Error Analysis of the Exponential RLS Algorithms Applied to Speech Signal Processing

*Kyung Yul Yoo

Abstract

The set of admissible time-variations in the input signal can be separated into two categories: slow parameter changes and large parameter changes which occur infrequently. A common approach used in the tracking of slowly time-varying parameters is the exponential recursive least-squares (RLS) algorithm. There have been a variety of research works on the error analysis of the exponential RLS algorithm for the slowly time-varying parameters. In this paper, the focus has been given to the error analysis of exponential RLS algorithms for the input data with abrupt property changes. The voiced speech signal is chosen as the principal application. In order to analyze the error performance of the exponential RLS algorithm, deterministic properties of the exponential RLS algorithms is first analyzed for the case of abrupt parameter changes, the convergence of the algorithm are verified and then related errors are analyzed. It has been theoretically shown that the impulsive input (or error variance) synchronous to the abrupt change of parameter vectors actually enhances the convergence of the exponential RLS algorithm. The analysis has also been verified through simulations on the synthetic speech signal.

I. Introduction

The ability to track time-variations in the signal model is one of the main advantages of an adaptive filter, compared with one with fixed parameters. Estimation of time-varying parameters is therefore a key issue in adaptive filtering. The set of admissible time-variations in the input signal can be separated into two categories: slow parameter changes and large parameter changes which occur infrequently[1].

A common approach used in the tracking of slowly time-varying parameters is to introduce an exponential forgetting factor, $\lambda \in (0, 1)$, to the input data, which leads to the exponential recursive least-squares (RLS) algorithm. In the exponential RLS algorithm the data are exponentially weighted and the time constant of the data weighting (i.e., roughly the number of significant data points) is $1/(1-\lambda)$. However, a fixed value of λ together with not persistently excited input data are known to cause problems such as the exponential growth of the covariance matrix, which result in "burst" phenomena (also known as covariance wind-up problem)[2][3]. This problem becomes a severe one if the parameter vector describing the input data undergoes abrupt changes. One way to overcome this problem is to vary the value of the forgetting factor with time, where we define an information content of the data and choose the forgetting factor at each iteration in such a way that this is kept constant. This leads to the design of a variable forgetting factor (VFF) RLS algorithm[4].

In this paper, the focus has been given on the error analysis of exponential RLS algorithms for the input data with abrupt property changes. Situations where the signal parameters experience abrupt changes could be found in various applications. A sudden change of load variance and a shift of operating point in a non-linear system are examples in the adaptive control. In the signal processing, the speech signal and heart cardiogram are typical examples. It is interesting that, in the latter examples, it is reasonable to assume that the signal parameters vary in the quasi-periodic way and the sudden changes of the system parameter often accompany a train of impulsive input at the moment of changes. The voiced speech signal is chosen as the principal application in this work. The deterministic property of the exponential RLS algorithms is first analyzed for the case of abrupt parameter changes. The convergence of the algorithm are then verified and related errors are analyzed. It has been theoretically shown that the impulsive input (or error variance) synchronous to the abrupt change of parameter vectors actually enhances the convergence of the exponential RLS algor-

*Department of Electrical Engineering, Hanyang University
Manuscript Received August 14, 1996.

ithm using the idea of persistent excitation. The analysis has also been verified through simulations on the synthetic speech signal.

Ⅱ. Error Analysis of RLS Algorithm

Assume that the input signal x(n) is characterized by an AR(p) process such that

$$x(n)=\mathbf{w}^T(n)x(n-1)+e(n) \quad (1)$$

where $\mathbf{w}^T(n-1)=[x(n-1), \cdots, x(n-p)]$ is an input data vector, $\mathbf{w}^T(n)=[w_1(n), \cdots, w_p(n)]$ is a vector for the time-varying filter weight, and the measurement error $e(n)$ is a centered white Gaussian sequence with variance σ_0^2. When an RLS algorithm is used to estimate w(n) (this corresponds to the linear prediction), the least squares estimate $\hat{w}(n)$ with exponential forgetting is defined by the following RLS structure;

$$\hat{w}(n)=\hat{w}(n-1)+\lambda\mathbf{P}(n)x(n-1)(x(n)-x^T(n-1)\hat{w}(n-1)) \quad (2)$$

$$\mathbf{P}^{-1}(n)=\lambda(\mathbf{P}^{-1}(n-1)+x(n-1)x^T(n-1)) \quad (3)$$

It starts from arbitrary initial values and then converges to its steady-state, whose speed (i.e. convergence rate) partially characterizes its performance. Convergence, tracking, and estimation error represent different, though related, properties of an exponential RLS algorithm. That is, all adaptive filters experience losses in performance, which are expressed through the excess mean-squared error (MSE), and are a result of two main sources of error: self noise and lag error. Since they are conflicting each other, one should find a trade-off in terms of λ among the convergence speed, tracking efficiency, and the estimation error. There have been numerous theoretical works for the derivation of an optimum trade-off for λ for various applications, such as in [5-9].

