New Developments and Applications in Experimental Design
IMS Lecture Notes – Monograph Series (1998) Volume 34

SOME APPLICATIONS OF CANONICAL MOMENTS IN FOURIER REGRESSION MODELS

BY HOLGER DETTE

Ruhr-Universität Bochum

This paper applies recent results on canonical moments for the determination of optimal designs for multivariate Fourier regression models. Optimal designs for discriminating between different Fourier regression models can be found explicitly. It is also demonstrated that these designs may be useful in orthogonal series estimation and for testing additivity in nonparametric regression. In contrast to many other optimality criteria for the trigonometric regression model, the discrimination designs are *not* necessarily uniformly distributed on equidistant points.

1. Introduction. Consider a Fourier regression model in q variables

$$g_{2\lambda}(x) = \sum_{\substack{0\le i_1,\dots,i_q\le\lambda \\ 0\le\sum i_j\le\lambda}} \beta_{i_1,\dots,i_q} \prod_{j=1}^{q} C_{i_j}(x_j) + \sum_{\substack{0\le i_1,\dots,i_q\le\lambda \\ 1\le\sum i_j\le\lambda}} \alpha_{i_1,\dots,i_q} \prod_{j=1}^{q} S_{i_j}(x_j) \tag{1.1}$$

where $q, \lambda \in \mathbb{N}, x \in [-\pi,\pi]^q$, $C_k(z) = \cos(kz) \quad (k = 0,\dots,\lambda); S_k(z) = \sin(kz) \quad (k = 1,\dots,\lambda)$ and $S_0(z) = 1$. A simple example is the case $\lambda = 2, q = 2$, where there are 11 regression functions of the form

$$1,\ \cos x,\ \cos y,\ \cos x\cos y,\ \cos(2x),\ \cos(2y),\ \sin x,\ \sin y,\ \sin x\sin y,\ \sin(2x),\ \sin(2y).$$

Functions of this type are frequently used in orthogonal series estimation of a regression function $g : [-\pi,\pi]^q \to \mathbb{R}$ [see e.g. Eubank (1988, p. 66), Müller (1988, p. 21 and the discussion in Section 2)] where the unknown parameters are estimated by least squares. It is well known that the crucial point in the application of these methods is the appropriate choice of the "smoothing parameter" λ or the degree of the regression [see e.g. Hart (1985)].

Received July 1997; revised March 1998.
AMS 1991 subject classifications. 62K05, 62G05.
Key words and phrases. Discrimination designs, orthogonal series estimation.

In this paper we consider the corresponding design problem in this context. More specifically, define

$$(1.2) \qquad g_{2\lambda-1}(x) = \sum_{\substack{0\le i_1,\ldots,i_q\le\lambda-1 \\ 0\le\sum i_j\le\lambda-1}} \beta_{i_1,\ldots,i_q} \prod_{j=1}^{q} C_{i_j}(x_j) + \sum_{\substack{0\le i_1,\ldots,i_q\le\lambda \\ 1\le\sum i_j\le\lambda}} \alpha_{i_1,\ldots,i_q} \prod_{j=1}^{q} S_{i_j}(x_j)$$

and consider the class of models

$$(1.3) \qquad \mathcal{F}_{2d} = \{g_0, g_1, \ldots, g_{2d}\}$$

where d is a given bound. Note that the model $g_{2\lambda-1}$ in (1.2) is obtained from $g_{2\lambda}$ in (1.1) by omitting the terms $\prod_{j=1}^{q} \cos(i_j x_j)$ of the form $\sum_{j=1}^{q} i_j = \lambda$. If we omit the corresponding sine terms in $g_{2\lambda-1}$ we obtain the model $g_{2\lambda-2}$ etc.

For the selection of the appropriate model in the class $\mathcal{F}_{2d}$ Anderson (1962) proposed to test successively the hypotheses

$$\begin{array}{llllll}
H_0^{(2d)}: & \beta_{i_1,\ldots,i_q} = 0 & \forall & i_1,\ldots,i_q: & \sum_{j=1}^{q} i_j = d \\
H_0^{(2d-1)}: & \alpha_{i_1,\ldots,i_q} = 0 & \forall & i_1,\ldots,i_q: & \sum_{j=1}^{q} i_j = d \\
H_0^{(2d-2)}: & \beta_{i_1,\ldots,i_q} = 0 & \forall & i_1,\ldots,i_q: & \sum_{j=1}^{q} i_j = d-1 \\
\vdots & \vdots & & \vdots & \vdots \\
H_0^{(1)}: & \alpha_{i_1,\ldots,i_q} = 0 & \forall & i_1,\ldots,i_q: & \sum_{j=1}^{q} i_j = 1
\end{array}$$

and to decide for the model g_{k_0} where k_0 is the first index for which $H_0^{(k_0)}$ is rejected. Anderson (1962) and Spruill (1990) proved several optimality properties of this procedure. Roughly speaking it minimizes the probability of the error of choosing a too high degree Fourier regression model. Under the assumption of normally distributed errors it is well known that the quality of the F-test for the hypothesis $H_0^{(\ell)} (\ell = 1, \ldots, 2d)$ depends on the given design, and a good discriminating design should improve the power of these tests.

