Robustness in Experiment Design

Cristian R. Rojas, Juan-Carlos Agüero, *Member, IEEE*, James S. Welsh, *Member, IEEE*, Graham C. Goodwin, *Fellow, IEEE*, and Arie Feuer, *Fellow, IEEE*

Abstract—This paper focuses on the problem of robust experiment design, i.e., how to design an input signal which gives relatively good estimation performance over a large number of systems and model structures. Specifically, we formulate the robust experiment design problem utilizing fundamental limitations on the variance of estimated parametric models as constraints. Using this formulation we design an input signal for situations where only diffuse a priori information is known about the system. Furthermore, we present a robust version of the unprejudiced optimal input design problem. To achieve this, we first develop a closed form solution for the input spectrum which minimizes the maximum weighted integral of the variance of the frequency response estimate over all model structures.

Index Terms—Optimal input design, robust input design, system identification.

I. INTRODUCTION

ADVANCED control design is based on the availability of models for the system under study [1]. The success of these techniques depends on the quality of the models utilized to design the control strategy. This has, inter-alia, inspired interest in the area of *system identification* over the last three decades (e.g., see [2]–[6]).

A general requirement in system identification is to learn as much about a system as possible from a given observation period. This has motivated substantial interest in the area of optimal experiment design. Optimal experiment design has been studied both in the statistics literature [7]–[9] and in the engineering literature [3], [10]–[14], primarily focusing on the goals of system identification.

Most of the existing literature is based on designing the experiment to optimize a given scalar function of the Fisher Information Matrix [3]. This presents a fundamental difficulty, namely, when the system response depends nonlinearly on the parameters, the Information Matrix depends on the true system parameters. Moreover, models for dynamical systems (even if

linear) typically have the characteristic that their response depends nonlinearly on the parameters. Hence, the information matrix for models of dynamical systems generally depends upon the true system parameters. Therefore experiment designs based on the Fisher Information Matrix will, in principle, depend upon knowledge of the true system parameters. This is paradoxical since the "optimal experiment" then depends on the very thing that the experiment is aimed at estimating [13].

The above reasoning has motivated the study of "robust" experiment design with respect to uncertain *a priori* information. Work in this area has been growing in recent years [15]–[22]. In particular, in [15] and [22] experiments are designed for the purpose of robust control and the term *least-costly identification experiment* was coined. Based on geometric properties of the information matrix, a modification of the least-costly identification experiment approach is analyzed in [19] and [20]. Finally, in [21] the experiments are designed in a game-theory framework where the "true" system parameters belong to a compact set. Here, we propose an alternative approach motivated by the analysis presented in [23], and the recent results in [24].

In general, the choice of the "best" experiment to identify a process depends on the prior knowledge we have about the process. In this paper we analyze, and solve, the following problem: *Say we are just beginning to experiment on a system and thus have very little (i.e., diffuse) prior knowledge about it. What would be a "good" initial experiment in order to estimate the parameters of the system?*

To this end, we build on works such as [12] and [23], which assume that both the true system and the noise dynamics are known (at the time of designing the experiment). In this paper, we do not assume knowledge of the true system, but (for the results of Section VI) we do assume that the noise spectrum is known. Basic prior knowledge about the plant can be obtained, for example, by using nonparametric frequency domain methods based on a simple experiment [4]–[6], [25], [26]; however, the use of this kind of prior knowledge for robust experiment design is not considered in the present contribution, and it will be explored in a future publication.

The results derived in this paper are valid for models with a finite number of parameters, but where the number of samples N is sufficiently large. The possibility of removing this last condition and keeping at the same time the simplicity of our expressions, which have a very intuitive interpretation, is beyond the scope of this paper, since they are much more difficult to obtain, and most approaches to finite sample analysis are done using numerical techniques [27]. Moreover, exact variance results for finite samples also rely on higher order moments of the underlying distribution of the data, so they are inherently less robust with respect to the assumptions on the true system than asymptotic results.

The paper is structured as follows. Section II formulates the problem and describes the notation used in the paper. In

Section III, the basic tools used in the sequel are presented. Section IV deals with the design of an input signal based on diffuse prior information. In Section V we study the problem of designing an input spectrum having normalized power which minimizes a weighted integral of the variance of the frequency response of a model. Section VI revisits the problem of unprejudiced input design. Section VII presents a simple numerical example. Finally, Section VIII provides conclusions.

II. SYSTEM DESCRIPTION AND NOTATION

Consider a single-input single-output (SISO) linear system given by

$$y(t) = G_0(q^{-1})u(t) + H_0(q^{-1})w(t)$$

where $\{u(t)\}$ is a quasi-stationary signal [5], and $\{w(t)\}$ is a zero mean Gaussian white noise sequence with variance σ^2. We denote the unit delay operator by q^{-1} and assume H_0 to be a stable minimum phase transfer function with $H_0(0) = 1$. To simplify the notation, we denote $H_0(q^{-1})w(t)$ by $v(t)$.

Given N input-output data pairs $\{u(t), y(t)\}_{t=1}^{N}$, a model of the form

$$y(t) = G(q^{-1}, \theta)u(t) + H(q^{-1}, \theta)\epsilon(t)$$

will be estimated. Here the dimension of θ will specify the *order* of the model.

We assume that the estimators for G_0 and H_0 are asymptotically efficient (e.g., Maximum Likelihood, or PEM for Gaussian disturbances). Note that this is not a limitation, since there are standard estimation methods which satisfy this condition. This assumption allows us to decouple the problems of experiment design and estimation (c.f. [3, Sec. 6.2]).

The spectrum of a quasi-stationary signal $\{x(t)\}$ [5] is defined as

$$\Phi_x(\omega) := \sum_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau}, \quad \omega \in [-\pi, \pi]$$

where $R_x(\tau) := \bar{E}\{x(t)x(t-\tau)\}$ is the autocovariance of $\{x(t)\}$ [5], and where $\bar{E}\{f_t\} := \lim_{N\to\infty}(1/N)\sum_{t=1}^{N} E\{f_t\}$.

Notation: If $x \in \mathbb{C}^{n \times m}$, then $\bar{x}$, x^T and x^H denote its complex conjugate, transpose and complex conjugate transpose, respectively. Let $\mathbb{E} := \{z \in \mathbb{C} : |z| > 1\}$ and $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$. The Hardy space of analytic functions f on $\mathbb{E}$ taking values on $\mathbb{R}^n$ such that $\lim_{r \to 1_+} \int_{-\pi}^{\pi} \|f(re^{j\omega})\|_2^2 d\omega < \infty$ is denoted as $\mathcal{H}_2^n$ [28], [29]. Now define $C^1(X, Y)$ as the space of all functions from $X \subseteq \mathbb{R}$ to $Y \subseteq \mathbb{R}$ having a continuous derivative, and $C(\mathbb{T}, \mathbb{R}_0^+)$ as the space of all continuous functions $f : \mathbb{T} \to \mathbb{R}_0^+$ such that $f(\bar{z}) = \overline{f(z)}$ for every $z \in \mathbb{T}$.

In the sequel, quantities with a hat "$\hat{}$" correspond to estimators (of their respective "unhatted" quantities), which implicitly depend on the data length, N. Covariance expressions are valid as $N \to \infty$ [5] (i.e., they are correct up to order[1] $1/N$).

III. TECHNICAL PRELIMINARIES

The results presented below depend upon a fundamental limitation result developed in [24]. For completeness, we state the

[1]Loosely speaking, this means that all expressions in the sequel which involve variances, of the form $A = (\geq)B$, should be interpreted as $\lim_{N\to\infty} NA = (\geq)\lim_{N\to\infty} NB$.

main result of [24]. In this section, we assume there is no undermodelling, i.e., there exists a $\theta = \theta_0$ such that $G_0(q^{-1}) = G(q^{-1}, \theta_0)$ and $H_0(q^{-1}) = H(q^{-1}, \theta_0)$. In addition, G and H are independently parameterized,[2] and the vector of true parameters θ_0 is split into two components, i.e., $\theta_0 := [\theta_{G_0}^T \; \theta_{H_0}^T]^T$. Under these and some additional mild assumptions [5, Sec. 9.4], the estimate of $G_0(e^{j\omega})$, $\hat{G}(e^{j\omega}) := G(e^{j\omega}, \hat{\theta})$, has an asymptotic variance satisfying

$$\text{Var}\left[\hat{G}(e^{j\omega})\right] = \Gamma^H(e^{j\omega})$$
$$\times \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau\right]^{-1} \Gamma(e^{j\omega})$$

where $\Gamma(e^{j\omega}) := \partial G(e^{j\omega}, \theta)/\partial\theta|_{\theta=\theta_0}$.

Theorem 1: If the parameter vector of G, θ_G, has dimension n_G, and $G(q, \theta_G)$ is *parameter identifiable* under Φ_u for the maximum likelihood method [4],[3] then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)}\text{Var}\left[\hat{G}(e^{j\omega})\right] d\omega = \frac{n_G}{N}.$$

Proof: See [24]. ∎

As explained in detail in [24], Theorem 1 shows that it is not possible to reduce the variance of $\hat{G}$ uniformly at all frequencies by choosing a suitable model structure, since if we reduce the variance at some frequencies, it will necessarily increase at others, thus implying a "water-bed" effect [30]. Additionally, any over-parameterization of G results in an increase in the integrated variance of its estimate.

The following converse to Theorem 1 will prove useful in the sequel.

Theorem 2: Let $\Phi_u, \Phi_v : [-\pi, \pi] \to \mathbb{R}^+$ be continuous and even. Also, let $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ be such that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)}V(e^{j\omega})d\omega = \frac{n}{N} \qquad (1)$$

where $n, N \in \mathbb{N}$. Then, there exists a function $\Gamma \in \mathcal{H}_2^n$ such that

$$\Gamma^H(z)\left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau\right]^{-1} \Gamma(z) = V(z)$$
$$(2)$$

for every $z \in \mathbb{T}$.

Proof: See Appendix A. ∎

Theorem 2 shows that if a function $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ satisfies relation (1), which is similar to the fundamental limitation of Theorem 1, then it is possible to find a model structure for G with n parameters for which V is the variance of $\hat{G}$. In this case, the resulting model structure is characterized by Γ, the gradient of G with respect to θ_G. For instance, given Γ from

[2]Having an independent parameterization for G and H means that θ can be split into two components, θ_G and θ_H, such that $G(q, \theta)$ functionally depends only on θ_G, and $H(q, \theta)$ functionally depends only on θ_H.

[3]The assumption that $G(q, \theta_G)$ is parameter identifiable under Φ_u for the maximum likelihood (ML) method means the ML estimator of θ_G converges almost surely to θ_{G_0}, where $G(q, \theta_{G_0}) = G_0$.

Theorem 2, a model structure for which $\mathrm{Var}[\hat{G}(e^{j\omega})] = V(e^{j\omega})$ is[4] $G(q, \theta_G) = \theta_G^T \Gamma(q)$.

From Theorems 1 and 2, we have that (1) provides a complete characterization of those functions V which could correspond to the variance of $\hat{G}$.

Note that the parameters involved in the fundamental limitations of Theorems 1 and 2 must be evaluated at their true values. Also notice that the assumption that Φ_u and Φ_v are continuous and nonzero in Theorem 2 might seem restrictive, since it does not allow, for example, multisine inputs. However, it is a standard assumption for the derivation of several variance results, e.g., see [31].

