Robust Optimal Experiment Design for System Identification

Graham C. Goodwin [a] Cristian R. Rojas [a] James S. Welsh [a] Arie Feuer [b]

[a] *EE & CS, The University of Newcastle, Australia 2308.*

[b] *EE, Technion, Israel.*

Abstract

This paper develops the idea of min-max robust experiment design for dynamic system identification. The idea of min-max experiment design has been explored in the statistics literature. However, the technique is virtually unknown by the engineering community and, accordingly, there has been little prior work on examining its properties when applied to dynamic system identification. This paper initiates an exploration of these ideas. The paper considers linear continuous time systems with energy (or power) bounded inputs. We assume that the parameters lie in a given compact set and optimise the worst case over this set. We also provide a detailed analysis of the solution for an illustrative one parameter example and propose a convex optimisation algorithm that can be applied more generally to a discretised approximation to the design problem. We explore the role played by different design criteria and present a simulation example illustrating the merits of the proposed approach.

Key words: Robust Experiment Design, Optimal Input Design.

1 Introduction

The goal of experiment design is to adjust the experimental conditions so that maximal information is gained from the experiment. Background to this problem can be found in early statistics literature (Wald, 1943; Cox, 1958; Kempthorne, 1952; Kiefer and Wolfowitz, 1960; Karlin and Studden, 1966; Fedorov, 1972; Whittle, 1973; Wynn, 1972) as well as in the engineering literature (Levadi, 1966; Gagliardi, 1967; Goodwin and Payne, 1973; Goodwin et al., 1973b,a; Arimoto and Kimura, 1973; Mehra, 1974; Goodwin and Payne, 1977; Zarrop, 1979; Hildebrand and Gevers, 2003a). A recent survey is contained in (Gevers, 2005) where many additional references can be found. The focus in the engineering literature has been predominately on experiment design for dynamic system identification.

A key issue with experiment design for dynamic systems is that the model is typically nonlinearly parameterised. This means, amongst other things, that the Fisher information matrix (Goodwin and Payne, 1977), which is typically used as the basis for experiment design, depends, inter alia, on the true system parameters (i.e. the nominal optimal experiment depends on the very thing that the experiment is aimed at finding).

This issue has been recognised in the statistics literature where several approaches have been explored. These include:

- Sequential design, where one iterates between parameter estimation, on the one hand, and experiment design using the current parameter estimates, on the other - see (Chernoff, 1975; Ford and Silvey, 1980; Ford et al., 1985; Wu, 1985; Müller and Pötscher, 1992; Walter and Pronzato, 1997).
- Bayesian design (Chaloner and Larntz, 1989; Atkinson and Doner, 1992; Atkinson et al., 1993; Chaloner and Verdinelli, 1995; Sebastiani and Wynn, 2000; El-Gamal and Palfrey, 1996). The Bayesian approach is characterised by the minimisation of the expected value (over the prior parameter distribution) of a local optimality criterion related to the information matrix.
- Min-max design (Pronzato and Walter, 1988; Landaw, 1984; D'Argenio and Van Guilder, 1988; Melas, 1978; Fedorov, 1980; Biedermann and Dette, 2003; Dette et al., 2003).

However, there has been little work on robust experiment design for engineering problems. This has been highlighted in the recent survey paper (Hjalmarsson, 2005), where it is stated that "...as usual in experiment design, in order to compute the optimal design the true system has to be known. Methods that are robust with respect to uncertainty about the system is a wide open research field."

Preliminary work in the engineering literature on robust experiment design includes substantial work on iterative design (Hjalmarsson, 2005; Gevers, 2005) and an insightful

sub-optimal min-max solution for a one parameter problem (Walter and Pronzato, 1997). Actually the latter problem will be discussed in detail in section 3 of the current paper. Also, a number of very recent engineering papers refer to the idea of min-max optimal experiment design - see for example papers presented at SYSID'06, e.g., (Gevers and Bombois, 2006; Goodwin et al., 2006; Mårtensson and Hjalmarsson, 2006).

Our goal in the current paper is to develop the idea of min-max optimal experiment design for dynamic system identification. To gain insight into this approach, we explore an illustrative example in depth.

We assume prior knowledge in the form that the system parameters, θ, are contained in a given compact set Θ. We then choose a design criterion $f(M(\theta), \theta)$ where $M(\theta)$ is the Fisher information matrix, evaluated at θ, and design the experiment to optimise the worst case of $f(M(\theta), \theta)$ over Θ. Notice that this differs from the usual approaches to experiment design *in the engineering literature* which typically optimise $f(M(\theta_0), \theta_0)$ for some given nominal value θ_0.

Our approach is more akin to the usual formulation of robust optimal control which typically considers the worst case (Zhou et al., 1996). Indeed, there are substantial links between the work presented here and continuous game theory (Başar and Olsder, 1995; Başar and Bernhard, 1995; Fudenberg and Tirole, 1991; Owen, 1995; Szép and Forgó, 1985). We explore some of these connections below.

The merits of the approach proposed in this paper are illustrated by an example (presented in Section 5) which shows, for a realistic second order system, that an order of magnitude improvement in the worst case performance in experiment design can be achieved at the expense of only a few percent degradation in the nominal performance.

The layout of the remainder of the paper is as follows: In Section 2 we give a general formulation of the min-max approach to robust optimal experiment design. Section 3 explores an illustrative one parameter example in considerable detail so as to give insight into the problem. In Section 4 we describe the extension to multi-parameter systems. In Section 5 we present several results illustrating the merits of the proposed approach. Finally, in Section 6 we draw conclusions.

2 Experiment Design Criteria

2.1 The Information Matrix

So as to be specific we first consider a single input single output linear discrete time system, with input $\{u_t\}$ and output $\{y_t\}$, of the form:

$$y_t = G_1(q)u_t + G_2(q)w_t \tag{1}$$

where G_1 and G_2 are rational transfer functions, q is the forward shift operator, $G_2(\infty) = 1$, and $\{w_t\}$ is zero mean Gaussian white noise of variance Σ. We let $\beta \triangleq [\theta^T,\ \gamma^T,\ \Sigma]^T$ where θ denotes the parameters in G_1 and γ denotes the parameters in G_2.

We recall that the log likelihood function (Goodwin and Payne, 1977) for data Y given parameters β, is given by

$$\log p(Y|\beta) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\Sigma - \frac{1}{2\Sigma}\sum_{t=1}^{N}\varepsilon_t^2 \tag{2}$$

where

$$\varepsilon_t \triangleq G_2(q)^{-1}[y_t - G_1(q)u_t]\ . \tag{3}$$

Fisher's information matrix is obtained by taking the following expectation (Goodwin and Payne, 1977)

$$M \triangleq \mathbb{E}_{Y|\beta}\left[\left(\frac{\partial \log p(Y|\beta)}{\partial\beta}\right)\left(\frac{\partial \log p(Y|\beta)}{\partial\beta}\right)^T\right] \tag{4}$$

where from (2)

$$\frac{\partial \log p(Y|\beta)}{\partial\beta} = -\frac{1}{\Sigma}\sum_{t=1}^{N}\varepsilon_t\frac{\partial\varepsilon_t}{\partial\beta} - \frac{1}{2\Sigma}\frac{\partial\Sigma}{\partial\beta}\left[N - \frac{1}{\Sigma}\sum_{t=1}^{N}\varepsilon_t^2\right], \tag{5}$$

from (3)

$$\frac{\partial\varepsilon_t}{\partial\beta} = -G_2(q)^{-1}\left\{\frac{\partial G_2(q)}{\partial\beta}\varepsilon_t + \frac{\partial G_1(q)}{\partial\beta}u_t\right\} \tag{6}$$

and where $\mathbb{E}_{Y|\beta}$ denotes the expectation over the distribution of the data given β.

We assume an open loop experiment so that w_t and u_t are uncorrelated. We also assume that G_1, G_2 and Σ are independently parameterised. Taking expectations, as in (4), M can be partitioned as

$$M = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix} \tag{7}$$

where M_1 is the part of the information matrix related to θ, and M_2 is independent of the input. Thus,

$$M_1 \triangleq \frac{1}{\Sigma}\sum_{t=1}^{N}\left(\frac{\partial\varepsilon_t}{\partial\theta}\right)\left(\frac{\partial\varepsilon_t}{\partial\theta}\right)^T \tag{8}$$

where $\partial\varepsilon_t/\partial\theta$ satisfies

$$\frac{\partial\varepsilon_t}{\partial\theta} = -G_2(q)^{-1}\frac{\partial G_1(q)}{\partial\theta}u_t\,. \tag{9}$$

Notice that M_1 depends on the full parameter vector β. Assuming N is large, it is more convenient to work with the scaled average information matrix for the parameters θ, (Goodwin and Payne, 1977; Walter and Pronzato, 1997),

$$\overline{M}(\beta, \phi_u) \triangleq \lim_{N\to\infty} \frac{1}{N} M_1 \Sigma \,. \tag{10}$$

Utilising Parseval's Theorem, we finally have that

$$\overline{M}(\beta, \phi_u) = \frac{1}{\pi} \int_0^\pi \widetilde{M}(\beta, \omega) \phi_u(e^{j\omega}) d\omega \tag{11}$$

where

$$\widetilde{M}(\beta, \phi_u) \triangleq \mathrm{Re} \left\{ \frac{\partial G_1(e^{j\omega})}{\partial \theta} \left| G_2(e^{j\omega}) \right|^{-2} \left[\frac{\partial G_1(e^{j\omega})}{\partial \theta} \right]^H \right\} \tag{12}$$

and ϕ_u is the discrete input spectral density (considered as a generalised function). Here, H is the conjugate transpose operator.

It is also possible to do a parallel development (Goodwin and Payne, 1977; Walter and Pronzato, 1997) for continuous time models. In the latter case, (11) is replaced by

$$\overline{M}(\beta, \phi_u) = \int_0^\infty \widetilde{M}(\beta, \omega) \phi_u(\omega) d\omega \tag{13}$$

where

$$\widetilde{M}(\beta, \phi_u) \triangleq \mathrm{Re} \left\{ \frac{\partial G_1(j\omega)}{\partial \theta} \left| G_2(j\omega) \right|^{-2} \left[\frac{\partial G_1(j\omega)}{\partial \theta} \right]^H \right\}, \tag{14}$$

G_1 and G_2 are continuous time transfer functions (assumed independently parameterised) and ϕ_u is the continuous time input spectral density.

Notice that the results presented below do not depend on Σ since it appears as a scaling factor in (8). Also, we see from (13) that, in $\overline{M}(\beta, \phi_u)$, G_2 simply plays the role of a frequency dependent weighting. This is easily included in the analysis. However, for simplicity we assume white noise, although the extension to non-white noise is straightforward. Hence in the sequel we refer only to θ.

2.2 Brief Review of Design Criteria for Nominal Experiment Design

Since $\overline{M}$ is a matrix, we need a scalar measure of $\overline{M}$ for the purpose of experiment design. In the nominal case typically treated in the engineering literature (i.e. when a fixed prior estimate of θ is used), several measures of the "size" of $\overline{M}$ have been proposed. Some examples are:

(i) D - optimality (Goodwin and Payne, 1977)

$$J_d(\theta, \phi_u) \triangleq [\det \overline{M}(\theta, \phi_u)]^{-1} \,. \tag{15}$$

(ii) Experiment design for robust control (Hjalmarsson, 2005; Hildebrand and Gevers, 2003b,a)

$$J_{rc}(\theta, \phi_u) \triangleq \sup_\omega g(\theta, \omega)^H \overline{M}^{-1} g(\theta, \omega) \tag{16}$$

where g is a frequency dependent vector related to the ν-gap (Hildebrand and Gevers, 2003b,a).

