2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

A New Fast Online Identification Method for Linear Time-Varying Systems

Amir Haddadi and Keyvan Hashtrudi-Zaad, *member, IEEE*

***Abstract*— A novel online identification algorithm is proposed, which addresses the problem of convergence rate in dynamic parameter estimation in the presence of abrupt variations as well as noise in time-varying systems. The proposed identification technique optimizes a mean fourth error cost function by virtue of steepest descent (SD) method. It is proven that a unique solution for the optimal correcting gain of the SD update law exists and a closed-form solution is derived. To obtain high sensitivity to parameter variations, a block-wise version of the proposed technique, that incorporates only a finite length window of data, is developed. The performance of the proposed method is compared to those of two benchmark identification techniques.**

I. Introduction

In many applications the dynamic behaviour of plants go under gradual or drastic change. This can be due to the aging process, change in the plant configuration or the loading effect. For instance due to the high degree of dynamic nonlinearity, the inertia matrix of a hydraulic excavator changes throughout its workspace [16], or the dynamics of a robot, once becomes in contact with hard environment, goes under abrupt changes [17]. If the estimation of the parameters of such systems are inaccurate, *e.g.* within a few milliseconds after contact for the latter application, either instability and/or performance degradation occurs. There has been a variety of online estimation methods proposed for the identification of such time-varying systems.

Least squares (LS) minimizes mean square error (MSE) between the observed system output and the output of the estimated system. Many variations of LS have been developed since it was first proposed by Karl Gauss in 1795 [1]-[4]. Since LS is a batch method, for online applications all data from the beginning should be included in each time iteration. Recursive Least Squares (RLS) was proposed in 1950 using matrix inversion lemma to overcome the problem of computational complexity of running LS algorithm in every time step [2], [3].

With the recent advancements in computer technology the computational complexity of batch methods such as Batch Least Squares (BLS) with sliding window have become less and less an issue. Therefore, these methods have been used and shown to be of significant advantage in their tracking ability and convergence rates in the presence of parameter jumps [5]. BLS method has been shown to be sensitive to measurement noise when the window of data is short and it becomes slow when larger window size is chosen. As a result, there should be a compromise between sensitivity to noise and convergence rate. In [5] a variable window length is chosen and the length is resized whenever a change is detected. This requires a mechanism for parameter change detection.

Considerable effort has been put towards the development of modified versions of RLS for tracking changes in system dynamic parameters. Among these, RLS methods with Forgetting Factor and Covariance Resetting or Perturbing [7] are the most common ones [3]. Selective weighting RLS has also been proposed that minimizes the mean weighted squares of error. A limited list of weight selection methods can be found in [6]. Other than LS-based methods, Steepest Descent (SD) optimization method is also used to minimize MSE cost function [10], [11]. These methods tend to offer faster initial convergence rates as reported in [11]. In these cases an optimal correcting gain for steepest descent update law can be calculated explicitly.

For time-varying systems with abrupt changes in their dynamic parameters, considering higher power rather than two for error in the cost function causes the optimization algorithm to better capture sudden changes. As a result, a number of researchers have studied the problem of Least Mean P^{th} (LMP) optimization in various applications of communications such as filter design [12], adaptive FIR filters [8], [9] and sinusoidal frequency estimation [13]. Since LS-based techniques cannot be used when the power "P" is greater than two, to date only SD-based optimization methods have been used to minimize Mean P^{th} Error (MPE) [8], [9], [10], [14]. However, the SD correcting gains that have been used are not optimal and are obtained by trial and error. In fact, a closed-form solution for an optimal SD correcting gain when power is larger than two has not been proposed before. In this paper, we will develop a closed-form solution for an optimal SD correcting gain for P=4 and will prove the uniqueness of such gain. We will also propose a block-wise version of SD-based Least Mean Fourth (LMF) identification technique, so-called BLMF, in which a block window of data is used for online identification.

This paper is organized as follows. In Section II, MPE cost function is defined and the convex property of this function is studied. Section III discusses the use of SD method in the Least Mean Square (LMS) problem. It further presents the optimal correcting gain for LMF and proves the uniqueness of this gain. Section IV compares BLMF with two benchmark methods SD-based Block-wise LMS and RLS with

This work was supported by Ontario Centres of Excellence (OCE), Industrial Research Assistance Program (IRAP) and Quanser Consulting Inc.

The authors are with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada. amir.haddadi@ece.queensu.ca, khz@post.queensu.ca.

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

forgetting factor, in terms of convergence and computational load as well as sensitivity to noise. Conclusions are drawn in Section V.

