INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2004; **18**:551–570 (DOI: 10.1002/acs.813)

Model-free recursive LQ controller design (learning LQ control)

Yoshiaki Kawamura[1,*,†], Masahiro Nakano[2] and Hiroyuki Yamamoto[1]

[1] *Osaka Electro-Communication University, 18-8 Hatsucho, Neyagawa City, Osaka 572-8350, Japan*
[2] *Mitsubishi Electric Co., Ltd. 8-1-1 tsukaguchihonmachi, Amagasaki City, Hyougo 661-8661, Japan*

SUMMARY

A new data-based iterative self-optimizing approach to practical design (learning/adaptive process) of the infinite-horizon LQ regulator is proposed. Optimality is given by a certain orthogonality condition of response signals, and the global convergence of feedback gain is proved for MIMO systems by an expansion of the Riccati equation. The design is applied to stabilizing control and steady state error-less control of physical systems. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: LQ control; Riccati equation; orthogonality; learning control; adaptive control; self-tuning regulator

1. INTRODUCTION

Although the use of a model had been regarded as essential for designing the LQ regulator, we recently found some propositions concerning model-free (data-based) designs of the LQ regulator [1–7] that give the regulator without knowing the system matrices. The model-free minimum-variance control [8, 9] seems applicable to the LQ problem as well. We also found a prototype of model-free design in the 1960s in relation to the early modern control theory [10, 11]. However there exist various difficulties to apply them to physical systems, and developing a useful method is desirable.

The purpose of this paper is to introduce a new data-based approach to the infinite-horizon LQ controller design. That is aimed at developing practical algorithm. Kawamura primitively noticed the key relation in the 1980s when he considered the dual relation of the well-known orthogonality related to the Kalman filter. That led to some orthogonality relations between the responses of the LQ regulator [12]. Those days were bright age of model-reference adaptive control and model-based self-tuning regulator [13, 14], and the iterative learning control by Arimoto *et al.* was also attractive [15, 16]. Those results were not applicable at all because they were essentially different from the model-free LQ design. Since then, Kawamura and some students have engaged in developing a related method, and they have reported many interim results mainly in Japanese. Although it gave an alternative approach to the previous

*Correspondence to: Y. Kawamura, Department of Electronics, Faculty of Engineering, Osaka Electro-Communication University, 18-8 Hatsucho, Neyagawa City, Osaka 572-8530, Japan.
† E-mail: kawamura@isc.osakac.ac.jp

Received 23 April 2003
Revised 1 March 2004
Accepted 16 March 2004
Published online 21 May 2004

identification-based LQ design [1, 2], (the latter is written in English), it could not alouse sufficient interest of researchers.

The original model-free LQ design by Kawamura is characterized by calculation of the inner product as sensitivity. The main drawback is its supersensitivity to a constant value, e.g. the static friction of physical systems, because of the long-term calculation of the inner product. Other different model-free LQ design methods have similar defect. We cannot obtain practicable model-free LQ design without the overcoming. The current approach has been developed as a countermeasure. It involves two new basic ideas:

- Introduction of a recursive (one-stage) data-based calculation of sensitivity of unknown systems.
- Introduction of statistical data processing and its convergence analysis.

We refer to this new approach as the 'recursive algorithm'.

2. RECURSIVE RELATION

2.1. Inner product of responses and sensitivity

First, we show some definitions. We study the linear time-invariant system

$$x(t+1) = Ax(t) + Bu(t) \tag{1}$$

defined on $t = 0, 1, 2, \ldots$ as the state $x(t) \in R^n$ and the input $u(t) \in R^m$. We assume that the state is directly measurable. Let

$$z(t) = \begin{pmatrix} Cx(t) \\ Du(t) \end{pmatrix} \tag{2}$$

be the generalized controlled output for $t \leqslant L-1$ where $z(t) \in R^l$, and let

$$z(L) = C_1 x(L). \tag{3}$$

Let $\mathbf{z}$ denote the whole signal $\{z(0), z(1), \ldots, z(L)\}$. Define an inner product [12] of two signals $\mathbf{z}_1$ and $\mathbf{z}_2$ by

$$\begin{aligned} \langle \mathbf{z}_1, \mathbf{z}_2 \rangle_L &= \sum_{t=0}^{L-1} z_1(t)^{\mathrm{T}} z_2(t) + z_1(L)^{\mathrm{T}} z_2(L) \\ &= \sum_{t=0}^{L-1} \{x_1(t)^{\mathrm{T}} C^{\mathrm{T}} C x_2(t) + u_1(t)^{\mathrm{T}} D^{\mathrm{T}} D u_2(t)\} \\ &\quad + x_1(L)^{\mathrm{T}} C_1^{\mathrm{T}} C_1 x_2(L). \end{aligned} \tag{4}$$

Note that the standard LQ performance index is given by $V_L = \langle \mathbf{z}, \mathbf{z} \rangle_L$.

Consider the feedback input $u(t) = Gx(t) + v(t)$ where $G \in R^{m \times n}$ and $v(t) \in R^m$ is a pulse at $t = 0$. Since $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle_L$ is bi-linear with regard to the initial signals $\xi_i(0) = (x_i(0)^{\mathrm{T}} v_i(0)^{\mathrm{T}})^{\mathrm{T}}$, certain matrices Γ, Γ^{aa}, $\Gamma^{ab} = (\Gamma^{ba})^{\mathrm{T}}$ and Γ^{bb} satisfy

$$\begin{aligned}\langle \mathbf{z}_1, \mathbf{z}_2 \rangle_L &= \xi_1(0)^{\mathrm{T}} \Gamma \xi_2(0) \\ &= x_1(0)^{\mathrm{T}} \Gamma^{aa} x_2(0) + x_1(0)^{\mathrm{T}} \Gamma^{ab} v_2(0) \\ &\quad + v_1(0)^{\mathrm{T}} \Gamma^{ba} x_2(0) + v_1(0)^{\mathrm{T}} \Gamma^{bb} v_2(0). \end{aligned} \tag{5}$$

This relation means that the matrices Γ^{aa}, Γ^{ab} and Γ^{bb} indicate the zero-, first- and second-order sensitivities of the inner product with regard to the additional pulse $v(0)$. We call Γ the fundamental sensitivity matrix. It plays the central role of this paper. Note that the sensitivity is connected with the inner product though it is usually mentioned in relation to the output.

The problem is the infinite-horizon LQ control problem in which we minimize the total cost V_∞. We use the following standard assumption:

Assumption 1

(A, B) is stabilizable, (C, A) is detectable and $D^{\mathrm{T}} D > 0$.

Under Assumption 1, it is well known that a time-invariant state feedback $u(t) = Gx(t)$ comprises the stable infinite-horizon LQ optimal control [13]. Therefore we regard that the problem is optimizing the time-invariant gain $G \in R^{m \times n}$ without information about A and B. We study the problem in terms of the inner product.

2.2. Orthogonality condition

The fundamental problem of model-free design is how to judge the optimality of an unknown system. First, we briefly explain the orthogonality of the LQ regulator as the key of this approach. Let $G \in R^{m \times n}$, and let $u(t) = Gx(t) + v(t)$. Define two closed-loop response signals: $\mathbf{z}^a$ is an initial state response such that $x^a(0) \in R^n$ and $v^a(t) \equiv 0$; $\mathbf{z}^b$ is a pulse response such that $x^b(0) = 0$, $v^b(0) \in R^m$ and $v^b(t) \equiv 0$ for $t \geqslant 1$.

Under Assumption 1, it was shown [12, 2] (the latter is an English paper) that G is the infinite-horizon LQ gain if and only if

$$\lim_{L \to \infty} \langle \mathbf{z}^a, \mathbf{z}^b \rangle_L = 0 \tag{6}$$

for any $x^a(0)$ and $v^b(0)$.

Remark 1

Since we usually pay attention to the response $Cx(t)$, orthogonality does not hold for the LQ problem, and it was not mentioned in related literature. The vector output $z(t)$ contains not only $Cx(t)$ but also the input factor $Du(t)$ as in (2). The inner product of a signal with itself agrees with the standard LQ performance index. This relation is crucial for the orthogonality. Although the pulse is given at only $t = 0$, the time-invariance of the system gives the optimality at every t. The following is a brief explanation of (6): the optimal response (initial state response) is orthogonal to a perturbation (pulse response). This is the dual relation of the orthogonality between the optimal estimation error and the innovation [17]. The orthogonality led to the original algorithm [1, 2] (English).

