Fortgeschrittene Versuchsplanung

Prof. Dr. Christine Müller
Technische Universität Dortmund

Sommersemester 2019

Christine Müller, *Fortgeschrittene Versuchsplanung*, SS 2019

Contents

Chapter 1

Preface

This manuscript does not only deal with the theory of obtaining good designs but also provides some R functions for this task. The main R packages which are used here are `agricolae` and `conf.design`. The package `agricolae` is meanwhile a very complex package so that for its installation the following packages are needed: `AlgDesign, coda, combinat, deldir, expm, gdata, gmodels, gtools, klaR, LearnBayes, sp, spdep`.

Many futher R packages for designing experiments can be found on the webpage `https://cran.r-project.org/web/views/ExperimentalDesign.html`.

Chapter 2

The general linear model

In the general linear model, it is assumed that the data $y_1, \ldots, y_N$ are realizations of stochastically independent random variables $Y_1, \ldots, Y_N$ which satisfy

$$\mathrm{E}(Y_n) = x(t_n)^\top \theta \tag{2.1}$$

or, respectively,

$$Y_n = x(t_n)^\top \theta + Z_n \quad \text{with} \quad \mathrm{E}(Z_n) = 0 \tag{2.2}$$

for $n = 1, \ldots, N$. Thereby $\theta \in \mathbb{R}^R$ is an unknown parameter vector, $t_1, \ldots, t_N \in \mathcal{T}$ are known **experimental conditions**, also called **design points**, in the **design region** $\mathcal{T}$ and $x : \mathcal{T} \longrightarrow \mathbb{R}^R$ a known **regression function**. $Z_1, \ldots, Z_N$ are error variables which usually satisfy $\mathrm{var}(Z_n) = \sigma^2$ for all $n = 1, \ldots, N$.

Setting $Y = (Y_1, \ldots, Y_N)^\top$, $Z = (Z_1, \ldots, Z_N)^\top$, $X_d = (x(t_1), \ldots, x(t_N))^\top$, the model (2.1) or (2.2), respectively, can be written as

$$Y = X_d \theta + Z \quad \text{with} \quad \mathrm{E}(Z) = 0_N \quad \text{and} \quad \mathrm{Cov}(Z) = \sigma^2 \, I_{N\times N},$$

where $0_N \in \mathbb{R}^N$ is the N dimensional vector consisting only of zeros and $I_{N\times N}$ is the $N \times N$ identity matrix. $X_d = (x(t_1), \ldots, x(t_N))^\top$ is called **design matrix** and $d = (t_1, \ldots, t_N)$ the **design**.

2.1 Identifiability

In many examples, it happens that the design matrix $X_d = (x(t_1),\ldots,x(t_N))^\top \in \mathbb{R}^{N\times R}$ is of full rank R. However, there are also many examples where the rank of X_d is less than R, i.e. $\mathrm{rk}(X_d) < R$.

2.1.1 Example (One-way layout)
In the one-way layout, we assume that a qualitative factor A can attain A levels. These A levels provides A samples so that the two sample problem is a special case of the one-way layout with $A = 2$.

For the one-way layout several parameterizations are possible.

Non-singular parameterization:

$$Y_n = \mu_a + Z_n, \quad \text{if } t_n = a, \quad \text{for } n = 1,\ldots,N,$$

i.e.

$$\begin{aligned} x(t_n) &= (1\!\!1_{\{1\}}(t_n), 1\!\!1_{\{2\}}(t_n), \ldots, 1\!\!1_{\{A\}}(t_n))^\top \in \mathbb{R}^A, \\ \theta &= (\mu_1,\mu_2,\ldots,\mu_A)^\top \in \mathbb{R}^A. \end{aligned}$$

As soon as each level is observed at least once, then $X_d = (x(t_1),\ldots,x(t_N))^\top \in \mathbb{R}^{N\times A}$ is of full rank A. Sorting the observations/measurements with respect to the levels and assuming that each level is observed M times (i.e. we have balance design), then the design matrix can be written with the Kronecker product as

$$X_d = I_{A\times A} \otimes 1_M = \begin{pmatrix} 1_M & 0_M & 0_M & \ldots & 0_M & 0_M \\ 0_M & 1_M & 0_M & \ldots & 0_M & 0_M \\ 0_M & 0_M & 1_M & \ldots & 0_M & 0_M \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0_M & 0_M & 0_M & \ldots & 1_M & 0_M \\ 0_M & 0_M & 0_M & \ldots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times A}.$$

Control parameterization:
Assume without loss of generality that the first level is the control level, for example a placebo in clinical studies or the standard crop in agricultural studies. Then we can set

$$\begin{aligned} Y_n &= \mu + Z_n \quad \text{if } t_n = 1, \\ Y_n &= \mu + \alpha_a + Z_n \quad \text{if } t_n = a, \quad \text{for } a = 2,\ldots,A. \end{aligned}$$

Then we have

$$x(t_n) = (1, 1\!\!1_{\{2\}}(t_n), \ldots, 1\!\!1_{\{A\}}(t_n))^\top \in \mathbb{R}^A,$$
$$\theta = (\mu, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^A.$$

As soon as each level is observed at least once, then $X_d = (x(t_1), \ldots, x(t_N))^\top \in \mathbb{R}^{N\times A}$ is of full rank A. In balanced designs, the design matrix has now the form

$$X_d = \begin{pmatrix} 1_M & 0_{M\times(A-1)} \\ 1_{M(A-1)} & I_{(A-1)\times(A-1)} \otimes 1_M \end{pmatrix}$$
$$= \begin{pmatrix} 1_M & 0_M & 0_M & \ldots & 0_M & 0_M \\ 1_M & 1_M & 0_M & \ldots & 0_M & 0_M \\ 1_M & 0_M & 1_M & \ldots & 0_M & 0_M \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_M & 0_M & 0_M & \ldots & 1_M & 0_M \\ 1_M & 0_M & 0_M & \ldots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times A}.$$

Singular parameterization:
This parameterization is preferred in applications although the corresponding design matrix is not of full rank:

$$Y_n = \mu + \alpha_a + Z_n \quad \text{if } t_n = a, \quad \text{for } a = 1, \ldots, A,$$
$$x(t_n) = (1, 1\!\!1_{\{1\}}(t_n), 1\!\!1_{\{2\}}(t_n), \ldots, 1\!\!1_{\{A\}}(t_n))^\top \in \mathbb{R}^{A+1},$$
$$\theta = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^{A+1}.$$

Here the design matrix $X_d \in \mathbb{R}^{N\times(A+1)}$ is never of full rank since it has $A + 1$ columns. In balanced designs, it has the form:

$$X_d = \left(\begin{array}{cc} 1_{MA} & I_{A\times A} \otimes 1_M \end{array} \right)$$

$$= \begin{pmatrix} 1_M & 1_M & 0_M & 0_M & \dots & 0_M & 0_M \\ 1_M & 0_M & 1_M & 0_M & \dots & 0_M & 0_M \\ 1_M & 0_M & 0_M & 1_M & \dots & 0_M & 0_M \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_M & 0_M & 0_M & 0_M & \dots & 1_M & 0_M \\ 1_M & 0_M & 0_M & 0_M & \dots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times(A+1)}.$$

X_d is not of full rank because there are two many parameters. Namely there are $A+1$ parameters where there are only A different experimental conditions. This means that not all of these $A+1$ parameter can be estimated by data, i.e. they are **not identifiable**. To avoid this problem one can use the side condition that

$$\sum_{a=1}^{A} \alpha_a = 0.$$

However, this requirement is not convenient mathematically. It is more convenient to use no side condition for the parameters. This is possible since the interest lies not in estimating (identifying) all parameters. The interest is here in estimating and testing the difference of level effects, i.e. we want to know whether the A levels provide different effects. This means that we are only interested in specific **aspects** of the unknown parameter vector θ. For example, we may only interested in $\lambda(\theta) = \alpha_1 - \alpha_2$, the difference of the effects of the first and the second level. Statistical methods as estimators and tests should provide for such aspects the same results independently of the parameterization which is used.

2.1.1 Definition (Linear Aspect)
If $L \in \mathbb{R}^{S\times R}$, then $\lambda(\theta) = L\theta$ is called linear aspect of $\theta \in \mathbb{R}^R$.

2.1.2 Definition (Linear identifiability)
A linear aspect $\lambda(\theta) = L\theta$ is called linear identifiable at X_d or $d = (t_1, \dots, t_N)$, respectively, if and only if for all $\theta \in \mathbb{R}^R$ it holds

$$X_d\theta = 0_N \implies L\theta = 0_S.$$

2.1.3 Theorem
The linear aspect $\lambda(\theta) = L\theta$ is linear identifiable at d if and only if there exists $K \in \mathbb{R}^{S\times N}$ such that $L = KX_d$.

Proof.
$\Leftarrow$: Clear.
$\Rightarrow$: Let be $b \in \mathbb{R}^R$ arbitrary and set

$$\theta = b - (X_d^\top X_d)^- X_d^\top X_d\, b.$$

Then Lemma 8.1.2 b) provides

$$X_d\theta = X_d(b - (X_d^\top X_d)^- X_d^\top X_d\, b) = X_d b - X_d(X_d^\top X_d)^- X_d^\top X_d\, b = X_d b - X_d b = 0.$$

The linear identifiability of $\lambda(\theta) = L\theta$ implies

$$0 = L\theta = Lb - L(X_d^\top X_d)^- X_d^\top X_d\, b$$

and thus

$$Lb = L(X_d^\top X_d)^- X_d^\top X_d\, b$$

for all $b \in \mathbb{R}^R$. This means

$$L = L(X_d^\top X_d)^- X_d^\top X_d = K X_d$$

with $K = L(X_d^\top X_d)^- X_d^\top$. □

2.1.4 Lemma
If the linear aspect $\lambda(\theta) = L\theta$ is linear identifiable at a subdesign $d(1)$ of $d = (d(1), d(2))$ then it is also identifiable at d.

2.2 Estimators

2.2.1 Definition
An estimator $\widehat{\theta} := \widehat{\theta}(y) := \widehat{\theta}(y,d)$ *satisfying*

$$\widehat{\theta} \in \arg\min_{\theta\in\mathbb{R}^R} (y - X_d\theta)^\top (y - X_d\theta)$$

is called a least squares estimator of θ*.*

2.2.2 Theorem
Let be $\widehat{\theta} = (X_d^\top X_d)^- X_d^\top y$*. Any estimate* $\widetilde{\theta}$ *with* $X_d\widetilde{\theta} = X_d\widehat{\theta}$ *satisfies*

$$\widetilde{\theta} \in \arg\min_{\theta\in\mathbb{R}^R} (y - X_d\theta)^\top (y - X_d\theta),$$

i.e. $\widetilde{\theta}$ *is a least squares estimator.*

2.2.3 Theorem
If $Y = X_d\theta + Z$*,* $E(Z) = 0_N$*,* $Cov(Z) = \sigma^2 I_{N\times N}$*,* $\widehat{\theta}(y) = (X_d^\top X_d)^- X_d^\top y$*, and* $\lambda(\theta) = L\theta$ *is identifiable at* X_d*, then:*

a) $L\widehat{\theta}(Y)$ *is unbiased estimator for* $L\theta$*.*

b) $\widehat{\sigma}^2(Y) = \frac{1}{N-\text{rk}(X_d)}\, Y^\top \left(I_{N\times N} - X_d(X_d^\top X_d)^- X_d^\top\right) Y$ *is unbiased estimator for* σ^2*.*

c) $L\widehat{\theta}(y)$ *and* $\widehat{\sigma}^2(y)$ *do not dependent on the choice of the g-inverse for all* $y \in \mathbb{R}^N$*.*

d) $Cov(L\widehat{\theta}(Y)) = L\,(X_d^\top X_d)^- L^\top\ \sigma^2$ *and* $Cov(L\widehat{\theta}(Y))$ *does not depend on the choice of the g-inverse.*

Estimators and designs can be compared by their covariance matrices.

Proof of Theorem 2.2.3 d). Since $\lambda(\theta) = L\theta$ is identifiable at d, there exists $K \in \mathbb{R}^{S\times N}$ with $L = KX_d$. Lemma 8.1.2 b) imply

$$\begin{aligned}
&\text{Cov}(L\widehat{\theta}(Y)) = \text{Cov}(KX_d\,(X_d^\top X_d)^- X_d^\top Y) = KX_d\,(X_d^\top X_d)^- X_d^\top \text{Cov}(Y) X_d\,(X_d^\top X_d)^- X_d^\top\, K^\top \\
&\quad= KX_d\,(X_d^\top X_d)^- X_d^\top \text{Cov}(Z) X_d\,(X_d^\top X_d)^- X_d^\top\, K^\top \\
&\quad= KX_d\,(X_d^\top X_d)^- X_d^\top \sigma^2 I_{N\times N} X_d\,(X_d^\top X_d)^- X_d^\top\, K^\top \\
&\quad= K\,\underbrace{X_d\,(X_d^\top X_d)^- X_d^\top X_d}_{=X_d}\,(X_d^\top X_d)^- X_d^\top\, K^\top\, \sigma^2 \\
&\quad= KX_d\,(X_d^\top X_d)^- X_d^\top\, K^\top\, \sigma^2 = L\,(X_d^\top X_d)^- L^\top\, \sigma^2.
\end{aligned}$$

□

2.2.1 Example (One-way layout: Continuation of Example 2.1.1)
Assume that level a_1 (shortly level 1) of factor A is the control level (the placebo, the standard crop etc.) and that the effects of the $A-1$ other levels of the factor should be estimated as

additional effect to the effect of level 1. These additional effects can be positive or negative. Assume that the observations/measurements are ordered according to the factor levels so that

$$\begin{aligned} d &= (\underbrace{1,\ldots,1}_{N_1},\underbrace{2,\ldots,2}_{N_2},\ldots,\underbrace{A,\ldots,A}_{N_A}) \\ y &= (y_1,\ldots,y_N)^\top = (y_{11},\ldots,y_{1N_1},\ldots,y_{A1},\ldots,y_{AN_A})^\top = (y_{1*}^\top, y_{2*}^\top,\ldots,y_{A*}^\top)^\top, \end{aligned}$$

where $N = N_1 + N_2 + \ldots + N_A$ and

$$y_{a*} = (y_{a1},\ldots,y_{aN_a})^\top$$

for all $a = 1,\ldots,A$. Set also

$$\begin{aligned} y_{a\bullet} &= 1_{N_a}^\top y_{a*} = \sum_{n=1}^{N_a} y_{an}, \\ \overline{y}_{a\bullet} &= \frac{1}{N_a}\, y_{a\bullet}, \end{aligned}$$

for $a = 1,\ldots,N$, and

$$\begin{aligned} y_{\bullet\bullet} &= \sum_{a=1}^{A}\sum_{n=1}^{N_a} y_{an} = \sum_{n=1}^{N} y_n, \\ \overline{y} = \overline{y}_{\bullet\bullet} &= \frac{1}{N}\, y_{\bullet\bullet}. \end{aligned}$$

Non-singular parameterization:
If

$$\begin{aligned} x(t_n) &= (1\!\!1_{\{1\}}(t_n), 1\!\!1_{\{2\}}(t_n),\ldots,1\!\!1_{\{A\}}(t_n))^\top \in \mathbb{R}^A, \\ \theta &= (\mu_1,\mu_2,\ldots,\mu_A)^\top \in \mathbb{R}^A, \end{aligned}$$

then the interesting aspect $\lambda(\theta)$ is

$$\lambda(\theta) = \begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = L\theta \in \mathbb{R}^{A-1}$$

with

$$L = \begin{pmatrix} -1_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \in \mathbb{R}^{(A-1)\times A}.$$

The design matrix X_d has here in the general case the form

$$X_d = \begin{pmatrix} 1_{N_1} & 0_{N_1} & 0_{N_1} & \dots & 0_{N_1} & 0_{N_1} \\ 0_{N_2} & 1_{N_2} & 0_{N_2} & \dots & 0_{N_2} & 0_{N_2} \\ 0_{N_3} & 0_{N_3} & 1_{N_3} & \dots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \dots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 0_{N_A} & 0_{N_A} & 0_{N_A} & \dots & 0_{N_A} & 1_{N_A} \end{pmatrix}$$

so that

$$\begin{aligned} (X_d^\top X_d)^- &= \begin{pmatrix} N_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & N_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & N_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N_{A-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & N_A \end{pmatrix}^{-1} \\ &= \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1} = \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right), \end{aligned}$$

where $\operatorname{diag}(b_1, b_2, \dots, b_N) \in \mathbb{R}^{N\times N}$ denotes a diagonal matrix with diagonal elements $b_1, b_2, \dots, b_N$. Since

$$X_d^\top y = \begin{pmatrix} y_{1\bullet} \\ y_{2\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain

$$\widehat{\theta} = (X_d^\top X_d)^- X_d^\top y = (X_d^\top X_d)^{-1} X_d^\top y = \begin{pmatrix} \frac{1}{N_1} y_{1\bullet} \\ \frac{1}{N_1} y_{2\bullet} \\ \vdots \\ \frac{1}{N_1} y_{A\bullet} \end{pmatrix} = \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} \\ \vdots \\ \overline{y}_{A\bullet} \end{pmatrix},$$

which is the unique estimator for θ. Then

$$L(X_d^\top X_d)^{-1} X_d^\top y = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}$$

is the unique estimator for $\lambda(\theta) = L\theta$. That $\lambda(\theta) = L\theta$ is identifiable at X_d follows with Theorem 2.1.3 from the fact that $X_d^\top X_d$ is non-singular since

$$L = L(X_d^\top X_d)^{-1} X_d^\top X_d = K X_d.$$

Control parameterization:
Here we have (see Example 2.1.1)

$$x(t_n) = (1, \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^A,$$
$$\theta = (\mu, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^A,$$

so that

$$\lambda(\theta) = (\alpha_2, \ldots, \alpha_A)^\top = L\theta$$

with

$$L = \begin{pmatrix} 0_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \in \mathbb{R}^{(A-1)\times A}$$

is the interesting aspect. The design matrix X_d has here in the general case the form

$$X_d = \begin{pmatrix} 1_{N_1} & 0_{N_1} & 0_{N_1} & \ldots & 0_{N_1} & 0_{N_1} \\ 1_{N_2} & 1_{N_2} & 0_{N_2} & \ldots & 0_{N_2} & 0_{N_2} \\ 1_{N_3} & 0_{N_3} & 1_{N_3} & \ldots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \ldots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 1_{N_A} & 0_{N_A} & 0_{N_A} & \ldots & 0_{N_A} & 1_{N_A} \end{pmatrix}$$

so that

$$(X_d^\top X_d)^- = \begin{pmatrix} N & N_2 & N_3 & \ldots & N_{A-1} & N_A \\ N_2 & N_2 & 0 & \ldots & 0 & 0 \\ N_3 & 0 & N_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ N_{A-1} & 0 & 0 & \ldots & N_{A-1} & 0 \\ N_A & 0 & 0 & \ldots & 0 & N_A \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} N & b^\top \\ b & \operatorname{diag}(N_2, N_3, \ldots, N_A) \end{pmatrix}^{-1},$$

where $b = (N_2, N_3, \ldots, N_A)^\top$. The inverse of $X_d^\top X_d$ is given by Lemma 8.3.1. For applying this lemma, set $B = b$, $C = \text{diag}(N_2, N_3, \ldots, N_A)$, and $A = N$. Then

$$\begin{aligned}
\widetilde{E} &= N - b^\top \text{diag}(N_2, N_3, \ldots, N_A)^{-1}\, b = N - b^\top \text{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \ldots, \frac{1}{N_A}\right) b \\
&= N - \sum_{a=2}^{A} N_a = N_1, \\
\widetilde{E}^- B^\top C^{-1} &= \frac{1}{N_1}\, b^\top \text{diag}(N_2, N_3, \ldots, N_A)^{-1} \\
&= \frac{1}{N_1}\, b^\top \text{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \ldots, \frac{1}{N_A}\right) = \frac{1}{N_1} 1_{A-1}^\top \\
& C^{-1} + C^{-1} B\, \widetilde{E}^- B^\top C^{-1} \\
&= \text{diag}(N_2, N_3, \ldots, N_A)^{-1} + \text{diag}(N_2, N_3, \ldots, N_A)^{-1} b\, \frac{1}{N_1}\, b^\top \text{diag}(N_2, N_3, \ldots, N_A)^{-1} \\
&= \text{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \ldots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{A-1} 1_{A-1}^\top \\
&= \text{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \ldots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{(A-1)\times(A-1)}
\end{aligned}$$

such that

$$(X_d^\top X_d)^{-1} = \begin{pmatrix} \frac{1}{N_1} & -\frac{1}{N_1} 1_{A-1}^\top \\ -\frac{1}{N_1} 1_{A-1} & \text{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \ldots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{(A-1)\times(A-1)} \end{pmatrix}.$$

With

$$X_d^\top y = \begin{pmatrix} y_{\bullet\bullet} \\ y_{2\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain

$$\widehat{\theta} = (X_d^\top X_d)^{-1} X_d^\top y = \begin{pmatrix} \frac{y_{\bullet\bullet}}{N_1} - \frac{1}{N_1} 1_{A-1}^\top (y_{2\bullet}, \ldots, y_{A\bullet})^\top \\ -\frac{y_{\bullet\bullet}}{N_1} 1_{A-1} + (\frac{y_{2\bullet}}{N_2}, \frac{y_{3\bullet}}{N_3}, \ldots, \frac{y_{A\bullet}}{N_A})^\top + \frac{1}{N_1} 1_{A-1}\, 1_{A-1}^\top (y_{2\bullet}, \ldots, y_{A\bullet})^\top \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{N_1}\left(y_{\bullet\bullet} - \sum_{a=2}^A y_{a\bullet}\right) \\ (\overline{y}_{2\bullet}, \overline{y}_{3\bullet}, \ldots, \overline{y}_{A\bullet})^\top - 1_{A-1} \frac{1}{N_1}\left(y_{\bullet\bullet} - \sum_{a=2}^A y_{a\bullet}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix}$$

which is the unique estimator for θ. Then

$$L(X_d^\top X_d)^{-1} X_d^\top y = \begin{pmatrix} 0_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}$$

is the same unique estimator for $\lambda(\theta) = L\theta$ as we obtained for the non-singular parametrization. That $\lambda(\theta) = L\theta$ is identifiable at X_d follows as for the non-singular parametrization from the fact that $X_d^\top X_d$ is non-singular.

Singular parameterization:
Here we have (see Example 2.1.1)

$$x(t_n) = (1, \mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^{A+1},$$
$$\theta = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^{A+1}.$$

so that

$$\lambda(\theta) = (\alpha_2 - \alpha_1, \ldots, \alpha_A - \alpha_1)^\top = L\theta$$

with

$$L = \begin{pmatrix} 0_{A-1} & -1_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \in \mathbb{R}^{(A-1)\times(A+1)}$$

is the interesting aspect. The design matrix X_d has here in the general case the form

$$X_d = \begin{pmatrix} 1_{N_1} & 1_{N_1} & 0_{N_1} & 0_{N_1} & \dots & 0_{N_1} & 0_{N_1} \\ 1_{N_2} & 0_{N_2} & 1_{N_2} & 0_{N_2} & \dots & 0_{N_2} & 0_{N_2} \\ 1_{N_3} & 0_{N_3} & 0_{N_3} & 1_{N_3} & \dots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 1_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \dots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 1_{N_A} & 0_{N_A} & 0_{N_A} & 0_{N_A} & \dots & 0_{N_A} & 1_{N_A} \end{pmatrix} \tag{2.3}$$

so that

$$\begin{aligned} X_d^\top X_d &= \begin{pmatrix} N & N_1 & N_2 & N_3 & \dots & N_{A-1} & N_A \\ N_1 & N_1 & 0 & 0 & \dots & 0 & 0 \\ N_2 & 0 & N_2 & 0 & \dots & 0 & 0 \\ N_3 & 0 & 0 & N_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ N_{A-1} & 0 & 0 & 0 & \dots & N_{A-1} & 0 \\ N_A & 0 & 0 & 0 & \dots & 0 & N_A \end{pmatrix} \\ &= \begin{pmatrix} N & b^\top \\ b & \operatorname{diag}(N_1, N_2, \dots, N_A) \end{pmatrix} \in \mathbb{R}^{(A+1)\times(A+1)}, \end{aligned} \tag{2.4}$$

where $b = (N_1, N_2, \dots, N_A)^\top$. Here $X_d^\top X_d$ is singular so that only a g-inverse can be calculated. The g-inverse of $X_d^\top X_d$ is given by Lemma 8.3.1. For applying this lemma, set $B = b$, $A = \operatorname{diag}(N_1, N_2, \dots, N_A)$, and $C = N$. Then

$$\begin{aligned} E &= N - b^\top \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1}\, b = N - b^\top \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right) b \\ &= N - \sum_{a=1}^{A} N_a = N - N = 0, \end{aligned}$$

Then E^- can be any value $c \in \mathbb{R}$ since $0c0 = 0$. Hence set $E^- = c \in \mathbb{R}$. Then

$$\begin{aligned} & E^- B^\top A^{-1} = c\, b^\top \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1} \\ &= c\, b^\top \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right) = c\, 1_A^\top \\ & A^{-1} + A^{-1} B\, E^- B^\top A^{-1} \\ &= \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1} + \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1} b\, c\, b^\top \operatorname{diag}(N_1, N_2, \dots, N_A)^{-1} \\ &= \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right) + c\, 1_A 1_A^\top \\ &= \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right) + c\, 1_{A\times A} \end{aligned}$$

such that

$$(X_d^\top X_d)^- = \begin{pmatrix} c & -c\,1_A^\top \\ -c\,1_A & \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right) + c\,1_{A\times A} \end{pmatrix}.$$

With

$$X_d^\top y = \begin{pmatrix} y_{\bullet\bullet} \\ y_{1\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain that

$$\widehat{\theta} = (X_d^\top X_d)^- X_d^\top y = \begin{pmatrix} c\,y_{\bullet\bullet} - c\,1_A^\top\,(y_{1\bullet}, \ldots, y_{A\bullet})^\top \\ -c\,y_{\bullet\bullet}\,1_A + (\frac{y_{1\bullet}}{N_1}, \frac{y_{2\bullet}}{N_2}, \ldots, \frac{y_{A\bullet}}{N_A})^\top + c\,1_A\,1_A^\top\,(y_{1\bullet}, \ldots, y_{A\bullet})^\top \end{pmatrix}$$

$$= \begin{pmatrix} c\left(y_{\bullet\bullet} - \sum_{a=1}^A y_{a\bullet}\right) \\ (\overline{y}_{1\bullet}, \overline{y}_{2\bullet}, \ldots, \overline{y}_{A\bullet})^\top - 1_A\,c\left(y_{\bullet\bullet} - \sum_{a=1}^A y_{a\bullet}\right) \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} \\ \vdots \\ \overline{y}_{A\bullet} \end{pmatrix}$$

is a least squares estimator for θ for all $c \in \mathbb{R}$. Hence $\widehat{\theta}$ is unique although $(X_d^\top X_d)^-$ is not unique. However, $\widehat{\theta}$ is not the only least squares estimator since $\widehat{\theta} + \mu\gamma$ with $\gamma = (1, -1, -1, \ldots, -1)^\top$ and $\mu \in \mathbb{R}$ is also a least squares estimator according to Theorem 2.2.2 since $X_d\gamma = 0$ so that $X_d\widehat{\theta} = X_d(\widehat{\theta} + \mu\gamma)$. The property $X_d\gamma = 0$ for $\gamma \neq 0_{A+1}$ means according to Definition 2.1.2 that θ is not identifiable at X_d. However, $\lambda(\theta) = L\theta$ is identifiable since

$$L = \begin{pmatrix} 0_{A-1} & -1_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} = K\,X_d$$

for

$$K = \begin{pmatrix} -\frac{1}{N_1}1_{N_1}^\top & \frac{1}{N_2}1_{N_2}^\top & 0 & 0 & \ldots & 0 \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & \frac{1}{N_3}1_{N_3}^\top & 0 & \ldots & 0 \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & 0 & \frac{1}{N_4}1_{N_4}^\top & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & 0 & 0 & \ldots & \frac{1}{N_A}1_{N_A}^\top \end{pmatrix} \in \mathbb{R}^{(A-1)\times N}.$$

In particular for any $\widetilde{\theta} = \widehat{\theta} + \mu\gamma$ with $\mu \in \mathbb{R}$ we have

$$
\begin{aligned}
L\widetilde{\theta} &= L\left((X_d^\top X_d)^{-1} X_d^\top y + \mu\gamma\right) \\
&= \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} \mu \\ \overline{y}_{1\bullet} - \mu \\ \overline{y}_{2\bullet} - \mu \\ \vdots \\ \overline{y}_{A\bullet} - \mu \end{pmatrix} = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}.
\end{aligned}
$$

This is the same unique estimator for $\lambda(\theta) = L\theta$ as we obtained for the other parametrizations.

Since all estimators for the additional effects of the $A-1$ factor levels which are not the control are unique and do not depend on the parametrization we can calculate the covariance matrix of the estimator with the non-singular parametrization:

$$
\begin{aligned}
\mathrm{Cov}(L\widehat{\theta}) &= \sigma^2\, L(X_d^\top X_d)^{-1} L^\top \\
&= \sigma^2 \begin{pmatrix} -1_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \mathrm{diag}\left(\frac{1}{N_1}, \dots, \frac{1}{N_A}\right) \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} -1_{A-1} & I_{(A-1)\times(A-1)} \end{pmatrix} \begin{pmatrix} -\frac{1}{N_1} 1_{A-1}^\top \\ \mathrm{diag}\left(\frac{1}{N_2}, \dots, \frac{1}{N_A}\right) \end{pmatrix} \\
&= \sigma^2 \left(\frac{1}{N_1} 1_{(A-1)\times(A-1)} + \mathrm{diag}\left(\frac{1}{N_2}, \dots, \frac{1}{N_A}\right)\right).
\end{aligned}
$$

For the special case of $A = 2$ we obtain

$$
\mathrm{Cov}(L\widehat{\theta}) = \mathrm{var}(\overline{Y}_{2\bullet} - \overline{Y}_{1\bullet}) = \sigma^2 \left(\frac{1}{N_1} + \frac{1}{N_2}\right).
$$

A further design question: Usually the N experiments are done in a specific temporal or spacial order. For examples plants are growing on specific positions of the field, animals are living in specific places of a cot, patients of a hospital arriving in a specific order to the hospital. Then the question is how to assign two different treatments to the experimental units. Since it is never clear if there are special spacial or temporal influences on the measurement, the assignments of the treatments to the experimental units should be done randomly. This allocation can be easily done by the function `design.crd` of the `agricolae` package.

To allocate 7 treatments `t1` and 5 treatments `t2` to 12 units, type:

```
> library(agricolae)
> design.crd(c("t1","t2"),c(7,5))$book
   plots r c("t1", "t2")
1    101 1            t1
2    102 2            t1
```

```
3     103 1            t2
4     104 3            t1
5     105 2            t2
6     106 3            t2
7     107 4            t1
8     108 4            t2
9     109 5            t1
10    110 6            t1
11    111 5            t2
12    112 7            t1
```

Then we get an order how to allocate the two treatments: at first `t2`, then `t2` again, then `t1` and so on.

If one do not need to allocate specific names for the treatments, then one can use also:

```
> r2dtable(1,rep(1,12),c(7,5))
[[1]]
      [,1] [,2]
 [1,]    1    0
 [2,]    1    0
 [3,]    1    0
 [4,]    0    1
 [5,]    0    1
 [6,]    0    1
 [7,]    0    1
 [8,]    1    0
 [9,]    1    0
[10,]    1    0
[11,]    0    1
[12,]    1    0
```

2.3 Tests

2.3.1 Example (One-way layout: Continuation of Example 2.2.1)
The aim is to test whether the different levels of the factor A have different effects for the oberservations. If they do not have different effects then we say that the factor A has no influence. Hence we have to decide between

$$H_0: \text{ Factor A has no influence}$$

versus

$$H_1: \text{ Factor A has an influence.}$$

The null hypothesis H_0 can be expressed in different forms for the different parametrizations:

Non-singular parametrization:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_A$$

$$\iff$$

$$H_0 : L\theta = \begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = 0 \text{ with } L = \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}.$$

Control parametrization:

$$H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_A = 0$$

$$\iff$$

$$H_0 : L\theta = \begin{pmatrix} \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_A \end{pmatrix} = 0 \text{ with } L = \left(0_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}.$$

Singular parametrization:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_A$$

$$\iff$$

$$H_0 : L\theta = \begin{pmatrix} \alpha_2 - \alpha_1 \\ \alpha_3 - \alpha_1 \\ \vdots \\ \alpha_A - \alpha_1 \end{pmatrix} = 0 \text{ with } L = \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times(A+1)}.$$

Hence in any parametrization, hypotheses can be formulated in form of

$$H_0: \; L\theta = l \quad \text{versus} \quad H_1: \; L\theta \neq l$$

with $L \in \mathbb{R}^{S \times R}$ of full rank, $l = Lb \in \mathbb{R}^S$ for some $b \in \mathbb{R}^R$. Thereby, $\lambda(\theta) = L\theta$ shall be identifiable at X_d.

Let be $\omega(X_d) := X_d(X_d^\top X_d)^- X_d^\top$ the perpendicular projection matrix onto $C(X_d) = \{X_d\theta;\; \theta \in \mathbb{R}^R\}$.

2.3.1 Theorem
Let be $Z \sim \mathcal{N}(0_N, \sigma^2 I_{N\times N})$, $q_{1-\alpha}$ the $1-\alpha$-quantile of the central F-distribution with rk(L) and rk$(I_{N\times N} - \omega(X_d))$ degrees of freedom, $L = KX_d$, and $l = L\,b$, then

$$1\!\!\mathrm{I}_{\{\widehat{V}(y) > q_{1-\alpha}\}}(y) \quad \textit{with} \quad \widehat{V}(y) = \frac{\frac{1}{rk(L)}\,(L\widehat{\theta} - l)^\top [L(X_d^\top X_d)^- L^\top]^- (L\widehat{\theta} - l)}{\frac{1}{rk(I_{N\times N} - \omega(X_d))}\; y^\top (I_{N\times N} - \omega(X_d))y}$$

is α-level test for $H_0: \; L\theta = l$ versus $H_1: \; L\theta \neq l$. Its β-error at θ with $L\theta \neq l$ is given by

$$F_{F\left(rk(L),\, rk(I_{N\times N} - P),\, \frac{(L\theta - l)^\top [L(X_d^\top X_d)^- L^\top]^- (L\theta - l)}{\sigma^2}\right)}(q_{1-\alpha})\,.$$

Thereby $F_{F(n,m,\delta)}$ is the distribution function of the F-distribution with n and m degrees of freedom and noncentrality parameter δ.

2.3.2 Remark (Designing experiments)
The aim of a good design is to minimize the β-error of the test. This means here that the non-centrality parameter

$$(L\widehat{\theta} - l)^\top [L(X_d^\top X_d)^- L^\top]^- (L\widehat{\theta} - l)$$

should be as large as possible for all θ with $L\theta \neq l$. This is achieved if

$$L(X_d^\top X_d)^- L^\top$$

is as small as possible. Since $L(X_d^\top X_d)^- L^\top$ is the covariance matrix of the estimator $L(X_d^\top X_d)^- X_d^\top y$ (see Theorem 2.2.3 d)), we realize that testing and estimation leads to the same optimization problem for designing experiments.

2.4 Partial parameter vectors

Often only a part of the parameter vector $\theta = (\theta_1, \dots, \theta_R)^\top \in \mathbb{R}^R$ is of interest. Without loss of generality, the parameter vector can be written as

$$\theta^\top = (\alpha^\top, \beta^\top) \quad \text{with} \quad \alpha = (\theta_1, \dots, \theta_A)^\top \in \mathbb{R}^A \quad \text{and} \quad \beta = (\theta_{A+1}, \dots, \alpha_R)^\top \in \mathbb{R}^B.$$

Then the linear model can be represented as

$$y = X_d\theta + z = (V_d, W_d) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + z = V_d\alpha + W_d\beta + z,$$

where $V_d = (v(t_1), \dots, v(t_N))^\top \in \mathbb{R}^{N\times A}$ and $W_d = (w(t_1), \dots, w(t_N))^\top \in \mathbb{R}^{N\times B}$ are the design matrices corresponding to α and β. In particular a single observation is given by

$$y_n = x(t_n)^\top\theta + z_n = (v(t_n)^\top, w(t_n)^\top) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + z_n = v(t_n)^\top\alpha + w(t_n)^\top\beta + z_n.$$

The aspect α is then a linear aspect of the form

$$\lambda(\theta) = (I_{A\times A}, 0_{A\times B})\,\theta = \alpha$$

where $I_{A\times A}$ is the $A\times A$ identity matrix and $0_{A\times B}$ is a $A\times B$ matrix consisting only of zeros. Hence L of the linear aspect satisfies $L = (I_{A\times A}, 0_{A\times B})$.

