Journal of Statistical Planning and Inference 184 (2017) 94–104

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Optimal designs for spline wavelet regression models

Jacob M. Maronge [a], Yi Zhai [a], Douglas P. Wiens [b], Zhide Fang [a,*]

[a] *Biostatistics Program, Louisiana State University Health Sciences Center, New Orleans, LA, 70112, USA*

[b] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, T6G 2G1, Canada*

A R T I C L E   I N F O

Article history:
Received 16 June 2016
Received in revised form 26 October 2016
Accepted 27 November 2016
Available online 20 December 2016

MSC:
primary 62K05
62G35
secondary 62G07

Keywords:
Haar wavelets
Minimax
Model robustness
Optimal designs
Spline wavelets
Trigonometric regression
Wavelet regression

A B S T R A C T

In this article we investigate the optimal design problem for some *wavelet regression* models. Wavelets are very flexible in modeling complex relations, and optimal designs are appealing as a means of increasing the experimental precision. In contrast to the designs for the Haar wavelet regression model (Herzberg and Traves 1994; Oyet and Wiens 2000), the *I*-optimal designs we construct are different from the *D*-optimal designs. We also obtain *c*-optimal designs. Optimal (*D*- and *I*-) *quadratic spline wavelet* designs are constructed, both analytically and numerically. A case study shows that a significant saving of resources may be realized by employing an optimal design. We also construct *model robust* designs, to address response misspecification arising from fitting an incomplete set of wavelets.

1. Introduction and summary

Although their history dates back to the early 20th century (Haar, 1910), wavelets have been particularly popular in recent decades, both in theory and in applications. Applications of wavelets, in basic and biomedical sciences, can be found in the fields of signal processing, numerical analysis, and function representation, to mention only a few. See Chui (1992), Alpert (1992), Daubechies (1993), Antoniadis and Oppenheim (1995), Strang and Nguyen (1996) and Wojtaszczyk (1997) for more details. The use of wavelets in functional data analysis leads to applications in regression and subsequently in design theory—in a compact design space, one can use finitely many wavelet basis functions as regressors, to provide an adequate fit to the true response function. The wavelet model fitting becomes more precise as the order of the wavelets increases.

Since the 1950s, design theory for regression models has been developed extensively, in numerous papers. See, for example, Fedorov (1972), Dette and Studden (1997), Pukelsheim (1993), Atkinson et al. (2007) and references therein for detailed definitions, results and discussions.

We entertain an experimental setup in which, given inputs x_i chosen from a compact 'design space' $\mathcal{X}$, the n outcomes Y_i depend on x_i via a regression model

$$Y_i = \sum_{j=0}^{p} \theta_j f_j(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

* Corresponding author.
E-mail address: zfang@lsuhsc.edu (Z. Fang).

http://dx.doi.org/10.1016/j.jspi.2016.11.005
0378-3758/

Here $\{f_j(x)\}_{j=0}^p$ are known functions, $\boldsymbol{\theta} := (\theta_0, \theta_1, \ldots, \theta_p)^T$ is a vector of unknown parameters, and the random errors ε_i are uncorrelated, with mean zero and variance σ^2. Given data $\{(x_i, y_i)\}_{i=1}^n$, $\boldsymbol{\theta}$ is estimated by the least squares estimate $\hat{\boldsymbol{\theta}}$. The design problem is to choose $\{x_i\}_{i=1}^n$, not necessarily distinct, in order to minimize some scalar function of the variance–covariance matrix $\text{cov}[\hat{\boldsymbol{\theta}}]$.

The *exact* experimental design ξ prescribes the number $n\xi(x)$ of observations to be taken at a point $x \in \chi$. Since the integer constraints typically make the exact design problem analytically intractable, we adopt the theory of *approximate design*, allowing ξ to be any probability measure supported on finitely many points in $\mathcal{X}$. Procedures to implement an approximate design ξ are discussed in Chapter 12 of Pukelsheim (1993). Intuitively, one makes $\lfloor n\xi(x)\rfloor$ observations at x and then distributes the remaining resources in a prescribed manner—see our case study in Section 5.

Define $\mathbf{f}(x) = (f_0(x), f_1(x), \ldots, f_p(x))^T$. The information matrix of a design ξ is

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}^T(x)d\xi,$$

and then $\text{cov}[\hat{\boldsymbol{\theta}}] = \left(\sigma^2/n\right)\mathbf{M}^+(\xi)$, where for definiteness we use the Moore–Penrose inverse $\mathbf{M}^+$. If $\mathbf{c} \in Range\,(\mathbf{M}(\xi))$ we say that $\mathbf{c}^T\boldsymbol{\theta}$ is *estimable*, and then the *c-optimal* design is that which minimizes $\mathbf{c}^T\mathbf{M}^+(\xi)\mathbf{c}$. When $\mathbf{M}(\xi)$ is non-singular, which we shall assume except when discussing c-optimality, $\mathbf{M}^+ = \mathbf{M}^{-1}$ and we can define optimality criteria through scalar functions of $\mathbf{M}^{-1}(\xi)$, as follows. Define $d(x, \xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x)$; note that the variance of the estimate $\mathbf{f}^T(x)\hat{\boldsymbol{\theta}}$ of $E[Y|x]$ is proportional to $d(x, \xi)$, with the constant σ^2/n of proportionality being independent of the design. A design ξ is called D-optimal, A-optimal, E-optimal, G-optimal or I-optimal if it minimizes $|\mathbf{M}^{-1}(\xi)|$, $trace(\mathbf{M}^{-1}(\xi))$, $\lambda_{\max}(\mathbf{M}^{-1}(\xi))$ (the largest eigenvalue), $\max_{x\in\mathcal{X}} d(x, \xi)$ or $\int_{\mathcal{X}} d(x, \xi)dx$, respectively.

In Section 3 we construct optimal designs for those wavelet regression models developed in Section 2. When $\mathbf{f}(x)$ is a vector of Haar wavelets, Herzberg and Traves (1994) obtained the D-optimal design. Oyet and Wiens (2000) show that this design is also A-optimal and I-optimal. Xie and Li (unpublished) find the D-optimal design for piecewise linear spline wavelet regression. In this article, we show that this design, given by Xie and Li, is A- and E-optimal but not I-optimal. We construct the I-optimal and c-optimal designs. The latter case includes optimal designs for interpolation and extrapolation. Since $\mathbf{f}(x)$ is no longer a Chebyshev system, one can view the c-optimality as a partial extension of Studden (1968).

We also obtain results for the quadratic spline wavelet model. The mathematical complexity restricts us to some simple cases, and so in Section 4 we discuss the numerical construction of these designs. We exhibit these in several cases which are apparently not yet in the literature. Then in Section 5 we test these designs on simulated data based on an experiment in automotive engineering, with a response function which is difficult to model in a more conventional way. We finish with a discussion of *model robustness* of design; again algorithms are given leading to the construction of both approximate and exact designs.

