

**Optimal Designs for Linear Models: Balancing
Efficiency and Independence among Parameter
Estimators**

by

Gun Ho Bae

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of
The University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Statistics

Faculty of Science

University of Manitoba

Winnipeg, Manitoba

Canada

© Gun Ho Bae, 2026

Abstract

Optimal design theory provides a systematic method for selecting experimental conditions that yield efficient and unbiased parameter estimates. While classical design criteria such as D -, A - and L -optimality have been extensively studied, they often focus on minimizing generalized variance or average variance without considering any specialized aspect of estimator behaviour. This dissertation develops methods for constructing optimal designs for several criteria based on the variances, covariances and correlations among the parameter estimators of linear models. It also develops methods for constrained optimal designs to maximize a set of compound criteria formed by combining a classical optimality criterion (D - or A -optimality) with correlation constraints.

Beyond classical variance-based objectives, special attention is given to minimizing the covariances and correlations among parameter estimators. Criteria are proposed that enforce balance among estimators, such as equalizing variances of specific parameter estimators or reducing correlation structures, thereby improving interpretability and stability of inference.

A further contribution is the study of constrained optimal designs. Many practical experiments must satisfy resource, ethical, or operational constraints

that restrict the design space or impose conditions on the design measure. Specifically, it introduces a general constrained optimization methodology, formulates first-order Lagrangian conditions, and implements iterative algorithms for constrained probability distributions. Numerical studies demonstrate how these constrained designs maintain efficiency while satisfying feasibility requirements, thereby extending the applicability of optimal design methodology to realistic experimental contexts.

Applications include polynomial regression models, chemistry-inspired regression models, and a radiation-dosage model examining the effect of treatment levels on tumor shrinkage in breast cancer patients. Across these examples, the proposed designs achieve lower estimator variance, reduced parameter correlation, and improved balance compared to standard optimal designs.

Overall, this research contributes to the theory and practice of optimal experimental design by (i) generalizing classical criteria to variance-, covariance-, and correlation-minimizing objectives, (ii) establishing rigorous optimality conditions and algorithmic construction methods, and (iii) extending the methodology to constrained optimization with correlation constraints.

Acknowledgements

My deepest gratitude and profound appreciation are first extended to my Ph.D. supervisor, Dr. Saumen Mandal, for his exceptional mentorship and unwavering support throughout my doctoral journey. Dr. Mandal not only provided the fascinating and challenging research topic that forms the basis of this dissertation but also fostered an environment conducive to intellectual and professional growth. His enthusiastic guidance, patience, and invaluable support were instrumental in the successful completion of this work. There were moments of considerable difficulty during this long and arduous endeavor, but his firm belief in my potential and steady direction kept me focused and motivated. His mentorship has been the single most crucial factor in realizing this academic milestone, and I believe the successful culmination of this thesis would not have been possible without his guidance.

I am also profoundly grateful to the members of my Ph.D. advisory committee: Dr. Ruppa Thulasiram (Department of Computer Science) and Dr. Po Yang (Department of Statistics). Their insightful questions, rigorous critiques, and constructive suggestions significantly sharpened the methodology and refined the presentation of this dissertation.

I extend my sincere thanks to the faculty members and staff of the Department of Statistics at the University of Manitoba for providing a stimulating and supportive academic environment throughout my graduate studies. Their collegiality, valuable discussions, and friendship greatly enriched my overall experience.

I gratefully acknowledge the financial support received from the Faculty of Graduate Studies, the Faculty of Science, the Department of Statistics, and from Dr. Mandal's NSERC research grant. Their generous contributions provided the essential resources necessary to conduct this research.

Finally, and most importantly, I reserve my deepest love and gratitude for my family. To my parents and my sister, thank you for your unwavering faith in me and for providing continuous, steadfast support throughout this entire academic endeavour. Above all, I owe an immeasurable debt of gratitude and affection to my cherished wife and our dear son. This doctoral journey became profoundly challenging as we navigated both our marriage and the birth of our son, yet their constant belief, encouraging smiles, and selfless sacrifice served as the pillar that sustained me. They are my ultimate motivation, and the completion of this thesis is dedicated to them.

Contents

Abstract	i
Acknowledgements	iii
Contents	v
List of Figures	ix
1 Introduction	1
1.1 Choice of a good design	3
1.2 Statistical model	4
1.3 The design space, types of designs, and the design measure . . .	8
2 Criteria, Optimality Conditions and Construction of Optimal Designs	15
2.1 Introduction	15
2.2 Criteria when interest is in inference about all of the parameters	16

2.2.1	D -optimality	16
2.2.2	A -optimality	17
2.2.3	G -optimality	17
2.2.4	E -optimality	18
2.3	Criteria when interest is only in some of the parameters or linear functions of the parameters	19
2.3.1	D_A -optimality	19
2.3.2	L -optimality	19
2.3.3	c -optimality	20
2.4	Optimality conditions	20
2.5	Construction of optimal designs	24
2.5.1	A class of algorithms	27
2.5.2	Choice and Scaling of Design Points	34
2.6	An application - Radiation dosage for tumour shrinkage	36
2.6.1	Introduction	36
2.6.2	Evaluating the criteria	38
2.6.3	Finding the optimal dosage and weights using the algorithm	41

3	Minimizing variance based criteria	44
3.1	Introduction	44
3.1.1	The General Equivalence Theorem and its Implications .	46
3.2	Variance based criteria for minimizing the average or total variance	47
3.2.1	Cross-efficiencies	57
3.3	Variance based criteria for equality of variances	66
4	Minimizing covariance based criteria	74
4.1	Introduction	74
4.2	Minimizing one covariance	76
4.3	Minimizing two or more covariances	84
5	Correlation based criteria: Estimating parameters with minimal correlation	94
5.1	Introduction	94
5.2	Minimizing one correlation	96
5.3	Compound optimization: Minimizing two or more correlations .	106
6	Optimal designs under constraints on independence estimation of parameters	114
6.1	Introduction	114

6.2	Constrained optimization with generalized or average variance subject to the constraints on covariance and correlation	117
6.2.1	General Problem (GP)	117
6.2.2	Methodology	118
6.3	Algorithms	129
6.3.1	Iterative Algorithm and Derivatives	129
6.3.2	Summary of Optimization Terms	133
6.4	Examples	136
6.4.1	A simple example on quadratic model	136
6.4.2	Results	140
7	Conclusion	174
	Bibliography	183

List of Figures

2.1	Simple linear Regression: Variance function $d(x, p^*)$ at the optimum design. The maximum value of 2 occurs at the support points.	32
2.2	Quadratic Regression: Variance function $d(x, p^*)$ at the optimum design. The maximum value of 3 occurs at the support points. . .	34
6.1	$\phi(p_2)$ versus p_2	165
6.2	$W_2(p_2)$ versus p_2	165
6.3	$W_4(p_2)$ versus p_2	166
6.4	$W_6(p_2)$ versus p_2	166
6.5	$W_7(p_2)$ versus p_2	167
6.6	$W_8(p_2)$ versus p_2	167
6.7	$W_9(p_2)$ versus p_2	168
6.8	$W_{10}(p_2)$ versus p_2	168
6.9	$W_{11}(p_2)$ versus p_2	169

6.10	$W_{12}(p_2)$ versus p_2 .	169
6.11	$\phi(p_3)$ versus p_3 .	170
6.12	$W_4(p_3)$ versus p_3 .	170
6.13	$W_5(p_3)$ versus p_3 .	171
6.14	$W_7(p_3)$ versus p_3 .	171
6.15	$W_8(p_3)$ versus p_3 .	172
6.16	$W_{10}(p_3)$ versus p_3 .	172
6.17	$W_{11}(p_3)$ versus p_3 .	173

Chapter 1

Introduction

This dissertation explores the theory and construction of optimal experimental designs for linear models, with a focus on going beyond traditional optimality criteria to include variance-, covariance-, and correlation-based criteria. The dissertation also explores the constrained optimal designs under explicit statistical constraints, specifically bridging the gap between efficiency and independence of the parameter estimators. As this study is based on optimal design, we begin with an in-depth discussion of optimal design theory.

Optimal designs are experimental designs that are optimal with respect to a given criterion. The development of optimal design theory is commonly attributed to the Danish statistician Kirstine Smith. In 1918, Smith wrote what is known as the first paper on optimal experimental design (Smith, 1918), a very detailed study in the context of polynomial models. Also, Kiefer (1959, 1974) is a renowned statistician for his studies in optimal design and has made many significant contributions, such as the general equivalence theorem, various

optimality criteria, and the construction of various optimal designs, to name a few. He introduced the formal “alphabetic optimality” terminology (e.g., *A*-optimality, *D*-optimality, *E*-optimality) to distinguish various design criteria. In fact, Kiefer is widely regarded as the “father of optimal design”.

In the design of experiments for estimating statistical models, optimal designs allow for parameters to be estimated without bias and with minimal variance. A non-optimal design requires a greater number of experimental runs to estimate the parameters with the same precision as an optimal design. Thus, optimal design can reduce the costs of experimentation as well as the amount of time required.

The goal of optimal design is to achieve precise parameter estimation while minimizing the required sample size, time, and cost. In order to achieve our desired goals, we define an appropriate criterion function, typically based on the information matrix. There are several design criteria that will lead to different optimal designs, such as *D*-optimality, *A*-optimality, *c*-optimality, and *G*-optimality. This research will focus on developing and constructing optimal designs based on some non-standard criteria. In addition, optimal designs are constructed subject to satisfying some constraints of interest.

We focus on optimal design theory for linear models. We will give some fundamental concepts of optimal design theory, such as the definition of a design, standardized variance function, information matrix, and various criterion functions and their properties. The relationship between the response variable and predictors is modeled through a design matrix and observed responses.

1.1 Choice of a good design

Many research problems can be formulated as relationships between a set of independent variables X and an outcome variable Y . The primary objective is to identify the relevant predictors and the response of interest.

A research design is implemented to organize the data collection process. It involves selecting the number and type of variables, as well as the number of measurements, time points, and subjects. For optimal design purposes, the statistical model must be carefully specified (Fedorov, 1972; Ghosh, 2025; Atkinson et al., 2007; Berger and Wong, 2009; López-Fidalgo, 2023).

To select an appropriate design, some knowledge of the true model or data-generating process is required. In many applications, the goal is to assess the effects of one or more predictors on a response variable.

Optimal design theory provides a systematic framework for utilizing available information to construct efficient designs. Once the statistical model, objectives, and constraints are specified, optimization methods are employed to determine the optimal design.

A researcher must also address the questions of how many independent variables to use and how many levels of these variables are needed to investigate the relationship between the independent variables and the dependent variable. If a whole range of levels of an independent variable is of interest and the study only includes a few of these levels, then it is best to randomly select these few levels from that range, such that independent variables are classified as random factors. In contrast, the so-called fixed factors are independent variables with a

fixed number of levels, such that inferences are only limited to these fixed levels. For example, the factor Age can be either fixed or random. Age is a fixed factor if persons from a fixed set of age groups are included in the study and the researcher is only interested in drawing conclusions for these fixed age groups. The factor Age is random if the age groups are randomly drawn from the complete range of all ages in the population so that conclusions apply to the whole range of ages. Fixed and random factors must be clearly identified in the statistical model because they affect our tests concerning model parameters. Once objectives and model specifications are determined, optimization procedures are employed to construct the optimal design.

1.2 Statistical model

Let us consider the problem of choosing an experimental plan that will give us information on models given by:

$$y \sim p(y|\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\sigma}), \quad (1.1)$$

where y is the response variable. The response y is a real-valued function evaluated at a design variable $\mathbf{x}$. Let $\mathbf{x} = (x_1, x_2, \dots, x_m)^\top$. The design variable $\mathbf{x}$ is chosen by the experimenter, the values belong to a space $\mathcal{X}$, called the design space. Typically this space is continuous but can be discrete as well. $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$ is a vector of k unknown parameters. The true value of $\boldsymbol{\theta}$ belongs to a set, say Θ . $p(\cdot)$ is a probability model. From the given design space $\mathcal{X}$, the experimenter can select the experimental conditions. In most situations,

$\mathcal{X}$ is taken to be compact. For each $\mathbf{x} \in \mathcal{X}$, an experiment is performed, and the outcome is a random variable $y = y(\mathbf{x})$, where $\text{Var}(y(\mathbf{x})) = \sigma^2$. σ is a fixed and unknown parameter. This is not of primary interest.

By linear model we mean the model is linear in the unknown parameters $\boldsymbol{\theta}$ but not necessarily linear in $\mathbf{x}$. So in linear models y has an expected value of the following form:

$$E(y|\mathbf{x}, \boldsymbol{\theta}, \sigma) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\theta} \quad (1.2)$$

where $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x}))^\top$ is a vector of k real valued functions. These functions are defined on the design space $\mathcal{X}$.

The design problem thus consists of selecting design points $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ on $\mathcal{X}$ to ensure the best inference or as reliable an inference as possible for all or some of the parameters $\boldsymbol{\theta}$. Such an allocation of observations over the design space $\mathcal{X}$ is referred to as a design.

Consider the model (1.2). Let y_i be the response (observation) obtained at $\mathbf{x}_i$ so that

$$E(y_i) = \mathbf{v}_i^\top \boldsymbol{\theta}, \quad \mathbf{v}_i = (f_1(\mathbf{x}_i), f_2(\mathbf{x}_i), \dots, f_k(\mathbf{x}_i))^\top, i = 1, 2, \dots, n. \quad (1.3)$$

Typically one or more observations are taken at the same $\mathbf{x}$ value. Suppose that $y_1, y_2, \dots, y_n$ are independent random variables with equal variance σ^2 . Thus we can write the standard linear model as given by

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\theta}, \quad D(\mathbf{Y}) = \sigma^2 \mathbf{I}_n \quad (1.4)$$

where $\mathbf{Y} = (y_1, y_2, \dots, y_n)^\top$, $\mathbf{X}$ is the $n \times k$ matrix whose (i, j) th element is $f_j(x_i)$, $D(\mathbf{Y})$ is the dispersion matrix (variance-covariance matrix) of $\mathbf{Y}$, and I_n is the $n \times n$ identity matrix. For this model, least squares estimators are a common choice because they are best linear unbiased estimators. Normal equations are given by

$$(\mathbf{X}^\top \mathbf{X}) \hat{\boldsymbol{\theta}} = \mathbf{X}^\top \mathbf{Y}. \quad (1.5)$$

The matrix $(\mathbf{X}^\top \mathbf{X})$ is the information matrix of $\boldsymbol{\theta}$. As we have a linear model, the information matrix does not depend on the parameters. The larger the matrix $(\mathbf{X}^\top \mathbf{X})$, the more information is in the experiment. If we are interested in all the parameters, then the selection of $\mathbf{x}$ must at least ensure that the information matrix is non-singular. In such case the unique solution for (1.5) is

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}. \quad (1.6)$$

Note that $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ and $D(\hat{\boldsymbol{\theta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.

The information matrix is symmetric and nonnegative definite. The information matrix is widely used in optimal experimental design. As we see the inverse of the covariance matrix is the information matrix. Because of this reciprocity property of estimator variance and information matrix, minimizing the variance corresponds to maximizing the information.

The predicted value of the response at $\mathbf{x}$ is

$$\begin{aligned} \hat{Y}(\mathbf{x}) &= f_1(\mathbf{x}) \hat{\theta}_1 + f_2(\mathbf{x}) \hat{\theta}_2 + \dots + f_k(\mathbf{x}) \hat{\theta}_k \\ &= \mathbf{f}^\top(\mathbf{x}) \hat{\boldsymbol{\theta}}. \end{aligned}$$

Note that the covariance matrix of $\hat{\theta}$, $D(\hat{\theta})$, does not depend on θ for linear models and only depends proportionally on σ^2 . In order to make better design, the matrix $D(\hat{\theta})$ must be as small as possible. This means that the matrix $(\mathbf{X}^\top \mathbf{X})$ must be as large as possible.

As an example, let us consider the simple linear regression model:

$$y = \theta_1 + \theta_2 x + \varepsilon = \mathbf{f}^\top(x) \boldsymbol{\theta} + \varepsilon$$

where $\mathbf{f}(x) = (1, x)^\top$, $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ where the intercept θ_1 and the slope θ_2 are unknown parameters and the ε 's are random errors assumed to be independent and normally distributed with zero mean and common variance σ^2 . The least squares estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are:

$$\hat{\theta}_1 = \bar{y} - \hat{\theta}_2 \bar{x}, \quad \hat{\theta}_2 = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \frac{SS_{xy}}{SS_x}$$

where the sum of squares between x and y is $SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$ and the sum of squares of x is $SS_x = \sum (x_i - \bar{x})^2$. The estimators are unbiased, i.e., $E(\hat{\theta}_1) = \theta_1$ and $E(\hat{\theta}_2) = \theta_2$. Under the normality assumption, these estimators have minimal variances among all linear unbiased estimators of θ_1 and θ_2 . Therefore, these estimators are efficient. The efficiency of an estimator increases as its variance becomes smaller. A larger variance of an estimator indicates more uncertainty about the estimators.

The variances of the two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ for the simple regression model are

$$\text{Var}(\hat{\theta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SS_x} \right), \quad \text{Var}(\hat{\theta}_2) = \frac{\sigma^2}{SS_x} = \frac{\sigma^2}{n \text{Var}(x)}. \quad (1.7)$$

Here, we can see the variances of the two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ will be decreasing, if the variance of the error terms in the model σ^2 decreases and/or the variation SS_x or variance $\text{Var}(x)$ increases and/or the sample size n increases. Increasing n beyond a certain value is not practical. However, we can select the values of x such that $\text{Var}(x)$ increases. This is one of the key points in optimal design.

Now, the question is, how do we select the values of x from the design space? We will discuss this in the following sections.

1.3 The design space, types of designs, and the design measure

When we construct optimal design in a continuous design space, we discretize the design space and construct the design accordingly. We do this in the following. The linear model in (1.2) can be written as:

$$E(y|\mathbf{v}, \boldsymbol{\theta}, \sigma) = \mathbf{v}^\top \boldsymbol{\theta} \quad (1.8)$$

where

$$\begin{aligned} \mathbf{v} &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x}))^\top, \mathbf{v} \in \mathcal{V}, \\ \mathcal{V} &= \{\mathbf{v} \in R^k : \mathbf{v} = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x}))^\top, \mathbf{x} \in \mathcal{X}\}. \end{aligned}$$

Here, $\mathcal{V}$ is the image under $\mathbf{f}$ of $\mathcal{X}$. So we can write $\mathcal{V} = \mathbf{f}(\mathcal{X})$, where $\mathbf{f}$ is the vector of real valued functions $(f_1, f_2, \dots, f_k)^\top$. Choosing a vector $\mathbf{x}$ in

the design space $\mathcal{X}$ is actually choosing a k -vector $\mathbf{v}$ in the closed bounded k -dimensional space $\mathcal{V}$. We call $\mathcal{V}$ is an induced design space. Suppose that $\mathcal{V}$ consists of J distinct vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_J$. To obtain observations on y , values for $\mathbf{v}$ must be chosen from the J elements of $\mathcal{V}$. So at this point we can focus on this induced design space.

Suppose that we will have n observations in total. So we need to decide how many of these, say n_j , to take at $\mathbf{v}_j$, $\sum_{j=1}^J n_j = n$. Given this setup, the matrix $(\mathbf{X}^\top \mathbf{X})$ can be written as:

$$\mathbf{X}^\top \mathbf{X} = \mathbf{M}(\mathbf{n}), \quad \mathbf{n} = (n_1, n_2, \dots, n_J)^\top \quad (1.9)$$

where

$$\begin{aligned} \mathbf{M}(\mathbf{n}) &= \sum_{j=1}^J n_j \mathbf{v}_j \mathbf{v}_j^\top \\ &= \mathbf{V} \mathbf{N} \mathbf{V}^\top \end{aligned}$$

and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_J]$, $\mathbf{N} = \text{diag}(n_1, n_2, \dots, n_J)$.

The objective is to choose n_j in order to make the matrix $\mathbf{M}(\mathbf{n})$ in a specific context. Since the n_j 's must be integers, this becomes an integer programming problem, which in the context of design is known as an exact design. Integer programming problems are challenging to solve because the principles of calculus cannot be utilized to determine the optimal solutions. Additionally, a distinct solution must be developed for each specific value of n . Due to this inherent characteristic of the problem, it is not possible to formulate an optimal exact design that can be expressed as a function of n . An alternative way is to represent

the matrix $M(\mathbf{n})$ differently. Let us define

$$p_j = n_j/n, \mathbf{p} = (p_1, p_2, \dots, p_J).$$

Then we can write the matrix $M(\mathbf{n})$ as

$$M(\mathbf{n}) = nM(\mathbf{p}) \tag{1.10}$$

where

$$M(\mathbf{p}) = \sum_{j=1}^J p_j \mathbf{v}_j \mathbf{v}_j^\top = \mathbf{V} \mathbf{P} \mathbf{V}^\top \tag{1.11}$$

and $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_J)$. Here, p_j is the proportion of observations taken at $\mathbf{v}_j$. So $p_j \geq 0$, $\sum_{j=1}^J p_j = 1$; and $\mathbf{p} = (p_1, p_2, \dots, p_J)$ is the resultant distribution on $\mathcal{V}$.

Thus, instead of dealing with the matrix $M(\mathbf{n})$, our problem is now to determine $\mathbf{p}$ to make the matrix $M(\mathbf{p})$ large subject to $p_j = n_j/n$. Relaxing $p_j = n_j/n$ to $p_j \geq 0$ and $\sum_{j=1}^J p_j = 1$ leads to an approximate design. Clearly, an approximate design problem is simpler and more flexible to solve than an exact design.

Therefore, an approximate solution would generally be preferred to the original exact design problem. Kiefer pioneered this approach and provided three powerful reasons for working with approximate designs instead of exact designs. The first is that it is highly challenging to identify optimal exact designs since they depend on the precise value of n in addition to the model and optimality criterion.

Because each value of the sample size n produces a new optimal design, we would need an unlimited list of optimal designs for practitioners for each model and criterion. The ideal approximate design, on the other hand, is independent of the value of n , making it simpler to deal with. Second, when we use an approximate design, rounding is necessary. The difference between the exact design and the implemented design disappears as n increases, and one can demonstrate that the implemented optimal approximate design is always close to the optimal exact design. The third reason is that many optimal exact designs still cannot be discovered for comparatively easy situations, as they often involve complex mathematical theory. In contrast, optimal approximation designs are either available analytically or can be discovered by iterative computer algorithms.

For example, we can discuss the exact and approximate designs in the context of a design problem for dose-response study. The dose levels, the number of dose levels, and the number of individuals to assign to each of these dose levels must all be determined by the researcher from a specified dose interval. Suppose that the resources at our disposal allow us to enroll in a set number of subjects, say n . Suppose two dosages are given. The design may specify that one third of the subjects be assigned to the first dose level and two third be given to the other dose level. If $n = 180$, then one-third to first dose and two-third to the second dose, resulting in the design that has 60 subjects at the first dose and 120 at the second dose. This is an exact design. But if $n = 100$, then 33.33333 subjects should be assigned at the first dose and 66.66666 at the second dose. This is an approximate design. Approximate design is rounded in some natural way so that it becomes an exact design before implementation. One of the following

forms may be used for the implemented design of the approximation design: 33 subjects at the first dose and 67 subjects at the second dose 2, or 34 subjects at the first dose and 66 subjects at the second dose.

We have used the term $\mathbf{p}$ to refer to both the vector $(p_1, p_2, \dots, p_J)$ and a probability distribution on the induced design space $\mathcal{V}$. Naturally, this results in a distribution or measure on the original design space $\mathcal{X}$. The full description of the design measure would be as follows:

$$\mathbf{p} = \begin{pmatrix} x_1 & x_2 & \dots & x_J \\ p_1 & p_2 & \dots & p_J \end{pmatrix}. \quad (1.12)$$

The first line provides the locations of the design points, while the second line provides the corresponding design weights. Note that, $\sum_{j=1}^J p_j = 1$ and $0 \leq p_j \leq 1$ for all j . Note also that, if $\mathbf{p}$ is a design measure or distribution, it is by definition defined on the design space. We will always be clear about what the design space is and about which design point receives which weight.

The support of the design measure $\mathbf{p}$ in the induced design space $\mathcal{V}$ is defined to be those vertices $\mathbf{v}_j$ with nonzero weights under $\mathbf{p}$. We denote it by

$$\text{Supp}(\mathbf{p}) = \{\mathbf{v}_j \in \mathcal{V} : p_j > 0, j = 1, 2, \dots, J\}.$$

The standardized variance of the predicted response on y at $\mathbf{x}$ for the design is given by

$$d(\mathbf{x}, \mathbf{p}) = \mathbf{f}^\top(\mathbf{x}) \mathbf{M}^{-1}(\mathbf{p}) \mathbf{f}(\mathbf{x}) \quad (1.13)$$

where $M(\mathbf{p})$ is the information matrix. As we can see that the standardized variance of the predicted response depends on the information matrix.

The information matrix plays a big role in optimal experimental design. Let us discuss some properties of this matrix. Note that, information matrix $M(\mathbf{p})$, as defined in (1.11). The information matrix $M(\mathbf{p})$ is symmetric and nonnegative definite. The symmetry of this matrix follows from its definition (1.10) and nonnegativeness can be verified by the following quadratic form.

$$\begin{aligned} \mathbf{a}^\top M(\mathbf{p}) \mathbf{a} &= \mathbf{a}^\top E_p[\mathbf{v} \mathbf{v}^\top] \mathbf{a} \\ &= E_p[\mathbf{a}^\top \mathbf{v} \mathbf{v}^\top \mathbf{a}] \\ &= E_p[(\mathbf{a}^\top \mathbf{v})^2] \geq 0. \end{aligned}$$

The information matrix is widely used in optimal experimental design. As mentioned before, the inverse of the variance-covariance matrix (variance-matrix or dispersion matrix) is actually the information matrix. Because of this reciprocity property of estimator-variance and information, minimizing the variance corresponds to maximizing the information. We will see that many of the design criteria are functions of the above information matrix $M(\mathbf{p})$.

We organize the rest of the chapters as follows. In Chapter 2, we introduce the theoretical background and notation for optimal design, including the information matrix and standard optimality criteria. Chapter 3 focuses on minimizing variance-based criteria. We study designs that minimize the variances of least-squares estimators for linear functions of the parameters as well as construct designs for equality of variances of certain parameter estimators of interest. In

Chapter 4, we explore covariance-based criteria. Here the aim is to construct designs that make the covariances of the parameter estimators zero or as small as possible. Chapter 5 turns to correlation-based criteria, providing an alternative but related perspective and illustrating how designs can be constructed to minimize correlations between estimators. Chapter 6 generalizes the methodology to constrained optimal designs. We incorporate restrictions into the optimization problem through Lagrangian methods combined with simultaneous optimization techniques to maximize compound criteria, and adapt multiplicative algorithms to accommodate covariance, and correlation constraints. Finally, Chapter 7 concludes with a summary of the contributions and directions for future research.

Chapter 2

Criteria, Optimality Conditions and Construction of Optimal Designs

2.1 Introduction

The goal of optimal experimental design is to choose experimental conditions so that the resulting parameter estimates are as informative and precise as possible. This is usually formalized by maximizing a real-valued function

$$\phi(\boldsymbol{p}) = \psi\{M(\boldsymbol{p})\}$$

where $M(\boldsymbol{p})$ is the information matrix associated with a design $\boldsymbol{p}$. A design maximizing $\phi(\boldsymbol{p})$ is called a ϕ -optimal design.

Numerous optimality criteria have been proposed in the literature; see, for example, Atkinson et al. (2007), Pukelsheim (2006), Fedorov (1972), Silvey

(1980), Berger and Wong (2009), and López-Fidalgo (2023). These criteria are usually named after a letter of the alphabet. These are the so-called alphabetic optimality criteria (Kiefer, 1959). These criteria can be broadly classified into two categories: (i) criteria that address all parameters in the model, and (ii) criteria that focus on a subset or on specific linear functions of the parameters.

2.2 Criteria when interest is in inference about all of the parameters

2.2.1 *D*-optimality

D-optimality is perhaps the most widely used criterion. It seeks to minimize the determinant of the covariance matrix of the parameter estimators, or equivalently, to maximize the determinant of the information matrix. The criterion is given by

$$\phi_D(\mathbf{p}) = \psi_D\{\mathbf{M}(\mathbf{p})\} = \log \det\{\mathbf{M}(\mathbf{p})\} = -\log \det\{\mathbf{M}^{-1}(\mathbf{p})\}. \quad (2.1)$$

In other words, *D*-optimality minimizes the generalized variance of the parameter estimates. Under a normal-error linear model, the joint confidence region for $\boldsymbol{\theta}$ is an ellipsoid with volume proportional to $[\det\{\mathbf{M}(\mathbf{p})\}]^{-1/2}$, so a *D*-optimal design yields the smallest such volume. *D*-optimality criterion is concave on the cone of positive-definite matrices and invariant under non-singular linear

transformations of the design space. Its first derivatives are

$$\frac{\partial \phi_D}{\partial p_j} = \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j. \quad (2.2)$$

2.2.2 A-optimality

A-optimality minimizes the average variance of the parameter estimators by minimizing the trace of the inverse information matrix, or equivalently by maximizing the following criterion:

$$\phi_A(\mathbf{p}) = -\text{tr}\{\mathbf{M}^{-1}(\mathbf{p})\}. \quad (2.3)$$

This criterion is especially useful when the overall average precision of all parameters is more important than their joint confidence region volume. Unlike *D*-optimality, *A*-optimality is not invariant to rescaling of factors. Its partial derivatives are

$$\frac{\partial \phi_A}{\partial p_j} = \mathbf{v}_j^\top \mathbf{M}^{-2}(\mathbf{p}) \mathbf{v}_j. \quad (2.4)$$

2.2.3 G-optimality

G-optimality minimizes the maximum variance of the predicted response over the design space.

$$\phi_G(\mathbf{p}) = -\max_{\mathbf{v} \in \mathcal{V}} \mathbf{v}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}. \quad (2.5)$$

This criterion is most appropriate when the goal is accurate prediction throughout the entire design region rather than only at the design points. In linear models,

G -optimality is equivalent to D -optimality (Kiefer and Wolfowitz, 1960). A convenient form of the derivative, when $\mathbf{v}_j$ achieves the maximum, is

$$\frac{\partial \phi_G}{\partial p_j} = [\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j]^2. \quad (2.6)$$

2.2.4 E -optimality

E -optimality maximizes the smallest eigenvalue of the information matrix (or equivalently minimizes the largest eigenvalue of the covariance matrix):

$$\phi_E(\mathbf{p}) = \lambda_{\min}\{\mathbf{M}(\mathbf{p})\}. \quad (2.7)$$

So, by the nature of this optimality, it complements D - and A -optimality by focusing on the “worst-case” variance rather than the determinant or trace of the information or the covariance matrix.

If $\mathbf{u}_{\min}$ denotes a normalized eigenvector corresponding to $\lambda_{\min}(\mathbf{M})$, then a first derivative of ϕ_E with respect to p_j can be written as

$$\frac{\partial \phi_E}{\partial p_j} = \mathbf{u}_{\min}^\top [\mathbf{v}_j \mathbf{v}_j^\top] \mathbf{u}_{\min} \quad (2.8)$$

which reflects the contribution of the j th design point to the least-informative direction of the parameter space.

2.3 Criteria when interest is only in some of the parameters or linear functions of the parameters

2.3.1 D_A -optimality

When interest is confined to s linear combinations $\alpha = A\theta$ with A of rank $s \leq k$, the covariance matrix of the estimator of $A\theta$ is proportional to $AM^{-1}(p)A^\top$. The D_A criterion maximizes the determinant of the inverse of this matrix. The criterion is given by

$$\phi_{D_A}(p) = -\log \det\{AM^{-1}(p)A^\top\}. \quad (2.9)$$

This extends D -optimality to cases where only part of the parameter vector is of primary interest. We will consider one such case in the next chapter. Its partial derivatives are

$$\frac{\partial \phi_{D_A}}{\partial p_j} = v_j^\top M^{-1}(p)A^\top [AM^{-1}(p)A^\top]^{-1} AM^{-1}(p)v_j. \quad (2.10)$$

2.3.2 L -optimality

Given a positive semi-definite matrix L , an L -optimal design minimizes

$$\phi_L(p) = -\text{tr}\{M^{-1}(p)L\}. \quad (2.11)$$

This criterion is linear in the entries of $M^{-1}(p)$ and is useful when different parameters or combinations have different weights of importance. If $L = A^\top A$,

then

$$\phi_L(\mathbf{p}) = -\text{tr}\{\mathbf{A}\mathbf{M}^{-1}(\mathbf{p})\mathbf{A}^\top\}. \quad (2.12)$$

Its partial derivatives are

$$\frac{\partial \phi_L}{\partial p_j} = \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A}^\top \mathbf{A} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j. \quad (2.13)$$

2.3.3 $\mathbf{c}$ -optimality

As a special case of L -optimality, consider estimating a single linear combination $\mathbf{c}^\top \boldsymbol{\theta}$. The $\mathbf{c}$ -optimal criterion minimizes the variance of the estimator of $\mathbf{c}^\top \boldsymbol{\theta}$.

So the criterion function is given by

$$\phi_c(\mathbf{p}) = -\mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{c}. \quad (2.14)$$

This is useful when one specific parameter (or a contrast, or a parametric function) is of interest. Its partial derivatives are

$$\frac{\partial \phi_c}{\partial p_j} = (\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{c})^2. \quad (2.15)$$

2.4 Optimality conditions

In this section we establish the conditions for optimality in our optimization problem, and then discuss how these conditions can be used to construct optimal

designs.

Our general problem is:

Maximize a criterion $\phi(\mathbf{p})$

$$\text{over } \mathcal{P} = \{\mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1\}. \quad (2.16)$$

Many well-known design problems fit this framework, including optimal regression designs, estimation problems and stratified sampling, where one seeks an optimal probability distribution over a finite set of design points.

We focus on a differential-calculus approach. A key tool is the directional derivative of the criterion function at a design $\mathbf{p}$ in the direction of another design. We denote this by $F_\phi\{\mathbf{p}, \mathbf{q}\}$. We also make use of a related derivative, convenient for computing partial derivatives of our criteria. These derivatives provide a unified way of expressing optimality conditions.

There are two main approaches to solving (2.16):

- (1) determine an optimizing design $\mathbf{p}^*$ directly, or
- (2) first determine an x^* maximizing a related criterion over $\mathcal{P}$ and then find $\mathbf{p}^*$ such that $x(\mathbf{p}^*) = x^*$. Here we adopt the first approach, which requires explicit conditions characterizing an optimal $\mathbf{p}^*$. We determine the optimality conditions in terms of point to point directional derivatives of the criterion function.

Definition 1. Following Whittle (1973), the directional derivative $F_\phi(\mathbf{p}, \mathbf{q})$ of a criterion function $\phi(\cdot)$ at $\mathbf{p}$ in the direction of $\mathbf{q}$ is

$$F_\phi(\mathbf{p}, \mathbf{q}) = \lim_{\varepsilon \downarrow 0} \frac{\phi\{(1 - \varepsilon)\mathbf{p} + \varepsilon\mathbf{q}\} - \phi(\mathbf{p})}{\varepsilon}. \quad (2.17)$$

This directional derivative can exist even if $\phi(\cdot)$ is not differentiable.

