Informative Input Design for Bayesian Identification of LPV Systems

Yusuke Fujimoto *, Wataru Kasai *, and Toshiharu Sugie *

Abstract: This paper proposes an input design method for identification of linear-parameter-varying (LPV) systems. In particular, this paper focuses on the Bayesian estimation of LPV systems, especially the impulse responses of LPV systems at the scheduling variable of interest. The mutual information is employed as a criterion, and a concrete procedure to obtain the local optimum is given. A numerical example is shown to demonstrate the effectiveness of the proposed input design.

Key Words: system identification, input design, linear-parameter-varying system.

1. Introduction

Most industrial processes have nonlinear dynamics, and it is important to capture their behavior for the controller design. Since the first principle modeling may become difficult for such nonlinear systems, system identification methods would be a good alternative for the modeling scheme [1]. In particular, identification with linear models is important and standard in the context of system control because the controller design for linear models is easier than the one for nonlinear models. Of course it is difficult to give a linear model which works as a good model for any operating points of the nonlinear system. Focusing on the specific operating point is important for the identification with a linear model when the system is nonlinear.

To make the operating point clear, we consider the case where the target system is described by a linear-parameter-varying (LPV) system [2], [3]. An LPV system is defined by a set of linear systems indexed by the so-called scheduling variables. A typical example of LPV system is a chemical plant whose inputs and outputs are amounts of reactants and products, respectively. The input-output relationship of this system is well described by a linear system with the fixed plant temperature. In this case, the temperature of the plant works as the scheduling variable, and the plant becomes a set of linear systems describing each temperature.

This paper focuses on constructing good linear models for the specific scheduling variables (i.e., operating points), under the conditions that i) the experiment is held with the varying scheduling variable, and ii) the available energy assigned to the identification input is limited. Figure 1 illustrates the condition i). The horizontal axis shows time, and the vertical axis shows the scheduling variables (e.g., temperature in a chemical plant). The solid line shows how the scheduling variable varies during the experiment, and the broken line shows the scheduling variable of interest. In this case, the input design problem plays a crucial role since the input energy assigned to the scheduling variable of interest affects the value of I/O data. An example of such a situation is the following; consider the example of chemical plant mentioned above. If the temperature is controlled by ON/OFF of the boiler, it is not easy to keep the temperature at the fixed value, e.g., 60°C. A possible solution is to perform an experiment with varying temperature; by switching off the boiler, the temperature drops from e.g., 100°C to 20°C like the solid line in Fig. 1. The purpose here is to construct a good linear model for 60°C, while the system behaves as a nonlinear system since the scheduling variable is varying.

This paper employs the Bayesian approach for stable LPV systems as an identification scheme [4]. One of the advantages of this approach is that it gives a simple idea to connect the model at the scheduling variables of interest and the I/O data far from such scheduling variables. In addition, [5] combines Kernel-Based system identification (see e.g., [6], [7]) to the work of [4]. Kernel-Based system identification estimates the impulse responses of stable systems in a Bayesian manner. Kernel is a prior covariance matrix of the impulse response, and Pillonetto and De Nicolao showed the way to encode bounded-input-bounded-output (BIBO) stability to this matrix.

This paper proposes an input design method under the above conditions. In particular, we employ the mutual information between the impulse response and the output as a criterion. The mutual information is known to be a good criterion in the Bayesian estimation framework, and several works have been reported [8]–[10]. In particular, [11] applied the mutual information to the standard Kernel-Based system identification, and showed its effectiveness. As far as the authors know, however,
few works applied the mutual information to the LPV system identification, and this is the first attempt on this topic.

Main contributions of this paper are the following.

1. An input design for Bayesian LPV system identification is mathematically formulated.

2. A concrete procedure to obtain the local optimal input is proposed.

3. The effectiveness of the proposed input design method is shown through the numerical simulation.

This work is based on our preliminary conference version [12]. In this journal version, reasons to employ the mutual information are shown, and the concrete procedure to obtain the local optimal input is given. In addition, the target system of the numerical simulation becomes significantly complex compared to the one in [12].

