WAVELET DESIGNS FOR NONPARAMETRIC REGRESSION MODELS WITH AUTOCORRELATED ERRORS

SELVAKKADUNKO SELVARATNAM

WAVELET DESIGNS FOR NONPARAMETRIC REGRESSION MODELS WITH AUTOCORRELATED ERRORS

by

©Selvakkadunko Selvaratnam

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirement for the Degree of

Master of Science in Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

St. John's Newfoundland, Canada June 2011

Abstract

We consider minimax designs for estimation of nonparametric regression models using wavelet approximations of the mean response function. We assume that the error terms are autocorrelated. Since the method of estimation depends on the choice of design, we argue that using ordinary least squares method (OLS) for estimation may lead to designs that are less efficient than designs constructed based on generalized least squares (GLS) or weighted least squares (WLS). A simulated annealing algorithm is developed to carry out the minimization problems to search for minimax designs. In this thesis we considered $AR(1)$ model for example. We found that the GLS method is good for the moderate level correlation and WLS or OLS is preferred for highly correlated data.

Acknowledgements

I would like to express my most sincere gratitude to my supervisor Dr. A. J. Oyet for his valuable guidance, constant encouragement, unbelievable patience, helpful suggestions and arrangements of adequate financial support during my program. I would also like to thank Dr. Z. Fan and Dr. A. M. Variyath for their comments and suggestions.

I am also grateful to the School of Graduate Studies and the Department of Mathematics and Statistics for their financial support in the form of Graduate Fellowship and Teaching Assistantship. I would like to thank all staff of the Department of Mathematics and Statistics for providing me with a friendly atmosphere and the necessary facilities to complete my program.

I am very much grateful to my parents, my brother and sisters and specially my wife, Ajitha and my little son, Nithulan for their strong belief in my capabilities and for all of their eternal love, support and encouragement throughout my study.

It is my great pleasure to thank my friends and well - wishers who directly or indirectly encouraged and helped me in the M.Sc program and contributed to this dissertation.

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

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Wavelet analysis has been found to be a powerful tool for the estimation of nonparametric mean response functions. During the 1990s, the nonparametric literature was dominated by (nonlinear) wavelet shrinkage and wavelet thresholding estimators. These estimators are a new subset of an old class of nonparametric regression estimators, namely orthogonal series methods. For more details see Antoniadis, Gregoire and Mckeague (1994)[2], and Wegman, Poston and Solka (1996) [40]. Several authors have used wavelets in nonparametric regression from the **minimax** viewpoint. See for example Donoho, Johnstone, Kerkyacharian and Picard (1995) [9], Oyet and Wiens (2000) [28] and Oyet (2003) [27].

Beginning with Herzberg and Traves (1994) [15], wavelets have continued to play

an important role in robust design of regression experiments. They assumed that the wavelet model was exact and thus applied classical methods for the design. Oyet and Wiens (2000) [28] considered the construction of designs for the estimation of a regression function, when it is anticipated that this function is to be approximated by the dominant terms in its wavelet expansion. They considered both the Haar wavelet basis and the multiwavelet system. The experimenter estimates the coefficients of those wavelets included in the approximation, hoping that the omitted terms will be inconsequential. This introduces bias into the least square estimates, which they proposed handling at the design stage by one of two methods;

(i) Implementing a *minimax robust* design, which enjoys the property of minimizing the maximum value of an *mse*-based loss function, with the maximum being taken over the remainder in the wavelet expansion as it varies over an $\mathbb{L}^2$ - neighborhood;

(ii) Implementing a *minimum variance unbiased* (*mvu*) design which, when employed with weighted least squares and weighs derived here, minimizes the variance subject to a side condition of unbiasedness.

Oyet (2003) [27] discussed the problem of constructing designs in order to maximize the accuracy of nonparametric curve estimation in the possible presence of

heteroscedastic errors. His approach was to exploit the flexibility of wavelet approximations to approximate the unknown response curve by its wavelet expansion thereby eliminating the mathematical difficulty associated with the unknown structure. It is expected that only finitely many parameters in the resulting wavelet response can be estimated by weighted least squares. The bias arising from this, compounds the natural variation of the estimates. Robust minimax designs and weights are then constructed to minimize mean squared error based loss functions of the estimates. He found the periodic and symmetric properties of the Euclidean norm of the multi-wavelet system useful in eliminating some of the mathematical difficulties involved. These properties lead him to restrict the search for robust minimax designs to a specific class of symmetric designs. He also constructed minimum variance unbiased designs and weights, which minimize the loss functions subject to a side condition of unbiasedness.

In this thesis, we propose minimax robust designs for estimation of nonparametric regression models using wavelet approximations of the mean response function. In particular, we use daubechies wavelets to approximate the mean response function. Here we assume that the error terms are autocorrelated. More specifically, we assume that the errors follow an autoregressive process of order 1, $AR(1)$. We allow for repeated observations in the optimal design and assume that there is no correlation

within observations from the same design points, but that the errors are autocorrelated between design points. This leads to a patterned block matrix for the covariance of the error vector. We study two estimators: the weighted least squares estimator (WLSE) using weights proposed by Oyet and Wiens (2000) [28], and the generalized least square estimator (GLSE). We consider the discrete version of Integrated mean squared error as a loss function and simulated annealing algorithm to search for integer-valued robust designs in finite design spaces. Previously, Zhou (2001) [47] had used the OLS method in constructing designs for autocorrelated models. It is well known that the choice of a method of estimation has some effect on the optimal design. Since the errors are autocorrelated we conjecture that it may be possible to find designs that have smaller loss (more efficient) than those of Zhou (2001) [47] by using GLS or WLS method of estimation rather than the OLS method.

There are two main approaches in the literature generally used in the construction of designs for regression experiments. If the regression model is assumed to be exact, classical design theory is commonly applied; whereas, when the model is only an approximation, robust design theory applied. The classical design problem in Section (1.1), the robust design problem in Section (1.2), and some background on wavelets in Section (1.3) are discussed. This is the most extensively studied of discussions and reviews; see Box and Draper (1959) [4], Kiefer (1959) [17], Fedorov (1972) [13], Silvey

(1980) [36], Shah and Sinha (1989) [35], Pukelsheim (1993) [33]. In the area of robust designs with serially correlated observations, there exists an abundant literature. Fang and Wiens (2000) [12] introduced a new approach to construct integer valued, minimax robust designs for approximately linear models with possible heteroscedastic errors. This approach uses a simulated annealing algorithm to search for integer valued, instead of continuous, robust design in finite design spaces. Zhou (2001)[47] applied this approach to construct integer valued, minimax robust design for approximately linear models with possible correlated errors, in particular, with moving average (MA) error processes. Minimax robust designs for approximately linear models with possible correlated errors have been studied by Wiens and Zhou (1996, 1997, 1999)([43], [44] and [45]) and Zhou (2001) [48] in infinite design spaces. Continuous robust designs have been derived for weakly stationary error process, Wiens and Zhou (1996) [43], for the first order MA $[MA(1)]$ process, Zhou (2001) [47], and for the first order autoregressive $[AR(1)]$ process, Wiens and Zhou (1999) [45]). Wiens and Zhou (1997)[44] introduced an *infinitesimal* approach to the construction of robust designs for linear models and used autocorrelated structure for errors. The idea of infinitesimal and minimax approaches were used by Zhou (2001) [48] to construct robust regression designs for linear models with correlated errors, particularly $MA(1)$ process. Recently Wiens and Zhou (2008) [46] and Ou and Zhou (2009) [25]

investigated robust estimators and designs for two-dimensional correlations. Also in that paper, Ou and Zhou (2009) [25] have found the result that the generalized least squares estimators is often more efficient than the least square estimator if the spatial correlation structure belongs to a small neighborhood of a covariance matrix.

1.1 The classical design problem

The standard nonlinear regression model involves observations of $\{\mathbf{x}_i, y(\mathbf{x}_i)\}$ where

$$y(\mathbf{x}_i) = \eta(\mathbf{x}_i; \boldsymbol{\theta}) + \epsilon_i, \ i = 1, 2, \cdots, n. \tag{1.1}$$

In (1.1), $y(\mathbf{x}_i) \in \Re$ is an observable random variable and $\mathbf{x}_i \in S \subseteq \Re$ is the i^{th} vector of some design variables. Thus S is called the design space. Typically S will be continuous but can also be a discrete space. The errors $\epsilon_i \in \Re$ are commonly assumed to be independent and identically distributed with mean zero and common variance σ^2 and $\eta(\mathbf{x}_i; \boldsymbol{\theta})$ is the value of some known square integrable, possibly nonlinear function of $\mathbf{x}_i$ upto the unknown parameter $\boldsymbol{\theta}$.

An experiment can be designed to answer a variety of questions of interest. It is clear that values or levels of inputs must be chosen before running an experiment and observing measurements on a variable of interest. Now, in order to obtain an observation on y, a value for $\mathbf{x}$ must first be selected from S where $\mathbf{x}$ can be set to any

value in S. Given this control over the selection of $\mathbf{x}$, a natural question to consider is at what values of $\mathbf{x}$ should observations y of size say, n, be taken in order to obtain the **best** inference. The **best** selection of $\mathbf{x}$ values or allocation of the n observations to the elements of S is commonly referred to as **an optimal design**.

Classical and robust design theories were developed to determine optimal design points for regression models. The difference between classical and robust design theory arises 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 concept of classical design of regression and optimality principle can be found in Smith (1918) [37], and Plackett and Burman (1946) [32]. Why has this concept come about? When we conduct an experiment, it is common to take time or expenditure into account. The motivation is to choose observations in order that a limited number of observations gives the maximum amount of information. Elfving (1952, 1956) [10], and [11], Chernoff (1953) [5] and others developed this principle of design theory. Kiefer (1959) [17], and Kiefer and Wolfowitz, J. (1959) [18] contributed significantly to the area by extending the previous work. The problem of nonlinear experimental

designs was first considered by Fisher (1922) [14] and White (1973) [41] who proved the nonlinear version of the Kiefer - Wolfowitz equivalence theorem. Though classical design theory was used in nonlinear design, the approach was to approximate the nonlinear function by a linear Taylor series expansion. Thus, nonlinear designs are in fact not optimal in the classical sense since the mean response is not exact. The first comprehensive volume on the theory of approximate continuous optimal experimental design was written by Fedorov (1972) [13], whereas the book by Silvey (1980) [36] gives a very concise description of the theory of optimal design for estimation in linear models. Discrete optimal designs are covered in the book by Shah and Sinba (1989) [35].

In estimation theory for statistical models with one real parameter, the reciprocal of the variance of an efficient estimator is called the "Fisher Information" for that estimator. Because of this reciprocity, minimizing the variance corresponds to maximizing the information. When the statistical model has several parameters, however, the mean of the parameter-estimator is a vector and its variance is a matrix. The inverse of the variance- matrix is called the "information matrix". In this case the problem of "minimizing the variance" becomes complicated. Using statistical theory, statisticians compress the information-matrix using real-valued summary statistics; being real-valued functions, these "information criteria" can be maximized. The

traditional optimality-criteria are invariants of the information matrix; algebraically, the traditional optimality-criteria are functionals of the eigenvalues of the information matrix. In order to apply optimal design theory to (1.1) an information criterion is required for comparing experiments and then selecting the **best** design with respect to the specified criterion. We define $M(\hat{\boldsymbol{\theta}})$ as the mean squared error (MSE) matrix of an estimator of $\boldsymbol{\theta}$ in (1.1). It may be possible to obtain a best inference for all or some of the unknown parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by making some function of the matrix $M(\hat{\boldsymbol{\theta}})$ large in some sense. Therefore, we consider various ways in which to make the matrix $M(\hat{\boldsymbol{\theta}})$ large, namely by maximizing some real valued function $\Phi(\hat{\boldsymbol{\theta}}) = \Psi(M(\hat{\boldsymbol{\theta}}))$. Here, the function Φ is called the criterion function, and in turn, the criterion defined by the function Φ is usually referred to as Φ-optimality criterion. A design maximizing $\Phi(\hat{\boldsymbol{\theta}})$ is said to be a Φ-optimal design. Mathematically, the classical design criteria can be stated as:

$$\min_{\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \in S\}} \Phi(V(\hat{\boldsymbol{\theta}})), \tag{1.2}$$

where $V(\hat{\boldsymbol{\theta}})$ is the covariance of $\hat{\boldsymbol{\theta}}$

The most commonly used criteria in design literature are:

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

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

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

(iv) G - optimality: Here, $\Phi(\cdot) = \max_{t \in T}(\cdot)$, where $\max_{t \in T}(\cdot)$ is the maximization function.

(v) Q - optimality: Here, $\Phi(\cdot) = \int(\cdot)$, where $\int(\cdot)$ is the integral function over a design space, say S.

Hoel (1958) noticed that the D- and G- optimum designs coincide in the model of a one - dimensional polynomial regression, and Kiefer and Wolfowitz (1960)[19] present extensive results on D- and G- optimality, including the Equivalence Theorem. The Equivalence Theorem established that a design is D-optimal if and only if it is G-optimal.

1.2 The robust design problem

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

(i) the model may not be exactly correct;

(ii) the errors ϵ may not be uncorrelated or normally distributed as assumed.

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 $\eta(x, \boldsymbol{\theta}_0)$ with the hope that the remainder terms are negligible. Under these conditions, the least squares estimator of $\boldsymbol{\theta}_0$ is biased and the classical designs which minimize variance alone are no longer "optimal" due to the bias.

Box and Draper (1959) [4] considered a response as polynomial of first degree when the true response was quadratic. They found that the optimal design in typical situations in which both variance and bias occur is nearly the same as would be obtained if *variance were ignored completely and the experiment designed so as to minimize bias alone.* They argued that a more appropriate optimality criterion 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,

$$J = \frac{n\Omega}{\sigma^2} \int_S E\{[\hat{\eta}(x) - \eta(x)]^2\} dx = ISB + IV \tag{1.3}$$

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

$$\Omega^{-1} = \int_S dx, \quad ISB = \frac{n\Omega}{\sigma^2} \int_S \{E[\hat{\eta}(x)] - \eta(x)\}^2 dx \tag{1.4}$$

and

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

Frequently the experimenter fits $E[y|x] = \mathbf{q}^T(x)\boldsymbol{\theta}$ when in fact

$$E[y|x] = \mathbf{q}^T(x)\boldsymbol{\theta}_0 + f(x) \tag{1.6}$$

for some unknown function f. The presence of f implies that, possibly, $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}$. In fact the meaning of the "true" parameter in a model like (1.6) is itself a problem to be addressed. A number of authors, such as Marcus and Sacks (1976) [23], Sacks and Ylvisaker (1978) [34], Pesotchinsky (1982) [30], Li and Notz (1982) [21], and Li (1984) [20] assumed f to be a member of the class of function defined by

$$\{f : |f(x)| \leq \phi(x)\}, \tag{1.7}$$

for a function $\phi(x)$ with specified properties. For example, $\phi(x)$ may be constant, or $\phi(x) = \|x\|^2$. The designs which are "robust" in this class are those which minimize some function of the MSE of $\hat{\boldsymbol{\theta}}$, and are quite sensitive to the choice of ϕ. They also tend to concentrate all mass on extreme points of S.

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 $\hat{\boldsymbol{\theta}}$, as a function of the contamination term f and a design measure ξ. To motivate (1.6), suppose that an experimenter fits the approximate linear model

$$E[y|x] \approx \mathbf{q}^T(x)\boldsymbol{\theta}_0 \tag{1.9}$$

for some parameter vector $\boldsymbol{\theta}_0$. If $E(y|x)$ is only approximately $\mathbf{q}^T(x)\boldsymbol{\theta}_0$ then what is $\boldsymbol{\theta}_0$? The "true" $\boldsymbol{\theta}_0$ is defined such that it makes (1.9) most accurate:

$$\boldsymbol{\theta}_0 = \arg\min \int_S \{E[y|x] - \mathbf{q}^T(x)\boldsymbol{\theta}\}^2 dx. \tag{1.10}$$

Then define

$$f(x) = E[y|x] - \mathbf{q}^T(x)\boldsymbol{\theta}_0. \tag{1.11}$$

Thus our approximate linear model is

$$y(x) = \mathbf{q}^T(x)\boldsymbol{\theta}_0 + f(x) + \epsilon, \quad x \in S \tag{1.12}$$

with

$$\int_S \mathbf{q}(x) f(x) dx = 0. \tag{1.13}$$

Huber (1975) [16] takes $f(x)$ from

$$\mathcal{F} = \left\{ f : \int_S q(x) f(x) dx = 0, \quad \int_S f^2(x) dx \leq \tau^2 \right\} \tag{1.14}$$

where radius τ of $\mathcal{F}$ is assumed fixed. The first condition in $\mathcal{F}$ says that f and $\mathbf{q}$ are orthogonal, so the parameter $\boldsymbol{\theta}$ is uniquely defined in model (1.12).

Wiens (1990) [42] generalized Huber's work from simple linear regression to multiple linear regression. In this thesis, we use Huber's type contamination term in (1.12)

1.2.1 The Robust designs with correlated errors

One approach to the construction of robust designs for linear models with autocorrelated errors was in two stages. These are:

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

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

Following this procedure, Berenblut and Webb (1974) [3] obtained robust D-optimal designs for the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{P}. \tag{1.15}$$

The correlation structure they considered was when $\mathbf{P} = \mathbf{V}(\rho)$, where ρ is the parameter of the first order autoregressive process. Thus, $\mathbf{V}(0) = \mathbf{I}$, the identity matrix. Wiens and Zhou (1996) [43] studied optimal designs for regression models under certain departures from the classical assumptions. They considered the usual formulation of the fixed-regressors linear regression model, which they wrote as

$$y_i = \mathbf{q}^T(x_i)\boldsymbol{\theta}_0 + n^{-\frac{1}{2}}f(x_i) + \epsilon_i, \quad i = 1, 2, \cdots, n \tag{1.16}$$

where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{P}$. Here, $\mathbf{P}$ is a positive semi-definite Toeplitz matrix with unit diagonal, i.e. the autocorrelation matrix of a weakly stationary

process. Then robust minimax designs were constructed by solving the problem

$$\min_{\xi} \max_{|\rho|<1} \max_{f \in \mathcal{F}} \Phi(M(f, \rho, \xi)) \tag{1.17}$$

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

In this thesis, we have applied this criterion and we used the loss function as the Integrated Mean Squared Error (IMSE) of the estimated response. More details of the Integrated Mean Squared Error can be found in Oyet and Wiens (2000) [28], Oyet (2003) [27] and Wiens and Zhou (1999) [45].

1.3 Some background on wavelets

In this section, we discuss some basic knowledge of wavelet systems related to our work. Additional reviews on wavelets and their statistical applications can be found in the papers by Mallat (1989), Meyer (1992), Daubechies (1992) and Härdle et al. (1988), among others.

The concept of wavelets was developed from the oldest and probably best known method, Fourier transform, which was developed in 1807 by Joseph Fourier. Wavelets as an alternative method to Fourier transform was first mentioned by Alfred Harr in 1910.

The development of wavelet theory has in recent years spawned applications in

signal processing, fast algorithms for integral transforms, and image and function representation methods. This last application has stimulated interest in wavelet applications to statistics and to the analysis of experimental data, with many successes in the efficient analysis, processing, and compression of noisy signals and images. See for instance Antoniadis (2007) [1].

In this thesis, we will focus on wavelets in the Hilbert space $\mathcal{L}_2(\Re)$. Additionally we are aiming to find some **simple** functions $\psi_{j,k}$ such that every function $f \in \mathcal{L}_2(\Re)$ has a representation of the type

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) \tag{1.18}$$

for some known coefficients $(d_{j,k})_{j,k\in\mathbb{Z}}$. The motivation comes from the approximation theory. The function f might be difficult to work with, but if such a representation exists, then we can hope that the finite partial sums of (1.18) can approximate f well, ie.

$$f(x) \approx \sum_{|j|\leq M} \sum_{|k|\leq N} d_{j,k}\psi_{j,k}(x). \tag{1.19}$$

By definition, 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

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \tag{1.20}$$

$$\text{and } \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \qquad j,k \in \mathbb{Z} \tag{1.21}$$

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

$$\phi(x) = \sum_{k\in\mathbb{Z}} \sqrt{2} h_k \phi(2x-k) \tag{1.22}$$

$$\text{and } \psi(x) = \sum_{r\in\mathbb{Z}} \sqrt{2} g_r \phi(2x-r), \tag{1.23}$$

$$\text{where } g_r = (-1)^r h_{-r+1} \tag{1.24}$$

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

$$\int \phi(x)dx = 1, \int \psi(x)dx = 1, \int \phi^2(x)dx = 1. \tag{1.25}$$

The condition

$$\sum_{k\in\mathbb{Z}} h_k = \sqrt{2} \tag{1.26}$$

ensures the existence of a unique solution to equations (1.22) and (1.23). Orthogonality of the translation of $\phi(x)$ is ensured by the condition

$$\sum_{k\in\mathbb{Z}} h_k h_{k-2j} = \delta_j; \;\; j \in \mathbb{Z} \tag{1.27}$$

where $\delta_j = \int \phi(x)\phi(x-j)dx$

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

$$\cdots\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots\cdots \tag{1.28}$$

These nested spaces have the properties that their intersection is trivial and their union is dense in $\mathcal{L}_2(\Re)$,

$$\bigcap_j V_j = \{0\}, \qquad \overline{\bigcup_j V_j} = \mathcal{L}_2(\Re). \tag{1.29}$$

Mallat (1989) [22] introduced the notion of a multiresolution analysis, the definition of which we recall here:

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

(a) $\bigcap_j V_j = \{0\}$

(b) $\overline{\bigcup_j V_j} = \mathcal{L}_2(\Re)$

(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

(d) $f(x) \in V_j \Longrightarrow f(x - 2^{-j}k) \in V_j \qquad \forall k \in \mathbb{Z}$

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

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) [22] 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 $\mathcal{L}_2(\Re)$. Thus, the representation in (1.18) is possible.

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

$$W_j \bigoplus V_j = V_{j+1}, \qquad W_j \perp V_j. \tag{1.30}$$

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

$$V_j = V_{j_0} \oplus \bigoplus_{i=j_0}^{j-1} W_i \qquad j > j_0 \in \mathbb{Z} \tag{1.31}$$

Iterating this infinitely many times, we find

$$\mathcal{L}_2(\Re) = V_{j_0} \oplus \bigoplus_{i=j_0}^{\infty} W_i = \bigoplus_{j\in\mathbb{Z}} W_j \tag{1.32}$$

This means that any $f \in \mathcal{L}_2(\Re)$ can be represented as

$$f(x) = \sum_{k\in\mathbb{Z}} c_{j_0k}\phi_{jk}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x) \tag{1.33}$$

or

$$f(x) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) \tag{1.34}$$

where c_{j_0k}, d_{jk} are some coefficients, and $\{\psi_{jk}\}, k \in \mathbb{Z}$ is a basis for W_j. The relation (1.33) 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 Construction of the scaling function ϕ

Strang (1989) [38] provides a brief introduction to the construction of scaling functions using three different approaches. The fourth construction of Daubechies scaling function will be discussed under the section on **some important wavelets**.

The basic dilation equation, described in (1.22) is in fact a two scale difference equation. We know from the condition in (1.26) that $\sum_{k\in\mathbb{Z}} h_k = \sqrt{2}$ ensures the uniqueness of ϕ but a smooth solution is not ensured. For a striking example, set $h_0 = \sqrt{2}$. Then the function ϕ, satisfies $\phi(x) = 2\phi(2x)$.

The dilation of ϕ is unfamiliar (but somehow very pleasing). For other h's, spline functions appear. For example $h_0 = \frac{1}{\sqrt{2}}$, and $h_1 = \frac{1}{\sqrt{2}}$, in this case the function ϕ satisfies $\phi(x) = \phi(2x) + \phi(2x-1)$, then we are able to find the function named as **Box function** and defined by

$$\phi(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

The algorithm for the construction of ϕ are outlined below.