Many other criteria have been described in the statistics literature, such as A-optimality ($\mathrm{tr}\, \overline{M}(\theta, \phi_u)^{-1}$), L-optimality ($\mathrm{tr}\, W \overline{M}(\theta, \phi_u)^{-1}$, for some $M \geq 0$) and E-optimality ($\lambda_{\max}(\overline{M}(\theta, \phi_u)^{-1})$); see (Kiefer, 1974). On the other hand, in the engineering literature, (Bombois et al., 2005), for example, proposed a criterion that leads to the required accuracy to achieve a given level of robust control performance. Other criteria will be discussed in Section 4.

A common feature of all these nominal experiment design approaches is that they are aimed at choosing ϕ_u to minimise a function of the type shown in (15) and (16). Notice, however, that the optimal input spectrum depends, inter-alia, on the unknown parameter vector θ.

2.3 Min-Max Robust Design

A min-max robust design criterion is the basis of the approach described in the current paper. Specifically, we assume that we have available a-priori information showing that the parameters can take any value in a compact set Θ. We also constrain the allowable set of input signals. A typical constraint (Goodwin and Payne, 1977; Zarrop, 1979; Walter and Pronzato, 1997) used in experiment design is that the input energy is constrained, i.e. we define [1]

$$\mathscr{S}(\mathbb{R}_0^+) \triangleq \left\{ \phi_u : \mathbb{R} \to [0, 1] : \begin{array}{c} \mathrm{supp}\, \phi_u \subset \mathbb{R}_0^+ \quad \text{and} \\ \int_{-\infty}^{\infty} \phi_u(\omega) d\omega = 1 \end{array} \right\}. \tag{17}$$

The min-max robust optimal input spectral density, ϕ_u^*, is then chosen as

$$\phi_u^* = \arg \min_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \sup_{\theta \in \Theta} J(\overline{M}(\theta, \phi_u), \theta) \tag{18}$$

where J is an appropriate scalar measure of $\overline{M}$. We are assuming for the moment that ϕ_u^* exists and is unique; these points will be studied in the next section. Notice also that we allow J to depend explicitly on θ; this point will be of practical importance – see discussion below.

[1] In general, given a set $X \subset \mathbb{R}$, we will denote by $\mathscr{S}(X)$ the set of all generalised functions ϕ_u on $\mathbb{R}$ (Rudin, 1973) such that ϕ_u is the derivative of some probability distribution function on $\mathbb{R}$, and $\mathrm{supp}\, \phi_u \subset X$, where $\mathrm{supp}\, \phi_u$ is the support of ϕ_u (i.e. roughly speaking, $\mathscr{S}(X)$ is the set of all (generalised) probability density functions on X).

2.4 A Mixed Policy Game Approach

An alternative approach to that described above would be to extend the space to include "mixed policies" (Başar and Bernhard, 1995) by introducing a (probability) density ξ on Θ. The counterpart of (18) would now take the form:

$$\phi_u^* = \arg \min_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \sup_{\xi \in \mathscr{S}(\Theta)} J'(\xi, \phi_u) \tag{19}$$

where J' is an appropriate scalar measure of the form:

$$J' \triangleq \int f \left[\int S_\theta \widetilde{M}(\theta, \omega) S_\theta^T \phi_u(\omega) d\omega \right] \xi(\theta) d\theta \tag{20}$$

where f is a scalar-valued function, e.g. $f(L) = \operatorname{tr} L^{-1}$ or $f(L) = \lambda_{\max}(L^{-1})$; $\widetilde{M}$ is the single frequency information matrix and S_θ is a parameter dependent scaling matrix (see Section 4.1 for further discussion of S_θ).

Notice that if f were linear, it could be introduced into the inner integral, and in that case it can be shown that this approach is equivalent to the previous one (see the proof of Theorem 3 for an idea of how the equivalence is established).

3 An Illustrative Example

Before delving into the general multi-parameter case, we first consider an illustrative continuous time one parameter problem to gain insight. We take $G_2(s) = 1$ and let

$$G_1(s) = \frac{1}{s/\theta + 1}\,. \tag{21}$$

Notice that this problem has also been discussed in Goodwin and Payne (1977) for the case of nominal experiment design. Also, the problem has been studied in Walter and Pronzato (1997) p339 in the context of min-max robust experiment design but where the input is restricted to a single sinusoid. Actually we will see below that the later restriction unduely limits the solution space and does not lead to the optimal strategy. This is heuristically reasonable since if θ lies in an interval, then it makes sense to spread the input energy in some sense to cover the possible θ scenarios.

For the model (21), it follows that

$$\overline{M}(\theta, \phi_u) = \int_0^\infty \widetilde{M}(\theta, \omega) \phi_u(\omega) d\omega \tag{22}$$

where $\widetilde{M}$ is the "single frequency" normalised information matrix given by

$$\begin{aligned} \widetilde{M}(\theta, \omega) &= \left| \frac{\partial G_1(\theta, \omega)}{\partial \theta} \right|^2 \\ &= \frac{\omega^2/\theta^4}{(\omega^2/\theta^2 + 1)^2}\,. \end{aligned} \tag{23}$$

3.1 Nominal Optimal Experiment Design for the Illustrative Example

Before turning to the robust design problem, we will briefly review the nominal experiment design problem for this case. Here one assumes that a prior estimate, $\widehat{\theta}$, of θ is available. Based on this information, the function ϕ_u is chosen so as to optimise some scalar-valued function of $\overline{M}(\widehat{\theta}, \phi_u)$ subject to a constraint on the input power. In the nominal case it can be shown that we only need to use a single frequency input for this example (Goodwin and Payne, 1977), namely, $\phi_u(\omega) = \delta(\omega - \omega^*)$. Moreover, by differentiation of the single frequency information matrix given in (23), it is readily seen that the optimal input frequency is

$$\omega^* = \theta\,. \tag{24}$$

This is an intuitively pleasing result, i.e. one places the test signal at the (nominal) 3dB break point. However, equation (24) reinforces the fundamental difficulty in nominal experiment design, namely, the optimal experiment depends on the very thing that the experiment is aimed at estimating.

To gauge the importance of the dependence on θ, we notice that $\widetilde{M}(\theta, \omega)$ in our example decays at the rate of 40dB per decade as a function of both θ and ω. Hence, given the prior estimate of the parameter, $\widehat{\theta}$, say we choose $\widehat{\omega} = \widehat{\theta}$ for the input signal frequency. Also, say that the true parameter lies in the range $(0.1\,\widehat{\theta} \le \theta \le 10\,\widehat{\theta})$, then $\min_{\theta \in \Theta} \widetilde{M}(\theta, \omega)$ is approximately $1/100^{th}$ of the nominal value! This seems to suggest that nominal experiment design is limited to those cases where an extremely good prior estimate is available. This point is reinforced in Figure 1 which shows a plot of $[\theta^2 \overline{M}(\theta, \phi_u)]^{-1}$ versus θ for the nominal optimal input.

Remark 1 *The reason for multiplying by θ^2 as in Figure 1 and then inverting is that $\overline{M}^{-1}$ is a variance measure and thus $[\theta^2 \overline{M}]^{-1}$ gives relative (mean square) errors. More will be said about this type of scaled cost function in the context of robust design below. Interestingly, the scaling turns out to be equivalent (for this example) to the MMDE cost function used in Walter and Pronzato (1997) p339.*

$\nabla\nabla\nabla$

3.2 Some Properties of the Min-Max Robust Optimal Experiment Design

We next turn to robust experiment design as described in Section 2. For the illustrative problem we use (see Remark 1 and Section 4)

$$J(\overline{M}(\theta, \phi_u), \theta) \triangleq [\theta^2\, \overline{M}(\theta, \phi_u)]^{-1}\,. \tag{25}$$

Thus, our min-max robust optimal experiment design can be stated as finding

$$\phi_u^* = \arg \min_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \overline{J}(\phi_u) \tag{26}$$

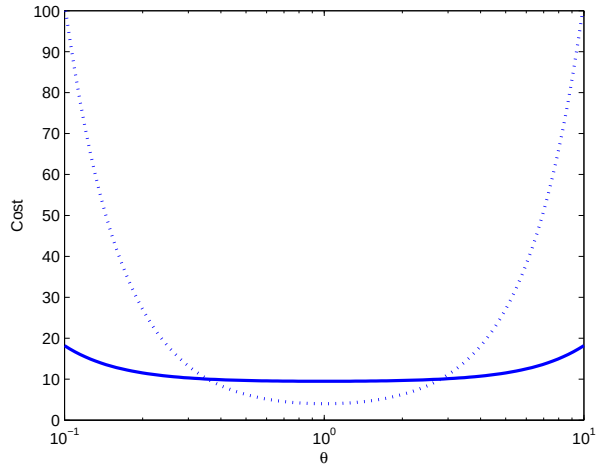

Fig. 1. $[\theta^2 \overline{M}(\theta, \phi_u)]^{-1}$ as a function of θ for nominal input (dotted) and '1/f' noise (solid).

where

$$\overline{J}(\phi_u) \triangleq \sup_{\theta \in \Theta} \left[\int_0^\infty \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2} \, \phi_u(\omega) \, d\omega \right]^{-1} \tag{27}$$

and

$$\Theta \triangleq \{\theta : \underline{\theta} \leq \theta \leq \overline{\theta}\} \, . \tag{28}$$

In the sequel, we will give further insights into the above design problem.

We first observe that, since the integrand in (27) is continuous in $\theta \in \Theta$ for every $\omega \in \mathbb{R}_0^+$ and it is bounded by an integrable function which is independent of θ (use e.g. C/ω^2, where C is large enough and independent of θ), the integral is continuous in θ; see (Bartle, 1966). This implies, with the compactness of Θ, that we can change the "sup" symbol in (27) to "max".

Furthermore, if we make the following changes of variables

$$\begin{aligned} x &\triangleq \frac{\ln\theta - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}} \\ y &\triangleq \frac{\ln\omega - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}} \\ \phi_u(\omega) &= \frac{2}{k\omega} \, \widetilde{\phi}_u \left(\frac{\ln\omega - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}} \right) \\ k &\triangleq 2(\ln\overline{\theta} - \ln\underline{\theta}) \end{aligned} \tag{29}$$

then the problem can be rewritten as

$$\widetilde{\phi}_u^* = \arg \max_{\widetilde{\phi}_u \in \mathscr{S}(\mathbb{R})} \min_{x \in [0,1]} \int_{-\infty}^{\infty} \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2} \, \widetilde{\phi}_u(y) dy \, . \tag{30}$$

To simplify the notation, let $F(x,y) \triangleq f(x-y)$, where $f(u) \triangleq e^{ku}/(e^{ku}+1)^2$.

The following theorems give some properties of ϕ_u^* and $\widetilde{\phi}_u^*$. Notice, however, that Lemma 2 and Theorem 5 are different in that Lemma 2 states that the optimal input has compact support, which is a technical requirement for proving other results. Theorem 5, on the other hand, states that the optimal input has finite support, which is a stronger result than Lemma 2, but its proof relies on the previous theorems.

