UNIVERSIDAD NACIONAL DE COLOMBIA

Penalized Bayesian Optimal Designs for Nonlinear Models of Continuous Response

Svetlana Ivanovna Rudnykh

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2018

Penalized Bayesian Optimal Designs for Nonlinear Models of Continuous Response

Svetlana Ivanovna Rudnykh

Tesis presentada como requisito parcial para optar al título de:
Doctora en Ciencias Estadística

Director:

Ph.D. Víctor Ignacio López Ríos

Línea de Investigación:

Diseño de Experimentos

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2018

Dedication

To my parents,

my daughter, Efigenia, and

my dear husband, Ramon.

Acknowledgments

I wish to thank several people for their support and contributions to this work.

First, I would like to express my sincere gratitude to Dr. Víctor I. López, my dissertation advisor, for his guidance, encouragement, and useful critiques of my research. I would also like to thank the rest of my Doctorate professors, Dr. Juan Carlos Salazar and Dr. Nelfi González for their teaching, wisdom, and patience.

I would additionally like to thank the administrative staff of the Department of Statistics at the Universidad Nacional de Colombia, especially Dr. Carlos Mario Lopera, Dr. Isabel Cristina Ramírez, and Olga Bústos, whose supports were invaluable throughout my time at the UNAL.

I also wish to thank the Universidad del Atlántico for its financial support in my academic training process.

I would also like to express my special gratitude to my classmate Karime García for her valuable help and friendship.

Finally, I would like to thank my husband, Rámon, and my daughter for their love, support, and patience during my years of study. The love and support of my family made it possible for me to make this dream come true.

Abstract

Experimental design is an important phase in both scientific and industrial research. In recent years, Bayesian optimal designs have become more and more popular, particularly in biomedical research and clinical trials. The Bayesian experimental design approach allows the prior information of unknown parameters to be incorporated into the design process in order to achieve a better design. The Bayesian optimal design theory can, however, produce inadequate designs from a practical perspective that conflict with common practice in laboratories or other guidelines established.

In this research, the penalized optimal design strategy with the Bayesian approach is suggested to reduce problems associated with the inadequacy of experimental designs from a practical perspective. New optimality criteria, which combine the use of desirability functions and the Bayesian approach, are constructed for linear and nonlinear regression models. The proposed technique based on the use of desirability functions helps to obtain optimal designs that fulfill Bayesian optimal design criteria and also satisfy practical preferences.

The proposed penalized strategy is illustrated with corresponding examples for both linear and nonlinear models. Furthermore, the methodology of choosing the appropriate desirability functions according to the practical design preferences is proposed and illustrated by an example of the Michaelis-Menten model.

Keywords: Bayesian optimal designs, Desirability functions, Penalized designs.

Resumen

El diseño experimental es una fase importante tanto en la investigación científica como en la industria. En los últimos años, los diseños óptimos bayesianos se han vuelto cada vez más populares, particularmente en la investigación biomédica y los ensayos clínicos. El enfoque de diseño experimental bayesiano permite incorporar la información previa disponible de parámetros desconocidos en el proceso de diseño y así poder obtener un mejor diseño. Sin embargo, la teoría del diseño óptimo bayesiano puede producir diseños inadecuados desde una perspectiva práctica que entran en conflicto con la práctica de laboratorio común u otras pautas establecidas.

Con el objetivo de reducir los problemas asociados con la inadecuación de los diseños experimentales desde una perspectiva práctica, en esta investigación, se proponen nuevos criterios de optimalidad que combinan el uso de funciones de deseabilidad y el enfoque bayesiano, tanto para modelos de regresión lineal, como no lineal. La técnica propuesta basada en el uso de las funciones de deseabilidad ayuda a obtener diseños óptimos penalizados que cumplen con los criterios de diseño óptimos bayesianos y también satisfacen preferencias prácticas.

La estrategia penalizada propuesta se ilustra con los respectivos ejemplos para modelos lineales y no lineales. Además, se propone y se ilustra una metodología para elegir las funciones de deseabilidad apropiadas de acuerdo con las preferencias experimentales desde un punto de vista práctico mediante un ejemplo del modelo de Michaelis-Menten.

Palabras clave: Diseños óptimos bayesianos, Funciones de deseabilidad, Diseños penalizados.

Contents

List of Figures

List of Tables

List of abbreviations and symbols

This section presents some abbreviations and symbols of frequent use along this document.

Abbreviations

CDF	Cumulative distribution function
DE	Differential Evolution
FDA	Federal Drug Administration
GA	Genetic Algorithm
GLP	Good Laboratory Practice
GV	Generalized Variance
i.i.d	independent and identically distributed
LSM	Least Squares Method
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MSE	Mean Square Errors
pdf	probability density function

Symbols in Latin letters

$\lceil \cdot \rceil$	Ceiling function that maps a real number to the smallest following integer
$\mathrm{E}[\cdot]$	Expectation with respect to the prior distribution π
d_i	ith desirability function
$D(\boldsymbol{x}, \boldsymbol{w})$	Overall desirability function
$d(\boldsymbol{x}, \xi)$	Standardized variance for continuous designs

det	Determinant of a matrix
$\boldsymbol{f}(\boldsymbol{x})$	Vector of known "basis" regression functions for linear models
$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$	Vector of partial derivatives of nonlinear model
$\boldsymbol{F}$	First partial derivative matrix of the nonlinear model with respect to $\boldsymbol{\theta}$
G	Number of simulations to obtain a random sample of parameters
i	Counter for the number of observations in the sample size N
$\boldsymbol{I}_N$	$N \times N$ identity matrix
k	Number of of features (factors) to penalize or number of desirability functions
l	Multiple or submultiple of corresponding minimum optimal criterion
m	Dimension of the explanatory variable
$\boldsymbol{M}$	Information matrix (Fisher)
$\boldsymbol{M}_B$	Bayesian information matrix
$\mathcal{M}$	Set of information matrices $\boldsymbol{M}$ corresponding to all possible continuous designs ξ
n	Number of design support points
N	Sample size of an experiment
p	Dimension of parameter vector $\boldsymbol{\theta}$
$\boldsymbol{R}$	Given positive definite $p \times p$ "precision" matrix
r_i	Number of repetitive experiments in corresponding point x_i
S	Counter for the number of simulations.
tr	Trace of a matrix
$U(d, \boldsymbol{\theta}, \xi, \boldsymbol{Y})$	Utility function
w_i	Weights of support points of design
x_i	Design points or support points of design if they are different
$\boldsymbol{X}$	$N \times p$ extended design matrix for linear models
$\mathscr{X}$	Design region
$\boldsymbol{Y}$	Vector of observations
Y_{iS}	Simulated observation at the design point x_i in the simulation S

Symbols in Greek letters

$\boldsymbol{\varepsilon}$	Vector of errors
$\eta(\boldsymbol{x}, \boldsymbol{\theta})$	Model function
$\boldsymbol{\theta}$	Vector of unknown parameters
$\boldsymbol{\theta}_t$	True vector of unknown parameters
Θ	Parameter space
ξ_N	Exact design of an experiment of size N
ξ	Continuous approximate design
Ξ_N	Set of all possible exact designs ξ_N
Ξ	Set of all possible continuous designs ξ
Λ	User specified constant necessary to control the weight of the penalty, relative to the minimum optimality criterion
$\pi(\cdot)$	Probability density function
ρ	Correlation coefficient
Σ	Covariance matrix
$\phi(\boldsymbol{x}, \xi)$	Derivative of Ψ in the direction $\bar{\xi}$
Ψ	Function of the information matrix that called Ψ-optimality criterion

1. Introduction

Experimental design plays a vital role in theoretical and applied scientific research. A well-designed experiment is an efficient method for learning about some phenomena, while a poorly designed experiment directly affects the quality of the conclusions derived from the experimental data. Due to cost or time, scientists want to design experiments that maximize the amount of information that can be retrieved from a finite amount of available resources. Statisticians have developed optimal design theory to generate efficiently designed experiments that satisfy the mentioned requirements.

Optimal designs provide a very efficient way to maximize the amount of information gained in an experiment. They are commonly derived using optimality criteria that are based on the Fisher information matrix (Atkinson, Donev & Tobias 2007, ch. 10). The D-, A-, G-optimal designs are some examples of such designs.

For nonlinear models, optimal designs generally depend on the true values of the model parameters. Since the parameter vector is not known, the researcher must postulate a best "guess" of the unknown parameter vector resulting in locally optimal designs (Chernoff 1953). The problem may arise when that guess is not close enough to the true parameter vector, and therefore, the design obtained may not be optimal. Therefore, if the model parameters are misspecified, the robustness of the non-Bayesian optimal design may be impaired.

Usually, prior information is available to the experimentation, for example in previous studies, specialized scientific literature or expert opinions. The Bayesian approach to experimental design allows using this available information into the design process. The Bayesian optimal design employes a prior distribution of the unknown parameters rather than guessed single values. Therefore, the incorporation of prior information on the unknown parameters would lead to more accurate and robust designs and increase their efficiency.

Furthermore, the optimal design theory can generate inadequate designs from a practical perspective. These designs can conflict with common practice in laboratories or other guidelines established. Many authors proposed several alternatives to generate optimal designs with the desired experimental properties, e.g. a combination of several crite-

ria and inclusion of restrictions or penalties in optimal design, both non-Bayesian and Bayesian.

The idea of incorporating constraints, such as costs or penalties, in optimal design returns to the paper by Elfving et al. (1952) on the geometrical construction of c-optimal designs. These authors introduced a constraint on the total costs instead of the constraint on the number of observations. Cook & Fedorov (1995) considered several types of constraints. They proposed that optimization problems with constraints can be embedded in the convex theory of experimental design. The constrained optimal design theory can be found in the book of Fedorov & Hackl (1997).

Dragalin & Fedorov (2006) and Dragalin, Fedorov & Wu (2008) applied this constrained optimal design theory to suggest compromise between individual and collective ethics in dose-finding studies. These authors showed that the resulting penalized optimal designs have achieved a reasonable balance between individual and collective ethics in dose-finding studies. Pronzato (2008, 2010) continued the research of previous authors and introduced flexibility in setting the compromise between the information gained and the cost of the experiment. He showed that, for suitable penalty functions, all doses in the experiment had a small cost; this allowed the avoidance of extreme doses generally suggested by optimal design for parameter estimation.

In recent years, Bayesian optimal designs have become more and more popular, particularly in biomedical research and clinical trials. Chaloner & Verdinelli (1995) presented a general overview of Bayesian experimental design, for both linear and nonlinear models. Chaloner & Larntz (1989) gave a unifying idea of Bayesian optimal design for nonlinear models.

Bayesian optimal design also has used the idea of incorporating constraints. Haines, Perevozskaya & Rosenberger (2003) formulated and used the criteria with a constraint to deal with the ethical dilemma of avoiding highly toxic doses. These authors proposed a constrained Bayesian design which maximizes Bayesian design criterion subject to a restricted dose space. This space was defined as the weighted sum of distribution function of the maximal allowed dose. They formulated the equivalence theorem for the constrained Bayesian optimal designs and developed the sequential optimal design scheme. The simulation results under this scheme indicated that this procedure performed effectively and efficiently.

Gao & Rosenberger (2013) proposed using the Bayesian approach to adaptive designs of Dragalin & Fedorov (2006) to penalize doses with too much toxicity or too little efficacy. These authors introduced a penalty coefficient to trade-off information goal and ethical

goal. Using simulation, they showed that a compromise between the two goals can be achieved by tuning this coefficient.

Parker & Gennings (2008) proposed a technique based on the desirability functions, as an alternative penalty approach. The use of desirability functions allows the user to obtain the experimental design with important particular features. These authors illustrated their proposed methodology for two nonlinear models: Gompertz nonlinear model and nonlinear threshold model. In the first example, penalized D_s-optimal design criteria included desirability functions that affect the assignment of replicates to the control group and the location of the maximum dose point. In the second example, penalized D-optimal design criteria included two desirability functions to define the minimum number of subjects in each dose group and the minimum difference between adjacent dose groups. Finally, these authors concluded that resulting penalized optimal designs had desirable practical characteristics defined by a researcher.

Furthermore, these designs had a significant enhancement in parameter variance over an arbitrary design with practical characteristics. Thus, the use of desirability functions allows the researcher to obtain the optimal design with desirable particular properties. The advantage of using them is that these functions can be represented as continuous functions and not as fixed restrictions.

Yeatts, Gennings & Crofton (2012) used the methodology described in Parker & Gennings (2008), to estimate the interaction threshold. The authors applied this penalty approach to a mixture of 18 Polyhalogenated Aromatic Hydrocarbons to obtain the D_s- and penalized D_s-optimal second-stage designs. The experimental data available from the first stage were incorporated in the second stage. The penalized D_s-optimal second-stage design criterion was applied to minimize the variance of the hypothesized interaction threshold given the design preferences of the researcher. These authors observed that the resulting penalized optimal design exhibited desirable practical characteristics, while it determined only a minimal increase in the variance of the interaction threshold. The authors concluded that this method is a useful tool to determine an experimental design that can be used to estimate the location of the interaction threshold precisely.

Inadequate experimental designs can cause problems in the estimation of the model parameters and/or generate conflicts with common laboratory practices or other established guidelines, as we mentioned before. The use of desirability functions to penalize the optimal design criterion can help overcome these difficulties.

There is not a methodology to define an appropriate desirability function per the practical design preferences, neither to construct augmented constrained optimal designs by

penalty functions, as far as we have reviewed, neither we have found works related with desirability functions applied to Bayesian constrained optimal designs. In this research, we attempt to fill this gap. More specifically, we propose a new optimality criterion to obtain penalized Bayesian D-optimal designs, combining the use of desirability functions and the Bayesian approach in the design construction for both linear and nonlinear regression models.

The purpose of this research is to establish a procedure that allows the construction of penalized Bayesian optimal experimental designs. We propose the methodology to determine appropriate desirability functions according to the practical design preferences and define the steps to follow to obtain the penalized optimal design, based on the predefined researcher experimental design preferences.

This dissertation is organized as follows. In Chapter 2 , literature on optimal design, augmented design, Bayesian optimal design, and computational methods in finding optimal designs are reviewed. This theoretical background is fundamental for the development of subsequent chapters.

In Chapter 3 , a summary of the desirability function theory and an approach of penalized optimal design strategy based on the use of desirability functions are presented. This strategy produces an experimental design having good statistical inference properties as well as desirable practical characteristics. The practical characteristics are determined by a penalty function through an overall desirability function. This penalty function is added to an "alphabetic" optimality criterion in order to penalize optimal designs that are unattractive from a practical point of view. The methodology of choosing the appropriate desirability functions according to the practical design preferences is proposed and illustrated with an example of the Michaelis-Menten model. In this example, the steps to follow to obtain the penalized optimal design are detailed.

The principal results of this dissertation are presented in Chapters 4 and 5 . An extension of the use of desirability functions in Bayesian optimal designs is proposed in this dissertation. Its purpose is to reduce problems associated with the inadequacy of experimental designs from a practical perspective. In Chapters 4 and 5 the new penalty criteria, which combine the use of desirability functions and the Bayesian approach, are constructed for linear and nonlinear regression models, respectively. Thus, the researcher can incorporate prior information of the unknown parameters by using a Bayesian approach and also satisfy practical preferences by applying the penalty through desirability functions. In each case, the justification of the proposed criterion is presented.

The penalized Bayesian D-optimal designs for both linear and nonlinear regression mod-

els, by incorporating our proposed penalty criterion in the design construction are obtained and discussed in Chapters 4 and 5 , respectively. The proposed methodology is illustrated with two examples of regression models. Chapter 4 presents the penalized Bayesian D-optimal design procedures and results for the quadratic regression model; and Chapter 5 presents the penalized Bayesian D-optimal design procedures and results for the exponential regression model. The recommended steps to follow to obtain the penalized Bayesian optimal design for nonlinear models are detailed in Chapter 5. Corresponding programs are developed in the statistical software R (Team 2018) to calculate the penalized designs.

Conclusions, recommendations, and possible future work are discussed in Chapter 6 . Computer codes implemented in R-project (Team 2018) to generate penalized and penalized Bayesian D-optimal designs may be found in Appendix B.

2. Background of Optimal Designs

This chapter presents the most important concepts and results of optimal design theory for regression models used in this dissertation. First, it covers classical linear and nonlinear optimal design theory described in Section 2.1 . Most of the content of this section has been taken from the books of Atkinson et al. (2007), Ermakov & Zhiglijavsky (1987), Fedorov & Hackl (1997), and (Fedorov & Leonov 2013).

The ways to solve the shortcomings of the classical theory of optimal design, such as augmented designs and Bayesian optimal designs, are presented in Section 2.2 and Section 2.3, respectively. The addition of extra points to an existing optimal design has the objective of improving the properties of the design. Meanwhile, the Bayesian optimal design uses available prior information of unknown parameters into the design process in order to obtain accurate and robust designs and increase their efficiency. Furthermore, the computational methods for finding optimal designs are reviewed in Section 2.4.

2.1. Optimal Experimental Design

Experimental design plays a vital role in theoretical and applied scientific research. Usually, its costs are high, and its resources are limited. A well-designed experiment is an efficient method for maximizing the amount of information that can be retrieved from a limited amount of available resources in the experiment. Statisticians have developed optimal design theory to generate efficiently designed experiments that satisfy the mentioned requirements.

2.1.1. Model

Suppose that observations y_i of an experiment satisfy

$$y_i = \eta(\boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \qquad i = 1, 2, \ldots, N, \tag{2-1}$$

where y_i are continuous response (or dependent or observed) variables, $\boldsymbol{x}_i$ is a $m \times 1$ vector of explanatory variables, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, where $\boldsymbol{\theta} \in \Theta$, here Θ is an open convex set in $\mathbb{R}^p$, ε_i are uncorrelated normally distributed random variables with zero mean and constant variance σ^2. Further, suppose that the vector $\boldsymbol{x}_i \in \mathscr{X}$ called space or design region, where $\mathscr{X}$ is a compact set in $\mathbb{R}^m$ and $\eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ is

a known continuous in $\mathscr{X}$ function that can be linear or nonlinear with respect to the model parameters $\boldsymbol{\theta}$.

For linear regression models the function $\eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ is given by

$$\eta(\boldsymbol{x}_i, \boldsymbol{\theta}) = \boldsymbol{f}(\boldsymbol{x}_i)^T \boldsymbol{\theta}, \tag{2-2}$$

where $\boldsymbol{f}(\boldsymbol{x}_i) = (f_1(\boldsymbol{x}_i), f_2(\boldsymbol{x}_i), \ldots, f_p(\boldsymbol{x}_i))^T$ is a vector of known "basis" regression functions. They are p linearly independent real-valued continuous functions on the design space $\mathscr{X}$. The linear model may be written in the matrix form (Ermakov & Zhiglijavsky 1987, p. 16)

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{2-3}$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_N)^T$ is the vector of observations, $\boldsymbol{X} = (\boldsymbol{f}(\boldsymbol{x}_1), \boldsymbol{f}(\boldsymbol{x}_2), \ldots, \boldsymbol{f}(\boldsymbol{x}_N))^T$ is the $N \times p$ extended design matrix and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T$ is the $N \times 1$ vector of the errors.

Under the above assumptions $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\theta}$ and $\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix (Fedorov & Hackl 1997, p. 7).

2.1.2. Estimation

Least squares are used as an estimation method to estimate parameters in linear and nonlinear models when the assumption of constant variance is valid. The least-squares estimate of $\boldsymbol{\theta}$ is the $p \times 1$ vector $\hat{\boldsymbol{\theta}}$ that minimizes the error sum of squares criterion with respect to $\boldsymbol{\theta}$, given in (Atkinson et al. 2007, p. 45):

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Omega} \sum_{i=1}^{N} (y_i - \eta(\boldsymbol{x}_i, \boldsymbol{\theta}))^2. \tag{2-4}$$

For linear model the least-squares estimator of the parameters is (Atkinson et al. 2007, p. 52)

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}, \tag{2-5}$$

where $\boldsymbol{X}^T \boldsymbol{X}$ is the $p \times p$ matrix. The covariance matrix of the least-squares estimator is

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}. \tag{2-6}$$

The value of σ^2 is not relevant in optimal design theory, and its value is chosen the same for all proposed designs for a given experiment.

For nonlinear model it is not possible to find a closed-form expression for $\hat{\boldsymbol{\theta}}$ like (2-5) in the linear case. The least-squares estimators are strongly consistent with a probability

of 1 when a large number of observations N (Seber & Wild 2003, p. 565). Therefore $\boldsymbol{\theta}$ is located in a sufficiently small neighborhood of $\boldsymbol{\theta}_t$, where $\boldsymbol{\theta}_t$ is the true vector of the unknown parameters. Then the model function in that neighborhood can be expressed as follows (Fedorov & Hackl 1997, p. 17):

$$\eta(\boldsymbol{x}, \boldsymbol{\theta}) \cong \eta(\boldsymbol{x}, \boldsymbol{\theta}_t) + (\boldsymbol{\theta} - \boldsymbol{\theta}_t)^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_t), \tag{2-7}$$

where $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \dfrac{\partial \eta(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, i.e., the model can be linearized. The least-squares estimator obtained in the linearized regression model is (Ermakov & Zhiglijavsky 1987, p. 200):

$$\hat{\boldsymbol{\theta}} \cong \left(\boldsymbol{F}^T \boldsymbol{F}\right)^{-1} \boldsymbol{F}^T \boldsymbol{Y}, \tag{2-8}$$

where $\boldsymbol{F}$ is the first partial derivative matrix of the model with respect to $\boldsymbol{\theta}$, defined as

$$\boldsymbol{F} = \left[\left(\frac{\partial \eta(\boldsymbol{x}_i, \boldsymbol{\theta}_t)}{\partial \theta_j}\right)\right] \tag{2-9}$$

for $i = 1, 2, \ldots, N$ observations and $j = 1, 2, \ldots, p$ model parameters. The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by $\sigma^2 \left(\boldsymbol{F}^T \boldsymbol{F}\right)^{-1}$. The determinant of this variance-covariance matrix is called the generalized variance (GV) of $\hat{\boldsymbol{\theta}}$ (Sengupta 2004). This concept was introduced by Wilks (1967) as a scalar measure of global multidimensional dispersion.

2.1.3. Experimental Designs

The precision of parameter estimates depends on the choice of the experimental points $\boldsymbol{x}_i \in \mathscr{X}$, as it was shown in the previous paragraph. This dependence allows the problem of optimal point selection to be considered, that is, designing the optimal experiment.

Suppose that the total number of observations is N, and this number is usually predetermined by cost constraints. A sequence of experimental conditions, $\boldsymbol{x}_1 \ldots \boldsymbol{x}_N$, from a compact set $\mathscr{X}$ is referred to as a design of fixed size or *exact design* of size N (López-Fidalgo 2009). Some of these N points may be repeated, meaning that several observations are taken at the same value of $\boldsymbol{x}$. Therefore, a probability measure may be assigned to the design, assuming that only n of these points are different.

Kiefer (1959) and Kiefer, Wolfowitz et al. (1959) formed a theoretical framework of optimal design by expressing a design as a probability measure and representing the design as the allocations of observations at any particular point in the design space. If point $\boldsymbol{x}_i$ appears r_i times then weight $p_i = r_i/N$ is the probability of $\boldsymbol{x}_i$, that is the proportion of experiments to be made under these conditions. This exact design, realizable in integers r_i for a specific N, is defined by the following array:

$$\xi_N = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \\ p_1 & \ldots & p_n \end{array} \right\}, \tag{2-10}$$

i.e., is a probability discrete measure concentrated on points $\boldsymbol{x}_i$ with weights p_i. Here,

$$n \leq N, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i = \frac{r_i}{N}, \quad \sum_{i=1}^{n} r_i = N, \quad i = 1, 2, \ldots, n. \tag{2-11}$$

The set of all possible exact designs for a specific N is denoted by Ξ_N.

The p_i's in (2-10) are rational numbers. Removing this constraint, the weights are assumed to be real numbers in the interval $[0, 1]$. Designs that allow weights to vary continuously in $[0, 1]$ are called *continuous or approximate designs*. The continuous or approximate design is defined by a probabilistic measure (Atkinson et al. 2007, pp. 119,120):

$$\xi = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \\ w_1 & \ldots & w_n \end{array} \right\}, \tag{2-12}$$

where $\boldsymbol{x}_i$, $i = 1, 2, \ldots, n$, is a point of the compact set $\mathscr{X}$ with the w_i the associated design weights. Since ξ is a measure,

$$\int_{\mathscr{X}} \xi\,(d\boldsymbol{x}) = 1 \quad \text{and} \quad w_i > 0 \quad \text{for all} \quad i \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1. \tag{2-13}$$

If all points $\boldsymbol{x}_i$ in (2-12) are different, then the design is concentrated on n points called support points of design. The set of all possible continuous designs of the form (2-12) and (2-13) is denoted by Ξ.

However, all experimental designs are exact in practice. For moderate N, exact designs can be found by integer approximation to the optimal continuous design. The details of approximation rules are considered by Pukelsheim & Rieder (1992).

2.1.4. Information Matrix

For linear models the information matrix (Fisher) associated with the continuous design (2-12) defines the matrix (Atkinson et al. 2007, p. 121):

$$\boldsymbol{M}(\xi) = \int_{\mathscr{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T \xi(d\boldsymbol{x}) = \sum_{i=1}^{n} w_i \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)^T, \tag{2-14}$$

where $\boldsymbol{f}\,(\boldsymbol{x})$ is a vector of known "basis" regression functions. For an N-trial exact design the information matrix (normalized) for $\boldsymbol{\theta}$ in the linear model (2-3) is defined as (Ermakov & Zhiglijavsky 1987, p. 87)

$$\boldsymbol{M}(\xi_N) = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{X} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)^T. \tag{2-15}$$

If the experiment design has the form (2-10) and complies with the constraint (2-11), then the information matrix is equal to (Ermakov & Zhiglijavsky 1987, p. 87)

$$\boldsymbol{M}(\xi_N) = \sum_{i=1}^{n} \frac{r_i}{N} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)^T = \sum_{i=1}^{n} p_i \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)^T. \tag{2-16}$$

In linear models, the information matrices are determined only by design and are not dependent on observations y_i. A valuable property of linear models is that the information matrix does not depend on parameter values either.

For nonlinear models the information matrix (Fisher) associated with the continuous design (2-12) defines the matrix (Fedorov & Hackl 1997, p. 100):

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \int_{\mathscr{X}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})^T \xi(d\boldsymbol{x}) = \sum_{i=1}^{n} w_i \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})^T, \tag{2-17}$$

where $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \dfrac{\partial \eta(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the vector of partial derivatives of the model. For an N-trial exact design the information matrix (normalized) for $\boldsymbol{\theta}$ in nonlinear models is defined as (Fedorov & Leonov 2013, p. 18)

$$\boldsymbol{M}(\xi_N, \boldsymbol{\theta}) = \frac{1}{N} \boldsymbol{F}^T \boldsymbol{F} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})^T = \sum_{i=1}^{n} p_i \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})^T. \tag{2-18}$$

In nonlinear models the information matrices and, consequently, $Cov(\hat{\boldsymbol{\theta}})$ depend on the unknown parameters $\boldsymbol{\theta}$.

Two important properties of the information matrix are (Fedorov & Leonov 2013, pp. 6-7):

1. The information matrix is a symmetric nonnegative definite matrix.
2. The information matrix is additive, i.e., it is the sum of information matrices that corresponds to the individual observations.

The set of information matrices $\boldsymbol{M}(\xi, \boldsymbol{\theta})$ corresponding to all possible designs continuous ξ is denoted by $\mathcal{M}$. The goal of an optimal design is to select the values of the explanatory variables to maximize the information available from the experiment concerning the chosen optimality criterion.

2.1.5. Optimality Criteria

Some dimensional convex function of the information matrix, $\Psi\{\boldsymbol{M}(\xi)\}$ is used in optimal design theory. A function Ψ defined as follows

$$\Psi : \mathcal{M} \to \mathbb{R}^+ \cup \{+\infty\} \tag{2-19}$$

is called Ψ-optimality criteria. Three important properties of optimality criteria are (Fedorov & Leonov 2013, p. 61):

1. **Monotonicity.** Ψ is a monotonically nonincreasing function if

$$\Psi(\boldsymbol{M}) \leq \Psi(\boldsymbol{M}'), \quad \text{if} \quad \boldsymbol{M} \geq \boldsymbol{M}', \tag{2-20}$$

   where $\boldsymbol{M}$ and $\boldsymbol{M}'$ are nonnegative definite matrices. The second inequality in (2-20) is understood in terms of Loewner ordering.

   The ordering of nonnegative definite matrices, or Loewner ordering, is understood as $\boldsymbol{M} \geq \boldsymbol{M}'$, if $\boldsymbol{M}' = \boldsymbol{M} + \boldsymbol{C}$, $\boldsymbol{C} \geq 0$, where the latter inequality means that the square matrix $\boldsymbol{C}$ is nonnegative definite matrix of the same order as $\boldsymbol{M}$ and $\boldsymbol{M}'$ matrices, i.e., $\boldsymbol{z}^T \boldsymbol{C}\boldsymbol{z} \geq 0$ for any vector $\boldsymbol{z}$; see (Fedorov & Leonov 2013, p. 6).

2. **Homogeneity.**

$$\Psi(\gamma \boldsymbol{M}) = \gamma \Psi(\boldsymbol{M}), \tag{2-21}$$

   where $\gamma > 0$.

3. **Convexity.** A function $\Psi(\boldsymbol{M})$ defined on a convex set $\mathcal{M}$ is called convex if for any $\alpha \in [0, 1]$, and any $\boldsymbol{M}_1$, $\boldsymbol{M}_2 \in \mathcal{M}$

$$\Psi(\boldsymbol{M}) \leq (1 - \alpha)\Psi(\boldsymbol{M}_1) + \alpha \Psi(\boldsymbol{M}_2), \tag{2-22}$$

   where $\boldsymbol{M}$ is a convex combination of $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$, i.e., $\boldsymbol{M} = (1 - \alpha)\boldsymbol{M}_1 + \alpha \boldsymbol{M}_2$. The function $\Psi(\boldsymbol{M}) = \det \boldsymbol{M}^{-1}$ is not convex, but $\Psi(\boldsymbol{M}) = -\log\det \boldsymbol{M}$ is convex.

Several specific functions have been suggested as optimality criteria and have letters of the alphabet associated with them. Each optimality criterion identified a specific goal in the experiment. The most well known and widely used "alphabetic" design optimality criteria are:

- D**-optimality:** $\Psi(\boldsymbol{M}) = -\log\det \boldsymbol{M}$. It is the best known optimality criterion. D-optimality maximizes $\log\det \boldsymbol{M}(\xi)$ or minimizes $-\log\det \boldsymbol{M}(\xi)$. It is an estimation criterion which maximizes parameter information by minimizing variability of the parameter estimates (Fedorov & Leonov 2013, p. 53).

- D_s**-optimality:** $\Psi(\boldsymbol{M}) = -\log(\det \boldsymbol{M}/\det \boldsymbol{M}_{22})$, where $\boldsymbol{M}_{22}$ is the $(p-s)\times(p-s)$ lower right submatrix of $\boldsymbol{M}$. D_s-optimal designs are appropriate when interest is in estimating a subset of s of the parameters. The other $p - s$ parameters are then treated as nuisance parameters. The D_s-optimal design for the parameters of interest minimizes the optimality criterion Ψ (Atkinson et al. 2007, p. 138).

- A**-optimality:** $\Psi(\boldsymbol{M}) = -\mathrm{tr}\,(\boldsymbol{M}^{-1})$. This criterion minimizes the sum or average, of the variances of the parameter estimates (Atkinson et al. 2007, p. 137).
- c**-optimality:** $\Psi(\boldsymbol{M}) = c^T\boldsymbol{M}^{-1}c$, where c is a known vector of constants. In this criterion interest is in estimating the linear combination of the parameters $c^T\boldsymbol{\theta}$ with minimum variance (Rodríguez Torreblanca & Ortiz Rodríguez 1999, p. 39).
- G**-optimality:** $\Psi(\boldsymbol{M}) = \max_{\boldsymbol{x}\in\mathscr{X}} d(\boldsymbol{x},\xi)$, where $d(\boldsymbol{x},\xi)$ is the standardized variance of the predicted response. For linear models this variance is equal to

$$d(\boldsymbol{x},\xi) = \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{M}(\xi)^{-1}\boldsymbol{f}(\boldsymbol{x}) \tag{2-23}$$

  and for nonlinear models it is equal to

$$d(\boldsymbol{x},\xi) = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta})^T\boldsymbol{M}(\xi,\boldsymbol{\theta})^{-1}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta}). \tag{2-24}$$

  This criterion minimizes the largest possible value of the variance in the design space $\mathscr{X}$ and is called G-optimal (Rodríguez Torreblanca & Ortiz Rodríguez 1999, p. 43).

2.1.6. General Equivalence Theorem

The General Equivalence Theorem (Kiefer 1974) provides methods for the construction of optimal designs, and then it allows to check the optimality of the resulting designs.

Theorem 2.1. *Let $\mathscr{X}$ be a compact set. If Ψ is convex on $\mathcal{M}$, the space of design information matrices, and differentiable at $\boldsymbol{M}(\xi^*)$, then a Ψ-optimal design ξ^* can be equivalently characterized by any of the following three conditions* (Atkinson et al. 2007, p. 122)*:*

(i) The design ξ^* minimizes $\Psi\{\boldsymbol{M}(\xi)\}$.

(ii) The design ξ^* maximizes the minimum over $\mathscr{X}$ of $\phi(\boldsymbol{x},\xi)$.

(iii) The minimum over $\mathscr{X}$ of $\phi(\boldsymbol{x},\xi^*) = 0$, this minimum occurring at the points of support of the design. For any non-optimal design ξ the minimum over $\mathscr{X}$ of $\phi(\boldsymbol{x},\xi) < 0$.

Here $\phi(\boldsymbol{x},\xi)$ is the derivative of Ψ in the direction $\bar{\xi}$, defined as

$$\phi(\boldsymbol{x},\xi) = \lim_{\alpha\to 0^+}\frac{1}{\alpha}\left[\Psi\left\{(1-\alpha)\,\boldsymbol{M}(\xi) + \alpha\boldsymbol{M}(\bar{\xi})\right\} - \Psi\left\{\boldsymbol{M}(\xi)\right\}\right], \tag{2-25}$$

where the measure $\bar{\xi}$ puts unit mass at the point $\boldsymbol{x}\in\mathscr{X}$.

