Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Stochastic input design problems for the frequency response in Bayesian identification

Man Zheng * **Yoshito Ohta** **

* *Department of Applied Mathematics and Physics, Kyoto University, Japan (e-mail: zheng.man.82x@st.kyoto-u.ac.jp).*
** *Department of Applied Mathematics and Physics, Kyoto University, Japan (e-mail: yoshito_ohta@i.kyoto-u.ac.jp)*

Abstract: Recently, the research of identification input design for Bayesian methods has been actively investigated. Either the problem is formulated as a non-convex problem with difficulty in solving or relaxed as a convex problem with a price of some conservativeness. In this contribution, a new minimum power input design problem is formulated by viewing the input as a stochastic process. We seek the minimum energy input with variance constraints over a frequency band. By exploiting the generalized Kalman-Yakubovich-Popov lemma, the stochastic consideration facilitates the input design problem to be presented as a convex problem whose decision variables are a finite number of autocorrelation coefficients. We obtain the autocorrelation coefficients of the desired stochastic input signal by solving the convex problem and extend them by the maximum entropy extension. Then, a specific identification input is sampled from the obtained stochastic process. Simulations results demonstrate the effectiveness of the proposed method.

Keywords: system identification, Bayesian methods, frequency domain identification, input and excitation design, convex optimization.

1. INTRODUCTION

System identification is a traditional and practical research topic which aims to build the mathematical expressions for dynamic systems. An essential problem in this literature is to estimate linear transfer functions based on limited input-output data. A classical method for handling such problems is prediction error methods (PEMs). However, PEMs suffer from the critical model/order selection problem, which may lead to tedious identification, and the performance of PEMs may not always be satisfactory especially for short and noisy observations.

By contrast, another identification technique, i.e., the Bayesian identification, overcomes the shortcomings of PEMs and has therefore received much attention recently. In kernel-based identification, the system parameters are commonly considered to be a zero-mean Gaussian random vector whose covariance (kernel) is designed to contain system information. Via combining the prior knowledge and the data, the kernel-based approach turns out to be a more reliable identification method with a small data set than PEMs. Furthermore, the kernel-based identification almost focuses on the high-order finite-impulse-response (FIR) model which will reasonably approximate the true stable identified system provided that the model order is large enough. This formulating let kernel-based identification escape from heavy model/order selection problems and thus conduct a more concise identification than PEMs. Other details about the kernel-based identification can be referred to Chen et al. (2012).

The latest research for kernel-based identification virtually focuses on the kernel structure. Previous researches demonstrated that kernels should reflect the characteristics of the impulse response such as stability and smoothness. Hence, the kernels designed for stable linear systems typically have exponentially decreasing diagonal elements and positive correlation, which makes them different from kernels frequently-used in machine learning literature. The basic kernels, i.e., the diagonal/correlated (DC) kernel, the tuned/correlated (TC) kernel, and the stable/spline (SS) kernel, are all available for expressing stable and smooth impulse responses in the stable linear systems identification. Furthermore, the prior selection is another important topic for Bayesian identification. For example, the use of a noninformative prior is discussed in Zheng and Ohta (2018).

However, there are still lots of problems unsettled for the kernel-based identification including the input design problem. The research about the input design for PEMs has a long history and has led to a rich collection of theoretical results Ljung (1999). The classical approach to organizing the input design problem is to minimize some scalar function of the covariance matrix with constraints on input power. Meanwhile, the study by Jansson and Hjalmarsson (2005) shows that the classical input design problem allows a convex formulation via considering the input spectrum as the design variable. The framework of input design for PEMs identification is available to the kernel-based identification because the covariance matrix of the posterior is straightforward under the Bayesian consideration. There have been several results in relation to this idea. In Fujimoto and Sugie (2018), they proposed a non-convex problem which maximizes the mutual information between the output and the impulse response subject to the energy-constraint. Then, a gradient-based method is proposed to handle this problem. However, the proposed problem in Fujimoto and Sugie (2018) will suffer a local minimum issue. In another contribution Mu and Chen (2018), an input design problem is proposed via minimizing some scale measures of the posterior covariance subject to the input energy constraint. Then, a convex formulation is valid for the periodic input. On the other hand, the algorithm proposed in Mu and Chen (2018) will only be effective if the optimal input is a periodic input. The periodic signal facilitates the convex formulation of the input design problem proposed in this paper. However, the optimal input may not always be a periodic input. A simple counterexample is given in Section 3. Thus, the exploitation for a more general and practical input design problem formulation is essential.

In this paper, we propose a novel input design problem which is different from the previous work. Instead of designing a deterministic input, we first consider to select a stochastic process and then sample the identification input from the obtained stochastic process. In the first step, we formulate the problem for a stationary Gaussian stochastic process whose autocorrelation coefficients are design variables instead of a deterministic input signal. Then, we consider minimizing the average power of the designed stochastic process subject to a variance constraint of the frequency response. In the second step, we can obtain the input by sampling techniques. Although some samples from the captured stochastic process may not satisfy the design requirements, we can easily distinguish an unsatisfactory input and only accept a good one.

The paper is organized as follows. In Section 2, the Bayesian system identification approach on Gaussian prior is reviewed briefly. In Section 3, we first propose an input design problem with frequency-domain constraints considering a deterministic input signal. The property about the optimal input in this problem is also derived. Section 4 reviews several mathematical preliminaries. In Section 5, we transform the original problem proposed in Section 3 to a convex optimization problem using the stochastic consideration. In Section 6, simulated data is reported to demonstrate the effectiveness of the proposed approach. Our conclusions, given in Section 7, end the paper.

Notation: The symbol $O_{k\times l}$ denotes the zero matrix of order $k \times l$. For a complex number $c \in \mathbb{C}$, the real part and the imaginary part of c are denoted by $\mathrm{Re}(c)$ and $\mathrm{Im}(c)$, respectively. Meanwhile, $\mathbb{R}^n$ denotes the real vector space of n-dimension. For a vector $\beta \in \mathbb{R}^n$, β_k denotes the kth element of β. The symbol $\mathcal{H}_n$ stands for the space of n-dimensional Hermitian matrices. For a matrix $M \in \mathcal{H}_n$, $M \succeq 0$ and $M \preceq 0$ denote positive semidefiniteness and negative semidefiniteness, respectively. Let $\mathcal{T}(r_0, r_1, \ldots, r_{n-1})$ denote the symmetric Toeplitz matrix whose first row is $[r_0\ r_1\ \ldots\ r_{n-1}]$.

