Kernels and designs

Hugo Maruri-Aguilar and Henry Wynn
Department of Statistics, London School of Economics
Houghton Street WC2A 2AE London (UK), H.Maruri-Aguilar@lse.ac.uk

May 7, 2008

Abstract

This paper comprises a revision of literature relating designs and kernels. Polynomial, Fourier (trigonometric) and Haar wavelet kernels are studied in conjunction with D-optimal designs. New results are presented that link Sobol' sequences with a class of Haar models, and a general result is also established for arbitrary Haar models.

Keywords: Kernels, polynomial, Haar wavelets, Fourier series, low discrepancy, D-optimality

1 Introduction

In this report the main interest is the optimality relation between linear models constructed with a family of functions (kernels) and designs. The review presents results using classical optimality theory, that is, the model is composed by a linear combination of functions in the kernel class and thus the model is linear in the parameters. The kernels considered in this report are polynomials, trigonometric functions and Haar wavelets.

Section 2 presents the basic concepts of optimal design. Then in Section 3 the material concerning polynomial models is reviewed. Part of this report presents partial results from work in progress. The material linking optimality and Haar models in Section 4 contains new results and it is aimed to be part of a paper involving low discrepancy sequences. For reasons of clarity, most proofs are omitted. In Section 5 the material concerning trigonometric functions is reviewed. Finally, Section 6 includes some comments towards the future use of this material in the context of Gaussian processes. An Appendix summarizes orthonormality properties of Haar wavelets.

2 Designs, prediction variance and optimality

Along most of this material the designs considered are sets of n points, i.e. a design is $D_n = \{x_1, \ldots, x_n\}$ with $x_i \in [0,1]^d$ for $i = 1, \ldots, n$. No replications are considered.

However, for purposes of design optimality it is often useful to refer to design measures. Design measures have a long story of use within optimal design of experiments, stemming from the seminal work by Kiefer (1959) and Kiefer and Wolfowitz (1959). A design measure ξ is a probability measure with finite support defined over a set of points. That is, a design measure is a collection of points and weights $\xi = \left\{ \begin{smallmatrix} x_1 & \cdots & x_n \\ w_1 & & w_n \end{smallmatrix} \right\}$ with $0 \leq w_i \leq 1$. The weights satisfy $\sum_{i=1}^{n} w_i = 1$, which for measures with non finite support becomes $\int \xi(dx) = 1$

Here we only recall the basic optimality results, referring to the classical text by Silvey (1980). Consider a linear model

$$y_i = \sum_{j=1}^{p} \theta_j f_j(x_i) + \varepsilon_i = f(x_i)^T \Theta + \varepsilon_i,$$

where $f(x_i)^T = (f_1(x_i), \ldots, f_p(x_i))$ and $\Theta^T = (\theta_1, \ldots, \theta_p)$ and the errors ε_i are assumed to be independently distributed with zero mean and variance σ^2. The model coefficients are estimated by least squares, and the predictive model is

$$\hat{y}(x) = \sum_{j=1}^{p} \hat{\theta}_j f_j(x_i) = f(x_i)^T \hat{\Theta},$$

where $\hat{\Theta}$ is the vector of estimated parameters. The prediction variance is

$$V(\hat{y}(x)) = \sigma^2 f(x)^T \left(\frac{X^T X}{n} \right)^{-1} f(x), \tag{1}$$

where X is the design-model matrix and n is the number of design points. A design D_n^* is D-optimal if it maximises the determinant of the information matrix $X^T X$. We recall that, for design measures, the information matrix is defined as $M(\xi) = \int f(x) f(x)^T \xi(dx)$ and that $V(\hat{y}(x)) = \frac{\sigma^2}{n} d(x, \xi)$ with $d(x, \xi) = f(x) M(\xi)^{-1} f(x)^T$. A design is G-optimal if it minimises the maximum prediction variance over the design region $[0,1]^d$. In both D- and G-optimality, the minimisation is carried along over the set of all designs with n points in $[0,1]^d$.

The general equivalence theorem by Kiefer and Wolfowitz states that when a design measure is D-optimal then it is also G-optimal, and the

value attained equals the number of parameters is the model. By the above theorem, a design is easily checked for optimality if

$$\max \frac{n}{\sigma^2} V(\hat{y}(x)) = \max d(x, \xi) = p.$$

3 Polynomial regression models

Optimal designs for Polynomial models have been studied for a long time. The literature is quite extensive, stemming from the early paper by Smith (1918) and going through the classical papers with design measures, see Kiefer (1959), Kiefer and Wolfowitz (1959). The algorithm by Wynn (1972) and the contributions by Atkinson and Cox (1974) and Atkinson and Fedorov (1975) expanded further the field towards construction of designs and discrimination between models. More recently, the theory of canonical moments and its extensions, see Dette and Studden (1997), add another branch to the thriving research in optimal design theory. See the book by Atkinson et al. (2007) for a comprehensive review of the topic.

m	Support points
1	$-1, 1$
2	$-1, 0, 1$
3	$-1, -0.4472, 0.4472, 1$
4	$-1, -0.6547, 0, 0.6547, 1$
5	$-1, -0.7651, -0.2852, 0.2852, 0.7651, 1$
6	$-1, -0.8302, -0.4688, 0, 0.4688, 0.8302, 1$

Table 1: D-optimal designs for polynomial models

A hierarchical polynomial model of degree m in one variable is the model spanned by the terms $1, x, x^2, \ldots, x^m$. D-optimal designs for hierarchical polynomial models are computed using Legendre polynomials. The support points are given by the roots of the equation $(1-x^2)P'_d(x) = 0$, where $P_d(x)$ is the Legendre polynomial defined by the recurrence relation

$$(d+1)P_{d+1}(x) = (2d+1)xP_d(x) - dP_{d-1}(x)$$

with $P_0(x) = 1$ and $P_1(x) = x$. The optimal designs are uniform measures, i.e. the design weights are $w_i = \frac{1}{m+1}$. Table 1 gives results for univariate models with design region $[-1, 1]$.

It is always possible to construct optimal designs for multivariate polynomial models, although only special low order models have been developed in the literature. Optimal designs have been proposed for second order multivariate polynomial model of the type

$$g(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{i=1}^{d} \sum_{j=i}^{d} \beta_{ij} x_i x_j.$$

The optimal designs proposed in Atkinson and Donev (1992) for the above model are composed of factorial 2^d points plus star runs and centre points. D-optimal designs for multivariate polynomial models can be easily constructed with the aid of canonical moments, although they are not of minimal size, see Dette and Studden (1997) for details.

