University of Alberta

Library Release Form

Name of Author: Alwell Julius Oyet

Title of Thesis: Robust Designs for Wavelet Approximations of Nonlinear Models

Degree: Doctor of Philosophy

Year this Degree Granted: 1997

(Signed)

Permanent Address:
Department of Statistics
University of Toronto In Mississauga
3359 Mississauga Road North
Mississauga, Ontario L5L 1C6.

Date:

UNIVERSITY OF ALBERTA

Robust Designs for Wavelet Approximations of Nonlinear Models

by

Alwell Julius Oyet

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Statistics

Department of Mathematical Sciences

Edmonton, Alberta

Fall 1997

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Robust Designs for Wavelet Approximations of Nonlinear Models** submitted by **Alwell Julius Oyet** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Statistics**.

Dr. D. P. Wiens (Supervisor)

Dr. R. Karunamuni

Dr. Y. Wu

Dr. S. J. Titus

Dr. S. Portnoy (External Examiner)

Date:

Dedicated

To All Who Make This World Meaningful for Me.

Abstract

We consider the construction of designs for the general nonlinear model. Using multiresolution analysis in wavelet theory, the classical nonlinear design problem is transformed into a robust design problem for 'approximately linear' models with orthonormal wavelet basis on the design space S as regressors.

The minimax approach is used to construct designs which are robust against small departures from the finite wavelet representation of the general nonlinear model. We find that the D-optimal design obtained by Herzberg and Traves (1994) is also G-, Q- and A-optimal (in the classical sense) if the Haar wavelet basis is used in the approximation. We provide a proof which we feel is simpler than that of Herzberg and Traves (1994). On the other hand, if the multiwavelets with $N = 2$ is used, the design which chooses more points in a neighbourhood of the midpoint of the design space and a few at the extremes is shown to be Q- and D-optimal in the simplest case.

Using the nonparametric local averaging procedure with positive but unknown weights, we construct optimal weights and designs under the restriction unbiasedness. We show that under this constraint, the ordinary least squares method is optimal in estimating the parameters of the Haar regression model. In other words, the optimal weight and design obtained were each uniform. For the general $N = 2$ multiwavelet regression model, we show that the optimal weight and design density are concave and convex paraboloids respectively in each of the $2^{(m+1)}$ intervals of the design space $S = [0, 1]$ with maximum points at the midpoints and minimum points at the endpoints of each interval. We also show that the design is symmetric about the midpoint of the design space.

Strategies for implementing the designs are discussed. The question of how well these wavelets approximate nonlinear models is also considered using specific examples.

Acknowledgement

Throughout my studies, I have been helped in many ways, through different means, by everyone - professors, students, office staff, friends, my parents, brothers, sisters and neighbours. So, any acknowledgement will only be a scratch on the surface.

However, my sincere gratitude goes to my supervisor, Professor D. P. Wiens for his immense support, constant encouragement, guidance and friendship. He patiently read through my work several times and offered his very valuable suggestions. I would like to thank Dr. Y. Wu who introduced me to the subject of wavelets.

I wish to acknowledge with thanks the members of my examining committee for their helpful remarks and suggestions.

I would also like to thank the academic staff of the Statistics Centre, Department of Mathematical Sciences for making their time and talents available to me. The staff of the Mathematical Sciences general office were very helpful and nice. I do not fail to remember the prayerful support of the community of St. Paul's (Anglican) Church.

May I put on record, with a grateful heart, the financial support from Professors D. P. Wiens and P. Hooper, my parents and the University of Alberta.

Above all, I thank God for His faithfulness and grace which is sufficient unto me.

Table of Contents

List of Tables

List of Figures

Chapter 1

1. INTRODUCTION

The title of this dissertation indicates that we will focus our attention on constructing designs for nonlinear models using wavelet approximations. Wavelet theory, which has been developing over the years, has proved to be useful in signal processing, fast algorithms for integral transforms in numerical analysis and function representation (see Daubechies (1992), Strang (1989) and Alpert (1992)). This wide applicability has contributed to the growing interest in them.

Let $y(\mathbf{x}_i) \in \Re$ be an observable random variable; $\mathbf{x}_i \in S \subseteq \Re^q$ the ith vector of some control variables and $\varepsilon_i \in \Re$ a sequence of uncorrelated random unobservable errors with mean zero and common variance σ^2. In this work, our discussion will be centered on a model describing the relationship between the response $y(\mathbf{x})$ and the independent variables $\mathbf{x}$ in the following manner :

$$y(\mathbf{x}_i) = \eta(\mathbf{x}_i) + \varepsilon_i, \quad i \; = \; 1, 2, ..., n, \tag{1.1}$$

where $\eta(\mathbf{x}_i)$ is the value of some square integrable, possibly nonlinear function η at $\mathbf{x}_i$. We observe that the precise structure of the nonlinearity present need not be known in order to apply wavelet approximation techniques. The only requirement is the choice of an appropriate wavelet basis and the order m of the approximation. Often, if the form of $\eta(\mathbf{x})$ is assumed known, it is only a convenient approximation based on experience and prior information which assumes that the deviations from the "true" model is negligible. In this case, the assumed form of $\eta(\mathbf{x})$ also depends on some unknown vector of parameters, say $\theta_0 \in \Re^p$. These parameters in general, have some physical meaning which makes them interpretable and of interest in their own right. However, since information provided by the original parameters θ_0 cannot be obtained from the wavelet approximation, the formulation (1.1) provides a convenient framework for our discussion. In Chapter 4, we find that the multiwavelet basis with N = 1 and 2 can be used to approximate several nonlinear relationships. Thus wavelet approximation techniques provide a flexible tool for analysing unknown and approximately known nonlinear functional forms.

The three simplifications that arise from using wavelet expansions of nonlinear models are :

(1) the regressors no longer depend on the unknown parameter $\boldsymbol{\theta}_0$;

(2) the difficult problem of estimating the parameters of a nonlinear model is eliminated; and

(3) the problem is transformed from a nonlinear design problem to a linear one with disturbance function - the well known robust design problem.

Thus, the problem of constructing classical designs for nonlinear models is made equivalent to that of constructing robust designs for linear models where the regressors constitute a system of orthonormal wavelet basis of the design space S. Herzberg and Traves (1994) are probably the first to consider classical designs for wavelet regression models using the Haar wavelets as regressors. The approximation procedure is outlined in Section 1.3 of this chapter.

In what follows, we provide a brief introduction to the theory of classical and robust design and some background on wavelets. More detailed discussions and reviews can be found in Box and Draper (1959), Fedorov (1972), Steinberg and Hunter (1984), Ford, Kitsos and Titterington (1989), Daubechies (1992) and Pukelsheim (1993) amongst others. The classical design problem is defined in Section 1.1 with some examples. We discuss the various approaches in the literature for dealing with linear and nonlinear problems. The assumptions of classical design theory discussed in Section 1.1 leads us naturally into the robust design problem presented in Section 1.2. We review four types of robust designs in the literature. These are:

(1) *robust designs for approximately linear models*;

(2) *designs robust against autocorrelated errors*;

(3) *minimax robust weights and designs for approximately linear models*; *and*

(4) *model robust designs for nonparametric regression models.*

Two approaches, the minimax and infinitesimal approach, are reviewed. In Section 1.3 we provide some background on wavelet theory relevant to our work. The results obtained in our investigation of robust designs for biased wavelet regression models are summarized in Section 1.4.

1.1. The Classical Design Problem

We begin by considering an observable random variable $y \in \Re$ which depends on a design variable $\mathbf{x} \in S \subseteq \Re^q$ through the model (1.1).

An experimenter can take n independent observations on y at the points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, not necessarily distinct, choosen from the set S. Since the set S often consists of more than n-points, the question that arises naturally is - which n-point design, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, should we choose ? Classical design theory was developed in an attempt to answer this question. If the response surface $\eta(\mathbf{x})$ can be written as

$$\eta(\mathbf{x}) = \eta(\mathbf{x}; \boldsymbol{\theta}_0) = \mathbf{q}^T(\mathbf{x})\boldsymbol{\theta}_0 \tag{1.2}$$

the problem is said to be a linear design problem; it is nonlinear otherwise. The classical design problem for regression models with a pre-specified form for $\eta(\mathbf{x})$ has been discussed in great details by several authors. In this case, it is implicitly assumed that the model (1.1) representing $y(\mathbf{x})$ is exactly correct.

Smith (1918) was the first to consider the question of design optimality. Other early contributors include Wald (1943), Hotelling (1944) and Elfving (1952). However, Kiefer (1959) and Kiefer and Wolfowitz (1959) contributed significantly to the area by extending the previous work. The subject of nonlinear experimental design was perhaps first studied by Fisher (1922). White (1973) proved the nonlinear version of the Kiefer-Wolfowitz equivalence theorem.

In order to apply optimal design theory to (1.1) a criterion is required for comparing experiments. For parametric models, this criterion, sometimes called the loss function, is often taken in optimal design theory to be a monotonic increasing scalar valued function $\Phi(M(\hat{\boldsymbol{\theta}}))$ of the mean squared error (MSE) matrix of an estimator of $\boldsymbol{\theta}_0$. If $\hat{\boldsymbol{\theta}}$ is unbiased, the MSE reduces to the Covariance matrix. Mathematically, the classical design problem can be stated as:

$$\min_{\{\mathbf{x}_1 \ldots \mathbf{x}_n \in S\}} \Phi(M(\hat{\boldsymbol{\theta}})).$$

For mathematical convenience, we associate with the n-point design $\mathbf{x}_1$, $\mathbf{x}_2$,..., $\mathbf{x}_n$ a discrete design measure $\xi(\mathbf{x})$ on S, which places equal mass $\frac{1}{n}$ at each of the design points $\mathbf{x}_i$,

$$\xi(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i} \tag{1.3}$$

where $\delta_{\mathbf{x}}$ denotes the pointmass 1 at $\mathbf{x}$. Then, we rewrite $M(\hat{\boldsymbol{\theta}})$ as $M(\xi(\mathbf{x}))$. This transforms the n-observation design problem to that of finding a discrete probability measure $\xi^*(\mathbf{x})$ which minimizes $\Phi(M(\xi(\mathbf{x})))$. A common approach to this problem is to extend the definition of the MSE matrix to the set of all probability distributions, denoted by $\mathcal{X}$, on S and find $\xi^*(\mathbf{x})$ in $\mathcal{X}$ to minimize $\Phi(M(\xi(\mathbf{x})))$. We then hope that an n-observation design whose associated probability distribution approximates $\xi^*(\mathbf{x})$ will be close to optimal for the n-observation problem. We adopt this viewpoint, often called *approximate design theory*. For robust designs, this implies that admissible designs are necessarily absolutely continuous (see Wiens(1992)).

Some optimality criteria commonly found in the literature are :

(1) D-Optimality - when the loss function is the *determinant*.

(2) A-Optimality - when $\Phi(\cdot) \equiv trace(\cdot)$.

(3) E-Optimality - when $\Phi(\cdot) \equiv \lambda_{max}(\cdot)$ where λ_{max} represents the maximum eigenvalue.

(4) G-Optimality - when

$$\Phi(M(\xi)) = \max_{\mathbf{x}\in S} \mathbf{q}^T(\mathbf{x})M(\xi)\mathbf{q}(\mathbf{x}).$$

(5) Weighted-optimality - when $\Phi(M(\xi)) = tr\{WM^{-1}(\xi)\}$, where W is a known weight.

The Kiefer-Wolfowitz equivalence theorem mentioned earlier provides a link between D- and G-optimality. Chang (1979) constructed weighted optimal designs for a linear model with linearly independent regressors belonging to the reproducing kernel Hilbert space (RKHS) $H(R)$ generated by a known continuous and positive definite covariance reproducing kernel $R(s,t)$ on $S \times S$. For any symmetric positive definite matrix W the optimal design was simultaneously A-, D- and weighted optimal for $S = [0,1]$.

The linear design problem has been studied by several authors using various loss functions and variations of the linear models. Interested readers can refer to Fedorov (1972), Silvey (1980) and Pukelsheim (1983). Designs for nonparametric regression models have also been studied by a number of authors. Chan (1991) employs first order differences to construct designs for estimating variance in nonparametric models with model function $g(t)$ and independent errors having zero mean and constant variance. The estimate considered was of the form $\hat{\sigma}^2 = \frac{Y^T DY}{trD}$ for some symmetric non-negative matrix D. Assuming that g satisfies a uniform Lipschitz condition and $D = (\gamma_{ij})$, a tridiagonal symmetric matrix, the asymptotic variance V was obtained. The uniform design was found to be the minimizer of V if γ_i is a one to one function of $(t_{i+1} - t_i)$ where $\gamma_{ii} = (\gamma_{i-1} + \gamma_i)$, $\gamma_0 = \gamma_n = 0$, $(i = 1, .., n)$ and $\gamma_{i,i+1} = -\gamma_i$, $(i = 1, .., n-1)$. Muller (1984) considered a nonparametric model where the error $Z(t)$ is a sequence of stochastic processes with $E\{Z(t)\} = Cov\{Z(t_1), Z(t2)\} = O(n^{-1})$, $t_1 \neq t_2$, $Var\{Z(t)\} = \sigma^2(t) + O(n^{-1})$ and the response surface $g \in \mathcal{C}^k([0,1])$. Based on the Gasser-Muller estimator $g_{n,\nu}(t)$, $k > \nu$, of the function g, with uniform and non-uniform bandwidths, the asymptotic Integrated Mean Squared Error (IMSE) of the estimate was derived. Under appropriate conditions, the design densities minimizing IMSE in the class $\{f \in \mathcal{C}([0,1]) \mid |f(x) - f(y)| \leq L_f|x-y|^\alpha$, $L_f > 0$ and all x, $y \in [0,1]\}$ were

$$f^*(t) = \frac{\sigma(t)(h(t))^{1/2}}{\int_0^1 \sigma(x)(h(x))^{1/2}dx}; \quad \tilde{f}^*(t) = \frac{\phi(t)}{\int_0^1 \phi(x)dx}$$

for uniform and non-uniform bandwidths respectively, where h is the density of a probability measure H, $\phi(x) = [\sigma(x)^{4(k-\nu)} g^{(k)}(x)^{2(2\nu+1)} h(x)^{2k+1}]^{1/(4k+1-\nu)}$ and

$0 < \alpha \leq 1$. We observe that the optimal design density derived in Chapter 3 of this thesis has the above structure.

Adopting a Bayesian approach Mitchell, Sacks and Ylvisaker (1994) defined three design criteria, A-optimality (for average), G-optimality (for global) and D-optimality (for determinant), for constructing designs when the response is represented by a random function (stochastic process). In general, these criteria do not have the properties which obtain in the classical setting. The main interest of this work was to draw attention to the fact that certain asymptotics produce a class of tractable Bayesian design problems from hard ones. Necessary and sufficient conditions for D- and A-optimality were obtained and several examples considered. They also compared exact designs computed numerically for the asymptotic criteria with exact designs computed for the original nonasymptotic ones, in some simple cases. They found that designs based on the asymptotic criteria were easier to compute and were quite efficient over a wide range of the parameters of the prior process. Discussions on traditional Bayesian design theory, where the prior for the response function is a random finite linear combination of known functions can be found in Pilz (1983), Chaloner (1984) and Bandemer, Nather and Pilz (1987) amongst others. In a related development, Sacks, Welch, Mitchell and Wynn (1989) discuss the design and analysis of computer experiments. One feature of computer experiments is that the output is deterministic. Sacks et al (1989) treat the deterministic output as a realization of a stochastic process $Y(x)$ that includes a linear regression model with error variance $\sigma^2 R$, where R is the matrix of stochastic process correlations. Using the IMSE criterion and a method called Kriging to evaluate the MSE, they constructed sequential designs for a circuit-simulator model on the design space $[0.5, 0.5]$.

Van der Linde (1985) discussed the question of estimating an unknown regression function $g(t)$, $t \in [a, b]$ given a finite number of observations and invited studies on optimal designs based on the global generalized smoothing error $\int_a^b e_s^d(t) d\Omega(t)$ where Ω is a measure on $[a, b]$ and $e_s^d(t)$ is defined by Van der

Linde (1985) to be the generalized smoothing error. Assuming that $g \in H(K)$, a RKHS, the technique was to interpolate g and estimate the interpolating function. The Bayesian approach adopted was justified by interpreting interpolation in RKHS as a Bayesian procedure. Fan (1992) also considered the problem of estimating a nonparametric regression function $g(x)$. Let $f(x)$, the marginal density of the random variable X and $Var(Y|X = x) = \sigma^2(x)$ be independent of $g(x)$. Restricting to the class $\mathcal{C}_2 = \{f(\cdot, \cdot) \mid |g''(y) - g(x) - g'(x)(y - x)| \leq \frac{C}{2}(y - x)^2\}$, and under some regularity conditions, Fan (1992) obtained the best linear smoother to be the local linear regression smoother given by

$$\hat{g}(x) = \frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j}$$

where $w_j = K(\frac{x-x_j}{h_n})[s_{n,2} - (x - x_j)s_{n,1}]$, $s_{n,l} = \sum_1^n K(\frac{x-x_j}{h_n})(x - x_j)^l$, $l = 1, 2$ with $K(x) = \frac{3}{4}[1 - x^2]_+$ and $h_n = \frac{15\sigma^2(x)}{f(x)C^2 n}$. This method is sometimes called the design adaptive regression method because it adapts to various design densities, to both fixed and random designs and to both interior and boundary points.

For nonlinear models, the methods used in the literature so far produce a mean squared error matrix which is a function of the unknown "true" parameter $\boldsymbol{\theta}_0$. The dependence of the MSE on the unknown $\boldsymbol{\theta}_0$ has been a major difficulty in obtaining good optimal designs. The following approaches have been used in the literature to remedy this difficulty.

B1 Parameters are assumed to be close to certain specified values.
Designs obtained using this assumption are called locally optimal designs (Chernoff (1953)).

B2 Use prior estimates either from previous experiments, or from a pilot experiment conducted specially for the purpose, or merely guesses (Box and Lucas (1959)).

B3 Propose some weighting function W($\cdot$) on the parameter space, Ω, which may or may not be a formal prior density and construct

$$\mathbf{M}(\boldsymbol{\xi}) = \int_{\boldsymbol{\Omega}} \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\xi}) W(d\boldsymbol{\theta})$$

or a new criterion

$$\mathbf{\Phi}_W(\boldsymbol{\xi}) = \int_{\mathbf{\Omega}} \mathbf{\Phi}(M(\boldsymbol{\theta}, \boldsymbol{\xi}))W(d\boldsymbol{\theta}).$$

This approach has been criticized by Ford, Kitsos and Titterington (1989). They also emphasize the effect of changes in the prior estimates of $\boldsymbol{\theta}_0$ on the properties of locally optimal designs and static designs obtained by **B2**.

B4 Use a sequential strategy in which the parameter estimates are updated after each trial and the next design point is then choosen with the aid of the improved estimates (Box and Hunter (1965)). Box (1970) introduces a criterion as a guide to the time when it is no longer worth changing points.

Chernoff (1953) studied locally optimal designs using minimization of the trace of the inverse of Fisher's Information matrix as the design criterion. He used the design problem for quantal response data as an example. Box and Lucas (1959) obtained numerically a D-optimal design for a chemical reaction model

$$\eta(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) = \frac{\theta_1}{\theta_1 - \theta_2}[exp(-\theta_2 x_1) - exp(-\theta_1 x_1)] \tag{1.4}$$

by working with a linearized approximation and preliminary guesses $\theta_{10} = 0.7$, $\theta_{20} = 0.2$. The problem was to choose x_{1i}, i=1,2,...,n so as to maximize the determinant of F^TF where F is the matrix of partial derivatives of η with respect to $\boldsymbol{\theta}$. The significance of this criterion derives from the fact that if $\eta(\mathbf{x}, \boldsymbol{\theta})$ is assumed to be approximately linear in the neighbourhood of an initial point $\boldsymbol{\theta}_*$, the asymptotic variance of the estimate $\hat{\boldsymbol{\theta}}$ is proportional to $(F^T\text{F})^{-1}$ and $\det(F^T\text{F})^{-1}$ is proportional to the volume contained within any specific ellipsoidal probability contour for $\boldsymbol{\theta}$ about $\boldsymbol{\theta}_*$ in the space of the parameters. Thus the D-optimality criterion ensures that any such probability contour includes the smallest volume. That is, we minimize the volume of the linear approximation joint-confidence regions for the parameters. The problem was considered for $n = p = 2$ so that the design points were choosen to maximize det(F). Considering (1.4), Atkinson and Hunter (1968) describe a sequential procedure for obtaining the design points using the same criterion. They found in several chemical examples that with n a multiple of p, the optimum plan consist of $\frac{n}{p}$ replications at each of the p optimum sets of levels for the case $n = p$, under certain sufficient conditions on the design space. Box (1968) studied the case when n is not a multiple of p. Haines (1993) show that the replication of the p-point D-optimal designs obtained by Atkinson and Hunter (1968) and Box (1968) as well as the result of Velilla and Llosa (1992) for heteroscedastic nonlinear models follow directly from approximate design theory.

Several authors have applied optimal design theory to nonlinear models arising from clinical trials (Begg and Kalish (1984)), life testing (Maxim et al (1977)) and dynamic systems (Titterington (1980)). The design criterion most studied is D-optimality and the sequential procedure is most favoured.

Examples of Classical Optimal Designs

In this section, we give five examples of classical designs. The first two are based on the simple line fit model. Example 3 is based on the Haar wavelet regression model adapted from the paper by Herzberg and Traves (1994). The last two examples are on polynomial and nonlinear regression. The method of parameter estimation in all the examples in this section is that of least squares.

Example 1: Consider the simple line fit model

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \;\; x_i \in S \; = \; [-1, 1] \;\; i = 1, 2, .., n.$$

Using the least squares estimator of $\boldsymbol{\theta}^T = (\theta_0, \theta_1)$ and applying elementary methods it can be shown that the D-optimal design is:

(a) choose $\frac{n}{2}$ points x_i at each of ± 1 when n is even;

(b) choose $\frac{n}{2}$ points x_i at each of ± 1 and an extra point at either $+1$ or -1 when n is odd; and

(c) choose $\frac{n}{2}$ points x_i at each of ± 1 and an extra point at 0 when n is odd, if in addition we impose the condition $\sum x_i = 0$.

Such designs are said to be exact because they are discrete and implementable.

Example 2: We continue with the model in Example 1 with $x_i \in [-1, 0]$. Define the vectors

$$\mathbf{x}_i^T = (1, x_i) \;\; and \;\; \mathbf{c}^T = (1, 1).$$

Following Pukelsheim (1993) it can be shown that the unique optimal design minimizing the variance of $\mathbf{c}^T\hat{\boldsymbol{\theta}}$ is the design which takes one-third of the observations at $x_1 = -1$ and two-thirds at $x_2 = 0$. That is,

$$\xi(-1) = \frac{1}{3}, \quad \xi(0) = \frac{2}{3}.$$

Example 3 (Wavelet regression): Model (Herzberg and Traves (1994)):

$$E[y|x] = \beta_0 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \beta_{2^j+k}\psi_{j,k}(x) \tag{1.5}$$

where $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $S = [0, 1]$ and

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.6}$$

Under the Haar regression model of order m given by (1.5) the design which places equal mass $2^{-(m+1)}$ in the 2^{m+1} intervals

$$\left[2^{-(m+1)}k, 2^{-(m+1)}(k+1)\right), \quad k = 0, 1, ..., 2^{m+1} - 1$$

is D-optimal. For $m = 2$ and $n = 8$ the design is implemented by taking the eight observations at the points

$$x_i = \frac{2i+1}{16}, \quad i = 0, 1, ..., 7. \tag{1.7}$$

In Chapter 2, we show that the design described above is also Q-, A- and G-optimal. We give a proof which we feel is simpler than that of Herzberg and Traves (1994). We also provide a strategy for implementing the design, in Chapter 4, of which the choice (1.7) is a special case.

Example 4 (Polynomial regression): Model (Hoel (1958)):

$$y_i = \theta_0 + \sum_{j=1}^{d} \theta_j t^j + \varepsilon_i, \quad d \geq 1 \ \text{ and} S = [-1, 1]. \tag{1.8}$$

Hoel (1958) has shown that the unique D-optimal design places equal weight $\frac{1}{d+1}$ at the $(d+1)$ solutions of the equation

$$(1 \ - \ t^2)P_d'(t) \ = \ 0, \tag{1.9}$$

where $P_d'(t)$ is the first derivative of the dth order Legendre polynomial $P_d(t)$.

Example 5 (Nonlinear regression):

Hamilton and Watts (1985) proposed a quadratic design criterion for the strictly nonlinear model of type (1.1) based on a second-order approximation of the volume of the parameter inference region. The approximation is that the volume is proportional to

$$|V^T V|^{-\frac{1}{2}} |C|^{-\frac{1}{2}} (1 + k^2 tr\{C^{-1} M\})$$

where

$$k = \frac{\rho}{\sqrt{2(p+2)}}, \ \rho^2 = ps^2 F(p, \nu; \alpha), \ V = \frac{\partial \eta(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \ C = I - B;$$

α is a fixed significance level and s^2 is an estimate of variance. The matrix B measures the intrinsic curvature of the expectation surface. Under the assumption of no intrinsic curvature, the matrix C becomes the identity matrix. If σ^2 is known ρ^2 becomes $\sigma^2 \chi^2(p; \alpha)$. The matrix M is defined by

$$M = \mathbf{c}\mathbf{c}^T \ + \ H + \ [\mathbf{c}^T][A], \ h_{ij} = tr\{A_i A_j\}, \ \mathbf{c} = (tr\{A_1\}, ..., tr\{A_p\})^T$$

where $A \ = \ L^T[U^T][V_{..}]L$ is a three dimensional array. The matrices U and L are obtained from the decomposition $V \ = \ UL^{-1}$ and satisfies $U^T U = I$ and $LL^T = (V^T V)^{-1}$. The square bracket multiplication (used to define multiplication by a three dimensional array) $[U^T][V_{..}]$ reduces the $n \times p \times p$ array $V_{..}$ with elements

$\frac{\partial\eta}{\partial\theta_i\partial\theta_j}$ to a $p \times p \times p$ array. Hamilton and Watts(1985) proposed to iteratively minimize the approximation as follows:

(a) choose an initial estimate, σ_0^2, for σ^2; select α and set $k_0^2 = \frac{\sigma_0^2\chi^2(p;\alpha)}{2(p+2)}$;

(b) evaluate V_0, C_0 and M_0 using an initial estimate $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$;

(c) iteratively minimize the quadratic approximation to obtain an optimal design.

Using the nonlinear model (1.4) and the procedure described above, they obtained an optimal design under the assumption of no intrinsic curvature equivalent to replacing C by its expected value. They used as initial estimates $\boldsymbol{\theta}_0 = (0.7, 0.2)^T$ and the D-optimal design from Box and Lucas (1959), $\mathbf{x}_D = (1.23, 6.86)^T$. They choose $\sigma_0 = 0.1$, so that $\rho_0 = 0.25$ and $k_0 = 0.0884$. Their iterative procedure led to the optimal design $\mathbf{x}_Q = (1.04, 5.56)^T$.

Their work was motivated by the papers by Bates and Watts (1981) and Cochran (1973). Bates and Watts (1981) suggested choosing the design to minimize the parameter effects curvature. Cochran (1973) invited studies of the small sample performance of the D-optimal design criterion after noting its asymptotic nature.

1.2. Robust Designs

Robust designs became a subject of interest for two major reasons. These are

(i) the model may not be exactly correct; and

(ii) the errors ε_i may not be uncorrelated

as earlier assumed implicitly or explicitly. It is well known that in most cases where the form of $\eta(\mathbf{x})$ is pre-specified, the assumed form is the model builder's best mathematical description of the process under study and often a convenient approximation. We recall that in the nonlinear case, the designs constructed so far have used a linear approximation of $\eta(\mathbf{x}, \boldsymbol{\theta}_0)$ with the hope that the remainder terms are negligible. Under these conditions, the least squares estimator of $\boldsymbol{\theta}_0$ is biased and the classical designs which minimize variance alone are no longer "optimal" due to the bias. Several authors including Box (1971), Cook, Tsai and Wei (1986) and Bates and Watts (1980) have studied the problem of bias in nonlinear regression models and provided approximations.

In the linear case, Box and Draper (1959) outlined the effects of departures from the assumed linear model on the optimal design. They criticized the classical optimality criteria, some of which have been defined in Section 1.1, and argued that a more appropriate optimality criterion is the Integrated Mean Squared Error (IMSE) of the estimate $\hat{\eta}(x)$ of the "true" response surface $\eta(x)$ over the design space S. That is,

$$\mathcal{L} = \frac{n\Omega}{\sigma^2}\int_S E\{[\hat{\eta}(x) - \eta(x)]^2\}dx = ISB \quad + \quad IV \tag{1.10}$$

where Ω, the Integrated Square Bias (ISB) and the Integrated Variance (IV) are defined by

$$\begin{aligned} \Omega^{-1} \quad &= \quad \int_S dx, \quad ISB = \frac{n\Omega}{\sigma^2}\int_S \{E[\hat{\eta}(x)] - \eta(x)\}^2 dx \\ \text{and } IV \quad &= \quad \frac{n\Omega}{\sigma^2}\int_S E\{\hat{\eta}(x) - E[\hat{\eta}(x)]\}^2 dx. \end{aligned}$$

They showed that if the assumed model is the simple linear model when the true model is quadratic, the designs minimizing IMSE were similar to those that minimized the bias component alone, but were quite different from those that minimized the variance component.

Using the minimax approach, Huber (1975) studied the effect of departures from linearity and agreed with Box and Draper (1959) in his conclusion. Huber (1975) observes that deviations from linearity that are too small to be detected are already large enough to tip the balance away from the (classically) optimal designs shown in Example 1.

For nonlinear models no work has been found in the literature that is aimed at studying the effect of departures from the assumed model. This is not surprising since the study of designs for nonlinear models has lagged behind partly due to the inherent difficulty associated with the designs depending on the unknown parameter values. However, White (1981), Steinberg and Hunter (1984) and

Ford, Kitsos and Titterington (1989) have noted the consequences of pretending that a nonlinear regression model is exactly correct. Other related activities in this area include designs that facilitate improvements in nonlinear models by trying to highlight suspected inadequacies or that discriminate between competing models. Studies in this area have been done by Hunter and Reiner (1965), Box and Hill (1967), Hill , Hunter and Wichern (1968) and Atkinson and Fedorov (1975).

Another point mentioned earlier is that of correlated errors. In this case, the covariance matrix of the estimate $\hat{\boldsymbol{\theta}}$ is a function of the unknown correlation matrix of the errors. Therefore, designing to minimize an optimality criterion which is a function of the covariance matrix under the assumption of uncorrelated errors will lead to designs that will not optimize the loss function under correlated errors. Thus the subject of robust designs is aimed at constructing designs which are not sensitive to small departures from the assumptions on which the model is based.

There are two basic approaches in the literature for constructing robust designs. These approaches, the minimax and infinitesimal approach, were both adapted from the theory of robust estimation. Huber (1964, 1975) introduced the minimax approach to robust estimation and design. Hampel (1974) is the first to use the infinitesimal approach for robust estimation. The infinitesimal approach was first applied to robust design theory by Wiens and Zhou (1996c).

Approximately Linear Models

In order to investigate the sensitivity of designs to model misspecification, several authors have studied various versions of a modification to the assumed linear model. The "approximately linear model" is represented as

$$\begin{aligned} E[y|\mathbf{x}] &= \mathbf{q}^T(\mathbf{x})\boldsymbol{\beta}_0 + f(\mathbf{x}) \\ y(\mathbf{x}_i) &= E[y|\mathbf{x}_i] + \varepsilon_i, \quad i = 1, 2, ..., n \end{aligned} \tag{1.11}$$

where

$$\mathbf{q}(\mathbf{x}) \in \mathcal{R}^p, \ \boldsymbol{\beta}_0 \in \mathcal{R}^p, \ \mathbf{x} \in S \subseteq \mathcal{R}^q, \ \mathrm{Var}(\varepsilon_i) \ = \ \sigma^2 \tag{1.12}$$

and $f(\mathbf{x})$ is some unknown contamination term belonging to some class $\mathcal{F}$. Robust minimax designs were constructed by solving the problem

$$\min_{\xi} \max_{f \in \mathcal{F}} \Phi(M(f, \xi)) \tag{1.13}$$

for some loss function $\Phi(\cdot)$, where $M(f, \xi)$ is the MSE of $\hat{\boldsymbol{\beta}}$.

To motivate (1.11), suppose that an experimenter fits the model

$$E[y|\mathbf{x}] = \mathbf{q}^T(\mathbf{x})\boldsymbol{\beta}_0$$

knowing fully well that it is only a convenient approximation. Define

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} \int_S [E[y|\mathbf{x}] - \mathbf{q}^T(\mathbf{x})\boldsymbol{\beta}]^2 d\mathbf{x} \tag{1.14}$$

and set

$$f(\mathbf{x}) = E[y|\mathbf{x}] - \mathbf{q}^T(\mathbf{x})\boldsymbol{\beta}_0. \tag{1.15}$$

Then the equations

$$y(\mathbf{x}) = \mathbf{q}^T(\mathbf{x})\boldsymbol{\beta}_0 + f(\mathbf{x}) + \varepsilon$$

and

$$\int_S \mathbf{q}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0} \tag{1.16}$$

define $\boldsymbol{\beta}_0$ uniquely provided the matrix $\int_S \mathbf{q}(\mathbf{x})\mathbf{q}^T(\mathbf{x})d\mathbf{x}$ is non-singular. To prevent the error due to bias from dominating that due to variance, Huber (1975) and more recently Wiens (1990) placed a bound on the disturbance function $f(\mathbf{x})$ to obtain the condition

$$\int_S f^2(\mathbf{x}) d\mathbf{x} \ \leq \ \tau^2 \tag{1.17}$$

for some small and known number τ. From (1.16) and (1.17) the class

$$\mathcal{F}_1 = \left\{ f(\mathbf{x}) \ \middle| \ \int_S f^2(\mathbf{x}) d\mathbf{x} \ \leq \ \tau^2, \ \int_S \mathbf{q}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0} \right\} \tag{1.18}$$

used by Huber (1975) and Wiens (1990, 1992, 1994, 1996) in their investigation was constructed. The class $\mathcal{F}_1$ has been criticized by Marcus and Sacks (1976) and Li and Notz (1982) as being too large. They claimed that exact designs in this class have infinite maximum loss; a claim which has been proved by Wiens (1992). Therefore, robust designs constructed for deviations in $\mathcal{F}_1$ are continuous designs which are approximated by discrete designs in practice.