This paper is organized as follows. The problem formulation is described in Sect. 2. Then, Sect. 3 gives a brief introduction for the mutual information and shows an optimization procedure. Section 4 shows a numerical simulation. Section 5 gives some concluding remarks.

[Notations] The transposition of a matrix A is denoted by $A^\top$. For a matrix $A \in \mathbb{R}^{m \times n}$, $\text{Tr}(A)$ and $\text{det}(A)$ represent its trace and determinant, respectively. The identity matrix of size $n$ is denoted by $I_n$. Let $f(x)$ be a matrix-valued function of a scalar $x$, $f : \mathbb{R} \rightarrow \mathbb{R}^{x \times x}$. Then, $\frac{\partial f}{\partial x}$ is an $n \times n$ matrix whose $(l, m)$ element is $\frac{\partial (f(x))_{lm}}{\partial x}$. Also let $g$ be a function of a matrix $X \in \mathbb{R}^{x \times x}$, $g : \mathbb{R}^{x \times x} \rightarrow \mathbb{R}$. Then, $\frac{\partial g}{\partial X}$ is an $n \times n$ matrix whose $(l, m)$ element is $\frac{\partial g(X)}{\partial X_{lm}}$. For a continuous random variable $x$, its probability density function is denoted by $p(x)$. A joint probability density function of two random variables $x$ and $y$ are denoted by $p(x, y)$. $p(x \mid y)$ denotes the conditional probability density function of $x$ with given $y$. When $x \in \mathbb{R}^n$ is a Gaussian random variable whose mean and covariance are given by $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$, we denote $p(x) \sim \mathcal{N}(\mu, \Sigma)$. In this paper, log $x$ is the natural logarithm of $x > 0$, and $e$ denotes the base of the natural logarithm.

### 2. Problem Setting

This paper focuses on constructing the LPV finite-impulse-response (LPV-FIR) model described as

$$y(k) = \sum_{i=0}^{n-1} u(k-i) g_p(i) + w(k), \quad (1)$$

where $k, y(k) \in \mathbb{R}$ and $u(k) \in \mathbb{R}$ denote the time step, the output, and the input, respectively. We denote the scheduling variable at time step $k$ by $p(k) \in \mathbb{R}$, and $g_p$ shows the impulse response of the system with the fixed scheduling variable $p$. The noise at time step $k$ is denoted by $w(k)$, which is assumed to be generated from an independent and identically distributed Gaussian process whose variance is denoted by $\sigma^2$. Let $p^*$ be the scheduling variable of interest. Then, the identification problem of the LPV-FIR model is to estimate $g_p^* = [g_{p^*}(0), \ldots, g_{p^*}(n-1)]^\top$ from the observed data $[u^*(k), y^*(k), p^*(k)]_{k=1}^n$. The superscript $o$ indicates that the value is the one in the experiment.

To construct the LPV-FIR model, we employ the Bayesian approach [4], [5]. We introduce a prior distribution of the impulse responses $g_{p^*}$ and $g_{p^*}(k) = [g_{p^*(k)}(0), \ldots, g_{p^*(k)}(n-1)]^\top (k = 1, \ldots, N)$. In particular, the zero-mean $n(N+1)$ dimensional Gaussian

$$p \left( \begin{bmatrix} g_{p^*} \\ g_{p^*(1)} \\ \vdots \\ g_{p^*(N)} \end{bmatrix} \right) \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} K_{p^*p^*} & K_{p^*p^*} \top \\ K_{p^*p^*} & K_{p^*p^*} \end{bmatrix} \right) \quad (2)$$

is employed as a prior distribution. Here, $K_{p^*p^*} \in \mathbb{R}^{nN \times nN}$ and $K_{p^*p^*} \in \mathbb{R}^{Nn \times Nn}$ are defined by

$$K_{p^*p^*} = \begin{bmatrix} K_{p^*(1)p^*(1)} & \cdots & K_{p^*(1)p^*(N)} \\ \vdots & \ddots & \vdots \\ K_{p^*(N)p^*(1)} & \cdots & K_{p^*(N)p^*(N)} \end{bmatrix}, \quad (3)$$

