Statistica Sinica **13**(2003), 129-142

EQUIDISTRIBUTED DESIGNS IN NONPARAMETRIC REGRESSION

E. Rafajłowicz and R. Schwabe

Wrocław University of Technology and Otto von Guericke University

Abstract: In nonparametric regression estimation the best possible rate of convergence is attained when the design points are equidistributed. Whole sequences of equidistributed designs can be generated from vectors with algebraically independent components. Generators are determined which perform well already for moderate numbers of observations. The proposed design sequences are nearly optimal in classical settings and simultaneously appropriate for the nonparametric approach.

Key words and phrases: Additive regression, equidistributed sequence, lattice design, nonparametric regression, optimal design.

1. Introduction

In statistical applications there is often a strict distinction made between observational studies and experimental situations. While for the former only a *passive* collection is possible for the observational outcomes of the *input–output* pairs (x, y), there is a substantial advantage in experiments when the settings for the input x can be *actively* chosen. Typically, the method of statistical analysis which measures the influence of the input x on the output y does not depend on whether the data arise from observational studies or from planned experiments. However, the performance of the estimators, tests etc., is strongly influenced by the distribution of the input variables. In terms of sample sizes, savings of fifty percent and more are possible in reaching a prescribed accuracy when an optimal or efficient design for the input x is chosen (see Cox and Reid (2000). For further readings on the wide–spread scope of applications for experimental designs, we refer to the handbook by Ghosh and Rao (1996)).

In the present paper we propose a method for generating experimental designs based on equidistributed sequences (see Niederreiter (1992)). The method is designed for nonparametric regression with an orthogonal expansion for the estimator. It may be applied also to kernel type estimators and semiparametric approaches.

We investigate a *quasi least squares estimator* based on an orthogonal series expansion. The estimator proposed here for a fixed design is similar to estimators commonly used in nonparametric random design settings.

The quasi least squares estimator is particularly simple and intuitively appealing when applied to an additive regression function estimated from a multivariate equidistributed design. The construction of the additive quasi least square estimator follows an idea by Linton and Nielsen (1995) (see also Efromovich (1999)).

Most of the existing theory on optimal design of experiments is based on the assumption that the linear regression model under study is correctly specified. This fundamental assumption implies that the existing design criteria (such as those related to D–, A– or G–optimality) take only the variance of the estimator for the parameters of the response function into account.

When considering regression in a nonparametric setting we are in a quite different position since, in addition to the variance, there is also a bias term present. Thus, we have to measure the estimation accuracy in terms of the mean squared error. On the other hand, the bias term cannot be directly minimized for the obvious reason that the underlying regression function is unknown. Equidistributed sequences are well suited for approximate calculations of integrals in the sense that they automatically reduce the bias term.

In this paper we focus on the fixed design case. Few papers have been devoted to the question of choosing "good" input sequences in a nonparametric setting so far. For example, Müller (1984) proposed a choice of design points for kernel type estimation, and Rafajłowicz (1987) discussed the use of points provided by quadrature formulae in nonparametric estimation. Recently, experimental design problems for local least squares estimation were considered by Müller (1998) and Cheng, Hall and Titterington (1988).

In the next section the class of designs considered is described, and the main problem is stated in the context of nonparametric regression estimation. In Section 3 results on consistency of orthogonal series estimators with equidistributed designs are stated, together with results on their convergence rate. These asymptotic results allow for restricting the class of design generators to quadratic irrationals which, for example, have a one-periodic continued fractions representation. In Section 5 the algorithm for searching satisfactory generators is proposed. Finally, in Section 6 selected generators and the corresponding designs are studied via some simulation results that compare estimation accuracy obtained by using equidistributed designs with that of equidistant grid designs. Proofs are deferred to the Appendix.

2. Assumptions

The dependence of the output variable y on the input x is described by a functional relationship $y = f(x) + \varepsilon$, where ε is a random observational error. Classical design theory assumes a parametric model, i.e., the structure of the response function f is known up to a finite-dimensional parameter ϑ, $f(x) = g(x,\vartheta)$. In particular, most results are related to linear models, $f(x) = g(x)^\top \vartheta = \sum_{j=1}^p g_j(x)\vartheta_j$. In a nonparametric setting however, the response function may come from a larger, infinite-dimensional function space. Hence, a whole function must be estimated rather than merely a few parameters and input variables x_i should cover the whole design region as well as possible. The collection $x_1, \ldots, x_n$ of input variables is called the design of the experiment.

The function f is estimated from observations $y_i = f(x_i) + \varepsilon_i$ based on the design points x_i, where the errors ε_i are zero mean, homoscedastic and uncorrelated. For estimating f in a nonparametric setting we choose equidistributed design sequences (see Kuipers and Niederreiter (1974)). Denote by $\{\{\cdot\}\}$ the fractional part of a real number.

Definition 1. (Equidistributed designs) Let $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})^T$ be a vector of distinct irrational numbers satisfying affine linear independence over the rationals, i.e.,

$$\alpha_0 + \sum_{j=1}^{d} \alpha_j \theta^{(j)} = 0 \quad \text{implies} \quad \alpha_0 = \alpha_1 = \cdots = \alpha_d = 0 \tag{1}$$

for every set of rational numbers $\alpha_0, \ldots, \alpha_d$.

