Adaptive Tracking of Linear Time-Variant Systems by Extended RLS Algorithms

Simon Haykin, *Fellow, IEEE*, Ali H. Sayed, *Member, IEEE*, James R. Zeidler, *Fellow, IEEE*, Paul Yee, *Student Member, IEEE*, and Paul C. Wei, *Member, IEEE*

***Abstract*— In this paper, we exploit the one-to-one correspondences between the recursive least-squares (RLS) and Kalman variables to formulate extended forms of the RLS algorithm. Two particular forms of the extended RLS algorithm are considered: one pertaining to a system identification problem and the other pertaining to the tracking of a chirped sinusoid in additive noise. For both of these applications, experiments are presented that demonstrate the tracking superiority of the extended RLS algorithms compared with the standard RLS and least-mean-squares (LMS) algorithms.**

***Index Terms*— Extended RLS algorithms, the Kalman filter, LMS algorithm, RLS algorithm, tracking performance.**

I. Introduction

THE LEAST-MEAN-SQUARE (LMS) algorithm [1], [2] and the recursive least-squares (RLS) algorithm [2] have established themselves as the principal tools for linear adaptive filtering. The convergence behaviors of both of these algorithms are now well understood [2], [3]. Typically, the RLS algorithm has a faster rate of convergence than the LMS algorithm and is not sensitive to variations in the eigenvalue spread of the correlation matrix of the input vector. However, when operating in a nonstationary environment, the adaptive filter has the additional task of tracking the statistical variations in environmental conditions. In this context, it is well recognized that the convergence behavior of an adaptive filter is a transient phenomenon, whereas its tracking behavior is a steady-state phenomenon. This means that, in general, a good convergence behavior does not necessarily translate into a good tracking behavior.

In recent years, much has been written on a comparative evaluation of the tracking behaviors of the LMS and RLS algorithms [4]–[7]. The general conclusion drawn from the studies reported in the literature to date is that typically, the LMS algorithm exhibits a better tracking behavior than the RLS algorithm. This conclusion should not be surprising since the LMS algorithm is model independent, whereas the RLS algorithm is model dependent. Unless the multiparameter regression model assumed in the derivation of the standard RLS algorithm closely matches the underlying model of the environment in which it operates, we would expect a degradation in the tracking performance of the RLS algorithm due to a model mismatch.

In a recent paper, Sayed and Kailath [8] delineated the relationship between the RLS algorithm and the Kalman filter in precise terms. Although work on this relationship may be traced back to the seminal paper by Godard [9] and subsequently elaborated on by many other investigators, the exact nature of the relationship was put on a firm footing for the first time in [8]. Thus, recognizing that the RLS algorithm is a special case of the Kalman filter and recognizing that the Kalman filter is the optimum linear tracking device on the basis of second-order statistics, how then is it that the exponentially weighted RLS algorithm has *not* inherited the good tracking behavior of the Kalman filter? The answer to this fundamental question lies in the fact that in formulating the standard form of the RLS algorithm by incorporating an exponential weighting function into the cost function, the transition matrix of the RLS algorithm (using the language of Kalman filter theory) is, in reality, a constant, which is clearly not the way to solve the tracking problem for a nonstationary environment.

The purpose of this paper is twofold. First, we describe two different methods for the design of an RLS-type algorithm to cope with corresponding forms of nonstationary environmental conditions. Second, we present computer experiments: one on system identification assuming a first-order Markov model and the other on tracking a chirped sinusoid in additive noise. The experiments demonstrate the tracking superiority of the extended RLS algorithm(s) over the LMS algorithm when the right model for the RLS algorithm is chosen to suit the particular problem of interest.

II. The Standard RLS Algorithm and Extensions

According to Sayed and Kailath [8], a state-space model for the exponentially weighted RLS algorithm may be described as follows:

$$\begin{aligned} \boldsymbol{x}(n+1) &= \lambda^{-1/2}\boldsymbol{x}(n), \\ y(n) &= \boldsymbol{u}^H(n)\boldsymbol{x}(n) + v(n) \end{aligned} \tag{1}$$

Manuscript received May 1, 1995; revised September 16, 1996. This work was supported in part by a grant from the National Science Foundation under Award MIP-9409319. The associate editor coordinating the review of this paper and approving it for publication was Prof. Tyseer Aboulnasr.

S. Haykin and P. Yee are with the Communications Research Laboratory, McMaster University, Hamilton, Ont., Canada L8S 4K1A (e-mail: haykin@synapse.mcmaster.ca; yee@soma.mcmaster.ca).

A. H. Sayed is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

J. R. Zeidler is with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA (e-mail: zeidler@nosc.mil).

P. C. Wei is with LSI Logic, Carlsbad, CA 92008 USA (e-mail: pwei@lsil.com).

Publisher Item Identifier S 1053-587X(97)03344-8.

where

$\boldsymbol{x}(n)$	state vector of the model at time (iteration) n,
$y(n)$	observation (reference) signal,
$\boldsymbol{u}(n)$	input signal vector,
$v(n)$	white-noise measurement noise with unit variance,
λ	forgetting factor.

The superscript H denotes Hermitian conjugation (complex conjugation for scalars).

Two observations from the state equation, i.e., the first line of (1), are immediately apparent:

- The transition matrix of the standard RLS algorithm is a constant multiple of the identity matrix equal to $\lambda^{-1/2}\boldsymbol{I}$.
- The process (state) noise vector is zero.

Now, both of these conditions are synonymous with a stationary environment. Thus, although it is widely believed in the literature that by introducing the forgetting factor λ into the design of the RLS algorithm, the algorithm is enabled to track statistical variations of the environment, in reality, this is not so. It is therefore not surprising that the RLS algorithm, in its standard form, does not always measure up to the LMS algorithm when it comes to tracking considerations.

Kalman filter theory tells us that a more general form of the state-space model of the RLS algorithm should be as follows:

$$\begin{aligned}\boldsymbol{x}(n+1) &= \boldsymbol{F}(n+1,n)\boldsymbol{x}(n) + \boldsymbol{r}(n),\\ y(n) &= \boldsymbol{u}^H(n)\boldsymbol{x}(n) + v(n).\end{aligned} \tag{2}$$

The measurement equation, i.e., second line of (2), is the same as before. However, the state equation, i.e., first line of (2), differs from that of (1) in two aspects:

- The transition matrix $\boldsymbol{F}(n+1,n)$ is time variant.
- The process (state) noise vector $\boldsymbol{r}(n)$ is nonzero.

This, therefore, points to two special ways in which the RLS algorithm may be modified in order to cope with different nonstationary environments, as explained in the next two sections; in one case, we assume $\boldsymbol{F}(n+1,n)$ is *known* and present the proper extension of the RLS solution (referred to here as ERLS-1). We then apply the algorithm to a system identification problem assuming a Markov model. In the other case, we assume $\boldsymbol{F}(n+1,n)$ is *not known* and proceed to suggest a second extension of the RLS solution by invoking connections with extended Kalman filtering (the extension is referred to as ERLS-2). The algorithm is then applied to tracking a chirped sinusoid in additive noise.

The fundamental point to stress here is that in both cases, prior knowledge about the original dynamical system model is explicitly built into the formulation of the extended forms of the RLS algorithm, thus improving the tracking performance of the resulting adaptive filters.

III. A System Identification Problem

Consider a linear time-variant system described by a first-order Markov model. Specifically, we have the following pair of equations as the system description:

$$\begin{aligned}\boldsymbol{w}_o(n+1) &= \boldsymbol{F}(n+1,n)\boldsymbol{w}_o(n) + \boldsymbol{r}(n),\\ y(n) &= \boldsymbol{u}^H(n)\boldsymbol{w}_o(n) + v(n)\end{aligned} \tag{3}$$

where

$\boldsymbol{F}(n+1,n)$	known transition matrix,
$\boldsymbol{w}_o(n)$	optimum tap-weight vector of the model at time n,
$\boldsymbol{r}(n)$	process noise,
$y(n)$	desired response,
$\boldsymbol{u}(n)$	input vector,
$v(n)$	measurement noise.

