

Dette, Holger; Wiens, Douglas P.

Working Paper

Robust designs for 3D shape analysis with spherical harmonic descriptors

Technical Report, No. 2007,12

Provided in Cooperation with:

Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund

Suggested Citation: Dette, Holger; Wiens, Douglas P. (2007) : Robust designs for 3D shape analysis with spherical harmonic descriptors, Technical Report, No. 2007,12, Universität Dortmund, Sonderforschungsbereich 475 - Komplexitätsreduktion in Multivariaten Datenstrukturen, Dortmund

This Version is available at:

<https://hdl.handle.net/10419/24997>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Robust designs for 3D shape analysis with spherical harmonic descriptors

Holger Dette¹, Douglas P. Wiens²

¹ Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

e-mail: holger.dette@ruhr-uni-bochum.de

² Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta
Canada T6G 2G1
e-mail: doug.wiens@ualberta.ca

March 22, 2007

Abbreviated title Robust designs for 3D shape analysis

Abstract Spherical harmonic descriptors are frequently used for describing three-dimensional shapes in terms of Fourier coefficients corresponding to an expansion of a function defined on the unit sphere. In a recent paper Dette, Melas and Pepelysheff (2005) determined optimal designs with respect to Kiefer's Φ_p -criteria for regression models derived from a truncated Fourier series. In particular it was shown that the uniform distribution on the sphere is Φ_p -optimal for spherical harmonic descriptors, for all $p > -1$. These designs minimize a function of the variance-covariance matrix of the least squares estimate but do not take into account the bias resulting from the truncation of the series.

In the present paper we demonstrate that the uniform distribution is also optimal with respect to a minimax criterion based on the mean square error, and as a consequence these designs are robust with respect to the truncation error. Moreover, we also consider heteroscedasticity and possible correlations in the construction of the optimal designs. These features appear naturally in 3D shape analysis, and the uniform design again turns out to be minimax robust against erroneous assumptions of homoscedasticity and independence.

Key words and phrases Shape analysis; spherical harmonic descriptors; optimal designs; mean square error; 3D-image data; minimax optimal designs; robust designs; dependent data

AMS 2000 Subject Classification: Primary 62K05, 62D32; secondary 62J05.

1. Introduction

Spherical harmonic shape descriptors are widely used to visualize 3D data in many fields including medicine, chemistry, architecture, agriculture and biology because of their ability to describe and compare shapes of various structures in terms of a relatively small number of parameters [see e.g. Brechbühler, Gerig and Kübler (1995), Novotni and Klein (2003), Székely, Kelemen, Brechbühler and Gerig (1996), Ding, Nesumi, Takano and Ukai (2000), Funkhouser, Min, Kazhdan, Chen, Halderman and Dobkin (2003), Kazhdan, Funkhouser and Rusinkiewicz (2003) among many others]. In many cases 3D data appears in the form $\{Y_i\}_{i=1}^n = \{r(\psi_i) + error_i\}_{i=1}^n$, where

$$\psi_i = (\theta_i, \phi_i) \in \mathcal{S} = \{(\theta, \phi) \mid \theta \in [0, \pi], \phi \in (-\pi, \pi]\}, \quad (1.1)$$

and where θ_i and ϕ_i denote the polar angle and azimuthal angle of the i^{th} observation [in other words the corresponding point of the shape has spherical coordinates $(Y_i \sin \theta_i \cos \phi_i, Y_i \sin \theta_i \sin \phi_i, Y_i \cos \theta_i)^T$].

Let

$$\{Y_l^m(\boldsymbol{\psi}) \mid m \in \{-l, -l+1, \dots, l\}; l \in \mathbb{N}_0\}$$

denote a complete orthonormal basis with respect to the uniform distribution

$$\mu(\boldsymbol{\psi}) d\boldsymbol{\psi} = \frac{\sin \theta}{4\pi} d\theta d\phi \quad (1.2)$$

on the unit sphere. If $r : S \rightarrow \mathbb{R}$ denotes a (square integrable) function representing the radius at direction $\boldsymbol{\psi}$, the coefficients

$$c_l^m = \frac{1}{4\pi} \int_0^\pi \int_{-\pi}^\pi r(\boldsymbol{\psi}) Y_l^m(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}$$

of a spherical harmonic expansion

$$r(\boldsymbol{\psi}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_l^m Y_l^m(\boldsymbol{\psi}) \quad (1.3)$$

are estimated from the 3D data. Because n data points do not allow one to determine all coefficients in the expansion (1.3), the series is truncated at a specific level, say d , and the coefficients c_l^m in this approximation of r are estimated. In this paper we concentrate on the (weighted) least squares criterion

$$\min_{c_l^m} \left\{ \sum_{i=1}^n \left(Y_i - \sum_{l=0}^d \sum_{m=-l}^l c_l^m Y_l^m(\boldsymbol{\psi}_i) \right)^2 w(\boldsymbol{\psi}_i) \right\} \quad (1.4)$$

with an appropriate non-negative weight function w on \mathcal{S} . The resulting estimates of the coefficients (appropriately normalized) are finally used for describing and analyzing the 3D shapes [see Ding et al. (2000), Kazhdan, Funkhouser and Rusinkiewicz (2003), Kelemen, Szekely and Gerig (1999) among many others].

In a recent paper Dette, Melas and Pepelysheff (2005) considered the optimal design problem for estimating the coefficients in the truncated Fourier expansion. They demonstrated that the commonly used designs [either a uniform distribution on \mathcal{S} realized by a grid or a uniform design taking observations on several circles with equal distances on the z -axis - see e.g. Ding et al. (2000)] are rather inefficient, if the emphasis of the design of the experiment is the minimization of the variances of the least squares estimates. In particular it is shown that the uniform distribution on the sphere (1.2) is optimal with respect to all Φ_p -criteria proposed by Kiefer (1974). The Φ_p -optimal designs minimize a p -norm of the eigenvalues of the variance-covariance matrix but do not take into account the bias, which is incurred by the truncation of (1.3). However in 3D-shape analysis the bias has a serious impact on the quality of the estimates [see e.g. Pawlak and Liao (2002)].

In the present paper we consider the optimal design problem for statistical analysis with spherical harmonic descriptors taking a variety of possible model specification errors into account. It is demonstrated that the uniform distribution on the sphere remains optimal with respect to a minimax mean square error criterion, which takes into account bias, heteroscedasticity and as well correlations in the data. As a consequence this design is robust with respect to the bias arising from the truncation of the Fourier series and with respect to the standard assumptions of homoscedastic and uncorrelated data.

In Section 2 we introduce the model and some of its basic properties. Section 3 deals with the construction of continuous optimal designs with respect to the minimax criterion based on the integrated mean square error calculated over the full set \mathcal{S} . Optimal design problems with respect to the integrated mean square error calculated over a finite subset of \mathcal{S} are considered in Section 4 and some numerical comparisons with commonly used designs are given in Section 5. The more technical details are given in the Appendix.

2. Some basic facts about spherical harmonic descriptors

We suppose that the random variables Y_i are observed, with additive error, i.e.

$$Y_i = Y(\boldsymbol{\psi}_i) = r(\boldsymbol{\psi}_i) + \eta(\boldsymbol{\psi}_i), \quad i = 1, \dots, n \quad (2.1)$$

at ‘locations’ $\boldsymbol{\psi}_i = (\theta_i, \phi_i) \in \mathcal{S}$, where the $\eta(\boldsymbol{\psi}_i)$ denote centered random variables with constant variance, say $\sigma_\eta^2 > 0$, and the the polar angle and azimuthal angle satisfy $\theta_i \in [0, \pi]$, $\phi_i \in (-\pi, \pi]$, respectively. It is assumed that the function $r(\boldsymbol{\psi})$ has an L^2 -expansion of the form (1.3), where the Y_n^m denote the spherical harmonic descriptors defined by

$$Y_n^m(\boldsymbol{\psi}) = \begin{cases} \sqrt{2n+1}P_n^0(\cos\theta), & m=0, n \geq 0, \\ \sqrt{2(2n+1)}\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta)\cos(m\phi), & m=1, \dots, n, n > 0, \\ \sqrt{2(2n+1)}\frac{(n+m)!}{(n-m)!}P_n^{-m}(\cos\theta)\sin(m\phi), & m=-n, \dots, -1, n > 0. \end{cases}$$

Here P_n^m is the m^{th} associated Legendre function of degree n satisfying the differential equation

$$(1-x^2)P''(x) - 2xP'(x) + \left\{n(n+1) - \frac{m^2}{1-x^2}\right\}P(x) = 0.$$

[See Andrews, Askey and Roy (1999) for more details]. Note that the functions P_n^m can be calculated recursively, because they are related to orthogonal polynomials.

For the estimation of the 3D shape the expansion in (1.3) is truncated at a given resolution d and the linear regression model

$$E[Y|\boldsymbol{\psi}] = \mathbf{z}^T(\boldsymbol{\psi})\mathbf{c}, \quad \text{Var}[Y|\boldsymbol{\psi}] = \sigma^2 > 0$$

is fitted to the data [see e.g. Brechbühler, Gerig and Kübler (1995)]; here the vector of regressors is given by

$$\mathbf{z}(\boldsymbol{\psi}) = (Y_0^0(\boldsymbol{\psi}), Y_1^{-1}(\boldsymbol{\psi}), Y_1^0(\boldsymbol{\psi}), Y_1^1(\boldsymbol{\psi}), \dots, Y_d^{-d}(\boldsymbol{\psi}), \dots, Y_d^d(\boldsymbol{\psi}))^T \in \mathbb{R}^{(d+1)^2}.$$

We note for future reference that the spherical harmonic descriptors form a complete orthonormal basis with respect to the uniform distribution on the sphere defined by (1.2), that is

$$\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi})\mathbf{z}^T(\boldsymbol{\psi})\mu(\boldsymbol{\psi})d\boldsymbol{\psi} = \mathbf{I}, \quad (2.2)$$

where the set \mathcal{S} is defined at (1.1) and \mathbf{I} denotes the identity matrix of order $(d+1)^2$. Moreover it was shown in Dette, Melas, and Pepelyshev (2005) that

$$\|\mathbf{z}(\boldsymbol{\psi})\| = d+1, \quad (2.3)$$

where $\|\cdot\|$ denotes the Euclidean norm.