IV. EXPERIMENT DESIGN WITH DIFFUSE PRIOR INFORMATION

The problem of designing a good input signal with diffuse prior information was examined in [21], where results based on Ljung's asymptotic (in the number of data points, and also in the number of parameters in the model) variance expression were obtained. In this section the results of [21] are shown to be valid, even for finite model orders, if we consider the worst case over all model structures of a given order. Again, note that we assume no undermodelling.

Our aim is to design an experiment which is "good" for a very broad class of systems. This means that we need a measure of "goodness" of an experiment which is system independent.

As argued in [16], [32]–[34], absolute variances are not particularly useful when one wants to design an experiment that applies to a broad class of systems. Specifically, an error standard deviation of 10^{-2} in a variable of nominal size 1 would be considered to be insignificant, whereas the same error standard deviation of 10^{-2} in a variable of nominal size 10^{-3} would be considered catastrophic. Hence, it seems highly desirable to work with relative errors (see also [35]–[37]).

Rather than look at a single frequency ω, we will look at an "average" measure over a range of frequencies. This leads to a general measure of the "goodness" of an experiment, given by

$$J(\Phi_u) = \int_a^b F\left(\mathrm{Var}\left[\hat{G}(e^{j\omega})\right] \Big/ \left|G(e^{j\omega})\right|^2\right) W(\omega)d\omega \quad (3)$$

where $0 < a < b < 2\pi$ and

$$\mathrm{Var}\left[\hat{G}(e^{j\omega})\right] = \Gamma^H(e^{j\omega})$$
$$\times \left[\frac{N}{2\pi}\int_{-\pi}^{\pi}\Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau\right]^{-1}\Gamma(e^{j\omega}).$$

The functions F and W will be specified later.

Essentially W is a weighting function that allows one to specify at which frequencies it would be preferable to obtain a good model (depending on the ultimate use of the model, but not necessarily on the true system characteristics).

In [21] it is argued that F and W should satisfy the following criteria.

---

[4]$G(q, \theta_G) = \theta_G^T \Gamma(q)$ is not the only model structure for which $\mathrm{Var}[\hat{G}(e^{j\omega})] = V(e^{j\omega})$. However, any model structure with such variance must satisfy $G(q, \theta_G) \approx G(q, \theta_0) + \theta_G^T\Gamma(q)$ locally around $\theta = \theta_0$, for some Γ satisfying (2), by definition of the model gradient Γ and the smoothness of G.

A.1) The optimal experiment, Φ_u^{opt}, which minimizes $\sup_{\Gamma \in \mathcal{H}_2^n} J$ in (3), should be independent of the system G and the noise dynamics Φ_v.

A.2) The integrand in (3) should increase if the variance $\mathrm{Var}[\hat{G}(e^{j\omega})]$ increases at any frequency. This implies that F should be a monotonically increasing function.

B) The weighting function W should satisfy the following:

for every $0 < \alpha < \beta < 2\pi$ and every $k > 0$
s.t. $0 < k\alpha < k\beta < 2\pi$ we have

$$\int_\alpha^\beta W(\omega)d\omega = \int_{k\alpha}^{k\beta} W(\omega)d\omega.$$

Criteria A.1 and A.2 are based on the desire to design an input signal which is independent of the system and the noise dynamics. Criterion B, on the other hand, is based on the observation that many properties of linear systems depend on the *ratio* of poles and zeros rather than on their absolute locations in the frequency domain [1], [30], [38]. This implies that if we scale the frequency ω by a constant, the optimal input must keep the same shape and simply relocate on the frequency axis, since the poles and zeros of the new system will have the same ratios as before.

Note that it is not possible in our framework to consider the full interval $[0, \pi]$, since, as we will see later, the optimal signal which satisfies these criteria in the range $[0, \pi]$ is $1/f$ noise, which has infinite power over this range; hence it is unrealizable in practice. However, the assumption that $0 < a < b < \pi$ in (3) seems reasonable, since for control design, it is well known that knowledge of the plant at low frequencies is unimportant, as the controllers typically include an integrator which takes care of the steady state behavior of the closed loop. Similarly, it is not necessary to estimate the high frequency region of G, since plants are typically low-pass. What is actually required from the control designer in order to use the proposed input signal is a frequency range $[a, b]$ where the relevant dynamics of the plant are believed to be.

Note that Criterion A.1 is not the same as in [21], since we are considering the worst case of J over all possible systems and model structures (of order n).

The purpose of obtaining a robust input with respect to all model structures comes from the fact that the optimal input typically depends on the gradient of the model with respect to the parameter vector θ, evaluated at its true value θ_0. Therefore, for a nonlinearly parameterized model, even though the user knows the model structure (since it is a design variable), the gradient typically depends on the true system, hence it will be unknown prior to the experiment. Of course, the gradient cannot take any possible value in $\mathcal{H}_2^n$ for some particular model structures (e.g., linearly parameterized models, for which it is actually independent of θ). However, in the sense of a fundamental limitation, the results derived in this paper establish a lower bound (and an input spectrum which achieves it) on the performance of the parameter estimation of the system, even before the selection of a model structure.

The following lemma, from [21], describes how W must be chosen to satisfy Criterion B.

Lemma 1: For $0 < a < b < 2\pi$, let $W \in C^1([a,b], \mathbb{R}^+)$. If W satisfies

$$\int_\alpha^\beta W(\omega) d\omega = \int_{k\alpha}^{k\beta} W(\omega) d\omega \qquad (4)$$

for every $a \leq \alpha < \beta \leq b$ and every $k > 0$ such that $a \leq k\alpha < k\beta \leq b$, then there exists a $\lambda > 0$ such that $W(x) = \lambda/x$ for every $x \in [a,b]$.

Proof: Since W is continuous, we have from (4) that

$$
\begin{aligned}
W(a) &= \lim_{\varepsilon \to 0_+} \frac{\int_a^{a+\varepsilon} W(\omega) d\omega}{\varepsilon} \\
&= \lim_{\varepsilon \to 0_+} k \frac{\int_{ka}^{ka+k\varepsilon} W(\omega) d\omega}{k\varepsilon} \\
&= kW(ka)
\end{aligned}
$$

for $1 \leq k < b/a$. Thus

$$W(ka) = \frac{1}{k} W(a), \qquad a \leq ka < b$$

or, by defining $x = ka$ and $\lambda = aW(a)$

$$W(x) = \frac{a}{x} W(a) = \frac{\lambda}{x}, \qquad a \leq x < b.$$

By the continuity of W, we also have that $W(b) = \lambda/b$. This proves the lemma. ∎

Criteria A.1 and A.2 constrain W to have a very particular form, as shown in the following lemma.

Lemma 2: Consider the experiment design problem

$$
\min_{\Phi_u > 0} \sup_{\Gamma \in \mathcal{H}_2^n} \int_a^b F\left(\frac{\mathrm{Var}\left[\hat{G}(e^{j\omega})\right]}{|G(e^{j\omega})|^2} \right) W(\omega) d\omega
$$

$$
\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1 \qquad (5)
$$

where $0 < a < b < 2\pi$, $F \in C^1([a,b], \mathbb{R}_0^+)$, $W \in C^1([a,b], \mathbb{R}^+)$, $|G|^2$ is continuously differentiable on $\mathbb{T}$, and

$$
\mathrm{Var}\left[\hat{G}(z)\right] = \Gamma^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau}) \Gamma^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau \right]^{-1} \Gamma(z)
$$

for $z \in \mathbb{T}$. Let Φ_u^{opt} be a stationary point. If Φ_u^{opt} does not depend on G, then there exist constants $\alpha, \beta \in \mathbb{R}$ such that

$$F(y) = \alpha \ln y + \beta$$

$$
\inf_{\omega \in [a,b]} \frac{\mathrm{Var}\left[\hat{G}(e^{j\omega})\right]}{|G(e^{j\omega})|^2} \leq y \leq \sup_{\omega \in [a,b]} \frac{\mathrm{Var}\left[\hat{G}(e^{j\omega})\right]}{|G(e^{j\omega})|^2}
$$

and

$$
\Phi_u^{\mathrm{opt}}(\omega) := \begin{cases} \frac{W(\omega)}{\frac{1}{2\pi} \int_a^b W(\tau) d\tau}, & \omega \in [a,b], \\ 0, & \text{otherwise.} \end{cases}
$$

Proof: See Appendix B. ∎

In the following lemma we establish that, for the choice of F given in Lemma 2, Φ_u^{opt} actually corresponds to the global optimum of the experiment design problem (5).

Lemma 3: Consider the experiment design problem

$$
\min_{\Phi_u > 0} \sup_{\Gamma \in \mathcal{H}_2^n} \int_a^b \ln \left[\frac{\mathrm{Var}\left[\hat{G}(e^{j\omega})\right]}{|G(e^{j\omega})|^2} \right] W(\omega) d\omega
$$

$$
\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1
$$

where $0 < a < b < 2\pi$, $W \in C^1([a,b], \mathbb{R}^+)$, $|G|^2$ is continuously differentiable on $\mathbb{T}$, and

$$
\mathrm{Var}\left[\hat{G}(z)\right] = \Gamma^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau}) \Gamma^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau \right]^{-1} \Gamma(z)
$$

for $z \in \mathbb{T}$. The solution to this problem is given by

$$
\Phi_u^{\mathrm{opt}}(\omega) := \begin{cases} \frac{W(\omega)}{\frac{1}{2\pi} \int_a^b W(\tau) d\tau}, & \omega \in [a,b], \\ 0, & \text{otherwise.} \end{cases} \qquad (6)
$$

Proof: See Appendix C. ∎

Lemma 3 shows that, under Criteria A.1 and A.2, the optimal input has to be proportional to the weighting function W. This means that the input should excite precisely those frequencies where higher model quality is required. This agrees with intuition. Notice also that the optimal input does not depend on the noise spectrum (according to Criterion A.1).

By combining Lemmas 1 and 2, Criteria A.1, A.2, and B imply that, when only diffuse prior knowledge is available about the system and the noise, then a reasonable experiment design problem can be stated as

$$
\min_{\Phi_u > 0} \sup_{\Gamma \in \mathcal{H}_2^n} \int_a^b \ln \left[\frac{\mathrm{Var}\left[\hat{G}(e^{j\omega})\right]}{|G(e^{j\omega})|^2} \right] \frac{1}{\omega} d\omega
$$

$$
\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1.
$$

Moreover, by Lemma 3, the corresponding optimal input spectrum is given by

$$
\phi_u^{\mathrm{opt}}(\omega) = \frac{1/\omega}{\int_a^b \frac{d\omega}{\omega}} = \frac{1/\omega}{\ln b - \ln a}, \qquad \omega \in [a,b]
$$

which is bandlimited "$1/f$" noise [16]. This extends the results of [21] to finite model orders.

The result presented in this section can be explained in the following way [36]: Practitioners who perform experiments often say that step type test signals are good, but typically do not excite high frequencies terms well enough. On the other hand random signals such as PRBS are also considered good, but typically have too much energy in the high-frequency region. Step type inputs have power spectral density that decays as $1/(freq.)^2$ ($1/f^2$) whereas random signals have constant power spectral density. This implies that a signal having power spectral density that lies somewhere between $1/f^2$ and a constant might be a good open-loop test signal. This suggests $1/f$ noise (over a limited bandwidth) as a possible good choice.

