EXACT MINIMAX WAVELET DESIGNS FOR DISCRIMINATION

YI LIU

EXACT MINIMAX WAVELET DESIGNS FOR DISCRIMINATION

by

©Yi Liu

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

Department of Mathematics and Statistics
Memorial University of Newfoundland

January 2004 Submitted

St. John's Newfoundland Canada

Your file Votre référence
ISBN: 0-494-09926-7
Our file Notre référence
ISBN: 0-494-09926-7

Abstract

In this thesis, we consider the problem of constructing designs in order to determine the number of wavelet terms that should be included in the wavelet representation of unknown nonparametric response curves. Our approach is to choose designs that will maximize, in some sense, the difference between the better model and the other competing wavelet models. Simulated annealing algorithm is developed to carry out exact, rather than approximate, minimax designs for discrimination between competing wavelet regression models. Sequential and nonsequential designs are discussed along with some examples based on the multiwavelet system and Daubechies wavelet system.

Acknowledgement

I would like to express my sincere gratitude to Dr. A. J. Oyet for granting me the privilege to write this thesis under his supervision and for his support, both financial and otherwise, through out the course of my program. He patiently read through my work several times and offered his very valuable suggestions. This work would have been impossible without his guidance and help.

I thank my professors who have shared with me their wealth of statistical knowledge at various points in the course of my program. The staff of the Math. & Stat. general office were very helpful and nice.

Also, my sincere thanks to my parents, friends and neighbours whose assistance and encouragement have seen me to the completion of my program.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Several authors have used wavelets in representing nonparametric mean response functions. See for example Antoniadis, Gregoire and McKeague (1994), Wegman, Poston and Solka (1996) and Oyet (2003). One unresolved question is how to determine the number of wavelet terms one should use in the approximation or the maximum level m of approximation one should use. Antoniadis et al. (1994) discussed the wavelet version of the Gasser-Müller estimation and recommended examining $m = 3,\ 4$ and 5 for sample sizes between 100 and 200. The choice of m was then made based on the value of the cross validation function. Oyet (2002) in his study of a modified wavelet version of the Gasser-Müller estimator notes that this approach may lead to an estimated response function with spikes and wiggles that are not features of the "true" response. He then suggested that the structure of the response function should be considered in choosing a value for m. These two approaches assume implicity that data $\{x_i, y_i\}_{i=1}^n$ is available and the choice of m can be made at the analysis stage. In this study, we consider the problem of choosing m at the design stage. Our view is that if we can select the design points x_i in such a way that the observations $\{x_i, y_i\}$ can be used to maximize, in some sense, the difference between competing models, then the choice is made easier at

the analysis stage. Our main interest, in this study, is therefore the construction of designs for discrimination between competing non-parametric models, based on their wavelet representation. Throughout our discussion, we assume that there are only two competing models of order m and $m-1$. Certainly the approach can be easily extended to three models, of order m, $m-1$ and $m-2$ or m, $m+1$ and $m+2$.

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. For a recent survey on the use of wavelets in signal processing, see Rioul and Vetterli (1991). In recent years, there has been growing interest in the application of wavelets to statistical problems. Due to the flexible nature of wavelet systems , successful statistical applications have already been made in the analysis of time series, outlier detection, nonparametric curve estimation and in the construction of classical and robust designs. Herzberg & Traves (1994) are perhaps the first to discuss classical designs for wavelet regression models using the Haar wavelets as the regressors. A recent discussion connecting wavelets with problems in nonparametric statistical inference has been made by Wegman (1991).

Designs for estimation and discrimination between competing regression models with a prespecified parametric form for the mean response curve have been documented in a substantial body of literature. In his discussion, Anderson (1962) proposed the following decision rule. For a given set of levels $(\alpha_1, \alpha_2, \cdots, \alpha_n)$, the procedure chooses the largest integer in $\{1, 2, \cdots, n\}$, for which the F-test in the model $h_g(x)$, where $h_g(x)$ is a specified linear model, rejects the hypotheses $H_0 : v_g = 0$ at the level α_g, $g = 1, 2, \cdots, n$. In a paper, Spruill (1990) considered similar problems of testing the degree of a polynomial mean and determined the optimal approximate

design with respect to a maximum criterion which maximizes the local power of the F-test. Dette and Haller (1998) used the theory of canonical moments to obtain the optimal discriminating designs for a Fourier regression model. Atkinson and Cox (1974) developed experimental designs for discriminating between several regression models based on an extension of D-optimum design theory. Atkinson and Fedorov (1975) described T-optimal designs for discriminating between three or more rival regression models, which need not be linear in the parameters. The criterion used by these authors are all variance based criteria which assumes that the structure of the mean response function is correctly specified. That is, these criteria does not account for the uncertainty in the true structure of the response curve. The effect of slight deviations, from the assumed response, on designs has been highlighted by Box and Draper (1959). In this paper, we therefore use mean squared error based criteria. The maximin criterion we have used depends directly on the noncentrality parameters of the distribution of F-statistic for testing the degree of a wavelet approximation. We also considered a criterion which is the ratio of the determinant of mean squared errors. Details are provided in Section 2.3. For some alternative techniques of testing the degree of a polynomial mean, see Anderson (1962), Hoel (1968), Dette (1995) and their references. For more general models, see Stone (1981) and references.

Instead of prespecifying an assumed parametric form for the mean response, we adopt a nonparametric set-up. Now, due to the difficulties associated with the unknown mathematical structure of nonparametric response functions, studies has been limited in the nonparametric case. Chan (1991) used first order differences to construct designs for estimation of the error variance in nonparametric regression. A Bayesian approach was adopted by Mitchell, Sacks and Ylvisaker (1994) to represent

the unknown response by a random function. A study of the relative merits of the convolution and evaluation kernel approaches to nonparametric regression has been made by Chu and Marron (1991). We apply recent advances in wavelet theory to transform the classical nonparametric design problem into a robust minimax design problem with contamination, in the wavelet domain, through a wavelet expansion of the mean response curve $\eta(x)$. In this way, we avoid the difficulties associated with the unknown structure of $\eta(x)$. Due to the adaptive nature of wavelets, the exact structure of $\eta(x)$ need not be known. We note that, in this wavelet representation of $\eta(x)$ only a finite number of terms can be estimated in actual computation. Thus, the remainder terms we shall represent by $f(x)$ accounts for the uncertainty about the exact mathematical structure of the response function.

Our focus is on constructing exact integer valued rather than approximate continuous designs for discriminating between wavelet models for nonparametric response curves. For this reason, we employ the simulated annealing algorithm, which has previously been used for constructing designs by several authors. Bohachevsky, Johnson and Stein (1986) gave a general discussion of simulated annealing algorithm with an application to the optimization of functions having many local extrema. The advantages of their method are the ability to migrate through a sequence of local extrema in search of the global solution and to recognize when the global extremum has been located. Haines (1987) reported the application of the annealing algorithm to the construction of exact D-, I-, and G-optimal designs for polynomial regression of degree 5 on the interval $[-1, 1]$ and for the second-order model in two factors on the design space $[-1, 1] \times [-1, 1]$. Meyer and Nachtsheim (1988) have developed it to construct D-Optimal Experimental Designs. Fang and Wiens (2000) introduced a new approach using a simulated annealing algorithm to construct integer-valued,

minimax robust designs for approximately linear models with possible heteroscedastic errors. Zhou (2001) constructed integer-valued, minimax robust designs for approximately linear models with possible correlated errors. And Oyet and Wiens (2003) notes that the use of this algorithms has made it possible to exhibit exact designs for multiwavelet models of higher orders. Typically, experimenters adopt approximate design theory to construct approximate continuous designs. The number of observations n_i, out of a total of n, allocated to a particular design point x_i is, in this case, not an integer. The experimenter then has to approximate n_i while hoping that the design will be at least near optimal.

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), Huber (1975), Box and Draper (1975), Steinberg and Hunter (1984) and Pukelsheim (1993). In Section 1.1 the classical design problem is defined. The robust design problem is discussed in Section 1.2. Some background on wavelet theory relevant to our work is provided in Section 1.3.

1.1 The classical design problem

Consider the following model which describes the relationship between the response $y(\mathbf{x})$ and the independent variable x:

$$y(x_i) = \eta(x_i) + \varepsilon_i, \ i = 1, 2, ..., n, \tag{1.1}$$

where $y(x_i) \in \mathcal{R}$ is an observable random variable; $x_i \in \mathcal{S} \subseteq \mathcal{R}^p$ is the ith vector of some design variables; $\varepsilon_i \in \mathcal{R}$ is a sequence of uncorrelated random unobservable errors with mean zero and common variance σ^2 and $\eta(x_i)$ is the value of some square

integrable, possibly nonlinear function η at x_i. If the response surface $\eta(x)$ can be written as

$$\eta(x) = \eta(x; \beta_0) = q^T(x)\beta_0, \tag{1.2}$$

the problem is said to be a linear design problem; it is nonlinear otherwise.

The design problem of interest arises in the following way: an experimenter plans to observe $y(x)$ at n not necessarily distinct values of the design variable x chosen from a design space S in order to maximize the accuracy of estimating a regression function $\eta(x)$. Since S contains more than n points, the problem is that of finding the "best" (in some sense) design points at which $y(x)$ will be observed. Classical and robust design theories were developed to determine the best design points. The difference between classical and robust theory arise from their underlying assumptions. In classical design theory, it is assumed that

- The model representing $y(x)$ is exact and $\eta(x)$ is correctly specified;
- The errors ε_i are uncorrelated and have variance σ^2.

The earliest references to the classical theory of regression designs and related optimality result can be found in Smith (1918) and Plackett & Burman (1946). Subsequent work was done by Elfving (1952,1955,1956), Chernoff (1953) and others. The concept of continuous design was first introduced by Kiefer & Wolfowitz (1959). Fisher (1922) first studied the subject of nonlinear experimental design and White (1973) proved the nonlinear version of the Kiefer-Wolfowitz equivalence theorem. The first comprehensive volume on the theory of optimal experimental design was written by Fedorov (1972). The book by Silvey (1980) gives a very compact description of the theory of optimal design for estimation in linear models. Discrete optimal designs are covered in the book by Shah and Sinha (1989).

In order to apply optimal design theory to (1.1) a criterion is required for comparing experiments and then selecting the 'best' design with respect to the specified criterion. For parametric models, this criterion is often taken in optimal design theory to be a monotonic increasing function $\Phi(M(\hat{\beta}))$ of the mean squared error(MSE) matrix of an estimator of β_0. If the estimator $\hat{\beta}$ is unbiased, the MSE reduces to the covariance matrix. Mathematically, the classical design problem can be stated as:

$$\min_{\{x_1,\ldots,x_n \in S\}} \Phi(V(\hat{\beta})), \tag{1.3}$$

where $V(\hat{\beta})$ is the covariance of $\hat{\beta}$.

Some optimality criteria commonly found in the literature are:

(1) D-optimality: Here, $\Phi(\cdot) = det(\cdot)$, where $det(\cdot)$is the determinant function.

(2) A-optimality: In this case, $\Phi(\cdot) = tr(\cdot)$, where $tr(\cdot)$ is the trace function.

(3) E-optimality: Here, $\Phi(\cdot) = \lambda_{max}(\cdot)$, where λ_{max} is the maximum eigenvalue function.

(4) Φ_q-optimality: Here, $\Phi_q(\cdot) = \sum_{i=1}^{p} \lambda_i(\cdot)^q$, where $\lambda_i(\cdot)$ is the i-th eigenvalue function.

(5) G-optimality: Here, $\Phi(\cdot) = \max_{x \in \mathcal{S}}(\cdot)$, where $\max_{t \in T}(\cdot)$ is the maximization function subject to constraint $a(x)^T\beta$.

(6) Q-optimality: Here, $\Phi(\cdot) = \int_{a(x)^T\beta}(\cdot)\xi dx$, where $\int_{a(x)^T\beta}(\cdot)\xi dx$ is the integral function subject to constraint $a(x)^T\beta$.

Hoel (1958) noticed that the D- and the G-optimum designs coincide in the model of a one-dimensional polynomial regression, and Kiefer and Wolfowitz (1960) proved that this is true for every linear model. The Kiefer-Wolfowitz equivalence

theorem provides a tool for verifying whether a given design is D-optimal or not. This tool has been extensively used by others, subsequently leading to a vast development of the subject matter. Later, Kiefer (1961, 1962) extended the notion of D-optimality to D_S-optimality criterion when one is interested in a subset of parameters, and established the corresponding equivalence theorem. Silvey (1978) and Pukelsheim (1980) studied the problem of sigularity for this situation. Fedorov (1972) extended the notion of A-optimality to linear optimality criterion and derived the corresponding equivalence theorem. Whittle (1973) and Kiefer (1974) extended to general equivalence theorems.

1.2 The robust design problem

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

(1) the model may not be exactly correct;

(2) the errors ε may not be uncorrelated.

It is well known that in most cases where the form of $\eta(x)$ 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 $\eta(x, \theta_0)$ with the hope that the remainder terms are negligible. Under these conditions, the least squares estimator of θ_0 is biased and the classical designs which minimize variance alone are no longer "optimal" due to the bias.

Box and Draper (1959) revealed the inherent dangers of designing a regression experiment on the basis that (1.1) is exactly correct. They studied the case where the experimenter fits a polynomial of first degree whereas the correct model is quadratic.

The estimate is subject to "bias error" engendered by the model misspecification as well as "variance error" due to sampling. They draw the conclusion that any slight model misspecification can erode any supposed gains arising from the use of a design which minimize variance alone. They argued that a more appropriate optimality criteria is the Integrated Mean Squared Error (IMSE) of the estimate $\hat{\eta}$ of the "true" response surface η over the design space S. That is,

$$\mathcal{L} = \frac{n\Omega}{\sigma^2}\int_S E\{[\hat{\eta}(x)-\eta(x)]^2\}dx = ISB + IV \tag{1.4}$$

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

$$\Omega^{-1} = \int_S dx, \quad ISB = \frac{n\Omega}{\sigma^2}\int_S \{E[\hat{\eta}(x)] - \eta(x)\}^2\, dx, \tag{1.5}$$

and

$$IV = \frac{n\Omega}{\sigma^2}\int_S E\{\hat{\eta}(x) - E[\hat{\eta}(x)]\}^2\, dx. \tag{1.6}$$

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.

Subsequent to Box and Draper (1959), various authors have investigated realistic designs for linear models, in which case (1.1) is taken as an approximation of the true model, a more precise description being

$$E(y|x) = q^T(x)\beta_0 + f(x), \tag{1.7}$$

for some unknown but "small" function f belonging to some class $\mathcal{F}$. Robust minimax designs were constructed by solving the problem

$$\min_{\xi}\max_{f\in\mathcal{F}} \Phi(M(f,\xi)) \tag{1.8}$$

for some loss function $\Phi(\cdot)$, where $M(f,\xi)$ is the MSE of $\hat{\beta}$, and the design ξ is a probability distribution $\{p_i\}_{i=1}^N$ on the design space $\mathcal{S} = \{x_i\}_{i=1}^N$; if $p_i = n_i/n$ for integers $n_i \geq 0$, we say that the design ξ is integer-valued.

To motivate (1.7), suppose that an experimenter fits the linear model

$$E[y|x] = q^T(x)\beta_0 \tag{1.9}$$

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

$$\beta_0 = \arg\min_{\beta} \int_S [E[y|x] - q^T(x)\beta]^2 \, dx \tag{1.10}$$

and set

$$f(x) = E[y|x] - q^T(x)\beta_0. \tag{1.11}$$

Marcus and Sacks (1976), Li and Notz (1982), Pesotchinsky (1982) and Liu and Wiens(1997) have discussed designs in which $f(x)$ belongs to the class

$$\mathcal{F} = \{f : |f(x)| \leq \phi(x), \forall x \in \mathcal{S}\}, \tag{1.12}$$

with various assumptions being made about ϕ. 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.

Huber (1975) takes $f(x)$ from

$$\mathcal{F} = \left\{ f : \int_S q(x)f(x)dx = 0, \ \int_S f^2(x)dx \leq \tau^2 \right\}. \tag{1.13}$$

The radius τ of $\mathcal{F}$ is assumed fixed. The first condition in $\mathcal{F}$ says that f and q are orthogonal, so the parameter β is uniquely defined in model (1.7). To see this suppose that

$$E(y|x) = q^T(x)\beta_1 + f_1(x) \tag{1.14}$$

and that

$$E(y|x) = q^T(x)\beta_2 + f_2(x), \tag{1.15}$$

where $q(x)$ is orthogonal to f_1 and f_2. Thus

$$\int q(x)q^T(x)(\beta_1 - \beta_2)dx + \int q(x)(f_1(x) - f_2(x))dx = 0. \tag{1.16}$$

Using 1st condition in (1.13), we have

$$\int q(x)q^T(x)dx[\beta_1 - \beta_2] = 0. \tag{1.17}$$

It follows that $\beta_1 = \beta_2$ provided that $\int q(x)q^T(x)dx$ is positive definite. The second condition assumes that f is small otherwise the model (1.7) will be totally wrong.

Some controversy revolves around the choice of the class $\mathcal{F}$. Marcus and Sacks (1976) and Li and Notz (1982) have criticized the class (1.13) as being too full. Wiens (1992) reported that the class (1.12) is thin, since it seems invariably to lead to 'robust' designs, all of whose mass is concentrated at a small number of, generally extreme, points in the design space. Wiens (1992) reported that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighborhood which is unrealistically sparse.

1.3 Some background on wavelets

In this section we introduce some definitions and theories on wavelets relevant to our work. More detailed discussions can be found in Mallat (1989), Meyer (1992), Daubechies (1992) and Härdle et al. (1998).

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

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \tag{1.18}$$

and

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z} \tag{1.19}$$

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

$$\phi(x) = \sum_{p\in\mathbb{Z}} h_p \phi(2x - p), \tag{1.20}$$

$$\psi(x) = \sum_{r\in\mathbb{Z}} g_r \phi(2x - r), \tag{1.21}$$

and

$$g_r = (-1)^r h_{-r+1} \tag{1.22}$$

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

$$\int \phi(x)dx = 1,\ \int \psi(x)dx = 0,\ \int \phi^2(x)dx = 1. \tag{1.23}$$

The condition

$$\sum_{p\in\mathbb{Z}} h_p = 2 \tag{1.24}$$

ensures the existence of a unique solution to equations (1.20) and (1.21). Orthogonality of the translates of $\phi(x)$ is ensured by the condition

$$\sum_{p\in\mathbb{Z}} h_p h_{p-2j} = \delta_{0j} = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } j \neq 0. \end{cases} \tag{1.25}$$

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

$$\ldots\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots\ldots. \tag{1.26}$$

The nested spaces have an intersection that is trivial and a union that is dense in $\mathcal{L}_2(\mathbb{R})$,

$$\bigcap_j V_j = \{\mathbf{0}\},\ \overline{\bigcup_j V_j} = \mathcal{L}_2(\mathbb{R}). \tag{1.27}$$

Mallat (1989) introduced the notion of a multiresolution analysis, the definition of which we recall here.