II. MPE Cost Function

The system to identify should be linearly parameterizable in terms of its unknown parameters. We consider the following single-input single-output difference model for the system dynamics

$$y(k) = \boldsymbol{\phi}^T(k)\boldsymbol{\theta}(k) + n(k) \tag{1}$$

where $y(k)$ is the system output, $\boldsymbol{\theta}(k)$ is an $m \times 1$ vector of parameters, $\boldsymbol{\phi}(k)$ is the regressors vector, $n(k)$ denotes the measurement noise, and m is the number of unknown parameters. The output prediction error $e(k)$ is defined as

$$e(k) := y(k) - \hat{y}(k) \tag{2}$$

where $\hat{y}(k) = \boldsymbol{\phi}^T(k)\hat{\boldsymbol{\theta}}(k)$ is the estimated output using the estimated parameters $\hat{\boldsymbol{\theta}}(k)$. The goal of Least Mean Pth optimization is to estimate the parameters at each time step such that the MPE cost function

$$J[\hat{\boldsymbol{\theta}}(k)] = \frac{1}{p}\sum_{j=1}^{k}[y(j) - \boldsymbol{\phi}^T(j)\hat{\boldsymbol{\theta}}(k)]^p \tag{3}$$

is minimized, where p is an even number. The convexity of MPE cost function in parameter space R^m has been proven in [8] and [14]. Thus, every minimum of the corresponding function is a global minimum [8]. This fact shows that the iterative algorithm will converge to its true value given rich enough inputs.

In order to have fast parameter convergence rate, we propose the use of a window of data as opposed to the use of entire data so that the effect of recent parameter change is emphasized in the cost function. Therefore, the above cost function is modified to

$$J[\hat{\boldsymbol{\theta}}(k)] = \frac{1}{p}\sum_{j=k-L+1}^{k}[y(j) - \boldsymbol{\phi}^T(j)\hat{\boldsymbol{\theta}}(k)]^p \tag{4}$$

In this paper, the identification method using the above cost function is called Block Least Mean Pth method. For $p = 4$ the method is called Block Least Mean Fourth (BLMF).

III. LMP Method using Steepest Descent Optimization

The LMS optimization method ($p = 2$) updates the system parameters along the SD of the MSE cost function. This method can be generalized to the LMP iterative identification method. Although, there exists a closed-form solution for optimal correcting gain for LMS optimization, there is no such solution for LMP when $p > 2$. In this section, we will develop an analytical solution for the optimal LMF optimization ($p = 4$). However, the LMS iterative identification method is first introduced.

A. BLMS Online Identification with Optimal Correcting Gain

In this section, the SD-based Block Least Mean Squares (BLMS) online identification method with optimal correcting gain is presented. The block-wise MSE cost function is defined as

$$J[\hat{\boldsymbol{\theta}}(k)] = \frac{1}{2}\sum_{j=k-L+1}^{k}[y(j) - \boldsymbol{\phi}^T(j)\hat{\boldsymbol{\theta}}(k)]^2 \tag{5}$$

In the sequel, for simplicity we will show the argument of variables in time step time "k" with a subscript "k", *e.g.* $J(\hat{\boldsymbol{\theta}}_k)$ as J_k, $\boldsymbol{\Phi}(k)$ as $\boldsymbol{\Phi}_k$ and $\hat{\boldsymbol{\theta}}(k)$ as $\hat{\boldsymbol{\theta}}_k$. By defining the regressor matrix $\boldsymbol{\Phi}_k$ and the output vector $\mathbf{Y}_k$ as

$$\boldsymbol{\Phi}_\mathbf{k} = \begin{pmatrix} \boldsymbol{\phi}_k^T \\ \boldsymbol{\phi}_{k-1}^T \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\phi}_{k-L+1}^T \end{pmatrix}, \mathbf{Y}_k = \begin{pmatrix} y_k \\ y_{k-1} \\ \cdot \\ \cdot \\ \cdot \\ y_{k-L+1}, \end{pmatrix} \tag{6}$$

we can rewrite the MSE cost function as

$$\begin{aligned} J_k &= \frac{1}{2}(\mathbf{Y}_k - \boldsymbol{\Phi}_k\hat{\boldsymbol{\theta}}_k)^T(\mathbf{Y}_k - \boldsymbol{\Phi}_k\hat{\boldsymbol{\theta}}_k) \\ &= \frac{1}{2}(\mathbf{Y}_k^T\mathbf{Y}_k - 2\hat{\boldsymbol{\theta}}_k^T\boldsymbol{\Phi}_k^T\mathbf{Y}_k + \hat{\boldsymbol{\theta}}_k^T\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\hat{\boldsymbol{\theta}}_k). \end{aligned}$$

In SD method at each iteration, we calculate the parameter vector for the next time step according to the update law

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \lambda_k\mathbf{u}_k \tag{7}$$

where $\mathbf{u}_k$ is a unit vector in the opposite direction of the gradient vector, that is

$$\mathbf{u}_k = -\frac{\mathbf{g}_k}{\|\mathbf{g}_k\|}, \tag{8}$$

and

$$\mathbf{g}_k = \frac{\partial J_k}{\partial\hat{\boldsymbol{\theta}}_k} = -\boldsymbol{\Phi}_k^T\mathbf{Y}_k + \boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\hat{\boldsymbol{\theta}}_k = -\boldsymbol{\Phi}_k^T\mathbf{E}_k. \tag{9}$$

is the gradient vector and $\mathbf{E}_k := \mathbf{Y}_k - \boldsymbol{\Phi}_k\hat{\boldsymbol{\theta}}_k$ is the output prediction error. An optimal value for the correcting gain λ_k should satisfy

$$\frac{\partial J_{k+1}}{\partial\lambda_k} = 0, \tag{10}$$

which yields to the following analytical solution

$$\lambda_k = \frac{\mathbf{u}_k^T\boldsymbol{\Phi}_k^T\mathbf{E}_k}{\mathbf{u}_k^T\boldsymbol{\Phi}_k^T\boldsymbol{\Phi}_k\mathbf{u}_k}. \tag{11}$$