Examples which show the good performance of bandlimited "$1/f$" noise as a first experiment when compared with other typical input signals, such as bandlimited white noise or an optimally designed input (based on perfect knowledge of the plant and noise properties), have been presented by the coauthors in several publications, e.g., [16], [21], [33], [36], [39], and [40].

Remark 1: It is important to notice that the results of this section obviously do not imply that bandlimited "$1/f$" noise is the optimal input signal under every possible circumstance. The optimality of this signal has been established for the case when there is a lack of prior knowledge. If, for instance, the frequency response of the system were known to contain peaks in some frequency regions, then it is reasonable to incorporate this prior knowledge into the optimization problem of Lemma 2. This has already been discussed in [21], where it is noted that the results of this section resemble the development of the Principle of Maximum Entropy as given in [41] and [42], where the same issue regarding the incorporation of prior knowledge arises.

Remark 2: Since we are considering the case where there is very little information about the plant, we cannot expect the optimal input, i.e., bandlimited "$1/f$" noise, to have spectacular performance compared to a carefully designed input based on full knowledge of the plant. The input signal we have proposed is designed to be used as the first experiment on the plant, in order to determine its main features. As performance requirements on the closed loop are increased, more experiments can be performed on the plant, from which we can obtain a better model, based on which one can design a better experiment and so on.

V. MIN-MAX ROBUST EXPERIMENT DESIGN

Here we utilize the results of Section III to analyse the problem of designing an input signal which is robust, in an integrated variance sense, against all possible model structures (and also the true values of the system parameters). This analysis will then be used in the next section to design input signals which are optimally robust, in terms of both bias and variance errors, for a particular application.

Theorem 3: Consider the experiment design problem:[5]

$$\min_{\Phi_u > 0} \sup_{\Gamma \in \mathcal{H}_2^n} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Var}\left[\hat{G}(e^{j\omega})\right] W(e^{j\omega}) d\omega,$$

$$\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1$$

where $W \in C(\mathbb{T}, \mathbb{R}^+)$ and

$$\mathrm{Var}\left[\hat{G}(z)\right] := \Gamma^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau}) \Gamma^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau\right]^{-1} \Gamma(z)$$

[5]Note that the input power has been normalized to be less than or equal to 1. When the input power is constrained to be below some other value, it suffices, for the problems considered in this paper, to scale the optimal solution to satisfy that constraint. For other kinds of experiment design problems, the reader is referred to [22] which provides methods to renormalize the optimal input.

for $z \in \mathbb{T}$. The solution of this problem is given by

$$\Phi_u^{\mathrm{opt}}(\omega) := \frac{\Phi_v(\omega) W(e^{j\omega})}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_v(\tau) W(e^{j\tau}) d\tau}, \quad \omega \in [-\pi, \pi] \quad (7)$$

and the optimal cost is

$$\min_{\substack{\Phi_u \geq 0 \\ \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1}} \sup_{\Gamma \in \mathcal{H}_2^n} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Var}\left[\hat{G}(e^{j\omega})\right] W(e^{j\omega}) d\omega$$

$$= \frac{n}{2\pi N} \int_{-\pi}^{\pi} \Phi_v(\omega) W(e^{j\omega}) d\omega.$$

Proof: See Appendix D. ∎

Theorem 3 gives the solution to a robust experiment design problem. The *nonrobust* version of that optimization problem (i.e., without the maximization with respect to Γ) is a very standard problem in experiment design. It was probably first studied by Ljung in [31], where several choices for W were considered, depending on the specific application of the model to be estimated. For example, if the purpose of the model is simulation, then W could be chosen as

$$W(e^{j\omega}) = \Phi_u^{\mathrm{sim}}(\omega)$$

where Φ_u^{sim} is the spectrum of the input to be used during the simulation; if the model is to be used for (one step ahead) prediction, then the choice should be

$$W(e^{j\omega}) = \frac{\Phi_u^{\mathrm{pred}}(\omega)}{\Phi_v(\omega)}$$

where Φ_u^{pred} is the spectrum of the input to be used during the prediction stage, and Φ_v can be taken as an initial estimate of the noise spectrum. The interested reader is referred to [12], [31], and [5, Sec. 12.2], where these choices are studied in detail.

The original nonrobust version of the problem, studied in [31], has a nice explicit closed-form solution for the case when both the number of samples and the model order are very large. Solutions for models having a finite number of parameters can be obtained, in general, only numerically, by using convex optimization techniques [43].

Theorem 3 shows that it is possible to obtain analytic expressions for the robust version of the experiment design problem, which are valid even for models with a finite number of parameters (but which are still asymptotic in sample size). The optimal solution in this case, (7), also has a very nice interpretation: it is such that the frequency-wise signal-to-noise ratio, Φ_u/Φ_v, is proportional to the weighting function W. Hence it puts more power at those frequencies where the noise power is high and where a better model is required. Also notice that (7) does not explicitly depend on the number of parameters n of the model structures considered. Hence it is optimal for every fixed n. Finally, note that, due to the robustification of the experiment design problem (by considering the maximum over all model structures having n parameters), the optimal spectrum does not depend (explicitly) on the true system.[6]

[6]The optimal spectrum (7) might still depend on the true system through W.

VI. Unprejudiced Input Design for Finite Model Order

Finally we consider the unprejudiced optimal input design problem. It is known, [23], that prejudice is introduced into the parameter estimation problem due to the assumption that the system of interest belongs to a limited set of models. This has been addressed in [23], where Yuan and Ljung develop a framework for reducing the effect of this prejudice for experiment design. This is accomplished in two ways: a) by including a bias effect explicitly in the cost and b) by using an asymptotic variance expression in the cost.

Utilizing fundamental limitations on the variance [24], we revisit the approach in [23] and develop an unprejudiced optimal input for finite-order models.

The result of the min max robust experiment design problem in the previous section is used to obtain an improved unprejudiced open-loop input design, in the sense of Yuan and Ljung [23]. First we recall the concept of an unprejudiced input design.

The experiment design problem considered in [23] is of the form

$$
\min_{\Phi_u > 0} \int_{-\pi}^{\pi} E\left\{ \left| \hat{G}(e^{j\omega}) - G_0(e^{j\omega}) \right|^2 \right\} W(e^{j\omega}) d\omega
$$

$$
\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1
$$

where $W \in C(\mathbb{T}, \mathbb{R}^+)$, and undermodelling, i.e., bias in G can exist. To solve this problem, the mean square error in the estimation of G_0 can be decomposed into bias and variance terms

$$
E\left\{ \left| \hat{G}(e^{j\omega}) - G_0(e^{j\omega}) \right|^2 \right\}
$$
$$
= \left| G_0(e^{j\omega}) - G^*(e^{j\omega}) \right|^2 + \text{Var}\left[\hat{G}(e^{j\omega}) \right]
$$

where $G^*(e^{j\omega}) := \lim_{N \to \infty} \hat{G}(e^{j\omega})$ almost surely. This decomposition holds asymptotically in N, in the sense that for finite N, the bias term should consider $E\{\hat{G}(e^{j\omega})\}$ instead of the limit estimate $G^*(e^{j\omega})$. This approximation, however, allows further simplifications in the calculation of the optimal experiment. Minimization of the bias term leads to the following solution [5], [23]:

$$
\Phi_u^{\text{opt}}(\omega) = c_1 W(e^{j\omega}) \left| H^*(e^{j\omega}) \right|^2 \tag{8}
$$

where $H^*(e^{j\omega}) := \lim_{N \to \infty} \hat{H}(e^{j\omega})$ almost surely, and $c_1 > 0$ is a normalization constant. Notice that this solution is independent of both G_0 and G^*.

With respect to the variance term, an asymptotic (in model order) variance expression is used in [23], which is minimized by the following input spectrum:

$$
\Phi_u^{\text{opt}}(\omega) = c_2 \sqrt{W(e^{j\omega}) \left| H_0(e^{j\omega}) \right|^2}
$$

where $c_2 > 0$ is a normalization constant. Note that the asymptotic (in model order) variance expression [31] used to develop this equation for the input spectrum does not consider the effect of bias.

In order to reconcile both expressions for Φ_u^{opt}, H^* is considered as a prefilter (designed by the user), such that

$$
\left| H^*(e^{j\omega}) \right|^2 = c_3 \sqrt{\frac{\left| H_0(e^{j\omega}) \right|^2}{W(e^{j\omega})}}
$$

where $c_3 > 0$. This solution is dimensionally inconsistent, since it constrains the noise prefilter to be proportional to the square root of the true noise spectrum, creating a paradox.

This paradox arises due to the use of an asymptotic (in model order) variance expression, which only holds approximately for model sets with a shift structure [5, Sec. 9.4].

To solve this dilemma, we consider the following experiment design problem:

$$
\min_{\Phi_u > 0} \sup_{G \in \mathcal{M}_n} \int_{-\pi}^{\pi} E\left\{ \left| \hat{G}(e^{j\omega}) - G_0(e^{j\omega}) \right|^2 \right\} W(e^{j\omega}) d\omega
$$

$$
\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq 1
$$

where $\mathcal{M}_n$ is the set of all stable model structures with n parameters, i.e., $\mathcal{M}_n := \{ G : \mathbb{C} \times \Theta \to \mathbb{C} : G(z, \cdot)$ is differentiable in the connected open set $\Theta \subseteq \mathbb{R}^n$ for all $z \in \mathbb{T}$, and $G(\cdot, \theta) \in \mathcal{H}_2$ for all $\theta \in \Theta \}$.

In this problem formulation we consider the worst case of the (weighted) mean square error over all model structures of a given order. Again, the cost function can be decomposed into both bias and variance terms. The bias term is minimized by (8), since the solution is independent of G_0 and G^*. This implies that taking the supremum over all model structures in $\mathcal{M}_n$ does not affect the previous solution. The argument is formalized in the following theorem.

Theorem 4 (Optimality of Dominant Strategies): Let $J : X \times Y \to \mathbb{R}$ be an arbitrary function, where X and Y are arbitrary sets. Assume that there exists an $x^* \in X$ such that

$$
J(x^*, y) = \min_{x \in X} J(x, y) =: C_y \in \mathbb{R}, \quad y \in Y.
$$

Then

$$
\sup_{y \in Y} J(x^*, y) = \min_{x \in X} \sup_{y \in Y} J(x, y).
$$

Therefore x^* is an optimal solution of the min-max problem[7] $\min_{x \in X} \sup_{y \in Y} J(x, y)$.