A class $\mathcal{F}_2$ which has been used by Marcus and Sacks (1976), Sacks and Ylvisaker (1978) and Pesotchinsky (1982) is

$$\mathcal{F}_2 = \{f(\mathbf{x}) \mid |f(\mathbf{x})| \leq \phi(\mathbf{x}), \text{ for all } \mathbf{x} \in S\} \tag{1.19}$$

where the function $\phi(\mathbf{x})$ is known. This class often leads to designs whose mass is concentrated at a small number of points in the design space, hence have severely limited robustness against realistic departures from the assumed model (see Wiens (1992)).

The class of functions f or its derivatives f' satisfying a uniform Lipschitz condition are also found in the robustness literature:

$$\begin{aligned}
\mathcal{F}_3 &= \mathcal{F}_3(M) = \{f \mid |f(x) - f(y)| \leq M|x-y|, \text{ for all } x, y \in S\} \\
&\supset \mathcal{F}_3^0(M) = \{f \mid |f'(x)| \leq M, \text{ for all } x \in S\} \supset \{\text{constants}\} \\
\mathcal{F}_4 &= \mathcal{F}_4(M) = \{f \mid f'(x) \in \mathcal{F}_3(M)\} \\
&\supset \mathcal{F}_4^0(M) = \left\{f \mid f'(x) \in \mathcal{F}_3^0(M)\right\} \supset \{\text{linear functions}\}.
\end{aligned}$$

Using IMSE loss and $\mathcal{F}_1$, Huber (1975) constructed robust designs for the model (1.11) with $\mathbf{q}^T(\mathbf{x}) = (1, x)$ and $S = [-0.5, 0.5]$. He showed that the optimal design has density

$$m_0(x) = (ax^2 + b)^+ \tag{1.20}$$

with a and b choosen to satisfy

$$\int_S m(x)dx = 1 \text{ and } \int_S x^2 m(x)dx = \gamma. \tag{1.21}$$

The parameter γ is determined by $\nu = \frac{\sigma^2}{n\tau^2}$. As $\nu \to 0$, the loss function is dominated by the ISB and $m(x) \to 1$ - the uniform density on $[-0.5, 0.5]$. On the other hand, the IV dominates as $\nu \to \infty$. The solution then converges to the classical optimal design which places equal mass at each of the points $\pm\frac{1}{2}$.

Wiens (1990) extended the work of Huber (1975) to the case of multiple regression with $\mathbf{q}^T(\mathbf{x}) = (1, \mathbf{x}^T)$, $\mathbf{x}^T = (x_1, x_2, ..., x_p)$ and S the sphere of unit volume

$$S = \left\{ \mathbf{x} \mid \ \|\mathbf{x}\| \ \le r_p \ := \ \frac{\left[\Gamma\left(\frac{p+2}{2}\right)\right]^{\frac{1}{p}}}{\sqrt{\pi}} \right\}. \tag{1.22}$$

He also obtained the density of the optimal design for the case of two interacting regressors, where $\mathbf{q}^T(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$ and $S = [-0.5, 0.5] \times [-0.5, 0.5]$. The minimax design density for multiple regression was shown to be of the form

$$m(\mathbf{x}) = (a\|\mathbf{x}\|^2 + b)^+$$

where a and b are determined so that $m(\mathbf{x})$ is a density and $\int_S \|\mathbf{x}\|^2 m(\mathbf{x}) d\mathbf{x} = p\gamma$. For the range of values of the ratio $\frac{\gamma}{\gamma_0}$, where $\gamma_0 = \int_S x_1^2 d\mathbf{x} = \frac{r_p^2}{p+2}$, the values of the constants a and b as well as the form of the least favourable disturbance function were obtained. As $\frac{\gamma}{\gamma_0} \to 1$ the uniform design becomes minimax corresponding to $n \to \infty$ or $\nu \to 0$. In the other directions, as $\gamma \to \frac{\gamma_0(p+2)}{p} = \frac{r_p^2}{p}$ the optimal design places pointmass 1 at $\|\mathbf{x}\| = r_p$.

The density of the optimal design for the model with two interacting regressors was shown to be

$$m_0(x_1, x_2) = (\lambda + \mu(x_1^2 + x_2^2) + \delta x_1^2 x_2^2)^+$$

where the multipliers are determined to satisfy

$$\int_S m_0(\mathbf{x}) d\mathbf{x} = 1, \ \int_S x_1^2 m_0(\mathbf{x}) d\mathbf{x} = \gamma \text{ and } \int_S x_1^2 x_2^2 m_0(\mathbf{x}) d\mathbf{x} = \gamma_{12}.$$

The uniform design becomes the minimax design if $\gamma = \frac{1}{12}$. Using $\mathcal{F}_1$ and the loss functions

$\mathcal{L}_D(f,\xi) = det(M(f,\xi))$, $\mathcal{L}_A(f,\xi) = trace(M(f,\xi))$

$\mathcal{L}_E(f,\xi) = \lambda_{max}(M(f,\xi))$ and $\mathcal{L}_G(f,\xi) = \sup_{x\in S}\ d(\mathbf{x}; f,\xi))$

where $d(\mathbf{x}; f,\xi)) = \mathbf{q}^T(\mathbf{x})M(f,\xi)\mathbf{q}(\mathbf{x})$, optimal designs for the multiple regression model have also been constructed by Wiens (1992). Other robust designs constructed using the class $\mathcal{F}_1$ can be found in Wiens (1991, 1993, 1994, 1996).

Pesotchinsky (1982) considers the setup given by

$$y(\mathbf{x}_i) = \sum_{j=0}^{k}\theta_j f_j(\mathbf{x}_i) + \psi(\mathbf{x}_i) + \varepsilon_i, \quad i = 1,2,...,n \tag{1.23}$$
$$f_0(\mathbf{x}_i) \equiv 1,\ f_j(\mathbf{x}_i) = x_{ij},\ \mathbf{x}_i = (x_{i1},...,x_{ik}) \in S$$

where $\psi(\mathbf{x})$ varies in the class $\mathcal{F}_2$ and $\phi(\mathbf{x})$ is a convex function of $\|\mathbf{x}\|^2 = \sum_{j=1}^{k} x_j^2$. In constructing the designs it is assumed that the experimenter intends to estimate the parameters of the model by means of the least squares method. Using the optimality criterion

$$\mathcal{L}_p(\xi,\psi) = \left[\frac{1}{k+1}tr\{M^p(\xi,\psi)\}\right]^{\frac{1}{p}},\ 0 < p < \infty \tag{1.24}$$

designs were constructed for the cases

(1) $\mathcal{L}_0(\xi,\psi) = \lim_{p\to 0^+}\mathcal{L}_p(\xi,\psi) = \{detM(\xi,\psi)\}^{\frac{1}{k+1}}$

(2) $\mathcal{L}_\infty(\xi,\psi) = \lim_{p\to\infty}\mathcal{L}_p(\xi,\psi) = \lambda_{max}\{M(\xi,\psi)\}$; and

(3) $\mathcal{L}_1(\xi,\psi) = \frac{1}{k+1}trace\{M(\xi,\psi)\}$

corresponding to D-, E-, and A-optimality criteria respectively. Restricting to symmetric designs ξ with $E_\xi(x_i) = m_{ii} = const = m$ and

$$S = \{\mathbf{x} : \|\mathbf{x}\| = \sqrt{km} = R\}$$

he showed that any symmetric design ξ_0 supported only by points on the sphere S_R of radius $R = \sqrt{mk}$ is D-optimal. The uniform design on a sphere of radius $R_1 = \sqrt{k\nu_1}$ and $R_\infty = \sqrt{k\nu_\infty}$ were shown to be A- and E-optimal respectively, where ν_1 and ν_∞ are determined appropriately. Interested readers can see Pesotchinsky (1982, pg 521). Other authors who have used the class $\mathcal{F}_2$ are Li and Notz (1982), Li (1984) and Liu and Wiens (1994).

Sacks and Ylvisaker (1984) used $\mathcal{F}_3$ and $\mathcal{F}_4$ to construct minimax designs for the nonparametric model

$$y_t = f(t) + \sigma\varepsilon_t, \quad E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1, \quad f \in \mathcal{F}$$

where $\mathcal{F}$ is one of $\mathcal{F}_3$, $\mathcal{F}_4$. The exact design problem was converted to an easier to solve approximate problem of choosing a measure C to minimize

$$J(k,t,c) = \frac{\sigma^2}{n}\left(\sum_{i=1}^{k}|c_i|\right)^2 + \sup_{\mathcal{F}}(Cf - \Gamma f)^2$$

where $Cf = \sum_{i=1}^{k} c_i f(t_i) = \int_S dC$ and Γf is any of the following

(i) Discrete : $\Gamma f = \sum_{j=1}^{N} \gamma_j f(x_j) = \int_S f d\Gamma$;

(ii) Continuous : $\Gamma f = \int_S \gamma(x) f(x) dx$;

(iii) Derivatives : $\Gamma f = f'(x_0)$.

For each of (i), (ii) and (iii) they proved the existence of an optimum C^* under appropriate conditions on n and k. Several examples and efficiency calculations were provided to illustrate the construction of designs from the approximate problem for fixed Γf. In Sacks and Ylvisaker (1985), their previous work was extended to the case where $f(t)$ is a realization of a stochastic process F with mean 0 and covariance function $R(s,t) = EF(s)F(t)$, $s,t \in S$. A connection was made between the Bayesian approach in this study and the minimax approach through a transformation to the RKHS associated with R. The specification of $\mathcal{F}$ serves to represent the departure from the assumed "ideal" model or serves as an approximation for the "real" f whose explicit form may never be known.

A related problem which has been studied by Karson, Manson and Hader (1969), Sacks and Ylvisaker (1978) and Marcus and Sacks (1978) is that of finding new linear estimators of $\boldsymbol{\theta}$ when the model is approximately linear. The strategy adopted by Karson et al (1969) is to construct an estimator which for a given design

(1) minimizes ISB, the bias arising from terms of specified higher degree being omitted from the fitted equation; and

(2) subject to achieving minimum bias, the estimator achieves minimum IV.

Linear Models With Autocorrelated Errors

When observations are taken sequentially in time, as in time series, it is often a good idea to subject the errors in the model representing the data to a test of independence or serial correlation. In most cases, it turns out that the errors are correlated. However, the precise structure of the underlying correlation is either unknown or the errors are only known to behave as a weakly stationary stochastic process such as the autoregressive or moving average process. The covariance matrix of the parameters in the model then depends on the unknown autocorrelation matrix, which we shall denote by P. In the absence of knowledge of P, the covariance matrix cannot be minimized to obtain optimal designs.

In the literature, robust designs for linear models with autocorrelated errors are constructed in two stages. These are:

(1) find a design ξ^* which is optimal for uncorrelated errors;

(2) order the design points to minimize the covariance matrix of the parameter estimate under correlated errors.

Following this procedure, Berenblut and Webb (1974) obtained robust D-optimal designs for the model

$$\mathbf{y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon},\ Var(\boldsymbol{\varepsilon}) = \sigma^2 P.$$

The correlation structure they considered is when $P = V(\rho)$, where ρ is the parameter of the first order autoregressive process. Thus $V(0) = I$, the identity matrix. Constantine (1989) constructed A-optimal designs which are robust against autocorrelated errors when $P = I + M$ where $M = (m_{ij})$ and

$$m_{ij} = \begin{cases} \rho_i\,, & \text{if } i = j+1 \text{ or } j = i+1 \\ 0\,, & \text{otherwise.} \end{cases} \tag{1.25}$$

Other contributions to the study of designs for exactly linear models with autocorrelated errors can be found in Jenkins and Chanmugan (1962), Kiefer and Wynn (1981, 1984), Bischoff (1992, 1993) and Pukelsheim (1993). Williams (1952), Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979) view the error process $\varepsilon(t)$ as a time series with the experimenter sampling in time. The most commonly studied process is the first-order autoregressive (AR(1)) process.

From the stages outlined above for the construction of designs robust against autocorrelated errors, one can see that these designs are themselves classical designs. Some work has been done in the area of constructing minimax designs for approximately linear models of the type (1.11) with autocorrelated errors by Wiens and Zhou (1996a, 1996b). In their work of 1996a they considered an error process which follows a very general model of dependence. The error process was assumed to have

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad Cov(\boldsymbol{\varepsilon}) = \sigma^2 P, \quad P_{ij} = \rho(|i-j|) \tag{1.26}$$

where P is the autocorrelation matrix of a weakly stationary process. Under some assumptions which include

$$\sum_{s=-\infty}^{\infty} |\rho(s)| < \infty \tag{1.27}$$

they obtained the asymptotic MSE($\sqrt{n}\hat{\boldsymbol{\theta}}$). Based on the IMSE loss they showed that the asymptotically minimax design ξ_* for the approximately linear model with uncorrelated errors retains its optimality when the errors are correlated, if the design points are randomly sampled from ξ_*. The error process also have to satisfy (1.26), (1.27) and

$$P = (1-\alpha)I + \alpha Q, \quad \alpha \in [0, 1]. \tag{1.28}$$

Following this work, they discussed (1996b) minimax designs for (1.11) when the errors follow an AR(1) process. The parameter was estimated by the best linear unbiased estimate (BLUE). They showed that the design ξ_* with marginal density (1.20) is an asymptotically minimax design for the BLUE provided the sign of ρ, $|\rho| < 1$, is fixed.

For convex classes $\mathcal{F}$ and $\mathcal{P}$ of disturbance functions and autocorrelation matrices define

$$f_s = (1-s)f_0 + sf_1, \quad f_0 \equiv 0, \quad f_1 \in \mathcal{F} \tag{1.29}$$

$$P_t = (1-t)P_0 + tP_1, \quad P_0 = I, \quad P_1 \in \mathcal{P}, \; 0 \le s, t \le 1. \tag{1.30}$$

Let $\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f, \xi, P)$ be the mean squared error matrix of $\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}$. The change of variance function $CVF(\xi, I, P_1)$ for a design ξ at I in the direction P_1 is defined by

$$CVF(\xi, I, P_1) = \frac{\partial}{\partial t}\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f_0, \xi, P_t)) \mid_{t=0} . \tag{1.31}$$

It measures the rate of change of MSE in the direction of a particular autocorrelation structure under small departures from the ideal model. Changes in MSE due to increases in bias as departures from the ideal model towards a particular disturbance function occurs is measured by the change of bias function (CBF). The CBF of ξ in the direction of f_1 is defined as

$$CBF(\xi, f_1) = \frac{1}{2}\frac{\partial^2}{\partial s^2}\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, sf_1, \xi, I)) \mid_{s=0} . \tag{1.32}$$

Using the CVF and the CBF, Wiens and Zhou (1997) introduced the concept of infinitesimal robustness. They defined three types of infinitesimal robustness. A design ξ is said to be

(a) V-robust if it minimizes $\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f_0, \xi, I))$ subject to a bound on the change of variance sensitivity (CVS). That is,

$$CVS(\xi, I) = \sup_{P \in \mathcal{P}} \frac{CVF(\xi, I, P)}{\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f_0, \xi, I))} \le \alpha; \tag{1.33}$$

(b) B-robust if it minimizes $\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f_0, \xi, I))$ subject to a bound on the change of bias sensitivity (CBS). That is,

$$CBS(\xi, f_0) = \sup_{f \in \mathcal{F}} \frac{CBF(\xi, f)}{\mathcal{L}(\mathrm{MSE}(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f_0, \xi, I))} \le \beta; \tag{1.34}$$

(c) M-robust if ξ is simultaneously V- and B-robust.

Restricting to the class of designs with $\sum_{i=1}^{n} x_i = 0$ and taking $S = [-0.5, 0.5]$, V-robust designs for the simple linear regression model were constructed for various range of values of α. They also derived the density for the B-robust design and outlined an approach to constructing M-robust designs.

In a related development designs which are robust against heteroscedastic errors in approximately specified regression models of type (1.11) with

$$f \in \mathcal{F}_1, \quad E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2 g(x_i)$$
$$\int_S g^2(x) \leq \Omega^{-1}$$

were discussed by Wiens (1996). One of the main results of this work is that the density $k_0(x)$ of the optimal design for polynomial fit, subject to a side condition of unbiasedness, is proportional to the function

$$h_q(x) = 0.5(P_q(x)P'_{q+1}(x) - P'_q(x)P_{q+1}(x))^{2/3} \tag{1.35}$$

where $P_q(x)$ is the qth Legendre polynomial on $[-1, 1]$. It is not difficult to show that the local maxima of $h_q(x)$, hence those of $k_0(x)$, are the zeros of $(1 - x^2)P'_q(x)$. From Example 4 of section 1.1.1, these are the support points of the D-optimal design ξ_D of the exact polynomial regression model. Therefore, $k_0(x)$ can be viewed as a smoothed version of ξ_D.

1.3. Some Background On Wavelets

In this section we introduce some theory on wavelets relevant to our work. We restrict ourselves to the basic definitions and some of its properties. More extensive discussions and examples can be found in Mallat (1989), Meyer (1992), Chui (1992) and Daubechies (1992).

A wavelet system is the collection of dilated and translated versions of a scaling function $\phi(x)$ and the primary wavelet $\psi(x)$ defined by

$$\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k) \tag{1.36}$$

and

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \quad j, k \in \mathcal{Z} \tag{1.37}$$

respectively. The functions $\phi(x)$ and $\psi(x)$ are choosen to satisfy the equations

$$\phi(x) = \sqrt{2}\sum_{p\in\mathcal{Z}} h_p\phi(2x-p) \tag{1.38}$$

$$\psi(x) = \sqrt{2}\sum_{r\in\mathcal{Z}} g_r\phi(2x-r), \quad g_r = (-1)^r h_{-r+1} \tag{1.39}$$

for a sequence $\{h_r\}$ of constants, called filter coefficients, with

$$\int \phi(x)dx = 1, \quad \int \psi(x)dx = 0, \quad \int \psi^2(x)dx = 1.$$

The condition

$$\sum_{p\in\mathcal{Z}} h_p = \sqrt{2} \tag{1.40}$$

ensures the existence of a unique solution to equations (1.38) and (1.39) (see Daubechies and Lagarias (1988)). Orthogonality of the translates of $\phi(x)$ is ensured by the condition

$$\sum_{p\in\mathcal{Z}} h_p h_{p-2j} = \delta_{0j}, \quad j\in\mathcal{Z}. \tag{1.41}$$

In the theory of wavelets, the space of square integrable functions, $\mathcal{L}_2(S)$ ($S \subseteq \Re$), is written as the limit of a sequence of closed subspaces $\{V_j\}$ where

$$\ldots\ldots\ldots \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \ldots\ldots\ldots \subset V_{-m} \subset \ldots\ldots \subset \mathcal{L}_2(S) \tag{1.42}$$

$$\bigcap_j V_j = \{0\}, \quad \overline{\bigcup_j V_j} = \mathcal{L}_2(S). \tag{1.43}$$

Definition 1.1: Let $\{\chi_k(x),\, k\in\mathcal{Z}\}$ be a complete system of functions in $\mathcal{L}_2(S)$. The system $\{\chi_k(x),\ k\in\mathcal{Z}\}$ is a Riesz basis if

(1) for any function $f(x)\in\mathcal{L}_2(S)$ the series of the squares of the Fourier coefficients is absolutely convergent. That is,

$$\sum_{k=1}^{\infty} |d_k|^2 \ < \ \infty;$$

where

$$d_k = \int_S f(x)\chi_k(x)dx$$

(2) for any sequence of numbers $\{d_k\} \in l_2$, the set of square summable sequences, there exists a function $f(x)$ for which the $\{d_k\}$ are its Fourier coefficients with respect to the set $\{\chi_k(x),\ k \in \mathcal{Z}\}$.

The conditions (1.38) and (1.39) ensure that the set $\{\phi_{j,k}(x),\ k \in \mathcal{Z}\}$ is a Riesz basis in each V_j. That is,

$$V_j = \{r(x) : r(x) = \sum_k \phi_{j,k} d_k, \|r\|^2 < \infty\} \tag{1.44}$$

for any fixed $j \in \mathcal{Z}$. Jaffard and Laurencot (in Chui (1992)) have shown that if H is a Hilbert space and $(\mathbf{e_p})$ a Riesz basis of H and G the operator defined by

$$G(f) = \sum_p < f, \mathbf{e_p} > \mathbf{e_p}, \tag{1.45}$$

then $\mathbf{a_p} = G^{-1/2}(\mathbf{e_p})$ forms an orthonormal basis of H. The Gram matrix G is defined by $G = (g(j,k))_{j,k \in \mathcal{Z}}$, where $g(j,k) \ = < \mathbf{e}_j, \mathbf{e}_k >$ and $< .,. >$ is the inner product on the Hilbert space H (see Meyer (1992) pg. 25). If in addition to (1.42) and (1.43) the condition

$$r(x) \in V_j \iff r(2^j x) \in V_0 \tag{1.46}$$

is satisfied, the sequence of closed subspaces $\{V_j, j \in \mathcal{Z}\}$ is said to be a multiresolution analysis of $\mathcal{L}_2(S)$. Mallat (1989) has shown that given any multiresolution analysis, it is possible to derive a function $\psi(x)$ such that the family $\{\psi_{j,k}(x)\ :\ j,k \in \mathcal{Z}\}$ is an orthonormal basis of $\mathcal{L}_2(S)$.

To construct $\psi_{j,k}(x)$, we define for each $j \in \mathcal{Z}$ the difference space W_j to be the orthogonal complement of V_j such that

$$W_j \bigoplus V_j \ = \ V_{j-1}, \quad W_j \perp V_j. \tag{1.47}$$

That is, any function $r(x) \in V_{j-1}$ can be written as a linear combination or direct sum of functions in W_j and V_j. It can be verified that W_j is a dilate of W_0

$$r(x) \in W_j \iff r(2^j x) \in W_0 \tag{1.48}$$

where

$$W_j = \left\{ r(x) \in \mathcal{L}_2(S) \mid r(x) = \sum_k c_k \psi_{j,k}(x),\ \|r(x)\| < \infty \right\}. \qquad (1.49)$$

Using (1.47), $\mathcal{L}_2(S)$ can be decomposed into a direct sum of the spaces W_j, so that

$$\bigoplus_j W_j \ = \ \overline{\bigcup_j V_j} \ = \ \mathcal{L}_2(S). \qquad (1.50)$$

This implies that $\mathcal{L}_2(S)$ is spanned by the dilates and translates of $\psi(x)$,

$$\mathcal{L}_2(S) = \left\{ r(x) \mid r(x) = \sum_{j,k} c_{jk} \psi_{j,k}(x),\ \|r(x)\| < \infty \right\}. \qquad (1.51)$$

The normalized dilates and translates $\psi_{j,k}(x)$ form an orthonormal wavelet basis for $\mathcal{L}_2(S)$.

The Haar wavelet basis is the simplest example of a wavelet system on $\mathcal{L}_2(S)$. The scaling function is :

$$\phi(x) = I_{[0,1)}(x) = \begin{cases} 1\,, & \text{if } 0 \leq x < 1 \\ 0\,, & \text{otherwise.} \end{cases} \qquad (1.52)$$

The refining relations for the Haar wavelet basis are

$$\phi(x) = \phi(2x-1) + \phi(2x)$$

and

$$\psi(x) = \phi(2x) - \phi(2x-1).$$

The multiwavelet system constructed by Alpert (1992) will also be useful in our study. The multiwavelet basis differ from other wavelet bases in that instead of a single scaling function $\phi(x)$, there are several functions $\phi_0, \ldots\ldots, \phi_{N-1}$ whose translates span the space V_0. Each scaling function is a dilated, translated and normalized Legendre polynomial on the interval $[0,1)$:

$$\phi_i(x) = \begin{cases} \sqrt{2i+1}P_i(2x-1), & x \in [0,1) \\ 0 & \text{otherwise} \end{cases} \qquad (1.53)$$

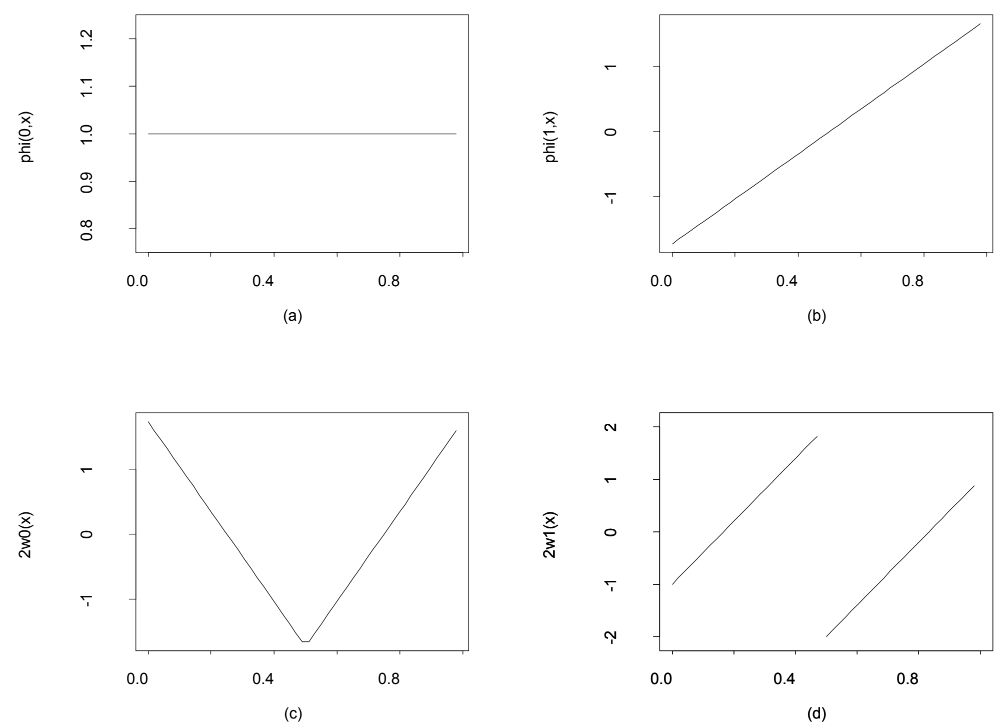


Figure 1: Scaling functions and Primary Wavelets: (a) $\phi_0(x)$; (b) $\phi_1(x)$; (c) ${}_2w_0(x)$; (d) ${}_2w_1(x)$.

where P_i, (i = 0,1,.....,N-1), are the Legendre polynomials. The space $V_n,\ n \in \mathcal{Z}$ are dilates of V_0 and the difference spaces W_n are as defined previously. The primary wavelets denoted by ${}_Nw_0, ..., {}_Nw_{N-1}$ vanish outside $[0, 1)$ and are orthogonal to polynomials of maximum degree,

$$\int_S {}_Nw_j(x)x^i dx = 0, \quad i = 0, 1, ..., N - 1 + j.$$

It turns out that the multiwavelets coincide with the Haar wavelet basis if N = 1. The procedure for constructing these wavelets are outlined in Alpert (1992, pgs. 197-199). For $N = 2$ the scaling functions and primary wavelets are

$$\phi_0(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.54}$$

$$\phi_1(x) = \begin{cases} \sqrt{3}(2x - 1), & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{1.55}$$

$${}_2w_0(x) = \begin{cases} \sqrt{3}(1 - 4x), & 0 \leq x < \frac{1}{2} \\ \sqrt{3}(4x - 3), & \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{1.56}$$

$${}_2w_1(x) = \begin{cases} 6x - 1, & 0 \leq x < \frac{1}{2} \\ 6x - 5, & \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{1.57}$$

The refining relations for these multiwavelets (N=2) are:

$$\begin{gathered} \phi_0(x) = \phi_0(2x) + \phi_0(2x - 1) \\ \phi_1(x) = \tfrac{\sqrt{3}}{2}(\phi_0(2x - 1) - \phi_0(2x)) + \tfrac{1}{2}(\phi_1(2x - 1) + \phi_1(2x)) \\ {}_2w_0(x) = \phi_1(2x - 1) - \phi_1(2x) \\ {}_2w_1(x) = \tfrac{1}{2}(\phi_0(2x) - \phi_0(2x - 1)) + \tfrac{\sqrt{3}}{2}(\phi_1(2x - 1) + \phi_1(2x)). \end{gathered} \tag{1.58}$$

The graphs of the scaling functions and primary wavelets are shown in Figure 1.

1.4. Summary of Results

Beginning from Chapter 2, we construct robust designs for the 'approximately linear wavelet regression model' of the form (1.11) with $q = 1$. We adopt the minimax approach. The wavelet bases on $S = [0, 1]$ used in the construction of the designs are the $N = 1$ (Haar) and $N = 2$ multiwavelets.

In Chapter 2, we transform the problem of finding the least favourable disturbance function $f(x)$ into an eigenvalue problem involving the symmetric positive definite root of a matrix G. We show that this matrix is at least positive semi-definite and propose a procedure for approximating G if it is singular.

To fix ideas, we proceeded step by step to construct minimax robust designs for the biased mth order Haar wavelet regression model. We found some evidence that no non-symmetric design is admissible (see Sections 3.1.1 and 3.1.2). Our conjecture is that this is true in general. Considering $m = 0$ and $m = 1$ we have shown, in Sections 3.1.1 and 3.1.2, that among symmetric and absolutely continous designs and for any $f \in \mathcal{F}$ the uniform design is A-, Q- and D-optimal. The results for $m = 0$ and $m = 1$ raised the suspicion that the uniform design might be minimax robust for the general problem. This suspicion led us to begin searching for a proof.

In the general case, we first considered the classical problem. We were able to show that any design ξ_0 with the property $B(\xi_0) = I_{2^{m+1}}$ is simultaneously A-, Q-, D- and G-optimal. It turns out that the design ξ_0^* which places equal mass $2^{-(m+1)}$ in each of the 2^{m+1} subintervals of S has this property. Using information from the results of the classical problem, it was not too difficult to show that the continuous version of ξ_0^*, the continuous uniform design is minimax robust in a strong sense (see the remarks after the proof of Theorem 2.2).

Under the assumption that the order of approximation for the two primary wavelets (m and p) of the $N = 2$ multiwavelet are equal ($m = p = 0$), we derived minimax robust A-, Q- and D-optimal design densities for the biased $N = 2$ multiwavelet regression model. The minimax design derived places more mass in a neighbourhood of the midpoint of the design space and a few at the extremes (see Figure 2). We are only able to provide solutions to this simplest case due to the complexity of the eigenvalue problem arising from the maximization of the loss function with respect to f.

In Chapter 3, we assume that the experimenter will use weighted least squares in estimating the parameters of the wavelet model. Under this assumption, we derived an optimal weight function and design density, with respect to the IMSE criterion, for the general wavelet regression model subject to the condition of unbiasedness. The condition of unbiasedness eliminates the complicated eigenvalue problem mentioned earlier. The optimal design density is shown to be a function of the squared Euclidean norm of the vector of wavelet basis used in the approximation and also inversely proportional to the optimal weight.

Using the Haar basis, we found that the optimal weight and design are each uniform. Implying that the ordinary least squares method is optimal in estimating the parameters of the Haar wavelet regression model if the model is unbiased.

For the $N = 2$ multiwavelet regression model we first derived a closed form for the squared norm of the vector whose components are the $N = 2$ multiwavelet basis for $m = p$. Using the closed form, we showed that the squared norm is a convex paraboloid in each of the 2^{m+1} subintervals of S with a maximum value of 2^{m+3} and a minimum value of 2^{m+1}, attained at the endpoints and midpoint, respectively, of each subinterval. The obvious implication of this finding is that the optimal weight and design are respectively concave and convex paraboloids in each of the 2^{m+1} subintervals. By deriving some identities, we are able to establish the fact that the optimal design is symmetric about $x = \frac{1}{2}$. And also that the value of the squared norm on the design space S is completely determined by its value in only one of the 2^{m+1} subintervals. The last result of this chapter is the derivation of a recursive relation for the squared norm when $m \neq p$.

In Chapter 4 we propose strategies for implementing the designs constructed in Chapters 2 and 3. We also consider how well the multiwavelets used in this study and the optimal weights derived can be used to approximate some commonly used nonlinear models and data with no pre-specified model. We find that the general features of the models were picked up by the fitted wavelet models with some features of the primary wavelet retained.

Chapter 2

ROBUST MINIMAX DESIGNS FOR BIASED MULTIWAVELET REGRESSION MODELS: ORDINARY LEAST SQUARES

1. Preliminaries

We continue our discussion by considering the model which describes the ith response $y_i \in \Re$ in a nonlinear experiment as follows:

$$y_i = \eta(x_i) + \varepsilon_i, \quad i = 1, 2, ..., n \tag{2.1}$$

where $x_i \in \Re$ is the ith design point of the explanatory variable x choosen from some design space $S \subseteq \Re$; $\eta(x_i) \in \Re$ is the value of some nonlinear function η at the design point x_i; and $\varepsilon_i \in \Re$ is a random sequence of uncorrelated and unobservable errors with mean zero and constant variance $\sigma^2 > 0$.

In this chapter, we begin the construction of designs for wavelet approximations to the nonlinear model (2.1). First, we discuss the structure of the approximation. Then, we provide some background on the general theory underlying the construction of robust designs in Section 2. The problem of finding the least favourable disturbance function $f(x)$ arising out of the wavelet approximation is considered in Section 2.1. It turns out that the general form of $f(x)$ can be obtained. However, we need a specific wavelet basis to construct the designs. The multiwavelets with parameter $N = 1, 2$ were used for this purpose. We observed in Section 1.3 of Chapter 1 that the multiwavelets with $N = 1$ coincides with the Haar wavelet basis.