2. BAYESIAN IDENTIFICATION FRAMEWORK

A linear single-input-single-output system is written as

$$y(t) = G(q)u(t) + v(t). \tag{1}$$

Here $\{u(t), y(t), t = 1, \ldots, N\}$ denote the input-output data and $v(t)$ denotes the additive noise, while q in (1) is the shift operator, $qu(t) = u(t+1)$. The Bayesian linear identification usually considers to build G like an n-order discrete FIR model, namely

$$G(\theta, q) = \sum_{k=1}^{n} g_k q^{-k}, \tag{2}$$

where $\theta := [g_1\ g_2\ \ldots\ g_n]^T$ is the estimation objective, i.e., the finite impulse response. The transfer function of the linear model (2) is obtained by the z-transformation of the impulse response

$$G(\theta, z) = \sum_{k=1}^{n} g_k z^{-k}, \tag{3}$$

where z is a complex variable. The corresponding frequency response of G is $G(\theta, z)\mid_{z=\mathrm{e}^{j\omega}}$, i.e.,

$$G\left(\theta, \mathrm{e}^{j\omega}\right) = \sum_{k=1}^{n} g_k \mathrm{e}^{-j\omega k}. \tag{4}$$

Via considering the FIR construction (2), the linear model (1) is rewritten in a matrix formulation:

$$Y_N = \Phi_N^T \theta + \Lambda_N,$$

where

$$Y_N := [y(n+1)\ y(n+2)\ \ldots\ y(N)]^T, \tag{5a}$$
$$\phi(t) := [u(t-1)\ \ldots\ u(t-n)]^T, \tag{5b}$$
$$\Phi_N := [\phi(n+1)\ \phi(n+2)\ \ldots\ \phi(N)], \tag{5c}$$
$$\Lambda_N := [v(n+1)\ v(n+2)\ \ldots\ v(N)]^T. \tag{5d}$$

An interpretation in a Bayesian perspective is that the parameter θ has a prior distribution which is a Gaussian distribution with mean zero and the covariance matrix K. Then under the assumption that the noise Λ_N is white Gaussian, i.e., $\Lambda_N \sim N\left(0, \sigma^2 I\right)$, the posterior distribution of θ given Y_N is

$$\theta \mid Y_N \sim N\left(\theta^B, P\right),$$

where θ^B is the maximum a posteriori (MAP) estimation calculated by

$$\begin{aligned}\theta^B &= \arg\min_{\theta}\ \| Y_N - \Phi_N^T\theta \|^2 + \sigma^2\theta^T K^{-1}\theta \\ &= \left(\Phi_N \Phi_N^T + \sigma^2 K^{-1}\right)^{-1} \Phi_N Y_N,\end{aligned}$$

and P is

$$P = \left(\frac{\Phi_N \Phi_N^T}{\sigma^2} + K^{-1}\right)^{-1}. \tag{6}$$

Here, the frequency response $G(\theta, \mathrm{e}^{j\omega})$ can also be realized as a random variable which is dependent on the frequency variable ω and the impulse response θ. When the distribution of θ is Gaussian, then the linear combination of g_k, i.e., $G(\theta, \mathrm{e}^{j\omega})$, is also Gaussian relying on ω. It should be noted that $G(\theta, \mathrm{e}^{j\omega})$ is a complex random variable which has real and imaginary parts. Thus, based on the prior and posterior distribution of θ, we can derive the posterior distribution of $G(\theta, \mathrm{e}^{j\omega})$ as follows:

$$\begin{bmatrix} \mathrm{Re}\left(G\left(\theta, \mathrm{e}^{j\omega}\right)\right) \mid Y_N \\ \mathrm{Im}\left(G\left(\theta, \mathrm{e}^{j\omega}\right)\right) \mid Y_N \end{bmatrix} \sim N\left(\begin{bmatrix} \rho_1\theta^B \\ -\rho_2\theta^B \end{bmatrix}, \begin{bmatrix} \rho_1 P\rho_1^T & -\rho_1 P\rho_2^T \\ -\rho_2 P\rho_1^T & \rho_2 P\rho_2^T \end{bmatrix}\right), \tag{7}$$

where

$$\Lambda_G := \left[\mathrm{e}^{-j\omega}\ \mathrm{e}^{-2j\omega}\ \ldots\ \mathrm{e}^{-nj\omega}\right], \tag{8a}$$

$$\rho_1 := \mathrm{Re}(\Lambda_G) = [\cos(\omega)\ \cos(2\omega)\ \ldots\ \cos(n\omega)], \tag{8b}$$

$$\rho_2 := -\mathrm{Im}(\Lambda_G) = [\sin(\omega)\ \sin(2\omega)\ \ldots\ \sin(n\omega)]. \tag{8c}$$

3. PROBLEM FORMULATION VIA DETERMINISTIC INPUT

3.1 Problem statement

A natural idea to formulate the input design problem is minimizing the energy of a deterministic input subject to a variance constraint. In this paper, we consider to use the trace of the covariance matrix in (7), i.e., $\rho_1 P\rho_1^T + \rho_2 P\rho_2^T$, as a scalar measure of the posterior covariance. Meanwhile, it should be noticed that this scalar measure is equivalent to a quadratic form of the covariance matrix P as follows:

$$\begin{aligned} \Lambda_G P \Lambda_G^* &= (\rho_1 - \rho_2 j)\, P\, (\rho_1 + \rho_2 j)^T \\ &= \rho_1 P\rho_1^T + \rho_2 P\rho_2^T - \rho_2 P\rho_1^T j + \rho_1 P\rho_2^T j \\ &= \rho_1 P\rho_1^T + \rho_2 P\rho_2^T, \end{aligned} \tag{9}$$

where Λ_G^* denotes the conjugate transpose of Λ_G.

