Optimal designs for statistical analysis with Zernike polynomials

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 3214 559

Viatcheslav B. Melas
St. Petersburg State University
Department of Mathematics
St. Petersburg
Russia
email: v.melas@pobox.spbu.ru

Andrey Pepelyshev
St. Petersburg State University
Department of Mathematics
St. Petersburg
Russia
email: andrey@ap7236.spbu.ru

June 16, 2006

Abstract

The Zernike polynomials arise in several applications such as optical metrology or image analysis on a circular domain. In the present paper we determine optimal designs for regression models which are represented by expansions in terms of Zernike polynomials. We consider two estimation methods for the coefficients in these models and determine the corresponding optimal designs. The first one is the classical least squares method and Φ_p-optimal designs in the sense of Kiefer (1974) are derived, which minimize an appropriate functional of the covariance matrix of the least squares estimator. It is demonstrated that optimal designs with respect to Kiefer's Φ_p-criteria ($p > -\infty$) are essentially unique and concentrate observations on certain circles in the experimental domain. E-optimal designs have the same structure but it is shown in several examples that these optimal designs are not necessarily uniquely determined. The second method is based on the direct estimation of the Fourier coefficients in the expansion of the expected response in terms of Zernike polynomials and optimal designs minimizing the trace of the covariance matrix of the corresponding estimator are determined. The designs are also compared with the uniform designs on a grid, which is commonly used in this context.

AMS Subject Classification: 62K05

Keywords and Phrases: Optimal design, Zernike polynomials, image analysis, D-optimality, E-optimality

1 Introduction

Consider a function f defined on the unit disc

$$D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\},$$

which belongs to the space $L^2(D)$ and the problem of estimating this function from the experimental data

$$Z_{ij} = f(x_i, y_j) + \varepsilon_{ij} \quad (1 \leq i \leq m; 1 \leq j \leq n) \tag{1.1}$$

where $(x_i, y_j) \in D$ are the experimental conditions and the random variables ε_{ij} are uncorrelated with zero mean and common variance. The problem of recovering the function f from the data $(Z_{ij})_{1 \leq j \leq m}$ on a circular domain arises in a wide range of applications including the diffraction theory of optical aberrations [see Born and Wolf (1975) or Wyant and Kreath (1992)], pattern recognition [see e.g. Abu-Mostafa and Psaltis (1984)], image analysis [see e.g. Kim and Kim

(1999) or Liao and Pawlak (1996)] and statistical models for circular data [see Fisher (1993)]. Numerous authors propose to expand the function f in (1.1) in terms of Zernike polynomials (or Zernike moments) and to estimate the coefficients in this expansion from the experimental data [see e.g. Liao and Pawlak (1998), Pawlak and Liao (2002), Hse and Newton (2004) among many others]. On the one hand the estimated coefficients of this expansion are used for reconstructing the unknown function [see e.g. Liao and Pawlak (1998)] and on the other hand these coefficients are also used for classification by the commonly used techniques as support vector machines, minimum distance or nearest neighbour classification [see e.g. Hse and Newton (2004)].

In the present paper we study the effect of the design [i.e. the choice of the location of the points (x_i, y_j) in the disc D] on the quality of the estimates for the coefficients in the expansion of the function f in terms of Zernike polynomials. In Section 2 we state some basic facts about these polynomials and explain very briefly how these functions are used in the disciplines mentioned in the previous paragraph. We consider two methods for estimating the coefficients in a truncated expansion of the regression function in terms of Zernike polynomials. The first one is the classical least squares method. In Section 3 we determine optimal designs with respect to Kiefer's Φ_p-criteria, which minimize a p-mean of the eigenvalues of the covariance matrix of the least squares estimate. It is shown that an optimal allocation of the explanatory variables (x_i, y_j) uses certain circles in the experimental domain D, and the number of circles increases with the degree in the expansion of f. We present several examples and compare the Φ_p-optimal designs with the uniform design, which is commonly used in these problems. In Section 4 we determine optimal designs if the coefficients in the truncated expansion in terms of Zernike polynomials are estimated directly by replacing the theoretical Fourier coefficients by a weighted mean of the observations. In this case the design appears nonlinearly in the (asymptotic) covariance matrix of the corresponding estimator and Φ_p-optimal designs have to be determined numerically in most cases. However, for the A-optimality criterion (corresponding to the case $p = -1$) the optimal designs can be found explicitly and are also compared with the uniform design.

2 Zernike moments

Zernike polynomials were first proposed by Zernike (1934) and are defined by

$$Z_n^{-m}(\rho,\varphi) = N_n^m R_n^m(\rho)\sin(m\varphi), \quad Z_n^m(\rho,\varphi) = N_n^m R_n^m(\rho)\cos(m\varphi), \tag{2.1}$$

where the function $R_n^m(\rho)$ is called radial function and defi ned for $n, m \in \mathbb{N}_0$ with $n \geq m \geq 0$ by

$$(2.2) \qquad R_n^m(\rho) = (-1)^{(n-m)/2}\rho^m P^{(m,0)}_{(n-m)/2}(1-2\rho^2)$$

$$= \begin{cases} \displaystyle\sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-l)!}{l!((n+m)/2-l)!((n-m)/2-l)!}\rho^{n-2l} & \text{for } n-m \text{ even} \\ 0 & \text{for } n-m \text{ odd} \end{cases}$$

and $P_k^{(\alpha,\beta)}(x)$ denotes the k-th Jacobi polynomial orthogonal with respect to the measure $(1-x)^\alpha(1+x)^\beta I_{[-1,1]}(x)\,dx$ [see Szegö (1975)]. In formula (2.1) the quantity

$$(2.3) \qquad N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}},$$

is a normalizing constant such that

$$(2.4) \qquad \frac{1}{\pi}\int_0^{2\pi}\int_0^1 Z_{n_1}^{m_2}(\rho,\varphi)Z_{n_2}^{m_1}(\rho,\varphi)\rho\,d\rho\,d\varphi = \delta_{n_1,n_2}\delta_{m_1,m_2}$$

for all $n_1, n_2 \in \mathbb{N}_0, 1 \leq |m_j| \leq n_j$ $(j = 1,2)$, where $n_j - m_j$ is even $(j = 1,2)$. Here and throughout this paper $\delta_{i,j}$ denotes Kronecker's symbol. Note also that the orthogonality relation for the polynomials $R_n^m(\rho)$ is given by

$$(2.5) \qquad \int_0^1 \rho R_{n_1}^m(\rho)R_{n_2}^m(\rho)\,d\rho = \frac{\delta_{n_1,n_2}}{2(n_1+1)}$$

