On Exponentially Weighted Recursive Least Squares for Estimating Time-Varying Parameters

Ta-Hsin Li

Department of Mathematical Sciences

IBM T. J. Watson Research Center

Yorktown Heights, NY 10598-0218

thl@watson.ibm.com

December 9, 2003

Abstract

The exponentially weighted recursive least-squares (RLS) has a long history as an algorithm to track time-varying parameters in signal processing and time series analysis. By reviewing the optimality conditions of RLS under a regression framework, possible sources of suboptimality of RLS for tracking time-varying parameters, especially when the parameters satisfy a state-space model, are identified. A straightforward relationship between the RLS variables and the Kalman filtering variables is established under the state-space model assumption. This relationship enables a unified development of several simple algorithms that generalize and extend the traditional RLS. Numerical examples are given to demonstrate the improved tracking performance of theses algorithms.

I. INTRODUCTION

The method of exponentially weighted recursive least-squares, or simply RLS, has long been employed as a simple alternative to Kalman filtering (KF) for tracking time-varying parameters (e.g., [1]–[6]). KF as an estimator is well known to be optimal under the state-space model (SSM) assumption ([7], [8]). One of the shortcomings of KF is the requirement of complete prior knowledge of the SSM and its parameters. Although under certain conditions the SSM parameters can be estimated from a set of observations via, for example, the Gaussian maximum likelihood approach, the problem of identifiability and numerical computation is not very easy to solve in general ([9, Sec. 3.4], [10, Chap. 13]). It is also not always easy to validate the SSM assumption based solely on the observations. RLS on the other hand does not require a dynamic model for the time-varying parameters except for a forgetting factor. This, coupled with lower computational costs, makes RLS an attractive alternative to KF.

The trade-off is, of course, in the performance of estimation. Being optimal, KF produces the most statistically efficient estimates of time-varying parameters under the SSM assumption. RLS on the other hand only produces suboptimal estimates in general, due to its simplified assumptions. It is therefore not surprising that KF often outperforms RLS in situations where a suitable state-space model is available for the time-varying parameters. In some standard literature of signal processing (e.g., [2]–[6], [11]), most of the performance analyses of RLS are focused either on the convergence and the steady-state behavior for constant rather than time-varying parameters, or on the computation of the error covariance matrix and its upper bounds in tracking time-varying parameters under the SSM assumption.

In this article, it is intended to highlight the compromise of RLS as an estimator of time-varying parameters under the SSM assumption when compared to the optimal KF solution. The relationship between RLS and KF was a focus in the earlier literature. For example, RLS was compared with KF in [1, pp. 93–95] for the special case of ordinary (nonweighted) least squares; a KF-like recursive algorithm was derived in [7, pp. 205–207 and 308–309] for the problem of exponentially weighted constrained least squares. The most recent work on this subject seems to be [12] and [13] (see also [11, Sec. 13.8]), where the variables of RLS algorithms are explicitly mapped to the variables of KF algorithms under a special unforced SSM formulation in which the state variable is exponentially time-varying. Although it may be necessary in unifying some KF algorithms that require a constant noise variance, this formulation has the possibility of being misunderstood as implying that RLS is optimal in estimating *time-varying state variables*, especially when reference to other possible formulations is lacking.

By revisiting the earlier literature, it is pointed out in this article that a more straightforward formulation for establishing the RLS-KF correspondence begins with a reference signal that comprises a linear function of the explanatory signal with *constant coefficient* plus a measurement error process that has *exponentially time-varying variance*. It is under this condition that RLS achieves its optimality according to the statistical estimation theory as well as its most natural correspondence with KF. This correspondence highlights the fact that the tracking capability of RLS for (slowly) time-varying parameters, as experienced in practice, can be properly explained only if the parameter-induced variability of the observed reference signal is treated as the time-varying 'measurement error' which, in general, is not white noise with exponential variance but becomes such exactly or approximately under suitable conditions. The correspondence also makes it easier to understand the necessity of rescaling the variables when employing some KF algorithms for time-invariant systems (such as the square-root algorithms [8, pp. 147–152]) to compute the RLS estimator.

Based on the RLS-KF correspondence, it is straightforward to obtain simple generalizations of RLS that preserve the exponential weighting but allow additional SSM structures to be incorporated possibly as extra tuning parameters. One such generalization is the so-called extended forgetting factor RLS (EFRLS) algorithm recently proposed in [14]. Another generalization equips RLS with an additional parameter to accommodate rapid random fluctuations of the time-varying parameters. Improved tracking performance of these algorithms are confirmed in two numerical examples that arise in mobile communications.

II. Review of Recursive Least Squares

Let $\{y_t\}$ be the reference signal and $\{\boldsymbol{x}_t\}$ be the explanatory (vector) signal. Given the observations $\{\boldsymbol{x}_t, y_t\}_{t=0}^{n}$ that are available at time n, it is well known (e.g., [7, pp. 151-152]) that the exponentially weighted RLS algorithm, with initial values $\boldsymbol{\mu}$ and $\mathbf{Q}_0$, produces the best linear functional, $\boldsymbol{\theta}_n$: $\boldsymbol{x}_t \mapsto y_t$ $(t = 0, 1, \ldots, n)$, that minimizes the exponentially weighted least-squares cost function

$$J_n(\boldsymbol{\theta}) := \sum_{t=0}^{n} \lambda^{n-t} |y_t - \boldsymbol{\theta}^* \boldsymbol{x}_t|^2 + \lambda^{n+1} (\boldsymbol{\theta} - \boldsymbol{\mu})^* \mathbf{Q}_0^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}), \tag{1}$$

where $\lambda \in (0,1]$ is the forgetting factor and the asterisk stands for Hermitian transpose. The optimal coefficient $\boldsymbol{\theta}_n := \arg\min_{\boldsymbol{\theta}} J_n(\boldsymbol{\theta})$ is a linear function of y_t and solves the normal equation

$$\boldsymbol{\Phi}_n \boldsymbol{\theta}_n = \boldsymbol{\psi}_n \quad \text{or} \quad \boldsymbol{\theta}_n = \boldsymbol{\Phi}_n^{-1} \boldsymbol{\psi}_n, \tag{2}$$

where $\boldsymbol{\Phi}_n := \sum_{t=0}^{n} \lambda^{n-t} \boldsymbol{x}_t \boldsymbol{x}_t^* + \lambda^{n+1} \mathbf{Q}_0^{-1}$ and $\boldsymbol{\psi}_n := \sum_{t=0}^{n} \lambda^{n-t} \boldsymbol{x}_t y_t^* + \lambda^{n+1} \mathbf{Q}_0^{-1} \boldsymbol{\mu}$.

Minimization of $J_n(\boldsymbol{\theta})$ is intuitively appealing from the viewpoint of predicting y_t from $\boldsymbol{x}_t$ because it produces the 'best' predictor among all linear functions of $\boldsymbol{x}_t$. The discounting mechanism, achieved via λ, is often justified in practice as a way of giving greater emphasis to the newer data and thereby providing the potential of tracking time-varying systems.

In order explain the optimality or suboptimality of $\boldsymbol{\theta}_n$ for estimating time-varying parameters, especially under the SSM assumption, let us first consider the following regression problem.

For fixed $n > 0$, let y_t and $\boldsymbol{x}_t$ satisfy

$$y_t = \boldsymbol{\beta}_n^* \boldsymbol{x}_t + e_t \quad (t = 0, 1, \ldots, n), \tag{3}$$

where $\boldsymbol{\beta}_n$ is an unknown vector and $\{e_t\}_{t=0}^{n}$ is a random process with zero mean, possibly time-varying variance σ_t^2, and, in general, nondiagonal covariance matrix $\mathbf{V}$. In the following, $\{e_t\}$ will be referred to as the *measurement error* in the regression model. Moreover, all probability distributions and expectations should be understood as being conditioned on $\{\boldsymbol{x}_t\}_{t=0}^{n}$ for which $\boldsymbol{\Phi}_n$ is nonsingular.