A special case of interest in the simulation examples of Section VI-A is when $\boldsymbol{F}(n+1,n) = a\boldsymbol{I}$, which is a constant multiple of the identity matrix, where a is further assumed to be less than unit magnitude in order to assure the stability of the model.

In the system identification problem described herein, we are given the input vectors $\{\boldsymbol{u}(n)\}$, the desired response $\{y(n)\}$, positive-definite matrices $\{\Pi_0, \boldsymbol{Q}(n)\}$, positive numbers $\{\sigma(n)\}$, and an initial guess $\overline{\boldsymbol{w}}_o$. The requirement is to estimate the unknown weight vector $\boldsymbol{w}_o(n)$ and to track its variation with time n.

This problem can be related to Kalman filter theory by posing the following optimization criterion:

$$\min_{\{\boldsymbol{w}_o(0),\boldsymbol{r}(n)\}_{n=0}^{N}} J[\boldsymbol{w}_o(0), \boldsymbol{r}(0), \boldsymbol{r}(1), \cdots, \boldsymbol{r}(n)] \tag{4}$$

subject to the state-equation constraint

$$\boldsymbol{w}_o(n+1) = \boldsymbol{F}(n+1,n)\boldsymbol{w}_o(n) + \boldsymbol{r}(n)$$

where the cost function J is quadratic in its arguments and is given by

$$\begin{aligned}J = {} & (\boldsymbol{w}_o(0) - \overline{\boldsymbol{w}}_o)^H \Pi_0^{-1} (\boldsymbol{w}_o(0) - \overline{\boldsymbol{w}}_o)\\ & + \sum_{n=0}^{N} \boldsymbol{r}^H(n)\boldsymbol{Q}^{-1}(n)\boldsymbol{r}(n)\\ & + \sum_{n=0}^{N} \frac{|y(n) - \boldsymbol{u}^H(n)\boldsymbol{w}_o(n)|^2}{\sigma^2(n)}.\end{aligned}$$

The unknown quantities in the above quadratic cost function are the initial weight-vector $\boldsymbol{w}_o(0)$ and the process noise sequence $\{\boldsymbol{r}(i)\}_{i=0}^{N}$,

$$\{\boldsymbol{w}_o(0), \boldsymbol{r}(0), \boldsymbol{r}(1), \cdots, \boldsymbol{r}(N)\}.$$

The solution of (4) can be shown (e.g., [8], [10], [11]) to lead to an iterative procedure that provides recursive estimates of the successive weight vectors $\boldsymbol{w}_o(n)$, which are denoted by $\hat{\boldsymbol{w}}(n|n-1)$, and it can be regarded as the Kalman filter corresponding to model (3) with the following (statistical) assumptions on the noise sequences:

- $\{\boldsymbol{r}(n)\}$ is assumed a zero-mean white noise sequence with covariance matrix $\boldsymbol{Q}(n), E\boldsymbol{r}(n)\boldsymbol{r}^H(n) = \boldsymbol{Q}(n)$.
- $\{v(n)\}$ is assumed a zero-mean white noise sequence with variance $\sigma^2(n), Ev(n)v^H(n) = \sigma^2(n)$.

- The initial state-vector $\boldsymbol{w}_o(0)$ is assumed random with mean $\overline{\boldsymbol{w}}_o$ and covariance matrix Π_0

$$E(\boldsymbol{w}_o(0)-\overline{\boldsymbol{w}}_o)(\boldsymbol{w}_o(0)-\overline{\boldsymbol{w}}_o)^H=\Pi_0.$$

- The random variables $\{\boldsymbol{r}(n), v(n), (\boldsymbol{w}_o(0)-\overline{\boldsymbol{w}}_o)\}$ are assumed uncorrelated.

We may rewrite these conditions, more compactly, as follows:

$$E\begin{bmatrix}\boldsymbol{r}(n)\\ v(n)\\ \boldsymbol{w}_o(0)-\overline{\boldsymbol{w}}_o\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{r}(m)\\ v(m)\\ \boldsymbol{w}_o(0)-\overline{\boldsymbol{w}}_o\end{bmatrix}^H = \begin{bmatrix}\boldsymbol{Q}(n)\delta(n,m) & 0 & 0\\ 0 & \sigma^2(n)\delta(n,m) & 0\\ 0 & 0 & \Pi_0\\ 0 & 0 & 0\end{bmatrix}$$

where $\delta(n,m)$ is the Kronecker delta function, which is equal to unity when $n=m$ and zero otherwise.

Building on the classical Kalman filter theory and exploiting the one-to-one correspondences that exist between Kalman variables and RLS variables [8], the RLS algorithm appropriate for the task (4) is the following so-called ERLS-1 solution:

Algorithm 1 (Extended RLS Solution—Version I)

The estimates of the weight-vector $\boldsymbol{w}_o(n)$ in (3), computed in the process of solving the optimization criterion given by (4), can be recursively evaluated as follows: Start with $\hat{\boldsymbol{w}}(0|-1)=\overline{\boldsymbol{w}}_o, \boldsymbol{P}(0,-1)=\Pi_0$, and repeat for $n\geq 0$:

$$\begin{aligned}\boldsymbol{k}(n)&=\boldsymbol{F}(n+1,n)\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)r_e^{-1}(n),\\ r_e(n)&=\boldsymbol{u}^H(n)\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)+\sigma^2(n),\\ \xi(n)&=y(n)-\boldsymbol{u}^H(n)\hat{\boldsymbol{w}}(n|n-1),\\ \hat{\boldsymbol{w}}(n+1|n)&=\boldsymbol{F}(n+1,n)\hat{\boldsymbol{w}}(n|n-1)+\boldsymbol{k}(n)\xi(n),\\ \boldsymbol{P}(n)&=\boldsymbol{P}(n,n-1)\\ &\quad-\frac{\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)\boldsymbol{u}^H(n)\boldsymbol{P}(n,n-1)}{r_e(n)},\\ \boldsymbol{P}(n+1,n)&=\boldsymbol{F}(n+1,n)\boldsymbol{P}(n)\boldsymbol{F}^H(n+1,n)+\boldsymbol{Q}(n)\end{aligned}$$

where

$\boldsymbol{k}(n)$	gain vector,
$\xi(n)$	*a priori* estimation error,
$\hat{\boldsymbol{w}}(n\|n-1)$	estimate of the unknown $\boldsymbol{w}_o(n)$ given the input data up to time $(n-1)$.

In the Kalman filtering context, the matrix $\boldsymbol{P}(n,n-1)$ is the covariance matrix of the predicted weight-error vector

$$\begin{aligned}\tilde{\boldsymbol{w}}(n,n-1)&=\boldsymbol{w}_o(n)-\hat{\boldsymbol{w}}(n|n-1),\\ \boldsymbol{P}(n,n-1)&=E\tilde{\boldsymbol{w}}(n,n-1)\tilde{\boldsymbol{w}}^H(n,n-1).\end{aligned}$$

In a stationary environment, the covariance matrix $\boldsymbol{Q}(n)$ is zero for all n, in which case, $\boldsymbol{P}(n+1,n)=\boldsymbol{P}(n)$, and the modified RLS algorithm reduces to its standard form (without exponential weighting). Under this condition, $\boldsymbol{P}(n)$ equals the inverse of the deterministic correlation matrix $\Phi(n)$ of the input vector:

$$\Phi(n)=\sum_{n=0}^{N}\boldsymbol{u}^H(n)\boldsymbol{u}(n).$$

In the special case $\boldsymbol{F}(n+1,n)=a\boldsymbol{I}$, with $\boldsymbol{Q}(n)=q^2\boldsymbol{I}$, and $\sigma^2=1$, the above recursions take the following form:

$$\begin{aligned}\boldsymbol{k}(n)&=a\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)r_e^{-1}(n),\\ r_e(n)&=\boldsymbol{u}^H(n)\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)+1,\\ \xi(n)&=y(n)-\boldsymbol{u}^H(n)\hat{\boldsymbol{w}}(n|n-1),\\ \hat{\boldsymbol{w}}(n+1|n)&=a\hat{\boldsymbol{w}}(n|n-1)+\boldsymbol{k}(n)\xi(n),\\ \boldsymbol{P}(n)&=\boldsymbol{P}(n,n-1)\\ &\quad-\frac{\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)\boldsymbol{u}^H(n)\boldsymbol{P}(n,n-1)}{r_e(n)},\\ \boldsymbol{P}(n+1,n)&=|a|^2\boldsymbol{P}(n)+q^2\boldsymbol{I}.\end{aligned}$$

