Statistica Sinica **19** (2009), 83-102

ROBUST DESIGNS FOR 3D SHAPE ANALYSIS WITH SPHERICAL HARMONIC DESCRIPTORS

Holger Dette and Douglas P. Wiens

Ruhr-Universität Bochum and University of Alberta

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: 3D-image data, dependent data, mean square error, minimax optimal designs, optimal designs, robust designs, shape analysis, spherical harmonic descriptors.

1. Introduction and Summary

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 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 others. In many cases 3D data appear in the form $Y_i = r(\boldsymbol{\psi}_i) + error_i$ $(i = 1, \ldots, n)$, where

$$\boldsymbol{\psi}_i = (\theta_i, \phi_i) \in \mathcal{S} = \{(\theta, \phi) \mid \theta \in [0, \pi],\ \phi \in (-\pi, \pi]\}\,, \tag{1.1}$$

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 $\{Y_l^m(\boldsymbol{\psi}) \mid m \in \{-l, -l+1, \ldots, l\}; l \in \mathbb{N}_0\}$ denote a complete orthonormal basis with respect to the uniform distribution on the unit sphere, with density

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

If $r : S \to \mathbb{R}$ denotes a (square integrable) function representing the radius in direction $\boldsymbol{\psi}$, the coefficients $c_l^m = \int_0^\pi \int_{-\pi}^\pi r(\boldsymbol{\psi}) Y_l^m(\boldsymbol{\psi}) \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}/4\pi$ of a spherical harmonic expansion

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

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

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

for an appropriate non-negative weight function w on $\mathcal{S}$. The resulting estimates of the coefficients are then used for describing and analyzing the 3D shapes. See Ding et al. (2000), Kazhdan et al. (2003), Kelemen, Szekely and Gerig (1999), among others.

In a recent paper Dette, Melas and Pepelyshev (2005) considered the optimal design problem for estimating the coefficients c_l^m. 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 (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). Moreover, 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, motivated by a desire for robustness, we consider this design problem while taking a variety of possible model specification errors into

account. 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) and 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 (2001), and more recently by Bischoff and Miller (2006).

To define our models precisely, we suppose that the random variables Y_i are observed with additive error, i.e.,

$$Y_i = Y(\boldsymbol{\psi}_i) = r(\boldsymbol{\psi}_i) + \eta(\boldsymbol{\psi}_i)\ , \quad i = 1, \ldots, n, \tag{1.5}$$

at 'locations' $\boldsymbol{\psi}_i = (\theta_i, \phi_i) \in \mathcal{S}$, where the $\eta(\boldsymbol{\psi}_i)$ denote centred random variables with constant variance, say $\sigma_\eta^2 > 0$, and the polar angle and azimuthal angle satisfy $\theta_i \in [0, \pi]$ and $\phi_i \in (-\pi, \pi]$, 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

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

Here P_n^m is the m^{th} associated Legendre function of degree n; see Andrews, Askey and Roy (1999) for more details. The expansion in (1.3) is truncated at a given resolution d, and the linear regression model

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

is fitted to the data; here the vector of regressors is given by

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

We frequently use the fact 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

$$\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi})\,\mathbf{z}^T(\boldsymbol{\psi})\,\mu(\boldsymbol{\psi})\,d\boldsymbol{\psi} = \mathbf{I}, \tag{1.6}$$

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

$$\|\mathbf{z}(\boldsymbol{\psi})\| = d+1, \tag{1.7}$$

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

We investigate several possible forms of model misspecification. First we note that truncation of the expansion (1.3) leads to a bias $f(\boldsymbol{\psi})$, so

$$r(\boldsymbol{\psi}) = \mathbf{z}^T(\boldsymbol{\psi})\,\mathbf{c} + f(\boldsymbol{\psi}), \tag{1.8}$$

and this must be addressed in the design. 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, viz.

$$f(\boldsymbol{\psi}) = \sum_{l=d+1}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m(\boldsymbol{\psi}). \tag{1.9}$$