Construction 1: Iterate $\phi_j(x) = \sum \sqrt{2} h_k \phi_{j-1}(2x - k)$ with the box function as $\phi_0(x)$. When $h_0 = \sqrt{2}$ the boxes get taller and thinner, approximating the delta function. For $h_0 = \frac{1}{\sqrt{2}}$, and $h_1 = \frac{1}{\sqrt{2}}$ the box is invariant: $\phi_j = \phi_0$. For $h_0 = \frac{1}{2\sqrt{2}}$, $h_1 = \frac{1}{\sqrt{2}}$, and $h_2 = \frac{1}{2\sqrt{2}}$ the hat function appears as $j \to \infty$, and $\frac{1}{8\sqrt{2}}, \frac{4}{8\sqrt{2}}, \frac{6}{8\sqrt{2}}, \frac{4}{8\sqrt{2}}, \frac{1}{8\sqrt{2}}$ yields the cubic B-spline. An example that will be important (denote this scaling function as D_4) has coefficients $\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}$, and $\frac{1-\sqrt{3}}{4\sqrt{2}}$. This scaling function D_4 leads to orthogonal wavelets. It is not as smooth as it looks. It is shown in Figure 1.1.

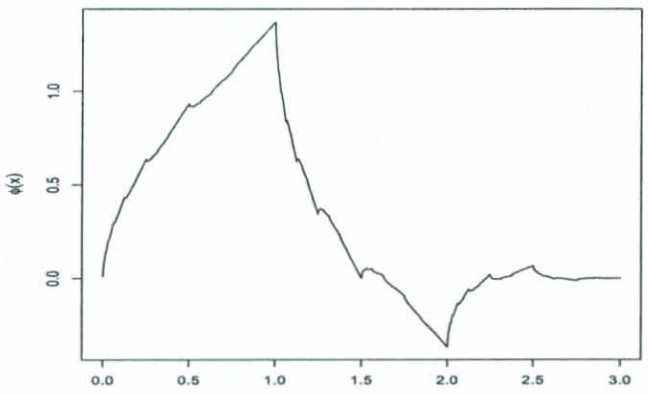

Figure 1.1: The scaling function $D_4(x)$

Construction 2: The second construction begins with the Fourier transform of the scaling function (1.22); given by

$$\Phi(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwx}\phi(x)dx.$$

It can then be shown that, using (1.22), we have

$$\Phi(w) = \frac{1}{\sqrt{2}}\sum h_k e^{-\frac{ikw}{2}}.\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{iyw}{2}}\phi(y)dy.$$

Now, define

$$m_0(w) = \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} h_k e^{-ikw} = \frac{1}{\sqrt{2}}H(w). \tag{1.35}$$

Then we can write

$$\Phi(w) = m_0(\frac{w}{2})\Phi(\frac{w}{2}). \tag{1.36}$$

The function in (1.35) is sometimes called the *transfer function* and it describes the behavior of the associated filter $\underline{h}$ in the Fourier domain. Notice that the function m_0 is periodic with the period 2π and that the filter taps $\{h_n, n \in \mathbb{Z}\}$ are the Fourier coefficients of the function $H(w) = \sqrt{2}m_0(w)$.

By iterating (1.36), one gets

$$\begin{aligned}\Phi(w) &= \left[\prod_{i=1}^{N} m_0\left(\frac{w}{2^i}\right)\right]\Phi(\frac{w}{2^N}) \\ &= \prod_{i=1}^{\infty} m_0(\frac{w}{2^i}).\end{aligned} \tag{1.37}$$

For $h_0 = \sqrt{2}$ we find $m_0 \equiv 1$ and $\Phi = 1$, the transform of the delta function. For $h_0 = h_1 = \frac{1}{\sqrt{2}}$ the products of the m_0's form a geometric series:

$$m_0(\frac{w}{2})m_0(\frac{w}{4}) = \frac{1}{4}(1+e^{-iw/2})(1+e^{-iw/4}) = \frac{1-e^{-iw}}{4(1-e^{-iw/4})}. \tag{1.38}$$

As $N \to \infty$ this approaches the infinite product $\dfrac{(1-e^{-iw})}{(iw)}$, which is the transform of the box function. The hat function comes from squaring $m_0(w)$, which by (1.37) also squares $\Phi(w)$. (Multiplication of m_0's is $\frac{1}{2}$ times convolution of h's.) The cubic B-spline comes from squaring again.

Construction 3: This construction of ϕ works directly with the recursion (1.22). Suppose ϕ is known at the integers $x = j$. The recursion (1.22) gives ϕ at the half-integers. Then it gives ϕ at the quarter-integers, and ultimately at all dyadic points $x = \dfrac{k}{2^j}$. *This is fast to program and all good wavelet calculations use recursion.* The values of ϕ at the integers come from an eigenvector. With the four Daubechies coefficients, set $x = 1$ and $x = 2$ in the dilation equation (1.22) and use the fact that $\phi = 0$ unless $0 < x < 3$:

$$\begin{aligned}
\phi(1) &= \frac{(3+\sqrt{3})}{4\sqrt{2}}\phi(1) + \frac{(1+\sqrt{3})}{4\sqrt{2}}\phi(2) \\
\phi(2) &= \frac{(1-\sqrt{3})}{4\sqrt{2}}\phi(1) + \frac{(3-\sqrt{3})}{4\sqrt{2}}\phi(2)
\end{aligned} \tag{1.39}$$

This is $\boldsymbol{\phi} = \mathbf{L}\boldsymbol{\phi}$, with matrix entries $L_{ij} = h_{2i-j}$, which is the classical eigenvalues problem. Compare with h_{i-j} for an ordinary difference equation. The eigenvalues

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

To repeat for emphasis: From $\phi(1)$ and $\phi(2)$ the recursion gives everything. In this construction the properties of $m_0(w) = \frac{1}{2}\sum \sqrt{2} h_k e^{-ikw}$ are decisive (1.37). The precise hypotheses are in flux, and infinitely many h_k can be allowed.

1.3.2 Some important wavelets

(i) Haar wavelets

The first example of a function f satisfying (1.18) was proposed by Alfred Harr in 1910. The Haar **mother** wavelet is a mathematical function defined by

$$\psi(x) = \begin{cases} 1, & \text{if } 0 \le x < 0.5\ ; \\ -1, & \text{if } 0.5 \le x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

The scaling function $\phi(x)$ can be described as

$$\phi(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1 ; \\ 0, & \text{elsewhere.} \end{cases}$$

The Haar wavelet basis has the following relations:

$$\phi(x) = \phi(2x-1) + \phi(2x) \tag{1.40}$$

$$\psi(x) = \phi(2x) - \phi(2x-1) \tag{1.41}$$

The Haar wavelet is easy to handle. However the only disadvantage is that it is discontinuous at $x = 0$, $x = 0.5$ and $x = 1$.

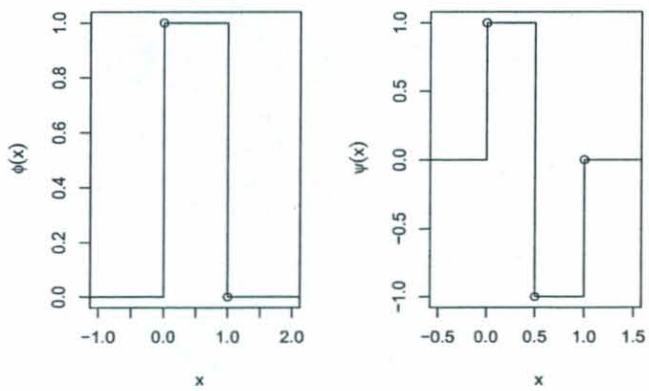

Figure 1.2: The scaling function and wavelet function of Haar wavelets

(ii) Daubechies compactly supported wavelets

One of the most important achievements in wavelet theory was the construction of orthogonal wavelets that were compactly supported but were smoother than Harr wavelets. Daubechies (1988) [6] constructed such wavelets by an ingenious solution of the dilation equation (1.22) that resulted in a family of orthonormal wavelets (several families actually). Each member of each family is indexed by a number N, which refers to the number of vanishing moments.

Daubechies-Lagarias Algorithm

In this section, we describe an algorithm for fast numerical calculation of wavelet values at a given point, based on the Daubechies-Lagarias *local pyramidal algorithm* [7] and [8]. The scaling function and wavelet function in Daubechies families have no explicit representations (except for the Haar wavelet). However, it is often necessary to find their values at arbitrary points.

The Daubechies-Lagarias algorithm enables us to evaluate ϕ and ψ at a point with preassigned precision. We will illustrate the algorithm on wavelets to form the Daubechies family; however, the algorithm works for all finite impulse response quadrature mirror filters.

Let ϕ be the scaling function of the **DAUB**N wavelet. The support of ϕ is

$[0, 2N-1]$. Let $x \in (0,1)$, and let $dyad(x) = \{d_1, d_2, \cdots, d_n, \cdots\}$ be the set of 0-1 digits in the dyadic representation of x ($x = \sum_{j=1}^{\infty} d_j 2^{-j}$). By $dyad(x,n)$, we denote the subset of the first n digits from $dyad(x)$, i.e., $dyad(x,n) = \{d_1, d_2, \cdots, d_n\}$.

Let $\underline{h} = (h_0, h_1, \cdots, h_{2N-1})$ be the wavelet filter coefficients. Define two $(2N-1) \times (2N-1)$ matrices as:

$$\mathbf{T}_0 = (\sqrt{2} \cdot h_{2i-j-1})_{1 \le i,j \le 2N-1} \quad \text{and} \quad \mathbf{T}_1 = (\sqrt{2} \cdot h_{2i-j})_{1 \le i,j \le 2N-1} \tag{1.42}$$

Theorem 1.3.1 [7]

$$\lim_{n \to \infty} \mathbf{T}_{d_1} \cdot \mathbf{T}_{d_2} \cdots \mathbf{T}_{d_n}$$

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

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

Example 1.3.1 Consider the **DAUB2** scaling function ($N = 2$). The corresponding filter is $\underline{h} = (\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}})$. According to (1.42) the matrices $\mathbf{T}_0$ and

$\mathbf{T}_1$ are given as

$$\mathbf{T}_0 = \begin{bmatrix} \frac{1+\sqrt{3}}{4} & 0 & 0 \\ \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} \\ 0 & \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} \end{bmatrix} \text{ and } \mathbf{T}_1 = \begin{bmatrix} \frac{3+\sqrt{3}}{4} & \frac{1+\sqrt{3}}{4} & 0 \\ \frac{1-\sqrt{3}}{4} & \frac{3-\sqrt{3}}{4} & \frac{3+\sqrt{3}}{4} \\ 0 & 0 & \frac{1-\sqrt{3}}{4} \end{bmatrix}$$

Let us evaluate the scaling function at an arbitrary point, say $x = 0.35$. Twenty **decimals** in the dyadic representation of 0.35 are

$dyad(0.35, 20) = \{0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1\}$. In addition to the value at 0.35, we get (for free) the values at 1.35 and 2.35 (the values 0.35, 1.35 and 2.35) are in the domain of ϕ, the interval [0, 3]. The values $\phi(0.35)$, $\phi(1.35)$ and $\phi(2.35)$ may be approximated as averages of the first, second, and third row, respectively in the matrix

$$\prod_{i \in dyad(0.35,20)} \mathbf{T}_i = \begin{bmatrix} 0.7243749 & 0.724373908 & 0.724372881 \\ 0.2672756 & 0.267276667 & 0.267277767 \\ 0.0083495 & 0.008349425 & 0.008349352 \end{bmatrix}$$

The Daubechies-Lagarias algorithm gives only the values of the scaling function. In applications, most of the evaluation needed involves the wavelet function. It turns out that another algorithm is unnecessary, due to the following result.

Theorem 1.3.2 [31] Let x be an arbitrary real number, let the wavelet be given by its filter coefficients, and let $\underline{u}$ with $2N-1$ be a vector defined as

$$\underline{u}(x) = \{(-1)^{1-[2x]} h_{i+1-[2x]}, i = 0, 1, \cdots, 2N-2\}.$$

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

Let the vector $\underline{v}$ be

$$\underline{v}(x,n) = \frac{1}{2N-1} \mathbf{1}' \prod_{i \in dyad(\{2x\},n)} \mathbf{T}_i,$$

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

$$\psi(x) = \lim_{n \to \infty} \underline{u}(x)' \underline{v}(x,n), \tag{1.43}$$

and the limit is constructive.

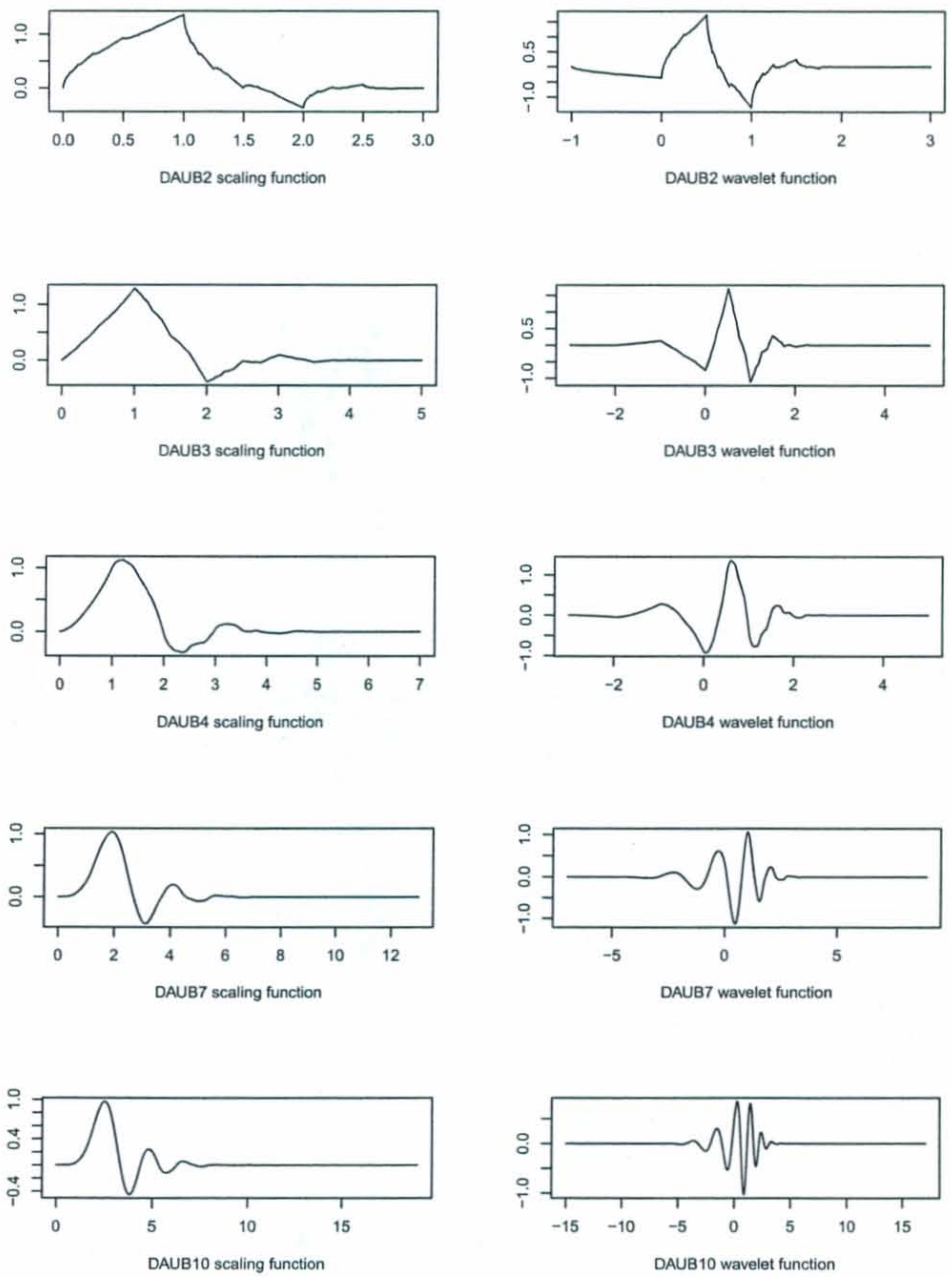

Figure 1.3: Graphs of scaling and wavelets function from Daubechies' family $N = 2$, 3, 4, 7, and 10

Chapter 2

Robust Design Using Weighted Least Squares Estimation

2.1 The model

Zhou (2001) [47] considered repeated observations from a finite design space to construct minimax robust designs for approximately linear models with correlated errors. We discuss designs for the nonparametric regression model with repeated designs

$$y_{ij} = \eta(x_i) + \epsilon_{ij} \tag{2.1}$$

where $i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, n_i; \quad n = \sum_{i=1}^{N} n_i$

The design variable x is assumed to belong to a design space S, and ϵ is random error. In order to investigate the expected response, we may denote the points $x_1, x_2, \cdots, x_N$ to make observations on y. Suppose we need n observations. The integer-valued design problem consists of determination of optimal allocation $n_1, n_2, \cdots, n_N$. If $n_i = 0$, the design point x_i is omitted. If $n_i \geq 1$, the design point x_i is repeated n_i times.

In particular we assume the random error term has the following properties

$$E(\epsilon_{ij}) = 0 \tag{2.2}$$

$$Cov(\epsilon_{ij}, \epsilon_{kl}) = \begin{cases} \sigma^2 & \text{if } i = k \text{ and } j = l, \\ 0 & \text{if } i = k \text{ and } j \neq l, \\ \sigma_{ik} & \text{if } i \neq k. \end{cases} \tag{2.3}$$

where $i, k = 1, 2, \cdots, N$ and $j, l = 1, 2, \cdots, n_i$

Oyet and Wiens (2000) [28] approximated the regression response by finitely many dominant terms of its wavelet representation, with remainder $f(x)$. We transform model (2.1) into an approximately linear model through the wavelet expansion:

$$\eta(x_i) = \mathbf{q}^T(x_i)\boldsymbol{\beta} + f(x_i) \tag{2.4}$$

That is our basic model (2.1) becomes

$$y_{ij} = \mathbf{q}^T(x_i)\boldsymbol{\beta} + f(x_i) + \epsilon_{ij} \tag{2.5}$$

where components of $\mathbf{q}^T(x_i)$ are dilated and translated version of a wavelet system. Since $\mathbf{q}^T(x)$ is made up of components of an orthogonal wavelet system, it is clear that

$$\int_S \mathbf{q}(x)\mathbf{q}^T(x) = I.$$

From orthogonality of wavelet system we obtain

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{q}(x_i)f(x_i) \;=\; 0, \tag{2.6}$$

In order to ensure that the bias in (2.4) is not too large we impose a bound on $f(x)$ defined by

$$\frac{1}{N}\sum_{i=1}^{N} f^2(x_i) \leq \tau^2 \tag{2.7}$$

for some, presumably small, constant τ. Thus errors due to bias will not swamp those due to random variation. We shall see that our results depend on τ and on the error variance σ^2 only through the quantity $\nu = \dfrac{\sigma^2}{\tau^2}$. To the experimenter, this quantity may represent the relative importance of variance versus bias with $\nu = 0$ corresponding to a 'pure bias' problem and $\nu = \infty$ to a 'pure variance' problem.

For simplicity purpose we re-define the equation (2.5) as

$$y_k = \mathbf{q}^T(u_k)\boldsymbol{\beta} + f(u_k) + \epsilon_k \tag{2.8}$$

where $k = 1, 2, \cdots, n$. In (2.8), the values of $u_1, u_2, \cdots, u_{n_1}$ are equal to x_1. Similarly, $u_{n_1+1}, u_{n_1+2}, \cdots, u_{n_1+n_2}$ are equal to x_2 and so on. Further, $y_1, y_2, \cdots, y_{n_1}$ are equal to $y_{11}, y_{12}, \cdots, y_{1n_1}$ respectively. Also $y_{n_1+1}, y_{n_1+2}, \cdots, y_{n_1+n_2}$ are equal to $y_{21}, y_{22}, \cdots, y_{2n_2}$ respectively, and so on. Now, define

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ \cdots \\ y_n \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{q}^T(u_1) \\ \mathbf{q}^T(u_2) \\ \cdots \\ \cdots \\ \mathbf{q}^T(u_n) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \cdots \\ \beta_r \end{bmatrix}, \quad \mathbf{f}_u = \begin{bmatrix} f(u_1) \\ f(u_2) \\ \cdots \\ \cdots \\ f(u_n) \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \cdots \\ \epsilon_n \end{bmatrix}$$

In matrix notations, we can re-write the model (2.8) as

$$\mathbf{Y} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{f}_u + \boldsymbol{\epsilon}, \tag{2.9}$$

where $E(\boldsymbol{\epsilon}) = \mathbf{0}_{n\times 1}$ and $Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma} = \sigma^2\mathbf{R}$. It can be shown that the correlation matrix $\mathbf{R}$ is a patterned block matrix of the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}_{n_1} & \rho_{12}\mathbf{J}_{n_1,n_2} & \rho_{13}\mathbf{J}_{n_1,n_3} & \cdots & \rho_{1N}\mathbf{J}_{n_1,n_N} \\ \rho_{21}\mathbf{J}_{n_2,n_1} & \mathbf{I}_{n_2} & \rho_{23}\mathbf{J}_{n_2,n_3} & \cdots & \rho_{2N}\mathbf{J}_{n_2,n_N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{N1}\mathbf{J}_{n_N,n_1} & \rho_{N2}\mathbf{J}_{n_N,n_2} & \rho_{N3}\mathbf{J}_{n_N,n_3} & \cdots & \mathbf{I}_{n_N} \end{bmatrix} \tag{2.10}$$

where,

$$Corr(\epsilon_{ij}, \epsilon_{kl}) = \begin{cases} 1 & \text{if } i = k \text{ and } j = l, \\ 0 & \text{if } i = k \text{ and } j \neq l, \\ \rho_{ik} & \text{if } i \neq k. \end{cases} \tag{2.11}$$

with $i, k = 1, 2, \cdots, N$ and $j, l = 1, 2, \cdots, n_i$. The constant matrix $\mathbf{J}_{n_i,n_j}$ in (2.10) is of dimension $n_i \times n_j$ and is defined by

$$\mathbf{J}_{n_i,n_j} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix},$$

and $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix.

Since the method of estimation influence the choice of a design an experimenter needs to decide on method of estimation prior to constructing the design. In this section, we will be using the wavelet version of weighted least squares estimate. The weights we shall use are given in the following theorem.

Theorem 2.1.1 *(Oyet and Wiens (2000) [28]) The minimum variance unbiased (mvu) design under the model (2.8) and integrated mean squared error loss has density* $k_0(x) = v_0(x)^{-1}$, *where the mvu weights are (proportional to)*

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

Thus, if we define $w_1, w_2, \cdots, w_N$ as weights related to $x_1, x_2, \cdots, x_N$ respectively, then, in general w_i can be computed as follows;

$$w_i = \frac{\int_0^1 \parallel \mathbf{q}(x) \parallel dx}{\parallel \mathbf{q}(x_i) \parallel}.$$

We shall denote by w_k^* the weights corresponding to the variable u_k, $k = 1, 2, \cdots, n$. Now, by defining the matrix $\mathbf{W}_u$

$$\mathbf{W}_u = \begin{bmatrix} w_1^* & 0 & 0 & \cdots & 0 \\ 0 & w_2^* & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & w_n^* \end{bmatrix}.$$

We can write the weighted least square estimate of the parameter $\boldsymbol{\beta}$ as

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left[\sum_{i=1}^{n} w_i^* \mathbf{q}(u_i)\mathbf{q}^T(u_i)\right]^{-1} \left(\sum_{i=1}^{n} w_i^* \mathbf{q}(u_i) y_i\right) \\ &= \left(\mathbf{Q}^T \mathbf{W}_u \mathbf{Q}\right)^{-1} \mathbf{Q}^T \mathbf{W}_u \mathbf{Y}. \end{aligned} \tag{2.12}$$

Oyet and Wiens (2002) [29] amongst others have used the Average Mean Squared Error ($AMSE$) as a design criterion. This measure of loss is the discrete version of the Integrated Mean Squared Error ($IMSE$). The $AMSE$ is defined as

$$\begin{aligned} AMSE &= \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\hat{Y}_{(x_i)} - E(Y|x_i)\right]^2\right\} \\ &= \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\hat{Y}_{(x_i)} - E(\hat{Y}_{(x_i)})\right]^2\right\} + \frac{1}{N}\sum_{i=1}^{N}\left\{E(\hat{Y}_{(x_i)}) - E(Y|x_i)\right\}^2 \\ &= AV + ASB \end{aligned} \tag{2.13}$$

where, $AV = \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\hat{Y}_{(x_i)} - E(\hat{Y}_{(x_i)})\right]^2\right\}$ and $ASB = \frac{1}{N}\sum_{i=1}^{N}\left\{E(\hat{Y}_{(x_i)}) - E(Y|x_i)\right\}^2$.

