Optimal Experiment Design with Diffuse Prior Information

Cristian R. Rojas Graham C. Goodwin James S. Welsh Arie Feuer

Abstract— **In system identification one always aims to learn as much as possible about a system from a given observation period. This has led to on-going interest in the problem of optimal experiment design. Not surprisingly, the more one knows about a system the more focused the experiment can be. Indeed, many procedures for 'optimal' experiment design depend, paradoxically, on exact knowledge of the system parameters. This has motivated recent research on, so called, 'robust' experiment design where one assumes only partial prior knowledge of the system. Here we go further and study the question of optimal experiment design when the a-priori information about the system is diffuse. We show that band-limited '$1/f$' noise is optimal for a particular choice of cost function.**

I. Introduction

In system identification, there is always a strong incentive to learn as much about a system as possible from a given observation period. This has motivated substantial interest in the topic of optimal experiment design. Indeed, there exists a body of work on this topic, both in the statistics literature [5, 14, 7] and in the engineering literature [17, 10, 27].

Much of the existing literature is based on designing the experiment to optimize some scalar function of the Fisher Information Matrix [10, pg. 6]. However, a fundamental difficulty is that when the system response depends non-linearly on the parameters, the Information Matrix depends, inter-alia, on the true system parameters. Moreover, we note that models for dynamical systems (even if linear) typically have the characteristic that their response depends non-linearly on the parameters. Hence, the information matrix for models of dynamical systems generally depends upon the true system parameters. This means that experiment designs which are 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, pg. 427].

The above reasoning has motivated the study of, so called, 'robust' optimal experiment designs with respect to uncertainty on *a priori* information. In this vein, various approaches have been proposed, e.g.

(i) Iterative design where one alternates between parameter estimation and experiment design based on the current estimates [4, 18, 25].

(ii) Bayesian design where one optimizes some function of the expected information matrix, with the expectation taken over some a-priori distribution of the parameters [1, 3, 6].

(iii) Min-Max design in which one optimizes the worst case over a bounded set of a-priori given parameter values [20, 8, 21].

The latter designs mentioned above are closely related to game theory. Indeed, game-theoretical ideas have been used to characterize the optimal robust (in the min-max sense) experiment. For example, several papers have studied different types of one-parameter robust experiment design problems [21, 11]. It has been shown for these problems that the optimal min-max experiment has many interesting properties, e.g. it exists, it is unique, it has compact support in the frequency domain and it is characterized by a line spectrum. For multi-parameter problems, one usually needs to use gridding strategies to carry out the robust designs numerically [21, 25].

A surprising observation from recent work on min-max optimal experiment design is that band-limited '$1/f$' noise is actually quite close to optimal for particular problems. Indeed, '$1/f$' noise has been shown to have a performance which is within a factor of 2 from the performance of robust optimal designs for first-order and resonant systems [21, 11]. It is important to note, however, that the proof of near optimality depends on a particular property of these systems which allows one to scale the parameters with respect to frequency.

Here we ask a more general question: 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 to use to estimate the system?

In this case we consider as diffuse prior information that the interesting part of the frequency response of the system lies in an interval $[a, b]$. This implies that we are seeking an experiment which is 'good' over a very broad class of possible systems. In this paper, we propose a possible solution to this problem, being that the experiment should consist of bandlimited '$1/f$' noise.

The paper is structured as follows. In Section II we discuss the problem of measuring the 'goodness' of an experiment by using a system independent criterion. Section III gives some desirable properties that such a measure would be expected to possess. In Section IV we consider a typical input constraint generally used in experiment design. Section V shows a preliminary result for choosing a suitable cost function which satisfies the properties developed in Section

C. R. Rojas, G. C. Goodwin and J. S. Welsh are with the School of Electrical Engineering & Computer Science, The University of Newcastle, NSW, Australia 2308. cristian.rojas@studentmail.newcastle.edu.au, james.welsh@newcastle.edu.au, graham.goodwin@newcastle.edu.au

A. Feuer is with the Department of Electrical Engineering, Technion, Haifa 32000, Israel. feuer@ee.technion.ac.il

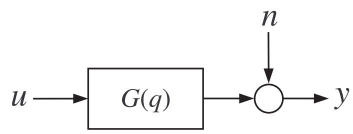

Fig. 1. Block diagram describing the relationship between the input u, the noise n and the output y of the system G to be identified.