Long-term responses ($L \to \infty$) are necessary to literally check this orthogonality, and considerable signal processing is unavoidable. As the result, we have not developed a useful model-free LQ design method. In order to solve it, we introduce the recursive algorithm with a finite L. Although we can obtain similar results for any $L = 1, 2, \ldots$, we restrict our discussion to $L = 1$. The reason is that the convergence proof is valid only if $L = 1$.

Let $P = C_1^{\mathrm{T}} C_1 \geqslant 0$ be the terminal cost matrix, and let $\mathbf{z}_i$ $(i = 1, 2)$ be a response given by $x_i(0)$ and $u_i(t) = Gx_i(t) + v_i(t)$ where $v_i(t)$ is a pulse at $t = 0$. If $L = 1$, the inner product has the form,

$$\langle \mathbf{z}_1, \mathbf{z}_2 \rangle_{L=1} = z_1(0)^{\mathrm{T}} z_2(0) + x_1(1)^{\mathrm{T}} P x_2(1). \tag{7}$$

We have the following relations (see Appendix A).

Lemma 1

Suppose that (1), (2) and the feedback relation hold. Then the sensitivity matrices defined in (5) are given for $L = 1$ by

$$\Gamma^{aa} = (A + BG)^{\mathrm{T}} P (A + BG) + C^{\mathrm{T}} C + G^{\mathrm{T}} D^{\mathrm{T}} D G$$

$$\Gamma^{ab} = (A + BG)^{\mathrm{T}} P B + G^{\mathrm{T}} D^{\mathrm{T}} D$$

$$\Gamma^{bb} = B^{\mathrm{T}} P B + D^{\mathrm{T}} D.$$

We transform the orthogonality to $L = 1$ provided that P has a special value indicated below (see Appendix B for the proof).

Theorem 1

Suppose Assumption 1. Let $P \geqslant 0$ be the (n, n) matrix used in the right-hand side of (7). Then $G \in R^{m \times n}$ is the infinite-horizon LQ gain if and only if the matrices P and G satisfy

$$x^a(0)^{\mathrm{T}} P x^a(0) = \langle \mathbf{z}^a, \mathbf{z}^a \rangle_{L=1}$$

$$\langle \mathbf{z}^a, \mathbf{z}^b \rangle_{L=1} = 0$$

for any $x^a(0) \in R^n$ and $v^b(0) \in R^m$.

In other words, a part of the sensitivity related to future data ($t \geqslant 1$) is replaced with P. Recalling definitions and Lemma 1, we also express the condition as $P = \Gamma^{aa}$ and $\Gamma^{ab} = 0$. This orthogonality condition corresponds to the algebraic Riccati equation as shown later.

We see this orthogonality in the simple first-order system:

$$x(t + 1) = x(t) + u(t), \quad C = 1, \quad D = \sqrt{2}. \tag{8}$$

The above responses ($t < L$) are

$$z^a(t)^{\mathrm{T}} = ((1 + G)^t x^a(0), \sqrt{2} G (1 + G)^t x^a(0))$$

$$z^b(t)^{\mathrm{T}} = \begin{cases} (0, \sqrt{2} v^b(0)) & (t = 0) \\ ((1 + G)^{t-1} v^b(0), \sqrt{2}\, G (1 + G)^{t-1} v^b(0)) & (t = 1, 2, \ldots). \end{cases}$$

For a stabilizing gain such that $-2 < G < 0$, we have

$$\langle \mathbf{z}^a, \mathbf{z}^b \rangle_{L=\infty} = x^a(0) \left\{ 2G + (1 + 2G^2) \sum_{t=0}^{\infty} (1 + G)^{2t+1} \right\} v^b(0).$$

On the other hand, the condition of Theorem 1 is

$$P = (1 + 2G^2) + (1 + G)^2 P$$

$$\langle \mathbf{z}^a, \mathbf{z}^b \rangle_{L=1} = x^a(0) \{2G + (1 + G)P\} v^b(0) = 0$$

by using (7). Applying the sum of the geometric series P, we have

$$\langle \mathbf{z}^a, \mathbf{z}^b \rangle_{L=\infty} = \langle \mathbf{z}^a, \mathbf{z}^b \rangle_{L=1} = x^a(0) \frac{(1 + 2G)(1 - G)}{1 - (1 + G)^2} v^b(0) = 0$$

Since $1 - G = 0$ means a divergence of $(1 + G)^{2t}$, the unique solution of these conditions is $1 + 2G = 0$, namely, $G = -0.5$. In comparison, the algebraic Riccati equation clearly gives the positive solution $P = 2$ and the time-invariant LQ gain $G = -0.5$. They give the same LQ gain.

3. RECURSIVE ALGORITHM

3.1. New algorithm

The theorem suggests that we can obtain the LQ regulator by finding the matrices P and G that satisfy the theorem. Because of (5), we directly obtain the sensitivities to optimize the system from response data. In this section, we explain the recursive algorithm given from these ideas. Detailed examination will be given later. Although we restrict the discussion to the basic state feedback, refer to Section 7 for the output feedback.

The point of the model-free design is the estimate of the sensitivity matrix. Let $k = 0, 1, 2, \ldots$ be the iteration number of changing gain. Let T_k be the kth sequential group of sampling times. Suppose that T_k has at least $n + m$ samples for each k (note that k is different from t). Suppose that P_k, H_k and G_k are given at the beginning of T_k. Feed the input

$$u(t) = G_k x(t) + v(t) \tag{9}$$

to the system on T_k, and observe the response where $v(t) \in R^m$ is a signal with sufficient richness (a white noise is typical).

Calculate a $(n + m, n + m)$ matrix $\hat{\Gamma}_k$ as

$$\hat{\Gamma}_k = \Xi_k \left[\left\{ \sum_{t \in T_k} \sum_{s \in T_k} \xi(t) \{ z(t)^{\mathrm{T}} z(s) .. + x(t+1)^{\mathrm{T}} P_k x(s+1) \} \xi(s)^{\mathrm{T}} \right\} + \delta^2 \hat{\Gamma}_{k0} \right] \Xi_k \tag{10}$$

on the basis of response data where

$$\Xi_k = \left\{ \sum_{t \in T_k} \xi(t) \xi(t)^{\mathrm{T}} + \delta I \right\}^{-1}, \quad \xi(t) = \begin{pmatrix} x(t) \\ v(t) \end{pmatrix} \tag{11}$$

are used to cancel individual data. The number $\delta \geqslant 0$ is sufficiently small, I is the identity matrix and $\hat{\Gamma}_{k0}$ is a matrix such that $\hat{\Gamma}_{k0}^{bb} > 0$. A typical $\hat{\Gamma}_{k0}$ is the block-diagonal matrix given by P_k and $\hat{\Gamma}_{k-1}^{bb}$ where

$$\hat{\Gamma}_k = \begin{pmatrix} \hat{\Gamma}_k^{aa} & \hat{\Gamma}_k^{ab} \\ \hat{\Gamma}_k^{ba} & \hat{\Gamma}_k^{bb} \end{pmatrix}. \tag{12}$$

The matrices G_{k+1}, P_{k+1} and H_{k+1} are given by response data on T_k as follows. Determine the following matrices:

$$\begin{aligned} G_{k+1} &= G_k + \Delta G_k \\ &= G_k - (\hat{\Gamma}_k^{bb} + H_k)^{-1} \hat{\Gamma}_k^{ba} \end{aligned} \tag{13}$$

$$\begin{aligned} P_{k+1} &= \hat{\Gamma}_k^{aa} + \hat{\Gamma}_k^{ab} \Delta G_k + \Delta G_k^{\mathrm{T}} \hat{\Gamma}_k^{ba} + \Delta G_k^{\mathrm{T}} \hat{\Gamma}_k^{bb} \Delta G_k \\ &= \hat{\Gamma}_k^{aa} - \hat{\Gamma}_k^{ab} (\hat{\Gamma}_k^{bb} + H_k)^{-1} (\hat{\Gamma}_k^{bb} + 2H_k)(\hat{\Gamma}_k^{bb} + H_k)^{-1} \hat{\Gamma}_k^{ba} \end{aligned} \tag{14}$$

at the end of T_k. Since the (m, m) matrix $H_k \geqslant 0$ decreases the gain change ΔG_k in (13), we can select H_k as long as it satisfies

$$H_{k+1} = \lambda_{1k} H_k + \lambda_{2k} \hat{\Gamma}_k^{bb} \tag{15}$$

where λ_{1k} and λ_{2k} are any numbers such that

$$0 \leqslant \lambda_{1k} \leqslant 1, \quad 0 \leqslant \lambda_{2k} \leqslant 1. \tag{16}$$

If G_0, P_0 and H_0 are given, repetition of the above process gives a sequence of gains $G_0, G_1, G_2, \ldots$ together with $P_0, P_1, \ldots$ and $H_0, H_1, \ldots$ as $k = 0, 1, 2, \ldots$ on the basis of on-line response data without information about A and B. The initial matrices need not be optimal. The initial conditions are

$$P_0 \geqslant 0, \quad H_0 \geqslant 0, \quad \hat{\Gamma}_{00}^{bb} > 0 \tag{17}$$

and an example of $\hat{\Gamma}_{00}^{bb}$ is $D^{\mathrm{T}} D$. The initial gain G_0 is any (m, n) matrix (typically $G_0 = 0$). We will prove that G_k converges to the LQ optimal gain as $k \to \infty$. Then we know that this model-free design corresponds to the direct (implicit) method of adaptive control.