An approximate design is a probability measure, say ξ , on the set \mathcal{S} . For a probability measure with finite support the support points, say $\boldsymbol{\psi}_i = (\theta_i, \phi_i)$, determine the points on the sphere, where the radius of the 3D shape is observed and the corresponding weights, say k_i , give the relative proportion of total observations, which are taken in a particular direction. Recently Dette, Melas and Pepelyshev (2005) determined optimal designs for statistical analysis with spherical harmonic descriptors, which minimize an appropriate function of the covariance matrix of the least squares estimate for the vector \mathbf{c} [see Silvey (1980) or Pukelsheim (1993)], where the observations in (2.1) are assumed to be homoscedastic and uncorrelated. In the present paper we investigate the problem of design of experiments, for 3D shape analysis with spherical harmonic descriptors, from several different perspectives.

1. First we note that, due to the truncation of the expansion (1.3), there appears a bias, say $f(\boldsymbol{\psi})$, i.e.

$$r(\boldsymbol{\psi}) = \mathbf{z}^T(\boldsymbol{\psi})\mathbf{c} + f(\boldsymbol{\psi}), \quad (2.4)$$

which somehow has to be addressed in the design of the experiment. In particular in 3D data the bias is often not negligible compared to the variance [see Pawlak and Liao (2002)]. Note that the function f refers to the remainder in the expansion (1.3) in an L^2 -sense, that is

$$f(\boldsymbol{\psi}) = \sum_{l=d+1}^{\infty} \sum_{m=-l}^l c_l^m Y_l^m(\boldsymbol{\psi}). \quad (2.5)$$

2. Secondly, the assumptions of homoscedastic and uncorrelated errors $\eta(\boldsymbol{\psi}_i)$ may be not realistic in some applications of shape analysis. For this reason we assume that the random error in (2.1) can be further decomposed as

$$\eta(\boldsymbol{\psi}) = \varepsilon(\boldsymbol{\psi}) + U(\boldsymbol{\psi}),$$

where the random variables $\varepsilon(\boldsymbol{\psi})$ are centered and uncorrelated with each other, but with possibly heterogeneous variances, say $\sigma_\varepsilon^2 g(\boldsymbol{\psi})$ for some function $g : \mathcal{S} \rightarrow \mathbb{R}^+$, and $U(\boldsymbol{\psi})$ is a random process, uncorrelated with $\varepsilon(\boldsymbol{\psi})$, with mean 0 and covariance function

$$\text{Cov}[U(\boldsymbol{\psi}), U(\boldsymbol{\psi}')] = h(\boldsymbol{\psi}, \boldsymbol{\psi}').$$

As a consequence of these assumptions the model (2.1) can be written as

$$Y_i = \mathbf{z}^T(\boldsymbol{\psi}_i) \mathbf{c} + f(\boldsymbol{\psi}_i) + \varepsilon(\boldsymbol{\psi}_i) + U(\boldsymbol{\psi}_i), \quad i = 1, \dots, n, \quad (2.6)$$

and in particular we have

$$E[Y(\boldsymbol{\psi})] = r(\boldsymbol{\psi}) = \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{c} + f(\boldsymbol{\psi}).$$

Moreover, it follows from (2.4) and (2.5) that

$$\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) f(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} = \mathbf{0}, \quad (2.7)$$

and that the vector \mathbf{c} is obtained as

$$\mathbf{c} = \arg \min_{\mathbf{t}} \int_{\mathcal{S}} (E[r(\boldsymbol{\psi})] - \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{t})^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi},$$

and given by the vector of the first $(d+1)^2$ Fourier coefficients of the function r . Let $\hat{\mathbf{c}}$ denote the estimate obtained from the least squares criterion (1.4) and define $\hat{Y}(\boldsymbol{\psi}) = \mathbf{z}^T(\boldsymbol{\psi}) \hat{\mathbf{c}}$ as the predicted response. In order to address possible bias, heteroscedasticity and correlation in the design of the experiment, we consider two integrated mean square error criteria:

$$\begin{aligned} IMSE_{1,f,g,h}(\xi) &= \int_{\mathcal{S}} E \left[\left\{ \hat{Y}(\boldsymbol{\psi}) - E[Y(\boldsymbol{\psi})] \right\}^2 \right] \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}, \\ IMSE_{2,f,g,h}(\xi) &= \int_{\mathcal{S}} E \left[\left\{ \hat{Y}(\boldsymbol{\psi}) - Y(\boldsymbol{\psi}) \right\}^2 \right] \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}, \end{aligned}$$

where ξ is the given experimental design. The criterion $IMSE_{2,f,g,h}$ measures the difference between the value of a ‘new’ observation r , observed at location $\boldsymbol{\psi}$, and its predicted value $\hat{r}(\boldsymbol{\psi})$, while $IMSE_{1,f,g,h}$ compares the prediction with the expectation of a new observation. Note also that both criteria depend on the bias f , variance function g and the correlation structure h of the errors η . For the case $f \equiv 0$, $h \equiv 0$ and $g \equiv 1$ Dette, Melas and Pepelyshev (2005) determined Φ_p -optimal designs in the sense of Kiefer (1974) minimizing a p -mean of the eigenvalues of the covariance matrix of the (unweighted) least squares estimate, but it is not clear if these designs are efficient or robust if any of these assumptions is violated. In order to obtain robust designs we propose a minimax approach, which seeks designs which minimize the worst $IMSE$ calculated over a certain class of functions. To be precise let η_f^2 and η_g^2 denote positive constants and consider the classes of functions

$$\mathcal{F} = \left\{ f \mid \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \leq \eta_f^2 \right\} \quad (2.8)$$

$$\mathcal{G} = \left\{ g \mid \sup_{\boldsymbol{\psi} \in \mathcal{S}} |g(\boldsymbol{\psi}) - g_0(\boldsymbol{\psi})| \leq \eta_g^2 \right\}. \quad (2.9)$$

For $j = 1, 2$ a design ξ_j^* is called minimax-optimal if it minimizes the maximum $IMSE_{j,f,g,h}$ over the classes \mathcal{F} , \mathcal{G} and \mathcal{H} , i.e.

$$\xi_j^* = \arg \min_{\xi} \max \left\{ IMSE_{j,f,g,h}(\xi) \mid f \in \mathcal{F}, g \in \mathcal{G}, h \in \mathcal{H} \right\}, \quad (2.10)$$

where the class \mathcal{H} will be specified in Section 3.1 and 3.2 corresponding to the cases of uncorrelated and correlated data, respectively. In the following two sections we will determine minimax-optimal designs with respect to the criterion (2.10). Section 4 is devoted to minimax optimal designs over a finite subset \mathcal{S}_0 of \mathcal{S} .

The need for robustness of design against a misspecified regression response was first elucidated in the seminal paper Box and Draper (1959). A minimax approach over a broad class of departures from the fitted response was subsequently formulated in Huber (1975). In a series of papers these notions were extended to multiple linear regression [Pesotchinsky (1982); Wiens (1992)] and to robustness against heteroscedasticity [Wiens (1998)]. A feature of this current article is that the continuous uniform design on the sphere is simultaneously optimal, and robust, in a variety of situations. The more ‘standard’ uniform distribution, i.e. with constant density, was previously shown to possess attractive robustness properties with respect to lack of fit testing in Wiens (1991); this work was subsequently extended by Biedermann and Dette (1992) and more recently by Bischoff and Miller (2006).

3. Minimax optimal designs on \mathcal{S}

In this section we suppose that any point $\boldsymbol{\psi} \in \mathcal{S} = [0, \pi] \times (-\pi, \pi]$ is a possible design point. We will demonstrate that for uncorrelated data the uniform distribution on the sphere remains optimal with respect to the minimax criterion (2.10), and that this conclusion remains valid if robustness against a broad class of correlation structures is required as well.

We will consider the uncorrelated and correlated cases separately. Before doing this we present the loss functions for fixed functions f , g and h , and then the maxima, over \mathcal{F} and \mathcal{G} , of these loss functions. It follows from Lemma 1 of Wiens (1992) that we may assume that the optimal design ξ_j^* is absolutely continuous with respect to Lebesgue measure, since otherwise the maximum loss (over \mathcal{F}) would be infinite. For a design ξ with density $k(\boldsymbol{\psi})$ define

$$m(\boldsymbol{\psi}) = k(\boldsymbol{\psi}) w(\boldsymbol{\psi}),$$

and assume that the average weight is one, so that

$$\int_{\mathcal{S}} w(\boldsymbol{\psi}) d\xi(\boldsymbol{\psi}) = \int_{\mathcal{S}} m(\boldsymbol{\psi}) d\boldsymbol{\psi} = 1$$

and $m(\cdot)$ is a probability density on \mathcal{S} . If the vector \mathbf{c} is estimated by weighted least squares (with weight function w), then a straightforward calculation shows that the resulting estimate can be expressed in terms of the design measure ξ (or equivalently m) as follows:

$$\hat{\mathbf{c}} = \mathbf{c} + \mathbf{B}_m^{-1} \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) Y(\boldsymbol{\psi}) d\boldsymbol{\psi},$$

where $Y(\boldsymbol{\psi})$ denotes the (continuous) data and the matrix \mathbf{B}_m is defined by

$$\mathbf{B}_m = \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi}.$$

The corresponding bias and covariance of this estimate are given by

$$\begin{aligned} E[\hat{\mathbf{c}}] - \mathbf{c} &= \mathbf{B}_m^{-1} \mathbf{b}_{f,m}, \\ \text{Cov}[\hat{\mathbf{c}}] &= \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1}, \end{aligned}$$

with

$$\begin{aligned}\mathbf{b}_{f,m} &= \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) f(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi}, \\ \mathbf{C}_{w,g,m} &= \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) w(\boldsymbol{\psi}) g(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi}, \\ \mathbf{D}_{h,m} &= \int_{\mathcal{S}} \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}'.\end{aligned}$$

Standard calculations now yield, for the integrated mean square errors,

$$\begin{aligned}IMSE_{1,f,g,h}(\xi) &= \mathbf{b}_{f,m}^T \mathbf{B}_m^{-2} \mathbf{b}_{f,m} + tr \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\} + \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}, \\ IMSE_{2,f,g,h}(\xi) &= \mathbf{b}_{f,m}^T \mathbf{B}_m^{-2} \mathbf{b}_{f,m} + tr \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\} + \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \\ &\quad - 2tr \mathbf{B}_m^{-1} \mathbf{E}_{h,m} + \sigma_\varepsilon^2 c_g + c_h,\end{aligned}$$

where

$$\begin{aligned}\mathbf{E}_{h,m} &= \int_{\mathcal{S}} \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}') h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}) m(\boldsymbol{\psi}') \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}' d\boldsymbol{\psi}, \\ c_g &= \int_{\mathcal{S}} g(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}, \\ c_h &= \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}.\end{aligned}$$

