ROBUST DESIGNS FOR WAVELET APPROXIMATIONS OF REGRESSION MODELS

ALWELL J. OYET[a] and DOUGLAS P. WIENS[b,*]

[a] *Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1C 5S7;*
[b] *Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

We consider the construction of designs for the estimation of a regression function, when it is anticipated that this function is to be approximated by the dominant terms in its wavelet expansion. We consider both the Haar wavelet basis and the multiwavelet system. The experimenter estimates the coefficients of those wavelets included in the approximation, hoping that the omitted terms will be inconsequential. This introduces bias into the least squares estimates, which we propose handling at the design stage by one of two methods: (i) implementing a *minimax robust* design, which enjoys the property of minimizing the maximum value of an *mse*-based loss function, with the maximum being taken as the remainder in the wavelet expansion varies over an L^2-neighbourhood; (ii) implementing a *minimum variance unbiased (mvu)* design which, when employed with weighted least squares and weights derived here, minimizes the variance subject to a side condition of unbiasedness. For the Haar wavelet system we show that the uniform design is both minimax robust and *mvu*. For multiwavelet approximations we give examples of both minimax robust and *mvu* designs. Two examples from the nonparametric regression literature are discussed, and designs are presented for each type of wavelet approximation.

Keywords: A-optimality; D-optimality; G-optimality; Haar basis; minimax robust; minimum variance unbiased; multiwavelet system; nonparametric regression; Q-optimality; regressogram; wavelets; weighted least squares

AMS 1991 Subject Classifications: Primary: 62K05, 62G35; Secondary: 62G07

*Corresponding author. Tel.: (780) 492-4406, Fax: (780) 492-4826, e-mail: doug.wiens@ualberta.ca

1. INTRODUCTION

The development of wavelet theory has in recent years spawned applications in signal processing, in fast algorithms for integral transforms in numerical analysis and in function representation methods (Daubechies, 1992; Strang, 1989; Alpert, 1992). This last application has stimulated interest in wavelet approximations of regression response functions for the analysis of experimental data – see Antoniadis, Gregoire and McKeague (1994) and Benedetto and Frazier (1994), and in the construction of designs to facilitate such approximations – see Herzberg and Traves (1994).

Traditionally, design theory has focused on the attainment of some form of a minimum variance property, assuming the fitted model to be exactly correct. This in particular was the approach of Herzberg and Traves (1994). Beginning with the seminal work of Box and Draper (1959), various authors – Huber (1975); Pesotchinsky (1982); Sacks and Ylvisaker (1984); Wiens (1992, 1998, 1999); Wiens and Zhou (1997) among others – have sought designs which are robust against various forms of model misspecification. Wavelet approximations suggest such an approach since, although one can in principle approximate any (sufficiently well-behaved) function arbitrarily closely through a wavelet expansion, in practice one would be willing to fit and estimate only a relatively small number of the components of such an expansion. The omitted terms then constitute a form of model misspecification, resulting in biased estimates of the included coefficients and of the regression response.

We start by considering a nonparametric regression model $Y(x) = \eta(x) + \varepsilon$ with additive, homoscedastic, uncorrelated errors. The experimenter is to observe $Y(x)$ at n, not necessarily distinct, values of the design variable x, chosen from the design space $S = [0, 1]$. The response $\eta(x) = E[Y|x]$ is assumed to be square integrable on S. This integrability allows one to approximate $\eta(x)$ by finitely many terms of its wavelet series; the theory relevant to this is outlined in Section 2 of this article. We consider in particular the Haar wavelets and the multiwavelets of Alpert (1992). In Section 3 we quantify the bias resulting from the least squares estimation of the coefficients in the approximating series. This bias depends on the magnitude $f(x)$ of the omitted terms and on the design, and forms a major component of the

Mean Squared Error (MSE) matrix of the coefficient estimates. We construct *minimax robust* designs, which minimize the maximum value of a scalar-valued function of the MSE matrix, with the maximum being evaluated as $f(\cdot)$ varies over an L^2-neighbourhood of the zero function. For the Haar wavelet approximation we establish a very strong and general robustness property of the (continuous) uniform design. For the multiwavelet system we find that the problem is much more difficult. We give the details for one particular multiwavelet approximation.

The difficulties associated with minimax designs for multiwavelet approximations motivate our consideration of *minimum variance unbiased (mvu)* designs in Section 4. When used with weighted least squares and the optimal weights derived here, these designs minimize functions of the covariance matrix of the coefficient estimates, subject to a side condition of unbiasedness. We are able to obtain *mvu* designs for multiwavelet approximations of any order.

In Section 5 we consider two data sets from the literature on non-parametric regression, and use them to illustrate the construction of the designs of Sections 3 and 4. Derivations are in the Appendix.

2. PRELIMINARIES

A wavelet system is a collection of dilated and translated versions of a *scaling function* $\phi(x)$ and a *primary wavelet* $\psi(x)$ defined for integers j and k by $\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k)$ and $\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$. The construction of the functions $\phi(x)$ and $\psi(x)$ is based on the concept of *multiresolution analysis* of the space of square integrable functions developed by Meyer (1986) and Mallat (1989) – see Chui (1992); Antoniadis *et al.* (1994) and Härdle *et al.* (1998) for details.

The *Haar* wavelet basis is the simplest example of a wavelet basis for the space $L^2(\mathcal{S})$ of square integrable functions on $\mathcal{S} = [0,1]$. The scaling function is $\phi(x) = I_{[0,1]}(x)$ and the primary wavelet is $\psi(x) = \phi(2x) - \phi(2x - 1) = I_{[0,1/2)}(x) - I_{[1/2,1)}(x)$, yielding

$$\psi_{j,k}(x) = 2^{-(j/2)}\{I_{[2^j k,\, 2^j(k+(1/2)))}(x) - I_{[2^j(k+(1/2)),\, 2^j(k+1))}(x)\},$$
$$j \leq 0, \ k \geq 0.$$

The *multiwavelet* system constructed by Alpert (1992) will also be used in our study. The multiwavelet basis differs from other wavelet bases in that instead of a single scaling function $\phi(x)$ there are several such functions $\phi_0, \ldots, \phi_{N-1}$. Each scaling function is a dilated, translated and normalized Legendre polynomial on $[0, 1)$ i.e., $\phi_i(x) = \sqrt{2i+1}P_i(2x-1)I_{[0,1)}(x)$ $(i = 1, \ldots, N)$, where $P_i(x)$ is the ith degree Legendre polynomial. In particular,

$$\phi_0(x) = I_{[0,1)}(x), \quad \phi_1(x) = 2\sqrt{3}(x - 1/2) \cdot I_{[0,1)}(x),$$
$$\phi_2(x) = 6\sqrt{5}((x - 1/2)^2 - 1/12) \cdot I_{[0,1)}(x).$$

The multiwavelets coincide with the Haar wavelet basis if $N = 1$. In this case we shall denote the primary wavelet by $_1w_0(x) = \psi(x)$. For $N = 2$ the primary wavelets are

$$_2w_0(x) = \sqrt{3}(4|x - 1/2| - 1) \cdot I_{[0,1)}(x),$$
$$_2w_1(x) = 2(1 - 3|x - 1/2|) \cdot (I_{[0,1/2)}(x) - I_{[1/2,1)}(x)).$$

For $N = 3$ they are

$$_3w_0(x) = {}_2w_1(x), \quad _3w_1(x) = -\sqrt{3}(30|x - 1/2|^2 - 16|x - 1/2| + 3/2)$$
$$\cdot I_{[0,1)}(x),$$
$$_3w_2(x) = -\sqrt{5}(24|x - 1/2|^2 - 12|x - 1/2| + 1)$$
$$\cdot (I_{[0,1/2)}(x) - I_{[1/2,1)}(x)).$$

See Figure 1.

