UNIVERSITY OF ALBERTA

Optimal Designs for Approximately Polynomial Regression Models

by

Giseon Heo

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

in

Statistics

Department of Mathematical Sciences

Edmonton, Alberta

Fall 1998

0-612-34778-8

Abstract

The contents of this dissertation may be divided into two cases. In Chapters 2 and 3 we construct robust designs for polynomial regression functions that may be contaminated by higher degree polynomials. The contaminating space is denoted by $\mathcal{F}$. In this setting we find minimax designs with a minimal number of support points. Such designs may not have enough support points to fit the models against which you wish to protect. We provide some guidelines to the experimenter who looks for good designs that also protect against the alternative model.

In Chapter 4 we consider approximately polynomial regression models where the true model is unspecified. The unspecified contaminating space is an L_2-type space and denoted by $\mathcal{F}_2$. For such problems, Wiens (1990, 1992) extends Huber (1975)'s minimax approach to simple linear regression as well as to bivariate linear regression. Although Wiens's minimax approach is not so straightforward to extend to higher degree polynomial regression functions, we construct a minimax design for an approximately quadratic polynomial regression model without constant term. For higher degree polynomial models we restrict to densities that are easy to work with and construct optimal designs. We compare these optimal designs to the minimax designs that are constructed by Wiens (1990) and to the optimal designs for an approximately quadratic regression model without constant term.

Dedicated

To my parents, my sisters – Namseon, Jiyeon, and Junam,

my husband Byron, and my daughter Grace Serome.

ACKNOWLEDGEMENTS

My sincere appreciation goes to my supervisor Professor D. P. Wiens who has guided me for many years and also I deeply thank him for summer financial support and travel grants. I also wish to acknowledge with thanks the members of my examining committee Dr. C. A. Field, Dr. R. J. Karunamuni, Dr. D. H. Kelker, Dr. S. J. Titus, and Dr. Y. Wu, for their suggestions and their time.

I thank my professors, Dr. Al-Hussaini, Dr. Elliott, Dr. Gombay and Dr. Hooper who taught not only Statistics but also indirectly their wisdom.

My deep appreciation goes to my colleagues, Naser Abd-Rabou, Emmanuel Benhin, Zhide Fang, Karla Nobrega, Alwell Oyet, Thuan Thach, and Eshetu Wondmagegnehu with whom I shared encouragement and humor.

I do not forget to give special thanks to Marion Benedict, Linda Drysdale, Christine Fischer, Laura Heiland and Patti Macdonald in our department for their kind words and consistent help throughout my school years.

Without the faith and unconditional love of my parents, my sisters, my husband, and my daughter, this work could not have been completed.

Contents

List of Tables

List of Figures

Conventions:

With the exception of the introduction and chapter 4, in our thesis the independent variable x for the regression problem assumes values anywhere in the interval $[-1/2, 1/2]$. For any choice of design points $(x_i)_{i \leq n}$ in this interval, the corresponding design measure is defined as $(1/n)\sum_{i=1}^{n}\delta_{x_i}$. However, our loss functions and optimization problem extend in a natural way to arbitrary distributions on $[-1/2, 1/2]$, and so we define a *design measure* to be any probability measure ξ over $[-1/2, 1/2]$. Our point of view is that the design is only a guide as to the placement and relative frequency of the design points. In practice, any design ξ will have to be approximated by an implementable design of the type $(1/n)\sum_{i=1}^{n}\delta_{x_i}$, via randomization for example. An important result of our convention is that the number of observations n and the design ξ are independent, and so n may be treated as a constant in the minimization over ξ.

Notation:

1. Unless otherwise noted, μ and ξ denote arbitrary probability measures on $[-1/2, 1/2]$, and all integrations in the x variable are over the range $[-1/2, 1/2]$. The moments of any measure will be denoted using subscripts, that is, $\xi_i = \int x^i \xi(dx)$ and $\mu_i = \int x^i \mu(dx)$ for all $i \geq 0$.

2. For a function $f : [-1/2, 1/2] \to \mathbb{R}$, we will sometimes use the shorthand $E_\xi f = \int f(x)\,\xi(dx)$. In addition, the symbol x may occasionally denote the identity function on $\mathbb{R}$; for example, xf refers to the function $x \mapsto xf(x)$.

3. We may use either $\mathrm{tr}(A)$ or $\mathrm{trace}(A)$ to denote the trace of a matrix, while $\det(A)$ or $|A|$ means its determinant. The eigenvalues of A are denoted λ and $\lambda_{\max}(A)$ means the largest eigenvalue of A.

Chapter 1

Literature review and objectives

1.1 Introduction

In the study of a dose-response relationship, it is common for an experimenter to apply equally spaced dosage levels in his experiment. But if he knows that the relation of dose and response is linear, then it would be more effective to apply minimum and maximum dosage only. To understand this, we must carefully distinguish between two types of error that can occur in such experiments : bias error, due to inadequacy of the model, and variance error, due to sampling. An equally spaced design is optimal when we are only concerned with the bias error, while the classical optimal design, which makes observations at the minimum and maximum values only, is optimal when we are only concerned about the variance error. In practice, it is realistic to assume that the model function is known only approximately and to use a design that minimizes a combination of these error terms. The goal of this thesis is to provide guidelines on choosing such designs.

1.2 Classical optimal designs

Many prairie farmers in Alberta have traditionally stocked dugouts with trout to provide summertime fishing and fun. but now a few are attempting to make a business out of culturing fish indoors. year-round. One problem that these farmers are facing is limited water supplies. so it is economical if the farmers can recycle waste water for raising fish. Most solids in wastewater from trout-rearing facilities settle readily. but a suspension of fine material remains. Several studies have shown that fine particulate adversely affects fish health and productivity. The wastewater engineering research team at Alberta Environmental Centre conducted a bench-scale experiment to find out the amount of total suspended solid (TSS) after applying ozone application rates (O_3) ranging from 0 to 2 mg/L. (see Heo and James (1995)). Because ozonation is to be used for disinfection and the associated capital cost is high. the team wants to determine O_3 rate in an optimal manner — minimizing the worst cost. If the research team knows by their previous experience that TSS and O_3 rate are linearly related. the team only needs to apply the lowest rate (no ozonation) half of the times and the highest rate half of the times in order to obtain the optimal design. This is quite unexpected to the layman who would normally expect that one should apply O_3 rates spread evenly over the range 0 to 2 mg/L. When observations (TSS) are subject to experimental error. the n observations are given by

$$y(x_i) \quad = \quad \theta_0 + \theta_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where x_i is the O_3 rate and $y(x_i)$ is the amount of TSS at rate $x_i \in [0, 2]$. We assume the error terms ε_i are uncorrelated with mean zero and variance σ^2. The parameters θ_0 and θ_1 are unknown regression coefficients. We will use the least squares estimates $\hat{\theta}_1 = \sum(x_i - \bar{x})(y_i - \bar{y})/\sum(x_i - \bar{x})^2$ and $\hat{\theta}_0 = \bar{y} - \hat{\theta}_1\bar{x}$. Our problem is to choose x_i's to improve the quality of this estimate $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1)^T$. One measure of how well the least squares line fits is the integrated mean squared error (IMSE).

$$\begin{aligned}\text{IMSE} &= \int_0^2 E(\theta_0 + \theta_1 x - \hat{\theta}_0 - \hat{\theta}_1 x)^2\, dx \\ &= \sigma^2\left(\frac{2}{n} + \frac{6(\bar{x}-1)^2 + 2}{3\sum(x_i - \bar{x})^2}\right).\end{aligned}$$

To minimize IMSE, the x_i's ought to be as spread out as much as possible, in other words, half of them at the lowest x-value and the other half at the highest x-value. This example demonstrates that we can improve the quality of our estimates by planning the locations of the x-values, rather than just choosing them haphazardly. We generalize the above statistical model to a regression problem with multiple variables as follows.

$$y(\boldsymbol{x}_i) = \boldsymbol{z}^T(\boldsymbol{x}_i)\boldsymbol{\theta} + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{1.2.1}$$

where the regressor $\boldsymbol{z}(\boldsymbol{x}) \in \mathbb{R}^p$ is a given function of $\boldsymbol{x}$. The design points $\boldsymbol{x}_i$ are confined to a subset $S \subset \mathbb{R}^q$, which we call the design space. The parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ is unknown, and the error terms ε_i are uncorrelated with mean zero and common variance σ^2.

The goal in the regression problem is to estimate the parameter $\boldsymbol{\theta}$.

1. One part of that problem is to select an estimator. Throughout we will simply use the least squares estimator $\hat{\boldsymbol{\theta}}$.

2. Another part of the problem is to choose the design points $\boldsymbol{x}_i$, $i = 1, \ldots, n$ in an optimal manner. This is equivalent to choosing their empirical measure $\xi := \frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{x}_i}$ on S.

For any potential design measure ξ, we define the following matrix,

$$A_\xi = \int \boldsymbol{z}(\boldsymbol{x})\boldsymbol{z}^T(\boldsymbol{x})\,\xi(d\boldsymbol{x}), \tag{1.2.2}$$

which is often called the per-observation information matrix of the design ξ. This matrix is related to the covariance of our estimator by

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = (\sigma^2/n)A_\xi^{-1}.$$

The classical optimality problem is to choose ξ to minimize a particular loss function. Examples of loss functions include the determinant, the trace, and the largest eigenvalue of the matrix A_ξ^{-1}, and these give the D-, A- , E-optimality criterion, respectively. In the fish example above, where $\boldsymbol{z}^T(x) = (1, x)$ and $S = [0, 2]$, we have

$$\begin{aligned} \det(A_\xi^{-1}) &= 1/(\xi_2 - \xi_1^2) \\ \mathrm{tr}(A_\xi^{-1}) &= (1 + \xi_2)/(\xi_2 - \xi_1^2). \end{aligned}$$

Using these expressions it is not hard to see that the design with half its design points at 0 and the other half at 2 is D-optimal, but not A-optimal.

It has been observed by Atkinson (1982) that although the loss functions for D-optimality and IMSE-optimality are invariant under scale changes in the x

variable, both A- and E-optimality suffer from the theoretical disadvantage that they are not invariant. Nevertheless it is usual to scale quantitative factors to lie between -1 and +1. In our work, we will always assume that the data have been scaled so that x lies between $-1/2$ and $+1/2$. In our notation, this means taking $S = [-1/2, 1/2]$.

With $S = [-1, 1]$, and $\boldsymbol{z}^T(x) = (1, x, \ldots, x^k)$, Hoel (1958) shows that the D-optimal design puts equal mass at the zeros of the polynomial $(1 - x^2)P_k'(x)$, where $P_k'(x)$ denotes the derivative of the kth Legendre polynomial $P_k(x)$ on $[-1, 1]$. If the experimenter wants to investigate $\boldsymbol{z}^T(x)\boldsymbol{\theta}$, the performance of a design ξ can be measured by the standardized variance $d(x; \xi) := \boldsymbol{z}^T(x)A_\xi^{-1}\boldsymbol{z}(x)$ of the optimal estimator $\boldsymbol{z}^T(x)\hat{\boldsymbol{\theta}}$. A design is called G-optimal if it minimizes $\max_{x\in S} d(x; \xi)$. Kiefer and Wolfowitz (1960) proved that a design is D-optimal if and only if it is G-optimal. For the readers who are interested in the classical optimal designs, we refer to Fedorov (1972), Silvey (1980), and Pukelsheim (1993).

1.3 Model robust designs

In the fish example, the experimenter assumes that TSS and O_3 are linearly related. What if he is mistaken about this relation? If TSS and O_3 are not exactly linearly related, then the optimal design cannot possibly detect the presence of any non-linear term in the regression function, no matter how large the sample size.

The classical optimal designs have the disadvantage that they are extremely model dependent. These designs provide no opportunity to check the model's adequacy. Box and Draper (1959) seem to have been the first to be concerned about the dangers of assuming that the model is known exactly when designing a regression experiment. They studied the case where the experimenter fits a polynomial of first degree whereas the true response is quadratic. The estimate is subject to both "bias error" due to the inadequacy of the linear function, as well as "variance error" due to sampling. They reach a somewhat unexpected conclusion that the designs minimizing bias alone are closer to minimizing both bias and variance than the designs minimizing variance alone.

In many practical situations, the model function is known only approximately so it is more realistic to consider a perturbed model

$$y(\boldsymbol{x}) = \boldsymbol{z}^T(\boldsymbol{x})\boldsymbol{\theta} + f(\boldsymbol{x}) + \varepsilon, \tag{1.3.1}$$

where f is an unknown perturbation function that lies in a contaminating space $\mathcal{F}$. The estimator $\hat{\boldsymbol{\theta}}$ is no longer unbiased, and the mean squared error matrix is, with $b(f,\xi) = \int_S \boldsymbol{z}(\boldsymbol{x})f(\boldsymbol{x})\,\xi(d\boldsymbol{x}) = (1/n)\sum \boldsymbol{z}(\boldsymbol{x}_i)f(\boldsymbol{x}_i)$, given by

$$\begin{aligned} \text{MSE}(f,\xi) &= E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \qquad (1.3.2) \\ &= (\sigma^2/n)A_\xi^{-1} + A_\xi^{-1}b(f,\xi)b^T(f,\xi)A_\xi^{-1}. \end{aligned}$$

The mean squared error (1.3.2) consists of two terms — the first is a variance error and the second is a bias error.

An optimal design obtained under the model (1.2.1) no longer is appropriate under the perturbed model (1.3.1). The design for the first problem makes the variance term $(\sigma^2/n)A_\xi^{-1}$ small but not the bias term $A_\xi^{-1}b(f,\xi)b^T(f,\xi)A_\xi^{-1}$. For any contaminating function f we could construct a design that minimizes MSE(f,ξ) for that specific f. However we are not looking for a design which is optimal for one f, but rather we are looking for a design which is reasonable for all $f \in \mathcal{F}$, in other words, a *robust* design. A design is said to be optimal for the perturbed model (1.3.1) if it minimizes the maximum loss over f. Beginning with Box and Draper (1959), designs for versions of (1.3.1) have been constructed in a large number of papers. These differ in the class $\mathcal{F}$, the design space S, the regressors $\boldsymbol{z}(\boldsymbol{x})$, and in the loss functions used. In the following subsections, we review some of these papers that have been classified into three groups for our convenience. The first two groups discuss L_2 type designs and L_∞ type designs which are named after the norms used in the bound on the disturbance term f. Prior to running the experiment, it is very common that the experimenter has little information about the regression model, $\boldsymbol{z}^T(\boldsymbol{x})\boldsymbol{\theta} + f(\boldsymbol{x})$. The experimenter might have, however, several possible models for the regression function, so it is plausible to put a weight on the different models. Läuter (1974, 1976) took this into account, proposed the generalized D-optimality criterion and proved an equivalence theorem similar to that of Kiefer-Wolfowitz. The third group consists of articles in which Läuter's idea was applied, and hence these designs are called Läuter type designs.

1.3.1 L_2 type designs

Huber (1975, 1981) and Wiens (1990, 1991, 1992, 1993, 1994, 1996) take the contaminating space $\mathcal{F}_2$

$$\mathcal{F}_2 = \{f : \|f\|_2^2 = \int_S f(\boldsymbol{x})^2\, d\boldsymbol{x} \le \eta^2, \int_S \boldsymbol{z}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} = \boldsymbol{0}\}, \quad (1.3.3)$$

where the radius η is assumed known. The first condition in the definition of $\mathcal{F}_2$ allows for finite bias and the second condition ensures the identifiability of $\boldsymbol{\theta}$. Huber restricts to symmetric designs on $[-1/2, 1/2]$ and obtains the minimax designs for the integrated MSE as a loss function. His optimal design measure ξ has a density of the form $\xi' = (ax^2 + b)^+$, where $a > 0$ and b depend on the ratio $(\sigma^2/n\eta^2)$. As $(\sigma^2/n\eta^2) \longrightarrow 0$, the limiting distribution is uniform ($\xi' = 1$) and as $(\sigma^2/n\eta^2) \longrightarrow \infty$, the limiting distribution is a measure that puts all its mass at the extreme points, i.e., $\xi = (\delta_{-1/2} + \delta_{1/2})/2$.

Wiens (1990) extends Huber's result to the case of multiple linear regression, $\boldsymbol{z}(\boldsymbol{x}) = (1, x_1, \ldots, x_p)^T$, where S is a sphere of unit volume in $\mathbb{R}^p$. He also gave robust designs for a bivariate model with interaction, $\boldsymbol{z}(\boldsymbol{x}) = (1, x_1, x_2, x_1 x_2)^T$, $S = [-1/2, 1/2] \times [-1/2, 1/2]$.

Under the approximately linear regression model , $\boldsymbol{z}^T(\boldsymbol{x})\boldsymbol{\theta} = \theta_0 + \sum_{j=1}^p \theta_j x_j + f$, with the contaminating space $\mathcal{F}_2$ in (1.3.3), Wiens (1992) constructs designs that minimize the maximum loss over f. He considers any loss function $\mathcal{L}(f, \xi)$ satisfying

(i) Monotonicity: If MSE(f_1, ξ) − MSE(f_2, ξ) is nonnegative definite, then $\mathcal{L}(f_1, \xi) \ge \mathcal{L}(f_2, \xi)$;

(ii) Unboundedness: $\mathcal{L}(f_n, \xi) \to \infty$, if $\lambda_{\max}(\mathrm{MSE}(f_n, \xi)) \to \infty$, as $n \to \infty$.

He searches for a design ξ^* that satisfies

$$\sup_{\mathcal{F}_2} \mathcal{L}(f, \xi^*) = \inf_{\Xi} \sup_{\mathcal{F}_2} \mathcal{L}(f, \xi)$$

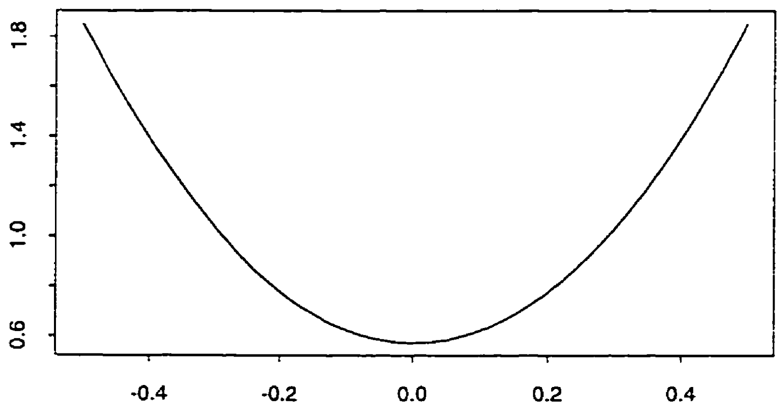
where Ξ is the space of all probability measures on S. He also proves that in order that $\sup_{\mathcal{F}_2} \mathcal{L}(f, \xi)$ be finite, ξ must be absolutely continuous.

In the case that the fitted response is a plane, and the design space is a sphere of unit volume in $\mathbb{R}^p$, Wiens presents explicit designs corresponding to the following five loss functions:

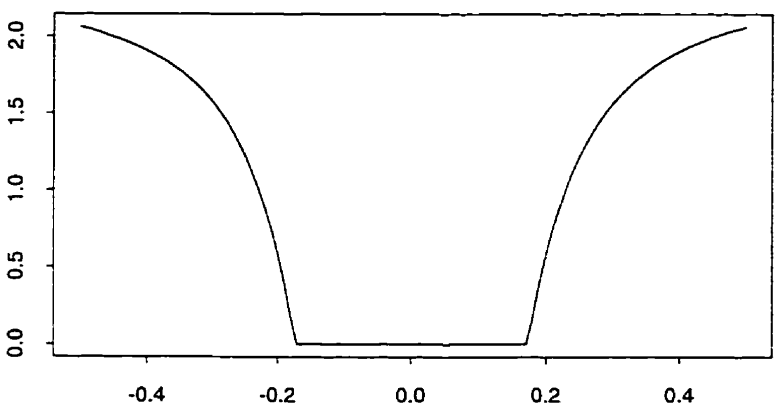
$$\begin{aligned}
\mathcal{L}_D &= \det(\mathrm{MSE}(f, \xi)) && D\text{-optimality,} \\
\mathcal{L}_A &= \mathrm{trace}(\mathrm{MSE}(f, \xi)) && A\text{-optimality,} \\
\mathcal{L}_E &= \lambda_{\max}(\mathrm{MSE}(f, \xi)) && E\text{-optimality,} \\
\mathcal{L}_Q &= \textstyle\int_S d(\boldsymbol{x}; f, \xi)\, d\boldsymbol{x} && Q\text{-optimality,} \\
\mathcal{L}_G &= \sup_{\boldsymbol{x} \in S} d(\boldsymbol{x}; f, \xi) && G\text{-optimality.}
\end{aligned}$$

See Figure 1.1 for the optimal design densities in the case of D- and A-optimality criteria with $p = 1$ and $\nu := (\sigma^2/n\eta^2) = 1$.

Figure 1.1: D- and A-optimal design densities for straight line regression, minimax in contaminating space $\mathcal{F}_2$ as in (1.3.3) with $\nu = 1$: (a) D-density$= 5.12x^2 + 0.573$; (b) A-density$= (2.345 - 0.07/x^2)^+$.



(a)



(b)

1.3.2 L_∞ type designs

Marcus and Sacks (1976) and Li and Notz (1982) made the criticism that using the contaminating space $\mathcal{F}_2$ means that no discrete design, and hence no implementable design, can have finite loss. They, as well as Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li (1984), Liu and Wiens (1997), take a smaller contaminating space

$$\mathcal{F}_\infty = \{ f : |f(\boldsymbol{x})| \leq \phi(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in S\}, \tag{1.3.4}$$

with various assumptions about ϕ.

Marcus and Sacks (1976) take $S = [-1, 1]$ and $\boldsymbol{z}^T(x) = (1, x)$, and let ϕ be a given function on S with $\phi(0) = 0$. They then look for designs that minimize

$$\sup_{\mathcal{F}_\infty} E[(\hat{\theta}_0 - \theta_0)^2 + b(\hat{\theta}_1 - \theta_1)^2],$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ denote the estimates of θ_0 and θ_1, and b is a specified constant. For instance, if $\phi(x) \geq mx$, then the unique optimal design has support only on the points $\{-1, 0, 1\}$. If ϕ is convex, the best design is supported on two points $\{-z, z\}$, where z depends on ϕ and b.

Li and Notz (1982) extend the work of Marcus and Sacks (1976) to the multivariate case where $\boldsymbol{z}(\boldsymbol{x}) = (1, x_1, x_2, \ldots, x_p)^T$, and

$$\mathcal{F}_\infty = \{ f : \ |f(\boldsymbol{x})| \leq c, \int_S \boldsymbol{z}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} = 0\}.$$

When the estimates $\hat{\theta}_i$ are linear (but are not necessarily the LSE) and the designs are restricted to have finite support, they show that the designs that minimize the weighted MSE, with weights (b_i),

$$\sup_{\mathcal{F}_\infty} E[(\hat{\theta}_0 - \theta_0)^2 + \sum b_i(\hat{\theta}_i - \theta_i)^2],$$

have support on the extreme points of $S \subset I\!R^p$.

Pesotchinsky (1982) also extends the results of Marcus and Sacks. He considers $(p+1)$ dimensional linear regression and for the construction of the optimality criteria he uses the Φ_k-family. More precisely,

$$y(\boldsymbol{x}_i) = \theta_0 + \sum_{j=1}^{p} \theta_j x_{ij} + f(\boldsymbol{x}_i) + \varepsilon_i,$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$, ϕ is a convex function, and he assumes that $|f(\boldsymbol{x})| \leq \phi(\|\boldsymbol{x}\|^2)$. For $0 < k < \infty$, his optimality functionals $\Phi_k(f, \xi)$ are derived from the MSE matrix via

$$\Phi_k(f, \xi) = \left\{ \frac{\text{trace}[\text{MSE}(f, \xi)^k]}{p+1} \right\}^{1/k} = \left\{ \frac{1}{p+1} \sum_{j=0}^{p} \lambda_j^k(f, \xi) \right\}^{1/k},$$

where $\lambda_0(f, \xi) \leq \lambda_1(f, \xi) \leq \ldots \leq \lambda_p(f, \xi)$ are the eigenvalues of $\text{MSE}(f, \xi)$. Then

$$\Phi_0(f, \xi) = \lim_{k \to 0^+} \Phi_k(f, \xi) = \{\det(\text{MSE}(f, \xi))\}^{1/(p+1)},$$

and

$$\Phi_\infty(f, \xi) = \lim_{k \to \infty} \Phi_k(f, \xi) = \max_{0 \leq j \leq p} \{\lambda_j(f, \xi)\} = \lambda_{\max}(\text{MSE}(f, \xi)).$$

Thus Φ_0, Φ_1, and Φ_∞ give the D-, A-, and E-optimality criteria, respectively. Pesotchinsky applies a minimax approach and defines a Φ_k- optimal design $\xi_k^*(\phi)$ as one that minimizes the supremum of $\Phi_k(f, \xi)$ over $\{f : |f(\boldsymbol{x})| \leq \epsilon\phi(\|\boldsymbol{x}\|^2)\}$. He proved the following two facts in the class $\Xi(m)$ of all symmetric designs ξ

with fixed $E_\xi(\boldsymbol{x}_i^2) = m$. First, any symmetric design $\xi \in \Xi(m)$ supported only by the points of the sphere S_r of radius $r = \sqrt{mp}$ is D-optimal in $\Xi(m)$ if $\phi(\|\boldsymbol{x}\|^2)$ is convex. Secondly, A- and E-optimal symmetric designs are unique and correspond to the uniform continuous measures on appropriate spheres.

Li (1984) studies robust regression using a design space S consisting of finitely many points, symmetrically distributed on the interval $[-1/2, 1/2]$.

Liu and Wiens (1997) study the regression model $E(Y \mid x) = \sum_{j=0}^{p-1} \theta_j x^j + x^p \psi(x)$. Here $S = [-1, 1]$ and ψ is unknown but $|\psi(x)| \leq \phi(x)$, where ϕ is known. With three criteria

(1) Choose ξ to maximize $\det(A_\xi)$, subject to bounding the normalized bias.

(2) Choose ξ to minimize the maximum bias, subject to bounding the variance.

(3) Choose ξ to minimize the maximum determinant of the MSE matrix.

the optimal designs are given for $p = 2, 3$ for general ϕ and for $p \geq 4$ if ϕ is constant. When $p = 2$ and $\phi(x) \equiv 1$ the optimal design corresponding to (3) is $(\delta_{\min(1,\sigma/\sqrt{n})} + \delta_{-\min(1,\sigma/\sqrt{n})})/2$. The L_∞ designs are generally supported on a small number of points and thus do not allow the exploration of models larger than the fitted ones.

1.3.3 Läuter type designs

Dette (1990, 1991, 1992) worked extensively on polynomial models under the Läuter type of criterion. Before reviewing his papers, we need to introduce the D_s-optimality criterion, which is useful when only a subset of the parameters is of interest. We first need to look at the papers written earlier by Stigler (1971), Atwood (1971), Studden (1982), and Cook and Nachtsheim (1982).

Let $z^T(x)\theta = (z_1^T(x)\theta_1, z_2^T(x)\theta_2)$, where θ_2 contains the parameters of interest. The corresponding information matrix A_ξ is split into block matrices

$$A_\xi = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \text{ where } A_{22} \text{ is } s \times s. \tag{1.3.5}$$

The covariance matrix of θ_2 is proportional to Σ^{-1}, where $\Sigma = A_{22} - A_{21}A_{11}^{-1}A_{12}$. A D_s-optimal design maximizes $\det(\Sigma)$. But $\det(A) = \det(A_{11})\det(\Sigma)$, thus this corresponds to maximizing the ratio $\det(A)/\det(A_{11})$.

Since the classical designs and the minimum bias designs do not allow for model adequacy checking, Stigler (1971) sought a criterion which enables us to check whether the model is appropriate, and to provide efficient inferences about the model if it is appropriate. He introduced C-restricted D-optimal designs and C-restricted G-optimal designs, for kth degree polynomial regression on $[-1, 1]$. The C-restricted D-optimal design for the kth degree polynomial is one which maximizes $\det[A_\xi]$ among all the designs ξ satisfying $\det[A_\xi(k)] \leq C\det[A_\xi(k+1)]$. The C-restricted G-optimal design for the kth degree polynomial is one which minimizes $\max_{-1\leq x\leq 1} d(x, \xi)$ among all the designs ξ satisfying $\det[A_\xi(k)] \leq C\det[A_\xi(k+1)]$. For $C \geq 4$, the C-restricted D- and G-optimal designs for the linear model is given by

$$\xi(-1) = \xi(1) = 1/4 + (1/2)\sqrt{1/4 - 1/C}, \quad \xi(0) = (1/2) - \sqrt{1/4 - 1/C}.$$

We split the regression function $z^T(x)\theta$ into two parts as we did in (1.3.5), $z^T(x)\theta := z_1^T(x)\theta_1 + z_2^T(x)\theta_2 = \sum_{i=0}^{s} \theta_i x^i + \sum_{i=s+1}^{k} \theta_i x^i$. The objective of Atwood (1971) is to derive good estimators and designs for estimating the regression function $z^T(x)\theta$. As an estimator, he used a weighted average of the best linear unbiased estimators of the degree s polynomial and the degree k polynomial and as a design, he used a combination $(1-a)\,\xi(s) + a\,\xi(k)$ of the known optimal designs $\xi(s)$ and $\xi(k)$ for the polynomials of degree s and k, respectively. In his derivation, when $k = s+1$ or $k = s+2$, Atwood applied the fact that the best degree s polynomial approximating the regression function $z^T(x)\theta$ is the Chebyshev (or Zolotarev) polynomial of degree k.

Studden (1982) applies Stigler's technique in a general setting. The model is assumed to be an rth degree polynomial but the coefficients of the higher powers might not be zero. So the regressor $z^T(x)$ is decomposed into two parts, $z_1^T(x) = (1, x, \ldots, x^r)$, $z_2^T(x) = (x^{r+1}, \ldots, x^m)$. So then the covariance matrix of $\hat{\theta}_{r+1}, \ldots, \hat{\theta}_m$ is proportional to Σ^{-1}. He formulated the D_{rm} problem which is to maximize $\det(A_{11})$ subject to a bound on the determinant of Σ. The D_{1m} and D_{2m}-optimal designs are obtained in terms of the canonical moments.

The work of Cook and Nachtsheim (1982) was motivated by a problem concerning the estimation of uranium content in calibration standards. The regression function can be approximated by a polynomial of a finite degree, but the degree is not known in advance. Since the experimenter was certain that a polynomial of degree less than equal to six would be an adequate model, they applied Läuter's idea to the integrated variance criterion and presented the iterative methods for design construction.

Dette (1990) assumed that the unknown model belongs to the class of polynomials $\mathcal{P}_k = \{\, g_j \,|g_j = \sum_{i=0}^{j} \alpha_{ji}\, x^i,\ j = 0, \ldots, k \,\}$ and determined designs that do well for each member of $\mathcal{P}_k$. He not only considered all the information matrices $(A_\xi(j))_{j=0}^k$, but he also introduced a family of priors $w = (w_j)_{j=0}^k$.

A design ξ is optimal for $\mathcal{P}_k$ with respect to the prior w if ξ maximizes the function $\Psi_w(\xi) = \sum_{j=0}^k \frac{w_j}{j+1} \log[\det(A_\xi(j))]$. The support points of the optimal designs are given by ± 1 and the zeros of a Jacobi polynomial. The masses at the interior support points are equal but the masses at the boundary points ± 1 are somewhat larger. The D- and D_1-optimal designs are identified as special cases. The one dimensional results are generalized to multivariate polynomial regression on the q-cube.

In polynomial regression models with Läuter's optimality criterion, Dette (1991) identifies robust designs given by Stigler (1971), Studden (1982) and Cook and Nachtsheim (1982), as D-optimal designs in the sense of Läuter.

Dette (1992) dealt with a situation where extra information about the model could be given. For instance, the experimenter is quite sure that the degree of the polynomial model which has to be fitted is even (or odd) and could provide the upper bound of the degree, $k = 2r$ (or $k = 2r - 1$), say. He chooses the loss functions to be $\sum_{l=1}^r w_l \log[\det A_{2l}(\xi)/\det A_{2l-1}(\xi)]$ in the case of polynomial models of even degree and $\sum_{l=1}^r w_l \log[\det A_{2l-1}(\xi)/\det A_{2l-2}(\xi)]$ in the case of polynomial models of odd degree. For the class of odd degree polynomials, with $w_l = 1/r$, the D_1-optimal design puts equal masses at the zeros of the polynomial $(1-x^2)[C_{r-1}^{(3/2)}(T_2(x)) + C_{r-2}^{(3/2)}(T_2(x))]$, where $C_l^{(a)}(x)$, $a > -1/2$, denotes the lth ultraspherical polynomial which is the lth orthogonal polynomial with respect to the measure $(1-x^2)^{a-1/2}\,dx$ and $T_l(x)$ denotes the lth Chebyshev polynomial of the first kind orthogonal with respect to the measure $(1-x^2)^{-1/2}\,dx$. He also considered the case of polynomials with only even (or odd) powers. For the even powers of polynomials, $\mathcal{P}_r = \{\,\sum_{j=0}^{i} \alpha_{ij}\, x^{2j} \mid i = 1,\ldots,r\,\}$, the D-optimal design puts equal masses $1/(2r+2)$ at the zeros of the polynomial $(1-x^2)C_{r-1}^{\prime(3/2)}(T_2(x))$ and mass $1/(r+1)$ at the point 0.

The papers involved with Läuter type designs apply extensively the theory of the canonical moments based on the work of Skibinsky (1968). The interested readers are referred to Dette (1993), Lau and Studden (1985), and Läuter (1976).

1.4 Summary of results

In this dissertation we construct optimal designs for approximately polynomial regression functions. In Chapter 2 and Chapter 3 we construct minimax designs when an experimenter fits a polynomial of degree p although the true model is a polynomial of degree q, $q > p$. In polynomial regression the supremum of the loss function depends on ξ only through its first $p+q$ moments. Searching for an optimal set of set of moments is quite complex, so we look only at discrete measures with the minimal number of support points.

We review a result by Wald (1939) which says that for any probability measure on $[0,1]$, there is a probability measure with $[(s+2)/2]$ or fewer support points, with the same first s moments. In Chapter 3 we offer an independent proof of this result. In fact we show that for any measure μ on $[a,b]$ not supported by p or fewer points, there exists a measure ξ on $[a,b]$ with $p+1$ support points, for which μ and ξ have the same first $2p+1$ moments.

Rychlik (1987) constructs Q-optimal designs for a linear regression function that is contaminated by higher degree polynomials. Applying Rychlik's approach to cases under D- and A-optimality criteria, we construct optimal designs in Chapter 2.

The optimal design depends not only on the form of the loss function but also on how we define "original part" and "contamination part". In Chapter 2, we consider that, even though the experimenter has misjudged the exact nature of the response function, he wants a fitted response function that will be useful in predicting response values in the future. In Chapter 3 we consider that the original model function, a pth order polynomial, has been contaminated by the addition of some higher order terms only. We want the fitted response function to estimate the original model as closely as possible. The two different situations are distinguished by denoting them polynomial model I and polynomial model II.

In this polynomial regression setting the main drawback is that the designs don't have enough support points to fit the models against which we wish to protect. In Chapter 2 we provide a guideline to remedy this situation.

It is not straightforward to apply Huber's or Wiens's minimax technique to an approximately quadratic polynomial regression function. In Chapter 4, using an ad-hoc approach, we construct optimal designs for approximately polynomial regression models. We compare these designs to the minimax designs for bivariate linear model obtained by Wiens (1990). We obtain a minimax density for an approximately quadratic regression model without constant term and compare the minimax design to the ad-hoc optimal design.

Chapter 2

Model robust designs in polynomial regression I

Suppose that an experimenter fits, by least squares, a linear regression model $E(Y \mid x) = \theta_0 + \theta_1 x$, but is concerned that true model might be contaminated by some unknown function f, that is, $E(Y \mid x) = \theta_0 + \theta_1 x + f(x)$. In this situation the experimenter would like to choose design points x_i that yield good estimates of θ_0 and θ_1 while offering some protection against the possible contamination. The optimal placement of design points was found by Huber (1975, 1981) and Wiens (1990, 1992) under the assumption that the contamination function f belongs to the L_2 type space

$$\mathcal{F}_2 = \{f : \int f^2(x)\,dx \leq \eta^2, \int f(x)\,dx = \int xf(x)\,dx = 0\}.$$

For this contaminating space, Huber (Q-optimality) and Wiens (D- and A-optimality) construct robust designs by minimizing, over a space of designs, the maximum loss as f ranges over $\mathcal{F}_2$. However, a criticism of this approach is that the space $\mathcal{F}_2$ is so wide that any discrete design, and hence any implementable design, has infinite maximum loss over $\mathcal{F}_2$. One possible remedy is considered by Rychlik (1987), who restricts the class of contamination functions to the space $\mathcal{F}$, which consists of all qth order polynomials in $\mathcal{F}_2$. Rychlik then constructs a minimax symmetric discrete design for Q-optimality case for the contaminating space $\mathcal{F}$. The main result in Rychlik (1987) is that any symmetric design is minimax if its even moments are identical to the corresponding even moments of Huber's minimax continuous design.

In this chapter we work with three different types of loss functions based on the D-, A-, and Q-optimality criteria. Since the normalized Legendre polynomials form an orthonormal basis in L_2, we can rewrite Rychlik's contamination space as

$$\mathcal{F} = \{\eta \sum_{i=2}^{q} \alpha_i l_i : \sum \alpha_i^2 \leq 1\},$$

where l_i is the normalized Legendre polynomial of degree i. In other words, the fitted model is $E(Y \mid x) = \theta_0 + \theta_1 x$, but the true model is of the form $E(Y \mid x) = \theta_0 + \theta_1 x + \eta \sum_{i=2}^{q} \alpha_i l_i(x)$.

