Bias Compensation of Recursive Least Squares Estimate in Closed Loop Environment

K. Ikeda*, Y. Mogami*, and T. Shimomura*

Abstract: In this paper, an asymptotic bias of the recursive least squares (RLS) estimate in the closed loop environment is analyzed and its compensation method is proposed under the assumption that the noise is white. Namely, a bias compensated RLS method in the closed loop environment based on output error (OE) model is proposed. A posteriori error is also analyzed for the estimation of the noise variance.

Keywords: Recursive Least Squares Estimate, Bias Compensation, Closed Loop Identification.

1 Introduction

There are increasing demands for the identification under the feedback control[8, 5, 7] due to the safety or economic reasons[1]. The difficulty of the closed loop identification arises from the fact that the input and the output of the plant correlates with the noise because of the feedback loop. In order to remove the asymptotic bias caused by the correlations, a special treatment will be required for the closed loop identification.

In order to obtain an unbiased estimate in the closed loop environment, the instrumental variable (IV) methods [2, 3, 4] are proposed, which require a few iterations to obtain unbiased estimate. These methods are based on the indirect approach of the closed loop identification or at least the instrumental variables are produced based on the reference input signal. Thus, the estimation error depends on the informativeness of the reference input. For less informative reference input, direct approach of the closed loop identification will be required.

*The University of Tokushima
Unbiased estimate is also obtained by using the bias compensated least squares (BCLS) method\cite[9, 10]{9, 10}, which is based on the analysis of the noise effect on the LS estimate and on the estimation of the noise variance. Therefore, BCLS will be applicable for the direct approach of the closed-loop identification.

One of the authors proposes an iterative BCLS method for the estimation of the plant parameters in the closed loop environment\cite[6]{6}, in which prefileters are iteratively redesigned in order to make the noise white. When the length of the I/O data is very large or estimation is done on-line, recursive least squares (RLS) method will be preferable. In this paper, the asymptotic bias of the RLS estimate in closed loop environment is analyzed and its compensation method is proposed. In order to estimate the noise variance, a \textit{posteriori} error is analyzed.

The paper is organized as follows. The problem is formulated in section II, and RLS method is briefly summarized in section III. Section IV analyzes the asymptotic bias of the RLS estimate. In section V, the noise variance is analyzed and the bias compensated recursive least square method is proposed. Section VI shows a numerical example for the illustration of the proposed method. Section VII concludes the paper.

\textbf{Notation:}
Let \( E\{x\} \) denote an expectation of random variable \( x \).
Let \( q \) denote a shift operator \( i.e. \ x_{k+1} = qx_k \).

\section{Problem Formulation}
Consider a single-input single-output (SISO) \textit{n}-th order discrete time plant:

\begin{equation}
y_k = \frac{b_p(q)}{a_p(q)} u_k + \nu_k,
\end{equation}

where \( u_k \in \mathbb{R}, y_k \in \mathbb{R}, \) and \( \nu_k \in \mathbb{R} \) are the input, the output, and the observation noise, respectively. The polynomial \( a_p(q) = q^n + a_1q^{n-1} + \cdots + a_n \) is monic and of \( n \)-th order, while \( b_p(q) = b_1q^{n-1} + \cdots + b_n \) has order at most \( n - 1 \).

The plant to be estimated is assumed to be controlled by the following feedback compensator:

\begin{equation}
u_k = \frac{b_c(q)}{a_c(q)} (r_k - y_k),\end{equation}

where \( r_k \in \mathbb{R} \) is a reference input, \( a_c(q) = q^m + a_1 q^{m-1} + \cdots + a_m \), and \( b_c(q) = b_c 0 + b_{c1} q^{m-1} + \cdots + b_{cm} \).

The following assumptions are made:

\begin{enumerate}
\item[(A1)] \( a_p(q) \) and \( b_p(q) \) do not have a common zero outside of the open unit disc.
\item[(A2)] an upper bound of the plant degree is known to be \( n \).
\item[(A3)] the observation noise \( \{\nu_k\} \) is a zero mean white noise with variance \( E\{\nu_k \nu_l\} = \sigma_n^2 \delta_{kl} \).
\end{enumerate}

where \( \delta_{kl} \) denotes a Kronecker delta.
(A4) the I/O data is collected in the closed loop environment and the feedback loop is asymptotically stable.