$$K_{p^*p^*}^\top = \begin{bmatrix} K_{p^*(1)p^*(1)} & \cdots & K_{p^*(1)p^*(N)} \\ \vdots & \ddots & \vdots \\ K_{p^*(N)p^*(1)} & \cdots & K_{p^*(N)p^*(N)} \end{bmatrix}, \quad (4)$$

where $K_{uq3} \in \mathbb{R}^{nN \times N}$ is a positive semidefinite matrix which shows the covariance of impulse responses $g_{q1}$ and $g_{q2}$, and determined by the so-called kernel function $k(q_1, q_2, i_1, i_2, \eta)$ as

$$\begin{bmatrix} K_{q1} \mid i_1, i_2 \end{bmatrix} = k(q_1, q_2, i_1, i_2, \eta), \quad (5)$$

where $\eta$ denotes the parameter of the kernel. Let $U \in \mathbb{R}^{Nn \times n}$ be the matrix defined by

$$U = \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ 0 & u_2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_N \end{bmatrix}, \quad (6)$$

$$u_k = [u(k), \ldots, u(k-n+1)] \in \mathbb{R}^{1 \times n}, \quad (7)$$

with $u(k) = 0$ for $k < 0$. Then, the joint distribution of $g^* = g_{p^*}$ and $Y = [y^*(1), \ldots, y^*(N)]^\top$ becomes

$$p \left( \begin{bmatrix} g^* \\ Y \end{bmatrix} \right) \sim \mathcal{N} \left( \begin{bmatrix} K_{p^*p^*} & K_{p^*p^*} U^\top \\ U K_{p^*p^*} & U K_{p^*p^*} U^\top + \sigma^2 I_N \end{bmatrix} \begin{bmatrix} g^* \\ Y \end{bmatrix}, \quad (8)$$

thus the posterior distribution of $g^*$ is given as

$$p(g^* \mid Y) \sim \mathcal{N} \left( \hat{g}^*, \tilde{K}_{p^*p^*} \right), \quad (9)$$

$$\tilde{K}_{p^*p^*} = K_{p^*p^*} - K_{p^*p^*} U^\top \left( U K_{p^*p^*} U^\top + \sigma^2 I_N \right)^{-1} U K_{p^*p^*}, \quad (10)$$

$$\hat{g}^* = K_{p^*p^*} U^\top \left( U K_{p^*p^*} U^\top + \sigma^2 I_N \right)^{-1} Y. \quad (11)$$

In the Bayesian approach, $\hat{g}^*$ is the estimate of the impulse response. Note that $g^*$ is regarded as a random variable in the Bayesian estimation framework although its true value is a constant. This fact becomes important when considering the entropy; if $g^*$ is constant, the entropy of $g^*$ becomes $-\infty$.

This procedure is easily extended to the case with several scheduling parameters of interest $p_1^*, \ldots, p_m^*$. Replace $K_{p^*p^*}$ and $K_{p^*p^*}$ by

$$\begin{bmatrix} K_{p_1^*p_1^*} & \cdots & K_{p_1^*p_m^*} \\ \vdots & \ddots & \vdots \\ K_{p_m^*p_1^*} & \cdots & K_{p_m^*p_m^*} \end{bmatrix}, \quad (12)$$
and
\[
\begin{bmatrix}
K_{p_1^T p'(1)} & \cdots & K_{p_1^T p'(N)} \\
\vdots & \ddots & \vdots \\
K_{p_m^T p'(1)} & \cdots & K_{p_m^T p'(N)}
\end{bmatrix}
\]  \tag{13}
respectively. Then, the posterior distribution of \( g'_p = [g'_p, \ldots, g'_p] \) is also given by (9) through (11).