Suppose that $\{e_t\}$ is a serially uncorrelated process and $\sigma_t^2 = c\lambda^t$ for some $c := \sigma_0^2 > 0$, i.e.,

$$e_t \sim \mathrm{WN}(0, c\lambda^t) \quad (t = 0, 1, \ldots, n). \tag{4}$$

Under this assumption, $\boldsymbol{\theta}_n$ is known to be the best (i.e., minimum variance) linear unbiased estimator (BLUE) of $\boldsymbol{\beta}_n$ (e.g., [15, p. 301]). If, in addition, $\{e_t\}$ is Gaussian, and $\boldsymbol{\beta}_n$ is also Gaussian with the prior distribution $\boldsymbol{\beta}_n \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{P}_0)$, where $\mathbf{P}_0 := c\lambda^{-1}\mathbf{Q}_0$, then the posterior distribution of $\boldsymbol{\beta}_n$ given $\mathcal{Y}_n := \{y_t\}_{t=0}^{n}$ is $\mathrm{N}(\boldsymbol{\Phi}_n^{-1}\boldsymbol{\psi}_n, \sigma_n^2 \boldsymbol{\Phi}_n^{-1})$, which gives $\boldsymbol{\theta}_n = E(\boldsymbol{\beta}_n | \mathcal{Y}_n)$, so $\boldsymbol{\theta}_n$ is the minimum mean-square error (MMSE) estimator and the maximum *a posteriori* probability (MAP) estimator of $\boldsymbol{\beta}_n$. If no prior distribution is imposed on $\boldsymbol{\beta}_n$, which is equivalent to assuming $\mathbf{Q}_0^{-1} = \mathbf{0}$, then $\boldsymbol{\theta}_n$ becomes the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}_n$ which has minimum variance among all unbiased estimators of $\boldsymbol{\beta}_n$ (e.g., [1, pp. 33–34], [15, p. 319]), where the unbiasedness of $\boldsymbol{\theta}_n$ is due to the fact that $E(e_t) = 0$ and the minimum variance property of $\boldsymbol{\theta}_n$ depends crucially on two assumptions: the e_t must be uncorrelated (i.e., $\mathbf{V}$ must be diagonal) and σ_t^2 must be an exponential function of t determined by λ via (4). Generally, if σ_t^2 is nonexponential and/or if $\mathbf{V}$ is nondiagonal, $\boldsymbol{\theta}_n$ remains unbiased for estimating $\boldsymbol{\beta}_n$ in the case of $\mathbf{Q}_0^{-1} = \mathbf{0}$ because $E(e_t) = 0$, but the statistical efficiency of $\boldsymbol{\theta}_n$ is not as good as MLE, or BLUE, that minimizes the generalized least-squares cost function $\sum_{s=0}^{n} \sum_{t=0}^{n} \lambda_{st} (y_s - \boldsymbol{\theta}^* \boldsymbol{x}_s)^* (y_t - \boldsymbol{\theta}^* \boldsymbol{x}_t)$, where λ_{st} is the (s,t) entry of $\mathbf{V}^{-1}$.

The most important feature in (3)–(4) is that the variance of e_t grows exponentially as t goes always from n into the past whereas the coefficient $\boldsymbol{\beta}_n$ remains constant for all $t = 0, 1, \ldots, n$. These are the minimum requirements for RLS to be optimal. However, to achieve the optimality, the forgetting factor must match the growth rate of the measurement error variance, as specified by (4).

III. SSM for Time-Varying Parameters

Now assume that the linear relationship between y_t and x_t is time-varying, i.e., $y_t = \boldsymbol{\beta}_t^* \boldsymbol{x}_t + z_t$ $(t = 0, 1, \ldots, n)$, where $\{z_t\}$ is a zero-mean random process which is independent of $\{\boldsymbol{\beta}_t\}$ and $\{\boldsymbol{x}_t\}$. The question is: under what conditions is the RLS solution $\boldsymbol{\theta}_n$ optimal as an estimator of $\boldsymbol{\beta}_n$? This question is investigated in the following under the SSM assumption.

Thanks to KF for providing an efficient way of estimating the hidden state variable, SSM is widely used to model time-varying systems, not only in physical sciences and engineering (e.g., [8]) but also in observational sciences such as econometrics (e.g., [9]). A popular SSM is of the form

$$\begin{aligned} \boldsymbol{\beta}_{t+1} &= \mathbf{F}_t \boldsymbol{\beta}_t + \boldsymbol{v}_t, \\ y_t &= \boldsymbol{\beta}_t^* \boldsymbol{x}_t + z_t, \end{aligned} \tag{5}$$

where the transition matrices $\mathbf{F}_t$ are nonsingular, the system noise $\boldsymbol{v}_t$ and the observation noise[1] z_t are mutually uncorrelated and satisfy

$$\boldsymbol{v}_t \sim \mathrm{WN}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad z_t \sim \mathrm{WN}(0, \varsigma_t^2). \tag{6}$$

Note that unlike the regression model (3), the relationship between y_t and $\boldsymbol{x}_t$ in (5) is characterized by a time-varying, rather than constant, coefficient $\boldsymbol{\beta}_t$.

[1] The process $\{z_t\}$ is referred to as observation noise instead of measurement error because the latter is reserved for the regression model (3). The distinction will become apparent in Sec. IV-C.

If the model parameters in (5)–(6), which include $\mathbf{F}_t$, $\boldsymbol{\Sigma}_t$, and ς_t^2, are completely known *a priori* as functions of t, and if the computational power is sufficient, then the KF algorithm, given in Table I, should be used to estimate the $\boldsymbol{\beta}_t$ (as did in [16] for the special case where $\mathbf{F} \equiv \mathbf{I}$, $\boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}_0$, and $\varsigma_t^2 \equiv \varsigma_0^2$), because KF provides the optimal solution under the SSM assumption [8]. However, if the model parameters are not completely known and are too difficult to estimate, which is true in many applications, then RLS provides a simple alternative for tracking $\boldsymbol{\beta}_t$. In order to understand the performance of RLS in such cases, it is necessary to reconcile the difference between (3)–(4) and (5)–(6) and to identify the compromise of RLS in achieving its computational simplicity.

A. RLS and KF Correspondence

The estimation problems of RLS and KF have the following correspondence.

Proposition 1 *Assume that* y_t, $\boldsymbol{x}_t$ *and* $\boldsymbol{\beta}_t$ *satisfy the SSM* (5)–(6), *where* $\mathbf{F}_t \equiv \mathbf{I}$, $\boldsymbol{\Sigma}_t \equiv \mathbf{0}$, *and* $\varsigma_t^2 = c\lambda^t$ *for some* $c := \varsigma_0^2 > 0$. *Then, the RLS solution* $\boldsymbol{\theta}_n$ *defined by* (2) *coincides with the KF estimator of* $\boldsymbol{\beta}_n$, *i.e.,* $\boldsymbol{\theta}_n = \boldsymbol{\beta}_{n|n} = \boldsymbol{\beta}_{n+1|n}$, *and the correspondence between RLS and KF variables are given in Table I.*

Proof. Recall that the KF equations for the 'filtering' problem under (5)–(6) are given in the first column of Table I (e.g., [8, p. 44], [17, p. 478]). The key to establishing the RLS-KF correspondence is to introduce the new variables

$$\mathbf{Q}_{t|t} := \mathbf{P}_{t|t}/\varsigma_t^2, \quad \mathbf{Q}_{t+1|t} := \mathbf{P}_{t+1|t}/\varsigma_t^2. \tag{7}$$

If $\varsigma_t^2 = c\lambda^t$, then $\varsigma_t^2/\varsigma_{t-1}^2 = \lambda$. Under the additional assumption that $\mathbf{F}_t \equiv \mathbf{I}$ and $\boldsymbol{\Sigma}_t \equiv \mathbf{0}$, the KF equations in the second column of Table I become those in the second column, which are readily recognized as forming the RLS algorithm (e.g., [11, p. 569]). Note that the initial values of KF at $t = 0$ are $\boldsymbol{\beta}_{0|-1} = \boldsymbol{\mu}$ and $\mathbf{P}_{0|-1} = \mathbf{P}_0$, where $\boldsymbol{\mu} := E(\boldsymbol{\beta}_0)$ and $\mathbf{P}_0 := \mathrm{Cov}(\boldsymbol{\beta}_0)$. Since (2) gives $\boldsymbol{\theta}_0 = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\psi}_0 = \boldsymbol{\mu} + \mathbf{Q}_0\boldsymbol{x}_0\,(\boldsymbol{x}_0^*\mathbf{Q}_0\boldsymbol{x}_0 + \lambda)^{-1}(y_0 - \boldsymbol{\mu}^*\boldsymbol{x}_0)^*$, the initial values of RLS must be $\boldsymbol{\beta}_{0|-1} = \boldsymbol{\mu}$ and $\mathbf{Q}_{0|-1} = \mathbf{Q}_0$. On the other hand, the new notation in (7) requires $\mathbf{Q}_{0|-1} = \mathbf{P}_{0|-1}/\varsigma_{-1}^2 = c^{-1}\lambda\mathbf{P}_0$. Combining these expressions gives $\mathbf{Q}_0 = c^{-1}\lambda\mathbf{P}_0$. Under this condition, $\boldsymbol{\theta}_n = \boldsymbol{\beta}_{n|n} = \boldsymbol{\beta}_{n+1|n}$ for all n. Q.E.D.