Proof: By definition of the infimum of a function, we have that

$$
\inf_{x \in X} \sup_{y \in Y} J(x, y) \leq \sup_{y \in Y} J(x^*, y) = \sup_{y \in Y} C_y. \tag{9}
$$

On the other hand, by the definition of the supremum,

$$
\sup_{y \in Y} J(x, y) \geq J(x, y^\circ), \quad x \in X, y^\circ \in Y.
$$

[7]In game-theoretical terms [44], Theorem 4 establishes that a dominating strategy is an equilibrium strategy.
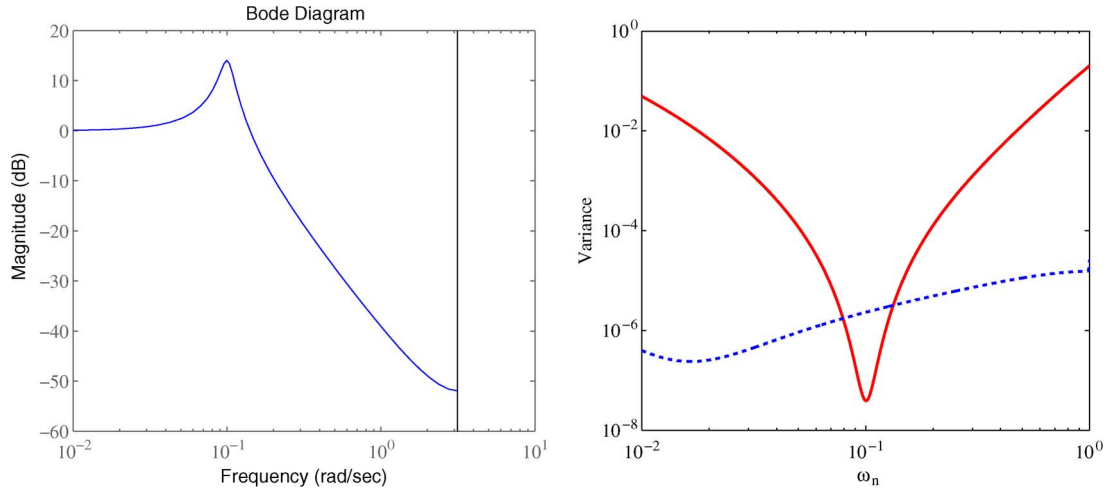
Fig. 1.   (Left) Magnitude Bode plot of G_0, from Section VII. (Right) Normalized variance of $\hat{\theta}$, as a function of the parameter ω_n, obtained from an experiment based on a sinusoid of unit power and frequency $\omega_0 = 1$ (red solid), and from an experiment based on bandlimited "$1/f$" noise of unit power localized between the frequencies 0:01 and 1 (blue dotted).

Thus, by taking the infimum over $x \in X$, we obtain

$$\inf_{x \in X} \sup_{y \in Y} J(x, y) \geq \inf_{x \in X} J(x, y^\circ)$$
$$= \min_{x \in X} J(x, y^\circ)$$
$$= C_{y^\circ}, \quad y^\circ \in Y. \tag{10}$$

Since (10) holds for every $y^\circ \in Y$, we can take the supremum over this quantity, which gives [45, Lemma 36.1]

$$\inf_{x \in X} \sup_{y \in Y} J(x, y) \geq \sup_{y^\circ \in Y} C_{y^\circ}. \tag{11}$$

Combining (9) and (11) and replacing inf by min (since the infimum is attained with $x = x^*$) concludes the proof.   ∎

For the variance term, we consider the true asymptotic (in sample size) variance expression

$$\mathrm{Var}\left[\hat{G}(e^{j\omega})\right] = \Gamma^H(e^{j\omega})$$
$$\times \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau}) \Gamma^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau\right]^{-1} \Gamma(e^{j\omega}) \tag{12}$$

which is asymptotic only in N. Notice, however, that we are still not considering the effect of bias on the variance of $\hat{G}$.

The variance term, based on expression (12), corresponds exactly to the min-max robust optimal experiment design problem considered in the previous section; hence the solution (from Theorem 3) is

$$\Phi_u^{\mathrm{opt}}(\omega) = c_4 W(e^{j\omega}) \left|H_0(e^{j\omega})\right|^2 \tag{13}$$

where $c_4 > 0$ is a normalization constant.

Remark 3: Notice that (13) and (8) can be naturally combined by letting $H^* = H_0$!

Just as in the robust experiment design problem considered in Theorem 3, the optimal input obtained here has a nice interpretation, namely it is chosen such that the signal-to-noise ratio is proportional, at each frequency, to the weighting function W.

VII. NUMERICAL EXAMPLE

Consider the following discrete-time linear time-invariant system of second order:

$$G_0(q^{-1}) = \frac{(1 + \xi^2 - 2\xi \cos \omega_n)q}{(q - \xi e^{j\omega_n})(q - \xi e^{-j\omega_n})}$$
$$= \frac{(1 + \xi^2 - 2\xi \cos \omega_n)q}{q^2 - 2\xi \cos(\omega_n)q + \xi^2}, \xi \in [0, 1), \omega_n \in (0, \pi).$$

Notice that $G_0(1) = 1$, and for $\xi \approx 1$, $G_0(q^{-1})$ is highly resonant, with a peak at $\omega_r \approx \omega_n$. The magnitude Bode plot of G_0 is shown in Fig. 1.

In order to verify the results of Section IV, let us consider a model of the form

$$G_1(q^{-1}, \theta) = \frac{(1 + \xi^2 - 2\xi\theta)q}{q^2 - 2\xi\theta q + \xi^2}$$

where ξ is known, and we need to estimate θ (whose true value is $\theta_0 = \cos \omega_n$). The output measurements of system are contaminated with white noise of variance σ^2. The information matrix for θ is

$$M = \frac{2N}{\pi\sigma^2} \int_{-\pi}^{\pi} \frac{\xi^2 \left|e^{2j\omega} - (1 + \xi^2)e^{j\omega} + \xi^2\right|^2}{\left|e^{2j\omega} - 2\xi\cos(\omega_n)e^{j\omega} + \xi^2\right|^4} \Phi_u(\omega)d\omega.$$

The maximum of M is achieved by choosing $\Phi_u(\omega) = \alpha[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$, where

$$\omega_0 = \arg \max_{\omega \in (0, \pi)} \frac{\xi^2 \left|e^{2j\omega} - (1 + \xi^2)e^{j\omega} + \xi^2\right|^2}{\left|e^{2j\omega} - 2\xi\cos(\omega_n)e^{j\omega} + \xi^2\right|^4}$$

and $\alpha > 0$ is chosen to satisfy an input power constraint. If $\xi \approx 1$, then $\omega_0 \approx \omega_n$, which means that the optimal input should be a sinusoid of frequency approximately equal to the resonance frequency of G_0. Furthermore, as $\xi \to 1_-$, the shape of

$$\frac{\xi^2 \left|e^{2j\omega} - (1 + \xi^2)e^{j\omega} + \xi^2\right|^2}{\left|e^{2j\omega} - 2\xi\cos(\omega_n)e^{j\omega} + \xi^2\right|^4}$$
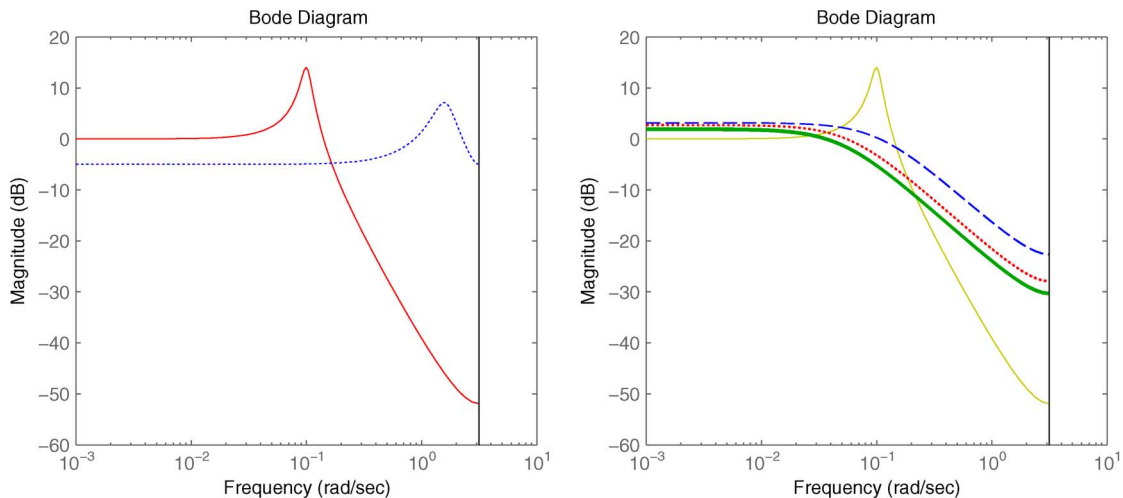
Fig. 2. (Left) Magnitude Bode plot of G_0 (red solid), and the spectrum Φ_u^{sim} (blue dotted). (Right) Magnitude Bode plots of 50 models estimated from experiments based on a white noise input (red dotted), a "$1/f$" noise input (blue dashed), and the unprejudiced optimal input (green solid); for comparison, the Bode plot of G_0 has also been included (yellow solid).

becomes sharper; hence missing the value of ω_0 (which depends on the true value of θ_0) may cause a huge performance degradation for $\xi \approx 1$. For example, let $\xi = 0.95$, $\omega_n \in [0.01, 1]$ and $\sigma = 0.1$. The variance of $\hat{\theta}$ obtained from an experiment based on a sinusoid of frequency $\omega_0 = 0.1$ (instead of ω_n) of unit power, as a function of the true value of $\theta_0 = \omega_n$, is shown in Fig. 1. In the same figure the normalized (i.e., multiplied by N) variance obtained using "$1/f$" noise (in the frequency range $[0.01, 1]$) of unit power, is presented. As it can be seen, the signal proposed in Section IV has superior robustness properties compared to the nominal optimal input, since its resulting variance is less sensitive to the knowledge of θ_0 than with the latter signal. In fact, the maximum variance obtained with the sinusoid input is $2.0229 \cdot 10^{-1}$, while the maximum variance obtained with "$1/f$" noise is just $2.6174 \cdot 10^{-5}$.

The advantages of "$1/f$" noise become apparent in situations where there is lack of enough prior knowledge to design an "optimal" experiment and it is required to obtain a good "general-purpose" model for control. However, when there is a more precise specification of the final application for the model, the results of Section VI become relevant.

Say that a model of G_0 is required for simulating the output of the system when the input $u_{\mathrm{sim}}(t)$ has a spectrum given by

$$\Phi_u^{\mathrm{sim}}(\omega) = \left| \frac{(e^{j\omega} - 0.4)(e^{j\omega} + 0.4)}{e^{2j\omega} + 0.49} \right|^2.$$

To study the properties of the inputs used during the identification in the presence of undermodelling, the following first-order model structure is considered:

$$G_2(q^{-1}, \theta) = \frac{b_1 q^{-1}}{1 + a_1 q^{-1}}$$

where $\theta = [a_1/b_1]^T$. According to the result of Section VI, the unprejudiced optimal input to use has a spectrum proportional to Φ_u^{sim}. The results of 50 Monte Carlo simulations (with $N = 10000$) are presented in Fig. 2, where the input signals considered are white noise, "$1/f$" noise (with the same characteristics as before) and the unprejudiced optimal input, normalized to unit power. As it can be seen from the figure,

TABLE I
MEAN PERFORMANCE OF THE SIMULATION EXPERIMENT FOR DIFFERENT INPUT SIGNALS

Input Signal	Mean Performance
White noise	$7.38 \cdot 10^{-2}$
Bandlimited '$1/f$' noise	$9.34 \cdot 10^{-2}$
Unprejudiced optimal input	$7.26 \cdot 10^{-2}$

none of the identified models can appropriately capture the shape of G_0. However, the models estimated using the unprejudiced optimal input give a slightly better fit in the relevant frequency range of Φ_u^{sim}. This reasoning is corroborated by Table I, which shows the mean performance of the experiments, $E\{[G_0(q^{-1})u_{\mathrm{sim}}(t) + w(t) - \hat{G}(q^{-1})u_{\mathrm{sim}}(t)]^2\}$, obtained by Monte Carlo averages. This table reveals the benefits of using the unprejudiced optimal input obtained in Section VI.[8]

VIII. CONCLUSIONS

In this paper we have introduced a variational approach to robust experiment design. Based on a fundamental limitation on the variance of parametric estimated models, a closed form expression is developed for several experiment design problems where the variance of the frequency response model is maximized over all model structures of a given finite order. In particular, we have revisited the problem of experiment design with diffuse prior information, i.e., where an input spectrum is designed which is independent of the true system and the noise dynamics. We have also studied the problem of unprejudiced input design, following Yuan and Ljung's formulation. Both problems have been investigated in the literature, however the approach of the current paper leads to results which are valid, not only for high order models, but also for models of finite order.