Since the covariance matrix of any Gauss-Markov estimator $L(X_d^\top X_d)^- X_d^\top y$ for $L\theta$ is given by $L(X_d^\top X_d)^- L^\top \sigma^2$ and the β-error for a test on $L\theta$ depends also only $L(X_d^\top X_d)^- L^\top$ we have to derive $L(X_d^\top X_d)^- L^\top$ for $L = (I_{A\times A}, 0_{A\times B})$ and $X_d = (V_d, W_d)$.

2.4.1 Definition
Let be $X_d \in \mathbb{R}^{N\times R}$.
a) $\omega(X_d) := X_d(X_d^\top X_d)^- X_d^\top$ denotes the perpendicular projection matrix onto $C(X_d)$.
b) $\omega^\perp(X_d) := I_{N\times N} - X_d(X_d^\top X_d)^- X_d^\top$ denotes the perpendicular projection matrix onto $C(X_d)^\perp$.

2.4.2 Lemma
If $X_d = (V_d, W_d)$ with $V_d \in \mathbb{R}^{N\times A}$ and $W_d \in \mathbb{R}^{N\times B}$, then it holds

$$(X_d^\top X_d)^- = \begin{pmatrix} (V_d^\top V_d)^- + (V_d^\top V_d)^- V_d^\top W_d\, E^-\, W_d^\top V_d (V_d^\top V_d)^- & -(V_d^\top V_d)^- V_d^\top W_d\, E^- \\ -E^-\, W_d^\top V_d (V_d^\top V_d)^- & E^- \end{pmatrix}$$

$$= \begin{pmatrix} F^- & -F^-\, V_d^\top W_d (W_d^\top W_d)^- \\ -(W_d^\top W_d)^- W_d^\top V_d\, F^- & (W_d^\top W_d)^- + (W_d^\top W_d)^- W_d^\top V_d\, F^-\, V_d^\top W_d (W_d^\top W_d)^- \end{pmatrix}$$

with $E = W_d^\top \omega^\perp(V_d)\, W_d$ and $F = V_d^\top \omega^\perp(W_d)\, V_d$.

Proof. At first note that

$$X_d^\top X_d = \begin{pmatrix} V_d^\top V_d & V_d^\top W_d \\ W_d^\top V_d & W_d^\top W_d \end{pmatrix}.$$

Then the first assertion follows from Lemma 8.3.1 setting there $A = V_d^\top V_d$, $B = W_d^\top V_d$ and $C = W_d^\top W_d$. Thereby, the conditions $AA^- B^\top = V_d^\top V_d (V_d^\top V_d)^- V_d^\top W_d = V_d^\top W_d = B^\top$ and $BA^- A = W_d^\top V_d (V_d^\top V_d)^- V_d^\top V_d = W_d^\top V_d$ are satisfied because of Lemma 8.1.2 b). In particular it holds

$$E = C - B\, A^-\, B^\top = W_d^\top W_d - W_d^\top V_d (V_d^\top V_d)^-\, V_d^\top W_d = W_d^\top (I_{N\times N} - V_d (V_d^\top V_d)^- V_d^\top)\, W_d.$$

The second equality follows by interchanging the role of A and C in Lemma 8.3.1. □

2.4.3 Theorem
If $X_d = (V_d, W_d)$ with $V_d \in \mathbb{R}^{N\times A}$ and $W_d \in \mathbb{R}^{N\times B}$ and $L = (I_{A\times A}, 0_{A\times B})$, then it holds

$$\begin{aligned} & L(X_d^\top X_d)^- L^\top \\ = \ & (V_d^\top V_d)^- + (V_d^\top V_d)^- V_d^\top W_d\, (W_d^\top \omega^\perp(V_d)\, W_d)^-\, W_d^\top V_d (V_d^\top V_d)^- = (V_d^\top \omega^\perp(W_d)\, V_d)^-. \end{aligned}$$

Proof. The both equalities follow directly from Lemma 2.4.2. □

2.4.4 Corollary
If $L = (I_{A\times A}, 0_{A\times B})$ and $X_d = (V_d, W_d)$, where $V_d \in \mathbb{R}^{N\times A}$ and $W_d \in \mathbb{R}^{N\times B}$ satisfy $V_d^\top W_d = 0_{A\times B}$, then it holds

$$L(X_d^\top X_d)^- L^\top = (V_d^\top V_d)^-.$$

Corollary 2.4.4 means that estimating α or testing a hypothesis on α in a model given by $Y = X_d\theta + Z = (V_d, W_d) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + Z$ with $V_d^\top W_d = 0_{A\times B}$ leads to the same covariance matrix and β-error of a test using the smaller model $Y = V_d\alpha + Z$.

2.4.5 Theorem
If $Y = X_d\theta + Z = V_d\alpha + W_d\beta + Z$ and $\lambda(\theta) = L\theta = L_\alpha\alpha$ with $L_\alpha \in \mathbb{R}^{s\times A}$, then it holds:
a) $\lambda(\theta)$ is identifiable at X_d if and only if there exists a $K \in \mathbb{R}^{s\times A}$ such that $L_\alpha = K\, V_d^\top \omega^\perp(W_d) V_d$.
b) If $\lambda(\theta)$ is identifiable at X_d, then the Gauss-Markov estimator for $\lambda(\theta)$ satisfies

$$L\widehat{\theta}(y) = L_\alpha \left[V_d^\top \omega^\perp(W_d)\, V_d\right]^-\; V_d^\top\, \omega^\perp(W_d)\, y$$

with

$$Cov(L\widehat{\theta}(Y)) = L_\alpha \left[V_d^\top \omega^\perp(W_d)\, V_d\right]^-\; L_\alpha^\top\, \sigma^2.$$

In particular, it holds

$$L(X_d^\top X_d)^- L^\top = L_\alpha \left[V_d^\top \omega^\perp(W_d)\, V_d\right]^-\; L_\alpha^\top.$$

Proof.
a) $\Rightarrow$: If $\lambda(\theta)$ is identifiable then there exists $\tilde{K} \in \mathbb{R}^{s\times N}$ with

$$L = (L_\alpha, 0_{s\times B}) = \tilde{K}\,(V_d, W_d) = (\tilde{K}\, V_d, \tilde{K}\, W_d)$$

which means

$$L_\alpha = \tilde{K}\,V_d \quad \text{and} \quad \tilde{K}\,W_d = 0_{s\times B}.$$

This implies

$$\begin{aligned} L_\alpha &= L_\alpha - 0_{s\times A} \\ &= \tilde{K}\,V_d - 0_{s\times B}(W_d^\top W_d)^-\,W_d^\top V_d = \tilde{K}\,V_d - \tilde{K}\,W_d(W_d^\top W_d)^-\,W_d^\top V_d \\ &= \tilde{K}\,(I_{N\times N} - W_d(W_d^\top W_d)^-\,W_d^\top)\,V_d = \tilde{K}\,\omega^\perp(W_d)\,V_d. \end{aligned}$$

Lemma 8.1.2 b) with $A = \omega^\perp(W_d)\,V_d$ yields, because $\omega^\perp(W_d)$ is idempotent,

$$\begin{aligned} L_\alpha &= \tilde{K}\,\omega^\perp(W_d)\,V_d\,\left[V_d^\top\omega^\perp(W_d)^\top\omega^\perp(W_d)V_d\right]^-\,V_d^\top\omega^\perp(W_d)^\top\omega^\perp(W_d)\,V_d \\ &= \tilde{K}\,\omega^\perp(W_d)\,V_d\,\left[V_d^\top\omega^\perp(W_d)V_d\right]^-\,V_d^\top\omega^\perp(W_d)\,V_d \\ &= K\,V_d^\top\omega^\perp(W_d)\,V_d. \end{aligned}$$

$\Leftarrow$: $L_\alpha = K\,V_d^\top\omega^\perp(W_d)\,V_d$ implies $L_\alpha = \tilde{K}\,V_d$ with $\tilde{K} = K\,V_d^\top\omega^\perp(W_d)$. Since $\omega^\perp(W_d)\,W_d = (I_{N\times N} - W_d(W_d^\top W_d)^-\,W_d^\top)W_d = 0_{N\times B}$ according to Lemma 8.1.2 b), we obtain

$$\tilde{K}(V_d, W_d) = (L_\alpha, 0_{s\times B}) = L.$$

b) Lemma 2.4.2 provides

$$\begin{aligned} L\widehat{\theta}(y) &= L(X_d^\top X_d)^-\,X_d^\top y \\ &= (L_\alpha, 0_{s\times B})\begin{pmatrix} F^- & -F^-\,V_d^\top W_d(W_d^\top W_d)^- \\ -(W_d^\top W_d)^- W_d^\top V_d\,F^- & G \end{pmatrix}\begin{pmatrix} V_d^\top \\ W_d^\top \end{pmatrix} y \end{aligned}$$

with $G = (W_d^\top W_d)^- + (W_d^\top W_d)^- W_d^\top V_d\,F^-\,V_d^\top W_d(W_d^\top W_d)^-$ and $F = V_d^\top\omega^\perp(W_d)V_d$ so that

$$\begin{aligned} &L\widehat{\theta}(y) \\ &= \left[L_\alpha\,F^- V_d^\top - L_\alpha\,F^-\,V_d^\top W_d(W_d^\top W_d)^- W_d^\top\right]\,y \\ &= L_\alpha\,F^- V_d^\top\left[I_{N\times N} - W_d(W_d^\top W_d)^- W_d^\top\right]\,y = L_\alpha\,\left[V_d^\top\omega^\perp(W_d)\,V_d\right]^-\,V_d^\top\omega^\perp(W_d)\,y. \end{aligned}$$

Moreover, Theorem 2.2.3 d) yields

$$\begin{aligned} \mathrm{Cov}(L\widehat{\theta}(Y)) &= L(X_d^\top X_d)^-\,L^\top\,\sigma^2 \\ &= (L_\alpha, 0_{s\times B})\begin{pmatrix} F^- & -F^-\,V_d^\top W_d(W_d^\top W_d)^- \\ -(W_d^\top W_d)^- W_d^\top V_d\,F^- & G \end{pmatrix}\begin{pmatrix} L_\alpha^\top \\ 0_{B\times s} \end{pmatrix}\sigma^2 \\ &= L_\alpha\,F^-\,L_\alpha^\top\,\sigma^2 = L_\alpha\,\left[V_d^\top\omega^\perp(W_d)\,V_d\right]^-\,L_\alpha^\top\,\sigma^2. \end{aligned}$$

□

2.4.1 Example (One-way layout: Continuation of Example 2.2.1)
In the singular parametrization of the one-way layout we have

$$y_n = \mu + \alpha_a + z_n = x(t_n)^\top \theta \quad \text{if } t_n = a, \quad \text{for } n = 1, \ldots, N,$$

with

$$x(t_n) = (1, \mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^{A+1},$$
$$\theta = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^{A+1}.$$

Then according to (2.3) the design matrix is given by

$$X_d = (1_N, V_d)$$

with

$$V_d = \begin{pmatrix} 1_{N_1} & 0_{N_1} & 0_{N_1} & \ldots & 0_{N_1} & 0_{N_1} \\ 0_{N_2} & 1_{N_2} & 0_{N_2} & \ldots & 0_{N_2} & 0_{N_2} \\ 0_{N_3} & 0_{N_3} & 1_{N_3} & \ldots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & & \vdots & \vdots & \\ 0_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \ldots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 0_{N_A} & 0_{N_A} & 0_{N_A} & \ldots & 0_{N_A} & 1_{N_A} \end{pmatrix}.$$

Because

$$\omega^\perp(1_N) = I_{N\times N} - 1_N(1_N^\top 1_N)^- 1_N^\top = I_{N\times N} - \frac{1}{N} 1_{N\times N}$$

and

$$1_N^\top V_d = (N_1, N_2, \ldots, N_A) =: b^\top$$

we have

$$H := V_d^\top \omega^\perp(1_N)\, V_d = V_d^\top V_d - \frac{1}{N}\, b\, b^\top = \text{diag}(N_1, N_2, \ldots, N_A) - \frac{1}{N}\, b\, b^\top.$$

Hence, according to Lemma 2.4.5, a linear aspect $\lambda(\theta) = L_\alpha \alpha$ with $\alpha = (\alpha_1, \ldots, \alpha_A)^\top$ is identifiable, if there exists K with $L_\alpha = K\, H$. From Example 2.2.1, we know that already that

$$\lambda(\theta) = (\alpha_2 - \alpha_1, \ldots, \alpha_A - \alpha_1)^\top$$

is identifiable. But what about

$$\lambda(\theta) = \alpha_1 + \alpha_2 + \ldots + \alpha_A = 1_A^\top \alpha?$$

Assume there exists $K \in \mathbb{R}^{1\times A}$ with $1_A^\top = K\, H$. But since $V_d\, 1_A = 1_N$ and $\omega^\perp(1_N)\, 1_N = 0_N$ we get the contradiction

$$0 \neq A = 1_A^\top 1_A = K\, H\, 1_A = K\, V_d^\top \omega^\perp(1_N)\, V_d\, 1_A = K\, V_d^\top \omega^\perp(1_N)\, 1_N = 0,$$

so that $\lambda(\theta) = \alpha_1 + \alpha_2 + \ldots + \alpha_A$ is not identifiable. Are there others vectors $v = (v_1, \ldots, v_A)^\top$ satisfying $H\,v = 0_A$? Since $H\,v = 0_A$ means

$$
\begin{aligned}
&\left(\text{diag}(N_1, N_2, \ldots, N_A) - \frac{1}{N}\,b\,b^\top\right)\,v = 0_A \\
\iff\quad & \begin{pmatrix} N_1\,v_1 \\ \vdots \\ N_A\,v_A \end{pmatrix} - \frac{1}{N}\begin{pmatrix} N_1 \\ \vdots \\ N_A \end{pmatrix} b^\top\,v = 0_A \\
\iff\quad & \begin{pmatrix} N_1(v_1 - \frac{1}{N}\,b^\top v) \\ \vdots \\ N_A(v_A - \frac{1}{N}\,b^\top v) \end{pmatrix} = 0_A \\
\iff\quad & v_a = \frac{1}{N}\,b^\top v, \quad \text{for } a = 1, \ldots, A,
\end{aligned}
$$

the only v satisfying $H\,v = 0_A$ is $v = c\,1_A$. Hence $C(H)$ is $A-1$ dimensional. Since

$$
v \in C(H) \iff H\,K^\top = v \iff v^\top = K\,H,
$$

any $v \in C(H)$ satisfies $v^\top 1_A = K\,H\,1_A = K 0_N = 0$ so that $C(H)$ is the space of all vectors v which are orthogonal to 1_A. This means that any $\lambda(\theta) = v^\top\alpha$ with $v^\top 1_A = 0$ is identifiable. In particular any $L_\alpha\alpha$ with $L_\alpha 1_A = 0_s$ is identifiable. Such linear aspects are called **contrasts**.

Chapter 3

Optimal designs

The first aim of an experimental design is that all interesting aspects of the assumed model are identifiable. If there are many factors of quantitative and qualitative type, then this is no simple task. The theory of fractional factorial designs leads to designs where given aspects of the model are identifiable (see the book of Mukerjee and Wu, 2006). For more complex models also algebraic methods are helpful. In particular it can decided with the theory of Gröbner bases which models and which aspects of models are identifiable if a design is already given (see the book of Pistone, Riccomagno, and Wynn 2001). These concepts however are beyond this lecture.

Another practical aspect of designing experiments are balanced designs. Balanced complete designs provide the property that the parameter can be estimated independently of the other parameters. Because of this property, the analysis of variance is not dependent of the order of the factors. Moreover models with random effects can be treated with the analysis of variance like models with fixed effects.

Additionally, a good design should provide precise estimates and a small β error for testing. A general setup for achieving this, is treated in this chapter.

3.1 Generalized designs

3.1.1 Definition
Let be $\mathcal{T}$ the experimental region and $x : \mathcal{T} \longrightarrow \mathbb{R}^R$ the known regression function.

$$d = (t_1, \ldots, d_N) \in \mathcal{T}^N$$

is called **concrete design** *and*

$$X = X_d = \begin{pmatrix} x(t_1)^\top \\ x(t_2)^\top \\ \vdots \\ x(t_N)^\top \end{pmatrix} \in \mathbb{R}^{N \times R}$$

is the corresponding design matrix.

The aim of a good design d is to maximize the power of the test for testing $H_0 : L\theta = l$ or/and to minimize the covariance matrix of the Gauss-Markov estimator $L(X_d^\top X_d)^- X_d^\top y$ for $\lambda(\theta) = L\,\theta$. Fortunately the problem for testing as well as the problem for estimation leads to the same optimization problem, namely to minimize

$$L(X_d^\top X_d)^- L^\top,$$

see Remark 2.3.2.

The design problem
Find a design

$$d \in \Delta \subset \Delta_{N,\lambda} := \left\{ d_N \in \mathcal{T}^N;\ \lambda(\theta) \text{ is identifiable at } d \right\}$$

such that $L(X_d^\top X_d)^- L^\top$ is minimal. If $L(X_d^\top X_d)^- L^\top$ is of full rank then the minimzation of $L(X_d^\top X_d)^- L^\top$ is equivalent with the maximization of $(L(X_d^\top X_d)^- L^\top)^{-1}$.

The only problem is, that $L \in \mathbb{R}^{S\times R}$ implies $L(X_d^\top X_d)^- L^\top \in \mathbb{R}^{S\times S}$ so that we have to minimize a matrix as soon as $S > 1$. On the set of $S \times S$-matrices we have only a partial ordering given for example by the Loewner ordering which means

$$A \le B \iff c^\top (B - A)\, c \ge 0 \text{ for all } c \in \mathbb{R}^S.$$

Partial ordering means that there are matrices which cannot be compared. If we have a special set of matrices $\{A;\ A \in \mathcal{A}\}$, it could be that there is no matrix $A_0 \in \mathcal{A}$ with

$$A_0 \le A \quad \text{for all} \quad A \in \mathcal{A}.$$

This happens in particular, if we compare different designs.

To reduce the dimension S, we always will assume here, that L is of full rank and that $\lambda(\theta) = L\theta$ is identifiable at d.

3.1.2 Lemma
If $\lambda(\theta) = L\theta$ is identifiable at d and $L \in \mathbb{R}^{S\times R}$ is of full rank, i.e. $\text{rk}(L) = S$, then $L(X_d^\top X_d)^- L^\top$ is invertible.

Proof. We always have $\text{rk}(L(X_d^\top X_d)^- L^\top) \le \text{rk}(L) = S$. Conversely, the identifiability implies with Theorem 2.1.3

$$\begin{aligned} S = \text{rk}(L) = \text{rk}(K\,X_d) &\overset{(\text{ Lemma 8.1.2 b)})}{=} \text{rk}(K\,X_d(X_d^\top X_d)^- X_d^\top X_d) \\ &\le \text{rk}(K\,X_d(X_d^\top X_d)^- X_d^\top) \overset{\text{rg}(AA^\top)=\text{rg}(A)}{=} \text{rk}(K\,X_d(X_d^\top X_d)^- X_d^\top X_d\,(X_d^\top X_d)^- X_d^\top K^\top) \\ &= \text{rk}(L(X_d^\top X_d)^- L^\top). \end{aligned}$$

Hence $L(X_d^\top X_d)^- L^\top$ is of rank S and thus invertible. □

3.1.3 Definition (Information matrix)
a) $I_\lambda(d) := (L(X_d^\top X_d)^- L^\top)^{-1}$ *is called information matrix for* $\lambda(\theta) = L\theta$ *at* d.

b) $I_\theta(d) := X_d^\top X_d$ *is called information matrix for* θ *at* d.

3.1.4 Remark
If $Y \sim \mathcal{N}(X_d\theta,\ \sigma^2 I_{N\times N})$ and σ^2 is known, then

$$\sigma^{-2}\, I_\theta(d) = \left(\mathrm{E}_\theta \left(\frac{\partial \ln f_\theta(Y)}{\partial \theta_s} \cdot \frac{\partial \ln f_\theta(Y)}{\partial \theta_r}\right)\right)_{r,s=1,\ldots,R},$$

i.e. $\sigma^{-2}\, I_\theta(d)$ is the Fisher information matrix.

The maximization of the information matrix $I_\lambda(d)$ within the concrete designs is a complicated task. Therefore the designs are generalized:

$$d = (t_1, \ldots, t_N) \;\longrightarrow\; \delta_N = \frac{1}{N}\sum_{n=1}^{N} e_{t_n} \;\longrightarrow\; \delta \text{ probability measure } on (\mathcal{T}, \mathcal{D}),$$

where $\mathcal{D}$ is a σ-algebra on $\mathcal{T}$ and e_t denotes the Dirac measure, the one-point measure, at t, i.e. $e_t(A) = 1\!\!1_A(t)$ for all $A \in \mathcal{D}$.

3.1.5 Definition (Generalized design)
A probability measure δ *on* $(\mathcal{T}, \mathcal{D})$ *is called generalized design.*

3.1.6 Remark
The information matrix of a concrete design can be expressed as

$$I_\theta(d) := X_d^\top X_d = (x(t_1), \ldots, x(t_N)) \begin{pmatrix} x(t_1)^\top \\ x(t_2)^\top \\ \vdots \\ x(t_N)^\top \end{pmatrix} = \sum_{n=1}^{N} x(t_n) x(t_n)^\top = N \int x(t) x(t)^\top\, \delta_N(dt)$$

if $\delta_N = \frac{1}{N}\sum_{n=1}^{N} e_{t_n}$.

3.1.7 Definition (Information matrix for generalized designs)
a) $I_\theta(\delta) := \int x(t)\, x(t)^\top \delta(dt)$ *is called information matrix for* θ *at* δ.

b) $I_\lambda(\delta) := \left(L\, I_\theta(\delta)^- L^\top\right)^{-1}$ *is called information matrix for* $\lambda(\theta) = L\theta$ *at* δ.

3.1.8 Remark
If $G : \mathcal{T} \to \mathbb{R}^{N\times M}$ so that

$$G(t) = \begin{pmatrix} g_{11}(t) & g_{12}(t) & \dots & g_{1M}(t) \\ g_{21}(t) & g_{22}(t) & \dots & g_{2M}(t) \\ \vdots & \vdots & & \vdots \\ g_{N1}(t) & g_{N2}(t) & \dots & g_{NM}(t) \end{pmatrix}$$

then

$$\int G(t)\,\delta(dt) = \begin{pmatrix} \int g_{11}(t)\,\delta(dt) & \int g_{12}(t)\,\delta(dt) & \dots & \int g_{1M}(t)\,\delta(dt) \\ \int g_{21}(t)\,\delta(dt) & \int g_{22}(t)\,\delta(dt) & \dots & \int g_{2M}(t)\,\delta(dt) \\ \vdots & \vdots & & \vdots \\ \int g_{N1}(t)\,\delta(dt) & \int g_{N2}(t)\,\delta(dt) & \dots & \int g_{NM}(t)\,\delta(dt) \end{pmatrix}.$$

Thereby the notation $\int g(t)\,\delta(dt)$ denotes the Lebesgue integral of a function $g : \mathbb{R} \to \mathbb{R}$ with respect to the measure δ: If δ is a discrete probability measure given by the discrete density $\delta(\{t_i\}) = p_i \geq 0$ for $i = 1, 2, \dots$, then

$$\int g(t)\,\delta(dt) = \sum_{i=1}^{\infty} g(t_i)\,p_i.$$

If δ is an absolute continuous probability measure given by the density $f : \mathbb{R}^p \to \mathbb{R}$, then

$$\int g(t)\,\delta(dt) = \int g(t)\,f(t)\,dt$$

is the Riemann integral.

To define the identifiability for generalized designs, note:

3.1.9 Lemma
$\lambda(\theta) = L\theta$ is identifiable at the concrete design d if and only if there exists $K \in \mathbb{R}^{S\times R}$ such that $L = K\,I_\theta(d)$.

Proof. According to Theorem 2.1.3, $\lambda(\theta) = L\theta$ is identifiable at d if and only if $L = K_0\,X_d$ for some $K_0 \in \mathbb{R}^{S\times N}$. Hence, if $L = K\,I_\theta(d) = K\,X_d^\top X_d = K_0\,X_d$, then $\lambda(\theta) = L\theta$ is identifiable at d. Conversely, if $\lambda(\theta) = L\theta$ is identifiable at d, then there exists $K_0 \in \mathbb{R}^{S\times N}$ with

$$L = K_0\,X_d \overset{(\text{ Lemma 8.1.2 b)})}{=} K_0\,X_d(X_d^\top X_d)^- X_d^\top X_d = K\,I_\theta(d).$$

□

3.1.10 Definition (Identifiability at generalized designs)
$\lambda(\theta) = L\theta$ is called identifiable at the generalized design δ if and only if there exists $K \in \mathbb{R}^{S\times R}$ such that $L = K\,I_\theta(\delta)$.

3.1.11 Lemma
If $\text{rk}(L) = S$ *and* $L = K\, I_\theta(\delta)$*, then* $\text{rk}(L\, I_\theta(\delta)^- L^\top) = S$ *and* $L\, I_\theta(\delta)^- L^\top$ *is independent of the choice of the g-inverse.*

Proof. At first note, that the identifiability implies

$$L\, I_\theta(\delta)^- L^\top = K\, I_\theta(\delta) I_\theta(\delta)^- I_\theta(\delta)\, K^\top = K\, I_\theta(\delta)\, K^\top,$$

so that $L\, I_\theta(\delta)^- L^\top$ does not depend on the choice of the g-inverse. Since

$$a^\top I_\theta(\delta) a = \int a^\top x(t)\, x(t)^\top a\, \delta(dt) = \int (a^\top x(t))^2\, \delta(dt) \geq 0,$$

$I_\theta(\delta)$ is positive semidefinite and symmetric so that there exists $A \in \mathbb{R}^{Q\times R}$ such that $I_\theta(\delta) = A^\top A$. Hence the assertion follows as in the proof of Lemma 3.1.2. □

The design problem for generalized designs
Find a generalized design

$$\delta \in \Delta \subset \Delta_\lambda := \{\delta;\ \lambda(\theta) \text{ is identifiable at } \delta\}$$

such that $(L\, I_\theta(\delta)^- L^\top)^{-1}$ is maximal.

If an optimal generalized design δ is found and if $x : \mathcal{T} \to \mathbb{R}^R$ is continous, $(\mathcal{T}, d_m)$ is a compact metric space with metric d_m and corresponding Borel-σ-algebra $\mathcal{D}$, then there exists a discrete probability measure (discrete design) $\overline{\delta}$ with

$$I_\theta(\delta) = I_\theta(\overline{\delta})$$

and finite support $\{\tau_1, \ldots, \tau_I\}$ with $I \leq \frac{R(R+1)}{2}$. This is a consequence of the Theorem of Caratheodory (see e.g. the book of Silvey 1980, P. 72) and the fact that the set of all probability measures with finite support is dense within all probability measures under the weak topology on the space of all probability measures on $(\mathcal{T}, \mathcal{D})$ (see e.g. the book of Billingsley 1968, P. 237).

Often it is also possible to find a concrete design d for a discrete design $\overline{\delta}$ such that

$$L\, I_\theta(\overline{\delta})^- L^\top = N\, L\, I_\theta(d)^- L^\top.$$

If this is not possible, then $\overline{\delta}$ must be approximated by an appropriate concrete design d.

The main reason for regarding the generalized designs is that the set of generalized designs is convex.

3.1.12 Lemma
If $\text{rk}(L) = S$*, then*

$$\Delta_\lambda := \{\delta;\ \lambda(\theta) \textit{ is identifiable at } \delta\}$$

is convex. In particular we have

$$\alpha\,\delta_1 + (1-\alpha)\,\delta_2 \in \Delta_\lambda$$

for all $\alpha \in (0,1)$*, if* $\delta_1 \in \Delta_\lambda$ *and* $I_\theta(\delta_2)$ *is finite.*

Proof. If $\delta_1 \in \Delta_\lambda$, then $I_\theta(\delta_1)$ is in particular finite. Since $I_\theta(\delta_2)$ is finite, $I_\theta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2) = \alpha\,I_\theta(\delta_1) + (1-\alpha)\,I_\theta(\delta_2)$ exists. Moreover, we have

$$I_\theta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2) = \alpha\,I_\theta(\delta_1) + (1-\alpha)\,I_\theta(\delta_2) \geq \alpha\,I_\theta(\delta_1)$$

since $I_\theta(\delta_2)$ is positive semidefinite. Lemma 8.1.7 implies $C(I_\theta(\delta_1)) \subset C(I_\theta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2))$ so that with $L = K\,I_\theta(\delta_1)$ also a $\overline{K}$ exists with $L = \overline{K}\,I_\theta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2)$. Hence, $\alpha\,\delta_1 + (1-\alpha)\,\delta_2 \in \Delta_\lambda$. $\square$

3.2 Optimality criteria for designs

3.2.1 Definition
Let be $\Delta \subset \Delta_\lambda$. The generalized design δ_* is called

a) U_λ-optimal in $\Delta :\Longleftrightarrow \quad I_\lambda(\delta_*)^{-1} \leq I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

b) D_λ-optimal in $\Delta :\Longleftrightarrow \quad \det\, I_\lambda(\delta_*)^{-1} \leq \det\, I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

c) A_λ-optimal in $\Delta :\Longleftrightarrow \quad tr\, I_\lambda(\delta_*)^{-1} \leq tr\, I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

d) E_λ-optimal in $\Delta :\Longleftrightarrow \quad \lambda_{max} I_\lambda(\delta_*)^{-1} \leq \lambda_{max} I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$.

Thereby det *denotes the determinat, tr the trace, and λ_{max} the maximum eigenvalue of a matrix.*

3.2.2 Remark
a) If $L = c \in \mathbb{R}^{1\times R}$, then the four optimality criteria coincides and a design δ_* which minimizes $cI_\theta(\delta)^- c^\top$ in Δ is called *c-optimal* in Δ.
b) A design δ_* with $\max_{x\in\mathcal{T}} x^\top I_\theta(\delta_*)^{-1} x \leq \max_{x\in\mathcal{T}} x^\top I_\theta(\delta)^{-1} x$ for all $\delta \in \Delta$ is called a *G-optimal design* in Δ. A modification of this criterion for concrete designs is to maximize the diagonal elements of the *hat matrix* $X_d(X_d^\top X_d)^{-1} X_d^\top$.

3.2.3 Lemma (See Lemma 8.4.3 in the Appendix)
Let A and B be symmetric $S \times S$ matrices with $A \geq B > 0$. Then it holds

a) $A^{-1} \leq B^{-1}$,

b) $\det A \geq \det B$,

c) $tr\, A \geq tr\, B$,

d) $\lambda_{max}\, A \geq \lambda_{max}\, B$.

3.2.4 Theorem
Let be $\lambda(\theta) = L\,\theta$ with $rk(L) = S$ and $\widetilde{\lambda}(\theta) = \widetilde{L}\theta$ with $\widetilde{L} = H\,L$ for $H \in \mathbb{R}^{S\times S}$.

a) If H is a regular matrix, then δ_* is D_λ-optimal in Δ if and only if δ_* is $D_{\widetilde{\lambda}}$-optimal in Δ.

b) If H is an orthogonal matrix, then δ_* is A_λ-optimal in Δ if and only if δ_* is $A_{\widetilde{\lambda}}$-optimal in Δ.

Proof.

$$\begin{aligned}
a)\quad & \det I_{\widetilde{\lambda}}(\delta)^{-1} = \det \widetilde{L}\, I_\theta(\delta)^- \widetilde{L}^\top = \det H\, L\, I_\theta(\delta)^- L^\top H^\top \\
& = (detH)^2\, detL\, I_\theta(\delta)^- L^\top = (detH)^2\, detI_\lambda(\delta)^{-1}. \\
b)\quad & \mathrm{tr}\, I_{\widetilde{\lambda}}(\delta)^{-1} = \mathrm{tr}\, \widetilde{L}\, I_\theta(\delta)^- \widetilde{L}^\top = \mathrm{tr}\, H\, L\, I_\theta(\delta)^- L^\top H^\top \\
& \overset{\text{Lemma 8.4.1}}{=} \mathrm{tr}\, L\, I_\theta(\delta)^- L^\top H^\top H \overset{H \text{orthogonal}}{=} \mathrm{tr}\, L\, I_\theta(\delta)^- L^\top = \mathrm{tr}\, I_\lambda(\delta)^{-1}.
\end{aligned}$$

□

3.2.5 Remark
Besides the invariance with respect to regular transformations of $\lambda(\theta) = L\,\theta$, the D_λ-optimal designs have the advantage that they minimize the volume of the confidence ellipsoid which

is derived from the F-test. Because of the relation between tests and confidence regions, this confidence ellipsoid is given according to Theorem 2.3.1 by

$$\widehat{B}_d(y) = \left\{ l \in \mathbb{R}^S;\ \frac{(L\widehat{\theta} - l)^\top I_\lambda(d)\,(L\widehat{\theta} - l)\,/\,\mathrm{rk}(L)}{\widehat{\sigma}^2(y)} \leq q_{1-\alpha} \right\}, \tag{3.1}$$

where $q_{1-\alpha}$ is the $1-\alpha$-quantile of the central F-distribution with $\mathrm{rk}(L)$ and $N - \mathrm{rk}(X_d)$ degrees of freedom. The volume of this confidence ellipsoid depends only via $\det I_\lambda(d)^{-1}$ on the design, since in general the volume $V^S(E)$ of an ellipsoid

$$E = \left\{ x \in \mathbb{R}^S;\ (x-\mu)^\top \Sigma^{-1}\,(x-\mu) \leq q \right\} \tag{3.2}$$

is

$$V^S(E) = (q\pi)^{S/2} \left(\Gamma\left(\frac{S}{2} + 1\right)\right)^{-1} (\det \Sigma)^{1/2},$$

where Γ is the Γ-function. (see e.g. the book of Pazman 1986, P. 79).

Moreover, D_λ-optimal designs minimizes the volume of ellipsoids where the power of the F-test given in Theorem 2.3.1 is bounded by given values. Namely, on the ellipsoid

$$E_d(\sigma^2\,k) := \left\{ L\theta \in \mathbb{R}^s;\ (L\theta - l)^\top I_\lambda(d)\,(L\theta - l) \leq \sigma^2\,k \right\}, \tag{3.3}$$

the power function is given by

$$\gamma_d(\theta) \leq 1 - F_{F\left(\mathrm{rk}(L),\, \mathrm{rk}(I-P),\, k\right)}(q_{1-\alpha})$$

(see Theorem 2.3.1).

The largest main axis of an ellipsoid E given by (3.2) is proportional to the maximum eigen value of Σ so that E_λ-optimal designs minimizes the largest main axis of the confidence region given by (3.1) and the power region (3.3). The sum of the main axes of the ellipsoid E is proportional to the sum of the eigen values of Σ which is the trace of Σ. Hence A_λ-optimal designs minimize the sum of the main axes of the confidence region given by (3.1) and the power region (3.3).

3.3 Characterizations of optimal designs

The characterizations of optimal designs based on the fact that the optimality criteria leads to convex functionals on the set of generalized designs. We consider here the following functionals:

$$\Phi_{A,\lambda}: \quad \Delta_\lambda \ni \delta \longrightarrow \Phi_{A,\lambda}(\delta) := \operatorname{tr} I_\lambda(\delta)^{-1} = \operatorname{tr} L\, I_\theta(\delta)^{-} L^\top \in \mathbb{R},$$

$$\Phi_{D,\lambda}: \quad \Delta_\lambda \ni \delta \longrightarrow \Phi_{D,\lambda}(\delta) := \ln\det I_\lambda(\delta)^{-1} = \ln\det\; L\, I_\theta(\delta)^{-} L^\top \in \mathbb{R}.$$

Note that minimizing $\Phi_{D,\lambda}(\delta)$ leads to the D_λ-optimal designs since the logarithm is a monotone increasing function. However the logarithm is necessary to provide the convexity of the functional. To prove the convexity of the functionals, we need the following lemmas.

3.3.1 Lemma (See Lemma 8.4.4 in the Appendix)
If $M_1,\ M_2 \in \mathbb{R}^{R\times R}$ are symmetric and positive semidefinite and $L \in \mathbb{R}^{S\times R}$ with $L = K_1 M_1$ and $L = K_2 M_2$, then

$$L\ \left(\alpha\, M_1 + (1-\alpha)\, M_2\right)^{-}\ L^\top \ \le\ \alpha\, L\, M_1^{-}\, L^\top + (1-\alpha)\, L\, M_2^{-}\, L^\top.$$

3.3.2 Lemma (See Lemma 8.4.5 in the Appendix)
If $A,\ B \in \mathbb{R}^{R\times R}$ are symmetric and positive definite and $\alpha \in (0,1)$, then

$$\det(\alpha\, A + (1-\alpha)\, B) \ge (\det A)^{\alpha}\ (\det B)^{1-\alpha}.$$

3.3.3 Theorem

a) $\Phi_{A,\lambda}$ is convex on Δ_λ.

b) $\Phi_{D,\lambda}$ is convex on Δ_λ.