Remark 1. Unlike the SSM in [11]–[13], the SSM (5)–(6) under the conditions of Proposition 1 highlights the underlying estimation problem of RLS, namely, the estimation of a *time-invariant coefficient* $\boldsymbol{\beta}_t$

TABLE I

SUMMARY OF ALGORITHMS

KF	RLS	EFRLS
Assumptions: $\mathbf{F}_t$, $\boldsymbol{\Sigma}_t$, ς_t^2 known	$\mathbf{F}_t \equiv \mathbf{I}$, $\boldsymbol{\Sigma}_t \equiv \mathbf{0}$, $\varsigma_t^2 = c\lambda^t$	$\mathbf{F}_t$ known, $\boldsymbol{\Sigma}_t \equiv \mathbf{0}$, $\varsigma_t^2 = c\lambda^t$
$\boldsymbol{g}_t = \mathbf{P}_{t\|t-1}\boldsymbol{x}_t(\boldsymbol{x}_t^*\mathbf{P}_{t\|t-1}\boldsymbol{x}_t + \varsigma_t^2)^{-1}$	$\boldsymbol{g}_t = \mathbf{Q}_{t\|t-1}\boldsymbol{x}_t(\boldsymbol{x}_t^*\mathbf{Q}_{t\|t-1}\boldsymbol{x}_t + \lambda)^{-1}$	$\boldsymbol{g}_t = \mathbf{Q}_{t\|t-1}\boldsymbol{x}_t(\boldsymbol{x}_t^*\mathbf{Q}_{t\|t-1}\boldsymbol{x}_t + \lambda)^{-1}$
$\mathbf{P}_{t\|t} = \mathbf{P}_{t\|t-1} - \boldsymbol{g}_t\boldsymbol{x}_t^*\mathbf{P}_{t\|t-1}$	$\mathbf{Q}_{t\|t} = \lambda^{-1}\mathbf{Q}_{t\|t-1} - \lambda^{-1}\boldsymbol{g}_t\boldsymbol{x}_t^*\mathbf{Q}_{t\|t-1}$	$\mathbf{Q}_{t\|t} = \lambda^{-1}\mathbf{Q}_{t\|t-1} - \lambda^{-1}\boldsymbol{g}_t\boldsymbol{x}_t^*\mathbf{Q}_{t\|t-1}$
$\mathbf{P}_{t+1\|t} = \mathbf{F}_t\mathbf{P}_{t\|t}\mathbf{F}_t^* + \boldsymbol{\Sigma}_t$	$\mathbf{Q}_{t+1\|t} = \mathbf{Q}_{t\|t}$	$\mathbf{Q}_{t+1\|t} = \mathbf{F}_t\mathbf{Q}_{t\|t}\mathbf{F}_t^*$
$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\boldsymbol{x}_t)^*$	$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\boldsymbol{x}_t)^*$	$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\boldsymbol{x}_t)^*$
$\boldsymbol{\beta}_{t+1\|t} = \mathbf{F}_t\boldsymbol{\beta}_{t\|t}$	$\boldsymbol{\beta}_{t+1\|t} = \boldsymbol{\beta}_{t\|t}$	$\boldsymbol{\beta}_{t+1\|t} = \mathbf{F}_t\boldsymbol{\beta}_{t\|t}$

Initial values are $\boldsymbol{\beta}_{0|-1} = \boldsymbol{\mu} := E(\boldsymbol{\beta}_0)$, $\mathbf{P}_{0|-1} = \mathbf{P}_0 := \mathrm{Cov}(\boldsymbol{\beta}_0)$, and $\mathbf{Q}_{0|-1} = \mathbf{Q}_0 := c^{-1}\lambda\mathbf{P}_0$.

based on observations in which the noise has an *exponentially time-varying variance*. This is consistent with the regression model (3)–(4) under which RLS is proven optimal.

Remark 2. The RLS-KF correspondence given in Proposition 1 does not require any transformation of the original variables y_t or $\boldsymbol{\beta}_t$ — the SSM (5)–(6) is directly for the original variables. Moreover, no KF variables, except $\mathbf{P}_{t|t}$ and $\mathbf{P}_{t+1|t}$, need to be transformed to become RLS variables. The new RLS variables $\mathbf{Q}_{t|t}$ and $\mathbf{Q}_{t+1|t}$, defined in (7), can be interpreted, from the KF viewpoint, as the *standardized* prediction and estimation error covariance matrices, respectively — standardized by the variance of the observation noise. This transformation was also suggested in [8, p. 136].

It is important to note that the initial value $\mathbf{Q}_0$ in RLS is different from the initial value $\mathbf{P}_0$ in KF. Therefore, one cannot interpret $\mathbf{Q}_0$ as the covariance matrix of $\boldsymbol{\beta}_0$. In fact, Table I shows that $\mathbf{Q}_0$ is a scaled version of $\mathbf{P}_0$ — scaled by the product of the forgetting factor and the reciprocal of the observation noise variance at $t = 0$. In practice, the latter can be treated as a tuning parameter which, together with the forgetting factor, controls the performance of RLS.

For computational reasons, one may want to obtain the RLS solution by employing some alternative KF algorithms that are applicable only to time-invariant systems. In this case, one just need to realize that the only time-varying parameter in (5)–(6) under the assumptions in Proposition 1 is the variance of the observation noise z_t, which is proportional to λ^t. Given λ, a simple standardization transform, with $z_t \mapsto \tilde{z}_t := z_t\lambda^{-t/2}$, $y_t \mapsto \tilde{y}_y := y_t\lambda^{-t/2}$, and $\boldsymbol{\beta}_t \mapsto \tilde{\boldsymbol{\beta}}_t := \boldsymbol{\beta}_t\lambda^{-t/2}$, will convert the time-varying system into a time-invariant one: $\tilde{\boldsymbol{\beta}}_{t+1} = \lambda^{-1/2}\tilde{\boldsymbol{\beta}}_t$, $\tilde{y}_t = \tilde{\boldsymbol{\beta}}_t^* x_t + \tilde{z}_t$, which is just the formulation in [11]–[13]. This formulation should only be regarded as a computational device, rather than the underlying estimation problem.