Lemma 2 *Considering the problem stated in (30), the optimal input $\widetilde{\phi}_u^*$, if it exists, has all its energy inside $[0,1]$. Namely,*

$$\int_{\mathbb{R}-[0,1]} \widetilde{\phi}_u^*(y) dy = 0 \, . \tag{31}$$

Thus, the spectral density of the optimal input has compact support, i.e. $\widetilde{\phi}_u^* \in \mathscr{S}([0,1])$ (or, equivalently, $\phi_u^* \in \mathscr{S}(\Theta)$), so we can replace (30) *with*

$$\widetilde{\phi}_u^* = \arg \max_{\widetilde{\phi}_u \in \mathscr{S}([0,1])} \min_{x \in [0,1]} \int_0^1 \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2} \, \widetilde{\phi}_u(y) dy \, . \tag{32}$$

PROOF. Notice [2] that $\partial F/\partial y > 0$ for $y < x$ and $\partial F/\partial y < 0$ for $y > x$. It follows that for any $x \in [0,1]$ we have $\int_{-\infty}^{\infty} F(x,y)\widetilde{\phi}_u^*(y)dy \leq \int_{-\infty}^{\infty} F(x,y)\widetilde{\phi}_u'(y)dy$, where $\widetilde{\phi}_u'$ is given by

$$\begin{aligned} \widetilde{\phi}_u'(y) \triangleq \widetilde{\phi}_u^*(y)\mathcal{X}_{[0,1]}(y) &+ \delta(y) \int_{-\infty}^{0} \widetilde{\phi}_u^*(\tau)d\tau \\ &+ \delta(y-1) \int_1^{\infty} \widetilde{\phi}_u^*(\tau)d\tau \end{aligned} \tag{33}$$

and $\mathcal{X}_{[0,1]}$ denotes the indicator function of $[0,1]$. The result follows.

$\nabla\nabla\nabla$

Theorem 3 *For the problem stated in* (26) *or* (30)*, there exists at least one optimal input, that is, there exists a $\phi_u^* \in \mathscr{S}(\mathbb{R}_0^+)$ such that for every $\phi_u \in \mathscr{S}(\mathbb{R}_0^+)$,*

$$\overline{J}(\phi_u^*) \leq \overline{J}(\phi_u) \, . \tag{34}$$

PROOF. By Lemma 2, (30) can be related to a *two-person zero-sum game on the unit square* with *kernel* F, such that

[2] This proof was suggested by an anonymous reviewer.

the player x tries to minimise F by using a *pure strategy*, and player y tries to maximise this quantity by using a *mixed strategy* (Başar and Olsder, 1995). Hence, in order to prove the existence of $\widetilde{\phi}_u^*$ (or, which is the same, of ϕ_u^*), we can make use of a version of the Minimax Theorem, due to (Glicksberg, 1950), which states that if F is an upper or lower semicontinuous function on $[0,1] \times [0,1]$, then

$$\begin{aligned}
&\inf_{\mu_x \in \mathscr{S}([0,1])} \sup_{\mu_y \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 F(x,y) \mu_x(x) \mu_y(y)\, dy\, dx \\
&= \sup_{\mu_y \in \mathscr{S}([0,1])} \inf_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 F(x,y) \mu_x(x) \mu_y(y)\, dy\, dx \\
&\triangleq V_m \qquad (35)
\end{aligned}$$

where V_m is called the *average value* of the game. Furthermore, if F is continuous then, by a usual compactness argument (such as the one given in the paragraph before (29)), there exist $\mu_x^*,\ \mu_y^* \in \mathscr{S}([0,1])$ such that for every $\mu_x,\ \mu_y \in \mathscr{S}([0,1])$,

$$\begin{aligned}
&\int_0^1 \int_0^1 F(x,y) \mu_x^*(x) \mu_y(y)\, dy\, dx \\
&\qquad \le \int_0^1 \int_0^1 F(x,y) \mu_x^*(x) \mu_y^*(y)\, dy\, dx \qquad (36) \\
&\qquad = V_m \\
&\qquad \le \int_0^1 \int_0^1 F(x,y) \mu_x(x) \mu_y^*(y)\, dy\, dx\,.
\end{aligned}$$

This last expression says that (μ_x^*, μ_y^*) defines a *saddle point solution in mixed strategies* for the game (Başar and Olsder, 1995).

Since in our case F is continuous, these results apply. Furthermore, by (36) and the compactness of $[0,1]$,

$$\begin{aligned}
&\int_0^1 \int_0^1 F(x,y) \mu_x^*(x) \mu_y^*(y)\, dy\, dx \\
&\qquad = \min_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 F(x,y) \mu_x(x) \mu_y^*(y)\, dy\, dx \qquad (37) \\
&\qquad = \min_{x \in [0,1]} \int_0^1 F(x,y) \mu_y^*(y)\, dy\,.
\end{aligned}$$

Hence, by (35), (36) and (37),

$$\begin{aligned}
&\min_{x \in [0,1]} \int_0^1 F(x,y) \mu_y^*(y)\, dy \\
&\qquad = \max_{\mu_y \in \mathscr{S}([0,1])} \min_{x \in [0,1]} \int_0^1 F(x,y) \mu_y(y)\, dy \qquad (38)
\end{aligned}$$

so if we take $\widetilde{\phi}_u^* = \mu_y^*$, we have an optimal solution to (30). This proves the existence of an optimal input.

$\nabla\nabla\nabla$

Theorem 4 *For the problem stated in* (26) *or* (30)*, there is a unique optimal input. Moreover,* 0 *and* 1 *do not belong to the support of* $\widetilde{\phi}_u^*$ *(or, equivalently,* $\underline{\theta}, \overline{\theta} \notin \operatorname{supp} \phi_u^*$*), and* $\widetilde{\phi}_u^*$ *is symmetric with respect to* $1/2$*, that is,* $\widetilde{\phi}_u^*(y) = \widetilde{\phi}_u^*(1-y)$ *(i.e.,* $\phi_u^*(\omega) = \phi_u^*(\overline{\theta}\underline{\theta}/\omega)$*).*

PROOF. To prove this, we will make use of some results from (Karlin, 1957). However, to this end it is necessary to show first that f is a strictly positive definite function.

By Bochner's Theorem (Bochner, 1959), f is strictly positive definite if and only if its Fourier transform is positive everywhere. Now, the Fourier transform of f is

$$\begin{aligned}
\hat{f}(\omega) &\triangleq \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \\
&= \int_{-\infty}^{\infty} \frac{e^{kt}}{(e^{kt}+1)^2} e^{-j\omega t} dt \qquad (39) \\
&= \frac{1}{k^2} \frac{\pi\omega}{\sinh \pi\omega/k} > 0, \qquad \omega \in \mathbb{R}\,.
\end{aligned}$$

Hence, Bochner's Theorem applies.

Now, since f is strictly positive definite, analytic and $L^1(\mathbb{R})$, it is a *positive analytic regular Pólya frequency function* (Karlin, 1957, 1959). Together with the fact that $f'(0) = 0$, it implies by Theorems 1 and 2 of (Karlin, 1957) that $\widetilde{\phi}_u^*$ is unique, that $0, 1 \notin \operatorname{supp} \widetilde{\phi}_u^*$ and that $\widetilde{\phi}_u^*$ is symmetric with respect to $1/2$.

$\nabla\nabla\nabla$

Theorem 5 *For the problem stated in* (26) *or* (30)*, the optimal input (*ϕ_u^* *or* $\widetilde{\phi}_u^*$*, respectively) has finite support. That*

is, if $\widetilde{\phi}_u^*$ is such that

$$\min_{x\in[0,1]} \int_{-\infty}^{\infty} f(x-y)\widetilde{\phi}_u^*(y)dy = \max_{\widetilde{\phi}_u\in\mathscr{S}([0,1])} \left[\min_{x\in[0,1]} \int_{-\infty}^{\infty} f(x-y)\widetilde{\phi}_u(y)dy\right] \quad (40)$$

then $\operatorname{supp}\widetilde{\phi}_u^*$ *is finite.*

PROOF. This proof is based on a result in (Karlin, 1957), which is included here for the sake of completeness. We will focus on the problem stated in (30).

We will first show that if μ_x^* is defined as in the proof of Theorem 3, and $y_0 \in [0,1]$ is in the support of $\widetilde{\phi}_u^*$, then

$$\int_0^1 f(x-y_0)\mu_x^*(x)dx = V_m\,. \quad (41)$$

By (36), we have that

$$\int_0^1 f(x-y)\mu_x^*(x)dx \le V_m, \qquad y\in[0,1]\,. \quad (42)$$

If this inequality were strict for $y=y_0$, then by the continuity of f there would be an interval $[a,b]\subset[0,1]$ for which $a\le y_0\le b$ and

$$\int_0^1 f(x-y)\mu_x^*(x)dx < V_m, \qquad y\in[a,b]\,. \quad (43)$$

Thus, integrating both sides of (42) weighted by $\widetilde{\phi}_u^*$, and taking (43) into account, we obtain

$$\int_0^1\int_0^1 f(x-y)\mu_x^*(x)\widetilde{\phi}_u^*(y)dydx < V_m \quad (44)$$

which contradicts the definition of V_m. This proves (41).

Now, if $\operatorname{supp}\widetilde{\phi}_u^*$ is infinite, then (41) holds for an infinite number of points in a compact interval, so those points have at least one limit point. On the other hand, the integral of the left side of this expression is an analytic function of y in $\mathbb{R}$, and its right side is constant, so by a well-known result of complex analysis (Rudin, 1987) we have that

$$\int_0^1 f(x-y)\mu_x^*(x)dx = V_m, \qquad y\in\mathbb{R}\,. \quad (45)$$

However, since f is bounded and $f(u)\to 0$ for $|u|\to\infty$,

$$\lim_{y\to\infty}\int_0^1 f(x-y)\mu_x^*(x)dx = 0 \ne V_m \quad (46)$$

which contradicts (45). Thus, $\widetilde{\phi}_u^*$ (as well as ϕ_u^*) has finite support.

$\nabla\nabla\nabla$

Remark 6 *Theorem 5 basically says that the robust optimal input is a finite linear combination of sinusoids. This is a rather surprising result, since the nominal optimal input is a single sinusoid of frequency equal to θ, so one would expect that the robust optimal signal should have a continuous spectrum to account for all the possible values of this parameter. On the other hand, this property also says that it is very easy to implement such a signal; the only remaining problem is to determine the amplitudes and frequencies of its sinusoids. This will be addressed in Sections 3.3 (analytically) and 3.5 (numerically).*

$\nabla\nabla\nabla$

3.3 Explicit Solution of the Robust Experiment Design Problem

In order to obtain an explicit solution to the robust experiment design problem for the illustrative example, we will use some ideas [3] from (Bohnenblust et al., 1950; McKinsey, 1952; Karlin, 1957, 1959).

According to the results of the previous section, the problem can be stated as

$$\mu_y^* = \arg \max_{\mu_y\in\mathscr{S}([0,1])} \min_{\mu_x\in\mathscr{S}([0,1])} \int_0^1 \left[\int_0^1 \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2}\,\mu_x(x)\,dx\right]\mu_y(y)\,dy\,. \quad (47)$$

As before, the kernel of the associated game is $F(x,y)\triangleq f(x-y)$, where $f(u)\triangleq e^{ku}/(e^{ku}+1)^2$.

Let us denote the optimal mixed strategies of players x and y of game (47) by μ_x^* and μ_y^* ($=\widetilde{\phi}_u^*$), respectively. Also, let

$$\Psi_x(x)\triangleq\int_0^1 \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2}\,\mu_y^*(y)\,dy$$
$$\Psi_y(y)\triangleq\int_0^1 \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2}\,\mu_x^*(x)\,dx\,. \quad (48)$$

[3] The idea behind this method was suggested by an anonymous reviewer.