Definition 2. Another useful derivative of $\phi(\cdot)$ is defined by

$$G_\phi\{\mathbf{p}, \mathbf{m}\} = \lim_{\varepsilon \downarrow 0} \frac{\phi\{\mathbf{p} + \varepsilon \mathbf{m}\} - \phi(\mathbf{p})}{\varepsilon}. \quad (2.18)$$

It is called the Gâteaux derivative of $\phi(\cdot)$ at $\mathbf{p}$ in the direction $\mathbf{m}$. Clearly $F_\phi\{\mathbf{p}, \mathbf{q}\} = G_\phi\{\mathbf{p}, \mathbf{m}\}$ with $\mathbf{m} = \mathbf{q} - \mathbf{p}$, and conversely $G_\phi\{\mathbf{p}, \mathbf{m}\} = F_\phi\{\mathbf{p}, \mathbf{p} + \mathbf{m}\}$. Note also that $G_\phi\{\mathbf{p}, \mathbf{e}_j\} = \frac{\partial^+ \phi}{\partial p_j}$, the right-hand partial derivative of $\phi(\cdot)$ with respect to p_j , where $\mathbf{e}_j$ is the j th unit vector. This is why we make use of the Gâteaux derivative to derive the partial derivatives of our criterion functions in the following chapters.

We formulate our optimality conditions in terms of the directional derivative $F_\phi(\mathbf{p}, \mathbf{q})$, which, by allowing the direction to be determined by a point $\mathbf{q}$, leads to a convenient generalization of standard calculus rules. If $\phi(\cdot)$ is differentiable at $\mathbf{p}$, then

$$F_\phi\{\mathbf{p}, \mathbf{q}\} = (\mathbf{q} - \mathbf{p})^\top \frac{\partial \phi}{\partial \mathbf{p}} = \sum_{i=1}^J (q_i - p_i) d_i, \quad d_i = \frac{\partial \phi}{\partial p_i}, \quad i = 1, \dots, J.$$

In particular,

$$F_j = F_\phi(\mathbf{p}, \mathbf{e}_j) = \frac{\partial \phi}{\partial p_j} - \sum_{i=1}^J p_i \frac{\partial \phi}{\partial p_i} = d_j - \sum_{i=1}^J p_i d_i.$$

Here $\mathbf{e}_j$ is the j th unit vector in $\mathbb{R}^J$; F_j is the *vertex directional derivative* of $\phi(\cdot)$. We call this as vertex directional derivative because it gives the directional derivative towards each vertex of the feasible region of the problem.

If $\phi(\cdot)$ is differentiable at an optimal distribution $\mathbf{p}^*$, the first-order conditions for $\mathbf{p}^*$ to be a local maximizer are:

$$F_j^* = F_\phi(\mathbf{p}^*, \mathbf{e}_j) = \begin{cases} 0 & \text{if } p_j^* > 0, \\ \leq 0 & \text{if } p_j^* = 0. \end{cases} \quad (2.19)$$

If $\phi(\cdot)$ is concave over its feasible region, these first-order conditions are both necessary and sufficient for optimality, a result known as the general equivalence theorem in optimal design (Kiefer, 1974). If the criterion function is not concave, first-order conditions are not sufficient for optimality. So we need to check the second-order condition. In order to check the second order condition, considering the nonzero components of $\mathbf{p}^*$, we need to choose free values for the first $J - 1$ components of $\mathbf{p}^*$, and set the final component in terms of the others. That is, we consider the substitution $p_J = 1 - (p_1 + p_2 + \cdots + p_{J-1})$. Let the criterion function $\phi(\mathbf{p})$ in terms of the free components $(p_1, p_2, \dots, p_{J-1})$ be $\tilde{\phi}(\mathbf{p})$. These conditions are expressed in terms of the Hessian:

$$H(\mathbf{p}) = (d_{ij}^{\tilde{\phi}}), \quad i, j = 1, 2, \dots, J - 1, \quad (2.20)$$

where $\tilde{\phi}$ is the criterion expressed in terms of the positive components of $\mathbf{p}$, and

$$d_{ij}^{\tilde{\phi}} = d_{ij}^{\phi} - d_{iJ}^{\phi} - d_{Jj}^{\phi} + d_{JJ}^{\phi}, \quad d_{ij}^{\phi} = \frac{\partial^2 \phi}{\partial p_i \partial p_j}.$$

The general approach is to first apply an iterative algorithm, which has at least the potential of converging to a candidate local maximum $\mathbf{p}^*$. We check

that $\mathbf{p}^*$ is a critical point using first order conditions. We then check that $\mathbf{p}^*$ satisfies a second order condition for constituting a local maximum. This second order condition is expressed in terms of the Hessian. The above Hessian matrix must be negative definite. That is, all of its eigenvalues must be negative. In the special case $J - 1 = 2$, this reduces to checking $\det H(\mathbf{p}^*) > 0$ and $d_{ii}^{\tilde{\phi}}(\mathbf{p}^*) < 0$ for $i = 1, 2$.

2.5 Construction of optimal designs

An analytic solution of the problem of constructing optimal designs is possible only in simple cases. If we are given the support points, we could find out an analytic solution of the optimal weights for some regression models. We first consider some polynomial regression problems for which explicit solutions can be obtained.

In polynomial regression in one variable of order $k - 1$ the model is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_{k-1} x^{k-1} = \mathbf{v}_x^\top \boldsymbol{\theta} \quad (2.21)$$

where $\mathbf{v}_x = (1, x, x^2, \dots, x^{k-1})^\top$, $x \in [-1, 1]$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{k-1})^\top$.

Fedorov (1972) reports that the discrete D -optimal design is unique, having a minimal support of k points which are the k roots of the polynomials

$$(1 - x^2)P'_{k-1}(x) \quad (2.22)$$

where $P_k(x)$ is the k^{th} Legendre polynomial:

$$P_k(x) = \sum_{n=0}^N \frac{(-1)^n (2k-2n)! x^{k-2n}}{2^k n! (k-n)! (k-2n)!}$$

where

$$N = \begin{cases} k/2 & \text{if } k \text{ is even} \\ (k-1)/2 & \text{if } k \text{ is odd.} \end{cases}$$

Note that, in a minimal support design, since the support of the optimal design $\mathbf{p}^*$ contains k points, the D -optimal design assigns weight $(1/k)$ to each of these.

We will find out analytic solutions of optimal designs for some polynomial regression models, namely, the quadratic and cubic regression models. For quadratic regression, we put $k = 3$ in model (2.21). The model is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

Using (2.22), we obtain

$$(1-x^2)\mathbf{p}'_2(x) = (3x)(1-x^2).$$

Hence, the support points of $\mathbf{p}^*$ are given by

$$x = 0, \pm 1.$$

Hence, we obtain the D -optimal design for the quadratic regression model as:

$$\mathbf{p}^* = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}.$$

For cubic regression model, we have $k = 4$ in model (2.21). The model is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3.$$

Using (3.4), we obtain

$$(1 - x^2)p'_3(x) = \frac{(15x^2 - 3)(1 - x^2)}{2}.$$

Hence, the support points of $\mathbf{p}^*$ are given by

$$x = \pm 1, \quad \pm 1/\sqrt{5} = \pm 0.447.$$

Thus, we obtain the D -optimal design for the cubic regression model as:

$$\mathbf{p}^* = \begin{Bmatrix} -1 & -0.447 & 0.447 & 1 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}.$$

If the construction of optimal designs is not possible analytically, we can do this by using some algorithms. It is typically not possible to evaluate an explicit solution $\mathbf{p}^*$ or in particular to derive an optimal regression design explicitly. Iterative techniques must be employed and so special algorithms could be used for a constrained optimization problem which requires the calculation of an

optimizing probability distribution. If the design space is continuous, then before we run an algorithm we need to discretize the design space.

2.5.1 A class of algorithms

Problem (2.16) involves a specific set of constraints: the weights $p_1, p_2, \dots, p_J$ must all be nonnegative and sum to one. An iterative procedure which automatically satisfies these constraints and has many nice properties can be written as:

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j^{(r)}, \delta)}{\sum_{i=1}^J p_i^{(r)} f(x_i^{(r)}, \delta)}, \quad (2.23)$$

where $x_i^{(r)}$ denotes either $d_j^{(r)}$ or $F_j^{(r)}$, and

$$\begin{aligned} d_j^{(r)} &= \left. \frac{\partial \phi}{\partial p_j} \right|_{\mathbf{p}=\mathbf{p}^{(r)}}, \\ F_j^{(r)} &= d_j^{(r)} - \sum_{i=1}^J p_i^{(r)} d_i^{(r)} \quad (\text{vertex directional derivatives at } \mathbf{p} = \mathbf{p}^{(r)}). \end{aligned}$$

The function $f(\cdot, \cdot)$ is assumed to satisfy:

- (i) $f(\cdot, \cdot) > 0$ for all arguments;
- (ii) $f(\cdot, \cdot)$ is strictly increasing in its argument;
- (iii) $f(\cdot, \cdot)$ may depend on a tuning parameter δ .

Such multiplicative updates go back to Torsney (1977) who first introduced this type of algorithm, and considered $f(d, \delta) = d^\delta$ with $\delta > 0$. Later studies explored variants: Silvey et al. (1978) investigated the effect of δ for

$f(d, \delta) = d^\delta = d^\delta$; Torsney (1988) examined $f(d, \delta) = d^\delta = e^{\delta d}$ to handle cases where $\phi(\cdot)$ has negative derivatives; Mandal and Torsney (2000) discussed systematic choices of $f(\cdot, \cdot)$; Torsney (1983) established monotonicity properties; and Mandal et al. (2005); Torsney and Mandal (2001) applied such iterations to designs with additional constraints. When the optimal design includes many zero weights, some modification is required; Mandal and Torsney (2006) proposed a clustering-based algorithm. Yu (2010) proved the monotonic convergence of the general class of the multiplicative algorithm.

We now review key properties of the directional derivatives which guide the choice of $f(\cdot, \cdot)$.

(a) Feasibility of iterates: Because we always start from an initial design in $\mathcal{P}$ and $f(\cdot, \cdot)$ is positive, the update in (2.23) preserves the nonnegativity and unit-sum constraints. In other words, each iterate $\mathbf{p}^{(r)}$ remains feasible.

(b) Nonnegativity of the directional derivative towards next iteration: Let $F_\phi\{\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}\}$ denote the directional derivative of $\phi(\cdot)$ at $\mathbf{p}^{(r)}$ toward $\mathbf{p}^{(r+1)}$. We can show that

$$F_\phi\{\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}\} \geq 0.$$

This can be seen by introducing a random variable D taking value d_j with probability $p_j = p_j^{(r)}$. Then

$$F_\phi\{\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}\} = \frac{\text{Cov}[D, f(D, \delta)]}{\mathbb{E}[f(D, \delta)]}. \quad (2.24)$$

Proof:

$$\begin{aligned}
F_\phi\{\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}\} &= [\mathbf{p}^{(r+1)} - \mathbf{p}^{(r)}]^\top \mathbf{d} \\
&= \sum_{i=1}^J (p_i^{(r+1)} - p_i^{(r)}) d_i \\
&= \sum_{i=1}^J p_i^{(r+1)} d_i - \sum_{i=1}^J p_i^{(r)} d_i \\
&= \frac{\sum_{i=1}^J p_i f(d_i, \delta) d_i}{\sum_{i=1}^J p_i f(d_i, \delta)} - \sum_{i=1}^J p_i d_i \\
&= \frac{[\sum_i p_i f(d_i, \delta) d_i] - [\sum_i p_i d_i] [\sum_i p_i f(d_i, \delta)]}{\sum_i p_i f(d_i, \delta)} \\
&= \frac{\text{Cov}[D, f(D, \delta)]}{\mathbb{E}[f(D, \delta)]}.
\end{aligned}$$

Since $f(D, \delta)$ is increasing in D , the covariance is nonnegative, ensuring a nonnegative directional derivative and hence an improvement of the criterion.

(c) Fixed-point property: An iterate $\mathbf{p}^{(r)}$ is a fixed point of (2.23) if the derivatives $\partial\phi/\partial p_j^{(r)}$ are equal for all nonzero components, equivalently if the corresponding vertex directional derivatives $F_j^{(r)}$ are zero. Thus, any local maximizer of (2.16) is a fixed point of the iteration.

To accelerate convergence, it is useful to adopt objective choices for $f(F, \delta)$. Since any criterion produces both positive and negative vertex directional derivatives and $\sum p_j F_j = 0$, a suitable $f(F, \delta)$ should be centered around zero and change rapidly near $F = 0$. Using F_j as the first argument of $f(\cdot, \cdot)$ can thus improve convergence. The convergence properties depend jointly on the criterion function $\phi(\cdot)$, the form of $f(\cdot, \cdot)$, and the parameter δ . In later chapters we explore different choices of $f(\cdot, \cdot)$ and its arguments for constructing optimal designs.

Now we consider some examples and construct some D -optimal designs. In simple linear regression, the model is

$$E(y|x) = \theta_0 + \theta_1 x = \mathbf{v}_x^\top \boldsymbol{\theta}$$

where $\mathbf{v}_x = (1, x)^\top, x \in [-1, 1]$ and $\boldsymbol{\theta} = (\theta_0, \theta_1)^\top$.

$\mathbf{v}_x \in \mathcal{V} = \{\mathbf{v}_x : \mathbf{v}_x = (1, x)^\top, -1 \leq x \leq 1\}$, the induced design space.

We approximate the design interval by a grid of 201 points equally spaced at intervals of 0.01 between -1 and 1 . We take $f(d, \delta) = d^\delta$ for some value of δ , say 0.5. We start with the initial values of $\mathbf{p}$ as the equal weights, that is, $1/201$ each. Then we run the algorithm until the first-order conditions (2.19) are satisfied. The algorithm converges to the following optimal design for the simple linear regression model

$$\mathbf{p}^* = \begin{Bmatrix} -1 & 1 \\ 0.5 & 0.5 \end{Bmatrix}.$$

We construct the optimal design by considering some other choices of the function $f(d, \delta)$ (subject to the function being positive and increasing) and obtained the same solution as above. We construct optimal designs for the quadratic regression model. The model is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 = \mathbf{v}_x^\top \boldsymbol{\theta}$$

where $\mathbf{v}_x = (1, x, x^2)^\top, x \in [-1, 1]$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^\top$. We approximate the design interval by a grid of 201 points equally spaced at intervals of 0.01 between -1 and 1 . We take $f(d, \delta) = e^{d\delta}$ for some value δ , say 1.2. The algorithm

converges to the following optimal design:

$$\mathbf{p}^* = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}.$$

Note that this is the same optimal design that we obtained analytically before.

Now we obtain the expressions for the standardized variance of the predicted response (also called the variance function) for the above designs and confirm that the optimal designs are D -optimal designs. The standardized variance of the predicted response is given by (1.3).

Simple linear regression:

We obtain the information matrix at optimal design as given by

$$\begin{aligned} \mathbf{M}(\mathbf{p}) &= \sum_{j=1}^2 p_j \mathbf{v}_j \mathbf{v}_j^\top \\ &= \frac{1}{2} \sum_{j=1}^2 \mathbf{v}_j \mathbf{v}_j^\top \\ &= \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} d(x, \mathbf{p}) &= \mathbf{f}^\top(x) \mathbf{M}^{-1} \mathbf{f}(x) \\ &= \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \\ &= 1 + x^2. \end{aligned}$$

The plot of the variance function is given in Figure (2.1). The maximum value of 2 occurs at the two support points -1 and 1. This confirms the D -optimality of the above regression model. The plot visually confirms this finding, as the variance function reaches its maximum at the design's support points, which is a key requirement of the General Equivalence Theorem for D -optimality.

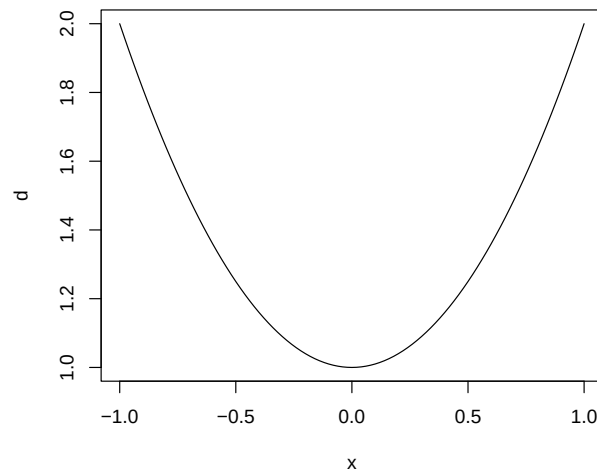


Figure 2.1: Simple linear Regression: Variance function $d(x, p^*)$ at the optimum design. The maximum value of 2 occurs at the support points.

Quadratic regression:

We obtain the information matrix at optimal weights as given by

$$\begin{aligned}
 M(\mathbf{p}) &= \sum_{j=1}^3 p_j \mathbf{v}_j \mathbf{v}_j^\top \\
 &= \frac{1}{3} \sum_{j=1}^3 \mathbf{v}_j \mathbf{v}_j^\top \\
 &= \frac{1}{3} \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}.
 \end{aligned}$$

The inverse information matrix

$$M^{-1}(\mathbf{p}) = 3 \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{pmatrix}.$$

So,

$$\begin{aligned}
 d(x, \mathbf{p}) &= \mathbf{f}^\top(x) M^{-1} \mathbf{f}(x) \\
 &= \begin{pmatrix} 1 & x & x^2 \end{pmatrix} 3 \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \\
 &= 3 - \frac{9}{2}x^2 + \frac{9}{2}x^4.
 \end{aligned}$$

The plot of the variance function is given in Figure (2.2). The maximum value of 3 occurs at the three support points -1, 0 and 1, confirming the D -

optimality of the above regression model. The plot also illustrates the variance function reaching its maximum at the three design points, further confirming the D -optimality of this design.

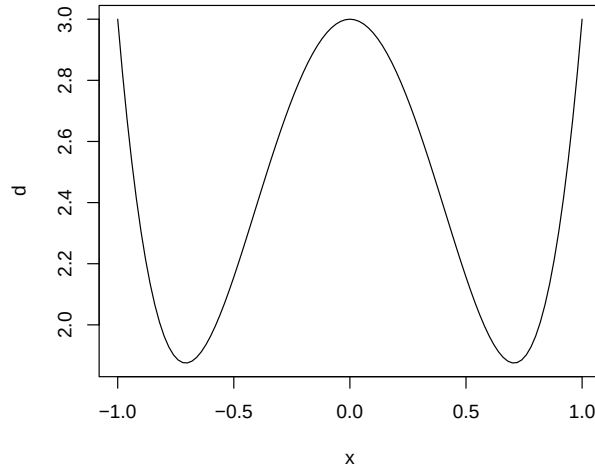


Figure 2.2: Quadratic Regression: Variance function $d(x, p^*)$ at the optimum design. The maximum value of 3 occurs at the support points.

2.5.2 Choice and Scaling of Design Points

In practice, the first step in constructing optimal designs often involves selecting and scaling a set of candidate design points. Traditionally, this step is performed manually, by specifying a small number of representative points and then rescaling them to a standard design region such as $[-1, 1]$. For example, Berger and Wong (2009) illustrate how a set of eight equally spaced points can be linearly transformed to lie within the interval $[-1, 1]$. For instance, in the dosage example, the eight design points $\{1, 2, \dots, 8\}$ are rescaled to the standard-

ized values $\{-1, -0.7143, -0.4286, -0.1429, 0.1429, 0.4286, 0.7143, 1\}$ using the transformation

$$d'_j = \frac{2d_j - (d_{\max} + d_{\min})}{d_{\max} - d_{\min}}, \quad -1 \leq d'_j \leq 1. \quad (2.25)$$

This transformation is a linear rescaling of the interval $[d_{\min}, d_{\max}]$:

- The midpoint $\tilde{d} = (d_{\max} + d_{\min})/2$ is mapped to 0.
- The left endpoint $d_{\min}$ is mapped to -1 .
- The right endpoint $d_{\max}$ is mapped to $+1$.

Thus, this is exactly the affine transformation that linearly maps the original range $[d_{\min}, d_{\max}]$ onto the symmetric interval $[-1, 1]$. Every value d_j is expressed in this normalized coordinate system, centered at the midpoint and scaled by half the range. Such manual scaling ensures interpretability and reduces computational effort, but it restricts the design space to a limited and predetermined set of candidates.

With modern computational resources, it is no longer necessary to restrict attention to such a small set of manually chosen points. Instead, we may construct a fine grid of candidate design points over the entire region of interest. For instance, if the design region is $[-1, 1]$ and we use a step size of 0.01, the candidate set will consist of 201 equally spaced points. If we use a step size of 0.1, the candidate set will consist of 21 equally spaced points. This algorithmic approach has two major advantages. First, it greatly increases flexibility, as the true optimal design is sought over a much richer set of potential points. Second, it allows the same algorithm to be applied regardless of the model or design space, avoiding the

need for ad hoc manual selection.

Thus, while manual scaling of a small number of design points remains useful for pedagogical purposes and simple illustrations, algorithmic scaling via dense grids provides a more systematic and powerful foundation for modern optimal design computations. In subsequent chapters, we shall make use of this powerful grid-based strategy, which ensures that the proposed multiplicative algorithms can search over the full design region in a uniform manner.

2.6 An application - Radiation dosage for tumour shrinkage

2.6.1 Introduction

We consider the radiation dosage model of Berger and Wong (2009). A radiologist is interested in the linear effect of radiation dosage (denoted by x) on tumour shrinkage (denoted by y). A simple linear model is assumed for studying the effect of radiation dosage on the reduction of tumours in 16 breast cancer patients. Assuming that there are no ethical objections to assigning $n = 16$ breast cancer patients to eight radiation-dosage levels, the following four designs are considered.

d_j	Design 1	Design 2	Design 3	Design 4
1	2	2	3	8
2	1	2	2	0
3	2	2	2	0
4	3	2	1	0
5	4	2	1	0
6	1	2	2	0
7	1	2	2	0
8	2	2	3	8
n	16	16	16	16
SS_x	70	84	108	196
$\text{var}(\hat{\theta}_1)$	0.0143	0.0119	0.0093	$0.0051(\times \sigma^2)$
$\text{var}(\hat{\theta}_0)$	0.3441	0.3036	0.25	$0.1658(\times \sigma^2)$

Equation (1.2) tells us that when both the sample size n and error variance σ^2 are fixed and the variance of the dosage levels $\text{var}(x)$ is maximized, the variance $\text{var}(\hat{\theta}_1)$ will be minimal. This is the same as saying $\text{var}(\hat{\theta}_1)$ is minimized when the sum of squares of x , say SS_x , is maximized. The above table shows how the sum of squares SS_x and the variance $\text{var}(x)$ influences the variance $\text{var}(\hat{\theta}_1)$ using the four designs all with $n = 16$ patients. Since Design 4 has the largest value of $\text{var}(x) = 196/16 = 12.25$, Design 4 is the optimal design for estimating the slope parameter.

2.6.2 Evaluating the criteria

We consider the linear model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

where the data vector $\mathbf{Y}$ is modelled as the product of a design matrix $\mathbf{X}$ and a vector of regression parameters $\boldsymbol{\theta} = (\theta_0, \theta_1)^\top$, and $\boldsymbol{\epsilon}$ is an $n \times 1$ error vector.

Design matrices for the four radiation dosage designs are given by:

Design 1	Design 2	Design 3	Optimal Design
$\mathbf{X}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 8 \end{bmatrix}$	$\mathbf{X}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 6 \\ 1 & 7 \\ 1 & 7 \\ 1 & 8 \\ 1 & 8 \end{bmatrix}$	$\mathbf{X}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \\ 1 & 6 \\ 1 & 7 \\ 1 & 7 \\ 1 & 8 \\ 1 & 8 \end{bmatrix}$	$\mathbf{X}_4 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \\ 1 & 8 \end{bmatrix}$

From equation (1.6), the least squares estimator $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

and the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ in matrix notation is

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Instead of considering the covariance matrix, we consider here the determinant of the information matrix for D -optimal design. Similarly, for A -optimality

we consider the negative trace of the inverse information matrix, $M^{-1}(\mathbf{p})$.

First, we obtain the information matrix for each design as given by

$$M(\mathbf{p}) = \sum_{j=1}^J p_j \mathbf{v}_j \mathbf{v}_j^{\top}$$

where $J = 8$ as we have 8 distinct points, $p_j = n_j/n$.

Design 1:

$$M(\mathbf{p}) = \begin{pmatrix} 1 & 4.4375 \\ 4.4375 & 24.0625 \end{pmatrix}.$$

Using the D - and A -optimality criteria, we obtain

D-criterion: 4.371094

A-criterion: -5.733691.

Design 2:

$$M(\mathbf{p}) = \begin{pmatrix} 1 & 4.5 \\ 4.5 & 25.5 \end{pmatrix}.$$

Using the D - and A -optimality criteria, we obtain

D-criterion: 5.25

A-criterion: -5.047619.

Design 3:

$$M(\mathbf{p}) = \begin{pmatrix} 1 & 4.5 \\ 4.5 & 27 \end{pmatrix}.$$

Using the D - and A -optimality criteria, we obtain

D -criterion: 6.75

A -criterion: -4.148148.

Optimal design (Design 4):

$$M(\mathbf{p}) = \begin{pmatrix} 1 & 4.5 \\ 4.5 & 32.5 \end{pmatrix}.$$

Using the D - and A -optimality criteria, we obtain

D -criterion: 12.25

A -criterion: -2.734694.

As we can see, Design 4 (optimal design) has the largest determinant of the information matrix for D -optimality. Similarly in terms of trace, Design 4 is the A -optimal design. This means that the parameters θ_0 and θ_1 can simultaneously be estimated as efficiently as possible if the design points were chosen at both ends of the x variable (dosage levels) (that is, $\text{Var}(x)$ is maximized) and if the observations would be equally divided over the two design points.

2.6.3 Finding the optimal dosage and weights using the algorithm

As discussed in the previous section, we can obtain the D -optimal design in a more general and powerful way (confirming optimality using the directional derivatives and the general equivalence theorem) rather than restricting attention to a small set of candidate designs (e.g., only four design points). Using the rescaling transformation (2.25), the design points can be standardized to the set

$$\{-1, -0.7143, -0.4286, -0.1429, 0.1429, 0.4286, 0.7143, 1\}.$$

Based on these eight candidate points, we apply algorithm (2.23) to compute the D -optimal design.

To implement the algorithm, we require a function $f(x, \delta)$ that is positive, strictly increasing in x , and depends on a free parameter δ . In this dissertation, we consider three families: the exponential function $e^{\delta x}$, the standard normal c.d.f., and the logistic c.d.f. $e^{\delta x}/(1 + e^{\delta x})$, where x denotes either a partial or directional derivative. For the present example, we use the exponential function $e^{\delta x}$ with $\delta = 0.5$, taking x to be the partial derivative.

After 70 iterations, the algorithm converges to the following design:

$$p^* = \begin{pmatrix} -1 & -0.7143 & -0.4286 & -0.1429 & 0.1429 & 0.4286 & 0.7143 & 1 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & -0.490 & -0.816 & -0.980 & -0.980 & -0.816 & -0.490 & 0 \end{pmatrix}.$$

The first row gives the candidate design points, the second row the corresponding design weights, and the third row the values of the first-order conditions for $\phi(\mathbf{p}^*)$ to be a local maximum in the feasible region (see 2.19). These conditions are satisfied, confirming the optimality of the solution.

Next, we consider a finer discretization of the design space $[-1, 1]$ into 21 equally spaced points with step size 0.1. Using the same exponential multiplicative function as above, we again apply algorithm (2.23) to find the D -optimal design.

After 172 iterations, the optimal design is given by:

Optimal Design ($J = 21$ candidate points)

$$\mathbf{p}^* = \left\{ \begin{array}{cccccccccc} -1.0 & -0.9 & -0.8 & -0.7 & -0.6 & -0.5 & -0.4 & -0.3 & -0.2 & -0.1 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.00 & -0.19 & -0.36 & -0.51 & -0.64 & -0.75 & -0.84 & -0.91 & -0.96 & -0.99 \\ 0.0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1.00 & -0.99 & -0.96 & -0.91 & -0.84 & -0.75 & -0.64 & -0.51 & -0.36 & -0.19 \\ 1.0 & & & & & & & & & \\ 0.5 & & & & & & & & & \\ 0.00 & & & & & & & & & \end{array} \right\}.$$

Again, the optimal design places all its weight on the boundary points -1 and 1 , with equal weights of 0.5 . The third row shows the corresponding vertex directional derivatives, which are negative for all inactive points and zero at

the two active support points. This verifies that the D -optimal for the radiation-dosage example assigns half weight at the minimum and maximum dosage levels.

Chapter 3

Minimizing variance based criteria

3.1 Introduction

One of the central objectives in the theory of optimal experimental design is to minimize the variance of the parameter estimators. Variance reflects the degree of uncertainty in parameter estimation: the smaller the variance, the more precise the estimates. In the design of experiments, where the allocation of experimental resources is under the researcher's control, the ability to strategically choose design points to minimize estimator variance plays a critical role in improving the reliability and efficiency of statistical inference.

As mentioned earlier, under least squares estimation of a linear regression model, the covariance matrix of the estimators is proportional to the inverse of the information matrix, which depends on the design points. Hence, minimizing variance amounts to maximizing information, which can be achieved by selecting design points that make the information matrix as large as possible in some

sense. This leads to well-established optimality criteria such as A -optimality (minimizing the trace of the inverse of the information matrix), and c -optimality (minimizing the variance of the estimator of a particular parameter or a specific linear combination of the parameters). Various researchers have worked on variance criteria, including López-Fidalgo (2023); Chowdhury et al. (2020); Berger and Wong (2009); Pukelsheim (2006); Fellman (1989) and among others.

Motivated by this, we explore a class of optimization problems in which we consider some variance-based criteria and construct such optimal designs. First, instead of considering directly an A -optimal design, that is, by minimizing the average or total variance of all of the parameter estimators, we minimize the average or total variance of a particular set of parameter estimators that are of interest for the research problem. Second, we establish measures of efficiency that enable meaningful comparisons across competing designs based on several parameters. This enables the evaluation of subset-focused designs in comparison to traditional A - and D -optimal designs, offering insight into the trade-offs between global optimality and targeted precision for the selected parameters. Third, we consider another class of problems where we need to treat two parameters (or linear parametric functions) equally due to their correspondence with two similar design variables. In such situations, our goal is to construct optimal designs that achieve equality of variances of the estimators of those parameters or parametric functions. This chapter explores the methodology for developing and constructing these optimal designs.

3.1.1 The General Equivalence Theorem and its Implications

Optimal design theories including the General Equivalence Theorem rely on a set of fundamental concepts, including the directional derivatives (2.17), and Gâteaux derivative (2.18). These are generalization of the concept of a derivative to functions defined on a vector or matrix space, which allows us to find a directional derivative. It helps us understand how a criterion function, such as the one we seek to maximize, changes as we move from one design to another. This is crucial for iterative algorithms that aim to find a better design in each step.

A cornerstone of optimal design theory is the General Equivalence Theorem Kiefer (1959). This theorem provides the necessary and sufficient conditions for a design to be optimal, and it links the three key concepts of a design problem:

1. The optimality of a design, p .
2. The maximization of a criterion function, $\phi(p)$.
3. The behavior of the directional derivatives, or the variance function $d(x, p)$.

In the context of D -optimality, the theorem states that a design p^* is D -optimal if and only if the maximum value of the standardized variance function, $d(x, p^*)$, over the design space is equal to the number of parameters, k . Furthermore, the maximum occurs at the support points of the optimal design.

While the theorems and proofs for non-standard criteria are more complex, the core principle remains similar. The iterative algorithms we use are designed

to find a solution where the directional derivatives are zero at the support points, satisfying the first-order conditions. If the concavity is lacking we check for second-order conditions. The behavior of the directional derivatives over the entire design space provides a powerful diagnostic tool to check the optimality of the converged solution. The examples in this chapter will visually and numerically confirm this principle, demonstrating how our solutions adhere to these foundational theoretical requirements.

3.2 Variance based criteria for minimizing the average or total variance

Instead of the A -optimality criterion, an optimization problem may involve minimizing the average or total variance of the estimators of a subset of parameters of interest. Note that minimizing the average variance is the same as minimizing the total variance.

Suppose that we are interested in a set of s ($s \leq k$) parameters or linear parametric functions in the model. This is particularly relevant in situations where not all parameters are of equal importance. For example, in a polynomial regression, the higher-order terms may be of specific interest, while the intercept and first-order terms serve as a baseline. This optimization problem addresses the above by focusing on minimizing the total variance of the estimators of these selected parameters.

Let us consider s parametric functions $\mathbf{a}_i^\top \boldsymbol{\theta}$, $i = 1, 2, \dots, s$. We would like

to minimize $\sum_{i=1}^s \text{Var}(\mathbf{a}_i^\top \hat{\boldsymbol{\theta}})$. Since the variance of the estimator of a linear function of parameters is given by $\text{Var}(\mathbf{a}_i^\top \boldsymbol{\theta}) = \mathbf{a}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}_i$, our objective is to minimize $\sum_{i=1}^s \mathbf{a}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}_i$. This is equivalent to maximizing the negative sum of these variances. Let $\phi_i(\mathbf{p}) = -\mathbf{a}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}_i$, $i = 1, 2, \dots, s$. Maximizing $\sum_{i=1}^s \phi_i(\mathbf{p})$ will give us the optimal design for minimizing the average or total variance of the estimators of $\mathbf{a}_i^\top \boldsymbol{\theta}$, $i = 1, 2, \dots, s$. Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_s]$. Therefore, we would like to maximize the criterion

$$\phi(\mathbf{p}) = -\text{tr}(\mathbf{A}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A}). \quad (3.1)$$

We call this A_s -optimality criterion. In order to construct the optimal design, we need to determine the partial and directional derivatives of this criterion function. We derive this in the following theorem.