The usefulness of wavelet methods in regression arises from the ability of linear combinations of wavelets to approximate arbitrarily closely any response function which is square integrable over the design space. Of course Fourier expansions yield this property as well, and depending on the complexity of the response function such an expansion might lead to a more economical representation, especially when the response is periodic. There is a substantial literature on design for such *trigonometric regression* models, well-reviewed by Dette et al. (2011). The situation is somewhat similar to that of design for quadratic spline wavelet regression as discussed in this article—except for quite simple models analytic results are sparse, but there are theorems allowing one to check whether or not a particular design is optimal. The numerical methods we discuss in Sections 4 and 5 are as applicable to trigonometric models as to wavelet models.

The designs presented here are constructed via computations implemented in MATLAB; the code is available from us.

2. Preliminaries

We use notation and definitions as in Wojtaszczyk (1997) and begin with the representation of the space $\mathcal{L}^2(\mathcal{R})$ of square integrable functions on $\mathcal{R}$ as the union of an increasing sequence of closed sets of functions:

$$\mathcal{L}^2(\mathcal{R}) = \cup_{r\in\mathcal{Z}} V_r, \cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots.$$

The sets V_r are dilates of one another, in that $f(x) \in V_r$ iff $f(2^{-r}x) \in V_0$, the set V_0 is closed under integer translations:

$$f(x) \in V_0 \Leftrightarrow f(x-k) \in V_0 \; \forall k \in \mathcal{Z} = \{0, \pm 1, \pm 2, \ldots\},$$

(i.e. $f(x) = f(x - [x])$) and there exists a *scaling function* (also known as a *father wavelet*) $\phi \in V_0$ whose integer translates span V_0. The functions $\phi_{r,k}(x) = 2^{r/2}\phi(2^r x - k)$, for $k \in \mathcal{Z}$, form an orthonormal basis for V_r, which we shall call the *natural* basis. Through a *multiresolution analysis* the space V_r can be represented as the direct sum of mutually orthogonal subspaces $V_r = V_0 \oplus \oplus_{l=0}^{r-1} W_l$, where W_l is the orthogonal complement of V_l in V_{l+1}. An orthonormal basis of W_l is $\left\{\psi_{l,k}(x) = 2^{l/2}\psi(2^l x - k)\right\}$ for a suitably constructed *mother wavelet* ψ. An alternate, orthonormal *wavelet basis* for V_r is then given by the union of these bases for $\{V_0, W_l : l = 0, \ldots, r-1\}$.

We are concerned here with the approximation of a regression response $E[Y|x] \in \mathcal{L}^2[0, 1]$. We restrict to wavelet classes for which the restriction $V_0[0, 1]$ of the members of V_0 to $\chi = [0, 1]$ is finite dimensional, by virtue of satisfying the conditions of the following lemma due to Wojtaszczyk (1997). Since it is not proved there we give a proof in the Appendix.

Lemma 1. *Suppose that V_0 contains a function $\gamma(x)$ vanishing off of $[K_1, K_2]$ for integers K_1, K_2 and is such that:*

(A1) *the set $\{\gamma(x-k) \mid k \in \mathcal{Z}\}$ is a basis for V_0;*
(A2) *the set $\left\{\gamma(x-k)I_{[0,1]}(x) \mid k \in \mathcal{K}_0\right\}$, where $\mathcal{K}_0 := \{-K_2+1, \ldots, -K_1\}$, is linearly independent.*
Then:
(i) *the set $\left\{\gamma(x-k)I_{[0,1]}(x) \mid k \in \mathcal{K}_0\right\}$ is a basis for $V_0[0,1]$;*
(ii) *a basis for $V_r[0,1]$ is the set $\mathcal{V}_r := \left\{2^{r/2}\gamma(2^r x-k)I_{[0,1]}(x) \mid k \in \mathcal{K}_r\right\}$, where*

$$\mathcal{K}_r := \left\{i+j \mid i = 0, \ldots, 2^r-1,\ j \in \mathcal{K}_0\right\} = \left\{-K_2+1, \ldots, -K_1+2^r-1\right\};$$

(iii) *there is a set $\mathcal{K}_r^* \subset \mathcal{Z}$, with the same cardinality $p_r := K_2 - K_1 + 2^r - 1$ as $\mathcal{K}_r$, such that a basis for $V_r[0,1]$ is*

$$\mathcal{V}_r^* := \left\{v_{r,k}(x) = \phi_{r,k}(x)I_{[0,1]}(x) \mid k \in \mathcal{K}_r^*\right\}.$$

We shall refer to the basis in (iii) of Lemma 1 as the natural basis for $V_r[0,1]$. In a similar manner the elements of a wavelet basis for V_r may be restricted to form a wavelet basis for $V_r[0,1]$. There will of course be other bases as well. If the vectors $\mathbf{v}(x)$ and $\mathbf{w}(x)$ consist of the elements of two bases, then there is a linear relationship between them: $\mathbf{v}(x) = \mathbf{P}\mathbf{w}(x)$ for a nonsingular matrix $\mathbf{P}$.

Example 2.1 (*Haar Wavelets*)**.** Here the scaling function is $\phi(x) = I_{[0,1]}(x)$ and $V_r[0,1]$ consists of those functions which are constant on all sub-intervals $\{[2^{-r}k, 2^{-r}(k+1)], k = 0, \ldots, 2^r-1\}$. Lemma 1 holds with $K_1 = 0,\ K_2 = 1,\ \gamma = \phi$. A wavelet basis is determined from $\phi(x)$ and the mother wavelet $\psi(x) = I_{[0,1/2)}(x) - I_{[1/2,1]}(x)$; see Chapter 1 of Wojtaszczyk (1997) for details.

Example 2.2 (*Spline Wavelets*)**.** Let S_d be the space of all splines of order d with integer nodes, so that if $f \in S_d$ then $f \in C^{d-1}$ and is a polynomial of degree at most d within each interval $[k, k+1]$. This space has as a (Riesz) basis the translates $\{N_d(x-k) \mid k = -d, \ldots, 0\}$ of the B-splines $N_d(x)$ defined recursively by $N_0(x) = I_{[0,1]}(x)$, and (with '$*$' denoting convolution), $N_d(x) = N_{d-1}(x) * N_0(x)$. The support of $N_d(x)$ is $[0, d+1]$; in the interior of this interval $N_d(x)$ is strictly positive. Thus (A1) of Lemma 1 holds with $K_1 = 0, K_2 = d+1,\ \gamma = N_d$. An induction on d combined with the identity $N_d'(x) = N_{d-1}(x) - N_{d-1}(x-1)$ shows that (A2) holds as well. Note that when $d = 0$ the spline wavelets coincide with the Haar wavelets and $N_0(x)$ is the scaling function. For $d > 0$ the scaling function and mother wavelet are substantially more complicated; this need not concern us here since the basis given above is the most convenient one for our purposes. Xie (unpublished) combines a wavelet basis for $\oplus_{l=0}^{n-1} W_l[0,1]$ with the basis of Lemma 1(i) to obtain a regression model. The father wavelet for the $d = 1$ spline wavelet (linear spline wavelet) is

$$\phi(x) = N_1(x) = xI_{[0,1)}(x) + (-x+2)I_{[1,2]}(x), \tag{1}$$

with support $[0,2]$. Similarly, the father wavelet for the quadratic spline wavelet ($d = 2$) is

$$\phi(x) = N_2(x) = \frac{x^2}{2}I_{[0,1)}(x) + \left(-x^2+3x-\frac{3}{2}\right)I_{[1,2)}(x) + \left(\frac{x^2}{2} - 3x + \frac{9}{2}\right)I_{[2,3]}(x), \tag{2}$$

with support $[0,3]$.

