Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 52 (2008) 4305–4324
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Robust designs for series estimation

Holger Dette[a], Douglas P. Wiens[b,*]

[a] *Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany*
[b] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

Received 29 November 2007; received in revised form 11 February 2008; accepted 11 February 2008
Available online 15 February 2008

Abstract

We discuss optimal design problems for a popular method of series estimation in regression problems. Commonly used design criteria are based on the generalized variance of the estimates of the coefficients in a truncated series expansion and do not take possible bias into account. We present a general perspective of constructing robust and efficient designs for series estimators which is based on the integrated mean squared error criterion. A minimax approach is used to derive designs which are robust with respect to deviations caused by the bias and the possibility of heteroscedasticity. A special case results from the imposition of an unbiasedness constraint; the resulting "unbiased designs" are particularly simple, and easily implemented. Our results are illustrated by constructing robust designs for series estimation with spherical harmonic descriptors, Zernike polynomials and Chebyshev polynomials.

1. Introduction

Series estimators represent a popular technique for estimating the conditional mean $E[Y|\mathbf{x}]$ in regression problems [see e.g. Efromovich (1999)]. They apply to such well-known regression functions as polynomial and trigonometric series, wavelets, spherical harmonic descriptors and Zernike polynomials [see e.g. Härdle et al. (1998), Brechbühler et al. (1995) or Pawlak and Liao (2002) among others]. In these models it is typically assumed that $E[Y|\mathbf{x}]$ is given by a function $\psi(\mathbf{x})$, defined on a bounded set $\mathcal{S} \subset \mathbb{R}^q$ and satisfying $\psi \in L^2_\mu(\mathcal{S})$, where μ denotes a density on the set $\mathcal{S}$ and $L^2_\mu(\mathcal{S})$ the corresponding space of square integrable, real-valued functions. If $z_1, z_2, \ldots$ defines a complete orthonormal system on $\mathcal{S}$ with respect to the measure μ, then the function ψ admits a series expansion of the form

$$\psi(\mathbf{x}) = \sum_{j=1}^{\infty} c_j z_j(\mathbf{x}),$$

where the functions $z_j(\mathbf{x})$ are orthonormal with respect to the weighting function μ, i.e.

$$\int_{\mathcal{S}} z_j(\mathbf{x}) z_k(\mathbf{x})\, \mu(\mathbf{x}) \mathrm{d}\mathbf{x} = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases} \tag{1.1}$$

* Corresponding author. Tel.: +1 780 492 4406; fax: +1 780 492 6826.
E-mail addresses: holger.dette@ruhr-uni-bochum.de (H. Dette), doug.wiens@ualberta.ca (D.P. Wiens).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2008.02.008

Suppose that data $Y_1, \ldots, Y_n$ are observed, with conditional means $E[Y_i|\mathbf{x}_i]$ and with additive error, at 'locations' $\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{S}$. If the experimenter is confident that $E[Y_i|\mathbf{x}_i] = \psi(\mathbf{x}_i)$ then he approximates the function $\psi(\mathbf{x})$ by a suitable truncation of the series, say $\sum_{j=1}^p c_j z_j(\mathbf{x})$. He then fits a linear regression model with p regressors $\{z_j(\mathbf{x})\}_{j=1}^p$, estimates the regression coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ (which can be assumed to agree with the c_j if indeed $E[Y|\mathbf{x}] = \psi(\mathbf{x})$) and predicts a response Y at an arbitrary location $\mathbf{x} \in \mathcal{S}$ by

$$\hat{Y}(\mathbf{x}) = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\theta}}. \tag{1.2}$$

Here $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), \ldots, z_p(\mathbf{x}))^{\mathrm{T}}$ denotes the vector of regression functions.

A commonly used method of estimation stems from the fact that the Fourier coefficients are given by

$$c_j = \int_{\mathcal{S}} \psi(\mathbf{x}) z_j(\mathbf{x})\, \mu(\mathbf{x}) \mathrm{d}\mathbf{x}, \quad j = 1, 2, \ldots.$$

Estimates may then be computed by discretizing these integrals, and replacing $\psi(\mathbf{x})$ by (weighted means of) the observations. This method is often preferred to least squares estimation because it does not involve the inversion of a possibly very high dimensional matrix "$\mathbf{X}^{\mathrm{T}}\mathbf{X}$", which in the intended applications can be numerically unstable. The corresponding estimates will be described in the following section and will be called "direct estimates" throughout this paper. These methods are often employed in shape and image analysis instead of least squares techniques — see, e.g. Brechbühler et al. (1995) and Pawlak and Liao (2002).

Most of the literature on optimal designs in this context concentrates on optimality criteria minimizing the generalized variance of the least squares estimator in the regression model $\psi(\mathbf{x}) = E[Y|\mathbf{x}] = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}$ [see e.g. Lau and Studden (1985), Herzberg and Traves (1994) and Dette et al. (2005) among others]. Recently, Dette et al. (2007) investigated optimal designs minimizing the generalized variance of the least squares and direct estimates of the parameters in truncated Fourier expansions resulting from the system of Zernike polynomials. On the other hand – to the knowledge of the authors – no results on optimal designs are available which address the problem of possible bias obtained by model misspecification, for example by the truncation of the series.

The present paper is devoted to problems of constructing robust and efficient designs for series estimation, which take a variety of possible model specification errors into account. These errors are caused on the one hand by misspecification of the regression function – either because $E[Y|\mathbf{x}]$ can be only approximated by $\psi(\mathbf{x})$, or through the truncation of the series expansion of $\psi(\mathbf{x})$ – and on the other hand by a misspecification of the stochastic error structure.

In Section 2 we review some results on the method of direct estimation, and present two motivating examples which arise from applications in shape and image analysis. A representation of the asymptotic integrated mean squared error is derived, and a minimax criterion is proposed for the determination of optimal designs which are robust with respect to the model assumptions, in particular with respect to the bias obtained from possible misspecification of the regression response, and heteroscedasticity in the data. Sections 3 and 4 consider the problem of determining optimal designs with respect to the new minimax criterion for the general series estimator. In Section 5 we discuss the problem of constructing robust designs explicitly, if a series estimate with Zernike polynomials as basis functions is used to recover a function on a circular domain [see Pawlak and Liao (2002)]. This exploits some of the special structure of Zernike polynomials, and so in Section 6 we discuss the treatment of general series models lacking the structure of these models for shape and image analysis. This treatment is implemented in the context of function approximation with Chebyshev polynomials. Finally, some technical details are given in the Appendix.

2. Direct estimation and robust designs

In this section we formulate an optimal design problem for direct estimation. Let $\|\cdot\|$ denote the Euclidean norm. We introduce a density $k(\mathbf{x})$ on $\mathcal{S}$ and generate a partition of $\mathcal{S}$ into n disjoint sets $\mathcal{S}_i$ such that, as $n \to \infty$,

$$\max_{1 \le i \le n} \left| \int_{\mathcal{S}_i} k(\mathbf{x})\, \mathrm{d}\mathbf{x} - \frac{1}{n} \right| = o\left(n^{-1}\right), \tag{2.1a}$$

$$\max_{1 \le i \le n} \sup_{\mathbf{x}, \mathbf{y} \in \mathcal{S}_i} \|\mathbf{x} - \mathbf{y}\| = o(1). \tag{2.1b}$$

For example, if $\mathcal{S} = [a, b]$ is an interval and k a positive continuous density on $[a, b]$ satisfying

$$\frac{1}{n} = \int_{q_{i-1:n}}^{q_{i:n}} k(\mathbf{x})\mathrm{d}\mathbf{x} + o(n^{-1}),$$

where $a = q_{0:n} < q_{1:n} < \cdots < q_{n:n} = b$ [see Sacks and Ylvisaker (1970)], then (2.1a) is obviously satisfied with $\mathcal{S}_i = (q_{i-1:n}, q_{i:n}](i = 1, \ldots, n)$. Moreover, (2.1b) holds as long as k is strictly positive on its support.

Similarly if $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| \leq 1\}$ is the unit disc, $n = n_1 n_2$ is such that $\min\{n_1, n_2\} \to \infty$ and k_1, k_2 denote positive continuous densities on $[0, 1]$ and $[0, 2\pi]$, respectively, such that

$$\frac{1}{n_1} = \int_{\rho_{i-1}}^{\rho_i} k_1(\rho)\mathrm{d}\rho + o(n_1^{-1}), \qquad \frac{1}{n_2} = \int_{\phi_{j-1}}^{\phi_j} k_2(\phi)\mathrm{d}\phi + o(n_2^{-1}),$$

(where $0 = \rho_0 < \rho_1 < \cdots < \rho_{n_1} = 1$, $0 = \phi_0 < \phi_1 < \cdots < \phi_{n_2} = 2\pi$), then the density $k(\rho, \phi) = k_1(\rho)k_2(\phi)$ satisfies (2.1a) and (2.1b) with

$$\mathcal{S}_{i,j} = \{(\rho\cos\phi, \rho\sin\phi) \mid \rho_{i-1} \leq \rho \leq \rho_i, \phi_{j-1} \leq \phi \leq \phi_j\}$$

$(i = 1, \ldots, n_1, j = 1, \ldots, n_2)$.

It follows that if the function $\alpha(\mathbf{x})$ is uniformly continuous on $\mathcal{S}$, and the points $\mathbf{x}_i$ are arbitrarily chosen representative elements of $\mathcal{S}_i$, then

$$\int_{\mathcal{S}} \alpha(\mathbf{x})k(\mathbf{x})\mathrm{d}\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} \alpha(\mathbf{x}_i) + o(1). \tag{2.2}$$

This discretization of the integral will be used frequently in the following discussion. Throughout this paper we assume that n independent observations from the model

$$Y_i = Y(\mathbf{x}_i) = E[Y|\mathbf{x}_i] + \varepsilon(\mathbf{x}_i) \quad i = 1, \ldots, n \tag{2.3}$$

are available, where the errors $\varepsilon(\mathbf{x}_1), \ldots, \varepsilon(\mathbf{x}_n)$ are centred random variables with third absolute moments.

For non-negative weights $\{w(\mathbf{x}_i)\}$ we propose the direct estimate

$$\hat{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)\frac{Y(\mathbf{x}_i)w(\mathbf{x}_i)\mu(\mathbf{x}_i)}{k(\mathbf{x}_i)}. \tag{2.4}$$

This is motivated by the fact that for constant weights $w(\mathbf{x}_i) \equiv 1$, if $E[Y|\mathbf{x}] = \psi(\mathbf{x})$ then the estimate is asymptotically unbiased for $\mathbf{c} = \left(c_1, \ldots, c_p\right)^{\mathrm{T}}$:

$$E[\hat{\theta}] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)E[Y|\mathbf{x}_i]\frac{\mu(\mathbf{x}_i)}{k(\mathbf{x}_i)} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\psi(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} + o(1) = \mathbf{c} + o(1).$$

We have used (1.1) and (2.2) to establish the last two equalities. We allow for non-constant weights in anticipation of possible heteroscedasticity, and also for a more precise tuning of the estimate. We will norm the weights to have a unit average, viz.

$$\int_{\mathcal{S}} w(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x} = 1,$$

which implies that the function

$$m(\mathbf{x}) \stackrel{\text{def}}{=} w(\mathbf{x})\mu(\mathbf{x}) \tag{2.5}$$

is a probability density on $\mathcal{S}$. It will prove convenient to optimize over $m(\cdot)$ rather than over $w(\cdot)$.

In the following discussion we shall frequently use the identity

$$\int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x})\,\mu(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbf{I}_p, \tag{2.6}$$

which follows directly from (1.1).

