

General Weighted Optimality of Designed Experiments

Jonathan W. Stallings

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

John P. Morgan, Chair
Xinwei Deng
Klaus Hinkelmann
Bradley Jones

March 21, 2014
Blacksburg, Virginia

Keywords: Optimal design, baseline parameterization, weighted variance, weighted information matrix, weighted optimality criteria, A_W -optimality, limiting weights, reduced models

Copyright 2014, Jonathan W. Stallings

General Weighted Optimality of Designed Experiments

Jonathan W. Stallings

ABSTRACT

Design problems involve finding optimal plans that minimize cost and maximize information about the effects of changing experimental variables on some response. Information is typically measured through statistically meaningful functions, or criteria, of a design's corresponding information matrix. The most common criteria implicitly assume equal interest in all effects and certain forms of information matrices tend to optimize them. However, these criteria can be poor assessments of a design when there is unequal interest in the experimental effects. Morgan and Wang (2010) addressed this potential pitfall by developing a concise weighting system based on quadratic forms of a diagonal matrix \mathbf{W} that allows a researcher to specify relative importance of information for any effects. They were then able to generate a broad class of weighted optimality criteria that evaluate a design's ability to maximize the weighted information, ultimately targeting those designs that efficiently estimate effects assigned larger weight.

This dissertation considers a much broader class of potential weighting systems, and hence weighted criteria, by allowing \mathbf{W} to be any symmetric, positive definite matrix. Assuming the response and experimental effects may be expressed as a general linear model, we provide a survey of the standard approach to optimal designs based on real-valued, convex functions of information matrices. Motivated by this approach, we introduce fundamental definitions and preliminary results underlying the theory of general weighted optimality.

A class of weight matrices is established that allows an experimenter to directly assign weights to a set of estimable functions and we show how optimality of transformed models may be placed under a weighted optimality context. Straightforward modifications to SAS PROC OPTEX are shown to provide an algorithmic search procedure for weighted optimal designs, including A -optimal incomplete block designs. Finally, a general theory is given for design optimization when only a subset of all estimable functions is assumed to be in the model. We use this to develop a weighted criterion to search for A -optimal completely randomized designs for baseline factorial effects assuming all high-order interactions are negligible.

Acknowledgements

One of the most difficult and important decisions a PhD student faces is choosing an advisor. I am fortunate to have had Dr. John P. Morgan as my PhD advisor, who has offered invaluable guidance and mentorship throughout my graduate career. His elegant insight and ability to push me to new levels were crucial to my success.

I would also like to extend my gratitude to Drs. Xinwei Deng, Klaus Hinkelmann, and Brad Jones, who were gracious enough to serve on my committee. Their advice and suggestions are greatly appreciated and I hope to work with them again in the future.

Finally, I want to thank my family, especially my father, Wade, brother, Robert, and sister, Amy, and beloved partner, Sarah, who provided constant support during many personal challenges, including my mother's unexpected passing. I would like to dedicate this dissertation to my mother, Delores Tiller Stallings, who would have been immensely proud to see it completed.

Contents

1	Introduction	1
1.1	General linear model and estimation space	1
1.1.1	The Moore-Penrose inverse and square-root decompositions	5
1.2	Standard optimality	5
1.2.1	Treatment-versus-control experiments	9
1.2.2	Factorial experiments	10
1.2.3	Baseline parameterization of factorials	15
1.3	General weighted optimality	21
1.4	Outline	22
2	Preliminary Results	23
2.1	Weight matrices and weighted variances	23
2.2	Weighted information matrices	25
2.3	Weighted eigenvalue criteria	28
2.3.1	Weighted- E criterion	28
2.3.2	Weighted- A criterion	29
2.3.3	Weighted- D criterion	31
2.4	Estimation equivalence	33
2.5	General weighted optimality criteria	37

2.5.1	Permutation invariance for connected designs	42
2.6	Matrix averaging, weight balance, and permutation invariance	45
3	Weighted optimality for a set of estimable functions	50
3.1	Model transformations and weighted optimality	51
3.1.1	The baseline and sum-to-zero transform	53
3.2	Assigning weights to estimable functions	54
3.2.1	Assigning weights to k orthonormal estimable functions	55
3.2.2	An admissibility criterion	55
3.2.3	Assigning weights to k estimable functions	57
3.2.4	Normalization	59
3.2.5	Assigning weights to $s > k$ estimable functions	62
3.3	A_W -optimality for $s \geq k$ estimable functions	63
3.3.1	Approximate and exact A_W -optimal CRDs	68
3.4	A_W -optimal CRDs for two-level OP and BP effects	72
4	Weighted optimality criteria and design search algorithms	85
4.1	Overview and modification of design search algorithms	86
4.2	Adjusted information matrices and limiting weights	88
4.3	A -optimal CRDs for two-level BP effects adjusted for $\theta_{0\dots 0}$	94
4.4	A -optimal incomplete block designs	103
5	Weighted optimality of reduced models	107
5.1	Reduced cell means model and weighting	108
5.2	A -optimal designs for reduced BP 2-f.i. model	111
5.2.1	Eigenvalues of $\mathbf{C}_{d\theta_2}$ for strength 4 BAs	124
5.2.2	A -optimal approximate BAs for reduced BP 2-f.i. model	137

5.2.3	Saturated A -optimal exact BAs for reduced BP 2-f.i. model	140
5.2.4	A -optimal exact BAs for reduced BP 2-f.i. model with $m = 4, 5$, and 6 and $N > \nu$	146
5.3	A -optimal designs for the reduced BP 2-f.i. model with weighted effects . . .	149
6	Future work	156
	Bibliography	158

List of Tables

3.1	Weights assigned by \mathbf{W}_{Hw} in Example 3.2.1 to pairwise comparisons for treatments between and within treatment groups V_1 and V_2 with $w_0 = 20$	66
3.2	Replication proportions for treatments of a given heft for A -optimal approximate CRD with respect to estimation of all non-normalized BP effects. . . .	76
3.3	A -optimal CRDs for estimation of all non-normalized BP effects for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3. . .	82
3.4	A -optimal CRDs for estimation of all normalized BP effects for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.	82
3.5	BA heft replications for A -optimal designs for estimation of all non-normalized BP effects for $m = 3$ to 5 factors with increasing N	83
3.6	BA heft replications for A -optimal designs for estimation of all normalized BP effects for $m = 3$ to 5 factors with increasing N	84
4.1	A -optimal CRDs for estimation of non-normalized BP effects, adjusted for the intercept, for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.	99
4.2	A -optimal CRDs for estimation of normalized BP effects, adjusted for the intercept, for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.	99
4.3	BA heft replications for A -optimal CRDs for estimation of all non-normalized BP effects, adjusted for the intercept, for $m = 3$ to 5 factors with increasing N .100	
4.4	BA heft replications for A -optimal CRDs for estimation of all normalized BP effects, adjusted for the intercept, for $m = 3$ to 5 factors with increasing N . .	101

4.5	Cases where $w < 0.99$ was required for SAS PROC OPTEX to find known A -optimal design for adjusted BP effects.	102
4.6	Number of A -optimal designs found of 100 modified Fedorov iterations for p treatments in h blocks of size k . Missing block weight indicates unweighted approach found the A -optimal design for all 100 iterations.	106
5.1	Optimal ratios for $m = 4, \dots, 10$ for saturated BP 2-f.i. model and optimal A -values and BA index sets for $N = \nu$	140
5.2	A -efficient strength 4 BA index sets for the saturated, reduced BP 2-f.i. model for $m = 4, \dots, 10$. Efficiencies are with respect to the approximate designs in Table 5.1.	142
5.3	Index sets for a 29×7 BA with strength 7 for decreasing strength.	143
5.4	Saturated designs constructed from the strength m BA with $n_0 = n_2 = n_{m-1} = 1$ and their A -efficiencies with respect to the A -optimal approximate BA and A -best exact BA index sets. Efficiencies for BP effects are based on the A -optimal approximate BA.	144
5.5	A -efficient strength 4 BAs with $m = 4$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEX. . .	147
5.6	Potential A -efficient strength 4 BAs with $m = 5$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEX.	148
5.7	Potential A -efficient strength 4 BAs with $m = 6$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEX.	149
5.8	Optimal ratios for the saturated, normalized BP 2-f.i. model, optimal A_W -values, and optimal BA index sets.	154
5.9	Optimal ratios for the saturated BP 2-f.i. model with $w_1 = 20$ and $w_0 = w_2 = 1$, optimal A_W -values, and BA index sets.	154

Chapter 1

Introduction

1.1 General linear model and estimation space

Suppose an experimenter has N experimental units, possibly subject to nuisance factors such as blocks, that are to be assigned some combination of settings for the experimental variables. An optimal design is the “best” allocation of these settings to the N units, relative to a statistically meaningful criterion associated with how the data will be analyzed once collected. For a given design, d , we represent the effects of the experimental variables on the response as parameters in the linear model

$$\mathbf{y} = \mathbf{X}_d \boldsymbol{\tau} + \mathbf{L} \boldsymbol{\beta} + \mathbf{e} , \tag{1.1}$$

where \mathbf{y} is the $N \times 1$ vector of responses; $\boldsymbol{\beta}$ is a $b \times 1$ vector containing all nuisance effects; \mathbf{L} (which often includes the all-ones vector, $\mathbf{1}$, for the intercept) is an $N \times b$ matrix that relates $\boldsymbol{\beta}$ to \mathbf{y} and is not design dependent; $\boldsymbol{\tau}$ is a $p \times 1$ vector of parameters that are the target of inference; \mathbf{X}_d is the $N \times p$ matrix which relates $\boldsymbol{\tau}$ to \mathbf{y} and is generated by design

d ; and \mathbf{e} is an $N \times 1$ vector of random errors. We assume that $E(\mathbf{e}) = \mathbf{0}$ and $\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}_N$, where \mathbf{I}_N is the $N \times N$ identity matrix. That is, the errors are taken to be uncorrelated and have constant variance, σ^2 . Without loss of generality, for the work to be undertaken here, we further assume that $\sigma^2 = 1$.

Model (1.1) encompasses a broad range of experimental scenarios, including regression models ($\boldsymbol{\tau}$ is a vector of slopes) and ANOVA setups like for completely randomized designs and blocked designs ($\boldsymbol{\tau}$ is a vector of treatment effects, $\boldsymbol{\beta}$ contains intercept and block effects). The *cell means model*, introduced by Hocking (1985, 2003), is a specific case of model (1.1) in which $\boldsymbol{\tau}$ is the vector of mean responses for each experimental setting, \mathbf{X}_d is a 0/1 matrix where each row has exactly one 1 that indicates which experimental setting is applied to the corresponding experimental unit, and $b = 0$. This model is one of several common, equivalent options when the experimental settings are comprised of m different factors each having some set number of levels and there are no nuisance factors.

Under model (1.1) and some design d , a least-squares estimator for $\boldsymbol{\tau}$, say $\hat{\boldsymbol{\tau}}$, is a solution to the system of linear equations $\mathbf{C}_d \hat{\boldsymbol{\tau}} = \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L) \mathbf{y}$ where $\mathbf{C}_d = \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L) \mathbf{X}_d$, $\mathbf{P}_L = \mathbf{L}(\mathbf{L}^T \mathbf{L})^{-} \mathbf{L}^T$ is the orthogonal projector onto the column space of \mathbf{L} , and $(\mathbf{L}^T \mathbf{L})^{-}$ denotes any generalized inverse of $\mathbf{L}^T \mathbf{L}$. We say that \mathbf{C}_d is the information matrix for $\boldsymbol{\tau}$ *adjusted* for $\boldsymbol{\beta}$. When the rank of \mathbf{C}_d , $r(\mathbf{C}_d)$, equals p there is a unique solution for $\hat{\boldsymbol{\tau}}$ and thus for every linear combination $\widehat{\mathbf{h}^T \boldsymbol{\tau}} = \mathbf{h}^T \hat{\boldsymbol{\tau}}$. When $r(\mathbf{C}_d) < p$, there are many possible solutions $\hat{\boldsymbol{\tau}}$, which are of the form $\hat{\boldsymbol{\tau}} = \mathbf{C}_d^{-} \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L) \mathbf{y}$. It can be shown that if \mathbf{h} , called a *coefficient vector*, is in the row space of \mathbf{C}_d , denoted $\mathcal{R}(\mathbf{C}_d)$, then the best linear unbiased estimator for $\mathbf{h}^T \boldsymbol{\tau}$ is $\widehat{\mathbf{h}^T \boldsymbol{\tau}}$, which is invariant to the chosen solution $\hat{\boldsymbol{\tau}}$. So long as $\mathbf{h} \in \mathcal{E}$, the variance of the estimator $\widehat{\mathbf{h}^T \boldsymbol{\tau}}$ under design d is

$$\text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = \mathbf{h}^T \mathbf{C}_d^{-} \mathbf{h} ,$$

which is also invariant to choice of solution $\hat{\boldsymbol{\tau}}$. Due to this invariance, we call such $\mathbf{h}^T \boldsymbol{\tau}$ *estimable functions* and so refer to $\mathcal{R}(\mathbf{C}_d)$ as the *estimation space*, \mathcal{E} .

Any estimation space is clearly a subset of \mathbb{R}^p , with equality when the assumed model allows us to estimate all linear combinations of $\boldsymbol{\tau}$. Let \mathcal{E}^* be the estimation space with maximal dimension, which equals the row space across all possible information matrices having maximal rank, say $k^* \leq p$. Typically only those designs having information matrices with this estimation space are considered, thus allowing the experimenter to estimate the largest possible number of estimable functions. However, further assumptions that reduce the set of estimable functions of interest may be made (see Chapter 5). For example, while for factorial experiments (see Section 1.2.2) there exist designs that can estimate all the factorial effects, we may often assume that high-order interactions are insignificant. If this assumption is justified, we should only consider designs that can estimate all and only the potentially significant estimable functions, lest we waste valuable resources. Hence, for the desired estimation space, $\mathcal{E} \subset \mathcal{E}^*$, which has some dimension $k \leq k^* \leq p$, we only consider those designs that are capable of estimating all and only functions generated by \mathcal{E} . That is, we only consider a design, d , if it has an information matrix where $\mathcal{R}(\mathbf{C}_d) = \mathcal{E}$, and so $r(\mathbf{C}_d) = k$. Denote the class of designs meeting this estimation space requirement by \mathcal{D} .

Let \mathbf{P}_τ be the orthogonal projector onto \mathcal{E} so that $\mathbf{P}_\tau \mathbf{C}_d = \mathbf{C}_d \mathbf{P}_\tau = \mathbf{C}_d$ (as $\mathcal{R}(\mathbf{C}_d) = \mathcal{E}$ for all $d \in \mathcal{D}$) and let $\mathbf{F}_1 = (\mathbf{f}_{11} | \dots | \mathbf{f}_{1k})$ be any $p \times k$ matrix satisfying $\mathbf{F}_1^T \mathbf{F}_1 = \mathbf{I}_k$ and $\mathbf{P}_\tau = \mathbf{F}_1 \mathbf{F}_1^T$. Obviously $(\mathbf{I} - \mathbf{P}_\tau)$ is the orthogonal projector onto the space orthogonal to \mathcal{E} . Let $\mathbf{F}_0 = (\mathbf{f}_{01} | \dots | \mathbf{f}_{0,p-k})$ be any $p \times (p - k)$ matrix satisfying $\mathbf{F}_0^T \mathbf{F}_0 = \mathbf{I}_{p-k}$ and $(\mathbf{I} - \mathbf{P}_\tau) = \mathbf{F}_0 \mathbf{F}_0^T$. Then $\mathbf{F}_1^T \mathbf{F}_0 = 0$ and $\mathbf{C}_d \mathbf{F}_0 = 0$ so that the vectors in \mathbf{F}_0 are a set of orthonormal eigenvectors with zero eigenvalues for all \mathbf{C}_d .

To demonstrate the above terminology, we turn to the commonly employed block design ANOVA model with p treatments and h blocks each of size $k \leq p$. If y_{uj} is the response for

the u -th unit in block j , then the statistical model relating y_{uj} to the treatment and block effects may be written in scalar form is

$$y_{uj} = \mu + \tau_{d[u,j]} + \beta_j + e_{uj} ,$$

where $d[u, j]$ denotes the treatment assigned to unit u in block j . In matrix form, $\boldsymbol{\beta}$ will have $h + 1$ parameters corresponding to the h block parameters β_1, \dots, β_h and the intercept μ . This model is often used in practice because it is justifiable by randomization alone in special cases, as shown by Hinkelmann and Kempthorne (2005) and Caliński and Kageyama (2000).

The maximal estimation space, \mathcal{E}^* , for $\boldsymbol{\tau}$ under this model contains all coefficient vectors satisfying $\mathbf{h}^T \mathbf{1} = 0$ and has dimension $p - 1$. Such estimable functions are called *contrasts* and such block designs are called *connected*. Connected designs allow estimation of all pairwise treatment comparisons, which is often the goal for comparative experiments. This estimation space has projector $\mathbf{P}_{\boldsymbol{\tau}} = \mathbf{I} - \frac{1}{p} \mathbf{J}$ where \mathbf{J} is the $p \times p$ matrix of all 1's. Any set of $v - 1$ orthonormal contrasts may be used to make up the matrix \mathbf{F}_1 , and $\mathbf{F}_0 = \frac{1}{\sqrt{p}} \mathbf{1}$. The general form of information matrices for $\boldsymbol{\tau}$ for such designs is

$$\mathbf{C}_d = \mathbf{R}_d - \frac{1}{k} \mathbf{N}_d \mathbf{N}_d^T$$

where, if r_{di} is the replication of treatment i under d , $\mathbf{R}_d = \text{Diag}(r_{d1}, \dots, r_{dp})$ and \mathbf{N}_d is the $p \times h$ treatment/block incidence matrix. That is, the element in the i -th row and j -th column of \mathbf{N}_d is non-zero if and only if treatment i appears in block j , otherwise it is 0.

1.1.1 The Moore-Penrose inverse and square-root decompositions

One generalized inverse for \mathbf{C}_d is the Moore-Penrose inverse which has appealing mathematical properties that will be used throughout this dissertation. We define such generalized inverses here only for symmetric, nonnegative definite (nnd) matrices.

Definition 1.1.1. *Let \mathbf{C} be a $p \times p$ symmetric, nnd matrix having $k \leq p$ eigenvalues, $\lambda_1, \dots, \lambda_k$. If the spectral decomposition of $\mathbf{C} = \mathbf{E}\mathbf{D}_\lambda\mathbf{E}^T$, where $\mathbf{D}_\lambda = \text{Diag}(\lambda_1, \dots, \lambda_k)$, then the Moore-Penrose inverse (or MP inverse) of \mathbf{C} , denoted \mathbf{C}^+ , is*

$$\mathbf{C}^+ = \mathbf{E}\mathbf{D}_\lambda^{-1}\mathbf{E}^T.$$

If \mathbf{C} has full rank then $\mathbf{C}^+ = \mathbf{C}^{-1}$ and we elect to use the latter notation. For information matrices, note that $\mathbf{C}_d^+\mathbf{C}_d = \mathbf{C}_d\mathbf{C}_d^+ = \mathbf{F}_1\mathbf{F}_1^T = \mathbf{P}_\tau$.

A symmetric, nnd matrix, \mathbf{C} , has a square-root decomposition $\mathbf{C} = \mathbf{C}^{1/2}\mathbf{C}^{1/2}$ for some $\mathbf{C}^{1/2}$, called a square-root matrix of \mathbf{C} . Letting $\mathbf{D}_\lambda^{1/2} = \text{Diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k})$, it is easily shown that $\mathbf{E}\mathbf{D}_\lambda^{1/2}\mathbf{E}^T$ is a square-root matrix of \mathbf{C} . Throughout this dissertation we assume all square-root matrices are constructed in this way. We choose to denote the MP inverse of a square root matrix $\mathbf{C}^{1/2}$ as $\mathbf{C}^{+1/2} = \mathbf{E}\mathbf{D}_\lambda^{-1/2}\mathbf{E}^T$ rather than $\mathbf{C}^{1/2+}$. If \mathbf{C} has full rank, we use the conventional $\mathbf{C}^{-1/2}$ instead of $\mathbf{C}^{+1/2}$. Finally, it is easily shown that $\mathbf{C}^+ = \mathbf{C}^{+1/2}\mathbf{C}^{+1/2}$.

1.2 Standard optimality

The “best” experimental design will maximize information that can be obtained on estimable functions of $\boldsymbol{\tau}$, the target of inference. Every \mathbf{C}_d is symmetric and nonnegative definite (nnd) so that it has k positive eigenvalues which may be ordered as $0 < z_{d1} \leq z_{d2} \leq \dots \leq z_{dk}$. It can

be shown that the variance of any normalized estimable function is a convex combination of the inverse eigenvalues of \mathbf{C}_d , hence we call $\frac{1}{z_{di}}$ the canonical variances for design d . It follows that a design whose \mathbf{C}_d has relatively small positive inverse eigenvalues will minimize the variance of estimable functions in some overall sense, and hence a design's overall effectiveness can be evaluated through a measure of the size of those eigenvalues. The three commonly used optimality criteria defined next are just such measures; there are many others.

Definition 1.2.1. *A design $d^* \in \mathcal{D}$ is E -optimal if it minimizes the maximal canonical variance, that is, if $\frac{1}{z_{d^*1}} = \min_{d \in \mathcal{D}} \frac{1}{z_{d1}}$, or $z_{d^*1} = \max_{d \in \mathcal{D}} z_{d1}$.*

Definition 1.2.2. *A design $d^* \in \mathcal{D}$ is D -optimal if it minimizes the product of the canonical variances, that is, if $\prod_{i=1}^k \frac{1}{z_{d^*i}} = \min_{d \in \mathcal{D}} \prod_{i=1}^k \frac{1}{z_{di}}$.*

Definition 1.2.3. *A design $d^* \in \mathcal{D}$ is A -optimal if it minimizes the average canonical variance, that is, if $\frac{1}{k} \sum_{i=1}^k \frac{1}{z_{d^*i}} = \frac{1}{k} \min_{d \in \mathcal{D}} \sum_{i=1}^k \frac{1}{z_{di}}$.*

The E -criterion is straightforward in its interpretation: it seeks to minimize the largest possible variance across all normalized estimable functions. Assuming $\mathbf{e} \sim N(\mathbf{0}_N, \mathbf{I}_N)$, the D -criterion aims to minimize the volume of the confidence ellipsoid of any k orthonormal estimable functions. Finally, the A -criterion measures the average variance of any k orthonormal estimable functions. In the case of the one-way ANOVA model where all treatment contrasts are estimable, it is also proportional to the average variance of all $\binom{p}{2}$ pairwise treatment contrasts. Morgan and Stallings (2013) give a necessary and sufficient condition for the average variance of a set of estimable functions to be proportional to the A -value for any design, showing that an A -optimal design accomplishes much more than may be initially apparent.

With many optimality criteria to choose from, a natural question to ask is whether there exist designs that are simultaneously optimal for multiple criteria, not just one criterion.

Kiefer (1975) approached this question within a broad class of criteria possessing three simple mathematical properties; his original framework is slightly generalized here. Given estimation space \mathcal{E} , let $\mathfrak{C}_{\mathcal{E}}$ denote the class of symmetric, nnd matrices with row space \mathcal{E} . Note that $\mathfrak{C}_{\mathcal{E}}$ contains all information matrices for designs in \mathcal{D} but is generally a much larger class. Let \mathbf{C}, \mathbf{C}_1 , and \mathbf{C}_2 be any matrices in $\mathfrak{C}_{\mathcal{E}}$ and let $\Pi_{\mathcal{E}}$ be the class of permutation matrices such that $\mathcal{R}(\Pi\mathbf{C}\Pi^T) = \mathcal{E}$. The criteria in Definition 1.2.4 may be used with many classes of statistical models, including the general linear model (1.1).

Definition 1.2.4. *The function $\Phi : \mathfrak{C}_{\mathcal{E}} \rightarrow \mathbb{R}$ is a conventional (or standard) optimality criterion if*

(i) *Φ is permutation invariant with respect to $\Pi_{\mathcal{E}}$: $\Phi(\Pi\mathbf{C}\Pi^T) = \Phi(\mathbf{C})$ for any permutation matrix, $\Pi \in \Pi_{\mathcal{E}}$,*

(ii) *Φ respects the nnd ordering: $\mathbf{C}_1 - \mathbf{C}_2$ is nnd $\Rightarrow \Phi(\mathbf{C}_1) \leq \Phi(\mathbf{C}_2)$,*

(iii) *Φ is convex: $\Phi(\alpha\mathbf{C}_1 + (1 - \alpha)\mathbf{C}_2) \leq \alpha\Phi(\mathbf{C}_1) + (1 - \alpha)\Phi(\mathbf{C}_2)$, $0 \leq \alpha \leq 1$.*

Although Φ has $\mathfrak{C}_{\mathcal{E}}$ as its domain of definition, we may assume that Φ is defined for all symmetric, nnd matrices having any row space of dimension k . For example, for $p > 0$, the family of eigenvalue-based criteria

$$\Phi_p(\mathbf{C}) = \left[k^{-1} \sum_{i=1}^k z_{di}^{-p} \right]^{1/p}$$

is comprised of standard optimality criteria that are clearly well-defined for all such matrices. The E -, D -, and A -criterion are special cases of such criteria, corresponding to $p \rightarrow \infty$, $p \rightarrow 0$, and $p = 1$, respectively.

Properties (i) and (iii) of Definition 1.2.4, permutation invariance and convexity, allow *matrix averaging* to establish a lower bound for a given $\Phi(\mathbf{C}_d)$. Define the averaged form

of an information matrix as $\bar{\mathbf{C}}_d = \sum_{\mathbf{\Pi} \in \Pi_\varepsilon} \mathbf{\Pi} \mathbf{C}_d \mathbf{\Pi}^T / |\Pi_\varepsilon|$. Restricting our attention to those $\mathbf{\Pi} \in \Pi_\varepsilon$ is necessary to guarantee that $\bar{\mathbf{C}}_d \in \mathfrak{C}_\varepsilon$. It then follows that

$$\Phi(\mathbf{C}_d) = \sum_{\mathbf{\Pi} \in \Pi_\varepsilon} \Phi(\mathbf{\Pi} \mathbf{C}_d \mathbf{\Pi}^T) / |\Pi_\varepsilon| \geq \Phi \left(\sum_{\mathbf{\Pi} \in \Pi_\varepsilon} \mathbf{\Pi} \mathbf{C}_d \mathbf{\Pi}^T / |\Pi_\varepsilon| \right) = \Phi(\bar{\mathbf{C}}_d) . \quad (1.2)$$

A given d may or may not achieve the lower bound $\Phi(\bar{\mathbf{C}}_d)$, which we emphasize applies only to d . That is, even if d does achieve the lower bound this in no way implies that d is optimal. There may exist another design, \tilde{d} , such that $\Phi(\mathbf{C}_{\tilde{d}}) < \Phi(\bar{\mathbf{C}}_d)$.

Consider the case where Π_ε contains all permutation matrices. A design d is said to be *symmetric* if it has an information matrix of the form $\mathbf{C}_d = a\mathbf{I} + b\mathbf{J}$. Such matrices, which include $\bar{\mathbf{C}}_d$, are said to be *completely symmetric* and are invariant to all permutation matrices, and so cannot be improved by matrix averaging. Kiefer (1975) proved that when working with the estimation space having projector $\mathbf{P}_\tau = \mathbf{I} - \frac{1}{p}\mathbf{J}$, which has the Π_ε considered here, if there exists a design, d^* , having a completely symmetric information matrix \mathbf{C}_{d^*} with maximal trace across all $d \in \mathcal{D}$, then that design is optimal *for all* Φ meeting the conditions of Definition 1.2.4. He called designs that are optimal with respect to all standard criteria *universally optimal*, presumably because the class of optimality criteria is quite broad. Note these are sufficient conditions (although the maximal trace condition is necessary for a universally optimal design), that is, there may exist a universally optimal design that is not symmetric. Kiefer's result is most useful when one already has a design in mind and wants to show that design is universally optimal.

The property of permutation invariance implicitly states that standard criteria do not advocate greater precision in estimating some estimable functions over others; it is the mathematical expression of equal interest in all estimable functions. This is not a desirable property if we are more interested in estimating some functions more precisely than others, which

has been mentioned by Glonek and Solomon (2004) and Gupta (2006). As will be seen, a technical foundation of this dissertation is the relaxation of permutation invariance in a way that is statistically meaningful for a design problem at hand.

It should be emphasized that the criteria considered throughout this dissertation are measures of variances alone. The statistical models are assumed to be correct (usually this is not in question) and thus there is no bias in the analysis. The “universal” part of Kiefer’s nomenclature is in reference to the standard criteria Φ , and not to all possible analysis goals. See Chapters 6 and 10 in Atkinson, Donev, and Tobias (2007) for a survey of other popular design criteria that focus estimation on a specific set of estimable functions or seek to minimize the average prediction variance across some regression range.

We now look at specific experimental scenarios where efficient estimation of a selected set of estimable functions, potentially at the expense of others, may be preferred. These will serve as examples throughout this dissertation.

1.2.1 Treatment-versus-control experiments

Under the framework of a comparative experiment with p treatments, let τ_0 be a control treatment and $\tau_1, \dots, \tau_{p-1}$ be test treatments. Suppose we are primarily interested in efficient estimation of the elementary treatment-versus-control (TvC) contrasts, $\tau_0 - \tau_i$, $i = 1, \dots, p-1$. If we could show that some standard optimality criterion was equivalent to measuring the variances for this set of contrasts alone then we would pursue optimality with respect to that criterion, but this is not the case. Instead, define $\mathbf{H} = (\mathbf{1}_{p-1} | -\mathbf{I}_{p-1})$, the $(p-1) \times p$ matrix where each row of $\mathbf{H}\boldsymbol{\tau}$ is an elementary TvC contrast. The covariance matrix for these contrasts for a design d is easily shown to be $\mathbf{H}\mathbf{C}_d^{-}\mathbf{H}^T$, and we call $(\mathbf{H}\mathbf{C}_d^{-}\mathbf{H}^T)^{-1}$ the TvC information matrix. Applying standard optimality criteria to the TvC information matrix,

we may find non-symmetric designs that are optimal; see Majumdar and Notz (1983), Cheng, Majumdar, Stufken, and Ture (1988), Jacroux (1993), Hedayat and Yang (2005), Yang and Stufken (2008), Hedayat, Jacroux, and Majumdar (1988), and Majumdar (1996) for optimal plans for TvC experiments.

One should take care in choosing a criterion for design comparison and be sure to understand the implications of what it measures, which is the underlying theme of this dissertation. For example, the E -criterion seeks to minimize the maximal variance across all orthonormal contrasts. This eigenvalue criterion can be applied to the TvC information matrix since it is an eigenvalue optimization problem, but this will not necessarily minimize the maximal variance of the elementary TvC contrasts, which is a more reasonable goal.

1.2.2 Factorial experiments

In many experiments, treatments are comprised of a combination of *factors* each having a certain number of settings, or *levels*. Much of the following notation is based on that from Hedayat, Sloane, and Stufken (1999). Let F_1, F_2, \dots, F_m denote the m factors of interest and let s_1, s_2, \dots, s_m be their respective number of levels, so the total number of treatment combinations is $\prod_{l=1}^m s_l$. We code the levels of factor F_l as $0, 1, \dots, s_l - 1$. A specific treatment combination is indexed using the m -tuples (j_1, j_2, \dots, j_m) , where $0 \leq j_l \leq s_l - 1, l = 1, \dots, m$. For convenience, we will often write the m -tuple as $j_1 j_2 \dots j_m$, which should not be confused with their product. In the factorial literature when $s_l = 2$ for all l , typically the number of $j_l = 1$ in the treatment index is called the weight of the treatment combination. This conflicts with our forthcoming use of the word, so we refer to this as the *heft* of the treatment combination.

Definition 1.2.5. *For a factorial experiment with m factors each with 2 levels, the heft of a treatment combination with index $j_1 \dots j_m$ is $\sum_{l=1}^m j_l$. That is, the heft is the number of*

$j_l = 1$.

Let $\tau_{j_1 j_2 \dots j_m}$ denote the mean response for treatment combination $j_1 j_2 \dots j_m$, and let $\boldsymbol{\tau}$ be the vector of $\tau_{j_1 j_2 \dots j_m}$ written in standard order, or *reverse lexicographic* (revlex) ordering, defined next for $s_l = 2$ for all l . The ordinary lexicographic (lex) ordering of treatment means in $\boldsymbol{\tau}$ for $m \geq 1$ factors is defined as a $2^m \times m$ array, denoted \mathbf{A}_m^* , and is constructed recursively as follows. For $m = 1$, $\mathbf{A}_1^* = (0, 1)^T$, which implies $\boldsymbol{\tau}^T = (\tau_0, \tau_1)$. For $m = 2$, the lex order is the ordering of the rows in \mathbf{A}_2^* given by

$$\mathbf{A}_2^* = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0}_2 & \mathbf{A}_1^* \\ \mathbf{1}_2 & \mathbf{A}_1^* \end{pmatrix}$$

so the ordering of these 2^2 treatment means will be $\boldsymbol{\tau}^T = (\tau_{00}, \tau_{01}, \tau_{10}, \tau_{11})$. The above is the first step in a recursion that gives, for any $m \geq 2$, the lex ordering as the rows of \mathbf{A}_m^* found by

$$\mathbf{A}_m^* = \begin{pmatrix} \mathbf{0}_{2^{m-1}} & \mathbf{A}_{m-1}^* \\ \mathbf{1}_{2^{m-1}} & \mathbf{A}_{m-1}^* \end{pmatrix}.$$

The *revlex* ordering is found in the same way, except that the pasting of columns shown above is done in the reverse order. So the revlex ordering for m -vectors on 2 symbols is given by the ordering of the rows of the $2^m \times m$ array, \mathbf{A}_m , where

$$\mathbf{A}_m = \begin{pmatrix} \mathbf{A}_{m-1} & \mathbf{0}_{2^{m-1}} \\ \mathbf{A}_{m-1} & \mathbf{1}_{2^{m-1}} \end{pmatrix}$$

with $\mathbf{A}_1 = \mathbf{A}_1^*$. Examples of the revlex order for $m = 2$ and $m = 3$ factors are:

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

When we have a factorial treatment structure, interest typically lies in treatment contrasts that measure the change in response with respect to changes in the levels of a subset of the factors. Those contrasts can be specified by adapting the approach taken for optimization of TvC contrasts via a properly defined transformation matrix \mathbf{H} . The commonly employed set of contrasts for factorial experiments are called *orthogonal factorial effects*, comprised of *main effects* and *interactions*. A *simple effect* of factor F_l is a contrast of treatment means with indices that differ only in index j_l . That is, it is a comparison of the treatment means for different levels of F_l but with the levels of the other $m - 1$ factors fixed. Main effect and interaction contrasts, which are linear combinations of simple effect contrasts, are well-known to all who work with factorial models and so will not be re-defined here; see for example Hedayat et al. (1999) or Hinkelmann and Kempthorne (2008).

The factorial contrasts for simple, main, and interaction effects in general depend on the levels of the factors involved. We now focus on the case of m factors each having $s_l = 2$ levels, so 2^m total treatment combinations. In this case there is a unique (up to sign change) main effect contrast for each factor, and likewise for the interaction of any g factors ($g = 2, 3, \dots, m$).

Moreover, these unique contrasts are mutually orthogonal and each involves all treatment combinations. It is also common to normalize such contrasts so that the coefficient vectors all have a length of 1. As there is a total of 2^m of these orthonormal contrasts, they provide an alternative parameterization of $\boldsymbol{\tau}$. Henceforth we refer to this orthogonal parameterization as OP and to these orthogonal factorial effects as OP effects.

Writing $\boldsymbol{\tau}$ in revlex order based on the treatment indices allows construction of the \mathbf{H} transformation matrix for the OP effects as a series of Kronecker products. For m factors each with 2 levels, define \mathbf{H}_{OP} to be the matrix whose rows consist of the coefficient vectors for the OP effects along with the normalized vector $\frac{1}{\sqrt{2^m}}\mathbf{1}_{2^m}$. Then we may write \mathbf{H}_{OP} as

$$\mathbf{H}_{OP} = \mathbf{Z}_m \otimes \mathbf{Z}_{m-1} \otimes \cdots \otimes \mathbf{Z}_1, \quad \mathbf{Z}_l = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ for } l = 1, 2, \dots, m, \quad (1.3)$$

where \otimes indicates the Kronecker product. The ordering of the \mathbf{Z}_l in the Kronecker product in (1.3) allows us to index the OP effects in $\mathbf{H}_{OP}\boldsymbol{\tau}$ using the revlex ordering. To demonstrate, let $\boldsymbol{\theta} = \mathbf{H}_{OP}\boldsymbol{\tau}$, a $2^m \times 1$ vector. Then the OP intercept, which we may denote as $\theta_{00\dots 0}$ appears first in $\boldsymbol{\theta}$, as does $\tau_{00\dots 0}$ in $\boldsymbol{\tau}$. The OP main effect for F_1 , which we may denote as $\theta_{10\dots 0}$, appears second in $\boldsymbol{\theta}$, as does $\tau_{10\dots 0}$ in $\boldsymbol{\tau}$.

Invoking the *effect hierarchy principle*, higher-order interactions are often assumed to be negligible. That is, we assume prior to any data collection that some of the factorial effects $\mathbf{H}_{OP}\boldsymbol{\tau}$ equal 0, or are sufficiently close to 0 to be of no concern. This can be a useful assumption, since estimation of all 2^m factorial effects and the intercept would require at least 2^m runs, a number which can get large very quickly with increasing m . If higher-order effects are negligible, then a *fractional factorial* experiment may be more appropriate, where fractional means we need only to estimate a fraction of the total factorial effects, and fewer runs are required. The literature on fractional factorial designs is rich, and we point the

uninformed reader to the texts Box and Draper (1987), Hedayat et al. (1999), Hinkelmann and Kempthorne (2008), Hinkelmann and Kempthorne (2005), and Myers, Montgomery, and Anderson-Cook (2009) which introduce both applications and theory of factorial experiments.

When there are no blocks to consider, it is well-known that an orthogonal array (OA) of appropriate strength is universally optimal for the estimation of the factorial effects included in the model. Existence of these designs requires, at the very least, that the number of EUs is a multiple of 4. Chakravarti (1956) generalized the structure of OA's with the concept of *balanced arrays* (although he called them partially balanced arrays at the time) which exist for all run sizes. While the concept generalizes to factors with more than two levels, we provide a definition for this specific case.

Definition 1.2.6. *Let \mathbf{V} be an $N \times m$ array where each column has symbols 0 and 1, and let \mathbf{V}_t be any $N \times t$ subarray of \mathbf{V} . By Definition 1.2.5, each row of \mathbf{V}_t has some heft between 0 and t . Suppose that for each $g = 0, \dots, t$, every possible row of heft g appears n_g times in \mathbf{V}_t . If the numbers n_g do not depend on the particular subarray \mathbf{V}_t , then we say that \mathbf{V} is a *balanced array (BA) of strength t with index set (n_0, n_1, \dots, n_t) .**

If a strength t BA has an index set where $n_0 = n_1 = \dots = n_t$ then \mathbf{V} is a strength t OA. It is straightforward to show that a strength t BA is also a strength $t - 1$ BA. Much of the theory on the optimality and construction of balanced arrays for fractional factorials, may be found in Chopra (1967), Srivastava (1961, 1965, 1972), and Srivastava and Chopra (1971a,b). Many of these results have been summarized in Rafter and Seiden (1974). These papers refer to fractional factorials derived from balanced arrays of appropriate strength as *balanced designs* since the covariance matrix for the factorial effects is invariant to factor permutations. Much of the work by Srivastava and Chopra establishes optimal designs within the class of balanced designs. Even though balanced arrays can sometimes be efficient with respect to all designs,

there are cases when they are very inefficient so that optimization within a class of strength t balanced arrays may yield poor overall designs.

1.2.3 Baseline parameterization of factorials

Mukerjee and Tang (2012) focus on a different set of factorial contrasts, corresponding to the *baseline parametrization* for factorial treatment effects. These contrasts, referred to as *baseline effects* or *BP effects*, define main effects and interactions relative to the baseline treatment effect $\tau_{00\dots 0}$. A major class of applications for BP effects is to the analysis of cDNA microarray experiments; see Mukerjee and Tang (2012) for other applications.

The baseline main effect for F_1 is defined to be

$$\theta_{10\dots 0} = \tau_{10\dots 0} - \tau_{00\dots 0} ,$$

which is the simple effect of F_1 with all other factors set at baseline. The BP 2 factor interaction (2-f.i.) between factors F_1 and F_2 is

$$\theta_{110\dots 0} = \tau_{110\dots 0} - \tau_{100\dots 0} - \tau_{010\dots 0} + \tau_{00\dots 0} .$$

Our use of the parameters $\theta_{j_1\dots j_m}$ in this section should not be confused with the parameters shown in the previous section after (1.3). Note that the $j_1 j_2$ components of the mean responses have the same form as an OP 2-f.i. for a 2^2 factorial. Proceeding in this way, define a BP g -f.i., $g = 2, 3, \dots, m$, as a non-normalized, OP g -f.i. for just those g factors, while holding the remaining $m - g$ factors at their baseline level (the OP g -f.i. averages over the levels of the remaining $m - g$ factors). It follows that the BP m -f.i. is equal to its non-normalized OP counterpart, since there are no factors left to hold at their baseline level.

The issue of normalizing BP effects will be taken up later.

A Kronecker product representation for the BP transformation matrix, say \mathbf{H}_{BP} , exists that is similar to that for \mathbf{H}_{OP} . Just as is done for $\boldsymbol{\tau}$, the BP parameters in $\boldsymbol{\theta}$ are indexed by m -tuples of 0's and 1's. The heft of a BP parameter index indicates the order of the BP interaction and those $j_l = 1$ correspond to the factors involved in that interaction. The BP intercept has heft 0, a BP main effect has heft 1, and all BP interactions have heft greater than or equal to 2. Equation (1) of Mukerjee and Tang (2012) says that any treatment mean can be written as the sum of $\theta_{00\dots 0}$ and all those BP parameters whose indices have 1's only in a subset of the positions having 1's in the subscript vector of that treatment mean. This is of sufficient usefulness that we state it, along with two related results, as a lemma.

Lemma 1.2.1. *The following relationships hold among the treatment means $\tau_{j_1 j_2 \dots j_m}$ and the BP parameters $\theta_{j_1^* j_2^* \dots j_m^*}$.*

- (i) *Let J be the set of positions in the index $j_1 j_2 \dots j_m$ of $\tau_{j_1 j_2 \dots j_m}$ containing non-zeros. Then $\tau_{j_1 j_2 \dots j_m}$ is the sum of exactly those $\theta_{j_1^* j_2^* \dots j_m^*}$'s whose indices have non-zeros in any subset of the positions J .*
- (ii) *Let the BP parameter $\theta_{j_1^* j_2^* \dots j_m^*}$ have heft h (that is, its index contains exactly h ones). Then there are exactly 2^{m-h} treatment means $\tau_{j_1 j_2 \dots j_m}$ that are a function of $\theta_{j_1^* j_2^* \dots j_m^*}$.*
- (iii) *Let $\tau_{j_1 j_2 \dots j_m}$ have heft h . Then for each $j = 0, 1, \dots, m-h$, there are exactly $\binom{m-h}{j}$ BP parameters $\theta_{j_1^* j_2^* \dots j_m^*}$ of heft $h+j$ that are a function of $\tau_{j_1 j_2 \dots j_m}$.*

Inverting the relationship in Lemma 1.2.1(i) gives the required BP transformation matrix, \mathbf{H}_{BP} .

Lemma 1.2.2. *For a 2^m factorial treatment structure under baseline parameterization, $\boldsymbol{\tau} =$*

$\mathbf{H}_{BP}^{-1}\boldsymbol{\theta}$, where

$$\mathbf{H}_{BP}^{-1} = \mathbf{Z}_m^{-1} \otimes \mathbf{Z}_{m-1}^{-1} \otimes \cdots \otimes \mathbf{Z}_1^{-1}, \quad \mathbf{Z}_l^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad l = 1, 2, \dots, m,$$

and where $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ have revlex ordering. Furthermore, $\mathbf{H}_{BP} = \mathbf{Z}_m \otimes \mathbf{Z}_{m-1} \otimes \cdots \otimes \mathbf{Z}_1$ for

$$\mathbf{Z}_l = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad l = 1, 2, \dots, m.$$

Proof. Write both $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ so that the indices are in revlex order. The result is clear for $m = 1$. For $m = 2$ we have

$$\begin{pmatrix} \tau_{00} \\ \tau_{10} \\ \tau_{01} \\ \tau_{11} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_{00} \\ \theta_{10} \\ \theta_{01} \\ \theta_{11} \end{pmatrix}$$

and clearly the inverse transformation matrix is the claimed Kronecker product. Now assume the result holds for $m = t$ (with parameters in revlex order) and denote its inverse transformation matrix by $\mathbf{H}_{BP_t}^{-1}$. Add a new factor and assign it to position $t + 1$ in the parameter index. There are now 2^{t+1} parameters, and when placed in revlex order, the first 2^t have factor $t + 1$ at level 0, and the remaining 2^t have factor $t + 1$ at level 1.

The inverse transformation matrix $\mathbf{H}_{BP_{t+1}}^{-1}$ is defined to be $\mathbf{Z}_{t+1}^{-1} \otimes \mathbf{H}_{BP_t}^{-1}$ and

$$\mathbf{Z}_{t+1}^{-1} \otimes \mathbf{H}_{BP_t}^{-1} = \begin{pmatrix} \mathbf{H}_{BP_t}^{-1} & 0 \\ \mathbf{H}_{BP_t}^{-1} & \mathbf{H}_{BP_t}^{-1} \end{pmatrix}.$$

By the induction hypotheses, the first 2^t rows of the displayed matrix produce the treatment means having factor $t+1$ at level 0. The other rows must be shown to produce the treatment means with factor $t+1$ at level 1. Select any one of the rows and consider the index vector \mathbf{s} (say) of the corresponding treatment mean. We know from the induction hypothesis that the first 2^t elements of the row are adding in exactly those θ 's having 1's only in positions corresponding to 1's in the first t positions of \mathbf{s} (all have 0 in position $t+1$). The remaining 2^t elements of that row are adding θ 's with exactly the same subscript vectors except they also all have 1 in position t . By Lemma 1.2.1, this is the selected treatment mean. \square

Again, the ordering of the \mathbf{Z}_l in the Kronecker product in Lemma 1.2.2 allows us to index the BP effects using the revlex order as was shown for the OP effects after (1.3). It is evident from Lemma 1.2.2 that \mathbf{H}_{BP} and \mathbf{H}_{BP}^{-1} , both being Kronecker products of m lower triangular matrices, have zeros in exactly the same positions. Combining this observation with Lemma 1.2.1(i) gives the following corollary.

Corollary 1.2.1. *Let the BP effect $\theta_{j_1^* j_2^* \dots j_m^*}$ have heft h . Then it is a linear combination of exactly those $\tau_{j_1 j_2 \dots j_m}$ whose subscript vectors have non-zeros in any subset of the positions J^* . Moreover, the coefficients in this linear combination are all 1's and -1 's.*

The following terminology will be useful in proofs throughout the dissertation. We say a BP effect with index $j_1 \dots j_m$ involves factor l if $j_l = 1$. If h is the heft of index $j_1 \dots j_m$, clearly the effect involves h factors. For an effect, define its *factor set* to be the set of factors involved in that effect. Hence the factor set has h elements. The *overlap set* between BP effects is the intersection of their factor sets. If the overlap set is nonempty, we say those effects *overlap*. We say two effects *overlap at factor l* if the overlap set includes factor l . Finally, two BP effects are said to *overlap by n factors* if the cardinality of the overlap set is n . The above definitions may also be used for factorial effects in the OP setup. For large m , it is more convenient to reference each BP effect by its factor set, lest we work with

cumbersome m -tuples. For example, for any m the BP main effect for F_1 could be indexed as either $\theta_{10\dots 0}$ or θ_1 , and the BP 2-f.i. for F_1 and F_2 could be indexed as $\theta_{110\dots 0}$ or θ_{12} . This more compact indexing was used in Mukerjee and Tang (2012).

