Good, Bad and Optimal Experiments for Identification

Graham C. Goodwin, Cristian R. Rojas and James S. Welsh

Abstract: The topic of experiment design has attracted on-going interest for almost half a century. There is always a strong incentive to learn the most about a system with minimal perturbation to normal operations. In the early engineering literature there was considerable interest in nominal experiment design based on a-priori estimates of parameters. However, this can be precarious if the a-priori estimates are poor. This chapter will discuss more recent results on robust experiment design which, inter-alia, account for cases where the prior knowledge is less precise. Two illustrative case studies will be used to highlight the differences between "good, bad and optimal" experiments. In experiment design, as in many engineering problems, it is often better to have an approximate answer to the correct question than an "optimal" answer to the wrong question.

1 Introduction

This chapter is dedicated to, and inspired by, the work of Lennart Ljung. For over three decades he has shown unparalleled leadership in the area of system identification.

Quoting from (Ljung, 1999), "the construction of a model from data involves three basic entities:

1. The data set.
2. A set of candidate models.
3. A rule by which candidate models can be assessed using the data."

In this chapter we will focus on the first of these entities. Again quoting (Ljung, 1999), "The input-output data are sometimes recorded during a specifically designed identification experiment, where the user may determine which signals to measure and when to measure them and may also choose the input signals. The objective with *experiment design* is thus to make these choices so that the data become maximally informative, subject to constraints that may be at hand."

This chapter will contribute to the understanding of what constitutes good, bad and optimal experiments.

Experiment design has been the subject of intense study in both the statistics literature (Wald, 1943; Cox, 1958; Kempthorne, 1952; Kiefer and Wolfowitz, 1960; Karlin and Studden, 1966; Fedorov, 1972; Whittle, 1973; Wynn, 1972) and the engineering literature (Levadi, 1966; Gagliardi, 1967; Goodwin and Payne, 1973; Goodwin, Payne and Murdoch, 1973; Goodwin, Murdoch and Payne, 1973; 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.

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

One may imagine that, provided a good a-priori estimate of the parameter vector is available, then an experiment designed on the basis of this parameter vector would still be useful in practice. However, one does need to be rather careful that the experiment does not simply reinforce what one already knows (or believes to know) instead of making a genuine exploration of "unchartered territories".

Indeed, these issues have inspired the title of this chapter which is intended to make a contrast between an "optimal experiment" based on some narrowly defined optimisation criterion and what we might think of heuristically as a good (or indeed bad) experiment. The title has been chosen to mirror the paper (Rosenbrock and McMorran, 1971), which discusses perceived inadequacies in "optimal" linear quadratic regulators.

The issue of robustness to nominal parameter values has been addressed in both the statistics and engineering literature. Suggested strategies for dealing with robustness 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; Hjalmarsson, 2005; Gevers, 2005).

- 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; Gevers and Bombois, 2006; Goodwin, Welsh, Feuer and Derpich, 2006; Mårtensson and Hjalmarsson, 2006; Goodwin, Rojas, Welsh and Feuer, 2006).

Here we build on and extend the latter work, where the basic idea is to assume that the system parameters, θ, are contained in a given compact set Θ. A design criterion $f(M(\theta), \theta)$ is chosen, where $M(\theta)$ is the Fisher information matrix evaluated at θ, and an experiment is designed to optimise the worst case of $f(M(\theta), \theta)$ over Θ. Notice that this differs from the nominal experiment design where one would optimise $f(M(\theta_0), \theta_0)$ for some given nominal value θ_0.

The main contribution of the current chapter is the detailed study of the design of a robust optimal input signal for the identification of the resonance frequency of a second order system. Furthermore, theoretical support is developed for the key properties of existence and finite support of this robust optimal input.

The layout of the chapter is as follows: In Section 2 we present some standard results on the information matrix for linear dynamic systems. A general formulation of the nominal and min-max approaches to optimal experiment design is given in Sections 3 and 4, respectively. Section 5 explores two illustrative one parameter examples: a first order system and a second order resonant system. In Section 6 we present a theoretical analysis of the second order resonant system. Finally, in Section 7 we draw conclusions.

2 The Information Matrix

To set the scene, 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.

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.

For simplicity, 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. Then, 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

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

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 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

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

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

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

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.

3 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 ($\operatorname{tr} \overline{M}(\theta, \phi_u)^{-1}$), L-optimality ($\operatorname{tr} W\overline{M}(\theta, \phi_u)^{-1}$, for some $W \geq 0$) 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.

We will argue below that beyond the intended use of the model it is also important that the test signal be "robust" to potential uncertainties regarding the true nature of the system.

4 Min-Max Robust Design

So as to reduce the sensitivity of the experiment to the nominal parameter values, 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]

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

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 later. Notice also that we allow J to depend explicitly on θ; this point will be of practical importance – see discussion below.