Theorem 3.2. For $\phi(\mathbf{p}) = -\text{tr}(\mathbf{A}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A})$, where $\mathbf{A}$ is a matrix of order $k \times s$, $s < k$, and $\text{rank}(\mathbf{A}) = s$, d_j^ϕ , $j = 1, 2, \dots, s$, are given by

$$d_j^\phi = \mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{A} \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{v}_j.$$

Proof:

Let $h(\mathbf{M}) = \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{A}$. As $\phi(\mathbf{p})$ is also a function of $\mathbf{M}(\mathbf{p})$, we can write our criterion function as

$$\phi(\mathbf{p}) = \psi(\mathbf{M}(\mathbf{p})), \quad \mathbf{M}(\mathbf{p}) = \sum_j p_j \mathbf{v}_j \mathbf{v}_j^\top$$

with

$$\psi(M) = -\text{tr}(h(M)).$$

The partial derivatives of ψ can be derived by using the Gâteaux derivative (2.18) that was defined earlier. So,

$$G_\psi\{M, N\} = -\text{tr}(G_h\{M, N\}). \quad (3.2)$$

From the definition of Gâteaux derivative

$$G_h\{M, N\} = \lim_{\varepsilon \downarrow 0} \frac{h(M + \varepsilon N) - h(M)}{\varepsilon} \quad (3.3)$$

$$= \lim_{\varepsilon \downarrow 0} \frac{A^\top (M + \varepsilon N)^{-1} A - A^\top M^{-1} A}{\varepsilon}. \quad (3.4)$$

Now

$$\begin{aligned} (M + \varepsilon N)^{-1} &= (I + \varepsilon M^{-1} N)^{-1} M^{-1} \\ &= M^{-1/2} (I + \varepsilon M^{-1/2} N M^{-1/2})^{-1} M^{-1/2} \\ &= M^{-1/2} (I + \varepsilon B)^{-1} M^{-1/2} \end{aligned} \quad (3.5)$$

where $B = M^{-1} N$ or $M^{-1/2} N M^{-1/2}$.

$$\begin{aligned} (I + \varepsilon B)^{-1} &= I - \varepsilon B + \varepsilon^2 B^2 - \varepsilon^3 B^3 + \varepsilon^4 B^4 - \dots \\ &= I - \varepsilon B \quad \text{for small } \varepsilon. \end{aligned} \quad (3.6)$$

Thus by using (3.5) and (3.6), $h(\mathbf{M} + \varepsilon \mathbf{N})$ in (3.3) can be written as

$$\begin{aligned}
h(\mathbf{M} + \varepsilon \mathbf{N}) &= \mathbf{A}^\top (\mathbf{M} + \varepsilon \mathbf{N})^{-1} \mathbf{A} \\
&= \mathbf{A}^\top \mathbf{M}^{-1/2} (\mathbf{I} - \varepsilon \mathbf{B}) \mathbf{M}^{-1/2} \mathbf{A} \\
&= \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{A} - \varepsilon \mathbf{A}^\top \mathbf{M}^{-1/2} \mathbf{B} \mathbf{M}^{-1/2} \mathbf{A} \\
&= \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{A} - \varepsilon \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{N} \mathbf{M}^{-1} \mathbf{A}.
\end{aligned}$$

Hence from (3.3) and (3.2)

$$G_h\{\mathbf{M}, \mathbf{N}\} = -\mathbf{A}^\top \mathbf{M}^{-1} \mathbf{N} \mathbf{M}^{-1} \mathbf{A} \quad (3.7)$$

$$\text{and } G_\psi\{\mathbf{M}, \mathbf{N}\} = -\text{tr}(-\mathbf{A}^\top \mathbf{M}^{-1} \mathbf{N} \mathbf{M}^{-1} \mathbf{A}). \quad (3.8)$$

Then from (3.8) with $\mathbf{M} = \mathbf{M}(p)$,

$$\begin{aligned}
\frac{\partial \phi}{\partial p_j} &= G_\psi\{\mathbf{M}, \mathbf{v}_j \mathbf{v}_j^\top\} \\
&= -\text{tr}(-\mathbf{A}^\top \mathbf{M}^{-1} \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{A}) \\
&= \mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{A} \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{v}_j.
\end{aligned}$$

Therefore, the corresponding vertex directional derivatives will be

$$F_j^\phi = \mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{A} \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{v}_j - \sum_{i=1}^J p_i \mathbf{v}_i^\top \mathbf{M}^{-1} \mathbf{A} \mathbf{A}^\top \mathbf{M}^{-1} \mathbf{v}_i. \quad (3.9)$$

We now consider an example of the quadratic regression model:

$$E(Y|X) = \theta_1 + \theta_2 x + \theta_3 x^2$$

with the design interval $[-1, 1]$. Suppose that we are interested in minimizing the total variance of the estimators of the parameters θ_2 and θ_3 . We then maximize $\phi(\mathbf{p}) = \phi_2(\mathbf{p}) + \phi_3(\mathbf{p})$ subject to the constraints $p_j \geq 0$ and $\sum p_j = 1$. So, here, $\mathbf{A} = [\mathbf{a}_2, \mathbf{a}_3]$, where $\mathbf{a}_2 = (0, 1, 0)^\top$, and $\mathbf{a}_3 = (0, 0, 1)^\top$. The design space is continuous. We discretize the design space to form a uniform grid on the continuous design space. In particular, we approximate the design interval by a grid of points equally spaced at intervals of 0.1, giving us $J = 21$. We take the initial design to be $p_j^{(0)} = 1/J, j = 1, 2, \dots, J$. We take $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, where x is the vertex directional derivatives of the criterion and $\delta = 0.1$. Starting with equal initial weights, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -12.08862, \\ -9.40704 &\leq F_j^\phi \leq 30.5007.\end{aligned}$$

After 4000 iterations, the algorithm converges to a solution.

$$\begin{aligned}\phi(\mathbf{p}) &= -5.828427, \\ -4.244622 &\leq F_j^\phi \leq 8.881784 \times 10^{-16}.\end{aligned}$$

The algorithm converges to a solution having three support points, namely -1.0, 0.0 and 1.0 with corresponding weights (0.2928932, 0.4142136, 0.2928932). The directional derivatives corresponding to the three support points are zero, and are negative at all the remaining design points with zero weights. This confirms that the directional derivatives meet the first-order conditions. We have found a design $\mathbf{p}$ for which the minimum total variance of the estimators of the

parameters θ_2 and θ_3 is 5.828427. This result demonstrates that the algorithm successfully finds the optimal design for this specific criterion and model.

Now we are going to see practical problem arising in Chemistry. The model was provided by chemistry in Scotland, considered by Torsney and Alahmadi (1995). They studied a three-parameter regression model that describes the relationship between the viscosity and the concentration (x) of a chemical solution. The regression model describes the relationship between the viscosity Y and the concentration x of a chemical solution. Viscosity is the response. The model is

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2.$$

The design interval is restricted to $(0, 0.2]$. First, note the design interval $(0, 0.2]$ is very short here. The chemists who supplied this model were more concerned with the terms $x^{1/2}$ and x^2 . We construct designs for minimizing the average or sum of the variances of the estimates of the corresponding parameters θ_2 and θ_3 . So we maximize $\phi(\mathbf{p}) = -\text{tr}(\mathbf{A}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A})$ subject to the constraints $p_j \geq 0$ and $\sum p_j = 1$, where $\mathbf{A} = [\mathbf{a}_2, \mathbf{a}_3]$, $\mathbf{a}_2 = (0, 1, 0)^\top$, and $\mathbf{a}_3 = (0, 0, 1)^\top$. According to the above design space, we approximate the design interval by a grid of points equally spaced at intervals of 0.01 between 0.02 and 0.2. By making an initial run of the algorithm (2.23) in this discretized space. We run algorithm with $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, where x are the vertex directional derivatives of the criterion with $\delta = 0.00001$.

Starting with equal initial weights, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -284606.8, \\ -283454.1 &\leq F_j^\phi \leq 713372.7,\end{aligned}$$

After 2000 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -121565.6, \\ -119365.6 &\leq F_j^\phi \leq 8.731149 \times 10^{-10},\end{aligned}$$

for $1 \leq i \leq J = 19$. We found that the algorithm converged to a solution with three active support points: 0.02, 0.12, and 0.20. The corresponding optimal weights for these points were determined to be (0.3486168, 0.4288118, 0.2225714). The directional derivatives corresponding to the three support points are zero, and are negative at all the remaining design points with zero weights. This confirms that the directional derivatives meet the first-order conditions. We have found a design measure $\mathbf{p}$ for which the minimum total variance of the estimators of the parameters θ_2 and θ_3 is 121565.6. This result demonstrates that we have found the optimal design for this specific criterion and specific model.

It is quite natural to think that D_s -optimality will be the counterpart of A_s -optimality. So we calculate the cross-efficiencies between these two optimality criteria. Before we do that we need to construct the D_s -optimal designs.

We have already discussed D and D_A -optimality (2.1) and (2.9) in section 2.2. D -optimality is concerned with estimating all k parameters of a model, whereas D_A -optimality is used when we are interested in estimating s linear

combinations of the parameters $\theta_1, \theta_2, \dots, \theta_k$. These s linear combinations can be expressed as the vector $\alpha = A\theta$, where A is a $s \times k$ matrix of rank $s \leq k$. D_s -optimality is an important special case of D_A -optimality, used when the interest lies in estimating a specific subset of s parameters. This criterion arises in D_A -optimality when the matrix A is given by $A = [I_s : O]$, where I_s is the identity matrix with dimension $s \times s$ and O is the zero matrix with dimension $s \times (k - s)$. In this situation, without loss of generality, our objective is to estimate only the first s parameters, from θ_1 to θ_s , in the parameter vector θ . Since $A = [I_s : O]$, we partition the information matrix $M(p)$ as:

$$M(p) = \begin{bmatrix} M_{11}^{s \times s} & M_{12}^{s \times (k-s)} \\ M_{12}^\top & M_{22}^{(k-s) \times (k-s)} \end{bmatrix}. \quad (3.10)$$

Then the matrix inverse matrix of $AM^{-1}(p)A^\top$ can be expressed as $(M_{11} - M_{12}M_{22}^{-1}M_{12}^\top)$ (Rohde, 1965). Therefore, in this special case, maximizing ϕ_{D_A} in (2.9) in this particular case is equivalent to maximizing the following D_s -optimality criterion function:

$$\phi_{D_s}(p) = \log \det\{(M_{11} - M_{12}M_{22}^{-1}M_{12}^\top)\}. \quad (3.11)$$

For further details, we refer to Karlin and Studden (1966); Atwood (1969); Silvey and Titterton (1973); Silvey (1980).

We will now apply the D_s -optimality criterion to the same quadratic regres-

sion model used for the A_s -optimality example:

$$E(y|x) = \theta_1 + \theta_2 x + \theta_3 x^2$$

with the design interval $[-1, 1]$. Our focus here is on finding the optimal design that maximizes the determinant of the sub-matrix corresponding to the parameters θ_2 and θ_3 . The regression vector is $(1, x, x^2)^\top$, and the sub-vector of interest is $\mathbf{v}_{j,2} = (x_j, x_j^2)^\top$. We discretize the design interval into a uniform grid of 21 points, from -1.0 to 1.0, and initialize all design weights to $p_j^{(0)} = 1/21$. Using the multiplicative algorithm (2.23), we take the logistic cdf $f(F, \delta) = \exp(\delta F)/(1 + \exp(\delta F))$, where F denotes the vertex directional derivatives of the criterion and $\delta = 0.3$. We find that the algorithm converges to a solution. After 3000 iterations,

$$\begin{aligned} \phi(\mathbf{p}) &= -1.909543, \\ -1.12455 &\leq F_j^\phi \leq 0. \end{aligned}$$

The optimal design has three support points at $x = -1.0, 0.0$, and 1.0 , with corresponding weights of $(1/3, 1/3, 1/3)$.

The directional derivatives for the support points are all equal to 2, which satisfies the equivalence theorem condition with $s = 2$. This contrasts with the A_s -optimal design for the same model, which had different weights at these same support points, highlighting a key difference between the two criteria.

Next, we apply the D_s -optimality criterion to the chemistry problem:

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2$$

with the design interval $(0.0, 0.2]$. The parameters of interest are again θ_2 and θ_3 . The sub-vector of interest for the directional derivative is $\mathbf{v}_{j,2} = (x_j^{1/2}, x_j^2)^\top$. We use the same discretized design space from the previous section, with points spaced at 0.01 between 0.02 and 0.2. We run the algorithm with $f(F, \delta) = \exp(\delta F)/(1 + \exp(\delta F))$, where F are the vertex directional derivatives of the criterion with $\delta = 0.4$.

After 4000 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -16.94491, \\ -1.670149 &\leq F_j^\phi \leq 2.708944 \times 10^{-14}. \end{aligned}$$

The algorithm converges to a solution with three active support points: 0.02, 0.10, and 0.20. The corresponding optimal weights are (0.4905948, 0.3380028, 0.1714025). The directional derivatives confirm optimality, as they are zero at the support points and negative at all other design points with zero weights.

The two examples demonstrate a crucial point in optimal design theory: the choice of criterion directly influences the resulting optimal design. While both A_s - and D_s -optimality focus on a subset of the parameters (θ_2 and θ_3 in our chemistry example), they lead to distinct optimal designs.

For the chemistry model, the A_s -optimal design had support points at 0.02, 0.12, and 0.20, with weights (0.3486, 0.4288, 0.2226). In contrast, the D_s -

optimal design had support points at 0.02, 0.10, and 0.20, with weights (0.491, 0.338, 0.171). While the endpoints (0.02 and 0.20) are common to both designs, the third, interior support point is different. This difference is not merely a minor rounding error; it reflects the fundamental disparity between the two criteria. Each criterion optimizes a different aspect of the estimation problem, leading to unique solutions.

While these optimal designs are “best” for their specific purpose, the question remains: how effective are they with respect to other criteria? For example, is an A_s -optimal design still highly efficient from a D_s perspective? The next section will address this by introducing the concept of efficiency to quantitatively compare how well a design, which is optimal for one criterion, performs under another. This allows us to assess the robustness of our designs and understand the trade-offs involved in criterion selection.

3.2.1 Cross-efficiencies

While we can determine which design is optimal for a specific criterion, it is also of great practical importance to evaluate how well that design performs with respect to other criteria. This is where the concept of cross-efficiency becomes essential. Cross-efficiency is a quantitative measure that evaluates how effectively a design, which is optimal for one criterion, performs under another. The efficiency of a design ξ_i with respect to a criterion ϕ is defined by comparing its performance to that of the optimal design for that criterion.

Mathematically, the efficiency is defined as:

$$\text{Eff}(\xi_i, \phi) = \frac{\text{Value of optimal design}}{\text{Value of evaluated design}}. \quad (3.12)$$

Efficiency is typically a value between 0 and 1, with a value closer to 1 indicating a more efficient design. For the A_s - and D_s -optimality criteria, the efficiency formulas are derived from their optimization objectives:

A_s -efficiency: This criterion minimizes the sum of the variances of a set of s estimators, i.e., it minimizes $\text{tr}(\mathbf{M}_s^{-1})$. Thus, a smaller trace value is better. The efficiency of a design ξ is calculated by dividing the trace of the optimal A_s design by the trace of the design ξ :

$$\text{Eff}_{A_s}(\xi) = \frac{\text{tr}(\mathbf{M}_{s,A_s}^{-1})}{\text{tr}(\mathbf{M}_{s,\xi}^{-1})}. \quad (3.13)$$

D_s -efficiency: This criterion maximizes the determinant of the information matrix corresponding to the specific parameters. Therefore, a larger determinant is better. The efficiency is calculated by taking the square root of the ratio of the determinants:

$$\text{Eff}_{D_s}(\xi) = \left(\frac{\det(\mathbf{M}_{s,\xi})}{\det(\mathbf{M}_{s,D_s})} \right)^{1/s}.$$

Here, s is the number of parameters of interest, which is $s = 2$ for both our examples.

Example 1: Cross-efficiencies for the Quadratic Regression Model

From our previous analysis of the quadratic model, we have:

A_s -optimal design: Support points at $(-1.0, 0.0, 1.0)$ with weights $(0.3, 0.4, 0.3)$.

D_s -optimal design: Support points at $(-1.0, 0.0, 1.0)$ with weights $(1/3, 1/3, 1/3)$. Let's now compute the cross-efficiencies to see which design is more robust.

The model is $E(y|x) = \theta_1 + \theta_2 x + \theta_3 x^2$ with the vector of basis functions $\mathbf{f}(x) = (1, x, x^2)^\top$. The parameters of interest are θ_2 and θ_3 . The information matrix $\mathbf{M}_{A_s}$ is calculated using the weights of the A_s -optimal design:

$$\mathbf{M}_{A_s} = 0.3 \mathbf{f}(-1) \mathbf{f}(-1)^\top + 0.4 \mathbf{f}(0) \mathbf{f}(0)^\top + 0.3 \mathbf{f}(1) \mathbf{f}(1)^\top$$

$$\mathbf{M}_{A_s} = 0.3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} (1, -1, 1) + 0.4 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (1, 0, 0) + 0.3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1, 1, 1)$$

$$\mathbf{M}_{A_s} = \begin{pmatrix} 1 & 0 & 0.6 \\ 0 & 0.6 & 0 \\ 0.6 & 0 & 0.6 \end{pmatrix}.$$

The parameters of interest are θ_2 and θ_3 . These correspond to the second and third rows/columns of the information matrix. We partition the matrix $\mathbf{M}_{A_s}$ into four sub-matrices:

$$\mathbf{M}_{A_s} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}$$

where

$$\begin{aligned} \mathbf{M}_{11} &= [1] \\ \mathbf{M}_{12} &= \begin{bmatrix} 0 & 0.6 \end{bmatrix} \\ \mathbf{M}_{21} &= \begin{bmatrix} 0 \\ 0.6 \end{bmatrix} \\ \mathbf{M}_{22} &= \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}. \end{aligned}$$

The A_s criterion value is calculated from the trace of the inverse of the partitioned information matrix, $\mathbf{M}_{s,A_s}$. We find it using the formula $\mathbf{M}_{s,A_s} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$. So we obtain

$$\mathbf{M}_{s,A_s} = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.24 \end{pmatrix}.$$

So, $\det(\mathbf{M}_{s,A_s}) = 0.144$ and $\text{tr}(\mathbf{M}_{s,A_s}^{-1}) = 5.8333$.

The D_s -optimal design for the parameters θ_2 and θ_3 in a quadratic model is supported at the points $x = \{-1, 0, 1\}$ with corresponding weights $(1/3, 1/3, 1/3)$. We construct the information matrix $\mathbf{M}_{D_s}$, using the basis function vector $\mathbf{f}(x) = (1, x, x^2)^\top$ and the weights of the D_s -optimal design.

$$\mathbf{M}_{D_s} = \frac{1}{3}\mathbf{f}(-1)\mathbf{f}(-1)^\top + \frac{1}{3}\mathbf{f}(0)\mathbf{f}(0)^\top + \frac{1}{3}\mathbf{f}(1)\mathbf{f}(1)^\top$$

$$\mathbf{M}_{D_s} = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{pmatrix}.$$

We partition the matrix $\mathbf{M}_{D_s}$ to isolate the information related to the parameters of interest, θ_2 and θ_3 .

$$\mathbf{M}_{D_s} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}.$$

The sub-matrices are of the same dimensions as above. We calculate the partitioned information matrix $\mathbf{M}_{s,D_s} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$.

$$\mathbf{M}_{s,D_s} = \begin{pmatrix} 2/3 & 0 \\ 0 & 2/9 \end{pmatrix}.$$

So, $\det(\mathbf{M}_{s,D_s}) = 4/27$ and $\text{tr}(\mathbf{M}_{s,D_s}^{-1}) = 6$.

A_s -efficiency is:

$$\text{Eff}_{A_s}(\xi_{D_s}) = \frac{5.8333}{6} \approx 0.9722 \quad (97.22\%).$$

D_s -efficiency is:

$$\text{Eff}_{D_s}(\xi_{A_s}) = \left(\frac{0.144}{4/27} \right)^{1/2} \approx 0.986 \quad (98.6\%).$$

These efficiency results show that both the A_s -optimal and D_s -optimal designs are highly efficient. The two designs perform very similarly, with only a small trade-off between minimizing average variance and maximizing joint precision.

Example 2: Cross-efficiencies for the Chemistry Model

For the Chemistry model, our previous results were:

A_s -optimal design: Support points at (0.02,0.12,0.20) with weights (0.3486, 0.4288, 0.2226).

D_s -optimal design: Support points at (0.02,0.12,0.20) with weights (0.491, 0.338, 0.171).

The model is $E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2$, and the parameters of interest are θ_2 and θ_3 . The basis function vector is $\mathbf{f}(x) = (x, x^{1/2}, x^2)^\top$, and the corresponding information matrix is $\mathbf{M}_{A_s}$.

The basis function vectors at the support points are:

$$\begin{aligned}\mathbf{f}(0.02) &= (0.02, \sqrt{0.02}, (0.02)^2)^\top = (0.02, 0.1414, 0.0004)^\top \\ \mathbf{f}(0.12) &= (0.12, \sqrt{0.12}, (0.12)^2)^\top = (0.12, 0.3464, 0.0144)^\top \\ \mathbf{f}(0.20) &= (0.20, \sqrt{0.20}, (0.20)^2)^\top = (0.20, 0.4472, 0.0400)^\top.\end{aligned}$$

The information matrix is computed by adding the outer products of these vectors, each scaled by its weight.

$$\begin{aligned}\mathbf{M}_{A_s} &= 0.3486 \mathbf{f}(0.02) \mathbf{f}(0.02)^\top + 0.4288 \mathbf{f}(0.12) \mathbf{f}(0.12)^\top + 0.2226 \mathbf{f}(0.2) \mathbf{f}(0.2)^\top \\ &= \begin{pmatrix} 0.015218 & 0.038691 & 0.002524 \\ 0.038691 & 0.102948 & 0.006139 \\ 0.002524 & 0.006139 & 0.000445 \end{pmatrix}.\end{aligned}$$

The matrix is partitioned with respect to the parameter of interest $\boldsymbol{\theta}_s = (\theta_2, \theta_3)$

and the nuisance parameter θ_1 . The sub-matrices M_{ij} have the same dimensions as in the previous example. The sub-matrices are:

$$\begin{aligned} M_{11} &= [0.015218] \\ M_{12} &= \begin{bmatrix} 0.038691 & 0.002524 \end{bmatrix} \\ M_{21} &= \begin{bmatrix} 0.038691 \\ 0.002524 \end{bmatrix} \\ M_{22} &= \begin{bmatrix} 0.102948 & 0.006139 \\ 0.006139 & 0.000445 \end{bmatrix}. \end{aligned}$$

Now we calculate the partitioned information matrix $M_{s,A_s} = M_{22} - M_{21}M_{11}^{-1}M_{12}$.

$$M_{s,A_s} = \begin{pmatrix} 0.004469 & -0.000272 \\ -0.000272 & 0.000028 \end{pmatrix}.$$

The determinant and trace of the inverse are: $\det(M_{s,A_s}) = 0.000000051148$ and $\text{tr}(M_{s,A_s}^{-1}) = 87920.8$. Calculating the D_s -optimality criterion value for the design with three support points $x = \{0.02, 0.10, 0.20\}$ and corresponding weights $(0.491, 0.338, 0.171)$. The basis function vector is $\mathbf{f}(x) = (x, x^{1/2}, x^2)^\top$,

and the corresponding information matrix is $\mathbf{M}_{D_s}$:

$$\begin{aligned} \mathbf{f}(0.02) &= (0.02, \sqrt{0.02}, (0.02)^2)^\top \approx (0.02, 0.1414, 0.0004)^\top \\ \mathbf{f}(0.10) &= (0.10, \sqrt{0.10}, (0.10)^2)^\top \approx (0.10, 0.3162, 0.0100)^\top \\ \mathbf{f}(0.20) &= (0.20, \sqrt{0.20}, (0.20)^2)^\top \approx (0.20, 0.4472, 0.0400)^\top. \end{aligned}$$

Using the given weights, the information matrix is computed as:

$$\begin{aligned} \mathbf{M}_{D_s} &= 0.491 \mathbf{f}(0.02) \mathbf{f}(0.02)^\top + 0.338 \mathbf{f}(0.10) \mathbf{f}(0.10)^\top + 0.171 \mathbf{f}(0.20) \mathbf{f}(0.20)^\top \\ &= \begin{pmatrix} 0.0104 & 0.0274 & 0.0017 \\ 0.0274 & 0.0778 & 0.0042 \\ 0.0017 & 0.0042 & 0.0003 \end{pmatrix}. \end{aligned}$$

The matrix is partitioned with respect to the parameter of interest $\boldsymbol{\theta}_s = (\theta_2, \theta_3)$ and the nuisance parameter θ_1 . Now we calculate the partitioned information matrix $\mathbf{M}_{s,D_s} = \mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^{-1} \mathbf{M}_{12}$.

$$\mathbf{M}_{s,D_s} = \begin{pmatrix} 0.0056 & -0.0003 \\ -0.0003 & 0.00003 \end{pmatrix}.$$

$\det(\mathbf{M}_{s,D_s}) = 0.000000078$ and $\text{tr}(\mathbf{M}_{s,D_s}^{-1}) = 456410.25$.

A_s -efficiency is:

$$\text{Eff}_{A_s}(\xi_{D_s}) = \frac{87920.8}{456410.25} \approx 0.1926355 \text{ or } 19.26\%.$$

D_s -efficiency is:

$$\text{Eff}_{D_s}(\xi_{A_s}) = \left(\frac{0.000000051148}{0.000000078} \right)^{1/2} \approx 0.81 \text{ or } 81\%.$$

In stark contrast to the quadratic model, the results for the chemistry model demonstrate a significant trade-off. The A_s -efficiency of the D_s -optimal design is extremely low at 19.26%. This means that a design optimized to minimize the confidence region volume is very inefficient at minimizing the sum of the variances. Conversely, the A_s -optimal design has a very high D_s -efficiency of 81%. It performs its intended task perfectly while still doing a great job at the other criterion. Therefore, for this model, the A_s -optimal design is clearly a superior and more robust choice.

The comparison between these two models highlights a crucial point: the relationship between different optimality criteria is highly dependent on the specific model being studied. In the quadratic model, the two criteria are largely aligned, and both designs are excellent. In the chemistry model, there is a strong trade-off, and one design is significantly better than the other across both criteria. This underscores the importance of not only calculating design efficiencies but also interpreting what these values mean in the context of the specific research goals.

3.3 Variance based criteria for equality of variances

In statistical inference, especially in the context of experimental design and estimation, the equality of variances of parameter estimators refers to situations where the variances (or mean squared errors) of estimators for different parameters are made as uniform as possible. This is closely linked to variance balance in optimal design, ensuring that no parameter is estimated with substantially less precision than another. A summary of the importance of equality of variances is as follows:

Fair precision across parameters: In regression or ANOVA-type models, researchers may wish to estimate multiple parameters simultaneously. If one parameter has a very small variance while another has a very large variance, the inference becomes unbalanced. Equalizing variances helps ensure all parameters are estimated with comparable reliability.

Robustness: Designs with equal (or nearly equal) variances often remain efficient across a broader range of true parameter values.

Interpretability: Equal precision simplifies comparison across parameters, as confidence intervals or hypothesis tests will not be distorted by heterogeneous uncertainty.

In regression design, one related criterion for balancing variances is E -optimality. The E -optimality criterion directly relates to balancing variances. It maximizes the minimum eigenvalue of the information matrix.

Suppose that the two linear functions of θ are $\mathbf{a}^\top \theta$ and $\mathbf{b}^\top \theta$, where $\mathbf{a}, \mathbf{b} \in R^k$.

Let $\mathbf{a}^\top \hat{\boldsymbol{\theta}}$ and $\mathbf{b}^\top \hat{\boldsymbol{\theta}}$ be the least squares estimators of $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$ respectively. We aim to construct a design that achieves equality of the variances of $\mathbf{a}^\top \hat{\boldsymbol{\theta}}$ and $\mathbf{b}^\top \hat{\boldsymbol{\theta}}$. The two variances are given by $v_1(\mathbf{p}) = \text{Var}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{a}$, $v_2(\mathbf{p}) = \text{Var}(\mathbf{b}^\top \hat{\boldsymbol{\theta}}) = \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{b}$. In order to convert this into a maximization problem, we set up the criterion as $\phi(\mathbf{p}) = -(v_1(\mathbf{p}) - v_2(\mathbf{p}))^2$, and maximize it over the constraint region $\mathcal{P} = \{\mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1\}$ (2.16). Note that $\phi(\mathbf{p}) \leq 0$. If the maximum value of zero is attained, then the variances of the estimators of the two linear functions of $\boldsymbol{\theta}$ are equal, and also the solution will generally form a hypersurface, defined by the equation $v_1(\mathbf{p}) - v_2(\mathbf{p}) = 0$, in a $(J - 1)$ -dimensional space. However, if no such design exists, maximizing this criterion will yield a design with the minimum numerical difference between the two variances. Note that the first- and second-order conditions should be used to confirm that the candidate point found by an algorithm is a local optimum.

In order to run the algorithm (2.23) and to find the directional derivatives, we need to find the partial derivatives of the criterion function. We derive the partial derivatives (d_j^ϕ) of $\phi(\mathbf{p})$ as follows. Let $V(\mathbf{p}) = v_1(\mathbf{p}) - v_2(\mathbf{p})$ then $\phi(\mathbf{p}) = -(V(\mathbf{p}))^2$. The partial derivatives of $\phi(\mathbf{p})$ with respect to p_j are $d_j^\phi = -2V(\mathbf{p})d_j^V$, where $d_j^V = \partial v_1(\mathbf{p})/\partial p_j - \partial v_2(\mathbf{p})/\partial p_j = -(\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{v}_j)^2 + (\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{v}_j)^2$.

If the criterion function is not concave, we need to check the first- and second-order conditions. The Hessian matrix of $\tilde{\phi}(\mathbf{p})$ is given by (2.20)

$$H(\mathbf{p}) = ((d_{ij}^{\tilde{\phi}})), i, j = 1, 2, \dots, J - 1$$

where $d_{ij}\tilde{\phi} = d_{ij}^{\phi} - d_{iJ}^{\phi} - d_{jJ}^{\phi} + d_{JJ}^{\phi}$, and $\partial^2\phi/\partial p_i\partial p_j = d_{ij}^{\phi} = -2d_i^V d_j^V - 2V(\mathbf{p})d_{ij}^V$,

where

$$\begin{aligned} d_{ij}^V &= 2\mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{a}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \\ &\quad - 2\mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \mathbf{b}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j. \\ &= 2\mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j [\mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{a}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j - \mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \mathbf{b}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j]. \end{aligned}$$

Thus d_{ij}^{ϕ} can be derived as

$$d_{ij}^{\phi} = -2d_i^V d_j^V - 4V(\mathbf{p}) \mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j [\mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{a}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j - \mathbf{v}_i^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \mathbf{b}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j].$$

Now we consider a two-variable model example by its design space, one of the five examples studied by Silvey et al. (1978) and Wu (1978) in a standard design context.

The model is $E(y|x) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 = \mathbf{v}^{\top} \boldsymbol{\theta}$, $\mathbf{v} \in \mathcal{V}$, where

$$\mathcal{V} = \{\mathbf{v} = (1, x_1, x_2)^{\top} : (x_1, x_2) \in \mathcal{Q}\},$$

and where $\mathcal{Q}$ is a quadrilateral with vertices $(-1, -2), (-1, 1), (1, -1), (2, 2)$.

Our goal is to obtain the estimators of the parameters θ_2 and θ_3 such that the variances of these estimators are equal. Thus we take $\phi(\mathbf{p}) = -(\mathbf{a}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} - \mathbf{b}^{\top} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b})^2$, where $\mathbf{a} = (0, 1, 0)^{\top}$ and $\mathbf{b} = (0, 0, 1)^{\top}$. We seek to maximize $\phi(\mathbf{p})$. If the maximum value of zero is attained, the parameters θ_2 and θ_3 are estimated with equal variance. We use $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $\delta = 0.5$ and x as the directional derivatives of $\phi(\mathbf{p})$. With equal initial weights

$(\mathbf{p}^{(0)} = (0.25, 0.25, 0.25, 0.25))$ we obtain at the initial iteration the following values.

$$\begin{aligned}\phi(\mathbf{p}) &= -0.04374386, \\ -0.04684159 &\leq F_i^\phi \leq 0.04579246.\end{aligned}$$

After 1000 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -3.081488 \times 10^{-33}, \\ -2.910879 \times 10^{-16} &\leq F_i^\phi \leq 4.115337 \times 10^{-17}.\end{aligned}$$

The value of $\phi(\mathbf{p})$ is very close to zero, and the directional derivatives are also practically zero. This indicates that the algorithm has successfully converged to a solution where the objective function is maximized at a value of zero, meaning the variances of the two estimators are equal. The optimal design is found to be:

$$\mathbf{p}^* = \begin{Bmatrix} (-1, -2) & (-1, 1) & (1, -1) & (2, 2) \\ 0.10557503 & 0.47462534 & 0.05701769 & 0.36278194 \end{Bmatrix}.$$

Since the criterion function lacks concavity, we must check the second-order conditions. We therefore check the Hessian matrix in the second-order condition (2.20) to confirm optimality.

$$H(\mathbf{p}^*) = \begin{Bmatrix} -6.57833945 & -0.0655397066 & -10.8546657 \\ -0.06553971 & -0.0006531007 & -0.1081443 \\ -10.85466566 & -0.1081443376 & -17.9108679 \end{Bmatrix}.$$

All the eigenvalues of the Hessian matrix are negative. The design satisfies both the first- and second-order conditions. We have found a design measure $\mathbf{p}$ for which the criterion function attains its maximum possible value of zero. This indicates that the parameters θ_2 and θ_3 are estimated with equal variance.

Now we consider a second example to demonstrate the flexibility of the proposed methodology. This time, we use a quadratic interaction model, which is a second-order model that includes an interaction term between the two independent variables. This type of model is commonly used in empirical research to capture more complex relationships between factors, a topic extensively discussed in the field of optimal design Box and Wilson (1992).

The model is defined as:

$$E(y|\mathbf{x}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 = \mathbf{v}^\top \boldsymbol{\theta}, \quad \mathbf{v} \in \mathcal{V},$$

where

$$\mathcal{V} = \left\{ \mathbf{v} = (1, x_1, x_2, x_1 x_2)^\top : (x_1, x_2) \in \mathcal{Q} \right\},$$

and where $\mathcal{Q}$ is a quadrilateral with vertices $(-1, -1), (-1, 1), (1, -1), (2, 3)$.

Our goal is to obtain the estimators of the parameters θ_2 and θ_3 with equal variance. We therefore define our criterion function as $\phi(\mathbf{p}) = -(\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{a} - \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{b})^2$, where $\mathbf{a} = (0, 1, 0, 0)^\top$ and $\mathbf{b} = (0, 0, 1, 0)^\top$. We seek to maximize $\phi(\mathbf{p})$ to its maximum possible value of zero. Using the same initial weights as in the previous section, we use $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $\delta = 0.01$ and x as the directional derivatives of $\phi(\mathbf{p})$.

We obtain the following results at the initial iteration:

$$\begin{aligned}\phi(\mathbf{p}) &= -0.02777778, \\ -0.5007868 &\leq F_i^\phi \leq 0.6103243.\end{aligned}$$

After 1500 iterations, the algorithm has converged, yielding the following values:

$$\begin{aligned}\phi(\mathbf{p}) &= -1.414481 \times 10^{-20}, \\ -4.855226 \times 10^{-10} &\leq F_i^\phi \leq 3.236817 \times 10^{-10}.\end{aligned}$$

The objective function value is extremely close to zero, and the vertex directional derivatives are also practically zero, indicating a successful convergence to an optimal solution where the variances are equal. The optimal design is found to be:

$$\mathbf{p}^* = \begin{Bmatrix} (-1, -1) & (-1, 1) & (1, -1) & (2, 3) \\ 0.2474316 & 0.2020547 & 0.3030821 & 0.2474316 \end{Bmatrix}.$$

Here again, we checked that the eigenvalues of the Hessian matrix are negative. Some eigenvalues are small in magnitude, however they are all negative. The design satisfies both the first- and second-order conditions. We can conclude that the parameters θ_2 and θ_3 are estimated with equal variance for this design.

The equality of variances of parameter estimators is not always an explicit goal, but ensuring variances are equal (or nearly equal) enhances fairness, interpretability, and robustness of inference. Designs such as balanced ANOVA and optimal polynomial regression designs support this principle and are widely used

in practice. In engineering experiments, balance ensures that main effects and interactions are estimated with equal precision, which is crucial for identifying key process factors.

In summary, this chapter has solely focused on minimizing variance-based criteria. Variance minimization is a fundamental objective in statistics, as it directly improves the precision of parameter estimation and supports the construction of efficient estimators such as the BLUE.

A key contribution of this chapter is that it moves beyond this global perspective by considering variance-based criteria for subsets of parameters or specific components of interest. In many practical applications, interest is not in all parameters equally, and the proposed work allows for more targeted design construction based on the parameters of primary importance.

In addition, this chapter develops efficiency measures to enable meaningful comparison between competing designs under different criteria. This allows the performance of subset-focused designs to be evaluated relative to classical A - and D -optimal designs, providing insight into the trade-offs between global optimality and targeted precision.

Furthermore, this chapter also introduces criteria aimed at achieving equality of variances among parameter estimators, incorporating a notion of balance or fairness. These criteria are generally not concave, unlike standard optimality criteria, and therefore require more advanced theoretical treatment. The chapter establishes second-order optimality conditions for such criteria, using Hessian-based conditions, to rigorously verify optimality in such cases.

Overall, Chapter 3 advances variance-based optimal design by (i) extending classical criteria to subset-focused objectives, (ii) introducing efficiency-based comparisons between designs, and (iii) developing theoretical tools for handling non-concave optimality criteria.

Chapter 4

Minimizing covariance based criteria

4.1 Introduction

In the previous chapter, we focused on optimal designs based on a number of variance-based criteria. In this chapter, we focus on covariance-based criteria.