Definition 1.3.1 *A multiresolution analysis of* $\mathcal{L}_2(\mathbb{R})$ *consists of an increasing sequence of closed subspaces* V_j, $j \in \mathbb{Z}$ *such that*

(a) $\cap V_j = \{0\}$;

(b) $\overline{\cup V_j} = \mathcal{L}_2(\mathbb{R})$;

(c) there exists a scaling function $\phi \in V_0$ *such that* $\{\phi(x-k), k \in \mathbb{Z}\}$ *is an orthonormal basis of* V_0;

and for all $f \in \mathcal{L}^2(\mathbb{R})$,

(d) for all $k \in \mathbb{Z}$, $f(x) \in V_j \iff f(x-k) \in V_j$, *and*

(e) $f(x) \in V_j \iff f(2x) \in V_{j+1}$.

The intuitive meaning of (e) is that in passing from V_j to V_{j+1}, the resolution of the approximation is doubled. Mallat (1989) has shown that given any multiresolution analysis, it is possible to derive a function $\psi(x)$ such that the family $\{\psi_{j,k}(x) : j, k \in \mathbb{Z}\}$ is an orthonormal basis of $\mathcal{L}_2(\mathbb{R})$.

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

$$W_j \oplus V_j = V_{j+1},\ W_j \perp V_j. \tag{1.28}$$

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

$$V_j = V_0 \oplus \bigoplus_{i=0}^{j-1} W_i. \tag{1.29}$$

Iterating this infinitely many times ,we find

$$\mathcal{L}_2(\mathbb{R}) = \bigcup_{j=0}^{\infty} V_j = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j. \tag{1.30}$$

This means that any $f \in \mathcal{L}_2(\mathbb{R})$ can be represented as a series(convergent in $\mathcal{L}_2(\mathbb{R})$):

$$f(x) = \sum_{k\in\mathbb{Z}} d_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}\psi_{jk}(x), \tag{1.31}$$

where $d_{j_0 k}, c_{jk}$ are some coefficients, and $\{\psi_{jk}\}, k \in \mathbb{Z}$ is a basis for W_j. The relation (1.31) is called a multiresolution expansion of f. The space W_j is called resolution level of multiresolution analysis. In Fourier analysis there is only one resolution level. In multiresolution analysis there are many resolution levels which is the origin of its name.

1.3.1 Wavelet system construction

The general framework of wavelet system construction is as follows.

1. Pick a scaling function ϕ such that $\{\phi_{0k}\}$ is an orthonormal system, and (1.27) is satisfied.

2. Find a primary function $\psi \in W_0$ such that $\{\psi_{0k}, k \in \mathbb{Z}\} = \{\psi(x-k)\}, k \in \mathbb{Z}$, is an orthonormal basis in W_0. Then, consequently, $\{\psi_{jk}, k \in \mathbb{Z}\}$ is orthonormal basis in W_j.

3. Conclude that any $f \in \mathcal{L}_2(\mathbb{R})$ has the unique representation in terms of an $\mathcal{L}_2$-convergent series:

$$f(x) = \sum_{k\in\mathbb{Z}} d_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}\psi_{jk}(x),$$

where the wavelet coefficients are

$$d_{j_0k} = \int f(x)\overline{\phi_{j_0k}(x)}dx,\ c_{jk} = \int f(x)\overline{\psi_{jk}(x)}dx. \tag{1.32}$$

We now outline four constructions of the "scaling function" ϕ found in the literature (see Strang (1989) and Pinheiro and Vidakovic (1997)). Once $\phi(x)$ is known, we can compute the primary wavelet ψ through (1.21).

CONSTRUCTION 1. Iterate $\phi_j(x) = \sum h_k\phi_{j-1}(2x-k)$ with the box function as $\phi_0(x)$. When $h_0 = 2$ the boxes get taller and thinner, approximating the delta function. For $h_0 = h_1 = 1$, the box is invariant: $\phi_1 = \phi_0$. For $\frac{1}{2}$, 1, $\frac{1}{2}$, the hat function appears. And $\frac{1}{8}$, $\frac{4}{8}$, $\frac{6}{8}$, $\frac{4}{8}$, $\frac{1}{8}$ yields the cubic B-spline. An example that will be important in our discussion has coefficients $\frac{1}{4}(1+\sqrt{3})$, $\frac{1}{4}(3+\sqrt{3})$, $\frac{1}{4}(3-\sqrt{3})$, and $\frac{1}{4}(1-\sqrt{3})$. This scaling function leads to orthogonal wavelets.

CONSTRUCTION 2. The second construction takes the Fourier transform of (1.20):

$$\begin{aligned}\hat{\phi}(\xi) &= \sum h_k \int \phi(2x-k)e^{i\xi x}\,dx \\ &= \frac{1}{2}\left(\sum h_k e^{ik\xi/2}\right)\int \phi(y)e^{iy\xi/2}\,dy \\ &= P\left(\frac{\xi}{2}\right)\hat{\phi}\left(\frac{\xi}{2}\right).\end{aligned} \tag{1.33}$$

The symbol $P(\xi) = \frac{1}{2}\sum h_k e^{ik\xi}$ is the crucial function in this theory. With $\xi = 0$ we find $P(0) = 1$ (see (1.24)). Now repeat (1.33) at $\xi/2$, $\xi/4, \ldots$ and recall $\hat{\phi}(0) = \int \phi(x)\, dx = 1$, we get an infinite product:

$$\hat{\phi}(\xi) = P\left(\frac{\xi}{2}\right)\ \hat{\phi}\left(\frac{\xi}{2}\right) = P\left(\frac{\xi}{2}\right)\ P\left(\frac{\xi}{4}\right)\ \hat{\phi}\left(\frac{\xi}{4}\right) = \cdots = \prod_{j=1}^{\infty} P\left(\frac{\xi}{2^j}\right). \quad (1.34)$$

For h_0=2 we find $P \equiv 1$ and $\hat{\phi} \equiv 1$, the transform of the delta function. For $h_0 = h_1 = 1$, the products of the P's are geometric series:

$$P\left(\frac{\xi}{2}\right) = P\left(\frac{\xi}{4}\right) = \frac{1}{4}(1 + e^{i\xi/2})(1 + e^{i\xi/4}) = \frac{1 - e^{i\xi}}{4(1 - e^{i\xi/4})}. \quad (1.35)$$

As $N \to \infty$ this approaches the infinite product $(1 - e^{i\xi})(-i\xi)$. This is $\int_0^1 e^{i\xi x}\, dx$, the transform of the box function. The hat function comes from squaring $P(\xi)$ which by (1.34) also squares $\hat{\phi}(\xi)$. The cubic B-spline comes from squaring again.

CONSTRUCTION 3. This construcion of ϕ works directly with the recursion (1.20). Suppose ϕ is known at the integer $x = j$. The recursion (1.20) gives ϕ at the half-integers. Then it gives ϕ at the quarter-integers, and ultimately at all dyadic point $x = k/2^j$. This is fast to program.

The values of ϕ at the integers come from an eigenvector. With the four Daubechies coefficients $h_0 = \frac{1}{4}(1 + \sqrt{3})$, $h_1 = \frac{1}{4}(3 + \sqrt{3})$, $h_2 = \frac{1}{4}(3 - \sqrt{3})$, $h_3 = \frac{1}{4}(1 - \sqrt{3})$, set $x = 1$ and $x = 2$ in the dilation equation (1.20) and use the fact that $\phi = 0$ unless $0 < x < 3$, we get:

$$\phi(1) = \frac{1}{4}(3 + \sqrt{3})\phi(1) + \frac{1}{4}(1 + \sqrt{3})\phi(2), \quad (1.36)$$

$$\phi(2) = \frac{1}{4}(1 - \sqrt{3})\phi(1) + \frac{1}{4}(3 - \sqrt{3})\phi(2). \quad (1.37)$$

This is the eigenvalue problem $\phi = L\phi$, with matrix entries $L_{ij} = h_{2i-j}$. The eigenvalues are 1 and $\frac{1}{2}$, and the corresponding eigenvector for $\lambda = 1$ has components

$\phi(1) = \frac{1}{2}(1+\sqrt{3})$, $\phi(2) = \frac{1}{2}(1-\sqrt{3})$, which are the heights on our graph of *Daub*2. The other eigenvalue $\lambda = \frac{1}{2}$ means that the recursion can be differentiated: $\phi'(x) = \sum h_k 2\phi'(2x-k)$ leads similarly to $\phi'(1)$ and $\phi'(2)$. In some weak sense, $\phi = D_4$ has a "dilative derivative." For the hat function, the recursion matrix again has $\lambda = 1,\ \frac{1}{2}$. From the cubic spline the eigenvalues are $1,\ \frac{1}{2},\ \frac{1}{4},\ \frac{1}{8}$.

When $\phi(1)$ and $\phi(2)$ is known, the dilation equation gives ϕ at half-integers, such as

$$\begin{aligned} \phi\left(\frac{1}{2}\right) &= \frac{1}{4}(1+\sqrt{3})\phi(1) = \frac{1}{4}(2+\sqrt{3}), \\ \phi\left(\frac{3}{2}\right) &= \frac{1}{4}(3+\sqrt{3})\phi(2) + \frac{1}{4}(3-\sqrt{3})\phi(1) = 0. \end{aligned}$$

Then the equation gives ϕ at quarter-integers as combinations of ϕ at half-integers.

CONSTRUCTION 4. The fourth construction is based on the Daubechies-Lagarias local pyramidal algorithm (see Daubechies and Lagarias (1991, 1992)). The Daubechies-Lagarias algorithm enables us to evaluate ϕ and ψ at a point with pre-assigned precision. We will illustrate the algorithm on wavelets from the Daubechies family; however, the algorithm works for all finite impulse response quadrature mirror filters.

Let ϕ be the scaling function of the D_N wavelet with support $[0, 2N-1]$. Let $x \in (0,1)$, and define $dyad(x) = \{d_1,\ d_2,\ \ldots,\ d_n,\ \ldots\}$ as the set of $0-1$ digits in the dyadic representation of x. That is $x = \sum_{j=1}^{\infty} d_j 2^{-j}$. By $dyad(x,n)$, we denote the subset of the first n digits from $dyad(x)$, i.e., $dyad(x,\ n) = \{d_1,\ d_2,\ \ldots,\ d_n\}$. Let $\mathbf{h} = (h_0, h_1, \ldots, h_{2N-1})$ be the wavelet filter coefficients. Define two $(2N-1) \times (2N-1)$ matrices as:

$$T_0 = (h_{2i-j-1})_{1 \le i,j \le 2N-1} \text{ and } T_1 = (h_{2i-j})_{1 \le i,j \le 2N-1}. \tag{1.38}$$

Then the local pyramidal algorithm can be constructed based on Theorem 1.3.1 (see Daubechies and Lagarias (1992) or Pinheiro and Vidakovic (1997)).

Theorem 1.3.1

$$\lim_{n\to\infty} T_{d_1} \cdot T_{d_2} \cdot \cdots \cdot T_{d_n}$$

$$= \begin{bmatrix} \phi(x) & \phi(x) & \cdots & \phi(x) \\ \phi(x+1) & \phi(x+1) & \cdots & \phi(x+1) \\ \vdots & & & \\ \phi(x+2N-2) & \phi(x+2N-2) & \cdots & \phi(x+2N-2) \end{bmatrix}. \quad (1.39)$$

The convergence of $\|T_{d_1} \cdot T_{d_2} \cdot \ldots \cdot T_{d_n} - T_{d_1} \cdot T_{d_2} \cdot \ldots \cdot T_{d_{n+m}}\|$ to zero, for fixed m, is exponential and constructive, i.e., effective decreasing bounds on the error can be established.

Example 1.3.1 *Again consider the Daub2 scaling function. The corresponding filter is* $\mathbf{h} = (\frac{1+\sqrt{3}}{4}, \frac{3+\sqrt{3}}{4}, \frac{3-\sqrt{3}}{4}, \frac{1-\sqrt{3}}{4})$. *According to (1.38) the matrices* T_0 *and* T_1 *are given as*

$$T_0 = \begin{bmatrix} \frac{1+\sqrt{3}}{4} & 0 & 0 \\ \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} \\ 0 & \frac{1-\sqrt{2}}{4} & \frac{3-\sqrt{3}}{4} \end{bmatrix} \text{ and } T_1 = \begin{bmatrix} \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} & 0 \\ \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} \\ 0 & 0 & \frac{1-\sqrt{3}}{4} \end{bmatrix}.$$

Let us evaluate the scaling function at an arbitrary point, for instance, $x = 0.45$. *Twenty "decimals" in the dyadic representation of* 0.45 *obtained through an s-plus code are* $dyad(0.45, 20) = \{0,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1\}$. *In addition to the value at 0.45, we get the values at* 1.45 *and* 2.45. *The values* $\phi(0.45)$, $\phi(1.45)$

and $\phi(2.45)$ *may be approximated as averages of the first, second, and third row, respectively in the matrix*

$$\prod_{i \in dyad(0.45,20)} T_i = \begin{bmatrix} 0.86480582 & 0.86480459 & 0.86480336 \\ 0.08641418 & 0.08641568 & 0.08641719 \\ 0.04878000 & 0.04877973 & 0.04877945 \end{bmatrix}.$$

The Daubechies-Lagarias algorithm gives only the values of the scaling function. The following theorem gives the values of the wavelet function.

Theorem 1.3.2 *Let* x *be an arbitrary real number. And let the wavelet be given by its filter coefficients* $\{h_0, h_1, \ldots, h_{2N-1}\}$. *Define vector* $\mathbf{u}$ *with* $2N-1$ *components as*

$$\mathbf{u}(x) = \{(-1)^{1-[2x]} h_{i+1-[2x]},\ i = 0, \ldots, 2N-2\}. \tag{1.40}$$

If for some i *the index* $i + 1 - [2x]$ *is negative or larger than* $2N-1$, *then the corresponding components of* $\mathbf{u}$ *is equal to 0.*

Let the vector $\mathbf{v}$ *be defined as*

$$\mathbf{v}(x,n) = \frac{1}{2N-1} \mathbf{1}' \prod_{i \in dyad(\{2x\},n)} T_i, \tag{1.41}$$

where $\mathbf{1}' = (1, 1, \ldots, 1)$ *is the row-vector of ones. Then,*

$$\psi(x) = \lim_{n \to \infty} \mathbf{u}(x)' \mathbf{v}(x,n), \tag{1.42}$$

and the limit is constructive.

Computationally, Construction 4 is the easiest to implement. Thus, this construction has been used in this thesis to construct the Daubechies wavelet systems in Figure (1.2).

1.3.2 Some important wavelet bases

In this section we describe some commonly used families of wavelets: Haar's, multiwavelet and the Daubechies wavelet system.

Haar System: The Haar wavelet basis is the simplest example of a wavelet system on $\mathcal{L}^2(S)$. The scaling function is:

$$\phi(x) = I_{[0,1)}(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.43}$$

The refining relations for the Haar wavelet basis are

$$\phi(x) = \phi(2x-1) + \phi(2x) \tag{1.44}$$

and

$$\psi(x) = \phi(2x) - \phi(2x-1). \tag{1.45}$$

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

$$\phi_i(x) = \begin{cases} \sqrt{2i+1} P_i(2x-1), & x \in [0,1), \\ 0, & \text{otherwise.} \end{cases} \tag{1.46}$$

where P_i $(i = 0, 1, \ldots, N-1)$, are the Legendre polynomials. The space V_n, $n \in \mathbb{Z}$ are dilates of V_0 and the difference spaces W_n are as defined previously. The primary

wavelets denoted by ${}_N\omega_0, \ldots, {}_N\omega_{N-1}$ vanish outside $[0,1)$ and are orthogonal to polynomials of maximum degree,

$$\int_S {}_N\omega_j(x) x^i dx = 0, \; i = 0, 1, \ldots, N-1+j. \tag{1.47}$$

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). For $N = 2$ the scaling functions and primary wavelets are

$$\phi_0(x) = \begin{cases} 1, & \text{if } 0 \le x < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.48}$$

$$\phi_1(x) = \begin{cases} \sqrt{3}(2x-1), & 0 \le x < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.49}$$

$${}_2\omega_0(x) = \begin{cases} \sqrt{3}(1-4x), & 0 \le x < \frac{1}{2} \\ \sqrt{3}(4x-3), & \frac{1}{2} \le x < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{1.50}$$

$${}_2\omega_1(x) = \begin{cases} 6x-1, & 0 \le x < \frac{1}{2} \\ 6x-5, & \frac{1}{2} \le x < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{1.51}$$

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

$$\phi_0(x) = \phi_0(2x) + \phi_0(2x-1) \tag{1.52}$$

$$\phi_1(x) = \frac{\sqrt{3}}{2}(\phi_0(2x-1) - \phi_0(2x)) + \frac{1}{2}(\phi_1(2x-1) + \phi_1(2x)) \tag{1.53}$$

$$_2\omega_0(x) = \phi_1(2x-1) - \phi_1(2x) \tag{1.54}$$

$$_2\omega_1(x) = \frac{1}{2}(\phi_0(2x) - \phi_0(2x-1)) + \frac{\sqrt{3}}{2}(\phi_1(2x-1) + \phi_1(2x)). \tag{1.55}$$

The graphs of the scaling function and primary wavelets are shown in Figure (1.1).

Daubechies System Daubechies was the first to construct compactly supported orthogonal wavelets with a preassigned degree of smoothness. The scaling functions and primary wavelets of the Daubechies (1992) wavelet systems, commonly represented as $_N\phi(x)$ and $_N\psi(x)$ respectively, have no closed forms. They are constructed numerically for different values of the wavelet number N. The algorithm we have used in this thesis is the Construction 4 we introduced in Section 1.3.1. Table 1.1 list the filter coefficients $_Nh_n$ for $N = 2$ through 10. Figure 1.2 shows the plots of the corresponding $_N\phi$, $_N\psi$ for $N = 2,\ 3,\ 4,\ 5$ and 7. Both $_N\phi$ and $_N\psi$ have support width $2N - 1$.

An important feature of the Daubechies wavelets in Figure 1.2 is their smoothness whereas in Figure 1.1 the multiwavelets have cusps and jumps. The choice of a wavelet system to be used will therefore depend on whether the experimenter expects the response to be a smooth function, contain discontinuities, or be a step function.