An equidistributed sequence in d dimensions is defined by

$$x_i = (\{\{i\theta^{(1)}\}\}, \ldots, \{\{i\theta^{(d)}\}\}), \tag{2}$$

$i = 1, 2, \ldots$, where $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})$ is called the generator of the sequence.

The first n members $x_1, \ldots, x_n$ of an equidistributed sequence constitute an equidistributed design of size n.

Note that the designs of Definition 1 are equidistributed in the sense of Kuipers and Niederreiter (1974), while one can consider other sequences of this type, for example Van der Corput sequences, we confine our attention to (2) with θ irrational.

Rational equidistributed designs have been shown to be optimal in Fourier regression models by Bates, Riccomagno, Schwabe and Wynn (1998).

In the sequel we consider generators with quadratic irrationals $\theta^{(j)}$: solutions of quadratic equations $ax^2 + bx + c = 0$, where a, b and c are integers and the determinant $b^2 - 4ac > 0$ is not a perfect square. From Lagrange's theorem (see Baker (1984)) it follows that $\theta^{(j)}$ is a quadratic irrational if and only if its continued fraction expansion

$$\theta^{(j)} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$$

is ultimately periodic with integer coefficients $a_0, a_1, \ldots$. To keep the presentation simple we confine ourselves to quadratic irrationals $CF(A)$ which are one–periodic continued fractions, thus $a_0 = 0$ and $a_1 = a_2 = \cdots = A$. For example, $CF(1) = (\sqrt{5}-1)/2$ is the famous golden section number and $CF(2) = \sqrt{2}-1$. Let $\mathcal{CF} = \{CF(A) |\, A = 1, 2, \ldots\}$. We denote by $\mathcal{CF}_d$ those generators $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})$ with components in $\mathcal{CF}$ satisfying the independence condition (1).

We restrict the design region to the unit cube $X = [0,1]^d$. Let $C(X)$ denote the space of continuous functions and $V(X)$ the class of functions f of bounded variation $\mathcal{V}(f)$ on X (in the sense of Hardy and Krauze for the multivariate case, see, e.g., Kuipers and Niederreiter (1974)). $L_2(X)$ is the space of square integrable functions f on X. A finite an orthonormal basis for $L_2(X)$ is denoted by $\bar{v}_N = (v_1, \ldots, v_N)^T$. $W^\mu(X)$ denotes the Sobolev space of functions on X with smoothness μ, i.e., for $\mu = p + \alpha$, $0 \le \alpha < 1$, $f \in W^\mu$ is p times differentiable and its pth derivative is Lipschitz continuous with exponent α. When $X = [0,1]$ we write $C(0,1)$, $V(0,1)$, etc. For $f \in W^\mu(0,1)$, $0 < \mu \le 1$, denote by $\mathcal{L}(f) > 0$ the Lipschitz constant, $|f(x') - f(x'')| \le \mathcal{L}(f)|x' - x''|^\mu$. A function $f \in W^\mu(0,1)$ is periodic if $f^{(p)}(0) = f^{(p)}(1)$, $0 \le p \le \mu$.

In the bivariate case we consider addditive regression functions, where f can be decomposed into marginal effects of two variables $x^{(1)}$, $x^{(2)}$ according to

$$f(x^{(1)}, x^{(2)}) = c_0 + g(x^{(1)}) + h(x^{(2)}), \tag{3}$$

where c_0 is the overall mean, and g and h are square integrable functions. In order to ensure identifiablity we assume g and h integrate to zero. Extension of (3) to the multivariate case is immediate.

Let $v_1, v_2, \ldots$ be a complete sequence of orthonormal functions in $L_2(X)$. Typically this will be a system of orthogonal polynomials like the Legendre polynomials, or the trigonometric system for periodic functions. Let $a_k = \int_X f(x) v_k(x)\, dx$. We define the estimator $\hat{f}_n$ of f in a natural way by

$$\hat{f}_n(x) = \sum_{k=1}^{N} \hat{a}_{kn} v_k(x), \qquad \hat{a}_{kn} = \frac{1}{n} \sum_{i=1}^{n} y_i v_k(x_i), \tag{4}$$

so $\hat{a}_{kn}$ is the estimator of the regression coefficients a_k at stage n. The estimators $\hat{a}_{kn}$ and $\hat{f}_n$ will be called *quasi least squares estimators*. They are obtained by replacing the inverse of the normalized Fisher information matrix $n^{-1} \sum_{i=1}^{n} \bar{v}_N(x_i) \bar{v}_N^T(x_i)$ by the $N \times N$ identity matrix for computational simplicity.

In (4), N plays the role of a smoothing parameter. In asymptotic considerations the degree $N = N(n)$ will depend on the number of observations. Typically, N increases more slowly than n, and we suppose

$$N(n) \to \infty, \qquad N^3(n)/n^{2-\varepsilon} \to 0, \tag{5}$$

for arbitrarily small $\varepsilon > 0$. For $f \in W^{\mu}(X)$, select $N(n)$ in such a way that

$$\gamma_1 n^{1/(1+2\mu/d)} \leq N(n) \leq \gamma_2 n^{1/(1+2\mu/d)}, \tag{6}$$

for some constants $0 < \gamma_1 \leq \gamma_2$.