We also note that the assumptions of homoscedastic and uncorrelated errors $\eta(\boldsymbol{\psi}_i)$ may be unrealistic in some applications of shape analysis. For this reason we assume that the random error in (1.5) can be further decomposed as

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

where the $\varepsilon(\boldsymbol{\psi})$ are centred and uncorrelated with each other but with possibly heterogeneous variances, say $\sigma_\varepsilon^2 g(\boldsymbol{\psi})$ for some function $g : \mathcal{S} \to \mathbb{R}^+$, and where $U(\boldsymbol{\psi})$ is a random process, uncorrelated with $\varepsilon(\boldsymbol{\psi})$, with mean 0 and covariance function $\mathrm{Cov}\left[U(\boldsymbol{\psi}), U(\boldsymbol{\psi}')\right] = h(\boldsymbol{\psi}, \boldsymbol{\psi}')$.

As a consequence of these assumptions, the model (1.5) can be written as

$$Y_i = \mathbf{z}^T(\boldsymbol{\psi}_i)\,\mathbf{c} + f(\boldsymbol{\psi}_i) + \varepsilon(\boldsymbol{\psi}_i) + U(\boldsymbol{\psi}_i), \quad i = 1, \ldots, n, \tag{1.10}$$

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 (1.8) and (1.9) that

$$\int_{\mathcal{S}} \mathbf{z}(\boldsymbol{\psi})\, f(\boldsymbol{\psi})\, \mu(\boldsymbol{\psi})\, d\boldsymbol{\psi} = \mathbf{0}, \tag{1.11}$$

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

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

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:

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

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. The designs constructed by Dette, Melas and Pepelyshev (2005) are Φ_p-optimal for the case $f \equiv 0$, $h \equiv 0$ and $g \equiv 1$; it is not clear if they 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 that minimize the worst $IMSE$ calculated over a certain class of functions. Thus, let η_f^2 and η_g^2 denote positive constants and consider the classes of functions

$$\mathcal{F} = \left\{f \;\middle|\; \int_{\mathcal{S}} f^2(\boldsymbol{\psi})\,\mu(\boldsymbol{\psi})\,d\boldsymbol{\psi} \le \eta_f^2\right\}, \tag{1.12}$$
$$\mathcal{G} = \left\{g \;\middle|\; \sup_{\boldsymbol{\psi}\in\mathcal{S}} |g(\boldsymbol{\psi}) - g_0(\boldsymbol{\psi})| \le \eta_g^2\right\}. \tag{1.13}$$

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.,

$$\xi_j^* \;=\; \mathrm{argmin}_\xi \max\left\{IMSE_{j,f,g,h}(\xi) \;\middle|\; f \in \mathcal{F}, g \in \mathcal{G}, h \in \mathcal{H}\right\}, \tag{1.14}$$

where the class $\mathcal{H}$ will be specified in Sections 2.1 and 2.2 corresponding to the cases of uncorrelated and correlated data, respectively.

In Section 2 we treat the construction of continuous optimal designs with respect to the minimax criterion based on *IMSE* calculated over the full set $\mathcal{S}$. Optimal design problems with respect to the integrated mean square error calculated over a finite subset $\mathcal{S}_0 = \{\boldsymbol{\psi}_i\}_{i=1}^N$ of $\mathcal{S}$ are considered in Section 3. It is demonstrated there that the uniform distribution on the sphere remains optimal with respect to a minimax mean square error criterion that takes into account bias, heteroscedasticity, and 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 violations of the standard assumptions of homoscedastic and uncorrelated data.

Some numerical comparisons of the designs constructed in Section 3 with some common competitors are given in Section 4. The examples of that section demonstrate that our 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 designs - the corresponding information matrix is a multiple of the identity, and thus the regression coefficients can be computed without matrix inversions: we have $\hat{\mathbf{c}} = \sum_{i=1}^{N} \mathbf{z}(\boldsymbol{\psi}_i) m_i \bar{Y}_{i\cdot}$, where $\bar{Y}_{i\cdot}$ is the average of the observations at $\boldsymbol{\psi}_i$, and m_i is the product of the design and regression weights at $\boldsymbol{\psi}_i$. 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)).