We now seek to maximize the integrated mean square error with respect to $f \in \mathcal{F}$ and $g \in \mathcal{G}$. This is straightforward for the function g and was carried out for the function f in general regression models with $\mu(\cdot) \equiv 1$ by Wiens (1992). The further generalization is straightforward and yields:

Proposition 1

(1) The maximum of $\mathbf{b}_{f,m}^T \mathbf{B}_m^{-2} \mathbf{b}_{f,m}$ with respect to $f \in \mathcal{F}$ satisfying (2.7) is given by

$$\eta_f^2 \left(ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] - 1 \right), \quad (3.1)$$

where $ch_{\max}[\mathbf{A}]$ denotes the maximum eigenvalue of the matrix \mathbf{A} and \mathbf{K}_m is defined by

$$\mathbf{K}_m = \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \frac{m^2(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} d\boldsymbol{\psi}.$$

The maximum value (3.1) is attained by any function f_* of the form

$$f_*(\boldsymbol{\psi}) = \eta_f \alpha_m^T \left[\frac{m(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} \mathbf{I} - \mathbf{B}_m \right] \mathbf{z}(\boldsymbol{\psi}),$$

where α_m is any solution of the equation $(\mathbf{K}_m - \mathbf{B}_m^2)^{1/2} \alpha_m = \beta_m$ and β_m is any eigenvector of the matrix $\mathbf{B}_m^{-1} (\mathbf{K}_m - \mathbf{B}_m^2) \mathbf{B}_m^{-1}$, corresponding to the maximum eigenvalue, normalized so that $\|\beta_m\| = 1$. Moreover,

$$\max_{f \in \mathcal{F}} IMSE_{j,f,g,h}(\xi) = \begin{cases} \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + tr \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\}, & \text{if } j = 1, \\ \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + tr \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\} \\ \quad - 2tr \mathbf{B}_m^{-1} \mathbf{E}_{h,m} + \sigma_\varepsilon^2 c_g + c_h, & \text{if } j = 2. \end{cases}$$

(2) The functions $\text{tr} \mathbf{B}_m^{-1} \mathbf{C}_{w,g,m} \mathbf{B}_m^{-1}$ and c_g are maximized, with respect to $g \in \mathcal{G}$, by $g_*(\boldsymbol{\psi}) = g_0(\boldsymbol{\psi}) + \eta_g^2$. Moreover,

$$\max_{f \in \mathcal{F}, g \in \mathcal{G}} IMSE_{j,f,g,h}(\xi) = \begin{cases} \text{tr} \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g_*,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\} + \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}], & \text{if } j = 1, \\ \text{tr} \left\{ \mathbf{B}_m^{-1} \left[\frac{\sigma_\varepsilon^2}{n} \mathbf{C}_{w,g_*,m} + \mathbf{D}_{h,m} \right] \mathbf{B}_m^{-1} \right\} + \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] \\ \quad - 2 \text{tr} \mathbf{B}_m^{-1} \mathbf{E}_{h,m} + \sigma_\varepsilon^2 c_{g_*} + c_h, & \text{if } j = 2. \end{cases}$$

3.1 Uncorrelated data

Throughout this subsection we assume that robustness with respect to the correlation structure of the errors does not have to be addressed in the optimality criterion (2.10), and so we take $h(\cdot, \cdot) \equiv 0$ in the expressions of Proposition 1. We begin our derivation with the case of classical least squares estimation based on uncorrelated and homoscedastic data, as considered in Dette, Melas, and Pepelyshev (2005) for Kiefer's (1974) Φ_p -optimality criteria. These authors however did not take the bias or possible heteroscedasticity into account (in other words they put $\eta_f^2 = 0$, $g_0(\boldsymbol{\psi}) = 1$, $\eta_g^2 = 0$ in our notation) and showed that the uniform distribution on the sphere defined by (1.2) is optimal with respect to all of Kiefer's Φ_p -criteria. The following result shows that this design is also optimal with respect to both $IMSE$ -criteria considered in the present paper, if the errors in model (2.1) are homoscedastic and uncorrelated.

Theorem 1 *If the errors in model (2.1) are homoscedastic (i.e. $\eta_g^2 = 0$; $g_0(\cdot) \equiv 1$) and uncorrelated (i.e. $h(\cdot, \cdot) \equiv 0$) and the weight function in the least squares criterion (1.4) is constant, then the uniform design on the sphere defined by (1.2) minimizes $\max_{f \in \mathcal{F}} IMSE_{j,f,1,0}(\xi)$ for $j = 1, 2$.*

If heteroscedasticity is a concern the experimenter will consider the use of weighted least squares. From (2) of Proposition 1 it follows that the maximum loss can be minimized with respect to the weight function w by minimizing

$$\text{tr} \mathbf{B}_m^{-1} \mathbf{C}_{w,g_*,m} \mathbf{B}_m^{-1} = \int_{\mathcal{S}} w(\boldsymbol{\psi}) g_*(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi} \quad (3.2)$$

over all non-negative functions $w(\cdot)$ subject to the requirement that the function $k(\boldsymbol{\psi}) = m(\boldsymbol{\psi})/w(\boldsymbol{\psi})$ be a density, that is

$$\int_{\mathcal{S}} \frac{m(\boldsymbol{\psi})}{w(\boldsymbol{\psi})} d\boldsymbol{\psi} = 1. \quad (3.3)$$

(Here we note that the weight function w occurs only in the expression $\mathbf{C}_{w,g_*,m}$ of $IMSE_j$, $j = 1, 2$). In (3.3) the integrand and $w(\boldsymbol{\psi})$ are defined to be zero off of the support of $m(\cdot)$. The following Proposition gives the solution to this minimization problem.

Proposition 2 *The quantity (3.2) is minimized over non-negative weights, subject to the normalizing condition (3.3), by*

$$w_*(\boldsymbol{\psi}) = \frac{\gamma_m}{\|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| \sqrt{g_*(\boldsymbol{\psi})}}$$

on the support of $m(\cdot)$, where $g_*(\boldsymbol{\psi}) = g_0(\boldsymbol{\psi}) + \eta_g^2$ and $\gamma_m = \int_{\mathcal{S}} \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| \sqrt{g_*(\boldsymbol{\psi})} m(\boldsymbol{\psi}) d\boldsymbol{\psi}$. Moreover, the minimum value of (3.2) is given by

$$\text{tr} \mathbf{B}_m^{-1} \mathbf{C}_{w_*,g_*,m} \mathbf{B}_m^{-1} = \gamma_m^2.$$

Note that the optimal weights depend on the given design m . As a consequence of Proposition 2 we obtain minimax-optimal designs for weighted least squares estimation with “optimal” weights, which are robust against bias and heteroscedasticity. We assume here that the experimenter fits a homoscedastic model, i.e. takes $g_0(\boldsymbol{\psi}) \equiv 1$.

Theorem 2 *If the errors in model (2.1) are uncorrelated (i.e. $h(\cdot, \cdot) \equiv 0$) and the function g_0 in (2.9) satisfies $g_0(\boldsymbol{\psi}) \equiv 1$, then the uniform design defined by (1.2) minimizes*

$$\min_w \max_{f \in \mathcal{F}, g \in \mathcal{G}} IMSE_{j,f,g,0}(\xi) \quad (3.4)$$

for $j = 1, 2$. In other words $k_*(\boldsymbol{\psi}) = \mu(\cdot)$ is the minimax robust design density, and constant regression weights are minimax in the presence of heteroscedastic (but uncorrelated) errors.

Remark 1 *Note that the optimality of the uniform design $\mu(\cdot)$, or the uniform design/weights combination has been established using only two properties of the spherical harmonic basis functions:*

- (i) the norm $\|\mathbf{z}(\boldsymbol{\psi})\|$ is constant on \mathcal{S} ;
- (ii) the regressors $\mathbf{z}(\boldsymbol{\psi})$ are orthonormal with respect to some measure $\mu(\cdot)$.

In fact the crucial requirement is that the function

$$\mathbf{z}^T(\boldsymbol{\psi}) \left[\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \right]^{-1} \mathbf{z}(\boldsymbol{\psi})$$

be constant on \mathcal{S} , since the transformed regressors

$$\tilde{\mathbf{z}}(\boldsymbol{\psi}) = \left[\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \right]^{-1/2} \mathbf{z}(\boldsymbol{\psi})$$

then satisfy conditions (i) and (ii) stated above. These properties, and hence the associated optimality of the design $m(\cdot) = \mu(\cdot)$, hold in several other regression models. Examples are Haar wavelet regression - see Herzberg and Traves (1994) and Oyet and Wiens (2000), and trigonometric regression - see Karlin and Studden (1966).

3.2 Correlated data

In the previous subsection we have shown that for uncorrelated data the uniform distribution on the sphere defined by (1.2) is robust in a minimax sense with respect to the bias in the expansion (2.4) and to heteroscedasticity in the data. We have not yet discussed robustness issues with respect to assumptions regarding the correlation structure of the data. Here we will demonstrate that the uniform distribution on the sphere is also optimally robust against a broad class of correlation structures. We consider, for a given constant $\eta_h^2 > 0$, the neighbourhood

$$\mathcal{H} = \left\{ h \mid 0 \leq \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') f(\boldsymbol{\psi}) f(\boldsymbol{\psi}') \mu(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \leq \eta_h^2 \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \forall f \in L_{\mu}^2(\mathcal{S}) \right\}. \quad (3.5)$$

Note that we then have, for any vector of functions $\mathbf{a}(\boldsymbol{\psi})$ with $\|\mathbf{a}\| \in L_{\mu}^2(\mathcal{S})$,

$$0 \leq \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{a}^T(\boldsymbol{\psi}) \mathbf{a}(\boldsymbol{\psi}') \mu(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \leq \eta_h^2 \int_{\mathcal{S}} \|\mathbf{a}(\boldsymbol{\psi})\|^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}. \quad (3.6)$$

From (2) of Proposition 1 we see that the terms in $\max_{f \in \mathcal{F}, g \in \mathcal{G}} IMSE_{j,f,g,h}(\xi)$ which contain h do not involve either w or g_* . The design $m = \mu$, already seen to be minimax robust in \mathcal{F} and \mathcal{G} , thus continues to be so in \mathcal{H} if it minimizes

$$\max_{h \in \mathcal{H}} \text{tr} \mathbf{B}_m^{-1} \mathbf{D}_{h,m} \mathbf{B}_m^{-1} = \max_h \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \quad (3.7)$$

in the case of $IMSE_{1,f,g,h}$, and

$$\begin{aligned} \max_{h \in \mathcal{H}} \{ \text{tr} \mathbf{B}_m^{-1} \mathbf{D}_{h,m} \mathbf{B}_m^{-1} - 2 \text{tr} \mathbf{B}_m^{-1} \mathbf{E}_{h,m} + c_h \} &= \max_{h \in \mathcal{H}} \left\{ \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \right. \\ &\quad \left. - 2 \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' + \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \right\} \end{aligned} \quad (3.8)$$

in the case of $IMSE_{2,f,g,h}$.