[see Szegö (1975)]. A function $f \in L^2(D)$ admits the expansion

$$(2.6) \qquad f(\rho,\varphi) = \sum_{k=0}^{\infty}\sum_{\substack{i=-k \\ k-|i|\text{even}}}^{k} \theta_{(k,i)} Z_k^i(\rho,\varphi), \quad \rho \in [0,1], \varphi \in [0,2\pi),$$

where the quantities $\theta_{(k,i)}$ are the usual Fourier coefficients given by

$$(2.7) \qquad \theta_{(k,i)} = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\varphi)Z_k^i(\rho,\varphi)\rho\,d\rho\,d\varphi \quad \text{if } k-|i| \text{ is even}$$

Assume that data according to the model (1.1) is available, where the explanatory variables (x_i, y_j) are represented in terms of polar coordinates (ρ_i, φ_j). Usually a truncated expansion of (2.6) up to a given order, say $d \in \mathbb{N}$, is used as an approximation of the function f and one way to estimate the coefficients in this expansion (and as a by-product the approximation of the function f) is the least squares method, which determines the parameter $\theta = (\theta_{(0,0)}, \theta_{(1,-1)}, \theta_{(1,1)}, \ldots, \theta_{(d,-d)}, \ldots, \theta_{(d,d)})^T \in \mathbb{R}^{(d+1)(d+2)/2}$ such that

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \Big(Z_{ij} - \sum_{k=0}^{d} \sum_{\substack{\ell=-k \\ k-|\ell| \text{ even}}}^{k} \theta_{(k,\ell)} Z_k^\ell(\rho_i, \varphi_j) \Big)^2 \tag{2.8}$$

becomes minimal. A second method of estimating the coefficients in the truncated expansion uses the direct estimation of the Fourier coefficients (2.7), that is

$$\tilde{\theta}_{(k,\ell)} = \frac{1}{\pi} \sum_{i=1}^{m} \sum_{j=1}^{n} Z_k^\ell(\rho_i, \varphi_j) \rho_i (\rho_i - \rho_{i-1})(\varphi_j - \varphi_{j-1}) Z_{ij} \tag{2.9}$$

($\ell = 0, \ldots d; |k| \leq \ell, \ell - |k|$ even), see Pawlak and Liao (2002), for example. The estimated coefficients in this expansion are finally used for several purposes, such as estimation of function f in the context of image reconstruction [see e.g. Pawlak and Liao (2002)] or symbol recognition by support vector machines, minimum mean distance or nearest neighbour methods [see Hse and Newton (2004)]. A common design in this context is to observe the data on a grid in the disc D [see Pawlak and Liao (2002)]. In the present paper we consider the problem of finding optimal designs for the estimation of the parameters in a truncated Fourier expansion of order $d \in \mathbb{N}$ derived from the infinite series (2.6). We first state some results for least squares estimation in Section 3 and consider afterwards the problem of finding efficient designs for the direct estimation of the Fourier coefficients in Section 4.

3 Optimal designs for least squares estimation in Zernike regression models

Consider the regression model corresponding to the least squares problem (2.8)

$$E[Y \mid \rho, \varphi] = \theta^T f_d(\rho, \varphi); \quad \text{Var}[Y \mid \rho, \varphi] = \sigma^2 > 0, \tag{3.1}$$

where

$$f_d(\rho, \varphi) = \Big(Z_0^0(\rho, \varphi), Z_1^{-1}(\rho, \varphi), Z_1^1(\rho, \varphi), \ldots, Z_d^{-d}(\rho, \varphi), \ldots, Z_d^d(\rho, \varphi) \Big)^T \in \mathbb{R}^{(d+1)(d+2)/2} \tag{3.2}$$

is the vector of Zernike polynomials of order d and

$$\theta = (\theta_{(0,0)}, \theta_{(1,-1)}, \theta_{(1,1)}, \theta_{(2,-2)}, \theta_{(2,0)}, \theta_{(2,2)}, \ldots, \theta_{(d,-d)}, \ldots \theta_{(d,d)})^T \in \mathbb{R}^{(d+1)(d+2)/2}$$

the corresponding vector of parameters. Note that there appear $(d+1)(d+2)/2$ Zernike polynomials in the regression model (3.1). In the present section an approximate design is a probability measure, say ξ, on the set $[0,1] \times [0, 2\pi]$ (note that we will present the design in polar coordinates). This concept is due to Kiefer (1974) and appropriate for least squares estimation. In the following Section 4 we will consider a slightly different definition of a design which is more appropriate for the purpose of direct estimation of the Fourier coefficients. For a probability measure with finite support the support points, say (ρ_i, φ_j), determine the points where observations are taken and the corresponding weights, say $w_{i,j}$, give the relative proportion of total observations, taken at the point (ρ_i, φ_j). For a given design ξ with finite support the covariance matrix of the least squares estimate for the vector θ is proportional to the inverse of the information matrix

$$M(\xi) = \int_0^{2\pi} \int_0^1 f(\rho, \varphi) f^T(\rho, \varphi)\, d\xi(\rho, \varphi), \tag{3.3}$$

and an optimal approximate design maximizes an appropriate function of this matrix. There are numerous criteria proposed in the literature, which can be used to discriminate between competing designs [see Silvey (1980) or Pukelsheim (1993)], and in this paper we will restrict ourselves to the famous family of Φ_p-criteria introduced by Kiefer (1974). Let $-\infty \leq p < 1$; following Kiefer (1974) we call the design ξ_p^* Φ_p-optimal for estimating the parameter θ in the regression model (3.1) if ξ_p^* maximizes the expression