1. Analysis of Speech Signal

An all-pole model such as in Eq. (1) has commonly been adopted in the parametric analysis of the speech signal. Here, the speech signal is generated by exciting an all-pole filter by either a white noise sequence or a quasi-periodic impulse train. Parameters in this model are then estimated through the linear prediction, which could be realized by a block algorithm such as the autocorrelation method or the covariance method[10][11] or by a sequential algorithm such as a class of RLS algorithms[12][13]. The basic idea of block algorithms is that the model parameters remain stationary within the window length, which usually encompasses several pitch periods. However, the speech signal is non-stationary and model parameters are time-varying. Therefore, a large window size often ends up with a poor parameter estimates in the block algorithms compared to the pitch synchronous covariance method[14]. In the sequential estimation, one possible approach for the time-varying case would be assuming that the parameters in Eq. (1) either undergoes an abrupt parameter changes synchronously to epoch location or the noise variance of the assumed speech sequence shows a significant variance at the time of epoch. The basis of the first assumption can be found on the pitch synchronous speech analysis method. It is known that the pitch synchronous estimation of the formant parameter provides the most reliable estimates among various block-oriented speech analysis methods. The second assumption is also valid from the parametric model of a voiced speech signal. Based on these assumptions, it is quite natural to assume that the speech signal in Eq. (1) has quasi-periodic parameter changes together with impulsive inputs.

Either impulses at each epoch location in every voiced sequence or the parameter change happening synchronously with the beginning of a pitch period causes another source of error for the exponential RLS algorithm when it is applied to speech processing. Since this error is due to the variation of signal characteristics, it might be categorized as a tracking error. However, if the effect of these variations subsides fast enough as usually does, say within a pitch period of a speech signal, then its effect on the final estimation error would be negligible and thus it is reasonable to categorize it separately. Therefore, we could divide the estimation error into three parts; self noise, lag error, and error due to sudden parameter variation. We define the third source of error as the excitation error, which is caused by the abrupt parameter changes and impulsive input components.

2. Deterministic Error Analysis

In this section, we first identify the source of the excitation error through a deterministic analysis based on the previous assumptions. Let the speech signal is stationary over an interval of interest, and thus be represented by the constant parameter vector, w, and let {h(n), n ≥ 0} be the impulse response of the system represented by w. Now, assume that two impulses are applied to the filter at times n = 0 and n = T_0 with magnitude g_1 and g_2, respectively. Then, the output speech sequence will be a superposition of two impulse responses such as

$$x(n) = g_1 h(n) + g_2 h(n - T_0) \tag{4}$$

For $n < T_0$, the least-squares (LS) estimate $\hat{w}_{LS}$ is the solution of the normal equations,

$$\hat{R}(n) \cdot \hat{w}_{LS} = \hat{p}(n) \tag{5}$$

where $\hat{R}(n)$ and $\hat{p}(n)$ are the sample correlation matrix and cross-correlation vector, respectively, given by

$$\hat{R}(n) = \sum_{k=1}^{n} \lambda^{n-k} x(n-1) x^T(n-1) \tag{6}$$

$$\hat{p}(n) = \sum_{k=1}^{n} \lambda^{n-k} x(n-1) x(n) \tag{7}$$

To prove the exactness of the LS estimate, it is sufficient to show that $\hat{R}(n) w = \hat{p}(n)$. Then, by the uniqueness of the solution under the assumption that the sample autocorrelation matrix $\hat{R}(n)$ is nonsingular, we have $\hat{w}_{LS} \equiv w$. Form the normal equations for linear prediction of the form

$$\sum_{k=1}^{n} \lambda^{n-k} x(k-\tau) \sum_{i=1}^{p} w_i x(k-i) = \sum_{k=1}^{n} \lambda^{n-k} x(k) x(k-\tau),$$

$$\tau = 1, \ldots, p \tag{8}$$

it is straightforward to show that the vector w exactly satisfies the normal equations. Hence, the LS estimate becomes exact. This also holds for the interval $n \geq T_0 + p$. This result can be simply extended to the equally spaced multi-pulse case. As a result, we can conclude that the normal equations generate the exact parameter vector estimate at times $kT_0 + p < n \leq kT_0 - 1$ and $n \geq kT_0 + p$ for $k \geq 0$. That is, if the input sequence is a superposition of an impulse response and its delayed, weighted copies, then the least-squares and thus the RLS algorithm can produce an exact solution, provided that an exact initialization scheme is employed and the real parameter vector remains constant.

Analysis of the normal equations in the regions $T_0 \leq n < T_0 + p$ is rather complicated. Within this interval, the normal equations do not produce an exact solution. When T_0 is sufficiently large, then it is easy to show that the tap weight error vector, $\tilde{w}(n) = w - \hat{w}(n)$ at time $T_0 + k$ could be approximated to

$$\tilde{w}(T_0 + k) \cong g_2 \lambda^{k-1} \hat{R}^{-1}(T_0 + k) x(T_0 - 1), \quad k \geq 1 \tag{9}$$

Eq. (9) describes a bias in the least-squares estimation due to the impulsive inputs and could be categorized into an excitation error. Note that the deterministic bias is proportional to the magnitude of the impulse. From the above analysis, it is clear that those parameters respond to the existence of the impulses quite sensitively, which are reflected in sharp peaks on the parameter trajectories, $\{w_i(n), n \geq 0\}_{i=1}^{p}$. The existence of this error source could be utilized on the design of an event detector using a class of RLS algorithms[15].