A practitioner will generally readily acknowledge that the truncated regression model may be only approximately correct. For instance, perhaps because only finitely many of the $z_j(\mathbf{x})$ are being fitted, or because of more general unease about the adequacy of the fitted model, the experimenter may feel that $\psi(\mathbf{x})$ is only an approximation to $E[Y|\mathbf{x}]$, which is in turn only approximately given by a linear combination of $z_1(\mathbf{x}), \ldots, z_p(\mathbf{x})$. If so, measures are needed which are robust against the ensuing prediction bias. As well, the meaning of the parameter $\boldsymbol{\theta}$ becomes unclear in this context. We shall thus *define*

$$\boldsymbol{\theta} = \arg\min_{\mathbf{t}} \int_{\mathcal{S}} \left(E[Y|\mathbf{x}] - \mathbf{z}^{\mathrm{T}}(\mathbf{x})\,\mathbf{t} \right)^2 \mu(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{2.7}$$

as the parameter corresponding to the best L^2_μ-approximation of the regression function by elements of the space spanned by the functions $z_1, \ldots, z_p$, and define

$$f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta} \tag{2.8}$$

as the deviation of this approximation from the true mean response. Furthermore, assumptions made by the experimenter concerning the homoscedasticity of the measurement errors (or a particular, assumed heteroscedasticity structure) may also be in doubt. An alternative description of the model (2.3) is then given by

$$Y(\mathbf{x}_i) = \mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}) + \varepsilon(\mathbf{x}_i), \quad i = 1, \ldots, n \tag{2.9}$$

where:

1. The function $f(\mathbf{x})$ accounts for inadequacies in the approximation $E[Y|\mathbf{x}] \approx \mathbf{z}^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\theta}$ obtained from model misspecification.
2. The random variables $\varepsilon(\mathbf{x})$ are additive measurement errors, pairwise independent, and with possibly heterogeneous variances $\sigma^2_\varepsilon g(\mathbf{x})$.

We impose bounds on the magnitudes of the approximation error f, and of the variance function g centred at the experimenter's best guess g_0, that is

$$\int_{\mathcal{S}} f^2(\mathbf{x})\mu(\mathbf{x})\,\mathrm{d}\mathbf{x} \le \eta_f^2, \tag{2.10}$$

$$\sup_{\mathbf{x}\in\mathcal{S}} |g(\mathbf{x}) - g_0(\mathbf{x})| \le \eta_g^2, \tag{2.11}$$

where η_f^2 and η_g^2 are given non-negative constants. Note that Eqs. (2.7) and (2.8) give

$$\int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x})\,\mu(\mathbf{x})\mathrm{d}\mathbf{x} = \mathbf{0}. \tag{2.12}$$

From (2.9) we see that if the specification $E[Y|\mathbf{x}] = \psi(\mathbf{x})$ is correct then $\boldsymbol{\theta} = \mathbf{c}$ and $f(\mathbf{x}) = \sum_{j=p+1}^{\infty} c_j z_j(\mathbf{x})$.

Let $\mathcal{F}$ and $\mathcal{G}$ be, respectively, the class of functions f constrained by (2.10) and (2.12), and the class of positive functions g constrained by (2.11). We assume throughout this paper that $f, g, m, (m/k)$ and $z_1, \ldots, z_p$ are uniformly continuous, which will allow us to apply the approximation (2.2) whenever it is necessary.

The fitted model assumes $f \equiv 0$; if the experimenter believes that the error variances are homogeneous then he will as well take $g_0 \equiv 1$. Violations of these specifications will however increase the mean squared error of the predictions, and in the following discussion we will determine designs which are robust with respect to such misspecifications of the model. The design and estimation problem is to choose a density $k(\cdot)$ and weights $w(\cdot) = m(\cdot)/\mu(\cdot)$ in order to minimize the asymptotic integrated mean squared error of the predictions. Before we derive an explicit expression for this mean squared error we shall illustrate the methodology of series estimation in two examples, which have found considerable interest in applications.

Example 2.1. For describing three-dimensional shapes a common model employs spherical harmonic descriptors [see for example Brechbühler et al. (1995) for some general discussion and Ding et al. (2000), who analyze the shapes of Citrus species via spherical harmonics]. In this model the distance Y from the origin to the boundary of the shape is

described in terms of two angles $\theta \in [0, \pi]$, $\phi \in (-\pi, \pi]$ and, with $\mathbf{x} = (\theta, \phi)$ and $p = (d+1)^2$ the regression functions are arranged as

$$\mathbf{z}(\mathbf{x}) = \left(\mathbf{z}_0^{\mathrm{T}}(\mathbf{x}), \mathbf{z}_1^{\mathrm{T}}(\mathbf{x}), \ldots, \mathbf{z}_d^{\mathrm{T}}(\mathbf{x})\right)^{\mathrm{T}}, \quad \text{where}$$
$$\mathbf{z}_k^{\mathrm{T}}(\mathbf{x}) = \left(Z_k^{-k}(\mathbf{x}), Z_k^{-k+1}(\mathbf{x}), \ldots, Z_k^{k-1}(\mathbf{x}), Z_k^{k}(\mathbf{x})\right).$$

Here

$$Z_k^m(\mathbf{x}) = \begin{cases} \sqrt{2k+1}\, P_k^0(\cos\theta), & m = 0,\ k \geq 0, \\ \sqrt{2(2k+1)\dfrac{(k-m)!}{(k+m)!}}\, P_k^m(\cos\theta)\cos(m\phi), & m = 0, \ldots, k,\ k > 0, \\ \sqrt{2(2k+1)\dfrac{(k+m)!}{(k-m)!}}\, P_k^{-m}(\cos\theta)\sin(m\phi), & m = -k, \ldots, -1,\ k > 0, \end{cases}$$

and P_k^m is the mth associated Legendre function of degree k. The functions Z_k^m are orthonormal over the design space $\mathcal{S} = \{\theta \in [0, \pi],\ \phi \in (-\pi, \pi]\}$ with respect to the density function $\mu(\mathbf{x}) = \sin\theta/4\pi$, where $\mathbf{x} = (\theta, \phi)$. We note for future reference that it is shown in Dette et al. (2005) that

$$\|\mathbf{z}(\mathbf{x})\| = d + 1. \tag{2.13}$$

Example 2.2. Let $\mathcal{S}$ be the unit disc centred at the origin, and suppose we make noisy observations of a function in $L^2(\mathcal{S}) = \{f : \mathcal{S} \to \mathbb{R} \mid \int_{\mathcal{S}} f^2(\mathbf{x})\mathrm{d}\mathbf{x} < \infty\}$. As pointed out by Pawlak and Liao (2002) such a function admits an expansion in terms of Zernike polynomials, and a truncated series with coefficients estimated from the data is used to reconstruct the function f. To be precise, we consider a radius $\rho \in [0, 1]$, an angle $\phi \in (0, 2\pi]$ and a predictor $\mathbf{x} = (\rho, \phi)$. The regression model has $p = (d+1)(d+2)/2$ regression functions defined by

$$\mathbf{z}(\mathbf{x}) = \left(\mathbf{z}_0^{\mathrm{T}}(\mathbf{x}), \mathbf{z}_1^{\mathrm{T}}(\mathbf{x}), \ldots, \mathbf{z}_d^{\mathrm{T}}(\mathbf{x})\right)^{\mathrm{T}}, \quad \text{where}$$
$$\mathbf{z}_k^{\mathrm{T}}(\mathbf{x}) = \left(Z_k^{-k}(\mathbf{x}), Z_k^{-k+2}(\mathbf{x}), \ldots, Z_k^{k-2}(\mathbf{x}), Z_k^{k}(\mathbf{x})\right).$$

Here $Z_k^m(\mathbf{x})$ is a Zernike polynomial (Zernike, 1934) defined by:

$$Z_k^m(\mathbf{x}) = \sqrt{k+1}\, R_k^{|m|}(\rho) \cdot a_k^m(\phi),$$
$$a_k^m(\phi) = \sqrt{\frac{2}{1 + I(m=0)}} \cdot \begin{cases} \sin(m\phi), & m < 0, \\ \cos(m\phi), & m \geq 0, \end{cases}$$
$$R_k^m(\rho) = \begin{cases} \displaystyle\sum_{l=0}^{\frac{k-m}{2}} \frac{(-1)^l (k-l)!}{l!\left(\frac{k+m}{2} - l\right)!\left(\frac{k-m}{2} - l\right)!}\rho^{k-2l}, & k - m \text{ even}, \\ 0, & k - m \text{ odd}. \end{cases}$$

The Zernike polynomials are orthonormal on the set $\mathcal{S} = \{(\rho, \phi) | \rho \in [0, 1],\ \phi \in (0, 2\pi]\}$ with respect to the density function

$$\mu(\mathbf{x}) = \mu^{(1)}(\rho)\, \mu^{(2)}(\phi),$$

where $\mu^{(2)}(\phi) = (2\pi)^{-1}$ is the uniform density on the interval $(0, 2\pi]$ and $\mu^{(1)}(\rho) = 2\rho$ is a density on the interval $[0, 1]$. It will prove helpful to represent the vector of regression functions $\mathbf{z}(\mathbf{x})$ as

$$\mathbf{z}(\mathbf{x}) = \mathbf{R}(\rho)\, \mathbf{a}(\phi),$$

where $\mathbf{R}(\rho)$ is the $p \times p$ diagonal matrix containing the terms $\sqrt{k+1}\, R_k^{|m|}(\rho)$ on its diagonal and $\mathbf{a}(\phi)$ is the p-dimensional vector containing the terms $a_k^m(\phi)$, all ordered as they appear in the vector $\mathbf{z}(\mathbf{x})$. Then

$$\int_0^{2\pi} \mathbf{a}(\phi)\mathbf{a}^{\mathrm{T}}(\phi)\, \mu^{(2)}(\phi)\mathrm{d}\phi = \mathbf{I}_p = \int_0^1 \mathbf{R}^2(\rho)\, \mu^{(1)}(\rho)\, \mathrm{d}\rho \tag{2.14}$$

[see Szëgo (1975)] and the orthogonality relations in (2.6) follow from Fubini's theorem. Another useful result is that the function

$$\|\mathbf{z}(\mathbf{x})\|^2 = \operatorname{tr} \mathbf{R}^2(\rho) \stackrel{\text{def}}{=} \Theta_d(\rho), \tag{2.15}$$

depends only on the parameter ρ alone, as shown in Dette et al. (2007, Theorem 4.1). Some applications and further details can be found in Pawlak and Liao (2002). Common applications of this model are in optics (Wyant and Creath, 1993) and image analysis (Liao and Pawlak, 1996; Kim and Kim, 1999).