Thus, we can summarize the proposed input design problem as follows:

$$\min_u\ u^T u \tag{10a}$$

$$\text{s.t. } \Lambda_G P\Lambda_G^* \le \alpha\ \forall\omega \in [\omega_1, \omega_2]. \tag{10b}$$

Here, the frequency band $[\omega_1, \omega_2]$ represents the set of important frequencies. For example, the band near the crossover frequency in feedback control is critical for the control robustness if the identified model is used for compensator design. This frequency band can be decided according to preliminary experiments or practical demands. Similarly, the error α is decided by practical requirements. Nevertheless, it should satisfy $\alpha < \sup_{\omega\in[\omega_1,\omega_2]} \Lambda_G K \Lambda_G^*$ to let (10b) work as an active constraint in the above problem.

Unfortunately, the problem (10) is a non-convex optimization problem and thus suffer the local minimum issue. Hence, it is necessary to exploit the convex reformulation for (10).

3.2 Discussion of previous approaches

A periodic input, proposed by Mu and Chen (2018), admits a convex construction of problem (10). However, the optimal input obtained by (10) may not always be a periodic signal. As an example, we consider a special case of problem (10). We assume that $N - 1 = n$, $K = I$, and ω is fixed to be one frequency. Then, problem (10) is rewritten as follows:

$$\min_u\ u^T u \tag{11a}$$

$$\text{s.t. } \Lambda_G \left(\frac{uu^T}{\sigma^2} + I\right)^{-1} \Lambda_G^* \le \alpha. \tag{11b}$$

Let J be a matrix that is all ones along the secondary diagonal and zero everywhere else. Then, a vector u is called symmetric if $Ju = u$ and skew symmetric if $Ju = -u$ respectively. The following theorem reveals the property about the optimal solution of problem (11).

Theorem 1. The optimal solution of (11) is symmetric or skew-symmetric.

Proof. Based on the so-called matrix inversion lemma, (11) is equivalent to

$$\min_u\ u^T u \tag{12a}$$

$$\text{s.t. } u^T \left(\frac{\Lambda_G^*\Lambda_G}{\sigma^2} - \frac{n-\alpha}{\sigma^2} I\right) u \ge n - \alpha, \tag{12b}$$

where $\alpha < \Lambda_G I \Lambda_G^* = n$.

Notice that the complex vector Λ_G is decomposed as

$$\Lambda_G = \rho_1 - \rho_2 j.$$

Then, because $u^T(\rho_2^T\rho_1 - \rho_1^T\rho_2)u = 0$, we have

$$\begin{aligned} u^T\Lambda_G^*\Lambda_G u &= u^T\left(\rho_1^T + \rho_2^T j\right)(\rho_1 - \rho_2 j)\, u & (13a) \\ &= u^T\left(\rho_1^T\rho_1 + \rho_2^T\rho_2 + \left(\rho_2^T\rho_1 - \rho_1^T\rho_2\right) j\right) u & (13b) \\ &= u^T\left(\rho_1^T\rho_1 + \rho_2^T\rho_2\right) u, & (13c) \end{aligned}$$

where $\rho_1^T\rho_1 + \rho_2^T\rho_2$ is verified to be a symmetric Toeplitz matrix. Thus, (12) is further simplified as follows:

$$\min_u\ u^T u \tag{14a}$$

$$\text{s.t. } u^T \left(\frac{\rho_1^T\rho_1 + \rho_2^T\rho_2}{\sigma^2} - \frac{n-\alpha}{\sigma^2} I\right) u \ge n - \alpha. \tag{14b}$$

The optimal solution of (14) lies on the eigenvector of $\rho_1^T\rho_1 + \rho_2^T\rho_2$ corresponding to the maximum eigenvalue.

It should be noticed that the paper Cantoni and Butler (1976) explores the eigenvectors of symmetric Toeplitz matrices whose eigenvalues are all distinct. Furthermore, they indicate that the eigenvectors are symmetric or skew-symmetric in this case. However, when $n \ge 4$, the matrix $\rho_1^T\rho_1 + \rho_2^T\rho_2$ always has the same eigenvalues 0, for which the eigenvectors are not discussed clearly in the previous research. Thus, we now turn to prove that all eigenvectors of $\rho_1^T\rho_1 + \rho_2^T\rho_2$ corresponding to the non-zero eigenvalues are symmetric or skew-symmetric, which leads to Theorem 1 immediately. First, let $x\rho_1^T + y\rho_2^T$ lie on the eigenvector of $\rho_1^T\rho_1 + \rho_2^T\rho_2$ where $x \neq 0$ and $y \neq 0$. Then, it follows that

$$\begin{aligned} &\left(\rho_1^T\rho_1 + \rho_2^T\rho_2\right)\left(x\rho_1^T + y\rho_2^T\right) \\ &= \left(x\rho_1\rho_1^T + y\rho_1\rho_2^T\right)\rho_1^T + \left(x\rho_2\rho_1^T + y\rho_2\rho_2^T\right)\rho_2^T. \end{aligned} \tag{15}$$

Since the right-hand side of (15) is in the same direction of $x\rho_1^T + y\rho_2^T$, it follows that

$$\frac{x\rho_1\rho_1^T + y\rho_1\rho_2^T}{x} = \frac{x\rho_2\rho_1^T + y\rho_2\rho_2^T}{y}. \tag{16}$$

Solving the equation (16) gives

$$\frac{y}{x} = \frac{-a \pm \sqrt{a^2 + 4b^2}}{2b}, \tag{17}$$

where $a := \rho_1\rho_1^T - \rho_2\rho_2^T$ and $b := \rho_1\rho_2^T$.

If we let $x = 1$, the vector $x\rho_1^T + y\rho_2^T$ becomes

$$\bar{\rho} := \rho_1^T + \frac{-a \pm \sqrt{a^2 + 4b^2}}{2b}\rho_2^T. \tag{18}$$

What is left is to show that $\bar{\rho}$ is symmetric or skew-symmetric.