The excitation error due to an impulse lasts only p iterations after the impulse is applied to the system. From this exactness of the LS estimates for $n > T_0 + p$, we could deduce that

$$\hat{R}^{-1}(T_0 + k) x(T_0 - 1) = 0_p, \quad k \geq 0 \tag{10}$$

where 0_p is a p-dimensional zero vector. Now, assume that the underlying parameter vector step-changes its value at time $T_0 + 1$ from $w + \delta w$ to w. Also assume that the WRLS algorithm has converged to generate the estimate $w + \delta w$ at time T_0. Then we can show that

$$\tilde{w}(T_0 + k) \cong g_2 \lambda^{k-1} \hat{R}^{-1}(T_0 + k) x(T_0 + k) \hat{R}(T_0 + 1) \delta w, \quad k \geq 0 \tag{11}$$

If δw represents bias or improper initialization, not the actual change of the tap weight vector, or if δw is small compared to the magnitude of impulse, then the first term in Eq. (11) would be the dominant factor on the interval $[T_0, T_0 + p - 1]$. Also in speech sequences, it is reasonable to assume that $\|\delta w\| < g_2$ for some vector norm, which makes the dominance of the first term clear. For $k > p$, the transient phenomenon caused by the impulsive input dies out, and then the second term in Eq. (11) becomes dominant in $\tilde{w}(T_0 + k)$ so that

$$\tilde{w}(T_0 + k) \cong \lambda^k \hat{R}^{-1}(T_0 + k) \hat{R}(T_0 + 1) \delta w, \quad k > p \tag{12}$$

Eq. (12) shows the transient behavior of the excitation error. If the pitch period is long, then this excitation error would vanish as the exponential RLS algorithm converges prior to the next epoch. Assuming 8 kHz sampling rate and a pitch period of 10 msec (corresponding to 80 samples within a pitch period), the exponential RLS algorithm would have plenty of time to converge within a pitch period. For short pitch periods (as in female voice), the excitation error may not have enough time to die out and thus can contribute to the overall estimation error.

3. Convergence of RLS algorithm

The convergence of the adaptive algorithms are closely related with the property of the input data sequence. One of the important properties is the persistent excitation condition[16]. Suppose that there exists a positive integer $N \geq p$ and two real values m and M, $0 < m < M < \infty$, such that the persistent excitation condition is satisfied, i.e.,

$$0 < m\mathbf{I} \leq \sum_{i=n+1}^{n+N} \mathbf{x}(i-1)\,\mathbf{x}^T(i-1) \leq M\mathbf{I} < \infty \tag{13}$$

for all $n > 0$. With this persistent excitation condition, we can find asymptotic lower- and upper- bounds on $\mathbf{P}(n)$ such that (derivation is given in the Appendix)

$$\left(\frac{1-\lambda^N}{M\lambda}\right)\mathbf{I} \leq \mathbf{P}(n) \leq \left(\frac{1-\lambda^N}{M\lambda^N}\right)\mathbf{I}, \text{ as } t \to \infty \tag{14}$$

From Eq. (14), it is easy to see $\mathbf{P}(n) \to 0$ as $t \to \infty$. This phenomenon is compatible with the well-known "burst" property of the RLS algorithm.

Let $\tilde{\mathbf{w}}(n) = \mathbf{w}(n) - \hat{\mathbf{w}}(n)$ be the estimation error at time n, and $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}(n-1)$ be the parameter's variation at time t. Then, from Eqns. (2) and (3), we can derive

$$\tilde{\mathbf{w}}(n) = [\mathbf{I} - \lambda\mathbf{P}(n)\,\mathbf{x}(n-1)\,\mathbf{x}^T(n-1)]\cdot(\mathbf{w}(n-1) + \Delta\mathbf{w}(n)) - \lambda\mathbf{P}(n)\,\mathbf{x}(n-1)\,e(n) \tag{15}$$

Using superposition, the estimation error, $\tilde{\mathbf{w}}(n)$, can be divided into three independent components, i.e. $\tilde{\mathbf{w}}(n) = \tilde{\mathbf{w}}^1(n) + \tilde{\mathbf{w}}^2(n) + \tilde{\mathbf{w}}^3(n)$, each of which characterizes the effects on the estimation error due to: (a) initial estimation error, (b) variations of the signal parameters, and (c) stochastic error and satisfies the following equations[17].

$$\tilde{\mathbf{w}}^1(n) = \lambda\mathbf{P}(n)\,\mathbf{P}^{-1}(n-1)\,\tilde{\mathbf{w}}^1(n-1),\quad \tilde{\mathbf{w}}^1(0) = \tilde{\mathbf{w}}(0) \tag{16a}$$

$$\tilde{\mathbf{w}}^2(n) = \lambda\mathbf{P}(n)\,\mathbf{P}^{-1}(n-1)\,(\tilde{\mathbf{w}}^2(n-1) + \Delta\mathbf{w}(n)),\quad \tilde{\mathbf{w}}^2(0) = 0 \tag{16b}$$

$$\tilde{\mathbf{w}}^3(n) = \lambda\mathbf{P}(n)\,\mathbf{P}^{-1}(n-1)\,\tilde{\mathbf{w}}^3(n-1) - \lambda\mathbf{P}(n)\,\mathbf{x}(n-1)\,e(n),\quad \tilde{\mathbf{w}}^3(0) = 0 \tag{16c}$$