Theorem 3 *If the sets \mathcal{F} , \mathcal{G} and \mathcal{H} are given by (2.8), (2.9) (with $g_0(\boldsymbol{\psi}) \equiv 1$) and (3.5), respectively, then the uniform design on the unit sphere defined by (1.2) minimizes*

$$\min_w \max_{f \in \mathcal{F}, g \in \mathcal{G}, h \in \mathcal{H}} IMSE_{j,f,g,h}(\xi)$$

for $j = 1, 2$. In other words $k_*(\boldsymbol{\psi}) = \mu(\cdot)$ is the minimax robust design density, and constant regression weights are minimax, in the presence of heteroscedastic and correlated errors as well as bias.

Remark 2 *If the class \mathcal{H} is instead defined by*

$$\tilde{\mathcal{H}} = \left\{ h \mid 0 \leq \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') f(\boldsymbol{\psi}) f(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \leq \eta_h^2 \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) d\boldsymbol{\psi}, \forall f \in L^2(\mathcal{S}) \right\} \quad (3.9)$$

we obtain by a similar argument as given in the proof of Theorem 3 that

$$\begin{aligned} \min_w \max_{f \in \mathcal{F}, g \in \mathcal{G}, h \in \tilde{\mathcal{H}}} IMSE_{1,f,g,h} &= \eta_f^2 c h_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + \frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2) \int_{\mathcal{S}} \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}) d\boldsymbol{\psi} \\ &\quad + \eta_h^2 \int_{\mathcal{S}} (\|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}))^2 d\boldsymbol{\psi}. \end{aligned} \quad (3.10)$$

Note that the first and second term in this expression are minimized by the uniform distribution on the sphere. However, this design does not minimize the second integral at (3.10). This can be seen even in the simplest case $d = 0$, where $z(\boldsymbol{\psi}) = 1$ and $\mathbf{B}_m = \int_{\mathcal{S}} m(\boldsymbol{\psi}) d\boldsymbol{\psi} = 1$. In this case the second integral is minimized by minimizing

$$\int_{\mathcal{S}} (\|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}))^2 d\boldsymbol{\psi} = \int_{\mathcal{S}} m^2(\boldsymbol{\psi}) d\boldsymbol{\psi}$$

among probability densities $m(\boldsymbol{\psi})$. The minimizer is easily seen to be the conventional uniform density on \mathcal{S} , i.e. $m(\boldsymbol{\psi}) \equiv (\int_{\mathcal{S}} d\boldsymbol{\psi})^{-1} = 1/(2\pi^2)$. As a consequence the uniform distribution on the sphere is in general not minimax optimal, if the class (3.9) is used to address for possible correlations in the data.

4. Minimax optimal designs on discrete subsets

The experimenter faces obvious difficulties in implementing a continuous design such as $\mu(\boldsymbol{\psi})$, for which there are no atoms at which to place the design points. In this section we address this problem in the following way. We restrict attention to a subset $\mathcal{S}_0 = \{\boldsymbol{\psi}_i\}_{i=1}^N \subset [0, \pi] \times (-\pi, \pi]$ of the design space \mathcal{S} , which is finite but sufficiently large as to contain all points at which one might contemplate making observations. We continue to analyze designs for the model (2.6), and assume that the experimenter takes $n_i = n(\boldsymbol{\psi}_i) \geq 0$ observations $Y_{ij} = Y_j(\boldsymbol{\psi}_i)$ ($j = 1, \dots, n_i$) at the location $\boldsymbol{\psi}_i$, $i = 1, \dots, N$. Then $n = \sum_{i=1}^N n_i$ denotes the total sample size. The weighted least squares estimate is given by

$$\hat{\mathbf{c}} = \left[\sum_{i=1}^N n_i \mathbf{z}(\boldsymbol{\psi}_i) w(\boldsymbol{\psi}_i) \mathbf{z}^T(\boldsymbol{\psi}_i) \right]^{-1} \sum_{i=1}^N \mathbf{z}(\boldsymbol{\psi}_i) w(\boldsymbol{\psi}_i) \sum_{j=1}^{n_i} Y_{ij}(\boldsymbol{\psi}_i),$$

where $w(\boldsymbol{\psi})$ is again a non-negative weight function. We endow \mathcal{S}_0 with a probability measure, say μ , where $\mu\{\boldsymbol{\psi}_i\} = \mu_i > 0$. Note that we again use the notation of approximate designs in the sense of Kiefer (1974) and call any probability measure with finite support \mathcal{S}_0 a design. If a design has masses k_1, \dots, k_N at the points $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N$ the experimenter takes approximately $n(\boldsymbol{\psi}_i) = nk_i$ observations at $\boldsymbol{\psi}_i$ ($i = 1, \dots, N$). As in Section 3 we consider, for a given design, the normalized probabilities $m_i = k_i w(\boldsymbol{\psi}_i)$. We assume that the average weight is one, i.e.

$$\sum_{i=1}^N w_i k_i = \sum_{i=1}^N m_i = 1,$$

so that $\{m_i\}_{i=1}^N$ is a probability distribution on \mathcal{S}_0 . The analogues of (2.7), (2.8) and (2.9) are

$$\sum_{i=1}^N \mu_i \mathbf{z}(\boldsymbol{\psi}_i) f(\boldsymbol{\psi}_i) = \mathbf{0}, \quad (4.1)$$

$$\mathcal{F}_0 = \left\{ f \mid \sum_{i=1}^N \mu_i f^2(\boldsymbol{\psi}_i) \leq \eta_f^2 \right\}, \quad (4.2)$$

$$\mathcal{G}_0 = \left\{ g \mid \sup_{\mathcal{S}_0} |g(\boldsymbol{\psi}) - g_0(\boldsymbol{\psi})| \leq \eta_g^2 \right\},$$

respectively. In analogy with the approach taken in Section 3.2 we consider the following class of covariance structures:

$$\mathcal{H}_0 = \left\{ h \mid 0 \leq \sum_{i=1}^N \sum_{j=1}^N h(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j) f(\boldsymbol{\psi}_i) f(\boldsymbol{\psi}_j) \mu_i \mu_j \leq \eta_h^2 \sum_{i=1}^N f^2(\boldsymbol{\psi}_i) \mu_i \right\},$$

for all functions $f(\cdot)$ bounded on \mathcal{S}_0 .

As in the previous sections we discuss the integrated mean square error criteria, but now evaluate the mean square error only at those points in \mathcal{S}_0 , that is

$$\begin{aligned} IMSE_{1,f,g,h}(\xi) &= \sum_{i=1}^N \mu_i E \left[\left\{ \hat{Y}(\boldsymbol{\psi}_i) - E[Y(\boldsymbol{\psi}_i)] \right\}^2 \right], \\ IMSE_{2f,g,h}(\xi) &= \sum_{i=1}^N \mu_i E \left[\left\{ \hat{Y}(\boldsymbol{\psi}_i) - Y(\boldsymbol{\psi}_i) \right\}^2 \right]. \end{aligned}$$

Our aim is to obtain optimal (minimax) weights $\{w_i\}$ and design probabilities $\{k_i\}$, where the maxima will be taken over the classes \mathcal{F}_0 , \mathcal{G}_0 and \mathcal{H}_0 . We use the following definitions:

$$\begin{aligned} \mathbf{f} &= (f(\boldsymbol{\psi}_1), \dots, f(\boldsymbol{\psi}_N))^T, & \mathbf{g} &= (g(\boldsymbol{\psi}_1), \dots, g(\boldsymbol{\psi}_N))^T, \\ \mathbf{P} &= \text{diag}(\mu_1, \dots, \mu_N), & \boldsymbol{\mu} &= (\mu_1, \dots, \mu_N)^T = \mathbf{P}\mathbf{1}_N, \\ \mathbf{Z} &= (\mathbf{z}(\boldsymbol{\psi}_1), \dots, \mathbf{z}(\boldsymbol{\psi}_N))^T, & \mathbf{A} &= \mathbf{Z}^T \mathbf{P} \mathbf{Z}, \\ \mathbf{M} &= \text{diag}(m_1, \dots, m_N), & \mathbf{B}_m &= \mathbf{Z}^T \mathbf{M} \mathbf{Z}, \\ \mathbf{Q} &= \mathbf{Z} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^T, & \mathbf{W} &= \text{diag}(w_1, \dots, w_N), \\ \mathbf{H} &= (h(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j))_{i,j=1, \dots, n}, & \mathbf{r} &= [\text{diag}(\mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{W})] \mathbf{1}_N. \end{aligned}$$

A straightforward calculation yields, for the integrated mean square errors,

$$\begin{aligned} IMSE_{1,f,g,h}(\xi) &= [\mathbf{f}^T \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{f} + \mathbf{f}^T \mathbf{P} \mathbf{f}] + \frac{\sigma_\varepsilon^2}{n} \mathbf{r}^T \mathbf{g} + \text{tr} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{H} \mathbf{M}, \\ IMSE_{2,f,g,h}(\xi) &= [\mathbf{f}^T \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{f} + \mathbf{f}^T \mathbf{P} \mathbf{f}] + \sigma_\varepsilon^2 \left(\boldsymbol{\mu} + \frac{1}{n} \mathbf{r} \right)^T \mathbf{g} + \text{tr} [(\mathbf{Q} \mathbf{M} - \mathbf{I}) \mathbf{H} (\mathbf{M} \mathbf{Q} - \mathbf{I}) \mathbf{P}]. \end{aligned}$$

As in Section 3 we begin with the calculation of the maximum loss over the different classes. The maximum over the class \mathcal{F}_0 is given by the following result.