If it is a–priori known that a bivariate function f is additive, then it is natural to construct its estimator $\hat{f}_n$ in an additive form. Let (ϕ_k) be a system of orthonormal functions, complete in $L^2(0,1)$, and which includes a constant term $\phi_1(x) = 1$. This holds for the Legendre polynomials and the trigonometric system.

We generate the design sequence $x_i = (\{\{i\theta^{(1)}\}\}, \{\{i\theta^{(2)}\}\})$ according to Definition 1. The estimator $\hat{f}_n$ of (3) has the form

$$\hat{f}_n(x^{(1)}, x^{(2)}) = \hat{c}_0 + \sum_{k=2}^{N_1} \hat{g}_k \phi_k(x^{(1)}) + \sum_{l=2}^{N_2} \hat{h}_k \phi_k(x^{(2)}),$$

where the coefficients are given by

$$\hat{c}_0 = \frac{1}{n} \sum_{i=1}^{n} y_i, \qquad \hat{g}_k = \frac{1}{n} \sum_{i=1}^{n} y_i \phi_k(x_i^{(1)}), \qquad \hat{h}_k = \frac{1}{n} \sum_{i=1}^{n} y_i \phi_k(x_i^{(2)}).$$

The sequences $N_1(n) \to \infty$, and $N_2(n) \to \infty$ play the role of smoothing parameters, and $N(n) = N_1(n) + N_2(n) - 1$ is the joint degree of the basis.

By $c, c_1, c_2, \ldots$ we will denote generic constants throughout. These may vary from line to line.

3. Asymptotics for Estimation Based on equidistributed designs

The problem of nonparametric estimation is to find a sequence of functions $\hat{f}_n$ based on the pairs $(x_1, y_1), \ldots, (x_n, y_n)$ such that $\hat{f}_n$ approaches f when the sample size increases.

In the nonparametric setting standard design criteria which are based on the information matrix cannot be applied because for those exact knowledge of the model structure is assumed. Moreover, designs which are optimal in the classical setting are typically concentrated on a relatively small number of design points and do not allow for checking the assumed model.

It is desirable to consider the integrated mean squared error as a design criterion:

$$IMSE(\hat{f}_n, f) = \int_X \operatorname{Var}(\hat{f}_n(x))\, dx + \int_X (E\hat{f}_n(x) - f(x))^2\, dx.$$

In the univariate case we consider the situation where the v_k are elements of the trigonometric system $\{1, \sqrt{2}\sin(2\pi x), \sqrt{2}\cos(2\pi x), \sqrt{2}\sin(2\pi 2x), \sqrt{2}\cos(2\pi 2x), \ldots\}$, $x \in [0,1]$.

Proposition 1. *Let f be in $V(0,1) \cap C(0,1)$ and suppose* (5) *is satisfied for the sequence $N(n)$. Assume that $x_i = \{\{i\theta\}\}$, $i = 1, \ldots, n$, is an equidistributed design with an algebraic irrational generator θ. If $\hat{f}_n$ is the estimator of f defined by* (4) *then $IMSE(\hat{f}_n, f) \to 0$ as $n \to \infty$.*

Under slightly more restrictive conditions on the degree $N(n)$ it can be shown that Proposition 1 remains valid when v_k is the system of Legendre polynomials.

Proposition 2. *Let $f \in W^{\mu}(0,1)$, $\sqrt{3}/2 < \mu \leq 1$ and let $N(n)$ be chosen such that* (6) *holds. Let $x_i = \{\{i\theta\}\}$, $i = 1, \ldots, n$, an equidistributed sequence, where the generator θ is an ultimately periodic continued fraction. If f is periodic, then $IMSE(\hat{f}_n, f) = O(n^{-\frac{2\mu}{2\mu+1}})$.*

Note that under the conditions of Proposition 2, $\hat{f}_n$ attains the best possible rate of convergence in the integrated mean squared error sense, uniformly in $W^{\mu}(0,1)$.

For additive functions f we assume that the components of $\hat{f}_n$ are both spanned by the trigonometric system. From Propositions 1 and 2 we can infer similar asymptotic properties of the estimators $\hat{f}_n$ constructed according to (7) in view of

$$\begin{aligned} &IMSE(\hat{f}_n, f) \\ &= E(\hat{c}_0 - c_0)^2 + E\int_0^1 (\hat{\hat{g}}_n(x^{(1)}) - g(x^{(1)}))^2 dx^{(1)} + E\int_0^1 (\hat{\hat{h}}_n(x^{(2)}) - h(x^{(2)}))^2 dx^{(2)}, \end{aligned}$$

where $\hat{\hat{g}}_n(x^{(1)}) = \sum_{k=2}^{N_1} \hat{g}_k \phi_k(x^{(1)})$ and $\hat{\hat{h}}_n(x^{(2)}) = \sum_{k=2}^{N_2} \hat{h}_k \phi_k(x^{(2)})$.