One can easily verify that the addition of three terms together leads to Eq. (15). The convergence of both $\tilde{\mathbf{w}}^1(n)$ and $\tilde{\mathbf{w}}^3(n)$ could be verified using the usual analytic procedure shown previous works[6][10]. Thus, it is sufficient to verify the convergence of $\tilde{\mathbf{w}}^2(n)$ to completely analyze the convergence of the exponential RLS algorithms for the abrupt parameter changes. Without loss of generality, suppose that the estimation started at time $t = -\infty$. This assumption is valid only if the step change of parameters are sufficiently spaced so that the exponential RLS algorithm have converged. Also assume that a change $\delta\mathbf{w}$ in the parameter vector occurs at time instant $T_0 + 1$, so that $\Delta\mathbf{w}(n) = \delta\mathbf{w}\cdot\delta(T_0 + 1)$, where $\delta(\cdot)$ is the Kroneker delta function. Since fully converged, we can assume that $\tilde{\mathbf{w}}^2(T_0) = 0$ and thus consider $\delta\mathbf{w}$ as an initial estimation error of type-a for the estimation algorithm starting at time $T_0 + 1$. To prove the convergence of $\tilde{\mathbf{w}}^2(n)$, define a matrix norm by;

$$\|\tilde{\mathbf{w}}(n)\|_{\mathbf{P}^{-1}(n)} = \tilde{\mathbf{w}}^T(n)\,\mathbf{P}^{-1}(n)\,\tilde{\mathbf{w}}(n) \tag{17}$$

Here, $\tilde{\mathbf{w}}(n) \equiv \tilde{\mathbf{w}}^2(n)$ for the notational simplicity. Then, using Eq. (3) and the matrix inversion lemma, it is straighftforward to show that

$$\tilde{\mathbf{w}}^T(n)\,\mathbf{P}^{-1}(n)\,\tilde{\mathbf{w}}(n) = \lambda\left(\tilde{\mathbf{w}}^T(n-1)\,\mathbf{P}^{-1}(n-1)\,\tilde{\mathbf{w}}(n-1) - \frac{\tilde{\mathbf{w}}^T(n-1)\,\mathbf{x}(n-1)\,\mathbf{x}^T(n-1)\,\tilde{\mathbf{w}}(n-1)}{1 + \mathbf{x}^T(n-1)\,\mathbf{P}(n-1)\,\mathbf{x}(n-1)}\right) \tag{18}$$

so that, for $n \gg T_0$,

$$\|\tilde{\mathbf{w}}(n)\|_{\mathbf{P}^{-1}(n)} = \lambda\left(\|\tilde{\mathbf{w}}(n-1)\|_{\mathbf{P}^{-1}(n-1)} - \frac{\{\tilde{\mathbf{w}}^T(n-1)\,\mathbf{x}(n-1)\}^2}{1 + \mathbf{x}^T(n-1)\,\mathbf{P}(n-1)\,\mathbf{x}(n-1)}\right) \tag{19}$$

and thus

$$\|\tilde{\mathbf{w}}(n)\|_{\mathbf{P}^{-1}(n)} \leq \lambda\,\|\tilde{\mathbf{w}}(n-1)\|_{\mathbf{P}^{-1}(n-1)} \leq \|\tilde{\mathbf{w}}(n-1)\|_{\mathbf{P}^{-1}(n-1)} \tag{20}$$

Also, from Eq. (16b), we get by letting $n = T_0 + \tau$, $\tau > 0$

$$\tilde{\mathbf{w}}(T_0 + \tau) = \lambda^{n}\,\mathbf{P}(T_0 + \tau)\,\mathbf{P}^{-1}(T_0)\,\delta\mathbf{w} \tag{21}$$

and

$$\lambda^{\tau}\,\frac{\lambda_{min}(\mathbf{P}(T_0 + \tau))}{\lambda_{max}(\mathbf{P}(T_0))}\,\|\delta\mathbf{w}\| \leq \|\tilde{\mathbf{w}}(T_0 + \tau)\| \leq \lambda^{\tau}\,\frac{\lambda_{max}(\mathbf{P}(T_0 + \tau))}{\lambda_{min}(\mathbf{P}(T_0))}\,\|\delta\mathbf{w}\| \tag{22}$$

Now, from Eq. (14), we can derive the upper and lower bound on the maximum and minimum eigenvalues of $\mathbf{P}(n)$, respectively, such that

$$\frac{1}{\lambda_{max}(\mathbf{P}(T_0 + \tau))} \geq \frac{m\lambda^N}{1 - \lambda^N} \quad \text{and}$$

$$\lambda_{min}(\mathbf{P}(T_0 + \tau)) \geq \frac{1 - \lambda^N}{M\lambda} \tag{23}$$

so that Eq. (22) becomes

$$\frac{m}{M}\lambda^{t+N-1} \le \frac{\|\tilde{w}(T_0+\tau)\|}{\|\delta w\|} \le \frac{M}{m}\lambda^{t+N-1} \qquad (24)$$

From Eq. (24), we can conclude that the estimation error component due to the parameter step change converges exponentially to zero, and $\tilde{w}^2(T_0+\tau)$ are changing within a strip, whose size is a function of the degree of persistent excitation of the input signal. Also, note that the size of this strip reduces exponentially. It is very delicate to measure how a stochastic sequence is persistently excited[16]. However, one thing obvious is that an impulsive component at the same instant as that of parameter step change enhances the degree of persistent excitation of the input signal.