The derivations are given in the on-line supplement at http://www.stat.sinica.edu.tw/statistica/.

2. 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 (1.14), 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 (equivalently, in terms of m) as

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

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

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

with

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

Standard calculations yield, for the integrated mean square errors,

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

where

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

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.

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* (1.11) *is given by*

$$\eta_f^2\left(ch_{\max}\left[\mathbf{B}_m^{-1}\mathbf{K}_m\mathbf{B}_m^{-1}\right] - 1\right), \tag{2.1}$$

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 (2.1) *is attained by any function* f_* *of the form*

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

where $\boldsymbol{\alpha}_m$ *is any solution of the equation* $\left(\mathbf{K}_m - \mathbf{B}_m^2\right)^{1/2}\boldsymbol{\alpha}_m = \boldsymbol{\beta}_m$, *and* $\boldsymbol{\beta}_m$ *is any eigenvector of the matrix* $\mathbf{B}_m^{-1}\left(\mathbf{K}_m - \mathbf{B}_m^2\right)\mathbf{B}_m^{-1}$ *corresponding to the maximum eigenvalue, normalized so that* $\|\boldsymbol{\beta}_m\| = 1$. *Moreover,*

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

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

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

2.1. 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 (1.14), and so we take $h(\cdot,\cdot)\equiv 0$ in the expressions of Proposition 1. We begin our development with the case of classical least squares estimation based on uncorrelated and homoscedastic data, as considered in Dette, Melas and Pepelyshev (2005). These authors however did not take the bias or possible heteroscedasticity into account (in other words they put $\eta_f^2 = 0$, $g_0(\boldsymbol{\psi}) = 1$, and $\eta_g^2 = 0$, in our notation) and showed that the uniform distribution on the sphere defined by (1.2) is optimal with respect to all 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 (1.5) are homoscedastic and uncorrelated.

Theorem 1. *If the errors in model* (1.5) *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

$$tr\mathbf{B}_m^{-1}\mathbf{C}_{w,g_*,m}\mathbf{B}_m^{-1} = \int_{\mathcal{S}} w(\boldsymbol{\psi})\, g_*(\boldsymbol{\psi})\, \mathbf{z}^T(\boldsymbol{\psi})\, \mathbf{B}_m^{-2}\mathbf{z}(\boldsymbol{\psi})\, m(\boldsymbol{\psi})\, d\boldsymbol{\psi} \tag{2.2}$$

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

$$\int_{\mathcal{S}} \frac{m(\boldsymbol{\psi})}{w(\boldsymbol{\psi})} d\boldsymbol{\psi} = 1. \tag{2.3}$$

(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 (2.3) the integrand and $w(\boldsymbol{\psi})$ are defined to be zero off the support of $m(\cdot)$.

Proposition 2. *The quantity* (2.2) *is minimized over non-negative weights, subject to the normalizing condition* (2.3), *by*

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

on the support of $m(\cdot)$, *where* $\gamma_m = \int_{\mathcal{S}} \left\|\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi})\right\| \sqrt{g_*(\boldsymbol{\psi})} m(\boldsymbol{\psi})\, d\boldsymbol{\psi}$ *and* $g_*(\boldsymbol{\psi}) = g_0(\boldsymbol{\psi}) + \eta$. *Moreover, the minimum value of* (2.2) *is given by*

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

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 that 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* (1.5) *are uncorrelated* (*i.e.,* $h(\cdot,\cdot) \equiv 0$) *and the function* g_0 *in* (1.13) *satisfies* $g_0(\boldsymbol{\psi}) \equiv 1$, *then the uniform design defined by* (1.2) *minimizes*

$$\min_{w} \max_{f\in\mathcal{F}, g\in\mathcal{G}} IMSE_{j,f,g,0}(\xi) \tag{2.4}$$

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 S;
(ii) the regressors $z(\boldsymbol{\psi})$ 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 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). 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).

