Chapter 4

Optimal Designs for Generalized Linear Models

John Stufken and Min Yang

4.1 Introduction

Both HK1 and HK2 deal with experiments in which the planned analysis is based on a linear model. Selecting designs for such experiments remains a critically important problem. However, there are many problems for which a linear model may not be a great choice. For example, if the response is a binary variable or a count variable rather than a continuous measurement, a linear model may be quite inappropriate. Experiments with such response variables are quite common. For example, in an experiment to compare different dosages of a drug, the outcome may be success (the dosage worked for a particular patient) or failure (it did not work). A design would consist of selecting the different dosages to be used in the experiment and the number of patients to be assigned to the selected dosages. How can one identify a good design for such a problem in which a linear model for the binary response is simply inadequate?

Another feature of the example in the previous paragraph is that the purpose of the experiment is not to compare different treatments, but to understand the relationship between the response and the dosage of the drug. Most of the chapters in HK1 and HK2 are devoted to comparative experiments in which various treatments are to be compared to each other. Typically, each treatment is associated with a treatment effect, and the

The research of John Stufken is supported by NSF grants DMS-0706917 and DMS-1007507. The research of Min Yang is supported by NSF grants DMS-0707013 and DMS-0748409.

purpose of the experiment is some form of comparison of these effects. An exception to this is Chapter 12 in HK1, in which the purpose of the experiment is to understand the relationship between the response and one or more regression variables. This latter scenario will also be the focus of the current chapter.

As noted in HK1, Section 2.10, there are many considerations that can and should go into the selection of an appropriate design for a particular experiment. These could be of a general scientific nature, of a purely statistical nature, or of an entirely practical nature, and are almost always a combination of considerations of all types. The single consideration in this chapter is that of identifying a design that is *optimal* according to a specified statistical optimality criterion. Some criteria of this type were briefly presented in HK2, Section 1.13 in the context of block designs. Thus, as a result of additional experiment-specific considerations, the designs identified in this chapter may not be the designs used in experiments. But even if one does not use a design that is optimal under a specific optimality criterion, the optimal design will still provide a benchmark under this criterion for all other designs. Thus we may be willing to sacrifice optimality for other desirable properties of a design provided that the design that we do use has still a reasonable performance with respect to the optimality criterion (or possibly with respect to multiple optimality criteria). Without knowing how to identify optimal designs, we will have no basis to assess whether a given design performs well under the criterion of interest.

The three previous paragraphs set the stage for this chapter. As a concise summary, we will provide a brief introduction to some recent results on finding optimal designs for experiments in which the focus is on studying the relationship between a response and one or more regression variables through a special class of nonlinear models, namely *Generalized Linear Models* (GLMs).

GLMs have found wide applicability in many fields, including drug discovery, clinical trials, social sciences, marketing, and many more. Methods of analysis and inference for these models are well established (see for example McCullagh and Nelder, 1989; McCulloch and Searle, 2001; and Agresti, 2002). The study of optimal designs for experiments that plan to use a GLM is however not nearly as well developed (see also Khuri, Mukherjee, Sinha and Ghosh, 2006), and tends to be much more difficult than the corresponding and better studied problem for the special case of linear models. (While linear models are a special case of GLMs, this chapter focuses on GLMs that correspond to nonlinear models.)

One of the challenges is that for a GLM an optimal design typically depends on the unknown parameters. This leads to the concept of locally optimal designs, which are optimal for *a priori* chosen values of the parameters. The designs may be poor if the choice of values is far from the true parameter values. Where feasible, a multistage approach could help with this, in which a small initial design is used to obtain some information about the parameters. We will briefly return to this later, but simply state for now that results presented in this chapter are also applicable for a multistage approach (see also Yang and

Stufken, 2009).

To illustrate some of the basic ideas we present a small example. Many of these ideas will be revisited in more detail in later sections.

Example 4.1. In dose-response studies and growth studies, a subject receives a stimulus at a certain level x. The binary response indicates whether the subject does or does not respond to the stimulus. The purpose of such an experiment is often to study the relationship between the probability that a subject responds and the level of the stimulus. The level can be controlled by the experimenter, and a judicious selection of the levels must be made prior to the experiment.

With Y_i and x_i denoting the response and stimulus level for the ith subject, we could use a logistic regression model of the form

$$logit[P(Y_i = 1)] = \alpha + \beta x_i, \tag{4.1}$$

where the *logit* function represents the log of the odds ratio. We consider two designs (without claiming that either is a good design). Design I uses level 0.067 for 53% of the subjects and level 0.523 for the other 47%. Design II uses each of the levels 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6 equally often, namely for 16.7% of the subjects. Thus we will think of a design as a probability measure on the possible levels. Such a design is also known as an *approximate design*. Whether it corresponds to a design that can be used in practice depends in part on the number of subjects in the experiment. For example, Design I can be used with 100 subjects (53 of them at level 0.523 and 47 at level 0.067), but couldn't be used exactly with, for example, only 30 subjects. Thus, for an experiment with 30 subjects, there is no *exact design* corresponding to Design I. An exact design corresponding to Design II will only exist if the number of subjects is a multiple of 6.

While we can only use an exact design in practice, the discreteness of the design space for a fixed number of subjects makes it difficult to identify optimal designs in such a space. Working with approximate designs circumvents this difficulty, but at the expense that we might not be able to use an exact design that corresponds precisely to an optimal approximate design.

Continuing with the example, which of the two designs is best?

As stated, this question cannot be answered. First, we need to specify what we mean by "best". We will do this by using an optimality criterion. For example, we could compare the designs in terms of $Var(\hat{\alpha}) + Var(\hat{\beta})$, where $\hat{\alpha}$ and $\hat{\beta}$ denote the maximum likelihood estimators (MLEs) for the unknown parameters α and β. We might then want to select a design that minimizes this criterion. But it is not quite that simple. For GLMs considered here, these variances depend on the unknown parameters. How the two designs compare in terms of $Var(\hat{\alpha}) + Var(\hat{\beta})$, or other commonly used criteria, will thus depend on the true, but unknown values of the parameters. This leads to the concept of *locally optimal designs*.

The experimenter offers one or more "guesses" for the true values of the parameters, and at each of these we can compare the two designs. The first may be better for some guessed values, the second for others. For any guess of α and β, the best design among all possible designs is said to be locally optimal for that guess.

Continuing with our example, suppose that we have guessed $\alpha = -3$ and $\beta = 10$. We will therefore compare the two designs under the assumption that these are the true parameter values. Under this assumption, it can be shown that $Var(\hat{\alpha})+Var(\hat{\beta})$ for Design I is 40% smaller than for Design II. This means that with Design I we need only 60% of the number of subjects that would be needed for Design II in order to achieve the same efficiency (in terms of $Var(\hat{\alpha}) + Var(\hat{\beta})$).

Moreover, it can be shown that Design I remains more efficient than Design II under this optimality criterion if our guess for the true parameter values is slightly off.

This example shows that judicious selection of a design can make a big difference, but also shows that the problem of selecting a good or optimal design is a fairly complicated one.

One of the difficulties compared to linear models is that GLMs present a much broader class of models. While linear models are all of the form $E(Y) = X\beta$, often accompanied by assumptions of independence and normality, in GLMs the form of the relationship between the response and the regression variables depends on a *link function* and on distributional assumptions for Y, which vary from one model to the next. This means that it is very difficult to establish unifying and overarching results, and that the mathematics becomes more difficult and messier. Until recently, many of the available results to identify optimal designs were obtained via the so-called *geometric approach*. This method is inspired by the pioneering work of Elfving (1952) for linear models. But for a long time it meant that results could only be obtained on a case-by-case basis. Each combination of model, optimality criterion and objective required its own proof. A number of very fine papers have appeared over the years that address optimal designs for several such combinations. But there is a limit to what can be handled with this approach.

Another difficulty, as already alluded to in Example 4.1, is that design comparisons depend on the unknown parameters. The principal cause for this is that the Fisher information matrix for the parameters in a GLM tends to depend on the parameters. Thus the challenge in designing an experiment for such a model is that one would like to find the best design for estimating the unknown parameters, and yet one has to know the parameters to find the best design. As explained in the context of the example, one way to solve this problem is to use locally optimal designs, which are based on the best guess of the parameters. Even when a good guess is not available, knowing locally optimal designs is valuable because they provide benchmarks for all other designs. An alternative that will not be pursued in this paper is to adopt a Bayesian approach, which naturally facilitates the inclusion of some uncertainty about a guess for the unknown parameters. This also requires some prior knowledge about the parameters, and may only lead to a computational solution for specific problems without providing broader insights.

Another way to deal with the dependence on unknown parameters is, whenever feasible, through a multistage approach (see Silvey, 1980; Sitter and Forbes, 1997). In a multistage approach we would start with a small initial design to obtain some information about the unknown parameters. This information is then used at the next stage to estimate the true parameter values and to augment the initial design in an optimal way. The resulting augmented design could be the final design, or there could be additional stages at which more design points are added. This is also a mathematically difficult problem, but as we will see it is not more difficult than using a single-stage approach. Since the multistage approach uses information obtained in earlier stages, one may hope that it will lead in the end to a better overall design than can be obtained with the single-stage approach.

Khuri, Mukherjee, Sinha, and Ghosh (2006) surveyed design issues for GLMs and noted (p. 395) that "The research on designs for generalized linear models is still very much in its developmental stage. Not much work has been accomplished either in terms of theory or in terms of computational methods to evaluate the optimal design when the dimension of the design space is high. The situations where one has several covariates (control variables) or multiple responses corresponding to each subject demand extensive work to evaluate "optimal" or at least efficient designs."

There are however a number of recent papers that have made major advances in this area. These include Biedermann, Dette and Zhu (2006), Yang and Stufken (2009), and (on nonlinear models in general) Yang (2010). These papers tackle some of the aforementioned difficulties, the first by further exploring the geometric approach and the other two by a new analytic approach. These papers convincingly demonstrate that unifying results for multiple models, multiple optimality criteria and multiple objectives can be obtained in the context of nonlinear models. While these papers do provide many answers, they also leave many open questions, especially with regard to slightly more complicated models.