The Equivalence Theorem of Kiefer-Wolfowitz (Kiefer & Wolfowitz 1960) establishes the equivalence between G- and D-optimality. This theorem provides methods for the construction and checking of D-optimal designs.

Theorem 2.2. *If the set of information matrices $\mathcal{M}$ is compact, then the following three statements are equivalent* (Ermakov & Zhiglijavsky 1987, p. 109)*:*

(i) $\xi^* = \arg\max_{\xi\in\Xi}[\det\{\boldsymbol{M}(\xi)\}]$*, i.e., the design ξ^* is D-optimal.*

(ii) $\xi^* = \arg\min_{\xi\in\Xi}[\max_{\boldsymbol{x}\in\mathscr{X}} d(\boldsymbol{x},\xi)]$*, i.e., the design ξ^* is G-optimal.*

(iii) $\max_{\boldsymbol{x}\in\mathscr{X}} d(\boldsymbol{x},\xi^*) = p$*.*

The information matrices of all designs that meet one of the three stated statements coincide with each other. In points $\boldsymbol{x}_i$ of these designs, $d(\boldsymbol{x}_i,\xi^*) = p$.

Continuous designs that are D-optimal are also G-optimal, that is they minimize the maximum over $\mathscr{X}$ of the variance $d(\boldsymbol{x},\xi)$. It does not, in general, hold for exact designs. One exact design will be D-optimal, but will not be G-optimal.

This theorem does not say anything about the number of support points of the optimal design. Usually, optimal designs contain fewer points. For many D-optimal designs, especially for models in one factor, the designs contain p points, each with weight $1/p$, where p is the number of unknown parameters in the model.

2.2. Design Augmentation

The addition of new trials to an existing experimental design is meant as design augmentation. These designs are especially useful when found models are inadequate, as shown in George & Ogot (2006). Their purpose is to construct an experiment that incorporates existing data.

Let an initial design ξ_0 with N_0 prior observations (Atkinson et al. 2007, p. 315)

$$\xi_0 = \left\{ \begin{array}{ccc} x_1^0 & \dots & x_q^0 \\ w_1^0 & \dots & w_q^0 \end{array} \right\}, \tag{2-26}$$

with information matrix $N_0\boldsymbol{M}_0$ where

$$\boldsymbol{M}_0 = \sum_{i=1}^{q} w_i^0 f(x_i^0) f^T(x_i^0), \tag{2-27}$$

where the weights w_i^0 are multiples of $1/N_0$.

The design is increased by N additional observations. Augmentation of a design of size N_0 to one of size $N + N_0$ provides the new information in the experiment that comes

from the new N-trial design ξ with information matrix $N\boldsymbol{M}(\xi)$. Combining the prior information with augmentation of N trials yields the posterior information matrix

$$\tilde{\boldsymbol{M}}(\xi) = N_0\boldsymbol{M}_0 + N\boldsymbol{M}(\xi), \tag{2-28}$$

where $N\boldsymbol{M}(\xi)$ is the information matrix of design ξ. The D-optimal designs maximize $\det \tilde{\boldsymbol{M}}(\xi)$.

Let weights

$$\alpha = \frac{N_0}{N_0 + N} \quad \text{and} \quad 1 - \alpha = \frac{N}{N_0 + N}. \tag{2-29}$$

The normalized information matrix for a continuous design ξ is

$$\boldsymbol{M}_\alpha(\xi) = \alpha\boldsymbol{M}_0 + (1 - \alpha)\boldsymbol{M}(\xi). \tag{2-30}$$

Maximizing $\Psi\{\boldsymbol{M}_\alpha(\xi)\}$ for given α is equivalent to maximizing $\Psi\{\tilde{\boldsymbol{M}}(\xi)\}$ (2-28) for given N_0 and N.

The Equivalence Theorem states that the standardized variance for continuous D-optimal augmentation designs ξ^* fulfills (Atkinson, Bogacka & Zocchi 2000) for linear models:

$$d_\alpha(x, \xi^*) = (1-\alpha) f^T(x)\{\boldsymbol{M}_\alpha(\xi^*)\}^{-1} f(x) + \alpha \sum_{i=1}^{q} w_i^0 f^T(x_i^0)\{\boldsymbol{M}_\alpha(\xi^*)\}^{-1} f(x_i^0) \leq p \tag{2-31}$$

for all $x \in \mathscr{X}$, where p is the number of parameters in the model, that is the dimension of the information matrix $\boldsymbol{M}_\alpha(\xi)$. The first variance term is the posterior variance at a point in $\mathscr{X}$ and the second a weighted sum of posterior variances at the points of the prior design.

Analogous expression of $d_\alpha(x, \xi^*)$ is obtained for nonlinear models by replacing the vector of basis functions $\boldsymbol{f}(x)$ in (2-31) with the vector of partial derivatives $\boldsymbol{f}(x, \boldsymbol{\theta})$.

2.3. Bayesian Optimal Design

Any prior information is usually available for the experimentation. The idea of the Bayesian optimal design is to use this prior distribution on the unknown parameters in the optimal design process.

2.3.1. Utility Function

A Bayesian design problem is a problem of statistical decision (Chaloner & Verdinelli 1995), involving the design space, the utility function, and the distribution of the random

variables. Lindley (1972) introduced a decision-theoretic approach. He provided a mathematical basis for selecting Bayesian optimal designs. The following is the argument of Lindley.

A design ξ is selected from a set Ξ while the data $\boldsymbol{Y}$ are observed from a sample space $\mathscr{Y}$. Then, a decision d will be chosen from the decision set $\mathscr{D}$ based on the observation $\boldsymbol{Y}$. Thus, a decision consists of two parts: first, the selection of a design ξ, and second, the choice of a terminal decision d. It is assumed that the unknown parameters are $\boldsymbol{\theta} \in \Theta$, where Θ is the parameter space. The utility function is of the form $U(d, \boldsymbol{\theta}, \xi, \boldsymbol{Y})$.

Denoting by $\pi(\cdot)$ a prior probability density, the expected utility of the best decision for any design ξ is given by

$$U(\xi) = \int_{\mathscr{Y}} \max_{d \in \mathscr{D}} \int_{\Theta} U(d, \boldsymbol{\theta}, \xi, \boldsymbol{Y})\, \pi(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi)\, \pi(\boldsymbol{Y} \mid \xi)\, d\boldsymbol{\theta} d\boldsymbol{Y}. \tag{2-32}$$

The Bayesian solution to the experimental design problem is the design ξ^* that maximizes the equation (2-32):

$$U(\xi^*) = \max_{\xi \in \Xi} \int_{\mathscr{Y}} \max_{d \in \mathscr{D}} \int_{\Theta} U(d, \boldsymbol{\theta}, \xi, \boldsymbol{Y})\, \pi(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi)\, \pi(\boldsymbol{Y} \mid \xi)\, d\boldsymbol{\theta} d\boldsymbol{Y}. \tag{2-33}$$

Thus, according to Chaloner & Verdinelli (1995) Lindley's argument suggests that a good way to design an experiment is to consider the design as a decision problem. The procedure consists in choosing a utility function that can adequately describe the objectives of the experiment and then selecting a design that maximizes the expected utility. Thereby, the Bayesian solution is to find the best design and the best decision rule that maximizes expected utility.

2.3.2. Bayesian Optimality Criteria for Linear Models

Consider the problem of choosing a design ξ for a normal linear regression model defined in (2-2). Suppose that a prior distribution $\pi(\boldsymbol{\theta}, \sigma^2)$ on $\boldsymbol{\theta}$, σ^2 is given such that the conditional prior distribution $\pi(\boldsymbol{\theta} \mid \sigma^2)$ of $\boldsymbol{\theta}$ given σ^2 is $\mathscr{N}(\mu, \sigma^2 \boldsymbol{R}^{-1})$, where $\boldsymbol{R}$ is a given positive definite $p \times p$ "precision" matrix.

Under the mentioned assumptions the posterior conditional distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2)$ of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$, σ^2 is normal with mean vector (Chaloner & Verdinelli 1995)

$$\widehat{\boldsymbol{\theta}}_B = E(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2) = (\boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{R})^{-1} (\boldsymbol{X}^T \boldsymbol{Y} + \boldsymbol{R}\mu) \tag{2-34}$$

and covariance matrix $\sigma^2 (\boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{R})^{-1}$ (see Appendix B).

The expected gain in Shannon information is extensively used as a utility function (Bernardo 1979). The Bayesian design is chosen to maximize the expected gain in Shannon information or, equally, maximize the expected Kullback-Leibler distance between the posterior and the prior distributions (Chaloner & Verdinelli 1995):

$$\int \log \frac{\pi\left(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi\right)}{\pi\left(\boldsymbol{\theta}\right)} \pi\left(\boldsymbol{Y}, \boldsymbol{\theta} \mid \xi\right) d\boldsymbol{\theta}\, d\boldsymbol{Y}. \tag{2-35}$$

The prior distribution is not dependent on the design ξ. Therefore, the design ξ maximizing the expected gain in Shannon information is actually the design that maximizes the following integral:

$$U_1\left(\xi\right) = \int \log \pi\left(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi\right) \pi\left(\boldsymbol{Y}, \boldsymbol{\theta} \mid \xi\right) d\boldsymbol{\theta}\, d\boldsymbol{Y}. \tag{2-36}$$

This integral represents the expected Shannon information of the posterior distribution. In the normal linear regression model

$$U_1\left(\xi\right) = -\frac{p}{2}\log(2\pi) - \frac{p}{2} + \frac{1}{2}\log\det\{\sigma^{-2}\left(N\boldsymbol{M}(\xi) + \boldsymbol{R}\right)\}. \tag{2-37}$$

After dropping the constant and multiplier terms in Equation (2-37), we can obtain the optimality criterion

$$\Psi_1\left(\xi\right) = \det\left\{\boldsymbol{M}(\xi) + \frac{1}{N}\,\boldsymbol{R}\right\} = \det \boldsymbol{M}_B(\xi) \tag{2-38}$$

and it is known as Bayesian D-optimality criterion for linear models, where non-Bayesian D-optimality maximizes the determinant of $\boldsymbol{M}(\xi)$ and Bayesian D-optimality maximizes the determinant of $\boldsymbol{M}_B(\xi)$, which does not depend on the prior value μ. When the sample size N is large or the matrix $\boldsymbol{R}$ corresponds to imprecise information, the difference between a Bayesian design and its corresponding non-Bayesian one can be small.

The equivalence theory (Kiefer 1974) supplies a method for verifying the optimality of any particular design. If ξ^* is the optimal design, the support points will be at the roots of the function $d(\xi^*, x)$, where $d(\xi^*, x) = f\left(x\right)^T \boldsymbol{M}_B\left(\xi\right)^{-1} f\left(x\right)$ is the directional derivative and $f\left(x\right)$ are the "basis" functions of the linear model.

Theorem 2.3. (Equivalence Theorem). *Let $\mathscr{X}$ be a compact set. If Ψ is convex on $\mathcal{M}$, the space of design information matrices, and differentiable at $\boldsymbol{M}(\xi^*)$, where ξ^* is the optimal design, then the following three conditions are equivalent* (Pilz 1991, p. 140) *:*

(i) The design $\xi^* \in \Xi$ is Bayesian D-optimal in Ξ if and only if

$$\sup_{x \in \mathscr{X}} f\left(x\right)^T \boldsymbol{M}_B\left(\xi\right)^{-1} f\left(x\right) = \operatorname{tr} \boldsymbol{M}\left(\xi^*\right) \boldsymbol{M}_B\left(\xi^*\right)^{-1}. \tag{2-39}$$

(ii) For all $\xi \in \Xi$ *and* $\xi^* \in \Xi_B$ *it holds*

$$1 \geq \frac{\det \boldsymbol{M}_B\left(\xi^*\right)^{-1}}{\det \boldsymbol{M}_B\left(\xi\right)^{-1}} \geq 1 + \operatorname{tr} \boldsymbol{M}(\xi)\, \boldsymbol{M}_B\left(\xi\right)^{-1} - \sup_{x \in \mathscr{X}} f\left(x\right)^T \boldsymbol{M}_B\left(\xi\right)^{-1} f\left(x\right). \quad \text{(2-40)}$$

(iii) The design $\xi^* \in \Xi$ *can be Bayesian D-optimal in* Ξ *only if*

$$\operatorname{supp} \xi^* \subseteq \left\{ x \in \mathscr{X} : f\left(x\right)^T \boldsymbol{M}_B\left(\xi\right)^{-1} f\left(x\right) = \operatorname{tr} \boldsymbol{M}(\xi^*)\, \boldsymbol{M}_B\left(\xi^*\right)^{-1} \right\}. \quad \text{(2-41)}$$

The proof of Theorem 2.3 can be found in Pilz (1991, p. 140).

2.3.3. Bayesian Optimality Criteria for Nonlinear Models

Experimental design is usually more difficult to find in nonlinear models than in linear models. The reason is that their Fisher information matrix usually depends on the unknown parameters, which can not be separated as a simple multiplier.

In non-Bayesian designs, the parameters in the Fisher information matrix are usually replaced by supposed values of the parameters, called "guesses" (Chernoff 1953). In the Bayesian optimal design approach, the assumptions do not concentrate on single values. Instead, a prior distribution is assigned to each unknown parameter. These distributions can be centered around the assumed parameter values. The Bayesian optimality criterion is to minimize the Bayes risk by integrating the risk function over the prior distribution (Chaloner & Verdinelli 1995).

The asymptotic approximations can be used for the nonlinear models since their exact posterior distributions are often intractable. The normal approximation to the posterior distribution is usually used. For a nonlinear model with unknown parameters $\boldsymbol{\theta}$, a design ξ and a sample size of N, the expected Fisher information matrix is denoted by $N\boldsymbol{I}(\xi, \boldsymbol{\theta})$, where

$$\boldsymbol{I}(\xi, \boldsymbol{\theta}) = \int_{\mathscr{X}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \eta(x; \boldsymbol{\theta})\right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} \eta(x; \boldsymbol{\theta})\right]^T d\xi\left(x\right). \quad \text{(2-42)}$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$. One normal approximation can be written as

$$\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi \sim \mathscr{N}\left(\hat{\boldsymbol{\theta}}, \left[N\boldsymbol{I}\left(\xi, \hat{\boldsymbol{\theta}}\right)\right]^{-1}\right), \quad \text{(2-43)}$$

where $N\boldsymbol{I}\left(\xi, \hat{\boldsymbol{\theta}}\right)$ is the observed Fisher information matrix inserting $\hat{\boldsymbol{\theta}}$. In (2-43) the posterior normal approximation only depends on the data through $\hat{\boldsymbol{\theta}}$.

If using this approximation in the Equation (2-36), $U_1(\xi)$ can be written as

$$U_1(\xi) = -\frac{p}{2}\log(2\pi) - \frac{p}{2} + \frac{1}{2}\int \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{2-44}$$

The Bayesian optimality criterion can be obtained dropping the constant and multiplier terms in Equation (2-44)

$$\Psi_1(\xi) = \int \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{2-45}$$

and it is known as Bayesian D-optimality criterion for nonlinear models, where Bayesian D-optimality maximizes the criterion $\Psi_1(\xi)$.

In the general case, using normal approximation can be written Bayesian optimality criterion as

$$\Psi_B(\xi) = \int \Phi\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{2-46}$$

where $\Phi\{\cdots\}$ corresponds to the convex functional, defined in the set of expected Fisher information matrices.

In order to examine whether a given design is Bayesian D-optimal or not, we use an equivalence theorem given in Dette & Neugebauer (1997), which characterizes Bayesian D-optimal design for nonlinear models.

Theorem 2.4. *Let Ξ be not empty and $|\Psi_1(\xi)| < \infty$ for all $\xi \in \Xi$, then the following conditions are equivalent:*

(i) a design $\xi^* \in \Xi$ is Bayesian D-optimal within the class Ξ;

(ii) ξ^* minimizes $\max_{x\in\mathscr{X}} d(\xi, x)$;

(iii) $\max_{x\in\mathscr{X}} d(\xi^*, x) = p$, where

$$d(\xi,x) = \int_\Omega f(x,\boldsymbol{\theta})^T\,\boldsymbol{I}(\xi,\boldsymbol{\theta})\,f(x,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} \tag{2-47}$$

and

$$\Xi := \left\{\xi \mid \max_{x\in\mathscr{X}} \int_\Omega f(x,\boldsymbol{\theta})^T\,\boldsymbol{I}(\xi,\boldsymbol{\theta})\,f(x,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} < \infty\right\}. \tag{2-48}$$

The proof of Theorem 2.4 follows directly from Chaloner & Larntz (1989).

2.4. Minimization Algorithms

The Bayesian methods for optimal design require numerical optimization and integration because the searching the optimum of the Bayesian optimality criteria $\Psi_B(\xi)$ cannot be performed analytically. Therefore we need to use the corresponding software to find such optimal designs. There are a lot of optimization algorithms. Three iterative algorithms are presented here.

2.4.1. Linearization Method

The Gauss-Newton method is a commonly used iterative linearization method for minimizing an objective function. In our case, an optimality criterion is minimized. Given (Lange 2010)

$$\xi^{(0)} = \left\{ \begin{array}{ccc} x_1^0 & \dots & x_n^0 \\ w_1^0 & \dots & w_n^0 \end{array} \right\}, \tag{2-49}$$

as an initial design, a Taylor series expansion is used to approximate the optimality criterion as follows:

$$\Psi(\xi) \approx \Psi(\xi^{(a)}) + \Delta^{(a)}(\xi - \xi^{(a)}), \tag{2-50}$$

where $\Delta^{(a)}$ is the gradient of the criterion Ψ in the point $\xi^{(a)}$ in the direction $(\xi - \xi^{(a)})$. This direction is chosen with the objective of achieving the maximum decrease of the criterion, $\Psi(\xi)$ to obtain $\delta^{(a)} = \xi - \xi^{(a)}$. The algorithm sequentially computes approximations, $\xi^{(1)}$, $\xi^{(2)}$, ..., where $\xi^{(a+1)} = \xi^{(a)} + \delta^{(a)}$, until the value of $\delta^{(a)}$ is negligible and $\Psi(\xi^{(a)})$ is minimized. The last iteration provides the optimal design ξ^*.

This algorithm is available using R software through some command lines as *nlm*, *nlminb* and *optim*. These commands are included in the R Stats Package that is available in the R standard installation (Team 2018). The *nlm* command carry out a minimization of the function using a Newton-type algorithm. The *nlminb* command allows to make an unconstrained and box-constrained optimization using PORT routines and finally *optim* command makes optimization of the function based on several methods (quasi-Newton, conjugate-gradient and Nelder-Mead algorithms) (Team 2018).

2.4.2. Direct Search Algorithm

The simplex algorithm by Nelder & Mead (1965) is one of the most well-known algorithms for unconstrained multidimensional optimization without derivatives. It can be employed for discontinuous function problems that are common in statistical research. It is often used for statistical parameter estimation problems when the function values are uncertain or noisy (Parker & Gennings 2008).

The Nelder-Mead method is *simplex*-based. A *simplex* $S \in \mathbb{R}^u$ is defined as the convex hull of $u+1$ vertices $x_0, x_1, \ldots, x_u \in \mathbb{R}^n$. For example, a simplex in $\mathbb{R}^2$ is a *triangle*, and a simplex in $\mathbb{R}^3$ is a *tetrahedron*.

A simplex-based direct search method begins with a set of $u+1$ points $x_0, x_1, \ldots, x_u \in \mathbb{R}^u$ considered to be the vertices of a working simplex S and the corresponding set values of the evaluation function at those vertices (Nelder & Mead 1965). The initial working simplex S does not have to be degenerate, i.e., the points $x_0, x_1, \ldots, x_u$ should not be in the same hyperplane. Then, the process executes a transformation sequence of the working simplex S in order to reduce the values of the function at its vertices. The transformation is tested in each phase by calculating the function values at one or more test points and comparing them with those at the vertices. This process is stopped when the working simplex S becomes sufficiently small and fulfills the test criterion. The points of the final simplex are proposed as optimal search points.

The Nelder-Mead method can be used in the optimal design theory to find the optimal design. First, the user must specify starting values for design support points, x_i^0, and their weights, w_i^0, to form an initial simplex, where $i = 1, 2, \ldots, n$. The design criterion, $\Psi(\boldsymbol{x}, \boldsymbol{w})$, is evaluated in these initial points and weights. Then the simplex moves towards the minimum of the criterion in the $2n$-dimensional space, following a set of rules until convergence. The design criterion values, $\Psi(\boldsymbol{x}, \boldsymbol{w})$, are calculated and compared in each step of this iterative process. The values of $\boldsymbol{x}$ and $\boldsymbol{w}$, which correspond to the found minimum value of the criterion, form the optimal experimental design ξ^*.

This algorithm is available through command lines such as *nelminb* and *neldermead* in R software (Team 2018) with the packages *adagio* (Borchers 2018) and *nloptr* (Ypma, Borchers & Eddelbuettel 2014) respectively.

2.4.3. Differential Evolution Algorithm

Differential Evolution (DE) algorithm is a stochastic, population-based optimization method minimizing an objective function that was introduced by Storn & Price (1997). DE belongs to the class of genetic algorithms (GAs) which are inspired by the principle of natural evolution and have advantages over traditional methods of optimization (Holland 1992).

DE approach can be used to find approximate solutions to such problems where objective functions are non-differentiable, non-continuous, nonlinear, noisy, flat, multidimensional or have many local minima, constraints or stochasticity. This approach eliminates the need for explicit consideration of a set of initial points and can be used in highly

constrained regions. Instead of the point-to-point search in traditional methods, this algorithm moves from one population to another and thus traverses the design space in many directions simultaneously, thus preventing local optimum from occurring.

The DE algorithm is an iterative optimization procedure like GAs using the similar three main types of operators at each step to create the next generation from the current population:

- *Selection operators* select the individuals, called parents, that contribute to the population at the next generation.
- *Crossover (recombination) operators* combine two parents to form children for the next generation.
- *Mutation operators* apply random changes to individual parents to form children.

The main difference is that the genetic algorithms are based on the crossover (recombination), while the DE is based on the mutation operation as a search mechanism.

The DE algorithm begins with a relatively large population that is randomly created at the start. Then, mutation, crossover, and selection operations are iteratively applied to the population in order to improve individual fitness of the individuals. Fitness can be evaluated by the value of a suitable objective function. In our case, this objective function is the design optimality criterion, Ψ, and the individuals are possible optimal designs. Finally, these iterative procedures converge to the global optimum regardless of the complexity of this objective function and the initial design.

The DE algorithm is available with the package *DEoptim* (Mullen, Ardia, Gil, Windover & Cline 2016) in R software through command line as *DEoptim* (Team 2018). Cardona, López & Correa (2012) implemented this DE algorithm for the construction of Bayesian D-optimal designs for parameters estimation in pharmacokinetics maximizing the expected utility.

The GA is a stochastic optimization method so that in the limited number of iterations, it hardly achieves the exact optimum. It can, however, usually reach near optimum. Hamada, Martz, Reese & Wilson (2001) used GAs to find near-optimal Bayesian design for several regression models. The found near-optimal design points can, therefore, be used as the starting points for other local optimization algorithms. Concerning our problem, the near-optimal design points of the GAs can be used as starting points, $\boldsymbol{x}_0$, for the initial design of the Nelder-Mead algorithm to avoid the local optimum getting stuck. Zhang (2006) implemented a hybrid method combining a GA and a local optimization algorithm to find Bayesian D-optimal design for generalized linear models.

3. Penalized Optimal Designs

Desirability functions have been mainly implemented in the manufacturing and industrial sectors. They are used by engineers to optimize product quality as the most popular way to analyze several results simultaneously (Wu 2005, Becerra Rodríguez, Zitzumbo Guzmán, Domínguez Domínguez, García Alcaraz & Alonso Romero 2014). However, the desirability functions usually are not used in statistical procedures, particularly in the construction of optimal designs. This chapter contains basic information about desirability functions and how to use them to obtain penalized optimal designs.

Section 3.1 provides a summary of the desirability function theory used in this dissertation. This summary is based on the works of Harrington (1965), Derringer & Suich (1980), and Gibb (1998). The desirability function concept and methodology of its use for optimizing quality characteristics were developed by Harrington (1965), and later they were modified by Derringer & Suich (1980). Finally, the application of this methodology was extended by Gibb (1998) introducing continuous and differentiable desirability functions.

Section 3.2 presents an approach of penalized optimal design strategy using desirability functions. This strategy enables the researcher to determine the specific characteristics of the experimental design, which are considered important from a practical perspective. As a result, the obtained design fulfills traditional optimal design criteria and also practical preferences imposed by a researcher. The penalized optimal criterion is described and supported in Section 3.3. The methodology of choosing the appropriate desirability functions is presented in Section 3.4, and it is also illustrated by an example of the Michaelis-Menten model in Section 3.5. This example describes the detailed steps to follow for obtaining the penalized optimal design with desired experimental characteristics.

3.1. Desirability Functions

Researchers in many areas of science are faced with the problem of simultaneously improving multiple responses that depend on a common set of controllable variables. Desirability optimization methodology addresses this problem. Harrington (1965) developed the concept of a desirability function to solve a multivariate optimization problem. This

concept combines the responses of several factors into a single function in order to optimize the outcome of a process.

Responses of each factor, X_i, $i = 1, 2, \ldots, k$, are converted to a dimensionless, ordinal measure d_i, $0 \leq d_i \leq 1$, where a value of 0 indicates that the response is undesirable and a value of 1 denotes a desirable response. According to Harrington (1965) an abbreviated summary of Harrington's guidelines for associating X_i with d_i is provided in Table **3-1**.

Table 3-1.: Abbreviated version of Harrington's guidelines for transforming levels of X_i to the desirability scale.

d_i	Quality Equivalent
1.00	The highest quality
1.00 - 0.80	Acceptable and excellent
0.80 - 0.63	Acceptable and good
0.63 - 0.40	Acceptable but poor
0.40 - 0.30	Borderline
0.30 - 0.00	Unacceptable
0.00	Completely unacceptable

The index i represents the ith desirability function or the ith response of interest. The shape of the desirability function is determined by whether one is trying to maximize or minimize the response or target a range of values.

The "target" desirability function is designed for cases where lower and upper "specification limits" exist, denoted by $X_{i,\min}$ and $X_{i,\max}$, respectively. The goal is to maintain X_i within these limits. A trivial "target" desirability function is given by

$$d_i = \begin{cases} 1 & X_{i,\min} < X_i < X_{i,\max} \\ 0 & \text{otherwise.} \end{cases} \tag{3-1}$$

This trivial desirability function has two shortcomings. First, a discrete change in desirability at the specification limits is not realistic. Second, this function does not account for uncertainty in the estimated response. Harrington (1965) formulated a more practical desirability function given by

$$d_i = \exp\left(- \left| \frac{2X_i - (X_{i,\max} + X_{i,\min})}{X_{i,\max} - X_{i,\min}} \right|^{\nu_i} \right). \tag{3-2}$$

The parameter ν_i, $0 < \nu_i < \infty$, is specified by the researcher to control both the slope of d_i and the interval over which $d_i \approx 1$. As $\nu_i \to \infty$, d_i converges to the trivial desirability function.

Harrington's "bigger-is-better" and "smaller-is-better" desirability functions are applicable where a single "specification limit" exists. The goal is for X_i, or $-X_i$ depending on the problem, to be above this limit. Harrington (1965) considered the smooth shape more appropriate. He formulated this shape using a special form of the Gompertz growth curve

$$d_i = \exp\left(-\exp\left(-\left(a_i + b_i X_i\right)\right)\right), \tag{3-3}$$

where a_i and b_i are specified so that $d_i = 1/e$ at the specification limit and the slope reflects the rate at which quality is affected by the level of the response.

Each single desirability function can then be combined into composite desirability, which allows the simultaneous consideration of multiple constraints. Harrington (1965) employed the geometric mean of individual desirability functions to define the overall desirability of the combined factors:

$$D = \left(d_1 \times d_2 \times \cdots \times d_k\right)^{1/k}. \tag{3-4}$$

By defining overall desirability as a geometric mean, Harrington (1965) transformed the multivariate optimization problem to a univariate maximization problem. Since the geometric mean heavily weights small d_i, this approach avoids factor levels unfavorable to even a single response. It is clear that if any individual desirability function d_i is zero, the overall desirability D will also be zero.

Derringer & Suich (1980) extended the methodology of Harrington by defining flexible, piecewise-continuous "bigger-is-better", "smaller-is-better" and "target" desirability functions with shape parameters specified by a user. For the "bigger-is-better" case, d_i increases as X_i increases and is employed when X_i is to be maximized and defined as

$$d_i = \begin{cases} 0 & X_i < X_{i,\min} \\ \left(\frac{X_i - X_{i,\min}}{X_{i,\max} - X_{i,\min}}\right)^{s_i} & X_{i,\min} < X_i < X_{i,\max},\ s_i > 0 \\ 1 & X_i > X_{i,\max} \end{cases}. \tag{3-5}$$

and graphed in Figure **3-1**(a).

The value $X_{i,\min}$ gives the minimum acceptable value of X_i. The value $X_{i,\max}$ gives the highest value of X_i. The value of s_i used in the transformation would again be specified by the user. Figure **3-1** indicates a large value of s_i would be specified if it were very

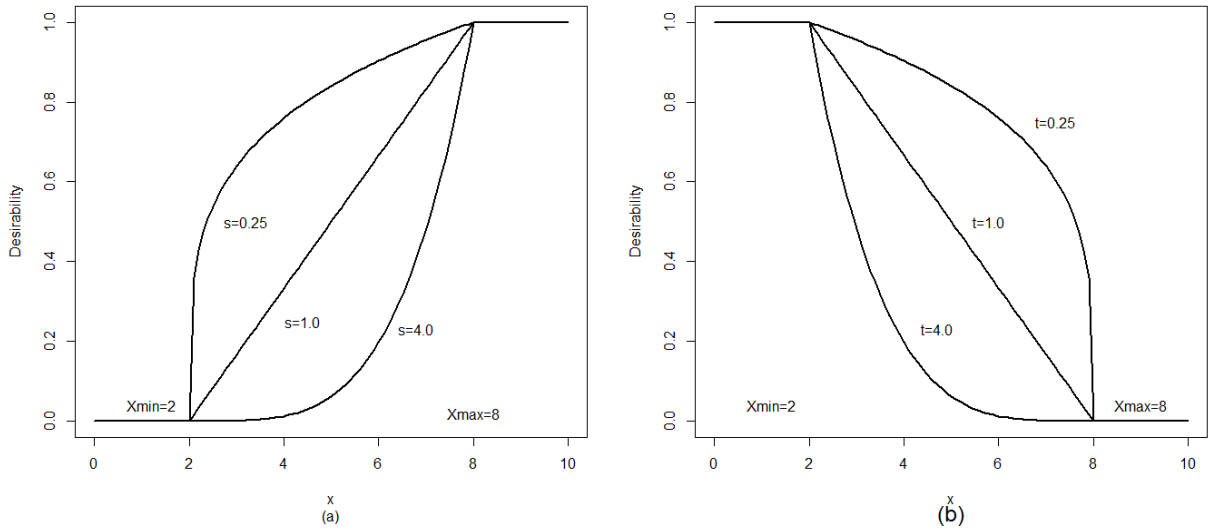

Figure 3-1.: (a) "Bigger-is-better" desirability function in (3-5) and (b) "smaller-is-better" desirability function in (3-6).

desirable for the value of X_i to increase rapidly above $X_{i,\min}$. X_i must be considerably greater than $X_{i,\min}$. On the other hand, a small value of s_i would be specified if having values of X_i considerably above $X_{i,\min}$ were not of critical importance.

For the "smaller-is-better" case, d_i decreases as X_i increases and is employed when X_i is to be minimized y defined as

$$d_i = \begin{cases} 1 & X_i < X_{i,\min} \\ \left(\frac{X_i - X_{i,\max}}{X_{i,\max} - X_{i,\min}}\right)^{t_i} & X_{i,\min} < X_i < X_{i,\max},\ t_i > 0 \\ 0 & X_i > X_{i,\max}, \end{cases} . \tag{3-6}$$

and graphed in Figure **3-1**(b).

The value $X_{i,\max}$ gives the maximum acceptable value of X_i. The value $X_{i,\min}$ gives the lowest value of X_i. The value of t_i used in the transformation would again be specified by the user. The values of t_i in the "smaller-is-better" function play the similar role as s_i does in the "bigger-is-better" function.

The "target" desirability function is employed when there are lower and upper "specification limits" designated $X_{i,\min}$ and $X_{i,\max}$ respectively, where c_i is the target response. The user-specified shape parameters, s_i and t_i, determine the exact form of the desir-

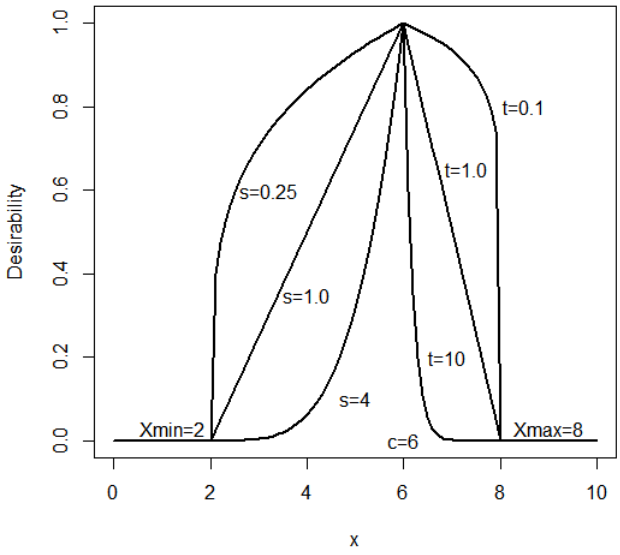

Figure 3-2.: "Target" desirability function in (3-7).

ability function as shown in Figure **3-2** and defined as

$$d_i = \begin{cases} 0 & X_i < X_{i,\min},\ X_i > X_{i,\max} \\ \left(\frac{X_i - X_{i,\min}}{c_i - X_{i,\min}}\right)^{s_i} & X_{i,\min} < X_i < c_i,\ s_i > 0 \\ \left(\frac{X_i - X_{i,\max}}{c_i - X_{i,\max}}\right)^{t_i} & c_i < X_i < X_{i,\max},\ t_i > 0. \end{cases} \quad . \tag{3-7}$$

The value selected for c_i would be that value of X_i, which was most desirable and could be selected anywhere between $X_{i,\min}$ and $X_{i,\max}$. The values of s_i and t_i in the "target" desirability function play much the same role as s_i does in the "bigger-is-better" function.