III. Sections VI and VII develop the form of the cost function which satisfies the properties in Section III. In Section VIII we show that bandlimited '$1/f$' noise is an optimal input signal according to this cost function, and Section IX clearly illustrates the advantages of bandlimited '$1/f$' noise by means of an example. We present conclusions in Section X.

II. A MEASURE OF THE 'GOODNESS' OF AN EXPERIMENT

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. To construct such a measure, we make use of the work of Ljung [15], who has shown that, for a broad class of linear systems, the variance of the error in the estimated discrete time frequency response takes the following asymptotic (in both system order and data points) form:

$$\mathrm{Var}(\hat{G}(e^{j\omega})) = K\frac{\phi_n(\omega)}{\phi_u(\omega)}; \qquad \omega \in [0, 2\pi], \tag{1}$$

where ϕ_n is the noise spectral density and ϕ_u is the input spectral density. Here K is a function of the number of system parameters and the number of observations. Figure 1 shows how the input u, the noise n and the output y of the system are related, i.e.

$$y(t) = G(q)u(t) + n(t) \tag{2}$$

where G is the transfer function of the system, and q is the forward shift operator.

Actually, it has been argued in [19] that better approximations exist to that given in (1) but the simpler expression (1) suffices for our purposes. In fact, the expressions given in [19] for Box-Jenkins and Output-Error models include a factor which is dependent, for some particular special cases, only upon the poles of G. We note that this can be incorporated into ϕ_n, thus obtaining a test signal which is independent on ϕ_n and the plant. This implies that the results given here are exact for some classes of models of finite order.

An interesting and highly desirable property of (1) is that it is essentially independent of the system parameters. This is because it depends only on ϕ_n and ϕ_u. Of course, ϕ_n is somewhat problematic since it would also be desirable to have (1) independent of the real characteristics of the noise. This will also be part of our consideration.

As argued in [12, 21, 11, 26], absolute variances are not particularly useful when one wants to carry out an experiment design 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 preferable to work with relative errors. Thus, if $|G(e^{j\omega})|$ is the magnitude of the frequency response of the plant at frequency ω, then equation (1) suggests that the relative variance at frequency ω is given by

$$\text{Rel. Var}(\hat{G}(e^{j\omega})) = K\frac{\phi_n(\omega)}{\phi_u(\omega)|G(e^{j\omega})|^2}; \qquad \omega \in [0, 2\pi]. \tag{3}$$

Finally, 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 of the form:

$$\begin{aligned} J(\phi_u) &= \int_a^b F(\mathrm{Var}(\hat{G}(e^{j\omega}))/|G(e^{j\omega})|^2)W(\omega)d\omega \\ &= \int_a^b F\left(\frac{K\phi_n(\omega)}{\phi_u(\omega)|G(e^{j\omega})|^2}\right)W(\omega)d\omega, \end{aligned} \tag{4}$$

where F and W are functions to be specified later, and $0 < a < b < 2\pi$. Here, W is a weighting function that allows the control engineer to define at which frequencies it would be preferable to obtain a better model (depending on the control requirements, but not necessarily on the true plant characteristics).

In the next Section we propose some desirable properties of the functions F and W.

III. DESIRABLE PROPERTIES OF THE COST FUNCTION

We consider two sets of criteria. The first relates principally to the function F, the second to the function W. In addition to these properties, we will also assume that $F \in C^1([a,b], \mathbb{R}_0^+)$ and $W \in C^1([a,b], \mathbb{R}^+)$, where $C^1(X,Y)$ is the space of all functions from $X \subseteq \mathbb{R}$ to $Y \subseteq \mathbb{R}$ having a continuous derivative.

Criteria A

It is reasonable to consider a cost function (4) whose minimum is achieved by a function which does not depend on the actual system characteristics. The reason being that these characteristics are typically unknown at the time the experiment is applied, and in fact it is the purpose of the experiment to reveal this information.