As specific examples we consider the cases $d \leq 4$, for which the Zernike polynomials are the elements of the vectors $\mathbf{z}_k(\mathbf{x})$ $(k = 0, \ldots, d)$ with

$$\begin{aligned}
\mathbf{z}_0^{\mathrm{T}}(\mathbf{x}) &= 1; \qquad \mathbf{z}_1^{\mathrm{T}}(\mathbf{x}) = \left(\sqrt{2}\rho\sin\phi, \sqrt{2}\rho\cos\phi\right);\\
\mathbf{z}_2^{\mathrm{T}}(\mathbf{x}) &= \left(\sqrt{3}\rho^2\sin(2\phi), \sqrt{3}\left(2\rho^2-1\right), \sqrt{3}\rho^2\cos(2\phi)\right);\\
\mathbf{z}_3^{\mathrm{T}}(\mathbf{x}) &= \left(\sqrt{4}\rho^3\sin(3\phi), \sqrt{4}\left(3\rho^3-2\rho\right)\sin\phi, \sqrt{4}\left(3\rho^3-2\rho\right)\cos\phi, \sqrt{4}\rho^3\cos(3\phi)\right);\\
\mathbf{z}_4^{\mathrm{T}}(\mathbf{x}) &= \left(\sqrt{5}\rho^4\sin(4\phi), \sqrt{5}\left(4\rho^4-3\rho^2\right)\sin(2\phi), \sqrt{5}\left(6\rho^4-6\rho^2+1\right),\right.\\
&\quad \left.\sqrt{5}\left(4\rho^4-3\rho^2\right)\cos(2\phi), \sqrt{5}\rho^4\cos(4\phi)\right).
\end{aligned}$$

In order to present an asymptotic representation of the integrated mean squared error we recall the notation (2.5), and define

$$\begin{aligned}
\mathbf{b}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)\, f(\mathbf{x}_i)\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)},\\
\mathbf{B}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)\, \frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i),\\
\mathbf{C}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)\, \frac{m^2(\mathbf{x}_i)g(\mathbf{x}_i)}{k^2(\mathbf{x}_i)}\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i).
\end{aligned}$$

Theorem 1. *Define, for* $\mathbf{x} \in \mathcal{S}$, $\tau^3(\mathbf{x}) = E\left[\left|\varepsilon(\mathbf{x})/\left(\sigma_\varepsilon\sqrt{g(\mathbf{x})}\right)\right|^3\right]$. *Assume that*

(i) $\tau^3(\mathbf{x})$ *is bounded on* $\mathcal{S}$, *and that*

(ii) $\mathbf{C}_n$ *is positive definite, and its minimum eigenvalue is bounded away from 0, for sufficiently large n.*

Then the direct estimate (2.4) is asymptotically normally distributed:

$$\hat{\boldsymbol{\theta}} \sim AN\left(\mathbf{B}_n\boldsymbol{\theta} + \mathbf{b}_n, \frac{\sigma_\varepsilon^2}{n}\mathbf{C}_n\right). \tag{2.16}$$

Remark 1. Assumption (i) of Theorem 1 is very mild; it is satisfied, for instance, if the errors are Gaussian since then $\tau^3(\mathbf{x})$ is constant. Assumption (ii) is of course standard.

Now define

$$\mathbf{b}_{f,m} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) m(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{2.17a}$$

$$\mathbf{B}_m = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x}) m(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{2.17b}$$

$$\mathbf{C}_{g,k,m} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) \frac{m^2(\mathbf{x})g(\mathbf{x})}{k(\mathbf{x})}\mathbf{z}^{\mathrm{T}}(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{2.17c}$$

Under the conditions leading to (2.2), these are the limits of $\mathbf{b}_n$, $\mathbf{B}_n$ and $\mathbf{C}_n$ respectively.

The loss function for the problem is taken to be the asymptotic integrated mean squared error of the predictions, defined as

$$\text{IMSE}(k, m; f, g) = \int_{\mathcal{S}} E\left[\left\{\hat{Y}(\mathbf{x}) - E[Y(\mathbf{x})]\right\}^2\right]\mu(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathcal{S}} \left\{\text{VAR}\left[\hat{Y}(\mathbf{x})\right] + \text{BIAS}^2\left[\hat{Y}(\mathbf{x})\right]\right\}\mu(\mathbf{x})\mathrm{d}\mathbf{x},$$

where

$$\text{BIAS}\left[\hat{Y}(\mathbf{x})\right] = E\left[\hat{Y}(\mathbf{x})\right] - \left[\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x})\right],$$

$\hat{Y}(\mathbf{x})$ is defined by (1.2), and the variance and bias are given their asymptotic values derived from (2.16) and (2.17). Straightforward calculations, utilizing (2.6) and (2.12), yield

$$\text{IMSE}(k, m; f, g) = \text{tr}\left[\frac{\sigma_\varepsilon^2}{n}\mathbf{C}_{g,k,m}\right] + \left\|\left(\mathbf{B}_m - \mathbf{I}_p\right)\boldsymbol{\theta} + \mathbf{b}_{f,m}\right\|^2 + \int_{\mathcal{S}} f^2(\mathbf{x})\mu(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Since IMSE depends on f and g we aim to solve the following minimax problem:

$$\min_{k,m}\max_{f\in\mathcal{F},g\in\mathcal{G}} \text{IMSE}(k, m; f, g).$$

The minimizing density is called a minimax design. In some cases $\mathbf{B}_m \neq \mathbf{I}_p$ and then IMSE also depends on $\boldsymbol{\theta}$; in such cases we shall endow $\boldsymbol{\theta}$ with a prior distribution $\pi(\boldsymbol{\theta})$ satisfying

$$E_\pi[\boldsymbol{\theta}] = \mathbf{0},\ E_\pi\left[\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right] = \boldsymbol{\Sigma} \leq \eta_{\boldsymbol{\theta}}^2\mathbf{I}_p, \tag{2.18}$$

(the last inequality refers to the Loewner ordering) and consider instead

$$\min_{k,m}\max_{f\in\mathcal{F},g\in\mathcal{G}}\max_{\boldsymbol{\Sigma}} E_\pi[\text{IMSE}(k, m; f, g)]. \tag{2.19}$$

Priors $\pi(\cdot)$ with non-zero means can also be considered here, although the maximization problem then becomes more complex.

3. Unbiased direct estimators

If constant weights $w(\mathbf{x}_i) = 1\ \ (i = 1, \ldots, n)$ are used in the direct estimate, it follows from Theorem 1 and the orthonormality conditions (1.1) that this estimate is asymptotically unbiased. Note that in the presence of (2.12), the condition $\mathbf{b}_{f,m} = \mathbf{0}$ is satisfied. On the other hand, if unbiasedness of $\hat{Y}(\mathbf{x})$ for $E[Y|\mathbf{x}]$ is *imposed*, so that the condition $\mathbf{b}_{f,m} = \mathbf{0}$ must hold for all f simultaneously with (2.12), then necessarily $m(\cdot) = \mu(\cdot)$; hence the weights must satisfy $w(\mathbf{x}) \equiv 1$. In such cases we have from the orthonormality relations (2.6) that $\mathbf{B}_m = \mathbf{I}_p$. Consequently, we obtain

$$\begin{aligned}\max_{f,g}\text{IMSE}(k, \mu; f, g) &= \max_{g}\text{tr}\left[\frac{\sigma_\varepsilon^2}{n}\mathbf{C}_{g,k,\mu}\right] + \max_{f}\int_{\mathcal{S}} f^2(\mathbf{x})\mu(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &= \frac{\sigma_\varepsilon^2}{n}\int_{\mathcal{S}}\|\mathbf{z}(\mathbf{x})\|^2\frac{\mu^2(\mathbf{x})\left[g_0(\mathbf{x}) + \eta_g^2\right]}{k(\mathbf{x})}\mathrm{d}\mathbf{x} + \eta_f^2.\end{aligned}$$

With the notation

$$g_*(\mathbf{x}) \overset{\text{def}}{=} g_0(\mathbf{x}) + \eta_g^2, \tag{3.1}$$

an optimal design density $k_*(\mathbf{x})$ will thus minimize

$$\int_{\mathcal{S}}\|\mathbf{z}(\mathbf{x})\|^2\frac{\mu^2(\mathbf{x})g_*(\mathbf{x})}{k(\mathbf{x})}\mathrm{d}\mathbf{x},$$

subject to the normalizing condition $\int_{\mathcal{S}} k(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. The solution to this problem is presented in the following proposition. In order that we may refer to it again later, we state it for general weights.

Proposition 1. *The density $k(\cdot)$ minimizing the functional*

$$\Psi_1(k; m) \stackrel{\text{def}}{=} \int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|^2 \frac{m^2(\mathbf{x})\, g_*(\mathbf{x})}{k(\mathbf{x})} \mathrm{d}\mathbf{x}$$

for fixed $m(\cdot)$ is given by

$$k_*(\mathbf{x}; m) = \frac{\|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x}) \sqrt{g_*(\mathbf{x})}}{\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x}) \sqrt{g_*(\mathbf{x})} \mathrm{d}\mathbf{x}}, \tag{3.2}$$

with minimal value

$$\Psi_1(k_*; m) = \left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x}) \sqrt{g_*(\mathbf{x})} \mathrm{d}\mathbf{x}\right]^2.$$

With $w(\mathbf{x}) \equiv 1$ we have $m = \mu$ in Proposition 1 and we obtain that the minimax design for the direct estimate, under the requirement of unbiasedness, is given by

$$k_*(\mathbf{x}; \mu) = \frac{\|\mathbf{z}(\mathbf{x})\|\, \mu(\mathbf{x}) \sqrt{g_*(\mathbf{x})}}{\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, \mu(\mathbf{x}) \sqrt{g_*(\mathbf{x})} \mathrm{d}\mathbf{x}}, \tag{3.3}$$

and

$$\min_k \max_{f,g} \mathrm{IMSE}(k, \mu; f, g) = \frac{\sigma_\varepsilon^2}{n} \left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, \mu(\mathbf{x}) \sqrt{g_*(\mathbf{x})} \mathrm{d}\mathbf{x}\right]^2 + \eta_f^2.$$

We illustrate these results in the two examples considered before.

Example 3.1 (Example 2.1 *Continued*). By virtue of (2.13), the minimax unbiased design for series estimation with direct estimates and spherical harmonic regressors is obtained from (2.1a) with density $k(\theta, \phi) \propto \sin\theta \sqrt{g_*(\theta, \phi)}$. If a homoscedastic model is fitted, so that g_* is constant, then the optimal density satisfies $k(\theta, \phi) \propto \sin\theta$ and $k(\cdot, \cdot)$ generates a product design. One factor has density $(\sin\theta)/2$ on the interval $[0, \pi]$ and the other is a uniform distribution on the interval $(-\pi, \pi]$. The design can be easily implemented by taking observations at

$$(\theta_i, \phi_j) = \left(\arccos\left(1 - \frac{2(i-1)}{n_1 - 1}\right), -\pi + \frac{2j}{n_2}\right)$$

for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ with $n = n_1 n_2$. This design was also found to possess further favourable robustness properties when used with the least squares estimator in this model — see Dette and Wiens (in press).

Example 3.2 (Example 2.2 *Continued*). Consider the problem of series estimation with Zernike polynomials using direct estimates with equal weights. By virtue of (2.15), the minimax unbiased design for observations on the unit disc is obtained from (2.1a) with $k(\rho, \phi) \propto \rho \sqrt{\Theta_d(\rho)\, g_*(\rho, \phi)}$. A natural assumption here is that the error variances depend on the regressors only through the distance ρ from the origin, so that $g_*(\rho, \phi) = g_*(\rho)$. In this case the minimax design is again a product design, with one factor having density $\rho\sqrt{\Theta_d(\rho)\, g_*(\rho)} / \int_0^1 \rho\sqrt{\Theta_d(\rho)\, g_*(\rho)}\mathrm{d}\rho$ for $\rho \in [0, 1]$ and the other being a uniform distribution on the interval $(0, 2\pi]$. For a constant function $g_*(\rho)$ this design was also shown to be A-optimal by Dette et al. (2007). See Figs. 1–3 for some comparative examples. In each of these figures plot (c) pertains to this example; plots (a) and (b) pertain to Example 5.1, which continues this Zernike polynomial application without the requirement of unbiasedness.