3.1 Smooth supersaturated models

Polynomial models have well known oscillation properties, which are not desired if the model is to be used for interpolation. To avoid this problem, smooth supersaturated models have been proposed. They are polynomial models in which extra terms are added while at the same time minimising a measure of smoothness.

The recent work on supersaturated models in Maruri-Aguilar and Wynn (2008) can be used to construct optimal designs. The basic premise is that a the fitted polynomial $\hat{y}(x) = \sum \hat{\theta}_\alpha x^\alpha$ is a linear function of the observations, i.e. $\hat{y}(x) = \sum_{i=1}^{n} c_i y_i$, where the coefficients c_i are smooth polynomials that depend on x. Ongoing work aims to use c_i as smooth kernels as basis for design.

4 Haar regression models

Haar wavelets are a set of orthogonal functions constructed by scaling and shifting of a *mother wavelet*, which is a step function. The mother wavelet is defined as

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Define $\psi_{j,k}(x) = \psi\left(2^k x - j\right)\sqrt{2^k}$ for $k = 0, 1, 2, \ldots$ and $j = 0, 1, \ldots, 2^k - 1$. For a wavelet $\psi_{j,k}(x)$, the interval $\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right)$ is the region on which $\psi_{j,k}(x)$

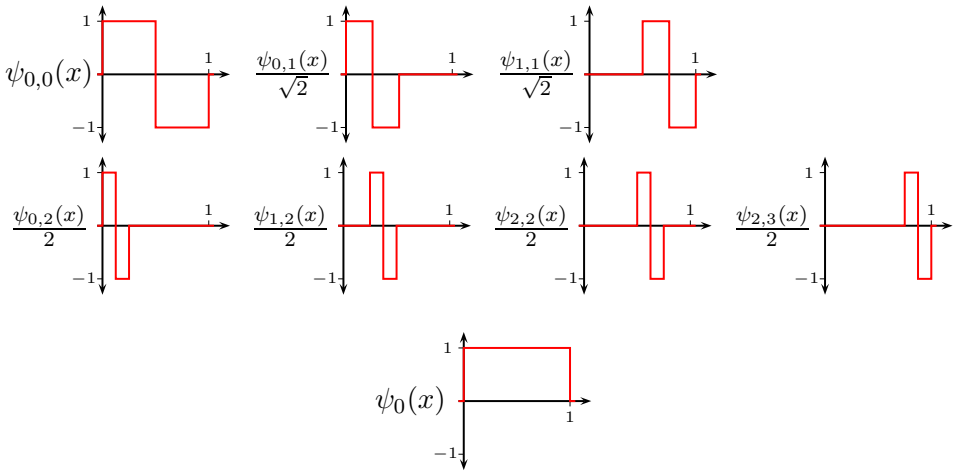

Figure 1: Examples of wavelets. The constant function $\psi_0(x)$ is also shown. In all cases the horizontal axis corresponds to x.

takes non-zero value and it is referred to as the support of $\psi_{j,k}(x)$. See examples of wavelets in Figure 1.

The constant function taking value of one over $[0, 1)$ is not a wavelet, but will be used later in the analysis. We refer to it as $\psi_0(x)$. Let Ψ be the set of all wavelets and the constant function, i.e. define

$$\begin{aligned} \Psi &= \left\{\psi_{j,k}(x) : k = 0, 1, \ldots, j = 0, 1, \ldots, 2^k - 1\right\} \cup \psi_0(x) \\ &= \{\psi_0(x), \psi_{0,0}(x), \psi_{0,1}(x), \psi_{1,1}(x), \psi_{0,2}(x), \psi_{1,2}(x), \ldots\} \end{aligned} \tag{2}$$

4.1 Univariate Haar models

The study of Haar models starts by considering certain subsets of Ψ which are termed "well formed sets of wavelets".

Definition 1 *Let A be a finite subset of $\Psi \setminus \psi_0(x)$. We say that A is a* well formed *set of wavelets when if the wavelet $\psi_{j,k}(x)$ belongs to A then all wavelets $\psi_{i,k}(x)$ are also in A for $i = 0, \ldots, 2^k - 1$, $i \neq j$. The set $G(A) = \{k : \psi_{j,k}(x) \in A\}$ is called the set of generators of A.*

For a well formed set of wavelets A, its cardinality is given by

$$|A| = \sum_{k \in G(A)} 2^k.$$

Its cardinality obbeys the bound $|A| \leq 2^{1+\max G(A)} - 1$.

In general, a well formed set of wavelets is uniquely related to its generator set. That is, a finite set of non-negative integers uniquely defines a well formed set of wavelets, i.e. for $B \subset \mathbb{Z}_{\geq 0}$, B finite, define

$$W(B) = \{\psi_{j,k}(x) : k \in B, j = 0, \ldots, 2^k - 1\}.$$

The set $W(B)$ is a well formed set of wavelets that also satisfies the following contention $W(B) \subset \Psi \backslash \psi_0(x)$. Clearly, the set of generators of $W(B)$ satisfies $G(W(B)) = B$.

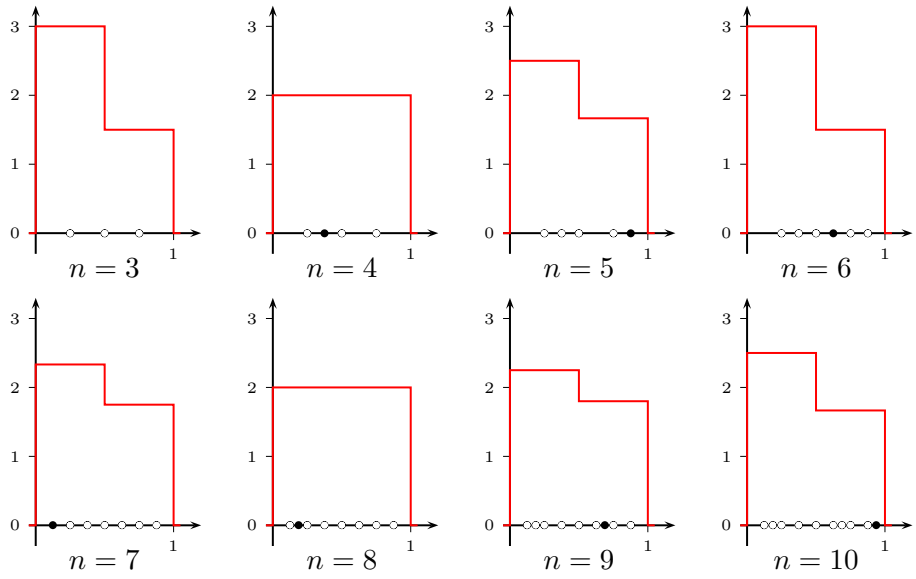

Figure 2: Variance of prediction against x for different sample sizes and using model A_1, see Example 2. The Sobol' design points are shown in each case, with the last point added in solid black.