The main result of this chapter is the finding that any design ξ_0 with $B(\xi_0) = I_{2^{m+1}}$ is simultaneously A-, D-, G- and Q-optimal for the classical design problem if the Haar wavelet basis is used in the approximation of any nonlinear function $\eta(x)$. That is, ξ_0 minimizes $trB^{-1}(\xi)$ and maximizes the *determinant* of $B(\xi)$, where the covariance matrix is proportional to $B^{-1}(\xi)$. We use the equivalence theorem and orthonormality of wavelet basis to obtain G- and Q-optimality. We also find that the continuous uniform design is Q- and D-optimal for the robust design problem. For the multiwavelets with $N = 2$ we are only able to provide solutions to the simplest case due to the complexity of the eigenvalue problem arising from the maximization of the loss function with respect to f.

Let $x \in S = [0,1]$ and $\eta(x) \in \mathcal{L}_2(S)$. The multiresolution analysis of $\mathcal{L}_2(S)$, discussed in Section 1.3 of Chapter 1, leads to two wavelet representations of $\eta(x)$ defined by Meyer (1992) and Walter (1995) as :

$$\eta(x) = \sum_{j,k \in \mathcal{Z}} c_{jk} \psi_{-j,k}(x) \tag{2.2}$$

and

$$\eta(x) = \sum_{l \in \mathcal{Z}} d_l \phi_{0,l}(x) + \sum_{j,k \in \mathcal{Z}_+} c_{jk} \psi_{-j,k}(x) \tag{2.3}$$

where

$$d_l = \int_S \eta(x) \phi_{0,l}(x) dx, \quad c_{jk} = \int_S \eta(x) \psi_{-j,k}(x) dx. \tag{2.4}$$

Meyer (1992) also showed that (2.3) implies (2.2) and states that it is not known whether (2.2) implies (2.3). Since actual computations require finite representations, we rewrite $\eta(x)$ as :

$$\eta(x) = \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} c_{jk} \psi_{-j,k}(x) + f(x) \tag{2.5}$$

and

$$\eta(x) = \sum_{l \in \mathcal{Z}} d_l \phi_{0,l}(x) + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} c_{jk} \psi_{-j,k}(x) + f(x) \tag{2.6}$$

respectively. The range of k has been restricted to $(k \geq 0)$ so that at any level j the orthonormal wavelet basis $\psi_{-j,k}(x)$ will be zero on the complement of the design space $S = [0,1]$, (see also (2.46) and (2.109)). The function f(x) is the remainder satisfying

$$\int_S f^2(x) dx \leq \tau^2$$

for some small, known value τ and m a finite non-negative integer. Define the $1 \times 2^{m+1}$ vectors

$$\begin{aligned} \boldsymbol{\beta}_0^T &= (d_0, c_{00}, c_{10}, c_{11}, ..., c_{m,2^m-1}); && (2.7) \\ \mathbf{q}^T(x) &= (\phi_{00}(x), \psi_{00}(x), \psi_{-1,0}(x), \psi_{-1,1}(x), ..., \psi_{-m,2^m-1}(x)) \\ &= (\phi(x),\ \psi(x),\ 2^{1/2}\psi(2x),\ 2^{1/2}\psi(2x-1), ... \\ &\quad ...,\ 2^{m/2}\psi(2^m x - 2^m + 1)). && (2.8) \end{aligned}$$

Then (2.1) can be written as

$$y(x) = \mathbf{q}^T(x)\boldsymbol{\beta}_0 + f(x) + \varepsilon. \qquad (2.9)$$

To estimate the parameters $\boldsymbol{\beta}_0$, we employ the least squares method because of it's classical nature and mathematical convenience. We note that the design problems discussed in this work remains the same if the robust M-estimate or the Mallows-type Generalized (or "Bounded Influence") M-estimate is used instead of the Least squares method. This is a consequence of the fact that the asymptotic variance of $\hat{\boldsymbol{\beta}}_M$ is a scalar multiple of the variance under ordinary least squares estimation. The asymptotic variance of $\hat{\boldsymbol{\beta}}_{GM}$ is proportional to the variance of $\hat{\boldsymbol{\beta}}_{WLS}$. In both cases, the multiples are independent of the weights and design. For details of the asymptotics see Wiens (1996a).

A special class of nonparametric regression smoothers of $\eta(x)$ is the local averaging procedure defined in general by

$$\hat{\eta}(x) = n^{-1} \sum_{i=1}^{n} W_{ni}(x) Y_i$$

where $\{W_{ni}(x)\}_{i=1}^n$ is a sequence of weights which may also depend on the points x_i. If the weights are positive and satisfy

$$n^{-1} \sum_{i=1}^{n} W_{ni}(x) = 1,$$

then $\hat{\eta}(x)$ is a least squares estimate. In this case, $\hat{\eta}(x)$ is the solution to

$$\min_{\mu} n^{-1} \sum_{i=1}^{n} W_{ni}(x)(Y_i - \mu)^2.$$

So, the least squares estimate is a special case of the nonparametric local averaging procedure for estimating $\eta(x)$. The ordinary least squares method of estimation which we have adopted in this chapter corresponds to the case where the weights are uniform. In Chapter 3, we consider the case when the weights are not uniform.

Antoniadis, Gregoire and Mckeague (1994) in their discussion of least squares wavelet regression, observed that the wavelets used for least squares regression should form a basis of the $\mathcal{L}_2$ space on the design region S. This explains why we considered a multiresolution of the design space S rather than of $\Re$ in the early part of this section. In recent years, several authors have considered the problem of constructing wavelets which form a basis of $\mathcal{L}_2$ on a closed interval [A,B]. They include Andersson et al (1993), Alpert (1992), Cohen, Daubechies and Vial (1992), Chui and Quak (1992), Daubechies (1993) and Jaffard and Meyer (1989). Antoniadis, Gregoire and Mckeague (1994) also examined the problem of the best value of m and state that in practice, for sample sizes between 100 and 200, it suffices to examine only m = 3,4 and 5.

2. General Theory

The wavelet equivalent (2.9) of model (2.1) is precisely the 'approximately linear model' discussed in Section 1.2.1 of Chapter 1. Following the technique outlined in that section, we approximate $E(y|x)$ by $\mathbf{q}^T(x)\boldsymbol{\beta}_0$ and define, by least squares, the "true" parameter in the wavelet approximation $\boldsymbol{\beta}_0$ by

$$\boldsymbol{\beta}_0 = arg\min_{\boldsymbol{\beta}} \int_S [E(y|x) - \mathbf{q}^T(x)\boldsymbol{\beta})]^2 dx. \tag{2.1}$$

Then with

$$f(x) = E(y|x) - \mathbf{q}^T(x)\boldsymbol{\beta}_0$$

we have that (2.9) holds, and that

$$\int_S \mathbf{q}(x) f(x) dx = \mathbf{0}. \tag{2.2}$$

We note that (2.9) and (2.11) define $\boldsymbol{\beta}_0$ uniquely since

$$\int_S \mathbf{q}(x)\mathbf{q}^T(x) dx = I, \tag{2.3}$$

implying

$$\boldsymbol{\beta}_0 = \int_S E(y|x)\mathbf{q}(x)dx.$$

Suppose that a sample of size n, $\{(x_i, y_i)\}_i^n$, is taken from the model (2.1), approximated by (2.9). The least squares estimate of $\boldsymbol{\beta}_0$ under the approximation $E(y|x) \approx \mathbf{q}^T(x)\boldsymbol{\beta}_0$ is

$$\hat{\boldsymbol{\beta}} = (Q^T Q)^{-1} Q^T \mathbf{y}(x) \tag{2.4}$$

where Q : $n \times 2^{m+1}$ is given by

$$Q = \begin{bmatrix} \phi(x_1) & \psi(x_1) & 2^{1/2}\psi(2x_1) & \dots & \dots \\ \phi(x_2) & \psi(x_2) & 2^{1/2}\psi(2x_2) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \phi(x_n) & \psi(x_n) & 2^{1/2}\psi(2x_n) & \dots & \dots \end{bmatrix} = \begin{bmatrix} \mathbf{q}^T(x_1) \\ \mathbf{q}^T(x_2) \\ \dots \\ \dots \\ \mathbf{q}^T(x_n) \end{bmatrix} \tag{2.5}$$

and

$$\begin{aligned} \mathbf{y}^T(x) &= (y(x_1), y(x_2), \dots\dots., y(x_n)) \\ \mathbf{f}^T(x) &= (f(x_1), f(x_2), \dots\dots., f(x_n)). \end{aligned} \tag{2.6}$$

Define

$$B = B(\xi) = \frac{Q^T Q}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{q}(x_i)\mathbf{q}^T(x_i) = \int_S \mathbf{q}(x)\mathbf{q}^T(x)d\xi(x) \tag{2.7}$$

$$\mathbf{b} = \mathbf{b}(f,\xi) = \frac{Q^T \mathbf{f}(x)}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{q}(x_i)f(x_i) = \int_S \mathbf{q}(x)f(x)d\xi(x) \tag{2.8}$$

and express $\hat{\beta}$ in terms of $\xi(x)$ to obtain

$$\hat{\boldsymbol{\beta}} = B^{-1}\int_S \mathbf{q}(x)y(x)d\xi(x) \tag{2.9}$$

where $\xi(x)$ is the distribution function of x_i given by (1.3). Under the model (2.9), the bias, μ and variance, Σ of $\hat{\boldsymbol{\beta}}$ are

$$\boldsymbol{\mu} = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = B^{-1}\mathbf{b}(f,\xi) \;\; and \;\; \Sigma = \frac{\sigma^2}{n}B^{-1} \tag{2.10}$$

respectively. For the uniform design, $d\xi(x) = dx$, $B(\xi) = I$ and $\mathbf{b}(f,\xi) = 0$. This implies that,

$$\boldsymbol{\mu} = 0, \;\; and \;\; \Sigma = \frac{\sigma^2}{n}I.$$

Denoting the mean squared error matrix of $\hat{\boldsymbol{\beta}}$ by $M(f,\xi)$ we have

$$\begin{aligned} M(f,\xi) &= \boldsymbol{\mu}\boldsymbol{\mu}^T + \Sigma \\ &= B^{-1}\mathbf{b}\mathbf{b}^T B^{-1} + \frac{\sigma^2}{n} B^{-1}. \end{aligned} \tag{2.11}$$

We recall that the idea behind (1.3) is to transform the n-observation design problem into that of finding a probability measure ξ^*, corresponding to an n-observation design, such that

$$\min_{\xi} \max_{f\in\mathcal{F}} \Phi(M(f,\xi)) = \max_{f\in\mathcal{F}} \Phi(M(f,\xi^*)) \tag{2.12}$$

for some real-valued monotone function Φ, and

$$\mathcal{F} = \left\{ \ f(x) \mid \textstyle\int_S f^2(x)dx \le \tau^2, \ \int_S \mathbf{q}(x)f(x)dx = \mathbf{0} \ \right\}. \tag{2.13}$$

If we can find a ξ^* to solve (2.21) then hopefully an n-observation design $\mathbf{x}_*$ whose associated probability distribution approximates ξ^* will be close to optimal for the n-observation design problem.

All loss functions we will be considering satisfy the following conditions:

(C1) Monotonicity: If $M(f_1,\xi) \ \ge \ M(f_2,\xi)$, in the sense of positive semidefiniteness, then $\Phi(M(f_1,\xi)) \ \ge \ \Phi(M(f_2,\xi))$;

(C2) Unboundedness: $\Phi(M(f_n,\xi)) \ \to \ \infty$ if $Ch_1(M(f_n,\xi)) \ \to \ \infty$ as $n \ \to \ \infty$, where Ch_1 denotes the maximum characteristic root.

We also assume, to avoid trivialities, that if there is a point $x_0 \in S$ with $\mathbf{q}(x_0) = \mathbf{0}$, (e.g. $\mathbf{q}(1) = \mathbf{0}$), then $\xi\{x_0\} = 0$. Otherwise, since such a point x_0 would contribute nothing to $\mathbf{b}$ or B, we could remove it from S and work with the conditional design on $S\backslash\{x_0\}$.

Under the above conditions, Wiens (1992) has shown that a necessary condition for $sup_{\mathcal{F}} \ \Phi(M(f,\xi))$ to be finite is the absolute continuity of the design measure ξ. The loss functions we shall consider are:

(1) Integrated mean squared error loss (IMSE),

$$\begin{aligned} \mathcal{L}_Q(f,\xi) = \Phi(M(f,\xi)) &= \int_S E[\{\hat{y}(x) - E(y|x)\}^2]dx \\ &= \int_S E[\{\hat{\boldsymbol{\beta}}^T\mathbf{q}(x) - \boldsymbol{\beta}_0^T\mathbf{q}(x) - f(x)\}^2]dx \\ &= \int_S \mathbf{q}^T(x)M(f,\xi)\mathbf{q}(x)dx + \int_S f^2(x)dx. \end{aligned}$$

Substitute for $M(f, \xi)$ and simplify to obtain

$$\mathcal{L}_Q(f, \xi) = \mathbf{b}^T B^{-2}\mathbf{b} + \frac{\sigma^2}{n} tr B^{-1} + \int_S f^2(x)dx. \tag{2.14}$$

(2) Trace of $M(f, \xi)$,

$$\mathcal{L}_A(f, \xi) = \mathbf{b}^T B^{-2}\mathbf{b} + \frac{\sigma^2}{n} tr B^{-1}. \tag{2.15}$$

(3) Determinant of $M(f, \xi)$,

$$\begin{aligned} \mathcal{L}_D(f, \xi) &= \Phi(M(f, \xi)) = |M(f, \xi)| \\ &= \left| B^{-1}\mathbf{b}\mathbf{b}^T B^{-1} + \frac{\sigma^2}{n} B^{-1} \right| \\ &= \left(\frac{\sigma^2}{n}\right)^{2^{m+1}} \left\{ \frac{1 + \left(\frac{n}{\sigma^2}\right) \mathbf{b}^T B^{-1}\mathbf{b}}{|B|} \right\}. \end{aligned} \tag{2.16}$$

Any design that is optimal with respect to these loss functions will be said to be Q-, A- and D-optimal respectively. By optimal we mean the design which minimizes the maximum (over f) loss. We observe that we can also define the IMSE as

$$\mathcal{L}_Q(f, \xi) = tr\{M(f, \xi)\} = \mathcal{L}_A(f, \xi), \tag{2.17}$$

since the maximum over $f \in \mathcal{F}$ lies on the boundary of the first constraint in (2.22) and therefore has no effect on the maximization problem (see discussion before and after (2.31)). Expression (2.26) then implies that

$$\min_{\xi} \max_{f \in \mathcal{F}} \mathcal{L}_Q(f, \xi) \equiv \min_{\xi} \max_{f \in \mathcal{F}} \mathcal{L}_A(f, \xi). \tag{2.18}$$

That is, Q- and A-optimality are equivalent for wavelet regression models.

2.1. Least Favourable Function $f(x)$

A-, Q-optimality

We first fix ξ and maximize $\mathcal{L}_Q$ over $\mathcal{F}$. We consider only designs with finite maximum loss; these are necessarily absolutely continuous. Denote by $m(x)$ the density of the distribution $\xi(x)$. Then, our problem is,

$$\begin{gathered} maximize \;\; J(f, \xi) = \mathbf{b}^T B^{-2}\mathbf{b} + \textstyle\int_S f^2(x)dx \;\; subject\ to \\ (i)\ K_1(f) = \textstyle\int_S \mathbf{q}(x) f(x)dx = \mathbf{0}, \;\; (ii)\ K_2(f) = \textstyle\int_S f^2(x)dx - \tau^2 \leq 0. \end{gathered} \tag{2.19}$$

To solve the above problem we either proceed as in Wiens (1990) or use the Fritz John's condition and the independence constraint qualification for mixed constraints (see Bazaraa and Shetty (1976)) to obtain the same solution. We employ the latter approach. Let $\boldsymbol{\nu}_1 \in \Re^{2^{m+1}}$, $\nu_2 \geq 0$ be Lagrange multipliers. The maximizing f_0 must then satisfy the equations

$$\begin{gathered} -\delta J(f_0, \Delta f, \xi) + \boldsymbol{\nu}_1^T \delta K_1(f_0, \Delta f) + \nu_2 \delta K_2(f_0, \Delta f) = 0 \\ K_1(f_0) = 0 \\ \nu_2 K_2(f_0) = 0, \text{ for all } \Delta f \in \mathcal{F} \end{gathered} \tag{2.20}$$

where $\delta J(f_0, \Delta f, \xi)$ is the Gateaux variation of the functional $J(f_0, \Delta f, \xi)$ in the direction of Δf at $f_0 \in \mathcal{F}$. Now,

$$\begin{gathered} \delta J(f_0, \Delta f, \xi) = 2 \int_S \mathbf{b}^T(f_0) B^{-2} \mathbf{q}(x) m(x) \Delta f(x) dx \\ \delta K_1(f_0, \Delta f) = \int_S \mathbf{q}(x) \Delta f(x) dx \\ \delta K_2(f_0, \Delta f) = 2 \int_S f_0(x) \Delta f(x) dx. \end{gathered} \tag{2.21}$$

There are two possibilities for the multiplier ν_2 :

(1) $\nu_2 = 0$ implying $K_2(f_0)$ is inactive or nonbinding.

(2) $\nu_2 > 0$ implying $K_2(f_0)$ is active.

We note from (2.28) that if case (1) holds then we have strict inequality in constraint (ii). Otherwise, we have equality and

$$\int_S f^2(x) dx - \tau^2 = 0.$$

Note that if $f_0(x)$ satisfies (i) but

$$\int_S f_0^2(x) dx = h^2 \tau^2 < \tau^2 \text{ for some } 0 < h^2 < 1,$$

then

$$\int_S (h^{-1} f_0)^2(x) dx = \tau^2.$$

The function $h^{-1} f_0(x)$ satisfies the constraints (i) and (ii), and

$$\begin{aligned} \mathbf{b}^T(h^{-1} f_0, \xi) B^{-1}(\xi) \mathbf{b}(h^{-1} f_0, \xi) &= h^{-2} \mathbf{b}^T(f_0, \xi) B^{-1}(\xi) \mathbf{b}(f_0, \xi) \\ &> \mathbf{b}^T(f_0, \xi) B^{-1}(\xi) \mathbf{b}(f_0, \xi). \end{aligned}$$

It follows that the maximum over f lies on the boundary of the second constraint and case (2) holds. So, our problem becomes,

$$\begin{array}{c} maximize\ \ J(f,\xi) \ = \ \mathbf{b}^T B^{-2}\mathbf{b}\ \ subject\ to \\ (i)\ K_1(f) = \int_S \mathbf{q}(x)f(x)dx = \mathbf{0}, \ \ (ii)\ K_2(f) = \int_S f^2(x)dx \ - \ \tau^2 \ = \ 0. \end{array} \tag{2.22}$$

The second term of the functional $J(f,\xi)$, namely $\int_S f^2(x)dx$, has been dropped because constraint (ii) implies that it is a constant, τ^2. Hence, the second term has no effect on the solution of the maximization problem.

Combining (2.29) and (2.30) we have

$$\int_S \{2\mathbf{b}^T(f_0)B^{-2}\mathbf{q}(x)m(x) - \boldsymbol{\nu}_1^T\mathbf{q}(x) - 2\nu_2 f_0(x)\}\Delta f(x)dx \ = \ 0.$$

If this equation holds for all Δf, then the maximizing f_0 must satisfy the equation

$$2\mathbf{b}^T(f_0)B^{-2}\mathbf{q}(x)m(x) - \boldsymbol{\nu}_1^T\mathbf{q}(x) - 2\nu_2 f_0(x) \ = \ 0. \tag{2.23}$$

That is,

$$f_0(x) = \mathbf{q}^T(x)[B^{-2}m(x)\mathbf{c} + \boldsymbol{\alpha}] \tag{2.24}$$

where

$$\mathbf{c} = \frac{\mathbf{b}(f_0)}{\nu_2}, \ \ \boldsymbol{\alpha} = -\frac{\boldsymbol{\nu}_1^T}{2\nu_2}.$$

Normalize (2.33) to satisfy (2.28i) to obtain

$$\boldsymbol{\alpha} = -B^{-1}\mathbf{c}.$$

Therefore,

$$f_0(x) = \mathbf{q}^T(x)B^{-1}[B^{-1}m(x) \ - \ I]\mathbf{c}. \tag{2.25}$$

To show that the maximizing $f_0(x)$ is of the form (2.34), see Wiens (1990).

From (2.24) we have that

$$\mathcal{L}_Q(f_0,\xi) = \mathbf{b}^T(f_0,\xi)B^{-2}(f_0)\mathbf{b}(f_0,\xi) + \left(\frac{\sigma^2}{n}\right) trB^{-1} + \tau^2. \tag{2.26}$$

If we define

$$C = \int_S \mathbf{q}(x)\mathbf{q}^T(x)m^2(x)dx,$$

then

$$\mathbf{b}(f_0, \xi) = (CB^{-2} \; - \; I)\mathbf{c}. \tag{2.27}$$

Problem (2.31) becomes

$$\begin{array}{c} maximize \; J(f_0, \xi) = \mathbf{c}^T(B^{-2}C - I)B^{-2}(CB^{-2} - I)\mathbf{c} \\ subject \;\; to \\ \tau^2 = \mathbf{c}^T B^{-2}(C - B^2)B^{-2}\mathbf{c}. \end{array} \tag{2.28}$$

Let $G = C$ - B^2 and $G^{\frac{1}{2}}$ the symmetric positive definite root of G. The question that arises in this context is whether G is positive definite. First, we show that it is at least positive semi-definite. To see this, we note that G can be written as,

$$G = \int_S \mathbf{q}(x)\mathbf{q}^T(x)m^2(x)dx \; - \; B^2 = \int_S (m(x)I - B)\mathbf{q}(x)\mathbf{q}^T(x)(m(x)I - B)dx.$$

So that,

$$\tilde{\mathbf{a}}^T G\tilde{\mathbf{a}} = \int_S \{\tilde{\mathbf{a}}^T(m(x)I - B)\mathbf{q}(x)\}^2 dx \; \geq \; 0,$$

for some vector $\tilde{\mathbf{a}} \in \Re^{2^{m+1}}$. Now, if G is not positive definite, we approximate it as follows:

(i) Take any density $m_1(x)$ for which $G > 0$ and put
$m_t(x) = (1 - t)m(x) + tm_1(x)$.

(ii) Evaluate $G_t = (C - B^2)|_{m(x)=m_t(x)}$.

(iii) Evaluate $P(t) = |G_t|$. Then,

(a) $P(t)$ is a polynomial in t.

(b) $P(0) = 0$, $P(1) = |G_{m_1(x)}| > 0$, which implies that $P(t)$ is positive (since P is non-negative) for all sufficiently small $t > 0$.

(iv) Put $G_n = G_{t_n} > 0$ for a sequence $t_n \downarrow 0$.

(v) Use G_n in place of G and take limits at (2.42).

If we define

$$\mathbf{a} = \frac{G^{\frac{1}{2}} B^{-2} \mathbf{c}}{\tau}, \tag{2.29}$$

then

$$\mathbf{c} = \tau B^2 G^{-\frac{1}{2}} \mathbf{a},$$

and

$$f_0(x) = \tau \mathbf{q}^T(x)[m(x)I - B]G^{-\frac{1}{2}}\mathbf{a} \tag{2.30}$$

for some $\mathbf{a}$ satisfying $\|\mathbf{a}\|^2 = 1$. We rewrite (2.39) as

$$f_0(x) = \mathbf{u}^T(x)\mathbf{a}$$

where

$$\mathbf{u}^T(x) = \tau \mathbf{q}^T(x)[m(x)I - B]G^{-\frac{1}{2}}.$$

It can be verified that the following hold :

$$\begin{array}{c} \int_S \mathbf{u}(x)\mathbf{u}^T(x)dx = \tau^2 I \\ \int_S \mathbf{u}(x)\mathbf{q}^T(x)m(x)dx = \tau G^{\frac{1}{2}} \\ \int_S f_0^2(x)dx = \tau^2 \\ \mathbf{b}(f_0, \xi) = \tau G^{\frac{1}{2}}\mathbf{a} \end{array} \tag{2.31}$$

and

$$\mathbf{b}^T(f_0, \xi)B^{-2}\mathbf{b}(f_0, \xi) = \tau^2 \mathbf{a}^T G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}} \mathbf{a}. \tag{2.32}$$

Our problem is then to maximize (2.41) subject to $\|\mathbf{a}\|^2 = 1$. Now,

$$\max_{\|\mathbf{a}\|=1} \tau^2 \mathbf{a}^T G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}} \mathbf{a} = \tau^2 Ch_1\{G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}}\} \tag{2.33}$$

where $\mathbf{a}$ is the eigenvector corresponding to the maximum characteristic root $Ch_1(\cdot)$. To find the eigenvalues of $G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}}$, we solve

$$|G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}} - \lambda I| = 0 \ \ or \ \ |G - \lambda B^2| = 0 \tag{2.34}$$

where $G = C - B^2$.

D-optimality

Here, maximizing $\mathcal{L}_D$ over $\mathcal{F}$ is equivalent to maximizing the functional $\mathbf{b}^T B^{-1}\mathbf{b}$. Proceeding as in Section 2.1.1, we have that the maximizing f_0 is of the form

$$f_0(x) = \mathbf{q}^T(x)[B^{-1}m(x) - I]\mathbf{c}.$$

We transform the maximization problem into an eigenvalue problem by observing that

$$\begin{aligned} \mathbf{b}(f_0, \xi) &= B(B^{-1}CB^{-1} - I)\mathbf{c} \\ \mathbf{b}^T B^{-1}\mathbf{b} &= \mathbf{c}^T(B^{-1}CB^{-1} - I)B(B^{-1}CB^{-1} - I)\mathbf{c} \\ \tau^2 &= \mathbf{c}^T B^{-1}(C - B^2)B^{-1}\mathbf{c}. \end{aligned} \tag{2.35}$$

So that we now solve

$$\max_{\|\mathbf{a}\|=1} \tau^2 \mathbf{a}^T G^{\frac{1}{2}} B^{-1} G^{\frac{1}{2}} \mathbf{a} = \tau^2 Ch_1\{G^{\frac{1}{2}} B^{-1} G^{\frac{1}{2}}\} \tag{2.36}$$

where $\mathbf{a}$ is defined by (2.38). That is, we solve $|G^{\frac{1}{2}}B^{-1}G^{\frac{1}{2}} - \lambda I| = 0$ or $|G - \lambda B| = 0$.

For the purpose of illustration, we now consider examples using specific type of wavelets which form an orthonormal basis of $\mathcal{L}_2(S)$.

3. Examples

3.1. The Haar Wavelet

The Haar wavelet basis for $\mathcal{L}_2(S)$ is given by (see Daubechies (1993), in Recent Advances In Wavelet Analysis)

$$\{\phi_{0,0}\} \cup \{\psi_{-j,k};\ j \le 0,\ 0 \le k \le 2^{-j} - 1\} \tag{2.1}$$

where the scaling function is defined by (1.52) and the primary wavelet is

$$\psi(x) = I_{[0 \le x < \frac{1}{2}]} - I_{[\frac{1}{2} \le x < 1]}. \tag{2.2}$$

The wavelet coefficients defined in (1.38) and (1.39) are

$$h_k = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0, 1 \\ 0 & \text{otherwise} \end{cases}, \quad g_k = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0 \\ -\frac{1}{\sqrt{2}} & \text{if } k = 1 \\ 0 & \text{otherwise.} \end{cases}$$

From the primary wavelet, we find that

$$\begin{aligned}\psi_{j,k}(x) &= 2^{-\frac{j}{2}}\left\{I_{[2^j k\le x<\frac{(2k+1)2^j}{2}]} - I_{[\frac{(2k+1)2^j}{2}\le x<(k+1)2^j]}\right\}\\ \psi^2_{j,k}(x) &= 2^{-j}\left\{I_{[2^j k\le x<\frac{(2k+1)2^j}{2}]} + I_{[\frac{(2k+1)2^j}{2}\le x<(k+1)2^j]}\right\}.\end{aligned} \tag{2.3}$$

We suppose that an experimenter plans to approximate a nonlinear regression model by the wavelet equivalent using the Haar wavelet basis. In the next section, we construct robust designs for the Haar wavelet regression model.

A-, Q-optimality

To fix ideas, we proceed step by step to consider the cases $m = 0$, 1 and the general case.

$\mathbf{m = 0}$: Here, $\mathbf{q}^T(x) = (\phi(x),\ \psi(x))$,

$$\mathbf{q}(x)\mathbf{q}^T(x) = \begin{bmatrix} \phi(x) & \psi(x) \\ \psi(x) & \psi^2(x) \end{bmatrix}, \tag{2.4}$$

$$B = \begin{bmatrix} 1 & \int_S \psi(x)m(x)dx \\ \int_S \psi(x)m(x)dx & 1 \end{bmatrix} \tag{2.5}$$

and

$$C = \begin{bmatrix} \int_S m^2(x)dx & \int_S \psi(x)m^2(x)dx \\ \int_S \psi(x)m^2(x)dx & \int_S m^2(x)dx \end{bmatrix}. \tag{2.6}$$

If for simplicity we set

$$e = \int_S \psi(x)m(x)dx,\ h = \int_S \psi(x)m^2(x)dx,\ and\ p = \int_S m^2(x)dx, \tag{2.7}$$

then

$$C - (1+\lambda)B^2 = \begin{bmatrix} p-(1+e^2)(1+\lambda) & h-2e(1+\lambda) \\ h-2e(1+\lambda) & p-(1+e^2)(1+\lambda) \end{bmatrix}. \tag{2.8}$$

There is some evidence, as seen in the theorem below, that no non-symmetric design exist for $\mathcal{L}_Q$, $\mathcal{L}_A$ loss.

Theorem 2.1. *Let y satisfy the biased $m = 0$ Haar regression model. Let $S = [0, 1]$. Then, among absolutely continuous designs ξ, non-symmetric designs are inadmissible for $\mathcal{L}_Q$ and $\mathcal{L}_A$. The uniform design is A- and Q-optimal for any $f \in \mathcal{F}$. Also,* $\min_\xi \max_f \mathcal{L}_Q(f,\xi) \equiv \min_\xi \max_f \mathcal{L}_A(f,\xi) = 2\frac{\sigma^2}{n}$.

Proof : The steps we follow in proving Theorem 2.1 are :

(i) Solve the eigenvalue problem (2.43).

(ii) From (2.35) and (2.42) set $max_f\ \mathcal{L}_Q(f,\xi) = \tau^2(\lambda_{max} + 1) + \frac{\sigma^2}{n} tr B^{-1}$ and minimize with respect to $m(x)$.

(iii) Use the result from (ii) to discuss the inadmissibility of non-symmetric designs.

(iv) Restrict to symmetric designs and show that the minimax robust design is uniform.