Remark 2

The generalized minimum variance self-tuning regulator [20] is well-known, where a fixed positive matrix P is used to calculate the performance. Roughly speaking, the matrix is replaced with the variable P_k in the above, and we can optimize P_k together with G_k as k increases.

Remark 3

Let N_k be the number of elements of T_k. The N_k elements need not be consecutive sampling times. If $(A + BG_k)$ is unstable, it is necessary to reset the state repeatedly in observation of responses. Note that (10) is composed of each data-pair at t and $t + 1$. We can continue the algorithm by excluding invalid data-pairs related to resets.

Remark 4

The small elements of δI and $\delta^2\hat{\Gamma}_{k0}$ usually play no role. However, they ($\delta > 0$) guarantee the existence of Ξ_k and $(\hat{\Gamma}_k^{bb} + H_k)^{-1}$ even if T_k does not have $n + m$ linearly independent $\xi(t)$. This is similar to the recursive least-square method. If $\delta = 0$, it is necessary to continue observation on T_k until the inverses exist or by selecting a large N_k.

3.2. Explanation using a gradient method

We often discuss adaptive schemes in terms of gradient methods. For the LQ problem, gradient methods are studied by Kawamura [18, 19]. The recursive algorithm occurred by extending the Riccati equation as a gradient method. The following is developments from them.

The algorithm is composed of the estimate of the sensitivity and the optimization process. Compare (10) with (5) by referring to (7) provided that $\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{z}$. The right-hand sides of them are similar where each $\{z(t), x(t+1)\}$ $(t \in T_k)$ fragment of the real-time response $\mathbf{z}$ replaces the unit-length response $\{z_i(0), x_i(1)\}$. Similarly, each $v(t)$ fragment of $\mathbf{v}$ replaces the unit-length pulse $v_i(0)$. Effect of individual signals are removed by multiplying Ξ_k. We use the following ideal assumption in the theoretical analysis:

Assumption 2

Data are given by (1), (2) and (9) exactly. The sum $\sum_{t \in T_k} \xi(t)\xi(t)^{\mathrm{T}}$ is positive, and the small δ is negligible in (10) and (11).

Under this assumption, we understand $\hat{\Gamma}_k$ as follows (see Appendix C):

Lemma 2

Suppose Assumption 2. Then it follows that

$$\hat{\Gamma}_k = \Gamma. \tag{18}$$

In other words, $\hat{\Gamma}_k$ is an estimate of the sensitivity matrix under such uncertain data that do not necessarily satisfy Assumption 2.

The remaining part is related to the optimization. We decompose the optimal control problem (multi-stage optimization problem) into recursive one-stage problems, which we usually solve backwards from the terminal time. Relation (14) is a gradient method for each one-stage problem under Assumption 2. Determine matrix P_{k+1} so that $x_k^a(0)^{\mathrm{T}} P_{k+1} x_k^a(0)$ is the total cost $\langle \mathbf{z}^a, \mathbf{z}^a \rangle_{L=1}$ provided that $x_k^a(1)$ is given by G_{k+1}. Since P_{k+1} is quadratic with respect to G_{k+1}, we have the first equality of (14) by taking the gain change ΔG_k into consideration. The second equality follows by substituting (13). The recursive algorithm is composed of repetition of this one-stage optimization, that is valid regardless of stability.

If $H_k \equiv 0$ as a special case, (14) gives the second-order gradient method (Newton method). If the response has no irregular behaviour, the Newton method minimizes P_{k+1}. In contrast, non-linearity of physical systems causes inconsistency with Assumption 2, that leads to a random error of $\hat{\Gamma}_k^{bb}$ in (13). Then random variation of ΔG_k does not vanish as $k \to \infty$. It becomes indispensable to construct the optimal system under uncertain information on system sensitivity. Therefore, statistical treatment is inevitable.

Stochastic approximation is the well-known statistical processing. However, it is especially inadequate for unstable or nearly unstable feedback. It is necessary to reduce $\|\Delta G_k\|$ with a suitable ratio regardless of stability of $A + BG_k$. We use $H_k \geqslant 0$ to solve this problem.

Remark 5

Hjalmarsson *et al*. [8, 9] proposed a considerably different iterative model-free method based on the sensitivity $\lim_{L\to\infty} \mathrm{d}V_k/\mathrm{d}G_k$, that is different from $\hat{\Gamma}_k$. Suppose that k is sufficiently large, $A + BG_k$ is stable and the change of gain is negligible. Then we derive their relevance

$$\lim_{L\to\infty} \mathrm{d}V_k = \hat{\Gamma}_k^{ab}\, \mathrm{d}G_k \left\{ \sum_{t=0}^{\infty} x^a(t) x^a(t)^{\mathrm{T}} \right\}.$$

These sensitivities have different dependence on G_k because $x^a(t)$ depends on G_k. Therefore they give quite different behaviours of convergence.

Remark 6

Introduction of H_k and conditions (15) and (16) are based on an analogy of Landau's adaptive scheme [21]. Roughly speaking, we use a similar idea for the LQ problem by replacing the system equation with the Riccati-like difference equation. Namely, the proof shown in Section 4.2 and Appendix E is related to [22, 23] (the latter is an extension of the former by regarding the coefficient of gain correction as variable). A peculiar discussion of the LQ design is caused by the fact that the error term of P_k is not necessary non-negative.

4. CONVERGENCE ANALYSIS

4.1. Relation to the Riccati equation

It is important for the model-free design to prove the convergence of G_k at the LQ gain. Using Lemma 2, we immediately obtain the following relation:

Theorem 2

Suppose Assumption 2. Then, the recursive algorithm (9)–(17) is equivalent to (12)–(17).

Referring to Lemmas 1 and 2, we know that the latter equations are written by the system matrices. We analyse the convergence in terms of the matrices. Consider a special case $H_0 = 0$ and $\lambda_{2k} \equiv 0$, namely, $H_k \equiv 0$. Then, we see the following relation by rewriting G_k (see Appendix D):

Corollary 1

Suppose Assumption 2 and $H_k \equiv 0$. Then, the recursive algorithm (9)–(17) is equivalent to the Riccati difference equation and the related gain equation:

$$P_{k+1} = A^{\mathrm{T}} P_k A + C^{\mathrm{T}} C - A^{\mathrm{T}} P_k B (B^{\mathrm{T}} P_k B + D^{\mathrm{T}} D)^{-1} B^{\mathrm{T}} P_k A \tag{19}$$

$$G_{k+1} = -(B^{\mathrm{T}} P_k B + D^{\mathrm{T}} D)^{-1} B^{\mathrm{T}} P_k A. \tag{20}$$

It is well known that the above equations give the infinite-horizon LQ gain as $k \to \infty$ under Assumption 1. Therefore, the recursive algorithm leads to the LQ gain provided that $H_k \equiv 0$ [18]. Unfortunately this algorithm ($H_k \equiv 0$) is not necessarily practical for actual systems with non-linearity as mentioned in Sections 6.1 and 6.2. The matrix H_k is a countermeasure.

4.2. Convergence of gain

It becomes clear that the proof of the convergence of gain without assuming $H_k \equiv 0$ is quite important. It is asserted by Corollary 1 that (12)–(17) express an extension of the well-known Riccati difference equation and the gain equation. Since such an extension has not been discussed, it requires new progress in analysis. In fact, the convergence efficiency of G_k is inferior to the Riccati difference equation for ideal systems. We obtain the following result (the derivation is proposed in Appendix E):

Theorem 3

Suppose Assumptions 1 and 2. Let P_k and G_k be given by (12)–(17). Then they converge at the solution $P^* \geqslant 0$ of the algebraic Riccati equation and the time-invariant LQ gain G^*.