It is interesting to note that overlapping defines an association scheme on BP effects. Such an association scheme may be called a multi-dimensional partially balanced association scheme, as described in Bose and Srivastava (1964).

Mukerjee and Tang (2012) found that if no BP 2-f.i.'s or higher are present, then the BP main effects are directly proportional to the OP main effects, so that orthogonal arrays are universally optimal designs for estimation of the BP main effects. Most of their work then looked at minimizing bias in the estimates if interactions were truly present, much like was done in Deng and Tang (1999) for OP effects. Once interactions are included in the estimation, we no longer have proportionality and the optimal designs for some criterion are likely different from OA's. At the end of their paper, they conjecture that certain balanced arrays may be A -optimal when 2-f.i.'s are included in the model, which we verify in Section 5.2.

Approximate A -optimal designs for $\mathbf{H}\boldsymbol{\tau}$

TvC and factorial experiments are specific applications of a general estimation problem for a set of estimable functions that span the estimation space. Rather than focusing optimality criteria on efficient estimation of all estimable functions, we instead want to find a design that best estimates $\mathbf{H}\boldsymbol{\tau}$, where $\mathbf{H}\boldsymbol{\tau}$ is some subset of estimable functions. It is required that $r(\mathbf{H}) = h \leq k$ and $\mathcal{R}(\mathbf{H})$ is a subset of the maximal estimation space. Just like for TvC, the covariance matrix for these contrasts is $\mathbf{H}\mathbf{C}_d^-\mathbf{H}^T$ and the A -criterion measures the trace of this matrix. Kao, Yates, Lewis, and Dean (1995) derived a general result for A -optimal block designs for estimating contrasts $\mathbf{H}\boldsymbol{\tau}$. Rather than searching for \mathbf{C}_d corresponding to

exact designs, they searched for an optimal *approximate information matrix*. Let \mathcal{D}_H be the class of designs such that \mathbf{H} is estimable, i.e. $d \in \mathcal{D}_H \Rightarrow \mathbf{H}\mathbf{C}_d^-\mathbf{C}_d = \mathbf{H}$. Note that if $h < k$ then \mathcal{D}_H is a larger class than the class of connected designs, hence disconnected designs may become A -optimal.

Define $C_H = \{\mathbf{C}_d : d \in \mathcal{D}_H\}$ and let c_{\max} denote the maximal trace of those $\mathbf{C}_d \in C_H$. The authors call \mathbf{M} an approximate information matrix if it lies in the set

$$\begin{aligned} \mathcal{M} = \{ \mathbf{M} : \mathbf{M} \text{ is a } p \times p \text{ symmetric, nnd matrix, with } \mathcal{R}(\mathbf{M}) \supseteq \mathcal{R}(\mathbf{H}), \\ \mathbf{M}\mathbf{1}_p = \mathbf{0}_p, \text{ and } \text{tr}(\mathbf{M}) \leq c_{\max} \} . \end{aligned} \quad (1.4)$$

The conditions $\mathcal{R}(\mathbf{M}) \supseteq \mathcal{R}(\mathbf{H})$ and $\mathbf{M}\mathbf{1}_p = \mathbf{0}_p$ imply $h \leq r(\mathbf{M}) \leq p-1$. Optimizing over this larger set of matrices (which contains C_H) allowed the authors to derive a form of an approximate information matrix to be A -optimal for $\mathbf{H}\boldsymbol{\tau}$.

Theorem 1.2.1. *Let \mathcal{M} be defined as in (1.4) and let $\mathbf{H}\boldsymbol{\tau}$ be a vector of contrasts where $r(\mathbf{H}) = h \leq p-1$. Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be a set of orthonormal eigenvectors of $\mathbf{H}^T\mathbf{H}$ and let $\theta_1, \dots, \theta_p$ be the corresponding eigenvalues. Then $\text{tr}(\mathbf{H}\mathbf{M}^{*-}\mathbf{H}^T) = \min\{\text{tr}(\mathbf{H}\mathbf{M}^-\mathbf{H}^T) : \mathbf{M} \in \mathcal{M}\}$ if and only if $\mathbf{M}^* = \alpha \sum_{i=1}^p \sqrt{\theta_i} \mathbf{u}_i \mathbf{u}_i^T$, where α is a constant such that $\text{tr}(\mathbf{M}^*) = c_{\max}$.*

It is unlikely that \mathbf{M}^* is in C_H , but Theorem 1.2.1 can be used to find a lower bound for the A -value, and we may assess efficiency of designs in C_H with respect to this bound. As approximate information matrices are much more flexible than \mathbf{C}_d in C_H , this bound may at times be unrealistic. Note well that the eigenspaces of \mathbf{M}^* and $\mathbf{H}^T\mathbf{H}$ are equivalent, implying that if a $\mathbf{C}_d \in C_H$ also has eigenspaces in common with $\mathbf{H}^T\mathbf{H}$ it may be efficient, although its trace must be large as well to guarantee this. Kao et al. (1995) refer to information matrices with some subset of eigenspaces in common with $\mathbf{H}^T\mathbf{H}$ as *aligned*

designs.

1.3 General weighted optimality

This dissertation introduces a flexible class of design criteria that both extend and unify standard optimality criteria by allowing an experimenter to assign relative importance to all estimable functions through a concise weighting system. Research on weighted optimality had been sparse leading up to the paper by Morgan and Wang (2010) and primarily focused on weighting of the A -criterion, see Gupta, Ramana, and Parsad (1999) and Shirakura and Tong (1996). In their paper, Morgan and Wang (2010) create “weighted variances” by multiplying the variance of an estimable function by a weight determined by a quadratic form of a positive definite, diagonal weight matrix \mathbf{W} . They then define the corresponding weighted optimality criteria as functions that seek to find designs that minimize these weighted variances in some overall sense. Their weight matrix is used to assign large weight to functions of greater importance and the weighted optimality criteria focus attention on designs that efficiently estimate these functions. Their weighted optimality criteria, which are constructed by applying standard optimality criteria to weighted information matrices, are shown to be summary measures of weighted variances.

The weighted optimality approach taken by Morgan and Wang (2010) was primarily explored in the context of comparative experiments with p treatments and connected designs. In this context, the weight w_i in the diagonal \mathbf{W} , while corresponding to a treatment effect τ_i , is not assigning weight to that effect since τ_i is not estimable (the model is overparameterized). Instead, the weight w_i is distributed to the contrasts involving the treatment τ_i . Hence larger w_i puts greater emphasis on contrasts involving τ_i and a weighted optimal design will typically estimate these contrasts with smaller variance than a universally optimal design

would. Indeed, the weighted optimal design will likely accomplish this by assigning more replicates to treatment i than a universally optimal design would.

1.4 Outline

The main aim of this dissertation is to generalize weighted optimality as developed by Morgan and Wang (2010), allowing all symmetric, positive definite matrices as candidates for weight matrices. The dissertation is organized as follows. Chapter 2 provides the preliminary results that generalize the theoretical foundation of weighted optimality to all symmetric, positive definite weight matrices. The chapter also introduces the concept of estimation equivalence, in which many weight matrices are shown to assign the same weights to every estimable function. Having the fundamental theory established, Chapter 3 applies the theory to show that eigenvalue-based optimization of information matrices for a set of estimable functions, like that shown in Section 1.2.1, is equivalent to a weighted optimality criteria. We then present a class of weight matrices that are able to directly assign weights to a set of estimable functions and a weighted criterion is shown to measure the average weighted variance of a large number of estimable functions with any chosen weights. Due to the complexity involved in finding analytical solutions to the design problem, Chapter 4 looks at implementing weighted optimality in SAS PROC OPTEX and an important theorem is presented that brings a new perspective to adjusted information. Finally, Chapter 5 details a linear model theory for reduced estimation spaces which we use to find A -optimal plans to estimate baseline factorial effects including up to 2-f.i.'s. We conclude the dissertation with a discussion of future work and open problems.

Chapter 2

Preliminary Results

This chapter motivates and develops the theory of general weighted optimality, beginning with the introduction of fundamental concepts like weighted variances and weighted information matrices. Establishment of the relationship between these two concepts is crucial to creating a statistically meaningful set of optimality criteria. Eigenvalue-based weighted optimality criteria are investigated first, and then shown to exist in a larger class of weighted optimality criteria. Conditions are given for any two weight matrices to assign the same weight (up to proportionality) to all estimable functions, a property that is used to establish which permutations the general weighted optimality criteria are invariant to. In seeking designs that are optimal under all weighted optimality criteria for some weight matrix \mathbf{W} , we conclude the chapter with a discussion on weight balanced designs and matrix averaging.

2.1 Weight matrices and weighted variances

Suppose that we wish to assign a weight $w > 0$ to some estimable function $\mathbf{h}^T \boldsymbol{\tau}$ in order to reflect its importance relative to other estimable functions. For this weight to impact

variance-based design criteria with a useful statistical interpretation, it is combined with the variance of the estimator $\widehat{\mathbf{h}^T \boldsymbol{\tau}}$ to create a “weighted variance”: $w \times \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$. We then seek to create a broad class of weighted optimality criteria that measure a design’s ability to minimize these weighted variances in some overall sense, implying that larger weight should be applied to those functions of greater interest.

In order to assign weights to all estimable functions, we will need a concise, yet flexible, weighting system. Take any $p \times p$ symmetric, positive definite matrix \mathbf{W} that also satisfies $\mathbf{1}^T \mathbf{W} \mathbf{1} = 1$. Such a matrix is called a *weight matrix*. The restriction on $\mathbf{1}^T \mathbf{W} \mathbf{1}$ is a simple scaling condition which, while unnecessary to the overall theory, imposes no loss of generality and is invoked for convenience. If $\mathbf{1}^T \mathbf{W} \mathbf{1} \neq 1$ then we need only multiply \mathbf{W} by $[\mathbf{1}^T \mathbf{W} \mathbf{1}]^{-1}$ to achieve this. Given \mathbf{W} , weights are assigned to variances in the following way:

Definition 2.1.1. *The weighted variance for the estimable function $\mathbf{h}^T \boldsymbol{\tau}$ for design d is*

$$\text{Var}_{dW}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = [\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}). \quad (2.1)$$

We call $[\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1}$ the weight assigned to $\mathbf{h}^T \boldsymbol{\tau}$.

It is obvious from (2.1) that any ordering of designs based on weighted variances is invariant to the choice of positive scaling of \mathbf{W} . Let $\mathbf{h}^* = \mathbf{h} / \sqrt{\mathbf{h}^T \mathbf{h}}$ be the normalized version of \mathbf{h} . Then we can decompose the weight assigned to $\mathbf{h}^T \boldsymbol{\tau}$ as

$$[\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} = [\mathbf{h}^{*T} \mathbf{W}^{-1} \mathbf{h}^*]^{-1} \times \frac{1}{\mathbf{h}^T \mathbf{h}}. \quad (2.2)$$

Hence assigned weights in general are made up of a normalization constant for \mathbf{h} and the weight assigned to the normalized version of \mathbf{h} . Even though weights depend on the length of the coefficient vector, weighted variances do not.

Lemma 2.1.1. *The weighted variance for $\mathbf{h}^T \boldsymbol{\tau}$ is invariant to the length of \mathbf{h} .*

Proof. Take any contrast $\mathbf{h}^T \boldsymbol{\tau}$ and define $\mathbf{h}^* = \mathbf{h} / \sqrt{\mathbf{h}^T \mathbf{h}}$ so that $\mathbf{h}^{*T} \mathbf{h}^* = 1$ and $\mathbf{h}^{*T} \boldsymbol{\tau}$ is a normalized estimable function. The weighted variance for $\mathbf{h}^{*T} \boldsymbol{\tau}$ is

$$\begin{aligned} \text{Var}_{dW}(\widehat{\mathbf{h}^{*T} \boldsymbol{\tau}}) &= [\mathbf{h}^{*T} \mathbf{W}^{-1} \mathbf{h}^*]^{-1} \text{Var}_d(\widehat{\mathbf{h}^{*T} \boldsymbol{\tau}}) \\ &= \mathbf{h}^T \mathbf{h} [\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} \frac{1}{\mathbf{h}^T \mathbf{h}} \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) \\ &= [\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = \text{Var}_{dW}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}), \end{aligned}$$

the weighted variance for the non-normalized estimable function. □

In light of Lemma 2.1.1, it is clear that the assigned weights and weighted variance may always be viewed in terms of the normalized estimable function. Further discussion of normalization may be found in Section 3.2.4.

Requiring \mathbf{W} to be positive definite, i.e. $[\mathbf{x}^T \mathbf{W}^{-1} \mathbf{x}]^{-1} > 0$ for all $p \times 1$ vectors $\mathbf{x} \neq \mathbf{0}$, is done primarily for mathematical convenience. Weights are only assigned to estimable functions so that we truly only need \mathbf{W} to satisfy $[\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} > 0$ for all $\mathbf{h} \in \mathcal{E}$. Consequences of this will be the main topic in Section 2.4.

2.2 Weighted information matrices

Assigning weights as defined in Definition 2.1.1 appears awkward at first, but serves to generate statistically meaningful weighted optimality criteria based on the standard criteria of Definition 1.2.4. Writing \mathbf{W} in spectral form as $\mathbf{W} = \mathbf{E} \mathbf{D}_\lambda \mathbf{E}^T$ allows us to define $\mathbf{W}^{-1/2}$ (see Section 1.1.1), which is used the following definition.

Definition 2.2.1. For a weight matrix \mathbf{W} and design $d \in \mathcal{D}$ with information matrix \mathbf{C}_d , the weighted information matrix for $\boldsymbol{\tau}$ is

$$\mathbf{C}_{dW} = \mathbf{W}^{-1/2} \mathbf{C}_d \mathbf{W}^{-1/2}.$$

Clearly all \mathbf{C}_{dW} are symmetric, nnd, and, because all competing designs are assumed to have information matrices with the same estimation space \mathcal{E} of dimension k , $r(\mathbf{C}_{dW}) = r(\mathbf{C}_d) = k$. The eigenvectors of \mathbf{C}_{dW} corresponding to the $p-k$ zero eigenvalues are not design dependent and are generated by $\mathbf{F}_{W0} = \mathbf{W}^{1/2} \mathbf{F}_0$, where \mathbf{F}_0 was introduced in Section 1.1. Hence $\mathcal{R}(\mathbf{C}_{dW})$ does not necessarily equal \mathcal{E} , although it is the same for all \mathbf{C}_{dW} . Write the spectral decomposition of \mathbf{C}_{dW} as

$$\mathbf{C}_{dW} = \sum_{i=1}^k \theta_{di} \mathbf{f}_{dWi} \mathbf{f}_{dWi}^T = \mathbf{F}_{dW1} \mathbf{D}_\theta \mathbf{F}_{dW1}^T \quad (2.3)$$

where the $p \times k$ matrix $\mathbf{F}_{dW1} = (\mathbf{f}_{dW1} | \mathbf{f}_{dW2} | \cdots | \mathbf{f}_{dWk})$ satisfies $\mathbf{F}_{dW1}^T \mathbf{F}_{dW1} = \mathbf{I}_k$, $\mathbf{F}_{dW1}^T \mathbf{F}_{W0} = 0$, and $0 < \theta_{d1} \leq \theta_{d2} \leq \cdots \leq \theta_{dk}$ are the positive eigenvalues of \mathbf{C}_{dW} . We do note that, in general, $\mathbf{F}_{W0}^T \mathbf{F}_{W0} \neq \mathbf{I}_{p-k}$. The next theorem relates the weighted variance of any estimable function to the eigenvalues of \mathbf{C}_{dW} .

Lemma 2.2.1. The weighted variance for any estimable function $\mathbf{h}^T \boldsymbol{\tau}$ is a convex combination of $\theta_{d1}^{-1}, \dots, \theta_{dk}^{-1}$.

Proof. Consider for $1 \leq i \leq k$ the vectors $\mathbf{l}_{di} = \mathbf{W}^{1/2} \mathbf{f}_{dWi}$. They satisfy $\mathbf{F}_0^T \mathbf{l}_{di} = \mathbf{F}_0^T \mathbf{W}^{1/2} \mathbf{f}_{dWi} = \mathbf{F}_{W0}^T \mathbf{f}_{dWi} = 0$, that is, they are estimable functions. Let $\mathbf{L}_d = \mathbf{W}^{1/2} \mathbf{F}_{dW1}$, which should not be confused with \mathbf{L} in (1.1), be the $p \times k$ matrix whose columns are $\mathbf{l}_{d1}, \dots, \mathbf{l}_{dk}$. The columns of \mathbf{L}_d are linearly independent (because the columns of \mathbf{F}_{dW1} are) and so are a basis for \mathcal{E} . This says that given any estimable function $\mathbf{h}^T \boldsymbol{\tau}$ there must be a $k \times 1$ vector \mathbf{q} such that

$\mathbf{h} = \mathbf{L}_d \mathbf{q}$. With this the variance for $\widehat{\mathbf{h}^T \boldsymbol{\tau}}$ may be written as

$$\text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = \mathbf{h}^T \mathbf{C}_d^- \mathbf{h} = \mathbf{q}^T \mathbf{L}_d^T \mathbf{C}_d^- \mathbf{L}_d \mathbf{q} = \mathbf{q}^T \mathbf{F}_{dW1}^T \mathbf{W}^{1/2} \mathbf{C}_d^- \mathbf{W}^{1/2} \mathbf{F}_{dW1} \mathbf{q} . \quad (2.4)$$

Since $\mathbf{W}^{-1/2} \mathbf{C}_d \mathbf{W}^{-1/2} \mathbf{f}_{dWi} = \theta_{di} \mathbf{f}_{dWi}$, the matrix $\mathbf{F}_{dW1}^T \mathbf{W}^{1/2} \mathbf{C}_d^- \mathbf{W}^{1/2} \mathbf{F}_{dW1}$ in the last quadratic form has (i, j) element

$$\begin{aligned} \mathbf{f}_{dWi}^T \mathbf{W}^{1/2} \mathbf{C}_d^- \mathbf{W}^{1/2} \mathbf{f}_{dWj} &= \frac{1}{\theta_{di} \theta_{dj}} [\mathbf{f}_{dWi}^T \mathbf{W}^{-1/2} \mathbf{C}_d \mathbf{W}^{-1/2}] \mathbf{W}^{1/2} \mathbf{C}_d^- \mathbf{W}^{1/2} [\mathbf{W}^{-1/2} \mathbf{C}_d \mathbf{W}^{-1/2} \mathbf{f}_{dWj}] \\ &= \frac{1}{\theta_{di} \theta_{dj}} \mathbf{f}_{dWi}^T \mathbf{W}^{-1/2} \mathbf{C}_d \mathbf{W}^{-1/2} \mathbf{f}_{dWj} = \begin{cases} 1/\theta_{di} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

It follows that

$$\text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = \sum_{i=1}^k \frac{q_i^2}{\theta_{di}},$$

where q_i is the i -th element of \mathbf{q} . Now $\mathbf{q}^T \mathbf{q} = \mathbf{q}^T \mathbf{F}_{dW1}^T \mathbf{W}^{1/2} \mathbf{W}^{-1} \mathbf{W}^{1/2} \mathbf{F}_{dW1} \mathbf{q} = \mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}$.

Putting this in (2.1) with the variance expression just derived gives

$$\text{Var}_{dW}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = [\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1} \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) = \frac{1}{\mathbf{q}^T \mathbf{q}} \sum_{i=1}^k \frac{q_i^2}{\theta_{di}} \quad (2.5)$$

which is a convex combination of the $1/\theta_{di}$, as claimed. \square

Since every weighted variance is a convex combination of the θ_{di}^{-1} , minimizing a relevant, summary function of the θ_{di}^{-1} is minimizing weighted variance in some overall sense. It shows that any optimality criterion, that in the unweighted case seeks to minimize variances as measured by a function of inverses of eigenvalues of \mathbf{C}_d , is minimizing weighted variance when used to evaluate the weighted information matrices \mathbf{C}_{dW} . The inverse eigenvalues θ_{di}^{-1} are *canonical weighted variances*, termed simply canonical variances in the unweighted

($\mathbf{W} = \frac{1}{p}\mathbf{I}$) case. If $\mathbf{W} \propto \mathbf{I}$ then $\mathbf{C}_{dW} \propto \mathbf{C}_d$ and the design ordering for any weighted criterion is equivalent to that for the standard criterion. Hence a standard optimality criterion is a special case of a weighted criterion.

We now investigate weighted analogues of standard eigenvalue criteria and discuss how they evaluate weighted variances. The following lemma, which is straightforward to show, is useful in providing meaningful interpretations of weighted eigenvalue criteria.

Lemma 2.2.2. *Let \mathbf{C}_d^- and \mathbf{C}_{dW}^- be arbitrary generalized inverses of \mathbf{C}_d and \mathbf{C}_{dW} , respectively. Then $\mathbf{W}^{-1/2}\mathbf{C}_{dW}^-\mathbf{W}^{-1/2}$ is a generalized inverse of \mathbf{C}_d , and $\mathbf{W}^{1/2}\mathbf{C}_d^-\mathbf{W}^{1/2}$ is a generalized inverse of \mathbf{C}_{dW} .*

2.3 Weighted eigenvalue criteria

We now define the weighted versions of the E -, A -, and D -criterion.

2.3.1 Weighted- E criterion

Definition 2.3.1. *The weighted E -value for design d , or E_{dW} , is the inverse of the smallest positive eigenvalue of the weighted information matrix \mathbf{C}_{dW} :*

$$E_{dW} = \frac{1}{\theta_{d1}} .$$

A design d^ is E_W -optimal in a design class \mathcal{D} if it minimizes the largest canonical weighted variance, that is, if $E_{d^*W} = \min_{d \in \mathcal{D}} E_{dW}$.*

Lemma 2.3.1 provides an interpretation for E_{dW} analogous to that for the unweighted E -criterion.

Lemma 2.3.1. *For a design d , E_{dW} is the largest weighted variance over all estimable functions.*

Proof. The maximal weighted variance is

$$\max_{\mathbf{h} \in \mathcal{E}} \left(\frac{\mathbf{h}^T \mathbf{C}_d^- \mathbf{h}}{\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}} \right) = \max_{\mathbf{h} \in \mathcal{E}} \left(\frac{\mathbf{h}^T \mathbf{W}^{-1/2} \mathbf{C}_{dW}^- \mathbf{W}^{-1/2} \mathbf{h}}{\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}} \right) = \max_{\mathbf{y}^T \mathbf{F}_{W0} = 0} \left(\frac{\mathbf{y}^T \mathbf{C}_{dW}^- \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right)$$

where $\mathbf{y} = \mathbf{W}^{-1/2} \mathbf{h}$ is an eigenvector for \mathbf{C}_{dW} having a positive eigenvalue as \mathbf{y} is orthogonal to \mathbf{F}_{W0} . It follows that the maximal weighted variance is equal to the largest eigenvalue of \mathbf{C}_{dW}^- , being $1/\theta_{d1}$. \square

2.3.2 Weighted- A criterion

Definition 2.3.2. *The weighted A -value for design d , or A_{dW} , is the average of the inverse of the positive eigenvalues of the weighted information matrix \mathbf{C}_{dW} :*

$$A_{dW} = \frac{1}{k} \sum_{i=1}^k \frac{1}{\theta_{di}} .$$

A design d^ is A_W -optimal in a design class \mathcal{D} if it minimizes the average (or equivalently the sum) of the canonical weighted variances, that is, if $A_{d^*W} = \min_{d \in \mathcal{D}} A_{dW}$.*

In providing additional interpretations for A_{dW} , we say two estimable functions $\mathbf{h}_i^T \boldsymbol{\tau}$ and $\mathbf{h}_j^T \boldsymbol{\tau}$ are *weighted orthogonal* if $\mathbf{h}_i^T \mathbf{W}^{-1} \mathbf{h}_j = 0$. A set of contrasts are mutually weighted orthogonal if any pair of contrasts in the set are weighted orthogonal.

Lemma 2.3.2. *For a design d , A_{dW} is proportional to the average weighted variance for any k weighted orthogonal contrasts.*

Proof. Let $\tilde{\mathbf{l}}_1, \tilde{\mathbf{l}}_2, \dots, \tilde{\mathbf{l}}_k$ be the coefficient vectors for any set of weighted orthogonal estimable functions with assigned weights $[\tilde{\mathbf{l}}_i^T \mathbf{W}^{-1} \tilde{\mathbf{l}}_i]^{-1} = w_i$. Further let \mathbf{D}_w be the $k \times k$ diagonal matrix of these weights. Writing $\tilde{\mathbf{L}}$ for the $p \times k$ matrix whose columns are the $\tilde{\mathbf{l}}_i$, then $\tilde{\mathbf{L}}^T \mathbf{W}^{-1} \tilde{\mathbf{L}} = \mathbf{D}_w^{-1} \Leftrightarrow \mathbf{D}_w^{1/2} \tilde{\mathbf{L}}^T \mathbf{W}^{-1} \tilde{\mathbf{L}} \mathbf{D}_w^{1/2} = \mathbf{I}_k$ which further shows that the columns of $\tilde{\mathbf{L}}$ are linearly independent. Consequently $\tilde{\mathbf{L}} \mathbf{D}_w^{1/2} = \mathbf{L}_d \mathbf{A}$ for $\mathbf{L}_d = \mathbf{W}^{1/2} \mathbf{F}_{dW1}$ and some nonsingular $\mathbf{A}_{k \times k}$. Since also $\mathbf{L}_d^T \mathbf{W}^{-1} \mathbf{L}_d = \mathbf{I}_k$, we have $\mathbf{D}_w^{1/2} \tilde{\mathbf{L}}^T \mathbf{W}^{-1} \tilde{\mathbf{L}} \mathbf{D}_w^{1/2} = \mathbf{I}_k \Leftrightarrow \mathbf{A}^T \mathbf{L}_d^T \mathbf{W}^{-1} \mathbf{L}_d \mathbf{A} = \mathbf{I}_k \Leftrightarrow \mathbf{A}^T \mathbf{A} = \mathbf{I}_k \Leftrightarrow \mathbf{A}$ is orthogonal. This implies that $\tilde{\mathbf{L}} \mathbf{D}_w \tilde{\mathbf{L}}^T = \mathbf{L}_d \mathbf{A} \mathbf{A}^T \mathbf{L}_d^T = \mathbf{L}_d \mathbf{L}_d^T$ so

$$\mathbf{W}^{-1/2} \tilde{\mathbf{L}} \mathbf{D}_w \tilde{\mathbf{L}}^T \mathbf{W}^{-1/2} = \mathbf{F}_{dW1} \mathbf{F}_{dW1}^T .$$

$$\begin{aligned} \sum_{i=1}^k w_i \text{Var}(\widehat{\tilde{\mathbf{l}}_i^T \tau}) &= \text{tr}(\mathbf{D}_w^{1/2} \tilde{\mathbf{L}}^T \mathbf{C}_d^- \tilde{\mathbf{L}} \mathbf{D}_w^{1/2}) = \text{tr}(\mathbf{D}_w^{1/2} \tilde{\mathbf{L}}^T \mathbf{W}^{-1/2} \mathbf{C}_{dW}^+ \mathbf{W}^{-1/2} \tilde{\mathbf{L}} \mathbf{D}_w^{1/2}) \\ &= \text{tr}(\mathbf{C}_{dW}^+ \mathbf{W}^{-1/2} \tilde{\mathbf{L}} \mathbf{D}_w \tilde{\mathbf{L}}^T \mathbf{W}^{-1/2}) \\ &= \text{tr}(\mathbf{C}_{dW}^+ \mathbf{F}_{dW1} \mathbf{F}_{dW1}^T) \\ &= \text{tr}(\mathbf{C}_{dW}^+) = \sum_{i=1}^k \frac{1}{\theta_{di}} = k A_{dW} . \end{aligned}$$

□

2.3.3 Weighted- D criterion

Definition 2.3.3. *The weighted D -value for design d , or D_{dW} , is the product of the inverse of the positive eigenvalues of the weighted information matrix \mathbf{C}_{dW} :*

$$D_{dW} = \prod_{i=1}^k \frac{1}{\theta_{di}} .$$

A design d^ is D_W -optimal in a design class \mathcal{D} if it minimizes the product of the canonical weighted variances, that is, if $D_{d^*W} = \min_{d \in \mathcal{D}} D_{dW}$.*

Interestingly, this weighted criterion does not respond to weights.

Lemma 2.3.3. *For any design $d \in \mathcal{D}$ and weight matrix \mathbf{W} , D_{dW} is proportional to D_d .*

Proof. Recall that \mathbf{P}_τ is the orthogonal projector onto \mathcal{E} . As is well-known, $\mathbf{B}\mathbf{B}^T$ and $\mathbf{B}^T\mathbf{B}$ have the same positive eigenvalues for any matrix \mathbf{B} , so the positive eigenvalues of \mathbf{C}_{dW} are equal to those of $\mathbf{C}_d^{1/2}\mathbf{W}^{-1}\mathbf{C}_d^{1/2} = \mathbf{C}_d^{1/2}\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau\mathbf{C}_d^{1/2}$. Note that $\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau$ is symmetric and nnd (with column space equal to \mathcal{E}), and so admits a square root decomposition $\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau = [\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2}$. Hence the positive eigenvalues of \mathbf{C}_{dW} are equal to those of the following matrices:

$$\mathbf{C}_d^{1/2}[\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2}[\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2}\mathbf{C}_d^{1/2} \Leftrightarrow [\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2}\mathbf{C}_d[\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2} \Leftrightarrow \mathbf{W}_*^{1/2}\mathbf{C}_d\mathbf{W}_*^{1/2}$$

where $\mathbf{W}_*^{1/2} = [\mathbf{P}_\tau\mathbf{W}^{-1}\mathbf{P}_\tau]^{1/2} + (\mathbf{I} - \mathbf{P}_\tau)$, a symmetric, positive definite matrix. Note that

$\mathcal{R}(\mathbf{W}_*^{1/2} \mathbf{C}_d \mathbf{W}_*^{1/2}) = \mathcal{E}$ so that we may finally write

$$\begin{aligned}
D_{dW} &= |\mathbf{W}_*^{1/2} \mathbf{C}_d \mathbf{W}_*^{1/2} + (\mathbf{I} - \mathbf{P}_\tau)|^{-1} \\
&\propto |\mathbf{W}_*^{1/2}| |\mathbf{W}_*^{1/2} \mathbf{C}_d \mathbf{W}_*^{1/2} + (\mathbf{I} - \mathbf{P}_\tau)|^{-1} |\mathbf{W}_*^{1/2}| \\
&= |\mathbf{C}_d + \mathbf{W}_*^{-1/2} (\mathbf{I} - \mathbf{P}_\tau) \mathbf{W}_*^{-1/2}|^{-1} \\
&= |\mathbf{C}_d + (\mathbf{I} - \mathbf{P}_\tau)|^{-1} = D_d .
\end{aligned}$$

□

Lemma 2.3.3 is related to the well-known fact that the D -criterion is invariant to full-rank, linear transformations. Proportionality to the standard D -criterion implies that the standard design ordering is equivalent to that for D_{dW} , which is important enough to state as a corollary.

Corollary 2.3.1. *If $D_{d_1} \leq D_{d_2}$, then $D_{d_1W} \leq D_{d_2W}$ for all weight matrices. Furthermore, the relative D_W -efficiencies equal those for the standard D -criterion.*

Lemma 2.3.3 and Corollary 2.3.1 are why there is no D -optimality work for design problems such as those discussed in sections 1.2.1 and 1.2.3. The D -criterion is unable to account for selected functions having greater importance and so is an inappropriate criterion to use when there is differential interest of estimable functions. The most enlightening part of the proof for Lemma 2.3.3 is that the eigenvalues of \mathbf{C}_{dW} for any weight matrix \mathbf{W} are equivalent to the eigenvalues of \mathbf{C}_{dW^*} where $\mathbf{W}_* = [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau + (\mathbf{I} - \mathbf{P}_\tau)]^{-1}$. This suggests that there may exist multiple weight matrices that assign the same weights to every estimable function. We explore the existence of non-unique weight matrices in Section 2.4.

2.4 Estimation equivalence

A statistically meaningful weighted criterion generates a design ordering based on weighted variances, comprised of variances $\text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$ multiplied by weights $[\mathbf{h}^T \mathbf{W}^{-1} \mathbf{h}]^{-1}$. The question taken up here is uniqueness, or lack thereof, of \mathbf{W} : can different weight matrices generate the same weights for all estimable functions? Consider, for example, comparative experiments where all treatment contrasts are estimable and $\mathbf{C}_d \mathbf{1} = \mathbf{0}$ for all information matrices. For a given \mathbf{W} form another weight matrix $\mathbf{W}_* = \mathbf{W} + q\mathbf{J}$, $q > 0$. This matrix is clearly not equal to \mathbf{W} , even if we do scale both matrices, but both of these weight matrices assign (up to a constant multiple) the same weights for all contrasts. Although this is a simple example, it demonstrates that a weight matrix can be contained in a family of weight matrices, each of which produces the same design ordering. This leads to what we call *estimation equivalence*.

Definition 2.4.1. *Two weight matrices \mathbf{W}_1 and \mathbf{W}_2 are estimation equivalent with respect to \mathcal{E} with projection matrix \mathbf{P}_τ if for some $q > 0$ and all $\mathbf{h} \in \mathcal{E}$*

$$\mathbf{h}^T \mathbf{W}_1^{-1} \mathbf{h} = q \mathbf{h}^T \mathbf{W}_2^{-1} \mathbf{h} .$$

If \mathbf{W}_1 and \mathbf{W}_2 are estimation equivalent, then clearly their corresponding design criteria, assuming they measure weighted variances, will produce the same design orderings. Hence working with weighted information matrices based on \mathbf{W}_1 would be equivalent to working with those based on \mathbf{W}_2 . The following are some results that can be used to characterize the family of estimation equivalent weight matrices for a given \mathbf{W} , starting with a general matrix result.

Lemma 2.4.1. *Let \mathbf{A} be any $p \times p$ symmetric matrix and let \mathbf{P}_V be the orthogonal projector onto a subspace V of \mathbb{R}^p . Then $\mathbf{h}^T \mathbf{A} \mathbf{h} = 0$ for all $\mathbf{h} \in V$ if and only if $\mathbf{P}_V \mathbf{A} \mathbf{P}_V = \mathbf{0}$.*

Furthermore, \mathbf{A} must be of the form

$$\mathbf{A} = (\mathbf{I} - \mathbf{P}_V)\mathbf{A}_1(\mathbf{I} - \mathbf{P}_V) + \mathbf{A}_2(\mathbf{I} - \mathbf{P}_V) + (\mathbf{I} - \mathbf{P}_V)\mathbf{A}_2^T \quad (2.6)$$

where \mathbf{A}_1 is symmetric.

Proof. Assume $\mathbf{P}_V\mathbf{A}\mathbf{P}_V = 0$ and rewrite \mathbf{A} as

$$\begin{aligned} \mathbf{A} &= (\mathbf{P}_V + (\mathbf{I} - \mathbf{P}_V))\mathbf{A}(\mathbf{P}_V + (\mathbf{I} - \mathbf{P}_V)) \\ &= \mathbf{P}_V\mathbf{A}\mathbf{P}_V + (\mathbf{I} - \mathbf{P}_V)\mathbf{A}(\mathbf{I} - \mathbf{P}_V) + \mathbf{P}_V\mathbf{A}(\mathbf{I} - \mathbf{P}_V) + (\mathbf{I} - \mathbf{P}_V)\mathbf{A}\mathbf{P}_V \\ &= (\mathbf{I} - \mathbf{P}_V)\mathbf{A}(\mathbf{I} - \mathbf{P}_V) + \mathbf{P}_V\mathbf{A}(\mathbf{I} - \mathbf{P}_V) + (\mathbf{I} - \mathbf{P}_V)\mathbf{A}\mathbf{P}_V, \end{aligned}$$

so $\mathbf{h}^T\mathbf{A}\mathbf{h} = 0$ for all $\mathbf{h} \in V$. Here $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_2 = \mathbf{P}_V\mathbf{A}$.

Now assume $\mathbf{h}^T\mathbf{A}\mathbf{h} = 0$ for all $\mathbf{h} \in V$. Take any row, say \mathbf{h}_1^T , of \mathbf{P}_V and we have $\mathbf{h}_1^T\mathbf{A}\mathbf{h}_1 = 0$ so the diagonal elements of $\mathbf{P}_V\mathbf{A}\mathbf{P}_V$ are 0. Now take any two rows, \mathbf{h}_1^T and \mathbf{h}_2^T , of \mathbf{P}_V and let $\mathbf{c}_0^T = \mathbf{c}_1^T + \mathbf{c}_2^T$. Then $\mathbf{c}_0 \in V$ and

$$\begin{aligned} 0 &= \mathbf{c}_0^T\mathbf{A}\mathbf{c}_0 \\ &= \mathbf{c}_1^T\mathbf{A}\mathbf{c}_1 + \mathbf{c}_2^T\mathbf{A}\mathbf{c}_2 + 2\mathbf{c}_1^T\mathbf{A}\mathbf{c}_2 \\ &= 2\mathbf{c}_1^T\mathbf{A}\mathbf{c}_2 \\ &= 2\mathbf{c}_2^T\mathbf{A}\mathbf{c}_1, \end{aligned}$$

since \mathbf{A} is symmetric. Hence $\mathbf{c}_1^T\mathbf{A}\mathbf{c}_2 = \mathbf{c}_2^T\mathbf{A}\mathbf{c}_1 = 0$ and all off-diagonals of $\mathbf{P}_V\mathbf{A}\mathbf{P}_V$ are 0.

The form of \mathbf{A} follows from the previous argument. \square

The following corollary to Lemma 2.4.1 is proven by setting $\mathbf{A} = \mathbf{W}_1^{-1} - \frac{1}{q}\mathbf{W}_2^{-1}$ for two

estimation equivalent weight matrices and taking $V = \mathcal{E}$ so that $\mathbf{P}_V = \mathbf{P}_\tau$.

Corollary 2.4.1. *The weight matrices \mathbf{W}_1 and \mathbf{W}_2 are estimation equivalent if and only if $\mathbf{P}_\tau \mathbf{W}_1^{-1} \mathbf{P}_\tau = \frac{1}{q} \mathbf{P}_\tau \mathbf{W}_2^{-1} \mathbf{P}_\tau$ for some $q > 0$.*

Using this corollary and the required form of \mathbf{A} in Lemma 2.4.1, we have the following lemma characterizing all weight matrices that are estimation-equivalent to a given \mathbf{W} .

Lemma 2.4.2. *The weight matrices that are estimation equivalent to \mathbf{W} are all matrices with inverses*

$$\mathbf{W}_*^{-1} = q \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau + \mathbf{A} \quad (2.7)$$

for $q > 0$ and any \mathbf{A} of the form

$$\mathbf{A} = (\mathbf{I} - \mathbf{P}_\tau) \mathbf{A}_1 (\mathbf{I} - \mathbf{P}_\tau) + \mathbf{A}_2 (\mathbf{I} - \mathbf{P}_\tau) + (\mathbf{I} - \mathbf{P}_\tau) \mathbf{A}_2^T \quad (2.8)$$

where \mathbf{A}_1 is symmetric and

$$q \mathbf{c}^T \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau \mathbf{c} > -\mathbf{c}^T \mathbf{A} \mathbf{c}$$

for all $\mathbf{c} \neq \mathbf{0}$.

Proof. All such \mathbf{W}_* having inverses of form (2.7) are clearly symmetric and estimation equivalent to \mathbf{W} . Next we need to show that all \mathbf{W}_* are positive definite. That is, for any $p \times 1$ $\mathbf{c} \neq \mathbf{0}$ we must show $\mathbf{c}^T \mathbf{W}_*^{-1} \mathbf{c} > 0$. In general, the quadratic form may be written as

$$\mathbf{c}^T \mathbf{W}_*^{-1} \mathbf{c} = q \mathbf{c}^T \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau \mathbf{c} + \mathbf{c}^T \mathbf{A} \mathbf{c} .$$

The first summand is non-negative so to be positive definite

$$q\mathbf{c}^T \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau \mathbf{c} > -\mathbf{c}^T \mathbf{A} \mathbf{c} .$$

Now we want to show that all estimation equivalent matrices may be written in the same form as \mathbf{W}_*^{-1} . If \mathbf{W}_0 is an estimation equivalent weight matrix to \mathbf{W} then there is some $q > 0$ such that $\mathbf{P}_\tau \mathbf{W}_0^{-1} \mathbf{P}_\tau = q \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$. We may rewrite \mathbf{W}_0^{-1} as

$$\begin{aligned} \mathbf{W}_0^{-1} &= \mathbf{P}_\tau \mathbf{W}_0^{-1} \mathbf{P}_\tau + (\mathbf{I} - \mathbf{P}_\tau) \mathbf{W}_0^{-1} (\mathbf{I} - \mathbf{P}_\tau) + (\mathbf{I} - \mathbf{P}_\tau) \mathbf{W}_0^{-1} \mathbf{P}_\tau + \mathbf{P}_\tau \mathbf{W}_0^{-1} (\mathbf{I} - \mathbf{P}_\tau) \\ &= \mathbf{P}_\tau \mathbf{W}_0^{-1} \mathbf{P}_\tau + \mathbf{A}_* \\ &= q \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau + \mathbf{A}_* \end{aligned}$$

where \mathbf{A}_* is of form (2.8) with $\mathbf{A}_1 = \mathbf{W}_0^{-1}$, and $\mathbf{A}_2 = \mathbf{P}_\tau \mathbf{W}_0^{-1}$. □

One of the most useful applications of these results was seen in Lemma 2.3.3 and is of sufficient use to state as a corollary.

Corollary 2.4.2. *For a given weight matrix \mathbf{W} , the weight matrix, \mathbf{W}_* , having inverse and inverse square root matrix*

$$\mathbf{W}_*^{-1} = \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau + (\mathbf{I} - \mathbf{P}_\tau) \tag{2.9}$$

$$\mathbf{W}_*^{-1/2} = [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{1/2} + (\mathbf{I} - \mathbf{P}_\tau) , \tag{2.10}$$

is estimation equivalent to \mathbf{W} . Furthermore, weighted information matrices under \mathbf{W}_ have row space \mathcal{E} .*

As will be seen in the following section, estimation equivalence will often be relied on for mathematical results. Given a family of estimation equivalent weight matrices one should

choose a form that either makes analytical results straightforward or is clearer in how it assigns weight to functions. For example, a diagonal weight matrix is easy to construct and the resulting weights are straightforward. Some may even choose to pick a \mathbf{W}^{-1} rather than a \mathbf{W} , which is acceptable. One has the option to start with an easily constructed weight matrix and may then use the flexibility afforded by estimation equivalence to derive a weight matrix that may make analytical results more straightforward.

2.5 General weighted optimality criteria

Weighted eigenvalue criteria are a rich class of design measures that clearly focus attention on designs that minimize weighted variances in some overall sense. There are many other weighted criteria, other than those considered to now, that have practical relevance. For example, with comparative experiments and connected designs, the standard MV -criterion measures the maximal variance of all pairwise comparisons. The weighted analogue of this criterion is presented in Definition 2.5.1.

Definition 2.5.1. *The weighted MV -value for design d , or MV_{dW} , is the largest weighted variance among all pairwise comparisons:*

$$MV_{dW} = \max_{i \neq j} Var_{dW} \left(\widehat{\frac{\tau_i - \tau_j}{\sqrt{2}}} \right).$$

A design d^ is MV_W -optimal in a design class \mathcal{D} if it minimizes the largest weighted variance of all pairwise comparisons, that is, if $MV_{d^*W} = \min_{d \in \mathcal{D}} MV_{dW}$.*

The importance of the MV_W -criterion motivates the need to define a broad class of weighted optimality criteria in the same vein as Definition 1.2.4. In that definition, conventional optimality functions were required to be convex, to preserve nonnegative definite ordering,

and to be permutation invariant. These properties allowed matrix averaging to establish lower bounds on $\Phi(\mathbf{C}_d)$ and led to sufficient conditions for universally optimal designs. Recall that these criteria were also defined as functions of all symmetric, nnd matrices with the same row space (\mathfrak{E}), this being the class of all possible information matrices. Here we want to evaluate weighted information through \mathbf{C}_{dW} , so we generate weighted optimality criteria by applying standard optimality criteria to the \mathbf{C}_{dW} .

Definition 2.5.2. *For given weight matrix \mathbf{W} , the weighted version of a standard criteria Φ , denoted Φ_W , is the mapping $\Phi_W : \mathfrak{E} \rightarrow \mathbb{R}$ given by*

$$\Phi_W(\mathbf{C}) = \Phi(\mathbf{C}_W) , \quad (2.11)$$

for all $\mathbf{C} \in \mathfrak{E}$, where $\mathbf{C}_W = \mathbf{W}^{-1/2} \mathbf{C} \mathbf{W}^{-1/2}$.

We note that weighted optimality criteria are well-defined for any weight matrix \mathbf{W} , as all standard optimality criteria are well-defined for all symmetric, nnd matrices having any row space of dimension k . It also follows that all valid weighted criteria must further satisfy $\Phi_{W_1}(\mathbf{C}) \propto \Phi_{W_2}(\mathbf{C})$ for all $\mathbf{C} \in \mathfrak{E}$ when \mathbf{W}_1 and \mathbf{W}_2 are estimation equivalent. Furthermore, we emphasize that weighted optimality criteria are a function of $\mathbf{C} \in \mathfrak{E}$ not \mathbf{C}_W . That is, the design optimization problem is to find an information matrix, not a weighted information matrix, that minimizes all or some Φ_W . When looking for an optimal design, however, it may be worthwhile to think of the design optimization problem in terms of finding an optimal weighted information matrix, \mathbf{C}_{dW^*} , and then taking the transformation $\mathbf{C}_{d^*} = \mathbf{W}^{1/2} \mathbf{C}_{dW^*} \mathbf{W}^{1/2}$ to find the optimal form of the information matrix. Although, there is no guarantee that the resulting \mathbf{C}_{d^*} will correspond to an information matrix.

Theorem 2.5.1 details the properties of Φ_W induced from Φ in terms of comparing $\mathbf{C}, \mathbf{C}_1, \mathbf{C}_2 \in \mathfrak{E}$. The only modification to the three basic properties of standard criteria as given in Defi-

dition 1.2.4 is to the class of permutations preserving criteria values: $\Phi_W(\Pi C \Pi^T) = \Phi_W(C)$ for all Φ_W . This permutation class, denoted Π_W , is shown to be those permutations in Π_ε satisfying

$$\Pi P_\tau W^{-1} P_\tau \Pi^T = P_\tau W^{-1} P_\tau ,$$

which is closely related to the concept of estimation equivalence.