Derringer (1994) suggested employing a weighted geometric mean, as an extension of the overall desirability. This mean allows each individual desirability function to be assigned different importance. The weighted overall desirability has the following form, in which an exponent, w_j weights each of the k responses by importance:

$$D = (d_1^{w_1} \times d_2^{w_2} \times \cdots \times d_k^{w_k})^{1/\sum_{j=1}^{k} w_j} . \tag{3-8}$$

Del Castillo, Montgomery & McCarville (1996) noted difficulties applying Derringer and Suich's desirability functions with common software applications, such as the electronic spreadsheet, which use gradient-based optimization algorithms because Derringer and Suich's d_i contain nondifferentiable points. They solved the problem by redefining d_i in

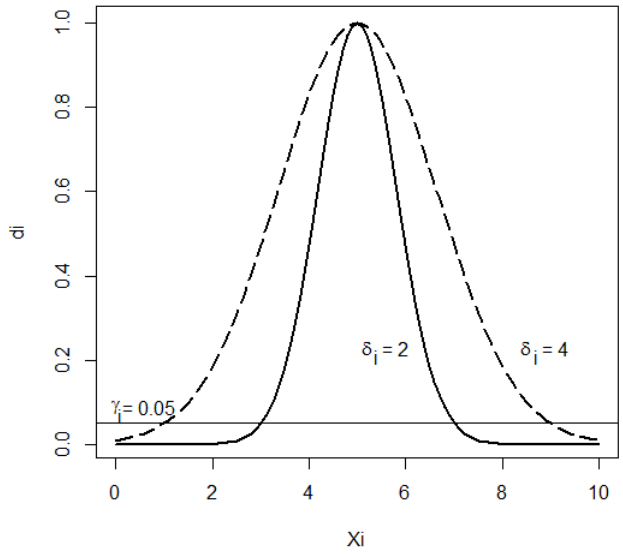

Figure 3-3.: Plots of the smooth "target" desirability function in (3-9).

nondifferentiable neighborhoods with cubic splines.

Gibb (1998) solves this problem differently by extending the desirability function methodology to comprise continuous and differentiable desirability functions. The responses of each factor are converted into a continuous and differentiable function with a range from 0 to 1. Gibb proposed the normal and logistic densities as reasonable candidates for desirability functions.

The desirability function defined in

$$d_i = \exp\left(-\frac{1}{2}\left(\frac{X_i - a_i}{b_i}\right)^2\right) \tag{3-9}$$

serves as a continuous, differentiable "target" desirability function. The parameter a_i is the target response level and the parameter b_i controls the function spread. To facilitate the specification of b_i, the parameter b_i is defined in terms of two other parameters, δ_i and γ_i as given in

$$b_i = \frac{\delta_i}{\sqrt{-2\log(\gamma_i)}}, \quad \gamma_i \in (0,1). \tag{3-10}$$

Under this parametrization $d_i = \gamma_i$ when $X_i = a_i \pm \delta_i$.

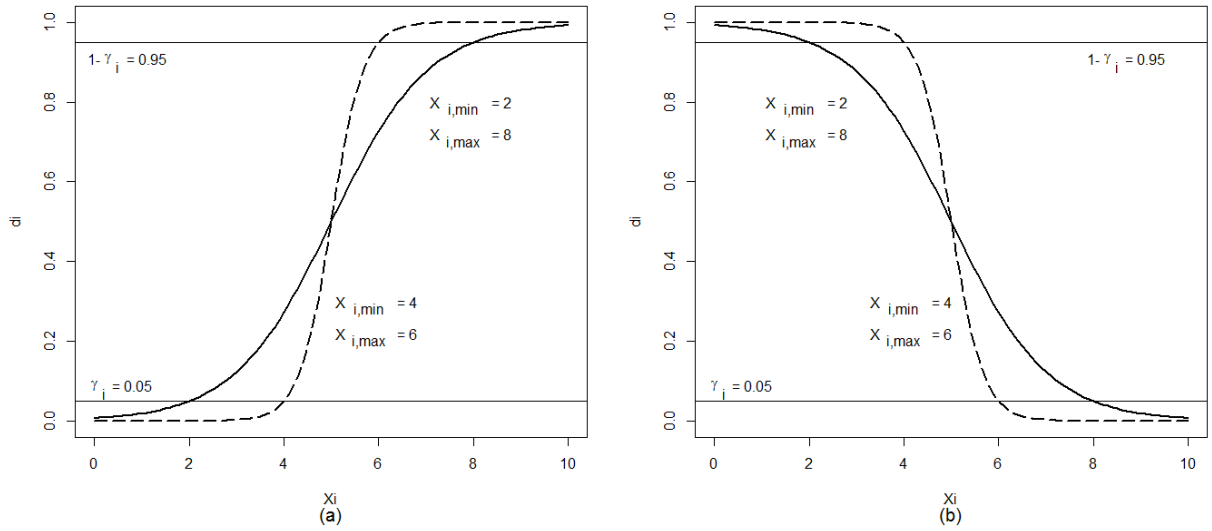

Figure 3-4.: Plots of (a) the smooth "bigger-is-better" desirability function in (3-11) and (b) the smooth "smaller-is-better" desirability function in (3-14).

Figure **3-3** presents plots of the smooth "target" desirability function, where $a_i = 5$, $\gamma_i = 0.05$, $\delta_i = 2$ (solid line) and $\delta_i = 4$ (dashed line). It observes that the interval width of desirability function increases as δ_i increases.

Gibb proposed the logistic cumulative distribution function (CDF) for the "bigger-is-better" desirability function:

$$d_i = \left(1 + \exp\left(-\frac{X_i - a_i}{b_i}\right)\right)^{-1}, \tag{3-11}$$

where

$$a_i = \frac{X_{i,\max} + X_{i,\min}}{2} \tag{3-12}$$

and

$$b_i = \frac{X_{i,\max} - X_{i,\min}}{2\log\left(\dfrac{1-\gamma_i}{\gamma_i}\right)}, \quad \gamma_i \in (0,1). \tag{3-13}$$

Under this parametrization $d_i(X_{i,\min}) = \gamma_i$ and $d_i(X_{i,\max}) = 1 - \gamma_i$.

Figure **3-4**(a) presents plots of the smooth "bigger-is-better" desirability functions, where $a_i = 5$, $\gamma_i = 0.05$, $(X_{i,\max}, X_{i,\min}) = (2, 8)$ (solid line) and $(X_{i,\max}, X_{i,\min}) = (4, 6)$ (dashed line). It observes that the slope of desirability function increases as the difference between $X_{i,\max}$ and $X_{i,\min}$ decreases.

The "smaller-is-better" or minimizing desirability function is obtained by simply changing the sign of the exponential argument in the logistic CDF, having the resulting form

$$d_i = \left(1 + \exp\left(\frac{X_i - a_i}{b_i}\right)\right)^{-1}. \tag{3-14}$$

In such a case, the roles of $X_{i,\min}$ and $X_{i,\max}$ are reversed. Figure **3-4**(b) presents plots of the smooth "smaller-is-better" desirability functions. It is observed that the slope of desirability function increases as the difference between $X_{i,\max}$ and $X_{i,\min}$ decreases.

An asymmetric "target" desirability function can be constructed as a product of a "bigger-is-better" desirability function (3-11) and a "smaller-is-better" desirability function (3-14): $d_{i(\text{target})} = d_{i(\max)} \times d_{i(\min)}$ (Gibb 1998). This procedure allows to incorporate asymmetry into the desirability function.

3.2. Use of Desirability Functions in Optimal Designs

The optimal design theory can produce inadequate designs from a practical point of view that can create conflict with common laboratory practice or other conventional guidelines. Parker (2005) proposed to combine optimal design theory with desirability functions integrating desired experimental characteristics into an optimal design.

An overall desirability function, $D(\boldsymbol{x}, \boldsymbol{w}) = (d_1 \times d_2 \times \cdots \times d_k)^{1/k}$, is defined in (3-4), for any k experimental design preferences as functions of the support points, $\boldsymbol{x}$, and their weights, $\boldsymbol{w}$. The function $(1 - D(\boldsymbol{x}, \boldsymbol{w}))$ is designated as the penalty function, where its value of 1 indicates an experimental design with undesirable properties.

Let $\Psi(\boldsymbol{x}, \boldsymbol{w})$ be an "alphabetic" optimality criterion for example, D-optimality criterion. The penalty function, $(1 - D(\boldsymbol{x}, \boldsymbol{w}))$, is added to the optimality criterion to penalize experimental designs. A user specified positive constant, Λ, is required to bring these two functions to similar scales, since the minimum value of the "alphabetical" optimality criterion is numerically small, while the penalty function has a range from 0 to 1. This constant can also control the penalty weight, relative to the minimum optimality criterion. Parker (2005) proposed to find a penalized optimal design by minimizing the new criterion

$$\Psi_P(\boldsymbol{x}, \boldsymbol{w}) = \Psi(\boldsymbol{x}, \boldsymbol{w}) + \Lambda(1 - D(\boldsymbol{x}, \boldsymbol{w})) \tag{3-15}$$

with respect to $\boldsymbol{x}$ and $\boldsymbol{w}$ for a given value of Λ.

The iterative optimization algorithms, described in Section 2.4, are used to evaluate $\boldsymbol{x}$ and $\boldsymbol{w}$ values that minimize the new criterion (3-15), given a value of Λ. The initial value

of Λ is selected as $\Lambda_0 = |\min\{\Psi(\boldsymbol{x}, \boldsymbol{w})\}|$, the absolute minimum value of the considered "alphabetic" optimality criterion, e.g. the D-optimality criterion. Penalized optimal designs are generated by minimizing the penalized optimal criterion (3-15) for values of Λ in multiples o submultiples of Λ_0. The final value of Λ is selected in the range in which stability is shown in the responses of the overall desirability function. The penalized optimal design obtained is determined by the $\boldsymbol{x}$ and $\boldsymbol{w}$ values associated with the minimum value of the penalized optimal criterion (3-15) for a value of Λ in this range. The resulting penalized optimal design is optimal according to the "alphabetic" design criterion and also has the desired practical characteristics.

3.3. Justification of New Criterion

Penalized optimal designs are obtained by minimizing the penalized optimal criterion (3-15) with respect to $\xi \in \Xi$ for a given value of $\Lambda > 0$, i.e.,

$$\xi_P^* = \arg\min_{\xi\in\Xi} \Psi_P(\xi) = \arg\min_{\xi\in\Xi} \{\Psi(\xi) + \Lambda(1 - D(\xi))\}. \tag{3-16}$$

The first term of the penalized optimal criterion in (3-16) represents an "alphabetic" optimality criterion for example, the D-optimality criterion, $\Psi(\xi) = -\log\det\{\boldsymbol{M}(\xi)\}$, that is a monotone and convex function (Fedorov & Leonov 2013, p. 61). The second term of the penalized optimal criterion in (3-16) is a bounded function between 0 and Λ (Harrington 1965), which is a penalty function representing constraints applied to the D-optimal designs, where Λ is a user-specified positive parameter that controls the balance between the overall desirability (or penalty) and optimality.

Minimization of the criterion (3-16) is similar to the constrained optimization of a scalar-valued objective function of several variables using Lagrange multipliers (Gavin & Scruggs 2012), where the objective function is modified by adding terms that describe the constraints. Thus, the search for the penalized D-optimal design can be considered as the minimization of the D-optimality criterion with respect to $\xi \in \Xi$ subject to the restrictions through an overall desirability function, $D(\xi)$, for a given value of Λ.

In general, the penalized optimal criterion in (3-16) is not a convex function of $\xi \in \Xi(\mathscr{X})$. However, this criterion is quasiconvex, i.e.,

$$\Psi_P[\alpha\xi_1 + (1-\alpha)\xi_2] \leq \max\{\Psi_P(\xi_1), \Psi_P(\xi_2)\},\ 0 \leq \alpha \leq 1; \tag{3-17}$$

see Fedorov & Leonov (2013, p. 113). For quasiconvex functions, most results used in convex optimization stay valid (Avriel 2003, ch. 6.1). This fact determines the possibility of finding the global minimum of the penalized optimal criterion (3-15).

3.4. Choice of Desirability Functions for Optimal Designs

The criterion of penalized optimality is based on desirability functions that summarize the preferences of a researcher in the experimental design. The proposed technique may include any design feature, that may be adjusted by defining some suitable desirability functions. We discussed the choice of appropriate desirability functions involved in obtaining penalized optimal designs with desirable characteristics in detail in the paper Rudnykh & López-Ríos (2018).

Initially, the non-penalized optimal design is analyzed. The amount of individual desirability functions is determined by the user, who chooses the number of required restrictions on the design. The design

$$\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{array} \right\}, \tag{3-18}$$

is characterized by n points x_i called support points of design and their respective weights w_i. Therefore, the restrictions can be applied to support points and their weights.

One of the common undesirable inconveniences of the optimal design is related to the range of experimentation; for example, the maximum support point is outside the region of experimental interest. The search region for the support points of design can be restricted by using suitable desirability functions described in Section 3.1. If large values of x are undesirable, a decreasing function is chosen. For example, the desirability functions defined in (3-6) and (3-14) can be selected as possible candidates for this purpose. If small values of x are undesirable, an increasing function is chosen and the desirability functions defined in (3-3), (3-5) and (3-11) are appropriate in this case. If both small and large values of x are not desirable, then the desirability function has a bell shape (not necessarily symmetric). The "target" desirability functions defined in (3-2), (3-7) and (3-9) reflect this practical concern. The user-specified maximum and minimum values of x help to define the explicit forms of the respective desirability functions. It is recommended to plot the resulting desirability function to check levels of x.

Another undesirable inconvenience of optimal design is related to the support points spacing. Accuracy constraints in the experimental method imply the minimum acceptable spacing between adjacent design points. In this case, the small values of this spacing are undesirable; therefore an increasing desirability function is chosen. The user-specified maximum and minimum values of the acceptable spacing between adjacent design points help to define the explicit form of this desirability function. A logistic function (3-11) can be used as a possible candidate to generate the required desirability functions (Parker &

Gennings 2008), but other functions can be used to obtain the appropriate form. This fact can be verified by plotting them.

Preferences for the number of observations per design point is related to the weights w_i associated with the support points. The number of observations at each point should be sufficient to estimate variability within them. A reasonable number of observations is defined for the design point of interest, taking into account the total sample size of the experiment. The "bigger-is-better" desirability function defined in (3-3), (3-5) and (3-11), or increasing function, reflects that the allocation of less than this reasonable number of observations to the design point is unacceptable. If the number of observations in the design point is restricted to a range of values, then the "target" desirability function defined in (3-2), (3-7) and (3-9) is used in this case. The exact shape of desirability functions can be obtained by defining the user-specified minimum and maximum values of the number of observations at the given point.

The nonlinear model of study may contain one or more threshold parameters used for defining two, or more, model regions. Researchers can be interested in placing a certain number of design points in each region, or a certain percentage of the sample, in assessing the model fitting.

The desirability function for the percentage of design points in the interest region, for validation purposes, can be a "target" desirability function defined in (3-2), (3-7) and (3-9) and has a bell shape (not necessarily symmetric). Yeatts et al. (2012) used an asymmetric "target" desirability function to place certain dose groups below the interaction threshold. This desirability function can be constructed by multiplying a set of logistic functions to obtain the resulting desirability function $d_{i(\text{target})} = d_{i(\text{max})} \times d_{i(\text{min})}$ (Gibb 1998), where $d_{i(\text{max})}$ is a maximizing desirability function defined in (3-11), and $d_{i(\text{min})}$ is a minimizing desirability function defined in (3-14). This allows the researcher to deal with asymmetric desirability functions. The appropriate numbers of design points above and below the interaction threshold are defined to evaluate the model fitting. The candidate design points are displaced through appropriate desirability functions. The user-specified maximum and minimum values of these design points help to define the explicit forms of the respective desirability functions.

The desirability function that defines the sample proportion below the threshold restricts the weights of the points in this region. To ensure the model fitting in this region, it is essential that the design allocates a sufficient amount of sample points below this threshold. This desirability function can be a "target" desirability function that limits a desirable number of observations above or below the interaction threshold (Yeatts et al. 2012). The permissible maximum and minimum weight values of the points in this

region help to define the explicit forms of the respective desirability functions.

In conclusion, the recommended steps to fulfill a specific design restriction are summarized as follows:

1. Initially, identify the characteristic that is subject to restriction according to the experimental design preferences.
2. Then, define the restriction according to the design preferences set by the researcher (maximize or minimize the characteristic, or target a range of its values).
3. Next, determine the range of acceptable values of this characteristic according to the design preferences.
4. To restrict this characteristic, select the appropriate desirability function among the functions described in Section 3.1.
5. Determine the explicit form of the appropriate desirability function according to the given restriction by replacing the permissible maximum and minimum values of this characteristic.
6. In order to verify the desirability levels of restricted characteristics of interest, it is recommended that the resulting desirability functions be plotted before they are used in the penalized optimal design strategy.

The penalized optimal design technique using desirability functions, as described in Section 3.2, is illustrated with the example of the nonlinear model proposed by Michaelis and Menten (Bates & Watts 1988, p. 33).

3.5. Example: Michaelis-Menten Model

The Michaelis-Menten model is one of the most widely used models in pharmacology, biology, and medical research. It is commonly used to describe saturation functions for numerous physical and biological phenomena.

The Michaelis-Menten model for enzyme kinetics relates the initial "velocity" of an enzymatic reaction to the substrate concentration x through the equation (Bates & Watts 1988, p. 33)

$$\eta(x;\theta) = \frac{\theta_1 x}{\theta_2 + x}, \quad x \geq 0,\ \theta_1 > 0,\ \theta_2 > 0, \tag{3-19}$$

where η is the predicted reaction rate, θ_1 is the ultimate velocity parameter and θ_2 is the half-velocity parameter. Nonlinear least-squares provide the following estimation of the

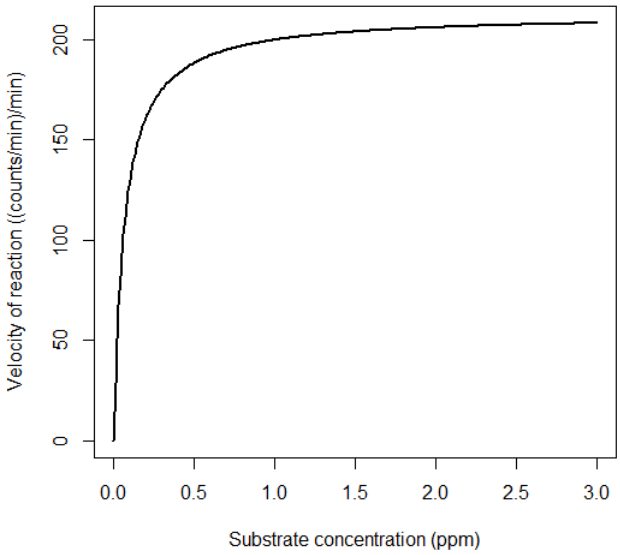

Figure 3-5.: Plot of the Michaelis-Menten model given in (3-19), where $\hat{\theta}_1 = 212.68$ and $\hat{\theta}_2 = 0.064$.

parameters of the Michaelis Menten model: $\hat{\theta}_1 = 212.68$ and $\hat{\theta}_2 = 0.064$. The expected Michaelis-Menten model shape is presented in Figure **3-5** .

The problem of designing experiments for the Michaelis-Menten model has studied by many researchers (see, e.g., López-Fidalgo & Wong (2002); Dette & Biedermann (2003); Trandafir & López-Fidalgo (2004); Dette, Kiss & Wong (2008); López-Fidalgo, Tommasi & Trandafir (2008); Dette & Kunert (2014)). It was demonstrated that the D-optimal design for this model is supported on two points and has equal weights at these points.

Iterative algorithms described in Section 2.4, are used to determine the D-optimal design for the Michaelis-Menten model given in (3-19), that minimize the D-optimality criterion. The resulting locally D-optimal design using the estimations of the unknown parameters is obtained as

$$\xi_D^* = \left\{ \begin{array}{cc} 0.064 & 2.056 \cdot 10^{13} \\ 0.5 & 0.5 \end{array} \right\}. \tag{3-20}$$

This design was obtained without taking into account an upper bound of x; therefore the maximum support point has a very large value according to the required computational precision. Thus, this point provides little information about the model curve (see Figure **3-5**) and does not aid in the detection of a concentration-velocity in the low concentration region. In practice, an upper bound is usually placed on the substrate

concentration (López-Fidalgo & Wong 2002, Dette et al. 2008), although theoretically, x is positive and unbounded. As an alternative approach to truncation, we use a desirability function technique to generate experimental designs with good statistical inference properties and practical characteristics defined by a researcher.

Given impractical features of the D-optimal design (3-20), a researcher wants to obtain a more practical design, such as one with the following features: three support points, where the first support point $x_1 = 0$ with 50% of observations with total sample size 20, the maximum design point within the region of experimental interest, and the minimum difference between adjacent design points at least 0.1 concentration units apart.

Initially, the control point (third point) is added, and the following augmentation D-optimal design is obtained:

$$\xi_A^* = \left\{ \begin{array}{ccc} 0 & 0.064 & 8.40 \cdot 10^8 \\ 0.0050 & 0.4975 & 0.4975 \end{array} \right\}. \tag{3-21}$$

The control point of this augmentation D-optimal design has a very small weight. A penalized-optimal design strategy is developed using three desirability functions. The function d_1 determines the minimum number of observations in the control point, the function d_2 defines the placement of the maximum design point within the research region, and the function d_3 characterizes the minimum difference between adjacent design points.

Considering the total size of the sample of 20 observations with a three-point design, it is reasonable to assume that at least six observations should be assigned to the control group, i.e., $r_{1,\min} = 6$ and $r_{1,\max} = 10$. Substituting these values of r_1 in (3-5), the explicit form of the "bigger-is-better" desirability function is obtained:

$$d_1(r_1) = \left\{ \begin{array}{cc} 0 & r_1 \leq 6 \\ \left(\frac{r_1 - 6}{4}\right)^4 & 6 < r_1 \leq 10 \\ 1 & r_1 > 10. \end{array} \right. \tag{3-22}$$

A large value of power $s_1 = 4$ indicates that it is very desirable for the value of r_1 to increase rapidly above $r_{1,\min} = 6$. A plot of this desirability function, shown in Figure **3-6**(a), confirms that assigning less than six observations to the control point is not acceptable.

Figure **3-5** shows that the concentration-velocity curve plateaus around 3; therefore, the maximum design point provides little information about this curve. Thus, it is reasonable that a researcher would consider a maximum concentration greater than 3 unacceptable.

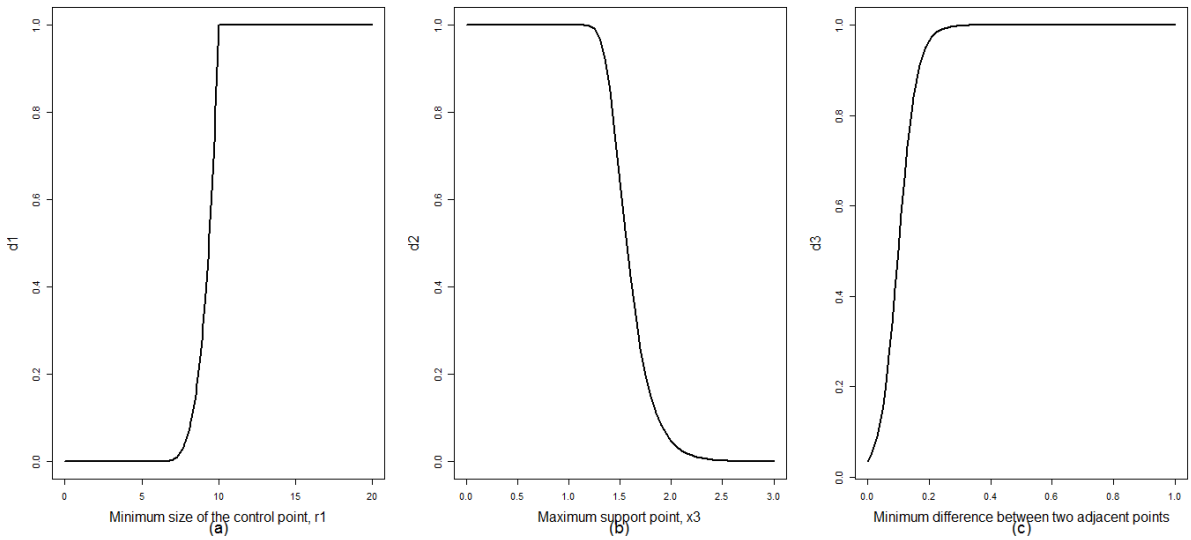

Figure 3-6.: Plots of desirability functions: (a) $d_1(r_1)$, (b) $d_2(x_3)$ and (c) $d_3(\text{diff})$.

The desirability function of the Gompertz form (3-3), described by Harrington (1965), reflects this practical interest:

$$d_2(x_3) = 1 - \exp(-\exp(9.25 - 6.15x_3)), \tag{3-23}$$

where $a_2 = -9.25$ and $b_2 = 6.15$ are specified such that $d_2(3) = 0.0001 \cong 0$ and $d_2(1.5) = 1 - 1/e = 0.63212$. A plot of this desirability function, shown in Figure **3-6**(b), exhibits that the maximum design point, x_3, is considered undesirable as it approaches 3.

In order to keep the appropriate concentration points spaced from a practical perspective, it is assumed that the adjacent concentration points would be optimally separated by at least 0.1 units of concentration. The function of desirability based on the logistic function (3-11), the smooth "smaller-is-better" desirability function proposed by Gibb (1998)

$$d_3(\text{diff}) = \frac{1}{1 + \exp(-(\text{diff} - a_3)/b_3)}, \tag{3-24}$$

where the parameter a_3 is an average of the upper ($\text{diff}_{\max}$) and lower ($\text{diff}_{\min}$) bounds of the response level being targeted, b_3 controls the function spread, and γ_3 is such that $d_3(\text{diff}_{\min}) = \gamma_3$ and $d_3(\text{diff}_{\max}) = 1 - \gamma_3$.

If $\gamma_3 = 0.05$ and $\text{diff}_{\min} = 0.01$, $\text{diff}_{\max} = 0.19$, then the desirability function $d_3(\text{diff})$ has the form (3-24). A plot of this desirability function, shown in Figure **3-6** (c), shows that the minimum difference between two adjacent design points is considered undesirable when approaching a 0.1 units of concentration.

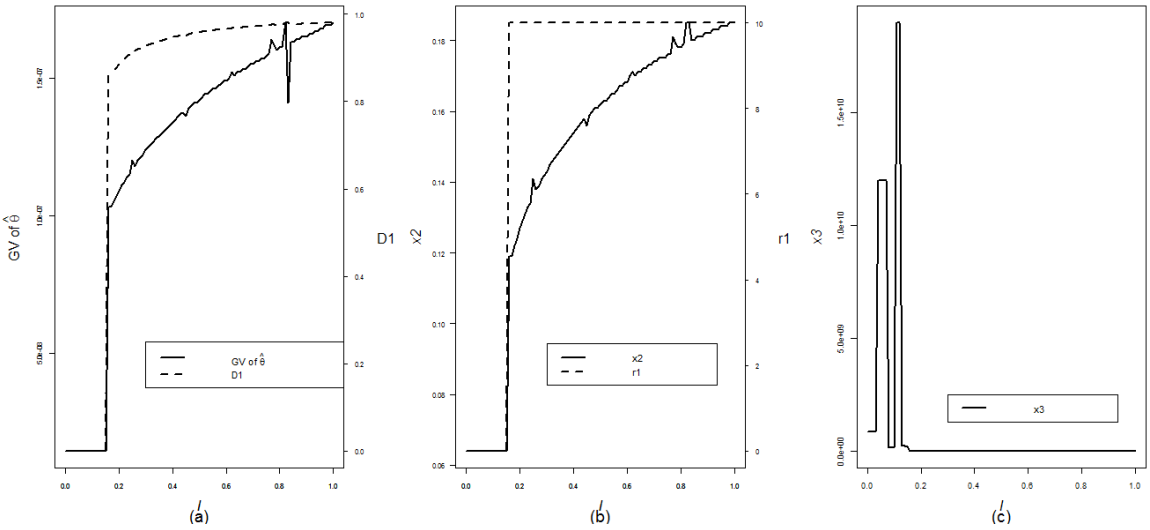

Figure 3-7.: Plots of penalized optimal design responses from the lth submultiple.

The overall desirability function of the form given in (3-4), is defined by

$$D_1(\xi) = (d_1 \times d_2 \times d_3)^{\frac{1}{3}} . \tag{3-25}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is used in R-project (Team 2018) to minimize the penalized D-optimal criterion

$$-\log\{\det \mathbf{M}(\xi)\} + l\Lambda_0 (1 - D_1(\xi)) \tag{3-26}$$

for the Michaelis-Menten model given in (3-19), where Λ_0 defines an absolute value of the minimum non-penalized D-optimal criterion. Computer code implemented in R-project (Team 2018) to generate penalized D-optimal designs may be found in Appendix B.

Below are the steps to follow to obtain the penalized D-optimal design:

1. Λ_0 is determined by minimizing the penalized D-optimal criterion (3-26) for $l = 0$. In this case $\Lambda_0 = |\min\{-\log\{\det \mathbf{M}(\xi)\}\}| = 12.04841$.

2. Penalized D-optimal designs are generated by minimizing the penalized D-optimal criterion (3-26) for different values $l = 0.01, 0.02, \ldots, 1$.

3. Penalized optimal design responses from the lth submultiple are plotted to observe their behavior. Figure **3-7** graphically presents the responses of the generalized variance (GV) of $\hat{\boldsymbol{\theta}}$, D_1, x_2, r_1 and x_3 of the penalized optimal designs from the lth submultiple.

Table 3-2.: Experimental designs for the Michaelis-Menten model given in (3-19) and generalized variance.

Design type	x_1 (r_1)	x_2 (r_2)	x_3 (r_3)	GV of $\hat{\boldsymbol{\theta}}$
D-optimal	0	0.064	$2.056 \cdot 10^{13}$	$1.45 \cdot 10^{-8}$
	(0)	(10)	(10)	
Equal spaced	0	0.60	1.20	$2.35 \cdot 10^{-6}$
	(10)	(5)	(5)	
Penalized optimal	0	0.127	1.278	$1.09 \cdot 10^{-7}$
	(10)	(5)	(5)	

4. The final value of l is selected within the range corresponding to the stability exposed in the responses of the overall desirability function. Figure **3-7**(a) shows that the desirability function responses become stable in the range of $l = 0.20$. Furthermore, when $l = 0.20$, the practical design characteristics improve with an acceptable increase in the generalized variance of $\hat{\boldsymbol{\theta}}$.

It is important to note that it is not essential to know the exact final value of l, as penalized optimal designs are similar within the stability range of overall desirability function responses.

A penalized optimal design from this range is

$$\xi^*_{\text{penal}} = \left\{ \begin{array}{ccc} 0.0 & 0.127 & 1.278 \\ 10 & 5 & 5 \end{array} \right\}. \tag{3-27}$$

This penalized optimal design exhibits the desirable practical properties as three different design points separated by at least 0.1 concentration units, and the first support point contains 50% of all observations, i.e., ten observations.

An alternative way to include desired constraints might be to limit the experimental region to $[0, 1.2]$ and place the three design points separated by the same distance. Thus, the design of the equally spaced points is obtained.

Table **3-2** presents the three different experimental designs for the Michaelis-Menten model, defined in (3-19) and their respective variances of the estimated parameters. It is noted that the smallest variance has the D-optimal design. Although the penalized design has a small increase in the variance of the estimated parameters as compared to D-optimal design, it has the practical characteristics desired by the researcher and

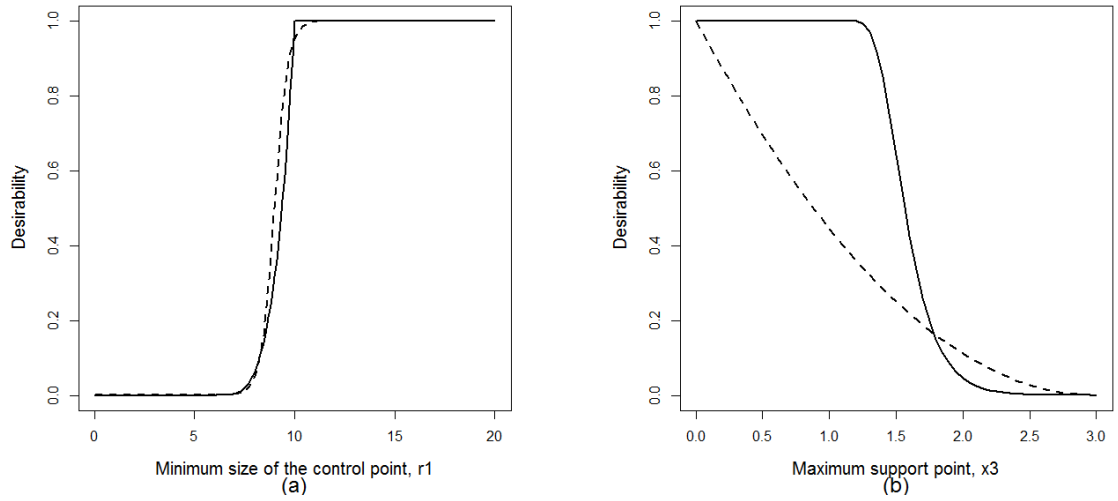

Figure 3-8.: Comparison of the functions of desirability: (a) $d_1(r_1)$ and $d_4(r_1)$; (b) $d_2(x_3)$ and $d_5(x_3)$.

represents an improvement in terms of the generalized variance of $\hat{\boldsymbol{\theta}}$ over the equally spaced design, which is arbitrarily chosen.