Deriving closed-form solution for the optimal gain for MPE optimization with powers greater than two has not been proposed before. In the next section, we will find a closed-form solution for the LMF (LMP for $p = 4$) optimal correcting gain, prove its uniqueness and propose the BLMF online identification method.

B. BLMF Online Identification with Optimal Correcting Gain

The LMF cost function is defined as

$$J_k = \frac{1}{4} \sum_{j=k-L+1}^{k} [y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k]^4. \tag{12}$$

As compared to the LMS cost function in (5), LMF imposes a larger cost associated with jumps in system dynamics. Therefore, this method is believed to display faster convergence in the presence of parameter jumps when error takes a large value. The gradient of cost function is derived as

$$\mathbf{g}_k = \frac{\partial J_k}{\partial \hat{\boldsymbol{\theta}}_k} = - \sum_{j=k-L+1}^{k} \boldsymbol{\phi}_j [y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k]^3. \tag{13}$$

As mentioned in the introduction, SD-based optimization methods have been used in communication applications to minimize Mean Pth Error (MPE), including when "$p = 4$" [8], [9], [10], [14]. In these cases, the correcting gain is chosen proportional to the gradient norm, that is $\lambda_k = \mu_k \|\mathbf{g}_k\|$, where the scale factor μ_k is a constant obtained by trial and error. This update law that is based on "scaled gradient norm" (SGN) correcting gain is not optimal.

Below we propose an optimal closed-form solution for the correcting gain λ_k. The optimization algorithm is performed via analytical minimization of the cost function. Thus, there is no numerical optimization involved to cause any delay in the estimation process. Using update law (7), the cost function is optimized for correcting gain λ_k when

$$\frac{\partial J_{k+1}}{\partial \lambda_k} = \sum_{j=k-L+1}^{k} (\boldsymbol{\phi}_j^T \mathbf{u}_k [y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k - \lambda_k \boldsymbol{\phi}_j^T \mathbf{u}_k]^3) = 0 \tag{14}$$

where $\mathbf{u}_k$ is defined in (8). Replacing for $\mathbf{u}_k$ from (8) and rearranging (14) the following cubic polynomial in terms of the optimal correcting gain is found that needs to be solved at every time step

$$G_k = A_k \lambda_k^3 + B_k \lambda_k^2 + C_k \lambda_k + D_k = 0, \tag{15}$$

where

$$A_k = \sum_{j=k-L+1}^{k} (\boldsymbol{\phi}_j^T \mathbf{u}_k)^4$$

$$B_k = -3 \sum_{j=k-L+1}^{k} (\boldsymbol{\phi}_j^T \mathbf{u}_k)^3 (y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k)$$

$$C_k = 3 \sum_{j=k-L+1}^{k} (\boldsymbol{\phi}_j^T \mathbf{u}_k)^2 (y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k)^2$$

$$D_k = - \sum_{j=k-L+1}^{k} (\boldsymbol{\phi}_j^T \mathbf{u}_k)(y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k)^3 = \|\mathbf{g}_k\|.$$

It is important to note that no approximation and no numerical method have been used in deriving the above equation, unlike many nonlinear optimization methods that have been used for system identification [1]. An important question that remains to be answered is: whether there exists a real root for the above third-order polynomial, and if it does, is it unique or multiple?

Theorem - Roots of G_k: There exists one and only one optimal real correcting gain for the BLMF problem.

Proof: We have to prove that the function G_k has one and only one real root λ_k. The function G_k is a 3^{rd}-order polynomial; thus, the existence of at least one real root is guaranteed. Therefore, if G_k is monotonically increasing (decreasing) it has always one and only one real root. In order to show the monotonic behavior, the derivative of G_k derived as

$$\frac{\partial G_k}{\partial \lambda_k} = 3 A_k \lambda_k^2 + 2 B_k \lambda_k + C_k \tag{16}$$

has to always be positive. Since A_k is always positive, if $B_k^2 - 3A_k C_k \leq 0$, then G_k is a monotonically increasing function. By defining $a_j = \boldsymbol{\phi}_j^T \mathbf{u}_k$ and $e_j = y_j - \boldsymbol{\phi}_j^T \hat{\boldsymbol{\theta}}_k$, the two terms $3A_k C_k$ and B_k^2 can be written as

$$3A_k C_k = 9 (\sum_{j=k-L+1}^{k} a_j^4)(\sum_{i=k-L+1}^{k} e_i^2 a_i^2) \tag{17}$$

$$B_k^2 = 9 (\sum_{j=k-L+1}^{k} e_j a_j^3)(\sum_{i=k-L+1}^{k} e_i a_i^3). \tag{18}$$