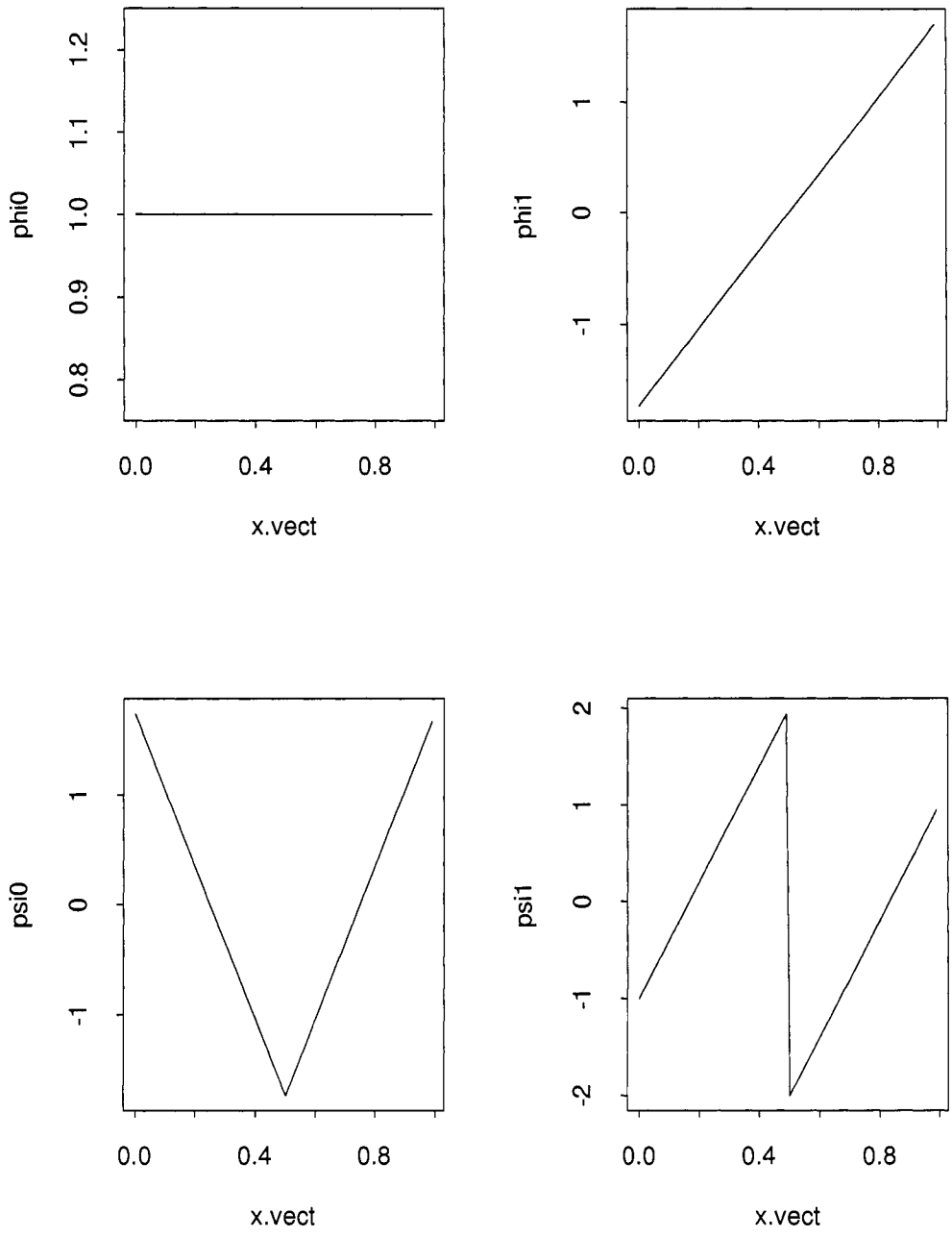

Figure 1.1: Plots of the scaling functions ϕ and primary wavelets ψ for the $N = 2$ Multiwavelets.

Table 1.1: The filter coefficients

	n	$_N h_n$
N=2	0	0.4829629131445341
	1	0.8365163037378077
	2	0.2241438680420134
	3	-.1294095225512603
N=3	0	0.3326705529500825
	1	0.8068915093110924
	2	0.4598775021184914
	3	-.1350110200102546
	4	-.0854412738820267
	5	0.0352262918857095
N=4	0	0.2303778133088964
	1	0.8068915093110924
	2	0.6308807479398587
	3	-.0279837694168599
	4	-.1870348118190931
	5	0.0308613818355607
	6	0.0328830116668852
	7	-.0105974017850690
N=5	0	.1601023979741929
	1	0.6038292697971895
	2	0.7243085284377726
	3	0.1384281459013203
	4	-.2422948870663823
	5	0.0322448695846381
	6	0.0775714938400459
	7	-.0062414902127983
	8	-.0125807519990820
	9	0.0033357252854738
N=6	0	0.1115407433501095
	1	0.4946238903984533
	2	0.7511339080210959
	3	0.3152503517091982
	4	-.2262646939654400
	5	0.1297668685672625
	6	0.0975016055873225
	7	0.0275228655303053
	8	-.0315820393174862
	9	0.0055538422011614
	10	0.0047772575109455
	11	-.0010773010853085
N=7	0	0.0778520540850037
	1	0.3965393194818912
	2	0.7291320908461957
	3	0.4697822874051889
	4	-.1439060039285212
	5	-.2240361849938412
	6	0.0713092192668272
	7	0.0806126091510774
	8	-.0380299369350104
	9	-.0165745416306655
	10	0.0125509985560986
	11	0.0044295779729214
	12	-.0018016407040473
	13	0.0003537137999745

	n	$_N h_n$
N=8	0	0.0544158422441072
	1	0.3128715909143166
	2	0.6756307362973195
	3	0.5853546836542159
	4	-.0158291052563823
	5	-.2840155429615824
	6	0.0004724845739124
	7	0.1287474266204893
	8	-.0173693010018090
	9	-.0440882539307871
	10	0.0139810279174001
	11	0.0087460940474065
	12	-.0048703529934520
	13	-.0003917403733770
	14	0.0006754494064506
	15	-.0001174767841248
N=9	0	0.0380779473638778
	1	0.2438346746125858
	2	0.6048231236900955
	3	0.6572880780512736
	4	0.1331983858249883
	5	-.2932737832791663
	6	-.0968407832229492
	7	0.1485407493381256
	8	0.0307256814793385
	9	-.0676328290613279
	10	0.0002509471148340
	11	0.0223616621236798
	12	-.0047232047577518
	13	-.0042815036824635
	14	0.0018476468830563
	15	0.0002303857635232
	16	-.0002519631889427
	17	0.0000393173203163
N=10	0	0.0266700579005473
	1	0.1881768000776347
	2	0.5272011889315757
	3	0.6884590394534363
	4	0.2811723436605715
	5	-.2498464243271598
	6	-.1959462743772862
	7	0.1273693403357541
	8	0.0930573646035547
	9	-.0713941471663501
	10	-.0294575368218399
	11	0.0332126740593612
	12	0.0036065535669870
	13	-.0107331754833007
	14	0.0013953517470688
	15	0.0019924052951925
	16	-.0006858566979564
	17	-.0001164668551285
	18	0.0000935886703202
	19	-.0000132642028945

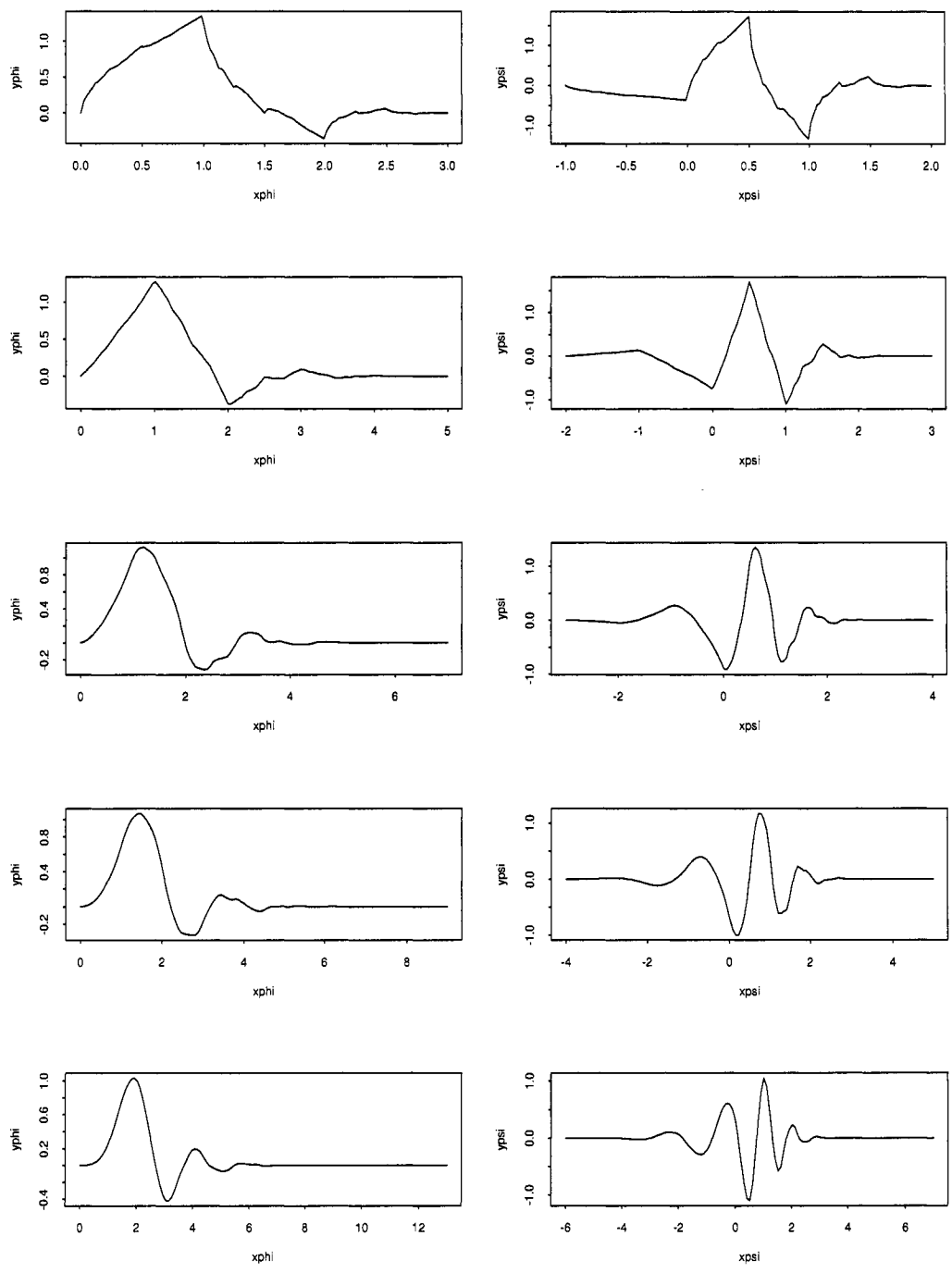

Figure 1.2: Plots of the scaling functions ${}_N\phi$ and wavelets ${}_N\psi$ for the Daubechies wavelets for N=2, 3, 4, 5, 7.

Chapter 2

NON-SEQUENTIAL WAVELET DESIGNS FOR MODEL DISCRIMINATION

2.1 Introduction

In this section we discuss optimum designs for discrimination between two competing models. We assume that the function $\eta_t(x)$ is one of two known mean response functions $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$ neither of which is, in general, a special case of the other. The optimum design for discrimination between two models will depend upon which model is true and, often, on the values of the parameters of the true model. Without loss of generality we suppose that the first model is true and write

$$\eta_t(x) = \eta_1(x, \theta_1). \tag{2.1}$$

A good design for discrimination between the models will then provide a large lack-of-fit sum of squares for the second model. When the second model is fitted to the data, the least squares parameter estimates will depend on the experimental

design as well as on both the value of θ_1 and the errors. In the absence of error the parameter estimates are

$$\hat{\theta}_2(\xi) = \min_{\theta_2} \int_{\mathcal{X}} \{\eta_t(x) - \eta_2(x, \theta_2)\}^2 \xi\,(dx), \tag{2.2}$$

yielding a residual sum of squares

$$\Delta_2(\xi) = \int_{\mathcal{X}} [\eta_t(x) - \eta_2\{x, \hat{\theta}_2(\xi)\}]^2 \xi\,(dx). \tag{2.3}$$

For linear models, $N\Delta_2(\xi)/\sigma^2$ is the non-centrality parameter of the χ^2 distribution of the residual sum of squares for the second model. Designs which maximize $\Delta_2(\xi)$ are called T-optimum, to emphasize the connection with testing for discriminating between models; the letters D and M have already been used , as we have seen, for other criteria (Kiefer 1959). The T-optimum design, obtained by maximizing (2.3), provides the most powerful F test for lack of fit of the second model when the first is true. If the models are non-linear in the parameters, the exact F test is replaced by asymptotic results, but we still design to maximize (2.3).

For linear models with extended design matrices X_1 and X_2 and parameter vectors θ_1 and θ_2 the least squares estimates $\hat{\theta}_2$ minimizing (2.2) are

$$\hat{\theta}_2 = (X_2^T X_2)^{-1} X_2^T X_1 \theta_1. \tag{2.4}$$

Provided that the two models do not contain any terms in common, the non-centrality parameter (2.3) for this exact design is

$$\frac{N\Delta_2(\xi_N)}{\sigma^2} = \frac{N}{\sigma^2} \theta_1^T \{X_1^T X_1 - X_1^T X_2 (X_2^T X_2)^{-1} X_2^T X_1\} \theta_1 \tag{2.5}$$

which makes explicit the dependence of $\Delta_2(\xi_N)$ on the parameters θ_1 of the true model, unless θ_1 is a scalar. In that case designs maximizing (2.5) minimize the variance of estimation of θ_1 in the combined model $E(Y) = X_1\theta_1 + X_2\theta_2$, a criterion which does not depend on the value of θ_1. If θ_1 is a vector, but the two models contain terms in common, θ_1 is reduced by the omission of the common terms. More detailed discussion of these topics is given in Section 2.3.

2.2 General Theory

We continue our discussion by considering the following model:

$$y_{ij} = \eta(x_i) + \varepsilon_{ij},\ i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, n_i, \tag{2.6}$$

where x_i is the ith design point of the explanatory variable x chosen from some design space $\mathcal{S}_* \subseteq \mathcal{R}$; $\eta(x_i) \in \mathcal{R}$ is the value of some nonlinear mean response function at the design point x_i; n_i is the number of observations at x_i and $n = \sum_{i=1}^{N} n_i$ is the total number of observations. And we assume that the error terms ε_{ij} are independent and identically distributed random variables with zero mean and constant variance $\sigma^2 > 0$.

Let $\mathcal{S}_* = [0, 1]$ and $\eta(x) \in \mathcal{L}_2(\mathcal{S}_*)$. The multiresolution analysis of $\mathcal{L}(\mathcal{S}_*)$, discussed in Section 1.3, leads to a wavelet representations of $\eta(x)$ as:

$$\eta(x) = \sum_{k\in\mathbb{Z}} d_k\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}\psi_{jk}(x), \tag{2.7}$$

where the wavelet coefficients are

$$d_k = \int \eta(x)\overline{\phi_{0k}(x)}dx,\ c_{jk} = \int \eta(x)\overline{\psi_{jk}(x)}dx. \tag{2.8}$$

Since in actual computations we can not use infinitely many terms in the representation, we need to decide on the maximum level m at which to terminate the approximation. When m has been determined, we can write (2.7) as:

$$\eta(x) = d_0\phi(x) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} c_{jk}\psi_{jk}(x) + f(x). \tag{2.9}$$

The term $f(x)$ represents components of the wavelet system not used in the approximation, and accounts for the uncertainty in the true structure of the response function. The presence of this term automatically introduces bias in the estimates of the response function $\eta(x)$. In order to control the magnitude of the bias, we impose a bound on $f(x)$. That is,

$$\frac{1}{N}\sum_{i=1}^{N} f^2(x_i) \leq \tau^2, \tag{2.10}$$

for a known constant τ. Define the $2^{m+1} \times 1$ dimensional vectors

$$\mathbf{q}_m(x) = (\phi(x), \psi_{00}(x), \psi_{10}(x), \psi_{11}(x), \ldots, \psi_{m,2^m-1})^T, \tag{2.11}$$

$$\boldsymbol{\beta}_m = (d_0, c_{00}, c_{10}, c_{11}, \ldots, c_{m,2^m-1})^T. \tag{2.12}$$

Then (2.6) can be written as

$$y_{ij} = \mathbf{q}_m^T(x_i)\boldsymbol{\beta}_m + f(x_i) + \varepsilon_{ij},\ i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, n_i. \tag{2.13}$$

If the components of the wavelet system used in (2.13) are orthogonal, we have that

$$\int_0^1 \mathbf{q}_m(s)f(s)\,ds = 0. \tag{2.14}$$

This in turn implies that $\boldsymbol{\beta}_m$ in the model (2.13) is identifiable, provided the matrix $\int_0^1 \mathbf{q}_m(s)\mathbf{q}_m^T(s)\,ds$ is invertible. In the following chapters, we have used the discrete analogue of the identifiability condition

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{q}_m(x_i)f(x_i) = 0, \tag{2.15}$$

to obtain an expression for the contamination term $\mathbf{f}(x)$. The minimax designs we construct will be robust against deviations in the class

$$\mathcal{F} = \left\{ f : \frac{1}{N}\sum_{i=1}^{N} \mathbf{q}_m(x_i)f(x_i) = 0,\ \frac{1}{N}\sum_{i=1}^{N} f^2(x_i) \leq \tau^2 \right\}. \tag{2.16}$$

Now, supposing the two competing wavelet models for representing (2.6) are the models with maximum level of approximations $m-1$ and m. Then, we embed the $(m-1)$th order model in the mth order model and write

$$y_{ij} = \mathbf{q}_{m-1}^T(x_i)\boldsymbol{\beta}_{m-1} + \mathbf{z}^T(x_i)\boldsymbol{\gamma} + f(x_i) + \varepsilon_{ij}, \tag{2.17}$$

where $i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, n_i$, and $\mathbf{q}_{m-1}(x_i)$ and $\mathbf{z}_(x_i)$ are both $2^m \times 1$ dimensional vectors. Clearly, $\mathbf{q}_m(x_i) = (\mathbf{q}_{m-1}^T(x_i),\ \mathbf{z}^T(x_i))^T$.

There are several methods in the literature that can be used in estimating the parameters $\boldsymbol{\beta}_m$. Some methods that have been widely studied include a wavelet version of the Gasser-Müller estimator (GM) (see Antoniadis et.al. (1994)), a modified wavelet version of the Gasser-Müller estimator (MGM) (see Oyet and Sutradhar (2003)), weighted least squares estimator (WLS) (see Oyet and Wiens (2000)), and the nonlinear method of thresholding (see Donoho and Johnstone (1995)). Oyet and

Sutradhar (2003) show that both the MGM and WLS estimators are more efficient than the GM estimator through a simulation study. In this study, we define the parameter $\boldsymbol{\beta}_m$ by

$$\boldsymbol{\beta}_m = arg \min_{\omega} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \varepsilon_{ij}^2(\omega). \tag{2.18}$$

Upon making uncorrelated observations $y_{ij}(j = 1, ..., n_i)$ at $x_i(i = 1, ..., N)$, the experimenter estimates $\boldsymbol{\beta}_m$ by weighted least squares

$$\hat{\boldsymbol{\beta}}_m = \left[\frac{1}{n} \sum_{i=1}^{N} n_i w_i \mathbf{q}_m(x_i) \mathbf{q}_m^T(x_i) \right]^{-1} \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{n_i} w_i \mathbf{q}_m(x_i) y_{ij}, \tag{2.19}$$

where the weights we shall use are those derived by Oyet and Wiens (2000), defined by $w(x_i) = \int_0^1 \|\mathbf{q}_m(s)\| ds / \|\mathbf{q}_m(x_i)\|$. Let $\{p_i = n_i/n\}$ be the integer-valued design on $\mathcal{S}$ and define $m_i = p_i w_i$. Here, $\mathcal{S}$ is a finite dimensional but dense space constructed by partitioning the $[0, 1]$ interval $\mathcal{S}_*$. Then we are seeking a probability distribution $\{m_i\}$ and possible weights $\{w_i\}$ which, subject to the constraint $\sum_{i=1}^{N}(m_i/w_i) = 1$ minimize the maximum, over f, value of the loss function.

