Chapter 2
Optimal Regression Designs

Abstract In this chapter, we review the theory of optimum regression designs. Concept of continuous design and different optimality criteria are introduced. The role of *de la Garza phenomenon* and *Loewner order domination* are discussed. Equivalence theorems for different optimality criteria, which play an important role in checking the optimality of a given otherwise prospective design, are presented. These results are repeatedly used in later chapters in the search for optimal mixture designs. We present standard optimality results for single variable polynomial regression model and multivariate linear and quadratic regression model. Kronecker product representation of the model(s) and related optimality results are also discussed.

Keywords Continuous design · Optimality criteria · de la Garza phenomenon · Loewner order domination · Polynomial regression models · Equivalence theorem · Optimum regression designs

2.1 Introduction

In this chapter, we will discuss optimality aspects of regression designs in an *approximate* (or, *continuous*) design setting defined below.

Let y be the observed response at a point $(x_1, x_2, \ldots, x_k) = \boldsymbol{x}$ varying in some k-dimensional experimental domain $\mathcal{X}$ following the general linear model

$$y(\boldsymbol{x}) = \eta(\boldsymbol{x}, \boldsymbol{\beta}) + e(\boldsymbol{x}), \qquad (2.1.1)$$

with usual assumptions on error component $e(\boldsymbol{x})$, viz. mean zero and uncorrelated homoscedastic variance σ^2; $\eta(\boldsymbol{x}, \boldsymbol{\beta})$ is the mean response function involving k or more unknown parameters. Once for all, we mention that $\boldsymbol{x}$ will represent a combination of the mixing components, the number of such components will be understood

from the context. Moreover, the same will be used to denote a row or a column vector, as the context demands.

Generally, it is assumed that in the region of immediate interest, $\eta(\boldsymbol{x}, \boldsymbol{\beta})$ can be approximated by a polynomial of certain degree and can be expressed as

$$\eta(\boldsymbol{x}, \boldsymbol{\beta}) = f'(\boldsymbol{x})\boldsymbol{\beta}. \tag{2.1.2}$$

The discrete or, exact designing problem is that of choosing N design points in the experimental domain $\mathcal{X}$ so that individually each of the t parameters of the mean response function can be estimated with satisfactory degree of accuracy. A *continuous* or an *approximate design* ξ for model (2.1.2), as introduced by Kiefer (1959), consists of finitely many distinct *support points* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, at which observations of the response are to be taken, and of corresponding design weights $\xi(\boldsymbol{x}_i) = p_i, i = 1, 2, \ldots, n$ which are positive real numbers summing up to 1. In other words, an approximate design ξ is a probability distribution with finite support on the factor space $\mathcal{X}$ and is represented by

$$\xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n; p_1, p_2, \ldots, p_n\}, \tag{2.1.3}$$

which assigns, respectively, masses $p_1, p_2, \ldots, p_n; p_i > 0, \sum p_i = 1$, to the n distinct support points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ of the design ξ in the experimental region [may be a subspace of the factor space $\mathcal{X}$]. Let $\mathcal{D}$ be the class of all competing designs. For a given N, a design ξ cannot, in general, be properly realized, unless its weights are integer multiples of $1/N$ i.e., unless $n_i = Np_i, i = 1, 2, \ldots, n$ are integers with $\sum n_i = N$. An approximate design becomes an *exact design* of size N in the special case when $n_i = Np_i, i = 1, 2, \ldots, n$ are integers.

The *information matrix* (often termed 'per observation moment matrix') for β, using a design ξ, is given by

$$M(\xi) = \sum p_i f(\boldsymbol{x}_i) f'(\boldsymbol{x}_i). \tag{2.1.4}$$

It may be noted that for unbiased estimation of the parameters in the mean response function, it is necessary that the number of 'support points' i.e., $\boldsymbol{x}_i$s must be at least 't,' the number of model parameters. It is tacitly assumed that the choice of a design leads to unbiased estimation of the parameters and as such the information matrix M of order $t \times t$ is a positive definite matrix. Let $\mathcal{M}$ denote the class of all positive definite moment matrices. As we will see in the rest of this monograph, the information matrix (2.1.4) of a design plays an important role in the determination of an optimum design. In fact, most of the optimality criteria are different functions of the information matrix.

