Bias Compensation of Recursive Least Squares Estimate in Closed Loop Environment

Kenji Ikeda *, Yoshio Mogami *, and Takao 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.

Key Words: recursive least squares estimate, bias compensation, closed loop identification.

1. Introduction

There are increasing demands for the identification under the feedback control[1]–[3] due to the safety or economic reasons [4]. 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 [5]–[7] have been 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.

Unbiased estimate is also obtained by using the bias compensated least squares (BCLS) method [8],[9], 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 has proposed an iterative BCLS method for the estimation of the plant parameters in the closed loop environment[10], in which prefilterers 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 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 squared method is proposed. Section VI shows a numerical example for the illustration of the proposed method. Section VII concludes the paper.

Notation:
Let $E[x]$ denote an expectation of random variable $x$. Let $q$ denote a shift operator i.e. $q y_k = y_{k+1}$.

2. Problem Formulation

Consider a single-input single-output (SISO) $n$-th order discrete time plant:

$$y_k = \frac{b_p(q)}{a_p(q)} u_k + \nu_k,$$  \hspace{1cm} (1)
$$a_p(q) = q^n + a_1 q^{n-1} + \cdots + a_n,$$  \hspace{1cm} (2)
$$b_p(q) = b_1 q^{m-1} + \cdots + b_m,$$  \hspace{1cm} (3)

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 $a_p(q)$ is monic and of $n$-th order, while $b_p(q)$ is polynomial whose degree is less than $n$.

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

$$u_k = \frac{b_c(q)}{a_c(q)} (r_k - y_k),$$  \hspace{1cm} (4)
$$a_c(q) = q^m + a_{1,a} q^{m-1} + \cdots + a_{m,a},$$  \hspace{1cm} (5)
$$b_c(q) = b_{0,c} q^m + b_{1,c} q^{m-1} + \cdots + b_{m,c},$$  \hspace{1cm} (6)

where $r_k \in \mathbb{R}$ is a reference input.

The following assumptions are made for the plant, the noise, the I/O data, and the compensator.

(A1) $a_p(q)$ and $b_p(q)$ do not have a common zero outside of the open unit disc.

(A2) an upper bound of the plant degree is known to be $n$.

(A3) the observation noise $\nu_k$ is a zero mean white noise with variance

$$E[\nu_k \nu_l] = \sigma^2 \delta_{kl}$$

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 \( v_k \).

(A6) \( a_r(q) \) and \( b_r(q) \) do not have a common zero and the degree of the compensator \( m \) is such that \( m > n - 1 \).

The assumption (A6) is a sufficient condition for the LS estimate to be determined uniquely when the reference input is zero [11]. From the assumption (A3) and (A6), 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

\[
d_c(q) = a_r(q) + b_r(q) b_o(q),
\]

all the zeros of \( d_c(q) \) lies in the open unit disc from the assumption (A4).

**Problem:** Estimate the unknown coefficients of \( a_r(q) \) and \( b_r(q) \) 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 prefilters are briefly summarized.

Define the characteristic polynomial of the prefilter as

\[
f(q) = q^n + f_1 q^{n-1} + \cdots + f_n,
\]

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 \( a_r(q)/f(q) \), we obtain

\[
y_k = \frac{f(q) - a_r(q)}{f(q)} y_{f,k} + \frac{b_r(q)}{f(q)} u_{f,k} + \frac{a_r(q)}{f(q)} v_k.
\]

From the equation above, the following linear regression formula is obtained:

\[
y_k = \varphi_k^T \theta + \epsilon_k,
\]

where

\[
\theta = [\theta_1, \ldots, \theta_n]^T = [1, a_1, \ldots, f_n - a_1, b_1, \ldots, b_n]^T,
\]

\[
\varphi_k = [y_{f,k-1}, \ldots, y_{f,k-n}, u_{f,k-1}, \ldots, u_{f,k-n}]^T,
\]

\[
\epsilon_k = \frac{a_r(q)}{f(q)} v_k.
\]