This algorithm has a single variable parameter, namely, q^2. Note also that this algorithm is the same as that described in [12], except that in that paper, a was taken as $a\approx 1$, and the unity term in the conversion factor $[\boldsymbol{u}^H(n)\boldsymbol{P}(n,n-1)\boldsymbol{u}(n)+1]$ was replaced by the minimum mean-squared error. This difference is attributed to a formulation of the correspondences between the Kalman variables and RLS variables that is different from the one derived in [8] and on which the extended version of the RLS algorithm (Version I above) is based. We may finally add that the computational complexity of the above algorithm (for $\boldsymbol{F}(n+1,n)=a\boldsymbol{I}$) is the same as the standard RLS recursion, viz., $O(M^2)$ operations per iteration for a weight vector of length M.

IV. Tracking of a Chirped Sinusoid in Noise

We now study a second example of a nonstationary environment, which arises while tracking chirped sinusoids in additive noise. Such an approach is frequently used to model Doppler shifts. The state-space model of interest in this case takes the following form (see Section VI-B for more details):

$$\begin{aligned}\boldsymbol{w}_o(n+1)&=\boldsymbol{F}(\psi)\boldsymbol{w}_o(n),\\ y(n)&=\boldsymbol{u}^H(n)\boldsymbol{w}_o(n)+v(n)\end{aligned}\tag{5}$$

where

$\boldsymbol{w}_o(n)$	optimum tap-weight vector that we wish to estimate,
$v(n)$	measurement noise,
$\{y(n)\}$	noisy measurements,
$\{\boldsymbol{u}(n)\}$	input tap vectors,
$\boldsymbol{F}(\psi)$	*unknown* diagonal matrix that is fully parameterized in terms of a single *unknown* parameter ψ.

This parameter is related to the linear shift of the center frequency in the chirped signal and the dependence of $\boldsymbol{F}$ on

it is as follows:

$$\boldsymbol{F}(\psi) \triangleq \begin{bmatrix} e^{j\psi} & & & \\ & e^{j2\psi} & & \\ & & \ddots & \\ & & & e^{jM\psi} \end{bmatrix}.$$

Here, M is the size of the tap-weight vector. If the parameter ψ is known, then $\boldsymbol{F}(\psi)$ will be a known transition matrix, and a standard least-squares (RLS) problem results, the solution of which can be written down as a special case of the standard Kalman filter recursions, as we have explained in the previous section.

We may note that we can as well include a process (state) noise vector $\boldsymbol{r}(n)$ into the first line of (5) with obvious changes to the discussion that follows. Here, however, we shall proceed with (5) for illustrative purposes. We also note that the argument given below is general and applies to other forms of transition matrices $\boldsymbol{F}(\cdot)$ and not necessarily with a diagonal structure as above. All we require is that the matrix $\boldsymbol{F}(\cdot)$ be parameterized by a collection of parameters and that the dependence of $\boldsymbol{F}(\cdot)$ on these parameters is known.

In any case, returning to (5), we see that both $\boldsymbol{w}_o(n)$ and ψ are unknowns that we wish to estimate. Ideally, we may want to determine these estimates to meet the optimality criterion

$$\min_{\{\boldsymbol{w}_o(0),\psi\}} J[\boldsymbol{w}_o(0),\psi]$$

subject to $\boldsymbol{w}_o(n+1) = \boldsymbol{F}(\psi)\boldsymbol{w}_o(n)$, where the cost function $J[\boldsymbol{w}_o(0),\psi]$ is given by

$$\begin{aligned} J[\boldsymbol{w}_o(0),\psi] &= [\boldsymbol{w}_o(0) - \overline{\boldsymbol{w}}_o \quad \psi - \overline{\psi}]^H \\ &\quad \cdot \begin{bmatrix} \Pi_0^{-1} & 0 \\ 0 & \pi_0^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_o(0) - \overline{\boldsymbol{w}}_o \\ \psi - \overline{\psi} \end{bmatrix} \\ &\quad + \sum_{n=0}^{N} \frac{|y(n) - \boldsymbol{u}(n)\boldsymbol{w}_o(n)|^2}{\sigma^2(n)}. \end{aligned}$$

Here, the $\{\sigma^2(n)\}$ denote weighting coefficients that we are free to choose. For example, in the exponentially weighted recursive least-squares problem, the $\sigma(n)$ are taken as exponentials of a forgetting factor λ. They may also designate the noise variance. The quantities $\{\Pi_0, \pi_0\}$ are also given and positive, and the $\overline{\boldsymbol{w}}_o$ and $\overline{\psi}$ are initial guesses for $\boldsymbol{w}_o(n)$ and ψ, respectively. The values of $\{\Pi_0, \pi_0\}$ indicate how confident we are about the initial guesses $\{\overline{\boldsymbol{w}}_o, \overline{\psi}\}$.

The above cost function is not quadratic in the unknowns $\{\boldsymbol{w}_o(0), \psi\}$ since the matrix $\boldsymbol{F}(\cdot)$ is a nonlinear function of the ψ. We may proceed here in two ways. We may first collect enough data in order to estimate the unknown ψ and then solve a standard RLS (or Kalman filtering) problem using the estimate $\hat{\psi}$ in the defining relation for $\boldsymbol{F}(\cdot)$. This would lead to a batch solution, i.e., an off-line solution for the determination of the ψ. Alternatively, we may devise a recursive solution that would estimate both ψ and the tap-weight vector on-line. We follow this second alternative here and borrow on connections with an approximation technique that is often used in the context of extended Kalman filtering [2], [11], [13].

For this purpose, we collect the unknowns into an extended (state) vector, say

$$\boldsymbol{x}(n) \triangleq \begin{bmatrix} \boldsymbol{w}_o(n) \\ \psi \end{bmatrix}$$

and note that it satisfies the following *nonlinear* (state-space) model:

$$\begin{aligned} \boldsymbol{x}(n+1) &= \begin{bmatrix} \boldsymbol{F}(\psi)\boldsymbol{w}_o(n) \\ \psi \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}(\psi) & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \boldsymbol{x}(n) \\ &\triangleq f(\boldsymbol{x}(n)), \\ y(n) &= [\boldsymbol{u}^H(n) \quad 0]\boldsymbol{x}(n) + v(n) \end{aligned} \tag{6}$$

where $f(\cdot)$ is clearly a nonlinear functional of the state vector. Note also that the last entry of the state vector does not change with time and is equal to ψ.

If we denote the individual entries of the state vector $\boldsymbol{x}(n)$ by $\{x_{n,i}\}$ then the nonlinear state-equation shows that the entries of the $(M+1)$-dimensional vector $\boldsymbol{x}(n+1)$ are time updated as follows (here, we invoke the fact that $x_{n,M+1} = \psi = x_{n+1,M+1}$):

$$\begin{aligned} \boldsymbol{x}(n+1) &= \begin{bmatrix} x_{n+1,1} \\ x_{n+1,2} \\ \vdots \\ \hline x_{n+1,M+1} \end{bmatrix} \\ &= \begin{bmatrix} e^{jx_{n,M+1}} x_{n,1} \\ e^{j2x_{n,M+1}} x_{n,2} \\ \vdots \\ \hline e^{jMx_{n,M+1}} x_{n,M} \\ x_{n,M+1} \end{bmatrix} = f(\boldsymbol{x}(n)). \end{aligned}$$

Let $\hat{\boldsymbol{x}}(n|n)$ denote the estimate of the vector $\boldsymbol{x}(n)$ that is based on the data $\{y(\cdot)\}$ available up to time n. Its individual entries are partitioned as follows:

$$\hat{\boldsymbol{x}}(n|n) = \begin{bmatrix} \hat{x}_{n,1|n} \\ \hat{x}_{n,2|n} \\ \vdots \\ \hline \hat{x}_{n,M|n} \\ \hat{x}_{n,M+1|n} \end{bmatrix} \triangleq \begin{bmatrix} \hat{\boldsymbol{w}}(n|n) \\ \hline \hat{\psi}_{|n} \end{bmatrix}$$

where we have written $\hat{\psi}_{|n}$ instead of $\hat{x}_{n,M+1|n}$ since the last entry of $\boldsymbol{x}(n)$ is ψ. We have also written $\hat{\boldsymbol{w}}(n|n)$ to refer to the leading M entries of $\hat{\boldsymbol{x}}(n|n)$ since the leading entries of $\boldsymbol{x}(n)$ correspond to the tap-weight vector $\boldsymbol{w}_o(n)$.