4. Minimax designs and weights — general theory

For general weights, the IMSE of the direct estimate (2.4) also depends on the parameter $\boldsymbol{\theta}$, and we invoke the prior distribution defined in (2.18). In this case we have

$$E_\pi\left[\left\|\left(\mathbf{B}_m - \mathbf{I}_p\right)\boldsymbol{\theta} + \mathbf{b}_{f,m}\right\|^2\right] = \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)\boldsymbol{\Sigma}\left(\mathbf{B}_m - \mathbf{I}_p\right)^{\mathrm{T}}\right] + \left\|\mathbf{b}_{f,m}\right\|^2,$$

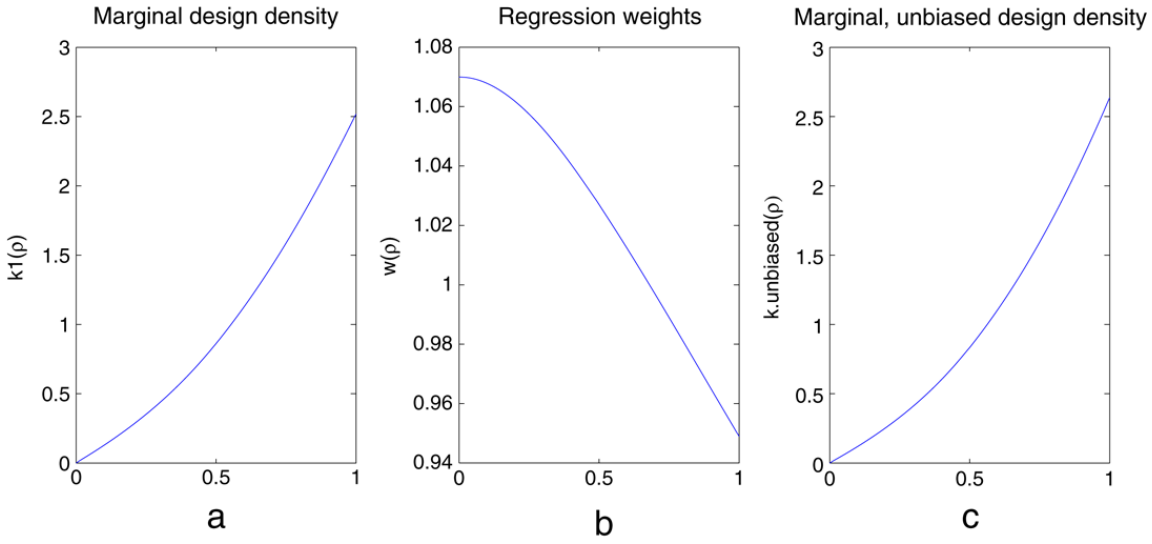

Fig. 1. (a) Marginal design density $k_*^{(1)}(\rho)$ and (b) weights $w_*(\rho)$ for $d = 1$, $s = .7$, $t = .05$; $I^* = \{1\}$. Plot (c) is the marginal of the unbiased design density (3.3).

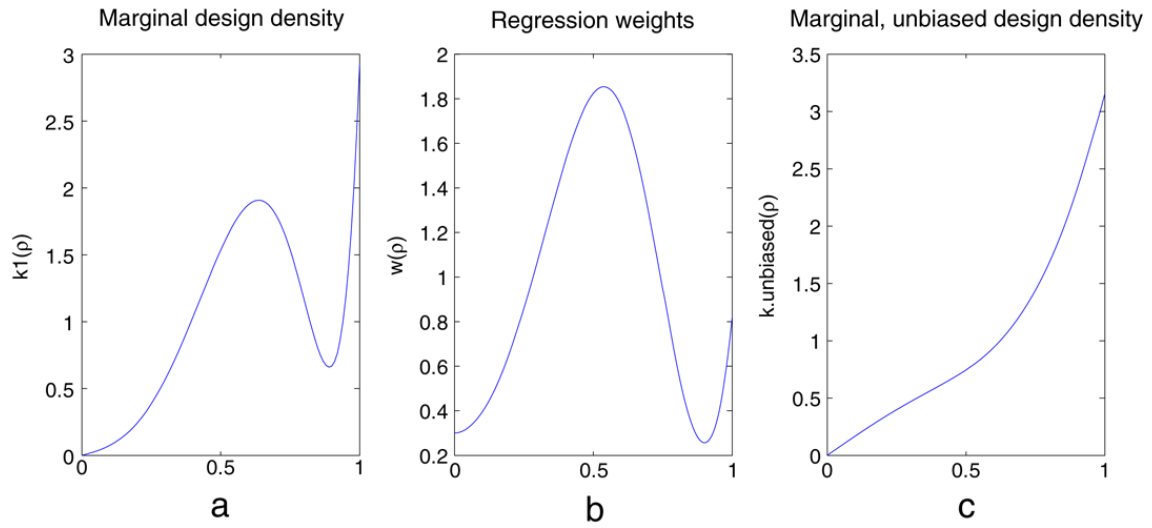

Fig. 2. (a) Marginal design density $k_*^{(1)}(\rho)$ and (b) weights $w_*(\rho)$ for $d = 2$, $s = .1$, $t = .4$; $i^* = 1$. Plot (c) is the marginal of the unbiased design density (3.3).

which yields

$$\max_{\boldsymbol{\Sigma}} E_\pi\left[\mathrm{IMSE}\left(k, m; f, g\right)\right] = \mathrm{tr}\left[\frac{\sigma_\varepsilon^2}{n}\mathbf{C}_{g,k,m}\right] + \eta_{\boldsymbol{\theta}}^2 \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right] + \left\|\mathbf{b}_{f,m}\right\|^2 + \int_{\mathcal{S}} f^2(\mathbf{x})\,\mu(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

For the determination of the optimal value in (2.19) we first maximize with respect to $f \in \mathcal{F}$, then with respect to $g \in \mathcal{G}$. Note that only two terms in the above expression depend on the function f. Consequently, the maximum over f is given in the following result. The proof is very similar to that of Theorem 1 in Wiens (1992) and is therefore omitted.

Proposition 2. *For fixed m the maximum of the function*

$$\Psi_2(f; m) \stackrel{\mathrm{def}}{=} \left\|\mathbf{b}_{f,m}\right\|^2 + \int_{\mathcal{S}} f^2(\mathbf{x})\mu(\mathbf{x})\mathrm{d}\mathbf{x}$$

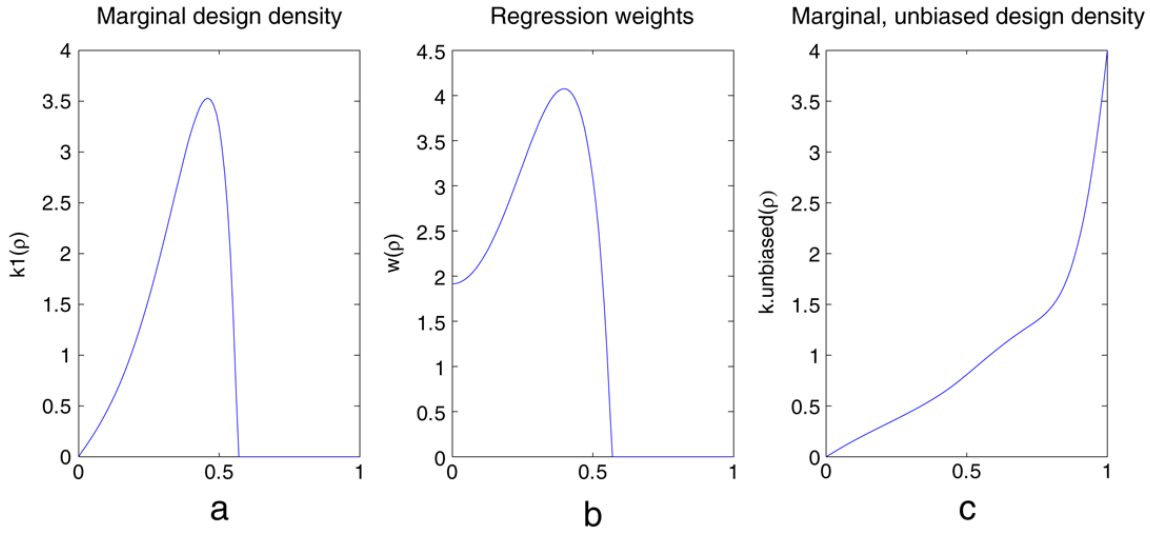

Fig. 3. (a) Marginal design density $k_*^{(1)}(\rho)$ and (b) weights $w_*(\rho)$ for $d = 4$, $s = .2$, $t = .8$; $i^* = 5$. Plot (c) is the marginal of the unbiased design density (3.3).

over the class of functions f satisfying (2.10) and (2.12) is given by

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

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

$$\mathbf{K}_m \overset{\text{def}}{=} \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\,\mathbf{z}^{\mathrm{T}}(\mathbf{x})\frac{m^2(\mathbf{x})}{\mu(\mathbf{x})}\mathrm{d}\mathbf{x}.$$

Moreover, the matrix $\mathbf{K}_m - \mathbf{B}_m^2$ is non-negative definite; in particular, we have for any vector $\boldsymbol{\alpha} \in \mathbb{R}^p$

$$\boldsymbol{\alpha}^{\mathrm{T}}\left(\mathbf{K}_m - \mathbf{B}_m^2\right)\boldsymbol{\alpha} = \int_{\mathcal{S}} \left\{ \boldsymbol{\alpha}^{\mathrm{T}}\left[\frac{m(\mathbf{x})}{\mu(\mathbf{x})}\mathbf{I}_p - \mathbf{B}_m\right]\mathbf{z}(\mathbf{x}) \right\}^2 \mu(\mathbf{x})\mathrm{d}\mathbf{x} \geq 0. \tag{4.1}$$

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

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

where $\boldsymbol{\alpha}_m$ is any solution to 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 $\mathbf{K}_m - \mathbf{B}_m^2$, corresponding to the maximum eigenvalue, normalized such that $\|\boldsymbol{\beta}_m\| = 1$.

The maximization with respect to the function $g \in \mathcal{G}$, and the minimization with respect to the density k, are straightforward. Recalling the definition of $g_*(\cdot)$ in (3.1) we have, using Proposition 1,

$$\max_g \operatorname{tr}\mathbf{C}_{g,k,m} = \int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|^2 \frac{m^2(\mathbf{x})}{k(\mathbf{x})} g_*(\mathbf{x})\mathrm{d}\mathbf{x},$$

$$\min_k \max_g \operatorname{tr}\mathbf{C}_{g,k,m} = \max_g \operatorname{tr}\mathbf{C}_{g,k_*,m} = \left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x})\sqrt{g_*(\mathbf{x})}\mathrm{d}\mathbf{x}\right]^2,$$

where the minimum is attained for the density $k_*(\mathbf{x}; m)$ given by (3.2). Combining these observations, we obtain

$$\begin{aligned}\Phi(m) &\overset{\text{def}}{=} \min_k \max_{f\in\mathcal{F}, g\in\mathcal{G}} \max_{\boldsymbol{\Sigma}} E_\pi\left[\mathrm{IMSE}(k, m; f, g)\right] \\ &= \eta_f^2 \cdot \left(ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1 \right) + \frac{\sigma_\varepsilon^2}{n}\left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x})\sqrt{g_*(\mathbf{x})}\mathrm{d}\mathbf{x}\right]^2 + \eta_{\boldsymbol{\theta}}^2 \operatorname{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right],\end{aligned}$$

and for the solution of the optimization problem (2.19) it remains to determine the optimal weights $w_*(\cdot) = m_*(\cdot)/\mu(\cdot)$ with

$$m_* = \arg\min_m \Phi(m).$$

Suppose now that the experimenter anticipates homoscedasticity, but seeks protection against violations of these assumptions. In this case we set $g_0(\cdot) \equiv 1$ and obtain

$$\Phi(m) = \eta_f^2 \cdot \left(ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1\right) + \frac{\sigma_\varepsilon^2}{n}\left(1 + \eta_g^2\right)\left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\| m(\mathbf{x})\mathrm{d}\mathbf{x}\right]^2 + \eta_{\boldsymbol{\theta}}^2 \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right].$$

With the notation $\eta_T^2 \overset{\text{def}}{=} \eta_f^2 + \frac{\sigma_\varepsilon^2}{n}\left(1 + \eta_g^2\right) + \eta_{\boldsymbol{\theta}}^2$ and

$$s = \frac{\eta_f^2}{\eta_T^2}, \qquad t = \frac{\frac{\sigma_\varepsilon^2}{n}\left(1 + \eta_g^2\right)}{\eta_T^2},$$

we have that $\Phi(m)$ is proportional (up to the constant η_T^2) to the function

$$\Phi_0(m; s, t) \overset{\text{def}}{=} s \cdot \left(ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1\right) + t \cdot \left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\| m(\mathbf{x})\mathrm{d}\mathbf{x}\right]^2 + (1 - s - t) \cdot \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right]. \quad (4.2)$$

The parameters $s, t \in [0, 1]$ (with $s + t \leq 1$) may be chosen by the experimenter according to his assessment of the relative importance of the various sources of IMSE from which he seeks protection.