This chapter will provide an introduction to the general problem and a peek at available results. The emphasis will be on the analytic approach. In Section 4.2, we introduce notation and basic concepts such as the information matrix and optimality criteria. Some tools, including Kiefer's equivalence theorem, Elfving's geometric approach, and the new analytic approach are presented in Section 4.3. Some optimality results for the simplest GLMs are introduced in Section 4.4. In Section 4.5, we study GLMs with multiple covariates and with block effects. We conclude with some brief remarks in Section 4.6.

4.2 Notation and Basic Concepts

While GLMs can be appropriate for many types of data, the focus in this chapter is on binary and count data. Let Y denote the response variable. In a GLM, a *link function* G

relates $E(Y)$ to a linear combination of the regression variables, i.e., $G(E(Y)) = \boldsymbol{X}^T\boldsymbol{\theta}$.

4.2.1 Binary Data

If $Y_1, \ldots, Y_n$ denote the binary response variables for n subjects, and $x_{i1}, \ldots, x_{ip}$ denote the values of p regression variables for subject i, then a class of GLMs can be written as

$$Prob(Y_i = 1) = P(\boldsymbol{X}_i^T\boldsymbol{\theta}). \tag{4.2}$$

Here the superscript T denotes matrix transposition, $\boldsymbol{X}_i = (1, x_{i1}, \ldots, x_{ip})^T$, the vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)^T$ contains the unknown regression parameters, and $P(x)$ is a cumulative distribution function (cdf). Popular choices for the latter include $P(x) = e^x/(1+e^x)$ for the logistic model as in (4.1) and $P(x) = \Phi(x)$, the cdf of the standard normal distribution, for the probit model. Other choices include the double exponential and double reciprocal models. The inverse of $P(x)$ is the *link function* for these GLMs, i.e. $P^{-1}(Prob(Y_i = 1)) = \boldsymbol{X}_i^T\boldsymbol{\theta}$. For the logistic model, this link function corresponds to the *logit* function of Example 4.1.

The likelihood function for $\boldsymbol{\theta}$ under (4.2) can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} P(\boldsymbol{X}_i^T\boldsymbol{\theta})^{Y_i}\left(1 - P(\boldsymbol{X}_i^T\boldsymbol{\theta})\right)^{(1-Y_i)}. \tag{4.3}$$

The resulting likelihood equations are

$$\sum_{i=1}^{n} \boldsymbol{X}_i \frac{[Y_i - P(\boldsymbol{X}_i^T\boldsymbol{\theta})]P'(\boldsymbol{X}_i^T\boldsymbol{\theta})}{P(\boldsymbol{X}_i^T\boldsymbol{\theta})\left(1 - P(\boldsymbol{X}_i^T\boldsymbol{\theta})\right)} = \boldsymbol{0}. \tag{4.4}$$

The MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, is obtained by numerically solving these nonlinear equations, such as in statistical software packages as SAS and SPSS. The likelihood function can also be used to obtain asymptotic covariance matrices for functions of $\hat{\boldsymbol{\theta}}$ that are of interest. To do this we will need a generalized inverse of the Fisher information matrix. The information matrix can be computed as

$$E\left(-\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) = \sum_{i=1}^{n} \boldsymbol{I}_{\boldsymbol{X}_i} = \sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i^T \frac{[P'(\boldsymbol{X}_i^T\boldsymbol{\theta})]^2}{P(\boldsymbol{X}_i^T\boldsymbol{\theta})\left(1 - P(\boldsymbol{X}_i^T\boldsymbol{\theta})\right)}. \tag{4.5}$$

$\boldsymbol{I}_{\boldsymbol{X}_i}$ is the information matrix for $\boldsymbol{\theta}$ at a single design point $\boldsymbol{X}_i$.

4.2.2 Count Data

For count data $Y_1, \ldots, Y_n$, we will assume the model for Y_i to be a Poisson regression model with mean λ_i. Using the same notation as in Subsection 4.2.1 and using the logarithm as the link function, we have

$$log(\lambda_i) = \boldsymbol{X}_i^T\boldsymbol{\theta}. \tag{4.6}$$

The likelihood function for $\boldsymbol{\theta}$ under (4.6) can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\exp(-\lambda_{\mathrm{i}})\lambda_{\mathrm{i}}^{\mathrm{Y_i}}}{Y_i!}, \tag{4.7}$$

which results in the likelihood equations

$$\sum_{i=1}^{n} \boldsymbol{X}_i \left(Y_i - \exp(\boldsymbol{X}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{\theta})\right) = \mathbf{0}. \tag{4.8}$$

The MLE for $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, can again be obtained numerically by solving these nonlinear equations. The information matrix for θ under Model (4.6) can be written as

$$E\left(-\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) = \sum_{i=1}^{n} \boldsymbol{I}_{\boldsymbol{X}_i} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \exp(\boldsymbol{X}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{\theta}). \tag{4.9}$$

4.2.3 Optimality Criteria

For an exact design with a total of n subjects we must select (1) distinct vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$, as defined in Subsection 4.2.1, in a design space, say $\mathcal{X}$; and (2) the number of subjects, n_i, to be assigned to $\boldsymbol{X}_i$ so that $n = \sum_{i=1}^{k} n_i$. The $\boldsymbol{X}_i$'s are also called the *support points* for the design. The optimal exact design problem is to make such selections so that the resulting design is best with respect to a certain optimality criterion. As already alluded to in Example 4.1, instead of this typically intractable exact design problem, the corresponding approximate design problem is considered. Thus we would like to find a design $\xi = \{(\boldsymbol{X}_i, \omega_i), i = 1, \ldots, k\}$, where the ω_i's are nonnegative weights that sum to 1. Thus ω_i represents the proportion of subjects that are to be assigned to $\boldsymbol{X}_i$. The corresponding information matrix for $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{I}_\xi = \sum_{i=1}^{k} \omega_i \boldsymbol{I}_{\boldsymbol{X}_i}, \tag{4.10}$$

where $\boldsymbol{I}_{\boldsymbol{X}_i}$ is again the information matrix for $\boldsymbol{\theta}$ for a one-point design that only uses $\boldsymbol{X}_i$. If there is an exact design for n subjects corresponding to ξ (which requires $n\omega_i$ to be integral for all i), then the information matrix for this exact design is n times the matrix shown in (4.10).

The interest of the experimenter may not always be in $\boldsymbol{\theta}$, but could be in a vector function of $\boldsymbol{\theta}$, say $g(\boldsymbol{\theta})$. By the Delta method, the approximate covariance matrix of $g(\hat{\boldsymbol{\theta}})$ under design ξ is equal to

$$\boldsymbol{\Sigma}_\xi = \left(\frac{\partial g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right) \boldsymbol{I}_\xi^- \left(\frac{\partial g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right)^T, \tag{4.11}$$

where $\boldsymbol{I}_\xi^-$ is a generalized inverse of the information matrix in (4.10). In some situations we could be interested in a function $g(\boldsymbol{\theta})$ for which the matrix $\boldsymbol{\Sigma}_\xi$ is singular for any design ξ. For example, this would happen if the elements of $g(\boldsymbol{\theta})$ are linearly dependent, such as for $g(\boldsymbol{\theta}) = (\theta_1, \theta_2, \theta_1+\theta_2)^T$. We will however limit our interest in this chapter to functions $g(\boldsymbol{\theta})$ for which there are designs that make $\boldsymbol{\Sigma}_\xi$ nonsingular. In particular, if we are interested in $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$ or if g is a one-to-one function, the parametrization should be such that there are designs for which $\boldsymbol{I}_\xi$ is nonsingular.

For designs for which the inverse of the matrix in (4.11) exists, this inverse is the information matrix for $g(\boldsymbol{\theta})$ under design ξ. We would like to select a design that, in some sense, maximizes this information matrix, or minimizes the covariance matrix. The following are some of the more prominent optimality criteria that suggest how we might want to do this. For the first three of these, the minimization is over those designs ξ for which $\boldsymbol{\Sigma}_\xi$ is nonsingular. For the fourth criterion, ξ should be a design so that the vector $\boldsymbol{c}$ is not in the null space of $\boldsymbol{\Sigma}_\xi$.

- *D-optimality.* A design is D-optimal for $g(\boldsymbol{\theta})$ if it minimizes $|\boldsymbol{\Sigma}_\xi|$ over all possible designs. Such a design minimizes the expected volume of the asymptotic $100(1-\alpha)\%$ joint confidence ellipsoid for the elements of $g(\boldsymbol{\theta})$. For a one-to-one transformation $h(\boldsymbol{\theta})$ of $g(\boldsymbol{\theta})$, if ξ is D-optimal for $g(\boldsymbol{\theta})$ then the same holds for $h(\boldsymbol{\theta})$. Thus D-optimality is invariant under such transformations. Many other optimality criteria do not have this property.

- *A-optimality.* A design is A-optimal for $g(\boldsymbol{\theta})$ if it minimizes $Tr(\boldsymbol{\Sigma}_\xi)$ over all possible designs. Such a design minimizes the sum of the asymptotic variances of the estimators of the elements of $g(\boldsymbol{\theta})$.

- *E-optimality.* A design is E-optimal for $g(\boldsymbol{\theta})$ if it minimizes the largest eigenvalue of $\boldsymbol{\Sigma}_\xi$ over all possible designs. Such a design minimizes the expected length of the longest semi-axis of the asymptotic $100(1-\alpha)\%$ joint confidence ellipsoid for the elements of $g(\boldsymbol{\theta})$.

- *c-optimality.* A design is c-optimal for $g(\boldsymbol{\theta})$ if it minimizes $\boldsymbol{c}^T\boldsymbol{\Sigma}_\xi\boldsymbol{c}$ over all possible designs, where $\boldsymbol{c}$ is a vector of the same length as $g(\boldsymbol{\theta})$. Such a design minimizes the asymptotic variance of $\boldsymbol{c}^T g(\hat{\boldsymbol{\theta}})$.