The extended RLS solution linearizes the state equation (6) around $\hat{\boldsymbol{x}}(n|n)$ as follows: Differentiating $f(\cdot)$ with respect to $\boldsymbol{x}(n)$ and evaluating at $\hat{\boldsymbol{x}}(n|n)$ defines a state transition matrix $\boldsymbol{F}(n+1,n)$,

$$\boldsymbol{F}(n+1,n) \triangleq \left.\frac{\partial f}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}(n|n)}$$

which can be easily seen, in view of the diagonal structure of $\boldsymbol{F}(\psi)$, to be of the following form:

$$\boldsymbol{F}(n+1,n) = \begin{bmatrix} \boldsymbol{F}(\hat{\psi}_{|n}) & \dot{\boldsymbol{F}}(\hat{\psi}_{|n})\hat{\boldsymbol{w}}(n|n) \\ 0 & 1 \end{bmatrix}.$$

Here, $\boldsymbol{F}(\hat{\psi}_{|n})$ corresponds to evaluating $\boldsymbol{F}(\cdot)$ at $\hat{\psi}_{|n}$, whereas $\dot{\boldsymbol{F}}(\hat{\psi}_{|n})$ corresponds to evaluating the derivative of $\boldsymbol{F}(\cdot)$, with respect to ψ, at $\hat{\psi}_{|n}$. Note that

$$\dot{\boldsymbol{F}}(\psi)=\begin{bmatrix} je^{j\psi} & & & \\ & j2e^{j2\psi} & & \\ & & \ddots & \\ & & & jMe^{jM\psi} \end{bmatrix}.$$

This leads to the following *linearized* model at time n

$$\begin{aligned}\boldsymbol{x}(n+1)&\approx\boldsymbol{F}(n+1,n)\boldsymbol{x}(n),\\ y(n)&=[\boldsymbol{u}^H(n)0]\boldsymbol{x}(n)+v(n)\end{aligned}$$

and we are now in a position to write down the corresponding RLS solution by invoking the correspondence between the RLS and Kalman variables [8]. The resulting equations have the same fundamental structure as the expressions of Algorithm 1 and will be referred to as the ERLS-2 version.

Algorithm 2 (Extended RLS Solution—Version II)

The estimates of the weight vector in (5) can be recursively evaluated as follows: Start with $\hat{\boldsymbol{w}}(0|-1)=\overline{\boldsymbol{w}}_o, \hat{\psi}_{|-1}=\overline{\psi}$

$$\boldsymbol{P}(0,-1)=\begin{bmatrix}\Pi_0 & 0\\ 0 & \pi_0\end{bmatrix}$$

and repeat for $n\geq 0$:

$$\begin{aligned}\boldsymbol{k}(n)&=\boldsymbol{P}(n,n-1)\begin{bmatrix}\boldsymbol{u}(n)\\0\end{bmatrix}r_e^{-1}(n),\\ r_e(n)&=\sigma^2(n)+[\boldsymbol{u}^H(n)\quad 0]\boldsymbol{P}(n,n-1)\begin{bmatrix}\boldsymbol{u}(n)\\0\end{bmatrix},\\ \xi(n)&=y(n)-\boldsymbol{u}^H(n)\hat{\boldsymbol{w}}(n|n-1),\\ \begin{bmatrix}\hat{\boldsymbol{w}}(n|n)\\ \hat{\psi}_{|n}\end{bmatrix}&=\begin{bmatrix}\hat{\boldsymbol{w}}(n|n-1)\\ \hat{\psi}_{|n-1}\end{bmatrix}+\boldsymbol{k}(n)\xi(n),\\ \hat{\boldsymbol{w}}(n+1|n)&=\boldsymbol{F}(\hat{\psi}_{|n})\hat{\boldsymbol{w}}(n|n),\\ \boldsymbol{P}(n,n)&=(\boldsymbol{I}-\boldsymbol{k}(n)[\boldsymbol{u}^H(n)\quad 0])\boldsymbol{P}(n,n-1),\\ \boldsymbol{P}(n+1,n)&=\boldsymbol{F}(n+1,n)\boldsymbol{P}(n,n)\boldsymbol{F}^H(n+1,n).\end{aligned}$$

The computational complexity of the above algorithm is again of the same order as the standard RLS recursion, viz., $O(M^2)$ operations per iteration.

V. Stability Considerations

Unlike the LMS and standard RLS algorithms, it is not possible to formulate conditions of a generic nature for which the stability of an extended RLS algorithm (say ERLS-1) is assured. Nevertheless, there are some useful results in the literature on Kalman filters that we may consider. In particular, we may invoke the following theorem [14]–[16].

Theorem 1: If the linear dynamical model, on which an extended RLS algorithm is based [say, model (3)], is stochastically observable and stochastically controllable, then the extended RLS algorithm is uniformly asymptotically globally stable.

In other words, as the number of iterations n increases without bound, the expected value of the squared Euclidean norm of the tap-weight error vector

$$\tilde{\boldsymbol{w}}(n,n-1)=[\boldsymbol{w}_o(n)-\hat{\boldsymbol{w}}(n|n-1)]$$

approaches zero, no matter what the initial conditions are. That is, the so-called *mean-square deviation*

$$D(n)=E[||\boldsymbol{w}_o(n)-\hat{\boldsymbol{w}}(n|n-1)||^2]$$

tends to zero. The theorem assumes that the vector $\boldsymbol{u}^H(n)$ in the model (3) is also stochastic. The statement therefore reflects a desired tracking capability.

It is important, however, to note that the sufficient conditions for stability of the extended RLS algorithm do not require stability of the original dynamical system [16]. We may thus identify four possible scenarios that can arise in practice:

1) The original dynamical system model and the extended RLS algorithm are both stable.
2) The original dynamical system model is unstable, but the extended RLS algorithm is stable.
3) The original dynamical system model is stable, but the extended RLS algorithm is unstable.
4) The original dynamical system model and the extended RLS algorithm are both unstable.

Clearly, only scenarios 1 and 2 correspond to a good tracking behavior. Scenario 2 is of noteworthy interest in that the extended RLS algorithm has the potential ability of tracking a linear dynamical system, even though the system is unstable (i.e., its state vector grows without bound). A similar remark applies to the standard RLS and LMS algorithms.

VI. Computer Experiments

We include here the results of several computer simulations that confirm the tracking superiority of the extended versions of the RLS algorithm.

