COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND
INFORMATICS

OPTIMAL EXPERIMENTAL DESIGNS FOR ESTIMATING LINEAR PARAMETER SUBSYSTEMS

Dissertation thesis

2018 **Mgr. Samuel ROSA**

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

OPTIMAL EXPERIMENTAL DESIGNS FOR ESTIMATING LINEAR PARAMETER SUBSYSTEMS

Dissertation thesis

Study programme:	Applied Mathematics
Branch of study:	1114 Applied Mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	doc. Mgr. Radoslav Harman, PhD.

Bratislava 2018 **Mgr. Samuel ROSA**

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

OPTIMÁLNE NÁVRHY EXPERIMENTOV PRE ODHADOVANIE SYSTÉMOV LINEÁRNYCH FUNKCIÍ PARAMETROV

Dizertačná práca

Študijný program:	Aplikovaná matematika
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	doc. Mgr. Radoslav Harman, PhD.

Bratislava 2018 **Mgr. Samuel ROSA**





Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Mgr. Samuel Rosa
Study programme:	Applied Mathematics (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	Applied Mathematics
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Optimal experimental designs for estimating linear parameter subsystems

Annotation: A common objective of an experiment is to estimate some system of linear functions of the parameters of a statistical model. In the dissertation thesis, we study and extend the theory of optimal experimental designs for such experimental objectives. In particular, we develop a theory for optimal approximate designs for estimating systems of treatments contrasts in models with nuisance effects; for comparing drug doses with the placebo in dose-escalation studies; and for estimating treatment contrasts in models with only treatment effects. We also compare the theory of optimality for estimating systems of interest with the recently developed weighted optimality theory.

Aim: The aim of the thesis is to provide analytical results on optimal experimental designs for estimating systems of linear functions of parameters in various statistical models, focusing on the approximate designs.

Tutor:	doc. Mgr. Radoslav Harman, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, DrSc.
Assigned:	06.02.2014

Approved: 20.02.2014 prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Mgr. Samuel Rosa
Študijný program: aplikovaná matematika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: aplikovaná matematika
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Optimal experimental designs for estimating linear parameter subsystems
Optimálne návrhy experimentov pre odhadovanie systémov lineárnych funkcií parametrov

Anotácia:

Častým cieľom experimentu je odhadnutie systému lineárnych funkcií parametrov štatistického modelu. V dizertačnej práci študujeme a rozširujeme teóriu optimálnych návrhov experimentov pre takéto experimentálne ciele. Konkrétne, rozvíjame teóriu optimálnych aproximatívnych návrhov pre odhadovanie systémov kontrastov ošetrení v modeloch s rušivými vplyvmi, pre porovnanie dávok lieku s placebom v štúdiách so zvyšujúcou sa dávkou lieku, a pre odhadovanie kontrastov ošetrení v modeloch, ktoré závisia len od vplyvov ošetrení. Tiež porovnávame teóriu optimality pre odhadovanie systémov záujmu s nedávno vyvinutou teóriou váženej optimality.

Cieľ: Cieľom dizertačnej práce je poskytnúť analytické výsledky o optimálnych návrhoch experimentov pre odhadovanie systémov lineárnych funkcií parametrov v rôznych štatistických modeloch, zamerajúc sa na aproximatívne návrhy.

Školiteľ: doc. Mgr. Radoslav Harman, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, DrSc.

Dátum zadania: 06.02.2014

Dátum schválenia: 20.02.2014 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
školiteľ

Acknowledgements I would like to express my gratitude to my supervisor doc. Mgr. Radoslav Harman, PhD. for his help and guidance.

Abstrakt

V tejto práci sa zaoberáme optimálnymi návrhmi experimentov pre odhadovanie systémov lineárnych funkcií neznámych parametrov modelu. Takýto systém vzniká napríklad ak cieľom experimentu je porovnať ošetrenia s kontrolným ošetrením. V práci formulujeme nutné a postačujúce podmienky optimality aproximatívnych návrhov pre odhadovanie systémov kontrastov ošetrení v modeli s vplyvmi ošetrení a s rušivými vplyvmi. Takisto poskytujeme pre tento model metódu konštrukcie optimálnych aproximatívnych návrhov s malým nosičom. Ďalej sa zaoberáme optimálnymi aproximatívnymi návrhmi pre porovnanie vplyvov dávok lieku s placebom v štúdiách so zvyšujúcou sa dávkou lieku. Poskytujeme optimálne návrhy pre tieto štúdie pre viaceré kritériá optimality. Tiež získavame reprezentáciu aproximatívnych návrhov pre porovnanie dvojíc ošetrení v modeli, v ktorom sú iba vplyvy ošetrení, pomocou teórie grafov. Na základe poskytnutej reprezentácie formulujeme viacero výsledkov o optimálnych návrhoch v tomto modeli. Nakoniec analyzujeme nedávno navrhnuté kritériá váženej optimality a v práci ich porovnávame so známymi kritériami pre odhadovanie systémov funkcií parametrov.

Kľúčové slová: Aproximatívny návrh, Optimálny návrh, Kontrast ošetrení, Štúdia so zvyšujúcou sa dávkou lieku, Teória grafov, Vážená optimalita

Abstract

In this thesis, we study optimal designs for estimating systems of interest, i.e., for estimating sets of linear functions of the unknown parameters of the model. Such a system arises, e.g., from the experimental objective of comparing test treatments with a control. We obtain necessary and sufficient conditions for optimality of approximate designs for estimating systems of treatment contrasts in a model with treatment and nuisance effects as well as a method for the construction of optimal approximate designs with small support in this model. Furthermore, we investigate optimal approximate designs for comparing the effects of the drug doses with the effect of the placebo in dose-escalation studies. We obtain optimal designs for these studies for multiple optimality criteria. We also provide graph representation of approximate designs for estimating systems of pairwise treatment comparisons in a model with treatment effects only. Based on the provided representation, we obtain various optimality results in such a model. Finally, we study the recently developed weighted optimality criteria and compare them to the well-known criteria for estimating systems of interest.

Keywords: Approximate design, Optimal design, Treatment contrast, Dose-escalation study, Graph theory, Weighted optimality

Preface

Before an experiment is performed, it is beneficial to choose experimental settings that allow one to acquire as much information as possible about the phenomenon under investigation. The branch of statistics that studies the process of choosing the 'best' experimental conditions is called (optimal) design of experiments.

The design of experiments as a statistical discipline has been developed since the beginning of the 20th century, with notable contribution of Fisher [1935]. Today, it is a well-developed field of statistics, covered in multiple monographs, e.g., Pázman [1986], Pukelsheim [1993] and Atkinson et al. [2007].

In this work, I examine optimal designs for estimating systems of interest in various settings, focusing mostly on the approximate designs.

To formalize which designs are 'best', the optimality criteria were introduced. For each competing design, the value of the chosen optimality criterion is a single number that represents the amount of relevant information that can be obtained if that design is used. Then, the design that attains the highest value of the optimality criterion is said to be optimal. Because the single number extremely compresses the information provided by an experiment, there are many criteria, each capturing a different aspect of the available information. This calls for the study of multiple criteria and their justification in the considered settings, as is the case in this thesis.

The approximate designs were proposed as an approach that makes the optimum design problem easier to tackle (see Kiefer [1959]). They are continuous relaxations of the exact designs – they specify only the proportions, rather than actual numbers, of all trials that are to be assigned to the particular experimental conditions.

It is often assumed that the objective of the experiment is the estimation of the vector of unknown parameters of the model. However, the experimenters may be interested in estimating a system of linear functions of these parameters. Some parameters may be of no interest (e.g., the effects of blocking or a time trend), and the experimental objective may consist of estimating specific comparisons of the other ones (e.g., comparing test treatments with a control). The experimental design should correspond to the stated experimental objective; therefore, if the interest lies in a system of linear functions of the parameters, the optimality considerations should reflect this. The reader can find a thorough and rigorous presentation of the theory on optimal designs for systems of interest in the monograph Pukelsheim [1993].

My interest in the approximate designs was motivated primarily by the fact that they tend to allow for fairly general and not too complicated results and formulas, circumventing the multitude of special cases and convoluted expressions that are often unavoidable in the exact design case. This tendency for 'nicer' results for approxi-

mate designs is demonstrated throughout this work. The reason for the theoretical (rather than focused on solving particular practical problems) and analytical (rather than algorithmic) nature of the thesis is perhaps the same. It is the kind of beauty and enjoyment that can be found in theory – particularly in simple and/or general theoretical results.

My interest in the general systems of linear functions of parameters arose from the original research interest in the study of optimal approximate designs for treatment contrasts in a model with treatment and nuisance effects (see Chapter 2). The rest of the work is all in some sense motivated by this original research interest: Chapter 3 deals with a special case of the aforementioned model, but with additional design constraints; in Chapter 4, a simple model that proves to play an important role in the more general model of Chapter 2 is studied; and Chapter 5 compares the criteria for estimating a system of interest, which are applied in the other chapters, with the recently developed weighted optimality criteria.

Contents

Preface **9**

Notation **13**

Introduction **14**

1 Design of experiments **16**
1.1 Statistical model 16
1.2 Design of experiments 17
1.3 Equivalence theorems 21
1.4 Rank-deficient systems 22
1.5 Design properties in treatment-nuisance model 23
1.6 Model without nuisance effects 24

2 Designs resistant to nuisance effects **26**
2.1 Introduction 26
2.2 Optimal approximate designs 29
2.2.1 Preliminaries 29
2.2.2 Conditions for optimality 32
2.2.3 Rank-deficient subsystems 36
2.2.4 Optimal treatment proportions for selected systems of contrasts 37
2.3 Examples 42
2.3.1 Trend-resistant designs 42
2.3.2 Block designs and row-column designs 45
2.4 Constructing efficient exact designs 46

3 Designs in dose-escalation studies **52**
3.1 Introduction 52
3.2 The model 53
3.3 E-optimality 55
3.3.1 Standard designs 55
3.3.2 Extended designs 57
3.3.3 Interpretation of E-optimality for dose-escalation 61
3.3.4 Comparing standard and extended designs 64
3.4 LV-optimality 66
3.5 MV-optimality 70
3.6 Discussion 70

4 Optimal treatment designs represented by graphs **73**

4.1 Introduction . . . 73
4.2 Experimental settings . . . 74
4.3 Graph representation . . . 75
4.3.1 Graph theory . . . 75
4.3.2 Representation of treatment proportions . . . 77
4.4 Optimality . . . 81
4.4.1 Permutation of treatments . . . 81
4.4.2 D-optimality . . . 81
4.4.3 A-optimality . . . 83
4.4.4 E-optimality . . . 85
4.4.5 Symmetric systems of contrasts . . . 90
4.5 Discussion . . . 93

5 Weighted optimality **96**
5.1 Introduction . . . 96
5.1.1 The model . . . 98
5.1.2 Weighted optimality . . . 99
5.1.3 Standard optimality for a system of interest . . . 100
5.2 Equivalence of weighted and standard optimality . . . 101
5.2.1 Equivalence . . . 101
5.2.2 Applications . . . 102
5.3 Extending the theory of weighted optimality . . . 104
5.3.1 General definition of the weight matrix . . . 104
5.3.2 Weighted optimality for any set of estimable functions . . . 106
5.3.3 Primary and secondary weights . . . 108
5.3.4 Applications . . . 110

Conclusion **113**

Resumé **116**

References **127**

Notation

1_n	the $n \times 1$ vector of ones
0_n	the $n \times 1$ vector of zeros
e_i	the i-th standard unit vector, i.e., a vector with the i-th element equal to one, and all other elements equal to zero
I_n	the $n \times n$ identity matrix
J_n	the $n \times n$ matrix of ones
$J_{m \times n}$	the $m \times n$ matrix of ones
$0_{m \times n}$	the $m \times n$ matrix of zeros
$\mathfrak{S}_+^k$	the set of all $k \times k$ symmetric nonnegative definite matrices
$\mathfrak{S}_{++}^k$	the set of all $k \times k$ symmetric positive definite matrices
$\succeq$	the Loewner ordering; $A \succeq B \Leftrightarrow A - B \in \mathfrak{S}_+^k$
$\mathcal{C}(A)$	the column space of matrix A, i.e., the linear space generated by the columns of A
$\mathcal{N}(A)$	the null space of matrix A, i.e., the set of vectors x satisfying $Ax = 0$
A^-	a generalized inverse of A: any matrix satisfying $AA^-A = A$
A^+	the Moore-Penrose pseudoinverse of A
$A^{+1/2}$	the matrix $(A^+)^{1/2}$
$\mathrm{tr}(A)$	trace of matrix A
$\mathrm{Det}(A)$	the pseudo determinant of $A \in \mathfrak{S}_+^k$ (i.e., the product of all nonzero eigenvalues of A)
$\lambda_1(A), \ldots, \lambda_k(A)$	the eigenvalues of $A \in \mathfrak{S}_+^k$ in descending order: $\lambda_1(A) \geq \ldots \geq \lambda_k(A)$
$\lambda_{\min}(A)$	the smallest *positive* eigenvalue of $A \in \mathfrak{S}_+^k$
$\lambda_{\max}(A)$	the largest eigenvalue of $A \in \mathfrak{S}_+^k$; $\lambda_{\max}(A) = \lambda_1(A)$
$\mathrm{diag}(v_1, \ldots, v_k)$	the diagonal matrix with diagonal elements corresponding to the elements of $v_1, \ldots, v_k$, where $v_1, \ldots, v_k$ are column or row vectors
x^{-1}	the vector of inverted components of vector x; $x = \left(x_1, \ldots, x_n\right)^T$, $x_i \neq 0$ for all i, then $x^{-1} := \left(x_1^{-1}, \ldots, x_n^{-1}\right)^T$
$A \otimes B$	the Kronecker product of matrices A and B
A_{-ij}	the matrix obtained from A by deleting its i-th row and j-th column
$A_{i,j}$	the element of the matrix A in the i-th row and j-th column

Introduction

Consider an experiment described by a linear regression model with a vector of unknown parameters β. The usual experimental objective is to estimate the parameter vector β; however, one might be more interested in estimating a system of parameters $K^T\beta$ given by a matrix K, as noted in Preface.

The aim of this work is to examine optimal designs for estimating systems of linear functions of parameters (systems of interest in short) in various settings, focusing on the approximate designs. In particular, the main objectives of this thesis are the following:

1. Develop a general theory of optimal approximate designs for estimating systems of treatment contrasts under the presence of nuisance effects.

2. Provide analytical results on optimal approximate designs for comparing drug doses with the placebo in dose-escalation studies as modeled by Bailey [2009a].

3. Characterize the problem of obtaining optimal approximate designs for estimating systems of treatment contrasts in the model with treatment effects only, and provide optimality results for various systems of contrasts and various optimality criteria in such a model.

4. Compare the theory of optimality for estimating a system of interest with the recently developed theory of weighted optimality as described by Stallings and Morgan [2015].

Organization of the thesis

The first chapter is an introductory one. We formulate some standard results on optimal approximate designs for the general linear model, mostly following Pukelsheim [1993]. Then, we further investigate the models that are studied in detail in the following chapters: a model with treatment and nuisance effects, and a model with treatment effects only.

The remaining four chapters deal with the four main objectives of this thesis, in the order stated above. Each of these chapters is based on a submitted or published paper (in the following order: Rosa and Harman [2016], Rosa and Harman [2017], Rosa [2017], Rosa [2016]).

In the second chapter, we study optimal approximate designs in a general model with treatment effects and nuisance effects. We provide necessary conditions for optimality and sufficient conditions for optimality for general optimality criteria, resulting in obtaining entire classes of optimal designs. We also provide a method for constructing efficient exact designs by using the obtained optimality conditions.

In Chapter 3, we examine dose-escalation studies, which are performed for assessing the safety and tolerability of newly developed drugs. These studies can be looked upon as blocking experiments, which are a special case of the model investigated in Chapter 2. However, the unique nature of the dose-escalation studies results in constraints on the set of considered designs, and thus prevents one from applying standard results on block designs or those from Chapter 2. We provide optimal approximate designs for comparing drug doses with the placebo for selected optimality criteria.

Motivated by the result of Chapter 2 that states that the optimal treatment proportions are crucial in obtaining optimal designs in the model considered therein, we further examine these treatment proportions in Chapter 4. We represent a common class of systems of contrasts by graphs and we represent the treatment proportions by weights on these graphs. We obtain interpretations of the considered optimal design problems using graph terminology, and we demonstrate the applicability of the obtained graph representation by providing some optimality results.

In Chapter 5, we study the recently proposed weighted optimality criteria. These criteria can be used to express various interests in particular treatments, and also to express the experimental objective of estimating a system of interest. We compare these criteria to the criteria for estimating a system of interest, which we employ in the other chapters, and we show the equivalence of these two approaches. Moreover, we extend the theory of weighted optimality so that it captures more experimental objectives than the current definition of weighted optimality available in the literature.

1 Design of experiments

1.1 Statistical model

Consider the following experiment consisting of N trials. For each trial $i = 1, \dots, N$, the experimenters choose experimental conditions x_i from a finite set of given experimental conditions $\mathfrak{X}$. In this work, we consider a linear regression model for the responses of the experiment $Y_1, \dots, Y_N$:

$$Y_i = f^T(x_i)\beta + \varepsilon_i, \quad i = 1, \dots, N, \tag{1.1}$$

where $\beta \in \mathbb{R}^m$ is the vector of unknown parameters, $f : \mathfrak{X} \to \mathbb{R}^m$ is the regression function and $\varepsilon_1, \dots, \varepsilon_N$ are independent and identically distributed random errors with zero mean and a finite variance $\sigma^2 > 0$. Moreover, suppose that the aim of the experiment is to estimate a set of linear functions of the unknown parameters $K^T\beta$, where K is an $m \times s$ matrix. Throughout this section, we assume that $s \leq m$ and that $\operatorname{rank}(K) = s$, i.e., $K^T\beta$ is a full-rank system, if not stated otherwise.

In the following chapters, we often consider a special case of model (1.1), where the response of the i-th trial is determined by the effect $\tau_{u(i)}$ of the chosen treatment $u(i)$ and by the effects of the nuisance experimental conditions $t(i)$ from a finite set $\mathfrak{T}$, $|\mathfrak{T}| =: n < \infty$. We assume that the model is additive in the treatment and nuisance effects and that it can be expressed as

$$Y_i = \tau_{u(i)} + h^T(t(i))\theta + \varepsilon_i, \quad i = 1, \dots, N, \tag{1.2}$$

where $\tau := (\tau_1, \dots, \tau_v)^T$ is the vector of treatment effects, $\theta \in \mathbb{R}^d$ is the vector of nuisance parameters and $h : \mathfrak{T} \to \mathbb{R}^d$ is the regression function of the nuisance experimental conditions. Then, x_i is the ordered pair $(u(i), t(i))$, $\beta = (\tau^T, \theta^T)^T$ and $f(u, t) = (e_u^T, h^T(t))^T$.

Suppose that the experimental objective is to estimate a set of contrasts $Q^T\tau$, where $Q \in \mathbb{R}^{v \times s}$. A contrast is a linear function whose coefficients sum to zero, i.e., $Q^T 1_v = 0_s$. In the notation of model (1.1), $K^T = (Q^T, 0_{s \times d})$. We assume that we are interested in all treatments $1, \dots, v$, i.e., each treatment is present in Q (no row of Q is 0_s^T). Note that the aim is to estimate a set of *treatment* contrasts; i.e., the effects of the other experimental conditions are indeed a nuisance.

As an example, consider the experiment given by Cox [1951]. The experimenters have v methods for processing wool, which they aim to compare by processing N batches of wool, one batch at a time. For each batch, they choose which method to process it with. In our notation, the treatments are the methods of processing wool and for each batch, the response is the quantity or quality of the processed wool obtained from that batch. The nuisance effects are the effects of the time trend, because the

particular methods of processing wool are applied in a time sequence and the wool degrades in time. Thus, the nuisance conditions are the times in which the trials are performed. The (exact) design of the experiment specifies the sequence in which the treatments are applied (note that one treatment may be applied to multiple batches), e.g., 1,2,3,3,2,1 for $N = 6$ and $v = 3$.

1.2 Design of experiments

An *exact design* for an experiment given by model (1.1) is a function $\xi : \mathfrak{X} \to \{0, 1, \ldots, N\}$ that satisfies $\sum_{x \in \mathfrak{X}} \xi(x) = N$. The exact design defines how many trials will be performed under the respective experimental conditions (also called the design points): in each design point $x \in \mathfrak{X}$, $\xi(x)$ trials are performed. In this work, however, we focus more on the approximate designs.

An *approximate design* is a function $\xi : \mathfrak{X} \to [0, 1]$ that satisfies $\sum_{x \in \mathfrak{X}} \xi(x) = 1$. Instead of the actual numbers of trials for the respective experimental conditions, the approximate design specifies the proportions of trials: in each design point $x \in \mathfrak{X}$, the proportion $\xi(x)$ of the total number of trials are performed. Henceforth, by a design we mean an approximate design, if not stated otherwise. For a monograph on approximate designs, see, e.g., Pázman [1986] or Pukelsheim [1993]. In this chapter, we mostly follow Pukelsheim [1993].

Formally, the exact design 1,2,3,3,2,1 in the wool processing example in the previous section is a function $\xi : \{1, 2, 3\} \times \{1, \ldots, 6\} \to \{0, 1, \ldots, 6\}$ given by Table 1. The value $\xi(u, t)$ for treatment u and time t specifies the number of trials performed in time t with treatment u. The approximate design describing the same experiment would be given by a table similar to Table 1, except every 1 would be replaced by $1/6$. An alternative approximate design ξ_2 for the wool processing experiment is given by Table 2. Note that ξ_2 does not define an exact design, unless multiple batches can be processed at the same time.

$u \backslash t$	1	2	3	4	5	6
1	1	0	0	0	0	1
2	0	1	0	0	1	0
3	0	0	1	1	0	0

Table 1: An exact design ξ for processing wool. The values in the table represent $\xi(u, t)$. u - treatment number, t - time

To measure the amount of information an experiment provides, and thus to compare multiple designs, the notion of an information matrix was introduced.

$u\backslash t$	1	2	3	4	5	6
1	1/12	0	2/18	0	1/18	1/12
2	1/12	1/6	0	0	1/18	1/12
3	0	0	1/18	1/6	1/18	0

Table 2: An approximate design for processing wool. u - treatment number, t - time

Definition 1.1 (by Pukelsheim [1993]). *Let ξ be a design in model* (1.1). *Then its* moment matrix *is $M(\xi) = \sum_{x\in\mathfrak{X}} \xi(x) f(x) f^T(x)$ and its* information matrix *for estimating $K^T\beta$ is*

$$N_K(\xi) = \min_{L\in\mathbb{R}^{s\times m}: LK=I_s} LM(\xi)L^T, \tag{1.3}$$

where the minimum is taken with respect to the Loewner ordering.

The justification of the definition of the information matrix can be found in Pukelsheim [1993]. Note that any moment matrix as well as any information matrix is nonnegative definite.

Usually, for a given experiment, only designs under which the system $K^T\beta$ is estimable are considered. Such designs are called feasible for $K^T\beta$.

Definition 1.2 (by Pukelsheim [1993]). *The system $K^T\beta$ is said to be* estimable *under design ξ if $\mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))$. If $K^T\beta$ is estimable under ξ, the design ξ is called* feasible *for $K^T\beta$.*

If $K^T\beta$ is estimable, we denote its least-squares estimator by $\widehat{K^T\beta}$. Pukelsheim [1993] provides a closed-form expression of the information matrix in the case that $K^T\beta$ is estimable.

Lemma 1.3 (by Pukelsheim [1993]). *The information matrix $N_K(\xi)$ is non-singular if and only if $K^T\beta$ is estimable under ξ. In that case $N_K(\xi)$ attains the form*

$$N_K(\xi) = (K^T M^-(\xi) K)^{-1}.$$

We remark that in the case that ξ is feasible, the information matrix $N_K(\xi)$ is positive definite. Also note that for feasible ξ, the expression $(K^T M^-(\xi)K)^{-1}$ does not depend on the choice of $M^-(\xi)$, because $\mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))$ in such a case.

Let Ξ be the set of all designs considered for the experiment. Then, we should find the 'best' or a 'good' feasible design $\xi \in \Xi$. If not specified otherwise, Ξ is assumed to be the set of all designs $\xi : \mathfrak{X} \to [0,1]$, $\sum_x \xi(x) = 1$. Although the information matrix measures the amount of information we obtain from an experiment, it is problematic to directly compare various information matrices (e.g., the Loewner ordering is only

a partial ordering). The optimality criteria, which are functions that map the set of information matrices to real numbers, provide a helpful tool which allows us to easily compare information matrices and thus experimental designs. Then, a design $\xi \in \Xi$ that is feasible for $K^T\beta$ and that attains the maximal value of an optimality criterion Φ is said to be Φ-optimal for $K^T\beta$.

Definition 1.4 (from Pukelsheim [1993]). *Let Φ be a function $\Phi : \mathfrak{S}_+^s \to \mathbb{R}$ and let Ξ be a set of competing designs. Then a design $\xi^* \in \Xi$ is said to be* Φ-optimal *for $K^T\beta$ in Ξ if it is feasible for $K^T\beta$ and maximizes the function $\Phi\big(N_K(\xi)\big)$ among all designs $\xi \in \Xi$.*

In particular, a feasible design $\xi^* \in \Xi$ is said to be

D-optimal *if it maximizes the determinant $\Big(\det\big(N_K(\xi)\big)\Big)^{1/s}$*

A-optimal *if it maximizes the harmonic mean of the eigenvalues of the information matrix $\frac{s}{1/\lambda_1+\ldots+1/\lambda_s}$*

E-optimal *if it maximizes the smallest eigenvalue $\lambda_{\min}(N_K(\xi))$*

among all feasible designs $\xi \in \Xi$. Moreover, let $c \in \mathbb{R}^m$ and let Ξ be a subset of all designs satisfying $c \in \mathcal{C}(M(\xi))$. Then, $\xi^* \in \Xi$ is said to be c-optimal *in Ξ if it maximizes $(c^T M^-(\xi)c)^{-1}$.*

For brevity, we often write $\Phi(\xi)$ instead of $\Phi(N_K(\xi))$. Although we may formally choose Φ to be any function, we usually require some properties to be satisfied. We call functions satisfying these desirable properties the information functions.

Definition 1.5 (by Pukelsheim [1993]). *A function $\Phi : \mathfrak{S}_+^s \to \mathbb{R}$ is said to be*

isotonic *with respect to the Loewner ordering if it satisfies $A \succeq B \succeq 0 \Rightarrow \Phi(A) \geq \Phi(B)$,*

concave *when it satisfies $\Phi\big((1-\alpha)A + \alpha B\big) \geq (1-\alpha)\Phi(A) + \alpha\Phi(B)$ for all $\alpha \in (0,1)$, and $A, B \in \mathfrak{S}_+^s$,*

positively homogeneous *when it satisfies $\Phi(\alpha A) = \alpha\Phi(A)$ for all $\alpha > 0$, $A \in \mathfrak{S}_+^s$,*

superadditive *when it satisfies $\Phi(A+B) \geq \Phi(A) + \Phi(B)$ for all $A, B \in \mathfrak{S}_+^s$,*

nonnegative *when it satisfies $\Phi(A) \geq 0$ for all $A \in \mathfrak{S}_+^s$,*

nonconstant *when there exist $A, B \in \mathfrak{S}_+^s$ such that $\Phi(A) \neq \Phi(B)$,*

upper semicontinuous *when the upper level sets $\{A \in \mathfrak{S}_+^s \mid \Phi(A) \geq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$.*

A function $\Phi : \mathfrak{S}_+^s \to \mathbb{R}$ *is called an* information function *if it is positively homogeneous, superadditive, nonnegative, nonconstant and upper semicontinuous.*

For more properties of the information functions, see Pukelsheim [1993]. Notably, a positively homogeneous function $\Phi : \mathfrak{S}_+^s \to \mathbb{R}$ is superadditive if and only if Φ is concave, and thus the information functions are concave. We say that Φ has a *strictly concave version* if there exists an equivalent version of Φ, say Φ_s, that satisfies $\Phi_s((1-\alpha)A + \alpha B) > (1-\alpha)\Phi_s(A) + \alpha\Phi_s(B)$ for any $A, B \in \mathfrak{S}_{++}^s$, $\alpha \in (0,1)$. For simplicity, we sometimes say that Φ is a strictly concave criterion. Note that such Φ_s cannot be positively homogeneous. Two criteria are equivalent if they result in the same ordering of the information matrices; e.g., if one is an increasing transformation of the other.

A generalization of the D-, A- and E- optimality criteria are the Kiefer's Φ_p-optimality criteria.

Definition 1.6 (by Pukelsheim [1993])**.** *Let* $p \in [-\infty, 0]$*. Then, a design* ξ^* *is said to be* Φ_p*-optimal for* $K^T\beta$ *in* Ξ *if it is feasible for* $K^T\beta$ *and maximizes the value of Kiefer's optimality criterion* $\Phi_p(N_K(\xi))$ *in* Ξ*. Let* N *be a nonnegative definite* $s \times s$ *matrix with eigenvalues* $\lambda_1, \ldots, \lambda_s$*, where* $\lambda_{\min}$ *is the smallest eigenvalue of* N*. Then if* N *is positive definite, the* Φ_p *criterion is defined as*

$$\Phi_p(N) = \begin{cases} \left(\frac{1}{s}\sum_{j=1}^{s} \lambda_j^p\right)^{1/p}, & p \in (-\infty, 0), \\ \left(\prod_{j=1}^{s} \lambda_j\right)^{1/s}, & p = 0, \\ \lambda_{\min}, & p = -\infty. \end{cases}$$

Otherwise $\Phi_p(N) = 0$*.*

Note that with $p = 0$, -1 and $-\infty$ in Φ_p-criterion we get the D-, A- and E- optimality criterion, respectively. The Φ_p criteria satisfy the properties of the information functions, and for $p \in (-\infty, 0]$ they have a strictly concave version.

Another commonly used optimality criterion in designing experiments for estimating systems of interest is MV-optimality. A design is MV-optimal if it minimizes the maximum variance of the least squares estimators of the linear functions of interest.

Definition 1.7. *Let* $K = (k_1, \ldots, k_s)$ *and let* Ξ *be a subset of designs feasible for* $K^T\beta$*. Then, a design* $\xi^* \in \Xi$ *is said to be* MV*-optimal for* $K^T\beta$ *in* Ξ *if it minimizes*

$$\Psi_{MV}(\xi) = \max_{i=1,\ldots,s} k_i^T M^-(\xi) k_i.$$

For a function $\Phi : \mathfrak{S}_{++}^s \to (0, \infty)$, the *polar function* $\Phi^\infty : \mathfrak{S}_+^s \to [0, \infty)$ is given by

$$\Phi^\infty(H) = \inf_{N \in \mathfrak{S}_{++}^s} \frac{\operatorname{trace}(NH)}{\Phi(N)} \text{ for } H \in \mathfrak{S}_+^s.$$

1.3 Equivalence theorems

Throughout this work, we employ the Equivalence Theorems, mostly as formulated by Pukelsheim [1993]: the General Equivalence Theorem (GET), the GET for Φ_p-optimality, the GET for E-optimality and the GET for c-optimality; see Theorems 7.14, 7.19 and 7.21 by Pukelsheim [1993], and Corollary 5.1 by Pukelsheim [1980]. Note that in this section, it is assumed that Ξ is convex and compact.

Lemma 1.8 (General Equivalence Theorem)**.** *Let $\xi \in \Xi$ be a feasible design for $K^T\beta$ and let Φ be an information function. Then, ξ is Φ-optimal for $K^T\beta$ in Ξ if and only if there exists $D \in \mathfrak{S}^s_+$ that solves the polarity equation*

$$\Phi(C)\Phi^\infty(D) = \operatorname{tr}(CD) = 1,$$

where Φ^∞ is the polar information function to Φ, and there exists a generalized inverse G of $M(\xi)$ that satisfies the normality inequality

$$\operatorname{tr}(M(\tilde{\xi})GKN_K(\xi)DN_K(\xi)K^TG^T) \leq 1 \text{ for all } \xi \in \Xi.$$

Lemma 1.9 (GET for Φ_p-optimality)**.** *Let $\xi \in \Xi$ be a feasible design for $K^T\beta$ and let $p \in (-\infty, 0]$. Then, ξ is Φ_p-optimal for $K^T\beta$ in Ξ if and only if there exists a generalized inverse G of $M(\xi)$ that satisfies the normality inequality*

$$\operatorname{tr}(M(\tilde{\xi})GKN_K^{p+1}(\xi)K^TG^T) \leq \operatorname{tr}(N_K^p(\xi)) \text{ for all } \tilde{\xi} \in \Xi.$$

Lemma 1.10 (GET for E-optimality)**.** *A feasible design $\xi \in \Xi$ with its moment matrix M and information matrix N_K is E-optimal for estimating $K^T\beta$ in Ξ if and only if there exist a generalized inverse G of M and $E \in \mathfrak{S}^s_+$, $\operatorname{tr}(E) = 1$, such that*

$$\operatorname{tr}(M(\tilde{\xi})GKN_KEN_KK^TG^T) \leq \lambda_{\min}(N_K) \text{ for all } \tilde{\xi} \in \Xi. \tag{1.4}$$

Lemma 1.11 (GET for c-optimality)**.** *A design $\xi \in \Xi$ satisfying $c \in \mathcal{C}(M(\xi))$ is c-optimal in Ξ if and only if there exists a generalized inverse G of M, such that $c^TG^TM(\tilde{\xi})Gc \leq c^TM^-c$ for all $\tilde{\xi} \in \Xi$.*

Using the Equivalence Theorems, sufficient and necessary conditions for optimality for a wide class of optimality criteria can be formulated, see Theorem 8.13 and Corollary 8.14 by Pukelsheim [1993].

Lemma 1.12. *Let Φ be an information function that has a strictly concave version and let $\xi^* \in \Xi$ be Φ-optimal for $K^T\beta$ in Ξ. Let G be a generalized inverse of $M(\xi^*)$ that satisfies the normality inequality of the General Equivalence Theorem (Lemma 1.8). Then, a design $\xi \in \Xi$ is Φ-optimal for $K^T\beta$ in Ξ if and only if $M(\xi)GK = K$.*

Lemma 1.13. *Let $p \in [-\infty, 0]$ and let ξ^* be a Φ_p-optimal design for $K^T\beta$ in Ξ. Then if $p > -\infty$, any other design $\tilde{\xi} \in \Xi$ is also Φ_p-optimal if and only if it satisfies $M(\xi)GK = K$, where G is a generalized inverse of $M(\xi^*)$ satisfying the normality inequality of the General Equivalence Theorem for Φ_p-optimality (Lemma 1.9). If $p = -\infty$, then if $\tilde{\xi}$ satisfies $M(\tilde{\xi})GK = K$, it is $\Phi_{-\infty}$-optimal. If $\tilde{\xi}$ is $\Phi_{-\infty}$-optimal, it satisfies $M(\tilde{\xi})GKE = KE$, where G is a generalized inverse of $M(\xi^*)$, $E \in \mathfrak{S}^s_+$, $\mathrm{tr}(E) = 1$, and G and E satisfy the normality inequality of the General Equivalence Theorem for E-optimality (Lemma 1.10).*

We will also make use of the relationship between E-optimality and c-optimality, see Pukelsheim and Studden [1993], Pukelsheim [1993].

Lemma 1.14 (Theorem 7.23 by Pukelsheim [1993])**.** *Let $\xi \in \Xi$ be a feasible design for $K^T\beta$ with its information matrix N_K and let h, $\|h\| = 1$, be an eigenvector of N_K corresponding to the smallest eigenvalue of N_K. Then ξ is E-optimal for $K^T\beta$ in Ξ and $E = hh^T$ satisfies the normality inequality* (1.4) *in the General Equivalence Theorem 1.10 if and only if ξ is c-optimal, where $c = Kh$. If the smallest eigenvalue of N_K has multiplicity 1, ξ is E-optimal for $K^T\beta$ if and only if it is c-optimal, where $c = Kh$.*

1.4 Rank-deficient systems

Up to this point, we have always assumed that $K^T\beta$ is of full rank. However, in the following chapters, we also examine rank-deficient systems, where $s \leq m$ need not hold and $\mathrm{rank}(K) < s$. For a detailed study of such systems, see Pukelsheim [1993]. One example of a rank-deficient system is the set of centered contrasts $K = (Q^T, 0_{s\times d})^T$, $Q = I_v - J_v/v$ (see, e.g., Pukelsheim [1983]), which aims at estimating $\tau_1 - \bar{\tau}, \ldots, \tau_v - \bar{\tau}$, where $\bar{\tau} = \sum_u \tau_u/v$ is the average treatment effect.

In the rank-deficient case, we consider information functions that depend only on the eigenvalues of N, i.e., orthogonally invariant information functions (see, e.g., Harman [2004] for further details). A functions Φ is *orthogonally invariant* if it satisfies $\Phi(N) = \Phi(UNU^T)$ for any orthogonal matrix U. One notable class of the orthogonally invariant information functions is the class of the Φ_p criteria.

For these systems, the information matrix $N_K(\xi)$ is not well defined. Instead, following Pukelsheim [1993], for a feasible design we define the matrix $C_K(\xi) := (K^T M^-(\xi)K)^+$, where the superscript $+$ denotes the Moore-Penrose inverse. Such a matrix can be viewed as the information matrix for rank-deficient $K^T\beta$. Then, if $\Phi(N)$ depends only on the eigenvalues of N, instead of maximizing Φ defined on all eigenvalues of $N_K(\xi)$, we maximize Φ defined on the positive eigenvalues of $C_K(\xi)$.

For the full-rank system, the eigenvalues of the information matrix $N_K(\xi)$ are the inverses of the eigenvalues of $K^T M^-(\xi)K$. Similarly, the matrix $C_K(\xi)$ satisfies that its nonzero eigenvalues are inverses of the nonzero eigenvalues of the matrix $K^T M^-(\xi)K$.

Thus, at least in the sense of their spectra, the matrices $C_K(\xi)$ are an analogue to the information matrices for full-rank systems.

1.5 Design properties in treatment-nuisance model

In this section, we investigate further the properties of experimental designs in model (1.2). The moment matrix of a design ξ may be expressed in the form

$$M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi) \end{bmatrix},$$

where

$$M_{11}(\xi) = \operatorname{diag}\left(\sum_{t \in \mathfrak{T}} \xi(1,t), \ldots, \sum_{t \in \mathfrak{T}} \xi(v,t) \right),$$
$$M_{12}(\xi) = \left(\sum_{t \in \mathfrak{T}} \xi(1,t)h(t), \ldots, \sum_{t \in \mathfrak{T}} \xi(v,t)h(t) \right)^T,$$
$$M_{22}(\xi) = \sum_{t \in \mathfrak{T}} \left(\sum_{u=1}^{v} \xi(u,t) \right) h(t)h^T(t).$$

Let us denote the *Schur complement* of the moment matrix $M(\xi)$ (formally, the Schur complement of $M_{22}(\xi)$ in $M(\xi)$) as $M_\tau(\xi) = M_{11}(\xi) - M_{12}(\xi)M_{22}^-(\xi)M_{12}^T(\xi)$. Using the Schur complement, a generalized inverse of $M(\xi)$ can be constructed.

Lemma 1.15 (by Harville [1997])**.** *Let M be a nonnegative definite matrix partitioned as*

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix}.$$

Then, G given by

$$G = \begin{bmatrix} M_\tau^- & -M_\tau^- M_{12} M_{22}^- \\ -M_{22}^- M_{12}^T M_\tau^- & M_{22}^- + M_{22}^- M_{12}^T M_\tau^- M_{12} M_{22}^- \end{bmatrix} \tag{1.5}$$

is a generalized inverse of M, where $M_\tau = M_{11} - M_{12}M_{22}^-M_{12}^T$ is the Schur complement.

Based on Lemma 1.15, the estimability of $Q^T\tau$ and the information matrices for $Q^T\tau$ can be characterized using the Schur complement.

Lemma 1.16. *Let $K = \left(Q^T, 0_{s\times d}\right)^T$. Then, $K^T\beta$ is estimable in model (1.2) under a design ξ if and only if $\mathcal{C}(Q) \subseteq \mathcal{C}(M_\tau(\xi))$. In that case, the information matrix $N_K(\xi)$ can be expressed as $N_K(\xi) = (Q^T M_\tau^-(\xi) Q)^{-1}$.*

Proof. For a thorough proof, see, e.g., Rosa [2014]. □

Similarly, the matrix $C_K(\xi)$ of a feasible design is $C_K(\xi) = (Q^T M_\tau^-(\xi) Q)^+$.

1.6 Model without nuisance effects

One simple case of model (1.2) is a model with no nuisance effects

$$Y_i = \tau_{u(i)} + \varepsilon_i, \quad i = 1, \ldots, N. \tag{1.6}$$

We denote an (approximate) design in such a model as w; i.e., w is a function from $\{1, \ldots, v\}$ to $[0, 1]$ that satisfies $\sum_u w(u) = 1$. We say that w is a *treatment proportions design* or simply a *treatment design*. Because a design w always provides v treatment weights $w(1), \ldots, w(v)$, we often equivalently express w as a $v \times 1$ vector of weights $w = \big(w_1, \ldots, w_v\big)^T$.

The properties of a design w in model (1.6) are generally very simple to analyze. For instance, it is straightforward to show that the moment matrix of w is $M(w) = \mathrm{diag}(w)$. Moreover, the set of contrasts $Q^T\tau$ is estimable in (1.6) under w if and only if $w_i > 0$ for all i. In such a case, the information matrix of w is evidently $N_Q(w) = \big(Q^T \mathrm{diag}(w^{-1})Q\big)^{-1}$.

In the case of rank-deficient systems, for a feasible design $w > 0$ the rank-deficient information matrix becomes $C_Q(w) = (Q^T M^{-1}(w) Q)^+$.

Note that instead of (1.2), one could consider the model

$$Y_i = \mu + \tau_{u(i)} + \varepsilon_i, \quad i = 1, \ldots, N, \tag{1.7}$$

where μ is the constant term. Then, the moment matrix of a design w is

$$M(w) = \begin{bmatrix} M_{11}(w) & M_{12}(w) \\ M_{12}^T(w) & M_{22}(w) \end{bmatrix},$$

where $M_{11}(w) = \mathrm{diag}(w)$, $M_{12}(w) = w$ and $M_{22}(w) = 1$. The Schur complement of $M_{22}(w)$ in $M(w)$ is $M_\tau(w) = M_{11}(w) - M_{12}(w)M_{22}^-(w)M_{12}^T(w) = \mathrm{diag}(w) - ww^T$.

To express feasibility and the information matrix for $Q^T\tau$, we employ a simple technical lemma:

Lemma 1.17 (by Harville [1997])**.** *Let A be an $m \times n$ matrix and let G be its generalized inverse; let B be an $m \times p$ matrix. Then, there exists an $n \times p$ matrix X that satisfies the equation $AX = B$ if and only if $AGB = B$.*

Lemma 1.18. *A design w is feasible for $Q^T\tau$ in model (1.7) if and only if $w > 0$. Moreover, if w is feasible, then $N_Q(w) = (Q^T \mathrm{diag}(w^{-1})Q)^{-1}$.*

Proof. A design w is feasible if and only if $\mathcal{C}(Q) \subseteq \mathcal{C}(M_\tau(w))$, see Lemma 1.16.

Let w be feasible and let $w_i = 0$ for some $i \in \{1, \ldots, v\}$. Then, the i-th row of $M_\tau(w)$ is 0_v^T and clearly $\mathcal{C}(Q) \subseteq \mathcal{C}(M_\tau(w))$ cannot hold.

For the converse part, let $w > 0$ and let $G := \mathrm{diag}(w^{-1})$. Then,

$$\begin{aligned} M_\tau(w)GM_\tau(w) &= (\mathrm{diag}(w) - ww^T)\mathrm{diag}(w^{-1})(\mathrm{diag}(w) - ww^T) \\ &= (I_v - w1_v^T)(\mathrm{diag}(w) - ww^T) = \mathrm{diag}(w) - ww^T - ww^T + w1_v^T ww^T \\ &= \mathrm{diag}(w) - ww^T, \end{aligned}$$

where the last equality follows from the fact that w is a design and thus $1_v^T w = 1$. Therefore, G is a generalized inverse of $M_\tau(w)$. The condition $\mathcal{C}(Q) \subseteq \mathcal{C}(M_\tau(w))$ holds if and only if there exists a matrix X, such that $M_\tau(w)X = Q$, which is equivalent to $M_\tau(w)GQ = Q$ (see Lemma 1.17). Moreover, we have

$$M_\tau(w)GQ = (\mathrm{diag}(w) - ww^T)\mathrm{diag}(w^{-1})Q = (I_v - w1_v^T)Q = Q - w1_v^T Q = Q,$$

because $1_v^T Q = 0_s^T$, as Q is a system of contrasts. Thus, w is feasible.

We have shown that $\mathrm{diag}(w^{-1})$ is a generalized inverse of $M_\tau(w)$ for any feasible w, which yields $N_Q(w) = (Q^T M_\tau^-(w)Q)^{-1} = (Q^T \mathrm{diag}(w^{-1})Q)^{-1}$. □

The feasibility condition and the information matrix given by Lemma 1.18 coincide with the feasibility condition and the information matrix in model (1.6), respectively. Similarly, $C_Q(w) = (Q^T \mathrm{diag}(w^{-1})Q)^+$. It follows that in (1.7), any optimality results are the same as in model (1.2), i.e., the results obtained in the following chapters for model (1.2) hold also for model (1.7) with the constant term.