Define a vector $\mathbf{f} = (f(x_1), f(x_2), \ldots, f(x_N))^T$, the $N \times N$ matrices $\mathbf{M} = diag(m_1, ..., m_N), \mathbf{W} = diag(w_1, ...w_N)$, the $N \times 2^m$ matrices

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{q}_{m-1}^T(x_1) \\ \mathbf{q}_{m-1}^T(x_2) \\ \cdots \\ \mathbf{q}_{m-1}^T(x_N) \end{pmatrix}, Z = \begin{pmatrix} \mathbf{z}^T(x_1) \\ \mathbf{z}^T(x_2) \\ \cdots \\ \mathbf{z}^T(x_N) \end{pmatrix}, \text{and the } N \times 2^{m+1} \text{ matrix } \mathbf{Q} = \begin{pmatrix} \mathbf{q}_m^T(x_1) \\ \mathbf{q}_m^T(x_2) \\ \cdots \\ \mathbf{q}_m^T(x_N) \end{pmatrix}.$$

Under the model (2.13) we can write expression (2.19) as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_m &= \left[\frac{1}{n}\sum_{i=1}^{N} n_i w_i \mathbf{q}_m(x_i)\mathbf{q}_m^T(x_i)\right]^{-1} \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{n_i} w_i \mathbf{q}_m(x_i)\left(\mathbf{q}_m^T(x_i)\boldsymbol{\beta}_m + f(x_i) + \varepsilon_{ij}\right) \\ &= \boldsymbol{\beta}_m + \mathbf{B}^{-1}\mathbf{b} + \frac{1}{n}\mathbf{B}^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n_i} w_i \mathbf{q}_m(x_i)\varepsilon_{ij}, \qquad (2.20)\end{aligned}$$

where

$$\mathbf{b} = \mathbf{Q}^T\mathbf{M}\mathbf{f} = \sum_{i=1}^{N} m_i \mathbf{q}_m(x_i) f(x_i), \qquad (2.21)$$

$$\mathbf{B} = \mathbf{Q}^T\mathbf{M}\mathbf{Q} = \sum_{i=1}^{N} m_i \mathbf{q}_m(x_i)\mathbf{q}_m^T(x_i). \qquad (2.22)$$

Then the bias vector $\mathbf{d}$ and the covariance matrix $\mathbf{K}$ of $\hat{\boldsymbol{\beta}}_m$ are

$$\mathbf{d} = E(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) = \mathbf{B}^{-1}\mathbf{b} \qquad (2.23)$$

and

$$\mathbf{K} = \frac{\sigma^2}{n}\mathbf{B}^{-1}\sum_{i=1}^{N} m_i w_i \mathbf{q}_m(x_i)\mathbf{q}_m^T(x_i)\mathbf{B}^{-1} = \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}, \qquad (2.24)$$

where

$$\mathbf{D} = \mathbf{Q}^T\mathbf{M}\mathbf{W}\mathbf{Q} = \sum_{i=1}^{N} m_i w_i \mathbf{q}_m(x_i)\mathbf{q}_m^T(x_i). \qquad (2.25)$$

Denoting the mean squared error matrix of $\hat{\boldsymbol{\beta}}_m$ in an mth order model by $M(\mathbf{f}, \xi_N)$, we have

$$M(\mathbf{f}, \xi_N) = \mathbf{B}^{-1}\mathbf{b}\mathbf{b}^{\mathbf{T}}\mathbf{B}^{-1} + \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}, \qquad (2.26)$$

where ξ_N is the N-point design

$$\xi_N = \left\{ \begin{array}{cccc} x_1, & x_2 & \ldots, & x_N \\ p_1, & p_2, & \ldots, & p_N \end{array} \right\}.$$

Similarly, the mean squared error of $\hat{\boldsymbol{\beta}}_{m-1}$ in an $(m-1)$th order model can then be written as

$$M^*(\mathbf{f}^*, \xi_N) = \mathbf{B}^{*-1}\mathbf{b}^*\mathbf{b}^{*\mathbf{T}}\mathbf{B}^{*-1} + \frac{\sigma^2}{n}\mathbf{B}^{*-1}\mathbf{D}^*\mathbf{B}^{*-1}, \tag{2.27}$$

where $\mathbf{b}^*$, $\mathbf{B}^*$ and $\mathbf{D}^*$ are the equivalents of $\mathbf{b}$, $\mathbf{B}$ and $\mathbf{D}$.

We assume that $\mathbf{Q}^*, \mathbf{Q}$ and $\mathbf{Z}$ are of full rank, and define the singular value decomposition of $\mathbf{Q}$ by

$$\mathbf{Q} = \mathbf{U}_{N\times r}\boldsymbol{\Lambda}_{r\times r}\mathbf{V}^T_{r\times r} \ (r = 2^{m+1}), \tag{2.28}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = I_r$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of singular values $\lambda_i(\mathbf{Q})$ $(i = 1, ..., r)$ of $\mathbf{Q}$. Observe from (2.15) that

$$\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{f} = 0. \tag{2.29}$$

This implies that $\mathbf{f}$ belongs to the orthogonal complement of the column space of $\mathbf{U}$ denoted by $[col(\mathbf{U})]^{\perp}$. Now let $\tilde{\mathbf{U}}_{N\times(N-r)}$ be a matrix whose columns form an orthogonal basis of $[col(\mathbf{U})]^{\perp}$. Then, $\mathbf{U}_* = [\mathbf{U} : \tilde{\mathbf{U}}]$ satisfies

$$\mathbf{U}_*\mathbf{U}_*^{\mathbf{T}} = \mathbf{U}\mathbf{U}^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I}_N. \tag{2.30}$$

And any vector $\mathbf{f}$ satisfying (2.10) and (2.15) is representable as

$$\mathbf{f} = \alpha\tilde{\mathbf{U}}\mathbf{e}, \tag{2.31}$$

where $\|\mathbf{e}\| = 1$ and α is a normalizing constant. The choice of $\alpha = \tau\sqrt{N}$ ensures equality in (2.10). That is,

$$\mathbf{f} = \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{e}, \ \|\mathbf{e}\| = 1. \tag{2.32}$$

2.3 Non-sequential Designs

We recall that our problem is to construct designs for discrimination between two competing wavelet models. One obvious approach to check whether the $(m-1)$th order wavelet model is better than the mth order model is to fit model (2.17), without the term $f(x)$ and then to test the null hypothesis

$$H_0 : \beta_{2^m+1} = \beta_{2^m+2} = \ldots = \beta_{2^{m+1}} = 0. \tag{2.33}$$

Let $\mathbf{O}_{2^m}$ be the $2^m \times 2^m$ zero matrix, $\mathbf{I}_{2^m}$ be the $2^m \times 2^m$ identity matrix and define the $2^m \times 2^{m+1}$ partitioned matrix $\mathbf{C} = (\mathbf{O}_{2^m}, \mathbf{I}_{2^m})$. Then, using matrix notations, the null hypothesis H_0 becomes

$$H_0 : \boldsymbol{\gamma} = \mathbf{C}\boldsymbol{\beta}_m = \mathbf{0}, \tag{2.34}$$

where $\mathbf{0}$ is the $2^m \times 1$ zero vector. By fitting (2.17) without the term $f(x)$, we have assumed that the mth order wavelet approximation is exact, when in fact it is not. Under this assumption, the numerator of the F-test for the hypothesis $\boldsymbol{\gamma} = \mathbf{0}$ is the mean square for regression on $\mathbf{q}_{m-1}^T(x)$ and $\mathbf{z}^T(x)$ adjusted for regression on $\mathbf{q}_{m-1}^T$. The power of the test depends on the non-centrality parameter (see Pukelshein (1993))

$$\delta(\xi_N, \mathbf{f}) = (\mathbf{C}\boldsymbol{\beta}_m)^T[\mathbf{C}\mathbf{K}\mathbf{C}^\mathbf{T}]^{-1}(\mathbf{C}\boldsymbol{\beta}_m). \tag{2.35}$$

Let $\mathbf{A} = [\mathbf{C}\mathbf{K}\mathbf{C}^\mathbf{T}]^{-1}$, which is the inverse of the covariance matrix of the least squares estimate of $\boldsymbol{\gamma}$.

In order to detect departures from the $(m-1)$th model, which is to reject the null hypothesis (2.34), experiments should be designed to make the quantity (2.35) large; see Kiefer (1959,§2) for further discussion on this design criterion. That is

$$\Delta(\xi_N^*) = \sup_{\xi_N} \delta(\xi_N, \mathbf{f}). \tag{2.36}$$

When $\boldsymbol{\gamma}$ is scalar, this is equivalent to estimating $\boldsymbol{\gamma}$ with minimum variance, whatever the value of the parameter. But if $\boldsymbol{\gamma}$ is not scalar, the value of (2.35) depends on the vector of unknown parameters. Several types of design procedure are possible. Atkinson and Fedorov (1975) introduced the T-optimum design, which will depend on which of the two models is true and on the values of the regression parameters. Suppose one of the model is true, the experiment should be designed to yield as large a value as possible of the sum of squares for lack of fit of the second model. Some further results are obtained for the linear model case by combining the Bayesian formulation with a maximin approach.

In the absence of specific knowledge about the departures, Atkinson (1972) introduced two design criteria depending solely on the dispersion matrix $A = [\mathbf{CKC^T}]^{-1}$.

Orthogonal designs. If the nuisance parameter are so scaled that departures of equal importance from $(m-1)$th are represented by equal changes in the values of the elements of $\boldsymbol{\gamma}$, we can follow Wald (1943), who was investigating the properties of balanced designs such as Latin squares, and consider the power of designs on the sphere

$$\boldsymbol{\gamma}^T\boldsymbol{\gamma} = 1. \tag{2.37}$$

The minimum value of (2.35) on this sphere is equal to the minimum eigenvalue of A.

Because of the scaling of the nuisance parameters, each element of γ should be estimated with equal variance. But this condition alone does not guarantee satisfactory power against all alternatives on (2.37). To ensure uniform power on this sphere, the design has in addition to be orthogonal, when A becomes a multiple of the identity matrix. Given a class of these designs we choose the one for which (2.35) is a maximum.

D-optimum designs. The D-optimal design for γ maximizes the determinant of A, i.e. the product of the eigenvalues is a maximum. This criterion has the advantage over orthogonality that it does not depend on the scaling of the factors. It also ensures that the power function of the F test has maximum Gaussian curvature at the null hypothesis among all locally unbiased tests of a given size. Thus we find the D-optimun design for which the generalized variance of the estimates of γ is a minimum.

2.3.1 T-optimum design

The extremum problem (2.36) will depend on which of the two models is true and on the values of the regression parameters. Atkinson and Fedorov (1975) introduced a locally optimum design called T-optimum designs. The T-optimal criterion we have used, is a slight modification of $\delta(\xi_N, \mathbf{f})$ which takes into consideration the fact

that the wavelet model is only an approximation. Instead of using the covariance matrix $\mathbf{K}$ of $\boldsymbol{\beta}_m$, we shall use the mean squared error matrix $M(\xi_N, \mathbf{f})$ to account for the bias in estimation by assuming the approximation is exact. Then our modified T-optimal criteria is

$$\Delta(\xi_N, \mathbf{f}) = (\mathbf{C}\boldsymbol{\beta}_\mathbf{m})^T[\mathbf{C}\mathbf{M}(\xi_\mathbf{N}, \mathbf{f})\mathbf{C}^T]^{-1}(\mathbf{C}\boldsymbol{\beta}_m). \tag{2.38}$$

The dependence of $\Delta(\xi_N, \mathbf{f})$ on $\boldsymbol{\beta}_m$ can be resolved by adopting a Bayesian approach(See Ankinson and Fedorov 1975). If it can be assumed that the experimenter has some prior information about the parameter $\boldsymbol{\beta}_m$ which can be expressed as a prior distribution $p_0(\boldsymbol{\beta}_m)$, it is natural to replace (2.38) by

$$\delta(\xi_N^*, \mathbf{f}) = \max_{\xi_N} \delta(\xi_N, \mathbf{f}), \tag{2.39}$$

where

$$\delta(\xi_N, \mathbf{f}) = \int p_0(\boldsymbol{\beta}_m)\Delta(\xi_N, \mathbf{f})\, d\boldsymbol{\beta}_m. \tag{2.40}$$

Substituting (2.38) into above expression, we have

$$\begin{aligned} \delta(\xi_N, \mathbf{f}) &= \int p_0(\boldsymbol{\beta}_m)(\mathbf{C}\boldsymbol{\beta}_m)^T[\mathbf{C}M(\xi, \mathbf{f})\mathbf{C}^T]^{-1}\mathbf{C}\boldsymbol{\beta}_m\, d\boldsymbol{\beta}_m \\ &= tr\{[\mathbf{C}\mathbf{M}(\xi_\mathbf{N}, \mathbf{f})\mathbf{C}^T]^{-1} \int \mathbf{C}\boldsymbol{\beta}_m(\mathbf{C}\boldsymbol{\beta}_m)^T p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m\}. \end{aligned} \tag{2.41}$$

Now, by setting

$$\beta_0 = \int \boldsymbol{\beta}_m p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m, \tag{2.42}$$

$$D_0 = \int \mathbf{C}(\boldsymbol{\beta}_\mathbf{m} - \beta_\mathbf{0})(\boldsymbol{\beta}_\mathbf{m} - \beta_\mathbf{0})^T\mathbf{C}^T p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m, \tag{2.43}$$

it is easy to show that

$$\begin{aligned}
& \delta(\xi_N, \mathbf{f}) \\
= \; & tr\{[\mathbf{C}M(\xi_N, \mathbf{f})\mathbf{C}^T]^{-1} \int \mathbf{C}[(\boldsymbol{\beta}_m - \beta_0) + \beta_0][(\boldsymbol{\beta}_m - \beta_0) + \beta_0]^T \mathbf{C}^T p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m\} \\
= \; & tr\{[\mathbf{C}M(\xi_N, \mathbf{f})\mathbf{C}^T]^{-1}[D_0 + 2\beta_0^T \mathbf{C}^T \int \mathbf{C}(\boldsymbol{\beta}_m - \beta_0) p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m \\
& + \mathbf{C}\beta_0\beta_0^T \mathbf{C}^T \int p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m]\} \qquad (2.44)
\end{aligned}$$

Observing the fact that

$$\int p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m = 1 \qquad (2.45)$$

and

$$\int \mathbf{C}(\boldsymbol{\beta}_m - \beta_0) p_0(\boldsymbol{\beta}_m)\, d\boldsymbol{\beta}_m = 0, \qquad (2.46)$$

we obtain

$$\delta(\xi_N, \mathbf{f}) = tr\{[\mathbf{C}M(\xi_N, \mathbf{f})\mathbf{C}^T]^{-1}(\mathbf{D}_0 + (\mathbf{C}\beta_0)(\mathbf{C}\beta_0)^T)\}. \qquad (2.47)$$

The new extremum problem (2.47) can be solved by the standard methods of convex design theory.

A difficulty of the Bayesian approach is the presence of the prior distribution $p_0(\boldsymbol{\beta}_m)$ in the criterion. One way of avoiding this problem is to combine the Bayesion formulation with a maximin approach and to solve the extremum problem

$$\gamma(\xi_N^*) = \max_{\xi_N} \gamma(\xi_N) \qquad (2.48)$$

where

$$\gamma(\xi_N) = \min_{p_0(\boldsymbol{\beta}_m)\in\mathcal{P}} \delta(\xi_N, \mathbf{f}) \tag{2.49}$$

and $\mathcal{P}$ is the class of density functions with a given measure of scattering $|\mathbf{D}_0| = d$. First we evaluate $\gamma(\xi_N)$. Since $\delta(\xi_N, \mathbf{f})$ depends on $\mathbf{D}_0$ and β_0, we can replace (2.49) by

$$\gamma(\xi_N) = \min_{\beta_0} \delta(\xi_N, \mathbf{f}), \tag{2.50}$$

where the minimization is subject to $|\mathbf{D}_0| = d$. From the definition of the MSE matrix, we have

$$\begin{aligned} |M(\xi_N, \mathbf{f})| &= |\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1} + \frac{\sigma^2}{n}\mathbf{H}^{-1}| \\ &= |(\frac{n}{\sigma^2}\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{H} + \mathbf{I}) \cdot \frac{\sigma^2}{n}\mathbf{H}^{-1}| \\ &= |\frac{\sigma^2}{n}\mathbf{H}^{-1}|\ |\mathbf{I} + \frac{n}{\sigma^2}\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{B}\mathbf{D}^{-1}\mathbf{B}| \\ &= \left(\frac{\sigma^2}{n}\right)^r \left\{\frac{1 + \frac{n}{\sigma^2}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}}{|\mathbf{H}|}\right\}. \end{aligned} \tag{2.51}$$

The matrix $\mathbf{H}$ as defined is positive definite, so that $M(\xi_N, \mathbf{f})$ is also positive semi-definite. That is $\theta^T M(\xi_N, \mathbf{f})\theta \geq 0$ for all $\theta \neq 0$. Then

$$\begin{aligned} & \min_{\beta_0} \delta_(\xi_N, \mathbf{f}) \\ = & \min_{\beta_0} tr\{[\mathbf{C}\mathbf{M}(\xi_\mathbf{N}, \mathbf{f})\mathbf{C}^T]^{-1}(\mathbf{D_0} + (\mathbf{C}\beta_\mathbf{0})(\mathbf{C}\beta_\mathbf{0})^T)\} \\ = & \min_{\beta_0} tr\{[\mathbf{C}\mathbf{M}(\xi_N, \mathbf{f})\mathbf{C}^T]^{-1}\mathbf{D}_0 + (\mathbf{C}\beta_0)^T[\mathbf{C}\mathbf{M}(\xi_N, \mathbf{f})\mathbf{C}^T]^{-1}(\mathbf{C}\beta_0)\}. \end{aligned} \tag{2.52}$$

Define $\alpha_0 = \mathbf{C}\beta_0$, $\mathbf{M}^* = [\mathbf{C}\mathbf{M}(\xi_\mathbf{N}, \mathbf{f})\mathbf{C}^T]^{-1}$, $\lambda_i(\mathbf{M}^*)$ as the i-th eigenvalue of $\mathbf{M}^*$ and

$\mathbf{e}_i$ as the normalized eigenvectors of $\mathbf{M}^*$, then

$$\begin{aligned}(\mathbf{C}\beta_0)^T[\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T]^{-1}(\mathbf{C}\beta_0) &= \alpha_0^T\sum\lambda_i(\mathbf{M}^*)\mathbf{e}_i\mathbf{e}_i^T\alpha_0 \\ &\geq \lambda_{min}(\mathbf{M}^*)\|\alpha_0\|^2, \end{aligned} \tag{2.53}$$

where $\lambda_{min}(\mathbf{M}^*)$ is the smallest eigenvalue of $\mathbf{M}^*$. Since $\|\alpha_0\|^2 \geq 0$ and $\mathbf{M}^*$ is positive semi-definite, we obtain $\lambda_{min}(\mathbf{M}^*)\|\alpha_0\|^2 \geq 0$. This implies that,

$$\min_{\beta_0}\delta(\xi_N,\mathbf{f}) = tr\{[\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T]^{-1}\mathbf{D}_0\}. \tag{2.54}$$

Using results obtained by Beckenhash & Bellman(1965, pg 70), we have

$$2^{-m}tr\{[\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T]^{-1}\mathbf{D}_0\} \geq |\mathbf{D}_0|^{1/2^m}|\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T|^{-1/2^m}. \tag{2.55}$$

Since the lower limit is to be achieved for some $\mathbf{D}_0$, the minimum problem (2.54) becomes

$$\min_{D_0}\min_{\beta_0}\delta(\xi_N,\mathbf{f}) = d^{1/2^m}|\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T|^{-1/2^m}. \tag{2.56}$$

Thus (2.39) can be replaced by the new extremum problem

$$d^{1/2^m}|\mathbf{CM}(\xi_\mathbf{N}^*,\mathbf{f})\mathbf{C}^T|^{-1/2^m} = \sup_{\xi_N} d^{1/2^m}|\mathbf{CM}(\xi_\mathbf{N},\mathbf{f})\mathbf{C}^T|^{-1/2^m}. \tag{2.57}$$

It is clear from (2.56) that the design ξ_N^* which maximizes $\min_{D_0}\min_{\beta_0}\delta(\xi_N,\mathbf{f})$ is equivalent to the design that minimizes $|\mathbf{CM}(\xi_\mathbf{n},\mathbf{f})\mathbf{C}^\mathbf{T}|$. Since $|\mathbf{CM}(\xi_\mathbf{n},\mathbf{f})\mathbf{C}^\mathbf{T}|$ is a function of the contamination vector $\mathbf{f}$, we adopt the minimax approach by computing the least favorable function $\mathbf{f_0}$ which maximizes $|\mathbf{CM}(\xi_\mathbf{n},\mathbf{f})\mathbf{C}^\mathbf{T}|$ before constructing ξ_N^*. That is, the minimax T-optimal design is a solution to

$$\min_{\xi_N}\max_{\mathbf{f}}|\mathbf{CM}(\xi_\mathbf{n},\mathbf{f})\mathbf{C}^\mathbf{T}|, \tag{2.58}$$

for every $\mathbf{f}$ satisfying (2.10) and (2.15).