An alternative approach to that described above would be to extend the space to include "mixed policies" (Başar and Bernhard, 1995; Goodwin, Rojas, Welsh and Feuer, 2006) by introducing a (probability) density ξ on Θ. The counterpart of (18) would now take the form:

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

[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 $\operatorname{supp} \phi_u \subset X$, where $\operatorname{supp} \phi_u$ is the support of ϕ_u (i.e. roughly speaking, $\mathscr{S}(X)$ is the set of all (generalised) probability density functions on X).

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.

Notice that if f were linear it could be introduced into the inner integral, and then it can be shown that this approach is equivalent to the previous one.

5 One Parameter Examples

To gain maximal insight into robust experiment design, we will study two simple continuous time problems in some detail. These are:

(i) A first order system with known gain but unknown time constant.

(ii) A resonant second order system with known damping but unknown resonance frequency.

5.1 First Order System

This system has been studied in detail in a recent paper by the same authors (Goodwin, Rojas, Welsh and Feuer, 2006). Thus, we will only provide a summary of the key results which will then form the basis for comparison with the second order system studied next.

The first order system is defined by $G_2(s) = 1$ and

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

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

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

It is well known (Goodwin and Payne, 1977) that the nominal optimal open loop input for this system can be realised by a single sinusoid of frequency $\omega^* = \theta_0$, where θ_0 is the given a-priori nominal parameter value.

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

We next turn to min-max robust experiment design. We use

$$J(\overline{M}(\theta,\phi_u),\theta) \triangleq [\theta^2\, \overline{M}(\theta,\phi_u)]^{-1}\,. \tag{24}$$

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

$$\phi_u^* = \arg \min_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \overline{J}(\phi_u)\,, \tag{25}$$

where

$$\overline{J}(\phi_u) \triangleq \sup_{\theta\in\Theta} \left[\int_0^\infty \frac{\omega^2/\theta^2}{(\omega^2/\theta^2+1)^2}\, \phi_u(\omega)\, d\omega \right]^{-1} \tag{26}$$

and

$$\Theta \triangleq \{\theta : \underline{\theta} \le \theta \le \overline{\theta}\}\,. \tag{27}$$

Remark 1. The reason for multiplying by θ^2 in (24) and then inverting is that $\overline{M}^{-1}$ is a variance measure and thus $[\theta^2\overline{M}]^{-1}$ gives relative (mean square) errors. This use of per-unit error concept is important in robust design to ensure that like quantities are compared. See (Goodwin, Rojas, Welsh and Feuer, 2006).

For the above problem, the following key properties have been established in (Goodwin, Rojas, Welsh and Feuer, 2006):

(i) Compact support: Every optimal input should have all its energy inside $[\underline{\theta},\overline{\theta}]$.

(ii) Existence: There exists at least one optimal input.

(iii) Uniqueness: The optimal input is unique, and $\underline{\theta}$ and $\overline{\theta}$ do not belong to the input spectrum.

(iv) Finite support: The optimal input has finite support in the frequency domain, and thus can be realised as a finite sum of sinusoids.

Another remarkable property (established in (Goodwin, Rojas, Welsh and Feuer, 2006)) is that bandlimited '$1/f$' noise, defined by the spectrum

$$\phi_u^{1/f}(\omega) \triangleq \begin{cases} \dfrac{1/\omega}{\ln\overline{\omega} - \ln\underline{\omega}}, & \omega \in [\underline{\omega},\overline{\omega}] \\ 0, & \text{otherwise} \end{cases}\,, \tag{28}$$

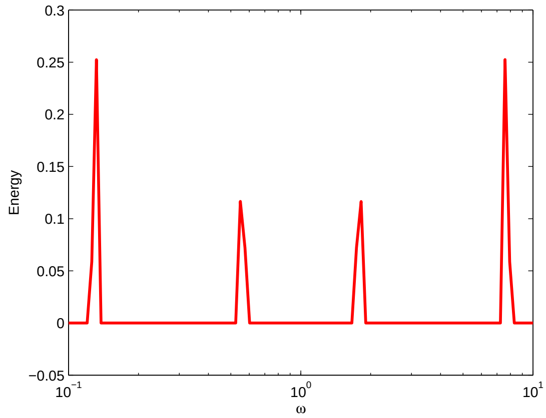

Figure 1.1: *Discretised spectrum of the robust optimal input for the first order system.*

where for this system we consider $\underline{\omega} = \underline{\theta}$ and $\overline{\omega} = \overline{\theta}$, is near optimal. In fact, it has been proven that the performance of bandlimited '$1/f$' noise is (at most) a factor of 2:1 away from the performance of the true optimal input.

A numerical procedure based on converting the min-max design problem into a standard LP problem is described in (Goodwin, Rojas, Welsh and Feuer, 2006). Choosing $\underline{\theta} = 0.1$, $\overline{\theta} = 10$ gives the *unique* optimal input spectrum shown in Figure 1.1.