Proposition 3. *Assume that the additive regression function f is periodic in both components. Let $\hat{f}_n$ be the estimator based on the equidistributed design $x_i = (\{\{i\theta^{(1)}\}\}, \{\{i\theta^{(2)}\}\})$, $i = 1, \ldots, n$, with generator $\theta = (\theta^{(1)}, \theta^{(2)}) \in \mathcal{CF}_2$. If g and h are in $V(0,1) \cap C(0,1)$, and if* (5) *holds for $N_1(n)$ and $N_2(n)$, then $IMSE(\hat{f}_n, f) \to 0$ as $n \to \infty$.*

Proposition 4. *Assume the conditions of Proposition* 3 *on the design $x_1, \ldots, x_n$ and the shape of f and $\hat{f}_n$. If $g, h \in W^{\mu}(X)$, $\sqrt{3}/2 < \mu \leq 1$, and* (6) *holds for $N_1(n)$ and $N_2(n)$, then $IMSE(\hat{f}_n, f) = O(n^{-\frac{2\mu}{2\mu+1}})$.*

Stone (1986) proved the best rate of convergence is the same for additive regression as for the univariate case. This is attained by $\hat{f}_n$ in view of Proposition 4.

Note that the results of Propositions 1 to 4 can be generalized to wider classes of generators which satisfy (1), such as ultimately periodic continued fractions with bounded coeffiecients.

4. Choice of the Approximation Order

In the univariate case we use the notation $\hat{f}_{n,N}(x)$ to display the dependence of $\hat{f}_n(x)$ on N. The best choice for the degree is $N^* = \arg\ \min_N IMSE(\hat{f}_{n,N}, f)$. However, this is not attainable in practice because N^* depends on the unknown function f. Therefore we have to derive an estimate $\hat{N}$ for the degree N.

Define the residuals by $r(n,N) = n^{-1}\sum_{i=1}^n (y_i - \hat{f}_{n,N}(x_i))^2$ and the mean squared error by $MSE(n,N) = E\, n^{-1}\sum_{i=1}^n (\hat{f}_{n,N}(x_i) - f(x_i))^2$. Then

$$E\, r(n,N) = MSE(n,N) + \sigma^2 - \frac{2\sigma^2}{n}\sum_{k=1}^{N} n^{-1}\sum_{i=1}^{n} v_k^2(x_i).$$

For the trigonometric system (v_k) the Lipschitz constant decomposes according to $\mathcal{L}(v_k^2) \le ck$. If x_i, $i = 1,\ldots,n$, is an equidistributed design with $\theta \in \mathcal{CF}_d$ then $n^{-1}\sum_{i=1}^n v_k^2(x_i) - \int_X v_k^2(x)\,dx = O(k\log(n)/n)$, which implies $E\, r(n,N) = MSE(n,N) + \sigma^2 - 2\sigma^2\frac{N}{n} + O(N^2\log(n)/n^2)$. Following an idea of Mallows, we define $crit(n,N) = r(n,N) - \sigma^2 + 2\sigma^2 N/n$. Let $\hat{N} = \arg\ \min_{\tau_1 n^\beta \le N \le \tau_2 n^\beta} crit(n,N)$ where $0 < \beta < 1/2$ and $0 < \tau_1 < \tau_2$. Then it can be shown that $IMSE(\hat{f}_{n,N^*}, f) - crit(n,\hat{N}) = o(n^{-(1/2-\delta)})$ a.s. for every $\delta > 0$, if $f \in W^\mu(0,1)$, $\mu > \sqrt{3}/2$, and x_i constitutes an equidistributed design with $\theta \in \mathcal{CF}$.

This result remains valid if σ^2 is replaced by a consistent estimator $\hat{\sigma}^2$ in the definition of the criterion function. Similar statements can be made in the additive case.

5. Selection of Generators for Experimental Designs

Denote by $M_n^{(N)}(\theta) = \frac{1}{n}\sum_{i=1}^n \bar{v}_N(x_i)\bar{v}_N^T(x_i)$ the normalized information matrix for estimating a response function spanned by $\bar{v}_N = (v_1,\ldots,v_N)^T$ when the equidistributed design $x_i = \{\{i\theta\}\}$, $i = 1,\ldots,n$, is used.

As a benchmark for the comparison of designs for different values of n and N we take the theoretically optimal uniform design measure for which $\mathrm{tr}((\int \bar{v}_N(x)\bar{v}_N^T(x)\,dx)^{-1}) = N$. Here, tr denotes trace. Denote by $q(\theta,n,N) = \frac{1}{N}\mathrm{tr}(M_n^{(N)}(\theta)^{-1}) - 1$ the standardized deviation of the design generated by θ from the ideal uniform design. Then the design $x_i = \{\{i\theta\}\}$, $i = 1,\ldots,n$, is considered to be sufficiently good if $|q(\theta,n,N)| \le \delta$ for a prespecified tolerance $\delta > 0$.

