\le N_\rho \text{ for all } t \quad (2)$$

Where the function *card*() represents the *cardinality* of the set.

(iii)The *pseudo metric* $\rho(t_1,t_2)$ is defined as

$$\rho(t_1,t_2) \overset{\triangle}{=} |\rho(t_1)-\rho(t_2)| \quad (3)$$

Proposition 1 When $\rho(t)$ as defined in Definition 1, when $t\to\infty$, we have $\rho(t)\to\infty$. *Proof.* This can be very easily proved using the Definition 1.

B. *RML Learning Algorithm*

Using the function □ (t) defined in Definition 1, and the assumption that the number of data samples up to time t is □ (t). we can formulate again the matrix L_t, E_t defined in [21] have the new expressions as follows.

$$\Lambda_t=\Lambda_{\rho(t)}=\begin{bmatrix}\lambda_{t1} & 0 & \cdots & 0\\ 0 & \lambda_{t2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \lambda_{t\rho(t)}\end{bmatrix} \quad (4)$$

where $\lambda_{ti}\ge 0$ for $i=1,2,\cdots,\rho(t)$.

$$Y_{\rho(t)}=[y_1,y_2,\cdots,y_{\rho(t)}]^T \quad (5)$$

$$\Phi_{\rho(t)}=[\varphi_1,\varphi_2,\cdots,\varphi_{\rho(t)}]^T \quad (6)$$

$$E_{\rho(t)}=[e_1,e_2,\cdots,e_{\rho(t)}]^T \quad (7)$$

$$P_t=\Phi^T_{\rho(t)}\Lambda_t\Phi_{\rho(t)}=\sum_{i=1}^{\rho(t)}\lambda_{ti}\varphi_i\varphi_i^T \quad (8)$$

$$Q_t=\Phi^T_{\rho(t)}\Lambda_t Y_{\rho(t)}=\sum_{i=1}^{\rho(t)}\lambda_{ti}y_i\varphi_i \quad (9)$$

$$R_t=Y^T_{\rho(t)}\Lambda_t Y_{\rho(t)}=\sum_{i=1}^{\rho(t)}\lambda_{ti}y_i^2 \quad (10)$$

$$J_t=\frac{1}{2}\theta^T_{t-1}P_t\theta_{t-1}-\theta^T_{t-1}Q_t+\frac{1}{2}R_t \quad (11)$$

Furthermore, we have the **General Form of RML learning algorithm** in Eq. (12).

$$\theta_t=\theta_{t-1}+\frac{\alpha J_t}{\beta+(Q_t-P_t\theta_{t-1})^T(Q_t-P_t\theta_{t-1})}(Q_t-P_t\theta_{t-1}) \quad (12)$$

where $\beta>0, 0<\alpha<4, t=1,2,\ldots,\rho(t)$.

C. *Properties*

For RML learning algorithm, we can assure that all of the previous theorems in [21] hold in case Λ_t in Eq. (4) satisfies the condition that

$$\sum_{i=1}^{\rho(t)}\lambda_{ti}\le M_\lambda \quad (13)$$

for all t where M_λ is a positive constant number.

Theorem 1 For any given initial value θ_0, the vector sequence θ_t generated in (12) has the following properties:

(i)

$$|\theta_t-\theta^\star|\le|\theta_0-\theta^\star| \quad (14)$$

(ii)

$$\lim_{t\to\infty}\frac{\alpha J_t}{\sqrt{\beta+(Q_t-P_t\theta_{t-1})^T(Q_t-P_t\theta_{t-1})}}=0 \quad (15)$$

(iii)

$$\lim_{t\to\infty}|\theta_t-\theta_{t-s}|=0 \text{ for any finite s.} \quad (16)$$

Proof. If we introduce a Lyapunov function V_t as

$$V_t=\tilde{\theta}^T_t\tilde{\theta}_t=(\theta_t-\theta^\star)^T(\theta_t-\theta^\star)$$

Then Theorem 1 can be carried out with the similar way as Lemma 2.1 in [20]. □

Theorem 2 For any set of noiseless data samples, RML learning algorithm (12) globally converges decreasingly in the sense that

$$\lim_{t\to\infty}J_t=0 \quad (17)$$

Proof. A similar proof like Lemma 2.2 in [20] can carry out this theorem. □

From (4) - (12), we can see that when Λ_t is computed recursively. Some specialized RML learning algorithms can be obtained by selecting various choices of Λ_t.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

III. VARIOUS SPECIALIZED RML LEARNING ALGORITHMS

In this section we will derive several RML learning algorithms by choosing different kinds of time-variant symmetric non-negative definite matrix function Λ_t.

A. Projection Learning Algorithm

If we choose $\rho(t)=t$ and

$$\Lambda_t=\begin{bmatrix} 0_{t-1} & 0 \\ 0 & 1 \end{bmatrix} \quad (18)$$

where 0_{t-1} is the $(t-1)$th-order square zero matrix. In this case, RML learning algorithm in (12) can reduce to the 1st-order learning algorithm as.

$$\theta_t=\theta_{t-1}+\frac{\alpha' e_t\varphi_t}{\beta+e_t^2\varphi_t^T\varphi_t}(y_t-\varphi_t^T\theta_{t-1}) \quad (19)$$

where

$t=1,2,\cdots,\beta>0,0<\alpha'=\frac{1}{2}<2$,

and $e_t=y_t-\varphi_t^T\theta_{t-1}$.