2 Designs resistant to nuisance effects

2.1 Introduction

In this chapter, we examine optimal designs in model (1.2), i.e., we provide optimal designs for estimating a system of s contrasts $Q^T\tau$ of treatment effects under the presence of nuisance effects. This chapter is based on paper Rosa and Harman [2016].

The results of an experiment may be affected both by conditions with effects that we wish to estimate and by other conditions with nuisance effects. For example, any experiment that consists of multiple trials performed in a time sequence may be subject to a nuisance time trend caused by the aging of the material used for the experiment, the wearing down of the experimental devices, changes in the temperature, etc. The objective of the experimental design in such cases is to eliminate these nuisance effects, or to provide as much information as possible on the effects of interest. In this chapter, we examine a simpler model without nuisance effects and then present designs that are optimal in this simpler model while retaining all of the useful information that is present in the general model with nuisance effects.

There is a large amount of literature on optimal designs in specific models with nuisance effects. Exact design theory has yielded many results on optimal block designs (e.g., Majumdar and Notz [1983], Majumdar [1996], Jacroux [2002]) and trend-resistant designs (e.g., Cox [1951], Cheng [1990], Atkinson and Donev [1996]) as well as block designs in the presence of a trend (e.g., Jacroux and Ray [1990], Jacroux et al. [1997]) and row-column designs (e.g., Jacroux [1982]). In the context of approximate design theory, Pukelsheim [1983] studied block designs that are universally optimal for estimating centered treatment contrasts, and Giovagnoli and Wynn [1985] obtained optimal block designs for comparing treatments with a control with respect to Kiefer's optimality criteria.

The approach of first finding a design in a simpler model and then ensuring that the information is retained in a finer model has been used, e.g., by Kunert [1983], Filipiak et al. [2009], Schwabe and Wierich [1995] and Schwabe [1996]. Kunert [1983] studied exact designs for estimating a maximal system of orthonormal contrasts in the case of universal optimality; Filipiak et al. [2009] considered multivariate models; Schwabe and Wierich [1995], Schwabe [1996] and others studied approximate product designs in additive models.

In this chapter, we provide a general theory for optimal approximate designs in models with both treatment and nuisance effects. We provide simple conditions for the optimality of approximate designs for the estimation of any system of contrasts in an additive model with any system of nuisance effects, unlike Pukelsheim [1983] and Giovagnoli and Wynn [1985], who restricted their attention to block designs, specific

systems of treatment contrasts, and specific optimality criteria.

One drawback of considering approximate designs is that they are not designs of actual finite-sized experiments, and it is sometimes difficult to construct exact designs from Φ-optimal approximate designs. In particular, Schwabe and Wierich [1995], Schwabe [1996] and others have provided optimal product designs in general settings. However, approximate product designs have highly regular structures and large support sizes and thus are difficult to convert into exact designs. We present an entire class of optimal approximate designs that always encompasses the class of optimal product designs. We show that the obtained class of optimal designs usually contains not only the optimal product designs but also designs with small support sizes, which can be employed to construct efficient exact designs. Moreover, we provide a method of obtaining these sparsely supported Φ-optimal approximate designs.

The results concerning approximate designs are usually simpler and more general than those concerning exact designs; therefore, they provide valuable insight into the qualitative aspects of the design problem. This is also the case in the model studied here. In contrast to the general settings considered in this chapter, the standard results regarding optimal exact designs (e.g., Majumdar and Notz [1983], Jacroux and Ray [1990] and others cited earlier) depend on the choices of the particular model with nuisance effects, the numbers of treatments and nuisance conditions, the systems of contrasts, and the optimality criterion. For instance, results on trend-resistant designs are usually tailored for evenly spaced time points, a specific time trend (often modeled using low-order polynomials), a given relationship between the numbers of times and treatments, and a symmetric system of contrasts, e.g., the maximum system of orthonormal contrasts. However, in application, these conditions often do not hold (see Atkinson and Donev [1996]).

In contrast to Kunert [1983], by employing the approximate theory, we provide a method for constructing Φ-optimal (approximate) designs for any system of contrasts and any optimality criterion, rather than limiting ourselves to universal optimality. Moreover, there always exists a design in the class of Φ-optimal designs presented in this chapter, which is not the case in Kunert [1983].

We show that a Φ-optimal approximate design may be obtained in two steps:

(i) Calculate the Φ-optimal proportions of treatment replications (treatment weights). These optimal proportions depend on the choice of the contrasts of interest and on the optimality criterion Φ; however, they do not depend on the nuisance effects.

(ii) Subject to maintaining the optimal proportions of treatment replications, distribute the treatments with respect to nuisance conditions such that the resulting design is *resistant to nuisance effects*.

Steps (i) and (ii) address two natural questions: how many times each treatment should be chosen and how these treatment replications should be arranged with respect to nuisance conditions. Designs that are resistant to nuisance effects represent an extension of designs that are orthogonal to the time trend (balanced for the trend, or trend free, cf. Cox [1951] and Jacroux and Ray [1990]) to a broader class of models and treatment contrasts.

Moreover, step (i) is a fairly simple, low-dimensional, convex optimization problem, which can often be solved analytically. Step (ii) is a problem of much larger scale, with the same number of variables as the original problem of finding a Φ-optimal design. However, we will show that step (ii) is linear, i.e., solvable through linear programming. Therefore, by constructing a Φ-optimal nuisance-resistant design in these two steps, we break down the process of finding Φ-optimal designs into two much simpler problems – one is much smaller, and the other is linear.

The set of optimal nuisance-resistant designs is typically large, and the vertex solutions to the linear problem of step (ii) tend to have small support sizes. Thus, once the optimal treatment proportions are calculated, we can employ the simplex method of linear programming to obtain optimal approximate designs with small support sizes. This allows us to construct efficient exact designs using a simple heuristic, especially in the presence of nuisance trends. Therefore, finding Φ-optimal nuisance-resistant designs not only breaks down the process of obtaining Φ-optimal designs into two much simpler steps but also tends to yield approximate designs that are suitable for practical use.

Often, some symmetric system of contrasts is considered, e.g., a system of centered treatment effects (e.g., Pukelsheim [1983]) or the set of all pairwise comparisons (e.g., Bailey and Cameron [2009]), resulting in optimal uniform treatment proportions. However, in general, the uniform treatment proportions need not be optimal, especially in non-symmetric systems of contrasts, most notably for the comparison of test treatments with a control (see Bechhofer and Tamhane [1981], Majumdar and Notz [1983]) or the comparison of two sets of treatments (see Majumdar [1986], Jacroux [2002]).

For selected systems of treatment contrasts and a wide class of optimality criteria Φ, we calculate Φ-optimal treatment proportions and thus obtain a class of Φ-optimal designs. For instance, to estimate the contrasts for comparing a set of test treatments with a set of controls, we provide MV-optimal designs and optimal designs with respect to Kiefer's Φ_p-optimality criteria, $p \in [-\infty, 0]$, including A- and E-optimality ($p = -1$ and $p = -\infty$, respectively). These results generalize the results given by Giovagnoli and Wynn [1985], who obtained Φ_p-optimal block designs for comparing treatments with one control. For any *completely symmetric* system of contrasts, we show that the uniform design is Φ-optimal for all orthogonally invariant information functions, which generalizes, for instance, a result reported by Pukelsheim [1983].

2.2 Optimal approximate designs

2.2.1 Preliminaries

Note that in this chapter, we assume that $K = (Q^T, 0_{s\times d})^T$, and that K is of full rank s, unless stated otherwise. Recall that by a design, we mean an approximate design, if not stated otherwise. For a design ξ in (1.2), the marginal design defined by $w(u) = \sum_t \xi(u,t)$ for all u represents the total weights of individual treatments in ξ and will be called the *treatment proportions design of* ξ. Such w is in fact a design in model (1.6). Note that if ξ is an exact design of size N and w is its treatment proportions design, then Nw is the vector of the numbers of replications of treatments in ξ.

Similarly to the treatment proportions design w, we define the *nuisance conditions design* α as a design in the marginal model without treatments:

$$Y_i = h^T(t(i))\theta + \varepsilon_i, \quad i = 1, \ldots, N. \tag{2.1}$$

That is, α is a function from $\mathfrak{T}$ to $[0,1]$ that satisfies $\sum_t \alpha(t) = 1$. For a design ξ in (1.2), the marginal design $\alpha(t) = \sum_u \xi(u,t)$ for all t defines the proportions of trials to be performed under particular nuisance conditions and will be called the *nuisance conditions design of* ξ.

Proposition 2.1. *Let ξ be a design in model (1.2), and let w be its treatment proportions design. Then, $N_K(\xi) \preceq N_Q(w)$.*

Proof. Let us partition the matrix L in

$$N_K(\xi) = \min_{L\in\mathbb{R}^{s\times m}: LK = I_s} LM(\xi)L^T$$

as $L = \big(L_1, L_2\big)$, where L_1 is an $s \times v$ matrix and L_2 is an $s \times d$ matrix. Then,

$$\begin{aligned} N_K(\xi) &= \min_{LK=I_s} LM(\xi)L^T = \min_{(L_1,L_2)(Q^T,0)^T=I_s} (L_1, L_2)M(\xi)(L_1, L_2)^T \\ &\preceq \min_{L_1Q=I_s} L_1 M_{11}(\xi) L_1^T = N_Q(w). \end{aligned}$$

□

This proposition implies that with the introduction of nuisance effects, the available information about the contrasts of interest cannot increase. However, as we will see, for a large class of designs, this information is retained exactly.

We will say that a design ξ with its treatment proportions design w is *resistant to nuisance effects*, or *nuisance resistant*, for a given system of contrasts Q if it satisfies

$$\left[\tfrac{1}{w_1}\sum_{t\in\mathfrak{T}} \xi(1,t)h(t), \quad \ldots, \quad \tfrac{1}{w_v}\sum_{t\in\mathfrak{T}} \xi(v,t)h(t)\right] Q = 0. \tag{2.2}$$

To show that nuisance resistant designs are indeed resistant to nuisance effects, we need some auxiliary lemmas.

Lemma 2.2. *Let $\tilde{M}$ be a nonnegative definite matrix. If a design ξ satisfies $M(\xi)\tilde{M}^{-}K = K$ for some generalized inverse $\tilde{M}^{-}$ of $\tilde{M}$, then (i) ξ is feasible for $K^T\beta$ and (ii) $K^TM^{-}(\xi)K = K^T\tilde{M}^{-}K$.*

Proof. The steps of the proof follow the proof of Theorem 8.13 from Pukelsheim [1993] (Lemma 1.12). Let $G := \tilde{M}^{-}$. Because $M(\xi)GK = K$, we obtain $M(\xi)X = K$, where $X = GK$. Therefore, $\mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))$, and hence, ξ is feasible. Let us premultiply the equation $M(\xi)GK = K$ by $K^TM^{-}(\xi)$ such that we obtain $K^TM^{-}(\xi)K$ on the right-hand side. The left-hand side is then equal to $K^TM^{-}(\xi)M(\xi)GK$. Note that $K^T = X^TM^T(\xi) = X^TM(\xi)$, and hence, the following holds:

$$K^TM^{-}(\xi)M(\xi)GK = X^TM(\xi)M^{-}(\xi)M(\xi)GK = X^TM(\xi)GK = K^TGK.$$

It follows that $K^TM^{-}(\xi)K = K^T\tilde{M}^{-}K$. □

Lemma 2.3. *Let $w > 0$ be a treatment proportions design, and let $G := \mathrm{diag}\big(w^{-1}, 0_d\big)$. Let ξ be a design in model (1.2). Then, ξ satisfies $M(\xi)GK = K$ if and only if (i) w is a treatment proportions design of ξ and (ii) ξ is resistant to nuisance effects.*

Proof. We may express $M(\xi)GK = K$ as

$$M_{11}(\xi)\mathrm{diag}\big(w^{-1}\big)Q = Q \quad \text{and} \quad M_{12}^T(\xi)\mathrm{diag}\big(w^{-1}\big)Q = 0.$$

Because both $M_{11}(\xi)$ and $\mathrm{diag}\big(w^{-1}\big)$ are diagonal matrices and all rows of Q are assumed to be nonzero vectors, the first equation is equivalent to $\frac{1}{w_u}\sum_t \xi(u,t) = 1$ for all u, which is (i). From the second equation, we find that every row of $M_{12}^T(\xi)\mathrm{diag}\big(w^{-1}\big) = \big[\frac{1}{w_1}\sum_t \xi(1,t)h(t), \ldots, \frac{1}{w_v}\sum_t \xi(v,t)h(t)\big]$ needs to be in $\mathcal{N}(Q^T)$, which is (ii). □

The following proposition justifies the definition of the nuisance resistant designs.

Proposition 2.4. *Let ξ be a nuisance-resistant design with its treatment proportions design $w > 0$; then, (i) ξ is feasible for $K^T\beta$; (ii) $K^TM^{-}(\xi)K = Q^TM^{-1}(w)Q$; and (iii) ξ has the same information matrix as w, i.e., $N_K(\xi) = N_Q(w)$.*

Proof. Let $\tilde{M} := \mathrm{diag}(w, 0_d)$. Then, $G := \mathrm{diag}(w^{-1}, 0_d)$ is a generalized inverse of $\tilde{M}$. From Lemma 2.3 it follows that $M(\xi)GK = K$ and from Lemma 2.2, it follows that (i) and (ii) hold. Statement (iii) is a direct consequence of (ii). □

Note that the conditions for resistance to nuisance effects also have another desirable property: they are invariant with respect to a regular reparametrization of the nuisance regressors. That is, a design is resistant to nuisance effects with respect to nuisance regressors h if and only if it is resistant to nuisance effects with respect to nuisance regressors $\tilde{h} = Rh$, where R is any non-singular $d \times d$ matrix.

In general, the class of designs that are resistant to nuisance effects depends on the chosen system of contrasts Q. Nevertheless, as we will show, there is a large subclass

of nuisance-resistant designs that is invariant with respect to the choice of Q, i.e., these designs satisfy (2.2) for any system of contrasts.

We will say that a design ξ in (1.2) with its treatment proportions design $w > 0$ is *balanced* if it satisfies

$$\frac{1}{w_1}\sum_{t\in\mathfrak{T}}\xi(1,t)h(t) = \frac{1}{w_2}\sum_{t\in\mathfrak{T}}\xi(2,t)h(t) = \ldots = \frac{1}{w_v}\sum_{t\in\mathfrak{T}}\xi(v,t)h(t). \tag{2.3}$$

If ξ is balanced, then for any $k \in \{1,\ldots,d\}$, the vector

$$s_k := \Big(\frac{1}{w_1}\sum_{t\in\mathfrak{T}}\xi(1,t)h_k(t), \ldots, \frac{1}{w_v}\sum_{t\in\mathfrak{T}}\xi(v,t)h_k(t)\Big)^T$$

satisfies $s_k = a_k 1_v$ for some $a_k \in \mathbb{R}$. Because Q is a matrix of contrasts, we have $1_v^T Q = 0_s^T$ and, hence, $s_k^T Q = 0_s^T$. It follows that a balanced design ξ is indeed nuisance resistant.

When the matrix of contrasts Q is of the maximum rank, $v-1$, the null space $\mathcal{N}(Q^T)$ has dimension 1 and consists of vectors of the form $a1_v$ for $a \in \mathbb{R}$. Hence, for such a Q, the balanced designs are the only nuisance-resistant designs, i.e., in this specific but frequent case, the concepts of resistance to nuisance effects and balancedness coincide. One consequence of this observation is that for given nuisance regressors h, the class of balanced designs is the intersection of the sets of nuisance-resistant designs with respect to all possible choices of the contrast matrix Q.

We remark that the conditions given in (2.3) mean that a design ξ is balanced with respect to nuisance effects. That is, for each regressor h_k, the weighted average of the values $h_k(t)$ with weights $\xi(u,t)/w_u$, $t \in \mathfrak{T}$, is the same for each treatment u. Balancedness can also be understood geometrically: Assume that for each u, we calculate the barycenter of the n points $h(t) \in \mathbb{R}^d$ with weights $\xi(u,t)/w_u$, $t \in \mathfrak{T}$. Then, these barycenters must be the same for all treatments u.

A typical experimental scenario is that we need to perform the same number of trials, usually one, under each nuisance condition $t \in \mathfrak{T}$. In this case, it is straightforward to show that the conditions given in (2.3) may be expressed in a more compact form, as follows.

Proposition 2.5. *Let ξ be a design in which the same weight is assigned to each nuisance condition, i.e., the nuisance conditions design of ξ is $\alpha = 1_n/n$. Then, ξ satisfies (2.3) if and only if it satisfies*

$$\frac{1}{w_u}\sum_{t\in\mathfrak{T}}\xi(u,t)h(t) = \frac{1}{n}\sum_{t\in\mathfrak{T}}h(t) \text{ for all } u \in \{1,\ldots,v\}. \tag{2.4}$$

In the case of an exact design ξ in which one trial is assigned to each nuisance condition, the balancedness of ξ has a straightforward interpretation. Suppose, for

instance, that the nuisance conditions represent time and that $h_1(t)$ is proportional to the temperature of the room at time t. For each treatment u, let T_u be the average temperature at the times of trials with treatment u. Then, the balance conditions for $h_1(t)$ mean that the temperature conditions are 'fair' for all treatments, in the sense that the average temperatures T_u are the same: $T_u \equiv T$ for all u.

Let w be a treatment proportions design, and let α be a nuisance conditions design. Then, a design ξ is the *product design* of w and α if it satisfies

$$\xi(u,t) = w(u)\alpha(t) \text{ for all } u \in \{1,\dots,v\}, t \in \mathfrak{T},$$

which we denote by $\xi = w \otimes \alpha$. Note that any product design $w \otimes \alpha$ satisfies $\frac{1}{w_u}\sum_t \xi(u,t)h(t) = \sum_t \alpha(t)h(t)$ for all u; therefore, any product design is balanced and, consequently, is also resistant to nuisance effects.

2.2.2 Conditions for optimality

The following theorem shows that the optimality of treatment proportions is a necessary condition for the optimality of a design in model (1.2).

Theorem 2.6. *Let $Q^T\tau$ be a full-rank system of contrasts, let Φ be an information function, let ξ^* be a Φ-optimal design in model (1.2) for $Q^T\tau$, and let w^* be its treatment proportions design. Then, w^* is Φ-optimal in (1.6) for $Q^T\tau$.*

Proof. Let ξ be a feasible design in (1.2). Using Proposition 2.1, we find that $N_K(\xi) \preceq N_Q(w)$, where w is the treatment proportions design of ξ. Moreover, because product designs are nuisance resistant, part (iii) of Proposition 2.4 implies that $N_Q(w) = N_K(w \otimes \alpha)$ for any nuisance conditions design α. Therefore, $N_K(\xi) \preceq N_K(w \otimes \alpha)$.

Suppose that w^* is not a Φ-optimal design. Then, there exists a design w_b in model (1.6) such that $\Phi(N_Q(w^*)) < \Phi(N_Q(w_\mathrm{b}))$. Then, $\Phi(N_K(\xi^*)) \leq \Phi(N_Q(w^*)) < \Phi(N_Q(w_\mathrm{b})) = \Phi(N_K(w_\mathrm{b} \otimes \alpha))$ for any nuisance conditions design α. This contradicts the Φ-optimality of ξ^*. □

From Theorem 2.6, it follows that to find an optimal approximate design, we need to break this process down into two steps: obtaining the optimal treatment weights followed by optimally allocating these weights to the nuisance conditions. Note that finding a Φ-optimal treatment proportions design is a convex v-dimensional optimization problem:

$$\max_{w>0,\, 1_v^T w=1} \Phi((Q^T \mathrm{diag}(w^{-1})Q)^{-1}), \tag{2.5}$$

which can usually be easily solved numerically, and often analytically, as we demonstrate in Section 2.2.4.

As noted in Section 2.1, although the optimal treatment proportions are often uniform, i.e., $\bar{w} = 1_v/v$, the optimality of uniform treatment proportions is not guaranteed in general. In the following simple example of the comparison of test treatments with a control, we demonstrate that the variance of the estimators can be reduced by using non-uniform treatment proportions.

Example 2.7. Consider a blocking experiment with $n = b$ blocks and $v = 5$ treatments for the purpose of comparing test treatments with a control (for a description of a blocking experiment, see Section 2.3.2). Let the first treatment be the control; then, the objective is to estimate $\tau_i - \tau_1$, $i = 2, \ldots, 5$, which can be expressed as $Q^T\tau$, where $Q = (-1_4, I_4)^T$.

Let $w_\gamma = (\gamma, 1_4^T(1-\gamma)/4)^T$ represent treatment proportions such that a weight $\gamma \in (0, 1)$ is assigned to the control and the remaining weight is split uniformly among the other treatments. Consider the product design $\xi = w_\gamma \otimes (1_b/b)$. Then, Proposition 2.4 states that $N_K(\xi) = N_Q(w_\gamma) = (Q^T M^{-1}(w_\gamma) Q)^{-1}$. The variance matrix of the least squares estimators of $Q^T\tau$ under ξ is equal to $\sigma^2/N \cdot N_K^{-1}(\xi)$, where N is the total number of trials. It can be calculated that $Q^T M^{-1}(w_\gamma) Q = 4I_4/(1-\gamma) + J_4/\gamma$, and thus, the variances of the estimators of $\tau_i - \tau_1$ are equal to each other and proportional to $V(\gamma) := \gamma^{-1} + 4/(1-\gamma)$. In Figure 1, the value of $V(\gamma)$ for $\gamma \in [1/6, 1/2]$ is displayed. In particular, for uniform treatment proportions, i.e., if $\gamma = 1/5$, the variances are proportional to $V(1/5) = 10$, whereas if $\gamma = 1/3$, the variances are proportional to $V(1/3) = 9$. Note that the minimum $V(\gamma)$ is attained for $\gamma = 1/3$ ($w_{1/3}$ represents the A-optimal treatment proportions; see Theorem 2.15). Thus, in this simple example, by using the A-optimal treatment proportions (i.e., by performing $1/3$ of all trials with the control treatment and $1/6$ of all trials with each of the other treatments), the variances of the contrasts of interest can be reduced to 90% of what they would be if uniform treatment proportions were chosen.

Once the optimal treatment proportions have been obtained, we need to allocate them to the nuisance conditions. Because nuisance-resistant designs retain all information on the contrasts of interest (see Proposition 2.4), such designs provide optimal allocations of these optimal treatment proportions.

Theorem 2.8. *Let $Q^T\tau$ be a full-rank system of contrasts, and let w^* be a Φ-optimal treatment proportions design for $Q^T\tau$. Let ξ^* be a nuisance-resistant design with its treatment proportions design w^*. Then, ξ^* is Φ-optimal for $Q^T\tau$ and has the same information matrix as w^*, i.e., $N_K(\xi^*) = \left(Q^T \mathrm{diag}((w^*)^{-1}) Q\right)^{-1}$.*

Proof. Let ξ be a feasible design in model (1.2), and let w be its treatment proportions design. Because Φ is isotonic, it follows from Proposition 2.1 that $\Phi(N_Q(w)) \geq \Phi(N_K(\xi))$. Since w^* is Φ-optimal, $\Phi(N_Q(w^*)) \geq \Phi(N_Q(w))$ and it is feasible; thus,

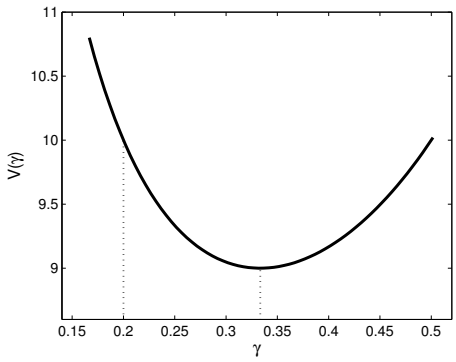


Figure 1: Dependence of the variances of the estimators in the case of comparisons with a control on the treatment proportions w_γ. For a given $\gamma \in (0,1)$, the value $V(\gamma)$ is proportional to the (common) variance of the estimators of $\tau_i - \tau_1$ $(i = 2, \ldots, 5)$ for a product design in which a proportion γ of the trials is assigned to the first treatment and proportions $(1-\gamma)/4$ are assigned to the other treatments. The dotted lines represent the contrast variances when w_γ is uniform ($\gamma = 1/5$) and when w_γ is A-optimal ($\gamma = 1/3$).

$w^* > 0$. Using Proposition 2.4, we find that $\Phi(N_K(\xi^*)) = \Phi(N_Q(w^*)) \geq \Phi(N_Q(w)) \geq \Phi(N_K(\xi))$, i.e., ξ^* is Φ-optimal. □

Balanced designs are nuisance resistant; therefore, balanced designs with Φ-optimal treatment weights w^* are Φ-optimal. Moreover, they have the same information matrix as w^*. Note that the set of optimal balanced designs is never empty, because it contains the set of product designs $w^* \otimes \alpha$ for all α. Since α is any nuisance conditions design, the class of Φ-optimal designs for model (1.2) is very large (unless $n = 1$).

Similar results on the optimality of product designs have been reported by Schwabe [1996] (cf. Theorem 3.2 therein) for a general additive model, $Y_i = \beta_0 + f_1^T(u_1(i))\beta_1 + f_2^T(u_2(i))\beta_2 + \varepsilon_i$. Note that general nuisance-resistant designs, because they need not have a product structure, are not covered by Schwabe [1996].

Lemma 1.12 enables a complete characterization of Φ-optimal designs for a strictly concave Φ. However, to use Lemma 1.12, we must obtain a matrix G that satisfies the normality inequality of the General Equivalence Theorem (Lemma 1.8).

Lemma 2.9. *Let Φ be an information function that has a strictly concave version. Then, there exists a unique Φ-optimal treatment proportions design for $Q^T\tau$, say w^*. Let $G := \mathrm{diag}((w^*)^{-1}, 0_d)$. Then, G satisfies the normality inequality of the General Equivalence Theorem for estimating $K^T\beta$ in model (1.2).*

Proof. Let w^* be a Φ-optimal treatment proportions design, and let $G_{11} := \mathrm{diag}\big((w^*)^{-1}\big)$. Then, the matrix G_{11}, which is the unique generalized inverse of $M(w^*)$, satisfies the normality inequality of the General Equivalence Theorem for model (1.6). By Lemma 1.12, a treatment proportions design w is Φ-optimal if and only if $M(w)M^{-1}(w^*)Q = Q$.

By the same argument as that used in the proof of Lemma 2.3, $M(w)M^{-1}(w^*)Q = Q$ holds if and only if $w = w^*$, which proves the uniqueness of w^*. Let $N^* := N_Q(w^*)$. Because G_{11} satisfies the normality inequality in model (1.6), there exists a matrix D that satisfies the polarity equation $\Phi(N^*)\Phi^\infty(D) = \mathrm{tr}(N^*D) = 1$ and the matrix $B_w = G_{11}QN^*DN^*Q^TG_{11}$ satisfies $\mathrm{tr}(M(\tilde{w})B_w) \leq 1$ for all $\tilde{w} > 0$.

Since $N_K(\xi^*)$ is positive definite and an equivalent version of Φ is strictly concave, there exists a unique Φ-optimal information matrix $N_K(\xi^*)$. The Φ-optimality of $N_K(w^* \otimes \alpha) = N_Q(w^*) = N^*$ for any nuisance conditions design α yields $N_K(\xi^*) = N^*$. Thus, the polarity equality holds in model (1.2) for the same matrix D. Let $\tilde{\xi}$ be a feasible design. Then, the left-hand side of the normality inequality in model (1.2) is $\mathrm{tr}(M(\tilde{\xi})B)$, where

$$B = \begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q \\ 0 \end{bmatrix} N^*DN^* \begin{bmatrix} Q^T & 0 \end{bmatrix} \begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} B_w & 0 \\ 0 & 0 \end{bmatrix}.$$

Then, because B_w satisfies the normality inequality in model (1.6), we obtain

$$\mathrm{tr}(M(\tilde{\xi})B) = \mathrm{tr}(M_{11}(\tilde{\xi})B_w) = \mathrm{tr}(M(\tilde{w})B_w) \leq 1,$$

where $\tilde{w}$ is the treatment proportions design of $\tilde{\xi}$. □

Theorem 2.6 provides the necessary conditions for optimality, and Theorem 2.8 provides sufficient conditions for optimality. It turns out that for the wide class of strictly concave optimality criteria, we can obtain conditions that are both necessary and sufficient for the optimality of a design ξ in model (1.2).

Theorem 2.10. *Let $Q^T\tau$ be a full-rank system of contrasts, let Φ be an information function that has a strictly concave version, and let w^* be the unique Φ-optimal treatment proportions design for $Q^T\tau$ in* (1.6)*. Then, a design ξ is Φ-optimal for $Q^T\tau$ in* (1.2) *if and only if (i) w^* is the treatment proportions design of ξ and (ii) ξ is resistant to nuisance effects.*

Proof. Lemma 2.9 ensures the uniqueness of w^*. Let $G = \mathrm{diag}((w^*)^{-1}, 0_d)$. Consider an extension of model (1.2), where $t \in \mathfrak{T}'$, $\mathfrak{T}' = \mathfrak{T} \cup \{z\}$, and $h(z) = 0_d$, and let ξ' be a design in this extended model, with its treatment proportions design w^*, that satisfies $\xi'(u,t) = 0$ for $t \neq z$. Then, $M(\xi') = \mathrm{diag}(w^*, 0_d)$ and G is a generalized inverse of $M(\xi')$. From Lemma 2.9, it follows that G satisfies the normality inequality of the General Equivalence Theorem for any nuisance regressors and, consequently, also for the extended model. By Lemma 1.12, a design ξ is Φ-optimal in the extended model if and only if $M(\xi)GK = K$. The equality $M(\xi)GK = K$ holds if and only if ξ satisfies (i) and (ii) from Lemma 2.3 in the extended model, i.e.,

$$\sum_{t \in \mathfrak{T}'} \xi(u,t) = w^*_u, \quad u = 1, \ldots, v, \tag{2.6}$$

$$\left[\tfrac{1}{w_1^*}\sum_{t\in\mathfrak{T}'}\xi(1,t)h(t),\quad\ldots,\quad\tfrac{1}{w_v^*}\sum_{t\in\mathfrak{T}'}\xi(v,t)h(t)\right]Q=0. \tag{2.7}$$

Let $\tilde{\Xi}$ be the set of feasible designs in the extended model that satisfy $\xi(u,z)=0$ for all u, which represents the set of all feasible designs in model (1.2). The Φ-optimal value in the extended model can be attained in $\tilde{\Xi}$, e.g., $\xi^*=w^*\otimes\alpha$, where α is any nuisance conditions design that satisfies $\alpha(z)=0$. Thus, a design ξ is Φ-optimal in $\tilde{\Xi}$ if and only if $\xi\in\tilde{\Xi}$ and it satisfies (2.6) and (2.7), i.e., $\sum_{t\in\mathfrak{T}}\xi(u,t)=w_u^*$ for $u=1,\ldots,v$, and (2.2). It follows that a design ξ is Φ-optimal in (1.2) if and only if its treatment proportions design is Φ-optimal and ξ is nuisance resistant. □

Because the balanced designs are the only nuisance-resistant designs for a system of contrasts of maximum rank, $v-1$, we obtain the following corollary.

Corollary 2.11. *Let Φ be an information function that has a strictly concave version, and let Q be a matrix of contrasts of rank $v-1$. Then, a design ξ is Φ-optimal for estimating $Q^T\tau$ if and only if its treatment proportions design is Φ-optimal in (1.6) and ξ is balanced.*

2.2.3 Rank-deficient subsystems

In the rank-deficient case, results analogous to those for the full-rank case hold. We will show that with the introduction of nuisance effects, the information about the treatment contrasts, as measured by $C_K(\xi)$, cannot increase. The ordering of the matrices $C_K(\xi)$ corresponds to the inverse ordering of the matrices $K^TM^-(\xi)K$. For any design ξ, we obtain $K^TM^-(\xi)K=Q^TM_\tau^-(\xi)Q$, and for its treatment proportions design, we obtain $Q^TM^-(w)Q=Q^TM_{11}^-(\xi)Q$. Moreover, $M_\tau(\xi)=M_{11}(\xi)-M_{12}(\xi)M_{22}^-(\xi)M_{12}^T(\xi)\preceq M_{11}(\xi)$; therefore, there exist generalized inverses that satisfy $M_\tau^-(\xi)\succeq M_{11}^-(\xi)$ (see Wu [1980]), and it follows that $K^TM^-(\xi)K\succeq Q^TM^-(w)Q$. Because $\Phi(N)$ depends only on the eigenvalues of N and the Moore-Penrose inverse X^+ has the inverse nonzero eigenvalues of X, it is implied that $\Phi\big((K^TM^-(\xi)K)^+\big)\leq\Phi\big((Q^TM^-(w)Q)^+\big)$, i.e., $\Phi(C_K(\xi))\leq\Phi(C_Q(w))$.

From part (ii) of Proposition 2.4, it follows that any nuisance-resistant design ξ has the same matrix $C_K(\xi)$ as its treatment proportions design, i.e., $C_K(\xi)=C_Q(w)$. Hence, Theorems 2.6 and 2.8 hold even in the rank-deficient case.

Theorem 2.12. *Let Φ be an information function, and let Q be a $v\times s$ matrix of contrasts with* $\operatorname{rank}(Q)<s$. *Let w^* be a Φ-optimal design for estimating $Q^T\tau$ in model (1.6). Then, the following statements hold:*

(i) Any nuisance-resistant design ξ whose treatment proportions design is w^* is Φ-optimal for estimating $Q^T\tau$ and $C_K(\xi)=\big(Q^T\operatorname{diag}((w^*)^{-1})Q\big)^+$.

(ii) If ξ is a Φ-optimal design in model (1.2) and w is its treatment proportions design, then w is Φ-optimal for estimating $Q^T\tau$ in model (1.6).

In particular, we obtain the optimality of balanced and product designs with optimal treatment weights.

2.2.4 Optimal treatment proportions for selected systems of contrasts

We say that the system of contrasts $Q^T\tau$ is *completely symmetric* if QQ^T is completely symmetric, i.e., if $QQ^T = aI_v + bJ_v$ for some $a, b \in \mathbb{R}$. It is easy to show that such a Q must satisfy $QQ^T = a(I_v - J_v/v)$ for some $a > 0$. We will show that several common systems of contrasts are completely symmetric.

We consider information functions Φ that are orthogonally invariant, i.e., $\Phi(UNU^T) = \Phi(N)$ for any $s \times s$ orthogonal matrix U and any $N \in \mathfrak{S}^s_+$. Note that a function Φ is orthogonally invariant if and only if $\Phi(N)$ depends only on the eigenvalues of N (see, e.g., Harman [2004] for further details).

Theorem 2.13. *Let $Q^T\tau$ be a completely symmetric system of contrasts. Then, the uniform treatment proportions design $\bar{w} = 1_v/v$ is Φ-optimal for estimating $Q^T\tau$ with respect to any orthogonally invariant information function Φ.*

Proof. First, let us assume that Q is of full column rank. Let w be a feasible treatment proportions design, and let P be a $v \times v$ permutation matrix. We define Pw as the design obtained via the P-permutation of the treatments in w, i.e., $Pw(u) = w(\pi_P(u))$ for $u \in \{1, \ldots, v\}$, where π_P is the permutation of the elements $\{1, \ldots, v\}$ that corresponds to the matrix P. Because $Pw > 0$, it is feasible, its moment matrix is $M(Pw) = PM(w)P^T$, and it has the information matrix $N_Q(Pw) = (Q^TPM^{-1}(w)P^TQ)^{-1}$.

We will use the well-known fact that for any matrix X, the nonzero eigenvalues of the matrices X^TX and XX^T are the same (e.g., 6.54(c) in Seber [2008]), including their multiplicities. Let us define $Y = Q^TM^{-1/2}(w)$ and $Z = Q^TPM^{-1/2}(w)$. Since QQ^T is completely symmetric, $Y^TY = Z^TZ$. Furthermore, $YY^T = Q^TM^{-1}(w)Q = N_Q^{-1}(w)$ and $ZZ^T = Q^TPM^{-1}(w)P^TQ = N_Q^{-1}(Pw)$; thus, $N_Q(w)$ and $N_Q(Pw)$ have the same set of nonzero eigenvalues. Because they are of the same (full) rank, it follows that $N_Q(w)$ and $N_Q(Pw)$ are orthogonally similar and that $\Phi(Pw) = \Phi(w)$. Note that analogous results hold in the rank-deficient case for the matrices $C_Q(w)$ and $C_Q(Pw)$.

The uniform treatment proportions design satisfies

$$\Phi\big(\bar{w}\big) = \Phi\left(\frac{1}{v!}\sum_{P-\text{perm.}} Pw\right) \geq \frac{1}{v!}\sum_{P-\text{perm.}} \Phi(Pw) =$$
$$= \frac{1}{v!}\sum_{P-\text{perm.}} \Phi(w) = \frac{1}{v!}v!\Phi(w) = \Phi(w),$$

where the inequality follows from the concavity of Φ. Thus, $\bar{w}$ is Φ-optimal. □

By a *maximal system of orthonormal contrasts*, we mean a set of $v-1$ contrasts that are orthogonal to each other and are of norm 1, i.e., $q_1, \ldots, q_{v-1}$ satisfy $q_i^T q_j = 0$ for $i \neq j$ and $q_i^T q_i = 1$ for all i. Note that a special case of a maximal system of orthonormal contrasts is the Helmert contrasts (see, e.g., Cox and Reid [2000], Appendix C). Because Q is a $v \times (v-1)$ matrix of orthonormal contrasts, the matrix $[Q, 1_v/\sqrt{v}]$ is orthogonal. It follows that $Q^T Q = I_{v-1}$ and $QQ^T = I_v - J_v/v$; thus, it is a completely symmetric system. It is easy to verify that the information matrix of a treatment proportions design $w > 0$ is $N_Q(w) = Q^T M(w) Q - Q^T M(w) J_v M(w) Q$ and, in particular, that $N_Q(\bar{w}) = v^{-1} I_{v-1}$.

Consider a system of centered treatment effects, or *centered contrasts*, which is the system of contrasts $\tau_1 - \bar{\tau}, \ldots, \tau_v - \bar{\tau}$, where $\bar{\tau}$ is the mean of the treatment effects. That is, $Q = I_v - J_v/v$, which is a $v \times v$ matrix of rank $v-1$, and thus, $Q^T\tau$ is a rank-deficient system. Note that $QQ^T = Q$ is completely symmetric, and hence, the centered contrasts constitute a completely symmetric system of contrasts. In Section 5 of Pukelsheim [1983], this system of contrasts was analyzed in great detail for a special case of model (1.2), namely, block designs, and the optimality of product designs with uniform treatment weights was obtained. The matrix $C_Q(\bar{w})$ of the uniform treatment proportions design satisfies $C_Q(\bar{w}) = (vQ^TQ)^+ = v^{-1}Q = I_v/v - J_v/v^2$.

By a *system of all pairwise comparisons*, we mean the system of $\tau_i - \tau_j$ for all $i > j$ (considered in, e.g., Bailey and Cameron [2009]). The corresponding $v \times \frac{v(v-1)}{2}$ matrix Q satisfies $QQ^T = vI_v - J_v$, and thus, the system is completely symmetric.

Corollary 2.14. *The uniform treatment proportions design $\bar{w}$ is Φ-optimal for estimating the system of orthonormal contrasts, the system of centered contrasts, and the system of all pairwise comparisons with respect to any orthogonally invariant information function Φ.*

Consider an experiment with the purpose of comparing two sets of treatments. Exact designs for such contrasts have been studied in multiple design settings, e.g., Majumdar [1986] and Jacroux [2002] performed such investigations for block experiments, Jacroux [1990] studied A- and MV-optimal designs in model (1.6), and Jacroux [1993] and Githinji and Jacroux [1998] considered the presence of trends. Without loss of generality, let the first set consist of the first g (control) treatments, $0 < g < v/2$, and let the second set be the set of the remaining $v-g$ treatments. Then, the objective is to estimate all treatment comparisons $\tau_j - \tau_i$, where $1 \leq i \leq g$ and $g+1 \leq j \leq v$, which gives rise to $Q = (-I_g \otimes 1_{v-g}, 1_g \otimes I_{v-g})^T$, where $\otimes$ denotes the Kronecker product. In Majumdar [1986], the author suggests that such a situation may arise when comparing two 'packages' of treatments or when comparing a set of test treatments with a set of standard (control) treatments. We will call such a system of treatment contrasts a *comparison of treatments with controls* or a *comparison of two groups of treatments*.

A system of contrasts of this type naturally generalizes the standard system for the comparison of $v-1$ treatments with one control, $\tau_2-\tau_1, \ldots, \tau_v-\tau_1$, where $g=1$.

Theorem 2.15. *Let $p \in [-\infty, 0]$. If $p > -\infty$, let γ_p be the unique solution to the equation*

$$(v-g-1)\gamma^{1-p} - (g-1)(1-\gamma)^{1-p} + 2\gamma - 1 = 0 \tag{2.8}$$

in the interval $(0, 1/2]$ and let $\gamma_{-\infty} = 1/2$. Then, the treatment proportions design that satisfies $w_1 = \ldots = w_g = \gamma_p/g$ and $w_{g+1} = \ldots = w_v = (1-\gamma_p)/(v-g)$ is the unique Φ_p-optimal treatment proportions design for the comparison of $(v-g)$ treatments with g controls, $0 < g < v/2$.

Proof. Note that for $Q = (-I_g \otimes 1_{v-g}, 1_g \otimes I_{v-g})^T$, we have

$$QQ^T = \begin{bmatrix} (v-g)I_g & -J_{g\times(v-g)} \\ -J_{(v-g)\times g} & gI_{v-g} \end{bmatrix}.$$

Let w be a treatment proportions design, let P_1 and P_2 be $g \times g$ and $(v-g)\times(v-g)$ permutation matrices, respectively, and let

$$\tilde{P} = \begin{bmatrix} P_1 & 0_{g\times(v-g)} \\ 0_{(v-g)\times g} & P_2 \end{bmatrix}. \tag{2.9}$$

Define $\tilde{P}w$ to be the design given by the $\tilde{P}$-permutations of the treatments. Then, $M(\tilde{P}w) = \tilde{P}M(w)\tilde{P}^T$ and $\tilde{P}^T QQ^T \tilde{P} = QQ^T$. From an argument analogous to that used in the proof of Theorem 2.13, $C_Q(\tilde{P}w)$ and $C_Q(w)$ are orthogonally similar; therefore $\Phi_p(\tilde{P}w) = \Phi_p(w)$.

Let $\tilde{w} := \frac{1}{(v-g)!g!}\sum_{\tilde{P}} \tilde{P}w$, where the sum is over all $v \times v$ permutation matrices $\tilde{P}$ of the form (2.9). Then, $\Phi_p(\tilde{w}) \geq \Phi_p(w)$. It follows that an optimal design exists in the class of designs that allocate one weight to each of the first g treatments, say γ_1 $(0 < \gamma_1 < 1/g)$, and another weight $\gamma_2 := (1-g\gamma_1)/(v-g)$ to each of the other treatments. We denote the total weight of the first g treatments as $\gamma := g\gamma_1$ and for a given γ, we denote such designs as w_γ.

The nonzero eigenvalues of $C_Q(w_\gamma)$ are inverse to the nonzero eigenvalues of $V(w_\gamma) := Q^T M^{-1}(w_\gamma)Q$, where $M(w_\gamma) = \mathrm{diag}(\gamma_1 1_g, \gamma_2 1_{v-g})$. Let $X = Q^T M^{-1/2}(w_\gamma)$. Then, the set of nonzero eigenvalues of $V(w_\gamma) = XX^T$ coincides with the set of nonzero eigenvalues of

$$X^T X = M^{-1/2}(w_\gamma)QQ^T M^{-1/2}(w_\gamma) = \begin{bmatrix} (v-g)\gamma_1^{-1}I_g & -(\gamma_1\gamma_2)^{-1/2}J_{g\times(v-g)} \\ -(\gamma_1\gamma_2)^{-1/2}J_{(v-g)\times g} & g\gamma_2^{-1}I_{v-g} \end{bmatrix}.$$

It can be verified that $X^T X$ has the following eigenvalues, listed with the corresponding eigenvectors of the form $x = (x_1^T, x_2^T)^T$, where $x_1 \in \mathbb{R}^g$ and $x_2 \in \mathbb{R}^{v-g}$: $\mu_1 = g\gamma_2^{-1}$ with

multiplicity (w.m.) $v-g-1$, $x_1 = 0_g$ and $1_{v-g}^T x_2 = 0$; $\mu_2 = (v-g)\gamma_1^{-1}$ w.m. $g-1$, $1_g^T x_1 = 0$ and $x_2 = 0_{v-g}$; $\mu_3 = (v-g)\gamma_1^{-1} + g\gamma_2^{-1}$ w.m. 1, $x_1 = -(v-g)\gamma_2^{1/2} 1_g$ and $x_2 = g\gamma_1^{1/2} 1_{v-g}$; and $\mu_4 = 0$ w.m. 1, $x_1 = \gamma_1^{1/2} 1_g$ and $x_2 = \gamma_2^{1/2} 1_{v-g}$.