$$\Phi_p(\xi) = \Big(\frac{2}{(d+1)(d+2)} \operatorname{tr}(M^p(\xi)) \Big)^{\frac{1}{p}} \tag{3.4}$$

among all designs with non-singular information matrix. Note that the cases $p = 0$ and $p = -\infty$ correspond to the frequently used D- and E-optimality criterion, that is

$$\begin{aligned} \Phi_0(\xi) &= |M(\xi)|^{2/(d+1)(d+2)}, \\ \Phi_{-\infty}(\xi) &= \lambda_{\min}(M(\xi)), \end{aligned}$$

respectively, where $\lambda_{\min}(B)$ denotes the minimum eigenvalue of the matrix B. Our main result of this section describes the structure of Φ_p-optimal designs for the least squares estimation of the coefficients in the regression model (3.1) with Zernike polynomials as regression functions. For this define for $0 \leq r \leq 1$ the measure $U(r)$ as the uniform distribution on the circle with radius r and center 0, that is

$$U(r) \sim \mathcal{U}(\{(\rho, \varphi) \in D \mid \rho = r\}).$$

The following result shows that Φ_p-optimal designs are specific convex combinations of designs of this type.

Theorem 3.1. *There exist radii* $0 \leq r_1 < r_2 < \ldots < r_{\lfloor d/2 \rfloor +1} = 1$ *and positive weights* $w_1, \ldots, w_{\lfloor d/2 \rfloor +1}$ *with* $\sum_{i=1}^{\lfloor d/2 \rfloor +1} w_i = 1$, *such that any* Φ_p*-optimal design* ξ_p^* *in the regression model (3.1) is of the form*

$$\xi_p^* = \sum_{i=1}^{\lfloor d/2 \rfloor +1} w_i U(r_i).$$

If d is even, we have $r_1 = 0$, *(in other words: observations have to be taken at the center), while* $r_1 > 0$ *if d is odd. Moreover, if* $p > -\infty$ *the radii* r_i *and weights* w_i *are uniquely determined.*

Proof. Note that the design problem is rotation invariant. In other words: Let ξ denote a Φ_p-optimal design on the disc D with support points (ρ_i, φ_j), weights w_{ij}, and assume that $\tilde{\xi}$ is obtained from ξ by rotating all support points of ξ with an angle of size α, then it follows for the vector of regression functions

$$f(\rho, \varphi + \alpha) = T f(\rho, \varphi) \tag{3.5}$$

where the matrix T is given by

$$T = \operatorname{diag}(T_0, T_1, \ldots, T_d) \in \mathbb{R}^{(d+1)(d+2)/2 \times (d+1)(d+2)/2} \tag{3.6}$$

with blocks $T_j \in \mathbb{R}^{(j+1)\times(j+1)}$ defined by

$$T_{2i+1} = \begin{pmatrix} \cos(2i+1)\alpha & 0 & 0 \ldots & 0 & 0 & \ldots & 0 & -\sin(2i+1)\alpha \\ 0 & \cos(2i-1)\alpha & 0 \ldots & 0 & 0 & \ldots & \sin(2i-1)\alpha & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 \ldots & \cos\alpha & -\sin\alpha & \ldots & 0 & 0 \\ 0 & 0 & 0 \ldots & \sin\alpha & \cos\alpha & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \sin(2i-1)\alpha & 0 \ldots & 0 & 0 & \ldots & \cos(2i-1)\alpha & 0 \\ \sin(2i+1)\alpha & 0 & 0 \ldots & 0 & 0 & \ldots & 0 & \cos(2i+1)\alpha \end{pmatrix}$$

$$T_{2i} = \begin{pmatrix} \cos 2i\alpha & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & -\sin 2i\alpha \\ 0 & \cos(2i-2)\alpha & 0 & \dots & 0 & 0 & 0 & \dots & \sin(2i-2)\alpha & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \cos 2\alpha & 0 & -\sin 2\alpha & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & \sin 2\alpha & 0 & \cos 2\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \sin(2i-2)\alpha & 0 & \dots & 0 & 0 & 0 & \dots & \cos(2i-2)\alpha & 0 \\ \sin 2i\alpha & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \cos 2i\alpha \end{pmatrix}.$$

It is easy to see that the matrix T is orthogonal and from (3.5) it follows that

$$M(\tilde{\xi}) = TM(\xi)T^T.$$

Consequently, the matrix $M(\tilde{\xi})$ has the same eigenvalues as $M(\xi)$ and $\tilde{\xi}$ is also Φ_p-optimal. We finally note that the group of matrices of the form (3.6) is a subgroup of the orthogonal group acting on $\mathbb{R}^{(d+1)(d+2)/2}$.

A straightforward calculation shows that the vector of regression function can be written as

$$f(\rho, \varphi) = K^{-1}g(x_1, x_2) = K^{-1}g(x) \tag{3.7}$$

where g is the vector of $\binom{d+2}{2}$ monomials of the form $x_i^{\alpha_i}x_j^{\alpha_j}$ $(i, j \in \{1, 2\}; \alpha_1, \alpha_2 \in \mathbb{N}_0; \alpha_1 + \alpha_2 \leq 2)$, $x_1 = \rho\cos\varphi$, $x_2 = \rho\sin\varphi$ and $K \in \mathbb{R}^{(d+1)(d+2)/2\times(d+1)(d+2)/2}$ is a suitable matrix which does not depend on the explanatory variable $(x_1, x_2) \in D$. Therefore the problem of finding a Φ_p-optimal design for the Zernike regression model (3.1) is equivalent to the problem of finding a Φ_p-optimal design for the parameter $K^T\theta$ in the two dimensional polynomial regression model

$$y = \theta^T g(x) + \varepsilon \tag{3.8}$$

on the "2-dimensional ball" D, which has been studied by numerous authors [see Box and Hunter (1957), Kiefer (1961), Galil and Kiefer (1977) among many others]. By the previous paragraph there exists a rotatable Φ_p-optimal design, say η_p^*, for this problem and by the general equivalence theorem for Φ_p-optimality [see Pukelsheim (1993), p. 180] it follows for $p > -\infty$ that a design η_p^* is Φ_p-optimal if and only if the inequality