The proof of Theorem 2.5.1 requires the following result concerning permutations of a symmetric matrix. First, we provide an equivalent expression of a symmetric, nnd matrix \mathbf{A} related to its spectral decomposition. Suppose \mathbf{A} has $m \leq k$ distinct positive eigenvalues, denoted λ_j . The *eigenspace* corresponding to λ_j is the set of all vectors satisfying $\mathbf{A}\mathbf{e} = \lambda_j\mathbf{e}$. Let \mathbf{P}_j be the orthogonal projector onto this subspace, that is $\mathbf{P}_j\mathbf{e} = \mathbf{e}$ if and only if \mathbf{e} is in the eigenspace of λ_j and $\mathbf{P}\mathbf{e} = \mathbf{0}$ for \mathbf{e} in any other eigenspace. It is well-known that if the \mathbf{P}_j are chosen to be symmetric and idempotent then are unique. We may then write

$$\mathbf{A} = \sum_{j=1}^m \lambda_j \mathbf{P}_j .$$

Lemma 2.5.1. *Let \mathbf{A} be a symmetric, nnd matrix described above. Then $\Pi \mathbf{A} \Pi^T = \mathbf{A}$ if and only if $\Pi \mathbf{P}_j \Pi^T = \mathbf{P}_j$ for all orthogonal projectors onto the n eigenspaces.*

Proof. Assume $\Pi \mathbf{P}_j \Pi^T = \mathbf{P}_j$ for $j = 1, \dots, m$, so

$$\Pi \mathbf{A} \Pi^T = \sum_{j=1}^m \lambda_j \Pi \mathbf{P}_j \Pi^T = \sum_{j=1}^m \lambda_j \mathbf{P}_j = \mathbf{A} .$$

If \mathbf{e} is in the eigenspace of \mathbf{A} corresponding to λ_j then $\mathbf{A}\mathbf{e} = \lambda_j\mathbf{e}$ and clearly $\Pi \mathbf{A} \Pi^T(\Pi \mathbf{e}) = \lambda_j \Pi \mathbf{e}$. If $\Pi \mathbf{A} \Pi^T = \mathbf{A}$ then $\mathbf{A}(\Pi \mathbf{e}) = \lambda_j(\Pi \mathbf{e})$ so $\Pi \mathbf{e}$ must also lie in the eigenspace of \mathbf{A} corresponding to λ_j . For any j , we have a unique \mathbf{P}_j that projects onto the eigenspace corresponding to λ_j . If λ_j has multiplicity n_j then any set of n_j orthonormal vectors $\mathbf{e}_1, \dots, \mathbf{e}_{n_j}$

lying in the eigenspace of λ_j satisfy $\mathbf{P}_j = \sum_{l=1}^{n_j} \mathbf{e}_l \mathbf{e}_l^T$. Clearly $\mathbf{\Pi} \mathbf{e}_l$ and $\mathbf{\Pi} \mathbf{e}_{l'}$, $l \neq l'$, are also orthonormal and lie in the eigenspace of λ_j so $\mathbf{P}_j = \sum_{l=1}^{n_j} \mathbf{\Pi} \mathbf{e}_l \mathbf{e}_l^T \mathbf{\Pi}^T = \mathbf{\Pi} \left(\sum_{l=1}^{n_j} \mathbf{e}_l \mathbf{e}_l^T \right) \mathbf{\Pi}^T = \mathbf{\Pi} \mathbf{P}_j \mathbf{\Pi}^T$.

□

Theorem 2.5.1. *Let \mathbf{W} be a chosen weight matrix. For every \mathbf{C} , \mathbf{C}_1 , \mathbf{C}_2 in \mathfrak{C}_ε , any weighted criterion Φ_W under \mathbf{W} satisfies*

(i) Φ_W is invariant to all $\mathbf{\Pi} \in \Pi_W$: $\Phi_W(\mathbf{\Pi} \mathbf{C} \mathbf{\Pi}^T) = \Phi_W(\mathbf{C})$ for all $\mathbf{\Pi} \in \Pi_W$,

(ii) Φ_W preserves the nnd ordering: $\mathbf{C}_1 - \mathbf{C}_2$ is nnd $\Rightarrow \Phi_W(\mathbf{C}_1) \leq \Phi_W(\mathbf{C}_2)$,

(iii) Φ_W is convex: $\Phi_W(\alpha \mathbf{C}_1 + (1 - \alpha) \mathbf{C}_2) \leq \alpha \Phi_W(\mathbf{C}_1) + (1 - \alpha) \Phi_W(\mathbf{C}_2)$, $0 \leq \alpha \leq 1$.

Proof. A weighted criterion, Φ_W , for the chosen \mathbf{W} is generated by a standard optimality criterion, Φ , which is defined on, and properties (1)-(3) of Definition 1.2.4 hold for, all symmetric, nnd matrices having any row space of dimension k . We first show (ii) and (iii). If $\mathbf{C}_1 - \mathbf{C}_2$ is nnd then so is $\mathbf{C}_{1W} - \mathbf{C}_{2W}$ so $\Phi(\mathbf{C}_{1W}) \leq \Phi(\mathbf{C}_{2W})$, or $\Phi_W(\mathbf{C}_1) \leq \Phi_W(\mathbf{C}_2)$, proving (ii). Next, $\Phi_W(\alpha \mathbf{C}_1 + (1 - \alpha) \mathbf{C}_2) = \Phi(\alpha \mathbf{C}_{1W} + (1 - \alpha) \mathbf{C}_{2W}) \leq \alpha \Phi(\mathbf{C}_{1W}) + (1 - \alpha) \Phi(\mathbf{C}_{2W}) = \alpha \Phi_W(\mathbf{C}_1) + (1 - \alpha) \Phi_W(\mathbf{C}_2)$, proving (iii).

Let \mathbf{W}_* be the estimation equivalent weight matrix for \mathbf{W} as described in Corollary 2.4.2. These matrices assign identical weights to all estimable functions, so $\Phi_W(\mathbf{C}) = \Phi_{W_*}(\mathbf{C})$ for all $\mathbf{C} \in \mathfrak{C}_\varepsilon$. Clearly the permutation class $\Pi_W = \{\mathbf{\Pi} : \mathbf{\Pi} \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau \mathbf{\Pi}^T = \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau\}$ equals that for \mathbf{W}_* . Also, if $\mathbf{\Pi} \in \Pi_W$ then it is easily shown that $\mathbf{\Pi}^T \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau \mathbf{\Pi} = \mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ and because $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ and $[\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{1/2}$ share the same eigenspaces, Lemma 2.5.1 says

$\Pi^T [P_\tau W^{-1} P_\tau]^{1/2} \Pi = [P_\tau W^{-1} P_\tau]^{1/2}$. Hence if $\Pi \in \Pi_W$ then

$$\begin{aligned}
\Phi_W(\Pi C \Pi^T) &= \Phi_{W_*}(\Pi C \Pi^T) = \Phi([P_\tau W^{-1} P_\tau]^{1/2} \Pi C \Pi^T [P_\tau W^{-1} P_\tau]^{1/2}) \\
&= \Phi(\Pi^T [P_\tau W^{-1} P_\tau]^{1/2} \Pi C \Pi^T [P_\tau W^{-1} P_\tau]^{1/2} \Pi) \\
&= \Phi([P_\tau W^{-1} P_\tau]^{1/2} C [P_\tau W^{-1} P_\tau]^{1/2}) \\
&= \Phi_{W_*}(C) = \Phi_W(C) .
\end{aligned}$$

It remains to be shown that if $\Phi_W(\Pi C \Pi^T) = \Phi_W(C)$ for all $\Pi \in \Pi_\mathcal{E}$, $C \in \mathfrak{C}_\mathcal{E}$, and Φ_W under W , then $\Pi \in \Pi_W$. Assume on the contrary that the result holds for some Π in $\Pi_\mathcal{E}$ but not in Π_W , and consider $C^* = [P_\tau W^{-1} P_\tau]^+$, the MP inverse (see Defintion 1.1.1) of $P_\tau W^{-1} P_\tau$. It follows that

$$[P_\tau W^{-1} P_\tau]^{1/2} C^* [P_\tau W^{-1} P_\tau]^{1/2} = P_\tau ,$$

as the eigenspaces of $C^* = [P_\tau W^{-1} P_\tau]^+$ corresponding to its positive eigenvalues equal those of $[P_\tau W^{-1} P_\tau]^{1/2}$ and are orthogonal to $(I - P_\tau)$. Hence

$$\begin{aligned}
\Phi_W(C^*) &= \Phi_{W_*}(C^*) = \Phi([P_\tau W^{-1} P_\tau]^{1/2} C^* [P_\tau W^{-1} P_\tau]^{1/2}) \\
&= \Phi(P_\tau) .
\end{aligned}$$

The matrix P_τ is symmetric and idempotent so its k positive eigenvalues equal 1 and it follows that the A_W - and E_W -criterion values for C^* equal 1. By assumption, the A_W - and E_W -criterion values for $\Pi C^* \Pi^T$ must also equal 1, which happens if and only if all the positive eigenvalues of $[P_\tau W^{-1} P_\tau]^{1/2} \Pi C^* \Pi^T [P_\tau W^{-1} P_\tau]^{1/2}$ equal 1. The row space of $[P_\tau W^{-1} P_\tau]^{1/2} \Pi C^* \Pi^T [P_\tau W^{-1} P_\tau]^{1/2}$ equals \mathcal{E} , so the eigenvectors of this matrix corresponding to its positive eigenvalues make up an orthonormal basis for P_τ . Hence if all the

positive eigenvalues equal 1 the spectral decomposition of $[P_\tau W^{-1} P_\tau]^{1/2} \Pi C^* \Pi^T [P_\tau W^{-1} P_\tau]^{1/2}$ says

$$\begin{aligned} [P_\tau W^{-1} P_\tau]^{1/2} \Pi C^* \Pi^T [P_\tau W^{-1} P_\tau]^{1/2} &= P_\tau \\ \Rightarrow P_\tau \Pi C^* \Pi^T P_\tau &= [P_\tau W^{-1} P_\tau]^{+1/2} [P_\tau W^{-1} P_\tau]^{+1/2} = [P_\tau W^{-1} P_\tau]^+ = C^* . \end{aligned}$$

As $\Pi \in \Pi_\varepsilon$, $P_\tau \Pi C^* \Pi^T P_\tau = \Pi C^* \Pi^T$ so $\Pi C^* \Pi^T = C^*$ and $\Pi \notin \Pi_W \Rightarrow \Pi P_\tau W^{-1} P_\tau \Pi^T \neq P_\tau W^{-1} P_\tau$. By Lemma 2.5.1, this implies $\Pi C^* \Pi^T \neq C^*$, a contradiction. \square

As seen in the proof of Theorem 2.5.1, the permutation class Π_W is a minimal class of permutation matrices that all Φ_W are invariant to. We cannot rule out the possibility of a particular weighted criterion to be permutation invariant to more than those permutations in Π_W . A trivial example of such a criterion is $\Phi_W(\mathbf{C}) = 1$ for all \mathbf{C} . However, based on the proof, Π_W is likely complete for all eigenvalue-based criteria.

Depending on the estimation space and chosen weight matrix, the condition for Π to be in Π_W may be simplified to something more intuitive. For example, in regression models and the cell means model, $P_\tau = \mathbf{I}_p$ and the condition reduces to $\Pi W^{-1} \Pi^T = W^{-1}$. Morgan and Wang (2010), who worked with diagonal weight matrices and $P_\tau = \mathbf{I} - \frac{1}{p} \mathbf{J}$, claimed Π_W consists of those permutation matrices satisfying $\Pi \mathbf{w} = \mathbf{w}$, the vector of diagonal elements in the diagonal \mathbf{W} . While intuitive, they did not provide a necessary and sufficient condition like that in Theorem 2.5.1, which we provide in the following section.

2.5.1 Permutation invariance for connected designs

The goal of this section is to explore whether we can simplify the permutation invariance condition established in the previous section when $P_\tau = \mathbf{I} - \frac{1}{p} \mathbf{J}$, for instance with

connected designs for comparative experiments. Note that in this case, $\Pi_{\mathcal{E}}$ contains all permutation matrices, which is the class that Kiefer (1975) worked with. This matrix is completely symmetric and $\mathbf{J} = \Pi\mathbf{J} = \mathbf{J}\Pi$ which shows that if $\Pi\mathbf{W}^{-1}\Pi^T = \mathbf{W}^{-1}$ then $\Pi\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}\Pi^T = \Pi\Pi^T\mathbf{P}_{\tau}\Pi\mathbf{W}^{-1}\Pi^T\mathbf{P}_{\tau}\Pi\Pi^T = \mathbf{P}_{\tau}\Pi\mathbf{W}^{-1}\Pi^T\mathbf{P}_{\tau} = \mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}$, so such permutations are in Π_W . We now investigate whether $\Pi\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}\Pi^T = \mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}$ implies $\Pi\mathbf{W}^{-1}\Pi^T = \mathbf{W}^{-1}$. If $\Pi\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}\Pi^T = \mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}$ then, multiplying this out for $\mathbf{P}_{\tau} = \mathbf{I} - \frac{1}{p}\mathbf{J}$,

$$\begin{aligned} \Pi\mathbf{W}^{-1}\Pi^T - \frac{1}{p}[\Pi\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}\Pi^T] + \frac{1}{p^2}\Pi\mathbf{J}\mathbf{W}^{-1}\mathbf{J}\Pi^T \\ = \mathbf{W}^{-1} - \frac{1}{p}[\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}] + \frac{1}{p^2}\mathbf{J}\mathbf{W}^{-1}\mathbf{J} \\ \Leftrightarrow \Pi\mathbf{W}^{-1}\Pi^T - \frac{1}{p}[\Pi\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}\Pi^T] = \mathbf{W}^{-1} - \frac{1}{p}[\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}]. \end{aligned}$$

A sufficient condition for $\Pi\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}\Pi^T = \mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau} \Rightarrow \Pi\mathbf{W}^{-1}\Pi^T = \mathbf{W}^{-1}$ is if $\Pi\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}\Pi^T = \mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1}$.

Lemma 2.5.2. *For $\mathbf{P}_{\tau} = \mathbf{I} - \frac{1}{p}\mathbf{J}$ and weight matrix \mathbf{W} satisfying $\mathbf{W}\mathbf{1} = \lambda\mathbf{1}$, Π_W are those weight matrices satisfying $\Pi\mathbf{W}^{-1}\Pi^T = \mathbf{W}^{-1}$.*

Proof. Follows immediately from the fact that $\mathbf{W}^{-1}\mathbf{J} + \mathbf{J}\mathbf{W}^{-1} = \frac{2}{\lambda}\mathbf{J}$ which is invariant to all permutations. \square

Such weight matrices in Lemma 2.5.2 will be thoroughly discussed in chapter 3. Corollary 2.4.2 states that every weight matrix \mathbf{W} has an estimation equivalent weight matrix \mathbf{W}_* that satisfies the eigenvector condition in Lemma 2.5.2, so there exists an estimation equivalent weight matrix for which we can simplify the condition.

Let $\mathbf{W}^{-1} = (w_{ij})$, and denote $\mathbf{W}^{-1}\mathbf{1} = \mathbf{w}^{-1} = (w_{i.})$ and $w_{..} = \mathbf{1}^T\mathbf{W}^{-1}\mathbf{1}$. Similarly let $\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau} = (p_{ij})$ and $\mathbf{p}^* = (p_{ii})$, the $p \times 1$ vector of the diagonal elements of $\mathbf{P}_{\tau}\mathbf{W}^{-1}\mathbf{P}_{\tau}$.

It is easy to show that

$$p_{ii} = w_{ii} - \frac{2}{p}w_{i.} + \frac{w_{..}}{p^2} .$$

Now if $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ is invariant to a permutation $\mathbf{\Pi}$ then clearly $\mathbf{\Pi} \mathbf{p}^* = \mathbf{p}^*$. Hence

$$\mathbf{\Pi} \mathbf{p}^* = \mathbf{p}^* \Rightarrow \mathbf{\Pi} \left(\mathbf{p}^* - \frac{w_{..}}{p^2} \mathbf{1} \right) = \mathbf{p}^* - \frac{w_{..}}{p^2} \mathbf{1} ,$$

since $\mathbf{\Pi} \mathbf{1} = \mathbf{1}$ for any permutation matrix.

Lemma 2.5.3. *For $\mathbf{P}_\tau = \mathbf{I} - \frac{1}{p} \mathbf{J}$ and a diagonal weight matrix \mathbf{W} , Π_W is comprised of those weight matrices satisfying $\mathbf{\Pi} \mathbf{w} = \mathbf{w}$ where \mathbf{w} is the $p \times 1$ vector of diagonal elements.*

Proof. Because \mathbf{W}^{-1} is diagonal we have that $w_{ii} = w_{i.}$, which implies that the i th element of $\mathbf{p}^* - \frac{w_{..}}{p^2} \mathbf{1}$ is equal to $w_{ii} - \frac{2}{p}w_{i.} = (1 - \frac{2}{p})w_{i.}$ and so $\mathbf{p}^* - \frac{w_{..}}{p^2} \mathbf{1}$ is proportional to $\mathbf{w}^{-1} = \mathbf{W}^{-1} \mathbf{1}$.

Hence

$$\begin{aligned} \mathbf{\Pi} \mathbf{W}^{-1} \mathbf{J} + \mathbf{J} \mathbf{W}^{-1} \mathbf{\Pi}^T &= \mathbf{\Pi} \mathbf{w}^{-1} \mathbf{1}^T + \mathbf{1} \mathbf{w}^{-1'} \mathbf{\Pi}^T \\ &= \mathbf{w}^{-1} \mathbf{1}^T + \mathbf{1} \mathbf{w}^{-1'} \\ &= \mathbf{W}^{-1} \mathbf{J} + \mathbf{J} \mathbf{W}^{-1} . \end{aligned}$$

Clearly permutation invariance also holds for \mathbf{w} , which simply inverts the elements in \mathbf{w}^{-1} .

□

2.6 Matrix averaging, weight balance, and permutation invariance

A crucial piece to Kiefer's theorem of universal optimality was complete symmetry of \mathbf{C}_{d^*} , which resulted from matrix averaging with respect to all permutation matrices. As pointed out in the previous section, matrix averaging for a weighted optimality criterion is done with respect to Π_W , a potentially smaller subset of all possible permutations. An information matrix then need not be completely symmetric in order to reach its lower bound for some weighted optimality criterion attained by matrix averaging.

The positive eigenvalues for completely symmetric information matrices are all equal, so every normalized estimable function has equal variance. Such designs are thus called *variance balanced*. If a design, d^* , has a completely symmetric weighted information matrix then the positive eigenvalues of \mathbf{C}_{d^*W} are all equal and so every weighted variance, which is a convex combination of the inverse eigenvalues of \mathbf{C}_{d^*W} , must be the same and we call d^* *weight balanced*.

Definition 2.6.1. *A design d is said to be weight balanced for \mathbf{W} if $\text{Var}_{dW}(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$ does not depend on the $\mathbf{h} \in \mathcal{E}$ selected.*

The following lemma characterizes the form of all \mathbf{C}_d for weight balanced designs for a given \mathbf{W} .

Lemma 2.6.1. *A design d is weight balanced for \mathbf{W} if and only if, for some $\epsilon > 0$, \mathbf{C}_d may be written as*

$$\mathbf{C}_d = \epsilon \mathbf{L}_d \mathbf{D}_w \mathbf{L}_d^T ,$$

where columns of \mathbf{L}_d are k weighted orthogonal functions satisfying $\mathbf{L}_d^T \mathbf{W}^{-1} \mathbf{L}_d = \mathbf{D}_w^{-1}$, a $k \times k$ diagonal matrix whose diagonal elements are the inverse weights assigned to the

corresponding estimable functions.

Proof. Assume d is weight balanced for \mathbf{W} so all the eigenvalues of \mathbf{C}_{dW} equal (say) $\epsilon > 0$. It follows that the spectral decomposition for \mathbf{C}_{dW} is $\mathbf{C}_{dW} = \epsilon \mathbf{F}_{dW1} \mathbf{F}_{dW1}^T$ where \mathbf{F}_{dW1} has k orthonormal $p \times 1$ vectors that are also orthogonal to \mathbf{F}_{W0} . Hence $\mathbf{C}_d = \mathbf{W}^{1/2} \mathbf{C}_{dW} \mathbf{W}^{1/2} = \epsilon \mathbf{W}^{1/2} \mathbf{F}_{dW1} \mathbf{F}_{dW1}^T \mathbf{W}^{1/2} = \epsilon \mathbf{L}_d \mathbf{L}_d^T$. Now $\mathbf{L}_d^T \mathbf{F}_0 = \mathbf{F}_{dW1}^T \mathbf{W}^{1/2} \mathbf{F}_0 = \mathbf{F}_{dW1}^T \mathbf{F}_0$ and so the columns of \mathbf{L}_d make up k linearly independent estimable functions. The functions satisfy $\mathbf{L}_d^T \mathbf{W}^{-1} \mathbf{L}_d = \mathbf{F}_{dW1}^T \mathbf{F}_{dW1} = \mathbf{I}_k$ and hence are weighted orthogonal with assigned weight 1.

Now assume that $\mathbf{C}_d = \epsilon \mathbf{L}_d \mathbf{D}_w \mathbf{L}_d^T$ where $\mathbf{L}_d^T \mathbf{W}^{-1} \mathbf{L}_d = \mathbf{D}_w^{-1}$. It then follows that $\mathbf{C}_{dW} = \epsilon \mathbf{W}^{-1/2} \mathbf{L}_d \mathbf{D}_w \mathbf{L}_d^T \mathbf{W}^{-1/2}$ and $\mathbf{W}^{-1/2} \mathbf{L}_d \mathbf{D}_w^{1/2}$ make up a set of k linearly independent eigenvectors for \mathbf{C}_{dW} , each having positive eigenvalue ϵ . Hence all the positive eigenvalues of \mathbf{C}_{dW} are equal and so d is weight balanced. \square

The impact a weight balanced design has on the unweighted variances is best understood by looking at the ratio of any two weighted variances under that design, which equals 1 by definition. Let $w_1 = [\mathbf{h}_1^T \mathbf{W}^{-1} \mathbf{h}_1]^{-1}$ and $w_2 = [\mathbf{h}_2^T \mathbf{W}^{-1} \mathbf{h}_2]^{-1}$ denote the weights assigned to the estimable functions $\mathbf{h}_1^T \boldsymbol{\tau}$ and $\mathbf{h}_2^T \boldsymbol{\tau}$. The ratio of these weighted variances is

$$\frac{w_1 \text{Var}_d(\widehat{\mathbf{h}_1^T \boldsymbol{\tau}})}{w_2 \text{Var}_d(\widehat{\mathbf{h}_2^T \boldsymbol{\tau}})} = 1 \Leftrightarrow \frac{\text{Var}_d(\widehat{\mathbf{h}_1^T \boldsymbol{\tau}})}{\text{Var}_d(\widehat{\mathbf{h}_2^T \boldsymbol{\tau}})} = \frac{w_2}{w_1}.$$

If the weights are equal, then the variances must be equal and so the estimable functions are estimated with equal efficacy. If $w_2 > w_1$ then d will estimate $\mathbf{h}_2^T \boldsymbol{\tau}$ with smaller variance than $\mathbf{h}_1^T \boldsymbol{\tau}$. Hence weight balanced designs estimate those estimable functions assigned the largest weight with the smallest, and equal, variance. This does not immediately say that the design will be optimal for all, or any, Φ_W , but it is behaving like we would want a weighted optimal design to behave.

Matrix averaging with respect to Π_W and the convexity property of weighted criteria can be used to create lower bounds for $\Phi_W(\mathbf{C}_d)$ just as in the standard approach. It then becomes desirable to identify which information matrices are invariant to all such permutations. Clearly completely symmetric matrices fall in this group, but other “less symmetric” matrices become candidates for weighted optimality due to the potentially smaller permutation group, Π_W . Indeed, it is this reduction in symmetry that allows such designs to place greater emphasis on some estimable functions.

Now we show that all weight balanced designs have information matrices that are invariant to all permutations in Π_W , further implicating their likelihood of containing a weighted optimal design for some weighted criterion.

Lemma 2.6.2. *Let $d^* \in \mathcal{D}$ be a weight balanced design for \mathbf{W} . Then \mathbf{C}_{d^*} is invariant to all simultaneous row and column permutations in Π_W .*

Proof. As shown in the proof of Lemma 2.3.3, the positive eigenvalues of \mathbf{C}_{d^*W} equal those of $\mathbf{W}_*^{1/2} \mathbf{C}_{d^*} \mathbf{W}_*^{1/2}$ where $\mathbf{W}_*^{1/2} = [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{1/2} + (\mathbf{I} - \mathbf{P}_\tau)$. Because $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ and $(\mathbf{I} - \mathbf{P}_\tau)$ are orthogonal and $[\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{1/2}$ is a square-root decomposition based on the spectral decomposition of $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$, $\mathbf{W}_*^{-1/2} = [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{+1/2} + (\mathbf{I} - \mathbf{P}_\tau)$. That d^* is weight balanced for \mathbf{W} implies that the positive eigenvalues of \mathbf{C}_{dW} , and hence $\mathbf{W}_*^{1/2} \mathbf{C}_{d^*} \mathbf{W}_*^{1/2}$, are all equal to (say) $\epsilon > 0$. Note that $\mathcal{R}(\mathbf{W}_*^{1/2} \mathbf{C}_d \mathbf{W}_*^{1/2}) = \mathcal{E}$ for all d and so

$$\begin{aligned} \mathbf{W}_*^{1/2} \mathbf{C}_{d^*} \mathbf{W}_*^{1/2} = \epsilon \mathbf{P}_\tau &\Rightarrow \mathbf{C}_d = \epsilon \mathbf{W}_*^{-1/2} \mathbf{P}_\tau \mathbf{W}_*^{-1/2} = \epsilon [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{+1/2} [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^{+1/2} \\ &= \epsilon [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^+ , \end{aligned}$$

the MP inverse of $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$. This matrix has the same eigenspaces as $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ and so permutation invariance over Π_W of $\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau$ implies that for $\mathbf{C}_d = \epsilon [\mathbf{P}_\tau \mathbf{W}^{-1} \mathbf{P}_\tau]^+$. \square

The second condition in Kiefer's result required \mathbf{C}_{d^*} to have maximal trace across all designs. In a weighted optimality context, however, a design, d , having \mathbf{C}_{dW} that has maximal trace across all weighted information matrices may be far from desirable.

Example 2.6.1. *For the cell means model the information matrix is $\mathbf{C}_d = \mathbf{R}_d$ where $\mathbf{R}_d = \text{Diag}(r_{d1}, \dots, r_{dp})$, a diagonal matrix of the number of replications assigned to each treatment. Assume we use a diagonal weight matrix $\mathbf{W} = \text{Diag}(w_1, \dots, w_p)$ to declare relative interest in estimating the treatment means. Clearly to achieve complete symmetry of \mathbf{C}_{dW} the replication numbers must be inversely proportional to their assigned weights. For the sake of argument, we allow the number of replication to be non-integers, but still greater than 0, so such a \mathbf{C}_{dW} exists. Such designs are called approximate or continuous designs (see Section 3.3.1).*

The trace of any \mathbf{C}_{dW} will be

$$\sum_{i=1}^p \frac{r_{di}}{w_i}$$

which we seek to maximize with respect to the replications subject to the constraint $\sum_i r_{di} = N$ and $r_{di} > 0$. This is a linear programming problem but the vertices, where the maximum solution(s) typically lie, are not included due to the non-negative constraints on the r_{di} . If we could include the vertices, the obvious maximal trace would occur with a design that assigns $r_{di} = N$ for the treatment having the smallest weight assigned to it, say $w_{(1)}$. There are approximate designs that can approach this value so we may consider $N/w_{(1)}$ as a tight upper bound on the maximal trace. It seems highly unlikely that for general diagonal \mathbf{W} there would exist even an approximate design that achieves both complete symmetry of \mathbf{C}_{dW} and also maximal trace.

As seen in the above example, once we introduce a desired weighting, maximizing trace and balancing weighted information matrices can often lead to very different designs. Balance of \mathbf{C}_{dW} is a superior property than maximal trace, which can direct attention to nearly

singular designs, as seen in Example 2.6.1. A Kiefer-like result for universally weighted optimal designs in terms of weighted information matrices seems unlikely as the two components of that result cannot be simultaneously achieved.

Weighted optimality for a given \mathbf{W} and criterion $\Phi_{\mathbf{W}}$ will need to be treated on a case-by-case basis, both in terms of the chosen estimation space and weight matrix. We now turn our attention to developing a class of weight matrices that allow direct assignment of weights to a set of estimable functions and finding their corresponding weighted optimal designs. The theory is demonstrated for a blocked treatment-versus-control experiment and completely randomized designs that efficiently estimate all OP and BP effects.

Chapter 3

Weighted optimality for a set of estimable functions

The main motivation behind weighted optimality criteria is that experimental goals sometime imply differential interest among the estimable functions. Diagonal weight matrices are most useful when the weighting is motivated solely based on the parameters τ_i involved. For comparative experiments with treatment effects, $\boldsymbol{\tau}$, this is appealing when the treatments may be partitioned into g groups of varying importance, which was an important application in Morgan and Wang (2010). This chapter focuses on a class of weight matrices that allows a researcher to directly assign weight to a chosen set of estimable functions.

We begin with an equivalence lemma that relates the standard approach to design optimization for a set of estimable functions and a class of weighted optimality criteria, demonstrating how weighted optimality brings new insight to a broad class of design problems. Then we develop a class of weight matrices that allow direct assignment of weights to a set of estimable function. One of the most interesting and useful incorporations of weighting presented in this chapter is through the A_W -criterion, which is shown to be able to measure

the average weighted variance for a large set of estimable functions with any desired weighting. Applications for treatment-versus-control experiments and factorial experiments serve to demonstrate the theory.

3.1 Model transformations and weighted optimality

Let \mathbf{H} be an $s \times p$ matrix satisfying $\mathcal{R}(\mathbf{H}) = \mathcal{E}$, so $r(\mathbf{H}) = k$ and each row, say \mathbf{h}_j^T , is a coefficient vector for an estimable function of interest. In Section 1.2.1 we presented the information matrix for the TvC contrasts in the form $(\mathbf{H}\mathbf{C}_d^-\mathbf{H}^T)^{-1}$. For a general \mathbf{H} described in this section, this matrix is invertible if and only if \mathbf{H} is $k \times p$. For $s > k$ estimable functions, $\mathbf{H}\boldsymbol{\tau}$, transform model (1.1) as follows

$$\begin{aligned}\mathbf{y} &= \mathbf{X}_d\boldsymbol{\tau} + \mathbf{L}\boldsymbol{\beta} + \mathbf{e} = \mathbf{X}_d(\mathbf{H}^T\mathbf{H})^+\mathbf{H}^T\mathbf{H}\boldsymbol{\tau} + \mathbf{L}\boldsymbol{\beta} + \mathbf{e} \\ &= \mathbf{X}_{dH}\boldsymbol{\tau}_H + \mathbf{L}\boldsymbol{\beta} + \mathbf{e}\end{aligned}\tag{3.1}$$

for $\mathbf{X}_{dH} = \mathbf{X}_d(\mathbf{H}^T\mathbf{H})^+\mathbf{H}^T$ and $\boldsymbol{\tau}_H = \mathbf{H}\boldsymbol{\tau}$. This transformation produces an equivalent model since $\mathbf{P}_\tau = (\mathbf{H}^T\mathbf{H})^+(\mathbf{H}^T\mathbf{H})$ and estimable functions of $\boldsymbol{\tau}^* = \mathbf{P}_\tau\boldsymbol{\tau}$ are exactly those of $\boldsymbol{\tau}$. The information matrix for $\boldsymbol{\tau}_H$ in (3.1), denoted \mathbf{C}_{dH} , is

$$\mathbf{C}_{dH} = \mathbf{X}_{dH}^T(\mathbf{I} - \mathbf{P}_L)\mathbf{X}_{dH} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^+\mathbf{C}_d(\mathbf{H}^T\mathbf{H})^+\mathbf{H}^T.\tag{3.2}$$

One may then apply standard optimality criteria to \mathbf{C}_{dH} , which lies in $\mathfrak{C}_\mathcal{E}$, to evaluate designs in terms of the estimable functions of interest.

Letting $\mathbf{B} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^+\mathbf{C}_d^{1/2}$, and using the fact that $\mathbf{B}\mathbf{B}^T (= \mathbf{C}_{dH})$ and $\mathbf{B}^T\mathbf{B}$ have the

same nonzero eigenvalues, any positive eigenvalue λ of \mathbf{C}_{dH} satisfies

$$\begin{aligned} |\mathbf{C}_{dH} - \lambda \mathbf{I}| = 0 &\Leftrightarrow |\mathbf{B}^T \mathbf{B} - \lambda \mathbf{I}| = 0 \Leftrightarrow |\mathbf{C}_d^{1/2} (\mathbf{H}^T \mathbf{H})^+ \mathbf{C}_d^{1/2} - \lambda \mathbf{I}| = 0 \\ &\Leftrightarrow |(\mathbf{H}^T \mathbf{H})^{+1/2} \mathbf{C}_d (\mathbf{H}^T \mathbf{H})^{+1/2} - \lambda \mathbf{I}| = 0. \end{aligned} \quad (3.3)$$

In light of (3.3), define the (unscaled) weight matrix \mathbf{W}_H to be

$$\mathbf{W}_H = \mathbf{H}^T \mathbf{H} + (\mathbf{I} - \mathbf{P}_\tau) \quad (3.4)$$

which has inverse $\mathbf{W}_H^{-1} = (\mathbf{H}^T \mathbf{H})^+ + (\mathbf{I} - \mathbf{P}_\tau)$ and $\mathbf{W}_H^{-1/2} = (\mathbf{H}^T \mathbf{H})^{+1/2} + (\mathbf{I} - \mathbf{P}_\tau)$. It follows that \mathbf{C}_{dH} , and \mathbf{C}_{dW} for $\mathbf{W} = \mathbf{W}_H$, have the same positive eigenvalues, giving us the following lemma.

Lemma 3.1.1. *For optimality criteria that depend only on the eigenvalues of \mathbf{C}_{dH} ,*

$$\Phi(\mathbf{C}_{dH}) = \Phi_W(\mathbf{C}_d),$$

where $\mathbf{W} = \mathbf{W}_H$ is defined in (3.4). Hence design optimization for $\boldsymbol{\tau}_H$ in model (3.1) is a weighted optimality problem for $\boldsymbol{\tau}$ in model (1.1) with weight matrix \mathbf{W}_H .

We encountered an important result involving the general form $\mathbf{H}^T \mathbf{H}$ in Section 1.2.3. Denote the k positive eigenvalues of $\mathbf{H}^T \mathbf{H}$ as $\lambda_1, \dots, \lambda_k$ and their corresponding eigenvectors as $\mathbf{e}_1, \dots, \mathbf{e}_k$. Theorem 1.2.1 states that the approximate information matrix of the form $\mathbf{M}^* = \alpha \sum_{i=1}^k \sqrt{\lambda_i} \mathbf{e}_i \mathbf{e}_i^T$, where α is chosen to give \mathbf{M}^* maximal trace across all possible $\mathbf{C}_d \in C_H$, is A -optimal for $\mathbf{H}\boldsymbol{\tau}$. The eigenvectors of $\mathbf{P}_\tau \mathbf{W}_H^{-1} \mathbf{P}_\tau$ corresponding to positive eigenvalues clearly coincide with $\mathbf{e}_1, \dots, \mathbf{e}_k$ so the eigenspaces of \mathbf{M}^* corresponding to its distinct positive eigenvalues equal those of $\mathbf{P}_\tau \mathbf{W}_H^{-1} \mathbf{P}_\tau$. Hence \mathbf{M}^* is permutation invariant to all $\boldsymbol{\Pi} \in \Pi_{W_H}$. We see that the concept of aligned designs is related to invariance to

weight-preserving permutations, and hence A_W -optimality with respect to \mathbf{W}_H . Moreover, Lemma 3.1.1 and the balance of aligned designs with respect to Π_W says that such designs (if they exist) are potentially optimal for many eigenvalue criteria, not just the A_W -criterion.

There are two common transformations used with ANOVA models arising from comparative experiments, not for the purpose of targeting specific parametric functions, but simply to make all treatment effects identifiable: the baseline transform and the sum-to-zero constraint. Evidently these transformation have the potential to impact the design optimization problem. We now examine the effects of performing such transformation from the weighted optimality point of view.

3.1.1 The baseline and sum-to-zero transform

Assume \mathcal{E} contains all treatment contrasts so $\mathbf{P}_\tau = \mathbf{I} - \frac{1}{p}\mathbf{J}$. Let τ_p in $\boldsymbol{\tau}$ be the baseline treatment and transform $\boldsymbol{\tau}$ to $\boldsymbol{\theta} = \mathbf{H}\boldsymbol{\tau}$ where $\mathbf{H} = (\mathbf{I}_{p-1} \mid -\mathbf{1}_{p-1})$. Note that $\boldsymbol{\theta}$ has $p-1$ parameters. Here

$$\mathbf{H}^T \mathbf{H} = \left(\begin{array}{c|c} \mathbf{I}_{p-1} & -\mathbf{1}_{p-1} \\ \hline \mathbf{1}_{p-1}^T & p \end{array} \right)$$

which has MP-inverse

$$(\mathbf{H}^T \mathbf{H})^+ = \left(\begin{array}{c|c} \mathbf{I}_{p-1} - \frac{1}{p-1}\mathbf{J}_{p-1} & \mathbf{0} \\ \hline \mathbf{0}^T & 0 \end{array} \right) + \frac{1}{p} \left(\begin{array}{c|c} \frac{1}{p(p-1)}\mathbf{J}_{p-1} & -\frac{1}{p}\mathbf{1} \\ \hline -\frac{1}{p}\mathbf{1}^T & \frac{p-1}{p} \end{array} \right).$$

The weight assigned to any contrast, $\mathbf{c}^T \boldsymbol{\tau}$ is then $[\mathbf{c}^T (\mathbf{H}^T \mathbf{H})^+ \mathbf{c}]^{-1}$. If $\mathbf{c}^T \boldsymbol{\tau}$ is a normalized contrast involving only the first $p-1$ treatments then clearly this weight will be 1. However, this is not the case for contrasts that involve the baseline treatment. For example, the contrasts $\frac{\tau_i - \tau_p}{\sqrt{2}}$ for $i = 1, \dots, p-1$ are assigned weight 2, implying that the comparisons

involving the baseline treatment are of greater interest. In fact, the ratio of these weights is exactly that produced in the limit as $w_1 \rightarrow 1$ by a diagonal weight matrix $Diag(\mathbf{w})$ where $\mathbf{w}^T = (w_2, w_2, \dots, w_2, w_1)$ and $w_2 = (1 - w_1)/(p - 1)$. This limit, as shown in Morgan and Wang (2010), produces the standard treatment-versus-control optimality setup. As is well known, optimality results for this setup are much different from that of the untransformed model.

The sum-to-zero transformation replaces τ_i by $\tau_i - \bar{\tau}$, where $\bar{\tau} = (\sum_{i=1}^p \tau_i)/p$. The necessary transformation matrix \mathbf{H} is then

$$\mathbf{H} = \mathbf{I}_p - \frac{1}{p}\mathbf{J}_p = \mathbf{P}_\tau.$$

Hence $\mathbf{H}^T \mathbf{H} = \mathbf{P}_\tau$ and $\mathbf{W} = \mathbf{I}$ so optimization for the sum-to-zero parameters is equivalent to the standard approach for optimization for all estimable functions.

3.2 Assigning weights to estimable functions

The goal of this section is to construct weight matrices that assign any chosen weights to any specified set of estimable functions. As will be shown, how this is done depends on how many estimable functions we want to directly assign weights to. When assigning weights to functions, we must also contemplate whether to assign them to the normalized or non-normalized versions, even though the resulting weighted variances will be equivalent. We begin by looking at assigning weights to a set of k orthonormal estimable functions.

3.2.1 Assigning weights to k orthonormal estimable functions

Weight matrices satisfying $\mathbf{W}(\mathbf{I} - \mathbf{P}_\tau) = q(\mathbf{I} - \mathbf{P}_\tau)$ have k orthonormal eigenvectors that make up an orthonormal basis of all estimable functions. Let \mathbf{H}^* be the $k \times p$ matrix whose rows are these eigenvectors so that $\mathbf{H}^*\boldsymbol{\tau}$ are k orthonormal estimable functions. Consider the weight matrix

$$\mathbf{W}_{H^*} = \mathbf{H}^{*T} \mathbf{D}_w \mathbf{H}^* + (\mathbf{I} - \mathbf{P}_\tau) , \quad (3.5)$$

where \mathbf{D}_w is the $k \times k$ diagonal matrix of the eigenvalues corresponding to the rows of \mathbf{H}^* . These eigenvalues may be freely chosen by the experimenter and are easily shown to be the weights assigned to the $\mathbf{H}^*\boldsymbol{\tau}$. The $\mathbf{H}^*\boldsymbol{\tau}$ are also clearly weighted orthogonal, and by Lemma 2.3.2, an A_W -optimal design for this weight matrix minimizes the average weighted variance of these k functions.

3.2.2 An admissibility criterion

Every information matrix with estimation space \mathcal{E} admits a spectral decomposition of the form $\mathbf{C}_d = \mathbf{H}_d^T \mathbf{D}_{dw} \mathbf{H}_d$ where the rows of \mathbf{H}_d are k orthonormal eigenvectors of \mathbf{C}_d and \mathbf{D}_{dw} is a $k \times k$ diagonal matrix of the positive eigenvalues of \mathbf{C}_d corresponding to the eigenvectors. We may then create a weight matrix of the form (3.5) from a chosen design, say d 's, information matrix, $\mathbf{W}_d = \mathbf{C}_d + (\mathbf{I} - \mathbf{P}_\tau)$. Hence $\mathbf{W}_d^{-1/2} = \mathbf{C}_d^{+1/2} + (\mathbf{I} - \mathbf{P}_\tau)$ and the weighted information matrix for design d is

$$\mathbf{C}_{dW} = \mathbf{C}_d^{+1/2} \mathbf{C}_d \mathbf{C}_d^{+1/2} = \mathbf{P}_\tau ,$$

making d weight balanced for \mathbf{W}_d . We now explore the E_W -criterion for such weight matrices generated by a specific type of design.

For given design class \mathcal{D} , a design $d \in \mathcal{D}$ is *admissible* if there is no $\tilde{d} \in \mathcal{D}$ for which $\mathbf{C}_{\tilde{d}} \neq \mathbf{\Pi} \mathbf{C}_d \mathbf{\Pi}^T$ for all permutation matrices $\mathbf{\Pi}$ and $\text{Var}_{\tilde{d}}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) \leq \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$ for every estimable function $\mathbf{h}^T \boldsymbol{\tau}$. Let \mathcal{C}^* be the subclass of information matrices corresponding to all admissible designs. Lemma 3.2.1 supplies a weighted criterion that takes admissible designs from design contender to optimum, showing that every admissible design maximizes information in some way.

Lemma 3.2.1. *For a given $d \in \mathcal{D}$ where $\mathbf{C}_d \in \mathcal{C}^*$, d is E_W -optimal for \mathbf{W}_d .*

Proof. As shown earlier, $\mathbf{C}_{dW} = \mathbf{P}_{\boldsymbol{\tau}}$ under \mathbf{W}_d and so the positive eigenvalues of \mathbf{C}_{dW} all equal 1, which is also equal to $1/E_{dW}$. If a competing design \tilde{d} is E_W -better than d then all the eigenvalues for $\mathbf{C}_{\tilde{d}W}$ are greater than 1. It follows from Lemma 2.2.1 that $\text{Var}_{\tilde{d}W}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) < \text{Var}_{dW}(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$ for every estimable function $\mathbf{h}^T \boldsymbol{\tau}$. But the weight applied to a particular $\mathbf{h}^T \boldsymbol{\tau}$ is constant for all designs, implying that $\text{Var}_{\tilde{d}}(\widehat{\mathbf{h}^T \boldsymbol{\tau}}) < \text{Var}_d(\widehat{\mathbf{h}^T \boldsymbol{\tau}})$ for every contrast and that d is inadmissible, a contradiction. \square

Recall that E_{dW} equals the largest weighted variance across all estimable functions for a design d . It is tempting to assume that for \mathbf{W}_d , d minimizes the maximal weighted variance across all the k orthonormal estimable functions that make up the spectral decomposition of \mathbf{C}_d . However, the following counterexample shows that is not always the case.

Counterexample 3.2.1. Take a completely randomized design for $p = 3$ treatments with $N = 7$ EU's. There are only 15 possible connected designs, being all permutations of the following replication vectors:

Design	Replication Vector	Permutations
d_1	(1, 2, 4)	6
d_2	(1, 3, 3)	3
d_3	(1, 1, 5)	3
d_4	(2, 2, 3)	3

We first show that design d_1 is admissible in order to apply Lemma 3.2.1. Consider the normalized eigenvector \mathbf{h}_2 corresponding to the maximum eigenvalue of \mathbf{C}_{d_1} . The variance of $\widehat{\mathbf{h}_2^T \boldsymbol{\tau}}$ under d_1 is 0.3628. Only design d_2 estimates $\mathbf{h}_2^T \boldsymbol{\tau}$ with smaller variance ($\text{Var}_{d_2}(\widehat{\mathbf{h}_2^T \boldsymbol{\tau}}) = 0.3456$), which can be checked manually across the other 14 designs. As d_1 estimates the pairwise contrast $\frac{\tau_1 - \tau_3}{\sqrt{2}}$ with smaller variance than d_2 (0.625 versus 0.667, respectively), d_1 is admissible.

It is straightforward to check that d_2 also estimates $\mathbf{h}_1^T \boldsymbol{\tau}$ given by the other (orthonormal) eigenvector of \mathbf{C}_{d_1} with smaller variance than d_1 , so for $i = 1$ and 2,

$$\frac{\text{Var}_{d_2}(\widehat{\mathbf{h}_i^T \boldsymbol{\tau}})}{\text{Var}_{d_1}(\widehat{\mathbf{h}_i^T \boldsymbol{\tau}})} = \frac{\text{Var}_{d_2 W}(\widehat{\mathbf{h}_i^T \boldsymbol{\tau}})}{\text{Var}_{d_1 W}(\widehat{\mathbf{h}_i^T \boldsymbol{\tau}})} < 1$$

and d_2 has smaller weighted variances than d_1 for the two estimable functions corresponding to the eigenvectors of \mathbf{C}_{d_1} even though d_1 is E_W -optimal with respect to \mathbf{W}_{d_1} .

3.2.3 Assigning weights to k estimable functions

The weight matrices constructed in (3.5) are straightforward in their derivation, but are less likely to be used in practice. Only in specific cases, such as factorial experiments, would a set

of k orthonormal contrasts be the main focus of estimation. In this section we derive weight matrices that can directly assign contrast weights to a set of at least k linearly independent, normalized estimable functions. These weight matrices will have a form similar to those in Section 3.2.1. Why these functions are normalized will be discussed later.

Let \mathbf{H} be a $k \times p$ matrix where $\mathbf{H}\boldsymbol{\tau}$ are k normalized, linearly independent estimable functions and choose w_1, \dots, w_k to be the k weights that are to be assigned to these k functions. Let \mathbf{D}_w be the diagonal matrix of these weights and write $\mathbf{H}_w = \mathbf{D}_w^{1/2} \mathbf{H}$.

Lemma 3.2.2. *For \mathbf{H}_w described above, the weight matrix $\mathbf{W}_{Hw} = \mathbf{H}^T \mathbf{D}_w \mathbf{H} + (\mathbf{I} - \mathbf{P}_\tau)$ assigns weight w_j to the contrast $\mathbf{h}_j^T \boldsymbol{\tau}$ where \mathbf{h}_j^T is the j -th row of \mathbf{H} .*

Proof. The assigned weight is the inverse of the j -th diagonal element of $\mathbf{H}\mathbf{W}_{Hw}^{-1}\mathbf{H}^T$. Now $\mathbf{H}_w \mathbf{W}_{Hw}^{-1} \mathbf{H}_w^T = \mathbf{H}_w (\mathbf{H}_w^T \mathbf{H}_w)^+ \mathbf{H}_w^T$ is symmetric, idempotent, and has full rank and so equals \mathbf{I}_k . This gives

$$\mathbf{H}_w (\mathbf{H}_w^T \mathbf{H}_w)^+ \mathbf{H}_w^T = \mathbf{I}_k \Leftrightarrow \mathbf{H} (\mathbf{H}_w^T \mathbf{H}_w)^+ \mathbf{H}^T = \mathbf{D}_w^{-1} \Leftrightarrow \mathbf{H} \mathbf{W}_{Hw}^{-1} \mathbf{H}^T = \mathbf{D}_w^{-1}$$

showing that the j -th diagonal element of $\mathbf{H}\mathbf{W}_{Hw}^{-1}\mathbf{H}^T$ is $1/w_j$, as required. \square

It is clear from the proof of Lemma 3.2.2 that the weight matrix \mathbf{W}_{Hw} makes the chosen k estimable functions $\mathbf{H}\boldsymbol{\tau}$ weighted orthogonal, which we now state as a corollary.