To facilitate simultaneous study of some of the above and additional criteria, Kiefer (1974) introduced, among others, the class of functions

$$\Phi_p(\boldsymbol{\Sigma}_\xi) = \left[\frac{1}{v} Tr\left((\boldsymbol{\Sigma}_\xi)^p\right)\right]^{1/p}, \ 0 < p < \infty. \tag{4.12}$$

Here v is the dimension of $\boldsymbol{\Sigma}_\xi$. A design is Φ_p*-optimal* for $g(\boldsymbol{\theta})$ if it minimizes $\Phi_p(\boldsymbol{\Sigma}_\xi)$ over all possible designs. In addition, we define $\Phi_0(\boldsymbol{\Sigma}_\xi) = \lim_{p\downarrow 0} \Phi_p(\boldsymbol{\Sigma}_\xi)$ and $\Phi_\infty(\boldsymbol{\Sigma}_\xi) =$

$\lim_{p\to\infty} \Phi_p(\boldsymbol{\Sigma}_\xi)$. Obviously, Φ_1-optimality is equivalent to A-optimality. It can be shown that Φ_0-optimality corresponds to D-optimality and Φ_∞-optimality to E-optimality.

Which optimality criterion one should use may depend on the objective of the experiment, but also on personal preference (see also HK2, Section 1.13.4). One compromise could be to find a design that performs well under multiple criteria. In order to assess whether a design performs well under different criteria, one would however first need to know optimal designs under these criteria. In this chapter, we will only focus on tools for identifying optimal designs.

To emphasize an earlier observation, the information matrix $\boldsymbol{I}_\xi$ and the covariance matrix $\boldsymbol{\Sigma}_\xi$ depend, for the GLMs considered here, on $\boldsymbol{\theta}$. Thus none of the above minimizations can be carried out, unless we have a "guess" for $\boldsymbol{\theta}$. This approach, resulting in locally optimal designs, is taken here.

4.3 Tools for Finding Locally Optimal Designs

Once we have decided on a model, a function $g(\boldsymbol{\theta})$ of interest, an optimality criterion, and a guess for the parameter values, we are ready to identify a locally optimal design by optimizing the objective function corresponding to the selected criterion. (We will often drop the adjective "locally" from hereon, but the reader should remember that the discussion in this chapter is always about locally optimal designs.) However, this is a very challenging problem. How many support points do we need in an optimal design? What are those points? What are the corresponding weights? Directly optimizing the objective function is generally not feasible because the objective function is too complicated and there are too many variables. Moreover, even if a purely numerical optimization were feasible, it might not provide enough inside into the structure and general features of optimal designs. There are two standard traditional tools for studying optimal designs that have inspired many researchers: Elfving's geometric approach and Kiefer's equivalence theorem. The emphasis in this chapter is however on a new analytical approach, although we will give the reader also a flavor of the traditional tools.

4.3.1 Traditional Approaches

The geometric approach proposed by Elfving (1952) for linear models has had a profound impact on optimal design theory. Whereas Elfving was interested in c- and A-optimal experimental designs for linear models in two dimensions, his work has proven to be inspirational for the development of optimal design theory in a much broader framework.

To describe the basic idea, write the information matrix for $\boldsymbol{\theta}$ as $\sum_i \omega_i f(\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{X}_i, \boldsymbol{\theta})^T$, where $f(\boldsymbol{X}_i, \boldsymbol{\theta})$ is a column vector that depends on the design point $\boldsymbol{X}_i$ and the parameter vector $\boldsymbol{\theta}$. Note that the information matrices in (4.5) and (4.9), or more precisely those for the corresponding approximate designs, are of that form with $f(\boldsymbol{X}, \boldsymbol{\theta})$ being a multiple of $\boldsymbol{X}$. Let $\mathcal{G}$ (which depends on $\boldsymbol{\theta}$) be the space that is generated through $f(\boldsymbol{X}, \boldsymbol{\theta})$ by letting $\boldsymbol{X}$ take all possible values in the design space $\mathcal{X}$. Assume that $\mathcal{G}$, which is called the *induced design space*, is closed and bounded. The support points for an optimal design must lie on the "smallest ellipsoid" centered at the origin that contains $\mathcal{G}$. The definition of "smallest ellipsoid" depends on the optimality criterion but can be stated explicitly. Studying these ellipsoids and their intersections with the induced design space provides therefore a method for determining possible support points for an optimal design. Many optimality results in the literature are based on variations of this approach. Some seminal contributions include Ford, Torsney, and Wu (1992) for c- and D-optimal designs; Dette and Haines (1994) for E-optimal designs; and Biedermann, Dette, and Zhu (2006) for Φ_p-optimal designs. We refer the reader to these articles and their references to learn more about this approach.

A second result that has had a profound impact on research in optimal design is the equivalence theorem. In a seminal contribution, Kiefer and Wolfowitz (1960) derived an equivalence theorem for D-optimality. Later, Kiefer (1974) presented a more general result that applies for Φ_p optimality. Pukelsheim (2006) contains very general results on equivalence theorems. All of these studies focus on optimality for linear functions of the parameters under linear models. However, once a value of $\boldsymbol{\theta}$ is fixed under the local optimality approach, then the results extend under mild conditions. The following result is formulated in the spirit of Pukelsheim (2006).

Theorem 4.1 (Equivalence theorem)**.** *Suppose that the information matrix for $\boldsymbol{\theta}$ under a design $\xi = \{(\boldsymbol{X}_i, \omega_i), i = 1, \ldots, k\}$ is given by $\sum_i \omega_i f(\boldsymbol{X}_i, \boldsymbol{\theta}) f(\boldsymbol{X}_i, \boldsymbol{\theta})^T$. A design ξ^* is Φ_p-optimal for $g(\boldsymbol{\theta})$, $0 \le p < \infty$, if and only if there exists a generalized inverse $\boldsymbol{I}_{\xi^*}^{-}$ of $\boldsymbol{I}_{\xi^*}$ so that*

$$f(\boldsymbol{X}, \boldsymbol{\theta})^T \boldsymbol{I}_{\xi^*}^{-} \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)^T (\boldsymbol{\Sigma}_{\xi^*})^{p-1} \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right) \boldsymbol{I}_{\xi^*}^{-} f(\boldsymbol{X}, \boldsymbol{\theta}) \le Tr\left((\boldsymbol{\Sigma}_{\xi^*})^p\right), \text{ for all } \boldsymbol{X} \in \mathcal{X}, \tag{4.13}$$

where $\boldsymbol{\Sigma}_\xi$ is as defined in (4.11) and $\mathcal{X}$ is the design space. Equality in (4.13) holds if and only if $\boldsymbol{X}$ is a support point for a Φ_p-optimal design.

The equivalence theorem is very useful to verify whether a candidate design is indeed optimal. For the case that $p = 0$, corresponding to D-optimality, the righthand side of (4.13) reduces simply to the dimension of $g(\boldsymbol{\theta})$.

4.3.2 An Analytical Approach

The equivalence theorem and the geometric approach are very powerful tools for studying optimal designs for GLMs. Nonetheless, there are many problems of great practical interest for which neither of these tools has, as of yet, helped to provide a solution. Yang and Stufken (2009) propose a new strategy for studying optimal designs for GLMs. The remainder of this chapter will discuss this strategy and will present some of the results that have been obtained by using it.

For a given model and design space, suppose that we can identify a subclass of designs, say Ξ, so that for any design $\xi \notin \Xi$, there is a design $\tilde{\xi} \in \Xi$, so that $\boldsymbol{I}_{\tilde{\xi}}(\boldsymbol{\theta}) \geq \boldsymbol{I}_{\xi}(\boldsymbol{\theta})$, i.e., the information matrix for $\boldsymbol{\theta}$ under $\tilde{\xi}$ dominates that under ξ in the Loewner ordering. Then $\tilde{\xi}$ is locally at least as good as ξ under the commonly used optimality criteria such as D-, A- and E-optimality, and more generally Φ_p-optimality. Moreover, if interest is in $g(\boldsymbol{\theta})$, design $\tilde{\xi}$ is also better than design ξ in the Loewner ordering since

$$\boldsymbol{\Sigma}_{\xi} = \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right) \boldsymbol{I}_{\xi}^{-} \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)^T \geq \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right) \boldsymbol{I}_{\tilde{\xi}}^{-} \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)^T = \boldsymbol{\Sigma}_{\tilde{\xi}}. \tag{4.14}$$

This would mean that the search for optimal designs could be restricted to a search in Ξ. We will refer to Ξ as a *complete class* for this problem. One trivial choice for Ξ is the class of all possible designs. This choice is not useful. In order for Ξ to be helpful, it needs to be a small class. For example, for some models Ξ might perhaps consist of "all designs with at most 3 support points". That would be an enormous reduction from having to consider all possible designs with any number of support points. In order to be useful, the approach must also work for any "guess" of $\boldsymbol{\theta}$ under the local optimality approach. Note also from our formulation that the choice of Ξ is not based on any of the common optimality criteria and does not depend on a particular choice for $g(\boldsymbol{\theta})$. So we would like to use the same complete class Ξ no matter what the guess for $\boldsymbol{\theta}$ is, no matter which function $g(\boldsymbol{\theta})$ we are interested in, and no matter which information-based optimality criterion we have in mind. Ξ will depend on the model and the design space.

Thus, for a given model and design space, the approach consists of identifying small complete classes Ξ.

Before we continue, we observe that this approach is also helpful for multistage experiments, where an initial experiment may be used to get a better idea about the unknown parameters. At a second or later stage, the question then becomes how to add more design points in an optimal fashion. If ξ_1 denotes the design used so far for n_1 design points, and we want to augment this design optimally by n_2 additional design points, then we are looking for an approximate design ξ_2 that maximizes the combined information matrix $n_1 I_{\xi_1} + n_2 I_{\xi_2}$. Because the first part of this matrix is fixed, if we have a complete class Ξ for the single-stage approach, it is immediately clear that we can restrict our choice for ξ_2 to Ξ in order to obtain a combined information matrix that can not be beaten in the Loewner ordering by a choice of ξ_2 that is not in Ξ. Thus Ξ is also a complete class for the multistage approach.