Since we assume the ideal linear system, this analysis has the same level as basic results of direct methods of adaptive control. This relation immediately gives the following important result for the model-free design:

Corollary 2

Suppose Assumptions 1 and 2. Let G_k be given by the recursive algorithm (9)–(17). Then, G_k converges at G^* as $k \to \infty$.

Remark 7

This result guarantees the convergence without assuming the matching condition, positive-real condition etc., that are necessary for Landau's scheme or direct methods of adaptive control. We can apply it to MIMO systems where the numbers of inputs and outputs are different.

Remark 8

If $(A + BG_0)$ is unstable, the matrix P_k once diverges until $(A + BG_k)$ becomes stable. After it is stabilized, P_k rapidly decreases. If the stabilizing of $(A + BG_k)$ is not easy, P_k and therefore H_k may become extremely large before stabilization. This often causes numerical errors. We avoid this as follows. Specify an upper limit $\zeta_{\max}$ so that $\zeta_{\max} \geqslant \mathrm{tr}(P^*)$. Replace P_{k+1} with $\{\zeta_{\max}/\mathrm{tr}(P_{k+1})\}P_{k+1}$ provided that $\mathrm{tr}(P_{k+1}) > \zeta_{\max}$.

Remark 9

The convergence rate is lowest if $\lambda_{1k} \equiv \lambda_{2k} \equiv 1$. In this case, H_k increases almost linearly as $k \to \infty$. It is shown in Appendix E that $\tilde{G}_k^{\mathrm{T}} H_k \tilde{G}_k \to 0$ where $\tilde{G}_k = G_k - G^*$. It follows that the lowest convergence rate of $\|G_k - G^*\|$ is, at least, $(1/k)^{1/2}$.

Remark 10

In order to optimize gain for stable but non-controllable modes, the state must be reset frequently to these modes so that the state is sufficiently rich. However, this process is unnecessary in many cases. In fact, the performance index is almost the same for sufficiently large k even if G_k is not optimized for unexcited modes.

5. NUMERICAL CALCULATION

The recursive algorithm includes matrix inversion and products of matrices. A direct calculation of (10) requires $O((n+m)^3)$ multiplications at the end of each T_k. Instead of the calculation, the following iterative signal processing for the single input system ($m=1$) requires $O((n+1)^2)$ multiplications for each sampling time, and this is similar to standard adaptive schemes.

Let $\Psi(t)$, $\Upsilon(t)$, $\Xi(t)$ and $\hat{\Gamma}(t)$ be $(l, n+m)$, $(n, n+m)$, $(n+m, n+m)$ and $(n+m, n+m)$ matrices, respectively. Their initial values are

$$\Psi(t_k) = 0, \quad \Upsilon(t_k) = 0, \quad \Xi(t_k) = \frac{1}{\delta} I, \quad \hat{\Gamma}(t_k) = \hat{\Gamma}_{k0} \tag{21}$$

for each k. Their numerical calculation at each $t \in T_k$ is (see Appendix F)

$$\Psi(t+1) = \Psi(t) + z(t)\xi(t)^{\mathrm{T}} \tag{22}$$

$$\Upsilon(t+1) = \Upsilon(t) + P_k x(t+1)\xi(t)^{\mathrm{T}} \tag{23}$$

$$\Xi(t+1) = \Xi(t) - \{\Xi(t)\xi(t)\}\{\xi(t)^{\mathrm{T}}\Xi(t)\xi(t)+1\}^{-1}\{\xi(t)^{\mathrm{T}}\Xi(t)\} \tag{24}$$

$$\begin{aligned}\hat{\Gamma}(t+1) = \hat{\Gamma}(t) &\\ &+ \Xi(t+1)\xi(t)[\{z(t)^{\mathrm{T}}\Psi(t) + x(t+1)^{\mathrm{T}}\Upsilon(t)\}\Xi(t+1)\\ &- \xi(t)^{\mathrm{T}}\hat{\Gamma}(t)] + [\Xi(t+1)\{\Psi(t)^{\mathrm{T}}z(t)\\ &+ \Upsilon(t)^{\mathrm{T}}x(t+1)\} - \hat{\Gamma}(t)\xi(t)]\xi(t)^{\mathrm{T}}\Xi(t+1)\\ &+ \Xi(t+1)\xi(t)[z(t)^{\mathrm{T}}z(t) + x(t+1)^{\mathrm{T}}P_k x(t+1)\\ &+ \xi(t)^{\mathrm{T}}\hat{\Gamma}(t)\xi(t)]\xi(t)^{\mathrm{T}}\Xi(t+1).\end{aligned} \tag{25}$$

We obtain the estimate $\hat{\Gamma}_k$ as the final value

$$\hat{\Gamma}_k = \hat{\Gamma}(t_{\mathrm{F}k}+1) \tag{26}$$

where $t_{\mathrm{F}k}$ is the final sampling time on each T_k. The total calculation is decreased at each sampling time because they have only products of vectors with vectors or matrices.

6. APPLICATIONS

6.1. Basic numerical simulation

For the simple example (8), the recursive algorithm gives the gain change shown in Figure 1(a) where $P_0 = G_0 = H_0 = 0$, $E(v(t)^2) = 1$, $N_k \equiv 2$ and $\delta = 0.000001$. The sequence of gains converges at the true LQ gain $G^* = -0.5$ regardless of λ_{1k} and λ_{2k}. This result agrees with the convergence analysis. The convergence speed is fastest when $\lambda_{1k} \equiv \lambda_{2k} \equiv 0$. Figure 1(b) shows the change of the same gain provided that the system equation (8) has an unknown system noise $w(t)$ such that $E(w(t)) = 0$ and $\sqrt{E(w(t)^2)} = 0.1$. The noise is used to describe uncertain behaviour in the corresponding physical system, and the result is different from analysis. The effect of the noise is smaller if λ_{1k} and λ_{2k} are larger, and G_k converges only if $\lambda_{1k} \equiv \lambda_{2k} \equiv 1$ in the figure. This result demonstrates the usefulness of H_k.

6.2. Stabilization of a physical inverted pendulum

As an application to stabilizing control, the recursive algorithm was applied to the control of a physical inverted pendulum. The system was composed of a pendulum hinged to the outer end of a rotary arm. A motor turns the arm horizontally around the central shaft of the arm. The friction between the arm and the pendulum was so small that the pendulum maintained standing only by the motion of the arm.

Let θ be the angle of the pendulum and let ϕ be the angle of the arm. The aim of this control is to stabilize the pendulum at the vertical position ($\theta = 0$) on a datum point ($\phi = 0$). This is a single-input-double-output control problem that is unfit for direct methods of adaptive control.

The input was the motor voltage given by a zero-order holder, and the outputs were $\theta(t)$ and $\phi(t)$ given by encoders. The input u was feedback of the state $x(t) = (\theta(t), \theta^{(1)}(t), \phi(t), \phi^{(1)}(t), u(t-\Delta t))^{\mathrm{T}}$ where $u(t-\Delta t)$ was to cope with delays in the pendulum system, and derivatives

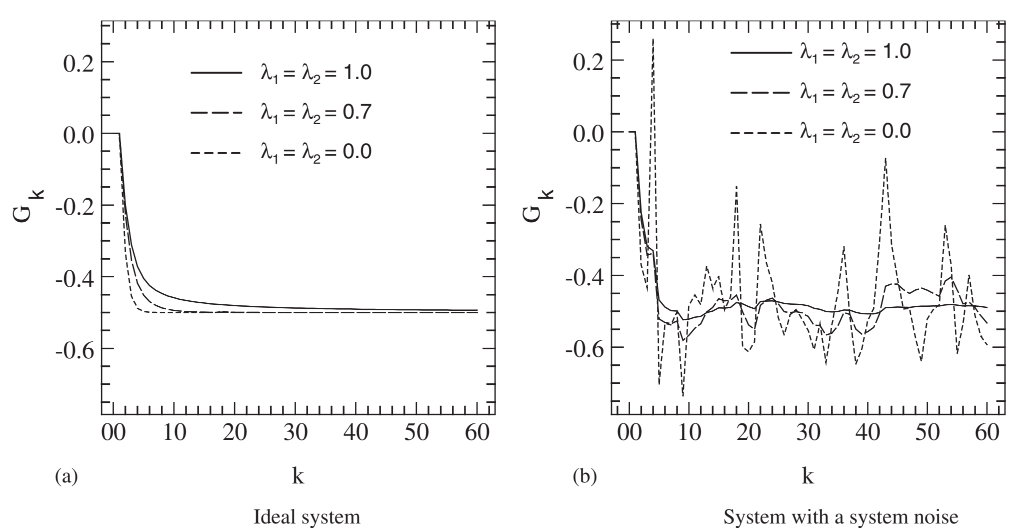

Figure 1. Gain changes of simple systems. (a) Ideal system, (b) system with a system noise.