We take the nominal parameter value as $\theta_0 = 1$ (the geometric mean of $\underline{\theta}$ and $\overline{\theta}$). Table 1.1 is reproduced from (Goodwin, Rojas, Welsh and Feuer, 2006).

Table 1.1: *Relative Values of Cost for the Different Input Signals*

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

We see from Table 1.1 that bandlimited white noise and the nominal optimal test signal achieve costs that are approximately an order of magnitude worse than those for the min-max optimal input. Surprisingly, bandlimited '$1/f$' noise achieves a performance which is within 40% of the optimal performance. More will be said about this in the context of the second order system discussed below.

5.2 A Resonant Second Order System

We next consider the following second order system: $G_2(s) = 1$,

$$G_1(s) = \frac{\theta^2}{s^2 + 2\xi\theta s + \theta^2}, \tag{29}$$

where $\xi \in (0, 1)$ is assumed known (to keep the example simple) and θ is assumed to lie in a given range $[\underline{\theta}, \overline{\theta}]$.

The scaled single frequency information matrix is readily seen to be

$$\theta^2 \widetilde{M}(\theta, \omega) = \frac{4[(\omega/\theta)^4 + \xi^2(\omega/\theta)^2]}{\{[1 - (\omega/\theta)^2]^2 + 4\xi^2(\omega/\theta)^2\}^2}\,. \tag{30}$$

The nominal optimal test can be realised by a single sinusoid at frequency (see section 6 for the details)

$$\omega^* = \sqrt{x}\,\theta, \tag{31}$$

where x is the unique positive root of

$$2x^3 + 3\xi^2 x^2 + 2(2\xi^4 - \xi^2 - 1)x - \xi^2 = 0\,. \tag{32}$$

Notice that $\omega^* \to \theta$ as $\xi \to 0$. Again, this result is heuristically reasonable since one places the optimal test signal very near the resonance frequency.

The above resonant system is an ideal test case to highlight the merits of robust experiment design. Indeed, whilst it makes sense to put the test signal energy near the resonance frequency, it can be reasonably assumed that this can totally "miss the target" if the resonance frequency is not quite where it was believed to be.

In the results presented below, we assume that $\theta \in [0.5, 2]$ with a nominal value of 1. Also, we consider 3 values for ξ, namely, 0.1, 0.05 and 0.02.

Figure 1.2 shows the sensitivity of the cost $[\theta^2 \widetilde{M}(\theta, \omega)]^{-1}$ with respect to θ when using the nominal optimal input designed for $\theta_0 = 1$. This figure provides a strong incentive to try to use robust designs for this example.

In the next section we will establish important properties of the min-max optimal design for this case. However, to motivate the reader, we first present the numerical results based on using the same algorithm as in (Goodwin, Rojas, Welsh and Feuer, 2006). See Remark 2 in Section 6 for details on the frequency ranges considered for the optimal input spectrum, bandlimited white noise and bandlimited '$1/f$' noise.

Figure 1.3 shows the spectrum of the robust optimal input for $\xi = 0.1$.

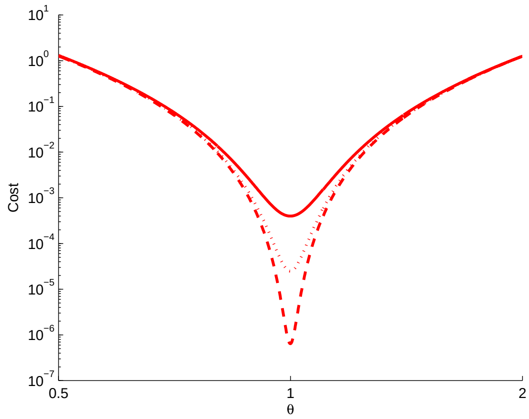

Figure 1.2: *Variation of the cost function,* $[\theta^2 \widetilde{M}(\theta,\omega)]^{-1}$*, with respect to* θ*, for* $\xi = 0.1$ *(solid),* $\xi = 0.05$ *(dotted) and* $\xi = 0.02$ *(dashed), when using the nominal optimal test signal designed for* $\theta_0 = 1$*.*

The results are summarised in Table 1.2.

(Note that the performance of the optimal input has been normalised to 1 for ease of comparison.)

A startling observation from Table 1.2 is how well bandlimited '$1/f$' noise performs. Specifically, we see that it is within a factor of 1.82 for $\xi = 0.1$, a factor of 1.90 for $\xi = 0.05$ and a factor of 1.97 for $\xi = 0.02$, of the optimal input.

Figure 1.4 exhibits the sensitivity of the cost $[\theta^2 \widetilde{M}(\theta,\omega)]^{-1}$ with respect to θ for the test signals considered in Table 1.2 for $\xi = 0.1$.

From Figure 1.4 it can be seen that the min-max optimal input is indeed robust for the parameter range $[0.5, 2]$ when contrasted against bandlimited '$1/f$' noise and bandlimited white noise. Furthermore, we observe that bandlimited '$1/f$' noise gives a performance that is nearly as good as the robust min-max optimal input.