Delete e_t^2 from both numerator and denominator from (19) and β is a any positive number, so we can obtain the Projection Learning Algorithm [1, 11] in adaptive control.

$$\theta_t=\theta_{t-1}+\frac{\alpha'\varphi_t}{\beta+\varphi_t^T\varphi_t}(y_t-\varphi_t^T\theta_{t-1}) \quad (20)$$

where

$t=1,2,\cdots,\beta>0,0<\alpha'=\frac{1}{2}<2$.

Thus, Projection learning algorithm is a special case of RML algorithm when we choose a set of specific parameters. The Projection learning algorithm minimizes the cost function $J_t=\frac{1}{2}e_t^2$.

B. Recursive Mixed L2-Linfty Learning Algorithm

The conventional recursive least squares (RLS) algorithm is a powerful learning algorithm in adaptive control. However, the algorithm has disadvantages that must be applied with the condition Φ to be of full rank, the convergence is very slow, and the computation is quite intensive. In this subsection, we propose the Recursive Mixed L2-Linfty learning algorithm, which is free of full rank and has much less computation.

Choose $\rho(t)=t$ and $\Lambda_t=I_t$ where I_t is the tth-order identity matrix, then the Eqs. of θ_t and J_t as.

$$\theta_t=\theta_{t-1}+\frac{\alpha J_t T}{\beta+(Q_t-P_t\theta_{t-1})^T T(Q_t-P_t\theta_{t-1})}(Q_t-P_t\theta_{t-1}) \quad (21)$$

$$J_t=\frac{1}{2}\theta_{t-1}^T P_t\theta_{t-1}-\theta_{t-1}^T Q_t+\frac{1}{2}R_t \quad (22)$$

where $\beta>0,0<\alpha<4,t=1,2,...,$, and T is chosen to be a symmetric positive definite matrix.

P_t, Q_t and R_t defined in (8) - (10) have the recursive computation formulas as

$$\begin{aligned} P_t&=P_{t-1}+\varphi_t\varphi_t^T \text{ with } P_0=0 \\ Q_t&=Q_{t-1}+y_t\varphi_t \text{ with } Q_0=0 \\ R_t&=R_{t-1}+y_t^2 \text{ with } R_0=0 \end{aligned} \quad (23)$$

This new Recursive Mixed L2-Linfty learning algorithm does not need to assume that Φ is of full rank and minimizes the cost function $J_t=\frac{1}{2}\sum_{i=1}^{t}e_i^2$, which is the same as the cost function of the conventional recursive least squares. Note that the matrix $\Lambda_t=I_t$ does not satisfy the condition in (13), so we must choose a small symmetric positive definite matrix T to avoid the burst phenomenon. This technique for the Recursive Mixed L2-Linfty is effective in practice of identification and control.

C. Recursive Mixed Mean L2-Linfty Learning Algorithm

In the preceding subsection, we have discussed the Recursive Dual Minimum learning algorithm, which is a powerful learning algorithm in identification. However, since the trace of $\Lambda_t=I_t$ is not uniformly bounded, P_t, Q_t and R_t may become very large when t increases. In this subsection, we will propose a Recursive Minimum Mean Squares learning algorithm.

Choosing $\rho(t)=t$ and $\Lambda_t=\frac{1}{t}I_t$, the Recursive Mixed Mean L2-Linfty learning algorithm is:

$$\theta_t=\theta_{t-1}+\frac{\alpha J_t}{\beta+(Q_t-P_t\theta_{t-1})^T(Q_t-P_t\theta_{t-1})}(Q_t-P_t\theta_{t-1}) \quad (24)$$

$$J_t=\frac{1}{2}\theta_{t-1}^T P_t\theta_{t-1}-\theta_{t-1}^T Q_t+\frac{1}{2}R_t \quad (25)$$

where $\beta>0,0<\alpha<4,t=1,2,....$

P_t, Q_t and R_t defined in (8) - (10) have the recursive computation formulas as

$$\begin{aligned} P_t&=\frac{t-1}{t}P_{t-1}+\frac{1}{t}\varphi_t\varphi_t^T \text{ with } P_0=0 \\ Q_t&=\frac{t-1}{t}Q_{t-1}+\frac{1}{t}y_t\varphi_t \text{ with } Q_0=0 \\ R_t&=\frac{t-1}{t}R_{t-1}+\frac{1}{t}y_t^2 \text{ with } R_0=0 \end{aligned} \quad (26)$$

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

The Recursive Mixed Mean L2-Linfty learning algorithm in (25, 26) minimizes the cost function of mean of squares of errors $J_t=\frac{1}{2t}\sum_{i=1}^{t}e_i^2\to 0$.

D. Recursive λ-weighted Mixed L2-Linfty Learning Algorithm

In [9, 12], the conventional recursive least squares algorithm with forgetting index λ (named as λ-RLS algorithm) is another powerful learning algorithm for adaptive control. However, the algorithm is applied with the condition that Φ must be of full rank and the computation is quite intensive. In this subsection, we propose a new Recursive λ-weighted Mixed L2-Linfty learning algorithm that is free of the full rank of Φ and has less computation compared to the conventional λ-RLS algorithm.