(A5) the reference input \( r_k \) is independent of the observation noise \( \nu_k \).

From the assumption (A3), the persistently excitation (PE) condition will be satisfied even if \( r_k = 0 \) because the closed loop is driven by a white noise.

Let the characteristic polynomial of the closed loop be denoted by
\[
d_{cl}(q) = a_p(q)a_c(q) + b_p(q)b_c(q),
\]
all the zeros of \( d_{cl}(q) \) lies in the open unit disc from the assumption (A4).

**Problem:** Estimate the unknown coefficient of \( a_p(q) \) and \( b_p(q) \) from the I/O data \( \{u_k, y_k\} \) \( k = 1, \ldots, N \).

### 3 Recursive Least Squares Estimate

In this section, recursive least squares estimate together with prefilterers are briefly summarized.

Define the characteristic polynomial of the prefilter as
\[
f(q) = q^n + f_1q^{n-1} + \cdots + f_n,
\]
where all the zeros of \( f(q) \) are selected to lie in the open unit disc. Define the filtered output \( y_{f,k} \) and the filtered input \( u_{f,k} \) as follows:
\[
y_{f,k} = \frac{q^n}{f(q)}y_k \quad \text{and} \quad u_{f,k} = \frac{q^n}{f(q)}u_k.
\]

Multiplying the both side of eq. (1) by \( a_p(q)/f(q) \), we obtain
\[
y_k = \frac{f(q) - a_p(q)}{f(q)}y_k + \frac{b_p(q)}{f(q)}u_k + \frac{a_p(q)}{f(q)}\nu_k.
\]

From the equation above, the following linear regression formula is obtained:
\[
y_k = \varphi_k^\top \theta + \varepsilon_k,
\]
where
\[
\theta = [\theta_1^\top, \theta_2^\top]^\top = [f_1 - a_1, \ldots, f_n - a_n, b_1, \ldots, b_n]^\top,
\]
\[
\varphi_k = [y_{f,k-1}, \ldots, y_{f,k-n}, u_{f,k-1}, \ldots, u_{f,k-n}]^\top,
\]
\[
\varepsilon_k = \frac{a_p(q)}{f(q)}\nu_k.
\]

Adopting the weighted least squares criterion:
\[
J_k(\hat{\theta}) = \sum_{i=1}^{k} \lambda^{k-i} \left[ y_i - \hat{\theta}^\top \varphi_i \right]^2
\]
where $1 \geq \lambda > 0$ is a design parameter, the parameter minimizing $J_k(\hat{\theta})$ can be given recursively by

$$
\hat{\theta}_k = \hat{\theta}_{k-1} + L_k e_k, \quad (12)
$$

$$
L_k = \frac{\Gamma_{k-1} \varphi_k}{\lambda + \varphi_k^\top \Gamma_{k-1} \varphi_k}, \quad (13)
$$

$$
e_k = y_k - \varphi_k^\top \hat{\theta}_{k-1}, \quad (14)
$$

$$
\Gamma_k = \frac{1}{\lambda} \left[ \Gamma_{k-1} - \frac{\Gamma_{k-1} \varphi_k \varphi_k^\top \Gamma_{k-1}}{\lambda + \varphi_k^\top \Gamma_{k-1} \varphi_k} \right], \quad (15)
$$

with initial values $\hat{\theta}_0$ and $\Gamma_0 = \Gamma_0^\top > 0$. It is well known that eq. (15) is obtained by applying the matrix inversion lemma recursively to the r.h.s of the following equation:

$$
\Gamma_k = \left( \lambda^k \Gamma_0^{-1} + \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^\top \right)^{-1} \quad (16)
$$

Define $\tilde{\theta}_k = \hat{\theta}_k - \theta$ and rewrite eq. (12) for $\tilde{\theta}_k$ by using eq. (7) and the relations $L_k = \Gamma_k \varphi_k$ and $\lambda \Gamma_k \Gamma_{k-1}^{-1} = I - L_k \varphi_k^\top$ which are obtained from eq. (15):

$$
\tilde{\theta}_k = \lambda \Gamma_k \Gamma_{k-1}^{-1} \tilde{\theta}_{k-1} + \Gamma_k \varphi_k e_k. \quad (17)
$$

Applying eq. (17) recursively, we finally obtain

$$
\hat{\theta}_k = \theta + \lambda^k \Gamma_k \Gamma_0^{-1} \tilde{\theta}_0 + \Gamma_k \sum_{i=1}^{k} \lambda^{k-i} \varphi_i e_i. \quad (18)
$$

Because the equation error $e_i$ is not white and has correlation with the regression vector $\varphi_i$ in general, the least squares estimator $\hat{\theta}_k$ has an asymptotic bias.