The main motivation for the work in this chapter and the next chapter came from a viscosity model provided by a chemist in Scotland and was implemented in response to a practical problem arising in chemistry. In some chemical models, it may be necessary to estimate certain parameters independently of the others. The above problem in chemistry is a study of the relationship between the viscosity and concentration of a chemical solution using a three-parameter regression model. Viscosity is the response variable, and concentration is the design variable. Torsney and Alahmadi (1995) analyzed a three-parameter

regression model relating viscosity to the concentration of a chemical solution. Building on such work, Mandal et al. (2017) developed an approach based on covariances between the parameter estimators. Since covariance-based criteria provide a first step toward estimating parameters independently, we further develop and explore such criteria in this chapter. In the next chapter, we will explore correlation-based criteria.

Another motivating problem in this context is the study of lipoprotein levels in relation to albumin concentration in childhood nephrotic syndrome (Noto et al., 1999). Over the course of five years, they investigated a group of 84 children who were in various clinical phases of the disease. Among the variables that they studied were lipoproteins, plasma albumin, and proteinuria. They carried out a number of multiple regression analyses to ensure that the resulting parameter estimators were not correlated with one another.

In this chapter, we construct approximate optimal designs for minimizing the absolute covariances between least-squares estimators of the parameters in a linear model. Specifically, we construct designs that ensure the covariances between the estimators of two or more components, or linear combinations of components, of θ are zero or as minimum as possible. To achieve this, we minimize the squared covariances between the estimators of the parameters of interest. Equivalently, we maximize the negative of the squared covariances, a technique that aligns with the maximization framework of our algorithms and is more convenient for optimization.

A key difficulty with this criterion is its lack of concavity. Unlike concave functions, which possess a single, easily identifiable global maximum, non-

concave functions may have multiple local maxima. This poses a significant challenge, as an algorithm may converge to a suboptimal solution even if the first-order conditions (where the directional derivatives are close to zero) are satisfied. To overcome this challenge, we adopt a two-step verification process. First, we employ an iterative numerical algorithm to find a candidate solution that we believe to be a local maximum. Second, we verify the solution by checking both the first-order and second-order conditions. The second-order condition, which requires establishing the negative definiteness of the Hessian matrix, is essential for confirming that the candidate point is indeed a local maximum. By adhering to this two-step process, we ensure the validity and reliability of the optimal designs we construct in this chapter.

4.2 Minimizing one covariance

Denote the two linear combinations of interest by $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$. Our objective is to find a design such that the covariance of the corresponding estimators $\mathbf{a}^\top \hat{\boldsymbol{\theta}}$ and $\mathbf{b}^\top \hat{\boldsymbol{\theta}}$ is as close to zero as possible. Formally, we seek to maximize the criterion:

$$\phi(\mathbf{p}) = - \left[\text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) \right]^2 = - \left[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \right]^2$$

over $\mathcal{P} = \{ \mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1 \}$.

If the maximum value of zero is attained, then the covariance between the estimators is zero. In many cases such a design may not exist, but maximizing

$\phi(\mathbf{p})$ nonetheless yields a design with minimal numerical covariance, ensuring that the estimators are as close to being uncorrelated as possible.

For convenience, let us denote the covariance by

$$k(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}.$$

Thus, $\phi(\mathbf{p}) = -k^2(\mathbf{p})$. The structure of $k(\mathbf{p})$ is similar to that of the $\mathbf{c}$ -optimality criterion, which is defined as

$$\phi_c(\mathbf{p}) = \mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{c},$$

where $\mathbf{c}$ is a fixed vector of constants. Minimizing $\phi_c(\mathbf{p})$ produces a $\mathbf{c}$ -optimal design, which is itself a special case of linear optimality. The latter refers to optimization problems where the criterion is linear in the entries of the covariance matrix $\mathbf{M}^{-1}(\mathbf{p})$. In that context, $\mathbf{c}$ -optimality corresponds to constructing a design that minimizes the variance of the estimator of $\mathbf{c}^\top \boldsymbol{\theta}$. The function $\phi_c(\mathbf{p})$ is known to be convex on the cone of positive definite symmetric matrices Fedorov (1972). In contrast, $k(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$ is not convex on the same space. It will be convex if $\mathbf{a} \propto \mathbf{b}$. Nevertheless, it shares an important property with $\mathbf{c}$ -optimality: it is linear in the entries of $\mathbf{M}^{-1}(\mathbf{p})$. This structural similarity makes $k(\mathbf{p})$ amenable to analysis within the same general framework as linear optimality criteria, even though optimization is more challenging due to the lack of convexity.

Since $\phi(\mathbf{p}) = -k^2(\mathbf{p})$, the first-order partial derivatives of $\phi(\mathbf{p})$ with respect

to p_j are given by

$$d_j^\phi = -2k(\mathbf{p})d_j^k,$$

where d_j^k denotes the corresponding derivative of $k(\mathbf{p})$. Similarly, the second-order derivatives are

$$d_{ij}^\phi = -2d_i^k d_j^k - 2k(\mathbf{p})d_{ij}^k.$$

To implement our algorithms, we therefore require expressions for the first- and second-order partial derivatives of $k(\mathbf{p})$ with respect to the design weights. These are given by

$$\begin{aligned} \frac{\partial k(\mathbf{p})}{\partial p_j} &= \mathbf{a}^\top \left(\frac{\partial \mathbf{M}^{-1}(\mathbf{p})}{\partial p_j} \right) \mathbf{b} \\ &= \mathbf{a}^\top \left(-\mathbf{M}^{-1}(\mathbf{p}) \frac{\partial \mathbf{M}(\mathbf{p})}{\partial p_j} \mathbf{M}^{-1}(\mathbf{p}) \right) \mathbf{b} \\ &= -\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \\ &= -\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j, \end{aligned} \quad (4.1)$$

and

$$\begin{aligned} \frac{\partial^2 k(\mathbf{p})}{\partial p_i \partial p_j} &= \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i \\ &\quad + \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i. \end{aligned} \quad (4.2)$$

In practice, both partial and directional derivatives can take very large numerical values in early iterations of the algorithm. This often slows down convergence and may even cause the criterion values to become unbounded. To address this,

we replace the raw partial derivatives d_j by standardized versions

$$\tilde{d}_j = \frac{d_j}{\sum_{i=1}^J p_i d_i},$$

where we note that $\sum_{i=1}^J p_i d_i = -2\phi(\mathbf{p})$, which is positive whenever $\phi(\mathbf{p})$ is nonzero. This normalization substantially improves the stability of the iterative procedure.

In the following subsections, we present several examples. The first example revisits the quadratic regression model $E(y|x) = \theta_1 + \theta_2 x + \theta_3 x^2$, considered in the previous chapter, but with design interval $[-1, 2]$. Here we aim to estimate the parameters θ_2 and θ_3 such that the covariance between the corresponding estimators is zero. To do so, we find the optimal design $\mathbf{p}$ by maximizing $\phi(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2$, where $\mathbf{a} = (0, 1, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$.

We start by taking $f(x, \delta) = \Phi(\delta x)$ and choose the x_j to be the partial derivatives d_j of $\phi(\cdot)$ with respect to the p_j . We take the initial design to be $p_j^{(0)} = 1/3$, $j = 1, 2, 3$, set $\delta = 0.001$ and apply the algorithm. After 864 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -2.406038 \times 10^{-17}, \\ -3.332671 \times 10^{-7} &\leq F_i^\phi \leq 9.871143 \times 10^{-9}. \end{aligned}$$

We now use the same settings as above but take $\delta = 0.005$. In this case, after

200 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -1.32381 \times 10^{-26}, \\ -7.817247 \times 10^{-12} &\leq F_i^\phi \leq 2.315139 \times 10^{-13}.\end{aligned}$$

Taking $f(x, \delta) = \exp(\delta x)/(1+\exp(\delta x))$, $x_j = F_j$ and $\delta = 0.001$, and starting with equal initial weights, after 1384 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -3.270146 \times 10^{-31}, \\ -3.885303 \times 10^{-14} &\leq F_i^\phi \leq 1.150813 \times 10^{-15}.\end{aligned}$$

We now use the same choices as in the third experiment but increase δ to 0.005.

We then obtain, after 300 iterations

$$\begin{aligned}\phi(\mathbf{p}) &= -2.899589 \times 10^{-22}, \\ -1.156936 \times 10^{-9} &\leq F_i^\phi \leq 3.426542 \times 10^{-11}.\end{aligned}$$

We observed that the choice of the function $f(\cdot, \cdot)$ and the parameter δ had a significant impact on the algorithm's convergence rate. While some combinations led to a faster approach to the optimal solution, others required more iterations. Despite these differences in convergence speed, a key finding was that the algorithm consistently converged to the same optimal design across all four experiments. The final design, represented by the weights $\mathbf{p}^* = (0.4729487, 0.4984561, 0.0285952)$, confirmed the theoretical expectation for the given support points $(-1, 0, 2)$. The most crucial outcome of this analysis is that the value of the criterion function $\phi(\mathbf{p})$ at this optimal design reached its global maximum of zero, which was verified with high precision (at least 11 dec-

imal places). Since this value represents the covariance between the estimators for parameters θ_2 and θ_3 , achieving a value of zero demonstrates that we have found a design that renders the estimators uncorrelated. The attainment of this global maximum obviates the need to examine the eigenvalues of the Hessian matrix, as the solution is confirmed to be optimal.

We now consider an example in which zero covariance cannot be attained but can still construct a design such that the parameter estimators of interest are as uncorrelated as possible. It involves a practical problem arising in chemistry Torsney and Alahmadi (1995). As we mentioned earlier, this regression model describes the relationship between the viscosity y and the concentration x of a chemical solution. Viscosity is the response. The model is given by

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2.$$

The design interval is restricted to a short interval $(0, 0.2]$. It was desired to minimize the covariance between $\hat{\theta}_3$ and each of $\hat{\theta}_1$ and $\hat{\theta}_2$. That is, we seek to minimize each of the two covariances. Since we cannot obtain a design that produces a common value of zero covariance.

In the first covariance case we take $\mathbf{a} = (1, 0, 0)^\top$, $\mathbf{b} = (0, 0, 1)^\top$. We discretize the design space at intervals of 0.01 between 0.02 and 0.20. We start with equal initial weights. Note that the values of both the partial and directional derivatives are very large in the initial stages of the algorithm so we employ the technique of standardizing the partial derivatives. We run algorithm (2.23) with $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $x_j = \tilde{d}_j$ and $\delta = 1.5$. After 2500 iterations,

we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -1487305833, \\ -1.025987 &\leq F_i^\phi \leq 2.056355 \times 10^{-12}.\end{aligned}$$

We found that the algorithm converged to a solution with three active support points: 0.02, 0.12, and 0.20. The corresponding optimal weights for these points were determined to be (0.4233560, 0.4049047, 0.1717393), which places the optimal design on a boundary of the design space $\mathcal{P}$. The optimality of this design was confirmed by checking both first- and second-order conditions. For the first-order conditions, the directional derivatives corresponding to the three active support points were zero, while they were negative for all other design points with zero weight. This indicates that the algorithm successfully located a stationary point. The value of the minimized covariance at this solution was found to be -38565.6. To verify that this stationary point is indeed a local maximum, we analyzed the Hessian matrix, $H(\mathbf{p}^*)$, which is of order 2×2 since the vector of positive weights, $\mathbf{p}^*$, has a length of 3. The determinant of this Hessian was calculated as $\det(H(\mathbf{p}^*)) = 1.202243 \times 10^{21}$. The specific elements are $d_{11}^{\tilde{\phi}}(\mathbf{p}^*) = -48693539565$, $d_{12}^{\tilde{\phi}}(\mathbf{p}^*) = d_{21}^{\tilde{\phi}}(\mathbf{p}^*) = -34641007532$, and $d_{22}^{\tilde{\phi}}(\mathbf{p}^*) = -49333905134$. Since the determinant is positive, the second-order conditions are satisfied, which confirms that the solution we found is a local maximum for the criterion function.

In the second covariance case we take $\mathbf{a} = (0, 1, 0)^\top$, $\mathbf{b} = (0, 0, 1)^\top$ and again start with equal initial weights. Again the values of both the partial and directional derivatives are very large in the initial stages of the algorithm so we

employ the same technique of standardizing the partial derivatives. The logistic c.d.f. seems to be a good choice for $f(\cdot, \cdot)$ so we use this function with $x_j = \tilde{d}_j$ and $\delta = 1.5$. After 1400 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -47739048, \\ -1.184627 &\leq F_i^\phi \leq 5.002221 \times 10^{-12}.\end{aligned}$$

We observed that the algorithm converged to a second solution with three support points: 0.02, 0.12, 0.20. The corresponding weights for this solution are (0.5089060, 0.3468095, 0.1442846). Similar to the previous case, this design also resides on a boundary of the design space $\mathcal{P}$. The optimality conditions were checked to confirm the nature of this solution. The first-order conditions were met, as the directional derivatives were zero at the three active support points and negative for all other zero-weighted points. The value of the covariance at this solution was 6909.344. To verify that this stationary point is indeed a local maximum, we analyzed the Hessian matrix, $H(\mathbf{p}^*)$. Here, $\mathbf{p}^*$ denotes the vector of positive components of the optimal design. Note that $H(\mathbf{p}^*)$ is of order 2×2 because in the Hessian we consider $J - 1$ components instead of J components. The determinant of this Hessian was calculated as $\det(H(\mathbf{p}^*)) = 1.43192 \times 10^{18}$. The specific elements are $d_{11}^{\tilde{\phi}}(\mathbf{p}^*) = -1698697895$, $d_{12}^{\tilde{\phi}}(\mathbf{p}^*) = d_{21}^{\tilde{\phi}}(\mathbf{p}^*) = -1323469058$, and $d_{22}^{\tilde{\phi}}(\mathbf{p}^*) = -1874077442$. The positive determinant, $\det(H(\mathbf{p}^*)) > 0$, confirms that the second-order conditions are satisfied. Thus, we have found a point where the criterion function attains a local maximum.

As demonstrated by the preceding two examples, the proposed algorithm successfully converges to solutions that satisfy both the first and second-order

optimality conditions, confirming that these points are local maxima. In both cases, the optimal designs were found to have three support points located on the boundary of the design space. This highlights the algorithm's capability to accurately identify optimal solutions even under these constraints. Based on these successful findings, we now extend our investigation from a single covariance or two covariances to the more complex case of minimizing covariances among more than two linear combinations of the model parameters. The following section will detail the formulation of this new optimization problem and present our findings.

4.3 Minimizing two or more covariances

This time we consider the case of more than two parameters or linear combinations of the parameters. That is, we minimize two or more covariances at the same time. Let us consider four linear combinations $\mathbf{a}^\top \boldsymbol{\theta}, \mathbf{b}^\top \boldsymbol{\theta}, \mathbf{c}^\top \boldsymbol{\theta}$ and $\mathbf{d}^\top \boldsymbol{\theta}$, where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^k$. We want to minimize the magnitude of the covariance between the estimators of $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$ and between the estimators of $\mathbf{c}^\top \boldsymbol{\theta}$ and $\mathbf{d}^\top \boldsymbol{\theta}$. The two covariances are given by

$$k_1(\mathbf{p}) = \text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) \text{ and } k_2(\mathbf{p}) = \text{cov}(\mathbf{c}^\top \hat{\boldsymbol{\theta}}, \mathbf{d}^\top \hat{\boldsymbol{\theta}}).$$

Let $\phi_i(\mathbf{p}) = -k_i(\mathbf{p})^2$, $i = 1, 2$. We now create a compound criterion involving both $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. We choose an appropriate compound criterion in such a way that if there exists a design $\mathbf{p}^*$ which simultaneously maximizes the two functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$ then this $\mathbf{p}^*$ should maximize the compound criterion.

Ideally we wish to find the $\mathbf{p}^*$ to produce a common value of zero for $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. The optimum solution of the compound criterion must satisfy the constraints $p_j \geq 0$ and $\sum_{j=1}^J p_j = 1$. We consider the following two compound criteria

$$\begin{aligned}\phi(\mathbf{p}) &= \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}), \\ \phi(\mathbf{p}) &= \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}.\end{aligned}$$

When we maximize any of these two compound criteria we consider two sets of partial derivatives d_j corresponding respectively to the two functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. We use the multiplicative algorithm. The choice of x in $f(x, \delta)$ depends on which of the compound criteria is used. According to the above compound criteria, we consider the following choices of x :

$$\begin{aligned}x_j &= d_j^{\phi_1} + d_j^{\phi_2} \text{ if we maximize } \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}) \\ x_j &= d_j^{\min\{\phi_1, \phi_2\}} \text{ if we maximize } \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}\end{aligned}$$

where $d_j^{\phi_i}$ denotes the partial derivative of ϕ_i with respect to p_j , $i = 1, 2$, $j = 1, 2, \dots, J$. Here again, we improve the rate of convergence of the algorithm by replacing $d_j^{\phi_i}$ by the corresponding directional derivatives.

The following example, discussed in Section 3.3, is one of five studied by Silvey et al. (1978) and Wu (1978) in the context of constructing D -optimal designs on a finite design space. The model is $E(y) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 = \mathbf{v}^\top \boldsymbol{\theta}$,

$\mathbf{v} \in \mathcal{V}$, where

$$\mathcal{V} = \{\mathbf{v} = (1, x_1, x_2)^\top : (x_1, x_2) \in \mathcal{Q}\},$$

and where $\mathcal{Q}$ is a quadrilateral with vertices $(-1, -1), (-1, 1), (1, -1), (2, 3)$. We take the design points to consist of the vertices of the quadrilateral, and wish to find the optimal weights for these four vertices according to the following covariance criteria. We wish to obtain an estimator of the parameter θ_1 which is uncorrelated with both of the estimators of the other two parameters θ_2 and θ_3 . Thus we take $\phi_1(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2$, and $\phi_2(\mathbf{p}) = -[\mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{d}]^2$, where $\mathbf{a} = (1, 0, 0)^\top$ and $\mathbf{b} = (0, 1, 0)^\top$, $\mathbf{c} = \mathbf{a}$ and $\mathbf{d} = (0, 0, 1)^\top$. We seek to maximize the compound criterion $\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p})$. We use $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $\delta = 0.05$ and $x_j = d_j^{\phi_1} + d_j^{\phi_2}$. With equal initial weights $(\mathbf{p}^{(0)} = (0.25, 0.25, 0.25, 0.25))$ we obtain at the initial iteration the following values

$$\begin{aligned} (\phi_1, \phi_2) &= (-0.001371742, -0.027777778), \\ -0.1031528 &\leq F_i^{\phi_1} \leq 0.08198155, \\ -0.2706929 &\leq F_i^{\phi_2} \leq 0.3507063. \end{aligned}$$

After 900 iterations, we obtain

$$\begin{aligned} (\phi_1, \phi_2) &= (-2.244726 \times 10^{-17}, -6.035377 \times 10^{-17}), \\ -1.040999 \times 10^{-8} &\leq F_i^{\phi_1} \leq 1.040999 \times 10^{-8}, \\ -1.768824 \times 10^{-8} &\leq F_i^{\phi_2} \leq 1.280213 \times 10^{-8}. \end{aligned}$$

We also consider maximizing the minimum of $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$, i.e. using compound criterion $\phi(\mathbf{p}) = \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}$ rather than maximizing their sum. We use the same $f(\cdot, \cdot)$ and $\delta = 0.05$, and the same starting values but replace the partial derivatives by the corresponding directional derivatives. After 1726 iterations we obtain

$$\begin{aligned}(\phi_1, \phi_2) &= (-3.465597 \times 10^{-17}, -3.550244 \times 10^{-17}), \\ -1.294161 \times 10^{-8} &\leq F_i^{\phi_1} \leq 1.294161 \times 10^{-8}, \\ -1.356006 \times 10^{-8} &\leq F_i^{\phi_2} \leq 9.824029 \times 10^{-9}.\end{aligned}$$

The algorithm converged to the following optimal design:

$$\mathbf{p}^* = \begin{Bmatrix} (-1, -1) & (-1, 1) & (1, -1) & (2, 3) \\ 0.350 & 0.220 & 0.290 & 0.140 \end{Bmatrix}.$$

The criterion functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$ are constrained to be non-positive, with a maximum of zero. The calculated values for the criteria, $-3.465597 \times 10^{-17}$ and $-3.550244 \times 10^{-17}$ respectively, are effectively zero within the limits of numerical precision. The attainment of this theoretical global maximum, where the criterion functions are at their highest possible value, provides a strong confirmation of the solution's optimality. This result implies that the estimators for the parameters θ_2 and θ_3 are uncorrelated with the estimator for θ_1 at this optimal design. Since the global maximum has been reached, the analysis of the Hessian matrix's eigenvalues is not required to confirm local optimality. However, we checked that the eigenvalues of the Hessian matrix are all negative.

Here we present another example, following Section 3.3. The model is

$$E(Y|\mathbf{x}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 = \mathbf{v}^\top \boldsymbol{\theta}, \quad \mathbf{v} \in \mathcal{V},$$

where

$$\mathcal{V} = \{ \mathbf{v} = (1, x_1, x_2, x_1 x_2)^\top : (x_1, x_2) \in \mathcal{Q} \},$$

and $\mathcal{Q}$ is a quadrilateral with vertices $(-1, -1)$, $(-1, 1)$, $(1, -1)$, and $(2, 3)$.

As in the previous example, our objective is to obtain an estimator of the parameter θ_1 that is uncorrelated with the estimators of the other two parameters, θ_2 and θ_3 . To achieve this, we define the criteria

$$\phi_1(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2, \quad \phi_2(\mathbf{p}) = -[\mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{d}]^2,$$

where $\mathbf{a} = (1, 0, 0, 0)^\top$, $\mathbf{b} = (0, 1, 0, 0)^\top$, $\mathbf{c} = \mathbf{a}$, and $\mathbf{d} = (0, 0, 1, 0)^\top$. We then consider two compound criteria: (i) maximizing $\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p})$, and (ii) maximizing $\phi(\mathbf{p}) = \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}$.

We first consider maximizing the sum of the two criteria. Using the multiplicative function (the logistic cdf)

$$f(x, \delta) = \frac{\exp(\delta x)}{1 + \exp(\delta x)}, \quad \delta = 0.05,$$

and setting $x_j = d_j^{\phi_1} + d_j^{\phi_2}$, we start with equal initial weights $\mathbf{p}^{(0)} =$

$(0.25, 0.25, 0.25, 0.25)$. At the initial iteration, the criterion values are

$$\begin{aligned}(\phi_1, \phi_2) &= (-0.0156250, -0.0017361), \\ -0.2766357 &\leq F_i^{\phi_1} \leq 0.5289198, \\ -0.0843902 &\leq F_i^{\phi_2} \leq 0.1447765.\end{aligned}$$

After 600 iterations, the algorithm converges to

$$\begin{aligned}(\phi_1, \phi_2) &= (0, -4.520293 \times 10^{-32}), \\ 0 &\leq F_i^{\phi_1} \leq 0, \\ -6.245819 \times 10^{-16} &\leq F_i^{\phi_2} \leq 7.208828 \times 10^{-16}.\end{aligned}$$

Next, we consider maximizing the minimum of $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. We use the same function $f(\cdot, \cdot)$ with $\delta = 0.12$ and the same starting weights, but replace the partial derivatives with the corresponding directional derivatives. After 1100 iterations, we obtain

$$\begin{aligned}(\phi_1, \phi_2) &= (-1.42900 \times 10^{-31}, -2.73519 \times 10^{-32}), \\ -1.11046 \times 10^{-15} &\leq F_i^{\phi_1} \leq 1.28165 \times 10^{-15}, \\ -4.85826 \times 10^{-16} &\leq F_i^{\phi_2} \leq 5.60721 \times 10^{-16}.\end{aligned}$$

In both cases, the optimal design converges to

$$\mathbf{p}^* = \left\{ \begin{array}{cccc} (-1, -1) & (-1, 1) & (1, -1) & (2, 3) \\ 0.2033924 & 0.2560077 & 0.2880087 & 0.2525912 \end{array} \right\}.$$

The above example demonstrates that both compound criteria lead to the same optimal design. The optimization algorithm reaches the theoretical maximum for both ϕ_1 and ϕ_2 with high numerical precision. We also checked the negative definiteness of the Hessian in both settings. Despite the different formulations (sum vs. maximin), both approaches converge to the same optimal design, highlighting the robustness of the method and the consistency of the resulting optimal design.

We now consider the quadratic full model given by

$$E(y \mid x_1, x_2) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 + \theta_6 x_2^2.$$

We consider a covariance criterion example using a slightly asymmetric design region. The design points are taken on a 3×3 grid with

$$x_1 \in \{-1, 0, 2\}, \quad x_2 \in \{-1, 0, 2\}.$$

The regression vector is defined as

$$\mathbf{v}(x_1, x_2) = (1, x_1, x_1 x_2, x_2, x_1^2, x_2^2)^\top,$$

so that the information matrix is

$$\mathbf{M}(\mathbf{p}) = \sum_{j=1}^J p_j \mathbf{v}(x_{1j}, x_{2j}) \mathbf{v}(x_{1j}, x_{2j})^\top,$$

where $J = 9$ is the number of design points. We take the initial design to be

uniform,

$$p_j^{(0)} = \frac{1}{9}, \quad j = 1, \dots, 9.$$

The vectors used to define the covariance criteria are

$$\mathbf{a} = (1, 0, 0, 0, 0, 0)^\top, \quad \mathbf{b} = (0, 0, 1, 0, 0, 0)^\top, \quad \mathbf{c} = \mathbf{a}, \quad \mathbf{d} = (0, 0, 0, 0, 0, 1)^\top.$$

The corresponding compound criterion is

$$\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}),$$

$$\phi_1(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2, \quad \phi_2(\mathbf{p}) = -[\mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{d}]^2.$$

Here, ϕ_1 corresponds to minimizing the squared covariance between the intercept and interaction term $x_1 x_2$, while ϕ_2 corresponds to between the intercept and the quadratic term x_2^2 . The design is updated iteratively using the function

$$f(x, \delta) = \frac{\exp(\delta x)}{1 + \exp(\delta x)},$$

with $\delta = 0.3$, where x represents the directional derivatives of $\phi(\mathbf{p})$. Starting from the uniform initial design, we have

$$\begin{aligned} (\phi_1, \phi_2) &= (-0.0021085, -2.250), \\ -.098117 &\leq F_i^{\phi_1} \leq 0.08687, \\ -13.42201 &\leq F_i^{\phi_2} \leq 15.16218. \end{aligned}$$

After 8,533 iterations, the algorithm converges to

$$\begin{aligned}(\phi_1, \phi_2) &= (-4.0679 \times 10^{-11}, -1.5279 \times 10^{-9}), \\ -2.08 \times 10^{-5} &\leq F_i^{\phi_1} \leq 1.46 \times 10^{-5}, \\ -9.73 \times 10^{-4} &\leq F_i^{\phi_2} \leq 9.74 \times 10^{-9}.\end{aligned}$$

The final optimal design is summarized below:

Index	Design Point (x_1, x_2)	Final p_j
1	$(-1, -1)$	0.10753
2	$(0, -1)$	0.00000
3	$(2, -1)$	0.11981
4	$(-1, 0)$	0.16896
5	$(0, 0)$	0.23881
6	$(2, 0)$	0.16678
7	$(-1, 2)$	0.10418
8	$(0, 2)$	0.00000
9	$(2, 2)$	0.09393

We see that both ϕ_1 and ϕ_2 converge to values numerically indistinguishable from zero. The compound optimization method successfully converges to the optimal design.

In summary, this chapter addresses the covariance-based criteria. While variance-based criteria, as we did in the previous chapter, improve the precision of the parameter estimators, they do not directly address the dependence structure

among them. In particular, classical criteria such as A - and D -optimality do not control the covariance between estimators.

This chapter addresses this limitation by introducing covariance-based optimality criteria. The key idea is to construct designs that minimize one or more covariances among parameter estimators, thereby reducing dependence and, in some cases, achieving approximate or complete uncorrelatedness between estimators. This is important for interpretability and stability of inference, especially when parameters are jointly estimated.

Building on earlier work (e.g., Mandal et al. (2017)), this chapter strengthens and generalizes covariance-based design methodology by providing a more systematic formulation, extending the criteria to multiple covariances, and developing corresponding optimality conditions and computational approaches.

Furthermore, the study of covariance naturally motivates the investigation of correlation structures, since correlation standardizes covariance and provides a scale-invariant measure of dependence. This progression forms the foundation for the developments in Chapter 5, where correlation-based criteria are studied in greater depth.

Chapter 5

Correlation based criteria: Estimating parameters with minimal correlation

5.1 Introduction

In the previous chapter, we considered covariance-based criteria. We considered criteria based on the covariance between two parameter estimators. However, it is more practical to consider correlations between parameter estimators. A key distinction exists between covariance and correlation that is critical to our work. A covariance of zero indicates the correlation is zero, and vice versa. However, if the measures are not zero, they do not provide the same interpretation. Covariance indicates the direction of the relationship between the parameter estimators

but not its strength. This can be problematic, as a large magnitude of covariance can make the partial or directional derivatives of the criterion very large, making it difficult to run optimization algorithms. In contrast, correlation indicates both the direction and the strength of the relationship, with a value guaranteed to fall between -1 and 1. This makes the correlation criterion advantageous for algorithmic stability. Furthermore, when the covariance or correlation between estimators is known to be non-zero, our next goal is to construct a design that makes the parameter estimators as uncorrelated as possible. Motivated by this, in this chapter, we consider a correlation-based criterion. The correlation criterion is much more challenging due to the complexity of the expressions for the criterion itself and its partial and directional derivatives. We have been able to cope with these challenges and constructed the optimal designs for a number of models including the model in chemistry.

More specifically, we will opt for a non-standard criterion, which is the negative of the squared correlation between two parameter estimators. We construct the optimal designs by maximizing the negative of the squared correlation. The advantage of taking this criterion is that the value of the criterion is always less than or equal to zero, and the maximum value of the criterion is zero. If the maximum value of zero is attained, the estimators of the two parameters or the linear combinations of the parameters will be uncorrelated. In that case, assuming a Gaussian model, the corresponding parameters or parametric functions are estimated independently. If any such design does not exist, maximizing the above criterion will produce a design with minimum correlation, meaning the estimators are as uncorrelated with each other as possible. In other words, the

estimators are as uncorrelated with each other as possible.

In addition to maximizing the above criterion for two parameter estimators, we consider the problem of obtaining more than two parameter estimators with more than one correlation among them. To accomplish this, we employ a compound optimization criterion and use simultaneous optimization techniques. In particular, we convert it into an optimization problem where we simultaneously maximize two criteria, i.e., two functions of the design weights. The criteria share a maximum of zero, which is simultaneously obtained at the optimal design weights. This optimization technique can be even extended to more than two correlations of interest.

5.2 Minimizing one correlation

We begin by considering the problem of minimizing the correlation between two specific linear combinations of the parameter estimators. Let the two linear combinations of interest be $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$. We wish to find a design such that the correlation of the estimators of $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$ is zero. When the zero correlation is not possible, we will find a design such that the correlation of the estimators is as close to zero as possible.

As established in Chapters 3 and 4, the variance-covariance matrix of the least squares parameter estimator $\hat{\boldsymbol{\theta}}$ is proportional to the inverse of the information matrix, $\mathbf{M}^{-1}(\mathbf{p})$. Building on this foundation, the covariance between the two

linear combinations is given by:

$$\text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}.$$

The variances of the individual linear combinations are similarly defined as:

$$\text{var}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \quad \text{and} \quad \phi_b(\mathbf{p}) = \text{var}(\mathbf{b}^\top \hat{\boldsymbol{\theta}}) = \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}.$$

Using these expressions, the correlation between the two linear combinations is defined as:

$$l(\mathbf{p}) = \frac{\text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}})}{\sqrt{\text{Var}(\mathbf{a}^\top \hat{\boldsymbol{\theta}})} \sqrt{\text{Var}(\mathbf{b}^\top \hat{\boldsymbol{\theta}})}} = \frac{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}{\sqrt{(\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a})(\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b})}}.$$

Our optimization problem is to find a design $\mathbf{p}$ that maximizes the negative of the squared correlation, $-l^2(\mathbf{p})$. We define our objective function as:

$$\phi(\mathbf{p}) = -l^2(\mathbf{p}) = -\frac{[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2}{[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}][\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]} \quad (5.1)$$

The goal is to find the optimal design weights p_j that maximize $\phi(\mathbf{p})$, subject to the standard constraints $\{\mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1\}$ (2.16). Since $\phi(\mathbf{p}) = -l^2(\mathbf{p})$, the partial derivatives of $\phi(\mathbf{p})$ with respect to the p_j are $d_j^\phi = -2l(\mathbf{p})d_j^l$, where the d_j^l are the corresponding partial derivatives of $l(\mathbf{p})$. As we mentioned in the introduction, this criterion function is not concave. We need to check for second-order conditions using the Hessian. In order to check second-order conditions we need to find the second order partial derivatives

of $l(\mathbf{p})$ with respect to p_i and p_j , this is $d_{ij}^\phi = -2d_i^l d_j^l - 2l(\mathbf{p})d_{ij}^l$. In order to run the algorithms so as to find the optimal design and to check conditions for optimality we need to find the first and second order partial derivatives d_j^l and d_{ij}^l respectively. Derivatives of $l(\mathbf{p})$ are provided by the following theorem. Let $k(\mathbf{p}) = \text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$, $\phi_a(\mathbf{p}) = \text{var}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}$ and $\phi_b(\mathbf{p}) = \text{var}(\mathbf{b}^\top \hat{\boldsymbol{\theta}}) = \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$, then $\phi(\mathbf{p})$ is $-\left(\frac{k(\mathbf{p})}{\sqrt{\phi_a(\mathbf{p})}\sqrt{\phi_b(\mathbf{p})}}\right)^2$.