In polynomial regression, the supremum over $\mathcal{F}$ of the loss function depends on the design measure only through its first $(q+1)$ moments. This means that our search for an optimal design measure is really a search for an optimal set of moments. This allows us to exploit Wald's (1939) result giving the minimum number of support points for a probability measure with a fixed set of moments. Our minimization problem is therefore reduced to a search over a finite dimensional subspace of design measures with small support. In Section 2 we prove that for D-optimality, the optimal design is always symmetric. For the Q- and A-optimality cases, it is not known whether the optimal design is symmetric or not, but we prove the existence of an optimal design in the class of symmetric designs.

Adapting Rychlik's method we obtain the minimax symmetric design under the D- and A-optimality criteria.

There are two main results in section 5. The first, Theorem 2.5.1, rephrases Rychlik's main theorem and says that any symmetric design measure ξ is minimax if the expectations under ξ of the first $[(q+1)/2]$ even-order Legendre polynomials are the same as for Wiens's design measure. This is the same as saying that ξ and Wiens's design share the first $[(q+1)/2]$ even moments. Since the moments in Theorem 2.5.1 are generated by a continuous density, they belong to the interior of the moment space. Applying Comment 3 in this chapter, the optimal design ξ can be chosen to have $[(q+1)/2]+1$.

As we mentioned in Chapter 1, the minimax design depends on the ratio $\nu = \sigma^2/n\eta^2$. When $\nu = 0$, only the "bias" term is involved in the minimization, and so the uniform density is optimal. On the other hand, as ν approaches infinity, the "variance" term swamps everything else, and we expect the optimal density to resemble the classical design, that is, all the mass is on the boundary.

The second main result of section 5 is Theorem 2.5.2, which says that for large ν the optimal design coincides exactly with the classical design $(\delta_{-1/2}+\delta_{1/2})/2$. The proof of Theorem 2.5.2 needs only simple algebra but is very lengthy. One of the reasons for the length of the proof is that the cases where q is even or odd must be treated separately.

2.1 Introduction

In this chapter, we consider the case where the experimenter fits a polynomial of order p, but where the true model function is only approximately a polynomial of order p. More precisely, we set $\boldsymbol{z}_1(x) = (1, x, \ldots, x^p)^T$, where the regressor x ranges over the interval $[-1/2, 1/2]$. The experimenter fits, by least squares, the model

$$E(Y \mid x) = \boldsymbol{z}_1^T(x)\,\boldsymbol{\theta}, \quad \boldsymbol{\theta} \in I\!R^{p+1}, \tag{2.1.1}$$

although the true model is

$$E(Y \mid x) = \boldsymbol{z}_1^T(x)\,\boldsymbol{\theta} + f(x), \quad f \text{ is unknown }. \tag{2.1.2}$$

For any choice of design points $(x_i)_{i=1}^n$ our observations will be given by

$$y_i = \boldsymbol{z}_1^T(x_i)\,\boldsymbol{\theta} + f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where we assume additive, uncorrelated errors ε_i with common variance σ^2.

Our fitted response function is a pth order polynomial whose coefficients are given by the usual least squares estimate

$$\hat{\boldsymbol{\theta}} = (nA_\xi)^{-1} \sum_{i=1}^n \boldsymbol{z}_1(x_i)\, y_i.$$

Here $\xi = (1/n)\sum_{i=1}^n \delta_{x_i}$ is the design measure, and $A_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x)\,\xi(dx)$ is the corresponding information matrix.

The definition of optimal design depends on the criteria used to judge the quality of the estimate $\hat{\boldsymbol{\theta}}$. This depends not only on the form of the loss function but also on how we define "original model" and "contamination". There are many ways to split the true response function $E(Y \mid x)$ into a lower order part and a remainder; this choice reflects the purpose of the estimation. Different choices, such as we make here and in Chapter 3, will lead to similar but different optimal designs.

Here in Chapter 2 we imagine that, even though the experimenter has misjudged the exact nature of the response function, he wants a fitted response function that will be useful in predicting future values of y. Accordingly, we define the true coefficient vector $\boldsymbol{\theta}_0$ to be the vector $\boldsymbol{\theta} \in I\!R^{p+1}$ that minimizes

$$\int [E(Y \mid x) - \boldsymbol{z}_1^T(x)\boldsymbol{\theta}]^2 \, dx.$$

Differentiating with respect to $\boldsymbol{\theta}$ we are led to

$$\boldsymbol{\theta}_0 = [\int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x) \, dx]^{-1} \int \boldsymbol{z}_1(x) E(Y \mid x) \, dx. \qquad (2.1.3)$$

(Note that all integrations in the x variable are assumed to be over the range $[-1/2, 1/2]$.) The uniqueness of $\boldsymbol{\theta}_0$ depends on the invertibility of the matrix $\int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x) \, dx$ which is guaranteed by Lemma 2.1.1.

Define the function $f(x)$ to be $E(Y \mid x) - \boldsymbol{z}_1^T(x)\boldsymbol{\theta}_0$, and assume as well that f is a polynomial of degree $q \geq p$. The true response $E(Y \mid x)$ in (2.1.2) then is a polynomial of degree q,

$$E(Y \mid x) = \boldsymbol{z}^T(x)\boldsymbol{\alpha}, \quad \text{where } \boldsymbol{z}(x) = (1, x, \ldots, x^q)^T, \boldsymbol{\alpha} \in I\!R^{q+1}. \qquad (2.1.4)$$

Then using this fact in equation (2.1.3),

$$f(x) = \boldsymbol{z}^T(x)\boldsymbol{\alpha} - \boldsymbol{z}_1(x)^T [\int \boldsymbol{z}_1(x)\boldsymbol{z}_1(x)^T \, dx]^{-1} \int \boldsymbol{z}_1(x)\boldsymbol{z}(x)^T \, dx \, \boldsymbol{\alpha}.$$

Splitting the vector $z(x)$ into two parts, $z_1(x) = (1, x, \ldots, x^p)^T$, and $z_2(x) = (x^{p+1}, \ldots, x^q)$, the contamination function $f(x)$ can be rewritten as follows.

$$f(x) = \left[\begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix} - \begin{pmatrix} z_1(x) \\ [\int z_1 z_2^T\, dx]^T [\int z_1 z_1^T\, dx]^{-1} z_1(x) \end{pmatrix}\right]^T \alpha. \qquad (2.1.5)$$

We now collect the non-zero components of the right hand side of the equation (2.1.5), the function $f(x)$ can be written

$$f(x) = \left(z_2(x) - [\int z_1(x) z_2^T(x)\, dx]^T [\int z_1(x) z_1^T(x)\, dx]^{-1} z_1(x)\right)^T \beta. \quad (2.1.6)$$

with $\beta \in I\!R^{q-p}$. Defining the function

$$u(x) = z_2(x) - [\int z_1(x) z_2^T(x)\, dx]^T [\int z_1(x) z_1^T(x)\, dx]^{-1} z_1(x). \qquad (2.1.7)$$

we can write the true model as "pure part" plus "contaminated part" in the following way,

$$E(Y \mid x) = z_1^T(x)\theta_0 + u^T(x)\beta. \qquad (2.1.8)$$

The quality of the least squares estimate $\hat{\theta}$ will depend on the size of the contamination term $u^T(x)\beta$ and on the placement of the design points. We assume the contamination is small, in the sense that for some known $\eta \geq 0$, the function $u^T\beta$ belongs to

$$\mathcal{F} = \{u^T\beta : \int [u^T(x)\beta]^2\, dx \leq \eta^2\}. \qquad (2.1.9)$$

It is easy to see from the equation (2.1.6) that $\int z_1(x) f(x)\, dx = 0$ which ensures the identifiability of θ_0. For a given loss function $\mathcal{L}$ then, our problem is to choose design points that are robust against the worst possible contamination in $\mathcal{F}$, that is, to minimize $\sup_{\mathcal{F}} \mathcal{L}$. Our loss function will always depend on the support points $(x_i)_{i=1}^n$ only through the design measure $\xi = (1/n)\sum_{i=1}^n \delta_{x_i}$, so we recast our problem in terms of measures on $[-1/2, 1/2]$. Setting $C_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{u}^T(x)\,\xi(dx)$, we can rewrite the estimate as

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= (nA_\xi)^{-1}\sum_{i=1}^n \boldsymbol{z}_1(x_i)y_i \\
&= (nA_\xi)^{-1}\sum_{i=1}^n \boldsymbol{z}_1(x_i)[\boldsymbol{z}_1^T(x_i)\boldsymbol{\theta}_0 + \boldsymbol{u}^T(x_i)\boldsymbol{\beta} + \varepsilon_i] \\
&= (nA_\xi)^{-1}[nA_\xi\boldsymbol{\theta}_0 + nC_\xi\boldsymbol{\beta} + \sum_{i=1}^n \boldsymbol{z}_1(x_i)\varepsilon_i] \\
&= \boldsymbol{\theta}_0 + A_\xi^{-1}C_\xi\boldsymbol{\beta} + (nA_\xi)^{-1}\sum_{i=1}^n \boldsymbol{z}_1(x_i)\varepsilon_i \\
&= \boldsymbol{\theta}_0 + A_\xi^{-1}C_\xi\boldsymbol{\beta} + A_\xi^{-1}\int \boldsymbol{z}_1(x)\varepsilon(x)\,\xi(dx). \qquad (2.1.10)
\end{aligned}$$

Taking expectations gives $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}_0 + A_\xi^{-1}C_\xi\,\boldsymbol{\beta}$, so we get a bias error due to the inappropriateness of the model function $\boldsymbol{z}_1^T(x)\,\boldsymbol{\theta}_0$. The covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{n}A_\xi^{-1},$$

and so for a given design ξ and contamination term $\boldsymbol{u}^T(x)\boldsymbol{\beta}$, the mean square error of $\hat{\boldsymbol{\theta}}$ is the matrix

$$\mathrm{MSE}(\boldsymbol{\beta},\xi) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T] = \frac{\sigma^2}{n}A_\xi^{-1} + A_\xi^{-1}C_\xi\boldsymbol{\beta}\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1}. \qquad (2.1.11)$$

We will consider each of the three loss functions

$$\mathcal{L}_D(\boldsymbol{\beta},\xi) = \text{determinant}\,[\mathrm{MSE}(\boldsymbol{\beta},\xi)], \qquad (2.1.12)$$

$$\mathcal{L}_A(\boldsymbol{\beta},\xi) = \text{trace}\,[\mathrm{MSE}(\boldsymbol{\beta},\xi)], \qquad (2.1.13)$$

$$\mathcal{L}_Q(\boldsymbol{\beta},\xi) = \mathrm{IMSE}(\hat{y}(x)) = \int E\left(\hat{Y}(x) - E(Y \mid x)\right)^2 dx \qquad (2.1.14)$$

and our problem becomes that of obtaining optimal designs in the sense of minimizing $\sup_{\mathcal{F}} \mathcal{L}_D$, $\sup_{\mathcal{F}} \mathcal{L}_A$, and $\sup_{\mathcal{F}} \mathcal{L}_Q$. We close this section by proving

Lemma 2.1.1 *The matrix A_μ is singular iff the measure μ has p or fewer support points.*

Proof: If A_μ is singular, let $\mathbf{c}$ be a non-zero vector so that $\mathbf{c}^T A_\mu \mathbf{c} = 0$. Then

$$\mathbf{c}^T A_\mu \mathbf{c} = \mathbf{c}^T \int \boldsymbol{b}\boldsymbol{b}^T \, \mu(dx) \, \mathbf{c} = \int (\mathbf{c}^T \boldsymbol{b})^2 \, \mu(dx) = 0,$$

where $\boldsymbol{b} = (1, \ldots, x^p)^T$. The measure μ must then be supported on the points where $(\mathbf{c}^T \boldsymbol{b})^2 = 0$, and since $\mathbf{c}^T \boldsymbol{b}$ is a polynomial of degree less than or equal to p, there are at most p such points.

On the other hand, if the points $(x_i)_{i=1}^r$ support μ, where $r \leq p$, define $o(x) = \prod_{i=1}^r (x - x_i) = c_0 + c_1 x + \cdots + c_p x^p$. Here we set $c_i = 0$ for $i = r+1, \ldots, p$. Let $\mathbf{c}$ be the vector $(c_0, c_1, \cdots, c_p)^T$, then

$$A_\mu \mathbf{c} = \left(\int o(x)\mu(dx), \int x o(x)\mu(dx), \ldots, \int x^p o(x)\mu(dx) \right)^T.$$

But $o = 0$ μ-a.e. and so $A_\mu \mathbf{c} = \mathbf{0}$, which shows that A_μ is singular. ∎

2.2 Suprema of loss functions

In this section we will find formulas to express the suprema of the loss functions over $\mathcal{F}$ in terms of the design measure ξ. Recall that for any probability measure

ξ on $[-1/2, 1/2]$ we define

$$A_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x)\,\xi(dx), \tag{2.2.1}$$

$$C_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{u}^T(x)\,\xi(dx). \tag{2.2.2}$$

For the special case when ξ is the Lebesgue measure, we use the notation A_0 and C_0. We also define a $(q-p)\times(q-p)$ matrix by $B_0 = \int \boldsymbol{u}(x)\boldsymbol{u}^T(x)\,dx$, $\boldsymbol{\alpha} = B_0^{1/2}\boldsymbol{\beta}/\eta$ and $\nu = \sigma^2/n\eta^2$.

We note that the matrix B_0 is invertible. Each component of the vector $\boldsymbol{u}(x)$ is a polynomial, so for any vector $\boldsymbol{c} \in \mathbb{R}^{q-p}$ such that $B_0\boldsymbol{c} = \boldsymbol{0}$, we have $0 = \boldsymbol{c}^T B_0 \boldsymbol{c} = \int \|\boldsymbol{u}^T\boldsymbol{c}\|^2\,dx$. This implies that the polynomial $\boldsymbol{u}^T\boldsymbol{c}$ vanishes almost everywhere on $[-1/2, 1/2]$, which can only happen if $\boldsymbol{c} = \boldsymbol{0}$, so B_0 is invertible.

2.2.1 D-optimality criterion

We begin by rewriting the loss function in terms of A_ξ and C_ξ.

$$\begin{aligned}\mathcal{L}_D(\boldsymbol{\beta}, \xi) &= \text{determinant}[\text{MSE}(\boldsymbol{\beta}, \xi)] \\ &= \left(\frac{\sigma^2}{n}\right)^{p+1} \frac{1}{|A_\xi|}\left(1 + \frac{n}{\sigma^2}\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1} C_\xi \boldsymbol{\beta}\right).\end{aligned}$$

The supremum over $\mathcal{F}$ can be written as

$$\begin{aligned}\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) &= \sup_{\boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \le \eta^2} \mathcal{L}_D(\boldsymbol{\beta}, \xi) \\ &= \sup_{\boldsymbol{\alpha}^T\boldsymbol{\alpha}=1} \left(\frac{\sigma^2}{n}\right)^{p+1} \frac{1}{|A_\xi|}\left(1 + \frac{1}{\nu}\boldsymbol{\alpha}^T B_0^{-1/2} C_\xi^T A_\xi^{-1} C_\xi B_0^{-1/2}\boldsymbol{\alpha}\right) \\ &= \left(\frac{\sigma^2}{n}\right)^{p+1} \frac{1}{|A_\xi|}\left(1 + \frac{1}{\nu}\lambda_{\max} G_\xi\right). \end{aligned} \tag{2.2.3}$$

where $G_\xi = B_0^{-1/2} C_\xi^T A_\xi^{-1} C_\xi B_0^{-1/2}$. Applying the fact that for $A_{n\times p}$ and $B_{p\times n}$ matrices, the non-zero eigenvalues of AB and BA are the same and have the same multiplicities, we write $\lambda_{\max} G_\xi = \lambda_{\max}(A_\xi^{-1/2} C_\xi B_0^{-1} C_\xi^T A_\xi^{-1/2})$, when $q \geq 2p+1$, so that it is clear there are $p+1$ non-zero eigenvalues in the matrix G_ξ.
We claim that the minimax design exists and is symmetric under the D-optimality criterion. This is not the case for the A- and Q-optimal designs as we will see in the next subsection.

To prove the claim in Theorem 2.2.4 below, we need a few definitions and lemmas. Let Ξ denote the space of probability measures on $[-1/2, 1/2]$, and equip Ξ with the topology of weak convergence. We also let $\Xi_0 = \{\xi \in \Xi : |A_\xi| > 0\}$ and $\Xi_0^s = \{\xi \in \Xi_0 : \xi \text{ is symmetric}\}$. For any probability measure ξ on $[-1/2, 1/2]$, we let ξ^- denote the image of ξ under the mapping $x \mapsto -x$. The symmetrized version of ξ is defined to be the measure $\tilde{\xi} = (\xi + \xi^-)/2$. Thus ξ is symmetric if and only if $\xi = \xi^- = \tilde{\xi}$.

Define $g : \Xi \longrightarrow I\!R \cup \{\infty\}$ by $g(\xi) = \sup_{\mathcal{F}} \mathcal{L}_D(\beta, \xi)$ if $|A_\xi| > 0$, and $g(\xi) = \infty$ otherwise.

Lemma 2.2.1 *The function g is continuous on Ξ.*

Proof: We start by noting that the matrix A_ξ is made up of moments of ξ, that is, $(A_\xi)_{ij} = \xi_{i+j-2}$ for $1 \leq i, j \leq p$. Since the map $x \mapsto x^{i+j-2}$ is bounded and continuous on $[-1/2, 1/2]$, Theorem 25.8 (see page 344 of Billingsley (1986)) tells us that $\xi \mapsto (A_\xi)_{ij}$ is continuous. Now the determinant of a matrix is a polynomial in its components, so $\xi \mapsto |A_\xi|$ is also continuous. Similar arguments show that $\xi \mapsto C_\xi$ is continuous.

Suppose now $\xi(n) \to \xi$ with $|A_\xi| > 0$. Since $|A_{\xi(n)}| \to |A_\xi|$, we have $|A_{\xi(n)}| > 0$ for large n so without loss of generality we will assume that $A_{\xi(n)}$ is invertible. The inverse map is continuous on the space of nonsingular matrices, so that $A_{\xi(n)}^{-1} \to A_\xi^{-1}$ and hence $G_{\xi(n)} \to G_\xi$, elementwise. This convergence along with the estimate

$$\sup_{\boldsymbol{\alpha}^T\boldsymbol{\alpha}\le 1} \boldsymbol{\alpha}^T(G_{\xi(n)} - G_\xi)\boldsymbol{\alpha} \le \sum_{i,j}(G_{\xi(n)} - G_\xi)_{ij}^2.$$

then convinces us that $g(\xi(n)) \to g(\xi)$.

Suppose, on the other hand, that $\xi(n) \to \xi$ with $|A_\xi| = 0$. Since $g(\xi(n))$ is bounded below by a constant times $|A_{\xi(n)}|^{-1}$, we have $g(\xi(n)) \to \infty = g(\xi)$ as $n \to \infty$, so g is continuous at ξ. ∎

We recall a lemma which was stated and proved in Wiens (1993).

Lemma 2.2.2 *If V and W are matrices each of whose elements is a linear function of a real variable t, and if W is positive definite, then $\phi(t) = \boldsymbol{a}^T V^T W^{-1} V \boldsymbol{a}$ is a convex function of t for each $\boldsymbol{a}$.*

We note that by Lemma 2.2.2, G_ξ is convex in ξ if $|A_\xi| \neq 0$, and it is easy to see that then $\sup_{\{\|\boldsymbol{a}\|=1\}} \boldsymbol{a}^T G_\xi \boldsymbol{a}$ is also a convex functional of ξ.

Lemma 2.2.3 *For any $\xi \in \Xi$, we have $g(\tilde{\xi}) \le g(\xi)$.*

Proof: Without loss of generality we may assume that $g(\xi) < \infty$, that is, A_ξ is invertible.

The measures ξ and ξ^- share the same even moments, while the odd moments of ξ^- carry a reversed sign, that is, $\xi_i = (-1)^i(\xi^-)_i$. Therefore we have $A_\xi = PA_{\xi^-}P$ where $P = \text{diag}((-1)^0, \ldots, (-1)^p)$. An immediate consequence is that $|A_\xi| = |A_{\xi^-}|$. The convexity of the map $\xi \mapsto \log(|A_\xi|^{-1})$ implies that $\log(|A_{\bar{\xi}}|^{-1}) \leq (1/2)\log(|A_\xi|^{-1}) + (1/2)\log(|A_{\xi^-}|^{-1}) = \log(|A_\xi|^{-1})$. It follows that $|A_{\bar{\xi}}|^{-1} \leq |A_\xi|^{-1}$.

We next show that $\lambda_{\max} G_\xi = \lambda_{\max} G_{\xi^-}$. Similar to the matrix P, we define the matrix $Q = \text{diag}((-1)^{p+1}, \ldots, (-1)^q)$. Note that $P = P^T = P^{-1}$ and $Q = Q^T = Q^{-1}$. First, we see that $\boldsymbol{z}_1(-x) = P\boldsymbol{z}_1(x)$ and $\boldsymbol{z}_2(-x) = Q\boldsymbol{z}_2(x)$. Therefore $\boldsymbol{z}_1(-x)\boldsymbol{z}_2^T(-x) = P\boldsymbol{z}_1(x)\boldsymbol{z}_2^T(x)Q$, and so $C_{\xi^-} = \int \boldsymbol{z}_1(x)\boldsymbol{z}_2^T(x)\,\xi^-(dx) = \int P\boldsymbol{z}_1(x)\boldsymbol{z}_2^T(x)Q\,\xi(dx) = PC_\xi Q$. Applying the equations $A_{\xi^-} = PA_\xi P$ and $C_{\xi^-} = PC_\xi Q$ when ξ is Lebesgue measure, we can show that $\boldsymbol{u}(-x) = Q\boldsymbol{u}(x)$ as follows:

$$\begin{aligned}
\boldsymbol{u}(-x) &= \boldsymbol{z}_2(-x) - C_0^T A_0^{-1}\boldsymbol{z}_1(-x) \\
&= Q\boldsymbol{z}_2(x) - QC_0^T PPA_0^{-1}PP\boldsymbol{z}_1(x) \\
&= Q\boldsymbol{z}_2(x) - QC_0^T A_0^{-1}\boldsymbol{z}_1(x) \\
&= Q\boldsymbol{u}(x).
\end{aligned}$$

This now shows that B_0 commutes with Q since $QB_0Q = \int Q\boldsymbol{u}(x)\boldsymbol{u}(x)^T Q\,dx = \int \boldsymbol{u}(-x)\boldsymbol{u}^T(-x)\,dx = B_0$. Consequently we have

$$\begin{aligned}
G_{\xi^-} &= B_0^{-1/2}C_{\xi^-}^T A_{\xi^-}^{-1}C_{\xi^-}B_0^{-1/2} \\
&= B_0^{-1/2}QC_\xi^T PPA_\xi^{-1}PPC_\xi QB_0^{-1/2}
\end{aligned}$$

$$
\begin{aligned}
&= QB_0^{-1/2}C_\xi^T A_\xi^{-1} C_\xi B_0^{-1/2} Q \\
&= QG_\xi Q.
\end{aligned}
$$

This proves $\lambda_{\max} G_\xi = \lambda_{\max} G_{\xi^-}$, and therefore

$$
\begin{aligned}
g(\tilde{\xi}) &\le \left(\frac{\sigma^2}{n}\right)^{p+1} \frac{1}{|A_\xi|}\{1 + \frac{1}{2\nu}[\lambda_{\max} G_\xi + \lambda_{\max} G_{\xi^-}]\} \\
&= \left(\frac{\sigma^2}{n}\right)^{p+1} \frac{1}{|A_\xi|}\{1 + \frac{1}{\nu}\lambda_{\max} G_\xi\} \\
&= g(\xi).
\end{aligned}
$$

■

Theorem 2.2.4 *There exists a D-optimal design $\tilde{\xi} \in \Xi_0^s$.*

Proof: Let $M = \inf_\Xi g(\xi) < \infty$ and let $\{\xi(n)\}$ be any sequence in Ξ with $g(\xi(n)) \to M$. Since Ξ is compact, there exist a subsequence $\{\xi(n_j)\}$ and ξ such that $\xi(n_j) \to \xi$. Because $\xi(n_j) \to \xi$ implies $\tilde{\xi}(n_j) \to \tilde{\xi}$, and since g is weakly continuous, $g(\tilde{\xi}(n_j)) \to g(\tilde{\xi})$. On the other hand, $M \le g(\tilde{\xi}(n_j)) \le g(\xi(n_j))$ and $g(\xi(n_j)) \to M$ so $g(\tilde{\xi}(n_j)) \to M$. Hence $g(\tilde{\xi}) = M$, so $\tilde{\xi}$ is optimal and $\tilde{\xi} \in \Xi_0^s$

.

■

2.2.2 A- and Q-optimality criteria

Rewriting the loss function $\mathcal{L}_A$ in terms of A_ξ and C_ξ gives

$$
\mathcal{L}_A(\boldsymbol{\beta}, \xi) = \text{trace}[\text{MSE}(\boldsymbol{\beta}, \xi)]
$$

$$
\begin{aligned}
&= \left(\frac{\sigma^2}{n}\right) \text{trace}(A_\xi^{-1}) + \text{trace}(A_\xi^{-1} C_\xi \boldsymbol{\beta}\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1}) \\
&= \left(\frac{\sigma^2}{n}\right) \text{trace}(A_\xi^{-1}) + \boldsymbol{\beta}^T C_\xi^T A_\xi^{-2} C_\xi \boldsymbol{\beta}.
\end{aligned}
$$

So, we get

$$
\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi) &= \sup_{\boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \leq \eta^2} \mathcal{L}_A(\boldsymbol{\beta}, \xi) \\
&= \sup_{\boldsymbol{\alpha}^T\boldsymbol{\alpha}=1} \left(\frac{\sigma^2}{n}\right) [\text{trace}(A_\xi^{-1}) + \frac{1}{\nu} \boldsymbol{\alpha}^T B_0^{-1/2} C_\xi^T A_\xi^{-2} C_\xi B_0^{-1/2} \boldsymbol{\alpha}] \\
&= \frac{\sigma^2}{n} [\text{trace}(A_\xi^{-1}) + \frac{1}{\nu} \lambda_{\max} H_\xi], \qquad (2.2.4)
\end{aligned}
$$

where $H_\xi = B_0^{-1/2} C_\xi^T A_\xi^{-2} C_\xi B_0^{-1/2}$.

Secondly we consider the loss function $\mathcal{L}_Q$.

$$
\begin{aligned}
\mathcal{L}_Q(\boldsymbol{\beta}, \xi) &= \int \text{MSE } (\hat{y}(x))\, dx \\
&= \int E\{\hat{y}(x) - E(Y|x)\}^2\, dx \\
&= \int E\{\boldsymbol{z}_1^T(x)\hat{\boldsymbol{\theta}} - \boldsymbol{z}_1^T(x)\boldsymbol{\theta}_0 - \boldsymbol{u}^T(x)\boldsymbol{\beta}\}^2\, dx \\
&= \int \boldsymbol{z}_1^T(x)\{E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\}\boldsymbol{z}_1(x)\, dx + \boldsymbol{\beta}^T\{\int \boldsymbol{u}(x)\boldsymbol{u}^T(x)\, dx\}\boldsymbol{\beta} \\
&= \int \text{trace}\{\boldsymbol{z}_1^T(x)\text{MSE}(\boldsymbol{\beta}, \xi)\boldsymbol{z}_1(x)\}\, dx + \boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \\
&= \text{trace}\{\text{MSE}(\boldsymbol{\beta}, \xi)A_0\} + \boldsymbol{\beta}^T B_0 \boldsymbol{\beta}.
\end{aligned}
$$

Substituting the expression for the MSE and taking the supremum over $\mathcal{F}$ gives

$$
\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_Q(\boldsymbol{\beta}, \xi) &= \sup_{\boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \leq \eta^2} \mathcal{L}_Q(\boldsymbol{\beta}, \xi) \\
&= \frac{\sigma^2}{n} \text{ trace } (A_\xi^{-1} A_0) + \eta^2 \lambda_{\max} B_0^{-1/2} [C_\xi^T A_\xi^{-1} A_0 A_\xi^{-1} C_\xi + B_0] B_0^{-1/2} \\
&= \left(\frac{\sigma^2}{n}\right) \text{trace}(A_\xi^{-1} A_0) + \eta^2 [1 + \lambda_{\max} K_\xi] \\
&= \left(\frac{\sigma^2}{n}\right) [\text{trace}(A_\xi^{-1} A_0) + \frac{1}{\nu}(1 + \lambda_{\max} K_\xi)]. \qquad (2.2.5)
\end{aligned}
$$

where $K_\xi = B_0^{-1/2} C_\xi^T A_\xi^{-1} A_0 A_\xi^{-1} C_\xi B_0^{-1/2}$.

Remark For the D-optimality case, we have shown that the optimal design exists and is symmetric. For the A- and Q-optimality cases, the loss is not convex and the question of whether or not the optimal design is symmetric remains open. Nevertheless we can prove the existence of an optimal design in the class of symmetric designs. The proof is essentially same as in Lemma 2.2.1 and Theorem 2.2.4 except that the space Ξ is replaced by Ξ^s.

Define a function $h : \Xi^s \rightarrow I\!R \cup \{\infty\}$ by $h(\xi) = \sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi)$ if $|A_\xi| > 0$, and $h(\xi) = \infty$ otherwise. As we proved in Lemma 2.2.1, the function h is weakly continuous. Only one thing we need to pay attention to is the fact that when $\xi(n) \rightarrow \xi$ with $|A_\xi| = 0$, $h(\xi(n)) \longrightarrow \infty$ as $n \rightarrow \infty$, since $h(\xi(n)) \geq \frac{\sigma^2}{n} \text{trace}(A_{\xi(n)}^{-1})$. We now prove the existence of an optimal design $\xi \in \Xi_0^s$. Let $M = \inf_{\Xi^s} h(\xi) < \infty$ and let $\{\xi(n)\}$ be any sequence in Ξ^s with $h(\xi(n)) \rightarrow M$. Since Ξ^s is compact, there exist a subsequence $\{\xi(n_j)\}$ and ξ such that $\xi(n_j) \rightarrow \xi$. Since h is weakly continuous, $h(\xi(n_j)) \rightarrow h(\xi)$ and so $h(\xi) = M$, and $\xi \in \Xi_0^s$.

2.3 Number of support points

Having found a tractable formula for the maximum loss arising from various loss functions, we turn our attention to finding the optimal design, that is, the probability measure ξ that minimizes $\sup_{\mathcal{F}} \mathcal{L}(\boldsymbol{\beta}, \xi)$. A closer inspection of the loss function reveals that, since we are doing polynomial regression, $\sup_{\mathcal{F}} \mathcal{L}(\boldsymbol{\beta}, \xi)$ depends on ξ only through its first $p + q$ moments. This shows us that there is no unique optimal measure. but rather an optimal set of moments. which in general may have several different corresponding measures. Therefore. in order to simplify the search for the optimal set of moments. we begin by looking only at discrete measures with the minimal number of support points. This section is dedicated to proving a modification of a result by Wald (1939) which says that for any probability measure on $[0, 1]$, there is a discrete probability measure with $[(s+2)/2]$ or fewer support points. with the same first s moments. where the square brackets $[\quad]$ mean the integer part of a number. By this reduction. we hope to simplify the problem and get concrete information on the optimal solution.

Moment space M_s: Let M_s be the sth moment space. that is. M_s is the image of the space of probability measures μ on $[0, 1]$ under the mapping $\mu \mapsto (\int x\, d\mu, \ldots, \int x^s\, d\mu)$. Then M_s is a compact. convex subset of $\mathbb{R}^s$ and its extreme points are the image of $\{\delta_t : t \in [0, 1]\}$. that is. the extreme points of M_s are $\{(t, t^2, \ldots, t^s) : 0 \leq t \leq 1\}$.

A natural question arises: What is the minimum number of support points needed to attain any possible set of moments? In other words. find the minimum value of k so that for every $(\mu_1, \ldots, \mu_s) \in M_s$. there exists a probability measure of the form $\mu = \sum_{i=1}^{k} \alpha_i \delta_{t_i}$ where $\alpha_i \geq 0$. $\sum_{i=1}^{k} \alpha_i = 1$. and $(t_i)_{1 \leq i \leq k}$. such that $\mu_i = \int x^i\, d\mu(x)$ for $1 \leq i \leq s$.

We study moment space M_s in detail and answer the above question in this section.

Definition 2.3.1 *The degree of a distribution function F with jump points $t_1, \ldots t_m$ in $[0, 1]$ is defined as the number of interior jump points. i.e.. those in* *the open interval* $(0, 1)$, *plus one-half the number of jump points at the endpoints, i.e., those at 0 or 1. In other words,*

$$degree(F) = |\{t_1, \ldots, t_m\} \cap (0, 1)| + (1/2)|\{t_1, \ldots, t_m\} \cap \{0, 1\}|,$$

where the bars $|\ \ |$ *denote cardinality.*

Theorem 2.3.2 *Wald (1939)*

Let F *be a discrete distribution function on* $[0, 1]$, *and* G *an arbitrary distribution function on* $[0, 1]$. *If* F *has degree* d, *then the number of changes in sign of* $F - G$ *is less than or equal to* $2d - 1$.

Proof: Let F have jumps at the interior points $0 < t_1 < t_2 < \cdots < t_{d'} < 1$. Since F is constant on each of the subintervals $(t_1, t_2), \ldots, (t_{d'-1}, t_{d'})$, the function $F - G$ is monotone and so it can have at most one change of sign on each of these $d' - 1$ intervals. Besides that, $F - G$ may have a change of sign at any of the endpoints $t_1, t_2, \ldots, t_{d'}$, giving $2d' - 1$ potential changes of sign.

It remains to check $F - G$ on the intervals $(0, t_1)$ and $(t_{d'}, 1)$. We consider three different cases.

1. $(d = d')$. In this case, F has no jumps at either 0 or 1. Since $-G \leq 0$ and $1 - G \geq 1$, there is no change in sign of $F - G$ in either $(0, t_1)$ or $(t_{d'}, 1)$. So the total number of changes is less than or equal to $2d' - 1 = 2d - 1$.

2. $(d = d' + (1/2))$. If F has a jump at 0 (resp. at 1), then there may be a change of sign in the interval $(0, t_1)$ (resp. $(t_{d'}, 1)$). So the total number of changes is less than or equal to $2d' = 2d - 1$.

3. ($d = d' + 1$). Since F has a jump both at 0 and 1, the sign of $F - G$ may change on either of $(0, t_1)$ or $(t_{d'}, 1)$. The total number of changes is less than or equal to $2d' + 1 = 2d - 1$.

■

Theorem 2.3.3 *Wald (1939)*

Let F, G be discrete distribution functions on $[0, 1]$ with degrees less than or equal to d, and both having jumps at 1. Then the number of changes in sign of $F - G$ is less than or equal to $2d - 2$.

Proof: Suppose that one of the functions has degree strictly less than d, and without loss of generality suppose that it is F. Then degree$(F) = m \leq d - (1/2)$ and so by Theorem 2.3.2, we find that the number of changes in sign of $F - G$ is less than or equal to $2m - 1 \leq 2(d - 1/2) - 1 = 2d - 2$.

So to prove the theorem, we may assume that degree(F) =degree$(G) = d$. Without loss of generality, we may assume that $t_{d'}$, the largest interior jump point of F, is greater than or equal to the last interior jump point of G. Therefore $F - G$ is constant on $(t_{d'}, 1)$, so no change in sign can occur over this interval.

If q is an integer, then both F and G have jumps at 0 and $d' = d - 1$ interior jump points. As in case 2 of the proof of Theorem 2.3.2, the number of changes in sign of $F - G$ is less than or equal to $2d' = 2d - 2$.

On the other hand, if q is not an integer, then neither F nor G have jumps at 0, and they both have $d' = d - (1/2)$ interior jump points. Since there are no jumps at 0, the function $F-G$ is equal to $-G$ on $(0, t_1)$, and so no change in sign is possible on either $(0, t_1)$ or $(t_{d'}, 1)$. As in case 1 of the proof of Theorem 2.3.2, the number of changes in sign of $F-G$ is less than or equal to $2d'-1 = 2d-2$. ■

Theorem 2.3.4 *Wald (1939)*

Let F, G be distribution functions on $[0, 1]$ with the same first s moments, then either $F = G$, or $F - G$ has at least s changes in sign.

Proof: For all $a \in \mathbb{R}^s$ we have

$$\int_0^1 (a_1 t + a_2 t^2 + \cdots + a_s t^s)(F - G)(dt) = 0,$$

and so using integration by parts,

$$\int_0^1 (a_1 + 2a_2 t + \cdots + s a_s t^{s-1})(F - G)(t)\, dt = 0.$$

Now suppose that $F - G$ only changes sign at $0 < t_1 < t_2 < \cdots < t_k$ where $k < s$. Put $a_{k+2} = \cdots = a_s = 0$, and consider the k equations

$$\begin{aligned}
a_1 + 2a_2 t_1 + \quad \cdots + s a_s {t_1}^{s-1} &= 0 \\
a_1 + 2a_2 t_2 + \quad \cdots + s a_s {t_2}^{s-1} &= 0 \\
&\cdots \\
a_1 + 2a_2 t_k + \quad \cdots + s a_s {t_k}^{s-1} &= 0.
\end{aligned}$$

Let $a_1, a_2, \ldots, a_{k+1}$ be a non-trivial solution to this system of equations and define a polynomial Q by $Q(t) = a_1 + 2a_2 t + \cdots + s a_s t^{s-1} = a_1 + 2a_2 t + \cdots + (k+1) a_{k+1} t^k$.