[8]From Table I it might seem that "$1/f$" noise is not a good input signal in this case. However, the derivation of such an input was based on the assumption that we need a good "general-purpose" model in a given frequency range. In the simulation example, we ask for a model which is better at high frequencies than at low ones (because of Φ_u^{sim}). "$1/f$" noise has less power at high frequencies than white noise or the unprejudiced optimal input; hence it is expected to give worse performance in this example.

Appendix A
Proof of Theorem 2

To prove Theorem 2, we first establish the following lemma.

Lemma 4 (Uniform Approximation of the Variance): Let $\Phi_u, \Phi_v : [-\pi, \pi] \to \mathbb{R}^+$ be continuous and even. Also, let $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ be such that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}) d\omega = \frac{n}{N}$$

where $n, N \in \mathbb{N}$. Then, for every $\epsilon > 0$ there exists a vector-valued polynomial in z^{-1}, $\Gamma \in \mathcal{H}_2^n$, such that

$$\left| \Gamma^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau}) \Gamma^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau \right]^{-1} \Gamma(z) - V(z) \right|$$
$$< \epsilon, \quad z \in \mathbb{T}.$$

Proof: The idea is to approximate Γ by a piecewise constant vector Γ_1, then by a piecewise linear (continuous) vector Γ_2, and finally by a trigonometric polynomial vector Γ_3 (using Theorem 5 of Appendix E).

In order to simplify the problem, we define the following:

$$\widetilde{\epsilon} := \frac{N}{2\pi} \left[\min_{\omega \in [-\pi, \pi]} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} \right] \epsilon,$$

$$\widetilde{\Gamma}(e^{j\omega}) := \Gamma(e^{j\omega}) \sqrt{\frac{\Phi_u(\omega)}{\Phi_v(\omega)}},$$

$$\widetilde{V}(e^{j\omega}) := \frac{N}{2\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}), \quad \omega \in [-\pi, \pi].$$

It can be readily seen that Lemma 4 would follow from establishing the existence of a function $\widetilde{\Gamma} \in \mathcal{H}_2^n$ such that

$$\left| \widetilde{\Gamma}^H(z) \left[\int_{-\pi}^{\pi} \widetilde{\Gamma}(e^{j\tau}) \widetilde{\Gamma}^H(e^{j\tau}) d\tau \right]^{-1} \widetilde{\Gamma}(z) - \widetilde{V}(z) \right| < \widetilde{\epsilon}, \quad z \in \mathbb{T}. \tag{14}$$

Let $m \geq 2n$ such that $|\widetilde{V}(e^{j\omega_1}) - \widetilde{V}(e^{j\omega_2})| < \widetilde{\epsilon}/6$ whenever $|\omega_1 - \omega_2| \leq 2\pi/m$ (for $\omega_1, \omega_2 \in [-\pi, \pi]$). According to Lemma 5 of Appendix E, there are n orthonormal vectors $v^i \in \mathbb{R}^m$ such that[9]

$$\sum_{i=1}^{n} \left(v_k^i \right)^2 = \frac{2\pi}{m} \widetilde{V} \left(\frac{2\pi}{m}[k - 1/2] - \pi \right) + \frac{\eta}{m}, \quad k = 1, \dots, m \tag{15}$$

where[10]

$$\eta := n - \frac{2\pi}{m} \sum_{k=1}^{m} \widetilde{V} \left(\frac{2\pi}{m}[k - 1/2] - \pi \right).$$

---

[9]The condition that $c_k \in [0, 1]$ in Lemma 5 is satisfied if we choose m large enough, since the right side of (15) converges uniformly (in k) to 0 as $m \to \infty$.

[10]The term η/m in (15) is due to the requirement that $\sum_{k=1}^{m} \sum_{i=1}^{n} |v_k^i|^2 = \sum_{i=1}^{n} \|v^i\|_2^2 = n$.

Thus, if we define the function $\Gamma_1 : \mathbb{T} \to \mathbb{R}^n$ by $[\Gamma_1(e^{j\omega})]_i = \sqrt{m/2\pi} v_k^i$ for $\omega \in [2\pi(k-1)/m - \pi, 2\pi k/m - \pi)$ and $i = 1, \dots, n$, then it holds that

$$\left[\int_{-\pi}^{\pi} \Gamma_1(e^{j\omega}) \Gamma_1^T(e^{j\omega}) d\omega \right]_{il} = \sum_{k=1}^{m} v_k^i v_k^l \frac{m}{2\pi} \frac{2\pi}{m}$$
$$= (v^l)^T v^i$$
$$= \delta_{i,l}, \quad i, l = 1, \dots, n$$

where $\delta_{i,l}$ is the Kronecker Delta function, and hence

$$\int_{-\pi}^{\pi} \Gamma_1(e^{j\omega}) \Gamma_1^T(e^{j\omega}) d\omega = I.$$

Now, for every $\omega \in [2\pi(k-1)/m - \pi, 2\pi k/m - \pi)$,

$$\Gamma_1^T(e^{j\omega}) \Gamma_1(e^{j\omega}) = \sum_{i=1}^{m} \left(v_k^i \right)^2 \frac{m}{2\pi}$$
$$= \widetilde{V} \left(\frac{2\pi}{m}[k - 1/2] - \pi \right) + \frac{\eta}{2\pi}.$$

Thus

$$\left| \Gamma_1^T(e^{j\omega}) \left[\int_{-\pi}^{\pi} \Gamma_1(e^{j\tau}) \Gamma_1^T(e^{j\tau}) d\tau \right]^{-1} \Gamma_1(e^{j\omega}) - \widetilde{V}(e^{j\omega}) \right|$$
$$= \left| \Gamma_1^T(e^{j\omega}) \Gamma_1(e^{j\omega}) - \widetilde{V}(e^{j\omega}) \right|$$
$$= \left| \widetilde{V} \left(\frac{2\pi}{m}[k - 1/2] - \pi \right) - \widetilde{V}(e^{j\omega}) + \frac{\eta}{2\pi} \right|$$
$$< \frac{\epsilon}{6} + \frac{|\eta|}{2\pi}$$
$$< \frac{\epsilon}{6} + \frac{\epsilon}{6}$$
$$= \frac{\epsilon}{3}, \quad \omega \in [-\pi, \pi] \tag{16}$$

since

$$\frac{|\eta|}{2\pi} = \frac{1}{2\pi} \left| \int_{-\pi}^{\pi} \widetilde{V}(e^{j\omega}) d\omega - \frac{2\pi}{m} \sum_{k=1}^{m} \widetilde{V} \left(\frac{2\pi}{m}[k - 1/2] - \pi \right) \right|$$
$$< \frac{1}{2\pi} m \frac{2\pi}{m} \frac{\epsilon}{6}$$
$$= \frac{\epsilon}{6}.$$

Let $\Gamma_2 : \mathbb{T} \to \mathbb{R}^n$ be a continuous function such that $\Gamma_2(\overline{z}) = \Gamma_2(z)$ for all $z \in \mathbb{T}$, and

$$\left| \Gamma_1^T(z) \Gamma_1(z) - \Gamma_2^T(z) \left[\int_{-\pi}^{\pi} \Gamma_2(e^{j\tau}) \Gamma_2^T(e^{j\tau}) d\tau \right]^{-1} \Gamma_2(z) \right|$$
$$< \frac{\epsilon}{2}, \quad z \in \mathbb{T}. \tag{17}$$

Here we replace Γ_1, for a given $\alpha > 0$, by a piecewise linear function Γ_2 such that

$$\left\| \left[\int_{-\pi}^{\pi} \Gamma_2(e^{j\tau})\Gamma_2^T(e^{j\tau})d\tau \right]^{-1} - I \right\|_{\infty} < \alpha$$

and $|\Gamma_2^T(z)\Gamma_2(z) - \Gamma_1^T(z)\Gamma_1(z)| < \epsilon/6$ for every $z \in \mathbb{T}$, the later being possible since $|\tilde{V}(e^{j\omega_1}) - \tilde{V}(e^{j\omega_2})| < \tilde{\epsilon}/6$ whenever $|\omega_1 - \omega_2| \leq 2\pi/m$. Thus, we can choose α small enough to ensure that (17) holds.

Finally, since $\Gamma \mapsto \Gamma^T(\int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^T(e^{j\tau})d\tau)^{-1}\Gamma$ is continuous with respect to the uniform norm on $C(\mathbb{T}, \mathbb{R}^n)$ in a neighborhood of Γ_2, by Theorem 5 of Appendix E, there exists a (vector-valued) trigonometric polynomial

$$\Gamma_3(e^{j\omega}) = \sum_{i=-p}^{p} a_{|i|}e^{j\omega i}, \quad \omega \in [-\pi, \pi]$$

with $a_i \in \mathbb{R}^n$ for $i = 0, \ldots, p$, such that

$$\left| \Gamma_3^T(z) \left[\int_{-\pi}^{\pi} \Gamma_3(e^{j\tau})\Gamma_3^T(e^{j\tau})d\tau \right]^{-1} \Gamma_3(z) - \right.$$
$$\left. \Gamma_2^T(z) \left[\int_{-\pi}^{\pi} \Gamma_2(e^{j\tau})\Gamma_2^T(e^{j\tau})d\tau \right]^{-1} \Gamma_2(z) \right| < \frac{\epsilon}{6} \quad (18)$$

for every $z \in \mathbb{T}$. Therefore, the function $\tilde{\Gamma} \in \mathcal{H}_2^n$ given by $\tilde{\Gamma}(z) = \Gamma_3(z)z^{-p}$ satisfies (14), as can be seen by combining (16)–(18). ∎

Proof of Theorem 2: We first outline the steps in the proof, and then elaborate on each step:

1) Construct a sequence of functions in $\mathcal{H}_2^n$, $\{\Gamma_k\}_{k=1}^{\infty}$, using Lemma 4.
2) Show that this sequence is a *normal family*, from which it follows that it has a subsequence converging to, say, Γ_{∞}.
3) Establish that the function

$$\Gamma \mapsto \Gamma^H \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1} \Gamma$$

is continuous in a neighborhood of Γ_{∞}, hence proving that Γ_{∞} satisfies the conditions of the theorem.