$$d(x, \eta_p^*) = g^T(x)\bar{M}^{-1}(\eta_p^*)K(K^T\bar{M}^{-1}(\eta_p^*)K)^{-p-1}K^T\bar{M}^{-1}(\eta_p^*)g(x) \leq \operatorname{tr}(K^T\bar{M}^{-1}(\eta_p^*)K)^{-p} \tag{3.9}$$

holds for all $x \in D$, where $\bar{M}(\eta)$ is the information matrix of the design η in the model (3.8). In the remaining case $p = -\infty$ the characterization of the $\Phi_{-\infty}$-optimal designs is slightly different. Here the E-optimality of the design $\eta^*_{-\infty}$ is equivalent to the existence of a matrix E with $\operatorname{tr} E = 1$ such that the inequality

$$(3.10)\quad g^T(x)\bar{M}^{-1}(\eta^*_{-\infty})K(K^T\bar{M}^{-1}(\eta^*_{-\infty})K)^{-1}E(K^T\bar{M}^{-1}(\eta^*_{-\infty})K)^{-1}K^T\bar{M}^{-1}(\eta^*_{-\infty})g(x) \\ \leq \lambda_{\min}((K^T\bar{M}^{-1}(\eta^*_{-\infty})K)^{-1})$$

holds for all $x \in D$ [see Pukelsheim (1993), p. 182]. We can now adapt the arguments of the proof of Theorem 3.2.1 in Kiefer (1961) to the present situation. Note that Kiefer (1961) considered the D-optimality criterion ($p = 0$) and the case $K = I_{(d+1)(d+2)/2}$, for which the inequality in (3.9) reduces to

$$g^T(x)\bar{M}^{-1}(\eta^*_0)g(x) \leq \binom{d+2}{2},$$

but the arguments can be directly transferred to the function $d(x, \eta^*_p)$ in (3.9) and (3.10). For example, it is easy to see that for a rotatable design η^*_p the function $d(x, \eta^*_p)$ depends only on the radius $\rho = \sqrt{x_1^2 + x_2^2}$ and is a polynomial of degree $2d$ with positive coefficient of ρ^{2d}, which is equal to the constant $\operatorname{tr}(K^T\bar{M}^{-1}(\eta^*_p)K)^{-p}$ (if $p > -\infty$) or $\lambda_{\min}((K^T\bar{M}^{-1}(\eta^*_p)K)^{-1})$ (if $p = -\infty$) at the support points of η^*_p.. A careful counting of the roots of this polynomial (including their multiplicities) now shows that an invariant Φ_p-optimal design concentrates masses at exactly $\lfloor \frac{d}{2} \rfloor + 1$ radii $0 \leq r_1 \leq \ldots \leq r_{\lfloor \frac{d}{2} \rfloor + 1} \leq 1$. The remaining statements of Theorem 3.1 follow similarly as in Kiefer (1961) and their proof is omitted for the sake of brevity.

□

Remark 3.2. Note that the uniform distribution $U(r)$ on a circle with radius $r \in [0, 1]$ and origin $(0, 0)$ is not directly implementable. However, it is easy to see that each uniform distribution $U(r_k)$ in Theorem 3.1 can be replaced by a discrete uniform measure with masses $1/m$ at $m \geq d$ points

$$\{(r \cos \frac{2\pi k}{m}, r \sin \frac{2\pi k}{m}) \mid k = 0, \ldots, m-1\}.$$

In general the determination of Φ_p-optimal designs for the Zernike-regression model (3.1) has to be performed numerically. In the following discussion we present some explicit examples for the D- ($p = 0$) and E-optimality criterion ($p = -\infty$).

3.1 D-optimal designs for least squares estimation

Recall the definition of the vector of regression functions $f(\rho, \varphi)$ in (3.2) and let $U(r)$ denote a uniform distribution on the circle with radius r and center $(0,0)$. It is easy to see that the matrix

$$I = I(\rho) = \int f(\rho,\varphi) f^T(\rho,\varphi) dU(r)(\rho,\varphi)$$

depends only on the squared radius r^2 and that there exists a permutation matrix, say P, such that this matrix can be represented as

$$PIP^T = \mathrm{diag}\{B_{-d}, \dots, B_{-1}, B_0, B_1, \dots, B_d\},$$

where the blocks B_i are defined by

$$B_i = B_i(r) = \int h_i(\rho,\varphi) h_i^T(\rho,\varphi) dU(r)(\rho,\varphi) \in \mathbb{R}^{\lfloor (d-|i|)/2 \rfloor \times \lfloor (d-|i|)/2 \rfloor}$$

with

$$h_i(\rho,\varphi) = \Big(\tilde{R}^i_{|i|}(\rho,\varphi), \tilde{R}^i_{|i|+2}(\rho,\varphi), \dots, \tilde{R}^i_{|i|+2\lfloor (d-|i|)/2 \rfloor}(\rho,\varphi) \Big)^T \in \mathbb{R}^{\lfloor (d-|i|)/2 \rfloor}$$

$(i = -d, -d+1, \dots, d)$ and $\tilde{R}^j_i(\rho,\varphi) = N^j_i R^j_i(\rho,\varphi)$. Therefore we obtain for the design $\xi = \sum_{k=1}^{\lfloor d/2 \rfloor + 1} w_k U(r_k)$

$$PM(\xi)P = \mathrm{diag}\Big(\sum_{k=1}^{\lfloor d/2 \rfloor + 1} w_k B_i(r_i) \Big)_{i=-d}^{d},$$

and observing the relation $B_{-i}(r) = B_i(r)$ it follows for the determinant of the information matrix

$$\det M(\xi) = \det\Big(\sum_{k=1}^{\lfloor d/2 \rfloor + 1} w_k B_0(r_k) \Big) \prod_{i=1}^{d} \Big(\det\Big(\sum_{k=1}^{\lfloor d/2 \rfloor + 1} w_k B_i(r_k) \Big) \Big)^2. \tag{3.11}$$

With this representation the D-optimal designs for least squares estimation in the Zernike regression model can be calculated by standard software such as Mathematica or Matlab and some D-optimal designs are given in Table 1.