TABLE II

SUMMARY OF ALGORITHMS – II

EFRLS-2	RLS-2	RLS-3
$\mathbf{F}_t$ known, $\mathbf{\Sigma}_t = \rho\,\varsigma_t^2\mathbf{I}$, $\varsigma_t^2 = c\lambda^t$	$\mathbf{F}_t \equiv \mathbf{I}$, $\mathbf{\Sigma}_t = \rho\,\varsigma_t^2\mathbf{I}$, $\varsigma_t^2 = c\lambda^t$	$\mathbf{F}_t = \alpha\mathbf{I}$, $\mathbf{\Sigma}_t = \rho\,\varsigma_t^2\mathbf{I}$, $\varsigma_t^2 = c\lambda^t$
$\boldsymbol{g}_t = \mathbf{Q}_{t\|t-1}\mathbf{x}_t(\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}\mathbf{x}_t+\lambda)^{-1}$	$\boldsymbol{g}_t = \mathbf{Q}_{t\|t-1}\mathbf{x}_t(\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}\mathbf{x}_t+\lambda)^{-1}$	$\boldsymbol{g}_t = \mathbf{Q}_{t\|t-1}\mathbf{x}_t(\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}\mathbf{x}_t+\lambda)^{-1}$
$\mathbf{Q}_{t\|t} = \lambda^{-1}\mathbf{Q}_{t\|t-1} - \lambda^{-1}\boldsymbol{g}_t\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}$	$\mathbf{Q}_{t\|t} = \lambda^{-1}\mathbf{Q}_{t\|t-1} - \lambda^{-1}\boldsymbol{g}_t\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}$	$\mathbf{Q}_{t\|t} = \lambda^{-1}\mathbf{Q}_{t\|t-1} - \lambda^{-1}\boldsymbol{g}_t\mathbf{x}_t^*\mathbf{Q}_{t\|t-1}$
$\mathbf{Q}_{t+1\|t} = \mathbf{F}_t\mathbf{Q}_{t\|t}\mathbf{F}_t^* + \rho\mathbf{I}$	$\mathbf{Q}_{t+1\|t} = \mathbf{Q}_{t\|t} + \rho\mathbf{I}$	$\mathbf{Q}_{t+1\|t} = \|\alpha\|^2\mathbf{Q}_{t\|t} + \rho\mathbf{I}$
$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\mathbf{x}_t)^*$	$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\mathbf{x}_t)^*$	$\boldsymbol{\beta}_{t\|t} = \boldsymbol{\beta}_{t\|t-1} + \boldsymbol{g}_t(y_t - \boldsymbol{\beta}_{t\|t-1}^*\mathbf{x}_t)^*$
$\boldsymbol{\beta}_{t+1\|t} = \mathbf{F}_t\boldsymbol{\beta}_{t\|t}$	$\boldsymbol{\beta}_{t+1\|t} = \boldsymbol{\beta}_{t\|t}$	$\boldsymbol{\beta}_{t+1\|t} = \alpha\boldsymbol{\beta}_{t\|t}$

B. Simple Generalizations of RLS

Based on the RLS-KF correspondence, it is straightforward to generalize RLS by considering other simple settings of the SSM (5)–(6). The generalizations are guaranteed to be optimal under the assumed conditions because of their equivalence to the KF estimator.

For example, if $\mathbf{F}_t$ is known, then, under the assumption that $\mathbf{\Sigma}_t \equiv 0$ and $\varsigma_t^2 = c\lambda^t$, the KF recursions become those in the third column of Table I. These equations constitute the extended forgetting factor RLS, or EFRLS, proposed in [14], that incorporates the known coupling-effect among the components of $\boldsymbol{\beta}_t$, as described by $\mathbf{F}_t$. This algorithm can eliminate the bias in RLS, as will be further discussed in Sec. IV-C.

As another example, suppose that $\mathbf{F}_t$ is known and that $\mathbf{\Sigma}_t = \rho\,\varsigma_t^2\mathbf{I}$ and $\varsigma_t^2 = c\lambda^t$. In this case, the third equation in KF (the first column in Table I) can be rewritten as $\mathbf{Q}_{t+1|t} = \mathbf{F}_t\mathbf{Q}_{t|t}\mathbf{F}_t^* + \rho\mathbf{I}$. Combining this equation with the others leads to a generalization of EFRLS shown in the first column of Table II. This algorithm, which may be called EFRLS-2, has an additional parameter $\rho > 0$ that can be interpreted as the ratio of the system noise variance to the observation noise variance. In the special case of $\mathbf{F}_t \equiv \mathbf{I}$, EFRLS-2 reduces to a two-parameter RLS algorithm, or RLS-2, given in the second column of Table II (the special case with $\lambda = 1$ was also suggested in [1, pp. 94–95]). This algorithm does not depend on any prior knowledge of the SSM when λ and ρ are regarded as tuning parameters.

Finally, as shown in the last column of Table II, a three-parameter RLS, or RLS-3, can be obtained by assuming $\mathbf{F}_t = \alpha\mathbf{I}$ in EFRLS-2. Under this assumption, the components of $\boldsymbol{\beta}_t$ are uncoupled autoregressive (AR) processes of order 1 with a common AR coefficient α. As with ρ and λ in RLS-2, α can also be regarded as a tuning parameter.

Note that $\boldsymbol{\beta}_t$ is treated by RLS inherently as a time-invariant parameter (which explains why RLS is successful mainly in tracking slowly time-varying parameters) and by EFRLS as a deterministic function of t which is known completely once the initial value is given. In contrast, $\boldsymbol{\beta}_t$ is treated by EFRLS-2, RLS-2, and RLS-3 as a random process, where the randomness comes from the system noise and is handled by the parameter ρ. One can see from Table II that the presence of the system noise in $\boldsymbol{\beta}_t$ is manifested as an extra additive term $\rho\mathbf{I}$ in the prediction error covariance matrix $\mathbf{Q}_{t+1|t}$. In calculating the gain vector $\boldsymbol{g}_t$, this extra term effectively boosts the influence of the most recent data and curtains the influence of the estimation error assessment from the past. It also increases the estimation error covariance matrix $\mathbf{Q}_{t|t}$, which is to be used in the next round of estimation. Overall, by incorporating the system noise via the parameter ρ, these algorithms are prevented from being overly optimistic about their success in the past and thus overly conservative in adjusting themselves to the new data. This gives rise to the potential for the algorithms in Table II to have better tracking performance than RLS for rapidly time-varying parameters.

Of course, in the most general case under (5)–(6), the KF recursions provide the optimal solution. The so-called extended RLS (ERLS) algorithm, proposed recently in [13] (see also [11, p. 727]), is nothing but KF for the one-step predictor $\boldsymbol{\beta}_{n+1|n}$. Note that KF in the most general case does not necessarily have the exponential weighting, which RLS has, in its underlying cost function. This special feature of RLS is preserved by EFRLS proposed in [14] and by EFRLS-2 (and its variations in Table II) described in this section. The key lies in their common assumption that ς_t^2 is an exponential function of t. Both EFRLS and EFRLS-2 are special cases of a modified KF algorithm described in [8, Sec. 6.2] in which exponential data weighting is employed as an additional feature in the KF cost function.

C. Sources of Suboptimality of RLS Under SSM

Since RLS is better understood from the regression point of view, another approach to interpreting RLS under the SSM assumption is to reformulate (5)–(6) as a regression model similar to (3)–(4). To that end, let us fix n and use (5) to propagate $\boldsymbol{\beta}_t$ backwards from n to $t \leq n$. This leads to

$$\boldsymbol{\beta}_t = \mathbf{H}_{nt}\boldsymbol{\beta}_n - \sum_{s=t}^{n-1} \mathbf{G}_{st}\boldsymbol{v}_s, \tag{8}$$

and therefore, for $t = 0, 1, \ldots, n$,

$$y_t = \boldsymbol{\beta}_n^* \mathbf{H}_{nt}^* \boldsymbol{x}_t + e_t, \tag{9}$$

$$e_t := z_t - \sum_{s=t}^{n-1} \boldsymbol{v}_s^* \mathbf{G}_{st}^* \boldsymbol{x}_t, \tag{10}$$

where $\mathbf{G}_{st} := (\mathbf{F}_s \ldots \mathbf{F}_t)^{-1}$ and $\mathbf{H}_{nt} := \mathbf{G}_{n-1,t}$. Note that the measurement error e_t in (10) depends not only on the observation noise z_t but also on the explanatory signal $\boldsymbol{x}_t$ as well as the current and future values of the system noise $\{\boldsymbol{v}_t, \ldots, \boldsymbol{v}_{n-1}\}$.

With the help of (9)–(10), it is easy to see that there are at least three ways in which the optimality conditions of RLS can be violated under the SSM assumption:

- Failure to incorporate nontrivial transition matrices into the functional relationship between y_t and $\boldsymbol{x}_t$: the actual functional is $\boldsymbol{\beta}_n^* \mathbf{H}_{nt}^* \boldsymbol{x}_t$ rather than $\boldsymbol{\beta}_n^* \boldsymbol{x}_t$;
- Failure to incorporate nonexponentially time-varying measurement error variance: σ_t^2 is not necessarily an exponential function of t;
- Failure to incorporate the serial correlation of the measurement error process: the e_t are in general serially correlated.

The first offense results in biased estimation when $\mathbf{F}_t \neq \mathbf{I}$ (i.e., when the components of $\boldsymbol{\beta}_t$ are scaled and/or coupled as they evolve over time). The bias can be eliminated with the knowledge of $\mathbf{F}_t$ by replacing $\boldsymbol{x}_t$ with $\mathbf{H}_{nt}^* \boldsymbol{x}_t$ in (1). It is from this modified cost function that EFRLS was originally derived in [14]. The second and third offenses are responsible for the lack of estimation accuracy required for rapid tracking, which can be improved in general situations only if RLS is replaced with KF or the algorithms in Table II.