Corollary 3.2.1. *For the weight matrix \mathbf{W}_{Hw} described in Lemma 3.2.2, the chosen k estimable functions are weighted orthogonal. Furthermore, A_{dW} measures the average weighted variance of these k functions with weights specified in \mathbf{D}_w :*

$$A_{dW} = \frac{1}{k} \sum_{j=1}^k w_j \text{Var}_d(\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}) .$$

One must take care not to read too much into Lemma 3.2.2. As an example, consider $\mathbf{D}_w = \mathbf{I}_k$ to place equal weight on the k normalized estimable functions. This gives the weight matrix $\mathbf{W}_H = \mathbf{H}^T \mathbf{H} + (\mathbf{I} + \mathbf{P}_\tau)$. An arbitrary estimable function may be written as $\mathbf{h}^T \boldsymbol{\tau} = \mathbf{m}^T \mathbf{H} \boldsymbol{\tau}$ for some $k \times 1$ vector \mathbf{m} , so the weight assigned to $\mathbf{h}^T \boldsymbol{\tau}$ is the inverse of $\mathbf{h}^T \mathbf{W}_H^{-1} \mathbf{h} = \mathbf{m}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^+ \mathbf{H}^T \mathbf{m} = \mathbf{m}^T \mathbf{m}$. This weight is 1 for any normalized \mathbf{m} , the same weight for the functions $\mathbf{H} \boldsymbol{\tau}$. But consider for $p = 4$ and $\mathbf{P}_\tau = \mathbf{I} - \frac{1}{4} \mathbf{J}$ the three contrasts $(\tau_1 - \tau_3)/\sqrt{2}$, $(\tau_1 - \tau_4)/\sqrt{2}$, and $(\tau_2 - \tau_3)/\sqrt{2}$, and their corresponding \mathbf{H} . Then $\mathbf{m}^T = (0, -1, 1, 1)$ assigns $(\tau_2 - \tau_4)/\sqrt{2}$ the weight $(\mathbf{m}^T \mathbf{m})^{-1} = 1/3$, which would be disconcerting if one preferred that all four of these contrasts be assigned the same weight. The non-orthogonal \mathbf{H} has allowed the non-normalized \mathbf{m} to produce a normalized contrast. The proof of Lemma 3.2.2 still holds even if the rows of \mathbf{H} are non-normalized. It is then necessary to investigate the impact of normalizing, or not normalizing, the rows of \mathbf{H} on the weights for other estimable functions and resulting weighted criteria.

3.2.4 Normalization

For simplicity, here we work with full rank information matrices so that all $\mathbf{h}^T \boldsymbol{\tau}$ are estimable. We take \mathbf{H} to be full rank, with rows $\mathbf{h}_1^T, \dots, \mathbf{h}_p^T$, at least one of which is not normalized. Let \mathbf{D}_H be the diagonal matrix with diagonal entries $\mathbf{h}_j^T \mathbf{h}_j$. Then the rows of $\tilde{\mathbf{H}} = \mathbf{D}_H^{-1/2} \mathbf{H}$, denoted $\tilde{\mathbf{h}}_j^T = \frac{1}{\sqrt{\mathbf{h}_j^T \mathbf{h}_j}} \mathbf{h}_j^T$, are normalized and at least one row of \mathbf{H} is not normalized. These two options correspond to two weight matrices $\mathbf{W} = \mathbf{H}^T \mathbf{H}$ and $\tilde{\mathbf{W}} = \tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \mathbf{H}^T \mathbf{D}_H^{-1} \mathbf{H}$ that are not estimation equivalent, which should generally lead to different assignment of weights to all estimable functions (they are not estimation equivalent). Hence normalization of the rows of \mathbf{H} will likely impact design ordering.

Lemma 3.2.3. *The weighted variance of $\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}$ with respect to $\tilde{\mathbf{W}}$ is the variance of $\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}$,*

that is

$$\text{Var}_{d\tilde{\mathbf{W}}}(\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}) = [\mathbf{h}_j^T \mathbf{h}_j]^{-1} \text{Var}_d(\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}) = \text{Var}_d(\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}) .$$

The weighted variance of $\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}$ with respect to \mathbf{W} is the variance of $\mathbf{h}_j^T \boldsymbol{\tau}$, that is

$$\text{Var}_{d\mathbf{W}}(\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}) = [\mathbf{h}_j^T \mathbf{h}_j] \text{Var}_d(\widehat{\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}}) = \text{Var}_d(\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}) .$$

Proof. The weights for the p estimable functions $\mathbf{h}_j^T \boldsymbol{\tau}$ with respect to $\tilde{\mathbf{W}}$ are the inverse diagonal elements of

$$\mathbf{H} \tilde{\mathbf{W}}^{-1} \mathbf{H}^T = \mathbf{H} (\mathbf{H}^T \mathbf{D}_H^{-1} \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{D}_H$$

and so are just the inverses of the squared lengths $\mathbf{h}_j^T \mathbf{h}_j$ of the coefficient vectors \mathbf{h}_j , giving the first result.

The weights for the p estimable functions $\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}$ with respect to \mathbf{W} are the inverses of the diagonal of

$$\tilde{\mathbf{H}} \mathbf{W}^{-1} \tilde{\mathbf{H}}^T = \mathbf{D}_H^{-1/2} \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{D}_H^{-1/2} = \mathbf{D}_H^{-1} ,$$

that is, they are the squared lengths $\mathbf{h}_j^T \mathbf{h}_j$. □

Lemma 3.2.3 says that if the rows of \mathbf{H} are normalized, then weighted evaluation of the non-normalized $\mathbf{h}_j^T \boldsymbol{\tau}$ is nevertheless based on its normalized version. On the other hand, if one does not normalize the rows of \mathbf{H} , then weighted evaluation of the normalized $\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}$ is equivalent to unweighted evaluation of the non-normalized $\mathbf{h}_j^T \boldsymbol{\tau}$. That is, the weight assigned to $\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}$ equals the length of the non-normalized coefficient vector \mathbf{h}_j .

The purpose of using a transformation \mathbf{H} is to directly target estimation on the transformed parameters $\boldsymbol{\theta} = \mathbf{H} \boldsymbol{\tau}$. What has just been shown is that if the rows of \mathbf{H} are normalized,

then the optimality problem is in terms of normalized functions of the original parameters. If the rows of \mathbf{H} are not normalized, then the optimality problem is in terms of non-normalized functions of the original parameters. In the latter case, the fact that weighted optimality automatically incorporates normalization (see (2.2)) does not “carry over” to the original parameters underlying transformed models, unless the rows of the transformation matrix were themselves normalized. Indeed, if the rows of \mathbf{H} are not normalized then the induced weighting $\mathbf{W} = \mathbf{H}^T \mathbf{H}$ “un-normalizes” the $\tilde{\mathbf{h}}_j^T \boldsymbol{\tau}$ when evaluating their weighted variances. That is, \mathbf{W} assigns weight equal to the length of the vector to the normalized functions $\tilde{\mathbf{h}}_j^T$ so that longer vectors will be deemed more important.

Our arguments up to this point have been focused on the weighted variances of $\mathbf{H}\boldsymbol{\tau}$ and $\tilde{\mathbf{H}}\boldsymbol{\tau}$ induced from the two corresponding weight matrices. In particular, non-normalizing was shown to induce weights based on the length of the vector, which may or may not be indicative of the importance of that function. These arguments do not immediately tell us anything about whether the design ordering changes with the weight matrices. In fact, the answer likely depends on many conditions such as how different the lengths of the rows are and the criterion chosen.

It is thus sufficient, and of greater clarity, to develop a general theory when the rows of \mathbf{H} are normalized. In this way, the weights we want to assign to the estimable functions are made explicit (they are not “tainted” by the length of the contrast). If estimation is desired specifically to the non-normalized version, we may introduce the length back by modifying the diagonal elements of the matrix \mathbf{D}_H in $\tilde{\mathbf{W}}$.

3.2.5 Assigning weights to $s > k$ estimable functions

Focusing weighted estimation on a set of k linearly independent estimable functions is certainly an important step in generating a flexible class of weighted criteria, but we have seen cases where this restriction may induce undesirable weights for other functions. To remedy this, we would like a weight matrix that allows us to assign weights to a larger set of estimable functions, say $s > k$. One may try to accomplish this through a weight matrix of the form in Lemma 3.2.2 with $\mathbf{H}\boldsymbol{\tau}$ being s estimable functions specified by \mathbf{H} (now an $s \times p$ matrix) and \mathbf{D}_w an $s \times s$ diagonal matrix of desired weights. This weight matrix will not in general give $[\mathbf{h}_j^T \mathbf{W}_{Hw}^{-1} \mathbf{h}_j]^{-1} = w_j$ for all $j = 1, \dots, s$, as seen in the following example.

Example 3.2.1. Suppose we partition 5 treatments in a comparative experiment into two groups: $V_1 = \{1, 2, 3\}$ and $V_2 = \{4, 5\}$. The estimation space is taken to be all contrasts, and so $k = 5 - 1 = 4$. Let $\mathbf{H}\boldsymbol{\tau}$ be the normalized pairwise contrasts so $s = \binom{5}{2} > 4$. We hope to assign weight w_0 to the within-group contrasts, and weight 1 to the between-group contrasts, using the appropriate \mathbf{D}_w . Then $\mathbf{H}_w^T = \mathbf{H}^T \mathbf{D}^{1/2}$ can be written as

$$\mathbf{H}_w^T = \frac{1}{\sqrt{2}} \left(\begin{array}{ccc|c|cccccc} \sqrt{w_0} & \sqrt{w_0} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -\sqrt{w_0} & 0 & \sqrt{w_0} & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -\sqrt{w_0} & -\sqrt{w_0} & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & \sqrt{w_0} & -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -\sqrt{w_0} & 0 & -1 & 0 & -1 & 0 & -1 \end{array} \right)$$

so that $\mathbf{W}_{Hw} = \mathbf{H}_w^T \mathbf{H}_w + \frac{1}{2} \mathbf{J}$ (the $\frac{1}{2}$ is chosen to get a nice form for \mathbf{W}_{Hw} but does not affect the weights) is equal to

$$\mathbf{W}_{Hw} = \frac{1}{2} \left(\begin{array}{c|c} (3w_0 + 2)\mathbf{I}_3 + (1 - w_0)\mathbf{J}_3 & 0 \\ \hline 0 & (2w_0 + 3)\mathbf{I}_2 + (1 - w_0)\mathbf{J}_2 \end{array} \right)$$

and thus

$$\mathbf{W}_{Hw}^{-1} = 2 \left(\begin{array}{c|c} \frac{1}{3w_0+2}[\mathbf{I}_3 - \frac{1-w_0}{5}\mathbf{J}_3] & 0 \\ \hline 0 & \frac{1}{2w_0+3}[\mathbf{I}_2 - \frac{1-w_0}{5}\mathbf{J}_2] \end{array} \right). \quad (3.6)$$

Using (3.6) shows that \mathbf{W}_{Hw} assigns weight $(3w_0 + 2)/2$ to normalized pairwise contrasts within V_1 and weight $(2w_0 + 3)/2$ to normalized pairwise contrasts within V_2 . Not only is this not w_0 , but the two values are not even the same for $w_0 \neq 1$.

Suppose we choose $s > k$ normalized, estimable functions that we want to assign equal weight to that also satisfy $\mathbf{H}^T \mathbf{H} = \alpha \mathbf{P}_\tau$. Then the weight matrix $\mathbf{W}_H = \mathbf{H}^T \mathbf{H} + \alpha(\mathbf{I} - \mathbf{P}_\tau) = \alpha \mathbf{I}$ assigns equal weight to those functions. In fact, it assigns equal weight to all normalized estimable functions, reducing weighted optimality to standard optimality. This is the case for $w_0 = 1$ in Example 3.2.1. On the other hand, suppose $\mathbf{H}^T \mathbf{H} \neq \mathbf{P}_\tau$ but there exists a diagonal matrix of unequal weights \mathbf{D}_w for which $\mathbf{H}^T \mathbf{D}_w \mathbf{H} = \mathbf{P}_\tau$. Then the weight matrix \mathbf{W}_{Hw} clearly won't assign the desired weights, rather it will assign equal weights!

From what we have seen, the constructed weight matrix \mathbf{W}_{Hw} will not in general assign the desired weights to the chosen $s > k$ estimable functions. Finding such a weight matrix for a freely chosen set of weights is a very difficult problem, and may be impossible. Even though the desired weights are not necessarily assigned to the $s > k$ functions, we now show that the weights are always preserved in the A_{dW} expression.

3.3 A_W -optimality for $s \geq k$ estimable functions

Section 3.2.3 focused on constructing weight matrices that assign selected weights to k estimable functions. Corollary 3.2.1 showed that such weight matrices make the k functions weighted orthogonal, implying by Lemma 2.3.2 that the A_W -criterion measures the average weighted variance of those functions with weights specified in \mathbf{D}_w . The following theorem

shows that regardless of the weights \mathbf{W}_{Hw} assigns to the $s \geq k$ chosen estimable functions, the \mathbf{W}_{Hw} -weighted A -criterion nonetheless measures the average weighted variance of the s functions with the assigned weights as specified in \mathbf{D}_w .

Theorem 3.3.1. *Let $\mathbf{W}_{Hw} = \mathbf{H}_w^T \mathbf{H}_w + (\mathbf{I} - \mathbf{P}_\tau)$ with $\mathbf{H}_w = \mathbf{D}_w^{1/2} \mathbf{H}$ where $\mathbf{H}\boldsymbol{\tau}$ are $s \geq k$ normalized estimable functions with $\mathcal{R}(\mathbf{H}) = \mathcal{E}$ and \mathbf{D}_w are the desired weights for $\mathbf{H}\boldsymbol{\tau}$. Then A_{dW} with respect to \mathbf{W}_{Hw} is proportional to the average weighted variance for the s estimable functions where the diagonal elements of \mathbf{D}_w are the weights assigned to $\mathbf{H}\boldsymbol{\tau}$.*

Proof. Rewrite \mathbf{W}_{Hw} using its spectral decomposition as $\mathbf{W}_{Hw} = \mathbf{E}\mathbf{D}_\lambda\mathbf{E}^T$ where the columns of \mathbf{E} are eigenvectors satisfying $\mathbf{E}^T\mathbf{E} = \mathbf{I}_p$. Define $\mathbf{M} = \mathbf{E}\mathbf{D}_\lambda^{1/2}$ and let \mathbf{M}_1 be the k columns of \mathbf{M} orthogonal to $(\mathbf{I} - \mathbf{P}_\tau)$. It is easily checked that $\mathbf{M}_1^T \mathbf{W}_{Hw}^{-1} \mathbf{M}_1 = \mathbf{I}_k$ so the columns of \mathbf{M}_1 make up a set of weighted orthogonal contrasts.

Applying Lemma 2.3.2, we have

$$\begin{aligned}
A_{dW} &\propto \text{tr}(\mathbf{M}_1^T \mathbf{C}_d^- \mathbf{M}_1) \\
&= \text{tr}(\mathbf{M}^T \mathbf{C}_d^- \mathbf{M}) \quad \text{since } \mathbf{C}_d(\mathbf{I} - \mathbf{P}_\tau) = 0 \\
&= \text{tr}(\mathbf{E}^T \mathbf{E} \mathbf{M}^T \mathbf{C}_d^- \mathbf{M}) \\
&= \text{tr}(\mathbf{E} \mathbf{M}^T \mathbf{C}_d^- \mathbf{M} \mathbf{E}^T) \\
&= \text{tr}(\mathbf{W}^{1/2} \mathbf{C}_d^- \mathbf{W}^{1/2}) \\
&= \text{tr}(\mathbf{H}^T \mathbf{D}_w \mathbf{H} \mathbf{C}_d^-) \\
&= \text{tr}(\mathbf{D}_w \mathbf{H} \mathbf{C}_d^- \mathbf{H}^T) = \sum_{j=1}^s w_j \text{Var}_d(\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}) .
\end{aligned}$$

The last expression is proportional to the average weighted variance of the s functions $\widehat{\mathbf{H}\boldsymbol{\tau}}$ with corresponding weights specified in \mathbf{D}_w . □

Theorem 3.3.1 provides a very powerful optimality criterion that measures a design's ability to estimate a large set of estimable functions given expressed differential interest among those functions. When applied to the weight matrix in Example 3.2.1, we have averaged weighted variance proportional to

$$w_0 \sum_{\substack{i,j \in V_1 \\ i \neq j}} \text{Var}_d \left(\frac{\widehat{\tau_i - \tau_j}}{\sqrt{2}} \right) + w_0 \text{Var}_d \left(\frac{\widehat{\tau_4 - \tau_5}}{\sqrt{2}} \right) + \sum_{\substack{i \in V_1 \\ j \in V_2}} \text{Var}_d \left(\frac{\widehat{\tau_i - \tau_j}}{\sqrt{2}} \right)$$

even though this weight matrix does not in general assign weight w_0 to the within-group contrasts.

Suppose we use the weighting scheme in Example 3.2.1 for a block design with 5 blocks each of size 4, in which case a balanced incomplete block design, say d_1 , exists and is shown below. When $w_0 = 20$, that is, larger weight is to be set to the within-group comparisons, then $A_{d_1 W} = 2.293$. The incomplete block design, d_2 , displayed below has $A_{d_2 W} = 2.285$ and so is A_W -better than d_1 . The designs below have blocks as columns and treatments are labeled by $1, \dots, 5$:

$$d_1 = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 2 \\ \hline 2 & 2 & 2 & 3 & 3 \\ \hline 3 & 3 & 4 & 4 & 4 \\ \hline 4 & 5 & 5 & 5 & 5 \\ \hline \end{array} \quad d_2 = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 \\ \hline 2 & 2 & 2 & 2 & 3 \\ \hline 3 & 3 & 3 & 4 & 4 \\ \hline 4 & 4 & 5 & 5 & 5 \\ \hline \end{array}.$$

Note that d_2 is nearly identical to the BIBD, except one of the replications for treatment 5 was replaced with one for treatment 1.

To see the impact of choosing d_2 over d_1 , we focus on the within and between-group variances rather than comparing the A_W -values. The BIBD is known to estimate all normalized pairwise comparisons with equal variance, which in this case is $\text{Var}_{d_1} \left(\frac{\widehat{\tau_i - \tau_j}}{\sqrt{2}} \right) = 0.267$. For

d_2 the average variance for the normalized pairwise contrasts of the treatments in V_1 is 0.243 and $\text{Var}\left(\frac{\tau_4 - \tau_5}{\sqrt{2}}\right) = 0.327$, while the average variances for the between-group comparisons is 0.284.

We emphasize that the desired weights in \mathbf{D}_w are maintained in the A_{dW} expression, but are not necessarily the weights individually assigned to the functions by \mathbf{W}_{Hw} . Table 3.1 shows the actual weights assigned to the within- and between-group contrasts. As anticipated, in light of the variances discussed above, the largest weight among these comparisons is for the within-group comparisons of V_1 and the next is for the comparison of treatments 4 and 5. The smallest weight is placed on the between-group contrasts. Perhaps most surprising is that larger weights than the desired $w_0 = 20$ are assigned to the within-group comparisons.

Table 3.1: Weights assigned by \mathbf{W}_{Hw} in Example 3.2.1 to pairwise comparisons for treatments between and within treatment groups V_1 and V_2 with $w_0 = 20$.

Comparison group	Weight
Within V_1	31.0
Within V_2	21.5
Between V_1 and V_2	5.3

There are ways other than through the user-specified estimable functions $\mathbf{H}\boldsymbol{\tau}$ and chosen weights \mathbf{D}_w in Lemma 3.3.1 to interpret the A_W -criterion as shown by the following corollary.

Corollary 3.3.1. *For given weight matrix $\mathbf{W}_{Hw} = \mathbf{H}^T \mathbf{D}_w \mathbf{H} + (\mathbf{I} - \mathbf{P}_\tau)$ consider any other factorization $\mathbf{W}_{Hw} = \tilde{\mathbf{H}}^T \tilde{\mathbf{D}}_w \tilde{\mathbf{H}} + (\mathbf{I} - \mathbf{P}_\tau)$ in which $\tilde{\mathbf{D}}_w$ is a $\tilde{s} \times \tilde{s}$ positive diagonal matrix and $\tilde{\mathbf{H}}\boldsymbol{\tau}$ are $\tilde{s} \geq k$ estimable functions with $r(\tilde{\mathbf{H}}) = k$. Then A_{dW} is proportional to the average weighted variance of the \tilde{s} estimable functions $\tilde{\mathbf{H}}\boldsymbol{\tau}$, with weights specified by the corresponding diagonal elements of $\tilde{\mathbf{D}}_w$.*

Apparently A_{dW} can measure average weighted variance for many different sets of functions.

We now extend Lemma 3.3.1 and Corollary 3.3.1 to a fully general result.

Fix any weight matrix \mathbf{W} . Let $\mathbf{M}_{s \times p}$ be a matrix whose rows are coefficient vectors for estimable functions of interest. Let the rows of \mathbf{M}_w be those of \mathbf{M} except that each is multiplied by the square root of its weight. That is, if \mathbf{m}^T is a row of \mathbf{M} then $\mathbf{m}^T \boldsymbol{\tau}$ is an estimable function of interest and the corresponding row of \mathbf{M}_w is $(\mathbf{m}^T \mathbf{W}^{-1} \mathbf{m})^{-1/2} \mathbf{m}^T$. Let $\mathbf{P}_{\boldsymbol{\tau}_W}$ be the orthogonal projector onto the column space of \mathbf{C}_{dW} , which is the same for all d .

Lemma 3.3.1. *Suppose for some scalar α the matrix \mathbf{M} described above satisfies*

$$\mathbf{W}^{-1/2} \mathbf{M}_w^T \mathbf{M}_w \mathbf{W}^{-1/2} = \alpha \mathbf{P}_{\boldsymbol{\tau}_W}.$$

Then the A_{dW} value is proportional to the average weighted variance of the s contrasts in $\mathbf{M}\boldsymbol{\tau}$.

Proof. The average weighted variance of the s contrasts is proportional to

$$\begin{aligned} \text{tr}(\mathbf{M}_w \mathbf{C}_d^- \mathbf{M}_w^T) &= \text{tr}(\mathbf{M}_w \mathbf{W}^{-1/2} \mathbf{C}_{dW}^+ \mathbf{W}^{-1/2} \mathbf{M}_w^T) \\ &= \text{tr}(\mathbf{C}_{dW}^+ \mathbf{W}^{-1/2} \mathbf{M}_w^T \mathbf{M}_w \mathbf{W}^{-1/2}) \\ &= \alpha \times \text{tr}(\mathbf{C}_{dW}^+ \mathbf{P}_{\boldsymbol{\tau}_W}) \\ &= \alpha \times \text{tr}(\mathbf{C}_{dW}^+) . \end{aligned}$$

□

Incidentally, there is no restriction that the rows of \mathbf{M} be unique; repeats are allowed and some may even be the zero vector. Standard optimality coincides with $\mathbf{W} = \mathbf{I}$ and the condition in Lemma 3.3.1 reduces to $\mathbf{M}^T \mathbf{M} = \alpha \mathbf{P}_{\boldsymbol{\tau}}$ where the rows of \mathbf{M} are normalized, a result reported in Morgan and Stallings (2013).

3.3.1 Approximate and exact A_W -optimal CRDs

Having demonstrated the wide applicability of the A_W -criterion, we now focus on deriving optimal designs with general \mathbf{W} and \mathbf{W}_{Hw} described in the previous section. Assume the one-way ANOVA model for a completely randomized design

$$\mathbf{y} = \mathbf{X}_d \boldsymbol{\tau} + \mathbf{1}\mu + \mathbf{e} . \quad (3.7)$$

It is known that regardless of the CRD d , the variance for a contrast $\widehat{\mathbf{h}^T \boldsymbol{\tau}}$ is of the form $\sum_{i=1}^p \frac{h_i^2}{r_{di}}$, where $r_{di} > 0$ is the replication for treatment i . We first focus on finding *approximate* A_W -optimal CRDs, meaning the replications are allowed to be any positive real number, not just integers. The following result resembles the corollary in Section 8.8 in Pukelsheim (2006).

Theorem 3.3.2. *For any weight matrix with a factorization $\mathbf{W}_{Hw} = \mathbf{H}^T \mathbf{D}_w \mathbf{H} + \frac{1}{p} \mathbf{J}$ described previously, denote the i -th element of the j -th row of \mathbf{H} as h_{ij} . Then the approximate CRD d^* with replications*

$$r_{d^*i} = N \frac{\sqrt{w_i^*}}{\sum \sqrt{w_i^*}} \quad (3.8)$$

where $w_i^* = \sum_{j=1}^s w_j h_{ij}^2$ is A_W -optimal. Moreover, this design is optimal regardless of the chosen factorization.

Proof. By assumption, \mathbf{W} admits a factorization as in Corollary 3.3.1 so A_{dW} is proportional to the sum of the weighted variances of the s contrasts $\mathbf{H} \boldsymbol{\tau}$ with weights equal to the elements

of \mathbf{D}_w ,

$$\begin{aligned}
A_{dW} &\propto \sum_{j=1}^s w_j \text{Var}_d(\widehat{\mathbf{h}_j^T \boldsymbol{\tau}}) = \sum_{j=1}^s w_j \sum_{i=1}^p \frac{h_{ij}^2}{r_{di}} \\
&= \sum_{i=1}^p \frac{\sum_{j=1}^s w_j h_{ij}^2}{r_{di}} \\
&= \sum_{i=1}^p \frac{w_i^*}{r_{di}},
\end{aligned}$$

where $w_i^* = \sum_{j=1}^s w_j h_{ij}^2$. A straightforward application of Lagrange multipliers gives the minimum solution

$$\begin{aligned}
r_{di} &= N \frac{\sqrt{w_i^*}}{\sum_i \sqrt{w_i^*}} \\
&= N \frac{\sqrt{\sum_{j=1}^s w_j h_{ij}^2}}{\sum_i \sqrt{\sum_{j=1}^s w_j h_{ij}^2}}.
\end{aligned}$$

We have chosen one factorization that gives us an equivalent expression to A_{dW} . Corollary 3.3.1 says that choosing another factorization gives another equivalent expression to A_{dW} . Hence the A_W -optimal approximate design is invariant to the factorization chosen. \square

The invariance property in Theorem 3.3.2 is not surprising; all factorizations come from the same weight matrix \mathbf{W} , which determines the design ordering. Notice that the weights in (3.8) are comprised of the weights assigned, w_j , and the squared coefficients in the corresponding estimable function. If a treatment is not involved in $\mathbf{h}_j^T \boldsymbol{\tau}$, then $h_{ij}^2 = 0$ so that the corresponding w_j does not apply to that treatment. Hence the weight, w_i^* , for a given treatment balances both the desired w_j and the relative involvement of that treatment in that contrast, h_{ij} . The following corollaries show that equally-replicated CRDs can be A_W -optimal designs for some \mathbf{W} .

Corollary 3.3.2. *For the conditions in Theorem 3.3.2, if the diagonal elements of \mathbf{W}_{Hw} are equal then an equally-replicated CRD is A_W -optimal.*

Proof. If the diagonal elements of \mathbf{W}_{Hw} are equal then the diagonal elements of $\mathbf{H}^T \mathbf{D}_w \mathbf{H}$ are equal. But these are $w_i^* = \sum_{j=1}^s w_j h_{ij}^2$ and so $A_{dW} \propto A_d$. \square

We now offer the following corollary to Theorem 3.3.2 for when a cell means model may be preferred. Under this model, all linear combinations $\mathbf{h}^T \boldsymbol{\tau}$ are estimable and also have variance $\sum_{i=1}^p \frac{h_i^2}{r_{di}}$.

Corollary 3.3.3. *Suppose we have a CRD scenario with p treatments and assume the cell means model. For any weight matrix \mathbf{W} with diagonal elements w_{ii} the approximate design with replications*

$$r_{di} = N \frac{\sqrt{w_{ii}}}{\sum \sqrt{w_{ii}}}$$

is A_W -optimal.

We call $\sqrt{w_{ii}} / \sum \sqrt{w_{ii}}$ the *replication proportion* for treatment i . Approximate designs are only applicable when their replications are integers. However they can be very informative by hinting at what the exact optimal design should be. For example, they can identify groups of treatments that should be equally replicated and provide an overall picture of relative treatment replication. This will be demonstrated for the baseline parameterization in Section 3.4.

The following theorem demonstrates how exact A_W -optimal CRDs, assuming either the one-way ANOVA model (3.7) or cell means model, can be built up from an existing A_W -optimal design with less N .

Theorem 3.3.3. *Let $r_{i,N}^*$ be the replication numbers for an exact A_W -optimal CRD for p treatments with N runs and consider optimizing a CRD for p treatments with $N + 1$ runs.*

Suppose the optimization target is to minimize

$$A_W = \sum_{i=1}^p \frac{w_i}{r_{i,N+1}}$$

for some fixed weights w_i and replication $r_{i,N+1}$ for treatment i . Then there is a treatment \tilde{i} for which

$$r_{i,N+1}^* = \begin{cases} r_{i,N}^* + 1 & \text{if } i = \tilde{i}, \\ r_{i,N}^* & \text{if } i \neq \tilde{i} \end{cases}$$

are the replications for an A_W -optimal exact CRD for p treatment with $N + 1$ runs. That is, an A_W -optimal exact CRD for $N + 1$ adds one replication to one of the treatments in an A_W -optimal CRD for N runs.

Proof. The proof is done by induction. The smallest possible number of runs is p , and obviously the result is true for $N = p + 1$. Provided the result is true for given $N > p + 1$, we need to show it also holds for $N + 1$. Let d_N^* be an A_W -optimal design for N runs. Clearly an A_W -optimal design for $N + 1$ runs, d_{N+1}^* , must for some treatment, call it \tilde{i} , have $r_{\tilde{i},N+1}^* \geq r_{\tilde{i},N}^* + 1$. The argument proceeds in two cases.

Case 1. Suppose $r_{\tilde{i},N+1}^* = r_{\tilde{i},N}^* + 1$. Then $\sum_{i \neq \tilde{i}} r_{i,N}^* = \sum_{i \neq \tilde{i}} r_{i,N+1}^*$. Given this sum, both d_N^* and d_{N+1}^* must achieve the minimal value of $\sum_{i \neq \tilde{i}} w_i / r_i$. Thus (up to permutation among replication numbers for treatments assigned the same weight) the two designs assign the same replication numbers to all treatments except \tilde{i} . This is the claim of the theorem.

Case 2. Suppose $r_{\tilde{i},N+1}^* = r_{\tilde{i},N}^* + x$ for some integer $x \geq 2$. Then $\sum_{i \neq \tilde{i}} r_{i,N+1}^* = \sum_{i \neq \tilde{i}} r_{i,N}^* - (x - 1)$ and relative to d_N^* , the design d_{N+1}^* has been found by increasing the replication of treatment \tilde{i} by x , and decreasing the total replication for all other treatments by $x - 1$. This implies that adding x replicates to \tilde{i} and removing a total of $(x - 1)$ replicates from the other treatments decreases A_W more than adding one replicate to treatment \tilde{i} and maintaining

the same replication sum for the other $p - 1$ treatments. Now removing one replicate of \tilde{i} from each of these two possible modifications of d_N^* , this in turn implies that adding $x - 1$ replicates to \tilde{i} and removing a total of $(x - 1)$ replicates from the other treatments decreases A_W more than adding no replicates to treatment \tilde{i} and maintaining the same replication sum for the other $p - 1$ treatments. That is, d_N^* is improved by adding $x - 1$ replicates to treatment \tilde{i} and, in some fashion, removing $x - 1$ replicates from among the other $p - 1$ treatments. This contradicts the fact that d_N^* is A_W -optimal. \square

Given d_N^* , the exact design d_{N+1}^* is found as follows. For each treatment i calculate

$$\frac{w_i}{r_{i,N}^*} - \frac{w_i}{r_{i,N}^* + 1} = \frac{w_i}{r_{i,N}^*(r_{i,N}^* + 1)} \quad (3.9)$$

which is the decrease in A_W for d_N^* by increasing replication of treatment i by one unit. The optimal replication increase is for the treatment \tilde{i} which maximizes expression (3.9). For the cell means model and any weight matrix \mathbf{W} , w_i in Theorem 3.3.3 equals the i -th diagonal element of \mathbf{W} . For the one-way ANOVA model and weight matrix \mathbf{W}_{Hw} , the w_i in the theorem equals $\sum_{j=1}^s w_j h_{ij}^2$.

3.4 A_W -optimal CRDs for two-level OP and BP effects

For treatments comprised of m factors each at 2 levels, we defined \mathbf{H}_{OP} and \mathbf{H}_{BP} in Section 1.2.2 using a series of Kronecker products of 2×2 matrices, which are reparameterizations performed on $\boldsymbol{\tau}$ in the cell means model. The weight matrix corresponding to the 2^m OP effects is $\mathbf{H}_{OP}^T \mathbf{H}_{OP} = \mathbf{I}_{2^m}$ and so the OP transformation places identity weights on the treatment means. Hence optimality for all OP effects is equivalent to standard optimality for estimation of all treatment means.

A natural additional weighting for the OP effects would be to assign more weight to the main effects than the interactions through the weight matrix $\mathbf{W}_{OPw} = \mathbf{H}_{OP}^T \mathbf{D}_w \mathbf{H}_{OP}$ which has diagonal elements $w_{ii} = \sum_{j=1}^{2^m} w_j h_{ij}^2$. For \mathbf{H}_{OP} , $h_{ij}^2 = 2^{-m}$ for all h_{ij} so $w_{ii} = 2^{-m} \sum_{j=1}^{2^m} w_j$ is the same for all treatments. By Corollary 3.3.2 an equally-replicated CRD will be A_W -optimal for \mathbf{W}_{OPw} with any chosen diagonal weight matrix \mathbf{D}_w . While an interesting observation, this result assumes we are interested in estimating all orthogonal factorial effects, including the intercept. This does not coincide with most applications of factorial effects where higher order interactions are assumed negligible, a scenario that falls under the theory to be established in Chapter 5.

Finding A_W -optimal CRDs for efficient estimation of all 2^m BP effects is a much more interesting problem. The Kronecker product representation for \mathbf{H}_{BP} makes finding the corresponding weight matrix and its inverse straightforward. Let $\mathbf{A}^{\otimes m} = \mathbf{A} \otimes \cdots \otimes \mathbf{A}$, the Kronecker product of m copies of \mathbf{A} . Then we have

$$\begin{aligned} \mathbf{W}_{BP} &= \mathbf{H}_{BP}^T \mathbf{H}_{BP} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{\otimes m} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^{\otimes m} \\ &= \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}^{\otimes m} \\ \mathbf{W}_{BP}^{-1} &= \mathbf{H}_{BP}^{-1} \mathbf{H}_{BP}^{-1T} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\otimes m} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\otimes m} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{\otimes m}. \end{aligned}$$

Recall that the Hadamard product of two column vectors $\mathbf{v}_1 = (v_{1i})$ and $\mathbf{v}_2 = (v_{2i})$, denoted $\mathbf{v}_1 \odot \mathbf{v}_2$, is the column vector having the i -th row element $(v_{1i}v_{2i})$. The matrix \mathbf{H}_{BP}^{-1} is easily

shown to have columns equal to $\mathbf{1}_{2^m}$, \mathbf{A}_m , and all possible Hadamard products of at least two columns of \mathbf{A}_m . This is clear when one thinks of \mathbf{H}_{BP}^{-1} as the transformation matrix of the BP effect parameter vector $\boldsymbol{\theta}$ back into the treatment parameter vector $\boldsymbol{\tau}$, which is discussed in Lemma 1.2.1. Finally, let $\mathbf{A}_{m;h}$, $h = 2, \dots, m$, denote the $2^m \times \binom{m}{h}$ matrix of all Hadamard products involving h columns of \mathbf{A}_m . The order of the columns in $\mathbf{A}_{m;h}$ is not important for the work that follows. Being comprised of Hadamard products of columns with either 0 or 1 elements, each matrix $\mathbf{A}_{m;h}$ has elements equal to 0 or 1.

There clearly exists a column permutation of \mathbf{H}_{BP}^{-1} giving $\mathbf{H}_{BP}^{-1}\boldsymbol{\Pi} = (\mathbf{1}_{2^m} \mid \mathbf{A}_m \mid \mathbf{A}_{m;2} \mid \dots \mid \mathbf{A}_{m;m})$. We may then write $\mathbf{W}_{BP}^{-1} = \mathbf{H}_{BP}^{-1}\mathbf{H}_{BP}^{-1T} = \mathbf{H}_{BP}^{-1}\boldsymbol{\Pi}\boldsymbol{\Pi}^T\mathbf{H}_{BP}^{-1T}$ as

$$\mathbf{W}_{BP}^{-1} = \mathbf{1}\mathbf{1}^T + \mathbf{A}_m\mathbf{A}_m^T + \mathbf{A}_{m;2}\mathbf{A}_{m;2}^T + \dots + \mathbf{A}_{m;m-1}\mathbf{A}_{m;m-1}^T + \mathbf{A}_{m;m}\mathbf{A}_{m;m}^T, \quad (3.10)$$

which will be a useful representation when finding the permutation class of $\boldsymbol{\Pi}_W$ corresponding to \mathbf{W}_{BP} .

With the cell means model, the inverse of the diagonal elements of \mathbf{W}_{BP}^{-1} are the weights assigned to the treatment means. It is straightforward to show that if a treatment mean has heft h then the corresponding weight for that treatment mean is 2^{-h} . The intercept, which has heft 0, has the largest weight with 1, and the treatment with all factors not set at baseline has the smallest weight of 2^{-m} .

A similar Kronecker representation exists for the normalized BP effects by multiplying the bottom row of \mathbf{Z}_{BP} by $2^{-1/2}$. It is easily shown that the corresponding weight matrix for the normalized BP effects, say \mathbf{W}_{BP^*} , assigns weight 3^{-h} to treatments of heft h , which is $(\frac{2}{3})^h$ less than for those under the original BP model. One should then anticipate that optimal CRDs for all BP effects will depend on whether we normalize the effects or not.

We now focus on finding A_W -optimal approximate CRDs for estimation of non-normalized

and normalized BP effects for the cell means model. We begin by deriving general forms for the diagonal elements of \mathbf{W}_{BP} and \mathbf{W}_{BP^*} , which are needed to find the optimal replications given by Corollary 3.3.3. It is easily shown that the diagonal element corresponding to a treatment of heft h is 2^{m-h} and $\frac{3^{m-h}}{2^m}$ for \mathbf{W}_{BP} and \mathbf{W}_{BP^*} , respectively. Note these are not the same weights assigned to the treatments by the weight matrix.

In order to calculate the optimal replications in Corollary 3.3.3 we need to find the sum of the square roots of the diagonal elements. There are $\binom{m}{h}$ different treatment means with heft h , each having diagonal element either 2^{m-h} or $\frac{3^{m-h}}{2^m}$, depending on whether we are interested in estimating the non-normalized or normalized BP effects. For the non-normalized BP effects, the sum of the square roots of the diagonal elements will then be

$$\sum_{h=0}^m \binom{m}{h} \sqrt{2^{m-h}} = (1 + \sqrt{2})^m ,$$

which may be verified using the binomial expansion theorem. It follows that the replication proportion for a treatment of heft h equals $\sqrt{2^{m-h}}/(1 + \sqrt{2})^m$. For the normalized BP effects, the replication proportion for a treatment of heft h are similarly shown to be $\sqrt{3^{m-h}}/(1 + \sqrt{3})^m$. It follows that the approximate A_W -optimal design will replicate the baseline mean the most and the treatment mean with all factors not set at baseline the least. The replication proportions for estimation of the non-normalized BP effects are broken down by heft for $m = 3$ to $m = 8$ in Table 3.2. The replication proportions for estimation of the normalized BP effects are similar, and so not shown here, although slightly greater emphasis is placed on replication of treatments of lower heft.

As the replication proportions are irrational there can never be an N such that the approximate design yields an exact design. However, the approximate theory sheds light onto the optimal forms of exact designs, such as the observation that the baseline treatment will likely be replicated the most. We also see that treatments having the same heft should be

Table 3.2: Replication proportions for treatments of a given heft for A -optimal approximate CRD with respect to estimation of all non-normalized BP effects.

Heft	m					
	3	4	5	6	7	8
0	0.201	0.118	0.069	0.040	0.024	0.014
1	0.142	0.083	0.049	0.029	0.017	0.010
2	0.101	0.059	0.034	0.020	0.012	0.007
3	0.071	0.042	0.024	0.014	0.008	0.005
4	—	0.029	0.017	0.010	0.006	0.003
5	—	—	0.012	0.007	0.004	0.002
6	—	—	—	0.005	0.003	0.002
7	—	—	—	—	0.002	0.001
8	—	—	—	—	—	0.001

replicated as equally as possible. We may think of the design in terms of the $N \times m$ matrix $\mathbf{V}_d = \mathbf{X}_d \mathbf{A}_m$ where \mathbf{A}_m , which was defined in Section 1.2.2, has rows organized under the revlex order. Then equal replication of treatments of similar heft in the approximate A_W -optimal CRD implies that an exact A_W -optimal CRD \mathbf{V}_d will likely be an $N \times m$ BA of strength m or at least “close” to one.

We now identify the class of permutations that satisfy $\mathbf{\Pi} \mathbf{W}_{BP} \mathbf{\Pi}^T = \mathbf{W}_{BP}$ and $\mathbf{\Pi} \mathbf{W}_{BP^*} \mathbf{\Pi}^T = \mathbf{W}_{BP^*}$ and show that the information matrix of the BP effects under a BA of strength m is invariant to such permutations. We begin by first proving the following lemma, which will be needed later.

Lemma 3.4.1. *For a given m , the columns of \mathbf{W}_{BP}^{-1} indexed by $\{2^h + 1 : h = 0, 1, \dots, m-1\}$ make up a $2^m \times m$ array equal to*

$$\tilde{\mathbf{A}}_m = \mathbf{A}_m + \mathbf{J}$$

where \mathbf{A}_m is the array of treatment indices under the revlex ordering.

Proof. The proof is done by induction. For $m = 1$ we have

$$\mathbf{W}_{BP_1}^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = (\mathbf{1}_2 \mid \tilde{\mathbf{A}}_1) ,$$

and clearly column 2 gives the result. For any $m > 1$, $\tilde{\mathbf{A}}_m$ may be written as

$$\tilde{\mathbf{A}}_m = \mathbf{A}_m + \mathbf{J} = \left(\begin{array}{c|c} \mathbf{A}_{m-1} & \mathbf{0}_{m-1} \\ \hline \mathbf{A}_{m-1} & \mathbf{1}_{m-1} \end{array} \right) + \mathbf{J} = \left(\begin{array}{c|c} \tilde{\mathbf{A}}_{m-1} & \mathbf{1}_{m-1} \\ \hline \tilde{\mathbf{A}}_{m-1} & 2\mathbf{1}_{m-1} \end{array} \right) .$$

For $m = 2$ we have

$$\begin{aligned} \mathbf{W}_{BP_2}^{-1} &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \otimes (\mathbf{1}_2 \mid \tilde{\mathbf{A}}_1) \\ &= \left(\begin{array}{cc|cc} \mathbf{1}_2 & \tilde{\mathbf{A}}_1 & \mathbf{1}_2 & \tilde{\mathbf{A}}_1 \\ \hline \mathbf{1}_2 & \tilde{\mathbf{A}}_1 & 2\mathbf{1}_2 & 2\tilde{\mathbf{A}}_1 \end{array} \right) , \end{aligned}$$

and columns 2 and 3 clearly give the required $\tilde{\mathbf{A}}_2$. Now assume the result holds for $m = t$ factors. It follows then that for $m = t + 1$ factors

$$\begin{aligned} \mathbf{W}_{BP_{t+1}}^{-1} &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \otimes \mathbf{W}_{BP_t}^{-1} \\ &= \left(\begin{array}{c|c} \mathbf{W}_{BP_t}^{-1} & \mathbf{W}_{BP_t}^{-1} \\ \hline \mathbf{W}_{BP_t}^{-1} & 2\mathbf{W}_{BP_t}^{-1} \end{array} \right) . \end{aligned}$$

By the induction hypothesis, the set of columns $\{2, 3, 5, \dots, 2^{t-1} + 1\}$ produce the first t columns of $\tilde{\mathbf{A}}_{t+1}$ and column $2^t + 1$, which is the first column of $(\mathbf{W}_{BP_t}^{-1} \mid 2\mathbf{W}_{BP_t}^{-1})^T$, is clearly equal to $(\mathbf{1}_t^T \mid 2\mathbf{1}_t^T)^T$, giving the result. \square

We say a $2^m \times 2^m$ permutation matrix $\mathbf{\Pi}$ is a *factor permutation* if there exists an $m \times m$ permutation matrix, $\mathbf{\Pi}_m$, such that $\mathbf{\Pi}\mathbf{A}_m = \mathbf{A}_m\mathbf{\Pi}_m$. That is, if the row permutation $\mathbf{\Pi}$ of \mathbf{A}_m is equivalent to a column permutation of \mathbf{A}_m , we say $\mathbf{\Pi}$ is a factor permutation. By definition, a factor permutation of a BA of strength m again produces a BA of strength m with the same index set.

Lemma 3.4.2. *If $\mathbf{P}_\tau = \mathbf{I}_p$, the permutation class Π_W for \mathbf{W}_{BP} equals that for \mathbf{W}_{BP^*} and is comprised solely of the $m!$ factor permutations.*

Proof. We are interested in finding those permutation matrices that satisfy $\mathbf{\Pi}\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T = \mathbf{W}_{BP}^{-1}$ and $\mathbf{\Pi}\mathbf{W}_{BP^*}^{-1}\mathbf{\Pi}^T = \mathbf{W}_{BP^*}^{-1}$. If two elements in \mathbf{W}_{BP}^{-1} are equal they will also be equal in $\mathbf{W}_{BP^*}^{-1}$ and if two elements are unequal in one they will also be unequal in the other. Hence the class Π_W will be the same for both weight matrices.

Recall the representation of \mathbf{W}_{BP}^{-1} in (3.10). Let $\mathbf{\Pi}$ be any factor permutation: $\mathbf{\Pi}\mathbf{A}_m = \mathbf{A}_m\mathbf{\Pi}_m$ for some $m \times m$ permutation matrix $\mathbf{\Pi}_m$. Then $\mathbf{\Pi}\mathbf{1}\mathbf{1}^T\mathbf{\Pi}^T = \mathbf{1}\mathbf{1}^T$ and $\mathbf{\Pi}\mathbf{A}_m\mathbf{A}_m^T\mathbf{\Pi}^T = \mathbf{A}_m\mathbf{\Pi}_m\mathbf{\Pi}_m^T\mathbf{A}_m^T = \mathbf{A}_m\mathbf{A}_m^T$. Each column of $\mathbf{A}_{m;h}$ for $h = 2, \dots, m$ is comprised of some order of Hadamard products of the columns of \mathbf{A}_m , so it follows that the columns of $\mathbf{\Pi}\mathbf{A}_{m;h}$ may be written as the same order of Hadamard products of the columns of $\mathbf{\Pi}\mathbf{A}_m$. But $\mathbf{\Pi}\mathbf{A}_m = \mathbf{A}_m\mathbf{\Pi}_m$, a column permutation of \mathbf{A}_m , so $\mathbf{\Pi}\mathbf{A}_{m;h}$ may be written as a column permutation of $\mathbf{A}_{m;h}$, say $\mathbf{A}_{m;h}\mathbf{\Pi}_{\binom{m}{h}}$. Hence $\mathbf{\Pi}\mathbf{A}_{m;h}\mathbf{A}_{m;h}^T\mathbf{\Pi}^T = \mathbf{A}_{m;h}\mathbf{\Pi}_{\binom{m}{h}}\mathbf{\Pi}_{\binom{m}{h}}^T\mathbf{A}_{m;h}^T = \mathbf{A}_{m;h}\mathbf{A}_{m;h}^T$ for all $h = 2, \dots, m$, giving $\mathbf{\Pi}\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T = \mathbf{W}_{BP}^{-1}$.

