STATE-SPACE RLS

Mohammad Bilal Malik
College of Electrical and Mechanical Engineering
National University of Sciences and Technology, Pakistan

ABSTRACT

Kalman filter is linear optimal estimator for random signals. We develop state-space RLS that is counterpart of Kalman filter for deterministic signals i.e. there is no process noise but only observation noise. State-space RLS inherits its optimality properties from the standard least squares. It gives excellent tracking performance as compared to existing forms of RLS. A large class of signals can be modeled as outputs of neutrally stable unforced linear systems. State-space RLS is particularly well suited to estimate such signals. The paper commences with batch processing the observations, which is later extended to recursive algorithms. Comparison and equivalence of Kalman filter and state-space RLS become evident during the development of the theory. State-space RLS is expected to become an important tool in estimation theory and adaptive filtering.

1. INTRODUCTION

Classical optimal filter theory revolves around Kalman filter, which is thought to be linear optimum tracker for random signals. In case of deterministic signals, the process noise is absent and the Kalman Gain approaches zero asymptotically. This in turn renders Kalman filter useless. The common way of getting around this limitation is to add some process noise [2]. However, this approach modifies the problem rather than finding an optimal solution. Recursive least-squares (RLS) appears to be a substitute of Kalman filter for such cases. However, a form of RLS that could become a true counterpart of Kalman filter for deterministic signals has not been developed previously. The best work so far is a state-space approach of RLS given by Sayed and Kailath [1]. A one-to-one correspondence between their RLS and Kalman filter can be drawn [9]. The main problem with this formulation is the loss of model information, which results in poor tracking performance. Haykin et al. [10] have given extended RLS algorithms that improve the performance.

We take a different route from the classical approach and develop a truly state-space formulation of RLS. The signal to be estimated is modeled as output of an unforced linear time-invariant system. The work begins with batch processing of noise corrupted observations, which is later extended to recursive algorithms whilst employing the concept of exponential forgetting. We cover core issues like stability conditions and initialization methods. Different forms of the filter are discussed. A steady state solution is derived which is computationally efficient as compared to most other estimation algorithms. Finally an example demonstrating the application and power of state-space RLS concludes the paper.

2. STATE-SPACE MODEL

Consider a discrete-time unforced system

$$\begin{aligned} x[k+1] &= A\,x[k] \\ y[k] &= C\,x[k] + v[k] \end{aligned} \tag{2.1}$$

where $x \in R^n$ and $y \in R^m$. Observation noise is represented by $v[k]$. We make no assumptions about the nature of this noise at this stage. Notice the absence of '*process noise*'. We assume that the pair *(A, C)* is *l*-step observable and the matrix *A* is invertible. If the system is discretized version of a continuous-time system, then matrix *A* is derived from the state-transition matrix, which is always full rank. At this point we make no assumptions about the stability of the system. However, as we progress certain restriction will have to be made.

3. LEAST SQUARES OBSERVATION

We begin our discussion by batch processing the observations. From (2.1), we can write *p* different equations as follows.

$$\mathbf{y}[k] = \begin{bmatrix} y[k-p+1] \\ y[k-p+2] \\ \vdots \\ y[k-2] \\ y[k-1] \\ y[k] \end{bmatrix} = \begin{bmatrix} CA^{-p+1}x[k] \\ CA^{-p+2}x[k] \\ \vdots \\ CA^{-2}x[k] \\ CA^{-1}x[k] \\ Cx[k] \end{bmatrix} + \mathbf{v}[k] \tag{3.1}$$

where $p \geq l$ and observation noise vector is given by $\mathbf{v}[k] = \begin{bmatrix} v[k] & v[k-1] & \cdots & v[k-p+2] & v[k-p+1] \end{bmatrix}^T$. We may write (3.1) as

$$\mathbf{y}[k] = Hx[k] + \mathbf{v}[k] \tag{3.2}$$

where $\mathbf{y}[k]$ is the observation vector and *H* is defined as

0-7803-7663-3/03/$17.00 ©2003 IEEE

$$H = \begin{bmatrix} CA^{-p+1} \\ CA^{-p+2} \\ \vdots \\ CA^{-1} \\ C \end{bmatrix}_{mp\times n} \tag{3.3}$$