It is of interest to compare the performance of the commonly used uniform design on the disc D with the D-optimal designs determined in this section. For this purpose we present in the last column of Table 1 the D-efficiencies

$$\mathrm{eff}_D(\xi_{\mathrm{uni}}) = \Big(\frac{|M(\xi_{\mathrm{uni}})|}{|M(\xi_0^*)|} \Big)^{2/(d+1)(d+2)}$$

of the uniform design ξ_{uni}. We observe a remarkable improvement by using D-optimal designs. Thus the application of these designs will yield a substantiable more accurate estimation of the parameters in the regression model (3.1).

Table 1: *D-optimal designs for the Zernike regression model (3.1). The D-optimal design ξ_0^* is of the form $\sum_{k=1}^{\lfloor d/2\rfloor+1} w_k U(r_k)$, where $U(r_k)$ is the uniform distribution on a circle with center $(0,0)$, radius r_k and the quantities w_k, r_k are given in the Table. The table also shows the D-efficiencies of the uniform design (right column).*

d		ξ_0^*			$\mathrm{eff}_D(\xi_{\mathrm{uni}})$
1	r_k	1			
	w_k	1			0.6300
2	r_k	0	1		
	w_k	1/6	5/6		0.5707
3	r_k	0.5155	1		
	w_k	0.3077	0.6923		0.5785
4	r_k	0	0.6784	1	
	w_k	0.0667	0.3439	0.5894	0.5801
5	r_k	0.3522	0.7739	1	
	w_k	0.1534	0.3354	0.5112	0.5910

3.2 E-optimal designs for least squares estimation

For the E-optimality criterion the situation is slightly more complicated because the E-optimal design is not necessarily unique. The following result characterizes an important class of E-optimal designs.

Theorem 3.3. *If ξ^* denotes a design such that its information matrix (3.3) in the Zernike regression model (3.1) is given by the identity matrix $I_{(d+1)(d+2)/2}$, then ξ^* is E-optimal. In particular the following two designs are E-optimal for the Zernike regression model*

(a) The uniform distribution on the disc D

(b) Any design of the form

$$\xi^* = \sum_{i=1}^{n} w_i U(r_i), \tag{3.12}$$

such that the radii $r_1, \ldots, r_n$ and weights $w_1, \ldots, w_n$ define a quadrature formula, which integrates polynomials up to the degree $2d$ with respect to the measure xdx on the interval

$[0,1]$ *exactly, i.e.*

$$\sum_{i=1}^{n} w_i r_i^k = \int_0^1 x^k x \, dx = \frac{1}{k+2}, \quad k = 0, \dots, 2d.$$

Proof. By the equivalence theorem for E-optimality [see Pukelsheim (1993), p. 182] it follows that a design ξ^* is E-optimal in the Zernike regression model (3.1), if and only if there exists a matrix E with trace $E = 1$ such that the inequality

$$f^T(\rho, \varphi) E f(\rho, \varphi) \leq \lambda_{\min}(M(\xi^*)) = 1 \tag{3.13}$$

holds for all $(\rho, \varphi) \in [0,1] \times [0, 2\pi]$ (note that $M(\xi^*) = I_{(d+1)(d+2)/2}$ by assumption). Now it is easy to see that the inequality (3.13) holds with $E = e_1 e_1^T$, where $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^{(d+1)(d+2)/2}$ denotes the first unit vector, and consequently any design with information matrix equal to $I_{(d+1)(d+2)/2}$ is E-optimal for Zernike regression model (3.1). This proves the first assertion. Observing the orthonormality relations in (2.4) it is easy to see that the uniform distribution on the disc D has this property, which proves part (a) of the second assertion. Finally, if ξ denotes a design as described in part (b) of Theorem 3.3, then it follows from (2.5) that the measure μ with masses w_i at the points r_i $(i = 1, \dots, n)$ satisfies

$$\sum_{i=1}^{n} w_i r_i R_{n_1}^m(r_i) R_{n_2}^m(r_i) = \int_0^1 \rho R_{n_1}^m(\rho) R_{n_2}^m(\rho) \, d\rho = \frac{\delta_{n_1 n_2}}{2(n_1+1)}.$$

Now a straightforward calculation shows that the information matrix of the corresponding design ξ^* in (3.12) is also given by the identity matrix, which proves the remaining statement of the theorem.

$\Box$

Example 3.4. It is worthwhile to mention that in general the information matrix of an E-optimal design for the Zernike regression model (3.1) is not uniquely determined. In other words it can be shown numerically that there exist E-optimal designs whose information matrices are not equal to the identity matrix.

Consider for example the case $d = 1$, then

$$f_1(\rho, \varphi) = (1, 2\rho \sin \varphi, 2\rho \cos \varphi)^T$$

and note that the uniform measure $U(r)$ on a circle with radius $r \in (0,1]$ has information matrix

$$M(U(r)) = I(r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2r^2 & 0 \\ 0 & 0 & 2r^2 \end{pmatrix}$$