The penalized optimal design (3-27) was generated using the individual desirability functions d_1, d_2 and d_3, defined in (3-22), (3-23) and (3-24) respectively. However, these desirability functions are not unique to reflect similar desirability. Sensitivity to select appropriate desirability functions has been tested by considering other functions of different shapes but with the same desirability requirements.

The logistic function, the form of the "bigger-is-better" or maximizing desirability function, defined by Gibb (1998) was used to restrict the minimum size of the control point. Replacing the upper and lower limits of r_1 in (3-11), (3-12) and (3-13), the explicit form of this desirability function is

$$d_4(r_1) = \frac{1}{1 + \exp\left(-\left(r_1 - 9\right)/0.34\right)}. \tag{3-28}$$

It expresses similar desirability of the function d_1 that restricts control group around 50% of all observations, see Figure **3-8**(a).

The "smaller-is-better" desirability function

$$d_5(x_3) = \begin{cases} \left(\frac{3-x_3}{3}\right)^2 & x_3 \leq 3 \\ 0 & x_3 > 3 \end{cases} \tag{3-29}$$

Table 3-3.: Penalized optimal designs using the Michaelis-Menten model given in (3-19) and generalized variance.

Desirability function	x_1 (r_1)	x_2 (r_2)	x_3 (r_3)	GV of $\hat{\boldsymbol{\theta}}$
D_1	0 (10)	0.127 (5)	1.278 (5)	$1.09{\cdot}10^{-7}$
D_2	0 (10)	0.126 (5)	1.279 (5)	$1.08{\cdot}10^{-7}$
D_3	0 (10)	0.110 (5)	0.978 (5)	$1.10{\cdot}10^{-7}$
D_4	0 (10)	0.134 (5)	0.777 (5)	$1.52{\cdot}10^{-7}$

expresses a similar desirability of the function d_2 for the maximum design point, x_3, where the maximum concentration greater than 3 is unacceptable, see Figure **3-8**(b).

Using the new individual desirability functions, d_4 and d_5, the three new overall desirability functions, $D_2(\xi) = (d_4 \times d_2 \times d_3)^{\frac{1}{3}}$, $D_3(\xi) = (d_1 \times d_5 \times d_3)^{\frac{1}{3}}$ and $D_4(\xi) = (d_4 \times d_2 \times d_5)^{\frac{1}{3}}$, are defined. Computer code implemented in R-project (Team 2018) using D_2, D_3 and D_4 to generate penalized D-optimal designs may be found in Appendix B. The optimal penalized designs generated by the desirability functions D_1, D_2, D_3 and D_4 are given in Table **3-3**. This table allows to verify the performance of these desirability functions comparing resulting penalized designs using D_1, D_2, D_3 and D_4.

It is important to note the remarkable similarity between these four penalized designs. It has been shown in this case that the penalized designs are similar for desirability functions with similar shapes. In all cases, the penalized design technique has provided an experimental design with the desirable characteristics established by the researcher. The resulting penalized optimal designs have had a reasonable increment in the generalized variance, but the design characteristics have been improved from a practical perspective, i.e., the control group has an adequate number of observations, and the maximum support point is placed within the experimental region. It has also been shown that the penalized optimal designs represent an improvement in terms of the generalized variance of the estimated parameters compared to an arbitrarily chosen experimental design.

The example of Michaelis-Menten model illustrates the proposed methodology that allows the researcher to find the penalized optimal design with desirable properties. This

example presents the procedure for selecting the appropriate desirability functions according to the constraints of the experimental design from a practical perspective. The practical characteristics are determined by a penalty function through an overall desirability function that is the geometric mean of the individual desirability functions. The penalty function is added to an "alphabetic" optimality criterion to penalize non-practical optimal designs. The recommended steps to follow to obtain the penalized optimal design are detailed. The resulting penalized optimal design is optimal according to the "alphabetic" design criteria and also has the desired practical characteristics.

4. Penalized Bayesian Optimal Designs for Linear Models

Optimal experimental designs for linear regression models continue to receive considerable attention in the statistical literature because these models are probably the most popular in applications. Linear models are useful in the biological, physical, and social sciences, as well as in business and engineering. Moreover, they are used in both stages of research, such as planning and analysis of the resulting data. Therefore, although this dissertation deals with optimal designs for nonlinear models, a chapter on penalized Bayesian optimal designs for linear regression models may be useful to better understand some features of these designs for nonlinear models better.

This chapter presents the construction methodology of penalized Bayesian D-optimal designs for linear regression models. The Bayesian approach to experimental design enables to incorporate available prior information of unknown parameters to the experimentation, and the penalized optimal design strategy based on the use of desirability functions helps to obtain optimal designs that fulfill practical design preferences imposed by a researcher. The methodology presented here combines the use of desirability functions and the Bayesian approach in the construction of optimal designs obtained through utility functions associated with a Bayesian optimality criterion (Bayesian D-optimality) for linear regression models.

The organization of this chapter is as follows. The linear model is introduced in Section 4.1. A summary of the Bayesian D-optimality criterion for linear models is provided in Section 4.2. This summary is based on the work of Chaloner & Verdinelli (1995). The Bayesian optimal design methodology for linear models is illustrated by an example of the quadratic regression model in Section 4.3. In Subsection 4.3.2, the resulting Bayesian designs are examined for different structures of the covariance matrix associated with the parameter prior distribution under consideration.

Section 4.4 presents the new criterion proposed in this dissertation to obtain penalized Bayesian D-optimal designs for linear regression models. The penalized Bayesian optimal design methodology for linear models is illustrated by an example of the quadratic regression model in Section 4.5. The penalized Bayesian D-optimal designs are obtained by

using appropriate desirability functions, described in Section 3.1. Furthermore, the mean square errors (MSE) are calculated via simulation in order to compare non-penalized and penalized Bayesian optimal designs. The MSE results and discussion are presented in Section 4.7. The procedure of the efficient design apportionment described in Pukelsheim (1993, p. 309) is presented in Section 4.6.

4.1. Linear Model

Consider the linear regression model (El-Krunz & Studden 1991)

$$y = \eta(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\theta} + \varepsilon, \tag{4-1}$$

where

- y is a continuous response (or dependent or observed) variable.
- $\boldsymbol{x}$ is a $m \times 1$ vector of explanatory (or predictors or independent or control or design) variables. Suppose that the vector $\boldsymbol{x} \in \mathscr{X}$ called space or design region, where $\mathscr{X}$ is a compact set in $\mathbb{R}^m$.
- $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, where $\boldsymbol{\theta} \in \Theta$, here Θ is an open convex set in $\mathbb{R}^p$.
- $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is a known continuous in $\mathscr{X}$ linear function with respect to the model parameters $\boldsymbol{\theta}$.
- $\boldsymbol{f}(\boldsymbol{x})$ is a $p \times 1$ vector of known "basis" regression functions. They are p linearly independent real-valued continuous functions on the design space $\mathscr{X}$.
- ε is an uncorrelated normally distributed error term with zero mean and constant variance σ^2 and independent of $\boldsymbol{x}$.

It assumes that an experiment is performed at points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathscr{X}$ (some of these points can be repeated). An independent observation y_i will be observed at each point of $\boldsymbol{x}_i$, for $i = 1, 2, \ldots, N$. In this case, the linear model can be written in the matrix form (Ermakov & Zhiglijavsky 1987, p. 84):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{4-2}$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_N)^T$ is the vector of N observations, $\boldsymbol{X} = (\boldsymbol{f}(\boldsymbol{x}_1), \boldsymbol{f}(\boldsymbol{x}_2), \ldots, \boldsymbol{f}(\boldsymbol{x}_N))^T$ is the $N \times p$ extended design matrix and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T$ is the vector of the errors. Under the above assumptions

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\theta} \quad \text{and} \tag{4-3}$$

$$\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_N, \tag{4-4}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

4.2. Bayesian D-Optimality Criterion for Linear Models

The idea of the Bayesian optimal design is to incorporate available prior knowledge about a studied phenomenon into the design process. The Bayesian design approach uses a prior distribution of unknown parameters rather than guessed single values. Now, from a Bayesian perspective $\boldsymbol{\theta}$ is taken as a random vector rather than a fixed vector.

Suppose that a prior distribution $\pi(\boldsymbol{\theta}, \sigma^2)$ on $\boldsymbol{\theta}$, σ^2 is given such that the conditional prior distribution $\pi(\boldsymbol{\theta} \mid \sigma^2)$ of $\boldsymbol{\theta}$ given σ^2 is $\mathcal{N}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{R}^{-1})$, where $\boldsymbol{R}$ is a given positive definite $p \times p$ "precision" matrix.

Under the above assumptions the posterior conditional distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2)$ of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$, σ^2 is normal with mean vector (Chaloner & Verdinelli 1995)

$$\widehat{\boldsymbol{\theta}}_B = E(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2) = (\boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{R})^{-1} (\boldsymbol{X}^T \boldsymbol{Y} + \boldsymbol{R}\boldsymbol{\mu}) \tag{4-5}$$

and covariance matrix $\sigma^2 (\boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{R})^{-1}$ (see Appendix B).

According to Chaloner & Verdinelli (1995) the Bayesian solution to the experimental design problem is to find the best design and the best decision rule that maximizes expected utility $U(\xi)$ in (2-32). More details are found in Section 2.3.2.

When the expected gain in Shannon information (2-35) is used as a utility function in the normal linear regression model, this Bayesian solution is reduced to maximizing the function

$$\Phi_B(\xi) = \det \left\{ \boldsymbol{M}(\xi) + \frac{1}{N} \boldsymbol{R} \right\} = \det \boldsymbol{M}_B(\xi), \tag{4-6}$$

and it is known as Bayesian D-optimality criterion for linear models.

Design ξ_B^* that satisfies the following condition (DasGupta & Studden 1991)

$$\xi_B^* = \arg\min_{\xi \in \Xi} \left\{ -\log\det \left\{ \boldsymbol{M}(\xi) + \frac{1}{N} \boldsymbol{R} \right\} \right\} = \arg\min_{\xi \in \Xi} \{ -\log\det \boldsymbol{M}_B(\xi) \} \tag{4-7}$$

is the Bayesian D-optimal design for linear regression models.

A characteristic of optimal Bayesian design is the dependence on the sample size N. If N is large, $\boldsymbol{M}_B(\xi)$ is approximately equal to $\boldsymbol{M}(\xi)$, i.e., any differences between a Bayesian design and its corresponding non-Bayesian one are not important. In this case, the information provided by the data overwhelms the initial information contained in the prior distribution. In contrast, if N is small, the prior information has an important effect on the posterior distribution and determines the design.

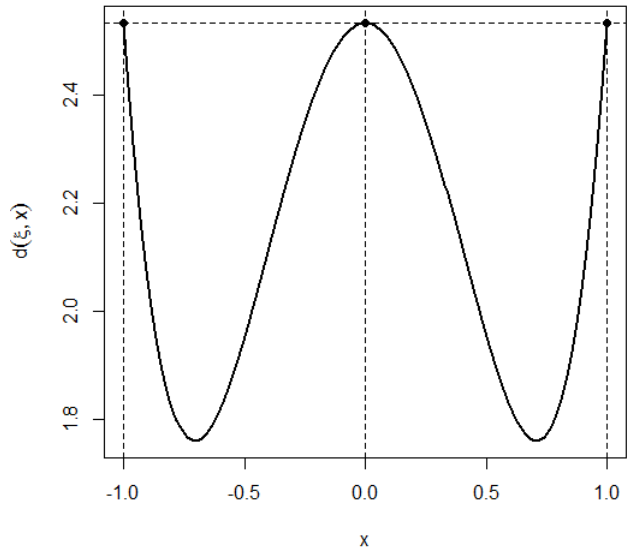

Figure 4-1.: Example: quadratic model (4-8). Bayesian D-optimality verification for the three-point Bayesian optimal design (4-9).

The methodology of the construction of the Bayesian D-optimal design for linear models is illustrated with an example of the quadratic regression model.

4.3. Example: Quadratic Model

Consider the quadratic regression model (DasGupta & Studden 1991)

$$E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 \tag{4-8}$$

and suppose $-1 \leq x \leq 1$, where the parameters θ_0, θ_1 and θ_2 are considered as random variables from the Bayesian perspective.

4.3.1. Bayesian D-Optimal Design

If the prior variances of θ_0, θ_1, θ_2 are 3, 5 and 1 with absent correlation, and if sample size $N = 9$, then the Bayesian D-optimal design is

$$\xi_B^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.369 & 0.261 & 0.369 \end{array} \right\}. \tag{4-9}$$

Figure **4-1** shows that the function $d(\xi, x) = \boldsymbol{f}(x)^T \boldsymbol{M}_B(\xi)^{-1} \boldsymbol{f}(x)$ achieves its maximum value tr $\boldsymbol{M}(\xi^*)\boldsymbol{M}_B(\xi^*) = 2.5335$ at the design points $x = -1$, 0 and 1, demonstrating

the D-optimality of the design ξ_B^* according to the equivalence theorem for Bayesian D-optimal designs for linear models (Pilz 1991, p. 140).

4.3.2. Optimal Designs for Different Structures of the Prior Covariance Matrix

A characteristic of Bayesian D-optimal design for linear models is the dependence of the structures of the prior covariance matrix, which can be observed in the example of the quadratic regression.

First, some combinations of the values of the parameter correlation coefficients do not generate a positive definite "precision" matrix $\boldsymbol{R}$. For example, the combination of the negative coefficients with absolute values greater than 0.5 generate the negative definite matrix $\boldsymbol{R}$.

Second, the majority of resulting Bayesian D-optimal designs have three support points, where the extremes -1 and 1 are preserved for all designs, while the midpoint of the design support x_2 oscillates around zero. If $\rho_{23} = 0$, then the midpoint of the design support x_2 is zero or very close to zero. $x_2 > 0$ if $\rho_{12} > 0$ and $x_2 < 0$ if $\rho_{12} < 0$. In this case the weights w_i tend to be uniform. If the absolute value of ρ_{23} is large, then the weight w_2 of the midpoint x_2 is small. The midpoint $x_2 > 0$ if $\rho_{23} < 0$ and $x_2 < 0$ if $\rho_{23} > 0$.

Third, when the correlation coefficient $\rho_{13} \geq 0.7$, and two other coefficients are approximately equal, the Bayesian D-optimal design has only two support points and places a half of the design runs at -1 and the other half at 1. For example, if the correlation coefficients $\rho_{13} = 0.9$ and $\rho_{12} = \rho_{23} = 0.5$ the resulting Bayesian D-optimal design is

$$\xi_B^* = \left\{ \begin{array}{cc} -1 & 1 \\ 0.5 & 0.5 \end{array} \right\}. \tag{4-10}$$

This is a surprising result since the optimal design to estimate a second-order model with three parameters contains only two design points (Hamada et al. 2001).

Figure **4-2** illustrates the Bayesian D-optimality verification of the two-point design ξ_B^* according to the equivalence theorem for Bayesian D-optimal designs for linear models (Pilz 1991, p. 140). The function $d(\xi, x)$ only achieves its maximum value at two support points $x = \pm 1$.

Finally, with absent correlation the Bayesian D-optimal design contains three design points, i.e., $x = \pm 1$ and $x = 0$. The weights of the support points depend on the prior

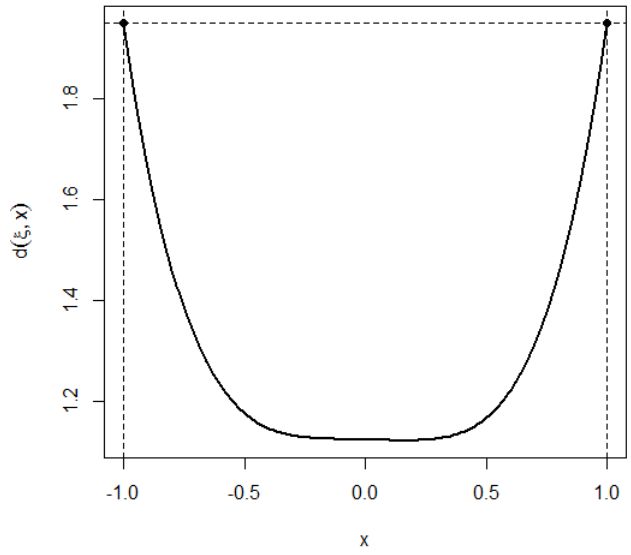

Figure 4-2.: Example: quadratic model (4-8). Bayesian D-optimality verification for the two-point Bayesian optimal design (4-10).

variances of θ_0, θ_1, θ_2. When the variances increase, the weights w_i tend to be uniform, i.e., Bayesian optimal designs approximate the corresponding non-Bayesian ones. When the variances decrease, the weights of the extremes -1 and 1 tend to 0.5, and the weight of the midpoint tends to zero, i.e., the optimal Bayesian designs approximate the two-point design (4-10).

4.4. New Criterion

Experimental designs constructed with optimal design theory may not be appropriate from a practical perspective. They can cause conflicts with the usual laboratory practice and established guidelines. Parker (2005) proposed a technique described in Section 3.2 that penalizes optimal designs through the use of desirability functions. The resulting penalized optimal design complies with the optimal design criterion and the practical design preferences.

We suggest extending the use of desirability functions, described in Section 3.1, in Bayesian optimal designs. Thus, the researcher can incorporate prior information of the unknown parameters by using a Bayesian approach and also satisfy practical design preferences imposed by a researcher.

We propose to find a penalized Bayesian D-optimal design for linear models by minimizing the new criterion:

$$\Psi_{BP}(\xi) = \Psi_B(\xi) + \Lambda(1 - D(\xi)) = -\log\{\det \boldsymbol{M}_B(\xi)\} + \Lambda(1 - D(\xi)) \tag{4-11}$$

with respect to $\xi \in \Xi$ for a given value of Λ, where Λ is a user-specified scale constant.

The first term of the penalized Bayesian optimal criterion in (4-11) represents a Bayesian "alphabetic" optimality criterion for example, the Bayesian D-optimality criterion for linear models, that is a monotone and convex function (DasGupta & Studden 1991). The second term of the penalized Bayesian optimal criterion in (4-11) is a bounded function between 0 and $\Lambda > 0$ (Harrington 1965), which is a penalty function representing constraints applied to the Bayesian D-optimal designs. The minimization of the criterion (4-11) is considered as the maximization of the expected utility (2-37), restricted by the penalty function.

The penalty function, $(1 - D(\xi))$, is added to the Bayesian D-optimality criterion for linear models, $-\log\{\det \boldsymbol{M}_B(\xi)\}$, to penalize experimental designs. A user-specified constant, Λ, is required to place these two terms on similar scales and manage the weight of the penalty, since the minimum value of the Bayesian D-optimality criterion is numerically small or negative, while the penalty function has a range of 0 to 1.

Minimization of the criterion (4-11) is similar to the constrained optimization of a scalar-valued objective function of several variables using Lagrange multipliers (Gavin & Scruggs 2012), where the objective function is modified by adding terms that describe the constraints. Thus, the search for the penalized Bayesian D-optimal design for linear models can be seen as minimizing the Bayesian D-optimality criterion for linear models with respect to $\xi \in \Xi$ subject to the restrictions through an overall desirability function, $D(\xi)$, for a given value of $\Lambda > 0$.

The penalized Bayesian D-optimal criterion in (4-11) is not usually a convex function of $\xi \in \Xi(\mathscr{X})$. However, this criterion is quasiconvex function (Fedorov & Leonov 2013, p. 113). For quasiconvex functions, most results used in convex optimization stay valid (Avriel 2003, ch. 6.1). This fact determines the possibility of finding the global minimum of the penalized Bayesian optimal criterion (4-11).

Iterative algorithms described in Section 2.4 , are used to determine the optimal design ξ_P^* that minimize the new criterion (4-11), given a value of Λ. The initial value of Λ is chosen by $\Lambda_0 = |\min\{-\log\{\det \boldsymbol{M}_B(\xi)\}\}|$, i.e., an absolute value of the minimum of the Bayesian D-optimality criterion. Penalized Bayesian optimal designs are generated by minimizing the penalized Bayesian optimal criterion (4-11) for values of Λ in multiples

or submultiples of Λ_0. The final value of Λ is selected in the range where stability is exhibited in the responses of the overall desirability function. The resulting optimal design is determined by the x and w values of the penalized optimal design ξ_P^* associated with the minimum value of the penalized Bayesian optimal criterion for linear models (4-11) for a value of Λ in this range. The resulting penalized Bayesian optimal design is optimal according to the Bayesian D-optimal design criterion for linear models and the practical design preferences.

The methodology of the construction of the penalized Bayesian D-optimal design for linear models is illustrated with an example of the quadratic regression model.

4.5. Penalized Bayesian D-Optimal Design

4.5.1. The two-point design (4-10)

The Bayesian D-optimal design (4-10) contains only two design support points to estimate a second-order model, which contains three unknown parameters. Therefore, we want to have the three-point distinct D-optimal design with an acceptable number of observations in new design point placed between -1 and 1.

Initially, the third point is added, and the following Bayesian D-optimal design is obtained:

$$\xi_A^* = \left\{ \begin{array}{ccc} -1 & 1 & 1 \\ 0.494 & 0 & 0.506 \end{array} \right\}. \tag{4-12}$$

It is observed that this Bayesian D-optimal design is actually supported by two design points, not three.

A penalized-optimal design strategy is developed using two desirability functions d_1 and d_2. d_1 characterizes the minimum number of observations in new design point x_2 and d_2 depicts the location of this point. The logistic cumulative distribution function, the form of the "bigger-is-better" or maximizing desirability function given in (3-11) is used to create the desirability function d_1. The normal density function, the bell shape or "target" desirability function given in (3-9) is used to create the desirability function d_2.

Considering the total sample size of 9 of for a three-point design, it is reasonable to assume that at least two observations should be assigned to the new design point. Replacing the values of $\gamma_1 = 0.05$ and $r_{2(\min)} = 2$, $r_{2(\max)} = 3$ in the formulas (3-11), (3-12) and (3-13), the desirability function $d_1(r_2)$ is obtained:

$$d_1(r_2) = \frac{1}{1 + \exp\left(-\left(r_2 - 2.5\right)/0.17\right)}. \tag{4-13}$$

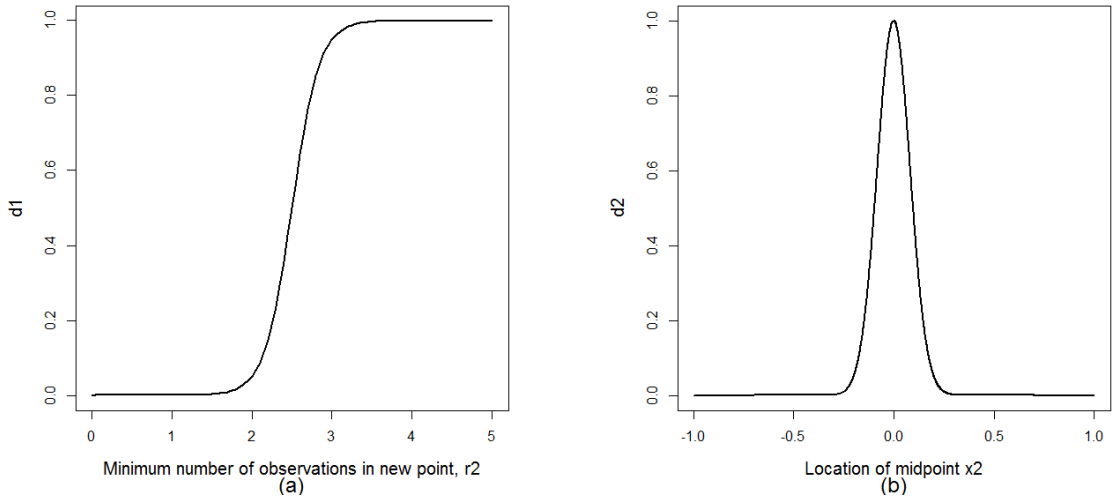

Figure 4-3.: Plots of desirability functions for the two-point design (4-10).

A plot of this desirability function, given in Figure **4-3**(a), shows that it is not acceptable to assign less than two observations to the new design point.

In order to keep the new design point approximately equidistant from the extremes of the design space, it is assumed that the target of value range of x_2 is 0 and the interval width of desirability function is 0.2. Replacing the values of $\gamma_2 = 0.05$, $a = 0$ and $\delta = 0.2$ in the formulas (3-9) and (3-10), the desirability function $d_2(x_2)$ is obtained:

$$d_2(x_2) = \exp\left\{-\frac{1}{2}\left(\frac{x_2}{0.082}\right)^2\right\}. \tag{4-14}$$

A plot of this desirability function, given in Figure **4-3**(b), shows that the location of the new design point x_2 outside the interval $[-0.2, 0.2]$ is unacceptable.

The overall desirability function is

$$D_{2p}(\xi) = (d_1 \times d_2)^{1/2}. \tag{4-15}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is used in R-project (Team 2018) to minimize the penalized Bayesian D-optimal criterion

$$-\log\{\det \boldsymbol{M}_B(\xi)\} + l\Lambda_0(1 - D_{2p}(\xi)) \tag{4-16}$$

for the quadratic model given in (4-8), where Λ_0 defines an absolute value of the minimum non-penalized Bayesian D-optimal criterion of the two-point design (4-10). The minimization of the criterion (4-16) is considered as the maximization of the expected

utility (2-37), restricted by the penalty function. Computer code implemented in R-project (Team 2018) to generate penalized Bayesian D-optimal designs may be found in Appendix B. The steps to follow to obtain the penalized D-optimal design are described in detail in Section 3.5.

Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (4-16) for different values $l = 0.01, 0.02, \ldots, 1$, where the user-specified constant $\Lambda_0 = |\min\{-\log\{\det \boldsymbol{M}_B(\xi)\}\}| = 0.3794315$ represents the absolute minimum value of the non-penalized Bayesian D-optimal criterion.

Initially penalized Bayesian D-optimal design is only supported by two extreme design points. The third distinct design point appears when $l = 0.72$. Figure **4-4** graphically displays the responses GV of $\hat{\boldsymbol{\theta}}$ and $D_{2p}(\xi)$ of the penalized Bayesian optimal designs from the lth submultiple. It shows the balance between the penalty imposed for Bayesian designs without the desired properties and the generalized variance. Figure **4-4** indicates stability in the responses of the overall desirability function around $l = 0.75$. The penalized Bayesian D-optimal design from this range is

$$\xi_{BP}^{*} = \left\{ \begin{array}{ccc} -1.000 & 0.000 & 1.000 \\ 0.354 & 0.278 & 0.368 \\ 3 & 3 & 3 \end{array} \right\}, \tag{4-17}$$

where the last row represents the number of observations at each design point for the sample size $N = 9$. The resulting design includes three distinct design points, and the new point placed between -1 and 1 has three observations. In summary, the properties of this penalized design are improved from a practical perspective, while the generalized variance is acceptably increased.

4.5.2. The three-point design (4-9)

Consider the Bayesian D-optimal design (4-9) for quadratic regression model (4-8). This design contains three design support points. We want to have the four-point distinct D-optimal design with the minimum two observations in new point placed between 0 and 1. Besides, the minimum difference between adjacent support points should be 0.3 units apart.

Initially, the fourth point is added to the initial design (4-9), and the following Bayesian D-optimal design is obtained:

$$\xi_{A}^{*} = \left\{ \begin{array}{cccc} -1 & 0 & 1 & 1 \\ 0.369 & 0.261 & 0.369 & 0.000 \end{array} \right\}. \tag{4-18}$$

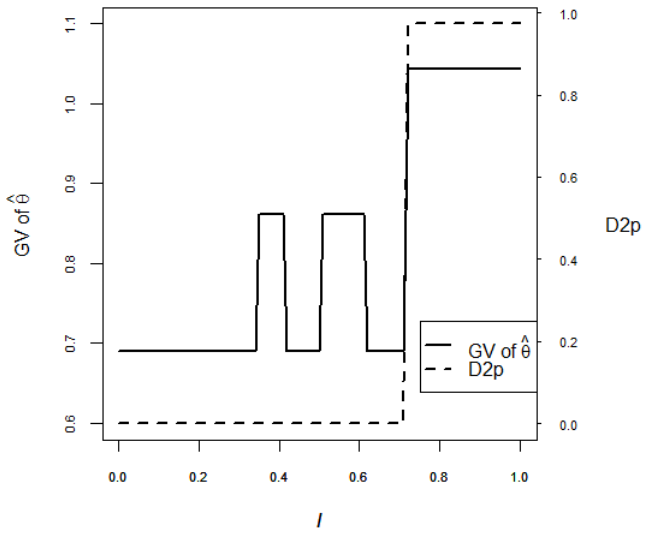

Figure 4-4.: Tradeoff between overall desirability D_{2p} and the generalized variance (GV) of $\hat{\boldsymbol{\theta}}$ from the lth submultiple using the criterion given in (4-16).

Clearly, this Bayesian D-optimal design is actually supported by three design points, not four.

A penalized-optimal design strategy is developed using three desirability functions to characterize the minimum number of observations in new point (d_1) and the minimum difference between adjacent points (d_2 and d_3). A logistic cumulative distribution function, of the type described by Gibb (1998), is used to create the desirability functions, though other functions can be used to achieve the appropriate shape. The logistic function, the form of the "bigger-is-better" or maximizing desirability function given in (3-11), captures the experimental design preferences.

Considering the total sample size of 9 for a four-point design, it is reasonable to assume that at least one observation should be assigned to the new design point. Replacing the values of $\gamma_1 = 0.05$ and $r_{3(\min)} = 1$, $r_{3(\max)} = 2$ in the formulas (3-11), (3-12), and (3-13), the desirability function $d_1(r_3)$ is obtained as:

$$d_1(r_3) = \frac{1}{1 + \exp\left(-\left(r_3 - 1.5\right)/0.17\right)}. \tag{4-19}$$

A plot of this desirability function, given in Figure **4-5**(a), it is not acceptable to allocate less than one observation to the new support point.

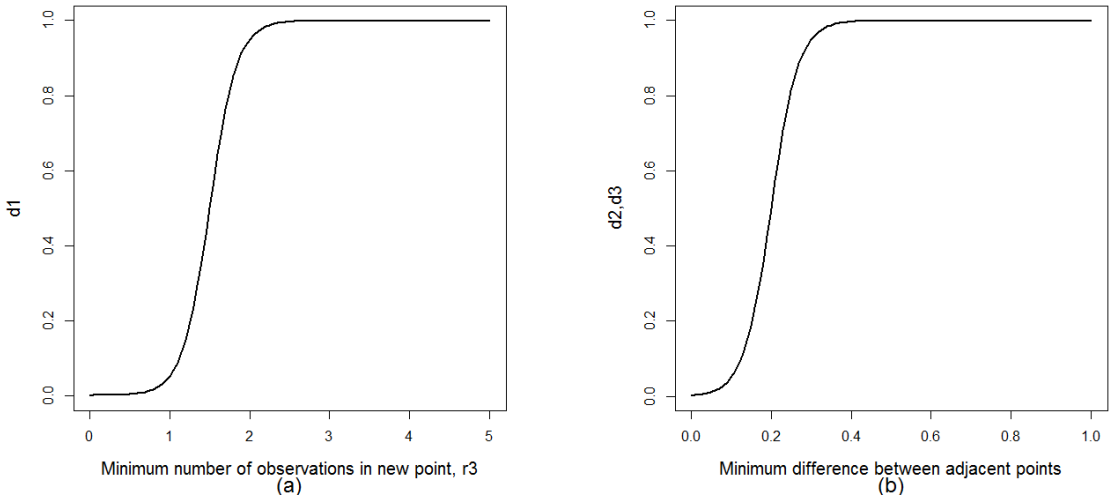

Figure 4-5.: Plots of desirability functions for the three-point design 4-9.

It is assumed that the adjacent support points should be optimally spaced at least 0.1 units apart in order to have four distinct support points. Replacing the values of $\gamma_{2,3} = 0.05$, $\text{diff}_{23(\min)} = \text{diff}_{34(\min)} = 0.1$, and $\text{diff}_{23(\max)} = \text{diff}_{34(\max)} = 0.3$ in the formulas (3-11), (3-12), and (3-13), the desirability functions $d_2\,(\text{diff}_{23})$ and $d_3\,(\text{diff}_{34})$ are obtained as:

$$d_2(\text{diff}_{23}) = d_3(\text{diff}_{34}) = \frac{1}{1 + \exp\left(-\left(\text{diff} - 0.2\right)/0.034\right)}. \tag{4-20}$$

A plot of these desirability functions, given in Figure **4-5**(b), shows that the spacing between support points of less than 0.1 units apart is unacceptable.

The overall desirability function is

$$D(\xi) = \left(d_1 \times d_2 \times d_3\right)^{1/3}. \tag{4-21}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is used in R-project (Team 2018) to minimize the penalized Bayesian optimal criterion

$$-\log\left\{\det \boldsymbol{M}_B\left(\xi\right)\right\} + l\Lambda_0\left(1 - D(\xi)\right) \tag{4-22}$$

for the quadratic model given in (4-8), where Λ_0 defines an absolute value of the minimum non-penalized Bayesian D-optimal criterion of the three-point design (4-9). The minimization of the criterion (4-22) is considered as the maximization of the expected utility (2-37), restricted by the penalty function. Computer code implemented in R-project (Team 2018) to generate penalized Bayesian D-optimal designs may be found

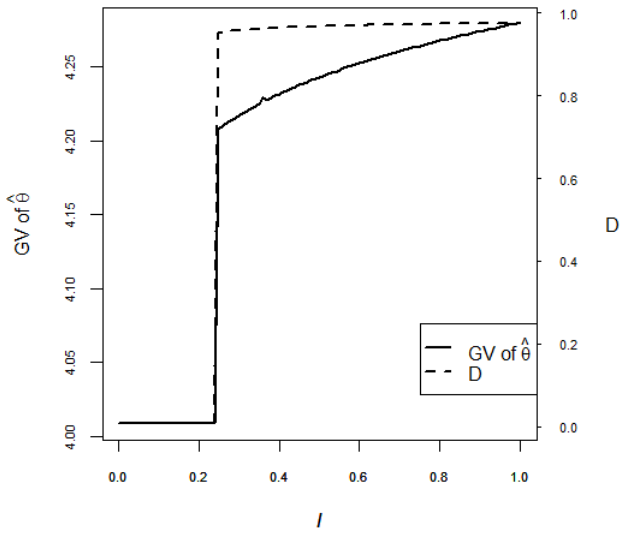

Figure 4-6.: Tradeoff between overall desirability $D(\xi)$ and the generalized variance (GV) of $\hat{\boldsymbol{\theta}}$ from the lth submultiple using the criterion given in (4-22).

in Appendix B. The procedure follows the steps, described in detail in Section 3.5, to obtain the penalized D-optimal design.