The polynomial Q can have at most k roots, and since $t_1 < t_2 < \cdots < t_k$ are k distinct roots of Q, these all must be simple roots. Thus the function Q changes sign at each t_i and so the product $Q(F-G)$ does not change sign. However, $\int Q(F-G) = 0$ and so $F = G$. ■

Theorem 2.3.5 *If x is a boundary point of M_s, then there exists a discrete probability measure μ whose first s moments are given by x and whose distribution function F has degree less than or equal to $s/2$.*

Proof: Let P be a supporting hyperplane of M_s at x. That is $P = \{z \in \mathbb{R}^s : (a, z)_{\mathbb{R}^s} = c\}$ for some $c \in \mathbb{R}$ and $a \in \mathbb{R}^s$, so that $(a, z)_{\mathbb{R}^s} \geq c$ for all $z \in M_s$ and equality is achieved at x. Now x can be written as a convex combination of extreme points of $M_s \cap P$, that is, of extreme points of M_s that also lie in P. In other words, there exists $m \in \mathbb{N}$, $t_1, \ldots, t_m \in [0, 1]$, and $\alpha_1, \ldots, \alpha_m \in [0, 1]$, so that $\sum \alpha_i = 1$, $x = \sum_{i=1}^m \alpha_i (t_i, t_i^2, \ldots, t_i^s)$, and $\sum_{j=1}^s a_j t_i^j - c = 0$ for $i = 1, 2, \ldots, m$.

If $\mu = \sum_{i=1}^m \alpha_i \delta_{t_i}$ then the first s moments of μ are given by x. We claim that the corresponding distribution function F has degree less than or equal to $s/2$.

The polynomial $P(t) = \sum_{j=1}^s a_j t^j - c$ has roots at each of $t_1, \ldots, t_m$. The inequality $(a, z)_{\mathbb{R}^s} \geq c$ applied to the extreme points in M_s shows that P is non-negative on $[0, 1]$. Thus, each root of P in $(0, 1)$ must be a double root. Since P is an sth degree polynomial, s must exceed 2 times the number of roots in $(0, 1)$ plus any roots at the endpoints 0 or 1. Dividing this inequality by 2 gives $|\{t_1, \ldots, t_m\} \cap (0, 1)| + (1/2)|\{t_1, \ldots, t_m\} \cap \{0, 1\}| \leq s/2$, which is the desired result. ■

Corollary 2.3.6 *The distribution function F of Theorem 2.3.5 is unique.*

Proof: Let G have same first s moments. By Theorem 2.3.2, the number of changes in sign of $F - G$ is less than or equal to $s - 1$. By Theorem 2.3.4, either $F = G$ or the number of changes in sign of $F - G$ is at least s. The distributions F and G must then coincide. ■

We are now able to prove the main result of this section, which gives the minimum number of support points of a probability measure on $[0, 1]$ with the first s moments specified.

Theorem 2.3.7 *If $x \in M_s$, then there exists a probability measure μ on $[0, 1]$, with $(s + 2)/2$ support points or fewer, whose first s moments are given by x.*

Proof: First suppose that x is a boundary point of M_s. Then by Theorem 2.3.5 there is a distribution function F whose first s moments are given by x and whose degree is less than or equal to $s/2$. Therefore the number of jump points of F is less than or equal to $(s + 2)/2$.

If x is an interior point of M_s, then the set $M_{s+1} \cap \{(x, t) : t \in [0, 1]\}$ is not a singleton. In particular, if we let $t_1 = \inf\{t : (x, t) \in M_{s+1}\}$ and $t_2 = \sup\{t : (x, t) \in M_{s+1}\}$, then $t_1 < t_2$. Since (x, t_1) and (x, t_2) are on the boundary of M_{s+1}, we may apply Theorem 2.3.5 and obtain distribution functions F_1 and F_2, so that the corresponding measures both have the first n

moments given by x, and so that the degrees of both F_1 and F_2 are less than or equal to $(s+1)/2$. Since $F_1 \neq F_2$, Proposition 3 tells us that $F_1 - F_2$ has at least s changes in sign.

Now, if both F_1 and F_2 have jumps at 1, then Theorem 2.3.3 says that the number of changes in sign would be less than or equal to $2((s+1)/2) - 2 = s - 1$. As this contradicts the fact that $F_1 - F_2$ has at least s changes in sign, we conclude that one of F_1 or F_2 has no jump at 1. For such a distribution function, the number of jump points is at most the degree plus one-half, that is, it is less than or equal to $(s+2)/2$. ∎

Comment 1. Theorem 2.3.7 says that every x in M_s gives the first s moments for some discrete distribution function with $[(s+2)/2]$ or fewer jump points. This is, in fact, the best result possible for x in the interior of M_s. That is because, for such x, the proof of Theorem 2.3.7 gives two distinct discrete distribution functions F_1 and F_2 whose first s moments are given by x. If F is another such function, then it must differ from one of F_1, F_2; without loss of generality suppose that $F \neq F_1$. Then Theorem 2.3.4 says that $F - F_1$ has at least s changes in sign, and so by Theorem 2.3.2 we find that $s \leq 2$ degree$(F) - 1$. In other words, $(s+1)/2 \leq$ degree(F) which implies that the number of jump points of F is greater than or equal to $(s+1)/2$. But the number of jump points is an integer so it also is greater than or equal to $[(s+2)/2]$.

Comment 2. Suppose $x \in \text{int}(M_s)$. Let $F_1 \neq F_2$ and the first s moments of $F_i = x$, $i = 1, 2$. The Theorem 2.3.7 says that $\deg(F_i) \leq (s+1)/2$, $i = 1, 2$. On the other hand, the comment above implies that $\deg(F_1) = \deg(F_2) = (s+1)/2$ and F_2 has jump at 1, but F_1 does not. We want to prove a result corresponding to Theorem 2.3.7 that deals with symmetric measures on $[-1, 1]$.

Theorem 2.3.8 *If $x \in M_s$, then there exists a symmetric probability measure $\tilde{\mu}$ on $[-1, 1]$, with $s+1$ support points or fewer, whose first s even moments are given by x.*

Proof: If ν is a probability measure on $[0, 1]$ whose first s moments are given by x, and if μ is the image measure of ν under the mapping $x \to x^2$, then the first s even moments of μ are given by x. Also, the degree of the corresponding distribution functions are the same.

If μ^- is the image of μ under the map $x \to -x$, then $\tilde{\mu} = (\mu + \mu^-)/2$ gives a symmetric probability measure on $[-1, 1]$. Also, the even moments of μ and $\tilde{\mu}$ coincide and the number of support points of $\tilde{\mu}$ is $2 \times \text{degree}(F)$, plus one if μ has a jump at 1. Here F is the distribution function corresponding to μ.

Now, if x is a boundary point of M_s, there is a probability measure ν on $[0, 1]$ whose first s moments are given by x, and whose distribution function has degree less than or equal to $s/2$. If x is an interior point of M_s, the proof of Theorem 2.3.7 gave ν whose distribution function's first s moments are given by x, which does not have a jump at 1, and which has degree less than or equal to $(s+1)/2$. In both cases, the number of jump points of $\tilde{\mu}$ is less than or equal to $s+1$. ∎

Comment 3. If x is an interior point of M_s, then Theorem 2.3.7, combined with the first comment, shows that the degree of F is equal to $(s+1)/2$, and so the number of support points of $\tilde{\mu}$ is equal to $s+1$. Therefore if s is even, the symmetric measure $\tilde{\mu}$ must have a support point at zero.

Our objective in this chapter is to obtain optimal designs minimizing the supremum of each loss function under the three criteria. In the second section, we obtained the suprema of three loss functions over $\mathcal{F}$ in terms of the design measure ξ. In polynomial regression the supremum of the loss function depends on ξ only through its first $p+q$ moments. Searching for an optimal set of moments is quite complex, so we looked only at discrete measures with the minimal number of support points. In the rest of this chapter, we apply the results from this section to polynomial regression. We will be searching for a design measure that minimizes the supremum of each loss function. It is of special interest with a interior point of M_s. We summarize above results in this setup.

Theorem 2.3.9 *Suppose we fit a polynomial of degree p although the true model is a polynomial of degree $q, q > p$. Each loss function depends on the design ξ through its even moments, $\xi_2, \xi_4, \ldots, \xi_{2s}$, where $s = [(p+q)/2]$. If the vector $(\xi_2, \xi_4, \ldots, \xi_{2s})$ belongs to the interior of the moment space M_s, then an optimal design is of the form $\xi = \sum_{k=1}^{s+1} \alpha_k \delta_{x_k}$, $x_k \in [-1/2, 1/2]$ with $(\alpha_k, x_k)_{k=1}^{s+1}$ determined by $\xi_2, \xi_4, \ldots, \xi_{2s}$.*

In the following three sections, we apply the results from the previous sections to solve our minimization problem under three optimality criteria. Following the ground work laid by Huber (1975), Rychlik (1987) obtained minimax designs for polynomial regression. We start the fourth section by reviewing Rychlik's work.

2.4 Q-optimal designs for approximately linear models

2.4.1 Huber's minimax design for simple linear regression (SLR)

As we mentioned in Section 1.3.1. Huber obtains the minimax design by fitting the linear model although the true response is only approximately linear. We recall the fitted model. the true model and the contaminating space.

$$\textbf{Fitted Model:}\quad E(Y \mid x) = \theta_0 + \theta_1 x. \tag{2.4.1}$$

$$\textbf{True Model:}\quad E(Y \mid x) = \theta_0 + \theta_1 x + f(x). \tag{2.4.2}$$

where the contamination function f lies in the contaminating space

$$\mathcal{F}_2 = \{f : \int f^2(x)\,dx \le \eta^2.\ \int f(x)\,dx = \int xf(x)\,dx = 0\}. \tag{2.4.3}$$

Under the Q-optimality criterion with the contaminating space $\mathcal{F}_2$. Huber's minimax design depends on the parameter $\nu = \sigma^2/n\eta^2$. If $0 \le \nu \le 162/25$. then for $t = t(\nu) \in (1, 9/5]$ such that

$$\frac{5}{2}t^2(t-1) = \nu. \tag{2.4.4}$$

the density is

$$m(t; x) = 1 + \frac{5}{4}(t-1)(12x^2 - 1). \tag{2.4.5}$$

If $\nu \ge 162/25$. then for $c = c(\nu) \in (0, 1)$ such that

$$\frac{18(3 + 6c + 4c^2 + 2c^3)^2}{25(1 + 2c)^3(1 - c)^2} = \nu. \tag{2.4.6}$$

the density is

$$m(c;x) = \frac{3}{(1+2c)(1-c)^2}(4x^2 - c^2)^+. \tag{2.4.7}$$

2.4.2 Rychlik's optimal design for SLR

The contaminating space $\mathcal{F}_2$ used above is so wide that the maximum loss is infinite. So Rychlik restricts the contaminating space to a finite dimensional set of disturbances, that is, the intersection of $\mathcal{F}_2$ and a finite dimensional subspace of $L^2(-1/2, 1/2)$. He takes

$$\mathcal{F} = \{f(x) = \sum_{i=1}^{q} \alpha_i x^i : \int f(x)\,dx = \int x f(x)\,dx = 0, \int f^2(x)\,dx \le \eta^2\}. \tag{2.4.8}$$

Applying the fact that the Legendre polynomials orthogonal to 1 and x form an orthogonal basis of the space spanned by $\mathcal{F}$, Rychlik proves that for $0 \le \nu \le 162/25$, any symmetric design is minimax if its even moments are identical to the corresponding even moments with respect to Huber's continuous design $m(t;x)$, where $m(t;x)$ and t are defined in (2.4.4) and (2.4.5). For $\nu \ge 162/25$, he proves that the minimax design has extreme support points, that is, $\xi = (\delta_{-1/2}+\delta_{1/2})/2$. As an example, we provide Rychlik's optimal design ξ^* for fitting a line although the true response might be a cubic or quartic.

$$\xi^* = \begin{cases} \frac{1}{2}\delta_{-\frac{1}{2}} + \frac{1}{2}\delta_{+\frac{1}{2}} & \text{if } \nu \ge 2970 \\ \frac{1}{2}\delta_{-\frac{1}{2}\sqrt{a}} + \frac{1}{2}\delta_{+\frac{1}{2}\sqrt{a}} & \text{if } 15.15 \le \nu \le 2970. \\ \frac{1}{2}\delta_{-\frac{1}{2}\sqrt{c(t)}} + \frac{1}{2}\delta_{+\frac{1}{2}\sqrt{c(t)}} & \text{if } \nu < 15.15. \end{cases}$$

where $a \in [0.742, 1]$ is the root of the equation $(315a^2/16)(105a^3 - 135a^2 + 51a - 5) = \nu$. The design point $c(t)$ is defined as $(t/3) \pm (-t^2 + 18t/7 - 3/35)^{1/2}$. At this point we note that there are four support points when $\nu < 15.15$ in the example above. We apply the Rychlik's approach to D- and A-optimality cases. in fact we will improve his results in terms of the number of support points - our minimax design will have fewer support points.

2.5 D-optimal designs for approximately linear model

2.5.1 Wiens's minimax design for SLR

Wiens (1992) obtains minimax designs for SLR with the contaminating space $\mathcal{F}_2$. under various optimality criteria. He extends these ideas to multiple regression as well. Wiens proves that in order that $\sup_{\mathcal{F}_2} \mathcal{L}(f, \xi)$ be finite. it is necessary that ξ be absolutely continuous. From now on we will consider measures ξ with density $\xi'(x) = m(x)$. unless otherwise mentioned. Under the D-optimality criterion. Wiens's (1992) minimax density depends on ν.

If $0 \le \nu \le 14/5$.

$$\nu = \frac{5}{4}(12\xi_2 - 1)(12\xi_2 + 1). \tag{2.5.1}$$

the density is

$$m(x; \xi_2) = 1 + \frac{5}{4}(12\xi_2 - 1)(12x^2 - 1). \quad 0 \le |x| \le 1/2. \tag{2.5.2}$$

If $\nu \ge 14/5$. then for $\xi_2(\nu)$ such that

$$\nu = \frac{2\xi_2[(1-\sqrt{b})J(\xi_2; b) - \sqrt{b}]}{(1-\sqrt{b})\xi_2 - (1/12)(1-b^{3/2})} - J(\xi_2; b). \tag{2.5.3}$$

$$J(b;\xi_2) = \frac{3(4\xi_2 - b)}{3(1-b) - 2(1-b^{3/2})}. \tag{2.5.4, 2.5.5}$$

the density is

$$m(x;\xi_2) = (4x^2 - b)^+/K(b), \quad \sqrt{b}/2 \le |x| \le 1/2, \tag{2.5.6}$$

where $K(b) = (1-b) - (2/3)(1-b^{3/2})$, and b is determined by the equation

$$\xi_2 = \frac{15(1-b) - 6(1-b^{5/2})}{180(1-b) - 120(1-b^{3/2})}. \tag{2.5.7}$$

2.5.2 Minimax design based on Rychlik's approach

Using Wiens's minimax density we construct the optimal density based on Rychlik's idea. We recall that the fitted model (2.4.1) and the true model (2.4.2) are

$$\textbf{Fitted Model:}\quad E(Y \mid x) = \theta_0 + \theta_1 x. \tag{2.5.8}$$

$$\textbf{True Model:}\quad E(Y \mid x) = \theta_0 + \theta_1 x + f(x). \tag{2.5.9}$$

where f belongs to the contaminating space $\mathcal{F}$ in (2.4.8). Again applying the fact that the Legendre polynomials orthogonal to 1 and x form an orthogonal basis of the space spanned by $\mathcal{F}$, we express this contaminating space as

$$\mathcal{F} = \{\eta \sum_{i=2}^{q} \alpha_i l_i \;:\sum \alpha_i^2 \le 1\}. \tag{2.5.10}$$

where l_i is the ith normalized Legendre polynomial, that is, $l_i(x) = (2i+1)^{1/2} P_i(2x)$, where P_i is the ith Legendre polynomial on $[-1, 1]$. For instance,

$$l_0(x) = 1,\ l_1(x) = 2\sqrt{3}x,\ l_2(x) = \sqrt{5}(6x^2 - 1/2),\ l_3(x) = \sqrt{7}(20x^3 - 3x).$$

The true model can be expressed as

$$E(Y \mid x) = \theta_0 + \theta_1 x + \eta \sum_{i=2}^{q} \alpha_i l_i(x) \tag{2.5.11}$$

The mean squared error matrix of $\hat{\boldsymbol{\theta}}$ is then

$$\text{MSE} = \eta^2 (\nu A_\xi^{-1} + A_\xi^{-1} C_\xi \boldsymbol{\alpha}\boldsymbol{\alpha}^T C_\xi^T A_\xi^{-1}).$$

where A_ξ and C_ξ are as in (2.2.1) and (2.2.2) with $\boldsymbol{u}^T(x)$ is replaced by $\boldsymbol{l}^T(x) = (l_2(x), \ldots, l_q(x))$. By virtue of Theorem 2.2.4, we assume that ξ is symmetric and so the matrices above are

$$A_\xi = \begin{pmatrix} 1 & 0 \\ 0 & \xi_2 \end{pmatrix}, \quad C_\xi = \begin{pmatrix} \boldsymbol{c}_1^T(\xi) \\ \boldsymbol{c}_2^T(\xi) \end{pmatrix}.$$

where

$$\begin{aligned} \boldsymbol{c}_1^T(\xi) &= E_\xi \boldsymbol{l}^T = (E_\xi l_2, \ldots, E_\xi l_q), \\ \boldsymbol{c}_2^T(\xi) &= E_\xi x\boldsymbol{l}^T = (E_\xi x l_2, \ldots, E_\xi x l_q). \end{aligned}$$

We observe that $\boldsymbol{c}_1^T(\xi)\boldsymbol{c}_2(\xi) = 0$, since l_j is an odd function for odd j and xl_j is an odd function for even j. We find that

$$\det(\text{MSE}) = \left(\frac{\sigma^2}{n}\right)^2 \frac{1}{\xi_2}(1 + \frac{1}{\nu}\boldsymbol{\alpha}^T C_\xi^T A_\xi^{-1} C_\xi \boldsymbol{\alpha}).$$

Put

$$\begin{aligned} \mathcal{L}_D(\boldsymbol{\alpha}, \xi) &= \frac{1}{\xi_2}(1 + \frac{1}{\nu}\boldsymbol{\alpha}^T C_\xi^T A_\xi^{-1} C_\xi \boldsymbol{\alpha}) \\ &= \frac{1}{\xi_2}\{1 + \frac{1}{\nu}[\boldsymbol{\alpha}^T(\boldsymbol{c}_1(\xi)\boldsymbol{c}_1^T(\xi) + \frac{1}{\xi_2}\boldsymbol{c}_2(\xi)\boldsymbol{c}_2^T(\xi))\boldsymbol{\alpha}]\}. \end{aligned}$$

Note that, for fixed second moment ξ_2, the loss function depends only on C_ξ and not A_ξ. We next want to find ξ^* minimizing the supremum of $\mathcal{L}(\boldsymbol{\alpha}, \xi)$:

$$\begin{aligned}
\xi^* &= \arg\min_{\xi} \sup_{\|\boldsymbol{\alpha}\|\le 1} \mathcal{L}(\boldsymbol{\alpha}, \xi) \\
&= \arg\min_{\xi} \frac{1}{\xi_2}[1 + \frac{1}{\nu}\lambda_{\max}(\boldsymbol{c}_1(\xi)\boldsymbol{c}_1^T(\xi) + \frac{1}{\xi_2}\boldsymbol{c}_2(\xi)\boldsymbol{c}_2^T(\xi))].
\end{aligned}$$

The non-zero roots of $\boldsymbol{c}_1(\xi)\boldsymbol{c}_1^T(\xi) + \frac{1}{\xi_2}\boldsymbol{c}_2(\xi)\boldsymbol{c}_2^T(\xi)$ are

$$\begin{aligned}
\lambda_1(\xi) &= \|\boldsymbol{c}_1(\xi)\|^2, && \text{with eigenvector } \boldsymbol{\alpha}_1(\xi) = \boldsymbol{c}_1/\|\boldsymbol{c}_1\|, \\
\lambda_2(\xi) &= (1/\xi_2)\|\boldsymbol{c}_2(\xi)\|^2, && \text{with eigenvector } \boldsymbol{\alpha}_2(\xi) = \boldsymbol{c}_2/\|\boldsymbol{c}_2\|.
\end{aligned}$$

Our minimization problem is now

$$\xi^* = \arg\min_{\xi} \frac{1}{\xi_2}\left(1 + \frac{1}{\nu}\max\{\lambda_1(\xi), \lambda_2(\xi)\}\right).$$

We state and prove a result that is similar to Rychlik's under Q-optimality. In the theorem below we prove that for small ν any design measure ξ is minimax if its expectations of even Legendre polynomials are identical to the corresponding expectations for Wiens's continuous design (2.5.2).

Theorem 2.5.1 *When $0 \le \nu \le 14/5$, any symmetric design measure ξ is minimax if the expectations under ξ of the first $s := [(q+1)/2]$ even-order Legendre polynomials are the same as for Wiens's design (2.5.2). Moreover, there exists a minimax design with $s+1$ design points on $[-1/2, 1/2]$. The design points and point masses are calculated, for $j = 1, \ldots, s$, by*

$$\xi_{2j} = \frac{60\, j\xi_2 - 3(j-1)}{2^{2j}(2j+1)(2j+3)}, \quad \textit{where } \xi_2 = \sqrt{\frac{4\nu+5}{720}}. \tag{2.5.12}$$

Proof: We first assume $\lambda_1(\xi) = \max\{\lambda_1(\xi), \lambda_2(\xi)\}$, and prove the first part of this theorem in four steps.

1. Find an absolutely continuous measure $M(d)$, with density $m(d)$, that minimizes $\lambda_1(\xi) = \|\boldsymbol{c}_1(\xi)\|^2 = \sum_{i=2}^{q}(E_\xi[l_i])^2$ subject to $\xi_2 = d$.

2. Choose d^* to minimize $\frac{1}{d}[1 + \frac{1}{\nu}\lambda_1(M(d))]$, and put $\xi^* = M(d^*)$.

3. Verify that $\lambda_1(\xi^*) \geq \lambda_2(\xi^*)$.

4. Verify that $m(d^*)$ is indeed a density, i.e., is non-negative with a total mass of unity.

These four steps imply then that ξ^* is minimax. The reason is that for any measure ξ, the maximum loss is proportional to

$$\begin{aligned} o(\xi) &:= \frac{1}{\xi_2}(1 + \frac{1}{\nu}\max\{\lambda_1(\xi), \lambda_2(\xi)\}) \\ &\geq \frac{1}{\xi_2}\{1 + \frac{1}{\nu}\lambda_1(\xi)\} \\ &\geq \frac{1}{\xi_2^*}\{1 + \frac{1}{\nu}\lambda_1(\xi^*)\} \qquad (\text{ by 1 and 2 }) \\ &= o(\xi^*) \qquad (\text{ by 3 }). \end{aligned}$$

Thus any distribution ξ with $C_\xi = C_{\xi^*}$ is also minimax. We now solve our minimization problem step by step.

The first step is to minimize $\lambda_1(\xi) = \sum_{i=2}^{q}(E_\xi[l_i])^2$ subject to $\xi_2 = d$. It is natural to conjecture that a minimizer is given by $m(x; d) = 1 + c\,l_2(x)$, (so then $E_{M(d)}(l_i) = 0$ if $i > 2$) where $c = c(d)$ satisfies $d = \int x^2(1 + c\,l_2(x))\,dx$. This implies $c = (\sqrt{5}/2)(12d - 1)$. On the other hand, $c = E_{M(d)}[l_2] = \sqrt{\lambda_1(M(d))}$.

This shows that $M(d)$ is the minimizer because if ξ is any distribution function with $\xi_2 = d$, then

$$\lambda_1(\xi) \geq (E_\xi[l_2])^2 = (5/4)(12d-1)^2 = c^2(d) = \lambda_1(M(d)).$$

The second step is to find d minimizing the loss, $(1/d)(1+\frac{1}{\nu}\lambda_1(M(d)))$. It turns out that the minimizer d^* is $[(4\nu+5)/720]^{1/2}$, so that

$$m(x; d^*) = 1 + \frac{5}{4}(12d^* - 1)(12x^2 - 1).$$

The third step is to verify that $\lambda_1(\xi^*) \geq \lambda_2(\xi^*)$. We have calculated $\lambda_1(\xi^*)$ above: $\lambda_1(\xi^*) = (5/4)(12d^* - 1)$, with $d^* = [(4\nu+5)/720]^{1/2}$. We recall that the second eigenvalue is given by $\|\mathbf{c}_2(\xi^*)\|^2/\xi_2^* = (\sum_{j=2}^{q}(E_{\xi^*} xl_j)^2)/\xi_2^*$. We observe that if j is even, then $E_{\xi^*}(xl_j) = 0$, since xl_j is a odd function and ξ^* is symmetric. When j is odd, $2k+1$, say, xl_{2k+1} can be expressed as a linear combination of the even Legendre polynomials, that is,

$$xl_{2k+1} = a_k l_{2k} + b_k l_{2(k+1)}, \tag{2.5.13}$$

where the coefficients a_k and b_k are

$$a_k = (2k+1)[4(4k+1)(4k+3)]^{-1/2}, \; b_k = (k+1)[(4k+3)(4k+5)]^{-1/2}.$$

Except for the first term, all other terms after taking expectations in (2.5.13) will disappear:

$$\begin{aligned} E_{\xi^*}(xl_{2k+1}) &= a_k E_{\xi^*}(l_{2k}) + b_k E_{\xi^*}(l_{2(k+1)}) \\ &= \begin{cases} a_1 E_{\xi^*}(l_2) & \text{if } k = 1 \\ 0 & \text{if } k > 1. \end{cases} \end{aligned}$$

Then $\sum_{j=2}^{q}(E_{\xi^*}(xl_j))^2 = a_1^2[\int l_2(x)(1 + c^* l_2(x))\,dx]^2 = (9/140)c^{*2}$, where $c^* = (\sqrt{5}/2)(12d^* - 1)$. Substituting d^*, the second eigenvalue becomes

$$\|c_2(\xi^*)\|^2/d^* = (27/35)\{c^{*2}/[(2/\sqrt{5})c^* + 1]\}. \tag{2.5.14}$$

The right hand side of the equation (2.5.14) is less than c^{*2}, which is $\lambda_1(\xi^*)$. This proves that $\lambda_1(\xi^*)$ is indeed larger than $\lambda_2(\xi^*)$. The final step is to show that $m(d^*)$ is in fact a density. Since $m(x; d^*)$ takes its minimum at $x = 0$, it suffices to show that $m(0; d^*) \geq 0$. Simple calculations show that this holds provided $0 \leq \nu \leq 14/5$.

For the second part, we notice that when $j \geq 2$, the expectations of the even Legendre polynomials with respect to ξ^* are equal to zero. A straightforward calculation provides, for $j = 1, \ldots, s$,

$$\xi_{2j} = \frac{60j\xi_2 - 3(j-1)}{2^{2j}(2j+1)(2j+3)}, \quad \text{where } \xi_2 = \sqrt{\frac{4\nu + 5}{720}},$$

and any symmetric measure ξ sharing these moments will also be minimax. This completes the proof. ■

Since the moments in Theorem 2.5.1 above are generated by a measure with a density, they belong to the interior of the moment space. Thus by applying Theorem 2.3.9, we characterize the optimal solution for arbitrary q in the Remark below.

Remark If we fit a linear model where the true response is a polynomial of degree $q \geq 2$, then for $0 \leq \nu \leq 14/5$ there exists an optimal design ξ of the form $\xi = \sum_{i=1}^{s+1} \alpha_i \delta_{x_i}$, where $x_i \in [-1/2, 1/2]$, $\alpha_i \geq 0$, and $\sum \alpha_i = 1$. The design ξ contains unknown parameters: the point masses α_i and the design points x_i. When s is even, since zero is a support point, there are $s/2$ point masses and $s/2$ design points. When s is odd, there are $(s-1)/2$ point masses and $(s+1)/2$ design points. Whether s is even or odd, there are s unknown parameters to be determined. But equation (2.5.12) gives s equations and thus we can theoretically determine all unknown parameters.

For very large ν, we show that the optimal design has two design points which are separated as widely as possible.

Theorem 2.5.2 *For every $s \geq 1$ there exists $\nu_s < \infty$ such that $(\delta_{-1/2} + \delta_{1/2})/2$ is minimax for $\nu \geq \nu_s$.*

The proof of Theorem 2.5.2 is not complicated but is very long. We need to give a proof separately when q is even or odd. We prepare some preliminary work to simplify the proof. The first thing we need to do is to write the eigenvalues $\lambda_1(\xi)$ and $\lambda_2(\xi)$ in terms of the moments ξ_i. It turns out that it is much easier to explain the following proof if we write the eigenvalues in terms of $\omega_i := \int (2x)^{2i} \xi(dx) = 2^{2i}\xi_{2i}$ instead of the moments ξ_i. Since ξ is symmetric, writing the eigenvalues in terms of ω_i will capture only the even Legendre polynomials. We dealt with the Legendre polynomials before, but now we need to write up the coefficients explicitly in the notation. We denote the $(2j)$th Legendre polynomial on $[-1, 1]$, by $P_{2j}(x) = \sum_{i=0}^{j} p_{2j,2i}\, x^{2i}$. The expectation of this polynomial is denoted by

$Q_{2j}(\boldsymbol{\psi}) = E_\xi[P_{2j}(2x)] = \sum_{i=0}^{j} p_{2j,2i}\,\psi_i$. As an illustration, we show some of the P_{2j}'s as well as the $Q_{2j}(\boldsymbol{\psi})$'s.

$$\begin{aligned}
P_2(x) &= (1/2)(-1+3x^2) := p_{2,0} + p_{2,2}\,x^2 \\
P_4(x) &= (1/8)(3-30x^2+35x^4) := p_{4,0} + p_{4,2}\,x^2 + p_{4,4}\,x^4 \\
P_6(x) &= (1/16)(-5+105x^2-315x^4+231x^6). \\
&:= p_{6,0} + p_{6,2}\,x^2 + p_{6,4}\,x^4 + p_{6,6}\,x^6
\end{aligned}$$

$$\begin{aligned}
Q_2(\boldsymbol{\psi}) &= \int_{-1/2}^{1/2} (1/2)(-1+3(2x)^2)\,\xi(dx) = (1/2)(-1+3\psi_1) := p_{2,0} + p_{2,2}\,\psi_1 \\
Q_4(\boldsymbol{\psi}) &= (1/8)(3-30\,\psi_1+35\,\psi_2) := p_{4,0} + p_{4,2}\,\psi_1 + p_{4,4}\,\psi_2 \\
Q_6(\boldsymbol{\psi}) &= (1/16)(-5+105\,\psi_1-315\,\psi_2+231\,\psi_3) \\
&:= p_{6,0} + p_{6,2}\,\psi_1 + p_{6,4}\,\psi_2 + p_{6,6}\,\psi_3.
\end{aligned}$$

The eigenvalues can be rewritten in terms of ψ_i, by

$$\begin{aligned}
\lambda_1(\boldsymbol{\psi}) &= \sum_{j=1}^{s}[E_\xi l_{2j}]^2 = (E_\xi\sqrt{4j+1}P_{2j}(2x))^2 \\
&= \sum_{j=1}^{s}(4j+1)[E_\xi\,P_{2j}(2x)]^2 = \sum_{j=1}^{s}(4j+1)Q_{2j}^2(\boldsymbol{\psi}).
\end{aligned}$$

$$\begin{aligned}
\lambda_2(\boldsymbol{\psi}) &= \frac{4}{\psi_1}\sum_{j=1}^{s-1}[a_jE_\xi l_{2j} + b_jE_\xi l_{2(j+1)}]^2 \\
&= \frac{4}{\psi_1}\sum_{j=1}^{s-1}[\frac{2j+1}{2\sqrt{4j+3}}Q_{2j}(\boldsymbol{\psi}) + \frac{j+1}{\sqrt{4j+3}}Q_{2(j+1)}(\boldsymbol{\psi})]^2 \\
&= \sum_{j=1}^{s-1}\frac{1}{(4j+3)\psi_1}[(2j+1)Q_{2j}(\boldsymbol{\psi}) + 2(j+1)Q_{2(j+1)}(\boldsymbol{\psi})]^2.
\end{aligned}$$

Secondly, we define a polynomial R_{2j} in the variable ψ_1 by taking $Q_{2j}(\boldsymbol{\psi})$ and replacing the ψ_i's in the following manner: if the preceding coefficient is positive we replace ψ_i by ψ_1^i but if it is negative we replace ψ_i by ψ_1. Some of the R_{2j}'s are

$$\begin{aligned}
R_2(\psi_1) &= (1/2)(-1+3\psi_1) \\
R_4(\psi_1) &= (1/8)(3-30\psi_1+35\psi_1^2) \\
R_6(\psi_1) &= (1/16)(-5+105\psi_1-315\psi_1+231\psi_1^3) \\
R_8(\psi_1) &= (1/128)(35-1260\psi_1+6930\psi_1^2-12012\psi_1+6435\psi_1^4).
\end{aligned}$$

Third and last, we define

$$r(\psi_1) = \sum_{j=1}^{s}(4j+1)R_{2j}^2(\psi_1). \qquad (2.5.15)$$

and we notice that $r(1) = \lambda_1(\mathbf{1})$, where $\mathbf{1}$ is the $s+1$ dimensional vector of ones. Let $g(\boldsymbol{\psi}) = \frac{1}{\psi_1}\{1 + \frac{1}{\nu}\lambda_1(\boldsymbol{\psi})\}$. We are now ready to prove the theorem when q is even.

Proof: We want to prove that for sufficiently large ν, $g(\boldsymbol{\psi}) \geq g(\mathbf{1})$ for all $\boldsymbol{\psi}$. Here we have assumed that $\lambda_1(\mathbf{1}) \geq \lambda_2(\mathbf{1})$. This is true because

$$\begin{aligned}
\lambda_1(\mathbf{1}) &= \sum_{j=1}^{s}(4j+1)^2 Q_{2j}^2(\mathbf{1}) = \sum_{j=1}^{s}(4j+1) = s(2s+3), \\
\lambda_2(\mathbf{1}) &= \sum_{j=1}^{s-1}\frac{1}{4j+3}[(2j+1)Q_{2j}(\mathbf{1}) + 2(j+1)Q_{2(j+1)}(\mathbf{1})]^2 \\
&= \sum_{j=1}^{s-1}\frac{1}{4j+3}[(2j+1)+2(j+1)]^2 = \sum_{j=1}^{s-1}(4j+3) = (s-1)(2s+3).
\end{aligned}$$

The inequality $g(\boldsymbol{\psi}) \geq g(\mathbf{1})$ is equivalent to

$$\nu \geq \frac{\psi_1 \lambda(\mathbf{1}) - \lambda_1(\boldsymbol{\psi})}{1 - \psi_1}$$

because

$$\begin{aligned} g(\boldsymbol{\psi}) \geq g(\mathbf{1}) \quad &\text{iff} \quad \frac{1}{\psi_1}(1 + \frac{1}{\nu}\lambda_1(\boldsymbol{\psi})) \geq 1 + \frac{1}{\nu}\lambda_1(\mathbf{1}) \\ &\text{iff} \quad \frac{\nu + \lambda_1(\boldsymbol{\psi}) - \psi_1 \nu - \psi_1 \lambda(\mathbf{1})}{\psi_1 \nu} \geq 0. \end{aligned}$$

The following lemma finds the maximum of $(\psi_1 \lambda_1(\mathbf{1}) - \lambda_1(\boldsymbol{\psi}))/(1 - \psi_1)$ over $\boldsymbol{\psi}$. Assuming this lemma for the moment we finish up the proof for the theorem and we will prove the lemma later.