In this paper we apply the theory of canonical moments [see Dette and Studden (1997)] in order to determine efficient designs for the discrimination between the different models of $\mathcal{F}_{2d}$. In the following section we present a concrete example from orthogonal series estimation in order to give some motivation for considering multivariate Fourier regression models and product designs on the q-dimensional cube. Section 3 introduces the corresponding design problem, while Section 4 demonstrates the application of canonical moments in this context. Finally, we present a solution of the discriminating design problem and illustrate the results by several examples. Especially we give some arguments for using uniform designs in orthogonal series estimation.

2. A motivating example. Our main interest in the models of the type (1.1) and (1.2) stems from the fact that these functions are used for orthogonal series estimation and certain tests of additivity in nonparametric regression [see e.g. Eubank (1988) and Eubank, Hart, Simpson and Stefanski (1995)]. In order to present a transparent discussion we consider in this section a regression model with two explanatory variables $(q = 2)$

$$y_{ij} = g(x_{1i}, x_{2j}) + \varepsilon_{ij} \quad i = 1, \ldots, n_1; j = 1, \ldots, n_2$$

where $\varepsilon_{11}, \ldots, \varepsilon_{n_1 n_2}$ are independent identically distributed random variables with mean 0 and positive variance σ^2. A typical problem in regression analysis is the estimation of the function g by nonparametric methods. It is well known that the efficiency of linear smoothers decreases rapidly with an increasing dimension of the explanatory variable. On the other hand, if $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$, the estimation error tends to 0 at the same rate as in the case of a single predictor [see Stone (1985)]. Therefore it is particularly important to perform a proper model check of additivity in a multivariate nonparametric regression. If the data confirms the hypothesis of additivity, further data analysis can be improved by using more efficient methods in the additive model. To be more precise we rewrite the regression function as

$$g(x_1, x_2) = f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2), \quad (x_1, x_2) \in [-\pi, \pi]^2,$$

then the problem of testing additivity is to check whether $f_{12} \equiv 0$ [see e.g. Barry (1993)]. A common assumption in these situations [see Barry (1993) or Eubank et al. (1995)] is that independent observations y_{ij} $i = 1, \ldots, n_1; j = 1, \ldots, n_2$ are available on a grid $\{(x_{1i}, x_{2j})\}_{i,j=1}^{n_1,n_2}$ with expectation $g(x_{1i}, x_{2j})$ and variance σ^2.

In order to estimate the regression function g it is common practice to use a two dimensional Fourier series estimation

$$\text{(2.1)} \quad \hat{g}_\lambda(x_1, x_2) = \sum_{0 \le j+k \le \lambda} \hat{\alpha}_{jk} \cos(jx_1)\cos(kx_2) + \sum_{2 \le j+k \le \lambda} \hat{\beta}_{jk} \sin(jx_1)\sin(kx_2)$$
$$\sum_{1 \le j \le \lambda} \hat{\gamma}_j \sin(jx_1) + \sum_{1 \le j \le \lambda} \hat{\delta}_j \sin(jx_2)$$

where the "sample Fourier" coefficients are defined by

$$\begin{aligned}
\hat{\alpha}_{jk} &= \frac{2}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} y_{rs} \cos(jx_{1r})\cos(kx_{2s}) \quad 1 \le j+k \le \lambda \\
\hat{\alpha}_{00} &= \frac{1}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} y_{rs} \\
\hat{\beta}_{jk} &= \frac{2}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} y_{rs} \sin(jx_{1r})\sin(kx_{2s}) \quad 1 \le j+k \le \lambda, \quad j, k \ge 1 \\
\hat{\gamma}_j &= \frac{2}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} y_{rs} \sin(jx_{1r}) \qquad 1 \le j \le \lambda \\
\hat{\delta}_j &= \frac{2}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} y_{rs} \sin(jx_{2s}) \qquad 1 \le j \le \lambda
\end{aligned}$$

and $n = n_1 n_2$ denotes the total sample size. Note that the expression in (2.1) contains also terms of the form $\cos(jx_1)$ and $\cos(kx_2)$ which are obtained by putting either

$j = 0$ or $k = 0$ in the first sum of (2.1). Eubank, Hart, Simpson, and Stefanski (1995) proposed a test of the interaction $f_{12}(x_1, x_2)$ which is based on a statistics of "sample Fourier coefficients" $\tilde{\alpha}_{jk}$ and $\tilde{\beta}_{jk}$ which are defined in a similar manner as $\hat{\alpha}_{jk}, \hat{\beta}_{jk}$. Note that we consider here the cube $[-\pi, \pi]^q$ as design space and we have to use a Fourier series estimator of the form $g_{2\lambda}$ or $g_{2\lambda-1}$ in contrast to Eubank, Hart, Simpson, and Stefanski (1995) who dealt with the problem on the cube $[0, \pi]^q$, where a Fourier series approximation, which involves only the cosine terms, is sufficient.