The solution of system (3.2) in terms of least squares is given as follows [5], [11]

$$\hat{x}[k] = (H^T H)^{-1} H^T \mathbf{y}[k] \tag{3.4}$$

The matrix $(H^T H)^{-1} H^T$ has dimensions $n \times mp$ and can be calculated off-line. This is an important formula though the computations could be intense if p is large. Certain optimality properties associated with this solution can be found in [5], [11]. Another variant of (3.4) could be weighted least-squares solution. In this case we define a weighting matrix W and the corresponding solution is

$$\hat{x}[k] = (H^T W H)^{-1} H^T W \mathbf{y}[k] \tag{3.5}$$

4. RECURSIVE ALGORITHM

We assume pre-windowing of the observations i.e.

$$y[k] = 0, \quad k < 0 \tag{4.1}$$

Let $k = p-1$. Define the weighting matrix as

$$W[k] = \begin{bmatrix} \lambda^k & 0 & \cdots & 0 & 0 \\ 0 & \lambda^{k-1} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \lambda & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{4.2}$$

where $0 < \lambda \le 1$. Starting as scalars for $k = 0$, the matrices H, W and $\mathbf{y}$ grow in size with time. We borrow a few symbols from classical formulation of RLS [9], however they will be used in a different meaning here. Let

$$\begin{aligned} \Phi[k] &= H^T[k] W[k] H[k] \\ z[k] &= H^T[k] W[k] \mathbf{y}[k] \end{aligned} \tag{4.3}$$

which gives

$$\Phi[k] x[k] = z[k] \quad \text{or} \quad x[k] = \Phi^{-1}[k] z[k] \tag{4.4}$$

4.1. Recursive Update of $\Phi[k]$

We have from (3.3), (4.2) and (4.3)

$$\Phi[k] = \Big[\lambda^k \left(A^T\right)^{-k} C^T C A^{-k} + \lambda^{k-1} \left(A^T\right)^{-k+1} C^T C A^{-k+1} + \cdots + \lambda A^{-T} C^T C A^{-1} + C^T C\Big] \tag{4.5}$$

Similarly

$$\Phi[k-1] = \Big[\lambda^{k-1} \left(A^T\right)^{-k+1} C^T C A^{-k+1} + \cdots + \lambda A^{-T} C^T C A^{-1} + C^T C\Big] \tag{4.6}$$

Comparing (4.5) and (4.6)

$$\begin{aligned} \Phi[k] &= \lambda A^{-T} \Phi[k-1] A^{-1} + C^T C \\ A^T \Phi[k] A &= \lambda \Phi[k-1] + A^T C^T C A \end{aligned} \tag{4.7}$$

4.2. Recursive Update of $\Phi^{-1}[k]$

We make use of matrix inversion lemma (see e.g. [9]) to arrive at the following result

$$\begin{aligned} \left[A^T \Phi[k] A\right]^{-1} = {} & \lambda^{-1} \Phi^{-1}[k-1] - \\ & \lambda^{-2} \Phi^{-1}[k-1] A^T C^T \times \\ & \left[I + \lambda^{-1} C A \Phi^{-1}[k-1] A^T C^T\right]^{-1} \times \\ & C A \Phi^{-1}[k-1] \end{aligned} \tag{4.8}$$

Rearranging we get the Riccati equation for state-space RLS

$$\begin{aligned} \Phi^{-1}[k] = {} & \lambda^{-1} A \Phi^{-1}[k-1] A^T - \\ & \lambda^{-2} A \Phi^{-1}[k-1] A^T C^T \times \\ & \left[I + \lambda^{-1} C A \Phi^{-1}[k-1] A^T C^T\right]^{-1} \times \\ & C A \Phi^{-1}[k-1] A^T \end{aligned} \tag{4.9}$$

If we are dealing with a single output system i.e. m=1, then equation (4.9) does not require matrix inversion. Otherwise, matrix of dimension $m \times m$ is required to be inverted.