2.2 Optimality Criteria

The utility of an optimum experimental design lies in the fact that it provides a design ξ^* that is best in some sense. Toward this let us bring in the concept of *Loewner ordering*. A design ξ_1 dominates another design ξ_2 in the Loewner sense if $M(\xi_1) - M(\xi_2)$ is a nonnegative definite (nnd) matrix. Thus, Loewner partial ordering among information matrices induces a partial ordering among the associated designs. We shall denote $\xi_1 \succ \xi_2$ when ξ_1 dominates ξ_2 in the Loewner sense. A design ξ^* that dominates over all other designs in $\mathcal{D}$ in the Loewner sense is called *Loewner optimal*. In general, there exists no Loewner optimal design ξ^* that dominates every other design ξ in $\mathcal{D}$ [vide Pukelsheim (1993)]. A popular way out is to specify an optimality criterion, defined as a real-valued function of $M(\xi)$. An optimal design is one whose moment matrix *minimizes* the criterion function $\phi(\xi)$ over a well-defined set of competing moment matrices (or designs); vide Shah and Sinha (1989) and Pukelsheim (1993) for details. Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_t$ be the t positive characteristic roots of the moment matrix M. It is essential that a reasonable criterion ϕ conforms to the Loewner ordering

$$M(\xi_1) \geq M(\xi_2) \Rightarrow \phi(M(\xi_1)) \geq \phi(M(\xi_2)).$$

The first original contribution on optimum regression design is by Smith (1918) who determined the G-optimum design for the estimation of parameters of a univariate polynomial response function. After a gap of almost four decades, a number of contributions in this area were made by, Elfving (1952), Chernoff (1953), Ehrenfield (1955), Guest (1958), Hoel (1958). Extending their results, Kiefer (1958, 1959, 1961), Kiefer and Wolfowitz (1959) developed a systematic account of different optimality criteria and related designs. These can be discussed in terms of maximizing the function $\phi(M(\xi))$ of $M(\xi)$.

The most prominent optimality criteria are the *matrix means* ϕ_p, for $p \in (-\infty, 2]$, which enjoy many desirable properties. These were introduced by Kiefer (1974, 1975):

$$\phi_p(M) = \left(\left(\frac{1}{t} \right) \sum_i \lambda_i^p \right)^{1/p}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_t$ denote the eigenvalues of the positive definite information matrix $M(\xi)$ of order $t \times t$. Excluding trivial cases, it is evident that an optimum design which satisfies all the criteria does not exist. The classical *A-, D- and E-optimality criteria* are special cases of ϕ_p-optimality criteria. The criterion $\phi_{-1}(M)$ is the A-optimality criterion. Maximizing $\phi_{-1}(M)$ is equivalent to minimizing the trace of the corresponding dispersion matrix (in the exact or asymptotic sense). The D-optimality criterion $\phi_0(M)$ is equivalent to maximizing the determinant det.(M). The extreme member of $\phi_p(M)$ for $p \to -\infty$ yields the smallest eigen (E-optimality criterion) $\phi_{-\infty}(M) = \lambda_{\min}(M)$. Besides, there are other optimality criteria for comparing designs viz. G-optimality criterion, D_s- and D_A-optimality

criteria, I-optimality criterion etc. [For different optimality criteria and their statistical significance, the readers are referred to Fedorov (1972), Silvey (1980), Shah and Sinha (1989), Pukelsheim (1993)].

In general, the direct search for optimum design may be prohibitive. The degree of difficulty depends on the nature of response function, criterion function and/or the experimental region. However, there are tools that can be used to reduce, sometimes substantially, the class of competing designs.

2.3 One Dimensional Polynomial Regression

In practice, polynomial models are widely used because they are flexible and usually provide a reasonable approximation to the true relationship among the variables. Polynomial models with low order, whenever possible, are generally recommended. Higher order polynomial may provide a better fit to the data and hence an improved approximation to the true relationship; but the numerous coefficients in such models make them difficult to interpret. Sometimes, polynomial models are used after an appropriate transformation has been applied on the independent variables to lessen the degree of nonlinearity. Examples of such transformations are the logarithm and square transformations. Box and Cox (1964), Carroll and Ruppert (1984) gave a detailed discussion on the use and properties of various transformations for improving fit in linear regression models.