The following three instances of well formed sets of wavelets will be used in studying optimality for Haar models.

Definition 2 *For $m > 0$ define A_m to be the well formed set of wavelets with generator set $\{m\}$.*

Definition 3 *For $m > 0$, let H_m be the* well formed *set of wavelets with generator set $\{0, \ldots, m\}$. We refer to H_m a* hierarchical *set of wavelets.*

The set A_m has elements $A_m = \{\psi_{0,m}(x), \ldots, \psi_{2^m-1,m}(x)\}$ and its cardinality is $|A_m| = 2^m$, i.e. A_m contains 2^m wavelets. A hierarchical set of wavelets H_m can be written as $H_m = \bigcup_{j=0}^{m} A_j$. The hierarchical set H_m has cardinality

$$|H_m| = \sum_{k=0}^{m} 2^k = 2^{m+1} - 1.$$

Definition 4 *Let I_m be a well formed set of wavelets that satisfies the strict contention*

$$A_m \subsetneq I_m \subsetneq H_m, \tag{3}$$

where $m = \max G(I_m)$.

Note that although the notation A_m and H_m refers to a single set, there can be several well formed sets I_m that satisfy Definition 4. This distinction has no effect in the following analysis as they all share the optimality properties we study and we refer generically by I_m to any of them. The cardinality of I_m is computed as $|I_m| = \sum_{k \in G(I_m)} 2^k$. As a result of the strict contention above, $|I_m|$ satisfies the strict inequality $|A_m| = 2^m < |I_m| < 2^{m+1} - 1 = |H_m|$.

Example 1 Let $m = 2$, then $A_2 = \{\psi_{0,2}(x), \psi_{1,2}(x), \psi_{2,2}(x), \psi_{3,2}(x)\}$ and $H_2 = A_2 \cup \{\psi_{0,0}(x), \psi_{0,1}(x), \psi_{1,1}(x)\}$. As an example of non-unicity of I_2, the sets $A_2 \cup \{\psi_{0,0}(x)\}$ and $A_2 \cup \{\psi_{0,1}(x), \psi_{1,1}(x)\}$ are well formed sets that satisfy Equation (3) and thus both belong to the class I_2. In other words, the generators for A_2 and H_2 are $\{2\}$ and $\{0, 1, 2\}$, while for intermediate well formed sets I_2, the options for generators are $\{0, 2\}$ and $\{1, 2\}$.

4.2 Sobol' sequences and Haar models

Sobol' (1967) proposed a sequence of points with good space filling properties (low discrepancy). The points in a Sobol' sequence are constructed using binary arithmetic operations. An implementation of Sobol' sequences is available in the computer language `R`, see R Development Core Team (2006) and Würtz (2008); for a description of the Sobol' algorithm used in the software implementation see Bratley and Fox (1988).

In the following two sections, the D-optimality relation between Sobol' designs and Haar regression models is studied. Sobol' sequences have optimality properties for well formed Haar regression models spanned by the sets A_m, I_m or H_m. A special consideration is made as well formed sets

Haar model		Design size			
Terms	Size	Sobol'		Sobol' plus origin	
		n_0	n^*	n_0	n^*
A_m	2^m	$2^{m+1}-1$	$2^{m+1}k$	2^m	$2^m k$
I_m	$\lvert I_m \rvert$	$2^{m+1}-1$	$2^{m+2}k$	$\lvert I_m \rvert$	$2^{m+1}k$
H_m	$2^{m+1}-1$	$2^{m+1}-1$	$2^{m+2}k$	$2^{m+1}-1$	$2^{m+1}k$
$A_m \cup \psi_0(x)$	2^m+1	$2^{m+1}-1$	$2^{m+2}k$	2^m+1	$2^{m+1}k$
$I_m \cup \psi_0(x)$	$\lvert I_m \rvert + 1$	$2^{m+1}-1$	$2^{m+2}k$	$\lvert I_m \rvert + 1$	$2^{m+1}k$
$H_m \cup \psi_0(x)$	2^{m+1}	$2^{m+2}-1$	$2^{m+2}k$	2^{m+1}	$2^{m+1}k$

Table 2: Optimality results for well formed Haar models.

of wavelets do not include the constant term $\psi_0(x)$. For this reason a distinction can be made between well formed Haar models with or without intercept term. The case without intercept is discussed in Section 4.4, while models with intercept are presented in Section 4.5. The next subsection summarizes the results of Sections 4.4 and 4.5.

4.3 Summary of optimality results

Table 2 reports the results obtained analysing Haar wavelet models using pure Sobol' designs and Sobol' designs with the origin added. The main result reported is the sample size n^* for which Sobol' designs are D-optimal for a series of values. This value n^* is expressed as a multiplier times an integer value k, i.e. the smallest design size for optimality is when $k = 1$. Also in Table 2 the minimal design size to identify model n_0 is presented.

Two important features of Table 2 must be stressed:

1. adding the origin to a Sobol' sequence makes identifiability of the model easier, e.g. in every row, the values of n_0 in the fifth column are strictly less than those in the third column, and

2. the frequency in which the designs are optimal is increased by a factor of two by adding the origin, i.e. the multiplier in the fifth column is half that for the third column. In particular, the minimal design size for optimality is reduced by adding the origin to half of what is required without the origin, in all cases considered.

We also refer to results previously available in the literature. Herzberg and Traves (1994) proposes optimal design for what we term as hierarchical

model with intercept $H_m \cup \psi_0(x)$ and in Tian and Herzberg (2007) they compute minimax designs using D- and A-optimality criteria for the same hierarchical model. The designs they find are uniformly spaced.

4.4 Optimality results for Haar models without intercept

The first result relates unscrambled Sobol' sequences with well formed sets of wavelets A_m.