It is well known that the crucial point in the application of these methods is the appropriate choice of the smoothing parameter λ [see Hart (1985) or Eubank (1988)]. In order to obtain efficient estimation and testing procedures it is therefore particularly important to be able to distinguish between regression models of different degree λ. The designs constructed in the following sections will be discriminating designs for the class of Fourier regression models $\mathcal{F}_{2d}$ and should therefore be useful for estimating and testing the additivity of nonparametric regression functions.

3. The design problem. We consider approximate designs here, i.e. probability measures on the cube $[-\pi, \pi]^q$, with the interpretation that the observations are taken at the support points in proportion to the corresponding masses [see Kiefer (1974)]. From a mathematical point of view it turns out to be useful to restrict ourselves to the class of product designs, i.e.

$$\Xi = \{\sigma_1 \times \ldots \times \sigma_q | \sigma_j \text{ probability measure on } [-\pi, \pi], j = 1, \ldots, q\},$$

and it is demonstrated by Lim and Studden (1988) that designs of this type have extremely high efficiencies compared to the optimal design in the class of all measures on the cube. Moreover, as pointed out in Section 2, such designs are useful for the estimation or for testing the additivity of a nonparametric regression, where Fourier regression models are commonly applied. Let f_ℓ denote the vector of regression functions in the model g_ℓ [see (1.1) and (1.2)] and

$$M_\ell(\sigma) = \int_{[-\pi,\pi]^q} f_\ell(x) f_\ell^T(x) d\sigma(x) \quad 1 \le \ell \le 2d \tag{3.1}$$

the information matrix of a design σ on the cube. It is well known that for an exact design the volume of the ellipsoid of concentration for the parameters corresponding to the hypothesis $H_0^{(\ell)}$ is inverse proportional to

$$\delta_\ell(\sigma) = \left\{ \frac{|M_\ell(\sigma)|}{|M_{\ell-1}(\sigma)|} \right\}^{1/2} \quad 1 \le \ell \le 2d \tag{3.2}$$

and consequently a good discrimination design should make these quantities as large as possible. Following Atkinson and Donev (1992) we introduce the following composite optimality criterion for model discrimination.

DEFINITION 3.1 *Let $\pi = (\alpha_1, \beta_1, \ldots, \alpha_d, \beta_d)$ denote a prior for the class $\mathcal{F}_{2d}$, then a design σ is called optimal discriminating design for the class $\mathcal{F}_{2d}$ with respect to the* *prior π if it maximizes the function*

$$\Phi(\sigma) = \sum_{\ell=1}^{d} \left(\frac{\beta_\ell}{N_{q-1,\ell}} \log \frac{|M_{2\ell}(\sigma)|}{|M_{2\ell-1}(\sigma)|} + \frac{\alpha_\ell}{N_{q-1,\ell}} \log \frac{|M_{2\ell-1}(\sigma)|}{|M_{2\ell-2}(\sigma)|} \right) \tag{3.3}$$

where

$$N_{q-1,\ell} := \binom{q-1+\ell}{\ell}$$

denotes the number of parameters in the hypothesis $H_0^{(2\ell)}$ and $H_0^{(2\ell-1)}$; $\ell = 1, \ldots, d$.

It can easily be shown that Φ is a concave function and that $\Phi(\sigma) = \Phi(\sigma^*)$ where σ^* is obtained from σ by reflecting components at the origin. Consequently, by standard arguments of design theory there exists a symmetric optimal discriminating design and it is reasonable to restrict the optimization to the class of product measures with symmetric components, i.e.

$$\Xi_s := \{\sigma_1 \times \ldots \times \sigma_q | \sigma_j \text{ is symmetric on } [-\pi, \pi] \ \forall\, j = 1, \ldots, q\}.$$

If $\sigma \in \Xi_s$ and $f_{c,\ell}; f_{s,\ell}$ denote the vectors collecting the cosine and sine terms of $f_{2\ell}$, respectively, then it is easy to see that

$$\begin{aligned} |M_{2\ell}(\sigma)| &= |M_\ell^c(\sigma)||M_\ell^s(\sigma)| \\ |M_{2\ell-1}(\sigma)| &= |M_\ell^s(\sigma)||M_{\ell-1}^c(\sigma)| \end{aligned} \tag{3.4}$$

where

$$M_\ell^c(\sigma) = \int_{[-\pi,\pi]^q} f_{c,\ell}(x) f_{c,\ell}^T(x) d\sigma(x) \tag{3.5}$$

$$M_\ell^s(\sigma) = \int_{[-\pi,\pi]^q} f_{s,\ell}(x) f_{s,\ell}^T(x) d\sigma(x) \tag{3.6}$$

denote the information matrices in the submodels $f_{c,\ell}$ and $f_{s,\ell}$ ($\ell = 1, \ldots, d$). Consequently, the optimality criterion in (3.3) can be rewritten as

$$\Phi(\sigma) = \sum_{\ell=1}^{d} \frac{\beta_\ell}{N_{q-1,\ell}} \log \left(\frac{|M_\ell^c(\sigma)|}{|M_{\ell-1}^c(\sigma)|} \right) + \frac{\alpha_\ell}{N_{q-1,\ell}} \log \left(\frac{|M_\ell^s(\sigma)|}{|M_{\ell-1}^s(\sigma)|} \right) \quad (\sigma \in \Xi_s). \tag{3.7}$$