Consequently, any design $U(r)$ with $r \geq 1/\sqrt{2}$ is E-optimal. Similarly, if $d = 2$ we have

$$f_2^T(\rho,\varphi) = (1, 2\rho\sin(\varphi), 2\rho\sin(\varphi), \sqrt{3}(2\rho^2-1), \sqrt{6}\rho^2\sin(2\varphi), \sqrt{6}\rho^2\cos(2\varphi))^T,$$

and the design

$$\xi_{w,r_1,r_2} = wU(r_1) + (1-w)U(r_2)$$

has information matrix

$$M(\xi_{w,r_1,r_2}) = P \begin{pmatrix} wB_0(r_1) + (1-w)B_0(r_2) & 0 \\ 0 & wB_1(r_1) + (1-w)B_1(r_2) \end{pmatrix} P,$$

where the matrices $B_i(r)\,(i = 1,2)$ are defined by

$$\begin{aligned} B_0(r) &= \begin{pmatrix} 1 & \sqrt{3}(2r^2-1) \\ \sqrt{3}(2r^2-1) & 3(2r^2-1)^2 \end{pmatrix}, \\ B_1(r) &= \operatorname{diag}\{2r^2, 2r^2, 3r^4, 3r^4\}, \end{aligned}$$

and P is an appropriate permutation matrix. If $(2r_1^2-1)w + (2r_2^2-1)(1-w) = 0$, and the conditions

$$\begin{aligned} 2r_1^2 w + 2r_2^2(1-w) &= 1, \\ 3(2r_1^2-1)^2 w + 3(2r_2^2-1)^2(1-w) &\geq 1, \\ 3r_1^4 w + 3r_2^4(1-w) &\geq 1, \end{aligned} \tag{3.14}$$

are satisfied, then it follows that the minimum eigenvalue of the matrix $M(\xi_{w,r_1,r_2})$ is 1 with multiplicity 3. Therefore the design ξ_{w,r_1,r_2} is E-optimal in this case.

In general the class of all E-optimal designs is difficult to describe, because relations of the type (3.14) become intractable for the case $d \geq 3$. Therefore we conclude this section with some numerical results which give the E-optimal designs with maximal determinant, that is

$$\begin{aligned} \xi^*_{\mathrm{DE}} &= \operatorname{argmax}\{|M(\xi)| \mid \xi \text{ is a design with } \lambda_{\min}(M(\xi)) = \operatorname{argmin}_\eta \lambda_{\min}(M(\eta))\} \\ &= \operatorname{argmax}\{|M(\xi)| \mid \xi \text{ is a design with } \lambda_{\min}(M(\xi)) = 1\} \end{aligned} \tag{3.15}$$

Some results for this type of design are given in Table 2. Note that these designs differ substantially from the D-optimal designs listed in Table 1. The E-optimal designs with maximal determinant put more mass in the central region of the disc.

Table 2: *E-optimal designs for the Zernike regression model with maximal determinant [see equation (3.15) for their definition].*

d		ξ^*_{DE}		
1	r_k	1		
	w_k	1		
2	r_k	0	1	
	w_k	1/2	1/2	
3	r_k	0.577	1	
	w_k	0.749	0.251	
4	r_k	0.0944	0.7098	1
	w_k	0.1738	0.6606	0.1656
5	r_k	0.3890	0.7987	1
	w_k	0.3688	0.5165	0.1147

4 Optimal designs for direct estimation of Fourier coefficients

Although least squares estimation is optimal with respect to minimizing the covariance matrix in the class of all unbiased (linear) estimators of the parameter θ, it could be numerically unstable because it relies on an inversion of a matrix of size $(d+1)(d+2)/2 \times (d+1)(d+2)/2$. In some applications as image analysis [see Pawlak and Liao (2002)] or symbol recognition [see Hse and Newton (2004)] the dimension in the truncated Fourier expansion could be rather large and therefore many authors propose to estimate the coefficients in this expansion (2.6) more directly using the definition of the Fourier coefficients $\theta_{(i,k)}$ in (2.7). In order to consider optimal design problems for this alternative estimation method we consider a slightly different notation of an approximate design, based on a concept introduced by Sacks and Ylvisaker (1970). In contrast to the previous section, where the design induced by these points was characterized by a uniform distribution on the set $\{(\rho_i, \varphi_j) \mid 1 \leq i \leq m; 1 \leq j \leq n\}$ we now characterize the design by a positive density h. To be precise let $(\rho_i, \varphi_j)_{i=1,\dots,m}^{j=1,\dots,n}$ in $[0,1] \times [0, 2\pi]$ denote the polar coordinates of the design points, where observations are taken and assume that

$$\frac{ij}{mn} = \int_0^{\rho_i} \int_0^{\varphi_j} h(\rho, \varphi)\, d\rho\, d\varphi \tag{4.1}$$

for a positive density h on $[0,1] \times [0, 2\pi]$. The function h is called design density and describes the optimal design. The coefficient $\theta_{(\ell,k)}$ in the expansion (2.6) is estimated by $(\rho_0 = 0, \varphi_0 = 0)$

$$\tilde{\theta}_{(\ell,k)} = \frac{1}{\pi} \sum_{i=1}^{m} \sum_{j=1}^{n} Z_\ell^k(\rho_i, \varphi_j) \rho_i (\rho_i - \rho_{i-1})(\varphi_j - \varphi_{j-1}) Z_{ij} \tag{4.2}$$

$(\ell = 0, \dots, d; |k| \leq \ell; \ell - |k|$ even). Note that Pawlak and Liao (2002) used a slightly different estimator. In the following we consider the problem of determining an optimal design, that is a design density h minimizing a certain functional of the covariance matrix of the estimator $\tilde{\theta} = (\tilde{\theta}_{(0,0)}, \tilde{\theta}_{(1,-1)}, \tilde{\theta}_{(1,1)}, \dots, \tilde{\theta}_{(d,-d)}, \dots, \tilde{\theta}_{(d,d)})^T$ for the parameter θ in the truncated Fourier expansion (3.1). For this we note that asymptotically the covariance matrix of this estimator is given by

$$\mathrm{Cov}(\tilde{\theta}) = \frac{\sigma^2}{nm\pi^2} N(h), \tag{4.3}$$

where the matrix $N(h)$ depends on the design density h and is defined as

$$N(h) = \Big(\int_0^1 \int_0^{2\pi} \frac{f_i(\rho, \varphi) f_j(\rho, \varphi) \rho^2}{h_1(\rho) h_2(\varphi)}\, d(\rho, \varphi) \Big)_{i,j=0}^{(d+1)(d+2)/2},$$