Adopting the weighted least squares criterion:

\[
J_\lambda(\hat{\theta}) = \sum_{i=1}^k \lambda^{k-i} \left| y_i - \hat{\varphi}_i^T \hat{\theta} \right|^2
\]

where \( 1 \geq \lambda > 0 \) is a design parameter, the parameter minimizing \( J_\lambda(\hat{\theta}) \) can be given recursively by

\[
\hat{\theta}_k = \hat{\theta}_{k-1} + L_k \epsilon_k,
\]

\[
L_k = \frac{\Gamma_{k-1} \hat{\varphi}_k}{\lambda + \hat{\varphi}_k^T \Gamma_{k-1} \hat{\varphi}_k},
\]

\[
e_k = y_k - \hat{\varphi}_k^T \hat{\theta}_{k-1},
\]

\[
\Gamma_k = \frac{1}{\lambda} \left[ \Gamma_{k-1} - \frac{\Gamma_{k-1} \hat{\varphi}_k \hat{\varphi}_k^T \Gamma_{k-1}}{\lambda + \hat{\varphi}_k^T \Gamma_{k-1} \hat{\varphi}_k} \right].
\]

with initial values \( \hat{\theta}_0 \) and \( \Gamma_0 = \Gamma_0^T > 0 \). It is well known that eq. (19) is obtained by applying the matrix inversion lemma recursively to the r.h.s. of the following equation:

\[
\Gamma_k = \left( \lambda \Gamma_{0}^{-1} + \sum_{i=1}^k \lambda^{k-i} \varphi_i \varphi_i^T \right)^{-1}.
\]

Define \( \hat{\theta}_k = \theta - \theta \) and rewrite eq. (16) for \( \hat{\theta}_k \) by using eq. (11) and the relations \( L_k = \Gamma_k \hat{\varphi}_k \) and \( \lambda \Gamma_k \Gamma_{k-1} = I - L_k \varphi_k^T \) which are obtained from eq. (19):

\[
\hat{\theta}_k = \hat{\theta}_{k-1} + L_k (y_k - \varphi_k^T \hat{\theta}_{k-1}) = \hat{\theta}_{k-1} + L_k (\epsilon_k - \varphi_k^T \hat{\theta}_{k-1}) = (I - \Gamma_k \varphi_k \varphi_k^T) \hat{\theta}_{k-1} + \Gamma_k \varphi_k \epsilon_k = \lambda \Gamma_k \Gamma_{k-1} \hat{\theta}_{k-1} + \Gamma_k \varphi_k \epsilon_k.
\]

Applying eq. (21) recursively, we finally obtain

\[
\theta_k = \theta + \lambda^k \Gamma_0^{-1} \hat{\theta}_0 + \sum_{i=1}^k \lambda^{k-i} \varphi_i \epsilon_i.
\]

Because the equation error \( \epsilon_i \) 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 (22) in the closed loop environment is investigated. Because \( r_k \) and \( v_k \) 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 \( y_{f,k-1} \) and the filtered input \( u_{f,k-1} \) are given by

\[
y_{f,k-1} = \frac{q^n a_r(q)}{d_r(q)} \frac{a_r(q)}{f(q)} v_k = \frac{q^n a_r(q)}{d_r(q)} \frac{d_r(q)}{f(q)} v_k,
\]

\[
u_{f,k-1} = \frac{q^n b_r(q)}{d_r(q)} \frac{a_r(q)}{f(q)} v_k = \frac{q^n b_r(q)}{d_r(q)} \frac{d_r(q)}{f(q)} v_k.
\]

These two equations together with eq. (14) have a state space representation as follows:

\[
\bar{X}_{k+1} = \bar{A} \bar{X}_k + \bar{b} \bar{v}_k,
\]

\[
\varphi_k = \left[ \begin{array}{c} C_{cl} \ 0_{2 \times 0} \end{array} \right] \bar{X}_k,
\]

\[
\epsilon_k = \left[ \begin{array}{c} 0_{1 \times (m+n)} \end{array} \right]^T \bar{X}_k + v_k,
\]