Solving for λ in the equation

$$|C - (1+\lambda)B^2| = [p - (1+e^2)(1+\lambda)]^2 - [h - 2e(1+\lambda)]^2 = 0,$$

we obtain two roots λ_1, λ_2 satisfying

$$(\lambda_1 + 1) = \frac{(p-h)}{(1-e)^2}, \quad (\lambda_2 + 1) = \frac{(p+h)}{(1+e)^2}.$$

Note that $|e| < 1$ and

$$\lambda_1 - \lambda_2 = \frac{-2(he^2 - 2pe + h)}{(1-e^2)^2}$$

is non-negative or negative depending on the sign of $(he^2 - 2pe + h)$. Suppose $(he^2 - 2pe + h) \leq 0$, so that $\lambda_{max} = \lambda_1$. We minimize

$$\max_{\mathcal{F}} \mathcal{L}_Q(f,\xi) = \tau^2 \frac{\int_S (1-\psi(x)) m^2(x) dx}{\alpha^2} + \frac{2\sigma^2}{\alpha(2-\alpha)n} \tag{2.9}$$

subject to the constraints

$$(i) \int_S m(x)dx = 1, \quad (ii) \int_S (1-\psi(x)) m(x) dx = \alpha. \tag{2.10}$$

That is, we minimize

$$\int_S [(1-\psi(x))m^2(x) - u(1-\psi(x))m(x) - vm(x)]dx,$$

for some multipliers u and v. We obtain

$$m(x) = \frac{u(1-\psi(x))+v}{2(1-\psi(x))}. \tag{2.11}$$

We note that the "density" (2.56) is not finite if $x \in [0, \frac{1}{2})$. Similarly, if $(he^2 - 2pe + h) \geq 0$ so that $\lambda_{max} = \lambda_2$, we obtain

$$m(x) = \frac{u(1+\psi(x))+v}{2(1+\psi(x))}. \tag{2.12}$$

This "density" is also not finite if $x \in [\frac{1}{2}, 1)$. Thus no non-symmetric density is admissible. Our conjecture is that this is true in the general case as well. So, restricting to absolutely continuous designs with density $m(x)$ which are symmetric about $x = \frac{1}{2}$, we have $e = h = 0$ and $\lambda_{max} = \lambda_1 = \lambda_2 = p - 1$. That is,

$$\lambda_{max} = \int_S m^2(x)dx - 1$$

and $\alpha = 1$. We then minimize

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \tau^2 \int_S m^2(x)dx + \frac{2\sigma^2}{n} \tag{2.13}$$

subject to the constraint

$$\int_S m(x)dx = 1 \tag{2.14}$$

to obtain $m(x) = I_S$, the continuous uniform design. To see this, we observe that if (2.59) holds then

$$\int_S m^2(x)dx = \int_S (m-1)^2(x)dx + 1$$

which is clearly minimized by $m(x) = I_S$.

m = 1: In this case,

$$\mathbf{q}^T(x) = (\phi(x), \psi(x), 2^{\frac{1}{2}}\psi(2x), 2^{\frac{1}{2}}\psi(2x-1))$$

$$\mathbf{q}(x)\mathbf{q}^T(x) = \begin{bmatrix} \phi(x) & \psi(x) & 2^{\frac{1}{2}}\psi(2x) & 2^{\frac{1}{2}}\psi(2x-1) \\ \psi(x) & \psi^2(x) & 2^{\frac{1}{2}}\psi(2x) & -2^{\frac{1}{2}}\psi(2x-1) \\ 2^{\frac{1}{2}}\psi(2x) & 2^{\frac{1}{2}}\psi(2x) & 2\psi^2(2x) & 0 \\ 2^{\frac{1}{2}}\psi(2x-1) & -2^{\frac{1}{2}}\psi(2x-1) & 0 & 2\psi^2(2x-1) \end{bmatrix} \quad (2.15)$$

Restricting to densities which are symmetric about x = $\frac{1}{2}$, we have

$$\begin{gathered} \int_S \psi(x)m(x)dx = \int_S \psi(x)m^2(x)dx = 0 \\ \int_S \psi^2(x)m(x)dx = \int_S m(x)dx = 1 \\ \int_S \psi^2(2x)m(x)dx = \int_0^{\frac{1}{2}} m(x)dx = \frac{1}{2} \\ \int_S \psi^2(2x-1)m(x)dx = \int_{\frac{1}{2}}^1 m(x)dx = \frac{1}{2}. \end{gathered} \quad (2.16)$$

Also,

$$\begin{gathered} \int_S \psi(2x)m(x)dx = \int_0^{\frac{1}{4}} m(x)dx - \int_{\frac{1}{4}}^{\frac{1}{2}} m(x)dx \\ \int_S \psi(2x-1)m(x)dx = \int_{\frac{1}{2}}^{\frac{3}{4}} m(x)dx - \int_{\frac{3}{4}}^{1} m(x)dx. \end{gathered} \quad (2.17)$$

It follows that

$$\int_S \psi(2x)m(x)dx = -\int_S \psi(2x-1)m(x)dx \quad (2.18)$$

and

$$B = \begin{bmatrix} 1 & 0 & \alpha & -\alpha \\ 0 & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & 0 \\ -\alpha & \alpha & 0 & 1 \end{bmatrix} \quad \textit{where} \ \ \alpha = 2^{\frac{1}{2}} \int_S \psi(2x)m(x)dx. \quad (2.19)$$

We also observe that

$$\begin{gathered} \int_S \psi^2(2x)m^2(x)dx = \int_S \psi^2(2x-1)m^2(x)dx \\ \int_S \psi(2x)m^2(x)dx = -\int_S \psi(2x-1)m^2(x)dx \\ \int_S \psi^2(2x)m^2(x)dx = \int_0^{\frac{1}{2}} m^2(x)dx = \int_{\frac{1}{2}}^1 m^2(x)dx \\ 2\int_S \psi^2(2x)m^2(x)dx = 2\int_0^{\frac{1}{2}} m^2(x)dx = \int_S m^2(x)dx. \end{gathered} \quad (2.20)$$

Therefore,

$$C = \begin{bmatrix} \gamma & 0 & \beta & -\beta \\ 0 & \gamma & \beta & \beta \\ \beta & \beta & \gamma & 0 \\ -\beta & \beta & 0 & \gamma \end{bmatrix} \tag{2.21}$$

where

$$\beta = 2^{\frac{1}{2}} \int_S \psi(2x) m^2(x) dx, \quad and\ \gamma = 2 \int_S \psi^2(2x) m^2(x) dx = \int_S m^2(x) dx.$$

Now,

$$B^2 = \begin{bmatrix} 1+2\alpha^2 & 0 & 2\alpha & -2\alpha \\ 0 & 1+2\alpha^2 & 2\alpha & 2\alpha \\ 2\alpha & 2\alpha & 1+2\alpha^2 & 0 \\ -2\alpha & 2\alpha & 0 & 1+2\alpha^2 \end{bmatrix} \tag{2.22}$$

$$C - (1+\lambda)B^2 = \begin{bmatrix} E & : & F \\ .. & .. & .. \\ F^T & : & E \end{bmatrix} \tag{2.23}$$

where

$$\begin{gathered} E = [\gamma - (1+\lambda)(1+2\alpha^2)]I \ \ and \\ F = \begin{bmatrix} \beta - 2(1+\lambda)\alpha & -[\beta - 2(1+\lambda)\alpha] \\ \beta - 2(1+\lambda)\alpha & \beta - 2(1+\lambda)\alpha \end{bmatrix}. \end{gathered} \tag{2.24}$$

So that

$$|C - (1+\lambda)B^2| = |E||E - F^T E^{-1} F|. \tag{2.25}$$

Now,

$$\begin{gathered} |E| = [\gamma - (1+\lambda)(1+2\alpha^2)]^2 \\ E^{-1} = [\gamma - (1+\lambda)(1+2\alpha^2)]^{-1} I \\ F^T E^{-1} F = 2[\gamma - (1+\lambda)(1+2\alpha^2)]^{-1}[\beta - 2(1+\lambda)\alpha]^2 I \\ E - F^T E^{-1} F = \{[\gamma - (1+\lambda)(1+2\alpha^2)] \\ -2[\gamma - (1+\lambda)(1+2\alpha^2)]^{-1}[\beta - 2(1+\lambda)\alpha]^2\} I. \end{gathered} \tag{2.26}$$

Therefore,

$$|C - (1+\lambda)B^2| = \{[\gamma - (1+\lambda)(1+2\alpha^2)]^2 - 2[\beta - 2(1+\lambda)\alpha]^2\}^2. \tag{2.27}$$

For ease of computation, we rewrite the matrix B as

$$B = \begin{bmatrix} I & : & P \\ .. & .. & .. \\ R & : & I \end{bmatrix}$$

with

$$B^{-1} = \begin{bmatrix} I + PH^{-1}R & : & -PH^{-1} \\ .... & .. & ...... \\ -H^{-1}R & : & H^{-1} \end{bmatrix}$$

where

$$H = I - RP = \begin{bmatrix} 1-2\alpha^2 & 0 \\ 0 & 1-2\alpha^2 \end{bmatrix}$$

and

$$I + PH^{-1}R = H^{-1} = (1-2\alpha^2)^{-1}I.$$

So that

$$trB^{-1} = \frac{4}{1-2\alpha^2}$$

Theorem 2.2. *Let y satisfy the biased $m = 1$ Haar regression model. Let $S = [0,1]$. Then, the continuous uniform design is robust minimax among symmetric and absolutely continuous designs ξ and for any $f \in \mathcal{F}$. Also,* $\min_\xi \max_f \mathcal{L}_Q(f,\xi) \equiv \min_\xi \max_f \mathcal{L}_A(f,\xi) = 4\frac{\sigma^2}{n}$.

Proof : The proof involves restricting to symmetric designs and using the steps outlined in the proof of Theorem 2.1. We begin by solving the eigenvalue equation

$$|C - (1+\lambda)B^2| = [\gamma - (1+\lambda)(1+2\alpha^2)]^2 - 2[\beta - 2(1+\lambda)\alpha]^2 = 0$$

to obtain

$$(\lambda_1 + 1) = \frac{\beta\sqrt{2} - \gamma}{2\alpha\sqrt{2} - (1+2\alpha^2)} \tag{2.28}$$

and

$$(\lambda_2 + 1) = \frac{\beta\sqrt{2} + \gamma}{2\alpha\sqrt{2} + (1 + 2\alpha^2)}. \tag{2.29}$$

Next, we find λ_{max}. To do this, we consider the difference

$$\lambda_1 - \lambda_2 = \frac{-2\sqrt{2}(2\alpha^2\beta - 2\alpha\gamma + \beta)}{(1 - 2\alpha^2)^2}. \tag{2.30}$$

From the above expression, $\lambda_{max} = \lambda_1$ if $f(\alpha) = 2\alpha^2\beta - 2\alpha\gamma + \beta \leq 0$. Otherwise, $\lambda_{max} = \lambda_2$. Let us assume, for a moment, that $\lambda_{max} = \lambda_2$. It follows that,

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \tau^2(\lambda_2 + 1) + \frac{\sigma^2}{n} trB^{-1}(\xi). \tag{2.31}$$

Using the expressions for λ_2 and trB^{-1} we obtain

$$\begin{aligned} \max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) &= \tau^2(\lambda_2 + 1) + \frac{4\sigma^2}{(1 - 2\alpha^2)n} \\ &= \tau^2 \frac{\beta\sqrt{2} + \gamma}{2\alpha\sqrt{2} + (1 + 2\alpha^2)} + \frac{4\sigma^2}{(1 - 2\alpha^2)n}. \end{aligned} \tag{2.32}$$

We then minimize,

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \tau^2 \frac{\int_S (1 + 2\psi(2x)) m^2(x) dx}{2\alpha\sqrt{2} + (1 + 2\alpha^2)} + \frac{4\sigma^2}{(1 - 2\alpha^2)n} \tag{2.33}$$

$$\textit{subject to } \ (i) \int_S m(x) dx = 1, \ (ii)\ 2^{\frac{1}{2}} \int_S \psi(2x) m(x) dx = \alpha \tag{2.34}$$

and from symmetry

$$(iii) \int_S I_{[\frac{1}{4} \leq x < \frac{1}{2}]} m(x) dx - \int_S I_{[\frac{1}{2} \leq x < \frac{3}{4}]} m(x) dx = 0. \tag{2.35}$$

Prior evaluation of the multipliers associated with the constraints show that the second symmetry constraint

$$\int_S I_{[0 \leq x < \frac{1}{4}]} m(x) dx - \int_S I_{[\frac{3}{4} \leq x < 1]} m(x) dx = 0$$

is inactive or nonbinding in the sense that the multiplier associated with this constraint is zero. It is therefore dropped from the list of constraints. For some multipliers u, v, w, we minimize

$$\begin{aligned} &\int_S (1 + 2\psi(2x)) m^2(x) dx - \int_S (w + \sqrt{2} u \psi(2x)) m(x) dx \\ &\qquad - \int_S v (I_{[\frac{1}{4} \leq x < \frac{1}{2}]} - I_{[\frac{1}{2} \leq x < \frac{3}{4}]}) m(x) dx \end{aligned} \tag{2.36}$$

to obtain

$$m(x) = \frac{\sqrt{2}u\psi(2x) + v(I_{[\frac{1}{4}\leq x<\frac{1}{2}]} - I_{[\frac{1}{2}\leq x<\frac{3}{4}]}) + w}{2(1+2\psi(2x))}. \tag{2.37}$$

We rewrite (2.82) as

$$m(x) = \frac{1}{6}(\sqrt{2}u + w)I_{[0\leq x<\frac{1}{4}]} + \frac{1}{2}(\sqrt{2}u - v - w)I_{[\frac{1}{4}\leq x<\frac{1}{2}]} \\ +\frac{1}{2}(w - v)I_{[\frac{1}{2}\leq x<\frac{3}{4}]} + \frac{1}{2}wI_{[\frac{3}{4}\leq x\leq 1]}. \tag{2.38}$$

Determining u, v and w to satisfy (2.79) and (2.80) we obtain

$$u = 2\sqrt{2} + 4\alpha,\ v = 4\sqrt{2}\alpha \text{ and } w = 2 + 2\sqrt{2}\alpha. \tag{2.39}$$

By substitution, the optimal design density is given by

$$m(x) = (1 + \alpha\sqrt{2})[I_{[0\leq x<\frac{1}{4}]} + I_{[\frac{3}{4}\leq x\leq 1]}] \\ +(1 - \alpha\sqrt{2})[I_{[\frac{1}{4}\leq x<\frac{1}{2}]} + I_{[\frac{1}{2}\leq x<\frac{3}{4}]}]. \tag{2.40}$$

We evaluate $f(\alpha) = 2\alpha^2\beta - 2\alpha\gamma + \beta$. Substituting for $m(x)$, we have

$$\beta = 2^{\frac{1}{2}}\left\{\int_0^{\frac{1}{4}} m^2(x)dx - \int_{\frac{1}{4}}^{\frac{1}{2}} m^2(x)dx\right\} = 2\alpha \text{ and} \tag{2.41}$$

$$\gamma = 2\left\{\int_0^{\frac{1}{4}} m^2(x)dx + \int_{\frac{1}{4}}^{\frac{1}{2}} m^2(x)dx\right\} = 1 + 2\alpha^2. \tag{2.42}$$

Therefore, $f(\alpha) = 0$. It follows from (2.75) that $\lambda_{max} = \lambda_2 = \lambda_1$ for $m(x)$ defined by (2.85). It turns out that using λ_1 or λ_2 as λ_{max} yields the same optimal design density (2.85). Substituting β and γ into (2.77) we have that

$$\min_{\xi} \max_{\mathcal{F}} \mathcal{L}_Q(f,\xi) = \frac{4\sigma^2}{(1-2\alpha^2)n} + \tau^2 \tag{2.43}$$

which we minimize with respect to α. Obviously, $\alpha = 0$, $\tau = 0$ minimizes (2.88). It follows that the uniform design is Q-optimal and

$$\min_{\xi} \max_{\mathcal{F}} \mathcal{L}_Q(f,\xi) = \frac{4\sigma^2}{n}. \tag{2.44}$$

Remark : We observe, from (2.85), that the minimax design for $m = 1$ is actually the sum of uniform designs over each of the 2^{m+1} intervals.

General Case

The designs which are optimal for any m are presented in Lemma 2.1 and Theorem 2.3. Lemma 2.1 provides the design ξ_0 which minimizes the *trace* of the covariance matrix $B^{-1}(\xi)$. The remark that follows shows that this design also maximizes the *determinant* of $B(\xi)$.

Lemma 2.1. *For the model (2.1) approximated by the wavelet model (2.9) with scaling function $\phi(x)$ and primary wavelet $\psi(x)$ defined by (1.52) and (2.47) respectively, any design ξ_0 with $B(\xi_0) = I_{2^{m+1}}$ minimizes $tr\{B^{-1}(\xi)\}$. In particular, any design ξ_0^* which places equal mass $2^{-(m+1)}$ in the 2^{m+1} intervals*

$$\{[(i-1)2^{-(m+1)},\ 2^{-(m+1)}i)\}_{i=1,2,\ldots,2^{m+1}} \tag{2.45}$$

is Q-optimal.

Proof : Consider the convex combination

$$\xi_t \ = \ (1-t)\xi_0 \ + \ t\xi_1, \text{ for any } \xi_1 \textit{ on } S$$

and define

$$\rho(t) \ = \ trB^{-1}(\xi_t).$$

Then $\rho(t)$ is a convex function (see Fedorov (1972) Theorem 2.9.1) and is minimized at t $= 0$ if and only if $\rho'(0) \geq 0$, for all ξ_1. In what follows, we evaluate $\rho'(0)$ and apply the condition $B(\xi_0) = I_{2^{m+1}}$. We then see that the result follows trivially if $||\mathbf{q}(x)||^2$ is a constant for all $x \in [0,1)$. We conclude the proof of the first part by using the properties of the Haar wavelets to show that $||\mathbf{q}(x)||^2 = 2^{m+1}$ for all $x \in [0,1)$. Finally, we show that $B(\xi_0^*) = I_{2^{m+1}}$. Now,

$$\rho'(t) = \frac{d}{dt}trB^{-1}(\xi_t) = tr\frac{d}{dt}B^{-1}(\xi_t)$$

where

$$\frac{d}{dt}B^{-1}(\xi_t) = -B^{-1}(\xi_t)\frac{d}{dt}B(\xi_t)B^{-1}(\xi_t).$$

Therefore,

$$\begin{aligned} \rho'(t) &= -tr\left\{B^{-1}(\xi_t)\frac{d}{dt}B(\xi_t)B^{-1}(\xi_t)\right\} \\ &= -tr\left\{B^{-1}(\xi_t)\int_S \mathbf{q}(x)\mathbf{q}^T(x)d(\xi_1-\xi_0)B^{-1}(\xi_t)\right\}. \end{aligned} \quad (2.46)$$

Now, if $B(\xi_0) = I_{2^{m+1}}$ then

$$\begin{aligned} \rho'(0) &= -tr\int_S \mathbf{q}(x)\mathbf{q}^T(x)d(\xi_1-\xi_0) \\ &= tr\int_S \mathbf{q}(x)\mathbf{q}^T(x)d(\xi_0-\xi_1). \end{aligned} \quad (2.47)$$

We now show that

$$\rho'(0) = tr\int_S \mathbf{q}(x)\mathbf{q}^T(x)d(\xi_0-\xi_1) \;\geq\; 0 \;\; for\ all\ \xi_1\ on\ S.$$

That is,

$$\int_S \mathbf{q}^T(x)\mathbf{q}(x)d\xi = \sum_{i=1}^{2^{m+1}}\int_S q_i^2(x)d\xi(x)$$

is maximized by ξ_0. The result follows if we can show that $\mathbf{q}^T(x)\mathbf{q}(x)$ is constant for all $x \in S$. From the definition of $\phi(x)$ and $\psi(x)$, we have that $q_i(1) = 0$ for all i. It follows from our assumption that $\xi(1) = 0$. Therefore, any design which concentrates mass at 1, cannot maximize $\int_S \mathbf{q}^T(x)\mathbf{q}(x)d\xi$. Hence, it is sufficient to show that $\mathbf{q}^T(x)\mathbf{q}(x)$ is constant for all $x \in [0,1)$.

At any level l, $(l = 0, 1, ..., m)$, there are 2^l ψ-functions and 2^{l+1} intervals. For any arbitrarily choosen $x_0 \in$ [0,1), only one of these ψ-functions is non-zero at a given level l, since the intervals are disjoint and x_0 belongs to one and only one of the intervals. In fact, the only non-zero ψ-function takes the value $\pm 2^{l/2}$. Therefore, for any $x_0 \in$ [0,1),

$$\sum_{k=0}^{2^l-1}\psi^2_{-l,k}(x_0) = 2^l \quad (2.48)$$

and

$$\mathbf{q}^T(x_0)\mathbf{q}(x_0) = \sum_{i=1}^{2^{m+1}} q_i^2(x_0)$$
$$= \phi^2(x_0) + \sum_{l=0}^{m}\sum_{k=0}^{2^l-1} \psi_{-l,k}^2(x_0)$$
$$= 1 \ + \ \sum_{l=0}^{m} 2^l = 2^{m+1}. \qquad (2.49)$$

Since x_0 is arbitrary, for any $x \in [0,1)$, $\mathbf{q}^T(x)\mathbf{q}(x) = 2^{m+1}$ and the first part of the theorem is proved.

To complete the proof we show that any design ξ_0^* which places equal mass $2^{-(m+1)}$ in the intervals (2.90) has the property $B(\xi_0^*) = I_{2^{m+1}}$. Now, at any level l, $(l = 0, 1, .., m)$, the diagonal elements of $\mathbf{q}(x)\mathbf{q}^T(x)$ are

$$\psi_{-l,k}^2(x) = 2^l \left\{ I_{[2^{-l}k \le x < (2k+1)2^{-(l+1)}]} + I_{[(2k+1)2^{-(l+1)} \le x < (k+1)2^{-l}]} \right\}$$

and the off diagonal elements are either zero or of the form

$$h[I_{[a_1,b_1)}(x) - I_{[a_2,b_2)}]$$

where h, a_1, a_2, b_1 and b_2 are constants such that $b_2 - a_2 = b_1 - a_1 = 2^{-(l+1)}$. So,

$$B(\xi_0^*) = \int_S \mathbf{q}(x)\mathbf{q}^T(x)d\xi_0^* = I_{2^{m+1}}.$$

Remark: In Fedorov (1972) it has been proved that the following assertions are equivalent:

(1) *the design* ξ^* *minimizes* $trB^{-1}(\xi)$

(2) *the design* ξ^* *minimizes* $max_x \phi(x,\xi)$, *where*

$$\phi(x,\xi) = tr[B^{-1}(\xi)\mathbf{q}(x)\mathbf{q}^T(x)B^{-1}(\xi)] = \mathbf{q}^T(x)B^{-2}(\xi)\mathbf{q}(x)$$

(3) $\max_x \ \phi(x,\xi^*) \ = \ trB^{-1}(\xi^*)$.

So, if $B^{-2}(\xi^*) \ = \ kB^{-1}(\xi^*)$, for some constant k, ξ^* is also G-optimal and hence D-optimal. Now, B$(\xi_0) = kI$ and $\mathcal{L}_Q \ = \ \mathcal{L}_A$ (see (2.26)). It follows that ξ_0 is simultaneously A-, D-, G- and Q-optimal considering only the variance. The proof of D-optimality of ξ_0 was also given by Herzberg and Traves (1994); we feel that our proof is much simpler.

Theorem 2.3. *For the model described in Lemma 2.1, the uniform design ξ^* minimizes $\mathcal{L}_Q(f,\xi)$ and $\mathcal{L}_A(f,\xi)$, among absolutely continuous designs ξ and for any $f \in \mathcal{F}$. Also,*

$$\mathcal{L}_Q(f,\xi^*) \equiv \mathcal{L}_A(f,\xi^*) = 2^{m+1}\frac{\sigma^2}{n}.$$

Proof: We shall show that ξ^* minimizes the two summands in $\mathcal{L}_Q(f,\xi)$ simultaneously. For the continuous uniform design,

$$\mathbf{b}(f,\xi^*) = \int_S \mathbf{q}(x)f(x)d\xi^*(x) = \int_S \mathbf{q}(x)f(x)dx = \mathbf{0},$$

for all $f \in \mathcal{F}$. In the proof of Theorem 2.1, we have shown that $\mathbf{q}^T(x)\mathbf{q}(x)$ is constant for all $x \in [0,1)$. It follows that the uniform design also maximizes

$$\int_S \mathbf{q}^T(x)\mathbf{q}(x)m(x)dx.$$

This implies that ξ^* minimizes $trB^{-1}(\xi)$. Therefore ξ^* minimizes $\mathcal{L}_Q(f,\xi^*)$ since it minimizes the two summands in $\mathcal{L}_Q(f,\xi^*)$ simultaneously. Under ξ^*, $\mathrm{B}(\xi^*) = I_{2^{m+1}}$ and

$$\mathcal{L}_Q(f,\xi^*) = \frac{\sigma^2}{n}trB^{-1}(\xi^*) = 2^{m+1}\frac{\sigma^2}{n}. \tag{2.50}$$

The results of Lemma 2.1, Theorems 2.3 and 2.6 imply that if any nonlinear regression model is approximated by the Haar wavelet basis, then the robust minimax design is uniform among symmetric and absolutely continuous designs.

D-optimaliy

m = 0: Using the notation in Section 3.1.1 we have

$$|C - B^2 - \lambda B| = \begin{vmatrix} p-(1+e^2)-\lambda & h-2e-e\lambda \\ h-2e-e\lambda & p-(1+e^2)-\lambda \end{vmatrix}. \tag{2.51}$$

Theorem 2.4. *Let y satisfy the biased $m=0$ Haar regression model. Let $S=[0,1]$. Then, among absolutely continuous designs ξ, non-symmetric designs are inadmissible for $\mathcal{L}_D$. The uniform design is D-optimal for any $f \in \mathcal{F}$. Also,* $\min_\xi \max_f \mathcal{L}_D(f,\xi) = \left(\frac{\sigma^2}{n}\right)^2$.

Proof : The main features of this proof involve determining the maximum eigenvalue and minimizing the maximum loss with respect to $m(x)$. To determine the maximum eigenvalue, we first solve $|C - B^2 - \lambda B| = 0$ for λ. We obtain two eigenvalues

$$\lambda_1 = \frac{(p-h)-(1+e)^2}{(1-e)} \tag{2.52}$$

and

$$\lambda_2 = \frac{(p+h)-(1+e)^2}{(1+e)}. \tag{2.53}$$

If $\lambda_{max} = \lambda_1$, we minimize

$$\begin{aligned}\max_{\mathcal{F}} \mathcal{L}_D(f,\xi) &= \left(\frac{\sigma^2}{n}\right)^2 \left\{\frac{1+\frac{n}{\sigma^2}\tau^2\lambda_1}{|B|}\right\} \\ &= \frac{\left(\frac{\sigma^2}{n}\right)^2}{\alpha(2-\alpha)}\left\{1+\frac{\tau^2 n}{\alpha\sigma^2}\left[\int_S (1-\psi(x))m^2(x)dx - (2-\alpha)^2\right]\right\}\end{aligned}$$

subject to the constraints (2.55). That is, we minimize

$$\int_S (1-\psi(x))m^2(x)dx. \tag{2.54}$$

The solution is given by (2.56). Using $\lambda_{max} = \lambda_2$, leads to (2.57). Implying that non-symmetric designs are inadmissible. Therefore restricting to absolutely continuous designs which are symmetric about $x = \frac{1}{2}$, we minimize

$$\max_{\mathcal{F}} \mathcal{L}_D(f,\xi) = \left(\frac{\sigma^2}{n}\right)^2 \left\{1+\frac{\tau^2 n}{\sigma^2}\left[\int_S m^2(x)dx - 1\right]\right\} \tag{2.55}$$

to obtain $m(x) = I_S$, the uniform design. That is,

$$\min_{\xi} \max_{\mathcal{F}} \mathcal{L}_D(f,\xi) = \left(\frac{\sigma^2}{n}\right)^2. \tag{2.56}$$

m = 1: In this case, we restrict to symmetric densities to obtain

$$|C - B^2 - \lambda B| = \begin{vmatrix} U & : & V \\ .. & .. & .. \\ V^T & : & U \end{vmatrix} \tag{2.57}$$

where

$$U = [(\gamma-\lambda)-(1+2\alpha^2)]I \textit{ and } V = \begin{bmatrix} \beta-(2+\lambda)\alpha & -[\beta-(2+\lambda)\alpha] \\ \beta-(2+\lambda)\alpha & \beta-(2+\lambda)\alpha \end{bmatrix}. \tag{2.58}$$

Theorem 2.5. *Let y satisfy the biased $m = 1$ Haar regression model. Let $S = [0, 1]$. Then, the continuous uniform design is robust minimax among symmetric and absolutely continuous designs ξ and for any $f \in \mathcal{F}$. Also,* $\min_\xi \max_f \mathcal{L}_D = \left(\frac{\sigma^2}{n}\right)^4$.

Proof : As usual, we begin by solving

$$|C - B^2 - \lambda B| = \{[(\gamma - \lambda) - (1 + 2\alpha^2)]^2 - 2[\beta - (2 + \lambda)\alpha]^2\}^2 = 0$$

to obtain

$$(\lambda_1 + 1) = \frac{\beta\sqrt{2} - \gamma}{\alpha\sqrt{2} - 1} + \alpha\sqrt{2} \tag{2.59}$$

and

$$(\lambda_2 + 1) = \frac{\beta\sqrt{2} + \gamma}{\alpha\sqrt{2} + 1} - \alpha\sqrt{2}. \tag{2.60}$$

Suppose that $\lambda_{max} = \lambda_2$, then we minimize

$$\begin{aligned}\max_{\mathcal{F}} \mathcal{L}_D(f, \xi) &= \left(\frac{\sigma^2}{n}\right)^4 \frac{1}{(1 - 2\alpha^2)^2}[1 + \frac{\tau^2 n}{\sigma^2}\lambda_2] \\ &= \frac{\left(\frac{\sigma^2}{n}\right)^4}{(1 - 2\alpha^2)^2}\left\{1 + \frac{\tau^2 n}{\sigma^2}\left[\frac{\int_S (1 + 2\psi(2x))m^2(x)dx}{(1 + \alpha\sqrt{2})} - \alpha\sqrt{2} - 1\right]\right\}\end{aligned}$$

subject to the constraints (2.79) and (2.80). That is, we minimize

$$\int_S (1 + 2\psi(2x))m^2(x)dx. \tag{2.61}$$

The solution is given by (2.82). It turns out that using $\lambda_{max} = \lambda_1$ leads to the same solution. Under this solution,

$$\beta = 2\alpha, \ \ \gamma = (1 + 2\alpha^2) \ \ and \ \ \lambda_1 = \lambda_2 = 0.$$

So that

$$\min_\xi \max_{\mathcal{F}} \mathcal{L}_D(f, \xi) = \left(\frac{\sigma^2}{n}\right)^4 \frac{1}{(1 - 2\alpha^2)^2}$$

is minimized by $\alpha = 0$. Therefore, the uniform design $m(x) = I_S$ is D-optimal and

$$\min_{\xi} \max_{\mathcal{F}} \mathcal{L}_D(f, \xi) = \left(\frac{\sigma^2}{n}\right)^4. \tag{2.62}$$

General Case : We state the result as a theorem.

Theorem 2.6. *For the model described in Lemma 2.1, the uniform design ξ^* minimizes $\mathcal{L}_D(f, \xi)$, among absolutely continuous designs ξ and for any $f \in \mathcal{F}$. Also,*

$$\mathcal{L}_D(f, \xi^*) = \left(\frac{\sigma^2}{n}\right)^{2^{m+1}}. \tag{2.63}$$

Proof : The statement of the theorem follows from the fact that ξ^* maximizes $|B(\xi)|$ and minimizes $\mathbf{b}^T B^{-1} \mathbf{b}$ simultaneously as shown in Lemma 2.1 and Theorem 2.3 respectively. To see this, we have shown in Theorem 2.3 that $\mathbf{b}(f, \xi^*) = \mathbf{0}$. Also, under ξ^*, $B(\xi^*) = I$. It therefore follows from Lemma 2.1 and the remarks that ξ^* maximizes $|B(\xi)|$ as well.

Remarks : The optimality of the continuous version ξ^* of ξ_0^* (see Lemma 2.1) stated in Theorems 2.3 and 2.6 is a particularly strong version of minimax robustness. This follows from the fact that

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

for any design ξ and any $f \in \mathcal{F}$ and $\mathcal{L}$ is any of $\mathcal{L}_Q, \mathcal{L}_A, \mathcal{L}_D$. The first inequality follows from Lemma 2.1 and the second inequality follows from the definitions of the loss functions. The equality is derived from (1.16).

3.2. Multiwavelets

For the purpose of our example, we take $N = 2$. We remind ourselves that when $N = 1$, the multiwavelets coincide with the Haar wavelet basis. For $N = 2$, the multiwavelet orthonormal basis for $\mathcal{L}_2([0, 1])$ is given by

$$\{\phi_0, \phi_1\} \cup \{{}_2w_0^{-j,k}(x), {}_2w_1^{-j,k}(x);\; j \leq 0,\; 0 \leq k \leq 2^{-j} - 1\} \tag{2.64}$$

where ${}_Nw_l^{j,k}(x) = 2^{-\frac{j}{2}}{}_Nw_l(2^{-j}x-k), \quad l=0,1$. The scaling functions and primary wavelets are defined by (1.54), (1.55), (1.56) and (1.57). For multiwavelets the representations discussed in Section 1 of this chapter can be written as,

$$\eta(x) = \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} c_{jk}\ {}_2w_0^{-j,k}(x) + \sum_{j=0}^{p}\sum_{k=0}^{2^j-1} e_{jk}\ {}_2w_1^{-j,k}(x) + f(x) \qquad (2.65)$$

and

$$\begin{aligned}\eta(x) &= \sum_{l=0}^{r} d_l\phi_{0,l}(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} c_{jk}\ {}_2w_0^{-j,k}(x) \\ &+ \sum_{j=0}^{p}\sum_{k=0}^{2^j-1} e_{jk}\ {}_2w_1^{-j,k}(x) + f(x). \qquad (2.66)\end{aligned}$$

We then write the model

$$y(x) = \eta(x) + \varepsilon$$

as

$$y(x) = \mathbf{q}^T(x)\boldsymbol{\beta}_0 + f(x) + \varepsilon \qquad (2.67)$$

where the vectors $\mathbf{q}(x)$ and $\boldsymbol{\beta}_0$ are defined in such a way that (2.112) is equivalent to one of the representations in (2.110) and (2.111). We limit our consideration to the representation described by (2.110) with $m=p=0$ due to the complexity of the eigenvalues arising from the maximization problem described previously for values of m, $p>0$. In this case,

$$\mathbf{q}^T(x) = ({}_2w_0(x), {}_2w_1(x)), \text{ and } \beta_0^T = (c_{00}, e_{00}). \qquad (2.68)$$

From previous results,

$$Var(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n}B^{-1}(\xi), \ \ bias(\hat{\boldsymbol{\beta}}) = B^{-1}(\xi)\mathbf{b}(f,\xi)$$

where

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

A-, Q-optimality

Recall that the Q-optimality problem is to minimize the maximum integrated mean squared error over some design space, where the maximum is evaluated over $\mathcal{F}$ the $\mathcal{L}_2$ contamination neighbourhood. We also recall that for orthonormal wavelets the Q-optimality and A-optimality problems are equivalent since

$$A = \int_S \mathbf{q}(x)\mathbf{q}^T(x)dx = I, \text{ the identity matrix.}$$

So, our problem is to solve

$$\min_{\xi} \max_{f\in\mathcal{F}} \mathcal{L}_Q(f,\xi)$$

where the loss function is defined by

$$\mathcal{L}_Q(f,\xi) = \mathbf{b}^T B^{-2}\mathbf{b} + \frac{\sigma^2}{n} trB^{-1} + \int_S f^2(x)dx. \tag{2.69}$$

From previous results, maximizing $\mathcal{L}_Q(f,\xi)$ over $\mathcal{F}$ involves evaluating the maximum characteristic root of the matrix $G^{\frac{1}{2}}B^{-2}G^{\frac{1}{2}}$. That is, we solve the equations $|G-\lambda B^2| = 0$ where $G = C - B^2$ and

$$C = \int_S \mathbf{q}(x)\mathbf{q}^T(x)m^2(x)dx.$$

Using (2.113) we have

$$\mathbf{q}(x)\mathbf{q}^T(x) = \begin{bmatrix} {}_2w_0^2(x) & {}_2w_0(x){}_2w_1(x) \\ {}_2w_0(x){}_2w_1(x) & {}_2w_1^2(x) \end{bmatrix}. \tag{2.70}$$

Restricting to densities that are symmetric about $x = \frac{1}{2}$ we can show that

$$\begin{array}{c} \int_0^{\frac{1}{2}} xm(x)dx = \frac{1}{2} - \int_{\frac{1}{2}}^1 xm(x)dx \\ \int_0^{\frac{1}{2}} x^2m(x)dx = \frac{1}{2} - 2\int_{\frac{1}{2}}^1 xm(x)dx + \int_{\frac{1}{2}}^1 x^2m(x)dx. \end{array} \tag{2.71}$$

Results simlar to (2.116) also hold if $m(x)$ is replaced by $\mathrm{m}^2(x)$. Using the result in (2.116), it can be shown that the following hold:

$$\begin{array}{c} \int_S {}_2w_0^2(x)m(x)dx = 3(32\beta - 48\alpha + 9) \\ \int_S {}_2w_0(x){}_2w_1(x)m(x)dx = 0 \\ \int_S {}_2w_1^2(x)m(x)dx = 72\beta - 120\alpha + 25 \\ \int_S {}_2w_0^2(x)m^2(x)dx = 6(16d - 24c + 9a) \\ \int_S {}_2w_0(x){}_2w_1(x)m^2(x)dx = 0 \\ \int_S {}_2w_1^2(x)m^2(x)dx = 2(36d - 60c + 25a) \end{array} \tag{2.72}$$

where

$$\begin{array}{c}\alpha = \int_{\frac{1}{2}}^{1} xm(x)dx,\ \beta = \int_{\frac{1}{2}}^{1} x^2m(x)dx,\ a = \int_{\frac{1}{2}}^{1} m^2(x)dx \\ c = \int_{\frac{1}{2}}^{1} xm^2(x)dx \text{ and } d = \int_{\frac{1}{2}}^{1} x^2m^2(x)dx.\end{array} \tag{2.73}$$

It follows that the matrices B and C can be written as

$$B = \begin{bmatrix} 3(32\beta - 48\alpha + 9) & 0 \\ 0 & 72\beta - 120\alpha + 25 \end{bmatrix} \tag{2.74}$$

$$C = \begin{bmatrix} 6(16d - 24c + 9a) & 0 \\ 0 & 2(36d - 60c + 25a) \end{bmatrix}. \tag{2.75}$$

The main idea behind (2.116) and (2.117) is to transform the problem from $S = [0, 1]$ to a smaller space $[0.5, 1]$ using the symmetry constraint. Solving the eigenvalue equation $|C - (1 + \lambda)B^2| = 0$ we obtain

$$\begin{array}{c}6(16d - 24c + 9a) - 9(1 + \lambda)(32\beta - 48\alpha + 9)^2 = 0 \\ 2(36d - 60c + 25a) - (1 + \lambda)(72\beta - 120\alpha + 25)^2 = 0.\end{array} \tag{2.76}$$

So, the eigenvalues of the matrix $G^{\frac{1}{2}}B^{-2}G^{\frac{1}{2}}$ are given by

$$\lambda_1 = \frac{2(16d - 24c + 9a)}{3(32\beta - 48\alpha + 9)^2} - 1 \tag{2.77}$$

and

$$\lambda_2 = \frac{2(36d - 60c + 25a)}{(72\beta - 120\alpha + 25)^2} - 1. \tag{2.78}$$

From this point we proceed as follows:

(i) assume that $\lambda_{max} = \lambda_1$

(ii) find the density $\mathrm{m}_0(x)$ minimizing the maximum loss; and

(iii) check the condition $\lambda_1(\mathrm{m}_0(x)) \geq \lambda_2(\mathrm{m}_0(x))$.