The Average Variance is abbreviated to AV and ASB is the Average Squared Bias.

We have also used the $AMSE$ as a design criterion. In what follows, we will derive expression for the AV and the ASB.

2.2 The Average Mean Squared Error

It is clear, from (2.13) that the $AMSE$ has two components namely, the AV and ASB. In order to derive an expression for the $AMSE$ we first need to derive the AV and the ASB. Now, by taking expectation of $\hat{\boldsymbol{\beta}}$ in (2.12) we obtain

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E\left\{\left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u\mathbf{Y}\right\} \\ &= \left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u E(\mathbf{Y}) \\ &= \left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u\left[\mathbf{Q}\boldsymbol{\beta}+\mathbf{f}_u\right] \\ &= \boldsymbol{\beta}+\left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u\mathbf{f}_u \end{aligned}$$

Thus, the bias of $\hat{\boldsymbol{\beta}}$ denoted by $\mathbf{d}$ can be written as

$$\begin{aligned}
\mathbf{d} &= bias(\hat{\boldsymbol{\beta}}) \\
&= E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} \\
&= \left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u\mathbf{f}_u \\
&= \left[\sum_{i=1}^{n} w_i^* \mathbf{q}(u_i)\mathbf{q}^T(u_i)\right]^{-1} \sum_{i=1}^{n} w_i^* f(u_i)\mathbf{q}(u_i) \\
&= \left[\sum_{i=1}^{N} n_i w_i \mathbf{q}(x_i)\mathbf{q}^T(x_i)\right]^{-1} \sum_{i=1}^{N} n_i w_i f(x_i)\mathbf{q}(x_i) \\
&= (\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{f}_N), \qquad (2.14)
\end{aligned}$$

where $\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & w_N \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{q}^T(x_1) \\ \mathbf{q}^T(x_2) \\ \cdots \\ \cdots \\ \mathbf{q}^T(x_N) \end{bmatrix}$, $\mathbf{f}_N = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \cdots \\ \cdots \\ f(x_N) \end{bmatrix}$ and

$$\mathbf{M} = \begin{bmatrix} n_1 & 0 & 0 & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & n_N \end{bmatrix}.$$

Now, we note that in actual computations the estimate of Y will be $\hat{Y}_{(x)} = \mathbf{q}^T(x)\hat{\boldsymbol{\beta}}$.

This implies that the ASB in (2.13) can be simplified to

$$\begin{aligned} ASB &= \frac{1}{N}\sum_{i=1}^{N}\left\{E(\hat{Y}_{(x_i)}) - E(Y|x_i)\right\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{q}_{(x_i)}^T E(\hat{\boldsymbol{\beta}}) - \mathbf{q}_{(x_i)}^T\boldsymbol{\beta} - f(x_i)\right\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{q}_{(x_i)}^T(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})\right\}^2 - \frac{2}{N}\sum_{i=1}^{N}\mathbf{q}_{(x_i)}^T(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})f(x_i) + \frac{1}{N}\sum_{i=1}^{N}f^2(x_i) \end{aligned}$$

Using the orthogonality condition in (2.6) we find that

$$\begin{aligned} \sum_{i=1}^{N}\mathbf{q}_{(x_i)}^T(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})f(x_i) &= \sum_{i=1}^{N}\mathbf{q}_{(x_i)}^T f(x_i)(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}) \\ &= 0. \end{aligned}$$

Consequently, the ASB reduces to

$$\begin{aligned} ASB &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{q}_{(x_i)}^T(E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})\right\}^2 + \frac{1}{N}\sum_{i=1}^{N}f^2(x_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right]^T\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T\left[E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right] \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}f^2(x_i) \end{aligned}$$

It follows from (2.14) that

$$\begin{aligned} ASB &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{d}^T\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T\mathbf{d} + \frac{1}{N}\sum_{i=1}^{N}f^2(x_i) \\ &= \frac{1}{N}\mathbf{d}^T\left(\sum_{i=1}^{N}\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T\right)\mathbf{d} + \frac{1}{N}\sum_{i=1}^{N}f^2(x_i) \\ &= \frac{1}{N}\mathbf{d}^T\mathbf{Z}^T\mathbf{Z}\mathbf{d} + \frac{1}{N}\parallel \mathbf{f}_N \parallel^2 . \end{aligned} \tag{2.15}$$

Having obtained an expression for the ASB we now proceed to derive an expression for the AV. For this purpose, we note that the variance of $\hat{\boldsymbol{\beta}}$ can be written as

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= Var\left\{\left(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{W}_u\mathbf{Y}\right\} \\ &= (\mathbf{Q}^T\mathbf{W}_u\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}_u Var(\mathbf{Y})\mathbf{W}_u\mathbf{Q}(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q})^{-1} \\ &= \sigma^2(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{W}_u\mathbf{R}\mathbf{W}_u\mathbf{Q}(\mathbf{Q}^T\mathbf{W}_u\mathbf{Q})^{-1}. \end{aligned}$$

Then, using kronecker products we find that

$$\begin{aligned} \mathbf{Q}^T\mathbf{W}_u &= [w_1^*\mathbf{q}(u_1),\ w_2^*\mathbf{q}(u_2),\ \cdots\cdots,\ w_n^*\mathbf{q}(u_n)] \\ &= [w_1(\mathbf{1}_{n_1}^T \otimes \mathbf{q}(x_1)),\ w_2(\mathbf{1}_{n_2}^T \otimes \mathbf{q}(x_2)),\ \cdots\cdots,\ w_N(\mathbf{1}_{n_N}^T \otimes \mathbf{q}(x_N))] \end{aligned}$$

where, $\mathbf{1}_{n_i}^T = (1, 1, \cdots, 1)_{1\times n_i}$, $i = 1, 2, \cdots, N$, and

$$
\begin{aligned}
\mathbf{Q}^T\mathbf{W}_u\mathbf{R} \;=\; & [w_1(\mathbf{1}_{n_1}^T \otimes \mathbf{q}(x_1)) + \sum_{(j=1)\ \&\ (j\neq 1)}^{N} \rho_{j1} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))\mathbf{J}_{n_j,n_1}, \\
& w_2(\mathbf{1}_{n_2}^T \otimes \mathbf{q}(x_2)) + \sum_{(j=1)\ \&\ (j\neq 2)}^{N} \rho_{j2} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))\mathbf{J}_{n_j,n_2}, \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\
& w_N(\mathbf{1}_{n_N}^T \otimes \mathbf{q}(x_N)) + \sum_{(j=1)\ \&\ (j\neq N)}^{N} \rho_{jN} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))\mathbf{J}_{n_j,n_N}].
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbf{Q}^T\mathbf{W}_u\mathbf{R}\mathbf{W}_u\mathbf{Q} \;=\; & w_1^2(\mathbf{1}_{n_1}^T \otimes \mathbf{q}(x_1))(\mathbf{1}_{n_1} \otimes \mathbf{q}^T(x_1)) \\
& + \; w_1 \sum_{(j=1)\ \&\ (j\neq 1)}^{N} \rho_{j1} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))(\mathbf{J}_{n_j,n_1} \otimes 1)(\mathbf{1}_{n_1} \otimes \mathbf{q}^T(x_1)) \\
& + \; w_2^2(\mathbf{1}_{n_2}^T \otimes \mathbf{q}(x_2))(\mathbf{1}_{n_2} \otimes \mathbf{q}^T(x_2)) \\
& + \; w_2 \sum_{(j=1)\ \&\ (j\neq 2)}^{N} \rho_{j2} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))(\mathbf{J}_{n_j,n_2} \otimes 1)(\mathbf{1}_{n_2} \otimes \mathbf{q}^T(x_2)) \\
& + \; \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + \; \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + \; w_N^2(\mathbf{1}_{n_N}^T \otimes \mathbf{q}(x_N))(\mathbf{1}_{n_N} \otimes \mathbf{q}^T(x_N)) \\
& + \; w_N \sum_{(j=1)\ \&\ (j\neq N)}^{N} \rho_{jN} w_j (\mathbf{1}_{n_j}^T \otimes \mathbf{q}(x_j))(\mathbf{J}_{n_j,n_N} \otimes 1)(\mathbf{1}_{n_N} \otimes \mathbf{q}^T(x_N))
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{Q}^T\mathbf{W}_u\mathbf{R}\mathbf{W}_u\mathbf{Q} \;=\; & \sum_{i=1}^{N} n_i w_i^2 \mathbf{q}(x_i)\mathbf{q}^T(x_i) \\
& + \sum_{(j=1)\ \&\ (j\neq 1)}^{N} \rho_{j1} w_j w_1 n_j n_1 \mathbf{q}(x_j)\mathbf{q}^T(x_1) \\
& + \sum_{(j=1)\ \&\ (j\neq 2)}^{N} \rho_{j2} w_j w_2 n_j n_2 \mathbf{q}(x_j)\mathbf{q}^T(x_2) \\
& + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& + \sum_{(j=1)\ \&\ (j\neq N)}^{N} \rho_{jN} w_j w_N n_j n_N \mathbf{q}(x_j)\mathbf{q}^T(x_N).
\end{aligned}
$$

Now define $\mathbf{R}^*$ as follows

$$
\mathbf{R}^* = \begin{bmatrix}
0 & \rho_{12} & \rho_{13} & \cdots & \rho_{1N} \\
\rho_{21} & 0 & \rho_{23} & \cdots & \rho_{2N} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\rho_{N1} & \rho_{N2} & \rho_{N3} & \cdots & 0
\end{bmatrix}. \tag{2.16}
$$

Then,

$$\begin{aligned}
\mathbf{Q}^T\mathbf{W}_u\mathbf{R}\mathbf{W}_u\mathbf{Q} &= \mathbf{Z}^T\mathbf{M}\mathbf{W}^2\mathbf{Z} \\
&\quad + [n_1w_1\mathbf{q}(x_1),\ n_2w_2\mathbf{q}(x_2),\ \cdots,\ n_Nw_N\mathbf{q}(x_N)]\mathbf{R}^* \begin{bmatrix} n_1w_1\mathbf{q}(x_1) \\ n_2w_2\mathbf{q}(x_2) \\ \cdots \\ \cdots \\ n_Nw_N\mathbf{q}(x_N) \end{bmatrix} \\
&= \mathbf{Z}^T\mathbf{M}\mathbf{W}^2\mathbf{Z} + \mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{R}^*\mathbf{W}\mathbf{M}\mathbf{Z}.
\end{aligned}$$

We already know that

$$\mathbf{Q}^T\mathbf{W}_u\mathbf{Q} = \mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{Z}.$$

So that by combining these results, $Var(\hat{\boldsymbol{\beta}})$ can be written as

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2[\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{Z}]^{-1}(\mathbf{Z}^T\mathbf{M}\mathbf{W}^2\mathbf{Z} + \mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{R}^*\mathbf{W}\mathbf{M}\mathbf{Z})[\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{Z}]^{-1} \quad (2.17)$$

The variance of $\hat{\boldsymbol{\beta}}$ depends on $\mathbf{R}^*$ but not $\mathbf{R}$. The matrix $\mathbf{R}^*$ is easier to handle compared to $\mathbf{R}$. Applying some algebra to the expression in (2.13), we can simplify

the AV in the following way.

$$\begin{aligned}
AV &= \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\hat{Y}_{(x_i)} - E(\hat{Y}_{(x_i)})\right]^2\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\mathbf{q}_{(x_i)}^T\hat{\boldsymbol{\beta}} - \mathbf{q}_{(x_i)}^T E(\hat{\boldsymbol{\beta}})\right]^2\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left\{\left[\mathbf{q}_{(x_i)}^T(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))\right]^2\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left\{(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))^T\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} traceE\left\{\mathbf{q}_{(x_i)}^T(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))^T\mathbf{q}_{(x_i)}\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} trace\left[\mathbf{q}_{(x_i)}^T E\left\{(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))^T\right\}\mathbf{q}_{(x_i)}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} trace\left\{\mathbf{q}_{(x_i)}^T Var(\hat{\boldsymbol{\beta}})\mathbf{q}_{(x_i)}\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} trace\left\{Var(\hat{\boldsymbol{\beta}})\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T\right\} \\
&= \frac{1}{N} trace\left\{\sum_{i=1}^{N} Var(\hat{\boldsymbol{\beta}})\mathbf{q}_{(x_i)}\mathbf{q}_{(x_i)}^T\right\}. \\
\text{Thus, AV} &= \frac{1}{N} trace\left\{\mathbf{Z}Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T\right\}, \qquad (2.18)
\end{aligned}$$

where $Var(\hat{\boldsymbol{\beta}})$ is given by (2.17). We assume that the $N \times p$ matrix $\mathbf{Z}$ is of full rank, so that the singular value decomposition (SVD) of $\mathbf{Z}$ is

$$\mathbf{Z}_{N\times p} = \mathbf{U}_{N\times p}\,\boldsymbol{\Lambda}_{p\times p}\,\mathbf{V}_{p\times p}^T \qquad (2.19)$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices such that $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_p$, and $\boldsymbol{\Lambda}$ is in

(2.19) a diagonal matrix of singular values of $\mathbf{Z}$. Using the SVD of $\mathbf{Z}$, we can now write $\mathbf{Z}^T\mathbf{MWZ} = \mathbf{V\Lambda U}^T\mathbf{MWU\Lambda V}^T$

for simplicity of notation we now define, $\mathbf{M}_1 = \mathbf{U}^T\mathbf{MWU}$ so that

$$\begin{aligned} \mathbf{Z}^T\mathbf{MWZ} &= \mathbf{V\Lambda M}_1\mathbf{\Lambda V}^T \\ (\mathbf{Z}^T\mathbf{MWZ})^{-1} &= \mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \end{aligned} \tag{2.20}$$

Since $\mathbf{V}^{-1} = \mathbf{V}^T$, it follows that the variance of $\hat{\boldsymbol{\beta}}$ can be written as,

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\left[\mathbf{V\Lambda U}^T\mathbf{MW}^2\mathbf{U\Lambda V}^T\right]\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &\quad + \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\left[\mathbf{V\Lambda U}^T\mathbf{MWR}^*\mathbf{WMU\Lambda V}^T\right]\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &= \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{V\Lambda U}^T\mathbf{MW}^2\mathbf{U\Lambda V}^T\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &\quad + \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{V\Lambda U}^T\mathbf{MWR}^*\mathbf{WMU\Lambda V}^T\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &= \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{MW}^2\mathbf{UM}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &\quad + \sigma^2\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{MWR}^*\mathbf{WMUM}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T \end{aligned}$$

Now, let us consider the argument in the trace function in (2.18),

$$\begin{aligned} \mathbf{Z}Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T &= \sigma^2\mathbf{U\Lambda V}^T\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{MW}^2\mathbf{UM}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{V\Lambda U}^T \\ &\quad + \sigma^2\mathbf{U\Lambda V}^T\mathbf{V\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{MWR}^*\mathbf{WMUM}_1^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{V\Lambda U}^T \\ &= \sigma^2\mathbf{UM}_1^{-1}\mathbf{U}^T\mathbf{MW}^2\mathbf{UM}_1^{-1}\mathbf{U}^T \\ &\quad + \sigma^2\mathbf{UM}_1^{-1}\mathbf{U}^T\mathbf{MWR}^*\mathbf{WMUM}_1^{-1}\mathbf{U}^T. \end{aligned}$$

If we define $\mathbf{A} = \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}$, then

$$\mathbf{Z}Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T = \sigma^2\left\{\mathbf{U}\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{U}^T + \mathbf{U}\mathbf{A}\mathbf{R}^*\mathbf{A}^T\mathbf{U}^T\right\}, \tag{2.21}$$

and the following expression for the AV follow immediately from (2.18) and (2.21)

$$\begin{aligned} AV &= \frac{1}{N}trace\left[\mathbf{Z}Var(\hat{\boldsymbol{\beta}})\mathbf{Z}^T\right] \\ &= \frac{\sigma^2}{N}\left\{trace\left[\mathbf{U}\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{U}^T\right] + trace\left[\mathbf{U}\mathbf{A}\mathbf{R}^*\mathbf{A}^T\mathbf{U}^T\right]\right\} \\ &= \frac{\sigma^2}{N}\left\{trace\left[\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\right] + trace\left[\mathbf{A}\mathbf{R}^*\mathbf{A}^T\right]\right\}. \end{aligned} \tag{2.22}$$

Concerning the ASB, we first note that the bias of $\hat{\boldsymbol{\beta}}$ which we denoted by $\mathbf{d}$ can also be written in terms of the SVD of $\mathbf{Z}$ as

$$\begin{aligned} \mathbf{d} &= (\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{f}_N) \\ &= \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{M}_1^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{f}_N \\ &= \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{f}_N. \end{aligned}$$

Using the expression for $\mathbf{d}$ in the ASB in (2.15) we obtain

$$\begin{aligned} ASB &= \frac{1}{N}\mathbf{f}_N^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{f}_N \\ &\quad + \frac{1}{N}\mathbf{f}_N^T\mathbf{f}_N \\ &= \frac{1}{N}\mathbf{f}_N^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{f}_N + \frac{1}{N}\mathbf{f}_N^T\mathbf{f}_N \\ &= \frac{1}{N}\mathbf{f}_N^T\left\{\mathbf{I}_N + \mathbf{A}^T\mathbf{A}\right\}\mathbf{f}_N \end{aligned} \tag{2.23}$$

where $\mathbf{A}$ is given in (2.21). Thus the $AMSE$ in (2.13) becomes

$$AMSE = \frac{1}{N}\mathbf{f}_N^T\mathbf{A}^T\mathbf{A}\mathbf{f}_N + \frac{1}{N}\mathbf{f}_N^T\mathbf{f}_N + \frac{\sigma^2}{N}\left\{trace\left[\mathbf{AWUM}_1^{-1}\right] + trace\left[\mathbf{AR}^*\mathbf{A}^T\right]\right\}, \quad (2.24)$$

which is a function of the unknown vector $\mathbf{f}_N$ and unknown matrix $\mathbf{R}^*$. Thus, in constructing the design based on the $AMSE$ as a criterion, we will use the MINIMAX approach.

2.3 Exact Minimax Robust Design

In this section we first maximize the $AMSE$ over $\mathbf{f}_N$ such that to the constraints (2.6) and (2.7). We will also obtain the maximum of the $AMSE$ over $\mathbf{R}^*$. Let $\tilde{\mathbf{U}}_{N\times(N-p)}$ be a matrix whose columns constitute an orthonormal basis for the orthogonal complement of the column space of $\mathbf{U}$, $[col(\mathbf{U})]^{\perp}$. Thus, $\tilde{\mathbf{U}}^T\mathbf{U} = 0$, and $\tilde{\mathbf{U}}^T\tilde{\mathbf{U}} = \mathbf{I}_{N-p}$. Now augment the columns of $\mathbf{U}$ and $\tilde{\mathbf{U}}$ to obtain the matrix $\mathbf{U}^*_{N\times N} = \left[\mathbf{U}_{N\times p} \ : \ \tilde{\mathbf{U}}_{N\times(N-p)}\right]$ such that

$$\begin{aligned}
\mathbf{U}^{*T}\mathbf{U}^* &= \begin{bmatrix}\mathbf{U}^T \\ \tilde{\mathbf{U}}^T\end{bmatrix}\begin{bmatrix}\mathbf{U} & \tilde{\mathbf{U}}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{U}^T\mathbf{U} & \mathbf{U}^T\tilde{\mathbf{U}} \\ \tilde{\mathbf{U}}^T\mathbf{U} & \tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\end{bmatrix} \\
&= \mathbf{I}_{N\times N}. \qquad (2.25)
\end{aligned}$$

Clearly, the condition (2.6) can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}(x_i)f(x_i) = \frac{1}{N}\mathbf{Z}^T\mathbf{f}_N = \frac{1}{N}\mathbf{V}\mathbf{\Lambda}\mathbf{U}^T\mathbf{f}_N = 0.$$

This, implies that, the function $\mathbf{f}_N$ belongs to the space defined by the orthogonal compliment of $\mathbf{U}$. That is,

$$\mathbf{f}_N \in \{Col(\mathbf{U})\}^{\perp}. \tag{2.26}$$

Since the columns of $\tilde{\mathbf{U}}$ is a basis for the space $\{Col(\mathbf{U})\}^{\perp}$ we can express $\mathbf{f}_N$ as a linear combination of columns of $\tilde{\mathbf{U}}$. Thus, we can write $\mathbf{f}_N = \tilde{\mathbf{U}}\boldsymbol{\alpha}$

Now, the condition (2.7) given by

$$\frac{1}{N}\sum_{i=1}^{N} f^2(x_i) = \frac{1}{N}\mathbf{f}_N^T\mathbf{f}_N \leq \tau^2.$$

That is,

$$\mathbf{f}_N^T\mathbf{f}_N \leq N\tau^2.$$

Thus, $\exists$ some $\mathbf{c}$ satisfying $\| \mathbf{c} \|^2 \leq 1$ such that

$$\| \mathbf{f}_N \|^2 = N\tau^2 \| \mathbf{c} \|^2$$

with equality if $\| \mathbf{c} \| = 1$. It follows that the maximum value of $\mathbf{f}_N$ is achieved when $\| \mathbf{c} \| = 1$. Thus, we have $\mathbf{f}_N = \sqrt{N}\tau\mathbf{c}$. Since $\tilde{\mathbf{U}}$ is an orthonormal transformation

matrix it does not alter the length of a vector. Thus, we have maximum of $\mathbf{f}_N$ over $\mathfrak{F}$ is

$$\mathbf{f}_N = \sqrt{N}\tau\tilde{\mathbf{U}}\mathbf{c} \quad \text{provided that} \quad \| \mathbf{c} \| = 1. \tag{2.27}$$

It is clear that the $AMSE$ depends on $\mathbf{f}_N$ only through the Average Square Bias (ASB). Thus

$$\max_{\mathbf{f}_N \in \mathfrak{F}} AMSE = \max_{\mathbf{f}_N \in \mathfrak{F}} ASB + AV$$

Using (2.27) we write

$$\begin{aligned} ASB &= \frac{1}{N} N\tau^2 \mathbf{c}^T \left\{ \tilde{\mathbf{U}}^T\tilde{\mathbf{U}} + \tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}} \right\} \mathbf{c} \\ &= \tau^2 \mathbf{c}^T \left\{ \mathbf{I}_{N-p} + \tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}} \right\} \mathbf{c}. \end{aligned}$$

Since $\tilde{\mathbf{U}}^T\tilde{\mathbf{U}} = \mathbf{I}_{N-p}$. Now, define

$$\mathbf{G} = \mathbf{I}_{N-p} + \tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}}. \tag{2.28}$$

Then ASB becomes

$$ASB = \tau^2 \mathbf{c}^T \mathbf{G} \mathbf{c}.$$

We maximize ASB with respect to $\mathbf{f}_N$ which is equivalent to maximize with respect to $\mathbf{c}$ subject to the constraint $\| \mathbf{c} \| = 1$. That is

$$\begin{aligned} \max_{\mathbf{f}_N} ASB &= \max_{\mathbf{f}_N} \tau^2 \mathbf{c}^T \mathbf{G} \mathbf{c} \\ &= \max_{\|\mathbf{c}\|=1} \tau^2 \mathbf{c}^T \mathbf{G} \mathbf{c}. \end{aligned} \tag{2.29}$$

Now, using the symmetric decomposition of $\mathbf{G}$ we can write

$$\mathbf{c}^T\mathbf{G}\mathbf{c} = \mathbf{c}^T \sum_{j=1}^{N-p} \lambda_j(\mathbf{G})\mathbf{e}_j\mathbf{e}_j^T\mathbf{c}$$

where $\mathbf{e}_j$ are normalized eigenvectors of $\mathbf{G}$. Clearly,

$$\begin{aligned} \mathbf{c}^T\mathbf{G}\mathbf{c} &\leq \lambda_{max}(\mathbf{G})\mathbf{c}^T \sum_{j=1}^{N-p} \mathbf{e}_j\mathbf{e}_j^T\mathbf{c} \\ &= \lambda_{max}(\mathbf{G})\mathbf{c}^T\mathbf{c} \\ &= \lambda_{max}(\mathbf{G}), \end{aligned}$$

since $\| \mathbf{c} \|^2 = 1$. It follows that

$$\max_{\mathbf{f}_N} ASB = \tau^2\lambda_{max}(\mathbf{G}), \tag{2.30}$$

where $\lambda_{max}(\mathbf{G})$ is the maximum eigenvalue of the matrix $\mathbf{G}$. We use properties of matrices and eigenvalues to simplify $\lambda_{max}(\mathbf{G})$ as follows:

$$\begin{aligned} \lambda_{max}(\mathbf{G}) &= \lambda_{max}\left[\mathbf{I}_{N-p} + \tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}}\right] \\ &= \lambda_{max}\left[\mathbf{I}_p + \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\right] \end{aligned}$$

We know that $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I}_N - \mathbf{U}\mathbf{U}^T$. As a result, $\lambda_{max}(\mathbf{G})$ becomes

$$\begin{aligned}
\lambda_{max}(\mathbf{G}) &= \lambda_{max}\left[\mathbf{I}_p + \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{I}_p + \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}(\mathbf{I}_N - \mathbf{U}\mathbf{U}^T)\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{I}_p + \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}^2\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1} - \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{W}\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{I}_p + \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}^2\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1} - \mathbf{M}_1^{-1}\mathbf{M}_1\mathbf{M}_1\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}^2\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{U}^T(\mathbf{M}\mathbf{W})^2\mathbf{U}\mathbf{M}_1^{-1}\right] \\
&= \lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right]. \qquad (2.31)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\max_{\mathbf{f}_N} AMSE &= \tau^2\lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right] + \frac{\sigma^2}{N}\left\{tr\left[\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\right] + tr\left[\mathbf{A}\mathbf{R}^*\mathbf{A}^T\right]\right\} \\
&= \tau^2\left\{\lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right] + \upsilon tr\left[\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\right] + \upsilon tr\left[\mathbf{A}\mathbf{R}^*\mathbf{A}^T\right]\right\} \qquad (2.32)
\end{aligned}$$

where $\upsilon = \dfrac{\sigma^2}{N\tau^2}$.