The strategy described in the previous paragraphs was used in Yang and Stufken (2009). They characterized complete classes Ξ for nonlinear models (including GLMs) with two parameters. We summarize their results here for GLMs only. For proofs and results for other nonlinear models with two parameters we refer the reader to Yang and Stufken (2009).

In the context of GLMs, and using the notation of Section 4.2, results are for two-parameter models with $\boldsymbol{X}_i^T\boldsymbol{\theta} = \theta_0 + \theta_1 x_{i1}$. We define $c_i = \theta_0 + \theta_1 x_{i1}$, and call this the *induced design point*. Note that under local optimality, using guessed values for θ_0 and θ_1, a design can be expressed in terms of choices for x_{i1} or, equivalently, in terms of the induced design points c_i.

For a design $\xi = \{(c_i, \omega_i), i = 1, ..., k\}$, write the information matrix for $\boldsymbol{\theta}$ as $\boldsymbol{I}_\xi = \sum_{i=1}^k \omega_i \boldsymbol{I}_{c_i}$, and write $\boldsymbol{I}_{c_i}$ as $\boldsymbol{A}^T\boldsymbol{C}(c_i)\boldsymbol{A}$. Here matrix $\boldsymbol{A}$ may depend on $\boldsymbol{\theta}$, but not on the induced design point c_i. We write the matrix $\boldsymbol{C}(c)$, which can depend on $\boldsymbol{\theta}$ and the induced design points, as

$$\boldsymbol{C}(c) = \begin{pmatrix} \Psi_1(c) & \Psi_2(c) \\ \Psi_2(c) & \Psi_3(c) \end{pmatrix}. \tag{4.15}$$

Define

$$F(c) = \Psi_1'(c)\left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)'\left(\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' \Big/ \left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)'\right)'. \tag{4.16}$$

We will assume that the design space, in terms of the induced design points, is an interval $[D_1, D_2]$ (where the endpoints can for some applications be $-\infty$ or ∞, respectively).

Theorem 4.2. *Assume that $F(c)$ is well-defined for $c \in [D_1, D_2]$. If $F(c) \le 0$ for all $c \in [D_1, D_2]$, then a complete class Ξ is obtained by taking all designs with at most two support points, with D_1 being one of them. If $F(c) \ge 0$ for all $c \in [D_1, D_2]$, then the class of all designs with at most two support points and D_2 being one of them forms a complete class.*

Theorem 4.2 requires $F(c)$ to be well defined and either positive or negative in the entire interval $[D_1, D_2]$. These requirements can be relaxed as stated in the following result.

Theorem 4.3. *For the matrix $\boldsymbol{C}(c)$ in (4.15), suppose that $\Psi_1(c) = \Psi_1(-c)$, $\Psi_2(c) = -\Psi_2(-c)$, and $\Psi_3(c) = \Psi_3(-c)$. Suppose further that $D_1 < 0 < D_2$ and that $F(c)$ is well-defined for $c \in (0, D_2]$. Let $F(c) < 0$ and $\Psi_1'(c)\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' < 0$ for $c \in (0, D_2]$. Then a complete class Ξ is obtained by taking all designs with at most two support points with the additional following restrictions: If $|D_1| = D_2$, the two points can be taken to be symmetric in the induced design space; if $|D_1| < D_2$, then the two points can be taken to be either symmetric or one of the points is taken as D_1 and the other in $(-D_1, D_2]$; if $|D_1| > D_2$, then the two points can be taken to be either symmetric or one of the points is taken as D_2 and the other in $[D_1, -D_2)$.*

For the proofs of these results we refer to Yang and Stufken (2009). The next sections will provide some applications of these results.

4.4 GLMs with Two Parameters

In this section we will focus on GLMs with $\boldsymbol{X}_i^T\boldsymbol{\theta} = \theta_0 + \theta_1 x_i$ for the binary data model (4.2) and the count data model (4.6). For simplicity of notation, we replace the parameters by α and β, that is we replace $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$ by $\boldsymbol{\theta} = (\alpha, \beta)^T$. By the expressions for the information matrix provided in (4.5) and (4.9), the information matrix for $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{I}_\xi = \sum_{i=1}^{k} \omega_i \Psi(\alpha + \beta x_i) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \tag{4.17}$$

where, $\xi = \{(x_i, \omega_i), i = 1, ..., k\}$. The function $\Psi(x)$ depends on the model and takes the following forms for the three models that we will focus on:

$$\Psi(x) = \begin{cases} \frac{e^x}{(1+e^x)^2} & \text{for the logistic model} \\ \frac{\phi^2(x)}{\Phi(x)(1-\Phi(x))} & \text{for the probit model} \\ e^x & \text{for the Poisson regression model,} \end{cases} \tag{4.18}$$

where $\phi(x)$ is the density function for the standard normal distribution.

These simple models have been studied extensively in the optimal design literature. For binary data, under the restriction of symmetry for the induced design space, Abdelbasit and Plackett (1983) identify a two-point D-optimal design for the logistic model. Minkin (1987) strengthens this result by removing the restriction on the design space. Using the geometric approach (see Subsection 4.3.1), Ford, Torsney, and Wu (1992) study c-optimal and D-optimal designs for this model. Using the same approach and model, Sitter and Wu (1993a, 1993b) study A- and F-optimal designs, while Dette and Haines (1994) investigate E-optimal designs. Mathew and Sinha (2001) obtain a series of optimality results for the logistic model by using an analytic approach, while Biedermann, Dette, and Zhu (2006) obtained Φ_p-optimal designs for a restricted design space using the geometric approach. Chaloner and Larntz (1989) and Agin and Chaloner (1999) study Bayesian optimal designs for the logistic model. For count data, Ford, Torsney, and Wu (1992) identified c- and D-optimal designs. Minkin (1993) studied optimal designs for $1/\beta$.

In this section, we will present optimal designs for the above models by using the analytic approach presented in Section 4.3. As we will see, this approach generalizes and extends most of the available results. Perhaps even more importantly, in the next section we will see that this approach also facilitates handling GLMs with more than two parameters.

From (4.17), the information matrix $\boldsymbol{I}_x$ for $(\alpha, \beta)^T$ under a one-point design with all weight at x can be written as

$$\boldsymbol{I}_x = \underbrace{\begin{pmatrix} 1 & 0 \\ -\alpha/\beta & 1/\beta \end{pmatrix}}_{\boldsymbol{A}^T} \begin{pmatrix} \Psi(c) & c\Psi(c) \\ c\Psi(c) & c^2\Psi(c) \end{pmatrix} \underbrace{\begin{pmatrix} 1 & -\alpha/\beta \\ 0 & 1/\beta \end{pmatrix}}_{\boldsymbol{A}}, \tag{4.19}$$

where $c = \alpha + \beta x$. Note that we are assuming that neither α nor β is equal to 0, as we will do throughout. Clearly, matrix $\boldsymbol{A}$ depends only on the parameters α and β, and not on x. Thus we can apply the analytic approach in Section 4.3 for these simple models.

We will first consider the Poisson regression model, where $\Psi_1(c) = e^c$, $\Psi_2(c) = ce^c$, and $\Psi_3(c) = c^2e^c$. We find that $\Psi_1'(c) = e^c$, $\left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)' = 1$, and $\left(\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' / \left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)'\right)' = 2$. Hence, the function $F(c)$ defined in (4.16) is positive in any interval. Applying Theorem 4.2, we reach the following conclusion.

Theorem 4.4. *For the Poisson regression model in (4.6) with* $c_i = \boldsymbol{X}_i^T\boldsymbol{\theta} = \alpha + \beta x_i$, *let* $c_i \in [D_1, D_2]$ *be the induced design space,* $D_1 < D_2 < \infty$. *Then the class of designs with at most two support points and* D_2 *being one of them forms a complete class.*

Thus Theorem 4.4 tells us that when searching for an optimal design, we need to look no further than the complete class described in the theorem. Given how simple the designs in this class are, it is easy to use a search algorithm to do this for a given function $g(\boldsymbol{\theta})$, a given optimality criterion, and a guess for the parameters. The algorithm would merely have to optimize a function of interest over two unknowns: the second support point and the weight for that point. Moreover, depending on the problem, it may actually be possible to derive an explicit form for the optimal design by using Theorem 4.4.

Example 4.2. By Theorem 4.4, for the Poisson regression model, a D-optimal design for $(\alpha, \beta)^T$ can be based on D_2 and $c < D_2$, where c needs to be determined. The weight for each of these support points must be equal to $\frac{1}{2}$ (see Silvey, 1980). Computing the determinant of the information matrix in (4.17), it is easily seen that the D-optimal design must maximize $e^{c/2}(D_2 - c)$. As a function of c, this is is an increasing function on $(-\infty, D_2 - 2]$ and a decreasing function on $[D_2 - 2, D_2]$. Thus, if $D_1 < D_2 - 2$, then a D-optimal design is given by $\xi = \{(D_2 - 2, 1/2), (D_2, 1/2)\}$. Otherwise, a D-optimal design is given by $\xi = \{(D_1, 1/2), (D_2, 1/2)\}$.

Theorem 4.4 unifies and extends a number of results that had already appeared in the literature. For example, Ford, Torsney, and Wu (1992) identified c- and D-optimal designs using the geometric approach. They showed that an optimal design has two support points and that one of them is D_2. Minkin (1993) studied optimal designs for $1/\beta$ for this model. He assumed $\beta < 0$ and used the induced design space $(-\infty, \alpha]$. He concluded that the optimal design has two support points, and that one of them is α.