Any permutation matrix satisfying $\mathbf{\Pi}\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T = \mathbf{W}_{BP}^{-1}$ must preserve the diagonal elements of \mathbf{W}_{BP}^{-1} . If two diagonal elements of \mathbf{W}_{BP}^{-1} are the same then the corresponding rows in \mathbf{A}_m must have the same heft. Hence $\mathbf{\Pi}\mathbf{A}_m$ may only permute rows that have the same heft. The index of the m columns in Lemma 3.4.1 is easily shown to index the m heft 1 rows of \mathbf{A}_m . Hence if $\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T$ permutes a column in this index, it must do so with another column that

makes up $\tilde{\mathbf{A}}_m$, as given in Lemma 3.4.1. Denote the permutation matrix of these m columns of \mathbf{W}_{BP}^{-1} corresponding to $\mathbf{\Pi}^T$ as $\mathbf{\Pi}_m$.

If $\mathbf{\Pi}\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T = \mathbf{W}_{BP}^{-1}$ then $\mathbf{W}_{BP}^{-1}\mathbf{\Pi}^T = \mathbf{\Pi}^T\mathbf{W}_{BP}^{-1}$ and clearly their $2^m \times m$ submatrices composed of the columns in Lemma 3.4.1 must be equal. That is

$$\tilde{\mathbf{A}}_m\mathbf{\Pi}_m = \mathbf{\Pi}^T\tilde{\mathbf{A}}_m \Leftrightarrow (\mathbf{A}_m + \mathbf{J})\mathbf{\Pi}_m = \mathbf{\Pi}^T(\mathbf{A}_m + \mathbf{J}) .$$

As \mathbf{J} is invariant to any row or column permutation, we have $\mathbf{A}_m\mathbf{\Pi}_m = \mathbf{\Pi}^T\mathbf{A}_m$, so $\mathbf{\Pi}$ must be a factor permutation. \square

Corollary 3.4.1. *For the cell means model, a design, d , has an information matrix invariant to all permutations in Π_W for \mathbf{W}_{BP} and \mathbf{W}_{BP^*} if and only if $\mathbf{V}_d = \mathbf{X}_d\mathbf{A}_m$ is an $N \times m$ BA of strength m .*

Proof. Information matrices under the cell means model are diagonal matrices of replications for each treatment, \mathbf{R}_d . For \mathbf{R}_d to be invariant to factor permutations, the replications for treatments of a given heft must be equal. Hence \mathbf{V}_d must be a BA of strength m . \square

Lemma 3.4.2 and Corollary 3.4.1 indicate that some form of a BA of strength m may be weighted optimal, or at least reasonably efficient, for estimation of non-normalized and normalized BP effects. However, the best BA will likely depend on the chosen weighted criterion. For the A_W -criterion, the replication proportions in Table 3.2 give insight into the relationship between the BA indices, assuming one is optimal. That is, we anticipate such a design will have larger BA indices for rows of lesser heft.

Let $w_{j_1 \dots j_m}$ denote the diagonal element of \mathbf{W}_{BP} corresponding to the treatment mean with index $j_1 \dots j_m$ and let $r_{dj_1 \dots j_m}$ denote the number of replications assigned to that treatment, so $\mathbf{R}_d = \text{Diag}(r_{d00\dots 0}, r_{d10\dots 0}, \dots, r_{d11\dots 1})$. Recall that J is the set of all indices $j_1 \dots j_m$.

For \mathbf{W}_{BP} , $\mathbf{H}_{BP}\boldsymbol{\tau}$ make up 2^m weighted orthogonal functions each assigned weight 1, so by Lemma 2.3.2 the A_{dW} expression is proportional to

$$\begin{aligned} A_{dW} &\propto \text{tr}(\mathbf{H}_{BP}\mathbf{R}_d^{-1}\mathbf{H}_{BP}^T) = \text{tr}(\mathbf{R}_d^{-1}\mathbf{W}_{BP}) \\ &= \sum_{j_1 \dots j_m \in J} \frac{w_{j_1 \dots j_m}}{r_{dj_1 \dots j_m}}. \end{aligned} \quad (3.11)$$

Applying Theorem 3.3.3 and the selection rule (3.9) explained at the end of the previous section to (3.11) proves the optimality of BAs for estimation of BP effects, regardless of whether they are normalized, which we now show.

Optimization for the BP effects, whether or not they are normalized, is equivalent to a weighted optimization in which every treatment mean of the same heft is assigned the same weight, say w_h . That is, $w_{j_1 \dots j_m} = w_h$ where $h = \sum j_l$. Suppose the A_W -optimal design for N runs d_N^* is a strength m BA, which is true for $N = 2^m$, the saturated case. This implies that treatments of the same heft will also be replicated the same number of times, say $r_{h,N}^*$. This assumption simplifies expression (3.11), so that in order to find the A_W -optimal design for $N + 1$ runs, we need only to maximize

$$\frac{w_h}{r_{h,N}^*(r_{h,N}^* + 1)} \quad (3.12)$$

with respect to h , rather than every possible treatment combination. That is, we need only to find the maximum of $m + 1$ different values. For whatever treatment heft h that maximizes (3.12), say h' , it does not matter which treatment of that heft we choose to increase the replication for, it will always produce an A_W -optimal design for $N + 1$ runs. Say we decide to replicate treatment i' having heft h' . Suppose this d_{N+1}^* is not a BA (we did not replicate the baseline treatment or the treatment with heft m). When we want to find the A_W -optimal

design for $N + 2$ runs we would not replicate treatment i' again, since

$$\frac{w_{h'}}{r_{i',N+1}^*(r_{i',N+1}^* + 1)} = \frac{w_{h'}}{(r_{i',N}^* + 1)(r_{i',N}^* + 2)} < \frac{w_{h'}}{r_{h',N}^*(r_{h',N}^* + 1)} = \frac{w_{h'}}{r_{h',N+1}^*(r_{h',N+1}^* + 1)} ,$$

which would be the decrease of $A_{d_{N+1}W}$ by replicating another treatment of heft h' . In fact, we would clearly want to replicate another treatment having heft h' as it is already known to produce a greater decrease than the other treatments. This continues until we have increased replication for every treatment having heft h' by 1, giving another BA.

To demonstrate this recursive relationship, we used Theorem 3.3.3 to find exact A -optimal designs for estimation of non-normalized and normalized BP effects for $m = 3$ factors and increasing run size. The same procedure was used to find the designs shown in Tables 3.3 and 3.4, with the necessary modification of weights. Both scenarios begin with the saturated design where $N = 8$, which must replicate each treatment once in order to estimate every BP effect. Overall we see the expected hierarchical structure of treatment replications corresponding to hefts proposed from the approximate design theory. We also see how designs that are not BAs for a given N are building up to a BA by adding replications to treatments of the same heft. The A -optimal designs for the normalized BP effects tend to replicate the baseline treatment more than for the non-normalized designs but in some cases the designs for a given N coincide, such as for $N = 8, 9, 13$, and 20 runs. As N increases, we anticipate the A -optimal designs for non-normalized and normalized BP effects to deviate slightly, based on the relative replications from the approximate design theory.

Tables 3.5 and 3.6 show the A -optimal CRDs that are BAs for non-normalized and normalized BP effects, respectively, for $m = 3$ to 5 factors and a given number of runs. The displayed run order depends on the number of factors and whether the effects are normalized. The replications are displayed as the BA index set for treatments of different hefts. Any two adjacent designs in the tables are nearly identical except the design on the right increases

Table 3.3: A -optimal CRDs for estimation of all non-normalized BP effects for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.

Treatment Index	N											
	8	9	10	11	12	13	14	15	16	17	...	20
000	1	2	2	2	2	3	3	3	3	4		4
100	1	1	2	2	2	2	2	2	2	2		3
010	1	1	1	2	2	2	2	2	2	2		3
001	1	1	1	1	2	2	2	2	2	2		3
110	1	1	1	1	1	1	2	2	2	2		2
101	1	1	1	1	1	1	1	2	2	2		2
011	1	1	1	1	1	1	1	1	2	2		2
111	1	1	1	1	1	1	1	1	1	1		1
BA	*	*			*	*			*	*		*

Table 3.4: A -optimal CRDs for estimation of all normalized BP effects for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.

Treatment Index	N											
	8	9	10	11	12	13	14	15	16	17	...	20
000	1	2	3	3	3	3	4	4	4	4		4
100	1	1	1	2	2	2	2	3	3	3		3
010	1	1	1	1	2	2	2	2	3	3		3
001	1	1	1	1	1	2	2	2	2	3		3
110	1	1	1	1	1	1	1	1	1	1		2
101	1	1	1	1	1	1	1	1	1	1		2
011	1	1	1	1	1	1	1	1	1	1		2
111	1	1	1	1	1	1	1	1	1	1		1
BA	*	*	*			*	*			*		*

replications for all treatments of some heft, say h' , by one. If the left design is for N^* then the right design is for $N^* + \binom{m}{h'}$ runs, where $\binom{m}{h'}$ is the number of treatments of heft h' . It follows that an A -optimal CRD for $N^* + x$ runs where $1 \leq x \leq \binom{m}{h'}$ adds one replicate to each treatment in any subset of x treatments of heft h' .

Across the m and N considered, there is a consistent pattern of which heft group is replicated with increasing N . For example, for both the non-normalized and normalized BP models,

the indices for the A -optimal BAs for $m = 3$ almost always equal the A -optimal BA indices for treatments of heft 0 to 3 for $m = 4$ and $m = 5$ factors. Furthermore, for the N considered, the heft 4 treatment replications for $m = 4$ equal the heft 4 treatment replications for $m = 5$. This equivalence does not hold for the $N = 32$ design for $m = 3$ factors and $N = 50$ design for $m = 4$ factors, implying that this equivalence will likely not hold for large N . The A -optimal CRDs are rarely the same for the non-normalized and normalized BP models, although they are similar for small N . As expected, we clearly see greater importance being placed on treatments of smaller heft for the normalized BP model as N increases.

Table 3.5: BA heft replications for A -optimal designs for estimation of all non-normalized BP effects for $m = 3$ to 5 factors with increasing N .

m	Treatment Heft	N												
		8	9	12	13	16	17	20	21	22	25	28	29	32
3	0	1	2	2	3	3	4	4	4	5	5	5	6	6
	1	1	1	2	2	2	2	3	3	3	4	4	4	5
	2	1	1	1	1	2	2	2	2	2	2	3	3	3
	3	1	1	1	1	1	1	1	2	2	2	2	2	2
	Treatment Heft	N												
		16	17	21	22	28	29	33	37	38	42	48	49	50
4	0	1	2	2	3	3	4	4	4	5	5	5	6	6
	1	1	1	2	2	2	2	3	3	3	4	4	4	4
	2	1	1	1	1	2	2	2	2	2	2	3	3	3
	3	1	1	1	1	1	1	1	2	2	2	2	2	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	2
	Treatment Heft	N												
		32	33	38	39	49	50	55	65	66	71	81	82	87
5	0	1	2	2	3	3	4	4	4	5	5	5	6	6
	1	1	1	2	2	2	2	3	3	3	4	4	4	4
	2	1	1	1	1	2	2	2	2	2	2	3	3	3
	3	1	1	1	1	1	1	1	2	2	2	2	2	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	2
	5	1	1	1	1	1	1	1	1	1	1	1	1	1

Although the examples in this section exemplify the process of creating a weight matrix

Table 3.6: BA heft replications for A -optimal designs for estimation of all normalized BP effects for $m = 3$ to 5 factors with increasing N .

m	Treatment Heft	N												
		8	9	10	13	14	17	20	21	22	25	26	29	30
3	0	1	2	3	3	4	4	4	5	6	6	7	7	7
	1	1	1	1	2	2	3	3	3	3	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	2	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	1	2
	Treatment Heft	N												
		16	17	18	22	23	27	33	34	35	39	40	46	50
4	0	1	2	3	3	4	4	4	5	6	6	7	7	7
	1	1	1	1	2	2	3	3	3	3	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	2	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	1	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	1
	Treatment Heft	N												
		32	33	34	39	40	45	55	56	57	62	63	73	83
5	0	1	2	3	3	4	4	4	5	6	6	7	7	7
	1	1	1	1	2	2	3	3	3	3	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	2	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	1	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1	1	1	1	1	1

and finding weighted optimal designs, they do not necessarily reflect the typical approach for estimation of OP and BP effects. Oftentimes the intercepts in both models are not necessarily of interest and interactions involving more than 2 factors are often assumed to be negligible. In chapter 4 we detail approaches to searching for weighted optimal designs in SAS PROC OPTEX and investigate the impact of setting weights arbitrarily close to zero on weighted information matrices. In chapter 5 we investigate the effects on the linear model when the estimation space is reduced to a smaller subset and use that theory to find A -optimal CRDs for estimation of BP effects under the assumption that higher order interactions are negligible.

Chapter 4

Weighted optimality criteria and design search algorithms

Optimal designs with respect to standard optimality criteria often appeal to our intuition regarding desirable design properties such as equal replication and variance balance. Kiefer's approach to design optimization provided a mathematical justification for why well-known designs should be preferred. However, his universal optimality theorem does not shed immediate insight into how to arrive at universally optimal designs; if a design cannot be found that satisfies his sufficient conditions, the theorem is of limited use. Even if we can identify the desired type of design, such as a balanced incomplete block design, it does not mean that one exists or is easily constructed. In these cases, analytical approaches to find optimal designs are intriguing but may become quite complex. Instead, researchers often rely on design search algorithms to seek out optimal, or at least highly efficient, designs with respect to some criterion.

We begin this chapter with a brief overview of design search algorithms, focusing on the general procedures commonly employed for approximate and exact designs, including se-

quential and exchange point algorithms. We will focus on algorithms that require a set of user-specified candidate points for the design matrix and describe how optimal designs are searched for given the candidate points and desired number of runs, N . A straightforward modification of these candidate points is easily shown to evaluate designs in terms of their weighted information matrices. These algorithms, however, commonly assume the information matrix has full rank which does not hold for all designs. We show how such algorithms may be “tricked” into evaluating rank-deficient designs through a limiting weight approach, where small weight is placed on model parameters that are not the target of estimation. These techniques are used to find A -optimal designs that estimate all BP effects adjusted for the intercept, and A -optimal incomplete block designs for comparative experiments.

4.1 Overview and modification of design search algorithms

Design search algorithms we consider begin with an assumed model, often linear, having a design matrix \mathbf{X}_d for some target parameters $\boldsymbol{\tau}$. The user specifies the total number of runs, N , and then chooses a set of candidate points, which are possible rows of \mathbf{X}_d . Continuous experimental variables are usually assumed to be centered and scaled so their values range between -1 and 1 . For such variables, there are an infinite number of candidate points that could be specified so an equally-spaced grid is often selected to represent the entire regression range. Design search algorithms systematically run through subsets of the points of size N , including replicated points, in order to find a design matrix \mathbf{X}_d that minimizes a chosen criterion. The criteria are often standard optimality criteria described in this dissertation, but other forms of criteria may be used as well. Standard design search algorithms assume all parameters in the model need to be estimated, so \mathbf{X}_d must have full column rank and

$\mathbf{C}_d = \mathbf{X}_d^T \mathbf{X}_d$. Modifications are then put in place to include block effects, but the author has not seen an algorithm in commercial software that operates for general information matrices of the form $\mathbf{C}_d = \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L) \mathbf{X}_d$.

How the search is carried out depends on whether we are interested in finding an optimal approximate or exact design. For approximate designs, the i -th candidate point, \mathbf{x}_i^T , is assumed to have some replication proportion, p_i , $0 \leq p_i \leq 1$ so that $\sum_i p_i = 1$. A general expression for the desired criteria, which is a function of $\mathbf{C}_d = \mathbf{X}_d^T \mathbf{X}_d = N \sum_i p_i \mathbf{x}_i \mathbf{x}_i^T$, may be derived and optimized with respect to the proportions p_i under the constraint that they must sum to 1. If there is no known closed-form expression, a steepest descent, sequential procedure is commonly used, as detailed in Chapter 9 of Atkinson et al. (2007). An approximate design approach will be taken in Section 5.2, but here we are more interested in finding optimal exact designs.

Finding exact optimal designs is more practical, but much harder since we are optimizing over a discrete set of candidate points rather than continuous replications. To begin, the user starts with an initial exact design with N_0 runs which may be less than, greater than, or equal to N . If less than N , candidate points that most improve the initial criterion value are added to the initial design until an N run design is reached. Similarly, if $N_0 > N$, candidate points that least affect the initial criterion are removed until an N run design is reached. Once the initial design is modified so that we have an N run design, a point within the design, says \mathbf{x}_k^T , is possibly exchanged for another point, \mathbf{x}_l^T . The exchange should be made between points that leads to the largest decrease in the chosen criterion. Points are exchanged until some stopping rule is reached, such as a maximum number of iterations or a minimal change occurs after points are exchanged.

In general, the algorithms for finding optimal exact designs are often slow and computationally inefficient. This is especially true if one must invert \mathbf{C}_d after every point exchange in

order to calculate the chosen criterion or if there are many possible point exchanges to consider. Furthermore, there is no guarantee that a global optimal design is reached as standard optimality criteria fail to be convex over the discrete domain of information matrices. For some criteria, the resulting change in the criterion value by adding, removing, or exchanging points has a straightforward expression, not requiring multiple inversions of the information matrix. This is true of the D -criterion, which appears frequently in the design literature; see Fedorov, Studden, and Klimko (1972), Mitchell (1974), Welch (1982), Cook and Nachtsheim (1980, 1989), and DuMouchel and Jones (1994). It would however be unwise to choose the D -criterion for design optimization simply because of its computational convenience. Its scale-invariance property is often suitable for regression models, but the A - and E -criteria can be more appropriate for categorical predictor variables. Moreover, we are not interested in the D_W -criterion in this dissertation for reasons shown earlier and so will seek a procedure to search for either A_W - or E_W -optimal designs.

As weighted optimality criteria are standard optimality criteria that evaluate weighted information matrices, we need only modify standard design search algorithms so that they optimize \mathbf{C}_{dW} . The necessary modification is straightforward: instead of inputting the candidate points, say as rows in \mathbf{X}_c , we input the weighted candidate points $\mathbf{X}_{cW} = \mathbf{X}_c \mathbf{W}^{-1/2}$. The search algorithm then proceeds as usual until it outputs the determined set of optimal weighted runs, \mathbf{X}_{d^*W} , which we transform into the optimal design, $\mathbf{X}_{d^*} = \mathbf{X}_{d^*W} \mathbf{W}^{1/2}$.

4.2 Adjusted information matrices and limiting weights

For the cell means model, the set of candidate points is \mathbf{I}_p , the $p \times p$ identity matrix, so the weighted candidate points are simply the rows of $\mathbf{W}^{-1/2}$. The same would be true for one-way ANOVA models with p treatments if the search algorithm were able to evaluate

information matrices from general linear models, $\mathbf{C}_d = \mathbf{X}_d^T(\mathbf{I} - \mathbf{P}_L)\mathbf{X}_d$. A simple, but computationally intensive algorithm could easily be programmed to accomplish this, but we are more interested in modifying existing software built on sophisticated computational techniques that speed up design selection.

The information matrix for $\boldsymbol{\tau}$ in model (1.1) adjusts for $\boldsymbol{\beta}$, the nuisance parameters. That is, we acknowledge that the parameters in $\boldsymbol{\beta}$ may explain variation but have no direct interest in their estimation. This perspective can be operationalized in a weighted context by thinking of functions of nuisance parameters as estimable effects that we assign very small weight to. So small, in fact, that their variances play no role in how designs are evaluated. The same idea can be employed with functions of the targeted parameters $\boldsymbol{\tau}$ that are of little or no interest, which is where we begin.

For a given general linear model and class of designs \mathcal{D} having information matrices with row space equal to \mathcal{E} , suppose we desire evaluation of designs in terms of a subspace of \mathcal{E} , say \mathcal{E}_1 , of dimension $k_1 < k$. That is, we acknowledge that all estimable functions in \mathcal{E} are in the model, but are primarily concerned with weighted evaluation of functions in \mathcal{E}_1 . We focus on the case where we have k_1 linearly independent estimable functions that span \mathcal{E}_1 to which we wish to assign weights w_1, \dots, w_{k_1} . Let \mathbf{H}_1 be the matrix with the coefficient vectors of these estimable functions along with $p - k$ orthonormal rows that span that null space of \mathbf{C}_d . Next let \mathcal{E}_2 be the subspace of \mathcal{E} having dimension k_2 ($k_1 + k_2 = k$) such that $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2$ and every vector in \mathcal{E}_2 is orthogonal to those in \mathcal{E}_1 . Choose any set of k_2 orthonormal estimable functions $\mathbf{H}_2\boldsymbol{\tau}$ that span \mathcal{E}_2 , which again we have minimal interest in estimating but are still assumed to be in the model.

If $\mathbf{H}^T = (\mathbf{H}_1^T | \mathbf{H}_2^T)$ then \mathbf{H} is invertible, $\mathbf{H}_1\mathbf{H}_2^T = 0$, and $\mathbf{H}^{-1} = (\mathbf{H}_1^T(\mathbf{H}_1\mathbf{H}_1^T)^{-1} | \mathbf{H}_2^T(\mathbf{H}_2\mathbf{H}_2^T)^{-1})$.

The matrix \mathbf{H} considered here is different from the general \mathbf{H} in Chapter 3, whose rows were made up of only coefficient vectors for estimable functions. By rewriting $\mathbf{X}_d\boldsymbol{\tau} =$

$\mathbf{X}_d \mathbf{H}^{-1} \mathbf{H} \boldsymbol{\tau}$ in the general model (1.1), the information matrix for $\mathbf{H} \boldsymbol{\tau}$ is easily shown to be $\mathbf{C}_{dH} = \mathbf{H}^{-1T} \mathbf{C}_d \mathbf{H}^{-1}$. Note that the non-estimable functions in \mathbf{H} will have no impact on the evaluation of \mathbf{C}_{dH} , as they lie in the null space of \mathbf{C}_d . Finally, the information matrix for $\mathbf{H}_1 \boldsymbol{\tau}$ adjusted for $\mathbf{H}_2 \boldsymbol{\tau}$, denoted $\mathbf{C}_{dH1(2)}$, is

$$\begin{aligned} \mathbf{C}_{dH1(2)} &= \mathbf{H}_1 \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L) [\mathbf{I} - (\mathbf{I} - \mathbf{P}_L) \mathbf{X}_d \mathbf{H}_2^T (\mathbf{H}_2 \mathbf{C}_d \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{X}_d^T (\mathbf{I} - \mathbf{P}_L)] (\mathbf{I} - \mathbf{P}_L) \mathbf{X}_d \mathbf{H}_1^T \\ &= \mathbf{H}_1 [\mathbf{C}_d - \mathbf{C}_d \mathbf{H}_2^T (\mathbf{H}_2 \mathbf{C}_d \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{C}_d] \mathbf{H}_1^T . \end{aligned} \quad (4.1)$$

Suppose that we wish to further weight the k_1 estimable functions in $\mathbf{H}_1 \boldsymbol{\tau}$ through a weight matrix, \mathbf{W}_1 . Then the weighted information matrix for which design optimization is to be performed will be $\mathbf{W}_1^{-1/2} \mathbf{C}_{dH1(2)} \mathbf{W}_1^{-1/2}$. We now show that design optimization with respect to the eigenvalues of this matrix is asymptotically equivalent to weighted estimation of $\boldsymbol{\tau}$ as we assign weights arbitrarily close to zero to $\mathbf{H}_2 \boldsymbol{\tau}$.

Theorem 4.2.1. *For the scenario described above, eigenvalue-based weighted optimization for the subset $\mathbf{H}_1 \boldsymbol{\tau}$ with weight matrix \mathbf{W}_1 is asymptotically equivalent to eigenvalue-based weighted optimization of $\boldsymbol{\tau}$ with weight matrix*

$$\mathbf{W}_{H_1} = \mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1 + w \mathbf{H}_2^T \mathbf{H}_2 , \quad (4.2)$$

as $w \rightarrow 0$.

Proof. From the discussion in Section 3.1, the positive eigenvalues of $\mathbf{C}_{dW} = \mathbf{W}_{H_1}^{-1/2} \mathbf{C}_d \mathbf{W}_{H_1}^{-1/2}$ equal those of $\tilde{\mathbf{C}}_{dW} = \tilde{\mathbf{W}}^{-1/2} \mathbf{C}_{dH} \tilde{\mathbf{W}}^{-1/2}$ where $\mathbf{C}_{dH} = \mathbf{H}^{-1T} \mathbf{C}_d \mathbf{H}^{-1}$ and

$$\tilde{\mathbf{W}} = \left(\begin{array}{c|c} \mathbf{W}_1 & \mathbf{0} \\ \hline \mathbf{0} & w \mathbf{I}_{k_2} \end{array} \right) .$$

This matrix may be partitioned as

$$\tilde{\mathbf{C}}_{dW} = \left(\begin{array}{c|c} \mathbf{W}_1^{-1/2} \mathbf{H}_1 \mathbf{C}_d \mathbf{H}_1^T \mathbf{W}_1^{-1/2} & \frac{1}{\sqrt{w}} \mathbf{W}_1^{-1/2} \mathbf{H}_1 \mathbf{C}_d \mathbf{H}_2^T \\ \hline \frac{1}{\sqrt{w}} \mathbf{H}_2 \mathbf{C}_d \mathbf{H}_1^T \mathbf{W}_1^{-1/2} & \frac{1}{w} \mathbf{H}_2 \mathbf{C}_d \mathbf{H}_2^T \end{array} \right) = \left(\begin{array}{c|c} \tilde{\mathbf{C}}_{dW11} & \tilde{\mathbf{C}}_{dW12} \\ \hline \tilde{\mathbf{C}}_{dW21} & \tilde{\mathbf{C}}_{dW22} \end{array} \right)$$

We investigate the behavior of $\tilde{\mathbf{C}}_{dW}^+$ as we let $w \rightarrow 0$. Note that the MP inverse of \mathbf{C}_{dW} under \mathbf{W}_{H_1} will have the same positive eigenvalues as $\tilde{\mathbf{C}}_{dW}^+$. Hence eigenvalue-based criteria evaluated on $\tilde{\mathbf{C}}_{dW}$ will equal that for \mathbf{C}_{dW} .

Let \mathbf{P}_0 be the orthogonal projector onto the null space of \mathbf{C}_d and define $\tilde{\mathbf{P}}_0 = \tilde{\mathbf{W}}^{1/2} \mathbf{H} \mathbf{P}_0 \mathbf{H}^T \tilde{\mathbf{W}}^{1/2}$. Then $\tilde{\mathbf{C}}_{dW}$ and $\tilde{\mathbf{P}}_0$ are orthogonal and $\left[\tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0 \right]^{-1} = \tilde{\mathbf{C}}_{dW}^+ + \tilde{\mathbf{P}}_0^+$. Partition \mathbf{C}_{dH} as

$$\mathbf{C}_{dH} = \left(\begin{array}{c|c} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \hline \mathbf{E}_{21} & \mathbf{E}_{22} \end{array} \right)$$

where $\mathbf{E}_{ij} = \mathbf{H}_i \mathbf{C}_d \mathbf{H}_j^T$ for $i, j = 1, 2$. Note that $\mathbf{E}_{11(2)} = \mathbf{E}_{11} - \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \mathbf{E}_{21} = \mathbf{C}_{dH11(2)}$ given in (4.1).

Rewrite the matrix $\tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0$ as

$$\begin{aligned} \tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0 &= \tilde{\mathbf{W}}^{-1/2} \left[\mathbf{C}_{dH} + \tilde{\mathbf{W}} \mathbf{H} \mathbf{P}_0 \mathbf{H}^T \tilde{\mathbf{W}} \right] \tilde{\mathbf{W}}^{-1/2} \\ &= \tilde{\mathbf{W}}^{-1/2} \left(\begin{array}{c|c} \mathbf{E}_{11} + \mathbf{W}_1 \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1 & \mathbf{E}_{12} \\ \hline \mathbf{E}_{21} & \mathbf{E}_{22} \end{array} \right) \tilde{\mathbf{W}}^{-1/2}, \end{aligned}$$

which follows because $\mathbf{H}_2 \mathbf{P}_0 = 0$.

Denote $\bar{\mathbf{E}}_{11} = \mathbf{E}_{11} + \mathbf{W}_1 \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1$ and $\bar{\mathbf{E}}_{11(2)} = \bar{\mathbf{E}}_{11} - \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \mathbf{E}_{21}$. The matrix $\bar{\mathbf{E}}_{11(2)}$ is

easily shown to be nonsingular, which we use to derive $\left[\tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0\right]^{-1}$:

$$\begin{aligned}
\left[\tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0\right]^{-1} &= \tilde{\mathbf{W}}^{1/2} \left(\begin{array}{c|c} \bar{\mathbf{E}}_{11} & \mathbf{E}_{12} \\ \hline \mathbf{E}_{21} & \mathbf{E}_{22} \end{array} \right)^{-1} \tilde{\mathbf{W}}^{1/2} \\
&= \tilde{\mathbf{W}}^{1/2} \left(\begin{array}{c|c} \bar{\mathbf{E}}_{11(2)}^{-1} & -\bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \\ \hline -\mathbf{E}_{22}^{-1} \mathbf{E}_{21} \bar{\mathbf{E}}_{11(2)}^{-1} & \mathbf{E}_{22}^{-1} + \mathbf{E}_{22}^{-1} \mathbf{E}_{21} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \end{array} \right) \tilde{\mathbf{W}}^{1/2} \\
&= \left(\begin{array}{c|c} \mathbf{W}_1^{1/2} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{W}_1^{1/2} & -\sqrt{w} \mathbf{W}_1^{1/2} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \\ \hline -\sqrt{w} \mathbf{E}_{22}^{-1} \mathbf{E}_{21} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{W}_1^{1/2} & w \left[\mathbf{E}_{22}^{-1} + \mathbf{E}_{22}^{-1} \mathbf{E}_{21} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{E}_{12} \mathbf{E}_{22}^{-1} \right] \end{array} \right).
\end{aligned}$$

Taking the limit of this matrix as $w \rightarrow 0$ clearly gives

$$\lim_{w \rightarrow 0} \left[\tilde{\mathbf{C}}_{dW} + \tilde{\mathbf{P}}_0\right]^{-1} = \left(\begin{array}{c|c} \mathbf{W}_1^{1/2} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{W}_1^{1/2} & 0 \\ \hline 0 & 0 \end{array} \right) = \tilde{\mathbf{C}}_{dW}^+ + \tilde{\mathbf{P}}_0^+ \quad (4.3)$$

Note that w is not involved in $\tilde{\mathbf{P}}_0$ so

$$\lim_{w \rightarrow 0} \tilde{\mathbf{P}}_0 = \left(\begin{array}{c|c} \mathbf{W}_1^{1/2} \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1^{1/2} & 0 \\ \hline 0 & 0 \end{array} \right) \Rightarrow \lim_{w \rightarrow 0} \tilde{\mathbf{P}}_0^+ = \left(\begin{array}{c|c} [\mathbf{W}_1^{1/2} \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1^{1/2}]^+ & 0 \\ \hline 0 & 0 \end{array} \right).$$

Plugging this into (4.3) and solving for $\tilde{\mathbf{C}}_{dW}^+$ gives

$$\lim_{w \rightarrow 0} \tilde{\mathbf{C}}_{dW}^+ = \left(\begin{array}{c|c} \mathbf{W}_1^{1/2} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{W}_1^{1/2} - [\mathbf{W}_1^{1/2} \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1^{1/2}]^+ & 0 \\ \hline 0 & 0 \end{array} \right).$$

Now $\bar{\mathbf{E}}_{11(2)} = \mathbf{E}_{11(2)} + \mathbf{W}_1 \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1$ is a sum of two orthogonal matrices so

$$\bar{\mathbf{E}}_{11(2)}^{-1} = \mathbf{E}_{11(2)}^+ + (\mathbf{W}_1 \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1)^+ = \mathbf{E}_{11(2)}^+ + \mathbf{W}_1^{-1/2} [\mathbf{W}_1^{1/2} \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1^{1/2}]^+ \mathbf{W}_1^{-1/2},$$

showing that $\mathbf{W}_1^{1/2} \bar{\mathbf{E}}_{11(2)}^{-1} \mathbf{W}_1^{1/2} - [\mathbf{W}_1^{1/2} \mathbf{H}_1 \mathbf{P}_0 \mathbf{H}_1^T \mathbf{W}_1^{1/2}]^+ = \mathbf{W}_1^{1/2} \mathbf{E}_{11(2)}^+ \mathbf{W}_1^{1/2}$, the MP inverse of $\mathbf{W}_1^{-1/2} \mathbf{C}_{dH11(2)} \mathbf{W}_1^{-1/2}$. \square

It at first seems that setting small weight onto parameters/effects in the model would imply that they are essentially zero. One must realize though that the weights only come into play when comparing information matrices, which are design/model dependent. Making decisions about the important effects in the model must be done prior to deciding on how to weight information of the target parameters. This will be the main topic in Chapter 5.

Setting $\mathbf{W}_1 = \mathbf{I}_{k_1}$ in Theorem 4.2.1 shows that unweighted estimation of $\mathbf{H}_1 \boldsymbol{\tau}$ is asymptotically equivalent to a weighted estimation problem for $\boldsymbol{\tau}$. The theorem also tells us that adjusting information for the nuisance parameters $\boldsymbol{\beta}$ may be thought of as a weighted optimization for the combined parameter vector $(\boldsymbol{\tau}^T | \boldsymbol{\beta}^T)^T$. In some cases, the theorem allows us to arrive at the necessary adjusted information matrix for $\boldsymbol{\tau}$ from a full rank matrix, \mathbf{C}_{dH} . This provides a computational technique to search for optimal adjusted information matrices in commonly employed design search algorithms through a modification of the candidate points.

To demonstrate this, reparameterize the treatment effects in the one-way ANOVA model using a $(p-1) \times p$ matrix \mathbf{H} whose rows are coefficient vectors for $p-1$ orthogonal contrasts:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_d \mathbf{H}^T \mathbf{H} \boldsymbol{\tau} + \mathbf{e} = \mathbf{X}_d^* \boldsymbol{\theta} + \mathbf{e}$$

where $\mathbf{X}_d^* = (\mathbf{1}_N | \mathbf{X}_d \mathbf{H}^T) = \mathbf{X}_d (\mathbf{1}_p | \mathbf{H}^T)$ has full column rank and $\boldsymbol{\theta}^T = (\mu | (\mathbf{H} \boldsymbol{\tau})^T)$. To show that eigenvalue optimization for the information matrix of $\boldsymbol{\tau}$ equals that for $\mathbf{H} \boldsymbol{\tau}$, we show that the positive eigenvalues for \mathbf{C}_d and \mathbf{C}_{dH} are equivalent. Let λ be a positive eigenvalue

for C_d . Then for some vector \mathbf{e} ,

$$\begin{aligned} C_d \mathbf{e} &= \lambda \mathbf{e} \\ \Leftrightarrow C_d \mathbf{H}^T \mathbf{H} \mathbf{e} &= \lambda \mathbf{e} \quad \text{as } \mathbf{H}^T \mathbf{H} = P_\tau \\ \Rightarrow \mathbf{H} C_d \mathbf{H}^T (\mathbf{H} \mathbf{e}) &= \lambda (\mathbf{H} \mathbf{e}) . \end{aligned}$$

Hence Theorem 4.2.1 says that eigenvalue-based optimization for $\boldsymbol{\tau}$ is asymptotically equivalent to a weighted optimization for $\boldsymbol{\theta}$ with $\mathbf{W}_1 = \mathbf{I}_{p-1}$ assigned to $\mathbf{H}\boldsymbol{\tau}$ and w assigned to μ . The set of candidate points will be all possible rows of \mathbf{X}_d^* and we must choose a sufficiently small value of w which we use to generate the weighted candidate points \mathbf{X}_{cw}^* . This weight, w , cannot be set too close to 0 lest the information matrix $\mathbf{X}_d^{*T} \mathbf{X}_d^*$ becomes nearly singular. On the other hand, w must be close enough to zero to ensure optimization is focused on $\mathbf{H}\boldsymbol{\tau}$.

The following sections apply this computational approach to search for A -optimal CRDs for BP effects, adjusted for the intercept, and A -optimal incomplete block designs. The latter requires modification of the model as was done for the one-way ANOVA model above. In both cases we compare the algorithm results to the theoretically derived optimal designs to show how well the limiting approach works.

4.3 A -optimal CRDs for two-level BP effects adjusted for $\theta_{0\dots 0}$

Suppose that we are interested in estimating all BP effects, either non-normalized or normalized, except for the baseline mean. The development to follow focuses on the non-normalized case; results may be similarly shown for the normalized BP effects by incorporating an ad-

ditional weighting into the information matrix for non-normalized effects.

As mentioned prior to (3.10), we may permute the columns of \mathbf{H}_{BP}^{-1} so that $\mathbf{H}_{BP}^{-1} = (\mathbf{1}_N | \mathbf{A}_{m;1} | \dots | \mathbf{A}_{m;m})$. This permutation does not affect weighted information, as it is equivalent to reordering the effects in $\boldsymbol{\theta} = \mathbf{H}_{BP}\boldsymbol{\tau}$, and does not permute $\boldsymbol{\tau}$. Let $\boldsymbol{\theta}_{(0)}$ be the vector of BP effect parameters excluding the baseline mean. The partition of $\boldsymbol{\theta}$ into the mean and $\boldsymbol{\theta}_{(0)}$ is reflected in the model thusly:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_d\boldsymbol{\tau} + \mathbf{e} = \mathbf{X}_d\mathbf{H}_{BP}^{-1}\mathbf{H}_{BP}\boldsymbol{\tau} + \mathbf{e} \\ &= \mathbf{X}_d(\mathbf{1}_p | \mathbf{A}_{m;1} | \dots | \mathbf{A}_{m;m})\boldsymbol{\theta} + \mathbf{e} \\ &= \mathbf{1}_N\theta_{0\dots 0} + \mathbf{X}_d(\mathbf{A}_{m;1} | \dots | \mathbf{A}_{m;m})\boldsymbol{\theta}_{(0)} + \mathbf{e} . \end{aligned} \tag{4.4}$$

The Section 4.2 computational approach to finding A -optimal designs, when applied to BP effects adjusted for θ_0 , requires us to specify a set of weighted candidate points \mathbf{X}_{cW} . Since we are starting from a cell means model, this will be $\mathbf{X}_{cW} = \mathbf{W}^{-1/2}$ for an appropriate weight matrix, \mathbf{W} . While we could use the weight matrix $\mathbf{W}_{BPw} = \mathbf{H}_{BP}^T \mathbf{D}_w \mathbf{H}_{BP}$ that is identical to \mathbf{W}_{BP} except for assigning w to the intercept portion of the matrix through \mathbf{D}_w , a clearer and equivalent approach is to use the weighted candidate points $\mathbf{X}_{cBP} = (w\mathbf{1}_p | \mathbf{A}_{m;1} | \dots | \mathbf{A}_{m;m})$ where w is sufficiently small.

As there is some uncertainty about which w will achieve the required optimization, we compare the computational approach to theoretically derived A -optimal CRDs that are known to perform the correct optimization. From (4.4), the adjusted information matrix

for $\boldsymbol{\theta}_{(0)}$ is

$$\mathbf{C}_{d\boldsymbol{\theta}_{(0)}} = \begin{pmatrix} \frac{\mathbf{A}_{m;1}^T}{\vdots} \\ \frac{\mathbf{A}_{m;m}^T}{\vdots} \end{pmatrix} \mathbf{X}_d^T \left(\mathbf{I} - \frac{1}{N} \mathbf{J} \right) \mathbf{X}_d (\mathbf{A}_{m;1} | \dots | \mathbf{A}_{m;m}) .$$

The adjusted information matrix $\mathbf{C}_{d\boldsymbol{\theta}_{(0)}}$ is the lower right $2^m - 1 \times 2^m - 1$ principal submatrix of $\mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d \mathbf{H}_{BP}^{-1}$ and the remaining elements of $\mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d \mathbf{H}_{BP}^{-1}$ are 0, since the first row of \mathbf{H}_{BP}^{-1T} is $\mathbf{1}$. Hence the positive eigenvalues of $\mathbf{C}_{d\boldsymbol{\theta}_{(0)}}$ equal those for $\mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d \mathbf{H}_{BP}^{-1}$. Note that $\mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d$ is the information matrix for $(\mathbf{I} - \frac{1}{p} \mathbf{J}) \boldsymbol{\tau}$ and the positive eigenvalues for $\mathbf{C}_{d\boldsymbol{\theta}_{(0)}}$ equal those for $\mathbf{W}_{BP}^{-1/2} \mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d \mathbf{W}_{BP}^{-1/2}$ where \mathbf{W}_{BP} has been defined previously. This implies that one may think of eigenvalue-based optimization for $\boldsymbol{\theta}_{(0)}$ as weighted optimization for contrasts of $\boldsymbol{\tau}$ with weight matrix \mathbf{W}_{BP} .

Working now under the contrasts framework, we take $\mathbf{C}_d = \mathbf{X}_d^T (\mathbf{I} - \frac{1}{N} \mathbf{J}) \mathbf{X}_d$. Let $\mathbf{H}_{BP(0)}$ be the $2^m - 1$ rows of \mathbf{H}_{BP} excluding that corresponding to the baseline mean, so $\boldsymbol{\theta}_{(0)} = \mathbf{H}_{BP(0)} \boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is the vector of treatment means. Clearly $\mathbf{H}_{BP(0)}$ are $2^m - 1$ coefficient vectors that make up $2^m - 1$ weighted orthogonal contrasts under \mathbf{W}_{BP} where each is assigned weight 1. Using Lemma 2.3.2 leads us to a similar expression seen in (3.11) in Section 3.4

$$\begin{aligned} A_{dW} &\propto \text{tr}(\mathbf{H}_{BP(0)} \mathbf{C}_d^{-1} \mathbf{H}_{BP(0)}^T) = \text{tr}(\mathbf{H}_{BP(0)} \mathbf{R}_d^{-1} \mathbf{H}_{BP(0)}^T) \\ &= \text{tr}(\mathbf{R}_d^{-1} \mathbf{H}_{BP(0)}^T \mathbf{H}_{BP(0)}) \\ &= \sum_{j_1 \dots j_m \in J} \frac{w_{j_1 \dots j_m}^*}{r_{dj_1 \dots j_m}}, \end{aligned} \tag{4.5}$$

where $w_{j_1 \dots j_m}^*$ simply counts the number of BP effects, excluding the baseline mean, that the treatment indexed by $j_1 \dots j_m$ is involved in. This is due to the fact that the squared elements of \mathbf{H}_{BP} are either 0 or 1. The weights $w_{j_1 \dots j_m}$ in (3.11) also counted the number of

BP effects the treatment was involved in, but included the baseline mean. By Lemma 1.2.1 (ii), a treatment of heft h is involved in 2^{m-h} BP effects when we include the baseline mean. When we exclude the baseline mean, this only affects the count for $\tau_{0\dots 0}$, since the first row of \mathbf{H}_{BP} is the unit vector with 1 in the first column. Hence the $w_{0\dots 0}^* = w_{00\dots 0} - 1 = 2^m - 1$ and the weight in (4.5) for a treatment with heft $h > 0$ is $w_{j_1\dots j_m}^* = w_{j_1\dots j_m} = 2^{m-h}$.

Optimizing (4.5) with respect to $r_{dj_1\dots j_m}$ produces the same form of solution as in Theorem 3.3.2, but using the weights just described. The impact of adjusting for the intercept is first investigated for the approximate A_W -optimal CRD, which requires us to calculate the sum of the square roots of the weights $w_{j_1\dots j_m}^*$. The square root weight for the baseline treatment is $w_{00\dots 0}^* = \sqrt{2^m - 1}$ while for the unadjusted case it was $w_{0\dots 0} = 2^m$. Noting that $w_{j_1\dots j_m}^* = w_{j_1\dots j_m}$ for all other treatments, a concise expression for the sum of the square roots of these weights may be found by adding and subtracting $w_{00\dots 0}^* = \sqrt{2^m}$:

$$\begin{aligned}
\sum_{j_1\dots j_m \in J} \sqrt{w_{j_1\dots j_m}^*} &= \sqrt{w_{00\dots 0}^*} - \sqrt{w_{00\dots 0}} + \sum_{j_1\dots j_m \in J} \sqrt{w_{j_1\dots j_m}^*} \\
&= \sqrt{w_{00\dots 0}^*} - \sqrt{w_{00\dots 0}} + \sum_{j_1\dots j_m \in J} \sqrt{w_{j_1\dots j_m}} \\
&= \sqrt{2^m - 1} - \sqrt{2^m} + (1 + \sqrt{2})^m.
\end{aligned} \tag{4.6}$$

As m increases this sum will approximately equal $(1 + \sqrt{2})^m$, which was the sum of the square roots of weights for the unadjusted case. Furthermore, only slightly smaller relative weight is placed on the baseline mean so we anticipate that designs optimal for all the BP effects will be similar to those that adjust for the baseline mean, and will become identical for all N as m grows large.

When working with the normalized BP effects, we premultiply $\mathbf{H}_{BP(0)}$ by $\mathbf{D}_{w(0)}^{1/2}$, a $2^m - 1 \times 2^m - 1$ diagonal matrix having the necessary normalization constants. Now letting $w_{j_1\dots j_m}^*$

denote the diagonal elements of $\mathbf{H}_{BP(0)}^T \mathbf{D}_{w(0)} \mathbf{H}_{BP(0)}$, we have $w_{j_1 \dots j_m}^*$ adds the inverse squared lengths for those BP effects that the treatment indexed by $j_1 \dots j_m$ is involved in. This is different from the non-normalized case, which simply counted the number of BP effects the treatment was involved in. This gives the same weights under \mathbf{W}_{BP^*} for the unadjusted case, but again we must modify the weight for the baseline treatment to be $w_{0 \dots 0}^* = \left(\frac{3}{2}\right)^m - 1$. The sum of the square roots of the weights is found similarly to that in (4.6):

$$\sum_{j_1 \dots j_m \in J} \sqrt{w_{j_1 \dots j_m}^*} = \sqrt{(3/2)^m - 1} - \sqrt{(3/2)^m} + (\sqrt{3/2} + 1/\sqrt{2})^m,$$

where $(\sqrt{3/2} + 1/\sqrt{2})^m$ equals the sum of the square roots of the diagonal elements of \mathbf{W}_{BP^*} .

As is apparent from a comparison of the adjusted and unadjusted weights, the A -optimal approximate CRDs for both the non-normalized and normalized BP effects adjusted for the intercept are nearly identical to those without adjustment. Again, because treatments of the same heft have the same weight we anticipate BAs are optimal or highly efficient exact CRDs. Theorem 3.3.3 was used to construct the A -optimal designs for $m = 3$ factors displayed in Tables 4.1 and 4.2. Compared to Tables 3.3 and 3.4 we see that most designs are similar, although replication of the baseline is less emphasized in the new tables. Greater emphasis is placed on replicating treatments having heft 1. For example, for $N = 17$ runs for the non-normalized BP effects, the number of replications is higher by 1 on a heft 1 treatment and lower by 1 on the baseline treatment, as compared to the $N = 17$ design in Table 3.3. For $N = 10$ runs the A -optimal CRD for the normalized BP effects shows the same shift of replication relative to its Table 3.4 counterpart.