Let $\eta(\cdot) \in L^2(\mathcal{S})$. The multiresolution analysis of $L^2(\mathcal{S})$ leads to the wavelet representations of $\eta(\cdot)$ given by Alpert (1992) and Walter (1995). For the Haar wavelet system we have

$$\eta(x) = d_0\phi_0(x) + \sum_{j,k \geq 0} c_{jk}\psi_{-j,k}(x)$$

where $d_0 = \int_0^1 \eta(x)\phi_0(x)dx$ and $c_{jk} = \int_0^1 \eta(x)\psi_{-j,k}(x)dx$. Similarly, for the multiwavelet system there exists a representation as a series in $\{\phi_l(x), {}_Nw_l^{-j,k}(x) | j, k \geq 0, l = 0, \ldots, N-1\}$, where $_Nw_l^{-j,k}(x) = 2^{j/2}{}_Nw_l(2^jx - k)$.
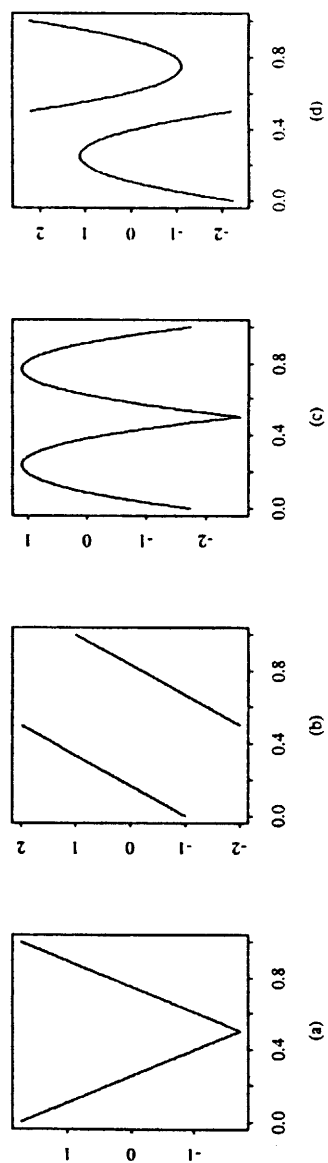
FIGURE 1 Primary multiwavelets: (a) $_2 w_0(x)$, (b) $_2 w_1(x) = _3 w_0(x)$, (c) $_3 w_1(x)$, (d) $_3 w_2(x)$.

We shall denote by $\mathbf{q}_{N,m}(x)$ the $N \cdot 2^{m+1} \times 1$ vector consisting of the wavelets $\{\phi_l(x), {}_Nw_l^{-j,k}(x) | j = 0, \ldots, m, k = 0, \ldots, 2^j - 1, l = 0, \ldots, N - 1\}$ in some order. The elements of $\mathbf{q}_{N,\infty}$ form a basis for $L^2(\mathcal{S})$. For the ranges of j and k in $\mathbf{q}_{N,m}(x)$ we find it convenient to write

$$
{}_Nw_l^{-j,k}(x) = 2^{(j/2)} {}_Nw_l(\{2^j x\}) I([2^j x] = k), \tag{1}
$$

where $\{x\} = x - [x]$ denotes the *fractional* part of x.

THEOREM 2.1   *The Euclidean norm of $\mathbf{q}_{N,m}(x)$ is given by*

$$
\|\mathbf{q}_{N,m}(x)\|^2 = \frac{2^{m-1}}{\sqrt{1 - \frac{1}{4N^2}}}(\phi_{N-1}(\{2^{m+1}x\})\phi_N'(\{2^{m+1}x\})
$$
$$
- \phi_{N-1}'(\{2^{m+1}x\})\phi_N(\{2^{m+1}x\})).
$$

In particular, $\|\mathbf{q}_{1,m}(x)\|^2 = 2^{m+1}$. The asymptotic value satisfies

$$
\lim_{N\to\infty} \frac{\|\mathbf{q}_{N,m}(x)\|^2}{N} = \frac{2^{m+1}}{\pi}\{2^{m+1}x\}^{-1/2}(1 - \{2^{m+1}x\})^{-1/2}. \tag{2}
$$

3. MINIMAX ROBUST DESIGNS

As in Daubechies (1993) we approximate the regression response by finitely many dominant terms of its wavelet representation, with remainder $f(x)$, viz. $E[Y|x] = \mathbf{q}_{N,m}^T(x)\theta_0 + f(x)$. Then $f(\cdot) \in L^2(\mathcal{S})$ and, with $\mathbf{q} = \mathbf{q}_{N,m}$, the model is

$$
Y(x) = \mathbf{q}^T(x)\theta_0 + f(x) + \varepsilon. \tag{3}
$$

We make frequent use of the relationships

$$
\text{(i)} \int_0^1 \mathbf{q}(x)f(x)dx = 0, \quad \text{(ii)} \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)dx = \mathbf{I}. \tag{4}
$$

We suppose that the experimenter is to take n observations (x_i, y_i) from (3) and then estimate θ_0 by ordinary least squares. Then $\hat{Y}(x) = \mathbf{q}^T(x)\hat{\theta}_0$ is a biased estimate of $E[Y|x]$. In order that errors due to bias

not swamp those due to random variation we assume that

$$\int_0^1 f^2(x)dx \le \tau^2 \tag{5}$$

for some, presumably small, constant τ. The integral above is the sum of squares of the coefficients of those wavelets *not* being fitted in the model. Our results depend on τ and on the error variance σ^2 only through the quantity $\nu = \sigma^2/n\tau^2$. This quantity may be interpreted as representing the relative importance to the experimenter of variance versus bias: $\nu = 0$ corresponding to a 'pure bias' problem, $\nu = \infty$ to a 'pure variance' problem.