Proof.
a) Lemma 3.3.1 provides for all $\delta_1, \delta_2 \in \Delta_\lambda$

$$L\, I_\theta(\alpha\delta_1 + (1-\alpha)\delta_2)^{-} L^\top \ \le\ \alpha\ L\, I_\theta(\delta_1)^{-} L^\top + (1-\alpha)\ L\, I_\theta(\delta_2)^{-} L^\top$$

so that with Lemma 3.2.3 we obtain

$$\begin{aligned}
\Phi_{A,\lambda}(\alpha\delta_1 + (1-\alpha)\delta_2) \ &= \ \operatorname{tr} L\, I_\theta(\alpha\delta_1 + (1-\alpha)\delta_2)^{-} L^\top \\
&\le\ \operatorname{tr}\ \left(\alpha\ L\, I_\theta(\delta_1)^{-} L^\top + (1-\alpha)\ L\, I_\theta(\delta_2)^{-} L^\top\right) \\
&=\ \alpha \operatorname{tr} L\, I_\theta(\delta_1)^{-} L^\top + (1-\alpha) \operatorname{tr} L\, I_\theta(\delta_2)^{-} L^\top \ = \ \alpha\ \Phi_{A,\lambda}(\delta_1) + (1-\alpha)\ \Phi_{A,\lambda}(\delta_2).
\end{aligned}$$

b) We show here the assertion only for $\lambda(\theta)=\theta$. Lemma 3.3.2 and the definition of $I_\theta(\delta)$ provide for all $\delta_1,\delta_2\in\Delta_\theta$

$$\begin{aligned}
\Phi_{D,\theta}(\alpha\delta_1+(1-\alpha)\delta_2) &= \ln\det\left(I_\theta(\alpha\delta_1+(1-\alpha)\delta_2)^{-1}\right)\\
&= \ln\left(\det\ I_\theta(\alpha\delta_1+(1-\alpha)\delta_2)\right)^{-1} = -\ln\left(\det\ I_\theta(\alpha\delta_1+(1-\alpha)\delta_2)\right)\\
&= -\ln\left(\det\ \alpha\ I_\theta(\delta_1)+(1-\alpha)\ I_\theta(\delta_2)\right) \le -\ln\left((\det\ I_\theta(\delta_1))^{\alpha}\ (\det\ I_\theta(\delta_2))^{1-\alpha}\right)\\
&= -\alpha\ln\left(\det\ I_\theta(\delta_1)\right) - (1-\alpha)\ \ln\left(\det\ I_\theta(\delta_2)\right)\\
&= \alpha\ln\left(\det\ I_\theta(\delta_1)^{-1}\right) + (1-\alpha)\ \ln\left(\det\ I_\theta(\delta_2)^{-1}\right)\\
&= \alpha\ \Phi_{D,\theta}(\delta_1)+(1-\alpha)\ \Phi_{D,\theta}(\delta_2).
\end{aligned}$$

The proof for general $\lambda(\theta)=L\theta$ is much more complicated and can be found in the books of Pázman (1986) and Pukelsheim (1993). □

Since $I_\theta(\delta)$ is a linear function in δ and the inverse, the trace, and the determinant are differentiable functions, the functionals $\Phi_{A,\lambda}$ and $\Phi_{D,\lambda}$ are Fréchet differentiable with respect to a metric on Δ_λ which provides the weak topology. The directional derivatives, the Gâteaux derivatives, have rather simple forms. To derive these forms, we need the following lemma.

3.3.4 Lemma (See Lemma 8.4.6 in the Appendix)

a) If $A:\mathbb{R}\ni t\to A(t)\in\mathbb{R}^{N\times M}$ and $B:\mathbb{R}\ni t\to A(t)\in\mathbb{R}^{M\times K}$ are differentiable in t_0, then

$$\frac{\partial}{\partial t}\,A(t)\;B(t)\Big|_{t=t_0}=\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)\;B(t_0)\;+\;A(t_0)\;\left(\frac{\partial}{\partial t}\,B(t)\Big|_{t=t_0}\right).$$

b) If $A:\mathbb{R}\ni t\to A(t)\in\mathbb{R}^{N\times N}$ is differentiable in t_0 and $A(t_0)$ is regular, then

$$\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0}=-A(t_0)^{-1}\;\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)\;A(t_0)^{-1}$$

and

$$\frac{\partial}{\partial t}\,\ln\det A(t)\Big|_{t=t_0}=tr\,\left(A(t_0)^{-1}\;\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)\right).$$

To define Gâteaux differentiability, let Δ_0 be the set of all probability measure on $(\mathcal{T},\mathcal{D})$, $\Delta\subset\Delta_0$, and define for $\delta_*\in\Delta$

$$\Delta(\delta_*):=\{\delta\in\Delta_0;\;\;\text{there exists } k>0 \text{ with } (1-\alpha)\,\delta_*+\alpha\,\delta\in\Delta \text{ for all } \alpha\le k\}\,.$$

3.3.5 Definition (Directional derivative and Gâteaux differentiability)
a) The directional derivative of the functional $\Phi : \Delta \longrightarrow \mathbb{R}$ at δ_* in direction of δ is defined as

$$\Phi'(\delta_*, \delta) := \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha} = \lim_{\alpha \downarrow 0} \frac{\Phi(\delta_* + \alpha\,(\delta - \delta^*)) - \Phi(\delta_*)}{\alpha}.$$

b) The functional $\Phi : \Delta \longrightarrow \mathbb{R}$ is Gâteaux differentiable at δ_* if and only if $\Phi'(\delta_*, \delta)$ exists for all $\delta \in \Delta(\delta_*)$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and

$$\Phi'(\delta_*, \delta) = \int \Phi'(\delta_*, e_t)\; \delta(dt)$$

for all $\delta \in \Delta(\delta_*)$, where e_t is the Dirac measure on t, i.e. $e_t(A) = 1\!\!1_A(t)$ for all $A \in \mathcal{D}$.

3.3.6 Theorem
If $\mathrm{rk}(L) = S$, then we have for all δ with $(1-\alpha)\,\delta_* + \alpha\,\delta \in \Delta_\lambda$ for sufficient small α the following directional derivatives

$$a) \quad \Phi'_{A,\lambda}(\delta_*, \delta) = tr\, L\, I_\theta(\delta_*)^- L^\top - \int |L\, I_\theta(\delta_*)^- x(t)|^2\; \delta(dt),$$

$$b) \quad \Phi'_{D,\lambda}(\delta_*, \delta) = S - \int x(t)^\top I_\theta(\delta_*)^- L^\top\; (L\, I_\theta(\delta_*)^- L^\top)^{-1}\; L\, I_\theta(\delta_*)^- x(t)\; \delta(dt).$$

Proof. Let be $\delta(\alpha) = (1-\alpha)\,\delta_* + \alpha\,\delta$. Since $I_\theta(\delta)$ is linear in δ we have

$$\begin{aligned}
\frac{\partial}{\partial \alpha} I_\theta(\delta(\alpha))\Big|_{\alpha=0} &= \lim_{\alpha \downarrow 0} \frac{I_\theta((1-\alpha)\,\delta_* + \alpha\,\delta) - I_\theta(\delta_*)}{\alpha} \\
&= \lim_{\alpha \downarrow 0} \frac{\int x(t)\,x(t)^\top\; ((1-\alpha)\,\delta_* + \alpha\,\delta)(dt) - \int x(t)\,x(t)^\top\; \delta_*(dt)}{\alpha} \\
&= \lim_{\alpha \downarrow 0} \frac{\alpha\,(\int x(t)\,x(t)^\top\; \delta(dt) - \int x(t)\,x(t)^\top\; \delta_*(dt))}{\alpha} \\
&= I_\theta(\delta) - I_\theta(\delta_*). \qquad (3.4)
\end{aligned}$$

At first we assume that $I_\theta(\delta_*)$ is regular.

a) Lemma 3.3.4 provides

$$\begin{aligned}
\Phi'_{A,\lambda}(\delta_*,\delta) &= \frac{\partial}{\partial\alpha}\operatorname{tr} L\, I_\theta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0}\\
&= \operatorname{tr} L \left(\frac{\partial}{\partial\alpha}\, I_\theta(\delta(\alpha))^{-1}\Big|_{\alpha=0}\right) L^\top\\
&\overset{\text{Lemma 3.3.4}}{=} -\operatorname{tr} L\, I_\theta(\delta_*)^{-1}\left(\frac{\partial}{\partial\alpha}\, I_\theta(\delta(\alpha))\Big|_{\alpha=0}\right) I_\theta(\delta_*)^{-1}\, L^\top\\
&\overset{(3.4)}{=} -\operatorname{tr} L\, I_\theta(\delta_*)^{-1}\,\left(I_\theta(\delta)-I_\theta(\delta_*)\right)\, I_\theta(\delta_*)^{-1}\, L^\top\\
&= \operatorname{tr} L\, I_\theta(\delta_*)^{-1}L^\top \;-\; \operatorname{tr} L\, I_\theta(\delta_*)^{-1}\, I_\theta(\delta)\, I_\theta(\delta_*)^{-1}\, L^\top\\
&= \operatorname{tr} L\, I_\theta(\delta_*)^{-1}L^\top \;-\; \operatorname{tr}\int L\, I_\theta(\delta_*)^{-1}\, x(t)\, x(t)^\top\, I_\theta(\delta_*)^{-1}\, L^\top\, \delta(dt)\\
&= \operatorname{tr} L\, I_\theta(\delta_*)^{-1}L^\top - \int |L\, I_\theta(\delta_*)^{-1}x(t)|^2\, \delta(dt).
\end{aligned}$$

b) Lemma 3.3.4 and Lemma 8.4.1 provide

$$\begin{aligned}
\Phi'_{D,\lambda}(\delta_*,\delta) &= \frac{\partial}{\partial\alpha}\ln\det\, L\, I_\theta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0}\\
&\overset{\text{Lemma 3.3.4}}{=} \operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\,\frac{\partial}{\partial\alpha}\, L\, I_\theta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0}\right)\\
&= \operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\left(\frac{\partial}{\partial\alpha}\, I_\theta(\delta(\alpha))^{-1}\Big|_{\alpha=0}\right) L^\top\right)\\
&\overset{\text{Lemma 3.3.4}}{=} -\operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}\left(\frac{\partial}{\partial\alpha}\, I_\theta(\delta(\alpha))\Big|_{\alpha=0}\right) I_\theta(\delta_*)^{-1}\, L^\top\right)\\
&\overset{(3.4)}{=} -\operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}\,\left(I_\theta(\delta)-I_\theta(\delta_*)\right)\, I_\theta(\delta_*)^{-1}\, L^\top\right)\\
&= \operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}L^\top\right)\\
&\quad - \operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}\, I_\theta(\delta)\, I_\theta(\delta_*)^{-1}\, L^\top\right)\\
&= \operatorname{tr} I_{S\times S} - \int \operatorname{tr}\left((L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}\, x(t)\, x(t)^\top\, I_\theta(\delta_*)^{-1}\, L^\top\right)\delta(dt)\\
&\overset{\text{Lemma 8.4.1}}{=} S - \int x(t)^\top I_\theta(\delta_*)^{-1}\, L^\top (L\, I_\theta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\theta(\delta_*)^{-1}\, x(t)\, \delta(dt).
\end{aligned}$$

The proof for singular $I_\theta(\delta_*)$ follows from the above properties with the fact that there exists always a regression function $\widetilde{x}:\mathcal{T}\to\mathbb{R}^Q$ and $\widetilde{L}\in\mathbb{R}^{S\times Q}$ so that $\widetilde{I}_\theta(\delta_*)=\int\widetilde{x}(t)\,\widetilde{x}(t)^\top\,\delta_*(dt)$ is regular and $L\, I_\theta(\delta_*)^- L^\top = \widetilde{L}\,\widetilde{I}_\theta(\delta_*)^{-1}\widetilde{L}^\top$ and $L\, I_\theta(\delta_*)^- x(t) = \widetilde{L}\,\widetilde{I}_\theta(\delta_*)^{-1}\widetilde{x}(t)$ for all $t\in\operatorname{supp}((1-\alpha)\,\delta_*+\alpha\,\delta)$, where $\operatorname{supp}(\delta)$ denotes the support of δ, i.e. the smallest set $A\in\mathcal{D}$ with $\delta(A)=1$.
□

3.3.7 Corollary

If $rk(L)=S$, then $\Phi_{A,\lambda}$ and $\Phi_{D,\lambda}$ are Gâteaux differentiable at δ_*.

Proof. Theorem 3.3.6 implies

$$\Phi'_{A,\lambda}(\delta_*, \delta) = \int \Phi'_{A,\lambda}(\delta_*, e_t)\, \delta(dt)$$

and

$$\Phi'_{D,\lambda}(\delta_*, \delta) = \int \Phi'_{D,\lambda}(\delta_*, e_t)\, \delta(dt),$$

which are the additional conditions for Gâteaux differentiability. □

3.3.8 Theorem (Theorem of Whittle)
Let be Δ a convex subset of all probability measure on $(\mathcal{T}, \mathcal{D})$, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $\Phi : \Delta \to \mathbb{R}$ convex and Gâteaux differentiable at δ_*. Then the following assertions are equivalent:

$$a) \quad \Phi(\delta_*) = \min_{\delta \in \Delta} \Phi(\delta),$$

$$b) \quad \Phi'(\delta_*, \delta) \geq 0 \quad \text{for all} \quad \delta \in \Delta(\delta_*),$$

$$c) \quad \Phi'(\delta_*, e_t) \geq 0 \quad \text{for all} \quad t \in \mathcal{T}.$$

Each of the assertion a), b), and c) implies

$$d) \quad \Phi'(\delta_*, e_t) = 0 \quad \text{for all} \quad t \in supp(\delta_*).$$

Thereby, $supp(\delta)$ denotes the support of δ, i.e. the smallest set $A \in \mathcal{D}$ with $\delta(A) = 1$.

Proof.
$a) \Longrightarrow b)$: If $\Phi(\delta_*) = \min_{\delta \in \Delta} \Phi(\delta)$, then

$$\Phi(\delta_*) \leq \Phi((1-\alpha)\,\delta_* + \alpha\,\delta)$$

for all $\delta \in \Delta(\delta_*)$ for sufficient small α. This implies

$$\Phi'(\delta_*, \delta) = \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha} \geq 0$$

for all $\delta \in \Delta(\delta_*)$.
$b) \Longrightarrow c)$: This follows at once with $\delta = e_t$ and $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$.
$c) \Longrightarrow a)$: Let be $\delta \in \Delta$ arbitrary. Since Δ and Φ are convex, it holds $(1-\alpha)\,\delta_* + \alpha\,\delta \in \Delta$ for all $\alpha \in [0, 1]$ and

$$\Phi(\delta) - \Phi(\delta_*) = \frac{(1-\alpha)\,\Phi(\delta_*) + \alpha\,\Phi(\delta) - \Phi(\delta_*)}{\alpha} \geq \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha}$$

for all $\alpha \in (0,1)$. The Gâteaux differentiability implies then

$$\Phi(\delta) - \Phi(\delta_*) \geq \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha} = \Phi'(\delta_*, \delta) = \int \Phi'(\delta_*, e_t)\,\delta(dt) \overset{c)}{\geq} 0.$$

$c) \Longrightarrow d)$: $\Phi'(\delta_*, e_t) \geq 0$ for all $t \in \mathcal{T}$ and

$$0 = \Phi'(\delta_*, \delta_*) = \int \Phi'(\delta_*, e_t)\,\delta_*(dt)$$

implies $\Phi'(\delta_*, e_t) = 0$ for all $t \in \text{supp}(\delta_*)$. □

3.3.9 Theorem (Equivalence theorem for A-optimality)
Let be $\Delta \subset \Delta_\lambda$ convex, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $rk(L) = S$. Then the following assertions are equivalent:

a) δ_* *is* A_λ *optimal in* Δ,

b) $|L\,I_\theta(\delta_*)^- x(t)|^2 \leq tr\,L\,I_\theta(\delta_*)^- L^\top$ *for all* $t \in \mathcal{T}$.

If δ_* is A_λ-optimal in Δ, then

c) $|L\,I_\theta(\delta_*)^- x(t)|^2 = tr\,L\,I_\theta(\delta_*)^- L^\top$ *for all* $t \in supp(\delta_*)$.

Proof. The assertion follows at once from Theorem 3.3.6 a) and Theorem 3.3.8. □

3.3.10 Theorem (Equivalence theorem for D-optimality)
Let be $\Delta \subset \Delta_\lambda$ convex, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $rk(L) = S$. Then the following assertions are equivalent:

a) δ_* *is* D_λ *optimal in* Δ,

b) $x(t)^\top I_\theta(\delta_*)^- L^\top\,(L\,I_\theta(\delta_*)^- L^\top)^{-1}\,L\,I_\theta(\delta_*)^- x(t) \leq S$ *for all* $t \in \mathcal{T}$.

If δ_* is D_λ-optimal in Δ, then

c) $x(t)^\top I_\theta(\delta_*)^- L^\top\,(L\,I_\theta(\delta_*)^- L^\top)^{-1}\,L\,I_\theta(\delta_*)^- x(t) = S$ *for all* $t \in supp(\delta_*)$.

Proof. The assertion follows at once from Theorem 3.3.6 b) and Theorem 3.3.8. □

3.3.11 Remark
The condition $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$ is satisfied for example for $\Delta = \Delta_\lambda$ (see Lemma 3.1.12).

3.3.12 Remark
There are algorithms based on the equivalence theorems to calculate D- and A-optimal designs numerically. Some of these algorithms are given in the R-package `AlgDesign`. In this package also a rounding function is provided since often it is not easy to find a concrete design d for an optimal generalized design δ_*.

Chapter 4

Optimal designs for models with one factor

4.1 Optimal designs for linear regression

In the linear regression model we have

$$x(t) = (1,t)^\top \quad \text{with} \quad t \in \mathcal{T} \subset \mathbb{R}$$

and

$$\theta = (\theta_0, \theta_1)^\top,$$

where θ_0 is the intercept and θ_1 the slope of the regression line.

4.1.1 Lemma
If $\mathcal{T} = [-a,a]$ for $0 < a \in \mathbb{R}$, then $\delta_* = \frac{1}{2}\,(e_{-a} + e_a)$ is A_θ-optimal and D_θ-optimal in Δ_θ.

Proof. We have $x(t) = (1,t)^\top$ so that

$$I_\theta(\delta_*) = \int x(t)\,x(t)^\top\,\delta_*(dt) = \frac{1}{2}\left(x(-a)\,x(-a)^\top + x(a)\,x(a)^\top\right)$$
$$= \frac{1}{2}\left(\begin{pmatrix} 1 \\ -a \end{pmatrix}(1\ -a) + \begin{pmatrix} 1 \\ a \end{pmatrix}(1\ a)\right) = \begin{pmatrix} 1 & 0 \\ 0 & a^2 \end{pmatrix}.$$

Since $L = I_{2\times 2}$, we obtain for all $t \in [-a,a]$

$$|L\,I_\theta(\delta_*)^{-1}x(t)|^2 = \left|\begin{pmatrix} 1 & 0 \\ 0 & a^{-2} \end{pmatrix}\begin{pmatrix} 1 \\ t \end{pmatrix}\right|^2$$
$$= 1 + a^{-4}t^2 \le 1 + a^{-2} = \operatorname{tr} I_\theta(\delta_*)^{-1}$$

and

$$x(t)^\top I_\theta(\delta_*)^{-1}L^\top\ (L\,I_\theta(\delta_*)^- L^\top)^{-1}\ L\,I_\theta(\delta_*)^- x(t) = x(t)^\top I_\theta(\delta_*)^{-1}x(t)$$
$$= (1\ t)\begin{pmatrix} 1 & 0 \\ 0 & a^{-2} \end{pmatrix}\begin{pmatrix} 1 \\ t \end{pmatrix} = 1 + a^{-2}t^2 \le 2$$

so that δ_* is A_θ-optimal and D_θ-optimal in Δ_θ according to Theorem 3.3.9 and Theorem 3.3.10, respectively. □

4.1.2 Lemma
If $\mathcal{T} = [0, a]$ for $0 < a \in \mathbb{R}$, then $\delta_D = \frac{1}{2}(e_0 + e_a)$ is D_θ-optimal in Δ_θ but not A_θ-optimal. The A_θ-optimal design in Δ_θ is $\delta_A = \frac{1}{\sqrt{1+a^2}+1}\left(\sqrt{1+a^2}\,e_0 + e_a\right)$.

Proof. Because of

$$I_\theta(\delta_D) = \int x(t)\,x(t)^\top\ \delta_D(dt) = \frac{1}{2}\left(x(0)\,x(0)^\top + x(a)\,x(a)^\top\right)$$
$$= \frac{1}{2}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}(1\ 0) + \begin{pmatrix} 1 \\ a \end{pmatrix}(1\ a)\right) = \frac{1}{2}\begin{pmatrix} 2 & a \\ a & a^2 \end{pmatrix},$$
$$I_\theta(\delta_D)^{-1} = \frac{2}{a^2}\begin{pmatrix} a^2 & -a \\ -a & 2 \end{pmatrix},$$

we have for all $t \in [0, a]$

$$x(t)^\top I_\theta(\delta_D)^{-1}L^\top\ (L\,I_\theta(\delta_D)^- L^\top)^{-1}\ L\,I_\theta(\delta_D)^- x(t) = x(t)^\top I_\theta(\delta_D)^{-1}x(t)$$
$$= (1\ t)\,\frac{2}{a^2}\begin{pmatrix} a^2 & -a \\ -a & 2 \end{pmatrix}\begin{pmatrix} 1 \\ t \end{pmatrix}$$
$$= \frac{2}{a^2}\,(1\ t)\begin{pmatrix} a^2 - a\,t \\ 2\,t - a \end{pmatrix} = \frac{2}{a^2}\,(a^2 - a\,t + 2t^2 - a\,t) = \frac{2}{a^2}\,(a^2 + 2\,t(t-a)) \le 2,$$

so that δ_D is D_θ-optimal according to Theorem 3.3.10. Moreover, we obtain

$$|L\,I_\theta(\delta_D)^{-1}x(t)|^2 = |I_\theta(\delta_D)^{-1}x(t)|^2$$
$$= \left|\frac{2}{a^2}\begin{pmatrix} a^2 & -a \\ -a & 2 \end{pmatrix}\begin{pmatrix} 1 \\ t \end{pmatrix}\right|^2 = \left|\frac{2}{a^2}\begin{pmatrix} a^2 - ta \\ 2t - a \end{pmatrix}\right|^2$$
$$= \frac{4}{a^4}\,\left((a^2 - ta)^2 + (2t-a)^2\right) \overset{t=0}{=} \frac{4}{a^4}\,\left(a^4 + a^2\right) = \frac{1}{a^2}\,\left(4a^2 + 4\right)$$
$$> \frac{1}{a^2}\,\left(2a^2 + 4\right) = \frac{2}{a^2}(a^2 + 2) = \operatorname{tr} I_\theta(\delta_D)^{-1},$$

so that condition b) of Theorem 3.3.9 is violated which means that δ_D is not A_θ-optimal. To show that δ_A is A_θ-optimal, set $\xi = \frac{1}{\sqrt{1+a^2}+1}$. Then we have

$$
\begin{aligned}
I_\theta(\delta_A) &= \int x(t)\,x(t)^\top\,\delta_A(dt) = \Big((1-\xi)\,x(0)\,x(0)^\top + \xi\,x(a)\,x(a)^\top\Big)\\
&= \left((1-\xi)\begin{pmatrix}1\\0\end{pmatrix}(1\;0) + \xi\begin{pmatrix}1\\a\end{pmatrix}(1\;a)\right) = \begin{pmatrix}1 & \xi\,a\\ \xi\,a & \xi\,a^2\end{pmatrix},\\
I_\theta(\delta_A)^{-1} &= \frac{1}{\xi\,a^2(1-\xi)}\begin{pmatrix}\xi\,a^2 & -\xi\,a\\ -\xi\,a & 1\end{pmatrix},
\end{aligned}
$$

so that

$$
\text{tr}\; I_\theta(\delta_A)^{-1} = \frac{\xi\,a^2+1}{\xi\,a^2(1-\xi)}.
$$

Now the condition b) of Theorem 3.3.9,

$$
|L\,I_\theta(\delta_A)^{-1}x(t)|^2 = |I_\theta(\delta_A)^{-1}x(t)|^2 \le \text{tr}\; I_\theta(\delta_A)^{-1}
$$

for all $t \in [0,a]$, is equivalent with

$$
\begin{aligned}
&\left|\frac{1}{\xi\,a^2(1-\xi)}\begin{pmatrix}\xi\,a^2 & -\xi\,a\\ -\xi\,a & 1\end{pmatrix}\begin{pmatrix}1\\t\end{pmatrix}\right|^2 = \left|\frac{1}{\xi\,a^2(1-\xi)}\begin{pmatrix}\xi\,a^2-\xi\,a\,t\\ t-\xi\,a\end{pmatrix}\right|^2\\
&\quad= \frac{1}{\xi^2\,a^4(1-\xi)^2}\,\Big((\xi\,a^2-\xi\,a\,t)^2+(t-\xi\,a)^2\Big) = \frac{\xi^2\,a^2(a-t)^2+(t-\xi\,a)^2}{\xi^2\,a^4(1-\xi)^2} \le \frac{\xi\,a^2+1}{\xi\,a^2(1-\xi)}\\
&\iff \xi^2\,a^2(a-t)^2+(t-\xi\,a)^2 \le \big(\xi\,a^2+1\big)\,\big(\xi\,a^2(1-\xi)\big)\\
&\iff \xi^2\,a^2(a^2-2\,a\,t+t^2)+(t^2-2\,t\,\xi\,a+\xi^2\,a^2) \le \big(\xi\,a^2+1\big)\,\big(a^2(\xi-\xi^2)\big)\\
&\iff t^2(\xi^2\,a^2+1)-2t\,\xi\,a\,(\xi\,a^2+1)+\xi^2\,a^4+\xi^2\,a^2 \le \xi^2\,a^4+\xi\,a^2-\xi^3\,a^4-\xi^2\,a^2\\
&\iff t^2(\xi^2\,a^2+1)-2t\,\xi\,a\,(\xi\,a^2+1)+2\,\xi^2\,a^2-\xi\,a^2+\xi^3\,a^4 \le 0. \qquad (4.1)
\end{aligned}
$$

Using the special form of

$$
\xi = \frac{1}{\sqrt{1+a^2}+1} = \frac{\sqrt{1+a^2}-1}{(\sqrt{1+a^2}+1)(\sqrt{1+a^2}-1)} = \frac{\sqrt{1+a^2}-1}{a^2}
$$

we obtain

$$
\begin{aligned}
2\,\xi^2\,a^2 - \xi\,a^2 + \xi^3\,a^4 &= \frac{2\,(\sqrt{1+a^2}-1)^2}{a^2} - \left(\sqrt{1+a^2}-1\right) + \frac{(\sqrt{1+a^2}-1)^2\,(\sqrt{1+a^2}-1)}{a^2}\\
&= \frac{2(1+a^2-2\sqrt{1+a^2}+1) - a^2\sqrt{1+a^2} + a^2 + (1+a^2-2\sqrt{1+a^2}+1)(\sqrt{1+a^2}-1)}{a^2}\\
&= \frac{1}{a^2}\,\Big(4 + 2\,a^2 - 4\sqrt{1+a^2} - a^2\sqrt{1+a^2} + a^2\\
&\quad + \sqrt{1+a^2} + a^2\sqrt{1+a^2} - 2 - 2\,a^2 + \sqrt{1+a^2} - 1 - a^2 + 2\sqrt{1+a^2} - 1\Big)\\
&= 0.
\end{aligned}
$$

Hence the inequality (4.1) is equivalent with

$$t^2(\xi^2\,a^2+1) - 2t\,\xi\,a\,(\xi\,a^2+1) \leq 0.$$

We see at once that equality holds if $t=0$. The second root of the quadratic function

$$f(t) = t^2(\xi^2\,a^2+1) - 2t\,\xi\,a\,(\xi\,a^2+1)$$

is given by

$$
\begin{aligned}
t &= \frac{2\xi\,a\,(\xi\,a^2+1)}{\xi^2\,a^2+1} = \frac{2\,a\,\frac{\sqrt{1+a^2}-1}{a^2}\,\sqrt{1+a^2}}{\frac{(\sqrt{1+a^2}-1)^2}{a^2}+1} = \frac{2\,a\,(\sqrt{1+a^2}-1)\,\sqrt{1+a^2}}{(\sqrt{1+a^2}-1)^2+a^2}\\
&= \frac{a\;2\,\left(1+a^2-\sqrt{1+a^2}\right)}{1+a^2-2\sqrt{1+a^2}+1+a^2} = \frac{a\;2\,\left(1+a^2-\sqrt{1+a^2}\right)}{2+2\,a^2-2\sqrt{1+a^2}} = a.
\end{aligned}
$$

This means that the quadratic function f is zero for $t=0$ and $t=a$ and smaller than zero for all $t\in[0,a]$. Hence inequality (4.1) holds for all $t\in[0,a]$, so that the criterion b) of Theorem 3.3.9 for A_θ-optimality of δ_A is satisfied. □

4.1.3 Lemma
Let be $\mathcal{T}=[0,a]$ for $0<a\in\mathbb{R}$.
a) If the interesting aspect is the slope of the regression line, i.e. $\lambda(\theta)=\theta_1$, then $\delta_*=\frac{1}{2}\,(e_0+e_a)$ is U_λ-optimal and thus A_λ-optimal and D_λ-optimal in Δ_λ.
b) If the interesting aspect is the intercept of the regression line, i.e. $\lambda(\theta)=\theta_0$, then $\delta_*=e_0$ is U_λ-optimal and thus A_λ-optimal and D_λ-optimal in Δ_λ.

Proof. Is an exercise.

4.2 Optimal designs for the one-way layout

In the one-way layout, we have only one factor with A levels so that

$$x(t) = (\mathbb{1}_{\{1\}}(t), \ldots, \mathbb{1}_{\{A\}}(t))^\top \in \mathbb{R}^A \quad \text{for} \quad t \in \mathcal{T} = \{1, \ldots, A\},$$
$$\theta = (\mu_1, \ldots, \mu_A)^\top \in \mathbb{R}^A.$$

If the first level is a control level (the standard crop, the placebo), then the interesting aspect of θ is

$$\lambda(\theta) = \begin{pmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = L\,\theta \quad \text{with} \quad L = (-1_{A-1} \mid I_{A-1\times A-1}) \in \mathbb{R}^{A-1\times A}. \tag{4.2}$$

4.2.1 Lemma
If $\mathcal{T} = \{1, \ldots, A\}$ and the interesting aspect is given by (4.2), then $\delta_D = \frac{1}{A}\sum_{a=1}^A e_a$ is D_λ-optimal in Δ_λ and $\delta_A = \frac{1}{\sqrt{A-1}+A-1}\left(\sqrt{A-1}\, e_1 + \sum_{a=2}^A e_a\right)$ is A_λ-optimal in Δ_λ.

Proof. Let u_a be the a'th unit vector in $\mathbb{R}^A$. For proving the D_λ-optimality of δ_D, note

$$I_\theta(\delta_D) = \sum_{a=1}^A u_a u_a^\top \frac{1}{A} = \frac{1}{A} I_{A\times A},$$

$$L\,I_\theta(\delta_D)^{-1} L^\top = A\,L\,L^\top = A\ (-1_{A-1} \mid I_{A-1\times A-1}) \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix},$$
$$= A\ (1_{A-1\times A-1} + I_{A-1\times A-1}),$$

$$\left(L\,I_\theta(\delta_D)^{-1} L^\top\right)^{-1} \overset{\text{Lemma 8.3.3}}{=} \frac{1}{A}\left(I_{A-1\times A-1} - \frac{1}{A} 1_{A-1\times A-1}\right)$$

$$I_\theta(\delta_D)^{-1} L^\top \left(L\,I_\theta(\delta_D)^{-1} L^\top\right)^{-1} L\,I_\theta(\delta_D)^{-1}$$
$$= \frac{A^2}{A} \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix} \left(I_{A-1\times A-1} - \frac{1}{A} 1_{A-1\times A-1}\right) (-1_{A-1} \mid I_{A-1\times A-1})$$
$$= A \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix} \left(-1_{A-1} + \frac{A-1}{A}\, 1_{A-1} \,\middle|\, I_{A-1\times A-1} - \frac{1}{A} 1_{A-1\times A-1}\right)$$
$$= A \begin{pmatrix} A - 1 - \frac{(A-1)^2}{A} & -1_{A-1}^\top + \frac{A-1}{A}\, 1_{A-1}^\top \\ -1_{A-1} + \frac{A-1}{A}\, 1_{A-1} & I_{A-1\times A-1} - \frac{1}{A} 1_{A-1\times A-1} \end{pmatrix}$$
$$= \begin{pmatrix} A-1 & -1_{A-1}^\top \\ -1_{A-1} & A\,I_{A-1\times A-1} - 1_{A-1\times A-1} \end{pmatrix}.$$

Hence for all $t \in \{1, \dots, A\}$, we have

$$x(t)^\top\, I_\theta(\delta_D)^{-1} L^\top \left(L\, I_\theta(\delta_D)^{-1} L^\top\right)^{-1} L\, I_\theta(\delta_D)^{-1}\, x(t) = A - 1,$$

so that δ_D is D_λ-optimal according to Theorem 3.3.10. The proof of the A_λ-optimality of δ_A bases on the following calculations:

$$\begin{aligned}
I_\theta(\delta_A) &= \frac{1}{\sqrt{A-1} + A - 1} \begin{pmatrix} \sqrt{A-1} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix}, \\
I_\theta(\delta_A)^{-1} &= \left(\sqrt{A-1} + A - 1\right) \begin{pmatrix} \frac{1}{\sqrt{A-1}} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix}, \\
L\, I_\theta(\delta_A)^{-1} &= \left(\sqrt{A-1} + A - 1\right) \left(-1_{A-1} \mid I_{A-1\times A-1}\right) \begin{pmatrix} \frac{1}{\sqrt{A-1}} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix} \\
&= \left(\sqrt{A-1} + A - 1\right) \left(-\frac{1}{\sqrt{A-1}}\, 1_{A-1} \,\middle|\, I_{A-1\times A-1}\right), \\
L\, I_\theta(\delta_A)^{-1} L^\top &= \left(\sqrt{A-1} + A - 1\right) \left(-\frac{1}{\sqrt{A-1}}\, 1_{A-1} \,\middle|\, I_{A-1\times A-1}\right) \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix} \\
&= \left(\sqrt{A-1} + A - 1\right) \left(\frac{1}{\sqrt{A-1}} 1_{A-1\times A-1} + I_{A-1\times A-1}\right), \\
\operatorname{tr} L\, I_\theta(\delta_A)^{-1} L^\top &= \left(\sqrt{A-1} + A - 1\right) \left(\frac{1}{\sqrt{A-1}}(A-1) + (A-1)\right) = \left(\sqrt{A-1} + A - 1\right)^2, \\
\left|L\, I_\theta(\delta_A)^{-1} x(1)\right|^2 &= \left|\left(\sqrt{A-1} + A - 1\right) \left(-\frac{1}{\sqrt{A-1}}\, 1_{A-1} \mid I_{A-1\times A-1}\right) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right|^2 \\
&= \left(\sqrt{A-1} + A - 1\right)^2 \frac{1}{A-1}\,(A-1), \\
\left|L\, I_\theta(\delta_A)^{-1} x(t)\right|^2 &= \left(\sqrt{A-1} + A - 1\right)^2 \quad \text{for} \quad t \in \{2, \dots, A\}.
\end{aligned}$$

Hence for all $t \in \{1, \dots, A\}$ it holds

$$\left|L\, I_\theta(\delta_A)^{-1} x(t)\right|^2 = \operatorname{tr} L\, I_\theta(\delta_A)^{-1} L^\top$$

so that Theorem 3.3.9 provides the A_λ-optimality of δ_A. □

Chapter 5

Optimal designs for models with two factors

5.1 Two quantitative factors

5.1.1 Lemma
In the multiple regression model

$$Y_n = \theta_0 + \theta_1\,\tau_{1n} + \theta_2\,\tau_{2n} + Z_n$$

with design region $\mathcal{T} = [-1,1] \times [-1,1]$, *the design*

$$\delta_* = \frac{1}{4}(e_{(-1,-1)} + e_{(-1,1)} + e_{(1,-1)} + e_{(1,1)})$$

is A_θ*- and* D_θ*-optimal in* Δ.

Proof. This was an exercise. □

5.2 Two-way layout

Here it is assumed that the data set contains three variables: one numeric variable concerning measurements and two factor variables A and B concerning treatments or blocks (groups). The factor variable A has A levels and the factor variable B has B levels. Let $N_{ab}(d)$ denote the sample size for factor level combination (a,b) in the concrete design d. Because of $A\,B$ level combinations, we have the following **incidence table** for the sample sizes:

		B			
		1	2	...	B
A	1	$N_{11}(d)$	$N_{12}(d)$	...	$N_{1B}(d)$
	2	$N_{21}(d)$	$N_{22}(d)$	...	$N_{2B}(d)$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	A	$N_{A1}(d)$	$N_{A2}(d)$	...	$N_{AB}(d)$

Table 5.1: Numbers of repetitions for each level combination

In **completely balanced designs** the sample sizes are all equal, i.e. $N_{ab}(d) = M$ for $a = 1, \ldots, A$ and $b = 1, \ldots, B$. But **unbalanced designs** with different sample sizes are also considered. In particular, some of the sample sizes $N_{ab}(d)$ can be zero.