On the other hand, the cost function (4) should be a measure of the 'size' of the variance in the estimation of the plant frequency response. Hence, loosely speaking, the cost function should increase accordingly to an increase of the variance at any frequency.

The above argument implies that the function F for measure (4) should be chosen so as to satisfy the following requirements:

A.1) The optimal experiment, ϕ_u^*, which minimizes J in (4), should be independent on the plant $|G(e^{j\omega})|^2$ and the noise variance ϕ_n.

A.2) The integrand in (4) 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.

Criterion B

Many properties of linear systems depend on the *ratio* of poles and zeros rather than on their absolute locations in the frequency domain [2, 9, 22]. This implies that if we scale the frequency ω by a constant, the optimal input must keep its shape, as the poles and zeros of the new plant will have the same ratios as before. This invariance property must be reflected in the weighting function W, which has to give equal weight to frequency intervals whose endpoints are in the same proportion.

Thus, the weighting function W should be such that for every $0 < \alpha < \beta < 2\pi$ and every $k > 0$ such that $0 < k\alpha < k\beta < 2\pi$ we have that

$$\int_{\alpha}^{\beta} W(\omega)d\omega = \int_{k\alpha}^{k\beta} W(\omega)d\omega. \tag{5}$$

IV. CONSTRAINTS

Our goal will then be to optimize a cost function as in (4) where ϕ_u is constrained in some fashion. A typical constraint used in experiment design is that the total input energy should be constrained [10, pg. 125]. Thus, we need to optimize $J(\phi_u)$ subject to a constraint of the form

$$\int_{a}^{b} \phi_u(\omega)d\omega = 1. \tag{6}$$

Specifically our goal is to adjust F and W such that the optimal experiment that minimizes (4) subject to the constraint (6) satisfies the criteria A.1, A.2 and B in Section III.

V. A PRELIMINARY TECHNICAL RESULT

Motivated by the need for a measure to be independent of the system and such that criteria A.1, A.2 and B are met subject to a constraint on the input, we have established the following result:

Lemma 1: For $0 < a < b < 2\pi$, let $g, F \in C^1([a,b], \mathbb{R}_0^+)$ and $W \in C^1([a,b], \mathbb{R}^+)$. Define, if it exists,

$$f^*(g) := \arg \min_{\substack{f \in C^1([a,b],\mathbb{R}^+) \\ \int_a^b f(x)dx = 1}} \int_{a}^{b} F\left(\frac{g(x)}{f(x)}\right) W(x)dx. \tag{7}$$

If $f^*(g)$ does not depend on g, then there are constants $\alpha, \beta, \gamma \in \mathbb{R}$ such that

$$F(y) = \alpha \ln y + \beta; \quad \inf_{x\in[a,b]} \frac{g(x)}{f(x)} \le y \le \sup_{x\in[a,b]} \frac{g(x)}{f(x)}, \tag{8}$$

and $f^* = \gamma W$.

Proof: Let $g, F \in C^1([a,b], \mathbb{R}_0^+)$ and $W \in C^1([a,b], \mathbb{R}^+)$ be fixed, and such that $f^*(g)$, as defined in (7), exists. Then, by [16, Section 7.7, Theorem 2], there is a constant $\lambda \in \mathbb{R}$ for which $f^*(g)$, is a stationary point of

$$J_\lambda(f) := \int_{a}^{b} F\left(\frac{g(x)}{f(x)}\right) W(x)dx + \lambda \int_{a}^{b} f(x)dx. \tag{9}$$

Thus, for any $h \in C^1([a,b], \mathbb{R}_0^+)$ we have that $\delta J_\lambda(f^*; h) = 0$, which means [16, Section 7.5] that

$$\int_{a}^{b} \left[F'\left(\frac{g(x)}{f^*(x)}\right)\left(-\frac{g(x)}{(f^*(x))^2}\right) W(x) + \lambda\right] h'(x)dx = 0, \tag{10}$$

thus, by [16, Section 7.5, Lemma 1],

$$F'\left(\frac{g(x)}{f^*(x)}\right) W(x) \frac{g(x)}{(f^*(x))^2} = \lambda; \quad x \in [a,b]. \tag{11}$$