Theorem 1 *Let the model be constructed with the set A_m, for $m > 0$; and let D_n be the first n points of the unscrambled Sobol' sequence. Then*

1. *the smallest sample size such that the model spanned by A_m is identifiable by D_n is $n_0 = 2^{m+1} - 1$; and*
2. *for $k = 1, \ldots,$ the design D_n is D-optimal for the model spanned by A_m when $n^* = 2^{m+1}k$.*

Example 2 Consider the well formed set $A_1 = \{\psi_{0,1}(x), \psi_{1,1}(x)\}$. For sample sizes $n^* = 2^2, 2(2^2), 3(2^2), \ldots = 4, 8, 12, \ldots$, the unscrambled Sobol' sequence is D-optimal for the Haar model spanned by A_1. In Figure 2 the variance of prediction is depicted for this model and sample sizes $n = 3, \ldots, 10$. In that figure, optimality can be seen when the maximum prediction variance over $[0, 1)$ equals the number of parameters, that is two. As another example, consider the regression model constructed with the well formed set $A_3 = \{\psi_{0,3}(x), \psi_{1,3}(x), \ldots, \psi_{7,3}(x)\}$. Unscrambled Sobol' sequences are optimal for sample sizes $n^* = 2^4, 2(2^4), 3(2^4), \ldots = 16, 32, 48, \ldots$ Figure 3 shows a graph of the maximum prediction variance against sample size, and this recurrent optimality behaviour of unscrambled Sobol' sequence becomes apparent.

Example 3 The use of scrambled Sobol' sequences may not neccesarily lead to optimal configurations with the same frequency as in Example 2. Consider a Faure-Tezuka scrambled Sobol' sequence and the same model A_3 of Example 2. D-optimality is achieved at sample sizes $n^* = 32, 152, 184, 280, \ldots$ In both unscrambled and scrambled Sobol', the maximum variance of prediction gets close to the number of parameters as sample size increases and thus we say the the designs are asymptotically optimal, that is $\lim_{n\to\infty} \max V(\hat{y}(x)) = 8$, where $V(\hat{y}(x))$ is evaluated in $[0, 1)$, see Figures 3 and 4.

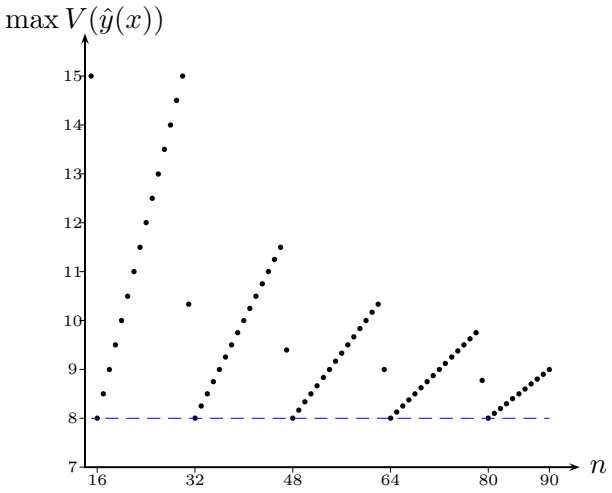

Figure 3: D-optimality of Sobol' unscrambled sequence with the model A_3 of Example 2.

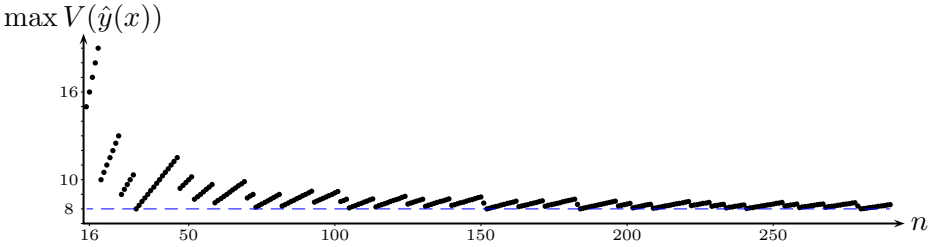

Figure 4: Asymptotic D-optimality of scrambled Sobol' sequence with the model A_3 of Example 3.

The other extreme of well formed sets of wavelets occurs when the model is constructed using the set H_m. The following result extends Theorem 1, while Theorem 3 covers all intermediate well formed sets of wavelets between A_m and H_m.

Theorem 2 *Let H_m be a hierarchical set of wavelets with $m > 0$; let D_n be the first n points of the unscrambled Sobol' sequence. Then*

1. *the smallest sample size such that the model spanned by H_p is identifiable by D_n is $n_0 = 2^{m+1} - 1$; and*
2. *for $k = 1, \ldots,$ the design D_n is D-optimal for the model spanned by H_m when $n^* = 2^{m+2}k$.*

Theorem 3 *Let I_m be a well formed set of wavelets. Let D_n be the first n points of an unscrambled Sobol' sequence. Then the two statements of Theorem 2 hold for the model spanned by I_m.*

Example 4 Consider $H_2 = \{\psi_{0,0}(x), \psi_{0,1}(x), \psi_{1,1}(x), \psi_{0,2}(x), \psi_{1,2}(x), \psi_{2,2}(x), \psi_{3,2}(x)\}$. For sample sizes $n^* = 16, 32, 48, \ldots$, the unscrambled Sobol' design D_n is D-optimal for the model spanned by H_2.

Other low discrepancy sequences can lead as well to D-optimal results for well formed Haar models. This fact is further explored in relation to Haar models in Section 4.7. Here a simple example is given.

Example 5 A univariate Weyl sequence is $\{i\sqrt{2} \mod 1 : i = 1, 2, \ldots\}$, which is uniformly distributed over $[0, 1]$. Take D_n as the first n points of the Weyl sequence, together with the univariate model spanned by H_1. The Weyl design D_n is optimal for the model H_1 for design sizes

$$n^* = 4, 12, 16, 24, 26, 28, 70, 72, 74, 82, 84, 86, 94, 96, 98, \ldots$$

and $n_0 = 3$.

4.5 Optimality results for Haar models with intercept

We initially consider adding the intercept to models with a single generator, that to models spanned by the well formed set of wavelets A_m with $m > 0$, together with $\psi_0(x)$. We recall that $A_m = \{\psi_{0,m}(x), \ldots, \psi_{2^m-1,m}(x)\}$.

Theorem 4 *Let the model be spanned by $A_m \cup \psi_0(x)$, for $m > 0$; that is, by adding the intercept to the well formed set A_m; let D_n be the first n points from the unscrambled Sobol' sequence. Then*

1. *the smallest sample size such that the model is identifiable by D_n is $n_0 = 2^{m+1} - 1$; and*
2. *for $k = 1, \ldots,$ the design D_n is D-optimal for the model when sample size is $n^* = 2^{m+2}k$.*

A natural idea is to add the origin to the Sobol' design, while considering the same model as in Theorem 4. This has the effect of improving the optimality performance by a factor of two.