By the formulas of trigonometric functions, it is trivial to confirm that

$$a = \sum_{k=1}^{n} \cos(2k\omega) = \mathrm{Re}\left(\sum_{k=1}^{n} \mathrm{e}^{2k\omega j}\right), \tag{19}$$

$$2b = \sum_{k=1}^{n} \sin(2k\omega) = \mathrm{Im}\left(\sum_{k=1}^{n} \mathrm{e}^{2k\omega j}\right). \tag{20}$$

Notice that

$$\sum_{k=1}^{n} \mathrm{e}^{2k\omega j} = \frac{\mathrm{e}^{2\omega j} - \mathrm{e}^{2\omega j + 2n\omega j} + \mathrm{e}^{2n\omega j} - 1}{2 - 2\cos(2\omega)}. \tag{21}$$

Thus,

$$a = \frac{\cos(2\omega) - \cos(2\omega + 2n\omega) + \cos(2n\omega) - 1}{2 - 2\cos(2\omega)}, \tag{22}$$

$$2b = \frac{\sin(2\omega) - \sin(2\omega + 2n\omega) + \sin(2n\omega)}{2 - 2\cos(2\omega)}, \tag{23}$$

and

$$\sqrt{a^2 + 4b^2} = \sqrt{\sum_{k=1}^{n} \mathrm{e}^{2k\omega j} \overline{\sum_{k=1}^{n} \mathrm{e}^{2k\omega j}}} \tag{24a}$$

$$= \frac{|\ 2\sin(n\omega)\sin(\omega)\ |}{1 - \cos(2\omega)}, \tag{24b}$$

where $\overline{\sum_{k=1}^{n} \mathrm{e}^{2k\omega j}}$ stands for the complex conjugate of $\sum_{k=1}^{n} \mathrm{e}^{2k\omega j}$.

Substituting (22), (23), and (24b) into $\bar{\rho}$, we have

$$\bar{\rho} = \rho_1^T + \frac{-\cos(2\omega) + 2\cos^2((n+1)\omega) - \cos(2n\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T$$
$$\pm \frac{4\sin(n\omega)\sin(\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T. \tag{25}$$

By the above, it is strightforward to confirm that

$$\bar{\rho}_k = \bar{\rho}_{n-k+1} \tag{26}$$

when

$$\bar{\rho} = \rho_1^T + \frac{-\cos(2\omega) + 2\cos^2((n+1)\omega) - \cos(2n\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T$$
$$+ \frac{4\sin(n\omega)\sin(\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T, \tag{27}$$

and

$$\bar{\rho}_k = -\bar{\rho}_{n-k+1} \tag{28}$$

when

$$\bar{\rho} = \rho_1^T + \frac{-\cos(2\omega) + 2\cos^2((n+1)\omega) - \cos(2n\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T$$
$$- \frac{4\sin(n\omega)\sin(\omega)}{\sin(2\omega) - \sin(2n\omega + 2\omega) + \sin(2n\omega)}\rho_2^T. \tag{29}$$

We conclude from (26) and (28) that $\bar{\rho}$ is symmetric or skew-symmetric.

Theorem 1 demonstrates that the optimal input of (10) is symmetric or skew-symmetric for a special case. Thus, a periodic input turns out to be unreasonable, though it can facilitate a convex reconstruction of (10).

Other contributions such as Fujimoto et al. (2018) proposed interesting input design criteria for the kernel-based identification. However, the problems they work on all suffer the non-convex issue, which leads to difficulties on qualifying the solution. In this paper, we tend to work on a convex input design problem which let us able to search the input in a more general vector space besides the periodic input.

4. MATHEMATICAL PRELIMINARIES FOR STOCHASTIC PROCESS

We firstly summarize the parameterization of a spectrum and some useful results respect to the signal spectrum as mathematical preliminaries.

4.1 Parameterization of the input spectrum

The spectrum of a stochastic signal u is defined as

$$\Phi_u(\omega) = \sum_{k=-\infty}^{\infty} r_{|k|}\mathrm{e}^{-j\omega k},$$

where $r_k := \mathrm{E}(u(t)u(t-k))$ denote the autocorrelation coefficients of u.

The spectrum of a real stochastic process is an even function of frequency which describes the expected energy distributed over the frequency domain. Thus, the coefficients r_k must satisfy

$$\Phi_u(\omega) \geq 0 \quad \forall w.$$

The spectrum is rewritten as

$$\Phi_u(\omega) = \Psi_u(\omega) + \Psi_u^*(\omega),$$

where

$$\Psi_u(\omega) := \frac{1}{2}r_0 + \sum_{k=1}^{\infty} r_k \mathrm{e}^{j\omega k}.$$

The positive constraint of the spectrum allows a linear matrix inequality (LMI) representation by the generalized Kalman-Yakubovich-Popov (KYP) lemma, which will be discussed in the next subsection.

However, in practice, the spectrum is hard to be parameterized by infinite autocorrelation coefficients. One possible approach to handle this problem is to use the finite-coefficients parameterization, which means that the spectrum is parameterized as

$$\Phi_{fu}(\omega) = \sum_{k=-n+1}^{n-1} r_{|k|}\mathrm{e}^{-j\omega k}.$$

Similarly, $\Phi_{fu}(\omega)$ should satisfy nonnegative constraints, i.e., $\Phi_{fu}(\omega) \geq 0\ \forall\omega$. Because the autocorrelation coefficients are symmetric, $r_i = r_{-i}$, we are able to merely consider the positive part of the input spectrum and parameterize the spectrum equivalently by

$$\Phi_{fu}(\omega) = \Psi\left(\mathrm{e}^{j\omega}\right) + \Psi\left(\mathrm{e}^{j\omega}\right)^*, \tag{30}$$

$$\Psi\left(\mathrm{e}^{j\omega}\right) = \frac{1}{2}r_0 + \sum_{k=1}^{n-1} r_k \mathrm{e}^{j\omega}. \tag{31}$$

4.2 The representation via KYP lemma

Several kinds of inequalities regarding the frequency variable can be boiled down to an LMI form. Here, we first introduce the following lemma for the positive definiteness on the whole frequencies.

