Scientific
Research

RLS Wiener Predictor with Uncertain Observations in Linear Discrete-Time Stochastic Systems

Seiichi Nakamori[1], Raquel Caballero-Águila[2], Aurora Hermoso-Carazo[3], Josefa Linares-Pérez[3]

[1]Department of Technical Education, Kagoshima University, Kagoshima, Japan; [2]Departamento de Estadística e I.O, Universidad de Jaén, Paraje Las Lagunillas, Jaén, Spain; [3]Departamento de Estadística e I.O, Universidad de Granada, Campus Fuentenueva S/N, Granada, Spain.
Email: nakamori@edu.kagoshima-u.ac.jp, raguila@ujaen.es, {ahermoso, jlinares}@ugr.es

ABSTRACT

This paper proposes recursive least-squares (RLS) l-step ahead predictor and filtering algorithms with uncertain observations in linear discrete-time stochastic systems. The observation equation is given by $y(k) = \gamma(k) z(k) + v(k)$, $z(k) = Hx(k)$, where $\{\gamma(k)\}$ is a binary switching sequence with conditional probability. The estimators require the information of the system state-transition matrix Φ, the observation matrix H, the variance $K(k,k)$ of the state vector $x(k)$, the variance $R(k)$ of the observation noise, the probability $p(k) = P\{\gamma(k) = 1\}$ that the signal exists in the uncertain observation equation and the $(2,2)$ element $\left[P(k \mid j)\right]_{2,2}$ of the conditional probability of $\gamma(k)$, given $\gamma(j)$.

Keywords: *Estimation Theory, Synthesis of Stochastic Systems, RLS Wiener Predictor, Uncertain Observations, Markov Probability*

1. Introduction

The estimation problem given uncertain observations is an important research topic in the area of detection and estimation problems in communication systems [1]. Nahi [2], assuming that the state-space model is given, proposes the RLS estimation method with uncertain observations, when the uncertainty is modeled in terms of independent random variables, and the probability that the signal exists in each observation is available. The term uncertain observations refers to the fact that some observations may not contain the signal and consist only of observation noise. In Hadidi and Schwartz [3], Nahi's results are extended to the case where the variables modeling the uncertainty are not necessarily independent.

In the above studies, it is assumed that the state-space model for the signal is given. However, in real applications, the state-space modeling errors might degrade the estimation accuracy. Nakamori [4] derived the RLS Wiener fixed-point smoothing and filtering algorithms, based on the invariant imbedding method, from uncertain observations with uncertainty modeled by independent random variables. In the derivation of such RLS Wiener

estimators, the state-transition matrix Φ, the observation matrix H, the variance $K(k,k)$ of the state vector $x(k)$, the variance $R(k)$ of the observation noise $v(k)$ and the observed values $y(k)$ are used. Moreover, Nakamori *et al.* [5], based on the innovation approach, proposed the RLS Wiener fixed-point smoother and filter in linear discrete-time stochastic systems. Here, the observation equation is given by

$y(k) = \gamma(k) z(k) + v(k),\ z(k) = Hx(k)$, where $\{\gamma(k)\}$ is a binary switching sequence with conditional probability. The innovation process is given by

$$\upsilon(s) = y(s) - \hat{y}(s, s-1),$$
$$\hat{y}(s, s-1) = P_{2,2}(s) H \Phi \hat{x}(s-1, s-1)$$

in terms of the $(2,2)$ element $\left[P(k \mid j)\right]_{2,2}$ of the conditional probability of $\gamma(k)$, given $\gamma(j)$ (see Nakamori *et al.* [5,6] for details).

In the current paper, under the same assumptions for the observation equation as in Nakamori *et al.* [5], an algorithm for the RLS Wiener l-step ahead predictor is derived, based on the invariant imbedding method. Thus, the observation equation is given by

$$y(k) = \gamma(k)z(k) + v(k), \quad z(k) = Hx(k),$$

where $\{\gamma(k)\}$ is a binary switching sequence with conditional probability. The observation equation adopted in this paper is suitable, for example, to model remote sensing situations with data transmission in multichannels, where the independence assumption of the variables describing the uncertainty in the observations is not realistic.