Now, the regression model we consider is

$$E[Y|x] = \mathbf{f}^T(x)\boldsymbol{\theta} = \sum_{k=0}^{p_r-1}\theta_k f_k(x), \quad x \in [0,1], \tag{3}$$

with $\mathbf{f}(x)$ consisting of the p_r elements of $\mathcal{V}_r^*$.

It is well known that in this framework D-optimality is equivalent to G-optimality, and that D-, G- and I-optimal designs are invariant under reparameterizations $\mathbf{f}^T(x) \mapsto \mathbf{f}^T(x)\mathbf{P}^{-1}$, $\boldsymbol{\theta} \mapsto \mathbf{P}\boldsymbol{\theta}$ that leave the mean response unchanged. However A- and E-optimal designs do not have this property of invariance. As to c-optimality, it is invariant when $\mathbf{c}$ is of the form $\mathbf{f}(x_0)$, corresponding to interpolation when $x_0 \in [0,1]$ and to extrapolation when $x_0 \notin [0,1]$.

3. Optimal designs

3.1. Linear spline wavelet regression model

We use the linear spline wavelets determined by (1) to define the regressors in the regression model (3) with design space $\chi = [0,1]$. Thus for a chosen non-negative integer r,

$$f_k(x) = \phi(2^r x - k + 1) \tag{4}$$

where $k = 0, 1, \ldots, p_r - 1 = 2^r$. Xie and Li (unpublished) show that the equispaced design

$$\xi_D = \frac{1}{2^r+1} \sum_{i=0}^{2^r} \delta_{\left\{\frac{i}{2^r}\right\}},$$

is D-optimal, with corresponding information matrix $\mathbf{M}(\xi_D) = \frac{1}{2^r+1}\mathbf{I}$, by verifying that $\max_{x\in[0,1]} d(x, \xi_D) = 2^r + 1$—this is the necessary and sufficient condition for optimality, from the celebrated Kiefer and Wolfowitz (1960) theorem.

In Theorems 1 and 2 we also invoke the Kiefer–Wolfowitz Theorem, by which a design ξ_0 is A- or I-optimal iff, for any design ξ_1, we have

$$tr(\mathbf{M}^{-1}(\xi_0)\mathbf{H}) - tr(\mathbf{M}^{-1}(\xi_0)\mathbf{M}(\xi_1)\mathbf{M}^{-1}(\xi_0)\mathbf{H}) \geq 0, \tag{5}$$

where $\mathbf{H} = \mathbf{I}$ in the case of A-optimality and $\mathbf{H} = \int_0^1 \mathbf{f}(x)\mathbf{f}^T(x)dx$ in the case of I-optimality. See, e.g., Pukelsheim (1993).

Theorem 1. *The design ξ_D is also A- and E-optimal for the regression model (3), in which the regressors are defined by (4) with $\phi(x)$ given by (1).*

Proof. For A-optimality, with $\xi_0 = \xi_D$, $\mathbf{M}(\xi_0) = \frac{1}{2^r+1}\mathbf{I}_{2^r+1}$ and $\mathbf{H} = \mathbf{I}_{2^r+1}$ in (5) we obtain the condition $tr(\mathbf{M}(\xi_1)) \leq 1$, verified as follows:

$$\begin{aligned} tr(\mathbf{M}(\xi_1)) &= \int_0^1 \mathbf{f}^T(x)\mathbf{f}(x)d\xi_1 = \sum_{k=0}^{2^r} \int_{\frac{k}{2^r}}^{\frac{k+1}{2^r}} [(2^r x - k)^2 + (1 + k - 2^r x)^2]d\xi_1 \\ &\leq (2^r+1)^{-2} tr(\mathbf{M}^{-1}(\xi_0)) = 1. \end{aligned}$$

For the E-optimality, since $\lambda_{\min}(\mathbf{M}(\xi_D)) = 1/(2^r+1)$, with the corresponding eigenvectors being the columns $\{e_k\}_{k=0}^{2^r}$ of $\mathbf{I}_{2^r+1}$, we have the following. With $w_k = 1/(2^r+1)$,

$$\mathbf{f}^T(x)(e_0, \ldots, e_{2^r})\text{diag}(w_0, \ldots, w_{2^r})(e_0, \ldots, e_{2^r})^T\mathbf{f}(x) = \frac{1}{2^r+1}\sum_{k=0}^{2^r} f_k^2(x) \leq \frac{1}{2^r+1},$$

for all $x \in [0, 1]$, with equality at the support points of ξ_D. Thus, by Theorem 2.1 in Dette and Grigoriev (2014), ξ_D is E-optimal. ■

We comment that A-optimality can also be proved by using Theorem 2.11.1 of Fedorov (1972). A consequence of the following theorem is that ξ_D is not I-optimal.

Theorem 2. *The I-optimal design for the linear spline wavelet regression is*

$$\xi_I = a(\delta_{\{0\}} + \delta_{\{1\}}) + b\sum_{k=1}^{2^r-1} \delta_{\left\{\frac{k}{2^r}\right\}},$$

where $a = 1/(\sqrt{2}(2^r - 1 + \sqrt{2}))$ and $b = \sqrt{2}a$.

Proof. We proceed as in the proof of Theorem 1, but with $\xi_0 = \xi_I$ and

$$\mathbf{H} = \int_0^1 \mathbf{f}(x)\mathbf{f}^T(x)dx = \frac{1}{2^r 6}\begin{pmatrix} 2\,1\,0\,0 & \cdots & 0\,0\,0 \\ 1\,4\,1\,0 & \cdots & 0\,0\,0 \\ 0\,1\,4\,1 & \cdots & 0\,0\,0 \\ \cdots & & \cdots \\ 0\,0\,0\,0 & \cdots & 1\,4\,1 \\ 0\,0\,0\,0 & \cdots & 0\,1\,2 \end{pmatrix},$$

so that $\mathbf{M}(\xi_0) = \text{diag}(a, b, \ldots, b, a)$. Then

$$tr\left(\mathbf{M}^{-1}(\xi_0)\mathbf{H}\right) = \frac{1}{2^r 6}\left[\frac{2}{a} + \frac{4}{b}(2^r - 1) + \frac{2}{a}\right] = \frac{2(2^r - 1 + \sqrt{2})^2}{2^r 3},$$

and

$$tr\left(\mathbf{M}^{-1}(\xi_0)\mathbf{M}(\xi_1)\mathbf{M}^{-1}(\xi_0)\mathbf{H}\right) = \int_0^1 \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi_0)\mathbf{H}\mathbf{M}^{-1}(\xi_0)\mathbf{f}(x)d\xi_1$$