## 4 Asymptotic bias in closed loop environment

In this section, the asymptotic bias of the recursive least squares estimator (18) in the closed loop environment is investigated. Because $\{r_k\}$ and $\{\nu_k\}$ are independent from the assumption (A5), the expectation of the correlation between the equation error and the $r$-dependent part of each regressor becomes zero. Thus, we assume $r_k = 0$ without loss of generality in this section.

The noise dependent parts of the filtered output $y_{f,k-i}$ and the filtered input $u_{f,k-i}$ are given by

$$
y_{f,k-i} = \frac{q^{n-i}a_c(q)}{d_c(q)} \cdot \frac{a_p(q)}{f(q)} \nu_k = \frac{q^{n-i}a_c(q)}{d_c(q)} \nu_k, \quad (19)
$$

$$
u_{f,k-i} = -\frac{q^{n-i}b_c(q)}{d_c(q)} \cdot \frac{a_p(q)}{f(q)} \nu_k = -\frac{q^{n-i}b_c(q)}{d_c(q)} \nu_k. \quad (20)
$$
These two equations together with eq. (10) have a state space representation as follows:

\[
\begin{align*}
\bar{X}_{k+1} &= \bar{A}\bar{X}_k + \bar{b}\nu_k, \\
\varphi_k &= [C_{cl} \ 0_{2n \times n}] \bar{X}_k, \\
\varepsilon_k &= [0_{1 \times (n+m)} \ h^\top] \bar{X}_k + \nu_k,
\end{align*}
\]  

(21)

(22)

(23)

where \( \bar{A} \) and \( \bar{b} \) are defined by

\[
\bar{A} = \begin{bmatrix} A_{cl} & b_{cl} h^\top \\ 0_{n \times (n+m)} & F \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} b_{cl} \\ g \end{bmatrix},
\]

(24)

and \((A_{cl}, b_{cl}, C_{cl})\) and \((F, g, h^\top, 1)\) are the system matrices of the state space representations of the transfer functions:

\[
\begin{bmatrix} A_{cl} & b_{cl} h^\top \\ C_{cl} & 0_{2n \times 1} \end{bmatrix} = \begin{bmatrix} s(q)a_c(q) \\ -s(q)b_c(q) \end{bmatrix}, \quad s(q) = [q^{n-1}, \ldots, q, 1]^\top,
\]

(25)

\[
\begin{bmatrix} F \\ h^\top \\ g \\ 1 \end{bmatrix} = \begin{bmatrix} a_p(q) \\ f(q) \end{bmatrix}.
\]

(26)

From eqs. (21), (22), and (23), and taking into account that \( E\{\bar{X}_k\nu_k\} = 0, E\{\varphi_k\varepsilon_k\} \) can be calculated as

\[
E\{\varphi_k\varepsilon_k\} = [C_{cl} \ 0] E\{\bar{X}_k\bar{X}_k^\top\} \begin{bmatrix} 0 \\ h \end{bmatrix}.
\]

(27)

Covariance matrix of \( \bar{X}_k \) is given by

\[
E\{\bar{X}_k\bar{X}_k^\top\} = P\sigma^2_
u
\]

(28)

where \( P = P^\top > 0 \) is a solution of the Lyapunov equation:

\[
P = \bar{A}P\bar{A}^\top + \bar{b}\bar{b}^\top.
\]

(29)

Finally, we obtain

\[
E\{\varphi_k\varepsilon_k\} = C_{cl} P_{12} h^\top \sigma^2_
u.
\]

(30)

where \( P_{12} \in \mathbb{R}^{(n+m) \times n} \) is a 1-2 block of \( P = P^\top > 0 \).

The asymptotic bias of the recursive least squares estimator in the closed loop environment is given by the following theorem.