Corollary 5 *Let the model be by $A_m \cup \psi_0(x)$, for $m > 0$; let the design D_n be constructed by adding the origin to the first $n - 1$ points from the unscrambled Sobol' sequence. Then*

1. *the smallest sample size such that the model is identifiable by D_n is $n_0 = 2^m + 1$; and*

2. *for $k = 1, \ldots$, the design D_n is D-optimal for the model when sample size is $n^* = 2^{m+1}k$.*

Example 6 The effect of adding the origin to the design greatly improves the performance of the Sobol' design for a Haar model with intercept. Consider the Haar model spanned by $A_2 \cup \psi_0(x)$. D-optimality is achieved for $n^* = 2^4, 2(2^4), 3(2^4), \ldots = 16, 32, 48, \ldots$ when using an unscrambled Sobol' sequence. For the same model, when the origin is added, D-optimality is achieved for $n^* = 2^3, 2(2^3), 2(2^4), \ldots = 8, 16, 24, \ldots$ In Figure 5 the prediction variances for the two cases are plotted.

We now extend the analysis to hierarchical Haar models with added intercept, i.e. models spanned by the set $H_m \cup \psi_0(x)$. Such a model has 2^{m+1} terms.

Theorem 6 *Let the model be $H_m \cup \psi_0(x)$, for $m > 0$; let D_n be the first n points of the unscrambled Sobol' sequence. Then*

1. *the smallest sample size such that the model is identifiable by D_n is $n_0 = 2^{m+2} - 1$; and*

2. *for $k = 1, \ldots$, the design D_n is D-optimal for the model when $n^* = 2^{m+2}k$.*

The previous result points to the relative robustness of Sobol' sequences when adding the intercept to a hierarchical set of wavelets, i.e. the design remains optimal at the same sample sizes as in Theorem 2. The following result shows the considerable improvement obtained when adding the origin to the unscrambled Sobol' sequence.

Corollary 7 *Let the model be $H_m \cup \psi_0(x)$, for $m > 0$; let the design D_n be constructed by adding the origin to the first $n - 1$ points of the unscrambled Sobol' sequence. Then*

1. *The smallest sample size that identifies the model is $n_0 = 2^m + 1$; and*

2. *for $k = 1, \ldots$, the design D_n is D-optimal for the model when $n^* = 2^{m+1}k$.*

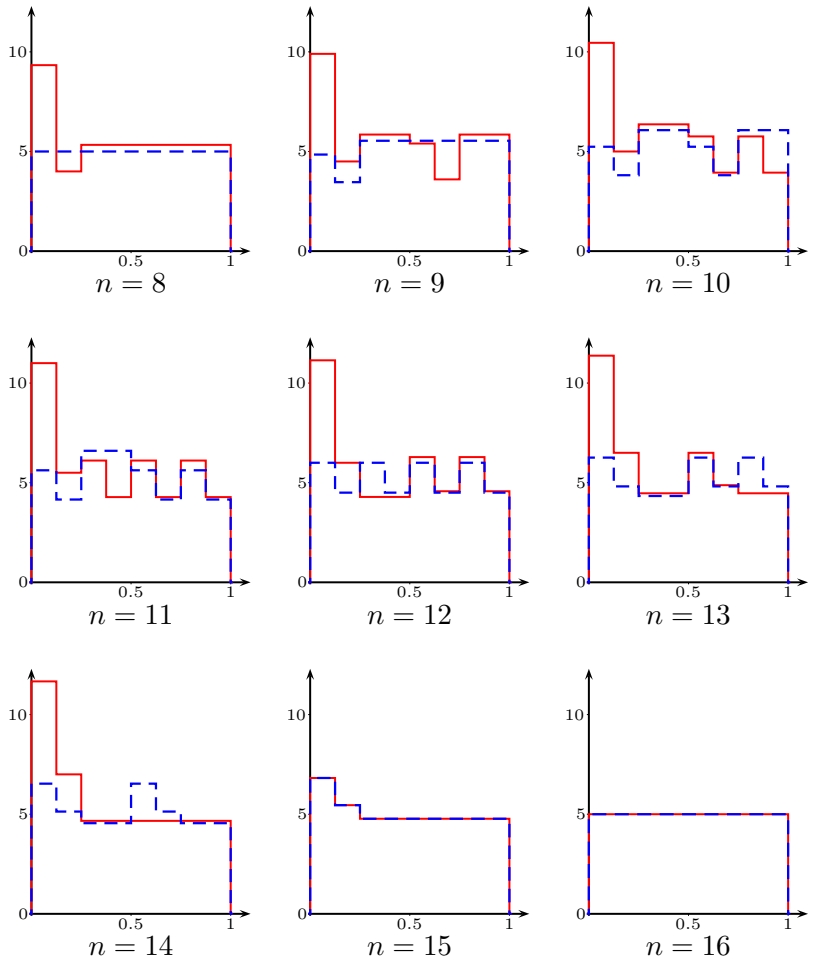

Figure 5: Variance of prediction for Haar model with intercept. The red line is for the Sobol' design, while the blue (dashed) curve is for the Sobol' design with the origin added, see Example 6.

Finally, we consider well formed Haar models together with an intercept term.

Theorem 8 *Let the model be spanned by $I_m \cup \psi_0(x)$ with $m > 0$; let D_n be the first n points of an unscrambled Sobol' sequence. Then*

1. *the smallest sample size that identifies the mode is $n_0 = 2^{m+1} - 1$*
2. *for $k = 1, \ldots,$ the design D_n is D-optimal for the model when $n^* = 2^{m+2}k$.*

As previously stated, robustness of the Sobol' sequence is greatly improved when the origin is added.

Corollary 9 *Let the model be $I_m \cup \psi_0(x)$ with $m > 0$ and let D_n be constructed by adding the origin to the first $n - 1$ points of the unscrambled Sobol' sequence. Then*

1. *the smallest sample size that identifies the model is $n = |I_m| + 1$; and*
2. *for $k = 1, \ldots,$ the design is D-optimal when $n^* = 2^{m+1}k$.*

The results of identifiability and optimality for unidimensional Haar models are summarized in Table 2. See Section 4.3 for a summary and comments on the results.

4.6 Multivariate Haar models

The results for univariate Haar regression models can be extended to linear multivariate Haar regression models. Ongoing promising research points to the existance of similar results to those given in Table 2 in Section 4.1. The following example shows similar optimality results as in Sections 4.4 and 4.5.