When k is very small, $F(x,y)$ is a strictly concave function in $y \in [0,1]$ for every $x \in [0,1]$. In that case, we have that $\Psi_y(y)$ (which appears in (47) by taking $\mu_x = \mu_x^*$) is a linear combination of strictly concave functions in y, so it is strictly concave as well, and it has a unique maximum at $y = 1/2$ (because of the symmetry of F). This means that

$$\widetilde{\phi}_u^*(y) = \mu_y^*(y) = \delta(y - 1/2)\,. \tag{49}$$

Hence, by the Minimax Theorem (Glicksberg, 1950), μ_x^* must satisfy

$$\mu_x^* = \arg \min_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \frac{e^{k(x-1/2)}}{(e^{k(x-1/2)}+1)^2}\, \mu_x(x)\, dx \tag{50}$$

that is,

$$\mu_x^*(x) = \frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-1)\,. \tag{51}$$

If we increase the value of k, $F(x,y)$ eventually ceases to be strictly concave in y for every x. This implies that there is a number $k_1 \in \mathbb{R}^+$ such that, for $k > k_1$, $\Psi_y(y)$ has at least two maxima. The value k_1 can be computed by setting the second derivative of $\Psi_y(y)$ equal to zero for $y = 1/2$:

$$\frac{\partial^2}{\partial y^2}\left[\int_0^1 \frac{e^{k_1(x-y)}}{(e^{k_1(x-y)}+1)^2}\, \mu_x^*(x)\, dx\right]\Bigg|_{y=1/2} = 0\,. \tag{52}$$

This implies

$$e^{3k_1/2} + e^{-3k_1/2} - 9e^{-k_1/2} - 9e^{k_1/2} = 16\,. \tag{53}$$

The only positive root of this equation is $k_1 = -2\ln(2-\sqrt{3}) \approx 2.6339$.

Thus, (49) and (51) hold for $0 < k \leq k_1$. However, for values of k slightly greater than k_1, (51) still holds, so $\Psi_y(y)$, with μ_x^* given by (51), has two maxima at, say, y' and $1-y'$, where

$$\frac{\partial}{\partial y}\left[\int_0^1 \frac{e^{k(x-y)}}{(e^{k(x-y)}+1)^2}\, \mu_x^*(x)\, dx\right]\Bigg|_{y=y'} = 0\,. \tag{54}$$

This equation has only one real solution y' between 0 and $1/2$, namely

$$y' = \frac{1}{k}\ln[2e^k(1+e^k)] - \frac{1}{k}\ln[e^{2k}-6e^k+1+\sqrt{e^{4k}-16e^{3k}+30e^{2k}-16e^k+1}\,] \tag{55}$$

from which we obtain

$$\widetilde{\phi}_u^*(y) = \mu_y^*(y) = \frac{1}{2}\delta(y-y') + \frac{1}{2}\delta(y-[1-y'])\,. \tag{56}$$

Expressions (51) and (56) hold as long as $\Psi_y(y)$ has two maxima, which is true while (51) satisfies

$$\begin{aligned}\mu_x^* &= \arg \min_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \Psi_x(x)\mu_x(x)\, dx \\ &= \arg \min_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \frac{1}{2}\left[\frac{e^{k(x-y')}}{(e^{k(x-y')}+1)^2}\right. \\ &\qquad \left. + \frac{e^{k(x-(1-y'))}}{(e^{k(x-(1-y'))}+1)^2}\right]\mu_x(x)\, dx\,.\end{aligned} \tag{57}$$

$\Psi_x(x)$ has local minima at $x = 0$, $x = 1$ and $x = 1/2$, so (51) and (56) hold for $k_1 < k \leq k_2$, where $k_2 \in \mathbb{R}^+$ is such that

$$\begin{aligned}&\frac{\partial}{\partial y}\left[\int_0^1 \frac{e^{k_2(x-y)}}{(e^{k_2(x-y)}+1)^2}\, \mu_x^*(x)\, dx\right]\Bigg|_{y=y'} = 0 \\ &\frac{e^{k_2(1/2-y')}}{(e^{k_2(1/2-y')}+1)^2} + \frac{e^{k_2(y'-1/2)}}{(e^{k_2(y'-1/2)}+1)^2} \\ &\qquad = \frac{e^{-k_2 y'}}{(e^{-k_2 y'}+1)^2} + \frac{e^{-k_2(1-y')}}{(e^{-k_2(1-y')}+1)^2}.\end{aligned} \tag{58}$$

The first equation gives y' in terms of $k = k_2$, and the last equation gives the minimal value of k, say k_2, for which $x = 1/2$ is a global minimum of $\Psi_x(x)$. This system of equations gives $k_2 \approx 3.6855$.

For values of k slightly higher than k_2, (56) still holds, but (51) changes to

$$\mu_x^*(x) = \alpha\delta(x) + (1-2\alpha)\delta(x-1/2) + \alpha\delta(x-1) \tag{59}$$

where $\alpha \in [0, 0.5]$. This expression satisfies (57) for every α, but we must assure that (56) still satisfies (47) when using (59), which happens for a particular choice of α. One way to find the optimal value of α is to substitute (59) into (47) and to force the derivative of the integral in (47) with respect to y equal to zero for $y = y'$. (See the proof of Theorem 12.5 of (McKinsey, 1952) for an example of how to use this idea in convex games.)

Continuing in this way, we can see that it is possible, at least in principle, to get an "explicit" solution to the robust experiment design problem. However, the expressions get rather complicated as we increase the value of k, so this is probably not a practical way of solving the problem.

In Figure 2 we can see the shapes of Ψ_x and Ψ_y for various values of k. These figures have been generated by the

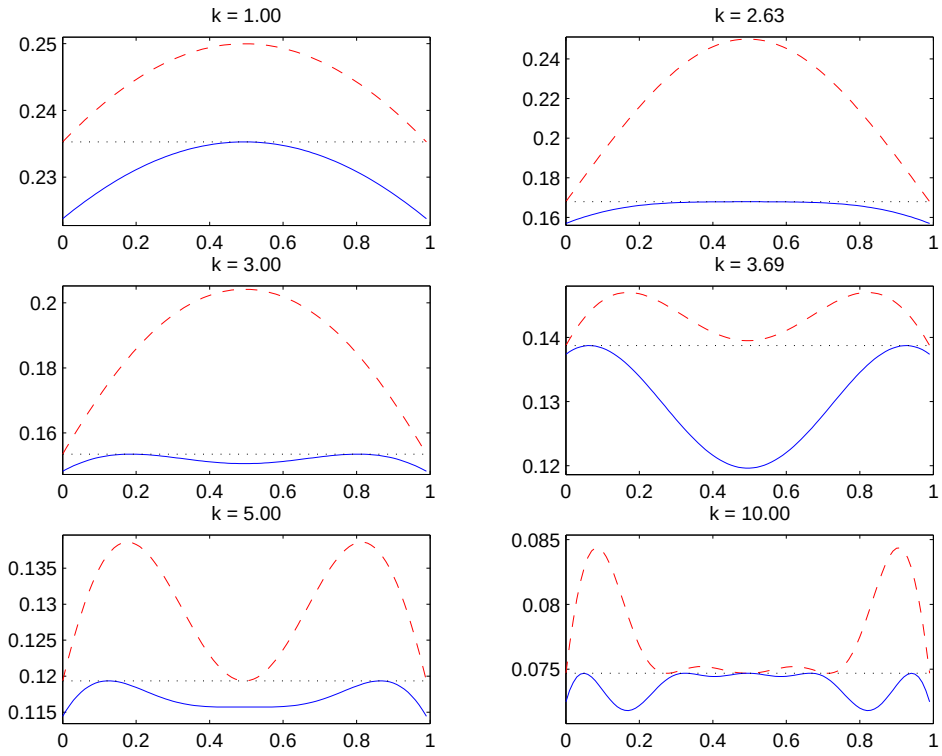

Fig. 2. Ψ_x (dashed) and Ψ_y (solid) for various values of k. The dotted line corresponds to the average value, V_m.

algorithm developed in Section 3.5. We can also see from the figures that the minimum value of Ψ_x coincides with the maximum value of Ψ_y; this is a consecuence of the Minimax Theorem, which states that both values are equal, and their common value is the so-called average value, V_m, of a game on the unit square with kernel F.

It can be further shown (Karlin, 1957) that the number of support points of both μ_x^* and μ_y^* go to infinity as $k \to \infty$, that they are lower semicontinuous functions of k and that they differ by at most 1. Namely, the number of support points of μ_x^* is always greater than or equal to that of μ_y^*.

3.4 Bandlimited '$1/f$' Noise Input

The results presented above are concerned with the optimal solution to the problem. We will also explore sub optimal solutions. In the latter context, the following result will be useful.

Lemma 7 *Let $\overline{\phi}_u \in \mathscr{S}([0,1])$. Also let*

$$\alpha_{min}(\overline{\phi}_u) \triangleq \min_{\theta\in\Theta} J(\overline{M}(\theta,\overline{\phi}_u),\theta) \tag{60}$$

$$\alpha_{max}(\overline{\phi}_u) \triangleq \max_{\theta\in\Theta} J(\overline{M}(\theta,\overline{\phi}_u),\theta)\,. \tag{61}$$

Then

$$\alpha_{min}(\overline{\phi}_u) \le \min_{\phi_u\in\mathscr{S}(\mathbb{R}_0^+)} \max_{\theta\in\Theta} J(\overline{M}(\theta,\phi_u),\theta) \le \alpha_{max}(\overline{\phi}_u)\,. \tag{62}$$

PROOF. The second inequality follows from the definition of the optimisation problem.

To establish the first inequality, we notice from (60) that

$$\begin{aligned}\frac{1}{\alpha_{min}(\overline{\phi}_u)} &= \frac{1}{\min_{\theta\in\Theta} J(\overline{M}(\theta,\overline{\phi}_u),\theta)}\\ &= \max_{\theta\in\Theta}\,[J(\overline{M}(\theta,\overline{\phi}_u),\theta)]^{-1}\\ &= \max_{\theta\in\Theta} \int_{\underline{\theta}}^{\overline{\theta}} \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\,\overline{\phi}_u(\omega)d\omega\,.\end{aligned} \tag{63}$$

Thus, for any feasible function ϕ_u, we must have

$$\frac{1}{\alpha_{min}(\overline{\phi}_u)} \ge \int_{\underline{\theta}}^{\overline{\theta}}\int_{\underline{\theta}}^{\overline{\theta}} \phi_u(\theta)\frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\,\overline{\phi}_u(\omega)d\omega\,d\theta\,. \tag{64}$$

Now let us assume that the first inequality in (62) is false; i.e. assume that

$$\begin{aligned}\alpha_{min}(\overline{\phi}_u) &> \min_{\phi_u\in\mathscr{S}(\mathbb{R}_0^+)} \max_{\theta\in\Theta} J(\overline{M}(\theta,\phi_u),\theta)\\ &= \max_{\theta\in\Theta} J(\overline{M}(\theta,\phi_u^*),\theta)\end{aligned} \tag{65}$$

and therefore,

$$\frac{1}{\alpha_{min}(\overline{\phi}_u)} < \min_{\theta\in\Theta} [J(\overline{M}(\theta,\phi_u^*),\theta)]^{-1} = \min_{\theta\in\Theta} \int_{\underline{\theta}}^{\overline{\theta}} \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\, \phi_u^*(\omega)d\omega\,. \quad (66)$$

Hence, if we form a convex combination of the integrals on the right hand side of (66) using $\overline{\phi}_u(\theta)$, we must have

$$\frac{1}{\alpha_{min}(\overline{\phi}_u)} < \int_{\underline{\theta}}^{\overline{\theta}} \overline{\phi}_u(\theta) \int_{\underline{\theta}}^{\overline{\theta}} \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\, \phi_u^*(\omega)d\omega\, d\theta\,. \quad (67)$$

However,

$$\frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2} = \frac{\theta^2/\omega^2}{(\theta^2/\omega^2+1)^2}\,. \quad (68)$$

Thus, changing the order of the variables of integration in (67) and using (68) gives

$$\frac{1}{\alpha_{min}(\overline{\phi}_u)} < \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} \phi_u^*(\theta) \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\, \overline{\phi}_u(\omega)d\theta\, d\omega\,. \quad (69)$$

We see that (69) contradicts (64) if we choose ϕ_u in (64) as ϕ_u^*. This contradiction establishes the result.