Theorem 5.2. For $l(\mathbf{p}) = \frac{\text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}})}{\sqrt{\text{Var}(\mathbf{a}^\top \hat{\boldsymbol{\theta}})}\sqrt{\text{Var}(\mathbf{b}^\top \hat{\boldsymbol{\theta}})}}$, $d_j^l, d_{ij}^l, i, j = 1, 2, \dots, J$, are given by

$$\begin{aligned} d_j^l &= \frac{\frac{\partial k}{\partial p_j} \phi_a^{1/2} \phi_b^{1/2} - k \left[\frac{1}{2} \phi_a^{-1/2} \phi_b^{1/2} \frac{\partial \phi_a}{\partial p_j} + \frac{1}{2} \phi_a^{-1/2} \phi_b^{1/2} \frac{\partial \phi_b}{\partial p_j} \right]}{\phi_a \phi_b} \\ &= \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{1}{2}} d_j^k - \frac{1}{2} k \phi_a^{-\frac{3}{2}} \phi_b^{-\frac{1}{2}} d_j^{\phi_a} - \frac{1}{2} k \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{3}{2}} d_j^{\phi_b} \\ &= \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{1}{2}} \left[d_j^k - \frac{1}{2} k \phi_a^{-1} d_j^{\phi_a} - \frac{1}{2} k \phi_b^{-1} d_j^{\phi_b} \right], \end{aligned}$$

$$\begin{aligned} d_{ij}^l &= -\frac{1}{2} \phi_a^{-\frac{3}{2}} d_j^{\phi_a} \phi_b^{-\frac{1}{2}} \left[d_i^k - \frac{1}{2} k \phi_a^{-1} d_i^{\phi_a} - \frac{1}{2} k \phi_b^{-1} d_i^{\phi_b} \right] \\ &\quad - \frac{1}{2} \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{3}{2}} d_j^{\phi_b} \left[d_i^k - \frac{1}{2} k \phi_a^{-1} d_i^{\phi_a} - \frac{1}{2} k \phi_b^{-1} d_i^{\phi_b} \right] \\ &\quad + \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{1}{2}} \left[d_{ij}^k - \frac{1}{2} \phi_a^{-1} (d_j^k d_i^{\phi_a} - k \phi_a^{-1} d_i^{\phi_a} d_j^{\phi_a} + k d_{ij}^{\phi_a}) \right. \\ &\quad \left. - \frac{1}{2} \phi_b^{-1} (d_j^k d_i^{\phi_b} - k \phi_b^{-1} d_i^{\phi_b} d_j^{\phi_b} + k d_{ij}^{\phi_b}) \right] \\ &= -\frac{1}{2} \phi_a^{-\frac{1}{2}} \phi_b^{-\frac{1}{2}} \left[-2d_{ij}^k + \phi_a^{-1} \left(d_i^k d_j^{\phi_a} + d_j^k d_i^{\phi_a} - \frac{3}{2} k \phi_a^{-1} d_i^{\phi_a} d_j^{\phi_a} - \frac{1}{2} k \phi_a^{-1} d_i^{\phi_b} d_j^{\phi_a} + k d_{ij}^{\phi_a} \right) \right. \\ &\quad \left. + \phi_b^{-1} \left(d_i^k d_j^{\phi_b} + d_j^k d_i^{\phi_b} - \frac{1}{2} k \phi_b^{-1} d_i^{\phi_a} d_j^{\phi_b} - \frac{3}{2} k \phi_b^{-1} d_i^{\phi_b} d_j^{\phi_b} + k d_{ij}^{\phi_b} \right) \right], \end{aligned}$$

where

$$\begin{aligned}
d_j^k &= -\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j, \\
d_{ij}^k &= \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \\
&\quad + \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j, \\
d_j^{\phi_a} &= -(\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j)^2, \\
d_{ij}^{\phi_a} &= 2\mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j, \\
d_j^{\phi_b} &= -(\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j)^2, \\
d_{ij}^{\phi_b} &= 2\mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j.
\end{aligned}$$

Proof:

We prove the theorem by making use of the Gâteaux derivative of (2.18) that we defined earlier. We have $l(\mathbf{p}) = k(\mathbf{p}) / \sqrt{\phi_a(\mathbf{p})} \sqrt{\phi_b(\mathbf{p})}$. We first consider the function $k(\mathbf{p})$. Since $k(\mathbf{p})$ is a function of $\mathbf{p}$ as well as of the information matrix $\mathbf{M}(\mathbf{p})$, let $k(\mathbf{p}) = K(\mathbf{M}(\mathbf{p}))$. From the definition of the Gâteaux derivative (2.18), we can write

$$G_K\{\mathbf{M}, \mathbf{N}\} = \lim_{\varepsilon \downarrow 0} \frac{K(\mathbf{M} + \varepsilon \mathbf{N}) - K(\mathbf{M})}{\varepsilon}. \quad (5.2)$$

$G_K\{\mathbf{M}, \mathbf{N}\}$ is called the Gâteaux derivative of K at $\mathbf{M}$ in the direction of $\mathbf{N}$.

Now,

$$\begin{aligned}
(M + \varepsilon N)^{-1} &= (I + \varepsilon M^{-1}N)^{-1}M^{-1} \\
&= M^{-1/2}(I + \varepsilon M^{-1/2}NM^{-1/2})^{-1}M^{-1/2} \\
&= M^{-1/2}(I + \varepsilon B)^{-1}M^{-1/2}
\end{aligned} \tag{5.3}$$

where $B = M^{-1}N$ or $M^{-1/2}NM^{-1/2}$, and

$$\begin{aligned}
(I + \varepsilon B)^{-1} &= I - \varepsilon B + \varepsilon^2 B^2 - \varepsilon^3 B^3 + \varepsilon^4 B^4 - \dots \\
&= I - \varepsilon B \quad \text{for } \varepsilon \text{ small.}
\end{aligned} \tag{5.4}$$

Now, using (5.3) and (5.4), $K(M + \varepsilon N)$ is given by

$$\begin{aligned}
K(M + \varepsilon N) &= \mathbf{a}^\top (M + \varepsilon N)^{-1} \mathbf{b} \\
&= \mathbf{a}^\top M^{-1/2} (I - \varepsilon B) M^{-1/2} \mathbf{b} \\
&= \mathbf{a}^\top M^{-1/2} M^{-1/2} \mathbf{b} - \varepsilon \mathbf{a}^\top M^{-1/2} B M^{-1/2} \mathbf{b} \\
&= \mathbf{a}^\top M^{-1} \mathbf{b} - \varepsilon \mathbf{a}^\top M^{-1} N M^{-1} \mathbf{b}.
\end{aligned} \tag{5.5}$$

From (5.2) and (5.5), with $M = M(\mathbf{p})$, we obtain

$$\begin{aligned}
d_j^k = \frac{\partial k}{\partial p_j} &= G_K\{M(\mathbf{p}), \mathbf{v}_j \mathbf{v}_j^\top\} \\
&= -\mathbf{v}_j^\top M^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top M^{-1}(\mathbf{p}) \mathbf{v}_j.
\end{aligned}$$

Now, to prove d_{ij}^k , let $K'_j(M(\mathbf{p})) = -\mathbf{v}_j^\top M^{-1} \mathbf{a} \mathbf{b}^\top M^{-1} \mathbf{v}_j$. Now,

$$\begin{aligned}
K'_j(M + \varepsilon N) &= -\mathbf{v}_j^\top (M + \varepsilon N)^{-1} \mathbf{a} \mathbf{b}^\top (M + \varepsilon N)^{-1} \mathbf{v}_j \\
&= -\mathbf{v}_j^\top M^{-1/2} (I - \varepsilon B) M^{-1/2} \mathbf{a} \mathbf{b}^\top M^{-1/2} (I - \varepsilon B) M^{-1/2} \mathbf{v}_j \\
&= -\mathbf{v}_j^\top [M^{-1} \mathbf{a} - \varepsilon M^{-1} N M^{-1} \mathbf{a}] [\mathbf{b}^\top M^{-1} - \varepsilon \mathbf{b}^\top M^{-1} N M^{-1}] \mathbf{v}_j \\
&= -\mathbf{v}_j^\top M^{-1} \mathbf{a} \mathbf{b}^\top M^{-1} \mathbf{v}_j \\
&\quad + \varepsilon [\mathbf{v}_j^\top M^{-1} N M^{-1} \mathbf{a} \mathbf{b}^\top M^{-1} \mathbf{v}_j + \mathbf{v}_j^\top M^{-1} \mathbf{a} \mathbf{b}^\top M^{-1} N M^{-1} \mathbf{v}_j] \\
&\quad - \varepsilon^2 [\mathbf{v}_j^\top M^{-1} N M^{-1} \mathbf{a} \mathbf{b}^\top M^{-1} N M^{-1} \mathbf{v}_j]. \tag{5.6}
\end{aligned}$$

Thus, from (5.2) and (5.6), with $M = M(\mathbf{p})$,

$$\begin{aligned}
d_{ij}^k &= \frac{\partial^2 k}{\partial p_i \partial p_j} \\
&= G_{K'_j} \{M(\mathbf{p}), \mathbf{v}_i \mathbf{v}_i^\top\} \\
&= [\mathbf{v}_j^\top M^{-1}(\mathbf{p}) \mathbf{v}_i \mathbf{v}_i^\top M^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top M^{-1}(\mathbf{p}) \mathbf{v}_j + \mathbf{v}_j^\top M^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top M^{-1}(\mathbf{p}) \mathbf{v}_i \mathbf{v}_i^\top M^{-1}(\mathbf{p}) \mathbf{v}_j].
\end{aligned}$$

The proofs of $d_j^{\phi_a}$, $d_{ij}^{\phi_a}$, $d_j^{\phi_b}$, $d_{ij}^{\phi_b}$ follow similarly. Now, d_j^l and d_{ij}^l follow using the general rules of differentiation. Hence the theorem.

As we mentioned in section 3.2, our criterion function is not concave and (given that first order conditions are met) we will check for second-order conditions by dealing with our summation constraint via the substitution $p_J = 1 - (p_1 + p_2 + \dots + p_{J-1})$. Let the criterion function $\phi(\mathbf{p})$, in terms of $p_1, p_2, \dots, p_{J-1}$ be $\tilde{\phi}(\mathbf{p})$. We will check that standard conditions on the Hessian matrix of $\tilde{\phi}(\mathbf{p})$ (2.20) are satisfied.

Building upon our work from Chapter 4, we now apply our method-

ology to the quadratic regression model. Our goal is to find an optimal design $\mathbf{p}$ that minimizes the correlation between the estimators of the parameters θ_2 and θ_3 . To achieve this, we maximize our criterion function, $\phi(\mathbf{p}) = -\left(\frac{k(\mathbf{p})}{\sqrt{\phi_a(\mathbf{p})}\sqrt{\phi_b(\mathbf{p})}}\right)^2$, with the vectors $\mathbf{a} = (0, 1, 0)$ and $\mathbf{b} = (0, 0, 1)$. We conducted two separate experiments to evaluate how the multiplicative algorithm's convergence rate is affected by different choices of the transformation function $f(x, \delta)$ and the tuning parameter δ . In all experiments, we consider the case $J = 3$, and suppose that the support points are -1, 0, 2, started with an initial design where each weight was equal, $p_j^{(0)} = 1/3$, $j = 1, 2, 3$.

In the first experiment, we used the cumulative distribution function of the standard normal distribution, $f(x, \delta) = \Phi(\delta x)$, with a small $\delta = 0.001$. The initial values were:

$$\begin{aligned}\phi(\mathbf{p}) &= -0.7032967, \\ -0.4955923 &\leq F_i^\phi \leq 0.2869219.\end{aligned}$$

After 2771 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -1.029473 \times 10^{-16}, \\ -3.39566 \times 10^{-07} &\leq F_i^\phi \leq 9.999169 \times 10^{-9}.\end{aligned}$$

As a second experiment we take $f(F, \delta) = \exp(\delta F)/(1+\exp(\delta F))$, and $\delta = 0.005$.

Starting with equal initial weights, after 880 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -9.742137 \times 10^{-17}, \\ -3.303266 \times 10^{-7} &\leq F_i^\phi \leq 9.726687 \times 10^{-9}.\end{aligned}$$

The solution that was obtained is given by the design $\mathbf{p}^* = (0.4714865, 0.4999180, 0.0285955)$ with the corresponding support points -1, 0 and 2. Here $\mathbf{p}^*$ is of length 3, so the Hessian matrix ($H(\mathbf{p}^*)$) is of order 2×2 . At the optimum, we find that $\det(H(\mathbf{p}^*)) = 31.97052$, $d_{11}^{\tilde{\phi}}(\mathbf{p}^*) = -568.6079$, $d_{22}^{\tilde{\phi}}(\mathbf{p}^*) = -568.8143$, and the eigenvalues are $-1.14 \times 10^3, -2.81 \times 10^{-2}$. The design satisfies both the first- and second-order conditions. At the optimum $\mathbf{p}^*$, the criterion attains the maximum possible value of zero. Thus, we have a zero correlation between estimators of θ_2 and θ_3 . Note that different choices of the multiplicative function $f(\cdot, \cdot)$ and of the tuning parameter δ have an effect on the rate of convergence of the algorithm.

While our previous example successfully achieved a zero correlation between parameter estimators, this is not always possible in real-world applications. A practical problem arising in chemistry, as studied by Torsney and Alahmadi (1995) and Mandal et al. (2017), provides an excellent example where a zero correlation is not obtainable, and our goal shifts to constructing a design that makes the estimators as uncorrelated as possible.

The regression model in this example describes the relationship between the viscosity y and the concentration x of a chemical solution. Viscosity is the

response, and the model is given by:

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2.$$

The design interval is restricted to $(0, 0.2]$. The objective was to minimize the correlation between $\hat{\theta}_3$ and each of $\hat{\theta}_1$ and $\hat{\theta}_2$ as much as possible. In this example, we cannot obtain a design that produces a common value of zero correlation. We construct two optimal designs, one for minimizing correlation between $\hat{\theta}_1$ and $\hat{\theta}_3$, and the other for minimizing correlation between $\hat{\theta}_2$ and $\hat{\theta}_3$.

In the first case, we focused on minimizing the correlation between the estimators of $\hat{\theta}_1$ and $\hat{\theta}_3$. We defined the vectors for the linear combinations as $\mathbf{a} = (1, 0, 0)^\top$, $\mathbf{b} = (0, 0, 1)^\top$. The design interval is short and half-closed, so we discretized the design space at intervals of 0.01 between 0.02 and 0.20. By making an initial run of the algorithm (2.23) in this discretized space, we find that the support points are 0.02, 0.11, 0.20. For this example, we use a multiplicative function of the form $f(d, \delta) = (1 + s\alpha d)^{s\delta}$, $s = \text{sign}(d)$, a function with two free parameters δ and α . Starting with equal initial weights and running the algorithm with $\delta = 0.02$ and $\alpha = 0.1$, we obtain the following initial values

$$\begin{aligned} \phi(\mathbf{p}) &= -0.9210763, \\ -0.06178723 &\leq F_i^\phi \leq 0.03606589. \end{aligned}$$

After 56000 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -0.8037895, \\ -0.05926553 &\leq F_i^\phi \leq 5.768531 \times 10^{-5}.\end{aligned}$$

We obtain the solution as given by the design $\mathbf{p}^* = (0.00088024, 0.99902770, 0.00009207)$ with the corresponding support points 0.02, 0.11 and 0.20. Here the optimal $\mathbf{p}^*$ is of length 3, so $H(\mathbf{p}^*)$ is of order 2×2 . Based on the above solution, we find that $\det(H(\mathbf{p}^*)) = 1.085472 \times 10^{14}$, $d_{11}^{\tilde{\phi}}(\mathbf{p}^*) = -102393579$ and $d_{22}^{\tilde{\phi}}(\mathbf{p}^*) = -131637232$. Eigenvalues are $-233566071.9, -464738.7$. The design satisfies both the first- and second-order conditions. So we have found an optimal design for which the criterion function attains a local optimum.

In the second case, since we seek to minimize the correlation between $\hat{\theta}_2$ and $\hat{\theta}_3$, we take $\mathbf{a} = (0, 1, 0)^\top$, $\mathbf{b} = (0, 0, 1)^\top$ and again start with equal initial weights. We run the algorithm with the same setup as above and find that the support points are 0.02, 0.13 and 0.20. We use the same multiplicative function $f(d, \delta) = (1 + s\alpha d)^{s\delta}$. Starting with equal initial weights and running the algorithm with $\delta = 0.02$ and $\alpha = 0.1$, we obtain the the following initial values

$$\begin{aligned}\phi(\mathbf{p}) &= -0.6666912, \\ -0.1011418 &\leq F_i^\phi \leq 0.1348566.\end{aligned}$$

After 60000 iterations, we obtain

$$\begin{aligned}\phi(\mathbf{p}) &= -0.5491246, \\ -0.04986593 &\leq F_i^\phi \leq 6.467381 \times 10^{-5}.\end{aligned}$$

In this second case, we obtain the solution as given by the design $\mathbf{p}^* = (0.00115067, 0.99870472, 0.00014461)$ with the corresponding support points 0.02, 0.13 and 0.20. Here $\mathbf{p}^*$ is of length 3. Thus, $H(\mathbf{p}^*)$ is of order 2×2 . Based on the above solution, we find that $\det(H(\mathbf{p}^*)) = 7.333524 \times 10^{14}$, $d_{11}^{\tilde{\phi}}(\mathbf{p}^*) = -204861938$ and $d_{22}^{\tilde{\phi}}(\mathbf{p}^*) = -551774507$. Eigenvalues are $-755665972.7, -970471.6$. The design satisfies both the first and second order conditions. Therefore, we have determined an optimal design for which the criterion function attains a local optimum. In both cases, while a zero correlation was not achievable, the algorithms found designs that minimize the correlation to its lowest possible value, confirming the practicality of this methodology.

5.3 Compound optimization: Minimizing two or more correlations

We consider the problem of simultaneously driving the correlations between two or more pairs of parameter estimators as close to zero as possible. For instance, this arises when one parameter estimator needs to be independent of two other estimators in a regression model, or more generally, when two or more distinct pairs of estimators need to be made independent of each other. We

solve this optimization problem by employing a compound optimization criterion and then by using simultaneous optimization techniques. The core principle of transforming a minimization problem into a maximization problem using a squared penalty function. We now adapt this technique to the context of multiple correlations.

Suppose we have four linear combinations of parameters, given by $\mathbf{a}^\top \boldsymbol{\theta}$, $\mathbf{b}^\top \boldsymbol{\theta}$, $\mathbf{c}^\top \boldsymbol{\theta}$ and $\mathbf{d}^\top \boldsymbol{\theta}$, where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^k$. We wish to minimize the correlation between the estimators of $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$ and also between the estimators of $\mathbf{c}^\top \boldsymbol{\theta}$ and $\mathbf{d}^\top \boldsymbol{\theta}$. Let us denote the two correlations as given by

$$l_1(\mathbf{p}) = \text{corr}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) \text{ and } l_2(\mathbf{p}) = \text{corr}(\mathbf{c}^\top \hat{\boldsymbol{\theta}}, \mathbf{d}^\top \hat{\boldsymbol{\theta}}).$$

Suppose that we wish to find designs satisfying these two correlations equal to two fixed constants, say c_1 and c_2 , respectively. This can be achieved by transforming the problem into a maximization problem in which we simultaneously maximize

$$\phi_1(\mathbf{p}) = -[l_1(\mathbf{p}) - c_1]^2 \text{ and } \phi_2(\mathbf{p}) = -[l_2(\mathbf{p}) - c_2]^2. \quad (5.7)$$

Thus, $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$ are both zero at the optimum. If we want the two correlations to be zero, we simply set c_1 and c_2 to zero.

Here we will need two sets of derivatives for the criterion functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. For appropriate choices of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$, these can be obtained by using Theorem 5.2. The implication here is that if there is a solution to our problem, then the optimal design $\mathbf{p}^*$ should simultaneously maximize the two

functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$, with a common maximum of zero. Note that this is also subject to the remaining constraints $p_j \geq 0$ and $\sum p_j = 1$. The sharing of a common optimal value of zero means that we have transformed the original optimization problem to a compound optimization problem with various choices of $\phi(\mathbf{p})$, namely,

$$\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}) \quad (5.8)$$

$$\phi(\mathbf{p}) = \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\} \quad (5.9)$$

$$\phi(\mathbf{p}) = \alpha \phi_1(\mathbf{p}) + (1 - \alpha) \phi_2(\mathbf{p}), 0 < \alpha < 1. \quad (5.10)$$

In the case of minimizing two correlations, we transformed the problem to that of maximizing two functions of the design weights simultaneously. Because each of the criterion functions ϕ_1 and ϕ_2 is negative (rather than for those $\mathbf{p}$ in which they are zero) and has a common maximum of zero, the sum of these (5.8) will have a maximum of zero at the common optimizing $\mathbf{p}^*$.

Alternatively, we solve the problem as a *Maximin* problem in which we maximize the minimum of the functions (5.9), and, following the same arguments above, $\min\{\phi_1, \phi_2\}$ will be zero at the optimum $\mathbf{p}^*$.

Another alternative is that we can maximize a convex combination of the two criterion functions (5.10). That is, we maximize $\alpha \phi_1(\mathbf{p}) + (1 - \alpha) \phi_2(\mathbf{p})$ for an α , $0 < \alpha < 1$. Therefore, for these types of problems, we need to use appropriate extension of the algorithm (2.23). The choice of the first argument of the function $f(d, \delta)$ depends on the type of optimization problem. For the

above three types of optimization problems, we use the following

$$d_j = d_j^{\phi_1} + d_j^{\phi_2} \text{ if we maximize } \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}) \quad (5.11)$$

$$d_j = d_j^{\min\{\phi_1, \phi_2\}} \text{ if we maximize } \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\} \quad (5.12)$$

$$d_j = d_j^{\alpha\phi_1} + d_j^{(1-\alpha)\phi_2} \text{ if we maximize } \alpha\phi_1(\mathbf{p}) + (1-\alpha)\phi_2(\mathbf{p}) \quad (5.13)$$

where $d_j^{\phi_i}$ denotes the partial derivative of ϕ_i with respect to p_j , $i = 1, 2$, $j = 1, 2, \dots, J$. As we have seen previously, replacing these partial derivatives with their corresponding directional derivatives can significantly improve the convergence rate of the algorithm.

We now consider the following example, related to those in Sections 3.2 and 4.3. The design space is defined by a quadrilateral as given below. The model is given by

$$E(y|\mathbf{x}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 = \mathbf{v}^\top \boldsymbol{\theta}, \mathbf{v} \in \mathcal{V}$$

where

$$\mathcal{V} = \{\mathbf{v} = (1, x_1, x_2)^\top : (x_1, x_2) \in \mathcal{Q}\},$$

and $\mathcal{Q}$ is a quadrilateral with vertices $(-1, -1), (-1, 1), (1, -1), (2, 3)$. Our design points are the four vertices of the above quadrilateral. We would like to find an optimal design such that $\hat{\theta}_1$ is uncorrelated with both $\hat{\theta}_2$ and $\hat{\theta}_3$. So, in the first criterion $\phi_1(\mathbf{p})$, we take $\mathbf{a} = (1, 0, 0)^\top$, $\mathbf{b} = (0, 1, 0)^\top$, and in the second criterion $\phi_2(\mathbf{p})$, we take $\mathbf{c} = \mathbf{a} = (1, 0, 0)^\top$, $\mathbf{d} = (0, 0, 1)^\top$.

First we will consider the summation criterion (5.8), i.e., $\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p})$. We use the multiplicative function $f(d, \delta) = \exp(\delta d)/(1 + \exp(\delta d))$ where d are the partial derivatives d_j as given in (5.11). We take the initial design to be $p_j^{(0)} = 1/4$, $j = 1, \dots, 4$, set $\delta = 0.05$ and apply the algorithm (2.23). We obtain the initial values as given by

$$\begin{aligned}(\phi_1, \phi_2) &= (-0.001540832, -0.05084746), \\ -0.1174356 &\leq F_i^{\phi_1} \leq 0.09099693, \\ -0.4800664 &\leq F_i^{\phi_2} \leq 0.5745475.\end{aligned}$$

After 603 iterations, we obtain

$$\begin{aligned}(\phi_1, \phi_2) &= (-2.280025 \times 10^{-17}, -2.280025 \times 10^{-17}), \\ -1.161573 \times 10^{-8} &\leq F_i^{\phi_1} \leq 1.161573 \times 10^{-8}, \\ -1.437969 \times 10^{-8} &\leq F_i^{\phi_2} \leq 1.066154 \times 10^{-8}.\end{aligned}$$

The algorithm converged to the following optimal design:

$$\mathbf{p}^* = \begin{Bmatrix} (-1, -1) & (-1, 1) & (1, -1) & (2, 3) \\ 0.3547995 & 0.2161604 & 0.2871203 & 0.1419198 \end{Bmatrix}.$$

We now consider the *maximin* criterion (5.9), that is, we maximize the criterion $\phi(\mathbf{p}) = \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}$. We use the same multiplicative function $f(d, \delta)$, the same δ , and the same starting values as above. After 1194 iterations,

we obtain

$$\begin{aligned}(\phi_1, \phi_2) &= (-1.805071 \times 10^{-17}, -1.939839 \times 10^{-17}), \\ -1.034291 \times 10^{-8} &\leq F_i^{\phi_1} \leq 1.034291 \times 10^{-8}, \\ -1.325102 \times 10^{-8} &\leq F_i^{\phi_2} \leq 9.842485 \times 10^{-9}.\end{aligned}$$

The algorithm converged to the following optimal design:

$$\mathbf{p}^* = \begin{Bmatrix} (-1, -1) & (-1, 1) & (1, -1) & (2, 3) \\ 0.3551678 & 0.2158658 & 0.2868993 & 0.1420671 \end{Bmatrix}.$$

If we round to three decimal places, we obtain the same solution in both compound optimization problems, in which case the optimal weights are given by 0.355, 0.216, 0.287 and 0.142 corresponding to the four vertices $(-1, -1)$, $(-1, 1)$, $(1, -1)$, $(2, 3)$. In both compound optimization problems, the two criterion values are zero. We do not need to calculate the eigenvalues of the Hessian matrix as we achieved the global maximum in our optimal solution. Therefore, we found an optimal design at which both criterion functions attain their maximum possible value of zero, thereby demonstrating that we have simultaneously achieved both correlations of zero.

Motivated by a practical problem arising in chemistry, we have addressed an optimization problem in which we minimize correlations between parameter estimators of a linear model. If the maximum value of zero is attained the two parameters or the two linear combinations of the parameters are estimated with zero correlation. If there does not exist a solution such that the correlation is

zero, then maximizing the above criterion will give a design with minimum correlation. We also addressed situations where it was desired to minimize two or more correlations at the same time. In order to accomplish this, we explored a compound optimization problem and used simultaneous optimization techniques to solve it.

We have faced a number of challenges throughout the work. First, the correlation criterion is neither convex nor concave. So standard optimality conditions such as the general equivalence theorem are not sufficient. Second, both the partial and directional derivatives of the correlation criterion and the complex nature of the form of the criterion make it challenging to implement. In some instances we had to experiment with several multiplicative functions, the first argument of these functions, and the free (tuning) parameter of the algorithm in order to satisfy the optimality conditions. In some instances, we had to use two tuning parameters in the multiplicative function. We have successfully overcome these challenges and constructed the best possible designs for a number of interesting models.

Note that while independence of parameter estimators is the ideal, achieving this goal is far more challenging. Since a lack of correlation in this situation indicates that the parameter estimators are independent, we can achieve this under the assumption of a Gaussian distribution along with the methodology developed in this chapter.

As mentioned in the previous chapter, the study of covariance motivated the investigation of correlation structures of this chapter. The key difference between minimizing covariance-based criteria and minimizing correlation-based criteria

lies in both the scale of the dependence measure and the level of complexity involved in the analysis.

Covariance-based criteria aim to reduce the covariance between parameter estimators. This directly targets the dependence structure and can be used to construct designs where estimators are uncorrelated. However, covariance is scale-dependent, meaning that its magnitude is influenced by the units and variances of the parameters, which limits its interpretability when comparing different estimators.

In contrast, correlation-based criteria normalize covariance by the variances of the estimators, resulting in a scale-free measure of dependence. This makes correlation more appropriate for comparing relationships across parameters and provides a more interpretable notion of independence. For this reason, minimizing correlation can be seen as a stronger and more meaningful objective than minimizing covariance in many practical settings.

From a methodological perspective, correlation-based criteria are significantly more challenging to handle. Since correlation is a nonlinear function involving both covariance and variances. Furthermore, first-order optimality conditions are not sufficient, and second-order conditions must be derived and verified for this complex criterion, making the analysis more complex than in the covariance-based case.

In summary, correlation minimization provides a normalized and more interpretable measure of dependence, at the cost of increased mathematical and computational complexity. This progression from covariance to correlation represents an important extension of the methodology developed in this dissertation.

Chapter 6

Optimal designs under constraints on independence estimation of parameters

6.1 Introduction

In the preceding chapters we studied optimal designs in the absence of explicit constraints, focusing on criteria that minimize variances, covariances, or correlations of parameter estimators. In practice, however, one often encounters situations where designs must satisfy additional constraints. These may arise from scientific, practical, or ethical considerations. For instance, in a chemical experiment the range of concentrations may be physically limited; in clinical studies, the number of patients allocated to each treatment arm may be restricted

by ethical or budgetary factors. In such cases, designs that are optimal in an unconstrained sense may no longer be feasible or desirable.

While classical optimal design theory provides powerful, single-criterion methods like D -, G -, and A -optimality, these approaches are inherently insufficient when the design must satisfy a complex constraint, especially one related to the structure of the variance-covariance matrix of the parameter estimators (see, e.g., López-Fidalgo (2023), Fedorov and Leonov (2014), Berger and Wong (2009), Torsney and Mandal (2001), Wang et al. (2015)). One particular statistical requirement is the demand for a specific variance-covariance structure, such as ensuring that certain parameters or parametric functions are estimated independently (i.e., their correlation is precisely zero or as small as possible) under a Gaussian model.

A key motivation for moving beyond single-criterion methods involves the need for simultaneous optimization. This arises in situations where researchers must optimize more than one function at the same time. For example, one may wish to minimize covariance or correlation for a pair of estimators while simultaneously maximizing overall precision (e.g., through D -optimality). In such complex scenarios, standard criteria alone are inadequate, and specialized simultaneous optimization methods, often involving maximizing the sum or minimum of multiple criteria, become essential tools for resolving conflicts between competing objectives.

The core challenge in solving the general constrained optimization problem is maximizing a design criterion subject to a constraint (such as zero correlation) and the typical mass constraints on the design weights. To tackle this, we first

formulate the constrained optimization problem using Lagrangian theory. We invoke the general equivalence theorem of Kiefer (1974) using the directional derivatives of the Lagrangian function. We then transform the attainment of these conditions into a criterion whose optimal value is zero. In order to ensure the constraints and to satisfy the optimality conditions, we set up a compound criterion and solve the problem via simultaneous optimization. In order to solve the simultaneous optimization problem, we propose two methods, namely maximizing the sum of the functions and maximizing the minimum of the functions (maximin approach). We present both the mathematical formulation of these problems and iterative algorithms for their numerical solution. We adapt the multiplicative algorithms to accommodate constraints, and develop practical stabilization strategies to facilitate convergence. We also make use of the properties of the directional derivatives to facilitate convergence of the algorithms.

We first formulate the general constrained design problem and specify the forms of constraints of interest. We then present the appropriate algorithms and the corresponding partial and directional derivatives. We then provide illustrative examples, including interpretations of the resulting figures. These examples demonstrate how constrained designs differ from their unconstrained counterparts, and highlight the practical relevance of this methodology.

6.2 Constrained optimization with generalized or average variance subject to the constraints on covariance and correlation

6.2.1 General Problem (GP)

Let $\phi(\mathbf{p})$ denote an optimality criterion of interest, such as those considered in earlier chapters. In the unconstrained setting, the goal is to maximize (or minimize) $\phi(\mathbf{p})$ over the general problem (2.16)

$$\mathcal{P} = \left\{ \mathbf{p} = (p_1, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1 \right\}.$$

In the constrained setting, we introduce an additional condition of the form

$$c(\mathbf{p}) = 0,$$

where $c(\mathbf{p})$ is a scalar function depending on the design weights. The optimization problem then is to maximize $\phi(\mathbf{p})$ subject to $\mathbf{p} \in \mathcal{P}$ and $c(\mathbf{p}) = 0$. $\phi(\mathbf{p})$ could be one of the criteria A -, c -, D -, or D_A -optimality.

We consider the following types of constraints $c(\mathbf{p})$.

- Covariance constraint:

$$c(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} \quad \text{if we want } \text{Cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) = 0.$$

- Correlation constraint:

$$c(\mathbf{p}) = \frac{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}} \sqrt{\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}} \quad \text{if we want } \text{corr}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}}) = 0.$$

These criteria were already considered in Chapters 4 and 5 in the unconstrained setting.

6.2.2 Methodology

To solve the constrained problem, we introduce the Lagrangian. First, we establish first-order conditions for a constrained local maximum using Lagrangian theory for the equality constraints $c(\mathbf{p}) = 0$ and $\sum p_i = 1$.

Let

$$L(\phi, \mathbf{p}, \lambda) = \phi(\mathbf{p}) + \lambda c(\mathbf{p}) + \mu \left(\sum_i p_i - 1 \right) \quad (6.1)$$

where $\phi(\mathbf{p})$ is either the D - or A -optimality criterion. For D -optimality criterion we take $\phi(\mathbf{p}) = -\log \det\{\mathbf{M}^{-1}(\mathbf{p})\}$. For A -optimality we take $\phi(\mathbf{p}) = -\text{tr}(\mathbf{A}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A})$, where $\mathbf{A} = [\mathbf{a}, \mathbf{b}]$. The reason for defining the matrix $\mathbf{A}$ with this structure is that we wish to minimize the total or average variance of $\mathbf{a}^\top \hat{\boldsymbol{\theta}}$ and $\mathbf{b}^\top \hat{\boldsymbol{\theta}}$.

Then for $i = 1, 2, \dots, J$, the directional derivative of the Lagrangian with respect

to p_i is

$$\begin{aligned}
d_i^L &= \frac{\partial L}{\partial p_i} \\
&= \frac{\partial \phi}{\partial p_i} + \lambda \frac{\partial c}{\partial p_i} + \mu \\
&= d_i^\phi + \lambda d_i^c + \mu
\end{aligned} \tag{6.2}$$

where

$$\begin{aligned}
d_i^\phi &= \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i \quad \text{for } D\text{-optimality} \\
&= \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A}^\top \mathbf{A} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i \quad \text{for the above version of } A\text{-optimality}
\end{aligned}$$

and d_i^c depends on either the covariance or correlation constraint. These are given in the next section.

Invoking the general equivalence theorem (Kiefer, 1974), the necessary condition for a local maximum is that $d_i^L = 0$ for all design points i with $p_i > 0$. Summing $p_i d_i^L = 0$ over all i and applying the constraint $\sum p_i = 1$ allows us to solve for μ :

$$\mu = - \sum_{i=1}^J p_i (d_i^\phi + \lambda d_i^c).$$

Substituting this expression back into Equation (6.2) with $d_i^L = 0$, we obtain the condition that must be satisfied by an optimal design $\mathbf{p}^*$:

$$d_i^\phi + \lambda d_i^c = \sum_{k=1}^J p_k (d_k^\phi + \lambda d_k^c). \tag{6.3}$$

Using the derivative d_i^L obtained previously, the vertex directional derivatives of

the Lagrangian are defined as

$$\begin{aligned}
F_i^L &= d_i^L - \sum_{k=1}^J p_k d_k^L \\
&= d_i^\phi + \lambda d_i^c - \sum_{k=1}^J p_k (d_k^\phi + \lambda d_k^c) \\
&= \left(d_i^\phi - \sum_{k=1}^J p_k d_k^\phi \right) + \lambda \left(d_i^c - \sum_{k=1}^J p_k d_k^c \right) \\
&= F_i^\phi + \lambda F_i^c,
\end{aligned} \tag{6.4}$$

where we have defined the vertex directional derivatives $F_i^\phi = d_i^\phi - \sum_{k=1}^J p_k d_k^\phi$ and $F_i^c = d_i^c - \sum_{k=1}^J p_k d_k^c$. Equivalently, this relationship can be expressed in vector notation as

$$\mathbf{F}^L = \mathbf{F}^\phi + \lambda \mathbf{F}^c. \tag{6.5}$$

Applying the equivalence theorem, the condition for an optimal design $\mathbf{p}^*$ is $F_i^L = 0$ for all support points. This yields the system of equations:

$$\begin{aligned}
\mathbf{F}^L &= \mathbf{0} \\
\Rightarrow \quad \mathbf{F}^c \lambda &= -\mathbf{F}^\phi.
\end{aligned} \tag{6.6}$$

Suppose we define the vectors $\mathbf{A} = \mathbf{F}^c$ and $\mathbf{b} = -\mathbf{F}^\phi$ and treat λ as the unknown scalar $\lambda = \lambda$. Then λ must satisfy the linear system

$$\mathbf{A} \lambda = \mathbf{b}. \tag{6.7}$$

If a solution exists, the set of solutions to (6.6) is given by

$$\boldsymbol{\lambda} = \mathbf{A}^- \mathbf{b} + (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{z} \quad \text{for any vector } \mathbf{z}, \quad (6.8)$$

where $\mathbf{A}^-$ is any generalized inverse of $\mathbf{A}$. A solution exists if and only if the consistency condition $\mathbf{A} \mathbf{A}^- \mathbf{b} = \mathbf{b}$ is satisfied.