The details of each step are as follows:

Step 1) *Construction of a sequence $\{\Gamma_k\}_{k=1}^{\infty}$*
To proceed, we use Lemma 4 to construct a sequence of functions in $\mathcal{H}_2^n$, $\{\Gamma_k\}_{k=1}^{\infty}$, such that

$$\left| \Gamma_k^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma_k(e^{j\tau})\Gamma_k^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1} \Gamma_k(z) - V(z) \right|$$
$$< \frac{1}{k}, \quad z \in \mathbb{T}.$$

Since the Γ_k's are polynomials in z^{-1}, they are analytic in the set $\mathbb{E}_{1/2} := \{z \in \mathbb{C} : |z| > 1/2\}$, and,

in particular, are bounded in this set. This, together with the fact that

$$\Gamma \mapsto \Gamma^H \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1} \Gamma$$

is invariant under scaling of its argument, implies that we can assume

$$\lim_{r \to \frac{1}{2}+} \max_{z \in \mathbb{T}} \|\Gamma_k(rz)\|_2 = 1, \quad k \in \mathbb{N}. \quad (19)$$

Furthermore, by applying a suitable constant unitary linear transformation to each Γ_k, we can further assume that

$$\int_{-\pi}^{\pi} \Gamma_k(e^{j\tau})\Gamma_k^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau = \alpha_k I, \quad k \in \mathbb{N} \quad (20)$$

where $\alpha_k > 0$ for every $k \in \mathbb{N}$.

Step 2) *A converging subsequence in $\{\Gamma_k\}_{k=1}^{\infty}$*
From Theorem 6 of Appendix E, it follows that $\{\Gamma_k\}_{k=1}^{\infty}$ is uniformly bounded (by 1) on $\mathbb{E}_{1/2}$. Therefore, by Theorem 7 of Appendix E we have that $\{\Gamma_k\}_{k=1}^{\infty}$ is a *normal family* in $\mathbb{E}_{1/2}$, i.e., there exists a subsequence $\{\Gamma_{k_i}\}_{i=1}^{\infty}$ which converges uniformly on compact subsets of $\mathbb{E}_{1/2}$. Let Γ_{∞} be the limit of this subsequence. Note that Γ_{∞} is analytic in $\mathbb{E}_{1/2}$ by Theorem 8 of Appendix E, and belongs to $\mathcal{H}_2^n$ due to $\sup_{z \in \mathbb{E}_{1/2}} \|\Gamma_{\infty}(z)\|_2 \leq 1$. Since $\mathbb{T} \subset \mathbb{E}_{1/2}$ is compact, $\Gamma_{k_i} \to \Gamma_{\infty}$ uniformly in $\mathbb{T}$.

Step 3) *Continuity of the variance expression*
The function

$$\Gamma \mapsto \Gamma^H \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma(e^{j\tau})\Gamma^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1} \Gamma$$

is continuous in a neighborhood of Γ_{∞} if

$$\int_{-\pi}^{\pi} \Gamma_{\infty}(e^{j\tau})\Gamma_{\infty}^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau = \lim_{i \to \infty} \alpha_{k_i} I > 0.$$

Therefore, we have

$$\Gamma_{\infty}^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma_{\infty}(e^{j\tau})\Gamma_{\infty}^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1} \Gamma_{\infty}(z)$$
$$= \lim_{i \to \infty} \Gamma_{k_i}^H(z) \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma_{k_i}(e^{j\tau})\Gamma_{k_i}^H(e^{j\tau})\frac{\Phi_u(\tau)}{\Phi_v(\tau)}d\tau \right]^{-1}$$
$$\times \Gamma_{k_i}(z)$$
$$= V(z), \quad z \in \mathbb{T}.$$

Thus, in order to show that Γ_{∞} satisfies the condition of the Theorem, we need to show that

$\lim_{i\to\infty} \alpha_{k_i} > 0$ (where α_k has been defined in (20)). This can be seen from the expression

$$V_{k_i}(z) := \Gamma_{k_i}^H(z)$$

$$\times \left[\frac{N}{2\pi} \int_{-\pi}^{\pi} \Gamma_{k_i}(e^{j\tau}) \Gamma_{k_i}^H(e^{j\tau}) \frac{\Phi_u(\tau)}{\Phi_v(\tau)} d\tau \right]^{-1} \Gamma_{k_i}(z)$$

$$= \alpha_{k_i}^{-1} \Gamma_{k_i}^H(z) \Gamma_{k_i}(z)$$

$$= \alpha_{k_i}^{-1} \| \Gamma_{k_i}(z) \|_2^2$$

for $i \in \mathbb{N}$ and $z \in \mathbb{T}$, where $V_{k_i} \to V$ uniformly in $\mathbb{T}$ as $i \to \infty$. Therefore, by maximizing over $\mathbb{T}$ and letting $i \to \infty$, we obtain

$$\max_{z\in\mathbb{T}} V(z) = \lim_{i\to\infty} \max_{z\in\mathbb{T}} V_{k_i}(z)$$

$$= \lim_{i\to\infty} \alpha_{k_i}^{-1} \max_{z\in\mathbb{T}} \| \Gamma_{k_i}(z) \|_2^2$$

$$= \lim_{i\to\infty} \alpha_{k_i}^{-1} \max_{z\in\mathbb{T}} \| \Gamma_\infty(z) \|_2^2 .$$

This implies

$$\lim_{i\to\infty} \alpha_{k_i} = \frac{\max_{z\in\mathbb{T}} \| \Gamma_\infty(z) \|_2^2}{\max_{z\in\mathbb{T}} V(z)} > 0.$$

Otherwise (19) would not hold. This completes the proof. ∎

APPENDIX B
PROOF OF LEMMA 2

By Theorems 1 and 2, the experiment design problem is equivalent to

$$\min_{\Phi_u > 0} \sup_{V \in C(\mathbb{T}, \mathbb{R}_0^+)} \int_{-\pi}^{\pi} F\left(\frac{V(e^{j\omega})}{|G(e^{j\omega})|^2} \right) W(\omega) d\omega$$

$$\text{s.t. } \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega = 1$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}) d\omega = \frac{n}{N}. \qquad (21)$$

The idea here is that every $\Gamma \in \mathcal{H}_2^n$ gives rise to a variance $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ which satisfies the integral constraint established in Theorem 1, and conversely, every $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ which satisfies the integral constraint can be related, by Theorem 2, to at least[11] one $\Gamma \in \mathcal{H}_2^n$. Therefore, the maximization with respect to $\Gamma \in \mathcal{H}_2^n$ can be replaced by a maximization with respect to $V \in C(\mathbb{T}, \mathbb{R}_0^+)$ (imposing the integral constraint of Theorem 1).

Let G and W be fixed, and assume that Φ_u^{opt} exists. This problem can now be solved using standard tools from calculus of variations [46].

---

[11]The possibility of having more than one Γ associated with the same variance V is not an issue here, since the cost function of the experiment design problem depends on Γ only through $\text{Var}[\hat{G}(z)]$.

The Lagrangian of problem (21) is

$$\mathcal{L}(V, \Phi_u, \lambda_1, \lambda_2) := \frac{1}{2\pi} \int_a^b \left[2\pi F\left(\frac{V(e^{j\omega})}{|G(e^{j\omega})|^2} \right) W(\omega) \right.$$

$$\left. + \lambda_1 \Phi_u(\omega) + \lambda_2 \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}) \right] d\omega$$

where λ_1 and λ_2 are Lagrange multipliers. By [46, Sec. 7.7, Theorem 2], there exist constants $\lambda_1, \lambda_2 \in \mathbb{R}$ for which $(V^{\text{opt}}, \Phi_u^{\text{opt}})$, the solution of (21), is a stationary point of $\mathcal{L}(V, \Phi_u, \lambda_1, \lambda_2)$.

Thus, for every $h_1, h_2 \in C^1([a, b], \mathbb{R}_0^+)$ we have that $\delta\mathcal{L}(V^{\text{opt}}, \Phi_u^{\text{opt}}, \lambda_1, \lambda_2; [h_1 h_2]^T) = 0$, where

$$\delta\mathcal{L}\left(V^{\text{opt}}, \Phi_u^{\text{opt}}, \lambda_1, \lambda_2; [h_1 h_2]^T \right)$$

$$:= \lim_{\alpha\to 0} \frac{1}{\alpha} \left[\mathcal{L}\left(V^{\text{opt}} + \alpha h_1, \Phi_u^{\text{opt}} + \alpha h_2, \lambda_1, \lambda_2 \right) \right.$$

$$\left. - \mathcal{L}\left(V^{\text{opt}}, \Phi_u^{\text{opt}}, \lambda_1, \lambda_2 \right) \right]$$

is the *Fréchet differential of $\mathcal{L}$ at $(V^{\text{opt}}, \Phi_u^{\text{opt}})$ with increment* $[h_1 \, h_2]^T$. This means [46, Sec. 7.5] that

$$\int_a^b \left\{ \left[\frac{2\pi W(\omega)}{|G(e^{j\omega})|^2} F'\left(\frac{V^{\text{opt}}(e^{j\omega})}{|G(e^{j\omega})|^2} \right) + \lambda_2 \frac{\Phi_u^{\text{opt}}(\omega)}{\Phi_v(\omega)} \right] h_1(\omega) + \right.$$

$$\left. \left[\lambda_1 + \lambda_2 \frac{V^{\text{opt}}(e^{j\omega})}{\Phi_v(\omega)} \right] h_2(\omega) \right\} d\omega = 0.$$

Thus, by [46, Sec. 7.5, Lemma 1]

$$\frac{2\pi W(\omega)}{|G(e^{j\omega})|^2} F'\left(\frac{V^{\text{opt}}(e^{j\omega})}{|G(e^{j\omega})|^2} \right) + \lambda_2 \frac{\Phi_u^{\text{opt}}(\omega)}{\Phi_v(\omega)} = 0,$$

$$\lambda_1 + \lambda_2 \frac{V^{\text{opt}}(e^{j\omega})}{\Phi_v(\omega)} = 0, \quad \omega \in [a, b]. \qquad (22)$$

From (22) we have that $V^{\text{opt}}(e^{j\omega}) = -(\lambda_1/\lambda_2)\Phi_v(\omega)$. Thus, substituting this into the first equation of (22), and letting $l(e^{j\omega}) := V^{\text{opt}}(e^{j\omega})/|G(e^{j\omega})|^2$, we have

$$l(e^{j\omega}) F'\left(l(e^{j\omega}) \right) = \frac{\lambda_1}{2\pi} \frac{\Phi_u^{\text{opt}}(\omega)}{W(\omega)}, \quad \omega \in [a, b]. \qquad (23)$$

The left side of (23) depends on G (through l), but the right side does not (due to the assumption of the independence of Φ_u^{opt} to G). Thus, both sides are equal to a constant, say $\alpha \in \mathbb{R}$, which implies that

$$F'\left(l(e^{j\omega}) \right) = \frac{\alpha}{l(e^{j\omega})}, \quad \omega \in [a, b]. \qquad (24)$$

Now, integrating both sides with respect to l between $\inf_{\omega\in[a,b]} l(e^{j\omega})$ and $\sup_{\omega\in[a,b]} l(e^{j\omega})$, we obtain

$$F\left(l(e^{j\omega}) \right) = \alpha \ln l(e^{j\omega}) + \beta, \quad \omega \in [a, b] \qquad (25)$$

for a constant $\beta \in \mathbb{R}$.