The minimum Bayesian D-optimal criterion defines $\Lambda_0 = |\min\{-\log\{\det \boldsymbol{M}_B(\xi)\}\}| = 1.365036$. Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (4-22) for different values $l = 0.01, 0.02, \ldots, 1$.

Initially penalized Bayesian D-optimal design is supported by only three design points. The fourth distinct design point appears when $l = 0.25$. Figure **4-6** graphically displays the responses GV of $\hat{\boldsymbol{\theta}}$ and $D(\xi)$ of the penalized Bayesian optimal designs from the lth submultiple. It exhibits the balance between the imposed penalty for Bayesian designs without the desired characteristics and the generalized variance. Figure **4-6** indicates stability in the responses of the overall desirability function around $l = 0.4$. A penalized Bayesian D-optimal design from this range is

$$\xi^*_{BP} = \left\{ \begin{array}{cccc} -1 & 0 & 0.295 & 1 \\ 0.367 & 0.115 & 0.167 & 0.351 \\ 3 & 1 & 2 & 3 \end{array} \right\}, \tag{4-23}$$

where the last row represents the number of observations at each design point for the sample size $N = 9$. The resulting design contains four distinct design points at least 0.3

units apart, and the new point placed between 0 and 1 has two observations.

Although this design has a small increase in the generalized variance of the estimated parameters (4.232) as compared to Bayesian D-optimal design (4.008), this penalized optimal design has the practical characteristics desired by the researcher: four distinct support points with an acceptable number of observations assigned to each support point.

4.6. Efficient Design Apportionment

Continuous designs are an approximate solution to the original problem of design from the perspective of a practical experiment. For the continuous design ξ, given in (2-12), to become realizable for a fixed sample size N, the weights w_i must be rounded to integers r_i which sum to N. The usual approach is to calculate the quota Nw_i, and then round Nw_i to the closest integer r_i. However, the numbers r_i so obtained may not sum to N.

This section presents another method of rounding the weights of a continuous design which performs much better. This procedure called efficient design apportionment has the smallest loss of efficiency due to discretization under a wide family of optimality criteria (Pukelsheim 1993, ch. 12).

The procedure of efficient design apportionment is essentially the following. According to Pukelsheim (1993, p. 309), its implementation has two phases.

First, the multiplier of the efficient design apportionment is defined as

$$\nu = N - \frac{1}{2}n, \tag{4-24}$$

where N and n are sample size and number of design support points, respectively. This multiplier is used to calculate the frequencies (number of observations at each support point)

$$r_i = \left\lceil \left(N - \frac{1}{2}n \right) w_i \right\rceil, \qquad i = 1, 2, \ldots, n, \tag{4-25}$$

where w_i are the weights of a continuous design and $\lceil \cdot \rceil$ denotes ceiling function that maps a real number to the smallest following integer.

Second, the discrepancy defined as

$$d = \sum_{i \leq n} r_i - N \tag{4-26}$$

is calculated. In case d is negative, a frequency r_j such that r_j/w_j attains $\min_{i \leq n} r_i/w_i$ ought to be augmented to $r_j + 1$. Similarly, in case d is positive, a frequency r_k for which

$(r_k - 1)/w_k$ attains $\max_{i \le n} (r_i - 1)/w_i$ ought to be reduced to $r_k - 1$. The procedure is repeated until $d = 0$.

The procedure of the efficient design apportionment is illustrated by the example of the quadratic model (4-8). The Bayesian D-optimal design with absent correlation for the sample size $N = 9$ (see (4-9)) has three support points: ± 1 and 0 with their respective weights 0.36946 and 0.26108. In this case, the multiplier of the efficient design apportionment is

$$\nu = N - \frac{1}{2}n = 9 - \frac{1}{2}3 = 7.5. \tag{4-27}$$

The respective frequencies are

$$r_1 = r_3 = \lceil 7.5 \cdot 0.36946 \rceil = \lceil 2.77095 \rceil = 3, \quad r_2 = \lceil 7.5 \cdot 0.26108 \rceil = \lceil 1.9581 \rceil = 2. \tag{4-28}$$

Thus the discrepancy is

$$d = 3 + 3 + 2 - 9 = -1 < 0. \tag{4-29}$$

Now r_i/w_i are calculated:

$$\frac{r_1}{w_1} = \frac{r_3}{w_3} = \frac{3}{0.36946} = 8.11996, \qquad \frac{r_2}{w_2} = \frac{2}{2.77095} = 7.66049. \tag{4-30}$$

The frequency r_2 attains the minimum value of r_i/w_i, where $i = 1, 2, 3$. Hence the frequency r_2 would increase to $r_2 + 1 = 2 + 1 = 3$. The discrepancy with new frequencies r_i is 0. Finally, the given Bayesian D-optimal design ξ_B^* rounded by the efficient design apportionment to a design $\xi_{B,N}^*$ for sample size $N = 9$ is

$$\xi_{B,N}^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 3 & 3 & 3 \end{array} \right\}, \tag{4-31}$$

i.e., three observations at each design point.

4.7. Mean Square Error as an Evaluation Measure for Optimal Designs

In statistics, the mean squared error (MSE) of an estimator evaluates the average of the squares of the errors and measures the quality of an estimator; it is always non-negative. If the MSEs are close to zero, then it implies closeness between the estimator and the parameter. Values of MSE may be used for comparative purposes, for example, to compare the performance of optimal designs. This comparison is illustrated with an example of the quadratic regression model (4-8). The MSE associated with the estimated parameters of the quadratic model (4-8) are calculated via simulation in order to compare non-penalized (4-30) and penalized (4-23) Bayesian optimal designs.

4.7.1. MSE of the estimated parameters of prior distribution

Let the conditional prior distribution of unknown parameters vector $\boldsymbol{\theta}^T = (\theta_0, \theta_1, \theta_2)$ given σ^2 is the multivariate normal distribution $\mathcal{N}_3\left(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}\right)$ with mean vector $\boldsymbol{\mu}^T = (1, 1, 1)$, variance-covariance matrix $\boldsymbol{\Sigma} = \text{diag}\,(3, 5, 1)$ and $\sigma^2 = 0.01$. For sample size of 9 observations, then exact designs are

$$D\text{-optimal:} \quad x_D^* = \{-1, -1, -1, 0, 0, 0, 1, 1, 1\}, \tag{4-32}$$

$$\text{Bayesian } D\text{-optimal:} \quad x_B^* = \{-1, -1, -1, 0, 0, 0, 1, 1, 1\}, \tag{4-33}$$

$$\text{Penalized Bayesian } D\text{-optimal:} \quad x_{BP}^* = \{-1, -1, -1, 0, 0.295, 0.295, 1, 1, 1\}. \tag{4-34}$$

The experimental observations of three designs are simulated according to the quadratic model (4-8) in the assigned design points (4-32), (4-33) and (4-34), respectively. Then the parameters are estimated from the simulated samples according to each design. The algorithm to fit linear models is available using R software through the command line as *lm*. It can also be used to estimate unknown parameters in linear models. Later the squared differences between estimated parameters and parameters generated according to the multivariate normal distribution $\mathcal{N}_3\left(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}\right)$ are calculated. Finally, the results of the MSE are obtained by averaging the squared differences mentioned above.

The algorithm implemented to calculate the MSE of the estimated parameters of the quadratic model (4-8) is illustrated in Figure **4-7**. The results of the MSE calculations of the estimated parameters in these three designs are presented in Table **4-1**.

Table 4-1.: Results of the MSE calculations of the estimated parameters for the prior covariance matrix Σ.

Design / Parameter	D-optimal	Bayesian D-optimal	Penalized Bayesian
θ_0	0.003254675	0.003254675	0.00372378
θ_1	0.001617000	0.001617000	0.00160155
θ_2	0.004948199	0.004948199	0.00561510
$\|\boldsymbol{\theta}\|$	0.001132752	0.002402986	0.002618802

From Table **4-1** it is observed that the MSE values of the parameters θ_0, θ_1 and θ_2 of the D-optimal and Bayesian D-optimal designs are equal. For linear models, the MSEs do not depend on the parameters. They only depend on design points and experimental errors. The design points x_i, $i = 1, 2, \ldots, 9$, of the D-optimal (4-32) and Bayesian

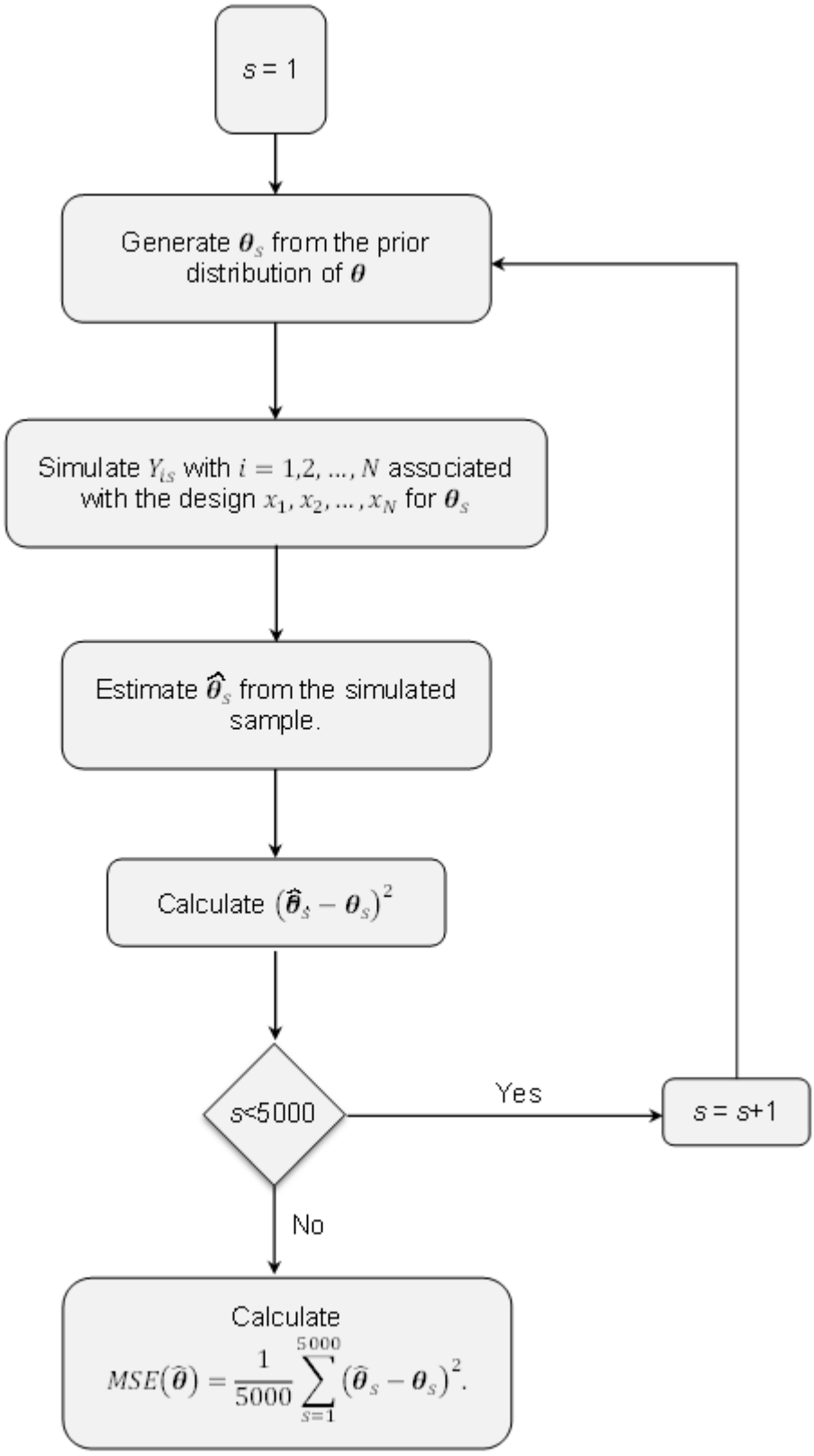

Figure 4-7.: Diagram of MSE calculations of the estimated parameters of the prior distribution in the Bayesian optimal design.

D-optimal (4-33) designs are the same, therefore, it is logical that the MSE values of both designs are equal. However, the norm of the parameter vector $\boldsymbol{\theta}$ has different MSE values of both designs because the quadratic model does not depend linearly on this norm vector.

From Table **4-1**, it can be seen that the MSE values of the penalized Bayesian design are greater than respective MSE values of the two previous designs, except for the MSE value of parameter θ_1. The D-optimal and non-penalized Bayesian designs contain three observations in each design support point $x = \{-1, 0, 1\}$, while the penalized Bayesian design has only one observation in the design midpoint. Two other observations are placed at 0.295. These observations are displaced from the optimal point 0. Therefore, this is expected to improve the accuracy of the parameter estimation $\hat{\theta}_1$ in the penalized Bayesian design.

The D-efficiency of the penalized Bayesian D-optimal design (4-34) with respect to the non-penalized Bayesian D-optimal design (4-33) is equal to 0.96244. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion evaluated for non-penalized design (4-33) and penalized Bayesian D-optimal design (4-34) such that this ratio is in the interval $[0, 1]$. The efficiency value shows that the penalized design is somewhat less efficient than the non-penalized design in question. A small loss of efficiency is observed when using the penalized design rather than the non-penalized design, but the penalized design fulfills the practical design preferences imposed by a researcher.

4.7.2. MSE for Different Prior Covariance Matrices

In Section 4.3.2, it is mentioned that Bayesian D-optimal design for linear models depends on the structures of the prior covariance matrix. In order to compare the performance of non-penalized and penalized Bayesian optimal designs, the MSEs of the parameters and D-efficiencies of penalized designs are calculated for two other prior covariance matrices:

$$\Sigma_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 3 \end{bmatrix}. \tag{4-35}$$

The non-penalized Bayesian D-optimal designs corresponding to the prior covariance matrices Σ_1 and Σ_2 are, respectively,

$$\xi_1^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.434 & 0.133 & 0.434 \end{array} \right\} \text{ and } \xi_2^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.346 & 0.308 & 0.346 \end{array} \right\}. \tag{4-36}$$

It is observed that the points of the designs do not change, but the weights do change. When the variances increase as in the case of Σ_2, the weights w_i tend to be uniform; this can be seen in the design ξ_2^*. When the variances decrease as in the case of Σ_1, the weights of the extremes -1 and 1 tend to 0.5, and the weight of the midpoint tends to zero; this can be seen in the design ξ_1^*.

The procedure of the efficient design apportionment (Pukelsheim 1993, p. 309) helps to obtain the respective exact designs for finite sample size $N = 9$:

$$\xi_{1,N=9}^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 4 & 1 & 4 \end{array} \right\} \quad \text{and} \quad \xi_{2,N=9}^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 3 & 3 & 3 \end{array} \right\}. \tag{4-37}$$

Thus, the exact design corresponding to the matrix Σ_2 is the same design of the matrix Σ, i.e., three observations in each support point. However, the exact design corresponding to the matrix Σ_1 is different. It has four observations in each extremal support point and only one observation in the design midpoint.

The same desirability functions (4-19) and (4-20) are used for designs ξ_1^* and ξ_2^* for the purpose of obtaining designs with the same practical characteristics desired as the ones of the penalized design (4-23). The respective penalized exact designs for sample size $N = 9$ are obtained as

$$\xi_{1P,N=9}^* = \left\{ \begin{array}{cccc} -1 & 0 & 0.307 & 1 \\ 3 & 1 & 2 & 3 \end{array} \right\} \quad \text{and} \quad \xi_{2P,N=9}^* = \left\{ \begin{array}{cccc} -1 & 0 & 0.294 & 1 \\ 3 & 1 & 2 & 3 \end{array} \right\}. \tag{4-38}$$

It can be observed that the resulting designs have the practical characteristics desired by the researcher, i.e., four distinct design points separated at least 0.3 units apart and new design point placed between 0 and 1 with two observations.

The results of the MSE calculations of the estimated parameters for the two prior variance matrices Σ_1 and Σ_2 are presented in Tables **4-2** and **4-3**, respectively.

From Table **4-2**, it can be seen that the MSE values of the penalized Bayesian design are less than respective MSE values of the non-penalized Bayesian design, except for the MSE value of parameter θ_1. It is greater than the respective MSE value of the non-penalized Bayesian design. In this case, the MSEs have an opposite behavior of the MSEs of the estimated parameters for the prior variance-covariance matrix Σ. A possible explanation of this behavior is the distribution of the points of both designs. The non-penalized Bayesian design has four observations in each extremal support point ± 1 and only one observation in the design midpoint, while the penalized Bayesian design has three observations around zero in two different points. Therefore, the 2 out of 3

parameters of the penalized Bayesian design have more accurate estimations than the parameters of the non-penalized Bayesian design.

Table 4-2.: Results of the MSE calculations of the estimated parameters for the prior covariance matrix Σ_1.

Parameter \ Design	Bayesian D-optimal	Penalized Bayesian
θ_0	0.009946392	0.003765882
θ_1	0.001221735	0.001600057
θ_2	0.011233774	0.005673496
$\|\boldsymbol{\theta}\|$	0.003741527	0.002141959

From Table **4-3**, it can be seen that the MSE values of the penalized Bayesian design are greater than respective MSE values of the non-penalized Bayesian design, except for the MSE value of parameter θ_1. It is less than the respective MSE value of the non-penalized Bayesian design. In this case, the MSEs have a similar behavior of the MSEs of the estimated parameters for the prior covariance matrix Σ. The non-penalized Bayesian design contains three observations in each design point $x = \{-1, 0, 1\}$, while the penalized Bayesian design has only one observation in the design midpoint. Two other observations are placed at 0.294; they are displaced from the optimal point 0. This affects the accuracy of the estimation of the parameters.

Table 4-3.: Results of the MSE calculations of the estimated parameters for the prior variance-covariance matrix Σ_2.

Parameter \ Design	Bayesian D-optimal	Penalized Bayesian
θ_0	0.003254675	0.003720369
θ_1	0.001617000	0.001601671
θ_2	0.004948199	0.005610360
$\|\boldsymbol{\theta}\|$	0.002624297	0.002869813

In summary, the MSE values of the penalized Bayesian designs are equal for these two prior variance-covariance matrices because the penalized exact designs are the same. While the MSE values of the non-penalized Bayesian designs are not equal for these the prior variance-covariance matrices.

The D-efficiencies of the penalized Bayesian D-optimal designs (4-38) relative to the non-penalized Bayesian D-optimal designs (4-37) are equal to 0.83048 and 0.95422, respectively. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion evaluated for non-penalized design (4-37) and penalized Bayesian D-optimal design (4-38) such that this ratio is in the interval $[0, 1]$. These D-efficiency values show that the penalized Bayesian designs for these two covariance matrices are somewhat less efficient than respective non-penalized Bayesian designs. Furthermore, the penalized design's D-efficiency corresponding to the covariance matrix Σ_1 is lower than the design's D-efficiency corresponding the covariance matrix Σ_2. The possible explanation for this result is the very different form of the first non-penalized Bayesian design with respect to the penalized one.

4.7.3. MSE of the estimated parameters of posterior distribution

In the normal linear regression model given in (4-2) under the assumptions mentioned in Section 4.2 the posterior distribution of $\boldsymbol{\theta}$ is also normal with mean

$$\widehat{\boldsymbol{\theta}}_B = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{R}\boldsymbol{\mu}) \tag{4-39}$$

and covariance matrix $\sigma^2(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R})^{-1}$, according to Chaloner & Verdinelli (1995). More details are found in Section 2.3.2.

The MSEs associated with $\widehat{\boldsymbol{\theta}}_B$ are calculated via simulation in order to compare the performance of non-penalized and penalized Bayesian D-optimal designs for the quadratic regression model given in (4-8).

The experimental observations of non-penalized and penalized Bayesian D-optimal designs are simulated according to the quadratic model given in (4-8) in the assigned design points of each design given in (4-33) and (4-34), respectively.

The mean vector $\widehat{\boldsymbol{\theta}}_B$ of the posterior distribution is estimated according to (4-39) from the simulated samples according to each design. Later the estimations $\widehat{\boldsymbol{\theta}}_B$ are compared with parameters generated by the conditional prior distribution. The estimations $\widehat{\boldsymbol{\theta}}_B$ are also compared with parameters previously estimated by the least-squares method (LSM) from the same simulated samples according to each design.

The algorithm implemented to calculate the MSE of the estimated parameters of the posterior distribution is illustrated in Figure **4-8**.

The results of the MSE calculations of the estimated parameters of the posterior distribution are presented in Table **4-4**.

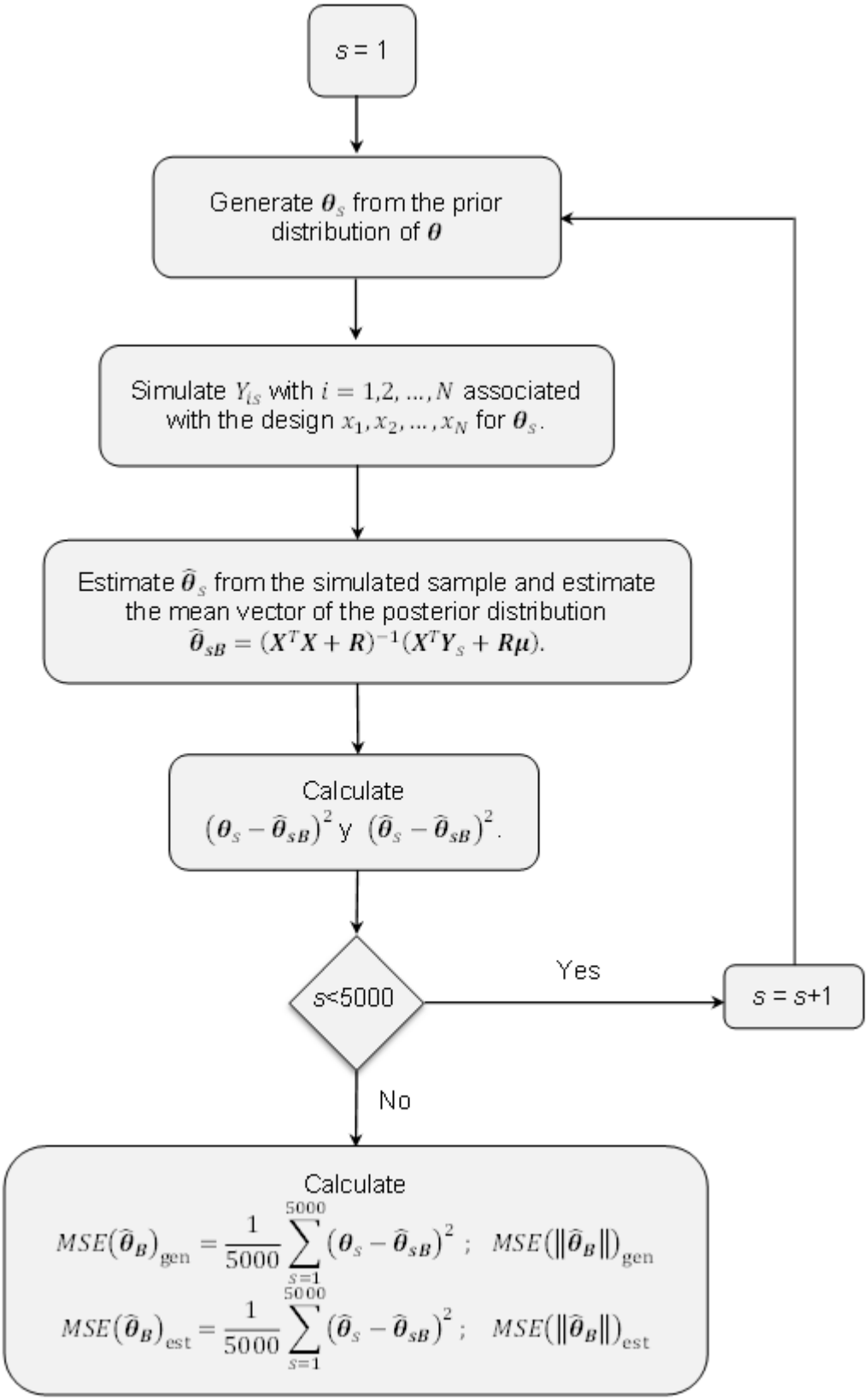

Figure 4-8.: Diagram of MSE calculations the estimated parameters of the posterior distribution in the Bayesian optimal design.

Table 4-4.: Results of the MSE calculations of the estimated parameters $\widehat{\boldsymbol{\theta}}_B$ of the posterior distribution.

Parameter \ Design	Bayesian D-optimal	Penalized Bayesian
θ_{0B}gen	0.063436953	0.077407304
θ_{0B}est	0.061664155	0.075681830
θ_{1B}gen	0.006403749	0.006720553
θ_{1B}est	0.004916254	0.005248749
θ_{2B}gen	0.117668875	0.137847546
θ_{2B}est	0.116274093	0.136663884
$\|\boldsymbol{\theta}_B\|$gen	0.07007669	0.08036217
$\|\boldsymbol{\theta}_B\|$est	0.06915855	0.07959425

From Table **4-4**, it can be seen that

- The MSE values of parameter θ_{1B} of both designs are less than the respective MSE values of the other parameters, while the MSE values of parameter θ_{2B} are greater than the others.
- The MSE values of the estimations $\widehat{\boldsymbol{\theta}}_B$ compared with the parameters generated $\boldsymbol{\theta}$ are greater than those compared with the estimated parameters $\widehat{\boldsymbol{\theta}}$ by LSM for both non-penalized and penalized Bayesian designs.
- All the MSE values of the penalized Bayesian design are greater than the respective MSE values of the non-penalized Bayesian design. This result confirms a small increase in the generalized variance of the estimated parameters (4.232) in the penalized Bayesian D-optimal design as compared to non-penalized design (4.008). The penalized Bayesian optimal designs change global Bayesian optimality in favor of practical design characteristics.

The example of the quadratic model, given in (4-8), illustrates the methodology that allows the researcher to find the penalized Bayesian D-optimal design for linear regression models. The penalized Bayesian optimal design technique allows to incorporate available prior information of unknown parameters to the design process and specify desirable characteristics of the experimental design that are essential for the researcher. These practical characteristics are defined by an overall desirability function that is the geometric mean of the individual desirability functions. The penalty function that is one minus the overall desirability function is added to an "alphabetic" optimality

criterion to penalize not practical optimal designs. The resulting penalized Bayesian D-optimal design is optimal according to the Bayesian D-optimal design criterion for linear regression models and also fulfills the practical design preferences. As a result, the penalized Bayesian D-optimal design provides the appropriate balance between Bayesian D-optimality and desirability.

5. Penalized Bayesian Optimal Designs for Nonlinear Models

Nonlinear regression models are widely used to describe the dependencies between a response and an explanatory variable (Bates & Watts 1988). They arise in scientific experiments in a variety of areas, such as pharmacology, biology, and agriculture. An important feature of nonlinear models is that, unlike for linear models, the Fisher information matrix and hence the optimal design depend on the values of the unknown parameters. Chernoff (1953) suggests adopting a "best guess" for the unknown parameters which leads to locally optimal designs. The problem may arise when that guess is not close enough to the true parameter vector because the design obtained may not be optimal. Bayesian design methodology has a lot to offer to experimental design practice, whereby prior knowledge of unknown parameters can always be used for the design process.

This chapter describes the methodology used to find penalized Bayesian D-optimal designs for nonlinear regression models. The Bayesian approach to experimental design allows to incorporate available prior knowledge (previous studies, literature, expert opinions) about unknown parameters into the design; and the penalized optimal design strategy based on the use of desirability functions helps to obtain optimal designs that fulfill practical design preferences imposed by a researcher. The methodology presented here combines the use of desirability functions and the Bayesian approach in the construction of penalized Bayesian D-optimal designs for the nonlinear regression model.

The organization of this chapter is as follows. The nonlinear model is introduced in Section 5.1. A summary of the Bayesian D-optimality criterion for nonlinear models is provided in Section 5.2. This summary is based on the work of Chaloner & Verdinelli (1995). The Bayesian optimal design methodology for nonlinear models is illustrated by an example of the two-parameter exponential regression model in Section 5.3. In Subsection 5.3.1, Bayesian designs are constructed for a series of prior distributions: gamma, lognormal, and uniform. The equivalence theorem for Bayesian D-optimal designs for nonlinear models is used to verify their design optimality. Design efficiency of these Bayesian designs are examined by comparing them with non-Bayesian designs.

The new criterion proposed in this dissertation is considered in Section 5.4. This criterion is used to obtain penalized Bayesian D-optimal designs for nonlinear models. The penalized Bayesian optimal design methodology for nonlinear models is detailed in Section 5.4. Then, this methodology is illustrated by an example of the two-parameter exponential regression model in Section 5.5. The penalized designs are constructed for different prior distributions of unknown parameters. Design D-efficiency and MSEs are evaluated based on simulation studies.

5.1. Nonlinear Model

Consider the nonlinear regression model

$$y = \eta(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon, \tag{5-1}$$

where

- y is a continuous response (or dependent or observed) variable.
- $\boldsymbol{x}$ is a $m \times 1$ vector of explanatory (or predictors or independent or control or design) variables. Suppose that the vector $\boldsymbol{x} \in \mathscr{X}$ called space or design region, where $\mathscr{X}$ is a compact set in $\mathbb{R}^m$.
- $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, where $\boldsymbol{\theta} \in \Theta$, here Θ is an open convex set in $\mathbb{R}^p$.
- $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is a known continuous nonlinear function in $\mathscr{X}$ with respect to the model parameters $\boldsymbol{\theta}$.
- ε is an uncorrelated normally distributed error term with zero mean and constant variance σ^2 and independent of $\boldsymbol{x}$.

It assumes that an experiment is performed at points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathscr{X}$ (some of these points can be repeated). An independent observation y_i will be observed at each point of $\boldsymbol{x}_i$, for $i = 1, 2, \ldots, N$. Under the above assumptions $\mathrm{E}(y_i) = \eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ for $i = 1, 2, \ldots, N$ and $\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_N$, where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_N)^T$ is the vector of N observations and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

5.2. Bayesian D-Optimality Criterion for Nonlinear Models

If the study model is nonlinear, it is usually harder to find the experimental design, than it is for a linear model. The search for optimal Bayesian designs is complicated in

nonlinear models because Equation (2-35) does not have an analytical solution. In this case, the asymptotic approximations may be used.

The normal approximation to the posterior distribution of $\boldsymbol{\theta}$ is often used. Using normal approximation (2-43) to the posterior distribution of parameters $\boldsymbol{\theta}$ in the Equation (2-36), the expected utility $U_1(\xi)$ can be written as

$$U_1(\xi) = -\frac{p}{2}\log(2\pi) - \frac{p}{2} + \frac{1}{2}\int_{\Theta} \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{5-2}$$

where $N\boldsymbol{I}(\xi,\boldsymbol{\theta})$ is the expected Fisher information matrix for nonlinear model with unknown parameters and $\pi(\boldsymbol{\theta})$ is the prior distribution of parameters $\boldsymbol{\theta}$.

The Bayesian D-optimality criterion for nonlinear models can be obtained by dropping the constant and multiplier terms in Equation (5-2), as follows

$$\Phi_B(\xi) = \int_{\Theta} \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{5-3}$$

According to Chaloner & Verdinelli (1995), the Bayesian solution to the experimental design problem is to find the best design and the best decision rule that maximizes expected utility. Hence, when the criterion (5-3) is maximized, the expected gain in the Shannon information of the posterior distribution (5-2) is also maximized or, equivalently, the variance of the posterior distribution is minimized.

Design ξ^*_{BD} that satisfies the following condition (Zhang 2006)

$$\xi^*_{BD} = \arg\min_{\xi\in\Xi} \int_{\Theta} \{-\log\det\{\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} \tag{5-4}$$

is the Bayesian D-optimal design for nonlinear regression models with respect to the prior distribution $\pi(\boldsymbol{\theta})$ of parameters $\boldsymbol{\theta}$. More details for Bayesian optimal designs for nonlinear models is found in Section 2.3.

To evaluate the integral in (5-4) the technique called the Monte Carlo (MC) method is used. The principle of the MC method for integral approximations in statistics is based on computer simulations of random variables according to the prior distribution of $\boldsymbol{\theta}$ to produce an approximation of integrals converging with the increasing number of simulations. Its justification is based on the Law of Large Numbers.

A random sample of parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_G)$ is generated from respective prior distribution $\pi(\boldsymbol{\theta})$, and then the empirical average

$$\overline{\Psi}_G(\xi) = \frac{1}{G}\sum_{j=1}^{G} \Psi(\xi,\boldsymbol{\theta}_j) = \frac{1}{G}\sum_{j=1}^{G} [-\log\det\{\boldsymbol{I}(\xi,\boldsymbol{\theta}_j)\}], \tag{5-5}$$

is proposed as an approximation of integral in (5-4). This average is computed in R software (Team 2018) through command lines as *mean*, since $\overline{\Psi}_G(\xi)$ converges almost surely to $\Psi(\xi)$ by the Strong Law of Large Numbers (Albert 2009).

The methodology of the construction of the Bayesian D-optimal design for nonlinear models is illustrated by an example of the exponential growth regression model with two parameters.

5.3. Example: Exponential Growth Model

Exponential functions are useful in modeling many biological, chemical and physical phenomena, such as populations growth, chemical reactions, radioactive decay, and drug concentration in the bloodstream.

Consider the exponential regression model with two parameters

$$\eta(x;\boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 x), \quad x \geq 0,\ \theta_1 > 0,\ \theta_2 > 0, \tag{5-6}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ denotes the unknown vector of parameters.