Choose $\rho(t)=t$ and Λ_t to be

$$\Lambda_t=\begin{bmatrix}\lambda^{t-1} & 0 & \cdots 0\\ 0 & \lambda^{t-2} & \cdots 0\\ 0 & 0 & \ddots 0\\ 0 & \cdots & 0\ 1\end{bmatrix} \tag{27}$$

where $0<\lambda<1$. Then the Recursive λ-weighted Mixed L2-Linfty learning algorithm included in (11),(12) and P_t, Q_t and R_t defined in (8) - (10) have the recursive computation formulas as

$$P_t=\lambda P_{t-1}+\varphi_t\varphi_t^T \text{ with } P_0=0$$
$$Q_t=\lambda Q_{t-1}+y_t\varphi_t \text{ with } Q_0=0 \tag{28}$$
$$R_t=\lambda R_{t-1}+y_t^2 \text{ with } R_0=0$$

It is easy to prove that the matrix Λ_t defined in Eq. (27) satisfies the condition in (13) because

$$tr(\Lambda_t)=\sum_{i=1}^{t}\lambda^{t-i}=\frac{1-\lambda^t}{1-\lambda}\le\frac{1}{1-\lambda} \tag{29}$$

is uniformly bounded for all t. Thus, the Recursive λ-weighted Mixed L2-Linfty learning algorithm globally minimizes the cost function $J_t=\frac{1}{2}\sum_{i=1}^{t}\lambda^{t-i}e_i^2$.

E. Instantaneous k-order RML Learning Algorithm

In this subsection, we derive a power instantaneous k-order dynamic RML learning algorithm from Eq. (12) for LIP models when choosing specific Λ_t. This learning algorithm updates at every step when the system has one more new data sample.

Choosing $\rho(t)=t$ and

$$\Lambda_t=\begin{bmatrix}0_{t-k} & 0\\ 0 & \Lambda(t;k)\end{bmatrix} \tag{30}$$

where 0_{t-k} is the (t-k)th-order zero matrix. And $\Lambda(t,k)$ is an k-order symmetric non-negative matrix satisfying

$$tr(\Lambda(t,k))\le M_\lambda \tag{31}$$

for all t for a positive constant number M_λ. One important matrix for Λ_t is when $\Lambda(t,k)=diag\{\lambda_{t-k+1},\lambda_{t-k+2},\cdots,\lambda_t\}$, then the Λ_t is

$$\Lambda_t=\begin{bmatrix}0_{t-k} & 0 & \cdots 0\\ 0 & \lambda_{t-k+1} & \cdots 0\\ \vdots & \vdots & \ddots \vdots\\ 0 & 0 & \cdots\lambda_t\end{bmatrix} \tag{32}$$

Our input and output matrices are with k rows from time $t-k+1$ to t.

$$\Phi_t=\Phi(t,k)=[\varphi_{t-k+1},\varphi_{t-k+2},\cdots,\varphi_t]^T \tag{33}$$

$$Y_t=Y(t,k)=[y_{t-k+1},y_{t-k+2},\cdots,y_t]^T \tag{34}$$

When $t<k$, λ_{t-k+1} to λ_0, φ_{t-k+1} to φ_0, and y_{t-k+1} to y_0 are arbitrarily set as the initial values.

P_t, Q_t and R_t defined in (8) - (10) have the recursive computation formulas as

$$P_t=P(t,k)=\Phi_t^T\Lambda(t,k)\Phi_t = \Phi^T(t,k)\Lambda(t,k)\Phi(t,k) \tag{35}$$

$$Q_t=Q(t,k)=\Phi_t^T\Lambda(t,k)Y_t = \Phi^T(t,k)\Lambda(t,k)Y(t,k) \tag{36}$$

$$R_t=R(t,k)=Y_t^T\Lambda(t,k)Y_t = Y^T(t,k)\Lambda(t,k)Y(t,k) \tag{37}$$

$$J_t=J(t,k)=\frac{1}{2}R_t-\theta_{t-1}^TQ_t+\frac{1}{2}\theta_{t-1}^TP_t\theta_{t-1} = \frac{1}{2}R(t,k)-\theta_{t-1}^TQ(t,k)+\frac{1}{2}\theta_{t-1}^TP(t,k)\theta_{t-1} \tag{38}$$

Then, the **instantaneous k-order Recursive Mixed L2-Linfty** learning algorithm

$$\theta_t=\theta_{t-1}+\frac{\alpha J_t}{\beta+(Q_t-P_t\theta_{t-1})^T(Q_t-P_t\theta_{t-1})}(Q_t-P_t\theta_{t-1}) \tag{39}$$

where $\beta>0, 0<\alpha<4, t=1,2,...$, and it minimizes the cost function $J_t=J(t,k)=\frac{1}{2}\sum_{i=t-k+1}^{t}\lambda_ie_i^2$ to its global minimum.

F. Batch k-order RML Learning Algorithm

The instantaneous k-order RML learning algorithm introduced in the preceding subsection updates the parameter vector at every step based on the current input and output data and the last $k-1$ data. However, the batch k-order RML learning algorithm developed in this subsection updates parameters at every k-step based on the last k data samples.

Choose $\rho(t)=kt$ and

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

$$\Lambda_t = \begin{bmatrix} 0_{kt-k} & 0 \\ 0 & \Lambda(kt;k) \end{bmatrix} \quad (40)$$

where 0_{kt-k} is the (kt-k)th-order zero matrix. And $\Lambda(kt,k)$ is an k-order symmetric non-negative matrix satisfying

$$tr(\Lambda(kt,k)) \leq M_\lambda \quad (41)$$

M_λ is a positive constant and $\Lambda_t = diag\{\lambda_{kt-k+1}\lambda_{kt-k+2},\cdots,\lambda_{kt}\}$, then the Λ_t is

$$\Lambda_t = \begin{bmatrix} 0_{kt-k} & 0 & \cdots & 0 \\ 0 & \lambda_{kt-k+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{kt} \end{bmatrix} \quad (42)$$