Lemma 2. (Lemma 2.1 in Jansson and Hjalmarsson (2005)) Let (A, B, C, D) be a controllable state-space realization of $\Psi\left(\mathrm{e}^{j\omega}\right) = \frac{1}{2}r_0 + \sum_{k=1}^{n-1} r_k \mathrm{e}^{j\omega}$. Then,

$$\Phi_{fu}(\omega) = \Psi\left(\mathrm{e}^{j\omega}\right) + \Psi(\mathrm{e}^{j\omega})^* \geq 0 \; \forall\omega,$$

if and only if there exists a symmetric matrix $Q = Q^T$ such that

$$\mathcal{K}\left(Q, (A, B, C, D)\right) := \begin{bmatrix} Q - A^T QA & -A^T QB \\ -B^T QA & -B^T QB \end{bmatrix} + \begin{bmatrix} 0 & C^T \\ C & D + D^T \end{bmatrix} \succeq 0. \quad (32)$$

Lemma 2 inspires us how to handle the nonnegative constraint of the input spectrum. If we have the input spectrum which is parameterized like (30) and (31), then a controllable canonical realization (A, B, C, D) of (31) is

$$A = \begin{bmatrix} O_{1\times n-2} & 0 \\ I_{n-2} & O_{n-2\times 1} \end{bmatrix}, \; B = [1 \; 0 \; \dots \; 0]^T, \quad C = [r_1 \; r_2 \; \dots \; r_{n-1}], \; D = \frac{1}{2}r_0. \quad (33)$$

Analogous to Lemma 2, there is a similar result to translate the inequality constraints on a finite frequency band to another form.

Lemma 3. Let (A, B, C, D) be a controllable state-space realization of $H\left(\mathrm{e}^{j\omega}\right)$. Define the finite frequency interval

$$\Lambda := \{\mathrm{e}^{j\omega} \mid \omega_2 \leq \omega \leq \omega_1\}.$$

Let Ω be the set of eigenvalues of A in Λ. Then, the following statements are equivalent.

i) For each $\lambda \in \Lambda\backslash\Omega$, we have

$$H(\lambda) + H^*(\lambda) \geq 0.$$

ii) There exist $P, Q \in \mathcal{H}_n$ such that $Q \succeq 0$ and

$$\mathcal{K}_a\left(P, Q, (A, B, C, D)\right) := \begin{bmatrix} 0 & C^T \\ C & D + D^T \end{bmatrix} - \begin{bmatrix} A & B \\ I & 0 \end{bmatrix}^* (\Sigma \otimes P + \Theta \otimes Q) \begin{bmatrix} A & B \\ I & 0 \end{bmatrix} \succeq 0. \quad (34)$$

Here

$$\Sigma := \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \; \Theta := \begin{bmatrix} 0 & \mathrm{e}^{j\omega_c} \\ \mathrm{e}^{-j\omega_c} & -2\cos\left(\omega_m\right) \end{bmatrix},$$

where $\omega_c := \left(\omega_1 + \omega_2\right)/2$, and $\omega_m := \left(\omega_2 - \omega_1\right)/2$.

Proof. This is a simple application of Theorem 4 in Iwasaki and Hara (2005).

Lemma 3 reveals the relationship between the constraints on a finite frequency interval and an equivalent condition. This translation facilitates a convex formulation of the input design problem.

4.3 Maximum entropy extension

When partial autocorrelation coefficients $\{r_0, r_1, \dots, r_{n-1}\}$ have been obtained, the extension $\{r_n, r_{n+1}, \dots\}$ should be explored to make sure that the corresponding parameterization can define a spectrum. One rational extension is called maximum entropy extension. The maximum entropy extension considers determining a spectrum $\Phi_u(\omega)$ of a discrete-time zero-mean stationary stochastic process u via maximizing the entropy subject to given partial autocorrelation elements. The unique solution to this problem is the spectrum given by

$$\Phi_u(\omega) = \mu^2 \mid a(\mathrm{e}^{j\omega}) \mid^{-2},$$

where

$$a\left(\mathrm{e}^{j\omega}\right) := \sum_{k=0}^{n-1} a_k \mathrm{e}^{jk\omega}, \;\; a_0 = 1.$$

The coefficients $\{a_1, a_2, \dots, a_{n-1}\}$ and μ are determined via the so-called Yule-Walker equations, i.e.,

$$[a_{n-1} \; \dots \; a_1 \; 1] \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_2 & \dots & r_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{bmatrix} = \mu^2 \left[0 \; \dots \; 0 \; 1\right].$$

In this paper, we choose to use the maximum entropy extension after we have obtained partial autocorrelation coefficients of the designed input.

5. PROBLEM FORMULATION VIA STOCHASTIC INPUT

5.1 Stochastic input design problem formulation

In this subsection, instead of designing a deterministic input signal, we alternatively consider deriving a stationary zero-mean Gaussian stochastic process, which will facilitate the convexity of the input design problem.

If we were able to use the law of large numbers, the long-time average converges in probability to the expected value as observed data size converges to infinity. Thus, the finite autocorrelations matrix, i.e.,

$$R := \mathcal{T}(r_0, r_1, \dots, r_{n-1}) = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_0 & \dots & r_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{bmatrix}, \quad (35)$$

can approximate the average of the matrix $\frac{1}{N-n}\Phi_N \Phi_N^T$ as N to be large enough. The limitation of this approximation is that it will only be valid with a sufficiently large data size which may be an unnatural setting in Bayesian identification. However, when N is finite, the stochastic input signal generated by the designed stationary process still satisfy the ordinary variance constraint to some extent. The probability of sampling such good inputs has been studied, which will be explained in our subsequent paper.

With the asymptotic approximation R of the matrix $\frac{1}{N-n}\Phi_N \Phi_N^T$, the original variance constraint in (10) admits the approximation as follows:

$$\Lambda_G P \Lambda_G^* \approx \Lambda_G \left(\frac{(N-n)R}{\sigma^2} + K^{-1}\right)^{-1} \Lambda_G^* \leq \alpha, \quad (36)$$

Then, we formulate the minimum power stochastic input design problem regarding a variance constraint.