A. System Identification

In this experiment, we consider the system identification of a simplified version of the first-order Markov model described in (3), viz.

$$\begin{aligned}\boldsymbol{w}_o(n+1)&=a\boldsymbol{w}_o(n)+\boldsymbol{r}(n),\\ y(n)&=\boldsymbol{u}^H(n)\boldsymbol{w}_o(n)+v(n),\qquad n\geq 0\end{aligned}\tag{7}$$

where, for all m and n

$$\begin{aligned}\overline{\boldsymbol{w}}_o(0)&=\boldsymbol{0},\quad \boldsymbol{u}(n)\sim N(\boldsymbol{0},\boldsymbol{R}),\\ E[\boldsymbol{u}(m)\boldsymbol{r}^H(n)]&=\boldsymbol{0},\quad E[\boldsymbol{u}(m)v^*(n)]=\boldsymbol{0},\\ \boldsymbol{r}(n)&\sim N(\boldsymbol{0},\boldsymbol{Q}),\quad v(n)\sim N(0,\sigma^2),\\ E[\boldsymbol{r}(m)v^*(n)]&=\boldsymbol{0}\end{aligned}$$

TABLE I
ABSOLUTE THEORETICAL PERFORMANCE OF RLS AND LMS ALGORITHMS

	Case 1: $\mathbf{R}^{-1} = c\mathbf{Q}$	Case 2: $\mathbf{R} = c\mathbf{Q}$
$D_{\min}^{\text{RLS}}$	$\sigma\sigma_Q^2\sqrt{c}\,(1+q_2)$	$\sigma\frac{1+q_2}{\sqrt{c(q_2-q_1^2)}}$
$D_{\min}^{\text{LMS}}$	$\sigma\sigma_Q^2\sqrt{2c\,(1+2q_1^2+q_2^2)}$	$2\sigma/\sqrt{c}$
$M_{\min}^{\text{RLS}}$	$2/\,(\sigma\sqrt{c})$	$\frac{\sigma_Q^2}{\sigma}\sqrt{2c\,(1+2q_1^2+q_2^2)}$
$M_{\min}^{\text{LMS}}$	$\frac{1+q_2}{\sigma\sqrt{c(q_2-q_1^2)}}$	$\frac{\sigma_Q^2}{\sigma}\sqrt{c}\,(1+q_2)$

TABLE II
RELATIVE THEORETICAL PERFORMANCE OF RLS AND LMS ALGORITHMS

	Case 1: $\mathbf{R}^{-1} = c\mathbf{Q}$	Case 2: $\mathbf{R} = c\mathbf{Q}$
$\frac{D_{\min}^{\text{RLS}}}{D_{\min}^{\text{LMS}}}$	$\frac{1+q_2}{\sqrt{2(1+2q_1^2+q_2^2)}}$	$\frac{1+q_2}{2\sqrt{q_2-q_1^2}}$
$\frac{M_{\min}^{\text{RLS}}}{M_{\min}^{\text{LMS}}}$	$\frac{2\sqrt{q_2-q_1^2}}{1+q_2}$	$\frac{\sqrt{2(1+2q_1^2+q_2^2)}}{1+q_2}$

and all vectors are M-dimensional. We shall restrict our attention to the case $M = 2$ for the following general form of $\boldsymbol{Q}$:

$$\boldsymbol{Q} = \sigma_Q^2\begin{bmatrix} 1 & q_1 \\ q_1 & q_2 \end{bmatrix}, \qquad |q_1| \le 1, q_2 > q_1^2$$

for the following two specific cases:

1) $\boldsymbol{R}^{-1} = c\boldsymbol{Q}$
2) $\boldsymbol{R} = c\boldsymbol{Q}$, both for $c > 0$.

Letting D represent the *mean-square deviation* and M the *relative mean-square misadjustment*, it can be shown [2], [6] that for each case, the results of Tables I and II hold.

Cases 1) and 2) are chosen because they represent a generalization of the example cases discussed in [2] that is sufficient to make their theoretical results hold. Our choices of the basic parameters for the experiments that follow are given in Table III. Note that although $\boldsymbol{Q}$ is constant throughout the experiments, c differs between experiments for cases 1) and 2). For a detailed discussion of these choices, refer to the Appendix.

As in [6], each result in Tables IV and V is obtained under the assumption of ergodicity of the instantaneous weight deviation $[\boldsymbol{w}_o(n) - \hat{\boldsymbol{w}}(n|n-1)]$ by time-averaging over one run of $N = 50\,000$ iterations in the steady state, i.e., after all transients have essentially dissipated; in the simulations, this is taken to occur at the iteration index $n = 50\,000$. The values of n and N so chosen can be justified by noting that plots of the simulated quantities as a function of N show no significant variation by that point.

The simulation results of Tables IV and V clearly show reasonable agreement between the experimentally and theoretically evaluated quantities of interest in both absolute and relative terms. Furthermore, the results demonstrate the superiority of the RLS algorithm over the LMS algorithm in case 1) and vice versa in case 2). This condition depends, of course, on the particular choice of experimental parameters, but what is constant is the reciprocal symmetry between cases 1) and 2) for the ratios $D_{\min}^{\text{RLS}}/D_{\min}^{\text{LMS}}$ and $M_{\min}^{\text{RLS}}/M_{\min}^{\text{LMS}}$ as given in Table II. In fact, following the analysis given in [2, ch. 16], it can be easily shown that this reciprocal symmetry holds generally in cases 1) and 2) for any positive definite covariance matrix $\boldsymbol{Q}$. If we had therefore chosen a set of parameters for which the RLS algorithm performs better than the LMS algorithm in terms of D in case 1), then it must be true that the LMS algorithm performs better than the RLS algorithm in terms of M in case 2), and vice versa.

TABLE III
BASIC PARAMETERS FOR EXPERIMENTS

σ_Q	q_1	q_2	σ	a	c in case 1	c in case 2
0.01	−0.75	1	0.2	0.9998	6.250×10^4	3.657×10^3

It is also interesting to note that the ERLS1 algorithm performs only marginally better than the optimal RLS/LMS algorithm in each case. Most likely, this situation is an artifact of the choice of experimental parameters; it makes both the relative mean-square weight deviation δ and relative mean-square misadjustment M sufficiently small so that differences between the performances of the algorithms are not easily discernible over what passes as normal simulation variance and numerical noise.

As a matter of interest, we have also repeated the experiments reported in [6], using the aforementioned algorithms. Here again, we have confirmed that the ERLS1 algorithm is the optimum linear tracker, irrespective of whether the LMS or the standard RLS is the better one of the two. We should add, however, that our experiments are designed from the outset to demonstrate the validity of the tracking theory under the meaningful tracking conditions described in Appendix A. For comparison, our simulations show a relative mean-square weight deviation δ of approximately 3% (cf. 2% theoretically), whereas those in [6] show approximately 35%. As a concrete example of the optimal tracking ability of the algorithms under this condition, we refer to Fig. 1, where we have plotted the first component $\hat{w}_1(n|n-1)$ of $\hat{\boldsymbol{w}}(n|n-1)$ over a range $n = 36\,001$ to $36\,500$. Again, we see that both of the optimally set algorithms track the actual optimal weight fairly well with little difference between them.

B. Tracking of Chirped Sinusoid

In this experiment, we consider the tracking of a chirped sinusoid. The deterministic shifts caused by the chirp represent the other extreme of the Markov model described in (2). The chirped input signal is given by

$$s(k) = \sqrt{P_s}e^{j(\omega+\psi k/2)k} \tag{8}$$

TABLE IV
ABSOLUTE EXPERIMENTAL PERFORMANCE OF RLS AND LMS ALGORITHMS (ERLS1 ALGORITHM RESULTS INSERTED FOR COMPARISON ONLY)

	Case 1: $\mathbf{R}^{-1} = c\mathbf{Q}$			Case 2: $\mathbf{R} = c\mathbf{Q}$		
	Exp. value	Theor. value	$\Delta\%$	Exp. value	Theor. value	$\Delta\%$
$D_{\min}^{\mathrm{RLS}}$	0.0103	0.01	+3.0%	0.0105	0.01	+5.0%
$D_{\min}^{\mathrm{LMS}}$	0.0135	0.0125	+8.0%	0.0071	0.0066	+7.6%
D^{ERLS1}	0.0102	–	–	0.0069	–	–
$M_{\min}^{\mathrm{RLS}}$	0.0423	0.04	+5.8%	0.0842	0.0756	+11.4%
$M_{\min}^{\mathrm{LMS}}$	0.0666	0.0605	+10.1%	0.0698	0.0605	+15.4%
M^{ERLS1}	0.0419	–	–	0.0673	–	–