For completeness, we now have to maximize (2.32) with respect to possible values of the correlation in the matrix $\mathbf{R}^*$.

$$\begin{aligned}
\max_{\mathbf{R}^*}\max_{\mathbf{f}_N} AMSE &= \tau^2\lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right] \\
&\quad + \frac{\nu\tau^2}{N}\left(trace\left[\mathbf{A}\mathbf{W}\mathbf{U}\mathbf{M}_1^{-1}\right] + \max_{\mathbf{R}^*} trace\left[\mathbf{A}\mathbf{R}^*\mathbf{A}^T\right]\right) \qquad (2.33)
\end{aligned}$$

The equation (2.33) is true for any $\mathbf{R}^*$, defined in (2.16) and the correlation matrix $\mathbf{R}$ is defined in (2.10).

In what follows, we maximize (2.33) over $\mathbf{R}^*$ and minimize it over the design space numerically. Previously, Zhou (2001) [47] had found, theoretically, that when the error terms in an approximately linear regression model is a moving average process of order 1, parameter θ, the correlation matrix is maximized at $\theta = 1$. It is clear that at $\theta = 1$, the $MA(1)$ process will not be invertible and hence the designs from such a process will not be useful in practice. For $AR(1)$ process with parameter ϕ, at $\phi = 1$, the $AR(1)$ process becomes non-stationary which also leads to the same problem. In constructing her designs, Zhou (2001) [47] used the OLS method of estimation. It is well known that such designs will be less efficient than designs based on WLS or GLS. We shall adapt the simulated annealing algorithm of Fang and Wiens (2000) [12] to demonstrate that the designs constructed by Zhou (2001) [47] can be improved.

2.4 The simulated annealing algorithm

In our example, we consider the model (2.5) and a design space S with equally spaced design points $x_i = (2i-1)/2N, i = 1, 2, \cdots, N$. For simplicity, we assume that one

of (n, N) is a multiple of the other. If $n < N$, then we also require that they have the same parity.

A simulated annealing algorithm, in general, consists of the following:

(a) A description of the initial state of the process; that is, of the starting vector of allocations $\mathbf{n} = \{n_1, n_2, \cdots, n_N\}$

(b) A scheme by which subsequent states are generated

(c) A criterion according to which these subsequent states are accepted or rejected.

If $n > N$, then the initial state is the uniform design, with $n_i = n/N$ for $i = 1, 2, \cdots, N$. If $n \leq N$, then this vector of frequencies assigned to $x_1, x_2, \cdots, x_N$ starts with the vector formed by repeating the vector $(1, 0, \cdots, 0)$ {with $N/n - 1$) 0's } $[n/2]$ times. This is followed by the same vector with the order of its elements reversed. If N is odd, then also a vector $(0, \cdots, 0, 1, 0, \cdots, 0)$ of length N/n is inserted in the middle. Thus in either case the initial design is symmetric and at least close to uniform. We impose symmetry on the designs largely for its intuitive appeal. However, we remark that searches for better, possibly asymmetric, designs have yielded no improvements.

To generate new designs, first define a $[N/2] \times 1$ vector $\mathbf{v} = (v_1, \cdots, v_{[N/2]}) =$

$(n_1, \cdots, n_{[N/2]})$ of the current allocation vector. Define

$$J_+ = \{i | v_i > 0\}, \quad J_0 = \{i | v_i = 0\} \tag{2.34}$$

with cardinalities $j_+ \geq 1$ and j_0. Generate a Bernoulli random variable,

$$B = \begin{cases} 1, & \text{with probability } \frac{j_0}{j_0+j_+}; \\ 0, & \text{with probability } \frac{j_+}{j_0+j_+}. \end{cases}$$

If $J_+ \geq 2$, then choose two indices (t_1, t_2) from J_+, at random and without replacement, and (if $B = 1$) pick an index t_0 form J_0, at random. Replace $\mathbf{v}$ by the vector $\tilde{\mathbf{v}}$ whose elements are those of $\mathbf{v}$ except for

$$\tilde{v}_{t_0} = v_{t_0} + B, \quad \tilde{v}_{t_1} = v_{t_1} - 1, \quad \& \quad \tilde{v}_{t_2} = v_{t_2} + 1 - B. \tag{2.35}$$

If $J_+ = 1$, then pick t_0 from J_0 at random, let t_1 be the element of J_+, and replace (2.35) by

$$\tilde{v}_{t_0} = v_{t_0} + 1, \quad \tilde{v}_{t_1} = v_{t_1} - 1. \tag{2.36}$$

If N is even, then $\tilde{\mathbf{n}} = (\tilde{n}_1, \cdots, \tilde{n}_N) = (\tilde{v}_1, \cdots, \tilde{v}_{N/2}, \tilde{v}_{N/2}, \cdots, \tilde{v}_1)$, thus preserving symmetry. If N is odd, then a further Bernoulli variable is simulated, with probability $1/N$ of "success". If a success is obtained, then, with probability $1/2$, $n_{[N/2]+1}$ (the frequency assigned to 0) is increased by 2, with these taken randomly and symmetrically from the remaining n_i. With the remaining probability $1/2$, $n_{[N/2]+1}$ is

reduced by 2, with these allocated randomly and symmetrically to the remaining n_i. If $n_{[N/2]+1} < 2$, then this step is omitted. Then $\tilde{\mathbf{n}}$ is constructed as earlier, with also the inclusion of the new frequency $\tilde{n}_{[N/2]+1}$.

To accept or reject $\tilde{\mathbf{n}}$ as the next state, first evaluate the loss $I = I(\tilde{\mathbf{n}})$. If $I(\tilde{\mathbf{n}}) < I(\mathbf{n})$, then $\tilde{\mathbf{n}}$ is accepted and iterations continue. If $\Delta I = I(\tilde{\mathbf{n}}) - I(\mathbf{n}) > 0$, then $\tilde{\mathbf{n}}$ is accepted with probability $exp(-\Delta I/T)$, where T is a user-chosen parameter. We initially choose $T = 7.5$, and we decrease T by a factor of .9 after each 100 iterations.

2.5 Example: AR(1) model

For the purpose of illustration, we assume that the errors are from an autoregressive process of order 1. That is, $\epsilon_i = \rho\epsilon_{i-1} + a_i; \quad i = 1, 2, \cdots, N,$

where a_i is a sequence of uncorrelated random variables with mean 0 and constant variance σ_a^2 called a white noise process. Thus we write $a_i \sim WN(0, \sigma_a^2)$.

We define the lag k autocovariance function $\gamma_k = E[\epsilon_{i-k}\epsilon_i]$, where $k \geq 1$. Then we have $\gamma_{k-1} = E[\epsilon_{i-k}\epsilon_{i-1}]$. By applying standard techniques we can show that the lag k autocovariance function γ_k satisfies the difference equation

$$\begin{aligned} E[\epsilon_{i-k}\epsilon_i] &= E[\epsilon_{i-k}(\rho\epsilon_{i-1} + a_i)] \\ &= \rho E[\epsilon_{i-k}\epsilon_{i-1}] + E[\epsilon_{i-k}a_i] \\ \gamma_k &= \rho\gamma_{k-1}\ ;\ k \geq 1 \end{aligned}$$

Using the fact that the autocorrelation function $\rho_k = \frac{\gamma_k}{\gamma_0}$, solving the difference equation we obtain

$$Corr(\epsilon_{i-k},\ \epsilon_i) = \rho^k\ ;\ k \geq 1.$$

Then, the autocorrelation matrix R in (2.10) becomes

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}_{n_1} & \rho\mathbf{J}_{n_1,n_2} & \rho^2\mathbf{J}_{n_1,n_3} & \cdots & \rho^{(N-1)}\mathbf{J}_{n_1,n_N} \\ \rho\mathbf{J}_{n_2,n_1} & \mathbf{I}_{n_2} & \rho\mathbf{J}_{n_2,n_3} & \cdots & \rho^{(N-2)}\mathbf{J}_{n_2,n_N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{(N-1)}\mathbf{J}_{n_N,n_1} & \rho^{(N-2)}\mathbf{J}_{n_N,n_2} & \rho^{(N-3)}\mathbf{J}_{n_N,n_3} & \cdots & \mathbf{I}_{n_N} \end{bmatrix} \tag{2.37}$$

and $\mathbf{R}^*$ in (2.16) is

$$\mathbf{R}^* = \begin{bmatrix} 0 & \rho & \rho^2 & \cdots & \rho^{j-1} & \cdots & \rho^{N-1} \\ \rho & 0 & \rho & \cdots & \rho^{j-2} & \cdots & \rho^{N-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ \rho^{i-1} & \rho^{i-2} & \rho^{i-3} & \cdots & \rho^{|i-j|} & \cdots & \rho^{N-i} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & \rho^{N-j} & \cdots & 0 \end{bmatrix}.$$

If we define the $N \times N$ matrix $\mathbf{G}_k$ with elements

$(\mathbf{G}_k)_{ij} = \delta_{|i-j|=k}$; $k = 1, 2, \cdots, N-1$, we can rewrite $\mathbf{R}^*$ as

$$\mathbf{R}^* = \rho\mathbf{G}_1 + \rho^2\mathbf{G}_2 + \cdots\cdots + \rho^i\mathbf{G}_i + \cdots\cdots + \rho^{N-1}\mathbf{G}_{N-1}.$$

Then,

$$\begin{aligned} trace(\mathbf{A}\mathbf{R}^*\mathbf{A}^T) &= trace\left\{\mathbf{A}\left(\rho\mathbf{G}_1 + \rho^2\mathbf{G}_2 + \cdots\cdots + \rho^i\mathbf{G}_i + \cdots\cdots + \rho^{N-1}\mathbf{G}_{N-1}\right)\mathbf{A}^T\right\} \\ &= \sum_{i=1}^{N-1} \rho^i trace(\mathbf{A}\mathbf{G}_i\mathbf{A}^T) \end{aligned}$$

and

$$\begin{aligned} \max_{\mathbf{R}^*} trace(\mathbf{A}\mathbf{R}^*\mathbf{A}^T) &= \max_{|\rho|<1} trace(\mathbf{A}\mathbf{R}^*\mathbf{A}^T) \\ &= \max_{|\rho|<1} \sum_{i=1}^{N-1} \rho^i trace(\mathbf{A}\mathbf{G}_i\mathbf{A}^T) \end{aligned} \tag{2.38}$$

Finally from (2.33) and (2.38) we get

$$\max_{\mathbf{f}_N} \max_{\mathbf{R}^*} AMSE = \tau^2 \lambda_{max} \left[\mathbf{M}_1^{-1} \mathbf{M}_2 \mathbf{M}_1^{-1} \right] + \frac{\nu \tau^2}{N} \left(trace \left[\mathbf{AWUM}_1^{-1} \right] + \max_{|\rho|<1} \sum_{i=1}^{N-1} \rho^i trace(\mathbf{AG}_i \mathbf{A}^T) \right).$$

Finally, the simulated annealing algorithm was used to construct optimal designs for the case of the $AR(1)$ model.

Chapter 3

Robust Design Using Generalized Least Squares Estimation

In Chapter 2, we used the weighted least squares criterion to estimate $\boldsymbol{\beta}$ in (2.9). In this section, we use the generalized least squares estimation to estimate $\boldsymbol{\beta}$ in (2.9). We use the same notations as in Chapter 2. Now, we re-write equation (2.9) as

$$\mathbf{Y} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{f}_u + \boldsymbol{\epsilon}, \tag{3.1}$$

where $Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma} = \sigma^2\mathbf{R}$ as in (2.10). Since $\boldsymbol{\Sigma}$ is a covariance matrix, it is at least positive semidefinite. For the existence of $\boldsymbol{\Sigma}^{-1}$, we assume that $\boldsymbol{\Sigma}$ is of positive definite matrix. In order to transform (3.1) into a model with constant variance we

premultiply (3.1) by $\Sigma^{-1/2}$ to obtain

$$\Sigma^{-1/2}\mathbf{Y} = \Sigma^{-1/2}\mathbf{Q}\boldsymbol{\beta} + \Sigma^{-1/2}\mathbf{f}_u + \Sigma^{-1/2}\boldsymbol{\epsilon}. \tag{3.2}$$

Now, define,

$$\mathbf{Y}^* = \Sigma^{-1/2}\mathbf{Y}, \quad \mathbf{Q}^* = \Sigma^{-1/2}\mathbf{Q}, \quad \mathbf{f}_u^* = \Sigma^{-1/2}\mathbf{f}_u \quad \text{and,} \quad \boldsymbol{\epsilon}^* = \Sigma^{-1/2}\boldsymbol{\epsilon}. \tag{3.3}$$

Then (3.1) becomes

$$\mathbf{Y}^* = \mathbf{Q}^*\boldsymbol{\beta} + \mathbf{f}_u^* + \boldsymbol{\epsilon}^*, \tag{3.4}$$

and the generalized least squares (GLS) estimate of the parameter $\boldsymbol{\beta}$ can be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_g &= (\mathbf{Q}^{*T}\mathbf{Q}^*)^{-1}\mathbf{Q}^{*T}\mathbf{Y}^* \\ &= (\mathbf{Q}^T\Sigma^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\Sigma^{-1}\mathbf{Y}. \end{aligned} \tag{3.5}$$

From (3.5) and the fact that $\Sigma^{-1} = (\sigma^2)^{-1}\mathbf{R}^{-1}$, the GLS estimate for $\boldsymbol{\beta}$ becomes

$$\hat{\boldsymbol{\beta}}_g = (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Y}. \tag{3.6}$$

In what follows, we use the Average Mean Squared Error (AMSE) design criterion defined in (2.13) to construct optimal designs. To do this, we first derive the bias and the variance of $\hat{\boldsymbol{\beta}}_g$.

3.1 The bias and Variance of $\hat{\beta}_g$

Taking expectation of both sides in equation (3.6) we obtain

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}_g] &= E[(\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Y}] \\ &= (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}E[\mathbf{Y}] \\ &= (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}[\mathbf{Q}\boldsymbol{\beta} + \mathbf{f}_u] \\ &= \boldsymbol{\beta} + (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{f}_u \end{aligned} \tag{3.7}$$

It follows that the bias of $\hat{\beta}_g$ denoted by d_g, is

$$\begin{aligned} \mathbf{d}_g &= bias(\hat{\boldsymbol{\beta}}_g) \\ &= E(\hat{\boldsymbol{\beta}}_g) - \boldsymbol{\beta} \\ &= (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{f}_u \end{aligned} \tag{3.8}$$

Again, from (3.6) we can show that

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}_g) &= Var\left\{\left(\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Y}\right\} \\ &= (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}Var(\mathbf{Y})\mathbf{R}^{-1}\mathbf{Q}(\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1} \\ &= \sigma^2(\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1} \end{aligned} \tag{3.9}$$

3.2 The Average Variance(AV_g) & Average Square Bias (ASB_g) of $\hat{\beta}_g$

Recall that in Chapter 2 we used the singular value decomposition method to decompose the matrix $\mathbf{Z}$ into

$$\mathbf{Z}_{N\times p} = \mathbf{U}_{N\times p}\ \mathbf{\Lambda}_{p\times p}\ \mathbf{V}^T_{p\times p},$$

$$\text{where}\quad \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_{\mathrm{p}}. \tag{3.10}$$

Similarly, the matrix $\mathbf{Q}_{n\times p}$ was decomposed using the singular value decomposition method as

$$\mathbf{Q}_{n\times p} = \mathbf{U}_q\ \mathbf{\Lambda}_q\ \mathbf{V}^T_q$$

$$\text{with}\quad \mathbf{U}^{\mathrm{T}}_{\mathrm{q}}\mathbf{U}_{\mathrm{q}} = \mathbf{V}^{\mathrm{T}}_{\mathrm{q}}\mathbf{V}_{\mathrm{q}} = \mathbf{I}_{\mathrm{p}}. \tag{3.11}$$

where $\mathbf{U}_q$ is a $n\times p$ dimensional matrix, $\mathbf{\Lambda}_q$ is a $p\times p$ diagonal matrix of singular values and, $\mathbf{V}_q$ is also a $p \times p$ dimensional matrix. Using the singular value decomposition of $\mathbf{Q}$ we write

$$\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q} = \mathbf{V}_q\mathbf{\Lambda}_q\mathbf{U}^T_q\mathbf{R}^{-1}\mathbf{U}_q\mathbf{\Lambda}_q\mathbf{V}^T_q$$

and define $\mathbf{M}_{1g} = \mathbf{U}_q^T \mathbf{R}^{-1} \mathbf{U}_q$. Then

$$\mathbf{Q}^T \mathbf{R}^{-1} \mathbf{Q} = \mathbf{V}_q \boldsymbol{\Lambda}_q \mathbf{M}_{1g} \boldsymbol{\Lambda}_q \mathbf{V}_q^T$$

$$\text{and } (\mathbf{Q}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Q})^{-1} = \mathbf{V}_{\mathrm{q}} \boldsymbol{\Lambda}_{\mathrm{q}}^{-1} \mathbf{M}_{1\mathrm{g}}^{-1} \boldsymbol{\Lambda}_{\mathrm{q}}^{-1} \mathbf{V}_{\mathrm{q}}^{\mathrm{T}}. \tag{3.12}$$

It follows from (3.9) that the variance of $\hat{\boldsymbol{\beta}}_g$ can be expressed as

$$Var(\hat{\boldsymbol{\beta}}_g) = \sigma^2 \mathbf{V}_q \boldsymbol{\Lambda}_q^{-1} \mathbf{M}_{1g}^{-1} \boldsymbol{\Lambda}_q^{-1} \mathbf{V}_q{}^T. \tag{3.13}$$

Recall from (2.18) that the Average Variance of $\hat{\boldsymbol{\beta}}_g$ can be written as

$$AV_g = \frac{1}{N} trace \left\{ \mathbf{Z} Var(\hat{\boldsymbol{\beta}}_g) \mathbf{Z}^T \right\}.$$

Using the SVD of $\mathbf{Z}$ in (3.10) and (3.13) we have

$$AV_g = \frac{\sigma^2}{N} trace \left\{ \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T \mathbf{V}_q \boldsymbol{\Lambda}_q^{-1} \mathbf{M}_{1g}^{-1} \boldsymbol{\Lambda}_q^{-1} \mathbf{V}_q^T \mathbf{V} \boldsymbol{\Lambda} \mathbf{U}^T \right\}.$$

Define the matrix $\mathbf{A}$ as

$$\mathbf{A} = \boldsymbol{\Lambda} \mathbf{V}^T \mathbf{V}_q \boldsymbol{\Lambda}_q^{-1}.$$

$$\text{so that } \mathbf{A}^{\mathrm{T}} = \boldsymbol{\Lambda}_{\mathrm{q}}^{-1} \mathbf{V}_{\mathrm{q}}^{\mathrm{T}} \mathbf{V} \boldsymbol{\Lambda}.$$

After some algebra we find that

$$AV_g = \frac{\sigma^2}{N} trace \left\{ \mathbf{A} \mathbf{M}_{1g}^{-1} \mathbf{A}^T \right\}. \tag{3.14}$$

When we substitute $\nu = \frac{\sigma^2}{\tau^2}$, we obtain

$$AV_g = \frac{\nu\tau^2}{N} trace\left\{\mathbf{A}\mathbf{M}_{1g}^{-1}\mathbf{A}^T\right\} \tag{3.15}$$

Remark that the Average variance AV_g depends on the correlation matrix $\mathbf{R}$ but does not depend on $\mathbf{f}_N$ or $\mathbf{f}_u$.