This could be seen at Figures 1 and 2. They show two trajectories, $\hat{w}_1(n)$ and $\hat{w}_2(n)$, which are linear predictive estimates of the signal generated by

$$x(n)=w_{1,k}(n)x(n-1)+w_{2,k}(n)x(n-2)+g_k\delta(n-kT_0)+e(n) \qquad (25)$$

where e(n) is a Gaussian noise sequence with variance 1, and $T_0=200$. The filter weight are adjusted periodically according to Table 1. Fig. 1 corresponds to the case with $g_k=0$ for all k, whereas Fig. 2 to those values in Table 1. The former one represents the parameter step change without accompanying impulsive inputs. In contrast, the latter one is with impulsive inputs. The exponential RLS algorithm with $\lambda=0.9833$ and 0.9672 are used, and it is compared to the block covariance algorithm implemented by the sliding window covariance (SWC) RLS algorithm with window size 60. Values of $\lambda=0.9833$ and 0.9672 are related to the window size 60 in some analytic sense, that is, $N=1/(1-\lambda_1)$ and $\lambda_2=(N-1)/(N+1)$[18]. One can clearly see that the exponential RLS algorithm converges much faster in the former case with impulsive inputs than the latter case. This is because the existence of the impul-

Table 1. Filter parameters and gains used to generate the sequence x(n).

Index	$\hat{w}_{1,k}(n)$	$\hat{w}_{2,k}(n)$	Gain, g_k
1	-1.3125	0.75	100
2	-0.8525	0.55	80
3	-0.56	0.4	60
4	-0.39	0.3	40
5	-0.3125	0.25	20

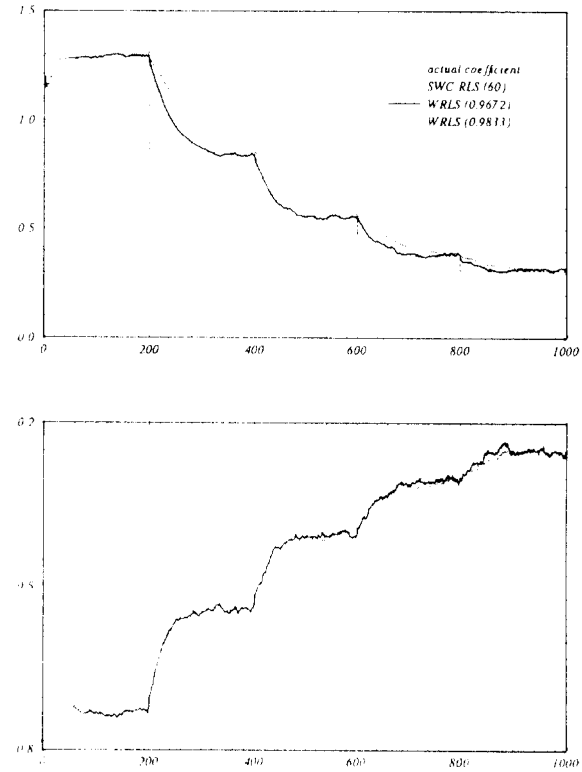

Figure 1. Parameter trajectories of $\hat{w}_1(n)$ and $\hat{w}_2(n)$, when the excitation signal is pure Gaussian noise sequence (without impulsive inputs).

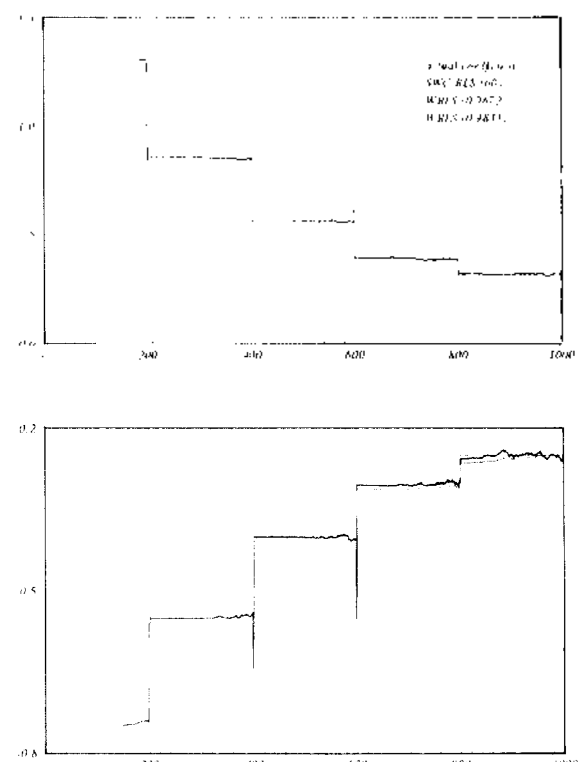

Figure 2. Parameter trajectories of $\hat{w}_1(n)$ and $\hat{w}_2(n)$, when the excitation signal is Gaussian noise sequence plus an impulsive inputs.

sive inputs actually reduce the size of the strip given on the Eq. (24).