In a one-dimensional polynomial regression, the mean response function is given by

$$\eta(x) = f'(x)\boldsymbol{\beta} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d, \qquad (2.3.1)$$

where $f'(x) = (1, x, x^2, \ldots, x^d)$ and $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_d)$. Several authors attempted to find optimum designs for the estimation of parameters of the above model. As mentioned earlier, Smith (1918) first obtained G-optimum designs for the estimation of parameters. de la Garza (1954) considered the estimation of parameters of the above model from N observations in a given range. By changing the origin and scale, the domain of experimental region, i.e., the factor space may be taken to be $\mathcal{X} = [-1, 1]$. Consider a design ξ given by (2.1.3) in the factor space $[-1, 1]$ with information matrix (2.1.4). de la Garza (1954) showed that corresponding to any arbitrary continuous design ξ as in (2.1.3) supported by n [$>d + 1$] distinct points, there exists a design with exactly $d + 1$ support points such as

$$\xi^* = \{x_1^*, x_2^*, \ldots, x_{d+1}^*; p_1^*, p_2^*, \ldots, p_{d+1}^*\}, \qquad (2.3.2)$$

for which the information matrices are the same, i.e, $M(\xi) = M(\xi^*)$. Moreover, $x_{\min} \leq x_{\min}^* \leq x_{\max}^* \leq x_{\max}$. This appealing feature of the two designs is referred to as *information equivalence*. Afterward, the de la Garza phenomenon has been extensively studied by Liski et al. (2002), Dette and Melas (2011), Yang (2010).

In addition to this, Pukelsheim (1993) extensively studied the phenomenon of *information domination* in this context.

Remark 2.3.1 The exact design analog of the feature of information equivalence is generally hard to realize. Mandal et al. (2014) provide some initial results in this direction.

In general, different optimal designs may require different number of design points. It is clear that in order to estimate '*t*' parameters in any model, at least '*t*' distinct design points are needed, and for many models and optimality criteria, the optimal number of distinct design points will be '*t*.' For nonlinear models, the information matrix has an interpretation in an approximate sense, as being the inverse of the asymptotic variance-covariance matrix of the estimates of the model parameters. An interesting result called *Caratheodory's Theorem* provides us with an upper bound on the number of design points needed for the existence of a positive definite information matrix. For many design problems with '*t*' parameters, this number is '*t*(*t* + 1)/2.' Thus the optimal number of distinct design points is between '*t*' and '*t*(*t* + 1)/2.' Finally, it should be noted that the upper bound does not hold for the *Bayesian* design criteria (Atkinson et al. 2007, Chap. 18).

Guest (1958) obtained general formulae for the distribution of the points of observations and for the variances of the fitted values in the minimax variance case, and compared the variances with those for the uniform spacing case. He showed that the values of $x_1, x_2, \ldots, x_{d+1}$, (with reference to the model (2.3.1)) that minimize the maximum variance of a single estimated ordinate are given by means of the zeros of the derivative of a Legendre polynomial. Hoel (1958) used the *D*-optimality criterion for determining the best choice of fixed variable values within an interval for estimating the coefficients of a polynomial regression curve of given degree for the classical regression model. Using the same criterion, some results are obtained on the increased efficiency arising from doubling the number of equally spaced observation points (i) when the total interval is fixed and (ii) when the total interval is doubled. Measures of the increased efficiency are found for the classical regression model and for models based on a particular stationary stochastic process and a pure birth stochastic process. Moreover, he first noticed that *D*- and *G*-optimum designs *coincide* in a one-dimensional polynomial regression model.

Kiefer and Wolfowitz (1960) extended and established this phenomenon of coincidence to any linear model through what is now known as 'Equivalence Theorem.' Writing $d(\boldsymbol{x}, \xi) = f'(\boldsymbol{x})M^{-1}(\boldsymbol{\beta}, \xi)f(\boldsymbol{x})$, the celebrated equivalence theorem of Kiefer and Wolfowitz (1960) can be stated as follows:

Theorem 2.3.1 *The following assertions:*

(i) the design ξ^* minimizes $\mid M^{-1}(\boldsymbol{\beta}, \xi) \mid$,
(ii) the design ξ^* minimizes $\max_{\boldsymbol{x}} d(\boldsymbol{x}, \xi)$,
(iii) $\max_{\boldsymbol{x}} d(\boldsymbol{x}, \xi^*) = t$

are equivalent. The information matrices of all designs satisfying (i)–(iii) coincide among themselves. Any linear combination of designs satisfying (i)–(iii) also satisfies (i)–(iii).