Before stating the problem, we introduce the mutual information, which is a fundamental quantity in information theory [13]. The mutual information between two random variables \( g^* \) and \( Y \) are defined by
\[
I(g^*; Y) = H(g^*) + H(Y) - H(g^* Y), \tag{14}
\]
where \( H(g^*) \) and \( H(Y) \) are Shannon’s entropies of \( g^* \) and \( Y \), respectively, and \( H(g^* Y) \) is a joint entropy of \( g^* \) and \( Y \). Here, Shannon’s entropy of a random variable \( X \) is defined as (with a slight abuse of notation)
\[
H(X) = -\int p_X(x) \log p_X(x) dx, \tag{15}
\]
where \( p_X \) denotes the probability density function of \( X \). In (15), the integral is taken over the support of \( p_X \). Similarly, the joint entropy of \( X_1 \) and \( X_2 \) is defined as
\[
H(X_1, X_2) = -\int p_{X_1X_2}(x_1, x_2) \log p_{X_1X_2}(x_1, x_2) dx_1 dx_2. \tag{16}
\]

Now we set the problem considered in this paper.

**Problem 1** Assume that the kernel function \( k \), the noise variance \( \sigma^2 \), the maximum energy for available input \( E \), the scheduling variables of interest \( p'_1, \ldots, p'_n \), the length of FIR \( n \), the length of the experiment \( N \), and the sequence of scheduling variable \( p'(1), \ldots, p'(N) \) are given. Find the input sequence \( u = [u(1), \ldots, u(N)]^T \in \mathbb{R}^N \) which maximizes the mutual information between \( g^* \) and \( Y \) subject to \( u^T u \leq E \).

Reasons to employ the mutual information as a criterion are shown in Sect. 3.

### 3. Mutual Information and Its Maximization

In Sect. 3, a solution for Problem 1 is shown. First, some reasons to employ the mutual information are given in Sect. 3.1. Then we give an algorithm which maximizes the mutual information in Sect. 3.2.

#### 3.1 Mutual Information for LPV System

In Sect. 3.1, we first show the explicit expression of \( I(g^*; Y) \), and then give some reasons to employ the mutual information as a criterion of the input sequence.

From the well-known property of the entropy of a multivariate Gaussian distribution [13], we have
\[
\begin{align*}
H(g^*) &= \frac{1}{2} \log(2\pi e)^n + \frac{1}{2} \log \det (K_{g'_p}) , \tag{17} \\
H(Y) &= \frac{1}{2} \log(2\pi e)^N + \frac{1}{2} \log \det (UK_{p'_p}U^T + \sigma^2 I_N) , \tag{18} \\
H(g^* Y) &= \frac{1}{2} \log(2\pi e)^{n+N} \\
&+ \frac{1}{2} \log \det \begin{bmatrix} K_{g'_p} & K_{g'_p}U^T \\ UK_{p'_p} & UK_{p'_p}U^T + \sigma^2 I_N \end{bmatrix} , \tag{19}
\end{align*}
\]
where \( K_{g'_p} \) and \( K_{g'_p} \) are given by (12) and (13), respectively. For the simplicity of notation, let \( K_y \in \mathbb{R}^{N \times N} \) and \( K_0 \in \mathbb{R}^{(n+N)(n+N)} \) be the matrices defined by
\[
\begin{align*}
K_y &= U K_{p'_p} U^T + \sigma^2 I_N , \tag{20} \\
K_0 &= \begin{bmatrix} K_{g'_p} & K_{g'_p}U^T \\ UK_{p'_p} & UK_{p'_p}U^T + \sigma^2 I_N \end{bmatrix} . \tag{21}
\end{align*}
\]

Then, from the formula of the determinants of block matrices,
\[
\det (K_0) = \det \left( K_{g'_p} \right) \det (K_y - U K_{p'_p} K_{p'_p} U^T) . \tag{22}
\]

By substituting (17), (18), (19) and (22) into (14),
\[
I(g^*; Y) = \frac{1}{2} \log \det (K_y) - \frac{1}{2} \log \det (K_0) . \tag{23}
\]

where \( \tilde{K} \in \mathbb{R}^{N \times N} \) is defined as
\[
\tilde{K} = U \left( K_{p'_p} - K_{p'_p} K_{p'_p} U^T + \sigma^2 I_N \right) U^T . \tag{24}
\]

It should be noted that \( K_y \) and \( \tilde{K} \) have the same structure with respect to \( U \). This fact becomes important when considering the gradient of \( I(g^*; Y) \) with respect to \( u \).