Let $l(x) := g(x)/f^*(x)$, then (11) can be written as

$$F'(l(x))l(x) = \lambda \frac{f^*(x)}{W(x)}; \quad x \in [a,b]. \tag{12}$$

The left side of (12) depends on g, but the right does not (because of the assumption on the independence of f^* upon g). Thus, both sides are equal to a constant, say, $\alpha \in \mathbb{R}$, which implies that

$$F'(l(x)) = \frac{\alpha}{l(x)}; \quad x \in [a,b]. \tag{13}$$

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

$$F(l(x)) = \alpha \ln l(x) + \beta; \quad x \in [a,b] \tag{14}$$

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

On the other hand, we have that

$$\lambda \frac{f^*(x)}{W(x)} = \alpha, \tag{15}$$

so if we define $\gamma := \alpha/\lambda$, we conclude that $f^* = \gamma W$. This concludes the proof. ∎

VI. CHOICE OF THE FUNCTION F

In this Section we use the result of the previous Section to find a suitable function F which satisfies Criteria A.1 and A.2, and to find the optimal input signal for the resulting cost function.

We first examine the choice of the function F in (4). Now, we may take, without loss of generality, $\alpha = 1$ and $\beta = 0$ for the function F given by Lemma 1. This is because, according to Lemma 1, every cost function (4) satisfying Criteria A.1 and A.2 is minimized by the same $f \in C^1([a,b], \mathbb{R}^+)$. Thus, such a cost function can be written as

$$J(\phi_u) = \int_{a}^{b} \ln\left(\frac{K\phi_n(\omega)}{\phi_u(\omega)|G(e^{j\omega})|^2}\right) W(\omega)d\omega. \tag{16}$$

It is then relatively straightforward to optimize (16) subject to the constraint given by (6). Indeed, by Lemma 1 the optimal experiment will be essentially given by a scaled version of W, i.e.

$$\phi_u^*(\omega) = \frac{1}{\int_a^b W(x)dx} W; \qquad \omega \in [a,b]. \tag{17}$$

The following Lemma establishes that ϕ_u^* gives not only an extremum, but a global minimum for the cost function (16).

Lemma 2: The function ϕ_u^* defined in (17) gives the global minimum of the cost function (16). In other words, for $0 < a < b < 2\pi$, let $W \in C^1([a,b],\mathbb{R}^+)$, then,

$$\phi_u^* = \tag{18}$$
$$\arg \min_{\substack{\phi_u \in C^1([a,b],\mathbb{R}^+) \\ \int_a^b \phi_u(\omega)d\omega = 1}} \int_a^b \ln\left(\frac{K\phi_n(\omega)}{\phi_u(\omega)|G(e^{j\omega})|^2}\right) W(\omega)d\omega.$$

Proof: The cost function (16) can be written as

$$J(\phi_u) = C - \int_a^b \ln(\phi_u(\omega))W(\omega)d\omega, \tag{19}$$

where C is a constant, independent of ϕ_u, given by

$$C := \int_a^b \ln\left(\frac{K\phi_n(\omega)}{|G(e^{j\omega})|^2}\right) W(\omega)d\omega. \tag{20}$$

Now, if ϕ_u is any function in $C^1([a,b],\mathbb{R}^+)$ such that $\int_a^b \phi_u(\omega)d\omega = 1$, then by (17) we have that

$$\begin{aligned}
J(\phi_u) &= C - \int_a^b \ln[\phi_u^*(\omega) + (\phi_u(\omega) - \phi_u^*(\omega))]W(\omega)d\omega \\
&= C - \int_a^b \ln(\phi_u^*(\omega))W(\omega)d\omega \\
&\quad - \int_a^b \frac{1}{\phi_u^*(\omega)}(\phi_u(\omega) - \phi_u^*(\omega))W(\omega)d\omega \\
&\quad - \int_a^b h(\phi_u(\omega), \phi_u^*(\omega))W(\omega)d\omega \qquad (21)\\
&= J(\phi_u^*) - \int_a^b h(\phi_u(\omega), \phi_u^*(\omega))W(\omega)d\omega \\
&\quad - \left(\int_a^b W(\omega)d\omega\right)\left(\int_a^b (\phi_u(\omega) - \phi_u^*(\omega))d\omega\right) \\
&= J(\phi_u^*) - \int_a^b h(\phi_u(\omega), \phi_u^*(\omega))W(\omega)d\omega,
\end{aligned}$$