**Theorem 1.** Consider the closed loop defined by eqs. (1) and (2) together with the assumptions (A1) to (A5). Assume that the correlation between \( \Gamma_k \) and \( \varphi_i\varepsilon_i \) \((i \leq k)\) is negligible. Then the expectation of the recursive least squares estimate defined by eq. (12) is given by

\[
E\{\hat{\theta}_k\} = \begin{cases} 
\theta + \lambda^k \Gamma_k \Gamma_0^{-1} \hat{\theta}_0 + \frac{1-\lambda^k}{\lambda} \Gamma_k C_{cl} P_{12} h \sigma^2_
u & \text{for } 0 < \lambda < 1, \\
\theta + \Gamma_k \Gamma_0^{-1} \hat{\theta}_0 + \lambda \Gamma_k C_{cl} P_{12} h \sigma^2_
u & \text{for } \lambda = 1,
\end{cases}
\]

(31)
where $P_{12}$ is a 1-2 block of $P$ defined by eqs. (24) to (29).

**Proof:** It is obvious from the discussions above. The matrix $P_{12}$ can be obtained by solving the following Sylvester equation instead of solving the Lyapunov equation (29):

$$P_{12} = A_d P_{12} F^\top + b_d h^\top P_{22} F^\top + b_d g^\top$$

(32)

where $P_{22}$ is a solution of the Lyapunov equation

$$P_{22} = F P_{22} F^\top + gg^\top,$$

(33)

which is independent of the plant parameters when $(F, g)$ is realized as a controller canonical form.

### 5 Estimation of the Noise Variance

Based on the similar idea of [9, 10], the noise variance $\sigma^2_\nu$ is to be estimated from the a posteriori error or the least squares residual.

Define $Q_k$ as

$$Q_k = \sum_{i=1}^k \lambda^{k-i} \hat{e}_i^2/k, \quad (34)$$

where $\hat{e}_{i/k}$ is the a posteriori error defined by $\hat{e}_{i/k} = y_i - \varphi_i^\top \hat{\theta}_k$. Also define $\hat{Q}_k$ recursively as

$$\hat{Q}_k = \lambda \hat{Q}_{k-1} + \frac{\lambda \hat{e}_k^2}{\lambda + \hat{\varphi}_k \Gamma_k^{-1} \hat{\varphi}_k}, \quad (35)$$

with the initial condition $\hat{Q}_0 = 0$. Then, the following relation is obtained:

$$\hat{Q}_k = Q_k + \lambda^k (\hat{\theta}_k - \hat{\theta}_0)^\top \Gamma_0^{-1} (\hat{\theta}_k - \hat{\theta}_0). \quad (36)$$

(See Appendix A for the derivation of eq. (36).)

By using the relation $\hat{e}_{i/k} = \varepsilon_i - \varphi_i^\top \hat{\theta}_k$, $Q_k$ becomes

$$Q_k = \sum_{i=1}^k \lambda^{k-i} \varepsilon_i^2 - 2\hat{\theta}_k^\top \sum_{i=1}^k \lambda^{k-i} \varphi_i \varepsilon_i + \hat{\theta}_k^\top \left( \sum_{i=1}^k \lambda^{k-i} \varphi_i \varphi_i^\top \right) \hat{\theta}_k. \quad (37)$$

By using eq. (16), we obtain

$$Q_k = \sum_{i=1}^k \lambda^{k-i} \varepsilon_i^2 - 2\hat{\theta}_k^\top \sum_{i=1}^k \lambda^{k-i} \varphi_i \varepsilon_i + \hat{\theta}_k^\top (\Gamma_k^{-1} - \lambda^k \Gamma_0^{-1}) \hat{\theta}_k, \quad (38)$$

where $\beta_k = \sum_{i=1}^k \lambda^{k-i} \varphi_i \varepsilon_i$. Define $\alpha_k = \lambda^k \Gamma_0^{-1} \hat{\theta}_0$. Then eq. (18) is rewritten as

$$\hat{\theta}_k = \Gamma_k (\alpha_k + \beta_k). \quad (39)$$
From this and eqs. (36) and (38), we obtain

$$\hat{Q}_k = \sum_{i=1}^{k} \lambda^{k-i} \varepsilon_i^2 - \beta_k^\top \Gamma_k \beta_k - \alpha_k^\top (\Gamma_k \alpha_k + 2 \Gamma_k \beta_k - \tilde{\theta}_0).$$

(40)

Expectation $E\{\varphi_i \varepsilon_i\}$ is already analyzed in the previous section while $E\{\varepsilon_i^2\}$ can be calculated similarly as

$$E\{\varepsilon_i^2\} = (1 + h^\top P_{22} h) \sigma^2_{\varphi}.$$  

(41)

where $P_{22}$ is a 2-2 block of $P$ defined by eq. (29), or, equivalently is a solution of the Lyapunov equation (33).