It could have been expected from the persistent excitation condition in Eq. (13). When m is fixed, smaller N implies faster convergence rate. Therefore, larger gain would decrease the value of N to satisfy this condition, and hence the convergence becomes much quicker. Therefore, we can conclude that the impulsive inputs enhances the convergence property of the RLS algorithm by increasing the degree of persistent excitation of the input signal.

4. Self Noise

Since the self noise is an internal problem from using an exponential windowing, it is dependent only on the forgetting factor (or window length) and the input signal statistics. The excess MSE due to this self noise can be written by $E[\tilde{\mathbf{w}}^T(n)\mathbf{R}\cdot\tilde{\mathbf{w}}(n)]$. Note that $\mathbf{R}$, $\mathbf{p}$, and $\mathbf{w}$ are assumed to be independent of time for the analysis of self noise. From the normal equations,

$$\tilde{\mathbf{w}}(n)=\hat{\mathbf{R}}^{-1}(n)\sum_{k=1}^{n}\lambda^{n-k}\mathbf{x}(k-1)e(k) \tag{26}$$

Then the variance can be approximated as

$$E[\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^T(n)]\cong\frac{\sigma_0^2\sum_{k=1}^{n}\lambda^{2(n-k)}}{\left(\sum_{k=1}^{n}\lambda^{n-k}\right)^2}\mathbf{R}^{-1} \tag{27}$$

Here, large value of n is assumed, so that, by the law of large numbers

$$\hat{\mathbf{R}}(n)\cong\sum_{k=1}^{n}\lambda^{n-k}(\mathbf{R}+\delta\mathbf{R}_1) \tag{28}$$

and

$$\sum_{k=1}^{n}\lambda^{2(n-k)}\mathbf{x}(k-1)\mathbf{x}^T(k-1)\cong\sum_{k=1}^{n}\lambda^{2(n-k)}(\mathbf{R}+\delta\mathbf{R}_2) \tag{29}$$

where the elements of $\delta\mathbf{R}_1$ and $\delta\mathbf{R}_2$ have zero means and small variations compared to the elements of $\mathbf{R}$. It is also assumed that $\|\delta\mathbf{R}_1\| << \|\mathbf{R}^{-1}\|^{-1}$ for some consistent matrix norm, and ignore high order terms to get

$$\begin{aligned}(\mathbf{R}+\delta\mathbf{R}_1)^{-1}&=\mathbf{R}^{-1}+\mathbf{R}^{-1}(\mathbf{R}^{-1}+\delta\mathbf{R}_1^{-1})\mathbf{R}^{-1}\\&\cong\mathbf{R}^{-1}+\mathbf{R}^{-1}\delta\mathbf{R}_1^{-1}\mathbf{R}^{-1}\end{aligned} \tag{30}$$

Now, we have

$$\Omega_{self}(n)=tr\{\mathbf{R}E[\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^T(n)]\}\cong\frac{1-\lambda}{1+\lambda}\,p\sigma_0^2 \tag{31}$$

Eq. (31) provides a quantitative insight to the excess MSE due to the exponential nature of the RLS estimators. Note that if $\lambda=1$ the excess MSE due to the self noise vanishes.

5. Excitation Error

Assume that the true tap coefficient vector undergoes a step change at time T_0+1 from $\mathbf{w}+\delta\mathbf{w}$ to $\mathbf{w}$, followed by the corresponding changes in the autocorrelation matrix and the cross-correlation vector from $\mathbf{R}+\delta\mathbf{R}$ and $\mathbf{p}+\delta\mathbf{p}$ to $\mathbf{R}$ and $\mathbf{p}$, respectively. For $n\le T_0$, $\hat{\mathbf{R}}(n)$ and $\hat{\mathbf{p}}(n)$ are assumed to have been adjusted according to the old signal characteristics. Thus, at time T_0,

$$E[\hat{\mathbf{R}}(T_0)]=\frac{1-\lambda^{T_0+1}}{1-\lambda}(\mathbf{R}+\delta\mathbf{R}) \tag{32}$$

and

$$E[\hat{\mathbf{p}}(T_0)]=\frac{1-\lambda^{T_0+1}}{1-\lambda}(\mathbf{p}+\delta\mathbf{p}) \tag{33}$$

Using the same technique as in Eq. (30), we have

$$E[\hat{\mathbf{w}}(T_0)]\cong\mathbf{w}+\delta\mathbf{w} \tag{34}$$

where

$$\delta\mathbf{w}=-\mathbf{R}^{-1}\delta\mathbf{R}\mathbf{R}^{-1}\mathbf{p}+\mathbf{R}^{-1}\delta\mathbf{p} \tag{35}$$

and $\mathbf{w}=\mathbf{R}^{-1}\mathbf{p}$. Note that it is further assume $\hat{\mathbf{R}}(n)$ be orthogonal to $\mathbf{w}(n)$.