Lemma 2.5.3 *For each $s \geq 1$, there exists $\psi_1^* < 1$ such that*

$$\lambda_1(\boldsymbol{\psi}) \geq r(\psi_1) \qquad \textit{for all } \boldsymbol{\psi} \textit{ with } \psi_1 \in (\psi_1^*, 1).$$

This lemma implies

$$\frac{\psi_1 \lambda_1(\mathbf{1}) - \lambda_1(\boldsymbol{\psi})}{1 - \psi_1} \leq \frac{\psi_1 \lambda_1(\mathbf{1}) - r(\psi_1)}{1 - \psi_1}.$$

and so

$$\sup_{\{\boldsymbol{\psi} : \psi_1^* \leq \psi_1 \leq 1\}} \frac{\psi_1 \lambda_1(\mathbf{1}) - \lambda_1(\boldsymbol{\psi})}{1 - \psi_1} \leq \sup_{\psi_1^* \leq \psi_1 \leq 1} \frac{\psi_1 \lambda_1(\mathbf{1}) - r(\psi_1)}{1 - \psi_1} := \nu_{s1}$$

The numerator of right hand side in the above inequality, $\psi_1 \lambda_1(\mathbf{1}) - r(\psi_1)$, is a polynomial in ψ_1 which vanishes at $\psi_1 = 1$, hence has a factor $(1 - \psi_1)$. Thus $\nu_{s1}(\psi_1^*) < \infty$ for each ψ_1^*. On the other hand, for $0 \leq \psi_1 \leq \psi_1^*$ we have

$$\sup_{\{\boldsymbol{\psi} : 0 \leq \psi_1 \leq \psi_1^*\}} \frac{\psi_1 \lambda_1(\mathbf{1}) - \lambda_1(\boldsymbol{\psi})}{1 - \psi_1} \leq \sup_{0 \leq \psi_1 \leq \psi_1^*} \frac{\psi_1 \lambda_1(\mathbf{1})}{1 - \psi_1} := \nu_{s2}.$$

Taking $\nu_s = \max\{\nu_{s1}, \nu_{s2}\}$ gives $g(\psi) \geq g(1)$ for all ψ whenever $\nu \geq \nu_s$. Consequently $(\delta_{-1/2} + \delta_{1/2})/2$ is minimax for $\nu \geq \nu_s$. This proves the theorem for q even. ■

Proof: (of Lemma 2.5.3)

We must show that ψ with $\psi_1 \in (\psi_1^*, 1)$,

$$\lambda_1(\psi) = \sum_{j=1}^{s}(4j+1)Q_{2j}^2(\psi) \geq \sum_{j=1}^{s}(4j+1)R_{2j}^2(\psi_1) = r(\psi_1).$$

Since $\psi_1 \geq \psi_2 \geq \cdots \geq \psi_s$, and $\psi_i \geq \psi_1^t$, for all i,

$$Q_{2j}(\psi) \geq R_{2j}(\psi_1), \text{ for all } \psi. \tag{2.5.16}$$

But we must take care of the case when $Q_{2j}(\psi)$ and $R_{2j}(\psi_1)$ are negative. Since $R_{2j}(\psi_1)$ is continuous in ψ_1 and $R_{2j}(1) = 1$, for each $j = 1, 2, \ldots, s$, there exists ψ_{j1}^* such that $R_{2j}(\psi_1) \geq 0$ for all $\psi_1 \in (\psi_{j1}^*, 1)$. For example, set ψ_{j1}^* to be the largest root of $R_{2j}(\psi_1)$. Taking ψ_1^* as the maximum of $\{\psi_{j1}^* : j = 1, \ldots, s\}$ gives

$$R_{2j}(\psi_1) \geq 0 \text{ for all } \psi_1 \in (\psi_1^*, 1). \tag{2.5.17}$$

The inequalities (2.5.16) and (2.5.17) imply that

$$Q_{2j}(\psi) \geq R_{2j}(\psi_1) \geq 0 \text{ for all } \psi \text{ with } \quad \psi_1 \in (\psi_1^*, 1).$$

and hence this shows $\lambda_1(\psi) \geq r(\psi_1)$ for all ψ with $\psi_1 \in (\psi_1^*, 1)$. ■

It remains to find $\sup_{\psi_1^* \leq \psi_1 \leq 1}(\psi_1\lambda(1) - r(\psi_1))/(1-\psi_1) := \nu_{s1}$ and determine whether or not $\nu_{s2} < \nu_{s1}$. The maximum of $\frac{\psi_1\lambda(1)-r(\psi_1)}{1-\psi_1}$ occurs at $\psi_1 = 1$, and we can formulate ν_{s1} in terms of s and the coefficients $p_{2j,2i}$ of $Q_{2j}(\psi)$. We write out the formula for ν_{s1} in the following lemma and give a proof.

Lemma 2.5.4

$$\sup_{\psi_1^* \le \psi_1 \le 1} \frac{\psi_1 \lambda(1) - r(\psi_1)}{1-\psi_1} = \frac{1}{2}s(2s+3)(2s^2+3s-1) - 2\sum_{j=3}^{s}(4j+1)\left(\sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right)$$

Proof: The proof is a collection of four facts.

Fact 1 $r(\psi_1) = \sum_{j=1}^{s}(4j+1)R_{2j}^2(\psi_1)$ is convex for all $\psi_1 \in (\psi_1^*, 1)$.

Fact 2 $h(\psi_1) := \frac{\psi_1\lambda(1) - r(\psi_1)}{1-\psi_1}$ is an an increasing function of $\psi_1 \in (\psi_1^*, 1)$.

Fact 3 $S_{2j}'(1) = \frac{1}{2}j(2j+1)$, where we define $S_{2j}(\psi_1) := P_{2j}(\sqrt{\psi_1})$. For example,

$$\begin{aligned} S_2(\psi_1) &= \frac{1}{2}(-1+3\psi_1) = R_2(\psi_1) \\ S_4(\psi_1) &= \frac{1}{8}(3-30\psi_1+35\psi_1^2) = R_4(\psi_1) \\ S_6(\psi_1) &= \frac{1}{16}(-5+105\psi_1-315\psi_1^2+231\psi_1^3) \ne R_6(\psi_1) \end{aligned}$$

Fact 4 For $j \ge 3$, we have

$$\begin{aligned} R_{2j}'(1) &= \begin{cases} S_{2j}'(1) - \sum_{k=1}^{(j-1)/2}(2k-1)\, p_{2j,4k} & j \text{ odd} \\ S_{2j}'(1) - \sum_{l=1}^{(j-2)/2}(2l)\, p_{2j,2(2l+1)} & j \text{ even} \end{cases} \\ &= S_{2j}'(1) - \sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}. \end{aligned}$$

We explain why these facts will provide the proof. Since $\lambda_1(1) = r(1)$, we rewrite $h(\psi_1)$ in terms of r, $h(\psi_1) = \frac{\psi_1 r(1) - r(\psi_1)}{1-\psi_1}$. By Fact 2,

$$\max_{\psi_1^* \le \psi_1 \le 1} h(\psi_1) = h(1) = r'(1) - r(1).$$

Combining the facts above gives an expression for $r'(1)$ as follows:

$$\begin{aligned}
r'(1) &= 2\sum_{j=1}^{s}(4j+1)R_{2j}(1)R'_{2j}(1) \\
&= 2\sum_{j=1}^{s}(4j+1)R'_{2j}(1) \quad \text{(because } R_{2j}(1)=1\text{)} \\
&= 2\sum_{j=1}^{2}(4j+1)S'_{2j}(1) + 2\sum_{j=3}^{s}(4j+1)R'_{2j}(1) \quad \text{(because } S_2 = R_2,\ S_4 = R_4\text{)} \\
&= 2\sum_{j=1}^{2}(4j+1)S'_{2j}(1) \\
&+ 2\sum_{j=3}^{s}(4j+1)\left(S'_{2j}(1) - \sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right) \\
&= 2\sum_{j=1}^{s}(4j+1)S'_{2j}(1) - 2\sum_{j=3}^{s}(4j+1)\left(\sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right) \\
&= \sum_{j=1}^{s}(4j+1)j(2j+1) - 2\sum_{j=3}^{s}(4j+1)\left(\sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
r'(1) - r(1) &= \sum_{j=1}^{s}(4j+1)(2j^2+j-1) \\
&- 2\sum_{j=3}^{s}(4j+1)\left(\sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right) \\
&= \frac{1}{2}s(2s+3)(2s^2+3s-1) \\
&- 2\sum_{j=3}^{s}(4j+1)\left(\sum_{k=1}^{j-2} k\, p_{2j,2k+2}\, I_{\{k+j \text{ even}\}}\right)
\end{aligned}$$

This proves Lemma 2.5.4. ∎

Proofs of Facts

1. It is enough to show that $R^2_{2j}(\psi_1)$ is convex for all $\psi_1 \in (\psi_1^*, 1)$. We will see

why $[R_{2j}(\psi_1)]'' = 2[R'_{2j}R'_{2j} + R_{2j}R''_{2j}] \geq 0$ for all $\psi_1 \in (\psi_1^*, 1)$. Remember that ψ_1^* was defined in such a way that $R_{2j}(\psi_1) \geq 0$ for all $\psi_1 \in (\psi_1^*, 1)$, and $R''_{2j}(\psi_1) \geq 0$ for all ψ_1 since all the negative terms in $R_{2j}(\psi_1)$ are linear in ψ_1. Consequently, $R^2_{2j}(\psi_1)$ is convex and so is $r(\psi_1)$ for all $\psi_1 \in (\psi_1^*, 1)$.

2. The convexity of r implies that $r'(\psi_1) < (r(\psi_1) - r(1))/(\psi_1 - 1)$, which is equivalent to $h'(\psi_1) > 0$, and this shows that $h(\psi_1)$ is an increasing function.

3. Taking the transformation $\psi_1 = y^2$, we have

$$\frac{d}{d\psi_1} S_{2j}(\psi_1) = \frac{1}{2\sqrt{\psi_1}} \frac{d}{dy} P_{2j}(y),$$

where P_{2j} is the $(2j)$th Legendre polynomial. Using the properties $(y^2 - 1)P'_j = j(yP_j - P_{j-1})$ and $yP'_j = jP_j + P'_{j-1}$, we obtain

$$\begin{aligned} P'_{2j}(y)|_{y=1} &= \frac{2j}{y^2 - 1}[yP_{2j}(y) - P_{2j-1}(y)]|_{y=1} \\ &= \frac{j}{y}[P_{2j}(y) + yP'_{2j}(y) - P'_{2j-1}(y)]|_{y=1} \\ &= \frac{j}{y}[P_{2j}(y) + 2jP_{2j}(y)]|_{y=1} \\ &= j(2j + 1). \end{aligned}$$

Thus, $S'_{2j}(1) = (j/2)(2j + 1)$.

4. Fact 4 is a simple observation and so the proof is omitted. We write out the values of ν_s for even q, in Table 2.1.

So far we have been concerned with only when q is even. How do we prove Theorem 2.5.2 for odd q? The proof will be parallel to the case of even q, and we are able to obtain a formula similar to Lemma 2.5.4. The only differences are in

Table 2.1: Values of $\psi_1^*, \nu_{s1}, \nu_{s2}$ and $\nu_s = \max\{\nu_{s1}, \nu_{s2}\}$ as in Theorem 2.5.2. Fitted response is a straight line: true response is a polynomial of degree even q. For $\nu \geq \nu_s$ the minimax design is $\delta_{\pm 1/2}$.

q	$s := q/2$	$\nu_{s1} = \nu_s$	ν_{s2}	ψ_1^*
2	1	10	2.46	0.33
4	2	91	40.17	0.7416
6	3	862.88	387.74	0.9652
8	4	7839.25	2841.27	0.995

$g(\boldsymbol{\psi})$ and $r(\psi_1)$, which depend on λ_2 :

$$g(\boldsymbol{\psi}) = \frac{1}{\psi_1}[1 + \frac{1}{\nu}\lambda_2(\boldsymbol{\psi})].$$

$$r(\psi_1) = \frac{1}{\psi_1}\sum_{j=1}^{s-1}\frac{1}{(4j+3)}[(2j+1)R_{2j}(\psi_1) + 2(j+1)R_{2(j+1)}(\psi_1)]^2.$$

Here $s = (q+1)/2$, and we are assuming that $\lambda_2(\mathbf{1}) \geq \lambda_1(\mathbf{1})$. This assumption is verified by noticing that

$$\begin{aligned}
\lambda_2(\mathbf{1}) &= \sum_{j=1}^{s-1}(1/(4j+3))[(2j+1)Q_{2j}(\mathbf{1}) + 2(j+1)Q_{2(j+1)}(\mathbf{1})]^2 \\
&= \sum_{j=1}^{s-1}(4j+3) = (s-1)(2s+3). \\
\lambda_1(\mathbf{1}) &= \sum_{j=1}^{s-1}(4j+1) = (s-1)(2s+1).
\end{aligned}$$

The only difficulty is to prove that $r(\psi_1)$ is convex as we proved in Fact 1 for the previous case. Here $r(\psi_1)$ is not so simple as before, and it is helpful to follow the proof if we rewrite $\lambda_2(\boldsymbol{\psi})$ in terms of the coefficients of the odd Legendre polynomials. Since

$$\begin{aligned} E_\xi\, xl_{2j+1} &= (\sqrt{4j+3})\, E_\xi[xP_{2j+1}(2x)] \\ &= \sqrt{4j+3}/2\, E_\xi[2x\, P_{2j+1}(2x)] \\ &= \sqrt{4j+3}/2\, Q_{2j+1}(\boldsymbol{\psi}). \end{aligned}$$

we can rewrite the second eigenvalue in terms of $Q_{2j+1}(\boldsymbol{\psi})$.

$$\lambda_2(\boldsymbol{\psi}) = (4/\psi_1)\sum_{j=1}^{s-1}[E_\xi\, xl_{2j+1}(x)]^2 = (1/\psi_1)\sum_{j=1}^{s-1}(4j+3)Q_{2j+1}^2(\boldsymbol{\psi}).$$

As an illustration. we mention a few terms above

$$\begin{aligned} E_\xi\, 2xP_3(2x) &= E_\xi\, 2x[(5/2)(2x)^3 - (3/2)(2x)] = E_\xi\,[(5/2)(2x)^4 - (3/2)(2x)^2]. \\ Q_3(\boldsymbol{\psi}) &= (1/2)(-3\psi_1 + 5\psi_2) := p_{3,1}\,\psi_1 + p_{3,3}\,\psi_2 \\ Q_5(\boldsymbol{\psi}) &= (1/8)(15\psi_1 - 70\psi_2 + 63\psi_3) := p_{5,1}\,\psi_1 + p_{5,3}\,\psi_2 + p_{5,5}\,\psi_3 \\ Q_7(\boldsymbol{\psi}) &= (1/16)(-35\psi_1 + 315\psi_2 - 693\psi_3 + 429\,\psi_4) \\ &:= p_{7,1}\,\psi_1 + p_{7,3}\,\psi_2 + p_{7,5}\,\psi_3 + p_{7,7}\,\psi_5. \end{aligned}$$

Similar to (2.5.15) we express $r(\psi_1)$ as

$$r(\psi_1) := (1/\psi_1)\sum_{j=1}^{s-1}(4j+3)R_{2j+1}^2(\psi_1).$$

Some of the $R_{2j+1}(\psi_1)$'s are

$$\begin{aligned} R_3(\psi_1) &= (1/2)(-3\psi_1 + 5\psi_1^2) := p_{3,1}\,\psi_1 + p_{3,3}\,\psi_1^2 \\ R_5(\psi_1) &= (1/8)(15\psi_1 - 70\psi_1 + 63\psi_1^3) := p_{5,1}\,\psi_1 + p_{5,3}\,\psi_1 + p_{5,5}\,\psi_1^3 \\ R_7(\psi_1) &= (1/16)(-35\psi_1 + 315\psi_1^2 - 693\psi_1 + 429\,\psi_1^4). \end{aligned}$$

We state a similar result to Lemma 2.5.4 and give a proof.

Lemma 2.5.5

$$\max_{\psi_1^* \le \psi_1 \le 1} \frac{\psi_1 \lambda_2(1) - r(\psi_1)}{1-\psi_1} = \frac{1}{2}(s-1)s(2s+1)(2s+3)$$
$$- \; 2\sum_{j=2}^{s-1}(4j+3)\left(\sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j\ odd\}}\right)$$

Proof: We first determine that $h(\psi_1) := \frac{\psi_1 r(1) - r(\psi_1)}{1-\psi_1}$ is an increasing function for $\psi_1^* \le \psi_1 \le 1$. This is true as long as $r(\psi_1) = (1/\psi_1)\sum(4j+3)R_{2j+1}^2(\psi_1)$ is convex when ψ_1 lies between ψ_1^* and 1. We claim it is convex and prove it by showing that $G(\psi_1) := (1/\psi_1)R_{2j+1}^2(\psi_1)$ is convex. We will show that $G''(\psi_1) \ge 0$ for all $\psi_1^* \le \psi_1 \le 1$.

A simple calculation shows that $G''(\psi_1) \ge 0$ iff

$$\psi_1^2 R'_{2j+1}(\psi_1)R'_{2j+1}(\psi_1) + \psi_1^2 R_{2j+1}(\psi_1)R''_{2j+1}(\psi_1)$$
$$-\psi_1 R_{2j+1}(\psi_1)R'_{2j+1}(\psi_1) + R_{2j+1}^2(\psi_1) \ge 0.$$

which is equivalent to

$$\psi_1 R'_{2j+1}(\psi_1)(\psi_1 R'_{2j+1}(\psi_1) - 2R_{2j+1}(\psi_1)) \ge 0.$$

Substituting $R_{2j+1}(1) = 1$ and using the definition of ψ_1^*, we find that $R'(\psi_1) > 0$, for all $\psi_1^* \le \psi_1 \le 1$. Our task is now to show that $(\psi_1 R'(\psi_1) - 2R(\psi_1)) \ge 0$. It is easier to see if we, in $R_{2j+1}(\psi_1)$, group the high power terms and linear terms separately.

$$R_{2j+1}(\psi_1) = \begin{cases} \sum_{i=1}^{(j+1)/2} p_{2j+1,4i-1}\, \psi_1^{2i} + \sum_{i=1}^{(j+1)/2} p_{2j+1,4i-3}\, \psi_1 & (j \text{ odd}) \\ \sum_{i=1}^{(j/2)} p_{2j+1,4i+1}\, \psi_1^{2i+1} + (p_{2j+1,1} + \sum_{i=1}^{j/2} p_{2j+1,4i-1})\, \psi_1 & (j \text{ even}). \end{cases}$$

We look only at the case when j is odd, as the other case is very similar. The relationship

$$\psi_1 R'_{2j+1}(\psi_1) = \sum_{i=1}^{(j+1)/2} (2i)\, p_{2j+1,4i-1}\, \psi_1^{2i} + \sum_{i=1}^{(j+1)/2} p_{2j+1,4i-3}\, \psi_1$$

implies that

$$\begin{aligned} \psi_1 R'_{2j+1}(\psi_1) - 2R_{2j+1}(\psi_1) &= \sum_{i=1}^{(j+1)/2} (2i-1)\, p_{2j+1,4i-1}\, \psi_1^{2i} \\ &- \sum_{i=1}^{(j+1)/2} p_{2j+1,4i-3}\, \psi_1. \end{aligned} \quad (2.5.18)$$

In equation (2.5.18), it is clear that the first term is positive; the second term is negative because the coefficients $p_{2j+1,4i-3}$ are negative. Thus the left hand side of (2.5.18) is positive, so that $r(\psi_1)$ is convex and $h(\psi_1)$ is an increasing function of ψ_1 for all $\psi_1^* \leq \psi_1 \leq 1$. So the maximum of $h(\psi_1)$ occurs at 1. That is,

$$\max_{\psi_1^* \leq \psi_1 \leq 1} h(\psi_1) = \frac{\psi_1 r(1) - r(\psi_1)}{1-\psi_1}\Big|_{\psi_1=1} = r'(1) - r(1).$$

It remains to determine $r'(1)$. Differentiating $r(\psi_1)$ with respect to ψ_1 gives $r'(\psi_1) = (1/\psi_1^2)(2\psi_1 \sum (4j+3) R_{2j+1}(\psi_1) R'_{2j+1}(\psi_1) - \sum (4j+3) R^2_{2j+1}(\psi_1))$. To calculate $r'(1)$, it is useful to define $S_{2j+1}(\psi_1)$ similar to what we did for even q. Define $S_{2j+1}(\psi_1) = \sqrt{\psi_1} P_{2j+1}(\sqrt{\psi_1})$. So then $R'_3(1) = S'_3(1)$, and

$$\begin{aligned} R'_{2j+1}(1) &= \begin{cases} S'_{2j+1}(1) - \sum_{k=1}^{j/2} (2k-1) p_{2j+1,4k-1} & \text{when } j \geq 2 \text{ is even} \\ S'_{2j+1}(1) - \sum_{l=1}^{(j-1)/2} (2l) p_{2j+1,4l+1} & \text{when } j \geq 3 \text{ is odd.} \end{cases} \\ &= S'_{2j+1}(1) - \sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j \text{ odd}\}}. \end{aligned}$$

Using the relation $P'_{2j+1}(y) = (2j+1)[y P_{2j+1}(y) - P_{2j}(y)]/(y^2-1)$, we obtain

$$S'_{2j+1}(1) = (1/2)(P_{2j+1}(1) + P'_{2j+1}(1)) = (1/2)(2j^2 + 3j + 2)).$$

We are now in a position to conclude

$$
\begin{aligned}
r'(1) - r(1) &= 2\sum_{j=1}^{s-1}(4j+3)R'_{2j+1}(1) - 2\sum_{j=1}^{s-1}(4j+3)R_{2j+1}(1) \\
&= 2\sum_{j=1}^{s-1}(4j+3)[-1 + S'_{2j+1}(1) - \sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j \text{ odd}\}}] \\
&= \sum_{j=1}^{s-1}(4j+3)(2j^2+3j) \\
&\quad -2\sum(4j+3)\left(\sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j \text{ odd}\}}\right) \\
&= \frac{1}{2}(s-1)s(2s+1)(2s+3) \\
&\quad -2\sum_{j=2}^{s-1}(4j+3)\left(\sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j \text{ odd}\}}\right)
\end{aligned}
$$

■

We present some values of ν_s when q is odd in Table 2.2.

Table 2.2: Values of $\omega_1^*, \nu_{s1}, \nu_{s2}$ and $\nu_s = \max\{\nu_{s1}, \nu_{s2}\}$ as in Theorem 2.5.2. Fitted response is a straight line; true response is a polynomial of degree odd q. For $\nu \geq \nu_s$ the minimax design is $\delta_{\pm 1/2}$.

q	$s := (q+1)/2$	$\nu_{s1} = \nu_s$	ν_{s2}	ω_1^*
3	2	35	10.5	0.6
5	3	381.5	82.0	0.82

So far we have analyzed and obtained the optimal designs for small ν. $(0 \leq \nu \leq (14/5))$, or large ν. $(\nu \geq \nu_s)$. What are the optimal designs for $(14/5) \leq \nu \leq \nu_s$?

If $\nu \in (14/5, \nu_s)$, the optimal solution lies on the boundary of the moment space M_s, and so by Corollary 2.3.6, the design distribution is uniquely determined. Before giving examples, we take a look at the eigenvalues λ_1, λ_2, when $2 \leq q \leq 6$. We make two observations. First, when q is odd, the second eigenvalue $\lambda_2(\psi)$ utilizes one more moment. Secondly, the eigenvalues repeat. The first eigenvalues $\lambda_1(\psi)$ are the same when $q = 2$ and 3, $q = 4$ and 5, the second eigenvalues are the same when $q = 3$ and 4, $q = 5$ and 6. We close this section by illustrating how we obtain the optimal designs for the cases $q = 2, 3, 4, 5,$ and 6.

Table 2.3: Eigenvalues when true response is Legendre polynomial of degree q where ω_i is denoted by $2^{2i}\xi_{2i}$.

q	$\lambda_1(\omega)$	$\lambda_2(\omega)$
2	$(5/4)(3\omega_1 - 1)^2$	none
3	$(5/4)(3\omega_1 - 1)^2$	$(7/4\omega_1)(5\omega_2 - 3\omega_1)^2$
4	$(5/4)(3\omega_1 - 1)^2+$ $(9/64)(3 - 30\omega_1 + 35\omega_2)^2$	$(7/4\omega_1)(5\omega_2 - 3\omega_1)^2$
5	$(5/4)(3\omega_1 - 1)^2+$ $(9/64)(3 - 30\omega_1 + 35\omega_2)^2$	$(1/\omega_1)[(7/4)(5\omega_2 - 3\omega_1)^2+$ $(11/64)(15\omega_1 - 70\omega_2 + 63\omega_3)^2]$
6	$(5/4)(3\omega_1 - 1)^2+$ $(9/64)(3 - 30\omega_1 + 35\omega_2)^2+$ $(13/256)(231\omega_3 - 315\omega_2 + 105\omega_1 - 5)^2$	$(1/\omega_1)[(7/4)(5\omega_2 - 3\omega_1)^2+$ $(11/64)(15\omega_1 - 70\omega_2 + 63\omega_3)^2]$

Example 1: $q = 2$.

The experimenter fits a linear model, although the true model might be quadratic. It is simple enough to obtain the optimal design by a direct calculation.

$$\xi^* = \begin{cases} \frac{1}{2}\delta_{-\frac{1}{2}} + \frac{1}{2}\delta_{+\frac{1}{2}} & \text{if } \nu \geq 10 \\ \frac{1}{2}\delta_{-\frac{1}{2}\sqrt{\omega_1^*}} + \frac{1}{2}\delta_{+\frac{1}{2}\sqrt{\omega_1^*}} & \text{if } \nu < 10 \end{cases} \qquad \omega_1^* = \sqrt{\frac{4\nu+5}{45}}$$

We notice that ν_1 also equals 10 in the table of the examples of ν_s.

Example 2 : q is 3 or 4.

In this case there are 3 support points including zero on $[-1/2, 1/2]$, for small $\nu \in (0, 14/5)$. From Table 2.1 and Table 2.2, we have $\nu_s = 35$ when $q = 3$, and $\nu_s = 91$ when $q = 4$. The optimal solution in terms of the moments ω_1 and ω_2, is given below. We state it for $q = 3$, with the relevant changes for $q = 4$ in brackets.

$$(\omega_1^*, \omega_2^*) = \begin{cases} (1, 1) & \text{if } \nu \geq 35 \quad (91) \\ (c^*, c^{*2}) & \text{if } 14/5 \leq \nu \leq 35 \quad (91) \\ \left(\sqrt{\frac{4\nu+5}{45}}, \frac{30\omega_1^*-3}{35}\right) & \text{if } \nu \leq 14/5 \end{cases}$$

where c^* is a minimizer in $[0, 1]$ of the loss function, ignoring the constant term, $l(\mathbf{c}) = (1/c)(\nu + \max\{\lambda_1(\mathbf{c}), \lambda_2(\mathbf{c})\})$, where $\mathbf{c} = (c, c^2)$ and $c \in [0, 1]$. Thus the optimal design is

$$\xi^* = \begin{cases} \frac{1}{2}\delta_{-\frac{1}{2}} + \frac{1}{2}\delta_{+\frac{1}{2}} & \text{if } \nu \geq 35 \quad (91). \\ \frac{1}{2}\delta_{-\frac{1}{2}\sqrt{c^*}} + \frac{1}{2}\delta_{+\frac{1}{2}\sqrt{c^*}} & \text{if } 14/5 \leq \nu \leq 35 \quad (91). \\ (1-\alpha)\delta_0 + \frac{\alpha}{2}\delta_{-\sqrt{x_1}} + \frac{\alpha}{2}\delta_{+\sqrt{x_1}} & \text{if } \nu < 14/5. \end{cases}$$

where

$$\alpha = \frac{35\psi_1^{*2}}{30\psi_1^* - 3} \text{ and } x_1 = \frac{30\psi_1^* - 3}{140\psi_1^*}.$$

Example 3 : q is 5 or 6.

There are 4 support points, $\pm x_1$ and $\pm x_2$, say. From the tables, we have ν_s = 381.5 or 862.88. The optimal solution in terms of moments, ψ_1, ψ_2, and ψ_3, is

$$(\psi_1^*, \psi_2^*, \psi_3^*) = \begin{cases} (1, 1, 1) & \text{if } \nu \geq 381.5 \quad (862.88) \\ \text{boundary of } M_3 & \text{if } 14/5 \leq \nu \leq 381.5 \quad (862.88) \\ \left(\sqrt{\frac{4\nu+5}{45}}, \frac{30\psi_1^*-3}{35}, \frac{15\psi_1^*-2}{21}\right) & \text{if } \nu < 14/5. \end{cases}$$

It is not easy to describe the boundary of M_3, but a portion of it is given by $(\psi_1, \psi_1^2, \psi_1^3)$ and this provides an optimal solution

$$\xi^* = \begin{cases} \frac{1}{2}\delta_{-\frac{1}{2}} + \frac{1}{2}\delta_{+\frac{1}{2}} & \text{if } \nu \geq 381.5 \quad (862.88). \\ \frac{1}{2}\delta_{-\frac{1}{2}\sqrt{c^*}} + \frac{1}{2}\delta_{+\frac{1}{2}\sqrt{c^*}} & \text{if } 14/5 \leq \nu \leq 381.5 \quad (862.88). \\ \alpha\delta_{\pm\sqrt{x_1}} + (1-\alpha)\delta_{\pm\sqrt{x_2}} & \text{if } \nu < 14/5. \end{cases}$$

where c^* is the minimizer of the loss function $l(\mathbf{c}) = (1/c)(\nu + \max\{\lambda_1(\mathbf{c}), \lambda_2(\mathbf{c})\})$, where $\mathbf{c} = (c, c^2, c^3)$. For small ν, the three parameters α, x_1, and x_2 are determined by the three equations

$$\begin{aligned} \psi_1^* &= \alpha(2x_1)^2 + (1-\alpha)(2x_2)^2, \\ \psi_2^* &= \alpha(2x_1)^4 + (1-\alpha)(2x_2)^4, \\ \psi_3^* &= \alpha(2x_1)^6 + (1-\alpha)(2x_2)^6. \end{aligned}$$

2.5.3 Suggested minimax design for arbitrary q

Instead of presenting the minimax design for each q, we conclude this section by providing some guidelines when we fit a linear model but we are concerned that the true response is a polynomial of some unknown but arbitrary degree q. There will be three different cases depending on the size of ν.

1. The minimax design for $0 \leq \nu \leq (14/5)$:

   The optimal design is of the form $\xi^* = \sum_{i=1}^{s+1} \alpha_i \delta_{x_i}$ which consists of s unknowns. These unknowns are determined by the s equations (2.5.12).

2. A minimax design for medium size of ν, $(14/5) \leq \nu \leq \nu_s$:

   The optimal solutions in terms of moments are on the boundary of the moment space M_s, which is not so easy to describe. But a part of the boundary can be expressed by $(\psi_1, \psi_1^2, \ldots, \psi_1^s)$ and this suggests that for each ν, the optimal design ξ^* is determined by the minimizer c^* of the loss function $l(\mathbf{c}) = (1/c)(\nu + \max(\lambda_1(\mathbf{c}), \lambda_2(\mathbf{c})))$, where $\mathbf{c} = (c, c^2, \ldots, c^s)$ and $c \in (0, 1)$, that is, $\xi^* = \delta_{\pm\sqrt{c^*}/2}$. This minimization can be easily solved numerically.

3. The minimax design for large $\nu \geq \nu_s$:

   By Theorem 2.5.2, the minimax design consists of the extreme points $\pm 1/2$ with equal mass.

2.6 A-optimal designs for approximately linear model

It is not as straightforward as for the D-optimal case to obtain the A-optimal minimax design. In the proof of Theorem 2.5.1 we set up the problem in four steps. In the first step, we assumed that the first eigenvalue λ_1 is larger than the second eigenvalue λ_2. In the third step we verified that the first eigenvalue is indeed larger at the minimax density. For A-optimality this third step fails to be satisfied for small ν. Wiens (1992) suggests that one may construct a minimax design by minimizing the loss function while one of two eigenvalues is held fixed. For $\nu \geq 4/9$, the second eigenvalue is larger at the minimax density. Wiens obtains minimax design of the form $m(x) = a(x^2 - b/4)^+/x^2$.

2.6.1 Wiens's minimax design for SLR

For small ν, Wiens first constructs a density to minimize the first (second) eigenvalue subject to the second (first) eigenvalue being fixed. The density is of the form

$$m_1(x; t, \gamma) = \left(\frac{a + bx^2}{1 + cx^2}\right)^+, \quad \left(m_2(x; t, \gamma) = \left(\frac{a_1 + b_1 x^2}{c_1 + x^2}\right)^+\right)$$

where the coefficients a, b, c satisfy

$$\int m(x; t, \gamma)\, dx = 1, \quad \int x^2 m(x; t, \gamma)\, dx = \gamma,$$

$$\lambda_1(m(x; t, \gamma)) = t \quad (\lambda_2(m(x; t, \gamma)) = t).$$

One then determines (t^*, γ^*) to minimize the supremum of the loss function. Then if $0 \le \nu \le 4/9$, the minimax density is

$$m(x; t^*, \gamma^*) = \begin{cases} m_1(x; t^*, \gamma^*) & \text{if } \sup_{\mathcal{F}} \mathcal{L}_A(m_1(x; t^*, \gamma^*)) \le \sup \mathcal{L}_A(m_2(x; t^*, \gamma^*)), \\ \\ m_2(x; t^*, \gamma^*) & \text{otherwise.} \end{cases}$$

2.6.2 Minimax design based on Rychlik's approach

For small ν, we are not able to obtain a result like Theorem 2.5.1 in this case, but for large ν, we have result similar to Theorem 2.5.2. That is, the optimal design has two design points which are extreme. The proof of this result follows very easily and so we present the result without proof. The eigenvalues are

$$\begin{aligned} \lambda_1(\psi) &= \sum_j [E_\xi l_{2j}(x)]^2 = (E_\xi \sqrt{4j+1} P_{2j}(2x))^2 \\ &= \sum_j (4j+1)[E_\xi P_{2j}(2x)]^2 = \sum_j (4j+1) Q_{2j}^2(\psi) \end{aligned}$$

$$\begin{aligned} \lambda_2(\psi) &= \frac{16}{\psi_1^2} \sum_{j=1}^{s-1} [a_j E_\xi l_{2j} + b_j E_\xi l_{2(j+1)}]^2 \\ &= \frac{16}{\psi_1^2} \sum_{j=1}^{s-1} [\frac{2j+1}{2\sqrt{4j+3}} Q_{2j}(\psi) + \frac{j+1}{\sqrt{4j+3}} Q_{2(j+1)}(\psi)]^2 \\ &= \sum_{j=1}^{s-1} \frac{4}{(4j+3)\psi_1^2} [(2j+1) Q_{2j}(\psi) + 2(j+1) Q_{2(j+1)}(\psi)]^2, \end{aligned}$$

where for the first eigenvalue, j runs from 1 to s or $s-1$, depending whether q is even or odd. We notice that the first eigenvalue is the same as the one for Q- and D-optimal cases, but the second eigenvalue is not. All the calculations are based on the fact that $\lambda_2(1) > \lambda_1(1)$ for all $q \ge 3$.

Theorem 2.6.1 *For every $s \geq 1$ there exists $\nu_s < \infty$ such that $(\delta_{-1/2} + \delta_{1/2})/2$ is minimax for $\nu \geq \nu_s$. The lower bound ν_s is determined by $\nu_s = \max\{\nu_{s1}, \nu_{s2}\}$, where*

$$\begin{aligned}
\nu_{s2} &= \max_{0 \leq \omega_1 \leq \omega^*} \frac{\omega_1 \lambda_2(\mathbf{1})}{4(1-\omega_1)}, \\
\nu_{s1} &= \max_{\omega_1^* \leq \omega_1 \leq 1} \frac{\omega_1}{4(1-\omega_1)}[\lambda_2(\mathbf{1}) - r(\omega_1)] \\
&= \frac{1}{2}(s-1)s(2s+1)(2s+3) \\
&\quad -2\sum_{j=2}^{s-1}(4j+3)\left(\sum_{k=1}^{j-1} k\, p_{2j+1,2k+1}\, I_{\{k+j \text{ odd}\}}\right)
\end{aligned}$$

Some values of ν_s are in Table 2.4.

Table 2.4: Values of $\omega_1^*, \nu_{s1}, \nu_{s2}$ and $\nu_s = \max\{\nu_{s1}, \nu_{s2}\}$ as in Theorem 2.6.1. Fitted response is a straight line: true response is a polynomial of degree even q. For $\nu \geq \nu_s$ the minimax design is $\delta_{\pm 1/2}$.

q	$s = [(q+1)/2]$	$\nu_{s1} = \nu_s$	ν_{s2}	ω_1^*
3	2	35	10.5	0.6
4	2	35	20.09	0.7416
5	3	381.5	82.0	0.82

2.6.3 Suggested minimax design for arbitrary q

We conclude this section by providing guidelines on how to fit a linear model when we are concerned that the true response is a polynomial of some unknown but arbitrary degree. There will be three different cases depending on the size of ν.

1. The minimax design for $0 \leq \nu \leq (4/9)$:

   We are not able to provide an optimal design. But we look at this problem in Chapter 4 with the contaminating space $\mathcal{F}_2$.

2. A minimax design for medium size of ν, $(4/9) \leq \nu \leq \nu_s$:

   An optimal design is of the form $\xi^* = \delta_{\pm\sqrt{c^*}/2}$, where c^* is the minimizer of the loss function $l(\mathbf{c}) = \nu(1 + 4/c) + \max\{\lambda_1(\mathbf{c}), \lambda_2(\mathbf{c})\}$, for each ν. The vector $\mathbf{c} = (c, c^2, \ldots, c^s)$ is as defined in the previous section.

3. The minimax design for large $\nu \geq \nu_s$:

   By Theorem 2.6.1, the optimal design consists of the extreme points with equal mass.

2.7 Final comment

The designs we have obtained may not have enough support points to fit the models against which we wish to protect, and so are clearly non-robust in this respect. For instance, as we have seen in Example 2, there is an optimal design with three support points when fitting a linear regression line although true model might be cubic or quartic. Since there are only three support points it is not possible to check whether or not the alternative model is appropriate. When $\nu \leq 14/5$, in this example we found an optimal design of the form

$$(1 - \alpha)\delta_0 + \alpha\delta_{\pm x_1}.$$

Consider the following perturbation of this optimal design for some small constant c

$$(1-\alpha)\delta_0 + \alpha\delta_{\pm(x_1 \pm c)}.$$

In this way, we obtain a close-to-optimal but safer design. The best way to choose the constant c will be the subject of future research.