Example 4.1 (Example 2.1 *Continued*). For spherical harmonic regression, note from (2.13) that $\|\mathbf{z}(\mathbf{x})\|$ is constant. Consequently the density $m(\mathbf{x}) = \mu(\mathbf{x})$ is minimax optimal. This follows from the fact that this choice yields $\mathbf{K}_m = \mathbf{B}_m = \mathbf{I}_p$, so that the first and third term on the right side of (4.2) are simultaneously minimized, while the second is independent of $m(\cdot)$. The corresponding minimax design for spherical harmonic regression is as described in Example 3.1.

It is worthwhile to mention that this conclusion is always possible, provided that $\|\mathbf{z}(\mathbf{x})\|$ is constant. Popular examples satisfying this property are the Haar wavelet model [see Herzberg and Traves (1994) and Oyet and Wiens (2000)] and the trigonometric regression model [see Karlin and Studden (1966)].

If $\|\mathbf{z}(\mathbf{x})\|$ is not constant, then the problem of determining the minimum of the function (4.2) for arbitrary $s, t \in [0, 1]$ is more complicated and will be investigated in the remaining part of this section. We begin with the discussion of several limiting values of the parameters s, t, for which the solution of the optimization problem

$$m_* = \arg\min_m \Phi_0(m; s, t)$$

is easy to obtain. For this purpose we will distinguish three cases.

Case 1: $s = 1$. In this case we have $\eta_f^2 = \eta_T^2$ and all loss is due to f. Here the function

$$\Phi_0(m; 1, 0) = ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1$$

is minimized by $m_{1,0}(\cdot) = \mu(\cdot)$, hence the constant weights $w(\mathbf{x}) \equiv 1$ are optimal, while the minimax optimal design density $k(\mathbf{x})$ is proportional to $\|\mathbf{z}(\mathbf{x})\| \mu(\mathbf{x})$. The minimum value is $\Phi_0(\mu; 1, 0) = 1$.

Case 2: $t = 1$. In this case we have $\eta_f^2 = \eta_{\boldsymbol{\theta}}^2 = 0$ and all loss is due to possible heteroscedasticity. Here

$$\Phi_0(m; 0, 1) = \left[\int_{\mathcal{S}} \|\mathbf{z}(\mathbf{x})\| m(\mathbf{x})\mathrm{d}\mathbf{x}\right]^2$$

is minimized in a limiting sense by the Dirac function $m_{0,1}(\cdot) = \delta_{\mathbf{x}_{\min}}$, placing all mass at those points $\mathbf{x}_{\min}$ at which $\|\mathbf{z}(\mathbf{x})\|$ is minimized.

Case 3: $1-s-t=1$. In this case all loss is due to misspecification of the parameter $\boldsymbol{\theta}$, more precisely to the matrix $\boldsymbol{\Sigma}_\pi$. Here

$$\Phi_0(m;0,0)=\operatorname{tr}\left[\left(\mathbf{B}_m-\mathbf{I}_p\right)^2\right]$$

is again minimized, with minimum value 0, by $m_{0,0}(\cdot)=\mu(\cdot)$.

For general $s,t\in[0,1]$ we propose to minimize the function $\Phi_0(m;s,t)$ by representing the minimizer – whose existence can be established using the methods of Theorom 2 of Heo et al. (2001) – as the solution to an auxiliary minimax problem. For vectors $\boldsymbol{\alpha}\in\mathbb{R}^p$ of unit norm, and fixed s and t, define the function

$$\mathcal{L}(m;\boldsymbol{\alpha})=s\cdot\left(\boldsymbol{\alpha}^{\mathrm{T}}\left[\mathbf{K}_m-\mathbf{B}_m^2\right]\boldsymbol{\alpha}+1\right)+t\cdot\left[\int_{\mathcal{S}}\|\mathbf{z}(\mathbf{x})\|\,m(\mathbf{x})\mathrm{d}\mathbf{x}\right]^2+(1-s-t)\cdot\operatorname{tr}\left[\left(\mathbf{B}_m-\mathbf{I}_p\right)^2\right].$$

Then $\Phi_0(m;s,t)=\max_{\|\boldsymbol{\alpha}\|=1}\mathcal{L}(m;\boldsymbol{\alpha})$ and the design problem is equivalent to the minimax problem

$$\min_m\max_{\|\boldsymbol{\alpha}\|=1}\mathcal{L}(m;\boldsymbol{\alpha}). \tag{4.3}$$

Theorem 2. *If $s>0$ then the density m_* which solves problem (4.3) is of the form*

$$m_*(\mathbf{x})=\frac{\left\{\lambda+\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\Gamma}\mathbf{z}(\mathbf{x})-\beta\|\mathbf{z}(\mathbf{x})\|\right\}^+}{\left(\boldsymbol{\alpha}_*^{\mathrm{T}}\mathbf{z}(\mathbf{x})\right)^2}\mu(\mathbf{x}), \tag{4.4}$$

(here q^+ denotes the positive part of a function q) for a symmetric matrix $\boldsymbol{\Gamma}$, scalars β and λ and a vector $\boldsymbol{\alpha}_*$ of unit norm. These parameters are determined by

$$\boldsymbol{\Gamma}=\frac{1}{2}\left(\mathbf{B}_{m_*}\boldsymbol{\alpha}_*\boldsymbol{\alpha}_*^{\mathrm{T}}+\boldsymbol{\alpha}_*\boldsymbol{\alpha}_*^{\mathrm{T}}\mathbf{B}_{m_*}\right)-\frac{1-s-t}{s}\left(\mathbf{B}_{m_*}-\mathbf{I}_p\right), \tag{4.5a}$$

$$\beta=\frac{t}{s}\int_{\mathcal{S}}\|\mathbf{z}(\mathbf{x})\|\,m_*(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{4.5b}$$

$$1=\int_{\mathcal{S}}m_*(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{4.5c}$$

$$\boldsymbol{\alpha}_*^{\mathrm{T}}\left[\mathbf{K}_{m_*}-\mathbf{B}_{m_*}^2\right]\boldsymbol{\alpha}_*=ch_{\max}\left[\mathbf{K}_{m_*}-\mathbf{B}_{m_*}^2\right]. \tag{4.5d}$$

If $s=0$ the solution is somewhat different, since in the absence of any bias the minimax design need no longer be absolutely continuous. This case is however of little interest from a robustness point of view.

Remark 2. See (4.4). In some models there may be points $\mathbf{x}$ at which $\boldsymbol{\alpha}_*^{\mathrm{T}}\mathbf{z}(\mathbf{x})=0$. It is evident from the proof of Theorem 2 that at such points it is required either that the numerator of $m_*(\mathbf{x})$ vanish, in which case we set $m_*(\mathbf{x})=0$, or that the numerator be positive and $m_*(\mathbf{x})=\infty$. The latter eventuality does not occur in any of the examples at which we have looked.

Remark 3. We shall discuss two possibilities for implementing Theorem 2. The first, implemented in Section 5 below, is to obtain $m_*(\mathbf{x})$ from Theorem 2 by solving Eq. (4.5). The second, deferred to Section 6, is to substitute (4.4) into $\Phi_0(m;s,t)$ at (4.2), and minimize the resulting expression over the parameters, subject to the constraints (4.5c) and $\|\boldsymbol{\alpha}\|=1$.

In the next section we continue with Example 2.2, and utilize the special structure of Zernike polynomials to simplify the requirements of Theorem 2.

5. Minimax designs and weights — Zernike polynomials

In this section we obtain minimax designs and weights for Zernike polynomial regression. First we write the bivariate density $m(\mathbf{x})$ in the form $m(\mathbf{x})=m^{(1)}(\rho)\,m^{(2)}(\phi|\rho)$. We will show, in Theorem 3, that under condition

(5.11) given there (and in (ii) of Lemma 1) we may restrict attention to the case in which the conditional density of ϕ, given ρ, is of the form

$$m^{(2)}(\phi|\rho) = \mu^{(2)}(\phi) = (2\pi)^{-1}. \tag{5.1}$$

When (5.1) holds, only the univariate density $m_*^{(1)}(\rho)$ need be determined numerically.

Note that, by virtue of (2.15), the objective function to be minimized is

$$\Phi_0(m; s, t) = s \cdot \left(ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1\right) + t \cdot \left[\int_0^1 \sqrt{\Theta_d(\rho)} m^{(1)}(\rho)\, \mathrm{d}\rho\right]^2 + (1 - s - t) \cdot \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right],$$

in which only the first and third terms depend on $m^{(2)}(\phi|\rho)$. We are not yet asserting that (5.1) holds, but when it does we will write

$$m_0(\mathbf{x}) = m^{(1)}(\rho)\, m_0^{(2)}(\phi|\rho) = m^{(1)}(\rho)\, \mu^{(2)}(\phi) \tag{5.2}$$

as a product measure and obtain (upon applying Fubini's Theorem)

$$\mathbf{B}_{m_0} = \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x}) m^{(1)}(\rho)\, m^{(2)}(\phi|\rho)\, \mathrm{d}\mathbf{x} = \int_0^1 \mathbf{R}^2(\rho)\, m^{(1)}(\rho)\, \mathrm{d}\rho \stackrel{\mathrm{def}}{=} \mathbf{H}; \tag{5.3}$$

here we have used the first identity in (2.14). Note that the matrix $\mathbf{H}$ is diagonal, with diagonal elements

$$h_i = \int_0^1 R_{ii}^2(\rho)\, m^{(1)}(\rho)\, \mathrm{d}\rho = \int_0^1 w(\rho)\, R_{ii}^2(\rho)\, \mu^{(1)}(\rho)\, \mathrm{d}\rho. \tag{5.4}$$

It will prove useful to interpret these diagonal elements as the expectations of the weights $w(\rho) = m^{(1)}(\rho)/\mu^{(1)}(\rho)$ when ρ has the density $R_{ii}^2(\rho)\, \mu^{(1)}(\rho)$ [note that $\int_0^1 R_{ii}^2(\rho)\, \mu^{(1)}(\rho)\, \mathrm{d}\rho = 1$ by (2.14)].

Calculations very similar to those leading to (2.15) give the useful identity

$$\mathbf{z}^{\mathrm{T}}(\mathbf{x})\mathbf{H}\mathbf{z}(\mathbf{x}) = \mathrm{tr}\,\mathbf{R}(\rho)\,\mathbf{H}\mathbf{R}(\rho). \tag{5.5}$$

As well, the matrix $\mathbf{K}_{m_0} - \mathbf{B}_{m_0}^2$ in the representation (4.1) becomes

$$\begin{aligned}\mathbf{K}_{m_0} - \mathbf{B}_{m_0}^2 &= \int_{\mathcal{S}} \left[\frac{m^{(1)}(\rho)}{\mu^{(1)}(\rho)}\mathbf{I}_p - \mathbf{H}\right] \mathbf{z}(\mathbf{x})\, \mathbf{z}^{\mathrm{T}}(\mathbf{x}) \left[\frac{m^{(1)}(\rho)}{\mu^{(1)}(\rho)}\mathbf{I}_p - \mathbf{H}\right] \mu^{(1)}(\rho)\, \mu^{(2)}(\phi)\mathrm{d}\phi\mathrm{d}\rho \\ &= \int_0^1 \left[w(\rho)\,\mathbf{I}_p - \mathbf{H}\right] \mathbf{R}^2(\rho) \left[w(\rho)\,\mathbf{I}_p - \mathbf{H}\right] \mu^{(1)}(\rho)\, \mathrm{d}\rho \\ &\stackrel{\mathrm{def}}{=} \mathbf{D},\end{aligned}$$

which is a diagonal matrix with diagonal elements

$$d_i = \int_0^1 (w(\rho) - h_i)^2\, R_{ii}^2(\rho)\, \mu^{(1)}(\rho)\, \mathrm{d}\rho. \tag{5.6}$$

As at (5.4), d_i can be interpreted as the variance of the weights $w(\rho)$ when ρ has the density $R_{ii}^2(\rho)\, \mu^{(1)}(\rho)$.