Mukhopadhyay & Haines (1995) show that the non-Bayesian D-optimal designs for the exponential model (5-6) are balanced on exactly two support points for *all* values of the parameters $\boldsymbol{\theta}$ and do not depend on the parameter θ_1, i.e.,

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1/\theta_2 \\ 0.5 & 0.5 \end{array} \right\}. \tag{5-7}$$

However, the Bayesian D-optimal designs for the exponential model (5-6) are not necessarily based on exactly two support points. The number of support points of Bayesian designs increases as the prior distribution for $\boldsymbol{\theta}$ becomes more dispersed (Firth & Hinde 1997, Duarte & Wong 2015). This behavior will be illustrated in the next Subsection 5.3.1.

5.3.1. Bayesian D-Optimal Design

Let π denote a bivariate prior distribution for $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ with independent marginal distributions π_1 and π_2. A design ξ_B^* is called Bayesian D-optimal (with respect to the prior π) if ξ_B^* maximizes

$$\Phi_\pi(\xi) = \mathrm{E}_\pi\left[\log\det\left\{\boldsymbol{I}(\xi,\boldsymbol{\theta})\right\}\right] = \int_\Theta \log\det\left\{\boldsymbol{I}(\xi,\boldsymbol{\theta})\right\} d\pi(\boldsymbol{\theta}) \tag{5-8}$$

and $\Phi_\pi(\xi_B^*) < \infty$. Suppose that, this prior distribution π is sufficiently regular such that there exists at least one design ξ with $\mathrm{E}_\pi[\,|\log\det\{\boldsymbol{I}(\xi,\boldsymbol{\theta})\}|\,] < \infty$.

For the model (5-6) a straightforward calculation (see Appendix A) gives

$$\mathrm{E}_\pi[\log\det\{\boldsymbol{I}(\xi,\boldsymbol{\theta})\}] = 2\mathrm{E}_\pi[\log(\theta_1)] + \mathrm{E}_\pi[\log\det\{\boldsymbol{I}(\xi,\boldsymbol{\theta}^*)\}], \tag{5-9}$$

where $\boldsymbol{\theta}^* = (1,\theta_2)^T$. Consequently the Bayesian D-optimal design for the model (5-6) depends on the prior distribution only through the marginal distribution π_2 of θ_2.

We begin to investigate the conditions under which the Bayesian D-optimal design for the exponential model (5-6) is based on two points. The following theorem gives the best Bayesian D-optimal design in the class of all designs supported at two points.

Theorem 5.1. *For the exponential model (5-6), the Bayesian D-optimal design among all designs with two support points puts equal masses at the points*

$$x_1 = 0, \qquad x_2 = [\mathrm{E}_{\pi_2}(\theta_2)]^{-1}, \tag{5-10}$$

where $\mathrm{E}_{\pi_2}[\cdot]$ *denotes the expectation with respect to the marginal distribution* π_2 *of the prior* π*.*

The proof of Theorem 5.1 can be found in Dette & Sperlich (1994).

This theorem shows that if a Bayesian D-optimal design for the exponential model in (5-6), supported at two points exists, then it must be of the form given in Theorem 5.1. Moreover, this design is also Bayesian D-optimal in the class of all two-point designs.

It is noted that the Bayesian D-optimal design is not dependent on the parameter θ_1. The resulting Bayesian optimal design depends on the marginal prior distribution π_2 of θ_2 and can vary substantially when different marginal prior distributions are used. The Bayesian D-optimal designs are constructed for a range of marginal prior distributions $\pi_2(\theta_2)$: gamma, lognormal, and uniform with different hyperparameters, in order to examine the influence of this marginal prior distribution. These designs are obtained, minimizing the following criterion of Bayesian D-optimality

$$\Psi_B = \int_\Theta \{-\log\det\{I(\xi,\boldsymbol{\theta})\}\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{5-11}$$

To evaluate the integral in (5-11), the random vectors $\boldsymbol{\theta}$ are generated according to the prior distribution π with the marginal prior distribution $\pi_2(\theta_2)$. Then the MC method is used to calculate this integral.

Since θ_2 is a scale parameter for the exponential model (5-6), it is admissible to adopt a gamma distribution for its prior (Mukhopadhyay & Haines 1995), i.e.

$$\pi_2(\theta_2) = \frac{\beta^\alpha \theta_2^{\alpha-1} \exp(-\beta\theta_2)}{\Gamma(\alpha)}, \quad \theta_2 > 0, \tag{5-12}$$

where the hyperparameters α and β are positive and known.

If the hyperparameter α is sufficiently large ($\alpha \geq 3$), the two-point design of Theorem 5.1 is globally Bayesian D-optimal design in the class of all designs (Dette & Sperlich 1994, Mukhopadhyay & Haines 1995, Dette & Neugebauer 1997).

The Equivalence Theorem for nonlinear models given by Chaloner & Larntz (1989) helps to check the D-optimality of resulting Bayesian designs. For any fixed prior π a design ξ_B^* is Bayesian D-optimal if and only if

$$\mathrm{E}_\pi\left[\mathrm{tr}\left(\boldsymbol{I}(\xi_B^*, \boldsymbol{\theta})^{-1} \boldsymbol{I}(x, \boldsymbol{\theta})\right)\right] = \mathrm{E}_{\pi_2}\left[\mathrm{tr}\left(\boldsymbol{I}(\xi_B^*, \boldsymbol{\theta}^*)^{-1} \boldsymbol{I}(x, \boldsymbol{\theta}^*)\right)\right] \leq 2 \tag{5-13}$$

for all x in the design space $\mathscr{X}$, with equality in (5-13) occurring if x belongs to the support of ξ_B^*.

The hyperparameters α and β in the gamma prior distribution (5-12) are chosen such that $\mathrm{E}_{\pi_2}(\theta_2) = 1$, that is, $\alpha = \beta$ in all cases. If the hyperparameter $\alpha = 4$, that is, according to Dette & Sperlich (1994), this hyperparameter has a sufficiently large value, therefore, the resulting Bayesian D-optimal design is balanced at two points: $x_1 = 0$ and $x_2 = [\mathrm{E}_{\pi_2}(\theta_2)]^{-1}$. In this case the variance of θ_2 is equal to 0.25. The hyperparameters of the other prior distributions of θ_2, lognormal and uniform, are chosen in such a way as to preserve the same mean and variance of θ_2 of the gamma prior distribution (5-12), i.e., $\mathrm{E}_{\pi_2}(\theta_2) = 1$ y $\mathrm{Var}_{\pi_2}(\theta_2) = 0.25$.

The continuous prior distributions of θ_2 (gamma, lognormal and uniform) are constructed through simulations in R (Team 2018). The number (1000) of simulations refers to the amount of θ_2 values generated from the respective prior distribution. These values of the gamma, lognormal and uniform prior distributions of θ_2 are generated via simulation by functions *rgamma*, *rlnorm* and *runif* in R (Team 2018), respectively. The gamma and uniform distributions are simulated with hyperparameters specified in Table **5-1**, while the lognormal distribution is simulated with hyperparameters $meanlog = -0.1115718$ and $sdlog = 0.47238073$, which correspond to the mean equal to 1 and variance equal to 0.25.

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is used in R-project (Team 2018) to minimize the Bayesian D-optimal criterion (5-11) to find the Bayesian

D-optimal two-point design for each prior distribution of θ_2: gamma, lognormal and uniform. The resulting Bayesian D-optimal two-points designs for these prior distributions are presented in Table **5-1**.

Table 5-1.: Bayesian D-optimal two-point designs and the efficiency of the D-optimal design (5-7) with respect to them for different prior distributions of θ_2.

Prior distribution of θ_2	Gamma (4,4)	Lognormal (1,0.25)	Uniform(0.134,1.866)
Support points x_i	0.0000; 0.9925	0.0000; 1.0131	0.0000; 0.9876
Weights w_i	0.5000; 0.5000	0.5000; 0.5000	0.5000; 0.5000
Efficiency	0.99998	0.99995	0.99995

The second row of this table shows the support points of the Bayesian D-optimal two-point design, while the third row contains the corresponding weights. It is noted that the resulting Bayesian two-point designs have the same characteristics of the two-point design of the Theorem 5.1. The inequality (5-13) of the Equivalence Theorem for nonlinear models is used to verify the D-optimality of these two-point designs. Figure **5-1** shows that the function $d\left(x, \xi_B^*\right)$ achieves its maximum value 2 at the support points 0 and 1, demonstrating the D-optimality of the resulting Bayesian two-point designs according to the Equivalence Theorem for Bayesian D-optimal designs for nonlinear models (Chaloner & Larntz 1989).

The fourth row of this table presents the efficiencies of the locally D-optimal design ξ^* given in (5-7) with respect to the Bayesian D-optimal two-point designs given in Table **5-1**. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion (5-11) evaluated for the Bayesian D-optimal two-point designs and the locally D-optimal design (5-7) such that this ratio is in the interval $[0, 1]$. It is observed that the D-efficiency of the locally D-optimal designs relative to the Bayesian D-optimal two-point designs are all greater than of 99.99%. Thus, the Bayesian D-optimal two-point designs coincide with the respective locally D-optimal designs.

If the hyperparameter α in (5-12) is not sufficiently large ($\alpha < 3$), the two-point design of Theorem 5.1 cannot be globally optimal with respect to the gamma prior (5-12). In this case, the Bayesian D-optimal designs have at least three support points because the variance of parameter θ_2 increases as hyperparameter α of the gamma prior (5-12) decreases. Therefore, more support points are required for the design of more dispersed prior distributions of θ_2.

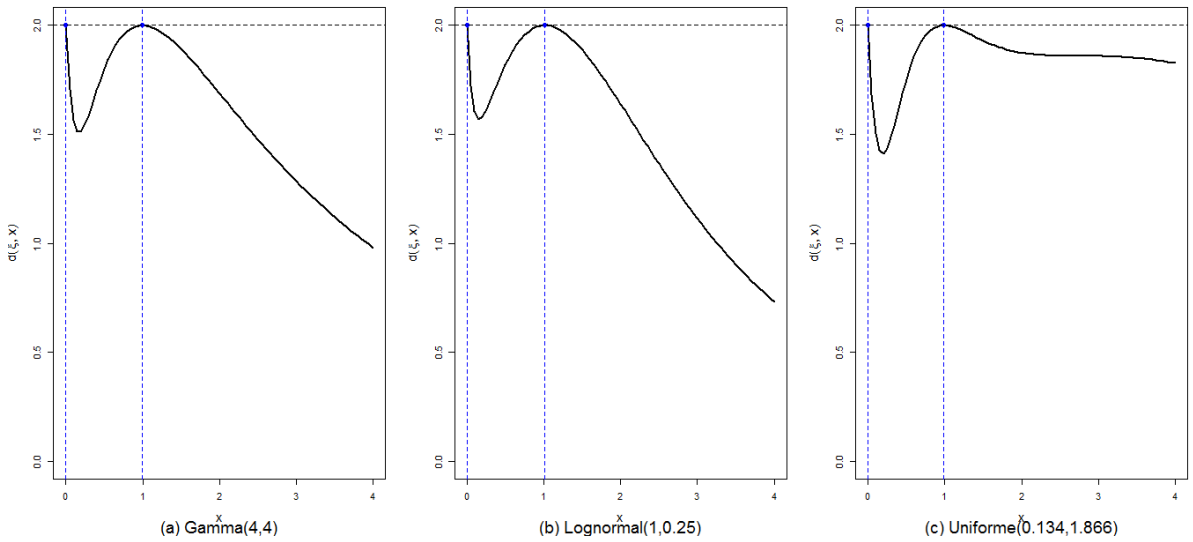

Figure 5-1.: Expected variance $d(x, \xi_B^*)$ for different prior distributions of θ_2 for the Bayesian D-optimal two-point designs given in Table **5-1**.

The hyperparameters of each prior distribution of θ_2, gamma, lognormal, and uniform, are chosen so that Bayesian D-optimal three-point design is obtained for each prior. The gamma and lognormal prior distributions have the same mean of the parameter θ_2, i.e., $\mathrm{E}_{\pi_2}(\theta_2) = 1$, but the mean of the uniform prior can not have the same value. The uniform distribution is symmetric, unlike the gamma and lognormal distributions. The maximum admissible variance for $\theta_2 > 0$ is equal to $1/3$, but the three-point designs for the uniform prior with the variance equal to $1/3$ are not globally optimal. Therefore, the mean of the uniform prior should be increased to 2. The gamma and lognormal prior distributions have the same variance of the parameter θ_2, $\mathrm{Var}_{\pi_2}(\theta_2) = 0.4$, but the variance of the uniform prior is equal to 1.08. The gamma and uniform distributions of θ_2 are simulated with hyperparameters specified in Table **5-2**, while the lognormal distribution is simulated with hyperparameters $meanlog = -0.1682361$ and $sdlog = 0.58006227$, which correspond to the mean equal to 1 and variance equal to 0.4. The resulting Bayesian D-optimal three-point designs for these prior distributions are presented in Table **5-2**.

The second row of this table shows the support points of the Bayesian D-optimal three-point design, while the third row contains the corresponding weights. Its fourth row also contains the efficiencies of the locally D-optimal designs ξ^* given in (5-7) with respect to the Bayesian D-optimal three-point designs given in this table. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion (5-11) evaluated for the Bayesian D-optimal three-point designs and the locally D-optimal design (5-7) such that this ratio is in the interval $[0, 1]$. The efficiency values show only a minor improvement

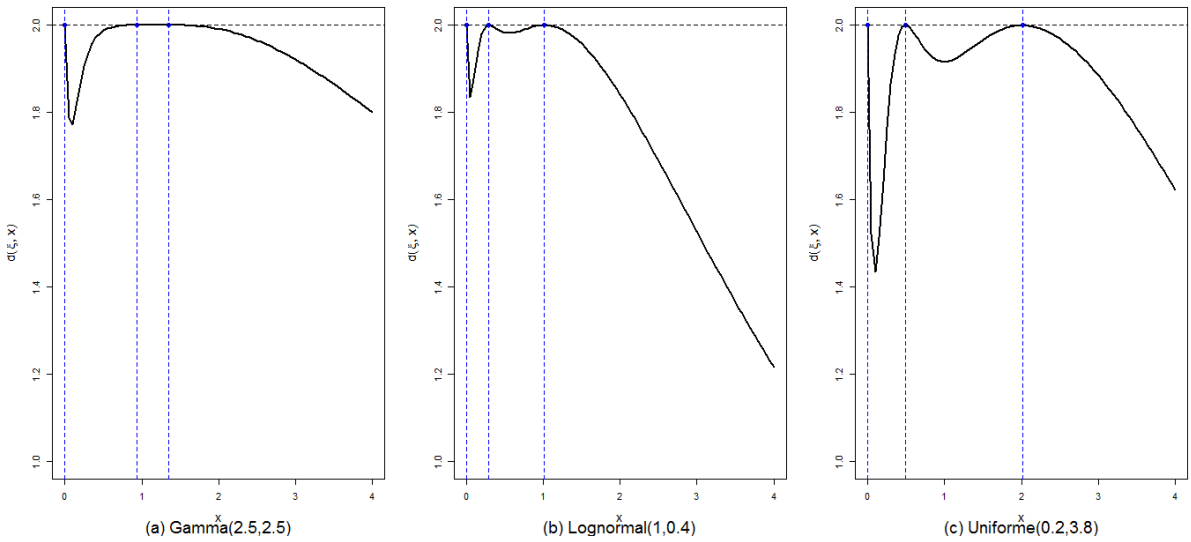

Figure 5-2.: Expected variance $d(x, \xi_B^*)$ for different prior distributions of θ_2 for the Bayesian D-optimal three-point designs given in Table **5-2**.

in the efficiency when one uses a Bayesian three-point design instead of the locally D-optimal two-point design.

Table 5-2.: Bayesian D-optimal three-point designs and the efficiency of the D-optimal design (5-7) with respect to them for different prior distributions of θ_2.

Prior of θ_2	Gamma (2.5,2.5)	Lognormal (1,0.4)	Uniform (0.2,3.8)
Support points x_i	0.000; 0.937; 1.351	0.000; 0.287; 1.017	0.000; 0.493; 2.011
Weights w_i	0.497; 0.373; 0.130	0.497; 0.010; 0.493	0.495; 0.484; 0.021
Efficiency	0.99964	0.99917	0.99922

The inequality (5-13) of the Equivalence Theorem for nonlinear models is used to verify the D-optimality of the resulting Bayesian three-point designs. Figure **5-2** shows that the function $d(x, \xi_B^*)$ achieves its maximum value 2 at the three respective support points, verifying the D-optimality of these Bayesian three-point designs according to the Equivalence Theorem for nonlinear models (Chaloner & Larntz 1989).

If the hyperparameters $\alpha = \beta = 1.5$, the Bayesian D-optimal three-point designs are not globally optimal with respect to the gamma prior (5-12). Here the Bayesian D-optimal design has at least four support points. This result is verified by inequality (5-13) of the Equivalence Theorem for nonlinear models.

Table 5-3.: Bayesian D-optimal four-point designs and the efficiency of the D-optimal design (5-7) with respect to them for different prior distributions of θ_2.

Prior of θ_2	Gamma (1.5,1.5)	Lognormal (1,1.2)
Support points x_i	0.000; 0.675; 1.726; 6.431	0.000; 0.274; 0.752; 2.094
Weights w_i	0.463; 0.307; 0.177; 0.052	0.447; 0.094; 0.207; 0.253
Efficiency	0.96706	0.94561

The parameters of each prior distribution of θ_2, gamma and lognormal, are chosen so that Bayesian D-optimal four-point design is obtained for each prior. The gamma and lognormal prior distributions have the same mean of the parameter θ_2, i.e., $\mathrm{E}_{\pi_2}(\theta_2) = 1$, but the variances are different of each prior distribution: $\sigma^2_{Ga} = 0.6667$ y $\sigma^2_{LN} = 1.2$. The gamma distribution of θ_2 is simulated with hyperparameters specified in Table **5-3**, while the lognormal prior is simulated with hyperparameters $meanlog = -0.39422868$ and $sdlog = 0.887951215$, which correspond to the mean equal to 1 and variance equal to 1.2. The Bayesian D-optimal four-point design can not be obtained for the uniform prior distribution of θ_2 for the exponential model (5-6), because this distribution is symmetric and their mean and variance increase simultaneously.

Bayesian D-optimal four-point designs for gamma and lognormal prior distributions of θ_2 are given in Table **5-3**, which also contains in its fourth row the efficiencies of the locally D-optimal designs ξ^* given in (5-7) with respect to the Bayesian D-optimal four-point designs given in this table. The efficiency values show that the Bayesian designs for these two prior distributions are more efficient than respective locally D-optimal designs. A small improvement in the efficiency is observed when the Bayesian four-point designs are used instead of the locally D-optimal two-point designs.

Figure **5-3** shows that the function $d(x, \xi^*_B)$ achieves its maximum value 2 at the four respective support points of design, verifying the D-optimality of the resulting Bayesian four-point designs for the gamma and lognormal prior distributions of θ_2 according to the Equivalence Theorem for Bayesian D-optimal designs for nonlinear models (Chaloner & Larntz 1989).

This example shows that the number of support points of Bayesian design for the exponential model (5-6) is not fixed. If the variance of the prior distribution of θ_2 tends to 0, then the Bayesian D-optimality criterion is reduced to the non-Bayesian D-optimality, and the Bayesian D-optimal design for a one-point prior is equal a the locally D-optimal design (5-7). When the prior distribution of θ_2 has support only over a small region,

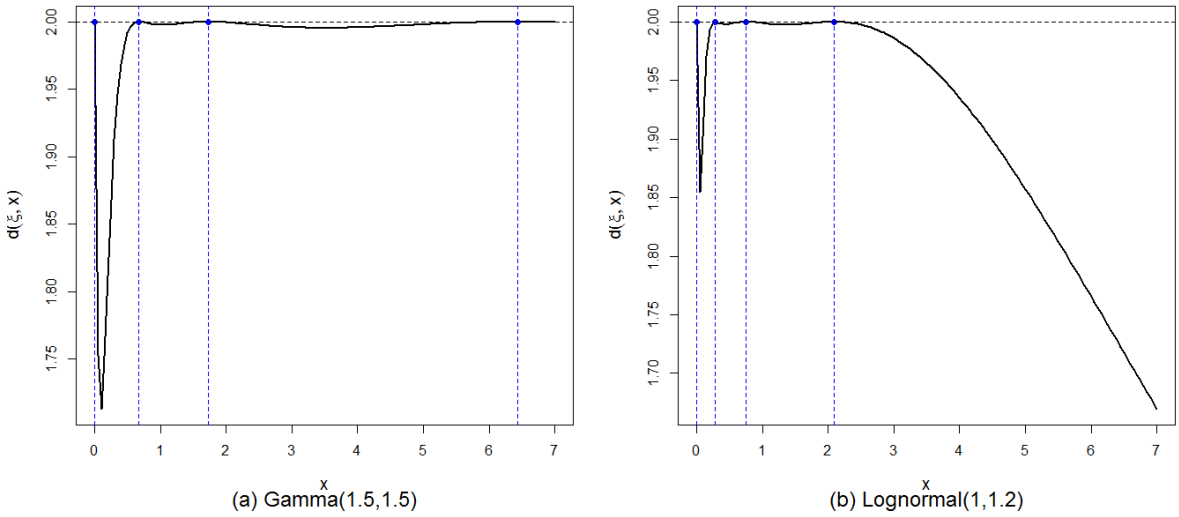

Figure 5-3.: Expected variance $d(x, \xi_B^*)$ for different prior distributions of θ_2 for the Bayesian D-optimal four-point designs given in Table **5-3**.

the Bayesian D-optimal designs have the same number of support points as locally D-optimal design, and that the number of support points increases as the prior becomes more dispersed. A prior distribution with significant variance requires more support points for the Bayesian D-optimal design than a distribution with smaller variance. The same result is obtained for the gamma, lognormal, and uniform prior distributions. Similar behavior was observed by Chaloner & Larntz (1989) in the case of logistic regression.

5.4. New Criterion for Nonlinear Models

In Section 4.4, the new criterion was introduced to find the penalized Bayesian optimal design for linear models. The resulting penalized design is optimal according to the Bayesian optimal design criterion and meets the practical design preferences. Furthermore, the Bayesian approach allows incorporating the available prior information of the unknown parameters into the design process.

The methodology proposed in Section 4.4, that combines the use of desirability functions and the Bayesian approach, can also be used in the construction of penalized Bayesian optimal designs for nonlinear regression models. The optimal design for nonlinear models depends on the unknown parameter values. The non-Bayesian approach commonly uses a "best guess" of the parameter values to design an experiment, which leads to "locally optimal" designs (Chernoff 1953). The problem may arise when that guess is

not close enough to the true parameter vector because the design obtained may not be optimal. In the Bayesian approach, the initial guesses of the unknown parameters do not concentrate on single values. Instead, a prior distribution, which can be centered around the guessed value, is assigned to each unknown parameter. If the optimal design obtained is inadequate from a practical perspective, the penalized technique based on the use of desirability functions can help to construct optimal designs that fulfill traditional optimal design criteria and also practical design preferences imposed by a researcher.

We suggest extending the use of desirability functions described in Section 3.1 in Bayesian optimal designs for nonlinear models. Thus, the researcher can incorporate prior information of the unknown parameters by using a Bayesian approach and also satisfy practical preferences. We propose that a penalized Bayesian D-optimal design for nonlinear models may be found by minimizing the new criterion:

$$\Psi_{BP}(\xi) = \int_{\Theta} \{-\log\det\{I(\boldsymbol{\theta},\xi)\}\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} + \Lambda(1-D(\xi)) = \Psi_B(\xi) + \Lambda(1-D(\xi)) \quad (5\text{-}14)$$

with respect to $\xi \in \Xi$ for a given value of Λ, where Λ is a user-specified scale constant.

The first term of the new criterion in (5-14) represents the Bayesian D-optimality criterion for nonlinear models, where $\Psi(\boldsymbol{\theta},\xi) = -\log\det\{I(\boldsymbol{\theta},\xi)\}$ is the D-optimality criterion for each $\boldsymbol{\theta} \in \Theta$. It follows from Jensen's inequality that if the functional $\Psi(\boldsymbol{\theta},\xi)$ is convex then the Bayesian D-optimality criterion $\Psi_B(\xi)$ is also convex (Ermakov & Zhiglijavsky 1987, p. 202). The second term of the new criterion in (5-14) is a bounded function between 0 and $\Lambda > 0$ (Harrington 1965), which is a penalty function representing constraints applied to the Bayesian D-optimal designs. The minimization of the criterion (5-14) is considered as the maximization of the expected utility (2-44), restricted by the penalty function.

The penalty function, $(1 - D(\xi))$, is added to the Bayesian D-optimality criterion for nonlinear models (5-11) to penalize experimental designs. The Bayesian D-optimality criterion (5-11) and the penalty function have very different values, so a user-specified positive constant, Λ, is required to place them on similar scales. In addition, this constant makes it possible to control the penalty weight in the new criterion (5-14).

Minimization of the criterion (5-14) is similar to the constrained optimization of a scalar-valued objective function of several variables using Lagrange multipliers (Gavin & Scruggs 2012), where the objective function is modified by adding terms that describe the constraints. Thus, the search for the penalized Bayesian D-optimal design for nonlinear models can be considered as minimizing the Bayesian D-optimality criterion for nonlinear models with respect to $\xi \in \Xi$ subject to the restrictions through an overall

desirability function, $D(\xi)$, for a given value of $\Lambda > 0$.

The penalized Bayesian D-optimal criterion in (5-14) is not usually a convex function of $\xi \in \Xi(\mathscr{X})$. However, this criterion is quasiconvex function (Fedorov & Leonov 2013, p. 113). For quasiconvex functions, most results used in convex optimization stay valid (Avriel 2003, ch. 6.1). This fact determines the possibility of finding the global minimum of the penalized Bayesian optimal criterion (5-14).

The methodology for the construction of penalized Bayesian D-optimal designs for nonlinear models is proposed as follows:

1. A non-penalized Bayesian D-optimal design is constructed by minimizing the Bayesian D-optimality criterion (5-11). To evaluate the integral in (5-11), the random variables $\boldsymbol{\theta}$ are generated according to respective prior distribution, and then the MC method, described in Section 5.2, is used to calculate this integral.

2. Appropriate desirability functions are chosen according to the methodology described in Section 3.4. A function, $D(\xi)$, defined in (3-4) as the geometric mean of individual desirability functions, represents the overall desirability for several k practical design preferences (Harrington 1965).

3. A penalized Bayesian D-optimal design is constructed by minimizing the new criterion (5-14), given a value of Λ, where the first term in (5-14) is evaluated by the Monte Carlo (MC) method, described in Section 5.2.

4. Iterative algorithms described in Section 2.4 , are employed to determine the penalized Bayesian D-optimal design ξ_P^*. The initial value of Λ is chosen by $\Lambda_0 = |\min\{\Psi_B(\xi)\}|$, which is an absolute value of the minimum of the non-penalized Bayesian D-optimality criterion (5-11), because Λ is required to bring the two terms in (5-14) to similar scales.

5. Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (5-14) for values of Λ in multiples or submultiples of Λ_0.

6. The final value of Λ is selected within the range in which stability is shown in the responses of the overall desirability function, $D(\xi)$. These responses can be plotted to better observe their behavior.

7. A penalized Bayesian optimal design is determined by the values of x and w of the resulting design ξ_P^*. These values are associated with the minimum value of the penalized Bayesian D-optimal criterion (5-14) for a given value of Λ in the stability range of $D(\xi)$. The resulting penalized Bayesian D-optimal design for a nonlinear

model is optimal according to the Bayesian D-optimal design criterion (5-14) and also fulfills the practical design preferences.

The methodology for the construction of penalized Bayesian D-optimal designs for nonlinear models is illustrated with an example of the exponential growth regression model. This methodology used here for penalized Bayesian design construction can be readily extended to other nonlinear models.

5.5. Penalized Bayesian D-Optimal Design for Exponential Model

This section describes the methodology used to find penalized Bayesian D-optimal designs for nonlinear regression models and illustrates this methodology through an example of the two-parameter exponential model (5-6). General guidelines for choosing appropriate desirability functions are presented to obtain penalized Bayesian optimal designs with desirable characteristics. Resulting penalized Bayesian designs are examined for a range of prior distributions of θ_2: gamma, lognormal, and uniform.

5.5.1. Penalized Designs with Three Support Points

The Bayesian two-point designs in Table **5-1** are examined. It is noted that these designs are balanced on exactly two support points: $x_1 = 0$ and $x_2 = [\mathrm{E}_{\pi_2}(\theta_2)]^{-1}$, i.e., the number of support points of design is the same as the number of unknown parameters of the exponential model (5-6).

We want to have the three-point distinct Bayesian D-optimal design with the minimum two observations in new point and the minimum difference 0.3 units between adjacent design points for three prior distributions of θ_2: gamma, lognormal, and uniform.

Initially, the third point is added, and the following Bayesian D-optimal design for the gamma prior distribution of θ_2 is obtained:

$$\xi_A^* = \left\{ \begin{array}{ccc} 0.0000 & 0.9925 & 0.9925 \\ 0.5000 & 0.3208 & 0.1792 \end{array} \right\}. \tag{5-15}$$

Clearly, this Bayesian D-optimal design is actually supported by two design points, not three.

A penalized-optimal design strategy is developed using two desirability functions $d_1(r_3)$ and $d_2(\min(\mathrm{diff}_{23}))$ to identify the minimum number of observations in the new point x_3 and the minimum difference between the design points x_2 and x_3. A logistic cumulative

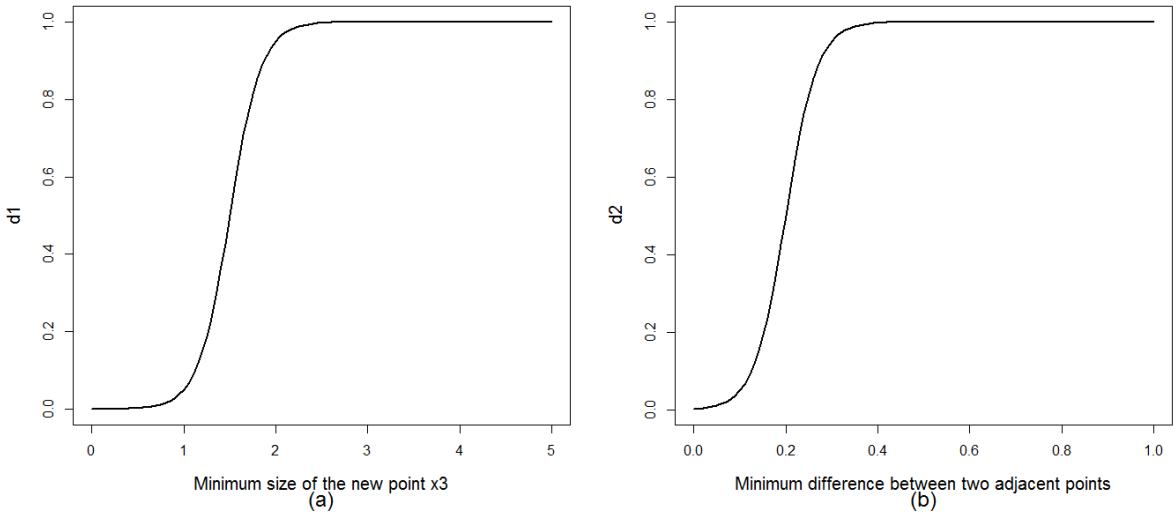

Figure 5-4.: Plots of desirability functions for the exponential model (5-6).

distribution function, of the type described by Gibb (1998), is used to generate the required desirability functions, but other functions can be used to obtain the appropriate shape. The logistic function, the form of the "bigger-is-better" or maximizing desirability function given in (3-11), captures the experimental design preferences.

The desirability functions $d_1(r_3)$ and $d_2(\min(\text{diff}_{23}))$ are defined analogously to the desirability functions (4-19) and (4-20) of the penalized Bayesian design (4-23) for quadratic regression model. Replacing the minimum and maximum values of r_3 and diff_{23} in the formulas (3-11), (3-12) and (3-13), the desirability functions $d_1(r_3)$ and $d_2(\min(\text{diff}_{23}))$ are obtained as:

$$d_1(r_3) = \frac{1}{1 + \exp\left(-\left(r_3 - 1.5\right)/0.17\right)}, \tag{5-16}$$

$$d_2(\min(\text{diff}_{23})) = \frac{1}{1 + \exp\left(-\left(\min(\text{diff}_{23}) - 0.2\right)/0.034\right)}. \tag{5-17}$$

Plots of these desirability functions can be found in Figure **5-4**. The plot of $d_1(r_3)$, given in Figure **5-4**(a), shows that the allocation of less than one observation to the new design point x_3 is unacceptable, and the plot of $d_2(\min(\text{diff}_{23}))$, given in Figure **5-4**(b), shows that the spacing between design points x_2 and x_3 of less than 0.1 units apart is unacceptable.

The overall desirability function is

$$D_1(\xi) = \left(d_1 \times d_2\right)^{1/2}. \tag{5-18}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is employed in R-project (Team 2018) to minimize the following penalized criterion

$$\int_{\Theta} \{-\log\det\{I(\boldsymbol{\theta},\xi)\}\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} + l\Lambda_0(1-D_1(\xi)) \tag{5-19}$$

for the exponential model given in (5-6), where Λ_0 defines an absolute value of the minimum non-penalized Bayesian D-optimal criterion of the two-point design (5-15). Computer code implemented in R-project (Team 2018) to generate penalized Bayesian D-optimal designs may be found in Appendix B. Procedure follows the steps, described in detail in Section 3.5, to obtain the penalized Bayesian D-optimal design for the exponential model (5-6).

The minimum Bayesian D-optimal criterion defines

$$\Lambda_0 = \left|\min_{\xi\in\Xi}\int_{\Theta}\{-\log\det\{I(\boldsymbol{\theta},\xi)\}\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}\right| = 3.401431. \tag{5-20}$$

Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (5-19) for different values $l = 0.01, 0.02, \ldots, 1$. The overall desirability function D_1 responses become stable from approximately $l = 0.10$. A resulting penalized Bayesian D-optimal design is given in Table **5-4** in its second column.

The same penalized-optimal design procedure is performed for lognormal and uniform prior distributions of θ_2, using the same two desirability functions $d_1(r_3)$ and $d_2(\min(\text{diff}_{23}))$ defined in (5-22) and (5-23), respectively. The resulting penalized Bayesian D-optimal designs for these prior distributions are given in Table **5-4** in its third and fourth columns, respectively.