Theorem 4 *Let $\tilde{\mathbf{Z}}$ be an $N \times (N - (d + 1)^2)$ matrix whose columns are orthonormal and form a basis for the orthogonal complement to the column space of the matrix \mathbf{Z} . Then the maximum of the function*

$$L(\mathbf{f}) = \mathbf{f}^T \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{f} + \mathbf{f}^T \mathbf{P} \mathbf{f}$$

over the set the set \mathcal{F}_0 defined in (4.2) subject to condition (4.1) is

$$\max_{\mathbf{f} \in \mathcal{F}_0} L(\mathbf{f}) = \eta_f^2 (1 + \lambda_m),$$

where

$$\lambda_m = ch_{\max} \left[\left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \right]. \quad (4.3)$$

If \mathbf{e} denotes an eigenvector of unit norm corresponding to this maximum eigenvalue, then the maximum of the function L is attained at $\mathbf{f}_{\max} = \eta_f \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \mathbf{e}$.

Remark 3 *Since $\mathbf{Q} = \mathbf{Z} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^T$, the $(N - (d + 1)^2) \times (N - (d + 1)^2)$ matrix in (4.3) factors as*

$$\left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Z} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{A} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} = \mathbf{C}^T \mathbf{C},$$

where the $(d + 1)^2 \times (N - (d + 1)^2)$ matrix \mathbf{C} is defined by

$$\mathbf{C} = \mathbf{A}^{1/2} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2}.$$

Thus there are at most only $(d + 1)^2$ nonzero eigenvalues. Note that $ch_{\max} [\mathbf{C}^T \mathbf{C}] = ch_{\max} [\mathbf{C} \mathbf{C}^T]$, and consequently, if $N - (d + 1)^2 > (d + 1)^2$, then it is more convenient to compute instead

$$\lambda_m = ch_{\max} [\mathbf{C} \mathbf{C}^T] = ch_{\max} \left[\mathbf{A}^{1/2} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1} \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Z} (\mathbf{Z}^T \mathbf{M} \mathbf{Z})^{-1} \mathbf{A}^{1/2} \right].$$

The matrix $\tilde{\mathbf{Z}}$ is computed by starting with the spectral decomposition of the projector $\mathbf{I}_N - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$, whose columns are orthogonal to those of \mathbf{Z} . The eigenvectors of this projector, associated with the $N - (d + 1)^2$ unit eigenvalues, are the columns of $\tilde{\mathbf{Z}}$.

The maximizations over the classes \mathcal{G}_0 and \mathcal{H}_0 are more straightforward.

Proposition 3 *Let $g_*(\boldsymbol{\psi})$ be as in Proposition 1 and let the vector \mathbf{g}_* have elements $g_*(\boldsymbol{\psi}_i)$. Then the maximum loss over $f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0$ is*

$$\begin{aligned} \max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h} &= \\ \eta_f^2 (1 + \lambda_m) &+ \begin{cases} \frac{\sigma_\varepsilon^2}{n} \mathbf{r}^T \mathbf{g}_* + \eta_h^2 \text{tr} [\mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{P}^{-1} \mathbf{M}], & \text{if } j = 1, \\ \sigma_\varepsilon^2 \left(\boldsymbol{\mu} + \frac{1}{n} \mathbf{r} \right)^T \mathbf{g}_* + \eta_h^2 \text{tr} [(\mathbf{Q} \mathbf{M} - \mathbf{I}) \mathbf{P}^{-1} (\mathbf{M} \mathbf{Q} - \mathbf{I}) \mathbf{P}], & \text{if } j = 2. \end{cases} \end{aligned}$$

If heteroscedasticity in the data cannot be excluded, weighted least squares estimation is a reasonable procedure and we now exhibit the optimal weights with respect to the minimax criterion. By virtue of Proposition 3, these are obtained by minimizing $\mathbf{r}^T \mathbf{g}_*$ subject to the constraint

$$\sum_{i=1}^N k_i = \sum_{i=1}^N \frac{m_i}{w_i} = 1. \quad (4.4)$$

For notational convenience we write $\mathbf{r}^T \mathbf{g}_* = \mathbf{s}^T \mathbf{w}$ for $\mathbf{w} = (w_1, \dots, w_N)^T$, $\mathbf{s} = (s_1, \dots, s_N)^T$ and $s_i = \|\mathbf{A}^{1/2} \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i)\|^2 m_i g_{*,i}$. The proof of the following result is analogous to that of Proposition 2 and is therefore omitted.

Proposition 4 *The quantity $\mathbf{s}^T \mathbf{w}$ is minimized over non-negative weights \mathbf{w} , subject to the normalizing condition (4.4) by*

$$w_{*,i} = \frac{\gamma_m}{\|\mathbf{A}^{1/2} \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i)\| \sqrt{g_{*,i}}}, \quad (4.5)$$

where

$$\gamma_m = \sum_{i=1}^N m_i \|\mathbf{A}^{1/2} \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i)\| \sqrt{g_{*,i}}.$$

The minimum value of $\mathbf{s}^T \mathbf{w}$ is given by $\mathbf{s}^T \mathbf{w}_* = \gamma_m^2$. Thus

$$\min_w \max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h} = \eta_f^2 (1 + \lambda_m) + \begin{cases} \frac{\sigma_\varepsilon^2}{n_b} \gamma_m^2 + \eta_h^2 tr [\mathbf{QPQMP}^{-1} \mathbf{M}], & \text{if } j = 1, \\ \frac{\sigma_\varepsilon^2}{n} \gamma_m^2 + \sigma_\varepsilon^2 \boldsymbol{\mu}^T \mathbf{g}_* + \eta_h^2 tr [(\mathbf{QM} - \mathbf{I}) \mathbf{P}^{-1} (\mathbf{MQ} - \mathbf{I}) \mathbf{P}], & \text{if } j = 2. \end{cases} \quad (4.6)$$

Recall that the matrix

$$\mathbf{A} = \sum_{i=1}^N \mu_i \mathbf{z}(\boldsymbol{\psi}_i) \mathbf{z}^T(\boldsymbol{\psi}_i) = \mathbf{Z}^T \mathbf{P} \mathbf{Z}$$

is the discrete analogue of the matrix $\int \mathbf{z}(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} = \mathbf{I}_{(d+1)^2}$. It was shown by Dette, Melas, and Pepelyshev (2005) using properties of quadrature formulas, that for sufficiently large N one can find points $\{\boldsymbol{\psi}_i\}_{i=1}^N$ and probabilities $\{\mu_i\}_{i=1}^N$ such that

$$\mathbf{A} = \mathbf{I}_{(d+1)^2}. \quad (4.7)$$

For the rest of this section we assume that $\mathcal{S}_0 = \{\boldsymbol{\psi}_i\}_{i=1}^N$ has been constructed in this manner and we denote the corresponding design by μ . As in Theorems 1 and 2 we shall also take $g_0(\boldsymbol{\psi}) \equiv 1$. Then (4.6) becomes

$$\min_w \max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h} = \eta_f^2 (1 + \lambda_m) + \frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2) \left(\sum_{i=1}^N m_i \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i)\| \right)^2 + \begin{cases} \eta_h^2 tr [\mathbf{QPQMP}^{-1} \mathbf{M}], & \text{if } j = 1, \\ \sigma_\varepsilon^2 (1 + \eta_g^2) + \eta_h^2 tr [(\mathbf{QM} - \mathbf{I}) \mathbf{P}^{-1} (\mathbf{MQ} - \mathbf{I}) \mathbf{P}], & \text{if } j = 2. \end{cases}$$

The following result shows that the design μ is the minimax robust design against any combination of bias, heteroscedasticity and dependence.

Theorem 5 *If condition (4.7) is satisfied and $g_0(\boldsymbol{\psi}) \equiv 1$ then the design $k_i^* = \mu_i$ ($i = 1, \dots, N$) together with constant weights (implying that $m = \mu$) minimizes*

$$\max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h}$$

for $j = 1, 2$ and any combination of $\eta_f^2 > 0$, $\eta_g^2 > 0$, $\eta_h^2 > 0$.

5. Numerical comparisons

In this section we construct discrete designs as described in §4, and compute, for these and some competing designs,

$$\max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h}$$

for a range of values of parameter values. Using $g_0(\boldsymbol{\psi}) \equiv 1$ and unit regression weights in all cases, this is (from Proposition 3) given by

$$\begin{aligned} & \max_{f \in \mathcal{F}_0, g \in \mathcal{G}_0, h \in \mathcal{H}_0} IMSE_{j,f,g,h} \\ & = \eta_f^2 (1 + \lambda_m) + \begin{cases} \frac{\sigma_\varepsilon^2}{n} \mathbf{r}^T \mathbf{g}_* + \eta_h^2 \text{tr} [\mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{P}^{-1} \mathbf{M}], & \text{if } j = 1, \\ \sigma_\varepsilon^2 (\boldsymbol{\mu} + \frac{1}{n} \mathbf{r})^T \mathbf{g}_* + \eta_h^2 \text{tr} [(\mathbf{Q} \mathbf{M} - \mathbf{I}) \mathbf{P}^{-1} (\mathbf{M} \mathbf{Q} - \mathbf{I}) \mathbf{P}], & \text{if } j = 2. \end{cases} \end{aligned} \quad (5.1)$$