$\nabla\nabla\nabla$

Remark 8 *It is impossible to find an input ϕ_u which brings $\alpha_{min}(\phi_u)$ equal to $\alpha_{max}(\phi_u)$. This is due to the fact that, for a fixed ϕ_u, the cost function $J(\overline{M}(\theta,\phi_u),\theta)$ is an analytic function of θ on $\mathbb{R}$, and it vanishes as $|\theta|\to\infty$; thus, if we force $\alpha_{min}(\phi_u) = \alpha_{max}(\phi_u)$, then this cost function would be constant in the interval $[\underline{\theta},\overline{\theta}]$, which implies, by its analyticity, that it would be constant in $\mathbb{R}$, and hence equal to 0. This is impossible, since the integral of ϕ_u over $[\underline{\theta},\overline{\theta}]$ is equal to 1, and ϕ_u can only take nonnegative values over that interval.*

$\nabla\nabla\nabla$

Remark 9 *We see from Lemma 7 that, if a feasible design, ϕ_u, is found such that $\alpha_{min}(\phi_u)$ and $\alpha_{max}(\phi_u)$ are "close", then the corresponding cost function will be "close" to optimal. In particular, if one could choose an input, ϕ_u, such that $\alpha_{min}(\phi_u) = \alpha_{max}(\phi_u)$, then this input would have been optimal. Alas, by Remark 8, there is no feasible input which brings $\alpha_{min}(\phi_u)$ to $\alpha_{max}(\phi_u)$. However, we will now examine a particular sub-optimal input such that $\alpha_{min}(\phi_u)$ and $\alpha_{max}(\phi_u)$ are within a factor of 2 of each other.*

$\nabla\nabla\nabla$

With the above as background, we next consider the following feasible input

$$\phi_u^{1/f}(\omega) \triangleq \begin{cases} \dfrac{1/\omega}{\ln\overline{\theta} - \ln\underline{\theta}}, & \omega\in[\underline{\theta},\overline{\theta}] \\ 0, & \text{otherwise} \end{cases}. \quad (70)$$

For this input, we have the following result.

Theorem 10 *Consider the bandlimited '$1/f$' noise input given in* (70)*. Let $\alpha_{min}(\phi_u^{1/f})$ and $\alpha_{max}(\phi_u^{1/f})$ be the corresponding limits as in* (60) *and* (61)*. Then*

$$\alpha_{min}(\phi_u^{1/f}) = 2\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{\overline{\theta}+\underline{\theta}}{\overline{\theta}-\underline{\theta}} \quad (71)$$

and

$$\alpha_{max}(\phi_u^{1/f}) = 4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{\overline{\theta}^2+\underline{\theta}^2}{\overline{\theta}^2-\underline{\theta}^2}\,. \quad (72)$$

PROOF. By direct calculation we obtain

$$\begin{aligned}\frac{1}{J(\overline{M}(\theta,\phi_u^{1/f}),\theta)} &= \frac{1}{\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\int_{\underline{\theta}}^{\overline{\theta}} \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\frac{d\omega}{\omega} \\ &= \frac{\theta^2}{\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\left(-\frac{1}{2(\theta^2+\omega^2)}\right)\Bigg|_{\theta=\underline{\theta}}^{\overline{\theta}} \\ &= \frac{\overline{\theta}^2-\underline{\theta}^2}{2\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\frac{\theta^2}{(\theta^2+\overline{\theta}^2)(\theta^2+\underline{\theta}^2)}\,. \end{aligned} \quad (73)$$

The function $f(\theta) = \theta^2/[(\theta^2+\overline{\theta}^2)(\theta^2+\underline{\theta}^2)]$, in Θ, increases to a maximum at $\theta = \sqrt{\underline{\theta}\,\overline{\theta}}$ and then decreases. Finally since $f(\underline{\theta}) = f(\overline{\theta})$ we conclude that

$$\begin{aligned}\frac{1}{\alpha_{max}(\phi_u^{1/f})} &= \frac{\overline{\theta}^2-\underline{\theta}^2}{2\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\frac{\theta^2}{(\theta^2+\overline{\theta}^2)(\theta^2+\underline{\theta}^2)}\Bigg|_{\theta=\underline{\theta}} \\ &= \frac{\overline{\theta}^2-\underline{\theta}^2}{4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\frac{1}{\overline{\theta}^2+\underline{\theta}^2} \end{aligned} \quad (74)$$

and

$$\begin{aligned}\frac{1}{\alpha_{min}(\phi_u^{1/f})} &= \frac{\overline{\theta}^2-\underline{\theta}^2}{2\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\frac{\theta^2}{(\theta^2+\overline{\theta}^2)(\theta^2+\underline{\theta}^2)}\Bigg|_{\theta=\sqrt{\underline{\theta}\overline{\theta}}} \\ &= \frac{1}{2\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)}\frac{\overline{\theta}-\underline{\theta}}{\overline{\theta}+\underline{\theta}}\,. \end{aligned} \quad (75)$$

which completes the proof.

$\nabla\nabla\nabla$

Corollary 11 *For bandlimited '$1/f$' noise input, the optimal cost, J^*, must satisfy*

$$\alpha_{min}(\phi_u^{1/f}) \leq J^* \leq 2\alpha_{min}(\phi_u^{1/f})\,. \tag{76}$$

PROOF. ¿From Theorem 10,

$$\begin{aligned}
\alpha_{max}(\phi_u^{1/f}) &= 4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{\overline{\theta}^2+\underline{\theta}^2}{\overline{\theta}^2-\underline{\theta}^2} \\
&\leq 4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{\overline{\theta}^2+\underline{\theta}^2+2\overline{\theta}\underline{\theta}}{\overline{\theta}^2-\underline{\theta}^2} \\
&= 4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{(\overline{\theta}+\underline{\theta})^2}{(\overline{\theta}+\underline{\theta})(\overline{\theta}-\underline{\theta})} \\
&= 4\ln\left(\frac{\overline{\theta}}{\underline{\theta}}\right)\frac{\overline{\theta}+\underline{\theta}}{\overline{\theta}-\underline{\theta}} \\
&= 2\alpha_{min}(\phi_u^{1/f})\,.
\end{aligned} \tag{77}$$

The result then follows from Lemma 7.

$\nabla\nabla\nabla$

Remark 12 *The above result is rather surprising since it shows that '$1/f$' noise performs very well for this problem. This is an interesting result since "conventional wisdom" probably suggests an input more akin to bandlimited white noise (e.g. a PRBS signal). However, one can easily verify that using $\underline{\theta} = 0.1$ and $\overline{\theta} = 10$, bandlimited '$1/f$' noise is almost an order of magnitude superior to bandlimited white noise – see Table 1 presented below.*

$\nabla\nabla\nabla$

3.5 Discrete Approximation to the Optimal Input

As we haven seen in Section 3.3, and as it is well known in the statistics literature (see e.g. Walter and Pronzato (1997)), finding an exact solution to problems of the type (26), (27) is, in general, extremely difficult. Some algorithms have been proposed, e.g. the relaxation algorithm of Shimizu and Aiyoshi (1980).

Here we pursue an alternative idea of finding an approximate design by discretisation of the design space (see also Walter and Pronzato (1997) p341). Since f is continuous, it is well known (Owen, 1995) that this approach can approximate the optimal solution as closely as desired.

To develop this approach, we first approximate the integral in equation (27) by a Riemann sum. Specifically, utilising Lemma 2, we choose a grid of N points $\underline{\theta} \leq \omega_m = \theta_m \leq \overline{\theta}$ for $0 \leq m \leq N$ so that $\omega_0 = \theta_0 = \underline{\theta}$, $\omega_N = \theta_N = \overline{\theta}$. Then

$$\begin{aligned}
J_m &\triangleq \left[\int_{\underline{\theta}}^{\overline{\theta}} \frac{\omega^2/\theta_m^2}{(\omega^2/\theta_m^2+1)^2}\,\phi_u(\omega)d\omega\right]^{-1} \\
&\approx \left[\sum_{n=0}^{N-1} \frac{\omega_n^2/\theta_m^2}{(\omega_n^2/\theta_m^2+1)^2}\,\phi_u(\omega_n)(\omega_{n+1}-\omega_n)\right]^{-1} \\
&= \left[\sum_{n=0}^{N-1} A_{m,n}E_n\right]^{-1}
\end{aligned} \tag{78}$$

where

$$A_{m,n} \triangleq \frac{\omega_n^2/\theta_m^2}{(\omega_n^2/\theta_m^2+1)^2} > 0 \tag{79}$$

and

$$E_n \triangleq \phi_u(\omega_n)(\omega_{n+1}-\omega_n)\,. \tag{80}$$

Notice that the matrix $A = \{A_{m,n}\}$ is symmetric and has positive entries.

We can now state the following discrete approximation to the optimisation problem in equation (26)

$$\mathbf{E}^* = \arg\min_{\mathbf{E}\in\mathscr{S}_d}\max_{0\leq m<N}(\mathbf{e}_m^T A\mathbf{E})^{-1} \tag{81}$$

where $\mathscr{S}_d \triangleq \{\mathbf{E} \in \mathbb{R}^N : \mathbf{1}^T\mathbf{E} = 1, E_n \geq 0\}$, $\mathbf{E} \triangleq [\,E_0\;E_1\;\cdots\;E_{N-1}\,]^T$, $\mathbf{e}_m$ is the m^{th} column of the N dimensional identity matrix, and $\mathbf{1}$ is an N dimensional vector of ones.

It is known that a finite dimensional min-max optimisation problem, such as (81), can be converted into a standard linear programming (LP) problem; see (McKinsey, 1952; Dantzig, 1951; Gale et al., 1951). For completeness, we present the details below. Let

$$\mathbf{F} \triangleq \begin{bmatrix} x \\ \mathbf{E} \end{bmatrix} \in \mathbb{R}^{N+1} \tag{82}$$

then we can readily show that (81) is equivalent to the following optimisation problem:

$$\max_{\mathbf{F}} \widetilde{C}\mathbf{F} \tag{83}$$

subject to

$$\begin{aligned}
\widetilde{A}\mathbf{F} &\geq 0 \\
\widetilde{B}\mathbf{F} &= 1
\end{aligned} \tag{84}$$

where

$$\widetilde{A} \triangleq \begin{bmatrix} -\mathbf{1} & A \\ \mathbf{0} & I \end{bmatrix} \in \mathbb{R}^{2N\times(N+1)}$$
$$\widetilde{B} \triangleq [\,0\ \mathbf{1}^T\,] \in \mathbb{R}^{1\times(N+1)} \qquad (85)$$
$$\widetilde{C} \triangleq [\,1\ 0\ \cdots\ 0\,] \in \mathbb{R}^{1\times(N+1)}$$

and $\mathbf{1} \in \mathbb{R}^{N\times 1}$.

This problem is readily recognised as an LP problem.