More specifically, the following results can be obtained (see Appendix for a proof).

Proposition 2 *Under the SSM* (5)–(6)*, the RLS solution* $\boldsymbol{\theta}_n$ *is, in general, a biased estimator of* $\boldsymbol{\beta}_n$*, and the bias, defined by* $\boldsymbol{b}_n := E(\boldsymbol{\theta}_n) - \boldsymbol{\mu}_n$*, with* $\boldsymbol{\mu}_n := E(\boldsymbol{\beta}_n)$*, can be expressed as*

$$\boldsymbol{b}_n = \boldsymbol{\Phi}_n^{-1} \left\{ \sum_{t=0}^{n} \lambda^{n-t} \boldsymbol{x}_t \boldsymbol{x}_t^* (\mathbf{H}_{nt} - \mathbf{I}) \boldsymbol{\mu}_n + \lambda^n \mathbf{P}_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_n) \right\},$$

which vanishes if $\mathbf{F}_t \equiv \mathbf{I}$ *and* $\boldsymbol{\mu} = \boldsymbol{\mu}_n$*, or if* $\boldsymbol{\mu} = \boldsymbol{\mu}_n = \mathbf{0}$*. The measurement error* $\{e_t\}$ *defined by* (10) *has mean zero and autocovariance function*

$$C_{t+\tau,t} := E(e_s e_t^*) = \varsigma_t^2 \delta_\tau + \sum_{s=t+\tau}^{n-1} \boldsymbol{x}_t^* \mathbf{G}_{st}^* \boldsymbol{\Sigma}_s \mathbf{G}_{s,t+\tau} \boldsymbol{x}_{t+\tau} \quad (\tau = 0, 1, \ldots),$$

where $\delta_0 := 1$ and $\delta_\tau := 0$ for all $\tau \neq 0$. Moreover, the e_t are uncorrelated if $\mathbf{\Sigma}_t \equiv \mathbf{0}$, or if $\mathbf{F}_t \equiv \mathbf{I}$ and

$$(\mathbf{\Sigma}_t + \cdots + \mathbf{\Sigma}_{n-1})\boldsymbol{x}_t = \mathbf{0} \quad (t = 0, 1, \ldots, n-1). \tag{11}$$

When $\mathbf{F}_t \equiv \mathbf{I}$, the variance of e_t is equal to $c\lambda^t$ if and only if

$$\varsigma_t^2 + \boldsymbol{x}_t^*(\mathbf{\Sigma}_t + \cdots + \mathbf{\Sigma}_{n-1})\boldsymbol{x}_t = c\lambda^t \quad (t = 0, 1, \ldots, n-1) \tag{12}$$

for some $c > 0$.

Remark 3. The conditions (11) and (12) are satisfied by the assumptions in Proposition 1. This explains, from a different standpoint, why RLS coincides with KF under these assumptions.

Remark 4. When $\mathbf{\Sigma}_t \neq \mathbf{0}$, the e_t can still satisfy (4) if (11) is true and if $\varsigma_t^2 = c\lambda^t$. Under these assumptions RLS is also optimal. This result is not easily seen by comparing the RLS and KF recursions.

As a simple example, consider a random-walk SSM where $\mathbf{F}_t \equiv \mathbf{I}$, $\mathbf{\Sigma}_t \equiv \rho\varsigma_0^2\mathbf{I}$, $\varsigma_t^2 \equiv \varsigma_0^2$, and $\boldsymbol{x}_t^*\boldsymbol{x}_t \equiv 1$. In this model, the variance of observation noise is constant, as commonly assumed in many applications, rather than exponential. Therefore, the necessity of nonconstant weighting in (1) can only be justified from the time-varying property of the functional relationship between y_t and $\boldsymbol{x}_t$ as a result of the random-walk coefficient. Is the exponential weighting in (1) optimal or nearly optimal?

This question can be answered with the help of Proposition 2, which asserts, for the given model, that $\sigma_t^2 = \varsigma_0^2 + (n-t)\rho\varsigma_0^2$. Clearly, σ_t^2 is a linear, rather than exponential, function of t, which rules out the possibility for RLS to be optimal. However, if λ is near unity so that $\lambda^{t-n} \approx 1 + (n-t)(1-\lambda)$, and if $\rho = 1 - \lambda$, then σ_t^2 can be approximated by the exponential function $c\lambda^t$ with $c := \varsigma_0^2\lambda^{-n}$. Moreover, λ being close to unity implies $\rho = 1 - \lambda$ is near zero. Therefore, for $\tau > 0$, the correlation coefficient

$$\frac{C_{t+\tau,t}}{\sigma_{t+\tau}\sigma_t} = \frac{(n-t-\tau)\rho\,\boldsymbol{x}_t^*\boldsymbol{x}_{t+\tau}}{\sqrt{(1+(n-t-\tau)\rho)\,(1+(n-t)\rho)}} = \mathcal{O}(\rho)$$

is small, meaning that the e_t are nearly uncorrelated. In summary, for the random-walk model, if the system-to-observation noise ratio ρ is small (hence $\boldsymbol{\beta}_t$ is nearly constant), then, by taking the forgetting factor to be $\lambda = 1 - \rho$ (which is near unity), the resulting e_t satisfy (4) approximately, and therefore the corresponding RLS estimator is nearly optimal.

V. NUMERICAL EXPERIMENTS

To demonstrate the potential of the generalized RLS algorithms in Table II, let us consider a numerical experiment, where

$$y_t = \beta_t + z_t$$

represents the noisy observation of a time-varying signal β_t which is a real-valued AR process of order 1 satisfying

$$\beta_t = a\beta_{t-1} + v_t \tag{13}$$

for some $a \in (0,1)$. The observation noise $\{z_t\}$ and the system noise $\{v_t\}$ are independent Gaussian white noise processes with mean zero and variances ς^2 and σ^2, respectively. In the experiment, $\varsigma^2 = 10^{-0.1}$ and $\sigma^2 = 1 - a^2$, so that $\text{Var}(\beta_t) = 1$ and hence the SNR, defined as $\text{Var}(\beta_t)/\varsigma^2$, is equal to 1 dB. This model, taken from [18], simulates the baseband form of the received (stationary) narrowband signal in a communications environment. It was also employed in [19] in a comparative study of adaptive algorithms. The problem is to estimate β_t (regarded as a time-varying parameter) on the basis of $\{y_t, y_{t-1}, \ldots, y_0\}$.

Fig. 1 shows a portion of a random realization of the signal and its estimates by different methods (all recursions begin with $\boldsymbol{\mu} = 0$, $\mathbf{P}_0 = 1$, and $\mathbf{Q}_0 = 1$). Since $\{y_t\}$ and $\{\beta_t\}$ are Gaussian and satisfy the SSM (5)–(6), with $\boldsymbol{x}_t \equiv 1$, $\mathbf{F}_t = a$, $\boldsymbol{\Sigma}_t \equiv \sigma^2$, and $\varsigma_t^2 \equiv \varsigma^2$, the KF estimates $\beta_{t|t}$ are optimal. In this one-dimensional case, KF coincides with RLS-3 that employs the true parameters $\lambda = 1$, $\rho = \sigma^2/\varsigma^2$, and $\alpha = a$; RLS is simply the exponentially weighted moving average (EWMA) of $\{y_t\}$, i.e., $c_t^{-1}\sum_{s=0}^{t}\lambda^{t-s}y_s$, where $c_t := \sum_{s=0}^{t}\lambda^{t-s} + \lambda^{t+1}$. As can be seen from Fig. 1, both RLS-2 and RLS-3 provide better tracking performance than RLS, as measured by the root mean-square error (RMSE) calculated over the entire data record of length 100. RLS, as expected, is only able to follow the slow trend in the data rather than the more rapid fluctuations. Of course, RLS-2 and RLS-3 are still suboptimal and inferior to KF because they do not use the complete information of the true SSM. Note that the tracking performance of RLS is notably improved by the introduction of ρ, but it is not further improved by the additional parameter α in this particular example, even though α takes the true value a of the SSM (as does KF).