Table 5-4.: Penalized Bayesian D-optimal two-point designs of Table **5-1** for different prior distributions of θ_2.

Prior of θ_2	Gamma (4,4)	Lognormal (1,0.25)	Uniform(0.134,1.866)
Support points x_i	0.000; 0.867; 1.222	0.000; 0.881; 1.256	0.000; 0.851; 1.245
Weights w_i	0.496; 0.296; 0.208	0.496; 0.296; 0.208	0.494; 0.297; 0.208
Efficiency	0.99726	0.99692	0.99699

The second row of this table shows the support points of the penalized Bayesian D-optimal two-point designs while the third row contains the corresponding weights. The efficiencies of these designs with respect to the non-penalized Bayesian D-optimal two-point designs given in Table **5-1** can be found in the fourth row of the table. Here the

efficiency is defined as the ratio of the Bayesian D-optimality criterion (5-11) evaluated for the non-penalized Bayesian designs and corresponding penalized Bayesian designs, such that this ratio is in the interval $[0, 1]$. It is observed that the D-efficiency of the penalized Bayesian designs relative to the non-penalized Bayesian D-optimal two-point designs are all greater than of 99%, indicating the irrelevance of the loss in this efficiency. In summary, the penalized Bayesian D-optimal designs in Table **5-4** are as efficient as non-penalized designs in Table **5-1** but they also have experimental characteristics desired by a researcher.

The MSE associated with the estimated parameters of the exponential model (5-6) are also calculated via simulation in order to compare the performance of non-penalized and penalized Bayesian optimal designs for several prior distributions of θ_2.

The non-penalized Bayesian D-optimal designs are the balanced two-point designs and presented in Table **5-1**. For a sample size of 12, then exact designs have 6 observations at each support point. The penalized Bayesian D-optimal designs are presented in Table **5-4**. The procedure of the efficient design apportionment (Pukelsheim 1993, p. 309) helps to obtain the distribution of observations for finite sample size 12 as follows: $5, 4, 3$.

The experimental observations of all designs are simulated according to the exponential model (5-6) in the corresponding assigned design points. Then the parameters are estimated from the simulated samples according to each design. The algorithm to fit nonlinear models is available using the package *minpack.lm* in R software through the command line as *nlsLM* (Team 2018), implementing the Levenberg-Marquardt nonlinear least-squares algorithm. It can also be used to estimate unknown parameters in nonlinear models. Later the squared differences between estimated parameters and parameters generated according to the respective prior distribution of θ_2 (gamma, lognormal, or uniform) are calculated. Finally, the results of the MSE are obtained by averaging the squared differences mentioned above.

The algorithm implemented to calculate the MSE of the estimated parameters of the exponential model (5-6) is similar to the algorithm used in Section 4.7.1. The results of the MSE calculations of the estimated parameters in these six designs are presented in Table **5-5**.

It is observed that the MSE values of the parameter θ_1 of the Bayesian D-optimal designs are practically equal for the different prior distributions of θ_2 for the same type of design. It is a logical result that these MSE values are equal because the parameter θ_1 is the linear term in the exponential model (5-6). The MSEs do not depend on the parameter values for linear terms. They only depend on design points.

Table 5-5.: Results of the MSE calculations of the estimated parameters for different prior distributions of θ_2.

Parameter \ Prior	Gamma (4,4)	Lognormal (1,0.25)	Uniform(0.134,1.866)
θ_1 non-penalized	0.001655299	0.001635315	0.001646491
θ_1 penalized	0.001984625	0.001951534	0.001947285
θ_2 non-penalized	0.023267231	0.026202415	0.018681273
θ_2 penalized	0.021190210	0.024442046	0.017512663
$\|\boldsymbol{\theta}\|$ non-penalized	0.01777523	0.02031373	0.01360342
$\|\boldsymbol{\theta}\|$ penalized	0.01686706	0.01973333	0.01340609

The MSE values of the parameter θ_2 of the Bayesian D-optimal designs are not equal for the different prior distributions of θ_2, because the exponential model (5-6) does not depend linearly on this parameter. The MSE values of the parameter θ_2 are the greatest for the lognormal prior distribution, while the MSE values of this parameter are the smallest for the uniform prior distribution. The skewness of these distributions has the same behavior.

From Table **5-5**, it can be seen that the MSE values of the parameter θ_1 of the non-penalized Bayesian designs are less than respective MSE values of the penalized Bayesian designs, while the MSE values of the parameter θ_2 have an opposite behavior. This behavior is observed for all prior distributions considered for the parameter θ_2. The non-penalized Bayesian designs contain six observations in each design support point $x = \{0, 1\}$, while the penalized Bayesian designs have only five observations in $x = 0$. Hence, it is expected that this affects the accuracy of the parameter θ_1 estimation in the penalized Bayesian design, because the observations at point $x = 0$ determine the estimation of this parameter. While the observations at point $x = 1$ determine the estimation of θ_2. The penalized Bayesian design has 7 observations distributed around $x = 1$; therefore the most accurate estimates of θ_2 are obtained in the penalized design.

5.5.2. Penalized Designs with Four Support Points

The penalized Bayesian D-optimal design with four support points is constructed to compare the penalized and non-penalized Bayesian designs. We want to obtain the penalized Bayesian D-optimal four-point design with similar characteristics of the designs in Table **5-3**.

Initially, two new points are added, and the following Bayesian D-optimal design for the gamma prior distribution of θ_2 is obtained:

$$\xi_A^* = \left\{ \begin{array}{cccc} 0.000 & 0.000 & 1.006 & 1.006 \\ 0.392 & 0.108 & 0.259 & 0.241 \end{array} \right\}. \tag{5-21}$$

Clearly, this Bayesian D-optimal design is actually supported by two design points, not four.

A penalized-optimal design strategy is developed using four desirability functions $d_1(r_2)$, $d_2(\min(\text{diff}_{23}))$, $d_3(x_2)$, and $d_4(x_4)$. A logistic cumulative distribution function, of the type described by Gibb (1998), is used to generate the required desirability functions $d_1(r_2)$ and $d_2(\min(\text{diff}_{23}))$, but other functions can be used to obtain the appropriate shape. The logistic function, the form of the "bigger-is-better" or maximizing desirability function given in (3-11), captures the experimental design preferences of the minimum number of observations in the point x_2 and the minimum difference between the support points x_2 and x_3. The normal density function, the bell shape or "target" desirability function given in (3-9) is used to create the desirability functions $d_3(x_2)$ and $d_4(x_4)$ to characterize the locations of x_3 and x_4, respectively.

Taking into consideration the characteristics of the non-penalized four-design in Table **5-3**, the minimum and maximum values of r_2 and diff_{23} are defined. Replacing the values of $\gamma_{1,2} = 0.05$, $r_{2(\min)} = 2$, $r_{2(\max)} = 3$, $\text{diff}_{23(\min)} = 0.8$, and $\text{diff}_{23(\max)} = 1$ in the formulas (3-11), (3-12) and (3-13), the desirability functions $d_1(r_2)$ and $d_2(\min(\text{diff}_{23}))$ are obtained as:

$$d_1(r_2) = \frac{1}{1 + \exp\left(-\left(r_2 - 2.5\right)/0.17\right)}, \tag{5-22}$$

$$d_2(\min(\text{diff}_{23})) = \frac{1}{1 + \exp\left(-\left(\min(\text{diff}_{23}) - 0.9\right)/0.034\right)}. \tag{5-23}$$

Plots of these desirability functions are shown in Figure **5-5**. The plot of $d_1(r_2)$, given in Figure **5-5**(a), shows that it is unacceptable to allocate less than two observations to the support point x_2; and the plot of $d_2(\min(\text{diff}_{23}))$, given in Figure **5-5**(b), exhibits that the spacing between design points of less than 0.8 units apart is not acceptable.

In order to keep the support point x_2 close to 0.675, it is assumed that the target of value range of x_2 is 0.675 and the interval width of desirability function is 0.2. Replacing the values of $\gamma_3 = 0.05$, $a_3 = 0.675$ and $\delta_3 = 0.2$ in the formulas (3-9) and (3-10), the desirability function $d_3(x_2)$ is obtained:

$$d_3(x_2) = \exp\left\{-\frac{1}{2}\left(\frac{x_2 - 0.675}{0.0817}\right)^2\right\}. \tag{5-24}$$

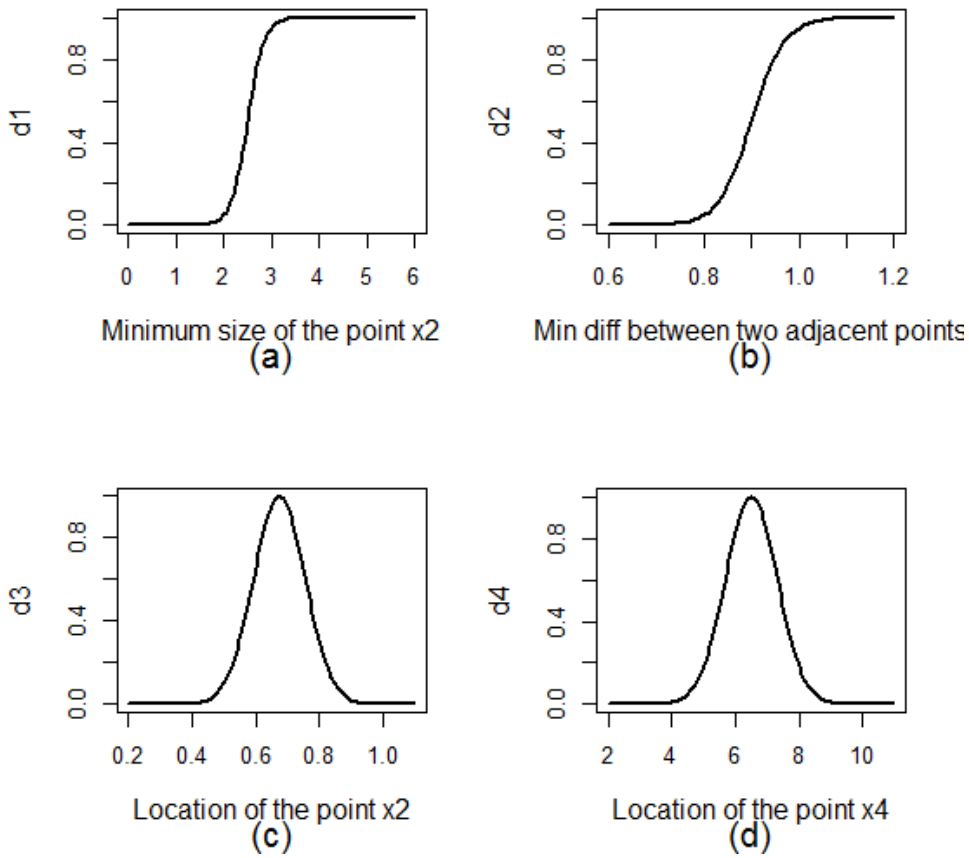

Figure 5-5.: Plots of desirability functions for the exponential model (5-6).

A plot of the last desirability function, given in Figure **5-5**(c), shows that the location of the design point x_2 outside the interval $[0.475, 0.875]$ is not acceptable.

In order to keep the support point x_4 close to 6.5, it is assumed that the target of value range of x_4 is 6.5 and the interval width of desirability function is 2. Replacing the values of $\gamma_4 = 0.05$, $a_4 = 6.5$ and $\delta_4 = 2$ in the formulas (3-9) and (3-10), the desirability function $d_4(x_4)$ is obtained:

$$d_4(x_4) = \exp\left\{-\frac{1}{2}\left(\frac{x_4 - 6.5}{0.817}\right)^2\right\}. \tag{5-25}$$

A plot of the last desirability function, given in Figure **5-5**(d), exhibits that the location of the design point x_4 outside the interval $[4.5, 8.5]$ is unacceptable.

The overall desirability function is

$$D_2(\xi) = \left(d_1 \times d_2 \times d_3 \times d_4\right)^{1/4}. \tag{5-26}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is employed in R-project (Team 2018) to minimize the following criterion

$$\int_{\Theta} \left\{-\log\det\left\{I\left(\boldsymbol{\theta},\xi\right)\right\}\right\} \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} + l\Lambda_0(1 - D_2(\xi)) \tag{5-27}$$

for the exponential model given in (5-6), where Λ_0 defines an absolute value of the minimum non-penalized Bayesian D-optimal criterion (5-20) of the two-point design (5-15). Computer code implemented in R-project (Team 2018) to generate penalized Bayesian D-optimal designs may be found in Appendix B. Procedure follows the steps, described in detail in Section 3.5, to obtain the penalized Bayesian D-optimal design with similar characteristics of the respective four-point design in Table **5-3**.

Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (5-27) for different values $l = 0.01, 0.02, \ldots, 1$. The responses of the overall desirability function D_2 become stable from approximately $l = 0.97$. A resulting penalized Bayesian D-optimal design is

$$\xi^*_{PB} = \left\{ \begin{array}{cccc} 0.000 & 0.677 & 1.779 & 6.502 \\ 0.478 & 0.336 & 0.181 & 0.005 \end{array} \right\}. \tag{5-28}$$

It is observed that this design has similar characteristics of the Bayesian D-optimal four-point design for gamma prior distribution of θ_2 with the hiperparameters $\alpha = \beta = 1.5$. The D-efficiency of the design (5-28) with respect to the non-penalized Bayesian D-optimal four-point design given in Table **5-3** is equal to 0.99143. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion (5-11) evaluated for the Bayesian D-optimal four-point design and the penalized Bayesian D-optimal design (5-28) such that this ratio is in the interval $[0, 1]$. It is observed that the D-efficiency of the penalized Bayesian D-optimal design relative to the Bayesian D-optimal four-point design is greater than 99%, indicating the irrelevance of the loss in this efficiency.

The MSEs associated with the estimated parameters of the exponential model (5-6) are also calculated via simulation in order to compare the performance of non-penalized and penalized Bayesian D-optimal designs. The procedure of the efficient design apportionment (Pukelsheim 1993, p. 309) helps to obtain the distribution of observations for finite sample size 12 for both designs as follows: $5, 4, 2, 1$.

The algorithm implemented to calculate the MSE of the estimated parameters of the exponential model (5-6) is similar to the algorithm used in Section 4.7.1. The results

Table 5-6.: Results of the MSE calculations of the estimated parameters.

Parameter \ Design	Non-penalized	Penalized
θ_1	0.001876112	0.001933713
θ_2	0.058906095	0.027197434
$\|\boldsymbol{\theta}\|$	0.05154846	0.02136317

of the MSE calculations of the estimated parameters in non-penalized and penalized designs are presented in Table **5-6**.

It is observed that the MSE values of the parameter θ_1 are practically equal for both designs because this parameter is the linear term in the exponential model (5-6). The MSE values of the parameter θ_2 of the penalized Bayesian design are less than corresponding MSE values of the non-penalized Bayesian design. The non-penalized Bayesian four-point design was constructed for gamma prior distribution of θ_2 with the hyperparameters $\alpha = \beta = 1.5$, while the penalized design (5-28) was obtained for gamma prior distribution of θ_2 with the hyperparameters $\alpha = \beta = 3$. Therefore, the parameter θ_2 with the larger variance of the non-penalized design has the MSE values larger than the one with the smaller variance of the penalized design. Thus, greater uncertainty produces greater errors in parameter estimation of θ_2.

In summary, the penalized Bayesian D-optimal design (5-28) has the same characteristics as the non-penalized Bayesian D-optimal four-point design; it is equally efficient, and its MSEs of the parameter θ_2 are lower compared to the non-penalized design.

Two impractical characteristics of the experimental design (5-28) are evident. First, the maximum support point is outside the experimental region. As a result, it provides little information about the model curve. Second, this support point has only one observation.

A penalized-optimal design strategy is developed using five desirability functions $d_1(r_2)$, $d_2(\min(\text{diff}_{23}))$, $d_3(x_2)$, $d_4(x_4)$, and $d_5(r_4)$. The first three desirability functions do not change with respect to the previous design. But the fourth desirability function $d_4(x_4)$ is modified by changing the target of value range of x_4 and the width of the interval of this function.

Figure **5-6** shows that the model curve plateaus around 4, therefore, the maximum support point of the design (5-28) provides little information about this curve. Thus, it is

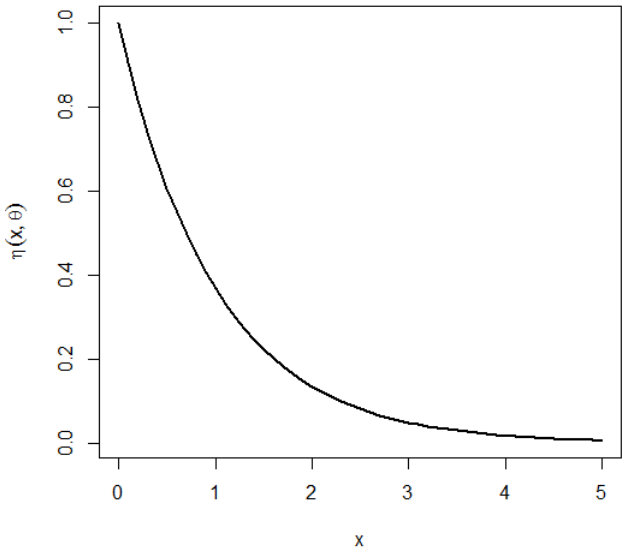

Figure 5-6.: Plot of the exponential regression model given in (5-6), where $\theta_1 = 1$ and $\theta_2 = 1$.

reasonable that a researcher would consider a maximum x_4 greater than 4 unacceptable. In order to keep the support point x_4 close to 3, it is assumed that the target of value range of x_4 is 3 and the interval width of desirability function is 1. Replacing the values of $\gamma_4 = 0.05$, $a_4 = 3$, and $\delta_4 = 1$ in the formulas (3-9) and (3-10), the desirability function $d_4(x_4)$ is obtained:

$$d_4(x_4) = \exp\left\{-\frac{1}{2}\left(\frac{x_4 - 3}{0.4085}\right)^2\right\}. \tag{5-29}$$

A plot of this desirability function, given in Figure **5-7**(a), shows that the location of the design point x_4 outside the interval $[2, 4]$ is unacceptable.

Considering the total sample size of 12 for a four-point design, it is reasonable to assume that at least one observation should be assigned to the new design point x_4. Replacing the values of $\gamma_5 = 0.05$, $r_{4(\min)} = 1$, $r_{4(\max)} = 3$ in the formulas (3-11), (3-12) and (3-13), the desirability functions $d_5(r_4)$ is obtained as:

$$d_5(r_4) = \frac{1}{1 + \exp\left(-\left(r_4 - 2\right)/0.34\right)}. \tag{5-30}$$

The plot of $d_5(r_4)$, given in Figure **5-7**(b), shows that the assignment of less than one observation to the support point x_4 is not acceptable.

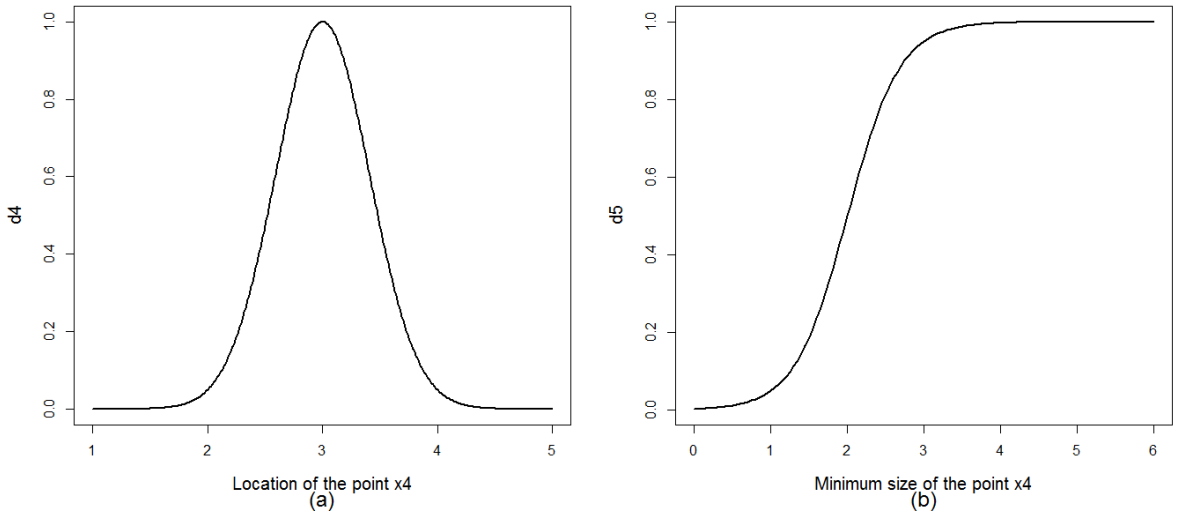

Figure 5-7.: Plots of desirability functions for the exponential model (5-6).

The overall desirability function is

$$D_3(\xi) = \left(d_1 \times d_2 \times d_3 \times d_4 \times d_5\right)^{1/5}. \tag{5-31}$$

The Nelder-Mead direct search algorithm (Nelder & Mead 1965) is employed in R-project (Team 2018) to minimize the following penalized criterion

$$\int_{\Theta} \left\{-\log\det\left\{I\left(\boldsymbol{\theta},\xi\right)\right\}\right\} \pi\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} + l\Lambda_0(1 - D_3(\xi)) \tag{5-32}$$

for the exponential model given in (5-6), where Λ_0 defines an absolute value of the minimum non-penalized Bayesian D-optimal criterion (5-20) of the two-point design (5-15). Computer code implemented in R-project (Team 2018) to generate penalized Bayesian D-optimal designs may be found in Appendix B. Procedure follows the steps, described in detail in Section 3.5, to obtain the penalized Bayesian D-optimal four-design with desired practical characteristics.

Penalized Bayesian D-optimal designs are generated by minimizing the penalized Bayesian D-optimal criterion (5-27) for different values $l = 0.01, 0.02, \ldots, 1$. The overall desirability function D_3 responses become stable from approximately $l = 0.30$. A resulting penalized Bayesian D-optimal design is

$$\xi^*_{PB} = \left\{ \begin{array}{cccc} 0.000 & 0.678 & 1.711 & 2.920 \\ 0.437 & 0.257 & 0.180 & 0.126 \\ 5 & 3 & 2 & 2 \end{array} \right\}, \tag{5-33}$$

where the last row represents the number of observations at each support point for the sample size $N = 12$. It is observed that this penalized optimal design has the desired practical characteristics since two observations are assigned to the new support point and the maximum support point, $x_4 = 2.920$, is within the experimental region. The D-efficiency of the design (5-33) with respect to the non-penalized Bayesian D-optimal four-point design given in Table **5-3** is equal to 0.99. It is observed that the D-efficiencies of the penalized Bayesian D-optimal designs, (5-28) and (5-33), relative to the Bayesian D-optimal four-point design are practically equal, indicating the irrelevance of the loss in this efficiency.

The MSEs of the estimated parameters are also calculated for the penalized design (5-33). The MSE value of the parameter θ_1 is equal to 0.001877552; it is approximately equal to the MSE value of the penalized design (5-28). It is logical because this parameter is the linear term in the exponential model (5-6). The MSE value of the parameter θ_2 is equal to 0.035074111; it is less than respective MSE value of the non-penalized Bayesian design, but it is greater than respective MSE value of the penalized Bayesian design (5-28).

In summary, the penalized Bayesian D-optimal design (5-33) has the desired practical characteristics, since two observations have been assigned to the maximum support point and also this point, $x_4 = 2.920$, is within the active experimental region. Although the MSE value of θ_2 is greater than in the penalized Bayesian design (5-28), both penalized Bayesian designs are almost equally efficient.

This example illustrates the methodology that allows the researcher to obtain the penalized Bayesian D-optimal design for nonlinear regression models. The penalized design technique proposed allows specifying particular characteristics of the experimental design without losing the efficiency of the design. If the optimal design theory leads from a practical perspective to an inappropriate design, this penalized technique may be used to improve the practical characteristics of the initial design. Thus, the resulting penalized Bayesian optimal design is optimal according to the corresponding Bayesian design criterion and fulfills the practical design preferences imposed by a researcher.

6. Conclusions, Recommendations, and Future Works

This last chapter provides general conclusions of the dissertation, suggests recommendations that should be taken into account in obtaining the penalized designs, and outlines possible lines of research for future work.

6.1. Conclusions

This research has suggested penalized optimal design strategy with the Bayesian approach to reduce issues related to experimental designs for linear and nonlinear models. The primary goal was to establish a procedure that would allow the construction of penalized Bayesian optimal experimental designs for these regression models.

A new optimality criterion was constructed with two simultaneous objectives: to incorporate the prior information of the unknown parameters and to satisfy practical design preferences imposed by a researcher. The proposed criterion combines the use of desirability functions and the Bayesian approach in the construction of penalized Bayesian D-optimal designs, which have good statistical inference properties and desirable practical characteristics. The justifications of the proposed criteria corresponding to each case were presented in their respective chapters.

The practical characteristics of experimental designs are delimited by a penalty function. This function includes an overall desirability function that is the geometric mean of the individual desirability functions. It is added to an "alphabetic" optimality criterion to penalize not practical optimal designs. The methodology of choosing the appropriate desirability functions according to the practical design preferences was proposed and illustrated with an example of the Michaelis-Menten model. The steps to follow to obtain the penalized optimal design were detailed in this example. The resulting penalized optimal design had a reasonable increment in the generalized variance but smaller compared to an arbitrarily chosen experimental design, and also there was an improvement in the practical design characteristics. This example additionally showed that resulting penalized designs were similar employing comparable shaped desirability functions.

The proposed penalized Bayesian optimal design methodology was illustrated by two examples of regression models: quadratic and exponential. The recommended steps to follow to obtain the penalized Bayesian optimal design for nonlinear models were detailed in Chapter 5. Corresponding routines were developed in the statistical software R (Team 2018) to calculate the penalized designs. It was shown that the D-efficiencies of the penalized Bayesian D-optimal designs relative to the Bayesian D-optimal designs were practically equal, indicating the irrelevance of the loss in this efficiency.

The mean squared errors (MSEs) of the parameters, which were calculated via simulation in order to compare the performance of non-penalized and penalized Bayesian optimal designs, showed that their values were comparable for linear terms. They are equal if the numbers of observations are equal in the corresponding support points. The lowest MSE value corresponds to the largest number of observations and vice versa. The MSE values of the scale term depend on the variance and skewness of the parameter prior distribution. They increase with the increase of these characteristics.

Finally, the proposed methodologies allowed the construction of penalized Bayesian D-optimal designs that provide a suitable balance between Bayesian D-optimality and overall desirability. In all examples, this proposed penalized methodology has provided an experimental design that was optimal according to the Bayesian criteria and with the design preferences established by the researcher.

6.2. Recommendations

The following recommendations are hereby made basing on the study findings. These recommendations relate to the issue of the penalized Bayesian design construction.

- The Bayesian design process involves the choice of a prior distribution for the statistical models. In the case of linear models, several values of the parameter correlation coefficients of normal prior distribution were considered for the quadratic model. It is recommended to check the combinations of these values because some combinations do not generate a positive definite "precision" matrix $\boldsymbol{R}$.

  In the case of nonlinear models, several hyperparameter values of gamma, lognormal, and uniform prior distributions of the parameter θ_2 were also examined for the exponential model to obtain the Bayesian optimal designs. Our findings suggest that the choice between gamma and lognormal prior distributions of θ_2 has little consequences for Bayesian D-optimal designs. However, uniform prior distribution with comparatively large variances produces very different Bayesian optimal designs due to the symmetry of this prior. If a researcher has poor information about

parameter mean values, then a prior distribution with comparatively large variance should be chosen. In contrast, if reasonable prior information is available about them, Bayesian designs based on this kind of prior distributions may direct to a loss of design efficiency. On the other hand, it must be taken into account that the number of support points for nonlinear models increases as the uncertainty of the prior distribution increases (Firth & Hinde 1997). Therefore, the prior distribution must be chosen with great prudence before implementing the Bayesian design.

- The proposed penalized optimal design criterion includes two terms, where the first term represents an "alphabetical" optimality criterion, and the second term is a penalty function that represents the constraints applied to impractical optimal designs. A parameter, Λ, specified by the researcher, manages the penalty and optimality contributions in the penalized optimal criterion. The procedure for selecting Λ in the penalized optimal design process is presented in Sections 3.5 and 5.4. The initial value of Λ is suggested as an absolute value of the minimum of the corresponding non-penalized optimality criterion. It is recommended to choose the final value of Λ in the range corresponding to the stability exposed in the overall desirability function responses, which can be plotted to observe their behavior better. It is emphasized that because resulting penalized optimal designs are similar in the range of Λ that produces stable responses of the overall desirability, it is not important to know the exact final value of Λ.

- If regression model under study is not linear, the first term of the penalized Bayesian optimality criterion has quadrature form. The non-Bayesian D-criterion is integrated over the parameter prior distribution, thereby giving more weight to more probable parameter values and less weight to improbable values. Random variable values are generated according to corresponding prior distribution, and then the MC method is used to calculate this integral. It is recommended to generate more than 1000 parameter values to guarantee the best fit to the integral.

- The proposed penalized design procedure permits the researcher to define practical characteristics of the experimental design through individual desirability functions. The methodology of choosing the appropriate desirability functions according to the practical design preferences is presented in Section 3.4. After defining the analytical expressions of these desirability functions, it is recommended to plot them to verify desirability levels of design characteristic subject to restriction; and then, it is suggested to use them in the proposed strategy. This allows for avoiding unpleasant errors when applying them in the penalty procedure.

6.3. Future Works

In light of this dissertation, the following lines of research are proposed.

- Our investigation was oriented to the study of the Bayesian D-optimality criterion. However, our proposed penalized design process could also be used for all other design criteria, such as Bayesian A-optimality. The penalized Bayesian A-optimal design could, therefore, be constructed in an analogous manner for both linear and nonlinear models.

- The Bayesian approach is appropriate for the sequential design process. It uses currently available data to choose the next design points, and can be divided into two categories: (i) sequential non-Bayesian approach where at each stage the local optimal design is computed at the current estimations of the parameters (Ford & Silvey 1980), and (ii) sequential Bayesian approach where at each stage the optimal design is obtained by optimizing the Bayesian optimality criterion over an updated prior distribution for the parameters (Roy, Ghosal & Rosenberger 2009).

  In our research, we have only studied the non-sequential Bayesian optimal designs. In some cases with poor prior information, sequential design can improve design efficiency with additional information introduced consecutively in the design construction process (Zhang 2006). Thus, the penalized Bayesian two-stage design procedure could be used to increase initial design efficiency.

- It is proposed to extend the developed penalized design methodology to generalized linear models, particularly, to the logistic and Poisson regression models. These models are widely used to model binary and count data in biological, pharmacological, and medical research (Bender 2009).

- It is interesting to incorporate the proposed penalized design methodology into mixed-effects models, where the applied restrictions could improve the practical characteristics of the optimal designs for these models.

- In this research, it was assumed that the error variance is constant. However, in practice, there are cases where this assumption is not satisfied necessarily. Therefore, it is proposed to extend the proposed penalized design methodology for heteroscedastic models.

- In this research, we have considered penalized experimental designs for parameter estimation. This work could be extended for model discrimination between two or more rival models.

Appendices

A. Deduction of the formula (5-9)

The Fisher information for $\boldsymbol{\theta}$ in the exponential regression model (5-6) for a single observation at $x \geq 0$ is given by the matrix

$$\boldsymbol{I}(x, \boldsymbol{\theta}) = \left[\frac{\partial}{\partial \boldsymbol{\theta}} \eta(x; \boldsymbol{\theta})\right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} \eta(x; \boldsymbol{\theta})\right]^T = \begin{bmatrix} \exp(-2\theta_2 x) & -\theta_1 x \exp(-2\theta_2 x) \\ -\theta_1 x \exp(-2\theta_2 x) & \theta_1^2 x^2 \exp(-2\theta_2 x) \end{bmatrix}, \tag{A-1}$$

where $\frac{\partial}{\partial \boldsymbol{\theta}} \eta(x; \boldsymbol{\theta})$ denotes the gradient of $\eta(x; \boldsymbol{\theta})$.

The information matrix of an approximate design ξ is

$$\boldsymbol{I}(\xi, \boldsymbol{\theta}) = \int_0^\infty \boldsymbol{I}(x, \boldsymbol{\theta})\, d\xi(x). \tag{A-2}$$

For the model (5-6) the calculation shows

$$\begin{aligned}
\mathrm{E}_\pi \left[\log\det\left\{\boldsymbol{I}(\xi, \boldsymbol{\theta})\right\}\right] &= \mathrm{E}_\pi \left[\log\det\left\{\int_0^\infty \boldsymbol{I}(x, \boldsymbol{\theta})\, d\xi(x)\right\}\right] \\
&= \mathrm{E}_\pi \left[\log\det\left\{\int_0^\infty \begin{bmatrix} 1 & -\theta_1 x \\ -\theta_1 x & \theta_1^2 x^2 \end{bmatrix} \exp(-2\theta_2 x)\, d\xi(x)\right\}\right] \\
&= \mathrm{E}_\pi \left[\log\det\begin{bmatrix} \int_0^\infty e^{-2\theta_2 x} d\xi(x) & -\theta_1 \int_0^\infty x e^{-2\theta_2 x} d\xi(x) \\ -\theta_1 \int_0^\infty x e^{-2\theta_2 x} d\xi(x) & \theta_1^2 \int_0^\infty x^2 e^{-2\theta_2 x} d\xi(x) \end{bmatrix}\right] \\
&= \mathrm{E}_\pi \left[\log\left\{\theta_1^2 \det\begin{bmatrix} \int_0^\infty e^{-2\theta_2 x} d\xi(x) & -\int_0^\infty x e^{-2\theta_2 x} d\xi(x) \\ -\int_0^\infty x e^{-2\theta_2 x} d\xi(x) & \int_0^\infty x^2 e^{-2\theta_2 x} d\xi(x) \end{bmatrix}\right\}\right] \\
&= \mathrm{E}_\pi \left[\log\left\{\theta_1^2 \det\left(\int_0^\infty \begin{bmatrix} 1 & -x \\ -x & x^2 \end{bmatrix} \exp(-2\theta_2 x)\, d\xi(x)\right)\right\}\right] \\
&= \mathrm{E}_\pi \left[\log\left\{\theta_1^2 \det\left(\int_0^\infty \boldsymbol{I}(x, \boldsymbol{\theta}^*)\, d\xi(x)\right)\right\}\right] \\
&= \mathrm{E}_\pi \left[\log\left\{\theta_1^2 \det\left\{\boldsymbol{I}(\xi, \boldsymbol{\theta}^*)\right\}\right\}\right] \\
&= 2\mathrm{E}_\pi \left[\log(\theta_1)\right] + \mathrm{E}_\pi \left[\log\det\left\{\boldsymbol{I}(\xi, \boldsymbol{\theta}^*)\right\}\right].
\end{aligned}$$

Therefore, it follows that

$$\mathrm{E}_\pi \left[\log\det\left\{\boldsymbol{I}(\xi, \boldsymbol{\theta})\right\}\right] = 2\mathrm{E}_\pi \left[\log(\theta_1)\right] + \mathrm{E}_\pi \left[\log\det\left\{\boldsymbol{I}(\xi, \boldsymbol{\theta}^*)\right\}\right], \tag{A-3}$$

where $\boldsymbol{\theta}^* = (1, \theta_2)^T$.