2.2. 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 (1.8), 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

$$\begin{aligned}\mathcal{H} = \Big\{h \;\Big|\; 0 \le &\int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right) f(\boldsymbol{\psi}) f\left(\boldsymbol{\psi}'\right) \mu(\boldsymbol{\psi})\mu(\boldsymbol{\psi}')d\boldsymbol{\psi}d\boldsymbol{\psi}' \\ &\le \eta_h^2 \int_{\mathcal{S}} f^2(\boldsymbol{\psi}) \mu(\boldsymbol{\psi})d\boldsymbol{\psi} \;\forall f \in L_\mu^2(\mathcal{S})\Big\}. \end{aligned} \tag{2.5}$$

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

$$0 \le \int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right) \mathbf{a}^T(\boldsymbol{\psi}) \mathbf{a}\left(\boldsymbol{\psi}'\right) \mu(\boldsymbol{\psi}) \mu\left(\boldsymbol{\psi}'\right) d\boldsymbol{\psi}d\boldsymbol{\psi}' \le \eta_h^2 \int_{\mathcal{S}} \|\mathbf{a}(\boldsymbol{\psi})\|^2 \mu(\boldsymbol{\psi}) d\boldsymbol{\psi}. \tag{2.6}$$

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

$$\max_{h\in\mathcal{H}} tr\mathbf{B}_m^{-1}\mathbf{D}_{h,m}\mathbf{B}_m^{-1}=\max_h \int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right)\mathbf{z}^T\left(\boldsymbol{\psi}'\right)\mathbf{B}_m^{-2}\mathbf{z}\left(\boldsymbol{\psi}\right)m\left(\boldsymbol{\psi}\right)m\left(\boldsymbol{\psi}'\right)d\boldsymbol{\psi}d\boldsymbol{\psi}' \tag{2.7}$$

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

$$\begin{aligned}&\max_{h\in\mathcal{H}}\left\{tr\mathbf{B}_m^{-1}\mathbf{D}_{h,m}\mathbf{B}_m^{-1}-2tr\mathbf{B}_m^{-1}\mathbf{E}_{h,m}+c_h\right\}\\ &=\max_{h\in\mathcal{H}}\Big\{\int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right)\mathbf{z}^T\left(\boldsymbol{\psi}'\right)\mathbf{B}_m^{-2}\mathbf{z}\left(\boldsymbol{\psi}\right)m\left(\boldsymbol{\psi}\right)m\left(\boldsymbol{\psi}'\right)d\boldsymbol{\psi}d\boldsymbol{\psi}'\\ &-2\int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right)\mathbf{z}^T\left(\boldsymbol{\psi}'\right)\mathbf{B}_m^{-1}\mathbf{z}\left(\boldsymbol{\psi}\right)\mu\left(\boldsymbol{\psi}\right)m\left(\boldsymbol{\psi}'\right)d\boldsymbol{\psi}d\boldsymbol{\psi}'+\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}\right)\mu\left(\boldsymbol{\psi}\right)d\boldsymbol{\psi}\Big\}\end{aligned} \tag{2.8}$$

in the case of $IMSE_{2,f,g,h}$. The proof of the following theorem consists of verifying that $m=\mu$ is indeed the minimizer in each case.

Theorem 3. *If the sets $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ are given by* (1.12), (1.13) (*with $g_0(\boldsymbol{\psi})\equiv 1$*) *and* (2.5), *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

$$\tilde{\mathcal{H}}=\left\{h\;\middle|\; 0\le\int_{\mathcal{S}}\int_{\mathcal{S}} h\left(\boldsymbol{\psi},\boldsymbol{\psi}'\right)f\left(\boldsymbol{\psi}\right)f\left(\boldsymbol{\psi}'\right)d\boldsymbol{\psi}d\boldsymbol{\psi}'\le\eta_h^2\int_{\mathcal{S}} f^2\left(\boldsymbol{\psi}\right)d\boldsymbol{\psi},\ \forall f\in L^2\left(\mathcal{S}\right)\right\}, \tag{2.9}$$