Since $3A_k C_k = B_k^2$ for $i = j$, we can say

$$B_k^2 - 3A_k C_k = 9 (\sum_{i,j=k-L+1, i \neq j}^{k} a_j^3 e_j a_i^3 e_i - \sum_{i,j=k-L+1, i \neq j}^{k} a_j^4 e_i^2 a_i^2).$$

Therefore, we have

$$\begin{aligned} \frac{B_k^2 - 3A_k C_k}{9} &\leq \sum_{i \neq j} |a_j|^3 |e_j| |a_i|^3 |e_i| - \sum_{i \neq j} a_j^4 e_i^2 a_i^2 \\ &= \sum_{i<j} 2 |a_j|^3 |e_j| |a_i|^3 |e_i| - (\sum_{i<j} a_j^4 e_i^2 a_i^2 + a_i^4 e_j^2 a_j^2) \\ &= \sum_{i<j} a_j^2 a_i^2 (2 |e_j| |e_i| |a_j| |a_i|) - \sum_{i<j} a_j^2 a_i^2 (e_j^2 a_i^2 + e_i^2 a_j^2) \\ &= \sum_{i<j} - a_j^2 a_i^2 (|e_j| |a_i| - |e_i| |a_j|)^2 \leq 0. \end{aligned}$$

■

In order to find the optimal correcting gain λ_k from G_k, either numerical or analytical methods can be used. The analytical roots of this cubic polynomial can be obtained using Cardano's equations [15]

$$\lambda_k = \frac{p_k}{3 m_k} - m_k - (\frac{B_k}{3 A_k}),$$

$$m_k = \sqrt[3]{\frac{q_k}{2} \pm \sqrt{\frac{q_k^2}{4} + \frac{p_k^3}{27}}},$$

$$p_k = \frac{3 A_k C_k - B_k^2}{3 A_k^2} = \frac{C_k}{A_k} - 3 (\frac{B_k}{3 A_k})^2,$$

$$q_k = \frac{D_k}{A_k} + \frac{2 B_k^3 - 9 A_k B_k C_k}{27 A_k^3} = \frac{D_k}{A_k} + 2 (\frac{B_k}{3 A_k})^3 - \frac{C_k}{A_k} (\frac{B_k}{3 A_k}).$$

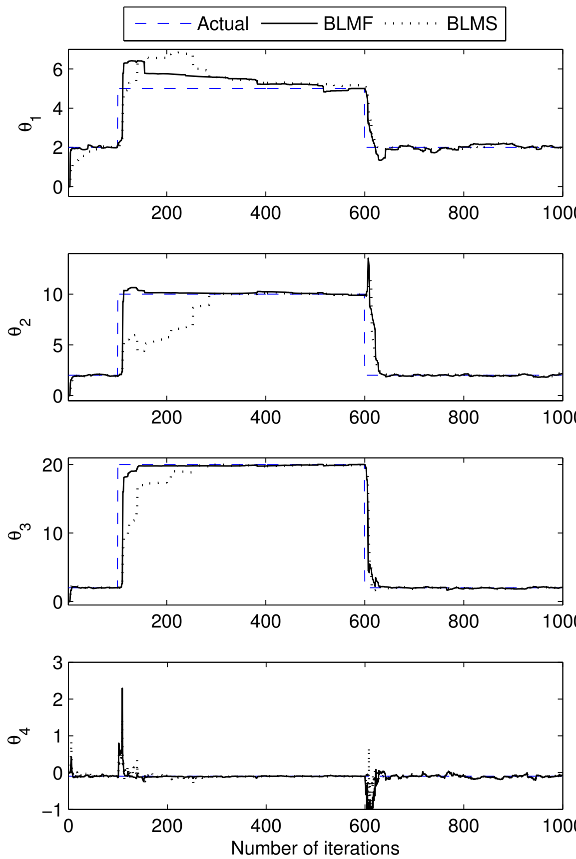

Fig. 1. Comparison between the proposed method (BLMF) and BLMS with optimal gain, $L = 10$.

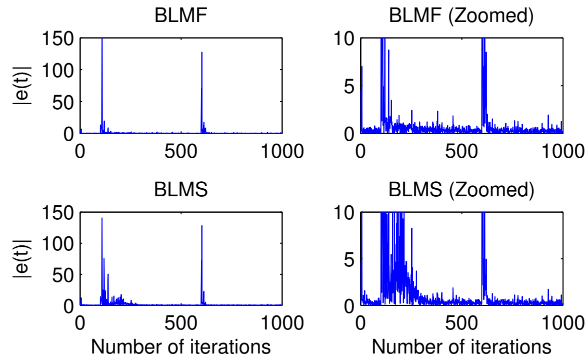

Fig. 2. Absolute Prediction error for BLMF and BLMS with optimal gain, $L = 10$. Right column shows the zoomed plots.