The main difference between the present and the classical A-optimal design problem statement is the restriction to equidistributed sequences. This side condition assures that the bias term becomes small for sufficiently large n. An essential part of the bias is given by $\left|n^{-1}\sum_{i=1}^n f(x_i)v_k(x_i) - \int_0^1 f(x)v_k(x)dx\right|$. Thus, it is important to select the generator θ in such a way that $\Delta_n(\theta,f) = \left|n^{-1}\sum_{i=1}^n f(x_i) - \int_0^1 f(x)dx\right|$ converges sufficiently fast to zero, uniformly in f.

Denote by $D_n^\star$ the discrepancy of the sequence x_i, $i = 1, \ldots, n$. The following bounds are known from the literature (see Kuipers and Niederreiter (1974)).

(**B1**) Koksma's inequality: For every $f \in V(0,1)$ we have $\Delta_n(\theta, f) \le V(f) D_n^\star$.

(**B2**) $\Delta_n(\theta, f) \le c(D_n^\star)^\mu$ for every $f \in W^\mu(0,1)$, $0 < \mu \le 1$.

(**B3**) For $\theta \in \mathcal{CF}$ the discrepancy $D_n^\star$ is of the order $O(\log(n)/n)$.

(**B4**) For $\theta \in \mathcal{CF}_d$ the discrepancy is of the order $O(n^{-1+\varepsilon})$ for every $\varepsilon > 0$.

Within the class $\mathcal{CF}_d$ we consider a selection algorithm for searching "good" generators because explicit minimization is too time consuming. The selection algorithm starts by combining generators for lower dimensions and separates those which are suitable for regression with larger numbers of variables. This algorithm resembles the procedure of winnowing, i.e., freeing grain from chaff by wind.

Algorithm for winnowing good design generators

Step 0. Specify the dimension D for which the search will be performed. Choose the range of search $K > 0$ and tolerance bounds $\delta = (\delta_1, \ldots, \delta_D)$ of acceptable values for $|q_d((\theta^{(1)}, \ldots, \theta^{(d)}), n, N)|$ for given dimension d.

Step 1. Set $d = 1$ and select the set Θ_1 of one–dimensional generators $\theta^{(1)} = CF(p)$, $p = 1, \ldots, K$, for which the corresponding sequence $x_i = \{\{i\theta^{(1)}\}\}$, $i = 1, \ldots, n$, satisfies $|q_1(\theta^{(1)}, n, N)| \le \delta_1$.

Step 2. If $d < D$ set $d := d + 1$ and go to Step 3, otherwise STOP the algorithm and provide the sets $\Theta_1, \ldots, \Theta_D$ of generators as the result.

Step 3. Form the set of candidate generators $\mathcal{G}_d = \Theta_{d-1} \times \Theta_1$. For each $\theta_d = (\theta^{(1)}, \ldots, \theta^{(d)}) \in \mathcal{G}_d$, calculate $q_d(\theta_d, n, N)$ and select the set Θ_d of those $\theta_d \in \mathcal{G}_d \cap \mathcal{CF}_d$ for which $|q_d(\theta_d, n, N)| \le \delta_d$ holds. Then go to Step 2.

Occasionally the algorithm might stop at a premature stage when Θ_d becomes void. In that case the range K or the tolerance bounds $\delta = (\delta_1, \ldots, \delta_D)$ have to be increased.

As a side effect the winnowing algorithm results in uniformly distributed projections onto lower–dimensional hypercubes. This is of particular importance in situations where not all input variables are essential for fitting the model.

6. Numerical Examples and Simulations

In this section we report the results of using the winnowing algorithm for selecting good design generators. Generators for the univariate case are mainly considered as ingredients for higher dimensional designs. We note only that the golden section number $CF(1) = (\sqrt{5} - 1)/2$ can be used as a universal design generator for regression functions spanned by orthonormal Legendre polynomials for various degrees N and sample sizes n, as exhibited in Figure 1 where $Nq(\theta, n, N)$ is depicted.

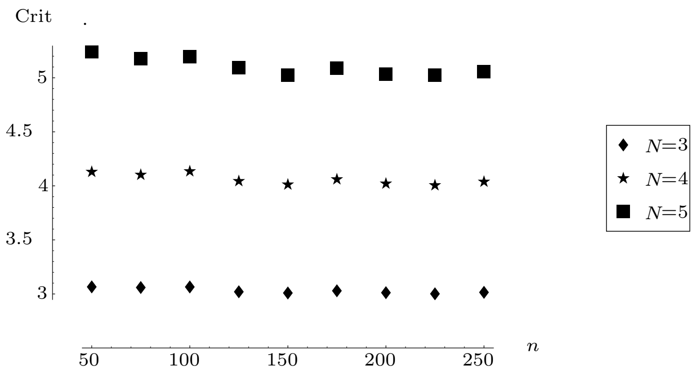

Figure 1. Performance of the golden section number $CF(1) = (\sqrt{5} - 1)/2$.

Table 1. Performance of $\theta = (CF(i), CF(j))$: $Nq(\theta, n, N)$.