Thus, the following theorem is obtained.

**Theorem 2.** Under the same assumptions as in Theorem 1, the expectation of $\hat{Q}_k$ defined in (35) is given by

$$E\{\hat{Q}_k\} = \frac{1 - \lambda^k}{1 - \lambda} (1 + h^\top P_{22} h) \sigma^2_{\varphi} - \left(\frac{1 - \lambda^k}{1 - \lambda}\right)^2 h^\top P_{12} C_{cl}^\top \Gamma_k C_{cl} P_{12} h \sigma^4_{\varphi}$$

$$- \alpha_k^\top \left[ \Gamma_k \alpha_k + 2 \left(\frac{1 - \lambda^k}{1 - \lambda}\right) \Gamma_k C_{cl} P_{12} h \sigma^2_{\varphi} - \tilde{\theta}_0 \right]$$

(42)

for $0 < \lambda < 1$ and

$$E\{\hat{Q}_k\} = k(1 + h^\top P_{22} h) \sigma^2_{\varphi} - k^2 h^\top P_{12} C_{cl}^\top \Gamma_k C_{cl} P_{12} h \sigma^4_{\varphi}$$

$$- \alpha_k^\top \left[ \Gamma_k \alpha_k + 2 k \Gamma_k C_{cl} P_{12} h \sigma^2_{\varphi} - \tilde{\theta}_0 \right]$$

(43)

for $\lambda = 1$.

**Proof:** It is obvious from the discussions above.

The noise variance will be estimated by solving the second order equation (42) or (43) for $\sigma^2_{\varphi}$. In many cases, $\alpha_k$ is negligible because $\Gamma_0$ is usually selected as $\Gamma_0 = \gamma I$, in which the design parameter $\gamma$ is set to be a large number, say $\gamma = 10^2 \sim 10^5$. Furthermore, $\alpha_k$ goes to 0 as $k \to \infty$ when $\lambda < 1$. As a result, an estimate of the noise variance will be defined as

$$\hat{\sigma}^2_{\varphi} = \begin{cases} \left(\frac{1 - \lambda^k}{1 - \lambda}\right) R_k & \text{for } 0 < \lambda < 1, \\ \frac{1}{k} R_k & \text{for } \lambda = 1, \end{cases}$$

(44)

where

$$R_k = \frac{2\hat{Q}_k / (1 + h^\top P_{22} h)}{1 + \sqrt{1 - \frac{4\hat{Q}_k h^\top P_{12} C_{cl}^\top \Gamma_k C_{cl} P_{12} h}{(1 + h^\top P_{22} h)^2}}}.$$  

(45)

However, $A_{cl}$ and $h$ are composed of the true value of the plant parameters. Thus, $A_{cl}$, $h$ together with $P_{12}$ and $R_k$ in eqs. (32) and (45) must be replaced by their
estimates, which will be denoted by $\hat{A}_{cl,k}$, $\hat{h}_k$, $\hat{P}_{12,k}$ and $\hat{R}_k$. These terms will be recursively estimated by using the bias compensated RLS estimate, which will be denoted by $\hat{\theta}_{BC,k}$. However, $\hat{\theta}_{BC,k}$ will not be required at each time step but at every $K$ steps. Thus, $\hat{A}_{cl,k}$ and $\hat{h}_k$ are to be defined by using $\hat{\theta}_{BC,k-K}$.