The allocation of the level combinations in a completely balanced design should be done randomly to the $N = M\,A\,B$ experimental units. Such designs are called **randomized designs for two factors** and can be created with the function `design.ab` of the library `agricolae`.

5.2.1 Example

If factor A has the 3 levels `1, 2, 3` and factor B has 4 the levels `1, 2, 3, 4` and $M = 2$, then we obtain for example the following allocation:

```
> library(agricolae)
> design.ab(c(3,4),2)$book
   plots block A B
1    101     1 1 1
2    102     1 2 4
3    103     1 2 1
4    104     1 2 3
5    105     1 2 2
6    106     1 1 3
7    107     1 3 4
8    108     1 3 2
9    109     1 3 1
10   110     1 1 2
11   111     1 1 4
12   112     1 3 3
13   113     2 2 2
14   114     2 1 3
15   115     2 3 3
16   116     2 3 4
17   117     2 2 3
18   118     2 2 4
19   119     2 2 1
20   120     2 3 2
21   121     2 1 4
22   122     2 1 2
23   123     2 3 1
```

```
24    124       2 1 1
```

We see that the second repetitions are given in a second block. This makes sense since if the experiment must be stopped before all measurements are done, then at least all level combinations were used at least one time.

5.3 Block designs

Usually one of the two factors is a nuisance factor or block factor so that the interest lies mainly in estimating and testing hypotheses on only one factor. Let be B the block factor and A the treatment factor which is of main interest. Usually blocks have the same size K so that $N_{\bullet b} := \sum_{a=1}^{A} N_{ab}(d) = K$ for $b = 1, \dots, B$. However, $N_{a\bullet} := \sum_{b=1}^{B} N_{ab}(d)$ may vary for $a = 1, \dots, A$. Another important quantity is $\lambda(a_1, a_2) := \sharp\{b;\ a_1, a_2 \text{ in block } b\} = \sharp\{b;\ N_{a_1 b} \geq 1, N_{a_2 b} \geq 1\}$. If $N_{ab}(d) \in \{0, 1\}$ for all $a = 1, \dots, A$ and $b = 1, \dots, B$, then $\lambda(a_1, a_2)$ can be also expressed as $\lambda(a_1, a_2) = \sum_{b=1}^{B} N_{a_1 b} N_{a_2 b}$.

5.3.1 Definition (Block designs)

a) A block design d for A treatments and B blocks of size K is given by

$$b_d : \{1, \dots, B\} \times \{1, \dots, K\} \longrightarrow \{1, \dots, A\}$$

or by the **incidence matrix** $(N_{ab}(d))_{a=1,\dots,A, b=1,\dots,B}$ *with* $N_{\bullet b} = K$ *for* $b = 1, \dots, B$ *where* $N_{ab}(d)$ *denotes the number of level combinations a and b.*

b) The set of all block designs for A treatments and B blocks of size K is denoted by $\boldsymbol{\Omega}_{\mathbf{A},\mathbf{B},\mathbf{K}}$.

c) A block design d is **complete** *if $N_{ab}(d) \geq 1$ for all $a \in \{1, \dots, A\}$ and all $b \in \{1, \dots, B\}$.*

d) A block design d is **incomplete** *if there exists level combinations (a, b) with $N_{ab}(d) = 0$.*

e) A block design d is **balanced** *if each treatment level a appears in the same number of blocks ($N_{1\bullet} = N_{2\bullet} = \ldots = N_{A\bullet}$) and each pair of different treatments $a_1, a_2 \in \{1, \dots, A\}$ appears in the same number of blocks and this number is not zero, i.e. $\lambda(a_1, a_2) = \lambda \neq 0$ for all $a_1, a_2 \in \{1, \dots, A\}$ with $a_1 \neq a_2$.*

f) A block design d is called **completely balanced block design** *if $N_{ab}(d) = M > 0$ for all $a \in \{1, \dots, A\}$ and all $b \in \{1, \dots, B\}$.*

g) An incomplete block design which is balanced is called **balanced incomplete block design (BIBD)**.

5.3.2 Remark
The sample size of a block design is

$$N = \sum_{b=1}^{B} N_{\bullet b} = B\,K.$$

Completely balanced block designs

Although with the block factor a temporal or spacial influence is taking into account, there may be also unknown temporal or spacial influence. Therefore the treatments should be allocated in a block randomly. Such designs are called **randomized complete block designs (RCBD)** and can be constructed with `design.rcbd` of the `agricolae` package.

5.3.3 Example

If there are 3 levels `T1, T2, T3` for the treatment and 4 levels for the block factor, we can use for example the following randomized complete block design:

```
> library(agricolae)
> design.rcbd(c("T1","T2","T3"),4)$book
   plots block c("T1", "T2", "T3")
1    101     1                  T1
2    102     1                  T3
3    103     1                  T2
4    201     2                  T3
5    202     2                  T2
6    203     2                  T1
7    301     3                  T2
8    302     3                  T3
9    303     3                  T1
10   401     4                  T1
11   402     4                  T2
12   403     4                  T3
```

Balanced incomplete block designs (BIBD)

Balanced incomplete block designs are needed when the block size is too small so that not all treatments can be applied in the block. The block size K can be even 2 in the extreme case. This is for example the case when the experimental units are the eyes of persons and $A > 2$ eye drops should be studied. Then each person provides a block of block size is 2 and the number of treatments levels is higher than the block size.

If $K < A$ then only $N_{ab}(d) \in \{0, 1\}$ for all $a = 1, \ldots, A$ and $b = 1, \ldots, B$ makes sense. The question is, for which block sizes and for which numbers of treatment levels a balanced incomplete block design exists for such situations. Here some necessary conditions for the existence are given. For balanced designs, let be

$$R \quad = N_{1\bullet} = \ldots = N_{A\bullet} \text{ the total number of repetitions of the treatment levels,} \tag{5.1}$$

$$K \quad = N_{\bullet 1} = \ldots = N_{\bullet B} \text{ the block sizes,}$$

$$\lambda \quad \text{the number of blocks in which a pair of different treatments } a_1,\ a_1 \in \{1, \ldots, A\} \text{ appears, i.e. } \lambda = \lambda(a_1, a_2) \text{ for all } a_1, a_2 \in \{1, \ldots, A\} \text{ with } a_1 \neq a_2. \tag{5.2}$$

Obviously, it holds:

$$A\ R = B\ K \tag{5.3}$$

Moreover, there are

$$\begin{pmatrix} A \\ 2 \end{pmatrix} = \frac{A\,(A-1)}{2} \qquad \text{different pairs of treatment levels,}$$
$$\begin{pmatrix} K \\ 2 \end{pmatrix} = \frac{K\,(K-1)}{2} \qquad \text{different pairs of treatments in each block,}$$

so that

$$\lambda\, A\,(A-1) = B\, K\,(K-1). \tag{5.4}$$

Substituting $B\ K$ by $A\ R$ and dividing by A, we obtain

$$\lambda\,(A-1) = R\,(K-1). \tag{5.5}$$

Conditions (5.3) and (5.5) are only necessary conditions for a balanced incomplete block designs but not sufficient conditions.

In the function `design.bib` of the `agricolae` package one can only specify the number of treatments and the block size.

5.3.4 Example (Balanced incomplete block design)
The balanced incomplete block design for $A = 3$ and $K = 2$ has the form:

```
> design.bib(c("T1","T2","T3"),k=2)$book
[1] "No improvement over initial random design."

Parameters BIB
==============
Lambda     : 1
treatmeans : 3
Block size : 2
Blocks     : 3
Replication: 2

Efficiency factor 0.75

<<< Book >>>
  plots block c("T1", "T2", "T3")
1   101     1                  T1
2   102     1                  T3
3   201     2                  T3
4   202     2                  T2
5   301     3                  T2
6   302     3                  T1
```

This means that we have $R = 2$, $B = 3$ and $\lambda = 1$. A BIB design with $A = 16$ and $K = 6$ is so large that we give here only the automatic printout:

```
> ddd<-design.bib(as.factor(1:16),k=6)

Parameters BIB
==============
Lambda     : 1
treatmeans : 16
Block size : 6
Blocks     : 8
Replication: 3

Efficiency factor 0.8888889

<<< Book >>>
```

5.4 Contrasts in block designs

If $N_{ab}(d) \geq 2$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$, then the following model for the distribution of the measurements variables can be used:

$$Y_n = \mu_{ab} + Z_n = \mu + \alpha_a + \beta_b + \gamma_{ab} + Z_n \ \text{ with } \ Z_n \sim \mathcal{N}(0, \sigma^2) \ \text{ if } \ t_n = (a, b)^\top,$$

where

1. Z_n is the measurement error,
2. μ the **average mean**,
3. α_a the **main effect** of level a of factor A,
4. β_b the **main effect** of level b of factor B,
5. γ_{ab} the **interaction** between the levels a and b of the factors A and B.

If the parameters μ, $\alpha_1, \ldots, \alpha_A$, $\beta_1, \ldots, \beta_B$, $\gamma_{11}, \ldots, \gamma_{1B}, \ldots, \gamma_{AB}$ can attain arbitrary values, then the model is overparametrized since there are $1 + A + B + A\,B$ different parameters while there are only $A\,B$ level combinations. Hence side conditions for the parameters are needed, which are:

$$\sum_{a=1}^{A} \alpha_a = 0, \ \sum_{b=1}^{B} \beta_b = 0, \ \sum_{a=1}^{A} \gamma_{ab} = 0 \text{ for all } b = 1, \ldots, B, \ \sum_{b=1}^{B} \gamma_{ab} = 0 \text{ for all } a = 1, \ldots, A.$$

The block size K is often smaller as the number A of treatments. Therefore, from here we will consider only models without interactions, i.e.

$$\begin{aligned} Y_n &= x(t_n)^\top \theta + Z_n = \mu + \alpha_a + \beta_b + Z_n, \ \text{ if } t_n = (a, b), \ \text{ or} \\ Y_{abn} &= \mu + \alpha_a + \beta_b + Z_{abn}. \end{aligned}$$

Then we can write

$$Y = V_d \alpha + 1_N \mu + \tilde{W}_d \beta + Z = V_d \alpha + W_d \tilde{\beta} + Z,$$

with

$$\begin{aligned} \alpha &= (\alpha_1, \ldots, \alpha_A)^\top \\ \beta &= (\beta_1, \ldots, \beta_A)^\top \\ \tilde{\beta} &= (\mu, \beta_1, \ldots, \beta_A)^\top. \end{aligned}$$

The design matrix of a block design is given up to permutations of the rows as

$$X_d = (V_d, W_d) \text{ with } V_d \in \mathbb{R}^{N \times A} \text{ and } W_d = (1_N, I_{B \times B} \otimes 1_K)\,.$$

If the design is completely balanced, then additionally we have

$$V_d = 1_B \otimes I_{A\times A} \otimes 1_M$$

as well as $N = A\,B\,M$ and $K = A\,M$.

Since the block factor is a nuisance parameter, the main interest lies in estimating and testing aspects of the effect α of the treatment factor A.

5.4.1 Lemma
a) Every block design d satisfies $V_d 1_A = 1_N$, $W_d 1_{B+1} = 2\,1_N$, $1_N^\top W_d = (N, K\,1_B^\top)$.
b) Every balanced block design d with $N_{a\bullet} = R$ for all $a \in \{1,\dots,A\}$ satisfies additionally $1_N^\top V_d = R\,1_A^\top$.

5.4.2 Definition
a) A linear aspect $\lambda(\theta) = L_\alpha \alpha$ with $L_\alpha \in \mathbb{R}^{S\times A}$ is called contrast if

$$L_\alpha 1_A = 0_S.$$

b) A contrast $\lambda(\theta) = L_\alpha \alpha$ is called a contrast between $a(1)$ and $a(2)$ if

$$L_\alpha = u_{a(1)}^\top - u_{a(2)}^\top \text{ or } \lambda(\theta) = \alpha_{a(1)} - \alpha_{a(2)}, \text{ respectively,}$$

where $u_a \in \mathbb{R}^A$ denotes the a'th unit vector.

5.4.3 Example
A contrast used already for the one-way layout is given by

$$L_\alpha = \left(-1_{A-1} | I_{(A-1)\times(A-1)}\right).$$

5.4.4 Definition (Comparable treatments in a block design)

a) Two treatments $a(1), a(2) \in \{1,\dots,A\}$ in a block design d are **directly comparable**, *if they appear in a block $b(0) \in \{1,\dots,B\}$, i.e. $N_{a(1)b(0)} \geq 1$ and $N_{a(2)b(0)} \geq 1$.*

b) Two treatments $a(1), a(2) \in \{1,\dots,A\}$ in a block design d are **comparable**, *if they are directly comparable or there exists $M \geq 1$ and $a_1,\dots,a_M \in \{1,\dots,A\}$ such that*

$a(1)$ and a_1 are directly comparable,
a_1 and a_2 are directly comparable,
$\vdots$
a_{M-1} and a_M are directly comparable,
a_M and $a(2)$ are directly comparable.

c) A block design d is **connected** *if all pairs $a(1), a(2) \in \{1,\dots,A\}$ are comparable.*

5.4.5 Remark
a) If two treatments $a(1), a(2) \in \{1,\dots,A\}$ in a block design d are directly comparable then there

exists $b_0 \in \{1,\dots,B\}$, $k(1), k(2) \in \{1\dots,K\}$ with $b_d(b_0,k(1)) = a(1)$ and $b_d(b_0,k(2)) = a(2)$.
b) In a complete block design, all $a(1), a(2) \in \{1,\dots,A\}$ are directly comparable.
c) In a balanced block design, all $a(1), a(2) \in \{1,\dots,A\}$ are directly comparable.

5.4.6 Theorem
a) A contrast $L_\alpha \alpha = \alpha_{a(1)} - \alpha_{a(2)}$ between $a(1)$ and $a(2)$ is identifiable at a block design d if and only if $a(1)$ and $a(2)$ are comparable in the design d.
b) All contrasts are identifiable at a block design d if and only if d is a connected block design.

Proof.
a) At first assume that $a(1)$ and $a(2)$ are directly comparable. Then there exists a block $b(0)$ so that $a(1)$ and $a(2)$ are in the block $b(0)$. Without loss of generality let $a(1) = 1$, $a(2) = 2$, $b(0) = 1$, $L_\alpha = u_1^\top - u_2^\top$ and $d = ((1,1)^\top, (2,1)^\top, (a_3,b_3)^\top, \dots, (a_N,b_N)^\top)$. Then a subdesign of d is $d(1) = ((1,1)^\top, (2,1)^\top)$ with corresponding design matrix

$$X_{d(1)} = \left(I_{2\times 2}, 0_{2\times(A-2)}, 1_2, 1_2, 0_{2\times(B-1)}\right).$$

Then it holds

$$(1,-1)X_{d(1)} = \left(1,-1,0_{1\times(A-2)},0,0,0_{1\times(B-1)}\right) = \left(u_1^\top - u_2^\top, 0_{1\times(B+1)}\right) = (L_\alpha, 0_{1\times(B+1)})$$

so that $L\theta = (L_\alpha, 0_{1\times(B+1)})\begin{pmatrix}\alpha\\ \tilde\beta\end{pmatrix}$ is identifiable at $d(1)$. With Lemma 2.1.4 it is also identifiable at d.

Now assume that $a(1)$ and $a(2)$ are comparable but not directly comparable. Then there exists $M \geq 1$ and $a_1,\dots,a_M \in \{1,\dots,A\}$ such that

$a(1)$ and a_1 are directly comparable,
a_1 and a_2 are directly comparable,
$\vdots$
a_{M-1} and a_M are directly comparable,
a_M and $a(2)$ are directly comparable.

In particular, we have

$$\begin{aligned} L_\alpha(0) := u_{a(1)}^\top - u_{a(2)}^\top &= u_{a(1)}^\top - u_{a_1}^\top + u_{a_1}^\top - u_{a_2}^\top \dots + u_{a_{M-1}}^\top - u_{a_M}^\top + u_{a_M}^\top - u_{a(2)}^\top \\ &= L_\alpha(1) + \dots + L_\alpha(M+1), \end{aligned}$$

where the aspects $L_m\theta = L_\alpha(m)\alpha$ are identifiable at d for $m = 1,\dots,M+1$. This means that there exists K_m with $L_m = K_m X_d$ for $m = 1,\dots,M+1$. This implies

$$L_0 := (L_\alpha(0), 0_{1\times(B+1)}) = (u_{a(1)}^\top - u_{a(2)}^\top, 0_{1\times(B+1)}) = \sum_{m=1}^{M+1} L_m = \sum_{m=1}^{M+1} K_m X_d$$

so that the contrast L_0 between $a(1)$ and $a(2)$ is identifiable at d.

For the reverse implication, assume that $\lambda(\theta) = \alpha_{a(1)} - \alpha_{a(2)}$ is identifiable at d but $a(1)$ and $a(2)$ are not comparable. Since $a(1)$ and $a(2)$ are not comparable, there exists blocks $b(1), \ldots, b(J)$ and $b(J+1), \ldots, b(B)$ so that all treatments comparable with $a(1)$ are contained in the blocks $b(1), \ldots, b(J)$ and do not appear in the blocks $b(J+1), \ldots, b(B)$ while all other treatments including treatment $a(2)$ are not included in the blocks $b(1), \ldots, b(J)$. Without loss of generality, let be $b(j) = j$ for $j = 1, \ldots, B$. Then the design matrix has the following form

$$X_d = \begin{pmatrix} V_{11} & 0_{JK \times (A-I)} & 1_{JK} & W_{11} & 0_{JK \times (B-J)} \\ 0_{(N-JK) \times I} & V_{22} & 1_{N-JK} & 0_{(N-JK) \times J} & W_{22} \end{pmatrix},$$

where I is the number of treatments which are comparable with $a(1)$ including $a(1)$. Write $u_{a(1)}^\top - u_{a(2)}^\top = (L_1, -L_2)$, where L_1 is a unit vector in $\mathbb{R}^I$ and L_2 is a unit vector in $\mathbb{R}^{A-I}$. Since $a(1) - a(2)$ is identifiable at d, there exists $K_1 \in \mathbb{R}^{1 \times JK}$ and $K_2 \in \mathbb{R}^{1 \times (N-JK)}$ such that

$$\begin{aligned} L = \left(L_1, -L_2, 0, 0_{1 \times J}, 0_{1 \times (B-J)}\right) &= (K_1, K_2) \\ &\cdot \begin{pmatrix} V_{11} & 0_{JK \times (A-I)} & 1_{JK} & W_{11} & 0_{JK \times (B-J)} \\ 0_{(N-JK) \times I} & V_{22} & 1_{N-JK} & 0_{(N-JK) \times J} & W_{22} \end{pmatrix}. \end{aligned}$$

This means in particular $L_1 = K_1 V_{11}$ and $0_{1 \times J} = K_1 W_{11}$. Multiplying with 1_I and 1_J, respectively, leads to the contradiction

$$1 = L_1 \, 1_I = K_1 V_{11} \, 1_I = K_1 \, 1_{JK} = K_1 W_{11} \, 1_J = 0_{1 \times J} \, 1_J = 0.$$

Hence $a(1)$ and $a(2)$ are comparable.

b) This is a direct consequence of a) since any one-dimensional contrast given by $L_\alpha \in \mathbb{R}^{1 \times A}$ can be expressed by a sum of contrasts between two treatments. $\square$

5.5 Optimality of completely balanced block designs

Consider again the block model without interactions

$$Y = X_d\theta + Z = V_d\alpha + 1_N\mu + \tilde{W}_d\beta + Z = V_d\alpha + W_d\tilde{\beta} + Z,$$

with

$$X_d = (V_d, W_d) \text{ with } V_d \in \mathbb{R}^{N\times A} \text{ and } W_d = (1_N|I_{B\times B}\otimes 1_K) = (1_B\otimes 1_K|I_{B\times B}\otimes 1_K)\,.$$

It is shown here that a completely balanced block design, i.e. a design with

$$V_d = 1_B \otimes I_{A\times A} \otimes 1_M,$$

is D-optimal for the contrast given by $\lambda(\theta) = L\theta = L_\alpha\alpha$ with

$$L_\alpha = \left(-1_{A-1}|I_{(A-1)\times(A-1)}\right).$$

Recall that D-optimal designs are invariant with respect to regular transformations of the linear aspect. Hence if the D-optimality for the above contrast is shown then it holds also for other contrast matrices with rank $A-1$.

The completely balanced block design d_* corresponds to a generalized design δ_* which is the uniform distribution on $\mathcal{T} = \{1,\dots,A\}\times\{1,\dots,B\}$, i.e.

$$\delta_*(a,b) = \frac{1}{A\,B} \text{ for all } (a,b) \in \{1,\dots,A\}\times\{1,\dots,B\}.$$

If $I_\theta(d_*)$ and $I_\lambda(d_*) = (L\,I_\theta(d_*)^-L^\top)^{-1}$ are the information matrices for the concrete design $d_* \in \Omega_{A,B,K}$ with $M = \frac{K}{A} \in \mathbb{N}$ and $I_\theta(\delta_*)$ and $I_\lambda(\delta_*) = (L\,I_\theta(\delta_*)^-L^\top)^{-1}$ are the information matrices for the generalized design δ_*, then the following relations hold

$$I_\theta(\delta_*) = \frac{1}{M\,A\,B}\,I_\theta(d_*) = \frac{1}{K\,B}\,I_\theta(d_*) \text{ and } I_\lambda(\delta_*) = \frac{1}{M\,A\,B}\,I_\lambda(d_*) = \frac{1}{K\,B}\,I_\lambda(d_*).$$

Hence the quantities $I_\lambda(d_*)$ and $L\,I_\theta(d_*)^-$ must be calculated for checking the conditions of the equivalence theorem for D-optimality (Theorem 3.3.10). According to Theorem 2.4.3 and Theorem 2.4.5, we have

$$I_\lambda(d_*)^{-1} = L_\alpha\,(V_{d_*}^\top\,\omega^\perp(W_{d_*})\,V_{d_*})^-\,L_\alpha^\top.$$

5.5.1 Lemma
a) Any block design $d \in \Omega_{A,B,K}$ satisfies

$$V_d^\top\,\omega^\perp(W_d)\,V_d = diag(N_{1\cdot}(d),\dots,N_{A\cdot}(d))$$
$$-\;\frac{1}{K}\sum_{b=1}^{B}\begin{pmatrix} N_{1b}(d)^2 & N_{1b}(d)N_{2b}(d) & \dots & N_{1b}(d)N_{Ab}(d)\\ N_{2b}(d)N_{1b}(d) & N_{2b}(d)^2 & \dots & N_{2b}(d)N_{Ab}(d)\\ \vdots & \vdots & & \vdots\\ N_{Ab}(d)N_{1b}(d) & N_{Ab}(d)N_{2b}(d) & \dots & N_{Ab}(d)^2\end{pmatrix}.$$

b) If $d_* \in \Omega_{A,B,K}$ is a completely balanced design, then

$$V_{d_*}^\top \omega^\perp(W_{d_*})\, V_{d_*} = \frac{K\,B}{A}\left(I_{A\times A} - \frac{1}{A} 1_A\, 1_A^\top\right),$$

$$I_\lambda(d_*) = \frac{K\,B}{A}\left(I_{(A-1)\times(A-1)} - \frac{1}{A} 1_{A-1} 1_{A-1}^\top\right),$$

$$L\, I_\theta(d_*)^- = \frac{A}{K\,B}\,\left(-1_{A-1}|I_{(A-1)\times(A-1)}|0_{A-1}|0_{(A-1)\times B}\right)$$

for an appropriate g-inverse $I_\theta(d_*)^-$.

Proof.
a) At first note

$$V_d^\top \omega^\perp(W_d)\, V_d = V_d^\top\left(I_{N\times N} - W_d\left(W_d^\top W_d\right)^- W_d^\top\right) V_d = V_d^\top V_d - V_d^\top W_d\left(W_d^\top W_d\right)^- W_d^\top V_d.$$

Write V_d blockwise, i.e.

$$V_d = \begin{pmatrix} V_{d1} \\ V_{d2} \\ \vdots \\ V_{dB} \end{pmatrix},$$

where $V_{db} \in \mathbb{R}^{K\times A}$ provides the design matrix for the treatments in block b. Then we have

$$V_d^\top V_d = \sum_{b=1}^B V_{db}^\top V_{db} = \sum_{b=1}^B \operatorname{diag}(N_{1b}(d),\ldots,N_{Ab}(d)) = \operatorname{diag}(N_{1\bullet}(d),\ldots,N_{A\bullet}(d)).$$

Moreover,

$$\begin{aligned} W_d^\top W_d &= \begin{pmatrix} 1_B^\top \otimes 1_K^\top \\ I_{B\times B}\otimes 1_K^\top \end{pmatrix} (1_B\otimes 1_K | I_{B\times B}\otimes 1_K) \\ &= \begin{pmatrix} B\,K & 1_B^\top K \\ 1_B\,K & I_{B\times B}\,K \end{pmatrix} = K\begin{pmatrix} B & 1_B^\top \\ 1_B & I_{B\times B} \end{pmatrix}. \end{aligned}$$

Lemma 8.3.2 provides

$$\left(W_d^\top W_d\right)^- = \frac{1}{K}\begin{pmatrix} 0 & 0_B^\top \\ 0_B & I_{B\times B}\end{pmatrix}. \tag{5.6}$$

Then

$$V_d^\top W_d = (V_{d1}^\top|\ldots|V_{dB}^\top)\ (1_B\otimes 1_K|I_{B\times B}\otimes 1_K) = \left(\sum_{b=1}^B V_{db}^\top 1_K \middle| V_{d1}^\top 1_K \middle| \ldots \middle| V_{dB}^\top 1_K\right) \tag{5.7}$$

yields

$$V_d^\top W_d \left(W_d^\top W_d\right)^- W_d^\top V_d$$
$$= \frac{1}{K}\left(0_A|V_{d1}^\top 1_K|\dots|V_{dB}^\top 1_K\right)\begin{pmatrix}\sum_{b=1}^B 1_K^\top V_{db}\\ 1_K^\top V_{d1}\\ \vdots\\ 1_K^\top V_{dB}\end{pmatrix} = \frac{1}{K}\sum_{b=1}^B V_{db}^\top 1_K\, 1_K^\top\, V_{db}.$$

Since

$$V_{db}^\top 1_K = \begin{pmatrix} N_{1b}(d)\\ \vdots \\ N_{Ab}(d)\end{pmatrix},$$

we obtain the second part of the assertion a) by

$$V_d^\top W_d \left(W_d^\top W_d\right)^- W_d^\top V_d$$
$$= \frac{1}{K}\sum_{b=1}^B \begin{pmatrix} N_{1b}(d)\\ \vdots \\ N_{Ab}(d)\end{pmatrix}(N_{1b}(d),\dots,N_{Ab}(d))$$
$$= \frac{1}{K}\sum_{b=1}^B \begin{pmatrix} N_{1b}(d)^2 & N_{1b}(d)N_{2b}(d) & \dots & N_{1b}(d)N_{Ab}(d)\\ N_{2b}(d)N_{1b}(d) & N_{2b}(d)^2 & \dots & N_{2b}(d)N_{Ab}(d)\\ \vdots & \vdots & & \vdots\\ N_{Ab}(d)N_{1b}(d) & N_{Ab}(d)N_{2b}(d) & \dots & N_{Ab}(d)^2\end{pmatrix}.$$

b) Since $N_{ab}(d) = M = \frac{K}{A}$ for all $(a,b) \in \{1,\dots,A\}\times\{1,\dots,B\}$ for completely balanced designs, we obtain from a)

$$V_{d_*}^\top\,\omega^\perp(W_{d_*})\,V_{d_*} = \frac{K\,B}{A}I_{A\times A} - \frac{1}{K}\frac{B\,K^2}{A^2}1_A\,1_A^\top = \frac{K\,B}{A}\left(I_{A\times A} - \frac{1}{A}1_A\,1_A^\top\right).$$

Lemma 8.3.2 provides

$$\left(V_{d_*}^\top\,\omega^\perp(W_{d_*})\,V_{d_*}\right)^- = \frac{A}{K\,B}\left(I_{A\times A} - \frac{1}{A}1_A\,1_A^\top\right)$$

so that with $L_\alpha\, 1_A = 0$ it holds

$$L_\alpha\,(V_{d_*}^\top\,\omega^\perp(W_{d_*})\,V_{d_*})^-$$
$$= \left(-1_{A-1}|I_{(A-1)\times(A-1)}\right)\frac{A}{K\,B}\left(I_{A\times A} - \frac{1}{A}1_A\,1_A^\top\right) = \frac{A}{K\,B}\left(-1_{A-1}|I_{(A-1)\times(A-1)}\right) \quad (5.8)$$

and with Lemma 2.4.5 b)

$$\begin{aligned} I_\lambda(d_*)^{-1} &= L_\alpha\, (V_{d_*}^\top\, \omega^\perp(W_{d_*})\, V_{d_*})^-\, L_\alpha^\top \\ &= \frac{A}{K\,B}\, \left(-1_{A-1}|I_{(A-1)\times(A-1)}\right) \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \\ &= \frac{A}{K\,B}\, \left(I_{(A-1)\times(A-1)} + 1_{A-1}1_{A-1}^\top\right). \end{aligned}$$

Lemma 8.3.3 yields

$$\begin{aligned} I_\lambda(d_*) &= \frac{K\,B}{A}\, \left(I_{(A-1)\times(A-1)} + 1_{A-1}1_{A-1}^\top\right)^{-1} \\ &= \frac{K\,B}{A} \left(I_{(A-1)\times(A-1)} - \frac{1}{1+1_{A-1}^\top 1_{A-1}} 1_{A-1}1_{A-1}^\top\right) = \frac{K\,B}{A} \left(I_{(A-1)\times(A-1)} - \frac{1}{A} 1_{A-1}1_{A-1}^\top\right). \end{aligned}$$

To show the last assertion of b) note that Lemma 2.4.2 provides

$$\begin{aligned} L\, I_\theta(d_*)^- &= \left(L_\alpha|0_{(A-1)\times(B+1)}\right) \begin{pmatrix} V_{d_*}^\top V_{d_*} & V_{d_*}^\top W_{d_*} \\ W_{d_*}^\top V_{d_*} & W_{d_*}^\top W_{d_*} \end{pmatrix}^- \\ &= \left(L_\alpha|0_{(A-1)\times(B+1)}\right) \begin{pmatrix} I^- & -I^- V_{d_*}^\top W_{d_*}\, (W_{d_*}^\top W_{d_*})^- \\ E_1 & E_2 \end{pmatrix} \\ &= \left(L_\alpha\, I^- \,\middle|\, -L_\alpha\, I^- V_{d_*}^\top W_{d_*}\, (W_{d_*}^\top W_{d_*})^-\right), \end{aligned}$$

where $I = V_{d_*}^\top\, \omega^\perp(W_{d_*})\, V_{d_*}$. Since $V_{d_*b}^\top 1_K = \frac{K}{A} 1_A$ for all $b = 1, \ldots B$ for completely balanced designs, (5.7) is now

$$V_{d_*}^\top W_{d_*} = \left(\sum_{b=1}^B V_{d_*b}^\top 1_K \,\middle|\, V_{d_*1}^\top 1_K \,\middle|\, \ldots \,\middle|\, V_{d_*B}^\top 1_K\right) = \frac{K}{A}\left(B\, 1_A | 1_A\, 1_B^\top\right)$$

so that we obtain with (5.6)

$$V_{d_*}^\top W_{d_*}\, (W_{d_*}^\top W_{d_*})^- = \frac{K}{A}\left(B\, 1_A|1_A\, 1_B^\top\right) \frac{1}{K} \begin{pmatrix} 0 & 0_B^\top \\ 0_B & I_{B\times B} \end{pmatrix} = \frac{1}{A}\left(0_A|1_A\, 1_B^\top\right).$$

The expression (5.8) provides then

$$\begin{aligned} L\, I_\theta(d_*)^- &= \left(L_\alpha\, I^- \,\middle|\, -L_\alpha\, I^- V_{d_*}^\top W_{d_*}\, (W_{d_*}^\top W_{d_*})^-\right) \\ &= \frac{A}{K\,B} \left(\left(-1_{A-1}|I_{(A-1)\times(A-1)}\right) \,\middle|\, \left(-1_{A-1}|I_{(A-1)\times(A-1)}\right) \frac{1}{A} \left(0_A|1_A\, 1_B^\top\right)\right) \\ &= \frac{A}{K\,B}\, \left(-1_{A-1}|I_{(A-1)\times(A-1)}|0_{A-1}|0_{(A-1)\times B}\right). \end{aligned}$$

Hence also the last assertion is proved. □

5.5.2 Theorem
The uniform design δ_* on $\mathcal{T} = \{1, \dots, A\} \times \{1, \dots, B\}$ and thus every completely balanced design $d_* \in \Omega_{A,B,K}$ is D_λ-optimal in $\Omega_{A,B,K}$.

Proof. Lemma 5.5.1 b) provides for $x(t) = x((a,b)) = (u_a^\top, 1, \tilde{u}_b^\top)^\top$, where u_a is a unit vector in $\mathbb{R}^A$ and $\tilde{u}_b$ a unit vector in $\mathbb{R}^B$,

$$\begin{aligned} &L\, I_\theta(d_*)^- x(t) \\ &= \frac{A}{K\,B} \left(-1_{A-1} | I_{(A-1)\times(A-1)} | 0_{A-1} | 0_{(A-1)\times B}\right) \begin{pmatrix} u_a \\ 1 \\ \tilde{u}_b \end{pmatrix} = \frac{A}{K\,B} \left(-1_{A-1} | I_{(A-1)\times(A-1)}\right) u_a. \end{aligned}$$

If $a = 1$ we obtain

$$L\, I_\theta(d_*)^- x(t) = \frac{-A}{K\,B}\, 1_{A-1}$$

so that using Lemma 5.5.1 b) we have

$$\begin{aligned} &x(t)^\top I_\theta(d_*)^- L^\top \left(L\, I_\theta(d_*)^- L^\top\right)^{-1} L\, I_\theta(d_*)^- x(t) \\ &= x(t)^\top I_\theta(d_*)^- L^\top\, I_\lambda(d_*) L\, I_\theta(d_*)^- x(t) \\ &= \frac{-A}{K\,B}\, 1_{A-1}^\top\, \frac{K\,B}{A} \left(I_{(A-1)\times(A-1)} - \frac{1}{A} 1_{A-1} 1_{A-1}^\top\right) \frac{-A}{K\,B}\, 1_{A-1} \\ &= \frac{A}{K\,B}\, 1_{A-1}^\top \left(1_{A-1} - \frac{A-1}{A} 1_{A-1}\right) = \frac{A}{K\,B}\, 1_{A-1}^\top\, \frac{1}{A}\, 1_{A-1} = \frac{1}{K\,B}\,(A-1). \end{aligned}$$

If $a \neq 1$ we obtain

$$L\, I_\theta(d_*)^- x(t) = \frac{A}{K\,B}\, \check{u}_{a-1},$$

where $\check{u}_{a-1}$ is a unit vector in $\mathbb{R}^{A-1}$, so that

$$\begin{aligned} &x(t)^\top I_\theta(d_*)^- L^\top \left(L\, I_\theta(d_*)^- L^\top\right)^{-1} L\, I_\theta(d_*)^- x(t) \\ &= \frac{A}{K\,B}\, \check{u}_{a-1}^\top \left(I_{(A-1)\times(A-1)} - \frac{1}{A} 1_{A-1} 1_{A-1}^\top\right) \check{u}_{a-1} = \frac{A}{K\,B} \left(1 - \frac{1}{A}\right) = \frac{1}{K\,B}\,(A-1). \end{aligned}$$

Since $I_\theta(\delta_*) = \frac{1}{K\,B} I_\theta(d_*)$ we have for all $t \in \mathcal{T}$

$$x(t)^\top I_\theta(\delta_*)^- L^\top \left(L\, I_\theta(\delta_*)^- L^\top\right)^{-1} L\, I_\theta(\delta_*)^- x(t) = A - 1$$

so that the equivalence theorem for D-optimality (Theorem 3.3.10) provides the D_λ-optimality of δ_* and d_*, respectively. Note that the last arguments would also appear in proving the D-optimality of the uniform design for the one-way layout if the singular parametrization would be used. □

5.6 Optimality of balanced incomplete block designs

At first we prove that $V_{d_*}^\top \omega^\perp(W_{d_*})\, V_{d_*}$ of a balanced incomplete block designs d_* has a similar structure as that for a completely balanced design.

5.6.1 Lemma
If $d_* \in \Omega_{A,B,K}$ is balanced incomplete block design, then it holds

$$V_{d_*}^\top \omega^\perp(W_{d_*})\, V_{d_*} = \frac{B\,(K-1)}{A-1}\left(I_{A\times A} - \frac{1}{A} 1_A\, 1_A^\top\right).$$

For the special case $K = A$ the expression for $V_{d_*}^\top \omega^\perp(W_{d_*})\, V_{d_*}$ coincides with that for completely balanced block designs (see Lemma 5.5.1 b)).