Concerning the ASB, we know from (3.8) that

$$bias(\hat{\boldsymbol{\beta}}_g) = \mathbf{d}_g = (\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}^{-1}\mathbf{f}_u. \tag{3.16}$$

Using the SVD of $\mathbf{Q}$ in (3.16) we obtain

$$\begin{aligned}\mathbf{d}_g &= \mathbf{V}_q\boldsymbol{\Lambda}_q^{-1}\mathbf{M}_{1g}^{-1}\boldsymbol{\Lambda}_q^{-1}\mathbf{V}_q^T\mathbf{V}_q\boldsymbol{\Lambda}_q\mathbf{U}_q^T\mathbf{R}^{-1}\mathbf{f}_u \\ &= \mathbf{V}_q\boldsymbol{\Lambda}_q^{-1}\mathbf{M}_{1g}^{-1}\mathbf{U}_q^T\mathbf{R}^{-1}\mathbf{f}_u. \end{aligned} \tag{3.17}$$

From equation (2.27) in Chapter 2, we have

$$\mathbf{f}_N = \sqrt{N}\tau\tilde{\mathbf{U}}\mathbf{c}. \tag{3.18}$$

By using the orthogonality of wavelet systems we have

$$\frac{1}{n}\sum_{i=1}^{n} q(u_i)f(u_i) = 0, \tag{3.19}$$

which implies that

$$\begin{aligned}
\frac{1}{n}\mathbf{Q}^T\mathbf{f}_u &= 0 \\
\frac{1}{n}\mathbf{V}_q\mathbf{\Lambda}_q\mathbf{U}_q^T\mathbf{f}_u &= 0 \\
\mathbf{V}_q^T\mathbf{V}_q\mathbf{\Lambda}_q\mathbf{U}_q^T\mathbf{f}_u &= 0 \\
\mathbf{\Lambda}_q^{-1}\mathbf{\Lambda}_q\mathbf{U}_q^T\mathbf{f}_u &= 0 \\
\Longrightarrow \mathbf{f}_u^T\mathbf{U}_q &= 0.
\end{aligned}$$

This implies that the vector $\mathbf{f}_u$ belongs to the space of orthogonal complement of the column space of $\mathbf{U}_q$.

$$\mathbf{f}_u \in \{Col(\mathbf{U}_q)\}^{\perp} \tag{3.20}$$

Define $n_m = \max\{n_1, n_2, \cdots, n_N\}$. Then, we can write

$$\begin{aligned}
\frac{1}{N}\sum_{k=1}^{n}[f(u_k)]^2 &= \frac{1}{N}\sum_{i=1}^{N} n_i[f(x_i)]^2 \\
&\leq \frac{1}{N} n_m \sum_{i=1}^{N}[f(x_i)]^2 \\
&\leq n_m\tau^2.
\end{aligned}$$

$$\begin{aligned}
\text{That is } \mathbf{f}_u^{\mathrm{T}}\mathbf{f}_u &\leq N n_m\tau^2 \\
\| \mathbf{f}_u \|^2 &\leq N n_m\tau^2.
\end{aligned}$$

Thus, there exits some $\mathbf{c}_q$ satisfying $\| \mathbf{c}_q \|^2 \leq 1$ such that

$$\begin{aligned} \| \mathbf{f}_u \|^2 &= N n_m \tau^2 \| \mathbf{c}_q \|^2 \\ \| \mathbf{f}_u \|^2 &= \| \sqrt{N n_m} \tau \mathbf{c}_q \|^2 \\ \Longrightarrow \quad \| \mathbf{f}_u \| &= \| \sqrt{N n_m} \tau \mathbf{c}_q \| . \end{aligned}$$

Suppose $\tilde{\mathbf{U}}_q$ is a matrix whose columns constitute on orthonormal basis for $\{Col(\mathbf{U}_q)\}^{\perp}$.

Then, since $\mathbf{f}_u \in \{Col(\mathbf{U}_q)\}^{\perp}$ we have that

$$\mathbf{f}_u = \sqrt{N n_m} \tau \tilde{\mathbf{U}}_q \mathbf{c}_q. \tag{3.21}$$

Similar to (2.15) we can obtain the formula for ASB_g as

$$\begin{aligned} ASB_g &= \frac{1}{N} \mathbf{d}_g^T \mathbf{Z}^T \mathbf{Z} \mathbf{d}_g + \frac{1}{N} \| \mathbf{f}_N \|^2 \\ &= \frac{1}{N} \mathbf{d}_g^T \mathbf{Z}^T \mathbf{Z} \mathbf{d}_g + \frac{1}{N} \mathbf{f}_N^T \mathbf{f}_N. \end{aligned} \tag{3.22}$$

From equations (3.10) and (3.17) we have

$$\begin{aligned} \mathbf{Z} &= \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \\ \mathbf{d}_g &= \mathbf{V}_q \mathbf{\Lambda}_q^{-1} \mathbf{M}_{1g}^{-1} \mathbf{U}_q^T \mathbf{R}^{-1} \mathbf{f}_u. \end{aligned} \tag{3.23}$$

By using equations (3.22) and (3.23) we obtain

$$\begin{aligned} ASB_g &= \frac{1}{N} \mathbf{f}_u^T \mathbf{R}^{-1} \mathbf{U}_q \mathbf{M}_{1g}^{-1} \mathbf{\Lambda}_q^{-1} \mathbf{V}_q^T \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \mathbf{V}_q \mathbf{\Lambda}_q^{-1} \mathbf{M}_{1g}^{-1} \mathbf{U}_q^T \mathbf{R}^{-1} \mathbf{f}_u + \frac{1}{N} \mathbf{f}_N^T \mathbf{f}_N \\ &= \frac{1}{N} \mathbf{f}_u^T \mathbf{R}^{-1} \mathbf{U}_q \mathbf{M}_{1g}^{-1} \mathbf{A}^T \mathbf{A} \mathbf{M}_{1g}^{-1} \mathbf{U}_q^T \mathbf{R}^{-1} \mathbf{f}_u + \frac{1}{N} \mathbf{f}_N^T \mathbf{f}_N. \end{aligned} \tag{3.24}$$

If we define

$$\mathbf{B} = \mathbf{A}\mathbf{M}_{1g}^{-1}\mathbf{U}_q^T\mathbf{R}^{-1}, \tag{3.25}$$

the ASB_g simplifies to

$$\begin{aligned} ASB_g &= \frac{1}{N}\mathbf{f}_u^T\mathbf{B}^T\mathbf{B}\mathbf{f}_u + \frac{1}{N}\mathbf{f}_N^T\mathbf{f}_N \\ &= n_m\tau^2\mathbf{c}_q^T\tilde{\mathbf{U}}_q^T\mathbf{B}^T\mathbf{B}\tilde{\mathbf{U}}_q\mathbf{c}_q + \tau^2\mathbf{c}^T\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\mathbf{c}. \end{aligned} \tag{3.26}$$

We note that the ASB_g depend on both $\mathbf{c}_q$ and $\mathbf{c}$.

3.3 Maximizing the Average Mean Square Error (AMSE) with respect to $\mathbf{f}_N$, $\mathbf{f}_u$, and $\mathbf{R}$

Following our approach in Chapter 2 we first decompose the $AMSE_g$ into 2 components

$$AMSE_g = ASB_g + AV_g. \tag{3.27}$$

and then maximize the $AMSE_g$ with respect to $\mathbf{f}_N$ and $\mathbf{f}_u$. That is, we solve

$$\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \max_{\mathbf{f}_u,\,\mathbf{f}_N} ASB_g + \max_{\mathbf{f}_u,\,\mathbf{f}_N} AV_g.$$

Since the AV_g depends on neither $\mathbf{f}_u$ nor $\mathbf{f}_N$, the above maximization problem becomes

$$\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \max_{\mathbf{f}_u,\,\mathbf{f}_N} ASB_g + AV_g. \tag{3.28}$$

Now,

$$\begin{aligned}
\max_{\mathbf{f}_u,\,\mathbf{f}_N} ASB_g &= \max_{\mathbf{c}_q,\mathbf{c}} ASB_g \\
&= \max_{\mathbf{c}_q,\mathbf{c}} \{n_m\tau^2\mathbf{c}_q^T\tilde{\mathbf{U}}_q^T\mathbf{B}^T\mathbf{B}\tilde{\mathbf{U}}_q\mathbf{c}_q + \tau^2\mathbf{c}^T\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\mathbf{c}\} \\
&= \tau^2\{n_m \max_{\|\mathbf{c}_q\|^2\le 1} \mathbf{c}_q^T\tilde{\mathbf{U}}_q^T\mathbf{B}^T\mathbf{B}\tilde{\mathbf{U}}_q\mathbf{c}_q + \max_{\|\mathbf{c}\|^2\le 1} \mathbf{c}^T\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\mathbf{c}\} \\
&= \tau^2\{n_m\lambda_{\max}[\tilde{\mathbf{U}}_q^T\mathbf{B}^T\mathbf{B}\tilde{\mathbf{U}}_q] + \lambda_{\max}[\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}]\} \\
&= \tau^2\{n_m\lambda_{\max}[\mathbf{B}\tilde{\mathbf{U}}_q\tilde{\mathbf{U}}_q^T\mathbf{B}^T] + \lambda_{\max}[\mathbf{I}_{N-r}]\} \\
&= \tau^2\{n_m\lambda_{\max}[\mathbf{B}\tilde{\mathbf{U}}_q\tilde{\mathbf{U}}_q^T\mathbf{B}^T] + 1\}.
\end{aligned}$$

Recall that $\tilde{\mathbf{U}}_q\tilde{\mathbf{U}}_q^T + \mathbf{U}_q\mathbf{U}_q^T = \mathbf{I}_n$. It follows that $\tilde{\mathbf{U}}_q\tilde{\mathbf{U}}_q^T = \mathbf{I}_n - \mathbf{U}_q\mathbf{U}_q^T$

$$\begin{aligned}
\text{Therefore } \lambda_{\text{mg}} &= \lambda_{\max}[\mathbf{B}\tilde{\mathbf{U}}_q\tilde{\mathbf{U}}_q^T\mathbf{B}^T] \\
&= \lambda_{\max}[\mathbf{B}(\mathbf{I}_n - \mathbf{U}_q\mathbf{U}_q^T)\mathbf{B}^T] \qquad (3.29)
\end{aligned}$$

$$\text{and } \max_{\mathbf{f}_u,\,\mathbf{f}_N} \text{ASB}_g = \tau^2\{n_m\lambda_{mg} + 1\} \qquad (3.30)$$

For a fixed correlation matrix $\mathbf{R}$, using the equations (3.28), (3.30) and (3.15) we have the following result

$$\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \tau^2\left[n_m\lambda_{mg} + 1 + \frac{\nu}{N}trace\left\{\mathbf{A}\mathbf{M}_{1g}^{-1}\mathbf{A}^T\right\}\right]. \qquad (3.31)$$

Next, we take the maximum with respect to the matrix $\mathbf{R}$:

$$\max_{\mathbf{R}}\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \tau^2\max_{\mathbf{R}}\left[n_m\lambda_{mg} + 1 + \frac{\nu}{N}trace\left\{\mathbf{A}\mathbf{M}_{1g}^{-1}\mathbf{A}^T\right\}\right] \qquad (3.32)$$

and minimize over the design to obtain the optimal design. Due to the nature of (3.32), we will maximize (3.32) over $\mathbf{R}$ and minimize over the design numerically.

3.4 Example: AR(1) model using GLS method

We have already discussed an example based on the AR(1) model using WLS method in Section 2.6 of Chapter 2. We will use the same notations in our example using the GLS method. The correlation matrix $\mathbf{R}$ depends only on ρ, which takes values between -1 and +1. Thus, maximizing over $\mathbf{R}$ is equivalent to maximizing over ρ. Thus,

$$\max_{\mathbf{R}} \max_{\mathbf{f}_u,\, \mathbf{f}_N} AMSE_g = \max_{|\rho|<1} \max_{\mathbf{f}_u,\, \mathbf{f}_N} AMSE_g = \tau^2 \max_{|\rho|<1} \left[n_m \lambda_{mg} + 1 + \frac{\nu}{N} trace \left\{ \mathbf{A} \mathbf{M}_{1g}^{-1} \mathbf{A}^T \right\} \right] \tag{3.33}$$

We do not have a closed form expression for the equation (3.33), so we take several values of ρ between -1 and +1 and we use the simulated annealing algorithm to obtain optimal design (say prior optimum designs) for each value of ρ. Also, we computed minimum loss for each ρ. Among the minimum losses, we pick maximum loss over all ρ values (say ρ_{opt}).

Chapter 4

Results and Discussions

In this section, we compare the performances of the OLS, WLS and, GLS methods by implementing the simulated annealing algorithm for various values of $\upsilon = \frac{\sigma^2}{\tau^2}$. We defined υ in Chapter 2 as $\upsilon = \frac{\sigma^2}{\tau^2}$. The values of υ used in constructing the designs are $\upsilon = 0.05, 1, 5, 10, 50$. Recall that no closed form expression for the maximum loss over the correlation matrix R was obtained in Chapter 3. Thus, the loss function depends on the correlation matrix R. In the special case of the $AR(1)$ model, the maximum loss was shown to be a function of the lag 1 autocorrelation parameter ρ. As a result, various values of ρ was used to construct a prior optimal design that depends on ρ. The maximum loss was then taken over the recorded minimum losses of various values of ρ and the corresponding prior optimal design was chosen as the

optimal design. In our examples, the Daubechies wavelet system with wavelet number 4 and the number of wavelet terms $m = 2$ was used.

We assume that given the data $[(x_i, y_i)]_{i=1}^{n}$ the experimenter will use either the ordinary least squares method or weighted least squares method or generalized least squares method to estimate the unknown mean response function via wavelet expansion. Under this assumption and the $AR(1)$ model we construct integer valued designs for fixed values of ρ based on the simulated annealing algorithm. The average squared bias (ASB), average variance and, minimum loss are reported in Table 4.1 - Table 4.15 for OLS, WLS and, GLS respectively.

4.1 Integer Valued Optimal Design for $\upsilon = 0.05$

The results in Table 4.1 show that when $\upsilon = 0.05$, the maximum of minimum loss under OLS is 2.049277 occurs at $\rho = 0.97$. Under WLS Table 4.2 shows that the maximum of minimum loss was 2.128778 at $\rho = 0.85$, whereas the maximum of minimum loss for GLS which occurs at $\rho = 0.99999$ is 6.117178. So the OLS method works well considering all range of ρ values when $\upsilon = 0.05$. But we look at Figure 4.4 the GLS method performs very well between -0.6 and 0.6 compared to OLS and WLS. Also OLS works better compared to WLS at entire range. In the Figure 4.5

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.00081	2.000057	0.000752968
-0.9	2.005967	2.003607	0.002360792
-0.8	2.005569	2.003309	0.002260222
-0.6	2.022668	2.018743	0.003925095
-0.4	2.026631	2.022610	0.004021718
-0.2	2.026935	2.02206	0.004875512
0	2.02783	2.021947	0.00588259
0.2	2.026449	2.019190	0.007258619
0.4	2.029613	2.020224	0.009388586
0.6	2.034541	2.021450	0.01309060
0.8	2.044767	2.022849	0.02191772
0.9	2.042100	2.011287	0.03081347
0.96	2.043253	2.008100	0.03515281
0.97	2.049277	2.013889	0.03538827
0.98	2.039064	2.004818	0.03424561
0.99	2.046060	2.009735	0.03632598
0.99999	2.046796	2.009735	0.03706117

Table 4.1: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 0.05$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.080894	2.079151	0.001742439
-0.9	2.073306	2.070749	0.002557343
-0.8	2.073343	2.070503	0.002840847
-0.6	2.068662	2.065604	0.003058168
-0.4	2.069462	2.065604	0.003858562
-0.2	2.065668	2.060600	0.005067228
0	2.109661	2.103610	0.006051221
0.2	2.069385	2.061999	0.00738538
0.4	2.071444	2.061999	0.00944496
0.6	2.116754	2.103621	0.01313386
0.8	2.12522	2.103482	0.02173836
0.85	2.128778	2.103482	0.02529681
0.9	2.107938	2.078494	0.02944403
0.96	2.08487	2.052372	0.03249821
0.97	2.084333	2.052372	0.03196113
0.98	2.074100	2.043588	0.0305121
0.99	2.071805	2.043588	0.02821649
0.99999	2.085199	2.060650	0.02454846

Table 4.2: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 0.05$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.085488	1.085488	5.02156e-08
-0.9	1.059457	1.058895	0.0005617386
-0.8	3.276169	3.256918	0.01925158
-0.6	1.271935	1.269581	0.002353984
-0.4	1.037739	1.029456	0.008282977
-0.2	1.007879	1.001206	0.006673024
0	1.004795	1	0.004794594
0.2	1.007828	1.001152	0.006676406
0.4	1.031120	1.022797	0.008322858
0.6	1.185932	1.173867	0.01206458
0.8	2.083669	2.062900	0.02076957
0.9	3.709693	3.682098	0.02759484
0.95	4.027648	3.998459	0.02918908
0.98	5.367551	5.342765	0.02478615
0.99999	6.117178	6.11714	3.815961e-05

Table 4.3: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 0.05$

the optimum design for OLS and WLS looks like same but the optimal design for GLS is completely different from OLS and WLS.

It is clear, from

$$\begin{aligned}\max_{\mathbf{R}^*}\max_{\mathbf{f}_N} AMSE &= \tau^2\lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right] \\ &\quad + \frac{\nu\tau^2}{N}\left(trace\left[\mathbf{AWUM}_1^{-1}\right] + \max_{\mathbf{R}^*} trace\left[\mathbf{AR}^*\mathbf{A}^T\right]\right)\end{aligned}$$

for weighted least squares method or

$$\max_{\mathbf{R}}\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \tau^2 \max_{\mathbf{R}}\left[n_m\lambda_{mg} + 1 + \frac{\nu}{N} trace\left\{\mathbf{AM}_{1g}^{-1}\mathbf{A}^T\right\}\right]$$

for generalized least squares method, that when υ is small, say $\upsilon = 0.05$, the bias component of the $AMSE$ becomes dominant; whereas as υ becomes larger the average variance AV becomes dominant, as seen in Table 4.1 - Table 4.15, irrespective of the value of ρ. Figures 4.1 - 4.4, 4.6 - 4.9, 4.11 - 4.14, 4.16 - 4.19, and 4.21 - 4.24 provide a picture of the patterns in the optimal $AMSE$

$$\begin{aligned}\max_{\mathbf{f}_N}\max_{\mathbf{R}^*} AMSE &= \tau^2\lambda_{max}\left[\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}\right] \\ &\quad + \frac{\nu\tau^2}{N}\left(trace\left[\mathbf{AWUM}_1^{-1}\right] + \max_{|\rho|<1}\sum_{i=1}^{N-1}\rho^i trace(\mathbf{AG}_i\mathbf{A}^T)\right).\end{aligned}$$

for weighted least squares method and

$$\max_{\mathbf{R}}\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \max_{|\rho|<1}\max_{\mathbf{f}_u,\,\mathbf{f}_N} AMSE_g = \tau^2 \max_{|\rho|<1}\left[n_m\lambda_{mg} + 1 + \frac{\nu}{N} trace\left\{\mathbf{AM}_{1g}^{-1}\mathbf{A}^T\right\}\right]$$

for generalized least squares method, for various values of υ and ρ. In Figures 4.4, 4.9, 4.14, 4.19, and 4.24 we compare the performance of the designs for OLS, WLS, and GLS methods of estimation. It can be seen that when $|\rho| < 0.6$, the designs from the GLS method are more efficient. However, when the $AR(1)$ process approaches non-stationary state $|\rho| \longrightarrow 1$, the OLS becomes more efficient.

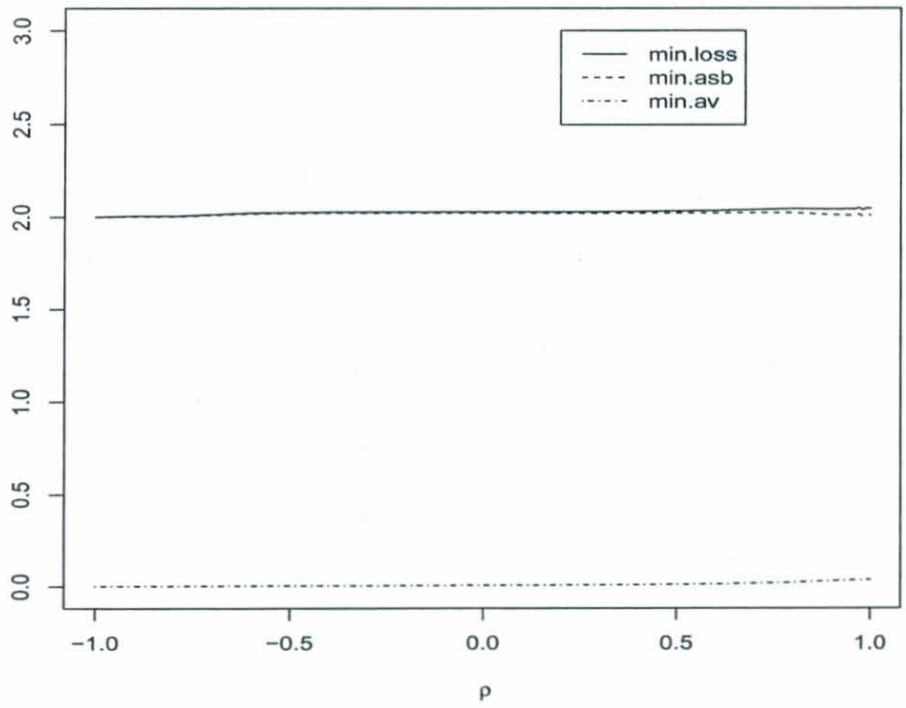

Figure 4.1: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 0.05$

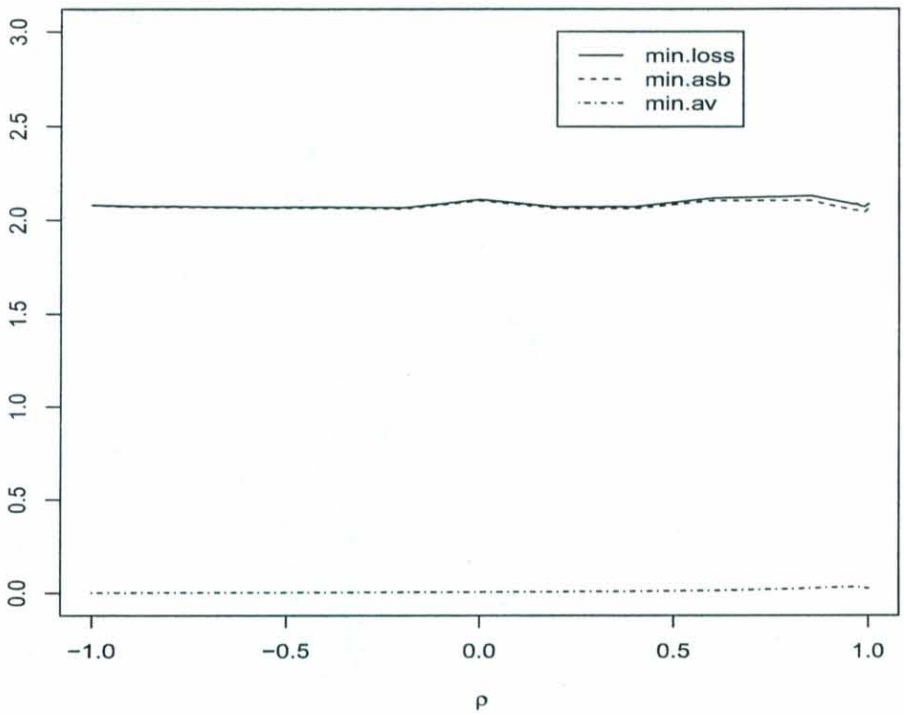

Figure 4.2: The comparison of *Minimum loss*, *ASB* and, *AV* of WLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 0.05$