The estimators require the information of the system state-transition matrix Φ, the observation matrix H, the variance $K(k,k)$ of the state vector $x(k)$, the variance $R(k)$ of the observation noise, the probability $p(k) = P\{\gamma(k) = 1\}$ that the signal exists in the uncertain observation equation and the $(2,2)$ element $\left[P(k\,|\,j) \right]_{2,2}$ of the conditional probability of $\gamma(k)$, given $\gamma(j)$. The RLS Wiener prediction and filtering algorithms are summarized in Theorem 1 and its proof is deferred to the Appendix. The main issues in this paper which are different from those in Nakamori *et al.* [5] are concerned with the algorithm derivation; namely:

1) The prediction estimate is given as a linear transformation of the observed values.

2) The prediction algorithms are derived on the basis of the invariant imbedding method.

The current paper's main contribution is the derivation of a recursive least-squares algorithm for the predictor and filter design in systems with non-independent uncertain observations, using covariance information. Without making use of the state-space model, the algorithm is obtained from the autocovariance functions of the signal and the observation noise, the probability that the signal exists in the observed values and the (2,2) element of the conditional probability matrices of the sequence which describes the uncertainty in the observations. This approach is suitable in many practical situations where the equation generating the signal process is unknown, thus being not possible to use the state-space model to address the estimation problem. The deduction of the algorithm is mainly based on an invariant imbedding method.

2. Problem Formulation

Consider the following observation equation

$$y(k) = \gamma(k)z(k) + v(k), \quad z(k) = Hx(k), \quad (1)$$

where $z(k)$ is a signal, $x(k)$ is the $n \times 1$ zero-mean state vector and H is the $m \times n$ observation matrix.

- The sequence $\{v(k)\}$ is a white noise with zero mean and the variance of $v(k)$ is $R(k)$; that is,

$$E\left[v(k)v^T(s) \right] = R(k)\delta_K(k-s), \quad (2)$$

where $\delta_K(\cdot)$ denotes the Kronecker delta function.

- The random sequence $\{\gamma(k)\}$, which describes the

uncertainty in the observations, has the following stochastic properties [3]:

(P-1) $\gamma(k)$ is a discrete-time random variable taking the values 0 or 1 with $P\{\gamma(k) = 1\} = p(k)$. Therefore, $p(k)$ represents the probability that the observed value $y(k)$ contains the signal $z(k)$; this probability is assumed to be nonzero.

(P-2) The noise $\{\gamma(k)\}$ is a sequence of random variables with initial probability vector $(1 - p(0), p(0))^T$ and conditional probability matrix $P(k\,|\,j)$. The $(2,2)$ element of the conditional probability matrix of $\gamma(k)$ given $\gamma(j)$, is independent of j, for $j < k$; that is,

$$\left[P(k\,|\,j) \right]_{2,2} = \frac{E\left[\gamma(j)\gamma(k) \right]}{P\{\gamma(j) = 1\}} = P_{2,2}(k), \quad (3)$$

$$j = 0, \cdots, k-1.$$

- The state process $\{x(k)\}$ and the sequences $\{\gamma(k)\}$ and $\{v(k)\}$ are mutually independent.

Let us introduce the system matrix Φ in the state-space model for the state vector $x(k)$ and the variance $K(s,s)$ of the state vector $x(s)$. Then the autocovariance function $K_z(k,s)$ of the signal $z(k)$ is factorized as