From the definition of the $\mathbf{M}(\xi_\mathbf{n}, \mathbf{f})$ matrix, we obtain

$$\begin{aligned} & |\mathbf{C}\mathbf{M}(\xi_\mathbf{n}, \mathbf{f})\mathbf{C}^T| \\ = \; & |\mathbf{C}\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{C}^T + \frac{\sigma^2}{n}\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T| \\ = \; & |\frac{\sigma^2}{n}\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T| \cdot |\mathbf{I} + \frac{n}{\sigma^2}\mathbf{C}\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}| \\ = \; & \left(\frac{\sigma^2}{n}\right)^{2^{m+1}} |\mathcal{H}| \left(1 + \frac{n}{\sigma^2}\mathbf{b}^T\mathcal{D}\mathbf{b}\right), \end{aligned} \tag{2.59}$$

where $\mathcal{H} = \mathbf{C}\mathbf{H}^{-1}\mathbf{C}^\mathbf{T}$ and $\mathcal{D} = \mathbf{B}^{-1}\mathbf{C}^\mathbf{T}(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^\mathbf{T})^{-1}\mathbf{C}\mathbf{B}^{-1}$.

To find the determinant of $\mathbf{C}\mathbf{M}(\xi_\mathbf{n}, \mathbf{f})\mathbf{C}^\mathbf{T}$, first we need to find the determinant of $\mathcal{H}$, which has the form

$$\begin{aligned} |\mathcal{H}| &= |\mathbf{C}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}\mathbf{C}^T| \\ &= |\mathbf{C}(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{W}\mathbf{Q}(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}\mathbf{C}^T|. \end{aligned} \tag{2.60}$$

Using the singular value decomposition (2.28), and defining

$$\mathbf{M}_j = \mathbf{U}^T\mathbf{M}^j\mathbf{U}, \; \mathbf{M}_W = \mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{U}, \tag{2.61}$$

we obtain

$$|\mathcal{H}| = |\mathbf{C}(\mathbf{V}^T)^{-1}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{M}_W\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}\mathbf{C}^T|. \tag{2.62}$$

For simplicity, we define

$$\mathbf{F} = \mathbf{\Lambda}^{-1}\mathbf{V}^{-1}\mathbf{C}^T, \tag{2.63}$$

and use the property that the determinant of a matrix is equal to the product of eigenvalues of this matrix to write

$$|\mathcal{H}| = \prod_{i=1}^{2^m} \lambda_i(\mathbf{F}^T\mathbf{M}_1^{-1}\mathbf{M}_w\mathbf{M}_1^{-1}\mathbf{F}), \tag{2.64}$$

where $\lambda_i(\mathbf{F}^T\mathbf{M}_1^{-1}\mathbf{M}_w\mathbf{M}_1^{-1}\mathbf{F})$ is the ith eigenvalue of $\mathbf{F}^T\mathbf{M}_1^{-1}\mathbf{M}_w\mathbf{M}_1^{-1}\mathbf{F}$.

Next, we look at the term $\mathbf{b}^T\mathcal{D}\mathbf{b}$ in (2.59) which is a function of the contamination vector $\mathbf{f}$. It can be shown that

$$\begin{aligned} \mathbf{b}^T\mathcal{D}\mathbf{b} &= \mathbf{f}^T\mathbf{M}\mathbf{Q}\mathbf{B}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}\mathbf{C}\mathbf{B}^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{f} \\ &= \mathbf{f}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{F}(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}\mathbf{F}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{f}. \end{aligned} \tag{2.65}$$

It is clear that for all the $\mathbf{f}$ satisfying (2.10) and (2.15), the least favorable function $\mathbf{f}_0$ which maximizes $|\mathbf{C}M(\xi_N, \mathbf{f})\mathbf{C}^T|$ is equivalent to the function that maximizes $|\mathbf{b}^T\mathcal{D}\mathbf{b}|$. Using (2.32), we have

$$\max_{\mathbf{f}} \mathbf{b}^T\mathcal{D}\mathbf{b} = \max_{\mathbf{f}} \tau^2 N\mathbf{e}^T\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{F}(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}\mathbf{F}^T\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}\mathbf{e}. \tag{2.66}$$

If for simplicity, we set

$$\mathbf{G} = \tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{F}(\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^T)^{-1}\mathbf{F}^T\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}, \tag{2.67}$$

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

$$\max_{\mathbf{f}} \mathbf{b}^T\mathcal{D}\mathbf{b} = \max_{\mathbf{e}} \tau^2 N\mathbf{e}^T\mathbf{G}\mathbf{e} = \tau^2 N\lambda_{max}(\mathbf{G}), \tag{2.68}$$

where $\lambda_{max}(\mathbf{G})$ is the largest eigenvalue of $\mathbf{G}$. We define the square-root matrix of $(\mathbf{CH}^{-1}\mathbf{C}^T)^{-1}$ as

$$\mathbf{E} = (\mathbf{CH}^{-1}\mathbf{C}^T)^{-\frac{1}{2}}. \tag{2.69}$$

It follows that

$$\lambda_{max}(\mathbf{G}) = \lambda_{max}(\mathbf{EF}^T\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{MUM}_1^{-1}\mathbf{FE}). \tag{2.70}$$

From (2.30), we obtain $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I} - \mathbf{UU}^T$. Substituting $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T$ into (2.70), we have

$$\begin{aligned}\lambda_{max}(\mathbf{G}) &= \lambda_{max}(\mathbf{EF}^T\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{MMUM}_1^{-1}\mathbf{FE} - \mathbf{EF}^T\mathbf{M}_1^{-1}\mathbf{M}_1\mathbf{M}_1\mathbf{M}_1^{-1}\mathbf{FE}) \\ &= \lambda_{max}(\|\mathbf{MUM}_1^{-1}\mathbf{FE}\|^2 - \|\mathbf{FE}\|^2), \end{aligned} \tag{2.71}$$

where $\|\mathbf{A}\|^2 = \mathbf{A}^T\mathbf{A}$.

Now, we summarize the solution to the minimax T-optimal design in the following Theorem.

Theorem 2.3.1 *Let* $\lambda_i(\mathbf{F^T M_1^{-1} M_W F})$ $(i = 1, ..., 2^m)$ *be the eigenvalues of the matrix* $\mathbf{F^T M_1^{-1} M_W F}$ *and let* $\lambda_{max}(\mathbf{G})$ *be the maximum eigenvalue of* $\mathbf{G} = \|\mathbf{MUM}_1^{-1}\mathbf{FE}\|^2 - \|\mathbf{FE}\|^2$ *with corresponding eigenvector* $\mathbf{e}$. *Then, for fixed* $\upsilon = \frac{\sigma^2}{n\tau^2}$,

$$\max_{\mathbf{f}\in\mathcal{F}} |\mathbf{CM}(\xi_\mathbf{N}, \mathbf{f})\mathbf{C}^T| = \left(\frac{\sigma^2}{n}\right)^r \prod_{i=1}^{2^m} \lambda_i(\mathbf{F^T M_1^{-1} M_W F})(1 + \upsilon^{-1} N\lambda_{max}(\mathbf{G})). \tag{2.72}$$

In Section 2.4, we will use the simulated annealing algorithm to determine the N-point minimax robust T-optimal design ξ_N^* which minimizes (2.72) under a Daubechies and a multiwavelet model.

2.3.2 D-ratio Optimal Designs

Some analysts may object to using T-optimal designs due to the dependence of the criterion on the unknown parameter $\boldsymbol{\beta}_m$. The experimenter then has to consider designs constructed under an alternative criterion. A criterion that is commonly used is that which maximizes the accuracy in estimating the vector $\boldsymbol{\gamma}$ in (2.17) such as the D-optimality criterion. The D-optimality criterion is the determinant of the inverse of covariance matrix of $\hat{\boldsymbol{\gamma}}$ or in the robust case is the determinant of the MSE matrix of $\hat{\boldsymbol{\gamma}}$. The criterion we are about to discuss is partly motivated by the fact that, under ordinary least squares, we have that

$$\begin{aligned} |cov(\hat{\boldsymbol{\gamma}})|^{-1} &= |\mathbf{C}\ cov(\hat{\boldsymbol{\beta}}_m)\mathbf{C}^T|^{-1} \\ &= |\mathbf{Z}^T\mathbf{M}\mathbf{Z} - \mathbf{Z}^T\mathbf{M}\mathbf{Q}^*(\mathbf{Q}^{*T}\mathbf{M}\mathbf{Q}^*)^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{Z}| \\ &= \frac{|\mathbf{Q}^T\mathbf{M}\mathbf{Q}|}{|\mathbf{Q}^{*T}\mathbf{M}\mathbf{Q}^*|}. \end{aligned} \tag{2.73}$$

We also observe that

$$cov(\hat{\boldsymbol{\beta}}_m) = \sigma^2(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}, \tag{2.74}$$

and

$$cov(\hat{\boldsymbol{\beta}}_{m-1}) = \sigma^2(\mathbf{Q}^{*T}\mathbf{M}\mathbf{Q}^*)^{-1}. \tag{2.75}$$

Thus, we can rewrite (2.73) as

$$|cov(\hat{\boldsymbol{\gamma}})|^{-1} = \frac{|\mathbf{Q}^T\mathbf{M}\mathbf{Q}|}{|\mathbf{Q}^{*T}\mathbf{M}\mathbf{Q}^*|} = \frac{|cov(\hat{\boldsymbol{\beta}}_m)|^{-1}}{|cov(\hat{\boldsymbol{\beta}}_{m-1})|^{-1}}. \tag{2.76}$$

The designs which maximize $|cov(\hat{\gamma})|$ will also maximize $|cov(\hat{\beta}_m)|/|cov(\hat{\beta}_{m-1})|$. Now, due to the uncertainty in the true structure of the mean response, the experimenter may adopt a design criterion which accounts for this uncertainty by using the mean squared error matrix in place of the covariance matrix. Using the mean squared error matrix, the D-optimality criterion for the full mth order model is $|M(\mathbf{f}, \xi_N)|$, and the criterion for the reduced $(m-1)$th order model is $|M^*(\mathbf{f}^*, \xi_N)|$. We then defined the D-ratio optimality criterion as

$$\mathcal{L}(\xi_N) = \frac{|M(\mathbf{f}, \xi_N)|^{-1/2^{m+1}}}{|M^*(\mathbf{f}^*, \xi_N)|^{-1/2^m}}. \tag{2.77}$$

Then a design ξ_N^* is called a N-point D-ratio optimal design if

$$\Delta_D(\xi_N^*) = \max_{\xi_N} \mathcal{L}(\xi_N). \tag{2.78}$$

We have raised the determinant to the $1/2^{m+1}$ and $1/2^m$ power to adjust for the difference in the number of parameters between the mth order model and the $(m-1)$th order model. A second motivation for this criterion comes from the approach commonly used in constructing likelihood ratio tests for a specified null hypothesis. In that case, the basis for constructing a LRT is the magnitude of the ratio of the likelihood function under the null hypothesis to the likelihood function with no restriction. We have not used the term standardized or D-efficient to refer to the criterion (2.77) because these terms have been used by Dette (1997) and Pukelsheim and Rosenberger (1993) to define similar by not exactly the same criterion.

It is clear from (2.77) that the design ξ_N^* which maximizes $\mathcal{L}(\xi_N)$ is equivalent to

the design that minimizes $|M(\mathbf{f},\xi_N)|^{1/2^{m+1}}/|M^*(\mathbf{f}^*,\xi_N)|^{1/2^m}$. Since $|M(\mathbf{f},\xi_N)|^{1/2^{m+1}}$ and $|M^*(\mathbf{f}^*,\xi_N)|^{1/2^m}$ are functions of the contamination vector $\mathbf{f}$ and $\mathbf{f}^*$, respectively, we again adopt the minimax approach by computing the least favorable functions $\mathbf{f}_0$ and $\mathbf{f}_0^*$ which maximizes $|M(\mathbf{f},\xi_N)|^{1/2^{m+1}}$ and $|M^*(\mathbf{f}^*,\xi_N)|^{1/2^m}$ respectively before constructing ξ_N^*. Then, our minimax D-ratio design is a solution to

$$\min_{\xi_N}\left\{\frac{\max_{\mathbf{f}}|M(\mathbf{f},\xi_N)|^{1/2^{m+1}}}{\max_{\mathbf{f}^*}|M^*(\mathbf{f}^*,\xi_N)|^{1/2^m}}\right\}. \tag{2.79}$$

To solve this minimization problem, we first need to solve the two maximization problems respectively. First we look at the maximization problem for the mth order model. From (2.51) we observe that

$$\max_{\mathbf{f}}|M(\mathbf{f},\xi_N)| = \max_{\mathbf{f}}\left(\frac{\sigma^2}{n}\right)^{2^{m+1}}\cdot\left\{\frac{1+\frac{n}{\sigma^2}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}}{|\mathbf{H}|}\right\}. \tag{2.80}$$

From the definition of the $\mathbf{H}$ matrix, we know the term $|\mathbf{H}|$ in the above expression does not contain the contamination vector $\mathbf{f}$, which means we can skip the maximization step for this term. Furthermore, use the same definition for $\mathbf{M}_1$ and $\mathbf{M}_W$ as for the T-optimal design and substitute the singular value decomposition of $\mathbf{Q}$ into $|\mathbf{H}|$, we obtain

$$\begin{aligned}
|\mathbf{H}| &= |\mathbf{B}\mathbf{D}^{-1}\mathbf{B}| \\
&= |\mathbf{Q}^T\mathbf{M}\mathbf{Q}(\mathbf{Q}^T\mathbf{M}\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{Q}| \\
&= |\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\mathbf{M}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T(\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\mathbf{M}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T| \\
&= |\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T|.
\end{aligned} \tag{2.81}$$

Therefore,

$$|\mathbf{H}| = \prod_{i=1}^{2^{m+1}} \lambda_i(\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T), \tag{2.82}$$

where $\lambda_i(\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T)$ is the ith eigenvalue of $\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T$.

Next, we try to solve the maximization problem for the term $\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}$ subject to the vector $\mathbf{f}$. To do this, we consider

$$\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} = \mathbf{f}^T\mathbf{M}\mathbf{Q}(\mathbf{Q}^T\mathbf{M}\mathbf{W}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{f}. \tag{2.83}$$

Substituting the singular value decomposition of $\mathbf{Q}$ into the the above expression and simplify it, we have that

$$\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} = \mathbf{f}^T\mathbf{M}\mathbf{U}\mathbf{M}_W^{-1}\mathbf{U}^T\mathbf{M}\mathbf{f}. \tag{2.84}$$

Using the fact that

$$\mathbf{f} = \tau\sqrt{N}\tilde{\mathbf{U}}\mathbf{e}, \tag{2.85}$$

we have

$$\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} = \tau^2 N\mathbf{e}^T\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_W^{-1}\mathbf{U}^T\mathbf{M}\tilde{U}\mathbf{e}. \tag{2.86}$$

Thus,

$$\begin{aligned} \max_{\mathbf{f}} \mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} &= \tau^2 N\lambda_{max}(\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_W^{-1}\mathbf{U}^T\mathbf{M}\tilde{U}) \\ &= \tau^2 N\lambda_{max}(\mathbf{M}_W^{-1}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}). \end{aligned} \tag{2.87}$$

Use the fact that

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I} - \mathbf{U}\mathbf{U}^T, \tag{2.88}$$

and define

$$\mathbf{M}_2 = \mathbf{U}^T\mathbf{M}^2\mathbf{U}, \tag{2.89}$$

to obtain

$$\max_{\mathbf{f}} \mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} \;=\; \tau^2 N\lambda_{max}(\mathbf{M}_W^{-1}\mathbf{M}_2 - \mathbf{M}_W^{-1}\mathbf{M}_1^2). \tag{2.90}$$

Thus, the solution to the maximization problem (2.80) for the mth order wavelet model is

$$\max_{\mathbf{f}} |M(\mathbf{f}, \xi_N)| = \left(\frac{\sigma^2}{n}\right)^{2^{m+1}} \cdot \frac{[1 + \frac{nN\tau^2}{\sigma^2}\lambda_{max}(\mathbf{M}_W^{-1}\mathbf{M}_2 - \mathbf{M}_W^{-1}\mathbf{M}_1^2)]}{[\prod_{i=1}^{2^{m+1}} \lambda_i(\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T)]} \tag{2.91}$$

Similarly, we can obtain the solution to the maximization problem for the $(m-1)$th order model as

$$\max_{\mathbf{f}^*} |M(\mathbf{f}^*, \xi_N)| = \left(\frac{\sigma^2}{n}\right)^{2^m} \cdot \frac{[1 + \frac{nN\tau^2}{\sigma^2}\lambda_{max}(\mathbf{M}_W^{*-1}\mathbf{M}_2^* - \mathbf{M}_W^{*-1}\mathbf{M}_1^{*2})]}{[\prod_{i=1}^{2^m} \lambda_i(\mathbf{V}^*\mathbf{\Lambda}^*\mathbf{M}_1^*\mathbf{M}_W^{*-1}\mathbf{M}_1^*\mathbf{\Lambda}^*\mathbf{V}^{*T})]}, \tag{2.92}$$

where $\mathbf{M}_1^*$, $\mathbf{M}_2^*$ and $\mathbf{M}_W^*$ are the equivalents of $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_W$.

Now, we summarize the solution to the D-ratio optimal design in the following theorem.