Table I. Parameters of algorithm.

$C^{\mathrm{T}}C = \mathrm{diag}\ (25\ 0\ 9\ 0\ 0)$,	$D^{\mathrm{T}}D = 0.04$
$P_0 = \mathrm{diag}\ (0\ 0\ 0\ 0\ 0)$,	$G_0 = (0\ 0\ 0\ 0\ 0)$
$H_0 = 0$, $\lambda_{1k} \equiv 0.97$,	$\lambda_{2k} \equiv 0.5$
$\Delta t = 50$ ms,	$N_k \equiv 20$ (G_k changes at every 1.0 s)
$\delta = 10^{-6}$,	$v(t)$: random numbers less than ± 1 V

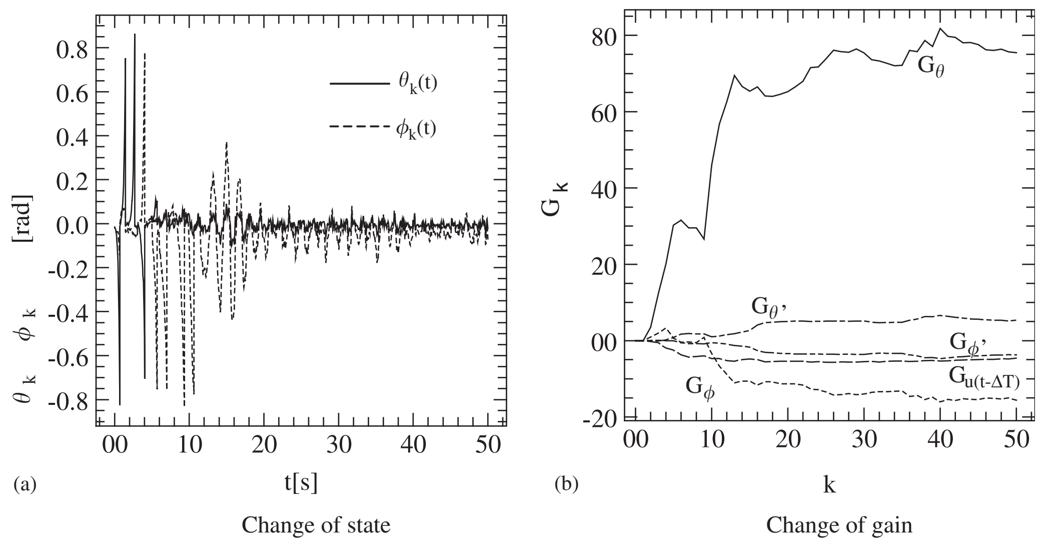

Figure 2. Learning control of a physical pendulum system. (a) Change of the state, (b) change of the gain.

were given by backward difference approximation. The recursive algorithm was calculated assuming that physical parameters are unknown.

The algorithm could not obtain a stabilizing gain by the Newton method ($H_k \equiv 0$) because of uncertain behaviour in the physical system. The stabilization was always accomplished, for example, under the parameters listed in Table I. The long Δt was to distinguish $x(t + \Delta t)$ from $x(t)$.

Figure 2(a) shows a typical experimental result of $\theta(t)$ and $\phi(t)$ through the optimization process. If one of the angles reached ± 0.75 rad, the calculation was automatically stopped and it was continued after a manual reset of the state (refer to Remark 3). Since the original system was unstable ($G_0 = 0$), the pendulum repeatedly fell down at the beginning. The figure excludes the periods used for manual resets. It illustrates that the pendulum fell down at about 0.7, 1.45 and 2.7 s. After that, the arm swung to boundaries at 4.05, 5.7, 7.0, 9.35, and 10.65 s. We see that stabilization of the pendulum caused the stabilization of the arm. Finally, the angles never reached ± 0.75 rad, and the system maintained inverted standing without a reset. A stabilizing gain was obtained by data observed within 16 s ($k = 16$) in total. This convergence rate was almost the average. Vibration of the arm remained because of the input noise $v(t)$.

Figure 2(b) shows the corresponding change in the feedback gain. The subscribed letters on G express the gain elements. By the way, the Riccati equation of the continuous-time model gave the LQ gain $(+21.5 + 2.68 - 3.35 - 2.09)$ in the model-based design. Similarly, the equation of the discrete-time model gave $(+26.3 + 3.29 - 5.23 - 2.64)$. The inverted standing could not maintained for these gains. It was mainly caused by neglecting friction of the motor and the gear in the model.

6.3. Steady state error-less control of a physical servo

The classical integral-type optimal servo [24, 25] was examined for data-based design because of the simple structure. The device was a physical plant comprises three rotary disks combined with a flexible shaft as shown in Figure 3. It is usually a sixth order system, whose state variables are angles $\theta_i(t)$ $(i = 1, 2, 3)$ and the angular velocities. The input was the current of a motor attached to disk 1 and the output was the angle of disk 3. Figure 4 describes the pulse response θ_3 at $t = 0$ of the open-loop system. It took about 30 s to come to a standstill.

Let $e(t) = C_c x(t) - r = \theta_3(t) - r$ be the error of the servo system where r was a constant target value. The integral-type optimal control is as follows. We transform the state space model as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x^{(1)}(t) \\ e(t) \end{pmatrix} = \begin{pmatrix} A_c & 0 \\ C_c & 0 \end{pmatrix}\begin{pmatrix} x^{(1)}(t) \\ e(t) \end{pmatrix} + \begin{pmatrix} B_c \\ 0 \end{pmatrix} u^{(1)}(t) \tag{27}$$

with the performance index

$$V_\infty = \sum_{t=0}^{\infty} e(t)^2 + Ru^{(1)}(t)^2$$

by differentiating the plant equation $\mathrm{d}x(t)/\mathrm{d}t = A_c x(t) + B_c u(t)$ where $x^{(1)}(t)$ and $u^{(1)}(t)$ are the derivatives. We can expect the steady state error-less operation ($e(t) = C_c x(t) - r \to 0$ under a constant r) because the signals were derivatives in the LQ design. The usual input $u(t)$ for optimization was given by integrating the above state feedback $u^{(1)}(t)$ with $v(t)$ as $u(t) = G_{k1}x(t) + G_{k2}e^{(-1)}(t) + v^{(-1)}(t)$.

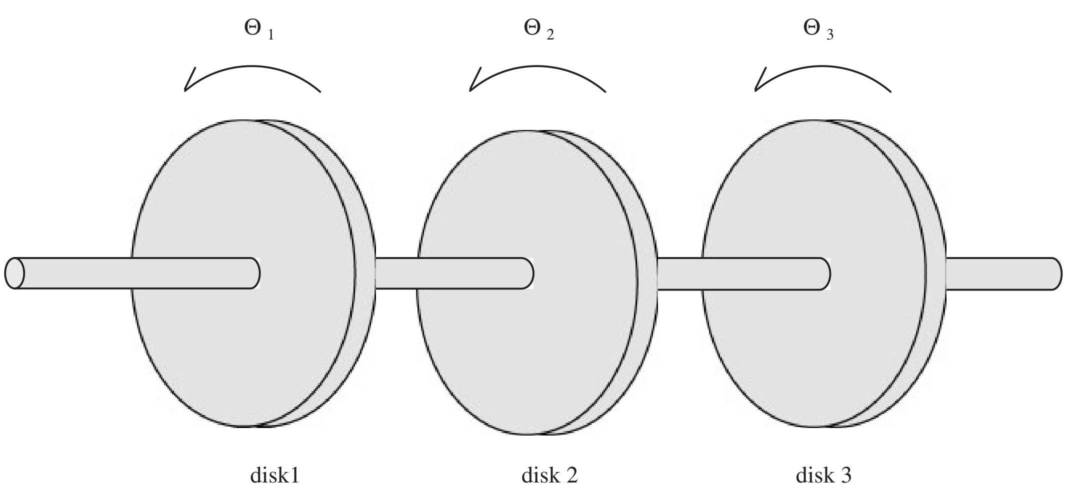

Figure 3. Experimental device. $Y = \theta_3$ (rad).