If step (iii) fails, we then hope that $\lambda_{max} = \lambda_2$. Taking $\lambda_{max} = \lambda_1$, we minimize

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \tau^2(\lambda_1 + 1) + \frac{\sigma^2}{n} trB^{-1} \tag{2.79}$$

subject to the conditions

$$(i) \int_{\frac{1}{2}}^{1} m(x)dx = \frac{1}{2},\ (ii) \int_{\frac{1}{2}}^{1} xm(x)dx = \alpha,\ (iii) \int_{\frac{1}{2}}^{1} x^2m(x)dx = \beta. \tag{2.80}$$

Using the expression for λ_1 we have

$$\max_{\mathcal{F}} \mathcal{L}_Q(f,\xi) = \frac{2\tau^2 \int_{\frac{1}{2}}^{1}(4x-3)^2 m^2(x)dx}{3(32\beta-48\alpha+9)^2} + \frac{\sigma^2}{n} trB^{-1}. \quad (2.81)$$

Define

$$\begin{aligned} m_t(x) &= (1-t)m_0(x) + tm_1(x) \\ \Delta(t) &= \int_{\frac{1}{2}}^{1} \{(4x-3)^2 m_t^2(x) + 2uwm_t(x) + 8vwxm_t(x) - 32wx^2 m_t(x)\}dx \end{aligned}$$

where $m_0(x)$, $m_1(x)$ satisfy (2.125), for some multipliers u, v and w. It is not too difficult to check that $\Delta''(t) \geq 0$. That is, $\Delta(t)$ is convex with a minimum at t = 0 if and only if $\Delta'(0) \geq 0$ for all $m_1(x)$, where

$$\Delta'(0) = 2\int_{\frac{1}{2}}^{1} \{(4x-3)^2 m_0(x) + uw + 4vwx - 16wx^2\}(m_1 - m_0)(x)dx. \quad (2.82)$$

It follows from (2.127) that the minimizing $m_0(x)$ must satisfy the equation

$$(4x-3)^2 m_0(x) + uw + 4vwx - 16wx^2 = 0, \;\; m_0(x) \geq 0.$$

That is,

$$m_0(x) = \left(\frac{16wx^2 - 4vwx - uw}{(4x-3)^2}\right)^+, \; x \in \left[\frac{1}{2}, 1\right] \quad (2.83)$$

with u, v and w determined to satisfy (2.125). We observe that if we write

$$16x^2 - 4vx - u = (4x-3)^2 - (v-6)(4x-3) - (u+3v-9)$$

and set $r = v - 6$, $\; t = u + 3v - 9$ then

$$\begin{aligned} m_0(x) &= w\left(1 - \frac{r}{(4x-3)} - \frac{t}{(4x-3)^2}\right)^+, \; x \in \left[\frac{1}{2}, 1\right] \\ &= max\left\{0, w\left(1 - \frac{r}{(4x-3)} - \frac{t}{(4x-3)^2}\right)\right\}, \; x \in \left[\frac{1}{2}, 1\right] \quad (2.84) \end{aligned}$$

where $w \geq 0$, $t \geq 0$ and r is arbitrary. For mathematical simplicity, we transform $m_0(x)$ by setting $y = 4x - 3$ to obtain

$$p_0(y) = \frac{1}{4} m_0\left(\frac{y+3}{4}\right) = \frac{w}{4}\left(1 - \frac{r}{y} - \frac{t}{y^2}\right)^+, \; y \in [-1, 1]. \quad (2.85)$$

We now express the conditions (2.125) in terms of $p_0(y)$. It is easy to show that

$$\int_{\frac{1}{2}}^{1} m_0(x)dx = \int_{-1}^{1} p_0(y)dy = \frac{1}{2}. \tag{2.86}$$

From (2.125) (ii) we have

$$\int_{\frac{1}{2}}^{1} xm_0(x)dx = \int_{-1}^{1} \left(\frac{y+3}{4}\right) m_0\left(\frac{y+3}{4}\right)\frac{dy}{4} = \alpha.$$

Use (2.131) and simplify to obtain

$$\int_{-1}^{1} yp_0(y)dy = \frac{1}{2}(8\alpha - 3). \tag{2.87}$$

Furthermore, (2.125) (iii) yields

$$\int_{\frac{1}{2}}^{1} x^2 m_0(x)dx = \int_{-1}^{1} \left(\frac{y+3}{4}\right)^2 m_0\left(\frac{y+3}{4}\right)\frac{dy}{4} = \int_{-1}^{1} \left(\frac{y+3}{4}\right)^2 p_0(y)dy = \beta.$$

Again we use (2.131), (2.132) and simplify to obtain

$$\int_{-1}^{1} y^2 p_0(y)dy = \frac{1}{2}(32\beta - 48\alpha + 9). \tag{2.88}$$

Also,

$$\begin{aligned}\int_{\frac{1}{2}}^{1} (4x-3)^2 m_0^2(x)dx &= \int_{-1}^{1} y^2 m_0^2\left(\frac{y+3}{4}\right)\frac{dy}{4} \\ &= \int_{-1}^{1} y^2 m_0\left(\frac{y+3}{4}\right) p_0(y)dy \\ &= w\int_{-1}^{1} y^2\left(1 - \frac{r}{y} - \frac{t}{y^2}\right) p_0(y)dy.\end{aligned}$$

Use (2.131), (2.132), (2.133) and simplify to obtain

$$\int_{\frac{1}{2}}^{1} (4x-3)^2 m_0^2(x)dx = \frac{w}{2}\{(32\beta - 48\alpha + 9) - r(8\alpha - 3) - t\}. \tag{2.89}$$

Substituting (2.134) into (2.126) we have

$$\min_{\xi}\max_{f} \mathcal{L}_Q(f,\xi) = \tau^2\left\{\frac{w[(32\beta - 48\alpha + 9) - r(8\alpha - 3) - t]}{3(32\beta - 48\alpha + 9)^2} + \nu tr B^{-1}\right\} \tag{2.90}$$

where r, t and w are chosen to satisfy (2.131), (2.132) and (2.133). Furthermore, on the interval $\left[\frac{1}{2}, 1\right]$ the inequality

$$\frac{1}{4} \leq x^2 \leq x \leq 1$$

holds. Multiply through by $m(x)$ and integrate to obtain

$$\frac{1}{8} \leq \beta \leq \alpha \leq \frac{1}{2}. \tag{2.91}$$

Similarly,

$$\frac{m(x)}{2} \leq xm(x), \text{ which implies } \frac{1}{4} \leq \alpha.$$

By the Cauchy-Schwarz inequality we have

$$\frac{1}{16} \leq \alpha^2 \leq \frac{\beta}{2}. \tag{2.92}$$

Together, (2.136) and (2.137) imply that α and β must satisfy the inequality

$$\frac{1}{8} \leq 2\alpha^2 \leq \beta \leq \alpha \leq \frac{1}{2}. \tag{2.93}$$

We summarize the results of this section in the following theorem:

Theorem 2.7. *For $\mathcal{A} = (\alpha, \beta)$ and $\frac{1}{8} \leq 2\alpha^2 \leq \beta \leq \alpha \leq \frac{1}{2}$ define the density*

$$m_0(x; \mathcal{A}) = w\left(1 - \frac{r}{(4x-3)} - \frac{t}{(4x-3)^2}\right)^+ \quad (0.5 \leq x \leq 1)$$

where the non-negative constants $w = w(\mathcal{A})$, $r = r(\mathcal{A})$, $t = t(\mathcal{A})$ are determined to satisfy (2.125). Then $m_0(x; \mathcal{A})$ minimizes $\int_{\frac{1}{2}}^{1}(4x-3)^2 m^2(x)dx$ for fixed $\mathcal{A}$. Define $\mathcal{A}_Q$ to be the minimizer of (2.135). Then $m_0(x; \mathcal{A}_Q)$ is minimax robust for $\mathcal{L}_Q$ and $\mathcal{L}_A$ for those values of ν for which $\lambda_1(m_0(x; \mathcal{A}_Q)) > \lambda_2(m_0(x; \mathcal{A}_Q))$. The eigenvalue inequality holds for $\nu \geq 4.45$.

We observe that if $r = t = 0$, the minimax robust design is uniform with

$$\alpha = \frac{3}{8},\ \beta = \frac{7}{24},\ B = I_{2\times 2} \text{ and } \mathcal{L}_Q(f, \xi) = \frac{2\sigma^2}{n} + \tau^2 \tag{2.94}$$

which is the minimum loss for the Haar wavelet basis. To see that the eigenvalue inequality holds for $\nu \geq 4.45$, we need to carry out the actual computations. Our problem now is to choose α and β to minimize

$$\frac{w[(32\beta - 48\alpha + 9) - r(8\alpha - 3) - t]}{3(32\beta - 48\alpha + 9)^2} + \nu\left\{\frac{1}{3(32\beta - 48\alpha + 9)} + \frac{1}{(72\beta - 120\alpha + 25)}\right\} \tag{2.95}$$

for fixed $\nu = \frac{\sigma^2}{n\tau^2}$ over the range

$$\frac{1}{8} \leq 2\alpha^2 \leq \beta \leq \alpha \leq \frac{1}{2}$$

where r, t and w are determined to satisfy

$$\begin{array}{c} (i)\int_{-1}^{1} p_0(y)dy = \frac{1}{2},\ (ii)\int_{-1}^{1} yp_0(y)dy = \frac{1}{2}(8\alpha - 3) \\ (iii)\int_{-1}^{1} y^2 p_0(y)dy = \frac{1}{2}(32\beta - 48\alpha + 9). \end{array} \tag{2.96}$$

In addition, we require that $\lambda_1(p_0(y)) \geq \lambda_2(p_0(y))$ at $w^* = w(\alpha^*, \beta^*)$, $r^* = r(\alpha^*, \beta^*)$ and $t^* = t(\alpha^*, \beta^*)$ where α^*, β^* are the values of α, β minimizing (2.140). Expressing λ_2 in terms of $(4x - 3)$ we have

$$\lambda_2(m_0(x)) = \frac{9\int_{\frac{1}{2}}^{1}(4x-3)^2 m_0^2(x)dx - 24c + 19a}{2(72\beta - 120\alpha + 25)^2} - 1 \tag{2.97}$$

$$\lambda_2(p_0(y)) = \frac{2\int_{-1}^{1}(3y-1)^2 p_0^2(y)dy}{(72\beta - 120\alpha + 25)^2} - 1. \tag{2.98}$$

The definition of $p_0(y)$ indicates that there exist some constants k, l such that $p_0(y) = 0$, $k \leq y \leq l$ where $-1 < k < 0$ and $0 < l < 1$. That is,

$$1 - \frac{r}{y} - \frac{t}{y^2} \leq 0, \quad k \leq y \leq l$$

with equality at k and l. This implies that k and l are given by

$$k = \frac{r - \sqrt{r^2 + 4t}}{2} \text{ and } l = \frac{r + \sqrt{r^2 + 4t}}{2}. \tag{2.99}$$

From (2.141) (i) we have

$$\frac{1}{2} = \int_{-1}^{1} p_0(y)dy = \frac{w}{4}\left[\int_{-1}^{k}\left(1 - \frac{r}{y} - \frac{t}{y^2}\right)dy + \int_{l}^{1}\left(1 - \frac{r}{y} - \frac{t}{y^2}\right)dy\right].$$

We simplify to obtain

$$w\left[(k-l+2)+rln\left(\left|\frac{l}{k}\right|\right)+t\left(2+\frac{1}{k}-\frac{1}{l}\right)\right]=2. \qquad (2.100)$$

Similarly, we obtain the following equations from (2.141) (ii) and (iii) respectively,

$$w\left[(k^2-l^2)+2r(l-k-2)+2tln\left(\left|\frac{l}{k}\right|\right)\right]=4(8\alpha-3) \qquad (2.101)$$

$$w[2(k^3-l^3+2)+3r(l^2-k^2)+6t(l-k-2)]=12(32\beta-48\alpha+9). \quad (2.102)$$

The problem defined above is solved numerically in the following way:

(a) Solve (2.144) for r and t to obtain

$$r=k+l, \quad t=-lk, \quad -1<k<0, \; 0<l<1. \qquad (2.103)$$

(b) Rewrite (2.145), (2.146) and (2.147) as follows:

$$w = \frac{2}{\left[(k-l+2)+rln\left(\left|\frac{l}{k}\right|\right)+t\left(2+\frac{1}{k}-\frac{1}{l}\right)\right]} \qquad (2.104)$$

$$\alpha = \frac{3}{8}+\frac{w}{32}\left[(k^2-l^2)+2r(l-k-2)+2tln\left(\left|\frac{l}{k}\right|\right)\right] \qquad (2.105)$$

$$\beta = \frac{3}{2}\alpha-\frac{9}{32}+\frac{w}{384}[2(k^3-l^3+2)+3r(l^2-k^2)+6t(l-k-2)] \qquad (2.106)$$

(c) Use (2.148), (2.149), (2.150) and (2.151) to write the objective function (2.140) as a function of k and l.

(d) Use the S-plus function nlminb($\cdot$) ("nonlinear minimization with box constraints") in an S-plus program to minimize the objective function with respect to k and l for fixed values of ν where $-1<k<0$ and $0<l<1$.

(e) Solve for r, t, w, α and β from (2.148) - (2.151) and check that the side conditions (2.138) and (2.141) are satisfied.

(f) Evaluate the eigenvalues $\lambda_1(p_0(y))$ and $\lambda_2(p_0(y))$ to check if $\lambda_1(p_0(y)) \geq \lambda_2(p_0(y))$.

Table 1: Some Parameter Values Minimizing (2.140)

ν	k	l	r	t	w	α	β	λ_1	λ_2
0.005	-0.0378	0.0390	0.001	0.002	1.081	0.375	0.292	1.000	1.013
0.05	-0.1156	0.1259	0.01	0.015	1.293	0.374	0.292	1.006	1.043
0.5	-0.3105	0.3682	0.058	0.114	2.270	0.364	0.282	1.093	1.164
1.0	-0.3939	0.4753	0.081	0.187	3.071	0.357	0.274	1.187	1.258
4.0	-0.5670	0.6920	0.125	0.392	6.858	0.335	0.244	1.619	1.629
4.45	-0.5797	0.7072	0.128	0.410	7.366	0.333	0.241	1.672	1.672
5.0	-0.5933	0.7234	0.130	0.429	7.975	0.330	0.238	1.736	1.722
10	-0.6694	0.8109	0.142	0.543	13.15	0.314	0.215	2.233	2.092
20	-0.7358	0.8824	0.147	0.649	22.54	0.295	0.188	3.017	2.621
50	-0.8087	0.9556	0.147	0.773	48.17	0.269	0.150	4.775	3.673

Some values of the constants and eigenvalues are shown in Table 1. Figure 2 shows the minimax densities for $\nu = 6$ and 20.

It turns out that $\lambda_1(p_0(y)) \geq \lambda_2(p_0(y))$ for values of $\nu \geq 4.45$ (approximated to two decimal places) and fails otherwise. So, we hope that $\lambda_{max} = \lambda_2$ for $\nu < 4.45$ and minimize

$$\max_f \mathcal{L}_Q(f,\xi) = \frac{2\tau^2 \int_{\frac{1}{2}}^{1}(6x-5)^2 m^2(x)dx}{(72\beta - 120\alpha + 25)^2} + \frac{\sigma^2}{n} trB^{-1} \tag{2.107}$$

subject to (2.125) to obtain

$$m_*(x) = w1\left(1 - \frac{r1}{(6x-5)} - \frac{t1}{(6x-5)^2}\right)^+, \quad x \in \left[\frac{1}{2}, 1\right] \tag{2.108}$$

where $w1 \geq 0$, $t1 \geq 0$ and $r1$ is an arbitrary constant. We proceed as before to put $z = 6x - 5$ in (2.153) to obtain

$$p_1(z) = \frac{w1}{6}\left(1 - \frac{r1}{z} - \frac{t1}{z^2}\right)^+, \quad z \in [-2, 1]. \tag{2.109}$$

The side conditions (2.125) now take the form

$$\begin{array}{c} (i) \int_{-2}^{1} p_1(z)dz = \frac{1}{2} \; (ii) \int_{-2}^{1} zp_1(z)dz = \frac{1}{2}(12\alpha - 5) \\ (iii) \int_{-2}^{1} z^2 p_1(z)dz = \frac{1}{2}(72\beta - 120\alpha + 25). \end{array} \tag{2.110}$$

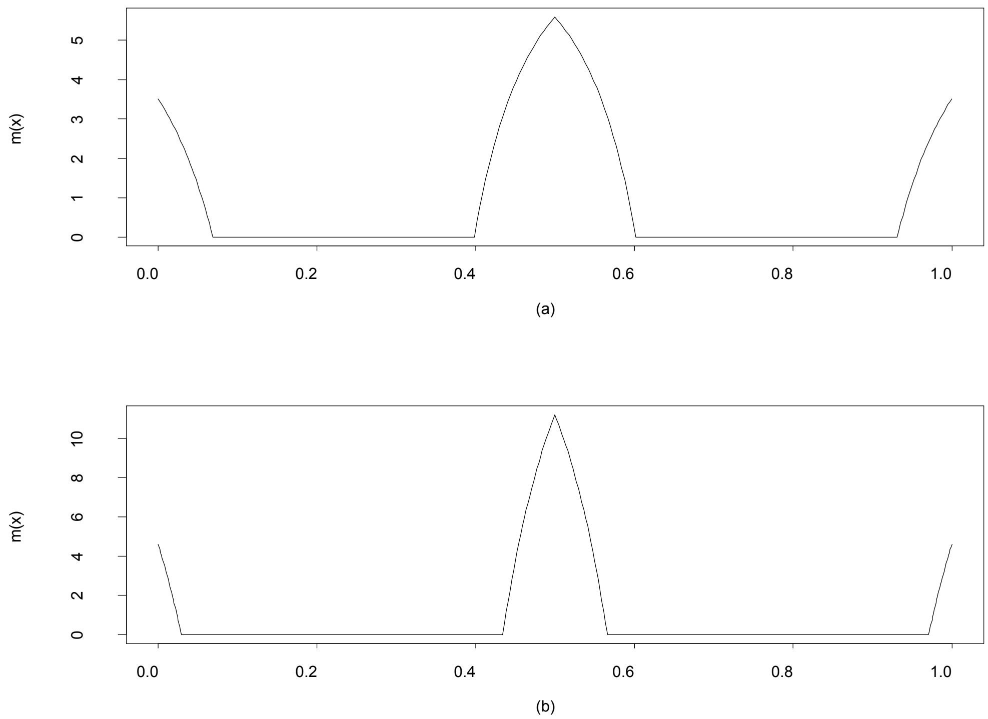


Figure 2: Q-optimal Density $m_0(x)$: (a) $\nu = 5$; (b) $\nu = 20$.

Also, it can be shown that

$$\int_{\frac{1}{2}}^{1} (6x-5)^2 m_*^2(x)dx = w1 \int_{-2}^{1} z^2 \left(1 - \frac{r1}{z} - \frac{t1}{z^2}\right) p_1(z)dz$$
$$= \frac{w1}{2}[(72\beta - 120\alpha + 25) - r1(12\alpha - 5) - t1] \quad (2.111)$$

Substituting (2.156) in (2.152) we have

$$\min_{\xi} \max_{f} \mathcal{L}_Q(f, \xi) = \tau^2 \left\{ \frac{w1[(72\beta - 120\alpha + 25) - r1(12\alpha - 5) - t1]}{(72\beta - 120\alpha + 25)^2} + \nu tr B^{-1} \right\} \quad (2.112)$$

where $r1$, $t1$ and $w1$ are chosen to satisfy (2.155). Proceeding as before, we obtain, from (2.155), the following equations :

$$w1\left[(k-l+3)+r1ln\left(\left|\frac{2l}{k}\right|\right)+t1\left(\frac{3}{2}+\frac{1}{k}-\frac{1}{l}\right)\right]=3 \tag{2.113}$$

$$w1\left[(k^2-l^2-3)+2r1(l-k-3)+2t1ln\left(\left|\frac{2l}{k}\right|\right)\right]=6(12\alpha-5) \tag{2.114}$$

$$\begin{aligned} &w1[2(k^3-l^3+9)+3r1(l^2-k^2+3)+6t1(l-k-3)] \\ &\qquad = 18(72\beta-120\alpha+25). \end{aligned} \tag{2.115}$$

So, we choose α and β to minimize

$$\begin{aligned} &\frac{w1[(72\beta-120\alpha+25)-r1(12\alpha-5)-t1]}{(72\beta-120\alpha+25)^2} \\ &+\nu\left\{\frac{1}{3(32\beta-48\alpha+9)}+\frac{1}{(72\beta-120\alpha+25)}\right\} \end{aligned} \tag{2.116}$$

for fixed values of ν over the range (2.138), with $r1$, $t1$ and $w1$ choosen to satisfy (2.158), (2.159) and (2.160). Some numerical solution to this problem is shown in Table 2. The numerical solution shows that the condition $\lambda_2(p_1(y)) \geq \lambda_1(p_1(y))$ fails to hold for all values of ν. It follows that for $\nu < 4.45$ no minimax solution to the Q-optimality problem exist. We now discuss how this difficulty can be overcome.

As at (2.129), we found that using $\lambda_{max} = \lambda_1$,

$$m_0(x;\alpha,\beta) = w\left(1-\frac{r}{(4x-3)}-\frac{t}{(4x-3)^2}\right)^+, \ \ x\in\left[\frac{1}{2},1\right]$$

minimizes $max_f \mathcal{L}_Q(f,\xi)$ where r, t and w satisfy (2.125).

Similarly, with $\lambda_{max} = \lambda_2$,

$$m_*(x;\tilde{\alpha},\tilde{\beta}) = w1\left(1-\frac{r1}{(6x-5)}-\frac{t1}{(6x-5)^2}\right)^+, \ \ x\in\left[\frac{1}{2},1\right]$$

minimizes $max_f \mathcal{L}_Q(f,\xi)$ where $r1$, $t1$ and $w1$ also satisfy (2.125). For ease of notation let us denote the expression (2.140) by $\mathcal{J}_0(\alpha,\beta)$ and (2.161) by $\mathcal{J}_*(\alpha,\beta)$. We then proceed as follows:

Table 2: Some Parameter Values Minimizing (2.161)

ν	k	l	r	t	w	α	β	λ_1	λ_2
0.005	-0.0667	0.0635	-0.003	0.004	1.093	0.372	0.287	1.020	1.000
0.05	-0.2193	0.1912	-0.028	0.042	1.348	0.364	0.278	1.064	1.006
0.5	-0.6549	0.4935	-0.161	0.323	2.674	0.344	0.252	1.239	1.095
1.0	-0.8503	0.6143	-0.236	0.522	3.909	0.333	0.236	1.382	1.191
5.0	-1.3105	0.8733	-0.437	1.144	13.42	0.291	0.178	2.263	1.747
10	-1.4748	0.9596	-0.515	1.415	25.35	0.270	0.148	3.133	2.235
20	-1.6053	0.9999	-0.605	1.605	46.67	0.262	0.137	4.061	2.872
50	-1.7353	1.00	-0.735	1.735	107.3	0.258	0.133	5.688	4.105

(1) Determine (α_*, β_*) to minimize (2.140) subject to $\mathcal{J}_0(\alpha,\beta) \geq \mathcal{J}_*(\alpha,\beta)$.

(2) Determine $(\tilde{\alpha_*}, \tilde{\beta_*})$ to minimize (2.161) subject to $\mathcal{J}_*(\alpha,\beta) \geq \mathcal{J}_0(\alpha,\beta)$.

The minimax design is then defined by

$$m(x) = \begin{cases} m_0(x;\alpha_*,\beta_*)\ , & \text{if } \mathcal{J}_0(\alpha_*,\beta_*) \leq \mathcal{J}_*(\tilde{\alpha}_*,\tilde{\beta}_*) \\ m_*(x;\tilde{\alpha}_*,\tilde{\beta}_*)\ , & \text{otherwise.} \end{cases} \tag{2.117}$$

We have not provided explicit and numerical solutions for the above discussion because the solutions are probably too complicated to be useful in practice.

D-optimality

Again, we recall that the D-Optimality problem involves minimizing the maximum determinant of the mean squared error matrix. In mathematical terms, we solve

$$\min_{\xi} \max_{f} \mathcal{L}_D(f,\xi),$$

where

$$\mathcal{L}_D(f,\xi) = \left(\frac{\sigma^2}{n}\right)^p \left\{\frac{1+\left(\frac{n}{\sigma^2}\right)\mathbf{b}^T B^{-1}\mathbf{b}}{|B|}\right\}. \tag{2.118}$$

The maximization problem, as seen previously, leads to solving the equation $|G - \lambda B| = 0$ where $G = C - B^2$. From (2.119) and (2.120) the matrix G

Table 3: Some Parameter Values Minimizing (2.168)

ν	k	l	r	t	w	α	β	λ_1	λ_2
0.005	-0.0380	0.0392	0.001	0.002	1.08	0.375	0.292	0.0002	0.014
0.05	-0.1179	0.1293	0.011	0.015	1.30	0.373	0.292	0.007	0.053
0.5	-0.3244	0.3967	0.072	0.129	2.41	0.361	0.278	0.193	0.302
1.0	-0.4115	0.5156	0.104	0.212	3.37	0.351	0.266	0.447	0.561
4.0	-0.5844	0.7435	0.159	0.435	7.95	0.321	0.224	1.813	1.848
5.04	-0.6109	0.7763	0.165	0.474	9.36	0.315	0.215	2.225	2.225
6.0	-0.6302	0.7998	0.170	0.504	10.6	0.310	0.208	2.588	2.553
10	-0.6835	0.8627	0.179	0.590	15.6	0.295	0.186	3.959	3.768
20	-0.7480	0.9350	0.187	0.699	26.8	0.273	0.155	6.798	6.199
30	-0.7818	0.9726	0.191	0.760	37.3	0.263	0.140	9.066	8.122
40	-0.8047	0.9999	0.195	0.805	47.2	0.259	0.134	10.767	9.666

is given by

$$\begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}$$

where $\mu_1 = 6(16d - 24c + 9a) - 9(32\beta - 48\alpha + 9)^2$ and $\mu_2 = 2(36d - 60c + 25a) - (72\beta - 120\alpha + 25)^2$. So, the equation $|G - \lambda B| = 0$ leads to two linear equations in λ given by

$$\begin{array}{c} 6(16d - 24c + 9a) - 9(32\beta - 48\alpha + 9)^2 - 3\lambda(32\beta - 48\alpha + 9) = 0 \\ 2(36d - 60c + 25a) - (72\beta - 120\alpha + 25)^2 - \lambda(72\beta - 120\alpha + 25) = 0. \end{array} \tag{2.119}$$

Solving these equations we obtain

$$\lambda_1 = \frac{2(16d - 24c + 9a)}{(32\beta - 48\alpha + 9)} - 3(32\beta - 48\alpha + 9) \tag{2.120}$$

and

$$\lambda_2 = \frac{2(36d - 60c + 25a)}{(72\beta - 120\alpha + 25)} - (72\beta - 120\alpha + 25). \tag{2.121}$$

We now assume that λ_1 is the maximum eigenvalue and minimize

$$\max_f \mathcal{L}_D(f, \xi) = \left(\frac{\sigma^2}{n}\right)^2 \left\{\frac{1 + \left(\frac{n}{\sigma^2}\right)\tau^2\lambda_1}{|B|}\right\} \tag{2.122}$$

Table 4: Some Optimal Parameter Values for $\lambda_{max} = \lambda_2$

ν	k	l	r	t	w	α	β	λ_1	λ_2
0.005	-0.0670	0.0637	-0.003	0.004	1.09	0.372	0.287	0.021	0.0002
0.05	-0.2240	0.1938	-0.030	0.043	1.36	0.364	0.278	0.078	0.008
0.5	-0.6880	0.5021	-0.186	0.346	2.81	0.345	0.253	0.404	0.197
1.0	-0.8883	0.6207	-0.268	0.304	4.17	0.334	0.239	0.720	0.445
5.0	-1.3249	0.8711	-0.454	1.154	13.9	0.292	0.180	2.796	2.230
10	-1.4751	0.9595	-0.516	1.415	25.4	0.270	0.148	4.975	4.033
20	-1.5987	0.9999	-0.599	1.599	45.1	0.262	0.137	7.450	6.348

with respect to the density m(x), subject to the constraints (2.125). Under these constraints the matrix B is fixed. So that minimizing (2.167) is equivalent to minimizing

$$\int_{\frac{1}{2}}^{1} (4x-3)^2 m^2(x) dx.$$

This problem has been solved in Section 3.2.1 and the minimizing density is given by (2.129). Using (2.134), we then choose α and β to minimize

$$|B|^{-1}\left\{\nu + \frac{w[(32\beta-48\alpha+9)-r(8\alpha-3)-t]}{(32\beta-48\alpha+9)} - 3(32\beta-48\alpha+9)\right\} \quad (2.123)$$

for fixed values of ν over the range (2.138) with r, t and w satisfying (2.141) where $|B| = 3(32\beta-48\alpha+9)(72\beta-120\alpha+25)$. Also, the minimizing α and β is choosen such that the inequality $\lambda_1(m_0(x)) \geq \lambda_2(m_0(x))$ is satisfied. Some numerical solution to this problem is shown in Table 3. We observe that $\lambda_1(m_0(x)) \geq \lambda_2(m_0(x))$ for values of $\nu \geq 5.04$ and fails otherwise. The results for D-optimality are also summarized in the following theorem:

Theorem 2.8. *For $\mathcal{B} = (\alpha, \beta)$ and $\frac{1}{8} \leq 2\alpha^2 \leq \beta \leq \alpha \leq \frac{1}{2}$ define the density*

$$m_0(x;\mathcal{B}) = w\left(1 - \frac{r}{(4x-3)} - \frac{t}{(4x-3)^2}\right)^+ \quad (0.5 \leq x \leq 1)$$

where the non-negative constants $w = w(\mathcal{B})$, $r = r(\mathcal{B})$, $t = t(\mathcal{B})$ are determined to satisfy (2.125). Then $m_0(x;\mathcal{B})$ minimizes $\int_{\frac{1}{2}}^{1}(4x-3)^2 m^2(x)dx$ for fixed $\mathcal{B}$.

Define $\mathcal{B}_D$ to be the minimizer of (2.168). Then $m_0(x;\mathcal{B}_D)$ is minimax robust for $\mathcal{L}_D$ for those values of ν for which $\lambda_1(m_0(x;\mathcal{B}_D)) > \lambda_2(m_0(x;\mathcal{B}_D))$. The eigenvalue inequality holds for $\nu \geq 5.04$.

Proceeding as in Section 3.2.1 we find that no minimax solution exist for $\nu < 5.04$. In Table 4 we present some of the parameters when $\lambda_{max} = \lambda_2$ was used. An approach similar to the discussion at the end of Section 3.2.1 can be adopted to overcome this difficulty. We observe that as the size of n becomes larger and larger, one would require τ^2 to be of the order of n^{-1} (i.e $\tau^2 = \mathrm{O}(\frac{1}{n})$), so that the error due to random variation and that due to bias will be of the same magnitude. In this case, $\nu = \frac{\sigma^2}{n\tau^2}$ will be bounded away from 0 as $n \to \infty$. That is, $\nu >> 0$ as $n \to \infty$. Thus the lack of definitive results for ν near zero isn't a problem.