By defining

$$\alpha = \frac{\eta_f^2}{\eta_f^2 + \frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2) + \eta_h^2}, \quad \beta = \frac{\frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2)}{\eta_f^2 + \frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2) + \eta_h^2}, \quad \gamma = 1 - \alpha - \beta,$$

we can write (5.1) as $(\eta_f^2 + \frac{\sigma_\varepsilon^2}{n} (1 + \eta_g^2) + \eta_h^2) \times \mathcal{L}(m; \alpha, \beta)$, for

$$\begin{aligned} \mathcal{L}(m; \alpha, \beta) & = \alpha (1 + \lambda_m) \\ & + \begin{cases} \beta \text{tr} [\mathbf{B}_m^{-1} \mathbf{A}] + \gamma \text{tr} [(\mathbf{B}_m^{-1} \mathbf{Z}^T \mathbf{M})^T \mathbf{A} (\mathbf{B}_m^{-1} \mathbf{Z}^T \mathbf{M}) \mathbf{P}^{-1}], & \text{if } j = 1, \\ \beta \{ \text{tr} [\mathbf{B}_m^{-1} \mathbf{A}] + n \} + \gamma \{ \text{tr} [(\mathbf{B}_m^{-1} \mathbf{Z}^T \mathbf{M})^T \mathbf{A} (\mathbf{B}_m^{-1} \mathbf{Z}^T \mathbf{M}) \mathbf{P}^{-1}] + N - 2(d+1)^2 \}, & \text{if } j = 2. \end{cases} \end{aligned}$$

Here α, β , and γ may be interpreted as representing the relative importance of errors due to bias, variance, and dependence in the mind of the experimenter. For our designs with $m = \boldsymbol{\mu}$ we attain

$$\min_m \mathcal{L}(m; \alpha, \beta) = \alpha + \begin{cases} (1 - \alpha)(d+1)^2, & \text{if } j = 1, \\ (1 - \alpha)(d+1)^2 + n\beta + \gamma \{ N - 2(d+1)^2 \}, & \text{if } j = 2. \end{cases} \quad (5.2)$$

Minimax designs satisfying our criteria may be constructed, for certain values of N , by the methods of Dette, Melas and Pepelysheff (2005). For the sake of completeness we briefly describe the method used here. Let $P_{d+1}(x)$ ($= P_{d+1}^0(x)$, in the notation of §2) be the Legendre polynomial of degree $d+1$. Let $x_1 < \dots < x_{d+1}$ be the zeros, all of which lie in $(-1, 1)$, of this polynomial. For $j = 1, \dots, d+1$ compute probabilities

$$v_j = \frac{1}{2} \int_{-1}^1 \prod_{k=1, k \neq j}^{d+1} \frac{x - x_k}{x_j - x_k} dx.$$

For $N = n_0 \cdot (d+1)$ and $n_0 \geq 2d+1$, let $\boldsymbol{\mu}$ be the product measure $\boldsymbol{\mu} = \boldsymbol{\mu}_1 \otimes \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_1$ places mass v_j on each of the points $\theta_j = \arccos(x_j)$, and $\boldsymbol{\mu}_2$ places mass n_0^{-1} on each of n_0 equally spaced points $\phi_k = \delta + (2\pi k/n_0) \in (-\pi, \pi]$ ($k = 1, \dots, n_0$). Here δ is any value in $(-(n_0+1)\pi/n_0, -\pi]$; in our examples we use $\delta = -\pi$. Then (4.7) holds and the design $\boldsymbol{\mu}$ on the set $\{\boldsymbol{\psi}_{j,k} = (\theta_j, \phi_k)\}$ is minimax. Note that $N > (d+1)^2$; this is of course desirable from a robustness standpoint.

We compare the minimax design (denoted by ‘ M ’) constructed as above with three more conventional uniform designs. The first (denoted by ‘ U_1 ’) places mass N^{-1} on each of the same support points as M . The second (denoted by ‘ U_2 ’) is given by $\boldsymbol{\mu}' = \boldsymbol{\mu}'_1 \otimes \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}'_1$ places mass $(d+1)^{-1}$ on each of $d+1$ equally spaced points $\theta_j = \pi j/(d+1)$ ($j = 1, \dots, d+1$). The third (denoted by ‘ U_3 ’) is given by $\boldsymbol{\mu}'' = \boldsymbol{\mu}''_1 \otimes \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}''_1$ places mass $(d+1)^{-1}$ on each of $\theta_j = \arccos(1 - (2j/(d+1)))$ ($j = 1, \dots, d+1$). The design U_3 was used by Ding et al. (2000) for a principal component analysis of data from a spherical harmonic regression analysis, and takes observations on several circles, equally spaced on the z -axis. Note that all of M, U_2 and U_3 satisfy $\mathbf{B}_m = \mathbf{A}$ and $\mathbf{M} = \mathbf{P}$; these properties ensure that (5.2) holds for all three designs; i.e. all are minimax with respect to their underlying measures $\boldsymbol{\mu}, \boldsymbol{\mu}', \boldsymbol{\mu}''$. However, of

Table 1. Values of $\mathcal{L}(m; \alpha, \beta)$ for the minimax design M and conventional uniform design U_1 from Example 5.1 ($d = 2$).

α	$\mathcal{L}(M; \alpha, \beta)$	$\mathcal{L}(U_1; \alpha, \beta)$					
		$\beta = 0$	$\beta = .2$	$\beta = .4$	$\beta = .6$	$\beta = .8$	$\beta = 1.0$
0	9.00	9.19	9.22	9.24	9.27	9.29	9.31
.2	7.40	7.57	7.59	7.61	7.64	7.66	*
.4	5.80	5.94	5.96	5.99	6.01	*	*
.6	4.20	4.31	4.33	4.36	*	*	*
.8	2.60	2.68	2.71	*	*	*	*
1.0	1.00	1.06	*	*	*	*	*

Table 2. Values of $\mathcal{L}(m; \alpha, \beta)$ for the minimax design M and conventional uniform design U_1 from Example 5.2 ($d = 6$).

α	$\mathcal{L}(M; \alpha, \beta)$	$\mathcal{L}(U_1; \alpha, \beta)$					
		$\beta = 0$	$\beta = .2$	$\beta = .4$	$\beta = .6$	$\beta = .8$	$\beta = 1.0$
0	49.00	50.42	51.38	52.34	53.30	54.26	55.22
.2	39.40	40.59	41.55	42.51	43.47	44.43	*
.4	29.80	30.75	31.71	32.68	33.64	*	*
.6	20.20	20.92	21.88	22.84	*	*	*
.8	10.60	11.09	12.05	*	*	*	*
1.0	1.00	1.26	*	*	*	*	*

the three, only the design M also satisfies (4.7). For the numerical comparison 3 levels of resolution - $d = 2, 6, 13$ - are considered.

Example 5.1. $d = 2$. In this case the support points of μ_1 , and corresponding probabilities, are

$$\begin{aligned} \theta: & \arccos\left(\sqrt{3/5}\right), \quad \pi/2, \quad \arccos\left(-\sqrt{3/5}\right); \\ v: & \quad 5/18, \quad 4/9, \quad 5/18. \end{aligned}$$

Example 5.2. $d = 6$. Here the support points of μ_1 , and corresponding probabilities, are

$$\begin{aligned} \theta: & 0.320, \quad 0.735, \quad 1.153, \quad 1.571, \quad 1.989, \quad 2.406, \quad 2.821; \\ v: & .065, \quad .140, \quad .191, \quad .209, \quad .191, \quad .140, \quad .065. \end{aligned}$$

Example 5.3. $d = 13$. The support points of μ_1 , and corresponding probabilities, are

$$\begin{aligned} \theta: & 0.166, \quad 0.381, \quad 0.597, \quad 0.813, \quad 1.030, \quad 1.246, \quad 1.463, \\ & 1.679, \quad 1.896, \quad 2.112, \quad 2.329, \quad 2.545, \quad 2.761, \quad 2.976; \\ v: & .018, \quad .040, \quad .061, \quad .079, \quad .093, \quad .103, \quad .108, \\ & .108, \quad .103, \quad .093, \quad .079, \quad .061, \quad .040, \quad .018. \end{aligned}$$

Table 3. Values of $\mathcal{L}(m; \alpha, \beta)$ for the minimax design M and conventional uniform design U_1 from Example 5.3 ($d = 13$).

α	$\mathcal{L}(M; \alpha, \beta)$	$\mathcal{L}(U_1; \alpha, \beta)$					
		$\beta = 0$	$\beta = .2$	$\beta = .4$	$\beta = .6$	$\beta = .8$	$\beta = 1.0$
0	196.00	200.22	206.30	212.38	218.46	224.54	230.62
.2	157.00	160.50	166.58	172.66	178.74	184.82	*
.4	118.00	120.78	126.86	132.94	139.02	*	*
.6	79.00	81.06	87.14	93.22	*	*	*
.8	40.00	41.34	47.42	*	*	*	*
1.0	1.00	1.62	*	*	*	*	*

Table 4. Values of the efficiencies of M relative to U_j when $\eta_f = \eta_g = \eta_h = 0$.

U_k	d	p			
		0	1	2	∞
U_1	2	1.017	1.035	1.107	1.333
U_2		1.216	1.400	2.493	3.238
U_3		1.319	1.858	7.104	7.074
U_1	6	1.072	1.127	1.364	1.463
U_2		1.167	1.269	1.888	4.113
U_3		1.440	16.33	8999.2	661.5
U_1	13	1.108	1.177	1.495	1.507
U_2		1.154	1.239	1.716	4.813
U_3		1.596	31,630	1.80×10^{11}	5.94×10^6

Tables 1 - 3 contain the values of $\mathcal{L}(m; \alpha, \beta)$ for M (or U_2 or U_3), U_1 and a range of values of α and β . We present only the results for $j = 1$ since the difference $\mathcal{L}(U_1; \alpha, \beta) - \mathcal{L}(M; \alpha, \beta)$ is the same for both values of j . We observe that with respect to the minimax criterion the design U_1 is rather efficient if the dimension is small. For example, the loss of efficiency in the case $d = 2$ when using U_1 instead of M is usually less than 5%. As d increases, the difference between M and U_1 is more substantial. For example in the cases $d = 6$ and $d = 13$ the losses in efficiency are about 12% and 17%, respectively.

In view of the fact that each of M , U_2 and U_3 is a minimax design (with respect to their underlying measures), one might choose between them on the basis of their efficiency in the ideal model, with $\eta_f = \eta_g = \eta_h = 0$. The natural measure is then Kiefer's Φ_p criterion, defined in terms of the eigenvalues $\{\lambda_i(\mathbf{C})\}_{i=1}^{(d+1)^2}$ of the covariance matrix \mathbf{C} of a design ξ by $\Phi_p(\xi) = ((d+1)^{-2} \sum_{i=1}^{(d+1)^2} \lambda_i^p)^{1/p}$, ($1 \leq p < \infty$) with $\Phi_0(\xi) = (\det(\mathbf{C}))^{1/(d+1)^2}$ and $\Phi_\infty(\xi) = \max(\lambda_i(\mathbf{C}))$. In Table 4 we give the relative efficiencies

$$\text{eff}(k) = \frac{\Phi_p(U_k)}{\Phi_p(M)} = \Phi_p(U_k)$$

for the uniform designs U_1 , U_2 and U_3 and a range of values of d and p . Clearly, by this criterion the design M performs substantially better. However, the design U_1 , which takes 1 observation at each of the same design points as M , yields reasonable efficiencies and is substantially easier to implement.