Denote by ξ the design measure, *i.e.*, the empirical distribution function of $\{x_i\}_{i=1}^n$. Define

$$\mathbf{B} = \mathbf{B}(\xi) = \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)d\xi(x), \quad \mathbf{b} = \mathbf{b}(f,\xi) = \int_0^1 \mathbf{q}(x)f(x)d\xi(x).$$

In terms of the model matrix $\mathbf{Q}$, with rows $\mathbf{q}^T(x_i)$, this is $\mathbf{B} = n^{-1}\mathbf{Q}^T\mathbf{Q}$. The least squares estimate is $\hat{\boldsymbol{\theta}}_0 = \mathbf{B}^{-1}\int_0^1 \mathbf{q}(x)y(x)d\xi(x)$, with bias $\mathbf{B}^{-1}\mathbf{b}$ and covariance matrix $(\sigma^2/n)\mathbf{B}^{-1}$. The Mean Squared Error matrix $\mathbf{M}(f,\xi)$ of $\hat{\boldsymbol{\theta}}_0$ is

$$\mathbf{M}(f,\xi) = \mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1} + \frac{\sigma^2}{n}\mathbf{B}^{-1}.$$

Antoniadis *et al.* (1994) have obtained bounds on the rate of convergence of the *mse* of $\hat{Y}$, as the sample size and number of fitted wavelets increase.

We adopt a *minimax* approach to the design problem, and seek a design measure ξ_* such that $\min_\xi \max_{f \in \mathcal{F}} \mathcal{L}(f,\xi) = \max_{f \in \mathcal{F}} \mathcal{L}(f,\xi_*)$, for some scalar-valued loss function $\mathcal{L}(f,\xi)$ depending of f and ξ through $\mathbf{M}(f,\xi)$. Here $\mathcal{F}$ is the class of functions f satisfying (4(i)) and (5). The loss functions we shall consider are

1. Integrated mean squared error: $\mathcal{L}_Q(f,\xi) = \int_0^1 E[(\hat{Y}(x) - E[Y|x])^2] dx = \mathrm{tr}\mathbf{M}(f,\xi) + \int_0^1 f^2(x)dx = \mathbf{b}^T\mathbf{B}^{-2}\mathbf{b} + (\sigma^2/n)\mathrm{tr}\mathbf{B}^{-1} + \int_0^1 f^2(x)dx$.
2. Trace: $\mathcal{L}_A(f,\xi) = \mathrm{tr}\mathbf{M}(f,\xi) = \mathbf{b}^T\mathbf{B}^{-2}\mathbf{b} + (\sigma^2/n)\mathrm{tr}\mathbf{B}^{-1}$.
3. Determinant: $\mathcal{L}_D(f,\xi) = |\mathbf{M}(f,\xi)| = (\sigma^2/n)^{2^{m+1}}\{(1+(n/\sigma^2))\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}/|\mathbf{B}|\}$.

We call a design *Q-variance optimal* if it minimizes $\mathcal{L}_Q(\mathbf{0}_{\mathcal{F}}, \xi)$, where $\mathbf{0}_{\mathcal{F}}$ is the zero function in $\mathcal{F}$; this coincides with the classical optimality criterion of variance minimization, assuming the fitted model to be exactly correct. We call a design *Q-minimax robust* if it minimizes $\max_{f \in \mathcal{F}} \mathcal{L}_Q(f, \xi)$. Notions of *A*- and *D*-variance optimality and minimax robustness are defined in an entirely analogous manner.

We adopt the viewpoint of *approximate* design theory, allowing as a design *any* distribution function on $\mathcal{S}$. We shall assume that $\xi(\{1\}) = 0$ for every design ξ. This is because all of the wavelet approximations considered here vanish off of $[0, 1)$. Lemma 1 of Wiens (1992) then applies, and states that a necessary condition for $\sup_{\mathcal{F}} \mathcal{L}(f, \xi)$ to be finite is the absolute continuity of ξ. We implement such a design by placing the design points at the quantiles – see the examples of Section 5.

Denote the density $\xi'(x)$ by $m(x)$. The maximum of $\mathcal{L}(f, \xi)$ may then be obtained in a manner similar to that in Theorem 1 of Wiens (1992); see Oyet (1997) for details. In terms of $\mathbf{C} = \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)m^2(x)dx$ and $\mathbf{G} = \mathbf{C} - \mathbf{B}^2$ we find (denoting the largest characteristic root by ch_1) that

$$\max_{f \in \mathcal{F}} \mathcal{L}_Q(f, \xi) = \tau^2(\nu \cdot \mathrm{tr}\mathbf{B}^{-1} + \mathrm{ch}_1 \mathbf{G}\mathbf{B}^{-2} + 1), \tag{6}$$

$$\max_{f \in \mathcal{F}} \mathcal{L}_A(f, \xi) = \tau^2(\nu \cdot \mathrm{tr}\mathbf{B}^{-1} + \mathrm{ch}_1 \mathbf{G}\mathbf{B}^{-2}), \tag{7}$$

$$\max_{f \in \mathcal{F}} \mathcal{L}_D(f, \xi) = \tau^2 \left(\frac{\sigma^2}{n}\right)^{2^{m+1}-1} \left(\frac{\nu + \mathrm{ch}_1 \mathbf{G}\mathbf{B}^{-1}}{|\mathbf{B}|}\right). \tag{8}$$

Note from (6) and (7) that *Q*- and *A*-minimax robust designs are necessarily identical, as are *Q*- and *A*-variance optimal designs. This is a consequence of the orthogonality property (4(ii)); it does not hold for non-orthogonal bases.

3.1. Minimax Designs for Haar Wavelet Approximations

The following lemma provides a condition under which a design ξ_0 will be *Q*- and *A*-variance optimal for the Haar wavelet approximation. It turns out that any design having the property prescribed by this theorem is also *D*- and *G*-variance optimal.

LEMMA 3.1 *For the Haar wavelet model with regressors $q(x) = q_{1,m}(x)$, any design ξ_0 with $B(\xi_0) = I_{2^{m+1}}$ minimizes $tr B^{-1}(\xi)$. In particular, any design ξ_* which places mass $2^{-(m+1)}$ in each of the 2^{m+1} intervals $\{[2^{-(m+1)}k, 2^{-(m+1)}(k+1))\}_{k=0,1,\dots,2^{m+1}-1}$ has $B(\xi_*) = I_{2^{m+1}}$, hence is Q- and A-variance optimal.*

Remarks

1. It follows from Lemma 3.1 together with Theorem 2.11.1 of Fedorov (1972) (since $B^{-1}(\xi_*) = B^{-2}(\xi_*)$) that ξ_* is D-variance optimal as well as A-variance optimal. It then follows from the Equivalence Theorem (Theorem 2.2.1 of Fedorov, 1972) that ξ_* is G-variance optimal as well, in that it minimizes $\max_{x \in \mathcal{S}} \text{var}[\hat{Y}(x)]$.
2. The D-variance optimality of ξ_* was also established by Herzberg and Traves (1994) using a completely different method of proof.