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

\[
\bar{A} = \left[ \begin{array}{c} A_{cl} \ b_h \bar{h}^T \ F \end{array} \right],
\]

\[
\bar{b} = \left[ \begin{array}{c} b_h \ g \end{array} \right].
\]
and \((A_{cl}, b_{cl}, C_{cl})\) and \((F, g, h^T, 1)\) are the system matrices of the state space representations of the transfer functions:

\[
\begin{bmatrix}
A_{cl} & b_{cl} \\
C_{cl} & 0
\end{bmatrix}_{10^{10}} = \begin{bmatrix}
s(q)a_i(q) - s(q)b_i(q) \\
d_i(q)
\end{bmatrix}_{10^{10}}, \tag{30}
\]

\[
Fh = a_i(q) f_i(q). \tag{32}
\]

From eqs. (25), (26), and (27), and taking into account that \(E[\dot{X}_k\dot{X}_k^T] = 0, E[\dot{\varphi}_k\dot{\varphi}_k^T]\) can be calculated as

\[
E(\dot{\varphi}_k\dot{\varphi}_k^T) = \begin{bmatrix} C_{cl} & 0 \end{bmatrix}E(\dot{X}_k\dot{X}_k^T) \begin{bmatrix} 0 \\ h \end{bmatrix}. \tag{33}
\]

Covariance matrix of \(X_k\) is given by

\[
E(\dot{X}_k\dot{X}_k^T) = P\sigma_e^2, \tag{34}
\]

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

\[
P = A_P A_P^T + \tilde{b}b^T, \tag{35}
\]

Finally, we obtain

\[
E(\dot{\varphi}_k\dot{\varphi}_k^T) = C_{cl}P_{12}h^T\sigma_e^2, \tag{36}
\]

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

Let \(A_{cl}\) and \(b_{cl}\) be realized as a controller canonical form. Then \(C_{cl} \in \sigma_e\times (m+n)\) becomes as follows:

\[
C_{cl} = \begin{bmatrix}
1 & \cdots & a_{cm} & 0 \\
0 & \cdots & 1 & a_{cm} \\
-b_{01} & -b_{11} & \cdots & 0 \\
0 & \cdots & -b_{0m} & -b_{cm}
\end{bmatrix}. \tag{37}
\]

Note that \(b_{cl}\) and \(C_{cl}\) are independent of the plant parameters. Also, when \(F\) and \(g\) are realized as a controller canonical form, \(F\) and \(g\) are independent of the plant parameters and \(h\) becomes the coefficient vector of \(a_i(q) - f(q)\), i.e. \(h = -\theta_1\).

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 (4) together with the assumptions (A1) to (A6). Assume that the correlation between \(\Gamma_k\) and \(\sum_{i=1}^{k}\lambda^{k-i}\varphi_i\varphi_i\) is negligible. Then the expectation of the recursive least squares estimate defined by eq. (16) is given by

\[
E\left[\hat{\theta}_k\right] = \theta + \lambda^k \Gamma_k \Gamma_k^{-1} \hat{\theta}_0 + \frac{1 - \lambda^k}{1 - \lambda} \Gamma_k C_{cl} P_{12} h \sigma_e^2 \tag{38}
\]

for \(0 < \lambda < 1\) and

\[
E\left[\hat{\theta}_k\right] = \theta + \Gamma_k \Gamma_k^{-1} \hat{\theta}_0 + \lambda^k \Gamma_k C_{cl} P_{12} h \sigma_e^2 \tag{39}
\]

for \(\lambda = 1\), where \(P_{12}\) is a 1-2 block of \(P\) defined by eqs. (28) to (35).