Proof of Lemma 5.6.1. Since d_* is a balanced incomplete block design, it holds $N_{ab}(d_*) \in \{0,1\}$ for all $(a,b) \in \{1,\dots,A\} \times \{1,\dots,B\}$ and there exists $R, \lambda \in \mathbb{N}$ such that (see (5.1) and (5.2))

$$N_{a\cdot}(d_*) = R \text{ for all } a \in \{1,\dots,A\},$$
$$\sum_{b=1}^B N_{ab}(d_*)^2 = \sum_{b=1}^B N_{ab}(d_*) = N_{a\cdot}(d_*) = R \text{ for all } a \in \{1,\dots,A\},$$
$$\sum_{b=1}^B N_{a(1)\,b}(d_*)\, N_{a(2)\,b}(d_*) = \lambda \text{ for all } a(1),\, a(2) \in \{1,\dots,A\} \text{ with } a(1) \neq a(2).$$

Then Lemma 5.5.1 a) provides

$$\begin{aligned}
V_{d_*}^\top \omega^\perp(W_{d_*})\, V_{d_*} &= \operatorname{diag}(N_{1\cdot}(d_*),\dots,N_{A\cdot}(d_*)) \\
&\quad - \frac{1}{K}\sum_{b=1}^B \begin{pmatrix} N_{1b}(d_*)^2 & N_{1b}(d)N_{2b}(d_*) & \dots & N_{1b}(d_*)N_{Ab}(d_*) \\ N_{2b}(d_*)N_{1b}(d_*) & N_{2b}(d_*)^2 & \dots & N_{2b}(d_*)N_{Ab}(d_*) \\ \vdots & \vdots & & \vdots \\ N_{Ab}(d_*)N_{1b}(d_*) & N_{Ab}(d_*)N_{2b}(d_*) & \dots & N_{Ab}(d_*)^2 \end{pmatrix} \\
&= R\, I_{A\times A} - \frac{1}{K}\begin{pmatrix} R & \lambda & \dots & \lambda) \\ \lambda & R & \dots & \lambda \\ \vdots & \vdots & & \vdots \\ \lambda & \lambda & \dots & R \end{pmatrix} \\
&= \left(R - \frac{R}{K} + \frac{\lambda}{K}\right) I_{A\times A} - \frac{\lambda}{K} 1_A\, 1_A^\top.
\end{aligned}$$

The necessary conditions $A\,R = B\,K$ and $\lambda = \frac{B\,K\,(K-1)}{A\,(A-1)}$ (see (5.3) and (5.4)) yield

$$
\begin{aligned}
& V_{d_*}^\top \,\omega^\perp(W_{d_*})\, V_{d_*} \\
&= \left(R\,\frac{K-1}{K} + \frac{B\,(K-1)}{A\,(A-1)} \right) I_{A\times A} - \frac{B\,(K-1)}{A\,(A-1)}\, 1_A\, 1_A^\top \\
&= \frac{B\,(K-1)}{A-1} \left(\left(\frac{R\,(A-1)}{B\,K} + \frac{1}{A} \right) I_{A\times A} - \frac{1}{A}\, 1_A\, 1_A^\top \right) \\
&= \frac{B\,(K-1)}{A-1} \left(\left(\frac{R\,(A-1)}{A\,R} + \frac{1}{A} \right) I_{A\times A} - \frac{1}{A}\, 1_A\, 1_A^\top \right) \\
&= \frac{B\,(K-1)}{A-1} \left(I_{A\times A} - \frac{1}{A}\, 1_A\, 1_A^\top \right).
\end{aligned}
$$

□

Balanced incomplete block designs will be used in particular when the block size K is smaller than the number A of treatments. Hence competing block designs d will also satisfy $N_{ab}(d) \in \{0, 1\}$ for all $(a, b) \in \{1, \ldots, A\} \times \{1, \ldots, B\}$.

5.6.2 Lemma
If a block design $d \in \Omega_{A,B,K}$ satisfies $N_{ab}(d) \in \{0, 1\}$ for all $(a, b) \in \{1, \ldots, A\} \times \{1, \ldots, B\}$, then

$$
tr\left(V_d^\top \,\omega^\perp(W_d)\, V_d \right) = N \left(1 - \frac{1}{K} \right).
$$

Proof. Lemma 5.5.1 a) yields

$$
\begin{aligned}
\operatorname{tr}\left(V_d^\top \,\omega^\perp(W_d)\, V_d \right) &= \operatorname{tr}\Bigg(\operatorname{diag}(N_{1\cdot}(d), \ldots, N_{A\cdot}(d)) \\
&\quad - \frac{1}{K} \sum_{b=1}^{B} \begin{pmatrix} N_{1b}(d)^2 & N_{1b}(d) N_{2b}(d) & \ldots & N_{1b}(d) N_{Ab}(d) \\ N_{2b}(d) N_{1b}(d) & N_{2b}(d)^2 & \ldots & N_{2b}(d) N_{Ab}(d) \\ \vdots & \vdots & & \vdots \\ N_{Ab}(d) N_{1b}(d) & N_{Ab}(d) N_{2b}(d) & \ldots & N_{Ab}(d)^2 \end{pmatrix} \Bigg) \\
&= \sum_{a=1}^{A} N_{a\cdot}(d) - \frac{1}{K} \sum_{b=1}^{B} \sum_{a=1}^{A} N_{ab}(d)^2.
\end{aligned}
$$

Since $N_{ab}(d) \in \{0, 1\}$ for all $(a, b) \in \{1, \ldots, A\} \times \{1, \ldots, B\}$ we have

$$
\operatorname{tr}\left(V_d^\top \,\omega^\perp(W_d)\, V_d \right) = \sum_{a=1}^{A} N_{a\cdot}(d) - \frac{1}{K} \sum_{b=1}^{B} \sum_{a=1}^{A} N_{ab}(d) = N - \frac{1}{K}\, N
$$

so that the assertion is proved. □

5.6.3 Lemma
Any block design $\Omega_{A,B,K}$ satisfies

$$rk(V_d^\top \, \omega^\perp(W_d) \, V_d) \le A - 1.$$

Proof. Since $1_N = 1_{B\,K}$ is a column of W_d, it holds $\omega^\perp(W_d)1_N = 0_N$. Because of $V_d \, 1_A = 1_{B\,K} = 1_N$ we have

$$V_d^\top \, \omega^\perp(W_d) \, V_d \, 1_A = 0_A$$

so that $V_d^\top \, \omega^\perp(W_d) \, V_d$ has rank of at most $A-1$. □

5.6.4 Theorem
Let be $\lambda(\theta) = L_\alpha \alpha$ with $L_\alpha = (-1_{A-1} | I_{(A-1)\times(A-1)})$ the interestimg linear aspect and

$$\begin{aligned} \Omega^0_{A,B,K} &= \{d \in \Omega_{A,B,K};\ N_{ab}(d) \in \{0,1\} \text{ for all } (a,b) \in \{1,\ldots,A\} \times \{1,\ldots,B\} \\ &\qquad \text{and } rk(V_d^\top \, \omega^\perp(W_d) \, V_d) = A-1 \Big\}. \end{aligned}$$

If $d_* \in \Omega_{A,B,K}$ is a balanced incomplete block design, then it holds for all $d \in \Omega^0_{A,B,K}$

$$\det(I_\lambda(d)^{-1}) \ge \det(I_\lambda(d_*)^{-1}),$$

i.e. d_* is D_λ-optimal in $\Omega^0_{A,B,K}$.

Proof. Let be $d \in \Omega^0_{A,B,K}$ arbitrary. The proof of Lemma 5.6.3 shows that $u = \frac{1}{\sqrt{A}} 1_A$ is an eigenvector of $V_d^\top \, \omega^\perp(W_d) \, V_d$ with eigenvalue 0. Hence the spectral decomposition of $V_d^\top \, \omega^\perp(W_d) \, V_d$ has the form

$$V_d^\top \, \omega^\perp(W_d) \, V_d = (u|\tilde{U}) \operatorname{diag}(0, \xi_1, \ldots, \xi_{A-1}) \begin{pmatrix} u^\top \\ \tilde{U}^\top \end{pmatrix},$$

where $0, \xi_1, \ldots, \xi_{A-1}$ are the eigenvalues and $U = (u|\tilde{U})$ contains the eigenvectors of $V_d^\top \, \omega^\perp(W_d) \, V_d$. In particular, we have $U\,U^\top = I_{A\times A} = U^\top U$. Also note that all eigenvalues are not negative since $V_d^\top \, \omega^\perp(W_d) \, V_d$ is positive semidefinite. Moreover, $\operatorname{rk}\left(V_d^\top \, \omega^\perp(W_d) \, V_d\right) = A-1$ implies $0 < \xi_1 \le \ldots \le \xi_{A-1}$. Lemma 5.6.2 provides

$$\begin{aligned} N\left(1 - \frac{1}{K}\right) &= \operatorname{tr}\left(V_d^\top \, \omega^\perp(W_d) \, V_d\right) \\ &= \operatorname{tr}(U \operatorname{diag}(0, \xi_1, \ldots, \xi_{A-1}) \, U^\top) = \operatorname{tr}(\operatorname{diag}(0, \xi_1, \ldots, \xi_{A-1}) \, U^\top \, U) \\ &= \operatorname{tr}(\operatorname{diag}(0, \xi_1, \ldots, \xi_{A-1})) = \xi_1 + \ldots + \xi_{A-1}. \end{aligned} \tag{5.9}$$

Additionally, we have

$$\left(V_d^\top \, \omega^\perp(W_d) \, V_d\right)^- = (u|\tilde{U}) \operatorname{diag}(0, \xi_1^{-1}, \ldots, \xi_{A-1}^{-1}) \begin{pmatrix} u^\top \\ \tilde{U}^\top \end{pmatrix}.$$

Since $L_\alpha u = \frac{1}{\sqrt{A}} L_\alpha 1_A = 0_{A-1}$ because $\lambda(\theta) = L_\alpha \alpha$ is a contrast, the inverse of the information matrix has the form (see Lemma 2.4.5)

$$\begin{aligned}
I_\lambda(d)^{-1} &= L_\alpha \left[V_d^\top \omega^\perp(W_d)\, V_d \right]^- L_\alpha^\top \\
&= L_\alpha\, (u|\tilde{U})\ \mathrm{diag}(0, \xi_1^{-1}, \ldots, \xi_{A-1}^{-1}) \begin{pmatrix} u^\top \\ \tilde{U}^\top \end{pmatrix} L_\alpha^\top \\
&= (0_{A-1}|L_\alpha \tilde{U})\ \mathrm{diag}(0, \xi_1^{-1}, \ldots, \xi_{A-1}^{-1}) \begin{pmatrix} 0_{A-1}^\top \\ \tilde{U}^\top L_\alpha^\top \end{pmatrix} \\
&= L_\alpha \tilde{U}\ \mathrm{diag}(\xi_1^{-1}, \ldots, \xi_{A-1}^{-1})\, \tilde{U}^\top L_\alpha^\top .
\end{aligned}$$

Hence the determinant is given by

$$\begin{aligned}
&\det(I_\lambda(d)^{-1}) = \det(L_\alpha \tilde{U})\ \det(\mathrm{diag}(\xi_1^{-1}, \ldots, \xi_{A-1}^{-1}))\ \det(\tilde{U}^\top L_\alpha^\top) \\
&= \frac{\det(L_\alpha \tilde{U} \tilde{U}^\top L_\alpha^\top)}{\xi_1 \cdot \ldots \cdot \xi_{A-1}} = \frac{1}{\xi_1 \cdot \ldots \cdot \xi_{A-1}} \det\left((0_{A-1}|L_\alpha \tilde{U}) \begin{pmatrix} 0_{A-1}^\top \\ \tilde{U}^\top L_\alpha^\top \end{pmatrix} \right) \\
&= \frac{1}{\xi_1 \cdot \ldots \cdot \xi_{A-1}} \det\left(L_\alpha\, (u|\tilde{U}) \begin{pmatrix} u^\top \\ \tilde{U}^\top \end{pmatrix} L_\alpha^\top \right) = \frac{\det(L_\alpha U U^\top L_\alpha^\top)}{\xi_1 \cdot \ldots \cdot \xi_{A-1}} = \frac{\det(L_\alpha L_\alpha^\top)}{\xi_1 \cdot \ldots \cdot \xi_{A-1}}.
\end{aligned}$$

Hence $\det(I_\lambda(d)^{-1})$ is minimized if $\xi_1 \cdot \ldots \cdot \xi_{A-1}$ is maximized under the side condition (5.9), i.e. $\xi_1 + \ldots + \xi_{A-1} = N\left(1 - \frac{1}{K}\right)$. Since $\xi_1 \cdot \ldots \cdot \xi_{A-1}$ is maximized if and only if $\ln(\xi_1) + \ldots + \ln(\xi_{A-1})$ is maximized we consider the Lagrange multiplier function

$$L(\xi_1, \ldots, \xi_{A-1}, \eta) = \sum_{a=1}^{A-1} \ln(\xi_a) + \eta \left(\sum_{a=1}^{A-1} \xi_a - N\left(1 - \frac{1}{K}\right) \right).$$

Since

$$\begin{aligned}
\frac{\partial}{\partial \xi_a} L(\xi_1, \ldots, x_{A-1}, \eta) &= \frac{1}{\xi_a} + \eta = 0 \text{ for } a = 1, \ldots, A-1, \\
\frac{\partial}{\partial \eta} L(\xi_1, \ldots, x_{A-1}, \eta) &= \sum_{a=1}^{A-1} \xi_a - N\left(1 - \frac{1}{K}\right) = 0
\end{aligned}$$

if and only if

$$\xi_a = \frac{1}{A-1} N \left(1 - \frac{1}{K}\right) = \frac{B\,K}{A-1} \frac{K-1}{K} = \frac{B\,(K-1)}{A-1}, \tag{5.10}$$

$\det(I_\lambda(d)^{-1})$ is minimized if all eigenvalues ξ_a satisfy (5.10). This is the case for the balanced incomplete block design d_* according to Lemma 5.6.1 . Hence d_* is D_λ-optimal in $\Omega^0_{A,B,K}$. □

As in the one-way layout, a balanced design is not A_λ-optimal if the aspect $\lambda(\theta) = L_\alpha \alpha$ is given by $L_\alpha = (-1_{A-1}|I_{(A-1)\times(A-1)})$. However, if L_α consists of all contrasts between two treatments, i.e. $L_\alpha \in \mathbb{R}^{A(A-1)\times A}$, then it can be proved that the balanced incomplete block design is A-optimal.

Chapter 6

Optimal designs for models with more than two factors

6.1 Latin square and graeco-latin square designs

If there are two block factors and one treatment factor and all three factors have the same number of levels, than a latin square design is the best design.

6.1.1 Definition
A **latin square design** *is a design which allocates to each level combination of two block factors exactly one treatment level such that for each block level all treatment levels are used. Thereby the numbers of treatment and block levels are the same.*

Latin square designs are produced by `design.lsd` of the `agricolae` package. For example,

```
> library(agricolae)
> design.lsd(c("A","B","C","D"))$book
   plots row col c("A", "B", "C", "D")
1    101   1   1                     D
2    102   1   2                     C
3    103   1   3                     A
4    104   1   4                     B
5    201   2   1                     A
6    202   2   2                     D
7    203   2   3                     B
8    204   2   4                     C
9    301   3   1                     C
10   302   3   2                     B
11   303   3   3                     D
12   304   3   4                     A
13   401   4   1                     B
```

```
14    402   4   2                      A
15    403   4   3                      C
16    404   4   4                      D
```

provides the following latin square

	1	2	3	4
1	D	C	A	B
2	A	D	B	C
3	C	B	D	A
4	B	A	C	D

Here the rows are the levels of the first block factor and the columns are the levels of the second block factor. The capital letters A,B,C,D denote the four levels of the treatment.

If there are two block factors and two treatment factors and all four factors have the same number of levels, than a graeco-latin square design is the best design.

6.1.2 Definition
A **graeco-latin square** *design is a design which allocates to each level combination of two block factors exactly one combination of levels of two treatment factors such that all levels of the first and the second treatment factor are used for each block level and all combinations of treatments appear. Thereby the numbers of treatment and block levels are the same.*

Graeco-latin square designs are produced by `design.graeco` of the `agricolae` package. For example,

```
> library(agricolae)
> design.graeco(c("A","B","C","D"),c("a","b","c","d"))$book
  plots row col c("A", "B", "C", "D") c("a", "b", "c", "d")
1     101   1   1                      D                     d
2     102   1   2                      B                     a
3     103   1   3                      A                     b
4     104   1   4                      C                     c
5     201   2   1                      B                     b
6     202   2   2                      D                     c
7     203   2   3                      C                     d
8     204   2   4                      A                     a
9     301   3   1                      A                     c
10    302   3   2                      C                     b
11    303   3   3                      D                     a
12    304   3   4                      B                     d
13    401   4   1                      C                     a
14    402   4   2                      A                     d
15    403   4   3                      B                     c
16    404   4   4                      D                     b
```

provides the following graeco-latin square

	1	2	3	4
1	Dd	Ba	Ab	Cc
2	Bb	Dc	Cd	Aa
3	Ac	Cb	Da	Bd
4	Ca	Ad	Bc	Db

Here the small letters stands for the graeco letters.

The level combinations of a latin square or graeco-latin square design can have repetitions or not. If they have repetitions, the number of repetitions should be equal for all combinations of the design. Many latin and graeco latin square designs have no repetitions. Then not all interactions can be estimated or tested. Even in the case of 4 factors and 3 levels for each factor, only an additive model, i.e. a model without interactions, can be estimated and no hypothesis about this model can be tested.

Generalizations of the latin square or graeco-latin square designs are the Youden designs and generalized Youden designs. For all these designs, D- and A-optimality can be shown.

6.2 Factorial designs

Latin square designs and graeco-latin square designs are special fractional factorial designs.

6.2.1 Definition
Assume that there are K factors $F_1, \ldots, F_K$, where factor F_k has p_k levels, $k = 1, \ldots, K$.
A **complete factorial design** *or* **full factorial design** *is a design where each of the $p_1 \cdot p_2 \cdot \ldots \cdot p_K$ level combinations is realized. If $p_1 = p_2 = \ldots = p_K = p$, then such a design is also called a p^K-design.*
A **fractional factorial design** *is a design where only some of the $p_1 \cdot p_2 \cdot \ldots \cdot p_K$ level combinations are realized.*

The level combinations of a full factorial design or fractional factorial design can be repeated $M \geq 1$ times.

As in latin square and graeco-latin square designs, not all interactions can be estimated in fractional factorial designs. However, some linear combinations of the interaction parameters can be estimated. If interaction parameters are only estimable within such linear combinations, then they are called **confounded**.

However, in complete factorial, all interactions are estimable. The function `design.ab` of the `agricolae` package provides also complete factorial designs, where the allocation to the experimental units within blocks is done randomly.

6.2.2 Example (Complete factorial design)
If there are three factors, each with two levels, which should be allocated to the experimental units of four blocks, type

```
> design.ab(c(2,2,2),4)$book
   plots block A B C
1     101     1 2 2 1
2     102     1 2 1 1
3     103     1 1 1 2
4     104     1 1 2 2
5     105     1 1 1 1
6     106     1 2 1 2
7     107     1 1 2 1
8     108     1 2 2 2
9     109     2 2 1 2
10    110     2 1 1 2
11    111     2 1 1 1
12    112     2 2 1 1
13    113     2 2 2 2
14    114     2 2 2 1
15    115     2 1 2 1
16    116     2 1 2 2
17    117     3 1 2 2
```

```
18    118      3 1 1 1
19    119      3 2 1 2
20    120      3 2 1 1
21    121      3 2 2 1
22    122      3 1 2 1
23    123      3 1 1 2
24    124      3 2 2 2
25    125      4 2 1 2
26    126      4 1 2 1
27    127      4 1 1 2
28    128      4 2 2 2
29    129      4 2 2 1
30    130      4 2 1 1
31    131      4 1 2 2
32    132      4 1 1 1
```

In each block we have $2^3 = 8$ level combinations of the three factors.

6.3 Factorial designs for factors with two levels

The simplest case is the case of K factors $F_1, F_1, \ldots, F_K$ each with two levels. The levels can be given as $0, 1$ or $-1, 1$.

6.3.1 Example
A full factorial 2^3-design for 3 factors A, B, C is given by:

A	B	C		A	B	C
0	0	0		1	1	1
1	0	0		−1	1	1
0	1	0		1	−1	1
1	1	0	or	−1	−1	1
0	0	1		1	1	−1
1	0	1		−1	1	−1
0	1	1		1	−1	−1
1	1	1		−1	−1	−1

2^K-designs can be used also for multiple regression. Lemma 4.1.1 shows that the 2^1-designs with equal number of realizations at $-a = -1$ and $a = 1$ is D- and A-optimal within all designs on $[-1, 1]$. Moreover, the design of Lemma 5.1.1 is a 2^2-design. This design is D- and A-optimal within all designs on $[-1, 1] \times [-1, 1]$. These examples show that the designs points can be chosen at the border of the design region. This holds also for multiple regression with more than two factors on $[-1, 1]^K$ or $[0, 1]^K$, respectively ($K > 2$). Hence designs for multiple regression can be restricted on $\{-1, 1\}^K$ or $\{0, 1\}^K$.

Here we will consider only $\{0, 1\}^K$ as design region. Hence $t = (\tau_1, \ldots, \tau_K)^\top \in \{0, 1\}^K$ is an experimental condition. There are 2^K different experimental conditions. In a full 2^K-design all main effects and all interactions are identifiable and thus estimable. Consider the following model for $t = (\tau_1, \ldots, \tau_K)^\top \in \{0, 1\}^K$:

$$\begin{aligned} \mu(t) &= \tilde{\mu}_0 + \tilde{\mu}_1\tilde{\tau}_1 + \ldots + \tilde{\mu}_K\tilde{\tau}_K + \tilde{\mu}_{12}\tilde{\tau}_1\tilde{\tau}_2 + \tilde{\mu}_{13}\tilde{\tau}_1\tilde{\tau}_3 + \ldots \\ &\quad + \tilde{\mu}_{(K-1)K}\tilde{\tau}_{K-1}\tilde{\tau}_K + \ldots + \tilde{\mu}_{12\ldots K}\tilde{\tau}_1\tilde{\tau}_2\ldots\tilde{\tau}_K \end{aligned} \tag{6.1}$$

with $\tilde{\tau}_k = 2(\frac{1}{2} - \tau_k)$, i.e. $\tilde{\tau}_k = 1$ for $\tau_k = 0$, $\tilde{\tau}_k = -1$ for $\tau_k = 1$.
Thereby we have

$\tilde{\mu}_1, \ldots, \tilde{\mu}_K$	are the main effects of the factors $F_1, \ldots, F_n$,
$\tilde{\mu}_{ij}$	are interactions between F_i and F_j,
$\tilde{\mu}_{ijk}$	are interaction of order three,
$\vdots$	
$\tilde{\mu}_{12\ldots K}$	are interactions of order K.

There are 1 overall mean $+K$ main effects $+\binom{K}{2}$ interactions of order two $+\binom{K}{3}$ interactions of order $3 + \ldots + \binom{K}{K}$ interactions if order K

$$= \sum_{k=0}^{K} \binom{K}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{K-k} 2^K = \left(\frac{1}{2}+\frac{1}{2}\right)^K 2^K = 2^K$$

parameters. Hence the number of parameters coincides with the number of different experimental conditions.

6.3.2 Definition

A model which includes all interactions is called a saturated model.

6.3.3 Example (Continuation of Example 6.3.1)

In the model of Example 6.3.1 with 3 Factors A, B, C, there are $2^3 = 8$ parameters $\tilde{\mu}_0,\ \tilde{\mu}_1,\ \tilde{\mu}_2,\ \tilde{\mu}_3,\ \tilde{\mu}_{12},\ \tilde{\mu}_{13},\ \tilde{\mu}_{23},\ \tilde{\mu}_{123}$. They are all identifiable in the 2^3-design, because of

design			design matrix							
τ_1	τ_2	τ_3	I	$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2$	$\tilde{\tau}_1\tilde{\tau}_3$	$\tilde{\tau}_2\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2\tilde{\tau}_3$
0	0	0	1	1	1	1	1	1	1	1
1	0	0	1	-1	1	1	-1	-1	1	-1
0	1	0	1	1	-1	1	-1	1	-1	-1
1	1	0	1	-1	-1	1	1	-1	-1	1
0	0	1	1	1	1	-1	1	-1	-1	-1
1	0	1	1	-1	1	-1	-1	1	-1	1
0	1	1	1	1	-1	-1	-1	-1	1	1
1	1	1	1	-1	-1	-1	1	1	1	-1

The columns of this design matrix are pairwise orthogonal so that $X_d^\top X_d = 8I_{8\times 8}$. Hence all parameters are identifiable at d.

6.3.4 Lemma

Any 2^K-design has a design matrix with columns $1_{2^K}, X_d^1, \ldots, X_d^{2^K-1}$ which are pairwise orthogonal and satisfy $\|X_d^n\|^2 = 2^K$ for $n = 1, \ldots, 2^K - 1$.

Proof. Since the elements of the design matrix are 1 and -1, the assertion $\|X_d^n\|^2 = 2^K$ is obviously true for $n = 1, \ldots, 2^K - 1$.

To prove the orthogonality of the columns, let be $B_K = (B_K^1|\ldots|B_K^K)$ the design matrix of the main effects. Then the column of the design matrix X_d for the interaction term $\tilde{\mu}_{i_1 i_2 \ldots i_k}$ is given by

$$B_K^{i_1} * B_K^{i_2} * \ldots * B_K^{i_k},$$

where $B_K^i * B_K^j$ denotes the pointwise multiplication of two vectors. Now use induction. The assertion is obviously true for $K = 1, 2, 3$ (for $K = 3$ see Example 6.3.3). Assume that the assumption is true for K. Then B_{K+1} can be written as

$$B_{K+1} = (B^1_{K+1}|\ldots|B^{K+1}_{K+1}) = \begin{pmatrix} B_K & 1_{2^K} \\ B_K & -1_{2^K} \end{pmatrix} = \begin{pmatrix} B^1_K & B^2_K & \ldots B^K_K & 1_{2^K} \\ B^1_K & B^2_K & \ldots B^K_K & -1_{2^K} \end{pmatrix},$$

where $B_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Since the assertion is true for K, we have for all different terms $\tilde{\mu}_{i_1 i_2 \ldots i_k}$ and $\tilde{\mu}_{j_1 j_2 \ldots j_l}$

$$(B^{i_1}_K * B^{i_2}_K * \ldots * B^{i_k}_K)^\top (B^{j_1}_K * B^{j_2}_K * \ldots * B^{j_l}_K) = 0$$

and

$$(B^{i_1}_K * B^{i_2}_K * \ldots * B^{i_k}_K)^\top 1_{2^K} = 0.$$

Then we obtain for $K+1$ for all terms involving the factor $K+1$

$$\begin{aligned}
&(B^{i_1}_{K+1} * \ldots * B^{i_k}_{K+1} * B^{K+1}_{K+1})^\top (B^{j_1}_{K+1} * \ldots * B^{j_l}_{K+1} * B^{K+1}_{K+1}) \\
&= \left(\begin{pmatrix} B^{i_1}_K \\ B^{i_1}_K \end{pmatrix} * \ldots * \begin{pmatrix} B^{i_k}_K \\ B^{i_k}_K \end{pmatrix} * \begin{pmatrix} 1_{2^K} \\ -1_{2^K} \end{pmatrix}\right)^\top \left(\begin{pmatrix} B^{j_1}_K \\ B^{j_1}_K \end{pmatrix} * \ldots * \begin{pmatrix} B^{j_l}_K \\ B^{j_l}_K \end{pmatrix} * \begin{pmatrix} 1_{2^K} \\ -1_{2^K} \end{pmatrix}\right) \\
&= \begin{pmatrix} B^{i_1}_K * \ldots * B^{i_k}_K * 1_{2^K} \\ B^{i_1}_K * \ldots * B^{i_k}_K * -1_{2^K} \end{pmatrix}^\top \begin{pmatrix} B^{j_1}_K * \ldots * B^{j_l}_K * 1_{2^K} \\ B^{j_1}_K * \ldots * B^{j_l}_K * -1_{2^K} \end{pmatrix} \\
&= \begin{pmatrix} *B^{i_1}_K * \ldots * B^{i_k}_K \\ -B^{i_1}_K * \ldots * B^{i_k}_K \end{pmatrix}^\top \begin{pmatrix} *B^{j_1}_K * \ldots * B^{j_l}_K \\ - * B^{j_1}_K * \ldots * B^{j_l}_K \end{pmatrix} \\
&= \left(B^{i_1}_K * \ldots * B^{i_k}_K\right)^\top \left(B^{j_1}_K * \ldots * B^{j_l}_K\right) + \left(-B^{i_1}_K * \ldots * B^{i_k}_K\right)^\top \left(-B^{j_1}_K * \ldots * B^{j_l}_K\right) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
&(B^{i_1}_{K+1} * \ldots * B^{i_k}_{K+1} * B^{K+1}_{K+1})^\top (B^{K+1}_{K+1}) \\
&= \left(\begin{pmatrix} B^{i_1}_K \\ B^{i_1}_K \end{pmatrix} * \ldots * \begin{pmatrix} B^{i_k}_K \\ B^{i_k}_K \end{pmatrix} * \begin{pmatrix} 1_{2^K} \\ -1_{2^K} \end{pmatrix}\right)^\top \left(\begin{pmatrix} 1_{2^K} \\ -1_{2^K} \end{pmatrix}\right) \\
&= \begin{pmatrix} B^{i_1}_K * \ldots * B^{i_k}_K \\ -B^{i_1}_K * \ldots * B^{i_k}_K \end{pmatrix}^\top \begin{pmatrix} 1_{2^K} \\ -1_{2^K} \end{pmatrix} \\
&= \left(B^{i_1}_K * \ldots * B^{i_k}_K\right)^\top 1_{2^K} - \left(-B^{i_1}_K * \ldots * B^{i_k}_K\right)^\top 1_{2^K} \\
&= 0.
\end{aligned}$$

The same arguments provide the assertion for terms not involving the factor $K+1$. Since additionally $(1_{2^K}| - 1_{2^K})^\top 1_{2^{K+1}} = 0$, all columns of the design matrix for $K+1$ factors are orthogonal. □

6.3.5 Corollary
If d is a 2^K-design with columns $1_{2^K}, X_d^1, \ldots, X_d^{2^K-1}$ and $X_d^k = (\tilde{\tau}_k^{(1)}, \ldots, \tilde{\tau}_k^{(2^K)})^\top$ for $k = 1, \ldots, K$, then for all $0_K \neq \alpha \in \{0,1\}^K$ there exists $\{n_1, \ldots, n_{2^{K-1}}\} \subset \{1, \ldots, 2^K\}$ with

$$\begin{aligned}
(\tilde{\tau}_1^{(n)})^{\alpha_1} \cdot (\tilde{\tau}_2^{(n)})^{\alpha_2} \cdot \ldots \cdot (\tilde{\tau}_K^{(n)})^{\alpha_K} &= -1 \quad \text{for} \quad n \in \{n_1, \ldots, n_{2^{K-1}}\},\\
(\tilde{\tau}_1^{(n)})^{\alpha_1} \cdot (\tilde{\tau}_2^{(n)})^{\alpha_2} \cdot \ldots \cdot (\tilde{\tau}_K^{(n)})^{\alpha_K} &= 1 \quad \text{for} \quad n \notin \{n_1, \ldots, n_{2^{K-1}}\}.
\end{aligned}$$

6.3.6 Theorem
Let be $\theta = (\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_K, \tilde{\mu}_{12}, \tilde{\mu}_{13}, \ldots, \tilde{\mu}_{1K}, \ldots, \tilde{\mu}_{(K-1)K}, \ldots, \tilde{\mu}_{12\ldots K})^\top = (\theta_1, \ldots, \theta_{2^K})^\top$ and $X_d = (X_d^1|\ldots|X_d^{2^K})$ the design matrix for the saturated model with columns corresponding to θ. If d is a 2^K-design then it holds

$$\begin{aligned}
&\widehat{\theta_j} = \frac{1}{2^K}(X_d^j)^\top y,\\
&var(\widehat{\theta_j}) = \frac{1}{2^K}\sigma^2,\\
&cov(\widehat{\theta_j}, \widehat{\theta_l}) = 0,
\end{aligned}$$

for all $j, l \in \{1, \ldots, 2^K\}$. This holds also for any submodel

$$y = X_d^J \theta^J + z,$$

with $j, l \in J = \{j_1, \ldots, j_k\} \subset \{1, \ldots, 2^K\}$, $\theta^J = (\theta_{j_1}, \ldots, \theta_{j_k})^\top$, and $X_d^J = (X_d^{j_1}|\ldots|X_d^{j_k})$.

Proof. Let be $J = \{j_1, \ldots, j_k\} \subset \{1, \ldots, 2^K\}$ arbitrary and without loss of generality $j = j_1$. Then we have $u_1^\top \theta^J = \theta_j$ if u_1 is the first unit vector in $\mathbb{R}^k$. Because of Lemma 6.3.4, the Gauss-Markov estimator for θ_j satisfies

$$\widehat{\theta_j} = u_1^\top \left((X_d^J)^\top X_d^J\right)^- (X_d^J)^\top y = u_1^\top \frac{1}{2^K} I_{k\times k} (X_d^J)^\top y = \frac{1}{2^K}(X_d^j)^\top y.$$

The other assertions are an exercise. □

6.4 Fractional factorial designs for factors with two levels

It is clear as soon as we have more and more factors, then the number of experimental units explodes. For example, for optimization the wheel for a car, there were 14 factors which could be varied. This would lead to a 2^14-design which has $2^{14} = 16384$ different experimental conditions. Such designs are usually not possible, the reason why fractional factorial designs are needed. The R package `conf.design` provides fractional factorial designs. In particular with this package designs can be constructed where specified treatment contrasts are confounded with blocks.

6.4.1 Example (Continuation of Example 6.3.3)
The 2^3-design in Example 6.3.3 can be reduced to a design with four units by dropping all design points $t = (\tau_1, \tau_2, \tau_3)^\top \in \{0,1\}^3$ with one or three values equal to 1, i.e. by using

design			design matrix							
τ_1	τ_2	τ_3	I	$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2$	$\tilde{\tau}_1\tilde{\tau}_3$	$\tilde{\tau}_2\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2\tilde{\tau}_3$
0	0	0	1	1	1	1	1	1	1	1
0	1	1	1	1	-1	-1	-1	-1	1	1
1	0	1	1	-1	1	-1	-1	1	-1	1
1	1	0	1	-1	-1	1	1	-1	-1	1

This is a incomplete factorial design. In such design not all parameters of the full (saturated) model (6.1) are identifiable. For example $\tilde{\mu}_3$ and $\tilde{\mu}_{12}$ are not identifiable since the columns for $\tilde{\tau}_3$ and $\tilde{\tau}_1\tilde{\tau}_2$ are identical. But $\tilde{\mu}_3 + \tilde{\mu}_{12}$ is identifiable. We say that $\tilde{\mu}_3$ and $\tilde{\mu}_{12}$ are confounded. The same holds for $\tilde{\mu}_2$ and $\tilde{\mu}_{13}$, $\tilde{\mu}_1$ and $\tilde{\mu}_{23}$, as well as $\tilde{\mu}_0$ and $\tilde{\mu}_{123}$. But if we use the smaller model given by

$$\mu(t) = \tilde{\mu}_0 + \tilde{\mu}_1\tilde{\tau}_1 + +\tilde{\mu}_2\tilde{\tau}_2 + \tilde{\mu}_3\tilde{\tau}_3$$

then all four parameters of this model are identifiable since

design			design matrix				removed			
τ_1	τ_2	τ_3	I	$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2$	$\tilde{\tau}_1\tilde{\tau}_3$	$\tilde{\tau}_2\tilde{\tau}_3$	$\tilde{\tau}_1\tilde{\tau}_2\tilde{\tau}_3$
0	0	0	1	1	1	1				
0	1	1	1	1	-1	-1				
1	0	1	1	-1	1	-1				
1	1	0	1	-1	-1	1				

i.e. the design matrix has columns which are pairwise orthogonal.

Use again

$$\theta = (\theta_1, \ldots, \theta_{2^K})^\top = (\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_K, \tilde{\mu}_{12}, \ldots, \tilde{\mu}_{(K-1)K}, \ldots, \tilde{\mu}_{12\ldots K})^\top$$

and

$$\begin{aligned} x(t) &= (x_1(t), \ldots, x_{2^K}(t))^\top \\ &= (\tilde{x}_0(t), \tilde{x}_1(t), \ldots, \tilde{x}_K(t), \tilde{x}_{12}(t), \ldots, \tilde{x}_{(K-1)K}(t), \ldots, \tilde{x}_{12\ldots K}(t))^\top \\ &= (1, \tilde{\tau}_1, \ldots, \tilde{\tau}_K, \tilde{\tau}_1\tilde{\tau}_2, \ldots, \tilde{\tau}_{K-1}\tilde{\tau}_K, \ldots, \tilde{\tau}_1\tilde{\tau}_2 \cdot \ldots \cdot \tilde{\tau}_K)^\top. \end{aligned}$$

Then $\mu(t) = x(t)^\top \theta$ is the **saturated model** (complete model).