Summarizing above, the bias compensated RLS estimate will be defined as

$$\hat{\theta}_{BC,k} = \hat{\theta}_k - \Gamma_k C_{cl} \hat{P}_{12,k} \hat{h}_k \hat{R}_k,$$  \hspace{1cm} (46)

where $\hat{\theta}_k$, $\Gamma_k$, and $\hat{Q}_k$ are defined in eqs. (12) to (15) and (35), $\hat{P}_{12,k}$ is a solution of the Sylvester equation

$$\hat{P}_{12,k} = \hat{A}_{cl,k} \hat{P}_{12,k} F^\top + b_{cl} \hat{h}_k^\top P_{22} F^\top + b_{cl} g^\top,$$  \hspace{1cm} (47)

$\hat{R}_k$ is defined

$$\hat{R}_k = \frac{2 \hat{Q}_k / (1 + \hat{h}_k^\top P_{22} \hat{h}_k)}{1 + \sqrt{1 - \frac{4 \hat{Q}_k \hat{h}_k^\top \hat{P}_{12,k} C_{cl} \Gamma_k C_{cl} \hat{P}_{12,k} \hat{h}_k}{(1 + \hat{h}_k^\top P_{22} \hat{h}_k)^2}}},$$  \hspace{1cm} (48)

and $\hat{A}_{cl,k}$ and $\hat{h}_k$ are defined by using $\hat{\theta}_{BC,k-K}$.

Consider the following first order plant:

$$y_k = \frac{b}{q-a} u_k + \nu_k = \frac{0.5}{q-0.8} u_k + \nu_k,$$

where $\nu_k$ is a white gaussian noise with variance 1. The control input $u_k$ is defined by the following feedback compensator:

$$u_k = \frac{0.2289}{q-0.09004} (r_k - y_k).$$  \hspace{1cm} (49)

The reference input is a constant $r_k = 1$, which is not informative enough for large $k$. The number of the samples is $N = 65536$. The initial design parameter of the prefilter is $f^{(0)} = 0.76$. The forgetting factor is $\lambda = 0.9998$ while $\hat{\theta}_{BC,k}$ is updated at every 5000 samples ($K = 5000$). The initial value of the design parameter $\Gamma_0$ is $10^4 I$. One hundred pairs of I/O data are prepared for the estimation of the parameters.

The proposed method is compared with the RLS estimate. The proposed estimates are plotted in fig.1, while the RLS estimates are plotted in fig.2. The ellipses calculated by the estimated covariance matrices are drawn by solid lines in both figures. In fig.1, the ellipse of the RLS estimate is plotted by a dashed line. The bias of the RLS estimate is successfully eliminated by the proposed method.

In order to see the effect of the forgetting factor $\lambda$, the bias compensated RLS estimate when $\lambda = 0.9999$ and $\lambda = 1.0$ are plotted in figs. 3 and 4, respectively. It is seen that the variance of the estimate becomes smaller as $\lambda$ goes to 1.
6 Conclusion

The asymptotic bias of the RLS estimate in closed loop environment is analyzed and its compensation method is proposed. For the purpose of on-line monitoring of the plant, the forgetting factor is introduced in the RLS estimate. Numerical example shows that the bias of the RLS estimate is successfully reduced even if the reference input is not informative enough.
Bibliography


Appendix A

In this appendix, eq. (36) is to be derived in the following 3 steps.

(i) Define $X_k$ as

$$X_k = \sum_{i=1}^{k} \lambda^{k-i} \phi_i \hat{e}_{i/k}. \quad (50)$$

Then, the following equation holds:

$$X_k = \lambda^k \Gamma_0^{-1} (\tilde{\theta}_k - \tilde{\theta}_0). \quad (51)$$

Proof: Note that $\hat{e}_{i/k} = \hat{e}_{i/k-1} - \varphi_i^\top (\tilde{\theta}_k - \tilde{\theta}_{k-1})$ and $e_k = \hat{e}_{k/k-1}$. Applying these equations for (50), the following equation is obtained

$$X_k = \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \hat{e}_{i/k-1} - \left( \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^\top \right) (\tilde{\theta}_k - \tilde{\theta}_{k-1}),$$

$$= \lambda X_{k-1} + \varphi_k e_k - \left( \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^\top \right) (\tilde{\theta}_k - \tilde{\theta}_{k-1}). \quad (52)$$

By using eq. (16),

$$X_k = \lambda X_{k-1} + \lambda^k \Gamma_0^{-1} (\tilde{\theta}_k - \tilde{\theta}_{k-1}) + \Gamma_k^{-1} \left[ \Gamma_k \varphi_k e_k - (\tilde{\theta}_k - \tilde{\theta}_{k-1}) \right]. \quad (53)$$

The third term of the r.h.s. of the equation above is zero from the definition of $\tilde{\theta}_k$. Thus, the following recursive form of $X_k$ is obtained

$$X_k = \lambda X_{k-1} + \lambda^k \Gamma_0^{-1} (\tilde{\theta}_k - \tilde{\theta}_{k-1}). \quad (54)$$

Applying this equation recursively, we obtain

$$X_k = \lambda^k X_0 + \lambda^k \Gamma_0^{-1} (\tilde{\theta}_k - \tilde{\theta}_0). \quad (55)$$

The initial condition $X_0 = 0$ is consistent with the definition of $X_1$ and the recursive form of $X_k$. This proves eq. (51).