Chapter 3

Model robust designs in polynomial regression II

In Chapter 2 the experimenter wanted a fitted response function that would be useful in predicting future y-values. Here in Chapter 3, we imagine that the original model function, a pth order polynomial, has been contaminated by the addition of some higher order terms only. We want the fitted response function to ignore the vagaries of the contamination and estimate the original model as closely as possible. Otherwise the problem is the same: to find a design that is optimal. It simplifies the mathematics to restrict our search to designs with minimal support. This is possible by using Theorem 3.3.2, which says that for any measure μ on $[a, b]$ not supported by p or fewer points, there exists a unique measure ξ on $[a, b]$ with $p+1$ support points, for which μ and ξ have the same first $2p+1$ moments. When the original model function is a pth order polynomial that might be contaminated by a $p+1$ order term only, for the D- and A-optimality cases, the maximum loss functions are simplified. This simplification makes it easy to prove the existence of a symmetric optimal design for the D-optimality criterion. In Section 2.2, we mentioned that for the A- and Q-optimality cases, it is not known if the optimal design is symmetric. When the contamination term is quadratic while the true model function is linear, we are able to prove that the optimal designs are symmetric.

3.1 Introduction

In this chapter the true coefficients of the lower order model are simply the first $p+1$ coefficients from the contaminated model.

More precisely, we set $\boldsymbol{z}_1(x) = (1, x, \ldots, x^p)^T$ and $\boldsymbol{z}_2(x) = (x^{p+1}, \ldots, x^q)^T$, where the regressors x range over the interval $[-1/2, 1/2]$. The experimenter fits, by least squares, the model

$$E(Y \mid x) = \boldsymbol{z}_1^T(x)\,\boldsymbol{\theta}, \quad \boldsymbol{\theta} \in \mathbb{R}^{p+1},$$

although the true model is

$$E(Y \mid x) = \boldsymbol{z}_1^T(x)\boldsymbol{\theta}_1 + \boldsymbol{z}_2^T(x)\boldsymbol{\beta}, \quad \boldsymbol{\theta}_1 \in \mathbb{R}^{p+1}, \ \boldsymbol{\beta} \in \mathbb{R}^{q-p}.$$

In contrast to Chapter 2, here the true coefficient vector coincides with $\boldsymbol{\theta}_1$, and the contamination term is simply $\boldsymbol{z}_2^T\boldsymbol{\beta}$. The parameter $\boldsymbol{\theta}_1$ is identifiable because any two polynomials that agree on a neighbourhood of zero have the same coefficients. For any choice of design points $(x_i)_{i=1}^n$ our observations will be given by

$$y_i = \boldsymbol{z}_1^T(x_i)\,\boldsymbol{\theta}_1 + \boldsymbol{z}_2^T(x_i)\,\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n.$$

where we assume additive, uncorrelated errors ε_i with common variance σ^2.

The quality of the least squares estimate $\hat{\boldsymbol{\theta}}$ will depend on the size of the contamination term $\boldsymbol{z}_2^T\boldsymbol{\beta}$ and on the placement of the design points. We assume that the contamination is small, in the sense that for some known $\eta > 0$, the function $\boldsymbol{z}_2^T\boldsymbol{\beta}$ belongs to

$$\mathcal{F} = \{\boldsymbol{z}_2^T\boldsymbol{\beta} : \int (\boldsymbol{z}_2^T(x)\boldsymbol{\beta})^2\, dx \leq \eta^2\}.$$

As in Chapter 2, the contamination space $\mathcal{F}$ defined above is a space of polynomials whose degree is less than or equal to q, but it is not the same contamination space $\mathcal{F}$ used in Chapter 2. In Chapter 3, the notation $\mathcal{F}$ refers to the space above.

For a given loss function $\mathcal{L}$ then, our problem is to choose design points that are robust against the worst possible contamination in $\mathcal{F}$, that is, to minimize $\sup_{\mathcal{F}} \mathcal{L}$. Our loss function will always depend on the support points $(x_i)_{i=1}^n$ only through the design measure $\xi = (1/n)\sum_{i=1}^n \delta_{x_i}$, so we recast our problem in terms of measures on $[-1/2, 1/2]$. In this chapter we let A_ξ and C_ξ, for any measure ξ on $[-1/2, 1/2]$, be given by

$$A_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x)\,\xi(dx), \quad C_\xi = \int \boldsymbol{z}_1(x)\boldsymbol{z}_2^T(x)\,\xi(dx).$$

This is the same as in Chapter 2 except $\boldsymbol{u}(x)$ is replaced by $\boldsymbol{z}_2(x)$. Note that A_ξ is a $(p+1)\times(p+1)$ matrix, while C_ξ is a $(p+1)\times(q-p)$ matrix.

The estimate $\hat{\boldsymbol{\theta}}$, the mean squared error of $\hat{\boldsymbol{\theta}}$ and three loss functions are derived in Chapter 2. See the equations from (2.1.10) to (2.1.14). We now want to obtain optimal designs in the sense of minimizing $\sup_{\mathcal{F}} \mathcal{L}_D$, $\sup_{\mathcal{F}} \mathcal{L}_A$ and $\sup_{\mathcal{F}} \mathcal{L}_Q$.

3.2 Suprema of loss functions

In this section we review the supremum of the loss function over $\mathcal{F}$ in terms of the design measure ξ. For convenience we define a $(q-p)\times(q-p)$ matrix by $B_0 = \int \boldsymbol{z}_2(x)\boldsymbol{z}_2^T(x)\,dx$, and let $\boldsymbol{\alpha} = B_0^{1/2}\boldsymbol{\beta}/\eta$, and $\nu = \sigma^2/n\eta^2$. The invertibility of the matrix B_0 follows from the same reasoning as in Chapter 2.

3.2.1 D-optimality criterion

We derived $\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta},\xi)$ in (2.2.3), here we merely rewrite it after rearranging the term $\nu = \sigma^2/n\eta^2$.

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta},\xi) = \left(\frac{\sigma^2}{n}\right)^p \frac{\eta^2}{|A_\xi|}\left(\nu + \lambda_{\max} G_\xi\right). \tag{3.2.1}$$

where $G_\xi = B_0^{-1/2} C_\xi^T A_\xi^{-1} C_\xi B_0^{-1/2}$.

3.2.2 A-optimality criterion

Similarly,

$$\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta},\xi) = \eta^2\left[\nu \,\mathrm{trace}\,(A_\xi^{-1}) + \lambda_{\max} H_\xi\right]. \tag{3.2.2}$$

where $H_\xi = B_0^{-1/2} C_\xi^T A_\xi^{-2} C_\xi B_0^{-1/2}$.

3.2.3 Q-optimality criterion

Although the idea is the same in Q-optimality case we obtain a little different formula and so we write it out properly. We begin by expanding

$$[\boldsymbol{z}_1^T\hat{\boldsymbol{\theta}} - E(Y \mid x)]^2 = [\boldsymbol{z}_1^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \boldsymbol{z}_2^T\boldsymbol{\beta}]^2$$

$$= z_1^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T z_1 - 2z_1^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)z_2^T\boldsymbol{\beta} + \boldsymbol{\beta}^T z_2 z_2^T \boldsymbol{\beta}.$$

Taking expected values gives

$$E[z_1^T\hat{\boldsymbol{\theta}} - E(Y \mid x)]^2 = z_1^T \mathrm{MSE}(\boldsymbol{\beta}, \xi) z_1 - 2\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1} z_1 z_2^T \boldsymbol{\beta} + \boldsymbol{\beta}^T z_2 z_2^T \boldsymbol{\beta}.$$

Recall that A_0 and C_0 denote the A_μ and C_μ matrices when μ is Lebesgue measure on $[-1/2, 1/2]$. Then we can rewrite the loss function as

$$\begin{aligned}
\mathcal{L}_Q(\boldsymbol{\beta}, \xi) &= \int E[z_1^T(x)\hat{\boldsymbol{\theta}} - E(Y \mid x)]^2\, dx \\
&= \mathrm{trace}[\mathrm{MSE}(\boldsymbol{\beta}, \xi) A_0] - 2\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1} C_0 \boldsymbol{\beta} + \boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \\
&= \mathrm{trace}[A_\xi^{-1} A_0] + \boldsymbol{\beta}^T C_\xi^T A_\xi^{-1} A_0 A_\xi^{-1} C_\xi \boldsymbol{\beta} \\
&\quad -2\boldsymbol{\beta}^T C_\xi^T A_\xi^{-1} C_0 \boldsymbol{\beta} + \boldsymbol{\beta}^T B_0 \boldsymbol{\beta}.
\end{aligned}$$

Taking the supremum over $\mathcal{F}$ gives

$$\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_Q(\boldsymbol{\beta}, \xi) &= \sup_{\boldsymbol{\beta}^T B_0 \boldsymbol{\beta} \le \eta^2} \mathcal{L}_Q(\boldsymbol{\beta}, \xi) \\
&= \left(\frac{\sigma^2}{n}\right) \mathrm{trace}(A_\xi^{-1} A_0) + \eta^2 (1 + \lambda_{\max} J_\xi) \\
&= \eta^2 \left[\nu\, \mathrm{trace}(A_\xi^{-1} A_0) + (1 + \lambda_{\max} J_\xi)\right]. \qquad (3.2.3)
\end{aligned}$$

where

$$J_\xi = B_0^{-1/2}(C_\xi^T A_\xi^{-1} A_0 A_\xi^{-1} C_\xi - C_0^T A_\xi^{-1} C_\xi - C_\xi^T A_\xi^{-1} C_0) B_0^{-1/2}.$$

Using the results of the next section we hope to get more concrete information on the nature of the optimal solution.

3.3 Number of support points

Having found tractable formulas for various loss functions, we turn our attention to finding the optimal design, that is the probability measure ξ that minimizes $\sup_{\mathcal{F}} \mathcal{L}(\beta, \xi)$. A closer inspection of the loss function reveals that, since we are doing polynomial regression, $\sup_{\mathcal{F}} \mathcal{L}(\beta, \xi)$ depends on ξ only through its first $p + q$ moments.

The main result of this section is Theorem 3.3.2 which gives the minimum number of support points necessary to define a measure whose first few moments are specified. The proof depends on Lemma 2.1.1 which clarifies the relationship between the value of p, the matrix A_μ, and the number of support points of μ.

Lemma 3.3.1 *If the measure μ has exactly $p + 1$ support points, i.e., $\mu = \sum_{i=0}^{p} \alpha_i \delta_{x_i}$, then* $\det A_\mu = \prod_{i=0}^{p} \alpha_i \prod_{i \neq j} (x_i - x_j)^2$.

Proof: When μ has exactly $p + 1$ support points, then we may rewrite A_μ as

$$A_\mu = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_p \\ \vdots & \vdots & \ddots & \vdots \\ x_0^p & x_1^p & \cdots & x_p^p \end{pmatrix} \begin{pmatrix} \alpha_0 & 0 & \cdots & 0 \\ 0 & \alpha_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_p \end{pmatrix} \begin{pmatrix} 1 & x_0 & \cdots & x_0^p \\ 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_p & \cdots & x_p^p \end{pmatrix}.$$

The matrix on the left is a Vandermonde matrix, and so its determinant is given by $\prod_{i \neq j} (x_i - x_j)$. The result now follows easily. ■

Theorem 3.3.2 *For any measure μ on $[a, b]$ not supported by p or fewer points, there exists a unique measure ξ on $[a, b]$ with $p + 1$ support points, for which μ and ξ have the same first $2p + 1$ moments.*

Proof: We begin by defining

$$\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix}^{-1} \begin{pmatrix} \mu_{p+1} \\ \mu_{p+2} \\ \vdots \\ \mu_{2p+1} \end{pmatrix}. \tag{3.3.1}$$

where the invertibility of the matrix of moments is guaranteed by our assumption on μ and Lemma 2.1.1. Setting $\phi(x) = c_p x^p + \cdots + c_1 x + c_0$, we may rewrite the matrix equation above as the system

$$\int \phi(x) x^i \, \mu(dx) = \int x^{p+1+i} \, \mu(dx), \quad i = 0, 1, \ldots, p. \tag{3.3.2}$$

Setting $\omega(x) = x^{p+1} - \phi(x)$, this system tells us that

$$\int \omega(x) x^i \, \mu(dx) = 0, \quad i = 0, 1, \ldots, p$$

in other words, that $\int \omega r \, d\mu = 0$ for any polynomial r of degree less than or equal to p. In particular, if ω can be factored non-trivially as $\omega = rs$, then

$$\int r^2(x) s(x) \, \mu(dx) = 0. \tag{3.3.3}$$

We claim that the polynomial ω has $p+1$ distinct real roots $(x_i)_{i=0}^p$ lying in the interval $[a, b]$. First of all, if ω had a complex root we could find a factorization of the type $\omega(x) = r(x)(x^2 + d^2)$ with $d \neq 0$. Then (3.3.3) implies that μ is supported by the $p-1$ or fewer roots of r, which contradicts our assumption about μ. Similarly, a multiple root would allow a factorization $\omega(x) = r(x)(x - d)^2$, which leads to the same contradiction. Finally, for any root d of ω we have $\omega(x) = r(x)(x - d)$ and so

$$\int r^2(x)(x - d) \, \mu(dx) = 0.$$

Since μ is not supported by the roots of r, it follows that $x - d$ must vanish somewhere in the interval $[a, b]$, in other words, $a \le d \le b$.

The points $(x_i)_{i=0}^{p}$ will be the support points for our measure ξ. We find the masses by solving the equation

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_p \\ \vdots & \vdots & \ddots & \vdots \\ x_0^p & x_1^p & \cdots & x_p^p \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}.$$

Note that since the x_i's are distinct, the matrix above is a Vandermonde matrix and hence invertible. Define the measure $\xi = \sum_{i=0}^{p} \alpha_i \delta_{x_i}$. Clearly, $\omega(x) = \prod_{i=0}^{p}(x - x_i)$ vanishes ξ-almost everywhere and so

$$\int \omega(x) x^i \, \xi(dx) = \int x^{p+1+i} \, \xi(dx), \quad i = 0, 1, \ldots. \tag{3.3.4}$$

By construction, ξ has the same first p moments as μ, and so setting $i = 0$ in (3.3.2) and (3.3.4), we get

$$\mu_{p+1} = \int \omega(x) \mu(dx) = \int \omega(x) \, \xi(dx) = \xi_{p+1},$$

which means that the $(p+1)$th moment is also the same. Continuing in this way for i from 1 to p, we conclude that ξ has the same first $2p+1$ moments as μ. We now prove that the weights α_i are positive. Since μ and ξ share the same first $2p+1$ moments, they give the same integral for any polynomial of degree $2p+1$ or less. For any i between 0 and p, define $r_i(x) = \prod_{j \ne i}(x - x_j)$. Then r_i is a polynomial of degree p, so r_i^2 has degree $2p$ and thus

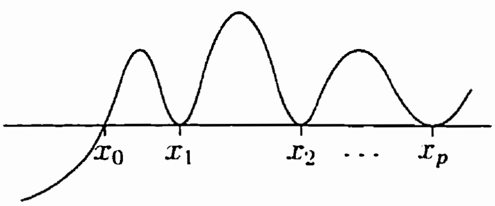
$$\alpha_i \prod_{j \ne i}(x_i - x_j)^2 = \int r_i^2(x) \, \xi(dx) = \int r_i^2(x) \, \mu(dx) > 0.$$

where the strict inequality follows from our assumption on the support of μ.

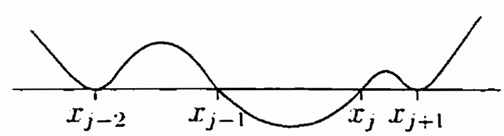
The only thing left to show is uniqueness. Let ξ be a measure with $p+1$ support points at $x_0 < x_1 < \cdots < x_p$ and τ any other measure with the same first $2p+1$ moments as ξ. We consider the following polynomials, each of which has degree of $2p+1$ or less.

Figure 3.1 Some of ψ functions.

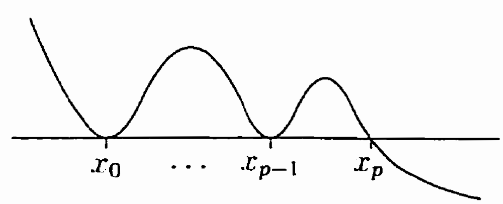
$$\psi_0(x) = (x - x_0)\prod_{i=1}^{p}(x - x_i)^2$$



$$\psi_j(x) = (x_{j-1} - x)(x - x_j)\prod_{i \neq j-1, j}(x - x_i)^2$$



$$\psi_{p+1}(x) = (x_p - x)\prod_{i=0}^{p-1}(x - x_i)^2$$



We have $\psi_i \equiv 0$, ξ—almost everywhere, $i = 0, 1, \ldots, p+1$ and by the shared moments $\int \psi_i \, d\tau = 0$, $i = 0, 1, \ldots, p+1$. Suppose $\tau(\mathbb{R} \setminus \{x_0, \ldots, x_p\}) > 0$ so that none of ψ_i is identically equal to zero τ-a.e. Then τ must assign mass to the $p+2$ disjoint intervals $\{\psi_i < 0\}_{i=0}^{p+1}$. Conversely, this shows that if τ has $p+1$ support points, then they must be precisely $x_0, \ldots, x_p$. In this case, the masses of τ are found by solving

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_p \\ \vdots & \vdots & \ddots & \vdots \\ x_0^p & x_1^p & \cdots & x_p^p \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_p \end{pmatrix}.$$

so that $\tau = \xi$. This concludes the proof. ∎

In our regression setting we work on the space of the symmetric measures and thus we must clarify that the above theorem holds there as well.

Theorem 3.3.3 *If the measure μ in Theorem 3.3.2 is symmetric about 0, then so is ξ.*

Proof: The measure ξ^- has $p+1$ support points and shares the first $2p+1$ moments with μ, the odd moments being zero. By the uniqueness in Theorem 3.3.2, we have $\xi^- = \xi$, in other words, ξ is symmetric. ∎

From now on ξ refers to the measure with $p+1$ support points whose existence is guaranteed by Theorem 3.3.2. We denote by ξ_k the kth moment of the measure ξ.

3.4 Some results when $q - p = 1$

Suppose now that $q - p = 1$ so that $C_\xi = (\xi_{p+1}, \ldots, \xi_{2p+1})^T$, and B_0 is a positive scalar.

Lemma 3.4.1 *If ξ has $p+1$ support points $(x_i)_{i=0}^p$, then*

$$\begin{aligned} A_\xi^{-1} C_\xi &= -\mathbf{c}, \\ C_\xi^T A_\xi^{-1} C_\xi &= \xi_{2p+2}, \end{aligned}$$

where the elements of the vector $\mathbf{c}$ *are the non-leading coefficients of the polynomial* $\prod_{i=0}^p (x - x_i)$.

Proof: Define $o(x) = \prod_{i=0}^p (x - x_i) = c_0 + c_1 x + \cdots + c_p x^p + x^{p+1}$ and note that o vanishes μ-almost everywhere. Letting $\mathbf{c} = (c_0, \ldots, c_p)^T$ and integrating gives

$$0 = \begin{pmatrix} \int o(x)\xi(dx) \\ \vdots \\ \int x^p o(x)\xi(dx) \end{pmatrix} = A_\xi \mathbf{c} + \begin{pmatrix} \xi_{p+1} \\ \vdots \\ \xi_{2p+1} \end{pmatrix} = A_\xi \mathbf{c} + C_\xi.$$

and multiplying the equation above by A_ξ^{-1} gives the first result. Multiplying the equation above by $C_\xi^T A_\xi^{-1}$ gives $C_\xi^T A_\xi^{-1} C_\xi = -C_\xi^T \mathbf{c}$. But

$$\begin{aligned} -C_\xi^T \mathbf{c} &= -(c_0 \xi_{p+1} + \cdots + c_p \xi_{2p+1}) \\ &= \xi_{2p+2} - (c_0 \xi_{p+1} + \cdots + c_p \xi_{2p+1} + \xi_{2p+2}) \\ &= \xi_{2p+2} - \int x^{p+1} o(x) \xi(dx) \\ &= \xi_{2p+2}. \end{aligned} \tag{3.4.1}$$

■

Observations when $q - p = 1$

1. For any measure μ, the supremum of the loss function $\sup_{\mathcal{F}} \mathcal{L}(\boldsymbol{\beta}, \mu)$ is a function of the first $2p+1$ moments of μ. But applying Theorem 3.3.3 we obtain a discrete measure ξ with $p+1$ support points that shares the first $2p+1$ moments, and hence gives the same loss. For the measure ξ, Lemma 3.4.1 tells us that $C_\xi^T A_\xi^{-1} C_\xi = \xi_{2p+2}$ and so

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \mu) = \sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) = \left(\frac{\sigma^2}{n}\right)^p \frac{\eta^2}{|A_\xi|} \left(\nu + B_0^{-1} \xi_{2p+2}\right). \quad (3.4.2)$$

For any measure ξ, let ξ^- stand for the image measure under the mapping $x \mapsto -x$, and let $\tilde{\xi}$ be the symmetrized measure $(1/2)(\xi + \xi^-)$. It is known that $|A_\xi| \leq |A_{\tilde{\xi}}|$, and since ξ_{2p+2} is the same for ξ and $\tilde{\xi}$, we conclude that the D-optimal measure must be symmetric. But it doesn't make the minimization procedure easier.

2. We notice that for the A-and Q-optimality cases the "bias" term is independent of point masses, α_i. From Lemma 3.4.1, $C_\xi^T A_\xi^{-2} C_\xi = \mathbf{c}^T \mathbf{c}$ and $A_\xi^{-1} C_\xi = -\mathbf{c}$, and the vector $\mathbf{c}$ depends only on design points x_i. And hence the bias terms H_ξ and J_ξ become

$$B_0^{-1/2} \mathbf{c}^T \mathbf{c} B_0^{-1/2} \text{ and } B_0^{-1/2} (\mathbf{c}^T A_0 \mathbf{c} + C_0^T \mathbf{c} + \mathbf{c}^T C_0) B_0^{-1/2}.$$

3. For any matrix map $t \to A(t)$ that is linear in t, the mapping $t \mapsto \text{trace}(A^{-1}(t))$ is convex because

$$(A^{-1})'' = 2A^{-1}A'A^{-1}A'A^{-1}$$

and

$$\frac{\partial^2}{\partial^2 t}[\text{trace}(A^{-1}(t))] = \text{trace}[\frac{\partial^2}{\partial^2 t}(A^{-1}(t))] \geq 0.$$

This means that the "variance" term of $\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi)$ is convex. (we make a note that this claim is true for arbitrary p and q not necessarily $q - p = 1$.)

1. A counterexample shows that "bias" terms of $\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi)$ as well as $\sup_{\mathcal{F}} \mathcal{L}_Q(\boldsymbol{\beta}, \xi)$ are not convex. Let $p = 1$, $q = 2$ and consider the measure $\xi = \frac{1}{10}\delta_{-.4} + \frac{9}{10}\delta_{.5}$. This gives $\xi_1 = .4100$, $\xi_2 = .241$, and $\xi_3 = .1061$ and so $\text{bias}_A(\xi) = 4.0$, $\text{bias}_Q(\xi) = 0.73$ whereas $\text{bias}_A(\tilde{\xi}) = 4.6465$, $\text{bias}_Q(\tilde{\xi}) = 2.433$, where $\tilde{\xi}$ is the symmetrized version of ξ. Nevertheless, when $p = 1$ and $q = 2$, we can prove that optimal design for $\mathcal{L}_A$ as well as for $\mathcal{L}_Q$ is symmetric.

Lemma 3.4.2 *If $p = 1$ and $q = 2$, for any measure ξ with $|A_\xi| \neq 0$ we can find a symmetric measure ξ^* such that, for any β,*

$$\mathcal{L}_A(\boldsymbol{\beta}, \xi^*) \leq \mathcal{L}_A(\boldsymbol{\beta}, \xi) \quad \textit{and} \quad \mathcal{L}_Q(\boldsymbol{\beta}, \xi^*) \leq \mathcal{L}_Q(\boldsymbol{\beta}, \xi)$$

and therefore the A-optimal solution and the Q-optimal solution are symmetric.

Proof: Since $p = 1$ and $q = 2$, we have $z_2(x) = x^2$.

$$A_\xi = \begin{pmatrix} 1 & \xi_1 \\ \xi_1 & \xi_2 \end{pmatrix} \quad \text{and} \quad A_\xi^{-1} = \frac{1}{\xi_2 - \xi_1^2} \begin{pmatrix} \xi_2 & -\xi_1 \\ -\xi_1 & 1 \end{pmatrix}.$$

By Theorem 3.3.2, we may assume that ξ has 2 support points on $[-1/2, 1/2]$, say, $\xi = \alpha\delta_x + (1-\alpha)\delta_y$, for some $x, y \in [-1/2, 1/2]$. Then Lemma 3.3.1 tells us that

$$\xi_2 - \xi_1^2 = \alpha(1-\alpha)(x-y)^2.$$

We now define

$$\xi^* = (\delta_{\frac{x-y}{2}} + \delta_{\frac{y-x}{2}})/2.$$

Note that $(y-x)^2 = (\frac{x-y}{2} - \frac{y-x}{2})^2$ and $\xi_1^* = 0$ so we get

$$\xi_2^* = \frac{1}{4}(x-y)^2 = \frac{\xi_2 - \xi_1^2}{4\alpha(1-\alpha)}.$$

First we look at the "variance" part of $\mathcal{L}_A$(resp. $\mathcal{L}_Q$). that is. trace(A_ξ^{-1})(resp. trace$[A_\xi^{-1}A_0]$). This is equal to $(\xi_2^* + d)/(\xi_2^* - \xi_1^{*2})$. where $d = 1$ for A-optimality and $d = 1/12$ for Q-optimality. Therefore.

$$\begin{aligned} \text{"var"}(\xi^*) &= \frac{\xi_2^* + d}{\xi_2^* - \xi_1^{*2}} = 1 + \frac{d}{\xi_2^*} = \frac{\xi_2 - \xi_1^2 + 4d\alpha(1-\alpha)}{\xi_2 - \xi_1^2} \\ &= \frac{\xi_2 + d}{\xi_2 - \xi_1^2} - \left(\frac{\xi_1^2 + d(1-2\alpha)^2}{\xi_2 - \xi_1^2}\right) \le \frac{\xi_2 + d}{\xi_2 - \xi_1^2} = \text{"var"}(\xi). \end{aligned}$$

Let us look at the bias term in $\mathcal{L}_A$. that is. $C_\xi^T A_\xi^{-2} C_\xi$. By Lemma 3.4.1. for any measure supported by two points x and y we have. $A_\xi^{-1}C_\xi = (-xy. x+y)$. so

$$\begin{aligned} \text{"bias"}(\xi) - \text{"bias"}(\xi^*) &= (xy)^2 + (x+y)^2 - \left(\frac{x-y}{2}\right)^4 \\ &= \frac{1}{16}(x+y)^2[16 - (x^2 - 6xy + y^2)] \end{aligned}$$

The minimum of $16 - (x^2 - 6xy + y^2)$ over the range $-1/2 \le x. y \le 1/2$ occurs at $x = -1/2. y = 1/2$ and the minimum value is 14. Since this is always positive. and since var$(\xi^*) \le$ var(ξ). we get $\mathcal{L}_A(\beta. \xi^*) \le \mathcal{L}_A(\beta. \xi)$.

For Q-optimality, it is useful to use the formula $A_0^{-1}C_0 = (1/12, 0)$, then rewrite $\mathcal{L}_Q(\boldsymbol{\beta}, \xi)$ as

$$\begin{aligned}
\mathcal{L}_Q(\boldsymbol{\beta}, \xi) &= \operatorname{tr}(A_\xi^{-1}A_0) \\
&\quad + \boldsymbol{\beta}^T\left\{[A_\xi^{-1}C_\xi - A_0^{-1}C_0]^T A_0 [A_\xi^{-1}C_\xi - A_0^{-1}C_0] + B_0 - C_0^T A_0^{-1}C_0\right\}\boldsymbol{\beta} \\
&= \operatorname{tr}(A_\xi^{-1}A_0) + \boldsymbol{\beta}^T\left\{(xy + 1/12)^2 + (x+y)^2/12 + 1/180\right\}\boldsymbol{\beta}.
\end{aligned}$$

Looking only at the term involving $\boldsymbol{\beta}$ and subtracting we get

$$\begin{aligned}
\text{"bias"}(\xi) - \text{"bias"}(\xi^*) &= (xy + 1/12)^2 + (x+y)^2/12 - [1/12 - (x-y)^2/4]^2 \\
&= (1/16)(x+y)^2\left[2 - (x^2 - 6xy + y^2)\right]
\end{aligned}$$

The minimum of $2-(x^2-6xy+y^2)$ occurs at $x = -1/2, y = 1/2$ and the minimum value is 0. As for A-optimality, this shows that $\mathcal{L}_Q(\boldsymbol{\beta}, \xi^*) \le \mathcal{L}_Q(\boldsymbol{\beta}, \xi)$. ∎

3.5 Optimal designs

We will illustrate how we can obtain optimal designs. We have proved that the D-optimal measure is symmetric when $q - p = 1$. In general, this follows from Theorem 2.2.4. We have also proved in Lemma 3.4.2 that the A- and Q-optimal measures are symmetric when $p = 1$ and $q = 2$. For the other cases, since we are not able to prove the optimal design is symmetric, we restrict to the class of symmetric designs.

Applying Theorem 3.3.3 we see that the minimization problem depends on the design points x_i and their masses α_i, $i = 1, \ldots, [(p+1)/2]$. Writing these in vector form gives

$$\boldsymbol{x} := (x_1, \ldots, x_{[(p+1)/2]}), \text{ and } \boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{[(p+1)/2]}).$$

To emphasize that we are viewing the maximum loss as a function of $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ we write $l(\boldsymbol{\alpha}, \boldsymbol{x}) = \sup_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}, \sum_i \alpha_i \delta_{x_i})$. We formulate the minimization problem as follows.

$$\text{Minimize } l(\boldsymbol{\alpha}, \boldsymbol{x}): \text{ subject to } \begin{cases} 0 \le \alpha_i \le 1 \\ \sum \alpha_i = 1 \\ 0 \le x_i \le 1/2. \end{cases}$$

Since we have assumed that ξ is symmetric, we only consider the design points $x_i \in [0, 1/2]$; the remaining ones are obtained by reflection. The number of support points will vary according to number of moments involved in the loss function.

For small values of p and q, the minimization problem can be solved numerically. Under the D- and A-optimality criteria we obtain the minimax designs for approximately linear, approximately quadratic and approximately cubic regression models. All the D-optimal design points and their masses are in Table 3.6, Table 3.8, Table 3.10, Table 3.12 and Table 3.14.

3.5.1 Approximately linear model

D-optimality criterion

Example 1: $p = 1$ (linear) and $q = 2$ (quadratic).

The experimenter fits a linear regression although the true model might be quadratic. By Theorem 3.3.3 and Lemma 3.4.1, there exists a design with two support points $\pm x_1$ on $[-1/2, 1/2]$ with equal mass and so ignoring the constant term in the equation (3.4.2), the target function becomes

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) = \frac{1}{\xi_2}(\nu + 80\,\xi_4) = \frac{1}{x_1^2}\left(\nu + 80\,x_1^4\right) := l_D(x_1).$$

Consequently the optimal design is, with $x_1^* = (\nu/80)^{1/4}$,

$$\xi^* = \begin{cases} (\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}})/2 & \text{if } \nu \geq 5 \\ (\delta_{-x_1^*} + \delta_{x_1^*})/2 & \text{if } \nu < 5 \end{cases}$$

Example 2: $p = 1$ and $q = 3$ (cubic).

The true model might be a cubic polynomial, but we fit a linear regression, so the regressors are $\boldsymbol{z}_1(x) = (1, x)^T$ and $\boldsymbol{z}_2(x) = (x^2, x^3)^T$. The supremum of the loss function $\sup_{\mathcal{F}} \mathcal{L}(\boldsymbol{\beta}, \xi)$ in (3.2.1) is a function of the first 4 moments. So in Theorem 3.3.3, the number of moments, $2p+1$ must be greater or equal 4, which implies that p needs to be 2, and thus we obtain a discrete measure ξ with 3 support points. Since the measure is symmetric, ξ is of the form $\xi = (1-\alpha)\delta_0 + (\alpha/2)\delta_{x_1} + (\alpha/2)\delta_{-x_1}$. For this measure

$$A_\xi = \begin{pmatrix} 1 & 0 \\ 0 & \xi_2 \end{pmatrix}, \text{ and } \quad G_\xi = \begin{pmatrix} 80\,\xi_2^2 & 0 \\ 0 & 448\,(\xi_4^2/\xi_2) \end{pmatrix}.$$

The supremum of the loss function becomes

$$l_D(\alpha, x_1) = \frac{1}{\alpha\,x_1^2}[\nu + \max\{80\,\alpha^2\,x_1^4,\ 448\,\alpha\,x_1^6\}].$$

The minimizer (α^*, x_1^*) is obtained by minimizing $l_D(\alpha, x_1)$ over the domain $[0, 1] \times [0, 1/2]$.

Example 3 : $p = 1$ and $q = 4$ (quartic).

We fit a linear regression although the true model is a quartic polynomial, so the regressors are $\mathbf{z}_1(x) = (1, x)^T$ and $\mathbf{z}_2(x) = (x^2, x^3, x^4)^T$. The supremum of the loss function $\sup_{\mathcal{F}} \mathcal{L}(\boldsymbol{\beta}, \xi)$ involves only the first 5 moments. Theorem 3.3.3 says that an optimal measure ξ has 3 support points 0 and $\pm x_1$. This measure is of the form $\xi = (1-\alpha)\delta_0 + (\alpha/2)\delta_{x_1} + (\alpha/2)\delta_{-x_1}$ for some $x_1 \in [0, 1/2]$. The matrix A_ξ is same as in Example 2. The eigenvalues of the matrix G_ξ are $448\frac{\xi_4^2}{\xi_2}$ and $980\xi_2^3 - 10080\xi_2\xi_4 + 28224\xi_4^2$. Thus the maximum loss is

$$l_D(\alpha, x_1) = \frac{1}{\alpha\, x_1^2}[\nu + \max\{980\alpha^2 x_1^4 - 10080\alpha^2 x_1^6 + 28224\alpha^2 x_1^8,\ 448\,\alpha\, x_1^6\}].$$

A-optimality criterion

Example 4: $p = 1$ and $q = 2$.

Similar to the D-optimal case, there are two support points, $\pm x_1$, with point mass $1/2$. Ignoring the constant term, the loss function in equation (3.2.2) is, by Lemma 3.4.1,

$$\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi) &= \nu\,\mathrm{trace}(A_\xi^{-1}) + B_0^{-1/2} C_\xi^T A_\xi^{-2} C_\xi B_0^{-1/2} \\
&= \frac{\nu(\xi_2 + 1)}{\xi_2} + B_0^{-1/2}\mathbf{c}^T\mathbf{c}B_0^{-1/2} \quad (\text{since } A_\xi^{-1}C_\xi = -\mathbf{c}) \\
&= \frac{\nu(x_1^2 + 1)}{x_1^2} + 80\, x_1^4 := l_A(x_1) \quad (\text{since } \mathbf{c}^T = -x_1^2).
\end{aligned}$$

Thus an optimal design is $\xi^* = (\delta_{\min\{x_1^*, 1/2\}} + \delta_{-\min\{x_1^*, 1/2\}})/2$, with $x_1^* = (\nu/160)^{1/6}$. That is,

$$\xi^* = \begin{cases} (\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}})/2 & \text{if } \nu \geq 5/2 \\ (\delta_{-x_1^*} + \delta_{x_1^*})/2 & \text{if } \nu < 5/2. \end{cases}$$

Table 3.1: A-optimal design when $p = 1,\ q = 2$

ν	(design point : mass)	minimax loss
0	no optimal design exists	0.0000
1	$(\pm 0.4292 : 0.5)$	9.14
2.5	$(\pm 0.5 : 0.5)$	17.5
5	$(\pm 0.5 : 0.5)$	30
10	$(\pm 0.5 : 0.5)$	55
30	$(\pm 0.5 : 0.5)$	155

The optimal design points and point masses are in Table 3.1 below.

Example 5: $p = 1, q = 3$ and $p = 1, q = 4$

There is an optimal measure ξ with three support points $\pm x_1$ and zero, and the suprema of the loss functions are respectively

$$
\begin{aligned}
l_A(\alpha, x_1) &= \frac{\nu(\alpha x_1^2 + 1)}{\alpha x_1^2} + \max\{80\alpha^2 x_1^4, 448 x_1^4\}, \\
l_A(\alpha, y) &= \frac{\nu(\alpha x_1^2 + 1)}{\alpha x_1^2} + \max\{448 x_1^2, 980\alpha^2 x_1^4 - 10080\alpha^3 x_1^6 + 28224\alpha x_1^8\}.
\end{aligned}
$$

Again the minimizer (α^*, x_1^*) is obtained by minimizing the loss functions above over a square $[0, 1] \times [0, 1/2]$. We present the optimal design for some ν in Table 3.2.

3.5.2 Approximately quadratic model

D- and *A*-optimality criteria

Example 6 : $p = 2$ and $q = 3$.

The experimenter fits a quadratic model although the true model might be cubic.