Chapter 3

OPTIMAL WEIGHTS AND DESIGNS FOR MULTIWAVELET REGRESSION MODELS: WEIGHTED LEAST SQUARES

1. Preliminaries

In Chapter 2, we discussed briefly the relationship between the nonparametric local averaging procedures and the least squares method. We noted that if the weights are positive and uniform then the local averaging procedure is equivalent to the ordinary least squares method which we used to estimate the parameters in the wavelet regression model defined in Chapter 2. In this chapter we use the local averaging procedure with positive weights. That is, the weighted least squares regression method. The problem is then to find minimax weights and designs under the wavelet regression model :

$$E[y|x] = \mathbf{q}^T(x)\boldsymbol{\beta}_0 + f(x) \quad (3.1)$$

$$y(x_i) = E(y|x_i) + \varepsilon_i, \quad (3.2)$$

where $f(x) \in \mathcal{F}$,

$$\mathcal{F} = \left\{ f(x) : \int_S f^2(x)dx \leq \tau^2, \int_S \mathbf{q}(x)f(x)dx = \mathbf{0} \right\}$$

and the errors ε_i, $(i = 1, 2, ..., n)$ are uncorrelated with mean 0 and constant variance σ^2. A similar problem has been considered by Wiens (1996) for approximately specified multiple and polynomial regression models with heteroscedastic errors. We have already mentioned one of the main results of that report in Chapter 1.

The main results of this chapter are presented in Theorem 3.1, Lemma 3.1 and Theorem 3.2. We find that for the Haar model the optimal design and weights are both uniform under the restriction of unbiasedness. This translates into requiring the product of the optimal design and weights to be the uniform density. We begin with a mathematical formulation of the problem we shall be discussing in this chapter.

Problem

Find optimal weights and design if weighted least squares regression is used to estimate the parameter $\boldsymbol{\beta}_0$.

Applying weighted least squares regression the estimate of β_0 can be expressed as

$$\hat{\boldsymbol{\beta}}_{WLS} = \left(\sum_{i=1}^{n} \mathbf{q}(x_i)\mathbf{q}^T(x_i)w(x_i) \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{q}(x_i)w(x_i)y(x_i) \right). \tag{3.3}$$

In matrix notation, we can write $\hat{\beta}_{WLS}$ as

$$\hat{\boldsymbol{\beta}}_{WLS} = (Q^T W Q)^{-1}(Q^T W \mathbf{y}) \tag{3.4}$$

where

$$\begin{aligned} Q^T &= (\mathbf{q}(x_1), \mathbf{q}(x_2), ..., \mathbf{q}(x_n)) \\ W &= diag(w(x_1), w(x_2), ..., w(x_n)) \end{aligned}$$

and

$$\mathbf{y}^T = (y(x_1), y(x_2), ..., y(x_n)).$$

Using (3.2) we have that

$$\hat{\boldsymbol{\beta}}_{WLS} = (Q^T W Q)^{-1} Q^T W (Q\beta + \mathbf{f}(x) + \boldsymbol{\varepsilon}) \tag{3.5}$$

where $\mathbf{f}^T(x) = (f(x_1), ..., f(x_n))$.

Let $\xi(x)$ be the discrete design measure on the design space defined by (1.3). Then, we can write expression (3.5) as

$$\hat{\boldsymbol{\beta}}_{WLS} = \boldsymbol{\beta} + B^{-1}\mathbf{b} + \frac{B^{-1}}{n} \sum_{i=1}^{n} \mathbf{q}(x_i)w(x_i)\varepsilon(x_i) \tag{3.6}$$

where

$$\begin{aligned} B = B(w, \xi) = \frac{Q^T W Q}{n} &= \int_S \mathbf{q}(x)\mathbf{q}^T(x)w(x)d\xi(x) \\ \mathbf{b} = \mathbf{b}(w, f, \xi) = \frac{Q^T W \mathbf{f}}{n} &= \int_S \mathbf{q}(x)w(x)f(x)d\xi(x). \end{aligned}$$

From (3.6) the bias and covariance matrix of the estimate $\hat{\boldsymbol{\beta}}_{WLS}$ are

$$bias(\hat{\boldsymbol{\beta}}_{WLS}) = B^{-1}\mathbf{b}, \text{ and } cov(\hat{\boldsymbol{\beta}}_{WLS}) = \frac{\sigma^2}{n} B^{-1} D B^{-1} \tag{3.7}$$

where

$$D = D(w, \xi) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{q}(x_i)\mathbf{q}^T(x_i) w^2(x_i) = \int_S \mathbf{q}(x)\mathbf{q}^T(x) w^2(x) d\xi(x).$$

It follows that the mean squared error matrix of $\hat{\beta}_{WLS}$ is

$$MSE(\hat{\beta}_{WLS}) = B^{-1}\mathbf{b}\mathbf{b}^T B^{-1} + \frac{\sigma^2}{n} B^{-1} D B^{-1}. \tag{3.8}$$

The loss function of interest in this study is the Integrated Mean Squared Error loss defined by

$$IMSE = \int_S E[(\hat{y}(x) - E[y|x))^2]dx = trMSE(\hat{\beta}_{WLS})A + \int_S f^2(x)dx. \tag{3.9}$$

However, since $\mathbf{q}(x)$ is a vector of orthonormal wavelets, the matrix

$$A = \int_S \mathbf{q}(x)\mathbf{q}^T(x)dx = I \text{ (the identity matrix).}$$

Thus the loss function becomes

$$\mathcal{L}(f, w, \xi) = \mathbf{b}^T B^{-2}\mathbf{b} + \frac{\sigma^2}{n} tr B^{-2} D + \int_S f^2(x)dx. \tag{3.10}$$

Following Wiens (1992) it can be shown that $\sup_f \mathcal{L}(f, w, \xi)$ is not finite unless the distribution function $\xi(x)$ is absolutely continuous. If we define

$$\xi'(x) = k(x) \text{ and } p(x) = k(x)w(x), \tag{3.11}$$

then

$$\begin{aligned} B = B(p(x)) = B(w,\xi) &= \int_S \mathbf{q}(x)\mathbf{q}^T(x)w(x)k(x)dx \\ &= \int_S \mathbf{q}(x)\mathbf{q}^T(x)p(x)dx \qquad (3.12) \\ \mathbf{b} = \mathbf{b}(f, p(x)) = \mathbf{b}(f, w, \xi) &= \int_S \mathbf{q}(x)f(x)w(x)k(x)dx \\ &= \int_S \mathbf{q}(x)f(x)p(x)dx \qquad (3.13) \\ D = D(w, p(x)) = D(w,\xi) &= \int_S \mathbf{q}(x)\mathbf{q}^T(x)w^2(x)k(x)dx \\ &= \int_S \mathbf{q}(x)\mathbf{q}^T(x)w(x)p(x)dx. \qquad (3.14) \end{aligned}$$

By the transformation (3.11) the loss function now depends on $w(x)$ explicitly only through the matrix D. Also, since $\hat{\boldsymbol{\beta}}_{WLS}$ remains invariant if the weights are multiplied by a scalar, ($i.e.$ $w(x) \rightarrow aw(x)$), we can assume that the average of the weights is 1. By this assumption, $p(x)$ is a density on S and

$$1 = \int_S k(x)dx = \int_S \frac{p(x)}{w(x)}dx. \tag{3.15}$$

The problems defined in the early part of this section can be discussed in the following ways:

P1 : *Take a fixed weight, $w_0(x)$ and minimize(over p) the maximum(over f) of the loss function.*

P2 : *Take a fixed density $p_0(x)$ on S and minimize(over w) the maximum(over f) loss. Then $k_0(x) = \frac{p_0(x)}{w_0(x)}$.*

P3 : *Solve $min_p \, min_w \, max_f \, \mathcal{L}(f, w, \xi)$. Then $k_0(x) = \frac{p_0(x)}{w_0(x)}$.*

Problem **P1** has been discussed for the multiwavelets constructed by Alpert (1992) with uniform weights when N is equal to 1 and 2. The simplest case was discussed in the case of $N = 2$ due to the complexity of the eigenvalues arising from the maximization of the loss function with respect to f. In the next section of this study, we discuss **P2** for these multiwavelets as well. We find that we are able to obtain optimal designs for the general case due to the simplicity arising from fixing the density $p_0(x)$. Results from Section 2.1 of Chapter 2 show that

$$\max_f \mathcal{L}(f, w, \xi) = \tau^2[1 + \lambda_{max} + \nu tr B^{-2} D] \tag{3.16}$$

where λ_{max} is the maximum characteristic root of $G^{\frac{1}{2}} B^{-2} G^{\frac{1}{2}}$ for

$$G = C - B^2 \quad \text{and} \quad C = \int_S \mathbf{q}(x)\mathbf{q}^T(x)p^2(x)dx.$$

Now

$$trB^{-2}D = trB^{-2}\int_S \mathbf{q}(x)\mathbf{q}^T(x)w(x)p(x)dx = \int_S l_p(x)w(x)p(x)dx \tag{3.17}$$

where $l_p(x) = \mathbf{q}^T(x)B^{-2}\mathbf{q}(x)$. Therefore

$$\max_f \mathcal{L}(f, w, \xi) = \tau^2 \left[1 + \lambda_{max} + \nu \int_S l_p(x)w(x)p(x)dx\right]. \tag{3.18}$$

2. Optimal Weight and Design for Fixed Density $p_0(x)$

In this section, we minimize (3.18) with respect to $w(x)$ subject to the constraint (3.15) for fixed $p_0(x)$.

Theorem 3.1. *For $p_0(x)$ a fixed density on S, define*

$$w_0(x;u) = \frac{u}{[l_{p_0}(x)]^{\frac{1}{2}}}$$

where u is determined to satisfy (3.15) and S is such that $\int_S dx = 1$. Define $u_Q = \int_S p_0(x)[l_{p_0}(x)]^{\frac{1}{2}} dx$. Then $w_0(x;u_Q)$ is the optimal weight for $\mathcal{L}_Q$ and $\mathcal{L}_A$. The optimal design density $k_0(x;u_Q)$ is inversely proportional to $w_0(x;u_Q)$. If in particular we take $p_0(x) = \gamma > 0$, then $l_{p_0}(x) = l_\gamma(x) = ||\mathbf{q}(x)||^2$ and $w_0(x;u_Q)$, $k_0(x;u_Q)$ are optimal subject to the condition of unbiasedness.

Proof : Define

$$\begin{aligned} w_t(x) &= (1-t)w_0(x) + tw_1(x) \\ \Phi(t) &= \int_S l_{p_0}(x)w_t(x)p_0(x)dx + u^2 \int_S \frac{p_0(x)}{w_t(x)} dx \end{aligned} \quad (3.1)$$

for $w_0(x)$, $w_1(x)$ satisfying (3.15) and an arbitrary multiplier u. Differentiating $\Phi(t)$ twice with respect to t we obtain

$$\Phi'(t) = \int_S \left(l_{p_0}(x) - \frac{u^2}{w_t^2(x)} \right) p_0(x)(w_1(x) - w_0(x))dx \quad (3.2)$$

$$\Phi''(t) = 2u^2 \int_S \frac{p_0(x)(w_1(x) - w_0(x))^2}{w_t^3(x)} dx. \quad (3.3)$$

Clearly, $\Phi''(t) \geq 0$. So, $\Phi(t)$ is a convex function of t which is minimized at $t = 0$ if and only if $\Phi'(0) \geq 0$ for all $w_1(x)$. The equation below, obtained from (3.20),

$$\Phi'(0) = \int_S \left(l_{p_0}(x) - \frac{u^2}{w_0^2(x)} \right) p_0(x)(w_1(x) - w_0(x))dx \quad (3.4)$$

suggests that we choose $w_0(x)$ to satisfy the equation

$$l_{p_0}(x) - \frac{u^2}{w_0^2(x)} = 0, \quad w_0(x) > 0. \quad (3.5)$$

That is,

$$w_0(x) = \frac{u}{[l_{p_0}(x)]^{\frac{1}{2}}} \tag{3.6}$$

where u is determined to satisfy (3.15). Using (3.15), it is easy to show that

$$u = \int_S p_0(x)[l_{p_0}(x)]^{\frac{1}{2}} dx. \tag{3.7}$$

It follows that the optimal weight under fixed $p_0(x)$ is

$$w_0(x) = \frac{\int_S p_0(x)[l_{p_0}(x)]^{\frac{1}{2}} dx}{[l_{p_0}(x)]^{\frac{1}{2}}} \tag{3.8}$$

and the optimal design density is

$$k_0(x) = \frac{p_0(x)[l_{p_0}(x)]^{\frac{1}{2}}}{\int_S p_0(x)[l_{p_0}(x)]^{\frac{1}{2}} dx}. \tag{3.9}$$

It can be seen from (3.26) and (3.27) that

$$w_0(x) \propto [l_{p_0}(x)]^{-\frac{1}{2}} \quad \text{while} \quad k_0(x) \propto \frac{1}{w_0(x)}. \tag{3.10}$$

Substitute for $w_0(x)$ and simplify to see that

$$\int_S l_{p_0}(x) w_0(x) p_0(x) dx = \left(\int_S [l_{p_0}(x)]^{\frac{1}{2}} p_0(x) dx \right)^2. \tag{3.11}$$

Therefore we have, from (3.18), that

$$\min_w \max_f \mathcal{L}(f, w, \xi) = \tau^2 \left[1 + \lambda_{max} + \nu \left(\int_S [l_{p_0}(x)]^{\frac{1}{2}} p_0(x) dx \right)^2 \right]. \tag{3.12}$$

Now we take $p_0(x) = \gamma$, where $\gamma > 0$ is some constant chosen such that $p_0(x)$ is a density on S. **This choice of $p_0(x)$ is equivalent to the side condition of unbiasedness** and results in (3.30) reducing to

$$\min_w \mathcal{L}(w, \xi) = \min_w \max_f \mathcal{L}(f, w, \xi) = \frac{\sigma^2}{n} \gamma^2 \left(\int_S [l_\gamma(x)]^{\frac{1}{2}} dx \right)^2 \tag{3.13}$$

where

$$l_\gamma(x) = \frac{\mathbf{q}^T(x)\mathbf{q}(x)}{\gamma^2}.$$

That is,

$$\min_w \mathcal{L}(w, \xi) = \frac{\sigma^2}{n} \left(\int_S [\mathbf{q}^T(x)\mathbf{q}(x)]^{\frac{1}{2}} dx \right)^2. \tag{3.14}$$

3 Examples

In our examples we will use the multiwavelets constructed by Alpert (1992) with $N = 1$ and 2. We recall that when $N = 1$ these wavelets coincide with the Haar wavelet. To avoid trivialities, we assume (as in Section 2 of Chapter 2) that if there is a point $x_0 \in S$ with $\mathbf{q}(x_0) = \mathbf{0}$ (e.g. $\mathbf{q}(1) = \mathbf{0}$), then $\xi\{x_0\} = 0$. Otherwise, since such a point x_0 would contribute nothing to $\mathbf{b}$, B or D, we could remove it from S and work with the conditional design on $S \backslash \{x_0\}$.

3.1. $N = 1$ (Haar Wavelet)

As usual, our design space is $S = [0, 1]$ which implies that $\gamma = 1$. Results for **P1**, with uniform weights, show that for any $x \in [0, 1)$

$$\mathbf{q}^T(x)\mathbf{q}(x) = 2^{m+1}$$

for the Haar wavelet basis of order m. It follows that applying the above assumption we have, from (2.94), that

$$l_\gamma(x) = 2^{m+1}, \quad w_0(x) = 1, \quad k_0(x) = 1, \quad x \in [0, 1) \tag{3.15}$$

and

$$\min_w \mathcal{L}(w, \xi) = \frac{2^{m+1}\sigma^2}{n}. \tag{3.16}$$

Thus we have :

Corollary 3.1. *Let $p_0(x) = \gamma$, $\gamma > 0$ a constant, be a fixed density on the design space $S = [0, 1]$. Let the components of the vector $\mathbf{q}(x)$ be the Haar wavelet basis of order m. Then, the design and weight minimizing $\mathcal{L}_Q$ and $\mathcal{L}_A$ are each uniform. In other words, ordinary least squares is optimal for estimating the parameters of the mth order Haar regression model under the condition of unbiasedness.*

3.2. $N = 2$ Multiwavelet

For $N = 2$ the vector $\mathbf{q}(x)$, as a function of the order m of the wavelet basis is given by

$$\mathbf{q}_m^T(x) = \Big(\phi_0(x), \phi_1(x), {}_2w_0(x), {}_2w_1(x), ..., {}_2w_0^{-m,0}(x), ...$$
$$..., {}_2w_0^{-m,2^m-1}(x), {}_2w_1^{-m,0}(x), ..., {}_2w_1^{-m,2^m-1}(x)\Big)\,. \tag{3.17}$$

Here, we have used the expression (2.111) with $m = p$ to define the vector (3.35). The results for $m \neq p$ are presented later in this section. If we set

$$\mathbf{a}_m^T(x) = \Big({}_2w_0(x), {}_2w_1(x), ..., {}_2w_0^{-m,0}(x), ...$$
$$..., {}_2w_0^{-m,2^m-1}(x), {}_2w_1^{-m,0}(x), ..., {}_2w_1^{-m,2^m-1}(x)\Big) \tag{3.18}$$

then

$$\mathbf{q}_m^T(x) = \Big(\phi_0(x), \phi_1(x), \mathbf{a}_m^T(x)\Big)\,. \tag{3.19}$$

It follows that

$$\mathbf{q}_{m+1}^T(x)\mathbf{q}_{m+1}(x) = \phi_0^2(x) + \phi_1^2(x) + \mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) \tag{3.20}$$

where

$$\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) = \mathbf{a}_m^T(x)\mathbf{a}_m(x) \;+\; 2^{m+1}\sum_{k=0}^{2^{m+1}-1} {}_2w_0^2(2^{m+1}x-k)$$
$$+\, 2^{m+1}\sum_{k=0}^{2^{m+1}-1} {}_2w_1^2(2^{m+1}x-k). \tag{3.21}$$

We now use the expressions for the primary wavelets ${}_2w_0(x)$ and ${}_2w_1(x)$ to obtain

$$\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) \;=\; \mathbf{a}_m^T(x)\mathbf{a}_m(x) \;+$$
$$2^{m+1}\sum_{k=0}^{2^{m+1}-1} 3\Big\{[1-4(2^{m+1}x-k)]^2 I_{\left[2^{-(m+1)}k,2^{-(m+1)}(k+\frac{1}{2})\right)} +$$
$$[4(2^{m+1}x-k)-3]^2 I_{\left[2^{-(m+1)}(k+\frac{1}{2}),2^{-(m+1)}(k+1)\right)}\Big\} \;+$$
$$2^{m+1}\sum_{k=0}^{2^{m+1}-1} \Big\{[6(2^{m+1}x-k)-1]^2 I_{\left[2^{-(m+1)}k,2^{-(m+1)}(k+\frac{1}{2})\right)} +$$
$$[6(2^{m+1}x-k)-5]^2 I_{\left[2^{-(m+1)}(k+\frac{1}{2}),2^{-(m+1)}(k+1)\right)}\Big\}\,. \tag{3.22}$$

We expand the first and third squared terms in (3.40) to obtain

$$3[1 - 4(2^{m+1}x - k)]^2 = 3 - 24(2^{m+1}x - k) + 48(2^{m+1}x - k)^2$$

$$[6(2^{m+1}x - k) - 1]^2 = 1 - 12(2^{m+1}x - k) + 36(2^{m+1}x - k)^2.$$

Adding these expressions we have

$$\begin{aligned} & 3[1 - 4(2^{m+1}x - k)]^2 + [6(2^{m+1}x - k) - 1]^2 \\ & \quad = 4 - 36(2^{m+1}x - k) + 84(2^{m+1}x - k)^2. \end{aligned} \tag{3.23}$$

Similarly,

$$\begin{aligned} & 3[4(2^{m+1}x - k) - 3]^2 + [6(2^{m+1}x - k) - 5]^2 \\ & \quad = 52 - 132(2^{m+1}x - k) + 84(2^{m+1}x - k)^2. \end{aligned} \tag{3.24}$$

Using (3.41) and (3.42) in (3.40) we have that

$$\begin{aligned} \mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) \;=\; & \mathbf{a}_m^T(x)\mathbf{a}_m(x) \;+ \\ & 2^{m+1}\sum_{k=0}^{2^{m+1}-1}[4 - 36(2^{m+1}x - k) + \\ & 84(2^{m+1}x - k)^2]I_{\left[2^{-(m+1)}k, 2^{-(m+1)}(k+\frac{1}{2})\right)} \;+ \\ & 2^{m+1}\sum_{k=0}^{2^{m+1}-1}[52 - 132(2^{m+1}x - k) + \\ & 84(2^{m+1}x - k)^2]I_{\left[2^{-(m+1)}(k+\frac{1}{2}), 2^{-(m+1)}(k+1)\right)}. \end{aligned} \tag{3.25}$$

Now, for any $x \in [0, 1)$ we can find one and only one value of k, $0 \leq k \leq 2^{m+1} - 1$, say k_*, such that

$$x \in \left[2^{-(m+1)}k_*, 2^{-(m+1)}(k_* + \frac{1}{2})\right)$$

or

$$x \in \left[2^{-(m+1)}(k_* + \frac{1}{2}), 2^{-(m+1)}(k_* + 1)\right)$$

since the intervals are disjoint. So we can write

$$\begin{aligned}\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) &= \mathbf{a}_m^T(x)\mathbf{a}_m(x) + \\ &2^{m+1}[4 - 36(2^{m+1}x - k_*) + \\ &84(2^{m+1}x - k_*)^2]I_{\left[2^{-(m+1)}k_*, 2^{-(m+1)}(k_*+\frac{1}{2})\right)} + \\ &2^{m+1}[52 - 132(2^{m+1}x - k_*) + \\ &84(2^{m+1}x - k_*)^2]I_{\left[2^{-(m+1)}(k_*+\frac{1}{2}), 2^{-(m+1)}(k_*+1)\right)}.\end{aligned} \tag{3.26}$$

Then, it can easily be verified that

$$\mathbf{q}_m^T(x)\mathbf{q}_m(x) = 4(3x^2 - 3x + 1)I_{[0,1)} + \sum_{j=0}^{m} e(x;j) \tag{3.27}$$

where

$$\begin{aligned}e(x;j) = 2^j[4 - 36(2^jx - k_*) + 84(2^jx - k_*)^2]I_{\left[2^{-j}k_*, 2^{-j}(k_*+\frac{1}{2})\right)} + \\ 2^j[52 - 132(2^jx - k_*) + 84(2^jx - k_*)^2]I_{\left[2^{-j}(k_*+\frac{1}{2}), 2^{-j}(k_*+1)\right)}.\end{aligned} \tag{3.28}$$

It follows from (3.45) that

$$\mathbf{q}_m^T(x)\mathbf{q}_m(x) = \mathbf{q}_{m-1}^T(x)\mathbf{q}_{m-1}(x) + e(x;m). \tag{3.29}$$

Next, we state and prove a theorem which provides a closed form for the squared Euclidean norm of the vector $\mathbf{q}_m(x)$ for an arbitrary order m.

Theorem 3.2. *Let the components of the vector $\mathbf{q}(x)$ be the multiwavelets of order m with $N = 2$. Then,*

$$||\mathbf{q}_m(x)||^2 = \sum_{k=0}^{2^{m+1}-1} l_m(x;k) \tag{3.30}$$

where

$$\begin{aligned}l_m(x;k) = 2^{m+3}[12(2^mx)^2 - 6(2^mx)(1 + 2k) + \\ (1 + 3k(k + 1)]I_{\left[2^{-(m+1)}k, 2^{-(m+1)}(k+1)\right)}.\end{aligned} \tag{3.31}$$

The squared norm can also be written as,

$$||\mathbf{q}_m(x)||^2 = 2^{m+3} \min_{0\le k\le 2^{m+1}-1} [1 - 3h_m(x;k)] \tag{3.32}$$

where

$$h_m(x;k) = (2^{m+1}x - k)(k + 1 - 2^{m+1}x). \tag{3.33}$$

Proof : If (3.48) is true, it implies that from (3.38) we can write

$$\phi_0^2(x) + \phi_1^2(x) + \mathbf{a}_m^T(x)\mathbf{a}_m(x) = \sum_{k=0}^{2^{m+1}-1} l_m(x;k). \tag{3.34}$$

It then follows that

$$\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) = \sum_{r=0}^{2^{m+2}-1} l_{m+1}(x;r) - \sum_{k=0}^{2^{m+1}-1} l_m(x;k). \tag{3.35}$$

Combining (3.38) and (3.47), it is easily verified that

$$\begin{aligned}\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) &= \mathbf{q}_{m+1}^T(x)\mathbf{q}_{m+1}(x) - \mathbf{q}_m^T(x)\mathbf{q}_m(x) \\ &= e(x; m+1). \end{aligned} \tag{3.36}$$

Therefore from (3.53) and (3.54), showing that (3.48) holds is equivalent to showing that

$$\sum_{r=0}^{2^{m+2}-1} l_{m+1}(x;r) - \sum_{k=0}^{2^{m+1}-1} l_m(x;k) = e(x; m+1). \tag{3.37}$$

So, if we are able to show that the identity (3.55) holds then it follows that (3.48) also holds. It is not too difficult to see, from (3.55), that for an arbitrary value of m, if $x \in \left[2^{-(m+2)}r, 2^{-(m+2)}(r+1)\right), \quad r = 0, 1, 2, ..., 2^{m+2} - 1$ then

$$k = k_* = \begin{cases} \frac{r}{2} & \text{if r is even} \\ \frac{r-1}{2} & \text{if r is odd} \end{cases} \tag{3.38}$$

We shall show that (3.55) holds for r even and r odd. Now, if $x \in \Big[2^{-(m+2)}r,$ $2^{-(m+2)}(r+1)\Big)$ and r is even we have

$$\begin{aligned}&\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) = l_{m+1}(x;r) - l_m\left(x; \frac{r}{2}\right) \\ &= 2^{m+1}\, 8[12(2^{m+1}x)^2 - 6(2^{m+1}x)(1+2r) + (1 + 3r(r+1))] \\ &- 2^m\, 8\left[12(2^m x)^2 - 6(2^m x)\left(1 + \frac{2r}{2}\right) + \left(1 + \frac{3r}{2}\left(\frac{r}{2} + 1\right)\right)\right]. \end{aligned} \tag{3.39}$$

We simplify to obtain

$$\begin{aligned}\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) &= l_{m+1}(x;r) - l_m\left(x;\frac{r}{2}\right)\\ = 2^{m+3}&\left[84(2^m x)^2 - 6(2^m x)(3+7r) + \left(1 + \frac{3r(6+7r)}{4}\right)\right].\end{aligned} \tag{3.40}$$

Also, from (3.46) we have that

$$e(x;m+1) = 2^{m+1}\left[4 - 36\left(2^{m+1}x - \frac{r}{2}\right) + 84\left(2^{m+1}x - \frac{r}{2}\right)^2\right]. \tag{3.41}$$

We simplify (3.59) to obtain

$$e(x;m+1) = 2^{m+3}\left[84(2^m x)^2 - 6(2^m x)(3+7r) + \left(1 + \frac{3r(6+7r)}{4}\right)\right]. \tag{3.42}$$

Comparing (3.58) and (3.60) we observe that

$$l_{m+1}(x;r) - l_m(x;k) = e(x;m+1)$$

when r is even and $x \in \left[2^{-(m+2)}r, 2^{-(m+2)}(r+1)\right)$. Therefore, we have shown that (3.55) holds when r is even and $x \in \left[2^{-(m+2)}r,\ 2^{-(m+2)}(r+1)\right)$. Now if $x \in \left[2^{-(m+2)}r, 2^{-(m+2)}(r+1)\right)$ and r is odd then

$$\begin{aligned}\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) &= l_{m+1}(x;r) - l_m\left(x;\frac{r-1}{2}\right)\\ = 2^{m+1}\ 8[12(2^{m+1}x)^2 &- 6(2^{m+1}x)(1+2r) + (1+3r(r+1))]\\ - 2^m\ 8&\left[12(2^m x)^2 - 6(2^m x)\left(1 + \frac{2(r-1)}{2}\right)\right.\\ &\left.+ \left(1 + \frac{3(r-1)}{2}\left(\frac{r-1}{2} + 1\right)\right)\right].\end{aligned} \tag{3.43}$$

Simplifying, we obtain

$$\begin{aligned}\mathbf{a}_{m+1}^T(x)\mathbf{a}_{m+1}(x) - \mathbf{a}_m^T(x)\mathbf{a}_m(x) &= l_{m+1}(x;r) - l_m\left(x;\frac{r-1}{2}\right)\\ = 2^{m+3}&\left[84(2^m x)^2 - 6(2^m x)(4+7r) + \left(\frac{7+3r(8+7r)}{4}\right)\right].\end{aligned} \tag{3.44}$$

We also have, from (3.46), that

$$e(x;m+1) = 2^{m+1}\left[52 - 132\left(2^{m+1}x - \frac{r-1}{2}\right)\right.$$
$$\left. +84\left(2^{m+1}x - \frac{r-1}{2}\right)^2\right]$$
$$= 2^{m+3}\left[84(2^m x)^2 - 6(2^m x)(4+7r)\right.$$
$$\left. + \left(\frac{7+3r(8+7r)}{4}\right)\right]. \qquad (3.45)$$

Again, comparing (3.62) and (3.63) we see that

$$l_{m+1}(x;r) - l_m(x;k) = e(x;m+1),$$

r odd and $x \in \left[2^{-(m+2)}r, 2^{-(m+2)}(r+1)\right)$. Implying that, (3.55) holds when r is odd and $x \in \left[2^{-(m+2)}r,\ 2^{-(m+2)}(r+1)\right)$. We therefore conclude that (3.55) holds for all r, $(r = 0, 1, 2, ..., 2^{m+2} - 1)$ which completes the proof for (3.48).

Next, to simplify notations, let us set $y = 2^m x$ and define

$$g_k(y) = 12y^2 - 6(1+2k)y + (1+3k(k+1)). \qquad (3.46)$$

At the endpoints of the interval, $x = 2^{-(m+1)}k$ and $x = 2^{-(m+1)}(k+1)$. When $x = 2^{-(m+1)}k$, $y = \frac{k}{2}$ and $g_k(\frac{k}{2}) = 1$. Also, when $x = 2^{-(m+1)}(k+1)$, $y = \frac{k+1}{2}$ and $g_k(\frac{k+1}{2}) = 1$. Therefore we have that

$$g_k\left(\frac{k}{2}\right) = g_k\left(\frac{k+1}{2}\right) = 1 \qquad (3.47)$$

which is also the maximum value of $g_k(y)$ attained at the extremes of each interval. We can then express $g_k(y)$ as

$$g_k(y) = 1 + 12\left(y - \frac{k}{2}\right)\left(y - \frac{k+1}{2}\right)$$
$$= 1 + 3(2y-k)(2y-(k+1)), \qquad (3.48)$$

and write the function $l_m(x;k)$ as follows

$$l_m(x;k) = 2^{m+3}[1 + 3(2^{m+1}x - k)(2^{m+1}x - (k+1))]I_{[2^{-(m+1)}k, 2^{-(m+1)}(k+1))}$$
$$= 2^{m+3}[1 - 3h_m(x;k)]I_{[k \le 2^{m+1}x < (k+1))} \qquad (3.49)$$

where $h_m(x;k) = (2^{m+1}x - k)(k + 1 - 2^{m+1}x)$. As a function of y, l_m can be written as

$$l_m(y;k) = 2^{m+3}[1 + 3(2y - k)(2y - (k+1))]I_{[\frac{k}{2},\frac{k+1}{2})}. \tag{3.50}$$

Then, it is not too difficult to see that $||\mathbf{q}_m(x)||^2$ can be written as

$$||\mathbf{q}_m(x)||^2 = 2^{m+3}min_{0\leq k\leq 2^{m+1}-1}[1 - 3h_m(x;k)]. \tag{3.51}$$

Lemma 3.1. *Let the variable y satisfy the relationship described by the model (3.2). Let the design space S be normalized such that $\int_S dx = 1$. Let the components of the vector $\mathbf{q}(x)$ be the multiwavelets of order m with $N = 2$. If the density $p_0(x) \equiv \gamma := (\int_S dx)^{-1}$, then the optimal weight $w_0(x)$ and design minimizing $\mathcal{L}_Q$ and $\mathcal{L}_A$ is given by*

$$w_0(x) = \frac{\mathcal{K}_m}{||\mathbf{q}_m(x)||} \text{ and } k_0(x) = \frac{||\mathbf{q}_m(x)||}{\mathcal{K}_m} \tag{3.52}$$

respectively, where $||\mathbf{q}_m(x)||^2$ is defined by (3.48) and

$$\begin{aligned}\mathcal{K}_m = \int_S ||\mathbf{q}_m(x)||dx &= \int_S \left[\sum_{k=0}^{2^{m+1}-1} l_m(x;k)\right]^{\frac{1}{2}} dx \\ &= 5.520692\left(2^{\frac{(m-3)}{2}}\right).\end{aligned} \tag{3.53}$$

Proof : The fact that the optimal weight and design are

$$w_0(x) = \frac{\mathcal{K}_m}{||\mathbf{q}_m(x)||} \text{ and } k_0(x) = \frac{||\mathbf{q}_m(x)||}{\mathcal{K}_m} \tag{3.54}$$

respectively, where

$$\mathcal{K}_m = \int_S ||\mathbf{q}_m(x)||dx = \sum_{k=0}^{2^{m+1}-1} \int_S [l_m(x;k)]^{\frac{1}{2}} dx \tag{3.55}$$

is a direct consequence of Theorems 3.1 and 3.2. Since the intervals are disjoint, the expression (3.73) for $\mathcal{K}_m$ holds. To prove (3.71) we use (3.49) to write $\mathcal{K}_m$ as

$$\mathcal{K}_m = \sum_{k=0}^{2^{m+1}-1} 2^{\frac{m+3}{2}} \int_{2^{-(m+1)}k}^{2^{-(m+1)}(k+1)} [g_m(x;k)]^{\frac{1}{2}} dx,$$

where

$$g_m(x;k) = 3(2^{2(m+1)})x^2 - 3(2^{m+1})(1+2k)x + (1+3k(k+1)).$$

To simplify notations, let us define

$$a_* = 3(2^{2(m+1)}), \quad b_* = -3(2^{m+1})(1+2k) \text{ and } c_* = (1+3k(k+1)).$$

Since

$$a_* > 0 \text{ and } \Delta = 4a_*c_* - b_*^2 = 3(2^{2(m+1)}) > 0,$$

it follows, from Gradshteyn and Ryzhik (1980), that

$$\int [g_m(x;k)]^{\frac{1}{2}} dx = \frac{(2a_*x + b_*)}{4a_*}[g_m(x;k)]^{\frac{1}{2}} + \frac{\Delta}{8a_*}\frac{1}{\sqrt{a_*}} Arsh\left(\frac{2a_*x+b_*}{\sqrt{\Delta}}\right).$$

Substitute and simplify to show that when $x = 2^{-(m+1)}k$,

$$2a_*x + b_* = -3(2^{m+1}) \text{ and } g_m(x;k) = 1.$$

For $x = 2^{-(m+1)}(k+1)$, we have

$$2a_*x + b_* = 3(2^{m+1}) \text{ and } g_m(x;k) = 1.$$

The fact that $g_m(x;k) = 1$ at the boundaries has been shown in the proof to Theorem 3.2. Therefore, by substitution we can verify that

$$\begin{aligned} \mathcal{K}_m &= \frac{2^{\frac{(m-3)}{2}}}{\sqrt{3}}[4\sqrt{3} + Arsh(\sqrt{3}) - Arsh(-\sqrt{3})] \\ &= \frac{2^{\frac{(m-3)}{2}}}{\sqrt{3}}[4\sqrt{3} + 2Arsh(\sqrt{3})] = 5.520692\left(2^{\frac{(m-3)}{2}}\right). \end{aligned}$$

We observe that the value of $\mathcal{K}_m$ is independent of k. The lemma that follows provides a justification for this. We show that once the value of $l_m(x;k)$ is known in any of the intervals, it is completely determined over the interval $[0,1]$. Figure 3 shows a plot of the squared norm of $\mathbf{q}_m(x)$, $w_0(x)$ and $k_0(x)$ for $m = 1$.