(ii) In the second step, a recursive form of $Q_k$ is to be derived. Recall $\hat{e}_{i/k} = \varepsilon_i - \varphi_i^\top \tilde{\theta}_k$, $Q_k$ is calculated as follows:

$$Q_k = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k}^2 = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} (\varepsilon_i - \varphi_i^\top \tilde{\theta}_k) = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} \varepsilon_i - X_k^\top \tilde{\theta}_k \quad (56)$$

Applying $\hat{e}_{i/k} = \hat{e}_{i/k-1} - \varphi_i^\top (\tilde{\theta}_k - \tilde{\theta}_{k-1})$, we obtain

$$Q_k = \hat{e}_{k/k} \varepsilon_k - \sum_{i=1}^{k-1} \lambda^{k-i} \varepsilon_i \varphi_i^\top (\tilde{\theta}_k - \tilde{\theta}_{k-1}) + \lambda Q_{k-1} + \lambda X_{k-1}^\top \tilde{\theta}_{k-1} - X_k^\top \tilde{\theta}_k \quad (57)$$
Applying (18) for $\tilde{\theta}_{k-1}$ together with equations $\hat{e}_{k/k} = \frac{\lambda e_k}{\lambda + \varphi_k^T \Gamma_k^{-1} \varphi_k}$ and $e_k = \varepsilon_k + \varphi_k^T \tilde{\theta}_{k-1}$, the first term of the r.h.s. of the equation above is calculated as

$\hat{e}_{k/k} \varepsilon_k = \frac{\lambda \varepsilon_k^2}{\lambda + \varphi_k^T \Gamma_k^{-1} \varphi_k} + \lambda^k (\tilde{\theta}_k - \tilde{\theta}_{k-1})^T \Gamma_k^{-1} \tilde{\theta}_0 + (\tilde{\theta}_k - \tilde{\theta}_{k-1})^T \sum_{i=1}^{k-1} \lambda^{k-i} \varphi_i \varepsilon_i.$

Finally, we obtain the recursive form of $Q_k$ as

$Q_k = \lambda Q_{k-1} + \frac{\lambda e_k^2}{\lambda + \varphi_k^T \Gamma_k^{-1} \varphi_k} + \lambda X_k^T \tilde{\theta}_k - \lambda \tilde{\theta}_0 X_0^T \Gamma_0^{-1} \tilde{\theta}_0.$

The initial condition $Q_0 = 0$ is consistent with the definition of $Q_1$ and this recursive equation.

(iii) Define $\tilde{Q}_k = \hat{Q}_k - Q_k$, then $\tilde{Q}_k$ obeys the following equation:

$\tilde{Q}_k = \lambda \tilde{Q}_{k-1} + X_k^T \tilde{\theta}_k - \lambda X_k^T \tilde{\theta}_{k-1} - \lambda^k (\tilde{\theta}_k - \tilde{\theta}_{k-1})^T \Gamma_0^{-1} \tilde{\theta}_0.$

Applying this equation recursively, we obtain

$\tilde{Q}_k = \lambda^k \tilde{Q}_0 + X_k^T \tilde{\theta}_k - \lambda X_k^T \tilde{\theta}_0 - \lambda^k (\tilde{\theta}_k - \tilde{\theta}_0)^T \Gamma_0^{-1} \tilde{\theta}_0.$

Recall that $\tilde{Q}_0 = 0$, $X_0 = 0$, and eq. (51), we obtain

$\tilde{Q}_k = \lambda^k (\tilde{\theta}_k - \tilde{\theta}_0)^T \Gamma_0^{-1} (\tilde{\theta}_k - \tilde{\theta}_0).$

This yields eq. (36).