Turning now to the logistic and probit models, both have the properties that $\Psi_1(c) = \Psi_1(-c)$, $\Psi_2(c) = -\Psi_2(-c)$, and $\Psi_3(c) = \Psi_3(-c)$. Theorems 4.2 and 4.3 can thus be applied as long as other conditions in the theorems are satisfied. For the logistic model it is easily seen that

$$
\begin{aligned}
\Psi_1'(c) &= -\frac{e^c(e^c-1)}{(e^c+1)^3}, \\
\left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)' &= \frac{e^{2c}+1}{(e^c-1)^2}, \\
\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' &= \frac{2((c-1)e^{2c}+c+1)}{(e^c-1)^2}, \\
\left(\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' \Big/ \left(\frac{\Psi_2'(c)}{\Psi_1'(c)}\right)'\right)' &= \frac{2(e^{2c}-1)^2}{(e^{2c}+1)^2}.
\end{aligned}
\tag{4.20}
$$

Hence, if the induced design space $[D_1, D_2] \subset (-\infty, 0]$, then $F(c) > 0$; if the induced design space $[D_1, D_2] \subset [0, \infty)$, then $F(c) < 0$; further, $\Psi_1'(c)\left(\frac{\Psi_3'(c)}{\Psi_1'(c)}\right)' < 0$ for $c > 0$. These same conclusions are also valid for the probit model, but we will skip the much more tedious details. Applying Theorems 4.2 and 4.3, we reach the following conclusion.

Theorem 4.5. *For the logistic or probit model as in (4.2) with $c_i = \boldsymbol{X}_i^T\boldsymbol{\theta} = \alpha + \beta x_i$, let $c_i \in [D_1, D_2]$ be the induced design space. Then the following complete class results hold:*

(i) If $D_2 \leq 0$, then the class of designs with at most two support points and D_2 being one of them forms a complete class;

(ii) if $D_1 \geq 0$, then the class of designs with at most two support points and D_1 being one of them forms a complete class;

(iii) if $D_2 = -D_1$, then the class of designs with at most two support points that are symmetric forms a complete class;

(iv) if $0 < -D_1 < D_2$, then the class of designs with at most two support points, where either one of the points is D_1 and the other point is larger than $-D_1$ or the two points are symmetric, forms a complete class;

(v) if $0 < D_2 < -D_1$, then the class of designs with at most two support points, where either one of the points is D_2 and the other point is smaller than $-D_2$ or the two points are symmetric, forms a complete class.

Example 4.3. By Theorem 4.5, for the logistic and probit models, D-optimal designs for $(\alpha, \beta)^T$ can be based on two support points, say c_1 and c_2. The weights for these support points must be equal to $\frac{1}{2}$ (see Silvey, 1980). Using (4.17), it is easily seen that the D-optimal design must maximize $(c_1 - c_2)^2\Psi(c_1)\Psi(c_2)$. Because of the relationship between c_1 and c_2 provided in Theorem 4.5, there is essentially one unknown to be decided. For example, if $D_2 = -D_1$, so that $c_1 = -c_2 = c$, say, we need to maximize $4c^2\Psi^2(c)$. For the logistic model, this is an increasing function for $c \in [0, 1.5434]$ and a decreasing function for $c > 1.5434$. (For the probit model the critical value for c is 1.1381.) Thus if $D_2 > 1.5434$, then a D-optimal design is given by $\xi = \{(-1.5434, 1/2), (1.5434, 1/2)\}$. Otherwise, a D-optimal design is given by $\xi = \{(-D_2, 1/2), (D_2, 1/2)\}$.

Example 4.4. Again using the logistic or probit model, suppose that $D_1 > 0$. By Theorem 4.5, we can take D_1 as one of the two support points, and with c, $D_1 < c \le D_2$, denoting the other support point we need to maximize $(c - D_1)^2\Psi(c)$ for a D-optimal design. This is an increasing function for $c \in [D_1, c^*]$ and a decreasing function when $c > c^*$. For the logistic model, c^* is the solution of $e^c(2 - c + D_1) + c - D_1 + 2 = 0$. If $c^* < D_2$, then a D-optimal design is given by $\xi = \{(D_1, 1/2), (c^*, 1/2)\}$. Otherwise, a D-optimal design is given by $\xi = \{(D_1, 1/2), (D_2, 1/2)\}$.

Example 4.5. Consider the logistic regression model in (4.2) with $c_i = \alpha + \beta x_i$. Suppose that we are interested in a locally D-optimal design for $\alpha = 2$ and $\beta = -1$, with the design space restricted to $x \in [0, 1]$. Then $[D_1, D_2] = [1, 2]$. We can apply the conclusion in Example 4.4 since $D_1 > 0$. Simple computation shows that $c^* = 3.1745$. Thus a D-optimal design in terms of the induced design space is given by $\xi = \{(1, 1/2), (2, 1/2)\}$, which is $\{(1, 1/2), (0, 1/2)\}$ in the original x-space.

In Subsection 4.5.2 we will return to finding D-optimal designs in a slightly more complicated setting. Here we show how Theorem 4.5 can be used to identify A-optimal designs for $(\alpha, \beta)^T$ under the logistic and probit models when there is no constraint on the design space, i.e., $D_1 = -\infty$ and $D_2 = \infty$. This question was posed by Mathew and Sinha (2001) and was studied by Yang (2008). By Theorem 4.5, we can focus on designs with two symmetric induced design points, i.e., $\xi = \{(x_1, \omega_1), (x_2, \omega_2)\}$, where $\alpha + \beta x_1 = -\alpha - \beta x_2$. With $c = \alpha + \beta x_1$, and using (4.17) and the observation that Ψ is an even function for these two models, it follows that the information matrix for $(\alpha, \beta)^T$ under ξ can be written as

$$\boldsymbol{I}_\xi = \Psi(c) \begin{pmatrix} 1 & \omega_1(\frac{c-\alpha}{\beta}) + \omega_2(\frac{-c-\alpha}{\beta}) \\ \omega_1(\frac{c-\alpha}{\beta}) + \omega_2(\frac{-c-\alpha}{\beta}) & \omega_1(\frac{c-\alpha}{\beta})^2 + \omega_2(\frac{-c-\alpha}{\beta})^2 \end{pmatrix}. \tag{4.21}$$

For a locally A-optimal design, we need to minimize the following trace as a function of c and ω_1 (with $\omega_2 = 1 - \omega_1$):

$$\begin{aligned} Tr\left(\boldsymbol{I}_\xi^{-1}\right) &= \frac{1 + \omega_1(\frac{c-\alpha}{\beta})^2 + \omega_2(\frac{-c-\alpha}{\beta})^2}{\left[\omega_1(\frac{c-\alpha}{\beta})^2 + \omega_2(\frac{-c-\alpha}{\beta})^2 - \left(\omega_1(\frac{c-\alpha}{\beta}) + \omega_2(\frac{-c-\alpha}{\beta})\right)^2\right]\Psi(c)} \\ &= \frac{[\beta^2 + (c+\alpha)^2]/\omega_1 + [\beta^2 + (c-\alpha)^2]/\omega_2}{4c^2\Psi(c)} \\ &\ge T^2(c). \end{aligned} \tag{4.22}$$

Here,

$$T(c) = \frac{\sqrt{\beta^2 + (c+\alpha)^2} + \sqrt{\beta^2 + (c-\alpha)^2}}{2c\sqrt{\Psi(c)}}. \tag{4.23}$$

Note that we do not have to worry about $c = 0$ since that corresponds to a 1-point design, which cannot be optimal for $(\alpha, \beta)^T$. The last inequality of (4.22) gives equality for

$$\omega_1 = 1 - \omega_2 = \frac{\sqrt{\beta^2 + (c+\alpha)^2}}{\sqrt{\beta^2 + (c+\alpha)^2} + \sqrt{\beta^2 + (c-\alpha)^2}} =: \omega_1(c). \tag{4.24}$$

We are now ready to present locally A-optimal designs under the logistic and probit models.

Theorem 4.6. *For the logistic or probit model as in (4.2) with* $\boldsymbol{X}_i^T\boldsymbol{\theta} = \alpha + \beta x_i$ *and an unrestricted design space, the design* $\xi^* = \{(x_1^*, \omega_1^*), (x_2^*, \omega_2^*)\}$ *is A-optimal for* $(\alpha, \beta)^T$*, where* $x_1^* = (c^* - \alpha)/\beta$*,* $x_2^* = (-c^* - \alpha)/\beta$*,* $\omega_1^* = \omega_1(c^*)$ *as defined in (4.24),* $\omega_2^* = 1 - \omega_1^*$*, and* c^* *is the only positive solution of the equation*

$$\frac{c^2 - \alpha^2 - \beta^2}{\sqrt{\beta^2 + (c+\alpha)^2}\sqrt{\beta^2 + (c-\alpha)^2}} = 1 + \frac{c\Psi'(c)}{\Psi(c)}. \tag{4.25}$$

Proof. Starting from (4.22) and (4.24), we will show that c^* is the unique positive design point that minimizes $T^2(c)$. The restriction to positive design points is justified because $T^2(c)$ is an even function. Since $T(c)$ in (4.23) is positive, we can instead focus on minimizing $T(c)$ for $c > 0$. By straightforward computations, we obtain that

$$T'(c) = \frac{T_1(c)}{2c^2},$$

where

$$\begin{aligned} T_1(c) = &\left(c\left(\Psi^{-1/2}(c)\right)' - \Psi^{-1/2}(c)\right)\left(\sqrt{\beta^2 + (c+\alpha)^2} + \sqrt{\beta^2 + (c-\alpha)^2}\right) \\ &+ \Psi^{-1/2}(c)\left(\frac{c^2 + \alpha c}{\sqrt{\beta^2 + (c+\alpha)^2}} + \frac{c^2 - \alpha c}{\sqrt{\beta^2 + (c-\alpha)^2}}\right). \end{aligned} \tag{4.26}$$