Example 7 Let $M_i = \{\psi_0(x_i), \psi_{0,0}(x_i)\} = \{1, \psi_{0,0}(x_i)\}$ for $i = 1, 2$, that is, the constant term and mother wavelet in x_i. A multivariate Haar regression is constructed with the terms in the (Kronecker) product

$$M = M_1 \times M_2 = \{1, \psi_{0,0}(x_1), \psi_{0,0}(x_2), \psi_{0,0}(x_1)\psi_{0,0}(x_2)\}$$

This bivariate model consists of the intercept term, one mother wavelet in each direction x_i and their interaction. Consider D_n as the first n points of a bidimensional unscrambled Sobol' sequence. It turns that the minimal size

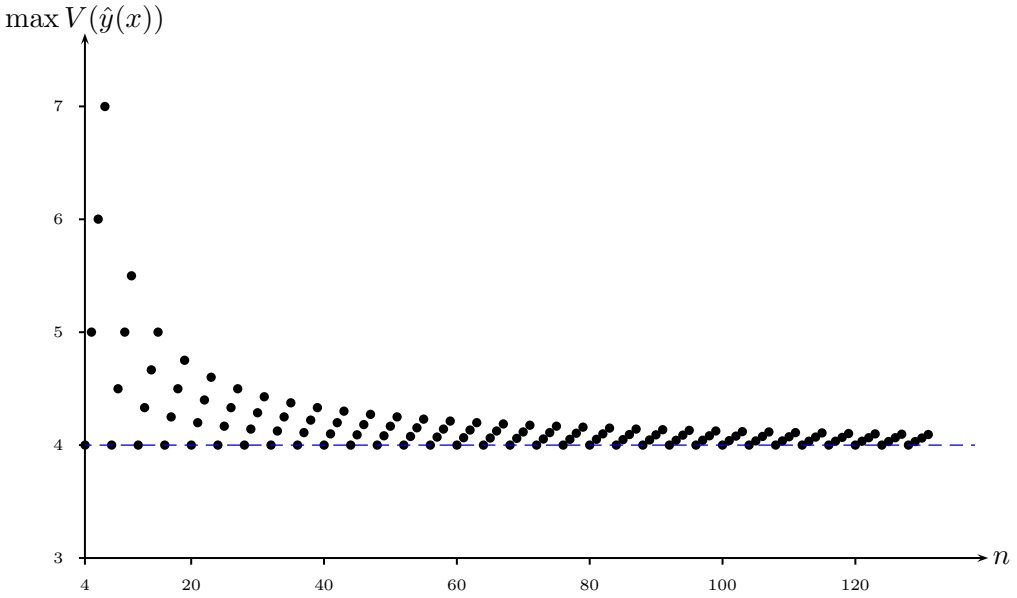

Figure 6: Optimality for the bidimensional Haar model of Example 7.

to identify the model is $n_0 = 4$, while the Sobol' configuration is D-optimal for sizes $n^* = 4, 8, 12, \ldots$, see Figure 6. As in previous univariate examples, the design is also asymptotically D-optimal. An interesting feature of this example is that adding the origin to the Sobol' design gives the same results obtained without it.

Example 8 Consider adding the terms $\psi_{0,1}(x_1)$ and $\psi_{1,1}(x_1)$ to the model in example 7. Then an unscrambled bidimensional Sobol' sequence with minimal sample size of $n_0 = 7$ is required to identify it; and it is D-optimal for sample sizes $n^* = 8, 16, 24, \ldots$ Adding the origin to the Sobol' design decreases n_0 to 6, while mantaining n^* unchanged.

The bivariate model of Example 7 is defined as a product model, and in that sense it mirrors the polynomial model $\{1, x_1, x_2, x_1x_2\}$, which has a classical optimal design in the 2^2 factorial design. An interesting future research direction is to find out to what extent Sobol' sequences for multivariate Haar models parallel the role which classical factorial designs have for product polynomial models.

4.7 Extension to arbitrary Haar models

All the optimality results presented are referred to well formed Haar models, for which the basic building block is the set of wavelets A_j. Haar regression models can be constructed for non well formed sets of wavelets, but then finding optimal designs will not necessarily lead to space filling designs or Sobol' sequences.

Example 9 Consider the univariate Haar model with terms $\psi_{1,1}(x)$,$\psi_{2,2}(x)$ and $\psi_{3,2}(x)$. We find that Sobol' sequences are not anymore D-optimal for this model. In fact, an unscrambled Sobol' sequence is not even asymptotically optimal, as inspection of the results show that $\lim_{n\to\infty} \max V(\hat{y}(x)) = 6$.

However, close study to the behaviour of Example 9 suggests a general link between the support of a wavelet model and the D-optimal design. We have the following theorem for a unidimensional wavelet model.

Theorem 10 *Let $M \subset \Psi$ be a finite set of wavelets and let the model be spanned by M; let*

$$g_M(x) = \sum_{\psi_{j,k}\in\Psi} \frac{\psi_{j,k}(x)^2}{2^k},\ f_M(x) = \frac{g_M(x)}{\int_0^1 g_M(x)dx}$$

and call $f_M(x)$ the model measure. *Then*

1. *the model measure is a probability density over $[0,1)$, and*
2. *the model measure is a D-optimal design measure for the model M.*

All well formed models considered in Sections 4.4 and 4.5 have uniform model measures over $[0,1)$ and this would explain recurrent optimality results obtained with Sobol' sequences, that is the Sobol' sequence define a uniform measure for such sample sizes. The binary construction of Sobol' sequences makes them specially suitable for well formed Haar models. However, other uniform measures are expected to yield D-optimal designs for well formed models, see the univariate Example 5 and the following multivariate example.

Example 10 Consider the model M of Example 7, for which the bivariate model measure $f_M(x_1, x_2)$ is a uniform model measure. Take a Weyl sequence with prime generators $2, 3$, i.e. $D_n = \{(i\sqrt{2} \mod 1, i\sqrt{3} \mod 1) :$

$i = 1, \ldots, n\}$. The Weyl sequence is D-optimal for sizes $n^* = 4, 12, 28$. For bigger sample sizes, D_n is not longer D-optimal, although asymptotical optimality is observed.

The most important contribution of model measures is for non well formed Haar models. In this case space filling designs cannot be expected to be optimal, and indeed the model measure indicates which regions are to be sampled and the frequency of sampling.

Example 11 (Continuation of Example 9) The maximum variance of prediction obtained with the Sobol' design indicates unnecessary sampling over $[0, \frac{1}{2})$ and correspondingly undersampling of the interval $[\frac{1}{2}, 1)$. Turning our attention to the model measure for the non well formed model in Example 9, it occurs that $f_m(x)$ is the uniform distribution over $[\frac{1}{2}, 1)$. This suggest to instead transform the unscrambled Sobol' design to lie in $[\frac{1}{2}, 1)$. This transformation yields D-optimal results for $n^* = 8, 16, 24, \ldots$ with $n_0 = 3$.

4.7.1 Optimality for Haar series

Theorem 10 allows for the construction of D-optimal designs for finite Haar models with no restrictions on the wavelets included. An important case would be to relax the assumption of M finite to allow design for infinite Haar series, i.e. the model spanned by Ψ.