$$= \frac{1}{2^r 6}\int_0^1 \Bigg(\frac{2}{a^2}(1-2^r x)^2 I_{\left[0,\frac{1}{2^r}\right]}(x) + \frac{4}{b^2}\sum_{k=1}^{2^r-1}\Big[(2^r x - k + 1)^2 I_{\left[\frac{k-1}{2^r},\frac{k}{2^r}\right]}(x)$$

$$+ (k+1-2^r x)^2 I_{\left[\frac{k}{2^r},\frac{k+1}{2^r}\right]}(x)\Big] + \frac{2}{a^2}(2^r x - 2^r + 1)^2 I_{\left[\frac{2^r-1}{2^r},1\right]}(x)$$

$$+ \frac{2}{ab}(1-2^r x)(2^r x) I_{\left[0,\frac{1}{2^r}\right]}(x) + \frac{2}{b^2}\sum_{k=1}^{2^r-1}(k+1-2^r x)(2^r x - k) I_{\left[\frac{k}{2^r},\frac{k+1}{2^r}\right]}(x)$$

$$+ \frac{2}{ab}(2^r - 2^r x)(2^r x - 2^r + 1) I_{\left[\frac{2^r-1}{2^r},1\right]}(x)\Bigg) d\xi_1$$

$$\le \frac{1}{2^r 6}\frac{2}{a^2} = \frac{2(2^r - 1 + \sqrt{2})^2}{2^r 3};$$

now (5) follows. ■

We have obtained the c-optimal designs under the condition that the signs of the elements of $\mathbf{c}$ be either constant or alternating. Define

$$S_1 = \{\mathbf{c}: \ all\ c_k \ge 0,\ or\ all\ c_k \le 0\},$$
$$S_2 = \{\mathbf{c}: \ all\ (-1)^k c_k \ge 0,\ or\ all\ (-1)^k c_k \le 0\}.$$

The proof of the following theorem depends on Elfving's Theorem (Elfving, 1952) and Lemma 2.1 of Studden (1968). For the readers' convenience, we state these two theorems in the Appendix.

Theorem 3. *For any* $\mathbf{c} \in S_1 \cup S_2$ *for which* $\mathbf{c}^T\boldsymbol{\theta}$ *is estimable, the design* ξ_c*, with support points* $\{2^{-r}k\}_{k=0}^{2^r}$ *and corresponding weights*

$$w_k = \frac{|c_k|}{\sum_{l=0}^{2^r}|c_l|}, \quad k = 0, 1, \ldots, 2^r,$$

is c*-optimal.*

Proof. For any vector $\mathbf{c}$, we have $\mathbf{c} = \sum_{k=0}^{2^r} c_k \mathbf{f}(k/2^r) = \sum_{k=0}^{2^r} sign(c_k)|c_k|\mathbf{f}(k/2^r)$. Let $\beta = (\sum_{l=0}^{2^r}|c_l|)^{-1}$, then $\beta\mathbf{c} = \sum_{k=0}^{2^r} sign(c_k) w_k \mathbf{f}(k/2^r)$, which implies that $\beta\mathbf{c} \in V$, defined in Elfving's Theorem in the Appendix. Here, we set $sign(c_k) = 1$ or -1 if $c_k = 0$.

If $\mathbf{c} \in S_2$, without loss of generality, we assume $(-1)^k sign(c_k) = 1$, $k = 0, 1, \ldots, 2^r$. Define a function $e(x) := \sum_{k=0}^{2^r}(-1)^k f_k(x)$, then $e(2^{-r}k) = (-1)^k$, $k = 0, 1, \ldots, 2^r$ and $|e(x)| \le 1$. By applying $e(x)$ to Lemma 2.1 of Studden (1968), we can show that $\beta\mathbf{c}$ is a boundary point of V. If $\mathbf{c} \in S_1$, by applying the function $U(x) := \sum_{k=0}^{2^r} f_k(x) \equiv 1$ to the same lemma, we show that $\beta\mathbf{c}$ is a boundary point of V. Thus, by Elfving's Theorem, $\xi_{\mathbf{c}}$ is c-optimal if $\mathbf{c} \in S_1 \cup S_2$. ■

As examples, if $\mathbf{c} = (1, 1, \ldots, 1)^T$, then $\mathbf{c} \in S_1$, and $\xi_{\mathbf{c}}$ coincides with ξ_D. If we want to compare the average of odd-order coefficients and the average of even-order coefficients, that is, $\mathbf{c} = \left(-\frac{1}{2^{r-1}+1}, \frac{1}{2^{r-1}}, -\frac{1}{2^{r-1}+1}, \ldots, \frac{1}{2^{r-1}}, -\frac{1}{2^{r-1}+1}\right)^T$, then $\mathbf{c} \in S_2$, and $\xi_{\mathbf{c}}$ has weight $\frac{1}{2^r+2}$ on design points $\frac{2k}{2^r}$, where $k = 0, 1, \ldots, 2^{r-1}$, and weight $\frac{1}{2^r}$ on design points $\frac{2k+1}{2^r}$, where $k = 0, 1, \ldots, 2^{r-1} - 1$. Note that, for $\mathbf{c}_t = \mathbf{f}(t)$, $t \in [2^{-r}k, 2^{-r}(k+1)]$, the optimal interpolation design $\xi_{\mathbf{c}_t}$ has only two support points $2^{-r}k$ and $2^{-r}(k+1)$, with weights $f_k(t)/(f_k(t) + f_{k+1}(t))$ and $f_{k+1}(t)/(f_k(t) + f_{k+1}(t))$ separately. This is a singular design. For $\mathbf{c} \notin S_1 \cup S_2$, we have not found c-optimal designs, partly because for such $\mathbf{c}$, the union of non-empty domains of the regressors corresponding to one sign (+ or −) is only a subset of [0, 1].

3.2. Quadratic spline wavelet regression model

When using the quadratic spline wavelet (2) to define the wavelet regression model (3) of order r, the number of regressors is, from Lemma 1(iii), $p_r = 3 - 0 + 2^r - 1 = 2^r + 2$. The model is more complicated than the linear spline model in that each regressor is quadratic, and in the linear spline model there is only one regressor having a non-zero value (actually maximum value) at each of the points $\{k/2^r\}_{k=0}^{2^r}$. This makes it difficult to construct the optimal designs. We only have analytic results for $r \in \{0, 1\}$; in Section 4 we discuss numerical approaches.

Theorem 4. *For the wavelet regression model (3) of order r, with regressors defined by the father wavelet (2), we have*

(i) *when* $r = 0$, *the mean response has three parameters and the D-optimal design* ξ_D *puts equal weight on* $0, \frac{1}{2}, 1$;

(ii) *when* $r = 1$, *the mean response has four parameters and the D-optimal design* ξ_D *puts equal weight on* $0, \frac{9-\sqrt{17}}{16}, \frac{7+\sqrt{17}}{16}, 1$.