4. Fourier regression, polynomial models and canonical moments. It is well known that the set of symmetric probability measures on the interval $[-\pi, \pi]$ can be mapped in a one to one manner on the set of all designs on the interval $[-1, 1]$ by the transformation $t = \cos x$ [see e.g. Lau and Studden (1985)]. For a symmetric product measure $\sigma = \sigma_1 \times \ldots \times \sigma_q \in \Xi_s$ we therefore define a projection $\xi_\sigma = \xi_1 \times \ldots \times \xi_q$ by transforming each factor σ_j via $t = \cos x$ to ξ_j ($j = 1, \ldots, q$). Throughout this paper let $T_j(z)$ and $U_j(z)$ denote the jth Chebyshev polynomial of the first and second kind, then it is well known that

$$\begin{aligned} \cos(j \arccos z) &= T_j(z) \\ \sin(j \arccos z) &= \sqrt{1-z^2}\, U_{j-1}(z) \quad (z \in [-1, 1]) \end{aligned} \tag{4.1}$$

[see e.g. Rivlin (1990)]. This means that a projection of the product measure $\sigma \in \Xi_s$ on a measue on the cube $[-1,1]^q$ induces a transformation of the vector $f_{c,\ell}$ into a vector $h_{c,\ell}$ which contains the $N_{q,\ell}$ functions

$$\left\{ \prod_{j=1}^{q} T_{i_j}(x_j) | 0 \leq \sum_{j=1}^{q} i_j \leq \ell \right\}$$

and similarly $f_{s,\ell}$ is mapped into a vector $h_{s,\ell}$ containing the $N_{q,\ell}-1$ functions

$$\left\{ \prod_{j=1}^{q} (1-x_j^2)^{\alpha(i_j)/2} U_{i_j-1}(x_j) | 1 \leq \sum_{j=1}^{q} i_j \leq \lambda \right\}$$

where $U_{-1}(z) = 1$ and $\alpha(i_j) = 1$ or 0 according to $i_j \geq 1$ or $i_j = 0$. Because the determinant criterion is invariant with respect to reparametrizations it follows that the optimality criterion Φ in (11) can be rewritten as

$$\Phi(\sigma) = \sum_{\ell=1}^{d} \frac{\beta_\ell}{N_{q-1,\ell}} \log\left(\frac{|B_\ell(\xi_\sigma)|}{|B_{\ell-1}(\xi_\sigma)|}\right) + \frac{\alpha_\ell}{N_{q-1,\ell}} \log\left(\frac{|A_\ell(\xi_\sigma)|}{|A_{\ell-1}(\xi_\sigma)|}\right) + C \tag{4.2}$$

where the constant C does not depend on the design, $B_\ell(\xi_\sigma)$ and $A_\ell(\xi_\sigma)$ are the information matrices corresponding to the regression functions

$$\left\{ \prod_{j=1}^{q} x_j^{i_j} | 0 \leq \sum_{j=1}^{q} i_j \leq \ell \right\} \quad \text{and} \quad \left\{ \prod_{j=1}^{q} (1-x_j)^{\alpha(i_j)/2} x_j^{i_j-1} | 1 \leq \sum_{j=1}^{q} i_j \leq \ell \right\} \tag{4.3}$$

$(z^{-1} = z^0 := 1)$ respectively, and $\xi_\sigma = \xi_1 \times \ldots \times \xi_q$ denotes the projection of $\sigma = \sigma_1 \times \ldots \times \sigma_q$. Therefore the discriminating design problem is equivalent to a problem of designing an experiment for the discrimination between certain homo- and heteroscedastic multivariate polynomial regression models.

An important tool used in optimal design for polynomials is the theory of canonical moments which was introduced by Studden (1980, 1982a, 1982b, 1989) in this context. We will only give a very brief heuristical introduction into this concept which should be sufficient for the purpose of this paper. For more details we refer to the work of Lau (1983, 1988), Skibinsky (1986) or to the recent monograph of Dette and Studden (1997). It is well known that a probability measure on the interval $[-1,1]$, say ξ, is determined by its sequence of moments $(c_1, c_2, \ldots)$. Skibinsky (1967) defined a one to one mapping from the sequences of ordinary moments onto sequences $(p_1, p_2, \ldots)$ whose elements vary independently in the interval $[0,1]$. For a given probability measure ξ on the interval $[-1,1]$ the element p_j of the corresponding sequence is called the jth canonical moment ξ. In order to indicate the dependence on ξ we use at some places the notation $p_j(\xi)$. The dependence on the design is omitted whenever it is clear from the context. If j is the first index for which $p_j \in \{0,1\}$, then the sequence of canonical moments terminates at p_j and the measure is supported at a finite number of points. The support points and corresponding masses can be found explicitly by evaluating certain orthogonal polynomials [see Skibinsky (1986), Lau (1988) or Dette and Studden (1997, Chapter 3)]. The set of probability measures on the interval $[-1,1]$ with first k canonical moments equal to $(p_1,\dots,p_k) \in (0,1)^{k-1} \times [0,1]$ is a singleton if and only if $p_k \in \{0,1\}$. Otherwise there exists an uncountable number of probability measures corresponding to $(p_1,\dots,p_k)$ [see Skibinsky (1986) or Dette and Studden (1997)].