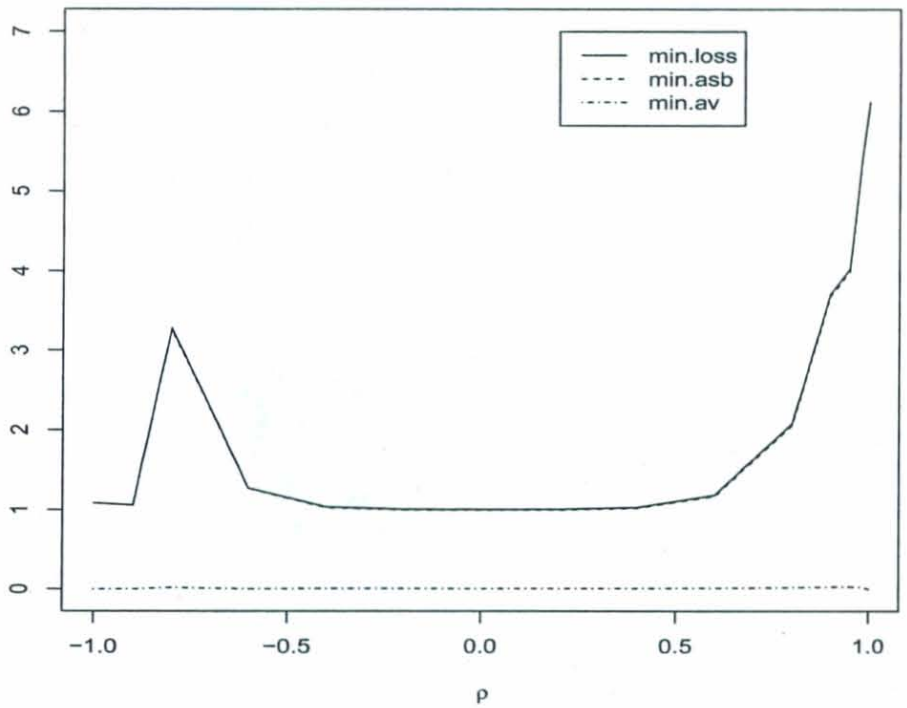

Figure 4.3: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 0.05$

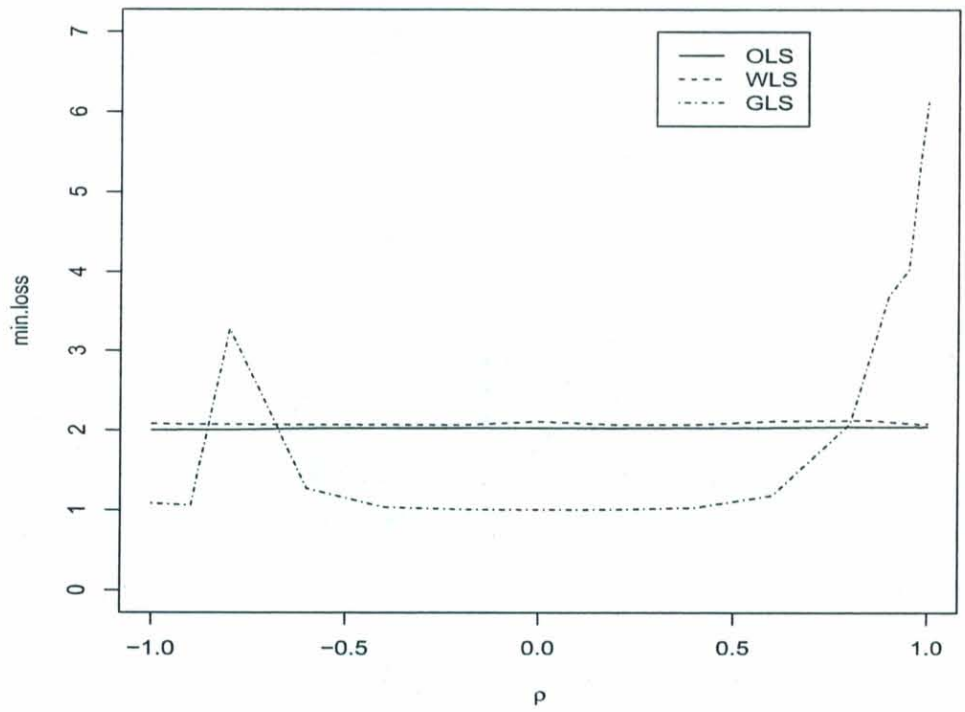

Figure 4.4: The comparison of *Minimum loss* for OLS, WLS and, GLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 0.05$

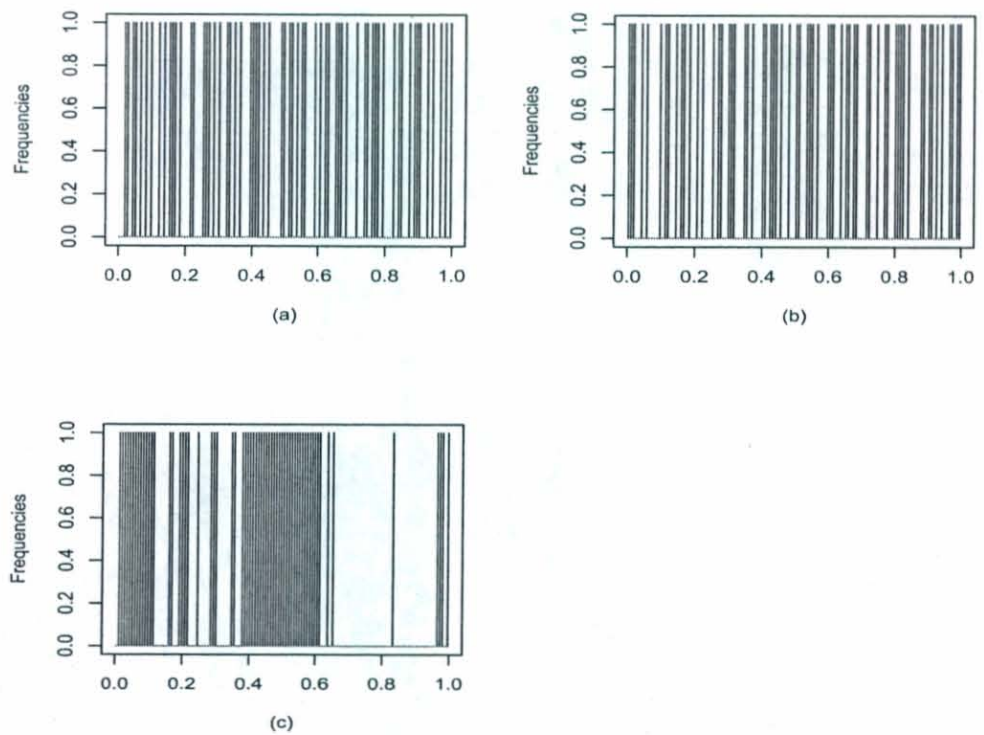

Figure 4.5: Minimax design points: (a) OLS (b) WLS (c) GLS when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 0.05$

4.2 Integer Valued Optimal Design for $v = 1$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.996743	1.995523	0.001219876
-0.9	2.033878	2.006919	0.02695928
-0.8	2.066264	2.036592	0.02967136
-0.6	2.05416	2.005176	0.04898433
-0.4	2.088217	2.022613	0.06560409
-0.2	2.093798	2.000279	0.09351933
0	2.118930	2.003297	0.1156327
0.2	2.155274	2.013068	0.1422062
0.4	2.196515	2.009910	0.1866055
0.6	2.269324	2.012412	0.2569118
0.8	2.450034	2.017605	0.4324294
0.9	2.604849	2.013777	0.5910713
0.96	2.699093	2.018787	0.6803062
0.97	2.706874	2.058684	0.6481904
0.98	2.679406	2.060381	0.619025
0.99999	2.597526	2.106485	0.4910409

Table 4.4: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

In Table 4.4 the maximum of minimum loss for *OLS* is 2.706874 when $v = 1$ at $\rho = 0.97$. Similarly, from Table 4.5 the maximum of minimum loss for *WLS* is 2.710629 when $v = 1$ at $\rho = 0.97$. Furthermore, from Table 4.6 the maximum of minimum loss for *GLS* is 6.348578 at $\rho = 0.97$. So the *OLS* method works well considering all range of ρ values when $v = 1$. But looking at Figure 4.9 we note that the *GLS* method performs very well between -0.99999 and 0.8 and compared to

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.077370	2.071862	0.005507841
-0.9	2.09058	2.065656	0.02492445
-0.8	2.11256	2.083059	0.02950096
-0.6	2.141584	2.085153	0.05643037
-0.4	2.133177	2.066520	0.06665684
-0.2	2.177348	2.077710	0.09963783
0	2.168872	2.051462	0.1174097
0.1	2.210527	2.080389	0.1301374
0.2	2.230831	2.084857	0.1459744
0.4	2.283002	2.098379	0.1846235
0.6	2.318458	2.061046	0.2574113
0.8	2.487547	2.05188	0.4356671
0.9	2.643451	2.058292	0.5851591
0.94	2.686622	2.053949	0.6326732
0.96	2.703887	2.066719	0.6371685
0.97	2.710629	2.085311	0.6253184
0.98	2.647005	2.056689	0.5903156
0.99999	2.521115	2.135155	0.3859602

Table 4.5: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.085489	1.085488	1.004312e-06
-0.9	1.091835	1.080426	0.01140855
-0.8	1.165675	1.139410	0.02626474
-0.6	1.194062	1.147588	0.04647377
-0.4	1.195115	1.029456	0.1656595
-0.2	1.113398	1.023462	0.0899358
0	1.095892	1	0.09589189
0.2	1.101511	1.001926	0.09958534
0.4	1.189254	1.022797	0.1664572
0.6	1.415159	1.173867	0.2412917
0.8	2.253745	1.843812	0.4099328
0.9	3.233730	2.701286	0.5324431
0.95	4.725332	4.137803	0.5875288
0.96	5.000686	4.426609	0.5740775
0.97	6.348578	5.78259	0.5659878
0.98	5.771601	5.275205	0.496396
0.99	5.97794	5.58029	0.3976498
0.99999	6.117903	6.11714	0.0007631921

Table 4.6: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 1$

OLS and WLS. Also the OLS works slightly well compared to WLS except at high positive correlation. In Figure 4.10 the optimum design for OLS, WLS and, GLS are different from each other. Also the optimal design for GLS is approximately the uniform design.

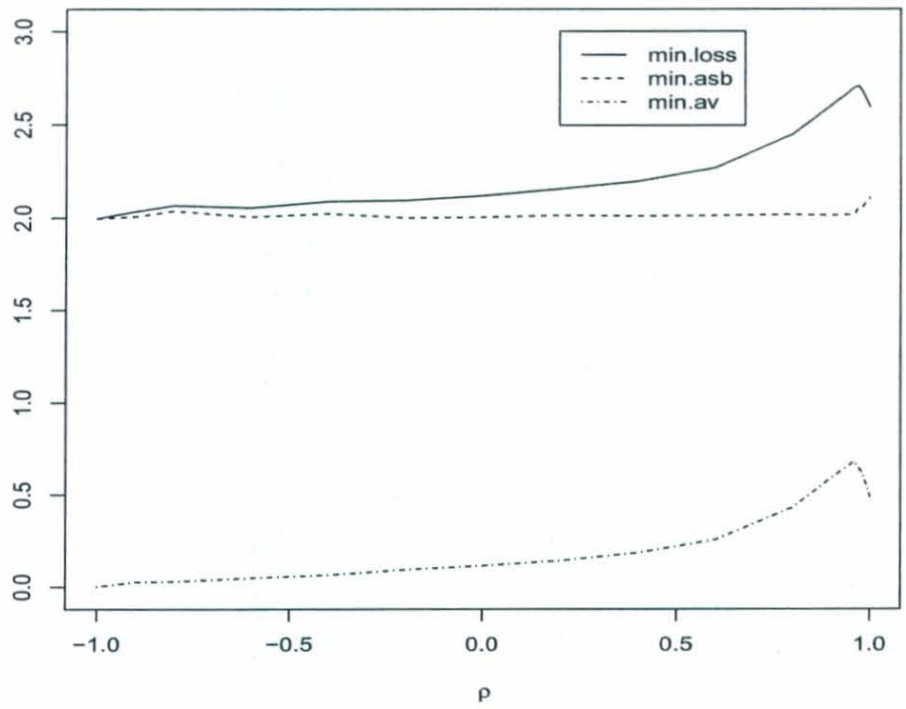

Figure 4.6: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

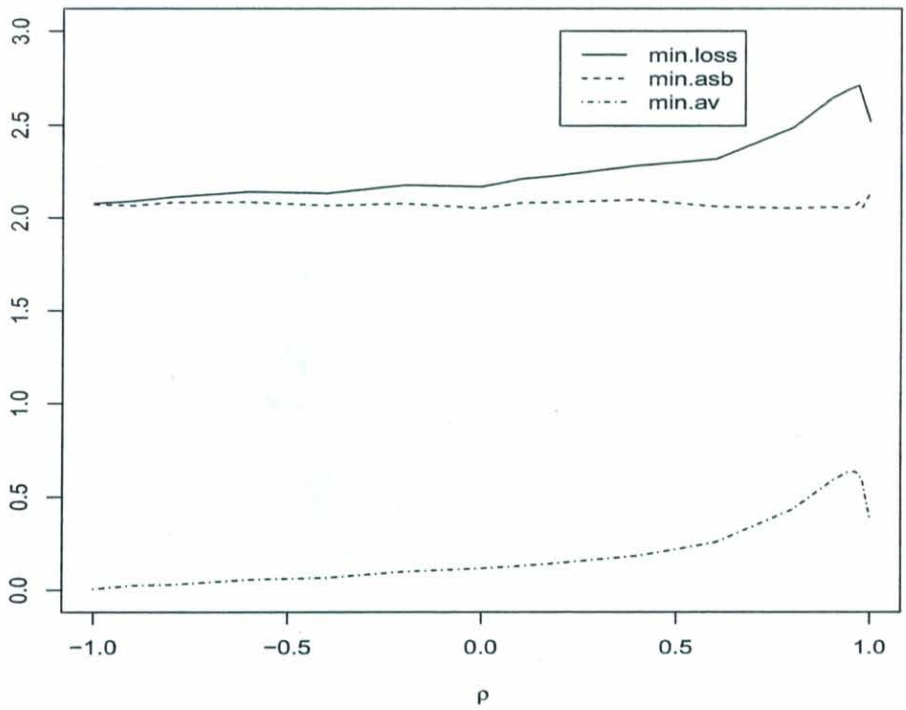

Figure 4.7: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

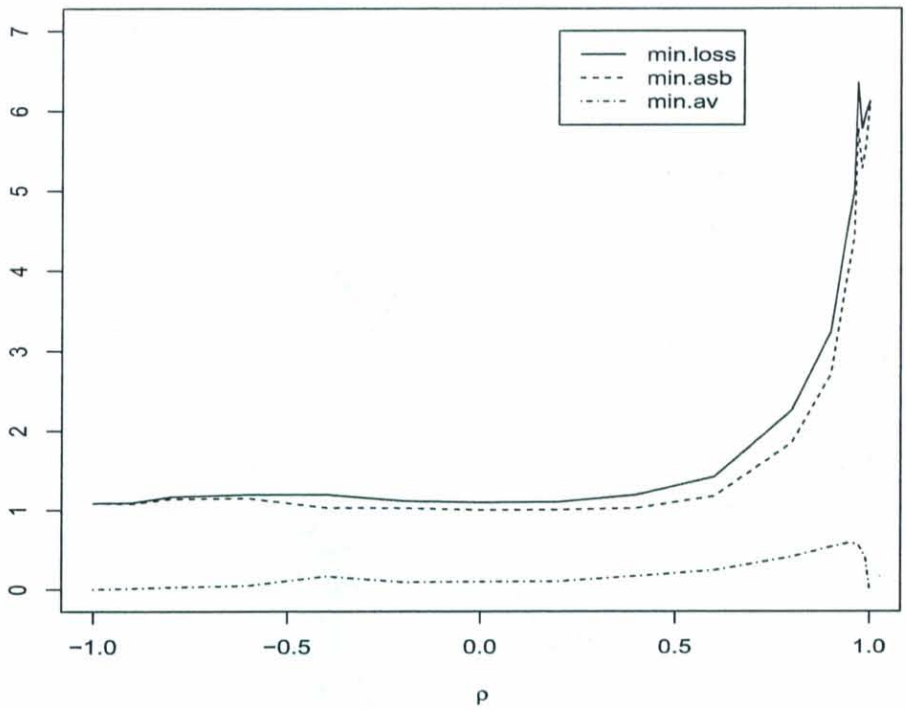

Figure 4.8: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

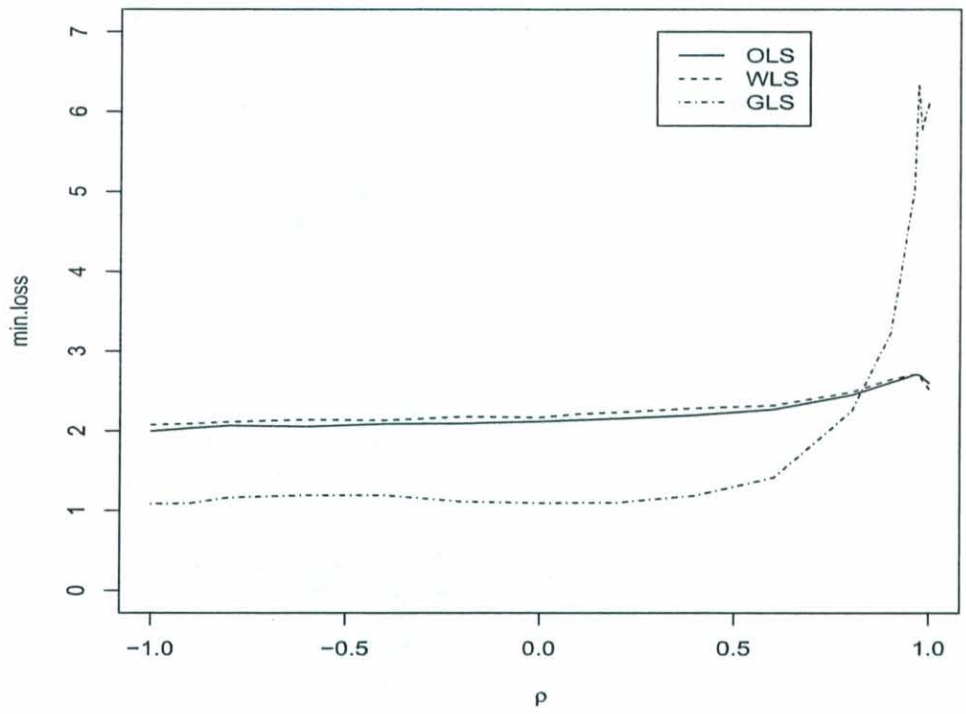

Figure 4.9: The comparison of *Minimum loss* for OLS, WLS and, GLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

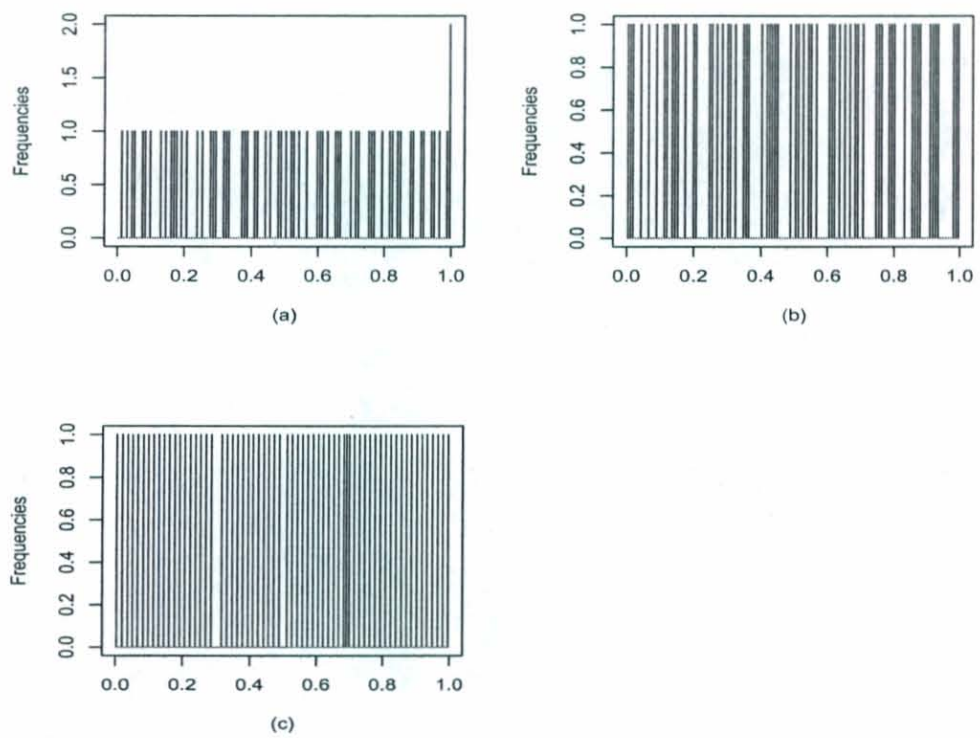

Figure 4.10: Minimax design points: (a) OLS (b) WLS (c) GLS when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 1$

4.3 Integer Valued Optimal Design for $\upsilon = 5$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.047865	2.057248	-0.0093831
-0.9	2.100265	2.049366	0.05089852
-0.8	2.132511	2.011909	0.1206018
-0.6	2.241812	2.018743	0.2230689
-0.4	2.314149	2.024323	0.2898256
-0.2	2.430382	2.010393	0.4199898
0	2.579118	2.010932	0.5681863
0.2	2.702937	2.011419	0.691518
0.4	2.871270	2.010999	0.8602715
0.6	3.257095	2.014997	1.242099
0.8	4.146442	2.061918	2.084524
0.9	4.922302	2.046661	2.875641
0.94	5.209525	2.116116	3.093409
0.95	5.272871	2.190409	3.082462
0.96	5.263621	2.20552	3.058101
0.97	5.169232	2.139847	3.029384
0.98	5.062085	2.25283	2.809255
0.99999	4.208087	2.499849	1.708238

Table 4.7: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $\upsilon = 5$

In Table 4.7 the maximum of minimum loss for *OLS* is 5.272871 when $\upsilon = 5$ at $\rho = 0.95$. Similarly, from Table 4.8 the maximum of minimum loss for *WLS* is 5.203846 when $\upsilon = 5$ at $\rho = 0.96$. Furthermore, from Table 4.9 the maximum of minimum loss for *GLS* is 8.846287 at $\rho = 0.98$. So the *WLS* method works well considering all range of ρ values when $\upsilon = 5$. But looking at Figure 4.14 we note

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.131281	2.124741	0.006540674
-0.9	2.153697	2.080342	0.0733552
-0.8	2.186272	2.062570	0.1237025
-0.6	2.259083	2.057157	0.2019262
-0.4	2.349457	2.044107	0.3053502
-0.1	2.575831	2.053423	0.5224084
0	2.636004	2.056528	0.579476
0.1	2.678510	2.048461	0.6300489
0.4	2.924343	2.053382	0.8709615
0.6	3.319283	2.076050	1.243233
0.8	4.185344	2.063325	2.122020
0.9	4.945942	2.06253	2.883413
0.94	5.191955	2.067705	3.12425
0.96	5.203846	2.171627	3.032218
0.97	5.073718	2.110858	2.96286
0.98	4.911075	2.173541	2.737534
0.99	4.594742	2.27139	2.323352
0.99999	3.862797	2.423530	1.439267

Table 4.8: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.085493	1.085488	5.02156e-06
-0.9	2.513215	2.443624	0.06959139
-0.87	2.005000	1.944012	0.06098852
-0.85	3.519706	3.537542	-0.0178355
-0.83	5.999396	3.914483	2.084914
-0.81	4.084916	3.891627	0.1932882
-0.8	1.417761	1.320471	0.09728996
-0.6	1.407637	1.16602	0.2416169
-0.4	1.332202	1.085325	0.2468768
-0.2	1.384103	1.040957	0.3431456
0.0	1.479459	1	0.4794594
0.4	1.805948	1.095238	0.71071
0.6	2.300031	1.282751	1.01728
0.8	4.139856	2.062900	2.076957
0.9	6.441582	3.682098	2.759484
0.97	8.612529	5.78259	2.829939
0.975	8.726667	5.996503	2.730165
0.98	8.846287	6.266682	2.579605
0.985	8.797336	6.445915	2.351421
0.99	7.568464	5.580743	1.987721
0.99999	6.120956	6.11714	0.003815961

Table 4.9: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

that the GLS method performs very well between -0.8 and 0.8 compared to OLS and WLS. Also the OLS are slightly better than WLS when ρ is between -0.99999 and 0.95 but between 0.95 and 0.99999, WLS performs well compared to OLS. In the Figure 4.15 the optimum design for OLS, WLS and, GLS is different from each other. Also the optimal design for GLS is approximately the uniform design.