4.3. Kalman Gain

Define the Kalman Gain as

$$\begin{aligned} K[k] = {} & \lambda^{-1} A \Phi^{-1}[k-1] A^T C^T \times \\ & \left[I + \lambda^{-1} C A \Phi^{-1}[k-1] A^T C^T\right]^{-1} \end{aligned} \tag{4.10}$$

Equation (4.9) can now be written as

$$\begin{aligned} \Phi^{-1}[k] = {} & \lambda^{-1} A \Phi^{-1}[k-1] A^T - \\ & \lambda^{-1} K[k] C A \Phi^{-1}[k-1] A^T \end{aligned} \tag{4.11}$$

Rearranging (4.10)

$$\begin{aligned} K[k] &= \Big[\lambda^{-1} A \Phi^{-1}[k-1] A^T + \\ &\quad \lambda^{-1} K[k] C A \Phi^{-1}[k-1] A^T \Big] C^T \\ &= \Phi^{-1}[k] C^T \end{aligned} \tag{4.12}$$

4.4. Recursive Computation of $z[k]$

From equation (3.1), (4.3) etc.

$$z[k] = \Big[\lambda^k \left(A^T\right)^{-k} C^T y[0] + \cdots + \lambda A^T C^T y[k-1] + C^T y[k]\Big] \tag{4.13}$$

which gives us

$$z[k] = \lambda A^{-T} z[k-1] + C^T y[k] \tag{4.14}$$

Equations (4.9) and (4.14) constitute RLS in its general form. We will later discuss special cases of RLS that would be computationally efficient.

4.5. Kalman Estimator Form

We derive a form that is similar to the usual Kalman Estimator form discussed in the literature. Define the predicted states as

$$\bar{x}[k] = A\hat{x}[k-1] \quad (4.15)$$

Similarly the predicted output is

$$\bar{y}[k] = C\bar{x}[k] \quad (4.16)$$

From equation (4.4), (4.11), (4.12) and (4.14)

$$\hat{x}[k] = \bar{x}[k] + K[k]\big(y[k] - \bar{y}[k]\big) \quad (4.17)$$

The final expression in (4.17) is the same as Kalman estimator form. We define the prediction error as

$$\varepsilon[k] = y[k] - \bar{y}[k] \quad (4.18)$$

The prediction error is also referred to as innovations in Kalman filter theory.

4.6. State-Space Representation of RLS estimator

Since State-Space RLS estimator is a linear time-varying filter, the map from y to $\hat{x}$ can be represented in state-space. It can be shown that the quadruplet $\left\{\lambda A^{-T}, C^{T}, \lambda\Phi^{-1}[k]A^{-T}, K[k]\right\}$ constitutes the requisite state-space matrices.

4.7. Stability and Convergence of $z[k]$

The stability of (4.14) is guaranteed if the eigenvalues of λA^{-1} have magnitude less than or equal to unity. We have to impose a restriction on the upper bound of λ

$$\lambda \le \left|\min \mathit{Eigenvalues}(A)\right| \quad (4.19)$$

An interesting observation is that an unstable system with all its eigenvalues outside the unit circle will not cause a problem for $\lambda \le 1$.

4.8. Stability and Convergence of $\Phi[k]$

The properties of the following matrix difference equation are discussed in [8]

$$\Phi[k+1] = P^{T}\Phi[k]P + Q \quad (4.20)$$

where $Q \ge 0$ is the forcing function. With any initial condition $\Phi[0] \ge 0$, the solution of (4.20) is well behaved for all $k \ge 0$ if all the Eigenvalues of P are strictly less than unity. Furthermore the steady state solution given by

$$\lim_{k\to\infty} \Phi[k] = \Phi \quad (4.21)$$

satisfies the discrete Lyapunov equation

$$P^{T}\Phi P - \Phi = -Q \quad (4.22)$$

However, if all the Eigenvalues of P are on the unit circle then the unique steady state solution of (4.20) is $\lim_{k\to\infty} \Phi[k] \to \infty$ or equivalently $\lim_{k\to\infty} \Phi^{-1}[k] = 0$. Let

$$P = \sqrt{\lambda}A^{-1}$$
$$Q = C^{T}C \quad (4.23)$$

Comparing (4.7) with (4.20), we see that the condition of convergence of (4.7) is

$$\sqrt{\lambda} \le \left|\min \mathit{Eigenvalues}(A)\right| \quad (4.24)$$

Condition (4.19) implies (4.24) because $\lambda \le 1$, therefore we take (4.19) as the condition of stability of the estimator.