We further introduce the following notations:

$$\Phi_t = \Phi(kt,k) = [\varphi_{kt-k+1}, \varphi_{kt-k+2}, \cdots, \varphi_{kt}]^T \quad (43)$$

$$Y_t = Y(kt,k) = [y_{kt-k+1}, y_{kt-k+2}, \cdots, y_{kt}]^T \quad (44)$$

$$P_t = P(kt,k) = \Phi_t^T \Lambda(kt,k) \Phi_t = \Phi^T(kt,k)\Lambda(kt,k)\Phi(kt,k) \quad (45)$$

$$Q_t = Q(kt,k) = \Phi_t^T \Lambda(kt,k) Y_t = \Phi^T(kt,k)\Lambda(kt,k)Y(kt,k) \quad (46)$$

$$R_t = R(kt,k) = Y_t^T \Lambda(kt,k) Y_t = Y^T(kt,k)\Lambda(kt,k)Y(kt,k) \quad (47)$$

$$J_t = J(kt,k) = \frac{1}{2}R_t - \theta_{t-1}^T Q_t + \frac{1}{2}\theta_{t-1}^T P_t \theta_{t-1} = \frac{1}{2}R(kt,k) - \theta_{t-1}^T Q(kt,k) + \frac{1}{2}\theta_{t-1}^T P(kt,k)\theta_{t-1} \quad (48)$$

Then, **Batch k-order Recursive Mixed L2-Linfty** learning algorithm

$$\theta_t = \theta_{t-1} + \frac{\alpha J_t}{\beta + (Q_t - P_t\theta_{t-1})^T (Q_t - P_t\theta_{t-1})}(Q_t - P_t\theta_{t-1}) \quad (49)$$

where $\beta > 0, 0 < \alpha < 4, t = 1,2,...,$ and it minimizes the cost function $J_t = J(kt,k) = \frac{1}{2}\sum_{i=kt-k+1}^{kt} \lambda_t e_i^2$ to its global minimum. The RQ learning algorithm we developed in [20] is one special case of batch k-order RML learning algorithm, where the $\Lambda(kt,k)$ is chosen to be k-order identity matrix.

IV. SOME APPLIED MODELS

System identification is the process of developing a mathematical model of a dynamic physical system using experimental data to describe the input, output and noise relationship. A family of candidate models are available and the particular member in this family is determined that satisfactorily describes the observed data. This selecting criterion is based on some error criterion such as minimizing the measurement residuals due to the input and output signal and noises. In this section, we discuss some models that RML algorithms can be applied for system identification.

A. *Continuous Time Models*

In the recursive schemes, the variables have been indexed by a discrete parameter t, where the notation t was chosen because in many applications it denotes time. In some cases it is natural to use continuous time observations [19]. It is straightforward to generalize the results to this case. Equations (5) to (10) of LIP in Section 2.2 are still used, but now assumed to be in real time domain. Assuming with exponential forgetting factor, the parameter should be determined such that the criterion

$$J_t = \int_0^t (Y_\tau - \Phi_\tau \theta_\tau) d\tau$$

is minimized.

Therefore, the form of **Continuous time RML learning algorithm** is:

$$\frac{d\theta_t}{dt} = P_t \Phi_t J_t$$

$$\frac{dP_t}{dt} = P_t - P_t \Phi_t^T \Lambda_t \Phi_t P_t$$

$$J_t = Y_t - \Phi_t \theta_t$$

where the forgetting factor matrix Λ_t is defined in equation (4).

B. *State Space Models*

A state space model of a linear system describes the system input and output via a quantity called **the state vector** [8]. In general, the state space models are not accessible for RML or other derived algorithms. Re-parameterization is needed before RML algorithm can be applied. Next we use an example to explain the state space model.

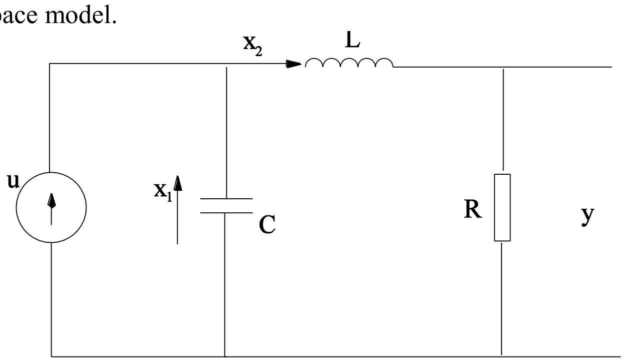

Fig 1: The circuit in State Space Model

Considering the circuit in Fig. 1, the state space representation is

$$\frac{dx}{dt} = \begin{bmatrix} 0 & -\frac{1}{C} \\ \frac{1}{L} & -\frac{R}{L} \end{bmatrix} x + \begin{bmatrix} \frac{1}{c} \\ 0 \end{bmatrix} u$$

$$y = \begin{bmatrix} 0 & R \end{bmatrix} x$$

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

and the transfer function is

$$G(s)=\frac{\frac{R}{LC}}{s^2+\frac{R}{L}s+\frac{1}{LC}}$$

Let $\theta_1=R,\theta_2=\frac{1}{L}$, and $\theta_3=\frac{1}{C}$. Then

$$G(s)=\frac{\theta_1\theta_2\theta_3}{s^2+\theta_1\theta_2 s+\theta_2\theta_3}$$

The coefficients are nonlinear (although of special structure) in the physical parameters $R,\frac{1}{L}$, and $\frac{1}{C}$. The system can be written as

$$G(s)=\frac{k_1}{s^2+k_2 s+k_3} \quad (50)$$

and it is possible to modify (50) by making an estimation of θ_1,θ_2, and θ_3, where the estimates must be constrained by the following relations.