Theorem 2.3.2 *Let M_l, $l = 1, 2$ and M_W be as defined before and $\mathcal{L}(\xi_N)$ be defined by (2.77). Then, for fixed $\nu = \frac{\sigma^2}{n\tau^2}$ and $\nu^* = \frac{\sigma^2}{n\tau^{*2}}$, an N-point D-ratio optimal design is any design ξ_N which minimizes*

$$\begin{aligned}\mathcal{L}(\xi_N) \;=\; & \frac{[1 + \frac{N}{\nu}\lambda_{max}(\mathbf{M}_W^{-1}\mathbf{M}_2 - \mathbf{M}_W^{-1}\mathbf{M}_1^2)]^{1/2^{m+1}}}{[1 + \frac{N}{\nu^*}\lambda_{max}(\mathbf{M}_W^{*-1}\mathbf{M}_2^* - \mathbf{M}_W^{*-1}\mathbf{M}_1^{*2})]^{1/2^m}} \\ & \cdot \frac{[\prod_{i=1}^{2^m} \lambda_i(\mathbf{V}^*\mathbf{\Lambda}^*\mathbf{M}_1^*\mathbf{M}_W^{*-1}\mathbf{M}_1^*\mathbf{\Lambda}^*\mathbf{V}^{*T})]^{1/2^m}}{[\prod_{i=1}^{2^{m+1}} \lambda_i(\mathbf{V}\mathbf{\Lambda}\mathbf{M}_1\mathbf{M}_W^{-1}\mathbf{M}_1\mathbf{\Lambda}\mathbf{V}^T)]^{1/2^{m+1}}}.\end{aligned} \tag{2.93}$$

We will discuss examples of these designs in Section 2.4 using the simulated annealing algorithm.

2.4 Minimax wavelet designs

In this section, we use the simulated annealing algorithm to determine the distribution $\mathbf{p} = (p_1, \cdots, p_N)$ which minimizes the maximum loss in Theorems 1 and 2.

The name of the simulated annealing algorithm originates from the analogy with the cooling process in thermodynamics, specifically to the way that metals, and some liquids, cool and crystalize. The cooling process is called annealing if the temperature is lowered slowly. A characteristic property of annealing is lowering the temperature gradually, allowing thermal equilibrium to be attained. At high temperatures the molecules move about freely, but when the temperature decreases they gradually lose mobility and form a pure crystal, which is, in fact, a state of global minimum energy. And , as long as the temperature is decreased slowly, nature is almost certain to find it. If, however, the temperature is decreased rather more rapidly, a local minimum energy state with higher energy may be found instead.

Simulated annealing is analogous to decreasing the temperature slowly, allowing ample time for the redistribution of the molecules. By analogy with the physical process, the temperature T is initially high. Therefore, the probability of accepting a move that increase the objective function is initially high. The temperature is

gradually decreased as the search progresses. In the end, the probability of accepting a move that increase the objective function becomes vanishingly small. It allows occasional uphill jumps, which make it possible to hop out of local minima and finally find the global minimum.

In general, a simulated annealing algorithm consists of the following:

1. A specification of the initial configuration $\mathbf{n} = (n_1, ..., n_N)$ for the system.

2. A scheme by which subsequent configurations are randomly generated.

3. A criterion according to which the new configurations are rejected or accepted.

Through a large number of iteration using steps 2 & 3, it is expected to reach the minimax design or near minimax design.

For simplicity, we restrict our discussion to the case where one of (n, N) is a multiple of the other. If $n < N$, then we also require that they have the same parity. We also assume that the experiment is to be carried out on the interval $[0, 1)$ and assume $\mathcal{S}$ to have equally spaced design points:

$$x_i = \frac{2i-1}{2N},\ i = 1,\ 2, \cdots, N. \tag{2.94}$$

The initial configuration, and method of generating new configurations, are as used by Fang and Wiens (2000). We proceed as follows.

Step 1. If $n > N$, we take the uniform design, with $n_i = n/N$ for $i = 1, ..., N$, as the initial configuration. If $n \leq N$, the initial configuration is constructed by

repeating the vector $(1, 0, \cdots, 0)$ (with $N/n - 1$ zeros) $[n/2]$ times followed by the same vector with the order of its elements reversed. If N is odd, we insert a vector $(0, \cdots, 0, 1, 0, \cdots, 0)$ of length N/n in the middle. Thus, the initial design is symmetric and at least close to uniform.

Step 2. Define $\mathbf{v} = (n_1, \cdots, n_{[N/2]})$, $V_+ = \{i | v_i > 0\}$ and $V_0 = \{i | v_i = 0\}$. Denote by n_+ and n_0 the number of elements in V_+ and V_0 respectively. Generate a Bernoulli random variable

$$B = \begin{cases} 1, & \text{with probability } \frac{n_0}{n_0+n_+}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.95}$$

If $n_+ \geq 2$, we pick two indices (u_1, u_2) from V_+, at random and without replacement. If $B = 1$ we also randomly select one element u_0 from V_0. We then construct a new vector $\tilde{\mathbf{v}}$ whose elements are those of $\mathbf{v}$ except for

$$\tilde{v}_{u_0} = v_{u_0} + B, \ \tilde{v}_{u_1} = v_{u_1} - 1, \ \tilde{v}_{u_2} = v_{u_2} + 1 - B. \tag{2.96}$$

If V_+ contains only one element u_1, we randomly pick an index u_0 from V_0 and replace (2.96) by

$$\tilde{v}_{u_0} = v_{u_0} + 1, \ \tilde{v}_{u_1} = v_{u_1} - 1.$$

If N is even, we construct a new configuration $\tilde{\mathbf{n}} = (\tilde{n}_1, \cdots, \tilde{n}_N) = (\tilde{v}_1, \cdots, \tilde{v}_{[N/2]}, \tilde{v}_{[N/2]}, \cdots, \tilde{v}_1)$. If N is odd, we generate one additional Bernoulli random vari-

able

$$B_1 = \begin{cases} 1, & \text{with probability } \frac{1}{N}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.97}$$

If a failure is obtained, then $\tilde{n}_{[N/2]+1}$ is the same as $n_{[N/2]+1}$. If a success is obtained, then $\tilde{n}_{[N/2]+1}$ is set equal to $n_{[N/2]+1} + 2$ with probability $1/2$. To account for this increase by 2, we randomly and symmetrically reduce the total remaining frequency by 2. With the remaining probability $1/2$, we have $\tilde{n}_{[N/2]+1} = n_{[N/2]+1} - 2$ where the total remaining frequencies are increased randomly and symmetrically by 2 to compensate for the reduction. This step is omitted if $n_{[N/2]+1} < 2$. The new configuration $\tilde{\mathbf{n}}$ is obtained as above by inserting $\tilde{n}_{[N/2]+1}$ in the middle.

Step 3. The new configuration is accepted, and iterations continue, if the difference in loss $\Delta\mathcal{A}(\tilde{n}) - \mathcal{A}(n)$ is negative. Otherwise, $\tilde{n}$ is accepted if the value $\exp\{-\Delta\mathcal{A}/T\}$ of the Boltzmann acceptance probability exceeds .5, where T is a user-chosen parameter. This means we always move to the new configuration if the corresponding loss is lower than that of the old configuration. Better points are always accepted. And new configurations with higher loss are accepted only with a smaller, temperature-dependent probability, which enables the algorithm to leave local minima. At any temperature there is a chance for the system to move 'upwards'.

If the initial value of the 'temperature' T is too large, the algorithm may take a long

time to converge. If the values of T is too small, the algorithm may yield a local minimax design instead of a global minimax design. Following Haines (1987), we initially choose T in such a way that at least 50 percent of the new configurations are accepted, which allow the algorithm escape from a local minimum in a few steps. In order to obtain a global minimum, a proper cooling schedule is also important. Following the discussion by Press, Flannery, Teukolsky and Vetterling (1989), we decrease T by a factor of 0.9 after each 100 iterations.

2.4.1 Designs for Multiwavelet Models

In this section, we present integer-valued designs which minimize (2.72) and (2.93) when the model (2.17) is based on the multiwavelet system and the parameters are to be estimated by the weighted least squared estimator. Note that in (2.72) and (2.93), only ν need to be specified, which can be viewed as a weighting factor supplied by the experimenter that measures the relative importance of loss due to variation versus that due to bias. Smaller ν gives more emphasis to the bias.

For the purpose of our example, we take $N = 2$. For $N = 2$, the multiwavelet orthonormal basis for $\mathcal{L}_2([0,1])$ is given by

$$\{\phi_0,\, \phi_1\} \cup \{{}_2\omega_0^{j,k}(x),\, {}_2\omega_1^{j,k}(x),\, j \geq 0, 0 \leq k \leq 2^j - 1\}, \tag{2.98}$$

where ${}_N\omega_l^{j,k}(x) = 2^{j/2}{}_N\omega_l(2^j x - k)$, $l = 0, 1$. The scaling functions and primary wavelets are defined by (1.52), (1.53), (1.54) and (1.55). For multiwavelets the

representations discussed in Section 2.2 can be written as,

$$\eta(x) = \sum_{l=0}^{1} d_l \phi_{0l}(x) + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} c_{jk\,2} \omega_0^{j,k}(x) + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} e_{jk\,2} \omega_1^{j,k}(x) + f(x). \quad (2.99)$$

In our examples we have found that the designs change slowly with changes in ν, which gives us flexibility to choose ν in practical applications. Minimax designs are derived for various values of ν, for both the T-optimal and D-ratio criteria. Also performance of the annealing program is examined using various values of constant T.

Multiwavelet systems: We exhibit the minimax T-optimal and D-ratio designs for the multiwavelet approximation obtained by implementing the annealing described above for both the case (i) $n = 64,\ N = 32,\ m = 2$ and (ii) $n = 48,\ N = 96,\ m = 3$. To see the influence of ν on the designs, we derive designs for $\nu = 0.5$, $\nu = 5$ and $\nu = 10$, respectively. We also examine the performance of the annealing program by using different values of T.

Figure 2.1 illustrate examples of T-optimal designs for case (i) obtained using $T = 1.5$. Figure 2.2 display T-optimal designs for case (ii) obtained using $T = 1$. Figure 2.3 and Figure 2.4 shows the D-ratio designs for case (i) and case (ii), respectively. These figures show the plots of loss function versus the number of iterations for various values of constant T in (a). These plots indicate that the search for the minimax design is more intensive and the convergence rate is slower for larger values of T. The integer-valued designs for $\nu = 0.5,\ 5$ and 10 have been shown in (b), (c), and (d), respectively. Optimal allocations is indicated in the

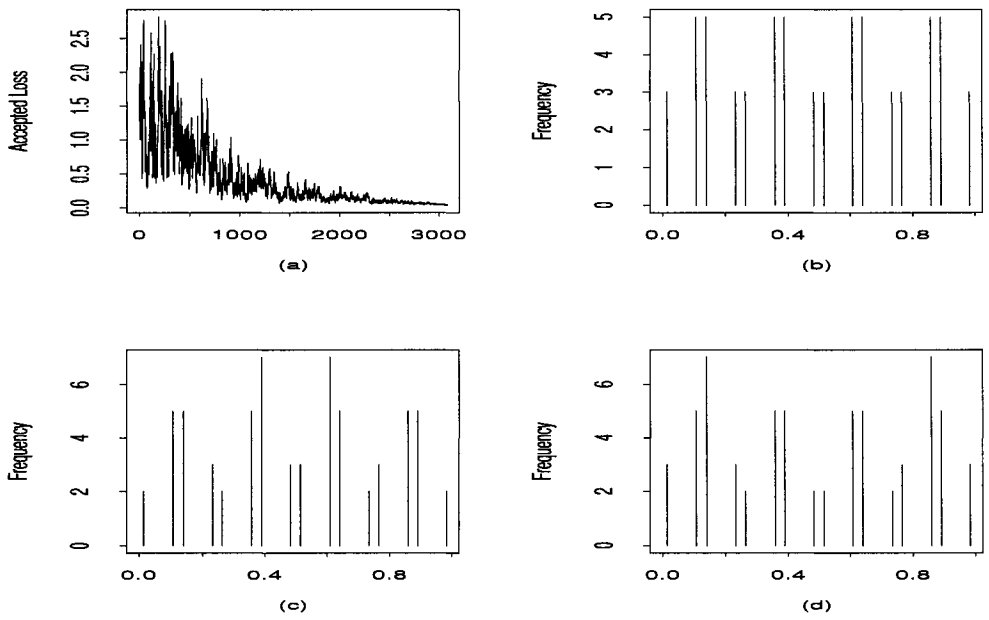

Figure 2.1: T-optimal exact integer-valued designs of a multivavelet model with $m = 2$, $n = 64$, $N = 32$ and $T = 1.5$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

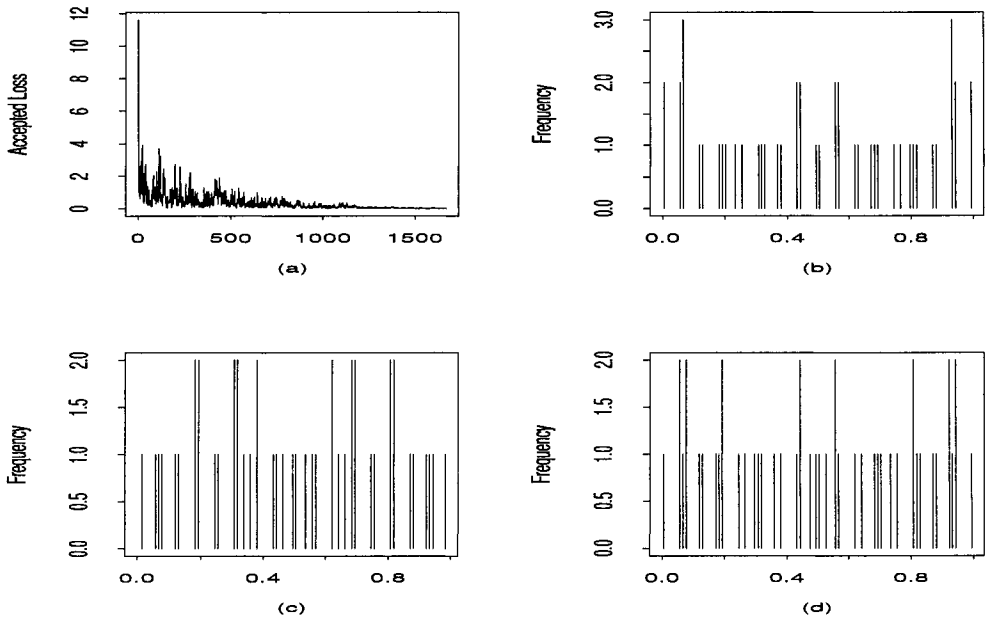

Figure 2.2: T-optimal exact integer-valued designs of a multivavelet model with $m = 3$, $n = 48$, $N = 96$ and $T = 1$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

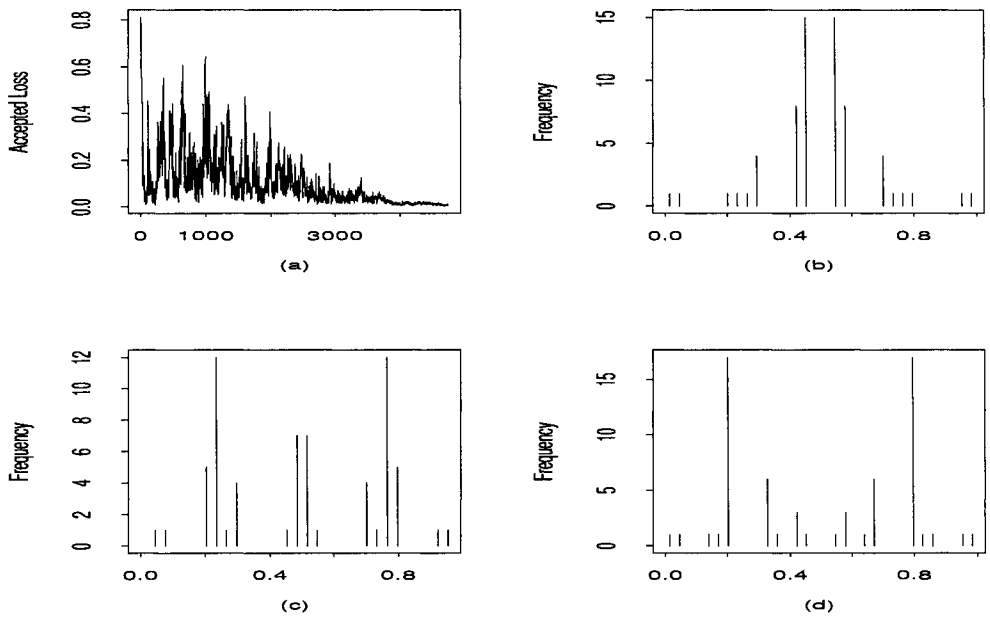

Figure 2.3: D-ratio exact integer-valued designs of a multiwavelet model with $m = 2$, $n = 64$, and $N = 32$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

y-axis by frequencies, and the optimal ordering is indicated by the numbers in the brackets. We have found that the designs change slowly with changes in ν and are symmetry about $x = \frac{1}{2}$

2.4.2 Designs for Daubechies Models

In this section, we exhibit both the T-optimal and D-ratio designs for the Daubechies wavelets approximation obtained by implementing the simulated annealing algorithm. The Daubechies scaling functions and primary wavelets have no closed form. Thus, we use Construction 4 described in Section 1.3.1 to construct them numerically. We use the Daubechies wavelet ${}_5\psi(x)$ in our examples. This choice is based on

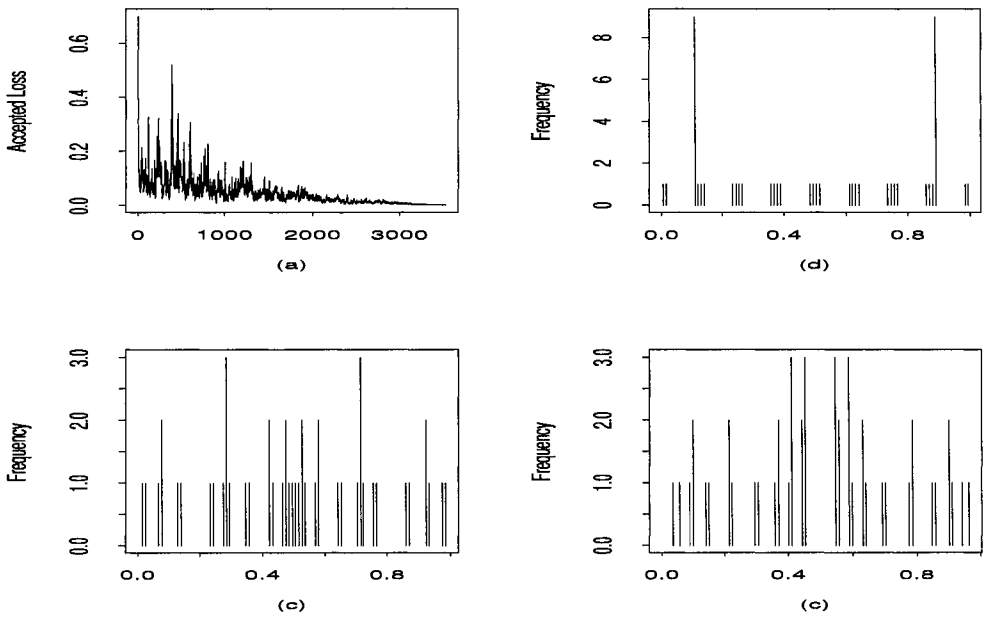

Figure 2.4: D-ratio exact integer-valued designs of a multiwavelet model with $m = 3$, $n = 48$, and $N = 96$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

the fact that it appears to perform better than the others in approximating curves based on weighted least squares; see Oyet (2002).

Daubechies wavelets: We exhibit the T-optimal and D-ratio designs for the Daubechies approximation obtained by implementing the annealing described above for both the case (i) $n = 64,\ N = 32,\ m = 2$ and (ii) $n = 48,\ N = 96,\ m = 3$.