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

**Remark 1** The condition that the correlation between \(\Gamma_k\) and \(\sum_{i=1}^{k}\lambda^{k-i}\varphi_i\varphi_i\) is negligible is a sufficient condition for the expectation of the product of \(\Gamma_k\) and \(\sum_{i=1}^{k}\lambda^{k-i}\varphi_i\varphi_i\) becomes the product of the expectations of these terms. This condition is satisfied when \(k\) is large and \(\lambda\) is nearly 1.

The matrix \(P_{12}\) can be obtained by solving the following Sylvester equation instead of solving the Lyapunov equation (35):

\[
P_{12} = A_{cl}P_{12}F^T + b_{cl}h^TP_{12}F^T + b_{cl}g^T \tag{40}
\]

where \(P_{22}\) is a solution of the Lyapunov equation

\[
P_{22} = F P_{22} F^T + g g^T, \tag{41}
\]

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 [8],[9], the noise variance \(\sigma^2_e\) is to be estimated from the *a posteriori* error or the least squares residual.

Define \(Q_k\) as

\[
Q_k = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i|k}^2, \tag{42}
\]

where \(\hat{e}_{i|k}\) is the *a posteriori* error defined by

\[
\hat{e}_{i|k} = y_i - \varphi_i^T \hat{\theta}_k. \tag{43}
\]

Also define \(\hat{Q}_k\) recursively as

\[
\hat{Q}_k = \lambda \hat{Q}_{k-1} + \frac{\lambda \sigma^2_e}{\lambda + \varphi_i^T (\Gamma_{k-1}^{-1} \varphi_i)}, \tag{44}
\]

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

\[
\hat{Q}_k = Q_k + \lambda^k (\hat{\theta}_k - \theta_0)^T \Gamma_k^{-1} (\hat{\theta}_k - \theta_0) \tag{45}
\]

(See Appendix A for the derivation of eq. (45).)

By using the relation \(\hat{e}_{i|k} = y_i - \varphi_i^T \hat{\theta}_k\), \(Q_k\) becomes

\[
Q_k = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i|k}^2 = 2 \hat{\beta}_k + \hat{\theta}_k^T (\Gamma_k^{-1} - \lambda \Gamma_k^{-1}) \hat{\theta}_k. \tag{46}
\]

By using eq. (20), we obtain

\[
Q_k = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i|k}^2 = 2 \hat{\beta}_k + \hat{\theta}_k^T (\Gamma_k^{-1} - \lambda \Gamma_k^{-1}) \hat{\theta}_k, \tag{47}
\]

where

\[
\hat{\beta}_k = \lambda^k \varphi_i \varphi_i \tag{48}
\]

Define

\[
\alpha_k = \lambda^k \Gamma_k^{-1} \hat{\theta}_0. \tag{49}
\]
Then eq. (22) is rewritten as
\[ \tilde{\theta}_k = \Gamma_k(\alpha_k + \beta_k). \]  
(50)

From this and eqs. (45) and (47), we obtain
\[ \dot{Q}_k = \sum_{i=1}^{k} \lambda^{i-k} \epsilon_i^2 - \beta_k^2 \Gamma_k \dot{R}_k - \alpha_k^2 (\Gamma_k \alpha_k + 2 \Gamma_k \beta_k - \tilde{\theta}_0). \]  
(51)

Expectation \( E[\epsilon_i^2] \) is already analyzed in the previous section while \( E[\epsilon_i^2] \) can be calculated similarly as
\[ E[\epsilon_i^2] = (1 + h^2 P_{22} h) \sigma_v^2. \]  
(52)

where \( P_{22} \) is a 2-2 block of \( P \) defined by eq. (35), or, equivalently, is a solution of the Lyapunov equation (41).

Thus, the following theorem is obtained.