The minimax solution for Zernike polynomial regression is motivated by the following result.

Lemma 1. *For the Zernike polynomial regression model, let $m^{(1)}(\rho)$ be a fixed but arbitrary marginal density of ρ. Then:*

1. *The quantity $\mathrm{tr}\,[(\mathbf{B}_m - \mathbf{I}_p)^2]$ is minimized by $m^{(2)}(\phi|\rho) = \mu^{(2)}(\phi)$.*
2. *Define $I^* = \{1, 5, 13, \ldots\}$ to be the set of indices, in the ordering of the elements of $\mathbf{z}(\mathbf{x})$, corresponding to terms R_k^m with $m = 0$ and hence $a(\phi) \equiv 1$. If $m^{(1)}(\rho)$ is such that the maximum diagonal element d_{i^*} of the matrix $\mathbf{D}$ satisfies $i^* \in I^*$, then $ch_{\max}\left[\mathbf{K}_m - \mathbf{B}_m^2\right]$ is also minimized by $m^{(2)}(\phi|\rho) = \mu^{(2)}(\phi)$.*

Let $\mathbf{h} = \mathbf{H1}$ be the p-vector with elements h_i given by (5.4) and define

$$\mathbf{r}(\rho) = \mathbf{R}^2(\rho)\,\mathbf{1} = \left(R_{11}^2(\rho), \ldots, R_{pp}^2(\rho)\right)^{\mathrm{T}},$$

so that $\mathbf{h} = \int_0^1 \mathbf{r}(\rho)\, m^{(1)}(\rho)\,\mathrm{d}\rho$. By virtue of Lemma 1, if $m_0(\mathbf{x})$ defined by (5.2) is such that the associated weights $w(\rho)$ minimize

$$\Phi_0(m_0; s, t) = s \cdot \left(\max_i d_i + 1\right) + t \cdot \left[\int_0^1 \sqrt{\Theta_d(\rho)}\, m^{(1)}(\rho)\,\mathrm{d}\rho\right]^2 + (1 - s - t) \cdot \|\mathbf{h} - \mathbf{1}\|^2, \tag{5.7}$$

with $\max_i d_i = d_{i^*}$ for some $i^* \in I^*$, then to prove that $m_0(\mathbf{x})$ is the desired minimax density one need only check that it is of the form specified by Theorem 2. In some cases this procedure can be accomplished by minimizing

$$s \cdot (d_{i^*} + 1) + t \cdot \left[\int_0^1 \sqrt{\Theta_d(\rho)}\, m^{(1)}(\rho)\,\mathrm{d}\rho\right]^2 + (1 - s - t) \cdot \|\mathbf{h} - \mathbf{1}\|^2 \tag{5.8}$$

for a *particular* $i^* \in I^*$. The next result gives the details.

Theorem 3. *For $s > 0$ and for each $i^* \in I^*$, define a weight function $w_{i^*}(\rho)$, and densities $m_{i^*}^{(1)}(\rho)$ and $m_{i^*}(\mathbf{x})$, by*

$$w_{i^*}(\rho) = \frac{\left\{r_{i^*}(\rho) + \lambda_{i^*} + \left[\mathbf{e}_{i^*} r_{i^*}(\rho) - \frac{1-s-t}{s}\mathbf{r}(\rho)\right]^{\mathrm{T}}\left(\mathbf{h}^{(i^*)} - \mathbf{1}\right) - \beta_{i^*}\sqrt{\Theta_d(\rho)}\right\}^+}{r_{i^*}(\rho)}, \tag{5.9a}$$

$$m_{i^*}^{(1)}(\rho) = w_{i^*}(\rho)\,\mu^{(1)}(\rho), \tag{5.9b}$$

$$m_{i^*}(\mathbf{x}) = m_{i^*}^{(1)}(\rho)\,\mu^{(2)}(\phi). \tag{5.9c}$$

Here $\mathbf{e}_{i^*}$ is the unit vector in $\mathbb{R}^p$ with a 1 in position i^* and zeros elsewhere and the p-dimensional vector $\mathbf{h}^{(i^*)}$ and scalars β_{i^*} and λ_{i^*} are to be determined from

$$\mathbf{h}^{(i^*)} = \int_0^1 \mathbf{r}(\rho)\, m_{i^*}^{(1)}(\rho)\,\mathrm{d}\rho, \tag{5.10a}$$

$$\beta_{i^*} = \frac{t}{s}\int_0^1 \sqrt{\Theta_d(\rho)}\, m_{i^*}^{(1)}(\rho)\,\mathrm{d}\rho, \tag{5.10b}$$

$$1 = \int_0^1 m_{i^*}^{(1)}(\rho)\,\mathrm{d}\rho. \tag{5.10c}$$

If

$$i^* = \arg\max_{1 \le i \le p} \int_0^1 \left(w_{i^*}(\rho) - h_{i^*}^{(i^*)}\right)^2 R_{ii}^2(\rho)\,\mu^{(1)}(\rho)\,\mathrm{d}\rho, \tag{5.11}$$

then the solution to the minimax problem (4.3) is given by $(m, \boldsymbol{\alpha}) = (m_*, \mathbf{e}_{i^*})$. The minimax design density is then given by Proposition 1, i.e.

$$k_*(\rho, \phi) = k_*^{(1)}(\rho)\,\mu^{(2)}(\phi),$$

where

$$k_*^{(1)}(\rho) = \frac{t\sqrt{\Theta_d(\rho)}\, w_{i^*}(\rho)\,\mu^{(1)}(\rho)}{s\beta_{i^*}},$$

and the design points may be obtained from

$$\int_0^{\rho_i} k_*^{(1)}(\rho)\,\mathrm{d}\rho = \frac{i}{n_1}, \qquad \phi_j = \frac{2\pi j}{n_2},$$

for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, with $n = n_1 n_2$.

Example 5.1 (Example 2.2 *Continued*). We present the results of applying Theorem 3 in three cases. The constants are determined by using the Matlab function `fsolve` to simultaneously solve Eq. (5.10). All integrals are evaluated by Simpson's Rule, over a grid of 101 equally spaced points in the interval [0, 1]. Our method is to compute $w_{i^*}(\cdot)$, for each $i^* \in I^*$. We then check that one such function satisfies (5.11); if so then it is the desired minimax weighting function $w_*(\cdot)$. For the values $d = 1$ (hence $i^* = 1$), $s = .7, t = .05$ we obtain output as illustrated in Fig. 1, with $\beta = .12, \lambda = .19$. In this and the other two cases we also plot the marginal, unbiased design density (3.3).

For $d = 2, s = .1, t = .4$ this procedure gives the design and weights displayed in Fig. 2, with $i^* = 1$, $\beta = 8.40$, $\lambda = 13.22$. For $d = 4, s = .2, t = .8$ the results are shown in Fig. 3 ($\beta = 11.40$, $\lambda = 35.35$, $i^* = 5$). In this case the design is only supported in the interval [0, .6], approximately; recall Remark 2 and note that $r_5(\rho) = 0$ only for $\rho = 1/\sqrt{2} > .6$. For some other values of d, s and t condition (5.11) is not satisfied; when this is the case the more involved minimization of (5.7) – with the maximization being over $i \in I^*$ – rather than the simpler (5.8), appears necessary.

6. Minimax designs and weights — general models

In this section we illustrate the second method of applying Theorem 2, as discussed in Remark 3. The method applies in general regression models, and is implemented here in the context of function approximation via orthogonal polynomials.

Consider approximating a function on $\mathcal{S} = [-1, 1]$ using Chebyshev polynomials. To satisfy (2.6) we take

$$\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}, \quad \text{and}$$
$$\mathbf{z}(x) = \left(T_0(x), \sqrt{2}T_1(x), \sqrt{2}T_2(x), \ldots, \sqrt{2}T_d(x)\right)^{\mathrm{T}},$$

where $T_j(x)$ is the jth Chebyshev polynomial of the first kind, given by $T_j(x) = \cos(j \arccos x)$. Some particular values are

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_2(x) = 2x^2 - 1, \qquad T_3(x) = 4x^3 - 3x.$$

The optimization described in Remark 3 is carried out using Matlab's `fminimax` routine. First define $\boldsymbol{\gamma} = \text{vecs}(\boldsymbol{\Gamma})$ to be the vector, of dimension $p(p+1)/2$, formed from the lower triangle of the matrix $\boldsymbol{\Gamma}$. The problem is then

$$\min_{\mathbf{v}} \max_{[F_1,\ldots,F_p]} \{F_i(\mathbf{v})\}$$

where $\mathbf{v} = \left(\boldsymbol{\alpha}^{\mathrm{T}}, \beta, \boldsymbol{\gamma}^{\mathrm{T}}, \lambda\right)^{\mathrm{T}}$ and, with $ch_{(i)}[\cdot]$ denoting the ith largest eigenvalue,

$$F_i(\mathbf{v}) = s \cdot \left(ch_{(i)}\left[\mathbf{K}_m - \mathbf{B}_m^2\right] + 1\right) + t \cdot \left[\int_{-1}^{1} \|\mathbf{z}(x)\| m(x)\, \mathrm{d}x\right]^2 + (1-s-t) \cdot \mathrm{tr}\left[\left(\mathbf{B}_m - \mathbf{I}_p\right)^2\right]$$

with

$$m(x) = \frac{\left\{\lambda + \mathbf{z}^{\mathrm{T}}(x)\boldsymbol{\Gamma}\mathbf{z}(x) - \beta\|\mathbf{z}(x)\|\right\}^+}{\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{z}(x)\right)^2}\mu(x).$$

This is solved subject to the constraints

$$\left(\int_{-1}^{1} m(x)\,\mathrm{d}x - 1, \|\boldsymbol{\alpha}\| - 1\right) = \mathbf{0}^{\mathrm{T}}.$$

Some representative designs and weighting functions are given in Fig. 4 ($d = 3, s = t = .3$) and Fig. 5 ($d = 5, s = .2, t = .4$). In this (and the other) examples there is considerable numerical instability; in particular the constants obtained here do not quite satisfy the equations of Theorem 2. However, the designs themselves change very little under changes of the constants; we attribute this to the response surface over which we are optimizing being quite flat.

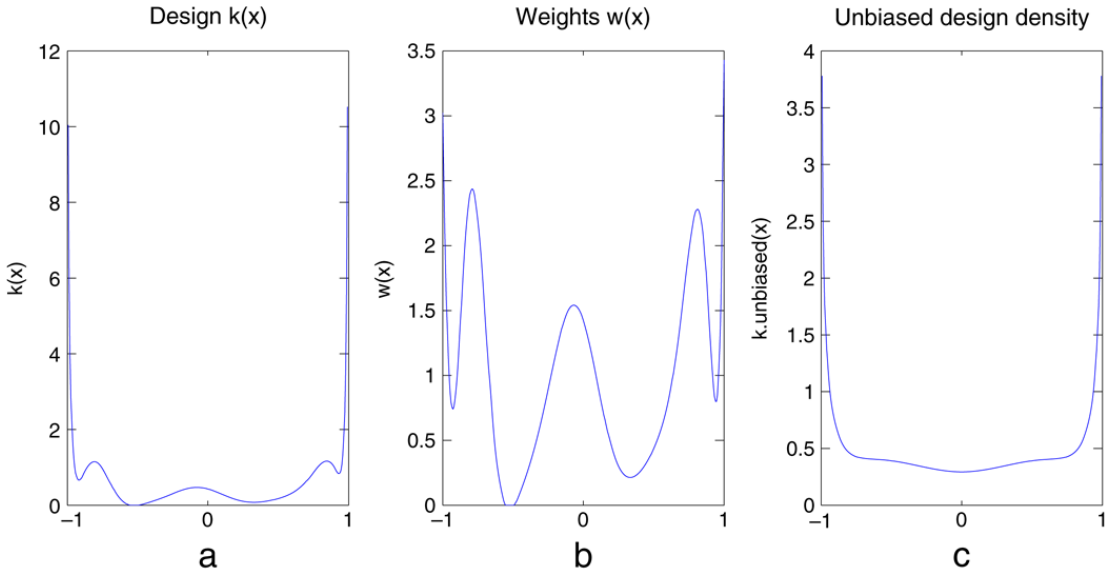

Fig. 4. (a) Design density, (b) weights and (c) unbiased design density for Chebyshev polynomial approximation; $d = 3$, $s = t = .3$.