Recall that the set Ψ constitutes a set of independent orthonormal functions over $[0, 1)$, see Appendix 7.1. For this reason any integrable function $\eta(x) : [0, 1) \mapsto \mathbb{R}$ can be represented (interpolated) with a series of wavelet functions, that is

$$\eta(x) = \sum_{\psi_\alpha \in \Psi} c_\alpha \psi_\alpha(x),$$

where the coefficients are $c_\alpha = \langle \eta(x), \psi_\alpha(x) \rangle$. Remarkably, the properties of Ψ mirror those of the system of trigonometric functions, i.e. Fourier analysis. Similarly to the Fourier transform, a wavelet transform can be defined, see Strang (1993) for comparisons between both Fourier and wavelet transform. Ongoing research is extending Theorem 10 in order to optimally design for a function $\eta(x)$ approximated using Haar series.

5 Fourier regression models

The trigonometric functions $\cos(2\pi x), \cos(4\pi x), \ldots, \sin(2\pi x), \sin(4\pi x), \ldots,$ together with the constant function form a set of orthonormal functions over

$[0,1)$ and thus they can be used to represent continuous functions with finite energy over $[0,1)$. This is the basis of the classical Fourier analysis.

In this report the interest lies in Fourier regression models and optimal designs associated to them. The main results are that for a given Fourier regression model, a D-optimal design can be constructed using lattice designs. The main references used are Hoel (1965), Riccomagno et al. (1997) and Bates et al. (1998).

5.1 Unidimensional Fourier model

Consider the univariate Fourier model of order m on $[0,1)$:

$$E(Y(x)) = \theta_0 + \sqrt{2}\sum_{k=1}^{m}\left(\theta_k \sin(2\pi kx) + \theta_{k+m}\cos(2\pi kx)\right) \tag{4}$$

A design consisting of at least $2m+1$ equidistant points on $[0,1)$ is D-optimal for the model in Equation (4), see Riccomagno et al. (1997). If the design has exactly $2m+1$ points then it is optimal and of minimal size. The designs considered for the above result are uniform measures.

5.2 Multivariate Fourier models

A simple multivariate Fourier regression model can be constructed with the (Kronecker) product of univariate Fourier models. Then finding an optimal design for the product model is reduced to the problem of finding designs for each marginal model and then computing the product of them, see Hoel (1965) and Schwabe (1996a).

Theorem 11 *(Schwabe (1996a)) Let the design measure ξ_i be D-optimal for the model with support vector $f_i(x_i)$, where $f_i(x_i)$ contains a constant term and $i = 1,\ldots,d$. Then the product measure $\xi_1 \times \cdots \times \xi_d$ is optimal for the model with support $f_1(x_1) \times \cdots \times f_i(x_i)$.*

When each marginal model $f_i(x_i)$ in Theorem 11 is a Fourier regression model, then the restiction of adding a constant term is not necessary, see Schwabe (1996b). Generally, product designs are of large size, and unless the model has a product form, product designs are not preferred and alternatives such a lattice designs must be considered.

5.3 Lattice D-optimal designs for Fourier models

For certain types of multivariate Fourier models, D-optimal lattices can be constructed. In what follows we review the material by Riccomagno et al. (1997) and Bates et al. (1998). We recall the following definition of a lattice design.

Definition 5 *The n point lattice design generated by $(g_1, \ldots, g_d) \in \mathbb{Z}^d$ is the uniform design supported on the grid*

$$\left(\frac{j}{n} g_1 \bmod 1, \ldots, \frac{j}{n} g_d \bmod 1\right) = \left(\frac{j g_1 \bmod n}{n}, \ldots, \frac{j g_d \bmod n}{n}\right)$$

for $j = 0, \ldots, n-1$.

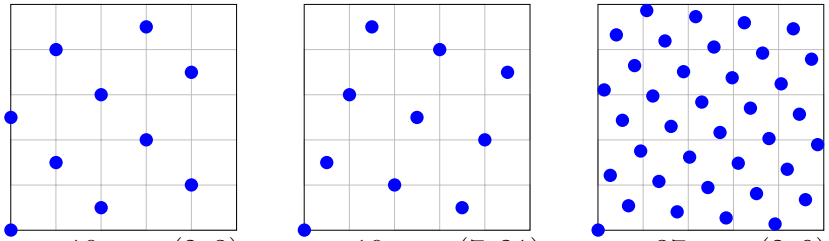

Figure 7: Examples of bidimensional lattices.

The unidimensional design consisting of n equally spaced points is a lattice design with generator $g = (1)$. See Figure 7 for examples of bidimensional lattices.

The models which can be related to lattices are Fourier regression models with interaction. The following notation is used.

Notation 12 *Let $F(d; m_1, \ldots, m_d; M)$ denote the complete Fourier model in d dimensions with one-dimensional "marginal" models for $x_1, \ldots, x_d$ up to orders $m_1, \ldots, m_d$, respectively, and with interactions up to M factors included.*

A number of results are available for Fourier models, see Riccomagno et al. (1997) for details. The lattice design generated by

$$(g_1, \ldots, g_d) = \left(1, 2m_1 + 1, (2m_1 + 1)(2m_2 + 1), \ldots, \prod_{k=1}^{d-1} (2m_k + 1)\right)$$

with size $n = \prod_{k=1}^{d}(2m_k + 1)$ is D-optimal for all Fourier models $F(d; m_1, \ldots, m_d; M)$ with $1 \leq M \leq d$. However, unless $M = d$, this grid has a large number of points, relative to the number of parameters. For the Fourier model $F(d; m_1, \ldots, m_d; 1)$, the following generator can be used

$$(g_1, \ldots, g_d) = \left(1, m_1 + 1, (m_1 + 1)(m_2 + 1), \ldots, \prod_{k=1}^{d-1}(m_k + 1)\right)$$

with size $n = 2m_d \prod_{k=1}^{d}(m_k + 1) + 1$. The two previous are examples of power-type generators.

A special type of generators are defined recursively, for instance using the formula $g_d = m_{d-1}g_{d-1} + 1$, with the initial value $g_1 = 1$. For the model $F(d; m, \ldots, m; 1)$ it is obtained $g_d = (m^d - 1)/(m - 1)$ and size $n = (2m^{d+1} - m - 1)/(m - 1)$. If $m = 1$ this is simplified to $g_d = d$, with size $n = 2d+1$. The general case for establishing general properties of generators is still subject of research see Schwabe and Wynn (2008).