**Theorem 2** Under the same assumptions as in Theorem 1, the expectation of \( \dot{Q}_k \) defined in (44) is given by
\[ E[\dot{Q}_k] = \frac{1 - \lambda^k}{1 - \lambda} (1 + h^2 P_{22} h) \sigma_v^2 \]
\[ - \left(1 - \frac{1}{1 - \lambda^2}\right) h^2 P_{12}^* \Gamma_k \Gamma_k P_{12} h \sigma_v^2 \]
\[ - \alpha_k^2 \left[ \Gamma_k \alpha_k + 2 \left(1 - \lambda^k\right) \Gamma_k \Gamma_k P_{12} h \sigma_v^2 - \tilde{\theta}_0 \right] \]
for \( 0 < \lambda < 1 \) and
\[ E[\dot{Q}_k] = k (1 + h^2 P_{22} h) \sigma_v^2 - k^2 h^2 P_{12}^* \Gamma_k \Gamma_k P_{12} h \sigma_v^2 \]
\[ - \alpha_k^2 \left[ \Gamma_k \alpha_k + 2 \Gamma_k \Gamma_k P_{12} h \sigma_v^2 - \tilde{\theta}_0 \right] \]
for \( \lambda = 1 \).

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

The noise variance will be estimated by solving the second order equation (53) or (54) for \( \sigma_v^2 \). 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
\[ \sigma_v^2 = \left(1 - \frac{1}{1 - \lambda^2}\right) R_k \]
for \( 0 < \lambda < 1 \) and
\[ \sigma_v^2 = \frac{1}{k} R_k \]
for \( \lambda = 1 \) where
\[ R_k = \frac{2 \dot{Q}_k / (1 + h^2 P_{22} h)}{1 + \sqrt{1 - \frac{4 \dot{Q}_k h^2 P_{12}^* \Gamma_k \Gamma_k P_{12} h}{(1 + h^2 P_{22} h)^2}}} \]
(57)

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. (40) and (57) must be replaced by their estimates, which will be denoted by \( \hat{A}_{cl}, \hat{h}, \hat{P}_{12}, \) 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, \( \hat{A}_{cl}, \hat{h} \) and \( \hat{h} \) are to be defined by using \( \hat{\theta}_{BC,k} \).

Summarizing above, the bias compensated RLS estimate will be defined as
\[ \hat{\theta}_{BC,k} = \dot{\theta}_k - \Gamma_k C_k \hat{P}_{12} \hat{h} \hat{R}_k. \]
(58)

where \( \dot{\theta}_k, \Gamma_k, \) and \( \hat{P}_{12} \) are defined in eqs. (16) to (19) and (44), \( \hat{P}_{12,k} \) is a solution of the Sylvester equation
\[ \hat{P}_{12,k} = A_{cl} \hat{P}_{12,k} F^T + b_k h_k \hat{P}_{22} F^T + h_k R_k, \]
\( \hat{R}_k \) is defined
\[ \hat{R}_k = \frac{2 \dot{Q}_k / (1 + h_k^2 P_{22} h_k)}{1 + \sqrt{1 - \frac{4 \dot{Q}_k h_k^2 \hat{P}_{12,k}^* \Gamma_k \Gamma_k \hat{P}_{12,k} h_k}{(1 + h_k^2 P_{22} h_k)^2}}} \]
and \( \hat{A}_{cl} \) and \( \hat{h} \) are defined by using \( \hat{\theta}_{BC,k} \).

**Remark 2** Neglecting the initial term \( \lambda^k \Gamma_k^{-1} \dot{\theta}_0 \) in the expectation of \( \dot{\theta}_k \) in eqs. (38) or (39), the expectation of \( \dot{\theta}_{BC,k} \) in eq. (58) can be calculated as follows:
\[ E[\dot{\theta}_{BC,k}] = E[\dot{\theta}_k] - \Gamma_k C_k \hat{P}_{12,k} \dot{h} \hat{R}_k. \]