Figure 2.5 display the T-optimal designs for case (i) obtained using $\nu = 0.5$, $\nu = 5$ and $\nu = 10$, respectively. And Figure 2.6 display the D-ratio designs for case (ii) obtained using $\nu = 0.5$, $\nu = 5$ and $\nu = 10$, respectively. In contrast to the designs based on the multiwavelet system, the optimal designs for the Daubechies model are non-symmetric;

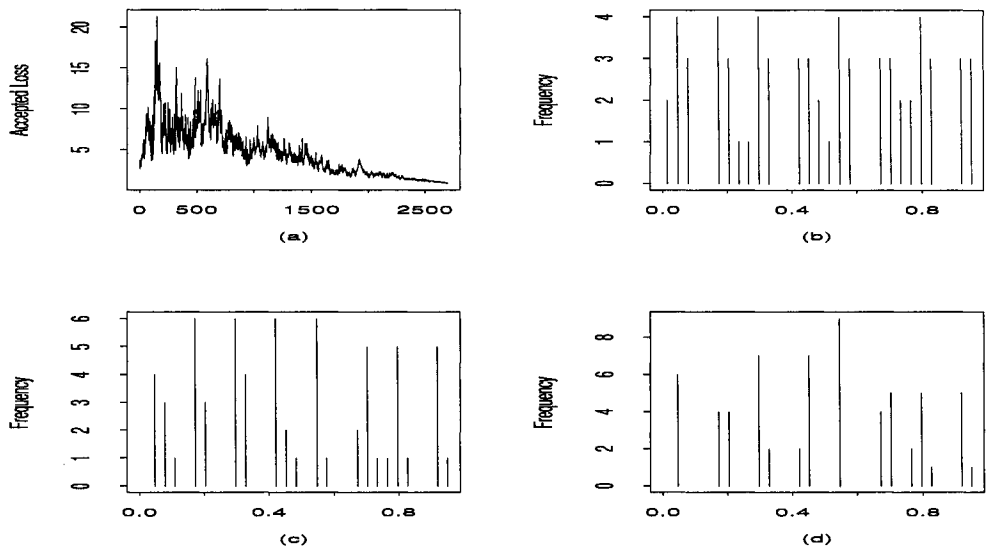

Figure 2.5: T-ratio exact integer-valued designs of a Daubechies model with $m = 2$, $n = 64$, and $N = 32$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

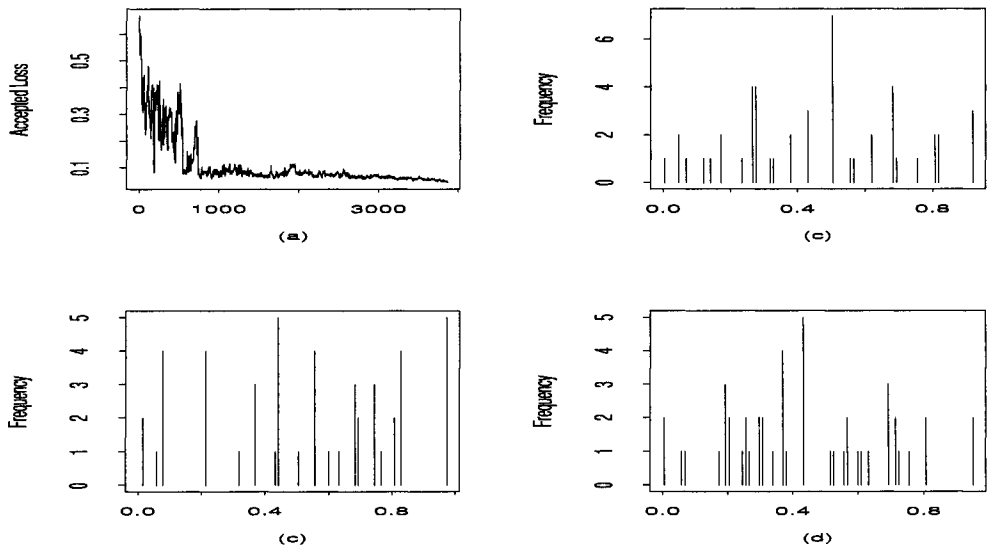

Figure 2.6: D-ratio exact integer-valued designs of a Daubechies model with $m = 3$, $n = 48$, and $N = 96$. (a) Accepted loss versus iteration number; (b-d) Design points and frequencies for $\nu = 0.5,\ 5,\ 10$.

Chapter 3

SEQUENTIAL WAVELET DESIGNS FOR MODEL DISCRIMINATION

3.1 Introduction

Sequential designs use accruing data to select future design points. They are important in many problems. In nonlinear situations the optimal experimental designs depend on unknown parameters to be estimated. A sequential approach is then naturally suggested: one should choose design points so as to maximize a measure of performance evaluated at the estimates obtained from observations made at previous designs points.

Sequential designs for linear models have been studied by, among others, Fedorov (1972), Gebhardt and Heckendorff (1983) and Schwabe (1991). Schwabe (1990) establishes optimality properties of such designs in exactly linear models. Fordorov (1971), Wynn (1970) and (1972) used sequential methods for computing D-optimum

designs. The basic idea of such methods is that we start from an initial design and we add new points to the design following a certain rule. The main problem in the theory of such methods is proving that they allow to approximate the optimum design as closely as is necessary. And Pázman (1974) proved that in a sequential design of a regression problem the variance of the least squares estimate for the response in the nth sequential point tends to zero with $n \to \infty$ which allows the proof of the convergence of certain procedures for computing D_a-optimum designs. Chaudhuri and Mykland (1993) developed a fully adaptive sequential design on a very general nonlinear setup that includes many models commonly encountered in practice. Subsequent design points were chosen to maximize the determinant of the Fisher information matrix of the design, evaluated at the current parameter estimated by maximum likelihood method. He also showed that the sequential designs was asymptotically D-optimal. Wiens (1996) considered the problem of the sequential choice of design points in an approximately linear model. He assumed that the parameters are estimated by M-estimation and chose the next design point to minimize the resulting integrated squared bias of the estimated response. He also showed that the sequential designs compare favorably with some fixed-sample-size designs in a simulation study. Sinha and Wiens (2002) introduced the formal notation of an approximately specified nonlinear regression model, and investigated sequential design methodologies when the fitted model is possibly of an incorrect parametric form.

Sequential experimental designs for discrimination between two regression models in the absence of constraints on the values of the parameters have been investigated by several authors. Hunter and Reiner (1965) proposed the idea to select the point at which the models differ the most by maximizing the residual sum of squares of the incorrect model. Atkinson and Cox (1974) use the equivalence of G-optimum and D-optimum designs in the sequential construction of D-optimum designs for model discrimination. Atkinson and Fedorov (1975) considered procedures which lead to designs which are asymptotically T-optimum and which give, at each trial, the largest increase in the expected value of the sum of squares of differences between the responses from the two models.

3.2 Sequential Designs

A possible disadvantage of the designs of the previous section is that the procedure is not directly adaptable to be sequential. In this section we therefore look at an iterative algorithm for constructing designs in which the design is built up one trial at a time using the equivalence of G-optimum and D-optimum designs. A design for all the parameters in a regression model is G-optimum if the maximum variance of the predicted response over the experimental region is a minimum. The iterative algorithm we have used, is a slight modification of Atkinson and Cox (1974) procedure which takes into consideration the fact that the wavelet model is only an approximation. Instead of using the covariance matrices of $\boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_{m-1}$, we shall

use the mean squared error matrixes $M(\mathbf{f}, \xi_N)$ and $M^*(\mathbf{f}^*, \xi_N)$ to account for the bias in estimation by assuming the approximation is exact.Thus, the modified equivalent optimum design for the subset of parameters $\boldsymbol{\gamma}$ minimizes the maximum of

$$d^{(n)}(x, z) = q^T(x)M(\mathbf{f}, \xi_N)q(x) - q^{*T}(x)M^*(\mathbf{f}^*, \xi_N)q^*(x), \tag{3.1}$$

where $q^T(x)$ and $q^{*T}(x)$ are the components of the wavelet system used in the decomposition for order m and m-1, respectively. Recall that $M(\mathbf{f}, \xi_N)$ and $M^*(\mathbf{f}^*, \xi_N)$ are the mean squared error matrices for wavelet models of order m and m-1,respectively.

To solve the minimax problem, we need to solve the maximization for the mth order model and the $(m-1)$th order model, respectively. First we look at the maximization problem for the mth order model

$$\max_{\mathbf{f}} q^T(x)M(\mathbf{f}, \xi_N)q(x). \tag{3.2}$$

Substituting (2.26) into (3.2), we obtain

$$\begin{aligned} & \max_{\mathbf{f}} q^T(x)M(\mathbf{f}, \xi_N)q(x) \\ = \; & \max_{\mathbf{f}} \left\{ q^T(x)\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}q(x) + \frac{\sigma^2}{n} q^T(x)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}q(x) \right\} \end{aligned} \tag{3.3}$$

First, we look at the term $q^T(x)\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}q(x)$ in the above expression. Clearly,

$$\begin{aligned} & q^T(x)\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}q(x) \\ = \; & q^T(x)(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{f}(\mathbf{Q}^T\mathbf{M}\mathbf{f})^T(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}q(x). \end{aligned} \tag{3.4}$$

We can rewrite (3.4) as

$$\mathbf{f}^T\mathbf{M}\mathbf{Q}(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}q(x)q^T(x)(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{f}. \tag{3.5}$$

Substituting the singular value decomposition (2.28) of Q into (3.5) and simplifying the result, we obtain

$$\begin{aligned} & q^T(x)\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}q(x) \\ = \; & \mathbf{f}^T(x)\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}q(x)q^T(x)(\mathbf{V}^T)^{-1}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{f}(x). \end{aligned} \tag{3.6}$$

For simplicity of notation, we define

$$\mathbf{a}(\mathbf{x}) = \mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}q(x). \tag{3.7}$$

Then, we can rewrite (3.6) as

$$q^T(x)\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}q(x) = \mathbf{f}^T\mathbf{a}(\mathbf{x})\mathbf{a}^{\mathbf{T}}(\mathbf{x})\mathbf{f}. \tag{3.8}$$

So, our problem is trying to maximize (3.8) subject to the constraint (2.10) and (2.15). According to (2.32), our problem become

$$\begin{aligned} \max_{\mathbf{f}} \mathbf{f}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\mathbf{f} &= \max_{\mathbf{e}} \tau^2 N\mathbf{e}\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}\mathbf{e} \\ &= \max_{\mathbf{e}} \tau^2 N\mathbf{e}^T \sum_{i=1}^{2^{m+1}} \lambda_i(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}})\mathbf{v}\mathbf{v}^T\mathbf{e}, \end{aligned} \tag{3.9}$$

subject to $\|\mathbf{e}\| = 1$, where $\lambda_i(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}})$ is the ith eigenvalue, and $\mathbf{v}$ is the corresponding normalized eigenvector. It is easy to see that

$$\tau^2 N\mathbf{e}^T \sum_{i=1}^{2^{m+1}} \lambda_i(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}})\mathbf{v}\mathbf{v}^T\mathbf{e} \leq \tau^2 N\lambda_{max}(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}). \tag{3.10}$$

Thus,

$$\max_{\mathbf{f}} \mathbf{f}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\mathbf{f} = \tau^2 N\lambda_{max}(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}), \tag{3.11}$$

where $\lambda_{max}(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}})$ is the maximum eigenvalue of $\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}$. From the fact that

$$\lambda_{max}(\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}) = \lambda_{max}(\mathbf{a}^T(\mathbf{x})\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{a}(\mathbf{x})) \tag{3.12}$$

and

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I} - \mathbf{U}\mathbf{U}^T, \tag{3.13}$$

we can rewrite (3.11) as

$$\max_{\mathbf{f}} \mathbf{f}^T\mathbf{a}(\mathbf{x})\mathbf{a}^T(\mathbf{x})\mathbf{f} = \tau^2 N[\mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x}) - \mathbf{a}^T(\mathbf{x})\mathbf{U}\mathbf{U}^T\mathbf{a}(\mathbf{x})]. \tag{3.14}$$

Now, we look at the second part of (3.3)

$$\frac{\sigma^2}{n}\, q^T(x)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}q(x). \tag{3.15}$$

Substituting $\mathbf{B}$ and $\mathbf{D}$ into the above expression, we obtain

$$\frac{\sigma^2}{n} q^T(x)(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{M}\mathbf{W}\mathbf{Q}(\mathbf{Q}^T\mathbf{M}\mathbf{Q})^{-1}q(x). \tag{3.16}$$

Using the singular value decomposition of $\mathbf{Q}$ and simplifying the result, we have

$$\frac{\sigma^2}{n} q^T(x)[(\mathbf{V}^T)^{-1}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{M}_W\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}]q(x). \tag{3.17}$$

Therefore, combine (3.14) and (3.17) to obtain

$$\begin{aligned}\max_{\mathbf{f}} q^T(x)M(\mathbf{f},\xi_N)q(x) = {} & \tau^2 N[\mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x}) - \mathbf{a}^T(\mathbf{x})\mathbf{U}\mathbf{U}^T\mathbf{a}(\mathbf{x})] \\ & + \frac{\sigma^2}{n} q^T(x)[(\mathbf{V}^T)^{-1}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{M}_W\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}]q(x).\end{aligned} \tag{3.18}$$

Similarly, the maximization problem for the $(m-1)$th order model can then be solved as

$$\max_{\mathbf{f}^*} q^{*T}(x)M^*(\mathbf{f}^*,\xi_N)q^*(x) = \tau^2 N[\mathbf{a}^{*T}(\mathbf{x})\mathbf{a}^*(\mathbf{x}) - \mathbf{a}^{*T}(\mathbf{x})\mathbf{U}^*\mathbf{U}^{*T}\mathbf{a}^*(\mathbf{x})]$$
$$+\frac{\sigma^2}{n}q^{*T}(x)[(\mathbf{V}^{*T})^{-1}(\mathbf{\Lambda}^*)^{-1}(\mathbf{M}_1^*)^{-1}\mathbf{M}_W^*(\mathbf{M}_1^*)^{-1}(\mathbf{\Lambda}^*)^{-1}(\mathbf{V}^*)^{-1}]q^*(x), \quad (3.19)$$

where $\mathbf{a}^*(\mathbf{x})$, $\mathbf{M}_1^*$ and $\mathbf{M}_W^*$ are the equivalents of $\mathbf{a}(\mathbf{x})$, $\mathbf{M}_1$ and $\mathbf{M}_W$.

For the purpose of simplicity, we define

$$\mathbf{A}_1 = \mathbf{a}^T(\mathbf{x})\mathbf{a}(\mathbf{x}) - \mathbf{a}^T(\mathbf{x})\mathbf{U}\mathbf{U}^T\mathbf{a}(\mathbf{x});$$

$$\mathbf{A}_1^* = \mathbf{a}^{*\mathbf{T}}(\mathbf{x})\mathbf{a}^*(\mathbf{x}) - \mathbf{a}^{*\mathbf{T}}(\mathbf{x})\mathbf{U}^*\mathbf{U}^{*\mathbf{T}}\mathbf{a}^*(\mathbf{x});$$

$$\mathbf{A}_2 = q^T(x)[(\mathbf{V}^T)^{-1}\mathbf{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{M}_W\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}]q(x);$$

$$\mathbf{A}_2^* = q^{*T}(x)[(\mathbf{V}^{*T})^{-1}(\mathbf{\Lambda}^*)^{-1}(\mathbf{M}_1^*)^{-1}\mathbf{M}_W^*(\mathbf{M}_1^*)^{-1}(\mathbf{\Lambda}^*)^{-1}(\mathbf{V}^*)^{-1}]q^*(x).$$

Using these notations, we can summarize the solution to the G-optimal design in the following Theorem.

Theorem 3.2.1 *Let* $\mathbf{A}_1$, $\mathbf{A}_1^*$, $\mathbf{A}_2$ *and* $\mathbf{A}_2^*$ *be as defined above. Then, for fixed* $\nu = \frac{\sigma^2}{n\tau^2}$ *and* $\nu^* = \frac{\sigma^2}{n\tau^{*2}}$, *the equivalent optimal design minimizes:*

$$\max_{\mathbf{f},\mathbf{f}^*} \mathbf{d}(\mathbf{q}(\mathbf{x}),\mathbf{q}^*(\mathbf{x})) = \mathbf{N}\tau[\mathbf{A_1} + \frac{\nu}{\mathbf{N}}\mathbf{A_2}] - \mathbf{N}\tau^{*\mathbf{2}}[\mathbf{A_1^*} + \frac{\nu^*}{\mathbf{N}}\mathbf{A_2^*}]. \quad (3.20)$$

As we mentioned in Chapter 2, ν can be viewed as a weighting factor reflecting the relative importance of bias versus variance to the experimenter. The experimenter may chose to assign the same weight to ν in the two competing models. That is $\frac{\nu}{\nu^*} = 1$. In that case, we only need to focus on a single value of ν in (3.20).

This equivalence of D- and G-optimum designs is used in the sequential construction of D-optimum designs. To construct a D-optimum design iteratively for the subset γ we proceed one trial at a time adding for the $(n+1)th$ observation a trial at the point where (3.20) is a maximum. We will discuss examples of these designs in the following section.

3.3 Simulation Study

In this section, we investigate the sequential design method by simulation. The idea of sequential method is that we start from an initial design, in which design points are chosen with little or no knowledge of the unknown parameter, followed by a fully adaptive sequential stage in which the design point are chosen sequentially, exploiting a G-optimality criterion and using the mean squared error evaluated at the current parameter estimates.

The sequential design approach can also be combined with a non-sequential design approach. Suppose after taking n observations using the non-sequential design approach, the results are not conclusive as to which model is best. In order to discriminate among these two rival models the experimenter may wish to perform further experimental runs. If experiments are conducted in sequence, the experimenter can consider each result before he runs the next experiment. In particular, computations prior to each new experiment can indicate where the next experiment ought to be conducted to provide the maximum discrimination between two rival

models.

For the purpose of example, we use one of the D-ratio optimal designs we obtained in Chap 2.4 as the initial design. We seek to minimize (3.20) resulting from one further observation at a point $x \in [0, 1)$.

Example I: We exhibit the sequential design for the multiwavelet approximation. We consider initial design of size $n = 64$ with $N = 32$ design points, where x_i are evenly distributed on $[0, 1)$. The D-ratio initial design points are listed in Table 3.1. We seek to add another 16 observations using the sequential approach. Table 3.2 shows the sequential design points and frequency for discriminating between the two models, respectively. From the table, we found the sequential design has added 2 observations to each of the 8 new design points at $x = 0.109375,\ 0.140625,\ 0.390625,\ 0.390625,\ 0.609375,\ 0.640625,\ 0.859375$ and 0.890625. Figure 3.1 shows the sequential minimax designs. Optimal allocations indicated in the y-axis by frequencies, and the optimal ordering is indicated in the x-axis. It is clear that the sequential design maintains the symmetry of the initial design.