6 Gaussian processes and Karhunen-Loeve expansions

Consider the model

$$Y(x) = f(x)^T\beta + Z(x),$$

where β is random vector, with $\mathrm{E}(\beta) = \mu$ and $\mathrm{Cov}(\beta) = \sigma^2\Sigma$. $Z(x)$ is a stochastic process independent of β, with zero mean and covariance structure depending only on the locations, i.e. $\mathrm{Cov}(Z(s), Z(t)) = \sigma^2 R(s,t)$ and $R(s,t)$ defines a Gaussian correlation structure $R(s,t) = e^{-\sum \theta_i |s_i - t_i|^{p_i}}$. Consequently, $\gamma(s,t) = \mathrm{Cov}(Y(s), Y(t)) = \sigma^2\left(f(s)^T\Sigma f(t) + R(s,t)\right)$ and $\mathrm{E}Y(x) = \mu$.

6.1 KL expansion

The stochastic process $Y(x)$ can be approximated through a Karhunen-Loeve expansion, i.e.

$$Y(x) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} q_i \phi_i(x) \tag{5}$$

where q_i are independent standard normal random variables, and the eigenvalues and eigenfunctions λ_i and $\phi_i(x)$ are obtained by solving the Fredholm equation $\int \gamma(s,t)\phi_i(s)ds = \lambda_i\phi_i(x)$.

6.2 Future work scope

Future work involves using the truncation the Karhunen-Loeve expansion. That is, Equation (5) can be approximated through the finite sum $\tilde{Y}(x) = \sum_{i=1}^{m} \sqrt{\lambda_i} q_i \phi_i(x)$. If the covariance function $\gamma(s,t)$ is smooth enough, then a good approximation is achievable with a small number of terms. The approximated expansion needs a careful balance between the number of terms used m and the number of design points used n. When $n > m$ an equivalence for Gaussian processes has been found between maximum entropy sampling and D-optimality see Wang (2008).

7 Appendix

7.1 Orthonormality properties of Ψ

The elements of Ψ satisfies orthonormality properties, that is

1. $\langle \psi_0(x), \psi_{j,k}(x) \rangle = 0$ for $k = 0, 1, 2, \ldots$ and $j = 0, 1, \ldots, 2^k - 1$,
2. $\langle \psi_{j,k}(x), \psi_{j,k}(x) \rangle = 1$ for $k = 0, 1, 2, \ldots$ and $j = 0, 1, \ldots, 2^k - 1$,
3. $\langle \psi_{j_1,k_1}(x), \psi_{j_2,k_2}(x) \rangle = 0$ for $(j_1, k_1) \neq (j_2, k_2)$

For the above properties recall that the inner product between two integrable functions $f = f(x), g = g(x),\ f, g : [0,1) \mapsto \mathbb{R}$ is defined as $\langle f, g \rangle = \int_0^1 fg dx$.

References

A. Atkinson and D. Cox. Planning experiments for discriminating between models. *J. Roy. Statist. Soc. Ser. B*, 36:321–348, 1974. With discussion by H. P. Wynn, D. M. Titterington, P. J. Laycock, D. V. Lindley, D. H. Hill, A. M. Herzberg, P. A. Tukey, A. O'Hagan, V. V. Fedorov, J. Dickey, J. Kiefer and C. A. B. Smith.

A. Atkinson and A. N. Donev. *Optimum Experimental Designs*. Number 8 in Oxford statistical science series. Oxford University Press, Oxford, 1992.

A. Atkinson and V. Fedorov. Optimal design: experiments for discriminating between several models. *Biometrika*, 62(2):289–303, 1975.

A. C. Atkinson, A. N. Donev, and R. D. Tobias. *Optimum experimental designs, with SAS*, volume 34 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2007.

R. A. Bates, E. Riccomagno, R. Schwabe, and H. P. Wynn. The use of lattices in the design of high-dimensional experiments. In *New developments and applications in experimental design (Seattle, WA, 1997)*, volume 34 of *IMS Lecture Notes Monogr. Ser.*, pages 27–35. 1998.

P. Bratley and B. L. Fox. ALGORITHM 659 Implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Soft.*, 14(1):88–100, 1988.

H. Dette and W. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1997.

A. M. Herzberg and W. N. Traves. An optimal experimental design for the Haar regression model. *Canad. J. Statist.*, 22(3):357–364, 1994.

P. G. Hoel. Minimax designs in two dimensional regression. *Ann. Math. Statist.*, 36:1097–1106, 1965.

J. Kiefer. Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B*, 21: 272–319, 1959.

J. Kiefer and J. Wolfowitz. Optimum designs in regression problems. *Ann. Math. Statist.*, 30:271–294, 1959.

H. Maruri-Aguilar and H. Wynn. Smooth supersaturated models. MUCM Technical report, 2008.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

E. Riccomagno, R. Schwabe, and H. P. Wynn. Lattice-based D-optimum design for Fourier regression. *Ann. Statist.*, 25(6):2313–2327, 1997.

R. Schwabe. *Optimum designs for multi-factor models*, volume 113 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1996a.

R. Schwabe. Model robust experimental design in the presence of interactions: the orthogonal case. *Tatra Mt. Math. Publ.*, 7:97–104, 1996b. ProbaStat '94 (Smolenice Castle, 1994).

R. Schwabe and H. Wynn. Self avoiding generator sequences for D-optimum lattice designs. Forthcoming preprint, IWAP, 2008.

S. Silvey. *Optimal design*. Chapman & Hall, London, 1980. An introduction to the theory for parameter estimation, Monographs on Applied Probability and Statistics.

K. Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12:1–85, 1918.

I. M. Sobol'. Distribution of points in a cube and approximate evaluation of integrals. *Akademija Nauk SSSR. Žurnal Vyčislitel'noĭ Matematiki i Matematičeskoĭ Fiziki*, 7:784–802, 1967.

G. Strang. Wavelet transforms versus Fourier transforms. *Bull. Amer. Math. Soc. (N.S.)*, 28(2):288–305, 1993.

Y. Tian and A. M. Herzberg. Estimation and optimal designs for linear Haar-wavelet models. *Metrika*, 65(3):311–324, 2007.

L. Wang. Karhunen-Loeve expansions and their applications. Ph. D. thesis, London School of Economics, 2008.

D. Würtz. High dimensional scrambled sobol sequences an r and splus software implementation. Journal of Statistical Software (preprint), 2008.

H. P. Wynn. Results in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B*, 34:133–147, 170–186, 1972. With discussion by M. J. Box, P. Whittle, S. D. Silvey, A. A. Greenfield, A. M. Herzberg, J. Kiefer, D. R. Cox, L. V. White, A. C. Atkinson, R. J. Brooks, C. L. Atwood, and R. Sibson, and replies by H. P. Wynn and P. J. Laycock.