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

$$\begin{aligned}&\min_w \max_{f\in\mathcal{F},g\in\mathcal{G},h\in\tilde{\mathcal{H}}} IMSE_{1,f,g,h}=\eta_f^2 ch_{\max}\left[\mathbf{B}_m^{-1}\mathbf{K}_m\mathbf{B}_m^{-1}\right]\\ &+\frac{\sigma_\varepsilon^2}{n}\left(1+\eta_g^2\right)\int_{\mathcal{S}}\left\|\mathbf{B}_m^{-1}\mathbf{z}\left(\boldsymbol{\psi}\right)\right\|m\left(\boldsymbol{\psi}\right)d\boldsymbol{\psi}+\eta_h^2\int_{\mathcal{S}}\left(\left\|\mathbf{B}_m^{-1}\mathbf{z}\left(\boldsymbol{\psi}\right)\right\|m\left(\boldsymbol{\psi}\right)\right)^2 d\boldsymbol{\psi}.\end{aligned} \tag{2.10}$$

Note that the first and second terms in this expression are minimized by the uniform distribution on the sphere. However, this design does not minimize the

second integral at (2.10). This can be seen even in the simplest case $d = 0$, where $z(\boldsymbol{\psi}) = 1$ and $B_m = \int_{\mathcal{S}} m(\boldsymbol{\psi})\, d\psi = 1$. In this case the second integral is minimized by minimizing

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

among probability densities $m(\boldsymbol{\psi})$. The minimizer is easily seen to be the conventional uniform density on S, i.e., $m(\boldsymbol{\psi}) \equiv \left(\int_{\mathcal{S}} d\boldsymbol{\psi} \right)^{-1} = 1/\left(2\pi^2\right)$. As a consequence the uniform distribution on the sphere is in general not minimax optimal, if the class (2.9) is used to address possible correlations in the data.

3. Minimax Optimal Designs on Discrete Subsets

The experimenter faces obvious difficulties in implementing a continuous design such as $\mu(\boldsymbol{\psi})$, for there are then 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 (1.10), and assume that the experimenter takes $n_i = n(\boldsymbol{\psi}_i) \geq 0$ observations $Y_{ij} = Y_j(\boldsymbol{\psi}_i)$ $(j = 1, \ldots n_i)$ at the location $\boldsymbol{\psi}_i$, $i = 1, \ldots, N$. 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 notion 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 $k_1, \ldots, k_N$ at the points $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N$ the experimenter takes approximately $n(\boldsymbol{\psi}_i) = nk_i$ observations at $\boldsymbol{\psi}_i$ $(i = 1, \ldots, N)$. As in Section 2 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 (1.11), (1.12) and (1.13) are

$$\sum_{i=1}^N \mu_i \mathbf{z}(\boldsymbol{\psi}_i) f(\boldsymbol{\psi}_i) = \mathbf{0}, \tag{3.1}$$

$$\mathcal{F}_0 = \left\{ f \;\middle|\; \sum_{i=1}^N \mu_i f^2(\boldsymbol{\psi}_i) \leq \eta_f^2 \right\}, \tag{3.2}$$

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

respectively. In analogy with the approach taken in Section 2.2, we consider the class of covariance structures

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

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

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

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

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

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

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

As in Section 2 we begin with the calculation of the maximum loss over the different classes.

Theorem 4. *Let $\tilde{\mathbf{Z}}$ be an $N\times\left(N-(d+1)^2\right)$ 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* $L(\mathbf{f}) = \mathbf{f}^T\mathbf{MQPQMf} + \mathbf{f}^T\mathbf{Pf}$ *over the set the set* $\mathcal{F}_0$ *defined in* (3.2), *subject to condition* (3.1), *is* $\max_{f\in\mathcal{F}_0} L(\mathbf{f}) = \eta_f^2(1+\lambda_m)$, *where*

$$\lambda_m = ch_{\max}\left[\left(\tilde{\mathbf{Z}}^T\mathbf{P}^{-1}\tilde{\mathbf{Z}}\right)^{-\frac{1}{2}}\tilde{\mathbf{Z}}^T\mathbf{P}^{-1}\mathbf{MQPQMP}^{-1}\tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}^T\mathbf{P}^{-1}\tilde{\mathbf{Z}}\right)^{-\frac{1}{2}}\right]. \quad (3.3)$$