A more comprehensive comparison of the methods is shown in Table III, where the RMSE is calculated on the basis of 5,000 independent runs of $\{y_t\}$, each being of length 100. Different values of a are employed to investigate the tracking performance under different rates of fluctuation of the signal β_t (small a for rapid

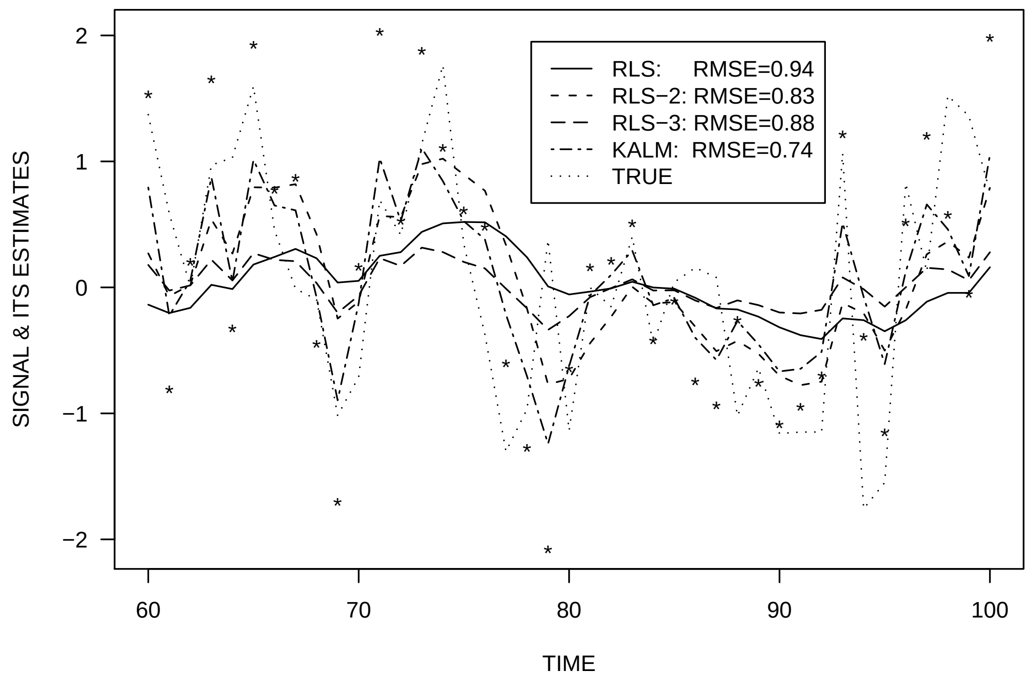


Figure 1: Comparison of tracking performance for the lowpass signal in (13): solid line, RLS ($\lambda = 0.9$); dashed line, RLS-2 ($\lambda = 0.9$, $\rho = 0.1$); long-dashed line, RLS-3 ($\lambda = 0.9$, $\rho = 0.1$, $\alpha = a = 0.5$); dot-dashed line, KF (same as RLS-3 with $\lambda = 1$, $\rho = \sigma^2/\varsigma^2 = 0.94$, $\alpha = a = 0.5$). The dotted line is the actual signal, and the dots represent the noisy observations. The SNR is equal to 1 dB.

fluctuation and large a for slow fluctuation). In all cases, the tuning parameters are: $\lambda = 0.9$ for RLS; $\lambda = 0.9$ and $\rho = 0.1$ for RLS-2; $\lambda = 1$ and $\rho = 0.1$ for RLS-2b; $\lambda = 0.9$, $\rho = 0.1$, and $\alpha = a$ for RLS-3; $\lambda = 1$, $\rho = 0.1$, and $\alpha = a$ for RLS-3b.

As can be seen, the RLS estimates in this experiment only achieves 50% efficiency of the KF estimates, as measured by the RMSE ratio η(RLS) := RMSE(KF)/RMSE(RLS) $\times$ 100%. With the additional parameters, especially with ρ, the RLS-2 and RLS-3 estimates are more efficient: for $a = 0.2$, 0.5, and 0.8, $\eta(\text{RLS-2}) =$ 65%, 72%, 84%, respectively, and $\eta(\text{RLS-3}) =$ 54%, 58%, 75%, respectively. Note that in all these cases the tuning parameters of RLS-2 and RLS-3 are chosen rather arbitrarily just to demonstrate the potential benefit of these parameters in improving the tracking performance of RLS. Fine tuning these parameter may result in further improvement.

Moreover, by comparing the RMSE of RLS-2 and RLS-3 with the RMSE of RLS-2b and RLS-3b,

TABLE III

RMSE FOR LOWPASS SIGNAL (13)

a	RLS	RLS-2	RLS-2b	RLS-3	RLS-3b	KF
0.2	0.86	0.66	0.70	0.80	0.82	0.43
0.5	0.80	0.57	0.60	0.71	0.74	0.41
0.8	0.63	0.39	0.42	0.44	0.48	0.33

respectively, we find that although KF requires $\lambda = 1$, the exponential weighting with $\lambda < 1$ in the generalized RLS algorithms is necessary when the complete information about the SSM is not available for the application of KF. This result justifies the algorithms in Table II as worthy alternatives to KF for estimating time-varying parameters, even if the SSM assumption is satisfied and the observation noise variance is nonexponential.

Finally, let us consider a nonlinear problem of tracking the amplitude and phase of a sinusoid in noise. This is a more general model than that discussed in [13] for simulating a mobile communications environment that suffers from not only the Doppler shift (due to the relative motion between the transmitter and the receiver) but also random channel fading. In this example,

$$y_t = (A + a_t)\sin\phi_t + z_t, \tag{14}$$

where a_t is the amplitude fluctuation around the average value of A and ϕ_t is the time-varying phase of the sinusoid. Assume that $\{a_t\}$ is a Gaussian AR(1) process with AR coefficient $a \in (0,1)$ and input noise variance σ_1^2. Assume further that $\{\phi_t\}$ satisfies the second-order difference equation

$$\phi_t - 2\phi_{t-1} + \phi_{t-2} = \varepsilon_t,$$

where $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma_2^2)$ is Gaussian and independent of $\{a_t\}$. Note that if ϕ_t is a pure quadratic function $\phi_t = \phi_0 + \omega_c t + \frac{1}{2}\psi t^2$, then $\phi_t - 2\phi_{t-1} + \phi_{t-2} = \psi$ for all t. Therefore, the signal in (14) is a randomly chirped sinusoid with time-varying amplitude $A + a_t$ and instantaneous chirp rate ε_t. The problem is to estimate the time-varying parameters a_t and ϕ_t on the basis of $\{y_t, y_{t-1}, \ldots, y_0\}$. Note that the ERLS-2 algorithm in [13] is not directly applicable to this problem because it assumes zero amplitude fluctuation and purely deterministic quadratic phase.

An equivalent (nonlinear) SSM for this problem is

$$\boldsymbol{\beta}_{t+1} = \mathbf{F}_t \boldsymbol{\beta}_t + \boldsymbol{v}_t, \quad y_t = h(\boldsymbol{\beta}_t, \mathbf{x}_t) + z_t,$$

TABLE IV

RMSE FOR CHIRPED SINUSOID (14)

	Amplitude Estimation						Phase Estimation					
Method	Min	Q1	Median	Mean	Q3	Max	Min	Q1	Median	Mean	Q3	Max
RLS	0.63	3.08	4.19	3.99	5.09	7.57	0.09	0.71	1.00	0.97	1.28	1.76
RLS-2	0.57	0.93	1.13	1.18	1.37	2.72	0.05	0.19	0.41	0.46	0.71	1.27
RLS-3	0.56	0.81	0.90	0.93	1.03	1.56	0.10	0.21	0.36	0.44	0.63	1.22
KF	0.42	0.65	0.76	0.79	0.89	1.60	0.09	0.16	0.22	0.26	0.31	1.06

where $\boldsymbol{\beta}_t := [a_t, \phi_t, \phi_{t-1}]^T$,

$$\mathbf{F}_t \equiv \mathbf{F} := \begin{bmatrix} a & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix},$$

$\boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma} := \mathrm{diag}(\sigma_1^2, \sigma_2^2, 0)$, $h(\boldsymbol{\beta}_t, \boldsymbol{x}_t) := (A + a_t)\sin\phi_t$, and $\boldsymbol{x}_t \equiv 1$. Because of the nonlinearity, one has to employ the so-called *extended* Kalman filter (e.g., [8, p. 195]), which takes the same form as the ordinary KF in Table I, except that $\boldsymbol{x}_t$ in the first two equations should be replaced by