We can extend the variance optimality of the continuous version of ξ_* to a particularly strong form of minimax robustness. For this, let $\mathcal{L}$ be any of $\mathcal{L}_Q$, $\mathcal{L}_A$, $\mathcal{L}_D$. Denote by ξ_U the continuous uniform design, with density $m(x) \equiv 1$. Then for any design ξ and any $f \in \mathcal{F}$ we have

$$\mathcal{L}(f, \xi_U) = \mathcal{L}(0_{\mathcal{F}}, \xi_U) \leq \mathcal{L}(0_{\mathcal{F}}, \xi) \leq \mathcal{L}(f, \xi).$$

The equality above is (4(i)). The first inequality is the variance optimality of ξ_U and the second inequality follows from the definitions of the loss functions. (When $\mathcal{L} = \mathcal{L}_Q$ one should add $\int_0^1 f^2(x)dx$ to each occurrence of $\mathcal{L}(0_{\mathcal{F}}, \cdot)$.) Thus we have:

THEOREM 3.2 *For the Haar wavelet approximation with regressors $q_{1,m}(x)$, the continuous uniform design ξ_U minimizes $\mathcal{L}_A(f, \xi)$, $\mathcal{L}_Q(f, \xi)$ and $\mathcal{L}_D(f, \xi)$, among all designs ξ and for any $f \in \mathcal{F}$. In particular ξ_U is the minimax robust design for these loss functions.*

3.2. Minimax Designs for Multiwavelet Approximations

For the multiwavelet approximation our design results are less complete than those for the Haar wavelets. To illustrate some of the issues involved we exhibit minimax robust designs for (3) with $q(x) = q_{2,0}(x)$.

We shall consider only designs with densities in

$$\mathcal{M}_S = \{m(\cdot)|m(x) = m(1-x) = m(1/2-x) = m(1/2+x),\ 0 \leq x \leq 1/2\}.$$

Such densities are symmetric within each of $[0, 1]$, $[0, 1/2]$ and $[1/2, 1]$, hence are periodic with period $1/2$. They can also be characterized as symmetric functions of $x \in [0, 1]$ which depend on x only through $\{2x\}$. Our motivation for the restriction to $\mathcal{M}_S$ rests on the observations that $\|\mathbf{q}(x)\|$, the minimax designs of Section 3.1 and the *mvu* densities of Section 4 are of this form, and that for $m \in \mathcal{M}_S$ the matrices $\mathbf{B}$ and $\mathbf{C}$ are nearly diagonal, leading to easily determined eigenvalues with tractable structure.

THEOREM 3.3 *For the multiwavelet approximation with regressors* $\mathbf{q}_{2,0}(x) = (\phi_0(x), \phi_1(x), _2w_0(x), _2w_1(x))^T$ *the Q-, A- and D-minimax robust designs in* $\mathcal{M}_S$ *have densities*

$$m_0(x; s) = m_0(1 - x; s) = r\left(\left(\frac{1}{4} - x\right)^2 - \frac{s}{16}\right)^+, \quad 0 \leq x \leq 1/2, \quad (9)$$

where $r = 48/\{1 - 3s + (2s^{3/2}I(s \geq 0))\}$ *is a normalizing constant and* $s = s(\nu) \in [-\infty, 1]$ *is chosen to minimize the maximum loss for fixed* ν. *The parameters* $\nu = \nu_{QA}$ *for Q- and A-minimax robustness, and* $\nu = \nu_D$ *for D-minimax robustness, satisfy*

$$\nu_{QA} = \begin{cases} \dfrac{9(3 - 5s)^2}{25(1 - 3s)^3}, & s \leq 0, \\[4mm] \dfrac{9(2s^{3/2} + 4s + 6\sqrt{s} + 3)^2}{25(1 + 2\sqrt{s})^3(1 - \sqrt{s})^2}, & 0 \leq s \leq 1; \end{cases}$$

$$\nu_D = \begin{cases} \dfrac{1}{(1 - 3s)}, & s \leq 0, \\[4mm] \dfrac{1 + 2s^{3/2}}{(1 + 2\sqrt{s})(1 - \sqrt{s})^2}, & 0 \leq s \leq 1. \end{cases} \qquad (10)$$

Some values of the constants are given in Tables I and II. The limiting values $s = -\infty$, 1 correspond to $\nu = 0$, ∞. The corresponding limiting designs are the uniform ($\nu = 0$) and the discrete design

TABLE I Some Q-, A-optimal parameter values for designs (9) for multiwavelet approximations

ν	r	s	Maximum loss[1]
100	663.7	.698	281.6
10	99.92	.263	31.04
5	62.21	.096	16.26
3.24	48.00	0.00	10.88
1.00	24.31	−.325	3.63
.50	15.26	−.715	1.88
.10	4.19	−3.48	.39
.01	.473	−33.53	.04

[1] See exhibit (A.4).

TABLE II Some D-optimal parameter values for designs (9) for multiwavelet approximations

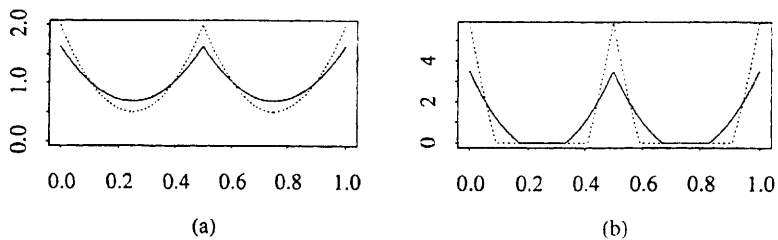| ν	r	s	Maximum loss[1]
100	1930	.821	14.27
10	269.5	.534	2.24
5	158.8	.403	1.40
1.00	48.00	0.00	.56
.5	24.00	−.333	.36
.1	4.80	−3.00	.09
.01	.480	−33.0	.01

[1] See exhibit (A.5).

FIGURE 2 Q- and A-minimax robust (solid line) and D-minimax robust (broken line) densities $m_0(x)$ for multiwavelet approximations with regressors $\mathbf{q}_{2,0}(x)$: (a) $\nu = .5$; (b) $\nu = 5$.

$(1/4)\delta_0 + (1/2)\delta_{1/2} + (1/4)\delta_1$ ($\nu = \infty$). See Figure 2 of plots of the Q- and D-minimax robust design densities for $\nu = .5$ and $\nu = 5$.

Note that $m_0(x; s)$ can be written as $p(\{2x\} - 1/2; s)$, where $p(t; s) = r((t^2/4) - (s/16))^+$ is a symmetric density on $[-1/2, 1/2]$. Motivated by

the structure revealed in the proof of Theorem 3.3, and by a comparison with the *mvu* designs, we make the following conjecture. Within the class of densities which are symmetric functions of $t = \{2^{m+1}x\} - 1/2$, the minimax density for regressors $\mathbf{q}_{2,m}(x)$ is $p(t; s)$, with s given by (10). Calculations carried out by computer algebra have given strong support to this conjecture.