where $h : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is given by

$$h(x,y) := \ln x - \ln y - \frac{1}{y}(x-y). \tag{22}$$

Thus, since $w > 0$, to prove that ϕ_u^* gives the global minimum for the cost function (16), it suffices to show that $h(x,y) < 0$ for every $x, y \in \mathbb{R}^+$ such that $x \neq y$. To this end, notice that

$$\frac{\partial h}{\partial x}(x,y) = \frac{1}{x} - \frac{1}{y}, \tag{23}$$

thus if $x > y$, then

$$h(x,y) = h(y,y) + \int_y^x \frac{\partial h}{\partial x}(\tilde{x}, y)d\tilde{x} < 0, \tag{24}$$

and similarly for $x < y$. This proves the Lemma. ∎

The relationship given in (17) highlights the importance of choosing the correct function W so as to reflect the desired relative frequency weighting. The choice of W will be explored in the next Section.

VII. Choice of the function W

A weighting function which is reasonable in the sense that it satisfies Criterion B is described below:

Lemma 3: For $0 < a < b < 2\pi$, let $W \in C^1([a,b],\mathbb{R}^+)$. If W satisfies

$$\int_\alpha^\beta W(\omega)d\omega = \int_{k\alpha}^{k\beta} W(\omega)d\omega \tag{25}$$

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

Proof: Since W is continuous, we have from (25) that

$$\begin{aligned}
W(a) &= \lim_{\varepsilon \to 0_+} \frac{\int_a^{a+\varepsilon} W(\omega)d\omega}{\varepsilon} \\
&= \lim_{\varepsilon \to 0_+} k \frac{\int_{ka}^{ka+k\varepsilon} W(\omega)d\omega}{k\varepsilon} \qquad (26)\\
&= kW(ka)
\end{aligned}$$

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

$$W(ka) = \frac{1}{k}W(a); \quad a \leq ka < b, \tag{27}$$

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

$$W(x) = \frac{a}{x}W(a) = \frac{\lambda}{x}; \quad a \leq x < b. \tag{28}$$

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

With this last result, and those of the previous Sections, we can now proceed to establish the form of a suitable measure of the 'goodness' of an experiment, and an optimal input signal according to this cost function. This will be done in the next Section.

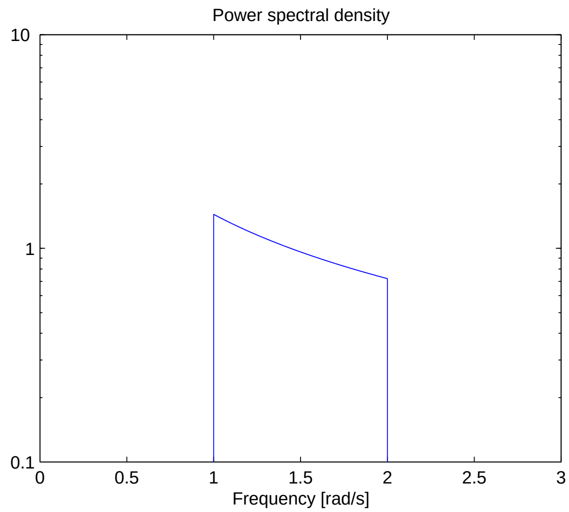

Fig. 2. Power spectral density of bandlimited '$1/f$' noise signal for $a = 1$ and $b = 2$.