The costs shown in Figures 1.2 and 1.4 give compelling support for classifying the input signals as good, bad and optimal. Clearly, for this example, the *bad* relates to the nominal optimal input, *good* the bandlimited '$1/f$' noise (and to a lesser extent the bandlimited white noise) and *optimal* the min-max input.

Recall that bandlimited '$1/f$' noise also performs extremely well in the case of the first order system studied in section 5.1. This leads us to the following controversial conjecture:

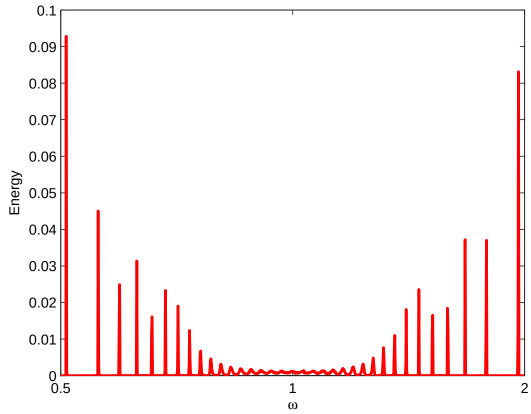

Figure 1.3: *Discretised spectrum of the robust optimal input designed for the resonant second order system with* $\xi = 0.1$.

A signal such as bandlimited '$1/f$' noise is actually a *good* (as opposed to optimal) test signal in system identification. The reason is that it is not only robust with respect to parameter variations but is also robust with respect to model structure.

6 Supporting Theory

Motivated by the results of section 5.2, we provide proofs of key properties for the second order resonant system, i.e. compact support, existence and finite support of the robust optimal test input. However, to provide some insight into the nature of the solution, we will consider first the nominal experiment design problem, which consists in finding a $\phi_u^* \in \mathscr{S}(\mathbb{R}_0^+)$, if it exists, such that

$$\phi_u^* = \arg \max_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \int_0^\infty \widetilde{M}(\theta_0, \omega)\, \phi_u(\omega) d\omega \tag{33}$$

for a fixed $\theta_0 \in \mathbb{R}^+$, where $\widetilde{M}$ is given by

$$\widetilde{M}(\theta, \omega) \triangleq \frac{1}{\theta^2} \frac{4[(\omega/\theta)^4 + \xi^2(\omega/\theta)^2]}{\{[1-(\omega/\theta)^2]^2 + 4\xi^2(\omega/\theta)^2\}^2}\,. \tag{34}$$

The details are somewhat more involved than those of the first order system analysed in (Goodwin, Rojas, Welsh and Feuer, 2006).

Table 1.2: *Relative Values of Cost* ($\max_{\theta\in\Theta}\ [\theta^2\overline{M}(\theta,\phi_u)]^{-1}$) *for Different Input Signals*

	Relative cost $\xi = 0.1$	Relative cost $\xi = 0.05$	Relative cost $\xi = 0.02$
Sinusoid at $\omega = 1$	338.16	$2,756.10$	$44,003.41$
Bandlimited white noise	3.69	4.00	3.65
Bandlimited '$1/f$' noise	1.82	1.90	1.97
Min-max optimal input	1	1	1

6.1 Nominal Experiment Design

To solve problem (33), we need to understand how $\widetilde{M}(\theta_0,\omega)$ behaves as a function of ω.

First notice that, from (34), $\widetilde{M}(\theta_0,\omega)$ is differentiable and nonnegative for every $\omega,\theta_0 \in \mathbb{R}^+$, while $\widetilde{M}(\theta_0,0) = 0$ and $\lim_{\omega\to\infty}\widetilde{M}(\theta_0,\omega) = 0$, and $\widetilde{M}(\theta_0,\cdot)$ does not have zeros in $\mathbb{R}^+$, so it must have at least one maximum in $\mathbb{R}^+$. Moreover,

$$\frac{\partial\widetilde{M}(\theta_0,\omega)}{\partial\omega} = -\frac{8\widetilde{\omega}\{2\widetilde{\omega}^6 + 3\xi^2\widetilde{\omega}^4 + [4\xi^4 - 2\xi^2 - 2]\widetilde{\omega}^2 - \xi^2\}}{\theta_0^5[\widetilde{\omega}^4 + 2(2\xi^2-1)\widetilde{\omega}^2 + 1]^3}\,. \tag{35}$$

where $\widetilde{\omega} \triangleq \omega/\theta_0$.

The second order factor in the denominator of (35) is of the form $x^2+2(2\xi^2-1)x+1$ (where $x \triangleq \widetilde{\omega}^2$), and it has roots at

$$x_{1,2} = 1 - 2\xi^2 \pm 2\xi^2\sqrt{\xi^2-1} \notin \mathbb{R}\,. \tag{36}$$

Hence, the denominator of (35) does not change sign. We next evaluate it at $\omega = 0$ and see that it is then always positive.