6.4.2 Definition

a) Factorial effects $\tilde{\mu}_{i_1 \ldots i_g}$ and $\tilde{\mu}_{j_1 \ldots j_h}$ with $\{i_1, \ldots, i_g\} \neq \{j_1, \ldots, j_h\}$ are called confounded in the incomplete 2^K-design $d = \{t^{(1)}, \ldots, t^{(N)}\}$ with $t_n \in \{0,1\}^K$ (briefly $F_{i_1} \ldots F_{i_g} = F_{j_1} \ldots F_{j_h}$ or $i_1 \ldots i_g = j_1 \ldots j_h$) if

$$\tilde{\tau}_{i_1} \cdot \ldots \cdot \tilde{\tau}_{i_g} = \tilde{\tau}_{j_1} \cdot \ldots \cdot \tilde{\tau}_{j_h} \text{ for all } t \in d.$$

b) A factorial effect $\tilde{\mu}_{i_1 \ldots i_g}$ is confounded with the overall mean $\tilde{\mu}_0$ in the incomplete 2^K-design $d = \{t^{(1)}, \ldots, t^{(N)}\}$ with $t_n \in \{0,1\}^K$ (briefly $F_{i_1} \ldots F_{i_g} = I$ or $i_1 \ldots i_g = I$) if

$$\tilde{\tau}_{i_1} \cdot \ldots \cdot \tilde{\tau}_{i_g} = 1 \text{ for all } t \in d.$$

6.4.3 Lemma

If $\tilde{\mu}_{i_1 \ldots i_g}$ and $\tilde{\mu}_{j_1 \ldots j_h}$ are confounded in an incomplete 2^K-design d then $\lambda(\theta) = (\tilde{\mu}_{i_1 \ldots i_g}, \tilde{\mu}_{j_1 \ldots j_h})^\top$ is not identifiable at d in the saturated model. The same holds if $\tilde{\mu}_{j_1 \ldots j_h}$ is replaced by $\tilde{\mu}_0$.

Proof. Let be $\theta_i = \tilde{\mu}_{i_1 \ldots i_g}$, $\theta_j = \tilde{\mu}_{j_1 \ldots j_h}$ and u_i, u_j the i'th, j'th unit vectors in $\mathbb{R}^{2^K}$. Since $\tilde{\mu}_{i_1 \ldots i_g}$ and $\tilde{\mu}_{j_1 \ldots j_h}$ are confounded, it holds

$$x(t)^\top u_i = x_i(t) = x_j(t) = x(t)^\top u_j \text{ for all } t \in d$$

and thus $X_d u_i = X_d u_j$ or $X_d(u_i - u_j) = 0$, respectively. Since $\lambda(u_i - u_j) = (1, -1)^\top \neq 0_2$, $\lambda(\theta)$ is not identifiable at d in the saturated model. □

Now the confoundness of factorial effects is expressed by the design points $t \in \{0,1\}^K$ itself.

6.4.4 Theorem

Factorial effects $\tilde{\mu}_{i_1 \ldots i_g}$ and $\tilde{\mu}_{j_1 \ldots j_h}$ with $\{i_1, \ldots, i_g\} \neq \{j_1, \ldots, j_h\}$ are confounded in the incomplete 2^K-design d if and only if one of the following three conditions is satisfied

$$\begin{aligned} &a) \quad \tilde{\tau}_{i_1} \cdot \ldots \cdot \tilde{\tau}_{i_g} \cdot \tilde{\tau}_{j_1} \cdot \ldots \cdot \tilde{\tau}_{j_h} = 1 \text{ for all } t \in d, \\ &b) \quad mod_2(\tau_{i_1} + \ldots + \tau_{i_g} + \tau_{j_1} + \ldots + \tau_{j_h}) = 0 \text{ for all } t \in d, \\ &c) \quad mod_2(\tau_{i_1} + \ldots + \tau_{i_g}) = mod_2(\tau_{j_1} + \ldots + \tau_{j_h}) \text{ for all } t \in d. \end{aligned}$$

Proof.
$\Longleftrightarrow$ a): The assertion follows from the fact that $\tilde{\tau}_{i_1} \cdot \ldots \cdot \tilde{\tau}_{i_g}, \tilde{\tau}_{j_1} \cdot \ldots \cdot \tilde{\tau}_{j_h} \in \{-1, 1\}$.

a) $\Longleftrightarrow$ b): We have
$\text{mod}_2(t_{i_1} + \ldots + t_{i_g} + t_{j_1} + \ldots + t_{j_h}) = 0$
$\Longleftrightarrow t_{i_1} + \ldots + t_{i_g} + t_{j_1} + \ldots + t_{j_h} = 2M,\ M \in \mathbb{N}$
$\Longleftrightarrow$ exact $2M$ components of $(t_{i_1}, \ldots, t_{i_g}, t_{j_1}, \ldots, t_{j_h}) \in \{0,1\}^{g+h}$ equal 1
$\Longleftrightarrow$ exact $2M$ components of $(\tilde{\tau}_{i_1}, \ldots, \tilde{\tau}_{i_g}, \tilde{\tau}_{j_1}, \ldots, \tilde{\tau}_{i_h}) \in \{-1,1\}^{g+h}$ equal -1
$\Longleftrightarrow \tilde{\tau}_{i_1} \cdot \ldots \cdot \tilde{\tau}_{i_g} \cdot \tilde{\tau}_{j_1} \cdot \ldots \cdot \tilde{\tau}_{j_h} = 1$.
b) $\Longleftrightarrow$ c): This follows from the property $\text{mod}_p(a) = \text{mod}_p(b) \Longleftrightarrow \text{mod}_p(a + (p-1)b) = 0$ for any $a, b \in \mathbb{Z}$ (see Lemma 8.5.1). □

6.4.5 Example (Continuation of Example 6.4.1)
In the incomplete 2^3-design of Example 6.4.1, it holds

$$\tilde{\tau}_3 \cdot \tilde{\tau}_1\tilde{\tau}_2 = 1 \Longleftrightarrow \text{mod}_2(\tau_1 + \tau_2 + \tau_3) = 0 \Longleftrightarrow \tau_3 = \text{mod}_2(\tau_2 + \tau_3)$$

Denoting the factors 1,2,3 with A, B, C this can be written shortly as $C = AB$. Similarly, it holds $B = AC$, $A = BC$, $I = ABC$. Another abbreviation is $3 = 12$, $2 = 13$, $1 = 23$, $I = 123$.

$I = 123$ or $I = ABC$ or $\text{mod}_2(\tau_1 + \tau_2 + \tau_3) = 0$, respectively, is called the defining equation for the design of Example 6.4.1. The defining equation can be also given by $\text{mod}_2(b^\top t) = 0$ where $t = (\tau_1, \tau_2, \tau_3)^\top$ and $b = (1,1,1)^\top$ is called defining pencil. Since there is only one defining equation, this design is called a 2^{3-1} design.

At this 2^{3-1} design, $(\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2)$ are identifiable in the model $\mu(t) = \tilde{\mu}_0 + \tilde{\mu}_1\tilde{\tau}_1 + \tilde{\mu}_2\tilde{\tau}_2 + \tilde{\mu}_3\tilde{\tau}_3 + \tilde{\mu}_{12}\tilde{\tau}_1\tilde{\tau}_2$. Such design and such model make in particular sence if C is a block factor such that the effect of C itself is not of interest.

6.4.6 Theorem
Let be $B = (b_1, \ldots, b_M)^\top \in \{0,1\}^{M\times K}$ with linearly independent rows $b_1^\top, \ldots, b_M^\top$. Then the design given by

$$d(B) = \{t = (\tau_1, \ldots, \tau_K)^\top \in \{0,1\}^K;\ \text{mod}_2(Bt) = 0_M\}$$

has 2^{K-M} elements.
Thereby $\text{mod}_2(Bt) = (\text{mod}_2(b_1^\top t), \ldots, \text{mod}_2(b_M^\top t))^\top \in \{0,1\}^M$.

Proof. Without loss of generality assume $B = (B_1|B_2)$ where $B_1 \in \{0,1\}^{M\times M}$ is nonsingular and $B_2 \in \{0,1\}^{M\times(K-M)}$. Then

$$Bt = (B_1, B_2)\begin{pmatrix} \tau_1 \\ \vdots \\ \tau_M \\ \tau_{M+1} \\ \vdots \\ \tau_K \end{pmatrix} = B_1 \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_M \end{pmatrix} + B_2 \begin{pmatrix} \tau_{M+1} \\ \vdots \\ \tau_K \end{pmatrix} = 0_M$$

is equivalent with

$$\begin{pmatrix} \tau_1 \\ \vdots \\ \tau_M \end{pmatrix} = -B_1^{-1}B_2 \begin{pmatrix} \tau_{M+1} \\ \vdots \\ \tau_K \end{pmatrix}.$$

Hence $(\tau_1, \dots, \tau_M)^\top$ is uniquely determined by $(\tau_{M+1}, \dots, \tau_K) \in \{0,1\}^{K-M}$. Since there are 2^{K-M} choices for $(\tau_{M+1}, \dots, \tau_K)$, the assertion is proved. □

6.4.7 Remark
There are two possibilities to construct a 2^{K-M}-design $d(B)$:
a) Create a full 2^K design and drop all rows which do not satisfy $\text{mod}_2(b_m^\top t) = 0$ for all $m = 1, \dots, M$.
b) Let $B = (B_1|B_2)$ where $B_1 = I_{M\times M} \in \{0,1\}^{M\times M}$ is the identity matrix. Set $L = K - M$ and create a full 2^L-design d_L. For all $t_l^{(L)} \in d_L$ set $t_l = (\text{mod}_2(B_2^\top t_l^{(L)}), t_l^{(L)})^\top$, $l = 1, \dots, L$. Then $d(B) = (t_1, \dots, t_L)$.

6.4.8 Definition
Let be $B = (b_1, \dots, b_M)^\top \in \{0,1\}^{M\times K}$ *with linearly independent rows* $b_1^\top, \dots, b_M^\top$. *Then the design given by*

$$d(B) = \{t = (\tau_1, \dots, \tau_K)^\top \in \{0,1\}^K;\ mod_2(Bt) = 0_M\}$$

is called **regular fraction of the** 2^K**-design** *or shortly* 2^{K-M}**-design**.
The rows $b_1^\top, \dots, b_M^\top$ *are called* **defining pencils** *of the design* $d(B)$.

The R function `conf.design` of the R-package `conf.design` provides fractional factorial designs via vectors (pencils) $b_1, \dots, b_M$ where only the part of b_m, $m = 1, \dots, M$, is used which concern those factors which are not block factors and each pencil b_m corresponds to a block factor B_m. Hence there must be M block factors $B_1, \dots, B_M$ and $L = K - M$ factors $F_1, \dots, F_L$ of interest. If the defining pencils have the form

$$\begin{aligned}
b_1 &= (1, 0, \dots, 0, b_{1(M+1)}, b_{1(M+2)}, \dots, b_{1K})^\top \ \ (B_1 = F_1^{b_{1(M+1)}} F_2^{b_{1(M+2)}} \dots F_L^{b_{1K}}) \qquad (6.2)\\
b_2 &= (0, 1, \dots, 0, b_{2(M+1)}, b_{2(M+2)}, \dots, b_{2K})^\top \ \ (B_2 = F_1^{b_{2(M+1)}} F_2^{b_{2(M+2)}} \dots F_L^{b_{2K}})\\
&\vdots\\
b_M &= (0, 0, \dots, 1, b_{M(M+1)}, b_{M(M+2)}, \dots, b_{MK})^\top \ \ (B_M = F_1^{b_{M(M+1)}} F_2^{b_{M(M+2)}} \dots F_L^{b_{MK}})
\end{aligned}$$

then only $(b_{1(M+1)}, b_{1(M+2)}, \dots, b_{1K})$, ..., $(b_{M(M+1)}, b_{M(M+2)}, \dots, b_{MK})$ are used in the call of `conf.design`. Thereby F_k^0 means that the factor F_k does not appear.

6.4.9 Example
2^{3-1}-design:

```
> library(conf.design)
> conf.design(c(1,1),p=2,
               treatment.names=LETTERS[1:2])
  Blocks A B
1        0 0 0
2        0 1 1
3        1 1 0
4        1 0 1
```

This is the same design as in Example 6.4.1. Thereby, it holds:

$M = 1$, $L = K - M = 2$ so that $K = 3$, $(b_{12}, b_{13}) = (1, 1)$ so that $b_1 = (1,\ 1, 1)^\top$, which means $B_1 = F_1\, F_2$ in the notation of (6.2) or `Block` $= AB = A^1 B^1$ with `Block` $= B_1$, $A = F_1$, $B = F_2$.

6.4.10 Example
2^{5-2}-design:

```
> conf.design(rbind(c(0,1,1), c(1,1,0)),p=2,
          treatment.names=LETTERS[1:3])
  Blocks A B C
1      00 0 0 0
2      00 1 1 1
3      01 1 0 0
4      01 0 1 1
5      10 1 1 0
6      10 0 0 1
7      11 0 1 0
8      11 1 0 1
```

Thereby it holds:
$M = 2$, $L = K - M = 3$ so that $K = 5$,

$$b_1 = (1, 0,\ 0, 1, 1)^\top,\ b_2 = (0, 1,\ 1, 1, 0)^\top,\ B = \begin{pmatrix} 1 & 0 & & 0 & 1 & 1 \\ 0 & 1 & & 1 & 1 & 0 \end{pmatrix}.$$

Hence $B_1 = F_1^0 F_2^1 F_3^1$ and $B_2 = F_1^1 F_2^1 F_3^0$ in the notation of (6.2) or `Block1`$= BC = A^0 B^1 C^1$, `Block2`$= AB = A^1 B^1 C^0$.

6.4.11 Example
2^{6-3}-design:

```
> conf.design(rbind(c(0,1,1),c(1,1,0),c(1,1,1)),
          p=2,treatment.names=LETTERS[1:3])
  Blocks A B C
1     000 0 0 0
2     001 1 1 1
3     010 0 1 1
4     011 1 0 0
5     100 1 1 0
6     101 0 0 1
7     110 1 0 1
8     111 0 1 0
```

But this not a good design. Why?

6.4.12 Definition
The **alias set** *of a design d is*

$$A(d) := \{b \in \{0, 1\}^K;\ mod_2(b^\top t) = 0 \text{ for all } t \in d\}.$$

All vectors $b \in A(d)$ are called **defining pencils** *of the design d.*

6.4.13 Lemma
Let be $b_1^\top, \ldots, b_M^\top$ the rows of $B \in \{0,1\}^{M \times K}$ with full rank. The 2^{K-M} design $d(B)$ satisfies

$$A(d(B)) = \mathcal{R}(B) := \left\{b \in \{0,1\}^K;\ b = mod_2(\lambda_1 b_1 + \ldots + \lambda_M b_M),\ (\lambda_1, \ldots, \lambda_M) \in \{0,1\}^M\right\}$$

where $\mathcal{R}(B)$ is the row space of B modulo 2 and

$$mod_2(\lambda_1 b_1 + \ldots + \lambda_M b_M) = \begin{pmatrix} mod_2(\lambda_1 b_{11} + \ldots + \lambda_M b_{M1}) \\ \vdots \\ mod_2(\lambda_1 b_{1K} + \ldots + \lambda_M b_{MK}) \end{pmatrix}.$$

Proof. Lemma 8.5.2 provides for $b \in \mathcal{R}(B)$

$$\text{mod}_2(b^\top t) = \text{mod}_2(\lambda_1 b_1^\top t + \ldots + \lambda_M b_M^\top t)) = \text{mod}_2(\lambda_1 \text{mod}_2(b_1^\top t) + \ldots + \lambda_M \text{mod}_2(b_M^\top t))) = 0$$

for all $t \in d(B)$ so that $b \in A(d(B))$. Assume there is a $b \in A(d(B))$ with $b \notin \mathcal{R}(B)$. Then $b^\top$ is linearly independent of the rows of B and can be added to B. Then Theorem 6.4.6 provides that $d(B)$ has $2^{K-(M+1)}$ design points which is a contradiction. Hence any $b \in A(d(B))$ must satisfy $b \in \mathcal{R}(B)$. □

The R-function `conf.set` of the R-package `conf.design` provides the alias set $A(d(B))$. However the output shows only the first l components of the vectors (pencils) b of $A(d(B))$, i.e. the components which concerns the interesting factors and not the block factors. Therefore it is important that there is no b of $A(d(B))$ which contains only one 1 in the first l components since this would mean that there is an interesting factor which has a main effect which is confounded with the block factors.

6.4.14 Example (Continuation of Example 6.4.11)
The alias set of the 2^{6-3}-design given by $(0,1,1),\ (1,1,0),\ (1,1,1)$ are obtained by

```
>conf.set(rbind(c(0,1,1),c(1,1,0),c(1,1,1)),p=2)
     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    1    1    0
[3,]    1    0    1
[4,]    1    1    1
[5,]    1    0    0
[6,]    0    0    1
[7,]    0    1    0
```

Since this alias set includes $(1,0,0),\ (0,1,0),\ (0,0,1)$, each main effect of the three interesting factors are confounded with the block factors. Hence the main effects of the three interesting factors are not identifiable at this 2^{6-3}-design. Note that the obvious member of $A(d(B))$, namely 0_k is not shown in the output.

A better 2^{6-3}-design is given by $(0,1,1),\ (1,1,0),\ (1,0,1)$

```
>conf.set(rbind(c(0,1,1),c(1,1,0),c(1,0,1)),p=2)
     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    1    1    0
```

```
[3,]    1    0    1
```

Here we see that `conf.set` provides only the first l components of the $b \in A(d(B))$ which are different. Hence only three vectors are given although $A(d(B))$ has $2^3 = 8$ elements, namely

$$\begin{aligned}
\text{generating pencil: } b_1 &= (1,0,0,\ \ 0,1,1)^\top & \widehat{=} \quad & B_1 = F_2F_3,\\
\text{generating pencil: } b_2 &= (0,1,0,\ \ 1,1,0)^\top & \widehat{=} \quad & B_2 = F_1F_2,\\
\text{generating pencil: } b_3 &= (0,0,1,\ \ 1,0,1)^\top & \widehat{=} \quad & B_3 = F_1F_3,\\
\text{mod}_2(b_1+b_2) &= (1,1,0,\ \ 1,0,1)^\top & \widehat{=} \quad & B_1B_2 = F_1F_3,\\
\text{mod}_2(b_2+b_3) &= (0,1,1,\ \ 0,1,1)^\top & \widehat{=} \quad & B_2B_3 = F_2F_3,\\
\text{mod}_2(b_1+b_3) &= (1,0,1,\ \ 1,1,0)^\top & \widehat{=} \quad & B_1B_3 = F_1F_2,\\
\text{mod}_2(b_1+b_2+b_3) &= (1,1,1,\ \ 0,0,0)^\top & \widehat{=} \quad & B_1B_2B_3 = I,\\
\text{mod}_2(0\cdot b_1+0\cdot b_2+0\cdot b_3) &= (0,0,0,\ \ 0,0,0)^\top. & &
\end{aligned}$$

Hence only interactions of the 2 interesting factors are confounded with the block factors.

Define $\mu_b = \tilde{\mu}_{i_1 i_2 \dots i_g}$ and $\tilde{t}^b = \tilde{\tau}_{i_1}\tilde{\tau}_{i_2}\cdot \ldots \cdot \tilde{\tau}_{i_g}$ for $b = (b_1,\dots,b_K)^\top$ with $b_i = 1$ if $i \in \{i_1,\dots,i_g\}$ and $b_i = 0$ else. Hence b denotes the factorial effect $F_1^{b_1}F_2^{b_2}\dots F_K^{b_K}$. Thereby again F_k^0 means that the factor F_k does not appear. Also, we set $\mu_b = \tilde{\mu}_0$ and $\tilde{t}^b = 1$ if $b = 0_K$.

6.4.15 Lemma
μ_b and μ_{b^*} confounded at d (briefly $F_1^{b_1}F_2^{b_2}\dots F_K^{b_K} = F_1^{b_1^*}F_2^{b_2^*}\dots F_K^{b_K^*}$) if and only if

$$b - b^* \in A(d).$$

Proof. $b - b^* \in A(d)$ means $\text{mod}_2((b-b^*)^\top t) = 0$ for all $t \in d$. According to Theorem 6.4.4 this is equivalent with $\tilde{t}^b = \tilde{t}^{b^*}$ or with the fact that μ_b and μ_{b^*} are confounded in the 2^K-design $d(B)$. □

6.4.16 Definition
Let be $b^* \in \{0,1\}^K$. Then

$$A(b^*,d) = \left\{a \in \{0,1\}^K;\ a = mod_2(b^*+b),\ b \in A(d)\right\}$$

is called the **alias set** *of the pencil b^* at the design d.*

In particular we have

$$A(0_K,d) = A(d).$$

6.4.17 Lemma
If $B \in \{0,1\}^{M\times K}$ is of full rank, then there are 2^{K-M} different alias sets $A(b^*,d(B))$ and each alias set has 2^M elements.

Proof. Lemma 6.4.13 provides $A(d(B)) = \mathcal{R}(B)$. There are 2^M choices for $\lambda_1, \ldots, \lambda_M \in \{0,1\}$ in the set $\mathcal{R}(B)$. Hence $A(d(B)) = \mathcal{R}(B)$ has 2^M elements and thus each alias set has 2^M elements. Since the alias sets provide equivalence classes for the equivalence relation $b - b^* \in A(d(B))$, the alias sets are pairwise disjoint. Since there are at all 2^K vectors (pencils) in $\{0,1\}^K$, there are $2^K/2^M = 2^{K-M}$ alias sets. □

6.4.18 Remark
Lemma 6.4.17 yields the proposal that only aspects of the form

$$\sum_{a \in A(b,d(B))} \mu_a$$

are identifiable at $d(B)$ in the saturated model since there are 2^{K-M} different aspects of this type and the 2^{K-M}-design $d(B)$ has 2^{K-M} design points. If a submodel is considered then only those μ_b are identifiable for which all other μ_a with $a \in A(b, d(B))$ are not included in the model. For a more rigorous proof of these proposals see Section 6.6.

6.4.19 Definition
A 2^{K-M}-design $d(B)$ has resolution R if each $0_K \neq b \in \mathcal{R}(B)$ has at least R ones.

6.4.20 Lemma
If $d(B)$ has resolution R then interactions of order r are only confounded with interactions of order $R - r$.
In particular we have:
If $d(B)$ has resolution III then main effects can be confounded only with interactions of order two, but they are not confounded with other main effects.
If $d(B)$ has resolution IV then main effects can be confounded only with interactions of order three, but they are not confounded with interactions of order two.
If $d(B)$ has resolution V then no interaction of order two is confounded with another interaction of order two.

Proof. Lemma 6.4.15 provides that μ_b and μ_{b^*} are confounded if and only if $b - b^* \in A(d)$. Hence $\text{mod}_2(b - b^*)$ must consists of at least R ones. Hence if b has r ones then b^* must contain at least $R - r$ ones at different positions. □

6.4.21 Example (Continuation of Example 6.4.14)
The second 2^{6-3}-design in Example 6.4.14 given by $(0,1,1)$, $(1,1,0)$, $(1,0,1)$, i.e. generated by $(1,0,0,\ \ 0,1,1)$, $(0,1,0,\ \ 1,1,0)$, $(0,0,1,\ \ 1,0,1)$, has resolution III since all defining pencils $b \neq 0_k$ consist of at least 3 ones.

6.5 Optimality of fractional factorial design

Consider here only a model with main effects of K factors, i.e.

$$\mu(t) = x(t)^\top \theta \ \text{ with } x(t) = (1, \tilde{\tau}_1, \ldots, \tilde{\tau}_K)^\top,\ \theta = (\tilde{\mu}_0, \tilde{\mu}_1, \ldots, \tilde{\mu}_K)^\top. \tag{6.3}$$

Then

$$X_d = (1_N, V_d)$$

where $V_d \in \mathbb{R}^{N\times K}$ is the design matrix of the main effects.

6.5.1 Theorem
Let be $\lambda(\theta) = (\tilde{\mu}_1, \ldots, \tilde{\mu}_K)^\top$. Then a design d_* is D_λ-optimal in $\{0,1\}^{N\times K}$ for the model (6.3) if

$$V_{d_*}^\top V_{d_*} = N\, I_{K\times K} \ \text{ and } \ 1_N^\top V_{d_*} = 0_K^\top.$$

Proof. For any design $d \in \{0,1\}^{N\times K}$, we have according to Lemma 2.4.5

$$I_\lambda(d) = (L\, I_\theta(d)^-\, L^\top)^{-1} = \left(I_{K\times K} \left[V_d^\top \omega^\perp(1_N) V_d^\top \right]^- I_{K\times K} \right)^{-1} = V_d^\top V_d - \frac{1}{N} V_d^\top 1_N 1_N^\top V_d.$$

Since $\mathrm{tr}(V_d^\top 1_N 1_N^\top V_d) \geq 0$, it holds

$$\mathrm{tr}(I_\lambda(d)) = \mathrm{tr}(V_d^\top V_d) - \mathrm{tr}\left(\frac{1}{N} V_d^\top 1_N 1_N^\top V_d \right) \leq \mathrm{tr}(V_d^\top V_d) = \sum_{k=1}^K \sum_{n=1}^N b_{nk}^2 = N\,K$$

where $b_{nk} \in \{-1, 1\}$ is the (n,k)'th element of V_d. Moreover,

$$c := \mathrm{tr}(I_\lambda(d)) = \sum_{k=1}^K \xi_k \leq NK$$

and

$$\det(I_\lambda(d)) = \prod_{k=1}^K \xi_k,$$

where $\xi_1, \ldots, \xi_K$ are the eigenvalues of $I_\lambda(d)$. As in the proof of Theorem 5.6.4, $\det(I_\lambda(d)) = \prod_{k=1}^K \xi_k$ is maximized under the side condition $\sum_{k=1}^K \xi_k = c$ if and only if $\xi_k = \frac{c}{K}$ for $k = 1, \ldots, K$. Hence

$$\det(I_\lambda(d)) \leq \left(\frac{c}{K}\right)^K \leq \left(\frac{NK}{K}\right)^K \leq N^K.$$

If d_* satisfies $V_{d_*}^\top V_{d_*} = N\, I_{K\times K}$ and $1_N^\top V_{d_*} = 0_K$, then

$$\det(I_\lambda(d_*)) = \det(V_{d_*}^\top V_{d_*}) = \det(N\, I_{K\times K}) = N^K$$

so that

$$\det(I_\lambda(d_*)) \geq \det(I_\lambda(d))$$

for all $d \in \{0,1\}^{N\times K}$. Hence d_* is D_λ-optimal in $\{0,1\}^{N\times K}$. □

6.5.2 Lemma
If d is a 2^{K-M}-design with resolution II then $1_N^\top V_d = 0_K^\top$.

Proof. The proof is an exercise. □

6.5.3 Lemma
If d is a 2^{K-M}-design with resolution III then $V_d^\top V_d = N\, I_{K\times K}$.

Proof. Let be $i,j \in \{1,\ldots,K\}$ with $i \neq j$ arbitrary. Since $d = \{t^{(1)},\ldots,t^{(N)}\}$ with $N = 2^{K-M}$ is a 2^{K-M}-design, without loss of generality there exists $B = (B_1|B_2) \in \mathbb{R}^{M\times K}$ with $d = d(B)$ and

$$\begin{pmatrix} \tau_1 \\ \vdots \\ \tau_M \end{pmatrix} = -B_1^{-1}B_2 \begin{pmatrix} \tau_{M+1} \\ \vdots \\ \tau_K \end{pmatrix}.$$

for all $t \in d$ (see the proof of Theorem 6.4.6). Since there are no restriction for the choice of $\tau_{M+1},\ldots,\tau_K$, the design given by

$$d_L = \begin{pmatrix} \tau_{M+1}^{(1)} & \tau_{M+2}^{(1)} & \cdots & \tau_K^{(1)} \\ \tau_{M+1}^{(2)} & \tau_{M+2}^{(2)} & \cdots & \tau_K^{(2)} \\ \vdots & \vdots & & \vdots \\ \tau_{M+1}^{(N)} & \tau_{M+2}^{(N)} & \cdots & \tau_K^{(N)} \end{pmatrix}$$

with $N = 2^{K-M}$ is a regular 2^L-design with $L = K - M$. According to Corollary 6.3.5, for all $0_K \neq \alpha \in \{0,1\}^L$ there exists $\{n_1,\ldots,n_{N/2}\} \subset \{1,\ldots,N\}$ with

$$\begin{aligned} (\tilde{\tau}_{M+1}^{(n)})^{\alpha_1} \cdot (\tilde{\tau}_{M+2}^{(n)})^{\alpha_2} \cdot \ldots \cdot (\tilde{\tau}_K^{(n)})^{\alpha_L} = -1 \;&\Leftrightarrow\; \mathrm{mod}_2(\alpha_1\tau_{M+1}^{(n)} + \ldots + \alpha_L\tau_K^{(n)}) = 1 \\ &\quad\;\text{for}\quad n \in \{n_1,\ldots,n_{N/2}\}, \\ (\tilde{\tau}_{M+1}^{(n)})^{\alpha_1} \cdot (\tilde{\tau}_{M+2}^{(n)})^{\alpha_2} \cdot \ldots \cdot (\tilde{\tau}_K^{(n)})^{\alpha_L} = 1 \;&\Leftrightarrow\; \mathrm{mod}_2(\alpha_1\tau_{M+1}^{(n)} + \ldots + \alpha_L\tau_K^{(n)}) = 0 \\ &\quad\;\text{for}\quad n \notin \{n_1,\ldots,n_{N/2}\}. \end{aligned} \tag{6.4}$$

Case 1: $i,j \in \{M+1,\ldots,K\}$.
The property (6.4) implies at once $\sum_{n=1}^N \tilde{\tau}_i^{(n)} \cdot \tilde{\tau}_j^{(n)} = 0$. Hence the i'th and the j'th columns of V_d are orthogonal.

Case 2: $i,j \in \{1,\ldots,M\}$.
In this case, $\tau_i = \sum_{k=M+1}^K \beta_k\tau_k$ and $\tau_j = \sum_{k=M+1}^K \gamma_k\tau_k$ for some $\beta,\gamma \in \mathbb{R}^{K-M}$ for all $t \in d$. If

β would be equal to γ, then $\tau_i = \tau_j$ for all $t \in d$ so that a $b \in \mathcal{R}(B)$ with only two ones would exists which is a contradiction to the assumption that d has resolution III. Hence there exists $k \in \{M+1, \ldots, K\}$ with $\beta_k \neq \gamma_k$. Without loss of generality, let $k = M+1$, $\beta_{M+1} = 0$, and $\gamma_{M+1} = 1$. Then $\tau_i = \sum_{k=M+2}^K \beta_k \tau_k$ and $\tau_j = \tau_{M+1} + \sum_{k=M+2}^K \gamma_k \tau_k$ so that

$$\tau_j - \tau_i = \tau_{M+1} + \sum_{k=M+2}^K (\gamma_k - \beta_k)\tau_k.$$

Hence $\alpha \in \{0,1\}^L$ with $\alpha_1 = 1$, $\alpha_k = \gamma_{M+k} - \beta_{M+k}$ for $k = 2, \ldots, L$ satisfies $\alpha \neq 0_L$ so that according to (6.4) there exists $\{n_1, \ldots, n_{N/2}\} \subset \{1, \ldots, N\}$ with

$$\begin{aligned} \mathrm{mod}_2(\tau_j^{(n)} - \tau_i^{(n)}) &= \mathrm{mod}_2\left(\tau_{M+1}^{(n)} + \sum_{k=M+2}^K (\gamma_k - \beta_k)\tau_k^{(n)}\right) = 1 \quad \text{for} \quad n \in \{n_1, \ldots, n_{N/2}\}, \\ \mathrm{mod}_2(\tau_j^{(n)} - \tau_i^{(n)}) &= \mathrm{mod}_2\left(\tau_{M+1}^{(n)} + \sum_{k=M+2}^K (\gamma_k - \beta_k)\tau_k^{(n)}\right) = 0 \quad \text{for} \quad n \notin \{n_1, \ldots, n_{N/2}\}. \end{aligned}$$

Hence $\tilde{\tau}_i^{(n)} = \tilde{\tau}_j^{(n)}$ for $n \in \{n_1, \ldots, n_{N/2}\}$ and $\tilde{\tau}_i^{(n)} \neq \tilde{\tau}_j^{(n)}$ for $n \notin \{n_1, \ldots, n_{N/2}\}$ so that $\sum_{n=1}^N \tilde{\tau}_i^{(n)} \cdot \tilde{\tau}_j^{(n)} = 0$.

Case 3: $j \in \{1, \ldots, M\}$ and $i \in \{M+1, \ldots, K\}$.
Here we have $\tau_j = \sum_{k=M+1}^K \beta_k \tau_k$ for some $\beta \in \mathbb{R}^{K-M}$ for all $t \in d$. Since d is of resolution III, there exists $k \in \{M+1, \ldots, K\} \setminus \{i\}$ with $\beta_k = 1$. Without loss of generality, let be $k = M+1$ and $i = M+2$. Then $\tau_j = \tau_{M+1} + \beta_{M+2}\tau_{M+2} + \sum_{k=M+3}^K \beta_k \tau_k$ so that

$$\tau_j - \tau_i = \tau_j - \tau_{M+2} = \tau_{M+1} + (\beta_{M+2} - 1)\tau_{M+2} + \sum_{k=M+3}^K \beta_k \tau_k.$$

Hence the assertion follows as in Case 2. □

6.5.4 Corollary
If d_* is a 2^{K-M}-design with resolution III then it is D_λ-optimal in $\{0,1\}^{N\times K}$ for the model (6.3).

Proof. The assertion follows at once from Theorem 6.5.1, Lemma 6.5.2 and Lemma 6.5.3. □

6.6 Fractional factorial designs for factors with p levels

The construction of fractional factorial designs for factors with more than two levels is similar to that for factors with two levels. Only the interpretation of factorial effects is more complicated. Since the construction bases on the theory of Galois fields, the number p of levels must be prime. Regular fraction of the complete p^K-design are defined as for $p = 2$:

6.6.1 Definition
Let be p prime and $B = (b_1, \ldots, b_M)^\top \in \{0, 1, \ldots, p-1\}^{M\times K}$ with linearly independent rows $b_1^\top, \ldots, b_M^\top$. Then the design given by

$$d(B) = \{t = (\tau_1, \ldots, \tau_K)^\top \in \{0, 1, \ldots, p-1\}^K;\ mod_p(Bt) = 0_M\}$$

is called **regular fraction of the p^K-design** *or shortly* **p^{K-M}-design**.
Thereby we have $mod_p(Bt) = (mod_p(b_1^\top t), \ldots, mod_p(b_M^\top t))^\top \in \{0, 1, \ldots, p-1\}^M$.
The rows $b_1^\top, \ldots, b_M^\top$ are called **defining pencils** *of the design $d(B)$.*

6.6.2 Remark
A p^{K-M}-design has p^{K-M} design points. This can be proved as for $p = 2$ (see Theorem 6.4.6).

6.6.1 Example
A 3^{5-2}-design given by $(0, 1, 1),\ (1, 2, 0)$ is generated as follows:

```
> conf.design(rbind(c(0,1,1),c(1,2,0)),p=3, treatment.names=LETTERS[1:3])
```

	Blocks	A	B	C		Blocks	A	B	C		Blocks	A	B	C
1	00	0	0	0	11	10	0	0	1	21	20	0	0	2
2	00	2	2	1	12	10	2	2	2	22	21	0	2	0
3	00	1	1	2	13	11	2	1	0	23	21	2	1	1
4	01	1	0	0	14	11	1	0	1	24	21	1	0	2
5	01	0	2	1	15	11	0	2	2	25	22	1	2	0
6	01	2	1	2	16	12	0	1	0	26	22	0	1	1
7	02	2	0	0	17	12	2	0	1	27	22	2	0	2
8	02	1	2	1	18	12	1	2	2					
9	02	0	1	2	19	20	2	2	0					
10	10	1	1	0	20	20	1	1	1					

Here the defining pencils are $(2, 0,\ \ 0, 1, 1)$ and $(0, 2,\ \ 1, 2, 0)$ with corresponding defining equations `Block1`$= BC = B^1C^1$, `Block2`$= AB^2$.
Thereby `Block1`$= BC = B^1C^1$ means $\tau_1 = \text{mod}_3(\tau_4 + \tau_5)$ for all $t \in d$ or equivalently, according to Lemma 8.5.1, $0 = \text{mod}_3(2 \cdot \tau_1 + \tau_4 + \tau_5)$ for all $t \in d$.
Similarly, `Block2`$= AB^2$ means $\tau_2 = \text{mod}_3(\tau_3 + 2 \cdot \tau_4)$ for all $t \in d$, or equivalently $0 = \text{mod}_3(2 \cdot \tau_2 + \tau_3 + 2 \cdot \tau_4)$ für alle $t \in d$.
The alias set $A(d)$ is here produced by:

```
> conf.set(rbind(c(0,1,1),c(1,2,0)),p=3)
```

```
     [,1] [,2] [,3]
[1,]    0    1    1
[2,]    1    2    0
[3,]    1    0    1
[4,]    1    1    2
```

To understand what are factorial effects in such models, recall from Chapter 5 that only treatment contrast are identifiable in a block model. If there are more than two factors and also interactions are considered then also only contrasts can be used. Even all factorial effects for $p = 2$ can be interpreted as contrasts as the following example shows.