Table 3.2: A-optimal design when $p = 1$, $q = 3$ or 4

ν	(design point : mass)	minimax loss
0	no optimal design exists	0.0000
1	$(\pm 0.2174 : 0.5)$, $(0 : 0)$	43.33
5	$(\pm 0.3250 : 0.5)$, $(0 : 0)$	99.66
10	$(\pm 0.3865 : 0.5)$, $(0 : 0)$	143.87
30	$(\pm 0.5 : 0.5)$, $(0 : 0)$	262.00

By Theorem 3.3.3 there is an optimal measure ξ with three support points on $[-1/2, 1/2]$ including zero, so $\xi = (1-\alpha)\delta_0 + (\alpha/2)\delta_{x_1} + (\alpha/2)\delta_{-x_1}$. Thus the loss function, by (3.4.2), becomes

$$
\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) &= \frac{1}{\xi_2(\xi_4 - \xi_2^2)}(\nu + 448\xi_6) \\
&= \frac{1}{\alpha^2(1-\alpha)x_1^6}\left(\nu + 448\alpha\, x_1^6\right) := l_D(\alpha, x_1).
\end{aligned}
$$

Hence the optimal design is

$$
\xi^* = (1-\alpha)\delta_0 + (\alpha/2)\delta_{\frac{1}{2}} + (\alpha/2)\delta_{-\frac{1}{2}} \quad \text{with} \quad \alpha = \frac{1}{28}\{7 - 3\nu + \sqrt{9\nu^2 + 70\nu + 49}\}.
$$

The maximum loss for A-optimality is, by Lemma 3.4.1,

$$
\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi) &= \nu\left[\frac{\xi_4 + 1}{\xi_4 - \xi_2^2} + \frac{1}{\xi_2}\right] + \lambda_{\max} H_\xi \\
&= \nu\left[\frac{1 + \alpha x_1^4 + (1-\alpha)x_1^2}{\alpha(1-\alpha)x_1^4}\right] + 448 x_1^4.
\end{aligned}
$$

The optimal designs for some ν are in Table 3.3.

Table 3.3: A-optimal design when $p = 2$, $q = 3$

ν	(design point : mass)	minimax loss
0	no optimal design exists	0.0000
1	(±0.5 : 0.2602), (0 : 0.4797)	101.8782
5	(±0.5 : 0.2602), (0 : 0.4797)	397.3909
10	(±0.5 : 0.2602), (0 : 0.4797)	766.7818
30	(±0.5 : 0.2602), (0 : 0.4797)	2244.345

3.5.3 Approximately cubic model

The last example in this section uses the regressors $z_1(x) = (1, x, x^2, x^3)^T$, and $z_2(x) = x^4$. That is, we fit a cubic regression whereas the true model might be quartic.

Example 7: $p = 3$ and $q = 4$.

We consider an optimal design measure of the form

$$\xi = (\alpha/2)(\delta_{x_1} + \delta_{-x_1}) + ((1-\alpha)/2)(\delta_{x_2} + \delta_{-x_2}).$$

The maximum loss functions are easily obtained from Lemma 3.4.1.

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) = \frac{1}{\xi_4^2\xi_2^2 - \xi_6\xi_2^3 + \xi_6\xi_4\xi_2 - \xi_4^3}[\nu + 2304\xi_8].$$

$$\sup_{\mathcal{F}} \mathcal{L}_A(\boldsymbol{\beta}, \xi) = \nu[\frac{1 + \xi_4}{\xi_4 - \xi_2^2} + \frac{\xi_6 + \xi_2}{\xi_6\xi_2 - \xi_4^2}] + 2304 * (x_1^4 + 2x_1^2x_2^2 + x_2^4 + x_1^4x_2^4).$$

The optimal design points and their masses are in Table 3.4.

Table 3.4: A-optimal design when $p = 3$, $q = 4$

ν	(design point : mass)	minimax loss
0	no optimal design exists	0.0000
1	$(\pm 0.2348 : 0.3341)$, $(\pm 0.5,0.1659)$	1414.387
5	$(\pm 0.2416 : 0.3311)$, $(\pm 0.5,0.1690)$	6203.047
10	$(\pm 0.2425 : 0.3307)$, $(\pm 0.5,0.1694)$	12186.19
30	$(\pm 0.2431 : 0.3304)$, $(\pm 0.5,0.1696)$	36117.73

3.6 Comparisons between polynomial models I and II

We present numerical values for optimal designs for the polynomial model I of Chapter 2 and the polynomial model II of this chapter. We first recall the notation $\boldsymbol{z}_1(x)$, $\boldsymbol{z}_2(x)$, and $\boldsymbol{u}(x)$ from the first section of Chapter 2:

$$\boldsymbol{z}_1^T(x) = (1, x, \ldots, x^p), \quad \boldsymbol{z}_2^T(x) = (x^{p+1}, \ldots, x^q),$$

$$\boldsymbol{u}(x) = \boldsymbol{z}_2(x) - [\int \boldsymbol{z}_1(x)\boldsymbol{z}_2^T(x)\,dx]^T[\int \boldsymbol{z}_1(x)\boldsymbol{z}_1^T(x)\,dx]^{-1}\boldsymbol{z}_1(x).$$

We notice that the contamination part $\boldsymbol{u}(x)$ of the model I is a polynomial containing all powers up to the qth. On the other hand, the contamination part $\boldsymbol{z}_2(x)$ of model II contains only terms of degree greater than p. The D-optimal design points and masses for the polynomial model I are in Table 3.5, Table 3.7, Table 3.9, Table 3.11 and Table 3.13. We also remember that the contaminating space $\mathcal{F}$ in this section is same as in 2.5.10 in Chapter 2.

***D*-optimal criterion**

Example 8 : $p = 1$ and $q = 2$.

The regressors are $\boldsymbol{z}_1(x)^T = (1, x)$, $\boldsymbol{z}_2^T(x) = x^2$.

It is easy to calculate $\boldsymbol{u}(x) = (x^2 - 1/12)$, $B_0 = \int \boldsymbol{u}(x)\boldsymbol{u}^T(x)\, dx = 1/180$, and $G_\xi = \frac{5}{4}(12\xi_2 - 1)^2$. The supremum of the loss function is

$$\sup_{\mathcal{F}} \mathcal{L}_D = \frac{1}{\xi_2}[\nu + \frac{5}{4}(12\xi_2 - 1)^2].$$

Consequently the optimal design is, $\xi^* = (\delta_{\min\{x_1^*, 1/2\}} + \delta_{-\min\{x_1^*, 1/2\}})/2$, with $x_1^* = (4\nu + 5/720)^{1/4}$, that is

$$\xi^* = \begin{cases} (\delta_{-\frac{1}{2}} + \delta_{\frac{1}{2}})/2 & \text{if } \nu \geq 10 \\ (\delta_{-x_1^*} + \delta_{x_1^*})/2 & \text{if } \nu < 10 \end{cases}$$

Example 9: $p = 1$ and $q = 3$.

The regressors are $\boldsymbol{z}_1(x)^T = (1, x)$, $\boldsymbol{z}_2^T(x) = (x^2, x^3)$. The contamination part is $\boldsymbol{u}^T(x) = (x^2 - 1/12, x^3 - 3x/20)$. The matrices are calculated as follows

$$B_0 = \int \mathrm{diag}(1/180, 1/2800), \quad G_\xi = \mathrm{diag}(\frac{5}{4}(12\xi_2 - 1)^2, \frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2),$$

hence we obtain the maximum of the loss function

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) = \frac{1}{\xi_2}[\nu + \max\{\frac{5}{4}(12\xi_2 - 1)^2, \frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2\}].$$

There are 3 support point including zero, and the optimal design is of the form $\xi^* = (1 - \alpha^*)\delta_0 + (\alpha^*/2)(\delta_{x_1^*} + \delta_{-x_1^*})$, where α^* and x_1^* are the minimizer of the maximum loss.

Example 10 : $p = 1$ and $q = 4$.

The regressors are $\boldsymbol{z}_1(x)^T = (1, x)$, $\boldsymbol{z}_2^T(x) = (x^2, x^3, x^4)$.

The contamination part consists of three polynomials.

$$\boldsymbol{u}^T(x) = (x^2 - 1/12, x^3 - 3x/20, x^4 - 1/80).$$

The matrix $B_0 = \int \boldsymbol{u}(x)\boldsymbol{u}^T(x)\,dx$ is a 3 by 3 matrix

$$B_0 = \begin{pmatrix} 1/180 & 0 & 1/840 \\ 0 & 1/2800 & 0 \\ 1/840 & 0 & 1/3600 \end{pmatrix}.$$

Two eigenvalues of the matrix G_ξ are

$$44100\xi_4^2 + \frac{945}{2}\xi_4 + 2205\xi_2^2 - \frac{525}{4}\xi_2 - 18900\xi_4\xi_2 + \frac{161}{64}.$$

and

$$\frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2.$$

Again there are three support points and so the optimal design is of the form $\xi^* = (1 - \alpha^*)\delta_0 + (\alpha^*/2)(\delta_{x_1^*} + \delta_{-x_1^*})$, where α^* and x_1^* are the minimizer of the maximum loss.

Example 11 : $p = 2$ and $q = 3$.

The regressors are $\boldsymbol{z}_1(x)^T = (1, x, x^2)$, $\boldsymbol{z}_2^T(x) = x^3$.

The contamination is a polynomial of degree three, $\boldsymbol{u}(x) = x^3 - 3x/20$. The constant matrix is $B_0 = 1/2800$, and the bias term has a single eigenvalue, $G_\xi = \frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2$. The maximum of the loss function is of the form

$$\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) = \frac{1}{\xi_2(\xi_4 - \xi_2^2)}[\nu + \frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2].$$

The optimal design is of the form $\xi^* = (1-\alpha^*)\delta_0 + (\alpha^*/2)(\delta_{x_1^*} + \delta_{-x_1^*})$, where α^* and x_1^* are the minimizer of the maximum loss.

Example 12 : $p = 3$ and $q = 4$.

The regressors are $\boldsymbol{z}_1(x)^T = (1, x, x^2, x^3)$, $\boldsymbol{z}_2^T(x) = x^4$.

The contamination is a polynomial of degree four, $\boldsymbol{u}(x) = x^4 - 3x^2/14 + 3/560$.

The constant matrix is $B_0 = 1/44100$. The maximum of the loss function is

$$\begin{aligned}
\sup_{\mathcal{F}} \mathcal{L}_D(\boldsymbol{\beta}, \xi) &= \frac{1}{\xi_2\xi_4\xi_6 - \xi_4^3 - \xi_2^3\xi_6 + \xi_2^2\xi_4^2}[\nu + (-9/64)(313600\xi_4^3 + 17760\xi_4^2 \\
&+ 9\xi_4 - 720\xi_2\xi_4 - 627200\xi_2\xi_4\xi_6 + 134400\xi_2^2\xi_6 - 17760\xi_2^2\xi_4 \\
&- 9\xi_2^2 + 720\xi_2^3 + 313600\xi_6^2 - 134400\xi_4\xi_6)/(\xi_2^2 - \xi_4)].
\end{aligned}$$

We provide the optimal designs for the polynomial model I and the polynomial model II in the following tables.

Table 3.5: D-optimal design for model I when $p = 1$, $q = 2$

ν	(design point : mass)	minimax loss
0.0	($\pm$0.2887 : 0.5)	0.0000
0.1	($\pm$0.2943 : 0.5)	1.1769
0.5	($\pm$0.3140 : 0.5)	5.4965
1.0	($\pm$0.3344 : 0.5)	10.2492
5	($\pm$0.4317 : 0.5)	37.0820
10	($\pm$0.5 : 0.5)	60
30	($\pm$0.5 : 0.5)	140

Table 3.6: D-optimal design for model II when $p = 1$, $q = 2$

ν	(design point : mass)	minimax loss
0.0	no optimal design exists	0.0000
0.1	($\pm$0.1880 : 0.5)	5.6569
0.5	($\pm$0.2812 : 0.5)	12.6491
1.0	($\pm$0.3344 : 0.5)	17.8885
5	($\pm$0.5 : 0.5)	40
10	($\pm$0.5 : 0.5)	60
30	($\pm$0.5 : 0.5)	140

Table 3.7: D-optimal design for model I when $p = 1$, $q = 3$

ν	(design point : mass)	minimax loss
0.0	(±0.3873 : 0.2778), (0.0000 : 0.4444)	0.0000
0.1	(±0.3844 : 0.2930), (0.0000 : 0.4140)	1.1769
0.5	(±0.3726 : 0.3551), (0.0000 : 0.2897)	5.4965
1.0	(±0.3598 : 0.4318), (0.0000 : 0.1364)	10.2492
2.0	(±0.3747 : 0.4786), (0.0000 : 0.0428)	18.3736
2.8	(±0.3933 : 0.4849), (0.0000 : 0.0302)	24.0000
5.0	(±0.4317 : 0.5), (0.0000 : 0.0000)	37.0820
10	(±0.4755 : 0.5), (0.0000 : 0.0000)	60.4531
30	(±0.4916 : 0.5), (0.0000 : 0.0000)	147.6658

Table 3.8: D-optimal design for model II when $p = 1$, $q = 3$

ν	(design point : mass)	minimax loss
0.0	no optimal design exists	0.0000
0.1	(±0.2656 : 0.2506), (0.0000 : 0.4988)	5.6569
0.5	(±0.3247 : 0.3748), (0.0000 : 0.2504)	12.6491
1.0	(±0.3344 : 0.5), (0.0000 : 0.0000)	17.8885
5.0	(±0.4226 : 0.5), (0.0000 : 0.0000)	42.2857
10	(±0.4727 : 0.5), (0.0000 : 0.0000)	67.1213
30	(±0.5000 : 0.5), (0.0000 : 0.0000)	148.000

Table 3.9: D-optimal design for model I when $p = 1$, $q = 4$

ν	(design point : mass)	minimax loss
0.0	($\pm$0.3873 : 0.2778), (0.0000 : 0.4444)	0.0000
0.1	($\pm$0.3904 : 0.2841), (0.0000 : 0.4318)	1.1769
0.5	($\pm$0.3999 : 0.3082), (0.0000 : 0.3836)	5.4965
1.0	($\pm$0.4079 : 0.3360), (0.0000 : 0.3280)	10.2492
2.0	($\pm$0.4176 : 0.3852), (0.0000 : 0.2296)	18.3736
2.8	($\pm$0.4226 : 0.4200), (0.0000 : 0.1600)	24.0000
5.0	($\pm$0.4307 : 0.5), (0.0000 : 0.0000)	37.0828
10	($\pm$0.4411 : 0.5), (0.0000 : 0.0000)	63.3668
30	($\pm$0.4652 : 0.5), (0.0000 : 0.0000)	160.2507

Table 3.10: D-optimal design for model II when $p = 1$, $q = 4$

ν	(design point : mass)	minimax loss
0.0	no optimal design exists	0.0000
0.1	($\pm$0.3432 : 0.1433), (0.0000 : 0.7134)	9.1803
0.5	($\pm$0.3854 : 0.3153), (0.0000 : 0.3694)	15.2197
1.0	($\pm$0.4073 : 0.4395), (0.0000 : 0.1211)	19.1857
5.0	($\pm$0.4226 : 0.5), (0.0000 : 0.0000)	42.2857
10	($\pm$0.4412 : 0.5), (0.0000 : 0.0000)	68.3665
30	($\pm$0.4679 : 0.5), (0.0000 : 0.0000)	164.609

Table 3.11: D-optimal design for model I when $p = 2$, $q = 3$

ν	(design point : mass)	minimax loss
0.0	(±0.3873 : 0.2724), (0.0000 : 0.4552)	0.0000
0.1	(±0.3918 : 0.3314), (0.0000 : 0.3371)	193.1609
0.5	(±0.4080 : 0.3251), (0.0000 : 0.3498)	853.4112
1.0	(±0.4254 : 0.3191), (0.0000 : 0.3618)	1500.122
5.0	(±0.5000 : 0.3032), (0.0000 : 0.3936)	4087.935
10	(±0.5000 : 0.3144), (0.0000 : 0.3711)	6279.972
30	(±0.5000 : 0.3257), (0.0000 : 0.3487)	14953.00

Table 3.12: D-optimal design for model II when $p = 2$, $q = 3$

ν	(design point : mass)	minimax loss
0.0	(±0.2500 : 0.2500), (0.0000 : 0.5000)	1792.000
0.1	(±0.5000 : 0.2534), (0.0000 : 0.4932)	1842.854
0.5	(±0.5000 : 0.2641), (0.0000 : 0.4719)	2040.824
1.0	(±0.5000 : 0.2735), (0.0000 : 0.4531)	2280.155
5.0	(±0.5000 : 0.3032), (0.0000 : 0.3936)	4087.935
10	(±0.5000 : 0.3144), (0.0000 : 0.3711)	6279.972
30	(±0.5000 : 0.3257), (0.0000 : 0.3487)	14953.00

Table 3.13: D-optimal design for model I when $p = 3$, $q = 4$

ν	(design point : mass)	minimax loss
0.0	(±0.1700 : 0.3551), (± 0.4306; 0.1449)	0.00
0.1	(±0.1724 : 0.2525), (± 0.4339 : 0.2475)	475078
0.5	(±0.1810 : 0.2613), (± 0.4462 : 0.2387)	2003326
1.0	(±0.1903 : 0.2706), (± 0.4609 : 0.2294)	3338269
5.0	(±0.2166 : 0.2795), (± 0.5 : 0.2205)	7523991
10	(±0.2191 : 0.2689), (± 0.5 : 0.2311)	11602484
30	(±0.2217 : 0.2578), (± 0.5 : 0.2422)	27685078

Table 3.14: D-optimal design for model II when $p = 3$, $q = 4$

ν	(design point : mass)	minimax loss
0.0	(±0.2225 : 0.3328), (± 0.5 : 0.1672)	3047007
0.1	(±0.2226 : 0.3294), (± 0.5 : 0.1706)	3146930
0.5	(±0.2228 : 0.3188), (± 0.5 : 0.1812)	3531119
1.0	(±0.2229 : 0.3097), (± 0.5 : 0.1903)	3989385
5.0	(±0.2233 : 0.2809), (± 0.5 : 0.2191)	7385404
10	(±0.2234 : 0.2697), (± 0.5 : 0.2303)	11464446
30	(±0.2235 : 0.2582), (± 0.5 : 0.2418)	27547100

Chapter 4

Optimal designs for approximately polynomial regression

Huber (1975) and Wiens (1990, 1992) obtained minimax densities for an approximately linear regression model with the contamination space $\mathcal{F}_2$. It is very natural to extend the minimax density approach to an approximately quadratic regression. It turns out that the minimax density for higher degree polynomials is not tractable. In this chapter we restrict to a class of densities that is tractable, practical and can be generalized to higher degree polynomial models. We present optimal designs for approximately quadratic regression and approximately cubic regression for this restricted class. These optimal designs are easily generalized to multiple regression as well. We not only describe the optimal designs for an approximately quadratic bivariate regression with interaction terms but we also explain how the densities might be implemented in practice.

4.1 Review for approximately linear regression and motivation

As we mentioned in Chapter 1, Wiens (1990, 1992) obtains minimax designs for approximately linear regression. Our objective in this chapter is to extend his idea to higher degree polynomial models. We present designs for approximately quadratic and cubic regression models as well as bivariate models with interaction terms.

We use the perturbed model and the contaminating space $\mathcal{F}_2$,

$$\begin{aligned} E(y \mid \boldsymbol{x}) &= \boldsymbol{z}^T(\boldsymbol{x})\boldsymbol{\theta} + f(\boldsymbol{x}), \qquad (4.1.1) \\ f \in \mathcal{F}_2 &= \{f : \|f\|_2^2 = \int_S f(\boldsymbol{x})^2\, d\boldsymbol{x} \leq \eta^2, \int_S \boldsymbol{z}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} = \boldsymbol{0}\}. \end{aligned}$$

The regressor $\boldsymbol{z} \in I\!R^p$ is given function of $\boldsymbol{x}$, where $\boldsymbol{x}$ varies on a design space $S \subseteq I\!R^q$. Wiens calculates the loss functions explicitly. With

$$b(f,\xi) = \int_S \boldsymbol{z}(\boldsymbol{x}) f(\boldsymbol{x})\, d\xi(\boldsymbol{x})$$

and A_ξ as defined in (1.2.2) and A_0 corresponding to the Lebesgue measure, Wiens finds that

$$\begin{aligned} \mathcal{L}_Q(f,\xi) &= \text{IMSE}(\hat{y}(\boldsymbol{x})) \\ &= \left(\frac{\sigma^2}{n}\right) \text{tr}[A_\xi^{-1} A_0] + b^T(f,\xi) A_\xi^{-1} A_0 A_\xi^{-1} b(f,\xi) + \int f^2(\boldsymbol{x})\, d\boldsymbol{x}, \\ \mathcal{L}_D(f,\xi) &= \det(\text{MSE}(f,\xi)) = \left(\frac{\sigma^2}{n}\right)^p \frac{1}{|A_\xi|}(1 + \frac{n}{\sigma^2} b^T(f,\xi) A_\xi^{-1} b(f,\xi)), \\ \mathcal{L}_A(f,\xi) &= \text{tr}(\text{MSE}(f,\xi)) = \left(\frac{\sigma^2}{n}\right) \text{tr} A_\xi^{-1} + b^T(f,\xi) A_\xi^{-2} b(f,\xi). \end{aligned}$$

Wiens applies the minimax approach, that is, looks for $\xi^* \in \Xi$ such that $\sup_{\mathcal{F}_2}\mathcal{L}(f,\xi)$ is minimized by ξ^*. It is necessary that ξ be absolutely continuous for $\sup_{\mathcal{F}_2}\mathcal{L}(f,\xi)$ to be finite (see Lemma 1 in Wiens (1992)) and so without loss of generality we may restrict our attention to absolutely continuous measures ξ. Let $m(\boldsymbol{x}) = \xi'(\boldsymbol{x})$ be the density with respect to Lebesgue measure. For the maximization part, it is enough to look for f in a finite dimensional space in $\mathbb{R}^p$. To state the result precisely, we need a few definitions.

$$\begin{aligned} K_\xi &= \int \boldsymbol{z}(\boldsymbol{x})\boldsymbol{z}^T(\boldsymbol{x})m^2(\boldsymbol{x})\,d\boldsymbol{x}, \\ H_\xi &= A_\xi A_0^{-1}A_\xi, \quad G_\xi = K_\xi - H_\xi, \\ r(\boldsymbol{x}) &= \eta\, G_\xi^{-1/2}(m(\boldsymbol{x})I - A_\xi A_0^{-1})\boldsymbol{z}(\boldsymbol{x}). \end{aligned}$$

Wiens establishes an important result in Theorem 1 of Wiens (1992):

$$\sup_{\mathcal{F}_2}\mathcal{L}(f,\xi) = \max_{\|\beta\|=1}\mathcal{L}(h_\beta,\xi) \quad \text{for all} \quad h_\beta = r^T(\boldsymbol{x})\beta,\ \|\beta\| = 1.$$

By this result, we are maximizing a continuous function over the compact set $\{\beta : \|\beta\| = 1\}$ and hence the maximum is attained. This result leads to

$$\sup_{\mathcal{F}_2}\mathcal{L}_Q(f,\xi) = \mathcal{L}_Q(h_\beta,\xi) = \eta^2\left[\nu\,\mathrm{tr}(A_\xi^{-1}A_0) + \lambda_{\max}K_\xi H_\xi^{-1}\right]. \quad (4.1.2)$$

$$\sup_{\mathcal{F}_2}\mathcal{L}_D(f,\xi) = \mathcal{L}_D(h_\beta,\xi) = \left(\frac{\sigma^2}{n}\right)^{p-1}\frac{\eta^2}{|A_\xi|}\left[\nu + \lambda_{\max}(G_\xi^{1/2}A_\xi^{-1}G_\xi^{1/2})\right]. \quad (4.1.3)$$

$$\sup_{\mathcal{F}_2}\mathcal{L}_A(f,\xi) = \mathcal{L}_A(h_\beta,\xi) = \eta^2\left[\nu\,\mathrm{tr}(A_\xi^{-1}) + \lambda_{\max}(G_\xi^{1/2}A_\xi^{-2}G_\xi^{1/2})\right]. \quad (4.1.4)$$

For the minimization part, Wiens fits a plane, $\boldsymbol{z}(\boldsymbol{x}) = (1,\boldsymbol{x}^T)^T$, where $\boldsymbol{x} = (x_1,\ldots,x_q)$, and the design space $S \subseteq \mathbb{R}^q$ is a sphere of unit volume with radius

ω. He restricts to densities $m(\boldsymbol{x})$ that are symmetric in each variable and each of the variables is exchangeable. There are two non-zero eigenvalues in each matrix

$$G_\xi^{1/2} A_\xi^{-1} G_\xi^{1/2}, \quad K_\xi H_\xi^{-1}, \quad G_\xi^{1/2} A_\xi^{-2} G_\xi^{1/2},$$

which are of the form

$$\lambda_1(m(\boldsymbol{x})) = \int m^2(\boldsymbol{x})\, d\boldsymbol{x}, \quad \lambda_2(m(\boldsymbol{x})) = \frac{\xi_0}{\xi_2^2} \int x_1\, m^2(\boldsymbol{x})\, d\boldsymbol{x},$$

where

$$\xi_0 = \int x_1^2\, d\boldsymbol{x} = \frac{\omega^2}{q+2}, \quad \xi_2 = \int x_1^2 m(\boldsymbol{x})\, d\boldsymbol{x}. \tag{4.1.5}$$

To proceed to the minimization, we need to know which eigenvalue is larger. Let us assume that $\lambda_1 \geq \lambda_2$. First hold ξ_2 fixed, and minimize $\int m^2(\boldsymbol{x})\, d\boldsymbol{x}$ over all continuous densities. Let $m(\boldsymbol{x}; \xi_2)$ denote the minimizer. Next, minimize the maximum loss function

$$\sup_{\mathcal{F}_2} \mathcal{L}_Q(f, \xi) = \eta^2 \left[\int m^2(\boldsymbol{x}; \xi_2)\, d\boldsymbol{x} + \nu(1 + q\frac{\xi_0}{\xi_2}) \right].$$

Let $m(\boldsymbol{x}; \xi_2^*)$ be the minimizer. Finally verify that indeed the first eigenvalue is larger at the minimizer, that is, $\lambda_1(m(\boldsymbol{x}; \xi_2^*)) \geq \lambda_2(m(\boldsymbol{x}; \xi_2^*))$. For Q- and D-optimal designs, the first eigenvalue at the minimizer is larger. For A-optimal design the second eigenvalue is larger at the minimizer. This leads to the optimal density functions, for Q- and D-optimality cases,

$$m(\boldsymbol{x}, \xi_2^*) = a(\|\boldsymbol{x}\|^2 - b)^+,$$

where a and b are determined by the equations in (4.1.5). For A-optimality case

$$m(\boldsymbol{x}, \xi_2^*) = a(\|\boldsymbol{x}\|^2 - b)^+/\|\boldsymbol{x}\|^2.$$

For higher degree polynomial regression models, not only are there more than two eigenvalues, but also the eigenvalues are more complicated than the ones in the linear model and so obtaining the optimal density function is not so easy. We will illustrate these obstacles by fitting a quadratic polynomial under the Q-optimality criterion. In the perturbed model (4.1.1) with the contaminating space $\mathcal{F}_2$, take the regressor, $\boldsymbol{z}(x) = (1, x, x^2)^T$, where $-1/2 \le x \le 1/2$. Ignoring the term η^2, the maximum IMSE is

$$\sup_{\mathcal{F}_2} \mathcal{L}_Q(f, \xi) = \nu \operatorname{tr}\left(A_\xi^{-1} A_0\right) + \lambda_{\max}(K_\xi H_\xi^{-1}).$$

Letting $\xi_i = \int x^i m(x)\, dx$, $k_i = \int x^i m^2(x)\, dx$, we calculate matrices

$$A_0 = \begin{pmatrix} 1 & 0 & 1/12 \\ 0 & 1/12 & 0 \\ 1/12 & 0 & 1/80 \end{pmatrix}, \quad A_\xi = \begin{pmatrix} 1 & 0 & \xi_2 \\ 0 & \xi_2 & 0 \\ \xi_2 & 0 & \xi_4 \end{pmatrix},$$

$$K_\xi = \begin{pmatrix} k_0 & 0 & k_2 \\ 0 & k_2 & 0 \\ k_2 & 0 & k_4 \end{pmatrix}, \quad H_\xi = \begin{pmatrix} h_0 & 0 & h_1 \\ 0 & h_2 & 0 \\ h_1 & 0 & h_3 \end{pmatrix},$$

where the elements of matrix H_ξ are functions of certain moments:

$$\begin{aligned} h_0 &= 9/4 - 30\xi_2 - 180\xi_2^2, \\ h_1 &= 9\xi_2/4 - 15\xi_4 - 15\xi_2^2 + 180\xi_2\xi_4, \\ h_2 &= 12\xi_2^2, \; h_3 = 9\xi_2^2/4 - 30\xi_2\xi_4 + 180\xi_4^2. \end{aligned}$$

There are three non-zero eigenvalues in the matrix $K_\xi H_\xi^{-1}$:

$$\begin{aligned}\lambda_1 &= k_2/h_2,\\ \lambda_2 &= (\mathcal{B}+\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}})/2\mathcal{A},\\ \lambda_3 &= (\mathcal{B}-\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}})/2\mathcal{A},\end{aligned}$$

where $\mathcal{A}, \mathcal{B}, \mathcal{C}$ depend on h_i and k_i.

$$\begin{aligned}\mathcal{A} &= h_0h_3 - h_1^2 = \det(H_\xi)/h_2,\\ \mathcal{B} &= k_0h_3 + k_4h_0 - 2k_2h_1,\\ \mathcal{C} &= k_0k_4 - k_2^2 = \det(K_\xi)/k_2.\end{aligned}$$

We are now looking for the density m that minimizes

$$\max\{\lambda_1(m), \lambda_2(m)\} + \nu\,\mathrm{tr}(A_\xi^{-1}A_0).$$

We first assume $\lambda_1 > \lambda_2$, when evaluated at the minimizing density. Fix ξ_2, ξ_4: this fixes the matrices A_ξ and H_ξ so that the problem is to:

1. Minimize $\lambda_1(m) = \int x^2m^2(x; \xi_2, \xi_4)\,dx$ for fixed ξ_2, and ξ_4.

2. Vary the parameters of the minimizing m so as to minimize $\lambda_1(m) + \nu\mathrm{tr}(A_\xi^{-1}A_0)$.

3. Verify that for the final m^*, we have indeed $\lambda_1(m^*) \geq \lambda_2(m^*)$.

From the first step, minimizing k_2 for fixed ξ_2, ξ_4 gives

$$m(x) = (ax^2 - b/x^2 + c)^+, \qquad b > 0.$$

This m can never place mass near $0, \pm\frac{1}{2}$ only as it should when $\nu \to \infty$. It turns out that

$$\lambda_2(m(x; a^*, b^*, c^*)) > \lambda_1(m(x; a^*, b^*, c^*)),$$

where $a^* = a(\xi_2^*, \xi_4^*)$, $b^* = b(\xi_2^*, \xi_4^*)$, $c^* = c(\xi_2^*, \xi_4^*)$ are the minimizers at step 2 above. Now then minimizing λ_2 for fixed ξ_2, ξ_4 gives the minimax density of the form, for positive a, and real coefficients b, c, d, e, and f,

$$m(x) = a\left(\frac{1 + bx^2 + cx^4}{d + ex^2 + fx^4}\right)^+. \tag{4.1.6}$$

But the density above is very complicated. As a remedy, we restrict to densities m which are reasonably tractable and have the correct limits. In this restricted class it is not meaningful to pick a largest eigenvalue, and now we write the loss function to be minimized as

$$l_Q(m) = \max\{\lambda_1(m), \lambda_2(m)\} + \nu \mathrm{tr}(A_\xi^{-1} A_0).$$

We propose densities, with coefficients $a \geq 0$, $-\infty < b < \infty$, $0 \leq s \leq t \leq 1/2$,

$$m(x) = a\{(x^2 - s^2)(x^2 - t^2) + b\}^+ \tag{4.1.7}$$

To carry out the minimization process using the "nlmin" function in Splus, we need to take transformations on t, $t' = (t - s)/(1 - 2s)$, and so that the domain is a rectangle shape, $0 \leq s \leq \frac{1}{2}$ and $0 \leq t' \leq \frac{1}{2}$. From the fact $\int m(x)\, dx = 1$, the constant a is written in terms of s, t, and b. Thus our optimization problem becomes to

$$\text{Minimize } l_Q\{s, s + t'(1 - 2s), b\}: \text{ subject to } \begin{cases} 0 \leq s \leq \frac{1}{2} \\ 0 \leq t' \leq \frac{1}{2} \\ |b| < +\infty. \end{cases}$$

In the optimization problem, it is complicated and time consuming to obtain the integrals ξ_i and k_i. We have applied Simpson's rule (see p.266 of Thisted (1988)) to approximate these integrations. We now provide explanations why we choose to restrict to the densities in (4.1.7). First, the densities of this form are not only easily constructed but also can be generalized to the higher polynomial models. For a general polynomial regressors, $\boldsymbol{z}(x) = (1, x, x^2, \ldots, x^p)^T$, take the densities of the form

$$a\{(x^2 - s_1^2)(x^2 - s_2^2)\cdots(x^2 - s_p^2) + b\}^+.$$

Transformation on the domain $0 \leq s_1 < s_2 < \cdots < s_p < 1/2$ can be easily done via

$$s_1' = s_1, s_2' = (s_2 - s_1)/(1 - 2s_1), \ldots, s_p' = (s_p - s_{p-1})/(1 - 2s_{p-1}) \quad (4.1.8)$$

so then $0 \leq s_i' \leq \frac{1}{2}$ for all $i = 1, \ldots, p$. We verify that this transformation works. Let

$$S' = \{\boldsymbol{s}' \in I\!R^p \ : \ 0 < s_i' < 1/2 : i = 1, \ldots, p\}$$

and

$$S = \{\boldsymbol{s} \in I\!R^p \ : \ 0 < s_1 < s_2 < \cdots < s_p < 1/2\}.$$

Define $\phi : S \longrightarrow S'$ by

$$(\phi(\boldsymbol{s}))_j = (s_j - s_{j-1})/(1 - 2s_{j-1})$$

for $j = 1, \ldots, p$, where $s_0 = 0$. We prove that ϕ is a one-to-one mapping of S onto S' : Let $\boldsymbol{s} \in S$, we show that $\phi(\boldsymbol{s}) \in S'$. The condition $s_{j-1} < s_j < 1/2$

implies that $0 < (s_j - s_{j-1})/(1 - 2s_{j-1}) < 1/2$, that is, $0 < (\varphi(\boldsymbol{s}))_j < 1/2$, and thus, $\phi(\boldsymbol{s}) \in S'$. Now for any $\boldsymbol{s}' \in S'$ define coordinates inductively $s_1 = s'_1$: $s_j = s_{j-1} + s'_{j+1}(1 - 2s_{j-1}), j = 2, \ldots, p$. By induction $0 < s_{j-1} < 1/2$. Thus $s_{j-1} < s_j < s_{j-1} + (1/2)(1 - 2s_{j-1})$. This shows that $\boldsymbol{s} = (s_1, \ldots, s_p) \in S$ and clearly $\phi(\boldsymbol{s}) = \boldsymbol{s}'$. The same inductive argument shows that if $\phi(\boldsymbol{s}) = \phi(\boldsymbol{t})$, then $\boldsymbol{s} = \boldsymbol{t}$ so ϕ is one-to-one.

Also, for $\nu = 0$, the minimizing density $m^*(x)$ is approximately uniform. As ν gets larger, the mass is concentrated around 0 and $\pm\frac{1}{2}$. This means that the densities have the correct limits because (i) for $\nu = 0$, only the bias term is minimized and hence the optimal design is uniform, (ii) for large ν, the variance swamps the bias term, and so the optimal density is as in the classical Q-optimal case, that is, all masses are at 0 and the boundary points.

We sum up the algorithm to obtain an optimal density for an approximately polynomial regression, $\boldsymbol{z}(x) = (1, x, \ldots, x^p)^T$.

1. Obtain the matrices A_0, A_ξ, K_ξ and set $H_\xi = A_\xi A_0^{-1} A_\xi$.

2. Obtain the $(p+1)$ eigenvalues of the matrix $K_\xi H_\xi^{-1}$.

3. make the transformation (4.1.8).

4. Obtain the coefficient a in (4.1.7) as a function of $\boldsymbol{s}'$ and b as a result of $\int m(x)\,dx = 1$.

5. Write out the loss function l in terms of $\boldsymbol{s}'$ and b.

$$l_Q(\boldsymbol{s}', b) = \max_{1 \le l \le p+1}\{\lambda_l(\boldsymbol{s}', b)\} + \nu \operatorname{tr}[A^{-1}(\boldsymbol{s}', b)A_0].$$

Using Simpson's rule, calculate the ξ_i's and k_i's and apply the "nlmin" function in Splus to minimize $l(s', b)$ over p dimensional rectangle and $-\infty < b < \infty$ for all ν.

In the following section we provide optimal density functions for some ν's under the three optimality criteria.

4.2 Numerical results for quadratic and cubic regression models

4.2.1 Design densities for approximately quadratic regression

Q-optimal design

As illustrated in the first section of this chapter, Q-optimality leads to the following type of density

$$m(x; a, s, t, b) = a\{(x^2 - s^2)(x^2 - t^2) + b\}^+ \tag{4.2.1}$$

to minimize the loss function $l_Q(m(x; a, s, t, b))$.