TABLE V
RELATIVE EXPERIMENTAL PERFORMANCE OF RLS AND LMS ALGORITHMS

	Case 1: $\mathbf{R}^{-1} = c\mathbf{Q}$			Case 2: $\mathbf{R} = c\mathbf{Q}$		
	Exp. value	Theor. value	$\Delta\%$	Exp. value	Theor. value	$\Delta\%$
$\frac{D_{\min}^{\mathrm{RLS}}}{D_{\min}^{\mathrm{LMS}}}$	0.7578	0.8	−5.3%	1.4806	1.5119	−2.1%
$\frac{M_{\min}^{\mathrm{RLS}}}{M_{\min}^{\mathrm{LMS}}}$	0.6347	0.6614	−4.0%	1.2064	1.25	−3.5%

where $\sqrt{P_s}$ denotes the signal amplitude. Noisy measurements of $s(k)$ are available, say

$$y(k) = s(k) + n(k)$$

where $n(k)$ denotes a white-noise sequence with power P_n. The signal-to-noise ratio (SNR) is denoted by $\rho = P_s/P_n$. A prediction problem is formulated with the objective of estimating $s(k)$ from the noisy data $\{y(k)\}$. More specifically, the "prediction error" $v(k)$ is defined as $v(k) = y(k) - \hat{s}(k)$, where

$$\hat{s}(k) = \boldsymbol{u}^H(k)\boldsymbol{w}_o$$

and

$$\boldsymbol{u}^H(k) = [y(k-1) \quad y(k-2) \quad \cdots \quad y(k-M)].$$

The prediction weight-vector $\boldsymbol{w}_o$ is chosen to minimize $E|v(k)|^2$. Because of the nonstationarity of the chirped signal $s(n)$, the optimal weight vector needs to be time variant and is shown in [17] to be given by

$$\boldsymbol{w}_0(n) = K\boldsymbol{F}(\psi)^n\boldsymbol{D}, \quad K = \frac{\rho}{1+\rho M}$$

where

$$\boldsymbol{D} = [e^{j(\omega-\psi/2)}, \cdots, e^{j(M\omega-\psi M^2/2)}].$$

Consequently, the state-space model of interest in this case takes the form

$$\begin{aligned} \boldsymbol{w}_0(n+1) &= \boldsymbol{F}(\psi)\boldsymbol{w}_0(n), \\ y(n) &= \boldsymbol{u}^H(n)\boldsymbol{w}_0(n) + v(n) \end{aligned} \tag{9}$$

TABLE VI
THEORETICAL RESULTS OF THE RLS AND LMS ALGORITHMS FOR A CHIRPED TONE WHERE $\xi_0 = E|y - u^H w_0|^2$ IS THE MINIMUM PREDICTION ERROR, AND $\eta = 1/M - 3(M+1)/(\rho M^2)$

	M_{min}	μ	μ_{opt}
LMS	$\frac{3M}{4}\xi_0\mu_{opt}$	$\mu P_n(1+\rho)$	$\left(\frac{(1-\eta)(1+\rho)^2\psi^2}{3}\right)^{1/3}$
RLS	$\frac{3M}{4}\xi_0\mu_{opt}$	β	$[(M+1)\rho\psi^2]^{1/3}$

which is in agreement with the model studied in Section IV. The sequence $v(\cdot)$ is taken as a white-noise process with variance $\sigma^2 \approx P_n$ [17]. With this model, the LMS and RLS algorithms are used in an adaptive line enhancer (ALE) configuration, predicting $y(k)$ by using the vector of past inputs $\boldsymbol{u}(k)$.

The relative performance of the RLS and LMS algorithms for tracking a chirped sinusoid in noise is given by Macchi *et al.* [17]–[19]. Table VI summarizes the theoretical misadjustment errors of the LMS and RLS algorithms for a chirped tone and the optimum adaptation constants (μ denotes the LMS step-size, and $\beta = 1 - \lambda$). The ratio of the excess errors is

$$\frac{M_{\min}^{\mathrm{LMS}}}{M_{\min}^{\mathrm{RLS}}} = \left(\frac{(1-\eta)(1+\rho)^2}{3(M+1)\rho}\right)^{1/3} \approx \left(\frac{\rho}{3M}\right)^{1/3}. \tag{10}$$

The factor η is small for large ρ and M. It can be seen that when the input chirped SNR (ρ) is less than $3M$, the performance of the LMS algorithm is superior to that of the RLS algorithm; for $\rho \gg 3M$, the reverse is true.

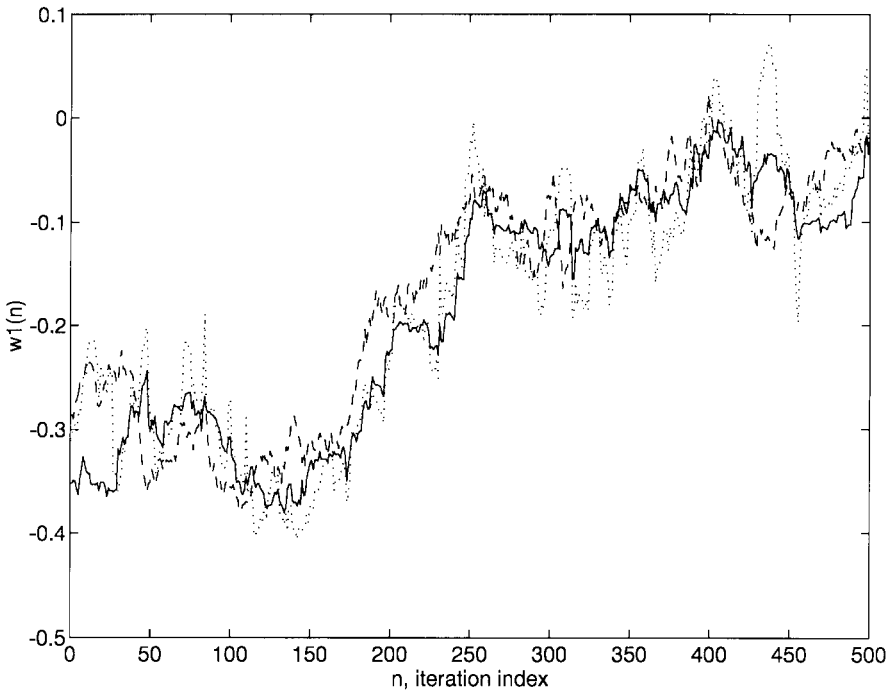

Fig. 1. Tracking ability of RLS and LMS algorithms in case 1) for optimal D ('- - -' is actual, '—' is optimal RLS algorithm, '· · ·' is optimal LMS algorithm).

TABLE VII
MISADJUSTMENT OF LMS, RLS, AND ERLS2 FOR THE CHIRPED SINUSOID PROBLEM

	Case I ($\rho = 2$)				Case II ($\rho = 100$)			
	μ_{min}		M_{min}/P_n		μ_{min}		M_{min/P_n}	
	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.
LMS	0.0037	$2\mu_{opt}$	-20.85dB	-24.37dB	0.0261	$2\mu_{opt}$	-12.31dB	-15.11dB
RLS	0.0039	μ_{opt}	-20.55dB	-22.19dB	0.0144	μ_{opt}	-14.89dB	-16.23dB
ERLS2	-	-	-	-27.64dB	-	-	-	-31.33dB

In the chirped-tone case, normal operation is in the region where the weight difference between the filter weights and the optimal weights is dominated by a nonzero mean-lag weight. In this case, the weight fluctuations caused by the adaptation can be neglected. This explains why a similar result for $D_{\min}$, which measures the total weight noise from the mean lag and the mean lag fluctuation component due to the nonstationarity and the adaptation process respectively, is not available.

To compare the performance of the RLS and LMS algorithms with the ERLS2 algorithm, we simulate two different cases indicated by (10): $\rho \ll 3M$, and $\rho \gg 3M$. The parameters are chosen to complement those in [17] and [19] and are given by the following:

1) $(\rho \ll 3M)$: $\rho = 2, M = 2, \psi = 10^{-4}$
2) $(\rho \gg 3M)$: $\rho = 100, M = 2, \psi = 10^{-4}$.