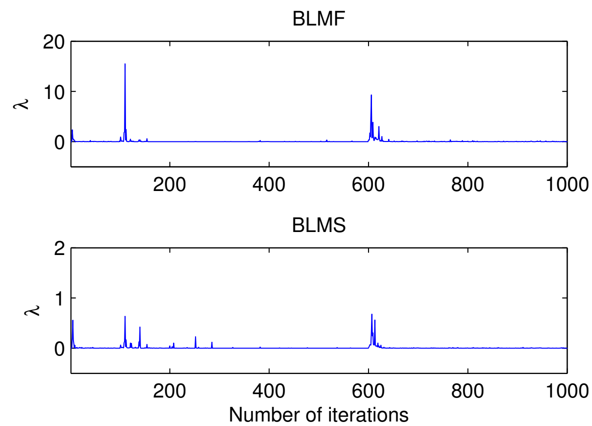

Fig. 3. Correcting gain λ_k in the BLMF and BLMS methods.

Since there are two solutions to the square root, and three complex solutions to the cubic root, six possibilities exist in computing m_k. However, only one solution is real and can be used for identification. Therefore, using the principal solution of the cubic root, the real λ_k can be calculated.

IV. Numerical Evaluation

Consider the following nonlinear time-varying system with input u_k and "measurement noise" n_k.

$$y_k{=}\theta_{1k} + \theta_{2k}u_{k-1} + \theta_{3k}u_{k-1}^2 + \theta_{4k}y_{k-1} + n_k \qquad (19)$$

This system is linear in its parameters and can be reformulated as in (1). For our simulations, the system parameters are chosen in a way that the system remains stable during the entire process in the input domain range. By defining the time periods $K_1 = \{k|k < 100, k > 600\}$ and $K_2 = \{k|100 \le k \le 600\}$, the system parameters are defined as:

$$\theta_{1k} = \begin{cases} 2 & k \in K_1 \\ 5 & k \in K_2 \end{cases}, \ \theta_{2k} = \begin{cases} 2 & k \in K_1 \\ 10 & k \in K_2 \end{cases}$$

$$\theta_{3k} = \begin{cases} 2 & k \in K_1 \\ 20 & k \in K_2 \end{cases}, \ \theta_{4k} = -0.1 \quad \forall k \ .$$

The above parameters show a sudden drastic change in the system dynamics at $k = 100$ and a sudden return to the original form at $k = 600$. All the parameters are initialized to zero in this paper, otherwise stated.

A. Comparison of BLMF and BLMS

Using (7), (11) and (15) we can simulate BLMS and BLMF algorithms with optimal correcting gains. Here n_k is a Gaussian white noise with zero mean and standard deviation 0.3 and the input signal is a Gaussian white noise with zero mean and standard deviation 1.

Figure 1 shows the parameter tracking performance for BLMS and BLMF with optimal gain when the window length is $L = 10$. For BLMF with SGN either the gain is normally chosen very small resulting in a very low convergence rate compared to BLMF with optimal gain, or the algorithm diverges. Thus, the results of BLMF with SGN is not provided in this paper. It is clear from Figure 1 that BLMF with optimal gain shows superior performance compared to BLMS with optimal gain, when an abrupt change occurs in the system parameters. The first parameter converges to 1.8 in about 10 time steps using BLMF method, whereas with

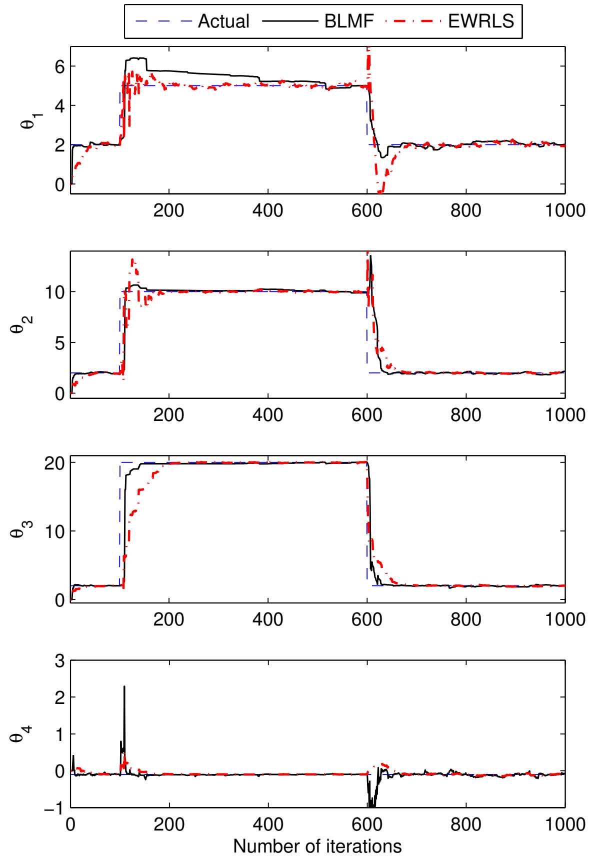

Fig. 4. Comparison between the proposed method (BLMF) and EWRLS, $L = 10$.