$$K_z(k,s) = HK(k,s)H^T,$$

$$K(k,s) = A(k)B^T(s), \quad 0 \le s \le k \quad (4)$$

$$A(k) = \Phi^k, B^T(s) = \Phi^{-s}K(s,s).$$

The purpose of this paper is to design a covariance-based recursive algorithm to obtain the $l-$step ahead prediction estimate of $x(k+l)$ from uncertain observations $y(i)$, $1 \le i \le k$. Due to the presence of a multiplicative noise component in the observation Equation (1), even if the additive noise is Gaussian, the conditional expectation of $x(k+l)$ given $y(i)$, $1 \le i \le k$, which provides the least-squares estimator, is not a linear function of the observations and its computation can be very complicated requiring, in general, an exponentially growing memory. For this reason, our attention is focused on the least-squares linear estimation problem. Specifically, we are interested in obtaining the least-squares linear estimator of the state vector $x(k+l)$ based on the observations $y(i)$, $1 \le i \le k$. This estimator, $\hat{x}(k+l,k)$, is the orthogonal projection of $x(k+l)$ on the space of $n-$dimensional linear transformations of the observations. So, $\hat{x}(k+l,k)$ is given by

$$\hat{x}(k+l,k) = \sum_{i=1}^{k} h(k+l,i,k)y(i) \quad (5)$$

as a linear transformation of the observed values $y(i)$, $1 \le i \le k$, where $h(k+l,i,k)$, $1 \le i \le k$, denotes the

impulse-response function.

Let us consider the least-squares prediction problem, which minimizes the criterion

$$J = E\left[\left(x(k+l)-\hat{x}(k+l,k)\right)^T\left(x(k+l)-\hat{x}(k+l,k)\right)\right]. \tag{6}$$

The orthogonal projection lemma [7] assures that $x(k+l,k)$ is the only linear combination of the observations $y(i)$, $1 \le i \le k$ such that the estimation error is orthogonal to them,

$$x(k+l)-\hat{x}(k+l,k) \perp y(s), \quad 1 \le s \le k, \tag{7}$$

that is,

$$E\left[\left(x(k+l)-\sum_{i=1}^{k}h(k+l,i,k)y(i)\right)y^T(s)\right]=0,$$
$$1 \le s \le k.$$

This condition is equivalent to the Wiener–Hopf equation

$$E\left[x(k+l)y^T(s)\right]=\sum_{i=1}^{k}h(k+l,i,k)E\left[y(i)y^T(s)\right], \tag{8}$$
$$1 \le s \le k,$$

useful to determine the optimum impulse-response function $h(k+l,i,k)$, $1 \le i \le k$, which minimizes the cost function (6). From $P\{\gamma(k)=1\}=p(k)$, the left-hand side of (7) is written as

$$E\left[x(k+l)y^T(s)\right]=K(k+l,s)H^T p(s). \tag{9}$$

Let $E_\gamma[\cdot]$ denote the statistical expectation with respect to $\gamma(\cdot)$. Then, from the observation equation (1) and the covariance function (2) for white observation noise $v(k)$, $E\left[y(i)y^T(s)\right]$ is reduced to

$$E\left[y(i)y^T(s)\right]=E_\gamma\left[\gamma(i)\gamma(s)\right]HK(i,s)H^T$$
$$+R(i)\delta_K(i-s). \tag{10}$$

Substituting (9) and (10) into (8), we have

$$h(k+l,s,k)R(s)=K(k+l,s)H^T p(s)$$
$$-\sum_{i=1}^{k}h(k+l,i,k)E_\gamma\left[\gamma(i)\gamma(s)\right]HK(i,s)H^T. \tag{11}$$

Under these conditions, in Section 3 the RLS Wiener prediction and filtering algorithms are presented.

3. RLS Wiener Prediction and Filtering Algorithm

Nakamori *et al.* [5,6], based on the innovation approach, proposed the algorithms for the fixed-point smoothing estimate and the filtering estimate. These algorithms are derived taking into account that the innovation process is

expressed as

$$\upsilon(s)=y(s)-\hat{y}(s,s-1),$$
$$\hat{y}(s,s-1)=P_{2,2}(s)H\Phi\hat{x}(s-1,s-1).$$

Under the preliminary assumptions made in Section 2, Theorem 1 proposes the RLS Wiener algorithms for the $l-$step ahead prediction estimates of the signal $z(k+l)$ and the state vector $x(k+l)$. These algorithms are derived, starting with (11), by iterative use of the invariant imbedding method.