6. Summary and conclusions

We have exhibited designs, appropriate for the analysis of spherical harmonic shape descriptors. These designs are minimax robust against a variety of model misspecifications - errors in the specification of the response function, in the assumption of homoscedasticity, and in the assumption of independence of the errors across observations. The examples of §5 have demonstrated that these designs enjoy a considerable advantage over more conventional uniform designs, especially for large values of d . There is as well a computational benefit to the minimax design satisfying (4.7) - since the corresponding information matrix is a multiple of the identity, the regression coefficients can be computed without matrix inversions: we have $\hat{\mathbf{c}} = \sum_{i=1}^N \mathbf{z}(\psi_i) m_i \bar{Y}_i$, where \bar{Y}_i is the average of the observations at ψ_i . Especially for the large values of d often used in practice, this confers a considerable advantage to these designs [see Brechbühler, Gerig and Kübler (1995)].

Appendix: Derivations

Proof of Theorem 1. If $w \equiv 1$, $\eta_g^2 = 0$; $g_0 = 1$ $h(\cdot, \cdot) \equiv 0$ we have $k(\cdot) = m(\cdot)$ and as a consequence $\mathbf{C}_{w,g,m} = \mathbf{B}_m$. Therefore we obtain for the maximal integrated mean square errors

$$\max_{f \in \mathcal{F}} IMSE_{j,f,1,0}(\xi) = \begin{cases} \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + \frac{\sigma_\varepsilon^2}{n} tr [\mathbf{B}_m^{-1}], & \text{if } j = 1, \\ \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + \frac{\sigma_\varepsilon^2}{n} tr [\mathbf{B}_m^{-1}] + \sigma_\varepsilon^2, & \text{if } j = 2. \end{cases}$$

We now shows that both terms in this expression are minimized by the uniform distribution (1.2) on the unit sphere. For this first note that $tr [\mathbf{B}_m^{-1}]$ corresponds to Kiefer's A -optimality criterion, which was considered in Dette, Melas and Pepelyshev (2005) and is minimal for the uniform distribution (1.2) on the unit sphere. Secondly, note that $ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] - 1 = ch_{\max} [\mathbf{B}_m^{-1} (\mathbf{K}_m - \mathbf{B}_m^2) \mathbf{B}_m^{-1}]$. Moreover, for any vector \mathbf{a} we have

$$\mathbf{a}^T (\mathbf{K}_m - \mathbf{B}_m^2) \mathbf{a} = \int_{\mathcal{S}} \left\{ \mathbf{a}^T \left[\frac{m(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} \mathbf{I} - \mathbf{B}_m \right] \mathbf{z}(\boldsymbol{\psi}) \right\}^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \geq 0,$$

and consequently the matrix $\mathbf{K}_m - \mathbf{B}_m^2$ is non-negative definite, which implies $ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] \geq 1$. But this minimum value of 1 is attained by $m(\cdot) = \mu(\cdot)$, for which $\mathbf{K}_m = \mathbf{B}_m = \mathbf{I}$. ■

Proof of Proposition 2: For weight functions $w_0(\boldsymbol{\psi})$ and $w_1(\boldsymbol{\psi})$ and $t \in [0, 1]$ define $w_t(\boldsymbol{\psi}) = (1-t)w_0(\boldsymbol{\psi}) + tw_1(\boldsymbol{\psi})$. In order that the function $w_0(\boldsymbol{\psi})$ minimize (3.2) subject to the normalizing conditions (3.3) it is sufficient that the function

$$\phi(t; \lambda) = \int_{\mathcal{S}} w_t(\boldsymbol{\psi}) g_*(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi} + \lambda \left[\int_{\mathcal{S}} \frac{m(\boldsymbol{\psi})}{w_t(\boldsymbol{\psi})} d\boldsymbol{\psi} - 1 \right] \quad (\lambda \geq 0)$$

be minimal at $t = 0$ for any $w_1(\cdot)$, and that $w_0(\cdot)$ satisfies (3.3). For this, since $\phi(t; \lambda)$ is a convex function of t , the first order condition is necessary and sufficient, i.e.

$$\phi'(0; \lambda) = \int_{\mathcal{S}} [w_1(\boldsymbol{\psi}) - w_0(\boldsymbol{\psi})] \left[g_*(\boldsymbol{\psi}) \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) - \lambda \frac{m(\boldsymbol{\psi})}{w_0^2(\boldsymbol{\psi})} \right] d\boldsymbol{\psi} \geq 0$$

for all $w_1(\cdot)$. This condition is satisfied if

$$w_0(\boldsymbol{\psi}) = \frac{\lambda}{\|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| \sqrt{g_*(\boldsymbol{\psi})}}$$

(on the support of $m(\cdot)$ - we can define $w_0(\boldsymbol{\psi})$ arbitrarily elsewhere), and it remains only to determine the constant λ to satisfy (3.3). ■

Proof of Theorem 2. From Proposition 2 we obtain

$$\min_w \max_{f,g} IMSE_{j,f,g,0}(\xi) = \begin{cases} \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + \frac{\sigma_\varepsilon^2}{n} \gamma_m^2, & \text{if } j = 1, \\ \eta_f^2 ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}] + \frac{\sigma_\varepsilon^2}{n} \gamma_m^2 + \sigma_\varepsilon^2, & \text{if } j = 2. \end{cases}$$

Because $g_0(\boldsymbol{\psi}) \equiv 1$, it follows that

$$\gamma_m = \sqrt{1 + \eta_g^2} \int_{\mathcal{S}} \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}) d\boldsymbol{\psi}. \quad (\text{A.1})$$

It was shown in the proof of Theorem 1 that the maximum eigenvalue $ch_{\max} [\mathbf{B}_m^{-1} \mathbf{K}_m \mathbf{B}_m^{-1}]$ is minimized by the uniform distribution on the sphere $\mu(\cdot)$, for which the corresponding minimax weights are, by Proposition 2, proportional to $\|\mathbf{z}(\boldsymbol{\psi})\|^{-1}$, hence by (2.3) are constant. If this choice of design can be shown to minimize (A.1) as well, then the assertion of the Proposition follows, i.e. $\mu(\cdot)$ minimizes (3.4). Showing this requires proving the inequality

$$\int_{\mathcal{S}} \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}) d\boldsymbol{\psi} \geq \int_{\mathcal{S}} \|\mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi})\| \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} = d + 1, \quad (\text{A.2})$$

where we have used (2.3) and (2.2) for the last equality. However, the inequality in (A.2) is a direct consequence of the Cauchy-Schwarz inequality:

$$\|\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})\| \geq \frac{\mathbf{z}^T(\boldsymbol{\psi})\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})}{\|\mathbf{z}(\boldsymbol{\psi})\|} = \frac{\mathbf{z}^T(\boldsymbol{\psi})\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})}{d+1};$$

this gives

$$\begin{aligned} \int_{\mathcal{S}} \|\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}) d\boldsymbol{\psi} &\geq \frac{1}{d+1} \int_{\mathcal{S}} \mathbf{z}^T(\boldsymbol{\psi})\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi} \\ &= \frac{1}{d+1} \text{tr}\mathbf{B}_m^{-1} \int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi})\mathbf{z}^T(\boldsymbol{\psi}) m(\boldsymbol{\psi}) d\boldsymbol{\psi} \\ &= \frac{1}{d+1} \text{tr}\mathbf{I}_{(d+1)^2} = d+1. \end{aligned}$$

■

Proof of Theorem 3: First take $j = 1$. From (3.7) and (3.6) we are to show that

$$\begin{aligned} &\max_h \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2}\mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\ &= \eta_h^2 \int_{\mathcal{S}} \mathbf{z}^T(\boldsymbol{\psi}) \mathbf{B}_m^{-2}\mathbf{z}(\boldsymbol{\psi}) \frac{m^2(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} d\boldsymbol{\psi} \\ &= \eta_h^2 \int_{\mathcal{S}} \left\| \mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}) \frac{m(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} \right\|^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \end{aligned}$$

is minimized by $m(\cdot) = \mu(\cdot)$. By the Cauchy-Schwarz inequality and (A.2),

$$\begin{aligned} \int_{\mathcal{S}} \left\| \mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}) \frac{m(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} \right\|^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} &\geq \left\{ \int_{\mathcal{S}} \left\| \mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}) \frac{m(\boldsymbol{\psi})}{\mu(\boldsymbol{\psi})} \right\| \mu(\boldsymbol{\psi}) d\boldsymbol{\psi} \right\}^2 \\ &= \left\{ \int_{\mathcal{S}} \|\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})\| m(\boldsymbol{\psi}) d\boldsymbol{\psi} \right\}^2 \\ &\geq (d+1)^2. \end{aligned} \tag{A.3}$$

But this lower bound $(d+1)^2$ is attained by $m(\cdot) = \mu(\cdot)$; this establishes Theorem 3 in the case $j = 1$.