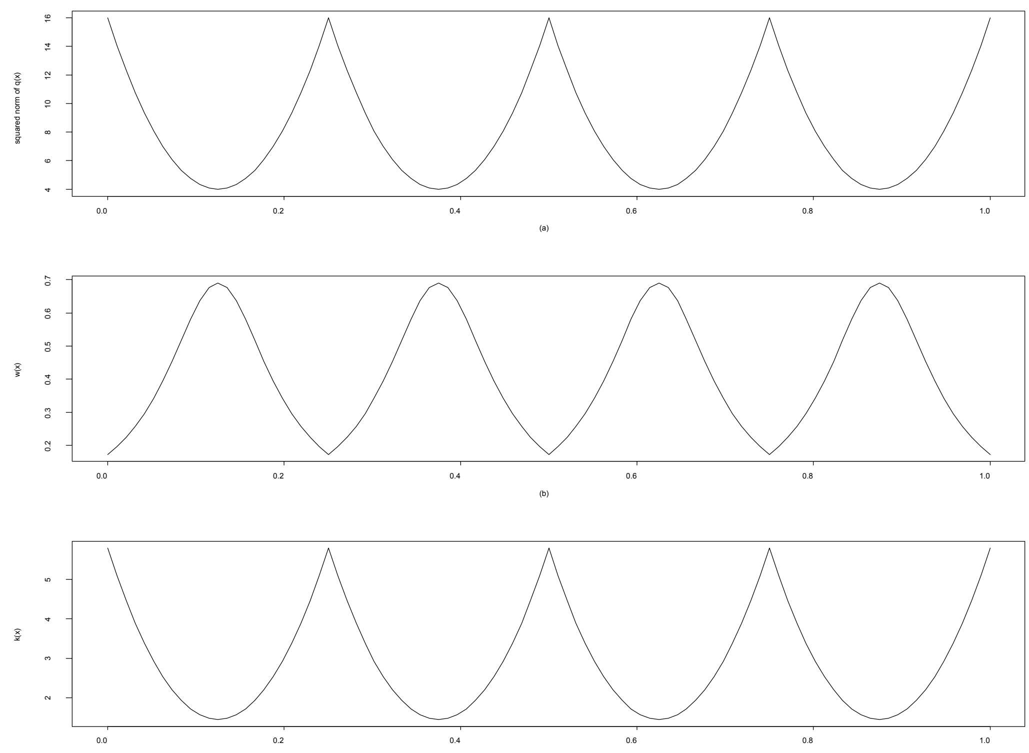


Figure 3: A Plot of $\|\mathbf{q}(x)\|^2$, Optimal Weight and Density for $m = 1$: (a) $\|\mathbf{q}_1(x)\|^2$; (b) $w_0(x)$; (c) $k_0(x)$.

Lemma 3.2. *The squared norm* $||\mathbf{q}_m(x)||^2$ *satisfies the inequality,*

$$2^{m+1} \leq ||\mathbf{q}_m(x)||^2 \leq 2^{m+3}\,, \quad \textit{for all } x \in [0, 1). \tag{3.56}$$

The function $l_m(x; k)$ *satisfies the identities,*

$$l_m(x + 2^{-(m+1)}; k + 1) = l_m(x; k), \;\; k \;=\; 0, 1, 2, .., 2^{m+1} - 1 \tag{3.57}$$

and

$$l_m(1 - x; r) = l_m(x; 2^{m+1} - (r + 1)), \;\; r \;=\; 0, 1, 2, .., 2^m - 1. \tag{3.58}$$

Proof : First, we show that in each interval $[2^{-(m+1)}k, 2^{-(m+1)}(k + 1))$, $(k = 0, 1, 2, ..., 2^{m+1} - 1)$, the function l_m is a minimum at the midpoint of the interval.

From elementary calculus, the stationary points of $g_k(y)$ must satisfy

$$\frac{d}{dy} g_k(y) = 24y - 6(1+2k) = 0.$$

This implies that

$$y = \frac{1+2k}{4}$$

or in terms of x

$$x = 2^{-(m+1)} \left(\frac{1+2k}{2}\right).$$

Since $\frac{d^2}{dy^2} g_k(y) > 0$, it follows that the point $y = \frac{1+2k}{4}$ is a minimum point of $g_k(y)$ with minimum value $g_k\left(\frac{1+2k}{4}\right) = \frac{1}{4}$. This shows that, in each interval $[2^{-(m+1)}k, 2^{-(m+1)}(k+1))$, $(k = 0, 1, 2, .., 2^{m+1} - 1)$, the squared norm of $\mathbf{q}_m(x)$ is a paraboloid (see Figure 3) with a minimum value of 2^{m+1} attained at the midpoint $2^{-(m+1)}\left(\frac{1+2k}{2}\right)$ of each interval and a maximum value of 2^{m+3} attained at the endpoints $2^{-(m+1)}k$ and $2^{-(m+1)}(k+1)$ (see (3.67) and the discussion before and after (3.67)). That is,

$$2^{m+1} \leq ||\mathbf{q}_m(x)||^2 \leq 2^{m+3}, \quad \text{for all } x \in [0,1). \tag{3.59}$$

Next, we show that the identity (3.75) holds. We observe that if $x \in [2^{-(m+1)}k,\ 2^{-(m+1)}(k+1))$ then $x + 2^{-(m+1)} \in [2^{-(m+1)}(k+1), 2^{-(m+1)}(k+2))$, $(k = 0, 1, 2, .., 2^{m+1} - 2)$. So transforming x to $x + 2^{-(m+1)}$ implies $y = 2^m x$ becomes $y + \frac{1}{2}$ and k takes the value $k + 1$. Based on the above, we can show that

$$g_{k+1}\left(y + \frac{1}{2}\right) = g_k(y).$$

In terms of the function l_m this means that

$$l_m(x + 2^{-(m+1)}; k+1) = l_m(x; k).$$

The second identity,

$$l_m(1-x; r) = l_m(x; 2^{m+1} - (r+1)), \quad r = 0, 1, 2, .., 2^m - 1$$

follows from the fact that,

$$I_{\left[2^{-(m+1)}r, 2^{-(m+1)}(r+1)\right)}(1-x) = I_{\left[1-2^{-(m+1)}(r+1), 1-2^{-(m+1)}r\right)}.$$

From the first identity (3.75), we can see that for $x \in [0, 2^{-(m+1)})$

$$\begin{aligned} l_m(x;0) &= l_m(x + 2^{-(m+1)}; 1) \\ &= l_m(x + 2\,2^{-(m+1)}; 2) = l_m(x + 3\,2^{-(m+1)}; 3) \\ &= \ldots\ldots\ldots = l_m(x + (2^{m+1}-1)2^{-(m+1)}; 2^{m+1}-1). \end{aligned} \tag{3.60}$$

This implies that the value of the squared norm of the vector $\mathbf{q}_m(x)$, $||\mathbf{q}_m(x)||^2$, is completely determined by its value in only one of the intervals $[2^{-(m+1)}k,$ $2^{-(m+1)}\ (\ k+1))$, $(k = 0, 1, 2, .., 2^{m+1}-1)$. The second identity (3.76) provides the following relationships :

$\mathbf{m = 0}$:

$$l_0(1-x;0) = l_0(x;1).$$

$\mathbf{m = 1}$:

$$l_1(1-x;0) = l_1(x;3)\ ;\ l_1(1-x;1) = l_1(x;2).$$

$\mathbf{m = 2}$:

$$l_2(1-x;0) = l_2(x;7)\ ;\ l_2(1-x;1) = l_2(x;6)$$
$$l_2(1-x;2) = l_2(x;5)\ ;\ l_2(1-x;3) = l_2(x;4).$$

and so on. This shows that $||\mathbf{q}_m(x)||^2$ is symmetric about $x = \frac{1}{2}$ (see Figure 3). It follows that the optimal weight and design derived using the multiwavelets when $N = 2$ is also symmetric about $x = \frac{1}{2}$.

As at (3.35) the vector $\mathbf{q}(x)$ was defined for $m = p$ and the results that followed were based on this definition. Theorem 3.3 provides the results for m $\neq p$.

Theorem 3.3. *Let the variable y satisfy the relationship described by the model (3.2). Let the components of the vector* $\mathbf{q}(x)$ *be the* $N = 2$ *multiwavelets with* $m \neq p$. *Then,*

$$||\mathbf{q}(x;m,p)||^2 = \begin{cases} ||\mathbf{q}_m(x)||^2 + \mathcal{H}(x;m,p) & \text{for } m < p \\ ||\mathbf{q}_p(x)||^2 + \mathcal{D}(x;m,p) & \text{for } m > p, \end{cases}$$

where

$$\begin{aligned} \mathcal{H}(x;m,p) &= \mathcal{H}(x;m,p-1) + b1(x;p), \;\; \mathcal{H}(x;m,m) = 0 \\ \mathcal{D}(x;m,p) &= \mathcal{D}(x;m-1,p) + b2(x;m), \;\; \mathcal{D}(x;p,p) = 0. \end{aligned}$$

and $||\mathbf{q}_m(x)||^2$, $||\mathbf{q}_p(x)||^2$ *satisfy (3.48). For* $k = 0, 1, ..., 2^{p+1} - 1$ *and* $x \in [2^{-(p+1)}k, 2^{-(p+1)}(k+1))$, $b1(x;p)$ *is defined by*

$$b1(x;p) = \begin{cases} 2^p\left[6\left(2^p x - \frac{k}{2}\right) - 1\right]^2, & \text{if } k \text{ is even} \\ 2^p\left[6\left(2^p x - \frac{k-1}{2}\right) - 5\right]^2, & \text{if } k \text{ is odd.} \end{cases}$$

For $k = 0, 1, ..., 2^{m+1} - 1$ *and* $x \in [2^{-(m+1)}k, 2^{-(m+1)}(k+1))$, $b2(x;m)$ *is defined by*

$$b2(x;m) = \begin{cases} 2^m\left[1 - 4\left(2^m x - \frac{k}{2}\right)\right]^2, & \text{if } k \text{ is even} \\ 2^m\left[4\left(2^m x - \frac{k-1}{2}\right) - 3\right]^2, & \text{if } k \text{ is odd.} \end{cases}$$

Proof : For $m \neq p$, we have,

$$\begin{aligned} \mathbf{q}^T(x;m,p) = \Big(&\phi_0(x), \phi_1(x), {}_2w_0(x), {}_2w_1(x), ..., {}_2w_0^{-m,0}(x), ... \\ &..., {}_2w_0^{-m,2^m-1}(x), {}_2w_1^{-p,0}(x), ..., {}_2w_1^{-p,2^p-1}(x)\Big). \end{aligned} \tag{3.61}$$

So that,

$$\begin{aligned} ||\mathbf{q}(x;m,p)||^2 = \phi_0^2(x) + \phi_1^2(x) &+ \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} 2^j {}_2w_0^2(2^j x - k) \\ &+ \sum_{j=0}^{p}\sum_{k=0}^{2^j-1} 2^j {}_2w_1^2(2^j x - k). \end{aligned} \tag{3.62}$$

Let us, for a moment, assume that m is less than p. Then, we can write (3.80) as

$$||\mathbf{q}(x;m,p)||^2 = ||\mathbf{q}_m(x)||^2 + \sum_{j=m+1}^{p} \sum_{k=0}^{2^j-1} 2^j {}_2w_1^2(2^j x - k). \tag{3.63}$$

Define,

$$b1(x;j) = \sum_{k=0}^{2^j-1} 2^j {}_2w_1^2(2^j x - k),$$

and substitute for ${}_2w_1(2^j x - k)$ to obtain

$$b1(x;j) = 2^j \sum_{k=0}^{2^j-1} \Big\{ [6(2^j x - k) - 1]^2 I_{\left[2^{-j}k, 2^{-j}\left(k+\frac{1}{2}\right)\right)} + [6(2^j x - k) - 5]^2 I_{\left[2^{-j}\left(k+\frac{1}{2}\right), 2^{-j}(k+1)\right)} \Big\}. \tag{3.64}$$

This implies that

$$||\mathbf{q}(x;m,p)||^2 = ||\mathbf{q}_m(x)||^2 + \sum_{j=m+1}^{p} b1(x;j). \tag{3.65}$$

Following previous discussions we can find some k, $0 \le k \le 2^j - 1$, say k_*, such that

$$b1(x;j) = 2^j \Big\{ [6(2^j x - k_*) - 1]^2 I_{\left[2^{-j}k_*, 2^{-j}\left(k_*+\frac{1}{2}\right)\right)} + [6(2^j x - k_*) - 5]^2 I_{\left[2^{-j}\left(k_*+\frac{1}{2}\right), 2^{-j}(k_*+1)\right)} \Big\}. \tag{3.66}$$

It is not too difficult to see, from (3.84), that for $k = 0, 1, ..., 2^{p+1} - 1$ and $x \in [2^{-(p+1)}k,\, 2^{-(p+1)}(k+1))$ we have,

$$k_* = \begin{cases} \frac{k}{2}, & \text{if } k \text{ is even} \\ \frac{k-1}{2}, & \text{if } k \text{ is odd} \end{cases} \tag{3.67}$$

and

$$b1(x;p) = \begin{cases} 2^p \left[6\left(2^p x - \frac{k}{2}\right) - 1\right]^2, & \text{if } k \text{ is even} \\ 2^p \left[6\left(2^p x - \frac{k-1}{2}\right) - 5\right]^2, & \text{if } k \text{ is odd.} \end{cases} \tag{3.68}$$

Alternatively, for $l = 0, 1, 2, ..., 2^p - 1$, we have

$$b1(x;p) = \begin{cases} 2^p\left[6\left(2^p x - l\right) - 1\right]^2, & \text{if } x \in [2^{-p}l,\ 2^{-(p+1)}(2l+1)) \\ 2^p\left[6\left(2^p x - l\right) - 5\right]^2, & \text{if } x \in [2^{-(p+1)}(2l+1),\ 2^{-p}(l+1)). \end{cases} \quad (3.69)$$

If we define the recursive relation,

$$\mathcal{H}(x;m,p) = \mathcal{H}(x;m,p-1) + b1(x;p), \quad \text{where } \mathcal{H}(x;m,m) = 0, \qquad (3.70)$$

then, for $m < p$ we have

$$||\mathbf{q}(x;m,p)||^2 = ||\mathbf{q}_m(x)||^2 + \mathcal{H}(x;m,p) \qquad (3.71)$$

where $b1(x;p)$ is given by (3.86) or (3.87) and $||\mathbf{q}_m(x)||^2$ is defined by (3.48).

On the other hand, if $m > p$ we have, from (3.80)

$$||\mathbf{q}(x;m,p)||^2 = ||\mathbf{q}_p(x)||^2 + \sum_{j=p+1}^{m} b2(x;j), \qquad (3.72)$$

where

$$b2(x;j) = 2^j\left\{[1 - 4(2^j x - k_*)]^2 I_{\left[2^{-j}k_*, 2^{-j}\left(k_* + \frac{1}{2}\right)\right)} + [4(2^j x - k_*) - 3]^2 I_{\left[2^{-j}\left(k_* + \frac{1}{2}\right), 2^{-j}(k_*+1)\right)}\right\} \qquad (3.73)$$

for some $k_* \in (0, 1, ..., 2^j - 1)$. From previous discussions, it can be easily verified that for $m > p$,

$$||\mathbf{q}(x;m,p)||^2 = ||\mathbf{q}_p(x)||^2 + \mathcal{D}(x;m,p) \qquad (3.74)$$

where

$$\mathcal{D}(x;m,p) = \mathcal{D}(x;m-1,p) + b2(x;m), \quad \mathcal{D}(x;p,p) = 0, \qquad (3.75)$$

and for $k = 0, 1, ..., 2^{m+1} - 1$ and $x \in [2^{-(m+1)}k, 2^{-(m+1)}(k+1))$,

$$b2(x;m) = \begin{cases} 2^m\left[1 - 4\left(2^m x - \frac{k}{2}\right)\right]^2, & \text{if } k \text{ is even} \\ 2^m\left[4\left(2^m x - \frac{k-1}{2}\right) - 3\right]^2, & \text{if } k \text{ is odd.} \end{cases} \qquad (3.76)$$

Chapter 4

APPLICATIONS

1. Preliminaries

Thus far we have only constructed optimal designs for wavelet approximations to the nonlinear regression model (4.1) without suggesting strategies for implementing these designs. This chapter is devoted to strategies for implementation. We recall that the model of interest is

$$E(y|x) = \eta(x), \quad x \in S$$

approximated by

$$E(y|x) = \eta(x) = \mathbf{q}^T(x)\boldsymbol{\beta}_0 + f(x) \tag{4.1}$$

where $f(x)$ is the remainder term arising from the approximation and the elements of the vector $\mathbf{q}(x)$ form a wavelet basis on the design space S.

In the introduction to Chapter 1 we mentioned that the precise mathematical structure of $\eta(x)$ need not be known in order to apply wavelet approximation techniques. We only need to decide on the appropriate wavelet basis to be used and the order m of the approximation. In situations where the precise form of $\eta(x)$ is assumed known with parameters having some physical interpretation, experimenters may wish to consider using the techniques outlined in this work to design their experiments rather than choosing their design points in an arbitrary fashion. Then the assumed form of $\eta(x)$ can be used to estimate the parameters after the experiment has been performed and measurements taken at the design points.

The design space S we have considered in this study is the unit interval, $[0, 1]$. To apply our results to a more general design space $S^* = [a, b]$, we transform any point $x^* \in S^*$ to $x \in S$ by,

$$x^* \longrightarrow x = \frac{x^* - a}{b - a}. \tag{4.2}$$

So that if x_i is an optimal design point in S, then

$$x_i^* = a + (b - a)x_i \tag{4.3}$$

is optimal in S^*.

2. Multiwavelets with Ordinary Least Squares

2.1. $N = 1$ (Haar Wavelet)

We have shown in chapter 2 that the design which minimizes the trace of the covariance function is that which places uniform weight $2^{-(m+1)}$ in the 2^{m+1} intervals

$$\{[2^{-(m+1)}k,\ 2^{-(m+1)}(k+1))\}_{k=0,1,\ldots,2^{m+1}-1}. \tag{4.1}$$

We also showed that this design is A-, D-, and G-optimal. Introducing the bias term, we found that the optimal design density is the uniform density.

To implement this design, the number of design points n has to be a multiple of the number of sub-intervals 2^{m+1}. The points are then selected uniformly from each of the 2^{m+1} intervals. In the literature, if the design is uniform over $[-1,1]$, the n-observation design is choosen to satisfy

$$x_i^* = \frac{2(i-1)}{n-1} - 1, \quad i = 1, 2, ..., n. \tag{4.2}$$

This transforms to

$$x_i = \frac{i-1}{n-1}, \quad i = 1, 2, ..., n \tag{4.3}$$

on $[0,1]$. However, this will include the point $x_n = 1$ for which $\mathbf{q}(x_n) = \mathbf{0}$. Since this point does not contribute to the estimation of $\boldsymbol{\beta}_0$ we avoid $x_n = 1$ by choosing the points as follows:

$$x_i = \frac{2i-1}{2n}, \quad i = 1, 2, ..., n \tag{4.4}$$

on $[0,1]$, where $n = 2^{m+1}a$ for some integer constant, $a > 0$. One approach will be to take "a" repeated measurements at x_i's choosen for $n = 2^{m+1}$. It turns out that the optimal n-observation design proposed by Herzberg and Traves (1994) is a special case of (4.7) with $a = 3$ and $m = 2$.

Another approach for selecting uniform design points will be to use the fact that $x \in \{[2^{-(m+1)}k,\ 2^{-(m+1)}(k+1))\}$ implies $x + 2^{-(m+1)} \in \{[2^{-(m+1)}(k +$

1), $2^{-(m+1)}(k+2))\}$. So, for $n = 2^{m+1}$ we first choose any $x_1 \in [0, 2^{-(m+1)})$ then subsequently we choose

$$x_i = x_1 + (i-1)2^{-(m+1)}, \quad i = 2, ..., n. \tag{4.5}$$

If we choose $x_1 = 2^{-(m+2)}$, then we have (4.7) as proposed previously. More generally, if $n = 2^{m+1}a$ we propose to choose,

$$\begin{array}{c} x_1 \in \left[0, \frac{2^{-(m+1)}}{a}\right), \quad \text{and} \\ x_i = x_1 + (i-1)\frac{2^{-(m+1)}}{a}, \quad i = 2, ..., n. \end{array} \tag{4.6}$$

One can also consider taking "a" repeated measures at x_i's choosen from (4.8). If in (4.9) we take $x_1 = \frac{2^{-(m+2)}}{a}$, then we obtain (4.7) as before.

2.2. $N = 2$ Multiwavelet

Instead of the Haar wavelet basis, one may decide to use the multiwavelets of order m with $N = 2$ in the approximation (4.1). Under the transformation, $y = 4x - 3$, we showed that for $m = 0$, the minimax design ξ_0 has density

$$p_0(y) = \frac{w}{4}\left(1 - \frac{r}{y} - \frac{t}{y^2}\right)^+, \quad y \in [-1, 1]$$

That is,

$$p_0(y) = \begin{cases} 0\,, & k < y < l \\ \frac{w}{4}\left(1 - \frac{r}{y} - \frac{t}{y^2}\right)\,, & -1 \le y \le k,\ l \le y \le 1 \end{cases} \tag{4.7}$$

for some k and l satisfying $-1 \le k < 0$, $0 < l \le 1$, where w, r and t depends on $\nu = \frac{\sigma^2}{n\tau^2}$. Some optimal values of w, r and t are given in Table 1 for fixed ν. As $\nu \longrightarrow 0$, k, $l \longrightarrow 0$ and $p_0(y) \longrightarrow 1$, the uniform density. On the other hand, as $\nu \longrightarrow \infty$, $k \longrightarrow 1$ and $l \longrightarrow -1$. However, k goes faster to 1 than l goes to -1. That is, the design chooses most of its points at the middle of the interval and a few at the extremes.

To implement the minimax design with density $p_0(y)$ we randomly sample design points from ξ_0 as follows:

(i) Let $P(y)$ be the distribution function of y corresponding to $p_0(y)$.

(ii) Select

$$y_i = P^{-1}\left(\frac{2i-1}{4n_1}\right), \quad i = 1, 2, ..., n_1.$$

(iii) Then, $x_i = \frac{3+y_i}{4}$, $i = 1, 2, ..., n_1$.

We observe that choosing $y \in [-1, 1]$ is equivalent to choosing $x \in \left[\frac{1}{2}, 1\right]$. Due to symmetry about $x = \frac{1}{2}$ the points in $\left[0, \frac{1}{2}\right]$ can be obtained by using the fact that $(1-x) \in \left[0, \frac{1}{2}\right]$ for every $x \in \left[\frac{1}{2}, 1\right]$. So, if we require a total of n points, we first choose $n_1 = \frac{n}{2}$ points in $\left[\frac{1}{2}, 1\right]$ then the other $n_2 = n_1$ points are obtained by symmetry.

For $\nu = 5$, we have

$$p_0(y) = \begin{cases} \frac{7.975}{4}\left(1 - \frac{0.1301}{y} - \frac{0.4292}{y^2}\right), & \text{if } -1 \le y \le -0.5933 \\ & \text{and } 0.7234 \le y \le 1 \\ 0, & -0.5933 < y < 0.7234. \end{cases} \tag{4.8}$$

A set of Q- and A-optimal design points randomly selected as described above for $n_1 = 16$ ($n = 32$) are (approximated to four decimal places):

0.5335	0.5767	0.5561	0.5028	0.599	0.5479	0.9954	0.9745	
0.9857	0.5653	0.5205	0.5269	0.9607	0.5404	0.5144	0.5085	

When $\nu = 50$, the Q- and A-optimal density is

$$p_0(y) = \begin{cases} \frac{48.17}{4}\left(1 - \frac{0.1469}{y} - \frac{0.7728}{y^2}\right), & \text{if } -1 \le y \le -0.8087 \\ & \text{and } 0.9556 \le y \le 1 \\ 0, & -0.8087 < y < 0.9556. \end{cases} \tag{4.9}$$

The points randomly choosen from the distribution function of this density are:

0.5196	0.5125	0.5326	0.5009	0.5027	0.5171	0.5223	0.9948	
0.5045	0.5147	0.5104	0.5378	0.5064	0.5083	0.5286	0.5253	

Similarly, we choose D-optimal design points for $\nu = 6$ and 40 with $n_1 = 8$ and 16 respectively,

0.5345	0.9566	0.5616	0.5467	0.5138	0.5045	0.5237	0.9902	

and

0.5009	0.5243	0.5190	0.5026	0.5350	0.5044	0.5308	0.5166	
0.5143	0.5122	0.5062	0.5101	0.5410	0.5081	0.5273	0.5215	

3. Multiwavelets with Weighted Least Squares

In Chapter 3, we constructed optimal weights and designs for multiwavelet approximations when the method of estimating the parameters is weighted least squares. For $N = 1$, we found that the optimal weight and design were uniform. We have discussed strategies for implementing uniform designs in Section 2 of this chapter. Our discussion in this section is therefore restricted to strategies for implementing the optimal weight and design constructed for $N = 2$ in Section 2.1.2 of Chapter 3.

We recall that in Section 2.1.2, we found the optimal weight and design to be,

$$w_0(x) = \frac{\mathcal{K}_m}{||\mathbf{q}_m(x)||}, \quad k_0(x) = \frac{||\mathbf{q}_m(x)||}{\mathcal{K}_m}$$

where

$$\mathcal{K}_m = 5.520692\left(2^{\frac{(m-3)}{2}}\right) \text{ and } ||\mathbf{q}_m(x)||^2 = 2^{m+3}\sum_{k=0}^{2^{m+1}-1} g_m(x;k)$$

with

$$g_m(x;k) = [a_* x^2 + b_* x + c_*] I_{\left[2^{-(m+1)}k, 2^{-(m+1)}(k+1)\right)},$$
$$a_* = 3\left(2^{2(m+1)}\right),\ b_* = -3\left(2^{m+1}\right)(1+2k),\ c_* = 1 + 3k(k+1).$$

To implement the design and weight, we select n design points from the distribution function of $k_0(x)$, then evaluate the corresponding weights at the selected points. For fixed m we proceed as follows:

(i) Let $K_0(x)$ be the distribution function of x corresponding to $k_0(x)$.

It can be shown that

$$K_0(x) = \mathcal{K}_m^{-1}\sum_{k=0}^{2^{m+1}-1} 2^{\frac{(m+3)}{2}}\left[\frac{(2a_* y + b_*)}{4a_*}\sqrt{g_m(y;k)}\right.$$
$$\left. + \frac{1}{8\sqrt{a_*}} Arsh\left(\frac{2a_* y + b_*}{\sqrt{a_*}}\right)\right]_{[0,x]\cap[2^{-(m+1)}k, 2^{-(m+1)}(k+1)]}.$$

(ii) Select $x_i = K_0^{-1}\left(\frac{2i-1}{2n}\right), \quad i = 1, 2, ..., n.$

Table 5: Randomly Selected Design Points and Weights for $m = 0$, $n = 16$

Design Point	0.98	0.928	0.023	0.793	0.708	0.523	0.868	0.478
Weight	0.859	0.932	0.861	1.150	1.151	0.860	1.034	0.859
Design Point	0.428	0.293	0.573	0.133	0.208	0.633	0.073	0.368
Weight	0.932	1.150	0.933	1.036	1.151	1.035	0.934	1.033

Table 6: Randomly Selected Design Points and Weights for $m = 1$, $n = 24$

Design Point	0.403	0.598	0.051	0.485	0.654	0.7	0.736	0.153
Weight	1.133	1.136	0.982	0.869	1.133	0.979	0.868	1.134
Design Point	0.45	0.2	0.097	0.516	0.95	0.551	0.301	0.265
Weight	0.977	0.978	1.135	0.872	0.978	0.98	0.983	0.871
Design Point	0.985	0.235	0.347	0.766	0.016	0.848	0.801	0.904
Weight	0.87	0.869	1.135	0.872	0.873	1.137	0.981	1.132

A set of $n = 16$ and 24 randomly choosen points for $m = 0$ and 1 respectively are shown in Tables 5 and 6.

Sometimes, experimenters partition the design space S into two subspaces S_1 and S_2 such that $S = S_1 \bigcup S_2$. Then, based on prior experience and knowledge of the experiment, they perform more experiments at points choosen from, say S_1, and the remainder at points choosen from S_2. That is, they choose n_1 design points from S_1 and n_2 from S_2 where $n_1 >> n_2$ and $n = n_1 + n_2$. Such designs are common in chemical kinetics and drug related experiments in pharmacology. The strategy we propose for choosing the design points as described above is:

(1) Let $S_1 = [a, b_1)$, $S_2 = [b_1, c]$ and $p_1 = K_0(b_1)$.

(2) Select n_1 design points from S_1 to satisfy

$$x_i = K_0^{-1}\left(\frac{(i-0.5)p_1}{n_1}\right), \quad i = 1, 2, ..., n_1.$$

(3) Select the remaining n_2 design points from S_2 to satisfy

$$y_j = K_0^{-1}\left(p_1 + \frac{(j-0.5)(1-p_1)}{n_2}\right)$$

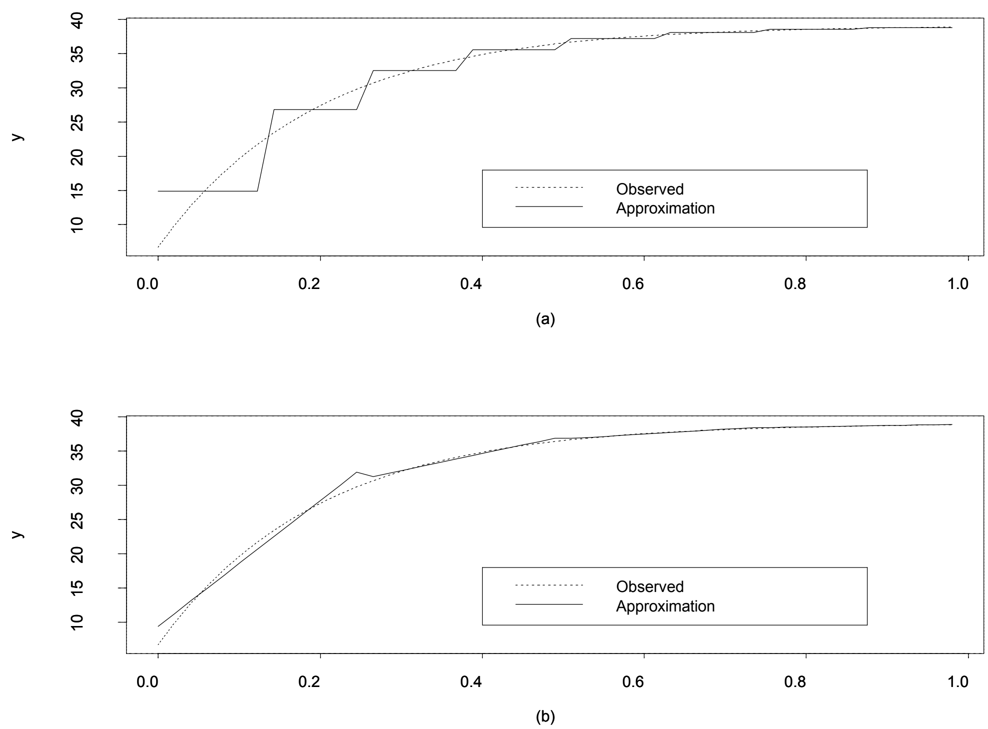


Figure 4: The Ion Transport Model Function: (a) Haar Wavelet $(m = 2)$; (b) $N = 2$ Multiwavelet With Optimal Weights $(m = 1)$.

$$= \quad K_0^{-1}\left(\frac{(n_2 - j + 0.5)p_1 + (j - 0.5)}{n_2}\right), \quad j = 1, 2, ..., n_2.$$

4. Some Examples of Multiwavelet Approximations

In what follows, we consider some nonlinear models commonly used in practice and see how well the wavelets used in this work can approximate these models. We also show how well the multiwavelet regression models fit a data set with no underlying pre-specified model. For each of the nonlinear models we proceed as follows:

(i) Generate n values of the nonlinear function $\eta(x, \theta)$ for fixed θ and $x_i,\ i = 1, 2, ..., n,\ x \in [a, b]$;

(ii) Transform x as in (4.2);

(iii) Fit the Haar wavelet regression model by ordinary least squares;

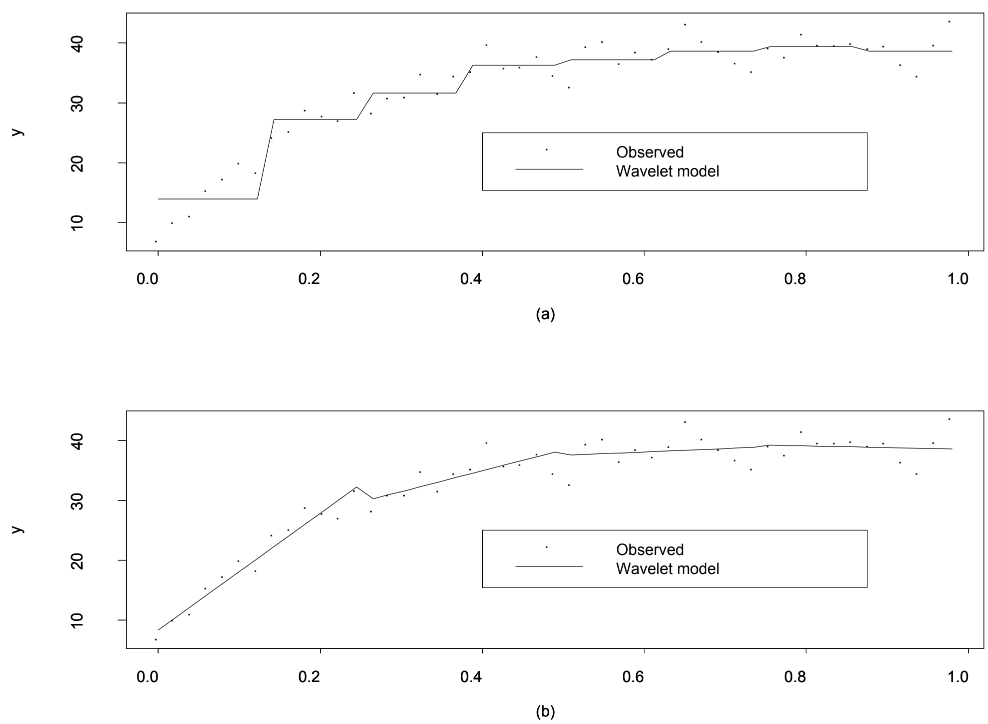


Figure 5: Simulated Data from the Ion Transport Model Function: (a) Haar Wavelet ($m = 2$) ; (b) $N = 2$ Multiwavelet With Optimal Weights ($m = 1$).