Theorem 1. Consider the observation equation described in (1) and assume that the probability $p(k)$ and the $(2,2)$ element $P_{2,2}(k)$ of the conditional probability matrix $P(k|j)$ are given. Let the system state-transition matrix Φ, the observation matrix H, the autovariance function $K(s,s)$ of the state vector $x(s)$, the variance $R(k)$ of the white observation noise $v(k)$ and the observed value $y(k)$ be given. Then the RLS Wiener algorithms for the $l-$step ahead prediction estimate $\hat{z}(k+l,k)$ of the signal $z(k+l)$ and the $l-$step ahead prediction estimate $\hat{x}(k+l,k)$ of the state vector $x(k+l)$ consist of (12)-(17).

$l-$step ahead prediction estimate of the signal $z(k+l):\hat{z}(k+l,k)$

$$\hat{z}(k+l,k)=H\hat{x}(k+l,k) \tag{12}$$

$l-$step ahead prediction estimate of the state vector $x(k+l):\hat{x}(k+l,k)$

$$\hat{x}(k+l,k)=\Phi\hat{x}(k+l-1,k-1)$$
$$+\Phi^l h(k,k,k)\left(y(k)-P_{2,2}(k)H\Phi\hat{x}(k-1,k-1)\right), \tag{13}$$
$$\hat{x}(l,0)=0$$

Filtering estimate of $z(k):\hat{z}(k,k)$

$$\hat{z}(k,k)=H\hat{x}(k,k) \tag{14}$$

Filtering estimate of $x(k):\hat{x}(k,k)$

$$\hat{x}(k,k)=\Phi\hat{x}(k-1,k-1)$$
$$+h(k,k,k)\left(y(k)-P_{2,2}(k)H\Phi\hat{x}(k-1,k-1)\right), \tag{15}$$
$$\hat{x}(0,0)=0$$

Filter gain: $h(k,k,k)$

$$h(k,k,k)=\left(K(k,k)H^T p(k)\right.$$
$$-P_{2,2}(k)\Phi S(k-1)\Phi^T H^T\Big)$$
$$\times\Big(R(k)+p(k)HK(k,k)H^T$$
$$\left.-P_{2,2}^2(k)H\Phi S(k-1)\Phi^T H^T\right)^{-1} \tag{16}$$

$$S(k) = \Phi S(k-1)\Phi^T + h(k,k,k)$$
$$\times \left(HK(k,k) p(k) - P_{2,2}(k) H\Phi S(k-1)\Phi^T \right), \quad (17)$$
$$S(0) = 0$$

Proof of Theorem 1 is detailed in the Appendix.

Clearly, the algorithms for the filtering estimate are the same as those proposed in Nakamori *et al.* [5]. From Theorem 1, the innovation process $\upsilon(k)$ is represented by

$$\upsilon(k) = y(k) - P_{2,2}(k) H\Phi \hat{x}(k-1,k-1). \quad (18)$$

4. A Numerical Simulation Example

In order to illustrate the application of the RLS Wiener prediction algorithm proposed in Theorem 1, we consider a scalar signal $z(k)$ whose autocovariance function $K_z(m)$ is given as follows [8]

$$K_z(0) = \sigma^2,$$
$$K_z(m) = \sigma^2 \left\{ \alpha_1 \left(\alpha_2^2 - 1 \right) \alpha_1^m / \left[(\alpha_2 - \alpha_1)(\alpha_2\alpha_1 + 1) \right] \right. $$
$$\left. - \alpha_2 \left(\alpha_1^2 - 1 \right) \alpha_2^m / \left[(\alpha_2 - \alpha_1)(\alpha_2\alpha_1 + 1) \right] \right\}, \quad (19)$$
$$0 < m,$$

with $\alpha_1, \alpha_2 = \left(-a_1 \pm \sqrt{a_1^2 - 4a_2} \right) / 2$,

where $a_1 = -0.1$, $a_2 = 0.8$ and $\sigma = 0.5$.