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

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* $g_*(\boldsymbol{\psi}_i)$. *The maximum loss over* $f\in\mathcal{F}_0, g\in\mathcal{G}_0, h\in\mathcal{H}_0$ *is*

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

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

$$\sum_{i=1}^N k_i = \sum_{i=1}^N \frac{m_i}{w_i} = 1. \quad (3.4)$$

For notational convenience we write $\mathbf{r}^T\mathbf{g}_* = \mathbf{s}^T\mathbf{w}$ for $\mathbf{w} = (w_1,\ldots,w_N)^T$, $\mathbf{s} = (s_1,\ldots,s_N)^T$, and $s_i = \left\|\mathbf{A}^{1/2}\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}_i)\right\|^2 m_i g_{*,i}$. 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* (3.4), *by*

$$w_{*,i} = \frac{\gamma_m}{\left\|\mathbf{A}^{1/2}\mathbf{B}_m^{-1}\mathbf{z}(\boldsymbol{\psi}_i)\right\|\sqrt{g_{*,i}}}, \quad (3.5)$$

where

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

The minimum value of $\mathbf{s}^T\mathbf{w}$ *is given by* $\mathbf{s}^T\mathbf{w}_* = \gamma_m^2$. *Thus* $\min_w \max_{f\in\mathcal{F}_0, g\in\mathcal{G}_0, h\in\mathcal{H}_0} IMSE_{j,f,g,h}$ *is given by*

$$\begin{aligned}&\min_w \max_{f\in\mathcal{F}_0, g\in\mathcal{G}_0, h\in\mathcal{H}_0} IMSE_{j,f,g,h} = \eta_f^2\left(1+\lambda_m\right)\\ &+\begin{cases}\frac{\sigma_\varepsilon^2}{n}\gamma_m^2 + \eta_h^2 tr\left[\mathbf{QPQMP}^{-1}\mathbf{M}\right], & \text{if } j=1,\\ \frac{\sigma_\varepsilon^2}{n}\gamma_m^2 + \sigma_\varepsilon^2\boldsymbol{\mu}^T\mathbf{g}_* + \eta_h^2 tr\left[\left(\mathbf{QM}-\mathbf{I}\right)\mathbf{P}^{-1}\left(\mathbf{MQ}-\mathbf{I}\right)\mathbf{P}\right], & \text{if } j=2.\end{cases}\end{aligned} \tag{3.6}$$

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

$$\mathbf{A} = \mathbf{I}_{(d+1)^2}. \tag{3.7}$$

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 also take $g_0(\boldsymbol{\psi}) \equiv 1$. Then (3.6) becomes

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

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

Theorem 5. *If* (3.7) *is satisfied and* $g_0(\boldsymbol{\psi}) \equiv 1$, *then the design* $k_i^* = \mu_i$ $(i = 1, \ldots, N)$ *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* $\eta_f^2 > 0$, $\eta_g^2 > 0$, *and* $\eta_h^2 > 0$.

4. Numerical Comparisons

In this section we construct discrete designs as described in Section 3 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 parameter values. Using $g_0(\boldsymbol{\psi}) \equiv 1$ and unit regression weights in

all cases, this is (from Proposition 3) given by

$$\max_{f\in\mathcal{F}_0,g\in\mathcal{G}_0,h\in\mathcal{H}_0} IMSE_{j,f,g,h} = \eta_f^2\left(1+\lambda_m\right) + \begin{cases} \frac{\sigma_\varepsilon^2}{n}\mathbf{r}^T\mathbf{g}_* + \eta_h^2 tr\left[\mathbf{QPQMP}^{-1}\mathbf{M}\right], & \text{if } j=1, \\ \sigma_\varepsilon^2\left(\boldsymbol{\mu}+\frac{1}{n}\mathbf{r}\right)^T\mathbf{g}_* + \eta_h^2 tr\left[\left(\mathbf{QM}-\mathbf{I}\right)\mathbf{P}^{-1}\left(\mathbf{MQ}-\mathbf{I}\right)\mathbf{P}\right], & \text{if } j=2. \end{cases} \tag{4.1}$$