For a proof of the result in the case $j = 2$ we recall (3.8) and consider the function

$$\begin{aligned} \Phi(h; m) &= \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2}\mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\ &\quad - 2 \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' + \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}. \end{aligned}$$

We have to show that

$$\max_h \Phi(h; m) \geq \max_h \Phi(h; \mu). \tag{A.4}$$

To establish (A.4), it is clearly sufficient to show that for *any* function $h \in \mathcal{H}$,

$$0 \leq \Phi(h; m) - \Phi(h; \mu) \text{ for any } m(\cdot). \tag{A.5}$$

For this, note that

$$\begin{aligned}
\Phi(h; m) - \Phi(h; \mu) &= \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\
&\quad - 2 \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) m(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\
&\quad - \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\
&\quad + 2 \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}' \\
&= \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') [\mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-2} \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi}) m(\boldsymbol{\psi}') \\
&\quad - 2\mathbf{z}^T(\boldsymbol{\psi}') \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) m(\boldsymbol{\psi}') + \mathbf{z}^T(\boldsymbol{\psi}') \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}')] d\boldsymbol{\psi} d\boldsymbol{\psi}' \\
&= \int_{\mathcal{S}} \int_{\mathcal{S}} h(\boldsymbol{\psi}, \boldsymbol{\psi}') \mathbf{a}^T(\boldsymbol{\psi}) \mathbf{a}(\boldsymbol{\psi}') d\boldsymbol{\psi} d\boldsymbol{\psi}',
\end{aligned}$$

with $\mathbf{a}(\boldsymbol{\psi}) = \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) - \mathbf{z}(\boldsymbol{\psi}) m(\boldsymbol{\psi})$. Now (A.5) follows from the non-negative definiteness of the kernel $h(\cdot, \cdot)$, i.e. from the first inequality in (3.6). \blacksquare

Proof of Theorem 4: The constraint (4.1) on \mathbf{f} is given by the equation $\mathbf{Z}^T \mathbf{P} \mathbf{f} = \mathbf{0}$. Equivalently, $\mathbf{P} \mathbf{f}$ lies in the orthogonal complement to the column space of \mathbf{Z} , so that $\mathbf{f} = \mathbf{P}^{-1} \tilde{\mathbf{Z}} \mathbf{c}$ for some vector \mathbf{c} . We have to maximize the expression

$$\mathbf{f}^T \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{f} + \mathbf{f}^T \mathbf{P} \mathbf{f} = \mathbf{c}^T \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \mathbf{c} + \mathbf{c}^T \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \mathbf{c}$$

subject to condition (4.2), which is $\mathbf{f}^T \mathbf{P} \mathbf{f} = \mathbf{c}^T \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \mathbf{c} \leq \eta_f^2$. Equivalently, with $\mathbf{e} = \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{1/2} \mathbf{c} / \eta_f$, we maximize

$$\mathbf{e}^T \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{P}^{-1} \tilde{\mathbf{Z}} \left(\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \tilde{\mathbf{Z}} \right)^{-1/2} \mathbf{e} + \mathbf{e}^T \mathbf{e}$$

subject to $\mathbf{e}^T \mathbf{e} \leq 1$. This is a standard problem whose solution is as described in the Theorem. \blacksquare

Proof of Proposition 3: In light of Theorem 4 we have only to show that

- (i) $\mathbf{r}^T \mathbf{g} = \sum r_i g(\boldsymbol{\psi}_i)$ and $(\boldsymbol{\mu} + \frac{1}{n} \mathbf{r})^T = \sum (\mu_i + \frac{r_i}{n}) g(\boldsymbol{\psi}_i)$ are both maximized over $g \in \mathcal{G}_0$ by $g = g_*$;
- (ii) $\text{tr}[\mathbf{Q} \mathbf{P} \mathbf{Q} \mathbf{M} \mathbf{H} \mathbf{M}]$ and $\text{tr}[(\mathbf{Q} \mathbf{M} - \mathbf{I}) \mathbf{H} (\mathbf{M} \mathbf{Q} - \mathbf{I}) \mathbf{P}]$ are both maximized over $h \in \mathcal{H}_0$ by $\mathbf{H} = \eta_h^2 \mathbf{P}^{-1}$.

The first of these is immediate from the definition of \mathcal{G}_0 and the fact that

$$r_i = \mathbf{z}^T(\boldsymbol{\psi}_i) \mathbf{B}_m^{-1} \mathbf{A} \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i) m_i w_i = \left\| \mathbf{A}^{1/2} \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i) \right\|^2 m_i w_i \geq 0.$$

The second follows from the fact that both traces are maximized by choosing \mathbf{H} to be maximal with respect to the Loewner ordering. But from the definition of \mathcal{H}_0 it follows that $\mathbf{H} \leq \eta_h^2 \mathbf{P}^{-1}$ in this ordering. \blacksquare

Proof of Theorem 5. If $\eta_f^2 > 0$ then we are to minimize λ_m . But λ_m is minimized by $m = \mu$, with minimum value $\lambda_\mu = 0$. This is because for $m = \mu$ we have $\mathbf{M} = \mathbf{P}$, and consequently the matrix (4.3) contains a factor

$$\tilde{\mathbf{Z}}^T \mathbf{P}^{-1} \mathbf{M} \mathbf{Q} = \tilde{\mathbf{Z}}^T \mathbf{Q} = \tilde{\mathbf{Z}}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{P} \mathbf{Z})^{-1} \mathbf{Z}^T = \mathbf{0}.$$

If $\eta_g^2 > 0$ then we are to show that $\sum_{i=1}^N m_i \left\| \mathbf{B}_m^{-1} \mathbf{z}(\boldsymbol{\psi}_i) \right\|^2$ is also minimized by $m = \mu$. But this is merely the discrete analogue of (A.2), and is proven in an identical manner. That the minimax weights (4.5) are constant follows from (4.7), $\mathbf{B}_\mu = \mathbf{I}$ and the constancy of g_0 .

It remains to show that the design $m = \mu$ is also optimal when $\eta_h^2 > 0$, i.e. that

$$\text{tr} [\mathbf{QPQMP}^{-1}\mathbf{M}] = \sum_{i=1}^N \left\| \mathbf{B}_m^{-1} \mathbf{z}(\psi_i) \frac{m_i}{\mu_i} \right\|^2 \mu_i$$

and

$$\text{tr} [(\mathbf{QM} - \mathbf{I})\mathbf{P}^{-1}(\mathbf{MQ} - \mathbf{I})\mathbf{P}] = \text{tr} [\mathbf{QPQMP}^{-1}\mathbf{M}] + N - 2(d+1)^2.$$

are both minimized by $m = \mu$. The first of these is proven as at (A.3), and implies the second. ■

Acknowledgements

The support of the Deutsche Forschungsgemeinschaft (SFB 475, ‘‘Komplexitätsreduktion in multivariaten Datenstrukturen’’) is gratefully acknowledged. The work of H. Dette was also supported in part by a NIH grant award IR01GM072876:01A1. The work of D. Wiens was supported by the Natural Sciences and Research Council of Canada, and was partially carried out while enjoying the hospitality of the Fakultät für Mathematik, Ruhr-Universität, Bochum. The authors are grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

References

- Andrews, G.E., Askey, R., and Roy, R. (1999), *Special functions*. Encyclopedia of Mathematics and Its Applications, Cambridge.
- Biedermann, S., and Dette, H. (2001), ‘‘Optimal Designs for Testing the Functional Form of a Regression via Nonparametric Estimation Techniques,’’ *Statistics & Probability Letters*, 52, 215-224.
- Bischoff, W. and Miller, F. (2006), ‘‘Optimal Designs Which are Efficient for Lack of Fit Tests,’’ *Annals of Statistics*, 34, 2015-2025.
- Box, G.E.P., and Draper, N.R. (1959), ‘‘A Basis for the Selection of a Response Surface Design,’’ *Journal of the American Statistical Association*, 54, 622-654.
- Brechbühler, Ch., Gerig, G., and Kübler, O. (1995), ‘‘Surface Parametrization and Shape Description,’’ *Computer Vision and Image Understanding*, 61, 154-170.
- Dette, H., Melas, V., and Pepelyshev, A. (2005), ‘‘Optimal Designs for Three-Dimensional Shape Analysis with Spherical Harmonic Descriptors,’’ *Annals of Statistics*, 33, 2758-2788.
- Ding, W., Nesumi, H., Takano, Y., and Ukai, Y. (2000), ‘‘Quantitative Evaluation of the Three-Dimensional Fruit Shape and Size of *Citrus* Based on Spherical Harmonic Descriptors,’’ *Euphytica*, 114, 103-116.
- Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., and Dobkin, D. (2003), ‘‘A Search Engine for 3D models,’’ *ACM Trans. Graph*, 22, 83-105.
- Herzberg, A. M., and Traves, W. N. (1994), ‘‘An Optimal Experimental Design for the Haar Regression Model,’’ *Canadian Journal of Statistics*, 22, 357-364.
- Huber, P.J. (1975), ‘‘Robustness and Designs,’’ in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.
- Karlin, S., and Studden, W.J. (1966), *Tchebycheff Systems: With Applications in Analysis and Statistics*, New York: Interscience Publishers.

- Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2003), "Rotation Invariant Spherical Harmonic Representation of 3D-Shape Descriptors," *Eurographics Symposium on Geometry processing* (eds. L. Kobbelt, P. Schröder, H. Happe). <http://www.cs.princeton.edu/mkazhdan/My-Papers/SGP03.pdf>
- Kelemen, A., Székely, G., and Gerig, G. (1999), "Elastic Model-Based Segmentation of 3D-Neuroradiological Data Sets," *IEEE-TMI*, 18, 828-839.
- Kiefer, J. (1974), "General Equivalence Theory for Optimum Designs (Approximate Theory)," *Ann. Statist.*, 2, 849-879.
- Novotni, M., and Klein, R. (2003), "3D Zernike Descriptors for Content Based Shape Retrieval," *Proceedings of the eighth ACM symposium on solid modeling and applications*, 216-225.
- Oyet, A.J., and Wiens, D.P. (1997), "Robust Designs for Wavelet Approximations of Regression Models," *Journal of Nonparametric Statistics*, 12, 837-859.
- Pawlak, M., and Liao, S.X. (2002), "On the Recovery of a Function on a Circular Domain," *IEEE Transactions on Information Theory*, 48, 2736-2753.
- Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k ," *The Annals of Statistics*, 10, 511-525.
- Pukelsheim, F. (1993), *Optimal Design of Experiments*, Toronto: John Wiley.
- Székely, G., Kelemen, A., Brechbühler, Ch., and Gerig, G. (1996), "Segmentation of 2D and 3D Objects from MRI Volume Data Using Constrained Elastic Deformations of Flexible Contour and Surface Models," *Medical Image Analysis*, 1, 19-34.
- Silvey, S.D. (1980). *Optimal Design*, London: Chapman and Hall.
- Wiens, D.P. (1991), "Designs for Approximately Linear Regression: Two Optimality Properties of Uniform Designs," *Statistics and Probability Letters*; 12, 217-221.
- Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.
- Wiens, D.P. (1998), "Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroscedastic Errors," *Journal of the American Statistical Association*, 93, 1440-1450.