Problem 4. Suppose that the kernel K, the noise variance σ^2, and the credible frequency interval $[\omega_1, \omega_2]$ are given. Let $\alpha > 0$ denote a system error which should be decided preliminarily. The identification input is generated from a stationary zero-mean Gaussian process $u(t)$. Our objective is to design the spectrum $\Phi_u(\omega)$ of $u(t)$ which minimizes the input power r_0 subject to the variance constraint (36) for all $\omega \in [\omega_1, \omega_2]$.

The problem 4 is expressed in a mathematical way like the following optimization problem:

$$\min_{r} r_0 \tag{37a}$$

$$\text{s.t. } \Lambda_G \left(\frac{(N-n)R}{\sigma^2} + K^{-1} \right)^{-1} \Lambda_G^* \leq \alpha \ \forall \omega \in [\omega_1, \omega_2], \tag{37b}$$

$$\Phi_{fu}(\omega) \geq 0 \ \forall \omega. \tag{37c}$$

Here, the variable r is defined to be $r := [r_0 \ r_1 \ \dots \ r_{n-1}]$. Via solving this problem (37), we will obtain finite autocorrelation coefficients r. However, the problem (37) is not a convex problem per se. We will explain how to convert this problem to a convex problem on r via Lemma 2 and Lemma 3.

5.2 Convex representation of input design problem

Section 4.2 reveals that the frequency inequalities (37b) and (37c) are equivalent to certain linear matrix inequalities.

First, let us concentrate on the constraint (37c). Using the controllable canonical realization (33) of the positive real part (31), the problem (37) is equivalent to the following form:

$$\min_{r,Q} r_0 \tag{38a}$$

$$\text{s.t. } \Lambda_G \left(\frac{(N-n)R}{\sigma^2} + K^{-1} \right)^{-1} \Lambda_G^* \leq \alpha \ \forall \omega \in [\omega_1, \omega_2], \tag{38b}$$

$$\mathcal{K}\left(Q, (A, B, C, D)\right) \succeq 0, \tag{38c}$$

$$Q = Q^T, \tag{38d}$$

where the function $\mathcal{K}\left(Q, (A, B, C, D)\right)$ is defined like (32).

With the realization (A, B, C, D) like (33), the inequality regarding the infinite frequency domain, i.e., (37c), becomes an LMI (38c) where the decision variables are the symmetric matrix Q and autocorrelation coefficients $r = [r_0 \ r_1 \ \dots \ r_{n-1}]$.

Then let us explain how to convert the constraint (38b) to another LMI. The Schur complement shows that the finite frequency inequality (38b) is equal to

$$T\left(\mathrm{e}^{j\omega}\right) := \begin{bmatrix} \frac{\alpha}{\sigma^2} & \Lambda_G \\ \Lambda_G^* & (N-n)R + \sigma^2 K^{-1} \end{bmatrix} \succeq 0 \ \forall \omega \in [\omega_1, \omega_2]. \tag{39}$$

Then, it is straightforward to check that $T^* = T$ and $T + T^* \succeq 0 \ \forall \omega \in [\omega_1, \omega_2]$. Thus, Lemma 3 is used here to derive (39) to an equivalent LMI like (34), which is explained in the following paragraphs.

Let (A_G, B_G, C_G, D_G) be a controllable canonical realization of the multi-input-single-output transfer function Λ_G. Then a controllable canonical realization of the positive real part of $T(\mathrm{e}^{j\omega})$ is $(A_\Gamma, B_\Gamma, C_\Gamma, D_\Gamma)$, where

$$A_\Gamma = A_G, \ \ B_\Gamma = [O_{n\times 1} \ B_G], \ \ C_\Gamma = \begin{bmatrix} C_G \\ O_{n\times n} \end{bmatrix}, \\ D_\Gamma + D_\Gamma^T = \begin{bmatrix} \frac{\alpha}{\sigma^2} & D_G \\ D_G^T & (N-n)R + \sigma^2 K^{-1} \end{bmatrix}. \tag{40}$$

Thus, the inequality (39) holds if and only if there exist $P_1, Q_1 \in \mathcal{H}_n$ such that $Q_1 \succeq 0$ and

$$\mathcal{K}_a\left(P_1, Q_1, (A_\Gamma, B_\Gamma, C_\Gamma, D_\Gamma)\right) \succeq 0. \tag{41}$$

Considering the definition of Λ_G, one example of the controllable canonical realization (A_G, B_G, C_G, D_G) is

$$A_G = \begin{bmatrix} O_{n-1\times 1} & I_{n-1} \\ 0 & O_{1\times n-1} \end{bmatrix}, \ \ B_G = I_n, \\ C_G = [1 \ O_{1\times n-1}], \ \ D_G = O_{1\times n}. \tag{42}$$

Applying the realization (42) to (41), one can check that (41) is also an LMI in P_1, Q_1, and r.

Based on the explanation above, finally we can summarize the following convex optimization problem which is equal to the original problem (37):

$$\min_{r,Q,Q_1,P_1} r_0 \tag{43a}$$

$$\text{s.t. } \mathcal{K}_a\left(P_1, Q_1, (A_\Gamma, B_\Gamma, C_\Gamma, D_\Gamma)\right) \succeq 0, \tag{43b}$$

$$\mathcal{K}\left(Q, (A, B, C, D)\right) \succeq 0, \tag{43c}$$

$$Q = Q^T, \tag{43d}$$

$$P_1 = P_1^T, \tag{43e}$$

$$Q_1 = Q_1^T, \tag{43f}$$

$$Q_1 \succeq 0. \tag{43g}$$

The problem (43) is a positive semidefinite problem which is efficiently solved by a Matlab package named 'YALMIP'.

After solving the problem (43), the complete spectrum $\Phi_u(\omega)$ is obtained through the maximum entropy extension according to the finite autocorrelation coefficients r.