The following Lemma provides an explicit representation of the determinants $|B_\ell(\xi)|$ and $|A_\ell(\xi)|$ in terms of the canonical moments corresponding to the factors of the product design $\xi = \xi_1 \times \ldots \times \xi_q$ and is the basic ingredient for the solution of the optimal discriminating design problem.

LEMMA 4.1. *Let $\xi = \xi_1 \times \ldots \times \xi_q$ denote a product measure on the cube $[-1,1]^q$ and $p_i^j = p_i(\xi_j)$ the ith canonical moment of ξ_j $(j = 1,\dots,q)$, then*

$$|A_d(\xi)| = c_{q,d} \prod_{i=1}^{q} \prod_{\ell=1}^{d} (\gamma_{i,2\ell-1}\gamma_{i,2\ell})^{N_{q,d-\ell}} \tag{4.4}$$

$$|B_d(\xi)| = \tilde{c}_{q,d} \prod_{i=1}^{q} \prod_{\ell=1}^{d} (\zeta_{i,2\ell-1}\zeta_{i,2\ell})^{N_{q,d-\ell}} \tag{4.5}$$

where $\gamma_{j,1} = 1-p_1^j$, $\zeta_{j,1} = p_1^j$, $\gamma_{j,i} = (1-p_i^j)p_{i-1}^j$, $\zeta_{j,i} = (1-p_{i-1}^j)p_i^j$ $\quad (i \geq 2), j = 1,\dots,q.$

PROOF: By the preceeding discussion $B_\ell(\xi)$ is the information matrix of a product design in a multivariate polynomial regression of degree ℓ using the first set in (4.3) as regression functions. This problem was considered by Lim and Studden (1988) and the representation (4.5) follows from their Lemma 4.3 and Lemma 5.1.

For the first part define $d\tilde{\xi}_i(x_i) = \frac{1}{\mu_i}(1-x_i^2)d\xi_i(x_i)$, $\mu_i = \int_{-1}^1 (1-x_i^2)d\xi_i(x_i) = 4\gamma_{i,1}\gamma_{i,2}$ $(i = 1,\dots,q)$ and $\tilde{\xi} = \tilde{\xi}_1 \times \ldots \times \tilde{\xi}_q$. Then it is easy to see that $|A_d(\xi)|$ is essentially the determinant of the information matrix of $\tilde{\xi}$ in a multivariate polynomial regression of degree $d-1$, i.e. $|A_d(\xi)| = C|B_{d-1}(\tilde{\xi})| \left(\prod_{i=1}^q \mu_i\right)^{N_{q,d-1}}$. If $\tilde{p}_j^i, \tilde{\zeta}_{i,j}$ denote the quantities of Lemma 4.1 corresponding to the product measure $\tilde{\xi}$, then it follows from Corollary 1.5.6 in Dette and Studden (1997) that

$$\tilde{\zeta}_{i,2j-1}\tilde{\zeta}_{i,2j} = \gamma_{i,2j+1}\gamma_{i,2j+2}$$

$(j \geq 1, i = 1,\dots,q)$. Therefore assertion (4.5) shows

$$\begin{aligned}
|A_d(\xi)| &= C\left(\prod_{i=1}^{q} \mu_i\right)^{N_{q,d-1}} |B_{d-1}(\tilde{\xi})| \\
&= \tilde{c}_{q,d} \prod_{i=1}^{q} \left(\prod_{\ell=1}^{d-1} \gamma_{i,2\ell+1}\gamma_{i,2\ell+2}\right)^{N_{q,d-1-\ell}} \prod_{i=1}^{q} (\gamma_{i,1}\gamma_{i,2})^{N_{q,d-1}} \\
&= \tilde{c}_{q,d} \prod_{i=1}^{q} \prod_{\ell=1}^{d} (\gamma_{i,2\ell-1}\gamma_{i,2\ell})^{N_{q,d-\ell}}
\end{aligned}$$

which proves the remaining statement of the lemma. □

With the aid of Lemma 4.1 the optimal discriminating designs can be described explicitly in terms of the canonical moments of the corresponding projections $\xi_1, \ldots, \xi_q$.