In Eq. (34), the high order term $\mathbf{R}^{-1}\delta\mathbf{R}\mathbf{R}^{-1}\delta\mathbf{p}$ is ignored. As we note, an extra error occurs for $n>T_0$, because the algorithm cannot adjust the coefficient vector from $\mathbf{w}+\delta\mathbf{w}$ to $\mathbf{w}$ instantly. The excess MSE error due to the excitation error defined by $\Omega_{exc}(n)=E[\tilde{\mathbf{w}}^T(n)\mathbf{R}\tilde{\mathbf{w}}(n)]$ is equal to, at time $n=T_0+1$

$$\Omega_{exc}(T_0+1)=\delta\mathbf{w}^T\mathbf{R}\delta\mathbf{w} \tag{36}$$

For $n>T_0$, we have

$$E[\hat{\mathbf{R}}(n)]\cong\frac{\mathbf{R}+\lambda^{n-T_0}\delta\mathbf{R}}{1-\lambda} \tag{37}$$

and

$$E[\mathbf{p}(n)]\cong\frac{\mathbf{p}+\lambda^{n-T_0}\delta\mathbf{p}}{1-\lambda} \tag{38}$$

Ignoring the higher order term during calculation, we have

$$E[\hat{\mathbf{w}}(n)] = \mathbf{w} + \lambda^{n-T_0}\delta\mathbf{w} \quad (39)$$

so that

$$\Omega_{exc}(n) = \lambda^{2(n-T_0)}\delta\mathbf{w}^T\mathbf{R}\delta\mathbf{w} \cong e^{2(1-\lambda)(n-T_0)}\Omega_{exc}(T_0+1) \quad (40)$$

This gives the time constant equal to $1/2(1-\lambda)$.

When a periodic pulse is added to the white-noise-driving process at the instant of coefficient change, the transient behavior of the exponential RLS algorithm becomes different. Assume that an impulse of amplitude g_0 is applied to the filter synchronized to the step change of parameter vector at time T_0+p+1. After the irregular effect of the impulsive input at the interval $[T_0+1, T_0-p]$, the bias in the tap weight vector at time T_0+p is given approximately by Eq. (12). Then the initial value of the excitation error will be

$$\Omega_{exc}(T_0+p) = E[\tilde{\mathbf{w}}^T(T_0+p)\mathbf{R}\tilde{\mathbf{w}}(T_0+p)] \quad (41)$$

If we assume that the magnitude of the impulsive input is large enough, then we have $\|\hat{\mathbf{R}}(T_0+p)\| > \|\hat{\mathbf{R}}(T_0)\|$ for some consistent matrix norm so that

$$\|\lambda^p\hat{\mathbf{R}}^{-1}(T_0+p)\hat{\mathbf{R}}(T_0)\delta\mathbf{w}\| < \lambda^p\|\hat{\mathbf{R}}^{-1}(T_0+p)\hat{\mathbf{R}}(T_0)\|\|\delta\mathbf{w}\| < \|\delta\mathbf{w}\| \quad (42)$$

From Eq. (40) it has been shown that the lag error decays exponentially with the time constant $1/2(1-\lambda)$ and the initial condition $\delta\mathbf{w}\mathbf{R}\delta\mathbf{w}$. Also, Eq. (42) shows that the impulsive input scales the initial value of the lag error approximately by the amount $\lambda^{2p}\|\hat{\mathbf{R}}^{-1}(T_0+p)\hat{\mathbf{R}}(T_0)\delta\mathbf{w}\|^2$. Therefore, together with the result in the convergence analysis (see Eq. (24)), we can conclude that an impulsive component synchronized to the parameter step change not only enhances the convergence characteristics, but also decrease the excitation error in a substantial degree.

Ⅲ. Conclusion

In this work, the performance of the exponential RLS algorithm is analyzed when it is applied to speech analysis. The existence of a peculiar source of excess MSE, denoted as the excitation error, is discussed when the exponential RLS algorithm is applied to estimate abruptly changing signal parameters. For the analysis of the excitation error, a deterministic is first taken. Then, the convergence property of the exponential RLS algorithm for this particular situation is investigated, followed by the study on the excess MSE due to the excitation error. It has been theoretically shown in this paper that the impulsive input (or error variance) synchronous to the abrupt change of parameter vectors actually enhances the convergence of the RLS algorithm. The analysis has also been simulated on the synthetic speech signal. This result would provide a guide in understanding the reliability of parameter estimates provided by the exponential RLS algorithm as well as in the design of a VFF-RLS algorithm.