The numerator of (35), on the other hand, vanishes at $\omega = 0$ and $\omega = \sqrt{x}\theta$, where x is a positive root of

$$2x^3 + 3\xi^2x^2 + 2(2\xi^4 - \xi^2 - 1)x - \xi^2 = 0\,. \tag{37}$$

Now, the linear coefficient of (37) has roots at $\xi = \pm j/\sqrt{2}, \pm 1$, hence for $\xi \in (0,1)$ it is negative. Thus, the coefficients of (37) have only one change of sign for every $\xi \in (0,1)$, so by Descartes' rule of signs (Dickson, 1914), (37) has exactly one positive root for every $\xi \in (0,1)$.

By the preceding analysis, $\widetilde{M}(\theta_0,\omega)$ has exactly one maximum at $\omega = \sqrt{x}\theta$, where x is the unique positive root of (37). Thus, the optimal nominal input

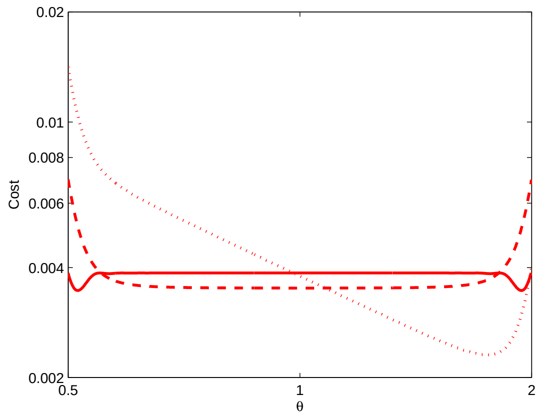

Figure 1.4: *Variation of cost versus θ for the robust optimal input (solid), bandlimited white noise (dotted) and bandlimited '$1/f$' noise (dashed).*

should have all its energy at this particular frequency, i.e.

$$\phi_u^* = \arg \max_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \int_0^\infty \widetilde{M}(\theta_0, \omega)\, \phi_u(\omega) d\omega = \delta_{\sqrt{x}\theta}, \tag{38}$$

where δ_α is the Dirac distribution with support in α.

6.2 Robust Experiment Design

The robust optimal experiment design problem for the resonant second order system consists in finding a $\phi_u^* \in \mathscr{S}(\mathbb{R}_0^+)$, if it exists, such that

$$\phi_u^* = \arg \max_{\phi_u \in \mathscr{S}(\mathbb{R}_0^+)} \inf_{\theta \in \Theta} \int_0^\infty \frac{4(\omega/\theta)^2[\xi^2 + (\omega/\theta)^2]}{\{[1-(\omega/\theta)^2]^2 + 4\xi^2(\omega/\theta)^2\}^2}\, \phi_u(\omega) d\omega\, . \tag{39}$$

We first observe that, since the integrand in (39) 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 "$\inf$" symbol in (39) to "$\min$".

Furthermore, if we make the following changes of variables

$$\begin{aligned} x &\triangleq \frac{\ln\theta - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}}, \\ y &\triangleq \frac{\ln\omega - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}}, \\ \phi_u(\omega) &= \frac{2}{k\omega}\,\widetilde{\phi}_u\left(\frac{\ln\omega - \ln\underline{\theta}}{\ln\overline{\theta} - \ln\underline{\theta}}\right), \\ k &\triangleq 2(\ln\overline{\theta} - \ln\underline{\theta}), \end{aligned} \tag{40}$$

then the problem can be rewritten as

$$\widetilde{\phi}_u^* = \arg \max_{\widetilde{\phi}_u \in \mathscr{S}(\mathbb{R})} \min_{x\in[0,1]} \int_0^\infty \frac{4e^{k(y-x)}[\xi^2 + e^{k(y-x)}]}{\{[1-e^{k(y-x)}]^2 + 4\xi^2 e^{k(y-x)}\}^2}\,\widetilde{\phi}_u(y)dy\,. \tag{41}$$

To simplify the notation, let $F(x,y) \triangleq f(y-x)$, where

$$f(u) \triangleq \frac{4e^{ku}[\xi^2 + e^{ku}]}{\{[1-e^{ku}]^2 + 4\xi^2 e^{ku}\}^2}\,. \tag{42}$$

We next establish the properties of existence and finite support for ϕ_u^* and $\widetilde{\phi}_u^*$ in the following theorems.