BLMS the convergence takes about 50 iterations. Although, the rise time (the time for the estimate to reach 90 percent of its final value) for both methods for θ_1 is almost equal, it is 8 to 18 times faster for BLMF considering θ_2 and θ_3. As shown in Figure 1 all parameters are initialized to zero and converge to 2. At $k = 100$ the change occurs and the parameters approach their final values, 5, 10 and 20 for θ_1, θ_2 and θ_3, from 2 which was their latest estimate. As a result BLMF performs well for different initial conditions, including zero and non-zero values. Figure 2 shows the absolute value of prediction error for both methods. As can be seen, when the system parameters change (at $k = 100$ and $k = 600$), the error increases. Following that the prediction error drops down based on the convergence rate of the estimation process. The results indicate that compared to BLMS, BLMF with optimal gain provides faster convergence rate. For BLMF, the error drops to less than 1 in about 50 time steps whereas for BLMS the convergence time is about 200 iterations. It is observed that whenever a high jump occurs in the system parameters, a high correcting gain causes faster convergence rate. Figure 3 shows that λ in BLMF is 25 times bigger than that of BLMS at the beginning of the parameter change. This rate is almost 17 at the end of the change period.

TABLE I

CONVERGENCE SPEED AND SENSITIVITY TO NOISE OF BLMF, BLMS AND EWRLS METHODS

	Convergence - θ_2 (*Number of iterations*)			STD of error in θ_2 estimate		
L	BLMF	BLMS	EWRLS	BLMF	BLMS	EWRLS
10	10	180	21	0.107	0.105	0.071
50	32	226	121	0.069	0.058	0.039
100	185	251	230	0.054	0.039	0.015

B. Comparison with Exponentially Weighted Recursive Least Squares (EWRLS)

In this part, the proposed method is compared with Recursive Least Squares (RLS) method with Forgetting Factor, known as Exponentially Weighted RLS (EWRLS). EWRLS method has been widely employed for parameter estimation of time-varying systems. Therefore, this method has also been selected as a benchmark to evaluate BLMF. EWRLS update equations can be written as

$$\mathbf{L}_{k+1} = \frac{\mathbf{P}_k \boldsymbol{\phi}_{k+1}}{\lambda + \boldsymbol{\phi}_{k+1}^T \mathbf{P}_k \boldsymbol{\phi}_{k+1}}$$

$$\mathbf{P}_{k+1} = \frac{1}{\lambda}[\mathbf{P}_k - \mathbf{L}_{k+1}\boldsymbol{\phi}_{k+1}^T \mathbf{P}_k]$$

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \mathbf{L}_{k+1}[F_{k+1} - \boldsymbol{\phi}_{k+1}^T \hat{\boldsymbol{\theta}}_k],$$

where matrix $\mathbf{P}$ is the covariance matrix and λ is the forgetting factor, which effectively puts emphasis on the last $\frac{\lambda}{1-\lambda}$ data points. Figure 4 shows the result of identification, when $L = 10$ for BLMF and $\lambda = 0.9091$ for EWRLS. The chosen forgetting factor causes EWRLS to put emphasis on the last 10 samples; therefore, the two methods are compared under similar conditions. As can be seen, although the long term convergence rates of BLMF and EWRLS are similar or even sometimes better for EWRLS depending on the input, noise and the type of system, BLMF shows a faster initial response to the abrupt changes with less fluctuations at $k = 0$ and $k = 100$, as shown in Figure 4. The quick response is of high importance for preventing instability in online control systems, such as adaptive control of robotic contact tasks.

C. The Effect of Window Length

In this part, we compare the performance of BLMF, BLMS and EWRLS methods under various window lengths, L. Window length can affect performance in three different ways:

1) *Convergence speed.* As discussed, short window lengths result in higher convergence rates for all block-wise methods. This is intuitive, since lesser data points collected before parameter change are included in the identification process. The first three columns of Table I show the convergence rates for θ_2 for the three methods with various window sizes. The table entries are mean values of the corresponding data collected