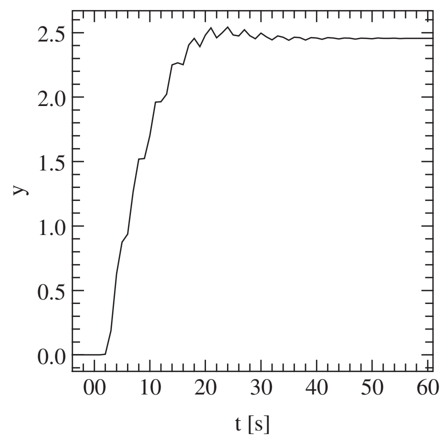

Figure 4. Impulse response of the open-loop system. $Y = \theta_3$ (rad).

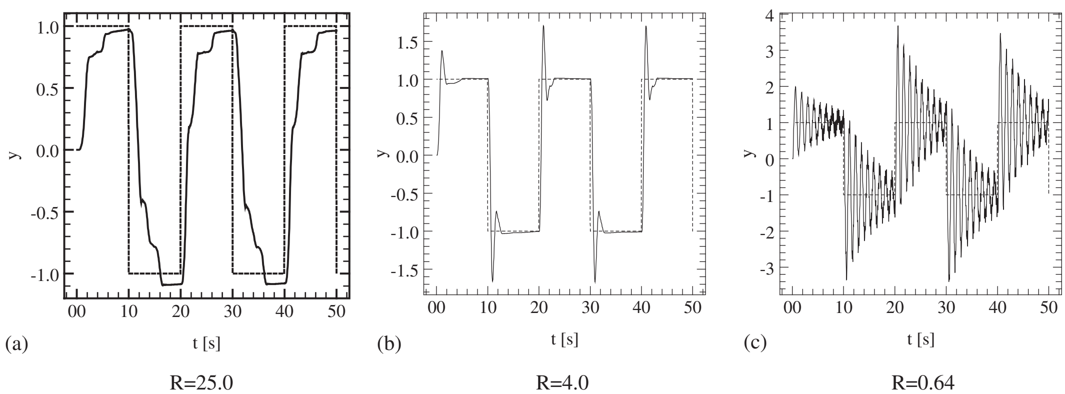

Figure 5. Optimized responses. $Y = \theta_3$ (rad). (a) $R = 25.0$, (b) $R = 4.0$, (c) $R = 0.64$.

The recursive algorithm used the provisional signals in (27) under the assumption that system parameters were unknown where R was fixed individually. Three encoders gave sampled data of the angles, and derivatives were given by backward difference approximation. The additional signal $v(t)$ was a difference of a random signal in order that $v^{(-1)}(t)$ may not become too large. The target signal r was a rectangular wave, and data pairs for which r changed the value between t and $t + 1$ were excluded from the calculation. Other parameters of the algorithm were basically the same as Table I. Since G_0 and P_0 were 0 for each optimization, G_k sometimes made the system unstable, and a reset of the state was required. Thus we really obtained the optimal control over a wide range of R.

Figure 5 illustrates experimental results of the optimized responses $\theta_3(t)$ as a function of the weight R. The signal $v(t)$ was removed in the responses for simple understanding. The dotted lines are the rectangular target $r(t) = \pm 1$ with the period 20 s. It seems that these responses fulfilled the steady state position error-less operation if r is constant permanently. Approaching level of (a) is low because of the large R. Conversely, the response in (c) is rapid but it contains the high frequency overshoot. By comparing them, we see that (b) $R = 4$ is a good value.

7. CONCLUSION

The theoretical analysis and experimental results show that the proposed model-free approach is simple and promising for practical controller design. The convergence analysis guarantees that the stabilizability and detectability are sufficient even for unstable multi-input-multi-output systems. It is desired that these results contribute to development of learning or adaptive LQ control design. Although the study is restricted to the basic problem, various alteration may be possible. Some modification is effective to obtain a suitable sampling period for many practical applications.

Extension of these results to output feedback is quite interesting because the standard observer is difficult to construct for unknown systems. Since it has been shown that we can transform the output feedback control problem of an unknown nth-order system into a known $2n$th-order state feedback problem [2] (English), the recursive algorithm seems applicable to the output feedback control of unknown systems as well.

The theoretical derivation is mainly given for ideal linear systems. It is a future study to compare this approach with the standard LQ design as to non-linearity, etc.

ACKNOWLEDGEMENTS

The authors want to express their appreciation to many students who partly joined the related work.

Appendix A: Proof of Lemma 1

Refer to (11). We rewrite (1), (2) and $u(t) = Gx(t) + v(t)$ as

$$x(t+1) = (A + BGB)\xi(t) \tag{A1}$$

$$z(t) = \begin{pmatrix} C & 0 \\ DG & D \end{pmatrix} \xi(t). \tag{A2}$$

Substitute these relations, $t = 0$, $v^a(0) = 0$ for $\mathbf{z}^a$ and $x^b(0) = 0$ for $\mathbf{z}^b$ into (7). Refer to (C1) because the right-hand sides are the same except $\xi(0)^{\mathrm{T}}$ and $\xi(0)$. Then we obtain the relations of the lemma as respective elements of the matrix because the relation holds for any $x^a(0)$ and $v^b(0)$. □

Appendix B: Proof of Theorem 1

Though (6) implies Theorem 1, we prove it from the algebraic Riccati equation. Under Assumption 1, it is well known that there exists the LQ optimal control with the gain G if and only if $P \geqslant 0$ satisfies the algebraic Riccati equation and the gain equation. They are obtained by replacing P_{k+1} and P_k with P and G_k with G, respectively, in (19) and (20). Since we use the latter to rewrite the last term of the former as $A^T PBG + G^T B^T PA + G^T(B^T PB + D^T D)G$, we rewrite the former as

$$P = (A + BG)^T P(A + BG) + C^T C + G^T D^T DG. \tag{B1}$$

Multiply both sides by $x^a(0)^T$ from the left and by $x^a(0)$ from the right. Substitute $(A + BG)x^a(0) = x^a(1)$ and $(C^T \; G^T D^T)^T x^a(0) = z^a(0)$, and refer to (7). Then, the multiplied (B1) agrees with the first equality of the theorem.

Multiply both sides of the gain equation by $(B^T PB + D^T D)$, and transpose the right-hand side to the left. Then we can rewrite it as

$$B^T P(A + BG) + D^T DG = 0. \tag{B2}$$

Multiply this equation by $v^b(0)^T$ from the left-hand side and by $x^a(0)$ from the right-hand side. Refer to the definition of $z^b(t)$. Then, the multiplied (B2) agrees with the second equality of the theorem.

In other words, the equalities of Theorem 1 are equivalent to the algebraic Riccati equation and the gain equation. □

Appendix C: Proof of Lemma 2

Assumption 2 implies Lemma 1. Using Lemma 1, we see that

$$\Gamma = \begin{pmatrix} (A + BG_k)^T \\ B^T \end{pmatrix} P_k (A + BG_k \;\; B) + \begin{pmatrix} C^T & G_k^T D^T \\ 0 & D^T \end{pmatrix} \begin{pmatrix} C & 0 \\ DG_k & D \end{pmatrix}. \tag{C1}$$

Multiply both sides by $\xi(t)\xi(t)^T$ from the left and by $\xi(s)\xi(s)^T$ from the right. Substitute (A1) and (A2). Then we have

$$\xi(t)\xi(t)^T \Gamma \xi(s)\xi(s)^T = \xi(t)\{z(t)^T z(s) + x(t+1)^T P_k x(s+1)\}\xi(s)^T. \tag{C2}$$

Add both sides with respect to t and s. Then Ξ_k offsets the sum of $\xi\xi^T$ in the left-hand side. We obtain the lemma because the left-hand side agrees with $\hat{\Gamma}_k$ by neglecting the small terms with δ. □

Appendix D: Proof of Corollary 1

Let $H_k \equiv 0$. Note the relation $\hat{\Gamma}_k^{ab} = A^T P_k B + G_k^T \hat{\Gamma}_k^{bb}$. Take

$$G_{k+1} = G_k - (\hat{\Gamma}_k^{bb})^{-1}(B^T P_k A + \hat{\Gamma}_k^{bb} G_k) \tag{D1}$$

$$P_{k+1} = \hat{\Gamma}_k^{aa} - (A^T P_k B + G_k^T \hat{\Gamma}_k^{bb})(\hat{\Gamma}_k^{bb})^{-1}(B^T P_k A + \hat{\Gamma}_k^{bb} G_k) \tag{D2}$$

from the second equalities of (13) and (14). Referring to Lemma 1, we can cancel all G_k by some terms in $\hat{\Gamma}_k^{aa}$, etc. We thus obtain the equations of the corollary. □

Appendix E: Proof of Theorem 3

The basic idea of the following is in common with [22, 23] as mentioned in Remark 6 by replacing the system equation with the Riccati-like equation.