Proof. We give the proof of (ii); that of (i) – which is somewhat well-known – is similar but simpler. After some algebra, we find that the information matrix $\mathbf{M}(\xi_D)$ is of the form $(1/4096)\left(\mathbf{M}_1 + \sqrt{17}\mathbf{M}_2\right)$, with

$$\mathbf{M}_1 = \begin{pmatrix} 561 & 454 & 297 & 0 \\ * & 3269 & -1236 & 297 \\ * & * & 3269 & 454 \\ * & * & * & 561 \end{pmatrix}, \qquad \mathbf{M}_2 = \begin{pmatrix} -9 & 42 & -65 & 0 \\ * & -381 & 436 & -65 \\ * & * & -381 & 42 \\ * & * & * & -9 \end{pmatrix}.$$

The variance function $d(x, \xi_D) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi_D)\mathbf{f}(x)$ is

$$\begin{aligned} d(x, \xi_D) &= \left(ax^4 + bx^3 + cx^2 + dx + e\right) I_{[0,1/2)}(x) \\ &\quad + \left(fx^4 + gx^3 + hx^2 + ix + j\right) I_{[1/2,1]}(x), \end{aligned}$$

in which the coefficients are

Coefficient	Value	Coefficient	Value
a	$125\sqrt{17}/2 + 965/2$	f	$125\sqrt{17}/2 + 965/2$
b	$-\left(55\sqrt{17} + 655\right)$	g	$-\left(195\sqrt{17} + 1275\right)$
c	$10\sqrt{17} + 290$	h	$220\sqrt{17} + 1220$
d	-40	i	$-\left(105\sqrt{17} + 505\right)$
e	4	j	$35\sqrt{17}/2 + 163/2$

Now, it can be shown that in $(0, 1)$, the function $d(x, \xi_D)$ has the stationary points $\left\{\frac{41-\sqrt{17}}{416}, \frac{9-\sqrt{17}}{16}, \frac{1}{2}, \frac{7+\sqrt{17}}{16}, \frac{375+\sqrt{17}}{416}\right\}$. Furthermore, by checking the second derivative of $d(x, \xi_D)$, we obtain the local minima at $\frac{41-\sqrt{17}}{416}$, $\frac{1}{2}$, and $\frac{375+\sqrt{17}}{416}$, and local maxima at $\frac{9-\sqrt{17}}{16}$ and $\frac{7+\sqrt{17}}{16}$. These and the fact that $d(0, \xi_D) = d\left(\frac{9-\sqrt{17}}{16}, \xi_D\right) = d\left(\frac{7+\sqrt{17}}{16}, \xi_D\right) = d(1, \xi_D) = 4$ imply that $d(x, \xi_D) \le 4, x \in [0, 1]$, with equality being held at the support points of ξ_D. By the Kiefer–Wolfowitz Theorem, we conclude that ξ_D is a D-optimal design for the quadratic spline wavelet regression model of order $r = 1$. This completes the proof. ∎

Remark. Optimal design for spline regression has been considered by Kaishev (1989) and Heiligers (1999), and part (ii) of Theorem 4 can also be derived from the Lemma of Kaishev (1989, p. 43). We present it as above partly to motivate our numerical treatment, for $r > 1$, in the next section.

4. Numerical construction of quadratic spline wavelet designs

We have derived and implemented algorithms for the construction of D- and I-optimal designs for quadratic spline wavelets. For this, we first discretize the design space as $\chi = \left\{x_i = \frac{i-1}{N-1}; i = 1, \dots, N\right\}$, with $N = 1001$ for a close approximation to a continuous interval.

Let $\mathbf{F}_{N\times p_r}$ have rows $\left\{\mathbf{f}^T(x_i); i = 1, \dots, N\right\}$ and for a design ξ let $\mathbf{D}(\xi)$ be the diagonal matrix with diagonal $\{\xi(x_1), \dots, \xi(x_N)\}$. It is convenient to work in a canonical form. Let $\mathbf{Q}_{N\times p_r}$ be such that its columns form an orthogonal basis for the column space of $\mathbf{F}$—this is 'Q' in the QR-decomposition of $\mathbf{F}$. Define $\mathbf{R}(\xi) = \mathbf{Q}^T\mathbf{D}(\xi)\mathbf{Q}$. Then the D- and (scaled) I-criteria to be minimized are, respectively,

$$\begin{aligned} \mathcal{D}_\xi &= \left|\mathbf{M}^{-1}(\xi)\right|^{1/p_r} = \left|\mathbf{R}^{-1}(\xi)\right|^{1/p_r}, \\ \mathcal{I}_\xi &= tr\mathbf{F}\mathbf{M}^{-1}(\xi)\mathbf{F}^T/p_r = tr\mathbf{R}^{-1}(\xi)/p_r. \end{aligned}$$

Denote by $\left\{\mathbf{q}_i^T; i = 1, \dots, N\right\}$ the rows of $\mathbf{Q}$. As in the Kiefer–Wolfowitz Theorem, the necessary and sufficient conditions for the optimality of a design $\boldsymbol{\xi}^0$ are

D-criterion: $\mathbf{q}_i^T\mathbf{R}^{-1}\left(\xi^0\right)\mathbf{q}_i - p_r \le 0$ for all $i = 1, \dots, N$, with equality if $\xi_i^0 > 0$;

I-criterion: $\mathbf{q}_i^T\mathbf{R}^{-2}\left(\xi^0\right)\mathbf{q}_i - tr\mathbf{R}^{-1}\left(\xi^0\right) \le 0$ for all $i = 1, \dots, N$, with equality if $\xi_i^0 > 0$.

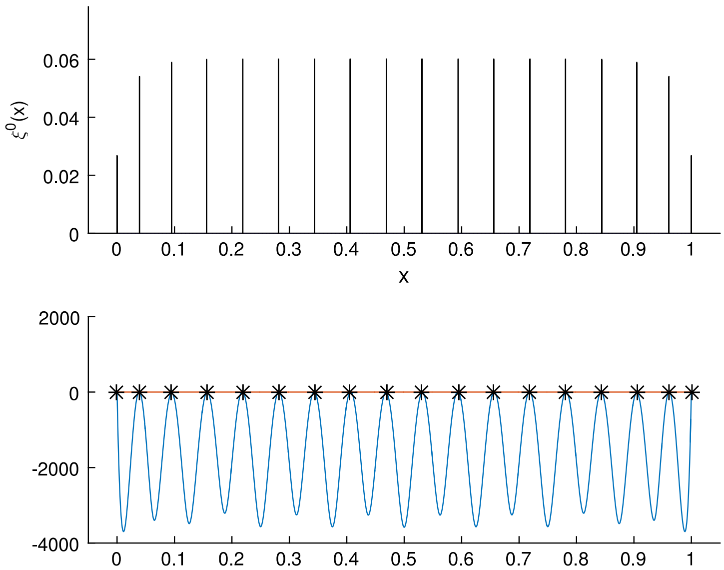

Fig. 1. Top: I-optimal design for the quadratic spline wavelet model of order $r = 4$; $p_r = 18$. Bottom: I-criterion $\mathbf{q}_i^T \mathbf{R}^{-2}(\xi^0)\mathbf{q}_i - tr\mathbf{R}^{-1}(\xi^0)$. Asterisks are at the p_r design points.

Table 1
D- and I-optimal quadratic spline wavelet designs ξ_D and ξ_I.