By defining

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

we can write (4.1) as $\left(\eta_f^2+\frac{\sigma_\varepsilon^2}{n}\left(1+\eta_g^2\right)+\eta_h^2\right)\times\mathcal{L}\left(m;\alpha,\beta\right)$, with

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

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 $m=\mu$ we have

$$\min_m \mathcal{L}\left(m;\alpha,\beta\right) = \alpha + \begin{cases} \left(1-\alpha\right)\left(d+1\right)^2, & \text{if } j=1, \\ \left(1-\alpha\right)\left(d+1\right)^2+n\beta+\gamma\left\{N-2\left(d+1\right)^2\right\}, & \text{if } j=2. \end{cases} \tag{4.2}$$

Minimax designs satisfying our criteria may be constructed, for certain values of N, by the methods of Dette, Melas and Pepelyshev (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 Section 1) be the Legendre polynomial of degree $d+1$. Let $x_1<\cdots<x_{d+1}$ be the zeros, all of which lie in $(-1,1)$, of this polynomial. For $j=1,\ldots,d+1$, 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 μ be the product measure $\mu=\mu_1\otimes\mu_2$, where μ_1 places mass v_j on each of the points $\theta_j=\arccos\left(x_j\right)$, and μ_2 places mass n_0^{-1} on each of n_0 equally spaced points $\phi_k=\delta+\left(2\pi k/n_0\right)\in(-\pi,\pi]$

Table 1. Values of $\mathcal{L}(m;\alpha,\beta)$ for the minimax design M and conventional uniform design U_1 from Example 4.1 ($d=2$).

		$\mathcal{L}(U_1;\alpha,\beta)$					
α	$\mathcal{L}(M;\alpha,\beta)$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$	$\beta=1.0$
0	9.00	9.19	9.22	9.24	9.27	9.29	9.31
0.2	7.40	7.57	7.59	7.61	7.64	7.66	$*$
0.4	5.80	5.94	5.96	5.99	6.01	$*$	$*$
0.6	4.20	4.31	4.33	4.36	$*$	$*$	$*$
0.8	2.60	2.68	2.71	$*$	$*$	$*$	$*$
1.0	1.00	1.06	$*$	$*$	$*$	$*$	$*$

($k=1,\ldots,n_0$). Here δ is any value in $(-(n_0+1)\pi/n_0,-\pi]$; in our examples we use $\delta=-\pi$. Then (3.7) holds and the design μ 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 $\mu'=\mu_1'\otimes\mu_2$, where μ_1' places mass $(d+1)^{-1}$ on each of $d+1$ equally spaced points $\theta_j=\pi j/(d+1)$ ($j=1,\ldots,d+1$). The third (denoted by U_3) is given by $\mu''=\mu_1''\otimes\mu_2$, where μ_1'' places mass $(d+1)^{-1}$ on each of $\theta_j=\arccos(1-(2j/(d+1)))$ ($j=1,\ldots,d+1$). 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 (4.2) holds for all three designs, i.e., all are minimax with respect to their underlying measures μ,μ',μ''. However, of the three, only the design M also satisfies (3.7). For the numerical comparisons three levels of resolution - $d=2,6,13$ - are considered.

Example 4.1. $d=2$. In this case the support points of μ_1, and corresponding probabilities, are

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

Example 4.2. $d=6$. Here the support points of μ_1, and corresponding probabilities, are

$$\begin{array}{l}\theta\text{: } 0.320,\ 0.735,\ 1.153,\ 1.571,\ 1.989,\ 2.406,\ 2.821;\\ v\text{: } 0.065,\ 0.140,\ 0.191,\ 0.209,\ 0.191,\ 0.140,\ 0.065.\end{array}$$

Table 2. Values of $\mathcal{L}(m;\alpha,\beta)$ for the minimax design M and conventional uniform design U_1 from Example 4.2 ($d=6$).