D-optimal design

There are three eigenvalues in the matrix $G_\xi^{1/2} A_\xi^{-1} G_\xi^{1/2}$, and they are

$$\begin{aligned}
\lambda_1 &= \frac{k_2 - h_2}{\xi_2}, \\
\lambda_2 &= \frac{\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}, \\
\lambda_3 &= \frac{\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}.
\end{aligned}$$

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are

$$\begin{aligned}
\mathcal{A} &= \xi_4 - \xi_2^2, \\
\mathcal{B} &= \xi_4(k_0 - h_0) + k_4 - h_3 - 2\xi_2(k_2 - h_1), \\
\mathcal{C} &= (k_0 - h_0)(k_4 - h_3) - (k_2 - h_1)^2.
\end{aligned}$$

$\mathcal{A}$-optimal design

The three eigenvalues of the matrix $G_\xi^{1/2} A_\xi^{-2} G_\xi^{1/2}$ are

$$\begin{aligned}
\lambda_1 &= \frac{k_2 - h_2}{\xi_2^2}, \\
\lambda_2 &= \frac{\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}, \\
\lambda_3 &= \frac{\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}},
\end{aligned}$$

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are

$$\begin{aligned}
\mathcal{A} &= (1 + \xi_2^2)(\xi_2^2 + \xi_4^2) + \xi_2^2(1 + \xi_4)^2, \\
\mathcal{B} &= (\xi_2^2 + \xi_4^2)(k_0 - h_0) + (1 + \xi_2^2)(k_4 - h_3) - 2\xi_2(1 + \xi_4)(k_2 - h_1), \\
\mathcal{C} &= (k_0 - h_0)(k_4 - h_3) - (k_2 - h_1)^2.
\end{aligned}$$

We present constants for densities for a few values of ν in Figure 4.1 and Table 4.1, Table 4.2 and Table 4.3.

Figure 4.1: Q-optimal design densities for an approximately quadratic regression model with contaminating space $\mathcal{F}_2$.

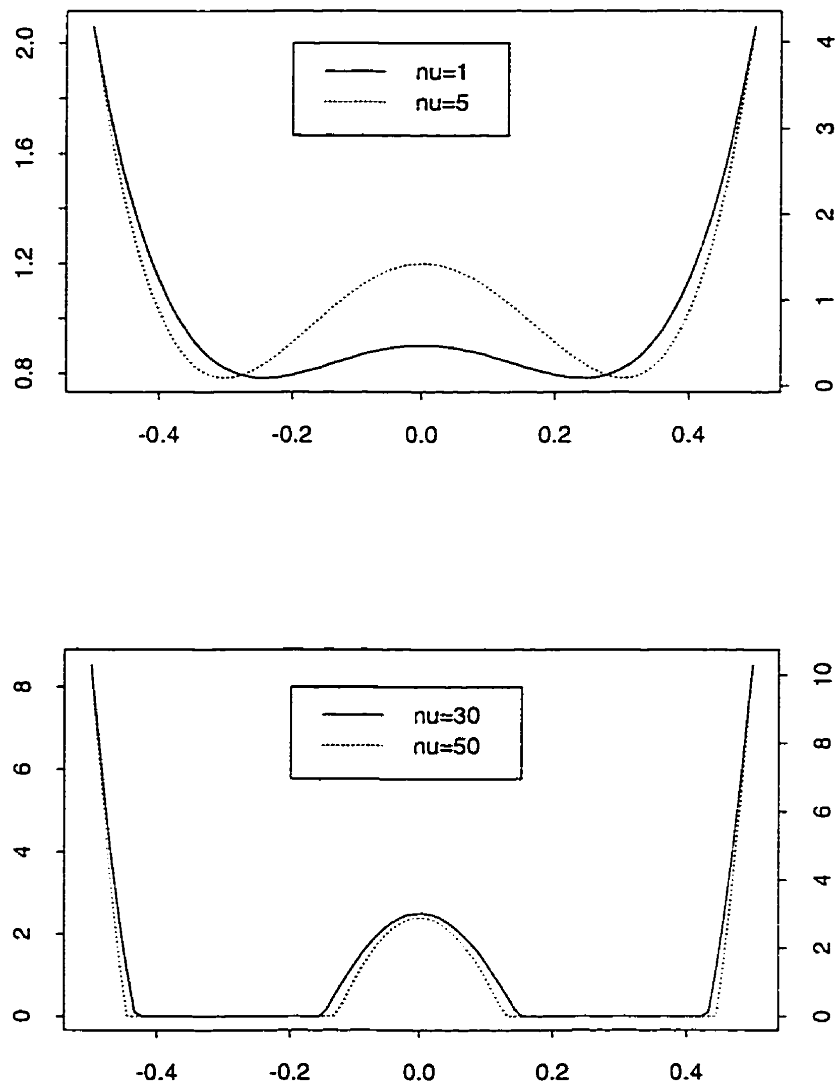


Table 4.1: Values of the constants for the Q-optimal design with the regressor $(1, x, x^2)^T$ and density $a\{(x^2 - s^2)(x^2 - t^2) + b\}^+$.

ν	a	s	t	b	minimax loss
0.0	0.9935	0.3055	0.4051	1.0002	1.0000
0.1	3.7502	0.0000	0.0000	0.2542	1.2949
0.5	12.7553	0.0000	0.0000	0.0659	2.4199
1.0	34.8431	0.2384	0.2448	0.0225	3.7771
5	161.6086	0.2296	0.3588	0.0020	13.7980
10	252.2335	0.2402	0.3658	-0.0008	25.5299
30	595.1804	0.2581	0.37763	-0.0053	70.8142
50	858.2380	0.2617	0.3832	-0.0067	115.2659
100	1649.753	0.2735	0.3864	-0.0090	225.1967
300	4561.569	0.2777	0.3958	-0.0109	659.8831

Table 4.2: Values of the constants for the D-optimal design with the regressor $(1, x, x^2)^T$ and density $a\{(x^2 - s^2)(x^2 - t^2) + b\}^+$.

ν	a	s	t	b	minimax loss
0.0	0.0575	0.00	0.50	17.4020	0.0002
0.1	4.1404	0.00	0.00	0.2290	204.233
0.5	18.5807	0.0000	0.0000	0.0413	862.4813
1.0	35.0934	0.1487	0.1489	0.0192	1534.574
5	242.8386	0.2213	0.3531	-0.0004	5011.176
10	416.5016	0.2388	0.3639	-0.0034	8295.715
30	1026.31	0.2777	0.3638	-0.0078	19687.41
50	1589.44	0.2660	0.3823	-0.0085	30198.11
100	2935.27	0.2725	0.3880	-0.0099	55263.38
300	8528.96	0.2399	0.4212	-0.0096	150218.8

Table 4.3: Values of the constants for the A-optimal design with the regressor $(1, x, x^2)^T$ and density $a\{(x^2 - s^2)(x^2 - t^2) + b\}^+$.

ν	a	s	t	b	minimax loss
0.0	0.0219	0.3598	0.3602	45.65	0.0
0.1	2.5611	0.00	0.00	0.37796	19.0908
0.5	10.5537	0.00	0.00	0.0823	90.8195
1.0	178.3140	0.2384	0.3627	0.0013	150.8177
5	343.3443	0.2460	0.33721	-0.0025	554.8372
10	569.2337	0.2548	0.3785	-0.0050	1014.474
30	1311.34	0.2743	0.3816	-0.0084	2724.04
50	2107.822	0.2920	0.3755	-0.0102	4363.187
100	3742.865	00.2776	0.3934	-0.0106	8344.859
300	10357.91	0.2806	0.3999	-0.0119	23899.08

4.2.2 Design densities for approximately cubic regression

We fit a cubic polynomial although the true model might be only approximately cubic. The regressor $z(x) = (1, x, x^2, x^3)^T$ and $-1/2 \le x \le 1/2$. Similar to the quadratic case, here are four matrices:

$$A_0 = \begin{pmatrix} 1 & 0 & 1/12 & 0 \\ 0 & 1/12 & 0 & 1/80 \\ 1/12 & 0 & 1/80 & 0 \\ 0 & 1/80 & 0 & 1/448 \end{pmatrix}, \quad A_\xi = \begin{pmatrix} 1 & 0 & \xi_2 & 0 \\ 0 & \xi_2 & 0 & \xi_4 \\ \xi_2 & 0 & \xi_4 & 0 \\ 0 & \xi_4 & 0 & \xi_6 \end{pmatrix},$$

$$K_\xi = \begin{pmatrix} k_0 & 0 & k_2 & 0 \\ 0 & k_2 & 0 & k_4 \\ k_2 & 0 & k_4 & 0 \\ 0 & k_4 & 0 & k_6 \end{pmatrix}, \quad H_\xi = \begin{pmatrix} h_0 & 0 & h_1 & 0 \\ 0 & h_2 & 0 & h_3 \\ h_1 & 0 & h_4 & 0 \\ 0 & h_3 & 0 & h_5 \end{pmatrix},$$

where

$$\begin{aligned} h_0 &= 9/4 - 30\xi_2 + 180\xi_2^2, \\ h_1 &= 9\xi_2/4 - 15\xi_2^2 - 15\xi_4 + 180\xi_2\xi_4, \\ h_2 &= 75\xi_2^2 - 840\xi_2\xi_4 + 2800\xi_4^2, \\ h_3 &= 75\xi_2\xi_4 - 420\xi_4^2 - 420\xi_2\xi_6 + 2800\xi_4\xi_6 \\ h_4 &= 9\xi_2^2/4 - 30\xi_2\xi_4 + 180\xi_4^2, \\ h_5 &= 75\xi_4^2 - 840\xi_4\xi_6 + 2800\xi_6^2. \end{aligned}$$

We first recall the suprema of three loss functions in (4.1.2), (4.1.3) and (4.1.4). There are four eigenvalues of each of the matrices $K_\xi H_\xi^{-1}$, $G_\xi^{1/2} A_\xi^{-1} G_\xi^{1/2}$ and $G_\xi^{1/2} A_\xi^{-2} G_\xi^{1/2}$. It can be shown that two larger eigenvalues are of the form, the super script Q is replaced by D, and A.

$$\lambda_1^Q = \frac{\mathcal{B}_1^Q + \sqrt{\mathcal{B}_1^{Q^2} - 4\mathcal{A}_1^Q\mathcal{C}_1^Q}}{2\mathcal{A}_1^Q},$$
$$\lambda_2^Q = \frac{\mathcal{B}_2^Q + \sqrt{\mathcal{B}_2^{Q^2} - 4\mathcal{A}_2^Q\mathcal{C}_2^Q}}{2\mathcal{A}_2^Q},$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$'s are respectively

$$\mathcal{A}_1^Q = h_0h_4 - h_1^2, \quad \mathcal{B}_1^Q = k_0h_4 + k_4h_0 - 2k_2h_1, \quad \mathcal{C}_1^Q = k_0k_4 - k_2^2,$$
$$\mathcal{A}_2^Q = h_2h_5 - h_3^2, \quad \mathcal{B}_2^Q = k_2h_5 + k_6h_2 - 2k_4h_3, \quad \mathcal{C}_2^Q = k_2k_6 - k_4^2.$$

For the D- and A-optimality cases $\mathcal{A}_1 = 1 = \mathcal{A}_2$. The other constants $\mathcal{B}, \mathcal{C}$ are

$$\mathcal{B}_1^D = [-\xi_6(k_2 - h_2) + 2\xi_4(k_4 - h_3) - \xi_2(k_6 - h_5)]/(\xi_4^2 - \xi_2\xi_6),$$
$$\mathcal{C}_1^D = \left[(k_2 - h_2)(k_6 - h_5) - (k_4 - h_3)^2\right]/(\xi_2\xi_6 - \xi_4^2),$$
$$\mathcal{B}_2^D = [\xi_4(k_0 - h_0) - 2\xi_2(k_2 - h_1) + k_4 - h_4]/(\xi_4 - \xi_2^2),$$
$$\mathcal{C}_2^D = \left[(k_0 - h_0)(k_4 - h_4) - (k_2 - h_1)^2\right]/(\xi_4 - \xi_2^2),$$
$$\mathcal{B}_1^A = \left[(k_2 - h_2)(\xi_6^2 + \xi_4^2) - 2(k_4 - h_3)(\xi_4\xi_6 + \xi_2\xi_4) + (k_6 - h_5)(\xi_2^2 + \xi_4^2)\right]/(\xi_4^2 - \xi_2\xi_6)^2,$$
$$\mathcal{C}_1^A = \left[(k_2 - h_2)(k_6 - h_5) - (k_4 - h_3)^2\right]/(\xi_4^2 - \xi_2\xi_6)^2,$$
$$\mathcal{B}_2^A = [(k_0 - h_0)(\xi_2^2 + \xi_4^2) - 2\xi_2(1 + \xi_4)(k_2 - h_1) + (1 + \xi_2^2)(k_4 - h_4)]/(\xi_4 - \xi_2^2)^2,$$
$$\mathcal{C}_2^A = \left[\xi_4^2 + \xi_2^4 - 2\xi_2^2\xi_4\right]\left[(k_0 - h_0)(k_4 - h_4) - (k_2 - h_1)^2\right]/(\xi_4 - \xi_2^2)^4.$$

In this cubic regression, we are working on the density of the form

$$a[(x^2 - s^2)(x^2 - t^2)(x^2 - u^2) + b]^+. \tag{4.2.2}$$

The coefficient a is non-negative, b is real and $0 \le s \le t \le u \le 1/2$. Taking transformations on s, t, u we rewrite the supremum of loss function l, say, in terms of s, t', u', and b. Consequently our optimization problem, for each fixed ν, is to

$$\text{Minimize } l[s, s + t'(1 - 2s), t + u'(1 - 2t), b]; \text{ subject to } \begin{cases} 0 \le s \le \frac{1}{2} \\ 0 \le t' \le \frac{1}{2} \\ 0 \le u' \le \frac{1}{2} \\ |b| < +\infty. \end{cases}$$

We present the design densities for some values of ν in Figure 4.2, Table 4.4, Table 4.5 and Table 4.6.

Figure 4.2: Q-optimal design densities for an approximately cubic regression model with contaminating space $\mathcal{F}_2$.

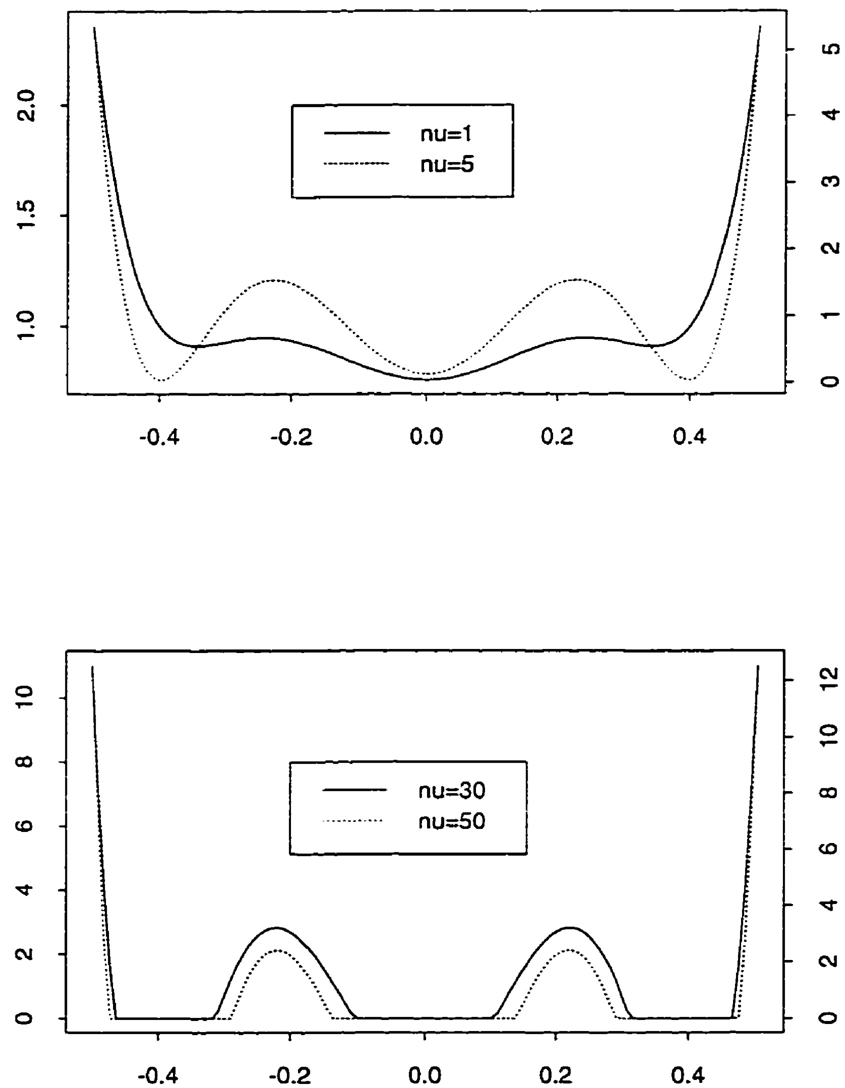


Table 4.4: Values of the constants for the Q-design with regressor $(1, x, x^2, x^3)^T$ and density $a[(x^2 - s^2)(x^2 - t^2)(x^2 - u^2) + b]^+$.

ν	a	s	t	u	b	minimax loss
0	0.9931	0.1768	0.3255	0.4122	1.0069	1.0
1	375.6921	0.2089	0.2783	0.3789	0.0025	4.6856
5	2502.225	0.1354	0.3075	0.4489	0.0004	18.2199
10	3950.191	0.1129	0.3164	0.4532	0.0002	34.2523
30	9116.176	0.0766	0.3275	0.4609	-0.0001	96.7066
50	13387.3	0.0846	0.3206	0.4671	-0.0002	158.3118
100	25209.03	0.1064	0.3066	0.4762	-0.0001	311.0753
300	65091.77	0.0818	0.3181	0.4786	-0.0003	917.073

Table 4.5: Values of the constants for the D-design with regressor $(1, x, x^2, x^3)^T$ and density $a[(x^2 - s^2)(x^2 - t^2)(x^2 - u^2) + b]^+$.

ν	a	s	t	u	b	minimax loss
0	1.9989	0.0786	0.3903	0.4544	0.5	2.4950
1	772.11	0.2154	0.3201	0.394	0.00128	3682302
5	3886.56	0.1284	0.3365	0.4347	0.0001	11006605
10	5705.22	0.1666	0.2766	0.4567	0.0002	17407275
30	5103.67	0.1868	0.3493	0.4290	0.0001	54582464
50	20954.8	0.1537	0.2846	0.4740	0.0000	59985266
100	38705.3	0.1680	0.2726	0.4813	0.0000	107854319
300	105929.5	0.1850	0.2582	0.4890	0.0000	287322642

Table 4.6: Values of the constants for the A-design with regressor $(1, x, x^2, x^3)^T$ and density $a[(x^2 - s^2)(x^2 - t^2)(x^2 - u^2) + b]^+$.

ν	a	s	t	u	b	minimax loss
0	0.9905	0.2191	0.3745	0.4407	1.0098	0.0000
1	22769.3	0.1278	0.3210	0.4510	0.0003	2416.38
5	4914.62	0.1458	0.3442	0.4491	0.0001	8975.609
10	7849.92	0.1281	0.3802	0.4395	-0.0002	16507.76
30	19640.54	0.14445	0.3187	0.4733	-0.0000	43453.84
50	31476.4	0.1129	0.3418	0.4697	-0.0003	69737.21
100	57743.0	0.1147	0.3427	0.4733	-0.0004	133714.1
300	160703.2	0.0710	0.3710	0.4662	-0.0000	383964.4

4.3 Bivariate regression model

4.3.1 Approximately linear with interaction terms

Wiens (1990) also fits the bivariate surface with interactions. That is, the regressor is $\boldsymbol{z}(\boldsymbol{x}) = (1, x_1, x_2, x_1x_2)^T$, and the design space is a rectangle $S = [-1/2, 1/2] \times [-1/2, 1/2]$. Wiens restricts to symmetric designs as well as exchangeable m. We find then

$$
\begin{aligned}
A_0 &= \operatorname{diag}(1, 1/12, 1/12, 1/144), \\
A_\xi &= \operatorname{diag}(1, \xi_2, \xi_2, \xi_{22}), \\
K_\xi &= \operatorname{diag}(k_0, k_2, k_2, k_{22}), \\
H_\xi &= \operatorname{diag}(1, 12\xi_2^2, 12\xi_2^2, 144\xi_{22}^2).
\end{aligned}
$$

where

$$\begin{aligned}
\xi_2 &= \int x_1^2 m(\boldsymbol{x})\, d\boldsymbol{x}, \quad \xi_{22} = \int x_1^2 x_2^2 m(\boldsymbol{x})\, d\boldsymbol{x}, \\
k_0 &= \int m^2(\boldsymbol{x})\, d\boldsymbol{x}, \quad k_2 = \int x_1^2 m^2(\boldsymbol{x})\, d\boldsymbol{x}, \\
k_{22} &= \int x_1^2 x_2^2 m^2(\boldsymbol{x})\, d\boldsymbol{x}.
\end{aligned}$$

The maximum IMSE is now

$$\lambda_{\max}(K_\xi H_\xi^{-1}) + \nu\left(1 + \frac{1}{6\xi_2} + \frac{1}{144\xi_{22}}\right).$$

There are three eigenvalues in the matrix $K_\xi H_\xi^{-1}$:

$$\begin{aligned}
\lambda_1 &= \int m^2(\boldsymbol{x})\, d\boldsymbol{x}, \\
\lambda_2 &= \frac{1}{12\xi_2^2} \int x_1^2 m^2(\boldsymbol{x})\, d\boldsymbol{x}, \\
\lambda_3 &= \frac{1}{144\xi_{22}^2} \int x_1^2 x_2^2 m^2(\boldsymbol{x})\, d\boldsymbol{x}.
\end{aligned}$$

Wiens (1990) obtains the minimax density of the form

$$m_0(\boldsymbol{x}) = \{a + b(x_1^2 + x_2^2) + c x_1^2 x_2^2\}^+, \tag{4.3.1}$$

with the coefficients determined by the three equations

$$\int m_0(\boldsymbol{x})\, d\boldsymbol{x} = 1, \int x_1^2 m_0(\boldsymbol{x})\, d\boldsymbol{x} = \xi_2, \int x_1^2 x_2^2 m_0(\boldsymbol{x})\, d\boldsymbol{x} = \xi_{22}.$$

For small values of ν ($0 \le \nu \le 1.2758$) Wiens calculates the minimax densities explicitly (see Table 4.7).

Our goal which as we introduced in Section 4.1 is to look for tractable densities for all ν. The algorithm in Section 4.1 can be generalized to this case as

Table 4.7: Values of the constants for the Q-design with regressors $(1, x_1, x_2, x_1x_2)^T$ and density $\{a + b(x_1^2 + x_2^2) + cx_1^2x_2^2\}^+$.

ν	a	b	c	ξ_2	ξ_{22}	minimax loss
0.0	1	0	0	0.0833	0.0069	1.0000
0.1026	0.8716	0.5409	5.5091	0.0888	0.0080	1.3957
0.2452	0.7379	1.1455	10.2538	0.0944	0.0091	1.9125
0.4307	0.6004	1.7947	14.4636	0.1000	0.0102	2.5491
0.6621	0.4603	2.4761	18.2869	0.1056	0.0112	3.3057
0.9427	0.3182	3.1816	21.8213	0.1111	0.0122	4.1837
1.2758	0.1745	3.9055	25.134	0.1166	0.0133	5.2412

long as we restrict to symmetric and exchangeable densities. We determine the minimizing coefficients of a general density of the form

$$m(x_1, x_2) = a_0\{(x_1^2 - s^2 + b_0)^+(x_2^2 - s^2 + b_0)^+\} \tag{4.3.2}$$

over the domain $0 \le s \le 1/2$, $-\infty < b_0 < \infty$. We show the optimal densities for $\nu = 2$ and $\nu = 10$ in Figure 4.3. We compare our optimal designs in Table 4.8 and Wiens's minimax designs in Table 4.7.

We write out the maximum loss functions for D- and A-optimality cases.

$$\begin{aligned}
\sup_{\mathcal{F}_2} \mathcal{L}_D(f, \xi) &= \frac{1}{\xi_1\xi_{22}}(\nu + \max\{k_0 - 1, \frac{k_2 - 12\xi_2^2}{\xi_2}, \frac{k_{22} - 144\xi_{22}^2}{\xi_{22}}\}) \\
\sup_{\mathcal{F}_2} \mathcal{L}_A(f, \xi) &= \nu(1 + \frac{2}{\xi_2} + \frac{1}{\xi_{22}}) + \max\{k_0 - 1, \frac{k_2 - 12\xi_2^2}{\xi_2^2}, \frac{k_{22} - 144\xi_{22}^2}{\xi_{22}^2}\}.
\end{aligned}$$

We present the optimal designs in Table 4.9 and Table 4.10.

Figure 4.3: Q-optimal design densities for an approximately linear bivariate regression model $\mathcal{F}_2$: (a) $m = 51.37(x^2 + 0.05619)(y^2 + 0.05619)$ when $\nu = 2$; (b) $m = 178.92(x^2 - 0.00988)^+(y^2 - 0.00988)^+$ when $\nu = 10$.

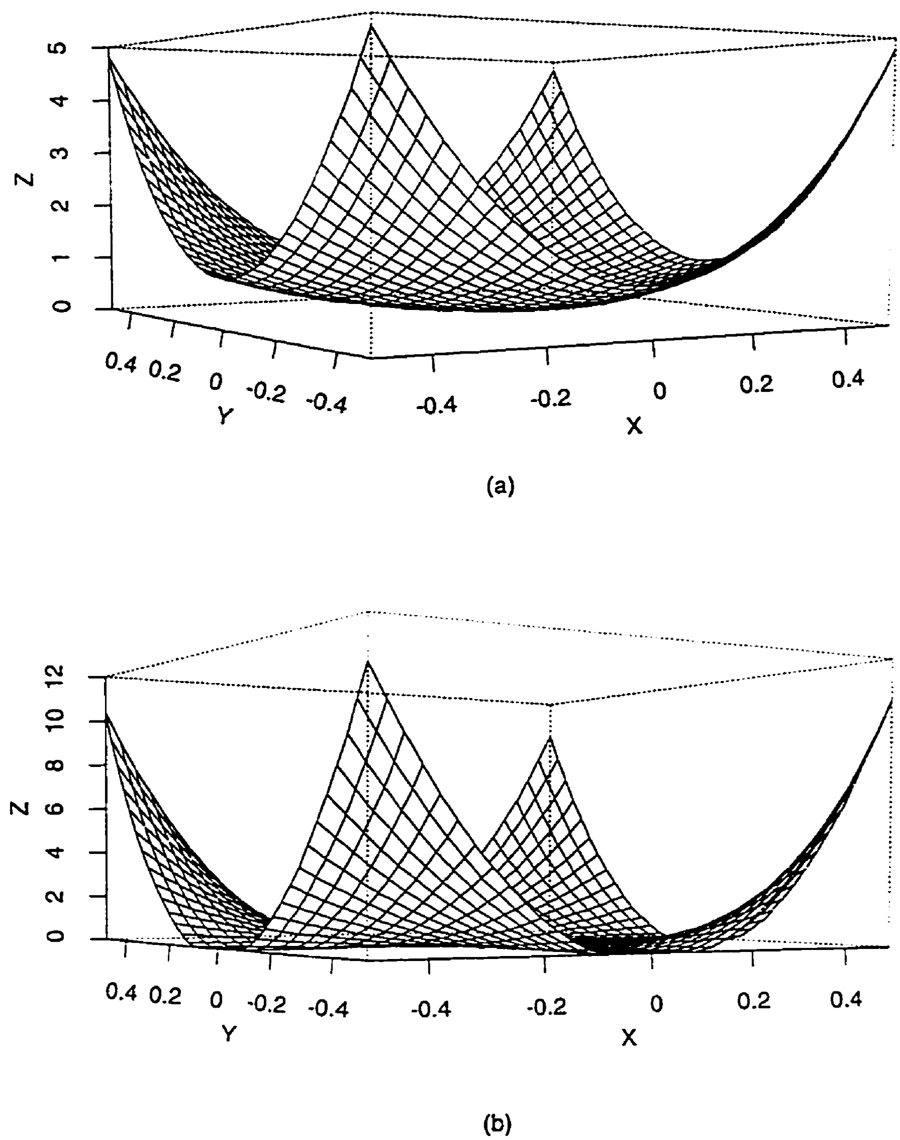


Table 4.8: Values of the constants for Q-design regressor $(1, x_1, x_2, x_1x_2)^T$ and density $a_0\{(x_1^2 - s^2 + b_0)^+(x_2^2 - s^2 + b_0)^+\}$.

ν	a_0	b_0	s	ξ_2	minimax loss
0.0	$1.05e^{-6}$	975.672	0.49997	0.0833	1.0000
0.2452	4.3300	0.3972	0.0000	0.0949	1.9136
0.6621	16.3292	0.1641	0.0000	0.1058	3.3058
1.2758	33.4823	0.0895	0.0000	0.1155	5.1881
5	108.1208	0.0171	0.0650	0.1411	15.2121
10	178.9152	0.0322	0.2051	0.1562	27.3458
30	471.074	0.0082	0.2494	0.1776	71.9586
100	1597.25	-0.0284	0.2700	0.1970	217.1245

Table 4.9: Values of the constants for the D-design with regressor $(1, x_1, x_2, x_1x_2)^T$, density $a_0\{(x_1^2 - s^2 + b_0)^+(x_2^2 - s^2 + b_0)\}^+$.

ν	a_0	b_0	s	ξ_2	minimax loss
0.0	0.0000	13336.25	0.0000	0.0833	0.0001
0.2452	12.445	0.2000	0.0001	0.1029	3458.756
0.6621	85.600	0.02475	0.000	0.1347	5581.7
1.2758	181.02	0.04400	0.2333	0.1565	6896.11
5	550.666	0.0011	0.2483	0.1804	11230.34
10	1005.57	-0.0158	0.2624	0.1904	15411.77
30	2869.85	-0.0457	0.2731	0.2043	28271.81
100	10249.13	-0.1170	0.1927	0.2168	63007.22

Table 4.10: Values of the constants for the A-design with regressor $(1, x_1, x_2, x_1x_2)^T$, density $a_0\{(x_1^2 - s^2 + b_0)^+(x_2^2 - s^2 + b_0)\}^+$.

ν	a_0	b_0	s	ξ_2	minimax loss
0.0	0.0000	401.455	0.1977	0.0833	0.0000
0.2452	4.2392	0.4024	0.0000	0.0948	36.7937
0.6621	28.8611	0.1028	0.0000	0.1132	84.001
1.2758	96.26	0.01859	0.0000	0.1378	133.635
5	244.56	0.0326	0.2389	0.1639	338.74
10	421.61	0.0127	0.2487	0.1755	574.96
30	1216.18	-0.0204	0.2669	0.1932	1389.13
100	4506.07	-0.05693	0.2765	0.2092	3848.60

4.3.2 Approximately quadratic with interaction terms

We can extend the idea above to the quadratic regression model with interaction, that is, $\boldsymbol{x} = (x_1, x_2)^T, \boldsymbol{z}(\boldsymbol{x}) = (1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^T$. The design domain is again a rectangle $[-1/2, 1/2] \times [-1/2, 1/2]$. We obtain the basic matrices with $\xi_{22} = \int x_1^2x_2^2 m(\boldsymbol{x})\, d\boldsymbol{x}$, $k_{22} = \int x_1^2x_2^2 m^2(\boldsymbol{x})\, d\boldsymbol{x}$.

$$A_0 = \begin{pmatrix} 1 & 0 & 0 & 1/12 & 1/12 & 0 \\ 0 & 1/12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/12 & 0 & 0 & 0 \\ 1/12 & 0 & 0 & 1/80 & 1/144 & 0 \\ 1/12 & 0 & 0 & 1/144 & 1/80 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/144 \end{pmatrix}.$$

$$
A_\xi = \begin{pmatrix} 1 & 0 & 0 & \xi_2 & \xi_2 & 0 \\ 0 & \xi_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \xi_2 & 0 & 0 & 0 \\ \xi_2 & 0 & 0 & \xi_4 & \xi_{22} & 0 \\ \xi_2 & 0 & 0 & \xi_{22} & \xi_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & \xi_{22} \end{pmatrix}.
$$

$$
K_\xi = \begin{pmatrix} k_0 & 0 & 0 & k_2 & k_2 & 0 \\ 0 & k_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_2 & 0 & 0 & 0 \\ k_2 & 0 & 0 & k_4 & k_{22} & 0 \\ k_2 & 0 & 0 & k_{22} & k_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{22} \end{pmatrix}, \quad H_\xi = \begin{pmatrix} h_1 & 0 & 0 & h_2 & h_2 & 0 \\ 0 & h_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_3 & 0 & 0 & 0 \\ h_2 & 0 & 0 & h_4 & h_5 & 0 \\ h_2 & 0 & 0 & h_5 & h_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & h_6 \end{pmatrix}.
$$

where

$$
\begin{aligned}
h_1 &= 7/2 - 60\xi_2 + 360\xi_2^2, \\
h_2 &= 7\xi_2/2 - 30\xi_2^2 - 15\xi_4 + 180\xi_2\xi_4 - 15\xi_{22} + 180\xi_2\xi_{22}, \\
h_3 &= 12\xi_2^2, \\
h_4 &= 7\xi_2^2/2 - 30\xi_2\xi_4 - 30\xi_2\xi_{22} + 180\xi_4^2 + 180\xi_{22}^2, \\
h_5 &= 7\xi_2^2/2 - 30\xi_2\xi_4 - 30\xi_2\xi_{22} + 360\xi_4\xi_{22}, \quad h_6 = 144\xi_{22}^2.
\end{aligned}
$$

There are 6 eigenvalues in the matrix $K_\xi H_\xi^{-1}$. It can be shown the maximum is among the following 4 eigenvalues.

$$
\lambda_1 = \frac{k_2}{h_3}, \quad \lambda_2 = \frac{k_{22}}{h_6}, \quad \lambda_3 = \frac{k_4 - k_{22}}{h_4 - h_5}, \quad \lambda_4 = \frac{\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}},
$$

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are

$$
\mathcal{A} = h_1 h_4 + h_1 h_5 - 2h_2^2,
$$

$$\mathcal{B} = k_0 h_4 + k_0 h_5 + k_4 h_1 + k_{22} h_1 - 4k_2 h_2,$$
$$\mathcal{C} = k_0 k_4 + k_0 k_{22} - 2k_2^2.$$

The maximum IMSE is, with above 4 eigenvalues,

$$\begin{aligned}\sup_{\mathcal{F}_2} \mathcal{L}_Q(f,\xi) &= \max\{\lambda_i\}_{i=1}^4 + \nu \mathrm{tr}(A_0 A_\xi^{-1}) \\ &= \max\{\lambda_i\}_{i=1}^4 + \nu[\frac{1}{6\xi_2} + \frac{1}{144\xi_{22}} \\ &+ \frac{\xi_2 - 6\xi_4 - 6\xi_{22}}{6(2\xi_2^2 - \xi_{22} - \xi_4)} - \frac{4\xi_2^2 + 60\xi_2\xi_4 - 60\xi_2\xi_{22} - 9\xi_4 + 5\xi_{22}}{360(\xi_4^2 - 2\xi_2^2\xi_4 + 2\xi_2^2\xi_{22} - \xi_{22}^2}))].\end{aligned}$$

For the D- and A-optimality cases, the maximum loss functions are

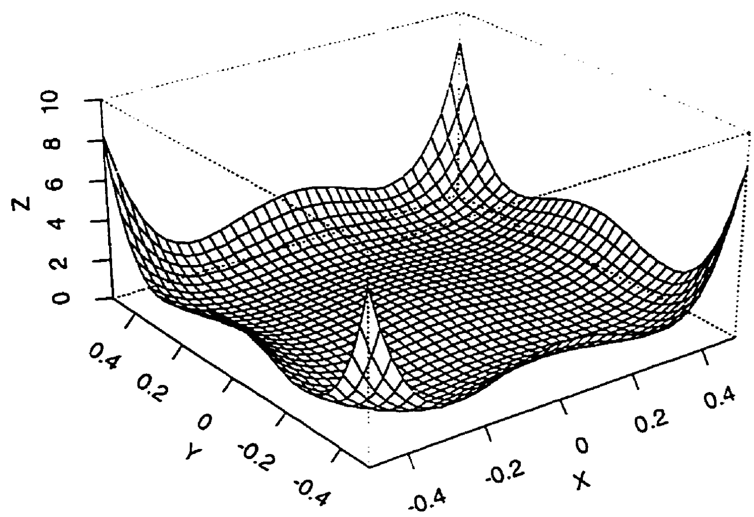
$$\begin{aligned}\sup_{\mathcal{F}_2} \mathcal{L}_D(f,\xi) &= \frac{1}{\xi_2^2\xi_{22}(\xi_4^2 - 2\xi_4\xi_2^2 + 2\xi_{22}\xi_2^2 - \xi_{22}^2)}(\nu + \max(\lambda_1^D, \lambda_2^D, \lambda_2^D, \lambda_2^D, \})), \\ \sup_{\mathcal{F}_2} \mathcal{L}_A(f,\xi) &= \nu(\frac{2}{\xi_2} + \frac{1}{\xi_{22}} + \frac{\xi_4 + \xi_{22}}{\xi_4 - 2\xi_2^2 + \xi_{22}} + \frac{2(\xi_4 - \xi_2^2)}{\xi_4^2 - \xi_4\xi_2^2 + 2\xi_{22}\xi_2^2 - \xi_{22}^2}) \\ &+\max\{\lambda_1^A, \lambda_2^A, \lambda_3^A, \lambda_4^A, \}.\end{aligned}$$

where the eigenvalues are function of $\xi_2, \xi_4, \xi_{22}, k_0, k_2, k_4$ and k_{22}. We omit the expressions of the eigenvalues because they are very lengthy.