THEOREM 4.2 *A design $\sigma = \sigma_1 \times \ldots \times \sigma_q \in \Xi_s$ is an optimal discriminating product design for the class of Fourier regression models $\mathcal{F}_{2d}$ with respect to the prior $\pi = (\alpha_1, \beta_1, \ldots, \alpha_d, \beta_d)$ if and only if the canonical moments $p_j^i = p_j(\xi_i)$ of the corresponding projection $\xi_\sigma = \xi_1 \times \ldots \times \xi_q$ satisfy*

$$p_{2\ell-1}^i = \frac{1}{2} \quad \ell = 1, \ldots, d$$

$$p_{2\ell}^i = \frac{\sum_{j=\ell}^d (\beta_j N_{q-1,j-\ell} + \alpha_j N_{q-1,j-\ell-1})/N_{q-1,j}}{\sum_{j=\ell}^d (\alpha_j + \beta_j)(N_{q-1,j-\ell} + N_{q-1,j-\ell-1})/N_{q-1,j}} \quad \ell = 1, \ldots, d$$

$(i = 1, \ldots, q)$ where $N_{q-1,-1} := 0$.

PROOF: Observing (4.2) and Lemma 4.1 it follows that the optimal product discriminating designs $\sigma = \sigma_1 \times \ldots \times \sigma_q$ have corresponding projection $\xi_\sigma = \xi_1 \times \ldots \times \xi_q$ such that all factors ξ_j $(j = 1, \ldots, q)$ have identical canonical moments up to the order $2d$. Therefore the optimality criterion in (4.2) reduces to

$$\exp\left[\frac{1}{q}\Phi(\sigma)\right] = C' \prod_{\ell=1}^{d} \prod_{j=1}^{\ell} \left[(\zeta_{2j-1}\zeta_{2j})^{\beta_\ell \frac{N_{q-1,\ell-j}}{N_{q-1,\ell}}} (\gamma_{2j-1}\gamma_{2j})^{\alpha_\ell \frac{N_{q-1,\ell-j}}{N_{q-1,\ell}}} \right]$$

where we used $N_{q,\ell-j} - N_{q,\ell-j-1} = N_{q-1,\ell-j}$, the common canonical moments of $\xi_1, \ldots, \xi_q$ are denoted by $\{p_j\}_{j\geq 1}$ and $\zeta_1 = p_1, \gamma_1 = 1 - p_1$ and $\zeta_j = (1 - p_{j-1})p_j, \gamma_j = (1 - p_j)p_{j-1}$ $(j \geq 2)$. The assertion now follows by a straightforward optimization, observing that the canonical moments vary independently in the interval $[0, 1]$. □

It is worthwhile to mention that the optimal discriminating designs are not necessarily unique. Theorem 4.2 characterizes the first $2d$ canonical moments of the corresponding projections and every product measure whose projections have the same canonical moments will also be an optimal discriminating (product) design for the class $\mathcal{F}_{2d}$ with respect to the prior π. On the other hand if $\alpha_d = 0$ or $\beta_d = 0$ it follows from Theorem 4.2 that $p_{2d}^i = 1$ or 0 $(i = 1, \ldots, q)$ and therefore all projections ξ_j are unique [see Dette and Studden (1997)]. In these cases there exists a unique symmetric optimal discriminating (product) design for the class $\mathcal{F}_{2d}$ with respect to the prior π.

Theorem 4.1 provides a complete solution of the discriminating design problem for the class of Fourier regression models in the set of all symmetric product measures on the cube $[-\pi, \pi]^q$. The corresponding projection ξ can be found from the optimal canonical moments by standard methods [see e.g. Dette and Studden (1997), Ch. 3] and the factors σ_j of the optimal product design $\sigma = \sigma_1 \times \ldots \times \sigma_q$ are obtained from ξ_j via the projection $t = \cos x$. We will illustrate this method in the following section.

5. Applications. In this section we demonstrate the application of the results of Section 4 in two examples. The first example is directly related to a given 2-dimensional

Fourier regression model while the second one demonstrates a potential application of our results for estimating and testing additivity of a nonparametric regression as described in Eubank, Hart, Simpson and Stefanski (1995).

5.1. *Fourier regression of degree 2 in two variables.* Consider the case $d = 2; q = 2$ introduced in Section 1. The projection $\xi = \xi_1 \times \xi_2$ of the optimal discriminating product design $\sigma = \sigma_1 \times \sigma_2$ for the class $\mathcal{F}_4 = \{g_1, g_2, g_3, g_4\}$ is characterized by the property that the canonical moments up to the order 4 of ξ_1 and ξ_2 are $p_1 = p_3 = 1/2$ and

$$p_2 = \frac{1}{3}\frac{3\beta_1 + 4\beta_2 + 2\alpha_2}{1 + \alpha_2 + \beta_2}, \quad p_4 = \frac{\beta_2}{\alpha_2 + \beta_2}.$$