	$x_i \in [0, .5)$	$\xi_D(x_i) = \xi_D(1 - x_i)$	$x_i \in [0, .5)$	$\xi_I(x_i) = \xi_I(1 - x_i)$
$r = 1$	{0, .305}	1/4	{0, .300}	{.1638, .3362}
$r = 2$	{0, .158, .381}	1/6	{0, .155, .380}	{.0943, .1944, .2113}
$r = 3$	{0, .079, .191,	1/10	{0, .078, .190,	{.0509, .1043, .1136,
	.313, .438}	...	.313, .438}	.1155, .1157}
$r = 4$	{0, .040, .096,	1/18	{0, .039, .095,	{.0267, .0540, .0589,
	.157, .219, .281,	...	.156, .219, .281,	.0599, .0600, .0601,
	.344, .406, .469}	...	.344, .406, .469}	.0601, .0601, .0601}
$r = 5$	{0, .020, .048,	1/34	{0, .019, .047,	{.0139, .0276, .0300
	.078, .109, .141,	...	.078, .109, .141,	.0305, .0306, .0306,
	.172, .203, .234,	...	.172, .203, .234,	.0306, .0306, .0306,
	.266, .297, .328,	...	.266, .297, .328,	.0306, .0306, .0306,
	.359, .391, .422,	...	.359, .391, .422,	.0306, .0305, .0300,
	.453, .484}	...	.453, .484}	.0276, .0139}

Details are in Wiens (submitted for publication), where the theory and algorithms described here are extended to include the case of possibly misspecified regression responses.

The algorithms are sequential, and in the current situation closely parallel that of Wynn (1970). At each step a design point is added to the current design, in order to effect the greatest possible decrease in $\mathcal{D}_\xi$ or $\mathcal{I}_\xi$. Upon convergence the D-criterion or I-criterion is checked. See Fig. 1, where the checking of the I-criterion is illustrated in the case $r = 4, p_r = 18$.

The quadratic spline wavelet designs, for $r \leq 5$, are detailed in Table 1. From there it is seen that as r increases the D- and I-optimal designs become increasingly similar to each other—except at $x = 0, 1$, where the I-optimal design places about half of much mass as at its other support points. As expected, the designs are symmetric about $x = .5$.

5. Case study

Brinkman (1981) carried out an experiment to investigate how the variable NOx (the concentration of the sum of nitric oxide and nitrogen dioxide in engine exhaust) is related to the variable E (an equivalence ratio as a measure of the richness of the air/ethanol mix). There were 88 measurements in the dataset. We have translated and scaled E to $x \in \chi = [0, 1]$, a transformation justified by the invariance of our procedures.

To illustrate our methods we smoothed the Brinkman data with a cubic spline, and used this to predict a mean response function $m(x)$ at $N = 1001$ equally spaced points spanning χ. We then simulated data at all points in χ by adding i.i.d. Normal errors, with mean 0 and standard deviation .2, to these 'true' mean responses $m(x)$. All designs used these data, so as to eliminate design-specific random error. See Fig. 2 for plots of $m(x)$ and fitted response $\hat{m}(x)$ computed by least squares from the Brinkman design on 83 points (some of which were replicated) and a quadratic wavelet model of order $r = 3$. This order was chosen from $1 \leq r \leq 5$ by AICc (Akaike's Information Criterion with the finite sample

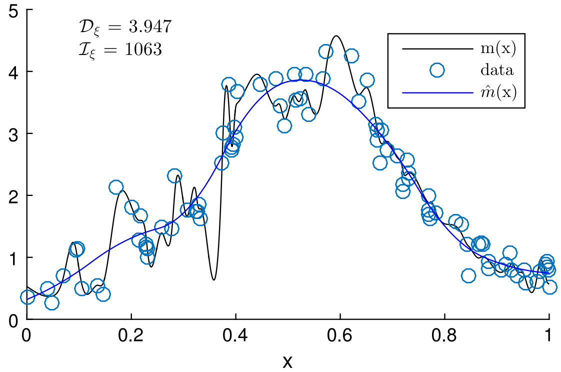

Fig. 2. Response $m(x)$ created by smoothing the Brinkman (1981) data; simulated data at Brinkman design points; fitted quadratic wavelet ($r = 3$) response $\hat{m}(x)$.

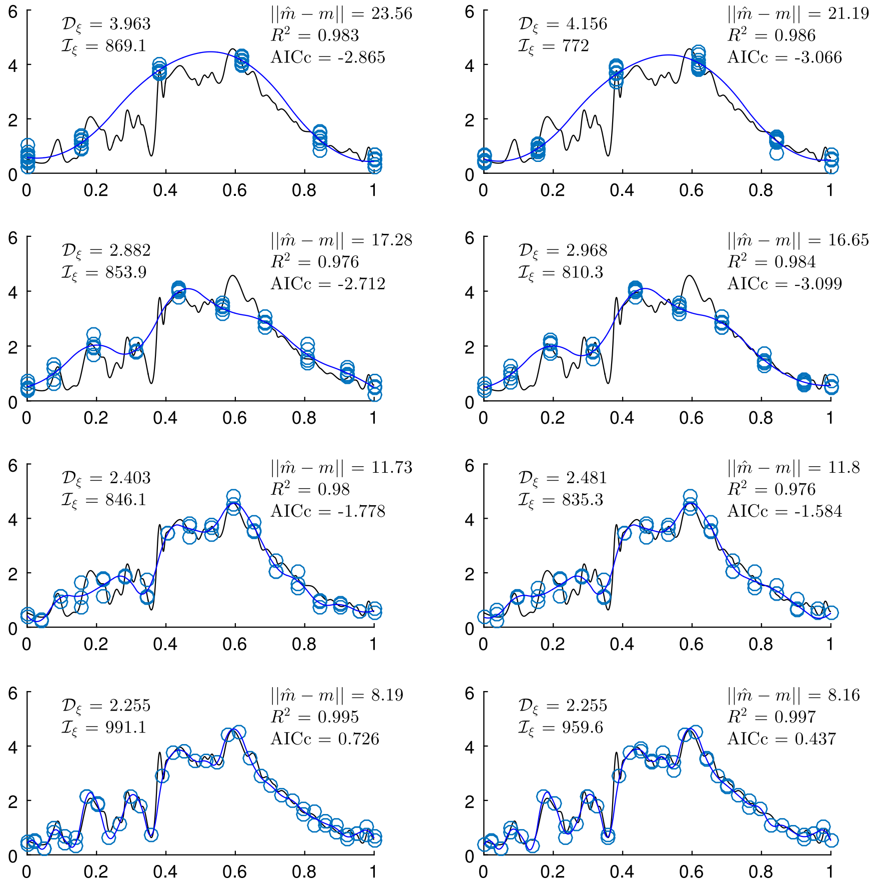

Fig. 3. Implementations of D- (left panel) and I- (right panel) optimal designs for quadratic spline wavelet models of orders $r = 2$ (top) to $r = 5$ (bottom). Legend as in Fig. 2.

correction). Rafajłowicz and Schwabe (2003) employ a similar residual-based criterion in their study of equally spaced designs in nonparametric regression.