On the other hand, we have that

$$\frac{\lambda_1}{2\pi} \frac{\Phi_u^{\text{opt}}(\omega)}{W(\omega)} = \alpha, \quad \omega \in [a, b]. \qquad (26)$$

Hence Φ_u^{opt} is proportional to W in $[a, b]$, and can be made equal to 0 outside this interval. This concludes the proof. ∎

APPENDIX C
PROOF OF LEMMA 3

To simplify the development, we extend W to a periodic function of period 2π in $\mathbb{R}$ by making it equal to 0 in $[0, 2\pi] \setminus [a, b]$. Then, as in the proof of Lemma 2, the experiment design problem is equivalent to

$$\min_{\Phi_u > 0} \sup_{V \in C(\mathbb{T}, \mathbb{R}_0^+)} \int_{-\pi}^{\pi} \ln \left[\frac{V(e^{j\omega})}{|G(e^{j\omega})|^2} \right] W(\omega) d\omega,$$

$$\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega = 1$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}) d\omega = \frac{n}{N}.$$

Let Φ_u be fixed. The cost function can then be written as

$$\int_{-\pi}^{\pi} \ln \left[\frac{V(e^{j\omega})}{|G(e^{j\omega})|^2} \right] W(\omega) d\omega = C_1 + \int_{-\pi}^{\pi} \ln \left[\widetilde{V}(e^{j\omega}) \right] W(\omega) d\omega$$

where C_1 is a constant, independent of V, given by

$$C_1 = - \int_{-\pi}^{\pi} \ln \left[\frac{N |G(e^{j\omega})|^2 \Phi_u(\omega)}{2\pi n \Phi_v(\omega)} \right] W(\omega) d\omega$$

$$\widetilde{V}(e^{j\omega}) = \frac{N \Phi_u(\omega)}{2\pi n \Phi_v(\omega)} V(e^{j\omega}), \quad \omega \in [-\pi, \pi].$$

Note that, due to the constraint on V, $\widetilde{V}$ should satisfy

$$\int_{-\pi}^{\pi} \widetilde{V}(e^{j\omega}) d\omega = 1. \tag{27}$$

Let $\widetilde{V}^{\mathrm{opt}} = cW$, where c is chosen to satisfy (27), and let $\widetilde{V}$ be any function in $C(\mathbb{T}, \mathbb{R}_0^+)$ which satisfies (27). Since $\ln(1 + x) \leq x$ for all $x \in (-1, \infty)$, with equality if and only if $x = 0$, we have

$$\int_{-\pi}^{\pi} \ln \left[\widetilde{V}(e^{j\omega}) \right] W(\omega) d\omega$$

$$= \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) + \left(\widetilde{V}(e^{j\omega}) - \widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right) \right] W(\omega) d\omega$$

$$= \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] W(\omega) d\omega$$

$$+ \int_{-\pi}^{\pi} \ln \left[1 + \frac{\widetilde{V}(e^{j\omega}) - \widetilde{V}^{\mathrm{opt}}(e^{j\omega})}{\widetilde{V}^{\mathrm{opt}}(e^{j\omega})} \right] W(\omega) d\omega$$

$$\leq \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] W(\omega) d\omega$$

$$+ \int_{-\pi}^{\pi} \frac{\widetilde{V}(e^{j\omega}) - \widetilde{V}^{\mathrm{opt}}(e^{j\omega})}{\widetilde{V}^{\mathrm{opt}}(e^{j\omega})} W(\omega) d\omega$$

$$= \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] W(\omega) d\omega$$

$$+ \frac{1}{c} \int_{-\pi}^{\pi} \left[\widetilde{V}(e^{j\omega}) - \widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] d\omega$$

$$= \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] W(\omega) d\omega \tag{28}$$

with equality if and only if $\widetilde{V} = \widetilde{V}^{\mathrm{opt}}$. This implies that, for a given Φ_u, we have that

$$\sup_{V \in C(\mathbb{T}, \mathbb{R}_0^+)} \int_{-\pi}^{\pi} \ln \left[\frac{V(e^{j\omega})}{|G(e^{j\omega})|^2} \right] W(\omega) d\omega$$

$$= C_1 + \int_{-\pi}^{\pi} \ln \left[\widetilde{V}^{\mathrm{opt}}(e^{j\omega}) \right] W(\omega) d\omega$$

$$= C_2 - \int_{-\pi}^{\pi} \ln \left[\Phi_u(\omega) \right] W(\omega) d\omega$$

where the supremum is taken over all V satisfying the integral constraint on the experiment design problem, and C_2 is given by

$$C_2 = \int_{-\pi}^{\pi} \ln \left[\frac{2\pi n c \Phi_v(\omega)}{N |G(e^{j\omega})|^2} \right] W(\omega) d\omega.$$

Now, take Φ_u^{opt} as in (6). Then, following a similar derivation to that in (28), we have

$$C_2 - \int_{-\pi}^{\pi} \ln[\Phi_u(\omega)] W(\omega) d\omega \leq C_2 - \int_{-\pi}^{\pi} \ln \left[\Phi_u^{\mathrm{opt}}(\omega) \right] W(\omega) d\omega$$

with equality if and only if $\Phi_u = \Phi_u^{\mathrm{opt}}$. This proves that Φ_u^{opt} is the optimal solution of the experiment design problem. ∎

APPENDIX D
PROOF OF THEOREM 3

As in the proof of Lemma 2, by Theorems 1 and 2, the experiment design problem is equivalent to

$$\min_{\Phi_u > 0} \sup_{V \in C(\mathbb{T}, \mathbb{R}_0^+)} \frac{1}{2\pi} \int_{-\pi}^{\pi} V(e^{j\omega}) W(e^{j\omega}) d\omega,$$

$$\text{s.t.} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega = 1,$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} V(e^{j\omega}) d\omega = \frac{n}{N}. \tag{29}$$

Note that we have changed the $\leq$ sign in the input power constraint to an equality, since it is an active constraint.

We now fix Φ_u and define

$$\widetilde{V}(e^{j\omega}) := \frac{N}{2\pi n}\frac{\Phi_v(\omega)}{\Phi_u(\omega)}V(e^{j\omega}).$$

Then, problem (29) for $\widetilde{V}$ becomes

$$\sup_{\widetilde{V}\in C(\mathbb{T},\mathbb{R}_0^+)} \frac{n}{N}\int_{-\pi}^{\pi}\frac{\Phi_v(\omega)W(e^{j\omega})}{\Phi_u(\omega)}\widetilde{V}(e^{j\omega})d\omega,$$

$$\text{s.t.} \quad \int_{-\pi}^{\pi}\widetilde{V}(e^{j\omega})d\omega = 1.$$

This is a mass distribution problem (e.g., see of [16, eq. (17)]); hence the optimal cost is

$$\sup_{\substack{\widetilde{V}\in C(\mathbb{T},\mathbb{R}_0^+) \\ \int_{-\pi}^{\pi}\widetilde{V}(e^{j\omega})d\omega = 1}} \frac{n}{N}\int_{-\pi}^{\pi}\frac{\Phi_v(\omega)W(e^{j\omega})}{\Phi_u(\omega)}\widetilde{V}(e^{j\omega})d\omega$$

$$= \frac{n}{N}\max_{\omega\in[-\pi,\pi]}\frac{\Phi_v(\omega)W(e^{j\omega})}{\Phi_u(\omega)}.$$

Now, if it were not true that $\Phi_u = \Phi_u^{\text{opt}}$ for almost every $\omega \in [-\pi,\pi]$, as defined in (7), then $\Phi_u(\omega) < \Phi_u^{\text{opt}}(\omega)$ for some $\omega = \omega^* \in [-\pi,\pi]$. Otherwise, $\Phi_u > \Phi_u^{\text{opt}}$ in a set of positive measure, which implies that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)d\omega > \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_v(\omega)W(e^{j\omega})d\omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_v(\tau)W(e^{j\tau})d\tau} = 1$$

thus contradicting the constraint on Φ_u. Therefore

$$\sup_{\substack{\widetilde{V}\in C(\mathbb{T},\mathbb{R}_0^+) \\ \int_{-\pi}^{\pi}\widetilde{V}(e^{j\omega})d\omega = 1}} \frac{n}{N}\int_{-\pi}^{\pi}\frac{\Phi_v(\omega)W(e^{j\omega})}{\Phi_u(\omega)}\widetilde{V}(e^{j\omega})d\omega$$

$$= \frac{n}{N}\max_{\omega\in[-\pi,\pi]}\frac{\Phi_v(\omega)W(e^{j\omega})}{\Phi_u(\omega)}$$

$$\geq \frac{n}{N}\frac{\Phi_v(\omega^*)W(e^{j\omega^*})}{\Phi_u(\omega^*)}$$

$$> \frac{n}{2\pi N}\int_{-\pi}^{\pi}\Phi_v(\omega)W(e^{j\omega})d\omega$$

and the cost is minimized with $\Phi_u = \Phi_u^{\text{opt}}$. ∎

APPENDIX E
ADDITIONAL RESULTS

Lemma 5 (Lieb's Lemma): Let $\{c_k\}_{k\in\mathbb{N}}$ be a monotonically nonincreasing sequence in $[0,1]$ such that $\sum_{k=1}^{\infty}c_k = N \in \mathbb{N}$. Then there exist N orthonormal elements of $l_2(\mathbb{R})$, $v^i = (v_1^i, v_2^i, \ldots)$ $(i = 1, \ldots, N)$, such that $\sum_{i=1}^{N}(v_k^i)^2 = c_k$ for all $k \in \mathbb{N}$.

Proof: See lemma from [47, p. 458]. ∎

Theorem 5 (Weierstrass' Second Theorem): If $f \in C(\mathbb{T},\mathbb{R}^n)$ then, for every $\epsilon > 0$ there is a (vector-valued) trigonometric polynomial

$$g(e^{j\omega}) = \sum_{i=-p}^{p}a_{|i|}e^{j\omega i}, \qquad \omega \in [-\pi,\pi]$$

with $a_i \in \mathbb{R}^n$ $(i = 0, \ldots, p)$, such that $|f(z) - g(z)|_2 \leq \epsilon$ for every $z \in \mathbb{T}$.

Proof: This is a simple extension of [48, Ch. I, Th. 21] to vector-valued functions on $\mathbb{T}$. ∎

Theorem 6 (Maximum Modulus Theorem): Let f be an analytic function in a region Ω, which contains a closed disk of radius r and center a. Then

$$|f(a)| \leq \max_{\theta\in[-\pi,\pi]}\left|f(a + re^{j\theta})\right|. \tag{30}$$

Moreover, equality in (30) occurs if and only if f is constant in Ω.

Proof: See [49, Th. 10.24]. ∎

Theorem 7 (Montel's Theorem): Let $\mathcal{F}$ be a set of analytic functions in a region Ω. Then $\mathcal{F}$ is uniformly bounded on each compact subset of Ω (or, equivalently, *locally bounded*), if and only if $\mathcal{F}$ is a *normal family*, i.e., every sequence of functions in $\mathcal{F}$ contains a subsequence which converges uniformly on compact subsets of Ω.