The mutual information is often employed as a criterion for the Bayesian experiment design since 1950’s (see e.g., [8]–[10]). In the rest of this subsection, some reasons to employ the mutual information for the experiment design are shown.

First, the mutual information is equivalent to the expected Kullback-Leibler divergence between the prior and the posterior distribution (see e.g., [10], [13]). Let \( p_1(x) \) and \( p_2(x) \) be probability density functions over \( X \). Then, the Kullback-Leibler (KL) divergence between these two probability density functions is defined as
\[
D(p_1 \| p_2) = \int_X \log \frac{p_1(x)}{p_2(x)} p_1(x) dx . \tag{25}
\]

See e.g., [13], [14] for more details. The KL divergence is not symmetric, however, it is a kind of pseudo-distance because \( D(p_1 \| p_2) \geq 0 \) holds for all \( p_1 \) and \( p_2 \), and \( D(p_1 \| p_2) = 0 \) if and only if \( p_1(x) = p_2(x) \) for almost everywhere over \( X \). In the context of the Bayes estimation, the expected value of the KL divergence with respect to the output is equivalent to the mutual information; i.e.,
\[
\mathbb{E}_Y \left[ D(p(g^* \| Y) \| p(g^*)) \right] = I(g^*; Y). \tag{26}
\]

This suggests that a difference between the prior and the posterior distribution can be measured by the mutual information between \( g^* \) and \( Y \). It should be noted that the experiment has no value when the prior and the posterior distribution is identical, i.e., \( D(p(g^* \| Y) \| p(g^*)) = 0 \). Hence, maximizing the KL divergence is a natural and intuitive way in the Bayesian experiment design.

Second, maximizing the mutual information is identical to minimizing the conditional entropy \( H(g^* \| Y) = H(g^*, Y) - I(g^*; Y) \).
\( H(Y) \), which is one measure of uncertainty of \( g' \) under the observation \( Y \). In other words, the determinant of the posterior covariance matrix of \( g' \) is minimized by maximizing the mutual information between \( g' \) and \( Y \). Minimizing such a covariance leads to reduce the effect of noise, thus this is also a reason to employ the mutual information as a criterion.

For these reasons, this paper employs the mutual information as a criterion for the input design problem.

### 3.2 Maximization of Mutual Information

In Sect. 3.2, a concrete procedure to generate the maximizer of the mutual information is given. To this end, we first show the gradient of the mutual information with respect to each \( u(t) \).

Recall that the mutual information given by (23) is described by a difference of \( \frac{1}{2} \log \det \left( UXU^T + \sigma^2 I_N \right) \), where \( X \in \{K_{p'p'}, K_{p'p'} - K_{p'p'}^{-1} K_{p'p'} K_{p'p'}^{-1}, K_{p'p'} \} \). Hence we focus on the gradient of \( J(u) = \log \det \left( UXU^T + \sigma^2 I_N \right) \) with respect to \( u \) in the following.

Let \( Z = UXU^T + \sigma^2 I_N \) for the ease of notation. Then,

\[
\frac{\partial J}{\partial u(t)} = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial J}{\partial (Z_{ij})} \frac{\partial (Z_{ij})}{\partial u(t)} = \text{Tr} \left( \frac{\partial J}{\partial Z} \frac{\partial Z}{\partial u(t)} \right),
\]

(27)

from the chain rule. The derivative of the log-determinant function is given by

\[
\frac{\partial J}{\partial Z} = Z^{-1},
\]

(28)

and

\[
\frac{\partial Z}{\partial u(t)} = U_X U^T + UXU^T,
\]

(29)

where \( (i, j) \) element of \( U_t \in \mathbb{R}^{N \times N} \) is defined by

\[
[U_{t}]_{ij} = \begin{cases} 1 & \text{if } t \leq i \leq N, j \leq ir, j = r(i - 1) + i - t + 1, \\ 0 & \text{otherwise}. \end{cases}
\]

(30)

See [15] for more detail about the derivatives of functions of matrices.