The covariance function (19) corresponds to a signal process generated by a second-order AR model. Therefore, according to Nakamori [4], the observation vector H, the variance $K(k,k) = K(0)$ of the state vector $x(k)$ and the system matrix Φ in the state equation are as follows:

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad K(k,k) = \begin{bmatrix} K_z(0) & K_z(1) \\ K_z(1) & K_z(0) \end{bmatrix},$$
$$\Phi = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix}, \quad K_z(0) = 0.25, \quad K_z(1) = 0.125. \quad (20)$$

As in Nakamori *et al.* [5], we consider that the signal $z(k)$ is transmitted through one of two channels, characterized by its observation equation as follows:

Channel 1: $y(k) = z(k) + v(k)$,
Channel 2: $y(k) = U(k)z(k) + v(k)$,

where $\{v(k)\}$ is a zero-mean white observation noise and $\{U(k)\}$ is a sequence of independent random variables taking values 0 or 1 with

$$P\{U(k) = 1\} = \breve{p} = 0.8, \text{ for all } k.$$

We assume that channel 1 is chosen at random with probability $1 - q = 0.7$ and, hence, channel 2 is selected with probability $q = 0.3$. Then, the observation equation is described by

$$y(k) = \gamma(k)z(k) + v(k), \quad (21)$$

where $\gamma(k) = (1-\alpha)1 + \alpha U(k)$ and α is a random variable, independent of $\{U(k)\}$, taking values 0 or 1 with $P\{\alpha = 1\} = q = 0.3$. Clearly, $\{\gamma(k)\}$ is a sequence of random variables which take values 0 or 1 with

$$p(k) = P\{\gamma(k) = 1\}$$
$$= P\{\alpha = 1, U(k) = 1\} + P\{\alpha = 0\}$$
$$= \breve{p}q + (1-q) = 0.94$$

for all k, and conditional probability matrix

$$P(k \mid j) = \begin{bmatrix} 1 - \breve{p} & \breve{p} \\ \dfrac{q\breve{p}(1-\breve{p})}{1 - q(1-\breve{p})} & \dfrac{1 - q(1-\breve{p}^2)}{1 - q(1-\breve{p})} \end{bmatrix}$$
$$= \begin{bmatrix} 0.2 & 0.8 \\ 0.0510638 & 0.9489362 \end{bmatrix},$$

for all $k, j = 0, \cdots, k-1$.

From (3), $\left[P(k \mid j) \right]_{2,2} = P_{2,2}(k) = 0.9489362$, for all $k, j = 0, \cdots, k-1$.

Substituting H, $K(k,k)$ and Φ, given by (20), into the prediction algorithm of Theorem 1, the prediction estimate of the signal has been calculated recursively.

Figure 1 illustrates the signal $z(k)$ and its prediction estimate $\hat{z}(k+3,k)$ for zero-mean white observation noise with variance 0.3^2. **Figure 2** illustrates the mean-square values (MSVs) of the filtering and prediction errors for zero-mean white observation noises with variances 0.1^2, 0.3^2, 0.5^2 and 0.7^2, comparing both the uncertain and certain observations cases (the latter cor-
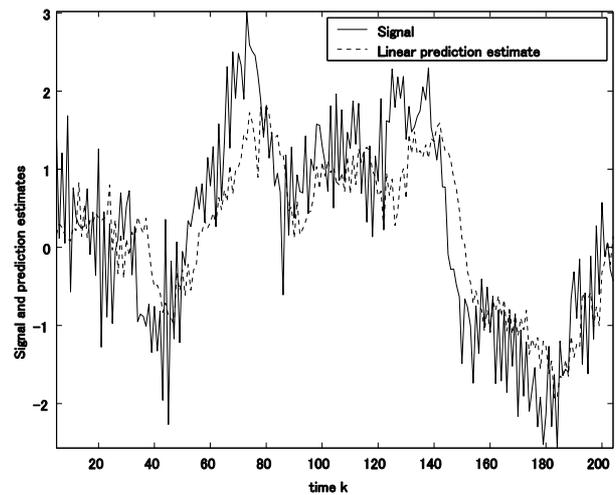
Figure 1. Signal $z(k)$ and its prediction estimate $\hat{z}(k+3,k)$ for the zero-mean white observation noise with the variance 0.3^2.

responds to the case $p(k) = P_{2,2}(k) = 1$). The MSVs of the filtering and prediction errors are evaluated by

$$\sum_{i=1}^{2000} \left(z(i) - \hat{z}(i+l,i) \right)^2 \Big/ 2000, \quad l = 1, 2, \cdots, 5.$$

Here, $l = 0$ corresponds to the calculation of the MSVs of the filtering errors.