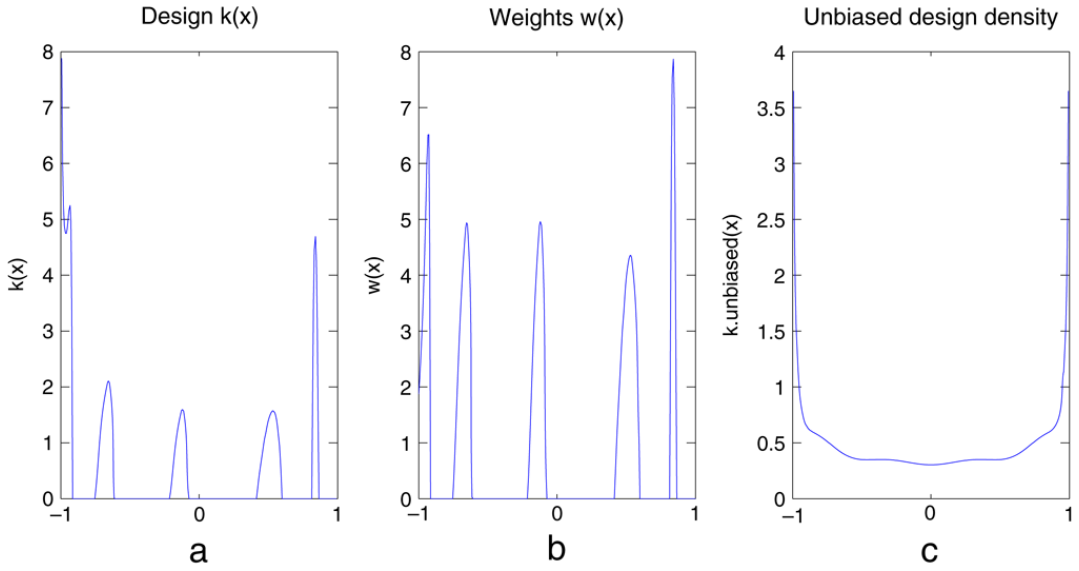

Fig. 5. (a) Design density, (b) weights and (c) unbiased design density for Chebyshev polynomial approximation; $d = 5$, $s = .2$, $t = .4$.

7. Summary

We have exhibited robust designs and weighting functions, appropriate for the "direct" series estimation technique often favoured by practitioners. These are solutions to two minimax problems, one of which is constrained by a requirement that the estimates be unbiased throughout a neighbourhood of contaminated response functions. In this case the results are particularly appealing — the optimal weights are constant, and the minimax optimal design density (3.3) is proportional to $\|\mathbf{z}(\mathbf{x})\|\,\mu(\mathbf{x})$ under the assumption of homoscedasticity, where $\mathbf{z}(\mathbf{x})$ denotes the vector of functions used in the series expansion. Given the numerical complexity of the minimax designs without this requirement of unbiasedness, and recognizing the dominant contribution, asymptotically, of bias over variance to mean squared error, it is our opinion that designs for unbiased estimation are to be preferred.

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 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 also grateful to Martina Stein, who typed parts of this paper with considerable technical expertise. The authors would also like to thank the reviewers for their constructive comments on an earlier version of this manuscript.

Appendix. Derivations

Proof of Theorem 1. From (2.4) and (2.9) we obtain

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \{\mathbf{B}_n\boldsymbol{\theta} + \mathbf{b}_n\}\right) = \sqrt{n}\mathbf{M},$$

where

$$\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}(\mathbf{x}_i)\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\varepsilon(\mathbf{x}_i).$$

We must show that $\sqrt{n}\mathbf{M}$ is $AN\left(\mathbf{0}, \sigma_\varepsilon^2\mathbf{C}_n\right)$. For this we invoke the Cramér–Wold device (Serfling, 1982, p. 18) and define, for an arbitrary vector $\mathbf{t}$,

$$X_i = \mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\varepsilon(\mathbf{x}_i),$$

with mean 0, variance

$$\sigma_i^2 = \sigma_\varepsilon^2\left(\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right)^2\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^2 g(\mathbf{x}_i)$$

and (with $\tau_i^3 = \tau^3(\mathbf{x}_i)$) third absolute moment

$$E\left[\left|X_i^3\right|\right] = \sigma_\varepsilon^3\tau_i^3\left|\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right|^3\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^3 g^{3/2}(\mathbf{x}_i).$$

By Liapounov's Central Limit Theorem (Serfling, 1982, p. 30), as long as

$$\frac{\sqrt{n}\sum_{i=1}^{n}E\left[\left|X_i^3\right|\right]}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\tau_i^3\left|\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right|^3\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^3 g^{3/2}(\mathbf{x}_i)}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right)^2\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^2 g(\mathbf{x}_i)\right)^{3/2}} = o(\sqrt{n}) \tag{A.1}$$

we have that

$$\frac{1}{n}\sum_{i=1}^{n}X_i \sim AN\left(0, \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right).$$

Equivalently,

$$\frac{\mathbf{t}^{\mathrm{T}}\mathbf{M}}{\sqrt{\mathbf{t}^{\mathrm{T}}\frac{\mathbf{C}_n}{n}\mathbf{t}}} \xrightarrow{d} N(0,1),$$

as required. To assess (A.1), we apply elementary upper and lower bounds on the numerator and denominator, yielding

$$\sup_{\|\mathbf{t}\|=1}\frac{\frac{1}{n}\sum_{i=1}^{n}\tau_i^3\left|\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right|^3\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^3 g^{3/2}(\mathbf{x}_i)}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{t}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)\right)^2\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^2 g(\mathbf{x}_i)\right)^{3/2}} \le \frac{\frac{1}{n}\sum_{i=1}^{n}\tau_i^3\|\mathbf{z}(\mathbf{x}_i)\|^3\left(\frac{m(\mathbf{x}_i)}{k(\mathbf{x}_i)}\right)^3 g^{3/2}(\mathbf{x}_i)}{(ch_{\min}\mathbf{C}_n)^{3/2}} < \infty$$

by assumptions (i), (ii) and our previous assumptions that $\mathbf{z}$, g and m/k be uniformly continuous, hence bounded on the bounded set $\mathcal{S}$; (A.1) follows. □

Proof of Proposition 1. For densities $k_0(\mathbf{x})$ and $k_1(\mathbf{x})$ and $t \in [0,1]$ define the convex combination $k_t(\mathbf{x}) = (1-t)\,k_0(\mathbf{x}) + tk_1(\mathbf{x})$. In order that $k_0(\cdot)$ minimize the function $\Psi_1(k;m)$ subject to the constraint that it integrate to 1 it is sufficient that the function

$$\phi(t;\lambda) = \Psi_1(k_t;m) + \lambda^2 \int_{\mathcal{S}} k_t(\mathbf{x})\mathrm{d}\mathbf{x}$$

be minimal at the point $t=0$ for any $k_1(\cdot)$, and that the density $k_0(\cdot)$ satisfy the constraint. For this, since $\phi(t;\lambda)$ is a convex function of t, the first order condition is necessary and sufficient, which means that the inequality

$$\phi'(0;\lambda) = \int_{\mathcal{S}} [k_1(\mathbf{x}) - k_0(\mathbf{x})] \left[\lambda^2 - \|\mathbf{z}(\mathbf{x})\|^2 \frac{m^2(\mathbf{x})}{k_0^2(\mathbf{x})} g_*(\mathbf{x})\right] \mathrm{d}\mathbf{x} \geq 0$$

holds for all $k_1(\cdot)$. This condition is satisfied if

$$k_0(\mathbf{x}) = \frac{\|\mathbf{z}(\mathbf{x})\|\, m(\mathbf{x})\sqrt{g_*(\mathbf{x})}}{\lambda},$$

and it remains only to determine λ to satisfy the constraint. □

Proof of Theorem 2. We will show that the pair $(\boldsymbol{\alpha}_*, m_*)$ in the statement of the theorem furnishes a *saddlepoint* solution to (4.3); viz. its members satisfy

$$\max_{\|\boldsymbol{\alpha}\|=1} \mathcal{L}(m_*;\boldsymbol{\alpha}) = \mathcal{L}(m_*;\boldsymbol{\alpha}_*) = \min_m \mathcal{L}(m;\boldsymbol{\alpha}_*). \tag{A.2}$$

Note that the first equality in (A.2) is equivalent to (4.5d). To handle the second equality we introduce the convex combination

$$m_t = (1-t)\,m_* + tm_1$$

of the minimizing m_* and an arbitrary competing density m_1. We are to determine a non-negative function $m_*(\cdot)$ for which the normalizing condition $\int_{\mathcal{S}} m_*(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ is satisfied and such that $\mathcal{L}(m_t;\boldsymbol{\alpha}_*)$ is minimized at $t=0$ for all m_1. It follows from (4.1) and Lemma 4.4 of Huber (1981) that the objective function is a convex function of the parameter t. Thus, with

$$\Psi(t) \stackrel{\text{def}}{=} \mathcal{L}(m_t;\boldsymbol{\alpha}_*) - 2\lambda \int_{\mathcal{S}} m_t(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

the first order condition "$\Psi'(0) \geq 0$ for all densities m_1" is necessary and sufficient for the density m_* to be the optimum. A straightforward calculation shows that

$$\Psi'(0) = 2\int_{\mathcal{S}} (m_1 - m_*)(\mathbf{x}) \cdot \{p(\mathbf{x})m_*(\mathbf{x}) - q(\mathbf{x})\}\,\mathrm{d}\mathbf{x},$$

where the functions p and q are defined by

$$p(\mathbf{x}) = \frac{s\left(\boldsymbol{\alpha}_*^{\mathrm{T}}\mathbf{z}(\mathbf{x})\right)^2}{\mu(\mathbf{x})} \geq 0,$$

$$q(\mathbf{x}) = \lambda + \mathbf{z}^{\mathrm{T}}(\mathbf{x})\left[s\mathbf{B}_{m_*}\boldsymbol{\alpha}_*\boldsymbol{\alpha}_*^{\mathrm{T}} - (1-s-t)\left(\mathbf{B}_{m_*} - \mathbf{I}_p\right)\right]\mathbf{z}(\mathbf{x}) - t\,\|\mathbf{z}(\mathbf{x})\| \int_{\mathcal{S}} \|\mathbf{z}(\mathbf{y})\|\, m_*(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

In other words: the condition $\Psi'(0) \geq 0$ is satisfied for all m_1 if and only if

$$m_*(\mathbf{x}) = \frac{q(\mathbf{x})^+}{p(\mathbf{x})} = \frac{\left\{\lambda + \mathbf{z}^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\Gamma}\mathbf{z}(\mathbf{x}) - \beta\,\|\mathbf{z}(\mathbf{x})\|\right\}^+}{\left(\boldsymbol{\alpha}_*^{\mathrm{T}}\mathbf{z}(\mathbf{x})\right)^2}\mu(\mathbf{x}),$$

where the parameters $\boldsymbol{\Gamma}$, β and λ are determined by (4.5a)–(4.5c). □

Proof of Lemma 1. Define $m_t(\mathbf{x}) = (1-t)\,m_0(\mathbf{x}) + tm_1(\mathbf{x})$, where $m_0(\mathbf{x})$ is as at (5.2) and $m_1(\mathbf{x}) = m^{(1)}(\rho)\,m_1^{(2)}(\phi|\rho)$, for an arbitrary conditional density $m_1^{(2)}(\phi|\rho)$.