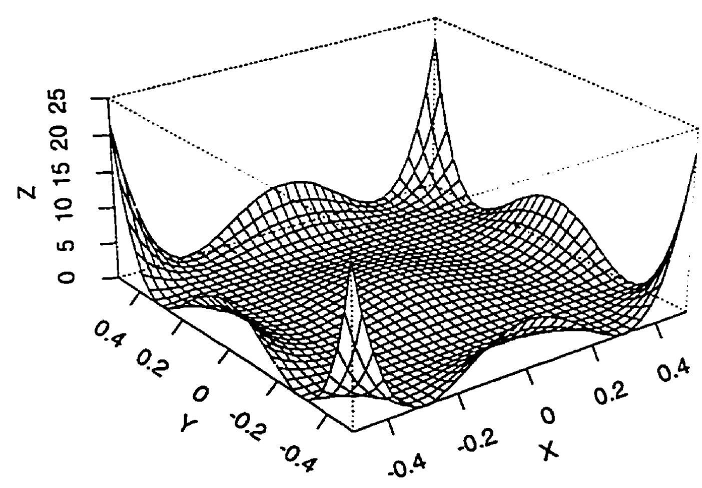
We provide designs for some ν in Figure 4.4, Table 4.11, Table 4.12 and Table 4.13. In this case, the density is of the form

$$m(\boldsymbol{x}) = a\{(x_1^2 - s^2)(x_1^2 - t^2) + b\}^+\{(x_2^2 - s^2)(x_2^2 - t^2) + b\}^+. \qquad (4.3.3)$$

Figure 4.4: Q-optimal design densities for an approximately quadratic bivariate regression model with $\mathcal{F}_2$: (a) $m = 5467((x^2 - 0.036)(x^2 - 0.111) + 0.0093)^+((y^2 - 0.036)(y^2 - 0.111) + 0.0093)^+$ when $\nu = 2$; (b) $m = 32042.64((x^2 - 0.2256^2)(x^2 - 0.3565^2) + 0.0015)^+((y^2 - 0.2256^2)(y^2 - 0.3565^2) + 0.0015)^+$ when $\nu = 10$.



(a)



(b)

Table 4.11: Values of the constants for the Q-design with regressor $(1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^T$ and density $a\{(x_1^2 - s^2)(x_1^2 - t^2) + b\}^+\{(x_2^2 - s^2)(x_2^2 - t^2) + b\}^+$.

ν	a	b	s	t	ξ_2	ξ_4	minimax loss
0	0.0005	43.3672	0.0000	0.4863	0.0833	0.0125	1.0000
1	23349.22	0.01617	0.1829	0.3226	0.1039	0.0180	6.2334
5	16814.33	0.0035	0.2150	0.3474	0.1175	0.0228	24.6463
10	32042.64	0.0015	0.2256	0.3565	0.1194	0.0243	46.2915
30	93752.19	-0.0018	0.2386	0.3671	0.1239	0.0268	128.866
100	363563.8	-0.0051	0.2517	0.3782	0.1274	0.0289	406.7444

Table 4.12: Values of the constants for the D-design with regressor $(1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^T$ and density $a\{(x_1^2 - s^2)(x_1^2 - t^2) + b\}^+\{(x_2^2 - s^2)(x_2^2 - t^2) + b\}^+$.

ν	a	b	s	t	ξ_2	ξ_4	minimax loss
0	0.0002	64.54	0.3137	0.4216	0.0833	0.0125	9.86
1	3658.1	0.0103	0.1765	0.2654	0.1212	0.0220	211048448
5	31846.77	0.0005	0.2286	0.3116	0.1477	0.0303	378480288
10	52476.31	-0.0003	0.2135	0.3372	0.1530	0.0323	507068013
30	149875.1	-0.0028	0.2247	0.3578	0.1605	0.0354	882575359
100	532912.5	-0.0057	0.2436	0.3704	0.1674	0.0383	1840399532

Table 4.13: Values of the constants for the A-design with regressor $(1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^T$ and density $a\{(x_1^2 - s^2)(s_1^2 - t^2) + b\}^+\{(x_2^2 - s^2)(x_2^2 - t^2) + b\}^+$.

ν	a	b	s	t	ξ_2	ξ_4	minimax loss
0	0.0009	32.606	0.1538	0.4432	0.0833	0.0125	0.0000
1	91.1818	0.0922	0.0000	0.0000	0.0947	0.0152	485.15
5	1502.37	0.0204	0.1809	0.2951	0.1037	0.0178	2191.14
10	5791.90	0.0087	0.2168	0.3363	0.1062	0.0191	4105.3
30	28588.87	0.001618	0.2368	0.3599	0.1103	0.0221	10696.53
100	86241.37	-0.00165	0.2430	0.3674	0.1186	0.0255	30281

Remark

Although in general the minimax density (4.1.6) is complicated, the quadratic model without a constant term is analytically solvable. Setting the regressor $z(x) = (x, x^2)^T$, $-1/2 \le x \le 1/2$, we calculate the matrices

$$A_0 = \operatorname{diag}(1/12, 1/80), \quad A_\xi = \operatorname{diag}(\xi_2, \xi_4),$$
$$K_\xi = \operatorname{diag}(k_2, k_4), \quad H_\xi = \operatorname{diag}(1/(12\xi_2^2), 1/(80\xi_4^2)).$$

The two eigenvalues of $K_\xi H_\xi^{-1}$ are

$$\lambda_1(m(x;\xi_2,\xi_4)) := \frac{1}{12\xi_2^2}\int x^2 m^2(x)\,dx,$$
$$\lambda_2(m(x;\xi_2,\xi_4)) := \frac{1}{80\xi_4^2}\int x^4 m^2(x)\,dx.$$

The maximum IMSE, ignoring the term η^2, is now

$$\max_{\mathcal{F}_2} \mathcal{L}_Q(f,\xi) = \max\{\lambda_1(m(x;\xi_2,\xi_4)), \lambda_2(m(x;\xi_2,\xi_4))\}$$

$$+\nu(\frac{1}{12\xi_2}+\frac{1}{80\xi_4}).$$

In three steps, we obtain the minimax density as follows.

1. First we assume that $\lambda_1(m(x)) \geq \lambda_2(m(x))$. Holding ξ_2, ξ_4 fixed we

$$\text{minimize} \int x^2 m^2(x)\,dx, \text{ subject to } \begin{cases} \int m(x)\,dx = 1, \\ \int x^2 m(x)\,dx = \xi_2, \\ \int x^4 m(x)\,dx = \xi_4. \end{cases} \qquad (4.3.4)$$

   Let $m(x; \xi_2, \xi_4)$ be the minimizer.

2. Minimize

$$\frac{1}{12\xi_2^3}\int x^2 m^2(x; \xi_2, \xi_4)\,dx + \nu(\frac{1}{12\xi_2}+\frac{1}{80\xi_4}) \quad \text{over}(\xi_2, \xi_4),$$

   whose domain is $0 \leq \xi_2 \leq \sqrt{\xi_4} \leq \sqrt{\xi_2} \leq 1$. Let (ξ_2^*, ξ_4^*) be the minimizers.

3. Verify that

$$\frac{1}{12\xi_2^*}\int x^2 m^2(x; \xi_2^*, \xi_4^*)\,dx \geq \frac{1}{80\xi_4^{*2}}\int x^4 m^2(x; \xi_2^*, \xi_4^*)\,dx.$$

For the first step, put

$$m_t(x) = (1-t)m_0(x) + tm_1(x),$$

where m_0, m_1 satisfy the three conditions in (4.3.4). Set

$$L(t) = \int x^2 m_t^2(x)\,dx + a\int m_t(x)\,dx - b\int x^2 m_t(x)\,dx - c\int x^4 m_t(x)\,dx.$$

for Lagrange multipliers $a, b, c \geq 0$. Since $L(t)$ is a convex function of t for all m_0 and m_1, the density m_0 is optimal if and only if $L'(0) \geq 0$ for each m_1. This means that m_0 must satisfy

$$\begin{aligned} L'(t)|_{t=0} &= 2\int x^2 m_t(x) m_t'(x)\,dx + a\int m_t'(x)\,dx - b\int x^2 m_t'(x)\,dx \\ &\quad -c\int x^4 m_t'(x)\,dx \\ &= \int (m_1 - m_0)\{2m_0(x)x^2 + a - bx^2 - cx^4\}\,dx \geq 0. \end{aligned}$$

The above inequality requires that

$$m_0(x) = b(cx^2 + 1 - a/x^2)^+. \tag{4.3.5}$$

The second step is equivalent to finding (ξ_2^*, ξ_4^*) minimizing

$$\begin{aligned} &\frac{1}{12\xi_2^2}\int x^2 b(cx^2 + 1 - a/x^2)^+ m_0(x)\,dx + \nu\left(\frac{1}{12\xi_2} + \frac{1}{80\xi_4}\right) \\ = &\frac{b}{12\xi_2^2}(c\xi_4 + \xi_2 - a) + \nu\left(\frac{1}{12\xi_2} + \frac{1}{80\xi_4}\right). \end{aligned}$$

In the third step, we must verify that for the minimax $m(x; a^*, b^*, c^*)$,

$$\frac{1}{12\xi_2^2}\int b^* x^2(c^* x^2 + 1 - a^*/x^2)^+ m(x)\,dx \geq \frac{1}{80\xi_4^2}\int b^* x^4(c^* x^2 + 1 - a^*/x^2)^+ m(x)\,dx.$$

This is equivalent to

$$\frac{1}{12\xi_2^{*2}}(c^*\xi_4^* + \xi_2^* - a^*) \geq \frac{1}{80\xi_4^{*2}}(c^*\xi_6^* + \xi_4^* - a^*\xi_2^*).$$

Suppose that $m(x; a, b, c) \geq 0$ for $x \in (s, 1/2)$. Since $\int m(x; a, b, c) = 1$, the coefficients a, b, c and the moments ξ_4, ξ_6 may be written in terms of s and ξ_2, whose domain is $0 \leq s \leq 1/2, 0 \leq \xi_2 \leq 1$. For very small ν all three steps can be carried out successfully. For larger ν, $\lambda_1 \geq \lambda_2$ is not satisfied at the minimizing density. We present some of the minimax densities for $0 \leq \nu \leq 0.4$ in Table 4.14 and we compare in Table 4.15 the densities obtained by method in this chapter.

Table 4.14: Values of the constants and eigenvalues for the regressor $(x, x^2)^T$ and the density $b(cx^2 + 1 - a/x^2)^+$.

ν	a	b	c	λ_1	λ_2	minimax loss
0.0	1	0.0000	1.00	0.0000	1	1
0.1	0.0078	1.4361	0.2151	1.0160	1.0023	1.1599
0.2	0.0146	1.6561	0.3538	1.0373	1.0071	1.2956
0.4	0.0264	2.0029	0.5418	1.0809	1.0200	1.5370

Table 4.15: Values of the constants and eigenvalues for the regressor $(x, x^2)^T$ and the density $a\{(x^2 - s^2)(x^2 - t^2) + b\}^+$.

ν	a	s	t	b	λ_1	λ_2	minimax loss
0.0	0.9923	0.2333	0.4263	1.0003	1.0000	1.0000	1.0000
0.1	5.3934	0.0000	0.0000	0.1729	1.0087	1.0069	1.1909
0.2	11.1534	0.0000	0.0000	0.0772	1.0322	1.0237	1.3650
0.4	24.2226	0.0000	0.0000	0.0288	1.4849	1.2462	1.6699
1.0	80.0000	0.0000	0.0000	0.0000	1.4849	1.2462	2.3115
5.0	118.9624	0.0000	0.0000	-0.0078	1.7716	1.3814	5.4289
10	270.7635	0.1989	0.2876	-0.0033	2.1854	1.6088	8.9431
30	399.4985	0.0000	0.0000	-0.02986	3.1812	2.1740	21.7318

4.4 Applications on ozonation data

We have obtained optimal designs that are continuous. In practice we need to implement these continuous designs. We illustrate an implementation technique using the ozonation experiment that was introduced in Chapter 1. Before the experiment the researcher was convinced that TSS and O_3 were linearly related, and that as O_3 increases TSS decreases. After a bench-scale experiment he realized that this assumption is questionable and it is more reasonable to assume that the relationship is only approximately linear. That is, we adopt the approximately linear regression model with $\mathcal{F}_2$ type contamination. Suppose that $n = 16$ and $\nu = 10$. From Section 2.4.1 the minimax density is then, after taking a transformation on O_3 so that it lies between $-1/2$ and $1/2$,

$$m_0(x) = 15.55(x^2 - 0.024)^+.$$

The design points might be chosen by selecting the n points $M_0^{-1}(\frac{i-1}{n-1}), i = 1, \ldots, n$. For this particular example the minimax design points are

$$\pm 0.5, \pm .4802, \pm 0.4583, \pm 0.4338,$$

$$\pm 0.4057, \pm 0.3720, \pm 0.3286, \pm 0.2608.$$

The ozonation is not the only factor useful to remove suspended solids. The other important factor is the Gas to Liquid ratio which we denote GL. Three different levels, 0.2, 0.4, 0.6 of GL were applied in this experiment. It might be useful to adopt an approximately linear model in a bivariate regression setting with possible interaction terms. After taking a linear transformation on

GL so that it lies between $-1/2$ and $1/2$. we consider an approximately linear regression model with regressors $\boldsymbol{z}(\boldsymbol{x}) = (1, x_1, x_2, x_1x_2)^T$ and the design space $[-1/2, +1/2] \times [-1/2, +1/2]$. The optimal density is of the form (4.3.2). We assume again that $\nu = 10$. and the sample size is 16. The corresponding density is

$$m_0(x_1, x_2) = 178.91(x_1^2 - 0.0098827)^+(x_2^2 - 0.0098827)^+.$$

By independence the corresponding distribution function M_0 can be written as the product of M_1 and M_2. The optimal design points can be chosen from

$$(M_1^{-1}(\frac{4(i-1)}{n-1}), M_2^{-1}(\frac{4(j-1)}{n-1})), \; i, j = 1, \ldots, 4.$$

For this particular example the optimal design points are

$(-0.5, +0.2313)$, $(-0.5, +0.4272)$, $(-0.3960, +0.2313)$, $(-0.3960, +0.4272)$,

$(+0.3960, +0.2313)$, $(+0.3960, +0.4272)$, $(+0.5, +0.2313)$, $(+0.5, +0.4272)$,

$(-0.5, -0.2313)$, $(-0.5, -0.4272)$, $(-0.3960, -0.2313)$, $(-0.3960, -0.4272)$,

$(+0.3960, -0.2313)$, $(+0.3960, -0.4272)$, $(+0.5, -0.2313)$, $(+0.5, -0.4272)$.

Chapter 5

Conclusions

Our goal in this thesis is to find a minimax design ξ^*, that is,

$$\sup_{\mathcal{F}} \mathcal{L}(f,\xi^*) = \inf_{\Xi} \sup_{\mathcal{F}} \mathcal{L}(f,\xi),$$

where three loss functions and three different contaminating spaces are considered. With respect to contaminating spaces, we distinguish these problems by denoting them (P1), (P2), and (P3). In the first two cases, the fitted model is $\hat{y}(x) = \boldsymbol{z}_1^T(x)\hat{\boldsymbol{\theta}}$ the differences are found in the true model as shown below.

	True Model	Contaminating space
$(P1)$:	$E(Y \mid x) = \boldsymbol{z}_1^T(x)\boldsymbol{\theta}_0 + \boldsymbol{u}^T(x)\boldsymbol{\beta}$	$\mathcal{F} = \{\int(\boldsymbol{u}^T(x)\boldsymbol{\beta})^2\,dx \le \eta^2,$ $\int \boldsymbol{u}(x)\boldsymbol{z}_1^T(x)\,dx = 0\}$
$(P2)$:	$E(Y \mid x) = \boldsymbol{z}_1^T(x)\boldsymbol{\theta}_1 + \boldsymbol{z}_2^T(x)\boldsymbol{\beta}$	$\mathcal{F} = \{\int(\boldsymbol{z}_2^T(x)\boldsymbol{\beta})^2\,dx \le \eta^2\}$

where $\boldsymbol{z}_1^T(x) = (1, x, \ldots, x^p)$ and $\boldsymbol{z}_2^T(x) = (x^{p+1}, \ldots, x^q)$.

In the multiple regression case (P3) we have the true model

$$E(Y \mid \boldsymbol{x}) = \boldsymbol{z}^T(\boldsymbol{x})\boldsymbol{\theta}_0 + f(\boldsymbol{x}).$$

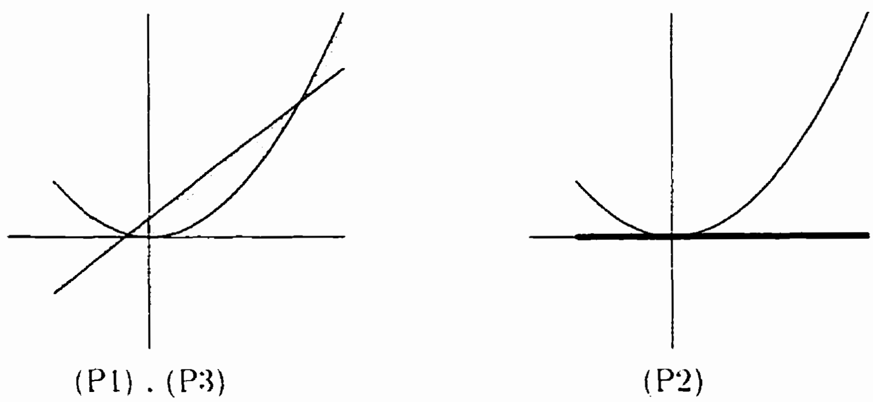
and contaminating space

$$\mathcal{F}_2 = \{\int f^2(\boldsymbol{x})\,d\boldsymbol{x} \le \eta^2, \int \boldsymbol{z}_1(\boldsymbol{x}) f(\boldsymbol{x})\,d\boldsymbol{x} = \boldsymbol{0}\}.$$

Here the bold $\boldsymbol{x}$ reminds us that the variable is multidimensional and the regressors are given, for example, by $\boldsymbol{z}^T(\boldsymbol{x}) = (1, x_1, x_2, x_1x_2)$ or
$\boldsymbol{z}^T(\boldsymbol{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$.

Let's explain the difference between problems (P1) and (P2), which on the surface look identical.

Figure 5.1: True model function $E(Y \mid x) = 0 + 0x + 1x^2$.



The difference is simply the way in which the function $E(Y \mid x)$ is decomposed into "lower order part" and "contamination part". In (P1), the true coefficient $\boldsymbol{\theta}_0$ and $\boldsymbol{u}$ are defined so that a fitted response function will be useful in predicting future values of y. On the other hand, in (P2), $\boldsymbol{\theta}_1$ and $\boldsymbol{\beta}$ are simply the coefficients of low powers and high powers of x, respectively. In this case the experimenter simply wants to estimate the original model function as closely as possible regardless of the contamination. We illustrate the difference with a simple (and artificial) example: Suppose that the true model function is $E(Y \mid x) = 0 + 0x + 1x^2$ and we want to fit a linear model. In (P1) the fitted line will follow the contamination term x^2. whereas in (P2) it deliberately ignores the contamination and follows the horizontal $0 + 0x$ as closely as possible.

We now summarize our main results and explain what can be extended from them. There are three main results:

(R1) The suprema of loss functions are expressed in terms of moments. Since there might be many design measures corresponding optimal set of moments. with help of Wald (1939). we looked only at discrete measures with the minimal number of support points.

For (P1) under the D-optimality criterion with $p = 1$. arbitrary q and small ν. any symmetric measure μ with $\mu_{2i} = m_{2i}$ is minimax. where $i = 1, \ldots, [(q+1)/2]$ and m is Wiens's measure as found in (2.5.2). We observe that any optimal measure tends to Wiens's measure as $q \to \infty$.

(R2) Independently of Wald. we have shown for (P1) and (P2) with arbitrary p and q. that for any symmetric optimal measure μ not supported by p or fewer points. there exists an optimal measure ξ with $p + 1$ support points that shares the first $2p + 1$ moments of μ.

(R3) For (P3) under a restricted class of densities. we obtained continuous optimal designs for multiple regression with interaction terms.

We illustrate how the results can applied to our regression problems. Suppose that an experimenter fits a linear regression model although the true model might be cubic, i.e., $p = 1$, $q = 3$. Define $g : \Xi \to I\!R \cup \{\infty\}$ by $g(\xi) = \frac{1}{\xi_2}[\nu + \max\{\frac{5}{4}(12\xi_2 - 1)^2, \frac{7}{\xi_2}(20\xi_4 - 3\xi_2)^2\}]$ if $\xi_2 \neq 0$ and $g(\xi) = \infty$ otherwise. As we have shown in Lemma 2.2.1, the function g is continuous on the compact set Ξ, and hence there exists an optimal measure μ^*, say, on Ξ. Applying (R2), there is an optimal design measure ξ with 3 support points, $\xi = (1-\alpha)\delta_0 + (\alpha/2)\delta_{\pm x_1}$. The design point x_1 and its mass α are chosen to minimize g. For instance, when $\nu = 1$, one choice is $\alpha = 0.8636$, and $x_1 = 0.3598$. On the other hand, from (R1), we know that Wiens's measure m is also optimal: it has density $m(x) = 1 + \frac{5}{4}(12\sqrt{(4\nu+5)/720} - 1)(12x^2 - 1)$ on $[-1/2, 1/2]$. By Theorem 3.3.2 and Theorem 3.3.3, there is a measure of the form $\xi = (\alpha/2)\delta_{-x_1} + (1-\alpha)\delta_0 + (\alpha/2)\delta_{x_1}$, so that $\xi_i = m_i$ for $i = 0, 1, \ldots, 5$ and thus ξ is also optimal. Solving for x_1 and α gives $x_1 = \sqrt{m_4/m_2}$, and $\alpha = \sqrt{(4\nu+5)/(720x_1^4)}$. When $\nu = 1$, since $m_2 = 1/\sqrt{80}$, and $m_4 = 0.0186$ and so this gives $x_1 = 0.4079$ and $\alpha = 0.672$. So we have two distinct optimal measures with three support points.

In the first two results, the number of support points of these optimal designs is too low to provide an opportunity to assess the higher order model. We present an example of how we might overcome this obstacle.

Estimation in higher order models

Using Wiens's optimal measure, the number of support points might be added so that we estimate the coefficients in higher order models. For instance, when

$p = 1$ and $q = 3$. we have seen an optimal measure with three support points. Since there are four coefficients in the model. we need a minimum of four support points in our design to estimate them all. The error variance is estimated by replicating the observations.

Now. because Wiens's measure m is absolutely continuous. Theorem 3.3.2 tells us that there exists a unique measure with 4 support points with same first seven moments as Wiens's. This measure is necessarily optimal as well. It has the form

$$\xi = [(1-\alpha)\delta_{-x_2} + \alpha\delta_{-x_1} + \alpha\delta_{x_1} + (1-\alpha)\delta_{x_2}]/2.$$

where α. x_1. and x_2 are determined by three equations. $\xi_2 = \sqrt{(4\nu+5)/720}$. $\xi_4 = (3/14)\xi_2 - 3/560$. and $\xi_6 = (5/112)\xi_2 - 1/672$.

Lack of Fit (LOF) test for (P1)

The LOF test can be performed by realizing that Theorem 2.5.1 also proves that any optimal design measure has the same second moment as Wiens's. The reason for this is following. We know from Section 2.5.2. that $\lambda_1(\xi) = \sum_{i=2}^{q}(E_\xi[l_i])^2 \geq (5/4)(12\xi_2 - 1)^2$. Let $h(\xi_2) = (1/\xi_2)[1 + (5/4\nu)(12\xi_2 - 1)^2]$. We observe that h is strictly convex in ξ_2. and so h has a unique minimum. We also know that $\sup_{\mathcal{F}} \mathcal{L}_D(\xi) \geq h(\xi_2)$ and we found ξ^* in Theorem 2.5.1 so that

$$\sup_{\mathcal{F}} \mathcal{L}_D(\xi^*) = \min_{\xi_2} h(\xi_2) = h(\xi_2^*).$$

If another $\dot{\xi}$ minimizes $\sup_{\mathcal{F}} \mathcal{L}_D(\xi)$ then

$$h(\xi_2^*) = \sup_{\mathcal{F}} \mathcal{L}_D(\xi^*) = \sup_{\mathcal{F}} \mathcal{L}_D(\dot{\xi}) \geq h(\dot{\xi}_2) \geq h(\xi_2^*).$$

Since it is a unique minimum, $\hat{\xi}_2 = \xi_2^*$.

We point out that higher moments may not be the same as Wiens's. This enables us to carry out the LOF test.

We demonstrate how this claim can be used in the LOF test for fitting a straight line whereas true response might be quadratic. In (P1) this means $\boldsymbol{z}_1^T(x) = (1, x)$ and $\boldsymbol{u}^T(x)\boldsymbol{\beta} = (x^2 - 1/12)\boldsymbol{\beta}$. We assume that the errors are normally distributed. Using the equation (2.2) in Wiens (1991) we obtain the non-centrality parameter in the test of LOF, $\mathcal{P}(\boldsymbol{\beta}, \xi) := (\xi_4 - \xi_2^2)\boldsymbol{\beta}^2$. We now want to find ξ^* such that

$$\min_{\{\beta^2=1\}} \mathcal{P}(\beta, \xi^*) = \max_{\Xi'} \min_{\{\beta^2=1\}} \mathcal{P}(\beta, \xi). \tag{5.0.1}$$

where $\Xi' = \{\xi \in \Xi : \ \xi_2 \text{ minimizes } \sup_{\mathcal{F}} \mathcal{L}_D(\xi)\}$. This is equivalent to find ξ^* maximizing ξ_4 over $\Xi' = \{\xi \in \Xi : \xi_2 = \sqrt{(4\nu + 5)/720}\}$. Then the solution for (5.0.1) is of the form $\xi^* = (1 - \alpha)\delta_0 + (\alpha/2)\delta_{\pm 1/2}$, where $\alpha = 4\xi_2^*$. This is so because $\xi_4^* = \alpha^*/16 = \xi_2^*/4$ is the maximum.

Applications of (R3) to Growth Models

We provide guidelines on how the optimal designs for an approximately polynomial with interaction terms might be adopted in other areas of science.

(1) Yield-Density Curves (see Seber and Wild (1989))

In agriculture, several models are used for quantifying the relationship between the density of crop planting and crop yield. The common Yield-Density models are

$$\text{Shinozaki and Kira:} \quad E(Y \mid x) = (\beta_0 + \beta_1 x)^{-1} \tag{5.0.2}$$

$$\text{Holliday:} \quad E(Y \mid x) = (\beta_0 + \beta_1 x + \beta_2 x^2)^{-1}. \tag{5.0.3}$$

where x and w denote the density of planting and the yield per unit area and $y := w/x$ represents the average yield per plant if all plants survived. It seems to be quite usual that the agronomist, after collecting his data, fits the models above to predict the yield or to find the density of planting maximizes the yield. But the main drawbacks in this situation are that first the experimenter can never be sure the model used is correct and secondly, for the models above, $E(Y \mid x)$ decreases as x increases when x is small, however in reality this decrease will not occur. We might be able to overcome these difficulties by looking at this as a design problem as follows. Suppose the density of planting x lies between $[a - b/2, a + b/2]$, for $a, b \in I\!R$. The nonlinear model (5.0.2) can be linearized by a Taylor series expansion and the change of variables $x' = (x - a)/b$.

$$\begin{aligned} E(Y \mid x) &= (\beta_0 + \beta_1 a)^{-1} - (\beta_0 + \beta_1 a)^{-2}(x - a) + f(x) \quad (5.0.4) \\ &= \theta_0 + \theta_1 x' + f(x'), \quad x' \in [-1/2, 1/2]. \quad (5.0.5) \end{aligned}$$

Similarly, the Holliday model transforms to $E(Y \mid x') = \theta_0 + \theta_1 x' + \theta_2 x'^2 + f(x')$, where $\theta_0 = (\beta_0 + \beta_1 a + \beta_2 a^2)^{-1}, \theta_1 = -b(\beta_0 + \beta_1 a + \beta_2 a^2)^{-2}$ and $\theta_2 = b^2(\beta_0 + \beta_1 a + \beta_2 a^2)^{-3}$. Before planting the crop, by applying the algorithm in Section 4.1 and the technique of implementing the densities in Section 4.4, the agronomist may choose the optimal x_i's so that the estimates for θ's are the most efficient.

Crop yield is not only affected by the density of planting but also the shape of the area available to each plant. A bivariate model is recommended considered for this case. $E(Y \mid x_1, x_2) = [\beta_0 + \beta_1(1/x_1 + 1/x_2) + \beta_2/(x_1 x_2)]^{-1/\gamma}$, where x_1 is the spacing between plants within a row and x_2 is the distance between the rows. Using a similar idea to the one above, the optimal design for bivariate regression model with interaction terms can be chosen by the experimenter.

We close this section by mentioning applications on growth models in Forest Science.

(2) Growth Models

Predicting total tree height based on observed diameter at breast height outside bark is routinely required in practical management and silvicultural work (see page 2 of Huang, Titus, and Wiens (1991)). Many nonlinear height-diameter functions are available to predict height growth. The Chapman-Richards function has been used extensively in describing the height-diameter as well as a base function for developing more complicated models (Huang and Titus (1994)). These functions are given by the expression

$$E(Y \mid x) = 1.3 + \beta_0(1 - e^{-\beta_1 x})^{\beta_2}, \tag{5.0.6}$$

where y is the total tree height (meters), x is the diameter (meters) of the tree at breast height. The technique described in (1) above might be applied when $\beta_2 \geq 1$.

Reference

Atkinson, A. C. (1982). "Developments in the design of experiments." *International Statistical Review* **50**, 161–177.

Atwood, C. L. (1971). "Robust procedures for estimating polynomial regression." *Journal of the American Statistical Association* **66**, 855–860.

Billingsley, P. (1986). *Probability and Measure*. Wiley, New York.

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.

Cook, R. D. and Nachtsheim, C. J. (1982). "Model robust, linear-optimal designs." *Technometrics* **24**, 49–54.

Dette, H. (1990). "A generalization of D- and D_1-optimal designs in polynomial regression." *Annals of Statistics* **18**, 1784–1804.

Dette, H. (1991). "A note on robust designs for polynomial regression." *Journal of Statistical Planning and Inference* **28**, 223–232.

Dette, H. (1992). "Optimal designs for a class of polynomials of odd or even degree." *Annals of Statistics* **20**, 238–259.

Dette, H. (1993). "On a mixture of the D- and D_1-optimality criterion in polynomial regression." *Journal of Statistical Planning and Inference* **35**, 233–249.

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

Heo, G. and James, W. (1995). "The effects of Ozone Application Rate and Gas:Liquid Ratio on removing fine suspended particles in a bubble column reactor." Unpublished.

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

Huang, S. and Titus, S. J. (1994). "An age-independent individual tree height prediction model for boreal spruce-aspen stands in Alberta." *Canadian Journal of Forest Research* **24**, 1295–1301.

Huang, S., Titus, S. J., and Wiens, D. P. (1991). "Comparison of nonlinear height-diameter function for major Alberta species." *The University of Alberta Technical Report 91.15.*

Huber, P. J. (1975). "Robustness and designs." in: J. N. Srivastava, ed. *A Survey of Statistical Design and Linear Models*, North-Holland, Amsterdam, 287–303.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Karlin, S. and Shapley, L. S. (1953). "Geometry of moment spaces." *Memoirs of the American Mathematical Society*, 12.

Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*, Interscience, New York.

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

Lau, T.-S. and Studden, W. J. (1985). "Optimal designs for trigonometric and polynomial regression using canonical moments." *Annals of Statistics* **13**, 383–394.

Läuter, E. (1974). "Experimental design in a class of models." *Mathematische Operationsforschung und Statistik* **5**, 379–398.

Läuter, E. (1976). "Optimal multipurpose designs for regression models." *Mathematische Operationsforschung und Statistik* **7**, 51–68.

Li, K.-C. (1984). "Robust regression designs when the design space consists of finitely many points." *Annals of Statistics* **12**, 269–282.

Li, K.-C. and Notz, W. (1982). "Robust designs for nearly linear regression." *Journal of Statistical Planning and Inference* **6**, 135–151.

Liu, S. X. and Wiens, D. P. (1997). "Robust designs for approximately polynomial regression." *Journal of Statistical Planning and Inference* **64**, 369–381.

Marcus, M. B. and Sacks, J. (1976). "Robust designs for regression problems." in: S. S. Gupta and D. S. Moore, eds. *Statistical Decision Theory and Related Topics II*. Academic Press, New York, 245–268.

Pesotchinsky, L. (1982). "Optimal robust designs: Linear regression in $\mathbb{R}^k$." *Annals of Statistics* **10**, 511–525.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Rychlik, T. (1987). "Robust experimental design: A comment on Huber's result." *Zastosowania Matematyki Applicationes Mathematicae* **19**, 93–107.

Sacks, J. and Ylvisaker, D. (1978). "Linear estimation for approximately linear models." *Annals of Statistics* **6**, 1122–1137.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Skibinsky, M. (1968). "Extreme nth moments for distributions on $[0,1]$ and the inverse of a moment space map." *Journal of Applied Probability* **5**, 693–701.

Stigler, S. (1971). "Optimal experimental design for polynomial regression." *Journal of the American Statistical Association* **66**, 311–318.

Studden, W. J. (1982). "Some robust-type D-optimal designs in polynomial regression." *Journal of the American Statistical Association* **77**, 916–921.

Thisted, R. A. (1988). *Elements of Statistical Computing*. Chapman and Hall.

Wald, A. (1939). "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments." *Transactions of the American Mathematical Society* **46**, 280–306.

Wiens, D. P. (1990). "Robust minimax designs for multiple linear regression." *Linear Algebra and its Applications* **127**, 327–340.

Wiens, D. P. (1991). "Designs for approximately linear regression: Two optimality properties of uniform designs." *Statistics & Probability Letters* **12**, 217-221.

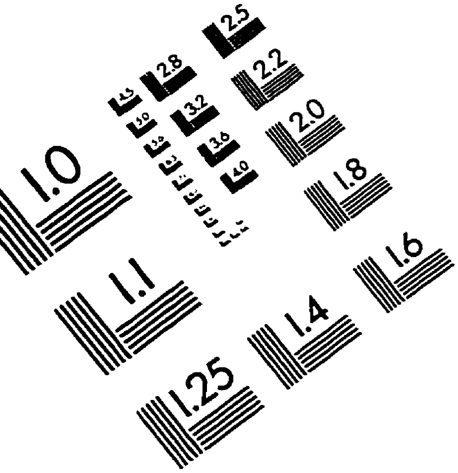
Wiens, D. P. (1992). "Minimax designs for approximately linear regression." *Journal of Statistical Planning and Inference* **31**, 353-371.

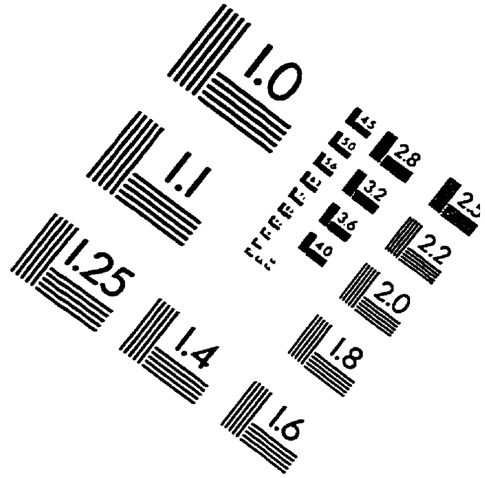
Wiens, D. P. (1993). "Designs for approximately linear regression: Maximizing the minimum coverage probability of confidence ellipsoids." *Canadian Journal of Statistics* **21**, 59-70.

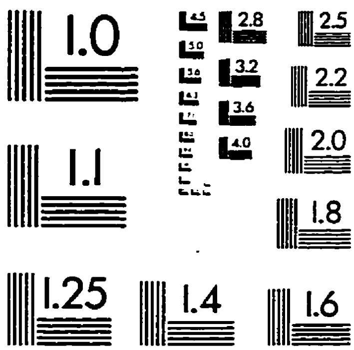
Wiens, D. P. (1994). "Robust designs for approximately linear regression: M-estimated parameters." *Journal of Statistical Planning and Inference* **40**, 135-160.

Wiens, D. P. (1996). "Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors." *The University of Alberta Technical Report 96.01*.

IMAGE EVALUATION TEST TARGET (QA-3)

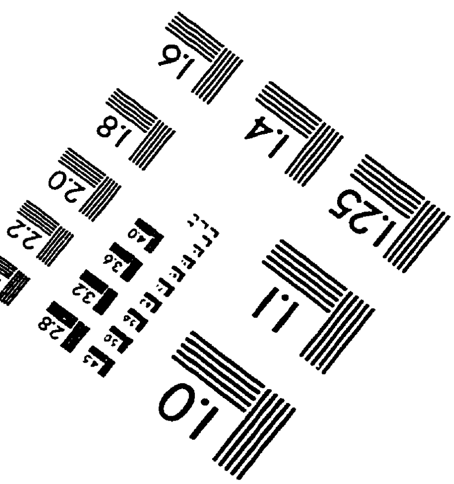


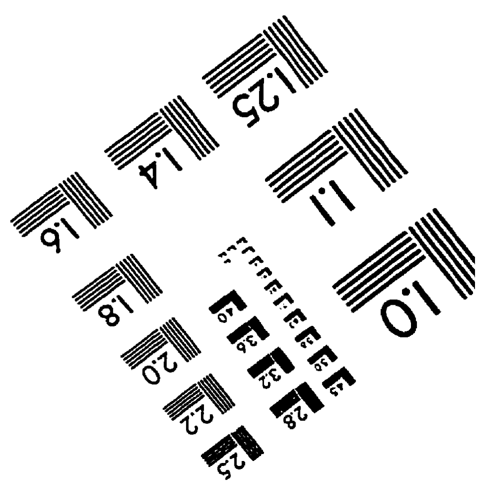




150mm

6"





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989