Every design ξ_1 with these first four canonical moments corresponds via $\xi = \xi_1 \times \xi_1$ and the projection $t = \cos x$ to an optimal discriminating design $\sigma = \sigma_1 \times \sigma_1$. Uniqueness occurs if and only if $\beta_2 = 0$ or $\alpha_2 = 0$ because in this case $p_4 = 0$ or $p_4 = 1$. For example, if we assume a uniform prior for $\mathcal{F}_4$, i.e. $\alpha_j = \beta_j = 1/4 \quad j = 1, 2$, we obtain $p_j^* = 1/2 \quad j = 1, \dots, 4$. Terminating this sequence with $p_5 = \frac{1}{2}, p_6 = 0$ yields a three point design ξ_1 on $[-1, 1]$ with equal masses at the points $-\sqrt{3/4}, 0, \sqrt{3/4}$. The transformation $t = \cos x$ of this design to a symmetric measure on $[-\pi, \pi]$ shows that the design $\sigma_1^* \times \sigma_1^*$ with

$$\sigma_1^* = \begin{Bmatrix} -\frac{5}{6}\pi & -\frac{3}{6}\pi & -\frac{1}{6}\pi & \frac{1}{6}\pi & \frac{3}{6}\pi & \frac{5}{6}\pi \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{Bmatrix} \tag{5.1}$$

is an optimal discriminating design with respect to the uniform prior. Alternatively we could terminate at any other index $j \geq 5$, which gives a sequence of canonical moments $(p_1^*, p_2^*, p_3^*, p_4^*, p_5, \dots, p_{j-1}, p_j)$ where $p_k \in (0, 1) \quad (k = 5, \dots, j-1)$ and $p_j \in \{0, 1\}$. We can also obtain designs with an infinite number of support points by continuing $(p_1^*, p_2^*, p_3^*, p_4^*)$ with an infinite sequence of canonical moments. All these designs will be optimal because the first four canonical moments of the corresponding projection are given by $1/2, 1/2, 1/2, 1/2$. For example, the uniform distribution on $[-\pi, \pi]^2$ is an optimal discriminating (product) design because the corresponding projection onto $[-1, 1]^2$ via $t = \cos x$ yields a product of arcsin-distributions and these have canonical moments $p_j \equiv 1/2$ for all $j \geq 1$.

As an example where the optimal design is unique, consider the prior $\pi^* = (\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$. In this case the canonical moments of the optimal discriminating design are given by $p_1 = p_3 = 1/2, p_2 = 7/12, p_4 = 1$ which corresponds to the unique optimal projection $\xi^* = \xi_2^* \times \xi_2^*$ determined by the measure ξ_2^* with masses $7/24, 5/12, 7/24$ at the points $-1, 0, 1$. The optimal discriminating (product) design for the class $\mathcal{F} = \{g_1, g_2, g_4\}$ with respect to the prior π^* is given by $\sigma_2^* \times \sigma_2^*$ where (σ_2^* is obtained from ξ_2^* via $t = \cos x$ as)

$$\sigma_2^* = \begin{Bmatrix} -\pi & -\frac{\pi}{2} & 0 & \frac{\pi}{2} & \pi \\ \frac{7}{48} & \frac{5}{24} & \frac{7}{24} & \frac{5}{24} & \frac{7}{48} \end{Bmatrix} \tag{5.2}$$

5.2. *Good designs for testing additivity.* As pointed out in Section 2 the optimal product discriminating designs may be useful for the estimation and certain tests of additivity in nonparametric regression. For the sake of simplicity we consider again the case $q = 2$. Additionally we note now, that a common assumption in this situation is that the grid of the design points $\{(x_{1i}, x_{2j})\}_{i,j=1}^{n_1,n_2}$ is generated by two positive densities h_1 and h_2 on the interval $[-\pi, \pi]$, such that

$$\int_{-\pi}^{x_{ij}} h_i(u)du = \frac{2j-1}{2n_i} \quad j = 1, \ldots, n_i \quad i = 1, 2 \tag{5.3}$$

(note that we have adjusted the assumptions of Eubank et al. (1995) to the interval $[-\pi, \pi]$). In this case repeated observations in the marginal distributions of the design are not possible. An interesting question is now, which is an appropriate design density for the nonparametric estimation of the regression and for the testing procedure of Eubank et al. (1995). Because the main problem is the choice of an appropriate λ in (2.1), this is essentially the problem of finding an optimal discriminating design for the class of Fourier regression models $\{g_{2j} | j \geq 1\}$. Note that the number of observations is given by $n = n_1 n_2$ and as a consequence the number of parameters in the model is limited. Therefore we consider the discrimination design problem for the class

$$\tilde{\mathcal{F}} = \{g_2, \ldots, g_{2m}\}$$

where $m = \min\{n_1, n_2\} - 1$ [see also Eubank et al. (1995)] which corresponds to the discriminating design problem for the class $\mathcal{F}_{2m}$ with respect to a uniform prior on the models of even degrees, i.e. $\alpha_\ell = 0$, $\ell = 1, \ldots, m; \beta_\ell = 1/m$, $\ell = 1, \ldots, m$. By Theorem 4.2 the optimal discriminating product design is given by $\sigma^* \times \sigma^*$ where the corresponding projection ξ^* of σ^* has canonical moments