(i, j)	(10,36)	(5,43)	(3,44)	(14,44)	(10,14)
Legendre polynomials	0.000	-0.001	0.001	0.001	0.001
trigonometric system	0.005	0.000	-0.013	0.012	0.017

In Table 1 some designs are presented which were selected by the winnowing algorithm for the additive bivariate third order Legendre polynomial with $N = 7$ and $n = 225$.

By construction the winnowing algorithm selects generators, which are of high quality when the orthonormal system (v_k), the sample size n and the degree N are all specified. We investigate the performance of the associated designs if some or all experimental conditions are modified. For example, it is shown in Table 1 that the generators selected for the Legendre polynomials also perform well if they are used for a trigonometric system consisting of cosine terms only.

To illustrate the uniformity of the selected designs we exhibit the design generated by $\theta = (CF(5), CF(43))$ in Figure 2. It performs almost optimally for both Legendre polynomials and a trigonometric system.

The winnowing algorithm reduces the computational burden for selecting generators, but it does not guarantee that the generator is optimal. To verify the efficiency of the winnowing algorithm the quality of winnowed generators is compared with those found by exhaustive search. This is done here for comparison only, since it is applicable for relatively small problems.

The winnowing algorithm started from the set $CF(i)$, $i = 1, \ldots, 55$, of univariate generators. Thus, exhaustive search had to be conducted over the generators $\theta = (CF(i), CF(j))$, i, $j = 1, \ldots, 55$. As a basis, additive bivariate third order Legendre polynomials were used. The winnowing algorithm was performed with two different values for the tolerance bound δ_1. In the first case (strategy I), δ_1 was selected in such a way that the computational time was reduced by a factor of 1/2 compared to exhaustive search. In the second case (strategy II), the reduction in computational time was by a factor of 1/20. The results of comparisons are summarized in Table 2 for various n. Column 2 gives the performance of the optimal generator found by exhaustive search, while the results for the winnowing algorithm are listed in columns 3 and 4, respectively. For strategy I the winnowing algorithm found the optimal generators across the whole range of n. Strategy II resulted in a small loss of quality for small sample sizes, and in optimal values for larger n.

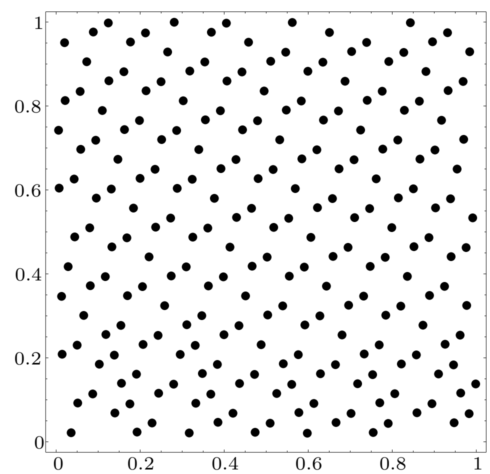

Figure 2. Design generated by $\theta = (CF(5), CF(43))$.

A simple competitor to equidistributed designs could be equidistant grid designs concentrated on a $\sqrt{n} \times \sqrt{n}$ square lattice. These designs perform poorly as is indicated in the last column of Table 2. Hence, they are not to be recommended, at least for small to moderate sample sizes.

Table 2. Comparison of the winnowing algorithm with exhaustive search and equidistant designs.

n	exhaustive search	winnowing strategy I	winnowing strategy II	equidistant design
25	0.314	0.314	0.341	13.10
49	-0.007	0.017	0.037	4.57
100	0.000	0.000	0.112	1.847
225	0.000	0.000	0.000	0.762

Another commonly used competitor is a design generated by random (or pseudo-random) numbers from a uniform distribution. For comparison we generated 100 random designs of size $n = 225$ from the uniform distribution on the unit square. For each of these random designs the performance measure $Nq((x_1, \ldots, x_n), n, N) = \mathrm{tr}(M_n^{(N)}(x_1, \ldots, x_n)^{-1}) - N$ was calculated. The best random design had a value of 0.0002, but proved to be worse still by an order of magnitude than the best design from Table 2. The mean performance measure was 0.21 with a standard deviation of 0.28, which shows large variation. Thus, equidistributed designs do better than randomly generated designs.

We also investigated the performance of equidistributed designs compared to equdistant grid designs and randomly generated designs when the test function to be estimated does not come from the space spanned by the basis functions. As a measure of performance we calculated the empirical integrated mean squared error

$$EIMSE\,(x_1, \ldots, x_n) = \frac{1}{mL} \sum_{j=1}^{m} \sum_{\ell=1}^{L} \left[f(\kappa_\ell) - \hat{f}_n^{(j)}(\kappa_\ell) \right]^2,$$

where $m = 100$ replicates of the observations $y_i^{(j)} = f(x_i) + \varepsilon_i^{(j)}$ at the design points x_i, $i=1, \ldots, n$, were generated according to a Gaussian noise with variance σ^2, and $\hat{f}_n^{(j)}$ denotes the estimate based on the jth series of observations $(x_i, y_i^{(j)})$, $j = 1, \ldots, m$. Moreover, to avoid evaluation of the integral, $L = 500$ points κ_ℓ were randomly generated from the uniform distribution, and the integral was estimated by the average of the squared deviations.