4. MINIMUM VARIANCE UNBIASED DESIGNS

In this section we suppose that the coefficient vector θ_0 is to be estimated by *weighted* least squares, and seek designs and weights which minimize $\mathcal{L}_Q$, hence $\mathcal{L}_A$, subject to a condition of unbiasedness. For a non-negative weighting function $v(x)$ define the vectors and matrices

$$\mathbf{b} = \mathbf{b}(v, f, \xi) = \int_0^1 \mathbf{q}(x)v(x)f(x)d\xi(x),$$

$$\mathbf{B} = \mathbf{B}(v, \xi) = \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)v(x)d\xi(x),$$

$$\mathbf{D} = \mathbf{D}(v, \xi) = \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)v^2(x)d\xi(x).$$

In a more familiar regression notation these are $\mathbf{B} = n^{-1}\mathbf{Q}^T\mathbf{V}\mathbf{Q}$ and $\mathbf{D} = n^{-1}\mathbf{Q}^T\mathbf{V}^2\mathbf{Q}$, where the model matrix $\mathbf{Q}$ is as in Section 3 and $\mathbf{V}$ is the $n \times n$ diagonal matrix with diagonal elements $v(x_i)$. Then the weighted least squares estimate $\hat{\theta}_{\mathrm{WLS}} = (\mathbf{Q}^T\mathbf{V}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{V}\mathbf{Y}$ has bias vector and covariance matrix $\mathbf{B}^{-1}\mathbf{b}$ and $(\sigma^2/n)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}$ respectively. Let $k(x) = \xi'(x)$ be the design density and set $m = kv$. Assume, without loss of generality, that the average weight is $\int_0^1 v(x)d\xi(x) = 1$, so that m is a density on S. We say that a design/weights pair is *minimum variance unbiased (mvu)* if it minimizes $\mathcal{L}_Q$ and $\mathcal{L}_A$ subject to the unbiasedness condition

$$\mathbf{b}(v, f, \xi) = \int_0^1 \mathbf{q}(x)f(x)m(x)dx = \mathbf{0} \quad \text{for all } f \in \mathcal{F}. \tag{11}$$

By virtue of (4(i)), condition (11) holds if m is the uniform density of S. This uniformity has been shown to be necessary as well – see

Theorem 2b in Wiens (1998). It in turn implies that $\mathbf{B} = \mathbf{I}$ and then, apart from an additive constant $\int_0^1 f^2(x)dx$,

$$\mathcal{L}_Q = \mathcal{L}_A = \frac{\sigma^2}{n}\,\mathrm{tr}\mathbf{D} = \frac{\sigma^2}{n}\int_0^1 \|\mathbf{q}(x)\|^2 v(x)dx.$$

An *mvu* design then has

$$v_0 = \mathrm{argmin}\left\{\int_0^1 \|\mathbf{q}(x)\|^2 v(x)dx \Big| \int_0^1 v(x)^{-1}dx = 1\right\}.$$

Standard variational arguments give:

THEOREM 4.1 *The minimum variance unbiased design has density* $k_0(x) = v_0(x)^{-1}$, *where the mvu weights are (proportional to)*

$$v_0(x) = \frac{\int_0^1 \|\boldsymbol{q}(x)\|dx}{\|\boldsymbol{q}(x)\|}.$$

4.1. MVU Designs for Haar Wavelet Approximations

By Theorem 3.2 the uniform design, with constant weights, is unbiased as well as minimax robust, hence is *mvu* for the Haar wavelet approximation. This also follows from Theorem 4.1 since, by Theorem 2.1, $\|\mathbf{q}_{1,m}(x)\|$ is constant.

4.2. MVU Designs for Multiwavelet ($N \geq 2$) Approximations

We consider the multiwavelet approximation with regressors $\mathbf{q}(x) = \mathbf{q}_{N,m}(x)$. The *mvu* design densities are periodic functions, with period $2^{-(m+1)}$. The following theorem is an immediate consequence of Theorems 4.1 and 2.1.

THEOREM 4.2 *For the multiwavelet $q_{N,m}$-approximation the mvu design density is*

$$k_{N,m}(x) = \kappa_N \cdot [\phi_{N-1}(\{2^{m+1}x\})\phi_N'(\{2^{m+1}x\}) \\ - \phi_{N-1}'(\{2^{m+1}x\})\phi_N(\{2^{m+1}x\})]^{(1/2)},$$
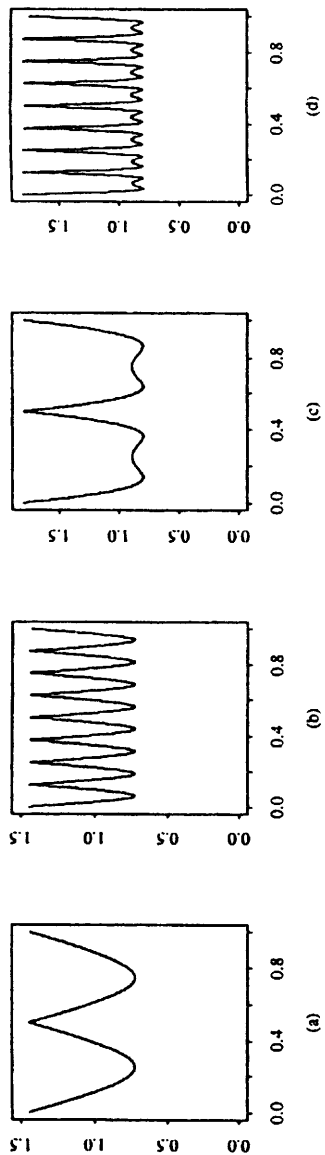
FIGURE 3 Minimum variance unbiased design densities for multiwavelet approximations: (a) $k_{2,0}(x)$; (b) $k_{2,2}(x)$; (c) $k_{3,0}(x)$; (d) $k_{3,2}(x)$.

where the normalizing constant is

$$\kappa_N = \left(\int_0^1 [\phi_{N-1}(x)\phi'_N(x) - \phi'_{N-1}(x)\phi_N(x)]^{(1/2)} dx \right)^{-1}.$$

The mvu weights are $v_{N,m}(x) \propto k_{N,m}(x)^{-1}$. *The limiting density is*

$$\lim_{N \to \infty} k_{N,m}(x) = \frac{\{2^{m+1}x\}^{-1/4}(1 - \{2^{m+1}x\})^{-1/4}}{\beta((3/4),(3/4))}.$$

It can be shown that the local maxima of $k_{N,m}(x)$ are the zeros of the function $\{2^{m+1}x\}(1 - \{2^{m+1}x\})\phi'_{N-1}(\{2^{m+1}x\})$. Recall that the design ξ_D which is D-variance optimal for $(N-1)$th degree polynomial regression places mass N^{-1} at each of the zeros of $x(1 - x)\phi'_{N-1}(x)$. The *mvu* design can be viewed as a smoothed, translated and dilated version of ξ_D. The design ξ_D has a limiting *beta*(1/2, 1/2) density.

Some particular cases are:

$$k_{2,m}(x) = 2.5099 \cdot \left[\left(\{2^{m+1}x\} - \frac{1}{2} \right)^2 + \frac{1}{12} \right]^{(1/2)},$$

$$k_{3,m}(x) = 8.0024 \cdot \left[\left(\left(\{2^{m+1}x\} - \frac{1}{2} \right)^2 - \frac{1}{20} \right)^2 + \frac{1}{100} \right]^{(1/2)}.$$

See Figure 3 for plots in the cases $m = 0$ and $m = 2$.