$k_1=\theta_1\theta_2\theta_3$

$k_2=\theta_1\theta_2$

$k_3=\theta_2\theta_3$

Then the state spaces model is changed to be a continuous time Autoregressive Moving Average (ARMA) process model, when the model is sampled at the chosen time interval, it is the discrete ARMA model stated in [20]. We can then use RML learning algorithms on it for parameter estimation.

V. CASE STUDIES

In this section, we will use the various RML learning algorithms on several examples to show the algorithms' efficiency and effectiveness. The first example illustrates the robustness of the RML learning algorithm in the presence of noisy data.

Example 1 The system model is

$y'_t+0.4y'_{t-1}=x_t+0.6x_{t-1}$

$y'_t=y_t-w_t$

where x_t is a random signal sequence and w_t is nonmeasurable, normally distributed, statistically independent Gaussian noise with variance 0.36. Using the RML learning algorithm with k=10, we get the simulation results shown in Fig. 2 and Fig. 3. In this simulation, the parameters converge to [−0.4084,0.5916,1.0014].

Next we use the RML learning algorithm on two examples of industrial applications.

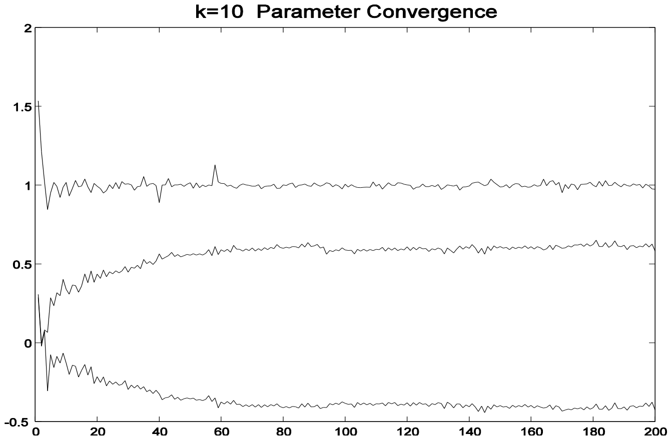

Fig 2: Parameters convergence of Example 1 using RML when k=10.

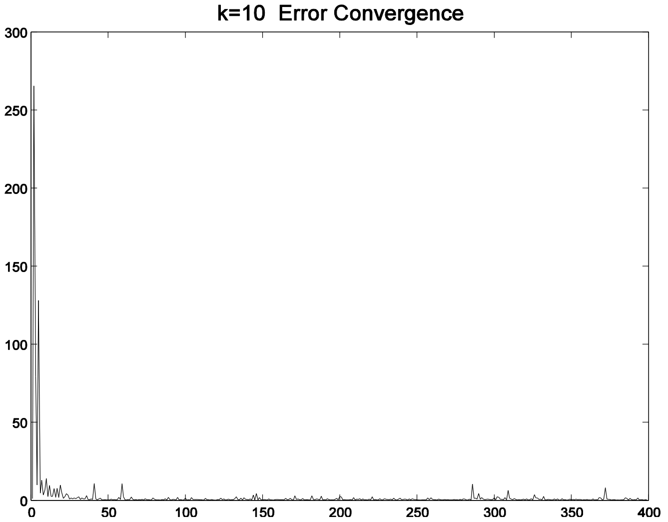

Fig 3: Error convergence of Example 1 using RML when k=10.

Example 2 Suppose that we are analysts for a client - American Manufacturing Company, a major manufacturer of a wide variety of commercial and industrial products. American Manufacturing owns a large nine-building complex in Central City and heats this complex by using a modern coal-fueled heating system. In the past, American Manufacturing has encountered problems in determining the proper amount of coal to order each week to heat the complex adequately. Because of this, we are requested to develop an accurate way to predict the amount of fuel (in tons of coal) that will be used to heat the nine-building complex in future weeks. The experience indicates that weekly fuel consumption depends substantially on (1) the average hourly temperature (in degrees Fahrenheit) during the week and (2) factors other than average hourly temperature that contribute to an overall "chill factor" such as:

- Wind velocity (in miles per hour) during the week
- "Cloud cover" during the week
- Variations in temperature, wind velocity, and cloud cover during the week (perhaps caused by the movement of weather fronts).

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

Week i	Average hourly temperature, x_i	Weekly fuel consumption, y_i
1	x_1= 28.0	y_1 = 12.4
2	x_2 = 28.0	y_2 = 11.7
3	x_3= 32.5	y_3 = 12.4
4	x_4 = 39.0	y_4 = 10.8
5	x_5 = 45.9	y_5 = 9.4
6	x_6 = 57.8	y_6 = 9.5
7	x_7 = 58.1	y_7 = 8.0
8	x_8 = 62.5	y_8 = 7.5

Table 1: Fuel consumption data of Example 2

In this example we use regression analysis to predict the *dependent variable* weekly fuel consumption y, on the basis of the *independent variable* average hourly temperature x. Table 1 lists the gathered data of the average hourly temperature (x_i) and the fuel consumption (y_i) observed in week *i* for the total *n*=8 weeks prior to the current week. Here the letter *i* denotes the time order of a previously observed week. Then we will use additional independent variables, which measure the effects of factors such as wind velocity and cloud cover, to help us predict weekly fuel consumption. It should be noted that it would, of course, be better to have more than eight weeks of data. However, sometimes data availability is initially limited. Furthermore, we have purposely limited the amount of data to simplify subsequent discussions in our regression example.