It is also quite straightforward to compute a discrete approximation to the Bayesian optimal input for the example problem. For example, say that we use $J(\overline{M}(\theta,\phi_u),\theta)$ as a Bayesian risk and, for the sake of illustration assume that θ has a uniform distribution on Θ. Then, the Bayesian design problem becomes

$$\phi_u^B = \arg \min_{\phi_u\in\mathscr{S}(\mathbb{R}_0^+)} \frac{1}{\overline{\theta}-\underline{\theta}} \cdot \int_{\underline{\theta}}^{\overline{\theta}} \left[\int_0^\infty \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\,\phi_u(\omega)d\omega\right]^{-1} d\theta\,. \qquad (86)$$

We can approximate this, as in (78), by

$$\mathbf{E}^B = \arg \min_{\mathbf{E}\in\mathscr{I}_d} \frac{1}{N}\sum_{k=0}^{N-1} \exp\left\{\left[\frac{\log\overline{\theta}-\log\underline{\theta}}{N}\right]k+\log\underline{\theta}\right\} \cdot (\mathbf{e}_k^T A\mathbf{E})^{-1}\,. \qquad (87)$$

In the next section, we will also consider a Bayesian design for the case when $\log\theta$ has a uniform distribution on $(\log\underline{\theta}\ ,\ \log\overline{\theta})$.

3.6 Numerical Results

We present below numerical results for the problem described above where we take $\underline{\theta}=0.1$, $\overline{\theta}=10$, $N=100$ and compare

(i) A nominal input of frequency 1 [rad/s] (Notice that this is the optimal input if the initial estimate of the parameter is $\widehat{\theta}=1$).
(ii) Band limited white noise input, limited to the frequency range $[0.1, 10]$ [rad/s].
(iii) Band limited '$1/f$' noise input, limited to the frequency range $[0.1, 10]$ [rad/s].
(iv) The approximate discretised robust optimal input generated by Linear Programming.
(v) The approximate discretised Bayesian optimal input for a uniform distribution on θ.

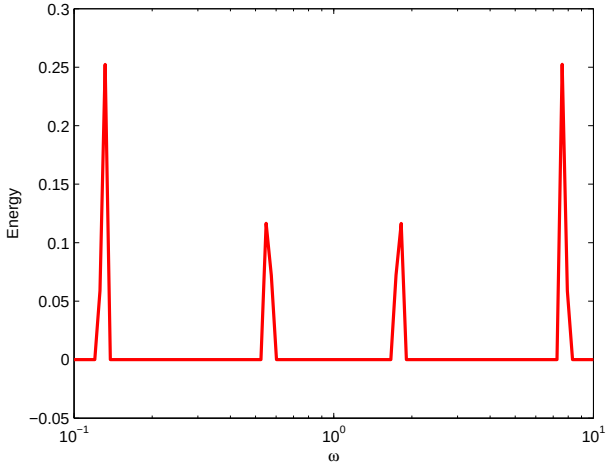

Fig. 3. Values of E for discretised robust optimal input.

(vi) The approximate discretised Bayesian optimal input for a uniform distribution on $\log\theta$.

Relative costs for the different experimental conditions are shown in Table 1. Notice that the costs have been scaled so that the optimal value is 1.

We see from Table 1 that bandlimited white noise gives poor performance under all criteria. Indeed, we see from the table that bandlimited '$1/f$' noise is almost an order of magnitude better than a bandlimited white noise input for all cost functions. Furthermore, going to the discretised min-max optimum gives a further 40% improvement for the min-max cost function. The discretised min-max optimal input energy, ϕ_u^*, is shown in Figure 3. Notice that the above results are consistent with Theorem 5, which asserts that the unique optimal input has finite support. Figure 4 shows the corresponding values of $[\theta^2\overline{M}(\theta,\phi_u^*)]^{-1}$ as a function of θ. It is interesting to notice from Figure 4 that $[\theta^2\overline{M}(\theta,\phi_u^*)]^{-1}$ is an almost constant function of θ. This should be compared with the comments in Remark 9. The comparative costs are given in Table 1.

4 Generalisation to Multi-parameter Problems

For the multi-parameter case we return to the general expression for $\overline{M}(\beta,\phi)$ given in (11) and (13). Again for simplicity, we assume white noise and hence refer only to θ although the extension to coloured noise offers no additional difficulties. We convert this problem into an approximate discrete form as was done in Section 3 by discretising the input and parameter spaces. We write

$$Q_k(\mathbf{E}) \triangleq \sum_m A_{km}E_m, \qquad \theta_k\in\Theta \qquad (88)$$

as an approximation to the integral in (13) i.e. Q_k is the information *matrix* corresponding to the k^{th} (discretised)

Table 1
Relative Values of Cost for the Different Input Signals

	$\max_{\theta\in\Theta}\,[\theta^2\overline{M}(\theta,\phi_u)]^{-1}$	Bayesian cost on θ	Bayesian cost on $\log\theta$
Single frequency at $\omega = 1$	7.75	4.8	2.26
Bandlimited white noise	12.09	9.05	2.96
Bandlimited '$1/f$' noise	1.43	1.51	1.07
Robust min-max optimal input	1.00	1.45	1.12
Bayesian design (for uniform distribution on θ)	5.4	1.00	1.61
Bayesian design (for uniform distribution on $\log\theta$)	1.53	1.46	1.00

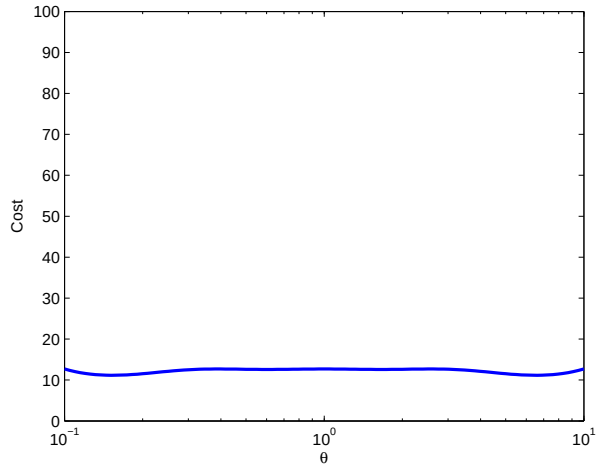

Fig. 4. Variation of cost versus θ for robust optimal input.

element θ_k of the parameter set Θ, the index m denotes the frequency and E_m is the input energy at the m^{th} frequency.

There exist many possible choices for the inner design criterion $J(\overline{M}(\theta,\phi_u),\theta)$ in the multi-parameter case – see the discussion in Section 2.2 and recent results in (Welsh et al., 2006). Three other alternatives are discussed below.

4.1 Minimal Eigenvalue

The use of the minimum eigenvalue of the information matrix as a design criterion for *nominal* experiment design has previously been studied (Mareels et al., 1987). For the *robust* case, we propose to optimise the worst case of the following related criterion which uses the minimum eigenvalue of a scaled version of the information matrix,

$$J_1(\overline{M}(\theta,\phi_u),\theta) \triangleq (\lambda_{\min}\{S_\theta\overline{M}(\theta,\phi_u)S_\theta\})^{-1} \qquad (89)$$

where $\lambda_{\min}$ denotes the minimum eigenvalue and S_θ is a parameter dependent scaling matrix. One possible choice for S_θ is $\mathrm{diag}[\theta_1,\ldots,\theta_m]$. The motivation for this choice is that $\overline{M}(\theta,\phi_u)^{-1}$ is a measure of the parameter covariance matrix. Hence $S_\theta^{-1}\overline{M}(\theta,\phi_u)^{-1}S_\theta^{-1}$ is the covariance normalised by the nominal values of each parameter. Therefore it is a measure of the *relative error*. This seems to be an important property *in the robust design context* (where we maximise over $\theta\in\Theta$) since it ensures that one is maximising (over Θ) the relative errors. These errors are normalised and thus better scaled for comparison purposes.

Another useful property of $J_1(\overline{M}(\theta,\phi_u),\theta)$ is that, due to the normalisation by S_θ, the scaled information matrix does not depend on the system gain. This simplifies the problem of discretisation of the set Θ by eliminating one degree of freedom (the gain). This also makes sense, heuristically speaking, since the system gain simply scales the output.

4.2 Relative Frequency Domain Errors

This criterion is motivated by robust control (Goodwin et al., 2001). It is well known (Zhou et al., 1996), that the achieved sensitivity, S, is related to the nominal sensitivity, S_0, via

$$S = \frac{S_0}{1 + T_0\,\Delta G/G} \qquad (90)$$

where T_0 is the nominal complementary sensitivity and $\Delta G/G$ is the relative error in G. Indeed, this leads to the well known sufficient condition for robust stability, namely $\|T_o\,\Delta G/G\|_\infty < 1$.

Say we put an upper bound on $\|T_0\|_\infty$, then we see that what is important is the infinity norm of the relative error in G, $\Delta G/G$. Then, noting that the covariance of all unbiased estimates of θ are lower bounded by $(N\overline{M})^{-1}$ where N is the number of data points (see (Goodwin and Payne, 1977)), we can obtain a measure of the size of $\Delta G/G$ as

$$\left\|\frac{N\mathbb{E}|\Delta G|^2}{|G|^2}\right\|_\infty = \max_\omega\left\{\frac{\frac{\partial G(j\omega)^T}{\partial\theta}\,\overline{M}^{-1}\,\frac{\partial G(-j\omega)}{\partial\theta}}{|G(j\omega)|^2}\right\} \triangleq J_2(\overline{M}(\theta,\phi_u),\theta)\,. \qquad (91)$$

Notice that here we use the per-sample information matrix $\overline{M}$.

It is readily seen that $J_2(\overline{M}(\theta,\phi_u),\theta)$ is a dimensionless quantity. Thus the associated experiment design is independent of the system gain in the same way that this was true for J_1 (see (89)).

Remark 13 *We see that the criterion J_2 has the form*

$$J_2(\overline{M}(\theta,\phi_u),\theta) = \max_{\omega} g(\omega)^* \overline{M}^{-1} g(\omega) \qquad (92)$$

where

$$g(\omega) \triangleq \frac{\partial G(-j\omega)/\partial\theta}{|G(j\omega)|}\,. \qquad (93)$$

Thus we see that (91) *maximises $x^*\overline{M}^{-1}x$ where x is restricted to the particular set of vectors given in* (93)*. This can be compared with $J_1(\overline{M}(\theta,\phi_u),\theta)$ which is actually equivalent to maximising $y^*\overline{M}^{-1}y$ over the set of vectors y where $y = S_\theta Z$ and $Z^*Z = 1$.*

$\nabla\nabla\nabla$

4.3 A Criterion Related to the ν Gap

Hildebrand and Gevers (2003b,a) have suggested the following criterion for nominal experiment design such that the worst case ν gap is minimised,

$$\begin{aligned} &J_3(\overline{M}(\theta,\phi_u),\theta) \\ &\quad\triangleq \max_{\omega} \left.\frac{\lambda_{\max}\left\{\begin{bmatrix}\mathrm{Re}\,\frac{\partial G}{\partial\theta}\\ \mathrm{Im}\,\frac{\partial G}{\partial\theta}\end{bmatrix}\overline{M}^{-1}\begin{bmatrix}\mathrm{Re}\,\frac{\partial G}{\partial\theta} & \mathrm{Im}\,\frac{\partial G}{\partial\theta}\end{bmatrix}\right\}}{[1+|G|^2]^2}\right|_{\omega} \\ &\quad= \max_{\omega} \frac{\lambda_{\max}\begin{bmatrix} R_\omega^T\overline{M}^{-1}R_\omega & R_\omega^T\overline{M}^{-1}I_\omega \\ I_\omega^T\overline{M}^{-1}R_\omega & I_\omega^T\overline{M}^{-1}I_\omega\end{bmatrix}}{[1+|G_\omega|^2]^2} \end{aligned} \qquad (94)$$

where the subscript ω denotes 'frequency ω' and

$$\begin{aligned} R_\omega &\triangleq \mathrm{Re}\left\{\frac{\partial G(\omega)}{\partial\theta}\right\} \\ I_\omega &\triangleq \mathrm{Im}\left\{\frac{\partial G(\omega)}{\partial\theta}\right\}. \end{aligned} \qquad (95)$$

Remark 14 *Not surprisingly, there is a connection between $J_2(\overline{M}(\theta,\phi_u),\theta)$ and $J_3(\overline{M}(\theta,\phi_u),\theta)$ since both are motivated by robust control. Specifically, it is readily seen that*

$$\begin{aligned} \frac{\partial G}{\partial\theta}^T\overline{M}^{-1}\frac{\partial\bar{G}}{\partial\theta} &= R_\omega^T\overline{M}^{-1}R_\omega + I_\omega^T\overline{M}^{-1}I_\omega \\ &= \mathrm{tr}\,L_\omega \end{aligned} \qquad (96)$$

where L_ω appears in (94)*, i.e.*

$$L_\omega \triangleq \begin{bmatrix} R_\omega^T\overline{M}^{-1}R_\omega & R_\omega^T\overline{M}^{-1}I_\omega \\ I_\omega^T\overline{M}^{-1}R_\omega & I_\omega^T\overline{M}^{-1}I_\omega\end{bmatrix}. \qquad (97)$$

We notice that

$$\lambda_{\max}(L_\omega) \le \mathrm{tr}\,L_\omega \le 2\lambda_{\max}(L_\omega)\,. \qquad (98)$$

Hence, we see that the criteria $J_2(\overline{M}(\theta,\phi_u),\theta)$ and $J_3(\overline{M}(\theta,\phi_u),\theta)$ are loosely connected. Moreover, Remark 13 links both criteria to $J_1(\overline{M}(\theta,\phi_u),\theta)$.