VIII. Band-limited '$1/f$' noise

If we apply the results of the previous sections to the cost function (16), then we immediately see that a reasonable cost function for measuring the 'goodness' of an experiment when having only diffuse prior knowledge about a plant is

$$J(\phi_u) = \int_a^b \ln\left(\frac{K\phi_n(\omega)}{\phi_u(\omega)|G(e^{j\omega})|^2}\right)\frac{1}{\omega}d\omega. \tag{29}$$

Therefore, according to (17) and Lemma 2, the optimal input spectrum is given by

$$\phi_u^*(\omega) = \frac{1/\omega}{\int_a^b \frac{d\omega}{\omega}} = \frac{1/\omega}{\ln b - \ln a}; \quad \omega \in [a, b]. \tag{30}$$

Figure 2 shows the spectral density of this type of signal, known as *bandlimited '$1/f$' noise*, for $a = 1$ and $b = 2$.

Thus we see that, subject to the assumptions introduced above, i.e. Criteria A.1, A.2 and B, '$1/f$' noise is the robust input signal for identifying a system when one has only diffuse prior knowledge.

Remark 1: The fact that bandlimited '$1/f$' noise is the solution of a variational problem means that it is possible to consider additional prior information by imposing constraints in the optimisation problem. In this sense, the problem of experiment design resembles the development of the Principle of Maximum Entropy as given in [23, 24].

IX. Example

We have seen above that bandlimited '$1/f$' noise can be regarded as a robust optimal test signal in the sense described in Section VIII. This result is consistent with earlier findings in the literature, which show that bandlimited '$1/f$' noise has near optimal properties for specific classes of systems. For example, it is known to yield a performance which is within a factor of 2 of the optimum for certain families of one-parameter problems [12, 21, 11], although general results for multi-parameter problems are not yet available.

TABLE I

Relative Values of Cost for Different Input Signals

	$\max_{\theta\in\Theta} [\theta^2 \overline{M}(\theta, \phi_u)]^{-1}$
Single frequency at $\omega = 1$	7.75
Bandlimited white noise	12.09
Bandlimited '$1/f$' noise	1.43
Robust min-max optimal input	1.00

Table I, reproduced from [21], shows some interesting results. In particular, this Table shows the numerical results for the problem of designing an input signal to identify the parameter θ of the plant

$$G(s) = \frac{1}{s/\theta + 1}, \tag{31}$$

where it is assumed *a priori* that θ lies in the range $\Theta := [0.1, 10]$. The cost function used for comparison is the worst case normalized variance of an efficient estimator of θ,

$$J'(\phi_u) := \max_{\theta\in\Theta} \left[\int_0^\infty \frac{\omega^2/\theta^2}{(\omega^2/\theta^2 + 1)^2}\phi_u(\omega)d\omega\right]^{-1}, \tag{32}$$

where the inputs being compared are

(i) A sine wave of frequency 1 (this is the optimal input if the true parameter is $\widehat{\theta} = 1$).

(ii) Bandlimited white noise input, limited to the frequency range $[0.1, 10]$.

(iii) Bandlimited '$1/f$' noise input, limited to the frequency range $[0.1, 10]$.

(iv) The approximate discretised robust optimal input generated by Linear Programming [21].

Notice that, for ease of comparison, the costs in Table I have been normalized so that the robust optimal input has cost 1.00. Figure 3 shows the performance of these signals according to the normalized variance obtained as a function of the true value of θ. Both Table I and Figure 3 demonstrate that bandlimited '$1/f$' noise does indeed yield good performance at least in terms of an specific example. The results presented in the current paper give theoretical support to these earlier observations.

X. Conclusions

In this paper, we have studied the problem of robust experiment design in the face of diffuse prior information. We have analysed a general class of criteria for measuring how good an experiment is, and have found that there is a specific measure within this class that gives a system independent optimal experiment design, which is suitable for the case when one only has a vague idea about the plant to be identified. We have also shown that '$1/f$' noise is optimal according to this cost function.