From **Figure 2**, it is deduced that, as l becomes larger, the prediction accuracy worsens for both the uncertain and the certain observations cases, with each different observation noise. It might also be noticed that the MSVs with uncertain observations are almost equal to those with certain observations except for the observation noise with variance 0.1^2. For the observation noise with variance 0.1^2, the MSVs of the prediction errors with the certain observations are smaller than those with the uncertain observations, particularly for the 2 and 4-step ahead predictions.

For reference, the autoregressive (AR) model used to generate the signal process in the simulations, is given by

$$z(k+1) = -a_1 z(k) - a_2 z(k-1) + w(k+1),$$
$$E\left[w(k)w(s) \right] = \sigma^2 \delta_K (k-s). \tag{22}$$

5. Conclusions

Under the preliminary assumptions of Section 2, for the
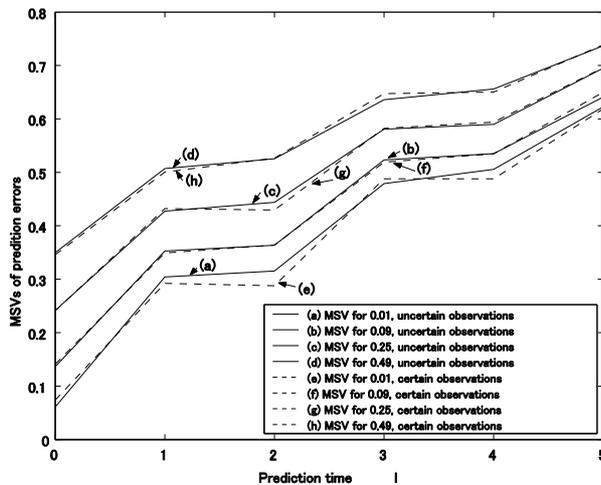
observation Equation (1), the RLS Wiener algorithms for the $l-$step ahead prediction estimates $\hat{z}(k+l,k)$ of the signal $z(k+l)$ and $\hat{x}(k+l,k)$ of the state vector $x(k+l)$ are derived by iterative use of the invariant imbedding method. The prediction algorithms take into account the stochastic properties of the random variables $\{\gamma(k)\}$ in the observation Equation (1) such as the probability $p(k) = P\{\gamma(k) = 1\}$ that the signal exists in the uncertain observation equation, and the $(2,2)$ element $\left[P(k|j) \right]_{2,2}$ of the conditional probability of $\gamma(k)$, given $\gamma(j)$.

A numerical simulation example in Section 4 shows that the prediction algorithm proposed in this paper is feasible.

REFERENCES

[1] H. L. Van Trees, "Detection, Estimation and Modulation Theory (Part I)," Wiley, New York, 1968.

[2] N. Nahi, "Optimal Recursive Estimation with Uncertain Observation," *IEEE Transactions on Information Theory*, Vol. IT-15, No. 4, 1969, pp. 457-462. doi:10.1109/TIT.1969.1054329

[3] N. Hadidi and S. Schwartz, "Linear Recursive State Estimators under Uncertain Observations," *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 6, 1979, pp. 944-948. doi:10.1109/TAC.1979.1102171

[4] S. Nakamori, "Estimation Technique Using Covariance Information in Linear Discrete-Time Systems," *Signal Processing*, Vol. 58, No. 3, 1997, pp. 309-317. doi:10.1016/S0165-1684(97)00032-7