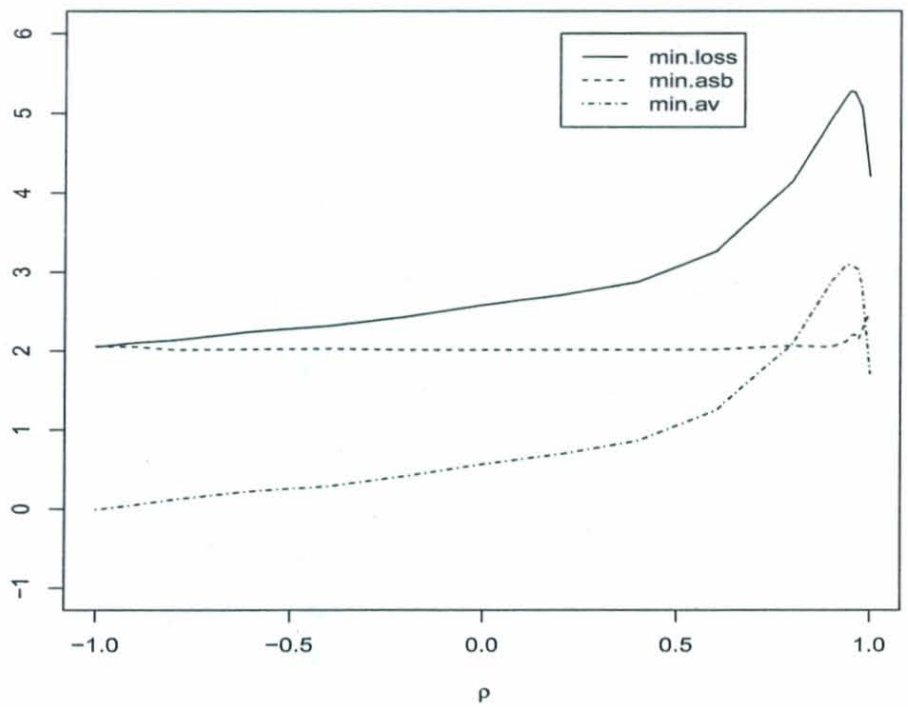

Figure 4.11: The comparison of *Minimum loss*, ASB and, AV of OLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

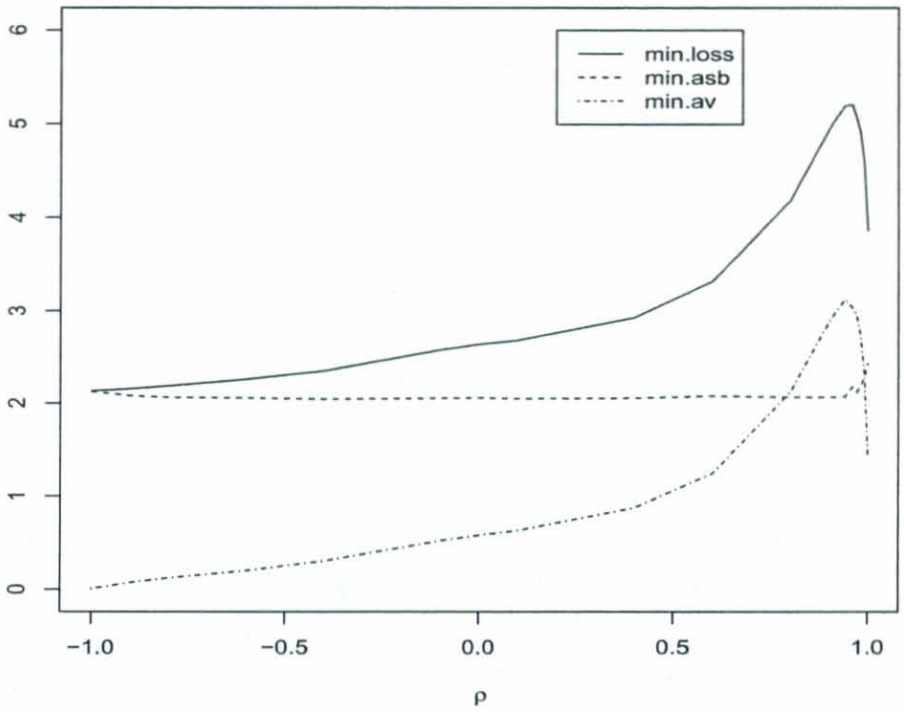

Figure 4.12: The comparison of *Minimum loss*, ASB and, AV of WLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

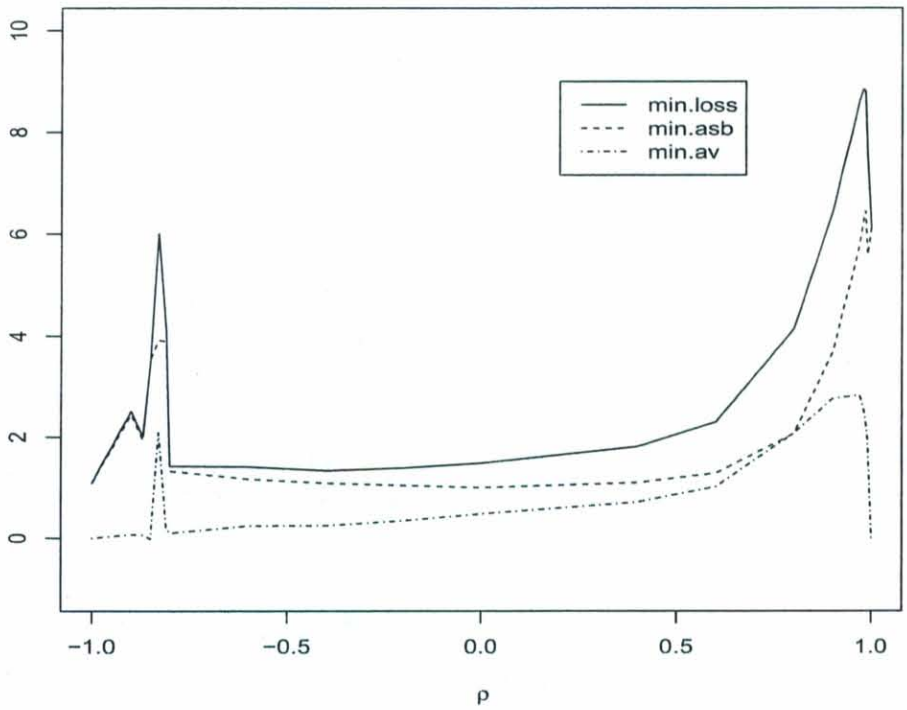

Figure 4.13: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

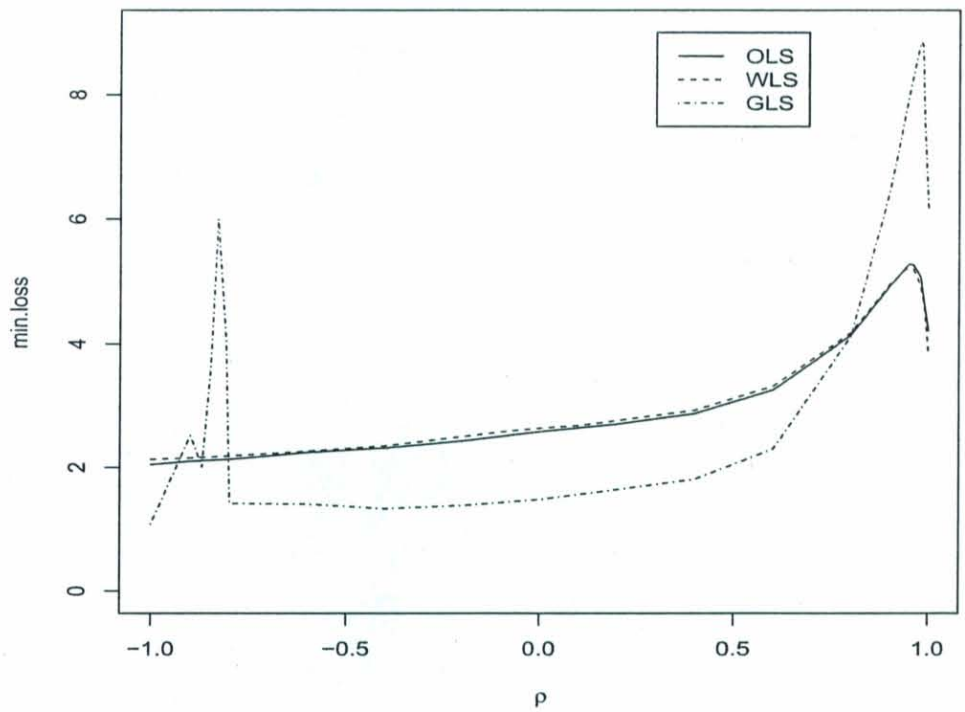

Figure 4.14: The comparison of *Minimum loss* for OLS, WLS and, GLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

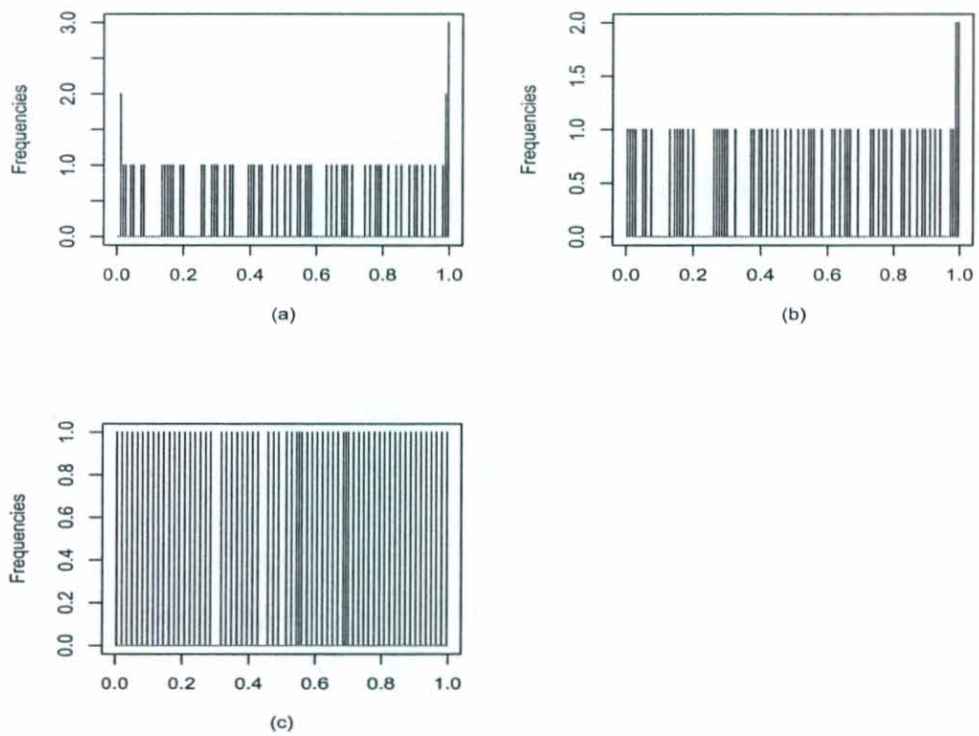

Figure 4.15: Minimax design points: (a) OLS (b) WLS (c) GLS when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 5$

4.4 Integer Valued Optimal Design for $v = 10$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.045167	2.0705	-0.02533313
-0.9	2.269027	2.241803	0.02722357
-0.8	2.322824	2.123792	0.1990314
-0.6	2.488752	2.088627	0.4001257
-0.4	2.612679	2.014389	0.5982902
-0.2	2.794901	2.002144	0.7927574
0	3.077763	2.001131	1.076632
0.2	3.365560	2.00663	1.358930
0.4	3.731137	2.027582	1.703555
0.6	4.479028	2.022361	2.456667
0.8	6.233229	2.070706	4.162522
0.9	7.864589	2.316972	5.547617
0.93	8.329474	2.430321	5.899152
0.94	8.362358	2.240780	6.121579
0.95	8.33708	2.264812	6.072267
0.96	8.333771	2.326118	6.007653
0.97	8.209657	2.490442	5.719215
0.98	7.83935	2.56329	5.27606
0.99999	5.771057	3.438719	2.332339

Table 4.10: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

In Table 4.10 the maximum of minimum loss for *OLS* is 8.362358 when $v = 10$ at $\rho = 0.94$. Similarly, from the Table 4.11 the maximum of minimum loss for *WLS* is 8.327955 when $v = 10$ at $\rho = 0.94$. Furthermore from Table 4.12 the maximum of minimum loss for *GLS* is 11.44247 at $\rho = 0.97$. So the *WLS* method works well considering all range of ρ values when $v = 10$. But looking at Figure 4.19 we note that

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	2.267489	2.581732	-0.3142424
-0.9	2.247185	2.112259	0.1349265
-0.8	2.331291	2.120416	0.2108748
-0.6	2.464161	2.063245	0.4009151
-0.4	2.667517	2.063233	0.6042834
-0.2	2.870352	2.049251	0.8211004
0	3.157555	2.046277	1.111278
0.2	3.445544	2.073824	1.371721
0.4	3.816696	2.080104	1.736593
0.6	4.666314	2.271805	2.394509
0.8	6.317123	2.13009	4.187033
0.9	7.932483	2.497829	5.434654
0.92	8.164618	2.347608	5.81701
0.94	8.327955	2.357453	5.970502
0.96	8.275183	2.289933	5.98525
0.97	8.072333	2.305684	5.766649
0.98	7.752204	2.531776	5.220427
0.99999	5.225318	3.38924	1.836078

Table 4.11: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.085498	1.085488	1.004312e-05
-0.9	1.56146	1.482761	0.07869866
-0.8	2.295938	2.139232	0.1567057
-0.6	1.585500	1.244839	0.340662
-0.4	1.849210	1.129927	0.7192823
-0.2	1.570897	1.088429	0.4824679
0	1.958919	1	0.9589189
0.2	2.002956	1.001531	1.001425
0.4	2.26098	1.030917	1.230064
0.6	3.373752	1.612591	1.761161
0.8	6.091392	2.112047	3.979345
0.9	8.214476	2.744597	5.469879
0.97	11.44247	5.78259	5.659878
0.98	11.42589	6.266682	5.159209
0.99	9.53389	5.55796	3.975931
0.99999	6.124772	6.11714	0.007631921

Table 4.12: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

the GLS method performs very well between -0.99999 and 0.8 compared to OLS and WLS. Also OLS and WLS is approximately same for entire range. In the Figure 4.20 the optimum design for OLS, WLS and, GLS is different from each other. Also the optimal design for GLS is approximately the uniform design.

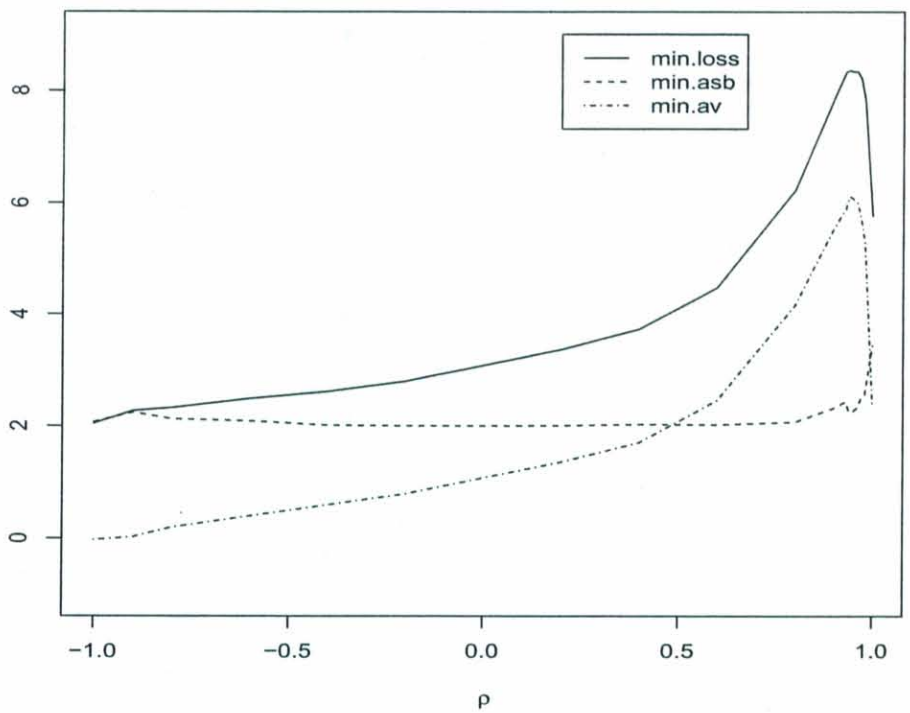

Figure 4.16: The comparison of *Minimum loss*, ASB and, AV of OLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

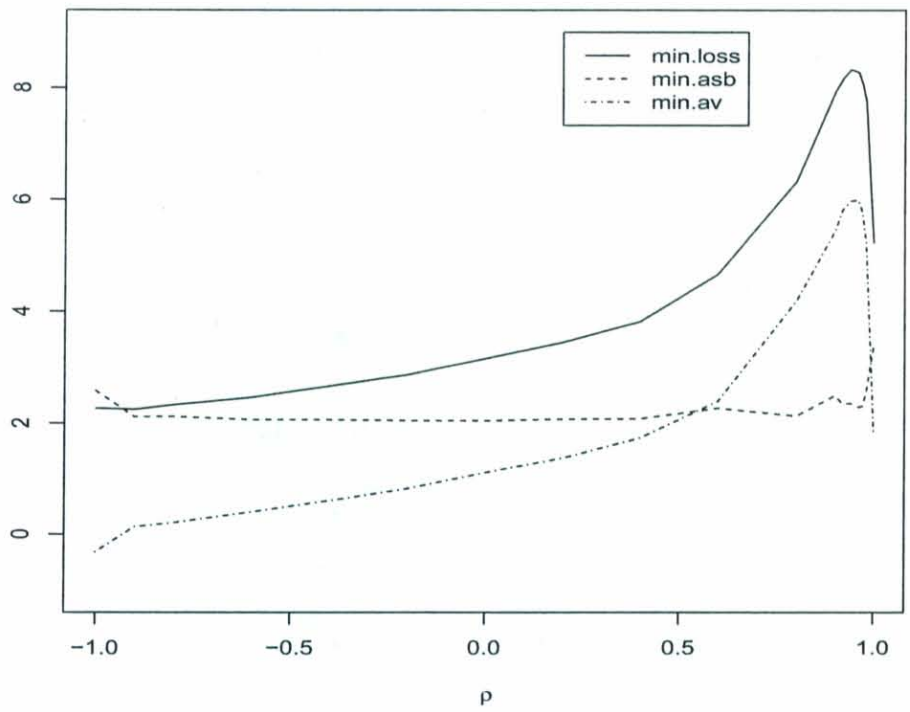

Figure 4.17: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

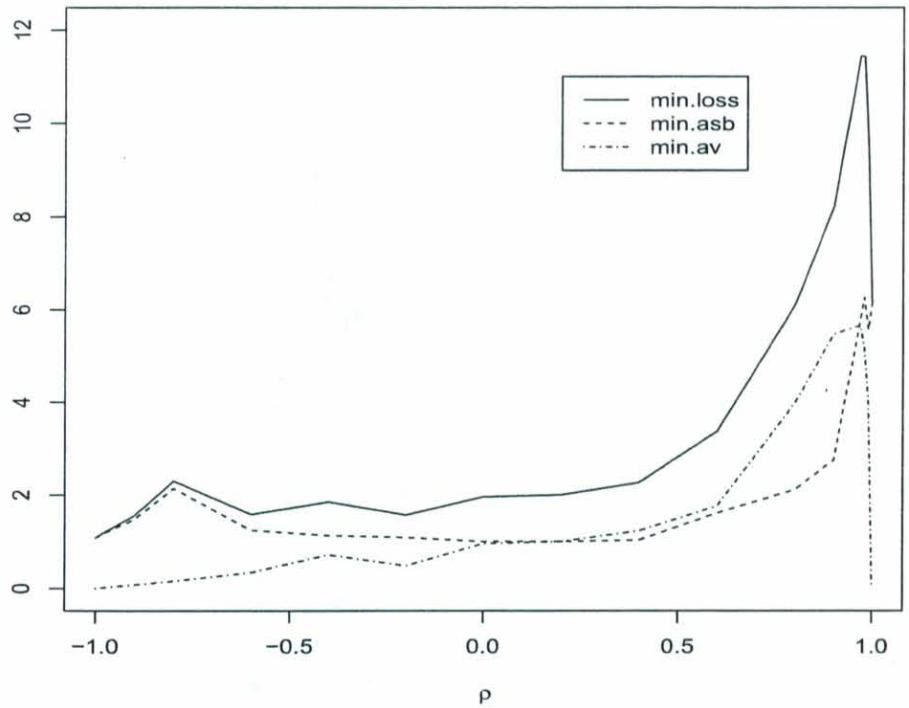

Figure 4.18: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

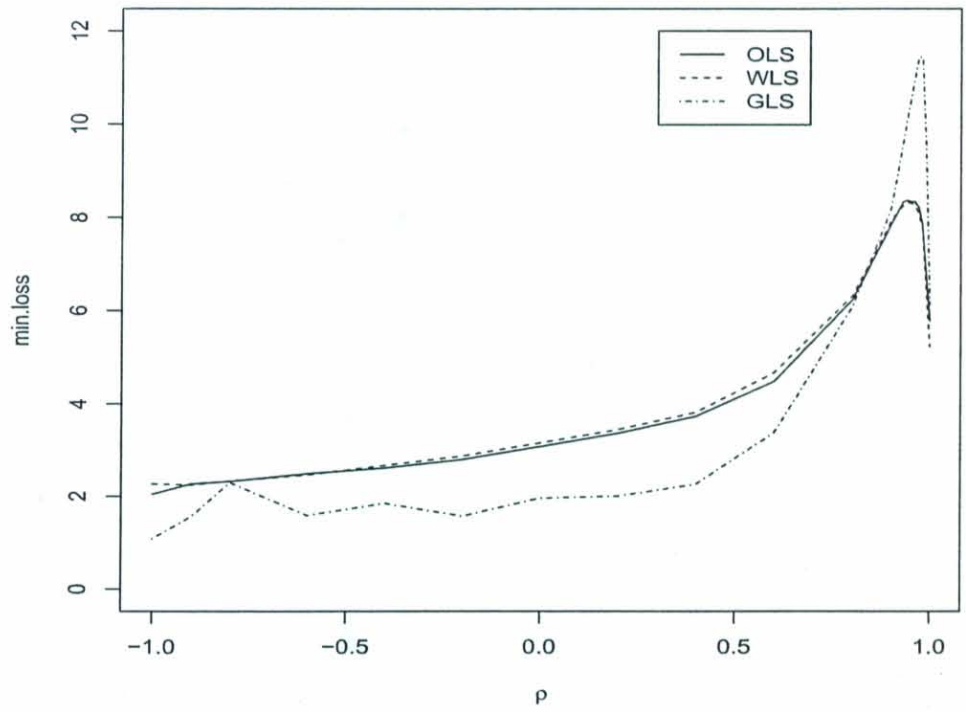

Figure 4.19: The comparison of *Minimum loss* for *OLS*, *WLS* and, *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