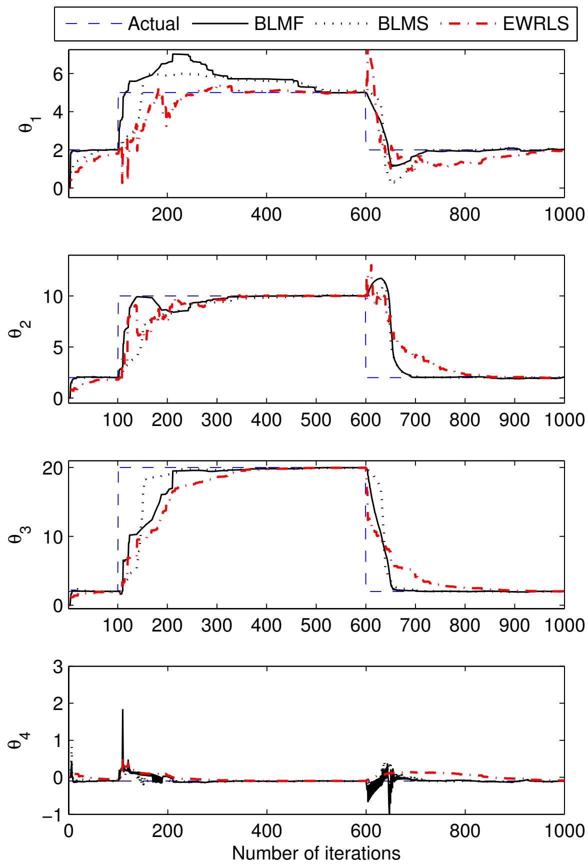

Fig. 5. Comparison between the proposed method (BLMF), the conventional BLMS, and EWRLS, $L = 50$.

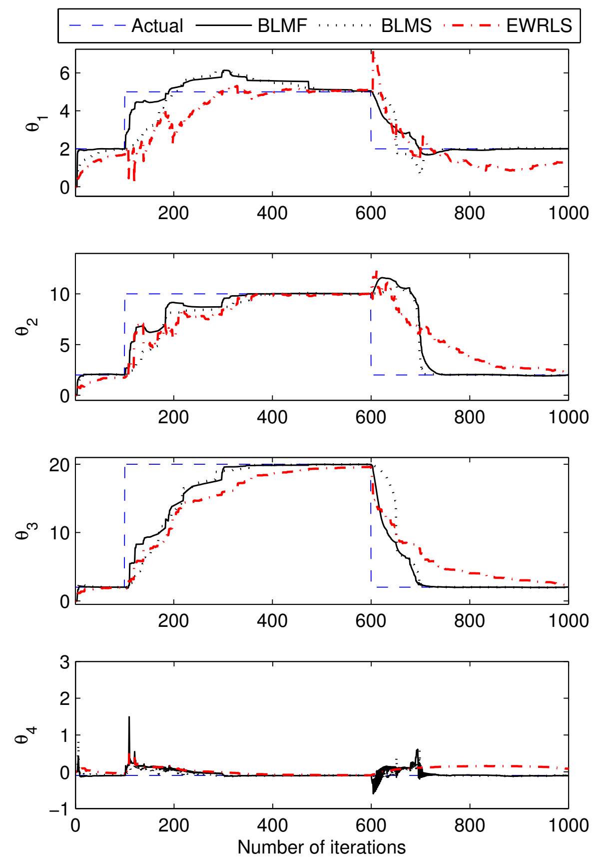

Fig. 6. Comparison between the proposed method (BLMF), the conventional BLMS, and EWRLS, $L = 100$.

from 5 separate simulations when white noise with the same standard deviation is the identification input for all simulations. Although the convergence rate for each parameter is different, increase in convergence rate for decreased window length is experienced for all parameters. Figures 5 and 6 show the results for BLMF and BLMS methods when the window length is 50 and 100, respectively. It is clear that although in both cases the convergence speed becomes slower as the window length increases, BLMF shows faster initial convergence rate in all cases. Increasing window length to 100 closes the gap between BLMF, BLMS and EWRLS as can be seen in Figure 6 and Table I. Here window length for EWRLS method refers to the effect of lambda in creating an equivalent window length $1/(1-\lambda)$.

2) *Effect of noise*. The effect of noise on parameter estimates is expected to increase with decrease in window size. The last three columns of Table I show the standard deviation of error in θ_2 after the convergence as an indication of the effect of noise in the estimation of θ_2. It reveals that EWRLS has the minimum sensitivity to noise; thus, it is more robust. This is due to the fact that all past collected data points with lower weights are included in the estimation process.
It can be seen that the sensitivity to noise in BLMF is more than that of EWRLS whereas, it is close to BLMS. For window size of $L = 10$ the STD of error in θ_2 for BLMF is 0.107, which is a relatively small number and is practically negligible. For larger window sizes STD becomes even smaller. Similar patterns can be derived for the other identified parameters of the system.

3) *Computational complexity*. Increasing window size makes computations more complex in BLMS and BLMF. However, for applications, in which the input signal is not persistently exciting for small window lengths, or the measurement noise is very high, larger window sizes are required. In such cases, computational complexity can become an issue in BLMF which requires calculation of the cubic polynomial coefficients in (15). Table II compares the performance of BLMF, BLMS and EWRLS methods for various window sizes in terms of computational load. The "Elapsed Time" denotes the time it takes in seconds for any of the algorithms to go through 1500 iterations.

TABLE II
COMPUTATIONAL LOAD OF BLMF, BLMS AND EWRLS METHODS

L	Elapsed Time (*sec./1500 iterations*)			Maximum Sampling Rate (Hz)		
	BLMF	BLMS	EWRLS	BLMF	BLMS	EWRLS
10	0.6238	0.1760	0.1344	2439	8333	11110
50	1.4997	0.1924	0.1322	1000	7692	11110
100	2.5635	0.2160	0.1351	589	7143	11110

This value affects the minimum sampling time or maximum sampling frequency at which the identification process can be implemented in real time.