(iv) Fit the $N = 2$ multiwavelet regression model by weighted least squares using the optimal weights constructed in Chapter 3;

(v) Overlay plots of the values of $\eta(x, \theta)$ from (i) and the fitted values from (iii) and (iv) on same page

(vi) Simulate the model

$$y_i = \eta(x_i, \theta) + \varepsilon, \;\; \varepsilon \sim N(0, \, \sigma^2),$$

for some known value of σ^2 and repeat (ii) - (v).

We observe that the fitted wavelet regression models picked up the main features of the data in all our examples. However, the fitted models appear to have retained some features of the primary wavelets used in the approximation. For instance, the fitted Haar wavelet regression models exhibit the step function

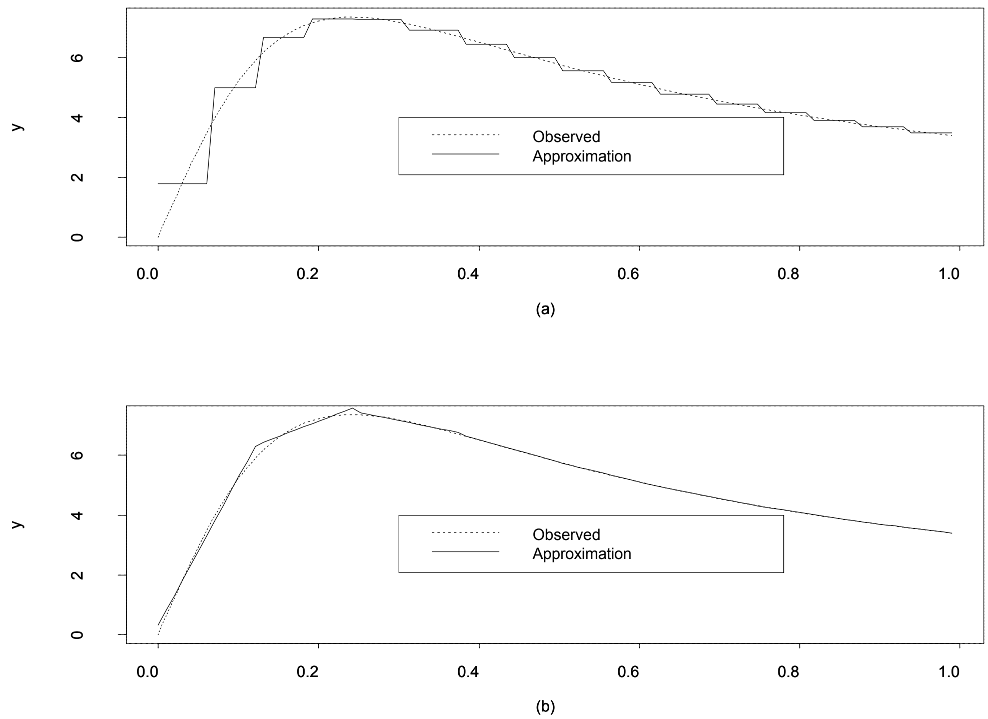


Figure 6: The Quadratic Michaelis-Menten Model Function: (a) Haar Wavelet ($m = 3$) ; (b) $N = 2$ Multiwavelet With Optimal Weights ($m = 2$).

feature of its primary wavelet $\psi(x)$ (see (2.47)). On the other hand the fitted $N = 2$ multiwavelet regression models appear to exhibit the sharp-curve feature of one of its primary wavelets ${}_2w_0(x)$ (see (1.56) and Figure 1) especially at points where the functions being approximated change direction.

1. Ion Transport Model:

The model function

$$\eta(x, \theta) = \theta_1(1 - \theta_2 exp(-\theta_3 x))$$

is commonly used to describe data from ion transport experiments (e.g. chloride ions), through blood cell walls.The function $\eta(x, \theta)$ measures the concentration of the ions at time x. In this model, the parameters have physical meanings.

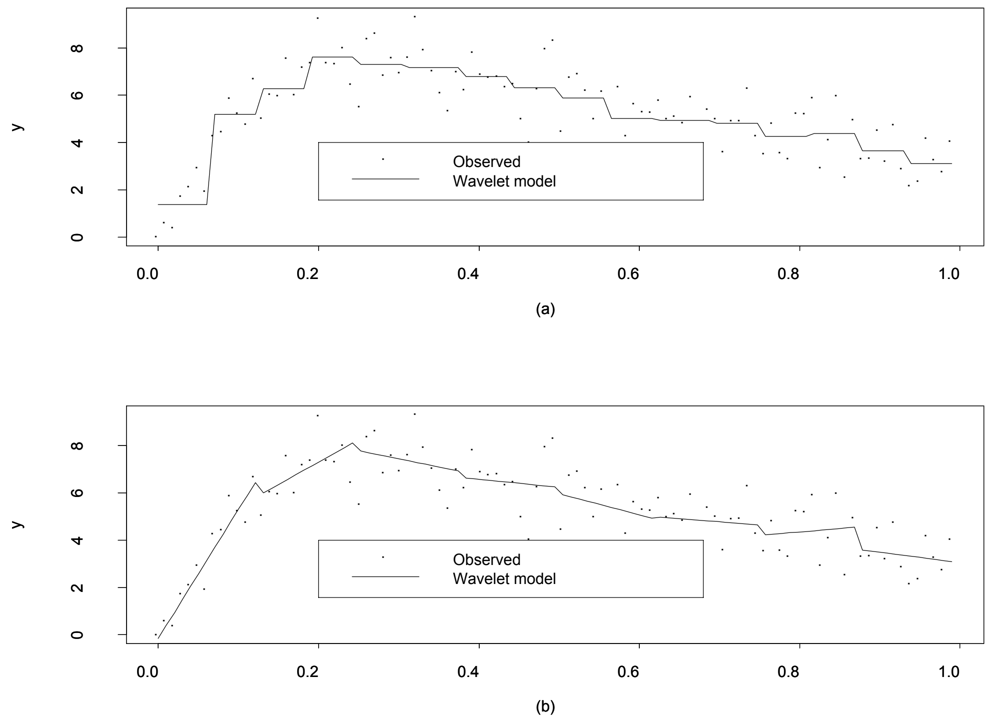
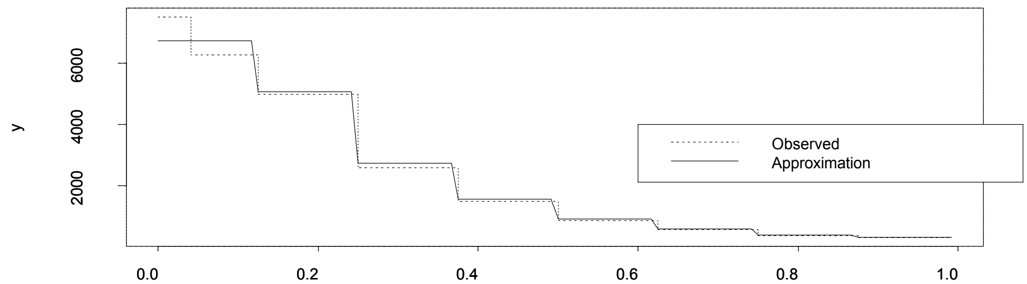


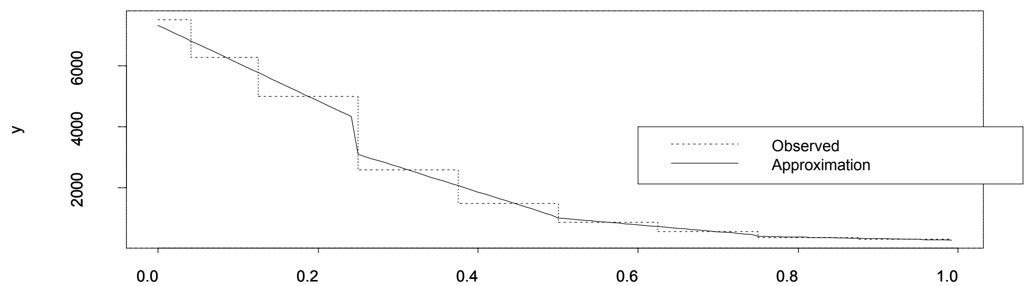
Figure 7: Simulated Data from the Quadratic Michaelis-Menten Model: (a) Haar Wavelet ($m = 3$) ; (b) $N = 2$ Multiwavelet With Optimal Weights ($m = 2$).

The parameter θ_1 is interpreted as the final percentage concentration, θ_3 is a rate constant and θ_2 accounts for the unknown initial and final concentrations and the unknown initial reaction time. The parameters, $\hat{\theta} = (39.09,\ 0.828,\ 0.159)^T$, we used are the estimates obtained by Bates and Watts (1988, pg. 93) from real data with $n = 50$. The range of x, $x \in [0, 32]$, which is slightly different from the interval used in Bates and Watts (1988), was transformed as discussed earlier. Figure 4 shows the plots obtained by following steps (i) - (v) outlined above. Using the values generated from $\eta(x;\theta)$ (see step (i)) as data we calculated the mean squared error (MSE). The MSE from the Haar fit is approximately 5.24 and 0.509 from the $N = 2$ multiwavelets.

Following step (vi) with $\hat{\sigma}^2 = 3.534$ we obtained Figure 5. The MSE from the Haar fit is approximately 8.732 and 5.313 from the $N = 2$ multiwavelets.



(a)



(b)

Figure 8: Amount of Saccharin versus Time(hrs) Interval: (a) Haar Wavelet (m = 2) ; (b) $N = 2$ Multiwavelet With Optimal Weights ($m = 1$).

2 The Quadratic Michaelis - Menten Model:

The Michaelis - Menten Model

$$\eta(x, \theta) = \frac{\theta_1 x}{\theta_2 + x + \theta_3 x^2}$$

is popular in enzyme kinetic experiments. It relates the "velocity" of an enzymatic reaction to the substrate concentration x. Bates and Watts (1988 pg. 114) also used this model to analyse data from an experiment on the utilization of nitrite in bush beans. The parameters, $\theta = (1254, 20.5, 350.44)$ were carefully choosen to obtain the shape seen in Figure 6. The sample size was $n = 100$ and $x \in [0, 16]$ was transformed as in Section 1. The mean squared error for the Haar and $N = 2$ multiwavelets models are approximately 0.162 and 0.008 respectively.

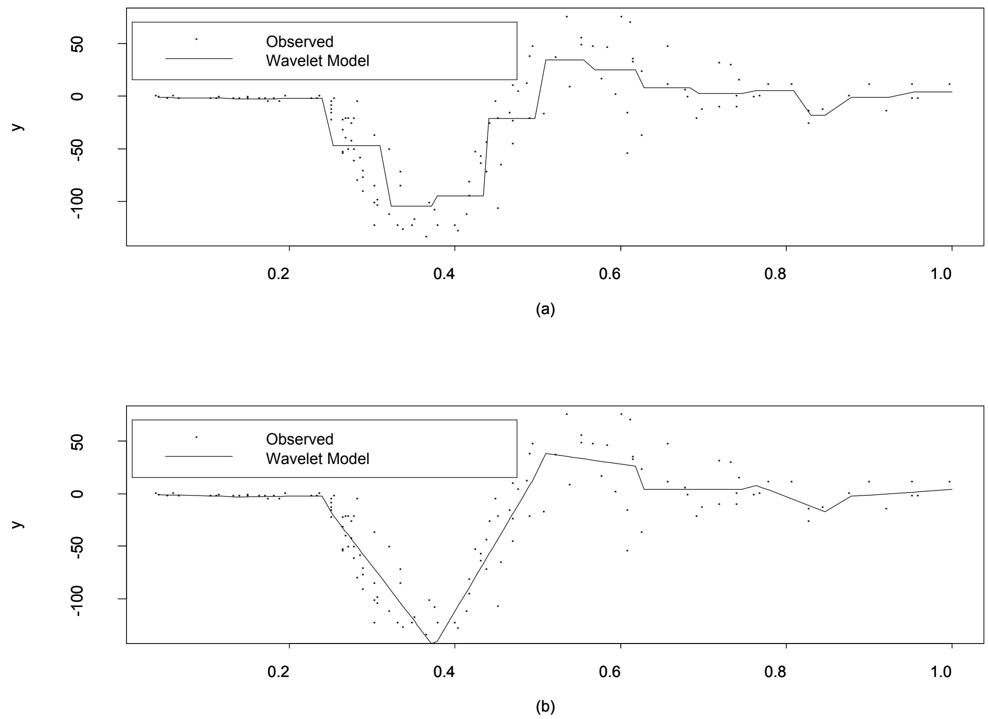


Figure 9: A Plot of the Motorcycle Data Set: (a) Haar Wavelet ($m = 3$) ; (b) $N = 2$ Multiwavelet With Optimal Weights ($m = 2$).

Again, we apply step (vi) with $\hat{\sigma}^2 = 0.694$ to obtain Figure 7. The MSE from the Haar fit is approximately 1.136 and 1.102 from the $N = 2$ multiwavelets.

3 Metabolism of Saccharin Compounds:

Here, we use real data from an experiment on the metabolism of saccharin compounds provided by Renwick (1982). A rat is given a single bolus of saccharin. At given time intervals, the amount of saccharin accumulated in the urine of the rat is measured. The response is the level of radioactivity of the urine which was converted to amount of saccharin in micrograms (μg). The proposed integrated model is

$$\eta(\mathbf{x}, \theta) = \frac{\theta_3}{\theta_1} e^{-\theta_1 x_1}(1 - e^{-\theta_1 x_2}) + \frac{\theta_4}{\theta_2} e^{-\theta_2 x_1}(1 - e^{-\theta_2 x_2})$$

where η is the amount of saccharin excreted during an interval, x_1 is interval starting time, and x_2 is the length of the interval.

In Figure 8, we plot the observed excreted amount versus the time interval, (e.g. 0 - 5), scaled as discussed in step (ii) above. This example is aimed at showing that the Haar model performs better in approximating step functions.

4. The Motorcycle Data:

Our next example illustrates the flexibility of wavelets in describing nonlinear experiments even when the precise mathematical structure of the function describing the experiment is unknown. The motorcycle data set taken from Hardle (1990) are measurements of the head acceleration (y) of a post mortem human test object after a simulated impact with motorcycles in a given time (x). The nonlinear model which describes the experiment is unknown. A plot of the fitted models are shown in Figure 9.

We observe that the higher the degree of nonlinearity of the experiment, the higher the value of m, hence more design points required to obtain a good approximation.

5. Concluding Remarks

In this work we have outlined the results of our investigation into the use of wavelets in designing nonlinear experiments. Wavelets was introduced into the problem by transforming the nonlinear regression model describing the experiment into a wavelet regression model with disturbance function. Throughout our discussion the mathematical structure of the underlying nonlinearity was not assumed known. Using examples, we have shown that the multiwavelet bases is capable of capturing the general features of any nonlinearity in an experiment, though still retaining some features of its primary wavelets. This calls for further studies into the use of wavelet bases with smoother properties in designing nonlinear experiments.

For the simplest case ($m = 0$) of the mth order Haar wavelet regression model we have been able to show that no non-symmetric design will be optimal. We also observe that the optimal weights and designs constructed in Chapter 3 (with no symmetry constraint) were symmetric. Our conjecture is that, in general, no non-symmetric design will be optimal. Investigation into a formal proof is also another area for further research. We have also shown that the classical D-optimal design proposed by Herzberg and Traves (1994) is simultaneously Q-, A-, D- and G-optimal and provided a proof we sense is much simpler. The continuous uniform design which we showed to be optimal for the biased Haar model can be considered a smoothed version of this design.

The optimal weights and designs constructed in Chapter 3 are for general wavelet regression models. Therefore, once the wavelets to be used in the approximation is choosen, the corresponding optimal weights and designs can be constructed. Strategies for implementing the designs have been outlined including a situation where the experimenter wishes to take more observations from some region of the design space.

Apart from the problems mentioned earlier which require further studies, some other open problems are :

(1) D-, E- and G-optimal weights and designs for wavelet regression

models subject to the condition of unbiasedness.

(2) Robust minimax weights and designs for the general $N = 2$ multiwavelet regression model.

(3) Robust infinitesimal designs for wavelet regression models.

(4) Robust designs for nonlinear models when the nonlinear function $\eta(x)$ is estimated by wavelet versions of kernel estimators (see Antoniadis et al (1994)).

(5) Robust weights and designs for wavelet regression models when the estimators of the parameters are generalized M-estimators or other robust estimators.

(6) Robust designs for biased wavelet regression models with autocorrelated errors.

(7) Robust minimax weights and designs for biased wavelet regression models with heteroscedastic errors.

We hope that this work will motivate further research in the direction of constructing designs for wavelet regression models and the construction of wavelets for design purposes.

Reference

Alpert, B. K. (1992). Wavelets And Other Bases For Fast Numerical Linear Algebra. *In Wavelets: A Tutorial In Theory And Applications, C. K. Chui (ed.). Academic Press, Boston,* 181-216.

Andersson, L., Hall, N., Jawerth, B., and Peters, G. (1993). Wavelets On Closed Subspaces Of The Real Line. *In Recent Advances In Wavelet Analysis, L. L. Schumaker and G. Webb (eds.). Academic Press, Boston*, 1-61.

Antoniadis, A., Gregoire, G., and McKeague, I. W. (1994). Wavelet Methods For Curve Estimation. *J. Amer. Statist. Assoc.* **89**, 1340-1353.

Atkinson, A. C., and Fedorov, V. V. (1975). The Design of Experiments for Discriminating Between Two Rival Models. *Biometrika* **62** , 57-70.

Atkinson, A. C., and Hunter, W. G. (1968). The Design of Experiments for Parameter Estimation. *Technometrics* **10**, 271-289.

Bandemer, H., Nather, W., and Pilz, J. (1987). Once more: Optimal Experimental Design for Regression Models (with discussion). *Statistics* **18**, 171-217.

Bates, D. M., and Watts, D. G. (1980). Relative Curvature Measures of Nonlinearity (with discussion). *J. Roy. Stat. Soc. Ser.* **B 42**, 1-25.

Bates, D. M., and Watts, D. G. (1981). Parameter Transformations for Improved Approximate Confidence Regions in Nonlinear Least Squares. *Ann. Statist.* **9**, 1152-1167.

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Applications.* Wiley, New York.

Bazaraa, M. S., and Shetty, C. M. (1976). *Foundations of Optimization. Lecture Notes In Economics And Mathematical Systems.* M. Beckmann and H. P. Kunzi (*eds.*), Springer-Verlag, New York.

Begg, C. B., and Kalish, L. A. (1984). Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model. *Biometrics* **40**, 409-420.

Berenblutt, I. I., and Webb, G. I. (1974). Experimental Design in the Presence of Autocorrelated Errors. *Biometrika* **61**, 427-437.

Bickel, P. J., and Herzberg, A. M. (1979). Robustness of Design Against Autocorrelation in Time I: Asymptotic Theory, Optimality for Location and Linear Regression. *Ann. Statist.* **7**, 77-95.

Bischoff, W. (1992). On Exact D-optimal Designs for Regression Models With Correlated Observations. *Ann. Inst. Statist. Math* **44**, 229-238.

Bischoff, W. (1993). On D-optimal Designs for Linear Models Under Correlated Observations With an Application to a Linear Model With Multiple Response. *J. Statist. Plann. Inference* **37**, 69-80.

Box, M. J. (1968). The Occurence of Replications in Optimal Designs of Experiments to Estimate Parameters in Nonlinear Models. *J. Roy. Statist. Soc. Ser.* **B 30**, 290-302.

Box, M. J. (1970). Some Experiences with a Nonlinear Experimental Design Criterion. *Technometrics* **12**, 569-589.

Box, M. J. (1971). Bias in Nonlinear estimation (with discussion). *J. Roy. Statist. Soc. Ser.* **B 32**, 171-201.

Box, G. E. P., and Draper, N. R. (1959). A Basis for the Selection of a Response Surface Design. *J. Amer. Statist. Assoc.* **54**, 622-654.

Box, G. E. P., and Hill, W. J. (1967). Discrimination Among Mechanistic Models. *Technometrics* **9**, 57-71.

Box, G. E. P., and Hunter, W. G. (1965). Sequential Design of Experiments for Nonlinear Models. *Proc. of the IBM Scientific Computing Symposium on Statistics. October 21-23, 1963*, 113-137.

Box, G. E. P., and Lucas, H. L. (1959). Design of Experiments in Nonlinear Situations. *Biometrika* **46**, 77-90.

Chaloner, K. (1984). Optimal Bayesian Experimental Design for Linear Models. *Ann. Statist.* **12**, 283-300.

Chan, L. (1991). Optimal Design for Estimation of Variance in Nonparametric Regression Using First Order Differences. *Biometrika* **78**, 926-929.

Chang, D. (1979) Design of optimal control for a regression problem. *Ann. Statist.* **7**, 1078-1085 (Corr: V8 p1402).

Chernoff, H. (1953). Locally Optimal Designs for Estimating Parameters. *Ann. Math. Statist.* **30**, 586-602.

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

Chui, C. K., and Quak, E. (1992). *Wavelets On A Bounded Interval. In Numerical Methods Of Approximation Theory.* D. Braess and L. L. Schumaker (*eds.*), Birkhäuser, Basel, 1-24.

Cochran, W. G. (1973). Experiments for Nonlinear Functions. *J. Amer. Statist. Assoc.* **68**, 771-778.

Cohen, A., Daubechies, I., and Vial, P. (1992). Wavelets And Fast Wavelet Transforms On The Interval. *Submitted To Applied And Computational Harmonic Analysis.*

Constantine, G. M. (1989). Robust Designs for Serially Correlated Observations. *Biometrika* **76**, 245-251.

Cook, R. D., Tsai, C. L., and Wei, B. C. (1986). Bias in Nonlinear Regression Models. *Biometrika* **73**, 615-623.

Daubechies, I. (1992). *Ten Lectures On Wavelets.* Society For Industrial And Applied Mathematics(*SIAM*) (*Pubs.*).

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.* L. L. Schumaker and G. Webb (*eds.*). Academic Press, Boston, 237-258.

Daubechies, I., and Lagarias, J. C. (1988). Two-Scale Difference Equations I: Global Regularity of Solutions. *Technical Report, Bell Laboratories.*

Elfving, G. (1952). Optimum Allocation in Linear Regression Theory. *Ann. Math. Statist.* **23**, 255-262.

Fan, J. (1992). Design-adaptive Nonparametric Regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.

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

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. Roy. Soc. of London, Ser. A* **222**, 309-368.

Ford, I., Kitsos, C. P., and Titterington, D. M. (1989). Recent advances in Nonlinear Experimental Design. *Technometrics* **31**, 49-60.

Ford, I., Titterington, D. M., and Wu, C. F. J. (1985). Inference and Sequential Design. *Biometrika* **72**, 545-551.

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

Haines, L. M. (1993). Optimal Designs for Nonlinear Regression. *Comm. In Statist., Theory and Methods* **22**, 1613-1627.

Hamilton, D. C., and Watts. D. G. (1985). A Quadratic Design Criterion for Precise Estimation in Nonlinear Regression Models. *Technometrics* **27**, 241-250.

Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *J. Amer. Statist. Assoc.* **69**, 383-393.

Hărdle, W. (1990). *Applied Nonparametric Regression.* Econometric Society Monographs No. 19. Cambridge University Press, New York.

Herzberg, A. M., and William, T. N. (1994). An Optimal Experimental Design For The Haar Regression Model. *Can. J. of Statist.* **22**, 357-364.

Hill, W. J., Hunter, W. G., and Wichern, D. W.(1968). A Joint Design Criterion for the Dual Problem of Model Discrimination and Parameter Estimation. *Technometrics* **10**, 145-160.

Hoel, P. G. (1958). Efficiency Problems in Polynomial Estimation. *Ann. Math. Statist.* **29**, 1134-1145.

Hotelling, H. (1944). Some Improvements in Weighing and Other Experimental Techniques. *Ann. Math. Statist.* **15**, 297-306.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **35**, 73-101.

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

Hunter, W. G., and Reiner, A. M. (1965). Designs for Discriminating Between Two Rival Models. *Technometrics* **7**, 307-323.

Jaffard, S., and Laurencot, Ph. (1992). *Orthonormal Wavelets, Analysis Of Operators And Applications To Numerical Analysis. In Wavelets: A Tutorial In Theory And Applications.* C. K. Chui (*ed.*), Academic Press, Boston, 543-601.

Jaffard, S., and Meyer, Y. (1989). Wavelet Bases In Open Sets Of $\Re^n$. *J. Math. Pures et Appl.* **68**, 95-108.

Jenkins, G. M., and Chanmugam, J. (1962). The Estimation of Slope When Errors are Autocorrelated. *J. Roy. Statist. Soc. Ser.* **B 24**, 199-214.

Karson, M. J., Manson, A. R., and Hader, R. J. (1969). Minimum Bias Estimation and Experimental Design for Response Surfaces. *Technometrics* **11**, 461-475.

Kiefer, J. (1959). Optimal Experimental Designs. *J. Roy. Statist. Soc. Ser.* **B 21**, 273-319.

Kiefer, J., and Wolfowitz, J. (1959). Optimal Designs in Regression Problems. *Ann. Math. Statist.* **30**, 271-294

Kiefer, J., and Wynn, H. P. (1981). Optimum Balanced Block and Latin Squares Designs for Correlated Observations. *Ann. Statist.* **9**, 737-757.

Kiefer, J., and Wynn, H. P. (1984). Optimum and Minimax Exact Treatment designs for One-dimensional Autoregressive Error Process. *Ann. Statist.* **12**, 414-450.

Li, K. C. (1984). Robust Regression Designs when the Design Space Consists of Finitely Many Points. *Ann. Statist.* **12**, 269-282.

Li, K. C., and Notz, W. (1982). Robust Designs for Nearly Linear Regression. *J. Statist. Plann. Inference* **6**, 135-151.

Liu, S. X., and Wiens, D. P. (1998). Robust Designs for Approximately Polynomial Regression. *To appear in J. Statist. Plann. Inference.*

Mallat, S. (1989). Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\Re)$. *Trans. Amer. Math. Soc.* **315**, 69-87.

Marcus, M. B., and Sacks, J. (1976). *Robust Designs for Regression Problems. In Statistical Theory and Related Topics II.* S. S. Gupta and D. S. Moore (*eds.*), Academic Press, New York, 245-268.

Maxim, L. D., Hendrickson, A. D., and Cullen, D. E. (1977). Experimental Design for Sensitivity Testing: The Weibull Model. *Technometrics* **19**, 405-412.

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

Mitchell, T., Sacks, J., and Ylvisaker, D. (1994). Asymptotic Bayes Criteria for Nonparametric Response Surface Design. *Ann. Statist.* **22**, 634-651.

Müller, H. (1984). Optimal designs for Nonparametric Kernel Regression. *Statist. Probab. Lett.* **2**, 285-290.

Pesotchinsky, L. (1982). Optimal Robust Designs: Linear Regression in R^k. *Ann. Statist.* **10**, 511-525.

Pilz, J. (1983). Bayesian Estimation and Experimental Design in Linear Regression Models. *Teubner-Texte zur Math.* **55**. Teubner, Leipzig.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Renwick, A. G. (1982). *Pharmacokinetics in Toxicology. In Principles and Methods of Toxicology.* Wallace Hayes (*ed.*), Raven Press, New York, 659-710.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science* **4**, 409-423 (C/R: 423-435).

Sacks, J., and Ylvisaker, D. (1966). Designs for Regression Problems With Correlated Errors. *Ann. Math. Statist.* **37**, 66-89.

Sacks, J., and Ylvisaker, D. (1968). Designs for Regression Problems With Correlated Errors; Many Parameters. *Ann. Math. Statist.* **39**, 49-69.

Sacks, J., and Ylvisaker, D. (1978). Linear Estimation for Approximately Linear Models. *Ann. Statist.* **6**, 1122-1137.

Sacks, J., and Ylvisaker, D. (1984). Some Model Robust Designs in Regression. *Ann. Statist.* **12**, 1324-1348.

Sacks, J., and Ylvisaker, D. (1985). Model Robust Design in Regression: Bayes Theory. *Proc. Berk. Conf. in Honor of Jerzy Neyman and Jack Kiefer* **2**, 667-679.

Smith, K. (1918). On the Standard Deviations of Adjusted and Interpolated values of an Observed Polynomial Function and its Constants and the Guidance they Give Towards a Proper Choice of the Distribution of Observations. *Biometrika* **12**, 1-85.

Steinberg, D. M., and Hunter, W. G. (1984). Experimental Design: Review and Comment. *Technometrics* **26**, 71-97.

??trang, G. (1989). Wavelets and Dilation Equations: A Brief Introduction. *SIAM Review* **31**, 614-627.

??itterington, D. M. (1980). Aspects of Optimal Design in Dynamic Systems. *Technometrics* **22**, 287-299.

??an der Linde, A. (1985). Interpolation of Regression Functions in Reproducing Kernel Hilbert Spaces. *Statistics 16*, 351-361.

??ellila, S., and Llosa, R. (1992). A Result in Optimal Design with Application in Particle Physics. *Comm. In Statist., Theory and Methods* **21**, 203-212.

Wald, A. (1943). On the Efficient Design of Statistical Investigations. *Ann. Math. Statist.* **14**, 134-140.

Walter, G. G. (1995). Pointwise Convergence Of Wavelet Expansions. *J. of Approximation Theory* **80**, 108-118.

White, L. V. (1973). An Extension to the General Equivalence Theorem for Nonlinear Models. *Biometrika* **60**, 345-348.

White, H. (1981). Consequences and Detection of Misspecified Nonlinear Regression Models. *J. Amer. Statist. Soc.* **76**, 419-433.

Wiens, D. P. (1990). Robust Minimax Designs For Multiple Linear Regression. *Linear Algebra Appl.* **127**, 327-340.

Wiens, D. P. (1991). Designs for Approximately Linear Regression: Two Optimality Properties of Uniform Designs. *Statist. Probab. Lett.* **12**, 217-221.

Wiens, D. P. (1992). Minimax Designs for Approximately Linear Regression. *J. Statist. Plann. and Inference* **31**, 353-371.

Wiens, D. P. (1993). Robust Sequential Designs for Approximately Linear Models. *Can. J. of Statist.* **24**, 67-79.

Wiens, D. P. (1994). Robust Designs for Approximately Linear Regression: M-estimated Parameters. *J. Statist. Plann. Inference* **40**, 135-160.

Wiens, D. P. (1996). Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroscedastic Errors. *Technical Report 96.01, University of Alberta.*

Wiens, D. P. (1996a). Asymptotics of Generalized M-Estimation of Regression and Scale With Fixed Carriers in an Approximately Linear Model. *Statist. Probab. Lett.* **30**, 271-285.

Wiens, D. P., and Zhou, J.(1996a). Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors. *J. Statist. Plann. Inference* **55**, 95-106.

Wiens, D. P., and Zhou, J. (1996b). Minimax Designs for Approximately Linear Models With AR(1) Errors. *Submitted to the J. Statist. Plann. Inference.*

Wiens, D. P., and Zhou, J.(1997). Robust Designs Based on the Infinitesimal Approach. *To appear in J. Amer. Statist. Assoc.*.

Williams, R. M.(1952). Experimental Designs for Serially Correlated Observations. *Biometrika* **39**, 151-167.