To develop a regression model describing the fuel consumption value, we first consider the second week in Table 1. In the second week the average hourly temperature was x_2=28.0, and the fuel consumption was y_2=11.7. If we were to observe the first week having the same average hourly temperature of 28.0, however the fuel consumption 12.4 is different from 11.7. This is because factors other than average hourly temperature - factors such as average hourly wind velocity and average hourly thermostat setting - affect weekly fuel consumption. It follows that there is an infinite population of potential weekly fuel consumptions that could be observed when the average hourly temperature is 28.0. To generalize the proceeding discussion, considering all eight fuel consumptions in Table 1, we may express for *i*=1,2,...,8 y_i in the form

$$y_i=\theta_0+\theta_1 x_i+e_i \quad (51)$$

Here, e_i is the error term describes the effect on y_i of all factors that have occurred in the *i*th week other than the average hourly temperatures x_i. The θ_0 is the *y-intercept* of the straight line, and θ_1 is the *slope* of the straight line.

We refer to the equation (51) as the *simple linear* (or *straight-line*) *regression model* relating y_i to x_i in this example, and θ_0,θ_1 is the *parameters* of the model. Note that y_i is assumed to be randomly selected from the infinite population of potential values of the dependent variable that could be observed when the value of the independent variable *x* is x_i.

Let's use RML learning algorithm, due to the finite data, we use them repeatedly in our recursive algorithm. When we choose *k*=4 and set the initial parameter values $[\theta_0 \quad \theta_1]=[1 \quad 1]$ and the error bound 0.3, then we can get that θ_0=15.6631 and θ_1=–0.1243.

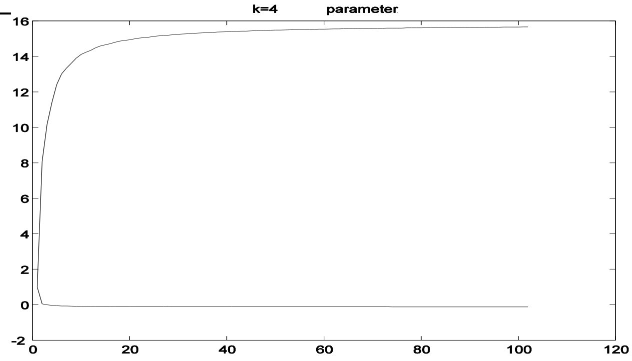

Fig **4: Parameters convergence of Example 2 using RML when *k*=4.**

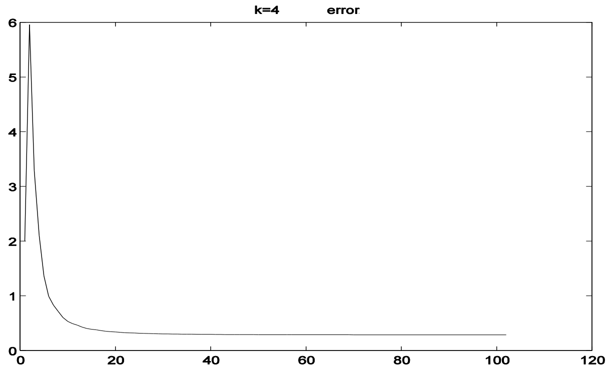

Fig 5: Error convergence of Example 2 using RML. Error drops below 0.3 after 101 steps.

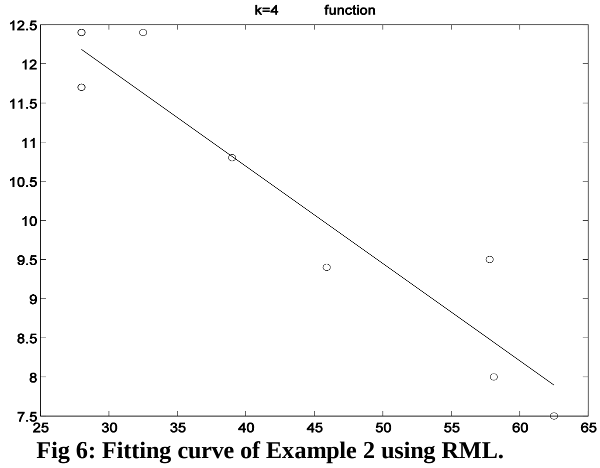

Fig 6: Fitting curve of Example 2 using RML.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

Fig. 4 - Fig. 6 show the results of the simulations using RML learning algorithm. Then using the values of θ_0 and θ_1, we can predict the amount of fuel for the nine-building complex of American Manufacturing Company in future weeks.

Example 3 Suppose that a department store tried out a television advertising campaign in order to see what effect, if any, television advertising had on its sales. Data on the number of minutes' advertising per week and the sales in the store were recorded for 20 successive weeks, beginning a week ahead of the advertising in Table 2.