		$\mathcal{L}(U_1;\alpha,\beta)$					
α	$\mathcal{L}(M;\alpha,\beta)$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$	$\beta=1.0$
0	49.00	50.42	51.38	52.34	53.30	54.26	55.22
0.2	39.40	40.59	41.55	42.51	43.47	44.43	*
0.4	29.80	30.75	31.71	32.68	33.64	*	*
0.6	20.20	20.92	21.88	22.84	*	*	*
0.8	10.60	11.09	12.05	*	*	*	*
1.0	1.00	1.26	*	*	*	*	*

Table 3. Values of $\mathcal{L}(m;\alpha,\beta)$ for the minimax design M and conventional uniform design U_1 from Example 4.3 ($d=13$).

		$\mathcal{L}(U_1;\alpha,\beta)$					
α	$\mathcal{L}(M;\alpha,\beta)$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$	$\beta=1.0$
0	196.00	200.22	206.30	212.38	218.46	224.54	230.62
0.2	157.00	160.50	166.58	172.66	178.74	184.82	*
0.4	118.00	120.78	126.86	132.94	139.02	*	*
0.6	79.00	81.06	87.14	93.22	*	*	*
0.8	40.00	41.34	47.42	*	*	*	*
1.0	1.00	1.62	*	*	*	*	*

Example 4.3. $d=13$. The support points of μ_1, and corresponding probabilities, are

$$\begin{aligned}
\theta\text{:}\ & 0.166,\ 0.381,\ 0.597,\ 0.813,\ 1.030,\ 1.246,\ 1.463,\\
& 1.679,\ 1.896,\ 2.112,\ 2.329,\ 2.545,\ 2.761,\ 2.976;\\
v\text{:}\ & 0.018,\ 0.040,\ 0.061,\ 0.079,\ 0.093,\ 0.103,\ 0.108,\\
& 0.108,\ 0.103,\ 0.093,\ 0.079,\ 0.061,\ 0.040,\ 0.018.
\end{aligned}$$

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

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

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

measure is then Kiefer's Φ_p criterion. 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.

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 an NIH grant award. 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 would also like to thank the reviewers for their constructive comments on an earlier version of this manuscript.

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. *Statist. Probab. Lett.* **52**, 215-224.

Bischoff, W. and Miller, F. (2006). Optimal designs which are efficient for lack of fit tests. *Ann. Statist.* **34**, 2015-2025.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **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. *Ann. Statist.* **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. Graphics* **22**, 83-105.

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

Huber, P. J. (1975). Robustness and Designs. In *A Survey of Statistical Design and Linear Models* (Edited by J. N. Srivastava), 287-303. North Holland, Amsterdam

Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York.

Kazhdan, M., Funkhouser, T. and Rusinkiewicz, S. (2003). Rotation invariant spherical harmonic representation of 3D-shape descriptors. *Eurographics Symposium on Geometry Processing* (Edited by L. Kobbelt, P. Schröder and H. Happe). http://www.cs.princeton.edu/~mkazhdan/My-Papers/SGP03.pdf

Kelemen, A., Szekely, G. and Gerig, G. (1999). Elastic Model-Based Segmention 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. (2000). Robust designs for wavelet approximations of regression models. *J. Nonparametr. Stat.***12**, 837-859.

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

Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k. *Anna. Statist.* **10**, 511-525.

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.

Wiens, D. P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Statist. Probab. Lett.* **12**, 217-221.

Wiens, D. P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference***31**, 353-371.

Wiens, D. P. (1998). Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. *J. Amer. Statist. Assoc.* **93**, 1440-1450.

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.

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

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1.

E-mail: doug.wiens@ualberta.ca

(Received March 2007; accepted July 2007)