5.3 Procedures for stochastic input design

Based on the explanation above, we can summarize the procedures for sampling a stochastic input as follows:

(1) Step 1: Perform a preliminary experiment to estimate the kernel K, the noise variance σ^2, and the credible frequency interval $[\omega_1, \omega_2]$.
(2) Step 2: Solve the optimization problem (43) to obtain the finite autocorrelation coefficients r via convex optimization techniques such as 'YALMIP' in Matlab. Here, the matrices (A, B, C, D) and $(A_\Gamma, B_\Gamma, C_\Gamma, D_\Gamma)$ have been defined in (33) and (40) respectively.
(3) Step 3: Calculate the complete spectrum $\Phi_u(\omega)$ using the maximum entropy extension (Section 4.3).
(4) Step 4: Sample a signal from $\Phi_u(\omega)$ as the identification input.

6. SIMULATIONS

6.1 Test functions

In numerical experiments, a test system is defined as follows:

$$F(z) = \frac{z-2}{(z-0.62-0.62j)(z-0.62+0.62j)}.$$

The test system $F(z)$ has a pair of complex eigenvalues which will lead to a resonance phenomenon in Fig. 1.

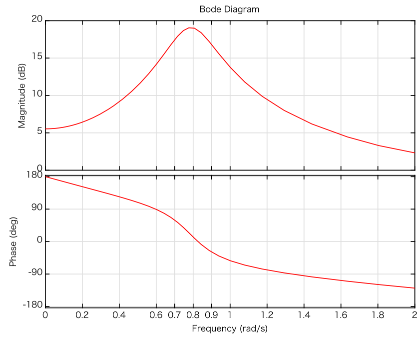


Fig. 1. Bode diagram for the test system $F(z)$.

We also plot the impulse response of $F(z)$ in Fig. 2. It can be seen from Fig. 2 that the impulse response decreases to zero at around the 50th step. Thus, the system order n is set to be 50, while the observed data size N should be fixed bigger than 50 in the following experiments.

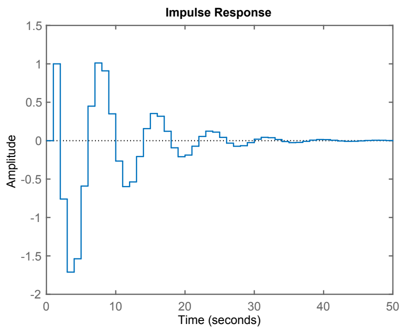


Fig. 2. The impulse response for the test system $F(z)$.

6.2 Preliminary setting

A preliminary experiment is needed to obtain the kernel K and the noise variance σ in the input design problem 4. In this paper, we select the kernel to be the DC kernel which is written as follows:

$$K(i,j) := k_1 k_2^{\frac{i+j}{2}} k_3^{|i-j|},$$

where k_1, k_2, and k_3 are hyperparameters. The hyperparameters and σ are effectively estimated by the empirical Bayes method. Then, the credible frequency set is decided to be $[0.6, 0.9]$ in all of the following experiments.

It should be noticed that the preliminary experiment is conducted with a random small energy input which may lead to a rough identification. However, the rough identification result also contains some system information which can instruct the input design problem.

6.3 Numerical experiment: bode diagrams

In this experiment, we first solve the problem (43) with $n = 50$, $\alpha = 20$, while N is set to be 51, 100, and 1000, respectively. Then, the finite correlation coefficients r are extended using the technique mentioned in Section 4. The complete spectrum $\Phi_u(\omega)$ is obtained after the extension. The proposed identification inputs denoted by filtered G are generated by sampling $\Phi_u(\omega)$. By comparison, we consider another white Gaussian process which has the same power r_0 with $\Phi_u(\omega)$ and denote the samples from the white Gaussian process by white G. We conduct 1000 identification experiments using the filtered G and the white G respectively. Then, the estimations of $F(z)$ using the filtered G and the white G are denoted by $\hat{G}_f$ and $\hat{G}_w$, respectively. The bode diagrams of $\hat{G}_f$ and $\hat{G}_w$ are shown in Fig. 3.