6.6.3 Example
It holds in the 2^2-model $\mu(t) = \tilde{\mu}_0 + \tilde{\mu}_1\tilde{\tau}_1 + \tilde{\mu}_2\tilde{\tau}_2 + \tilde{\mu}_{12}\tilde{\tau}_1\tilde{\tau}_2$ and we have for $\mu = (\mu(0,0), \mu(1,0), \mu(0,1), \mu(1,1))^\top$

$$\begin{aligned}
\tilde{\mu}_0 &= \frac{1}{4}(\mu((0,0)) + \mu((1,0)) + \mu((0,1)) + \mu((1,1))) = \frac{1}{4}(1,1,1,1)\cdot\mu = L_0^\top\mu,\\
\tilde{\mu}_1 &= \frac{1}{4}(\mu((0,0)) - \mu((1,0)) + \mu((0,1)) - \mu((1,1))) = \frac{1}{4}(1,-1,1,-1)\cdot\mu = L_1^\top\mu,\\
\tilde{\mu}_2 &= \frac{1}{4}(\mu((0,0)) + \mu((1,0)) - \mu((0,1)) - \mu((1,1))) = \frac{1}{4}(1,1,-1,-1)\cdot\mu = L_2^\top\mu,\\
\tilde{\mu}_{12} &= \frac{1}{4}(\mu((0,0)) - \mu((1,0)) - \mu((0,1)) + \mu((1,1))) = \frac{1}{4}(1,-1,-1,1)\cdot\mu = L_{12}^\top\mu,
\end{aligned}$$

where

$$\begin{aligned}
L_0 &= (l_0((0,0)), l_0((1,0)), l_0((0,1)), l_0((1,1)))^\top = \left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)^\top,\\
L_1 &= (l_1((0,0)), l_1((1,0)), l_1((0,1)), l_1((1,1)))^\top = \left(\frac{1}{4},-\frac{1}{4},\frac{1}{4},-\frac{1}{4}\right)^\top,\\
L_2 &= (l_2((0,0)), l_2((1,0)), l_2((0,1)), l_2((1,1)))^\top = \left(\frac{1}{4},\frac{1}{4},-\frac{1}{4},-\frac{1}{4}\right)^\top,\\
L_{12} &= (l_{12}((0,0)), l_{12}((1,0)), l_{12}((0,1)), l_{12}((1,1)))^\top = \left(\frac{1}{4},-\frac{1}{4},-\frac{1}{4},\frac{1}{4}\right)^\top.
\end{aligned}$$

This means that $\tilde{\mu}_0$, $\tilde{\mu}_1$, $\tilde{\mu}_1$, $\tilde{\mu}_{12}$ are linear aspects of $\mu((0,0))$, $\mu((1,0))$, $\mu((0,1))$, $\mu((1,1))$ and vice versa. Moreover,

$$\begin{aligned}
\tilde{\mu}_1 &= \frac{1}{4}(\mu((0,0)) - \mu((1,0)) + \mu((0,1)) - \mu((1,1))),\\
&= \frac{1}{4}(\mu((0,0)) + \mu((0,1))) - \frac{1}{4}(\mu((1,0)) + \mu((1,1))),\\
\tilde{\mu}_2 &= \frac{1}{4}(\mu((0,0)) + \mu((1,0)) - \mu((0,1)) - \mu((1,1))),\\
&= \frac{1}{4}(\mu((0,0)) + \mu((1,0))) - \frac{1}{4}(\mu((0,1)) + \mu((1,1))),\\
\tilde{\mu}_{12} &= \frac{1}{4}(\mu((0,0)) - \mu((1,0)) - \mu((0,1)) + \mu((1,1))),
\end{aligned}$$

and

$$L_1^\top 1_4 = L_2^\top 1_4 = L_{12}^\top 1_4 = 0.$$

Hence $\tilde{\mu}_1$, $\tilde{\mu}_1$, $\tilde{\mu}_{12}$ are treatment contrasts.

6.6.4 Definition
Let be $F_1, \dots, F_K$ K factors with levels $p_1, \dots, p_K$.
Then the linear aspect of $\mu = (\mu((\tau_1, \dots, \tau_K)))_{\tau_i \in \{0,1,\dots,p_i-1\}, i=1,\dots,K}$ given by

$$\sum_{\tau_1=0}^{p_1-1} \cdots \sum_{\tau_K=0}^{p_K-1} l((\tau_1, \dots, \tau_K))\, \mu((\tau_1, \dots, \tau_K))$$

is called treatment contrast if
(i) there exists $(\tau_1, \dots, \tau_K)$ with $l((\tau_1, \dots, \tau_K)) \neq 0$, and
(ii)

$$\sum_{\tau_1=0}^{p_1-1} \cdots \sum_{\tau_K=0}^{p_K-1} l((\tau_1, \dots, \tau_K)) = 0.$$

6.6.5 Definition
Let be $F_1, \dots, F_K$ K factors with levels $p_1, \dots, p_K$.
Then the treatment contrast

$$\sum_{\tau_1=0}^{p_1-1} \cdots \sum_{\tau_K=0}^{p_K-1} l((\tau_1, \dots, \tau_K))\, \mu((\tau_1, \dots, \tau_K))$$

belongs to the factorial effect $F_{i_1} \dots F_{i_g}$ if
(i) $l((\tau_1, \dots, \tau_K))$ depends only on $\tau_{i_1}, \dots, \tau_{i_g}$,
(ii) for all $\gamma \in \{i_1, \dots, i_g\}$, we have

$$\sum_{\tau_\gamma=0}^{p_\gamma-1} l(\tau_1, \dots, \tau_\gamma, \dots, \tau_K) = 0$$

for all $\tau_1, \dots, \tau_{\gamma-1}, \tau_{\gamma+1}, \dots, \tau_K$.

6.6.6 Example (Continuation of Example 6.6.3)
In the case of two factors F_1 and F_2, each with two levels, we have:

$\tilde{\mu}_1$ is treatment contrast belonging to F_1,
since $l_1((0, \tau_2)) = \frac{1}{4}$, $l_1((1, \tau_2)) = -\frac{1}{4}$, and $l_1((0, \tau_2)) + l_1((1, \tau_2)) = 0$ for $\tau_2 = 0, 1$.

$\tilde{\mu}_2$ is treatment contrast belonging to F_2,
since $l_2((\tau_1, 0)) = \frac{1}{4}$, $l_2((\tau_1, 1)) = -\frac{1}{4}$, and $l_2((\tau_1, 0)) + l_2((\tau_1, 1)) = 0$ for $\tau_1 = 0, 1$.

$\tilde{\mu}_{12}$ is treatment contrast belonging to $F_1 F_2$,
since $l_{12}((0,0)) = \frac{1}{4}$, $l_{12}((0,1)) = -\frac{1}{4}$, $l_{12}((1,0)) = -\frac{1}{4}$, $l_{12}((1,1)) = \frac{1}{4}$, and
$l_{12}((0,0)) + l_{12}((0,1)) = 0$, $l_{12}((0,0)) + l_{12}((1,0)) = 0$,
$l_{12}((1,1)) + l_{12}((0,1)) = 0$, $l_{12}((1,1)) + l_{12}((1,0)) = 0$.

Now let be $F_1, \ldots, F_K$ K factors, each with p levels, and $\mu = (\mu(t))_{t\in\{0,1,\ldots,p-1\}^K}$ the vector which provides the effects at the experimental conditions $t \in \{0, 1, \ldots, p-1\}^K$.

6.6.7 Definition
A treatment contrast $L\mu$ belongs to the pencil $b \in \{0, 1, \ldots, p-1\}^K$, if

$$L\mu = \sum_{j=0}^{p-1} l_j \left\{ \sum_{t\in V_j(b)} \mu(t) \right\}$$

with

$$V_j(b) = \{t \in \{0, 1, \ldots, p-1\}^K;\ mod_p(b^\top t) = j\},\quad j = 0, 1, \ldots, p-1$$

and

$$0_p \neq (l_0, l_1, \ldots, l_{p-1})^\top \in \mathbb{R}^p \ \text{ satisfy } \sum_{j=0}^{p-1} l_j = 0.$$

6.6.8 Lemma
Let be $b = (b_1, \ldots, b_K) \in \{0, 1, \ldots, p-1\}^K$ a pencil with $b_i \neq 0$ for $i \in \{i_1, \ldots, i_g\}$ and $b_i = 0$ else. Then the treatment contrast $L\mu$ belonging to b belongs to the factorial effect $F_{i_1} \ldots F_{i_g}$.

Proof. Since the components $b_i = 0$ do not influence the determination of $V_0, V_1, \ldots, V_{p-1}$, also the determination of $l_0, l_1, \ldots, l_{p-1}$ does not depend on these components of t, i.e. it holds

$$l((\tau_1, \ldots, \tau_K)) = \bar{l}(\tau_{i_1}, \ldots, \tau_{i_g}) = l_j \text{ if } \mathrm{mod}_p(\sum_{\gamma=1}^{g} b_{i_\gamma}\tau_{i_\gamma}) = j.$$

Hence we have for $\gamma = 1, \ldots, g$ and fixed $\tau_{i_1}, \ldots, \tau_{i_{\gamma-1}}, \tau_{i_{\gamma+1}}, \ldots, \tau_{i_g}$

$$\sum_{\tau_{i_\gamma}=0}^{p-1} \bar{l}(\tau_{i_1}, \ldots, \tau_{i_g}) = l_0 + l_1 + \ldots + l_{p-1} = 0.$$

□

6.6.9 Lemma
A treatment contrast $L\mu = \sum_{t\in\{0,\ldots,p-1\}^K} l(t)\mu(t)$ is not identifiable at $d \subset \{0, \ldots, p-1\}^K$ if there exists $c \in \mathbb{R}$ with $l(t) = c$ for all $t \in d$.

Proof. If $x(t) = (1\!\!1_{\{\tilde{t}\}}(t))_{\tilde{t}\in\{0,1,\ldots,p-1\}^K}$, where $1\!\!1_A$ denotes the indicator function of the set A, then we have $\mu(t) = x(t)^\top\mu$ and the design matrix has the form $X_d = (x(t))_{t\in d}$. Since $L\mu$ is a treatment contrast, it holds $\sum_{t\in\{0,\ldots,p-1\}^K} l(t) = 0$ and there exists a $t \in \{0, 1, \ldots, p-1\}^K$ with $l(t) \neq 0$. Then there exists also a $t_0 \in \{0, 1, \ldots, p-1\}^K$ with $0 \neq l(t_0) \neq c$. The property $l(t_0) \neq c$ implies $t_0 \notin d$. Let be $\mu(t_0) = a \neq 0$ and $\mu(t) = 0$ else. Then it holds

$$X_d\mu = 0 \text{ and } L\mu = l(t_0)a \neq 0,$$

so that $L\mu$ is not identifiable at d.

□

6.6.10 Theorem
If a pencil b_* is an element of the alias set of the design d, i.e.

$$b_* \in A(d) = \{b \in \{0,1,\ldots,p-1\}^n;\ mod_p(b^\top t) = 0 \text{ für alle } t \in d\},$$

then the treatment contrast $L\mu$ belonging to b_* is not identifiable at d.

Proof. It holds $t \in V_0(b)$ for all $t \in d$. Hence we have $l(t) = l_0$ for all $t \in d$ so that Lemma 6.6.9 provides the assertion. □

Define again the alias set of b_* at a design d as

$$A(b_*, d) := \{b \in \{0,1,\ldots,p_1\}^K;\ b_* - b \in A(d)\}$$

and

$$V_j(b, B) := \left\{t \in \{0,1,\ldots,p_1\}^K;\ \text{mod}_p(b^\top t) = j \ \text{ and } \ \text{mod}_p(Bt) = 0_M\right\},$$

where $j \in \{0,1,\ldots,p-1\}$ and $B \in \{0,1,\ldots,p-1\}^{M\times K}$. Then we have $V_j(b,B) = V_j(b) \cap d(B)$.

6.6.11 Lemma
Let be p prime, $B = (b_1,\ldots,b_M)^\top \in \{0,1,\ldots,p-1\}^{M\times K}$ with linearly independent rows $b_1^\top,\ldots,b_M^\top$, and $b \in \{0,1,\ldots,p_1\}^K$. Then for all $t \in \{0,1,\ldots,p_1\}^K$ and $j \in \{0,1,\ldots,p-1\}$

$$\sum_{a\in A(b,d(B))} 1\!\!1_{\{mod_p(a^\top t)=j\}}(a) = \begin{cases} p^M & \text{if } t \in V_j(b,B), \\ 0 & \text{if } t \in d(B) \setminus V_j(b,B), \\ p^{M-1} & \text{if } t \notin d(B). \end{cases}$$

Proof. As for $p = 2$, any $a \in A(b, d(B))$ can be written as $a = b + B^\top\lambda$ with $\lambda = (\lambda_1,\ldots,\lambda_M)^\top \in \{0,1,\ldots,p-1\}^M$. Then we have

$$\sum_{a\in A(b,d(B))} 1\!\!1_{\{\text{mod}_p(a^\top t)=j\}}(a) = \sharp\left\{\lambda \in \{0,1,\ldots,p-1\}^M;\ \text{mod}_p(b^\top t + \lambda^\top Bt) = j\right\},$$

where $\sharp A$ denotes the cardinality of the set A. Then using Lemma 8.5.2 we obtain:
a) If $t \in V_j(b,B)$, then it holds

$$\text{mod}_p(b^\top t + \lambda Bt) = \text{mod}_p(\text{mod}_p(b^\top t) + \text{mod}_p(\lambda^\top \text{mod}_p(Bt))) = j + 0 = j$$

for all $\lambda \in \{0,1,\ldots,p-1\}^M$. Hence $\sum_{a\in A(b,d(B))} 1\!\!1_{\{\text{mod}_p(a^\top t)=j\}}(a) = p^M$.
b) If $t \in d(B) \setminus V_j(b,B)$, then $\text{mod}_p(b^\top t) \neq j$ and $\text{mod}_p(\lambda^\top \text{mod}_p(Bt)) = 0$ such that

$$\text{mod}_p(b^\top t + \lambda Bt) = \text{mod}_p(\text{mod}_p(b^\top t) + \text{mod}_p(\lambda^\top \text{mod}_p(Bt))) \neq j$$

for all $\lambda \in \{0,1,\ldots,p-1\}^M$. This implies $\sum_{a\in A(b,d(B))} 1\!\!1_{\{\text{mod}_p(a^\top t)=j\}}(a) = 0$.
c) If $t \notin d(B)$ then $\text{mod}_p(b^\top t) \neq 0_M$ and we have $\text{mod}_p(b^\top t + \lambda^\top Bt) = j$ if and only if $\text{mod}_p(\lambda^\top Bt) = \text{mod}_p(j - b^\top t)$. Since $v = (v_1,\ldots,v_M)^\top := \text{mod}_p(B^\top t) \neq 0_M$ we can assume without loss of generality $v_1 \neq 0$. Then $\text{mod}_p(\lambda_1 v_1) + \text{mod}_p(\lambda_2 v_2 + \ldots + \lambda_M v_M) = \text{mod}_p(\lambda^\top v) = \text{mod}_p(j - b^\top t)$ implies $\lambda_1 = \text{mod}_p(v_1^{-1}(j - b^\top t - \lambda_2 v_2 - \ldots \lambda_M v_M))$. Hence there are p^{M-1} free choices for $\lambda_1,\ldots,\lambda_M \in \{0,1,\ldots,p-1\}$ so that $\sum_{a\in A(b,d(B))} 1\!\!1_{\{\text{mod}_p(a^\top t)=j\}}(a) = p^{M-1}$. □

6.6.12 Lemma
Let be p prime, $B = (b_1, \ldots, b_M)^\top \in \{0, 1, \ldots, p-1\}^{M \times K}$ with linearly independent rows $b_1^\top, \ldots, b_M^\top$, and $\sum_{j=0}^{p-1} l_j = 0$ so that

$$L_a\mu = \sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(a)} \mu(t) \right\}$$

are treatment contrasts belonging to pencils $a \in A(d(B))$. Then

$$\sum_{a \in A(b,d(B))} \left[\sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(a)} \mu(t) \right\} \right] = p^M \sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(b,B)} \mu(t) \right\}. \tag{6.5}$$

Proof. Lemma 6.6.11 provides

$$\begin{aligned}
&\sum_{a \in A(b,d(B))} \left[\sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(a)} \mu(t) \right\} \right] \\
&= \sum_{a \in A(b,d(B))} \left[\sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in \{0,1,\ldots,p-1\}^K} 1\!\!1_{\{\mathrm{mod}_p(a^\top t)=j\}}(a)\, \mu(t) \right\} \right] \\
&= \sum_{j=0}^{p-1} l_j \left[\sum_{t \in \{0,1,\ldots,p-1\}^K} \left\{ \sum_{a \in A(b,d(B))} 1\!\!1_{\{\mathrm{mod}_p(a^\top t)=j\}}(a) \right\} \mu(t) \right] \\
&= \sum_{j=0}^{p-1} l_j \left[p^M \sum_{t \in V_j(b,B)} \mu(t) + p^{M-1} \sum_{t \notin d(B)} \mu(t) \right] \\
&= p^M \sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(b,B)} \mu(t) \right\} + p^{M-1} \left\{ \sum_{t \notin d(B)} \mu(t) \right\} \sum_{j=0}^{p-1} l_j \\
&= p^M \sum_{j=0}^{p-1} l_j \left\{ \sum_{t \in V_j(b,B)} \mu(t) \right\}
\end{aligned}$$

since $\sum_{j=0}^{p-1} l_j = 0$. □

6.6.13 Remark
Since the right-hand side of (6.5) in Lemma 6.6.12 depends only on treatment effects included in $d(B)$ it is identifiable at $d(B)$. This implies also that the linear aspect $\sum_{a \in A(b,d(B))} L_a\mu$ as defined in Lemma 6.6.12 is identifiable at $d(B)$. But $L_a\mu$ and $L_{\tilde{a}}\mu$ with different $a, \tilde{a} \in A(d(B))$ are confounded at $d(B)$.

6.6.14 Theorem
Let be p prime, $B = (b_1, \ldots, b_M)^\top \in \{0, 1, \ldots, p-1\}^{M \times K}$ with linearly independent rows $b_1^\top, \ldots, b_M^\top$. A factorial effect belonging to a pencil $b \notin A(d(B))$ is identifiable at $d(B)$ if and only if all factorial effects belonging to $a \in A(b, d(B))$ with $b \neq a$ are not included in the model.

Proof. This is a direct consequence of Lemma 6.6.12. See also Lemma 6.6.8 for the relation between factorial effects and pencils. □

Chapter 7

Optimal designs for nonlinear problems

In this chpater we regard the problem of estimating a nonlinear aspect of a linear model and of estimating the whole parameter θ of a nonlinear model. Thereby a nonlinear model is given by

$$y(t) = \mu(t, \theta) + z,$$

where $\mu(t, \theta)$ cannot be expressed in the form $\mu(t, \theta) = x(t)^\top \theta$ as this is the case for linear models. Similarly, an aspect $\varphi(\theta)$ of θ is nonlinear, if it cannot be expressed as $\varphi(\theta) = L\theta$. Identifiability can be extended to nonlinear aspects and nonlinear models as follows:

7.0.1 Definition
An aspect $\varphi : \Theta \to \mathbb{R}^S$ is identifiable at d if for all $\theta_1,\ \theta_2 \in \Theta$ we have the implication

$$\mu(t, \theta_1) = \mu(t, \theta_2) \text{ for all } t \in d \Longrightarrow \varphi(\theta_1) = \varphi(\theta_2).$$

We start with the problem of estimating a nonlinear aspect in a linear model and continue with the problem of estimating the whole parameter θ of a nonlinear model.

In both situations the information matrix can be only approximated or given asymptotically. Since we will use the asymptotic approach, we add here the sample size N to the estimators and designs.

7.1 Nonlinear aspects in linear models

At first note that identifiability of a nonlinear aspect $\varphi(\theta)$ in a linear model $y(t) = x(t)^\top\theta + z$ can be characterized similarly as for linear aspects.

7.1.1 Lemma
The aspect φ is identifiable at d if and only if there exists $\varphi^* : \mathbb{R}^N \to \mathbb{R}^S$ with $\varphi(\theta) = \varphi^*(X_d\theta)$ for all $\theta \in \Theta$.

Moreover the identifiability of a nonlinear aspect $\varphi(\theta)$ can be attributed to linear identifiability. For that we need the partial derivatives with respect to θ of φ which depend on θ because of the nonlinearity of φ. Then we have the following matrix of derivatives

$$\dot{\varphi}_\theta := \frac{\partial}{\partial\tilde{\theta}}\varphi(\tilde{\theta})/_{\tilde{\theta}=\theta} \in \mathbb{R}^{S\times R}.$$

With these derivatives the identifiability of φ at a set $d \subset \mathcal{T}$ can be characterized.

7.1.2 Lemma
Let Θ be a convex and open subset of $\mathbb{R}^R$ and φ continuously differentiable on Θ. Then the aspect φ is identifiable at d if and only if for all $\theta \in \Theta$ the linear aspect φ_θ with $\varphi_\theta(\tilde{\theta}) = \dot{\varphi}_\theta\tilde{\theta}$ is identifiable at d.

Proof. Without loss of generality, we can assume that φ is a one-dimensional function, i.e. $S = 1$. Let $\theta_1,\ \theta_2 \in \Theta$ be any parameters so that $x(t)^\top\theta_1 = x(t)^\top\theta_2$ for all $t \in d$. The mean value theorem and the convexity of Θ provide at once that the identifiability of all linear aspects φ_θ implies the identifiability of φ. To show the converse implication assume that $\varphi(\theta_1) = \varphi(\theta_2)$ and $\dot{\varphi}_\theta\theta_1 \neq \dot{\varphi}_\theta\theta_2$ for some $\theta \in \Theta$. Then for all $\lambda > 0$ we have $\dot{\varphi}_\theta\lambda(\theta_1 - \theta_2) \neq 0$. Because Θ is open there exists λ_0 with $\theta + \lambda(\theta_1 - \theta_2) \in \Theta$ for $\lambda \in [0, \lambda_0]$. The identifiability of φ and $x(t)^\top(\theta + \lambda(\theta_1 - \theta_2)) = x(t)^\top\theta$ implies $\varphi(\theta + \lambda(\theta_1 - \theta_2)) = \varphi(\theta)$ for all $\lambda \in [0, \lambda_0]$. Hence the derivative of φ at θ in direction of $\theta_1 - \theta_2$ must be equal to 0 which is a contradiction. □

As the Gauss-Markov estimator for a linear aspect $\lambda(\theta) = L\theta$ is based on the least squares estimator by $\hat{\lambda}_N(y_N, d_N) = \lambda(\hat{\theta}_N(y_N, d_N)) = L\hat{\theta}_N(y_N, d_N))$ we similarly can base an estimator for a nonlinear aspect on the least squares estimator.

7.1.3 Definition (Gauss-Markov estimator for a nonlinear aspect)
An estimator $\hat{\varphi}_N : \mathbb{R}^N \times \mathcal{T}^N \to \mathbb{R}^S$ is a Gauss-Markov estimator for the nonlinear aspect $\varphi : \Theta \to \mathbb{R}^S$ and denoted by $\hat{\varphi}_N$ if $\hat{\varphi}_N(y_N, d_N) = \varphi(\hat{\theta}_N(y_N, d_N))$ for all $y_N \in \mathbb{R}^N$ and $d_N \in \mathcal{T}^N$, where $\hat{\theta}_N$ is the least squares estimator for θ.

If $(d_N)_{N\in\mathbb{N}}$ is converging to a design measure δ or d_N is the realization of a random design D_N with distribution $\bigotimes_{n=1}^N \delta$, then under the classical assumptions for the error distribution the Gauss-Markov estimator is asymptotically normally distributed.

7.1.4 Theorem
If Θ is a convex and open subset of $\mathbb{R}^R$, φ is continuously differentiable on Θ and identifiable at δ with finite support and $(d_N)_{N\in\mathbb{N}}$ are deterministic designs converging to δ or $(D_N)_{N\in\mathbb{N}}$ are random designs given by δ, then

$$\mathcal{L}(\sqrt{N}(\hat{\varphi}_N - \varphi(\theta))|P_\theta^N) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0,\, \sigma^2\, \dot{\varphi}_\theta\, \mathcal{I}(\delta)^- \dot{\varphi}_\theta^\top)$$

for all $\theta \in \Theta$.

Proof. The identifiability of φ at the finite set $\mathcal{D} = \text{supp}(\delta)$ provides according to Lemma 7.1.1 that $\varphi(\theta) = \varphi^*(X_d\theta)$ for all $\theta \in \Theta$. Because the Gauss-Markov estimator $X_\mathcal{D}\hat{\theta}_N$ is a consistent estimator of $X_\mathcal{D}\theta$ and Θ is open, for every $\theta \in \Theta$ and $\epsilon > 0$ there exists N_ϵ such that a version of $\hat{\theta}_N$ is lying in Θ with probability greater than $1-\epsilon$ for $N \geq N_\epsilon$. Note that if θ is not identifiable at δ, then $\hat{\theta}_N$ is not unique and the different versions of $\hat{\theta}_N$ are given by the different versions of the generalized inverse of $X_{d_N}^\top X_{d_N}$. If $\hat{\theta}_N \in \mathcal{B}$, then according to Lemma 7.1.1 we have $\hat{\varphi}_N = \varphi^*(X_\mathcal{D}\hat{\theta}_N)$. Moreover, the differentiability of φ provides also the differentiability of φ^* with $\dot{\varphi}_\theta = \dot{\varphi}_\theta^* X_d$, where $\dot{\varphi}_\theta^* := \frac{\partial}{\partial\eta}\varphi^*(\eta)/_{\eta = X_d\theta}$. Then the differentiability of φ^* and the consistency and asymptotic normality of the Gauss-Markov estimator $X_d\hat{\theta}_N$ provides

$$\begin{aligned}
&\sqrt{N}(\hat{\varphi}_N - \varphi(\theta)) - \sqrt{N}(\dot{\varphi}_\theta\hat{\theta}_N - \dot{\varphi}_\theta\theta)\\
&= \sqrt{N}(\varphi^*(X_d\hat{\theta}_N) - \varphi^*(X_d\theta)) - \sqrt{N}(\dot{\varphi}_\theta\hat{\theta}_N - \dot{\varphi}_\theta\theta)\\
&= \sqrt{N}|X_d\hat{\theta}_N - X_d\theta| \frac{\varphi^*(X_d\hat{\theta}_N) - \varphi^*(X_d\theta) - \dot{\varphi}_\theta^*(X_d\hat{\theta}_N - X_d\theta)}{|X_d\hat{\theta}_N - X_d\theta|}\\
&\to 0
\end{aligned}$$

in probability for $(P_\theta^N)_{N\in\mathbb{N}}$ so that the assertion follows from the asymptotic normality of the Gauss-Markov estimators $\dot{\varphi}_\theta\hat{\theta}_N$. □

The estimator based on the least squares estimator is the estimator which minimize uniformly the asymptotic covariance matrix for all $\theta \in \Theta$ within all estimators of the form $\varphi(\hat{\theta}_N)$ where $\hat{\theta}_N$ is some estimator for θ.

For deriving optimal designs it is as for linear aspect in general not possible to find a design which minimizes the asymptotic covariance matrix uniformly. But additionally, because the covariance matrix depends on θ, it is in general also not possible to find a design which minimizes some one-dimensional function of the covariance matrix simultaneously for all $\theta \in \Theta$. Therefore only *locally optimal* designs can be derived. Because the asymptotic covariance matrix of the Gauss-Markov estimator for a nonlinear aspect at θ is equal to the covariance matrix of the Gauss-Markov estimator for φ_θ given by $\varphi_\theta(\tilde{\theta}) = \dot{\varphi}_\theta\tilde{\theta}$, the locally optimal designs are optimal designs for estimation of φ_θ. Let be

$$\Delta(\varphi) = \{\delta \in \Delta_0;\ \varphi \text{ is identifiable at } \delta\}$$

and Δ_0 the set of all design measures on the design region $\mathcal{T}$.

7.1.5 Definition (Local optimality for a nonlinear aspect)
a) $\delta_{\theta,D}$ is locally D-optimal for φ in Δ at θ if

$$\delta_{\theta,D} \in \arg\min\{\det(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top);\ \delta \in \Delta \cap \Delta(\varphi_\theta)\}.$$

b) $\delta_{\theta,A}$ is locally A-optimal for φ in Δ at θ if

$$\delta_{\theta,A} \in \arg\min\{\mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top);\ \delta \in \Delta \cap \Delta(\varphi_\theta)\}.$$

Several strategies were involved to overcome the θ-dependence of the designs. One strategy is to use a prior π for the unknown parameter vector θ. This leads to so-called Bayesian designs.

7.1.6 Definition (Bayesian optimality for a nonlinear aspect)
a) $\delta_{B,D}$ is Bayesian D-optimal for φ in Δ with respect to π if

$$\delta_{B,D} \in \arg\min\left\{\int \det(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top)\ \pi(d\theta);\ \delta \in \Delta \cap \Delta(\varphi_\theta)\right\}.$$

b) $\delta_{B,A}$ is Bayesian A-optimal for φ in Δ with respect to π if

$$\delta_{B,A} \in \arg\min\left\{\int \mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top)\ \pi(d\theta);\ \delta \in \Delta \cap \Delta(\varphi_\theta)\right\}.$$

Another strategy is to use the locally optimal designs for efficiency comparisons. Here only the locally A-optimal designs will be of interest so that we define *relative efficiency* of a design δ at θ as

$$E(\delta,\theta) := \frac{\mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta_{\theta,A})^- \dot\varphi_\theta^\top)}{\mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top)} \in [0,1].$$

A design δ_M which maximizes the minimum relative efficiency is called a *maximin efficient design*.

7.1.7 Definition (Maximin efficient design)
δ_M is maximin efficient for φ in Δ if

$$\delta_M \in \arg\max\left\{\min_{\theta\in\Theta} \frac{\mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta_{\theta,A})^- \dot\varphi_\theta^\top)}{\mathrm{tr}(\dot\varphi_\theta\, \mathcal{I}(\delta)^- \dot\varphi_\theta^\top)};\ \delta \in \Delta \cap \Delta(\varphi)\right\}.$$

Maximin efficient designs can be derived in particular if we restrict ourselves to designs with a support included in a finite set $\mathcal{D}$, i.e. to

$$\Delta_{\mathcal{D}}^* := \{\delta \in \Delta_0;\ \mathrm{supp}(\delta) \subset \mathcal{D}\},$$

because in many situations all locally A-optimal designs have a support which is included in a set $\mathcal{D} = \{\tau_1, ..., \tau_I\}$. Usually in this situations, the regressors $x(\tau_1), ..., x(\tau_I)$ are linearly independent so that $\mathcal{D} = \{\tau_1, ..., \tau_I\}$ is a minimum support of a generalized design. Then the locally A-optimal designs $\delta_{\theta,A}$ in $\Delta^*_{\mathcal{D}} \cap \Delta(\varphi_\theta)$ have a very simple form.

7.1.8 Lemma
If $\mathcal{D} = \{\tau_1, ..., \tau_I\}$ and $x(\tau_1), ..., x(\tau_I)$ are linearly independent, then the locally A-optimal design $\delta_{\theta,A}$ in $\Delta^*_{\mathcal{D}} \cap \Delta(\varphi_\theta)$ is given by

$$\delta_{\theta,A}(\{t\}) = \frac{|\dot{\varphi}_\theta (X_{\mathcal{D}}^\top X_{\mathcal{D}})^- x(t)|}{\sum_{\tau \in \mathcal{D}} |\dot{\varphi}_\theta (X_{\mathcal{D}}^\top X_{\mathcal{D}})^- x(\tau)|} \qquad \text{for } t \in \mathcal{D},$$

where $X_{\mathcal{D}} = (x(\tau_1), ..., x(\tau_I))^\top$.

Proof. See Müller (1997), p. 21. □

7.1.9 Theorem
Let φ be identifiable at $\mathcal{D} = \{\tau_1, ..., \tau_I\}$, $x(\tau_1), ..., x(\tau_I)$ are linearly independent and for all $t \in \mathcal{D}$

$$\max\{\delta_{\theta,A}(\{t\});\ \theta \in \Theta\} = 1.$$

Then δ_M is maximin efficient for φ in $\Delta^*_{\mathcal{D}}$ if and only if $\delta_M = \frac{1}{I}\sum_{t\in\mathcal{D}} e_t$, and the maximin efficiency is equal to $\frac{1}{I}$.

Proof. See Müller (1997), p. 21. □

7.1.1 Example (Linear calibration)
In the linear calibration problem a linear regression model is assumed, where the observation at t_{nN} is given by

$$Y_{nN} = \theta_0 + \theta_1 t_{nN} + Z_{nN} = x(t_{nN})^\top \theta + Z_{nN}$$

for $n = 1, ..., N$, where $\theta = (\theta_0, \theta_1)^\top \in \Theta = \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ and $x(t) = (1, t)^\top \in \mathbb{R}^2$. Assume that the experimental region is $\mathcal{T} = [-1, 1]$. Then the interesting aspect of θ is that experimental condition t_y which would provide some given value y if there are no errors Z_{nN}, i.e. $\theta_0 + \theta_1 t_y = y$. But this means that the interesting aspect is $\varphi(\theta) = t_y = \frac{y-\theta_0}{\theta_1}$ which is a nonlinear aspect of θ. Then we have $\dot{\varphi}_\theta = \frac{-1}{\theta_1^2}(\theta_1, y - \theta_0)^\top \in \mathbb{R}^2$. For deriving the locally A-optimal design with Lemma 7.1.8 in $\Delta^*_{\mathcal{D}}$ with $\mathcal{D} = \{-1, 1\}$ note

$$(X_{\mathcal{D}}^\top X_{\mathcal{D}})^- x(t) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ t \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ t \end{pmatrix}$$

so that

$$\dot{\varphi}_\theta\,(X_{\mathcal{D}}^\top X_{\mathcal{D}})^- x(t) = \frac{-1}{2\,\theta_1^2}\,(\theta_1 + (y-\theta_0)t) = \begin{cases} \frac{-1}{2\,\theta_1^2}\,(\theta_1 - y + \theta_0) & \text{for } t=-1 \\ \frac{-1}{2\,\theta_1^2}\,(\theta_1 + y - \theta_0) & \text{for } t=1. \end{cases}$$

Hence the locally A-optimal designs in $\Delta_{\mathcal{D}}^*$ are given by

$$\delta_{\theta,A} = \frac{1}{|\theta_1 - y + \theta_0| + |\theta_1 + y - \theta_0|}\,(|\theta_1 - y + \theta_0|\,e_{-1} + |\theta_1 + y - \theta_0|\,e_1)\,, \tag{7.1}$$

where e_a denotes here again the Dirac measure on a. It can be even proved that these designs are also locally A-optimal within all designs on $\mathcal{T} = [-1,1]$, where $\varphi(\theta)$ is identifiable (see Exercise 7.1.2). Note that these designs are c-optimal with $c = \dot{\varphi}_\theta = \frac{-1}{\theta_1^2}(\theta_1, y-\theta_0)^\top$. Hence they are not only locally A-optimal but also locally D-optimal.

Moreover, we have $\delta_{\theta,A}(\{-1\}) = 1$ for $\theta_0 = y + \theta_1$ and $\delta_{\theta,A}(\{1\}) = 1$ for $\theta_0 = y - \theta_1$ so that $\max\{\delta_{\theta,A}(\{t\});\ \theta \in \Theta\} = 1$ for $t \in \{-1,1\} = \mathcal{D}$. Hence according to Theorem 7.1.9 the maximin efficient design in Δ_d^* is $\delta_M = \frac{1}{2}(e_{-1} + e_1)$, and the maximin efficiency is $\frac{1}{2}$. □

7.1.2 Exercise
Prove for the calibration problem of Example 7.1.1 that the design in (7.1) is locally A- and D-optimal in the set of all generalized designs on $\mathcal{T} = [-1,1]$ at which $\varphi(\theta) = t_y = \frac{y-\theta_0}{\theta_1}$ is identifiable.

7.2 Nonlinear models

For estimating θ in a nonlinear model given by the nonlinear response function $\mu(t,\theta)$ we also can use a least squares estimator.

7.2.1 Definition (Least squares estimator for a nonlinear model)
An estimator $\hat{\theta}_N : \mathbb{R}^N \times \mathcal{T}^N \to \mathbb{R}^R$ is a least squares estimator for θ and denoted by $\hat{\theta}_N$ if

$$\hat{\theta}_N(y_N, d_N) \in \arg\min \left\{ \sum_{n=1}^{N} (y_{nN} - \mu(t_{nN}, \theta))^2;\ \theta \in \Theta \right\}$$

for all $y_N \in \mathbb{R}^N$ and $d_N \in \mathcal{T}^N$.