A potential issue with the criterion $J_3(\overline{M}(\theta,\phi_u),\theta)$ is that, unlike $J_1(\overline{M}(\theta,\phi_u),\theta)$ and $J_2(\overline{M}(\theta,\phi_u),\theta)$, it is not dimensionless. This is not an issue in the case of nominal experiment design. However, it could be a problem with respect to robust design when one wishes to compare the criteria for different values of $\theta \in \Theta$. A possible normalisation for $J_3(\overline{M}(\theta,\phi_u),\theta)$ is given in (Welsh et al., 2006).

$\nabla\nabla\nabla$

Remark 15 *Notice that the above criteria are convex in terms of ϕ_u. This follows since the supremum of a family of convex functions is itself convex.*

$\nabla\nabla\nabla$

5 Numerical Example

To illustrate the merits of the robust optimal experiment design procedure on a realistic example, we have evaluated a discretised approximation to each of the criteria $J_1(\overline{M},\theta)$, $J_2(\overline{M},\theta)$ and $J_3(\overline{M},\theta)$ on a multi-parameter design example. The system is

$$G_1(s) = \frac{K}{s^2 + a_1 s + a_0} \qquad (99)$$

and, as before, we take $G_2(s) = 1$.

We assume prior knowledge of the parameters as follows:

$$\begin{aligned} \theta_1 &\triangleq a_1 \in [1,2] \\ \theta_2 &\triangleq a_0 \in [1,9] \\ \theta_3 &\triangleq K \in [1,2]\,. \end{aligned} \qquad (100)$$

Notice that the above parameter ranges correspond to the pole locations shown in Figure 5.

The parameter and frequency ranges were all divided into logarithmically spaced grids for the optimisation. For our example we chose each range to contain 20 values. We also used the Matlab® optimisation toolbox to carry out the minmax designs.

In all our simulations we approximated the integral in (13) by the following discretisation

$$\begin{aligned}\overline{M}(\theta,\phi_u) &= \int_0^\infty \mathrm{Re}\{Q(\omega)\}\ \phi_u(\omega)d\omega \\ &= \sum_n \int_{\Delta n} \mathrm{Re}\{Q(\omega)\}\ \phi_u(\omega)d\omega \qquad (101)\\ &\approx \sum_{n=1}^{20} \mathrm{Re}\{Q(\omega_n)\} \int_{\Delta n} \phi_u(\omega)d\omega \\ &= \sum_{n=1}^{20} \mathrm{Re}\{Q(\omega_n)\}E_n\end{aligned}$$

where

$$Q(\omega) \triangleq \frac{\partial G_1(j\omega)}{\partial\theta}|G_2(j\omega)|^{-2}\left[\frac{\partial G_1(j\omega)}{\partial\theta}\right]^H \qquad (102)$$

and

$$E_n \triangleq \int_{\Delta_n} \phi_u(\omega)d\omega \qquad (103)$$

is the input energy in the frequency range Δ_n. We have chosen

$$\Delta_n \triangleq \omega_{n+1} - \omega_n \qquad (104)$$

where $\omega_n = 0.3(10)^{(n-1)/20}$.

The discrete approximation to the robust optimal input was found for each of the criteria $\max_\theta J_1$, $\max_\theta J_2$ and $\max_\theta J_3$. For those criteria depending on a maximum over a frequency range (i.e. $\max_\theta J_2$ and $\max_\theta J_3$), we limited ω to $[0.3, 3]$ [rad/s]. (This choice was motivated by the region of possible pole locations.)

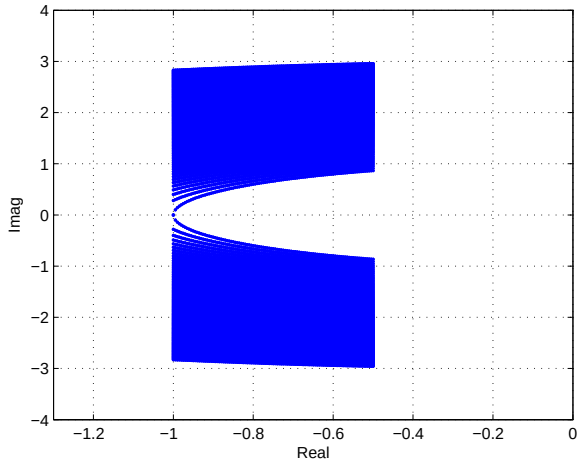

Fig. 5. Region of pole locations.

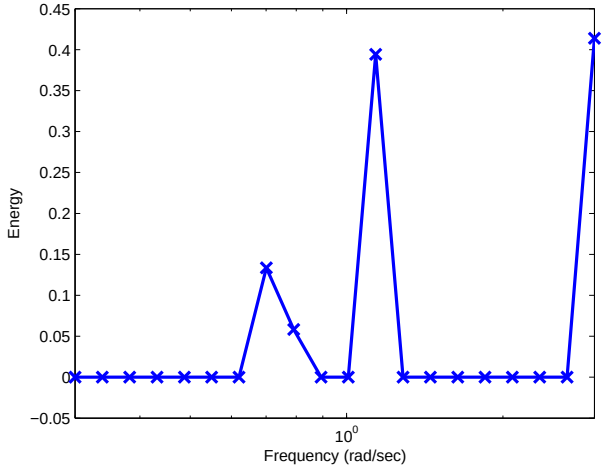

Fig. 6. Values of E for the discretised robust optimal input obtained from criterion $\max_\theta J_1$.

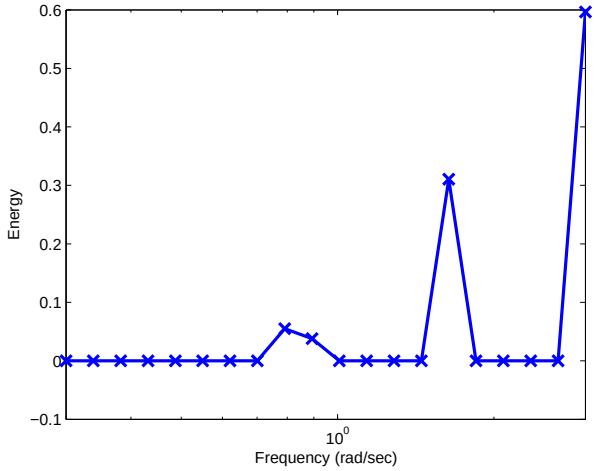

Fig. 7. Values of E for the discretised robust optimal input obtained from criterion $\max_\theta J_2$.

Sample results are shown in Figures 6 to 8. Figure 6 shows the discretised optimal input energy distribution for criterion $\max_\theta J_1$. Notice again that the input has finite support.

Figure 7 shows the discretised optimal input energy for $\max_\theta J_2$. We see from Figures 6 and 7 that the optimal input is roughly the same whether we use $\max_\theta J_1$ or $\max_\theta J_2$.

Figure 8 shows the discretised optimal input energy distribution for $\max_\theta J_3$.

Finally, Table 2 compares the evaluated cost functions obtained with different design criteria for different inputs. The table also shows the values of the corresponding cost functions for bandlimited '$1/f$' noise and bandlimited white noise.

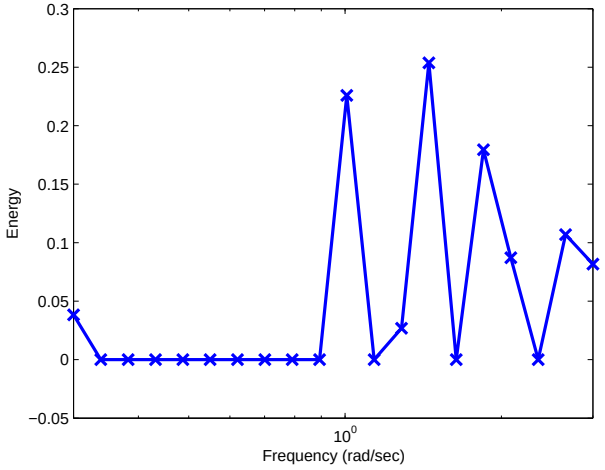

Fig. 8. Values of E for the discretised robust optimal input obtained from criterion $\max_\theta J_3$.

Table 2
Values of Cost for the Criteria

	$\max_\theta J_1$	$\max_\theta J_2$	$\max_\theta J_3$
Optimal input for $\max_\theta J_1$	1	1.73	1.55
Optimal input for $\max_\theta J_2$	1.14	1	1.32
Optimal input for $\max_\theta J_3$	1.78	2.43	1
Bandlimited '$1/f$' noise	2.08	3.13	1.22
Bandlimited white noise	5	8.8	1.93

Our choice of bandlimited '$1/f$' noise is motivated by earlier results in Section 3 where we showed that bandlimited '$1/f$' noise was near optimal for robust experiment design for the illustrative problem.

The results in Table 2 have been normalised so that the optimal design gives a value of 1. Observations from this table are:

(1) Bandlimited '$1/f$' noise gives much better results in the multi-parameter case for all criteria than does bandlimited white noise. We believe this to be a surprising and interesting observation!
(2) For $\max_\theta J_1$, the discretised robust optimal input is approximately twice as good as bandlimited '$1/f$' noise and about 5 times as good as bandlimited white noise.
(3) For $\max_\theta J_2$, the discretised optimal input is about 1.7 times better than the optimal input for $\max_\theta J_1$. The robust optimal input for $\max_\theta J_2$ is about 3 times as good as bandlimited '$1/f$' noise and almost 9 times as good as bandlimited white noise.
(4) For $\max_\theta J_3$, the criterion seems to be less sensitive to the test signal.
(5) We also notice that the discretised optimal inputs for $\max_\theta J_1$ and $\max_\theta J_2$ are quite similar whilst the discretised optimal result for $\max_\theta J_3$ is considerably different.