An approximate value for the optimal adaptation constant is defined in [17] and [19] and summarized in Table VI. This value was used as a first approximation to the optimal value, and a more exact value was determined experimentally by minimizing the output prediction error with respect to the adaptation parameter. It is known that in the chirped case, the adaptive-algorithm-update behaves as a feedback path, which not only updates the filter weights (FIR filter) but also fluctuates the weights to create poles (IIR) in the filter transfer function [20], [21]. Therefore, the filters generally perform better than the linear theory predicts. (Note the discrepancy between the theoretical optimum and the measured value in Fig 6(b) of [18].)

The same chirped signal was used for all three algorithms: LMS, standard RLS, and ERLS2. The estimates of the misadjustment were measured as the mean of $|y(k) - \boldsymbol{u}^H(k)\hat{\boldsymbol{w}}(k|k-1)|^2$ over 1500 iterations in steady state. This was repeated 10 times to find a mean value for the misadjustment. In order to illustrate that the ERLS2 algorithm can estimate an unknown chirp rate, the initial value of Π_0 was set to the identity matrix, and the initial guess for the chirp ψ was set to 0. The fact that Π_0 was incorrect was compensated by using a larger value for σ^2 in the algorithm. We found that the error decreased as σ^2 is increased for this case. The simulations results shown here are for σ^2 that is 200 times larger than the actual.

The results are summarized in Table VII. It can be seen that neither the LMS nor the RLS is superior in all cases; in this chirped-tone example, however, the LMS might be favored since the region of interest is usually low SNR. Note also that the misadjustments for the LMS and RLS actually increase as the SNR (ρ) is increased since neither algorithm is estimating the chirp rate. The ERLS2 algorithm uses the additional SNR

TABLE VIII
CONSTRAINT VERIFICATION TABLE FOR BASIC EXPERIMENTAL PARAMETERS OF TABLE III (NOTE NUMBERS REFER TO CORRESPONDING RELATIONS)

	Case 1: $\mathbf{R}^{-1} = c\mathbf{Q}$,		Case 2: $\mathbf{R} = c\mathbf{Q}$,		
	Exp. value	Condition	Exp. value	Condition	Note
$\lambda_{D\min}$	0.98	≈ 1	0.98	≈ 1	(A.1)
$\lambda_{M\min}$	0.98	≈ 1	0.9622	≈ 1	(A.1)
$\mu_{D\min}$	0.1562	$\ll 25.0$	0.0827	$\ll 15.9$	(A.2)
$\mu_{M\min}$	0.0827	$\ll 1.367$	0.0827	$\ll 1.367$	(A.2)
α	0.0283	$\ll 1$	0.0535	$\ll 1$	(A.3)

to improve its estimate of the chirp rate, thereby decreasing the misadjustment. It was observed that on a run-by-run basis, the ERLS2 always performed better.

The discrepancy between the theoretical LMS and RLS results are due to non-Wiener effects explained earlier. The weight update behaves as a feedback path, which produces poles in the transfer function. As in [18], we observed that both the LMS and RLS performed better than the theory predicted at low chirp rates. The difference, however, is greater for the LMS.

VII. CONCLUSIONS

The Kalman filter is known to be the linear optimum tracker on the basis of second-order statistics. Building on this fact and exploiting the one-to-one correspondences between the RLS and Kalman variables, we may derive extended forms of the RLS algorithm that inherit the good tracking behavior of the Kalman filter. In this paper, we have considered two particular forms of this extension:

- ERLS-1, pertaining to a system identification problem, and
- ERLS-2, pertaining to the tracking of a chirped sinusoid in noise.

In each case, prior knowledge about the original dynamical system model is built into the formulation of the extended form of the RLS algorithm, making it the optimum linear tracking device for the particular application of interest. We should add, though, that if sufficient prior knowledge is not available, say, if some of the parameters needed in the formulations of ERLS-1 and ERLS-2 are not available, then the extended algorithms would have to be expanded to include provisions for estimating these quantities.

Finally, we can go one step further and say that, by exploiting the time-shift structure of the input data in the manner described in [2] and [8], we may develop order-recursive realizations of extended RLS algorithms that are linear in their order of computational complexity; this is an issue that we will leave for future considerations.

APPENDIX
BACKGROUND THEORY FOR TRACKING EXPERIMENT

For cases 1) and 2) described in Section VI-A, the free parameters σ, σ_Q, c, q_1, and q_2 must be chosen such that we have the following:

1) The optimal RLS parameters $\lambda_{D\min}, \lambda_{M\min}$ and the optimal LMS parameters $\mu_{D\min}, \mu_{M\min}$, achieving the minima specified in Table II satisfy

$$\lambda_{D,M\min} = 1 - \epsilon, \qquad 1 \gg \epsilon \gg 1 - a, \tag{A.1}$$

$$\mu_{D\min} \ll (\|\boldsymbol{R}\boldsymbol{K}(n)\boldsymbol{R}\|)^{-1/2}, \quad \mu_{M\min} \ll \mathrm{tr}[\boldsymbol{R}]^{-1} \tag{A.2}$$

where

$$\boldsymbol{K}(n) \triangleq E[(\boldsymbol{w}_o(n) - \hat{\boldsymbol{w}}(n|n-1))(\boldsymbol{w}_o(n) - \hat{\boldsymbol{w}}(n|n-1))^H].$$

2) The Markov model in (7) satisfies the condition of *slow statistical variation*, i.e.,

$$\alpha = \frac{1}{\sigma}\sqrt{\mathrm{tr}[\boldsymbol{R}\boldsymbol{Q}]} \ll 1. \tag{A.3}$$

The rationale for these conditions can be found in [2] and [6]. In addition, we like to select a set of conditions that not only verifies the theoretical predictions of Tables I and II but also does so under meaningful tracking conditions, i.e.,

$$\frac{D_{\min}^{\mathrm{RLS,LMS}}}{E[\|\boldsymbol{w}_o(n)\|^2]} = \frac{D_{\min}^{\mathrm{RLS,LMS}}}{\mathrm{tr}[\boldsymbol{Q}]}(1 - a^2) = \delta \ll 1. \tag{A.4}$$

Under these constraints, we simplify the problem of assigning good values to the free parameters of the experiment by considering the case where σ_Q, q_1, q_2, and σ have been set *a priori*. We call this set of parameters the *primary parameters* of the experiment, as opposed to the others, which we call the *secondary parameters*. Assume that $D_{\min}^{\mathrm{RLS}} = d$ is then selected to satisfy condition (A.4). Since [2]

$$\lambda_{D\min} = 1 - \frac{1}{\sigma}\sqrt{\frac{\mathrm{tr}[\boldsymbol{Q}]}{\mathrm{tr}[\boldsymbol{R}^{-1}]}} \tag{A.5}$$

conditions (A.3) and (A.1) imply that

$$\sigma^2 = \frac{1}{\epsilon^2}\frac{\mathrm{tr}[\boldsymbol{Q}]}{\mathrm{tr}[\boldsymbol{R}^{-1}]} = K\mathrm{tr}[\boldsymbol{R}\boldsymbol{Q}], \qquad K \gg 1 \tag{A.6}$$

where K can be considered an "overdrive" factor that measures the degree to which the condition of slow statistical variation holds. Specializing to case (1)

$$K = \frac{1}{M\epsilon^2} \gg 1 \quad \text{for} \quad \epsilon \ll \frac{1}{\sqrt{M}} = \frac{1}{\sqrt{2}}.$$

On the other hand, from the relation for $D_{\min}^{\mathrm{RLS}}$ in Table I, we must choose c and ϵ such that

$$c = \left(\frac{d}{\sigma\sigma_Q^2(1+q_2)}\right)^2, \quad \epsilon = \frac{\sigma_Q^2(1+q_2)}{d}.$$