[5] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo and J. Linares-Pérez, "Linear Recursive Discrete-Time Estimators Using Covariance Information under Uncertain Observations," *Signal Processing*, Vol. 83, No. 7, 2003, pp. 1553-1559. doi:10.1016/S0165-1684(03)00056-2

[6] S. Nakamori, R. Caballero-Águila, A. Hermoso-Carazo and J. Linares-Pérez, "Fixed-Point Smoothing with Non-Independent Uncertainty Using Covariance Information," *International Journal of Systems Science*, Vol. 34, No. 7, 2003, pp. 439-452. doi:10.1080/00207720310001636390

[7] A. P. Sage and J. L. Melsa, "Estimation Theory with Applications to Communications and Control," McGraw-Hill, New York, 1971.

[8] S. Haykin, "Adaptive Filter Theory," Prentice-Hall, New Jersey, 2003.

Figure 2. Mean-square values (MSVs) of the filtering and prediction errors for the zero-mean white observation noises with the variances 0.1^2, 0.3^2, 0.5^2 and 0.7^2 for both the uncertain and certain observations.

Appendix A. Proof of Theorem 1

Let us introduce the equation concerned with the function $J(k,s)$ as

$$J(k,s)R(s) = p(s)B^T(s)H^T$$
$$-\sum_{i=1}^{k} J(k,i)E_\gamma[\gamma(i)\gamma(s)]HK(i,s)H^T. \quad \text{(A-1)}$$

From (11) and (A-1) it follows that

$$h(k+l,s,k) = A(k+l)J(k,s). \quad \text{(A-2)}$$

Subtracting the equation obtained by putting $k \to k-1$ in (A-1) from (A-1) yields

$$(J(k,s)-J(k-1,s))R(s)$$
$$=-J(k,k)E_\gamma[\gamma(k)\gamma(s)]HK(k,s)H^T$$
$$-\sum_{i=1}^{k-1}(J(k,i)-J(k-1,i)) \quad \text{(A-3)}$$
$$\times E_\gamma[\gamma(i)\gamma(s)]HK(i,s)H^T.$$

From (A-1), (A-3) and the relationship $E[\gamma(k)\gamma(s)] = P_{2,2}(k)p(s)$, it follows that

$$J(k,s)-J(k-1,s)$$
$$=-J(k,k)P_{2,2}(k)HA(k)J(k-1,s), \quad \text{(A-4)}$$
$$1 \le s < k.$$

Putting $s=k$ in (A-1) yields

$$J(k,k)R(k) = p(k)B^T(k)H^T$$
$$-\sum_{i=1}^{k} J(k,i)E_\gamma[\gamma(i)\gamma(k)]HK(i,k)H^T$$
$$= p(k)B^T(k)H^T$$
$$-J(k,k)E_\gamma[\gamma(k)\gamma(k)]HK(k,k)H^T$$
$$-\sum_{i=1}^{k-1} J(k,i)E_\gamma[\gamma(i)\gamma(k)]HK(i,k)H^T \quad \text{(A-5)}$$
$$= p(k)B^T(k)H^T - p(k)J(k,k)HK(k,k)H^T$$
$$-\sum_{i=1}^{k-1}(J(k-1,i)-J(k,k)P_{2,2}(k)HA(k)J(k-1,i))$$
$$\times p(i)P_{2,2}(k)HB(i)A^T(k)H^T.$$

Here, the relationship $E_\gamma[\gamma(k)\gamma(k)] = p(k)$, (4) and (A-4) are used. Let us introduce a function

$$r(k) = \sum_{i=1}^{k} J(k,i)p(i)HB(i). \quad \text{(A-6)}$$

Hence,

$$J(k,k) =$$
$$\left(p(k)B^T(k)H^T - r(k-1)A^T(k)H^T P_{2,2}(k)\right)$$
$$\left(R(k)+p(k)HK(k,k)H^T\right. \quad \text{(A-7)}$$
$$\left.-P_{2,2}^2(k)HA(k)r(k-1)A^T(k)\right)^{-1}.$$