We have constructed and implemented a variety of D- and I-optimal designs for quadratic spline wavelets, using each of $r = 1, \ldots, 5$ and simulated data as described above. In all cases we used a design size of $n = 50$; this called for some replications and so the frequencies used were determined by applying the 'efficient design apportionment' method of Pukelsheim and Rieder (1992). See Fig. 3. We have not plotted the results for $r = 1$ since the fits were so poor. The AICc criterion picks $r = 2$ for the D-optimal design and $r = 3$ for the I-optimal design. In each case a substantially better fit is

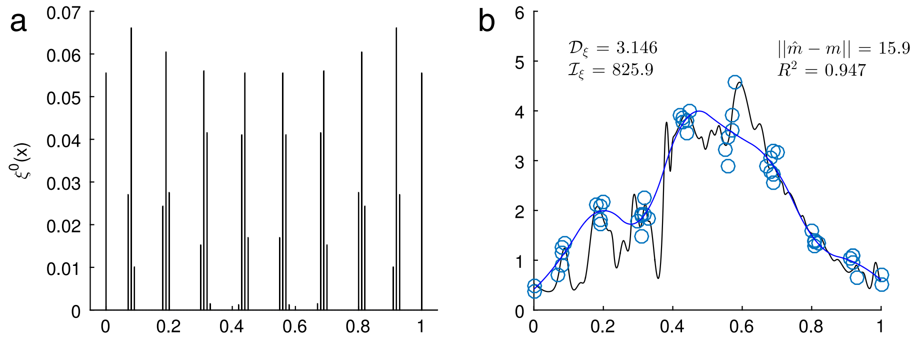

Fig. 4. (a) I-robust design for order $r = 3$ quadratic spline wavelet models. (b) Implementation of this design as with those of Fig. 1.

obtained by using larger values of r. This is borne out by the values of

$$\|\hat{m} - m\| = \left(\sum_{x\in\chi} \left(\hat{m}(x) - m(x)\right)^2\right)^{1/2};$$

these would of course not be available to the experimenter.

5.1. Robustness considerations

As in Section 2, the wavelets which form the regressors constitute only the first few terms of a series expansion converging to a given response function in $\mathcal{L}^2[0,1]$. A natural problem is then that of designing in such a way as to control a measure of the mean squared error, rather than merely the variation, in the face of biases arising from fitting an insufficient collection of wavelets. We address this problem in the following way. We write the mean response as $E[Y|x] = \mathbf{f}^T(x)\boldsymbol{\theta} + \psi(x)$, where $\mathbf{f}$ contains the wavelets employed in the experimenter's intended model, $\boldsymbol{\theta} := \arg\min_{\boldsymbol{\eta}} \int_\chi \left(E[Y|x] - \mathbf{f}^T(x)\boldsymbol{\eta}\right)^2 dx$, and $\psi(x) := E[Y|x] - \mathbf{f}^T(x)\boldsymbol{\theta}$ is 'model error', orthogonal to $\mathbf{f}(x)$ and bounded in $\mathcal{L}^2(\chi)$. Thus ψ is representable as a series in those wavelets not being fitted. For a constant $\tau \geq 0$ we impose a bound of $\tau/\sqrt{n}$ on the $\mathcal{L}^2$ norm of ψ; this ensures that the contribution of (squared) bias to the mean squared error drops at the same rate as that of variation. We adopt a *minimax* approach, and define the D-robust (respectively, I-robust) design problem as that of finding a design so as to minimize, respectively,

$$\mathcal{D}_{\max}(\boldsymbol{\xi}) = \max_{\psi}\left(\det E\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right]\right)^{1/p},$$

$$\mathcal{I}_{\max}(\boldsymbol{\xi}) = \max_{\psi}\int_\chi E\left[\left(\mathbf{f}^T(x)\,\hat{\boldsymbol{\theta}} - E[Y|x]\right)^2\right]dx.$$

See Wiens (2015) for the general theory of such problems in model robustness of design, and Oyet and Wiens (2000) for solutions in *multiwavelet* models.

Here we apply the algorithms developed in Wiens (submitted for publication), employed above in the D- and I-optimality problems and leading to the designs of Fig. 3, but extended to D- and I-robustness. We again discretize the design space and as an example let $\mathbf{f}(x)$ contain the quadratic spline wavelets of order $r = 3$. The solutions turn out to depend on τ and σ only through their ratio, which in this case we take to be 1. Then the I-robust design is pictured in Fig. 4(a); compare with the I-optimal design in Fig. 1.

A common feature, illustrated here, is that the robust designs can roughly be described as optimal designs in which replicates are replaced by clusters of nearby but distinct points. The exact implementation of the I-robust design, with $n = 50$ and applied to the same data as used in Fig. 3, is in Fig. 4(b). Compare with the I-optimal design for $r = 3$ from Fig. 3 ($\mathcal{D}_\xi = 2.968$, $\mathcal{I}_\xi = 810.3$, $\|\hat{m} - m\| = 16.65$). Again we see very common features of robust solutions—the I-robust design results in smaller biases, as measured by $\|\hat{m} - m\|$, in the face of model misspecification, at the cost of somewhat larger measures of variation when the fitted model is exactly correct.

Acknowledgments

The research of Zhide Fang is supported by Grant 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center of Pennington Biomedical Research Center. That of Douglas Wiens is supported by the Natural Sciences and Engineering Research Council of Canada.

We are grateful to two anonymous referees, whose insightful comments have greatly improved this work.

Appendix

Proof of Lemma 1. (i) Let $v \in V_0[0,1]$, so that $v(x) = v_0(x)I_{[0,1]}(x)$ for a function $v_0 \in V_0$. Then v_0 may be written as $v_0(x) = \sum_{k\in\mathbb{Z}} \alpha_k\gamma(x-k)$. Since $k \notin \mathcal{K}_0 \Rightarrow \gamma(x-k) = 0$ *a.e.* $x \in [0,1]$, we have $v(x) = \sum_{k\in\mathcal{K}_0} \alpha_k\gamma(x-k)I_{[0,1]}(x)$. Thus the independent set $\left\{\gamma(x-k)I_{[0,1]}(x) \mid k \in \mathcal{K}_0\right\}$ spans $V_0[0,1]$.