To further motivate our robust design approach, we also tried nominal experiment design for this example. Here we assumed nominal parameter values in the centre of the a priori region, i.e. we chose $\hat{\theta}_1 = 1.5$, $\hat{\theta}_2 = 5$ and $\hat{\theta}_3 = 1.5$. We then found the corresponding exact nominal optimal input using J_1 as our design criterion. For this input, in the case where the true parameters take any value in the a priori region, we found that the range of the cost is 30 to 2700. This lies in stark contrast to the range of cost for the discretised robust optimal input which turns out to be 26 to 400. Thus we see that the discretised robust optimal input gives almost 700% improvement in the worst case performance relative to the nominal optimal input. However, this is achieved with a negligible change (10%) in the best case performance. This provides a strong incentive to move to a robust design criterion since we see from Figure 5 that the region of possible pole locations is quite realistic for this example.

6 Conclusion

This paper has described and analysed a min-max approach to robust optimal experiment design for dynamic system identification. The paper has also evaluated and compared several different design criteria. Two illustrative examples have been presented, one with a scalar parameter and the other with multiple parameters, showing that substantial improvements in the worst case performance are achieved using a discretised robust design procedure relative to what is achieved via nominal experiment design procedure. The paper also opens up many potential research issues. For example, it is surprising that bandlimited '$1/f$' noise performs so well. This suggests that we should not use (near) white inputs such as PRBS. Instead, it may be valuable to investigate binary inputs whose energy distribution approximates bandlimited '$1/f$' noise. Also, the results in Section 5 suggest that the problem of linking robust control and experiment design may still offer many interesting research challenges.

Acknowledgements

The authors acknowledge helpful comments made by Juan Carlos Agüero, Milan Derpich, Michel Gevers, Arkadi Nemirovski and Brett Ninness on earlier drafts of this paper. Also, valuable suggestions were made by two anonymous reviewers. These are gratefully acknowledged.

References

Arimoto, S., Kimura, H., 1973. Optimal input test signals for system identification - an information theoretic approach. Int. J. Sytems Sci. 1 (3), 279–290.

Atkinson, A., Chaloner, K., Juritz, J., Herzberg, A., 1993. Optimum experiment design for properties of a compartmental model. Biometrica 29, 325–337.

Atkinson, A., Doner, A., 1992. Optimum Experiment Design. Clarendon Press, Oxford.

Başar, T., Bernhard, P., 1995. H^∞ - Optimal Control and Related Minmax Design Problems: A Dynamic Game Approach. Birkhauser, Boston-Basel-Berlin.

Başar, T., Olsder, G., 1995. Dynamic Noncooperative Game Theory., 2nd Edition. Academic Press, London / New York.

Bartle, R. G., 1966. The Elements of Integration. John Wiley & Sons, New York.

Biedermann, S., Dette, H., 2003. A note on maximin and bayesian d-optimal designs in weighted polynomial regression. Mathematical Methods of Statistics 12 (3), 358–370.

Bochner, S., 1959. Lectures on Fourier Integrals. Princeton University Press, Princeton, New Jersey.

Bohnenblust, H. F., Karlin, S., Shapley, L. S., 1950. Games with a continuous, convex pay-off. In: Contributions to the Theory of Games, vol. I. No. 24 in Annals of Mathematics Studies. Princeton University Press, Cambridge, pp. 181–192.

Bombois, X., Scorletti, G., Gevers, M., Van den Hof, P., Hildebrand, R., 2005. Least costly identification experiment for control. Tech. rep., CESAME, Louvian-la-Neuve.

Chaloner, K., Larntz, K., 1989. Optimal bayesian designs applied to logistic regression experiments. J. Statistical Planning and Inference 21, 191–208.

Chaloner, K., Verdinelli, I., 1995. Bayesian experiment design: A review. Statistical Science.

Chernoff, H., 1975. Approaches in sequential design of experiments. In: Srivastava, J. (Ed.), Survey of Statistical Design. North–Holland, Amsterdam, pp. 67–90.

Cox, D., 1958. Planning of Experiments. Wiley, New York.

Dantzig, G. B., 1951. A proof of the equivalence of the programming problem and the game problem. In: Koopmans, T. C. (Ed.), Activity Analysis of Production and Allocation: Proceedings of a Conference. Cowles Commission for Research in Economics. John Wiley & Sons, New York, pp. 330–335.

D'Argenio, D., Van Guilder, M., 1988. Design of experiments for parameter estimation involving uncertain systems, with application to pharmacokinetics. 12th IMACS World Congress on Scientific Computation 2, 511–513.

Dette, H., Melas, V., Pepelyshev, A., 2003. Standardized maximin e-optimal designs for the michaelis-menten model. Statistica Sinica 13 (4), 1147–1163.

El-Gamal, M., Palfrey, T., 1996. Economical experiments: Bayesian efficient experimental design. Int. J. Game Theory 25 (4), 495–517.

Fedorov, V., 1972. Theory of Optimal Experiments. Academic Press, New York and London.

Fedorov, V., 1980. Convex design theory. Math. Operationsforsch. Statist. Ser. Statistics 11 (3), 403–413.

Ford, I., Silvey, S., 1980. A sequentially constructed design for estimating a nonlinear parametric function. Biometrika 67 (2), 381–388.

Ford, I., Titterington, D., Wu, C.-F., 1985. Inference and sequential design. Biometrika 72 (3), 545–551.

Fudenberg, D., Tirole, J., 1991. Game Theory. The MIT Press, Cambridge, Massachusetts.

Gagliardi, R., 1967. Input selection for parameter idenfication in discrete systems. IEEE Transactions on Automatic Control AC-12 (5), 133–147.

Gale, D., Kuhn, H. W., Tucker, A. W., 1951. Linear programming and the theory of games. In: Koopmans, T. C. (Ed.), Activity Analysis of Production and Allocation: Proceedings of a Conference. Cowles Commission for Research in Economics. John Wiley & Sons, New York, pp. 317–329.

Gevers, M., 2005. Identification for control: From the early achievements to the revival of experiment design. European Journal of Control 11, 1–18.

Gevers, M., Bombois, X., 2006. Input design: From open-loop to control-oriented design. In: 14th IFAC Symposium on System Identification (SYSID'06). Newcastle (Australia), pp. 1329–1334.

Glicksberg, I. L., 1950. Minimax theorem for upper and lower semicontinuous payoffs. Rand Corporation Research Memorandum RM-478.

Goodwin, G., Graebe, S., Salgado, M., 2001. Control System Design. Prentice Hall, Upper Saddle River, New Jersey.

Goodwin, G., Murdoch, J., Payne, R., 1973a. Optimal test signal design for linear single input - single output system identification. Int. J. Control 17 (1), 45–55.

Goodwin, G., Payne, R., 1973. Design and characterisation of optimal test signals for linear single input - single output parameter estimation. 3rd IFAC Symposium, The Hague/Delft.

Goodwin, G., Payne, R., 1977. Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, New York.

Goodwin, G., Payne, R., Murdoch, J., 1973b. Optimal test signal design for linear single input - single output closed loop identification. CACSD Conference, Cambridge.

Goodwin, G., Welsh, J., Feuer, A., Derpich, M., 2006. Utilizing prior knowledge in robust optimal experiment design. Proc. of the 14th IFAC SYSID, Newcastle, Australia, 1358–1363.

Hildebrand, R., Gevers, M., 2003a. Identification for control: Optimal input design with respect to a worst-case ν-gap cost function. SIAM J. Control Optim. 41 (5), 1586–1608.

Hildebrand, R., Gevers, M., 2003b. Minimizing the worst-case ν-gap by optimal input design. Proc. of 13th IFAC Symposium on System Identification, 665–670.

Hjalmarsson, H., March 2005. From experiment design to closed-loop control. Automatica 41 (3), 393–438.

Karlin, S., 1957. On games described by bell shaped kernels. In: Contributions to the Theory of Games, vol. III. No. 39 in Annals of Mathematics Studies. Princeton University Press, New Jersey, pp. 365–391.

Karlin, S., 1959. Mathematical Methods and Theory in Games, Programming, and Economics. Addison-Wesley, Reading, Massachussets.

Karlin, S., Studden, W., 1966. Optimal experimental designs. Annals. Math. Stat. 37, 783–815.

Kempthorne, O., 1952. Design and Analysis of Experiments. Wiley, New York.

Kiefer, J., 1974. General equivalence theory for optimum designs (aproximate theory). The Annals of Statistics 2 (5), 849–879.

Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. Can. J. Math. 12, 363–366.

Landaw, E., 1984. Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. In: Eisenfeld, J., DeLisi, C. (Eds.), Mathematics and Computers in Biomedical Applications. Elsevier North-Holland, Amsterdam, pp. 181–187.

Levadi, V., 1966. Design of input signals for parameter estimation. IEEE Transactions on Automatic Control AC-11 (2), 205–211.

Mårtensson, J., Hjalmarsson, H., 2006. Robust input design using sum of squares constraints. In: 14th IFAC Symposium on System Identification (SYSID'06). Newcastle (Australia), pp. 1352–1357.

Mareels, I., Bitmead, R., Gevers, M., Jr, C. J., Kosut, R., Poubelle, M., 1987. How exciting can a signal really be? Systems & Control Letters 8, 197–204.

McKinsey, J. C. C., 1952. Introduction to the Theory of Games. McGraw-Hill, New York.

Mehra, R., 1974. Optimal inputs for system identification. IEEE Transactions on Automatic Control AC-19, 192–200.

Melas, V., 1978. Optimal designs for exponential regression. Mathematische Operationsforschung und Statistik Series Statistics 9 (1), 45–59.

Müller, W., Pötscher, B., 1992. Batch sequential design for a nonlinear estimation problem. In: Fedorov, V., Müller, W., Vuchkov, I. (Eds.), Model-Orientated Data Analysis, a Survey of Recent Methods. Physica-Verlag, Heidelberg, pp. 77–87.

Owen, G., 1995. Game Theory, 3rd Edition. Academic Press, San Diego, California.

Pronzato, L., Walter, E., 1988. Robust experiment design via maxmin optimisation. Mathematical Biosciences 89, 161–176.

Rudin, W., 1973. Functional Analysis. McGraw-Hill, New York.

Rudin, W., 1987. Real and Complex Analysis, 3rd Edition. McGraw-Hill, Singapore.

Sebastiani, P., Wynn, H., 2000. Maximum entropy sampling and optimal bayesian experimental designs. J. Royal Statistical Society: Series B 62 (1), 145–157.

Shimizu, K., Aiyoshi, E., 1980. Necessary conditions for min-max problems algorithms by a relaxation procedure. IEEE Transactions on Automatic Control 25 (1), 62–66.

Szép, J., Forgó, F., 1985. Introduction to the Theory of Games. D. Reidel Publishing Company, Budapest, Hungary.

Wald, A., 1943. On the efficient design of statistical investigations. Annals. Math. Stat. 14, 134–140.

Walter, E., Pronzato, L., 1997. Identification of Parametric Models from Experimental Data. Springer-Verlag, Berlin, Heidelberg, New York.

Welsh, J., Goodwin, G., Feuer, A., 2006. Evaluation and comparison of robust optimal experiment design criteria. Accepted to the American Control Conference, Minneapolis, USA (http://www.eecs.newcastle.edu.au/reports/EE05044.pdf).

Whittle, P., 1973. Some general points in the theory of optimal experimental design. J.R. Stat. Soc. (1), 123–130.

Wu, C.-F., 1985. Asymptotic inference from sequential design in a nonlinear situation. Biometrika 72 (3), 553–558.

Wynn, H., 1972. Results in the theory and construction of d-optimum experimental designs. J.R. Stat. Soc. (2), 133–147.

Zarrop, M., 1979. Optimal Experiment Design for Dynamic System Identification. Vol. 21 of Lecture Notes in Control and Information. Springer, Berlin, New York.

Zhou, K., Doyle, J., Glover, K., 1996. Robust and Optimal Control. Prentice Hall.