Assuming $\mathbf{A}^\top \mathbf{A}$ is nonsingular, one choice of the generalized inverse $\mathbf{A}^-$ is the Moore-Penrose inverse, given by

$$\begin{aligned} \mathbf{A}^- &= (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \\ &= [(\mathbf{F}^c)^\top \mathbf{F}^c]^{-1} (\mathbf{F}^c)^\top. \end{aligned} \quad (6.9)$$

Then, the particular solution for $\boldsymbol{\lambda}$ is

$$\mathbf{A}^- \mathbf{b} = -\frac{(\mathbf{F}^c)^\top \mathbf{F}^\phi}{(\mathbf{F}^c)^\top \mathbf{F}^c} = \hat{\boldsymbol{\lambda}}.$$

This particular solution $\hat{\boldsymbol{\lambda}}$ satisfies the consistency condition $\mathbf{A} \mathbf{A}^- \mathbf{b} = \mathbf{b}$ if and only if $\mathbf{A} \hat{\boldsymbol{\lambda}} = \mathbf{b}$, which simplifies to

$$\mathbf{A} \mathbf{A}^- \mathbf{b} = \hat{\boldsymbol{\lambda}} \mathbf{A} = \hat{\boldsymbol{\lambda}} \mathbf{F}^c.$$

Thus, an optimal design $\mathbf{p}^*$ should be such that the necessary condition is met:

$$\hat{\boldsymbol{\lambda}} \mathbf{F}^c = -\mathbf{F}^\phi \quad \text{i.e.,} \quad \hat{\boldsymbol{\lambda}} \mathbf{F}^c + \mathbf{F}^\phi = \mathbf{0}. \quad (6.10)$$

Substituting the expression for $\hat{\lambda}$ into the condition (6.10) yields

$$-\frac{(\mathbf{F}^c)^\top \mathbf{F}^\phi}{(\mathbf{F}^c)^\top \mathbf{F}^c} \mathbf{F}^c + \mathbf{F}^\phi = \mathbf{0}. \quad (6.11)$$

To eliminate the denominator and obtain an equivalent vector condition, we multiply by $[(\mathbf{F}^c)^\top \mathbf{F}^c]$:

$$\mathbf{w} = [(\mathbf{F}^c)^\top \mathbf{F}^c] \mathbf{F}^\phi - [(\mathbf{F}^c)^\top \mathbf{F}^\phi] \mathbf{F}^c = \mathbf{0}, \quad (6.12)$$

where the i -th component of the vector $\mathbf{w}$ is

$$\begin{aligned} w_i &= [(\mathbf{F}^c)^\top \mathbf{F}^c] F_i^\phi - [(\mathbf{F}^c)^\top \mathbf{F}^\phi] F_i^c \\ &= \left[\sum_{k=1}^J (F_k^c)^2 \right] F_i^\phi - \left[\sum_{k=1}^J F_k^c F_k^\phi \right] F_i^c. \end{aligned} \quad (6.13)$$

Since the optimal design $\mathbf{p}^*$ must satisfy $\mathbf{w} = \mathbf{0}$, this is equivalent to minimizing the squared Euclidean norm of $\mathbf{w}$:

$$\mathbf{w}^\top \mathbf{w} = 0. \quad (6.14)$$

Since $\mathbf{w}^\top \mathbf{w} \geq 0$ always holds, the optimal design $\mathbf{p}^*$ should minimize $\mathbf{w}^\top \mathbf{w}$ (with a minimum value of zero) or maximize the criterion $W = W_1 = -\mathbf{w}^\top \mathbf{w}$ with a maximum value of zero.

Alternatively, we can expand the squared norm $\mathbf{w}^\top \mathbf{w}$:

$$\begin{aligned}
\mathbf{w}^\top \mathbf{w} &= \left[(\mathbf{F}^c)^\top \mathbf{F}^c \mathbf{F}^\phi - (\mathbf{F}^c)^\top \mathbf{F}^\phi \mathbf{F}^c \right]^\top \left[(\mathbf{F}^c)^\top \mathbf{F}^c \mathbf{F}^\phi - (\mathbf{F}^c)^\top \mathbf{F}^\phi \mathbf{F}^c \right] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] \\
&\quad - 2 [(\mathbf{F}^c)^\top \mathbf{F}^c] [(\mathbf{F}^c)^\top \mathbf{F}^\phi] [(\mathbf{F}^\phi)^\top \mathbf{F}^c] \\
&\quad + [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 [(\mathbf{F}^c)^\top \mathbf{F}^c] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] - [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 [(\mathbf{F}^c)^\top \mathbf{F}^c] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c] \left([(\mathbf{F}^c)^\top \mathbf{F}^c] [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] - [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 \right). \tag{6.15}
\end{aligned}$$

This can be factored using the squared coefficient of determination (R_1) between the vectors $\mathbf{F}^\phi$ and $\mathbf{F}^c$, and is given below

$$\begin{aligned}
\mathbf{w}^\top \mathbf{w} &= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] \left[1 - \frac{[(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2}{[(\mathbf{F}^\phi)^\top \mathbf{F}^\phi][(\mathbf{F}^c)^\top \mathbf{F}^c]} \right] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] [1 - R_1], \tag{6.16}
\end{aligned}$$

where R_1 is defined as

$$R_1 = \frac{[(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2}{[(\mathbf{F}^\phi)^\top \mathbf{F}^\phi][(\mathbf{F}^c)^\top \mathbf{F}^c]} = \frac{\xi_1^2}{\xi_2 \xi_3}, \tag{6.17}$$

and the scalar components are given by

$$\xi_1 = (\mathbf{F}^c)^\top \mathbf{F}^\phi, \quad \xi_2 = (\mathbf{F}^\phi)^\top \mathbf{F}^\phi, \quad \xi_3 = (\mathbf{F}^c)^\top \mathbf{F}^c. \tag{6.18}$$

In terms of ξ_1, ξ_2 and ξ_3 , the squared norm is

$$\mathbf{w}^\top \mathbf{w} = \xi_2 \xi_3^2 - \xi_3 \xi_1^2. \tag{6.19}$$

Clearly $0 \leq R_1 \leq 1$. By Cauchy–Schwarz inequality, the condition for $\mathbf{w} = \mathbf{0}$ is $R_1 = 1$. Thus, $\mathbf{p}^*$ should maximize the criterion $W = W_2 = [R_1 - 1]$, which attains its maximum value of zero at the optimal design.

Additionally, more cases for choosing the optimal criterion W can be considered. From the first order optimality conditions,

$$F_j^L = \begin{cases} 0, & \text{if } p_j^* > 0, \\ \leq 0, & \text{if } p_j^* = 0, \end{cases} \quad j = 1, 2, \dots, J.$$

There are two ways to adapt the above approach. One is to utilize the condition $F_j^L = 0$ when $p_j^* > 0$, which implies that the vector $\mathbf{w}$ given in (6.12) must be zero at all support points. Thus, we can define a weighted zero-condition $\mathbf{l} = \mathbf{P}\mathbf{w} = \mathbf{0}$, where $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_J)$.

$$\begin{aligned} l_i &= p_i w_i = 0 \quad \text{for } i = 1, 2, \dots, J, \\ \text{i.e., } \mathbf{l} &= \mathbf{P}\mathbf{w} = \mathbf{0}. \end{aligned} \tag{6.20}$$

By replacing $\mathbf{w}$ with $\mathbf{l}$ in the squared norm argument, we conclude that the optimal design $\mathbf{p}^*$ should maximize the criterion

$$W = W_3 = -\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}. \tag{6.21}$$

We rewrite the criterion $\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}$ using the vector definitions:

$$\begin{aligned}
\mathbf{w}^\top \mathbf{P}^2 \mathbf{w} &= \left[((\mathbf{F}^c)^\top \mathbf{F}^c) \mathbf{F}^\phi - ((\mathbf{F}^c)^\top \mathbf{F}^\phi) \mathbf{F}^c \right]^\top \mathbf{P}^\top \mathbf{P} \left[((\mathbf{F}^c)^\top \mathbf{F}^c) \mathbf{F}^\phi - ((\mathbf{F}^c)^\top \mathbf{F}^\phi) \mathbf{F}^c \right] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{P}^2 \mathbf{F}^\phi] + [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^c] \\
&\quad - 2 [(\mathbf{F}^c)^\top \mathbf{F}^c] [(\mathbf{F}^c)^\top \mathbf{F}^\phi] [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^\phi] \\
&= [(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{P}^2 \mathbf{F}^\phi] + [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^c] [1 - R_2],
\end{aligned} \tag{6.22}$$

where R_2 is defined as

$$\begin{aligned}
R_2 &= \frac{2 [(\mathbf{F}^c)^\top \mathbf{F}^c] [(\mathbf{F}^c)^\top \mathbf{F}^\phi] [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^\phi]}{[(\mathbf{F}^c)^\top \mathbf{F}^c]^2 [(\mathbf{F}^\phi)^\top \mathbf{P}^2 \mathbf{F}^\phi] + [(\mathbf{F}^c)^\top \mathbf{F}^\phi]^2 [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^c]} \\
&= \frac{2\xi_1 \xi_3 \xi_4}{\xi_3^2 \xi_5 + \xi_1^2 \xi_6},
\end{aligned} \tag{6.23}$$

with the new scalar components:

$$\xi_4 = (\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^\phi, \quad \xi_5 = (\mathbf{F}^\phi)^\top \mathbf{P}^2 \mathbf{F}^\phi, \quad \xi_6 = (\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^c, \tag{6.24}$$

and ξ_1, ξ_2, ξ_3 as defined in (6.18). Thus, the squared norm can be concisely written as

$$\mathbf{w}^\top \mathbf{P}^2 \mathbf{w} = \xi_3^2 \xi_5 + \xi_1^2 \xi_6 - 2\xi_1 \xi_3 \xi_4. \tag{6.25}$$

We can now prove that $R_2 \leq 1$. Let the weighted directional derivative vectors be defined as

$$\mathbf{s} = \mathbf{P} \mathbf{F}^\phi, \quad \mathbf{t} = \mathbf{P} \mathbf{F}^c.$$

Then R_2 can be written in terms of $\mathbf{s}$ and $\mathbf{t}$ as

$$\begin{aligned}
R_2 &= \frac{2\xi_3\xi_1(\mathbf{s}^\top\mathbf{t})}{\xi_3^2(\mathbf{s}^\top\mathbf{s}) + \xi_1^2(\mathbf{t}^\top\mathbf{t})} \\
&= \frac{2\mathbf{u}^\top\mathbf{v}}{\mathbf{u}^\top\mathbf{u} + \mathbf{v}^\top\mathbf{v}} \quad [\mathbf{u} = \xi_3\mathbf{s}, \quad \mathbf{v} = \xi_1\mathbf{t}] \\
&= \frac{2\sum_{i=1}^J u_i v_i}{\sum_{i=1}^J u_i^2 + \sum_{i=1}^J v_i^2}, \tag{6.26}
\end{aligned}$$

where $\mathbf{u} = (u_1, u_2, \dots, u_J)^\top$ and $\mathbf{v} = (v_1, v_2, \dots, v_J)^\top$. Since u_i and v_i are real numbers, by the arithmetic-geometric mean inequality (or direct application of the Cauchy–Schwarz inequality to $\sum(u_i - v_i)^2 \geq 0$):

$$2u_i v_i \leq u_i^2 + v_i^2 \quad \text{which means} \quad 2\sum_{i=1}^J u_i v_i \leq \sum_{i=1}^J u_i^2 + \sum_{i=1}^J v_i^2.$$

Hence, from (6.26) it is proven that $R_2 \leq 1$. The optimal design $\mathbf{p}^*$ must maximize W_3 to attain its maximum value of zero.

This can also be seen directly from (6.22), since $\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}$ is non-negative. Because the first factor in the decomposition

$$\mathbf{w}^\top \mathbf{P}^2 \mathbf{w} = (\xi_3^2 \xi_5 + \xi_1^2 \xi_6)[1 - R_2]$$

is also non-negative, the second factor, $[1 - R_2]$, must be non-negative. Thus, $\mathbf{p}^*$ should maximize $W = W_4 = [R_2 - 1]$, yielding a maximum value of zero.

Another option for satisfying the optimality conditions is to consider the product of the directional derivative and the design weight, which is always zero

at the optimal design due to the support conditions:

$$\begin{aligned} E_j^L &= p_j F_j^L = 0 \quad \text{for } j = 1, 2, \dots, J, \\ \text{or } \mathbf{E}^L &= \mathbf{P} \mathbf{F}^L = \mathbf{0}. \end{aligned} \quad (6.27)$$

Substituting $\mathbf{E}^L$ for $\mathbf{F}^L$ in the previous derivation, we define the new vectors $\mathbf{E}^\phi = \mathbf{P} \mathbf{F}^\phi$ and $\mathbf{E}^c = \mathbf{P} \mathbf{F}^c$. We conclude that at $\mathbf{p}^*$ the new optimality vector is

$$\mathbf{w} = \mathbf{w}(\mathbf{E}) = \mathbf{0}. \quad (6.28)$$

Since $[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})] \geq 0$, $\mathbf{p}^*$ should maximize

$$W = W_5 = -[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})] \quad (6.29)$$

yielding a maximum of zero. Here we can also write the squared norm using the $\mathbf{E}$ vectors:

$$[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})] = [(\mathbf{E}^c)^\top \mathbf{E}^c]^2 [(\mathbf{E}^\phi)^\top \mathbf{E}^\phi] [1 - R_3], \quad (6.30)$$

where

$$R_3 = \frac{[(\mathbf{E}^c)^\top \mathbf{E}^\phi]^2}{[(\mathbf{E}^\phi)^\top \mathbf{E}^\phi][(\mathbf{E}^c)^\top \mathbf{E}^c]} = \frac{\xi_4^2}{\xi_5 \xi_6}. \quad (6.31)$$

In terms of the ξ_i 's defined in (6.24),

$$[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})] = \xi_5 \xi_6^2 - \xi_6 \xi_4^2. \quad (6.32)$$

Clearly from (6.31), $0 \leq R_3 \leq 1$. Thus, the maximum of W_5 (or minimum of

$[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})]$ is attained when R_3 is maximized, i.e.

$$[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})] = 0 \quad \Leftrightarrow \quad R_3 = 1.$$

Therefore $\mathbf{p}^*$ should maximize $W = W_6 = [R_3 - 1]$ yielding a maximum value of zero.

Therefore, we have used substitution to deal with the Lagrange multipliers in terms of $\mathbf{p}$ and transformed attainment of the first order conditions in the Lagrangian to an example of problem (2.16), where $\phi_L(\mathbf{p}) = W$ with several possible choices of $W(W_1, W_2, \dots, W_6)$ and the optimal value is zero. Alternatively, $\mathbf{p}^*$ must maximize any increasing function of W . We consider some such transformations, choosing them such that the maximum of the revised criterion is still zero, e.g., $(e^W - 1)$ or $[-\log(1 - W)]$.

There still remains the requirement of meeting the first-order conditions with respect to the Lagrange multipliers, that is, of ensuring the constraints. We consider constraints of the form $c(\mathbf{p}) = 0$. We deal with this by transformation to an optimization problem too; namely $\mathbf{p}^*$ should maximize $C(\mathbf{p}) = -[c(\mathbf{p})]^2$, subject to $p_j \geq 0, \sum p_j = 1$. So $C(\mathbf{p}^*) = 0$.

Thus, $\mathbf{p}^*$ should simultaneously maximize $C(\mathbf{p})$ and $W(\mathbf{p})$ with a common maximum of zero if $W(\mathbf{p}^*) = 0$, subject to $p_j \geq 0, \sum p_j = 1$. The sharing of a common optimal value of zero means that we have transformed the original constrained optimization problem to an example of (2.16) with various choices

of $\phi(\mathbf{p})$, namely

$$\begin{aligned}\phi_L(\mathbf{p}) &= \phi_1(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p}) \\ \phi_L(\mathbf{p}) &= \phi_2(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}.\end{aligned}$$

Hence, we can consider applying algorithms to solve the problem (2.16). We now consider appropriate algorithms to construct the optimal designs.

6.3 Algorithms

6.3.1 Iterative Algorithm and Derivatives

First we should maximize $W(\mathbf{p}) + C(\mathbf{p})$, or equivalently, maximize the minimum of the set of $\{W(\mathbf{p}), C(\mathbf{p})\}$, with two sets of derivatives, each of which must satisfy the optimality conditions and must share a common value at the optimum.

The iteration at the $(r+1)$ th step is

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(z_j^{(r)}, \delta)}{\sum_{j=1}^J p_j^{(r)} f(z_j^{(r)}, \delta)}, \quad (6.33)$$

where $f(z, \delta)$ is a positive and strictly increasing function in z which may depend on a positive parameter δ . The choice of z depends on the type of optimization problem.

$$z_j = F_j^W + F_j^C \quad \text{if we maximize } W + C, \quad (6.34)$$

$$z_j = F_j^{\min\{W, C\}} \quad \text{if we maximize } \min\{W, C\}, \quad (6.35)$$

where F_j^W, F_j^C are the directional derivatives of W, C , respectively, as given by

$$F_j^W = d_j^W - \sum_{k=1}^J p_k d_k^W, \quad F_j^C = d_j^C - \sum_{k=1}^J p_k d_k^C,$$

where d_j^W, d_j^C are first partial derivatives for the two functions W and C , respectively.

The first partial derivative $d_j^C = \frac{\partial C}{\partial p_j}$ is given by:

$$d_j^C = -2c(\mathbf{p}) \frac{\partial c(\mathbf{p})}{\partial p_j},$$

where the expression for $\frac{\partial c(\mathbf{p})}{\partial p_j}$ depends on the constraint $c(\mathbf{p})$. We focus on the covariance and correlation constraints. Let $\mathbf{M}^{-1} = \mathbf{M}^{-1}(\mathbf{p})$.

1. **Covariance constraint:** $c(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1} \mathbf{b}$. The first partial derivative of the constraint function is

$$\frac{\partial c(\mathbf{p})}{\partial p_j} = -\mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1} \mathbf{v}_j.$$

2. **Correlation constraint:** $c(\mathbf{p}) = \frac{\mathbf{a}^\top \mathbf{M}^{-1} \mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{M}^{-1} \mathbf{a}} \sqrt{\mathbf{b}^\top \mathbf{M}^{-1} \mathbf{b}}}$. The first partial derivative d_j^c is significantly more complex. We define the necessary scalar components:

$$\begin{aligned} v_a &= \mathbf{a}^\top \mathbf{M}^{-1} \mathbf{a}, \quad d_i^{v_a} = -(\mathbf{a}^\top \mathbf{M}^{-1} \mathbf{v}_i)^2 \\ v_b &= \mathbf{b}^\top \mathbf{M}^{-1} \mathbf{b}, \quad d_i^{v_b} = -(\mathbf{b}^\top \mathbf{M}^{-1} \mathbf{v}_i)^2 \\ k &= \mathbf{a}^\top \mathbf{M}^{-1} \mathbf{b}, \quad d_j^k = -\mathbf{v}_j^\top \mathbf{M}^{-1} \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1} \mathbf{v}_j. \end{aligned}$$

As derived in Chapter 5, d_j^c is given by:

$$d_j^c = v_a^{-\frac{1}{2}} v_b^{-\frac{1}{2}} \left[d_j^k + \frac{1}{2} k v_a^{-1} d_j^{v_a} + \frac{1}{2} k v_b^{-1} d_j^{v_b} \right].$$

The first partial derivative $d_j^W = \frac{\partial W}{\partial p_j}$ depends on the choice of the optimal criterion $W \in \{W_1, W_2, \dots, W_6\}$. We have explored various choices of W .

- For $W = -[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})]$,

$$\begin{aligned} d_j^W &= \frac{\partial W}{\partial p_j} \\ &= 2\xi_4\xi_6 \frac{\partial \xi_4}{\partial p_j} - \xi_6^2 \frac{\partial \xi_5}{\partial p_j} + [\xi_4^2 - 2\xi_5\xi_6] \frac{\partial \xi_6}{\partial p_j}. \end{aligned}$$

- For $W = -\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}$,

$$\begin{aligned} d_j^W &= \frac{\partial W}{\partial p_j} \\ &= 2[\xi_3\xi_4 - \xi_1\xi_6] \frac{\partial \xi_1}{\partial p_j} + 2[\xi_1\xi_4 - \xi_3\xi_5] \frac{\partial \xi_3}{\partial p_j} + 2\xi_1\xi_3 \frac{\partial \xi_4}{\partial p_j} - \xi_3^2 \frac{\partial \xi_5}{\partial p_j} - \xi_1^2 \frac{\partial \xi_6}{\partial p_j}. \end{aligned}$$

- For $W = -\mathbf{w}^\top \mathbf{w}$,

$$\begin{aligned} d_j^W &= \frac{\partial W}{\partial p_j} \\ &= 2\xi_1\xi_3 \frac{\partial \xi_1}{\partial p_j} - \xi_3^2 \frac{\partial \xi_2}{\partial p_j} + [\xi_1^2 - 2\xi_2\xi_3] \frac{\partial \xi_3}{\partial p_j}. \end{aligned}$$

where $\xi_1, \xi_2, \dots, \xi_6$ are given in (6.18) and (6.24), and $\frac{\partial \xi_i}{\partial p_j}$'s are given by

$$\begin{aligned}
\frac{\partial \xi_1}{\partial p_j} &= \sum_{i=1}^J \left[F_i^\phi \frac{\partial F_i^c}{\partial p_j} + F_i^c \frac{\partial F_i^\phi}{\partial p_j} \right] \\
\frac{\partial \xi_2}{\partial p_j} &= 2 \sum_{i=1}^J \left[F_i^\phi \frac{\partial F_i^\phi}{\partial p_j} \right] \\
\frac{\partial \xi_3}{\partial p_j} &= 2 \sum_{i=1}^J \left[F_i^c \frac{\partial F_i^c}{\partial p_j} \right] \\
\frac{\partial \xi_4}{\partial p_j} &= 2 p_j F_j^c F_j^\phi + \sum_{i=1}^J p_i^2 \left[F_i^\phi \frac{\partial F_i^c}{\partial p_j} + F_i^c \frac{\partial F_i^\phi}{\partial p_j} \right] \\
\frac{\partial \xi_5}{\partial p_j} &= 2 p_j (F_j^\phi)^2 + 2 \sum_{i=1}^J \left[p_i^2 F_i^\phi \frac{\partial F_i^\phi}{\partial p_j} \right] \\
\frac{\partial \xi_6}{\partial p_j} &= 2 p_j (F_j^c)^2 + 2 \sum_{i=1}^J \left[p_i^2 F_i^c \frac{\partial F_i^c}{\partial p_j} \right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial F_j^\phi}{\partial p_i} &= \frac{\partial^2 \phi}{\partial p_i \partial p_j} - \left[\frac{\partial \phi}{\partial p_i} + \sum_{t=1}^J p_t \frac{\partial^2 \phi}{\partial p_t \partial p_i} \right] \\
\frac{\partial F_j^c}{\partial p_i} &= \frac{\partial^2 c}{\partial p_i \partial p_j} - \left[\frac{\partial c}{\partial p_i} + \sum_{t=1}^J p_t \frac{\partial^2 c}{\partial p_t \partial p_i} \right].
\end{aligned}$$

Derivatives $\frac{\partial \phi}{\partial p_j}$ and $\frac{\partial^2 \phi}{\partial p_i \partial p_j}$ depend on the choice of the criterion ϕ . For D -optimality:

$$\begin{aligned}
\frac{\partial \phi}{\partial p_j} &= \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \\
\frac{\partial^2 \phi}{\partial p_i \partial p_j} &= -(\mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j)(\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i).
\end{aligned}$$

For A-optimality:

$$\begin{aligned}\frac{\partial \phi}{\partial p_j} &= \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}^\top \mathbf{a} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \\ \frac{\partial^2 \phi}{\partial p_i \partial p_j} &= -2[\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}^\top \mathbf{a} \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i][\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i].\end{aligned}$$

The derivative forms of different criteria in the unconstrained setting were already derived in Chapters 3, 4 and 5. More details can be found there. Next section we will see simple examples and some empirical results, but before that we summarize the choices of the various terms involved.

6.3.2 Summary of Optimization Terms

$\phi(\mathbf{p})$:

The choices of $\phi(\mathbf{p})$ (the criterion to be maximized) are:

$$\begin{aligned}\phi(\mathbf{p}) &= -\log \det\{\mathbf{M}^{-1}(\mathbf{p})\} && [D\text{-optimality}] \\ \phi(\mathbf{p}) &= -\text{tr}(\mathbf{A}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{A}) && [A\text{-optimality}]\end{aligned}$$

$c(\mathbf{p})$:

We consider:

$$\begin{aligned}c(\mathbf{p}) &= \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b} && [\text{Covariance}] \\ c(\mathbf{p}) &= \frac{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}} \sqrt{\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}} && [\text{Correlation}]\end{aligned}$$

$C(\mathbf{p})$:

We consider $C(\mathbf{p}) = -c^2(\mathbf{p})$. Note that we could also consider $C(\mathbf{p}) = [-c(\mathbf{p})]^{2k}$, where k is a positive integer.

$\phi_L(\mathbf{p})$:

$$\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$$

$$\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$$

$W(\mathbf{p})$: We have considered a variety of choices of $W(\mathbf{p})$. Here is a list of the choices considered in the next section where we report the results:

$$W_1 = -[\mathbf{w}^\top \mathbf{w}]$$

$$W_2 = [R_1 - 1]$$

$$W_3 = -[\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}]$$

$$W_4 = [R_2 - 1]$$

$$W_5 = -[[\mathbf{w}(\mathbf{E})]^\top [\mathbf{w}(\mathbf{E})]]$$

$$W_6 = [R_3 - 1]$$

$$W_7 = -[1 - R_1]^2$$

$$W_8 = -[1 - R_2]^2$$

$$W_9 = -[1 - R_3]^2$$

$$W_{10} = -[\ln(1 - W_5)]^2$$

$$W_{11} = -[\ln(1 - W_3)]^2$$

$$W_{12} = -[1 - \exp(W_5)]^2$$

where the ratios R_1, R_2, R_3 are defined as:

$$\begin{aligned} R_1 &= \frac{\xi_1^2}{\xi_2 \xi_3} \\ R_2 &= \frac{2\xi_1 \xi_3 \xi_4}{\xi_3^2 \xi_5 + \xi_1^2 \xi_6} \\ R_3 &= \frac{\xi_4^2}{\xi_5 \xi_6} \end{aligned}$$

and the ξ_i components are defined as:

$$\begin{aligned} \xi_1 &= [(\mathbf{F}^c)^\top \mathbf{F}^\phi] \\ \xi_2 &= [(\mathbf{F}^\phi)^\top \mathbf{F}^\phi] \\ \xi_3 &= [(\mathbf{F}^c)^\top \mathbf{F}^c] \\ \xi_4 &= [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^\phi] \\ \xi_5 &= [(\mathbf{F}^\phi)^\top \mathbf{P}^2 \mathbf{F}^\phi] \\ \xi_6 &= [(\mathbf{F}^c)^\top \mathbf{P}^2 \mathbf{F}^c]. \end{aligned}$$

Note the maximum value of all choices of W is zero. The reason for using various choices of W is that the convergence of algorithm (6.33) is likely to depend on the curvature of the function W .

$f(z, \delta)$:

The choice of the function $f(z, \delta)$ and its first argument play an important role in the convergence of the algorithm (6.33). We attempt to improve convergence by exploring both of these choices. We use the vertex directional derivatives in the first argument. The vertex directional derivatives are both positive and negative. This requires a function $f(z, \delta)$ which is defined for positive and negative z . Also

from the conditions of the algorithmic, the function $f(z, \delta)$ should be increasing in z . It can be proved that, $\sum p_j F_j = 0$. Thus, a suitable function should be centered on zero and change quickly about 0. With these above points in mind, one potential choice of $f(z, \delta)$ would be $f(z, \delta) = \exp(\delta z) / (1 + \exp(\delta z))$, the logistic c.d.f. Convergence rates depend on the parameter δ in addition to the choice of the function $f(z, \delta)$ and its first argument. Therefore, we will explore the choice of this tuning parameter as well.

6.4 Examples

6.4.1 A simple example on quadratic model

First we illustrate the zero covariance case by a simple example in the context of a quadratic regression model on three points: $-1, 0, 2$.

$$E(y|x) = \theta_1 + \theta_2 x + \theta_3 x^2, \quad -1 \leq x \leq 2.$$

The design is supported at $x = \{-1, 0, 2\}$ with weights $\mathbf{p} = (p_1, p_2, p_3)^\top$. The information matrix $\mathbf{M}(\mathbf{p})$ is: $\mathbf{M}(\mathbf{p}) = \sum_{j=1}^3 p_j \mathbf{v}_j \mathbf{v}_j^\top$.

We focus on the zero covariance constraint $C_0 = 0$, where $c(\mathbf{p}) = (\mathbf{M}^{-1}(\mathbf{p}))_{23}$, so that we want $\text{cov}(\hat{\theta}_2, \hat{\theta}_3) = 0$. The numerator of $c(\mathbf{p})$ is quadratic in the weights: $\det(\mathbf{M}(\mathbf{p})) \cdot c(\mathbf{p}) = (p_2 - 1/2)^2 - 9(p_3 - 1/2)^2 + 2$.

To handle the zero covariance constraint ($C_0 = 0$) and the $\sum_j p_j = 1$ constraint, we substitute the weights to transform the constrained optimization problem into

an unconstrained optimization with respect to one weight (p_2). Considering the nonnegativity constraints on p_1, p_2, p_3 , the only valid substitution formulas are

$$\begin{aligned} p_1 &= \left(\frac{1}{2} - p_2\right) + \frac{1}{3}\sqrt{\left(p_2 - \frac{1}{2}\right)^2 + 2}, \\ p_3 &= \frac{1}{2} - \frac{1}{3}\sqrt{\left(p_2 - \frac{1}{2}\right)^2 + 2}. \end{aligned}$$

These formulas ensure $C(\mathbf{p}) = 0$ is satisfied and are valid for the full range $0 < p_2 < 1$.

Figures 6.1–6.10 display the behavior of the transformed criteria $W_j(p_2)$ under both D -optimality and A -optimality, combined into a single plot for each W_j . Combining the two optimality criteria into one visualization provides a direct comparison of their behavior under identical constraints, a valuable insight that is obscured when plotted separately.

A key challenge in this visualization arises because the raw scales of the underlying criteria (ϕ_D and ϕ_A) differ substantially, which would cause severe visual distortion if plotted directly. To overcome this and enable a rigorous comparison of their relative performance and functional shape, all A -optimality curves have been normalized to the scale of the D -optimality criterion using the transformation. This approach, which rescales one criterion to the range of another for direct comparison, is a variation of the widely used Min-Max scaling method (Han et al., 2012) and is standard practice in optimal design studies (Pukelsheim, 2006).

$$\tilde{A}(\mathbf{p}) = \frac{A(\mathbf{p}) - \min_{\mathbf{p}} A(\mathbf{p})}{\max_{\mathbf{p}} A(\mathbf{p}) - \min_{\mathbf{p}} A(\mathbf{p})} \cdot \left[\max_{\mathbf{p}} D(\mathbf{p}) - \min_{\mathbf{p}} D(\mathbf{p}) \right] + \min_{\mathbf{p}} D(\mathbf{p}). \quad (6.36)$$

This normalization allows for a direct, visually consistent comparison of the two curves within the same plot. Across all W_j , the D -optimality and A -optimality curves exhibit a highly similar functional shape.

All normalized curves attain their standardized maximum value of zero, at $p_2^* \approx 0.4925$. This finding confirms the strong consistency of the constrained optimum across different W -functions and across the two major estimation-oriented optimality criteria. Nevertheless, the figures also reveal important algorithmic features. In particular, several of the $W_j(p_2)$ curves remain extremely flat near their maximum, indicating clear difficulty in satisfying the first-order optimality conditions. In such cases, the functions appear almost non-differentiable on either side of the optimum, which directly explains the slow convergence of the iterative algorithm. For some choices of W , large initial derivatives cause functional values to become temporarily unstable, reinforcing the importance of using standardized derivatives in the algorithm.

It is worth noting that for certain criteria, the location of the maximum is visually difficult to determine, even with the normalized scale, because the curves are nearly flat near the optimum. Without the numerical results, it would be extremely challenging to identify p^* with confidence. This highlights a limitation of graphical inspection: although the true maximum is known numerically, the visual evidence alone may be misleading due to the inherent functional flatness.

We also consider the constraint $\det(\mathbf{M}(\mathbf{p}))C_0 = 2$, which leads to linear constraints on the weights. It is simpler to substitute for p_1 and p_2 in terms of p_3 based on the two possible solution regions:

1. For $\frac{1}{3} \leq p_3 \leq \frac{1}{2}$:

$$\begin{aligned} p_1 &= -2(2p_3 - 1), \\ p_2 &= 3p_3 - 1. \end{aligned}$$

2. For $\frac{1}{2} \leq p_3 \leq \frac{2}{3}$:

$$\begin{aligned} p_1 &= 2p_3 - 1, \\ p_2 &= 2 - 3p_3. \end{aligned}$$

Figures 6.11–6.17 show analogous results for $W_j(p_3)$ under the same combined (D - and A -optimal) setup with standardized scales. Again, the curves attain the common maximum at p_3^* , demonstrating the robustness of the constrained solution across criteria. However, their shapes differ noticeably from the p_2 case. They exhibit two distinct turning points, one in each of the regions $p_3 < 0.5$ and $p_3 > 0.5$, and the functions are more sharply peaked, indicating stronger constraint effects along the p_3 direction. Near the optimum, some curves resemble piecewise linear functions, which makes the numerical derivatives less stable and slows convergence even further.

These graphical patterns highlight both the robustness of the method, in that all criteria lead to the same constrained solution, and the inherent numerical challenges that arise when optimizing under the sharply defined constraints of the p_3 formulation.

6.4.2 Results

6.4.2.1 Results for zero covariance and correlation

We continue with the quadratic model, where the information matrix $M(\mathbf{p})$ is defined on a three-point support. We first derive the D -optimal design subject to the constraints $c(\mathbf{p})$ where:

$$c(\mathbf{p}) = \mathbf{a}^\top M^{-1}(\mathbf{p}) \mathbf{b} \text{ where } \mathbf{a} = (0, 1, 0)^\top, \mathbf{b} = (0, 0, 1)^\top \text{ [Covariance]}$$

$$c(\mathbf{p}) = \frac{\mathbf{a}^\top M^{-1}(\mathbf{p}) \mathbf{b}}{\sqrt{\mathbf{a}^\top M^{-1}(\mathbf{p}) \mathbf{a}} \sqrt{\mathbf{b}^\top M^{-1}(\mathbf{p}) \mathbf{b}}} \\ \text{where } \mathbf{a} = (0, 1, 0)^\top, \mathbf{b} = (0, 0, 1)^\top \text{ [Correlation].}$$

Note that both of these constraints require $\text{cov}(\hat{\theta}_2, \hat{\theta}_3) = 0$ (i.e., the estimators for the linear and quadratic coefficients are uncorrelated). To run the algorithm, we choose the smoothing function $f(z, \delta) = \exp(\delta z) / (1 + \exp(\delta z))$, where $z = \mathbf{F}^{\phi_L}$. The function $\phi_L(\mathbf{p})$ is defined as $\min\{C(\mathbf{p}), W(\mathbf{p})\}$ (as per Equation (6.35)) starting from equal initial weights $\mathbf{p}$.