B. Deduction of the posterior distribution of θ in the normal linear regression model

Consider the linear regression model

$$\mathrm{E}\left(\boldsymbol{Y}|\boldsymbol{\theta},\sigma^2\right)=\boldsymbol{X\theta}, \tag{B-1}$$

where $\boldsymbol{Y}|\boldsymbol{\theta},\sigma^2 \sim \mathcal{N}\left(\boldsymbol{X\theta},\sigma^2\boldsymbol{I}\right)$ and the conditional prior distribution of $\boldsymbol{\theta}$ given σ^2 is $\mathcal{N}\left(\boldsymbol{\mu},\sigma^2\boldsymbol{R}^{-1}\right)$ with $\boldsymbol{\theta}\in\mathbb{R}^p$, where $\boldsymbol{R}$ is a given positive definite $p\times p$ "precision" matrix, i.e.,

$$\pi(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)^T\boldsymbol{R}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)\right]. \tag{B-2}$$

The likelihood probability density function is

$$L(\boldsymbol{Y}|\boldsymbol{\theta},\xi) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)\right]. \tag{B-3}$$

In order to calculate the posterior distribution of $\boldsymbol{\theta}$, we use Bayes theorem

$$\begin{aligned}\pi\left(\boldsymbol{\theta}|\,\boldsymbol{Y},\xi\right) &\propto \pi(\boldsymbol{\theta})\times L(\boldsymbol{Y}|\boldsymbol{\theta},\xi)\\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)^T\boldsymbol{R}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)+\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)\right]\right\}\\ &\propto \exp\left\{-\frac{\boldsymbol{A}}{2\sigma^2}\right\},\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{A} &= \left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)^T\boldsymbol{R}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)+\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)^T\left(\boldsymbol{Y}-\boldsymbol{X\theta}\right)\\ &= \boldsymbol{\theta}^T\boldsymbol{R\theta}-\boldsymbol{\theta}^T\boldsymbol{R\mu}-\boldsymbol{\mu}^T\boldsymbol{R\theta}+\boldsymbol{\mu}^T\boldsymbol{R\mu}+\boldsymbol{Y}^T\boldsymbol{Y}-\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{Y}-\boldsymbol{Y}^T\boldsymbol{X\theta}+\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X\theta}\\ &= \boldsymbol{\theta}^T\left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)\boldsymbol{\theta}-\boldsymbol{\theta}^T\left(\boldsymbol{X}^T\boldsymbol{Y}+\boldsymbol{R\mu}\right)-\left(\boldsymbol{Y}^T\boldsymbol{X}+\boldsymbol{\mu}^T\boldsymbol{R}\right)\boldsymbol{\theta}+\boldsymbol{Y}^T\boldsymbol{Y}+\boldsymbol{\mu}^T\boldsymbol{R\mu}.\end{aligned}$$

For convenience, let $\boldsymbol{V}=\left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)^{-1}$ a $p\times p$ matrix and $\boldsymbol{\theta}^*=\left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{Y}+\boldsymbol{R\mu}\right)$ a $p\times 1$ vector. Hence,

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{\theta}^T\boldsymbol{V}^{-1}\boldsymbol{\theta}-\boldsymbol{\theta}^T\boldsymbol{V}^{-1}\boldsymbol{\theta}^*-\boldsymbol{\theta}^{*T}\boldsymbol{V}^{-1}\boldsymbol{\theta}+\boldsymbol{Y}^T\boldsymbol{Y}+\boldsymbol{\mu}^T\boldsymbol{R\mu}\\ &= \left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)^T\boldsymbol{V}^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)-\boldsymbol{\theta}^{*T}\boldsymbol{V}^{-1}\boldsymbol{\theta}^*+\boldsymbol{Y}^T\boldsymbol{Y}+\boldsymbol{\mu}^T\boldsymbol{R\mu}.\end{aligned}$$

Finally, dropping terms that do not involve $\boldsymbol{\theta}$, the posterior distribution for $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\pi\left(\boldsymbol{\theta}|\,\boldsymbol{Y},\xi\right) &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)^T\,\boldsymbol{V}^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)^T\left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)\left(\boldsymbol{\theta}-\boldsymbol{\theta}^*\right)\right]\right\},\end{aligned} \tag{B-4}$$

where

$$\boldsymbol{\theta}^* = \left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)^{-1}\left(\boldsymbol{X}^T\,\boldsymbol{Y}+\boldsymbol{R}\boldsymbol{\mu}\right).$$

Therefore, the equation (B-4) shows that the posterior distribution $\pi\left(\boldsymbol{\theta}|\,\boldsymbol{Y},\xi\right)$ of $\boldsymbol{\theta}$ is proportional to the kernel of the $\mathcal{N}\left(\boldsymbol{\theta}^*,\sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}\right)^{-1}\right)$.

C. Computer codes implemented in R for obtaining the penalized designs

Example: Michaelis-Menten Model in Section 3.5

```
## Penalized D-optimal designs for Michaelis-Menten model
# The model is theta1*x/(theta2+x)
require(nloptr)

# 1. The function gradients for Michaelis-Menten model are defined as
fx <- function(t)
{
f1 <- t/(theta[2]+t)                 # derivative with respect to theta1
f2 <- -theta[1]*t/((theta[2]+t)^2)   # derivative with respect to theta2
return(as.matrix(c(f1,f2),ncol=1))
}

# 2. Function that calculates the information matrix M(xi)
Mom <-function(xi)
{
x <- as.matrix(xi[1,]); w <- xi[2,]  # points and weights of design xi
X <- t(apply(x,1,fx)); W <- diag(w)
Mom <- t(X)%*%W%*%X
return(Mom)
}

# 3. Function that calculates the penalty function
L=(12.04841)  # Absolute value of
              # the minimum non-penalized D-optimal criterion
k<-0.20       # 0.35 for D4
Pen<-function(xi)
{
p1<-xi[2,1];x3<-xi[1,3];x2<-xi[1,2];x1<-xi[1,1]
diff<-abs(x2-x1); n1<-round((20*p1),0)
 if(n1>10) d1=1 else
```

```
 if(n1<6) d1=0 else
d1<-((n1-6)/4)^(4)                    # desirability function d1
d2<-1-exp(-exp(9.25-6.15*x3))         # desirability function d2
d3<-1/(1+exp(-(diff-0.1)/0.03))       # desirability function d3
#d4<-1/(1+exp(-(n1-9)/0.34))          # desirability function d4
#d5<-((3-x3)/3)^(2)                   # desirability function d5
D1<-(d1*d2*d3)^(1/3)                  # overall desirability D1
#D2<-(d4*d2*d3)^(1/3)                 # overall desirability D2
#D3<-(d1*d5*d3)^(1/3)                 # overall desirability D3
#D4<-(d4*d5*d3)^(1/3)                 # overall desirability D4
Pen1<-L*(1-D1)
return(Pen1)
}

# 4. The functional of the information matrix is defined in (3-15) as
phi<-function(xi)
{
M=Mom(xi); Psi<-(-log(det(M)))+k*Pen(xi)
return(Psi)
}

# 5. The appropriate function to minimize is defined as
G<-function(dis)
{
ss<-length(dis); m<-ss/2
xi<-matrix(c(dis[1:m],dis[(m+1):ss]/sum(dis[(m+1):ss])),ncol=m,byrow=T)
au<-phi(xi)
return(au)
}

# 6. The initial values of the start design are given as
xi0<-c(0,0.1,1,0.3333,0.3333,0.3333)
ssm=length(xi0); m=ssm/2
# The local values of the parameters are
theta=c(212.68,0.064)      # theta=(theta1,theta2)

# 7. The process of minimization restricted to the regression range
#    is performed by Nelder-Mead direct search algorithm
resp=neldermead(xi0,G,lower=c(rep(0,3),rep(0.01,3)),
                      upper=c(0,rep(Inf,2),rep(0.99,3)))
```

```
# 8. Penalized D-optimal design is
soportep<-round(c(resp$par[1:3]),3)
pesosp<-c(resp$par[4:6]/sum(resp$par[4:6]))
n=20; obsp<-round(n*pesosp,0)
xioptimopen<-matrix(c(soportep,obsp),ncol=3,byrow=T)
xioptimopen
```

Example: Quadratic Model in Section 4.5

```
## Penalized Bayesian D-optimal designs for quadratic model
require(nloptr)

# Bayesian two-point design

#1. Function f(x) for quadratic model is defined as
fx <- function(x) as.matrix(c(1,x,x^2))

#2. Function that calculates the information matrix
Mom <- function(xi)
{
x <- as.matrix(xi[1,]); w <- xi[2,]     # points and weights of design xi
X <- t(apply(x,1,fx)); W <- diag(w)
Mom <- t(X)%*%W%*%X
return(Mom)
}

#3. Function that calculates the "precision" matrix R
n<-9   # sample size
var1<-3; var2<-5; var3<-1
ro12<- 0.5; ro13<- 0.9; ro23<- 0.5
cov12<-ro12*sqrt(var1*var2)
cov13<-ro13*sqrt(var1*var3)
cov23<-ro23*sqrt(var2*var3)
V<-matrix(c(var1,cov12,cov13,cov12,var2,cov23,cov13,cov23,var3),ncol=3)
R<-solve(V)

#4. Function that calculates the penalty function
L=(0.3794315)     # Absolute value of
                  # the minimum non-penalized D-optimal Bayesian criterion
k<-0.75           # submultiple of L
```

```
Pen<-function(xi)
{
x1<-xi[1,1]; x2<-xi[1,2]; x3<-xi[1,3]
p2<-xi[2,2]; n2<-round((9*p2),0)
d1<-1/(1+exp(-(n2-2.5)/0.17))      # desirability function d1
d2<-exp(-((x2/0.082)^2)/2)         # desirability function d2
D<-sqrt(d1*d2)                     # overall desirability D
Pen1<-L*(1-D)                      # penalty function
return(Pen1)
}

#5. The functional of the information matrix is defined in (4-16) as
phi <- function(xi) -log(det(Mom(xi)+R/n))+k*Pen(xi)

#6. The appropriate function to minimize is defined as
G <- function(dis)
{
ss <- length(dis); s <- ss/2
xi <- matrix(c(dis[1:s],dis[(s+1):ss]/sum(dis[(s+1):ss])),ncol=s,byrow=T)
au <- phi(xi)
return(au)
}

# 7. The initial values of the start design are given as
x0<-c(-1,0,0.9,0.4,0.3,0.3)
ss <-length(x0); s <-ss/2

# 8. The process of minimization restricted to the regression range
#    is performed by Nelder-Mead direct search algorithm
resBp<-neldermead(x0,G,lower=c(rep(-1,s),rep(0,s)),upper=c(rep(1,ss)))

# 9. Penalized Bayesian D-optimal design is
soportexiBp<- round(resBp$par[1:s],3)
pesosxiBp <- round(resBp$par[(s+1):ss]/sum(resBp$par[(s+1):ss]),3)
DesignBp<-rbind(soportexiBp,pesosxiBp)
n=9; obsBp<-round(n*pesosxiBp,0)
DesignBp; obsBp
```

```
# Bayesian three-point design

#1. Function f(x) for quadratic model is defined as
fx <- function(x) as.matrix(c(1,x,x^2))

#2. Function that calculates the information matrix
Mom <- function(xi)
{
x <- as.matrix(xi[1,]); w <- xi[2,]         # points and weights of design xi
X <- t(apply(x,1,fx)); W <- diag(w)
Mom <- t(X)%*%W%*%X
return(Mom)
}

#3. Function that calculates the "precision" matrix R
n<-9   # sample size
var1<-3; var2<-5; var3<-1
ro12<- 0; ro13<- 0; ro23<- 0
cov12<-ro12*sqrt(var1*var2); cov13<-ro13*sqrt(var1*var3)
cov23<-ro23*sqrt(var2*var3)
V<-matrix(c(var1,cov12,cov13,cov12,var2,cov23,cov13,cov23,var3),ncol=3)
R<-solve(V)

#4. Function that calculates the penalty function
L=(1.365036)      # Absolute value of
                  # the minimum non-penalized D-optimal Bayesian criterion
k<-0.4            # submultiple of L
Pen<-function(xi)
{
x1<-xi[1,1];x2<-xi[1,2];x3<-xi[1,3];x4<-xi[1,4]
p2<-xi[2,2];n2<-round((9*p2),0)
p3<-xi[2,3];n3<-round((9*p3),0)
diff2<-abs(x3-x2);diff3<-abs(x4-x3)
d1<-1/(1+exp(-(n3-1.5)/0.17))           # desirability function d1
d2<-1/(1+exp(-(diff2-0.2)/0.034))       # desirability function d2
d3<-1/(1+exp(-(diff3-0.2)/0.034))       # desirability function d3
D<-(d1*d2*d3)^(1/3)                     # overall desirability D
Pen1<-L*(1-D)                           # penalty function
return(Pen1)
}
```

```
#5. The functional of the information matrix is defined in (4-22) as
phi <- function(xi) -log(det(Mom(xi)+R/N))+k*Pen(xi)

#6. The appropriate function to minimize is defined as
G <- function(dis)
{
ss <- length(dis); s <- ss/2
xi <- matrix(c(dis[1:s],dis[(s+1):ss]/sum(dis[(s+1):ss])),ncol=s,byrow=T)
au <- phi(xi)
return(au)
}

# 7. The initial values of the start design are given as
x0<-c(-1,0,0.3,0.9,0.2,0.3,0.3,0.2)
ss <-length(x0); s <-ss/2

# 8. The process of minimization restricted to the regression range
#    is performed by Nelder-Mead direct search algorithm
resBp<-neldermead(x0,G,lower=c(-1,0,-1,-1,rep(0.01,s)),
                       upper=c(1,0,1,1,rep(1,s)))

# 9. Penalized Bayesian D-optimal design is
soportexiBp<- round(resBp$par[1:s],3)
pesosxiBp <- round(resBp$par[(s+1):ss]/sum(resBp$par[(s+1):ss]),3)
DesignBp<-rbind(soportexiBp,pesosxiBp)
n=9; obsBp<-round(n*pesosxiBp,0)
DesignBp; obsBp
```

Example: Exponential Growth Model in Section 5.5

```
## Penalized Bayesian D-optimal designs for exponential model
# Model is theta1*exp(-theta2*x)
# theta=(theta1,theta2)
require(nloptr)

# Bayesian two-point design

# 1. The function gradients for exponential model are defined as
fx <- function(t,theta1,theta2)
{
f1 <- exp(-theta2*t)                    # derivative with respect to theta1
f2 <- -(theta1*t*exp(-theta2*t))        # derivative with respect to theta2
return(as.matrix(c(f1,f2),ncol=1))
}
# 2. Function that calculates the information matrix M(xi,theta1,theta2)
Mom <-function(xi,theta1,theta2)
{
x <- as.matrix(xi[1,]); w <- xi[2,]     # points and weights of design xi
X <- t(apply(x,1,fx,theta1,theta2))
W <- diag(w); Mom <- t(X)%*%W%*%X
return(Mom)
}

# 3. The simulated sample of prior distributions of theta is imported
theta<-read.table("C:/MuestrasTes/MuestraGA-2p.txt")    # with Gamma(4,4)
#theta<-read.table("C:/MuestrasTes/MuestraLN-2p.txt")   # with Lognormal
#theta<-read.table("C:/MuestrasTes/MuestraU-2p.txt")    # with Uniforme
theta<-as.matrix(theta)
theta1<-theta[,1]
theta2<-theta[,2]

# 4. Function that calculates the penalty function
# for Gamma(4,4):
L=3.401431        # Absolute value of
                  # the minimum non-penalized D-optimal Bayesian criterion
k<-0.1            # submultiple of L

#L=3.360336; k<-0.2   # for Lognormal(1,0.25)
#L=3.411332; k<-0.3   # for Uniforme(0.134,1.866)
```

```
Pen<-function(xi)
{
x1<-xi[1,1];x2<-xi[1,2];x3<-xi[1,3]
diff23<-abs(x3-x2)
p3<-xi[2,3];n3<-round((12*p3),0)
p2<-xi[2,2];n2<-round((12*p2),0)
d1<-1/(1+exp(-(n3-1.5)/0.17))           # desirability function d1
d2<-1/(1+exp(-(diff23-0.2)/0.034))      # desirability function d2
D1<-(d1*d2)^(1/2)                       # overall desirability D1
Pen1<-L*(1-D1)
return(Pen1)
}

# 5. Function that calculates the utility function (5-19)
phi<-function(xi,theta1,theta2)
{
 N<-length(theta1)
 j<-0
 detM<-integer(N)
 for(j in 1:N){
  detM[j]<-det(Mom(xi,theta1[j],theta2[j]))
  j<-j+1
  }
phi1<-c(log(detM))
phi1<- -mean(phi1)+k*Pen(xi)   # Monte Carlo
return(phi1)
}

# 6. The appropriate function to minimize is defined as
G<-function(dis)
{
ss<-length(dis); m<-ss/2
xi<-matrix(c(dis[1:m],dis[(m+1):ss]/sum(dis[(m+1):ss])),ncol=m,byrow=T)
au<-phi(xi,theta1,theta2)
return(au)
}

# 7. The initial values of the start design are given as
xi0<-c(0,1,2,0.4,0.3,0.3)
ss=length(xi0); m=ss/2
```

```
# 8. The process of minimization restricted to the regression range
# is performed by Nelder-Mead direct search algorithm
set.seed(1234)
res3Bp=neldermead(xi0,G,lower=c(rep(0,3),rep(0.01,3)),
                        upper=c(rep(Inf,3),rep(0.99,3)))

# 9. Penalized Bayesian D-optimal design is
soporte3Bp<-res3Bp$par[1:3]
pesos3Bp<-res3Bp$par[4:6]/sum(res3Bp$par[4:6])
xioptimo3Bp<-round(matrix(c(soporte3Bp,pesos3Bp),ncol=3,byrow=T),3)
xioptimo3Bp



# Penalized Designs with Four Support Points

# 1. The function gradients for exponential model are defined as
fx <- function(t,theta1,theta2)
{
f1 <- exp(-theta2*t)                      # derivative with respect to theta1
f2 <- -(theta1*t*exp(-theta2*t))          # derivative with respect to theta2
return(as.matrix(c(f1,f2),ncol=1))
}

# 2. Function that calculates the information matrix M(xi,theta1,theta2)
Mom <-function(xi,theta1,theta2)
{
x <- as.matrix(xi[1,]); w <- xi[2,]       # points and weights of design xi
X <- t(apply(x,1,fx,theta1,theta2))
W <- diag(w); Mom <- t(X)%*%W%*%X
return(Mom)
}

# 3. The simulated sample of prior distributions of theta is imported
theta<-read.table("C:/MuestrasTes/MuestraGA-2pb.txt")  # with Gamma(3,3)
theta<-as.matrix(theta)
theta1<-theta[,1]
theta2<-theta[,2]
```

```
# 4. Function that calculates the penalty function
L=(3.37449)       # Absolute value of
                  # the minimum non-penalized D-optimal Bayesian criterion
k<-0.97           # submultiple of L
# k=0.30 para D3
Pen<-function(xi)
{
x1<-xi[1,1];x2<-xi[1,2];x3<-xi[1,3];x4<-xi[1,4]
diff2<-abs(x3-x2)
diff3<-abs(x4-x3)
p4<-xi[2,4];n4<-round((12*p4),0)
p3<-xi[2,3];n3<-round((12*p3),0)
p2<-xi[2,2];n2<-round((12*p2),0)
d1<-1/(1+exp(-(n2-2.5)/0.17))           # desirability function d1
d2<-1/(1+exp(-(diff2-0.9)/0.034))       # desirability function d2
d3<-exp(-(1/2)*((x2-0.675)/0.0817)^2)   # desirability function d3
d4<-exp(-(1/2)*((x4-6.5)/0.817)^2)      # desirability function d4
#d42<-exp(-(1/2)*((x4-3)/0.4085)^2)     # desirability function d42
#d5<-1/(1+exp(-(n4-2)/0.34))            # desirability function d5
D2<-(d1*d2*d3*d4)^(1/4)                 # overall desirability D2
#D3<-(d1*d2*d3*d42*d5)^(1/5)            # overall desirability D3
Pen1<-L*(1-D2)
return(Pen1)
}

# 5. Function that calculates the utility function (5-27) o (5-32)
phi<-function(xi,theta1,theta2)
{
 N<-length(theta1)
 j<-0
 detM<-integer(N)
 for(j in 1:N){
  detM[j]<-det(Mom(xi,theta1[j],theta2[j]))
  }
phi1<-c(log(detM))
phi1<- -mean(phi1)+k*Pen(xi)   # Monte Carlo
return(phi1)
}
```

```
# 6. The appropriate function to minimize is defined as
G<-function(dis)
{
ss<-length(dis); m<-ss/2
xi<-matrix(c(dis[1:m],dis[(m+1):ss]/sum(dis[(m+1):ss])),ncol=m,byrow=T)
au<-phi(xi,theta1,theta2)
return(au)
}

# 7. The initial values of the start design are given as
xi0<-c(0,1,1.5,6,0.3,0.2,0.3,0.2)
ss=length(xi0); m=ss/2

# 8. The process of minimization restricted to the regression range
# is performed by Nelder-Mead direct search algorithm
set.seed(1234)
res4Bp=neldermead(xi0,G,lower=c(rep(0,4),rep(0.01,4)),
                       upper=c(rep(Inf,4),rep(0.99,4)))

# 9. Penalized Bayesian D-optimal design is
soporte4Bp<-res4Bp$par[1:4]
pesos4Bp<-res4Bp$par[5:8]/sum(res4Bp$par[5:8])
xioptimo4Bp<-matrix(c(soporte4Bp,pesos4Bp),ncol=4,byrow=T)
n<-12; obs4Bp<-round(n*pesos4Bp,0)
xioptimo4Bp; obs4Bp
nu<-10;obs4ef<-ceiling(nu*pesos4Bp);obs4ef
```

Bibliography

Albert, J. (2009), *Bayesian computation with R*, Springer Science & Business Media.

Atkinson, A., Bogacka, B. & Zocchi, S. (2000), 'Equivalence theory for design augmentation and parsimonious model checking: response surfaces and yield density models', *Listy Biometryczne-Biometrical Letters* **37**(2), 67–95.

Atkinson, A., Donev, A. & Tobias, R. (2007), *Optimum experimental designs, with SAS*, Oxford Univ. Press, UK.

Avriel, M. (2003), *Nonlinear programming: analysis and methods*, Courier Corporation.

Bates, D. M. & Watts, D. G. (1988), *Nonlinear regression analysis and its applications*, John Wiles & Sons, Inc.

Becerra Rodríguez, M. B., Zitzumbo Guzmán, R., Domínguez Domínguez, J., García Alcaraz, J. L. & Alonso Romero, S. (2014), 'Aplicación de la función de deseabilidad para optimizar un producto vulcanizado', *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia* **37**(1), 85–94.

Bender, R. (2009), Introduction to the use of regression models in epidemiology, *in* 'Cancer Epidemiology', Springer, pp. 179–195.

Bernardo, J. M. (1979), 'Expected information as expected utility', *The Annals of Statistics* pp. 686–690.

Borchers, H. W. (2018), 'Package 'adagio': Discrete and global optimization routines', Available: https://CRAN.R-project.org/package=adagio.

Cardona, J., López, V. I. & Correa, J. C. (2012), 'Diseños óptimos bayesianos para estimación de parámetros en farmacocinética', *Comunicaciones en Estadística* **5**(1), 97–112.

Chaloner, K. & Larntz, K. (1989), 'Optimal bayesian design applied to logistic regression experiments', *Journal of Statistical Planning and Inference* **21**(2), 191–208.

Chaloner, K. & Verdinelli, I. (1995), 'Bayesian experimental design: A review', *Statistical Science* pp. 273–304.

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

Cook, D. & Fedorov, V. (1995), 'Invited discussion paper constrained optimization of experimental design', *Statistics* **26**(2), 129–148.

DasGupta, A. & Studden, W. (1991), 'Robust bayesian experimental designs in normal linear models', *The Annals of Statistics* pp. 1244–1256.

Del Castillo, E., Montgomery, D. & McCarville, D. (1996), 'Modified desirability functions for multiples response optimization', *Journal of quality technology* **28**(3), 337–345.

Derringer, G. C. (1994), 'A balancing act-optimizing a products properties', *Quality Progress* **27**(6), 51–58.

Derringer, G. & Suich, R. (1980), 'Simultaneous optimization of several response variables', *Journal of quality technology* **12**, 214–219.

Dette, H. & Biedermann, S. (2003), 'Robust and efficient designs for the michaelis–menten model', *Journal of the American Statistical Association* **98**(463), 679–686.

Dette, H., Kiss, C. & Wong, W. K. (2008), Robustness of optimal designs for the michaelis-menten model under a variation of criteria, Technical report, Technical Report//Sonderforschungsbereich 475, Komplexitätsreduktion in â.

Dette, H. & Kunert, J. (2014), 'Optimal designs for the michaelis–menten model with correlated observations', *Statistics* **48**(6), 1254–1267.

Dette, H. & Neugebauer, H.-M. (1997), 'Bayesian d-optimal designs for exponential regression models', *Journal of Statistical Planning and Inference* **60**(2), 331–349.

Dette, H. & Sperlich, S. (1994), 'A note on bayesian d-optimal designs for a generalization of the exponential growth model', *South African Statist. J* **28**, 103–117.

Dragalin, V. & Fedorov, V. (2006), 'Adaptive designs for dose-finding based on efficacy–toxicity response', *Journal of Statistical Planning and Inference* **136**(6), 1800–1823.

Dragalin, V., Fedorov, V. & Wu, Y. (2008), 'Adaptive designs for selecting drug combinations based on efficacy–toxicity response', *Journal of Statistical Planning and Inference* **138**(2), 352–373.

Duarte, B. P. & Wong, W. K. (2015), 'Finding bayesian optimal designs for nonlinear models: a semidefinite programming-based approach', *International Statistical Review* **83**(2), 239–262.

El-Krunz, S. M. & Studden, W. J. (1991), 'Bayesian optimal designs for linear regression models', *The Annals of Statistics* pp. 2183–2208.

Elfving, G. et al. (1952), 'Optimum allocation in linear regression theory', *The Annals of Mathematical Statistics* **23**(2), 255–262.

Ermakov, S. M. & Zhiglijavsky, A. A. (1987), *The Mathematical Theory of Optimum Experiments*, Nauka, Moscow. (In Russian).

Fedorov, V. V. & Hackl, P. (1997), *Model-oriented design of experiments*, Berlin.

Fedorov, V. V. & Leonov, S. L. (2013), *Optimal design for nonlinear response models*, CRC Press.

Firth, D. & Hinde, J. (1997), 'On bayesian d-optimum design criteria and the equivalence theorem in non-linear models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(4), 793–797.

Ford, I. & Silvey, S. (1980), 'A sequentially constructed design for estimating a nonlinear parametric function', *Biometrika* **67**(2), 381–388.

Gao, L. & Rosenberger, W. F. (2013), Adaptive bayesian design with penalty based on toxicity-efficacy response, *in* D. Ucinski, A. C. Atkinson & M. Patan, eds, 'mODa 10–Advances in Model-Oriented Design and Analysis', Springer, Heidelberg, pp. 91–98.

Gavin, H. P. & Scruggs, J. T. (2012), 'Constrained optimization using lagrange multipliers', *CEE 201L. Duke University* .

George, P. & Ogot, M. M. (2006), 'A compromise experimental design method for parametric polynomial response surface approximations', *Journal of Applied Statistics* **33**(10), 1037–1050.

Gibb, R. D. (1998), Optimal treatment combination estimation for univariate and multivariate response surface applications, PhD thesis, Virginia Commonwealth University.

Haines, L. M., Perevozskaya, I. & Rosenberger, W. F. (2003), 'Bayesian optimal designs for phase i clinical trials', *Biometrics* **59**(3), 591–600.

Hamada, M., Martz, H., Reese, C. & Wilson, A. (2001), 'Finding near-optimal bayesian experimental designs via genetic algorithms', *The American Statistician* **55**(3), 175–181.

Harrington, E. (1965), 'The desirability function', *Industrial quality control* **21**(10), 494–498.

Holland, J. H. (1992), *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, The MIT press, USA.

Kiefer, J. (1959), 'Optimum experimental designs', *Journal of the Royal Statistical Society: Series B (Methodological)* **21**(2), 272–304.

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

Kiefer, J. & Wolfowitz, J. (1960), 'The equivalence of two extremum problems', *Canadian Journal of Mathematics* **12**(5), 363–365.

Kiefer, J., Wolfowitz, J. et al. (1959), 'Optimum designs in regression problems', *The Annals of Mathematical Statistics* **30**(2), 271–294.

Lange, K. (2010), *Numerical analysis for statisticians*, Springer Science & Business Media.

Lindley, D. (1972), *Bayesian Statistics, A Review*, Vol. 2, SIAM.

López-Fidalgo, J. (2009), 'A critical overview on optimal experimental designs', *Boletín de Estadística e Investigación Operativa* **25**(1), 14–21.

López-Fidalgo, J., Tommasi, C. & Trandafir, P. C. (2008), 'Optimal designs for discriminating between some extensions of the michaelis–menten model', *Journal of Statistical Planning and Inference* **138**(12), 3797–3804.

López-Fidalgo, J. & Wong, W. K. (2002), 'Design issues for the michaelis–menten model', *Journal of Theoretical Biology* **215**(1), 1–11.

Mukhopadhyay, S. & Haines, L. M. (1995), 'Bayesian d-optimal designs for the exponential growth model', *Journal of Statistical Planning and Inference* **44**(3), 385–397.

Mullen, K., Ardia, D., Gil, D. L., Windover, D. & Cline, J. (2016), 'Deoptim: An r package for global optimization by differential evolution', Available: https://CRAN.R-project.org/package=DEoptim.

Nelder, J. A. & Mead, R. (1965), 'A simplex method for function minimization', *The computer journal* **7**(4), 308–313.

Parker, S. & Gennings, C. (2008), 'Penalized Locally Optimal Experimental Designs for Nonlinear Models', *Journal of Agricultural, Biological, and Environmental Statistics* **13**(3), 334–354.

Parker, S. M. (2005), Solutions to reduce problems associated with experimental designs for nonlinear models: Conditional analyses and penalized optimal designs, PhD thesis, Virginia Commonwealth University.

Pilz, J. (1991), *Bayesian estimation and experimental design in linear regression models*, Vol. 212, John Wiley & Sons Inc.

Pronzato, L. (2008), Asymptotic properties of adaptive penalized optimal design with application to dose-finding, Technical report.

Pronzato, L. (2010), 'Penalized optimal designs for dose-finding', *Journal of Statistical Planning and Inference* **140**(1), 283–296.

Pukelsheim, F. (1993), *Optimal design of experiments*, John Wiley & Sons Inc, New York.

Pukelsheim, F. & Rieder, S. (1992), 'Efficient rounding of approximate designs', *Biometrika* **79**(4), 763–770.

Rodríguez Torreblanca, C. & Ortiz Rodríguez, I. M. (1999), *Diseño óptimo de experimentos para modelos de regresión*, Universidad Almería, Servicio de publicaciones, Almería.

Roy, A., Ghosal, S. & Rosenberger, W. F. (2009), 'Convergence properties of sequential bayesian d-optimal designs', *Journal of Statistical Planning and Inference* **139**(2), 425–440.

Rudnykh, S. I. & López-Ríos, V. I. (2018), 'Elección de la función de deseabilidad para diseños óptimos bajo restricciones', *Revista EIA* **15**(30), 13–24.

Seber, G. & Wild, C. (2003), *Nonlinear regression*, John Wiley & Sons.

Sengupta, A. (2004), 'Generalized variance', *Encyclopedia of statistical sciences* .

Storn, R. & Price, K. (1997), 'Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces', *Journal of global optimization* **11**(4), 341–359.

Team, R. C. (2018), 'R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013 (version 3.5.1)', Available: http://www.R-project.org.

Trandafir, C. & López-Fidalgo, J. (2004), Locally optimal designs for an extension of the michaelis-menten model, *in* A. Di Bucchianico, H. Ler & H. P. Wynn, eds, 'mODa 7–Advances in Model-Oriented Design and Analysis', Physica-Verlag, Heidelberg, pp. 173–181.

Wilks, S. S. (1967), *Collected papers; contributions to mathematical statistics*, Wiley.

Wu, F.-C. (2005), 'Optimization of correlated multiple quality characteristics using desirability function', *Quality Engineering* **17**(1), 119–126.

Yeatts, S. D., Gennings, C. & Crofton, K. M. (2012), 'Optimal design for the precise estimation of an interaction threshold: The impact of exposure to a mixture of 18 polyhalogenated aromatic hydrocarbons', *Risk Analysis* **32**(10), 1784–1797.

Ypma, J., Borchers, H. & Eddelbuettel, D. (2014), 'R package 'nloptr': R interface to nlopt', Available: https://CRAN.R-project.org/package=nloptr.

Zhang, Y. (2006), Bayesian D-optimal design for generalized linear models, PhD thesis, Virginia Tech.