In this context, Fedorov (1972) also serves as a useful reference. This theorem plays an important role in establishing the D-optimality of a given design obtained from intuition or otherwise. Moreover, it gives the nature of the support points of an optimum design. For example, let us consider a quadratic regression given by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i, i = 1, 2, \ldots, n \qquad (2.3.3)$$

where β_i s are fixed regression parameters and e_i s are independent random error with usual assumptions, viz. mean 0 and variance σ^2. We assume, as before, that the factor space is $\mathcal{X} = [-1, 1]$.

The information matrix for an arbitrary design $\xi = \{x_1, x_2, \ldots, x_n; p_1, p_2, \ldots, p_n\}$ can be readily written down, and it involves the moments of x-distribution, i.e., $\mu'_r = \sum p_i x_i^r; r = 1, 2, 3, 4$. It is well known that the information matrix M is positive definite iff $n > 2$, since the x_i s are assumed to be all distinct without any loss of generality and essentially we are restricting to this class of designs. It is clearly seen that $d(x, \xi) = f'(x) M^{-1}(\boldsymbol{\beta}, \xi) f(x)$, with $f(x) = (1, x, x^2)'$ is *quartic in x* so that the three support points of the D-optimum design are at the two extreme points ± 1 and a point lying in between. Since the D-optimality criterion, for the present problem, is invariant with respect to sign changes, the interior support point must be at 0. Thus, the three support points of the D-optimum design in $[-1, 1]$ are 0 and ± 1. The weights at the support points are all equal since here the number of support points equals the number of parameters. It may be noted that for D-optimality, whenever the number of support points is equal to the number of parameters, the weights at the support points are necessarily equal.

Remark 2.3.2 The above result can be derived using altogether different arguments. In view of de la Garza phenomenon, given the design ξ with $n > 3$, there exists a three-point design $\xi^* = \{(a, P), (b, Q), (c, R)\}$, where $-1 \leq a < b < c \leq 1$, and $0 < P, Q, R < 1$, $P + Q + R = 1$, such that $M(\xi) = M(\xi^*)$. Again, referring to Liski et al. (2002), we may further 'improve' ξ^* to $\xi^{**} = \{(-1, P), (c, Q), (1, P)\}$ by proper choice of 'c' in the sense of Loewner Domination. It now follows that for D-optimality, det.$M(\xi^{**}) \leq (4/27)(1 - c^2)^2 \leq 4/27$ with '=' if and only if $c = 0$ and $P = Q = R = 1/3$.

Atwood (1969) observed that in several classes of problems an optimal design for estimating all the parameters is supported only on certain points of symmetry. Moreover he considered the optimality when nuisance parameters are present and obtained a new sufficient condition for optimality. He corrected a version of the condition which Karlin and Studden (1966) stated as equivalent to optimality, and proved the natural invariance theorem involving this condition. He applied these results to the problem of multi-linear regression on the simplex (introduced by Scheffé 1958) when estimating all or only some of the parameters. This will be discussed in detail in Chaps. 4–6.

Fedorov (1971, 1972) developed the equivalence theorem for *Linear optimality* criterion in the lines of equivalence theorem of Kiefer and Wolfowitz (1960) for

D-optimality. Assuming estimability of the model parameters, we denote the dispersion matrix by $D(\xi)$. It is known that $D(\xi) = M^{-1}(\xi)$. A design ξ^* is said to be linear optimal if it minimizes $L(D(\xi))$ over all ξ in Ξ where L is a linear optimality functional defined on the dispersion matrices satisfying

$$L(A + B) = L(A) + L(B)$$

for any two nnd matrices A and B and

$$L(cA) = cL(A)$$

for any scalar $c > 0$. Then, the equivalence theorem for linear optimality can be stated as follows.