Example II: In this example, we exhibit the sequential design for the Daubechies wavelet approximation. We start with an initial design of size $n = 48$ with $N = 96$ design points, where x_i are evenly distributed on $[0, 1)$. The D-ratio initial design points are listed in Table 3.3. We seek to generate another 16 observations one at a

Table 3.1: Initial Design Points

value of x	number of trials	value of x	number of trials
0.015625	1	0.546875	15
0.046875	1	0.578125	8
0.203125	1	0.703125	4
0.234375	1	0.734375	1
0.265625	1	0.765625	1
0.296875	4	0.796875	1
0.421875	8	0.953125	1
0.453125	15	0.984375	1

time using the sequential approach. Table 3.4 shows the sequential design points and frequency for discriminating between the two models, respectively. From the table, we found the sequential design has added observations to 12 new design points at $x =$ 0.098958333, 0.338541667, 0.463541667, 0.557291667, 0.588541667, 0.651041667, 0.713541667, 0.776041667, 0.848958333, 0.869791667, 0.963541667 and 0.984375000. Figure 3.2 shows the sequential minimax designs. Optimal allocations are indicated in the y-axis by frequencies, and the optimal ordering is indicated in the x-axis. It is clear that the sequential design is not symmetric for the Daubechies wavelet.

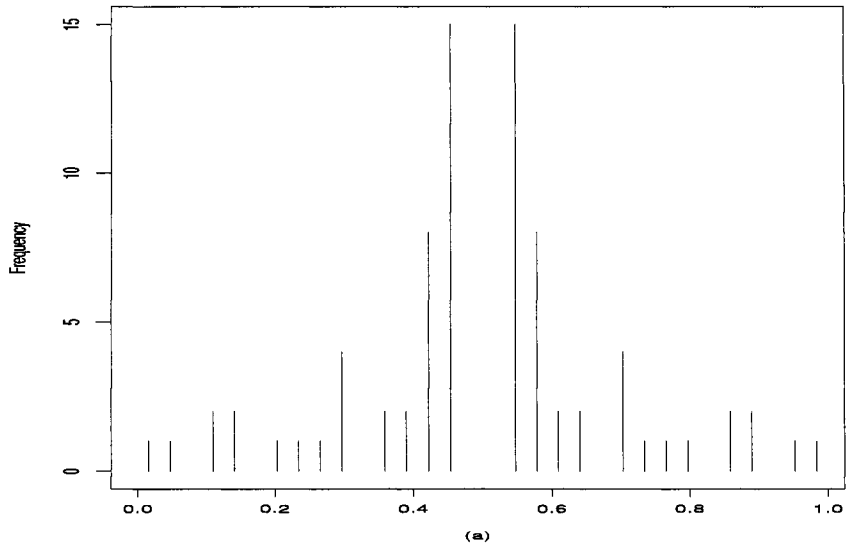

Figure 3.1: Sequential design points and frequencies for a multiwavelet model with $m = 2$, $n = 32$, and $N = 64$.

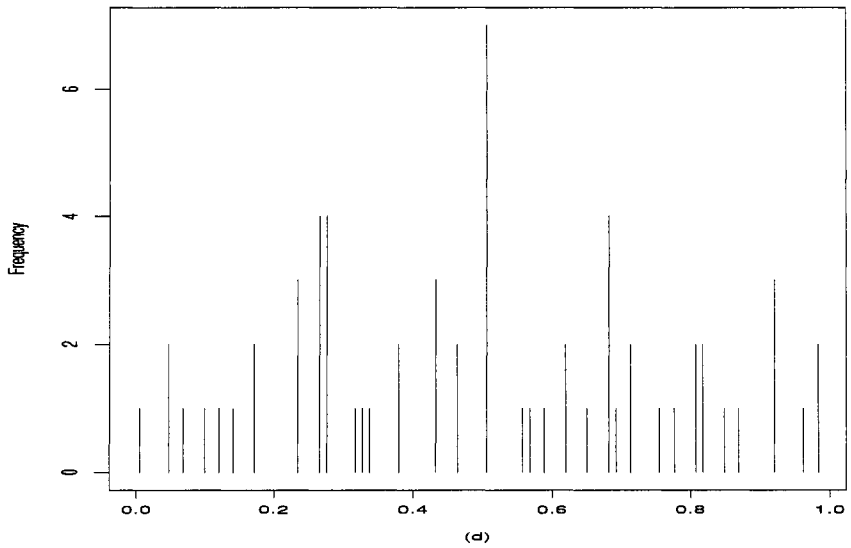

Figure 3.2: Sequential design points and frequencies for a Daubechies wavelet model with $m = 3$, $n = 48$, and $N = 96$.

Table 3.2: Final Design Points

value of x	number of trials	value of x	number of trials
0.015625	1	0.546875	15
0.046875	1	0.578125	8
0.109375	2	0.609375	2
0.140625	2	0.640625	2
0.203125	1	0.703125	4
0.234375	1	0.734375	1
0.265625	1	0.765625	1
0.296875	4	0.796875	1
0.359375	2	0.859375	2
0.390625	2	0.890625	2
0.421875	8	0.953125	1
0.453125	15	0.984375	1

Table 3.3: Initial Design Points

value of x	number of trials	value of x	number of trials
0.005208333	1	0.432291667	3
0.046875000	2	0.505208333	7
0.067708333	1	0.557291667	1
0.119791667	1	0.567708333	1
0.140625000	1	0.619791667	2
0.171875000	2	0.682291667	4
0.234375000	1	0.692708333	1
0.265625000	4	0.755208333	1
0.276041667	4	0.807291667	2
0.317708333	1	0.817708333	2
0.328125000	1	0.921875000	3
0.380208333	2		

Table 3.4: Final Design Points

value of x	number of trials	value of x	number of trials
0.005208333	1	0.557291667	1
0.046875000	2	0.567708333	1
0.067708333	1	0.588541667	1
0.098958333	1	0.619791667	2
0.119791667	1	0.651041667	1
0.140625000	1	0.682291667	4
0.171875000	2	0.692708333	1
0.234375000	3	0.713541667	2
0.265625000	4	0.755208333	1
0.276041667	4	0.776041667	1
0.317708333	1	0.807291667	2
0.328125000	1	0.817708333	2
0.338541667	1	0.848958333	1
0.380208333	2	0.869791667	1
0.432291667	3	0.921875000	3
0.463541667	2	0.963541667	1
0.505208333	7	0.984375000	2

Chapter 4

CONCLUDING REMARKS

In this study, we have developed exact integer-valued designs for discriminating between two competing nonparametric models, based on their wavelet representations. Wavelets was introduced into the problem by exploiting the flexibility of wavelet approximations to approximate the unknown response curve by its wavelet expansion. We noticed that in the representation only finite number of terms can be estimated by weighted least squares. The bias arising from this, compounds the natural variation of the estimates. The objective of our design is to select the correct wavelet representation in the design stage. We determined the robust design with respect to a minimax criterion which minimize a mean squared error based loss function of the estimates.

We have exhibited two non-sequential design criteria based on the mean squared error in Chapter 2. The modified T-optimal criteria depends directly on the noncentrality parameters of the F-statistics. We adopted the Bayesian approach to solve the dependence of the criterion on the unknown parameter $\boldsymbol{\beta}_m$. A D-ratio opti-

mality criterion was defined as the ratio of the determinants of the mean squared error for the two competing wavelet models. The use of a finite design space and of a simulated annealing algorithm has greatly facilitated the construction of the designs. In particular, the simulated annealing algorithm has allowed us to present exact integer-valued designs in situations in which only continuous designs could previously be considered. Examples show that robust optimal designs are symmetric for both the criterion using the multiwavelet approximation. We also observed that the designs constructed using the Daubechies wavelet are non-symmetric.

The optimal designs in Chapter 3 are constructed sequentially using the equivalence of the G-optimum and D-optimum designs. We proceeded one trial at a time, adding one new observation at the point at which the G-optimal criterion was maximized using the mean squared error evaluated at current data.

Apart from the problems we have discussed, designs which required further studies are

1. Exact minimax wavelet designs for discrimination and estimation

2. Exact robust designs for model discrimination when the estimators of the parameter are generalized M-estimators or other robust estimators.

3. Exact robust designs for model discrimination for biased wavelet regression models with autocorrelated errors.

4. Exact robust designs for model discrimination for biased wavelet regression

models with heteroscedastic errors.

5. Exact robust designs for discriminating between multivariate wavelet models.

We hope that this work will motivate further research in the direction of constructing designs for wavelet regression models.

REFERNCES

Alpert, B. K. (1992), "Wavelets and other bases for fast numerical linear algebra," *Wavelets: A Tutorial In Theory And Application*, Academic Press, Boston, 181-216.

Anderson, T. W. (1962), "The chioce of the degree of a polynomial regression as a multiple decision problem," *Annals of Mathematical Statistics,* 33, 255-265.

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

Atkinson, A. C. (1972), "Planning experiments to detect inadequate regression models," *Biometrika*, 59, 275-293.

Atkinson, A. C. and Cox, D. R. (1974), "Planning experiments for discrimination between models," *Journal of the Royal Statistical Society, Series B,* 47, 1-52.

Atkinson, A. C. and Fedorov, V. V. (1975), "The design of experiments for discrimination between two rival models," *Biometrika*, 62, 57-70.

Beckenbach, E. F. and Bellman, R. (1965), *Inequalities*, Berlin: Springer-Verlag, New York.

Bohachevsky, I. O., Johnson, M. E. and Stein, M. L. (1986), "Generalized Simulated Annealing for Function Optimization," *Technometrics*, 28, 209-217.

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

Box, G. E. P. and Draper, N. R. (1975), "Robust designs," *Biometrika*, 62, no. 2, 347–352.

Chan, L. (1991), "Optimal design for estimation of variance in nonparametric regression using first order differences," *Biometrika,* 78, 926-926.

Chaudhuri, P. and Mykland, P. A. (1993), "Nonlinear experiments: optimal design and inference based on likelihood," *Journal of the American Statistical Association* , 88, no. 422, 538-546.

Chernoff, H. (1953), "Locally optimal designs for estimating parameters," *Annals of Mathematical Statistics*, 24, 586–602.

Chu, C.K., Marron, J. S. (1991), "Choosing a kernel regression estimator," With comments and a rejoinder by the authors. *Statist. Sci.,*6, no. 4, 404–436.

Daubechies, I. (1992), *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics (SIAM).

Daubechies, I. and Lagarias, J. C. (1991), "Two-scale difference equations. I. Existence and global regularity of solutions," *SIAM Journal on Mathematical Analysis*, 22, no. 5, 1388-1410.

Daubechies, I. and Lagarias, J. C. (1992), "Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals," *SIAM Journal on Mathematical Analysis*, 23, no. 4, 1031-1079.

Dette, H. (1995), "Optimal designs for identifying the degree of a polynomial regression," *The Annals of Statistics,* 23, 1248-1267.

Dette, H. (1997), "Designing experiments with respect to "standardized" optimal-

ity criteria," *Journal of the Royal Statistical Society. Series B. Methodological*, 59, no. 1, 97-110.

Dette, H. and Haller, G. (1998), "Optimal designs for the identification of the order of a fourier regression," *The Annals of Statistics*, 26, no.4, 1496-1521.

Donoho, D. L. and Johnstone, I. M. (1995), "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, 90, no. 432, 1200-1224.

Elfving, G. (1952), "Optimum allocation in linear regression theory," *Annals of Mathematical Statistics* 23, 255262.

Elfving, G. (1955), "Geometric allocation theory," *Skandinavisk Aktuarietidskrift* , 37 170190 .

Elfving, G. (1956), "Selection of nonrepeatable observations for estimation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 6975.

Fang, Z. and Wiens, D. P. (2000), "Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models," *Journal of the American Statistical Association*, 95, 807-818.

Fedorov, V. V. (1971), "Design of experiments in the case of the simultaneous measurement of several response surfaces (Russian)," *Teor. Verojatnost. i Primenen*, 16, 319-327.

Fedorov, V. V. (1972), "Theory of Optimal Experiments," *New York: Academic Press.*

Fedorov, V. V. and Malyutov, M. B. (1972), "Optimal designs in regression problems," *Mathematische Operationsforschung und Statistik,* 3, 281-308.

Gebhardt, R.and Heckendorff, H. (1983), " Zur sequentiellen Versuchsplanung für das Regressionsproblem," *Mathematische Operationsforschung und Statistik Series Statistics*, 14, no. 3, 355366.

Haines, L. M. (1987), "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models," *Technometrics*, 29, 439-447.

Härdle, W., Kerkyacharian, G. , Picard, D. and Tsybakov A. (1998), *Wavelets, approximation, and statistical applications*, Springer-Verlag, New York.

Herzberg, A. M., and Traves, W. N. (1994), "An optimal experimental design for the Haar regression model," *The Canadian Journal of Statistics*, 22, 357-364.

Hoel, P. G. (1958), "Efficiency problems in polynomial estimation," *Annals of Mathematical Statistics*, 29, 1134-1145.

Hoel, P. G. (1968), "On testing for the degree of a polynomial", *Technometrics*, 10, 757–767.

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

Hunter, W. G. and Reiner, A. M. (1965), "Designs for discriminating between two rival models," *Technometrics*, 7, 307-323.

Kiefer, J. (1959), "Optimum experimental designs" *Journal of the Royal Statistical Society. Series B. Methodological* 21, 272-319.

Kiefer, J. (1961), "Optimum designs in regression problems, II," *Annals of Mathematical Statistics*, 32, 298-325.

Kiefer, J. (1962), "An extremum result," *Canadian Journal of Mathematics,* 14, 597-601.

Kiefer, J. (1974), "General equivalence theory for optimum designs (approximate theory)," *Annals of Statistics*, 2, 849–879.

Kiefer, J. and Wolfowitz, J (1959), "Optimum designs in regression problems," *Annals of Mathematical Statistics*, 30, 271-294.

Kiefer, J.and Wolfowitz, J. (1960), "The equivalence of two extremum problems," *Canadian Journal of Mathematics*, 12, 363-366.

Li, K. C. and Notz, W. (1982), "Robust designs for nearly linear regression," *Journal of Statistical Planning and Inference*, 6, no. 2, 135-151.

Liu, Shawn X. and Wiens, D. P. (1997), "Robust designs for approximately polynomial regression," *Journal of Statistical Planning and Inference*, 64, no. 2, 369-381.

Mallat, S. G.(1989), "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$," *Transactions of the American Mathematical Society*, 315, no. 1, 69-87.

Marcus, M. B. and Sacks, J. (1976), "Robust designs for regression problems," *Statistical Decision Theory and Related Topics II*, Academic, New York, 245-268.

Mayer, Y. (1992), *Wavelets and Operators*, Cambridge University Press, Cambridge.

Meyer, R. K. and Nachtsheim, C. J. (1988), "Constructing D-Optimal Experimental Designs by Simulated Annealing," *American Journal of Mathematical and Management Sciences*, 3 & 4, 329-259.

Mitchell, T., Sacks, J. and Ylvisaker, D. (1994), "Asymptotic Bayes criteria for nonparametric response surface design," *The Annals of Statistics,* 22, 634-651.

Oyet, A. J. (2002), "Minimax A- and D-Optimal Integer-Valued Wavelet Designs for Estimation," *Canadian Journal of Statistics,* 30(2), 301-316

Oyet, A. J. (2003), "Wavelet Designs for Estimating Nonparametric Curves with Heteroscedastic Error," *Journal of Statistical Planning and Inference*, 117(2), 323-343.

Oyet, A. J. and Sutradhar, B. (2003), "Testing variances in wavelet regression models," *Statistics and Probability Letter*, 61, no. 1, 97-109.

Oyet, A. J. and Wiens, D. P. (2000), "Robust designs for wavelet approximations of regression models," *Journal of Nonparametric Statistics*, 12, no. 6, 837-859.

Oyet, A. J. and Wiens, D. P. (2003), "On Exact Minimax Wavelet Designs for Estimating Nonparametric Regression Models Obtained By Simulated Annealing," *Statistics and Probability Letters*, 61(1), 111-121.

Pázman, A. (1974), "A convergence theorem in the theory of D-optimum experimental designs," *Annals of Statistics*, 2, 216-218.

Pesotchinsky, L. (1982), "Optimal robust designs: linear regression in R^k, *Annals*

of Statistics, 10, no. 2, 511-525.

Pinheiro, A. and Vidakovic, B. (1997), "Estimating the square root of a density via compactly supported wavelets," *Computational Statistics and Data Analysis*, 25, no. 4, 39–415.

Plackett, R. L. and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika,* 33, 305-325.

Pukelsheim, F. (1980), "On linear regression designs which maximize information," *Journal of Statistical Planning and Inference*, 4, no. 4, 339-364.

Pukelsheim, F. (1993), *Optimal design of experiments,* Wiley, New York.

Pukelsheim, F. and Rosenberger, J. L. (1993), "Experimental designs for model discrimination," *Journal of the American Statistical Association* , 88, no. 422, 642-649.

Rioul, O., and Vetterli,M.(1991), "Wavelets and Signal Processing," *IEEE Signal Processing Magazine,* 8, 14-38.

Schwabe, R. (1990), "Local optimality and global superiority of adaptive designs in linear models," *Computational Statistics and Data Analysis*, 10, no. 3, 325-330.

Schwabe, R. (1991), "Adaptive design in the one-way layout," *Journal of Statistical Planning and Inference* , 28, no. 3, 319-338.

Shah, K. R. and Sinha, Bikas K (1989), *Theory of optimal designs,* Lecture Notes in Statistics, 54, Springer, New York.

Silvey, S. D. (1978), "Optimal design measures with singular information matrices," *Biometrika*, 65, no. 3, 553-559.

Silvey, S. D. (1980), *Optimum design*, Chapman & Hall, London.

Sinha, S. and Wiens, D. P. (2002), "Robust sequential designs for nonlinear regression," *Canadian Journal of Statistics*, 30, no. 4, 601-618.

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.

Spruill, M. (1990), "Good designs for testing the degree of a polynomial mean", *Sankhya Ser. B,* 52, 67-74.

Steinberg, M. and Hunter, G. (1984), "Experimental design: review and comment", *Technometrics*, 26, no. 2, 71130.

Strang, G. (1989), "Wavelets and dilation equations: a brief introduction," *SIAM Review*, 31, no. 4, 614-627.

Wald, A. (1943), "On the efficient design of statistical investigations.," *Annals of Mathematical Statistics*, 14, 134-140.

Wegman, E. J.(1991), "The Straight Scoop on Wavelets and Nonparametric Function Estimation," *Proceedings of the 37th Conference on Design of Experiments in Army, Development and Testing.*

Wegman, E. J., Poston, W. L. and Solka, J. L. (1996),"Wavelets and Nonparametric Function Estimation," *Research developments in probability and statistics*,

257274.

White, V. (1973), "An extension of the general equivalence theorem to nonlinear models," *Biometrika*, 60, 345-348.

Whittle, P. (1973), "Some general points in the theory of optimal experimental design," *Journal of the Royal Statistical Society, Series B, Methodological*, 35, 123-130.

Wiens, D. P. (1992), "Minimax designs for approximately linear regression," *Journal of Statistical Planning and Inference*,0 31, no. 3, 353-371.

Wiens, D. P. (1996), "Robust sequential designs for approximately linear models," *Canadian Journal of Statistics*, 24, no. 1, 67-79.

Wynn, Henry P. (1970), "The sequential generation of D-optimum experimental designs," Annals of Mathematical Statistics, 41, 1655-1664.

Wynn, Henry P. (1972), "Results in the theory and construction of D-optimum experimental designs," *Journal of the Royal Statistical Society. Series B. Methodological*, 34, 133-147.

Zhou, J. (2001), "Integer-Valued, Minimax Robust Designs for Approximately Linear Models With Correlated Errors," *Communications in Statistics: Theory and Methods*, 30, 21-39.