Week	Minutes of TV Advertising (x)	Weekly Sales (y) (1000 dollars)	Residual (e)
1	$x_1 = 0$	$y_1 = 8.8$	$e_1 = -0.5363$
2	$x_2 = 5$	$y_2 = 9.0$	$e_2 = -0.9553$
3	$x_3 = 9$	$y_3 = 9.2$	$e_3 = -1.2505$
4	$x_4 = 4$	$y_4 = 10.0$	$e_4 = 0.1685$
5	$x_5 = 6$	$y_5 = 9.6$	$e_5 = -0.4791$
6	$x_6 = 13$	$y_6 = 10.4$	$e_6 = -0.5457$
7	$x_7 = 18$	$y_7 = 10.8$	$e_7 = -0.7647$
8	$x_8 = 5$	$y_8 = 9.8$	$e_8 = -0.1553$
9	$x_9 = 4$	$y_9 = 9.4$	$e_9 = -0.4315$
10	$x_{10} = 9$	$y_{10} = 10.2$	$e_{10} = -0.2505$
11	$x_{11} = 17$	$y_{11} = 12.0$	$e_{11} = 0.5591$
12	$x_{12} = 13$	$y_{12} = 11.0$	$e_{12} = 0.0543$
13	$x_{13} = 7$	$y_{13} = 10.8$	$e_{13} = 0.5971$
14	$x_{14} = 9$	$y_{14} = 10.8$	$e_{14} = 0.3495$
15	$x_{15} = 6$	$y_{15} = 10.6$	$e_{15} = 0.5209$
16	$x_{16} = 6$	$y_{16} = 11.0$	$e_{16} = 0.9209$
17	$x_{17} = 9$	$y_{17} = 11.2$	$e_{17} = 0.7495$
18	$x_{18} = 7$	$y_{18} = 11.0$	$e_{18} = 0.7971$
19	$x_{19} = 5$	$y_{19} = 10.4$	$e_{19} = 0.4447$
20	$x_{20} = 15$	$y_{20} = 11.4$	$e_{20} = 0.2067$

Table 2: Data for Television Advertising and Sales of Example 3

If we want to use only television advertising to estimate sales, we must adjust the data so as to eliminate the time effect. This way we will not violate the assumption of independent error terms. Thus, we perform a **first-order autoregressive** model that is a regression on first differences. A first difference is the difference between a value in one time period and the value in the preceding time period. These first differences should take the time effect into account if the value in one time period is affected by the value in the preceding time period. Our original first-order autoregressive model is

$$y_t=\theta_0+\theta_1 x_t+e_t$$

where

$$e_t=e_{t-1}+w_t$$

Here, we see that the error term at time t is a function of the error at the previous time period, plus some random fluctuation, w_t. If we let x' and y' denote the first differences for the x and y values, respectively, then

$$x'_t=x_t-x_{t-1}$$

$$y'_t=y_t-y_{t-1}$$

Furthermore,

$$y_t-y_{t-1}=\theta_0+\theta_1 x_t+e_t-(\theta_0+\theta_1 x_{t-1}+e_{t-1})$$

$$=\theta_1(x_t-x_{t-1})+(e_t-e_{t-1})$$

$$=\theta_1(x_t-x_{t-1})+w_t$$

Thus, we have the system model

$$y_t-y_{t-1}=\theta_1(x_t-x_{t-1})+w_t$$

i.e.,

$$y'_t=\theta_1 x'_t+w_t$$

and the residuals in the first-difference model are purely random.

Using RML learning algorithm, note that the first-differences model contains no intercept term, the slope of our fitted regression equation is θ_1=0.093 in Fig. 7, the error convergence trend is in Fig. 8. Thus, if in week 21, the store plans to purchase x_t=20 minutes of advertising on television, how would we estimate sales in week 21? In week 20, x_t=15 minutes, so $x't=x_t-x_{t-1}$=20–15=5 more minutes will be purchased in week 21 than in week 20. We estimate the growth in sales from week 20 to week 21 to be $\hat{y}'_t$=0.093*(5)=0.465=$465. In week 20, sales were $11400, we estimate actual sales in week 21 to be $11400+$465=$11865. Therefore, we can estimate sales in any week if we know its previous data.

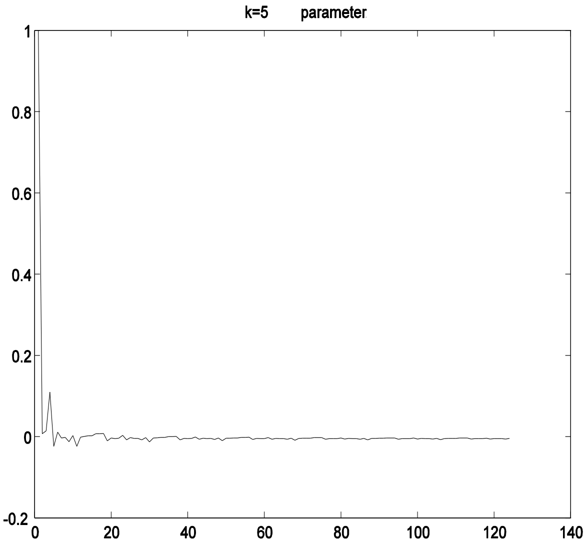

Fig 7: Parameters convergence of Example 3 using RML when *k*=5.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

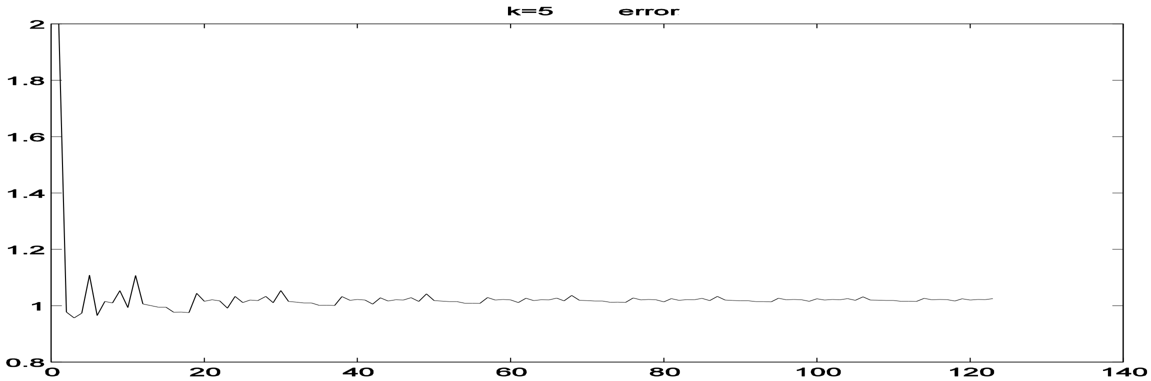

Fig 8: Error convergence of Example 3 using RML when *k*=5.