Taking the derivative of $T_1(c)$ we find that

$$\begin{aligned} T_1'(c) =& c\left(\Psi^{-1/2}(c)\right)''\left(\sqrt{\beta^2 + (c+\alpha)^2} + \sqrt{\beta^2 + (c-\alpha)^2}\right) \\ &+ 2c\left(\Psi^{-1/2}(c)\right)'\left(\frac{c+\alpha}{\sqrt{\beta^2 + (c+\alpha)^2}} + \frac{c-\alpha}{\sqrt{\beta^2 + (c-\alpha)^2}}\right) \\ &+ c\Psi^{-1/2}(c)\left(\frac{\beta^2}{(\sqrt{\beta^2 + (c+\alpha)^2})^3} + \frac{\beta^2}{(\sqrt{\beta^2 + (c-\alpha)^2})^3}\right). \end{aligned} \tag{4.27}$$

By studying the function $f(x) = x/\sqrt{\beta^2 + x^2}$, it is easily seen that for $c > 0$ it holds that

$$f(c + \alpha) + f(c - \alpha) > 0.$$

Combined with the observation that, for both the logistic and probit model $\Psi^{-1/2}(c)$ has positive first and second order derivatives for $c > 0$, we can conclude that $T_1'(c) > 0$. Thus $T_1(c)$ is an increasing function for $c > 0$ and has at most one positive root. It follows that $T'(c)$ also has at most one positive root. But since $T(c)$ goes to $+\infty$ for both $c \downarrow 0$ and $c \to \infty$, $T'(c)$ must also have at least one positive root, which must thus be unique and must minimize $T(c)$. It is fairly straightforward to show that $T_1(c) = 0$ is equivalent to (4.25). The result follows. □

Example 4.6. Consider the probit model in (4.2) with $c_i = \alpha + \beta x_i$. Suppose that we are interested in a locally A-optimal design for $\alpha = 1$, $\beta = 2$, and that there is no restriction on the design space. Using (4.25) with these values for α and β and with the choice of Ψ for the probit model as in (4.18), we obtain that c^* in Theorem 4.6 is equal to 1.3744. It follows that an A-optimal design is given by $\{(0.1872, 0.6041), (-1.1872, 0.3959)\}$ in the x-space.

4.5 GLMs with Multiple Parameters

Whereas a model with only two parameters is adequate for some applications, other situations call for models with more parameters. For example, the subjects in an experiment may have different characteristics that can be captured by one or more qualitative variables, such as race, gender or age category. To allow for differences in the relationship between the response variable and a covariate for subjects belonging to different groups, effects associated with these qualitative variables could be included in the model in addition to a covariate such as stimulus level (see, for example, Tighiouart and Rogatko, 2006). As another example, an experimenter may be able to control more than one covariate that has a relationship with the response variable. Optimal design results for GLMs with more than two parameters are relatively rare. Selected results include Sitter and Torsney (1995a, 1995b), who study D-optimal designs for binary data with two and more covariates. Under certain constraints, Haines, Kabera, Ndlovu, and O'Brien (2007) study D-optimal designs for logistic regression in two variables. Russell, Woods, Lewis, and Eccleston (2009) consider multivariate GLMs for count data. Computationally oriented contributions include Woods, Lewis, Eccleston, and Russell (2006) and Dror and Steinberg (2006) for studying robust designs as well as Dror and Steinberg (2008) for studying sequential designs. They all use D-optimality and provide algorithms for finding desirable designs. For example, Dror and Steinberg (2006) provide computer programs for deriving D-optimal designs for general models. This works best for smaller values of p.

In this section, we will systematically study (i) GLMs with multiple independently chosen covariates; and (ii) GLMs with group effects.

4.5.1 GLMs with Multiple Covariates

We will focus on the model in (4.2) for a binary response. We will again take $\boldsymbol{X}_i^T = (1, x_{i1}, \ldots, x_{ip})$, where $x_{i1}, \cdots, x_{ip}$ denote the values of p regression variables for subject i that can be selected by the experimenter. We will assume that these values can be selected independently. This implies in particular that there cannot be any functional relationships (such as $x_2 = x_1^2$ or $x_3 = x_1 x_2$) between these covariates. The design space $\mathcal{X}$ is thus a subset of $\mathbb{R}^p$.

An approximate design can be presented as $\xi = \{(\boldsymbol{X}_i, \omega_i), i = 1, ..., k\}$, where ω_i is the weight for the vector $\boldsymbol{X}_i$. For the parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)^T$, we will assume that $\theta_p \neq 0$. With $\boldsymbol{C}_i^T = (1, x_{i1}, \ldots, x_{i,p-1}, c_i)$ and $c_i = \boldsymbol{X}_i^T \boldsymbol{\theta}$, there is then a one-to-one relationship between $\boldsymbol{X}_i$ and $\boldsymbol{C}_i$. A design may thus also be written as $\xi = \{(\boldsymbol{C}_i, \omega_i), i = 1, ..., k\}$.

Using (4.5), and writing $\boldsymbol{A}^T = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{A}_1^T & 1/\theta_p \end{pmatrix}$ where $\boldsymbol{A}_1^T = \frac{1}{\theta_p}(-\theta_0, -\theta_1, \ldots, -\theta_{p-1})$, the information matrix for $\boldsymbol{\theta}$ under model (4.2) can be written as

$$\boldsymbol{I}_\xi = \sum_{i=1}^{k} \omega_i \boldsymbol{X}_i \Psi(c_i) \boldsymbol{X}_i^T = \boldsymbol{A}^T \left(\sum_{i=1}^{k} \omega_i \boldsymbol{C}_i \Psi(c_i) \boldsymbol{C}_i^T \right) \boldsymbol{A}, \tag{4.28}$$

where $\Psi(x) = [P'(x)]^2 / [P(x)(1 - P(x))]$. Note that for the logistic and probit models, $\Psi(x)$ was presented explicitly in (4.18). Also note that the case of $p = 1$ corresponds to the models in Section 4.4. The above matrix $\boldsymbol{A}^T$ reduces in that case to the corresponding matrix in (4.19).

As noted in Sitter and Torsney (1995a), unless appropriate constraints are placed on the design space, the information matrix can become arbitrarily large for the case that $p \geq 2$. In applications it may indeed be quite reasonable that each of the covariates can only take values in a bounded interval. In this chapter, we will assume that there are constraints on the first $p - 1$ covariates, but not on x_p. Specifically, we will assume that, for all i, $x_{ij} \in [U_j, V_j]$, $j = 1, \ldots, p - 1$, but x_{ip} can take any value in $(-\infty, \infty)$. The main reason for not placing a constraint on x_p is of a technical nature.

We will now show that there are, just as for the two-parameter models in Section 4.4, relatively simple complete classes for the multi-parameter models in this section. Consider a design $\xi = \{(\boldsymbol{C}_i, \omega_i), i = 1, \ldots, k\}$, with $\boldsymbol{C}_i^T = (1, x_{i1}, \ldots, x_{i,p-1}, c_i)$. We focus for the moment on x_{ij} for a fixed i and j with $j \leq p - 1$. Define $r_{ij} = \frac{V_j - x_{ij}}{V_j - U_j}$. Note that $r_{ij} \in [0, 1]$. Then, using the convexity of the function x^2, it is easy to see that

$$\begin{aligned} r_{ij} U_j + (1 - r_{ij}) V_j &= x_{ij}, \\ r_{ij} U_j^2 + (1 - r_{ij}) V_j^2 &\geq x_{ij}^2. \end{aligned} \tag{4.29}$$

It is now easy to see that if we replace $\boldsymbol{C}_i$ in ξ by $\boldsymbol{C}_{ij1} = (1, x_{i1}, \ldots, U_j, \ldots, x_{i,p-1}, c_i)^T$ and $\boldsymbol{C}_{ij2} = (1, x_{i1}, \ldots, V_j, \ldots, x_{i,p-1}, c_i)^T$ with weights $\omega_{ij1} = r_{ij}\omega_i$ and $\omega_{ij2} = (1 - r_{ij})\omega_i$, respectively, then the matrices $\omega_{ij1}\boldsymbol{C}_{ij1}\Psi(c_i)\boldsymbol{C}_{ij1}^T + \omega_{ij2}\boldsymbol{C}_{ij2}\Psi(c_i)\boldsymbol{C}_{ij2}^T$ and $\omega_i\boldsymbol{C}_i\Psi(c_i)\boldsymbol{C}_i^T$ have the same elements except for the $(j+1)$th diagonal element corresponding to covariate x_j. That diagonal element is larger in the former matrix than in the latter based on (4.29). Repeating this argument for other values of j, $1 \le j \le p-1$, we conclude that for any design point $\boldsymbol{C}_i$ there exist weights ω_i^ℓ, $\ell = 1, \ldots, 2^{p-1}$, so that

$$\omega_i\boldsymbol{C}_i\Psi(c_i)\boldsymbol{C}_i^T \le \sum_{\ell=1}^{2^{p-1}} \omega_i^\ell \boldsymbol{C}_i^\ell \Psi(c_i)(\boldsymbol{C}_i^\ell)^T. \tag{4.30}$$

The design points $\boldsymbol{C}_i^\ell$ are of the form $\left(\boldsymbol{C}_i^\ell\right)^T = (1, a_{\ell 1}, \ldots, a_{\ell,p-1}, c_i)$, where $a_{\ell j}$ is either U_j or V_j and $\sum_{\ell=1}^{2^{p-1}} \omega_i^\ell = \omega_i$. Now we are ready to present a complete class result.