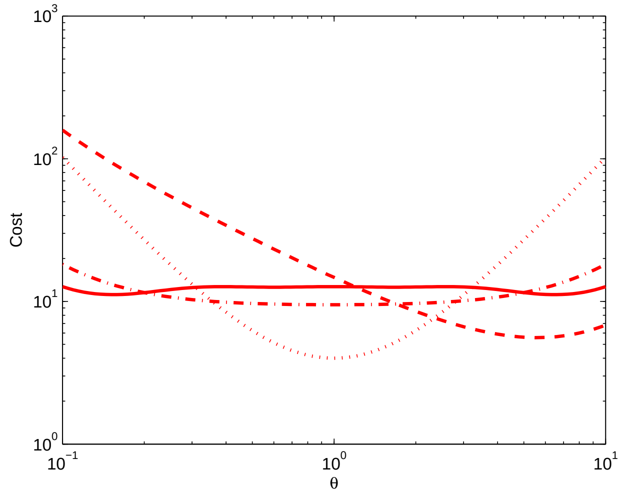

Fig. 3. Variation of the normalized variance $\left[\int_0^\infty \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\phi_u(\omega)d\omega\right]^{-1}$ as a function of θ for various input signals: the robust optimal input (solid), a sine wave of frequency 1 (dotted), bandlimited white noise (dashed) and bandlimited '$1/f$' noise (dash-dotted).

References

[1] A. C. Atkinson and A. N. Doner. *Optimum Experiment Design*. Clarendon Press, Oxford, 1992.

[2] H. W. Bode. *Network Analysis and Feedback Amplifier Design*. D. Van Nostrand, 1945.

[3] K. Chaloner and I. Verdinelli. Bayesian experiment design: A review. *Statistical Science*, 1995.

[4] H. Chernoff. Approaches in sequential design of experiments. In J.N. Srivastava, editor, *Survey of Statistical Design*, pages 67–90. North–Holland, Amsterdam, 1975.

[5] D. R. Cox. *Planning of Experiments*. Wiley, New York, 1958.

[6] M. A. El-Gamal and T. R. Palfrey. Economical experiments: Bayesian efficient experimental design. *Int. J. Game Theory*, 25(4):495–517, 1996.

[7] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York and London, 1972.

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

[9] G. C. Goodwin, S. F. Graebe, and M. E. Salgado. *Control System Design*. Prentice Hall, Upper Saddle River, New Jersey, 2001.

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

[11] G. C. Goodwin, C. R. Rojas, and J. S. Welsh. Good, bad and optimal experiments for identification. In T. Glad, editor, *Forever Ljung in System Identification - Workshop on the occasion of Lennart Ljung's 60th birthday*. September 2006.

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

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

[14] O. Kempthorne. *Design and Analysis of Experiments*. Wiley, New York, 1952.

[15] L. Ljung. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions on Automatic Control*, 30:834–844, September 1985.

[16] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[17] R. K. Mehra. Optimal inputs for system identification. *IEEE Transactions on Automatic Control*, AC-19:192–200, 1974.

[18] W. G. Müller and B. M. Pötscher. Batch sequential design for a nonlinear estimation problem. In V.V. Fedorov, W.G. Müller, and I.N. Vuchkov, editors, *Model-Orientated Data Analysis, a Survey of Recent Methods*, pages 77–87. Physica-Verlag, Heidelberg, 1992.

[19] B. Ninness and H. Hjalmarsson. Variance error quantifications that are exact for finite-model order. *IEEE Transactions on Automatic Control*, 49(8):1275–1291, 2004.

[20] L. Pronzato and E. Walter. Robust experiment design via maxmin optimisation. *Mathematical Biosciences*, 89:161–176, 1988.

[21] C. R. Rojas, G. C. Goodwin, J. S. Welsh, and A. Feuer. Robust optimal experiment design for system identification. *Automatica (accepted for publication)*, 2006.

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

[23] J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, 1980.

[24] J. Skilling. The axioms of maximum entropy. In G. J. Erickson and C. R. Smith, editors, *Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, pages 173–187. Kluwer Academic Publishers, 1988.

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

[26] J. S. Welsh, G. C. Goodwin, and A. Feuer. Evaluation and comparison of robust optimal experiment design criteria. *Proceedings of the American Control Conference, Minneapolis, USA*, pages 1659–1664, 2006.

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