$f(\rho, \varphi) = (f_0(\rho, \varphi), \dots, f_{(d+1)(d+2)/2}(\rho, \varphi))^T$ is the vector of Zernike polynomials of order d, and h_1 and h_2 denote the marginal distributions of the design density h, that is

$$h_1(\rho) = \int_0^{2\pi} h(\rho, \varphi)\, d\varphi, \;\; h_2(\varphi) = \int_0^1 h(\rho, \varphi)\, d\rho.$$

An optimal design for the direct estimation of the Fourier coefficients (specified by the density h) maximizes an appropriate (convex) function of the matrix $N^{-1}(h)$, which is (asymptotically) proportional to the inverse of the covariance matrix of the estimator

$$\tilde{\theta} = (\tilde{\theta}_{(0,0)}, \tilde{\theta}_{(1,-1)}, \tilde{\theta}_{(1,1)}, \dots, \tilde{\theta}_{(d,-d)}, \dots, \tilde{\theta}_{(d,d)})^T).$$

Note that the design density h appears nonlinearly in the (asymptotic) covariance matrix of the estimator $\tilde{\theta}$, and therefore Φ_p-optimal designs for the direct estimation of the Fourier coefficients have to be found numerically in most cases of practical interest. However, for the A-optimality criterion ($p = -1$) an explicit solution of the optimal design problem is possible.

Theorem 4.1. *The A-optimal design for the direct estimation of the Fourier coefficients in the Zernike model (3.2) minimizes* $\operatorname{tr} N(h)$ *with respect to the density h on* $[0,1]\times[0,2\pi]$ *and is given by any design h^* with marginal densities $h_1^*(\rho)$ and $h_2^*(\varphi)$, where h_2 is the density of the uniform distribution on the interval* $[0,2\pi]$,

$$h_1^*(\rho) = \frac{\rho\sqrt{\Theta_d(\rho)}}{\int_0^1 \rho\sqrt{\Theta_d(\rho)}\,d\rho},$$

the function Θ_d is defined by

$$\begin{aligned}\Theta_d(\rho) &= 4\sum_{k=0}^{d}\frac{(\lfloor\frac{d-k}{2}\rfloor+1)(\lfloor\frac{d-k}{2}\rfloor+k+1)}{(2\lfloor\frac{d-k}{2}\rfloor+k+2)}\rho^{2k}\\ &\qquad \times\Big(P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor}(x)\frac{d}{dx}P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor+1}(x)-P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor+1}(x)\frac{d}{dx}P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor}(x)\Big)\Big|_{x=1-2\rho^2}\end{aligned}$$

and $P^{(\alpha,\beta)}(x)$ denotes the k-th Jacobi polynomial orthogonal on the interval with respect to the measure $(1-x)^\alpha(1+x)^\beta dx$.

Proof. From the definition of the regression functions in the Zernike regression model (3.1) it follows that

$$\begin{aligned}\sum_{\ell=0}^{d}\sum_{k=-\ell}^{\ell}(Z_\ell^k(\rho,\varphi))^2 &= \sum_{\ell=0}^{d}\sum_{k=0}^{\ell}(N_\ell^k R_\ell^k(\rho))^2 = \sum_{\ell=0}^{d}\sum_{\substack{k=0\\ \ell-|k|\text{ even}}}^{\ell}(N_\ell^k)^2\rho^{2k}\Big(P^{(k,0)}_{\frac{\ell-k}{2}}(1-2\rho^2)\Big)^2\\ &= \sum_{k=0}^{d}2\rho^{2k}\sum_{m=0}^{\lfloor\frac{d-k}{2}\rfloor}\frac{2m+k+1}{1+\delta_{k,0}}\Big(P_m^{(k,0)}(1-2\rho^2)\Big)^2\\ &= 4\sum_{k=0}^{d}\frac{(\lfloor\frac{d-k}{2}\rfloor+1)(\lfloor\frac{d-k}{2}\rfloor+k+1)}{(2\lfloor\frac{d-k}{2}\rfloor+k+2)}\rho^{2k}\\ &\qquad\times\Big\{\frac{d}{dx}P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor+1}(x)P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor}(x)-P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor+1}(x)\frac{d}{dx}P^{(k,0)}_{\lfloor\frac{d-k}{2}\rfloor}(x)\Big\}_{x=1-2\rho^2}\\ &= \Theta_d(\rho),\end{aligned}$$

where $\Theta_d(\rho)$ is defined in Theorem 4.1 and we have used the Christoffel-Darboux formula for the Jacobi polynomials [see Szegö (1975) p. 71)]. Therefore we obtain

$$\begin{aligned}\operatorname{tr}(N(h)) &= \sum_{i=0}^{(d+1)(d+2)/2} \int_0^1 \int_0^{2\pi} \frac{f_i^2(\rho,\varphi)\rho^2}{h_1(\rho)h_2(\varphi)} d(\rho,\varphi) \\ &= \int_0^1 \int_0^{2\pi} \frac{\rho^2\Theta_d(\rho)}{h_1(\rho)h_2(\varphi)} d(\rho,\varphi) = \int_0^1 \frac{\rho^2\Theta_d(\rho)}{h_1(\rho)} d\rho \int_0^{2\pi} \frac{d\varphi}{h_2(\varphi)}\end{aligned}$$

and by Cauchy's inequality it follows that

$$\operatorname{tr}(N(h)) \geq \Big(\int_0^1 \rho\sqrt{\Theta_d(\rho)}\, d\varphi\Big)^2 (2\pi)^2$$

with equality if and only if

$$h_1(\rho) = \frac{\rho\sqrt{\Theta_d(\rho)}}{\int_0^1 \rho\sqrt{\Theta_d(\rho)}\, d\rho} = h_1^*(\rho)$$

and

$$h_2(\varphi) = \frac{1}{2\pi} I_{[0,2\pi]}(\varphi)$$

$\Box$

Note that the optimal density h^* in Theorem 4.1 is only specified by its marginal distributions. If a product design with marginals h_1^* and h_2^* is used, it is invariant with respect to rotations. In Figure 1 we show the function h_2^* for $d = 30$. We observe numerically that a limiting density exists, which can be used to construct optimal designs for the direct estimation of the Fourier coefficients in the Zernike regression model (3.1) with a very large degree.