Therefore, the nonzero eigenvalues of $C_Q(w_\gamma)$ are $\lambda_1 = \frac{1-g\gamma_1}{g(v-g)}$ w.m. $v-g-1$, $\lambda_2 = \frac{\gamma_1}{v-g}$ w.m. $g-1$ and $\lambda_3 = \frac{\gamma_1(1-g\gamma_1)}{v-g}$ w.m. 1. Thus, for $p \in (-\infty, 0)$, the Φ_p-optimal value of w_γ is obtained by minimizing the convex function

$$f_p(\gamma_1) = (v-g-1)\left(\frac{1-g\gamma_1}{g(v-g)}\right)^p + (g-1)\left(\frac{\gamma_1}{v-g}\right)^p + \left(\frac{\gamma_1(1-g\gamma_1)}{v-g}\right)^p$$

for $\gamma_1 \in (0, 1/g)$. Then, $f_p'(\gamma_1) = 0$ if and only if

$$-(v-g-1)(1-g\gamma_1)^{p-1} + (g-1)(g\gamma_1)^{p-1} + (1-2g\gamma_1)(g\gamma_1)^{p-1}(1-g\gamma_1)^{p-1} = 0,$$

which is equivalent to

$$-(v-g-1)(g\gamma_1)^{1-p} + (g-1)(1-g\gamma_1)^{1-p} + 1 - 2g\gamma_1 = 0.$$

Using $\gamma = g\gamma_1$, we obtain (2.8).

For $p = 0$ the unique maximum of the function $f_0(\gamma_1) = \lambda_1^{v-g-1}\lambda_2^{g-1}\lambda_3 = (1-g\gamma_1)^{v-g}\gamma_1^g/[g^{v-g-1}(v-g)^{v-1}]$ for $\gamma_1 \in (0, 1/g)$ is in $\gamma_1 = 1/v$, which means that w_γ is a uniform design. This corresponds to $\gamma = g/v$, which is obtained if we set $p = 0$ in (2.8).

The smallest nonzero eigenvalue of $C_Q(w_\gamma)$ is λ_3 and hence the $\Phi_{-\infty}$-optimal design can be obtained by maximizing

$$f_{-\infty}(\gamma) = \frac{\gamma(1-\gamma)}{g(v-g)},$$

which has a unique maximum in $\gamma = \frac{1}{2}$.

For $p \in (-\infty, 0]$, the uniqueness follows from Lemma 1.13 (as stated in Section 8.18 of Pukelsheim [1993], the General Equivalence Theorem also holds for the rank deficient Φ_p-criteria; however, instead of negative powers of $N_K(\xi^*)$, we have positive powers of $K^T M^-(\xi^*)K$ and instead of $\lambda_{\min}(N_K(\xi^*))$, we have $1/\lambda_{\max}(K^T M^-(\xi^*)K)$). Denote by w^* the Φ_p-optimal treatment proportions obtained from (2.8). The design w^* is Φ_p-optimal and the moment matrix $M(w^*)$ is non-singular, therefore there exists a *unique* generalized inverse $G = M^{-1}(w^*)$ satisfying the normality inequality. Then, $M(w)GQ = Q$ can be satisfied only by $w = w^*$, because $M(w)G$ is a diagonal matrix and no row of Q is equal to zero.

For $p = -\infty$ let $w > 0$. Then, the matrix $V(w) = Q^T M^{-1}(w) Q$ can be expressed as $V(w) = A_1(w) + A_2(w)$, where $A_1(w)$ is a block diagonal matrix with the blocks on the diagonal given by $w_1^{-1}J_{v-g}, \ldots, w_g^{-1}J_{v-g}$ and $A_2(w) = J_g \otimes \mathrm{diag}(w_{g+1}^{-1}, \ldots, w_v^{-1})$.

Let us partition $u \in \mathbb{R}^{g(v-g)}$ as $u^T = (u_1^T, \ldots, u_g^T)$, where $u_i \in \mathbb{R}^{v-g}$ for all i. Then, the largest eigenvalue $\mu_{\max}$ of $V(w)$ satisfies

$$\mu_{\max} = \max_{\|u\|=1} u^T V(w) u = \max_{\|u\|=1} \Big(\sum_{i=1}^{g} w_i^{-1} (1_{v-g}^T u_i)^2 + \sum_{i=1}^{g} \sum_{j=1}^{g} u_i^T \mathrm{diag}(w_{g+1}^{-1}, \ldots, w_v^{-1}) u_j \Big).$$

For the particular choice of $u = 1_{g(v-g)}/\sqrt{g(v-g)}$, we obtain

$$\begin{aligned} \mu_{\max} &\geq \sum_{i=1}^{g} w_i^{-1} \Big(\frac{v-g}{\sqrt{g(v-g)}} \Big)^2 + \sum_{i=1}^{g} \sum_{j=1}^{g} \sum_{k=g+1}^{v} \Big(\frac{1}{\sqrt{g(v-g)}} \Big)^2 w_k^{-1} \\ &= \frac{v-g}{g} \sum_{i=1}^{g} w_i^{-1} + \frac{g}{v-g} \sum_{i=g+1}^{v} w_i^{-1}. \end{aligned}$$

If w does not satisfy $w_1 = \ldots = w_g$ and $w_{g+1} = \ldots = w_v$, then define $\tilde{w}$ such that $\tilde{w}_1 = \ldots = \tilde{w}_g = \sum_{i=1}^{g} w_i/g$ and $\tilde{w}_{g+1} = \ldots = \tilde{w}_v = \sum_{i=g+1}^{v} w_i/(v-g)$. Then, from the first part of the proof, we have $\mu_{\max}(\tilde{w}) = \mu_3(\tilde{w}) = (v-g)\tilde{w}_1^{-1} + g\tilde{w}_{g+1}^{-1}$. From the inequality of arithmetic and harmonic means, it follows that $\frac{v-g}{g} \sum_{i=1}^{g} w_i^{-1} \geq (v-g)\tilde{w}_1^{-1}$ and $\frac{g}{v-g} \sum_{i=g+1}^{v} w_i^{-1} \geq g\tilde{w}_{g+1}^{-1}$, and at least one of these inequalities is strict. Hence, $\mu_{\max}(w) > \mu_{\max}(\tilde{w})$ and thus any $\Phi_{-\infty}$-optimal w must satisfy $w_1 = \ldots = w_g$ and $w_{g+1} = \ldots = w_v$. From the first part of the proof it follows that the only $\Phi_{-\infty}$-optimal design that satisfies these conditions is $w^* = (1_g^T/(2g), 1_{v-g}^T/(2(v-g)))^T$.

If $g = 1$, the matrix Q is of full rank $v-1$, and the eigenvalues of $V(w)$ are inverses of the eigenvalues of $N_Q(w)$. Therefore, the results hold also for $g = 1$. □

We note that for any $p \in (-\infty, 0]$ and $v > 2$, there exists a unique solution γ_p to equation (2.8) in the interval $(0, 1/2)$ because the function $F(\gamma) = (v-g-1)\gamma^{1-p} - (g-1)(1-\gamma)^{1-p} + 2\gamma - 1$ is an increasing function for $\gamma \in (0, 1/2)$ that satisfies $F(0) < 0$ and $F(1/2) > 0$.

The obtained optimal treatment proportions, which depend on the choice of the criterion Φ_p, generalize the results obtained for block designs by Giovagnoli and Wynn [1985] for comparison with one control. The optimal proportions given by Theorem 2.15 are characterized by a single value, γ_p, which is the total weight of the first g treatments. In particular, for D-optimality, $\gamma_0 = g/v$ and the optimal proportions are uniform; for A-optimality, $\gamma_{-1} = \frac{\sqrt{g(v-g)}-g}{v-2g}$, which lies in $(g/v, 1/2)$; and for E-optimality, $\gamma_{-\infty} = 1/2$, i.e., half of the total weight is allocated to each of the two sets of treatments.

For comparisons with controls, it is also common to use the criterion of MV-optimality, which minimizes the maximum variance of the contrasts of interest. It turns out that the MV- and A-optimal treatment proportions are identical. It follows that the sets of MV- and A-optimal nuisance-resistant designs are also identical.

Theorem 2.16. *Let* $\gamma = \frac{\sqrt{g(v-g)}-g}{v-2g}$, *and let* $w_1 = \ldots = w_g = \gamma/g$ *and* $w_{g+1} = \ldots = w_v = (1-\gamma)/(v-g)$. *Then,* w *is MV-optimal for the comparison of* $(v-g)$ *treatments with* g *controls,* $0 < g < v/2$.

Proof. This proof will closely follow the proof of Theorem 2.15. The covariance matrix of the least square estimator is proportional to $V(w) = Q^T M^{-1}(w) Q$. Note that the MV-optimality criterion Ψ_{MV} is permutationally invariant, because it depends only on the diagonal of the variance matrix.

Let w be a treatment proportions design and let $\tilde{P}$, $\tilde{w}$, γ_1, γ_2, γ and w_γ be defined as in the proof of Theorem 2.15. Then, $Q^T\tilde{P} = BQ^T$, where $B = P_1 \otimes P_2$, which is a permutation matrix. Thus, $V(\tilde{P}w) = BV(w)B^T$, $\Psi_{MV}(\tilde{P}w) = \Psi_{MV}(w)$ and $\Psi_{MV}(\tilde{w}) \leq \Psi_{MV}(w)$. It follows that an optimal design exists in the class of designs w_γ.

We have $V(w_\gamma) = \gamma_1^{-1} I_g \otimes J_{v-g} + \gamma_2^{-1} J_g \otimes I_{v-g}$ and all of its diagonal elements are $\gamma_1^{-1} + \gamma_2^{-1}$. Thus, the optimal γ_1 may be obtained by minimizing

$$f_{MV}(\gamma_1) = \gamma_1^{-1} + \frac{v-g}{1-g\gamma_1},$$

which has minimum in $\gamma_1^* = \frac{\sqrt{g(v-g)}-g}{g(v-2g)}$; therefore, $\gamma^* = \frac{\sqrt{g(v-g)}-g}{v-2g}$. □

Once optimal treatment proportions have been calculated, optimal nuisance-resistant (balanced, product) designs can be constructed that retain the same information matrices as those of their corresponding treatment proportions designs. Note that the matrix Q for a completely symmetric system of contrasts and for a comparison with controls is of rank $v-1$; therefore, for such systems, the balanced designs and nuisance-resistant designs coincide.

2.3 Examples

2.3.1 Trend-resistant designs

Let us consider a model in which we perform the trials in a time sequence, with exactly one trial at each time, and the nuisance effect is the effect of some time trend:

$$Y_i = \tau_{u(i)} + h_1(t(i))\theta_1 + \ldots + h_d(t(i))\theta_d + \varepsilon_i, i \in \{1, \ldots, n\}, \tag{2.10}$$

where $u(i) \in \{1, \ldots, v\}$ represents the chosen treatment for trial i and $t(i) \in \{1, \ldots, n\}$ denotes at which time that treatment is to be applied. The functions $h_1, \ldots, h_d : \mathbb{R} \to \mathbb{R}$ are the regressors of the time trend, which are often chosen to be polynomials of degrees $0, \ldots, d-1$, respectively.

The interest in designs that perform well under model (2.10) dates back to the mid-20th century, e.g., to Cox [1951]. The reader may find surveys of the literature

on trend-resistant experimental designs in papers by Cheng [1990] and Atkinson and Donev [1996]. The research focus is usually on the combinatorial construction of exact designs that are *orthogonal* to the time trend (or *trend free*). These are the designs for which no information is lost because of the time trend (see, e.g., Jacroux et al. [1997], Bailey et al. [1992]). Usually, the focus is on all parameters of interest, not on a system of contrasts Q, resulting in the condition that a design is trend free with respect to h_k in model (2.10) if $\sum_t \xi(u,t)h_k(t) = 0$ (see, e.g., Cox [1951]). Such trend-free designs satisfy $M_{12}(\xi) = 0$ and, thus, $M_\tau(\xi) = M_{11}(\xi)$.

Note that orthogonal designs satisfy (2.3), and thus, they are balanced. However, because our objective is to estimate a set of treatment contrasts Q, the stringent conditions for orthogonality need not hold for the information to be retained. If ξ is resistant to nuisance effects, then the equality $M_\tau(\xi) = M_{11}(\xi)$ in general does not hold, but such a ξ satisfies $N_K(\xi) = N_Q(w)$, i.e., designs that are resistant to nuisance effects eliminate the effects of the time trend. We remark that when $\sum_t h(t) = 0$, the conditions for orthogonality and the conditions for balancedness coincide.

We will examine a model with a trigonometric time trend of degree $D \in \mathbb{N}$, which can be used to model, for instance, circadian rhythms (cf. Kitsos et al. [1988]). For simplicity, let $\phi_n = 2\pi/n$, and let us consider the model

$$\begin{aligned} Y_t = \tau_{u(t)} + \theta_0 + \theta_1 \cos(\phi_n t) + \theta_2 \sin(\phi_n t) + \dots \\ + \theta_{2D-1} \cos(D\phi_n t) + \theta_{2D} \sin(D\phi_n t) + \varepsilon_t, \end{aligned} \tag{2.11}$$

where $t = 1, 2, \dots, n$.

An exact design ξ is represented by a sequence of treatments that determines which treatments are to be applied at which times. Note that the regression functions satisfy $\sum_t h_k(t) = 0$ for $k > 0$, i.e., the concepts of orthogonal and balanced designs coincide for this model.

The following lemma provides a method of constructing balanced designs in the model with trigonometric time trend described in (2.11).

Lemma 2.17. *Let $l, a \in \mathbb{N}$, let ξ_p be an exact design of size l and let $\xi = \xi_p\xi_p...\xi_p$ be the exact design of size $n = la$ formed through a-fold replication of ξ_p. Assume that $b \in \mathbb{N}$ is not an integer multiple of a. Then, ξ is balanced for the nuisance regressors of the form $\cos(b\phi_n t)$ and $\sin(b\phi_n t)$, $t = 1, ..., n$.*

Proof. Let $u \in \{1, ..., v\}$. Using the fact that $\xi(u, k + lj) = \xi_p(u, k)$ for all $k \in \{1, ..., l\}$ and $j \in \{0, ..., a - 1\}$, we obtain

$$\sum_{t=1}^{n} \xi(u,t) \cos(b\phi_n t) + \mathrm{i} \sum_{t=1}^{n} \xi(u,t) \sin(b\phi_n t) = \sum_{t=1}^{n} \xi(u,t) e^{b\phi_n t \mathrm{i}}$$

$$= \sum_{j=0}^{a-1} \sum_{k=1}^{l} \xi(u, k+lj) e^{b\phi_n (k+lj)\mathrm{i}} = \left(\sum_{k=1}^{l} \xi_p(u,k) e^{b\phi_n k\mathrm{i}} \right) \left(\sum_{j=0}^{a-1} e^{(b\phi_n l\mathrm{i})j} \right).$$

Note that if b is not an integer multiple of a, then $b\phi_n l = 2\pi(b/a)$ is not an integer multiple of 2π, which implies that $e^{b\phi_n l\mathrm{i}} \neq 1$. In that case

$$\sum_{j=0}^{a-1} e^{(b\phi_n l\mathrm{i})j} = \frac{1 - e^{b2\pi\mathrm{i}}}{1 - e^{b\phi_n l\mathrm{i}}} = 0.$$

□

From Theorem 2.8 and Lemma 2.17 it follows that by repeating a sequence of treatments with Φ-optimal treatment weights, we may obtain a Φ-optimal design for model (2.11) of high degree.

Proposition 2.18. *Let Φ be an information function. Let $l \in \mathbb{N}$, and let ξ_p be an exact design of size l with Φ-optimal treatment proportions for estimating contrasts $Q^T\tau$. Let $a \in \mathbb{N}$. Then, the exact design $\xi = \xi_p\xi_p...\xi_p$ of size $n = la$ that is formed through a-fold replication of ξ_p is Φ-optimal for all trigonometric models (2.11) of degrees $D < a$.*

It is, in fact, possible to show that the design ξ discussed in Proposition 2.18 is Φ-optimal for models of the type described by (2.11) of any degree, but they cannot include the terms $\cos(b\phi_n t)$ and $\sin(b\phi_n t)$, where b is an integer multiple of a.

Below, we demonstrate the results given by Proposition 2.18 in a simple example.

Example 2.19. Consider the model

$$Y_t = \tau_t + \theta_0 + \theta_1 \sin(r) + \theta_2 \cos(r) + \theta_3 \sin(2r) + \theta_4 \cos(2r) + \theta_5 \sin(3r) + \theta_6 \cos(3r) + \varepsilon_t,$$

where $r = \frac{2\pi}{n}t$ and $t = 1, 2, \ldots, n$. Let $n = 16$, $v = 3$, and $\xi_p = 2113$. Then, the design $\xi_1^* = 2113, 2113, 2113, 2113$ is E-optimal for comparison with one control. Let $n = 12$, $v = 3$, and $\xi_q = 321$. Then, the design $\xi_2^* = 321, 321, 321, 321$ is D-optimal for comparison with one control. Moreover, let Φ be an orthogonally invariant information function. Then, ξ_2^* is Φ-optimal for estimating any completely symmetric system of contrasts.

Another class of sequential designs consists of covariate-adaptive designs for randomized clinical trials, which are intended to sequentially balance patient covariates (see, e.g., Pocock and Simon [1975], Atkinson [1999]). Such designs provide a practical method for the randomized sequential allocation of the treatments to the patients to achieve high efficiency of the design or to balance the covariates across the treatments. By contrast, the propositions in this chapter simply state that balanced designs with optimal treatment proportions are optimal. Although covariate-adaptive designs usually attempt to achieve uniform treatment proportions, we note that the optimal treatment proportions need not be uniform (especially in clinical trials, in which a control is present; see Theorem 2.15).

2.3.2 Block designs and row-column designs

Consider an experiment in which the treatment units are arranged in b blocks. As usual, for each of the N treatment units, we choose one of v treatments. The response is then determined by the treatment effects and the block effects. We assume that the treatment and block effects do not interact, i.e, we obtain an additive blocking experiment:

$$Y_i = \mu + \tau_{u(i)} + \eta_{t(i)} + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.12}$$

where $u(i) \in \{1, \ldots, v\}$ and $t(i) \in \{1, \ldots, b\}$. Designs for blocking experiments are called *block designs*. There is a large amount of literature on this topic; in particular, the papers that consider treatment contrasts in block designs include Majumdar and Notz [1983] and Pukelsheim [1983].

Note that model (2.12) may be expressed as a special case of model (1.2), where $\mathfrak{T} = \{1, \ldots, b\}$, $n = b$, $\theta = (\mu, \eta_1, \ldots, \eta_b)^T$, and $h^T(t) = (1, e_t^T) \in \mathbb{R}^{b+1}$. For block designs, the balance conditions (2.3) become $\xi(1,t)/w_1 = \ldots = \xi(v,t)/w_v$ for all $t \in \{1, \ldots, b\}$, which leads to a product design $\xi = w \otimes \alpha$. That is, all balanced designs in model (2.12) are product designs. Therefore, for a system of contrasts of rank $v-1$ and a strictly concave criterion Φ, it follows from Theorem 2.10 that all Φ-optimal designs are product designs. Note that in general, balanced incomplete block designs and balanced treatment incomplete block designs (see, e.g., Majumdar and Notz [1983]) are not balanced in the sense of the conditions given in (2.3).

Block designs are often used to eliminate heterogeneity in one direction, e.g., heterogeneity caused by a nuisance time trend. If the position of a unit within a block also affects the response, or in general, if heterogeneity needs to be eliminated in two directions, then we may use *row-column designs* (see Jacroux [1982]). In such a design, N experimental units are arranged in b_1 rows and b_2 columns. The mean response is determined by the sum of the treatment, row and column effects, modeled as

$$Y_i = \tau_{u(i)} + \eta_{k(i)} + \varphi_{l(i)} + \varepsilon_i,\ i = 1, \ldots, N, \tag{2.13}$$

where $u(i) \in \{1, \ldots, v\}$; $k(i) \in \{1, \ldots, b_1\}$ and $l(i) \in \{1, \ldots, b_2\}$ represent the row and column, respectively, that have been chosen for the ith trial; and $\eta_{k(i)}$ and $\varphi_{l(i)}$ are the row and column effects.

This model can also be expressed as a special case of model (1.2), where $\mathfrak{T} = \{1, \ldots, b_1\} \times \{1, \ldots, b_2\}$, $n = b_1 b_2$, $\theta = (\eta_1, \ldots, \eta_{b_1}, \phi_1, \ldots, \phi_{b_2})^T$, and $h(k,l) = (e_k^T, e_l^T)^T$. The balance conditions for the row-column model become $w_1^{-1} \sum_l \xi(1,k,l) = \ldots = w_v^{-1} \sum_l \xi(v,k,l)$ for all $k = 1, \ldots, b_1$ and $w_1^{-1} \sum_k \xi(1,k,l) = \ldots = w_v^{-1} \sum_k \xi(v,k,l)$ for all $l = 1, \ldots, b_2$. That is, for any row (column), the ratio of the total weights of any two treatments i and j in that row (column) is given by the ratio of the treatment weights w_i/w_j. In other words, for the design ξ to be balanced (and hence optimal,

if ξ achieves the optimal treatment weights), the functions $\xi(u, \cdot, \cdot)/w_u$ must have the same row and column marginals for all $u = 1, \ldots, v$.

By combining models (2.12) and (2.10), a blocking experiment in the presence of a nuisance time trend is obtained (see, e.g., Bradley and Yeh [1980], Jacroux et al. [1997]), which we will examine further in Example 2.22.

2.4 Constructing efficient exact designs

By constructing product designs with optimal treatment proportions and calculating their criterial values (or by analytically deriving optimal criterial values), we may assess the quality of exact designs. More precisely, we can compute a lower bound on the efficiency of any given exact design ξ by calculating its approximate efficiency with respect to the positively homogeneous criterion Φ, $\text{eff}(\xi) = \frac{\Phi(\xi)}{\Phi(\xi^*)}$, where ξ^* is a Φ-optimal approximate design.

We will focus on exact designs of experiments in which exactly one trial is to be performed under each nuisance condition. The problem of finding such optimal designs is, in general, a difficult discrete optimization problem (see, e.g., Atkinson and Donev [1996], Harman and Sagnol [2015]).

In Section 2.2, we showed that in model (1.2), Φ-optimal designs can be constructed in two steps:

1. Find Φ-optimal treatment proportions for estimating $Q^T\tau$.

2. Arrange these proportions with respect to the nuisance conditions such that the resulting design is nuisance resistant.

Naturally, the problem of obtaining Φ-optimal nuisance-resistant designs is non-linear. In particluar, the problem of finding the Φ-optimal treatment proportions (2.5) in step 1 is a non-linear optimization problem. However, step 1 is a fairly simple convex programming problem, with only v variables, which can often even be solved analytically. Then, step 2 is a much larger problem, with vn variables and multiple constraints, but both the conditions for resistance to nuisance effects (2.2) and the conditions for achieving the given Φ-optimal treatment weights are linear. Therefore, step 2 is a linear programming problem.

Once the Φ-optimal treatment weights w^* have been obtained, a Φ-optimal design can be constructed as a product $w^* \otimes \alpha$ for any nuisance conditions design α, which seems to make step 2 redundant. However, in general, it is difficult to construct exact designs from product designs because of their regular structures and large support sizes, as noted in Section 2.1. For example, the efficient rounding method presented in Chapter 12 by Pukelsheim [1993] requires for the number of support points of the

approximate design to be at most the number of trials. Clearly, this is not satisfied by product trend-resistant designs if we require one treatment to be chosen in each time moment. We will show that nuisance-resistant designs with small support sizes, which can therefore be converted into highly efficient exact designs, can be obtained through linear programming in step 2.

For simplicity, consider the most common case of $\operatorname{rank}(Q) = v-1$, i.e., in the second step, we wish to construct a balanced design with the given Φ-optimal proportions obtained in step 1. Then, step 2 can be achieved through linear programming by solving the problem

$$\min\{c^T x | Ax = b, x \geq 0\}, \tag{2.14}$$

where $x \in \mathbb{R}^{vn}$ represents a design ξ in vector form and $Ax = b$ defines sufficient conditions for optimality. Because we, in fact, need to obtain any design that satisfies $Ax = b, x \geq 0$, we seek a *feasible* solution to (2.14) and thus are free to choose the vector c of the coefficients of the objective function. Let us denote the set of all feasible solutions to (2.14) by $\mathcal{P}$.

The system of equalities $Ax = b$ consists of

(i) v equalities $\sum_t \xi(u,t) = w_u$, $u = 1, \ldots, v$, i.e., ξ achieves the Φ-optimal treatment weights;

(ii) $d(v-1)$ equalities $w_1^{-1} \sum_t \xi(1,t)h(t) = w_u^{-1} \sum_t \xi(u,t)h(t)$, $u = 2, \ldots, v$, i.e., ξ is a balanced design; and

(iii) n equalities $\sum_u \xi(u,t) = 1/n$, $t = 1, \ldots, n$, i.e., exactly one trial is performed under each nuisance condition.

To obtain an optimal design with a small support size, it is beneficial to employ the simplex method of linear programming, the output of which is an optimal design ξ^* that represents a vertex in $\mathcal{P}$, the set of feasible solutions to (2.14).

Proposition 2.20. *Let ξ represent a vertex in $\mathcal{P}$. Then, ξ contains at most $v + (v-1)k + n - 1$ support points, where k is the affine dimension of the set $\{h(t)\}_{t \in \mathfrak{T}}$.*

Proof. It is well known that a point x is a vertex of the set $\{x | Ax = b, x \geq 0\}$ if and only if the system $\{A_j | x_j > 0\}$, where A_j is the jth column of A, is of full rank.

The matrix A consists of $v + (v-1)d + n$ rows, but these rows are linearly dependent. Let k be the affine dimension of $\{h(t)\}_{t \in \mathfrak{T}}$ and, without loss of generality, let $\mathfrak{T} = \{1, \ldots, n\}$. Then, the matrix $[h(2) - h(1), \ldots, h(n) - h(1)]$ is of rank k and thus its row space has dimension k. That is, without the loss of generality, we obtain that $h_i(t) - h_i(1) = \sum_{j=1}^{k} c_j^{(i)} \big(h_j(t) - h_j(1)\big)$ for some $c_1^{(i)}, \ldots, c_k^{(i)} \in \mathbb{R}$, for $i > k$ and $t \in \{1, \ldots, n\}$ (for $t = 1$, we formally obtain $0 = 0$). Let $u \in \{1, \ldots, v\}$. Then, if (ii)

is satisfied for $h_1, \ldots, h_k$, for all $i > k$ and $u \in \{1, ..., v\}$ we have:

$$\begin{aligned}
\frac{1}{w_1}\sum_t \xi(1,t)h_i(t) &= \frac{1}{w_1}\Big(h_i(1) - \sum_{j=1}^k c_j^{(i)} h_j(1)\Big)\sum_t \xi(1,t) + \sum_{j=1}^k \frac{c_j^{(i)}}{w_1}\sum_t \xi(1,t)h_j(t) \\
&= h_i(1) - \sum_{j=1}^k c_j^{(i)} h_j(1) + \sum_{j=1}^k \frac{c_j^{(i)}}{w_u}\sum_t \xi(u,t)h_j(t) \\
&= \frac{1}{w_u}\sum_t \xi(u,t)\Big(h_i(1) - \sum_{j=1}^k c_j^{(i)} h_j(1)\Big) + \sum_{j=1}^k \frac{c_j^{(i)}}{w_u}\sum_t \xi(u,t)h_j(t) \\
&= \frac{1}{w_u}\sum_t \xi(u,t)\Big[\Big(h_i(1) - \sum_{j=1}^k c_j^{(i)} h_j(1)\Big) + \sum_{j=1}^k c_j^{(i)} h_j(t)\Big] \\
&= \frac{1}{w_u}\sum_t \xi(u,t)h_i(t),
\end{aligned}$$

where the second and the third equalities also use (i). It follows that (ii) provides at most $k(v-1)$ additional linearly independent equalities.

If ξ satisfies (i), then $\sum_{u,t} \xi(u,t) = 1$ holds. Thus, if ξ satisfies (iii) for $t = 1, \ldots, n-1$, we have $1 = \sum_{t=1}^{n-1}\sum_u \xi(u,t) + \sum_u \xi(u,n) = \frac{n-1}{n} + \sum_u \xi(u,n)$ and therefore (iii) also holds for $t = n$. That is, (iii) provides only $n-1$ additional linearly independent equalities. Hence, A is of rank at most $v + (v-1)k + n - 1$ and a vertex x contains at most $v + (v-1)k + n - 1$ support points. □

From Proposition 2.20, it follows that by employing the simplex method, we can obtain a Φ-optimal design ξ^* with at most $(v-1)(k+1) + n$ support points. As a special case, when a constant term θ_0 is present in the time trend, it may be ignored in the conditions considered in (2.14) because it does not increase the affine dimension of $\{h(t)\}_{t\in\mathfrak{T}}$; thus, the upper bound on the number of support points is reduced by $v-1$.

Suppose that a ξ satisfying (iii) has a support size of n, the number of nuisance conditions. Then, ξ uniquely determines an exact design of size n. The number of support points in a design obtained using the simplex method is only slightly larger than n; it exceeds this minimum support size by $(v-1)(k+1)$. Note that this number of additional support points does not depend on n; thus, even with an increasing number of nuisance conditions, it remains small. For instance, if $v = 5$ and $n = 150$ and if we consider a nuisance time trend modeled by discrete orthogonal polynomials of degrees of up to 3 (i.e., $k = 3$), then the support size of an optimal design obtained using the simplex method is 166, whereas the minimum possible support size is 150. By comparison, the support size of a product design can be as large as 750.

We note that by using the Carathéodory Theorem (cf. Theorem 8.2. in Pukelsheim [1993]), it is possible to obtain results similar to Proposition 2.20; however, the Carathéodory Theorem does not provide an actual method of constructing a design with a small support size, unlike the simplex method.

From an optimal approximate design with a small support size, an efficient exact design can be constructed by rounding or, often, even by completely enumerating all treatments in a small number of nuisance conditions.

Example 2.21. Consider an experiment in which trials are performed in a time sequence

$$Y_i = \tau_{u(i)} + \theta_0 + \theta_1 h_1(t(i)) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $h_1(t) = e^t / \sum_j e^j$ represents an exponential time trend (e.g., the decay of wool during a wool processing experiment, as suggested by Atkinson and Donev [1996]). Assume that $v = 5$ and that the objective is to find an A-optimal design for comparing 3 treatments with 2 controls, i.e., $\gamma_{-1} = \sqrt{6} - 2 \approx 0.45$. We will generate optimal balanced designs with a small support size by employing the simplex method of linear programming (the *linprog* function of MATLAB, using the simplex algorithm). Note that the interior point (default) algorithm of MATLAB's *linprog* function also tends to provide optimal vertex solutions or optimal solutions with small support sizes; thus, it can be used instead of the simplex algorithm.

First, we consider a small n to demonstrate the proposed method in a simple setting. Let $n = 8$. Because θ_0 is the constant term, it follows from Proposition 2.20 that there are at most $n - 1 + v + (v - 1)D = 16$ linearly independent rows of A in (2.14), where $D = d - 1 = 1$. We remark that although we are free to choose the vector c in the linear programming problem, the support size of the design obtained using the simplex method does not appear to depend on the choice of c. Therefore, we choose each of the elements of c randomly from a uniform distribution on the interval $(0, 1)$.

We obtain a design ξ_1 that has a support size of 16 (where the minimum support size is 8) and is 'fixed' at 4 times (i.e., at each of these times, ξ_1 has only one nonzero element); see Table 3. The support size corresponds to the bounding value of 16 given by 2.20.

$u\backslash t$	1	2	3	4	5	6	7	8
1	0	0.1250	0.0560	0	0	0	0	0.0437
2	0	0	0.0690	0	0	0.1250	0.0059	0.0249
3	0.0245	0	0	0.1250	0	0	0	0.0340
4	0.0154	0	0	0	0.1250	0	0.0207	0.0224
5	0.0851	0	0	0	0	0	0.0984	0

Table 3: An A-optimal balanced approximate design obtained using the simplex method. The first two treatments are controls. u - treatment number, t - time

Based on a complete enumeration of the possible treatment combinations at the remaining 4 non-fixed times, we choose the design $\hat{\xi}_1 = 51234215$ that maximizes

the criterial value. For $v = 5$ and $n = 8$, it is possible to find the A-optimal exact design, $\xi^* = 41253214$, through complete enumeration. It turns out that design $\hat{\xi}_1$ has efficiency 1 relative to ξ^*, i.e., it is optimal; in fact, $\hat{\xi}_1$ can be obtained by relabeling treatments 3, 4, and 5 in ξ^*. Note that in general, the proposed heuristic does not produce optimal exact designs.

For $n = 100$ and the same model assumptions, we obtain a design ξ_2 with a support size of 108. That is, the number of support points of ξ_2 again exceeds the minimum support size by 8; moreover, ξ_2 contains only 5 non-fixed times. Therefore, even for $n = 100$, an efficient exact design may be constructed through a complete enumeration of the possible treatments at the non-fixed times. The resulting design $\hat{\xi}_2$ assigns 23, 22, 19, 18, and 18 trials to treatments $1, \ldots, 5$, respectively, which correspond to the A-optimal treatment weights given by $\gamma_{-1} \approx 0.45$. Moreover, $\hat{\xi}_2$ has an approximate efficiency of 0.994; its efficiency relative to the optimal exact design would be even higher, but for a problem of this size, it is unfeasible to compute an optimal exact design through complete enumeration.

Example 2.22. Consider the model described by Bradley and Yeh [1980]. We have a blocking experiment with b blocks, each of size l, where the response in a trial is also influenced by a common trend effect that is determined by the position of the unit within the block. In each block, exactly one trial is performed in each position. Moreover, the trend effect in position $t_2(i)$ does not depend on the particular block $t_1(i)$. We have

$$Y_i = \tau_{u(i)} + \eta_{t_1(i)} + p^T(t_2(i))\varphi + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.15}$$

where $t_1(i) \in \{1, \ldots, b\}$ is the block in which trial i is performed, η_{t_1} is the effect of the t_1th block, $t_2 \in \{1, \ldots, l\}$ denotes the position of the unit within the block, $n = bl$, φ is a $(D+1) \times 1$ vector of nuisance trend effects, and $p : \mathbb{R} \to \mathbb{R}^{D+1}$ is a regression function of the nuisance trend.

Let us assume that $v = 3$, $b = 3$ and $l = 8$, meaning that $n = 3 \times 8 = 24$, and that the time trend is modeled using discrete orthogonal polynomials p_0, p_1, and p_2 of degrees 0, 1, and 2, respectively, i.e., $D = 2$. We wish to find an E-optimal design for comparing the treatments with one control. The optimal weight of the first treatment is $\gamma^* = 1/2$, and the optimal weights of the other two are $1/4$.

Conditions (ii) can be expressed as two sets of conditions: (ii.a) $(v-1)b$ conditions $w_1^{-1} \sum_{t_2} \xi(1, t_1, t_2) = w_u^{-1} \sum_{t_2} \xi(u, t_1, t_2)$ for $u = 2, \ldots, v$ and $t_1 = 1, \ldots, b$ and (ii.b) $(v-1)(D+1)$ conditions $w_1^{-1} \sum_{t_1,t_2} \xi(1, t_1, t_2)p(t_2) = w_u^{-1} \sum_{t_1,t_2} \xi(u, t_1, t_2)p(t_2)$, $u = 2, \ldots, v$. By summing (ii.a) over all t_1 and using the fact that $\sum_t \xi(u, t) = 1$, we obtain (i), which reduces the number of linearly independent rows in A by v. Similarly to Proposition 2.20, it follows from (i) that there are at most $(v-1)(b+r)+n-1$ linearly independent rows in A, where r is the affine dimension of the set $\{p(t_2)\}_{t_2}$. Because

$p_0\varphi_0$ is the constant term, the number of linearly independent rows in A is at most $(v-1)(b+D)+n-1=33$. The minimum number of support points is $n=24$.

Using the simplex method, we obtain an E-optimal balanced approximate design ξ^*; see Table 4. The design ξ^* has a support size of 30, which exceeds the minimum support size by 6, and is fixed at 18 of the 24 positions.

block	$u\backslash t$	1	2	3	4	5	6	7	8
1	1	0.0417	0	0.0417	0	0.0417	0	0.0417	0
	2	0	0.0417	0	0	0	0.0417	0	0
	3	0	0	0	0.0417	0	0	0	0.0417
2	1	0.0417	0.0417	0	0.0324	0	0.0417	0.0093	0
	2	0	0	0	0	0.0417	0	0	0.0417
	3	0	0	0.0417	0.0093	0	0	0.0324	0
3	1	0.0046	0	0.0083	0	0.0417	0.0417	0.0370	0.0333
	2	0	0.0417	0.0333	0	0	0	0	0.0083
	3	0.0370	0	0	0.0417	0	0	0.0046	0

Table 4: An E-optimal balanced approximate design obtained using the simplex method for an experiment with 3 blocks, each of size 8, and a common trend effect. u - treatment number, t - time

Through complete enumeration of the possible treatments in the 8 non-fixed positions, we obtain an exact design $\xi : b_1 = 12131213, b_2 = 11312132, b_3 = 32231111$, where each sequence b_j determines the treatments in block j and their positions. Using Theorem 2.15, we find that ξ has efficiency of 0.999 with respect to the optimal approximate design.

3 Designs in dose-escalation studies

3.1 Introduction

As noted in the previous chapter, the blocking experiment is a special case of model (1.2). In this chapter, we examine optimal approximate designs in dose-escalation studies; and following Bailey [2009a], we model these studies as blocking experiments. Although the results from Chapter 2 cannot be applied directly, due to the design constraints that are inherent in the dose-escalation studies, some of these results will prove to be helpful. This chapter is based on paper Rosa and Harman [2017].

A dose-escalation study for a new drug consists of a series of trials, in which the drug is administered to cohorts of individuals, one cohort at a time. The experimenters select a finite number of increasing doses of the drug that are to be tested, and each subject is given either one of the doses or the placebo. In particular, the subjects in the first cohort may be given only the lowest dose or the placebo, the subjects in the second cohort only the two lowest doses or the placebo and so on. If an adverse effect is observed, the experiment is stopped. Therefore, no subjects are then exposed to higher doses. These studies are often performed when a new drug is tested on humans for the first time, and thus the safety of the subjects is the main concern; particularly if the drug is administered to healthy volunteers – hence the slowly escalating doses.

Bailey [2009a], motivated by the failure of the Te Genero trial (see Senn et al. [2007]), considered designs for dose-escalation studies with quantitative responses. In that paper, the dose-escalation studies were modeled as blocking experiments with design constraints, where the blocks represented the cohorts and the treatments were the placebo and the drug doses. The considered objective was to establish the safety and tolerability of the studied drug doses, which is reasonable in clinical trials with healthy volunteers. Indeed, the declared primary objective of the Te Genero trial was to 'establish the safety and tolerability' (see Senn et al. [2007]). The measured effects of the drug on the subjects were assumed to be quantitative, e.g., the peak flow rate in the lungs or the number of white cells in the blood. Some general principles that reduce variance were presented, as well as a wide range of well-performing designs for estimating a system of all pairwise comparisons of treatments. Haines and Clark [2014] provided some numerical methods for constructing A-, MV-, D- and E-optimal designs in this setting.

In this chapter, we adopt the model considered by Bailey [2009a], and we study the optimal approximate designs. The approximate, or continuous, designs determine the proportions of all subjects that are allocated to each combination of a cohort and a treatment. Note that when a total number of subjects is selected, optimal or efficient exact designs can generally be constructed from the optimal approximate designs, e.g., see Chapter 12 in Pukelsheim [1993].

Senn [2009] has argued that in each cohort, the comparison of the latest dose with the placebo is relevant and proposed a corresponding latest variance (LV) optimality criterion. This relevance follows from the fact that the crucial decision of whether to proceed to the next cohort or to stop the experiment is based on the effect of the latest dose relative to the placebo. Following the argued relevance of the comparisons with the placebo, we consider the problem of comparing doses of the drug with the placebo. Therefore, the system of contrasts studied here is in fact a comparison of test treatments (drug doses) with a control (placebo). In contrast, Bailey [2009a] studied the set of all pairwise comparisons of treatments.

There is a large amount of literature on comparisons with a control in blocking experiments, beginning with Bechhofer and Tamhane [1981] (e.g., see Majumdar and Notz [1983], Majumdar [1996] and Kunert et al. [2010]). However, these standard results are usually not applicable in the model considered in this chapter due to the dose-escalation constraints.

We consider the E-, LV- and MV-optimality criteria, and we obtain optimal designs for these criteria. In particular, the *Senn designs* (see Bailey [2009a], Senn [1997]) stand out because they are optimal with respect to all of the aforementioned criteria. These designs assign in each cohort half of the subjects to the placebo and the other half to the latest dose. Moreover, we obtain a complete characterization of the class of all E-optimal designs. This allows for selecting the best E-optimal design with respect to a secondary criterion chosen by the experimenter.

3.2 The model

We consider the model formulated by Bailey [2009a]. In this model, a new drug is to be tested on N individuals split into t cohorts. Each individual is given a treatment i, which is either the placebo (for $i = 0$) or one of $n \geq 2$ increasing doses of the drug (for $i = 1, \ldots, n$), i.e., $\text{dose}_1 < \ldots < \text{dose}_n$.

The measured safety or tolerability responses are assumed to be quantitative, and there are no assumptions on the values of the doses or on the relationship between the effects (e.g., linear or quadratic trend). This results in a general effect τ_i of dose i in the model. Furthermore, the additivity of treatment and cohort effects is assumed:

$$Y_j = \mu + \tau_{i(j)} + \eta_{k(j)} + \varepsilon_j, \quad j = 1, \ldots, N, \tag{3.1}$$

where N is the total number of subjects, Y_j is the measured safety or tolerability response for the jth subject, μ is the overall mean, $i(j) \in \{0, \ldots, n\}$ is the treatment administered to the jth subject, $k(j) \in \{1, \ldots, t\}$ is the cohort to which the jth subject belongs, τ_i is the effect of the ith treatment, η_k is the effect of the kth cohort, and ε_j is the random error. Treatment 0 represents the placebo and $1, \ldots, n$ denote the escalating doses. Adopting the notation of Bailey [2009a], we will consider two

cases: standard designs, where $t = n$, and extended designs, where $t = n + 1$. The dose-escalation model (3.1) in fact describes a blocking experiment (2.12) with cohorts as blocks and the doses and the placebo as treatments.

As noted by Bailey [2009a], this model can be employed to describe, e.g., the aforementioned Te Genero trial, where the safety and tolerability measurements included blood pressure and bodily temperature at certain times, quantity of selected cytokines and maximum concentration of the studied drug in serum. Model (3.1) was implicitly used by Senn et al. [2007] to evaluate the design of the Te Genero trial; see Sections 7.4.3.2-4 therein. Such a model for dose-escalation studies was also considered in the monograph Senn [1997]; see Section 20.2.5.

Based on the argued relevance of comparing the latest dose with the placebo, we assume that the objective of the experiment is to compare the drug doses with the placebo. That is, we intend to estimate $\tau_1 - \tau_0, \ldots, \tau_n - \tau_0$, which can be written as a system of contrasts $Q^T\tau$, where $Q^T = (-1_n, I_n)$, or in the terms of the general model (1.1), $K^T\beta$, where $K^T = \big(Q^T, 0_{n\times(t+1)}\big)$.

The value $\xi(i,k)$ determines the proportion of all subjects that are assigned to cohort k and given the treatment i. Let $w_i(\xi) := \sum_k \xi(i,k)$ be the total proportion of the trials with treatment i ($i = 0, \ldots, n$), $s_k(\xi) := \sum_i \xi(i,k)$ be the total proportion of the trials for cohort k ($k = 1, \ldots, t$), $w(\xi) := (w_0(\xi), \ldots, w_n(\xi))^T$, and $s(\xi) := (s_1(\xi), \ldots, s_t(\xi))^T$. Thus, $w(\xi)$ is the treatment proportions design corresponding to ξ and $s(\xi)$ is the vector of the relative sizes of the cohorts. We will omit the argument ξ in expressions in which it is clear from the context.

The moment matrix of a design ξ is

$$M(\xi) = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix}, \tag{3.2}$$

where $M_{11} = \operatorname{diag}(w)$, $M_{12} = \big(w, X\big)$, where $X = \big(\xi(i,k)\big)_{i,k}$, and

$$M_{22} = \begin{bmatrix} 1 & s^T \\ s & \operatorname{diag}(s) \end{bmatrix}.$$

Note that $w = X1_t$ and $s = X^T 1_{n+1}$.

In accordance with Bailey [2009a], we introduce some constraints on the set of considered designs Ξ. For each $k \in \{1, \ldots, n\}$, we have:

(i) $\xi(i,k) = 0$ for $i > k$. In cohort k, the highest allowed dose is k.

(ii) $\sum_i \xi(i,k) = 1/t$.

Constraint (i) formalizes the gradual dose-escalation requirement. Furthermore, in (ii), each cohort is assumed to be of the same size, say m. In the case of extended designs, the only condition for cohort $k = n + 1$ is (ii). In other words, cohort $n + 1$ of extended designs is an additional cohort that is subject to no constraints other than the fixed cohort size. Consequently, condition (ii) is assumed for all $k \in \{1, \ldots, t\}$.

Bailey [2009a] considered an additional constraint (iii) $\xi(k, k) \geq 1/N$. This means that in each cohort, the highest allowed dose must be administered to at least one subject. Since (iii) will naturally be satisfied by optimal designs, we disregard this constraint.

From condition (ii), there follows that $s = t^{-1}1_t$. Thus, the matrix M_{22} is of rank t, and the Schur complement of M_{22} in M is

$$M_\tau = \operatorname{diag}(w) - tXX^T.$$

It is easy to verify that the Schur complement for a given design ξ, and thus the information matrix of ξ, would be the same if the constant term μ was not present in model (3.1). It follows that our results hold even in a model without the constant term.

Recall that the experimental aim considered in this chapter is the comparison of the drug doses with the placebo, which can be expressed as comparing test treatments with a control. For a feasible design ξ, the information matrix for comparing the doses with the placebo $N_K(\xi)$ is obtained by deleting the first row and column of M_τ (e.g., see Bechhofer and Tamhane [1981] or Giovagnoli and Wynn [1985]).