Lemma 1 (Compact support of the optimal input spectrum)
For the problem stated in (41)*, the optimal input* $\widetilde{\phi}_u^*$*, if it exists, has compact support. Namely,*

$$\int_{\mathbb{R}-[\alpha,1+\alpha]} \widetilde{\phi}_u^*(y)dy = 0\,, \tag{43}$$

where $\alpha \in \mathbb{R}$ *is the only real solution of the equation*

$$2e^{3k\alpha} + 3\xi^2 e^{2k\alpha} + 2(2\xi^4 - \xi^2 - 1)e^{k\alpha} - \xi^2 = 0\,. \tag{44}$$

Thus, $\widetilde{\phi}_u^* \in \mathscr{S}([\alpha, 1+\alpha])$*, so we can replace* (41) *with*

$$\tilde{\widetilde{\phi}}_u^* = \arg \max_{\tilde{\widetilde{\phi}}_u \in \mathscr{S}([0,1])} \min_{x\in[0,1]} \int_0^1 \frac{4e^{k(y-x+\alpha)}[\xi^2 + e^{k(y-x+\alpha)}]}{\{[1-e^{k(y-x+\alpha)}]^2 + 4\xi^2 e^{k(y-x+\alpha)}\}^2}\,\tilde{\widetilde{\phi}}_u(y)dy\,, \tag{45}$$

where $\tilde{\widetilde{\phi}}_u(y) = \widetilde{\phi}_u(y+\alpha)$ *for every* $y \in [0,1]$.

Proof: The derivative of f is

$$\frac{df(u)}{du} = -\frac{4ke^{ku}[2e^{3ku} + 3\xi^2 e^{2ku} + (4\xi^4 - 2\xi^2 - 2)e^{ku} - \xi^2]}{[e^{2ku} + 2(2\xi^2 - 1)e^{ku} + 1]^3}, \qquad (46)$$

which has essentially the same form as $\partial\widetilde{M}(\theta_0, \omega)/\partial\omega$ in (35), after replacing $(\omega/\theta_0)^2$ by e^{ku}. As in section 6.1, by the application of Descartes' rule of signs we can show that $f(u)$ has a unique global and local maximum at $u = \alpha \triangleq k^{-1}\ln x$, where x is the unique positive root of

$$2x^3 + 3\xi^2 x^2 + 2(2\xi^4 - \xi^2 - 1)x - \xi^2 = 0\,. \qquad (47)$$

Moreover, since $\lim_{|u|\to\infty} f(u) = 0$, we have that $\partial f/\partial u > 0$ for $u < \alpha$ and $\partial f/\partial u < 0$ for $u > \alpha$. Thus for any $x \in [0, 1]$ we have

$$\int_{-\infty}^{\infty} F(x, y)\widetilde{\phi}_u^*(y)dy \le \int_{-\infty}^{\infty} F(x, y)\widetilde{\phi}_u'(y)dy\,, \qquad (48)$$

where $\widetilde{\phi}_u'$ is given by

$$\widetilde{\phi}_u' \triangleq \widetilde{\phi}_u^*\,\mathcal{X}_{[\alpha,1+\alpha]} + \delta_\alpha \int_{-\infty}^{\alpha_-} \widetilde{\phi}_u^*(\tau)d\tau + \delta_{1+\alpha} \int_{(1+\alpha)_+}^{\infty} \widetilde{\phi}_u^*(\tau)d\tau\,, \qquad (49)$$

and $\mathcal{X}_{[\alpha,1+\alpha]}$ denotes the indicator function of $[\alpha, 1 + \alpha]$.

Finally, a simple change of variables gives (45). □

Remark 2. Lemma 1 implies that

$$\operatorname{supp}\phi_u^* \subseteq \left[\underline{\theta}\left(\frac{\overline{\theta}}{\underline{\theta}}\right)^{\alpha}, \overline{\theta}\left(\frac{\overline{\theta}}{\underline{\theta}}\right)^{\alpha}\right]. \qquad (50)$$

This means that for e.g. $\underline{\theta} = 0.5$, $\overline{\theta} = 2$ and three values of ξ, namely $\xi = 0.1, 0.05, 0.02$, we have that the support of ϕ_u^* is contained in $[0.499975, 1.99990]$, $[0.499998, 1.99999]$ and $[0.499999, 1.99999]$, respectively.

Remark 3. Since ξ is usually unknown, it is convenient to find the smallest interval $[\underline{\omega}, \overline{\omega}]$ which contains $\operatorname{supp}\phi_u^*$ for every $\xi \in (0, 1)$. To this end, we determine the minimum and maximum values of the unique positive root of (47) for $\xi \in (0, 1)$. Now, $x = 1$ for $\xi = 0$ and $x = 1/2$ for $\xi = 1$, so, by the continuity of the roots of a polynomials (Horn and Johnson, 1985),

$$\underline{\omega} \le \underline{\theta}\left(\frac{\overline{\theta}}{\underline{\theta}}\right)^{-\frac{1}{k}\ln 2} < \overline{\theta} \le \overline{\omega}\,. \qquad (51)$$