References

1. G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Prentice Hall; Englewood Cliffs, New Jersey, 1984.
2. A. O. Cordero and D. Q. Mayne, "Deterministic convergence of a self-tuning regulator with variable forgetting factor," *IEE Proc.* Vol.128, Pt.D, No.1, pp.19-23, Jan. 1981.
3. G. E. Bottomley and S. T. Alexander, "A theoretical basis for the divergence of conventional recursive least squares filters," *Proc. '89 ICASSP*, Glasgow, Scotland, pp.908-911, 1989.
4. T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors," *Automatica*, Vol.17, pp.831,-835, 1981.
5. B. Widrow et. al., "Stationary and non-stationary learning characteristics of the LMS algorithm," *Proc. IEEE*, Vol.64 (8), pp.1151-1162, 1976.
6. E. Eleftheriou and D. D. Falconer, "Tracking properties and steady state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust. Speech Signal Process.*, Vol.ASSP-86, No.5, pp.1097-1109, 1986.
7. Q. T. Zhang and S. Haykin, "Tracking characteristics of the Kalman filter in a nonstationary environment for adaptive filter applications," *Proc. '83 ICASSP*, Boston, USA, pp. 671-674, 1983.
8. F. Ling and J. G. Proakis, "Nonstationary learning characteristics of least squares adaptive estimation algorithms," *Proc. '84 ICASSP*, San Diego, Cali., pp.30.3.1-4, 1984.
9. O. M. Macci and N. J. Bershad, "Adaptive recovery of a chirped sinusoid in noise, Part I: Performance of the RLS algorithm," *IEEE Trans. Signal Processing*, Vol.SP-39, No.3, pp.571-582, 1991.
10. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall; Englewood Cliffs, New Jersey, 1978.
11. J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, Vol.63, pp.561-580, 1975.
12. G. D. Gibson, J. L. Melsa, and S. K. Jones, "Digital speech analysis using sequential estimation techniques," *IEEE Trans. Acoust. Speech Signal Process.*, Vol.ASSP-27, No.4, pp.362-369, Aug. 1979.
13. X. Huang, G. Duncan, and M. Jack, "Formant estimation

system based on weighted least-squares lattice filter," IEE Proc. Vol.135, Pt.F, No.6, pp.539-546, Dec. 1988.
14. L. R. Rabiner, B. S. Atal, and M. R. Sambur, "LPC prediction error-Analysis of its variation with the position of the analysis frame," *IEEE Trans. Acoust. Speech Signal Process.*, Vol.ASSP-25, No.5, pp.434-442, Oct. 1977.
15. T. Hägglund, "Adaptive control of systems subject to large parameter changes," *Proc. 9th IFAC World Congress*, Budapest, Hungary, pp.993-998, 1984.
16. R. R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. on Info. Theory*, Vol.IT-30, No.2, pp.183-191, Mar. 1984.
17. R. Lozano, "Identification of time-variable linear models," Proc. 22nd CDC, San Antonio, Texas, pp.604-606, 1983.
18. S. Haykin, *Adaptive Filter Theory, 2nd Ed.*, Prentice Hall; Englewood Cliffs, New Jersey, 1991.

Appendix

Here we shall show upper- and lower-bound on the covariance matrix. Eq. (2) can be written as

$$\mathbf{P}^{-1}(n)=\lambda^{n}\mathbf{P}_0^{-1}+\sum_{i=1}^{n}\lambda^{n-i+1}\mathbf{x}(i-1)\mathbf{x}^{T}(i-1) \qquad \text{(A1)}$$

where $\mathbf{P}_0^{-1}=\mathbf{P}^{-1}(0)=\delta\mathbf{I}$ is the initial condition of the covariance matrix which is usually chosen to be a large number of δ. Since the covariance matrix $\mathbf{P}(n)$ is symmetric and positive definite, it satisfies

$$\lambda_{min}(\mathbf{P}^{-1}(n))\geq\lambda_{min}\left(\sum_{i=n-N+1}^{n}\lambda^{n-i+1}\mathbf{x}(i-1)\mathbf{x}^{T}(i-1)\right)+\cdots$$

$$+\lambda_{min}\left(\sum_{i=t-rN+1}^{t-(r-1)N}\lambda^{n-i+1}\mathbf{x}(i-1)\mathbf{x}^{T}(i-1)\right) \qquad \text{(A2)}$$

where $\lambda_{min}(A)$ denotes the minimum eigenvalue of the square matrix A and $r=[n/N]$ denotes the integer part of the quotient. Combining this with the persistent excitation condition, we can find an upper bound for $\mathbf{P}(n)$ given by;

$$\lambda_{min}(\mathbf{P}^{-1}(n))\geq m(\lambda^{N}+\lambda^{2N}+\cdots+\lambda^{rN})+\lambda_{min}(\lambda^{n}\mathbf{P}_0^{-1}) \qquad \text{(A3)}$$

Similarly, the lower bound can also be found from Eq. (27), which results

$$\left(\lambda^{n}\lambda_{min}(\mathbf{P}_0^{-1})+M\lambda\frac{1-\lambda^{(r+1)N}}{1-\lambda}\right)^{-1}\mathbf{I}\leq\mathbf{P}(n)\leq$$

$$\left(\lambda^{n}\lambda_{min}(\mathbf{P}_0^{-1})+m\lambda^{N}\frac{1-\lambda^{rN}}{1-\lambda^{N}}\right)^{-1}\mathbf{I} \qquad \text{(A4)}$$

for all $t>0$. Therefore, for $n>>1$, Eq. (A4) leads to Eq. (14).

▲Kyung Yul Yoo

Kyung Yul Yoo received the B.E. degree in electrical engineering from Hanyang university, Korea, in 1982, the MS degree in electrical engineering from the University of Arkansas, U.S.A, in 1987, the MA degree in mathematics from the University of Arkansas, U.S.A, in 1989, and the Ph.D. degree in electrical engineering from the University of Missouri-Rolla, U.S.A. in 1993.

From 1993 to 1995, he had been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, as a senior member of technical staff, developing various digital communication systems. Since 1995, he has been with the department of electrical engineering, Hanyang university, Korea, as a full-time lecturer. His current interest include statistical signal processing and adaptive filtering algorithms.