5. EXAMPLES

5.1. Example 1: Motorcycle Impact Data

The motorcycle impact data are discussed in Härdle (1990) and elsewhere in the nonparametric regression literature. In particular, Antoniadis *et al.* (1994) fit a wavelet version of a kernel estimate to these data. The response variable Y is a measure of the head acceleration of a post mortem human test object, at time x after a simulated motorcycle impact.

We have obtained three designs for the range of these data:

1. The minimax design of Section 3.1, for Haar wavelets $q_{1,4}(x)$. The 64 design points are equally spaced over the range $[2.4, 57.7)$ of the original data.
2. The *mvu* design of Section 4.2 for $q_{2,3}(x)$, with $n = 64$. The first four design points, in $[0, 1/16)$, are the $k/64$-quantiles of $k_{2,3}(x)$ of Theorem 4.2 ($k = 0, 1, 2, 3$). Explicitly, $x_1 = 0$, $x_2 = .013$, $x_3 = .031$, $x_4 = .050$. These are extended by periodicity to the remainder of $[0, 1)$ and then linearly transformed to span the range of the original data. The regression weights $v(x) = 1/k_{2,3}(x)$, normalized to have an average of one, are $v(x_1) = .692$, $v(x_2) = .962$, $v(x_3) = 1.384$, $v(x_4) = .962$; periodic thereafter.
3. The *mvu* design of Section 4.2 for $q_{3,2}(x)$, with $n = 48$. The first six design points, in $[0, 1/8)$, are $x_1 = 0$, $x_2 = .015$, $x_3 = .038$, $x_4 = .062$, $x_5 = .087$, $x_6 = .110$; these are extended periodically and transformed linearly as above. The corresponding regression weights are $v(x_1) = .563$, $v(x_2) = .906$, $v(x_3) = 1.249$, $v(x_4) = 1.126$, $v(x_5) = 1.249$, $v(x_6) = 0.906$.

We simulated data at these design points in the following way. First, a *loess* smoother was fitted to the original data, using S-PLUS software with a span of .1. Predicted values $\hat{y}_i^{(s)}$ of the smoother were then obtained at each design point, and a random error ε_i added to these values. The random errors were independently and normally distributed, with a standard deviation equal to that of the estimated standard deviation of $\hat{y}_i^{(s)}$.

The first two fits employ 32 regressors each; the third employs 24. The data and fits are shown in Figure 4. As remarked by
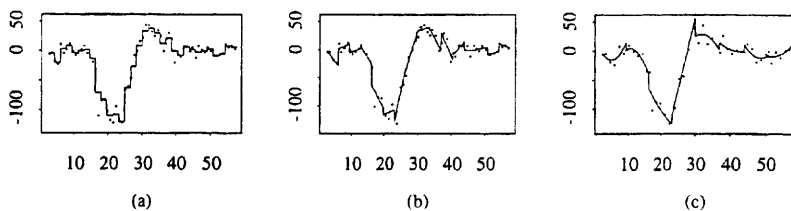

FIGURE 4 Designs and fitted response curves for the motorcycle impact data: (a) Haar wavelets $q_{1,4}(x)$; (b) multiwavelets $q_{2,3}(x)$; (c) multiwavelets $q_{3,2}(x)$.

Antoniadis *et al.* (1994), one purpose of the analysis is the estimation of the extremes of the response. Both *mvu* fits seems to be quite suitable for this. The Haar wavelet fit resembles Tukey's (1961) *regressogram*. Multiwavelet fits using $q_{2,4}(x)$ and $q_{3,3}(x)$ were obtained as well but seemed too oscillatory, especially in the range $x > 30$. Minimax designs for $q_{2,0}(x)$, as in Section 3.2, can be constructed by placing design points at the quantiles of $m_0(x)$. These are not shown since this approximation is not appropriate for these data.

5.2. Example 2: Sawtooth Data

McDonald and Owen (1986) discuss smoothing techniques for curves with steps, abruptly changing derivatives, or cusps. One example given is that of a sawtooth function on $[0, 1]$, for which simulated data are shown in Figure 5. The mean response is $E[Y|x] = \{2x\}$ and the additive random errors are independently and normally distributed, with a standard deviation equal to one-half that of the values of the mean response, as in McDonald and Owen (1986). Two designs for regressors $q_{2,0}(x)$ – an *mvu* design as at Section 4.2, and a minimax design as at Section 3.2 with $s = 0$, were constructed. Each has 20 sites. The *mvu* design points in $[0, 1/4)$ are $x_1 = 0$, $x_2 = .036$, $x_3 = .078$, $x_4 = .126$, $x_5 = .183$. These are extended to $[1/4, 1/2)$ by symmetry about $1/4$, then to $[1/2, 1)$ by periodicity. The minimax design points in $[0, 1/4)$ are $x_1 = 0$, $x_2 = .018$, $x_3 = .039$, $x_4 = .066$, $x_5 = .104$ and are extended in the same manner.

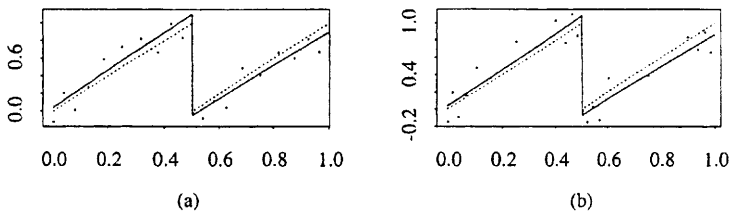
FIGURE 5   Multiwavelet fit to simulated sawtooth data at 20 sites. Regressors are $q_{2,0}(x)$. (a) *mvu* design; (b) minimax design. Broken line is the mean response $E[Y|x] = \{2x\}$.

6. SUMMARY

We have illustrated the construction of robust designs for wavelet approximations of regression functions. These designs are efficient in the presence of random variation, and robust against bias incurred by inadequacies in the wavelet approximation. A uniform design has been seen to be optimal for Haar wavelet approximations. For multiwavelet approximations we have given minimum variance unbiased designs and regression weights, for models of any order. Further work remains to be done on the construction of minimax designs for multiwavelet approximations.

References

Alpert, B. K. (1992) "Wavelets and Other Bases for Fast Numerical Linear Algebra", In: *Wavelets: A Tutorial in Theory and Applications*, Ed. Chui, C. K., Boston: Academic Press, pp. 181–216.

Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994) "Wavelet Methods for Curve Estimation", *Journal of the American Statistical Association*, **89**, 1340–1352.

Benedetto, J. J. and Frazier, M. W. (Eds.) (1994) *Wavelets: Mathematics and Applications*, Boca Raton, Fla.: CRC Press.

Box, G. E. P. and Draper, N. R. (1959) "A Basis for the Selection of a Response Surface Design", *Journal of the American Statistical Association*, **54**, 622–654.