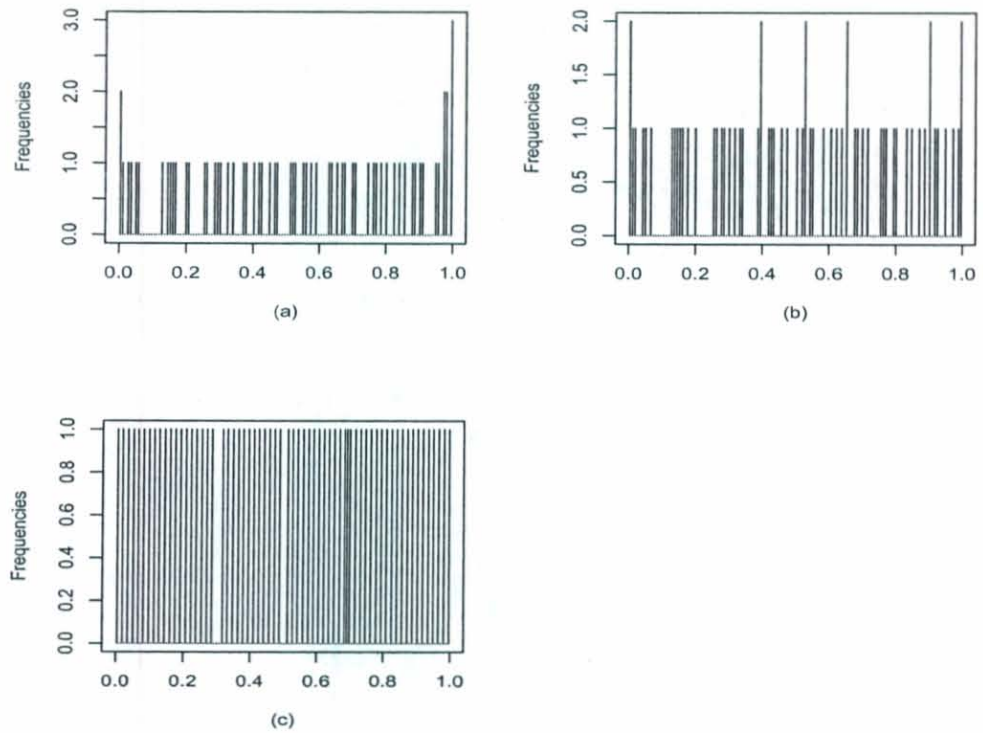

Figure 4.20: Minimax design points: (a) OLS (b) WLS (c) GLS when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 10$

4.5 Integer Valued Optimal Design for $v = 50$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	-10.55223	10.68958	-21.24181
-0.9	-3.608168	6.982474	-10.59064
-0.8	-0.7526654	6.001518	-6.754183
-0.6	1.69489	5.178896	-3.484006
-0.4	4.389297	3.702486	0.6868112
-0.2	5.684904	2.122914	3.56199
0	7.239887	2.074770	5.165118
0.2	8.70082	2.314758	6.386062
0.4	10.20747	3.288586	6.91888
0.6	13.13902	3.563669	9.575349
0.8	18.22139	7.716686	10.50471
0.9	21.46173	9.65862	11.80311
0.91	21.55166	9.96918	11.58248
0.92	21.23706	11.26862	9.968441
0.95	20.85874	11.65560	9.203142
0.96	19.73297	12.05363	7.67934
0.97	17.31635	12.66616	4.650195
0.98	15.54358	12.32861	3.214979
0.99999	-2.240714	17.10907	-19.34978

Table 4.13: The comparison of *Minimum loss*, *ASB* and, *AV* of *OLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

In the Table 4.13 the maximum of minimum loss for *OLS* is 21.55166 when $v = 50$ at $\rho = 0.91$. Similarly from Table 4.14 the maximum of minimum loss for *WLS* is 22.27768 when $v = 50$ at $\rho = 0.92$. Also from Table 4.15 the maximum of minimum loss for *GLS* is 33.49518 at $\rho = 0.95$. So the *OLS* method works well considering all range of ρ values when $v = 50$. But we looking Figure 4.24 we note that the *GLS*

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	-13.58427	11.65563	-25.23990
-0.9	-1.933222	6.253059	-8.186281
-0.8	-0.8325057	5.619122	-6.451628
-0.6	2.128258	4.73058	-2.602322
-0.4	4.31385	3.494529	0.8193212
-0.2	5.926634	2.536562	3.390073
0	7.421825	2.1856	5.236225
0.2	8.879873	2.462226	6.417647
0.4	10.26985	3.008363	7.261485
0.6	12.94254	4.014515	8.92802
0.8	18.58712	6.016187	12.57094
0.85	20.54168	7.607688	12.93399
0.9	21.79477	9.467093	12.32768
0.91	21.90699	8.990248	12.91674
0.92	22.27768	8.921036	13.35664
0.93	21.47031	11.26540	10.20491
0.94	20.94868	12.94772	8.000965
0.97	17.90256	11.67374	6.22882
0.99999	-4.566663	18.11551	-22.68218

Table 4.14: The comparison of *Minimum loss*, *ASB* and, *AV* of WLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

ρ	*Minimum loss*	*ASB*	*AV*
-0.99999	1.085538	1.085488	5.02156e-05
-0.9	1.833143	1.470647	0.3624959
-0.8	2.605071	1.910238	0.6948335
-0.6	2.301922	1.415000	0.8869212
-0.4	1.81329	1.155108	0.6581823
-0.2	2.890784	1.0707	1.820084
0	5.794594	1	4.794594
0.2	5.93516	1.002437	4.932723
0.4	6.150899	1.005317	5.145582
0.6	8.240323	1.732294	6.508028
0.8	20.30685	2.422327	17.88452
0.9	26.58375	2.829639	23.75411
0.94	32.71703	3.671975	29.04506
0.95	33.49518	4.493673	29.00151
0.97	32.46468	4.547815	27.91686
0.98	32.03664	6.215906	25.82074
0.99	25.38311	5.599105	19.78400
0.99999	6.1553	6.11714	0.03815961

Table 4.15: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

method performs very well between -0.4 and 0.6 compared to OLS and WLS. In the Figure 4.25 the optimum design for OLS and WLS takes more repeated observations compared to GLS method.

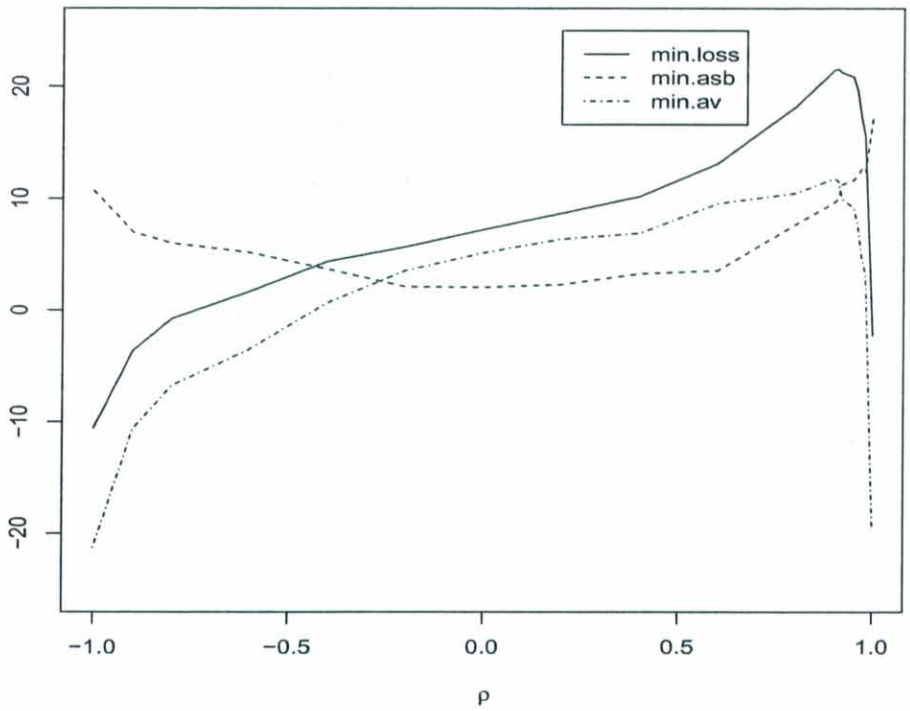

Figure 4.21: The comparison of *Minimum loss*, ASB and, AV of OLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

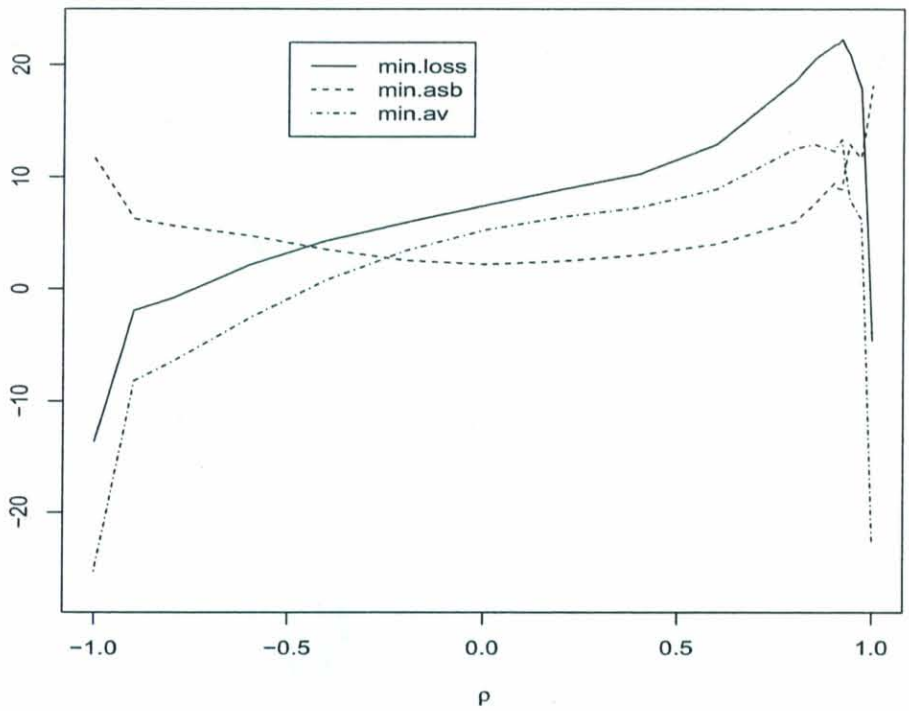

Figure 4.22: The comparison of *Minimum loss*, *ASB* and, *AV* of *WLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

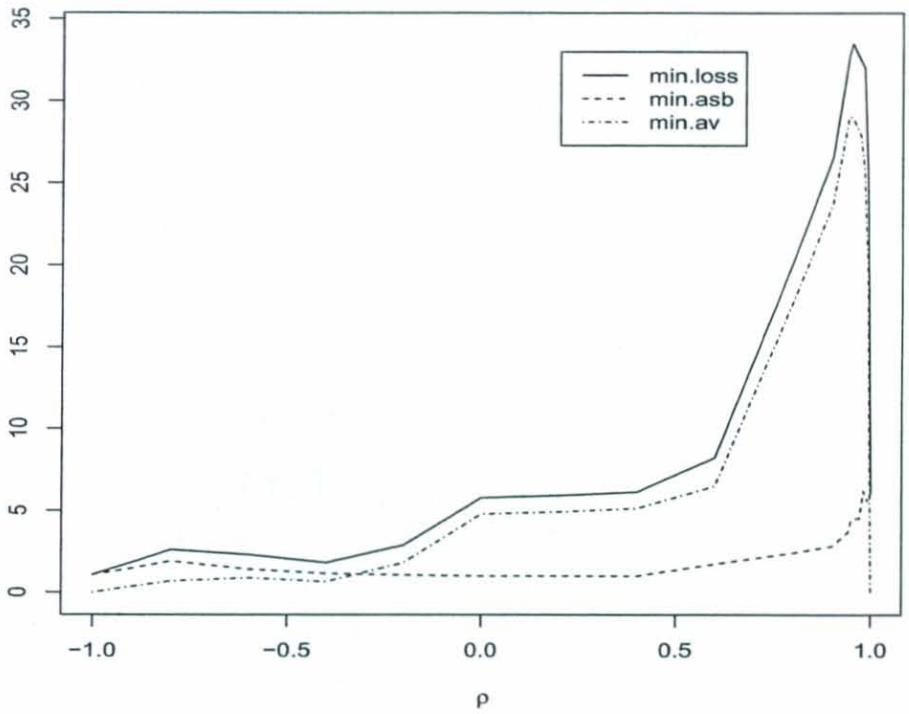

Figure 4.23: The comparison of *Minimum loss*, *ASB* and, *AV* of *GLS* method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

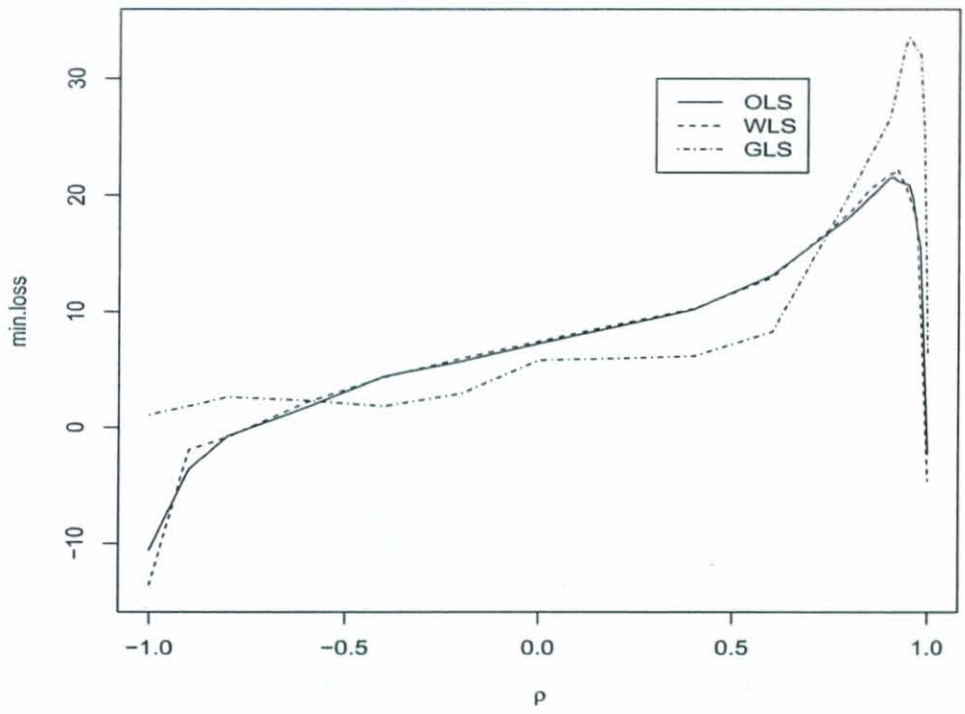

Figure 4.24: The comparison of *Minimum loss* for OLS, WLS and, GLS method for various ρ when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

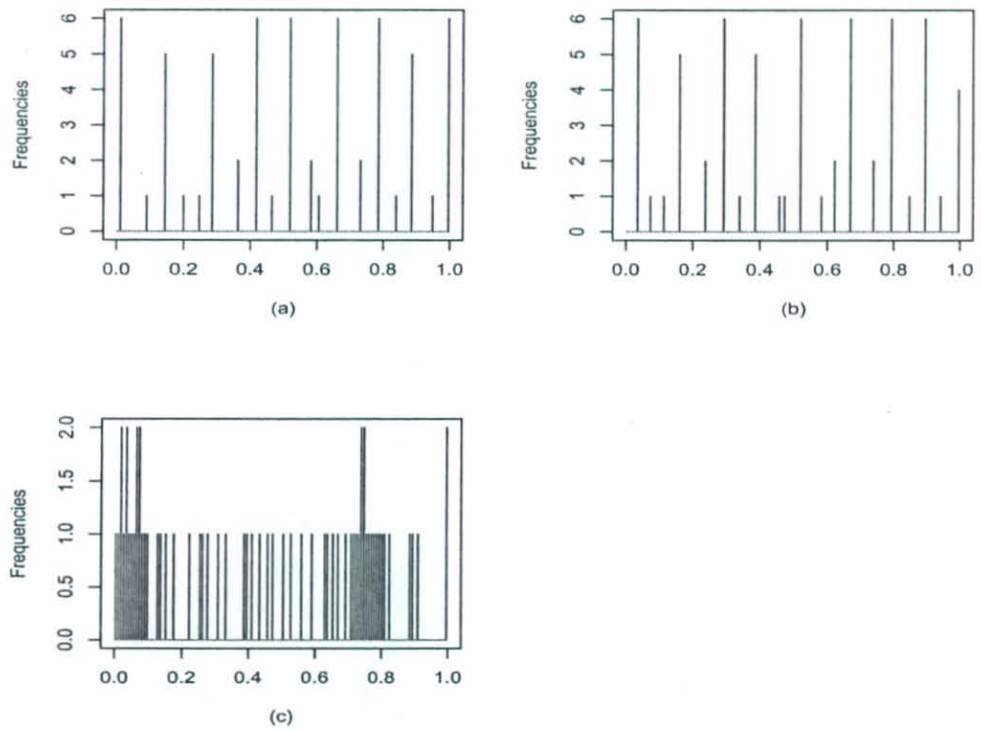

Figure 4.25: Minimax design points: (a) OLS (b) WLS (c) GLS when $N = 128$, $n = 64$, $m = 2$, DAUB4 and, $v = 50$

Chapter 5

Concluding Remarks

In this thesis, we considered minimax designs for estimation of nonparametric regression models with autocorrelated errors using wavelet approximation of the mean response function. We assumed that given the observed responses, as experimenter will estimate the parameters of the wavelet approximation by wavelet versions of ordinary least squares (OLS), weighted least squares (WLS) and generalized least squares (GLS). Based on this assumption we developed a simulated annealing algorithm to search for minimax designs under an AR(1) correlation structure.

In summary, the GLS method performed better than OLS and WLS methods, when $|\rho| \leq 0.6$, where ρ have been defined in Section 2.5 for AR(1) model. The performance of the OLS and WLS were similar. For practical reason we considered

the range of ρ between -0.99999 and 0.99999 since the stationarity condition for $AR(1)$ process is $|\rho| < 1$. When we have prior knowledge about a given data one can select a suitable method for constructing an optimal design for that particular type of data to improve the performance of the experiment. That is, GLS is the best for the moderate level correlation ($-0.6 < \rho < 0.6$) and WLS or OLS is preferred for highly correlated data.

We note that our techniques can be extended to Moving Average processes of order q to select which method of estimation is the best.

Bibliography

[1] Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developements and their applications. *Statistics Survey* **1**, 16 - 55.

[2] Antoniadis, A., Gregoire, G., and Mckeague, W. (1994). Wavelets methods for curve estimation. *Journal of the American Statistical Association* **89**, 1340 - 1352.

[3] Berenblut, I. I., and Webb, G. I. (1974). Experimental design in the presence of autocorrelated errors. *Biometrika* **61**, 427437.

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

[5] Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* **24**, 586 - 602.

[6] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* **41(7)**, 909 - 996.

[7] Daubechies, I., and Lagarias, J. (1991). Two-scale difference equations I. Existence and global regularity of solutions. *SIAM J. Math. Anal.* **22(5)**, 1388 - 1410.

[8] Daubechies, I., and Lagarias, J. (1992). Two-scale difference equations II. Lobal regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.* **23(4)**, 1031 - 1079.

[9] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D. (1995). Wavelet Shrinkage: Asymptopia?. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 301 - 369.

[10] Elfving, G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics* **23**, 255 - 262.

[11] Elfving, G. (1956). Selection of nonrepeatable observations for estimation. *The Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 69 - 75.

[12] Fang, Z., and Wiens, D.P. (2000). Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models **95**, 807 - 818.

[13] Fedorov, V. V. (1972). *Theory of optimal experiments.* W.J. Studden and E.M. Klimko (*trans. and eds.*). Academic Press, New York.

[14] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. of London, Ser.* A **222**, 309 - 368.

[15] Herzberg, A. M. and Traves, W. N. (1994). An optimal experimental design for the Haar regression model. *Canadian Journal of Statistics* **22(3)** , 357 - 364.

[16] Huber, P.J. (1975). *Robustness and designs. In: A Survey of statistical design and linear models.* J. N. Srivastava *(ed.)*, North Holland, Amsterdam, 287 - 303.

[17] Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society* **21**, 272 - 319.

[18] Kiefer, J., and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics* **32**, 271 294.

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

[20] Li, K.C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12**, 269 - 282.

[21] Li, K.C., and Notz, W. (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135 - 151.

[22] Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(\Re)$. *Transactions of the American Mathematical Society* **315**, 69 - 87.

[23] Marcus, M.B., and Sacks, J. (1976). *Robust designs for regression problems. In statistical theory and related topics II.* S.S. Gupta and D.S. Moore *(eds.)*, Academic Press, New York, 245 - 268.

[24] Nason, G.P. (2008). *Wavelet methods in statistics with R.* Springer Science, New York.

[25] Ou, B., and Zhou, J. (2009). Minimax robust designs for field experiments. *Metrika* **69**, 45 - 54.

[26] Oyet, A.J., (2002). Minimax $A-$ and $D-$ optimal integer-valued wavelet designs for estimation. *Canad. J. Statist* **30(2)**, 301316.

[27] Oyet, A.J., (2003). Wavelet designs for estimating nonparametric regression models with heteroscedastic error. *J. Statist. Plann. Inference*

[28] Oyet, A.J., and Wiens, D.P., (2000). Robust designs for wavelet approximations of regression models. *Journal of Nonparametric Statistics* **12**, 837-859.

[29] Oyet, A.J., and Wiens, D.P., (2002). On exact minimax wavelet designs obtained by simulated annealing. *Statistics and Probability Letters* **61**, 111-121.

[30] Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in $\Re^k$. *Ann. Statist.* **10**, 511 - 525.

[31] Pinheiro, A., and Vidakovic, B. (1997). Estimating the square root of a density via compactly supported wavelets. *Computational Statistics and Data Analysis* **25**, 399 - 415.

[32] Plackett, R.L., and Burman, J.P., (1946). The design of optimum multifactorial experiments. *Biometrika* **33**, 305 - 325.

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

[34] Sacks, J., and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6**, 1122 - 1137.

[35] Shah, K. R., and Sinha, B. K. (1989). *Theory of optimal designs.* Lecture Notes in Statistics, **54**, Springer-Verlag, New York.

[36] Silvey, S.D. (1980). *Optimal design (Monographs on Applied Probability and Statistics).* Chapman & Hall, London-New York.

[37] Smith, K., (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial functions and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* **20**, 1 - 85.

[38] Strang, G. (1989). Wavelets and dilation equations: a brief introduction. *SIAM Review* **31(4)**, 614 - 627.

[39] Vidakovic, B., (1999). *Statistical Modelling by Wavelets.* Wiley, New York.

[40] Wegman, E.J., Poston, W.L., and Solka, J.L., (1996). Wavelets and nonparametric function estimation. *Research Developments in Probability and Statistics*, VSP, Utrecht, pp 257 - 274.

[41] White, L. V. (1973). An extension to the general equivalence theorem for nonlinear models. *Biometrika* **60**, 345 - 348.

[42] Wiens, D.P. (1990). Robust minimax designs for multiple linear regression. *Linear Algebra Appl* **127**, 327340.

[43] Wiens, D.P., and Zhou, J. (1996). Minimax regression designs for approximately linear models with autocorrelated errors. *Journal of Statistical Planning and Inference* **55**, 95 - 106.

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

[45] Wiens, D.P., and Zhou, J. (1999). Minimax designs for approximately linear models with $AR(1)$ errors. *The Canadian Journal of Statistics* **27**, 781 - 794.

[46] Wiens, D.P., and Zhou, J. (2008). Robust estimators and designs for field experiments. *Journal of Statistical Planning and Inference* **138**, 93 - 104.

[47] Zhou, J., (2001). Integer-valued, minimax robust designs for approximately linear models with correlated errors. *Communication in Statistics*: Theory Methods **30**, 21 - 39

[48] Zhou, J., (2001). two robust design approaches for linear models with correlated errors. *Statistica Sinica* **11**, 261 - 272.