Based on these calculations, we have

\[
\frac{\partial J}{\partial u(t)} = \text{Tr} \left( \left( UXU^T + \sigma^2 I_N \right)^{-1} \left( U_X U^T + UXU^T \right) \right),
\]

(31)

thus the \( t \) th element of the gradient of \( I(g'; Y) \) with respect to \( u \) is given by

\[
\frac{\partial I(g'; Y)}{\partial u(t)} = \frac{1}{2} \text{Tr} \left( \left( UK_u U^T + \sigma^2 I_N \right)^{-1} \left( UK_u U^T + UK_u U^T \right) \right) - \frac{1}{2} \text{Tr} \left( \left( UK_u U^T + \sigma^2 I_N \right)^{-1} \left( UK_u U^T + UK_u U^T \right) \right).
\]

(32)

Based on these calculations, we have

\[
\frac{\partial I(g'; Y)}{\partial u} = \left[ \frac{\partial I(g'; Y)}{\partial u(1)}, \ldots, \frac{\partial I(g'; Y)}{\partial u(N)} \right]^T \in \mathbb{R}^N.
\]

(33)

The mutual information \( I(g'; Y) \) is not a concave function, hence its global maximization is difficult. However, the gradient of \( I(g'; Y) \) is available and the constraint \( u' u \leq E \) is convex, thus the projected gradient method can be employed for local maximization of \( I(g'; Y) \). Consider the maximization of the differentiable function \( f : \mathbb{R}^N \rightarrow \mathbb{R} \) over the convex set \( X \subset \mathbb{R}^N \). Then, the projected gradient method is described by the sequence

\[
x_{k+1} = P_X \left( x_k + \beta \frac{\partial f}{\partial x} \right),
\]

(34)

where \( x_k \) is a solution vector at \( k \) th step, \( \frac{\partial f}{\partial x} \) is the gradient of \( f \) at \( x_k \), and \( P_X \) denotes the orthogonal projection onto \( X \). The parameter \( \beta > 0 \) shows the step size. It is known that \( x_k \) converges to the stationary point under some mild conditions (see e.g., Proposition 2.3.2 in [16]).

Algorithm 1 shows the pseudo code of the optimization procedure. The parameter \( \Delta \) and \( N_{\text{max}} \) determine the terminal condition of the optimization. When the update becomes sufficiently small or the number of iterations reaches \( N_{\text{max}} \), the algorithm stops. We show the effectiveness of this optimization procedure through a numerical example in Sect. 4.

### 4. Numerical Simulation

This section provides a numerical example to demonstrate the effectiveness of the proposed method. The target system, which is taken from [3], is described by the difference equation

\[
\sum_{i=0}^{N} a_i(p_k) y(k+i) = b(p_k) u(k+4),
\]

(35)

where \( a_i(p_k) \) and \( b_k \) are defined by

\[
a_0(p_k) = -0.003, \quad a_1(p_k) = \frac{12}{125} - 0.1 \sin(p_k),
\]

\[
a_2(p_k) = -\frac{23}{85} + 0.2 \sin(p_k), \quad a_3(p_k) = \frac{61}{110} - 0.2 \sin(p_k),
\]

\[
a_4(p_k) = \frac{511 + 192 p_k^2 - 258(\cos(p_k) - \sin(p_k))}{860}, \quad a_5(p_k) = 0.58 - 0.1 p_k, \quad b(p_k) = \cos(p_k).
\]