Similarly, for case (2)

$$K = \frac{q_2 - q_1^2}{\epsilon^2(1 + 2q_1^2 + q_2^2)} \gg 1 \quad \text{for} \quad \epsilon \ll \sqrt{\frac{q_2 - q_1^2}{1 + 2q_1^2 + q_2^2}}$$

and

$$c = \frac{1}{(q_2 - q_1^2)}\left(\sigma\frac{(1+q_2)}{d}\right)^2.$$

It is important to note that had c been assumed to be unity for both cases 1) and 2), not all the required conditions for the validity of even just $D_{\min}^{\mathrm{RLS}}$ can be met for any choices of the primary and secondary parameters; in this sense, the generalization provided by the factor c is crucial. As long as ϵ [hence, a by condition (A.1)] and c are selected according to these relations for each case, conditions (A.1), (A.3), and (A.4) will be satisfied for $\lambda = \lambda_{D\min}$. In other words, the optimum mean-square deviations of the RLS algorithm for cases 1) and 2) with the primary parameters are realizable, albeit with different secondary parameters. The obvious question follows: Is this true with the other optima in Table I? Without going into the details, it is straightforward to verify that the answer is indeed yes for the numerical choices of the primary and secondary parameters given in Table III. Table VIII summarizes the results of the verification. In all cases, the formulae for the optimal parameters $\lambda_{D\min}, \lambda_{M\min}, \mu_{D\min}$, and $\mu_{M\min}$ can be found in [2] and [6]. Where possible, the constraint values are determined from the theoretical formulae (A.1) to (A.3); the exception is the constraint value for $\mu_{D\min}$, which can only be easily computed after the actual experiment is complete since we have no prior knowledge of the predicted weight sequence $\hat{\boldsymbol{w}}(n|n-1)$.

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996, 3rd ed.

[3] B. Widrow *et al.*, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, 1976.

[4] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1097–1110, 1986.

[5] A. Benveniste, "Design of adaptive algorithms for the tracking of time-varying systems," *Int. J. Adaptive Contr. Signal Processing*, vol. 1, pp. 3–29, 1987.

[6] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, 1994.

[7] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 283–287, 1986.

[8] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, pp. 18–60, July 1994.

[9] D. N. Godard, "Channel equalization using a Kalman filter for fast data transmission," *IBM J. Res. Dev.*, vol. 18, pp. 267–273, 1974.

[10] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. englewood Cliffs, NJ: Prentice-Hall, 1979.

[11] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[12] Q. T. Zhang and S. Haykin, "Tracking characteristics of the Kalman filter in a nonstationary environment for adaptive filter applications," *Proc. ICASSP*, Boston, MA, 1983, pp. 671–674.

[13] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA, MIT Press, 1983.

[14] J. J. Deyst and C. F. Price, "Conditions for the asymptotic stability of the discrete, minimum variance, linear estimator," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 702–705, 1968.

[15] T. P. McGarty, *Stochastic Systems and State Estimation*. New York: Wiley, 1974.

[16] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. New York; Academic, 1979, vol. 1.

[17] N. J. Bershad and O. M. Macchi, "Adaptive recovery of a chirped sinusoid in noise: Performance of the LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 595–602, Mar. 1991.

[18] O. Macchi, N. Bershad, and M. Mboup, "Steady-state superiority of LMS over LS for time-varying line enhancer in noisy environment," *Proc. Inst. Elec. Eng.*, pt. F (Radar and Signal Processing), vol. 138, pp. 354–360, Aug. 1991.

[19] O. M. Macchi and N. J. Bershad, "Adaptive recovery of a chirped sinusoid in noise: Performance of the RLS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 583–594, Mar. 1991.

[20] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 484–491, Feb. 1977.

[21] R. C. North, J. R. Zeidler, T. R. Albert, and W. H. Ku, "Comparison of adaptive lattice filters to LMS transversal filters for sinusoidal cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, Mar. 1992, vol. IV, pp. 33–36.

Simon Haykin (F'82) received the B.Sc. degree (First-Class Honors) in 1953, the Ph.D. degree in 1956, and the D.Sc. degree in 1967, all in electrical engineering from the University of Birmingham, U.K.

He is the founding Director of the Communications Research Laboratory, McMaster University, Hamilton, Ont., Canada. His research interests include nonlinear dynamics, neural networks, adaptive filters, and their applications in radar and communication systems. He is the Editor of *Adaptive and Learning Systems for Signal Processing, Communications and Control*, a new series of books for Wiley-Interscience.

In 1980, Dr. Haykin was elected Fellow of the Royal Society of Canada, he was awarded the McNaughton Gold Medal, IEEE (Region 7), in 1986, and he is a recipient of the Canadian Telecommunications Award from Queen's University. In 1996 he was awarded the title "University Professor."

Ali H. Sayed (M'92) was born in Sao Paulo, Brazil. In 1981, he graduated in first place in the National Lebanese Baccalaureat, with the highest score in the history of the examination. In 1987, he received the degree of Engineer in Electrical Engineering from the University of Sao Paulo, Brazil, and was the first-place graduate of the School of Engineering. In 1989, he received, with distinction, the M.S. degree in electrical engineering from the University of Sao Paulo, Brazil, and in 1992, he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

From September 1992 to August 1993, he was a Research Associate with the Information Systems Laboratory, Stanford University, after which he joined, as an Assistant Professor, the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. Since July 1996, he has been an Associate Professor in the Department of Electrical Engineering, University of California, Los Angeles. His research interests are in the areas of adaptive and statistical signal processing, linear and nonlinear filtering and estimation, interplays between signal processing and control methodologies, interpolation theory, and reliable and efficient algorithms for structured matrix problems.

Dr. Sayed was awarded the Institute of Engineering Prize and the Conde Armando Alvares Penteado Prize, both in 1987. In 1993, his Ph.D. thesis on structured algorithms in signal processing and mathematics received a special mention among the three best Ph.D. thesis written during the period from 1990 to 1993, for the Householder Prize in numerical linear algebra. He has written in over 70 publications and received a 1994 NSF Research Initiation Award and a 1996 IEEE Donald G. Fink Prize. He is a member of SIAM and ILAS.

James R. Zeidler (M'76–SM'84–F'94) received the B.A. degree in physics from MacMurray College, Jacksonville, IL, in 1966, the M.S. degree in physics from Michigan State University, East Lansing, in 1968, and the Ph.D. degree in physics from the University of Nebraska, Lincoln, in 1972.

In 1973, he was a postdoctoral research associate with the Department of Electrical Engineering, University of Nebraska and a lecturer in physics. Since 1974, he has been a scientist at the Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, San Diego, CA, where he has been responsible for fundamental research and new technology development for undersea surveillance, sonar and communications signal processing, communications signals exploitation, array processing, underwater acoustic communications, infrared image processing, electronic devices, and signal processor development. From 1983 to 1984, he was a technical advisor in the Office of the Assistant Secretary of the Navy (Research, Engineering, and Systems), Washington, DC. Since 1988, he has also been affiliated with the University of California, San Diego, where he is an Adjunct Professor of Electrical and Computer Engineering. His current interests are in wireless communications.

Dr. Zeidler was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING He received an award for the best unclassified paper at the IEEE Military Communications Conference in 1995.

Paul Yee (S'96) received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia, Vancouver, B.C., Canada, in 1989.

From 1989 to 1991, he worked as a Design Engineer at Newbridge Networks Corporation, Kanata, Ont., Canada. Since then, he has been with the Communications Research Laboratory, McMaster University, Hamilton, Ont., where he is currently a Ph.D. candidate. His research interests center around the theory and application of radial basis functions to probability estimation, classification, and time-series prediction.

Paul C. Wei (M'95) received the B.S. degree in electrical engineering from University of California, Los Angeles, in 1989. He received the M.S. and Ph.D. degrees in electrical engineering from University of California, San Diego, in 1991 and 1995, respectively.

He is currently working at the LSI Logic Wireless Design Center, Carlsbad, CA. From 1991 to 1995, he was a research assistant at the Industry/University Cooperative Research Center on Ultra High Speed Integrated Circuits and Systems, University of California, San Diego. He also worked part time at Hughes Network Systems, San Diego, during the last year of his graduate work. His research interests include signal and image processing algorithms, applications of signal processing techniques to broadband communications, and performance of adaptive algorithms in high-speed hardware implementations.