4.9. Neutrally Stable Systems

If the system (2.1) is Poisson or neutrally stable, A has Eigenvalues on the unit circle. This class of systems plays a central role in state space RLS because a large number of signals can be modeled by these systems. It is easy to verify that in case of neutrally stable systems the condition (4.19) translates to

$$\lambda \le 1 \quad (4.25)$$

4.10. Initializing RLS and Peaking Phenomenon

Proper initialization of any recursive algorithm is an important phase. There may be certain practical difficulties involved like the peaking phenomenon in high gain observers [3], [4]. However, the algorithm presented in equations (4.9) and (4.14) offers a self contained method for proper initialization and hence avoids problems like peaking. As the observations start to appear, we wait for l-samples ($H[k]$ becomes full rank at this instant because of l-step observability assumption). Using the definitions in (3.1) to (3.3) with $p=l$, we can calculate $\hat{x}[l-1]$. The matrix $(H^{T}H)^{-1}H^{T}$ in (3.4) can be calculated offline and the observation vector is $\mathbf{y}[l-1] = \begin{bmatrix} y[0] & y[1] & \cdots & y[l-1] \end{bmatrix}^{T}$. The only problem for this initialization scheme could be the observation noise. This would still give us a legitimate estimate to start off the recursion. It is worth mentioning that Kalman filter [7] and classical RLS [9] face certain difficulties in the context of proper initialization.

5. STEADY STATE SOLUTION

Steady state solutions in the realm of Kalman filtering have been important for their computational simplicity. Following somewhat similar lines, we investigate the steady state solution of the matrix difference equation (4.7), which is independent of the observations.

5.1. Computing Φ

We observe that Φ is a solution of [12]

$$\lambda A^{-T}\Phi A^{-1} - \Phi = -C^{T}C \quad (5.1)$$

if

$$\sqrt{\lambda} < \left|\min \mathit{Eigenvalues}(A)\right| \quad (5.2)$$

The significance of this result is that Φ and hence Φ^{-1} can be calculated offline.

5.2. Kalman Gain

Kalman gain (4.12) is time-invariant in this case and is given by $\Phi^{-1}C^{T}$.

5.3. Neutrally Stable Systems

For a neutrally stable system, the requirement (5.2) prohibits the use of infinite memory. For steady state solution the condition of convergence is therefore

$$\lambda < 1 \quad (5.3)$$

6. EXAMPLE

We want to track a sinusoidal wave $r(t) = \sin(\omega t + \phi)$ using steady state RLS filter. The discrete state-space model with sampling time T corresponding to equation (2.1) for this continuous signal is

$$x[k+1] = \begin{bmatrix} \cos(\omega T) & \sin(\omega T) \\ -\sin(\omega T) & \cos(\omega T) \end{bmatrix} x[k]; \quad x[0] = \begin{bmatrix} \sin(\phi) \\ \cos(\phi) \end{bmatrix}$$

$$y[k] = [1 \quad 0]x[k] + v[k]$$

The parameters used are $\omega = 0.1$, $\phi = \pi/3$, $T_S = 0.1$ and $\lambda = 0.95$. White zero-mean noise $v[k]$ with variance 0.001 corrupts the observations. The estimation performances for different signal models are illustrated in Figure 1. The estimation errors are $e_1[k] = x_1[k] - \hat{x}_1[k]$ and $e_2[k] = x_2[k] - \hat{x}_2[k]$, where $\hat{x}_1[k]$ and $\hat{x}_2[k]$ are the state estimates. The errors in constant velocity model are large as compared to constant acceleration model. However, since the models are not exact, exponential convergence is not observed for either case. For the case of exact model, perfect convergence in mean is achieved. In this case the prediction error process (4.18) asymptotically becomes white noise and can be regarded as innovations process in Kalman filtering. This in fact indicates the optimality of state-space RLS because it is impossible to reduce the prediction error to a level less than the observed white noise.

All of the signal models used in this example are neutrally stable. Improper initialization results in peaking which could have been avoided by using the method of Section 4.10.