Some results are exhibited in Figures 3 and 4 for the case of a trigonometric test function $f(x) = \sin(\pi x^{(1)}) + \cos(\pi x^{(2)})$, while additive bivariate third order Legendre polynomials were used as the basis. The sample size was $n = 121$ for all designs considered.

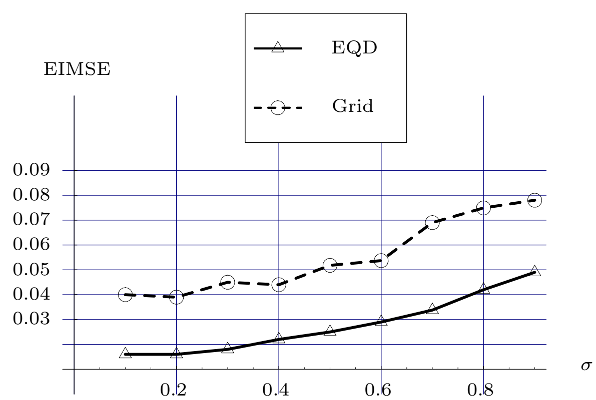

Figure 3. Comparison of equidistributed and equidistant designs.

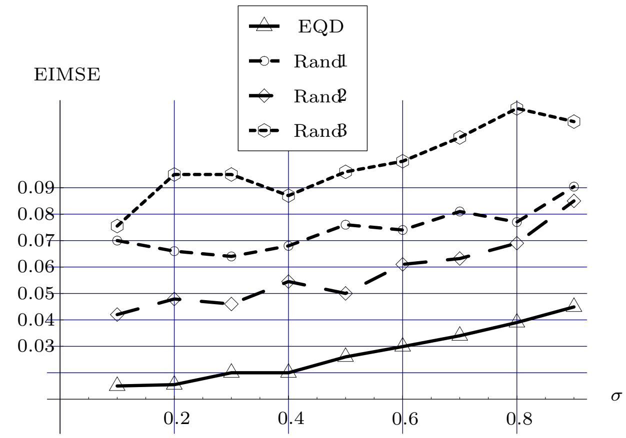

Figure 4. Comparison of equidistributed and randomly generated designs.

In Figure 3 the equidistributed design generated by $\theta = (CF(10), CF(14))$ is compared with an 11×11 eduidistant grid design. The picture shows that $EIMSE$ is smaller for the equidistributed design than for the equidistant grid design for all values of σ.

In Figure 4 the same equidistributed design is compared with three different random designs for various values of σ. Also, in this comparison, the equisdistributed design provides smaller estimation errors. Even if it may occasionally happen that a random design has a quality comparable to the selected equidistributed design, one can simultaneously observe a large variability of the quality of random designs (see run Rand. 3 in Figure 4 in which the random design has an $EIMSE$ about 5 times as large as that of the equidistributed design).

Acknowledgements

The research was partly supported by grants of the *Deutsche Forschungsgemeinschaft* and *The Council for Scientific Research of Poland*. We wish to thank some anonymous referees for their helpful comments.

Appendix: Proofs

Proof of Proposition 1. From the orthonormality and completeness of the sequence (v_k) it follows that the integrated mean squared error $IMSE(\hat{f}_n, f)$ can be decomposed according to

$$IMSE(\hat{f}_n, f) = W_n + B_n^2 + R(N, f), \tag{7}$$

a variance term $W_n = \sum_{k=1}^{N} \mathrm{Var}\,(\hat{a}_{kn})$, a bias term $B_n^2 = \sum_{k=1}^{N} (E(\hat{a}_{kn}) - a_k)^2$

and an approximation error $R(N, f) = \sum_{k=N+1}^{\infty} a_k^2$, respectively. Here and in the sequel we suppress the dependence of $N = N(n)$ on n.

For the variance of $\hat{a}_{kn}$ we have $\text{Var}(\hat{a}_{kn}) = \sigma^2 n^{-1}(n^{-1}\sum_{i=1}^{n} v_k^2(x_i))$. The term in the brackets converges to 1 due to the equidistribution of x_i, $i = 1, \dots, n$. This term is uniformly bounded in k. Hence, $W_n \leq c_1 N/n$.

From the definition of $\hat{a}_{kn}$ we obtain

$$(E(\hat{a}_{kn}) - a_k)^2 = \left(n^{-1}\sum_{i=1}^{n} f(x_i)v_k(x_i) - \int_X f(x)v_k(x)dx\right)^2$$

from which $(E(\hat{a}_{kn}) - a_k)^2 \leq (\mathcal{V}(fv_k)D_n^*)^2$ follows by the Koksma inequality (or the Koksma-Hlavka inequality, see Kuipers and Niederreiter (1974), p.143), for $f \in V(X) \cap C(X)$.

It can be verified that $\mathcal{V}(fv_k) \leq \sup_{x\in X}|v_k(x)|\mathcal{V}(f) + \sup_{x\in X}|f(x)|\mathcal{V}(v_k)$, (see e.g., Fang and Wang (1994), p.63). For continuously differentiable v_k the total variation equals $\int_0^1 |v_k'(x)|dx$. Hence, for the trigonometric system we get $\mathcal{V}(v_k) = ck$.