As it can be seen from the right three columns of Table II that BLMF can be implemented at the rate of 1 kHz for window sizes up to 50. For larger window sizes, a slower sampling frequency should be utilized. Although, BLMF is more computationally expensive for small window sizes, the convergence speed can go over 4 times higher than BLMS and EWRLS.

V. CONCLUSIONS

In this paper, the problems of convergence rate in dynamic parameter estimation in the presence of abrupt changes as well as noise in time-varying systems have been addressed. To this end, we have first derived a unique closed-form solution for the steepest descent correcting gain in the Least Mean Fourth optimization problem. We have also proposed a Block Least Mean Fourth (BLMF) iterative identification method and numerically compared its performance with those of the Block Least Mean Square and Exponentially Weighted RLS (EWRLS) identification methods in the presence of measurement noise in terms of convergence rate, computational load and sensitivity to noise.

Although the proposed BLMF method is computationally more expensive than BLMS and EWRLS due to the calculation of non-recursive cubic polynomial coefficients for correcting gain, it produces faster convergence rate up to certain window size (in our case 100 samples). For small window size of 10 samples, BLMF is about 3 times computationally slower than BLMS and about 4 times slower than EWRLS; yet, it converges 2 to 10 times faster than BLMS and 2 to 4 times faster than EWRLS for different system parameters.

The noise sensitivity analysis has revealed that EWRLS is more robust than the other methods, while BLMF and BLMS have similar robustness. However, sensitivity to noise in BLMF is acceptable for practical applications, even with small window size of 10.

If the measurement noise is high and large window sizes (more than 100) are required, EWRLS has shown better performance. However, for small window lengths BLMF is a better option, since it captures sudden drastic variations faster than the other methods and it can be implemented in real time using a sampling frequency of 1 kHz for window sizes of less than 50, which is needed in practical applications such as haptics.

Future work will aim toward i) developing faster methods of finding the correcting gain in order to decrease the computational complexity, ii) comparing BLMF with other iterative identification techniques for various changes in system dynamics, and iii) applying BLMF to online control applications such as impact and contact control systems.

REFERENCES

[1] O. Nelles, *Nonlinear System Identification*, Springer- Verlag, 2001

[2] K.J. Åström and B. Wittenmark, *Adaptive Control*, Addison-Wesley, Inc., 1995.

[3] G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[4] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification"- a survey, *Automatica*, vol. 26, no. 1, pp 7-21, Jan. 1990.

[5] J. Jiang and Y. Zhang, "A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters", *Int. J. Adapt. Control Signal Process.*, vol 18, pp 505-521, March 2004.

[6] T.H. Hunt, "A Comparative Study of System Identification Techniques in the Presence of Parameter Variation, Noise, and Data Anomalies", *39th IEEE Midwest symposium on Circuits and Systems*, vol. 2, pp.593-596, Aug. 1996.

[7] D.-J. Park and B.-E. Jun, "Selfperturbing Recursive Least Squares Algorithm with Fast Tracking Capability", *Electronics Letters,*, vol.28, no.6, pp 558-559, March 1992.

[8] S.C. Pei and C.C. Tseng, "Least Mean P-Power Error criterion for adaptive FIR filter" *IEEE Journal on Selected Areas In Communivations*, vol. 12, no. 9, pp. 1540-1547, Dec. 1994.

[9] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and its family", *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp 275-283, March 1984.

[10] H.C. Shin, W.J. Song, and A.H. Sayed, "Mean square performance of data reusing adaptive algorithms", *IEEE Signal Proc. Letters*, vol. 12, no. 12, pp 851-854, Dec. 2005.

[11] R.S. Sutton, "Gain adaptation beats Least Squares?", *Proc. Seventh Yale Workshop on Adap. Lear. Sys.* pp 161-165, 1992.

[12] C. S. Burrus and J. A. Barreto, "Least p-power error design of FIR filters", *IEEE Inr. Symp. Circuits, Syst. (San Diego, CA)*, pp 545-548, May 1992.

[13] J. Schroeder, R. Yarlagadda, and J. Hershey, "L_p normed minimization with applications to linear predictive modeling for sinusoidal frequency estimation," *Signal Processing*, vol. 24, pp. 193-216, Aug. 1991.

[14] B. Weng and K.E. Barner, "Nonlinear system identification in impulsive environments", *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp 2588- 2594, July 2005.

[15] L.N. Childs, *A Concrete Introduction to Higher Algebra*, Springer, 2000.

[16] S.E. Salcudean, S. Tafazoli, K. Hashtrudi-Zaad, P.D. Lawrence, and C. Reboulet, "Evaluation of impedance and teleoperation control of a hydraulic mini-excavator", *Proc. 5th Int. Symp. Exp. Rob. (ISER)*, pp 185-198, June 1997.

[17] L.J. Love and W.J. Book, "Environment estimation for enhanced impedance control", *Proc. 1995 IEEE Int. Conf. on Robotics and Automation*, Washington, DC, USA, vol. 2, pp 1854-1859, May 1995.