(i) $c(\mathbf{p}) = \mathbf{a}^\top M^{-1}(\mathbf{p}) \mathbf{b}$ (Covariance constraint) with $W = W_4 = [R_2 - 1]$ and $\delta = 0.05$ from equal initial weights,

$$(C, W) = (-1.777777777778, -0.031928029635), \\ -4.22222222 \leq F_i^C \leq 2.444444444, \\ -303453543431461 \leq F_i^W \leq 466427836802365.$$

After 167 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ 0 &\leq F_i^C \leq 0.00000001, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

Let $\delta = 0.01$. After 630 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ 0 &\leq F_i^C \leq 0, \\ -0.00000003 &\leq F_i^W \leq 0.00000003.\end{aligned}$$

Let $W = W_5 = -[[\mathbf{w}\mathbf{E}]^\top [\mathbf{w}\mathbf{E}]]$ and $\delta = 0.05$. After 75 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000000 &\leq F_i^C \leq 0.00000013, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

(ii) $c(\mathbf{p}) = \frac{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{a}}\sqrt{\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{b}}}$ (Correlation constraint) with $W = W_4 = [R_2 - 1]$ and $\delta = 0.05$ from equal initial weights,

$$\begin{aligned}(C, W) &= (-0.703296703297, -0.031928029635), \\ -0.49559232 &\leq F_i^C \leq 0.28692187, \\ -303453543431459 &\leq F_i^W \leq 466427836802363.\end{aligned}$$

After 190 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

Let $\delta = 0.01$. After 1012 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0, \\ -0.00000002 &\leq F_i^W \leq 0.00000002.\end{aligned}$$

Let $W = W_5 = -[[\mathbf{w}\mathbf{E}]^\top [\mathbf{w}\mathbf{E}]]$ and $\delta = 0.05$. After 422 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000002 &\leq F_i^C \leq 0, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

It is clear that all the values (C and W) and their corresponding directional derivatives (F_i^C and F_i^W) successfully converge to zero across the tested scenarios. For the D -optimal case, all calculated optimal designs converge to $\mathbf{p}^* = (0.472951, 0.498453, 0.028595)$.

We now consider deriving the A -optimal design subject to the constraints $c(\mathbf{p})$.

(i) $c(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p})\mathbf{b}$ (Covariance constraint) with $W = W_4 = [R_2 - 1]$ and

$\delta = 0.05$ from equal initial weights,

$$\begin{aligned}(C, W) &= (-1.777778, -0.018630683125), \\ -4.222222 &\leq F_i^C \leq 2.444444, \\ -0.51584143 \times 10^{-5} &\leq F_i^W \leq 0.62012732.\end{aligned}$$

After 196 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000000 &\leq F_i^C \leq 0.00000001, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

Let $W = W_{10} = -[\ln(1 - W_5)]^2$. After 40754 iterations,

$$\begin{aligned}(C, W) &= (-0.000000000003, -0.000000000007), \\ -0.00012729 &\leq F_i^C \leq 0.00000397, \\ -0.00000010 &\leq F_i^W \leq 0.00000010.\end{aligned}$$

(ii) $c(\mathbf{p}) = \frac{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a}} \sqrt{\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}}}$ (Correlation constraint) with $W = W_4 = [R_2 - 1]$ and $\delta = 0.05$ from equal initial weights,

$$\begin{aligned}(C, W) &= (-0.703296703297, -0.018630683125), \\ -0.49559232 &\leq F_i^C \leq 0.28692187, \\ -0.09104025 &\leq F_i^W \leq 0.15871285.\end{aligned}$$

After 354 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ 0 &\leq F_i^C \leq 0.00000002, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

Let $W = W_5 = -[[\mathbf{w}\mathbf{E}]^\top [\mathbf{w}\mathbf{E}]]$. After 197 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

It is clearly observable that all the values (C and W) and their corresponding directional derivatives (F_i^C and F_i^W) successfully converge to zero in the A -optimal design scenarios as well. All the optimal designs converge to $\mathbf{p}^* = (0.486197, 0.485220, 0.028583)$. The convergence of the D -optimal design, however, achieved slightly better residual values than the A -optimal design.

The second example is one of five studied by Silvey et al. (1978) and Wu (1978) in the context of constructing D -optimal designs on a finite design space. The linear model is defined as $E(y) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 = \mathbf{v}^\top \boldsymbol{\theta}$, where $\mathbf{v} \in \mathcal{V}$. The design space $\mathcal{V}$ is given by:

$$\mathcal{V} = \{\mathbf{v} = (1, x_1, x_2)^\top : (x_1, x_2) \in \mathcal{W}\},$$

where $\mathcal{W}$ is a quadrilateral with the following sets of vertices:

- $V_1: \{(-1, -1), (-1, 1), (1, -1), (2, 3)\}$

- $V_2: \{(-1, -1), (-1, 1), (1, -1), (2, 2)\}$
- $V_3: \{(-1, -2), (-1, 1), (1, -1), (2, 2)\}$

The design points are taken to consist of the four vertices of the quadrilateral, and we seek the optimal weights $\mathbf{p}$ for these four vertices under D -optimal and A -optimal criteria. The constraint $c(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$ is equivalent to requiring $\text{cov}(\hat{\theta}_1, \hat{\theta}_3) = 0$. We use the smoothing function $\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$ as in the first example.

Results for Design Space $V_1: \{(-1, -1), (-1, 1), (1, -1), (2, 3)\}$.

The constraint vectors are fixed at $\mathbf{a} = (1, 0, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$.

D -optimal design

(i) Covariance constraint with $W = W_3 = -[\mathbf{w}^\top \mathbf{E}^2 \mathbf{w}]$ and $\delta = 0.1$ from equal initial weights:

$$\begin{aligned} (C, W) &= (-0.02777778, 0.376001512919), \\ -0.2695473 &\leq F_i^C \leq 0.3518519, \\ -4.45834130 &\leq F_i^W \leq 5.06444430. \end{aligned}$$

After 754 iterations,

$$\begin{aligned} (C, W) &= (0, 0), \\ -0.00000002 &\leq F_i^C \leq 0.00000001, \\ 0 &\leq F_i^W \leq 0.00000001. \end{aligned}$$

(ii) Correlation constraint with $W = W_4 = [R_2 - 1]$ and $\delta = 0.1$ from equal initial

weights:

$$\begin{aligned}(C, W) &= (-0.05084746, -0.9928596), \\ -0.4800664 &\leq F_i^C \leq 0.5745475, \\ -1.01464994 &\leq F_i^W \leq 1.5296130.\end{aligned}$$

After 479 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000002 &\leq F_i^C \leq 0.00000003, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

It is clear that the functions C and W converge to zero, as do their directional derivatives F_i^C and F_i^W . The optimal designs for all scenarios converge to $\mathbf{p}^* = (0.178320, 0.150893, 0.376814, 0.293974)$.

A-optimal design

(i) Covariance constraint with $W = W_3 = -[\mathbf{w}^\top \mathbf{E}^2 \mathbf{w}]$ and $\delta = 0.05$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.02777778, -0.111519399679), \\ -0.2695473 &\leq F_i^C \leq 0.3518519, \\ -5.03256261 &\leq F_i^W \leq 2.16750924.\end{aligned}$$

After 1350 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000001, \\ -0.00000002 &\leq F_i^W \leq 0.00000003.\end{aligned}$$

(ii) Correlation constraint with $W = W_3 = -[\mathbf{w}^\top \mathbf{E}^2 \mathbf{w}]$ and $\delta = 0.05$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.05084746, -0.309888257530), \\ -0.4800664 &\leq F_i^C \leq 0.5745475, \\ -12.64704617 &\leq F_i^W \leq 6.65271222.\end{aligned}$$

After 716 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000001, \\ -0.00000002 &\leq F_i^W \leq 0.00000002.\end{aligned}$$

The optimal designs for the A-optimal case converge to $\mathbf{p}^* = (0.257656, 0.230345, 0.359538, 0.152461)$.

Results for Design Space V_2 : $\{(-1, -1), (-1, 1), (1, -1), (2, 2)\}$.

The constraint vectors are $\mathbf{a} = (1, 0, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$.

D-optimal design

(i) Covariance constraint with $W = W_1 = -[\mathbf{w}^\top \mathbf{w}]$ and $\delta = 0.005$ from equal

initial weights:

$$\begin{aligned}(C, W) &= (-0.011080332410, -5.142561186081), \\ -0.22044030 &\leq F_i^C \leq 0.22277300, \\ -58.69043241 &\leq F_i^W \leq 51.97674682.\end{aligned}$$

After 4831 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000012 &\leq F_i^C \leq 0.00000007, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

(ii) Correlation constraint with $W = W_1 = -[\mathbf{w}^\top \mathbf{w}]$ and $\delta = 0.005$ from equal

initial weights:

$$\begin{aligned}(C, W) &= (-0.014814814815, -8.552631547542), \\ -0.29739369 &\leq F_i^C \leq 0.28641975, \\ -102.02359941 &\leq F_i^W \leq 105.72264907.\end{aligned}$$

After 3954 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000011 &\leq F_i^C \leq 0.00000006, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

The optimal designs for V_2 under D -optimality converge to $\mathbf{p}^* = (0.1948654, 0.1871843, 0.3333333, 0.2846170)$.

A-optimal design

(i) Covariance constraint with $W = W_1 = -[\mathbf{w}^\top \mathbf{w}]$ and $\delta = 0.005$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.011080332410, -5.657959882030), \\ -0.22044030 &\leq F_i^C \leq 0.22277300, \\ -157.60236092 &\leq F_i^W \leq 93.65183378.\end{aligned}$$

After 5515 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000002 &\leq F_i^C \leq 0.00000002, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

(ii) Correlation constraint with $W = W_3 = -[\mathbf{w}^\top \mathbf{E}^2 \mathbf{w}]$ and $\delta = 0.005$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.014814814815, -0.609350100514), \\ -0.29739369 &\leq F_i^C \leq 0.28641975, \\ -17.26903766 &\leq F_i^W \leq 10.02343056.\end{aligned}$$

After 5037 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000001, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

The optimal designs for V_2 under A-optimality converge to $\mathbf{p}^* =$

(0.2371158, 0.2705397, 0.3299319, 0.1624126).

Results for Design Space V_3 : $\{(-1, -2), (-1, 1), (1, -1), (2, 2)\}$.

The constraint vectors are $\mathbf{a} = (1, 0, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$.

D-optimal design

(i) Covariance constraint with $W = W_1 = -[\mathbf{w}^\top \mathbf{w}]$ and $\delta = 0.01$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.006032613818, -0.614661380099), \\ -0.12358073 &\leq F_i^C \leq 0.17512151, \\ -16.04413364 &\leq F_i^W \leq 10.13946296.\end{aligned}$$

After 2954 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000002 &\leq F_i^C \leq 0.00000002, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

(ii) Correlation constraint with $W = W_4 = [R_2 - 1]$ and $\delta = 0.01$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.010973936900, -0.070260927925), \\ -0.22435604 &\leq F_i^C \leq 0.30483158, \\ -1.73968324 &\leq F_i^W \leq 2.08998839.\end{aligned}$$

After 1961 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000002, \\ -0.00000001 &\leq F_i^W \leq 0.00000001.\end{aligned}$$

The optimal designs for V_3 under D -optimality converge to $\mathbf{p}^* = (0.2613907, 0.2771813, 0.1826262, 0.2788018)$.

A -optimal design

(i) Covariance constraint with $W = W_1 = -[\mathbf{w}^\top \mathbf{w}]$ and $\delta = 0.01$ from equal initial weights:

$$\begin{aligned}(C, W) &= (-0.006032613818, -6.425464623777), \\ -0.12358073 &\leq F_i^C \leq 0.17512151, \\ -124.28277033 &\leq F_i^W \leq 97.85894380.\end{aligned}$$

After 3609 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000001, \\ -0.00000008 &\leq F_i^W \leq 0.00000007.\end{aligned}$$

(ii) Correlation constraint with $W = W_3 = -[\mathbf{w}^\top \mathbf{P}^2 \mathbf{w}]$ and $\delta = 0.01$ from equal

initial weights:

$$\begin{aligned}(C, W) &= (-0.010973936900, -1.275505215155), \\ -0.22435604 &\leq F_i^C \leq 0.30483158, \\ -25.95512327 &\leq F_i^W \leq 14.53719193.\end{aligned}$$

After 2259 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00000001 &\leq F_i^C \leq 0.00000001, \\ -0.00000003 &\leq F_i^W \leq 0.00000003.\end{aligned}$$

The optimal designs for V_3 under A-optimality converge to $\mathbf{p}^* = (0.2547348, 0.3546947, 0.2148403, 0.1757303)$.

6.4.2.2 Transition from zero to non-zero covariance and correlation

We see that, in all the above examples, the covariance and correlation constraint achieves the value zero at the optimum. That is, $C(\mathbf{p}^*) = -(c(\mathbf{p}^*))^2 = 0$. This naturally leads to the question: what if $c(\mathbf{p}^*)$ is not equal to zero? In practice, certain models may not have an optimal design that allows for zero correlation and covariance among parameter estimators. In such a case, instead of attaining a perfect solution, we aim for a pre-specified constant. Suppose that the minimal correlation and covariance between the parameter estimators can be determined as c_0 . This motivates the idea that we can optimize the D - or A -optimality criterion subject to the constraint $c(\mathbf{p}) - c_0 = 0$. Therefore, along with $W(\mathbf{p})$,

we optimize $C(\mathbf{p})$ in the following form

$$C(\mathbf{p}) = -[c(\mathbf{p}) - c_0]^2.$$

This makes an important transition. While the previous examples provide zero-covariance and correlation solutions, the following chemical model illustrates situations where approximate solutions are attainable.

The model was considered by (Torsney and Alahmadi, 1995; Mandal et al., 2017). It describes the relationship between the viscosity y and the concentration x of a chemical solution:

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2$$

with the design interval restricted to $(0, 0.2]$. The objective was to minimize the covariance or correlation between $\hat{\theta}_3$ and each of $\hat{\theta}_1$ and $\hat{\theta}_2$ as much as possible. In this example, we cannot obtain a design that produces a common value of zero covariance and correlation. First we are looking for two distinct designs, one to minimize $\text{cov}(\hat{\theta}_1, \hat{\theta}_3)$ and one to minimize $\text{cov}(\hat{\theta}_2, \hat{\theta}_3)$ and then similarly one to minimize $\text{corr}(\hat{\theta}_1, \hat{\theta}_3)$ and one to minimize $\text{corr}(\hat{\theta}_2, \hat{\theta}_3)$. We minimized the two covariance or correlations individually using the algorithm (6.33).

In the first covariance case we take $\mathbf{a} = (1, 0, 0)$ and $\mathbf{b} = (0, 0, 1)$. We consider the D -optimal design, where the constraint is equivalent to $\text{cov}(\hat{\theta}_1, \hat{\theta}_3) - c_0$. Let the target value for the covariance be $c_0 = -38565.6$. We define the penalty function for the constraint as $C(\mathbf{p}) = -[c(\mathbf{p}) + 38565.6]^2$, where $c(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$. In order to run the algorithm, we choose the smoothing function

$f(z, \delta) = \exp(\delta z) / (1 + \exp(\delta z))$ where $\delta = 0.01$ and $z = F^{\phi_L}$. We start with $\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$ (6.35) and $W = W_7$ at equal initial weights p .

Initial weights results:

$$\begin{aligned}(C, W) &= (-20715420, -0.002355114034), \\ -0.76257118 &\leq F_i^C \leq 0.44279721, \\ -2.5586559 \times 10^{36} &\leq F_i^W \leq 1.8641577 \times 10^{36}.\end{aligned}$$

After 1043 iterations with $\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$,

$$\begin{aligned}(C, W) &= (-0.000018342388, -0.000000013481), \\ -0.00007332 &\leq F_i^C \leq 0.00001600, \\ -0.21839289 &\leq F_i^W \leq 0.27342588.\end{aligned}$$

As we can see, the results for C and W are both getting close to zero, but the directional derivative of W is not close enough to zero. To obtain a better result, we tried the sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$ (6.34) and $W = W_8$. Other parameters remained the same.

After 1496 iterations,

$$\begin{aligned}(C, W) &= (-0.000018250402, 0), \\ -0.00000100 &\leq F_i^C \leq 0.00000023, \\ 0 &\leq F_i^W \leq 0.\end{aligned}$$

We also tried A -optimality with the sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$. We used the same setting as above.

Initial weights results for A -optimality:

$$\begin{aligned}(C, W) &= (-20715420, -0.010534893967), \\ -0.76257118 &\leq F_i^C \leq 0.44279721, \\ -0.64399844 &\leq F_i^W \leq 0.55516981.\end{aligned}$$

After 1670 iterations,

$$\begin{aligned}(C, W) &= (-0.000018250381, 0), \\ -0.00000005 &\leq F_i^C \leq 0.00000010, \\ 0 &\leq F_i^W \leq 0.\end{aligned}$$

The results show that for this case where a perfect zero-covariance solution does not exist, the sum function provides a more stable and effective algorithm. The directional derivatives for the sum function are much closer to zero, indicating better convergence.

In the second covariance case we take $\mathbf{a} = (0, 1, 0)$ and $\mathbf{b} = (0, 0, 1)$. We consider the D -optimal design with the constraint equivalent to $\text{Cov}(\hat{\theta}_2, \hat{\theta}_3) - c_0$. The value of c_0 is 6909.345, and the constraint penalty is $C(\mathbf{p}) = -[c(\mathbf{p}) - 6909.345]^2$. We initially use $\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$ (6.35), $W = W_8$, $\delta = 0.01$, and equal initial weights p .

Initial values are

$$\begin{aligned}(C, W) &= (-1914154, -0.1432652), \\ -0.8438959 &\leq F_i^C \leq 0.9420006, \\ -1.901875818 \times 10^{14} &\leq F_i^W \leq 1.473798307 \times 10^{14}.\end{aligned}$$

After 598 iterations,

$$\begin{aligned}(C, W) &= (-0.000870600773, -0.000201587316), \\ -0.00993072 &\leq F_i^C \leq 0.00218185, \\ -2.79133182 &\leq F_i^W \leq 4.55422469.\end{aligned}$$

As we can see, the results for C and W are both getting close to zero, but the directional derivatives of W are not close enough to zero. To make a better result, we tried the sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$ (6.34) with $W = W_8$. Other parameters remained the same.

After 940 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00009399 &\leq F_i^C \leq 0.00035693, \\ -0.00000001 &\leq F_i^W \leq 0.00000002.\end{aligned}$$

We also tried A-optimality with the sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$. We used the same setting as above.

Initial weights results for A-optimality:

$$\begin{aligned}(C, W) &= (-1914154, -0.4348073), \\ -0.84389594 &\leq F_i^C \leq 0.94200061, \\ -6.46927859 &\leq F_i^W \leq 4.46653022.\end{aligned}$$

After 1400 iterations,

$$\begin{aligned}(C, W) &= (0, 0), \\ -0.00013802 &\leq F_i^C \leq 0.00027205, \\ 0 &\leq F_i^W \leq 0.\end{aligned}$$

The final results show that the first case optimal designs approximately converge to weights $(0.423, 0.405, 0.172)$, and the second case optimal designs converge to $(0.508, 0.347, 0.145)$. These results are consistent under both D - and A -optimality. Overall, we find that using the sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$ gives better results than the minimum function $\phi_L(\mathbf{p}) = \min\{C(\mathbf{p}), W(\mathbf{p})\}$, and D -optimality tends to perform slightly better than A -optimality.

The choice between a minimum and a sum function for the compound optimality criterion $\phi_L(\mathbf{p})$ depends on whether the target value is exactly attainable. The sum function $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$ combines the two objectives into a single, smooth, continuously differentiable function. This allows the optimization algorithm to move reliably through the search space. As demonstrated in our results, the sum function consistently yields more stable convergence and superior designs, as evidenced by final directional derivatives that are closer to zero.

Following the covariance analysis, it is natural to extend our attention to the correlation coefficient. While covariance indicates the direction of a linear relationship, it is scale-dependent and difficult to compare across models or datasets. The correlation coefficient, by contrast, is a standardized measure between -1 and 1 , providing a more interpretable assessment of the strength and

direction of association. Similar to the covariance case, we are looking for two distinct designs, one to minimize $\text{corr}(\hat{\theta}_1, \hat{\theta}_3)$ and one to minimize $\text{corr}(\hat{\theta}_2, \hat{\theta}_3)$. We found the minimum values -0.8965431 and 0.7410294 respectively.

In the first correlation case, our target constant c_0 is -0.8965431 . We now consider the D -optimal design with the constraint $c(\mathbf{p}) - c_0 = 0$, i.e., $c(\mathbf{p}) + 0.8965431 = 0$. Let $C(\mathbf{p}) = [c(\mathbf{p}) + 0.8965431]^2$ where $c(\mathbf{p}) = (\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b})(\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a})^{-1/2}(\mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b})^{-1/2}$ with $\mathbf{a} = (1, 0, 0)$, $\mathbf{b} = (0, 0, 1)$.

To run the algorithm, we choose the smoothing function $f(z, \delta) = \exp(\delta z)/(1 + \exp(\delta z))$ with $\delta = 0.5$ and $z = F^{\phi_L}$. We use the compound criterion $\phi_L(\mathbf{p}) = C(\mathbf{p}) + W(\mathbf{p})$ (6.34) with W functions related to D -optimality. After 100000 iterations for D -optimality, we obtain the following results for different choices of the W function.

For $W = W_5$:

$$\begin{aligned} (C, W) &= (-0.000000301023, 0), \\ -0.00003642 &\leq F_i^C \leq 0.00000064, \\ -0.00000003 &\leq F_i^W \leq 0.00000018. \end{aligned}$$

For $W = W_{10}$:

$$\begin{aligned} (C, W) &= (-0.000000301634, 0), \\ -0.00003647 &\leq F_i^C \leq 0.00000065, \\ -0.00000000 &\leq F_i^W \leq 0.00000000. \end{aligned}$$

We also tried A -optimality with $\delta = 0.09$ and $W = W_{11}$. Other parameters

remained the same.

Initial state for A-optimality,

$$\begin{aligned}(C, W) &= (-0.003992227752, -5.368921711480), \\ -0.00406779 &\leq F_i^C \leq 0.00237441, \\ -1461.69864830 &\leq F_i^W \leq 1065.99708784.\end{aligned}$$

After 65167 iterations,

$$\begin{aligned}(C, W) &= (-0.004266054535, -0.000000006012), \\ -0.00335982 &\leq F_i^C \leq 0.00542318, \\ -0.00551506 &\leq F_i^W \leq 0.00331470.\end{aligned}$$

In the second correlation case we take $\mathbf{a} = (0, 1, 0)$ and $\mathbf{b} = (0, 0, 1)$. We consider the D -optimal design with the constraint equivalent to $\text{corr}(\hat{\theta}_2, \hat{\theta}_3) - c_0$. Our target constant c_0 is 0.74102941. We use the same algorithm parameters as in the first case. After 100000 iterations for D -optimality, we obtain the following results for different choices of the W function.

For $W = W_5$:

$$\begin{aligned}(C, W) &= (-0.000000283185, 0), \\ -0.00003689 &\leq F_i^C \leq 0.00000062, \\ -0.00000001 &\leq F_i^W \leq 0.00000129.\end{aligned}$$

For $W = W_{10}$:

$$\begin{aligned}(C, W) &= (-0.000000265508, 0), \\ -0.00003510 &\leq F_i^C \leq 0.00000058, \\ -0.00000000 &\leq F_i^W \leq 0.00000000.\end{aligned}$$

We also tried A -optimality with $\delta = 0.009$ and $W = W_{11}$. Other parameters remained the same.

Initial state for A -optimality,

$$\begin{aligned}(C, W) &= (-0.005697564300, -141.300554178879), \\ -0.00935003 &\leq F_i^C \leq 0.01246679, \\ -990.75781692 &\leq F_i^W \leq 1128.97939039.\end{aligned}$$

After 53063 iterations,

$$\begin{aligned}(C, W) &= (-0.005117461845, -0.000001523804), \\ -0.01769535 &\leq F_i^C \leq 0.01764363, \\ -0.00943790 &\leq F_i^W \leq 0.01785584.\end{aligned}$$

The optimal designs for the first case approximately converge to weights $(0.016, 0.982, 0.002)$ under D -optimality, and to $(0.3954, 0.2623, 0.3423)$ under A -optimality. The optimal designs for the second case approximately converge to $(0.015, 0.983, 0.002)$ under D -optimality, and to $(0.319, 0.433, 0.248)$ under A -optimality. Note that in the last case, the directional derivatives, as well as the functions C and W are close to zero; however, the results from the previous case are closer to zero.

In summary, the correlation cases show that the algorithm is effective in significantly reducing the initial non-zero correlation values to a magnitude very close to zero. The directional derivatives for both correlation cases demonstrate excellent convergence, with values of W reaching values very close to zero.

A final point to consider is the observation that the initial value of the constraint penalty C is much lower in the correlation example than in the covariance example. This is not a flaw in the algorithm but a direct consequence of the mathematical properties of the two metrics. The covariance is an unbounded value, meaning it can take on extremely large positive or negative magnitudes depending on the scale of the variables. This leads to an objective function value C , which is defined as the squared difference from the target, that can also be very large initially. The correlation, however, is a standardized measure that is inherently bounded between -1 and 1 . Because its values always fall within this finite range, the initial value of the objective function $C(\mathbf{p}) = -[c(\mathbf{p}) - c_0]^2$ is also much smaller and already closer to the optimal value of zero. This illustrates a fundamental difference in the nature of these two optimization problems and highlights the effectiveness of the algorithm even when starting from vastly different scales.

Recent research has also broadened the scope of constrained and compound optimal design problems beyond the classical linear regression setting considered in this thesis. For instance, Duarte et al. (2025) developed compound optimal designs via semidefinite programming, offering a flexible framework for balancing multiple optimality criteria simultaneously. In a broader survey, Huan et al. (2024) provided a synthesis of optimal experimental design formula-

tions and computational strategies, highlighting both Bayesian and frequentist approaches. From a practical perspective, Pauphilet (2025) demonstrated how discrete optimal design can be integrated with operational routing constraints, extending the applicability of these methods to real-world decision-making contexts. Moreover, Jiang and Yue (2021) investigated D -optimal designs in linear mixed models for longitudinal data, further illustrating the relevance of optimal design methodology in complex applied settings.

Together, these contributions show that the methods developed in this thesis connect naturally with an active and expanding body of work on modern extensions of constrained and compound optimal designs. They also suggest that the ideas presented here may serve as a basis for further theoretical developments and broader applications in future research.

In summary, this chapter contributes to the theory of optimal experimental design by extending correlation-based optimality criteria to settings where explicit constraints must be satisfied. While earlier chapters develop methods for minimizing variances, covariances, and correlations in an unconstrained setting, this chapter addresses a more realistic and practically relevant scenario in which designs are required to satisfy additional structural or scientific constraints. In many applications, such as chemistry or clinical studies, it is not sufficient to optimize a single criterion alone; instead, the design must simultaneously achieve desirable statistical properties, such as low correlation between parameter estimators, while also satisfying feasibility or structural restrictions.

More specifically, this chapter develops a unified methodology for constrained optimal design under correlation-based criteria. In particular, the chapter

formulates the problem of minimizing a standard optimality criterion subject to constraints such as minimizing correlation between parameter estimators using a Lagrangian approach. By incorporating the constraint into the objective through the Lagrangian function, the problem is transformed into one that can be solved using directional derivatives and optimality conditions.

A key methodological advancement is the extension of simultaneous optimization techniques to the constrained setting. Since the standard optimality criterion and the constraint based on the correlation criterion must both be satisfied, the problem is reformulated as a compound optimization problem in which multiple criteria are optimized together. This includes approaches based on maximizing the sum of criterion and maximizing the maximin criterion. These methods provide a systematic way to resolve conflicts between competing objectives.

Furthermore, the chapter adapts multiplicative algorithms to handle constrained optimization problems. This involves modifying the update rules and making use of directional derivatives of the Lagrangian to ensure convergence while maintaining feasibility. Additional stabilization strategies are also developed to address the numerical challenges arising from non-concavity and the presence of constraints.

Through illustrative examples, we construct the constrained optimal designs in models including one in chemical studies involving viscosity and concentration of a chemical solution, where minimal correlation between parameter estimators is essential.

Overall, Chapter 6 advances optimal design methodology by integrating stan-

dard optimality criteria and correlation-based criteria with constraint handling, thereby providing a more flexible and practically applicable methodology. This work bridges the gap between theoretical optimality and real-world experimental requirements, enabling the construction of designs that are both statistically efficient and feasible under practical constraints. In some practical applications such as the physical and biomedical sciences, highly correlated estimators complicate inference and interpretation, motivating the need for designs that simultaneously optimize efficiency and enforce independence among parameter estimators.

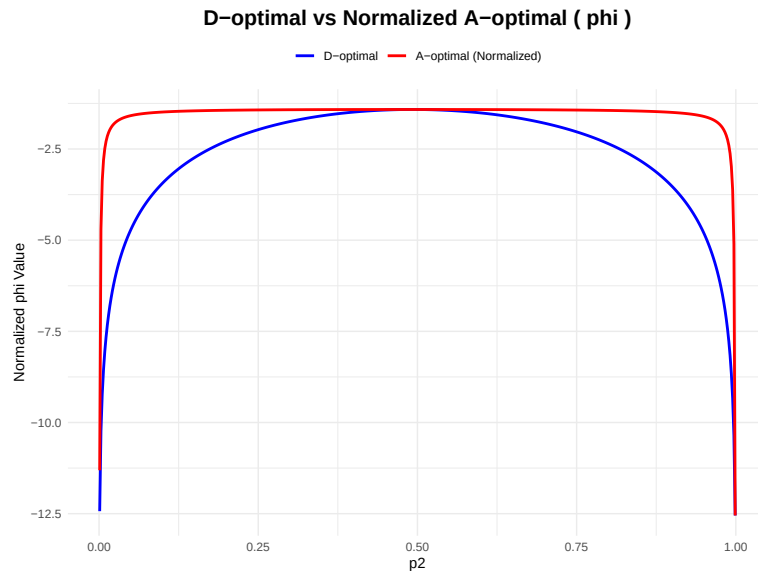


Figure 6.1: $\phi(p_2)$ versus p_2 .

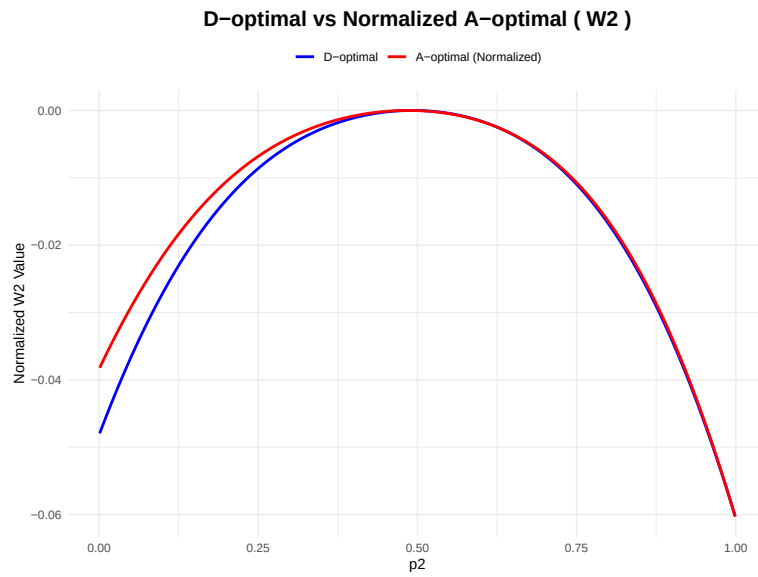


Figure 6.2: $W_2(p_2)$ versus p_2 .

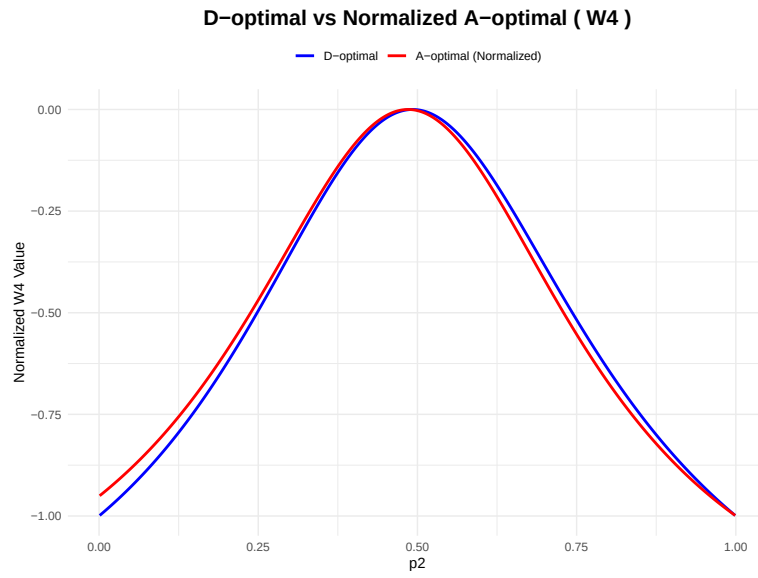


Figure 6.3: $W_4(p_2)$ versus p_2 .

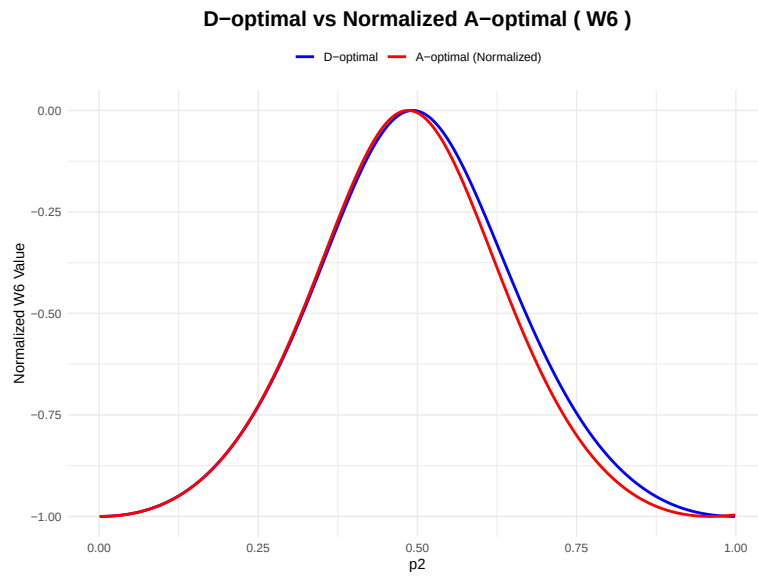


Figure 6.4: $W_6(p_2)$ versus p_2 .

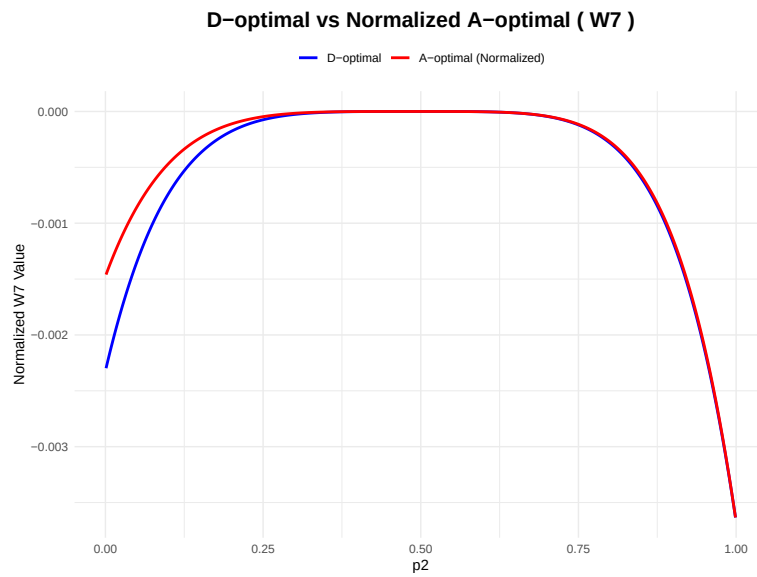


Figure 6.5: $W_7(p_2)$ versus p_2 .