Summarizing the above results we obtain $B_n^2 \leq c_2 N^3 (D_n^*)^2$ and, hence,

$$IMSE(\hat{f}_n, f) \leq c_1\frac{N}{n} + c_2\frac{N^3}{n^{2-\varepsilon}} + R(N, f).$$

The first term in this equation converges to zero due to the fact that condition $N^3/n^{2-\varepsilon} \to 0$ implies $N/n \to 0$. (5) ensures convergence of the second term to zero. Finally, also $R(N, f) \to 0$ as $N \to \infty$, since $f \in L_2(0, 1)$.

Proof of Proposition 2. Note that $fv_k \in W^{\mu}(0, 1)$ and that for the Lipschitz constant, $\mathcal{L}(fv_k) \leq \sup_{x\in[0,1]}|v_k(x)|\mathcal{L}(f) + \sup_{x\in[0,1]}|f(x)|\mathcal{L}(v_k)$. For the trigonometric system we have $\mathcal{L}(v_k) = 2\sqrt{2}\pi k$. This, together with (B2), implies $B_n^2 \leq cN^3(D_n^*)^{2\mu}$.

It remains to evaluate $R(N, f)$. Let $d_N(f)$ be the error of the best approximation of f by linear combinations of $\bar{v}_N = (v_1, \dots, v_N)$ in the supremum norm, $d_N(f) = \sup_{\gamma_N}||f - \gamma_N^T\bar{v}_N||_\infty$. Now $R(N, f) = \inf_{\gamma_N}||f - \gamma_N^T\bar{v}_N||_2^2 \leq \int_X(f(x) - \gamma_N^T\bar{v}_N(x))^2dx \leq d_N^2(f)$. For the trigonometric system it follows from Jackson's theorem (see Timan (1963)) that if f is periodic and $f \in W^{\mu}(0, 1)$ then $d_N(f) = O(N^{-\mu})$. From the above and (B3) resp. (B5), we get

$$IMSE(\hat{f}_n, f) \leq c_1\frac{N}{n} + c_2\frac{N^3}{n^{2\mu-\varepsilon}} + \frac{c_3}{N^{2\mu}}. \tag{8}$$

Choose $N(n)$ according to (6) with $d = 1$ to balance the convergence rate of the first and last term in (8), attaining the order $O(n^{-2\mu/(2\mu+1)})$. One can verify that the middle term on the r.h.s. of (8) converges to zero at the rate $O(n^{-\delta})$, where $\delta = (3-4\mu^2-2\mu+\varepsilon(2\mu+1))/(2\mu+1)$. For $\mu > \sqrt{3}/2$ this rate is faster than that attained by the first and the third terms. Thus, $IMSE(\hat{f}_n, f) = O(n^{-\frac{2\mu}{2\mu+1}})$ and the best rate of convergence is attained by $\hat{f}_n$.

References

Baker, A. (1984). *A Concise Introduction to the Theory of Numbers.* Cambridge University Press, Cambridge.

Bates, R. A., Riccomagno E., Schwabe R. and Wynn H. P. (1998). Lattices and dual lattices in optimal experimental design for Fourier models. *Comput. Statist. Data Anal.* **28**, 283-296.

Cheng, M., Hall, P. and Titterington, D. M. (1998). Optimal design for curve estimation by local linear smoothing. *Bernoulli* **4**, 3-14.

Cox, D. R. and Reid, N. (2000). *The Theory of the Design of Experiments.* Chapman and Hall CRC, Boca Raton.

Efromovich, S. (1999). *Nonparametric Curve Estimation.* Springer, New York.

Fang, K. T. and Wang, T. (1994). *Number-theoretic Methods in Statistics.* Chapman and Hall, London.

Ghosh, S. and Rao, C. R. (1996). *Design and Analysis of Experiments.* Handbook of Statistics 13. Elsevier, Amsterdam.

Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences.* Wiley, New York.

Linton, O. B. and Nielsen, J. B. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-100.

Müller, H.-G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Lett.* **2**, 285-290.

Müller, W. G. (1998). *Collecting Spatial Data* (*Optimum Design of Experiments for Random Fields*). Physica, Heidelberg.

Niederreiter, H. (1992). *Random Number Generation and Quasi–Monte Carlo Methods.* SIAM, Philadelphia.

Rafajłowicz, E. (1987). Nonparametric orthogonal series estimators of regression: a class attaining the optimal convergence rate in l_2. *Statist. Probab. Lett.* **5**, 219-224.

Stone, C. J. (1982). Optimal global rate of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 592-606.

Timan, A. F. (1963). *Theory of Approximation of Functions of a Real Variable.* Pergamon, Oxford.

Institute of Engineering Cybernetics, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland.

E-mail: raf@immt.pwr.wroc.pl

Otto von Guericke University, Institute for Mathematical Stochastics, PF 4120, D–39016 Magdeburg, Germany.

E-mail: rainer.schwabe@mathematik.uni-magdeburg.de

(Received February 2000; accepted July 2002)