Because \( \hat{P}_{12,k} \dot{h} \hat{R}_k \to P_{12} R_k \) as \( \lambda \to K \) the following equation is obtained
\[ \lim_{\hat{\theta}_{BC,k} \to \theta} E[\hat{\theta}_{BC,k} - \theta] = M_k (\hat{\theta}_{BC,k} - \theta), \]
where
\[ M_k = -\Gamma_k C_k \frac{\partial}{\partial \hat{\theta}_{BC,k}} P_{12} h R_k. \]
When the noise and the reference input are stationary, there exists \( M = \lim_{k \to \infty} M_k \). If \( |\lambda_i(M)| < 1 \) for \( i = 1, \ldots, 2n \) where \( \lambda_i(M) \) denotes the \( i \)-th eigenvalue of \( M \), then the algorithm proposed above is locally convergent. Eigen values of \( M_k \) depend on the reference input. When the reference input is informative enough and its magnitude is comparatively large, \( \Gamma_k \dot{Q}_k \) becomes small and \( |\lambda_i(M)| \) becomes less than 1. From some numerical studies, it is found that there are some cases when \( |\lambda_i(M)| \) becomes less than 1 even though the reference input is not persistently exciting of order \( 2n \). The condition for the proposed algorithm to be convergent when the reference input is not informative enough is under investigation and remains to be a future work. In practice, the eigen values of
\[ \tilde{M}_k = -\Gamma_k C_k \frac{\partial}{\partial \hat{\theta}_{BC,k}} P_{12} h R_k, \]
will be a criterion for the the proposed algorithm to be convergent or not.

**Remark 3** It is well known that larger \( \lambda \) will be required for the smaller variance of the estimate for time invariant systems. In order to compensate the asymptotic bias of the RLS estimate successfully, thousands of samples or more will be required in general. This means that the forgetting factor \( \lambda \) is requested to be more than 0.999. On the other hand, large \( \lambda \) will cause the large delay for the estimation. If the purpose of the estimation is online monitoring of the plant, precise estimate of the plant parameter and the real-time property of the estimation will be a trade-off.
of the covariance of the BCLS estimate $\|\text{Cov}(\hat{\theta}_{BC})\|_2$ is larger than that of the RLS estimate while the other singular values are comparative. On the other hand, the norm of the averaged BCLS estimation error $\|\bar{\theta}_{BC,N} - \theta\|_2 = 3.5043 \times 10^{-3}$ is very small compared with that of RLS estimation error $\|\bar{\theta}_N - \theta\|_2 = 0.7451$, where the averaged RLS and BCLS estimates are denoted by $\bar{\theta}_N$ and $\bar{\theta}_{BC,N}$, respectively. As a result, the BCLS estimate is closer to the true value than the RLS estimate because the norm of mean of the RLS estimation error (0.7451) is larger than $\sqrt{\|\text{Cov}(\hat{\theta}_{BC})\|_2}$.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Singular values of the covariance matrices of RLS and BCLS estimates.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>$2.0145 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$3.5841 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$2.4451 \times 10^{-7}$</td>
</tr>
<tr>
<td>$\sigma_4$</td>
<td>$6.6720 \times 10^{-9}$</td>
</tr>
<tr>
<td>$\sigma_5$</td>
<td>$6.8176 \times 10^{-10}$</td>
</tr>
<tr>
<td>$\sigma_6$</td>
<td>$2.1023 \times 10^{-13}$</td>
</tr>
</tbody>
</table>

7. Conclusion

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

**References**


Appendix A

In this appendix, eq. (45) is to be derived in the following 3 steps.

(i) Define $X_k$ as

$$X_k = \sum_{i=1}^{k} \lambda^{k-i} \Gamma_{i-1} \hat{e}_{i/k}. \quad (65)$$

Then, the following equation holds:

$$X_k = \lambda^{k} \Gamma_{0}^{-1} (\hat{b}_k - \bar{b}_0). \quad (66)$$

Proof: Note that $\hat{e}_{i/k} = \hat{e}_{i/k-1} - \varphi_i^T (\hat{b}_k - \bar{b}_{k-1})$ and $e_i = \hat{e}_{i/k-1}$. Applying these equations for (65), the following equation is obtained

$$X_k = \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \hat{e}_{i/k-1} - \left( \sum_{i=1}^{k} \lambda^{k-i} \varphi_i \varphi_i^T \right) (\hat{b}_k - \bar{b}_{k-1}).$$

$$= \lambda X_{k-1} + \lambda^{k} \Gamma_{0}^{-1} \hat{b}_k - \lambda X_{k-1} \bar{b}_{k-1} - \lambda^{k} \lambda^{k-1} \bar{b}_{k-1}. \quad (67)$$