In practice the degree d of the expansion in (3.1) and the number of total observations for estimating the regression function f, say nm, will be fixed. In this case we obtain from the assumption (4.1) and the optimal design density h^* a concrete design $(\rho_i, \varphi_j)_{i=1,\dots,n}^{j=1,\dots,m}$ as follows. Because the second component of the optimal design is a uniform distribution on the disc we use a discrete uniform distribution, say $\varphi_j = \frac{2\pi j}{m}$ $(j = 1, \dots, m)$, for the angle φ. Finally the values

$$\frac{i}{n} = \int_0^{\rho_i} h_1^*(\rho)\, d\rho \quad i = 1, \dots, n$$

are used for the radii $\rho_1, \dots, \rho_n$. This implementation of the A-optimal design for the direct estimation of the coefficients in the Zernike regression model is also concentrated on centered circles in the disc D. In Table 3 and 4 we show the optimal radii for $n = 5$ and $n = 10$ and various values of d.

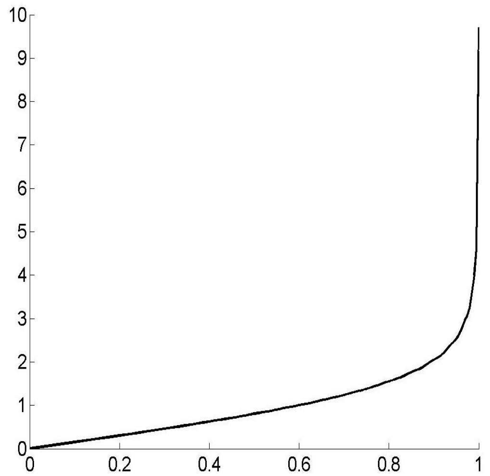

Figure 1: *The optimal design density h_d defined in Theorem 4.1 for $d = 30$.*

d	n	r_1	r_2	r_3	r_4	r_5
5	5	0.5088	0.7046	0.8412	0.9412	1
15	5	0.5078	0.7036	0.8408	0.9406	1
30	5	0.5076	0.7032	0.8402	0.9400	1

Table 3. *Optimal designs for direct estimation of the Fourier coefficients for $n = 10$ and various values of d.*

d	n	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
5	10	0.3614	0.5088	0.6188	0.7046	0.7768	0.8412	0.8972	0.9412	0.9744	1
15	10	0.3624	0.5078	0.6162	0.7036	0.7774	0.8408	0.8946	0.9406	0.9770	1
30	10	0.3622	0.5076	0.6156	0.7032	0.7770	0.8402	0.8944	0.9400	0.9768	1

Table 4. *Optimal designs for direct estimation of the Fourier coefficients for $n = 5$ and various values of d.*

Note that the A-optimal designs require a relatively large number of different radii in order to take observations close to the center of the disc D. For example, if $n = 10$ the minimal radius is given by $r_1 = 0.3614$ and no observations would be taken in the interior of the disc with this radius. In applications as image analysis or classification n can usually be chosen rather large such that this restriction causes no problems. It is also of interest to compare the A-optimal designs with the uniform design on the disc by its efficiency

$$\mathrm{eff}_A(h) = \Big(\frac{\mathrm{tr}(h^*)}{\mathrm{tr}(h)}\Big)^{-1}$$

(here h^* denotes an A-optimal design for the direct estimation of the Fourier coefficients). For the Zernike regression model of order $d = 5$, 10, 15 this efficiency is given by 0.912, 0.895, 0.892, respectively. This indicates that the uniform design is rather efficient for the direct estimation of the Fourier coefficients.

Acknowledgements. The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen") is gratefully acknowledged. The work by V.B.Melas was partly supported by the Russian Foundation of Basic Research (project 04-01-00519). The authors are also grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

References

Y.S. Abu-Mostafa, D. Psaltis (1984). Recognitive aspects of moment invariants. IEEE Trans. Pattern Anal. Machine Intell. 6, 698-706.

M. Born, E. Wolf (1975). Principles of Optics. Oxford, U.K., Pergamon Press.

G.E.P. Box, J.S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist. 28, 195-241.

N.I. Fisher (1993). Statistical Analysis of Circular Data. Cambridge, U.K., Cambridge Univ. Press.

Z. Galil, J. Kiefer (1977). Comparison of rotatable designs for regression on balls, I (quadratic). J. Statist. Plann. Inf. 1, 27-40.

H. Hse, A.R. Newton (2004). Sketched symbol recognition using Zernike moments. IEEE Pattern Recognition, 17th International Congress, 367-370.

J. Kiefer (1961). Optimum experimental designs V, with applications to systematic and rotatable designs. Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. 1, 381-405.

J. Kiefer (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849-879.

W.J. Kim, J.S. Kim (1999). Robust rotation angle estimator. IEEE Trans. Pattern Anal. Machine Intell. 21, 768-773.

S. Liao, M. Pawlak (1996). On image analysis by moments. IEEE Trans. Pattern Anal. Machine Intell. 18, 254-266.

S.X. Liao, M. Pawlak (1998). On the accuracy of Zernike moments for image analysis. IEEE Trans. Pattern Anal. Machine Intell. 20, 1358-1364.

F. Pukelsheim (1993). Optimal Design of Experiments. Wiley, N.Y.

M. Pawlak, S.X. Liao (2002). On the recovery of a function on a circular domain. IEEE Trans. Inform. Theory 48, 2736-2753.

J. Sacks, D. Ylvisaker (1970). Designs for regression problems for correlated errors. Ann. Math. Statist., 41, 2057-2074.

S.D. Silvey (1980). Optimal Design. Chapman and Hall, London.

G. Szegö (1975). Orthogonal polynomials. Am. Math. Soc. Colloqu. Publ. 23. Providence, RI.

J.C. Wyant, K. Creath (1993). Basic wavefront aberration theory for optical metrology. Applied Optics and Optical Engineering, Vol. XI, 2-58.

F. Zernike (1934). Diffraction theory of the cut procedure and its improved form, the phase contrast method (in German). Physica 1, 689-704.