Also, if we denote by $p_\xi(x)$ the left side of (47), then we have that

$$\begin{aligned} p_\xi(x+1/2) &= 2x^3 + 3(\xi^2+1)x^2 + (4\xi^4+\xi^2-1/2)x + \\ &\quad 2\xi^4 - (5/4)\xi^2 - (3/4) \\ p_\xi(x+1) &= 2x^3 + 3(\xi^2+2)x^2 + 4(\xi^4+\xi^2+1)x + 4\xi^4 \end{aligned} \tag{52}$$

The first two higher coefficients of $p_\xi(x+1/2)$ are positive for $\xi \in (0,1)$, and its linear coefficient is negative for $\xi \in (0,1)$ (since it has roots at $\xi = \pm 1, \pm j\sqrt{(6)}/4$), so by Descartes' rule of signs, the positive root of $p_\xi(x)$ is not less than $1/2$ for $\xi \in (0,1)$. Similarly, all coefficients of $p_\xi(x+1)$ are positive, so again by Descartes' rule of signs, the positive root of $p_\xi(x)$ is not greater than 1 for $\xi \in (0,1)$. This implies that

$$[\underline{\omega}, \overline{\omega}] = \left[\underline{\theta} \left(\frac{\overline{\theta}}{\underline{\theta}} \right)^{-\frac{1}{k}\ln 2}, \overline{\theta} \right]. \tag{53}$$

For our resonant system example where $\underline{\theta} = 0.5$ and $\overline{\theta} = 2$ we have that $\operatorname{supp} \phi_u^* \subseteq [2^{-3/2}, 2] \approx [0.3536, 2]$.

To simplify the notation used in establishing the properties of ϕ_u^* and $\widetilde{\phi}_u^*$, let us define

$$\widetilde{f}(u) \triangleq \frac{4e^{k(y-x+\alpha)}[\xi^2 + e^{k(y-x+\alpha)}]}{\{[1-e^{k(y-x+\alpha)}]^2 + 4\xi^2 e^{k(y-x+\alpha)}\}^2}, \tag{54}$$

so (45) can be rewritten as

$$\tilde{\tilde{\phi}}_u^* = \arg \max_{\tilde{\tilde{\phi}}_u \in \mathscr{S}([0,1])} \min_{x \in [0,1]} \int_0^1 \widetilde{f}(y-x) \tilde{\tilde{\phi}}_u(y) dy. \tag{55}$$

Theorem 1 (Existence of an optimal input)
For the problem stated in (55)*, there exists at least one optimal input.*

Proof: By Lemma 1, (55) can be related to a *two-person zero-sum game on the unit square* with *kernel* $\widetilde{F}(x,y) \triangleq \widetilde{f}(y-x)$, such that player x tries to minimise $\widetilde{F}$ 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 $\tilde{\tilde{\phi}}_u^*$, we will use a version of the Minimax Theorem due to (Glicksberg, 1950), which states that if $\widetilde{F}$ is an upper or lower semicontinuous function on $[0,1] \times$

$[0,1]$, then

$$\begin{aligned}
\inf_{\mu_x \in \mathscr{S}([0,1])} \sup_{\mu_y \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x(x)\mu_y(y)\; dy\; dx \\
&= \sup_{\mu_y \in \mathscr{S}([0,1])} \inf_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x(x)\mu_y(y)\; dy\; dx \quad (56)\\
&\triangleq V_m\,,
\end{aligned}$$

where V_m is called the *average value* of the game. Furthermore, if $\widetilde{F}$ is continuous then, by a compactness argument (such as the one given in the paragraph preceding equation (40)), there exist $\mu_x^*,\ \mu_y^* \in \mathscr{S}([0,1])$ such that for every $\mu_x,\ \mu_y \in \mathscr{S}([0,1])$,

$$\begin{aligned}
\int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x^*(x)\mu_y(y)\; dy\; dx &\leq \int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x^*(x)\mu_y^*(y)\; dy\; dx \quad (57)\\
&= V_m \\
&\leq \int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x(x)\mu_y^*(y)\; dy\; dx\,.
\end{aligned}$$

Since in our case $\widetilde{F}$ is continuous, these results directly apply. Furthermore, by (57) and the fact that $[0,1]$ is compact,

$$\begin{aligned}
\int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x^*(x)\mu_y^*(y)\; dy\; dx &= \min_{\mu_x \in \mathscr{S}([0,1])} \int_0^1 \int_0^1 \widetilde{F}(x,y)\mu_x(x)\mu_y^*(y)\; dy\; dx \quad (58)\\
&= \min_{x\in[0,1]} \int_0^1 \widetilde{F}(x,y)\mu_y^*(y)\; dy\,.
\end{aligned}$$

Hence, by (56), (57) and (58),

$$\min_{x\in[0,1]} \int_0^1 \widetilde{F}(x,y)\mu_y^*(y)\; dy = \max_{\mu_y \in \mathscr{S}([0,1])} \min_{x\in[0,1]} \int_0^1 \widetilde{F}(x,y)\mu_y(y)\; dy\,, \quad (59)$$

so if we take $\tilde{\tilde{\phi}}_u^* = \mu_y^*$, we have an optimal solution to (55). □

Theorem 2 (Finite support of the optimal input spectrum)
For the problem stated in (55)*, the optimal input has finite support, i.e.* $\operatorname{supp}\tilde{\tilde{\phi}}_u^*$ *is finite. This implies that the optimal solution of problem* (39) *has finite support as well.*

Proof: This proof is based on (Karlin, 1957), and it is included here for completeness.