To establish (i), we show that $\beta(t) = \mathrm{tr}[(\mathbf{B}_{m_t} - \mathbf{I}_p)^2]$ is minimized at $t = 0$, for any $m_1^{(2)}(\phi|\rho)$. Since $\beta(t)$ is convex, a necessary and sufficient condition for this is "$\beta'(0) \geq 0$ for all $m_1(\cdot)$". We will show that, in fact, $\beta'(0) \equiv 0$. For this we calculate, using (5.3), that

$$\beta'(0) = 2\mathrm{tr}\left[\left(\mathbf{H} - \mathbf{I}_p\right)\left(\mathbf{B}_{m_1} - \mathbf{H}\right)\right].$$

Using (2.15) and (5.5), we obtain

$$\begin{aligned}
\mathrm{tr}\left(\left(\mathbf{H} - \mathbf{I}_p\right)\mathbf{B}_{m_1}\right) &= \int_{\mathcal{S}} \mathbf{z}^{\mathrm{T}}(\mathbf{x})\left(\mathbf{H} - \mathbf{I}_p\right)\mathbf{z}(\mathbf{x}) m_1(\mathbf{x})\mathrm{d}\mathbf{x} \\
&= \int_{\mathcal{S}} \mathrm{tr}\left[\mathbf{R}(\rho)\left(\mathbf{H} - \mathbf{I}_p\right)\mathbf{R}(\rho)\right] m_1(\mathbf{x})\mathrm{d}\mathbf{x} \\
&= \mathrm{tr}\left[\left(\mathbf{H} - \mathbf{I}_p\right)\int_{\mathcal{S}} \mathbf{R}^2(\rho)\, m^{(1)}(\rho)\,\mathrm{d}\rho\right] \\
&= \mathrm{tr}\left[\left(\mathbf{H} - \mathbf{I}_p\right)\mathbf{H}\right],
\end{aligned}$$

so that $\beta'(0) = 0$, as required.

To establish (ii) it is sufficient to show that, for any $m_1(\mathbf{x})$ as above, if $i^* \in I^*$ then

$$ch_{\max}\left[\mathbf{K}_{m_1} - \mathbf{B}_{m_1}^2\right] \geq \left(\mathbf{K}_{m_1} - \mathbf{B}_{m_1}^2\right)_{i^*i^*} \geq d_{i^*} = ch_{\max}\left[\mathbf{K}_{m_0} - \mathbf{B}_{m_0}^2\right]. \tag{A.3}$$

Only the second inequality in (A.3) requires a proof — the first is true universally, and the equality is the definition of i^*.

A direct calculation, using (4.1), gives

$$\left(\mathbf{K}_{m_1} - \mathbf{B}_{m_1}^2\right)_{i^*i^*} = \int_0^1 \left[\int_0^{2\pi} l_{m_1}^2(\phi;\rho)\,\mu^{(2)}(\phi)\mathrm{d}\phi\right] R_{i^*i^*}^2(\rho)\,\mu^{(1)}(\rho)\,\mathrm{d}\rho,$$

where

$$l_{m_1}(\phi;\rho) = \frac{m^{(1)}(\rho)\, m_1^{(2)}(\phi|\rho)}{\mu^{(1)}(\rho)\,\mu^{(2)}(\phi)} a_{i^*}(\phi) - \frac{\mathbf{b}_{m_1,i^*}^{\mathrm{T}}\mathbf{R}(\rho)\,\mathbf{a}(\phi)}{R_{i^*i^*}(\rho)}$$

and $\mathbf{b}_{m_1,i^*}^{\mathrm{T}} = \int_{\mathcal{S}} \mathbf{z}_{i^*}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x})\, m_1(\mathbf{x})\mathrm{d}\mathbf{x}$ is the ith row of the matrix $\mathbf{B}_{m_1}$. Note that

$$l_{m_0}(\phi;\rho) = a_{i^*}(\phi)\left(\frac{m^{(1)}(\rho)}{\mu^{(1)}(\rho)} - h_{i^*i^*}\right), \tag{A.4}$$

and define

$$\gamma(t) = \int_0^1 \left[\int_0^{2\pi} l_{m_t}^2(\phi;\rho)\,\mu^{(2)}(\phi)\mathrm{d}\phi\right] R_{i^*i^*}^2(\rho)\,\mu^{(1)}(\rho)\,\mathrm{d}\rho,$$

with m_t as above. Using (A.4) and then (5.6) to evaluate $\gamma(0)$, we see that the required inequality in (A.3) is expressible as "$\gamma(1) \geq \gamma(0)$". By convexity (the integrand in $\gamma(t)$ is the square of a linear function of t) the condition "$\gamma'(0) \geq 0$ for every $m_1^{(2)}(\phi|\rho)$" is again necessary and sufficient for this. Since

$$\gamma'(0) = 2\int_0^1 \left[\int_0^{2\pi} l_{m_0}(\phi;\rho)\left(l_{m_1}(\phi;\rho) - l_{m_0}(\phi;\rho)\right)\mu^{(2)}(\phi)\,\mathrm{d}\phi\right] R_{i^*i^*}^2(\rho)\,\mu^{(1)}(\rho)\,\mathrm{d}\rho,$$

we are equivalently to show that

$$L_{i^*}(m_1) \stackrel{\mathrm{def}}{=} \int_0^1 \left[\int_0^{2\pi} l_{m_0}(\phi;\rho)\, l_{m_1}(\phi;\rho)\,\mu^{(2)}(\phi)\mathrm{d}\phi\right] R_{i^*i^*}^2(\rho)\,\mu^{(1)}(\rho)\,\mathrm{d}\rho \tag{A.5}$$

does not exceed $L_{i^*}(m_0) = d_{i^*}$. We will in fact show that $L_{i^*}(m_1)$ is *constant*, if $i^* \in I^*$.

In terms of

$$A_i(\rho; m) = \int_0^{2\pi} a_i^2(\phi) m^{(2)}(\phi|\rho)\, \mathrm{d}\phi$$

we calculate that (A.5) is

$$L_{i^*}(m_1) = \mathrm{COV}\left[\frac{m^{(1)}(\rho)}{\mu^{(1)}(\rho)}, \frac{m^{(1)}(\rho)}{\mu^{(1)}(\rho)} A_{i^*}(\rho; m_1)\right],$$

where the covariance is calculated with respect to the density $R^2_{i^* i^*}(\rho)\, \mu^{(1)}(\rho)$.

To now we have not used the assumption that $i^* \in I^*$. Under this condition $a^2_{i^*}(\phi) \equiv 1$ and so $A_{i^*}(\rho; m_1) = 1$ for all m_1; thus $L_{i^*}(m_1)$ does not depend upon m_1 and is constantly equal to $\mathrm{VAR}\left[m^{(1)}(\rho)/\mu^{(1)}(\rho)\right] = d_{i^*} = L_{i^*}(m_0)$. □

Proof of Theorem 3. Substituting (5.9b) and (5.9c) and $\boldsymbol{\alpha} = \mathbf{e}_{i^*}$ into the expressions of Theorem 2, and using (A.5), gives

$$w_{i^*}(\rho) = \frac{\left\{\lambda + \mathbf{z}^{\mathrm{T}}(\mathbf{x})\, \boldsymbol{\Gamma} \mathbf{z}(\mathbf{x}) - \beta \|\mathbf{z}(\mathbf{x})\|\right\}^+}{\left(\boldsymbol{\alpha}_*^{\mathrm{T}} \mathbf{z}(\mathbf{x})\right)^2} = \frac{\left\{\lambda_{i^*} + \mathbf{z}^{\mathrm{T}}(\mathbf{x})\, \boldsymbol{\Gamma} \mathbf{z}(\mathbf{x}) - \beta_{i^*} \sqrt{\Theta_d(\rho)}\right\}^+}{r_{i^*}(\rho)}. \tag{A.6}$$

Eqs. (4.5b) and (4.5c) of Theorem 2 become (5.10b) and (5.10c) respectively, and (4.5a) yields

$$\mathbf{z}^{\mathrm{T}}(\mathbf{x}) \boldsymbol{\Gamma} \mathbf{z}(\mathbf{x}) = r_{i^*}(\rho) + \left[\mathbf{e}_{i^*} r_{i^*}(\rho) - \frac{1-s-t}{s} \mathbf{r}(\rho)\right]^{\mathrm{T}} \left(\mathbf{h}^{(i^*)} - \mathbf{1}\right)$$

with $\mathbf{h}^{(i^*)}$ given by (5.10a). Substituting this last identity into (A.6) yields (5.9a). Thus $m_{i^*}(\mathbf{x})$ satisfies the conditions of Theorem 2 and so is the desired minimax density. □

References

Brechbühler, Ch., Gerig, G., Kübler, O., 1995. Parametrization of closed surfaces for 3-D shape description. Computer Vision and Image Understanding 61, 154–170.

Dette, H., Melas, V., Pepelyshev, A., 2005. Optimal designs for three-dimensional shape analysis with spherical harmonic descriptors. Annals of Statistics 33, 2758–2788.

Dette, H., Melas, V., Pepelyshev, A., 2007. Optimal designs for statistical analysis with Zernike polynomials. Statistics 41 (6), 453–470.

Dette, H., Wiens, D., 2007. Robust designs for 3d shape analysis with spherical harmonic descriptors. Statistica Sinica (in press).

Ding, W., Nesumi, H., Takano, Y., Ukai, Y., 2000. Quantitative evaluation of the three-dimensional fruit shape and size of citrus based on spherical harmonic descriptors. Euphytica 114, 103–116.

Efromovich, S., 1999. Nonparametric Curve Estimation. Springer, New York.

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A., 1998. Wavelets, Approximation and Statistical Applications. In: Lecture Notes in Statistics. Springer, NY.

Heo, G., Schmuland, B., Wiens, D.P., 2001. Restricted minimax robust designs for misspecified regression models. The Canadian Journal of Statistics 29, 117–128.

Herzberg, A.M., Traves, W.N., 1994. An optimal experimental design for the Haar regression model. Canadian Journal of Statistics 22, 357–364.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

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

Kim, W.J., Kim, J.S., 1999. Robust rotation angle estimator. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 768–773.

Lau, T.S., Studden, W.J., 1985. Optimal designs for trigonometric and polynomial regression. Annals of Statistics 13, 383–394.

Liao, S., Pawlak, M., 1996. On image analysis by moments. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 254–266.

Oyet, A.J., Wiens, D.P., 2000. Robust designs for wavelet approximations of regression models. Journal of Nonparametric Statistics 12, 837–859.

Pawlak, M., Liao, S., 2002. On the recovery of a function on a circular domain. IEEE Transactions on Information Theory 10, 2736–2753.

Sacks, J., Ylvisaker, D., 1970. Statistical design and integral approximation. In: Pyke, R. (Ed.), Proc. 12th Bienn. Semin. Can. Math. Congr. Can. Math. Soc., Montreal, pp. 115–136.

Serfling, R.J., 1982. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Szëgo, G., 1975. Orthogonal Polynomials. In: American Mathematical Society Colloquium Publications, vol. 23. American Mathematical Society, Providence, RI.

Wiens, D.P., 1992. Minimax designs for approximately linear regression. Journal of Statistical Planning and Inference 31, 353–371.

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

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