Theorem 2.3.2 *The following assertions:*

   (i) *the design ξ^* minimizes $L[D(\xi)]$,*
   (ii) *the design ξ^* minimizes $\max_x L[D(\xi)f(x)f'(x)D(\xi)]$,*
  (iii) $\max_x L[D(\xi^*)f(x)f'(x)D(\xi^*)] = L[D(\xi^*)]$

are equivalent. Any linear combination of designs satisfying (i)–(iii) also satisfies (i)–(iii).

Similar equivalence theorems are also available for E-optimality criterion (cf. Pukelsheim 1993).

Afterword, Kiefer (1974) introduced the ϕ-optimality criterion, a real-valued concave function defined on a set $\mathcal{M}$ of positive definite matrices. He then established the following equivalence theorem [cf. Silvey (1980), Whittle (1973)].

Let $\mathcal{M}$ be the class of all moment matrices obtained by varying design ξ in Ξ and ϕ is a real-valued function defined on $\mathcal{M}$. Then the Frechét derivative of ϕ at M_1 in the direction of M_2 is defined as

$$F_\phi(M_1, M_2) = \lim_{\alpha \to 0+} \frac{1}{\alpha} [\phi\{(1 - \alpha)M_1 + \alpha M_2\} - \phi(M_1)].$$

Theorem 2.3.3 *When ϕ is concave on $\mathcal{M}$, ξ^* is ϕ-optimal if and only if*

$$F_\phi(M(\xi^*), M(\xi)) \leq 0 \qquad (2.3.4)$$

for all $\xi \in \mathcal{D}$.

This theorem states simply that we are at the top of a *concave mountain* when there is no direction in which we can look forward to another point on the mountain. However, since it is difficult to check (2.3.4) for all $\xi \in \mathcal{D}$, Kiefer (1974) established the following theorem that is more useful in verifying the optimality or non-optimality of a design ξ^*.

Theorem 2.3.4 *When ϕ is concave on $\mathcal{M}$ and differentiable at $M(\xi^*)$, ξ^* is ϕ-optimal if and only if*

$$F_\phi\{M(\xi^*), f(\boldsymbol{x})f'(\boldsymbol{x})\} \leq 0 \qquad (2.3.5)$$

for all $\boldsymbol{x} \in \mathcal{X}$.

For the proof and other details, one can go through Kiefer (1974) and Silvey (1980). This result has great practical relevance because in many situations, the optimality problems may not be of the classical *A*-, *D*- or *E*-optimality type but fall under a wide class of ϕ-optimality criteria. The equivalence theorem above then helps us to establish the optimality of a design obtained intuitively or otherwise.

The equivalence theorem in some form or the other has been repeatedly used in subsequent chapters of this monograph. For the equivalence theorem for Loewner optimality or other specific optimality criteria, the readers are referred to Pukelsheim (1993).

Remark 2.3.3 An altogether different optimality criterion was suggested in Sinha (1970). Whereas all the traditional optimality criteria are exclusively functions of the (positive) eigenvalues of the information matrix, this one was an exception. In the late 1990s, there was a revival of research interest in this optimality criterion, termed as 'Distance Optimality criterion' or, simply, 'DS-optimality' criterion.

In the context of a very general linear model set-up involving a (sub)set of parameters θ admitting best linear unbiased estimator (blue) $\hat{\theta}$, it is desirable that the 'stochastic distance' between θ and $\hat{\theta}$ be the least. This is expressed by stating that the 'coverage probability'

$$\Pr[\|\hat{\theta} - \theta\| < \epsilon]$$

should be as high as possible, for every $\epsilon > 0$. As an optimal design criterion, therefore, we seek to characterize a design ξ_0 such that for every given ε, $\hat{\theta}$ based on ξ_0 provides largest coverage probability than any other competing ξ.

Sinha (1970) initiated study of DS-optimal designs for one-way and two-way analysis of variance (ANOVA) setup. Much later, the study was further continued in ANOVA and regression setup (Liski et al. 1998; Saharay and Bhandari 2003; Mandal et al. 2000). On the other hand, theoretical properties of this criterion function were studied in depth in a series of papers (Liski et al. 1999; Zaigraev and Liski 2001, 2006; Zaigraev 2005)

We will not pursue this criterion in the present Monograph.

Since in a mixture experiment, we will be concerned with a number of components, let us first review some results in the context of multi-factor experiment.

2.4 Multi-factor First Degree Polynomial Fit Models

Let us first consider a *k*-factor first degree polynomial fit model with no constant term, viz.,

$$y_{ij} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_{ij}, \tag{2.4.1}$$

with k regressor variables, n experimental conditions $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ik})$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, N_i$, $\sum N_i = N$. Most often we deal with a continuous or approximate theory version of the above formulation in which p_is are regarded as (positive) 'mass' attached to the points $\boldsymbol{x_i}$s, subject to the condition $\sum p_i = 1$.

In polynomial fit model with single factor, the experimental domain is generally taken as $\mathcal{X} = [-1, +1]$. For k-factor polynomial linear fit model (2.4.1), the experimental domain is a subset of the k-dimensional Euclidean space R^k. Generally, optimum designs are developed for the following two extensions of the one-dimensional domain $\mathcal{X} = [-1, +1]$: A *Euclidean ball* of radius $\sqrt{k}$ and a symmetric k-dimensional *hypercube* $[-1, +1]^k$. In practice, there may be other types of domains viz., a constrained region of the type $\mathcal{X}_R = [0 \le x_i \le 1, \sum x_i = \alpha \le 1]$. The mixture experiment, the optimality aspect of which will be considered in details in subsequent chapters, has domain that corresponds to $\alpha = 1$.

Below we develop the continuous design theory for the above model. Consider the experimental domain for the model (2.4.1), which is a k-dimensional ball of radius $\sqrt{k}$, that is, $\mathcal{X}(k) = [\boldsymbol{x} \in R_k, \|\boldsymbol{x}\| \le \sqrt{k}]$, where $\|.\|$ denotes the Euclidean norm. Set $\mu_{jm} = \sum_i p_i x_{ij} x_{im}$ for $j, m = 1, 2, \ldots, k$. This has the simple interpretation as the 'product moment' of jth and mth factors in the experiment. Then, the information matrix for an n-point $(n \ge k)$ design

$$\xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n; p_1, p_2, \ldots, p_n\}$$

is of the form

$$M(\xi) = \sum p_i f(\boldsymbol{x}_i) f'(\boldsymbol{x}_i) = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1k} \\ & \mu_{22} & \cdots & \mu_{2k} \\ & & \cdots & \cdots \\ & & & \mu_{kk} \end{pmatrix} \tag{2.4.2}$$

with $f'(\boldsymbol{x}_i) = (x_{i1}, x_{i2}, \ldots, x_{ik})$.

Using spectral decomposition of the matrix $M(\xi)$, it can be easily shown that $M(\xi)$ can equivalently be represented by a design ξ^* with k orthogonal support points in $\mathcal{X}(k)$:

$$\xi^* = \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_k^*; p_1^*, p_2^*, \ldots, p_k^*\} \tag{2.4.3}$$

i.e., $M(\xi^*) = M(\xi)$. Such a design is termed as orthogonal design (Liski et al. 2002). This incidentally demonstrates validity of the de la Garza phenomenon (DLG phenomenon) in the multivariate linear setup without the constant term. We can further improve over this design in terms of the Loewner order domination of the information matrix by stretching the mass at the boundary of $\mathcal{X}(k)$. In other words, given an orthogonal design ξ^* as in (2.4.3), there exists another k-point orthogonal design

$$\xi^{**} = \{\boldsymbol{x}_1^{**}, \boldsymbol{x}_2^{**}, \ldots, \boldsymbol{x}_k^{**}; p_1^*, p_2^*, \ldots, p_k^*\} \qquad (2.4.4)$$

with $\boldsymbol{x}_i^{**} = \sqrt{k}\boldsymbol{x}_i^*/\|\boldsymbol{x}_i^*\|$, such that $M(\xi^{**}) - M(\xi^*)$ is nnd i.e., $\xi^{**} \succ \xi^* \sim \xi$. One can now determine optimum designs in the class of designs (2.4.4) using different optimality criteria. Similar results hold for multi-factor linear model with constant term.

A symmetric k-dimensional unit cube $[-1, +1]^k$ is a natural extension of $[-1, +1]$. Note that $[-1, +1]^k$ is the convex hull of its extreme points, the 2^k vertices of $[-1, +1]^k$. It is known that in order to find optimal support points, we need to search the extreme points of the regression range only. If the support of a design contains other than extreme points, then it can be Loewner dominated by a design with extreme support points only. This result was basically presented by Elfving (1952, 1959). A unified general theory is given by Pukelsheim (1993).

A generalization of the model (2.4.1) incorporating the constant term has been studied in Liski et al. (2002). Also details for the latter factor space described above have been given there. We do not pursue these details here.

In the context of mixture models, as has been indicated before, we do not include a constant term in the mean model. So, the above study may have direct relevance to optimality issues in mixture models.

2.5 Multi-factor Second Degree Polynomial Fit Models

Consider now a second-degree polynomial model in k variables:

$$\eta_x = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \sum_{j>i}^k \beta_{ij} x_i x_j. \qquad (2.5.1)$$

For the second-degree model, in finding optimum designs, it is more convenient to work with the Kronecker product representation of the model (cf. Pukelsheim 1993).

For a k-factor second-degree model, $k \geq 2$, let us take the regression function to be

$$\eta(\boldsymbol{x}, \boldsymbol{\beta}) = g'(\boldsymbol{x})\boldsymbol{\beta} \qquad (2.5.2)$$

where

$$g'(\boldsymbol{x}) = (1, \boldsymbol{x}', \boldsymbol{x}' \otimes \boldsymbol{x}'), \qquad (2.5.3)$$

$\boldsymbol{\beta}$ is a vector of parameters and the factor space is given by

$$\mathcal{X}(k) = \{\boldsymbol{x} : \|\boldsymbol{x}\| \leq k\}.$$

To characterize the optimum design for the estimation of $\boldsymbol{\beta}$, let us consider the following designs:

$$\xi_0 = \{\boldsymbol{x} \mid \boldsymbol{x}'\boldsymbol{x} = 0\}$$

$$\xi_c = \frac{1}{2^r} \text{ fraction of a } 2^k \text{ factorial experiment with levels } \pm 1$$

$$\xi_s = \text{ set of star points of the form } (\pm\sqrt{k}, (0, 0, \ldots, 0), (0, \pm\sqrt{k}, 0, \ldots, 0), \ldots,$$
$$(0, 0, \ldots, \pm\sqrt{k})$$

$$\tilde{\xi}_{\sqrt{k}} = \frac{n_c\xi_c + n_s\xi_s}{n}, n_c = 2^{k-r}, n_s = 2k, n = n_c + n_s.$$

A design $\xi^* = (1 - \alpha)\xi_0 + \alpha \, \tilde{\xi}_{\sqrt{k}}$ is called a *central composite design* (CCD) (cf. Box and Wilson 1951). Such a design ξ^* is completely characterized by α. It is understood that $0 \le \alpha \le 1$.

Before citing any result on optimum design in the second-order case, let us first of all bring in the concept of Kiefer optimality. Symmetry and balance have always been a prime attribute of good experimental designs and comprise the first step of the Kiefer design ordering. The second step concerns the usual Loewner matrix ordering. In view of the symmetrization step, it suffices to search for improvement when the Loewner ordering is restricted only to exchangeable moment matrices.

Now, we cite a very powerful result on Kiefer optimality in the second-order model (2.5.1).

Theorem 2.5.1 *The class of CCD is complete in the sense that, given any design, there is always a CCD that is better in terms of*

(i) Kiefer ordering
(ii) ϕ-optimality, provided it is invariant with respect to orthogonal transformation.

There are many results for specific optimality criteria for the second-order model (see e.g., Pukelsheim 1993). We are not going to discuss the details.

It must be noted that in the context of mixture models, we drop the constant term β_0 from the mean model. Moreover, the factor space (constrained or not) is quite different from unit ball/unit cube. Yet, the approach indicated above has been found to be extremely useful in the characterization of optimal mixture designs. All these will be discussed in details from Chap. 4 onward.

References

Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs, in SAS*. Oxford: Oxford University Press.