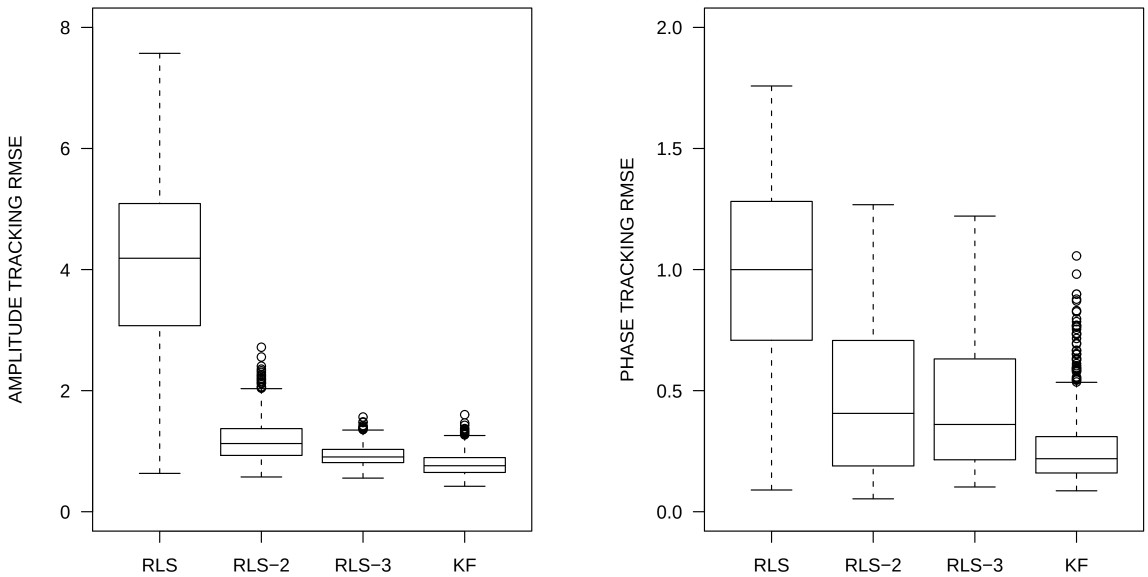
$$\boldsymbol{h}_t := \left.\frac{\partial h(\boldsymbol{\beta}, \boldsymbol{x})}{\partial \boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},\boldsymbol{x})=(\boldsymbol{\beta}_{t|t-1},\boldsymbol{x}_t)}$$

and that $\boldsymbol{\beta}_{t|t-1}^* \boldsymbol{x}_t$ in the fourth equation should be replaced by $h(\boldsymbol{\beta}_{t|t-1}, \boldsymbol{x}_t)$. For the problem in (14),

$$\boldsymbol{h}_t = [\sin\phi_{t|t-1}, (A + a_{t|t-1})\cos\phi_{t|t-1}, 0]^T.$$

Owing to the RLS-KF correspondence, the same method can be easily applied to the other algorithms in Tables I and II to obtain extended RLS algorithms for the nonlinear problem.

Fig. 2 and Table IV contain the results from a simulation study of these algorithms. Given each algorithm, RMSE is calculated for each of 1,000 realizations of $\{y_t\}$ in (14), each realization being of length 100. (To avoid ambiguity, realizations with $\phi_t \notin (-\pi, \pi)$ for some t are rejected and excluded from the total count.) The RMSEs for tracking a_t and ϕ_t are shown separately. The parameters employed to generate $\{y_t\}$ are: $A = 4$, $a = 0.85$, $\sigma_1^2 = 1 - a^2$ (so $\sigma_0^2 := \mathrm{Var}(a_t) = 1$), $\sigma_2^2 = 4 \times 10^{-6}$, $\phi_{-1} \sim \mathrm{N}(\xi, \sigma_2^2)$ and $\phi_{-2} \sim \mathrm{N}(2\xi, \sigma_2^2)$ with $\xi = -0.022$, $\varsigma^2 = 0.07225$ so the SNR, defined as $\frac{1}{2}(A^2 + \sigma_0^2)/\varsigma^2$, is equal to 15 dB. The tuning parameters of the algorithms are: $\lambda = 0.99$ for (extended) RLS; $\lambda = 0.99$ and $\rho = 0.1$



(a) (b)

Figure 2: Comparison of tracking performance for the chirped sinusoid in (14). (a) Boxplot of RMSE for tracking the amplitude fluctuation. (b) Boxplot of RMSE for tracking the phase. Results are based on 1,000 independent realizations of length 100. In a boxplot, the box contains the middle 50% of the sorted data with the middle bar representing the median; the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range (i.e., the length of the box) from the box.

for (extended) RLS-2; $\lambda = 0.99$, $\rho = 0.3$, and $\alpha = 0.8$ for (extended) RLS-3. The (extended) KF employs the exact model parameters $\mathbf{F}$, $\mathbf{\Sigma}$, and ς^2. The initial values for all recursions are $\boldsymbol{\mu} = [0, 0, \xi]^T$ and $\mathbf{P}_0 = \mathrm{diag}(0, 6\sigma_2^2, \sigma_2^2)$. Note that unlike the previous example, RLS-3 (and certainly RLS-2) does not become KF for this problem with any choice of the tuning parameters because of the inherent simplifications in RLS-3 (e.g., RLS-3 assumes $\mathbf{F} = \alpha\mathbf{I}$ but the actual $\mathbf{F}$ employed in KF does not have this form for any choice of α).

As can be seen, both RLS-2 and RLS-3 achieve a better tracking performance than RLS, thanks to the additional tuning parameters ρ and α. RLS fails completely in this case, giving essentially zero phase estimates and exploded amplitude estimates, as shown in Fig. 3, which cannot be improved by varying λ. As in the previous example, the parameters in RLS-2 and RLS-3 are not fine tuned to achieve the best performance, but the results are sufficient to demonstrate the main point: the generalized (and extended) RLS algorithms such as those in Table II should be considered as useful alternatives to KF in situations where KF is not optimal or not implementable because of incomplete knowledge of the SSM parameters.

VI. Concluding Remarks

In this article, a straightforward correspondence between RLS and KF has been identified that offers a unified development of several algorithms which generalize and extend the ordinary RLS. Since RLS inherently assumes a time-invariant functional relationship between the reference and explanatory signals observed with an measurement error that has exponentially time-varying variance, the tracking capability, or the lack of it, of RLS for time-varying parameters can only be explained by regarding the variability of the parameters as the time-varying measurement error which is not necessarily white noise with exponential variance but can become such exactly or approximately under suitable conditions.

Two numerical examples have demonstrated that with a few additional tuning parameters the generalized/extended RLS, easily derived from the RLS-KF correspondence, can achieve better tracking results than the ordinary RLS, but require less prior information than KF about the underlying system that governs the dynamics of the parameters under estimation.

Although the KF theory can be translated into the RLS language by the RLS-KF correspondence under the SSM assumption, the numerical and tracking properties of the generalized RLS algorithms such those in Table II need to be better understood analytically under more general assumptions about the time-varying parameters in order to further justify them as stand-alone adaptive algorithms for tracking these parameters. This subject seems worthwhile for future research.