Tables 4.3 and 4.4 show the A -optimal designs that are BAs of strength m when we adjust for the intercept for both the non-normalized and normalized BP effects. These tables are presented in a similar fashion to Tables 3.5 and 3.6. Not surprisingly many of the designs

Table 4.1: A -optimal CRDs for estimation of non-normalized BP effects, adjusted for the intercept, for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.

Treatment Index	N											
	8	9	10	11	12	13	14	15	16	17	...	20
000	1	2	2	2	2	3	3	3	3	3		4
100	1	1	2	2	2	2	2	2	2	3		3
010	1	1	1	2	2	2	2	2	2	2		3
001	1	1	1	1	2	2	2	2	2	2		3
110	1	1	1	1	1	1	2	2	2	2		2
101	1	1	1	1	1	1	1	2	2	2		2
011	1	1	1	1	1	1	1	1	2	2		2
111	1	1	1	1	1	1	1	1	1	1		1
BA	*	*			*	*			*			*

Table 4.2: A -optimal CRDs for estimation of normalized BP effects, adjusted for the intercept, for $m = 3$ factors. The last row indicates whether the design is a BA of strength 3.

Treatment Index	N											
	8	9	10	11	12	13	14	15	16	17	...	20
000	1	2	2	2	2	3	4	4	4	4		4
100	1	1	2	2	2	2	2	3	3	3		3
010	1	1	1	2	2	2	2	2	3	3		3
001	1	1	1	1	2	2	2	2	2	3		3
110	1	1	1	1	1	1	1	1	1	1		2
101	1	1	1	1	1	1	1	1	1	1		2
011	1	1	1	1	1	1	1	1	1	1		2
111	1	1	1	1	1	1	1	1	1	1		1
BA	*	*			*	*	*			*		*

are exactly the same as those for unadjusted estimation of the BP effects. For $m = 5$ factors all designs are exactly the same exact for $N = 50$ and $N = 54$.

The goal now is to determine whether the algorithm described in the previous section can produce the same designs for the A -criterion and, if so, what weight w needs to be placed on

Table 4.3: BA heft replications for A -optimal CRDs for estimation of all non-normalized BP effects, adjusted for the intercept, for $m = 3$ to 5 factors with increasing N .

m	Treatment	N													
	Heft	8	9	12	13	16	19	20	21	22	25	28	29	32	
3	0	1	2	2	3	3	3	4	4	5	5	5	6	6	
	1	1	1	2	2	2	3	3	3	3	4	4	4	5	
	2	1	1	1	1	2	2	2	2	2	2	3	3	3	
	3	1	1	1	1	1	1	1	2	2	2	2	2	2	
	Treatment	N													
	Heft	16	17	21	22	28	32	33	37	38	42	48	49	50	
4	0	1	2	2	3	3	3	4	4	5	5	5	6	6	
	1	1	1	2	2	2	3	3	3	3	4	4	4	4	
	2	1	1	1	1	2	2	2	2	2	2	3	3	3	
	3	1	1	1	1	1	1	1	2	2	2	2	2	2	
	4	1	1	1	1	1	1	1	1	1	1	1	1	2	
	Treatment	N													
	Heft	32	33	38	39	49	54	55	65	66	71	81	82	87	
5	0	1	2	2	3	3	3	4	4	5	5	5	6	6	
	1	1	1	2	2	2	3	3	3	3	4	4	4	4	
	2	1	1	1	1	2	2	2	2	2	2	3	3	3	
	3	1	1	1	1	1	1	1	2	2	2	2	2	2	
	4	1	1	1	1	1	1	1	1	1	1	1	1	2	
	5	1	1	1	1	1	1	1	1	1	1	1	1	1	

the intercept to achieve this goal. To implement the algorithm for non-normalized BP effects, we modified the candidate points by post-multiplying \mathbf{X}_d with \mathbf{H}_{BP}^{-1} instead of $\mathbf{W}_{BP}^{-1/2}$, yielding an information matrix with the same eigenvalues as if using $\mathbf{W}_{BP}^{-1/2}$. Because $\mathbf{X}_d = \mathbf{I}_p$, this means the candidate points are simply the rows of \mathbf{H}_{BP}^{-1} . It is unnecessary to input the entire matrix \mathbf{H}_{BP}^{-1} as the set of candidate points, since all columns corresponding to BP 2-f.i.'s or higher can be generated as Hadamard products of \mathbf{A}_m , defined previously. The input to generate candidate points, prior to additional weighting of the intercept or normalization, is thus simply $(\mathbf{1}|\mathbf{A}_m)$.

We construct \mathbf{A}_m using SAS PROC FACTEX by generating a saturated main effects plan

Table 4.4: BA heft replications for A -optimal CRDs for estimation of all normalized BP effects, adjusted for the intercept, for $m = 3$ to 5 factors with increasing N .

m	Treatment Heft	N												
		8	9	12	13	14	17	20	21	24	25	28	29	30
3	0	1	2	2	3	4	4	4	5	5	6	6	6	7
	1	1	1	2	2	2	3	3	3	4	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	3	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	2	2
	Treatment Heft	N												
		16	17	21	22	23	27	33	34	38	39	40	46	50
4	0	1	2	2	3	4	4	4	5	5	6	7	7	7
	1	1	1	2	2	2	3	3	3	4	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	2	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	1	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	1
	Treatment Heft	N												
		32	33	38	39	40	45	55	56	57	62	63	73	83
5	0	1	2	2	3	4	4	4	5	6	6	7	7	7
	1	1	1	2	2	2	3	3	3	3	4	4	4	4
	2	1	1	1	1	1	1	2	2	2	2	2	3	3
	3	1	1	1	1	1	1	1	1	1	1	1	1	2
	4	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1	1	1	1	1	1

and then recoding the variables to be 0 and 1 instead of ± 1 . We then append $\mathbf{1}_{2^m}$ to this matrix, additionally weighting the intercept column with the inverse weight $1/\sqrt{w}$, where w is close to zero in order to adjust for the intercept. Based on our approximate results, we anticipated that w may not need to be too much smaller than 1 to produce the designs in the above tables, although the theory dictates that w should be close to zero in order to guarantee we are evaluating adjusted information matrices.

The intercept term must be included in the SAS PROC OPTEX model statement, along with every possible product of the columns in \mathbf{A}_m , in order to specify the full BP model. SAS PROC OPTEX includes an intercept in the model by default, so the “noimt” model

option must be specified to prevent SAS from including an additional intercept in the model.

To achieve optimization for the normalized BP effects, we need only multiply each column of \mathbf{A}_m with $\sqrt{2}$, the inverse normalization constant for the BP main effects. Given this “normalized” \mathbf{A}_m , the procedure described above is otherwise unchanged.

Experimental results employing the above approach will now be described. The modified Fedorov algorithm based on the A -criterion with 500 iterations was run and the A -best design was output. Due to the large number of required runs we only performed the experiment for up to $m = 5$ factors. We initially started by setting $w = 0.99$, and when that did not produce the known A -optimal design then we searched for the largest $w < 0.99$ that would give the optimal design. In general, the weight $w = 0.99$ was able to find the designs listed in the tables above, but usually failed for large N . Some of these cases are listed in Table 4.5 along with the first weight found to produce the known A -optimal design. When the SAS PROC OPTEX design did not match the known A -optimal design, it often put replicates that belonged on the treatments with heft 1 on the baseline treatment. The weight $w = 0.99$ always gave the known A -optimal design for the non-normalized BP effects for $m = 5$ up to $N = 350$ runs.

Table 4.5: Cases where $w < 0.99$ was required for SAS PROC OPTEX to find known A -optimal design for adjusted BP effects.

m	N	Normalized BP (Y/N)	A -optimal BA index set (n_0, n_1, \dots, n_m)	Required w
3	60	N	(11, 9, 6, 4)	0.25
3	44	Y	(9, 7, 4, 2)	0.01
4	151	N	(17, 13, 9, 6, 4)	0.50
4	38	Y	(5, 4, 2, 1, 1)	0.15
4	93	Y	(13, 9, 5, 3, 2)	0.10
5	88	Y	(7, 5, 3, 2, 1, 1)	0.45

Although we see that our limiting weights approach does in fact lead us to the correct A -

optimal design for the adjusted BP effects, these examples are rather tame. Evaluating adjusted information is more interesting for blocked designs, which we now explore.

4.4 *A*-optimal incomplete block designs

SAS PROC OPTEX is capable of implementing blocking with fixed block sizes for the *D*-criterion only, using the algorithm by Cook and Nachtsheim (1989). For the *A*-criterion, there is no clear option in SAS PROC OPTEX for optimizing treatment estimation in the presence of blocks. We will return to this difficulty shortly. To be clear, the goal of this section is to use weighting to force OPTEX to find optimal block designs for the *unweighted* *A*-criterion.

We can implement Theorem 4.2.1 for the block design problem in SAS PROC OPTEX by including blocking factors in the candidate set and weighting them appropriately. We begin by generating $p+h$ columns for the treatments and blocks, where each column is an indicator variable for whether the unit is assigned treatment i or lies in block j , producing ph total candidate points. The model statement, which includes treatment and block effects, must give a full-rank information matrix. Let μ be the intercept parameter, $\boldsymbol{\tau}$ be the $p \times 1$ vector of treatment parameters, and $\boldsymbol{\beta}$ be the $h \times 1$ vector of block parameters. Instead of specifying the overparameterized model

$$E(\mathbf{y}) = \mu \mathbf{1}_N + \mathbf{X}_d \boldsymbol{\tau} + \mathbf{L} \boldsymbol{\beta},$$

in SAS PROC OPTEX, we use

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}_d \mathbf{H}^T \mathbf{H} \boldsymbol{\tau} + \mathbf{L} \boldsymbol{\beta} \\ &= \mathbf{X}_{dH} \boldsymbol{\tau}_H + \mathbf{L} \boldsymbol{\beta} , \end{aligned}$$

where $\boldsymbol{\tau}_H = \mathbf{H} \boldsymbol{\tau}$ is a vector of $p - 1$ orthonormal contrasts and $\mathbf{X}_{dH} = \mathbf{X}_d \mathbf{H}^T$ is now the $N \times (p - 1)$ design matrix. We maintain the indicator columns for the block effects to make weighting them straightforward. It follows that $\mathbf{X}_{dHL} = (\mathbf{X}_{dH} | \mathbf{L})$ has full column rank and hence $\mathbf{X}_{dHL}^T \mathbf{X}_{dHL}$ is invertible. Once an optimal design is found in terms of \mathbf{X}_{dHL} and output to a new data set, we perform this transformation to get back the original treatment design matrix: $\mathbf{X}_d; \mathbf{X}_{dH} \mathbf{H} + \frac{1}{p} \mathbf{J}$.

The sample size must also be specified in the procedure, which we take to be $N = hk$, where k is the proposed block size, not the dimension of the estimation space. Even though we attempt to specify a block size, there is no guarantee that this will be reflected in the most efficient design given. This is because when minimizing the trace of $(\mathbf{X}_{dHL}^T \mathbf{X}_{dHL})^{-1}$, all potential candidate points are considered and it is possible that unequal block sizes will be found optimal. One strategy for maintaining equal block sizes, the opposite of setting block weights close to zero, is to *increase* the block weight. This encourages “equal replication” of the block effects and hence equal block sizes. The concern though is whether the design is focusing too much attention on equal block sizes, and not on efficient estimation of the treatment effects.

A variety of incomplete block design scenarios were implemented in SAS PROC OPTEX for which an A -optimal IBD is known. Runs were carried out on both the unweighted and weighted candidate points to explore the efficacy of the weighted approach. Due to the previously mentioned computational issues, different block weights were used to get an A -

efficient design with equal block sizes and efficient estimation of the treatment effects. We performed 100 iterations of the modified Fedorov algorithm and recorded the number of times the unweighted and weighted results gave the known A -optimal IBD. The results are given in Table 4.6, which also includes the block weight which most often found the known A -optimal IBD.

Overall we see the weighted approach performed just as well as the unweighted approach and often outperformed it. The weighted approach worked particularly well when we had 6 and 7 treatments, often finding the A -optimal design twice as often as the unweighted approach. When the A -best design had unequal block sizes we considered every iteration a failure. For nearly saturated cases, we avoided this by setting large block weight, which shifted the algorithm's focus to having equal block sizes, leading to quick identification of the A -optimal design. Placing larger weight works well in these cases because there are relatively few saturated designs to consider and the algorithm is quickly able to find the one that is A -optimal.

In practice, when the A -optimal design is unknown, it is only when the block weight is “small”, and the best design gives equal block sizes, that we can be confident the design is A -efficient for treatment estimation. That is, when large weight must be placed on the blocks to achieve equal block sizes, the best design found may not be the most efficient for estimation of all contrasts. In this case, efficiency of the found design can be compared to known efficiency bounds to assess if it is reasonably good. Even though the unweighted approach was capable of finding the A -optimal design, introducing weights, either large or small, into the search proves to be a useful tool.

Table 4.6: Number of A -optimal designs found of 100 modified Fedorov iterations for p treatments in h blocks of size k . Missing block weight indicates unweighted approach found the A -optimal design for all 100 iterations.

p	h	k	Unweighted	Weighted	Block Weight
3	8	2	100	100	—
3	10	2	100	100	—
4	4	2	100	100	—
4	6	3	100	100	—
4	9	2	0	100	0.5
5	4	2	0	100	5
5	5	4	100	100	—
5	8	3	100	100	—
6	2	4	0	100	5
6	5	2	0	100	5
6	12	2	48	83	0.1
6	12	3	77	72	0.05
7	7	2	0	100	5
7	7	3	40	88	0.05
7	9	5	1	3	0.5
8	2	5	0	100	5
8	5	3	0	85	10
8	8	6	98	96	0.05
9	2	5	0	100	5
9	3	6	100	100	—
9	6	6	11	12	0.05
10	2	6	0	100	5
10	10	6	1	27	0.05
11	4	4	0	100	10
11	5	3	0	100	10

Chapter 5

Weighted optimality of reduced models

In Chapter 4 we showed that vanishing weights do not cause certain parameters to be dropped from the model. Rather, it leads to weighted estimation of non-vanishing parameters adjusted for those receiving vanishing weights. This chapter outlines work on the linear model theory of what we call *reduced models* resulting from constraints of the form $\mathbf{H}_2\boldsymbol{\tau} = \mathbf{0}$, where the rows of \mathbf{H}_2 are coefficient vectors for estimable functions. We call these reduced models as they assume only some functions of the potentially larger estimation space have nonzero effect sizes. Such models have also been referred to as *affine linear models* by Hinkelmann and Kempthorne (2008). We then demonstrate the theory using BP fractional designs and show that some form of strength 4 BAs are good candidates for A -optimality for reduced models including all the BP 2-f.i.'s.

5.1 Reduced cell means model and weighting

Consider the cell means model for some design d and any nonsingular transformation, \mathbf{H} :

$$\begin{aligned}\mathbf{y} &= \mathbf{X}_d \boldsymbol{\tau} + \mathbf{e} \\ &= \mathbf{X}_d \mathbf{H}^{-1} \mathbf{H} \boldsymbol{\tau} + \mathbf{e} .\end{aligned}$$

where the elements of $\boldsymbol{\tau}$ are the p treatment means, and \mathbf{X}_d is an $N \times p$ 0/1 unit/treatment incidence matrix. Suppose we partition \mathbf{H} and $\boldsymbol{\tau}$ this way

$$\mathbf{H} = \left(\begin{array}{c} \mathbf{H}_1 \\ \mathbf{H}_2 \end{array} \right) = \left(\begin{array}{c|c} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \hline \mathbf{H}_{21} & \mathbf{H}_{22} \end{array} \right) \quad \boldsymbol{\tau} = \left(\begin{array}{c} \boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_2 \end{array} \right) , \quad (5.1)$$

where $\boldsymbol{\tau}_1$ is $p_1 \times 1$ and $\boldsymbol{\tau}_2$ is $p_2 \times 1$, implying dimensions of the partitioned components of \mathbf{H} . Let $\boldsymbol{\gamma} = \mathbf{H} \boldsymbol{\tau}$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)^T$, where $\boldsymbol{\gamma}_i = \mathbf{H}_i \boldsymbol{\tau}$, for $i = 1$ or 2 . An experimenter may want to reparameterize the model for two reasons: (1) they are more interested in estimating $\boldsymbol{\gamma}$ precisely, not $\boldsymbol{\tau}$, or (2) they believe that some subset of the parameters $\boldsymbol{\gamma}$, say $\boldsymbol{\gamma}_2 = \mathbf{H}_2 \boldsymbol{\tau}$, is negligible and want to impose this constraint in the model. That is, they may want to declare that the potentially estimable functions $\boldsymbol{\gamma}_2$ are all zero.

If $\mathbf{H}_2 \boldsymbol{\tau} = \mathbf{0}$ then $\mathbf{H}_{21} \boldsymbol{\tau}_1 + \mathbf{H}_{22} \boldsymbol{\tau}_2 = \mathbf{0}$ or

$$\mathbf{H}_{21} \boldsymbol{\tau}_1 = -\mathbf{H}_{22} \boldsymbol{\tau}_2 \Leftrightarrow \boldsymbol{\tau}_2 = -\mathbf{H}_{22}^{-1} \mathbf{H}_{21} \boldsymbol{\tau}_1 ,$$

assuming that \mathbf{H}_{22} is nonsingular. That is, we can write $\boldsymbol{\tau}_2$ as a linear combination of $\boldsymbol{\tau}_1$, and so

$$\boldsymbol{\tau} = \left(\begin{array}{c} \mathbf{I}_{p_1} \\ -\mathbf{H}_{22}^{-1} \mathbf{H}_{21} \end{array} \right) \boldsymbol{\tau}_1 = \mathbf{Q} \boldsymbol{\tau}_1 ,$$

and

$$\boldsymbol{\gamma}_1 = \mathbf{H}_1 \boldsymbol{\tau} = (\mathbf{H}_{11} - \mathbf{H}_{12} \mathbf{H}_{22}^{-1} \mathbf{H}_{21}) \boldsymbol{\tau}_1 = \tilde{\mathbf{H}}_{11} \boldsymbol{\tau}_1 .$$

This says that we can estimate $\mathbf{H}_1 \boldsymbol{\tau}$ by estimating only $\boldsymbol{\tau}_1$, even if \mathbf{H}_1 contains functions involving the parameters in $\boldsymbol{\tau}_2$.

Assume now that we have initially defined \mathbf{H} so that \mathbf{H}_{22} is not invertible. Because \mathbf{H} has full rank, the rows of \mathbf{H}_2 are linearly independent and there exists a $p_2 \times p_2$ submatrix of \mathbf{H}_2 that is invertible. This submatrix, which is not necessarily unique, shows that we can always assume \mathbf{H}_{22} is invertible, for we need only to permute the columns of \mathbf{H} to make it so. Suppose a column permutation is necessary, $\mathbf{H}^* = \mathbf{H}\boldsymbol{\Pi}$, so that \mathbf{H}_{22}^* , the lower-right $p_2 \times p_2$ submatrix of \mathbf{H}^* , is invertible. This does not guarantee invertibility of $\tilde{\mathbf{H}}_{11}^*$, so if there are multiple permutations that give an invertible \mathbf{H}_{22} , we will choose one that gives a nonsingular $\tilde{\mathbf{H}}_{11}^*$, assuming one exists.

If we must permute the columns of \mathbf{H} we also must permute the elements of $\boldsymbol{\tau}$ since $\mathbf{H}_2 \boldsymbol{\Pi} \boldsymbol{\tau} \neq \mathbf{H}_2 \boldsymbol{\tau}$. The permutation $\mathbf{H}_2 \boldsymbol{\Pi}$ must switch a column in \mathbf{H}_{22} with one in \mathbf{H}_{21} , otherwise a permutation would not be necessary. It follows that at least one element in $\boldsymbol{\tau}_1$ must be exchanged with an element in $\boldsymbol{\tau}_2$ to get a new mean vector, $\boldsymbol{\tau}^*$. If we let $\boldsymbol{\tau}^* = (\boldsymbol{\tau}_1^{*T} \mid \boldsymbol{\tau}_2^{*T})^T = \boldsymbol{\Pi}^T \boldsymbol{\tau}$ then $\mathbf{H}_2 \boldsymbol{\Pi} \boldsymbol{\tau}^* = \mathbf{H}_2 \boldsymbol{\Pi} \boldsymbol{\Pi}^T \boldsymbol{\tau} = \mathbf{H}_2 \boldsymbol{\tau}$. Hence we want to define $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ so that \mathbf{H}_{22} and $\tilde{\mathbf{H}}_{11}$ are invertible. Indeed, if $\tilde{\mathbf{H}}_{11}$ is not invertible then our targeted parameters $\boldsymbol{\gamma}_1$ contain redundancies under the reduced model and should thus be redefined.

Assume that \mathbf{H} satisfies the necessary invertibility conditions. We may then rewrite the cell means model as

$$\mathbf{y} = \mathbf{X}_d \mathbf{Q} \boldsymbol{\tau}_1 + \mathbf{e}$$

and the information matrix for $\boldsymbol{\tau}_1$ under this model is $\mathbf{C}_{d\boldsymbol{\tau}_1} = \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q}$. Clearly \mathbf{Q} has full column rank so we may assume that there exist designs such that $\mathbf{C}_{d\boldsymbol{\tau}_1}$ is nonsingular;

an obvious choice is the design that replicates all treatments in $\boldsymbol{\tau}_1$ at least once. Hence every linear combination of $\boldsymbol{\tau}_1$ is an estimable function, which includes $\boldsymbol{\tau}_2 = -\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{21}\boldsymbol{\tau}_1$. It follows that the projector onto the estimation space for these information matrices will be $\boldsymbol{P}_{\boldsymbol{\tau}_1} = \boldsymbol{I}_{p_1}$.

To find optimal designs for $\boldsymbol{\gamma}_1$, given $\boldsymbol{\gamma}_2 = \mathbf{0}$, transform the cell means model in the following way

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}_d \boldsymbol{Q} \boldsymbol{\tau}_1 + \boldsymbol{e} \\ &= \boldsymbol{X}_d \boldsymbol{Q} \tilde{\boldsymbol{H}}_{11}^{-1} \tilde{\boldsymbol{H}}_{11} \boldsymbol{\tau}_1 + \boldsymbol{e} .\end{aligned}\tag{5.2}$$

We may also represent model (5.2) using the approach found in Hedayat et al. (1999) via a $p_1 \times p$, 0/1 matrix \boldsymbol{K} that selects the rows of \boldsymbol{H} corresponding to the \boldsymbol{H}_1 rows, giving $\boldsymbol{KH} = \boldsymbol{H}_1$, which implies $\boldsymbol{KH}\boldsymbol{\tau} = \boldsymbol{H}_1\boldsymbol{\tau} = \tilde{\boldsymbol{H}}_{11}\boldsymbol{\tau}_1$ under the constraint $\boldsymbol{H}_2\boldsymbol{\tau} = \mathbf{0}$. Note that if we similarly define $\tilde{\boldsymbol{H}}_{22} = \boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}$, and assume it is also invertible, then

$$\boldsymbol{H}^{-1} = \left(\begin{array}{c|c} \tilde{\boldsymbol{H}}_{11}^{-1} & -\boldsymbol{H}_{11}\boldsymbol{H}_{12}\tilde{\boldsymbol{H}}_{22}^{-1} \\ \hline -\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{21}\tilde{\boldsymbol{H}}_{11}^{-1} & \tilde{\boldsymbol{H}}_{22}^{-1} \end{array} \right) ,$$

and it is easily shown that $\boldsymbol{H}^{-1}\boldsymbol{K}^T = \boldsymbol{Q}\tilde{\boldsymbol{H}}_{11}^{-1}$. Using \boldsymbol{K} , rewrite model (5.2) as

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}_d \boldsymbol{H}^{-1} \boldsymbol{K}^T \boldsymbol{K} \boldsymbol{H} \boldsymbol{\tau} + \boldsymbol{e} \\ &= \boldsymbol{X}_d \boldsymbol{H}^{-1} \boldsymbol{K}^T \tilde{\boldsymbol{H}}_{11} \boldsymbol{\tau}_1 + \boldsymbol{e} .\end{aligned}\tag{5.3}$$

This gives us two equivalent representations, (5.2) and (5.3), for the transformed cell means model under the constraint $\boldsymbol{H}_2\boldsymbol{\tau} = \mathbf{0}$. We call these reduced models since the dimension of $\boldsymbol{H}\boldsymbol{\tau}$ has been reduced to the dimension of $\tilde{\boldsymbol{H}}_{11}\boldsymbol{\tau}_1$.

We then have two different representations for the information matrix for $\gamma_1 = \mathbf{H}_1 \boldsymbol{\tau} = \tilde{\mathbf{H}}_{11} \boldsymbol{\tau}_1$:

$$\begin{aligned} \mathbf{C}_{d\gamma_1} &= \tilde{\mathbf{H}}_{11}^{-1T} \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q} \tilde{\mathbf{H}}_{11}^{-1} \\ &= \mathbf{K} \mathbf{H}^{-1T} \mathbf{X}_d^T \mathbf{X}_d \mathbf{H}^{-1} \mathbf{K}^T, \end{aligned}$$

corresponding to models (5.2) and (5.3), respectively. Upon setting $\mathbf{W} = \tilde{\mathbf{H}}_{11}^T \tilde{\mathbf{H}}_{11}$, we have a weighted version of the information matrix $\mathbf{C}_{d\tau_1}$

$$\mathbf{C}_{dW\tau_1} = \mathbf{W}^{-1/2} \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q} \mathbf{W}^{-1/2},$$

which has the same eigenvalues as $\mathbf{C}_{d\gamma_1}$. Hence optimization with respect to eigenvalue criteria applied to $\mathbf{C}_{d\gamma_1}$ is equivalent to that for weighted eigenvalue criteria based on the weighted information matrix $\mathbf{C}_{dW\tau_1}$. Approaching this as a weighted optimization problem brings new insight into forms of optimal designs through an investigation of the weight-preserving permutations Π_W for which $\Phi_W(\Pi \mathbf{C}_{d\tau_1} \Pi^T) = \Phi_W(\mathbf{C}_{d\tau_1})$. This class of permutations identifies forms of $\mathbf{C}_{d\tau_1}$ that are permutation invariant, and so cannot be improved matrix averaging. Of course this is a preliminary step in identifying the optimal designs for γ_1 ; we would need to find a corresponding \mathbf{X}_d that gives $\mathbf{C}_{d\tau_1}$, which may or may not be straightforward.

5.2 A-optimal designs for reduced BP 2-f.i. model

Suppose we have m factors each with two levels, so 2^m total treatment combinations. Throughout this section k will be used as an index parameter and should not be confused with the dimension of the estimation space, which is always p_1 . Consider $\mathbf{H} = \mathbf{H}_{BP}$, specified in Lemma 1.2.2, with corresponding $\boldsymbol{\tau}$ under revlex ordering, and let $\boldsymbol{\theta} = \mathbf{H}_{BP} \boldsymbol{\tau}$. Note

that \mathbf{H}_{BP} is lower triangular, since it is a Kronecker product of m lower triangular matrices. Also, the first row of \mathbf{H}_{BP} is a unit vector and is the only normalized row. We first present results for the reduced BP model where all three or higher order interactions are assumed negligible and, given this constraint, focus estimation on the non-normalized BP effects including up to 2-f.i.'s. We then investigate estimation of the normalized BP effects including up to 2-f.i.'s. Normalizing the BP effects in the reduced model is a specific example of a weighting of the non-normalized BP effects where the same weight is applied to BP effects of the same heft. Hence we generalize our results corresponding to this type of weighting scheme and exemplify the approach with weights equal to the normalization constants.

Let \mathbf{A}_m be the $2^m \times m$ index matrix used to define the revlex ordering of $\boldsymbol{\tau}$. The order of the rows of \mathbf{A}_m corresponds to the order of the indices of the elements in both $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. Using Kronecker products was a convenient way to define \mathbf{H}_{BP} for any number of factors, but if we assume some BP effects are negligible, and want to use the results in Section 5.1, we will need to reorder the rows of \mathbf{H}_{BP} so that we can partition it to look like (5.1). In this section we want \mathbf{H}_2 to contain the rows of \mathbf{H}_{BP} corresponding to the higher order BP interactions. We must then find permutation matrices $\boldsymbol{\Pi}_r$ and $\boldsymbol{\Pi}_c$, for row and column permutations, such that $\tilde{\mathbf{H}}_{BP} = \boldsymbol{\Pi}_r \mathbf{H}_{BP} \boldsymbol{\Pi}_c$ has the BP g -f.i.'s for $g \geq 3$ in \mathbf{H}_2 and has invertible partitions, \mathbf{H}_{22} and $\tilde{\mathbf{H}}_{11}$. Note that if $\boldsymbol{\Pi}_c$ is not the identity then we must also permute $\boldsymbol{\tau}$ as $\tilde{\boldsymbol{\tau}} = \boldsymbol{\Pi}_c^T \boldsymbol{\tau}$ so $\mathbf{H}_2 \boldsymbol{\tau} = \mathbf{H}_2 \boldsymbol{\Pi}_c \tilde{\boldsymbol{\tau}}$.

Let $\boldsymbol{\Pi}_r$ be any row permutation matrix so that the index matrix $\mathbf{V}_m = \boldsymbol{\Pi}_r \mathbf{A}_m$ has rows of the same heft grouped together and the groups increase in heft as you go down the rows of \mathbf{V}_m . That is, \mathbf{V}_m starts with the only zero heft row, next contains all of the heft one rows,

and so on. For example, such a row permutation of \mathbf{A}_3 is

$$\mathbf{A}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \mathbf{V}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (5.4)$$

which was the order of treatment combinations in Table 3.3. Note that since the rows of \mathbf{A}_m index the order of parameters in $\boldsymbol{\theta}$, this is equivalent to putting the intercept first, then main effects, then two factor interactions, etc. in $\boldsymbol{\theta}^* = \mathbf{\Pi}_r \boldsymbol{\theta} = \mathbf{\Pi}_r \mathbf{H}_{BP} \boldsymbol{\tau}$. If the negligible effects are all 3-f.i.'s and higher this will give a row permutation so that $\mathbf{H}_2 \boldsymbol{\tau} = 0$ where \mathbf{H}_2 corresponds to the lower partition of $\mathbf{\Pi}_r \mathbf{H}_{BP}$, not \mathbf{H}_{BP} .

Let $\mathbf{\Pi}_r$ be the permutation matrix satisfying this left ordering and set $\mathbf{\Pi}_c = \mathbf{\Pi}_r^T$. Note that the ordering of the elements in $\tilde{\boldsymbol{\tau}} = \mathbf{\Pi}_r \boldsymbol{\tau}$ and $\boldsymbol{\theta}^* = \mathbf{\Pi}_r \boldsymbol{\theta} = \mathbf{\Pi}_r \mathbf{H}_{BP} \boldsymbol{\tau} = \mathbf{\Pi}_r \mathbf{H}_{BP} \mathbf{\Pi}_r^T \tilde{\boldsymbol{\tau}} = \tilde{\mathbf{H}}_{BP} \tilde{\boldsymbol{\tau}}$ both correspond to the ordering of \mathbf{V}_m . Take the row of $\tilde{\mathbf{H}}_{BP}$ corresponding to a g -f.i. where if $g = 0$ we mean the BP intercept and $g = 1$ means a BP main effect. This corresponds to the element $\boldsymbol{\theta}^*$ with index of heft g , say $\theta_{j_1 \dots j_m}$, where g j_k 's equal 1 and the rest are 0. By Corollary 1.2.1, the linear combination of $\tilde{\boldsymbol{\tau}}$ giving the g -f.i. includes only one element of $\tilde{\boldsymbol{\tau}}$ of heft g , $\tau_{j_1 \dots j_m}$, which is in the same position in $\tilde{\boldsymbol{\tau}}$ as $\theta_{j_1 \dots j_m}$ is in $\boldsymbol{\theta}^*$. Hence the corresponding diagonal element of \mathbf{H}_{BP} will be nonzero (in fact, it is 1). All elements in $\tilde{\boldsymbol{\tau}}$ following $\tau_{j_1 \dots j_m}$ will have heft greater than or equal to g and so cannot be involved in the linear combination. It follows that the diagonal elements of $\tilde{\mathbf{H}}_{BP}$ are all 1 and every

element after a diagonal element is 0, i.e. $\tilde{\mathbf{H}}_{BP}$ is lower triangular.

We use one particular $\mathbf{\Pi}_r$, giving a \mathbf{V}_m and information matrix for the BP significant effects that are consistent with the literature of balanced OP factorial designs. This permutation further defines the ordering of rows with the same heft. Essentially, the ordering takes the lex ordering of the rows of similar heft, but then reverses that order. For this reason, we will refer to it as the *contra-lexicographic ordering* (contralex). Note this is not revlex ordering, which is lex ordering from right to left. We define contralex ordering in a similar way as lexicographic, but with a minor modification. Let (j_1, \dots, j_m) and (k_1, \dots, k_m) be two rows of \mathbf{A}_m with the same heft. Then the row (j_1, \dots, j_m) comes before (k_1, \dots, k_m) in $\mathbf{V}_m = \mathbf{\Pi}_r \mathbf{A}_m$ if there exists a $v = 1, \dots, m$ where $j_v > k_v$ and for all $i < v$, $j_i = k_i$. In lexicographic ordering, the condition would be $j_v < k_v$.

One example of the contralex ordering is \mathbf{V}_3 in (5.4). For any m factors, the heft 1 rows in \mathbf{V}_m are organized into the identity \mathbf{I}_m . The heft 2 rows when $m = 4$ would be arranged as

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Henceforth, the order of the rows of \mathbf{V}_m will correspon to the contralex ordering of the treatment indices. We have shown that for this $\mathbf{\Pi}_r$, $\tilde{\mathbf{H}}_{BP} = \mathbf{\Pi}_r \mathbf{H}_{BP} \mathbf{\Pi}_r^T$ is lower triangular, so the submatrix \mathbf{H}_{22} of $\tilde{\mathbf{H}}_{BP}$ and $\tilde{\mathbf{H}}_{11} = \mathbf{H}_{11}$ clearly satisfy the invertibility conditions. The corresponding $\tilde{\boldsymbol{\tau}}_1$ will first be the baseline mean, then the heft 1 treatment combinations, and then the heft 2 treatment combinations.

For sake of notational brevity, henceforth drop the \sim notation so that \mathbf{H}_{BP} denotes the re-ordered transformation matrix and the subscripts of the parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ follow the contralex ordering. To avoid confusion about the highest order BP effects included in the reduced model, we denote the parameter vector of the BP intercept, main effects, and 2-f.i.'s as $\boldsymbol{\theta}_2$. While this conflicts with the general notation introduced in Section 5.1 where $\boldsymbol{\gamma}_2 = 0$, we feel that $\boldsymbol{\theta}_2$ better reflects the fact that we are including all BP effects up to the 2-f.i.'s. Here $\boldsymbol{\theta}_2$ plays the role of $\boldsymbol{\gamma}_1 = \mathbf{H}_1 \boldsymbol{\tau} = \mathbf{H}_{11} \boldsymbol{\tau}_1$ in the general framework of Section 5.1.

Let $\mathbf{V}_{m(h)}$ be the rows of \mathbf{V}_m with heft h . In light of Corollary 1.2.1 it is easy to see that

$$\mathbf{H}_{11} = \left(\begin{array}{c|c|c} 1 & \mathbf{V}_{m(0)} & \mathbf{0}_{\binom{m}{2}}^T \\ \hline -\mathbf{1}_m & \mathbf{V}_{m(1)} & \mathbf{0}_{m \times \binom{m}{2}} \\ \hline \mathbf{1}_{\binom{m}{2}} & -\mathbf{V}_{m(2)} & \mathbf{I}_{\binom{m}{2}} \end{array} \right).$$

Note that $\mathbf{V}_{m(0)} = \mathbf{0}_m^T$, $\mathbf{V}_{m(1)} = \mathbf{I}_m$, and $\mathbf{V}_{m(2)}^T \mathbf{V}_{m(2)} = (m-2)\mathbf{I}_m + \mathbf{J}_m$, so

$$\mathbf{W} = \mathbf{H}_{11}^T \mathbf{H}_{11} = \left(\begin{array}{c|c|c} \nu & -m\mathbf{1}_m^T & \mathbf{1}_{\binom{m}{2}}^T \\ \hline -m\mathbf{1}_m & (m-1)\mathbf{I}_m + \mathbf{J}_m & -\mathbf{V}_{m(2)}^T \\ \hline \mathbf{1}_{\binom{m}{2}} & -\mathbf{V}_{m(2)} & \mathbf{I}_{\binom{m}{2}} \end{array} \right). \quad (5.5)$$

We first find the class of permutation matrices Π_W , being those permutations that satisfy $\Pi \mathbf{W} \Pi^T = \mathbf{W}$ (see Theorem 2.5.1; here $\mathbf{P}_{\tau_1} = \mathbf{I}_{p_1}$). It follows from matrix averaging that finding those $\mathbf{C}_{d\tau_1} = \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q}$ that are invariant to such permutations may be good candidates for optimal designs.

If $\Pi \in \Pi_W$ then it must preserve the diagonal elements of \mathbf{W} , which are ν , m , and 1. This

implies that $\mathbf{\Pi}$ must permute within the partitions of \mathbf{W} , or that $\mathbf{\Pi}$ has the form

$$\mathbf{\Pi} = \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline 0 & \mathbf{\Pi}_m & 0 \\ \hline 0 & 0 & \mathbf{\Pi}_{\binom{m}{2}} \end{array} \right) ,$$

where $\mathbf{\Pi}_m$ and $\mathbf{\Pi}_{\binom{m}{2}}$ are permutation matrices of size m and $\binom{m}{2}$, respectively. Since $(m-1)\mathbf{I}_m + \mathbf{J}_m$ and $\mathbf{I}_{\binom{m}{2}}$ are completely symmetric, and so invariant to all permutations, we only need to find a $\mathbf{\Pi}_{\binom{m}{2}}$ for a given $\mathbf{\Pi}_m$ such that $\mathbf{\Pi}_{\binom{m}{2}} \mathbf{V}_{m(2)} \mathbf{\Pi}_m^T = \mathbf{V}_{m(2)}$. Equivalently, we must find a $\mathbf{\Pi}_{\binom{m}{2}}$ such that $\mathbf{\Pi}_{\binom{m}{2}} \mathbf{V}_{m(2)} = \mathbf{V}_{m(2)} \mathbf{\Pi}_m$. Note that $\mathbf{V}_{m(2)}$ is an $\binom{m}{2} \times m$ array containing all heft two rows exactly once. Hence it is a BA of strength m with $n_2 = 1$ and all other indices equal to zero. Clearly $\mathbf{V}_{m(2)} \mathbf{\Pi}_m$ is also a BA of strength m with the same heft index. Since the rows of $\mathbf{V}_{m(2)} \mathbf{\Pi}_m$ are all unique, there exists one and only one $\mathbf{\Pi}_{\binom{m}{2}}$ such that $\mathbf{\Pi}_{\binom{m}{2}} \mathbf{V}_{m(2)} = \mathbf{V}_{m(2)} \mathbf{\Pi}_m$. Hence there exists $m!$ permutation matrices $\mathbf{\Pi} \in \Pi_W$, which are in one-to-one correspondence with the $m!$ permutations $\mathbf{\Pi}_m$, and they all satisfy

$$\mathbf{\Pi} \left(\begin{array}{c} \mathbf{V}_{m(0)} \\ \hline \mathbf{V}_{m(1)} \\ \hline \mathbf{V}_{m(2)} \end{array} \right) = \left(\begin{array}{c} \mathbf{V}_{m(0)} \\ \hline \mathbf{V}_{m(1)} \\ \hline \mathbf{V}_{m(2)} \end{array} \right) \mathbf{\Pi}_m .$$

We previously called such permutation matrices *factor permutations*, although here they are defined only up to the heft two rows of \mathbf{V}_m .

Eigenvalues of the weighted information matrix $\mathbf{C}_{dW\tau_1}$ are equivalent to those of $\mathbf{C}_{d\theta_2} = \mathbf{H}_{11}^{-1T} \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q} \mathbf{H}_{11}^{-1}$. As $\mathbf{\Pi} \in \Pi_W$ must permute within the partitioned submatrices of \mathbf{W} shown in (5.5), and the same partitioned submatrices for \mathbf{H}_{11}^{-1} have a similar form as those for \mathbf{W} , it is easily shown that if $\mathbf{\Pi} \in \Pi_W$ then $\mathbf{\Pi} \mathbf{H}_{11}^{-1} \mathbf{\Pi}^T = \mathbf{H}_{11}^{-1}$. This may be used to prove the following lemma, which is offered without proof.

Lemma 5.2.1. *The matrix $\mathbf{C}_{d\tau_1} = \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q}$ is invariant to simultaneous row and column permutations for those $\mathbf{\Pi} \in \Pi_W$ if and only if $\mathbf{C}_{d\theta_2} = \mathbf{H}_{11}^{-1T} \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q} \mathbf{H}_{11}^{-1}$ is also invariant to the same permutations.*

It follows from Lemma 5.2.1 that instead of finding forms of $\mathbf{C}_{d\tau_1}$ that are invariant to permutations in Π_W , we may find forms of $\mathbf{C}_{d\theta_2}$. We choose to work with the alternate representation of $\mathbf{C}_{d\theta_2} = \mathbf{K} \mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T \mathbf{X}_d \mathbf{H}_{BP}^{-1} \mathbf{K}^T$. Write $\mathbf{V}_m = (\mathbf{v}_1 | \mathbf{v}_2 | \dots | \mathbf{v}_m)$ and define the vector $\mathbf{v}_{ij} = \mathbf{v}_i \odot \mathbf{v}_j$, the Hadamard product of \mathbf{v}_i and \mathbf{v}_j . Note that these are also 0/1 vectors, and elements of these vectors equal 0 unless the corresponding elements of \mathbf{v}_i and \mathbf{v}_j are both 1. Finally, let

$$\mathbf{V}_{m;2} = (\mathbf{v}_{12} | \mathbf{v}_{13} | \dots | \mathbf{v}_{1m} | \mathbf{v}_{23} | \dots | \mathbf{v}_{(m-1)m}) ,$$

which has all $\binom{m}{2}$ Hadamard products of any two columns of \mathbf{V}_m . This should not be confused with $\mathbf{V}_{m(2)}$, which is the $\binom{m}{2} \times m$ array of all left 2 rows of \mathbf{V}_m . The ordering of the \mathbf{v}_{ij} in $\mathbf{V}_{m;2}$ corresponds to the contralex ordering of the rows of $\mathbf{V}_{m;2}$. For example, the first row of $\mathbf{V}_{m;2}$ is $(1, 1, 0, \dots, 0)$ and the first column of $\mathbf{V}_{m;2}$ is $\mathbf{v}_1 \odot \mathbf{v}_2 = \mathbf{v}_{12}$.

Note that \mathbf{H}_{BP}^{-1} serves to transform $\boldsymbol{\theta}$ to $\boldsymbol{\tau}$. The reduced transformation matrix $\mathbf{H}_{BP}^{-1} \mathbf{K}^T$ is the first $\nu = 1 + m + \binom{m}{2}$ columns of \mathbf{H}_{BP}^{-1} , and by Lemma 1.2.1 (i), it follows that

$$\mathbf{H}_{BP}^{-1} \mathbf{K}^T = (\mathbf{1}_{2^m} | \mathbf{V}_m | \mathbf{V}_{m;2}) .$$

Let \mathbf{X}_d be the $N \times 2^m$, 0/1 design matrix where $\mathbf{X}_d \mathbf{1}_{2^m} = \mathbf{1}_N$, so

$$\begin{aligned} \mathbf{X}_d \mathbf{H}_{BP}^{-1} \mathbf{K}^T &= \mathbf{X}_d (\mathbf{1}_{2^m} \mid \mathbf{V}_m \mid \mathbf{V}_{m;2}) \\ &= (\mathbf{1}_N \mid \mathbf{X}_d \mathbf{V}_m \mid \mathbf{X}_d \mathbf{V}_{m;2}) \\ &= (\mathbf{1}_N \mid \mathbf{V}_d \mid \mathbf{V}_{d2}) , \end{aligned} \quad (5.6)$$

where $\mathbf{V}_d = \mathbf{X}_d \mathbf{V}_m = (\mathbf{v}_{d1} \mid \mathbf{v}_{d2} \mid \dots \mid \mathbf{v}_{dm})$ is an $N \times m$ array of 0's and 1's and $\mathbf{V}_{d2} = \mathbf{X}_d \mathbf{V}_{m;2} = (\mathbf{v}_{d12} \mid \mathbf{v}_{d13} \mid \dots \mid \mathbf{v}_{d(m-1)m})$, where $\mathbf{v}_{dij} = \mathbf{v}_{di} \odot \mathbf{v}_{dj}$. Finding a design in terms of \mathbf{X}_d is equivalent to finding a design in terms of $\mathbf{X}_d \mathbf{V}_m = \mathbf{V}_d$, which tells us the necessary treatment combinations to replicate. It follows that

$$\mathbf{C}_{d\theta_2} = \mathbf{K} \mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T \mathbf{X}_d \mathbf{H}_{BP}^{-1} \mathbf{K}^T = \left(\begin{array}{c|c|c} N & \mathbf{1}_N^T \mathbf{V}_d & \mathbf{1}_N^T \mathbf{V}_{d2} \\ \hline & \mathbf{V}_d^T \mathbf{V}_d & \mathbf{V}_d^T \mathbf{V}_{d2} \\ \hline & & \mathbf{V}_{d2}^T \mathbf{V}_{d2} \end{array} \right) , \quad (5.7)$$

where the empty partitions are implied by symmetry. Let $\mathbf{C}_{11} = \mathbf{V}_d^T \mathbf{V}_d$, $\mathbf{C}_{12} = \mathbf{V}_d^T \mathbf{V}_{d2}$, $\mathbf{C}_{21} = \mathbf{C}_{12}^T$, and $\mathbf{C}_{22} = \mathbf{V}_{d2}^T \mathbf{V}_{d2}$, where we drop the d for notational brevity. This notation will be used throughout this section and it should be understood that the submatrices pertain to a specific design.

If $\mathbf{\Pi} \in \Pi_W$ then

$$\begin{aligned} \mathbf{X}_d \mathbf{H}_{BP}^{-1} \mathbf{K}^T \mathbf{\Pi}^T &= (\mathbf{1}_N \mid \mathbf{V}_d \mid \mathbf{V}_{d2}) \mathbf{\Pi}^T \\ &= (\mathbf{1}_N \mid \mathbf{V}_d \mathbf{\Pi}_m^T \mid \mathbf{V}_{d2} \mathbf{\Pi}_{\binom{m}{2}}^T) \\ &= (\mathbf{1}_N \mid \mathbf{V}_d^* \mid \mathbf{V}_{d2}^*) , \end{aligned}$$

where \mathbf{V}_d^* is a column permutation of \mathbf{V}_d and will be written as $\mathbf{V}_d^* = (\mathbf{v}_{d1}^* \mid \mathbf{v}_{d2}^* \mid \dots \mid \mathbf{v}_{dm}^*)$.

Even though \mathbf{V}_{d2}^* is a column permutation of \mathbf{V}_{d2} , it can be constructed the same way as \mathbf{V}_{d2} but with the columns of \mathbf{V}_d^* rather than \mathbf{V}_d . We use this and the following lemma to derive those designs \mathbf{V}_d (which is equivalent to specifying an \mathbf{X}_d) having $\mathbf{C}_{d\theta_2}$ that are invariant to $\mathbf{\Pi} \in \Pi_W$.