Theorem 4.7. *For the logistic and probit model as in (4.2) with* $c_i = \boldsymbol{X}_i^T\boldsymbol{\theta}$*, let* $[U_j, V_j]$ *be the bounded interval for the* j*th covariate,* $1 \le j \le p-1$*. Then a complete class is formed by all designs with at most* 2^p *support points of the form* $\{(\boldsymbol{C}_{\ell 1}, \omega_{\ell 1})\ \&\ (\boldsymbol{C}_{\ell 2}, \omega_{\ell 2}), \ell = 1, \ldots, 2^{p-1}\}$*, where* $\boldsymbol{C}_{\ell 1}^T = (1, a_{\ell 1}, \ldots, a_{\ell,p-1}, c_\ell)$*,* $\boldsymbol{C}_{\ell 2}^T = (1, a_{\ell 1}, \ldots, a_{\ell,p-1}, -c_\ell)$*, and* $c_\ell > 0$*. Here* $a_{\ell j}$ *is either* U_j *or* V_j*, and* $(a_{\ell 1}, \ldots, a_{\ell,p-1})$*,* $l = 1, \ldots, 2^{p-1}$ *cover all such combinations.*

Proof. Let $\xi = \{(\boldsymbol{C}_i, \omega_i), i = 1, \ldots, k\}$ be an arbitrary design with $x_{ij} \in [U_j, V_j]$, $i = 1, \ldots, k$, $j = 1, \ldots, p-1$. By (4.28) and (4.30), we have

$$\boldsymbol{I}_\xi \le \boldsymbol{A}^T \left(\sum_{i=1}^{k} \sum_{\ell=1}^{2^{p-1}} \omega_i^\ell \boldsymbol{C}_i^\ell \Psi(c_i)(\boldsymbol{C}_i^\ell)^T \right) \boldsymbol{A}, \tag{4.31}$$

where $\boldsymbol{C}_i^\ell$ and ω_i^ℓ are as defined prior to the statement of the theorem. Notice that

$$\boldsymbol{C}_i^\ell \Psi(c_i)(\boldsymbol{C}_i^\ell)^T = \boldsymbol{B}_\ell^T \begin{pmatrix} \Psi(c_i) & c_i\Psi(c_i) \\ c_i\Psi(c_i) & c_i^2\Psi(c_i) \end{pmatrix} \boldsymbol{B}_\ell, \tag{4.32}$$

where $\boldsymbol{B}_\ell = \begin{pmatrix} 1 & a_{\ell 1} & \cdots & a_{\ell,p-1} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$. By (4.31) and (4.32), we have

$$\begin{aligned} \boldsymbol{I}_\xi &\le \boldsymbol{A}^T \left(\sum_{\ell=1}^{2^{p-1}} \boldsymbol{B}_\ell^T \left(\sum_{i=1}^{k} \omega_i^\ell \begin{pmatrix} \Psi(c_i) & c_i\Psi(c_i) \\ c_i\Psi(c_i) & c_i^2\Psi(c_i) \end{pmatrix} \right) \boldsymbol{B}_\ell \right) \boldsymbol{A} \\ &\le \boldsymbol{A}^T \left(\sum_{\ell=1}^{2^{p-1}} \sum_{i=1}^{2} \omega_{\ell i} \tilde{\boldsymbol{C}}_{\ell i} \Psi(\tilde{c}_l) \tilde{\boldsymbol{C}}_{\ell i}^T \right) \boldsymbol{A} \\ &= \boldsymbol{I}_{\tilde{\xi}} \end{aligned} \tag{4.33}$$

for a design $\tilde{\xi}$ that belongs to the complete class in the theorem with support points $\tilde{\boldsymbol{C}}_{\ell 1}$ and $\tilde{\boldsymbol{C}}_{\ell 2}$ and weights $\omega_{\ell 1}$ and $\omega_{\ell 2}$, $\ell = 1, \ldots .2^{p-1}$. The existence of these support points so that the second inequality in (4.33) holds is a consequence of Theorem 4.5. □

The complete class in Theorem 4.7 is simple, but finding an optimal design for a specific problem still requires the determination of 2^{p-1} values for the c_ℓ's and of the weights for the up to 2^p support points. This can easily be done by computer for small p, but is still challenging for larger p. For some problems it is again possible to find more explicit solutions for optimal designs. Below is a simple example without displaying all the computational details. We refer to Yang, Zhang, and Huang (2010) for more detailed examples of this kind.

Example 4.7. Consider a logistic model of the form (4.2) with $\boldsymbol{X}_i^T\boldsymbol{\theta} = c_i = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij}$. Suppose that we are interested in a locally D-optimal design for $(\beta_0, \beta_1, \beta_2, \beta_3) = (1, -0.5, 0.5, 1)$, and that the first two covariates are contained in the intervals $[-2, 2]$ and $[-1, 1]$, respectively. Suppose further that there is no restriction on the third covariate. Theorem 4.7 assures us that we can find an optimal design based on at most 8 design points for which the first two coordinates take all possible combinations of the limits of the respective intervals twice. Further computation gives a D-optimal design ξ^* for $(\beta_0, \beta_1, \beta_2, \beta_3)$ with equal weight 1/8 for each of the following eight points:

$$(-2, -1, -0.4564), \quad (-2, -1, -2.5436), \quad (-2, 1, -1.4564), \quad (-2, 1, -3.5436),$$
$$(2, -1, 1.5436), \quad (2, -1, -0.5436), \quad (2, 1, 0.5436), \quad (2, 1, -1.5436).$$

We can now compare this optimal design to any other design, for example to a full factorial design with 3 levels for each of the covariates, say -2, 0 and 2 for x_{i1}; -1, 0 and 1 for x_{i2}; and -3, -1 and 1 for x_{i3}, with equal weight for each of the 27 points. Simple computations show that the efficiency of ξ, defined as $\left(\frac{|I_\xi|}{|I_{\xi^*}|}\right)^{1/4}$, is only 70%.

We also point out that the support size for an optimal design need not be 2^p. This number is generally much larger than the number of parameters in $\boldsymbol{\theta}$, which is $p+1$, so that it may be possible to find optimal designs with a (much) smaller support size. The result in Theorem 4.7 does not exclude this possibility. Indeed, the statement of the theorem refers to "at most 2^p support points". This means that some of the weights can perhaps be taken as 0. It is our experience that this is often possible, but we do not have a general recipe for obtaining such smaller designs.

4.5.2 GLMs with Group Effects

In this section we return to the problem of a single covariate, but now in the presence of group effects. If the subjects in the study have different characteristics with respect to one or more classificatory variables, such as race, gender, age group, and so on, and if the relationship between the response variable and the covariate can be different for subjects with different characteristics, then this should be reflected in the model. This is the main problem studied in Stufken and Yang (2010). In this section we will present their main results, but refer for the proofs to the original paper. We will also restrict attention to a binary response variable, and refer to Stufken and Yang (2010) for Poisson regression models.

To present the model, it is convenient to change the notation slightly from that in previous sections. We will use double subscripts to denote the subjects. For example, Y_{ij} is now used to denote the response from the jth subject in the ith group. We assume that there is a single covariate for this subject, x_{ij}, the value of which can be selected by the experimenter. In addition, the relationship between Y_{ij} and x_{ij} may depend on the group, i.e., it may depend on i. As before, we want to model $Prob(Y_{ij} = 1)$.

We consider two different models, one with a common slope and one with a group-dependent slope. The first of these models can be written as

$$Prob(Y_{ij} = 1) = P(\alpha_i + \beta x_{ij}), \tag{4.34}$$

where β is a common slope parameter, the α_i's are group effects, and P is a cdf as in (4.2). We could have used the notation of Subsection 4.2.1 for this model with $\boldsymbol{X}_{ij}^T = (0, \ldots, 1, \ldots, 0, x_{ij})$ and $\boldsymbol{\theta}^T = (\alpha_1, \ldots, \alpha_k, \beta)$. Here k denotes the number of groups, and $\boldsymbol{X}_{ij}$ has a 1 in the ith position, with all other entries among the first k being 0. For the second model we can write

$$Prob(Y_{ij} = 1) = P(\alpha_i + \beta_i x_{ij}). \tag{4.35}$$

Now $\boldsymbol{X}_{ij}^T$ is of length $2k$ with a 1 in the ith position and with x_{ij} in position $k + i$, i.e., $\boldsymbol{X}_{ij}^T = (0, \ldots, 1, \ldots, 0, 0, \ldots, x_{ij}, \ldots, 0)$, and $\boldsymbol{\theta}^T = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k)$.

While these models include group effects, they do not attempt to model these effects. If appropriate, simpler models could be used. For example, for the models in (4.34) and (4.35), the α_i's could be modeled as a sum of an overall mean and main-effects for the different variables that induce the groups. This could also be done for the β_i's in (4.35). Such assumptions would reduce the number of parameters. For example, with two factors that have k_1 and k_2 levels, respectively, we have $k = k_1 \times k_2$. But if we can assume for modeling the α_i's, say, that the two factors do not interact, then we can use main-effects only and reduce the number of parameters from k to $k_1 + k_2 - 1$. We will not do this here, but merely note that the complete class results for the models in (4.34) and (4.35) that we will formulate in Theorem 4.8 also hold for these reduced models.

We will formulate the complete class results in terms of the induced design space, using $c_{ij} = \alpha_i + \beta x_{ij}$ for the model in (4.34) and $c_{ij} = \alpha_i + \beta_i x_{ij}$ for the model in (4.35). The complete class result can be formulated succinctly, with some loss of precision, by stating that for each group we should use a design of the form that we would have used if that group had been the only group (in which case we would have invoked Theorem 4.5). Here is the more precise statement.

Theorem 4.8. *For the logistic or probit models of the form (4.34) or (4.35), suppose that the design space is of the form $c_{ij} \in [D_{i1}, D_{i2}]$. A complete class is formed by all designs with at most two support points per group, with the additional restriction that if there are two support points in group i, then the five conditions in Theorem 4.5 apply with D_1 and D_2 replaced by D_{i1} and D_{i2}.*

The proof of this result is fairly simple and can be found in Stufken and Yang (2010). The result vastly reduces the search for optimal designs. Nonetheless, there are still up to $2k$ support points and corresponding weights, so that an optimization algorithm may still have difficulties solving this problem if k is large. Depending on the model (and the use of possible further simplifying assumptions based on main-effects and lower order interactions), it may be possible to find optimal designs with a smaller support size than $2k$. We would still search in the complete class specified in Theorem 4.8, but some of the weights can be taken as zero.