Let us denote

$$M_\tau = \begin{bmatrix} \alpha & b^T \\ b & C \end{bmatrix}, \quad X = \begin{bmatrix} z^T \\ Z \end{bmatrix}, \tag{3.3}$$

where C is an $n \times n$ matrix and Z is an $n \times t$ matrix. Then,

$$N_K(\xi) = C = \operatorname{diag}(w_1, \ldots, w_n) - tZZ^T, \tag{3.4}$$

where $Z = \big(\xi(i, k)\big)_{i=1,\ldots,n;k=1,\ldots,t}$.

3.3 E-optimality

3.3.1 Standard designs

To provide E-optimal designs, we characterize the E-optimal treatment proportions. From Theorem 2.6 and Theorem 2.15, we obtain the following lemma.

Lemma 3.1. *Let ξ be an E-optimal design for comparing the drug doses with the placebo in model (3.1) without design constraints. Then, ξ satisfies $w_0(\xi) = 1/2$ and $w_i(\xi) = 1/(2n)$ for $i = 1, \ldots, n$.*

A standard design that satisfies

$$\xi(0,k) = \xi(k,k) = \frac{1}{2n} \text{ for } k \in \{1,\dots,n\},$$

and is zero otherwise is a *Senn design*; see Table 5. In such a design, half of the subjects in each cohort are assigned to the placebo, and the other half are assigned to the highest allowed dose for that cohort. Note that Senn designs satisfy Lemma 3.1. In other words, they achieve the E-optimal treatment proportions. We show that Senn designs are the unique E-optimal standard dose-escalation designs.

$i\backslash k$	1	2	3	4
0	1/8	1/8	1/8	1/8
1	1/8	0	0	0
2	0	1/8	0	0
3	0	0	1/8	0
4	0	0	0	1/8

Table 5: Senn design for $n = 4$ doses. i - treatment number, k - cohort number

Theorem 3.2. *A standard design ξ is E-optimal for comparing drug doses with the placebo if and only if ξ is a Senn design.*

Proof. The Senn design ξ satisfies

$$w = (1/2, (2n)^{-1}1_n^T)^T, \quad X = (2n)^{-1}(1_n, I_n)^T$$

and thus $Z = (2n)^{-1}I_n$. Note that ξ is connected (see Eccleston and Hedayat [1974]) and therefore feasible. Then, equation (3.4) yields that $N_K = (4n)^{-1}I_n$, which has the smallest eigenvalue $1/(4n)$. The proof of Theorem 2.15 states that when there are no design constraints, the optimal value of the E-criterion is $1/[4(v-1)]$, where v is the number of treatments. In the dose-escalation model, $v = n+1$, and thus, the E-optimal value is $1/(4n)$, which the Senn design attains.

Now, consider the converse part. For a feasible design ξ to be E-optimal, it must have the smallest eigenvalue equal to that of the Senn design, which coincides with the optimal value of the E-optimality criterion in the model without the design constraints. Thus, such ξ is E-optimal in the model without design constraints, and from Lemma 3.1, it follows that ξ must satisfy

$$w_0(\xi) = 1/2, \quad w_i(\xi) = 1/(2n), \quad i = 1,\dots,n; \tag{3.5}$$

hence, $w = (1/2, (2n)^{-1}1_n^T)^T$.

Observe that $N_K(\xi) = C = (2n)^{-1}I_n - nZZ^T$, as defined in (3.3). Then, the smallest eigenvalue $\lambda_{\min}$ of $N_K(\xi)$ satisfies $\lambda_{\min} = 1/(2n) - n\mu_{\max}$, where $\mu_{\max}$ is the largest eigenvalue of ZZ^T. Let us denote the columns of matrix $Z = (z_{ij})$ by $z_1, \ldots, z_n$. Then,

$$\mu_{\max} = \max_{\|u\|=1} u^T ZZ^T u = \max_{\|u\|=1} \|Z^T u\|^2 = \max_{\|u\|=1} \sum_{j=1}^{n} (z_j^T u)^2,$$

and for the particular choice of $u = n^{-1/2}1_n$, we obtain

$$\mu_{\max} \geq \frac{1}{n} \sum_{j=1}^{n} (z_j^T 1_n)^2 = \frac{1}{n} \sum_{j=1}^{n} \Big(\sum_{i=1}^{n} z_{ij} \Big)^2 = \frac{1}{n} \sum_{j=1}^{n} q_j^2,$$

where $q_j := \sum_{i=1}^{n} z_{ij}$, $j = 1, \ldots, n$. From (i), it follows that $q_j = n^{-1} - \xi(0, j)$, and (3.5) yields $\sum_j q_j = 1 - 1/2 = 1/2$. Therefore, $\mu_{\max}$ is not smaller than the optimal value of the following convex optimization problem:

$$\min_{q \geq 0} \frac{1}{n} \sum_{j=1}^{n} q_j^2 \quad \text{subject to} \quad \sum_{j=1}^{n} q_j = \frac{1}{2}. \tag{3.6}$$

Because the objective function of (3.6) is strictly convex on the affine hyperplane formed by the equality constraints and invariant with respect to the permutations of $q_1, \ldots, q_n$, the *single* optimal solution q^* of (3.6) satisfies $q_1^* = \ldots = q_n^* = 1/(2n)$. Hence, the optimal value of (3.6) is $1/(4n^2)$. It follows that if q does not satisfy $q_1 = \ldots = q_n = 1/(2n)$, then $\mu_{\max} > 1/(4n^2)$. Therefore, the smallest eigenvalue $\lambda_{\min}$ of $N_K(\xi)$ for such ξ satisfies $\lambda_{\min} < 1/(4n)$, and thus, ξ is not E-optimal. Hence, any E-optimal design ξ must satisfy $q_1 = \ldots = q_n = 1/(2n)$, i.e., $\xi(0, 1) = \ldots = \xi(0, n) = 1/(2n)$.

From (3.5) and (i), ξ needs to satisfy $w_i(\xi) = 1/(2n)$ for $i > 0$ and $s_k(\xi) = \sum_i \xi(i, k) = 1/n$ for all k. Because $\xi(n, k)$ can be nonzero only for $k = n$, we obtain $\xi(n, n) = 1/(2n)$. Since $\xi(0, n) = 1/(2n)$, (ii) yields $\xi(i, n) = 0$ for $0 < i < n$. Then, $\xi(n-1, k)$ is nonzero only for $k = n-1$, which implies $\xi(n-1, n-1) = 1/(2n)$, $\xi(i, n-1) = 0$ for $0 < i < n-1$, and so forth. It follows that the single E-optimal standard design for given n is the Senn design. □

3.3.2 Extended designs

The extra cohort in extended designs provides additional freedom for the experimenter. Therefore, for each n, the class of E-optimal extended designs is richer compared to the single E-optimal standard design. The E-optimal extended designs can be completely characterized by the following conditions: the designs assign the placebo to half of the subjects in each cohort and allocate each dose to the same total number of subjects, $N/(2n)$. Note that the E-optimality of standard designs can be described by the same conditions. However, in the standard design case, the Senn designs are the only ones that satisfy these conditions. Recall that $t = n + 1$ in the extended design case.

Theorem 3.3. *An extended design ξ is E-optimal for comparing the drug doses with the placebo if and only if it satisfies $\xi(0,k) = 1/(2t)$ for $k = 1,\dots,n+1$ and $w_1(\xi) = \dots = w_n(\xi) = 1/(2n)$.*

Proof. Let ξ be a design that satisfies the conditions of Theorem 3.3. First, observe that ξ is connected (see Eccleston and Hedayat [1974]) and therefore feasible. The partitions of M_τ defined in (3.3) satisfy $z = (2(n+1))^{-1}1_{n+1}$, $X1_{n+1} = w = \big(1/2, (2n)^{-1}1_n^T\big)^T$, $Z1_{n+1} = (2n)^{-1}1_n$, $X^T1_{n+1} = (n+1)^{-1}1_{n+1}$ and $Z^T1_n = (2(n+1))^{-1}1_{n+1}$.

Thus, $N_K = C = (2n)^{-1}I_n - (n+1)ZZ^T$. Moreover, $N_K1_n = (4n)^{-1}1_n$. That is, 1_n is an eigenvector of N_K with eigenvalue $\lambda^* = 1/(4n)$. Analogous to the proof of Theorem 3.2, λ^* is equal to the optimal smallest eigenvalue in the model without constraints. Therefore, to prove that ξ is E-optimal, it suffices to prove that λ^* is the smallest eigenvalue of N_K. The smallest eigenvalue of N_K satisfies $\lambda_{\min}(N_K) = 1/(2n) - (n+1)\mu_{\max}$, where $\mu_{\max}$ is the largest eigenvalue of ZZ^T.

Let $\hat{Z} = \big(2nZ^T, (n+1)^{-1}1_{n+1}\big)$. The matrix $\hat{Z}$ satisfies $\hat{Z}1_{n+1} = 1_{n+1}$ and $1_{n+1}^T\hat{Z} = 1_{n+1}^T$, i.e., $\hat{Z}$ is doubly stochastic. Using the Birkhoff-von Neumann theorem, we obtain that

$$\hat{Z} = \sum_{\pi\in\mathfrak{P}_{n+1}} \alpha_\pi P_\pi,$$

where $\alpha_\pi \geq 0$ for all π, $\sum_\pi \alpha_\pi = 1$, $\mathfrak{P}_{n+1}$ denotes the set of all permutations of $n+1$ elements, and P_π is the permutation matrix given by π. Let us partition the matrices P_π as $P_\pi = \big(\tilde{P}_\pi, v_\pi\big)$, where $\tilde{P}_\pi$ is an $(n+1)\times n$ matrix. Then,

$$2nZ^T = \sum_{\pi\in\mathfrak{P}_{n+1}} \alpha_\pi\tilde{P}_\pi, \quad Z = \frac{1}{2n}\sum_{\pi\in\mathfrak{P}_{n+1}} \alpha_\pi\tilde{P}_\pi^T, \quad \frac{1}{n+1}1_{n+1} = \sum_{\pi\in\mathfrak{P}_{n+1}} \alpha_\pi v_\pi.$$

For any $i \in \{1,...,n+1\}$, let $\mathfrak{P}_{n+1}^{(i)}$ be the set of all permutations π of $n+1$ elements, the corresponding matrices of which have their last column equal to e_i, i.e., $\pi(n+1) = i$. Then, because the last column of $\hat{Z}$ is $(n+1)^{-1}1_{n+1}$, we have $\sum_{\pi\in\mathfrak{P}_{n+1}^{(i)}} \alpha_\pi = 1/(n+1)$.

The largest eigenvalue $\mu_{\max}$ of ZZ^T is identical to the largest eigenvalue of Z^TZ. Therefore, it satisfies

$$\begin{aligned}\mu_{\max} &= \max_{\|u\|=1} u^TZ^TZu = \max_{\|u\|=1}\|Zu\|^2 = \frac{1}{4n^2}\max_{\|u\|=1}\Big\|\sum_{\pi\in\mathfrak{P}_{n+1}}\alpha_\pi\tilde{P}_\pi^Tu\Big\|^2 \\ &= \frac{1}{4n^2}\max_{\|u\|=1}\Big\|\sum_{i=1}^{n+1}\sum_{\pi\in\mathfrak{P}_{n+1}^{(i)}}\alpha_\pi\tilde{P}_\pi^Tu\Big\|^2 \leq \frac{n+1}{4n^2}\max_{\|u\|=1}\sum_{i=1}^{n+1}\Big\|\sum_{\pi\in\mathfrak{P}_{n+1}^{(i)}}\alpha_\pi\tilde{P}_\pi^Tu\Big\|^2,\end{aligned}$$

where the last inequality follows from the fact that $\|\sum_{i=1}^k x_i\|^2 \leq k\sum_{i=1}^k\|x_i\|^2$ for any vectors $x_1,\dots,x_k$ of the same dimension.

Let $i \in \{1,\dots,n+1\}$. Then, the set of the extremal points of the convex set

$$\mathfrak{K}_i := \Big\{\sum_{\pi\in\mathfrak{P}_{n+1}^{(i)}}\alpha_\pi\tilde{P}_\pi^Tu \,\Big|\, \alpha_\pi \geq 0 \text{ for all } \pi\in\mathfrak{P}_{n+1}^{(i)}, \sum_{\pi\in\mathfrak{P}_{n+1}^{(i)}}\alpha_\pi = \frac{1}{n+1}\Big\}$$

is a subset of $\{(n+1)^{-1}\tilde{P}_\pi^T u \,|\, \pi \in \mathfrak{P}_{n+1}^{(i)}\}$, and the convex function $\|x\|^2, x \in \mathfrak{K}_i$, attains its maximum in at least one of the extremal points of $\mathfrak{K}_i$. It follows that

$$\Big\| \sum_{\pi \in \mathfrak{P}_{n+1}^{(i)}} \alpha_\pi \tilde{P}_\pi^T u \Big\|^2 \le \max_{\pi_i \in \mathfrak{P}_{n+1}^{(i)}} \Big\| \frac{1}{n+1} \tilde{P}_{\pi_i}^T u \Big\|^2 = \frac{1}{(n+1)^2} \max_{\pi_i \in \mathfrak{P}_{n+1}^{(i)}} \|\tilde{P}_{\pi_i}^T u\|^2,$$

and thus,

$$\mu_{\max} \le \frac{1}{4n^2(n+1)} \max_{\|u\|=1} \max_{\pi_1 \in \mathfrak{P}_{n+1}^{(1)}, \ldots, \pi_{n+1} \in \mathfrak{P}_{n+1}^{(n+1)}} \sum_{i=1}^{n+1} \|\tilde{P}_{\pi_i}^T u\|^2.$$

Consider $\pi_1 \in \mathfrak{P}_{n+1}^{(1)}, \ldots, \pi_{n+1} \in \mathfrak{P}_{n+1}^{(n+1)}$ and $u \in \mathbb{R}^{n+1}$, $\|u\| = 1$, that maximize the expression $\sum_{i=1}^{n+1} \|\tilde{P}_{\pi_i}^T u\|^2$, and let us denote $v_i = (v_{i,1}, \ldots, v_{i,n})^T := \tilde{P}_{\pi_i}^T u$, $i = 1, \ldots, n+1$. Then, $\mu_{\max}$ satisfies

$$\mu_{\max} \le \frac{1}{4n^2(n+1)} \sum_{i=1}^{n+1} \sum_{j=1}^{n} v_{i,j}^2 = \frac{1}{4n^2(n+1)} \sum_{i=1}^{n+1} S,$$

where S is the sum of squares of all elements of vectors $v_1, \ldots, v_{n+1}$. For each $i \in \{1, \ldots, n+1\}$, the vector $v_i = \tilde{P}_{\pi_i}^T u$ consists of all elements of u except u_i. Therefore, u_i^2 occurs n-times in S. Hence, $S = nu_1^2 + \cdots + nu_{n+1}^2 = n\|u\|^2 = n$, and it follows that $\mu_{\max} \le (4n(n+1))^{-1}$. Since $\mu_{\max} \le (4n(n+1))^{-1}$, the smallest eigenvalue of N_K is at least $1/(2n) - (n+1)/(4n(n+1)) = 1/(4n)$, and thus, λ^* is indeed the smallest eigenvalue of N_K.

For the converse part, note that similar to the proof of Theorem 3.2, the optimal value of the E-criterion is $1/(4n)$ and Lemma 3.1 yields that any E-optimal design ξ must satisfy (3.5). Let ξ be a design that satisfies (3.5) but that does not satisfy $\xi(0,k) = 1/(2(n+1))$, $k = 1, \ldots, n+1$. Let $q_j := 1/(n+1) - \xi(0,j)$, $j = 1, \ldots, n+1$; hence, $\sum_j q_j = 1 - 1/2 = 1/2$. Then, analogous to the proof of Theorem 3.2, the largest eigenvalue $\mu_{\max}$ of ZZ^T is not smaller than the optimal value of the following convex optimization problem:

$$\min_{q \ge 0} \frac{1}{n} \sum_{j=1}^{n+1} q_j^2 \quad \text{subject to} \quad \sum_{j=1}^{n+1} q_j = \frac{1}{2}.$$

This problem has a unique optimal solution q^* satisfying $q_1^* = \ldots = q_{n+1}^* = (2(n+1))^{-1}$, and the optimal value is $(4n(n+1))^{-1}$. It follows that if q does not satisfy $q_1 = \ldots = q_{n+1} = (2(n+1))^{-1}$, then $\mu_{\max} > (4n(n+1))^{-1}$. Therefore, the smallest eigenvalue $\lambda_{\min}$ of N_K for such ξ satisfies $\lambda_{\min} < (4n)^{-1}$, and thus, ξ is not E-optimal. □

The standard Senn designs can be extended to obtain E-optimal extended designs. We say that a *uniformly extended Senn design* is a design that satisfies

$$\xi(0,k) = \frac{1}{2t} \text{ for } k = 1, \ldots, n+1, \quad \xi(i,i) = \frac{1}{2t} \text{ and } \xi(i,n+1) = \frac{1}{2nt} \text{ for } i = 1, \ldots, n,$$

and is zero otherwise; see Table 6. Such a design is given by the standard Senn design in the first n cohorts. In the last cohort, half of the subjects are assigned to the control, and the other half are uniformly distributed among all other treatments. The uniformly extended Senn designs obviously satisfy the conditions of Theorem 3.3 and are thus E-optimal.

$i\backslash k$	1	2	3	4	5
0	1/10	1/10	1/10	1/10	1/10
1	1/10	0	0	0	1/40
2	0	1/10	0	0	1/40
3	0	0	1/10	0	1/40
4	0	0	0	1/10	1/40

Table 6: Uniformly extended Senn design for $n = 4$ doses. i - treatment number, k - cohort number

However, the class of E-optimal extended dose-escalation designs does not contain only the uniformly extended Senn designs. The complete characterization of E-optimal extended designs in Theorem 3.3 by linear constraints allows one to choose the best E-optimal design with respect to some secondary criterion, such as other optimality criteria. For example, one may choose the well-known criteria of D- and A-optimality.

Because the provided characterization of E-optimal extended designs is linear, we can employ an efficient convex optimization method from Harman and Sagnol [2015] to construct A-optimal or D-optimal designs within the class of all E-optimal designs:

Example 3.4. Let $n = 4$. Then, the A-optimal and D-optimal designs in the class of all E-optimal extended designs are given in Tables 7 and 8, respectively.

$i\backslash k$	1	2	3	4	5
0	0.1000	0.1000	0.1000	0.1000	0.1000
1	0.1000	0.0219	0.0031	0.0000	0.0000
2	0	0.0781	0.0287	0.0091	0.0091
3	0	0	0.0682	0.0284	0.0284
4	0	0	0	0.0625	0.0625

Table 7: The A-optimal design in the class of all E-optimal extended designs. i - treatment number, k - cohort number

$i\backslash k$	1	2	3	4	5
0	0.1000	0.1000	0.1000	0.1000	0.1000
1	0.1000	0.0248	0.0002	0.0000	0.0000
2	0	0.0752	0.0339	0.0079	0.0079
3	0	0	0.0659	0.0296	0.0296
4	0	0	0	0.0625	0.0625

Table 8: The D-optimal design in the class of all E-optimal extended designs. i - treatment number, k - cohort number

3.3.3 Interpretation of E-optimality for dose-escalation

Majumdar and Notz [1983], who pioneered the use of the optimality criteria in the design of experiments for comparing test treatments with a control, noted that E-optimality 'does not seem to posses a very natural statistical meaning [for the comparisons with a control]'. Since then, the criterion received little attention in the design literature on the test treatment-control comparisons, with a few exceptions (see Notz [1985], Morgan and Wang [2011]).

Nevertheless, the E-optimality does have a statistical interpretation for estimating a system of contrasts $Q^T\tau$ in terms of the confidence ellipsoid for $Q^T\tau$, which represents the amount of uncertainty about the studied system of contrasts. Under the assumption of normal errors, the largest eigenvalue of $N_K^{-1}(\xi)$ is proportional to the square of the length of the largest semi-axis of the confidence ellipsoid for $Q^T\tau$. Therefore, any E-optimal design for comparing drug doses with the placebo minimizes the length of the largest semi-axis of the confidence ellipsoid for $\tau_1 - \tau_0, \ldots, \tau_n - \tau_0$. This suggests that the E-optimality criterion may be relevant, as it minimizes the diameter of the confidence ellipsoid.

The E-optimality can also be expressed in terms of the variances of the drug dose-placebo contrasts in a straightforward manner, as by Morgan and Wang [2011], even without the normality assumption. We can write

$$\lambda_{\max}(N_K^{-1}(\xi)) = \max_{\|x\|=1} x^T N_K^{-1}(\xi) x.$$

Because $N_K^{-1}(\xi)$ is proportional to the covariance matrix of the least squares estimator of $Q^T\tau$, we have:

Proposition 3.5. *Any standard or extended E-optimal design for comparing the drug doses with the placebo minimizes the maximum variance among all least squares estimators of the drug dose-placebo contrasts of the form*

$$x^T Q^T \tau = \sum_{i=1}^{n} x_i \tau_i - (\sum_{i=1}^{n} x_i)\tau_0, \text{ where } \|x\| = 1. \tag{3.7}$$

Notz [1985] obtained a class of E-optimal exact row-column designs in a model without design constraints and provided two statistical interpretations for these designs using test treatment-control contrasts. One of these interpretations (Theorem 3.3 therein) was that the obtained designs minimized the maximum variance for the contrasts of the form (3.7). Morgan and Wang [2011] analyzed the relationship as in Proposition 3.5 further, using weighted optimality criteria (Theorem 2.1 therein).

Employing the relationship between E-optimality and c-optimality (see Pukelsheim and Studden [1993], Pukelsheim [1993]), it is possible to provide one more interpretation of the E-optimality in our model using the drug dose-placebo contrasts.

Theorem 3.6. *Let ξ be an E-optimal standard or extended design in model (3.1) for comparing the drug doses with the placebo. Then, ξ is c-optimal, where $c = (-1, n^{-1}1_n^T, 0_{t+1}^T)^T$. That is, ξ minimizes the variance of the least squares estimator of $\sum_i(\tau_i - \tau_0)/n$.*

Proof. Recall that $K^T = (Q^T, 0_{n\times(t+1)})$ and $Q^T = (-1_n, I_n)$. Let $\tilde{c} := (-1, n^{-1}1_n^T, 0_{t+1}^T)^T$. For the (standard) Senn design, we have $N_K = (4n)^{-1}I_n$. In the Equivalence Theorem (Lemma 1.10), let the generalized inverse G of M be given by (1.5), where

$$M_\tau^- = \begin{bmatrix} 0 & 0_n^T \\ 0_n & 4nI_n \end{bmatrix} \text{ and } M_{22}^- = \begin{bmatrix} 0 & 0_n^T \\ 0_n & nI_n \end{bmatrix}, \tag{3.8}$$

and $E = 1_n 1_n^T/n = hh^T$, where $h = n^{-1/2}1_n$. Then, it is possible to verify that there is equality in (1.4) for any permissible design $\tilde{\xi}$. Therefore, Lemma 1.14 implies that any E-optimal standard design (i.e., Senn design) is also c-optimal, where $c = n^{-1/2}K1_n$. This is equivalent to optimality for $n^{-1}K1_n = \tilde{c}$.

Let ξ be an E-optimal extended design for comparing the doses with the placebo, with its moment matrix M and information matrix N_K. Then, $h = n^{-1/2}1_n$ is an eigenvector of N_K corresponding to the smallest eigenvalue $\lambda_{\min} = 1/(4n)$ (see the proof of Theorem 3.3). The left-hand side of the normality inequality (1.4) for $E = hh^T$ and a given $\tilde{\xi} \in \Xi$ becomes

$$\operatorname{tr}(\tilde{M}GKN_Khh^TN_KK^TG^T) = \lambda_{\min}^2 h^TK^TG^T\tilde{M}GKh = \frac{1}{16n^3}1_n^TK^TG^T\tilde{M}GK1_n, \tag{3.9}$$

where $\tilde{M} := M(\tilde{\xi})$. Moreover, $K1_n = (-n, 1_n^T, 0_{n+2}^T)^T$.

Recall the partitioning of M_τ and X defined in (3.3). Let the generalized inverse G of M be given by (1.5), where

$$M_\tau^- = \begin{bmatrix} 0 & 0_n^T \\ 0_n & C^{-1} \end{bmatrix}.$$

Then,

$$GK1_n = \begin{bmatrix} 0 \\ C^{-1}1_n \\ 0 \\ -(n+1)Z^TC^{-1}1_n \end{bmatrix}.$$

Since $C^{-1} = N_K^{-1}$, we obtain $C^{-1}1_n = 4n1_n$. Then, $GK1_n = (0, 4n1_n^T, 0, -(n+1)4n1_n^TZ)^T = (0, 4n1_n^T, 0, -2n1_{n+1}^T)^T$ because $Z^T1_n = (2(n+1))^{-1}1_{n+1}$.

Let us partition $\tilde{M} = M(\tilde{\xi})$ using $\tilde{M}_{11}$, $\tilde{M}_{12}$, $\tilde{M}_{22}$ as in (3.2), and let us denote $\tilde{Z} = Z(\tilde{\xi})$ and $\tilde{w} = w(\tilde{\xi})$. Then, (3.9) becomes

$$\frac{1}{16n^3}\Big((0, 4n1_n^T)\tilde{M}_{11}\begin{pmatrix}0\\4n1_n\end{pmatrix} + (0, -2n1_{n+1}^T)\tilde{M}_{22}\begin{pmatrix}0\\-2n1_{n+1}\end{pmatrix} + 2(0, 4n1_n^T)\tilde{M}_{12}\begin{pmatrix}0\\-2n1_{n+1}\end{pmatrix}\Big),$$

which is equal to

$$\frac{1}{16n^3}\Big(16n^2\sum_{i>0}\tilde{w}_i + 4n^2 - 16n^2\sum_{i>0}\tilde{w}_i\Big) = \frac{1}{4n} = \lambda_{\min},$$

because $\tilde{Z}1_{n+1} = (\tilde{r}_1, \ldots, \tilde{r}_n)^T$. Therefore, the left-hand side of (1.4) is always equal to the right-hand side. Hence, Lemma 1.14 yields that ξ is $\tilde{c}$-optimal. □

Theorem 3.6 shows that if a standard or extended design is E-optimal for comparing the doses with the placebo, it minimizes the variance of the average dose effect compared with the placebo $(\sum_i \tau_i/n) - \tau_0$ or the average of the dose effects relative to the placebo $\sum_i(\tau_i - \tau_0)/n$. That is, the E-optimal designs also minimize

$$\mathrm{Var}\Big(\frac{1}{n}\widehat{\sum_{i=1}^{n}\tau_i - \tau_0}\Big), \tag{3.10}$$

where $\hat{\alpha}$ is the least squares estimator of α. The other interpretation of the class of E-optimal row-column designs obtained by Notz [1985] consisted of showing that the obtained designs minimized (3.10).

By contrast, the A-optimality criterion, which is often used in test treatment-control experiments, minimizes the average variance of the least squares estimators of the contrasts of interest. That is, it minimizes the average variance for the dose effects compared with the placebo effect, which can be expressed as minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\Big(\widehat{\tau_i - \tau_0}\Big);$$

compare with (3.10).

In fact, in the proof of Theorem 3.6, we show that the E-optimal designs for comparing drug doses with the placebo minimize $h^T N_K^{-1}(\xi)h$, where $h = 1_n$ (and this expression is equal to $n\lambda_{\min}^{-1}(N_K(\xi))$ for any E-optimal design ξ). Thus, the E-optimal designs minimize the sum of all elements of the inverse of the information matrix. It follows that the E-optimal designs for comparing drug doses with the placebo minimize the sum of all elements of the covariance matrix of $\widehat{Q^T\tau}$. In other words, such designs minimize the sum (or the average) of all variances $\mathrm{Var}(\widehat{\tau_i - \tau_0})$, $i = 1, \ldots, n$, and covariances $\mathrm{Cov}(\widehat{\tau_i - \tau_0}, \widehat{\tau_j - \tau_0})$, $i \neq j$, for the drug dose-placebo comparisons. By contrast, the A-optimal designs minimize the trace of $N_K^{-1}(\xi)$, i.e., the sum (or the average) of all variances $\mathrm{Var}(\widehat{\tau_i - \tau_0})$.

The following proposition shows that the covariances for the studied comparisons are nonnegative. This implies that the minimization of the sum of all covariances of $\widehat{\tau_i - \tau_0}$ is equivalent to the statistically meaningful minimization of the sum of absolute values of these covariances.

Proposition 3.7. *Let $Q = (-1_n, I_n)^T$ and let ξ be a standard or extended design that is feasible for $Q^T\tau$. Then, all elements of $N_K^{-1}(\xi)$ are nonnegative, i.e.,*

$$\mathrm{Cov}(\widehat{\tau_i - \tau_0}, \widehat{\tau_j - \tau_0}) \geq 0 \quad \textit{for all } i, j \in \{1, \ldots, n\}.$$

Proof. The information matrix $N_K(\xi)$ given by (3.4) clearly satisfies that its non-diagonal elements are non-positive. For any matrix H with non-positive non-diagonal elements, Theorem 6.2.3 of Berman and Plemmons [1994] yields that the following statements are equivalent: (D_{16}) every real eigenvalue of H is positive, and (N_{38}) every element of H^{-1} is nonnegative. Since $N_K(\xi)$ is a non-singular information matrix, it is positive definite, and thus $N_K(\xi)$ satisfies (D_{16}). Therefore, every element of $N_K^{-1}(\xi)$ is nonnegative. □

3.3.4 Comparing standard and extended designs

Bailey [2009a] considered extended designs because they often allow one to obtain lower variances of the treatment comparisons compared to the standard designs; see Section 7 therein. We demonstrate this phenomenon on the optimal designs obtained in this chapter. In particular, for a given number of doses n, consider the Senn design ξ_S and the uniformly extended Senn design ξ_{ES}. Then, from the proofs of Theorems 3.2 and 3.3, it follows that $\lambda_{\min}(N_K(\xi_S)) = 1/(4n) = \lambda_{\min}(N_K(\xi_{ES}))$, which suggests that their performance with respect to the E-optimality criterion is the same. Theorem 3.6 implies that both the standard and the extended E-optimal designs minimize the sum of all elements of the covariance matrix of $\widehat{Q^T\tau}$, which is proportional to $1_n^T N_K^{-1} 1_n = n\lambda_{\min}^{-1}(N_K)$. It follows that the sum of all variances and covariances of $\widehat{\tau_i - \tau_0}$ is the same under ξ_S and ξ_{ES}.

However, $N_K(\xi_S) = (4n)^{-1} I_n$ is a diagonal matrix, and $N_K(\xi_{ES}) = (2n)^{-1} I_n - (n+1) Z(\xi_{ES}) Z^T(\xi_{ES})$ is not. Specifically,

$$N_K(\xi_{ES}) = \frac{2(n+1)-n}{4n(n+1)} I_n - \frac{1}{4n^2(n+1)} J_n$$

and

$$N_K^{-1}(\xi_{ES}) = \frac{4n(n+1)}{2(n+1)-n} I_n + \frac{4}{2(n+1)-n} J_n.$$

Therefore, under the uniformly extended Senn design the variances of $\widehat{\tau_i - \tau_0}$ are proportional to $[4n(n+1)+4]/[2(n+1)-n] < 4n$ and their covariances are proportional to $4/[2(n+1)-n]$. That is, the uniformly extended Senn design provides lower variances of $\widehat{\tau_i - \tau_0}$ than the standard Senn design, but at the cost of their covariances being nonzero, unlike under the Senn design. Subsequently, it is up to the experimenter regarding whether they prefer uncorrelated estimates of $\tau_i - \tau_0$ with slightly higher variances, or smaller variances of the estimates at the cost of losing uncorrelation.

Moreover, all the eigenvalues of $N_K(\xi_{ES})$ except the smallest one are greater than all the eigenvalues of $N_K(\xi_S)$, because all the eigenvalues of $N_K(\xi_S)$ coincide. Therefore, the uniformly extended Senn design performs at least as good as the Senn design with respect to any eigenvalue-based optimality criterion. In fact, ξ_{ES} is strictly better than ξ_S with respect to any of the Kiefer's optimality criteria including A-optimality and D-optimality, with the exception of E-optimality.

Note that analogous results hold for any E-optimal extended design ξ that satisfies that $Z(\xi)Z^T(\xi)$ is not diagonal.

Example 3.8. Suppose that we wish to examine $n = 3$ doses on 24 subjects. Consider the Senn design ξ_S given in Table 9a and the uniformly extended Senn design ξ_{ES} given in Table 9b. Both of these designs use the same number of subjects to examine the three dose effects. Similar to (3.4), the information matrix of an exact design ξ is given by

$$N_K(\xi) = \mathrm{diag}(w_1, \ldots, w_n) - m^{-1} Z Z^T,$$

where w and Z are defined as for an approximate design and m is the size of each cohort. Then, $N_K(\xi_S) = 4I_3 - 8^{-1}4^2 I_3 = 2I_3$ and $N_K^{-1}(\xi_S) = 2^{-1} I_3$, $N_K(\xi_{ES}) = 4I_3 - 6^{-1}(9I_3 + J_3) = (5/2)I_3 - 6^{-1} J_3$ and $N_K^{-1}(\xi_{ES}) = (2/5)I_3 + (30)^{-1} J_3$. Thus, under ξ_S, the variances of $\widehat{\tau_i - \tau_0}$ are proportional to $1/2$, and the covariances are zero. Under ξ_{ES}, the variances are proportional to $2/5 + 1/30 = 13/30 < 1/2$, and the covariances are proportional to $1/30$. As shown above, there is a small trade-off between the variances and the covariances of the estimators. Note that both $N_K(\xi_S)$ and $N_K(\xi_{ES})$ have the same smallest eigenvalue $\lambda_{\min} = 2$. Therefore, ξ_S and ξ_{ES} are equally good with respect to the E-optimality criterion. Moreover, the sum of all variances and covariances is the same for these two designs, in accordance with Theorem 3.6.

$i\backslash k$	1	2	3
0	4	4	4
1	4	0	0
2	0	4	0
3	0	0	4

(a) Senn design

$i\backslash k$	1	2	3	4
0	3	3	3	3
1	3	0	0	1
2	0	3	0	1
3	0	0	3	1

(b) Uniformly extended Senn design

Table 9: Two E-optimal exact designs for $n = 3$ doses and 24 subjects. i - treatment number, k - cohort number

This demonstrates that the extended designs provide an alternative to the standard designs; they tend to imply slightly lower variances at the cost of increasing the covariances of the estimators of interest while maintaining the same numbers of subjects. Moreover, the E-optimal extended designs are better or at least as good as the standard Senn design with respect to any eigenvalue-based optimality criterion. One drawback of using E-optimal extended designs is their dependence on the entire study being conducted. If the study is stopped, the last cohort is never performed, and their optimality is generally lost. In contrast, the self-similarity of the (standard) Senn designs eliminates this problem.

3.4 LV-optimality

Senn [2009] suggested that what he calls the *latest variance* (LV) criterion may be more appropriate to consider than the traditional optimality criteria. Such a criterion seeks to minimize the variance of the estimator for comparing the latest dose with the placebo after each cohort. The reasoning is that after each cohort, the experimenter decides whether to continue with the experiment, and this decision is based on the results obtained thus far (the effect of the latest dose relative to the placebo). Therefore, the variance of the comparison of the latest dose with the placebo is crucial in the experiment, rather than all variances based on trials that may not even be performed.

Since the Senn designs place all the weight in a given cohort on the comparison of the latest dose with the placebo, intuitively, it is to be expected for them to be LV-optimal. We formalize this observation and the LV-optimality criterion itself.

We denote by $\mathrm{Var}_k(\widehat{\tau_k - \tau_0})$ the variance of the least squares estimator of $\tau_k - \tau_0$ based on the trials in the first k cohorts, $k = 1, \ldots, n$. Then, an *LV-optimal* design ξ for comparing the drug doses with the placebo minimizes each of $\mathrm{Var}_1(\widehat{\tau_1 - \tau_0})$, $\ldots$, $\mathrm{Var}_n(\widehat{\tau_n - \tau_0})$ among all feasible dose-escalation designs. By a feasible design, we mean here a design under which $\tau_1 - \tau_0$ is estimable after one cohort, $\tau_2 - \tau_0$ is estimable

after the first two cohorts and so forth. We further remark that the LV-optimality criterion is not defined for a general design problem; it is strongly tied to the particular properties of the dose-escalation studies.

The criterion of LV-optimality appears to call for a simultaneous multicriterial optimization (minimizing n variances) which need not have a solution in general. In other words, some trade-off between the various variances of interest might be required. However, we show that in our model, the Senn designs do have this strong property. That is, no other dose-escalation design has any of the variances in question lower.

Theorem 3.9. *Let ξ be a Senn design. Then, ξ is an LV-optimal standard design.*

Proof. Let ξ be a standard design and let $k \in \{1, \dots, n\}$. The covariance matrix of the least squares estimator of the contrasts of interest, $\mathrm{Var}_\xi(\widehat{Q^T\tau})$, is proportional to $N_K^{-1}(\xi) = Q^T M_\tau^- Q$. The variance calculated from the first k cohorts under a given design ξ, $\mathrm{Var}_k(\widehat{\tau_k - \tau_0})$, can be obtained from a stage-k 'design' $\xi^{(k)}$. Such a design is given by deleting all the trials in cohorts $k+1, \dots, n$ from ξ. For any i, we have $\xi^{(k)}(i,j) = \xi(i,j)$ for $j \le k$ and $\xi^{(k)}(i,j) = 0$ for $j > k$. Note that for $k < n$, $\xi^{(k)}$ is not a proper approximate design because all of its elements do not sum to 1, indicating that the trials in the future cohorts are ignored.

The stage-k moment matrix $M^{(k)}$ given by $\xi^{(k)}$ satisfies $M_{11}^{(k)} = \mathrm{diag}(w_0, \dots, w_k, 0_{n-k}^T)$,

$$M_{12}^{(k)} = \begin{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_k \end{bmatrix} & \begin{bmatrix} \xi(0,1) & \xi(0,2) & \dots & \xi(0,k) \\ \xi(1,1) & \xi(1,2) & \dots & \xi(1,k) \\ 0 & \xi(2,2) & \dots & \xi(2,k) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \xi(k,k) \end{bmatrix} & 0_{(k+1)\times(t-k)} \\ 0_{n-k} & 0_{(n-k)\times k} & 0_{(n-k)\times(t-k)} \end{bmatrix}$$

and

$$M_{22}^{(k)} = \begin{bmatrix} 1 & n^{-1}1_k^T & 0_{t-k}^T \\ n^{-1}1_k & n^{-1}I_k & 0_{k\times(t-k)} \\ 0_{(t-k)} & 0_{(t-k)\times k} & 0_{(t-k)\times(t-k)} \end{bmatrix},$$

where $t = n$, because we are considering standard designs. Then, the latest variance $\mathrm{Var}_k(\widehat{\tau_k - \tau_0})$ is proportional to $d_k(\xi) := (e_{k+1}^T - e_1^T, 0_{t+1}^T)\big(M^{(k)}\big)^-(e_{k+1}^T - e_1^T, 0_{t+1}^T)^T$, where e_1 and e_{k+1} are elementary unit vectors of length $n+1$.

To calculate $d_k(\xi)$, we disregard cohorts $k+1, \dots, n$ and we are not allowed to use doses $k+1, \dots, n$ in the first k cohorts. Such a model coincides with model (3.1), where $n = k$ and a design ξ' given by ξ for doses $0, \dots, k$ and cohorts $1, \dots, k$. Then, the latest variance under ξ is proportional to the inverse of the value of the c-optimality criterion for ξ', which is $\Phi_c(\xi') = \big(c^T M^-(\xi')c\big)^{-1}$ for $c^T = (e_{n+1}^T - e_1^T, 0_{n+1}^T)$. Therefore, without loss of generality, we may assume that $k = n$ and prove that the Senn design

has the highest value of c-optimality criterion, where $c^T = (e_{n+1}^T - e_1^T, 0_{n+1}^T)$. This can be proved using the General Equivalence Theorem for c-optimality; see Lemma 1.11.

Let ξ be a Senn design, let $c = (e_{n+1}^T - e_1^T, 0_{n+1}^T)^T$, and let us denote $M := M(\xi)$. Then, $w = 2^{-1}\big(1, n^{-1}1_n^T\big)^T$, $X = (2n)^{-1}\big(1_n, I_n\big)^T$,

$$M_\tau = \frac{1}{4n}\begin{bmatrix} n & -1_n^T \\ -1_n & I_n \end{bmatrix}, \text{ and let } M_\tau^- = \begin{bmatrix} 0 & 0_n^T \\ 0_n & 4nI_n \end{bmatrix}.$$

Thus, $c^T M^- c = (e_{n+1} - e_1)^T M_\tau^- (e_{n+1} - e_1) = 4n$.

Let G be the generalized inverse of M given by (1.5), where M_{22}^- is given by (3.8). It follows that

$$Gc = \begin{bmatrix} 0 & 0_n^T & \\ 0_n & 4nI_n & G_{12} \\ 0 & 0_n^T & \\ 0_n & -2nI_n & G_{22} \end{bmatrix} \begin{bmatrix} -1 \\ 0_{n-1} \\ 1 \\ 0_{n+1} \end{bmatrix} = \begin{bmatrix} 0_n \\ 4n \\ 0_n \\ -2n \end{bmatrix}$$

and therefore, for any feasible design ξ' satisfying conditions (i), (ii), we have

$$c^T G^T M(\xi') Gc = \begin{bmatrix} 0_n^T & 4n & 0_n^T & -2n \end{bmatrix} \begin{bmatrix} \operatorname{diag}(w') & w' & X' \\ w'^T & 1 & n^{-1}1_n^T \\ X'^T & n^{-1}1_n & n^{-1}I_n \end{bmatrix} \begin{bmatrix} 0_n \\ 4n \\ 0_n \\ -2n \end{bmatrix}$$

$$= (4n)^2 w_n' + (-2n)^2 \frac{1}{n} + 2 \times 4n(-2n)\xi'(n,n) = 4n,$$

because $\xi'(n,n) = w_n'$. Therefore, any design satisfying conditions (i), (ii) satisfies the desired inequality $4n \le c^T M^- c = 4n$. Hence, for any k, the Senn design attains the minimum possible latest variance. □

In the case of extended designs, a design is *LV-optimal* if it minimizes $\operatorname{Var}_1(\widehat{\tau_1 - \tau_0})$, $\ldots$, $\operatorname{Var}_n(\widehat{\tau_n - \tau_0}), \operatorname{Var}_{n+1}(\widehat{\tau_n - \tau_0})$, where $\operatorname{Var}_{n+1}(\widehat{\tau_n - \tau_0})$ is the variance of $\widehat{\tau_n - \tau_0}$ once the entire study has been performed. Unsurprisingly, the LV-optimal design is the extended Senn design ξ, constructed as follows: ξ is given by the Senn design in the first n cohorts, and its $(n+1)$-st cohort is a replication of the nth cohort in the sense of the treatment proportions. Formally,

$$\xi(0,k) = \xi(k,k) = \frac{1}{2(n+1)} \text{ for } k = 1, \ldots, n, \quad \xi(0,n+1) = \xi(n,n+1) = \frac{1}{2(n+1)},$$

and $\xi(i,k) = 0$ otherwise; see Table 10. We call such a design a *highest-dose extended Senn design.*

Note that the extended designs appear to be less meaningful when considering LV-optimality because in such a case, the additional cohort is only used to improve the estimate of the effect of the highest dose compared with the placebo.

$i\backslash k$	1	2	3	4	5
0	1/10	1/10	1/10	1/10	1/10
1	1/10	0	0	0	0
2	0	1/10	0	0	0
3	0	0	1/10	0	0
4	0	0	0	1/10	1/10

Table 10: Highest-dose extended Senn design for $n = 4$ doses. i - treatment number, k - cohort number

Theorem 3.10. *Let ξ be a highest-dose extended Senn design. Then, ξ is an LV-optimal extended design.*

Proof. For the first n latest variances, the argument is the same as in the standard design case; see the proof of Theorem 3.9. Optimality of ξ with respect to the $(n+1)$-st latest variance can be proved by using the General Equivalence Theorem for c-optimality, $c^T = (e_{n+1}^T - e_1^T, 0_{n+1}^T)$.

Design ξ satisfies

$$X = \frac{1}{2(n+1)} \begin{bmatrix} 1_{n-1}^T & 1_2^T \\ I_{n-1} & 0_{(n-1)\times 2} \\ 0_{n-1}^T & 1_2^T \end{bmatrix}, \quad w = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2(n+1)} 1_{n-1} \\ \frac{1}{n+1} \end{bmatrix}.$$

Then,

$$M_\tau = \begin{bmatrix} 1/4 & -\frac{1}{4(n+1)} 1_{n-1}^T & -\frac{1}{2(n+1)} \\ -\frac{1}{4(n+1)} 1_{n-1} & \frac{1}{4(n+1)} I_{n-1} & 0_{n-1} \\ -\frac{1}{2(n+1)} & 0_{n-1}^T & \frac{1}{2(n+1)} \end{bmatrix},$$

and

$$M_\tau^- = 2(n+1) \begin{bmatrix} 0 & 0_{n-1}^T & 0 \\ 0_{n-1} & 2I_{n-1} & 0_{n-1} \\ 0 & 0_{n-1}^T & 1 \end{bmatrix}$$

is a generalized inverse of M_τ. Hence, $c^T M^- c = (e_{n+1}^T - e_1^T) M_\tau^- (e_{n+1} - e_1) = 2(n+1)$. Furthermore,

$$M_{22}^- = \begin{bmatrix} 0 & 0_{n+1}^T \\ 0_{n+1} & (n+1) I_{n+1} \end{bmatrix}$$

is a generalized inverse of M_{22}. By choosing the generalized inverse G of M given by (1.5), after some calculations similar to the proof of Theorem 3.9, the normality inequality becomes $c^T G^T M(\xi') G c = 2(n+1) \leq c^T M^- c = 2(n+1)$ for any design ξ' satisfying (i), (ii). □

3.5 MV-optimality

A widely used optimality criterion in designing experiments for comparing the test treatments with a control is MV-optimality. A design is MV-optimal if it minimizes the maximum variance of the least squares estimators of the contrasts of interest. In the case of comparing drug doses with the placebo, a feasible design is MV-optimal if it minimizes

$$\Psi(\xi) = \max_{1 \le i \le n} \mathrm{Var}(\widehat{\tau_i - \tau_0}),$$

among all $\xi \in \Xi$.