State-space RLS has served as an optimal filter without any modification in the problem. Kalman filter on the other hand would have required addition of process noise which is equivalent to altering the problem or some other enhancement which may result in loss of optimality. Hence state-space RLS is the appropriate and optimal solution for deterministic signals.

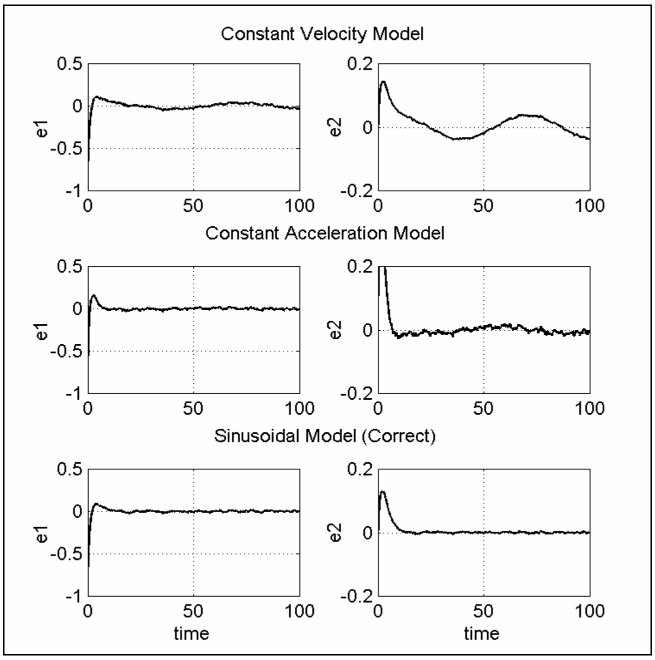

Figure 1

7. CONCLUSION

In this paper we have developed optimal linear estimator for deterministic signals. The method of least squares is one of the most common estimation schemes. However, our formulation of state-space RLS is much more flexible and powerful. Our work covers batch processing, recursive updates, stability conditions, initialization and steady state solutions etc. An example illustrates the power and usefulness of state-space RLS.

We have only covered preliminary details in this paper. The theory however, is expected to attract a much wider area of research. It will give rise to new algorithms and solutions for a diverse range of estimation and adaptive filtering problems.

8. REFERENCES

[1] A. H. Sayed and Thomas Kailath , "A state-space approach to adaptive RLS filtering", *IEEE Signal Processing Magazine*, Vol. 11 No. 3, 18-60, July 1994.

[2] Franklin, Powell and Workman, *Digital Control of Dynamical Systems, 2nd Edition.* Addison-Wesley, 1990

[3] H. K. Khalil, *Nonlinear Systems, 3rd Edition*. Prentice Hall, 2002

[4] H. K. Khalil, High-gain observers in nonlinear feedback control. In H. Nijmeijer and T. Fossen, editors, *New Directions in Nonlinear Observer Design,* Vol 244 of *Lecture Notes in Control and Information Science,* pages 249-268. Spinger, London, 1999

[5] Henry Stark and John Woods, *Probability, Random Processes, and Estimation Theory for Engrs.* Prentice Hall, 1994

[6] M. D. Srinath, P.K. Rajasekaran and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications.* Prentice Hall, 1996

[7] Robert Grover Brown and Patrick Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering* John Wiley & Sons, 1997

[8] R. R. Bitmead, M. R. Gevers, I. R. Petersen, "Monotonicity and Stabilization properties of the Riccati difference equation: Propositions, lemmas, theorems, fallacious conjectures and counter examples", *System and Control Letters 5:309-315, 1985*

[9] Simon Haykin, *Adaptive Filter Theory, 4th Edition.* Prentice Hall, 2001

[10] Simon Haykin, A. H. Syed, J. Zeidler, P. Yee, P. Wei , "Adaptive Tracking of Linear Time-Variant Systems by Extended RLS Algorithms", *IEEE Trans on Signal Processing*, Vol. 45 No. 6, 1118-1128, May 1997.

[11] Steven M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Prentice Hall, 1993

[12] Wilson J. Rugh, *Linear System Theory, 2nd Edition.* Prentice Hall, 1996