$$p_{2\ell} = \frac{1}{2} \cdot \frac{\sum_{j=\ell}^{m} \frac{j-\ell+1}{j+1}}{\sum_{j=\ell}^{m} \frac{j-\ell+1/2}{j+1}} \quad \ell = 1, \ldots, m-1 \quad (p_{2m} = 1)$$

and $p_{2\ell-1} = \frac{1}{2}$ $(\ell = 1, \ldots, m)$. Note that in this case the optimal discriminating design is unique [see Dette and Studden (1997)]. In order to obtain a design with a continuous density [as required by (5.3)] we note that $p_{2\ell} \to 1/2$ as $m \to \infty$ and as consequence, for moderate sample size, it makes sense to replace ξ^* by the arcsine distribution ξ_a on $[-1, 1]$ which satisfies $p_j(\xi_a) = 1/2$ for all $j \geq 1$. Note that $\xi_a \times \xi_a$ is also the projection of an optimal discriminating design for the class $\mathcal{F}_{2d}$ with respect to the uniform prior on all models for any $d \in \mathbb{N}$. This follows readily from Theorem 4.2 which shows that for $\alpha_\ell = \beta_\ell = \frac{1}{2d}$ $(\ell = 1, \ldots, d)$ the canonical moments of the corresponding projection satisfy $p_\ell = 1/2$ $(\ell = 1, \ldots, 2d)$ [see also the discussion in Section 5.1]. Moreover, ξ_a is the projection of the uniform distribution on $[-\pi, \pi]$ via the mapping $t = \cos x$. Consequently, an efficient choice of the densities h_1 and h_2 in (5.3) is the uniform measure on $[-\pi, \pi]$. For this reason we propose as an efficient design for the orthogonal series estimation and for testing the additivity of a nonparametric regression on the cube $[-\pi, \pi]^2$ the equidistant grid of the form

$$\left\{ \pi\left(\frac{2i-1}{n_1} - 1\right), \pi\left(\frac{2j-1}{n_2} - 1\right) \right\}_{i=1,\ldots,n_1}^{j=1,\ldots,n_2}.$$

Acknowledgement. The author is grateful to two referees for very helpful comments.

REFERENCES

ANDERSON, T.W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist.* **33** 255–265.

ATKINSON, A. C. AND DONEV, A. N. (1992). *Optimum Experimental Designs.* Oxford University Press, New York.

BARRY, D. (1993). Testing for additivity of a regression function. *Ann. Statist.* **21** 235-254.

DETTE, H. AND STUDDEN, W.J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis.* Wiley, New York.

EUBANK, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

EUBANK, R.L., HART, J.D., SIMPSON, D.G. AND STEFANSKI, L.S. (1995). Testing for additivity in nonparametric regression. *Ann. Statist.* **23** 1896–1920.

HART, J.D. (1985). On the choice of a truncation point in a Fourier series density estimation. *J. Statist. Comput. Simul.* **21** 95–116.

KIEFER, J.C. (1974). General equivalence theory for optimum designs. *Ann. Statist.* **2** 849-879.

LAU, T.S. (1983). Theory of canonical moments and its applications in polynomial regression I, II (Technical Reports # 83-23, # 83-24). Purdue University, West Lafayette

LAU, T.S. (1988). D-optimal designs on the unit q-ball. *J. Statist. Plan. Inf.***19** 299–315.

LAU, T.S. AND STUDDEN, W.J. (1985). Optimal designs for trigonometric and polynomial regression. *Ann. Statist.* **13** 383–394.

LIM, Y.B. AND STUDDEN, W.J. (1988). Efficient D_S-optimal designs for multivariate polynomial regression on the q-cube. *Ann. Statist.* **16** 383–393.

MÜLLER, H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer, New York.

RIVLIN, T.J. (1990). *Chebyshev Polynomials.* Wiley, New York.

SKIBINSKY, M. (1967). The range of the $(n+1)$th moment for distributions on $[0,1]$. *J. Appl. Prob.* **4** 543–552.

SKIBINSKY, M. (1986). Principal representations and canonical moment sequences for distributions on an interval. *J. Math. Anal. Appl.* **120** 95–120.

SPRUILL, M.C. (1990). Good designs for testing the degree of a polynomial mean. *Sankhya B* **52** 67–74.

STONE, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

STUDDEN, W.J. (1980). D_s-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* **8** 1132–1141.

STUDDEN, W.J. (1982a). Some robust-type D-optimal designs in polynomial regression. *J. Amer. Statist. Assoc.* **77** 916–921.

STUDDEN, W.J. (1982b). Optimal designs for weighted polynomial regression using canonical moments. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta, and J.O. Berger, eds.), 335-350. Academic Press, New York.

STUDDEN, W.J. (1989). Note on some Φ_p-optimal design for polynomial regression. *Ann. Statist.* **17** 618-623.

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY