

# Design of Experiments

Stefanie Biedermann and Min Yang



# Contents

- 1 Designs for selected non-linear models** **1**
- 1.1 Introduction . . . . . 1
- 1.2 Classical methods . . . . . 4
  - 1.2.1 Methods based on the Equivalence Theorem . . . . . 6
  - 1.2.2 Geometric approach . . . . . 8
  - 1.2.3 Functional approach . . . . . 12
- 1.3 General solutions . . . . . 14
  - 1.3.1 Algebraic method . . . . . 15
  - 1.3.2 Method based on Chebyshev systems . . . . . 18
- 1.4 Further examples . . . . . 20
  - 1.4.1 The two-parameter exponential model . . . . . 20
  - 1.4.2 The Emax model . . . . . 23
  - 1.4.3 A heteroscedastic linear model . . . . . 26
- 1.5 Model discrimination for nonlinear models . . . . . 29

1.6	Parameter robust approaches . . . . .	31
1.7	Summary . . . . .	36

# Chapter 1

## Designs for selected non-linear models

### 1.1 Introduction

This chapter is an example based guide to optimal design for nonlinear regression models. For clarity, we restrict ourselves to models with only one continuous explanatory variable. The theory presented below also holds for multi variable models. In practice, designs for multi variable models are usually found numerically due to the increased complexity; see e.g. [64] or [60] for some recent developments on algorithms. Some analytical results on finding optimal designs for multi variable models can be found in [63] and [10] and references therein.

Throughout this chapter, we assume we can make  $n$  observations  $y_1, \dots, y_n$ , at experimental conditions  $x_1, \dots, x_n \in \mathcal{X}$ , from a nonlinear model, i.e.

$$y_i = \eta(x_i, \boldsymbol{\theta}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n, \quad (1.1)$$

where  $\eta(x_i, \boldsymbol{\theta})$  is the regression function, known up to the vector of  $m$  unknown parameters,  $\boldsymbol{\theta}$ , and the errors  $\epsilon_i$ ,  $i = 1, \dots, n$ , are independent and identically distributed. The design space  $\mathcal{X}$  is usually an interval on the real axis, i.e.  $\mathcal{X} \subseteq \mathbb{R}$ .

Suppose without loss of generality that  $x_1, \dots, x_t$ ,  $t \leq n$ , are the distinct points among

$x_1, \dots, x_n$ . We consider approximate designs of the form  $\xi = \{(x_1, w_1), \dots, (x_t, w_t)\}$ , where the weight  $w_i$  gives the proportion of observations to be made in the corresponding support point  $x_i$ ,  $i = 1, \dots, t$ . We thus require  $0 < w_i \leq 1$ , for  $i = 1, \dots, t$ , and  $\sum_{i=1}^t w_i = 1$ . Note that  $nw_i$  is not restricted to be an integer to avoid cumbersome discrete optimization problems. In order to run an approximate design in practice a rounding procedure, see e.g. [50], is used.

We are concerned with the “optimal” choice of a design. A decision rule of what is deemed optimal is provided by an optimality criterion, which is selected to reflect the purpose of the experiment. In what follows, we assume that we want to estimate the model parameters as accurately as possible, and the estimation is either through maximum likelihood or nonlinear least squares. It is therefore natural to consider optimality criteria which are concerned with minimizing the (asymptotic) covariance matrix of the estimator  $\hat{\boldsymbol{\theta}}$  or, equivalently, maximizing the Fisher information,  $\mathbf{M}_\xi$ , which for model (1.1) is given by

$$\mathbf{M}_\xi = \sum_{i=1}^t w_i \mathbf{f}(x_i, \boldsymbol{\theta}) \mathbf{f}'(x_i, \boldsymbol{\theta}),$$

where

$$\mathbf{f}(x, \boldsymbol{\theta}) = \left( \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \theta_m} \right)', \quad (1.2)$$

is the vector of partial derivatives of  $\eta(x, \boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$ .

**Example.** The Michaelis-Menten model has expected response

$$\eta(x, \boldsymbol{\theta}) = \frac{\theta_1 x}{\theta_2 + x},$$

which is depicted in Figure 1.1 for parameter values  $\boldsymbol{\theta}' = (1, 0.6)$ . The parameter  $\theta_1$  gives the supremum of the curve, whereas  $\theta_2$  can be interpreted as the value of  $x$ , at which half the supremum is attained.

Figure 1.1 approximately here

For the Michaelis-Menten model, the Fisher information of a design  $\xi$  is

$$\mathbf{M}_\xi = \sum_{i=1}^t w_i \frac{x_i^2}{(\theta_2 + x_i)^2} \begin{bmatrix} 1 & \frac{-\theta_1}{(\theta_2 + x_i)} \\ \frac{-\theta_1}{(\theta_2 + x_i)} & \frac{\theta_1^2}{(\theta_2 + x_i)^2} \end{bmatrix}.$$

There is no total ordering on the nonnegative definite matrices of size  $(m \times m)$ . Hence they are mapped to the real axis through an objective function to make them comparable. A popular class of optimality criteria are the  $\phi_p$ -criteria, which maximize the corresponding matrix means. The  $p$ th matrix mean,  $p \in [-\infty, 1]$ , is defined as

$$\phi_p(\xi) = \begin{cases} \left( \frac{1}{m} \text{trace} \mathbf{M}_\xi^p \right)^{1/p} & p \neq -\infty, 0 \\ |\mathbf{M}_\xi|^{1/m} & p = 0 \\ \lambda_{\min}(\mathbf{M}_\xi) & p = -\infty, \end{cases}$$

where  $\lambda_{\min}(\mathbf{M}_\xi)$  is the minimal eigenvalue of  $\mathbf{M}_\xi$ . Well-known special cases are the  $D$ -,  $A$ - and  $E$ -criteria, where  $p = 0, -1, -\infty$ , respectively. A  $D$ -optimal design minimizes the volume of an asymptotic confidence ellipsoid for  $\hat{\boldsymbol{\theta}}$ , and an  $A$ -optimal design minimizes the average of the asymptotic variances for the estimators of the individual parameters.

If interest is in estimating a linear combination of the parameters,  $\mathbf{c}'\boldsymbol{\theta}$ , for a given vector  $\mathbf{c}$ , we use the  $c$ -optimality criterion, which minimizes the objective function  $\phi_c(\xi) = \mathbf{c}'\mathbf{M}_\xi^- \mathbf{c}$ , where  $\mathbf{M}_\xi^-$  is a generalized inverse of the information matrix  $\mathbf{M}_\xi$ . This corresponds to minimizing the asymptotic variance of  $\mathbf{c}'\hat{\boldsymbol{\theta}}$ .

We note that in nonlinear models at least some of the partial derivatives, and thus the Fisher information, depend on the unknown parameter vector  $\boldsymbol{\theta}$ . An optimal design with respect to some optimality criterion will therefore only be optimal for a specific value of  $\boldsymbol{\theta}$ , and is denoted a locally optimal design, see e.g. [14]. An important subclass of nonlinear models, the partially nonlinear models, are defined by [39] and [42] as models where some of the parameters appear linearly. For  $D$ -optimality, these linear parameters do not affect the maximization problem, and thus  $D$ -optimal designs depend only on the nonlinear parameters. The Michaelis-Menten model, for example, is a partially nonlinear model where the

parameter  $\theta_1$  appears linearly.

This chapter is organized as follows. In Section 1.2, we review three classical methods for finding optimal designs. Section 1.3 is devoted to a recent approach, shedding light on optimal design problems from a more general perspective. All these methods are illustrated through a running example, the Michaelis Menten model. Further models are discussed in section 1.4. For each of these, only the most suitable method is applied, including a discussion of the drawbacks of the other methods for this particular situation. While the main focus of this chapter is optimal design for parameter estimation, section 1.5 gives a brief overview of optimal design when the purpose of the experiment is discrimination between two or more models. All designs provided in this chapter are locally optimal in the sense of [14], i.e. they depend on a best guess of the unknown model parameters. In Section 1.6, we briefly discuss approaches to overcome this problem. In each section, we point the interested reader to further relevant articles from the recent literature on optimal design for nonlinear models.

## 1.2 Classical methods

In this section, we distinguish between three approaches to facilitate the computation of optimal designs.

The standard method in many situations is the use of an appropriate Equivalence Theorem in order to find certain properties, usually the number of support points and possibly the inclusion of end points of  $\mathcal{X}$  in the support of the optimal design. Equivalence Theorems are available for all commonly applied optimality criteria based on Fisher information, e.g. the  $\phi_p$ -criteria or  $c$ -optimality. The most important application of an Equivalence Theorem, however, is checking optimality of a given candidate design.

Similarly powerful methods, summarized as the geometric approach, use the visualization of what is called the induced design space, a combination of the model and the design space  $\mathcal{X}$ . Again, this often leads to finding the number of support points of an optimal design, and



to results concerning the inclusion of boundary points of  $\mathcal{X}$ . Results are available for  $\phi_p$ - and  $c$ -optimality. Since the plots used for visualization have as many axes as the underlying model has parameters this approach is most useful for models with two or at most three parameters.

A further method is the functional approach; see e.g. [47]. The main idea of this approach is to express the support points (and sometimes also the weights) of optimal designs as implicit functions of some auxiliary parameters. In many cases these functions, being real and analytic, can be expanded into Taylor series, for the coefficients of which recursive formulae are available. Results in this area cover the  $D$ -,  $E$ - and  $c$ -criteria and some parameter robust criteria.

Unlike some more recent methods, see section 1.3, which aim at finding complete classes of optimal designs that are dominating with respect to the Loewner ordering, the classical methods usually solve one design problem at a time. Some of these approaches, however, allow conclusions for a particular class of optimality criteria, the  $\phi_p$ -criteria.

In some situations, the first two methods provide (some of) the support points of an optimal design, but usually no characterization of the optimal weights. For the situation where the optimal design has the minimal number of support points,  $m$ , to estimate  $\boldsymbol{\theta}$ , [51] have developed a method to find optimal weights given the support points, applicable to many optimality criteria including the  $\phi_p$ -criteria. A similar result is available for  $c$ -optimality.

Define the model matrix  $\mathbf{X} = [\mathbf{f}(x_1, \boldsymbol{\theta}) \dots \mathbf{f}(x_m, \boldsymbol{\theta})]'$ , and let  $\mathbf{V} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}$ . Then the  $\phi_p$ -optimal weights  $w_i$  of a design with support points  $x_1, \dots, x_m$  can be obtained by solving the system of equations

$$w_i = \frac{\sqrt{u_{ii}}}{\sum_{j=1}^m \sqrt{u_{jj}}}, \quad i = 1, \dots, m, \quad (1.3)$$

where  $u_{ii}$  is the  $i$ th diagonal element of the matrix  $\mathbf{U} = \mathbf{V}\mathbf{M}_\xi^{p+1}\mathbf{V}'$ . Note that the information matrix can be expressed as  $\mathbf{M}_\xi = \mathbf{X}'\mathbf{W}\mathbf{X}$ , where  $\mathbf{W}$  is the diagonal matrix holding the weights. Hence for  $p = -1$ , i.e.  $A$ -optimality, we get an explicit solution since  $\mathbf{U}$  reduces

to  $(\mathbf{X}\mathbf{X}')^{-1}$ , and thus the right hand side of (1.3) does not depend on the weights.

For  $c$ -optimality with respect to a vector  $\mathbf{c}$ , we also obtain an explicit solution  $w_i = |v_i| / \sum_{j=1}^m |v_j|$ ,  $i = 1, \dots, m$ , where the vector  $\mathbf{v}$  is defined as  $\mathbf{v} = \mathbf{V}\mathbf{c}$ .

### 1.2.1 Methods based on the Equivalence Theorem

In what follows, we consider the situation where interest is in the whole parameter vector  $\boldsymbol{\theta}$ . Moreover, we restrict attention to the  $\phi_p$ -optimality criteria. Equivalence results for subsystems of  $\boldsymbol{\theta}$  and further criteria which are information functions in the sense of [49] can e.g. be found in [49], Ch. 7.

The Equivalence Theorem for  $\phi_p$ -optimality is as follows.

**Theorem 1.2.1** *The design  $\xi$  is  $\phi_p$ -optimal on  $\mathcal{X}$  if and only if*

$$\mathbf{f}'(x, \boldsymbol{\theta}) \mathbf{M}_\xi^{p-1} \mathbf{f}(x, \boldsymbol{\theta}) - \text{trace}(\mathbf{M}_\xi^p) \leq 0 \quad \forall x \in \mathcal{X}. \quad (1.4)$$

*Equality applies in (1.4) in the support points of  $\xi$ .*

**Example (continued).** Suppose we seek the  $D$ -optimal design for the Michaelis Menten model. This problem is tackled in three standard steps:

- Step 1: Use the Equivalence Theorem to show that for any value of  $\boldsymbol{\theta}$  the  $D$ -optimal design on  $\mathcal{X} = [0, B]$  has exactly two support points. Hence the weights of the  $D$ -optimal design are  $w_1 = w_2 = 0.5$ ; see e.g. [55].
- Step 2: Show that the larger support point of the  $D$ -optimal design is given by  $B$ , the upper boundary of the design region  $\mathcal{X}$ .
- Step 3: Find the smaller support point of the  $D$ -optimal design.

**Step 1:** For  $D$ -optimality, Theorem 1.2.1 simplifies to:

**Corollary 1.2.2** *The design  $\xi$  is  $D$ -optimal for  $\boldsymbol{\theta}$  if and only if inequality*

$$d(\xi, x, \boldsymbol{\theta}) = \mathbf{f}'(x, \boldsymbol{\theta}) \mathbf{M}_\xi^{-1} \mathbf{f}(x, \boldsymbol{\theta}) - m \leq 0 \quad (1.5)$$

*holds for all  $x \in \mathcal{X}$ , with equality in the support points of  $\xi$ .*

Consider inequality (1.5) for the Michaelis Menten model with arbitrary parameter value  $\boldsymbol{\theta}$  and a  $D$ -optimal design  $\xi$ . We require that  $\xi$  has at least two support points to have nonsingular Fisher information  $\mathbf{M}_\xi$ . Multiplying (1.5) through with  $(\theta_2 + x)^4$ , we obtain a polynomial of degree four,  $p_4(x)$  say, on the left hand side. We now count its possible number of roots, the support points. A polynomial of degree four can have at most four roots. However, if  $p_4(x)$  had four roots at least the two middle ones would have to be turning points, since  $p_4(x)$  must not become positive on  $\mathcal{X}$ . Hence the derivative of  $p_4(x)$ , a polynomial of degree three, would have at least five roots, which is a contradiction. Now suppose  $p_4(x)$  has three roots on  $\mathcal{X}$ . By the same argument as before, only the middle root may be a turning point, so the other two roots have to be the endpoints of  $\mathcal{X}$ . A schematic of such a polynomial is depicted in Figure 1.2. Now substitute the lower endpoint into  $d(\xi, x, \boldsymbol{\theta})$ . Since  $\mathbf{f}(0, \boldsymbol{\theta}) = 0$ , we find that  $d(\xi, 0, \boldsymbol{\theta}) = -2 \neq 0$ , so 0 cannot be a support point of the  $D$ -optimal design, which contradicts the assumption of a three-point design.

Figure 1.2 approximately here

**Step 2:** By a standard result in optimal design theory, see e.g. [55], Lemma 5.1.3., a  $D$ -optimal design with the number of support points equal to  $m$ , the number of model parameters, (often called a saturated design) must have equal weights  $w_1 = \dots = w_m = 1/m$ . Hence the objective function becomes

$$\phi_D(\xi, \boldsymbol{\theta}) = |\mathbf{M}_\xi| = \frac{1}{4} \frac{\theta_1^2 x_1^2 x_2^2 (x_2 - x_1)^2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4}. \quad (1.6)$$

We note that the linear parameter  $\theta_1$  comes out as a factor, and does therefore not affect the maximization of (1.6) with respect to the design. Without loss of generality, let  $x_2$  be

the larger support point, i.e.  $x_2 > x_1$ . For the derivative of  $\phi_D$  with respect to  $x_2$  we obtain

$$\frac{\partial \phi_D(\xi, \boldsymbol{\theta})}{\partial x_2} = \frac{\theta_1^2 x_1^2 x_2 (x_2 - x_1) [\theta_2 (x_2 - 0.5x_1) + 0.5x_1 x_2]}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^5} > 0,$$

so  $\phi_D$  is increasing as  $x_2$  increases, and is thus maximized at the upper boundary,  $B$ , of  $\mathcal{X}$ .

**Step 3:** Substitute  $x_2 = B$  into (1.6) and solve  $\partial \phi_D(\xi, \boldsymbol{\theta}) / \partial x_1 = 0$  for  $x_1$ . There are three solutions,  $0$ ,  $B\theta_2 / (B + 2\theta_2)$  and  $B$ . For both  $x_1 = 0$  and  $x_1 = B$ , the objective function becomes zero, so these points correspond to local (and global) minima. Hence the point  $x_1 = B\theta_2 / (B + 2\theta_2)$ , situated between the two, has to correspond to the only local maximum, which is also global on  $\mathcal{X}$  since the values attained at the end points are minima.

We finally check for one example that the design obtained is indeed  $D$ -optimal. Figure 1.3 shows  $d(\xi, x, \boldsymbol{\theta})$  for the  $D$ -optimal design  $\xi$  with parameter vector  $\boldsymbol{\theta}' = (1, 0.6)$  and design region  $\mathcal{X} = [0, 1]$ . The conditions of the Equivalence Theorem are clearly satisfied.

Figure 1.3 approximately here

Note that the same strategy can be applied to search for other  $\phi_p$ -optimal designs. One major difference is that, unlike the  $D$ -optimal design, other  $\phi_p$ -optimal designs may depend on the value of the linear parameter  $\theta_1$ . The other difference is that the optimal weights are not readily available in closed form, which requires either the use of formula (1.3) or a further variable over which to optimize in Step 3 above.

### 1.2.2 Geometric approach

For  $c$ -optimality, Elfving's Theorem (see [33]) is a powerful tool for characterizing optimal designs. This was generalized to  $D$ -optimality and Bayesian optimality criteria by [16] and [17], respectively. A related approach based on covering ellipses was introduced by [53], [54] and [56] for  $D$ -optimality and subsequently used by e.g. [35] and [38]. This method was extended to  $E$ -optimality for linear and nonlinear models by [28], [29] and [22], respectively.

In an integrated approach, [11] generalized this method, for two-parameter models, to the class of  $\phi_p$ -optimality criteria. We will briefly review the results by [11] and [33] and illustrate them through an example.

Both approaches use the concept of an induced design space,  $\mathcal{G}$ , where  $\mathcal{G} = \{\mathbf{f}(x, \boldsymbol{\theta}), x \in \mathcal{X}\}$  with  $\mathbf{f}(x, \boldsymbol{\theta})$  defined in (1.2). In what follows, we require  $\mathcal{G}$  to be compact, which is trivially satisfied if  $\mathcal{X}$  is compact and, for every admissible value of  $\boldsymbol{\theta}$ ,  $\mathbf{f}(x, \boldsymbol{\theta})$  is continuous in  $x$ .

Figure 1.4 shows a parametric plot of the induced design space for the Michaelis-Menten model with parameter vector  $\boldsymbol{\theta}' = (1, 0.6)$  and design space  $\mathcal{X} = [0, 1]$  as a solid line. Its reflection  $-\mathcal{G}$  has been added as a dotted line. The axes are given by the entries of the vector  $\mathbf{f}(x, \boldsymbol{\theta})$ , i.e. the horizontal axis is  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_1$ , and the vertical axis is  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_2$ .

Figure 1.4 approximately here

### Elfving's Theorem and its implications for $c$ -optimality

Consider a vector  $\mathbf{c}$  and designs  $\xi$  such that  $\mathbf{c} \in \text{range}(\mathbf{M}_\xi)$  to ensure estimability of  $\mathbf{c}'\boldsymbol{\theta}$ ; see e.g. [48]. Define the Elfving set,  $\mathcal{E}$ , by

$$\mathcal{E} = \text{co}(\mathcal{G} \cup -\mathcal{G}),$$

where  $\text{co}(A)$  means the convex hull of a set  $A \subset \mathbb{R}^m$ . Elfving's Theorem characterizes a  $c$ -optimal design in terms of the intersection of the halfline  $\{\gamma\mathbf{c} \mid \gamma > 0\}$  with the boundary of  $\mathcal{E}$ .

**Theorem 1.2.3** *A design  $\xi = \{(x_1, w_1), \dots, (x_t, w_t)\}$  is  $c$ -optimal for estimating  $\mathbf{c}'\boldsymbol{\theta}$  if and only if there exists a positive number  $\gamma^* > 0$  and real numbers  $\varepsilon_1, \dots, \varepsilon_t \in \{-1, 1\}$  such that the point  $\gamma^*\mathbf{c} = \sum_{i=1}^t w_i \varepsilon_i \mathbf{f}(x_i, \boldsymbol{\theta})$  is a boundary point of the Elfving set  $\mathcal{E}$ .*

To see how this result can be used to find  $c$ -optimal designs, consider the following example.

**Example (continued).** Suppose an experimenter is interested in estimating percentiles  $x_r$  of the Michaelis-Menten curve, i.e. values of  $x$ , for which one expects a proportion  $r$  of the supremum of the concentration,  $\theta_1$ , to be attained. For  $r \in (0, 1)$ , we solve the equation

$$\frac{\theta_1 x}{\theta_2 + x} = r\theta_1$$

for  $x$  to obtain  $x_r = r\theta_2/(1 - r)$ . The problem of estimating  $x_r$  for fixed but arbitrary  $r \in (0, 1)$  is therefore equivalent to estimating  $\theta_2$ , and we seek a  $c$ -optimal design where  $\mathbf{c} = (0, 1)'$ .

Now consider the shape of the Elfving set  $\mathcal{E}$  depicted in Figure 1.5 (for  $\boldsymbol{\theta} = (1, 0.6)'$  and  $\mathcal{X} = [0, 1]$ ). The halfline  $\{\gamma(0, 1)' \mid \gamma > 0\}$  intersects the boundary of  $\mathcal{E}$  for some positive  $\gamma^*$  at the point  $(0, \gamma^*)'$  on the vertical axis. This point is a convex combination of two points, one at the right hand boundary point of  $\mathcal{G}$ , and one on  $-\mathcal{G}$ , since they are all on the same straight line.

Figure 1.5 approximately here

This tells us that the  $c$ -optimal design must have two support points,  $x_1$  and  $x_2$  (w.l.o.g. let  $x_1 < x_2$ ) satisfying

$$\begin{aligned} \begin{pmatrix} 0 \\ \gamma^* \end{pmatrix} &= w_1 \varepsilon_1 \begin{pmatrix} \frac{x_1}{\theta_2 + x_1} \\ \frac{-\theta_1 x_1}{(\theta_2 + x_1)^2} \end{pmatrix} + (1 - w_1) \varepsilon_2 \begin{pmatrix} \frac{x_2}{\theta_2 + x_2} \\ \frac{-\theta_1 x_2}{(\theta_2 + x_2)^2} \end{pmatrix} \\ &= -w_1 \begin{pmatrix} \frac{x_1}{0.6 + x_1} \\ \frac{-x_1}{(0.6 + x_1)^2} \end{pmatrix} + (1 - w_1) \begin{pmatrix} \frac{1}{0.6 + 1} \\ \frac{-1}{(0.6 + 1)^2} \end{pmatrix}. \end{aligned}$$

We have  $x_2 = 1$  since the right hand side end point of  $\mathcal{G}$  is attained at the upper bound of  $\mathcal{X}$ , and  $\varepsilon_1 = -1$ ,  $\varepsilon_2 = 1$  since they correspond to points on  $-\mathcal{G}$  and  $\mathcal{G}$ , respectively. As we do not know the value of  $\gamma^*$  we cannot use this system of equations to find  $x_1$  and  $w_1$ , but we can substitute  $x_2 = 1$  into the objective function,  $\phi_c(\boldsymbol{\xi}) = \mathbf{c}' \mathbf{M}_{\boldsymbol{\xi}}^- \mathbf{c}$ , and minimize either analytically or numerically with respect to  $x_1$  and  $w_1$ . Alternatively, we can use the

weight formula by [51] to find the optimal weight  $w_1$  in terms of  $x_1$  to reduce the number of variables in the optimization problem.

### Characterization of $\phi_p$ -optimal designs via covering ellipses

[11] express the  $\phi_p$ -optimal design problem for two-parameter models in terms of a dual problem.

**Theorem 1.2.4** *Let  $\mathbf{N}$  be a nonnegative definite matrix with eigenvalues  $\lambda_1$  and  $\lambda_2$ , and let  $q$  be determined by the equation  $p + q = pq$ . Define the  $v_{2q}$ -content of the ellipse  $E_{\mathbf{N}} = \{\mathbf{u} \in \mathbb{R}^2 \mid \mathbf{u}'\mathbf{N}\mathbf{u} \leq 1\}$  as*

$$v_{2q}(E_{\mathbf{N}}) = \frac{\text{Vol}(E_{\mathbf{N}})}{l_{2q}(E_{\mathbf{N}})} = \frac{\pi/\sqrt{\lambda_1\lambda_2}}{[(2/\sqrt{\lambda_1})^{2q} + (2/\sqrt{\lambda_2})^{2q}]^{1/(2q)}},$$

where  $\text{Vol}(E_{\mathbf{N}})$  denotes the volume of the ellipse  $E_{\mathbf{N}}$ , and  $l_{2q}(E_{\mathbf{N}})$  is the  $l_{2q}$ -mean of the lengths of its major and minor diameter.

*Then the  $\phi_p$ -optimal design problem is the dual of finding a centered ellipse  $E_{\mathbf{N}}$  which covers the induced design space  $\mathcal{G}$  and has minimal  $v_{2q}$ -content. Moreover, this ellipse touches  $\mathcal{G}$  at the points  $\mathbf{f}(x_i^*, \boldsymbol{\theta})$ , where  $x_i^*$  are the support points of any  $\phi_p$ -optimal design.*

We consider the following example in order to illustrate how the calculation of optimal designs can be facilitated by this method.

**Example (continued).** For arbitrary  $p \in [-\infty, 1]$ , consider the aim is to find a  $\phi_p$ -optimal design for the Michelis Menten model. From Figure 1.4 we can see that a centered ellipse that covers the induced design space  $\mathcal{G}$  must touch  $\mathcal{G}$  in exactly two points to have minimal  $v_{2q}$ -content. One of these points is the right hand side end point of  $\mathcal{G}$  corresponding to the upper boundary of the design space  $\mathcal{X}$ . This general form of design does not depend on the value of  $p$ . An example showing the  $D$ -optimal ( $p = 0$ ) covering ellipse is depicted in Figure 1.6.

Figure 1.6 approximately here

This approach provides a geometric characterization of the optimal support points, but not of the optimal weights. [11] present a counterexample where all  $\phi_p$ -optimal designs have the same covering ellipse with minimal  $v_{2q}$ -content, but different weights. The standard strategy for finding a  $\phi_p$ -optimal design would therefore follow similar steps as we have seen in the section on the Equivalence Theorem:

1. Establish the number of support points by visual inspection. (Two, in this example.)
2. Identify as many support points as possible. (The upper boundary of  $\mathcal{X}$ , in this example.)
3. Substitute this information into the objective function, and optimize with respect to the remaining support points and the weights.

Again, if the optimal design is saturated, the method by [51] can be used to find the optimal weights.

### 1.2.3 Functional approach

The idea behind the functional approach is the following. Suppose the design problem has been reduced to an optimization problem, for which the optimum is attained in the interior of its domain. Hence the solution can be found by setting the gradient (with respect to the design variables) of the objective function to zero and solving for the design variables. By “design variables” we mean all support points and weights of the optimal design that require calculation. Now this gradient,  $g(\boldsymbol{\tau}, \boldsymbol{\theta})$  say, depends on two sets of “variables”, the design variables,  $\boldsymbol{\tau}$  say, and the model parameters held in the vector  $\boldsymbol{\theta}$ . Under some regularity assumptions, the implicit function theorem implies that in the neighborhood  $U$  of an arbitrary vector  $\boldsymbol{\theta}_0$  for which we have a vector  $\boldsymbol{\tau}_0$  with  $g(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0) = 0$  there exists a function  $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$  such that for all  $\boldsymbol{\theta} \in U$  we obtain  $g(\boldsymbol{\tau}(\boldsymbol{\theta}), \boldsymbol{\theta}) = 0$ , and  $\boldsymbol{\tau}(\boldsymbol{\theta})$  can be



expanded into a Taylor series. The coefficients for this series can be obtained by recursive formulae provided in [47].

This means that once we have found an optimal design  $\boldsymbol{\tau}_0$  with respect to one parameter vector  $\boldsymbol{\theta}_0$ , we can approximate optimal designs  $\boldsymbol{\tau}(\boldsymbol{\theta})$  for different values of  $\boldsymbol{\theta}$  by the first few terms of their Taylor polynomials about  $\boldsymbol{\theta}_0$ . An essential assumption for this approach is that there are as many design variables as there are model parameters to make the Jacobian of  $\boldsymbol{\tau}(\boldsymbol{\theta})$  invertible. A natural application is therefore to find saturated  $D$ -optimal designs. In many other situations, the approach can still be applied after using some properties of the optimal design and/or restricting/transforming the parameter space. More details can be found in [47] and the references therein. The method is easy to implement in software allowing symbolic calculations, such as Mathematica or Maple, and usually only a few coefficients are required for a good approximation to the true function  $\boldsymbol{\tau}(\boldsymbol{\theta})$ , provided the interval for each component of  $\boldsymbol{\theta}$  is not too wide.

**Example (continued).** Suppose we seek  $D$ -optimal designs for the Michaelis Menten model on the design space  $\mathcal{X} = [0, 1]$ , for different values of the parameter vector  $\boldsymbol{\theta}$ . Further assume we have already established that 1, the upper bound of  $\mathcal{X}$ , is a support point, and that the optimal designs depend only on  $\theta_2$ . Hence we aim to approximate the smaller support point  $x_1 = \tau(\theta_2)$  as a function of  $\theta_2$ . We note that for this particular example  $\tau(\theta_2) = \theta_2/(1 + 2\theta_2)$  can be found explicitly by a simple calculation, so an approximation would not be necessary in practice. It is still useful for illustration of the method.

From (1.6), we have to maximize  $x_1^2(1 - x_1)^2/(\theta_2 + x_1)^4$ , so we set its derivative with respect to  $x_1$  equal zero, which, after some algebra and observing that  $x_1 \neq 0$ ,  $1 - x_1 \neq 0$  and  $\theta_2 + x_1 \neq 0$ , is equivalent to

$$g(x_1, \theta_2) = x_1(1 + 2\theta_2) - \theta_2 = 0.$$

Assume we expand the series about  $\theta_{2,0} = 0.6$ , then the corresponding  $\tau_0 = \tau(\theta_{2,0})$  is  $3/11$ . The recursive formulae from [47], p. 34, yield for the  $(s + 1)$ th coefficient of the Taylor

expansion

$$\tilde{\tau}_{s+1} = -J_0^{-1} \frac{1}{(s+1)!} \frac{\partial^{s+1} g(\hat{\tau}_s(\theta_2), \theta_2)}{\partial \theta_2^{s+1}}, \quad s = 0, 1, \dots,$$

where  $J_0$  is the Jacobian of  $g(\tau, \theta_2)$  with respect to  $\tau$ , evaluated at  $(\tau_0, \theta_{2,0})$ , and

$$\hat{\tau}_s(\theta_2) = \tilde{\tau}_0 + \sum_{j=1}^s \tilde{\tau}_j (\theta_2 - \theta_{2,0})^j$$

is the Taylor expansion from the  $s$ th step. Therefore, the function  $g(\hat{\tau}_s(\theta_2), \theta_2)$  depends on  $\theta_2$  in both components (the first component is a polynomial in  $\theta_2$ ), which must be taken into account when working out its partial derivatives with respect to  $\theta_2$ .

We obtain  $J_0 = 1 + 2\theta_{2,0} = 2.2$ , so  $-J_0^{-1} = -1/2.2 = -0.45$  and  $\tilde{\tau}_1 = -0.45(2\tau_0 - 1) = 0.2066$  to 4 d.p. Hence  $\hat{\tau}_1(\theta_2) = 3/11 + 0.2066(\theta_2 - 0.6)$ . For  $\tilde{\tau}_2$  we require the second derivative of  $g(3/11 + 0.2066(\theta_2 - 0.6), \theta_2) = 3/11 + 0.2066(\theta_2 - 0.6) + 2\theta_2[3/11 + 0.2066(\theta_2 - 0.6)] + \theta_2$  with respect to  $\theta_2$ . This is given by  $4 \times 0.2066 = 0.8264$ , and we obtain for the second coefficient  $\tilde{\tau}_2 = -J_0^{-1} \times 0.8264/2! = -0.1878$  (4 d.p.). Continuing in this manner yields the next coefficients  $\tilde{\tau}_3 = 0.1708$ ,  $\tilde{\tau}_4 = -0.1552$ ,  $\tilde{\tau}_5 = 0.1411$  and  $\tilde{\tau}_6 = -0.1283$ .

Figure 1.7 shows two Taylor approximations about  $\theta_2 = 0.6$  to the true function  $x_1 = \tau(\theta_2)$  on the domain  $\theta_2 \in [0.1, 1.1]$ . The Taylor polynomial of degree four is virtually identical to  $\tau(\theta_2)$  across this interval. On a smaller interval, e.g.  $\theta_2 \in [0.4, 0.8]$ , the Taylor polynomial of degree two is already a good approximation.

Figure 1.7 approximately here

### 1.3 General solutions

The “classical” methods have in common that design problems are solved on a case-by-case basis. Each combination of model and optimality criterion requires its own proof. There is a recent development towards more general solutions of design problems. [61] considered nonlinear models with two parameters, and algebraically derived conditions, under which for

each design  $\xi$  there is a design  $\tilde{\xi}$  from a simple class, which is at least as good as  $\xi$  in terms of the Loewner ordering of the corresponding information matrices, i.e.  $\mathbf{M}_{\tilde{\xi}} - \mathbf{M}_{\xi}$  is positive semidefinite. For any reasonable optimality criterion maximizing an objective function  $\phi(\xi)$ ,  $\mathbf{M}_{\tilde{\xi}} \geq \mathbf{M}_{\xi}$  implies that  $\phi(\tilde{\xi}) \geq \phi(\xi)$ , i.e. optimality criteria should be isotonic relative to the Loewner ordering. The information functions in the sense of [49], which include the  $\phi_p$ -criteria as special cases, satisfy this condition. These results were subsequently generalized to models with an arbitrary number of parameters by [59] and [24]. In particular, [59] provides considerably tighter bounds on the maximal number of support points than Caratheodory's bound, for large classes of models and arbitrary optimality criteria based on the information matrix, thus greatly reducing the computational effort to find optimal designs. In many situations the optimal designs are saturated, i.e. they have as many support points as the model has parameters to be estimated. A similar result has been shown for polynomial models by [15], which has subsequently been named the de la Garza phenomenon. [24] provide a rigorous investigation, using the theory of Chebyshev systems, see e.g. [41], of the classes of models for which the de la Garza phenomenon holds. A further extension of these methods, which can result in finding even smaller complete classes for optimal designs, can be found in [62].

### 1.3.1 Algebraic method

The method proposed in [61] and [59] uses a transformation of the information matrix for a design  $\xi$ , of the form

$$\mathbf{M}_{\xi} = \mathbf{P}(\boldsymbol{\theta}) \left[ \sum_{i=1}^t w_i \mathbf{C}(\boldsymbol{\theta}, z_i) \right] \mathbf{P}(\boldsymbol{\theta})', \quad (1.7)$$

where

$$\mathbf{C}(\boldsymbol{\theta}, z_i) = \begin{bmatrix} \Psi_{1,1}(z_i) & \cdots & \Psi_{1,m}(z_i) \\ \vdots & \ddots & \vdots \\ \Psi_{1,m}(z_i) & \cdots & \Psi_{m,m}(z_i) \end{bmatrix}$$

and  $\mathbf{P}(\boldsymbol{\theta})$  is a  $m \times m$  nonsingular matrix that depends on the value of  $\boldsymbol{\theta}$  only. Typically, for fixed  $\boldsymbol{\theta}$  and an interval design space  $\mathcal{X}$ , the map from  $x \in \mathcal{X}$  to  $z \in [Z_l, Z_u]$  is one-to-one and onto, and a design  $\xi$  can be expressed in terms of its transformed support points

$z_i, i = 1, \dots, t$ , and its weights.

Now rename all nonconstant and distinct functions  $\Psi_{l,s}, 1 \leq l \leq s \leq m$ , to  $\Psi_1, \dots, \Psi_k$ , such that  $\Psi_k$  is one of  $\Psi_{l,l}, 1 \leq l \leq m$ , and that there is no  $\Psi_{l,s} = \Psi_k$  for  $l < s$ . The idea behind this approach is to show that for each design  $\xi = \{(z_1, w_1), \dots, (z_t, w_t)\}$  there exists a design  $\tilde{\xi} = \{(\tilde{z}_1, \tilde{w}_1), \dots, (\tilde{z}_t, \tilde{w}_t)\}$  from a simple class, for which  $\sum_{i=1}^t \tilde{w}_i \Psi_j(\tilde{z}_i) = \sum_{i=1}^t w_i \Psi_j(z_i)$  for  $j = 1, \dots, k-1$ , and  $\sum_{i=1}^t \tilde{w}_i \Psi_k(\tilde{z}_i) \geq \sum_{i=1}^t w_i \Psi_k(z_i)$ , which makes  $\tilde{\xi}$  at least as good as  $\xi$  in the Loewner ordering. The extension by [62] uses a similar idea where instead of just one entry,  $\Psi_k(z)$ , an arbitrary lower principal submatrix of the matrix  $\mathbf{C}(\boldsymbol{\theta}, z)$  is considered.

If  $\Psi_1, \dots, \Psi_k$  are differentiable infinitely often and the functions  $\psi_{l,s}$  recursively defined as

$$\psi_{l,s}(z) = \begin{cases} \Psi_l'(z), & s = 1, l = 1, \dots, k, \\ \left( \frac{\psi_{l,s-1}(z)}{\psi_{s-1,s-1}(z)} \right)', & 2 \leq s \leq k, s \leq l \leq k \end{cases} \quad (1.8)$$

have no zero value on the transformed design space  $[Z_l, Z_u]$ , the following result holds; see [59]:

**Theorem 1.3.1** *Let  $\Phi(z) = \prod_{l=1}^k \psi_{l,l}(z)$ ,  $z \in [Z_l, Z_u]$ . For any given design  $\xi$  there exists a design  $\tilde{\xi}$ , such that  $\mathbf{M}_\xi \leq \mathbf{M}_{\tilde{\xi}}$  in the Loewner ordering.*

- (a) *When  $k$  is odd and  $\Phi(z) < 0$ ,  $\tilde{\xi}$  has at most  $(k+1)/2$  support points including point  $Z_l$ .*
- (b) *When  $k$  is odd and  $\Phi(z) > 0$ ,  $\tilde{\xi}$  has at most  $(k+1)/2$  support points including point  $Z_u$ .*
- (c) *When  $k$  is even and  $\Phi(z) > 0$ ,  $\tilde{\xi}$  has at most  $k/2 + 1$  support points including points  $Z_l$  and  $Z_u$ .*
- (d) *When  $k$  is even and  $\Phi(z) < 0$ ,  $\tilde{\xi}$  has at most  $k/2$  support points.*

Note that the general formula for computing  $\Phi(z)$  can easily be implemented in software that is capable of symbolic calculations. Furthermore, even if  $\Phi(z)$  has a complicated structure, making it impossible to check directly if this function is positive/negative on  $[Z_l, Z_u]$ , we can easily obtain this information from visual inspection of its graph.

To fix ideas, consider the following example.

**Example (continued).** Consider the Michaelis Menten model. After some algebra, we find that the information matrix can be written in form (1.7) with

$$\mathbf{P}(\boldsymbol{\theta}) = \begin{bmatrix} 1/\theta_1 & 0 \\ -1/\theta_2 & 1/(\theta_1\theta_2) \end{bmatrix}, \quad \mathbf{C}(\boldsymbol{\theta}, z) = \begin{bmatrix} z^2 & z^3 \\ z^3 & z^4 \end{bmatrix},$$

where  $z = \theta_1 x / (\theta_2 + x)$ . Let  $\Psi_1(z) = \Psi_{1,1}(z) = z^2$ ,  $\Psi_2(z) = \Psi_{1,2}(z) = z^3$  and  $\Psi_3(z) = \Psi_{2,2}(z) = z^4$ . Then  $k = 3$ ,  $\Psi_k = \Psi_{l,l}$  for  $l = 2$ , and there is no  $l < s$  with  $\Psi_{l,s} = \Psi_k$ . From (1.8), we find that

$$\begin{aligned} \psi_{1,1} &= \Psi'_1(z) = 2z \\ \psi_{2,2} &= \left( \frac{\psi_{2,1}(z)}{\psi_{1,1}(z)} \right)' = \left( \frac{\Psi'_2(z)}{\Psi'_1(z)} \right)' = \left( \frac{3z^2}{2z} \right)' = \frac{3}{2} \\ \psi_{3,3} &= \left( \frac{\psi_{3,2}(z)}{\psi_{2,2}(z)} \right)' = \left( \frac{[\psi_{3,1}(z)/\psi_{1,1}(z)]'}{3/2} \right)' = \left( \frac{[\Psi'_3(z)/\Psi'_1(z)]'}{3/2} \right)' = \left( \frac{[4z^3/2z]'}{3/2} \right)' = \frac{8}{3}. \end{aligned}$$

Hence  $\Phi(z) = 2z \times 3/2 \times 8/3 = 8z > 0$  if  $z > 0$ . Since an observation in  $x = 0$  does not give any increase in information, we can choose  $\mathcal{X} = [A, B]$  where  $A$  is small but positive instead of the interval  $[0, B]$  we used before. Therefore  $x > 0$ , and we obtain that  $z > 0$  since  $\theta_1 > 0$  and  $\theta_2 > 0$ . That means we are in the situation of case (b) in Theorem 1.3.1. Hence for each optimality criterion based on the information matrix, there exists an optimal design with no more than  $(k + 1)/2 = 2$  support points, one of which is  $Z_u$ , which translates into  $B$  by the inverse map. This confirms our results for this example from previous sections. It remains to select an appropriate optimality criterion, and to use analytical or numerical search to obtain an optimal design from the class found.

Note that for larger values of  $k$ , we do not necessarily obtain the existence of a saturated optimal design. For example, if  $p = 6$ ,  $k$  can be as large as 21, provided all  $\Psi_{l,s}$ ,  $1 \leq l \leq s \leq m$  are distinct. If  $\Phi(z) \neq 0$  on  $[Z_l, Z_u]$ , there exists an optimal design with at most  $(k+1)/2 = 11$  support points, one of which is already determined. This leaves an optimization problem in 20 dimensions, 10 for the remaining support points, and 10 for the weights. This is still a considerable improvement on Caratheodory's lower bound,  $m(m + 1)/2$ , see e.g. [55] or

[49], where in the above situation we would have to solve an optimization problem in 41 dimensions (21 support points and 20 weights). However, for many models some  $\Psi_{l,s}$  occur repeatedly in the information matrix. For example, for polynomials of degree five, i.e.  $p = 6$ , the information matrix holds only 11 different functions of  $z = x$ , the monomials up to degree ten. The monomial of degree zero, i.e. the function constant to 1, does not depend on the design, and hence  $k = 10$ , and there exists an optimal design with six support points, two of which are the endpoints of the design interval  $\mathcal{X}$ , as given by the de la Garza phenomenon. In the next subsection, we will investigate in more detail for which classes of models the de la Garza phenomenon holds.

### 1.3.2 Method based on Chebyshev systems

Roughly speaking, [24] demonstrate that the de la Garza phenomenon occurs in any model, for which the functions in the information matrix, together with the constant function, form a Chebyshev system. Following [41], a set of  $k + 1$  continuous functions  $\{u_0, \dots, u_k\}$  is called a Chebyshev system on the interval  $[Z_l, Z_u]$  if

$$\begin{vmatrix} u_0(z_0) & \cdots & u_0(z_k) \\ \vdots & \ddots & \vdots \\ u_k(z_0) & \cdots & u_k(z_k) \end{vmatrix} > 0 \quad (1.9)$$

for all  $Z_l \leq z_0 < \dots < z_k \leq Z_u$ . Note that if the determinant in (1.9) is negative then the set  $\{u_0, \dots, -u_k\}$  is a Chebyshev system on  $[Z_l, Z_u]$ .

As in Section 1.3.1, denote the different elements of the transformed information matrix  $\mathbf{C}(\boldsymbol{\theta}, z)$  defined in (1.7) as  $\Psi_1, \dots, \Psi_k$ , such that  $\Psi_k$  is one of the diagonal elements, and that there is no  $\Psi_{l,s} = \Psi_k$  for  $l < s$ . Assume that  $\Psi_1, \dots, \Psi_k$  are all continuous, so infinite differentiability of the  $\Psi_i$  as in [59] is not required. Furthermore, let  $\Psi_0(z) = 1$ , and define the index,  $I(\xi)$ , of a design  $\xi$  on the interval  $[Z_l, Z_u]$  as the number of support points, where the boundary points,  $Z_l$  and  $Z_u$ , are only counted by 1/2.

Suppose that the sets  $\{\Psi_0, \Psi_1, \dots, \Psi_{k-1}\}$  and  $\{\Psi_0, \Psi_1, \dots, \Psi_k\}$  are Chebyshev systems.

Then the following result holds; see Theorem 3.1 in [24].

**Theorem 1.3.2** *For any design  $\xi$  there exists a design  $\tilde{\xi}$  with at most  $(k + 2)/2$  support points such that  $\mathbf{M}_{\tilde{\xi}} \geq \mathbf{M}_{\xi}$ . If  $I(\xi) < k/2$ , then  $\tilde{\xi} = \xi$ . Otherwise:*

- (a) *If  $k$  is odd,  $\tilde{\xi}$  has at most  $(k + 1)/2$  support points, and can be chosen such that its support contains the point  $Z_u$ .*
- (b) *If  $k$  is even,  $\tilde{\xi}$  has at most  $k/2 + 1$  support points, and can be chosen such that its support contains the points  $Z_l$  and  $Z_u$ .*

If the sets  $\{\Psi_0, \Psi_1, \dots, \Psi_{k-1}\}$  and  $\{\Psi_0, \Psi_1, \dots, -\Psi_k\}$  are Chebyshev systems, a similar result holds, with the point  $Z_u$  in (a) replaced by  $Z_l$ ,  $k/2 + 1$  in (b) replaced by  $k/2$ , and the points  $Z_l$  and  $Z_u$  crossed out in (b).

Several sets of functions, e.g. the monomials up to degree  $m$  for any integer  $m$ , are known to be Chebyshev systems. If this information is not available, using the definition given in (1.9) can be unwieldy, in particular if  $k$  is large. In this situation, it is usually easier to check the condition on  $\Phi(z)$  from the algebraic method described in subsection 1.3.1.

**Example (continued).** To apply this result to the Michaelis Menten model, we need to check if the sets of functions  $S_2 = \{\Psi_0(z), \Psi_1(z), \Psi_2(z)\}$  and  $S_3 = S_2 \cup \{\Psi_3(z)\}$  are Chebyshev systems on  $[Z_l, Z_u]$  where  $Z_l \geq 0$ ,  $\Psi_0(z) = 1$ ,  $\Psi_1(z) = z^2$ ,  $\Psi_2(z) = z^3$  and  $\Psi_3(z) = z^4$ . For  $S_2$  we obtain

$$\begin{vmatrix} 1 & 1 & 1 \\ z_0^2 & z_1^2 & z_2^2 \\ z_0^3 & z_1^3 & z_2^3 \end{vmatrix} = (z_2 - z_1)(z_2 - z_0)(z_1 - z_0)(z_1 z_2 + z_1 z_0 + z_2 z_0) > 0$$

for  $Z_l \leq z_0 < z_1 < z_2 \leq Z_u$ . Similarly, the determinant for  $S_3$  is  $(z_3 - z_1)(z_3 - z_0)(z_3 - z_2)(z_2 - z_1)(z_2 - z_0)(z_1 - z_0)(z_0 z_1 z_2 + z_1 z_2 z_3 + z_0 z_2 z_3 + z_0 z_1 z_3)$ , which is also positive for  $Z_l \leq z_0 < z_1 < z_2 < z_3 \leq Z_u$ . Hence  $S_2$  and  $S_3$  are both Chebyshev systems on  $[Z_l, Z_u]$ .

Here  $k = 3$ , so  $(k + 2)/2 = 2.5$ , and for any design  $\xi$  the dominating design  $\tilde{\xi}$  has at

most 2 support points. Only a design  $\xi$  with support points  $Z_l$  and  $Z_u$  can achieve an index  $I(\xi) < k/2 = 1.5$ . Hence such a design is dominated by itself. Any other design will be dominated by a design  $\tilde{\xi}$  with at most  $(k+1)/2 = 2$  support points, one of which is the upper bound of the transformed design interval,  $Z_u$ , which translates into the upper bound,  $B$ , of the original design space  $\mathcal{X}$ . Note that if interest is in estimating both model parameters any dominating design must have exactly two support points to ensure estimability. It remains to select an appropriate optimality criterion, and to use analytical or numerical search to obtain an optimal design from the class found.

## 1.4 Further examples

In this section, we will apply the methods described earlier to further examples. Unlike the previous sections, we will not apply every method to every model, but only a combination of the most suitable methods for each situation.

### 1.4.1 The two-parameter exponential model

Exponential growth models with expected response  $\eta(x, \boldsymbol{\theta})$  of the form

$$\eta(x, \boldsymbol{\theta}) = \sum_{l=1}^L a_l e^{-b_l x}, \quad a_l > 0, \quad l = 1, \dots, L, \quad (1.10)$$

occur in chemical kinetics, see e.g. [36], with particular emphasis on toxicology (see [5] and [6]) and microbiology (see [1]). Locally  $D$ -,  $c$ -, and  $E$ -optimal designs for this class of models have been found in [34], [25] and [26].

For the purpose of this example, we assume that  $L = 1$  in (1.10), and for consistency rename the parameters to obtain  $\eta(x, \boldsymbol{\theta}) = \theta_1 e^{-\theta_2 x}$ . We further let  $\theta_2 > 0$ , and  $\mathcal{X} = [0, B]$  for some  $B > 0$ . Note that knowledge of the sign of  $\theta_2$  is not a restrictive assumption, since the experimenter will usually know whether to expect growth or decline. The information



matrix for this model for a design  $\xi$  is given by

$$\mathbf{M}_\xi = \sum_{i=1}^t w_i \begin{bmatrix} e^{-2\theta_2 x_i} & -\theta_1 x_i e^{-2\theta_2 x_i} \\ -\theta_1 x_i e^{-2\theta_2 x_i} & \theta_1^2 x_i^2 e^{-2\theta_2 x_i} \end{bmatrix}. \quad (1.11)$$

Before selecting an optimality criterion, we seek the complete class of dominating designs in the Loewner sense. Following the approaches presented in section 1.3, we simplify the functions in the information matrix (1.11) using the transformation  $z = \theta_2 x$ ,  $z \in [0, \theta_2 B]$ , and defining

$$\mathbf{P}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 0 \\ 0 & -\theta_1/\theta_2 \end{bmatrix}.$$

This yields the functions  $\Psi_1(z) = e^{-2z}$ ,  $\Psi_2(z) = ze^{-2z}$  and  $\Psi_3(z) = z^2 e^{-2z}$ . The algebraic method from subsection 1.3.1 involves checking if  $\Phi(z)$ , a function consisting of ratios of  $\Psi_1, \Psi_2, \Psi_3$  and their derivatives, is positive/negative on  $[0, \theta_2 B]$ . We can see that the exponential term,  $e^{-2z}$ , will cancel in these ratios, and therefore expect  $\Phi(z)$  to have a simple form. Alternatively, we could use the method described in subsection 1.3.2, and show that  $\{1, \Psi_1, \Psi_2\}$  and  $\{1, \Psi_1, \Psi_2, \pm\Psi_3\}$  are Chebyshev systems, which appears to be harder. It turns out that  $\Phi(z) = -4e^{-2z}$ , which is negative for all  $z$ . Hence we stick to the algebraic method for this example.

Here,  $k = 3$  is odd and  $\Phi(z) < 0$  for all  $z \in [Z_l, Z_u]$ . Hence, from Theorem 1.3.1, we obtain that the dominating designs have at most  $(k+1)/2 = 2$  support points, one of which is  $Z_l$ , or zero in the original design space. It thus remains to find the other support point (if necessary), and one weight.

Note that if interest is in estimating both parameters, any optimal design will have two support points. If, however, we seek a  $c$ -optimal design, this may have just one support point, zero, depending on where the vector  $\mathbf{c}$  intersects the boundary of the Elfving set  $\mathcal{E}$ .

Assume the experimenter is only interested in estimating the rate,  $\theta_2$ , of exponential decay. In this case, the  $c$ -optimality criterion with  $\mathbf{c} = (0, 1)'$  will be appropriate. Now there are two simple ways to establish if a second support point is needed: A plot of the Elfving

set  $\mathcal{E}$  and the vector  $\mathbf{c}$ , similar to figure 1.5, or finding the weight  $w$  in the point zero by the method described in [51], and checking if this is always less than one. A weight equal one would indicate a design with only zero as its support point. For plotting the Elfving set, we would need to specify values for the two model parameters, hence to keep our results as general as possible, we use the latter method.

After some algebra, we find that the vector  $\mathbf{v} = \mathbf{V}\mathbf{c}$  is given by  $(1/(\theta_1 x_2), e^{\theta_2 x_2}/(\theta_1 x_2))'$ . Since neither of the two entries is equal zero, the weight  $w = |v_1|/(|v_1| + |v_2|) = 1/(1 + e^{\theta_2 x_2})$  is strictly between zero and one, hence a second support point is required. We further note that  $w$  does not depend on the value of  $\theta_1$ , and on  $\theta_2$  only through the product  $\theta_2 x_2$ .

We substitute the expression for  $w$  into the objective function and obtain

$$\phi_c(\xi, \boldsymbol{\theta}) = \frac{(1-w)e^{-2\theta_2 x_2} + w}{w(1-w)e^{-2\theta_2 x_2} \theta_1^2 x_2^2} = \frac{(1 + e^{\theta_2 x_2})^2}{\theta_1^2 x_2^2}.$$

Setting the derivative with respect to  $x_2$  equal zero is equivalent to solving

$$e^{\theta_2 x_2} (\theta_2 x_2 - 1) = 1,$$

which yields  $x_2 = 1.278/\theta_2$  (3 d.p.). Inspection of the second derivative reveals that this is indeed a minimum. Hence  $\theta_2 x_2 = 1.278$  is constant, and the weight  $w = 0.2178$  is constant, too, for any combination of  $\theta_2$  and the corresponding optimal value of  $x_2$ .

If  $\theta_2$  is relatively small, the optimal  $x_2$  becomes large, and may not be included in the design interval  $\mathcal{X} = [0, B]$ . In this case, inspection of the first derivative of the objective function with respect to  $x_2$  reveals that  $\phi_c$  is strictly decreasing on  $\mathcal{X}$ , and therefore the second support point has to be  $B$ . The corresponding optimal weight is then given by  $w = 1/(1 + e^{\theta_2 B})$ , and depends on the value of  $\theta_2$ .

Suppose now that the experimenter's interest is in  $\theta_1$  only, i.e.  $\mathbf{c} = (1, 0)'$ . This scenario is less likely to occur in practice, but we include it for illustration. In this case, it turns out that the second entry of the vector  $\mathbf{v}$ ,  $v_2$ , is equal zero, so in this situation the optimal weight at point  $x_1 = 0$  is one, and the  $c$ -optimal design has only one support point. This corresponds

to intuition, since  $\eta(0, \boldsymbol{\theta}) = \theta_1$ , so we expect that this parameter can be estimated well if we observe in  $x = 0$ . A plot of the Elfving set could reveal such information for all possible choices of vectors  $\mathbf{c}$  at one glance, but suffers from the dependence of  $\mathcal{E}$  on the unknown parameter values. In many cases, however, one can argue that the parameter values do not affect the general shape of  $\mathcal{E}$ , which can then be used to gain general insight into the design problem.

In this example, substituting the expression for the optimal weight into the objective function resulted in a considerable simplification of the optimization problem. This is not necessarily always the case. For example the objective function for  $A$ -optimality for the exponential model appears to become rather more complicated, possibly because of the square root terms involved in the weight formula.

#### 1.4.2 The Emax model

The Emax model is a generalization of the Michaelis Menten model, and is widely used in dose response studies. The expected response is given by

$$\eta(x, \boldsymbol{\theta}) = \theta_0 + \frac{\theta_1 x}{\theta_2 + x}, \quad \theta_1, \theta_2 > 0, \quad x \in [0, B],$$

where  $\theta_0$  represents the placebo response,  $\theta_1$  (often called  $E_{max}$ ) is the maximum achievable increase above the placebo response, and  $\theta_2$  is the dose which produces 50% of the  $E_{max}$  effect. [23] show that  $D$ - and  $ED_r$ -optimal designs for this model have the same support points but different weights, where  $ED_r$  is the smallest dose that achieves a proportion of  $r$ ,  $0 < r < 1$ , of the maximum effect in the observed dose range. Optimal designs for estimating the minimum effective dose are given in [21].

Again, we start with writing the information matrix in the form (1.7). Following [59], we

use the transformation  $z = 1/(\theta_2 + x)$  and

$$\mathbf{P}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -\theta_2 & 0 \\ 0 & -\theta_1 & \theta_1\theta_2 \end{bmatrix}$$

to obtain  $\Psi_1(z) = z$ ,  $\Psi_2(z) = z^2$ ,  $\Psi_3(z) = z^3$  and  $\Psi_4(z) = z^4$ . We know that the monomials of degree  $0, \dots, l$  form a Chebyshev system for every integer  $l$ , hence we can apply Theorem 1.3.2 immediately.

Here,  $k = 4$ , so for each design  $\xi$  there exists a dominating design  $\tilde{\xi}$  with at most  $(k + 2)/2 = 3$  support points. In particular, provided the index  $I(\xi) \geq k/2 = 2$ , from part (b), we obtain that  $Z_l$  and  $Z_u$  can be chosen as support points, which translate back into the endpoints of the design interval  $\mathcal{X}$ . A design with index strictly less than 2 has strictly less than three support points, and thus produces a singular information matrix. If interest is in estimating all parameters, we can thus restrict design search to designs with three support points, including zero and  $B$ .

The Emax model is a generalization of the Michaelis Menten model, having an additional parameter for a possible placebo effect. It may therefore be of interest how well an optimal design for the Emax model performs if the true model is the Michaelis Menten model. This corresponds to a situation where a placebo effect was anticipated, and therefore taken into account when designing the experiment, but then it turned out that this parameter was unnecessary in the model, so a Michaelis Menten model would be used in the analysis. For comparison with designs found previously, we seek the  $D$ -optimal design for the Emax model, on a design space  $\mathcal{X} = [0, B]$ . Since this model is partially nonlinear, the  $D$ -optimal design will not depend on the linear parameters  $\theta_0$  and  $\theta_1$ .

We know that the weights of this saturated  $D$ -optimal design will be equal, i.e.  $w_1 = w_2 = w_3 = 1/3$ . Substituting these weights, together with the known support points, 0 and  $B$ , into the objective function, we obtain

$$\phi_D(\xi, \boldsymbol{\theta}) = |\mathbf{M}_\xi| = \frac{1}{3^3} \frac{\theta_1^2 B^2 x_2^2 (B - x_2)^2}{(\theta_2 + x_2)^4 (\theta_2 + B)^4}.$$

This is proportional to the objective function of the Michaelis Menten model (1.6), and thus is also maximized by  $x_2 = \theta_2 B / (B + 2\theta_2)$ .

We find that two of the support points of the  $D$ -optimal design,  $\xi^*$  say, for the Emax model coincide with those of the  $D$ -optimal design for the Michaelis Menten model. The third support point of  $\xi^*$ ,  $x_1 = 0$ , however does not provide any information for the Michaelis Menten model, i.e. the information matrix in this point is the zero matrix. To assess the performance of  $\xi^*$  for estimating the Michaelis Menten model, we compute its  $D$ -efficiency, where the  $D$ -efficiency of a design  $\xi$  is defined as

$$\text{eff}_D(\xi) = \left( \frac{|\mathbf{M}_\xi|}{|\mathbf{M}_{\xi_D}|} \right)^{1/m}, \quad (1.12)$$

with  $\xi_D$  the  $D$ -optimal design for the true scenario.

For this example,  $\xi_D$  is the  $D$ -optimal design for the Michaelis Menten model, and  $m = 2$ , and we straightforwardly obtain that  $\mathbf{M}_{\xi^*} = 2/3 \times \mathbf{M}_{\xi_D}$  for all eligible values of  $\theta_2$  and  $B$ , provided these are the same for both designs. Hence, regardless of the parameter values or the upper endpoint of the design interval, we have that  $\text{eff}_D(\xi^*) = 2/3$ .

The  $D$ -efficiency of a design  $\xi$  can be interpreted as the proportion of observations one needs from the optimal design to get the same accuracy in parameter estimation as for design  $\xi$ . For example, if we conducted a clinical trial with 300 patients, using the  $D$ -optimal design for the Emax model, but then it turned out that there is no placebo effect, and the Michaelis Menten model is appropriate, we could get estimates for  $\theta_1$  and  $\theta_2$  with the same precision from a trial with 200 patients using the  $D$ -optimal design for the Michaelis Menten model. In practice, however, it would not be known before analysing the data from the trial that the placebo effect is not significant. At the planning stage, there are therefore two possible scenarios (placebo effect/no placebo effect) and two possible decisions (design for Emax/Michaelis Menten model) for the experimenter to make. Taking into account that the  $D$ -optimal design for the Michaelis Menten model is not capable of estimating/testing the presence/absence of the placebo effect, using design  $\xi^*$  seems to be the safer bet, even if some efficiency is lost if the smaller model is correct. In practice, a compromise design could

be employed, for example putting only weight 0.2 at point zero, and weight 0.4 at the other two support points. This design has higher efficiency in the smaller model than  $\xi^*$ , and is capable of estimating/testing all parameters. There will, however, be some loss in efficiency if the Emax model is correct.

### 1.4.3 A heteroscedastic linear model

In some situations, it is not realistic to assume that the variability of observations is constant throughout the design region, i.e. the variance function  $\text{Var}(\epsilon_i) = \sigma^2(x_i, \boldsymbol{\alpha})$  for some parameter vector  $\boldsymbol{\alpha}$ , where the functional form of  $\sigma^2(x_i, \boldsymbol{\alpha})$  is known. If this is not taken into account when planning the experiment, an inefficient analysis may result. Consider the class of heteroscedastic models where observations are described by

$$Y_i = \eta(x_i, \boldsymbol{\theta}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2(x_i, \boldsymbol{\alpha})), \quad i = 1, \dots, n. \quad (1.13)$$

The function  $\lambda(x_i, \boldsymbol{\alpha}) := 1/\sigma^2(x_i, \boldsymbol{\alpha})$  is called efficiency or intensity function. We note that even in the simplest case, i.e. when the parameters in the efficiency function do not overlap with the parameter vector  $\boldsymbol{\theta}$  in the expected response, and are not of interest, optimal designs will be affected by heteroscedasticity. In particular, design problems for linear models with nonconstant variance resemble those for nonlinear models in that they depend on the nuisance parameters held in  $\boldsymbol{\alpha}$ . In what follows, we will consider a simple example.

Let  $\eta(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2$ , and  $\lambda(x, \boldsymbol{\alpha}) = e^{-\alpha x}$  for some  $\alpha > 0$  and  $x \in \mathcal{X} = [0, \infty)$ . That means, we assume the variance is increasing exponentially as  $x$  increases. Here the information matrix for estimating  $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)'$  is given by

$$\mathbf{M}_{\boldsymbol{\xi}} = \sum_{i=1}^t w_i \lambda(x_i, \boldsymbol{\alpha}) \begin{bmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{bmatrix} = \sum_{i=1}^t w_i e^{-\alpha x_i} \begin{bmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{bmatrix},$$

which clearly depends on  $\alpha$ . Replacing  $\alpha x = z$ ,  $z \in [Z_l, Z_u) = [0, \infty)$ , and letting

$$\mathbf{P}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\alpha & 0 \\ 0 & 0 & 1/\alpha^2 \end{bmatrix}$$

we obtain  $\Psi_1(z) = e^{-z}$ ,  $\Psi_2(z) = ze^{-z}$ ,  $\Psi_3(z) = z^2e^{-z}$ ,  $\Psi_4(z) = z^3e^{-z}$  and  $\Psi_5(z) = z^4e^{-z}$ . We observe that the design problem closely resembles the problem for the exponential model. In fact, if we dropped the squared term,  $\theta_2 x^2$ , from the model equation, we would get the identical problem as in subsection 1.4.1 in terms of finding the simple complete class of Loewner dominating designs.

Again we consider the method by [59]. We find that  $\Phi(z) = -24e^{-z} < 0$ , which implies that the optimal design will have at most 3 support points, one of which is  $Z_l$ , translating into  $x_1 = 0$  in the original design interval. Suppose interest is in estimating all three parameters in  $\boldsymbol{\theta}$ , and we select the  $D$ -criterion. In this case, the optimal design will have exactly three support points, one of which is zero, and equal weights. The objective function to maximize is

$$|\mathbf{M}_\xi| = \frac{1}{27} e^{-\alpha(x_2+x_3)} x_2^2 x_3^2 (x_3 - x_2)^2.$$

A straightforward maximization yields the remaining  $D$ -optimal support points  $x_2 = (3 - \sqrt{3})/\alpha$  and  $x_3 = (3 + \sqrt{3})/\alpha$ .

[44] show a more general result, finding  $D$ -optimal designs for polynomials of any degree with efficiency function  $\lambda(x) = e^{-x}$  and several further functions commonly used to model heteroscedasticity. In particular, they find that  $D$ -optimal designs on  $\mathcal{X} = [0, \infty)$  for model (1.13) where  $\eta(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \dots + \theta_k x^k$ ,  $k \geq 1$ , and  $\lambda(x) = e^{-x}$  have  $k + 1$  equally weighted support points. These are given by the roots of  $xL_k^{(1)}(x)$  where  $L_k^{(1)}(x)$  is the  $k$ th generalized Laguerre polynomial. For more information on generalized Laguerre or indeed other classical orthogonal polynomials, see e.g. [57].

By a simple transformation, we see that for efficiency function  $\lambda(x, \alpha) = e^{-\alpha x}$  the support points of the  $D$ -optimal design are the roots of  $xL_k^{(1)}(\alpha x)$ . We apply this to our example for comparison. The generalized Laguerre polynomial  $L_2^{(1)}(x)$  is given by  $L_2^{(1)}(x) = 0.5x^2 - 3x + 3$ .

Solving

$$xL_2^{(1)}(\alpha x) = x(0.5\alpha^2 x^2 - 3\alpha x + 3) = 0$$

confirms the results we found before.

Note that [44] used a method not yet described in this chapter, the method of expressing the objective function in terms of canonical moments. We will only give a brief description of this approach. For further reading the interested reader is referred to [30] and the references therein.

The idea behind this approach is as follows. The entries in the information matrix  $\mathbf{M}_\xi$  can be viewed as the moments of the design  $\xi$ . This is particularly evident for polynomial models with constant variance. If we tried to maximize the determinant of  $\mathbf{M}_\xi$  with respect to the moments, we would get into trouble due to the complicated structure of the moment spaces. For example, the possible range for the second moment will depend on the value of the first moment in a nontrivial way.

Canonical moments are transformations of the ordinary moments of a probability measure. Roughly speaking, a canonical moment determines the relative position of the corresponding ordinary moment in its moment space, given the lower order moments. The big advantage of canonical moments is the simple structure of their moment spaces. In particular, they do not depend on the values of the lower order canonical moments. If it is possible to express the objective function in terms of canonical moments, it can be optimized over each of these quantities independently, which usually results in a considerable simplification of the problem. Once the optimal canonical moments have been found, the corresponding optimal design can be determined by applying results on continued fractions, the Stieltjes transform and orthogonal polynomials. The major limitation of canonical moments is that the objective function can only be expressed as a function of canonical moments in a few special cases, e.g. for  $D$ - or  $D_s$ -optimality for polynomial models with certain efficiency functions or trigonometric models with constant variance.



## 1.5 Model discrimination for nonlinear models

To discriminate between two nested models, a popular optimality criterion is  $D_s$ -optimality. Intuitively, this is related to  $D$ -optimality for the  $s$  entries in the parameter vector  $\boldsymbol{\theta}$  by which the models differ. Without loss of generality let  $\boldsymbol{\theta}' = (\boldsymbol{\theta}'_{(1)}, \boldsymbol{\theta}'_{(2)})$  where the  $s$  additional parameters are held in  $\boldsymbol{\theta}_{(1)}$ . Then a  $D_s$ -optimal design maximizes

$$\phi_{D_s}(\xi) = |(\mathbf{K}'\mathbf{M}_\xi^- \mathbf{K})^{-1}|,$$

where  $\mathbf{K}' = [\mathbf{I}_s \mathbf{0}_{s \times (m-s)}]$  and  $\mathbf{M}_\xi^-$  denotes a generalized inverse of the information matrix. The blocks in  $\mathbf{K}'$  are the identity matrix of size  $(s \times s)$  and the zero matrix of size  $(s \times (m-s))$ , respectively.

This criterion is motivated by the likelihood ratio test for the null hypothesis

$$H_0 : \mathbf{K}'\boldsymbol{\theta}_{(1)} = 0.$$

Since a  $D_s$ -optimal design minimizes the confidence ellipsoid for  $\mathbf{K}'\hat{\boldsymbol{\theta}}_{(1)}$  it is expected that using such a design will result in a high power for this test.

To discriminate between more than two nested models, compound or constrained criteria can be used. A compound criterion, see e.g. [43], optimizes a combination of  $l$  objective functions of the form  $\phi(\xi) = \phi_1(\xi)^{\beta_1} \times \dots \times \phi_l(\xi)^{\beta_l}$ , where  $l \geq 1$  is an integer, and the weights  $\beta_1, \dots, \beta_l$  sum to one. The weights are chosen to reflect the importance of each criterion, and the objective functions should be appropriately standardized to avoid some of them dominating the others just because they take values on a different scale. Constrained criteria optimize one objective function, subject to the constraints that the resulting design achieves at least given efficiencies for all the other criteria. A constrained optimal design does not necessarily exist for certain combinations of lower bounds for the efficiencies. An application of this method to a class of linear models can be found in [8].

Another popular optimality for model discrimination, which does not require the models to be nested, and is applicable directly to discriminate between more than two models, is  $T$ -

optimality; see [3, 4]. Suppose the aim of the experiment is to discriminate between models  $\eta_1(x, \boldsymbol{\theta}_1)$  and  $\eta_2(x, \boldsymbol{\theta}_2)$ . [3] suggest to fix a model,  $\eta_1(x, \boldsymbol{\theta}_1) = \eta_1(x)$ , say. A  $T$ -optimal design then maximizes the minimal deviation between the model  $\eta_1$  and the class of models defined by  $\eta_2$ , that is

$$\phi_T(\xi) = \inf_{\boldsymbol{\theta}_2 \in \Theta_2} \int_{\mathcal{X}} (\eta_1(x) - \eta_2(x, \boldsymbol{\theta}_2))^2 d\xi(x).$$

If the models are not nested, it may be difficult to decide which of them should be fixed, and thus assumed to be the “true” model. In this situation, a compound design for the two  $T$ -criteria with each model fixed in turn could be applied. If both models are linear and nested, and differ by a single term, the  $T$ -optimal design coincides with the  $D_s$ -optimal design for that term. For further properties of  $T$ -optimal designs in the context of approximation theory see [31]. Generally, analytical results for this criterion are hard to obtain, and usually optimal designs have to be found numerically.

A common drawback of both  $D_s$ - and  $T$ -optimal designs is that in some situations these designs have fewer support points than there are parameters in the larger model, so this model cannot be estimated if found preferable by the likelihood ratio test. Sometimes not even the smaller model can be estimated, see example below. In such a situation, again compound or constrained optimal designs, where the additional criteria are  $D$ -efficiencies for estimating each model, can be useful. If these turn out to be difficult to find, hybrid designs, which are weighted averages of optimal designs with respect to different criteria, can be a good compromise. Compound optimal designs combining  $T$ -optimality for model discrimination and  $D$ -optimality for estimation (also called  $DT$ -optimal designs) are described in [2], which also gives an overview of similar criteria used in the literature.

The optimal designs depend on the values of the unknown model parameters. For examples of discrimination designs for the Michaelis Menten model and exponential models, respectively, which have been made robust to parameter misspecifications, see e.g. [27] and [9].

For models with non-normal errors, [45] suggest an optimality criterion based on the Kullback Leibler distance, and show that this is consistent with  $T$ -optimality. To discriminate between different link functions for GLMs, see e.g. [58], who consider the difference in

deviances for the rival models.

**Example (continued).** We briefly discuss the discrimination problem between the Michaelis Menten and the Emax model with normally distributed errors. These are nested, so either the  $T$ - or the  $D_s$ -criterion can be used. For the  $T$ -criterion, an optimal design has to be found numerically. We will focus on  $D_s$ -optimality, since this criterion is based on the information matrix, so this will allow us to use results from previous sections.

For this example,  $\mathbf{K}' = (1, 0, 0)$ . Hence the  $D_s$ -criterion corresponds to the  $c$ -criterion for estimating  $\theta_0$  in the Emax model. From subsection 1.4.2 an optimal design  $\xi$  will have at most three support points, including the endpoints of  $\mathcal{X} = [0, B]$ , i.e.  $\xi = \{(0, w_1), (x_2, w_2), (B, 1 - w_1 - w_2)\}$  with  $x_2, w_1$  and  $w_2$  to be determined. Substituting this design into the objective function, yields that  $\phi_c(\xi) = 1/w_1$ , which is minimized for  $w_1 = 1$ . The optimal design is thus a one-point design in  $x_1 = 0$ .

The obvious drawback of this design is that it does not allow estimation of either of the two models. We consider hybrid designs, i.e. weighted averages of the  $D_s$ -optimal design for discrimination and the  $D$ -optimal designs in either model. In practice, the weighting is often selected to achieve certain values for the individual efficiencies. For example, if the resulting design has equal weights on the support points  $0, \theta_2 B / (2\theta_2 + B)$  and  $B$ , its  $D_s$ -efficiency is  $1/3$ , its  $D$ -efficiency for the Michaelis Menten model is  $2/3$ , and its  $D$ -efficiency for the Emax model is 1. Giving more weight to the point 0, e.g.  $1/2$ , and  $1/4$  to each of the other two support points, improves the efficiency for model discrimination to  $1/2$ , at the expense of reducing the  $D$ -efficiencies for estimating the Michaelis Menten and the Emax model, respectively, to  $1/2$  and  $0.945$ .

## 1.6 Parameter robust approaches

All design problems discussed so far have in common that the optimal designs found depend on at least some of the unknown model parameters. This leads to a chicken and egg situation: To get a good design, and thus precise estimates, we need to know the very quantities we

actually want to estimate from the data before these are collected. In this section, we will illustrate the effects of parameter misspecification when designing experiments, and then briefly discuss strategies to make designs robust. A more detailed investigation of this issue can be found in later chapters of this monograph.

**Example (continued).** Consider the Michaelis Menten model, and suppose an experiment was designed to be  $D$ -optimal for a specific value of  $\theta_2$ . However, the “true” value of this parameter is  $\theta_2^*$ . We can then find the  $D$ -efficiency of the misspecified design  $\xi$ , relative to the “correct”  $D$ -optimal design  $\xi^*$ . Substituting

$$\xi = \left\{ \left( \frac{\theta_2 B}{2\theta_2 + B}, 0.5 \right), (B, 0.5) \right\}, \quad \xi^* = \left\{ \left( \frac{\theta_2^* B}{2\theta_2^* + B}, 0.5 \right), (B, 0.5) \right\},$$

into the expression (1.12) for  $D$ -efficiency, we obtain that

$$\text{eff}_D(\xi) = \frac{\theta_2 \theta_2^* (\theta_2 + B) (\theta_2^* + B)}{[\theta_2 \theta_2^* + B(\theta_2 + \theta_2^*)/2]^2}.$$

Similarly, the  $D$ -efficiency of a misspecified  $D$ -optimal design  $\xi$  for the exponential model is given by

$$\text{eff}_D(\xi) = \frac{\theta_2^*}{\theta_2} e^{1 - \theta_2^*/\theta_2}.$$

Figure 1.8 shows  $D$ -efficiencies of the locally optimal designs for the Michaelis Menten model and the exponential model, respectively, when the value of  $\theta_2$  has been misspecified across a range of  $\theta_2 \in [0.1, 2]$ . In the upper panel, we see two scenarios for the Michaelis Menten model, where the true value,  $\theta_2^*$ , is 0.3 and 0.6, respectively. We see that the efficiencies appear to be reasonable even on this relatively wide range, with 0.764 and 0.628 the respective minimal efficiencies.

Figure 1.8 approximately here

The lower panel shows two scenarios for the exponential model, with  $\theta_2^*$  given by 0.6 and 1.2, respectively. Here, the drop in efficiency is dramatic. For example, if  $\theta_2^* = 1.2$ ,

but the experimenter designed the experiment for  $\theta_2 = 0.1$ , the efficiency of the  $D$ -optimal design is only 0.0002. This is intuitive, since the observations are taken at points 0 and 10, which is appropriate for a relatively slow decay with rate 0.1. However, if the true rate of decay is 1.2, the expected response decreases much faster than expected, and is almost zero at  $x = 10$ . Hence the design “misses” the “interesting” part of the experiment. This can also be seen from the information matrix, whose entries are almost zero for  $x = 10$  since  $e^{-1.2 \times 10} = 6.144 \times 10^{-6}$ , and thus  $x = 10$  provides almost zero information for the estimation of the model parameters.

These results show that parameter misspecification can be a serious issue, and robust designs are sought for experiments in practice. There are several different approaches, of which we will briefly introduce the four most common ones.

**Response-adaptive sequential/batch sequential experimentation.** If the nature of the experiment permits observations to be taken sequentially, future experimental conditions can be optimized based on the observations already made. Starting with an initial design  $\xi_0$ , which can e.g. be a locally optimal design, a robust design as described below, or an equidistant uniform design, we take some observations, from which the parameter vector  $\theta$  is estimated. This estimate,  $\hat{\theta}_1$  say, is then substituted into the objective function, to find the design  $\xi_1$  for the next stage, such that the combined design  $\xi_0 + \xi_1$  optimizes  $\phi(\xi, \hat{\theta}_1)$ . One or a whole batch of observations will be made according to the design  $\xi_1$ , from which an updated estimate for  $\theta$  is obtained. This procedure is repeated until the total sample size has been reached.

It is expected that by adopting this strategy the quality of the design, and thus the estimate, can be improved successively. However, there are a few drawbacks. Observations at experimental conditions from e.g.  $\xi_1$  depend on the estimate  $\hat{\theta}_1$  obtained from previous observations, i.e. the data from a sequential experiment are dependent, making inference more difficult. For many situations, it has been shown that under mild conditions the estimators obtained from such experiments are asymptotically efficient, and that the sequential designs converge to the locally optimal design for the “true” parameter value; see e.g. [52]. [20] show analytically that one can expect a benefit from two-stage adaptive designs over

non-adaptive designs for sufficiently large sample sizes. However, for small sample sizes the adaptive design may still be outperformed by non-adaptive designs, in particular if the initial design has been chosen poorly. Another open question in this context is how to choose the number of batches and observations per batch in the sequential procedure. Usually extensive simulations are required prior to experimentation to determine a good strategy.

**Bayesian/Pseudo-Bayesian designs.** Bayesian (also called fully Bayesian) designs are optimized for Bayesian inference, and are beyond the scope of this chapter. In the frequentist literature, a Bayesian (often called Pseudo-Bayesian)  $\phi$ -optimal design optimizes an objective function of the form

$$\int \phi(\xi, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (1.14)$$

where  $\phi(\xi, \boldsymbol{\theta})$  is the objective function of a local optimality criterion, and  $\pi(\boldsymbol{\theta})$  is a prior distribution summarizing the available knowledge for  $\boldsymbol{\theta}$ . This means that the local objective function  $\phi(\xi, \boldsymbol{\theta})$  is averaged over the plausible values for  $\boldsymbol{\theta}$ . The prior  $\pi(\boldsymbol{\theta})$  is specified solely for the purpose of finding a design that performs reasonably well across its domain, and is not used for data analysis. For further reading on relationships and differences of fully Bayesian and Pseudo-Bayesian designs, we refer to the review paper [13]. This topic will also be covered in more detail in a later chapter of this monograph.

A potential problem with Pseudo-Bayesian designs is the choice of prior distribution. If the domain is far from the “true” parameter value the same problems as for locally optimal designs arise. Moreover, even if the “true” value of  $\boldsymbol{\theta}$  is contained in the domain of  $\pi(\boldsymbol{\theta})$  it is not guaranteed that it can be estimated efficiently, since other values of the parameter vector may dominate the weighted average in (1.14).

Pseudo-Bayesian  $D$ -optimal designs for the Michaelis Menten model are found in [46]. Note that for numerical computation of a Pseudo-Bayesian optimal design the integral in (1.14) is usually replaced by a finite sum, which approximates the integral. For guidance on the choice of values for  $\boldsymbol{\theta}$  to be used in the summation, see e.g. [37].

**Standardized maximin optimal designs.** This approach is more cautious than the Pseudo-Bayesian, and addresses the problem of possibly low design efficiency in some regions

within the domain of the prior  $\pi(\boldsymbol{\theta})$  by optimizing the design for the worst case scenario. Moreover, it is not necessary to specify a prior distribution on  $\boldsymbol{\theta}$ , but only a plausible range  $\Theta$ .

Let  $\phi(\xi, \boldsymbol{\theta})$  be the objective function of a local optimality criterion, which without loss of generality must be maximized. Then a standardized maximin  $\phi$ -optimal design maximizes

$$\min_{\boldsymbol{\theta} \in \Theta} \frac{\phi(\xi, \boldsymbol{\theta})}{\phi(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta})}, \quad (1.15)$$

where  $\xi_{\boldsymbol{\theta}}^*$  is the locally  $\phi$ -optimal design for the value  $\boldsymbol{\theta}$ . Note that if  $\phi(\xi, \boldsymbol{\theta})$  were to be minimized the numerator and denominator in (1.15) would change places. The standardization with the maximal value,  $\phi(\xi_{\boldsymbol{\theta}}^*, \boldsymbol{\theta})$ , was first suggested by [18] to make the optimality criterion independent of the scale of  $\phi(\xi, \boldsymbol{\theta})$ , which can be considerably affected by the value of  $\boldsymbol{\theta}$ . For many local optimality criteria, the ratio in (1.15) is the  $\phi$ -efficiency of the design  $\xi$  by definition, or a one-to-one and onto map to this quantity. For example, for  $D$ -optimality, the ratio would be raised to the power  $1/m$  to obtain the  $D$ -efficiency as defined in (1.12).

One drawback of standardized maximin optimal designs is that they do not necessarily perform well if the “true” parameter value for  $\boldsymbol{\theta}$  is outside the range  $\Theta$  specified by the experimenter. If it is, we have a lower bound for the  $\phi$ -efficiency of the design for each value of  $\boldsymbol{\theta} \in \Theta$ . However, the efficiencies of standardized maximin optimal designs, in particular for large ranges  $\Theta$ , tend to be flat, i.e. may be close to the lower bound across the whole range. Another issue, preventing more widespread use of these designs is that they are usually very hard to compute. Examples where standardized maximin  $D$ -optimal designs with minimum support are found analytically for the Michaelis Menten model, the exponential model and polynomials with several different efficiency functions are given in [19], [40] and [7], respectively.

**Cluster designs.** These designs are used as an alternative if Pseudo-Bayesian and standardized maximin optimal designs are difficult to compute. A sample of  $J$  ( $J$  large, e.g. 1000) values of the parameter vector  $\boldsymbol{\theta}$  is drawn according to a prior distribution  $\pi(\boldsymbol{\theta})$ . A clustering algorithm is then applied to the support points of the corresponding  $J$  locally op-

timal designs. The cluster design is formed by taking the centroids of the resulting clusters as equally-weighted support points.

This basic method was first introduced in [32], and later modified by [12] to take non-equal weights of the locally optimal designs into account. Cluster designs are easy to compute, but suffer potentially from the same drawbacks as Pseudo-Bayesian designs. There is no general rule on how to select the number of support points for these designs. This is currently done through summary statistics for efficiencies relative to a large number of locally optimal designs, and simulations on a case by case basis.

## 1.7 Summary

We have outlined the most popular methods for finding optimal designs for nonlinear models, and illustrated them through examples. Some methods are particularly useful in specific situations. The general strategy, however, is as follows: First apply one of the “new” methods described in section 1.3 to identify a simple class for the optimal design. Secondly, select an appropriate optimality criterion and use one of the “classical” methods to further narrow down this class (if possible). Finally, use an optimization procedure, either analytically or numerically, to find the optimal design. Some powerful new algorithms for numerical calculation of optimal designs have been suggested in the recent literature, see e.g. [64] for  $D$ -optimal designs and [60] for  $\phi_p$ -optimal designs, the latter of which is also capable of finding response adaptive optimal designs and optimal designs for subsets or functions of the unknown model parameters.

## References

- [1] I. Alvarez, R. Virto, J. Raso and S. Condon, “Comparing predicting models for the *Escherichia coli* inactivation by pulsed electric fields”, *Innovative food science and emerging technologies*, vol. 4, pp. 195-202, 2003.



- [2] A.C. Atkinson, “ $DT$ -optimum designs for model discrimination and parameter estimation,” *J. Statist. Plann. Inf.*, vol. 138, pp. 56-64, 2008.
- [3] A.C. Atkinson and V.V. Fedorov, “The design of experiments for discriminating between two rival models,” *Biometrika*, vol. 62, pp. 57-70, 1975.
- [4] A.C. Atkinson and V.V. Fedorov, “Optimal design: Experiments for discriminating between several models,” *Biometrika*, vol. 62, pp. 289-303, 1975.
- [5] M. Becka, H.M. Bolt, and W. Urfer, “Statistical analysis of toxicokinetic data by non-linear regression (example: inhalation pharmacokinetics of propylene),” *Arch. toxicol.*, vol. 66, pp. 450-453, 1992.
- [6] M. Becka, H.M. Bolt, and W. Urfer, “Statistical evaluation of toxicokinetic data,” *Environmetrics*, vol. 4, pp. 311-322, 1993.
- [7] S. Biedermann and H. Dette, “A note on maximin and Bayesian  $D$ -optimal designs in weighted polynomial regression,” *Mathematical Methods of Statistics*, vol. 12, pp. 358-370, 2003.
- [8] S. Biedermann, H. Dette and P. Hoffmann, “Constrained optimal discriminating designs for Fourier regression models,” *Ann. Inst. Statist. Math.*, vol. 61, pp. 143-157, 2009.
- [9] S. Biedermann, H. Dette and A. Pepelyshev, “Optimal discrimination designs for exponential regression models,” *J. Statist. Plann. Inf.*, vol. 137, pp. 2579-2592, 2007.
- [10] S. Biedermann, H. Dette and D.C. Woods, “Optimal design for additive partially non-linear models,” *Biometrika*, vol. 98, pp. 449-458, 2011.
- [11] S. Biedermann, H. Dette and W. Zhu, “Optimal designs for dose-response models with restricted design spaces,” *J. Amer. Statist. Assoc.*, vol. 101, pp. 747-759, 2006.
- [12] S. Biedermann and D.C. Woods, “Optimal designs for generalised nonlinear models with application to second harmonic generation experiments,” *J. Roy. Statist. Soc., Ser. C*, vol. 60, pp. 281-299, 2011.
- [13] K. Chaloner and I. Verdinelli, “Bayesian experimental design: A review,” *Statistical Science*, vol. 10, pp. 273-304, 1995.
- [14] H. Chernoff, “Locally optimal designs for estimating parameters,” *Ann. Math. Statist.*, vol. 24, pp. 586-602, 1953.
- [15] A. de la Garza, “Spacing of information in polynomial regression,” *Ann. Math. Statist.*, vol. 25, pp. 123-130, 1954.

- [16] H. Dette, “Elfving’s Theorem for  $D$ -optimality,” *The Annals of Statistics*, vol. 21, pp. 753-766, 1993.
- [17] H. Dette, “A note on Bayesian  $c$ - and  $D$ -optimal designs in nonlinear regression models,” *The Annals of Statistics*, vol. 24, pp. 1225-1234, 1996.
- [18] H. Dette, “Designing experiments with respect to standardized optimality criteria,” *J. Roy. Statist. Soc., Ser. B*, vol. 59, pp. 97-110, 1997.
- [19] H. Dette and S. Biedermann, “Robust and efficient designs for the Michaelis-Menten model,” *J. Amer. Statist. Assoc.*, vol. 98, pp. 679-686, 2003.
- [20] H. Dette, B. Bornkamp and F. Bretz, “On the efficiency of adaptive designs,” *preprint*, 2010.  
[www.statistik.tu-dortmund.de/sfb823-dp2010.html](http://www.statistik.tu-dortmund.de/sfb823-dp2010.html)
- [21] H. Dette, F. Bretz, A. Pepelyshev and J. Pinheiro, “Optimal designs for dose-finding studies,” *J. Amer. Statist. Assoc.*, vol. 103, pp. 1225-1237, 2008.
- [22] H. Dette and L.M. Haines, “ $E$ -optimal designs for linear and nonlinear models with two parameters,” *Biometrika*, vol. 81, pp. 739-754, 1994.
- [23] H. Dette, C. Kiss, M. Bevanda and F. Bretz, “Optimal designs for the emax, log-linear and exponential models,” *Biometrika*, vol. 97, pp. 513-518, 2010.
- [24] H. Dette and V.B. Melas, “A note on the de la Garza phenomenon for locally optimal designs,” *The Annals of Statistics*, vol. 39, pp. 1266-1281, 2011.
- [25] H. Dette, V.B. Melas and A. Pepelyshev, “Local  $c$ - and  $E$ -optimal designs for exponential regression,” *Ann. Inst. Statist. Math.*, vol. 58, pp. 407-426, 2006.
- [26] H. Dette, V.B. Melas and W.K. Wong, “Locally  $D$ -optimal designs for exponential regression models,” *Statistica Sinica*, vol. 16, pp. 789-803, 2006.
- [27] H. Dette, V.B. Melas and W.K. Wong, “Optimal design for goodness-of-fit of the Michaelis-Menten enzyme kinetic function,” *J. Amer. Statist. Assoc.*, vol. 100, pp. 1370-1381, 2005.
- [28] H. Dette and W.J. Studden, “Geometry of  $E$ -optimality,” *The Annals of Statistics*, vol. 21, pp. 416-433, 1993.
- [29] H. Dette and W.J. Studden, “A geometric solution of the Bayesian  $E$ -optimal design problem,” in *Statistical Decision Theory and Related Topics V*, eds. S.S. Gupta and J.O. Berger, New York: Springer-Verlag, pp. 157-170, 1993.

- [30] H. Dette and W.J. Studden, "The theory of canonical moments with applications in statistics, probability and analysis," Wiley, New York, 1997.
- [31] H. Dette and S. Titoff, "Optimal discrimination designs," *The Annals of Statistics*, vol. 37, pp. 2056-2082, 2009.
- [32] H.A. Dror and D.M. Steinberg, "Robust experimental design for multivariate generalized linear models," *Technometrics*, vol. 48, pp. 520-529, 2006.
- [33] G. Elfving, "Optimum allocation in linear regression theory," *The Annals of Mathematical Statistics*, vol. 23, pp. 255-262, 1952.
- [34] S.M. Ermakov and V.B. Melas, "Design and Analysis of Simulation Experiments," Kluwer Academic Publishers, Dordrecht, London, 1995.
- [35] I. Ford, B. Torsney and C. F. J. Wu, "The use of a canonical form in the construction of locally optimal designs for non-linear problems," *J. Roy. Statist. Soc., Ser. B*, vol 54, pp. 569-583, 1992.
- [36] M. Gibaldi and D. Perrier, "Pharmacokinetics," Marcel Dekker, New York, 1982.
- [37] C.M. Gotwalt, B.A. Jones and D.M. Steinberg, "Fast computation of designs robust to parameter uncertainty for nonlinear settings," *Technometrics*, vol. 51, pp. 88-95, 2009.
- [38] L.M. Haines, "Optimal designs for nonlinear regression models," *Communications in Statistics, Part A - Theory and Methods*, vol. 22, pp. 1613-1627, 1993.
- [39] P. D. H. Hill, " $D$ -optimal designs for partially nonlinear regression models," *Technometrics*, vol. 22, pp. 275-276, 1980.
- [40] L. Imhof, "Maximin designs for exponential growth models and heteroscedastic polynomial models," *The Annals of Statistics*, vol. 29, pp. 561-576, 2001.
- [41] S. Karlin and W.J. Studden, "Tchebycheff systems: With applications in analysis and statistics," Pure and applied mathematics series, vol. XV, Wiley, New York, 1966.
- [42] A. I. Khuri, "A note on  $D$ -optimal designs for partially nonlinear regression models," *Technometrics*, vol. 26, pp. 59-61, 1984.
- [43] E. Läuter, "Experimental design in a class of models," *Math. Operationsforsch. Stat.*, vol. 5, pp. 379-398, 1974.
- [44] T.S. Lau and W.J. Studden, "On an extremal problem of Fejer," *J. Approx. Theory*, vol. 53, pp. 184-194, 1988.
- [45] J. López-Fidalgo, C. Tommasi and P. C. Trandafir, "An optimal experimental design

- criterion for discriminating between nonnormal models,” *J. Roy. Statist. Soc., Ser. B*, vol. 69, pp. 231-242, 2007.
- [46] J.N. Matthews and G.C. Allcock, “Optimal designs for Michaelis-Menten kinetic studies,” *Stat. Med.*, vol. 23, pp. 477-491, 2004.
- [47] V.B. Melas, “Functional approach to optimal experimental design,” *Lecture Notes in Statistics*, Springer, 2006.
- [48] F. Pukelsheim, “On linear regression designs which maximize information,” *J. Statist. Plann. Inf.*, vol. 4, pp. 339-364, 1980.
- [49] F. Pukelsheim, “Optimal design of experiments,” John Wiley & Sons, Inc., New York, 1993.
- [50] F. Pukelsheim and S. Rieder, “Efficient rounding of approximate designs,” *Biometrika*, vol. 79.4, pp. 763-770, 1992.
- [51] F. Pukelsheim and B. Torsney, “Optimal weights for experimental designs on linearly independent support points,” *The Annals of Statistics*, vol. 19, pp. 1614-1625, 1991.
- [52] A. Roy, S. Ghosal and W. Rosenberger, “Convergence properties of sequential Bayesian  $D$ -optimal designs,” *J. Statist. Plann. Inf.*, vol. 139, pp. 425-440, 2008.
- [53] R. Sibson, “Contribution to discussion of papers by H. P. Wynn and P. J. Laycock,” *J. Roy. Statist. Soc., Ser. B*, vol. 34, pp. 181-183, 1972.
- [54] S.D. Silvey, “Contribution to discussion of papers by H. P. Wynn and P. J. Laycock,” *J. Roy. Statist. Soc., Ser. B*, vol. 34, pp. 174-175, 1972.
- [55] S.D. Silvey, “Optimal Design,” Chapman and Hall, 1980.
- [56] S. D. Silvey and D. M. Titterington, “A geometric approach to optimum design theory,” *Biometrika*, vol. 60, pp. 21-32, 1973.
- [57] G. Szegő, “Orthogonal Polynomials,” 4th edition, American Mathematical Society Colloquium Publications, 23, Providence, 1975.
- [58] T.H. Waterhouse, D.C. Woods, D.C. Eccleston and S.M. Lewis, “Design selection criteria for discrimination/estimation for nested models and a binomial response,” *J. Statist. Plann. Inf.*, vol. 138, pp. 132-144, 2008.
- [59] M. Yang, “On the de la Garza phenomenon,” *The Annals of Statistics*, vol. 38, pp. 2499-2524, 2010.
- [60] M. Yang and S. Biedermann, “A new algorithm for deriving optimal designs: general

and efficient,” *preprint*, 2012.

- [61] M. Yang and J. Stufken, “Support points of locally optimal designs for nonlinear models with two parameters,” *The Annals of Statistics*, vol. 37, pp. 518-541, 2009.
- [62] M. Yang and J. Stufken, “Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences,” *preprint*, 2011.
- [63] M. Yang, B. Zhang and S. Huang, “Optimal designs for generalized linear models with multiple design variables,” *Statistica sinica*, vol. 21, pp. 1415-1430, 2011.
- [64] Y. Yu, “ $D$ -optimal designs via a cocktail algorithm,” *Statistics and Computing*, vol. 21, pp. 475-481, 2011.

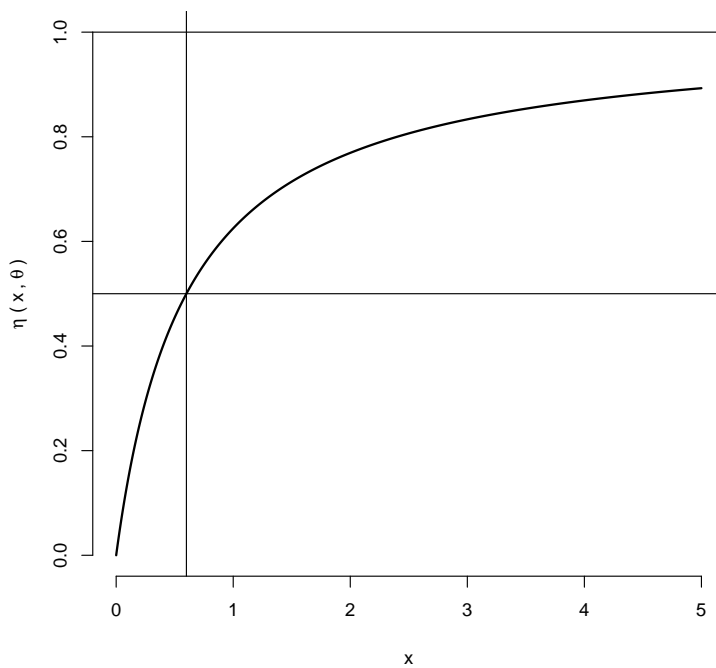


Figure 1.1: Plot of the expected response  $\eta(x, \theta)$  for the Michaelis Menten model with parameter vector  $\theta' = (1, 0.6)$ . For  $x \rightarrow \infty$ ,  $\eta(x, \theta)$  is asymptoting at  $\theta_1 = 1$ , and half of its supremum is attained at  $x = \theta_2 = 0.6$

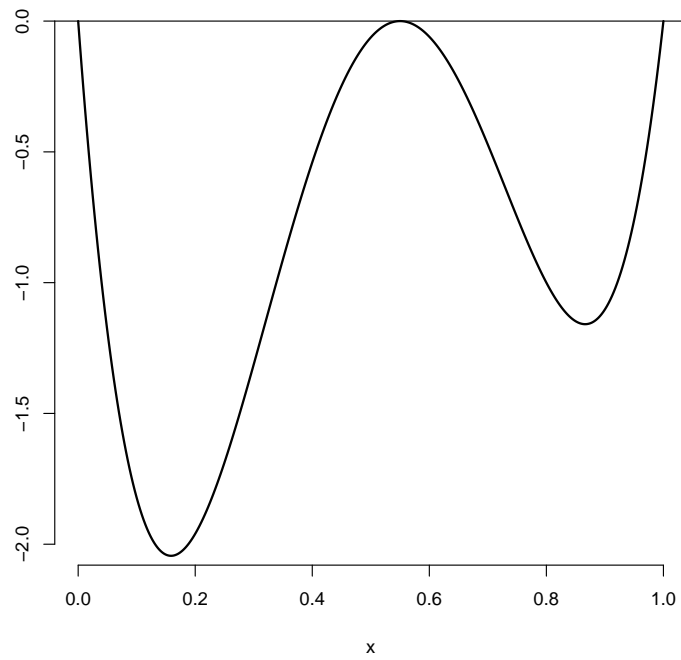


Figure 1.2: Plot of a polynomial of degree four, which is nonpositive on  $[0, 1]$  and attains its maximum, zero, at three points

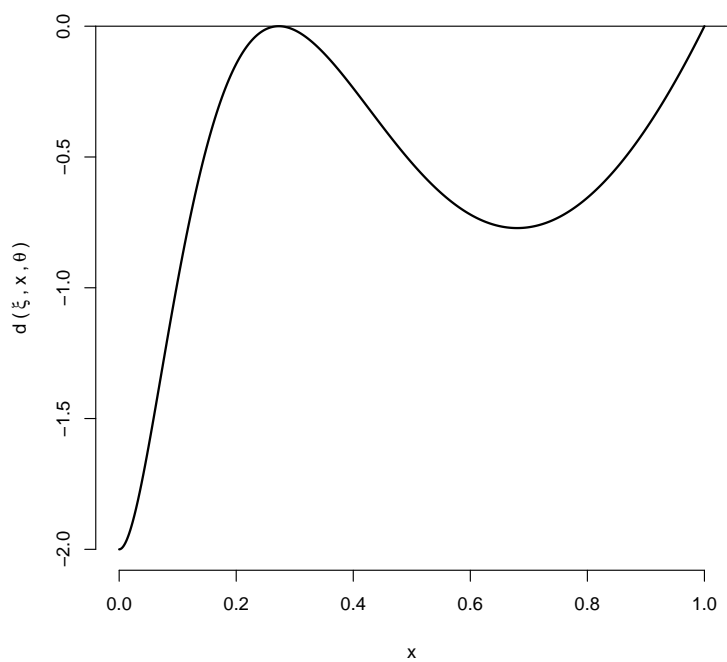


Figure 1.3: Plot of  $d(\xi, x, \theta)$  for the Michaelis-Menten model with parameter vector  $\theta' = (1, 0.6)$  and the  $D$ -optimal design  $\xi$  on the design space  $\mathcal{X} = [0, 1]$



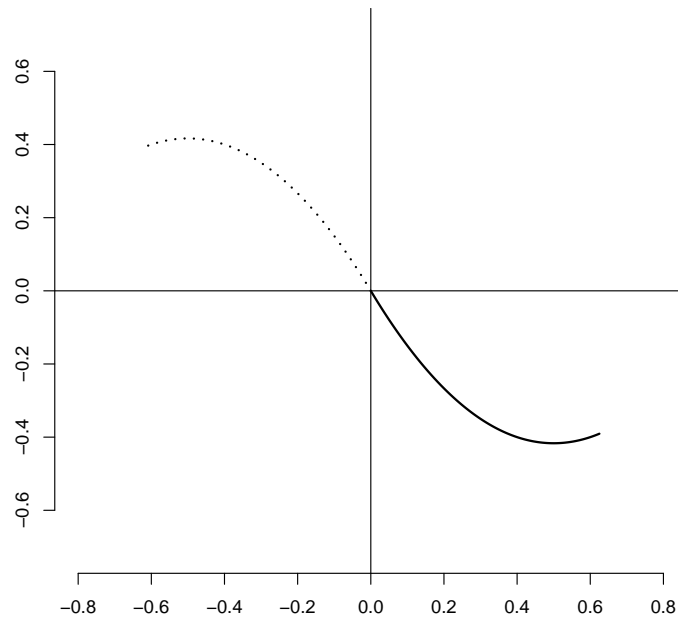


Figure 1.4: Parametric plot of the induced design space  $\mathcal{G}$  (solid line) for the Michaelis-Menten model with parameter vector  $\boldsymbol{\theta}' = (1, 0.6)$  and design space  $\mathcal{X} = [0, 1]$ , and its reflection  $-\mathcal{G}$  (dotted line). Horizontal axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_1$ , vertical axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_2$

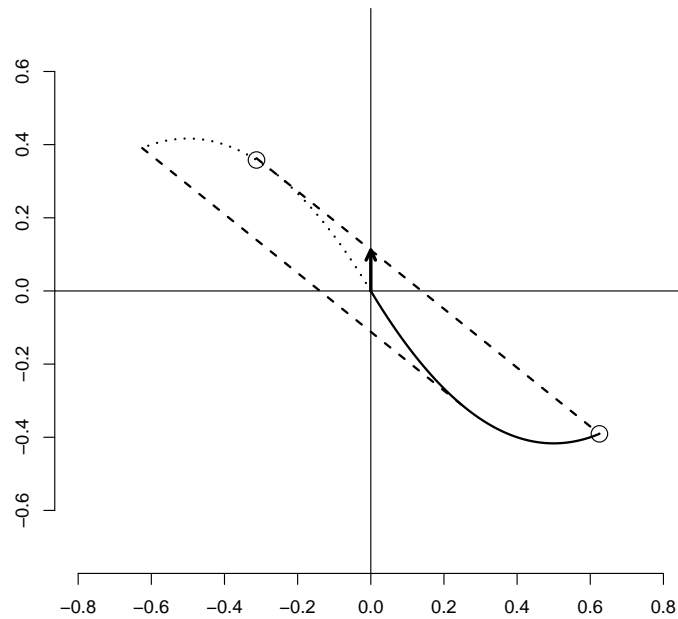


Figure 1.5: Parametric plot of the Elfving set,  $\mathcal{E}$ , for the Michaelis-Menten model with parameter vector  $\boldsymbol{\theta}' = (1, 0.6)$  and design space  $\mathcal{X} = [0, 1]$ . Solid line: induced design space  $\mathcal{G}$ , dotted line: its reflection  $-\mathcal{G}$ , dashed line: convex hull of  $\mathcal{G} \cup -\mathcal{G}$ , vertical arrow: vector  $\mathbf{c}$ , circles:  $c$ -optimal support points (or their reflections). Horizontal axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_1$ , vertical axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_2$

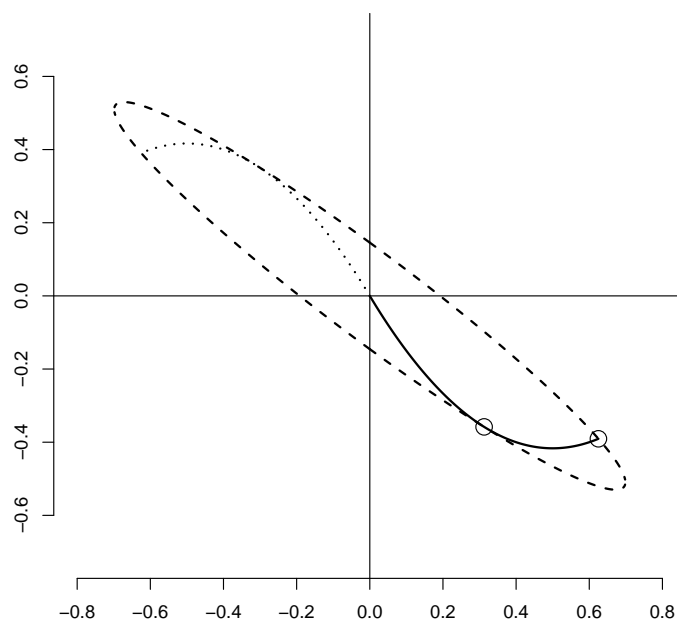


Figure 1.6: Parametric plot of the induced design space  $\mathcal{G}$  (solid line) for the Michaelis-Menten model with parameter vector  $\boldsymbol{\theta}' = (1, 0.6)$  and design space  $\mathcal{X} = [0, 1]$ , its reflection  $-\mathcal{G}$  (dotted line), with  $D$ -smallest covering ellipse (dashed line) and  $D$ -optimal support points. Horizontal axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_1$ , vertical axis:  $\partial\eta(x, \boldsymbol{\theta})/\partial\theta_2$

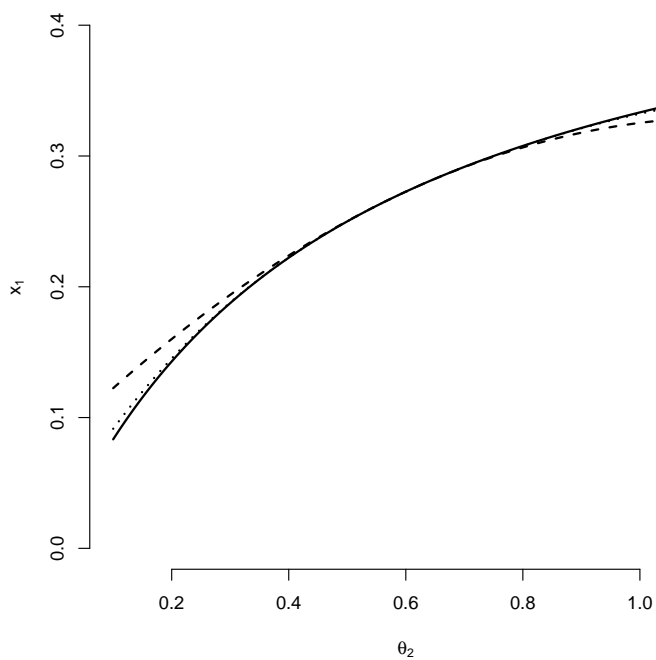


Figure 1.7: Taylor series approximations to the smaller support point,  $x_1$ , of the  $D$ -optimal design for the Michaelis Menten model with design space  $\mathcal{X} = [0, 1]$  in terms of the parameter  $\theta_2$ . Solid line: True curve, dotted line: Taylor polynomial of degree four, dashed line: Taylor polynomial of degree two. The Taylor series are centered about  $\theta_2 = 0.6$

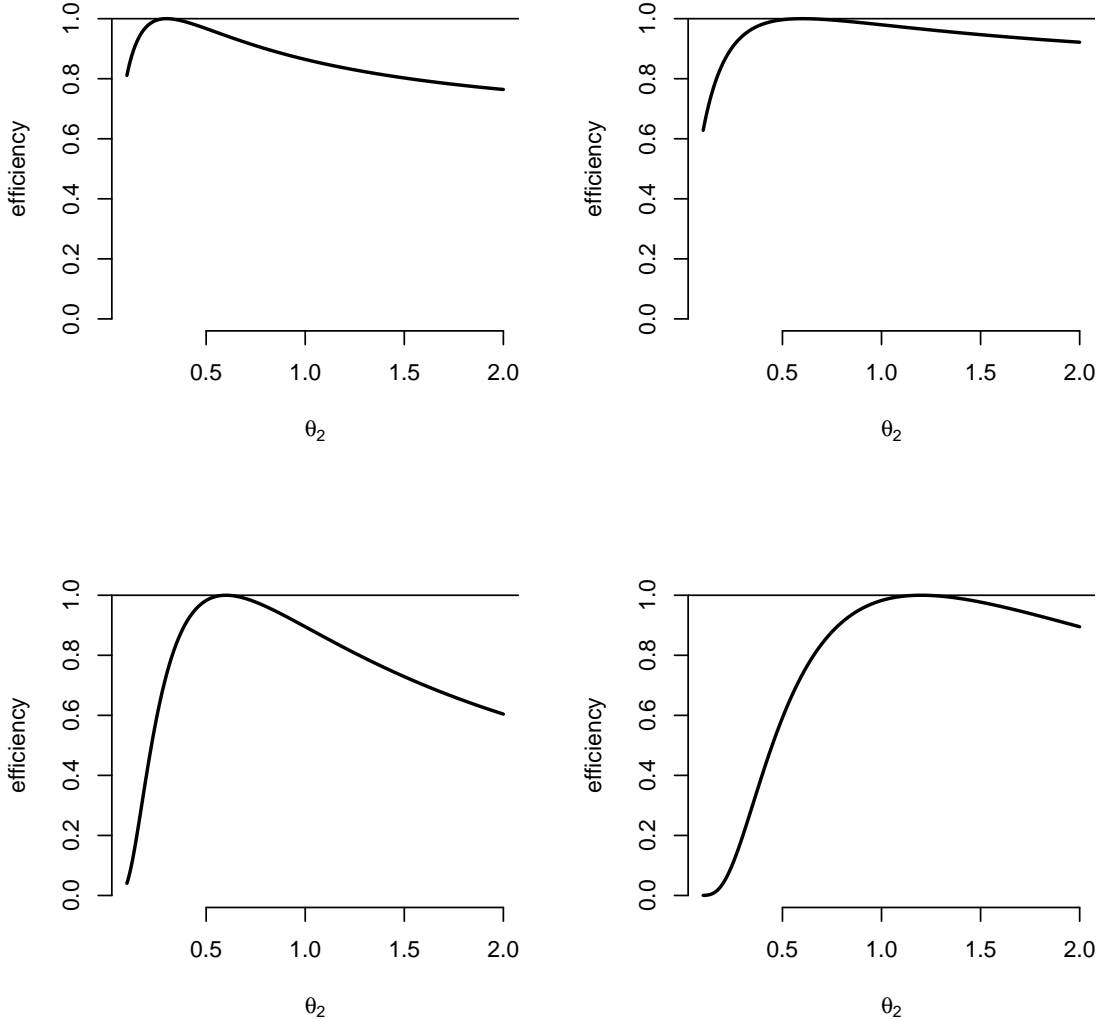


Figure 1.8: Upper panel: Efficiencies of the locally  $D$ -optimal designs for the Michaelis Menten model with parameter  $\theta_2 \in [0.1, 2]$ . Left: True parameter  $\theta_2^* = 0.3$ . Right: True parameter  $\theta_2^* = 0.6$ . Lower panel: Efficiencies of the locally  $D$ -optimal designs for the exponential model with parameter  $\theta_2 \in [0.1, 2]$ . Left: True parameter  $\theta_2^* = 0.6$ . Right: True parameter  $\theta_2^* = 1.2$