We rewrite (13) and (14) in terms of G_{k+1} as follows. Since Assumption 2 implies (C1), substitute (C1) and $G_k = G_{k+1} - \Delta G_k$ into the first equalities of them. Multiply both sides of the latter by $(B^{\mathrm{T}} P_k B + D^{\mathrm{T}} D + H_k)$. Then we can cancel all ΔG_k except $H_k \Delta G_k$ of the latter, and we obtain,

$$\begin{aligned} P_{k+1} = &(A + BG_{k+1})^{\mathrm{T}} P_k (A + BG_{k+1}) \\ &+ C^{\mathrm{T}} C + G_{k+1}^{\mathrm{T}} D^{\mathrm{T}} D G_{k+1} \end{aligned} \tag{E1}$$

$$H_k \Delta G_k = -\{B^{\mathrm{T}} P_k (A + BG_{k+1}) + D^{\mathrm{T}} D G_{k+1}\}. \tag{E2}$$

Now, define the error terms $\tilde{P}_k = P_k - P^*$ and $\tilde{G}_k = G_k - G^*$. Since (B1) and (B2) are equivalent to the algebraic Riccati equation, we can rewrite P as P^* and G as G^* in them. Consider the difference between (E1) and the rewritten (B1) and the difference between (E2) and rewritten (B2). Let $A^* = A + BG^*$. Substitute $A + BG_{k+1} = A^* + B\tilde{G}_{k+1}$ into the differences. We thus obtain the error equation

$$\begin{aligned} \tilde{P}_{k+1} &= A^{*\mathrm{T}} \tilde{P}_k A^* + g_k \tilde{G}_{k+1} + \tilde{G}_{k+1}^{\mathrm{T}} g_k^{\mathrm{T}} + \tilde{G}_{k+1}^{\mathrm{T}} h_k \tilde{G}_{k+1} \\ &= (A^* + B\tilde{G}_{k+1})^{\mathrm{T}} \tilde{P}_k (A^* + B\tilde{G}_{k+1}) \\ &\quad + \tilde{G}_{k+1}^{\mathrm{T}} h^* \tilde{G}_{k+1} \end{aligned} \tag{E3}$$

$$H_k \Delta G_k = -g_k^{\mathrm{T}} - h_k \tilde{G}_{k+1} \tag{E4}$$

where

$$\begin{aligned} g_k &= A^{*\mathrm{T}} P_k B + G^* D^{\mathrm{T}} D = A^{*\mathrm{T}} \tilde{P}_k B \\ h_k &= B^{\mathrm{T}} P_k B + D^{\mathrm{T}} D = B^{\mathrm{T}} \tilde{P}_k B + h^*, \\ h^* &= B^{\mathrm{T}} P^* B + D^{\mathrm{T}} D. \end{aligned}$$

We prove the convergence by using a Lyapunov function. Since $(A^*)^{\mathrm{T}}$ is a stable matrix under Assumption 1, there exist two regular matrices M and N such that the matrix Lyapunov equation

$$A^* M M^{\mathrm{T}} A^{*\mathrm{T}} - M M^{\mathrm{T}} = -N N^{\mathrm{T}} \tag{E5}$$

holds. Define

$$J_k = \mathrm{tr}\{M^{\mathrm{T}} (\tilde{P}_k + \tilde{G}_k^{\mathrm{T}} H_k \tilde{G}_k) M\}. \tag{E6}$$

We know that J_k is non-negative. Then change of J_k is given by

$$\begin{aligned} J_{k+1} - J_k = \mathrm{tr}\{M^{\mathrm{T}}(\tilde{P}_{k+1} - \tilde{P}_k \\ + \Delta G_k^{\mathrm{T}} H_k \tilde{G}_{k+1} + \tilde{G}_{k+1}^{\mathrm{T}} H_k \Delta G_k \\ + \tilde{G}_{k+1}^{\mathrm{T}}(H_{k+1} - H_k)\tilde{G}_{k+1} - \Delta G_k^{\mathrm{T}} H_k \Delta G_k)M\} \end{aligned} \tag{E7}$$

by substituting $\tilde{G}_k = \tilde{G}_{k+1} - \Delta G_k$. Substitute (E3) and (E4) into (E7) except the last term. We can cancel non-symmetric terms $g_k \tilde{G}_{k+1}$ and $\tilde{G}_{k+1}^{\mathrm{T}} g_k^{\mathrm{T}}$ as

$$\begin{aligned} J_{k+1} - J_k = \mathrm{tr}\{M^{\mathrm{T}}(A^{*\mathrm{T}} \tilde{P}_k A^* - \tilde{P}_k)M\} \\ - \mathrm{tr}\{M^{\mathrm{T}}(\tilde{G}_{k+1}^{\mathrm{T}}(h_k - H_{k+1} + H_k)\tilde{G}_{k+1} \\ + \Delta G_k^{\mathrm{T}} H_k \Delta G_k)M\}. \end{aligned} \tag{E8}$$

Since (15) and (16) mean $h_k - H_{k+1} + H_k \geqslant 0$, it follows that

$$J_{k+1} - J_k \leqslant \mathrm{tr}\{M^{\mathrm{T}}(A^{*\mathrm{T}} \tilde{P}_k A^* - \tilde{P}_k)M\}. \tag{E9}$$

Using the commutativity of matrices in the trace, we can rewrite the right-hand side as $\mathrm{tr}\{(A^* M M^{\mathrm{T}} A^{*\mathrm{T}} - M M^{\mathrm{T}})\tilde{P}_k\}$. Then (E5) implies

$$J_{k+1} - J_k \leqslant -\,\mathrm{tr}(N^{\mathrm{T}} \tilde{P}_k N). \tag{E10}$$

At this point, we first prove the theorem under the assumption that $P_0 \geqslant P^*$, that is $\tilde{P}_0 \geqslant 0$. Then, the second equality of (E3) guarantees that $\tilde{P}_k \geqslant 0$ for any $k = 0, 1, 2, \ldots$. Therefore, (E10) shows that J_k is monotone non-increasing. In addition, J_k is non-negative. As is known according to Weierstrass's theorem, the bounded monotone sequence J_k converges at a certain value. Then, $J_{k+1} - J_k \to 0$ and $\tilde{P}_k \to 0$ because N is regular. This convergence implies that $\tilde{G}_{k+1}^{\mathrm{T}} h^* \tilde{G}_{k+1} \to 0$ in (E3). From the observability of Assumption 1, we can conclude that $\tilde{G}_k \to 0$, namely $G_k \to G^*$.

Next is the study of the convergence without the assumption that $P_0 \geqslant P^*$. Necessity of this case distinguishes this problem from the usual adaptive control. Let $P_k^{\#}$ be the solution of the standard Riccati difference equation with the initial condition $P_0^{\#} = 0$. Let $\tilde{P}_k^{\#} = P_k^{\#} - P^*$. Let

$$\zeta_k = -\sum_{i=k}^{\infty} \mathrm{tr}(N^{\mathrm{T}} \tilde{P}_i^{\#} N). \tag{E11}$$

Let λ_i^* denote each eigenvalue of A^*. Since the standard solution $P_k^{\#}$ converges at P^* exponentially with $O(\max_i |\lambda_i^*|^k)$ [26], the infinite series in (E11) has the sum.