Subtracting the equation obtained by putting $k \to k-1$ in (A-6) from (A-6) yields

$$r(k)-r(k-1) = J(k,k)p(k)HB(k)$$
$$+\sum_{i=1}^{k-1}(J(k,i)-J(k-1,i))p(i)HB(i)$$
$$= J(k,k)p(k)HB(k) \quad \text{(A-8)}$$
$$-J(k,k)P_{2,2}(k)HA(k)\sum_{i=1}^{k-1} J(k-1,i)p(i)HB(i)$$
$$= J(k,k)(p(k)HB(k)-P_{2,2}(k)HA(k)r(k-1)).$$

Here, (A-5) and (A-7) have been used. Clearly, from (A-6), the initial condition for the recursive Equation (A-8) of $r(k)$ at $k=0$ is given by $r(0)=0$.

The $l-$ step ahead prediction estimate $\hat{x}(k+l,k)$ of $x(k+l)$ is given by (5). From (5) and (A-2), it follows that

$$\hat{x}(k+l,k) = A(k+l)\sum_{i=1}^{k} J(k,i)y(i). \quad \text{(A-9)}$$

Let us introduce a function

$$e(k) = \sum_{i=1}^{k} J(k,i)y(i). \quad \text{(A-10)}$$

Hence, the $l-$ step ahead prediction estimate $\hat{x}(k+l,k)$ of the state vector $x(k+l)$ and the filtering estimate $\hat{x}(k,k)$ of the state vector $x(k)$ are given by

$$\hat{x}(k+l,k) = A(k+l)e(k), \hat{x}(k,k) = A(k)e(k). \quad \text{(A-11)}$$

Subtracting the equation obtained by putting $k \to k-1$ in (A-10) from (A-10) yields

$$e(k)-e(k-1)$$
$$= J(k,k)y(k)+\sum_{i=1}^{k-1}(J(k,i)-J(k-1,i))y(i). \quad \text{(A-12)}$$

From (A-4) and (A-10), we get

$$e(k)-e(k-1) = J(k,k)y(k)$$
$$-J(k,k)P_{2,2}(k)HA(k)\sum_{i=1}^{k-1} J(k-1,i)y(i) \quad \text{(A-13)}$$
$$= J(k,k)(y(k)-P_{2,2}((k)HA(k)e(k-1)).$$

From (A-2), (A-11) and (A-13), it follows that

　　JSIP

$$\hat{x}(k,k) = \Phi \hat{x}(k-1,k-1)$$
$$+ h(k,k,k)\big(y(k) - P_{2,2}(k)H\Phi \hat{x}(k-1,k-1)\big),$$
$$\hat{x}(0,0) = 0,$$
$$\hat{x}(k+l,k) = \Phi \hat{x}(k+l-1,k-1)$$
$$+ \Phi^{l} h(k,k,k)\big(y(k) - P_{2,2}(k)H\Phi \hat{x}(k-1,k-1)\big),$$
$$\hat{x}(l,0) = 0. \tag{A-14}$$

Introducing

$$S(k) = A(k) r(k) A^{T}(k), \tag{A-15}$$

it follows, from (A-8) and (A-15), that

$$S(k) = A(k) r(k-1) A^{T}(k) + A(k) J(k,k)$$
$$\times \big(p(k) HB(k) - P_{2,2}(k)HA(k)r(k-1)\big) A^{T}(k)$$
$$= \Phi S(k-1)\Phi^{T} + h(k,k,k)$$
$$\times \big(p(k)HK(k,k) - P_{2,2}(k)H\Phi S(k-1)\Phi^{T}\big)$$
$$S(0) = 0.$$

Finally, the filter gain $h(k,k,k) = A(k) J(k,k)$ is expressed as follows

$$h(k,k,k) = \big(p(k)K(k,k)H^{T} - P_{2,2}(k)\Phi S(k-1)\Phi^{T}H^{T}\big)$$
$$\times \big(R(k) + p(k)HK(k,k)H^{T} - P_{2,2}^{2}(k)H\Phi S(k-1)\Phi^{T}\big)^{-1}.$$

(Q.E.D.)