Lemma 5.2.2. *Let \mathbf{B} be an $N \times m$ BA with two symbols (say 0 and 1) of strength $t < m$ with index set (n_0, n_1, \dots, n_t) . Suppose that for every $N \times (t+1)$ subarray the row $\mathbf{1}_{t+1}^T$ appears the same number of times, say \tilde{n}_{t+1} . Then \mathbf{B} is a BA of strength $t+1$ with index set $(\tilde{n}_0, \tilde{n}_1, \dots, \tilde{n}_{t+1})$ where $\tilde{n}_g = n_g + \sum_{i=g+1}^{t+1} (-1)^{i-g} n_i$ for $g = 0, \dots, t$.*

Proof. Consider any $N \times (t+1)$ subarray, \mathbf{B}_{t+1} , of \mathbf{B} . Then by assumption the row $\mathbf{1}_{t+1}^T$ appears n_{t+1} times. But \mathbf{B}_{t+1} must also be a BA of at least strength t and so every $N \times t$ subarray, say \mathbf{B}_t , of \mathbf{B}_{t+1} has the row $\mathbf{1}_t^T$ appear n_t times. Note that these n_t $\mathbf{1}_t^T$ rows in \mathbf{B}_t , either increase in heft in \mathbf{B}_{t+1} or maintain the same heft. Hence $n_{t+1} \leq n_t$ and we know that n_{t+1} of these n_t correspond to the $\mathbf{1}_{t+1}^T$ rows of \mathbf{B}_{t+1} . It then follows that there are $n_t - n_{t+1}$ rows of \mathbf{B}_{t+1} that have t ones (within the columns of \mathbf{B}_t) and one zero. Since each heft t run corresponds to a specific \mathbf{B}_t , all heft t rows appear $n_t - n_{t+1} = n_t + (-1)^{t+1-t} n_{t+1}$ times in \mathbf{B}_{t+1} .

Now assume for any fixed $g < t$ that every heft $g+1$ row appears $\tilde{n}_{g+1} = n_{g+1} + \sum_{i=g+1}^{t+1} (-1)^{i-g} n_i$ times in every $N \times (t+1)$ subarray. For the \mathbf{B}_{t+1} array above, since $g < t < t+1$, every heft g row in \mathbf{B}_{t+1} must correspond to a heft g row for some $N \times t$ subarray \mathbf{B}_t of \mathbf{B}_{t+1} . Conversely, any given heft g row in \mathbf{B}_t has either heft g or heft $g+1$ in \mathbf{B}_{t+1} . This of course depends on the column of \mathbf{B}_{t+1} not included in \mathbf{B}_t . Since \mathbf{B} is a BA of strength t , any given heft $g < t$ row appears n_g times in \mathbf{B}_t . By assumption every heft $g+1$ row appears in \mathbf{B}_{t+1} , \tilde{n}_{g+1} times, so there are $n_g - (n_{g+1} + \sum_{i=g+2}^{t+1} (-1)^{i-(g+1)} n_i)$ rows in \mathbf{B}_t that must have heft

g in \mathbf{B}_{t+1} . But this is equal to

$$n_g - \tilde{n}_{g+1} = (n_{g+1} + \sum_{i=g+2}^{t+1} (-1)^{i-(g+1)} n_i) = n_g + \sum_{i=g+1}^{t+1} (-1)^{i-g} n_i = \tilde{n}_g .$$

Hence every type of heft $g < t$ row in \mathbf{B}_{t+1} appears \tilde{n}_g times. By induction we have the result. \square

Theorem 5.2.1. *The reduced information matrix $\mathbf{C}_{d\theta_2} = \mathbf{K}\mathbf{H}_{BP}^{-1T}\mathbf{X}_d^T\mathbf{X}_d\mathbf{H}_{BP}^{-1}\mathbf{K}^T$ is invariant to all $\mathbf{\Pi} \in \Pi_W$ if and only if $\mathbf{X}_d\mathbf{V}_m = \mathbf{V}_d$ is a BA of at least strength 4.*

Proof. The m diagonals of \mathbf{C}_{11} (defined after (5.7)) are $\mathbf{v}_{di}^T\mathbf{v}_{di} = \mathbf{1}_N^T\mathbf{v}_{di}$ and the $\binom{m}{2}$ off-diagonals are $\mathbf{v}_{di}^T\mathbf{v}_{dj} = \mathbf{1}_N^T(\mathbf{v}_{di} \odot \mathbf{v}_{dj})$. If the diagonal elements are invariant to factor permutations then $\mathbf{1}_N^T\mathbf{v}_{di} = \mathbf{1}_N^T\mathbf{v}_{di'}$ for some i' , which may or may not be equal to i . Hence every column of \mathbf{V}_d must have the same number of 1's, so \mathbf{V}_d must be a BA of at least strength 1. Furthermore, the off-diagonals must satisfy $\mathbf{1}_N^T(\mathbf{v}_{di} \odot \mathbf{v}_{dj}) = \mathbf{1}_N^T(\mathbf{v}_{di'} \odot \mathbf{v}_{dj'})$ for some i', j' . Both of these count the number of (1, 1) in the $N \times 2$ subarray of \mathbf{V}_d . Since \mathbf{V}_d is at least a BA of strength 1, Lemma 5.2.2 implies that \mathbf{V}_d must be a BA of at least strength 2.

Assume $i < j$, $k < l$, $i \leq k$, and $j \leq l$. Then any element of \mathbf{C}_{22} can be written as

$$\begin{aligned} \mathbf{v}_{dij}^T\mathbf{v}_{dkl} &= (\mathbf{v}_{di} \odot \mathbf{v}_{dj})^T(\mathbf{v}_{dk} \odot \mathbf{v}_{dl}) \\ &= \mathbf{1}_N^T(\mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{dk} \odot \mathbf{v}_{dl}) . \end{aligned}$$

There are five cases to consider:

1. $i = k, j = l$
2. $i = k, j \neq l$
3. $i \neq k, j = k$

4. $i \neq k, j = l$

5. $i \neq k, j \neq l$.

Case (1) corresponds to the diagonal elements of \mathbf{C}_{22} , which are

$$\begin{aligned}
\mathbf{v}_{dij}^T \mathbf{v}_{dkl} &= \mathbf{v}_{dij}^T \mathbf{v}_{dij} \\
&= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{di} \odot \mathbf{v}_{dj}) \\
&= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{dj}) \\
&= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{dj}) ,
\end{aligned}$$

since the Hadamard product is commutative and $\mathbf{v}_{di} \odot \mathbf{v}_{di} = \mathbf{v}_{di}$ for all $i = 1, \dots, m$. These are equal to the off-diagonals of \mathbf{C}_{11} , and so must be equal for the reduced information matrix to be invariant to factor permutations.

Similarly, cases (2), (3), and (4) may be written as

$$\begin{aligned}
\mathbf{v}_{dij}^T \mathbf{v}_{dil} &= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{dl}) , \\
\mathbf{v}_{dij}^T \mathbf{v}_{djl} &= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{dl}) , \\
\mathbf{v}_{dij}^T \mathbf{v}_{dkj} &= \mathbf{1}_N^T (\mathbf{v}_{di} \odot \mathbf{v}_{dj} \odot \mathbf{v}_{dk}) ,
\end{aligned}$$

respectively. Note that cases (2) and (3) are equivalent and all three cases are the sum of the elements of the Hadamard product of three columns of \mathbf{V}_d . It follows that \mathbf{C}_{22} is invariant to all $\Pi_{\binom{m}{2}}$ if and only if the sum of the elements of the Hadamard product of *any three columns* of \mathbf{V}_d is the same. Since the Hadamard product gives a vector of 0's and 1's, every $N \times 3$ subarray of \mathbf{V}_d must have the row $(1, 1, 1)$ appear the same number of times. Lemma 5.2.2 then implies \mathbf{V}_d must be a BA of at least strength 3.

The same argument may be applied to case (5), and we may again use Lemma 5.2.2 to prove

that \mathbf{V}_d must be a BA of at least strength 4. It is straightforward to show the converse. \square

Unless otherwise stated, we will now index the BP main effects by the factors they involve, $i = 1, \dots, m$, and BP 2-f.i.'s by the two factors they involve jk , where $j = 1, \dots, m-1$; $k = 2, \dots, m$; and $j < k$. The concepts of factor sets and overlapping introduced in Section 1.2.3 will be needed for many proofs to follow and so will now be briefly demonstrated for a BP 2-f.i. denoted by jk . The factor set for a BP 2-f.i. denoted by jk will be $\{j, k\}$, as those are the two factors involved in that effect. This effect overlaps by two factors with itself; overlaps by one factor with all other 2-f.i.'s involving either factor j or k ; and does not overlap with 2-f.i.'s that involve neither j nor k .

Optimality criteria are frequently based on the eigenvalues of the matrix, so it is useful to derive a general expression for the eigenvalues of $\mathbf{C}_{d\theta_2}$ for m factors and a BA of strength 4, \mathbf{V}_d , with index set $(n_0, n_1, n_2, n_3, n_4)$. The form of such a $\mathbf{C}_{d\theta_2}$ may be represented using the following elements which are functions of the elements of the BA index set

$$\begin{aligned}
c_0 &= n_0 + 4n_1 + 6n_2 + 4n_3 + n_4 = N, & \text{the number of runs} \\
c_1 &= n_1 + 3n_2 + 3n_3 + n_4, & \# \text{ 1's per column} \\
c_2 &= n_2 + 2n_3 + n_4, & \# (1, 1)\text{'s per } N \times 2 \text{ subarray of } \mathbf{V}_d \\
c_3 &= n_3 + n_4, & \# (1, 1, 1)\text{'s per } N \times 3 \text{ subarray of } \mathbf{V}_d \\
c_4 &= n_4, & \# (1, 1, 1, 1)\text{'s per } N \times 4 \text{ subarray of } \mathbf{V}_d.
\end{aligned} \tag{5.8}$$

It is easy to see that $\mathbf{1}_N^T \mathbf{V}_d = c_1 \mathbf{1}_m^T$, $\mathbf{1}_N^T \mathbf{V}_{d2} = c_2 \mathbf{1}_{\binom{m}{2}}^T$, and $\mathbf{C}_{11} = (c_1 - c_2) \mathbf{I}_m + c_2 \mathbf{J}_m$. The $m \times \binom{m}{2}$ matrix \mathbf{C}_{12} has rows and columns corresponding to the BP main effects and 2-f.i.'s, respectively. The elements of \mathbf{C}_{12} are

$$(\mathbf{C}_{12})_{i,jk} = \begin{cases} c_2 & \text{if BP main effect } i \text{ and BP 2-f.i. } jk \text{ overlap} \\ c_3 & \text{otherwise} \end{cases}.$$

Clearly the effects overlap if $i = j$ or $i = k$. It follows that each row of \mathbf{C}_{12} has $m - 1$ c_2 's and $\binom{m-1}{2}$ c_3 's. Also, each column has 2 c_2 's and $m - 2$ c_3 's, which are the rows of \mathbf{C}_{12}^T .

The $\binom{m}{2} \times \binom{m}{2}$ symmetric matrix \mathbf{C}_{22} has rows and columns corresponding to the BP 2-f.i.'s. For brevity, we will simply say jk rather than BP 2-f.i. jk in the following representation of \mathbf{C}_{22}

$$(\mathbf{C}_{22})_{jk,j'k'} = \begin{cases} c_2 & \text{if } jk \text{ and } j'k' \text{ overlap by 2 factors} \\ c_3 & \text{if } jk \text{ and } j'k' \text{ overlap by 1 factor} \\ c_4 & \text{otherwise} \end{cases} \quad .$$

Note that the c_2 elements correspond to the diagonal elements of \mathbf{C}_{22} , since BP 2-f.i.'s overlap if and only if $jk = j'k'$. Each row/column in \mathbf{C}_{22} has one c_2 , $\binom{m-2}{2}$ c_4 's, and $\binom{m}{2} - \binom{m-2}{2} - 1 = 2m - 4 = 2(m - 2)$ c_3 's.

We demonstrate the form of $\mathbf{C}_{d\theta_2}$ when \mathbf{V}_d is a BA of strength 4, for $m = 4$ factors:

$$\mathbf{C}_{d\theta_2} = \left(\begin{array}{c|cccc|cccccc} c_0 & c_1 & c_1 & c_1 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 \\ \hline & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_3 & c_3 & c_3 \\ & c_2 & c_1 & c_2 & c_2 & c_2 & c_3 & c_3 & c_2 & c_2 & c_3 \\ & c_2 & c_2 & c_1 & c_2 & c_3 & c_2 & c_3 & c_2 & c_3 & c_2 \\ & c_2 & c_2 & c_2 & c_1 & c_3 & c_3 & c_2 & c_3 & c_2 & c_2 \\ \hline & & & & & c_2 & c_3 & c_3 & c_3 & c_3 & c_4 \\ & & & & & c_3 & c_2 & c_3 & c_3 & c_4 & c_3 \\ & & & & & c_3 & c_3 & c_2 & c_4 & c_3 & c_3 \\ & & & & & c_3 & c_3 & c_4 & c_2 & c_3 & c_3 \\ & & & & & c_3 & c_4 & c_3 & c_3 & c_2 & c_3 \\ & & & & & c_4 & c_3 & c_3 & c_3 & c_3 & c_2 \end{array} \right) , \quad (5.9)$$

5.2.1 Eigenvalues of $\mathbf{C}_{d\theta_2}$ for strength 4 BAs

The information matrix $\mathbf{C}_{d\theta_2}$ must be nonsingular to estimate all significant BP effects and so has ν positive eigenvalues that we wish to calculate. The following lemmas address this problem; their proofs rely on association schemes linked to the concept of overlapping introduced in Section 1.2.3. Many of the results are similar to those of Srivastava and Chopra (1971b), who worked with reduced models assuming OP 3-f.i.'s and higher were insignificant.

Lemma 5.2.3. *Three eigenvectors for a nonsingular $\mathbf{C}_{d\theta_2}$ where \mathbf{V}_d is a BA of strength 4 have the form $\mathbf{a} = (a_1 \mid a_2 \mathbf{1}_m^T \mid a_3 \mathbf{1}_{\binom{m}{2}}^T)^T$, where a_1, a_2, a_3 are constants.*

Proof. It is clear that for a given submatrix in the partitioning of $\mathbf{C}_{d\theta_2}$ shown in (5.7), the elements in each row sum to the same quantity. In light of this, consider a vector of the form $\mathbf{a} = (a_1 \mid a_2 \mathbf{1}_m^T \mid a_3 \mathbf{1}_{\binom{m}{2}}^T)^T$, which takes these row sums, so that

$$\mathbf{C}_{d\theta_2} \mathbf{a} = \begin{pmatrix} a_1^* \\ \frac{a_2^* \mathbf{1}_m}{a_3^* \mathbf{1}_{\binom{m}{2}}} \end{pmatrix} = \mathbf{a}^*$$

where

$$\begin{aligned} a_1^* &= c_0 a_1 + m c_1 a_2 + \binom{m}{2} c_2 a_3 \\ a_2^* &= c_1 a_1 + [c_1 + (m-1)c_2] a_2 + [(m-1)c_2 + \binom{m-1}{2} c_3] a_3 \\ a_3^* &= c_2 a_1 + [2c_2 + (m-2)c_3] a_2 + [c_2 + 2(m-2)c_3 + \binom{m-2}{2} c_4] a_3 \end{aligned} \quad .$$

If $\mathbf{a}^* = \lambda \mathbf{a}$ then \mathbf{a} is an eigenvector of $\mathbf{C}_{d\theta_2}$. That is, we want to find a_1 , a_2 , and a_3 satisfying $a_1^* = \lambda a_1$, $a_2^* = \lambda a_2$, and $a_3^* = \lambda a_3$. We can reduce the problem down to an

eigenvalue-eigenvector problem of the 3×3 matrix

$$\mathbf{M} = \begin{pmatrix} c_0 & mc_1 & \binom{m}{2}c_2 \\ c_1 & c_1 + (m-1)c_2 & (m-1)c_2 + \binom{m-1}{2}c_3 \\ c_2 & 2c_2 + (m-2)c_3 & c_2 + 2(m-2)c_3 + \binom{m-2}{2}c_4 \end{pmatrix}.$$

If λ and $\mathbf{b} = (b_1, b_2, b_3)$ is an eigenvalue-eigenvector pair of \mathbf{M} then λ and $\mathbf{a} = (b_1 \mid b_2 \mathbf{1}_m^T \mid b_3 \mathbf{1}_{\binom{m}{2}}^T)^T$ is an eigenvalue-eigenvector pair of $\mathbf{C}_{d\theta_2}$. Because \mathbf{M} is square, we know there are three eigenvalue-eigenvector pairs for \mathbf{M} and clearly every eigenvector is of the form $\mathbf{b} \in \mathcal{R}^3$. Hence three of the eigenvectors of \mathbf{C}_{θ_2} are of the form \mathbf{a} . \square

We know that $\mathbf{C}_{d\theta_2}$ is symmetric and positive definite, so its ν eigenvalues are real and positive, which implies that the eigenvalues of \mathbf{M} are real and positive. To find the corresponding eigenvalues, we turn to the characteristic polynomial of \mathbf{M} . Let \mathbf{M}_{ij} be the 2×2 submatrix of \mathbf{M} with rows/columns i and j removed. Then the characteristic polynomial for \mathbf{M} is

$$P_M(t) = t^3 - E_1 t^2 + E_2 t - E_3,$$

where

$$E_1 = \text{tr}(\mathbf{M})$$

$$E_2 = |\mathbf{M}_{11}| + |\mathbf{M}_{22}| + |\mathbf{M}_{33}|$$

$$E_3 = |\mathbf{M}|,$$

which are known as the principal minors of \mathbf{M} . If we focus on A -optimality, we do not need to solve for the eigenvalues individually. Instead, all we need is the sum of the inverse eigenvalues of \mathbf{M} , or the sum of the inverse roots of $P_M(t)$. Let $P(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_0$ be a polynomial of degree n with no roots equal to zero and consider the polynomial $Q(t) =$

$a_0t^n + a_1t^{n-1} + \cdots + a_n$. Then λ is a root of $P(t)$ if and only if λ^{-1} is a root of $Q(t)$.

The roots of $P_M(t)$ are positive reals, so the roots of the polynomial

$$Q_M(t) = -E_3t^3 + E_2t^2 - E_1t + 1$$

are the reciprocals of $P_M(t)$. By Vieta's formula, the sum of the roots of $Q_M(t)$ is

$$\sum_{i=1}^3 \lambda_i^{-1} = -\frac{E_2}{-E_3} = \frac{E_2}{E_3}.$$

It can be shown that since c_0 equals the number of available runs, N , and is fixed for a given design, the values of $2E_2$ and $2E_3$ are:

$$\begin{aligned} 2E_2 &= 2Nc_1 + 2mNc_2 + 4(m-2)Nc_3 + (m-2)(m-3)Nc_4 - 2m^2c_1^2 + 2c_1c_2 \\ &\quad + 4(m-2)c_1c_3 + (m-2)(m-3)c_1c_4 - (m+2)(m-1)c_2^2 \\ &\quad + (m-3)(m-2)(m-1)c_2c_4 - (m-2)^2(m-1)c_3^2 \end{aligned} \quad (5.10)$$

$$\begin{aligned} 2E_3 &= 2Nc_1c_2 + 4(m-2)Nc_1c_3 + (m-3)(m-2)Nc_1c_4 - 2(m-1)Nc_2^2 \\ &\quad + (m-3)(m-2)(m-1)Nc_2c_4 - (m-2)^2(m-1)Nc_3^2 + 2mc_1^2c_2 \\ &\quad - 4m(m-2)c_1^2c_3 - m(m-2)(m-1)c_1^2c_4 + 3(m+1)(m-1)c_1c_2^2 \\ &\quad + 2m(m-2)(m-1)c_1c_2c_3 - m(m-1)^2c_2^3. \end{aligned} \quad (5.11)$$

Now we must find the remaining $\nu - 3$ eigenvalues, which must have eigenvectors orthogonal to the three eigenvectors in Lemma 5.2.3. Vectors of the general form $\mathbf{e} = (0 \mid \mathbf{e}_m^T \mid \mathbf{e}_{\binom{m}{2}}^T)^T$, where $\mathbf{e}_m^T \mathbf{1}_m = 0$ and $\mathbf{e}_{\binom{m}{2}}^T \mathbf{1}_{\binom{m}{2}} = 0$ clearly satisfy this. We first look for potential eigenvectors

of the form $\mathbf{e} = (0 \mid \mathbf{0}_m^T \mid \mathbf{e}_{\binom{m}{2}}^T)^T$. Post-multiplying $\mathbf{C}_{d\theta_2}$ by such an \mathbf{e} gives

$$\mathbf{C}_{d\theta_2}\mathbf{e} = \begin{pmatrix} \frac{c_2 \mathbf{1}_{\binom{m}{2}}^T \mathbf{e}_{\binom{m}{2}}}{\mathbf{C}_{12} \mathbf{e}_{\binom{m}{2}}} \\ \frac{\mathbf{C}_{22} \mathbf{e}_{\binom{m}{2}}}{\mathbf{C}_{22} \mathbf{e}_{\binom{m}{2}}} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\mathbf{C}_{12} \mathbf{e}_{\binom{m}{2}}}{\mathbf{C}_{22} \mathbf{e}_{\binom{m}{2}}} \end{pmatrix}.$$

If \mathbf{e} is an eigenvector, then $\mathbf{C}_{12} \mathbf{e}_{\binom{m}{2}} = \mathbf{0}_m$ and $\mathbf{C}_{22} \mathbf{e}_{\binom{m}{2}} = \lambda \mathbf{e}_{\binom{m}{2}}$ for some λ , which implies that λ and $\mathbf{e}_{\binom{m}{2}}$ are an eigenvalue-eigenvector pair of \mathbf{C}_{22} . If $\mathbf{C}_{d\theta_2}$ is nonsingular then it must be positive definite, which implies that $\mathbf{C}_{22} = \mathbf{V}_{d2}^T \mathbf{V}_{d2}$ must also be positive definite and hence have $\binom{m}{2}$ positive eigenvalues. The goal now is to find eigenvectors of \mathbf{C}_{22} that also satisfy $\mathbf{C}_{12} \mathbf{e}_{\binom{m}{2}} = \mathbf{0}_m$ and $\mathbf{1}_{\binom{m}{2}}^T \mathbf{e}_{\binom{m}{2}} = 0$.

Recall that the elements of \mathbf{C}_{22} correspond to “products” of BP 2-f.i.’s. The diagonal elements all equal c_2 , and the off-diagonals equal c_3 if the corresponding 2-f.i.’s overlap by one factor, and c_4 if they do not overlap. Consider the matrix $\tilde{\mathbf{C}}_{22} = \frac{1}{n_3}(\mathbf{C}_{22} - n_4 \mathbf{J})$ which is easily shown to have elements

$$(\tilde{\mathbf{C}}_{22})_{jk,j'k'} = \begin{cases} \frac{n_2+2n_3}{n_3} & \text{if } jk \text{ and } j'k' \text{ overlap by 2 factors} \\ 1 & \text{if } jk \text{ and } j'k' \text{ overlap by 1 factor} \\ 0 & \text{otherwise} \end{cases}.$$

Let \mathbf{A}_{22} be the adjacency matrix for the BP 2-f.i.’s with elements

$$(\mathbf{A}_{22})_{jk,j'k'} = \begin{cases} 1 & \text{if } jk \text{ and } j'k' \text{ overlap by 1 factor} \\ 0 & \text{otherwise} \end{cases}.$$

It then follows that $\tilde{\mathbf{C}}_{22} = \frac{n_2+2n_3}{n_3} \mathbf{I} + \mathbf{A}_{22}$. If $\mathbf{e}_{\binom{m}{2}}$ is a contrast vector, and an eigenvector of

$\tilde{\mathbf{C}}_{22}$ with eigenvalue λ , then

$$\begin{aligned}\lambda \mathbf{e}_{\binom{m}{2}} &= \tilde{\mathbf{C}}_{22} \mathbf{e}_{\binom{m}{2}} \\ &= \frac{1}{n_3} \mathbf{C}_{22} \mathbf{e}_{\binom{m}{2}} ,\end{aligned}$$

implying that $n_3\lambda$ and $\mathbf{e}_{\binom{m}{2}}$ is an eigenvalue-eigenvector pair of \mathbf{C}_{22} . The goal then is to find eigenvectors of $\tilde{\mathbf{C}}_{22}$ that are contrasts and then check to see if they further satisfy $\mathbf{C}_{12} \mathbf{e}_{\binom{m}{2}} = \mathbf{0}_m$.

Clearly any eigenvector of \mathbf{A}_{22} is also an eigenvector of \mathbf{C}_{22} . If λ is an eigenvalue of \mathbf{A}_{22} , it follows that $n_3 \left(\frac{n_2+2n_3}{n_3} + \lambda \right)$ is an eigenvalue of \mathbf{C}_{22} . Note that $\mathbf{A}_{22} \mathbf{1} = (2m-4)\mathbf{1}$, since every 2-f.i. overlaps by one factor with $(m-2) + (m-2) = 2m-4$ other 2-f.i.'s. The other eigenvectors of \mathbf{A}_{22} must be orthogonal to $\mathbf{1}$ so they must be contrasts. We now derive the form of \mathbf{A}_{22}^2 to show that \mathbf{A}_{22} is the adjacency matrix for a strongly regular graph. The adjacency matrices for such graphs have a known form for the eigenvalues.

Denote the 0/1 columns of \mathbf{A}_{22} as \mathbf{a}_{jk} corresponding to the BP 2-f.i. ij . Clearly $\mathbf{a}_{ij}^T \mathbf{a}_{kl} = (\mathbf{a}_{ij} \odot \mathbf{a}_{kl})^T \mathbf{1}$. Note that all Hadamard products will be 0/1 vectors. These products have 1 elements for those 2-f.i.'s that overlap by one factor *for both* ij and kl . Assume $i < j$ and $k < l$, and we again have five cases to consider for the elements of \mathbf{A}_{22}^2 :

1. $i = k, j = l$
2. $i = k, j \neq l$
3. $i \neq k, j = k$
4. $i \neq k, j = l$
5. $i \neq k, j \neq l$.

Case (1) are the diagonal elements of \mathbf{A}_{22} and clearly $(\mathbf{a}_{ij} \odot \mathbf{a}_{ij})^T \mathbf{1} = 2m - 4$, the number of 2-f.i.'s that overlap by 1 factor for any given 2-f.i. .

Cases (2) to (4) correspond to the Hadamard product of 2-f.i.'s that overlap by one factor. Say one of these 2-f.i.'s is ij . Of the $2m - 4 = 2(m - 2)$ 2-f.i.'s that overlap by one factor with ij , only $m - 2$ of them also overlap by one factor with the other 2-f.i. in the product. Hence, the off-diagonals of \mathbf{A}_{22}^2 corresponding to cases (2) to (4) equal $m - 2$.

Finally, case (5) corresponds to Hadamard products $\mathbf{a}_{ij} \odot \mathbf{a}_{kl}$, being 2-f.i.'s that do not overlap. Of the $2m - 4$ 2-f.i.'s that overlap by one factor for ij , only 4 overlap by one factor for kl , being the 2-f.i.'s ik, il, jk , and jl (assuming $j < k$ and $j < l$, but there is no loss of generality).

This system of overlapping may be used to develop a triangular association scheme of the BP 2-f.i.'s which has two associate classes. Such association schemes were first introduced by Bose and Shimamoto (1952) and results for such association schemes are well-developed. Bose (1963) and Brouwer and van Lint (1984) established that an association scheme with two associate classes corresponds to a strongly regular graph. A graph with v vertices is known as a strongly regular graph if

- Each vertex has degree k
- Every two adjacent vertices have λ common neighbors
- Every two non-adjacent vertices have μ common neighbors

The eigenvalues of the adjacency matrix for a strongly regular graph are known to be

1. k with multiplicity 1
2. $\frac{1}{2} \left[(\lambda - \mu) + \sqrt{(\lambda - \mu)^2 + 4(k - \mu)} \right]$, with multiplicity $\frac{1}{2} \left[(v - 1) - \frac{2k + (v-1)(\lambda - \mu)}{\sqrt{(\lambda - \mu)^2 + 4(k - \mu)}} \right]$

$$3. \frac{1}{2} \left[(\lambda - \mu) - \sqrt{(\lambda - \mu)^2 + 4(k - \mu)} \right], \text{ with multiplicity } \frac{1}{2} \left[(v - 1) + \frac{2k + (v-1)(\lambda - \mu)}{\sqrt{(\lambda - \mu)^2 + 4(k - \mu)}} \right]$$

The eigenvector for eigenvalue k is clearly $\mathbf{1}_v$ so the other eigenvectors must be contrast vectors, since they must be orthogonal to $\mathbf{1}_v$.

The elements of \mathbf{A}_{22}^2 imply that \mathbf{A}_{22} is the adjacency matrix for a strongly regular graph with $v = \binom{m}{2}$, $k = 2m - 4$, $\lambda = m - 2$, and $\mu = 4$, where $m > 2$. Hence the eigenvalues for \mathbf{A}_{22} are

1. $2m - 4$, with multiplicity 1
2. $m - 4$, with multiplicity $m - 1$
3. -2 , with multiplicity $\frac{m(m-3)}{2}$

It follows that there are 2 distinct eigenvalues of \mathbf{C}_{22} corresponding to contrast eigenvectors, $n_3 \left(\frac{n_2 + 2n_3}{n_3} - 2 \right) = n_2$ and $n_3 \left(\frac{n_2 + 2n_3}{n_3} + (m - 4) \right) = n_2 + (m - 2)n_3$ with multiplicity $\frac{m(m-3)}{2}$ and $m - 1$, respectively.

The eigenvectors with eigenvalue $m - 4$ for \mathbf{A}_{22} are found in the following way. Group the $\binom{m}{2}$ BP 2-f.i.'s into m groups of $(m - 1)$, 2-f.i.'s each, where group g , $g = 1, \dots, m$, consists of those 2-f.i.'s that overlap by one factor with BP main effect g . Note that any specific 2-f.i. will be in exactly two groups, and any two groups share exactly one 2-f.i.. Let G_l be the sum of the BP 2-f.i.s, θ_{ij} , in group l . Consider any contrast of these G_l , written as $\sum_{l=1}^m d_l G_l$. In this contrast, the coefficient for the 2-f.i. θ_{ij} is $d_i + d_j$, since it is in G_i and G_j . Hence the

corresponding $\binom{m}{2} \times 1$ contrast vector for the 2-f.i.'s is

$$\mathbf{d} = \begin{pmatrix} d_1 + d_2 \\ d_1 + d_3 \\ \vdots \\ d_{m-1} + d_m \end{pmatrix}. \quad (5.12)$$

It follows that each element of $\mathbf{A}_{22}\mathbf{d}$, $\mathbf{a}_{ij}^T\mathbf{d}$, may be written as

$$\begin{aligned} \mathbf{a}_{ij}^T\mathbf{d} &= \sum_{l \neq i,j} (d_i + d_j) + \sum_{l \neq i,j} (d_j + d_l) \\ &= (m-2)d_i + (m-2)d_j + 2 \sum_{l \neq i,j} d_l \\ &= (m-2)d_i + (m-2)d_j - 2(d_i + d_j), \quad \text{since } 0 = \sum_{l=1}^m d_l = \sum_{l \neq i,j} d_l + d_i + d_j \\ &= (m-4)(d_i + d_j). \end{aligned}$$

Hence \mathbf{d} is an eigenvector of \mathbf{A}_{22} with eigenvalue $m-4$ and so \mathbf{d} is an eigenvector of \mathbf{C}_{22} with eigenvalue $n_2 + (m-2)n_3$.

Now consider $\mathbf{C}_{12}\mathbf{d}$. Recall that the i th row of \mathbf{C}_{12} has $m-1$ values of c_2 corresponding to 2-f.i.'s that overlap with main effect i , and $\binom{m-1}{2}$ values of c_3 corresponding to the 2-f.i.'s

that do not overlap with main effect i . It then follows that the i -th row of $\mathbf{C}_{12}\mathbf{d}$, $(\mathbf{C}_{12}\mathbf{d})_i$, is

$$\begin{aligned}
(\mathbf{C}_{12}\mathbf{d})_i &= c_2 \sum_{l \neq i} (d_i + d_l) + c_3 \sum_{\substack{l < l' \\ l, l' \neq i}} (d_l + d_{l'}) \\
&= c_2[(m-1)d_i - d_i] + c_3(m-2) \sum_{l \neq i} d_l \\
&= c_2(m-2)d_i - c_3(m-2)d_i \\
&= (c_2 - c_3)(m-2)d_i .
\end{aligned}$$

This implies that $\mathbf{C}_{12}\mathbf{d} = (c_2 - c_3)(m-2)\mathbf{d}_0$ where $\mathbf{d}_0 = (d_1, d_2, \dots, d_m)^T$ and so $\mathbf{e} = (0|\mathbf{0}_m^T|\mathbf{d}^T)^T$ is *not* an eigenvector of $\mathbf{C}_{d\theta_2}$ for any $m > 2$, unless $c_2 = c_3$, or $n_2 = n_3 = 0$. But n_2 is an eigenvalue of \mathbf{C}_{22} , a symmetric, positive definite matrix, so $n_2 > 0$. Hence if $\mathbf{C}_{d\theta_2}$ is positive definite then $\mathbf{C}_{12}\mathbf{d} \neq \mathbf{0}_m$ and \mathbf{d} cannot be an eigenvector of $\mathbf{C}_{d\theta_2}$.

Recall that \mathbf{d} was generated from a specific $m \times 1$ contrast vector \mathbf{d}_0 . The dimension of the eigenspace for \mathbf{A}_{22} corresponding to eigenvalue $m-4$ is $m-1$. There are $m-1$ orthogonal \mathbf{d}_0 contrast vectors that generate a \mathbf{d} eigenvector of \mathbf{A}_{22} corresponding to eigenvalue $m-4$. It can be shown that $m-1$ \mathbf{d} 's generated from $m-1$ orthogonal \mathbf{d}_0 's are also orthogonal. Hence the eigenspace of \mathbf{A}_{22} corresponding to eigenvalue $m-4$ can be generated from any such $m-1$ orthogonal contrast vectors. Now consider the projection matrix onto this eigenspace, $\mathbf{P}_1 = \sum_{i=1}^{m-1} \mathbf{d}_i \mathbf{d}_i^T$ for some orthonormal \mathbf{d}_i generated from a set of $(m-1)$, $m \times 1$ orthogonal contrast vectors. The following lemma gives the representation for \mathbf{P}_1 .

Lemma 5.2.4. *The orthogonal projection matrix onto the eigenspace of \mathbf{A}_{22} corresponding to eigenvalue $m-4$ is*

$$\mathbf{P}_1 = \frac{2}{m}\mathbf{I} + \frac{1}{m-2}\mathbf{A}_{22} - \frac{4}{m(m-2)}(\mathbf{J} - \mathbf{I}) . \quad (5.13)$$

Proof. The projection matrix onto the eigenspace of \mathbf{A}_{22} for eigenvalue $2m - 4$ is $\mathbf{P}_0 = \binom{m}{2}^{-1} \mathbf{J}$. Let \mathbf{e}_j be an eigenvector for \mathbf{A}_{22} with eigenvalue -2 . Note that \mathbf{e}_j is a contrast vector orthogonal to all \mathbf{d}_i and that there exists $m' = \frac{m(m-3)}{2}$ mutually orthogonal \mathbf{e}_j . Define $\mathbf{P}_2 = \sum_{j=1}^{m'} \mathbf{e}_j \mathbf{e}_j^T$ for any such m' orthonormal \mathbf{e}_j . Let \mathbf{P}_1^* be (5.13) and define $\mathbf{P}_1 = \sum_{i=1}^{m-1} \mathbf{d}_i \mathbf{d}_i^T$, for any $m-1$ orthonormal \mathbf{d}_i . We want to show that $\mathbf{P}_1^* = \mathbf{P}_1$. We know $\mathbf{I} = \mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2$ and pre-multiplying both sides by \mathbf{P}_1^* gives

$$\mathbf{P}_1^* = \binom{m}{2}^{-1} \mathbf{P}_1^* \mathbf{J} + \mathbf{P}_1^* \mathbf{P}_1 + \mathbf{P}_1^* \mathbf{P}_2 .$$

It is straightforward to show that $\mathbf{P}_1^* \mathbf{J} = 0 \mathbf{J} = 0$, $\mathbf{P}_1^* \mathbf{P}_1 = \sum_{i=1}^{m-1} (\mathbf{P}_1^* \mathbf{d}_i) \mathbf{d}_i^T = \sum_{i=1}^{m-1} \mathbf{d}_i \mathbf{d}_i^T = \mathbf{P}_1$, and $\mathbf{P}_1^* \mathbf{P}_2 = \sum_{j=1}^{m'} (\mathbf{P}_1^* \mathbf{e}_j) \mathbf{e}_j^T = 0$. Hence $\mathbf{P}_1^* = \mathbf{P}_1$. \square

This gives us the following corollary

Corollary 5.2.1. *The orthogonal projection matrix onto the eigenspace of \mathbf{A}_{22} for eigenvalue -2 is*

$$\mathbf{P}_2 = \frac{1}{(m-1)(m-2)} [(m-1)(m-4) \mathbf{I} + 2 \mathbf{J} - (m-1) \mathbf{A}_{22}] .$$

We want to show that $\mathbf{C}_{12} \mathbf{P}_2 = 0$. Consider $\mathbf{C}_{12} \mathbf{A}_{22}$, which has m rows and $\binom{m}{2}$ columns. Given a column of \mathbf{A}_{22} , \mathbf{a}_{jk} , and row of \mathbf{C}_{12} , \mathbf{c}_i^T , the product $\mathbf{c}_i^T \mathbf{a}_{jk}$ depends on whether the i -th BP main effect and 2-f.i. jk overlap by one factor. The entry in row i and column jk is

$$(\mathbf{C}_{12} \mathbf{A}_{22})_{i,jk} = \begin{cases} (m-2)c_2 + (m-2)c_3 & \text{if } i \text{ and } jk \text{ overlap,} \\ 2c_2 + 2(m-3)c_3 & \text{otherwise.} \end{cases} \quad (5.14)$$

$$(5.15)$$

This matrix has the same form as \mathbf{C}_{12} but different entries and so there are two possible

entries to check in $\mathbf{C}_{12}\mathbf{P}_2$,

$$\mathbf{C}_{12}\mathbf{P}_2 = \frac{1}{(m-1)(m-2)} \left[(m-1)(m-4)\mathbf{C}_{12} + 2 \left((m-1)c_2 + \binom{m-1}{2}c_3 \right) \mathbf{J} - (m-1)\mathbf{C}_{12}\mathbf{A}_{22} \right] .$$

Ignoring the scale factor $[(m-1)(m-2)]^{-1}$, if the entry in \mathbf{C}_{12} is c_2 then the corresponding element in $\mathbf{C}_{12}\mathbf{P}_2$ is

$$(m-1)(m-4)c_2 + 2 \left((m-1)c_2 + \binom{m-1}{2}c_3 \right) - (m-1)(m-2)(c_2 + c_3) = 0 .$$

If the entry in \mathbf{C}_{12} is c_3 then its corresponding element in $\mathbf{C}_{12}\mathbf{P}_2$ is

$$(m-1)(m-4)c_3 + 2 \left((m-1)c_2 + \binom{m-1}{2}c_3 \right) - (m-1)(2c_2 + 2(m-3)c_3) = 0 .$$

Hence, if $\mathbf{e}_{\binom{m}{2}}$ is in the eigenspace of \mathbf{A}_{22} with projection matrix \mathbf{P}_2 , then the vectors $\mathbf{e} = (0|\mathbf{0}_m^T|\mathbf{e}_{\binom{m}{2}}^T)^T$ are eigenvectors of $\mathbf{C}_{d\theta_2}$ with eigenvalue n_2 with multiplicity $\frac{m(m-3)}{2}$, thus proving the following lemma.

Lemma 5.2.5. *A nonsingular $\mathbf{C}_{d\theta_2}$ where \mathbf{V}_d is a BA of strength 4 has eigenvalue n_2 with multiplicity $\frac{m(m-3)}{2}$.*

Now consider vectors of the form $\mathbf{e} = (0|q\mathbf{d}_0^T|\mathbf{d}^T)^T$ where q is some constant to be determined, and \mathbf{d} is the vector created from \mathbf{d}_0 as in (5.12). Note that for a given q , we can generate $m-1$ of these eigenvectors which are orthogonal by choosing $m-1$ orthogonal \mathbf{d}_0 's. This vector is also clearly orthogonal to eigenvectors corresponding to eigenvalue n_2 . If \mathbf{e} is an eigenvector, then $\mathbf{C}_{d\theta_2}\mathbf{e} = \lambda\mathbf{e}$ for some λ , or

$$q\mathbf{C}_{11}\mathbf{d}_0 + \mathbf{C}_{12}\mathbf{d} = \lambda q\mathbf{d}_0 \tag{5.16}$$

$$q\mathbf{C}_{12}^T\mathbf{d}_0 + \mathbf{C}_{22}\mathbf{d} = \lambda\mathbf{d} . \tag{5.17}$$

We have already shown that $\mathbf{C}_{11} = c_1\mathbf{I} + c_2(\mathbf{J} - \mathbf{I})$ and $\mathbf{C}_{12}\mathbf{d} = (c_2 - c_3)(m - 2)\mathbf{d}_0$. Thus (5.16) is

$$\left[(c_1 - c_2) + \frac{(c_2 - c_3)(m - 2)}{q} \right] q\mathbf{d}_0 = \lambda q\mathbf{d}_0 ,$$

implying

$$\lambda = c_1 - c_2 + \frac{(c_2 - c_3)(m - 2)}{q} . \quad (5.18)$$

Next consider the $\binom{m}{2} \times m$ matrix \mathbf{C}_{12}^T , with rows and columns corresponding to 2-f.i.'s and main effects, respectively. Hence row ij will have c_2 in columns i and j and c_3 in the other columns. Thus a given entry in $\mathbf{C}_{12}^T\mathbf{d}_0$ is equal to

$$\begin{aligned} c_2(d_i + d_j) + c_3 \sum_{l \neq i, j} d_l &= c_2(d_i + d_j) - c_3(d_i + d_j) \\ &= (c_2 - c_3)(d_i + d_j) . \end{aligned}$$

Clearly $c_2 - c_3 > 0$, otherwise $n_2 = 0$. Thus (5.17) is

$$q(c_2 - c_3)\mathbf{d}_0 + (n_2 + n_3(m - 2))\mathbf{d} = \lambda\mathbf{d} ,$$

implying the corresponding eigenvalue is

$$\lambda = q(c_2 - c_3) + (n_2 + (m - 2)n_3) . \quad (5.19)$$

Setting the RHS of (5.18) and (5.19) equal, we solve for q :

$$\begin{aligned}
c_1 - c_2 + \frac{(c_2 - c_3)(m - 2)}{q} &= q(c_2 - c_3) + (n_2 + (m - 2)n_3) \\
\Leftrightarrow \quad q^2(c_2 - c_3) + q(-c_1 + c_2 + n_2 + (m - 2)n_3) - (m - 2)(c_2 - c_3) &= 0 \\
\Leftrightarrow \quad q^2(n_2 + n_3) + q(-n_1 - 2n_2 + n_2 + (m - 3)n_3) - (m - 2)(n_2 + n_3) &= 0 ,
\end{aligned}$$

implying there are two solutions for q ,

$$q = \frac{-(-n_1 - n_2 + (m - 3)n_3) \pm \sqrt{(-n_1 - n_2 + (m - 3)n_3)^2 + 4(m - 2)(n_2 + n_3)^2}}{2(n_2 + n_3)} .$$

Plugging the two values of q into (5.19) we get two eigenvalues

$$\lambda = \frac{1}{2} \left[n_1 + 3n_2 + (m - 1)n_3 \pm \sqrt{(n_1 + n_2 - (m - 3)n_3)^2 + 4(m - 2)(n_2 + n_3)^2} \right] .$$

Hence for a given \mathbf{d}_0 , we can generate two eigenvectors corresponding to the two solutions. There are $m - 1$ orthogonal \mathbf{d}_0 , and since the eigenvalues do not depend on the elements of \mathbf{d}_0 , the two eigenvalues each have multiplicity $m - 1$.

Lemma 5.2.6. *A nonsingular $\mathbf{C}_{d\theta_2}$ where \mathbf{V}_d is a BA of strength 4 has the two eigenvalues*

$$\begin{aligned}
\lambda_{q_1} &= \frac{1}{2} \left[n_1 + 3n_2 + (m - 1)n_3 + \sqrt{(n_1 + n_2 - (m - 3)n_3)^2 + 4(m - 2)(n_2 + n_3)^2} \right] \\
\lambda_{q_2} &= \frac{1}{2} \left[n_1 + 3n_2 + (m - 1)n_3 - \sqrt{(n_1 + n_2 - (m - 3)n_3)^2 + 4(m - 2)(n_2 + n_3)^2} \right] ,
\end{aligned}$$

each with multiplicity $m - 1$.

The three eigenvalues corresponding to the eigenvectors in Lemma 5.2.3 are equivalent to the eigenvalues of the matrix \mathbf{M} defined in the proof of that lemma, but do not have a nice form like in Lemmas 5.2.5 and 5.2.6. It is easier to derive a general expression for the *sum*

of the inverse of the three eigenvalues, so we can have a general form of the A -value for $\mathbf{C}_{d\theta_2}$. Even though this simplifies the problem, the expression is still quite involved. Taking E_2/E_3 and the other $\nu - 3$ eigenvalues in the previous lemmas, we have accounted for all the ν eigenvalues and can derive an expression of the A -value for $\mathbf{C}_{d\theta_2}$ when \mathbf{V}_d is a BA of strength 4.

Theorem 5.2.2. *Assume $\mathbf{C}_{d\theta_2}$ is the reduced information matrix for a BA of strength 4 and is nonsingular. Then the trace of $\mathbf{C}_{d\theta_2}^{-1}$ is*

$$\text{tr}(\mathbf{C}_{d\theta_2}^{-1}) = \frac{E_2}{E_3} + \left(\frac{m(m-3)}{2} \right) \frac{1}{n_2} + (m-1)(\lambda_{q_1}^{-1} + \lambda_{q_2}^{-1}) \quad (5.20)$$

where E_2 and E_3 are specified in (5.10) and (5.11) and the values λ_{q_1} and λ_{q_2} are given in Lemma 5.2.6.

5.2.2 A -optimal approximate BAs for reduced BP 2-f.i. model

The goal of this section is to identify the A -optimal BA index set (n_0, \dots, n_4) that minimizes (5.20). Here we do not restrict the elements of the index set to integer values, which they clearly must be in order to correspond to a design. Much like what was done previously with approximate CRDs, design optimization in this way can often help us identify potential structures that may lead to efficient and/or optimal designs.

Optimizing (5.20) is not a straightforward problem. It is complicated both by its nonlinearity, especially in E_2/E_3 , as well as the constraints on the elements of the index set. For example, the following constraints will need to be met for the optimal solution:

- $n_0, n_1, \dots, n_4 \geq 0$
- $n_0 + 4n_1 + 6n_2 + 4n_3 + n_4 = N$

- $\lambda_{q_1}, \lambda_{q_2} > 0$
- The eigenvalues of \mathbf{M} are all positive .

Srivastava and Chopra (1971b) derived a similar expression to (5.20) for the reduced OP 2-f.i. model and focused much of their attention on additional constraints for the BA index set that achieve positive eigenvalues of the comparable matrix to \mathbf{M} defined in the previous section. They did not attempt to optimize their equivalent expression, but rather used it to quickly calculate A -values for BAs of strength 4, finding optimal designs within this subclass. We could try to optimize with respect to c_1, \dots, c_4 ($c_0 = N$ which we assume is fixed) but the constraints are further complicated by the fact that $c_1 \geq c_2 \geq c_3 \geq c_4$.