(ii) Let $v \in V_r[0,1]$, and write $v(x) = \sum_{i=0}^{2^r-1} I_{[\frac{i}{2^r},\frac{i+1}{2^r}]}(x)v_r(x)$ for some $v_r \in V_r$. Then since $v_r(2^{-r}x) \in V_0$ we have

$$v(2^{-r}x) = \sum_{i=0}^{2^r-1} I_{[i,i+1]}(x)\sum_{j\in\mathbb{Z}}\alpha_j\gamma(x-j) = \sum_{i=0}^{2^r-1} f_i(x-i),$$

where $f_i(t) = \sum_{j\in\mathbb{Z}} \alpha_{i+j}\gamma(t-j)I_{[0,1]}(t) \in V_0[0,1]$. By (i) of this lemma we have $f_i(t) = \sum_{j\in\mathcal{K}_0}\alpha_{i+j}\gamma(t-j)I_{[0,1]}(t)$ and so

$$\begin{aligned} v(x) &= \sum_{i=0}^{2^r-1}\sum_{j\in\mathcal{K}_0}\alpha_{i+j}\gamma(2^rx-(i+j))I_{\left[\frac{i}{2^r},\frac{i+1}{2^r}\right]}(x) \\ &= \sum_{k\in\mathcal{K}_r}\alpha_k\gamma(2^rx-k)\sum_{i=0}^{2^r-1} I_{\left[\frac{i}{2^r},\frac{i+1}{2^r}\right]}(x) \\ &= \sum_{k\in\mathcal{K}_r}\alpha_k\gamma(2^rx-k). \end{aligned}$$

Thus the set $\left\{\gamma(2^rx-k)I_{[0,1]}(x) \mid k \in \mathcal{K}_r\right\}$ spans $V_r[0,1]$. To see that the elements are linearly independent, suppose that

$$\sum_{k\in\mathcal{K}_r}\alpha_k\gamma(2^rx-k) \equiv 0 \quad \text{for } x \in [0,1]. \tag{A.1}$$

For fixed $i = 0,\ldots,2^r-1$ let $x \in \left[\frac{i}{2^r},\frac{i+1}{2^r}\right]$ and put $y = 2^rx - i \in [0,1]$. Then with $j = k-i$, (A.1) gives $\sum_{j\in\mathcal{K}_r-i}\alpha_{j+i}\gamma(y-j) \equiv 0$ for $y \in [0,1]$. Since $\mathcal{K}_r - i = \{k-i \mid k \in \mathcal{K}_r\} \subset \mathcal{K}_0$, the above identity implies that $\alpha_{j+i} = 0$ for $j \in \mathcal{K}_r - i$. Since i is arbitrary, $\alpha_k = 0$ for all $k \in \mathcal{K}_r$.

(iii) The set $\left\{\phi_{r,k}(x)I_{[0,1]}(x) \mid k \in \mathbb{Z}\right\}$ spans $V_r[0,1]$ and hence contains a basis, necessarily of the same cardinality as $\mathcal{K}_r$. ■

Elfving's theorem. Let $V_+ = \{\mathbf{f}(x) : x \in [0,1]\}$, $V_- = \{-\mathbf{f}(x) : x \in [0,1]\}$ and $V = conv\{V_+ \cup V_-\}$. A design ξ is c-optimal if and only if there exists a measurable function $\eta(x)$ satisfying $|\eta(x)| \equiv 1$ such that (1) $\int_0^1 \eta(x)\mathbf{f}(x)d\xi = \beta\mathbf{c}$ for some constant β; (2) $\beta\mathbf{c}$ is a boundary point of V. Moreover $\beta\mathbf{c}$ lies on the boundary of V if and only if $\beta^{-2} = min_\xi\{\mathbf{c}^T\mathbf{M}^+(\xi)\mathbf{c}\}$.

Lemma 2 (*Of* Studden, 1968). *A vector* $\mathbf{c} \in V$, *which can be put in the form* $\mathbf{c} = \sum_{i=1}^n w_ih_i\mathbf{f}(x_i)$, *with* $|w_i| = 1,\ h_i > 0$ *and* $\sum_{i=1}^n h_i = 1$, *lies on the boundary of* V *if and only if there exists a nontrivial linear form* $u(x) = \sum_{k=0}^{2^r} a_kf_k(x)$ *such that* $|u(x)| \le 1$ *for* $x \in [0,1]$, $w_iu(x_i) = 1,\ i = 1,2,\ldots,n$, *and* $\sum_{k=0}^{2^r} a_kc_k = 1$.

References

Alpert, B.K., 1992. Wavelets and other bases for fast numerical linear algebra. In: Chui, C.K. (Ed.), Wavelets: A Tutorial in Theory and Applications. Academic Press, Boston, pp. 181–216.
Antoniadis, A., Oppenheim, G., 1995. Wavelets and Statistics. Springer-Verlag, New York.
Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs with SAS. Oxford University Press, Oxford.
Brinkman, N.D., 1981. Ethanola single-cylinder engine study of efficiency and exhaust emissions. SAE Trans. 90, 1410–1424.
Chui, C.K., 1992. Wavelets: A Tutorial in Theory and Applications. Academic Press, Boston.
Daubechies, I., 1993. Two recent results in wavelets: Wavelet bases on the interval, and biorthogonal wavelets diagonalizing the derivative operator. In: Schumaker, L.L., Webb, G. (Eds.), Recent Advances in Wavelet Analysis. Academic Press, Boston, pp. 237–258.
Dette, H., Grigoriev, Y., 2014. E-optimal designs for second-order response surface models. Ann. Statist. 42, 1635–1646.
Dette, H., Melas, V., Shpilev, P., 2011. Optimal designs for trigonometric regression models. J. Statist. Plann. Inference 141, 1343–1353.
Dette, H., Studden, W.J., 1997. The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis. Wiley, New York.
Elfving, G., 1952. Optimum allocation in linear regression theory. Ann. Math. Statist. 23, 255–262.
Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press, New York.
Haar, A., 1910. Zur Theorie der orthogonalen Funktionensysteme (Erste Mitteilung). Math. Ann. 69, 331–371.
Heiligers, B., 1999. E-optimal designs for polynomial spline regression. Tatra Mt. Math. Publ. 75, 159–172.
Herzberg, A.M., Traves, W.H., 1994. An optimal experimental design for the haar regression model. Canad. J. Statist. 22, 357–364.
Kaishev, V.K., 1989. Optimal experimental designs for the B-spline regression. Comput. Statist. Data Anal. 8, 39–47.
Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremal problems. Can. J. Math. 12, 363–366.
Oyet, A.J., Wiens, D.P., 2000. Robust designs for wavelet approximations of regression models. J. Nonparametr. 12, 837–859.
Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, New York.
Pukelsheim, F., Rieder, S., 1992. Efficient rounding of approximate designs. Biometrika 79, 763–770.
Rafajłowicz, E., Schwabe, R., 2003. Equidistributed designs in nonparametric regression. Statist. Sinica 13, 129–142.

Strang, G., Nguyen, F., 1996. Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley, MA, USA.
Studden, W.J., 1968. Optimal designs on Tchebycheff points. Ann. Math. Statist. 39, 1435–1447.
Wiens, D.P., 2015. Robustness of Design. In: Handbook of Design and Analysis of Experiments. Chapman & Hall/CRC (Chapter 20).
Wiens, D.P., 2016. I-Robust and D-Robust Designs on a Finite Design Space (submitted for publication).
Wojtaszczyk, P., 1997. A Mathematical Introduction to Wavelets. London Mathematical Society.
Wynn, H.P., 1970. The sequential generation of *D*-optimum experimental designs. Ann. Math. Statist. 5, 1655–1664.
Xie, M.Y., 1998. Some Optimalities of Uniform Designs and Projection Uniform Designs under Multi-Factor Models (Ph.D. dissertation), Hong Kong Baptist University, Department of Mathematics, (unpublished).
Xie, M., Li, Z., 2000. A D-Optimal Design for the Second Order Spline-Wavelet Model (Unpublished Manuscript).