Let us consider the standard designs. Then, as in the LV-optimal case, the Senn designs are optimal.

Theorem 3.11. *Let ξ be a Senn design. Then, ξ is an MV-optimal standard design for comparing drug doses with the placebo.*

Proof. Let ξ be a standard design, and let us denote $v_i(\xi) := \mathrm{Var}_\xi(\widehat{\tau_i - \tau_0})$, $i = 1, \ldots, n$. Then, $\Psi(\xi) = \max_i v_i(\xi)$. Theorem 3.9 states that the Senn design ξ_S is LV-optimal, i.e., $d_k(\xi) \ge d_k(\xi_S)$ for all $k = 1, \ldots, n$. If $k = n$, then the latest variance and the final variance (the variance when the entire design is carried out) coincide. It follows that $v_n(\xi) \ge v_n(\xi_S)$.

In the proof of Theorem 3.2, we can observe that $Q^T M_\tau^-(\xi_S) Q = 4nI_n$. It follows that $v_1(\xi_S) = \ldots = v_n(\xi_S) \propto 4n$, and thus $\Psi(\xi_S) = v_n(\xi_S)$. Because $v_n(\xi) \ge v_n(\xi_S)$, we obtain $\Psi(\xi) \ge v_n(\xi) \ge \Psi(\xi_S)$. □

From numerical studies it appears that in contrast to the MV-optimal standard designs, the MV-optimal extended designs do not follow a regular structure that we are able to describe analytically at this time.

3.6 Discussion

The optimality of the obtained designs depends on the considered model (3.1), which requires no assumptions on the relationship between the treatment effects. If some additional assumptions are made (e.g., linear or quadratic trend), in general, the optimal designs change (see Atkinson [2009]). However, as noted by Bailey [2009b], since there are no additional assumptions, the analysis and the results are quite general.

The optimality results for designs of experiments are also highly dependent on the chosen system of treatment contrasts of interest. The designs obtained in this chapter, which are optimal for the comparison of the drug doses with the placebo, would perform worse than many of the designs proposed by Bailey [2009a] if the aim of the experiment was to estimate the system of all pairwise comparisons. Ultimately, the choice of the treatment contrasts of interest depends on the experimenters. However, following Senn

[2009], we noted that the comparison of the drug doses with the placebo may be a relevant system of contrasts for the dose-escalation studies. Therefore, we intended to provide well-performing designs for the case when the dose effects relative to the placebo are the primary focus. In such a case, it appears that the Senn designs are a very reasonable choice.

Indeed, the Royal Statistical Society's Working Party (see Senn et al. [2007]) concluded that the Senn design may have been one of the designs better suited for the Te Genero trial compared with the design that was actually used. The actual design of the Te Genero trial assigned 2 patients to placebo and 6 patients to the latest dose in each of 3 cohorts of the study. The Senn designs were also considered and advocated for by Senn [1997]. In this chapter, in addition to other results, we show that these designs are *optimal* with respect to multiple criteria if comparisons of the doses with the placebo are the experimental objective.

Another desirable property of the Senn designs is their extremely simple and regular structure, which ensures that the approximate Senn designs can be easily converted into exact designs. Also, due to their simple structure, these designs can be easily described for the practitioners. That is: in each cohort, one assigns one half of the subjects to the placebo and the other half to the highest dose. Moreover, as noted by Senn [2009], the Senn designs are not 'anticipatory'. That is, due to their self-similar structure, their optimality properties hold even when the study is stopped, and thus the entire design is not performed. This 'non-anticipatory' property is crucial in the LV-optimality of the Senn designs. Dose-escalation designs usually do not possess this property, e.g., the design presented in Table 7 is E-optimal only if the entire experiment is performed.

If the extra cohort is considered and the main focus is on the comparisons of the doses with the placebo, one may choose one of the two extensions of the Senn designs that are provided in this chapter. These extensions have forms that are rather simple to describe, and they inherit the optimality properties of the Senn designs if the study is stopped before the entire design is carried out. Alternatively, we provided an entire class of E-optimal extended designs described by simple linear conditions, which allows one to calculate a design that is optimal in this class with respect to a secondary criterion, if some secondary criterion is deemed relevant.

In this chapter, we sought designs that allow one to estimate the safety and tolerability effects of the drug doses relative to placebo as precisely as possible. At the same time, the safety of the subjects (healthy volunteers) was provided by adhering to the dose-escalation scheme. Thus, only the placebo and the doses 1 to k were allowed to be administered to subjects in cohort k. In fact, the desire to obtain as much information as possible and the safety concerns imply conflicting objectives. Without the dose-escalation restrictions, the optimal designs would tend to give each dose to

subjects in all cohorts, which may clearly place subjects at risk.

However, one may consider the dose-escalation constraints to be insufficient for ensuring the safety of the subjects. For example, one may find that giving half of the subjects in each cohort the latest dose (as in Senn designs) is not safe enough. This can be achieved by introducing additional safety constraints, which ensure even slower escalation of the doses. For example, consider a rule that states that at most j individuals may be given a dose that was applied in $j-1$ previous cohorts. This means that when a given dose is administered for the first time, it may be given to only one patient. In the next cohort, this dose may be given to at most two patients, and so forth. Such a restriction would mean that we do not only escalate the allowed doses but also the numbers of patients. It may be interesting to numerically or analytically find optimal designs under such additional constraints. Note that the recent advances in algorithms for computing exact designs allow for such constraints (e.g., see Harman et al. [2016]).

4 Optimal treatment designs represented by graphs

4.1 Introduction

In this chapter, we study optimal designs for estimating a system of treatment contrasts in model (1.6). This chapter is based on paper Rosa [2017].

Consider model (1.2) and suppose that the experimental objective is to estimate a system of treatment contrasts $Q^T\tau$. The positions of any two treatments in the considered system of contrasts need not be symmetric. For example, in the experiment of comparing treatments with a control, it usually is beneficial to employ control in more trials than any of the other treatments (see Theorem 2.15).

In general, for a specified optimality criterion, the optimal numbers of treatment replications depend on the chosen set of contrasts. These optimal treatment replications in model (1.2) are given by optimal designs in model (1.6). Recall that we call such designs the treatment designs. As in the previous chapters, in this chapter we study optimal approximate designs.

In Chapter 2, it was shown that the attainment of optimal treatment proportions is a necessary condition of optimality (Theorem 2.6) in model (1.2). It was also noted that finding the optimal treatment proportions can therefore be thought of as the first step in obtaining optimal approximate designs under the presence of nuisance effects.

Nearly optimal or highly efficient treatment replications can be obtained by the standard rounding methods (e.g., Chapter 12 of Pukelsheim [1993]) from optimal approximate treatment designs. The advantage of the approximate designs is in obtaining general optimality results in simpler form, which can additionally provide an insight into the qualitative behavior of optimal exact designs. Indeed, the earlier mentioned optimal exact block designs for comparing treatments with a control by Majumdar and Notz [1983] and Jacroux [1987] tend to imply treatment proportions similar or equal to the optimal approximate treatment proportions given by Giovagnoli and Wynn [1985] or Theorem 2.15.

We primarily focus on a common class of treatment contrasts in which the aim of the experiment is to estimate a system of pairwise comparisons of treatments. We propose to represent such systems by graphs: the treatments are represented by vertices and the edges connecting pairs of vertices represent the particular pairwise comparisons. That is a rather straightforward relationship, but the graph representation can be exploited further, extensively employing the graph Laplacians.

We show that the treatment designs can then be expressed as (inverses of) the vertex weights; for an analysis of vertex-weighted graphs see Chung and Langlands [1996]. The main result of this chapter is then in proving that the value of any eigenvalue-based optimality criterion (e.g., the well-known Kiefer's Φ_p-optimality criteria, including the

widely used criteria of D-, A- and E-optimality) can be expressed using the eigenvalues of the Laplacian of the vertex-weighted graph.

This allows one to express the optimal design problem using the graph terminology and consequently obtain optimality results based on such a representation. To demonstrate its applicability, we employ this graph representation to provide optimality results for A- and E-optimality. Additionally, we provide a statistical interpretation of E-optimality in the considered experimental settings, based on the variances and the covariances of the contrasts of interest. This interpretation is analogous to the interpretation of the E-optimal dose-escalation designs in the previous chapter, and is similar to the interpretation of A-optimality, which is widely used for estimating systems of contrasts. We also study D-optimality for general systems of treatment contrasts. Moreover, we apply the obtained representation to provide an interpretation of the D-, A- and E-optimality using the graph terminology.

Although earlier we emphasized that the uniform treatment proportions need not be optimal, there are many systems of contrasts which imply the optimality of uniform treatment proportions for a wide range of optimality criteria. By employing some graph properties, we obtain a class of sets of pairwise comparisons in which the uniform treatment design is optimal with respect to any eigenvalue-based information function. We then extend these results to general sets of treatment contrasts.

It is well known that block designs can be represented by graphs (see, e.g., Cameron and Van Lint [1975], Cheng [1981], Bailey and Cameron [2009]), which may seem similar to the representation proposed here. However, these representations significantly differ, which will be demonstrated later.

The problem of obtaining optimal treatment designs is a convex optimization problem. Thus, it can usually be efficiently solved using numerical methods. However, the objective of this chapter is to provide analytical results on the optimal treatment designs. Such results allow one to observe the general behavior of the optimal designs with respect to the chosen optimality criteria and the chosen systems of contrasts of interest. Moreover, the optimality results provided in this chapter are often of a simple form, and thus the resulting optimal designs are easily calculated 'by hand' (e.g., the obtained A-optimal and E-optimal treatment proportions can be calculated simply by counting the degrees of the vertices in the corresponding graphs).

4.2 Experimental settings

The considered experimental objective is to estimate a system of s treatment contrasts $Q^T\tau$ in model (1.6). We denote the elements of Q as q_{ij} and its columns as $q_1, \ldots, q_s$; furthermore, we denote $r := \mathrm{rank}(Q)$.

Recall that in this chapter, we examine optimal approximate designs. Some basic

results on treatment designs are provided in Section 1.6. We also note that the results provided in this chapter for model (1.6) hold also for the model (1.7) with added constant term.

For easier interpretation, we alternatively use the criteria that are equivalent to the Φ_p criteria in the sense of the implied ordering of designs, but are to be *minimized*. Let $V_Q(w) := Q^T M^{-1}(w) Q$, let $Q^T\tau$ be a system with $\operatorname{rank}(Q) = r$ and let $w > 0$. Then,

$$\Psi_p(w) = \begin{cases} \sum_{j=1}^{r} \lambda_j^{-p}(V_Q(w)), & p \in (-\infty, 0), \\ \prod_{j=1}^{r} \lambda_j(V_Q(w)), & p = 0, \\ \lambda_{\max}(V_Q(w)), & p = -\infty. \end{cases}$$

In particular, $\Psi_{-1}(w) = \operatorname{tr}(V_Q(w))$, and $\Psi_0(w) = \operatorname{Det}(V_Q(w))$. A design is said to be Ψ_p-optimal for $Q^T\tau$ if it minimizes $\Psi_p(w)$ over all $w > 0$. Clearly, a design is Ψ_p-optimal for $Q^T\tau$ if and only if it is Φ_p-optimal for $Q^T\tau$.

4.3 Graph representation

4.3.1 Graph theory

We provide some basic terminology of graph theory; for more details, e.g., see Diestel [2000], Cvetković et al. [2010], and Bapat [2010]. A directed graph G is a pair $G = (V, E)$, where V is a finite set of vertices $V = \{1, \ldots, v\}$, and E is a set of edges $E = \{e_1, \ldots, e_s\}$, which are ordered pairs of vertices $e_k = (i_k, j_k)$, $i_k, j_k \in V$ for all k. If $e = (i, j)$, the edge e is directed from i to j. We do not allow loops or multiple edges between two vertices.

We say that two vertices $i, j \in V$ are adjacent, denoted as $i \sim j$, if there exists an edge $(i, j) \in E$ or $(j, i) \in E$. An edge e is incident with a vertex i if there exists a vertex $j \in V$ such that $e = (i, j)$ or $e = (j, i)$. We say that a vertex i has degree d_i, where d_i is the number of edges incident with i. If all edges incident with a vertex i are directed toward i, we say that i is a sink vertex; if all edges incident with i are directed from i, we say that i is a source vertex.

A graph G can be characterized by certain matrices. The adjacency matrix A of G is a $v \times v$ matrix with its rows and columns indexed by V and satisfying

$$A_{i,j} = \begin{cases} 1, & i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

The incidence matrix R of G is a $v \times s$ matrix with its rows and columns indexed by

V and E, respectively, that satisfies

$$R_{i,e} = \begin{cases} 1, & e = (i,j) \text{ for some } j \in V, \\ -1, & e = (j,i) \text{ for some } j \in V, \\ 0 & \text{otherwise.} \end{cases}$$

We say that the Laplacian matrix L of G is the $v \times v$ matrix $L := RR^T$, i.e.,

$$L_{i,j} = \begin{cases} d_i, & i = j, \\ -1, & i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $L = D - A$, where $D = \text{diag}(d_1, \ldots, d_v)$ and hence L does not depend on the orientation of the edges.

To specify that V is a set of vertices of G, we may write $V(G)$ instead of V, similarly for $E(G)$, $A(G)$, and so forth.

Following Chung and Langlands [1996], we say that G is a graph with vertex weights α if G is a graph and α is a function $\alpha : V \to \mathbb{R}_+$; for simplicity, we write $\alpha_i := \alpha(i)$ and $\alpha = (\alpha_1, \ldots, \alpha_v)^T$. Then, the vertex-weighted Laplacian $\mathcal{L}_\alpha$ (see Chung and Langlands [1996]) is a $v \times v$ matrix with its rows and columns indexed by V, satisfying

$$\mathcal{L}_{\alpha;i,j} = \begin{cases} d_i \alpha_i, & i = j, \\ -\alpha_i^{1/2} \alpha_j^{1/2}, & i \sim j, \\ 0 & \text{otherwise,} \end{cases}$$

which can be represented as $\mathcal{L}_\alpha = \text{diag}(\alpha^{1/2}) RR^T \text{diag}(\alpha^{1/2})$. A more common version of a weighted graph is one with edge weights, resulting in an edge-weighted Laplacian $\tilde{L}$, which differs from the vertex-weighted version, see, e.g., Merris [1994], Merris [1995]. Note that the often used normalized Laplacian (see, e.g., Chung [1997], Cvetković et al. [2010]) is a special case of the vertex-weighted Laplacian, with vertex weights equal to the inverses of the degrees of the vertices, $\alpha_i = d_i^{-1}$.

An automorphism of G is a bijection $\pi : V \to V$ satisfying $(i,j) \in E$ if and only if $(\pi(i), \pi(j)) \in E$. Since π is a bijection, it is in fact a permutation on V. Consider the permutation of vertex labels given by a permutation π. By P_π we denote the $v \times v$ permutation matrix given by π, i.e., P_π satisfies $P_\pi x = \pi(x)$ for any $x \in \mathbb{R}^v$. Then, under the relabeling given by π, R changes to $P_\pi R$ and A changes to $P_\pi A P_\pi^T$. It follows that if π is an automorphism, $P_\pi A P_\pi^T = A$ and $P_\pi R R^T P_\pi^T = RR^T$.

Let π be a permutation on V. A cycle c is a sequence of the form $(i, \pi(i), \ldots, \pi^{k-1}(i))$, where $i \in V$ and k is the smallest number such that $\pi^k(i) = i$. For any $j \in V$, we

say that the cycle $c = (i, \pi(i), \ldots, \pi^{k-1}(i))$ contains j, denoted as $j \in c$, if there exists $m \in \mathbb{N}$ such that $\pi^m(i) = j$. Then, any permutation π can be decomposed into its cycles $c_1, \ldots, c_K$. By a cyclic permutation π, we mean a permutation that consists of only one cycle, i.e., for any $i, j \in V$ there exists $m \in \mathbb{N}$ such that $\pi^m(i) = j$.

4.3.2 Representation of treatment proportions

A common class of systems of contrasts for treatment comparisons are systems of pairwise comparisons $\tau_j - \tau_i$, which are obtained when all the columns q_k of Q satisfy $q_k^T\tau = \tau_{j_k} - \tau_{i_k}$ for some $i_k, j_k \in \{1, \ldots, v\}$. We propose to represent any such system by a graph $G = (V, E)$, where V is the set of treatments and $e = (j, i) \in E$ if and only if $\tau_j - \tau_i$ is present in $Q^T\tau$. Intuitively, the representation is rather straightforward – the vertices represent the treatments and the edges represent the pairwise comparisons: if there is a comparison $\tau_j - \tau_i$ in $Q^T\tau$, then the corresponding vertices are connected by an edge directed from j to i. Then, it is easy to see that the incidence matrix R coincides with the coefficient matrix Q, i.e., $R = Q$. Therefore, any system of pairwise comparisons $Q^T\tau$ can be represented by a directed graph with the incidence matrix given by Q.

The directions of the edges depend only on the signs of the treatment contrasts ($\tau_j - \tau_i$ versus $\tau_i - \tau_j$), and therefore the optimality results obtained in this chapter do not depend on the directions of the edges in the corresponding graph.

Example 4.1. Let $v = 7$ and suppose that the aim of the experiment is to estimate the following treatment comparisons: $\tau_2 - \tau_1$, $\tau_3 - \tau_2$, $\tau_4 - \tau_3$, $\tau_5 - \tau_3$, $\tau_6 - \tau_5$, $\tau_7 - \tau_5$. Such a system of contrasts was considered by Mead [1990]. We denote these contrasts as $Q_1^T\tau$ and in the following sections we will use them to demonstrate our results. The matrix Q_1 is of the form

$$Q_1 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The corresponding graph G_1 (the graph in Figure 2, disregarding the vertex weights) allows for a clear representation of the system of contrasts.

Remark 4.2. A system of $v - 1$ pairwise comparisons $Q^T\tau$ is a full-rank system if and only if the corresponding graph is a tree (a connected graph without cycles; equivalently,

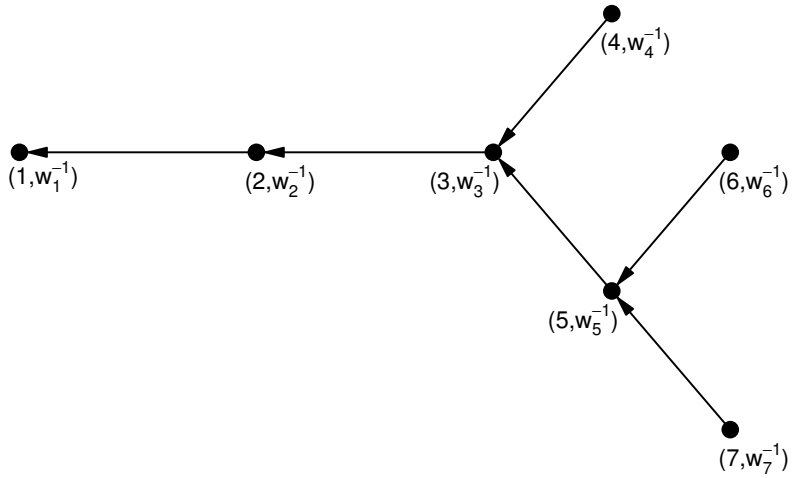


Figure 2: Graph representation of the system of pairwise comparisons $\tau_2-\tau_1$, $\tau_3-\tau_2$, $\tau_4-\tau_3$, $\tau_5-\tau_3$, $\tau_6-\tau_5$, $\tau_7-\tau_5$ and a design $w>0$ for this system of contrasts. The labels of the vertices are of the form (i,α_i), where i is the index of the vertex and $\alpha_i = w_i^{-1}$ is the vertex weight.

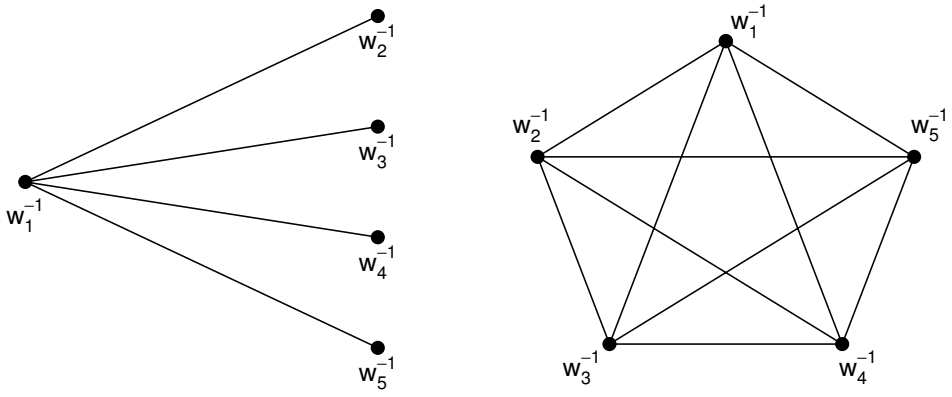
a connected graph with $v-1$ edges). It follows from the fact that the corresponding graph is connected if and only if $\mathrm{rank}(L) = v-1$ (e.g., Mohar [1991]). Furthermore, $\mathrm{rank}(Q) = \mathrm{rank}(QQ^T) = \mathrm{rank}(L)$ and the number of treatment contrasts is equal to the number of edges, i.e., $v-1$. Clearly, the graph G_1 in Example 4.1 is a tree, and thus the system of contrasts attains the full rank $v-1$.

For a feasible design $w>0$, consider the vertex weights α given by the inverse values of w, i.e., $\alpha_i = w_i^{-1}$. Then, $\mathcal{L}_w = M^{-1/2}(w)QQ^TM^{-1/2}(w)$; recall that $M(w) = \mathrm{diag}(w)$. Note the slight abuse of notation, where $\mathcal{L}_w$ should in fact be expressed as $\mathcal{L}_\alpha$, where $\alpha = w^{-1}$ or as $\mathcal{L}_{w^{-1}}$.

Let Φ be an orthogonally invariant information function. Then, we may express $\Phi(w)$ as $\Phi(w) = \phi(\lambda_1(N_Q(w)),\dots,\lambda_s(N_Q(w)))$ in the full-rank case and $\Phi(w) = \phi(\lambda_1(C_Q(w)),\dots,\lambda_r(C_Q(w)),0,\dots,0)$ in the rank-deficient case, for some function ϕ. The following theorem shows that the value of $\Phi(w)$ is determined by the spectrum of the Laplacian $\mathcal{L}_w$ of the corresponding vertex-weighted graph.

Theorem 4.3. *Let Φ be an orthogonally invariant information function and let w be a feasible treatment design for estimating a system of pairwise comparisons $Q^T\tau$. Then, if $Q^T\tau$ is a full-rank system, $\Phi(w) = \phi(1/\lambda_s(\mathcal{L}_w),\dots,1/\lambda_1(\mathcal{L}_w))$, and if $Q^T\tau$ is rank deficient, $\Phi(w) = \phi(1/\lambda_r(\mathcal{L}_w),\dots,1/\lambda_1(\mathcal{L}_w),0,\dots,0)$.*

Proof. Consider the full-rank case. Since Φ depends only on the eigenvalues of N_Q, we may write $\Phi(w) = \phi(\lambda_1(N_Q(w)),\dots,\lambda_s(N_Q(w))) = \phi(1/\lambda_s(V_Q(w)),\dots,1/\lambda_1(V_Q(w)))$



(a) Comparison with a control **(b)** All pairwise comparisons

Figure 3: Graph representations of treatment designs $w > 0$ for selected systems of treatment contrasts. For clarity, the directions of the edges are suppressed

because $\lambda_i(N_Q(w)) = 1/\lambda_{s-i+1}(V_Q(w))$. Moreover, recall that a matrix X^TX has the same positive eigenvalues as XX^T, including multiplicities. Defining $X = M^{-1/2}(w)Q$ yields that $\Phi(w) = \phi(1/\lambda_s(\mathcal{L}_w), \ldots, 1/\lambda_1(\mathcal{L}_w))$ because $\mathcal{L}_w = XX^T$. Since the positive eigenvalues of $V_Q(w)$ are inverses of the positive eigenvalues of $C_Q(w) = V_Q^+(w)$, analogous results hold in the rank-deficient case. □

Theorem 4.3 shows that not only can a system of pairwise comparisons be expressed as a graph G, but also that a feasible treatment design $w > 0$ can be represented by vertex weights on G given by the *inverse* values of w. For example, the system of contrasts for comparing test treatments with g controls $\tau_j - \tau_i$, $i \in \{1, \ldots, g\}$, $j \in \{g+1, \ldots, v\}$ can be represented by a complete bipartite graph with partitions $\{1, \ldots, g\}$ and $\{g+1, \ldots, v\}$, and the graph for all pairwise comparisons $\tau_j - \tau_i$, $j > i$ (e.g., Bailey and Cameron [2009]), is a complete graph (see Figure 3). The vertex weights implied by a design $w > 0$ for the system of contrasts $Q_1^T\tau$ from Example 4.1 are represented in Figure 2.

From Theorem 4.3 it follows that for a system of pairwise comparisons and an orthogonally invariant information function Φ, the treatment design problem

$$\begin{aligned} \max \quad & \Phi(N_Q(w)) \quad \text{or} \quad \Phi(C_Q(w)) \\ \text{s.t.} \quad & w > 0, \quad \sum_{i=1}^{v} w_i = 1 \end{aligned}$$

can be expressed as

$$\begin{aligned} \max \quad & \phi(1/\lambda_r(\mathcal{L}_w), \ldots, 1/\lambda_1(\mathcal{L}_w), 0, \ldots, 0) \\ \text{s.t.} \quad & w > 0, \quad \sum_{i=1}^{v} w_i = 1, \end{aligned}$$

where $\mathcal{L}_w$ is the Laplacian of the corresponding vertex-weighted graph.

Since $\alpha_i = w_i^{-1}$, the condition $\sum_i w_i = 1$ can be expressed as $\sum_i \alpha_i^{-1} = 1$ or $v/\sum_i \alpha_i^{-1} = v$, i.e., the harmonic mean of the vertex weights is equal to the number of vertices. Then, from the graph-theoretic point of view, the studied optimal design problem can be expressed as the problem of maximizing a given function defined on the eigenvalues of the Laplacian of a vertex-weighted graph over all vertex weights with a fixed harmonic mean:

$$\begin{aligned} \max \quad & \phi(1/\lambda_r(\mathcal{L}_\alpha), \ldots, 1/\lambda_1(\mathcal{L}_\alpha), 0, \ldots, 0) \\ \text{s.t.} \quad & \alpha > 0, \quad v/\sum_{i=1}^{v} \alpha_i^{-1} = v. \end{aligned}$$

Consider the Kiefer's Φ_p-optimality criteria. Using Theorem 4.3, $\Phi_p(w)$ may be expressed by employing the weighted Laplacian.

Proposition 4.4. *Let $Q^T\tau$ be a system of pairwise comparisons with* $\operatorname{rank}(Q) = r$ *and let $w > 0$ be a feasible design. Then, for any $p \in [-\infty, 0]$, the value of the Kiefer's Φ_p-optimality criterion can be expressed as*

$$\Phi_p(w) = \begin{cases} \left(\frac{1}{r}\sum_{j=1}^{r} \lambda_j^{-p}(\mathcal{L}_w)\right)^{1/p}, & p \in (-\infty, 0), \\ \left(\prod_{j=1}^{r} \lambda_j(\mathcal{L}_w)\right)^{-1/r}, & p = 0, \\ 1/\lambda_{\max}(\mathcal{L}_w), & p = -\infty. \end{cases}$$

In particular, $\Psi_0(w) = \prod_{j \le r} \lambda_j(\mathcal{L}_w)$, $\Psi_{-1}(w) = \operatorname{tr}(\mathcal{L}_w)$ and $\Psi_{-\infty}(w) = \lambda_{\max}(\mathcal{L}_w)$.

Note that the graph characterization of designs in model (1.6) significantly differs from the graph characterization of block designs. In particular, it is well known that the exact block designs with blocks of size two can be represented by (undirected) graphs $G = (V, E)$, where the vertices represent the particular treatments and there is an edge between two vertices i, j if and only if there is a block with treatments i and j, e.g., see Cameron and Van Lint [1975], Bailey and Cameron [2009]. An approximate block design can be represented by *edge weights*. Then, a Φ-optimal block design ξ for the treatment effects can be calculated by maximizing $\Phi(\tilde{L}_\xi)$, where $\tilde{L}_\xi$ is the corresponding edge-weighted Laplacian. Specifically, if Φ is orthogonally invariant, the optimality criterion can be calculated by using the eigenvalues of $\tilde{L}_\xi$ (i.e., not

through their inverses). For example, $\Psi_0(\xi) = \prod_{j \le v-1} \lambda_j^{-1}(\tilde{L}_\xi)$, $\Psi_{-1}(\xi) = \mathrm{tr}(\tilde{L}_\xi^{-1})$ and $\Psi_{-\infty}(\xi) = \lambda_{\min}^{-1}(\tilde{L}_\xi)$, which are, in a sense, opposite problems to the optimality of treatment proportions w; compare with Proposition 4.4.

In contrast to the block designs, the treatment designs are represented by the *inverses of the vertex weights* (the edges are fixed: they are specified by the considered system of contrasts) and the optimality criterion is calculated using the *inverses* of the eigenvalues of the vertex-weighted Laplacian.

4.4 Optimality

4.4.1 Permutation of treatments

To prove certain optimality results, we will consider treatment permutations; thus, we briefly examine such transformations. Let w be a design, let π be a permutation of $\{1, \dots, v\}$ and let P_π be the permutation matrix corresponding to π. We denote the design obtained from w by the π-permutation of the treatment labels as $P_\pi w$. Then, $M(P_\pi w) = P_\pi M(w) P_\pi^T$.

The vertex weights corresponding to $P_\pi w$ are $P_\pi w^{-1}$, the diagonal matrix of weights changes to $P_\pi \mathrm{diag}(w^{-1}) P_\pi^T$ and the vertex-weighted Laplacian changes to $\mathcal{L}_{Pw} = P_\pi \mathrm{diag}(w^{-1/2}) P_\pi^T R R^T P_\pi \mathrm{diag}(w^{-1/2}) P_\pi^T$. Note that we are permuting the design values $w \mapsto P_\pi w$, not the vertices of the graph. As noted in Section 4.3.1, if π is an automorphism of the graph G, we have $P_\pi R R^T P_\pi^T = R R^T$; thus $P_\pi^T R R^T P_\pi = R R^T$ and $\mathcal{L}_{Pw} = P_\pi \mathcal{L}_w P_\pi^T$. It follows that if π is an automorphism of the corresponding graph G, the Laplacians for w and for $P_\pi w$ are orthogonally similar.

4.4.2 *D*-optimality

The D-optimality criterion is of limited interest to us if $s = v - 1$ and Q is a full-rank system, because it is well known that in such a case, the uniform design $\bar{w} = 1_v / v$ is always D-optimal, which can be proved by a reparametrization of the system of interest. We extend this proposition to any system of $s \ge v - 1$ contrasts of rank $v - 1$. Of course, for the number of contrasts $s > v - 1$ we consider the rank-deficient version of D-optimality, $\Psi_0(w) = \prod_j \lambda_j(V_Q(w))$, where the product is over all positive eigenvalues of $V_Q(w)$. Thus, $\Psi_0(w)$ is the pseudo-determinant of $V_Q(w)$ (see Knill [2014]), which is the product of all nonzero eigenvalues of $V_Q(w)$. The reparametrization argument cannot be easily replicated in such a case because the pseudo-determinant does not satisfy $\mathrm{Det}(AB) = \mathrm{Det}(A)\mathrm{Det}(B)$ in general. First, we provide a technical lemma.

Lemma 4.5. *Let A be a $v \times v$ symmetric matrix satisfying $A 1_v = 0_v$. Then,* $\det(A_{-ij}) = (-1)^{i+j} \det(A_{-11})$ *for all $i, j \in \{1, \dots, v\}$.*

Proof. The Lemma follows from the well-known fact that for any matrix A satisfying

the conditions of this Lemma, the cofactors of any two elements of A are equal (e.g., see Lemma 4.2 by Bapat [2010]). □

Now, we may formulate the optimality of the uniform design.

Theorem 4.6. *Let $Q^T\tau$ be a system of $s \geq v-1$ contrasts of rank $v-1$. Then, the uniform design $\bar{w} = 1_v/v$ is Ψ_0-optimal for estimating $Q^T\tau$.*

Proof. Let $w > 0$, let π be a permutation of treatments, let $P := P_\pi$ and $M := M(w)$. The moment matrix of Pw is $M(Pw) = PMP^T$, and therefore $V_Q(Pw) = Q^TPM^{-1}P^TQ$. We use the facts provided by Knill [2014] that for any two $a \times b$ matrices F, G the following hold: $\mathrm{Det}(F^TG) = \mathrm{Det}(FG^T)$; and $\mathrm{Det}(F^TG) = \sum_X \det(F_X)\det(G_X)$, where the sum is over all $k \times k$ sub matrix masks X of F and F_X, G_X are the corresponding submatrices. Here, k is calculated so that $(-1)^k c_k \lambda^{b-k}$ is the smallest order entry in the characteristic polynomial of the $b \times b$ matrix F^TG. Then,

$$\Psi_0(Pw) = \mathrm{Det}(V_Q(Pw)) = \mathrm{Det}(QQ^TPM^{-1}P^T) = \sum_X \det(QQ^T)_X \det(PM^{-1}P^T)_X,$$

where the sum is over all $(v-1) \times (v-1)$ submatrices, because Q has rank $v-1$. Such a sum may be expressed as $\sum_{i,j} \det(QQ^T)_{-ij}\det(PM^{-1}P^T)_{-ij}$, where the sum is over all $i, j \in \{1, \ldots, v\}$. Because $\det(PM^{-1}P^T)_{-ij} = 0$ if $i \neq j$, we obtain $\Psi_0(Pw) = \sum_i \det(QQ^T)_{-ii}\det(PM^{-1}P^T)_{-ii}$.

Since Q is a matrix of contrasts, we have $1_v^TQ = 0_s^T$, and thus QQ^T satisfies the conditions of Lemma 4.5. It follows that $\det(QQ^T)_{-ii} = \det(QQ^T)_{-11}$; moreover, $\det(PM^{-1}P^T)_{-ii}$ is the product of all the w_j^{-1}-s except $w_{\pi^{-1}(i)}^{-1}$. Therefore,

$$\begin{aligned}\Psi_0(Pw) &= \det(QQ^T)_{-11}\sum_i \det(PM^{-1}P^T)_{-ii} = \det(QQ^T)_{-11}\sum_i \prod_{j\neq\pi^{-1}(i)} w_j^{-1} \\ &= \det(QQ^T)_{-11}\sum_i \prod_{j\neq i} w_j^{-1} = \det(QQ^T)_{-11}\sum_i \det(M^{-1})_{-ii} = \Psi_0(w).\end{aligned}$$

Hence, $\Phi_0(Pw) = \Phi_0(w)$, which yields

$$\Phi_0(\bar{w}) = \Phi_0\left(\frac{1}{v!}\sum_\pi P_\pi w\right) \geq \frac{1}{v!}\sum_\pi \Phi_0(P_\pi w) = \Phi_0(w),$$

where the inequality follows from the concavity of Φ_0. □

However, in general, the uniform treatment proportions need not be optimal, as is shown in the following example.

Example 4.7. Let $Q = (-1, 1_{v-1}^T/(v-1))^T$ (i.e., in fact, we are examining c-optimality), which aims at estimating the average comparison with control $\overline{\tau_i - \tau_0}$. Then, $\Psi_0(w) =$

$w_1^{-1} + (v-1)^{-2} \sum_{i>1} w_i^{-1}$ for any $w > 0$, and the unique optimal treatment design w^* satisfies $w_1^* = 1/2$ and $w_i^* = (2(v-1))^{-1}$ for $i > 1$, which obviously is not the uniform design.

Let us return to Example 4.1.

Example 4.8 (Example 4.1 cont.)**.** Since the system of contrasts $Q_1^T\tau$ in Example 4.1 is a full-rank system of $v-1$ contrasts, the D-optimal design for $Q_1^T\tau$ is $\bar{w} = 1_7/7$ and the corresponding graph G_1 has weight $\alpha_i = 7$ on each vertex.

The D-optimality criterion can be interpreted using the graph terminology, similarly to the well-known interpretation of D-optimal block designs, which maximize the number of spanning trees of a graph, see, e.g., Cheng [1981], Bailey and Cameron [2009]. In Chung and Langlands [1996], the authors provide an interpretation of the coefficients of the characteristic polynomial of the Laplacian of a vertex-weighted graph

$$\det(\lambda I - \mathcal{L}_\alpha) = \sum_{k=0}^{v} (-1)^k c_k \lambda^{v-k}(\mathcal{L}_\alpha),$$

with $c_0 = 1$, and since $\mathcal{L}_\alpha$ is always singular, $c_n = 0$. In particular, they obtain that $c_{v-1} = \prod_{j \le v-1} \lambda_j(\mathcal{L}_\alpha) = \kappa(G)$, where $\kappa(G)$ is the sum of weights of all rooted directed spanning trees, which is a generalization of the well-known Matrix-tree theorem (e.g., Mohar [1991]) for the vertex-weighted graphs. Furthermore, they obtain that $c_k = \kappa_{v-k}(G)$ for all $k = 1, \ldots, v-1$, where $\kappa_{v-k}(G)$ is the sum of weights of all rooted directed spanning forests with $v-k$ roots. For the formal definition of $\kappa_{v-k}(G)$ and $\kappa(G) = \kappa_1(G)$, see Chung and Langlands [1996].

Consider the vertex weights given by a feasible design w, $\alpha = w^{-1}$. It is well known that the coefficients of the characteristic polynomial satisfy

$$c_k = \sum_{i_1 < \ldots < i_k} \lambda_{i_1}(\mathcal{L}_w) \cdot \ldots \cdot \lambda_{i_k}(\mathcal{L}_w),$$

and therefore, for $\operatorname{rank}(Q) = r$, the $v-r$ smallest eigenvalues of $\mathcal{L}_w$ are zero and we obtain that $\kappa_{v-r}(G) = c_r = \prod_{i \le r} \lambda_i(\mathcal{L}_w) = \Psi_0(w)$. Thus, a D-optimal design minimizes the total weight of all rooted directed spanning forests with $v-r$ roots.

Proposition 4.9. *Let $Q^T\tau$ be a system of pairwise comparisons with* $\operatorname{rank}(Q) = r$*, let $w > 0$ be a feasible design and let G be the corresponding vertex-weighted graph. Then, $\Psi_0(w) = \kappa_{v-r}(G)$.*

4.4.3 *A*-optimality

The problem of obtaining A-optimal treatment designs is rather simple; these designs can be directly obtained from Corollary 8.8 of Pukelsheim [1993]. Thus, we focus on

the graph interpretation of the optimality problem and on the graph representation of the A-optimal treatment proportions.

Note that $\mathrm{tr}(\mathcal{L}_\alpha) = \sum_i d_i\alpha_i$, which can be expressed as $\sum_i \tilde{d}_i$, where $\tilde{d}_i := \alpha_i d_i$ is the weighted degree of the vertex i. Then, the A-optimality value $\Psi_{-1}(w) = \mathrm{tr}(\mathcal{L}_w)$ is equal to the total weighted degree of graph G, i.e., the sum of all weighted degrees of its vertices, $\Psi_{-1}(w) = \sum_i \tilde{d}_i = \sum_i w_i^{-1} d_i$, where $\tilde{d}_i = w_i^{-1} d_i$. As a consequence, any A-optimal treatment design w minimizes the total weighted degree of G with weights inverse to the design values, $\alpha_i = w_i^{-1}$. Then, it is straightforward to calculate the A-optimal design.

Theorem 4.10. *Let $Q^T\tau$ be a system of pairwise comparisons, let $w > 0$ be a feasible design and let G be the corresponding vertex-weighted graph. Then, $\Psi_{-1}(w) = \sum_i w_i^{-1} d_i = \sum_i \tilde{d}_i$. Furthermore, the A-optimal treatment proportions are*

$$w_i^* = \frac{\sqrt{d_i}}{\sum_j \sqrt{d_j}}, \quad i = 1, \ldots, v. \tag{4.1}$$

The formula (4.1) provides a straightforward characterization of A-optimal proportions using the graph terminology. The A-optimal design value for treatment i is proportional to the square root of the degree d_i of the vertex i (which is the number of occurrences of τ_i in $Q^T\tau$); i.e., the A-optimal proportions depend only on the numbers of times the particular treatments are present in the system $Q^T\tau$, the dependence being a square root. Interestingly, the A-optimal treatment proportions depend only on the *local* properties of the graph (specifically, the numbers of neighbors of the vertices), not on the global structure of the graph.

As noted earlier, the problem of calculating the A-optimal proportions for a general system of contrasts $Q^T\tau$ is quite trivial. It is easy to see that $\Psi_{-1}(w) = \mathrm{tr}(Q^T M^{-1}(w) Q) = \sum_i w_i^{-1} \sum_j q_{ij}^2$. Then, the A-optimal treatment proportions can be calculated:

$$w_i^* = \frac{\sqrt{\sum_k q_{i,k}^2}}{\sum_j \sqrt{\sum_k q_{j,k}^2}}, \quad i = 1, \ldots, v, \tag{4.2}$$

which can also be obtained from Corollary 8.8 of Pukelsheim [1993].

Example 4.11 (Example 4.1 cont.)**.** The results for the system of contrasts $Q_1^T\tau$ from Example 4.1 are demonstrated in Figure 4. In G_1, there are four vertices with degree 1, one vertex with degree 2 and two vertices with degree 3. Thus, $\sum_j \sqrt{d_j} = 4 + 2\sqrt{3} + \sqrt{2} =: S$ and the A-optimal design values are $w_1^* = w_4^* = w_6^* = w_7^* = 1/S \approx 0.11$, $w_2 = \sqrt{2}/S \approx 0.16$, $w_3^* = w_5^* = \sqrt{3}/S \approx 0.20$.

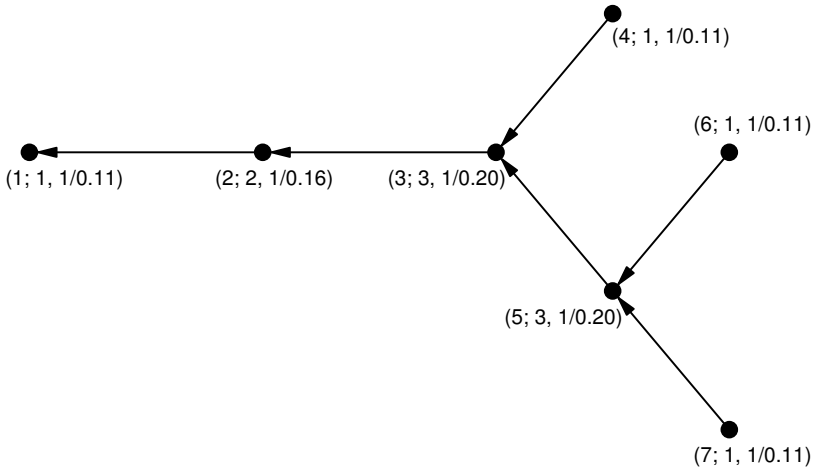


Figure 4: A-optimal design w^* for the system of contrasts $Q_1^T\tau$ from Example 4.1. The labels of the vertices are of the form $(i; d_i, \alpha_i)$, where i is the vertex index, d_i is the degree of vertex i, and $\alpha_i = 1/w_i^*$ is the vertex weight for vertex i.

4.4.4 E-optimality

The criterion of E-optimality does not seem to allow for such a direct formula for optimal weights as (4.1). However, we provide E-optimal treatment proportions for a specific class of systems of pairwise comparisons. We say that a system of pairwise comparisons $Q^T\tau$ is *bipartite* if the corresponding graph G is bipartite. It can be observed that a graph G is bipartite if and only if there exists an orientation of G, such that each vertex $i \in V$ is either a sink (with zero vertices directed from i) or a source (with zero vertices directed toward i). Note that if G is a tree, it is bipartite; thus, any full-rank system of $v-1$ pairwise comparisons is bipartite.

For proving E-optimality, we employ the Equivalence Theorem for E-optimality (Lemma 1.10). However, since we consider also the rank-deficient subsystems, we will slightly reformulate the Equivalence Theorem. As stated in Section 8.18 of Pukelsheim [1993], the Equivalence Theorem for Φ_p-optimality holds also in the rank-deficient case, only instead of $\lambda_{\min}(N_Q(w))$, we have $1/\lambda_{\max}(V_Q(w)))$. Moreover, if $E = hh^T$, where h is a normalized eigenvector of $V_Q(w)$ corresponding to $\lambda_{\max}(V_Q(w))$, we obtain a simpler expression, which holds for both the full-rank as well as the rank-deficient case.