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

$$\int_0^1 \widetilde{f}(x-y_0)\mu_x^*(x)dx = V_m\,. \tag{60}$$

By (57), we have that

$$\int_0^1 \widetilde{f}(x-y)\mu_x^*(x)dx \leq V_m, \qquad y \in [0,1]\,. \tag{61}$$

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

$$\int_0^1 \widetilde{f}(x-y)\mu_x^*(x)dx < V_m, \qquad y \in [a,b]\,. \tag{62}$$

Thus, integrating both sides of (61) weighted by $\tilde{\tilde{\phi}}_u^*$, and taking (62) into account, we obtain

$$\int_0^1 \int_0^1 \widetilde{f}(x-y)\mu_x^*(x)\tilde{\tilde{\phi}}_u^*(y)dydx < V_m\,, \tag{63}$$

which contradicts the definition of V_m. This proves (60).

Now, if $\operatorname{supp}\tilde{\tilde{\phi}}_u^*$ were infinite, then (60) would hold for an infinite number of points in a compact interval, so these points would 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 would have that

$$\int_0^1 \widetilde{f}(x-y)\mu_x^*(x)dx = V_m, \qquad y \in \mathbb{R}\,. \tag{64}$$

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

$$\lim_{y\to\infty} \int_0^1 \widetilde{f}(x-y)\mu_x^*(x)dx = 0 \neq V_m \tag{65}$$

which contradicts (64). Thus, $\tilde{\tilde{\phi}}_u^*$ (as well as ϕ_u^*) has finite support. □

7 Conclusions

This chapter has argued that one needs to be careful of so-called "optimal" experiments, especially when crucial pieces of a-priori information are implicit in the definition of "optimality". As an illustration, we have shown that robust optimal input signals can be designed which are insensitive with respect to prior information regarding the nominal parameter values. This means that we can associate the idea of good, bad and optimal to input signal design for system identification. In particular, we have seen that bandlimited '$1/f$' noise appears to have a very good degree of robustness in the face of parametric uncertainty for linear systems. However, one could equally argue that caution needs to be exercised in other areas, e.g. to be robust against possible nonlinearities of the system or regarding the model structure assumptions.

References

Arimoto, S. and 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. and Herzberg, A. (1993). Optimum experiment design for properties of a compartmental model, *Biometrica* **29**: 325–337.

Atkinson, A. and Doner, A. (1992). *Optimum Experiment Design*, Clarendon Press, Oxford.

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

Başar, T. and Olsder, G. (1995). *Dynamic Noncooperative Game Theory.*, 2nd edn, Academic Press, London / New York.

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

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

Bombois, X., Scorletti, G., Gevers, M., Van den Hof, P. and Hildebrand, R. (2005). Least costly identification experiment for control, *Technical report*, CESAME, Louvian-la-Neuve.

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

Chaloner, K. and Verdinelli, I. (1995). Bayesian experiment design: A review, *Statistical Science* .

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

Cox, D. (1958). *Planning of Experiments*, Wiley, New York.

D'Argenio, D. and 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. and Pepelyshev, A. (2003). Standardized maximin e-optimal designs for the michaelis-menten model, *Statistica Sinica* **13**(4): 1147–1163.

Dickson, L. E. (1914). *Elementary Theory of Equations*, John Wiley & Sons, New York.

El-Gamal, M. and 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. and Silvey, S. (1980). A sequentially constructed design for estimating a nonlinear parametric function, *Biometrika* **67**(2): 381–388.

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

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

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. and Bombois, X. (2006). Input design: From open-loop to control-oriented design, *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. C., Rojas, C. R., Welsh, J. S. and Feuer, A. (2006). Robust optimal experiment design for system identification, *Automatica (under revision)* .

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

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

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

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

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

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

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*, Cambridge University Press, Cambridge.

Karlin, S. (1957). On games described by bell shaped kernels, *Contributions to the Theory of Games, vol. III*, number 39 in *Annals of Mathematics Studies*, Princeton University Press, New Jersey, pp. 365–391.

Karlin, S. and 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. and 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* J. Eisenfeld and C. DeLisi (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.

Ljung, L. (1999). *System Identification: Theory for the User*, 2nd edn, Prentice Hall, Upper Saddle River, New Jersey.

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

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

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

Rosenbrock, H. H. and McMorran, P. D. (1971). Good, bad, or optimal?, *IEEE Transactions on Automatic Control* **16**(6): 552–554.

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

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

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

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

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

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.