This system is studied in Section 8.3.3.1 of [3], and it is stable over \( 0.6 \leq p_k \leq 0.8 \). Even though the dependence on scheduling variables of this system is different from (1), we employ the LPV-FIR model structure for the identification. This is because 1) the dependency of scheduling variables is unknown in practice, and 2) the stability of the system is known. With these conditions, the LPV-FIR model structure (1) is a reasonable choice. We set \( N = 100, n = 50, \sigma^2 = 10^{-2}, \) and \( E = 100. \)
The trajectory of the scheduling variable during the experiment is defined by
\[ p^*(k) = 0.8 - \frac{0.2}{1 + e^{0.1k}}, \]
and the scheduling variable of interest is set to \( p^* = 0.7 \). The transient of the scheduling variable is shown in Fig. 2. The horizontal axis shows the time step of the experiment, and the vertical axis shows the scheduling variable. The solid line and the broken line show the scheduling variable \( p^*(k) \) and \( p^* \), i.e., the scheduling variable through the experiment and that of interested, respectively. Figure 3 shows the impulse responses of the target system with fixed scheduling variables. The thick solid line, the thin solid line, and the thin broken line show the impulse responses with \( p(k) = 0.7, 0.6, \) and 0.8, respectively. Note that these impulse responses are significantly different. This suggests that identifying the impulse response for \( p(k) = 0.7 \) with the experiment with the scheduling variable depicted in Fig. 2 is not an easy task (recall \( n = 50 \) and \( N = 100 \)), hence the input design is crucial.

As a kernel, we employ the one proposed in [5] where the kernel function \( k(q_1, q_2, i_1, i_2, \eta) \) is defined as
\[ k(q_1, q_2, i_1, i_2, \eta) = \eta_1 \eta_2 \eta_3 \eta_4 e^{-\frac{(q_1 - q_2)^2}{2 \sigma^2}}. \]

The hyperparameter \( \eta \) is a four dimensional vector, and its elements satisfy the constraint \( |\eta_1| > 0, 0 < |\eta_2| < 1, |\eta_3| < 1, \) and \( |\eta_4| > 0 \). In this simulation, we set the hyperparameters \( \eta = [6, 0.6, 0.6, 0.01]^T \) by an oracle.

Based on the above setting, we run Algorithm 1 with \( \beta = 10^{-4}, \Delta = 10^{-4} \), and \( N_{\text{max}} = 30000 \). The resulting input sequence is shown in Fig. 4. The horizontal axis shows the time step, and the vertical axis shows the input. The solid line shows the input generated by Algorithm 1, and the broken line shows the random sequence for comparison. This random sequence is generated from the uniform white noise over \([-1, 1]\), and adjusted so as to satisfy \( u^T u = E \).

Figure 5 illustrates the noise-free outputs with these inputs. The horizontal axis shows the time, and the vertical axis shows the output. The solid line and the broken line show the outputs with the proposed and random inputs, respectively. The proposed input strongly excites the system around the 50 steps, where the scheduling variable \( p^*(k) \) becomes closer to \( p^* \).

To show the effectiveness of the proposed input, we run 300 Monte Carlo studies with independent noise realizations. Figures 6 and 7 show the estimated impulse responses with the proposed and random inputs, respectively. The horizontal axes show the time step, and the vertical axes show the impulse response. The solid lines show the impulse response of the true system with \( p^* \), and the gray lines show the estimated impulse responses (11). The proposed input clearly improves the identification accuracy especially for the first 20 steps.
The mutual information gives one solution for this accuracy. Note that how we should concentrate the energy is unclear. This paper proposes a new approach to this problem. The proposed input takes account of the trajectory of the scheduling variable during the experiment. The proposed input significantly outperforms the random one. This results with the proposed and random inputs respectively. The proposed input design problem for LPV system identification is discussed. In particular, this paper focuses on the case where the scheduling variable changes during the experiment, and the available energy for the identification input is constrained. One of the difficulties of this problem is that how we should distribute the feasible energy is unclear. This paper employs the mutual information as a criterion of identification input, and gives a concrete procedure to obtain the local maximum. A numerical simulation is shown to demonstrate the effectiveness of the proposed input design.

The mutual information is not concave and its maximization is not obvious. The mutual information gives one solution for this problem.

5. Conclusion

In this paper, the input design problem for LPV system identification is discussed. In particular, this paper focuses on the case where the scheduling variable changes during the experiment, and the available energy for the identification input is constrained. One of the difficulties of this problem is that how we should distribute the feasible energy is unclear. This paper employs the mutual information as a criterion of identification input, and gives a concrete procedure to obtain the local maximum. A numerical simulation is shown to demonstrate the effectiveness of the proposed input design.

The mutual information is not concave and its maximization is not an easy task. Hence, the concave relaxation of the mutual information is one of the future tasks.