By using (20),

$$X_k = \lambda X_{k-1} + \lambda^{k} \Gamma_{0}^{-1} (\hat{b}_k - \bar{b}_{k-1})$$

$$+ \lambda^{k-1} \Gamma_{0}^{-1} [ \varphi_i e_i - (\hat{b}_k - \bar{b}_{k-1})]. \quad (68)$$

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

$$X_k = \lambda X_{k-1} + \lambda^{k} \Gamma_{0}^{-1} (\hat{b}_k - \bar{b}_{k-1}). \quad (69)$$

Applying this equation recursively, we obtain

$$X_k = \lambda^k X_0 + \lambda^{k-1} \Gamma_{0}^{-1} (\hat{b}_k - \bar{b}_0). \quad (70)$$

The initial condition $X_0 = 0$ is consistent with the definition of $X_1$ and the recursive form of $X_k$. This proves eq. (66).

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

$$Q_k = \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k}^T \hat{e}_{i/k}$$

$$= \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} (e_i - \varphi_i^T \hat{b}_k)$$

$$= \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} e_i - X_k^T \hat{b}_k \quad (71)$$

Applying $\hat{e}_{i/k} = \hat{e}_{i/k-1} - \varphi_i^T (\hat{b}_k - \bar{b}_{k-1})$, we obtain

$$Q_k = \hat{e}_{k/k} e_k - \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} e_i - X_k^T \hat{b}_k$$

$$+ Q_{k-1} + \lambda X_{k-1} \bar{b}_{k-1} - X_k^T \hat{b}_k \quad (72)$$

Applying (22) for $\hat{b}_{k-1}$ together with equations $\hat{e}_{k/k} = \lambda \hat{e}_{k/k} + \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} e_i$ and $e_i = e_i + \varphi_i^T \hat{b}_{k-1}$, the first term of the r.h.s. of the equation above is calculated as

$$\hat{e}_{k/k} = \lambda \hat{e}_{k/k} + \sum_{i=1}^{k} \lambda^{k-i} \hat{e}_{i/k} e_i + \lambda \hat{b}_{k-1} - \bar{b}_{k-1} \Gamma_{0}^{-1} \bar{b}_0$$

$$+ (\hat{b}_{k-1} - \bar{b}_{k-1})^T \sum_{i=1}^{k} \lambda^{k-i} \varphi_i^T e_i. \quad (73)$$

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

$$Q_k = \lambda Q_{k-1} + \lambda \hat{e}_{k/k} \hat{e}_{k/k} + \lambda X_{k-1} \bar{b}_{k-1} - X_k^T \hat{b}_k + \lambda \hat{b}_{k-1} - \bar{b}_{k-1} \Gamma_{0}^{-1} \bar{b}_0. \quad (74)$$

Applying this equation recursively, we obtain

$$Q_k = \lambda^k \hat{Q}_0 + X_k^T \hat{b}_k - \lambda X_{k-1} \bar{b}_{k-1} - \lambda (\hat{b}_k - \bar{b}_k) \Gamma_{0}^{-1} \bar{b}_0. \quad (75)$$

Recall that $Q_0 = 0$, $X_0 = 0$, and eq. (66), we obtain

$$Q_k = \lambda^k (\hat{b}_k - \bar{b}_k) \Gamma_{0}^{-1} (\hat{b}_k - \bar{b}_0). \quad (76)$$

This yields eq. (45).