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *The Annals of Mathematical Statistics*, *40*, 1570–1602.

Box, G. E. P., & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *13*, 1–38.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211–252.

Carroll, R. J., & Ruppert, D. (1984). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*, *79*, 321–615.

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

Dette, H., & Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimum designs. *The Annals of Statistics*, *39*, 1266–1281.

Ehrenfield, S. (1955). On the efficiency of experimental designs. *The Annals of Mathematical Statistics*, *26*, 247–255.

Elfving, G. (1959). Design of linear experiments. In U. Grenander (Ed.), Probability and statistics. The Herald cramer volume (pp. 58–74). Stockholm: Almquist and Wiksell.

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

Fedorov, V. V. (1971). Design of experiments for linear optimality criteria. *Theory of Probability & its Applications*, *16*, 189–195.

Fedorov, V. V. (1972). *Theory of optimal experiments*. New York: Academic.

de la Garza, A. (1954). Spacing of information in polynomial regression. *The Annals of Mathematical Statistics*, *25*, 123–130.

Guest, P. G. (1958). The spacing of observations in polynomial regression. *The Annals of Mathematical Statistics*, *29*, 294–299.

Hoel, P. G. (1958). The efficiency problem in polynomial regression. *The Annals of Mathematical Statistics*, *29*, 1134–1145.

Karlin, S., & Studden, W. J. (1966). Optimal experimental designs. *The Annals of Mathematical Statistics*, *37*, 783–815.

Kiefer, J. (1958). On non-randomized optimality and randomized non-optimality of symmetrical designs. *The Annals of Mathematical Statistics*, *29*, 675–699.

Kiefer, J. (1959). Optimum experimental designs. Journal of the Royal Statistical Society. *Series B (Methodological)*, *21*, 272–304.

Kiefer, J., & Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, *30*, 271–294.

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

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

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

Kiefer, J. (1975). Optimum design: Variation in structure and performance under change of criterion. *Biometrika*, *62*, 277–288.

Liski, E. P., Luoma, A., Mandal, N. K., & Sinha, B. K. (1998). Pitman nearness, distance criterion and optimal regression designs. *Bulletin of the Calcutta Statistical Association*, *48*, 179–194.

Liski, E. P., Luoma, A., & Zaigraev, A. (1999). Distance optimality design criterion in linear models. *Metrika*, *49*, 193–211.

Liski, E. P., Mandal, N. K., Shah, K. R., & Sinha, B. K. (2002). *Topics in optimal design* (Vol. 163). Lecture notes in statistics New York: Springer.

Mandal, N. K., Pal, M., & Sinha, B. K. (2014). Some finer aspects of the de la Garza phenomenon : A study of exact designs in linear and quadratic regression models. To appear in *Statistics and Applications* (special volume in the memory of Prof. M.N. Das).

Mandal, N. K., Shah, K. R., & Sinha, B. K. (2000). Comparison of test vs. control treatments using distance optimality criterion. *Metrika*, *52*, 147–162.

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

Saharay, R., & Bhandari, S. K. (2003). D_S-optimal designs in one-way ANOVA. *Metrika*, *57*, 115–125.

Scheffé, H. (1958). Experiments with mixtures. Journal of the Royal Statistical Society. *Series B (Methodological)*, *20*, 344–360.

Shah, K. R., & Sinha, B. K. (1989). *Theory of optimal designs* (Vol. 54). Lecture notes in statistics New York: Springer.

Silvey, S. D. (1980). *Optimal designs*. London: Chapman Hall.

Sinha, B. K. (1970). On the optimality of some designs. *Calcutta Statistical Association Bulletin*, *20*, 1–20.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, *12*, 1–85.

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

Yang, M. (2010). On the de la Garza phenomenon. *The Annals of Statistics*, *38*, 2499–2524.

Zaigraev, A., & Liski, E. P. (2001). A stochastic characterization of Loewner optimality design criterion in linear models. *Metrika*, *53*, 207–222.

Zaigraev, A. (2005). Stochastic design criteria in linear models. *Austrian Journal of Statistics*, *34*, 211–223.

Zaigraev, A., & Liski, E. P. (2006). On D_S-optimal design matrices with restrictions on rows or columns. *Metrika*, *64*, 181–189.