Lemma 4.12. *Let $\xi \in \Xi$ be a feasible design for $Q^T\tau$ with its moment matrix M and $V_Q := Q^T M^- Q$, and let h be an eigenvector of V_Q corresponding to $\lambda_{\max}(V_Q)$, satisfying $\|h\| = 1$. Then, if there exists a generalized inverse G of M, such that*

$$h^T Q^T G^T M(\tilde{\xi}) G Q h \le \lambda_{\max}(V_Q) \text{ for all } \tilde{\xi} \in \Xi, \tag{4.3}$$

then ξ is E-optimal for $Q^T\tau$ in Ξ.

Proof. Let $E = hh^T$. Suppose that $Q^T\tau$ is a full-rank system. From $N_Q h = \lambda_{\max}^{-1}(V_Q)h$, it follows that the left-hand side of (1.4) is equal to $\lambda_{\max}^{-2}(V_Q)h^T Q^T G^T M(\tilde{\xi})GQh$. Then, (1.4) can be rearranged to (4.3).

If $Q^T\tau$ is rank deficient, the normality inequality of the Equivalence Theorem becomes $\operatorname{tr}(M(\tilde{\xi})GQV_Q^+EV_Q^+Q^TG^T) \leq 1/\lambda_{\max}(V_Q)$ for all $\tilde{w} \in \Xi$. Since h is an eigenvector of V_Q corresponding to $\lambda_{\max}(V_Q)$, it is an eigenvector of V_Q^+ corresponding to $1/\lambda_{\max}(V_Q)$, which yields the same inequality (4.3) as in the full-rank case. □

Now, the results on E-optimal treatment designs can be formulated. For the subclass of bipartite systems of pairwise comparisons, we obtain an analytical expression for E-optimal treatment proportions, similar to the formula (4.2) for A-optimality. Only instead of square roots of the vertex degrees, the E-optimal treatment designs are proportional directly to the vertex degrees, i.e., to the numbers of times the corresponding treatments are present in $Q^T\tau$.

Theorem 4.13. *Let $Q^T\tau$ be a bipartite system of pairwise comparisons and let G be the corresponding graph. Then, the treatment design given by*

$$w_i^* = \frac{d_i}{\sum_j d_j}, \quad i = 1, \ldots, v, \tag{4.4}$$

is E-optimal for $Q^T\tau$ with the optimal value $\Psi_{-\infty}(w^*) = \lambda_{\max}(V_Q(w^*)) = 2\sum_i d_i = 4s$. Moreover, if each vertex of G is either a sink or a source, then $h = 1_s/\sqrt{s}$ is an eigenvector of $V_Q(w^*)$ corresponding to $\lambda_{\max}(V_Q(w^*))$.

Proof. Suppose that each vertex of G is either a sink or a source. If it were not, we could change the orientations of the edges, which does not affect the optimality results. We denote $w := w^*$, $k := \sum_j d_j$ and $V_Q := V_Q(w)$. First, we calculate $Qh = Q1_s/\sqrt{s} = g/\sqrt{s}$, where g_i is either d_i or $-d_i$, which we denote as $g_i = \pm d_i$, $i = 1, \ldots, v$. Then,

$$V_Q h = Q^T M^{-1}(w)g/\sqrt{s} = 2k1_s/\sqrt{s} = 2kh,$$

because $w_i^{-1}g_i = \pm k$ for all i. Hence, h is an eigenvector of V_Q corresponding to $\lambda^* := 2k = 4s$. Now, we will prove that $\lambda^* = \lambda_{\max}(V_Q)$.

The k,ℓth element of V_Q, $v_{k,\ell}$, satisfies $v_{k,\ell} = \sum_i w_i^{-1} q_{i,k} q_{i,\ell}$. Indices k and ℓ represent edges of the graph, and the graph representation yields a formula for $v_{k,\ell}$. For $k = \ell$, we have

$$v_{k,k} = w_i^{-1} + w_j^{-1}, \text{ where } k = (i,j).$$

Since every vertex is a sink or a source, for $k \neq \ell$ we obtain

$$v_{k,\ell} = \begin{cases} w_i^{-1}, & \text{there exists a vertex } i, \text{ which is incident with both } k \text{ and } \ell, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

Moreover, $\lambda_{\max}(V_Q) = \max_{\|x\|=1} x^T V_Q x$. Let us alternatively denote the edges by the vertices they connect, i.e., the edge $e = (i, j)$ is denoted as an unordered pair ij. Similarly, we index the elements of x in $x^T V_Q x$ by the corresponding pairs of vertices. Then,

$$x^T V_Q x = \sum_{i_1 j_1 \in E} \sum_{i_2 j_2 \in E} v_{i_1 j_1, i_2 j_2} x_{i_1 j_1} x_{i_2 j_2} = \sum_{i=1}^{v} w_i^{-1} \sum_{p \sim i} \sum_{q \sim i} x_{ip} x_{iq} = k \sum_{i=1}^{v} d_i^{-1} \sum_{p \sim i} \sum_{q \sim i} x_{ip} x_{iq}$$

for any $x \in \mathbb{R}^s$. Let us denote as $n(i)$ the set of all vertices adjacent to vertex i. Then,

$$\lambda_{\max}(V_Q) = \max_{\|x\|=1} x^T V_Q x = \max_{\|x\|=1} k \sum_{i=1}^{v} d_i^{-1} \Big(\sum_{j \in n(i)} x_{ij} \Big)^2.$$

The Cauchy-Schwarz inequality yields $(\sum_{j \in n(i)} x_{ij})^2 \leq d_i \sum_{j \in n(i)} x_{ij}^2$, and thus

$$\lambda_{\max}(V_Q) \leq k \max_{\|x\|=1} \sum_{i=1}^{v} \sum_{j \in n(i)} x_{ij}^2 = 2k \max_{\|x\|=1} \sum_{(i,j) \in E} x_{ij}^2 = 2k = \lambda^*.$$

It follows that $\lambda^* = \lambda_{\max}(V_Q)$.

Let $E = hh^T$ and $G = M^{-1}(w)$, and recall that $Qh = g/\sqrt{s}$, where $g_i = \pm d_i$. Then, the left-hand side of (4.3) is

$$\frac{1}{s} g^T \mathrm{diag}(\tilde{w}_1 w_1^{-2}, \ldots, \tilde{w}_v w_v^{-2}) g = \frac{1}{s} k^2 \sum_{i=1}^{v} \tilde{w}_i d_i^{-2} g_i^2 = \frac{k^2}{s} \sum_{i=1}^{v} \tilde{w}_i = \frac{k^2}{s} = 4s,$$

which is equal to the right-hand side. Thus, w is E-optimal. □

Similar to the A-optimal treatment proportions, the E-optimal proportions for bipartite systems depend only on the local properties of the graph – the numbers of the neighbors of the vertices. However, for general systems of pairwise comparisons, this need not be true. The global properties of the graphs are embedded in the assumption that the graph is bipartite. If this assumption does not hold, the simple dependence of the E-optimal designs only on the degrees of the vertices given by (4.4) need not hold, as demonstrated in the following example.

Example 4.14. Let $v = 4$ and consider estimating $\tau_1 - \tau_2$, $\tau_2 - \tau_3$, $\tau_3 - \tau_1$ and $\tau_1 - \tau_4$. The degrees of the vertices in the corresponding graph G are $d_1 = 3$, $d_2 = 2$, $d_3 = 2$ and $d_4 = 1$. The treatment proportions given by (4.4) are $w = (3/8, 1/4, 1/4, 1/8)^T$, and it can be calculated that $\lambda_{\max}(V_Q(w)) \approx 13.8297$. For $\tilde{w} = (0.38, 0.23, 0.23, 0.16)^T$, it can be calculated that $\lambda_{\max}(V_Q(\tilde{w})) \approx 13.0435 < \lambda_{\max}(V_Q(w))$. Thus, w is not E-optimal.

Remark 4.15. The corresponding bipartite graph G can be transformed so that it consists of sinks and sources only simply by appropriately multiplying the particular

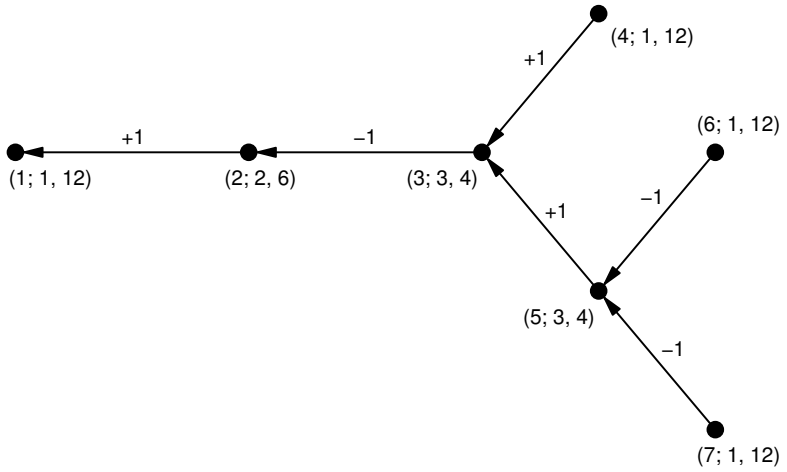


Figure 5: E-optimal design w^* for the system of contrasts $Q_1^T\tau$ from Example 4.1. The labels of the vertices are of the form $(i; d_i, \alpha_i)$, where i is the vertex index, d_i is the degree of vertex i, and $\alpha_i = 1/w_i^*$ is the vertex weight for vertex i. The values $+1$ or -1 on the edges represent the signs of the respective elements of the eigenvector h of $V_Q(w^*)$ corresponding to $\lambda_{\min}(V_Q(w^*))$.

pairwise comparisons by ± 1. However, if we do not wish to make such a transformation, a normalized eigenvector h corresponding to $\lambda_{\max}(V_Q(w^*))$ can still be obtained as follows. The elements of h belong to the set $\{-1/\sqrt{s}, 1/\sqrt{s}\}$, and for any vertex i the signs of all elements of h corresponding to the edges directed from i are the same, and they are opposite to the signs corresponding to the edges directed to i. Such an eigenvector h can be constructed by arbitrarily choosing one value $h_k \in \{-1/\sqrt{s}, 1/\sqrt{s}\}$ and then iteratively obtaining the signs of the incident edges. The choices $h_k = -1/\sqrt{s}$, $k \in K \subseteq \{1, \ldots, s\}$, can be thought of as reversing the directions of edges $k \in K$ in order to convert all vertices to sinks or sources; the choices $h_k = 1/\sqrt{s}$ mean maintaining the original directions of the corresponding edges.

Example 4.16 (Example 4.1 cont.)**.** The E-optimal design w^* for the system of contrasts $Q_1^T\tau$ from Example 4.1 is given in Figure 5. The total degree of G_1 is $\sum_j d_j = 12$, and thus the particular design values are $w_1^* = w_4^* = w_6^* = w_7^* = 1/12$, $w_2^* = 1/6$ and $w_3^* = w_5^* = 1/4$. The eigenvector h of $V_Q(w^*)$ corresponding to $\lambda_{\max}(V_Q(w^*)) = 2\sum_j d_j = 24$ is $h = (1, -1, 1, 1, -1, -1)^T/\sqrt{6}$, which is also represented in Figure 5 by values $+1$ or -1 on the edges. Notice that by reversing the directions of all edges for which $h_k = -1/\sqrt{6}$, all vertices of G_1 become sinks or sources.

The E-optimality criterion did not receive much attention in the literature on optimal designs, as noted in Section 3.3.3. Here, we provide a statistical interpretation

of E-optimality for the systems of treatment contrasts considered in Theorem 4.13, similar to that provided in Section 3.3.3.

Again, this can be achieved by employing the relationship between E-optimality and c-optimality (Lemma 1.14). For simplicity, let us first consider a graph G, in which each vertex is a sink or a source.

Proposition 4.17. *Let $Q^T\tau$ be a bipartite system of pairwise comparisons and let G be the corresponding graph. If each vertex of G is either a sink or a source, then the E-optimal treatment design given by (4.4) is also c-optimal, where $c = Q1_s$.*

Proof. In the proof of Theorem 4.13, we have shown that w^* satisfies (4.3) with $E = hh^T$, where $h = 1_s/\sqrt{s}$. Then, from Lemma 1.14 it follows that w^* is Qh-optimal, which is equivalent to the $Q1_s$-optimality of w^*. □

One consequence of the optimality of the considered designs for $Q1_s$ is that they minimize the variance of the least squares estimator of $\sum_i q_i^T\hat{\tau}$, or equivalently of $\sum_i q_i^T\hat{\tau}/s$. Thus, these E-optimal designs minimize the variance of the estimator of the average treatment contrast, which may be of interest in some settings, see Majumdar [1986] and Giovagnoli and Verdinelli [1988].

Furthermore, the $Q1_s$-optimality implies that the E-optimal designs minimize

$$1_s^T Q^T M^{-1}(w) Q 1_s = 1_s^T V_Q(w) 1_s \propto \sum_{i=1}^{s} \mathrm{Var}(q_i^T\hat{\tau}) + \sum_{i=1}^{s}\sum_{j\neq i} \mathrm{Cov}(q_i^T\hat{\tau}, q_j^T\hat{\tau}), \tag{4.6}$$

i.e., the sum of all variances and covariances for the contrasts of interest. Note that the covariances between $q_i^T\hat{\tau}$ and $q_j^T\hat{\tau}$ are nonnegative because of the nonnegativity of $V_Q(w)$, which was shown in (4.5). Thus, minimizing their sum is equivalent to minimizing the sum of the absolute values of these covariances.

In general, if not every vertex of G is a sink or a source, the eigenvector h is a vector with elements $\pm 1/\sqrt{s}$, with the signs representing a reorientation of the graph so that all vertices become sinks or sources, as noted in Remark 4.15. Then, analogously to (4.6), it can be shown that the obtained E-optimal designs minimize the sum of the *absolute values* of the variances and covariances for the contrasts of interest:

Theorem 4.18. *Let w be an E-optimal design for a bipartite system of pairwise comparisons $Q^T\tau$ given by Theorem 4.13. Then, w minimizes*

$$\sum_{i=1}^{s} \mathrm{Var}(q_i^T\hat{\tau}) + \sum_{i=1}^{s}\sum_{j\neq i} |\mathrm{Cov}(q_i^T\hat{\tau}, q_j^T\hat{\tau})|. \tag{4.7}$$

Theorem 4.18 provides a statistical interpretation of the E-optimal designs for the studied systems of interest similar to that of A-optimality: an A-optimal design for $Q^T\tau$ minimizes the sum of the variances of $q_i^T\hat{\tau}$. The E-optimal designs given by Theorem

4.13 seek to achieve a compromise between minimizing the variances of the estimators of the contrasts of interest and maintaining low correlation between these estimators.

The graph interpretation of E-optimality can be easily obtained through the quadratic form $x^T\mathcal{L}_w x$. For the unweighted graphs, the quadratic form given by the Laplacian $x^T L x = \sum_{(i,j)\in E}(x_i - x_j)^2$ measures the energy of a physical system represented by the graph, as noted by Mohar [1991]. For the vertex-weighted graphs, we have $\lambda_{\max}(\mathcal{L}_w) = \max_{\|x\|=1} x^T\mathcal{L}_w x = \max_{\|x\|=1}\sum_{(i,j)\in E}(w_i^{-1/2}x_i - w_j^{-1/2}x_j)^2$, which yields the following proposition.

Proposition 4.19. *Let $Q^T\tau$ be a system of pairwise comparisons, let $w > 0$ be a feasible design and let G be the corresponding vertex-weighted graph. Then,*

$$\Psi_{-\infty}(w) = \max_{\|y\|_w=1}\sum_{(i,j)\in E}(y_i - y_j)^2, \tag{4.8}$$

where $\|y\|_w := (\sum_i w_i y_i^2)^{1/2}$.

The expression $\sum_{(i,j)\in E}(y_i - y_j)^2$ measures the total variability (potential, energy) of a function y on vertices of G. Therefore, for each treatment design w, (4.8) represents the maximum total variability over all y, which are normalized in the weighted norm $\|y\|_w$. Then, the design under which this maximum total variability is the lowest is E-optimal, resulting in a weighted graph with the smallest worst variability as measured by (4.8). This seems to be a graph analogue to the interpretation of E-optimality for $Q^T\tau$, which states that E-optimal design minimizes the maximum possible variance (which depends on the design) of $(x_1 q_1^T + \ldots + x_s q_s^T)\hat{\tau}$ over all $\|x\| = 1$.

4.4.5 Symmetric systems of contrasts

If a system of contrasts is 'symmetric', intuitively, the uniform design $\bar{w} = 1_v/v$ should be optimal with respect to a wide range of optimality criteria. Using the graph representation, we may obtain such symmetric systems of contrasts.

In Section 4.4.1, we noted that if π is an automorphism, the matrix $\mathcal{L}_{P_\pi w}$ is orthogonally similar to $\mathcal{L}_w$. Therefore, Theorem 4.3 yields that $\Phi(P_\pi w) = \Phi(w)$ for any orthogonally invariant Φ. This observation is a useful tool in proving the optimality of $\bar{w}$ for a 'symmetric' system of contrasts.

Lemma 4.20. *Let w be a feasible design for estimating a system of pairwise comparisons $Q^T\tau$ and let π be an automorphism of the corresponding graph G. Then,*

$$\mathcal{L}_{P_\pi w} = P_\pi \mathcal{L}_w P_\pi^T. \tag{4.9}$$

and $\Phi(P_\pi w) = \Phi(w)$ for any orthogonally invariant information function Φ.

The following theorem shows that if a cyclic permutation is an automorphism of G, the uniform treatment design is Φ-optimal for $Q^T\tau$ for any orthogonally invariant information function. Recall that a permutation is cyclic if it consists of only one cycle.

Theorem 4.21. *Let $Q^T\tau$ be a system of pairwise comparisons $\tau_i - \tau_j$ and let G be the corresponding graph. Suppose that there exists a cyclic permutation π, which is an automorphism of G. Then, $\bar{w} = 1_v/v$ is Φ-optimal for estimating $Q^T\tau$ with respect to any orthogonally invariant information function Φ. Moreover, if Φ has a strictly concave version, then $\bar{w}$ is the unique Φ-optimal treatment design for $Q^T\tau$.*

Proof. Let $i \in 1, \dots, v-1$, and let $w > 0$ be a feasible treatment design. Note that if π is an automorphism, $\mathcal{L}_{P_\pi^i w} = P_\pi^i \mathcal{L}_w (P_\pi^i)^T$, and thus $\Phi(P_\pi^i w) = \Phi(w)$ holds for any $i \in \mathbb{N}$.

Since π is a cyclic permutation, the uniform treatment design can be expressed as $\bar{w} = \frac{1}{v}\sum_{i=0}^{v-1} P_\pi^i w$. Hence,

$$\Phi\big(\bar{w}\big) = \Phi\left(\frac{1}{v}\sum_{i=0}^{v-1} P_\pi^i w\right) \geq \frac{1}{v}\sum_{i=0}^{v-1}\Phi(P_\pi^i w) = \frac{1}{v}\sum_{i=0}^{v-1}\Phi(w) = \Phi(w),$$

where the inequality follows from the concavity of Φ. Thus, $\bar{w}$ is Φ-optimal. If Φ has a strictly concave version, say Φ_s, the inequality becomes strict for Φ_s. □

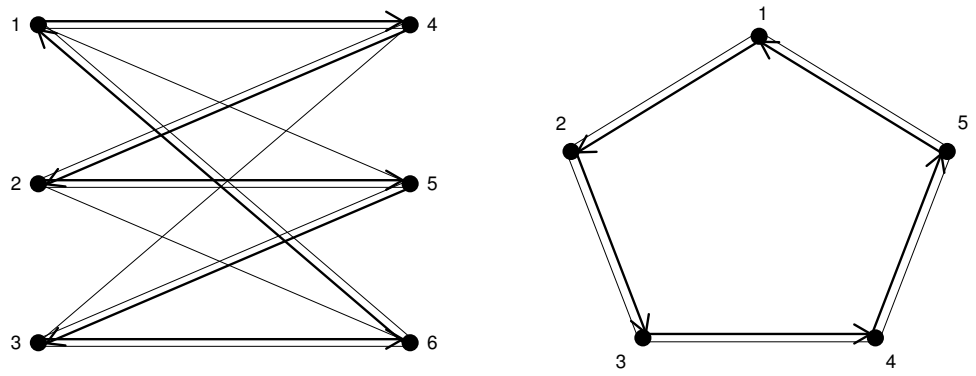
Note that for (4.9) to hold, it is sufficient for π to satisfy $P_\pi A P_\pi^T = A$, or equivalently, $P_\pi R R^T P_\pi^T = RR^T$. That is, it is not necessary to preserve the orientation of the edges. Furthermore, note that the proof of Theorem 4.21 does not employ the condition that $Q^T\tau$ is a system of pairwise comparisons. Hence, we may formulate a theorem for general systems of contrasts. We say that a system of contrasts $Q^T\tau$ (not necessarily a system of pairwise comparisons) is *cyclic* if there exists a cyclic permutation π satisfying

$$P_\pi Q Q^T P_\pi^T = QQ^T. \tag{4.10}$$

Theorem 4.22. *Let $Q^T\tau$ be a cyclic system of contrasts. Then, $\bar{w} = 1_v/v$ is Φ-optimal for estimating $Q^T\tau$ with respect to any orthogonally invariant information function Φ. Moreover, if Φ has a strictly concave version, then $\bar{w}$ is the unique Φ-optimal treatment design for $Q^T\tau$.*

Proof. The proof is analogous to the proof of Theorem 4.21; the matrix $\mathcal{L}_w$ is not considered to be the Laplacian matrix of the corresponding graph but formally, $\mathcal{L}_w := M^{-1/2}(w) Q Q^T M^{-1/2}(w)$ for any coefficient matrix Q and any $w > 0$. Similarly as in Theorem 4.3, the value $\Phi(w)$ is determined by the eigenvalues of $\mathcal{L}_w$. □

We note that Theorem 2.13, which states that if QQ^T is completely symmetric, then $\bar{w}$ is Φ-optimal for any orthogonally invariant criterion, is a corollary of Theorem 4.22.



(a) Comparing 3 treatments with 3 controls

(b) Cyclic comparisons

Figure 6: Graph representations of systems of contrasts; the directions of the edges (thin lines) are suppressed as they do not affect the results. The cyclic automorphisms are represented by bold arrows.

If QQ^T is completely symmetric, i.e., if $QQ^T = aI_v + bJ_v$ for some a, b, then (4.10) is satisfied for any permutation matrix and thus for any cyclic permutation. However, the completely symmetric systems do not cover all systems satisfying Theorem 4.22, as demonstrated in the following examples.

Example 4.23. Let $Q^T\tau$ be a system of contrasts for comparing two sets of treatments of equal size (say g), i.e., $\tau_j - \tau_i$, $i = 1, \ldots, g$, $j = g+1, \ldots, 2g$. This is a special case of *comparing $v - g$ treatments with g controls*, where $v - g = g$; see Section 2.2.4.

Then, $Q = (-I_g \otimes 1_g, 1_g \otimes I_g)^T$ and

$$QQ^T = \begin{bmatrix} gI_g & -J_g \\ -J_g & gI_g \end{bmatrix}.$$

Let π be the cyclic permutation $\pi : \pi(i) = g+i$ for $i \leq g$ and $\pi(i) = i-g+1$ for $i > g$, i.e., π can be represented by the cycle $(1, g+1, 2, g+2, 3, g+3, \ldots, g, 2g)$ (see Figure 6a, the arrows connect i to $\pi(i)$, $i = 1, \ldots, v$). It can be verified that $P_\pi QQ^T P_\pi^T = QQ^T$, and therefore the uniform design $\bar{w}$ is Φ-optimal for estimating $Q^T\tau$ with respect to any orthogonally invariant criterion.

Example 4.24. Let $Q^T\tau$ be a system of contrasts $\tau_2 - \tau_1$, $\tau_3 - \tau_2, \ldots, \tau_v - \tau_{v-1}$, $\tau_1 - \tau_v$. Clearly, $\pi = (1, 2, 3, \ldots, v)$, i.e., $\pi(i) = i+1$ for $i < v$ and $\pi(v) = 1$, is an automorphism of the corresponding graph (see Figure 6b). Thus, $\bar{w}$ is Φ-optimal for estimating $Q^T\tau$ with respect to any orthogonally invariant criterion.

If π is an arbitrary permutation consisting of cycles $c_1, \ldots, c_K$ that satisfies (4.10), the proof of Theorem 4.22 can be replicated, resulting in analogous results for each cycle c_i.

Theorem 4.25. *Let $Q^T\tau$ be a system of contrasts and let π be a permutation satisfying (4.10) that consists of cycles $c_1, \ldots, c_K$, i.e., $\pi = c_1 \ldots c_K$. Let Φ be an orthogonally invariant information function. Then, there exists a Φ-optimal design w that satisfies $w(i) = w(j)$ for all $i, j \in c_k$ for all $k \in \{1, \ldots, K\}$. Moreover, if Φ has a strictly concave version, then any Φ-optimal design must be uniform on each of the cycles c_k.*

Theorem 4.25 can be used to simplify the search for optimal treatment proportions. For example, let Φ be orthogonally invariant and consider comparing a set of test treatments with a set of controls. Then, to find a Φ-optimal treatment design w, it is sufficient to consider only two treatment proportions: one for the test treatments and one for the controls, as shown in Example 4.26. For the Kiefer's Φ_p-optimality criteria, these optimal treatment proportions are given in Theorem 2.15.

Example 4.26. Let $Q^T\tau$ be a system of contrasts for comparing $v - g$ treatments with g controls, $g < v/2$, i.e., $\tau_j - \tau_i$, $i = 1, \ldots, g$, $j = g + 1, \ldots, v$. Then, $Q = (-I_g \otimes 1_{v-g}, 1_g \otimes I_{v-g})^T$ and

$$QQ^T = \begin{bmatrix} (v-g)I_g & -J_{g\times(v-g)} \\ -J_{(v-g)\times g} & gI_{v-g} \end{bmatrix}.$$

Let $\pi = (1, \ldots, g)(g + 1, \ldots, v)$, i.e., $\pi(i) = i + 1$ for $i \neq g, i \neq v$; $\pi(g) = 1$ and $\pi(v) = g + 1$. Then, π satisfies (4.10). It follows that for any orthogonally invariant criterion Φ there exists a Φ-optimal design satisfying $w(i) = \gamma$ for all $i \in \{1, \ldots, g\}$ and $w(j) = (1 - g\gamma)/(v - g)$ for all $j \in \{g + 1, \ldots, v\}$, for some $\gamma \in (0, 1/g)$; cf. Theorem 2.15 for the Φ_p-criteria.

4.5 Discussion

The graph representation provided in this chapter establishes a clear approach to the study of optimal designs for estimating sets of pairwise treatment comparisons, based on the proposed straightforward visualization of the systems of interest. This representation provides a valuable insight into the studied problem, it is useful in proving the optimality statements, and it allows for a clear formulation of the obtained results. In particular, the results on E-optimality are strongly tied to the graph representation (see Theorem 4.13). The obtained simple formulas for A- and E-optimality in Section 4.4 allow one to use the graph approach to calculate the optimal designs 'by hand', just by counting the degrees of the vertices. This may be beneficial for the practitioners as the optimality results are comprehensible and easily applicable, without the need for

complex calculations. The applicability of the obtained results is demonstrated on a particular system of interest in Examples 4.1, 4.8, 4.11 and 4.16.

The E-optimality is less popular for systems of contrasts because it is often considered to be lacking a natural statistical meaning (see Majumdar and Notz [1983]). However, we showed that the provided E-optimal designs minimize (4.7), the sum of variances and of the absolute values of covariances for the comparisons of interest. This provides a meaningful interpretation of E-optimality for the studied systems of contrasts, similar to that of the widely used criterion of A-optimality.

We also studied the D-optimality criterion and obtained optimality results for general systems of contrasts without the need for graph representation. However, it has been argued that D-optimality may not be very suitable for estimating particular systems of interest; see Majumdar and Notz [1983], Hedayat et al. [1988], compare also with the weighted criteria (e.g., Lemma 6 by Stallings and Morgan [2015]). This follows from the fact that D-optimality usually does not distinguish between various systems of contrasts, resulting in uniform optimal proportions (see Theorem 4.6). Then, the D-optimal design for $Q^T\tau$ generally does not provide more information about the contrasts of interest $q_i^T\tau$ than about any other contrasts.

Motivated by the graph symmetry, we obtained a class of 'symmetric' systems of contrasts for which the uniform design is optimal with respect to any orthogonally invariant criterion. Although ultimately, the graph terminology is not necessary to formulate the optimality of the uniform designs (see Theorem 4.22), the graph representation can be used to find symmetry in the systems of pairwise comparisons more easily. The graph automorphisms, which represent the sought treatment permutations, can be obtained manually by examining the corresponding graph. Also, especially in the case of more complex systems of pairwise comparisons, one can employ the graph algorithms that provide all automorphisms of the graph (see, e.g., Darga et al. [2008]) to obtain all permutations that allow for a simplification of the optimal design problem using Theorem 4.25.

Employing the graph representation, we obtained a class of systems of pairwise comparisons, for which the E-optimal treatment proportions are given by the simple formula (4.4). It may be interesting to determine, whether this formula can be generalized to a wider range of systems of pairwise comparisons, or whether some other formula can be obtained for different systems of treatment contrasts. Another possibility for extending the provided results lies in studying other optimality criteria through the graph representation. In particular, the Kiefer's Φ_p criteria for p other than 0, -1, and $-\infty$ can be considered. We obtained that for $p = -\infty$, the optimal proportions depend directly on the degrees of the vertices (for a class of systems of interest); for $p = -1$, they depend on the square roots of the vertices; and for $p = 0$, the optimality of the uniform designs can be formulated as dependence on the 0th power of the vertex

degrees. Perhaps for other p, dependences on other powers of the degrees of the vertices can be proved.

5 Weighted optimality

5.1 Introduction

In this chapter, we study the weighted optimality criteria and their relationship with the criteria of optimality for $K^T\beta$. This chapter is based on a submitted paper Rosa [2016], currently available on arXiv. Note that in the current chapter, we do not focus on approximate designs; rather, we consider primarily the exact designs in order to make our results comparable to those by Stallings and Morgan [2015].

In a given experiment, certain parameters or functions of parameters may be of greater interest than others. In the optimal design theory, such objectives can be expressed using the weighted optimality criteria, developed by Morgan and Wang [2010] and Morgan and Wang [2011]. These authors consider models with treatment effects in which various weights are placed on the particular treatments. The so-called weighted information matrix, which captures the differential interest in the various treatments, is constructed using a diagonal matrix W of the treatment weights.

The most general definition of weighted optimality so far is provided by Stallings and Morgan [2015]. Their approach allows one to construct weighted optimality criteria that correspond to an experimental interest in a set of estimable functions $q_1^T\tau, \ldots, q_s^T\tau$ of the model parameters τ for (almost) arbitrary vectors q_i. We denote such an experimental objective as $Q^T\tau$, where Q is a coefficient matrix. Stallings and Morgan [2015] cover such objectives by allowing the matrix W of the weights of the model parameters to be any, generally non-diagonal, positive definite matrix. Then, the interest in $Q^T\tau$ can be expressed by constructing the corresponding weight matrix W_Q and then considering the weighted optimality with respect to W_Q.

Although the weighted optimality provides a meaningful approach to expressing the interest in a given set of linear estimable functions, such an objective can also be expressed directly, using the 'standard' theory of optimality for estimating the functions of interest $Q^T\tau$.

The optimality for a system of interest is a well-developed theory with a wide range of known results (e.g., see Pukelsheim [1993]). We mentioned some of these results in the previous chapters. Many of these results deal generally with the optimality criteria – notably the equivalence theorems (see Section 1.3); and many more provide optimal designs for particular models, systems of estimable functions and optimality criteria (e.g., see Chapters 2-4). There is also a multitude of results on algorithms for constructing optimal (or efficient) designs for a system of functions of interest: e.g., multiplicative algorithms (Yu [2010]), semidefinite programming (Harman and Sagnol [2015]), second-order cone programming (Sagnol [2011]) and mixed integer second-order cone programming (Sagnol and Harman [2015]).

The key statistical justification of weighted optimality criteria is the relationship

between the positive eigenvalues of the weighted information matrix and the weighted variances of estimable functions, see Lemma 1 by Stallings and Morgan [2015]. Therefore, in this chapter, we consider only eigenvalue-based optimality criteria, which are criteria that depend only on the positive eigenvalues of the information matrix.

Morgan and Wang [2010] and Stallings and Morgan [2015] observed that the optimality criteria for estimating the entire vector of parameters of interest τ are insufficient for expressing the interest in a set of estimable functions, or for assigning various weights to the particular parameters. This motivated the development of the weighted optimality criteria. However, while the new theory of weighted optimality is being developed, there already exists a similar well-developed theory of optimality for a system of interest. Therefore, a natural question arises – do these two approaches actually differ? The main result of this chapter lies in providing an answer to this question: in Section 5.2, we prove that these two approaches are equivalent; although, they have different statistical interpretations and reasoning.

In particular, we first show the equivalence of the standard optimality for $Q^T\tau$ and the corresponding weighted optimality; i.e., we show that they result in the same ordering of the competing designs. Conversely, we provide a formula for a system of interest $Q^T\tau$, so that optimality for $Q^T\tau$ is equivalent to the weighted optimality for an arbitrary weight matrix W.

The proved equivalence suggests that before extending or applying weighted optimality, it may be useful to examine whether some known results or algorithms from the rich theory of the standard optimality can be applied instead. Especially, it may often be unnecessary to use the weighted optimality corresponding to a (weighted) system $Q^T\tau$ to compute the optimal designs, when equivalently, the standard optimality for $Q^T\tau$ can be applied directly. We demonstrate this by showing that a recently proved result on weighted optimality (Theorem 2 by Stallings and Morgan [2015]) is, in fact, a special case of Corollary 8.8 by Pukelsheim [1993] in light of the provided equivalence.

We also provide a method based on the obtained equivalence that allows one to use the algorithms for optimality for a system of interest (e.g., the ones mentioned earlier) to compute optimal designs for weighted optimality criteria. Note that recently, Morgan and Stallings [2017] provided a method for constructing weighted-optimal designs that employs algorithms for calculating optimal designs for the vector τ by suitable transformations of the studied model.

Even if one computes the optimal designs for a system of interest by using the standard optimality, the approach of the weighted optimality can still be used as an analytical tool by calculating the weights of the estimable functions implied by the system of interest. Therefore, it seems important to be able to construct a weight matrix for any system of interest, as well as to interpret the implied weights for any

system of interest and any optimality criterion, neither of which was the case until now. In Section 5.3, we seek to remedy this.

Like Stallings and Morgan [2015] extended the theory of Morgan and Wang [2010] to capture more experimental situations, we extend this theory further. We propose a weight matrix W_Q of a simple form, which is able to capture *any* estimable system of interest $Q^T\tau$. Until now, the corresponding weighted optimality was provided only under certain conditions on Q.

Stallings and Morgan [2015] provide general interpretation of the weights implied by systems of interest of full rank. However, for rank-deficient systems, they observe some undesired behavior of weighted optimality and they interpret only the weighted A-optimality (see Lemma 7 and Lemma 8 therein). We provide a general interpretation of weighted optimality (even for rank-deficient systems) by distinguishing between two systems of weights of estimable functions: the primary weights, which are selected by the experimenter, and the secondary weights, which are implied by the weight matrix.

5.1.1 The model

Consider the vector formulation of model (1.1) with effects of interest $\tau \in \mathbb{R}^v$ and nuisance effects $\theta \in \mathbb{R}^m$:

$$Y = X(\xi)\tau + L(\xi)\theta + \varepsilon, \tag{5.1}$$

where $Y = (Y_1, \ldots, Y_n)^T$ is the vector of responses, and ε is the $n \times 1$ vector of uncorrelated errors with $E(\varepsilon) = 0_n$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$. Since the results in this chapter do not depend on the variance of the errors, for simplicity we assume $\sigma^2 = 1$. The (exact) design ξ determines the matrix $X(\xi)$ that relates τ to Y and the matrix $L(\xi)$ that relates θ to Y. In this chapter, by a *design* we mean an *exact design*.

The information matrix of ξ for τ is $C(\xi) = X^T(\xi)(I - P_L)X(\xi)$, where $L := L(\xi)$ and $P_L = L(L^TL)^-L^T$ is the matrix of the orthogonal projection on $\mathcal{C}(L(\xi))$. Given an optimality criterion Φ, a design is Φ-optimal for the vector τ if it maximizes $\Phi(C(\xi))$. Note that $C(\xi)$ is in fact the Schur complement $M_\tau(\xi)$ of the moment matrix. Let $X := X(\xi)$. Then, the moment matrix is

$$M(\xi) = \begin{bmatrix} X^T \\ L^T \end{bmatrix} \begin{bmatrix} X & L \end{bmatrix} = \begin{bmatrix} X^TX & X^TL \\ L^TX & L^TL \end{bmatrix}$$

and $M_\tau(\xi) = X^TX - X^TL(L^TL)^-L^TX = X^T(I - P_L)X$.

The approximate design ξ_a specifies the proportions of all trials that are performed under the particular experimental conditions. Analogously, the information matrix $C(\xi_a)$ of ξ_a for τ is the Schur complement $M_\tau(\xi)$, cf. Chapter 1. In the provided theoretical results, we consider exact designs, but these results extend naturally to the approximate ones because we study the optimality criteria in general, rather than particular models and designs.

A function $q^T\tau$, where $q \in \mathbb{R}^v$, is estimable under a design ξ if and only if $q \in \mathcal{C}(C(\xi))$. As in Stallings and Morgan [2015], we require for all competing designs to be able to estimate the same set of linear functions $q^T\tau$ of the parameters τ. We define $\mathcal{E}$ as the estimation space, which is the set of all vectors q corresponding to these estimable functions $q^T\tau$. That is, all competing designs satisfy $\mathcal{C}(C(\xi)) = \mathcal{E}$. We denote the $v \times v$ matrix of the orthogonal projection on $\mathcal{E}$ by P_τ.

Consider a system of s estimable functions of interest $q_1^T\tau, \ldots, q_s^T\tau$, together denoted as $Q^T\tau$, where $Q = (q_1, \ldots, q_s)$. We denote the rank of Q as r. The system $Q^T\tau$ is said to be normalized if $\|q_i\| = 1$ for all i. Recall that if Q has full column rank (i.e., if $s = r$), then we say that the system $Q^T\tau$ is of full rank; otherwise, $Q^T\tau$ is rank deficient.

Suppose that the interests in the functions $q_1^T\tau, \ldots, q_s^T\tau$ are represented by positive weights $b_1, \ldots, b_s$. This is easily incorporated into either the weighted or the standard optimality theory by rescaling the functions to $\tilde{q}_i^T\tau = \sqrt{b_i}q_i^T\tau$ and then working with the rescaled system of interest $\tilde{Q}^T\tau$. Equivalently, $\tilde{Q} := QB^{1/2}$, where $B := \text{diag}(b_1, \ldots, b_s)$ is the matrix of weights of the functions of interest.

Examples of model (5.1) include the model with only the treatment effects (1.6), and models with treatment and nuisance effects (1.2). In such cases, where τ are treatment effects, the estimation space $\mathcal{E}$ is the set of all vectors orthogonal to 1_v, representing the set of all treatment contrasts.

5.1.2 Weighted optimality

In this section, we briefly describe the weighted optimality criteria; for more details, see Stallings and Morgan [2015]. The weights of the parameters of interest are expressed using the weight matrix W, which is any $v \times v$ positive definite matrix. For a given W and any estimable function $q^T\tau$, the weight assigned to $q^T\tau$ is $(q^TW^{-1}q)^{-1}$ and the corresponding weighted variance is $\text{Var}_W(\widehat{q^T\tau}) = (q^TW^{-1}q)^{-1}\text{Var}(\widehat{q^T\tau})$. The higher the weight of an estimable function, the more interested we are in it. Notice the difference between the weights b_i of the *functions of interest* (which can easily be expressed by rescaling the functions of interest $q_i^T\tau$ to $\sqrt{b_i}q_i^T\tau$) and the matrix W of the weights of the model *parameters*, which implies weights for all estimable functions.

To capture the weights of the parameters of interest in evaluating the experimental designs, for a feasible design ξ, the information matrix $C(\xi)$ is transformed to the weighted information matrix

$$C_W(\xi) = W^{-1/2}C(\xi)W^{-1/2}.$$

Given an optimality criterion Φ, a design is Φ-optimal with respect to W (or in short, Φ_W-optimal) if it maximizes $\Phi_W(C(\xi)) := \Phi(C_W(\xi))$. Note that Stallings and Morgan consider optimality criteria that are to be minimized.

When the interest lies in a system $Q^T\tau$ that spans the entire $\mathcal{E}$ (i.e., if $r = \dim(\mathcal{E})$), the corresponding weight matrix is

$$W_Q = I - P_\tau + QQ^T,$$

where P_τ is the orthogonal projector on $\mathcal{E}$. For $r < \dim(\mathcal{E})$, the corresponding weight matrix is not provided.

If $Q^T\tau$ is normalized and of full rank, then the weight matrix $W_{\tilde{Q}}$ corresponding to the scaled system $\tilde{Q}^T\tau = B^{1/2}Q^T\tau$ places weight b_i on $q_i^T\tau$ $(1 \leq i \leq s)$, as desired. If the weights b_i are not specified, they are assumed to be 1 and indeed, the matrix W_Q places weight 1 on each of the functions of interest $q_i^T\tau$. However, in the rank-deficient case (i.e., if the system of interest $Q^T\tau$ has 'too many' functions of interest), the corresponding weight matrix W_Q does not generally place equal weights on each of the functions $q_i^T\tau$. That is not a desirable property; nevertheless, Stallings and Morgan [2015] show that the desired weights can be captured by the weighted A-optimality criterion even in such a case.

5.1.3 Standard optimality for a system of interest

As observed by Stallings and Morgan [2015], the standard optimality criteria defined on the information matrices $C(\xi)$ place equal emphasis on all normalized estimable functions, which is an evident limitation of such an approach. Consequently, the weighted optimality criteria present a way of eliminating these limitations because they are able to express various interest in the particular estimable functions.

However, this limitation of the standard optimality criteria arises primarily not because of the criteria considered, but rather it is due to the narrow definition of the information matrix $C(\xi)$, which captures the information about exactly the vector τ. When the interest lies in a set of s estimable functions $Q^T\tau$, this limitation can be completely circumvented by considering the well-established information matrix of a feasible design ξ for $Q^T\tau$:

$$N_Q(\xi) = (Q^T C^-(\xi) Q)^+. \tag{5.2}$$

If $Q^T\tau$ is of full rank, (5.2) reduces to $N_Q(\xi) = (Q^T C^-(\xi) Q)^{-1}$. As in the previous chapters, a design is said to be Φ-optimal for the system of interest $Q^T\tau$ if it maximizes $\Phi(N_Q(\xi))$. Note that for ease of notation, in this chapter, we denote the information matrices for both the full-rank and for the rank-deficient system by the same expression N_Q.

The information matrix $N_Q(\xi)$, unlike the weighted information matrix $C_{W_Q}(\xi)$, is well defined even for systems of interest that do not span the estimation space (i.e., for $r < \dim(\mathcal{E})$).

5.2 Equivalence of weighted and standard optimality

5.2.1 Equivalence

The following theorem shows that the standard optimality for $Q^T\tau$ and the corresponding weighted optimality are equivalent for any eigenvalue-based criterion (i.e., any criterion that depends only on the positive eigenvalues of the information matrix).

Theorem 5.1. *Let $Q^T\tau$ be a system of estimable functions that satisfies $r = \dim(\mathcal{E})$, and let W_Q be the corresponding weight matrix. Then, for any eigenvalue-based optimality criterion Φ, the Φ-optimality for $Q^T\tau$ and the Φ-optimality with respect to W_Q are equivalent: they result in the same ordering of the competing designs.*

Proof. Let us denote $C := C(\xi)$. In this proof, we employ the facts that the non-zero eigenvalues of AB are identical to the non-zero eigenvalues of BA, and that the non-zero eigenvalues of AA^T are identical to the non-zero eigenvalues of A^TA; see 6.54 of Seber [2008].

The eigenvalues of $C_{W_Q}(\xi)$ are identical to the eigenvalues of $W_Q^{-1}C$, which can be expressed as $(I - P_\tau + (QQ^T)^+)C = (QQ^T)^+C$ because $\mathcal{C}(C) \subseteq \mathcal{E}$. Therefore, the non-zero eigenvalues of $C_{W_Q}(\xi)$ are inverse to the non-zero eigenvalues of $((QQ^T)^+C)^+$.