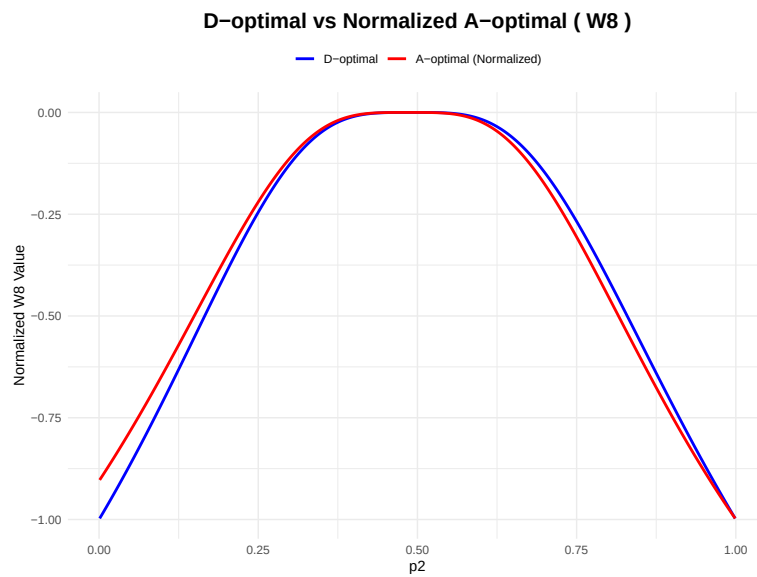


Figure 6.6: $W_8(p_2)$ versus p_2 .

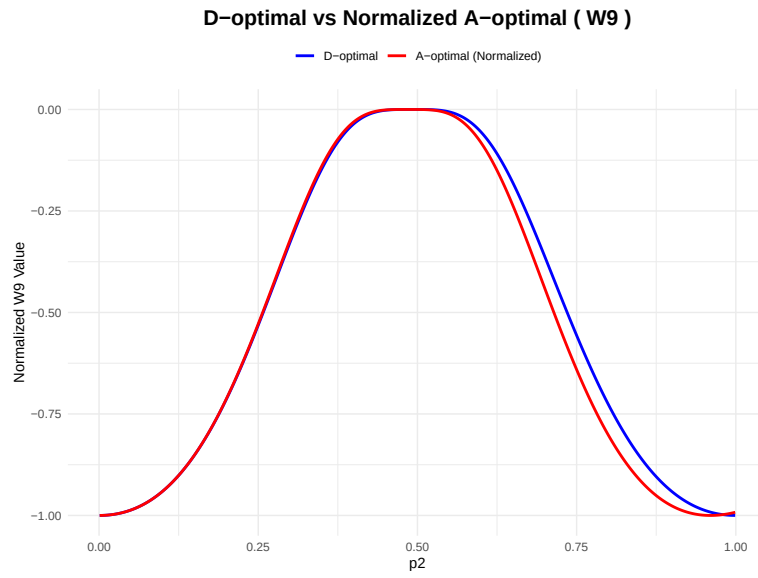


Figure 6.7: $W_9(p_2)$ versus p_2 .

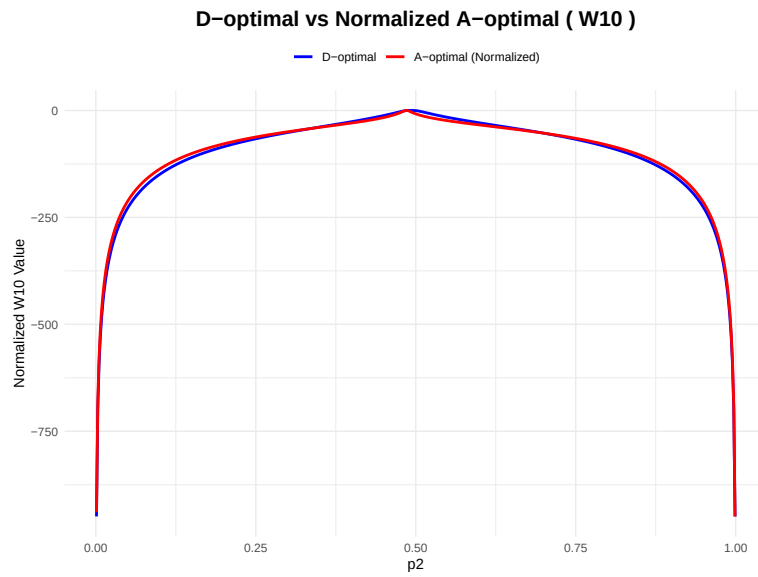


Figure 6.8: $W_{10}(p_2)$ versus p_2 .

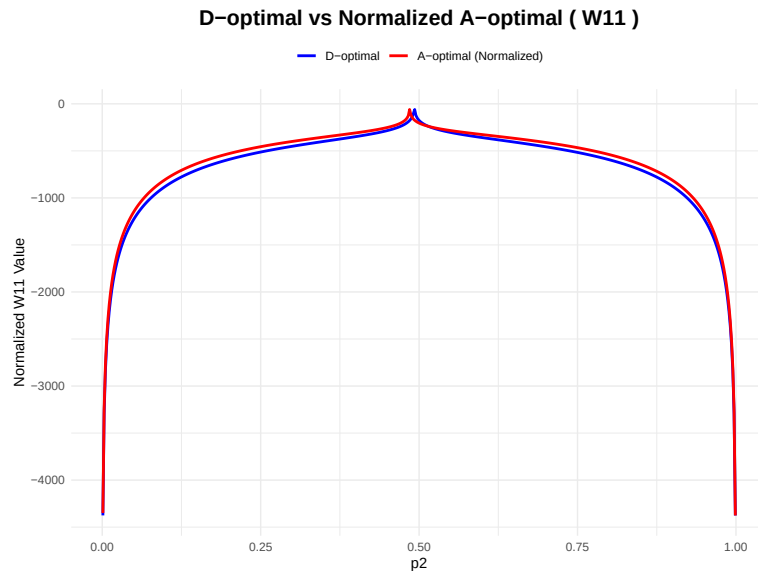


Figure 6.9: $W_{11}(p_2)$ versus p_2 .

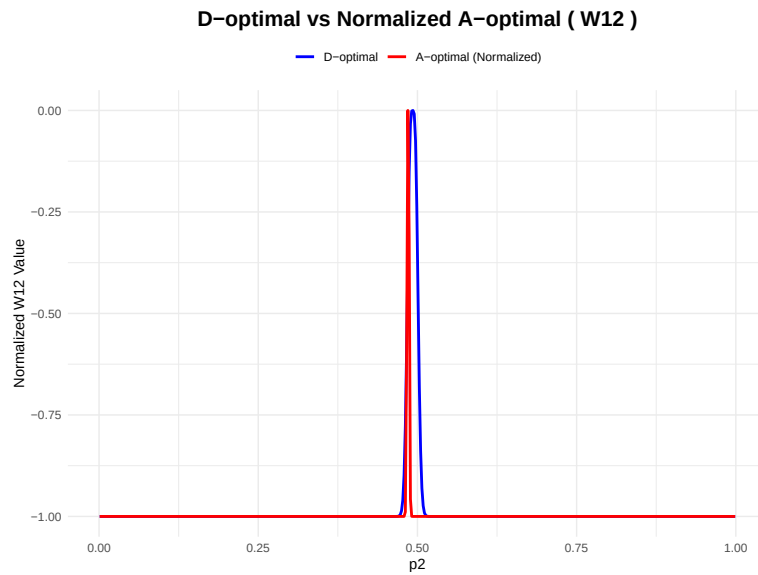


Figure 6.10: $W_{12}(p_2)$ versus p_2 .

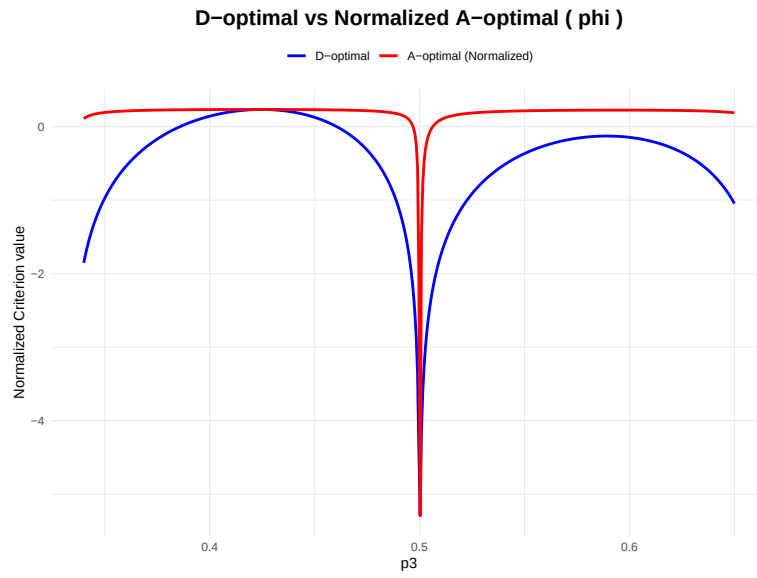


Figure 6.11: $\phi(p_3)$ versus p_3 .

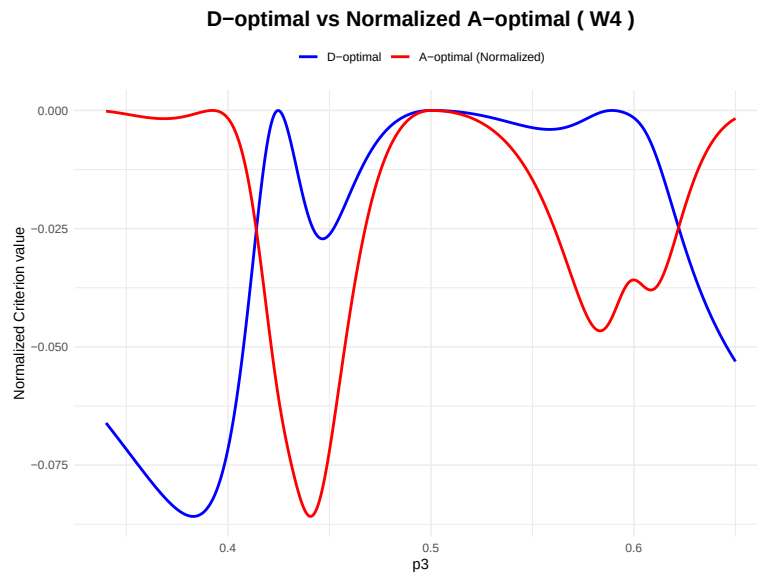


Figure 6.12: $W_4(p_3)$ versus p_3 .

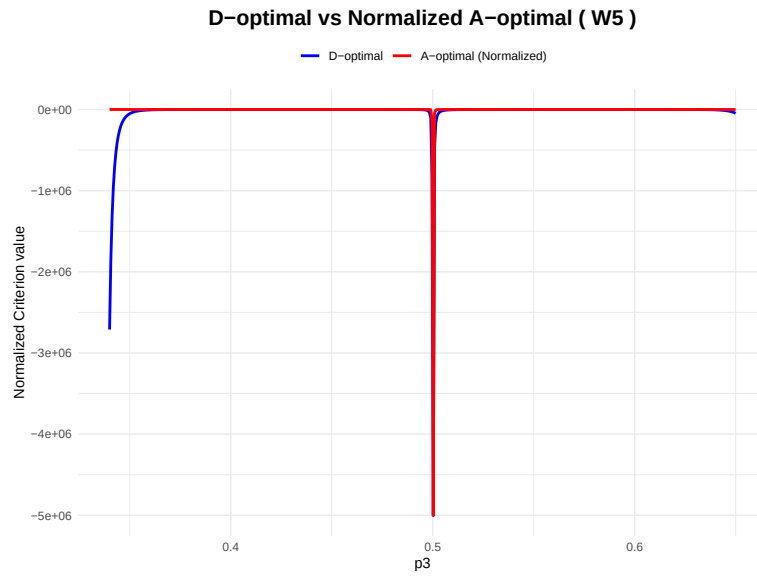


Figure 6.13: $W_5(p_3)$ versus p_3 .

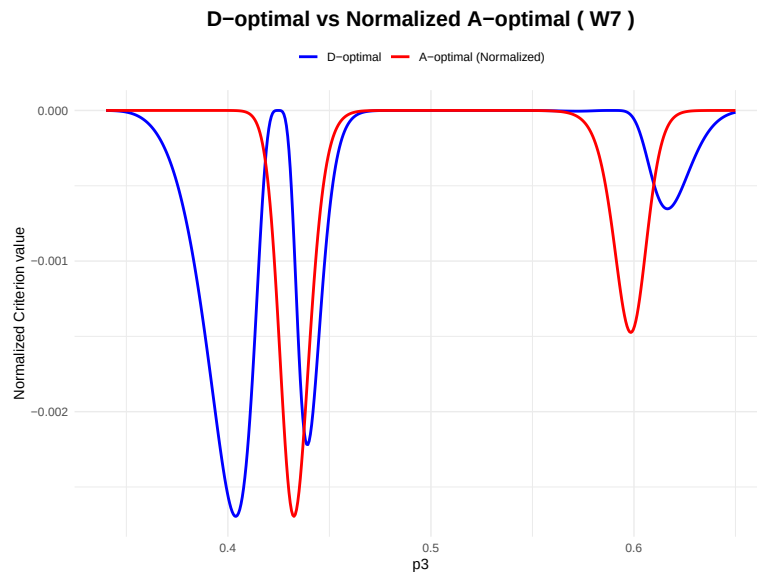


Figure 6.14: $W_7(p_3)$ versus p_3 .

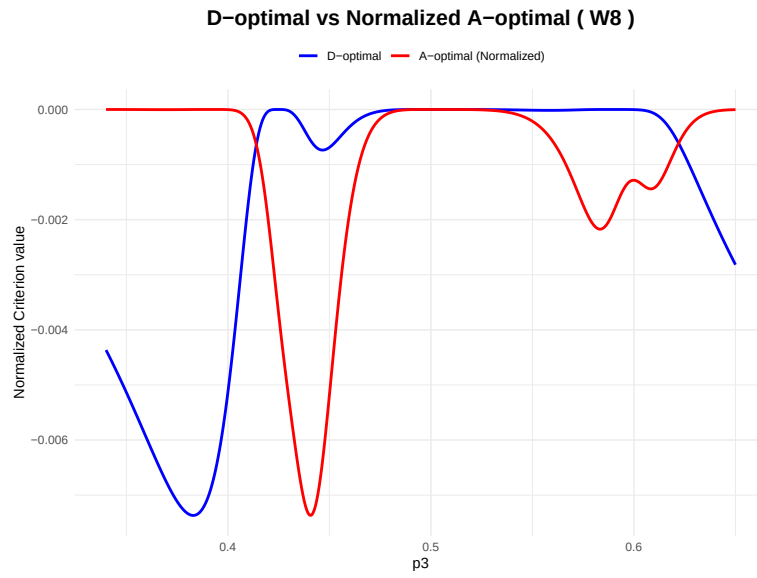


Figure 6.15: $W_8(p_3)$ versus p_3 .

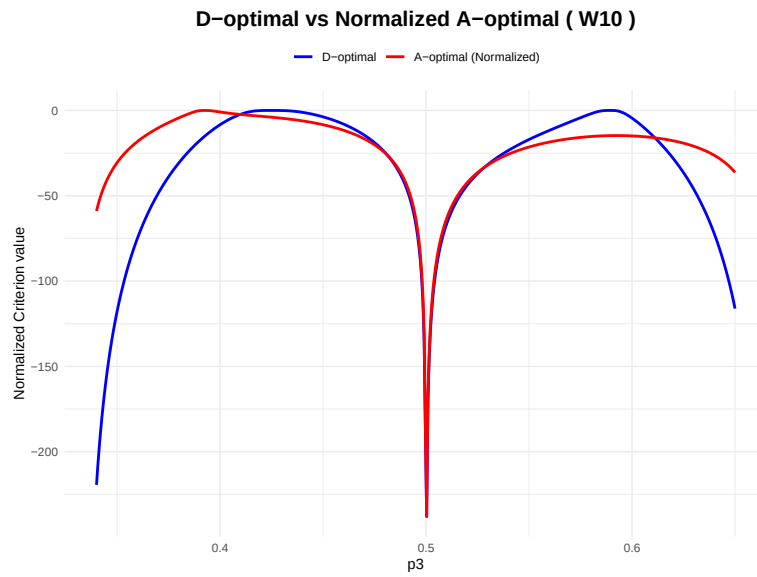


Figure 6.16: $W_{10}(p_3)$ versus p_3 .

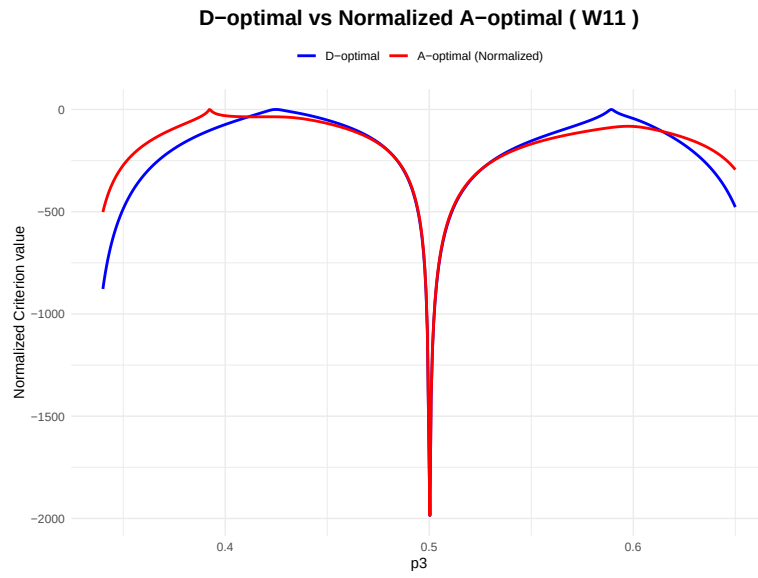


Figure 6.17: $W_{11}(p_3)$ versus p_3 .

Chapter 7

Conclusion

This dissertation has investigated the theory and construction of optimal experimental designs, with a focus on extending beyond classical optimality criteria to incorporate variance-, covariance-, and correlation-based objectives. The study began with a comprehensive review of the foundations of optimal design theory, including the role of the information matrix, equivalence theorems, and directional derivatives. Standard optimality criteria such as D -, A -, G -, D_A -, L -, and c -optimality were revisited in order to set the stage for the development of new non-standard criteria.

A first major contribution of this work is the development and construction of optimal designs that minimize variances of estimators, covariances between estimators, and correlations among estimators. These criteria allow for a more balanced allocation of uncertainty across parameters, addressing situations in which fairness or comparability of estimation precision is as important as overall efficiency. We demonstrated that such criteria can be expressed in terms of

the information matrix, and that their associated optimality conditions can be characterized through directional and Gâteaux derivatives as well as Hessian-based second-order conditions. This framework provides both necessary and sufficient conditions for optimality in many cases, thereby enriching the calculus of optimization for design problems.

A second major contribution is the construction of designs under constraints: balancing efficiency and independence among parameter estimators. Real experiments are rarely conducted in unconstrained settings; ethical considerations, limited resources, and operational restrictions often impose conditions on the feasible design space. To address this, we formulated a general constrained optimization problem. First-order Lagrangian conditions were derived, and iterative multiplicative algorithms were adapted and refined to respect feasibility while still improving efficiency. Numerical studies demonstrated that constrained optimal designs can achieve near-optimal performance relative to their unconstrained counterparts, while also satisfying real-world restrictions. This broadens the practical relevance of optimal design theory and makes it more suitable for applied experimental sciences.

The dissertation also provided a constructive component: both analytical solutions (where available, such as for simple polynomial regression models) and numerical algorithms (for more complex or constrained settings) were implemented. The algorithms were shown to converge under general conditions, and they were validated through empirical examples. Applications to polynomial regression, chemistry-inspired regression model, and a radiation-dosage study for breast cancer patients highlighted the value of the proposed framework. In

particular, the examples demonstrated how minimizing variance or correlation criteria can yield designs with superior balance among parameter estimators compared to standard D - or A -optimal designs.

The contributions of this dissertation can be summarized as follows:

1. Extension of optimal design criteria to variance-, covariance-, and correlation-minimizing objectives.
2. Derivation of rigorous optimality conditions using directional derivatives and Hessian-based second-order conditions, thereby extending the theory of optimal design.
3. Development of iterative algorithms for constructing both unconstrained and constrained designs, ensuring computational feasibility and convergence.
4. Applications in real and simulated contexts, demonstrating that the proposed designs can outperform classical optimal designs in terms of balancing efficiency and interpretability.

Future Work

Despite these advances, some limitations remain, which in turn suggest promising avenues for future research. The study was restricted to linear models, where the covariance structure of the estimators is fully determined by the information matrix. Extending the methodology to other types of models, such as nonlinear

and generalized linear models remains an important area for further research (Paul and Saha, 2007; Goos and Jones, 2011; Yi and Yuan, 2013; Fedorov and Leonov, 2014; Sinha and Xu, 2011; Wong et al., 2019; Xu and Sinha, 2021; Seurat et al., 2021; Shen et al., 2025). In particular, Paul and Saha (2007) provides a clear overview of generalized linear models and their extensions, in the context of biological and environmental applications. The methodology of estimating parameters independently of others in this thesis will be highly relevant to such biological and environmental applications. One possibility is to work on the polynomial models that are useful for computer experiments though Gaussian process models (Lin et al., 2010). Another possibility is to work on minimax A- and D-optimal wavelet designs (Oyet, 2002), and D-optimal minimax regression designs on discrete design space (Zhou, 2008). We can use the Lagrangian approach as wavelet design uses a combination of bias and variance-based optimality criteria. We can also apply the Lagrangian approach for model identification and discrimination purposes (Ghosh and Tian, 2006; Bingham and Chipman, 2007; López-Fidalgo, 2023). One limitation is computational: while the proposed algorithms are efficient for low- and moderate-dimensional problems, high-dimensional or large-scale experimental settings may require more advanced optimization techniques, including semidefinite programming and convex relaxation approaches recently explored in compound design settings (Duarte et al., 2025; Pauphilet, 2025). Finally, the works developed here are largely frequentist; integrating Bayesian perspectives and prior information into variance- and correlation-based criteria offers a promising direction for further research. Recent advances in Bayesian OED include scalable neural-operator-

based methods for high-dimensional problems (Go and Chen, 2025) and the use of Wasserstein-distance information criteria to stabilize inference in Gaussian settings (Helin et al., 2025). These developments suggest fruitful pathways for adapting the ideas of this dissertation to modern Bayesian contexts.

There is also significant scope for applying these methods in more complex applied settings. Recent work has extended optimal design methodology to mixed models and longitudinal data analysis, where correlation structures are more intricate (Jiang and Yue, 2021). Such extensions would allow variance- and correlation-minimizing design principles to be deployed in medical, social, and engineering contexts where repeated measures or hierarchical data are common. Incorporating these ideas could make the theoretical contributions of this dissertation even more broadly applicable.

Concluding Remarks

In conclusion, this dissertation has improved both the theoretical and practical aspects of optimal experimental design for linear models. By broadening the set of criteria beyond classical measures of efficiency, introducing methodologies for constrained optimal designs, and validating the work with illustrative applications, this work provides new tools for designing statistically efficient and practically feasible experiments. The results are expected to be of value not only to statisticians but also to researchers in applied fields where experimentation under constraints is needed. At the same time, the future directions outlined above connect this work to an extended research on compound criteria,

nonlinear models, and Bayesian approaches, ensuring that the ideas developed here can serve as a basis for further theoretical and applied research in optimal experimental design.

Contributions

The following two papers are under review in two journals. Both papers are original contributions in the context of independent estimation of parameters in regression models.

The first contribution, optimal designs to estimate parameters with minimal correlation (Bae and Mandal, 2025a), addresses the direct minimization of correlation between estimators, a critical requirement for valid interpretation in fields such as chemistry and pharmacy. This research tackles the mathematical challenge posed by the correlation criterion, which is neither convex nor concave. By employing compound optimization techniques, the study constructs optimal designs that achieve zero correlation or, where that is mathematically impossible, identifies the minimum achievable correlation to ensure parameter estimators remain as uncorrelated as possible (or as independent as possible under a Gaussian model).

The second contribution, optimal designs under statistical constraints: balancing efficiency and independence among parameter estimators (Bae and Mandal, 2025b), extends the constrained optimal design by incorporating independence as an explicit constraint within classical design theory. Recognizing that traditional D - or A -optimality focuses on information maximization or variance minimization but often ignores correlation, this work uses Lagrangian theory to solve a constrained optimization problem. This methodology allows for the maximization of compound criteria, successfully balancing the trade-off between high statistical efficiency and the enforced independence of parameter estimators.

We explore the proposed methodologies and construct the resulting optimal designs through illustrative examples, including applications in chemical studies.

The abstracts of the papers are given below.

Bae, G. H. and S. Mandal (2025a). Optimal designs to estimate parameters with minimal correlation. *Statistics and Computing*. Under review.

Abstract

In a scientific study, an experiment is designed to answer a variety of questions that are of interest to the researchers. One such interest arises in the fields of chemistry or pharmacy where there is a need to make some parameter estimators uncorrelated with each other in regression models. Owing to designing such experiments, we construct optimal designs to minimize the correlation between the estimators of parameters or parametric functions of a linear model. The optimization problem is quite challenging as the correlation criterion is neither convex nor concave. We first consider the case where we obtain the zero correlations for the parameters of interest. When the zero correlations are not possible to obtain, we find the minimum value so that the parameter estimators are as uncorrelated as possible. Sometimes it is desired to minimize two or more correlations at the same time. We accomplish this objective by considering a compound optimization problem and solving it by means of simultaneous optimization techniques. We explore the methodologies through some regression models of interest.

Bae, G. H. and S. Mandal (2025b). Optimal designs under statistical constraints: Balancing efficiency and independence among parameter estimators. *Journal of the Korean Statistical Society*. Under review.

Abstract

This paper develops a methodology for constructing optimal experimental designs under explicit statistical constraints on parameter estimators. Classical optimal design theory, such as D - or A -optimality, focuses on maximizing information or minimizing average variance, but it often fails when additional requirements, such as reducing correlations between parameter estimators, must be satisfied. In some practical applications, particularly in the physical and biomedical sciences, highly correlated estimators complicate inference and interpretation, motivating the need for designs that simultaneously optimize efficiency and enforce independence. We address this challenge by applying Lagrangian theory with simultaneous optimization techniques. Specifically, we maximize compound criteria formed by combining a classical optimality criterion (D - or A -optimality) with correlation constraints. Through illustrative examples, we construct the constrained optimal designs in models including one in chemical studies involving viscosity and concentration of a chemical solution, where minimal correlation between parameter estimators is essential. This work extends the scope of optimal design theory by providing a systematic approach for incorporating specific constraints into the optimization process.

Bibliography

Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum experimental designs, with SAS*. Oxford, UK: Oxford University Press. 3, 15

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Annals of Mathematical Statistics* 40(5), 1570–1602. 54

Bae, G. H. and S. Mandal (2025a). Optimal designs to estimate parameters with minimal correlation. *Statistics and Computing*. Under review. 180

Bae, G. H. and S. Mandal (2025b). Optimal designs under statistical constraints: Balancing efficiency and independence among parameter estimators. *Journal of the Korean Statistical Society*. Under review. 180

Berger, M. and W. Wong (2009). *An introduction to Optimal Designs for Social and Biomedical Research*, Volume 83. John Wiley & Sons. 3, 16, 34, 36, 45, 115

Bingham, D. R. and H. A. Chipman (2007). Incorporating prior information in optimal design for model selection. *Technometrics* 49(2), 155–163. 177

- Box, G. E. and K. B. Wilson (1992). On the experimental attainment of optimum conditions. In *Breakthroughs in statistics: methodology and distribution*, pp. 270–310. Springer. 70
- Campbell, D. T. and T. D. Cook (1979). Quasi-experimentation. *Chicago, IL: Rand Mc-Nally* 1(1), 1–384.
- Chowdhury, M., M. Chen, and S. Mandal (2020). A class of optimization problems on minimizing variance based criteria in respect of parameter estimators of a linear model. *Communications in Statistics-Simulation and Computation* 49(10), 2719–2731. 45
- Duarte, B. P., A. C. Atkinson, and N. Oliveira (2025). Compound optimal design of experiments–semidefinite programming formulations. *Optimization and Engineering*, 1–23. 161, 177
- Fedorov, V. V. (1972). *Theory of optimal experiments*. New York: Academic Press. 3, 15, 24, 77
- Fedorov, V. V. and S. L. Leonov (2014). *Optimal Design for Nonlinear Response Models*. Chapman & Hall/CRC Biostatistics Series. Boca Raton: CRC Press. 115, 177
- Fellman, J. (1989). An empirical study of a class of iterative searches for optimal designs. *Journal of Statistical Planning and Inference* 21, 85–92. 45
- Ghosh, S. (2025). *Statistical Planning and Inference: Concepts and Applications*. John Wiley & Sons. 3

- Ghosh, S. and Y. Tian (2006). Optimum two level fractional factorial plans for model identification and discrimination. *Journal of Multivariate Analysis* 97(6), 1437–1450. 177
- Go, J. and P. Chen (2025). Accurate, scalable, and efficient bayesian optimal experimental design with derivative-informed neural operators. *Computer Methods in Applied Mechanics and Engineering* 438, 117845. 178
- Goos, P. and B. Jones (2011). *Optimal design of experiments: a case study approach*. John Wiley & Sons. 177
- Han, J., M. Kamber, and J. Pei (2012). Data mining: Concepts and. *Techniques*, Waltham: Morgan Kaufmann Publishers. 137
- Helin, T., Y. Marzouk, and J. R. Rojo-Garcia (2025). Bayesian optimal experimental design with wasserstein information criteria. *arXiv preprint arXiv:2504.10092*. 178
- Huan, X., J. Jagalur, and Y. Marzouk (2024). Optimal experimental design: Formulations and computations. *Acta Numerica* 33, 715–840. 161
- Jiang, H. and R. Yue (2021). D-optimal population designs in linear mixed effects models for multiple longitudinal data. *Statistical Theory and Related Fields* 5(2), 88–94. 162, 178
- Karlin, S. and W. J. Studden (1966). Optimal experimental designs. *The Annals of Mathematical Statistics* 37(4), 783–815. 54
- Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)* 21(2), 272–304. 1, 16, 46

- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 849–879. 1, 23, 116, 119
- Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* 12, 363–366. 18
- Lin, C. D., D. Bingham, R. R. Sitter, and B. Tang (2010). A new and flexible method for constructing designs for computer experiments. *Annals of Statistics*, 1460–1477. 177
- López-Fidalgo, J. (2023). Optimal experimental design. *Lecture Notes in Statistics* 226. 3, 16, 45, 115, 177
- Mandal, S. and B. Torsney (2000). Algorithms for the construction of optimizing distributions. *Communications in Statistics-Theory and Methods* 29(5-6), 1219–1231. 28
- Mandal, S. and B. Torsney (2006). Construction of optimal designs using a clustering approach. *Journal of Statistical Planning and Inference* 136(3), 1120–1134. 28
- Mandal, S., B. Torsney, and K. Carriere (2005). Constructing optimal designs with constraints. *Journal of Statistical Planning and Inference* 128(2), 609–621. 28
- Mandal, S., B. Torsney, and M. Chowdhury (2017). Optimal designs for minimising covariances among parameter estimators in a linear model. *Australian & New Zealand Journal of Statistics* 59(3), 255–273. 75, 93, 103, 153

- Noto, D., C. M. Barbagallo, A. L. Cascio, A. B. Cefalù, G. Cavera, R. Caldarella, G. Marino, S. Travali, I. Cutaia, S. Maringhini, et al. (1999). Lipoprotein (a) levels in relation to albumin concentration in childhood nephrotic syndrome. *Kidney international* 55(6), 2433–2439. 75
- Oyet, A. J. (2002). Minimax A- and D-optimal integer-valued wavelet designs for estimation. *Canadian Journal of Statistics* 30(2), 301–316. 177
- Paul, S. and K. K. Saha (2007). The generalized linear model and extensions: a review and some biological and environmental applications. *Environmetrics: The official journal of the International Environmetrics Society* 18(4), 421–443. 177
- Pauphilet, J. (2025). Optimal experimental design with routing constraints. *Optimization Online preprint*. 162, 177
- Pukelsheim, F. (2006). Optimal design of experiments, Society for Industrial and Applied Mathematics. 15, 45, 137
- Rohde, C. A. (1965). Generalized inverses of partitioned matrices. *Journal of the Society for Industrial and Applied Mathematics* 13(4), 1033–1035. 54
- Seurat, J., Y. Tang, F. Mentré, and T. T. Nguyen (2021). Finding optimal design in nonlinear mixed effect models using multiplicative algorithms. *Computer Methods and Programs in Biomedicine* 207, 106126. 177
- Shen, W., J. Dong, and X. Huan (2025). Variational sequential optimal experimental design using reinforcement learning. *Computer Methods in Applied Mechanics and Engineering* 444, 118068. 177

- Sibson, R. (1974). D_A -optimality and duality. *Progress in Statistics* 2, 677–692.
- Silvey, S. (1980). *Optimal design: an introduction to the theory for parameter estimation*. London: Chapman and Hall. 15, 54
- Silvey, S., D. Titterington, and B. Torsney (1978). An algorithm for optimal designs on a design space. *Communications in Statistics-Theory and Methods* 7(14), 1379–1389. 27, 68, 85, 144
- Silvey, S. D. and D. M. Titterington (1973). A geometric approach to optimal design theory. *Biometrika* 60(1), 21–32. 54
- Sinha, S. K. and X. Xu (2011). Sequential D-optimal designs for generalized linear mixed models. *Journal of Statistical Planning and Inference* 141(4), 1394–1402. 177
- Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* 12(1-2), 1–85. 1
- Titterington, D. (1976). Algorithms for computing D-optimal designs on a finite design space. In *Proceedings of the 1976 Conference on Information Science and Systems*, Vol. 3, pp. 213–216. Johns Hopkins University.
- Torsney, B. (1977). Contribution to discussion of ‘maximum likelihood estimation via the EM algorithm’ by Dempster et al. *Journal of the Royal Statistical Society, Series B* 39, 26–27. 27

- Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. In *Semi-Infinite Programming and Applications*, pp. 249–260. Springer. 28
- Torsney, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. *Optimal Design and Analysis of Experiments* 215, 361–370. 28
- Torsney, B. and A. M. Alahmadi (1995). Designing for minimally dependent observations. *Statistica Sinica*, 499–514. 52, 74, 81, 103, 153
- Torsney, B. and S. Mandal (2001). Construction of constrained optimal designs. In *Optimum Design 2000*, pp. 141–152. Springer. 28, 115
- Wang, X., J. Lim, S. Kim, and K. Hahn (2015). Estimating cell probabilities in contingency tables with constraints on marginals/conditionals by geometric programming with applications. *Computational Statistics* 30, 107–129. 115
- Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society. Series B (Methodological)*, 123–130. 21
- Wong, W. K., Y. Yin, and J. Zhou (2019). Optimal designs for multi-response nonlinear regression models with several factors via semidefinite programming. *Journal of Computational and Graphical Statistics* 28(1), 61–73. 177
- Wu, C. (1978). Some iterative procedures for generating nonsingular optimal designs. *Communications in Statistics A* 14(14), 1399–1412. 68, 85, 144

- Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 133–147.
- Xu, X. and S. K. Sinha (2021). Robust designs for generalized linear mixed models with possible model misspecification. *Journal of Statistical Planning and Inference* 210, 20–41. 177
- Yi, Y. and Y. Yuan (2013). An optimal allocation for response-adaptive designs. *Journal of Applied Statistics* 40(9), 1996–2008. 177
- Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *Annals of Statistics* 38, 1593–1606. 28
- Zhou, J. (2008). D-optimal minimax regression designs on discrete design space. *Journal of Statistical Planning and Inference* 138(12), 4081–4092. 177