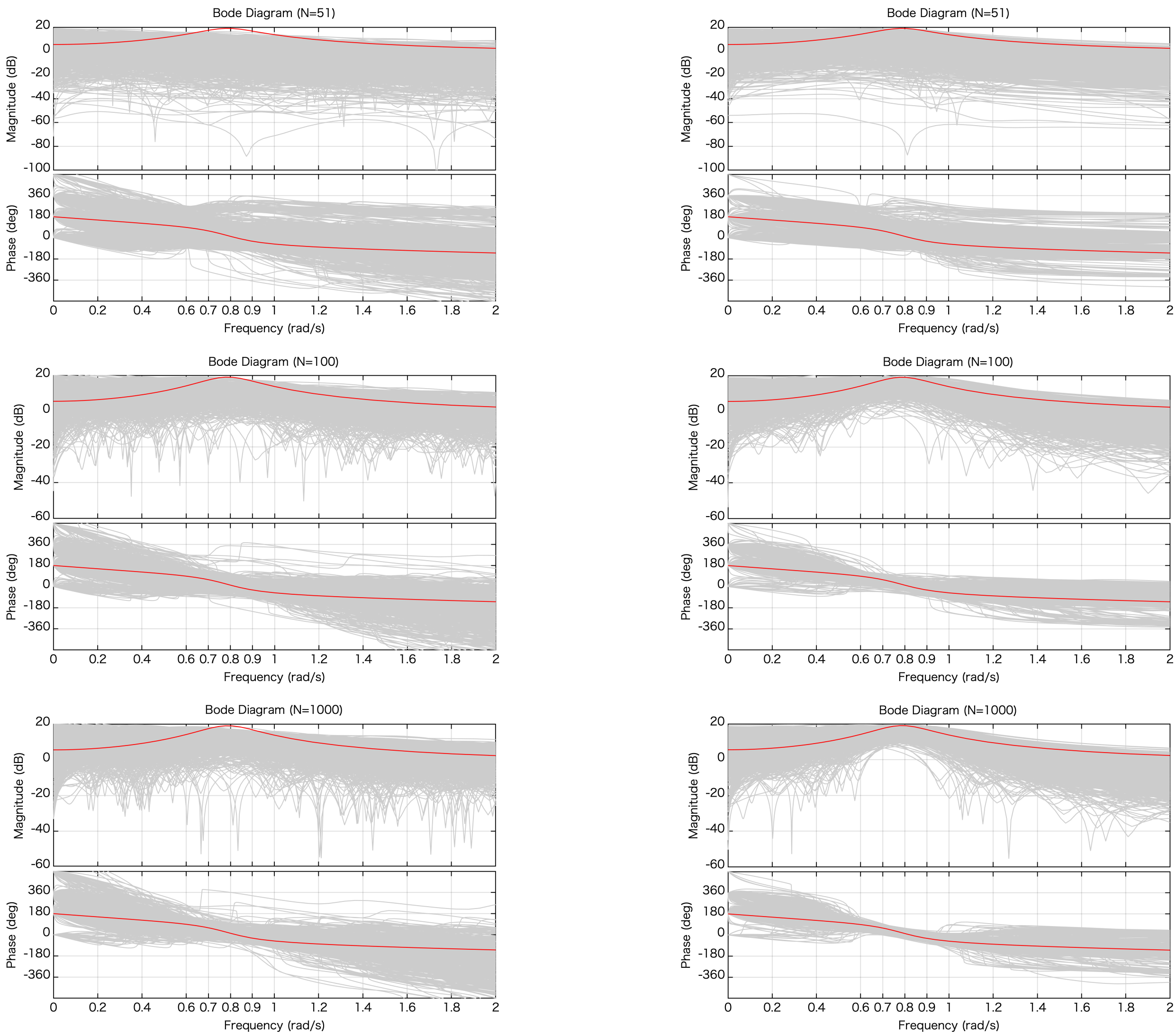
Generally speaking, in all experiments, it can be seen that the estimations $\hat{G}_f$ are more precise and more stable than the estimations $\hat{G}_w$ statistically. The identification when $N = 51$ is the most difficult case whose results are shown in the top row of Fig. 3. We can find that both filtered G and white G perform unsatisfactorily because the data is too few to provide enough information for the identification. However, the bode diagrams of $\hat{G}_f$ is closer to the true system than $\hat{G}_w$ generally. When $N = 100$, we give the identification results in the medium row of Fig. 3. As the increase of data size N, the power r_0 of filtered G and white G decreases. Nevertheless, the matrix $\frac{1}{N-n}\Phi_N\Phi_N^T$ made by filtered G is closer to the matrix R, while $\frac{1}{N-n}\Phi_N\Phi_N^T$ made by white G is closer to $r_0 I$ generally. Furthermore, the matrix R is the optimal solution of (43). Thus, filtered G performs more stable and more precise, especially on the credible frequency $[0.6, 0.9]$. When $N = 1000$, the matrices $\frac{1}{N-n}\Phi_N\Phi_N^T$ made by filtered G and white G are further close to R and $r_0 I$, respectively. However, the performance of white G does not improve significantly compared with the case of $N = 100$. By comparison, filtered G makes a further improvement, especially on the credible frequency $[0.6, 0.9]$.

Although a particular input in filtered G is not always better than white G, comparing the energy $u^T u$ and the posterior variance $\Lambda_G P \Lambda_G^*$ of the sampled u leads to a rough selection of the good inputs. A good input satisfies the variance constraint (10b) with the smaller energy. The probability of sampling such good input is discussed in the journal version of this paper.

7. CONCLUSION

In this research, we formulated an input design problem for Bayesian identification with variance constraints over a frequency band. When the input is a deterministic signal, the problem is a nonlinear optimization problem. In the first part of the paper, it was shown that the optimal solution is, in general, not a periodic function of time. This suggests that though the restriction of the input signals to periodic functions makes the problem convex, it may not yield a satisfactory signal.

In the second part of the paper, we proposed to use a stationary stochastic process to reformulate the problem into a convex problem. By exploiting the generalized KYP lemma, the input design problem was cast as an LMI problem whose decision variables are correlation coefficients. Then, the maximum entropy extension allows us to build the complete spectrum of the proposed stochastic process.



(a) white Gaussian input

(b) filtered Gaussian input

Fig. 3. Bode diagrams of estimated systems $\hat{G}_w$ (left) and $\hat{G}_f$ (right) without selection. The red line denotes the true system $F(z)$.

The simulation results demonstrate the effectiveness of our method.

ACKNOWLEDGEMENTS

This work was supported in part by JSPS Grants-in-Aid for Scientific Research (C) 20K04534.

REFERENCES

Cantoni, A. and Butler, P. (1976). Properties of the eigenvectors of persymmetric matrices with applications to communication theory. *IEEE Transactions on Communications*, 24(8), 804–809.

Chen, T., Ohlsson, H., and Ljung, L. (2012). On the estimation of transfer functions, regularizations and Gaussian processes — revisited. *Automatica*, 48(8), 1525–1535.

Fujimoto, Y., Maruta, I., and Sugie, T. (2018). Input design for kernel-based system identification from the viewpoint of frequency response. *IEEE Transactions on Automatic Control*.

Fujimoto, Y. and Sugie, T. (2018). Informative input design for kernel-based system identification. *Automatica*, 89, 37–43.

Iwasaki, T. and Hara, S. (2005). Generalized KYP lemma: unified frequency domain inequalities with design applications. *IEEE Transactions on Automatic Control*, 50(1), 41–59.

Jansson, H. and Hjalmarsson, H. (2005). Input design via LMIs admitting frequency-wise model specifications in confidence regions. *IEEE transactions on Automatic Control*, 50(10), 1534–1549.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, NJ.

Mu, B. and Chen, T. (2018). On input design for regularized LTI system identification: power-constrained input. *Automatica*, 97, 327–338.

Zheng, M. and Ohta, Y. (2018). Positive FIR system identification using maximum entropy prior. In *the 18th IFAC Symposium on System Identification*.