Chui, K. (1992) *Wavelets: A Tutorial in Theory and Applications*, Boston: Academic Press.

Daubechies, I. (1993) "Two Recent Results in Wavelets: Wavelet Bases on the Interval, and Biorthogonal Wavelets Diagonalizing the Derivative Operator", In: *Recent Advances in Wavelet Analysis*, Eds. Schumaker, L. L. and Webb, G., Boston: Academic Press, pp. 237–258.

Fedorov, V. V. (1972) *Theory of Optimal Experiments*, trans. and Eds. Studden, W. J. and Klimko, E. M., New York: Academic Press.

Gradshteyn, I. S. and Ryzhik, I. M. (1980) *Tables of Integrals, Series and Products*, Toronto: Academic Press.

Härdle, W. (1990) *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*, New York: Springer-Verlag.

Herzberg, A. M. and Traves, W. N. (1994) "An Optimal Experimental Design for the Haar Regression Model", *Canadian Journal of Statistics*, **22**, 357–364.

Huber, P. J. (1975) "Robustness and Designs", In: *A Survey of Statistical Design and Linear Models*, Ed. Srivastava, J. N., Amsterdam: North Holland, pp. 287–303.

Mallat, S. (1989) "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\Re)$", *Transactions of the American Mathematical Society*, **315**, 69–87.

McDonald, J. A. and Owen, A. B. (1986) "Smoothing With Split Linear Fits", *Technometrics*, **28**, 195–208.

Meyer, Y. (1986) *Ondellettes, fonctons splines et analyses graduees*, Lectures given at the University of Torino, Italy.

Meyer, Y. (1992) *Wavelets and Operators*, Eds. Garling, D. J. H., Gorenstein, D., Tom Dieck, T. and Walters, P., Cambridge University Press.

Oyet, A. J. (1997) "Robust Designs for Wavelet Approximations of Nonlinear Models", unpublished Ph.D. Thesis, University of Alberta, Department of Mathematical Sciences.

Pesotchinsky, L. (1982) "Optimal Robust Designs: Linear Regression in R^k", *The Annals of Statistics*, **10**, 511–525.

Sacks, J. and Ylvisaker, D. (1984) "Some Model Robust Designs in Regression", *The Annals of Statistics*, **12**, 1324–1348.

Strang, G. (1989) "Wavelets and Dilation Equations: A Brief Introduction", *SIAM Review*, **31**, 614–627.

Tukey, J. W. (1961) "Curves as Parameters and Touch Estimation", *Proceedings of the 4th Berkeley Symposium*, pp. 681–694.

Walter, G. G. (1995) "Pointwise Convergence of Wavelet Expansions", *Journal of Approximation Theory*, **80**, 108–118.

Wiens, D. P. (1992) "Minimax Designs for Approximately Linear Regression", *Journal of Statistical Planning and Inference*, **31**, 353–371.

Wiens, D. P. and Zhou, J. (1997) "Robust designs based on the infinitesimal approach", *Journal of the American Statistical Association*, **92**, 1503–1511.

Wiens, D. P. (1998) "Minimax Robust Designs and Weights for Approximately Specified Regression Models with Heteroscedastic Errors", *Journal of the American Statistical Association*, **93**, 1440–1450.

Wiens, D. P. (1999) "Robust Weights and Designs for Biased Regression Models: Least Squares and Generalized M-estimation", *Journal of Statistical Planning and Inference*, to appear.

APPENDIX: DERIVATIONS

Proof of Theorem 2.1 Define vectors $\phi(x) = (\phi_0(x), \ldots, \phi_{N-1}(x))^T$ and $\mathbf{w}(x) = (_N w_0(x), \ldots, _N w_{N-1}(x))^T$. Using (1) we derive

$$\|\mathbf{q}_{N,m}(x)\|^2 = \|\phi(x)\|^2 + \sum_{j=0}^m 2^j \|\mathbf{w}(\{2^j x\})\|^2. \qquad (A.1)$$

Now define vectors $\mathbf{s}(x) = (\phi(x)^T, \mathbf{w}(x)^T)^T$ and

$$\mathbf{t}(x) = (\phi(x)^T, -\phi(\{2x\})^T \cdot (I_{[0,1/2)}(x) - I_{[1/2,1)}(x)))^T.$$

The algorithm of Alpert (1992, p. 198) for the construction of multi-wavelets presents the orthonormal vector $\mathbf{s}(x)$ as a particular linear transformation $\mathbf{s}(x) = \mathbf{M}\mathbf{t}(x)$, where $\mathbf{M}$ is non-singular. The elements

of $t(x)$ are a basis for the space of functions that are polynomials of degree less than N on $[0, 1/2)$ and on $[1/2, 1)$. This space has as an orthonormal basis the elements of $\chi(x) = (\sqrt{2}\phi(\{2x\})^T I_{[0,1/2)}(x),$ $\sqrt{2}\phi(\{2x\})^T I_{[1/2,1)}(x))^T$. Thus $s(x) = P\chi(x)$ for some square matrix P. Since $s(x)$ and $\chi(x)$ are orthonormal, P is necessarily orthogonal and so

$$\|\phi(x)\|^2 + \|w(x)\|^2 = \|s(x)\|^2 = \|\chi(x)\|^2 = 2\|\phi(\{2x\})\|^2,$$

whence

$$\|w(x)\|^2 = 2\|\phi(\{2x\})\|^2 - \|\phi(x)\|^2.$$

Upon substituting this relationship into (A.1) the sum collapses to yield

$$\|q_{N,m}(x)\|^2 = 2^{m+1}\|\phi(\{2^{m+1}x\})\|^2.$$

This may be evaluated with the aid of formula 8.915.1 of Gradshteyn and Ryzhik (1980) to give the expression in the statement of the theorem. A standard asymptotic expansion for Legendre polynomials – formula 8.965 of Gradshteyn and Ryzhik (1980) – yields (2). ∎

Proof of Lemma 3.1 Consider the convex combination $\xi_t = (1 - t)\xi_0 + t\xi_1 (t \in [0,1])$ for any design ξ_1 on S, and define $\rho(t) = \text{tr}B^{-1}(\xi_t)$. The function $\rho(t)$ is convex and so is minimized at $t = 0$ if and only if $\rho'(0) \geq 0$ for all ξ_1. We calculate that

$$\rho'(t) = -\text{tr}\left\{B^{-1}(\xi_t) \int_0^1 q(x)q^T(x)d(\xi_1(x) - \xi_0(x))B^{-1}(\xi_t)\right\}.$$

The $\text{tr}B^{-1}(\xi_0)$ is a minimum if $B(\xi_0) = I_{2^{m+1}}$ and if as well

$$\text{tr} \int_0^1 q(x)q^T(x)d\xi(x) = \int_0^1 \|q(x)\|^2 d\xi(x)$$

is maximized by ξ_0. This latter condition follows trivially since, by Theorem 2.1, $\|q(x)\|^2$ is constant on $[0, 1)$.