Theorem 2 of Greville [1966] implies that for two symmetric matrices A and B of the same size, the formula $(AB)^+ = B^+A^+$ holds if and only if A^+ABB^T and A^TABB^+ are symmetric. Because $r = \dim(\mathcal{E})$ and $\mathcal{C}(Q) \subseteq \mathcal{C}(C) = \mathcal{E}$, it follows that $\mathcal{C}(Q) = \mathcal{C}(C) = \mathcal{E}$. Moreover, $\mathcal{C}((QQ^T)^+) = \mathcal{C}(QQ^T) = \mathcal{C}(Q) = \mathcal{C}(C) = \mathcal{C}(C^+)$, which yields that there exist matrices Z and V, such that $C = QQ^TZ$ and $(QQ^T)^+ = C^+V$. Since both $(QQ^T)^+$ and C^+ are symmetric, it follows that $(QQ^T)^+ = V^TC^+$. By setting $A = (QQ^T)^+$ and $B = C$ in Theorem 2 of Greville [1966], we obtain

$$A^+ABB^T = QQ^T(QQ^T)^+CC = QQ^T(QQ^T)^+QQ^TZC = C^2$$

and

$$A^TABB^+ = (QQ^T)^+(QQ^T)^+CC^+ = (QQ^T)^+V^TC^+CC^+ = (QQ^T)^+(QQ^T)^+,$$

which are both evidently symmetric. Therefore, the non-zero eigenvalues of $C_{W_Q}(\xi)$ are inverse to the eigenvalues of C^+QQ^T, which has the same non-zero eigenvalues as Q^TC^+Q. Thus, the matrix $C_{W_Q}(\xi)$ has the same non-zero eigenvalues as $(Q^TC^+Q)^+ = N_Q(\xi)$. □

Theorem 5.1 (and similarly the other theorems in this chapter establishing equivalence of the weighted and the standard optimality criteria) is proved by showing that for any competing design ξ, the information matrix $N_Q(\xi)$ for $Q^T\tau$ and the weighted information matrix $C_{W_Q}(\xi)$ have the same non-zero eigenvalues, including multiplicities. The condition $r = \dim(\mathcal{E})$ in this theorem is necessary for the weight matrix W_Q

to exist. As before, the weights b_i placed on the functions $q_i^T\tau$ can be expressed by appropriately rescaling the system $Q^T\tau$ to $\tilde{Q}^T\tau$.

Conversely, one may specify the weight matrix (as in Morgan and Wang [2010]), instead of considering a set of weighted estimable functions. This brings up an inverse problem: can a system of estimable functions that corresponds to the pre-specified weight matrix be constructed? We provide a positive answer to this question. The system $R^T\tau = R\tau$, where $R := (P_\tau W^{-1} P_\tau)^{+1/2}$, is the sought system of estimable functions. Note that any system $R_2^T\tau$ that satisfies $R_2 R_2^T = RR^T$ also corresponds to W.

Theorem 5.2. *Let W be a weight matrix and let $R = (P_\tau W^{-1} P_\tau)^{+1/2}$. Then, for any eigenvalue-based optimality criterion Φ, the Φ-optimality for $R\tau$ and the Φ-optimality with respect to W are equivalent: they result in the same ordering of the competing designs.*

Proof. In this proof, we employ the same techniques as in the proof of Theorem 5.1. Let us denote $C := C(\xi)$ and $W_2 := P_\tau W^{-1} P_\tau$. The non-zero eigenvalues of $C_W(\xi)$ are identical to those of $W^{-1}C = W^{-1}P_\tau C P_\tau$, which are in turn identical to the eigenvalues of $P_\tau W^{-1} P_\tau C = W_2 C$. The non-zero eigenvalues of $W_2 C$ are inverse to the non-zero eigenvalues of $(W_2 C)^+$.

Note that $\mathcal{C}(W_2^+) = \mathcal{C}(W_2) = \mathcal{E} = \mathcal{C}(C) = \mathcal{C}(C^+)$. Therefore, there exist matrices Z and H, such that $W_2 = C^+ Z$ and $C = W_2^+ H$. Thus, $W_2 = Z^T C^+$. Let us employ Theorem 2 of Greville [1966] for $A = W_2$ and $B = C$. Then, it can be shown that $A^T ABB^+ = W_2^2$ and $A^+ ABB^T = C^2$, which are both symmetric. Thus, $(W_2 C)^+ = C^+ W_2^+ = C^+ RR$. The non-zero eigenvalues of $C^+ RR$ are identical to those of $RC^+ R$, which are in turn inverse to the non-zero eigenvalues of $(RC^+ R)^+ = N_R(\xi)$. This shows that the non-zero eigenvalues of $C_W(\xi)$ and $N_R(\xi)$ are identical. □

5.2.2 Applications

Based on the equivalence provided in Theorem 5.1, it may often be unnecessary to consider the weighted optimality corresponding to a system of interest $Q^T\tau$, as the same results can be obtained directly by employing the optimality for $Q^T\tau$. For example Stallings and Morgan [2015] study the A_W-optimal designs in the one-way analysis of variance model (1.7). The value $\xi(i)$ $(1 \le i \le v)$ of an approximate design ξ determines the proportion of all trials performed with treatment i. Consider a system of interest $Q^T\tau$, let $W := W_{\tilde{Q}}$ be the weight matrix that corresponds to the system $Q^T\tau$ with weights B, and denote the elements of Q as q_{ik}. Then, Stallings and Morgan [2015] show that the approximate design ξ given by

$$\xi(i) = \frac{\sqrt{u_i}}{\sum_j \sqrt{u_j}}, \quad i = 1, \ldots, v, \tag{5.3}$$

where $u_i = \sum_k b_k q_{ik}^2$ $(1 \leq i \leq v)$, is A_W-optimal.

However, in light of Theorem 5.1, this formula is a special case of a more general result by Pukelsheim [1993]. Corollary 8.8 therein provides a formula for A-optimal designs for $Q^T\tau$ in a model with linearly independent regressors. It can be seen that a design ξ is Φ-optimal for $Q^T\tau$ in (1.7) if it is Φ-optimal in a simpler model with dropped constant term (1.6) (see Lemma 1.18), resulting in linearly independent regressors. The weights B can be represented by considering the scaled system of contrasts $\tilde{Q}^T\tau$, where $\tilde{Q} = QB^{1/2}$. By considering the system $\tilde{Q}^T\tau$ in model (1.6) in Corollary 8.8 by Pukelsheim [1993], one obtains (5.3).

Theorem 5.2 opens up the possibility to employ the theory of optimality for $Q^T\tau$ to construct designs that are optimal with respect to the weighted criteria. For example, if one seeks to obtain a Φ_W-optimal design for a given weight matrix W, this can be achieved by employing the already developed algorithms for optimality for $Q^T\tau$ by the following method:

1. Calculate $R = (P_\tau W^{-1} P_\tau)^{+1/2}$.

2. Calculate the Φ-optimal design ξ^* for $R\tau$ using the standard algorithms for optimal designs for a system of interest. Then, ξ^* is Φ_W-optimal.

Let us demonstrate the proposed method by computing optimal approximate designs for a dose-escalation study.

Example 5.3. Consider the standard designs for the dose-escalation study (3.1) (see Chapter 3), where $\tau = (\tau_0, \ldots, \tau_{v-1})^T$. Bailey [2009a] assumed that in a dose-escalation study all treatment comparisons $\tau_i - \tau_j$ $(0 \leq i < j \leq v-1)$ are of equal interest. However, Senn [2009] argued that the comparisons of the doses with the placebo $\tau_i - \tau_0$ $(i > 0)$ are the primary objective of such a study. In Chapter 3, we studied optimal designs for comparing the doses with the placebo, which corresponds to a limiting case of the weighted criteria, see Morgan and Wang [2010]. The resulting designs placed rather high proportions of the subjects on placebo.

Therefore, one may wish to place higher emphasis on the comparisons with placebo $\tau_i - \tau_0$ $(i > 0)$, but still maintain some level of interest in the between-dose comparisons $\tau_i - \tau_j$ $(i, j > 0)$. This can be achieved by assigning weight $w > 1$ to the placebo, and weight 1 to each of the doses. That is, $W = \mathrm{diag}(w, 1, \ldots, 1)$.

The value of an approximate design $\xi(i,k)$ represents the proportion of all individuals that are treated in cohort k with treatment i. The dose-escalation constraints yield $\sum_i \xi(i,k) = 1/(v-1)$, and $\xi(i,k) = 0$ for $i > k+1$. The semidefinite programming method of Harman and Sagnol [2015] allows for such design constraints. Thus, through the transformation given in Theorem 5.2 (in fact, we employ a full-rank system $R_2^T\tau$

that satisfies $R_2 R_2^T = RR^T$), we calculate the A_W- and E_W-optimal approximate dose-escalation designs using the algorithm of Harman and Sagnol [2015]. For three doses and for the weights of the placebo $w = 3$ and $w = 5$, the obtained designs are given in Tables 11 and 12. Naturally, the total number of proportions of the placebo $\sum_k \xi(0, k)$ increases as the weight w of the placebo increases, as can be observed in the last columns of these tables.

$i\backslash k$	1	2	3	$\sum$
0	0.167	0.110	0.082	0.359
1	0.167	0.081	0.042	0.290
2	0	0.143	0.064	0.207
3	0	0	0.144	0.144

$i\backslash k$	1	2	3	$\sum$
0	0.167	0.115	0.091	0.373
1	0.167	0.077	0.037	0.281
2	0	0.142	0.061	0.203
3	0	0	0.144	0.144

Table 11: The A_W-optimal approximate dose-escalation designs for 3 doses and for weight of the placebo $w = 3$ (left) and $w = 5$ (right). The value on position (i, k) represents $\xi(i, k)$ and the last column captures the row sums $\sum_k \xi(i, k)$. i - treatment number, k - cohort number

$i\backslash k$	1	2	3	$\sum$
0	0.167	0.109	0.100	0.377
1	0.166	0.081	0.033	0.280
2	0	0.143	0.033	0.177
3	0	0	0.167	0.167

$i\backslash k$	1	2	3	$\sum$
0	0.167	0.118	0.119	0.404
1	0.166	0.069	0.024	0.259
2	0	0.146	0.024	0.170
3	0	0	0.167	0.167

Table 12: The E_W-optimal approximate dose-escalation designs for 3 doses and for weight of the placebo $w = 3$ (left) and $w = 5$ (right). The value on position (i, k) represents $\xi(i, k)$ and the last column captures the row sums $\sum_k \xi(i, k)$. i - treatment number, k - cohort number

5.3 Extending the theory of weighted optimality

5.3.1 General definition of the weight matrix

Although Stallings and Morgan [2015] provide a fairly general definition of weighted optimality, they still capture only systems of interest of sufficient rank $r = \dim(\mathcal{E})$. This, for instance, does not cover the case of c-optimality; i.e., when only one function of interest $c^T\tau$ is selected. To capture such situations, we first slightly relax the definition of the weight matrix.

When the estimation space is not the entire $\mathbb{R}^v$, it is not necessary to weight all the functions with coefficients in $\mathbb{R}^v$. Moreover, the experimenters may wish to assign non-zero weights only to some subset of the estimable functions; we propose to represent

this set of functions by the column space of the weight matrix W. Therefore, W need not be non-singular. For possibly singular W, a natural relaxation of the weight of an estimable function is $(q^T W^- q)^{-1}$ for $q \in \mathcal{C}(W)$. Because $q \in \mathcal{C}(W)$, this weight does not depend on the choice of W^-. If $q \notin \mathcal{C}(W)$, the weight of $q^T\tau$ is zero.

Definition 5.4. *(i) The matrix W is a weight matrix if $W \in \mathfrak{S}_+^v$ and $\mathcal{C}(W) \subseteq \mathcal{E}$.*
(ii) For any $q \in \mathcal{C}(W)$, the weighted variance for $q^T\tau$ with respect to a weight matrix W is $\mathrm{Var}_W(\widehat{q^T\tau}) = (q^T W^- q)^{-1}\mathrm{Var}(\widehat{q^T\tau})$.

For any competing design, we require only $\mathcal{C}(W) \subseteq \mathcal{C}(C(\xi))$. This condition guarantees that all the functions $q^T\tau$ with non-zero weights are estimable under ξ. Following Definition 5.4, we must appropriately define the weighted information matrix for possibly singular W. We construct the weighted information matrix using a full-rank matrix K_W that satisfies $K_W K_W^T = W$ as follows. Given a weight matrix W, we denote its rank by d and its spectral decomposition by $W = S\Lambda S^T$, where $\Lambda = \mathrm{diag}(\lambda_1(W), \ldots, \lambda_v(W))$ and S is a $v \times v$ orthogonal matrix. We may also write $W = FDF^T$, where $D = \mathrm{diag}(\lambda_1(W), \ldots, \lambda_d(W))$ is a $d \times d$ matrix of the positive eigenvalues of W, and the $v \times d$ matrix F contains the first d columns of S. Then, $K_W = FD^{1/2}$.

Definition 5.5. *The weighted information matrix of a design ξ with respect to a weight matrix W is*

$$C_W(\xi) = (K_W^T C^-(\xi) K_W)^{-1},$$

where K_W is defined in the previous paragraph.

Both the matrices K_W and $C(\xi)$ are of rank d, and $\mathcal{C}(K_W) = \mathcal{C}(W) \subseteq \mathcal{E} = \mathcal{C}(C(\xi))$; thus, the weighted information matrix $C_W(\xi)$ is well defined, and is of full rank. A design is Φ-optimal with respect to W, or Φ_W-optimal in short, if it maximizes $\Phi(C_W(\xi))$.

Similarly to Stallings and Morgan [2015], we show the key property of the weight matrix W, which justifies its definition: we prove that W is relevant with respect to the weighted variances.

Proposition 5.6. *Let W be a weight matrix of rank d, let $q \in \mathcal{C}(W)$ and let ξ be a design that is feasible for $q^T\tau$. Then, the weighted variance of $\widehat{q^T\tau}$ under ξ is a convex combination of $\lambda_1^{-1}(C_W(\xi)), \ldots, \lambda_d^{-1}(C_W(\xi))$.*

Proof. Since $W = K_W K_W^T$, we have $\mathcal{C}(W) = \mathcal{C}(K_W)$, and thus $q \in \mathcal{C}(W) = \mathcal{C}(K_W)$ yields $q = K_W h$ for some $h \in \mathbb{R}^d$. Therefore,

$$\mathrm{Var}(\widehat{q^T\tau}) = q^T C^-(\xi) q = h^T C_W^{-1}(\xi) h = h^T U \Lambda^{-1} U^T h = \sum_{i=1}^{d} g_i^2 \lambda_i^{-1}(C_W(\xi)),$$

where $U\Lambda U^T$ is the spectral decomposition of $C_W(\xi)$, and $g := U^T h$. Then, $\sum_i g_i^2 = h^T U U^T h = h^T h$. The weight of $q^T\tau$ is the reciprocal of

$$q^T W^- q = h^T K_W^T (K_W K_W^T)^- K_W h = h^T h$$

because $K_W^T (K_W K_W^T)^- K_W$ is a symmetric idempotent matrix of full rank, i.e., I_d. It follows that

$$\mathrm{Var}_W(\widehat{q^T\tau}) = (h^T h)^{-1} \sum_{i=1}^{r} g_i^2 \lambda_i^{-1}(C_W(\xi)) = \sum_{i=1}^{r} \frac{g_i^2}{g^T g} \lambda_i^{-1}(C_W(\xi)).$$

□

5.3.2 Weighted optimality for any set of estimable functions

The relaxed definition of the weight matrix allows us to provide a corresponding weight matrix for *any* system of interest $Q^T\tau$. Moreover, this weight matrix is of a very simple form, with no need for the 'regularizing' term $I - P_\tau$.

We propose to assign weights only to estimable functions that can be constructed as linear combinations of $q_1, \ldots, q_s$. Formally, one weights only $q^T\tau$ with $q \in \mathcal{C}(Q)$. These are the functions that are guaranteed to be estimable under any design that is feasible for $Q^T\tau$. The corresponding weight matrix for any system of interest $Q^T\tau$ can then be constructed as follows.

Definition 5.7. *Let $Q^T\tau$ be a system of estimable functions with weights $b_1, \ldots, b_s$. Then, the corresponding unscaled weight matrix is $W_Q = QQ^T$ and the scaled weight matrix is $W_{\tilde{Q}} = QBQ^T$.*

As usual, the scaled weight matrix is obtained by considering the scaled system of functions of interest $\tilde{Q}^T\tau = B^{1/2}Q^T\tau$. Note that the set of all competing designs for W_Q is exactly the set of all designs that are feasible for $Q^T\tau$, which demonstrates the consistency of the respective definitions.

The following theorem shows that the weight matrix proposed here is indeed a generalization of that proposed by Stallings and Morgan [2015]. The former is defined for a wider class of systems $Q^T\tau$ (in fact, for any system of estimable functions) and when the latter exists (i.e., when $r = \dim(\mathcal{E})$), we show that they are equivalent with respect to the implied weights. This is called the estimation equivalence.

Theorem 5.8. *Let $Q^T\tau$ be a system of functions of interest, such that $r = \dim(\mathcal{E})$. Then, the weight matrices $W_1 = QQ^T$ and $W_2 = (I - P_\tau) + QQ^T$ are equivalent with respect to the implied weights of the estimable functions. That is, $q^T W_1^- q = q^T W_2^{-1} q$ for any $q \in \mathcal{E}$.*

Proof. Let $q \in \mathcal{E}$. Because $r = \dim(\mathcal{E})$, we have $\mathcal{C}(Q) = \mathcal{E}$, and thus $q^T W_1^- q$ does not depend on the choice of W_1^-. Then, $q^T W_2^{-1} q = q^T(I - P_\tau + (QQ^T)^+)q = q^T(QQ^T)^+ q = q^T W_1^- q$ since $P_\tau q = q$. □

As in Section 5.2, the equivalence of the standard criteria and the weighted criteria proposed here holds. We formalize this equivalence in the following two theorems.

Theorem 5.9. *Let $Q^T\tau$ be a system of estimable functions and let W_Q be the corresponding weight matrix. Then, for any eigenvalue-based optimality criterion Φ, the Φ-optimality for $Q^T\tau$ and the Φ-optimality with respect to $W_Q = QQ^T$ are equivalent: they result in the same ordering of designs feasible for $Q^T\tau$.*

Proof. The non-zero eigenvalues of $N_Q(\xi) = (Q^T C^- Q)^+$, where $C := C(\xi)$, are reciprocals of the non-zero eigenvalues of the matrix $Q^T C^- Q$. This matrix can be expressed as $Q^T C^- Q = (C^{+1/2}Q)^T(C^{+1/2}Q)$, which shows that it has the same non-zero eigenvalues as the matrix $C^{+1/2}QQ^TC^{+1/2}$. Because $QQ^T = W_Q = K_{W_Q}K_{W_Q}^T$, we have $C^{+1/2}QQ^TC^{+1/2} = C^{+1/2}K_{W_Q}K_{W_Q}^TC^{+1/2}$; and $C^{+1/2}K_{W_Q}K_{W_Q}^TC^{+1/2}$ has the same non-zero eigenvalues as $K_{W_Q}^TC^+K_{W_Q}$. Observing that the non-zero eigenvalues of the matrix $K_{W_Q}^TC^+K_{W_Q}$ are inverse to the non-zero eigenvalues of $C_{W_Q}(\xi)$ completes the proof. □

The system of estimable functions that corresponds to an arbitrary weight matrix W can be constructed by employing the matrix $W^{1/2}$.

Theorem 5.10. *Let W be a weight matrix. Then, for any eigenvalue-based optimality criterion Φ, the Φ-optimality for $W^{1/2}\tau$ and the Φ-optimality with respect to W are equivalent: they result in the same ordering of the competing designs.*

Proof. The system $W^{1/2}\tau$ is estimable because $\mathcal{C}(W^{1/2}) = \mathcal{C}(W) \subseteq \mathcal{E}$. Since $W^{1/2}$ satisfies $W^{1/2}(W^{1/2})^T = W$, the weight matrix W corresponds to the system $W^{1/2}\tau$. Theorem 5.9 then yields the identity of the non-zero eigenvalues. □

A proper weight matrix W_Q should place equal weight on each of the normalized functions $q_i^T\tau$. In the following proposition, we show that this is satisfied as long as $Q^T\tau$ is of full rank. We deal with the rank-deficient case in Section 5.3.3.

Proposition 5.11. *Let $Q^T\tau$ be a full-rank system of normalized estimable functions. Then, $W_Q = QQ^T$ places weight 1 on each of the functions $q_1^T\tau, \ldots, q_s^T\tau$. If $b_1, \ldots, b_s$ are the pre-specified weights of these functions, then $W_{\tilde{Q}}$ places weights $b_1, \ldots, b_s$ on $q_1^T\tau, \ldots, q_s^T\tau$.*

Proof. We have $Q^T W_Q^- Q = Q^T(QQ^T)^- Q = I_s$ because $Q^T(QQ^T)^- Q$ is a symmetric idempotent matrix of full rank, i.e., I_s. The second statement follows from the analogous observation that $\tilde{Q}^T(\tilde{Q}\tilde{Q}^T)^-\tilde{Q} = I_s$, which yields $B^{1/2}Q^T W_{\tilde{Q}}^- Q B^{1/2} = I_s$. □

Among the most common eigenvalue-based criteria, the D-optimality does not seem to be generally appropriate for weighted criteria, as shown in Lemma 6 of Stallings and Morgan [2015]. The interpretation of the weighted E-optimality is similar to that of Stallings and Morgan [2015], see Lemma 4 therein.

The value of the weighted A-optimality criterion, or A_W-optimality criterion in short, is $\Phi_{AW}(\xi) = d(\mathrm{tr}(C_W^+(\xi)))^{-1}$. Using the relationship between the weighted optimality and the optimality for systems of interest, the following proposition shows that A_W-optimality seeks to minimize the average variance for any system of estimable functions corresponding to W.

Proposition 5.12. *Let W be a weight matrix of rank d. Then, the inverse of the value of A_W-optimality for a design ξ is equal to the average variance for any system of d estimable functions $Q^T\tau$ that satisfy $W_Q = W$.*

Proof. We have $(\Phi_{AW}(\xi))^{-1} = \mathrm{tr}(K_W^T C^- K_W)/d$, where $C := C(\xi)$. Let $Q^T\tau$ be a system of d estimable functions, such that $QQ^T = W$. Then, the average weighted variance for $q_1^T\tau, \ldots, q_d^T\tau$ is

$$\frac{1}{d}\sum_{i=1}^{d}(q_i^T W^- q_i)^{-1}\mathrm{Var}(\widehat{q_i^T\tau}) = \mathrm{tr}(DQ^T C^- Q)/d,$$

where $D = \mathrm{diag}((q_1^T W^- q_1)^{-1}, \ldots, (q_d^T W^- q_d)^{-1})$. Moreover, $Q^T W^- Q = Q^T(QQ^T)^- Q = I_d$, as in the proof of Proposition 5.6. Thus, $D = I_d$ and the average weighted variance is equal to $\mathrm{tr}(Q^T C^- Q)/d$. From $QQ^T = K_W K_W^T$, it follows that $\mathrm{tr}(Q^T C^- Q)/d = \mathrm{tr}(K_W^T C^- K_W)/d$. □

5.3.3 Primary and secondary weights

If we build the weight matrix W_Q corresponding to a normalized full-rank system $Q^T\tau$ with weights B, the weights assigned to the functions $q_i^T\tau$ are b_i, as desired. However, as noted before, this is not true for rank-deficient systems. To explain this seeming deficiency of weighted optimality, we propose to distinguish between two systems of weights for a system of interest $Q^T\tau$. The *primary weights* $b_1, \ldots, b_s$ are pre-specified by the experimenters for a finite number of functions of interest $Q^T\tau$, and they represent the primary interest assigned to these functions. The secondary weights represent the amount of interest in any function $q \in \mathcal{C}(W)$ implied by the functions of interest and their primary weights: for any $q \in \mathcal{C}(W)$, the *secondary* (or *implied*) *weight* of $q^T\tau$ is $w(q) := (q^T W_{\tilde{Q}}^- q)^{-1}$.

Although in the rank-deficient case, the primary and the secondary weights generally do not coincide, this is not a deficiency of the theory of weighted optimality. In fact, such a phenomenon is natural and reasonable. From theoretical point of view, because of the confounding in the rank-deficient systems, the functions of interest provide each

other some of their weights. Therefore, the secondary weights of the functions of interest $q_i^T\tau$ should be greater than their primary weights. This is indeed the case, as shown in the following proposition.

Proposition 5.13. *Let $Q^T\tau$ be a system of estimable functions with weights $b_1, \ldots, b_s$, and let $W_{\tilde{Q}}$ be the corresponding weight matrix. Then, the secondary weight of any of the functions $q_i^T\tau$ with respect to $W_{\tilde{Q}}$ is at least b_i; i.e., $w(q_i) \geq b_i$. Moreover, if q_i can be expressed as a linear combination of the other columns of Q, then $w(q_i) > b_i$; otherwise $w(q_i) = b_i$.*

Proof. Let $P_Q := \tilde{Q}^T(\tilde{Q}\tilde{Q}^T)^-\tilde{Q}$. Observe that $w^{-1}(q_i) = q_i^T(\tilde{Q}\tilde{Q}^T)^-q_i = b_i^{-1}e_i^TP_Qe_i$, where e_i is the ith column of I_s. Because P_Q is idempotent, $\rho(P_Q) = 1$, where ρ is the spectral norm of a matrix. It follows that

$$e_i^TP_Qe_i = e_i^TP_QP_Qe_i = \|P_Qe_i\|_2^2 \leq \rho^2(P_Q)\|e_i\|_2^2 = 1.$$

Therefore, $w^{-1}(q_i) \leq b_i^{-1}$, which is equivalent to $w(q_i) \geq b_i$.

Equality is attained in $\|P_Qe_i\|_2^2 \leq \rho^2(P_Q)\|e_i\|_2^2$ if and only if

$$e_i \in \arg\max_x \frac{\|P_Qx\|_2}{\|x\|_2},$$

i.e., if and only if e_i is an eigenvector corresponding to $\lambda_1(P_Q) = 1$. The last condition is equivalent to $e_i \in \mathcal{C}(\tilde{Q}^T)$ because P_Q is the projector on $\mathcal{C}(\tilde{Q}^T)$. If q_i can be expressed as a linear combination of the other columns of Q, the same holds for $\tilde{q}_i$ and $\tilde{Q}$. Then, there exists $z \in \mathbb{R}^s$ such that $\tilde{Q}z = 0_v$ and $z_i = 1$. Therefore, $e_i^Tz = 1$, which means that $e_i \notin \mathcal{N}^\perp(\tilde{Q}) = \mathcal{C}(\tilde{Q}^T)$. That is, $w(q_i) > b_i$. Conversely, suppose that q_i cannot be expressed as a linear combination of the other columns of Q. Therefore, if $\tilde{Q}z = 0_v$ for some $z \in \mathbb{R}^s$, then $z_i = 0$, which can be rewritten as $e_i^Tz = 0$. It follows that $e_i \in \mathcal{N}^\perp(\tilde{Q}) = \mathcal{C}(\tilde{Q}^T)$; i.e., $w(q_i) = b_i$. □

From practical point of view, the optimal designs tend to place greater emphasis on the functions of interest with greater secondary weights even if all the functions of interest have the same primary weights. That is, the estimators of these functions tend to have smaller variances under optimal designs. Therefore, the *secondary* weights correspond to the actual amount of interest the system $Q^T\tau$ places on the particular functions of interest. We demonstrate this on Examples 5.15 and 5.16 in the next subsection. Hence, the discrepancy between the primary and the secondary weights is not a deficiency, but a useful feature of the weighted optimality theory. The primary weights are used for the construction of the weighted optimality, the secondary ones should be used for its interpretation.

5.3.4 Applications

The theory of weighted optimality can be used to analyze the amount of interest that a system $Q^T\tau$ implies for other estimable functions. Such an analysis was performed for a factorial experiment in Section 4.2 of Stallings and Morgan [2015].

The extended theory of weighted optimality proposed in this chapter allows one to perform a similar analysis even for systems of interest $Q^T\tau$ of rank $r < \dim(\mathcal{E})$. Let us demonstrate this on an example.

Example 5.14. Majumdar [1986], Giovagnoli and Verdinelli [1988] argue that in experiments with two groups of treatments, one may be interested in comparing the treatments in one group with the treatments in the other, or in comparing the average effects of the treatments in the groups. Consider an example similar to one given by Majumdar [1986]: suppose that a farm is using two 'standard' varieties of a plant, denoted 1 and 2, and is considering to start planting also one of three new varieties (3, 4 and 5). The yields Y of the varieties are given by model (1.6) (possibly with some nuisance effects), where τ_i is the effect of the ith variety ($i = 1, \ldots, 5$). The assumed objective of the experiment is to determine whether some or all of the new varieties have higher yields than the current yield (the average yield of the standard varieties). Formally, we wish to estimate $(\tau_i - (\tau_1 + \tau_2)/2) \cdot (\sqrt{2}/\sqrt{3})$, $i = 3, 4, 5$, which together form the system of interest $Q^T\tau$.

It follows that zero weight is assigned to the comparison of the standard varieties $(\tau_2 - \tau_1)/\sqrt{2}$ because this function cannot be constructed from $Q^T\tau$. On the other hand, it can be calculated that the weight of the comparison of the new varieties $(\tau_i - \tau_j)/\sqrt{2}$ $(i, j = 3, 4, 5)$ is $2/3$. The weight for the comparison of any new variety with any standard variety $q^T\tau = (\tau_i - \tau_j)/\sqrt{2}$ $(i = 3, 4, 5,\ j = 1, 2)$ is $4/3$, which is even higher. Therefore, the system $Q^T\tau$ seems to be a reasonable choice if the interest lies in comparing the new varieties with the standard ones (and possibly the new varieties with each other), but does not seem to be appropriate if significant interest actually lies in comparing the standard varieties with each other.

The following example demonstrates the difference between (and the usefulness of distinguishing between) the primary and the secondary weights.

Example 5.15. Consider the experimental settings as in Ver Hoef and Cressie [2001]. In each of n plots of land, the number of different plant species (species richness) is measured. But first, some of these plots are burned. The observed number of plant species in plot i is governed by equation (1.6), where treatment 1 is control (no burn), and treatments 2, 3, 4, 5 represent May, August, November, and February burn, respectively. The measurements then take place in the next summer. The contrasts of interest selected by Ver Hoef and Cressie [2001] are $q_1^T\tau = (\tau_2 + \tau_3)/2 - \tau_1$,

$q_2^T\tau = (\tau_4+\tau_5)/2-\tau_1$, $q_3^T\tau = (\tau_4+\tau_5)/2-(\tau_2+\tau_3)/2$, $q_4^T\tau = \tau_2-\tau_3$ and $q_5^T\tau = \tau_4-\tau_5$. To make the analysis clearer, let us consider the normalized versions of these contrasts, so that $\|q_i\| = 1$ for all i. Suppose that the primary weights are $b_1 = \ldots = b_5 = 1$.

This is clearly a rank-deficient system of interest. One can calculate the secondary weights: $w(q_1) = w(q_2) \approx 1.6$, $w(q_3) \approx 1.33$, $w(q_4) = w(q_5) \approx 1$. It follows that although the primary weights were uniform, the actual amount of interest this system implies for the first two contrasts is the greatest, the third contrast receives less weight, and the emphasis on the last two contrasts is the lowest.

Let us compare this result with the variances for the contrasts of interest implied by the A-optimal approximate design for $Q^T\tau$. Using formula (5.3), the values of the A-optimal design ξ^* are $\xi^*(1) \approx 0.23$, $\xi^*(2) = \ldots = \xi^*(5) \approx 0.19$. The variance for $q_i^T\tau$ is proportional to $v_i := q_i^T \mathrm{diag}(\xi^*(1)^{-1}, \ldots, \xi^*(5)^{-1}) q_i$. Therefore, the variances for $q_i^T\tau$ are proportional to $v_1 = v_2 \approx 4.61$, $v_3 = v_4 = v_5 \approx 5.21$. One can observe that the contrasts of interest with the greatest secondary weights have the smallest variances, and the contrasts with the smallest secondary weights have the largest variances under the A-optimal approximate design.

In the following example, we analyze the rank-deficient system of comparing two groups of treatments (see Theorem 2.15) by calculating the weights implied by this system.

Example 5.16. Suppose that in model (1.6), we wish to compare two groups of treatments, as in Theorem 2.15. For such a normalized system of interest $Q^T\tau$, where $Q = (I_g \otimes 1_{v-g}, -1_g \otimes I_{v-g})^T/\sqrt{2}$, it can be verified that

$$(QQ^T)^+ = 2\begin{bmatrix} \frac{1}{v-g} I_g - \frac{2v-g}{v^2(v-g)} J_{g\times g} & -\frac{1}{v^2} J_{g\times(v-g)} \\ -\frac{1}{v^2} J_{(v-g)\times g} & \frac{1}{g} I_{v-g} - \frac{v+g}{gv^2} J_{(v-g)\times(v-g)} \end{bmatrix}.$$

Consequently, one can calculate that the secondary weight of any of the between-group comparisons $(\tau_i - \tau_j)/\sqrt{2}$ $(i \le g,\ j > g)$ is $g(v-g)/(v-1) > 1$. Because of the confounding in the system $Q^T\tau$, the actual weight of each of the functions of interest is greater than 1. Meanwhile, the weight of any comparison of two treatments in the first group $(\tau_i - \tau_j)/\sqrt{2}$ $(i, j \le g)$ is $(v-g)/2$ and the weights of the comparisons in the second group $(\tau_i - \tau_j)/\sqrt{2}$ $(i, j > g)$ are $g/2$.

Therefore, the treatment comparisons in the group with *less treatments* attain *greater weights*. Moreover, usually the weights of the between-group comparisons are smaller than those of the in-group comparisons in one of the groups. This is true for the first group if $g > (v-1)/2$ and for the second group if $g < (v+1)/2$. This is in accordance with the behavior of the optimal designs. For example, consider the E-optimal design, which satisfies $\xi^*(j) = 1/(2g)$ for $j \le g$ and $\xi^*(j) = 1/[2(v-g)]$ for $j > g$ (see Theorem 2.15). Then, the variances for the between-group comparisons

Comparison	Size	Weight	Var
Between-group		7/4	9
Group 1 (in-group)	2	7/2	4
Group 2 (in-group)	7	1	14

Table 13: Comparing groups of sizes $g = 2$ and $v - g = 7$ in Example 5.16. The columns denote: the treatment comparisons; the sizes of the groups; the secondary weights of the comparisons; and the values proportional to the variances for these comparisons under ξ^*.

are proportional to $q_i^T \mathrm{diag}(\xi^*(1)^{-1}, \ldots, \xi^*(v)^{-1}) q_i = v$ and the variances for the comparisons in the first (second) group are proportional to $2g$ and $2(v - g)$, respectively. Clearly, the variances for the group with *less treatments* (i.e., with greater weights) are *smaller*. Moreover, if $v \neq 2g$, one of the in-group variances is smaller than the variance for the between-group comparisons. For $v = 9$ and $g = 2$, these observations are demonstrated in Table 13.

Conclusion

In this work, we provided analytical results on optimal approximate designs for estimating systems of interest in various models. We also compared the well-established theory of optimality for a system of interest (see Pukelsheim [1993]) with the recently developed weighted optimality criteria (see Stallings and Morgan [2015]).

In the first chapter, some elementary properties of experimental designs in the general linear model, and in models (1.2) and (1.6) are provided.

In Chapter 2, we investigated model (1.2) with treatment and nuisance effects. We formulated necessary conditions for optimality (Theorems 2.6, 2.12), sufficient conditions for optimality (Theorems 2.8, 2.12); and for the strictly concave optimality criteria, we provided conditions that are both necessary and sufficient (Theorem 2.10).

Through the aforementioned theorems, we showed that that the optimal nuisance-resistant designs may be obtained in two steps, thus simplifying the optimal design problem. In the first step, one calculates the optimal treatment proportions for the given system of contrasts and optimality criterion, which is a fairly simple convex optimization problem. The second step consists of arranging the given proportions in a particular manner (2.2), which eliminates the nuisance effects. The first step can often be solved analytically, which we demonstrated in Section 2.2.4 as well as in Chapter 4. We showed that the second step is in fact a linear programming problem, and thus it can be efficiently solved using the standard interior point or the simplex method.

One drawback of considering approximate designs is that it may be difficult to construct the exact designs from them (this is particularly true for product designs). However, the obtained class of optimal designs, which is usually richer than the class of optimal product designs, tends to contain designs with small supports. Hence, one can onstruct efficient exact designs from these designs more easily (see Example 2.21). Such designs can be obtained, e.g., by solving the second step using the simplex method (Proposition 2.20).

In Chapter 3, we examined optimal approximate designs in dose-escalation studies, modeled as blocking experiments with design constraints (see Bailey [2009a]). As suggested by Senn [2009], we considered the experimental objective of comparing doses with the placebo. We provided the classes of all E-optimal standard and extended designs (Theorems 3.2, 3.3) and we provided justifications for considering the E-optimality criterion for comparisons with the placebo (Section 3.3.3). We also provided LV- (Theorems 3.9, 3.10) and MV-optimal (Theorem 3.11) designs.

Among the obtained optimal approximate designs, the Senn designs stand out, due to their optimality to all of the investigated criteria as well as their simple and regular

structure. The regular structure allows for simple characterization of these designs for practical use, and allows one to construct optimal or efficient exact designs from such designs fairly easily.

In Chapter 4, we investigated model (1.6), in which the responses are affected only by treatment effects. Thus, in fact, we studied the optimal treatment proportions needed in the first step of obtaining optimal approximate designs for model (1.2). We showed that any system of pairwise comparisons can be represented by a graph and that the treatment designs can be represented by the (inverses of) the vertex weights in such graphs. We proved that for any eigenvalue-based criterion, the criterial value of the treatment design is then determined by the spectrum of the Laplacian of the corresponding vertex-weighted graph (Theorem 4.3).

To demonstrate the applicability of the obtained representation, we provided optimality results for A- (Proposition 4.10) and E-optimality (Theorem 4.13) using the graph representation. The obtained formulas are easy to calculate by hand, simply by counting the degrees of the vertices. We also formulated optimality results for D-optimality (Theorem 4.6). Using the graph theory, we provided a class of 'symmetric' systems of contrasts, for which the uniform treatment design is optimal with respect to any eigenvalue-based criterion; see Section 4.4.5. It seems that the graph representation allows for useful and rather 'elegant' (or easy to describe) optimality results.

In Chapter 5, we investigated a problem quite different to those in the previous chapters. In this chapter, we did not examine optimal approximate designs; rather, we studied the optimality criteria. In particular, we studied the criteria of weighted optimality, primarily following Stallings and Morgan [2015]. We showed the equivalence of the weighted optimality criteria and the criteria for estimating systems of interest, which we investigated in the previous chapters (Theorems 5.1 and 5.2). Therefore, one can use the known results on optimality for a system of interest (e.g., the ones mentioned in the previous chapters) instead of needing to formulate analogous theoretical results and algorithms for the weighted optimality.

Furthermore, we proposed a definition of weighted optimality that is slightly different to that of Stallings and Morgan [2015], see Definitions 5.4 and 5.5. This definition allows one to express a broader class of experimental objectives (see Definition 5.7), and is equivalent to the weighted optimality of Stallings and Morgan [2015] when the latter is defined (Proposition 5.8). Finally, we noted that it may be beneficial to differentiate between the primary weights, which are used to construct a weight matrix, and the secondary weights, which are implied by the weight matrix; see Section 5.3.3.

Although the individual chapters of this thesis deal with various (sometimes somewhat isolated) aspects of optimal designs for estimating systems of interest, by com-

bining the obtained results, some observations of a more general nature can be made. First, and least surprisingly, the various theorems for optimal approximate designs (e.g., Theorems 2.10, 3.3, 4.3) demonstrate that indeed the results on optimal approximate designs can often be of simple form despite capturing rather general settings.

Second, even though the E-optimality is usually not considered to possess a very natural statistical interpretation for estimating systems of interest (as mentioned earlier, cf. Majumdar and Notz [1983]), in many of the settings examined in this thesis, such an interpretation can be formulated. Most E-optimal designs for estimating a system of contrasts $q_1^T\tau, \ldots, q_s^T\tau$ provided in this work minimize the sum of variances and of the absolute values of covariances for these contrasts:

$$\sum_{i=1}^{s} \mathrm{Var}(\widehat{q_i^T\tau}) + \sum_{i=1}^{s}\sum_{j\neq i} |\mathrm{Cov}(\widehat{q_i^T\tau}, \widehat{q_j^T\tau})|. \tag{5.4}$$

We show that for the dose-escalation studies as well as in many settings in the model with only treatment effects, the E-optimal designs minimize (5.4); see Section 3.3.3 and Theorem 4.18. Because the optimal nuisance-resistant designs provided in Chapter 2 have the same information matrices as their treatment proportions (Theorem 2.8), the E-optimal nuisance-resistant designs for bipartite systems of contrasts given by Theorems 2.8 and 4.13 inherit the property of minimizing (5.4). This extends the statistical interpretation of E-optimal designs to the model with nuisance effects. The property of minimizing (5.4) suggests that the E-optimality criterion is often a reasonable and statistically justified choice when the experimental objective is to estimate a system of contrasts.

Third, the matrix QQ^T seems to play an important role (rather, a multitude of roles) when one wishes to estimate a system $Q^T\tau$. In Chapter 4, this matrix was shown to be the Laplacian of the graph that describes the system of interest, and it was crucial in characterizing the 'symmetric' systems of contrasts; see Theorem 4.22. Moreover, we have seen in Chapter 5 that the matrix QQ^T is also the weight matrix W_Q corresponding to any system of interest (see Definition 5.7), which describes how the system $Q^T\tau$ 'behaves': which weights it implies for the particular estimable functions. It is also not difficult to observe, using the techniques of Chapters 4 and 5, that for any eigenvalue-based criterion, its value depends on the system $Q^T\tau$ only through the matrix QQ^T. That is, for a system $Q_1^T\tau$, we have $\Phi(N_Q(\xi)) = \Phi(N_{Q_1}(\xi))$ for any Q that satisfies $QQ^T = Q_1Q_1^T$. Therefore, it seems that the matrix QQ^T (the Laplacian matrix, the weight matrix) is crucial in characterizing a system $Q^T\tau$ and in obtaining optimal designs for this system with respect to eigenvalue-based criteria.

Resumé

Úvod

Optimálne návrhy experimentov slúžia na zvolenie takých experimentálnych podmienok, ktoré umožnia získať z experimentu čo najviac relevantnej informácie. Od začiatku dvadsiateho storočia, keď táto problematika začala byť skúmaná s významným prispením od R. Fishera (viď Fisher [1935]), je dnes už navrhovanie experimentov rozvinutá disciplína, ktorej aspekty sú spracované v mnohých monografiách, napr. Pázman [1986], Pukelsheim [1993], Atkinson et al. [2007].

Okrem tzv. exaktných návrhov experimentov, ktoré presne určujú počty meraní, ktoré máme vykonať za jednotlivých experimentálnych podmienok, boli navrhnuté aj aproximatívne návrhy, ktoré predstavujú relaxáciu tých exaktných, viď Kiefer [1959]. Tieto návrhy špecifikujú iba pomery meraní pre jednotlivé experimentálne podmienky, čím sa úloha nájsť optimálny návrh zjednodušuje z numerického aj z analytického pohľadu.

Zvyčajne je množstvo získanej informácie merané presnosťou odhadov parametrov modelu. Možným experimentálnym záujmom však môže byť aj odhadovanie systému lineárnych funkcií parametrov modelu. Napríklad nemusíme pokladať všetky ošetrenia za rovnocenné, ale jedno z nich môže byť kontrolným - vtedy nás môžu zaujímať porovnania vplyvov ostatných ošetrení s kontrolným ošetrením. V takom prípade by sme mali zvoliť návrh experimentu, ktorý berie zvolený experimentálny cieľ do úvahy. V práci sa budeme venovať práve optimálnym návrhom pre odhadovnie systémov lineárnych funkcií parametrov (skrátene: systémov záujmu). Táto problematika je dôkladne teoreticky analyzovaná napr. v monografii Pukelsheim [1993], ale aj v množstve článkov. Za všetky spomeňme napríklad Majumdar and Notz [1983] a Hedayat et al. [1988], ktoré sa venujú práve porovnaniam s kontrolným ošetrením. Samozrejme, pre optimálne návrhy pre systémy záujmu boli vyvinuté viaceré algoritmy, napr. Yu [2010], Sagnol and Harman [2015].

Ciele

Cieľom tejto práce je analyticky skúmať optimálne návrhy pre systémy záujmu, s dôrazom na aproximatívne návrhy. Konkrétne, hlavnými cieľmi práce sú:

1. Skúmať optimálne aproximatívne návrhy pre systémy kontrastov ošetrení v modeloch s vplyvmi ošetrení aj s rušivými vplyvmi.

2. Získať optimálne aproximatívne návrhy pre porovnania s placebom v modeli z Bailey [2009a] pre štúdie so zvyšujúcou sa dávkou lieku.

3. Venovať sa problému hľadania optimálnych aproximatívnych návrhov v modeli s

iba vplyvmi ošetrení. Získať všeobecné výsledky o optimálnych návrhoch v tomto modeli, ako aj predpisy pre optimálne návrhy pre niektoré kritériá optimality.

4. Porovnať teóriu optimality pre systémy záujmu s váženou teóriou optimality navrhnutou v Morgan and Wang [2010]. Zároveň skúmať, prípadne rozšíriť váženú teóriu optimality.

Práca vychádza z publikovaných alebo zaslaných článkov Rosa and Harman [2016], Rosa and Harman [2017], Rosa [2017] a Rosa [2016]. Tieto štyri články sa v uvedenom poradí venujú štyrom vytýčeným cieľom práce.

Model

Vo všeobecnosti uvažujeme model

$$Y_i = f^T(x_i)\beta + \varepsilon_i, \quad i = 1, \ldots, N, \tag{R1}$$

kde Y_i je výsledok i-teho pokusu, ktorý je modelovaný pomocou neznámych parametrov β, regresných funkcií f a náhodných chýb ε, ktoré sú nekorelované a rovnako rozdelené s disperziou σ^2. Experimentálne podmienky zvolené pre i-ty pokus vyjadruje $x_i \in \mathfrak{X}$, kde $\mathfrak{X}$ je množina všetkých uvažovaných experimentálnych podmienok.