Proof: See [50, Th. 2.9]. ∎

Theorem 8 (Uniform Convergence of Analytic Functions): Let $\{f_k\}_{k\in\mathbb{N}}$ be a sequence of analytic functions in a region Ω, which converges to a function f uniformly on compact subsets of Ω. Then f is also analytic in Ω, and $\{f_k'\}_{k\in\mathbb{N}}$ converges to f' uniformly on compact subsets of Ω.

Proof: See [49, Th. 10.28]. ∎

REFERENCES

[1] G. C. Goodwin, S. F. Graebe, and M. E. Salgado, *Control System Design*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[2] P. Eykhoff, *System Identification: Parameter and State Estimation*. Hoboken, NJ: Wiley, 1974.

[3] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic Press, 1977.

[4] T. Söderström and P. Stoica, *System Identification*. Hertfordshire, U.K.: Prentice-Hall, 1989.

[5] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[6] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York: IEEE Press, 2001.

[7] D. R. Cox, *Planning of Experiments*. Hoboken, NJ: Wiley, 1958.

[8] O. Kempthorne, *Design and Analysis of Experiments*. Hoboken, NJ: Wiley, 1952.

[9] V. V. Fedorov, *Theory of Optimal Experiments*. New York: Academic Press, 1972.

[10] R. K. Mehra, "Optimal input signals for parameter estimation in dynamic systems—Survey and new results," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 753–768, Dec. 1974.

[11] M. Zarrop, *Optimal Experiment Design for Dynamic System Identification*. Berlin, Germany: Springer, 1979.

[12] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.

[13] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, Mar. 2005.

[14] J. C. Agüero and G. C. Goodwin, "Choosing between open- and closed-loop experiments in linear system identification," *IEEE Trans. Autom. Control*, vol. 52, no. 8, pp. 1475–1480, Aug. 2007.

[15] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, Oct. 2006.

[16] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.

[17] J. Mårtensson and H. Hjalmarsson, "Robust input design using sum of squares constraints," in *14th IFAC Symp. System Identification (SYSID'06)*, Newcastle, Australia, 2006, pp. 1352–1357.

[18] H. Hjalmarsson, J. Mårtensson, and B. Wahlberg, "On some robustness issues in input design," in *Proc. 14th IFAC Symp. System Identification*, Newcastle, Australia, 2006.

[19] J. Mårtensson and H. Hjalmarsson, "How to make bias and variance errors insensitive to system and model complexity in identification," *IEEE Trans. Autom. Control*, vol. 56, no. 1, pp. 100–112, Jan. 2011.

[20] J. Mårtensson, "Geometric analysis of stochastic model errors in system identification," Ph.D. dissertation, Royal Inst. Technol. (KTH), Stockholm, Sweden, Oct. 2007, tRITA-EE 2007:061.

[21] C. R. Rojas, G. C. Goodwin, J. S. Welsh, and A. Feuer, "Optimal experiment design with diffuse prior information," in *Proc. Eur. Control Conf. (ECC)*, Kos, Greece, Jul. 2007, pp. 935–940.

[22] C. R. Rojas, J. C. Agüero, J. S. Welsh, and G. C. Goodwin, "On the equivalence of least costly and traditional experiment design for control," *Automatica*, vol. 44, no. 11, pp. 2706–2715, 2008.

[23] Z. D. Yuan and L. Ljung, "Unprejudiced optimal open loop input design for identification of transfer functions," *Automatica*, vol. 21, no. 6, pp. 697–708, Nov. 1985.

[24] C. R. Rojas, J. S. Welsh, and J. C. Agüero, "Fundamental limitations on the variance of estimated parametric models," *IEEE Trans. Autom. Control*, vol. 54, no. 5, pp. 1077–1081, May 2009.

[25] J. S. Welsh and G. C. Goodwin, "Finite sample properties of indirect nonparametric closed-loop identification," *IEEE Trans. Autom. Control*, vol. 47, no. 8, pp. 1277–1292, Aug. 2002.

[26] T. Zhou, "Frequency response estimation for NCFs of an MIMO plant from closed-loop time-domain experimental data," *IEEE Trans. Autom. Control*, vol. 51, no. 1, pp. 38–51, Jan., 2006.

[27] M. C. Campi and E. Weyer, "Guaranteed non-asymptotic confidence regions in system identification," *Automatica*, vol. 41, pp. 1751–1764, 2005.

[28] P. L. Duren, *Theory of H^p Spaces*. San Diego, CA: Academic Press, 1970.

[29] P. Koosis, *Introduction to H_p Spaces*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[30] M. M. Seron, J. H. Braslavsky, and G. C. Goodwin, *Fundamental Limitations in Filtering and Control*. Berlin, Germany: Springer-Verlag, 1997.

[31] L. Ljung, "Asymptotic variance expressions for identified black-box transfer function models," *IEEE Trans. Autom. Control*, vol. AC-30, no. 9, pp. 834–844, Sep. 1985.

[32] G. C. Goodwin, J. S. Welsh, A. Feuer, and M. Derpich, "Utilizing prior knowledge in robust optimal experiment design," in *Proc. 14th IFAC SYSID*, Newcastle, Australia, 2006, pp. 1358–1363.

[33] G. C. Goodwin, C. R. Rojas, and J. S. Welsh, "Good, bad and optimal experiments for identification," in *Forever Ljung in System Identification—Workshop on the Occasion of Lennart Ljung's 60th Birthday*, T. Glad, Ed., Sep. 2006.

[34] J. S. Welsh, G. C. Goodwin, and A. Feuer, "Evaluation and comparison of robust optimal experiment design criteria," in *Proc. Amer. Control Conf.*, Minneapolis, MN, 2006, pp. 1659–1664.

[35] G. C. Goodwin, J. I. Yuz, and J. C. Agüero, "Relative error issues in sampled data models," in *IFAC World Congress*, Seoul, Korea, 2008.

[36] G. C. Goodwin, J. C. Agüero, J. S. Welsh, J. I. Yuz, G. J. Adams, and C. R. Rojas, "Robust identification of process models from plant data," *J. Process Control*, vol. 18, no. 9, pp. 810–820, 2008.

[37] J. C. Agüero, G. C. Goodwin, T. Söderström, and J. I. Yuz, "Sampled data errors-in-variables systems," in *Proc. 15th IFAC Symp. System Identification*, Saint Malo, France, 2009, pp. 1157–1162.

[38] H. W. Bode, *Network Analysis and Feedback Amplifier Design*. New York: Van Nostrand, 1945.

[39] G. Goodwin, J. Welsh, A. Feuer, and M. Derpich, "Utilizing prior knowledge in robust optimal experiment design," in *Proc. 14th IFAC Symp. System Identification*, Newcastle, Australia, 2006, pp. 1358–1363.

[40] C. R. Rojas, "Robust experiment design," Ph.D. dissertation, The Univ. Newcastle, Callaghan, Australia, Jun. 2008.

[41] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 1, pp. 26–37, Jan. 1980.

[42] J. Skilling, "The axioms of maximum entropy," in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, G. J. Erickson and C. R. Smith, Eds. Dordrecht, The Netherlands: Kluwer Academic, 1988, vol. 1, pp. 173–187.

[43] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1534–1549, Oct. 2005.

[44] S. Karlin, *Mathematical Methods and Theory in Games, Programming, and Economics*. Reading, MA: Addison-Wesley, 1959.

[45] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[46] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ: Wiley, 1969.

[47] E. H. Lieb, "Variational principle for many-fermion systems," *Phys. Rev. Lett.*, vol. 46, no. 7, pp. 457–459, Feb. 1981.

[48] N. I. Achieser, *Theory of Approximation*. New York: Frederick Ungar, 1956.

[49] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.

[50] J. B. Conway, *Functions of One Complex Variable*, 2nd ed. New York: Springer-Verlag, 1978.

Cristian R. Rojas was born in 1980. He received the M.S. degree in electronics engineering from the Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2004, and the Ph.D. degree in electrical engineering from The University of Newcastle, Newcastle, Australia, in 2008.

Since October 2008, he has been with the Royal Institute of Technology, Stockholm, Sweden, where he is currently an Assistant Professor of the Automatic Control Laboratory, School of Electrical Engineering. His research interest is in system identification.

Juan-Carlos Agüero (S'02–M'04) was born in Osorno, Chile. He received the professional title of Ingeniero civil electrónico and a Master of engineering from the Universidad Técnica Federico Santa María, Chile, in 2000, and the Ph.D. degree from The University of Newcastle, Callaghan, Australia, in 2006.

He gained industrial experience from 1997 to 1998 in the copper mining industry at El Teniente, Codelco, Chile. He is currently working as a Research Academic at The University of Newcastle. His research interests include system identification and control.

James S. Welsh (S'02–M'03) was born in Maitland, Australia, in 1965. He received the B.E. degree (Hons. I) in electrical engineering in 1997 and the Ph.D. degree, which studied ill-conditioning problems arising in system identification, in 2004, both from The University of Newcastle, Callaghan, Australia.

During the last several years, he has been actively involved in research projects at the Centre for Complex Dynamic Systems and Control, The University of Newcastle, including Powertrain Control, Model Predictive Control, and System Identification Toolboxes. His research interests include auto-tuning, system identification, and process control. He is currently a Senior Lecturer in the School of Electrical Engineering and Computer Science, The University of Newcastle.

Graham C. Goodwin (F'86) received the B.Sc. (physics), B.Eng. (electrical engineering), and Ph.D. degrees from the University of New South Wales, Sydney, Australia.

He is currently Professor Laureate of Electrical Engineering at the University of Newcastle, Callaghan, Australia and is Director of The University of Newcastle Priority Research Centre for Complex Dynamic Systems and Control. He holds Honorary Doctorates from the Lund Institute of Technology, Sweden and the Technion—Israel Institute of Technology. He is the coauthor of eight books, four edited books, and many technical papers.

Dr. Goodwin is the recipient of the Control Systems Society 1999 Hendrik Bode Lecture Prize, a Best Paper Award by the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, a Best Paper Award by the *Asian Journal of Control*, and two Best Engineering Text Book Awards from the International Federation of Automatic Control in 1984 and 2005. In 2008 he received the Quazza Medal from the International Federation of Automatic Control, and in 2010 he received the IEEE Control Systems Award. He is an Honorary Fellow of Institute of Engineers, Australia; a Fellow of the International Federation of Automatic Control; a Fellow of the Australian Academy of Science; a Fellow of the Australian Academy of Technology, Science and Engineering; a Member of the International Statistical Institute; a Fellow of the Royal Society, London; and a Foreign Member of the Royal Swedish Academy of Sciences.

Arie Feuer (F'04) received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1967 and 1973, respectively and the Ph.D. degree from Yale University, New Haven, CT, in 1978.

He has been with the Electrical Engineering Department, Technion since 1983 where he is currently a Chaired Professor. From 1967 to 1970 he was in industry working on automation design and between 1978 and 1983 with Bell Labs in Holmdel, N.J. In the last 22 years he has been regularly visiting the EE&CS department at the University of Newcastle. His current research interests include: medical imaging—in particular, ultrasound and CT; resolution enhancement of digital images and videos; 3-D video and multiview data; sampling and combined representations of signals and images; adaptive systems in signal processing and control; and applications of vector quantization.

Dr. Feuer received in 2009 an honorary doctorate from the University of Newcastle. Between the years 1994 and 2002 he served as the President of the Israel Association of Automatic Control and was a member of the IFAC Council during the years 2002–2005.