Thus ξ_0 is optimal. To establish the optimality of ξ_* we note that, since $\mathbf{q}(x)$ is constant in each interval $[2^{-(m+1)}k, 2^{-(m+1)}(k+1))$, we have

$$
\begin{aligned}
\mathbf{B}(\xi_*) &= \sum_{k=0}^{2^{m+1}-1} \int_{(k/2^{m+1})}^{(k+1/2^{m+1})} \mathbf{q}(x)\mathbf{q}^T(x)d\xi_*(x) \\
&= \sum_{k=0}^{2^{m+1}-1} \int_{(k/2^{m+1})}^{(k+1/2^{m+1})} \mathbf{q}(x)\mathbf{q}^T(x)dx \\
&= \int_0^1 \mathbf{q}(x)\mathbf{q}^T(x)dx = \mathbf{I}_{2^{m+1}}.
\end{aligned}
$$

∎

Proof of Theorem 3.3   First note that $m \in \mathcal{M}_S$ iff $m = \bar{m}$, where the 'doubly symmetrized' version $\bar{m}$ of m is defined by

$$
\bar{m}(x) = \bar{m}(1-x) = \frac{m(x) + m(1-x) + m(1/2-x) + m(1/2+x)}{4},
$$
$$
0 \le x \le 1/2.
$$

When $m \in \mathcal{M}_S$ we then have the identity

$$
\int_0^1 \phi(x)m(x)dx = 4\int_0^{1/4} \bar{\phi}(x)m(x)dx, \tag{A.2}
$$

valid for any $\phi(\cdot)$ for which the integrals exist. We find that $\overline{\phi_1}(x) = \overline{_2w_0}(x) = \overline{_2w_1}(x) = \overline{_2w_0\,_2w_1}(x) = \overline{\phi_1\,_2w_0}(x) = 0$, and that for $x \in [0, 1/2]$,

$$
\overline{_2w_0^2}(x) = 48\left(x - \frac{1}{4}\right)^2, \quad \overline{_2w_1^2}(x) = 36\left(x - \frac{1}{4}\right)^2 + \frac{1}{4},
$$
$$
\overline{\phi_1^2}(x) = 12\left(x - \frac{1}{4}\right)^2 + \frac{3}{4}, \quad \overline{\phi_1\,_2w_1}(x) = \sqrt{3}\left(12\left(x - \frac{1}{4}\right)^2 - \frac{1}{4}\right).
$$

Order the elements of $\mathbf{q}(x)$ as $\mathbf{q}(x) = (\phi_0(x), _2w_0(x), \phi_1(x), _2w_1(x))^T$ and define

$$
\beta_0 = 4\int_0^{1/4} \left(x - \frac{1}{4}\right)^2 m(x)dx,
$$
$$
\gamma_0 = 4\int_0^{1/4} \left(x - \frac{1}{4}\right)^2 m^2(x)dx, \quad \gamma = 4\int_0^{1/4} m^2(x)dx. \tag{A.3}
$$

Using (A.2) we obtain $\mathbf{B} = \mathrm{diag}(1, 48\beta_0, \mathbf{B}_2)$ and $\mathbf{C} = \mathrm{diag}(1, 48\gamma_0, \mathbf{C}_2)$, where

$$\mathbf{B}_2 = \begin{pmatrix} 12\beta_0 + 3/4 & \sqrt{3}(12\beta_0 - 1/4) \\ \sqrt{3}(12\beta_0 - 1/4) & 36\beta_0 + 1/4 \end{pmatrix},$$

$$\mathbf{C}_2 = \begin{pmatrix} 12\beta_0 + (3/4)\gamma & \sqrt{3}(12\beta_0 - (1/4)\gamma) \\ \sqrt{3}(12\beta_0 - (1/4)\gamma) & 36\beta_0 + (1/4)\gamma \end{pmatrix}.$$

We find that the characteristic roots of $\mathbf{GB}^{-2}$ are $\gamma - 1$ and $(\gamma_0/48\beta_0^2) - 1$, each appearing with multiplicity two; those of $\mathbf{GB}^{-1}$ also appear in pairs and are $\gamma - 1$ and $(\gamma_0/\beta_0) - 48\beta_0$. Thus from (6)–(8), ignoring some multiplicative and additive constants independent of $m(\cdot)$,

$$\max_{f \in \mathcal{F}} \mathcal{L}_Q(f, \xi) = \max_{f \in \mathcal{F}} \mathcal{L}_A(f, \xi) = 2\nu \left(1 + \frac{1}{48\beta_0} \right) \tag{A.4}$$
$$+ \max \left(\gamma - 1, \frac{\gamma_0}{48\beta_0^2} - 1 \right),$$

$$\max_{f \in \mathcal{F}} \mathcal{L}_D(f, \xi) = \frac{\nu}{(48\beta_0)^2} + \frac{1}{(48\beta_0)^2} \max \left(\gamma - 1, \frac{\gamma_0}{48\beta_0} - 48\beta_0 \right). \tag{A.5}$$

In each case the maximum characteristic root, evaluated at m_0, is $\gamma - 1$; this is easily verified once m_0 is determined.

We minimize the loss first for fixed β_0. For all three loss functions this requires the minimization of $\gamma = \gamma(m)$, subject to the constraint on β_0 and the requirement that $m(\cdot)$ be a density. It suffices that m minimize

$$\int_0^{1/4} \left\{ m^2(x) - 2r \left(x - \frac{1}{4} \right)^2 m(x) + \frac{rs}{8} m(x) \right\} dx \tag{A.6}$$

for Lagrange multipliers r and s, and satisfy the side conditions. The multipliers have been arranged in such a way that the integrand of (A.6) is minimized pointwise by (9). Since $m_0(x; s)$ is already symmetric within $[0, 1/2]$, membership in $\mathcal{M}_S$ is ensured by the requirement $m_0(x; s) = m_0(1 - x; s)$.

In the case $s \leq 0$ we calculate that

$$\beta_0 = \frac{3 - 5s}{80(1 - 3s)}, \quad \gamma_0 = \frac{3(15 - 42s + 35s^2)}{560(1 - 3s)^2}, \quad \gamma = \frac{3(3 - 10s + 15s^2)}{5(1 - 3s)^2};$$

for $s \geq 0$ these become

$$\beta_0 = \frac{2s^{3/2} + 4s + 6\sqrt{s} + 3}{80(1 + 2\sqrt{s})},$$

$$\gamma_0 = \frac{3(8s^2 + 24s^{3/2} + 48s + 45\sqrt{s} + 15)}{560(1 + 2\sqrt{s})^2(1 - \sqrt{s})},$$

$$\gamma = \frac{3(8s + 9\sqrt{s} + 3)}{5(1 + 2\sqrt{s})^2(1 - \sqrt{s})}.$$

In each case the loss may now be written in the form $\mathcal{L} = g(s)\nu + h(s)$ for appropriate functions g and h. The minimizing value of s satisfies $\nu = -h'(s)/g'(s)$, resulting in (10). ∎