Acknowledgments

This work is supported by JSPS Grant-in-Aid for JSPS fellow Number 15J05700 and JSPS KAKENHI Grant Number JP17H03281.

References

[1] L. Ljung: System Identification: Theory for the User, Prentice Hall, 1987.
[2] D.J. Leith and W.E. Leithead: Survey of gain-scheduling analysis and design, International Journal of Control, Vol. 73, No. 11, pp. 1001–1025, 2000.
[3] R. Tóth: Modeling and Identification of Linear Parameter-Varying Systems, Springer, 2010.
[4] A. Golabi, N. Meskin, R. Tóth, and J. Mohammadpour: A Bayesian approach for estimation of linear-regression LPV models, Proceedings of 53rd IEEE Conference on Decision and Control (CDC 2014), pp. 2555–2560, 2014.
[5] Y. Okabe and Y. Ohta: A new prior distribution for Bayesian approach in LPV system identification, Proceedings of SICE International Symposium on Control Systems 2016, 4A2-3, 2016.
[6] G. Pillonetto and G. De Nicolao: A new kernel-based approach for linear system identification, Automatica, Vol. 46, No. 1, pp. 81–93, 2010.
[7] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung: Kernel methods in system identification, machine learning and function estimation: A survey, Automatica, Vol. 50, No. 3, pp. 657–682, 2014.
[8] D.V. Lindley: On a measure of the information provided by an experiment, The Annals of Mathematical Statistics, Vol. 27, No. 4, pp. 986–1005, 1956.
[9] S. Arimoto and H. Kimura: Optimum input test signals for system identification: An information-theoretical approach, International Journal of Systems Science, Vol. 1, No. 3, pp. 279–290, 1971.
[10] K. Chaloner and I. Verdinelli: Bayesian experimental design: A review, Statistical Science, Vol. 10, No. 3, pp. 273–304, 1995.
[11] Y. Fujimoto and T. Sugie: Informative input design for kernel based-system identification, Proceedings of 55th IEEE Conference on Decision and Control, pp. 4636–4639, 2016.
[12] W. Kasai, Y. Fujimoto, and T. Sugie: On input design for linear parameter varying system identification, Proceedings of the 61st Annual Conference of the Institute of Systems, Control and Information Engineers (SCI’17), 146-1, 2017 (in Japanese).
[13] T.M. Cover and J.A. Thomas: Elements of Information Theory, 2nd Edition, Wiley-Interscience, 2006.
[14] A. Ullah: Entropy, divergence and distance measures with econometric applications, Journal of Statistical Planning and Inference, Vol. 49, No. 1, pp. 137–162, 1996.
[15] D.S. Bernstein: Matrix Mathematics: Theory, Facts, and Formulas, Princeton University Press, 2009.
[16] D.P. Bertsekas: Nonlinear Programming, 2nd Edition, Athena Scientific, 1999.

Yusuke Fujimoto (Student Member)

He received the Bachelor of Engineering and Master of Informatics degrees from Kyoto University, Kyoto, Japan, in 2013 and 2015, respectively. He is currently a Ph.D. student in the Department of Systems Science, Kyoto University. His current research interests include system identification and data-driven control. He is a student member of IEEE.

Wataru Kasai

He received the Bachelor of Engineering from Kyoto University, Kyoto, Japan, in 2017. He is currently a Master student in the Department of Systems Science, Graduate School of Informatics, Kyoto University. His current research interest is system identification.
Toshiharu Sugie (Member, Fellow)

He received the B.E., M.E., and Ph.D. degrees in engineering from Kyoto University, Japan, in 1976, 1978 and 1985, respectively. From 1978 to 1980, he was a research member of Musashino Electric Communication Laboratory in NTT, Musashino, Japan. From 1984 to 1988, he was a research associate of the Department of Mechanical Engineering, University of Osaka Prefecture, Osaka. In 1988, he joined Kyoto University, where he is currently a Professor of the Department of Systems Science. His research interests are in robust control, learning control, nonlinear control, identification for control, and control application to mechanical systems. He is an IEEE Fellow.