Aproximatívny návrh ξ je funkcia $\xi : \mathfrak{X} \to [0, 1]$ spĺňajúca $\sum_x \xi(x) = 1$. Pod návrhom experimentu budeme odteraz rozumieť aproximatívny návrh, ak nebude špecifikované inak. Momentová matica návrhu ξ je $M(\xi) = \sum_x \xi(x) f(x) f^T(x)$. Systém $K^T\beta$ pre $K \in \mathbb{R}^{v\times s}$ je odhadnuteľný pod ξ ak $\mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))$, kde $\mathcal{C}$ vyjadruje stĺpcový priestor. V takom prípade, ak K má plnú hodnosť, tak informačná matica návrhu ξ pre $K^T\beta$ je $N_K(\xi) = (K^T M^-(\xi) K)^{-1}$. Ak K nemá plnú hodnosť, tak v predpise informačnej matice vymeníme inverziu za Mooreovu-Penroseovu pseudoinverziu $(K^T M^-(\xi) K)^+$. Návrh ξ^*, pod ktorým je systém $K^T\beta$ odhadnuteľný, je Φ-optimálny pre $K^T\beta$ pre zvolené kritérium Φ, ak maximalizuje $\Phi(N_K(\xi))$.

Špeciálnym prípadom modelu (R1) je

$$Y_i = \tau_{u(i)} + h^T(t(i))\theta + \varepsilon_i, \quad i = 1, \ldots, N, \tag{R2}$$

kde $u(i) \in \{1, \ldots, v\}$ je ošetrenie zvolené pre i-ty pokus a $t(i) \in \mathfrak{T}$ sú ostatné (rušivé) experimentálne podmienky zvolené v tomto pokuse. Ďalej $\tau = (\tau_1, \ldots, \tau_v)^T$ sú vplyvy ošetrení, a vplyvy rušivých experimentálnych podmienok sú modelované ako $h^T(t(i))\theta$ pre zvolené regresné funkcie h. Potom $\beta = (\tau^T, \theta^T)^T$ a uvažujeme experimentálny cieľ odhadnúť systém kontrastov $Q^T\tau$ pre $Q \in \mathbb{R}^{v\times s}$, čo môžeme tiež vyjadriť ako $K^T\beta$, kde $K^T = (Q^T, 0)$. Informačná matica sa následne zjednoduší na $N_K(\xi) = (Q^T M_\tau^-(\xi) Q)^{-1}$, resp. $(Q^T M_\tau^-(\xi) Q)^+$, kde $M_\tau(\xi)$ je tzv. Schurov doplnok v matici $M(\xi)$, pozri Pukelsheim [1993]. Kontrasty sú lineárne funkcie, ktorých koeficienty sa nasčítajú na nulu, teda $Q^T 1_v = 0_s$.

Ešte jednoduchším modelom, ktorému sa budeme venovať, je model (R2) bez rušivých vplyvov, teda

$$Y_i = \tau_{u(i)} + \varepsilon_i, \quad i = 1, \ldots, N. \tag{R3}$$

Aproximatívne návrhy v tomto modeli značíme $w = (w(1), \ldots, w(v))^T$, ich momentové matice sú $M(w) = \mathrm{diag}(w)$. Tieto návrhy určujú aké relatívne množstvo pokusov vykonáme s jednotlivými ošetreniami, nazývame ich teda *návrhy proporcií (pomerov) ošetrení*. Ak predpokladáme, že odhadujeme $Q^T\tau$, tak informačná matica pre $Q^T\tau$ je $N_Q(w) = (Q^T M^{-1}(w) Q)^{-1}$.

Návrhy odolné voči rušivým vplyvom

Výsledky experimentu môžu byť určené vplyvmi, ktoré odhadujeme, ako aj ďalšími vplyvmi, o ktoré sa nezaujímame, ale nedokážeme ich z experimentu odstrániť. Tieto druhé vplyvy nazývame rušivé, môžu to byť napríklad blokové vyplvy, alebo rušivý časový trend. My skúmame model (R2), kde okrem rušivých vplyvov máme vplyvy ošetrení, pričom sa zaujímame o systém kontrastov ošetrení. Existuje veľké množstvo výsledkov o optimálnych návrhoch pre systém kontrastov ošetrení v prítomnosti rušivých vplyvov, napríklad o blokových návrhoch (napr. Majumdar and Notz [1983], Majumdar [1996], Jacroux [2002]), o návrhoch odolných voči vplyu trendu (Cox [1951], Cheng [1990], Atkinson and Donev [1996]), ako aj o optimálnych aproximatívnych návrhoch za rôznych podmienok (Pukelsheim [1983], Giovagnoli and Wynn [1985]).

Predpokladáme teda, že experimentálny cieľ je odhadovať $K^T\beta$, kde $K^T = (Q^T, 0)$. Najprv si všimnime, že každný aproximatívny návrh ξ v modeli (R2) má prislúchajúci návrh w v jednoduchšom modeli (R3). Prislúchajúci návrh w je určený celkovým (relatívnym) množstvom pokusov, ktoré ξ priraďuje jednotlivým ošetreniam. Môžeme sformulovať nasledujúce tvrdenie.

Tvrdenie 1. *Nech ξ je návrh v modeli* (R3) *a nech w je jeho prislúchajúci návrh pomerov ošetrení. Potom $N_K(\xi) \preceq N_Q(w)$.*

Teda v modeli s rušivými vplyvmi môžeme získať nanajvýš toľko informácie ako v modeli bez rušivých vplyvov, čo je intuitívne jasné. Následne môžeme sformulovať nutné podmienky optimality, ktoré hovoria, že optimálny návrh ξ nutne musí dosahovať optimálne pomery ošetrení.

Veta 1. *Nech $Q^T\tau$ je systém kontrastov a nech ξ^* je Φ-optimálny pre $Q^T\tau$ v modeli* (R2)*, kde Φ je informačná funkcia. Potom návrh proporcií w prislúchajúci ξ je Φ-optimálny v modeli* (R3).

Návrh ξ nazývame *odolný voči rušivým vplyvom*, ak spĺňa technickú podmienku

$$\left[\tfrac{1}{w_1} \sum_{t \in \mathfrak{T}} \xi(1,t) h(t), \quad \ldots, \quad \tfrac{1}{w_v} \sum_{t \in \mathfrak{T}} \xi(v,t) h(t) \right] Q = 0, \tag{R4}$$

kde w je príslušný návrh proporcií ošetrení. Dá sa ukázať, že každý návrh ξ odolný voči rušivým vplyvom spĺňa $N_K(\xi) = N_Q(w)$ pre prislúchajúci návrh w, teda ξ nestráca informáciu v dôsledku rušivých vplyvov. Získavame tak postačujúce podmienky optimality.

Veta 2. *Nech Φ je informačná funkcia, $Q^T\tau$ je systém kontrastov a nech w^* je Φ-optimálny pre $Q^T\tau$ v* (R3)*. Nech ξ^* je odolný voči rušivým vplyvom a nech mu prislúcha návrh w^*. Potom ξ^* je Φ-optimálny pre $Q^T\tau$ v* (R2)*.*

Navyše, ak kritérium Φ je ostro konkávne (teda ak existuje taká ekvivalentná forma funkcie Φ, ktorá je ostro konkávna) a Q má plnú hodnosť, tak získané podmienky sú nielen postačujúce, ale aj nutné.

Veta 3. *Nech Φ je informačná funkcia, pre ktorú existuje jej ostro konkávna verzia, nech $Q^T\tau$ je systém kontrastov s plnou hodnosťou a nech w^* je Φ-optimálny pre $Q^T\tau$ v* (R3)*. Potom ξ je Φ-optimálny pre $Q^T\tau$ v* (R2) *práve vtedy, keď ξ je odolný voči rušivým vplyvom a prislúcha mu w^*.*

Ukazuje sa, že optimálne návrhy v modeli (R2) môžeme získať v dvoch krokoch. V prvom kroku vypočítame optimálne proporcie ošetrení, čo je pomerne jednoduchý optimalizačný problém. V druhom kroku tieto proporcie rozložíme medzi rušivé podmienky tak, aby sa nestrácala informácia (teda aby spĺňali (R4)). Keďže podmienky (R4) sú lineárne, druhý krok, ktorý už je optimalizáciou vo veľkom počte rozmerov, môžeme riešiť pomocou lineárneho programovania.

Podobný prístup pozostávajúci z nájdenia optimálneho návrhu v jednoduchšom modeli a následnej konštrukcie návrhu, ktorý zachováva informáciu v zložitejšom modeli, bol použitý napríklad v Kunert [1983], Schwabe and Wierich [1995] a Schwabe [1996]. V Kunert [1983] boli skúmané exaktné návrhy pre univerzálnu optimalitu. Schwabe and Wierich [1995], Schwabe [1996] a ďalšie články študovali optimálne aproximatívne súčinové návrhy. Síce optimálne súčinové návrhy v modeli (R2) existujú vždy, ale vzhľadom na ich súčinovú formu je zvyčajne problematické získať z nich efektívne exaktné návrhy.

Nami poskytnutá trieda optimálnych návrhov zahŕňa optimálne súčinové návrhy, ale zvyčajne je omnoho širšia. To častokrát umožňuje vybrať taký optimálny aproximatívny návrh, ktorý je vhodný na konštrukciu efektívneho exaktného návrhu. Pre dané Φ a Q je množina všetkých Φ-optimálnych návrhov odolných voči rušivým vplyvom polyéder a návrhy vo vrcholoch tohto polyédra majú zvyčajne malý nosič, teda obsahujú veľa núl. Práve tieto návrhy sú obzvlášť vhodné na konštrukciu efektívnych exaktných návrhov a vieme ich získať simplexovou metódou. Príklad takto získaného návrhu je uvedený v tabuľke 1. Tento návrh sa zjavne dá pomerne jednoducho zaokrúhliť na exaktný návrh (napríklad zvolením v každom čase takého ošetrenia, ktoré nadobúda najvyššiu hodnotu návrhu v danom čase).

$u\backslash t$	1	2	3	4	5	6	7	8
1	0	0.1250	0.0560	0	0	0	0	0.0437
2	0	0	0.0690	0	0	0.1250	0.0059	0.0249
3	0.0245	0	0	0.1250	0	0	0	0.0340
4	0.0154	0	0	0	0.1250	0	0.0207	0.0224
5	0.0851	0	0	0	0	0	0.0984	0

Tabuľka 1: A-optimálny návrh odolný voči rušivým vplyvom za prítomnosti časového trendu. u - číslo ošetrenia, t - čas

Štúdie so zvyšujúcou sa dávkou lieku

Predpokladajme, že máme vopred zvolených $v-1$ dávok lieku a skúmame ich bezpečnosť a tolerovateľnosť tak, že ich podávame zdravým dobrovoľníkom. Okrem týchto dávok môžeme dať dobrovoľníkom aj placebo. Dobrovoľníkov (subjekty) rozdelíme do $v-1$ kohort, ktorým budeme ošetrenia dávať postupne: najprv subjektom v prvej kohorte môžeme dávať iba placebo, alebo najnižšiu dávku, potom subjektom v druhej kohorte dávame placebo a prvé dve dávky atď., až nakoniec subjektom v poslednej kohorte môžeme dať placebo alebo ľubovoľnú dávku. Vzhľadom na postupné zvyšovanie dávok lieku sa tieto štúdie nazývajú štúdie so zvyšujúcou sa dávkou lieku.

Motivovaná zlyhaním TeGenero štúdie (viď Senn et al. [2007]), Baileyová v článku Bailey [2009a] skúmala takéto štúdie a modelovala ich ako blokové experimenty:

$$Y_j = \mu + \tau_{i(j)} + \eta_{k(j)} + \varepsilon_j, \quad j = 1, \ldots, N, \tag{R5}$$

kde Y_j je výsledok merania (napr. krvný tlak, alebo iná kvantitatívna veličina indikujúca bezpečnosť a tolerovateľnosť ošetrenia), $i = 1$ je placebo, $i = 2, \ldots, v$ sú zvyšujúce sa dávky lieku, τ_i sú ich vplyvy a η_k je vplyv k-tej kohorty. Obmedzenia na postupne sa zvyšujúce dávky znamenajú lineárne ohraničenia pre návrhy experimentu. Okrem štandardných návrhov pozostávajúcich z vyššie popísaných $v-1$ kohort budeme uvažovať aj tzv. rozšírené návrhy (Bailey [2009a]), ktoré obsahujú ešte jednu poslednú kohortu navyše, v ktorej je možné aplikovať ľubovoľné ošetrenie.

Hoci článok Bailey [2009a] sa zaoberal porovnaniami všetkých dvojíc ošetrení, Senn v komentári Senn [2009] pokladal za najvýznamnejšie porovnania dávok s placebom $\tau_i - \tau_1$ $(i > 1)$ a navrhol kritérium LV-optimality postavené na týchto porovnaniach. My budeme uvažovať práve experimentálny cieľ porovnania dávok s placebom, čo sú vlastne porovnania ošetrení s kontrolným ošetrením. Existuje veľké množstvo výsledkov pre porovnania s kontrolným ošetrením (napr. Majumdar and Notz [1983], Majumdar [1996]), avšak tieto nie sú použiteľné v štúdiách so zvyšujúcou sa dávkou lieku vzhľadom na vyššie popísané ohraničenia pre návrhy experimentu.

Sennovým návrhom sa nazýva taký návrh ξ, ktorý spĺňa $\xi(0,k) = \xi(k,k) = 1/([2(v-1)]$ pre každé $k = 1, \ldots, v-1$ a $\xi(i,k) = 0$ inak. Hodnota $\xi(i,k)$ značí hodnotu návrhu ξ pre ošetrenie i a kohortu k. Pre kritérium E-optimality získavame úplnú charakterizáciu všetkých štandardných aj rozšírených návrhov.

Veta 4. *Štandardný návrh ξ je E-optimálny pre porovnania dávok s placebom práve vtedy, keď ξ je Sennov návrh.*

Veta 5. *Rozšírený návrh ξ je E-optimálny pre porovnania dávok s placebom práve vtedy, keď ξ spĺňa $\xi(0,k) = 1/[2(v-1)]$ pre $k = 1, \ldots, v$ a $\sum_k \xi(i,k) = 1/(2v)$ pre $i > 1$.*

Na základe vzťahu medzi E-optimalitou a c-optimalitou formulujeme štatistickú interpretáciu získaných E-optimálnych návrhov: minimalizujú súčet variancií a absolútnych hodnôt kovariancií pre kontrasty záujmu.

Veta 6. *Každý štandardný alebo rozšírený E-optimálny návrh minimalizuje*

$$\sum_{i=2}^{v} \mathrm{Var}(\widehat{\tau_i - \tau_1}) + \sum_{i=2}^{v} \sum_{j>1, j\neq i} |\mathrm{Cov}(\widehat{\tau_i - \tau_1}, \widehat{\tau_j - \tau_1})|.$$

Pre kritérium LV-optimality, ktoré minimalizuje $\mathrm{Var}(\widehat{\tau_i - \tau_1})$ po prvých $i-1$ kohortách, pre každé $i = 2, \ldots, v$, tiež získavame optimálne návrhy.

Veta 7. *Nech ξ je Sennov návrh. Potom ξ je LV-optimálny štandardný návrh.*

Rozšírený návrh je LV-optimálny, ak minimalizuje $\mathrm{Var}(\widehat{\tau_i - \tau_1})$ po prvých $i-1$ kohortách, pre každé $i = 2, \ldots, v$, a tiež minimalizuje $\mathrm{Var}(\widehat{\tau_v - \tau_1})$ po v kohortách. Pre rozšírené návrhy získavame podobné výsledky ako pre štandardné.

Veta 8. *Nech ξ spĺňa $\xi(1,k) = \xi(k,k) = 1/(2v)$ pre $k = 1, \ldots, v-1$, $\xi(1,v) = \xi(v,v) = 1/(2v)$ a $\xi(i,k) = 0$ inak. Potom ξ je LV-optimálny rozšírený návrh.*

Optimalita Sennových návrhov sa dá ukázať aj pre kritérium MV-optimality.

Veta 9. *Nech ξ je Sennov návrh. Potom ξ je MV-optimálny štandardný návrh pre porovnania dávok s placebom.*

Zdá sa, že Sennove návrhy sú rozumnou voľbou pre štúdie so zvyšujúcou sa dávkou lieku, pokiaľ sa zaujímame o porovnania dávok s placebom. Okrem optimality vzhľadom na viaceré kritériá majú veľmi jednoduchý tvar s veľkým množstvom nulových hodnôt, ktorý umožňuje ich priamočiare popísanie pre potreby praxe a zároveň umožňuje relatívne jednoduché získanie efektívnych alebo optimálnych exaktných návrhov zo Sennových návrhov.

Optimálne návrhy proporcií ošetrení reprezentované grafmi

Predpokladajme, že v modeli (R3) sa zaujímame o nejaký systém $Q^T\tau$ porovnaní s dvojíc ošetrení $\tau_j - \tau_i$, napríklad o porovnania s kontrolným ošetrením $\tau_i - \tau_1$ $(i > 1)$. Skonštruujeme reprezentáciu problému optimálneho návrhu pre $Q^T\tau$ v (R3) pomocou teórie grafov. Definujme (orientovaný) graf s vrcholmi $V = \{1, \ldots, v\}$, pričom existuje hrana idúca z vrcholu j do vrcholu i práve vtedy, keď porovnanie $\tau_j - \tau_i$ je súčasťou systému záujmu. Potom matica incidencie R tohto grafu je totožná s maticou Q.

Uvažujme ortogonálne invariantné kritérium Φ, teda také, ktoré závisí iba od vlastných hodnôt informačnej matice. Potom môžeme zapísať $\Phi(N) = \phi(\lambda_1(N), \ldots, \lambda_s(N))$ pre $N \in \mathfrak{S}^s_+$, pre nejakú funkciu ϕ. Ak pre daný návrh w vrcholom grafu priradíme váhy tak, že vrchol i má váhu $1/w(i)$, potom hľadanie Φ-optimálneho návrhu vieme zapísať pomocou Laplaciánu $\mathcal{L}_w$ tohto vrcholovo váženého grafu. Definícia Laplaciánu vrcholovo váženého grafu je poskytnutá v Chung and Langlands [1996].

Veta 10. *Nech Φ je ortogonálne invariantná informačná funkcia a nech w je návrh pre systém porovnaní dvojíc ošetrení $Q^T\tau$ v* (R3)*. Potom*

$$\Phi(N_Q(w)) = \phi(1/\lambda_r(\mathcal{L}_w), \ldots, 1/\lambda_1(\mathcal{L}_w), 0, \ldots, 0),$$

kde $r = \operatorname{rank}(Q)$.

Poznamenajme, že získaný vzťah sa líši od dobre známeho vzťahu medzi blokovými návrhmi a grafmi, ktorý je analyzovaný napr. v Cameron and Van Lint [1975], Cheng [1981].

Na základe odvodeného vzťahu medzi optimálnymi návrhmi proporcií ošetrení a teóriou grafov vieme sformulovať interpretácie A-, E- a D-optimality. Výrazom $\kappa_s(G)$ značíme súčet váh všetkých ku koreňom orientovaných lesov (rooted directed spanning forests) s s koreňmi, viď Chung and Langlands [1996]. Vážený stupeň vrcholu i s váhou α_i je $\tilde{d}_i = \alpha_i d_i$, kde d_i je stupeň vrcholu i.

Tvrdenie 2. *Nech $Q^T\tau$ je systém porovnaní dvojíc ošetrení s $\operatorname{rank}(Q) = r$ a nech w je návrh pre $Q^T\tau$ v* (R3)*. Potom ak G je príslušný vrcholovo vážený graf, tak*

(i) $\Phi_A(N_Q(w)) = \sum_{i=1}^{v} \tilde{d}_i$,

(ii) $\Phi_E(N_Q(w)) = \max_{\|y\|_w=1} \sum (y_i - y_j)^2$, *kde suma ide cez všetky hrany (i, j) grafu G a $\|y\|_w = (\sum_i w(i) y_i^2)^{1/2}$,*

(iii) $\Phi_D(N_Q(w)) = \kappa_{v-r}(G)$.

Pomocou grafovej reprezentácie môžeme vyjadriť A-optimálny návrh pre $Q^T\tau$:

$$w(i) = \frac{\sqrt{d_i}}{\sum_j \sqrt{d_j}}, \quad i = 1, \ldots, v,$$

kde d_i sú stupne vrcholov prislúchajúceho grafu G. Tento vzťah umožňuje veľmi jednoduché počítanie A-optimálnych návrhov, len na základe nakresleného grafu. Poznamenajme, že existuje všeobecný vzorec pre A-optimálne návrhy (Dôsledok 8.8 v Pukelsheim [1993]).

Systém porovnaní dvojíc ošetrení nazývame *bipartitný*, ak príslušný graf G je bipartitný. Pre takéto systémy vieme analyticky získať E-optimálne návrhy.

Veta 11. *Nech $Q^T\tau$ je bipartitný systém porovnaní dvojíc ošetrení a nech G je príslušný graf. Potom návrh w definovaný ako*

$$w(i) = \frac{d_i}{\sum_j d_j}, \quad i = 1, \ldots, v,$$

je E-optimálny pre $Q^T\tau$.

Poznamenajme, že z príslušného grafu G sa dá vyčítať aj vlastný vektor prislúchajúci k $\lambda_{\max}(N_Q(w))$ pre E-optimálny návrh w (pre bipartitné systémy). Následne získavame interpretáciu E-optimálnych návrhov podobnú ako v štúdiách so zvyšujúcou sa dávkou lieku.

Veta 12. *Nech w je E-optimálny návrh daný vetou 11. Potom w minimalizuje*

$$\sum_{i=1}^{s} \mathrm{Var}(\widehat{q_i^T\tau}) + \sum_{i=1}^{s}\sum_{j\neq i} |\mathrm{Cov}(\widehat{q_i^T\tau}, \widehat{q_j^T\tau})|,$$

kde $Q = (q_1, \ldots, q_s)$.

Je známe, že ak $s = v - 1$ a Q má plnú hodnosť, tak rovnomerný návrh $\bar{w}$ je D-optimálny. Rozšírme toto tvrdenie na $s \geq v - 1$ kontrastov (nie nutne porovnaní dvojíc ošetrení). Rovnomerný návrh spĺňa $\bar{w}(i) = 1/v$ pre každé i.

Veta 13. *Nech $Q^T\tau$ je systém $s \geq v - 1$ ošetrení a* $\mathrm{rank}(Q) = v - 1$. *Potom $\bar{w}$ je D-optimálny pre $Q^T\tau$.*

Pomocou grafovej reprezentácie vieme získať celú triedu systémov $Q^T\tau$, pre ktoré je rovnomerný návrh $\bar{w}$ optimálny vzhľadom na ľubovoľnú ortogonálne invariantnú informačnú funkciu Φ.

Veta 14. *Nech $Q^T\tau$ je systém porovnaní dvojíc ošetrení a nech G je príslušný graf. Ak existuje cyklická permutácia π, ktorá je automorfizmum grafu G, tak $\bar{w}$ je Φ-optimálny pre $Q^T\tau$ v* (R3) *vzhľadom na ľubovoľnú ortogonálne invariantnú informačnú funkciu Φ. Ak navyše existuje ostro konkávna verzia Φ, tak $\bar{w}$ je jediný Φ-optimálny návrh.*

Permutácia je cyklická, ak pozostáva iba z jedného cyklu. Analogický výsledok môžeme získať pre ľubovoľné systémy kontrastov, a môžeme ho rozšíriť aj na necyklické permutácie π. Najprv označme P_π permutačnú maticu prislúchajúcu k permutácii π.

Veta 15. *Nech $Q^T\tau$ je systém kontrastov, nech Φ je ortogonálne invariantná informačná funkcia a nech π je permutácia spĺňajúca $P_\pi QQ^T P_\pi^T = QQ^T$. Nech π pozostáva z cyklov $c_1, \ldots, c_K$. Potom existuje Φ-optimálny návrh w spĺňajúci $w(i) = w(j)$ pre každé $i, j \in c_k$ $(k = 1, \ldots, K)$. Ak navyše existuje ostro konkávna verzia Φ, tak každý Φ-optimálny návrh musí byť rovnomerný na každom cykle c_k.*

V modeli (R3) sa špeciálne pozrime na systém porovnaní $v - g$ ošetrení s g kontrolnými ošetreniami: $\tau_j - \tau_i$, $i = 1, \ldots, g$, $j = g + 1, \ldots, v$, pre nejaké $0 < g < v - 1$. Skúmame triedu kritérií tzv. Kieferovej Φ_p-optimality pre $p \in [-\infty, 0]$ (viď napr. Pukelsheim [1993]), ktorá zahŕňa známe kritériá D-, A- a E-optimality.

Veta 16. *Nech $p \in [-\infty, 0]$. Ak $p > -\infty$, nech γ_p je jedinečné riešenie rovnice*

$$(v - g - 1)\gamma^{1-p} - (g - 1)(1 - \gamma)^{1-p} + 2\gamma - 1 = 0$$

na intervale $(0, 1/2]$. Navyše položme $\gamma_{-\infty} = 1/2$. Potom návrh w daný $w(i) = \gamma_p/g$ pre $i = 1, \ldots, g$ a $w(i) = (1 - \gamma_p)/(v - g)$ pre $i = g + 1, \ldots, v$, je jediný Φ_p-optimálny návrh pre porovnanie $v - g$ ošetrení s g kontrolnými ošetreniami.

Vetou 16 zovšeobecňujeme výsledky Giovagnoli and Wynn [1985] pre porovnanie s jedným kontrolným ošetrením na všeobecný počet kontrolných ošetrení.

Vážená optimalita

Poznamenajme, že v tejto kapitole sa primárne budeme zaoberať exaktnými návrhmi, hoci výsledky sú prirodzene rozšíriteľné aj na aproximatívne návrhy. Morgan a Wang (Morgan and Wang [2010] a Morgan and Wang [2011]) navrhli téoriu váženej optimality pre návrhy experimentov, pričom dosiaľ najviac všeobecnej formulácie sa táto teória dočkala v Stallings and Morgan [2015]. Tento prístup umožňuje klásť rôzny dôraz na rôzne parametre, alebo systémy záujmu $Q^T\tau$. Technicky sa najprv zvolí kladne definitná matica váh W parametrov τ v modeli

$$Y = X(\xi)\tau + L(\xi)\theta + \varepsilon,$$

kde (exaktný) návrh ξ určuje maticu $X(\xi)$, ktorá vyjadruje vzťah medzi parametrami záujmu τ a Y, a maticu $L(\xi)$, ktorá vyjadruje vzťah medzi rušivými parametrami θ a Y. Následne namiesto informačnej matice $C(\xi) = X^T(\xi)(I - P_L)X(\xi)$ pre vektor τ pracujeme s váženou informačnou maticou $C_W(\xi) = W^{-1/2}C(\xi)W^{-1/2}$. Matica P_L je $P_L = L(L^TL)^-L^T$, kde $L = L(\xi)$.

Vyžadujeme, aby sme pre každý uvažovaný návrh experimentu mali rovnakú množinu odhadnuteľných funkcií $q^T\tau$, teda všetky zvažované návrhy majú rovnaký stĺpcový priestor $\mathcal{C}(C(\xi)) =: \mathcal{E}$. Napríklad v modeli (R2) je $\mathcal{E}$ množinou všetkých vektorov kolmých na 1_v a reprezentuje tak kontrasty ošetrení. Každá matica váh W priraďuje váhy

odhadnuteľným funkciám $q^T\tau$ (pre $q \in \mathcal{E}$) vyjadrené ako $w(q) = (q^T W^{-1} q)^{-1}$. Ak chceme vyjadriť záujem o systém $Q^T\tau$, môžeme to spraviť skonštruovaním príslušnej matice váh $W_Q = I - P_\tau + QQ^T$, kde P_τ je maticou ortogonálnej projekcie na $\mathcal{E}$. Táto matica váh je však poskytnutá v Stallings and Morgan [2015] iba pre systémy $Q^T\tau$ spĺňajúce $\text{rank}(Q) = \dim(\mathcal{E})$. Ak by sme jednotlivým funkciám $q_1^T\tau, \ldots, q_s^T\tau$ systému $Q^T\tau$ chceli priradiť váhy $b_1, \ldots, b_s$, ľahko to zabezpečíme transformáciou na systém $\tilde{Q}^T\tau$, kde $\tilde{q}_i = \sqrt{b_i} q_i$. V takom prípade dostávame žiaduce $w(q_i) = b_i$ pre $i = 1, \ldots, s$ pokiaľ Q má plnú hodnosť. Ak Q plnú hodnosť nemá, tento vzťah vo všeobecnosti neplatí.

Keďže vážená optimalita prislúchajúca k systému $Q^T\tau$ si kladie za cieľ vyjadriť experimentálny záujem o $Q^T\tau$, pričom tento experimentálny cieľ vie vyjadriť aj optimalita pre $Q^T\tau$ používaná v predchádzajúcich podkapitolách, vyvstáva otázka, či sa tieto dva prístupy líšia. Ukážeme, že sú ekvivalentné.

Veta 17. *Nech $Q^T\tau$ je systém odhadnuteľných funkcií, ktorý spĺňa* $\text{rank}(Q) = \dim(\mathcal{E})$ *a nech W_Q je príslušná matica váh. Potom pre akékoľvek kritérium Φ závisiace iba od vlastných hodnôt je 'štandardná' Φ-optimalita pre $Q^T\tau$ ekvivalentná s váženou Φ-optimalitou vzhľadom na W_Q. Teda tieto dva prístupy tvoria rovnaké usporiadanie návrhov.*

Naopak, ak by sme chceli pre danú maticu váh W získať systém záujmu $Q^T\tau$, ktorý prislúcha k W, stačí zvoliť $Q_W = ((P_\tau W^{-1} P_\tau)^+)^{1/2}$.

Veta 18. *Nech W je matica váh a nech $Q_W = ((P_\tau W^{-1} P_\tau)^+)^{1/2}$. Potom pre akékoľvek kritérium Φ závisiace iba od vlastných hodnôt je štandardná Φ-optimalita pre $Q_W^T\tau$ ekvivalentná s váženou Φ-optimalitou vzhľadom na W.*

Aby sme dokázali zachytiť váženou optimalitou ľubovoľný systém záujmu, rozšírime definíciu matice váh a následne poskytneme vhodnú relaxáciu váh odhadnuteľných funkcií a váženej informačnej matice. Pri hľadaní optimálnych návrhov vzhľadom na maticu W vyžadujeme iba, aby tieto návrhy spĺňali $\mathcal{C}(W) \subseteq \mathcal{C}(C(\xi))$. Pre maticu W hodnosti d vhodne definujeme $v \times d$ maticu K_W, ktorá je hodnosti d a spĺňa $K_W K_W^T = W$.

Definícia 1. *(i) Matica W je maticou váh, ak $W \in \mathfrak{S}_+^v$ a $\mathcal{C}(W) \subseteq \mathcal{E}$.*

(ii) Pre $q \in \mathcal{C}(W)$ je váha $q^T\tau$ daná ako $w(q) = (q^T W^- q)^{-1}$. Ak $q \notin \mathcal{C}(W)$, tak váha $q^T\tau$ je 0.

(iii) Vážená informačná matica návrhu ξ je $C_W(\xi) = (K_W^T C^-(\xi) K_W)^{-1}$.

Navrhnutá definícia váženej optimality spĺňa základnú vlastnosť, ktorá sa od váženej optimality požaduje:

Tvrdenie 3. *Nech W je matica váh hodnosti d, nech $q \in \mathcal{C}(W)$ a nech ξ je návrh spĺňajúci $q \in \mathcal{C}(C(\xi))$. Potom vážená variancia $\widehat{q^T\tau}$ pod ξ je konvexnou kombináciou $\lambda_1^{-1}(C_W(\xi)), \ldots, \lambda_d^{-1}(C_W(\xi))$.*

Následne môžeme skonštruovať maticu váh prislúchajúcu *ľubovoľnému* systému odhadnuteľných funkcií $Q^T\tau$, ktorá má navyše jednoduchší tvar než $I - P_\tau + QQ^T$.

Definícia 2. *Nech $Q^T\tau$ je systém s odhadnuteľných funkcií. Potom príslušná matica váh je $W_Q = QQ^T$. Ak funkciám $q_1^T\tau, \ldots, q_s^T\tau$ priraďujeme váhy $b_1, \ldots, b_s$, tak príslušná matica váh je $W_{\tilde{Q}} = \tilde{Q}\tilde{Q}^T$.*

Pôvodná príslušná matica váh a novo definovaná matica váh sú ekvivalentné, ak tá prvá existuje.

Veta 19. *Nech $Q^T\tau$ je systém záujmu, ktorý spĺňa* $\mathrm{rank}(Q) = \dim(\mathcal{E})$. *Potom $W_1 = QQ^T$ a $W_2 = I - P_\tau + QQ^T$ sú ekvivalentné vzhľadom na implikované váhy odhadnuteľných funkcií. Teda $q^T W_1^- q = q^T W_2^{-1} q$ pre ľubovoľné $q \in \mathcal{E}$.*

Analogicky k vetám 17 a 18 je navrhnutá vážená optimalita ekvivalentná so štandardnou optimalitou pre $Q^T\tau$. Pre danú maticu váh W je prisluchajúci systém záujmu $Q^T\tau$ pre $Q = W^{1/2}$. Ako v predchádzajúcom prípade, dá sa ukázať, že pokiaľ Q má plnú hodnosť, tak $w(q_i) = b_i$ pre $i = 1, \ldots, s$.

Ak Q nemá plnú hodnosť, tak vo všeobecnosti $w(q_i) = b_i$ neplatí. Aby sme tento vzťah osvetlili, rozlíšme medzi *primárnymi* váhami b_i funkcií $q_i^T\tau$ a *sekundárnymi* váhami $w(q)$ funkcií $q^T\tau$ pre $q \in \mathcal{C}(Q)$. Ak Q nemá plnú hodnosť, niektoré z funkcií $q_i^T\tau$ si navzájom poskytujú váhy, čo vyústi do vyšších sekundárnych váh. Navyše, ako sa ukazuje, aj keď majú funkcie $q_i^T\tau$ rovnaké primárne váhy, tak optimálne návrhy majú tendenciu klásť väčší dôraz na tie z nich s väčšími sekundárnymi váhami - teda odhady funkcií $q_i^T\tau$ s väčšími sekundárnymi váhami majú tendenciu mať nižšie variancie. Čiže na interpretáciu váženej optimality by zrejme mali byť používané sekundárne, nie primárne váhy.

References

A. C. Atkinson. Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine*, 18:1741–1752, 1999.

A. C. Atkinson. Commentary on 'Designs for dose—escalation trials with quantitative responses'. *Statistics in Medicine*, 28:3739–3741, 2009.

A. C. Atkinson and A. N. Donev. Experimental design optimally balanced for trend. *Technometrics*, 38:333–341, 1996.

A. C. Atkinson, A. Donev, and R. Tobias. *Optimum experimental designs, with SAS*. Oxford University Press, New York, 2007.

R. A. Bailey. Designs for dose-escalation trials with quantitative responses. *Statistics in Medicine*, 28:3721–3738, 2009a.

R. A. Bailey. Author's rejoinder to commentaries on 'Designs for dose—escalation trials with quantitative responses'. *Statistics in Medicine*, 28:3759–3760, 2009b.

R. A. Bailey and P. J. Cameron. Combinatorics of optimal designs. *Surveys in Combinatorics*, 365:19–73, 2009.

R. A. Bailey, C.-S. Cheng, and P. Kipnis. Construction of trend-resistant factorial designs. *Statistica Sinica*, 2:393–411, 1992.

R. B. Bapat. *Graphs and Matrices*. Springer, New York, 2010.

R. E. Bechhofer and A. C. Tamhane. Incomplete block designs for comparing treatments with a control: General theory. *Technometrics*, 23:45–57, 1981.

A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, Philadelphia, 1994.

R. A. Bradley and C. M. Yeh. Trend-free block designs: Theory. *The Annals of Statistics*, 8:883–893, 1980.

P. J. Cameron and J. H. Van Lint. *Graph theory, coding theory and block designs*. Cambridge University Press, London, 1975.

C.-S. Cheng. Maximizing the total number of spanning trees in a graph: Two related problems in graph theory and optimum design theory. *Journal of Combinatorial Theory, Series B*, 31:240–248, 1981.

C.-S. Cheng. Construction of run orders of factorial designs. In S. Ghosh, editor, *Statistical Design and Analysis of Industrial Experiments*, pages 423–439. Marcel-Dekker, New York, 1990.

F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, Providence, 1997.

F. R. K. Chung and R. P. Langlands. A combinatorial laplacian with vertex weights. *Journal of combinatorial theory (Series A)*, 75:316–327, 1996.

D. R. Cox. Some systematic experimental designs. *Biometrika*, 38:312–323, 1951.

D. R. Cox and N. Reid. *Theory of the Design of Experiments*. Chapman and Hall, London, 2000.

D. Cvetković, P. Rowlinson, and S. Simić. *An Introduction to the Theory of Graph Spectra*. Cambridge University Press, Cambridge, 2010.

P. T. Darga, Sakallah K. A., and Markov I. L. Faster symmetry discovery using sparsity of symmetries. In *Proceedings of the 45th Design Automation Conference*, pages 149–154, California, 2008. IEEE.

R Diestel. *Graph Theory*. Springer, New York, 2nd edition, 2000.

J. A. Eccleston and A. Hedayat. On the theory of connected designs: characterization and optimality. *The Annals of Statistics*, 2:1238–1255, 1974.

K. Filipiak, A. Markiewicz, and A. Szczepanska. Optimal designs under a multivariate linear model with additional nuisance parameters. *Statistical Papers*, 50:761–778, 2009.

R. A. Fisher. *The design of experiments*. Oliver and Boyd, Edinburgh, 1935.

A. Giovagnoli and I. Verdinelli. Optimal designs for comparing test treatments with controls: Comment. *Statistical Science*, 3:482–484, 1988.

A. Giovagnoli and H. P. Wynn. Schur-optimal continuous block designs for treatments with a control. In L. M. Le Cam and R. A. Olshen, editors, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, pages 651–666, California, 1985. Wadsworth.

F. Githinji and M. Jacroux. On the determination and construction of optimal designs for comparing a set of test treatments with a set of controls in the presence of a linear trend. *Journal of Statistical Planning and Inference*, 66:161–174, 1998.

T. N. E. Greville. Note on the generalized inverse of a matrix product. *SIAM Review*, 8:518–521, 1966.

L. M. Haines and A. E. Clark. The construction of optimal designs for dose-escalation studies. *Statistics and Computing*, 24:101–109, 2014.

R. Harman. Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika*, 60:137–153, 2004.

R. Harman and G. Sagnol. Computing D-optimal experimental designs for estimating treatment contrasts under the presence of a nuisance time trend. In A. Steland, E. Rafajlowicz, and K. Szajowski, editors, *Stochastic Models, Statistics and Their Applications*, pages 83–91. Springer, Wroclaw, 2015.

R. Harman, A. Bachratá, and L. Filová. Construction of efficient experimental designs under multiple resource constraints. *Applied Stochastic Models in Business and Industry*, 32:3–17, 2016.

D. A. Harville. *Matrix Algebra From A Statistician's Perspective*. Springer-Verlag, New York, 1997.

A. S. Hedayat, M. Jacroux, and D. Majumdar. Optimal designs for comparing test treatments with controls. *Statistical Science*, 3:462–476, 1988.

M. Jacroux. Some E-optimal designs for the one-way and two-way elimination of heterogeneity. *Journal of the Royal Statistical Society: Series B*, 44:253–261, 1982.

M. Jacroux. On the determination and construction of MV-optimal block designs for comparing test treatments with a standard treatment. *Journal of Statistical Planning and Inference*, 15:205–225, 1987.

M. Jacroux. Some optimal designs for comparing a set of test treatments with a set of controls. *Annals of the Institute of Statistical Mathematics*, 42:173–185, 1990.

M. Jacroux. On the construction of trend-resistant designs for comparing a set of test treatments with a set of controls. *Journal of the American Statistical Association*, 88:1398–1403, 1993.

M. Jacroux. A- and MV-efficient block designs for comparing a set of controls to a set of test treatments. *Sankhya: The Indian Journal of Statistics, Series B*, 64:141–161, 2002.

M. Jacroux and R. S. Ray. On the construction of trend-free run orders of treatments. *Biometrika*, 77:187–191, 1990.

M. Jacroux, D. Majumdar, and K. R. Shah. On the determination and construction of optimal block designs in the presence of linear trends. *Journal of the American Statistical Association*, 92:375–382, 1997.

J.. Kiefer. Optimum experimental designs. *Journal of Royal Statistical Society*, 21: 272–319, 1959.

C. P. Kitsos, D. M. Titterington, and B. Torsney. An optimal design problem in rhythmometry. *Biometrics*, 44:657–671, 1988.

O. Knill. Cauchy–Binet for pseudo-determinants. *Linear Algebra and its Applications*, 459:522–547, 2014.

J. Kunert. Optimal design and refinement of the linear model with applications to repeated measurements designs. *The Annals of Statistics*, 11:247–257, 1983.

J. Kunert, R. J. Martin, and J. Eccleston. Optimal block designs comparing treatments with a control when the errors are correlated. *Journal of Statistical Planning and Inference*, 140:2719–2738, 2010.

D. Majumdar. Optimal designs for comparisons between two sets of treatments. *Journal of Statistical Planning and Inference*, 14:359–372, 1986.

D. Majumdar. Optimal and efficient treatment-control designs. In S. Ghosh and C. R. Rao, editors, *Handbook of statistics 13: Design and Analysis of Experiments*, pages 1007–1053. North Holland, Amsterdam, 1996.

D. Majumdar and W. I. Notz. Optimal incomplete block designs for comparing treatments with a control. *The Annals of Statistics*, 11:258–266, 1983.

R. Mead. *Statistical methods for dividing sites into recommendation domains on the basis of experimental results.* International Maize and Wheat Improvement Center, Mexico City, 1990.

R. Merris. Laplacian matrices of graphs: A survey. *Linear Algebra and Its Applications*, 197:143–176, 1994.

R. Merris. A survey of graph laplacians. *Linear and Multilinear Algebra*, 39:19–31, 1995.

B. Mohar. The laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph theory, combinatorics, and applications 2*, pages 871–898. Wiley, New York, 1991.

J. P. Morgan and J. Stallings. Optimal experimental design that targets meaningful information. *WIREs Computational Statistics*, 9, 2017.

J. P. Morgan and X. Wang. Weighted optimality in designed experimentation. *Journal of the American Statistical Association*, 105:1566–1580, 2010.

J. P. Morgan and X. Wang. E-optimality in treatment versus control experiments. *Journal of Statistical Theory and Practice*, 5:99–107, 2011.

W. I. Notz. Optimal designs for treatment—control comparisons in the presence of two-way heterogeneity. *Journal of Statistical Planning and Inference*, 12:61–73, 1985.

A. Pázman. *Foundation of Optimum Experimental Design.* Reidel Publ., Dordrecht, 1986.

S. J. Pocock and R. Simon. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31:103–115, 1975.

F. Pukelsheim. On linear regression designs which maximize information. *Journal of Statistical Planning and Inference*, 4:339–364, 1980.

F. Pukelsheim. On optimality properties of simple block designs in the approximate design theory. *Journal of Statistical Planning and Inference*, 8:193–208, 1983.

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

F. Pukelsheim and W. J. Studden. E-optimal designs for polynomial regression. *The Annals of Statistics*, 21:402–415, 1993.

S. Rosa. Trend resistant experimental designs. Master's thesis, Comenius University in Bratislava, 2014.

S. Rosa. On weighted optimality of experimental designs. arXiv:1610.06427 [math.ST], 2016.

S. Rosa. Optimal designs for treatment comparisons represented by graphs. *AStA Advances in Statistical Analysis.*, 2017. doi: 10.1007/s10182-017-0312-5.

S. Rosa and R. Harman. Optimal approximate designs for estimating treatment contrasts resistant to nuisance effects. *Statistical Papers*, 57:1077–1106, 2016.

S. Rosa and R. Harman. Optimal approximate designs for comparison with control in dose-escalation studies. *TEST*, 26:638–660, 2017.

G. Sagnol. Computing optimal designs of multiresponse experiments reduces to second-ordercone programming. *Journal of Statistical Planning and Inference*, 141:1684–1708, 2011.

G. Sagnol and R. Harman. Computing exact D-optimal designs by mixed integer second-order cone programming. *The Annals of Statistics*, 43:2198–2224, 2015.

R. Schwabe. Optimal designs for additive linear models. *Statistics*, 27:267–278, 1996.

R. Schwabe and W. Wierich. D-optimal designs of experiments with non-interacting factors. *Journal of Statistical Planning and Inference*, 44:371–384, 1995.

G. A. Seber. *A Matrix Handbook for Statisticians.* John Wiley & Sons, New Jersey, 2008.

S. Senn. *Statistical Issues in Drug Development.* Wiley, Chichester, 1997.

S. Senn. Commentary on 'Designs for dose-escalation trials with quantitative responses'. *Statistics in Medicine*, 28:3754—3758, 2009.

S. Senn, D. Amin, R. A. Bailey, S. M. Bird, B. Bogacka, P. Colman, A. Garrett, A. Grieve, and P. Lachmann. Statistical issues in first-in-man studies. *Journal of the Royal Statistical Society: Series A*, 170:517–579, 2007.

J. W. Stallings and J. P. Morgan. General weighted optimality of designed experiments. *Biometrika*, 102:925–935, 2015.

J. M. Ver Hoef and N. Cressie. Spatial statistics: Analysis of field experiments. In S. M. Scheiner and J. Gurevitch, editors, *Design and analysis of ecological experiments*, pages 289–307. Oxford University Press, Oxford, 2nd edition, 2001.

C. F. Wu. On some ordering properties of the generalized inverses of non-negative definite matrices. *Linear Algebra and Its Applications*, 32:49–60, 1980.

Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, 38:1593–1606, 2010.