We now show that optimizing (5.20) with respect to certain ratios of c_0, \dots, c_4 simplifies the constraints and leads to quick optimization results in **Mathematica**. Define the following four ratios r_0, r_1, r_2 , and r_3 as

$$r_i = \frac{c_i}{c_{i+1}}$$

so that we have the following equivalent representations for c_1, \dots, c_4

$$c_i = c_0 \prod_{j=0}^{i-1} \frac{1}{r_j} .$$

These ratios will tell us the optimal relative sizes of c_1, \dots, c_4 to the sample size, c_0 , and their linear relationship with n_0, \dots, n_4 may be used to find the A -optimal approximate BA index set.

Mathematica was used to represent (5.20) in terms of these ratios and this expression was found to be proportional to the sample size, c_0 . Hence minimization of this expression with respect to the ratios does not depend on the sample size, much like how finding optimal designs with respect to replication weight for approximate CRDs does not depend on N . This

means that we may either remove the sample size from the expression or we can minimize it for a specifically chosen sample size (which must be greater than or equal to $\nu = 1 + m + \binom{m}{2}$) and those optimal values will hold for any other sample size. We elect to do the latter, otherwise the expression becomes close to zero and our experience shows that optimization programs may have difficulty with convergence.

We utilized the `Minimize` function in `Mathematica` to minimize the expression for a given m and set $N = \nu$, without restricting ourselves to integer values. Additional constraints were placed to guarantee the eigenvalues of $\mathbf{C}_{d\theta_2}$ were positive. We modified the constraints $1/n_2 > 0$, $\lambda_{q_1}^{-1} > 0$, and $\lambda_{q_2}^{-1} > 0$ by multiplying the left-hand side of each by E_2 . This was done to ensure these the optimization method was sensitive to the constraints, since E_2 grows larger than these other constraints with increasing m . Each ratio was also constrained to be greater than 1, that is, $r_i > 1$, based on the following relationships, which are easy to show

$$\begin{aligned} r_0 &= 1 + \frac{n_0 + 3n_1 + 3n_2 + n_3}{c_1} \\ r_1 &= 1 + \frac{n_1 + 2n_2 + n_3}{c_2} \\ r_2 &= 1 + \frac{n_2 + n_3}{c_3} \\ r_3 &= 1 + \frac{n_3}{c_4} . \end{aligned}$$

Sometimes this set of constraints was slightly modified to $r_i > 1.01$ to achieve convergence. The resulting optimal ratios were then transformed back to the optimal BA index set, giving an “approximate” BA of strength 4 and a lower bound for the A -criterion for any design, not just BAs. This follows because this class of approximate BAs includes the matrix-averaged versions of all designs with respect to $\mathbf{\Pi}_W$.

Table 5.1 has the optimal ratios for $m = 4, 5, \dots, 10$, which hold for any given $N \geq \nu$, and

the approximate optimal A -values and BA index set for saturated designs, i.e. $N = \nu$. For all the m shown, every ratio is close to 2 and consistently decrease as m increases. The ratio is largest for r_0 and smallest for r_3 , which is not surprising because $c_0 \geq c_1 \geq \dots \geq c_4$. There appears to be some hierarchical structure of the BA indices in Table 5.1, with n_0 greater than all other indices, followed by n_1 . For $m < 5$, $n_2 > n_4$, but thereafter n_4 becomes the third largest element in the BA index set, so the structure depends on the number of factors.

Table 5.1: Optimal ratios for $m = 4, \dots, 10$ for saturated BP 2-f.i. model and optimal A -values and BA index sets for $N = \nu$.

m	r_0	r_1	r_2	r_3	A -value	n_0	n_1	n_2	n_3	n_4
4	2.3819	2.2185	2.0456	1.8973	1.3066	1.4831	0.8897	0.5828	0.4813	0.5364
5	2.3630	2.1888	2.0163	1.8739	0.9875	2.1589	1.2745	0.8438	0.7155	0.8188
6	2.3459	2.1679	1.9983	1.8609	0.7617	2.9475	1.7313	1.1597	1.0014	1.1633
7	2.3306	2.1519	1.9860	1.8529	0.6009	3.8462	2.2594	1.5307	1.3402	1.5713
8	2.3168	2.1392	1.9770	1.8477	0.4840	4.8523	2.8580	1.9570	1.7324	2.0438
9	2.3045	2.1286	1.9701	1.8441	0.3970	5.9631	3.5267	2.4390	2.1788	2.5812
10	2.2933	2.1196	1.9647	1.8417	0.3309	7.1763	4.2649	2.9768	2.6798	3.1839

An orthogonal array of strength 4, which is a special case of a strength 4 BA, will have index set $(N/16, \dots, N/16)$ and is easily shown to have $r_i = 2$ for $i = 0, \dots, 3$. This implies that strength 4 orthogonal arrays may be A -efficient for the reduced BP 2-f.i. models, but may not be A -optimal. BAs that are “close” to orthogonal arrays, but with larger n_0 and smaller n_4 , are more likely to be A -optimal, as indicated in the approximate designs in Table 5.1.

5.2.3 Saturated A -optimal exact BAs for reduced BP 2-f.i. model

We first attempt to find A -optimal exact, saturated ($N = \nu$) BAs by searching for integer-valued indices of possible BAs that are “close” to the decimal values in Table 5.1 and so are

likely A -efficient. We call these possible BAs because establishing existence and construction of BAs from a given index can be a difficult task. A program was written in R that considers all integer-valued indices within some range of the decimal values found in Table 5.1 and then calculates the A -value for each one, assuming that design exists. We calculated the A -efficiency of each of these designs by dividing the optimal A -value found in Table 5.1 by the A -value of a BA with given indices. As the optimal A -value should be smaller than the A -value for any BA, these efficiencies will be less than 1; the closer the efficiency is to 1, the better the BA is in terms of the A -criterion, provided it exists.

The top 3 BA index sets for saturated designs with $m = 4, \dots, 10$ factors are given in Table 5.2. We also include the relative efficiencies with respect to BP effects of different orders, which were calculated by dividing the estimator variance under the optimal, approximate BA by the variance for the possible BA. All of these designs estimate all BP effects of the same order with the same variance. The top designs are typically around 90% efficient (except for $m = 4$ and 7). For large m the efficiency quickly approaches 1. In some cases the designs estimate the BP main effects better than the approximate optimal design (e.g. for $m = 7$ the BP main effect efficiency is 1.182). Efficiency tends to be much better for estimation of the BP 2-f.i.'s as m increases. This focus of estimation may be attributed to the fact that the number of 2-f.i.'s in the model grows rapidly as we increase m , hence designs that minimize the variance of estimates for these effects tend to minimize the A -criterion. An additional weighting approach is shown in the next section that allows us to place greater emphasis on the BP main effects.

Establishing existence of the possible BAs in Table 5.2 is a difficult task. Much of the theory on the existence and construction of BAs focuses on strength 2. For $m = 4$ there is no issue, since the indices describe the entire design (there is only one $N \times 4$ subarray, the array itself). The following lemma establishes that a strength t BA is also a strength $t - 1$ BA, a

Table 5.2: A -efficient strength 4 BA index sets for the saturated, reduced BP 2-f.i. model for $m = 4, \dots, 10$. Efficiencies are with respect to the approximate designs in Table 5.1.

m	A -eff	θ_0 -eff	θ_i -eff	θ_{ij} -eff	n_0	n_1	n_2	n_3	n_4
4	0.580	0.215	0.511	0.696	0	1	1	0	1
4	0.555	0.550	0.393	0.696	1	0	1	1	0
4	0.436	0.550	0.568	0.387	1	1	1	0	0
5	0.916	0.497	0.741	1.067	1	1	1	1	1
5	0.595	0.497	0.463	0.686	2	1	1	0	0
5	0.501	0.319	0.596	0.480	1	2	1	0	1
6	0.948	0.765	1.027	0.929	3	2	1	1	1
6	0.898	0.454	0.904	0.932	2	2	1	1	2
6	0.842	0.695	0.689	0.932	2	1	1	2	2
7	0.821	0.729	1.182	0.741	5	3	1	1	2
7	0.816	2.771	1.133	0.725	8	2	1	1	3
7	0.816	3.132	1.132	0.723	9	2	1	1	2
8	0.882	1.971	0.693	0.956	6	2	2	2	3
8	0.879	1.608	0.690	0.957	5	2	2	2	4
8	0.879	2.292	0.689	0.952	7	2	2	2	2
9	0.928	0.733	1.179	0.873	7	4	2	2	3
9	0.908	0.529	1.108	0.873	6	4	2	2	4
9	0.903	0.767	0.984	0.883	6	3	2	3	4
10	0.993	0.782	0.949	1.014	6	4	3	3	4
10	0.991	0.905	0.939	1.010	7	4	3	3	3
10	0.973	0.892	0.999	0.968	8	5	3	2	2

well-known known result which may be found in Rafter and Seiden (1974).

Lemma 5.2.7. *Let \mathbf{B} be an $N \times m$ BA with two symbols (say 0 and 1) of strength $t \leq m$ with index set (n_0, n_1, \dots, n_t) . Then \mathbf{B} is also a BA of strength $t - 1$ with elements of the index set*

$$n_{t-g}^* = n_{t-g+1} + n_{t-g} ,$$

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

One way to search for the optimal BAs of strength 4 would enumerate all $N \times m$ BAs of strength m , which are straightforward to construct, and then recursively use Lemma 5.2.7 to find its strength 4 index set. We demonstrate how this may be done for $m = 7$ in Example ??.

Example 5.2.1. Take the 29×7 BA of strength 7 with index set $(1, 0, 1, 0, 0, 0, 1, 0)$. This array is also of strength 4 with index set $(4, 3, 1, 1, 3)$. By recursively applying Lemma 5.2.7 we can produce the following table of decreasing strength: This demonstrates how the

Table 5.3: Index sets for a 29×7 BA with strength 7 for decreasing strength.

Strength	Index Set
7	$(1, 0, 1, 0, 0, 0, 1, 0)$
6	$(1, 1, 1, 0, 0, 1, 1)$
5	$(2, 2, 1, 0, 1, 2)$
4	$(4, 3, 1, 1, 3)$

higher indices “trickle down” to the lower ones when we consider lower strengths.

The strength 4 index set for the 29×7 BA in Example 5.2.1 appears “close” to the optimal index set found in Table 5.2. The A -value for this design is 0.7821, which is 77% efficient relative to the “optimal” approximate design found in Table 5.1 and 93% efficient relative to the A -best design in Table 5.2.

As demonstrated by Example 5.2.1, one way to construct an A -efficient BA of strength 4 for the saturated case is to start with a strength m BA with indices $n_0, n_2, n_{m-1} = 1$ and all

other indices equal to 0. Table 5.4 contains the indices for such designs for $m = 5, \dots, 10$ and their relative A -efficiencies to the approximate A -optimal BA and the A -best design in Table 5.2. We see that the BAs constructed in this way estimate BP main effects with variances similar to the approximate, optimal BA, but estimates for the other effects become less efficient. However, such designs may be appealing if estimation of the BP main effects is preferred. There may be a further weighting of the BP effects for which this design would be closer to optimality. For example, if we employ the techniques of this dissertation to place more weight on the BP main effects, then the design may become more attractive. The theory for additional weighting with the reduced BP model is the topic in Section 5.3. Another property of these BAs is that the strength 4 indices always have n_2 and $n_3 = 1$, while the approximate n_2 and n_3 in Table 5.1 become greater than 1 as m increases. This may explain why the designs drop in efficiency as we increase m .

Table 5.4: Saturated designs constructed from the strength m BA with $n_0 = n_2 = n_{m-1} = 1$ and their A -efficiencies with respect to the A -optimal approximate BA and A -best exact BA index sets. Efficiencies for BP effects are based on the A -optimal approximate BA.

m	n_0	n_1	n_2	n_3	n_4	θ_0 -eff	θ_i -eff	θ_{ij} -eff	Approximate A -eff	Exact BA A -eff
5	1	1	1	1	1	0.497	0.741	1.067	0.9159	1.0000
6	2	2	1	1	2	0.454	0.904	0.932	0.8984	0.9473
7	4	3	1	1	3	0.420	0.960	0.735	0.7683	0.9353
8	7	4	1	1	4	0.391	0.962	0.577	0.6337	0.7187
9	11	5	1	1	5	0.366	0.936	0.459	0.5205	0.5609
10	16	6	1	1	6	0.344	0.896	0.373	0.4304	0.4334

There are many other ways to construct BAs of strength 4 from larger and smaller arrays. Chakravarti (1961) details a construction method based on taking subarrays (both in terms of removing rows and columns) of OAs of some strength that produce BAs of the same strength. It may also be worthwhile to investigate necessary constraints on a BA index set to determine whether such a BA could exist. This may allow us to pare down the list in

Table 5.2 and get a better idea of the efficiency of the designs in Table 5.4.

We must be careful though in assuming a strength 4 BA is the A -optimal design. Certainly these designs are candidates for A -optimal designs, but there may exist designs that are “almost BAs” and that are truly A -optimal. For example, this was seen in Section 3.4 where for some N a strength m BA was not A -optimal, but arrays built up from certain BAs were optimal until they formed a new BA.

An ad hoc approach to finding optimal designs without restricting attention to only BAs is to use the design search capabilities of SAS PROC OPTEX. A modification of the model and candidate points with the procedure described in Section 4.3 allowed us to search for A -efficient saturated designs under the BP 2-f.i. model for $m = 4, \dots, 10$. We performed 5,000 iterations of the modified Fedorov algorithm for $m = 4$ to 6 and 10,000 iterations were performed for $m = 7$ to 10, due to the larger number of candidate points.

For $m = 4$, the A -best design SAS found was not a BA:

$$\mathbf{V}_d^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Its A -value for estimation of the BP effects is 2.0505 and its relative efficiency compared to the A -optimal approximate BA is 0.6372 which is better than the BAs listed in Table 5.2. We are cautiously confident that this design is A -optimal, as the search was extensive relative to the number of possible designs that could be considered.

For $m = 5$ and 6 the A -best designs were the BAs with index sets found in Table 5.4, which is not surprising given their high efficiencies. For $m > 7$, OPTEX was unable to find designs

with relative efficiencies close to what was seen in Table 5.2, even with a large number of iterations. For these large m , OPTEX has provided some evidence that the designs discussed previously may not exist.

5.2.4 *A*-optimal exact BAs for reduced BP 2-f.i. model with $m = 4, 5$, and 6 and $N > \nu$

We next investigated whether the issue of finding BAs may be due to the restriction $N = \nu$. That is, it may be that with larger N a BA would be easier to construct. For $m = 4, 5$, and 6, the ratios in Table 5.1 were used to calculate the optimal approximate indices for increasing N and we then searched for the corresponding BAs with the R code described in Section 5.2.3 as well as with the SAS PROC OPTEX macro described in the previous section. Constructing the BAs for $m = 4$ is trivial, since the indices will uniquely describe the $N \times 4$ BA we are looking for. Tables 5.5 to 5.7 list the optimal BA indices for $m = 4, 5$, and 6 as found with the R program. Their efficiencies relative to the best design found with the SAS PROC OPTEX macro is also noted in these tables.

The efficiency values for $m = 4$ are very encouraging. The worst *A*-efficiency (with respect to the approximate, optimal bound) was for the saturated case (58%). When comparing the design with the best BA index set found with R with the best design found in SAS, we see that the efficiency is 91%. As soon as we add one more run to the saturated case, the *A*-efficiency increases to 75% and continues to increase overall with N until it hits around 95%. The efficiencies with respect to the SAS-best design are even more encouraging. For $N = 16, 17, 22$, and 23 the same designs were found, and even when they were different the *A*-values were nearly the same. The high efficiencies carry over to $m = 5$ and $m = 6$, but there is still the issue of constructing these BAs. In some cases we may be evaluating

non-existent BAs, including those where the OPTEX A -best designs have efficiency greater than 1. In those cases, either the BA does not exist or OPTEX was unable to find the BA (or a better design).

Based on the results for $m > 6$ in the previous sections, we did not explore this approach for a larger number of factors. Other analytical tools will need to be developed to handle these larger problems and will likely require an exploration of the properties of strength 4 BAs.

Table 5.5: A -efficient strength 4 BAs with $m = 4$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEX.

N	A -eff	θ_0 -eff	θ_i -eff	θ_{ij} -eff	n_0	n_1	n_2	n_3	n_4	OPTEX A -eff
11	0.580	0.215	0.511	0.696	0	1	1	0	1	0.910
12	0.749	0.702	0.854	0.709	1	1	1	0	1	0.995
13	0.741	1.113	0.911	0.667	2	1	1	0	1	0.940
14	0.709	1.466	0.904	0.624	3	1	1	0	1	0.863
15	0.823	0.504	0.694	0.945	1	1	1	1	0	0.945
16	0.925	0.550	0.781	1.063	1	1	1	1	1	1.000
17	0.970	0.874	0.863	1.039	2	1	1	1	1	1.000
18	0.962	1.161	0.879	0.998	3	1	1	1	1	0.997
19	0.937	1.419	0.871	0.954	4	1	1	1	1	0.972
20	0.912	1.361	0.844	0.931	4	1	1	1	2	0.942
21	0.975	0.773	0.974	0.991	2	2	1	1	1	0.997
22	0.990	1.013	1.037	0.967	3	2	1	1	1	1.000
23	0.982	1.232	1.062	0.937	4	2	1	1	1	1.000
24	0.968	1.202	1.046	0.924	4	2	1	1	2	0.991
25	0.951	1.396	1.049	0.893	5	2	1	1	2	0.977
26	0.935	0.891	1.066	0.884	3	3	1	1	1	0.961
27	0.939	1.082	1.119	0.864	4	3	1	1	1	0.970
28	0.958	0.818	0.910	0.995	3	2	2	1	1	0.994
29	0.963	0.999	0.938	0.974	4	2	2	1	1	0.999
30	0.961	0.981	0.931	0.975	4	2	2	1	2	0.997
31	0.954	0.981	0.940	0.951	5	2	2	1	2	0.987
32	0.961	0.756	0.965	0.974	3	3	2	1	1	0.991

Table 5.6: Potential A -efficient strength 4 BAs with $m = 5$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEX.

N	A -eff	θ_0 -eff	θ_i -eff	θ_{ij} -eff	n_0	n_1	n_2	n_3	n_4	OPTEX A -eff
16	0.916	0.497	0.741	1.067	1	1	1	1	1	1.000
17	0.962	0.935	0.827	1.037	2	1	1	1	1	1.000
18	0.945	1.324	0.833	0.990	3	1	1	1	1	0.989
19	0.915	1.286	0.809	0.956	3	1	1	1	2	0.960
20	0.885	1.619	0.792	0.912	4	1	1	1	2	0.926
21	0.929	0.583	0.913	0.963	2	2	1	1	1	0.966
22	0.974	0.918	1.046	0.948	3	2	1	1	1	1.000
23	0.972	1.224	1.096	0.919	4	2	1	1	1	1.000
24	0.960	1.235	1.100	0.901	4	2	1	1	2	0.998
25	0.941	1.503	1.107	0.869	5	2	1	1	2	0.987
26	0.917	1.751	1.099	0.839	6	2	1	1	2	0.969
27	0.892	1.980	1.083	0.809	7	2	1	1	2	0.946
28	0.894	0.851	0.801	0.944	3	2	2	1	1	0.950
29	0.892	1.142	1.198	0.796	5	3	1	1	2	0.948
30	0.882	1.369	1.229	0.774	6	3	1	1	2	0.944
31	0.871	0.449	0.688	1.037	2	2	2	2	1	0.929
32	0.936	0.648	0.909	0.969	3	3	2	1	1	0.991
33	0.957	0.869	0.974	0.955	4	3	2	1	1	1.003
34	0.962	0.935	0.827	1.037	4	2	2	2	2	1.000
35	0.956	1.135	0.834	1.014	5	2	2	2	2	0.991
36	0.947	1.273	1.018	0.908	6	3	2	1	2	0.976
37	0.950	0.561	0.860	1.030	3	3	2	2	2	0.972
38	0.984	0.755	0.933	1.023	4	3	2	2	2	0.999
39	0.995	0.940	0.973	1.008	5	3	2	2	2	1.000
40	0.994	1.115	0.993	0.990	6	3	2	2	2	1.000

Table 5.7: Potential A -efficient strength 4 BAs with $m = 6$ factors for the reduced BP 2-f.i. model. Efficiencies for individual BP effects are based on the A -optimal approximate BAs. The last column is relative to the best design produced by OPTEx.

N	A -eff	θ_0 -eff	θ_i -eff	θ_{ij} -eff	n_0	n_1	n_2	n_3	n_4	OPTEx A -eff
22	0.948	0.765	1.027	0.929	3	2	1	1	1	1.056
23	0.960	1.167	1.122	0.903	4	2	1	1	1	1.005
24	0.951	1.249	1.153	0.882	4	2	1	1	2	1.000
25	0.931	1.599	1.164	0.850	5	2	1	1	2	0.993
26	0.906	1.922	1.155	0.820	6	2	1	1	2	0.977
27	0.880	2.221	1.136	0.791	7	2	1	1	2	0.957
28	0.854	2.180	1.109	0.765	7	2	1	1	3	0.934
29	0.859	0.969	1.222	0.766	5	3	1	1	2	0.945
30	0.854	1.270	1.291	0.745	6	3	1	1	2	0.941
31	0.841	1.552	1.322	0.724	7	3	1	1	2	0.926
32	0.908	0.454	0.707	1.070	2	2	2	2	2	0.992
33	0.949	0.743	0.772	1.057	3	2	2	2	2	1.000
34	0.953	1.016	0.790	1.034	4	2	2	2	2	1.000
35	0.943	1.272	0.792	1.009	5	2	2	2	2	0.987
36	0.929	1.274	0.785	0.991	5	2	2	2	3	0.965

5.3 A -optimal designs for the reduced BP 2-f.i. model with weighted effects

In the previous section, $\mathbf{C}_{d\theta_2}$ was shown to be equivalent to a weighted information matrix of the treatment means with left at most 2 with weight matrix $\mathbf{W} = \mathbf{H}_{11}^T \mathbf{H}_{11}$. Suppose we wish to further weight the BP intercept by w_0 , BP main effects by w_1 , and BP 2-f.i.'s by w_2 (this could be done for a variety of reasons). To achieve this weighting, let $\mathbf{w} = (w_0, w_1 \mathbf{1}_m^T, w_2 \mathbf{1}_{\binom{m}{2}}^T)^T$, $\mathbf{D}_w = \text{Diag}(\mathbf{w})$, $\mathbf{H}_{11}^* = \mathbf{D}_w^{1/2} \mathbf{H}_{11}$, and $\boldsymbol{\theta}_2^* = \mathbf{H}_{11}^* \boldsymbol{\tau}_1$. A natural selection of these weights is $\sqrt{w_0} = 1$, $\sqrt{w_1} = \frac{1}{\sqrt{2}}$, and $\sqrt{w_2} = \frac{1}{2}$, the normalization constants for the BP effects. We could also potentially use this technique to put more emphasis on

estimating main effects rather than 2-f.i.'s.

For some such \mathbf{D}_w , let $\mathbf{W}_* = \mathbf{H}_{11}^{*T} \mathbf{H}_{11}^*$, which has the form

$$\begin{aligned} \mathbf{W}_* &= \mathbf{H}_{11}^T \mathbf{D}_w \mathbf{H}_{11} \\ &= \left(\begin{array}{c|c|c} w_0 + w_1 m + w_2 \binom{m}{2} & -(w_1 + w_2(m-1)) \mathbf{1}_m^T & w_2 \mathbf{1}_{\binom{m}{2}}^T \\ \hline & (w_1 + w_2(m-2)) \mathbf{I}_m + w_2 \mathbf{J}_m & -w_2 \mathbf{V}_{m;2}^T \\ \hline & & w_2 \mathbf{I}_{\binom{m}{2}} \end{array} \right). \end{aligned} \quad (5.21)$$

Clearly w_0 , w_1 , and $w_2 > 0$, so the diagonal elements of the diagonal submatrices in the partition of (5.21) are the same within each submatrix but different between them. Using this fact and previous arguments used to find Π_W , where $\mathbf{W} = \mathbf{H}_{11}^T \mathbf{H}_{11}$, it is easily shown that $\Pi_{W_*} = \Pi_W$ since the respective matrices that make up the partitions for \mathbf{W}_* shown in (5.21) have a similar structure to the corresponding partitions in $\mathbf{W} = \mathbf{H}_{11}^T \mathbf{H}_{11}$. That is, Π_{W_*} contains all and only the $m!$ factor permutations.

The weighted information matrix for a design d corresponding to \mathbf{W}_* is defined to be $\mathbf{C}_{dW_*\tau_1} = \mathbf{W}_*^{-1/2} \mathbf{Q}^T \mathbf{X}_d^T \mathbf{X}_d \mathbf{Q} \mathbf{W}_*^{-1/2}$ and we want to find forms of $\mathbf{C}_{dW_*\tau_1}$ that are invariant to permutations in Π_W . Rather than working with this weighted information matrix, we instead focus on the following information matrix:

$$\begin{aligned} \mathbf{C}_{d\theta_2^*} &= \mathbf{D}_w^{-1/2} \mathbf{K} \mathbf{H}_{BP}^{-1T} \mathbf{X}_d^T \mathbf{X}_d \mathbf{H}_{BP}^{-1} \mathbf{K}^T \mathbf{D}_w^{-1/2} \\ &= \mathbf{D}_w^{-1/2} \mathbf{C}_{d\theta_2} \mathbf{D}_w^{-1/2}, \end{aligned}$$

which has the same eigenvalues as $\mathbf{C}_{d\mathbf{W}_*\tau_1}$. It is easily shown from (5.7) that

$$\mathbf{C}_{d\theta_2^*} = \left(\begin{array}{c|cc|cc} \frac{1}{w_0}N & \frac{1}{\sqrt{w_0w_1}} & \mathbf{1}_N^T \mathbf{V}_d & \frac{1}{\sqrt{w_0w_2}} & \mathbf{1}_N^T \mathbf{V}_{d2} \\ \hline & \frac{1}{w_1} & \mathbf{V}_d^T \mathbf{V}_d & \frac{1}{\sqrt{w_1w_2}} & \mathbf{V}_d^T \mathbf{V}_{d2} \\ \hline & & & \frac{1}{w_2} & \mathbf{V}_{d2}^T \mathbf{V}_{d2} \end{array} \right), \quad (5.22)$$

so the submatrices in the partitions of $\mathbf{C}_{d\theta_2}$ and $\mathbf{C}_{d\theta_2^*}$ found in (5.7) and (5.22), respectively, are proportional. The proof of Theorem 5.2.1 relied only on arguments for the submatrices in (5.7) and the added weights do not change these results, giving us the following corollary.

Corollary 5.3.1. *The information matrix $\mathbf{C}_{d\theta_2^*} = \mathbf{D}_w^{-1/2} \mathbf{C}_{d\theta_2} \mathbf{D}_w^{-1/2}$ is invariant to all $\mathbf{\Pi} \in \Pi_{W_*} = \Pi_W$ if and only if $\mathbf{X}_d \mathbf{V}_m = \mathbf{V}_d$ is a BA of strength 4.*

Even though BAs of strength 4 create optimal forms for both $\mathbf{C}_{d\theta_2}$ and $\mathbf{C}_{d\theta_2^*}$, we expect that for any substantially different values of w_0, w_1 , and w_2 that a BA that is optimal for one is not optimal for the other.

A BA of strength 4 will have $\mathbf{C}_{d\theta_2^*}$ of the following form

$$\begin{aligned} \mathbf{C}_{d\theta_2^*} &= \mathbf{D}_w^{-1/2} \mathbf{C}_{d\theta_2} \mathbf{D}_w^{-1/2} \\ &= \left(\begin{array}{c|cc|cc} \frac{c_0}{w_0} & \frac{c_1}{\sqrt{w_0w_1}} & \mathbf{1}_m^T & \frac{c_2}{\sqrt{w_0w_2}} & \mathbf{1}_{\binom{m}{2}}^T \\ \hline & \frac{1}{w_1} & \mathbf{C}_{11} & \frac{1}{\sqrt{w_1w_2}} & \mathbf{C}_{12} \\ \hline & & & \frac{1}{w_2} & \mathbf{C}_{22} \end{array} \right). \end{aligned} \quad (5.23)$$

Suppose we want to find an A -optimal design for $\mathbf{C}_{d\theta_2^*}$. If we knew the diagonal elements of $\mathbf{C}_{\theta_2}^{-1}$ then we could easily derive the trace of $\mathbf{C}_{d\theta_2^*}^{-1}$ given the BA indices using $\text{tr}(\mathbf{C}_{d\theta_2^*}^{-1}) = \text{tr}(\mathbf{D}_w^{1/2} \mathbf{C}_{d\theta_2}^{-1} \mathbf{D}_w^{1/2}) = \text{tr}(\mathbf{C}_{d\theta_2}^{-1} \mathbf{D}_w)$. We instead use the same eigenvalue approach that was used for $\mathbf{C}_{d\theta_2}$ in Section 5.2.1.

The eigenvectors for $\mathbf{C}_{d\theta_2}$ when we have a BA of strength 4 are clearly eigenvectors for

$\mathbf{D}_w^{-1/2}$ and so $\mathbf{C}_{d\theta_2}$ and $\mathbf{C}_{d\theta_2^*}$ have the same eigenvectors. First note that even though the submatrices in the partitions of $\mathbf{C}_{d\theta_2^*}$ seen in (5.22) are different from those in (5.7), the row sums of each partition are still constant. Hence we consider vectors of the form $\mathbf{a} = (a_1|a_2\mathbf{1}_m^T|a_3\mathbf{1}_{\binom{m}{2}}^T)^T$. For $\mathbf{C}_{d\theta_2}$, this gave a 3×3 matrix \mathbf{M} . For $\mathbf{C}_{d\theta_2^*}$, we get a similar matrix, \mathbf{M}^* , which is easily shown to be

$$\mathbf{M}^* = \begin{pmatrix} w_0^{-\frac{1}{2}} & 0 & 0 \\ 0 & w_1^{-\frac{1}{2}} & 0 \\ 0 & 0 & w_2^{-\frac{1}{2}} \end{pmatrix} \mathbf{M} \begin{pmatrix} w_0^{-\frac{1}{2}} & 0 & 0 \\ 0 & w_1^{-\frac{1}{2}} & 0 \\ 0 & 0 & w_2^{-\frac{1}{2}} \end{pmatrix},$$

a diagonal weighting of the original \mathbf{M} . `Mathematica` was used to find E_2^* and E_3^* , the analogs of E_2 and E_3 in (5.10) and (5.11), which are omitted from the dissertation.

The next eigenvectors to consider are those of the form $\mathbf{e} = (0|\mathbf{0}_m^T|\mathbf{e}_{\binom{m}{2}}^T)^T$. Note that any eigenvector of \mathbf{C}_{22} is also an eigenvector for $\frac{1}{w_2}\mathbf{C}_{22}$ with the original eigenvalues of \mathbf{C}_{22} scaled by $\frac{1}{w_2}$. The eigenvectors of $\frac{1}{w_2}\mathbf{C}_{22}$ with eigenvalue $\frac{n_2}{w_2}$ will clearly satisfy $\frac{1}{\sqrt{w_1w_2}}\mathbf{C}_{12}\mathbf{e}_{\binom{m}{2}} = 0$, which implies that $\frac{n_2}{w_2}$ is an eigenvalue of $\mathbf{C}_{d\theta_2^*}$ with multiplicity $\frac{m(m-3)}{2}$.

Now consider eigenvectors of the form $\mathbf{e} = (0|q\mathbf{d}_0^T|\mathbf{d}^T)^T$ defined previously for $\mathbf{C}_{d\theta_2}$. Then $\mathbf{C}_{d\theta_2^*}\mathbf{e}$ yields expressions analogous to (5.16) and (5.17)

$$\frac{q}{w_1}\mathbf{C}_{11}\mathbf{d}_0 + \frac{1}{\sqrt{w_1w_2}}\mathbf{C}_{12}\mathbf{d} = \lambda q\mathbf{d}_0 \quad (5.24)$$

$$\frac{q}{\sqrt{w_1w_2}}\mathbf{C}_{12}^T\mathbf{d}_0 + \frac{1}{w_2}\mathbf{C}_{22}\mathbf{d} = \lambda\mathbf{d} . \quad (5.25)$$

Solving for q as was done in Section 5.2.1, we get

$$q = \frac{-b \pm \sqrt{b^2 + 4(n_2 + n_3)^2(m-2)}}{2(n_2 + n_3)}$$

$$b = \sqrt{\frac{w_1}{w_2}}(n_2 + (m-2)n_3) - \sqrt{\frac{w_2}{w_1}}(c_1 - c_2) .$$

This form is similar to that for $\mathbf{C}_{d\theta_2}$, except the b term did not have the weights involved.

These q produce two different eigenvalues, each with multiplicity $m-1$

$$\lambda_{q_1}^* = \frac{-b + \sqrt{b^2 + 4(n_2 + n_3)^2(m-2)}}{2\sqrt{w_1 w_2}} + \frac{n_2 + (m-2)n_3}{w_2}$$

$$\lambda_{q_2}^* = \frac{-b - \sqrt{b^2 + 4(n_2 + n_3)^2(m-2)}}{2\sqrt{w_1 w_2}} + \frac{n_2 + (m-2)n_3}{w_2} .$$

Corollary 5.3.2. *Assume $\mathbf{C}_{d\theta_2^*}$ is the reduced information matrix for θ_2^* for a BA of strength 4 and is nonsingular. Then the trace of $\mathbf{C}_{d\theta_2^*}^{-1}$ is*

$$\text{tr}(\mathbf{C}_{d\theta_2^*}^{-1}) = \frac{E_2^*}{E_3^*} + \left(\frac{m(m-3)}{2} \right) \frac{w_2}{n_2} + (m-1)(\lambda_{q_1}^{*-1} + \lambda_{q_2}^{*-1}) . \quad (5.26)$$

Representing (5.26) in terms of the ratios r_0, \dots, r_3 again gives an expression proportional to c_0 . We again minimize this expression with respect to the ratios using the same approach taken to construct Table 5.1, but here for weights w_0, w_1 , and w_2 . We first investigate the A -optimal approximate BAs for the normalized BP effects under the reduced BP 2-f.i. model by setting $w_0 = 1$, $w_1 = \frac{1}{2}$, and $w_2 = \frac{1}{4}$. The results for $m = 4, \dots, 10$ are shown in Table 5.8. Focusing on the normalized BP effects caused r_0 and r_1 to increase over those in Table 5.1 while r_2 and r_3 decreased. As for the saturated A -optimal approximate BA, we see that greater emphasis is placed on n_0 and n_1 than found previously, implying that treatments of smaller heft will likely be replicated more in the exact A -optimal design for estimation of the normalized BP effects. This is consistent with what was found for the full model in

Section 3.4.

Table 5.8: Optimal ratios for the saturated, normalized BP 2-f.i. model, optimal A_W -values, and optimal BA index sets.

m	r_0	r_1	r_2	r_3	A_W -value	n_0	n_1	n_2	n_3	n_4
4	2.6614	2.3694	2.0623	1.8307	0.4193	2.0124	0.9755	0.5147	0.3838	0.4620
5	2.6189	2.3066	2.0118	1.8004	0.3153	2.9190	1.3819	0.7468	0.5853	0.7313
6	2.5812	2.2635	1.9813	1.7833	0.2416	3.9634	1.8626	1.0302	0.8348	1.0657
7	2.5483	2.2316	1.9609	1.7730	0.1893	5.1406	2.4166	1.3651	1.1338	1.4668
8	2.5196	2.2069	1.9464	1.7664	0.1514	6.4456	3.0431	1.7522	1.4833	1.9354
9	2.4944	2.1871	1.9357	1.7622	0.1234	7.8746	3.7417	2.1918	1.8841	2.4719
10	2.4722	2.1708	1.9275	1.7595	0.1022	9.4239	4.5115	2.6843	2.3368	3.0769

Now suppose we are more interested in precisely estimating the non-normalized BP main effects than we are the intercept and 2-f.i.'s. We reflect this relative importance by setting $w_1 = 20$ and $w_0 = w_2 = 1$ and again optimize the A_W expression in **Mathematica**; the results are reported in Table 5.9. As expected, the ratios r_0 and r_1 become even larger compared to the unweighted case but decrease much more dramatically as m increases. This gives a saturated A_W -optimal approximate BA that has large n_0 and n_1 values.

Table 5.9: Optimal ratios for the saturated BP 2-f.i. model with $w_1 = 20$ and $w_0 = w_2 = 1$, optimal A_W -values, and BA index sets.

m	r_0	r_1	r_2	r_3	A_W -value	n_0	n_1	n_2	n_3	n_4
4	3.4698	3.2830	1.9932	1.4402	7.8249	2.5114	1.3903	0.3331	0.1481	0.3364
5	3.4281	2.9003	1.8504	1.4458	5.3702	4.1093	1.8470	0.4714	0.2681	0.6015
6	3.3598	2.6789	1.7747	1.4380	3.8227	5.9223	2.3893	0.6475	0.4195	0.9577
7	3.2909	2.5394	1.7305	1.4317	2.8155	7.9523	3.0168	0.8602	0.6047	1.4006
8	3.2283	2.4449	1.7028	1.4279	2.1363	10.1983	3.7286	1.1099	0.8250	1.9280
9	3.1729	2.3774	1.6848	1.4260	1.6627	12.6574	4.5240	1.3972	1.0814	2.5382
10	3.1238	2.3268	1.6724	1.4255	1.3225	15.3246	5.4019	1.7226	1.3751	3.2318

It remains to be seen whether these weighted criteria lead to different exact designs than those

found in the previous section. We enumerated A -optimal BA indices using the previously mentioned program in **R** and found them to be different from those in Table 5.2. However, in many instances the A_W -best design produced by SAS PROC OPTEX was either the same or nearly identical to the unweighted case. As m or N increases we will likely see these designs diverge.

Chapter 6

Future work

Having developed a new general approach to design evaluation, there are many avenues of future work. Although the class of weight matrices we have focused on has many applications, it would be worthwhile to investigate other forms of weight matrices that give a desirable weighting of estimable functions. We have spent much time exploring the A_W -criterion and should develop properties and applications of other weighted criteria, including the E_W -criterion. Many of the results in this dissertation are for completely-randomized designs, although the theory generalizes outside this class of models, and we should begin to consider what happens when we introduce blocking.

For factorial experiments, it is sometimes desirable to estimate all OP main effects and 2-f.i.'s and it is known that a Resolution V design is universally optimal. If such a design does not exist, then typically an orthogonal array is chosen that maximizes the number of 2-f.i.'s. Alternatively, design search algorithms like OPTEx are used to find an efficient design that can estimate all effects. But, as mentioned before, the number of 2-f.i.'s can increase quickly and so standard optimization may implicitly focus its attention on estimating these effects well, instead of the main effects. By employing weights, we can search for a design that

not only estimates all the effects, but emphasizes precise estimation of main effects. Blocks could also be introduced, leading to optimal partially confounded factorial designs.

Much of the work in this dissertation has been devoted to developing a general theory for BP, and there is still much to be done. We could extend the baseline parameterization to factors with more than two levels by modifying the Kronecker construction in Lemma 1.2.2, following the approach in Hedayat et al. (1999). One specific application of BP is with cDNA microarray experiments, which often employ one or two blocking factors and small block size. Our modification of SAS PROC OPTEX to search for A -optimal incomplete block designs could be used to quickly search for good designs for cDNA microarray experiments.

Finally, general weighted optimality criteria has so far been defined in terms of the information matrix of a linear model. When working with nonlinear models, optimization is typically based on the Fisher information matrix, which itself is based on a linear approximation to the nonlinear model. Weighting this matrix like we have done for linear models will create weighted optimality criteria for estimation of parameters in a nonlinear model.

Bibliography

- Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007), *Optimum Experimental Designs, with SAS*, Oxford: Oxford University Press.
- Bose, R. C. (1963), “Strongly Regular Graphs, Partial Geometries, and Partially Balanced Designs,” *Pacific Journal of Mathematics*, 13, 389–419.
- Bose, R. C. and Shimamoto, T. (1952), “Classification and Analysis of Partially Balanced Incomplete Block Designs with Two Associate Classes,” *Journal of the American Statistical Association*, 47, 151–184.
- Bose, R. C. and Srivastava, J. N. (1964), “Multidimensional Partially Balanced Designs and their Analysis, with Applications to Partially Balanced Factorial Fractions,” *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 145–168.
- Box, G. E. P. and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley.
- Brouwer, A. E. and van Lint, J. H. (1984), “Strongly Regular Graphs and Partial Geometries,” In: D. M. Jackson and S. A. Vanstone *Enumeration and Design: Papers from the Conference on Combinatorics held at the University of Waterloo, Waterloo, Ontario, June 14–July 2 1982*, Academic Press, Toronto, Canada, 85–122.

- Caliński, T. and Kageyama, S. (2000), *Block Designs: A Randomization Approach; Volume I: Analysis*, New York: Springer.
- Chakravarti, I. M. (1956), “Fractional Replication in Asymmetrical Factorial Designs and Partially Balanced Arrays,” *Sankhyā: The Indian Journal of Statistics*, 17, 143–164.
- (1961), “On Some Methods of Construction of Partially Balanced Arrays,” *The Annals of Mathematical Statistics*, 32, 1181–1185.
- Cheng, C. S., Majumdar, D., Stufken, J., and Ture, T. E. (1988), “Optimal Step-Type Designs for Comparing Test Treatments With a Control,” *Journal of the American Statistical Association*, 83, 477–482.
- Chopra, D. V. (1967), “Investigation of the Construction and Existence of Balanced Fractional Factorial Designs of 2^m Series,” *Unpublished Ph.D. dissertation, University of Nebraska*.
- Cook, R. D. and Nachtsheim, C. J. (1980), “A Comparison of Algorithms for Constructing Exact D -Optimal Designs,” *Technometrics*, 22, 315–324.
- (1989), “Computer-Aided Blocking of Factorial and Response-Surface Designs,” *Technometrics*, 31, 339–346.
- Deng, L. Y. and Tang, B. (1999), “Generalized Resolution and Minimum Aberration Criteria for Plackett-Burman and Other Nonregular Factorial Designs,” *Statistica Sinica*, 9, 1071–1082.
- DuMouchel, W. and Jones, B. (1994), “A Simple Bayesian Modification of D -optimal Designs to Reduce Dependence on an Assumed Model,” *Technometrics*, 36, 37–47.
- Fedorov, V. V., Studden, W. J., and Klimko, E. M. (1972), *Theory of Optimal Experiments*, New York and London: Academic Press.

- Glonek, G. F. V. and Solomon, P. J. (2004), “Factorial and Time Course Designs of cDNA Microarray Experiments,” *Biostatistics*, 5, 89–111.
- Gupta, S. (2006), “Balanced Factorial Designs for cDNA Microarray Experiments,” *Communications in Statistics - Theory and Methods*, 35, 1469–1476.
- Gupta, V. K., Ramana, D. V. V., and Parsad, R. (1999), “Weighted A -efficiency of Block Designs for Making Treatment-Control and Treatment-Treatment Comparisons,” *Journal of Statistical Planning and Inference*, 77, 301–319.
- Hedayat, A. and Yang, M. (2005), “Optimal and Efficient Crossover Designs for Comparing Test Treatments with a Control Treatment,” *The Annals of Statistics*, 33, 915–943.
- Hedayat, A. S., Jacroux, M., and Majumdar, D. (1988), “Optimal Designs for Comparing Test Treatments with Controls,” *Statistical Science*, 3, 462–476.
- Hedayat, A. S., Sloane, N. J. A., and Stufken, J. (1999), *Orthogonal Arrays: Theory and Applications*, New York: Springer-Verlag.
- Hinkelmann, K. and Kempthorne, O. (2005), *Design and Analysis of Experiments, Vol. 2: Advanced Experimental Design*, Hoboken, NJ: John Wiley & Sons, Inc, 1st ed.
- (2008), *Design and Analysis of Experiments, Vol. 1: Introduction to Experimental Design*, Hoboken, NJ: John Wiley & Sons, Inc, 2nd ed.
- Hocking, R. R. (1985), *The Analysis of Linear Models*, Monterey, CA: Brooks/Cole.
- (2003), *Methods and Applications of Linear Models*, Hoboken, NJ: Wiley, 2nd ed.
- Jacroux, M. (1993), “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.

- Kao, L. J., Yates, P. K., Lewis, S. M., and Dean, A. M. (1995), “Approximate Information Matrices for Estimating a Given Set of Contrasts,” *Statist. Sinica*, 5, 593–598.
- Kiefer, J. (1975), “Construction and Optimality of Generalized Youden Designs,” In: Srivastava, J. N. *A Survey of Statistical Design and Linear Models*. North-Holland, Amsterdam, 333–353.
- Majumdar, D. (1996), “Optimal and Efficient Treatment-Control Designs,” In: S. Ghosh and C. R. Rao *Design and Analysis of Experiments. Handbook of Statistics*, North-Holland, Amsterdam, 1007–1053.
- Majumdar, D. and Notz, W. I. (1983), “Optimal Incomplete Block Designs for Comparing Treatments with a Control,” *The Annals of Statistics*, 11, 258–266.
- Mitchell, T. J. (1974), “An Algorithm for the Construction of D -Optimal Experimental Designs,” *Technometrics*, 16, 203–210.
- Morgan, J. P. and Stallings, J. W. (2013), “On the A -Criterion of Experimental Design,” *Journal of Statistical Theory and Practice (Forthcoming)*.
- Morgan, J. P. and Wang, X. (2010), “Weighted Optimality in Designed Experimentation,” *Journal of the American Statistical Association*, 105, 1566–1580.
- Mukerjee, R. and Tang, B. (2012), “Optimal Fractions of Two-Level Factorials under a Baseline Parameterization,” *Biometrika*, 99, 71–84.
- Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments (Wiley Series in Probability and Statistics)*, Wiley, 3rd ed.
- Pukelsheim, F. (2006), *Optimal Design of Experiments*, Society for Industrial and Applied Mathematics.

- Rafter, J. A. and Seiden, E. (1974), “Contributions to the Theory and Construction of Balanced Arrays,” *The Annals of Statistics*, 2, 1256–1273.
- Shirakura, T. and Tong, W.-P. (1996), “Weighted A -optimality for Fractional 2^m Factorial Designs of Resolution V,” *Journal of Statistical Planning and Inference*, 56, 243–256.
- Srivastava, J. N. (1961), *Contributions to the Theory of Construction and Analysis of Designs*, University of North Carolina at Chapel Hill.
- (1965), “Optimal Balanced 2^m Fractional Factorial Designs,” In: *S. N. Roy Memorial Volume*, Chapel Hill, NC, University of North Carolina and Calcutta: Indian Statistical Institute.
- (1972), “Some General Existence Conditions for Balanced Arrays of Strength t and 2 Symbols,” *Journal of Combinatorial Theory, Series A*, 13, 198–206.
- Srivastava, J. N. and Chopra, D. V. (1971a), “Balanced Optimal 2^m Fractional Factorial Designs of Resolution V, $m \leq 6$,” *Technometrics*, 13, 257–269.
- (1971b), “On the Characteristic Roots of the Information Matrix of 2^m Balanced Factorial Designs of Resolution V, with Applications,” *The Annals of Mathematical Statistics*, 42, 722–734.
- Welch, W. J. (1982), “Branch-and-Bound Search for Experimental Designs Based on D -Optimality and Other Criteria,” *Technometrics*, 24, 41–48.
- Yang, M. and Stufken, J. (2008), “Optimal and Efficient Crossover Designs for Comparing Test Treatments to a Control Treatment under Various Models,” *Journal of Statistical Planning and Inference*, 138, 278–285.