VI. CONCLUSION

In this paper, we have developed the various RML learning algorithms by choosing different forgetting factor matrix $\Lambda(\theta)$, and obtained several useful recursive learning algorithms for LIP models, such as Projection, Recursive Mixed L2-Linfty, Recursive Mixed Mean L2-Linfty, λ-weighted Mixed L2-Linfty, Instantaneous RML and Batch RML, etc. We have shown the effective of them by several simulation examples in adaptive identification and control fields.

Compared with other training methods, RML learning method has several distinct features. It can avoid the *windup* and *burst* phenomena which are the crucial drawbacks for the correspondent conventional learning algorithms. And the *k*-order RML has a faster training speed than the conventional ones when the *k* is chosen appropriately, the various new RML algorithms can be successfully applied in adaptive controller design.

In our future study, different various RML learning algorithms could be applied on more models and realistic industry applications in the fields of identification and adaptive control, to find the general rule on which algorithm is more appropriate in what kind of models by comparing the simulation results. In addition, an old business to study more is how to find the proper order k such that the convergent speed is fast and the computation at every iteration is reasonably simple.

REFERENCES

[1] K.J. Astrom and B. Witten mark, "Adaptive Control." Addison-Wesley, Mass., 1989.

[2] A. Bjorck, "Numerical Method for Least Squares Problems." Society for Industrial and Applied Mathematics (SIAM), 1996.

[3] S. Chu, S. Lo, "Application of the On-line Recursive Least-sqares Method to Perform Structural Damage Assessment." Journal of Structural Control Health Monitoring, 18(3), pp.241-264, 2011.

[4] J. Eriksson, M. Viberg, "Asymptotic Properties of Nonlinear Weighted Least Squares in Radar Array Processing." IEEE Transactions on Signal Processing, 52(11), pp. 3083-3095, 2004.

[5] D. Ernst, B. De Schutter, R. Babuska, "Online Least-squares Polcy Iteration for Reinforcement Learning Control." American Control Conference, pp.486-491, 2010.

[6] H.K. Fathy, K. Dongsoo, J.L. Stein, "Online vehicle mass estimation using recursive least squares and supervisory data extraction," American Control Conference, pp. 1842-1848, 2008.

[7] J.R. Deller, M. Nayeri, and S.F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," Proc. of the IEEE, 81(6), pp. 815-849, 1993.

[8] P.D. Gilbert, "State Space and ARMA Models: An Overview of the Equivalence." 1993.

[9] L.E. Ghaoui and H. Lebret, "Robust Least Squares and Applications." In Proc. CDC, pp. 249-254, Kobe, Japan, 1996.

[10] A.A. Giordano, F.M. Hsu, "Least Square Estimation with Applications to Digital Signal Processing." John Wiley & Sons, 1985.

[11] G.C. Goodwin and K.S. Sin, "Adaptive Filtering Prediction and Control." Prentice-Hall, Englewood Cliffs, N.J., 1984.

[12] C.C. Hang, T.H. Lee and W.K. Ho, "Adaptive Control." Instrument Society of America, USA, 1993.

[13] L. Li, M. L. Littman, and C. R. Mansley, "Online exploration in least-squares policy iteration," in Proceedings 8th International Joint Conference on Autonomous Agents an Mixed L2-Linftyultiagent Systems (AAMAS- 09), vol. 2, Budapest, Hungary, pp. 733-739, 10-15 May 2009.

[14] R. Johansson, "System Modeling and Identification." Prentice Hall, 1993.

[15] A. Maddi, A. Guessoum, D. Berkani, "Application to the recursive extended lest squares method for modeling a speech signal." Proc. of the 2nd ISCCP, 2006.

[16] Y. Nievergelt, "A Tutorial History of Least Squares with Applications to Astronomy and Geodesy." Journal of Computational and Applied Mathematics, 121(1), pp. 37-72, 2000.

[17] A. Riechman, and R.A. Scholtz, "Applications of Least-Square Lattice Filters in Adaptive Data Communication." Proc. Military Communications Conference, IEEE MILCOM, Vol. 3, pp. 47.5.1-47.5.4, 1986.

[18] M. Schuermans, P. Lemmerling, L. De Lathauwer, and S. Van Huffel, "Fast Communication: The Use of Total Least Squares Data Fitting in the Shape-from-moments

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 1, July 2015

Problem." Journal of Signal Processing, 86(5), pp. 1109-1115, 2006.

[19] C. Wang, D. Blei, and D. Heckerman, "Continuous time dynamic topic models," The 23rd Conference on Uncertainty in Artificial Intelligence, 2008.

[20] Q. Zhu, S. Tan and Y. Qiao, "A High-order Recursive Quadratic Algorithm of Linear-in-the-Parameters Models." Journal of Circuit, Systems, and Computers, 17 (2), pp. 1-28, 2008.

[21] Q. Zhu, Y. Qiao, and S. Tan, "A Robust High-order Mixed L2-Linfty Estimation for Linear-in-the-Parameters Models," Journal of Scientific Computing, 38 (2), pp. 185-206, 2009.

AUTHOR BIOGRAPHY

Dr. Qi Zhu earned his Ph.D. in Computer Science from University of Central Florida in 2002; currently he is an Associate Professor at University of Houston - Victoria. His research interests include operating system, queuing system, system modeling and analysis, software engineering, and web mining.