Appendix: Proof of Proposition 2

It is easy to see from (10) that $E(e_t) = E(z_t) - \sum_{s=t}^{n-1} E(\boldsymbol{v}_s^*)\,\mathbf{G}_{st}^*\,\boldsymbol{x}_t = 0$, which, combined with (9), leads to $E(y_t^*) = \boldsymbol{x}_t^*\mathbf{H}_{nt}\boldsymbol{\mu}_n$. Therefore, it follows from (2) and (8) that

$$E(\boldsymbol{\theta}_n) = \boldsymbol{\Phi}_n^{-1}\left\{\sum_{t=0}^{n}\lambda^{n-t}\mathbf{x}_t\,\mathbf{x}_t^*\,\mathbf{H}_{nt}\boldsymbol{\mu}_n + \lambda^n\mathbf{P}_0^{-1}\boldsymbol{\mu}\right\}.$$

Clearly, $E(\boldsymbol{\theta}_n) \neq \boldsymbol{\mu}_n$ in general, but $E(\boldsymbol{\theta}_n) = \boldsymbol{\mu}_n$ if $\mathbf{H}_{nt} \equiv \mathbf{I}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_n$, or if $\boldsymbol{\mu} = \boldsymbol{\mu}_n = \mathbf{0}$. Moreover, it follows from (10) that for any $\tau \geq 0$, $C_{t+\tau,t}$ can be expressed as

$$\varsigma_t^2 \delta_\tau + \sum_{s=t+\tau, s'=t}^{n-1} E(\boldsymbol{v}_s^* \mathbf{G}_{s,t+\tau} \boldsymbol{x}_{t+\tau} \boldsymbol{x}_t^* \mathbf{G}_{s,t}^* \boldsymbol{v}_{s'}) =$$

$$\varsigma_t^2 \delta_\tau + \sum_{s=t+\tau, s'=t}^{n-1} \mathrm{tr}[\mathbf{G}_{s,t+\tau} \boldsymbol{x}_{t+\tau} \boldsymbol{x}_t^* \mathbf{G}_{s',t}^* E(\boldsymbol{v}_{s'} \boldsymbol{v}_s^*)].$$

The first part of the assertion follows from the fact that $E(\boldsymbol{v}_{s'} \boldsymbol{v}_s^*) = \delta_{s-s'} \boldsymbol{\Sigma}_s$. The proof is complete upon noting that with $\mathbf{F}_t \equiv \mathbf{I}$ we have $C_{t+\tau,t} = \varsigma_t^2 \delta_\tau + \boldsymbol{x}_t^* (\boldsymbol{\Sigma}_{t+\tau} + \cdots + \boldsymbol{\Sigma}_{n-1}) \boldsymbol{x}_{t+\tau}$.

References

[1] T. C. Hsia, *System Identification*, Lexington, MA: Lexington Books, 1977.

[2] E. Eweda and O. Macchi, "Tracking error bounds of adaptive nonstationary filtering," *Automatica*, vol. 21, no. 3, pp. 293–302, 1985.

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

[4] L. Guo, L. Ljung, and P. Priouret, "Performance analysis of forgetting factor RLS algorithm," *Int. J. Adaptive Contr., Signal Processing*, vol. 7, pp. 525–537, 1993.

[5] S. Bittanti and M. Campi, "Bounded error identification of time-varying parameters by RLS techniques," *IEEE Trans. Automat. Contr.*, vol. 39, no. 5, pp. 1106–1110, 1994.

[6] M. C. Campi, "Exponentially weighted least squares identification of time-varying systems with white disturbances," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2906–2914, 1994.

[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, San Diego, CA: Academic Press, 1970.

[8] A. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs, NJ: Prentice Hall, 1979.

[9] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, UK: Cambridge University Press, 1989.

[10] H. Lütkepohl, *Introduction to Multiple Time Series Analysis*, 2nd edn., New York: Springer-Verlag, 1993.

[11] S. Haykin, *Adaptive Filter Theory*, 3rd edn., Upper Saddle River, NJ: Prentice Hall, 1996.

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

[13] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1118–1128, 1997.

[14] Y. Zhu, "Efficient recursive state estimator for dynamic systems without knowledge of noise covariances," *IEEE Trans. Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 102–114, 1999.

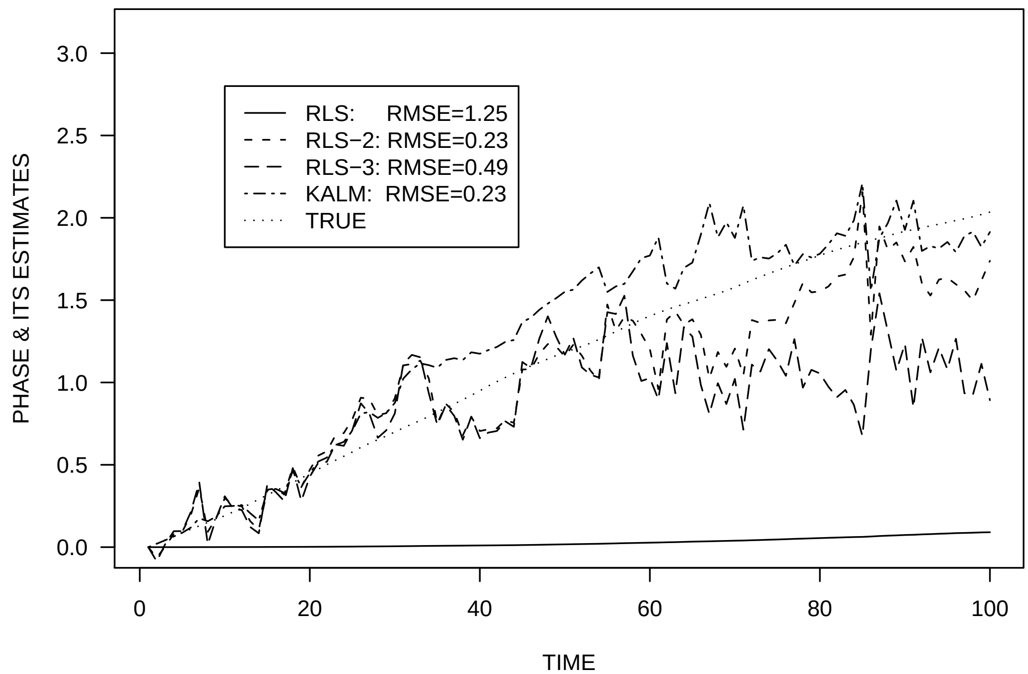
[15] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd edn., New York: Wiley & Sons, 1973.

[16] J. Wieslander and B. Wittenmark, "An approach to adaptive control using real-time identification" *Automatica*, vol. 7, pp. 211–217, 1971.

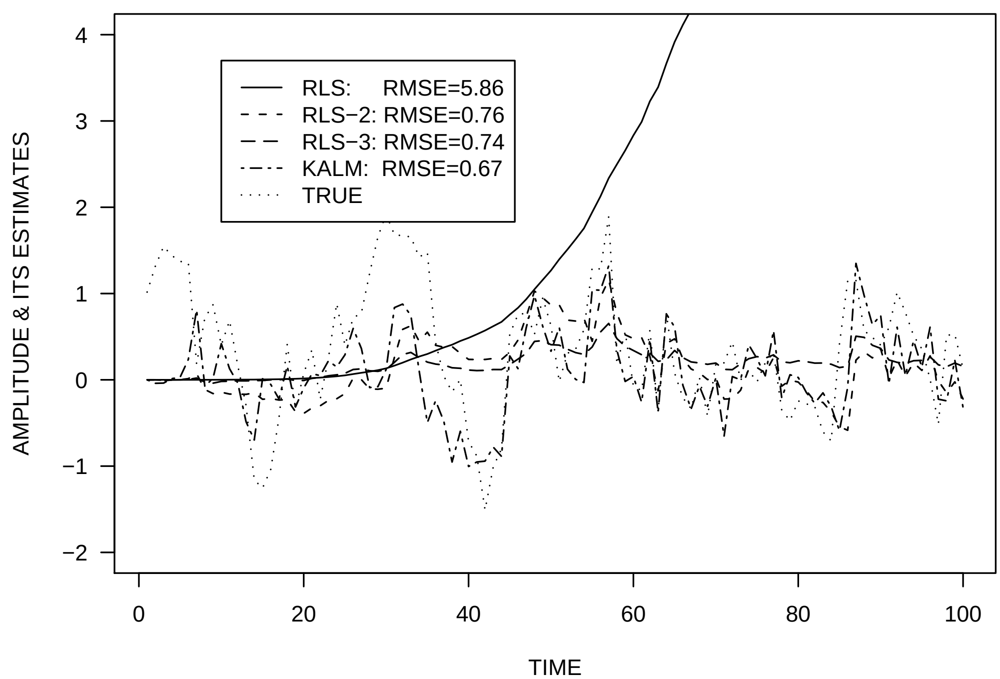
[17] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd edn., New York: Springer-Verlag, 1991.

[18] P. Wei, J. R. Zeidler, and W. H. Ku, "Adaptive recovery of a chirped signal using the RLS algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 363–376, 1997.

[19] E. Eweda, "Comparison of RLS, LMS, and Sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2937–2944, 1994.



(a)



(b)

Figure 3: Comparison of tracking performance for (a) the phase ϕ_t and (b) the amplitude a_t of a 100-point realization of the chirped sinusoid in (14). Parameters are the same as those for Fig. 2 and Table IV.