Since J_k is not a Lyapunov function yet, we use

$$\bar{J}_k = J_k + \zeta_k \tag{E12}$$

instead of J_k. Then, $\zeta_{k+1} - \zeta_k = \mathrm{tr}(N^{\mathrm{T}} \tilde{P}_k^{\#} N)$. By substituting these relations and (E10), it follows that

$$\bar{J}_{k+1} - \bar{J}_k \leqslant -\,\mathrm{tr}\{N^{\mathrm{T}}(\tilde{P}_k - \tilde{P}_k^{\#})N\}. \tag{E13}$$

From the definition, $\tilde{P}_k^{\#}$ is the minimum of $\tilde{P}_k$ for any possible control and for any $P_0 \geqslant 0$. Therefore, $\tilde{P}_k - \tilde{P}_k^{\#} \geqslant 0$. Since $\bar{J}_k$ is bounded below, (E13) plays the same role as (E10). Consequently, $\bar{J}_k$ converges, $\bar{J}_{k+1} - \bar{J}_k \to 0$ and $\tilde{P}_k - \tilde{P}_k^{\#} \to 0$. Then, $\tilde{P}_k \to 0$ because $\tilde{P}_k^{\#} \to 0$.

The rest of this proof is the same as the case $P_k \geqslant P^*$. Finally, we can conclude that $\tilde{G}_k \to 0$ without the assumption that $P_0 \geqslant P^*$. □

Appendix F: Calculation of $\hat{\Gamma}_k$

Let $t \in T_k$. Define parameters at each t as

$$\Psi(t) = \sum_{\tau \leqslant t-1 \cap \tau \in T_k} z(\tau)\xi(\tau)^{\mathrm{T}} \tag{F1}$$

$$\Upsilon(t) = \sum_{\tau \leqslant t-1 \cap \tau \in T_k} P_k x(\tau+1)\xi(\tau)^{\mathrm{T}} \tag{F2}$$

$$\Xi(t) = \left\{ \sum_{\tau \leqslant t-1 \cap \tau \in T_k} \xi(\tau)\xi(\tau)^{\mathrm{T}} + \delta I \right\}^{-1} \tag{F3}$$

$$\begin{aligned} \Theta(t) = \Bigg\{ & \sum_{\tau \leqslant t-1 \cap \tau \in T_k} \sum_{\sigma \leqslant t-1 \cap \sigma \in T_k} \xi(\tau)\{z(\tau)^{\mathrm{T}} z(\sigma) \\ & + x(\tau+1)^{\mathrm{T}} P_k x(\sigma+1)\}\xi(\sigma)^{\mathrm{T}} \Bigg\} + \delta^2 \hat{\Gamma}_{k0}, \end{aligned} \tag{F4}$$

$$\hat{\Gamma}(t) = \Xi(t)\Theta(t)\Xi(t). \tag{F5}$$

Note that $\Theta(t) = \Xi(t)^{-1}\hat{\Gamma}(t)\Xi(t)^{-1}$ and $\Xi(t)^{-1} = \Xi(t+1)^{-1} - \xi(t)\xi(t)^{\mathrm{T}}$. Thus, we have

$$\Theta(t+1) = \Theta(t) + \Delta\Theta(t) \tag{F6}$$

where

$$\begin{aligned} \Theta(t) = {} & \Xi(t+1)^{-1}\hat{\Gamma}(t)\Xi(t+1)^{-1} \\ & - \xi(t)\xi(t)^{\mathrm{T}}\hat{\Gamma}(t)\Xi(t+1)^{-1} - \Xi(t+1)^{-1}\hat{\Gamma}(t)\xi(t)\xi(t)^{\mathrm{T}} \\ & + \xi(t)\xi(t)^{\mathrm{T}}\hat{\Gamma}(t)\xi(t)\xi(t)^{\mathrm{T}} \end{aligned} \tag{F7}$$

$$\begin{aligned} \Delta\Theta(t) = {} & \xi(t)\{z(t)^{\mathrm{T}}\Psi(t) + x(t+1)^{\mathrm{T}}\Upsilon(t)\} \\ & + \{\Psi(t)^{\mathrm{T}} z(t) + \Upsilon(t)^{\mathrm{T}} x(t+1)\}\xi(t)^{\mathrm{T}} \\ & + \xi(t)\{z(t)^{\mathrm{T}} z(t) + x(t+1)^{\mathrm{T}} P_k x(t+1)\}\xi(t)^{\mathrm{T}}. \end{aligned} \tag{F8}$$

Their definitions give (21), (22), (23) and (26). The matrix inversion lemma [27] gives (24). Multiply (F6) by $\Xi(t+1)$ from both sides and substitute (F7) and(F8). Considering the relation between $\Theta(t)$ and $\hat{\Gamma}(t)$, we have (25) and (26). □.

REFERENCES

1. Kawamura Y. A basic algorithm of constructing the optimal regulator from input-output data. *Transactions of SICE* 1988; **24**:1216–1218 (Japanese).
2. Kawamura Y. Direct construction of LQ regulator base on orthogonalization of signals: dynamical output feedback. *Systems & Control Letters* 1998; **34**:1–9.

3. Furuta K, Wongsaisuwan M. Closed-form solution to discrete-time LQ optimal control and disturbance attenuation. *Systems & Control Letters* 1993; **20**:427–437.
4. Furuta K, Wongsaisuwan M. Discrete-time LQG dynamic controller design using plant Markov parameters. *Automatica* 1995; **38**:1325–1332.
5. Chan JT. Data-based synthesis of multivariable LQ regulator. *Automatica* 1996; **32**:403–407.
6. Chan JT. Output feedback realization of LQ optimal systems: general design using data-based synthesis. *International Journal of Control* 1999; **72**:1054–1064.
7. Fujisaki Y, Duan Y, Ikeda M, Fukuda M. A system representation and optimal control in input-output data space. *Transactions of SICE* 1999; **34**:1845–1853 (Japanese).
8. Hjalmarsson H, Gunnarsson S, Gevers M. Data-based tuning of a robust regulator for a flexible transmission system. *European Journal of Control* 1995; **1**:148–156.
9. Hjalmarsson H, Gevers M, Lequin O. Iterative feedback tuning: theory and applications in chemical process control. *Journal A* 1997; **38**:16–25.
10. Narendra KS, McBride LE. Multiparameter self-optimization system using correlation techniques. *IEEE Transactions on Automatic Control* 1964; **AC-9**:31–38.
11. Narendra KS, Streeter DN. An adaptive procedure for controlling undefined linear processes. *IEEE Transactions on Automatic Control* 1964; **AC-9**:545–548.
12. Kawamura Y. Duality on orthogonality conditions for discrete-time optimal control and optimal estimation. *Transactions of SICE* 1988; **24**:1360–1367 (Japanese).
13. Goodwin GC, Sin K. *Adaptive Filtering Prediction and Control*. Prentice-Hall: Englewood Cliffs, NJ, 1984.
14. Åström KJ, Wittenmark B. *Computer Controlled Systems*. Prentice-Hall: Englewood Cliffs, NJ, 1984.
15. Arimoto S, Kawamura S, Miyazaki F. Bettering operation of robots by learning. *Journal of Robotic Systems* 1984: **1**:123–140.
16. Xu J-X. *Iterative Learning Control*. Kluwer Academic: U.S.A., 1998.
17. Kailath T. An innovations approach to least-squares estimation part 1: linear filtering in additive noise. *IEEE Transactions on Automatic Control* 1963; **AC-13**:646–655.
18. Kawamura Y. Extension of the Riccati difference equation from robust gradient method. *Transactions of SICE* 1997; **33**:103–108 (Japanese).
19. Kawamura Y. Consideration on a fast learning scheme for the LQ optimal regulator. *Transactions of SICE* 1993; **29**:767–775 (Japanese).
20. Clarke DW, Gawthrop PJ. Self-tuning control. *Proceedings of the IEE* 1979; **126**-D:633–640.
21. Landau ID, Lozano R. Unification of discrete time explicit model reference adaptive control design. *Automatica* 1981; **17**:593–611.
22. Lin Y-H, Narendra KS. A new error model for adaptive systems. *IEEE Transactions on Automatic Control* 1980; **AC-25**:585–587.
23. Suzuki T, Sinnaka S, Tanaka K. Simple proof of the generalized adaptive law and some properties of the constant-trace algorithm. *Transactions of SICE* 1982; **18**:1074–1079 (Japanese).
24. Johnson CD. Optimal control of the linear regulator with constant disturbances. *IEEE Transactions Automatic Control* 1968; **AC-15**:416–421.
25. Smith HD, Davison EJ. Design of industrial regulators integral feedback and feedforward control. *Proceedings of the IEE* 1972; **119**:416–421.
26. Anderson BDO, Moore JB. *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, NJ, 1979; 81.
27. Ho YC. On the stochastic approximation method and optimal filtering theory. *Journal of Mathematical Analysis and Applications* 1962; **6**:152–154.