Example 4.8. Consider a logistic model of the form as in (4.34), with a common slope, for two factors that have three levels each. Assume that the two-factor interaction is negligible. Let $\alpha_1, \alpha_2, \ldots, \alpha_9$ correspond to the groups (1,1), (1,2), ..., (3,3), respectively. For finding a locally A-optimal design, suppose that $\boldsymbol{\theta}^T = (-0.95, -1, -0.9, -0.85, -0.9, -0.8, -1.05, -1.1, -1, 1)$, where the first nine entries correspond to $\alpha_1, \alpha_2, \ldots, \alpha_9$, and the last entry is for β. Note that these values are consistent with the assumption of no interaction between the two factors. We assume that there is no restriction on the design space. Suppose that the vector of interest, $\boldsymbol{\eta}$, consists of (1) the contrast of the average of the mean responses at levels 1 and 2 for the first factor versus the average at its 3rd level (i.e., $\frac{1}{6}(\sum_{i=1}^{6} \alpha_i - 2\sum_{i=7}^{9} \alpha_i)$); (2) the contrast of the average of the mean responses at levels 1 and 2 for the second factor versus the average at its 3rd level (i.e., $\frac{1}{6}(\sum_{i=1,2,4,5,7,8} \alpha_i - 2\sum_{i=3,6,9} \alpha_i)$); and (3) the slope parameter (i.e., β). An A-optimal design for $\boldsymbol{\eta}$ that uses fewer than 18 support points is shown in Table 4.1.

Table 4.1: Support Points and Weights for a Locally A-Optimal Design

Group	A-optimal Design
(1,1)	(1.7691, 0.0550); (0.1309, 0.0700)
(1,2)	No support points (weights are 0)
(1,3)	(1.7191, 0.0783); (0.0809, 0.0466)
(2,1)	No support points (weights are 0)
(2,2)	(1.7191, 0.0482); (0.0809, 0.0769)
(2,3)	(1.6191, 0.0852); (-0.0191, 0.0398)
(3,1)	(1.8691, 0.0658); (0.2309, 0.0591)
(3,2)	(1.9191, 0.0727); (0.2809, 0.0523)
(3,3)	(1.8191, 0.0949); (0.1809, 0.1552)

It is also possible to use Theorem 4.8 for some problems to find explicit optimal designs. Stufken and Yang (2010) prove such a result. To present it, let $\boldsymbol{\eta}_1 = (\alpha_1/\beta, \ldots, \alpha_k/\beta, \beta)^T$ and $\boldsymbol{\eta}_2 = ((\alpha_1 - \alpha_k)/\beta, \ldots, (\alpha_{k-1} - \alpha_k)/\beta, \beta)^T$.

Theorem 4.9. *For the model in (4.34), suppose that we have a single factor with k levels and no constraint on the design space. Let design* $\xi^* = \{(c_{i1} = c^*, \omega_{i1} = \frac{1}{2k}), (c_{i2} = -c^*, \omega_{i2} = \frac{1}{2k}), i = 1, \ldots, k\}$ *for some* c^*. *Then the following results hold:*

(i) ξ^* *is D-optimal for* $\boldsymbol{\eta}_1$ *if* c^* *maximizes* $c^2\Psi^{k+1}(c)$*; and*

(ii) ξ^* *is D-optimal for* $\boldsymbol{\eta}_2$ *if* c^* *maximizes* $c^2\Psi^k(c)$.

Computing the values of c^* for the logistic and probit models is easy to do with software like MATLAB and the expressions for $\Psi(x)$ in (4.18). The results are shown in Table 4.2.

Table 4.2: The value of c^* that maximizes $c^2\Psi^q(c)$

	c^*			c^*	
q	Logistic	Probit	q	Logistic	Probit
1	2.3994	1.5750	6	0.8399	0.6696
2	1.5434	1.1381	7	0.7744	0.6209
3	1.2229	0.9376	8	0.7222	0.5815
4	1.0436	0.8159	9	0.6793	0.5487
5	0.9254	0.7320	10	0.6432	0.5209

4.6 Summary and Concluding Comments

Selecting a good design is a complex problem that, typically, involves many different considerations. In the context of GLMs, this chapter focuses on one of these, namely design optimality with respect to some criterion based on information matrices. Rather than focusing on a single criterion, we focus on identifying complete classes of designs. For any optimality criterion that obeys the Loewner ordering of information matrices, we can always find a locally optimal design in these complete classes. Therefore, we can restrict searches for optimal designs to these classes. This is tremendously helpful if the complete classes are sufficiently small.

Identifying sufficiently small complete classes for GLMs is not a simple problem, but this chapter shows that the problem can be handled for a wide variety of models. By doing so, most results on optimal designs for GLMs obtained by other methods can be covered and extended. We have also shown how the determination of complete classes can at times be used to obtain explicit forms of optimal designs for specific criteria and objectives.

We reiterate that, in practice, one may wind up not using a locally optimal design. Practical considerations, considerations of robustness of a design to uncertainty in the "local values" of the parameters, robustness to model uncertainty, and other considerations may play a role in selecting the design that is eventually used. Nonetheless, whatever the considerations are, in the end one would hope to have a design that is efficient under reasonable criteria and assumptions. In order to assess the efficiency of a proposed design under optimality criteria, one has to compare it to an optimal design. That can only be done if one has the tools to identify an optimal design. Being able to identify optimal designs is thus important, irrespective of whether one plans to use an optimal design in an experiment or not.

While the analytic approach presented here, and in greater technical detail in some of the references given throughout this chapter, is very successful, there remain many open problems in the general area of identifying optimal designs for GLMs and other nonlinear models. The quote from Khuri, Mukherjee, Sinha, and Ghosh (2006) presented in Section 4.1 remains valid. There are still many dissertations to be written in this area.

4.7 References

Abdelbasit, K. M. and R. L. Plackett (1983). Experimental design for binary data. *J. Am. Statist. Assoc.*, **78**, 90–98.

Agin, M. and K. Chaloner (1999). Optimal Bayesian design for a logistic regression model: geometric and algebraic approaches. In: *Multivariate Analysis, Design of Experiments and Survey Sampling*, S. Ghosh (ed.). New York: Dekker, pp. 609–624.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edition. New York: Wiley.

Biedermann, S., H. Dette and W. Zhu (2006). Geometric construction of optimal designs for dose-response models with two parameters. *J. Am. Statist. Assoc.*, **101**, 747–759.

Chaloner, K. and K. Larntz (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inf.*, **21**, 191–208.

Dette, H. and L. M. Haines (1994). *E*-optimal designs for linear and nonlinear models with two parameters. *Biometrika*, **81**, 739–754.

Dror, H. A. and D. M. Steinberg (2006). Robust experimental design for multivariate generalized linear models. *Technometrics*, **48**, 520–529.

Dror, H. A. and D. M. Steinberg (2008). Sequential experimental designs for generalized linear models. *J. Am. Statist. Assoc.*, **103**, 288–298.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.*, **23**, 255–262.

Ford, I., B. Torsney and C. F. J. Wu (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc., B*, **54**, 569–583.

Haines L. M., M. G. Kabera, P. Ndlovu and T. E. O'Brien (2007). D-optimal designs for logistic regression in two variables. In: *MODA 8 - Advances in Model-Oriented Design and Analysis*, J. L. López-Fidalgo, J. M. Rodríguez-Díaz and B. Torsney (eds.). Heidelberg: Physica Verlag, pp. 91–98.

Khuri, A. I., B. Mukherjee, B. K. Sinha and M. Ghosh (2006). Design issues for generalized linear models: A review. *Statist. Sci.*, **21**, 376–399.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canad. J. Math.*, **12**, 363–366.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, **2**, 849-879.

Mathew, T. and B. K. Sinha (2001). Optimal designs for binary data under logistic regression. *J. Statist. Plann. Inf.*, **93**, 295–307.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* 2nd Edition. London: Chapman & Hall.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, Linear, and Mixed models.* New York: Wiley.

Minkin, S. (1987). Optimal designs for binary data. *J. Am. Statist. Assoc.*, **82**, 1098–1103.

Minkin, S. (1993). Experimental design for clonogenic assays in chemotherapy. *J. Am. Statist. Assoc.*, **88**, 410–420.

Pukelsheim, F. (2006). *Optimal Design of Experiments*, Philadelphia: SIAM.

Russell, K. G., D. C. Woods, S. M. Lewis and J. A. Eccleston (2009). D-optimal designs for Poisson regression models. *Statist. Sinica*, **19**, 721–730.

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

Sitter, R. R. and B. E. Forbes (1997). Optimal two-stage designs for binary response experiments. *Statist. Sinica*, **7**, 941–955.

Sitter, R. R. and B. Torsney (1995a). Optimal designs for binary response experiments with two design variables. *Statist. Sinica*, **5**, 405–419.

Sitter, R. R. and B. Torsney (1995b). Optimal designs for generalised linear models. In: *MODA 4 - Advances in Model-Oriented Design and Analysis*, C. P. Kitsos and W. G. Muller (eds.). Heidelberg: Physica-Verlag, pp. 87–102.

Sitter, R. R. and C. F. J. Wu (1993a). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scand. J. Statist.*, **20**, 329–341.

Sitter, R. R. and C. F. J. Wu (1993b). On the accuracy of Fieller intervals for binary response data. *J. Am. Statist. Assoc.*, **88**, 1021–1025.

Stufken, J. and M. Yang (2010). On locally optimal designs for generalized linear models with group effects. Submitted.

Tighiouart, M. and A. Rogatko (2006). Dose escalation with overdose control. In: *Statistical Methods for Dose-Finding Experiments*, S. Chevret (Ed.). Chichester: Wiley, pp. 173–188.

Woods, D. C., S. M. Lewis, J. A. Eccleston and K. G. Russell (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, **48**, 284–292.

Yang, M. (2008). A-optimal designs for generalized linear model with two parameters. *J. Statist. Plann. Inf.*, **138**, 624–641.

Yang, M. (2010). On the de la Garza phenomenon. *Ann. Statist.*, **38**, 2499–2524.

Yang, M. and J. Stufken (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.*, **37**, 518–541.

Yang, M., B. Zhang and S. Huang (2009). Optimal designs for binary response experiments with multiple variables. *Statist. Sinica*, in press.