Under some regularity conditions this least squares estimator is consistent and in particular asymptotically normally distributed and asymptotically optimal (see e.g. Seber and Wild (1989) pp. 563, Bunke and Bunke (1989) pp. 30).

7.2.2 Theorem
Under regularity conditions, the least squares estimator is asymptotically normally distributed at θ, i.e.

$$\mathcal{L}(\sqrt{N}(\hat{\theta}_N - \theta)|P_\theta^N) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0,\ \sigma^2\, \mathcal{I}_\theta(\delta)^{-1}),$$

where

$$\mathcal{I}_\theta(\delta) := \int \dot{\mu}(t,\theta)\, \dot{\mu}(t,\theta)^\top \delta(dt)$$

and

$$\dot{\mu}(t,\theta) := \left(\frac{\partial}{\partial \tilde{\theta}} \mu(t,\tilde{\theta})/_{\tilde{\theta}=\theta} \right)^\top \in \mathbb{R}^R.$$

Because the asymptotic covariance matrix of a least squares estimator in a nonlinear model depend on θ only locally optimal designs as for estimating nonlinear aspects can be defined.

7.2.3 Definition (Local optimality for a nonlinear model)
a) $\delta_{\theta,D}$ is locally D-optimal for φ in Δ at θ if

$$\delta_{\theta,D} \in \arg\min\{\det(\mathcal{I}_\theta(\delta)^{-1});\ \delta \in \Delta \cap \Delta(\theta)\}.$$

b) $\delta_{\theta,A}$ is locally A-optimal for φ in Δ at θ if

$$\delta_{\theta,A} \in \arg\min\{\mathrm{tr}(\mathcal{I}_\theta(\delta)^{-1});\ \delta \in \Delta \cap \Delta(\theta)\}.$$

As for nonlinear aspects, also Bayesian optimal designs and maximin efficient designs can be defined.

Again, A-optimal designs within $\Delta^*_{\mathcal{D}} \cap \Delta(\theta)$ can be easily derived if $\mathcal{D} = \{\tau_1, ..., \tau_I\}$ and $\dot{\mu}(\tau_1, \theta), \ldots, \dot{\mu}(\tau_I, \theta)$ are linearly independent, i.e. $\Delta^*_{\mathcal{D}}$ is the set of generalized designs with minimum support.

7.2.4 Lemma
If $\mathcal{D} = \{\tau_1, ..., \tau_I\}$ and $\dot{\mu}(\tau_1, \theta), \ldots, \dot{\mu}(\tau_I, \theta)$ are linearly independent, then the locally A-optimal design $\delta_{\theta,A}$ in $\Delta^*_{\mathcal{D}} \cap \Delta(\theta)$ is given by

$$\delta_{\theta,A}(\{t\}) = \frac{|(X_{\mathcal{D},\theta}^\top X_{\mathcal{D},\theta})^{-1}\, \dot{\mu}(t, \theta)|}{\sum_{\tau \in \mathcal{D}} |(X_{\mathcal{D},\theta}^\top X_{\mathcal{D},\theta})^{-1}\, \dot{\mu}(\tau, \theta)|}$$

for $t \in \mathcal{D}$, where

$$X_{\mathcal{D},\theta} = (\dot{\mu}(\tau_1, \theta), \ldots, \dot{\mu}(\tau_I, \theta))^\top.$$

Proof. See e.g. Müller (1997), p. 23. □

7.2.1 Example (Generalized linear model)
We regard a generalized linear model, where the observation at t_{nN} is given by

$$Y_{nN} = e^{\theta_0 + \theta_1 t_{nN}} + Z_{nN} = \mu(t_{nN}, \theta) + Z_{nN},$$

for $n = 1, ..., N$, where $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$, $\mu(t, \theta) = e^{\theta_0 + \theta_1 t}$ and $t \in \mathcal{T} = [0, 1]$. If we restrict ourselves to designs with a support included in $\mathcal{D} = \{0, 1\}$, then we have

$$X_{\mathcal{D},\theta} = \begin{pmatrix} e^{\theta_0} & 0 \\ e^{\theta_0+\theta_1} & e^{\theta_0+\theta_1} \end{pmatrix},$$

where

$$\dot{\mu}(t, \theta) = e^{\theta_0 + \theta_1 t} \begin{pmatrix} 1 \\ t \end{pmatrix}.$$

Hence according to Lemma 7.2.4 the locally A-optimal design in $\Delta^*_{\mathcal{D}}$ is given by

$$\delta_{\theta,A} = \frac{1}{\sqrt{2} + e^{-\theta_1}}\, (\sqrt{2}\, e_0 + e^{-\theta_1}\, e_1).\Box$$

7.2.2 Exercise
The kinetics of many enzymes is often modeled with the Michaelis-Menten model where the substrate concentration S_t at time t is given by

$$\frac{v_{max} S_t}{k + S_t}$$

where v_{max} is the maximum substrate concentration and k is a specific reaction parameter. This leads to a nonlinear model which has a response function of the form

$$\mu(t,\theta) = \frac{\theta_0\, t}{\theta_1 + t}$$

with $\theta \in [0,\infty]^2$ and $t \in [0,\infty]$. Determine for this response function the locally A-optimal designs for the design region $\mathcal{T} = \{0,1\}$.

Chapter 8

Appendix

8.1 G-inverses and projection matrices

8.1.1 Definition (g-inverse)
$A^- \in \mathbb{R}^{m\times n}$ is called g-inverse (generalized inverse) of $A \in \mathbb{R}^{n\times m}$ if and only if $A\,A^-\,A = A$.

If A is a regular matrix, then $A^- = A^{-1}$ and A^{-1} is the only g-inverse. Hence the g-inverse is really a generalization of the inverse for regular matrices. But note that, if A is not a regular matrix, then the g-inverse of A is not unique, i.e. there are several g-inverses. For the g-inverse of $A^\top A$ the following lemma holds.

8.1.2 Lemma
Let $(A^\top A)^-$ be a g-inverse of $A^\top A$. Then it holds:
a) $((A^\top A)^-)^\top$ is g-inverse of $A^\top A$.
b) $A^\top A(A^\top A)^- A^\top = A^\top$ and $A(A^\top A)^- A^\top A = A$.
c) $A(A^\top A)^- A^\top$ is idempotent, i.e. $A(A^\top A)^- A^\top\, A(A^\top A)^- A^\top = A(A^\top A)^- A^\top$.
d) $A(A^\top A)^- A^\top$ is independent of the choice of the g-inverse.
e) $A(A^\top A)^- A^\top$ is a symmetric matrix.

Proof.
a) $A^\top A((A^\top A)^-)^\top A^\top A = (A^\top A(A^\top A)^- A^\top A)^\top = (A^\top A)^\top = A^\top A$.
b) In general, it holds: $BD^\top D = CD^\top D$ implies $BD^\top = CD^\top$. For $BD^\top D = CD^\top D$ implies

$$\begin{aligned} 0 &= (BD^\top D - CD^\top D)(B-C)^\top = (BD^\top - CD^\top)D(B-C)^\top \\ &= (BD^\top - CD^\top)\,((B-C)D^\top)^\top = (BD^\top - CD^\top)\,(BD^\top - CD^\top)^\top \end{aligned}$$

Multiplying the last expression from both sides with an arbitrary vector of appropriate dimension yields $0 = x^\top(BD^\top - CD^\top)\,(BD^\top - CD^\top)^\top x$. This means $0 = (BD^\top - CD^\top)^\top x$ for all x and therefore $0 = BD^\top - CD^\top$.
Because of the definition of the g-Inverse, it holds $A^\top A(A^\top A)^- A^\top A = A^\top A$. Setting $B =$

$A^\top A(A^\top A)^-$, $C = I$ the identity matrix, and $D = A$ provides the first part of the assertion b). The second part follows from the first part by transposing the matrices and using a).
c) follows from b).
d) Let $(A^\top A)^\sim$ be another g-inverse of $A^\top A$. Assertion b) implies $A(A^\top A)^\sim A^\top A = A = A(A^\top A)^- A^\top A$. Setting $B = A(A^\top A)^\sim$, $C = A(A^\top A)^-$ and $D = A$, then the assertion shown in b) provides $A(A^\top A)^\sim A^\top = A(A^\top A)^- A^\top$. This means that $A(A^\top A)^- A^\top$ does not depend on the choice of the g-inverse.
e) The assertion a) implies $(A(A^\top A)^- A^\top)^\top = A((A^\top A)^-)^\top A^\top = A(A^\top A)^- A^\top$. □

8.1.3 Lemma
Let be $A \in \mathbb{R}^{n\times n}$ symmetric. If $rk(A) = m \le n$, then $rk(A^-) = m$.

Proof. Let be $A = U\text{diag}(\xi_1, \ldots, \xi_n)\, U^\top$ with $UU^\top = I_{n\times n} = U^\top U$ and $\xi_1 \ge \xi_2 \ge \ldots \ge \xi_m > \xi_{m+1} = \xi_{m+2} = \ldots = \xi_n = 0$ the spectral decomposition of A. Then we have $A = U\text{diag}(\xi_1^{-1}, \ldots, \xi_m^{-1}, 0, \ldots, 0)\, U^\top = U\text{diag}(\xi_1^-, \ldots, \xi_n^-)\, U^\top$. □

8.1.4 Definition (Idempotent matrix)
A $n \times n$ matrix A is called idempotent if and only if $A\,A = A$.

8.1.5 Lemma
If A is a symmetric and idempotent matrix of rank r, then A has r eigenvalues equal to 1 and $n - r$ eigenvalues equal to 0.

Proof. With the spectral decomposition of A. See e.g. Rencher 1998, P. 414. □

8.1.6 Definition (Column space)
Let be $X \in \mathbb{R}^{N\times R}$. Then

$$C(X) := \{X\beta;\ \beta \in \mathbb{R}^R\}$$

is called the column space of X. The set $C(X)^\perp := \{v \in \mathbb{R}^N;\ v^\top X = 0_R^\top\}$ is the orthogonal complement of $C(X)$.

8.1.7 Lemma
If A, $B \in \mathbb{R}^{R\times R}$ are symmetric matrices, then it holds:

$$A \ge B \ge 0 \implies C(B) \subset C(A). \tag{8.1}$$

Proof. If $x \in C(A)^\perp$, then $0_R^\top = x^\top A$ and thus $0 = x^\top A\, x \ge x^\top B\, x$, so that $0_R^\top = x^\top B$. Hence $x \in C(A)^\perp$ implies $x \in C(B)^\perp$. Since $C(A)^\perp \subset C(B)^\perp$ implies $C(B) \subset C(A)$, the assertion (8.1) is proved. □

8.1.8 Definition (Perpendicular projection matrix)
Let be U a subspace of $\mathbb{R}^N$. $P \in \mathbb{R}^{N\times N}$ is called perpendicular projection matrix onto U if and only if

$$Pu = u \text{ for all } u \in U,$$
$$Pv = 0 \text{ for all } v \in U^\perp = \{w \in \mathbb{R}^N;\ w^\perp u \text{ for all } u \in U\}.$$

8.1.9 Lemma
a) The perpendicular projection matrix $P \in \mathbb{R}^{N\times N}$ is idempotent.
b) Every idempotent and symmetric matrix $A \in \mathbb{R}^{N\times N}$ is a perpendicular projection matrix onto $C(A)$.

Proof.
a) It holds

$$PPu = Pu = u \text{ for all } u \in U,$$
$$PPv = P0 = 0 \text{ for all } v \in U^\perp.$$

Let be $w \in \mathbb{R}^N$ arbitrary. Then there exists $u \in U$ and $v \in U^\perp$ with $w = u + v$. Then we obtain

$$PPw = PP(u+v) = Pu = u = P(u+v) = Pw$$

and thus $PP = P$.
b) For $u = A\beta$ it holds $Au = AA\beta = A\beta = u$. If

$$v \in C(A)^\perp = \{w \in \mathbb{R}^N;\ w^\top A\beta = 0 \text{ for all } \beta \in \mathbb{R}^N\},$$

then $v^\top A\beta = 0$ for all $\beta \in \mathbb{R}^N$. The symmetry of A implies $\beta^\top A\, v = \beta^\top A^\top v = 0$ for all $\beta \in \mathbb{R}^N$ and thus $Av = 0$. □

8.1.10 Lemma
Let be $X \in \mathbb{R}^{N\times R}$.
a) $X(X^\top X)^- X^\top$ is the perpendicular projection matrix onto $C(X)$.
b) $I_{N\times N} - X(X^\top X)^- X^\top$ is the perpendicular projection matrix onto $C(X)^\perp$.
($I_{N\times N}$ denotes the $N \times N$ identity matrix).

Proof.
a) Lemma 8.1.2 c) and e) and Lemma 8.1.9 imply that $X(X^\top X)^- X^\top$ is the perpendicular projection matrix onto $C(X(X^\top X)^- X^\top)$. It remains to show $C(X) = C(X(X^\top X)^- X^\top)$. It is clear that $C(X(X^\top X)^- X^\top) \subset C(X)$. For the opposite inclusion note that for any $u \in C(X)$ there exists $\beta \in \mathbb{R}^p$ with $u = X\beta$. Then Lemma 8.1.2 b) implies

$$X(X^\top X)^- X^\top u = X(X^\top X)^- X^\top X\beta = X\beta = u$$

and thus $C(X) \subset C(X(X^\top X)^- X^\top)$.
b) Part a) implies $X(X^\top X)^- X^\top u = 0$ for all $u \in C(X)^\perp = C(X(X^\top X)^- X^\top)^\perp$ and thus

$$(I_{N\times N} - X(X^\top X)^- X^\top)u = u \text{ for all } u \in C(X)^\perp.$$

If $v \in (C(X)^{\perp})^{\perp} = C(X)$, then Part a) implies

$$(I_{N\times N} - X(X^\top X)^- X^\top)v = v - v = 0.$$

□

8.1.11 Lemma
For any $X \in \mathbb{R}^{N\times R}$, it holds

a) rk$(X(X^\top X)^- X^\top) =$ rk(X).

b) tr$(X(X^\top X)^- X^\top) =$ rk(X).

c) tr$(I_{N\times N} - X(X^\top X)^- X^\top) = N -$ rk(X).

d) rk$(I_{N\times N} - X(X^\top X)^- X^\top) = N -$ rk(X).

Thereby rk(A) denotes the rank of the matrix A.

Proof.
a) It is clear that $\text{rk}(X(X^\top X)^- X^\top) \le \text{rk}(X)$ holds. Because of Lemma 8.1.2 b) also the converse inequality holds:

$$\text{rk}(X) = \text{rk}(X(X^\top X)^- X^\top X) \le \text{rk}(X(X^\top X)^- X^\top).$$

b) $X(X^\top X)^- X^\top$ is a perpendicular projection matrix according to Lemma 8.1.10 a). Hence it is idempotent according to Lemma 8.1.9 a). This means according to Lemma 8.1.5 that $X(X^\top X)^- X^\top \in \mathbb{R}^{N\times N}$ has r eigenvalues equal to 1 and $N - r$ eigenvalues equal to 0, where $r = \text{rk}(X(X^\top X)^- X^\top)$. According to a) we have $r = \text{rk}(X)$ and according to Lemma 8.4.1 b) $\text{tr}(X(X^\top X)^- X^\top) = r$ such that $\text{tr}(X(X^\top X)^- X^\top) = \text{rk}(X)$.

c) The linearity of the trace provides

$$\text{tr}(I_{N\times N} - X(X^\top X)^- X^\top) = \text{tr}(I_{N\times N}) - \text{tr}(X(X^\top X)^- X^\top) = N - \text{rk}(X).$$

d) Since $I_{N\times N} - X(X^\top X)^- X^\top$ is also a perpendicular matrix according to Lemma 8.1.10 b), its rank coincide with its trace as in b) so that the assertion follows from c). □

8.2 Kronecker products

8.2.1 Definition
Let be $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M} \in \mathbb{R}^{N\times M}$ *and* $B \in \mathbb{R}^{I\times J}$. *The Kronecker product* $A \otimes B$ *is defined as*

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B & \ldots & A_{1M}B \\ A_{21}B & A_{22}B & \ldots & A_{2M}B \\ \vdots & \vdots & & \vdots \\ A_{N1}B & A_{N2}B & \ldots & A_{NM}B \end{pmatrix} \in \mathbb{R}^{N\,I\times M\,J}.$$

8.2.2 Definition
Let be $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M} \in \mathbb{R}^{N\times M}$. *The vec operator* $vec : \mathbb{R}^{N\times M} \longrightarrow \mathbb{R}^{N\cdot M}$ *is defined as*

$$vec(A) = (A_{11}, A_{21}, \ldots, A_{N1}, A_{12}, \ldots, A_{N2}, \ldots, A_{1M}, \ldots, A_{NM})^\top.$$

Note that if A is given columnwise by $A = (A_{\bullet 1}|A_{\bullet 2}|\ldots|A_{\bullet M})$, then

$$vec(A) = \begin{pmatrix} A_{\bullet 1} \\ A_{\bullet 2} \\ \vdots \\ A_{\bullet M} \end{pmatrix}.$$

8.2.3 Lemma
If $A \in \mathbb{R}^{N\times M}$, $B \in \mathbb{R}^{I\times J}$, $C \in \mathbb{R}^{M\times L}$, $D \in \mathbb{R}^{J\times K}$, *then*

a) $(A \otimes B)\,(C \otimes D) = A\,C \otimes B\,D \in \mathbb{R}^{N\,I\times L\,K}$,

b) $(A \otimes B)^\top = A^\top \otimes B^\top$.

Proof. Exercise. □

8.3 G-Inverses of partitioned matrices and other special matrices

8.3.1 Lemma (g-inverse of partitioned matrix)
If the symmetric matrix M is a partitioned matrix given by

$$M = \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix}$$

with $A \in \mathbb{R}^{K\times K}$, $B \in \mathbb{R}^{L\times K}$, and $C \in \mathbb{R}^{L\times L}$, where $AA^-B^\top = B^\top$ and $BA^-A = B$ or $B^\top = B^\top C^-C$ and $CC^-B = B$, respectively, are satisfied, then the g-inverse of M is given by

$$M^- = \begin{pmatrix} A^- + A^-B^\top E^- B\, A^- & -A^-B^\top E^- \\ -E^- B\, A^- & E^- \end{pmatrix}, \; M^- = \begin{pmatrix} \widetilde{E}^- & -\widetilde{E}^- B^\top C^- \\ -C^- B\, \widetilde{E}^- & C^- + C^- B\, \widetilde{E}^- B^\top C^- \end{pmatrix},$$

respectively, with $E = C - B\, A^- B^\top$, $\widetilde{E} = A - B^\top C^- B$.

Proof.

$$M\, M^- = \begin{pmatrix} A\, A^- + A\, A^- B^\top E^- B\, A^- - B^\top\, E^- B\, A^- & -A\, A^- B^\top E^- + B^\top\, E^- \\ B\, A^- + B\, A^-\, B^\top E^- B\, A^- - C\, E^- B\, A^- & -B\, A^- B^\top E^- + C\, E^- \end{pmatrix}$$

and

$$\begin{aligned} M\, M^- M = \Bigg(& \begin{matrix} A\, A^- A + A\, A^- B^\top E^- B\, A^- A - B^\top\, E^- B\, A^- A - A\, A^- B^\top E^- B + B^\top\, E^- B \\ B\, A^- A + B\, A^-\, B^\top E^- B\, A^- A - C\, E^- B\, A^- A - B\, A^- B^\top E^- B + C\, E^- B \end{matrix} \\ & \begin{matrix} A\, A^- B^\top + A\, A^- B^\top E^- B\, A^- B^\top - B^\top\, E^- B\, A^- B^\top - A\, A^- B^\top E^- C + B^\top\, E^- C \\ B\, A^- B^\top + B\, A^-\, B^\top E^- B\, A^- B^\top - C\, E^- B\, A^- B^\top - B\, A^- B^\top E^- C + C\, E^- C \end{matrix} \Bigg) \\ = \Bigg(& \begin{matrix} A + B^\top E^- B - B^\top\, E^- B - B^\top E^- B + B^\top\, E^- B \\ B + B\, A^-\, B^\top E^- B - C\, E^- B - B\, A^- B^\top E^- B + C\, E^- B \end{matrix} \\ & \begin{matrix} B^\top + B^\top E^- B\, A^- B^\top - B^\top\, E^- B\, A^- B^\top - B^\top E^- C + B^\top\, E^- C \\ B\, A^- B^\top + B\, A^-\, B^\top E^- B\, A^- B^\top - C\, E^- B\, A^- B^\top - B\, A^- B^\top E^- C + C\, E^- C \end{matrix} \Bigg) \\ = & \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix} = M \end{aligned}$$

since

$$\begin{aligned} & B\, A^- B^\top + (B\, A^-\, B^\top - C) E^- B\, A^- B^\top - (B\, A^- B^\top - C) E^- C \\ &= B\, A^- B^\top + (B\, A^-\, B^\top - C) E^- (B\, A^- B^\top - C) \\ &= B\, A^- B^\top + B\, A^-\, B^\top - C = C. \end{aligned}$$

□

8.3.2 Lemma
For $A \in \mathbb{N}$ and $P_A := \left(I_{A\times A} - \frac{1}{A} 1_A 1_A^\top\right)$ it holds:

a) $$\begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix}^- = \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix},$$

b) $P_A^- = P_A$.

Proof.
a)

$$\begin{aligned} &\begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \\ = &\begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} = \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix}. \end{aligned}$$

b) Since P_A is perpendicular projection matrix onto $C(1_A)^\perp$ (see Lemma 8.1.10 b)) we have $P_A P_A^- P_A = P_A P_A P_A = P_A$. □

8.3.3 Lemma
If $A \in \mathbb{R}^{A\times A}$ is regular and $b \in \mathbb{R}^A$, then

$$\begin{aligned} (A \mp b\,b^\top)^{-1} &= A^{-1} \pm \frac{A^{-1} b\, b^\top A^{-1}}{1 \mp b^\top A^{-1} b} \quad \textit{if} \quad 1 \mp b^\top A^{-1} b \neq 0, \\ (A - b\,b^\top)^- &= A^{-1} \quad \textit{if} \quad 1 - b^\top A^{-1} b = 0. \end{aligned}$$

Proof. If $1 \mp b^\top A^{-1} b \neq 0$, then

$$\begin{aligned} &(A \mp b\,b^\top)\left(A^{-1} \pm \frac{A^{-1} b\, b^\top A^{-1}}{1 \mp b^\top A^{-1} b}\right) \\ = \;& I_{A\times A} \mp b\,b^\top A^{-1} \pm \frac{b\,b^\top A^{-1}}{1 \mp b^\top A^{-1} b} - \frac{b\,(b^\top A^{-1} b)\, b^\top A^{-1}}{1 \mp b^\top A^{-1} b} \\ = \;& I_{A\times A} \mp b\,b^\top A^{-1} \pm \frac{b\,b^\top A^{-1}(1 \mp b^\top A^{-1} b)}{1 \mp b^\top A^{-1} b} \\ = \;& I_{A\times A}. \end{aligned}$$

If $1 - b^\top A^{-1} b = 0$, then

$$\begin{aligned} &(A - b\,b^\top)(A - b\,b^\top)^-(A - b\,b^\top) = (A - b\,b^\top)A^{-1}(A - b\,b^\top) \\ = \;& (AA^{-1} - b\,b^\top A^{-1})(A - b\,b^\top) = AA^{-1}A - b\,b^\top A^{-1} A - AA^{-1} b\,b^\top + b\,b^\top A^{-1} b\,b^\top \\ = \;& A - b\,b^\top - b(1 - b^\top A^{-1} b) b^\top = A - b\,b^\top. \end{aligned}$$ □

8.3.4 Lemma
For $A, B \in \mathbb{N}$ it holds

$$\begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}^- = \begin{pmatrix} 0 & 0_A^\top & 0_B^\top \\ 0_A & \frac{1}{B}\,I_{A\times A} & 0_{A\times B} \\ 0_B & 0_{B\times A} & \frac{1}{A}\,P_B \end{pmatrix},$$

where $P_B = (I_{B\times B} - \frac{1}{B}\,1_B\,1_B^\top)$.

Proof. Since P_B is perpendicular projection matrix onto $C(1_B)^\perp$ (see Lemma 8.1.10 b)), we have $P_B 1_B = 0_B$ and $1_B^\top P_B = 0_B^\top$. This implies

$$\begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top & 0_B^\top \\ 0_A & \frac{1}{B}\,I_{A\times A} & 0_{A\times B} \\ 0_B & 0_{B\times A} & \frac{1}{A}\,P_B \end{pmatrix} \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1_A^\top & 0_B^\top \\ 0_A & I_{A\times A} & 0_{A\times B} \\ 0_B & \frac{1}{B}1_B 1_A^\top & P_B \end{pmatrix} \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}$$

$$= \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & \underbrace{\frac{A}{B}\,1_B 1_B^\top + A\,P_B}_{=A\,I_{B\times B}} \end{pmatrix}.$$

□

8.4 Further results from linear algebra

8.4.1 Lemma
Let tr(A) denote the trace of a the matrix $A \in \mathbb{R}^{N\times N}$, i.e. the sum of the diagonal elements of A. Then we have:

a) tr(AB) = tr(BA) for all matrices $A \in \mathbb{R}^{N\times M}$, $B \in \mathbb{R}^{M\times N}$.

b) If A is symmetric, then tr(A) is the sum of the eigenvalues of the matrix A.

Proof.
a) Let $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M}$ and $B = (B_{mn})_{m=1,\ldots,M,n=1,\ldots,N}$. Then the n'the diagonal element of $AB \in \mathbb{R}^{N\times N}$ is $\sum_{m=1}^{M} A_{nm}B_{mn}$ and the m'th diagonal element of $BA \in \mathbb{R}^{M\times M}$ is $\sum_{n=1}^{N} B_{mn}A_{nm}$ so that

$$\text{tr}(AB) = \sum_{n=1}^{N}\sum_{m=1}^{M} A_{nm}B_{mn} = \sum_{m=1}^{M}\sum_{n=1}^{N} B_{mn}A_{nm} = \text{tr}(BA).$$

b) A has the spectral decomposition $PDP^\top$ where $P^\top P$ is the identity matrix I and D is a diagonal matrix consisting of the eigenvalues. According to a) it holds

$$\text{tr}(A) = \text{tr}(PDP^\top) = \text{tr}(DP^\top P) = \text{tr}(DI) = \text{tr}(D)$$

so that tr(A) is the sum of its eigenvalues. □

8.4.2 Lemma
Let A and B be symmetric $S \times S$ matrices and A positive definite, i.e. $A > 0$. Then there exists a regular matrix $U \in \mathbb{R}^{S\times S}$ with

$$A = U^\top U \quad \text{and} \quad B = U^\top \,\text{diag}(\mu_1, \ldots, \mu_S)\, U,$$

where $\mu_s \underset{(-)}{>} 0$ for $s = 1, \ldots, S$ if $B \underset{(-)}{>} 0$.

Proof. According to the spectral decomposition, there exists an orthogonal matrix P with

$$A = P^\top \text{diag}(\lambda_1, \ldots, \lambda_S)\, P,$$

where $\lambda_s > 0$ for $s = 1, \ldots, S$ because of $A > 0$. Set

$$D^{1/2} := \text{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_S}\right), \quad D^{-1/2} = (D^{1/2})^{-1},$$

and

$$C := D^{-1/2} P\, B\, P^\top D^{-1/2}.$$

C is symmetric and $C \underset{(-)}{\geq} 0$ if $B \underset{(-)}{\geq} 0$. There exists also an orthogonal matrix Q with

$$Q\,C\,Q^\top = \operatorname{diag}(\mu_1, \ldots, \mu_S),$$

where $\mu_s \underset{(-)}{\geq} 0$ for $s = 1, \ldots, S$ if $B \underset{(-)}{\geq} 0$. Set $U := Q\,D^{1/2}P$. Then we have

$$U^\top U = P^\top D^{1/2} Q^\top Q\,D^{1/2} P = P^\top \operatorname{diag}(\lambda_1, \ldots, \lambda_S) P = A$$

and

$$\begin{aligned} & U^\top \operatorname{diag}(\mu_1, \ldots, \mu_S)\,U = U^\top\,Q\,C\,Q^\top\,U = P^\top D^{1/2} Q^\top\,Q\,C\,Q^\top\,Q\,D^{1/2} P \\ = \quad & P^\top D^{1/2}\,C\,D^{1/2} P = P^\top D^{1/2}\,D^{-1/2} P\,B\,P^\top D^{-1/2}\,D^{1/2} P = P^\top P\,B\,P^\top P = B. \end{aligned}$$ □

8.4.3 Lemma
Let A and B be symmetric $S \times S$ matrices with $A \geq B > 0$. Then it holds

a) $A^{-1} \leq B^{-1}$,

b) $\det A \geq \det B$,

c) $\operatorname{tr} A \geq \operatorname{tr} B$,

d) $\lambda_{max} A \geq \lambda_{max} B$.

Proof. According to Lemma 8.4.2, there exists a regular matrix $U \in \mathbb{R}^{S \times S}$ with

$$A = U^\top U \quad \text{and} \quad B = U^\top \operatorname{diag}(\mu_1, \ldots, \mu_S)\,U.$$

$A \geq B$ implies

$$I_{S\times S} = (U^\top)^{-1} A\,U^{-1} \geq (U^\top)^{-1} B\,U^{-1} = \operatorname{diag}(\mu_1, \ldots, \mu_S),$$

i.e. $1 \geq \mu_s > 0$ for $s = 1, \ldots, S$.
$I_{S\times S} \leq \operatorname{diag}(\mu_1^{-1}, \ldots, \mu_S^{-1})$ implies

$$A^{-1} = U^{-1}\,I_{S\times S}\,(U^\top)^{-1} \leq U^{-1} \operatorname{diag}(\mu_1^{-1}, \ldots, \mu_S^{-1})\,(U^\top)^{-1} = B^{-1}.$$

Moreover, we have

$$\det A = (\det U)^2\,\det I_{S\times S} \geq (\det U)^2\,\det \operatorname{diag}(\mu_1, \ldots, \mu_S) = \det B,$$

and

$$\operatorname{tr} A = \sum_{s=1}^S u_s^\top A\,u_s \geq \sum_{s=1}^S u_s^\top B\,u_s = \operatorname{tr} B,$$

where $u_1, \dots, u_S$ are the unit vectors. □

8.4.4 Lemma
If $M_1,\ M_2 \in \mathbb{R}^{R\times R}$ are symmetric and positive semidefinite and $L \in \mathbb{R}^{S\times R}$ with $L = K_1 M_1$ and $L = K_2 M_2$, then

$$L\ \left(\alpha\, M_1 + (1-\alpha)\, M_2\right)^{-}\ L^\top\ \leq\ \alpha\, L\, M_1^{-}\, L^\top + (1-\alpha)\, L\, M_2^{-}\, L^\top.$$

Proof. At first let be $M_1,\ M_2$ positive definite. Then also $M := \alpha\, M_1 + (1-\alpha)\, M_2$ is symmetric and positive definite. Hence we have for all $x, y \in \mathbb{R}^R$

$$0 \leq \left(x^\top - y^\top M^{-1}\right)\ M\ \left(x - M^{-1}y\right) = y^\top M^{-1} y - (2\, x^\top y - x^\top M\, x),$$

where equality holds if and only if $x - M^{-1}y = 0$, i.e. $x = M^{-1}y$. This means

$$y^\top M^{-1} y = \max\left\{2\, x^\top y - x^\top M\, x;\ x \in \mathbb{R}^R\right\},$$

which implies for all $l \in \mathbb{R}^S$

$$\begin{aligned}
l^\top L\, M^{-1}\, L^\top l &= \max\left\{2\, x^\top L^\top l - x^\top M\, x;\ x \in \mathbb{R}^R\right\} \\
&= \max\left\{\alpha\,(2\, x^\top L^\top l - x^\top M_1\, x) + (1-\alpha)\,(2\, x^\top L^\top l - x^\top M\, x);\ x \in \mathbb{R}^R\right\} \\
&\leq \alpha\ \max\left\{2\, x^\top L^\top l - x^\top M_1\, x;\ x \in \mathbb{R}^R\right\} \\
&\quad + (1-\alpha)\ \max\left\{2\, x^\top L^\top l - x^\top M_2\, x;\ x \in \mathbb{R}^R\right\} \\
&\overset{x = M_1^{-1} L^\top l \text{ bzw. } x = M_2^{-1} L^\top l}{=} \alpha\ \left(2\, l^\top L\, M_1^{-1} L^\top l - l^\top L\, M_1^{-1} L^\top l\right) \\
&\quad + (1-\alpha)\ \left(2\, l^\top L\, M_2^{-1} L^\top l - l^\top L\, M_2^{-1} L^\top l\right) \\
&= \alpha\, l^\top L\, M_1^{-1} L^\top l + (1-\alpha)\, l^\top L\, M_2^{-1} L^\top l.
\end{aligned}$$

For the case that M_1 or M_2 is not positive definite, see Kiefer (Journal of the Royal Statistical Society, B 21, P. 272ff) or Gaffke/Krafft (Modern Applied Mathematics - Optimization and Operation Research, Korte (eds.), North Holland 1981). □

8.4.5 Lemma
If $A,\ B \in \mathbb{R}^{R\times R}$ are symmetric and positive definite and $\alpha \in (0,1)$, then

$$\det(\alpha\, A + (1-\alpha)\, B) \geq (\det A)^\alpha\ (\det B)^{1-\alpha}.$$

Proof. If A and B are diagonal matrices, then the concavity of the logarithm provides

$$\begin{aligned}
&\ln(\det(\alpha\, A + (1-\alpha)\, B)) = \ln \prod_{r=1}^{R} (\alpha\, A_{rr} + (1-\alpha)\, B_{rr})\\
&= \sum_{r=1}^{R} \ln\, (\alpha\, A_{rr} + (1-\alpha)\, B_{rr}) \;\geq\; \sum_{r=1}^{R} (\alpha\, \ln\; A_{rr} + (1-\alpha)\, \ln\; B_{rr})\\
&= \alpha\, \ln \prod_{r=1}^{R} A_{rr} + (1-\alpha)\, \ln \prod_{r=1}^{R} B_{rr} \;=\; \ln \left(\prod_{r=1}^{R} A_{rr}\right)^{\alpha} + \ln \left(\prod_{r=1}^{R} B_{rr}\right)^{1-\alpha}.
\end{aligned}$$

To prove the assertion for the general case, we use the fact that according to Lemma 8.4.2 there exists a regular matrix U and diagonal matrix D such that

$$A = U^\top U \quad \text{and} \quad B = U^\top D\, U.$$

Then we obtain with the above result

$$\begin{aligned}
&\det(\alpha\, A + (1-\alpha)\, B) = \det(U^\top (\alpha\, I_{R\times R} + (1-\alpha)\, D)\, U)\\
&= (\det U)^2 \;\det(\alpha\, I_{R\times R} + (1-\alpha)\, D) \;\geq\; (\det U)^2\; (\det I_{R\times R})^{\alpha}\; (\det D)^{1-\alpha}\\
&= \left((\det U)^2 \;\det I_{R\times R}\right)^{\alpha} \;\left((\det U)^2 \;\det D\right)^{1-\alpha} \;=\; (\det U^\top U)^{\alpha}\; (\det U^\top D\, U)^{1-\alpha}\\
&= (\det A)^{\alpha}\; (\det B)^{1-\alpha}.
\end{aligned}$$

□

8.4.6 Lemma

a) If $A : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{N\times M}$ *and* $B : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{M\times K}$ *are differentiable in* t_0, *then*

$$\frac{\partial}{\partial t}\, A(t)\; B(t)\Big|_{t=t_0} = \left(\frac{\partial}{\partial t}\, A(t)\Big|_{t=t_0}\right)\; B(t_0) \;+\; A(t_0)\; \left(\frac{\partial}{\partial t}\, B(t)\Big|_{t=t_0}\right).$$

b) If $A : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{N\times N}$ *is differentiable in* t_0 *and* $A(t_0)$ *is regular, then*

$$\frac{\partial}{\partial t}\, A(t)^{-1}\Big|_{t=t_0} = -A(t_0)^{-1}\; \left(\frac{\partial}{\partial t}\, A(t)\Big|_{t=t_0}\right)\; A(t_0)^{-1}$$

and

$$\frac{\partial}{\partial t}\; \ln\det A(t)\Big|_{t=t_0} = tr\; \left(A(t_0)^{-1}\; \left(\frac{\partial}{\partial t}\, A(t)\Big|_{t=t_0}\right)\right).$$

Proof.
a) The assertion follows from the product rule.

b) Set $B(t) = A(t)\,A(t)^{-1} = I_{N\times N}$. The assertion a) implies

$$\begin{aligned}
\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} &= \frac{\partial}{\partial t}\,A(t)^{-1}B(t)\Big|_{t=t_0}\\
&= \frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} + A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)A(t_0)^{-1} + A(t_0)^{-1}\,A(t_0)\left(\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0}\right)\\
&= 2\,\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} + A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)A(t_0)^{-1}.
\end{aligned}$$

To prove the second assertion in b), let Π the set of all permutations of $\{1,\ldots,N\}$. Then we have

$$\begin{aligned}
\det A(t) &= \sum_{\pi\in\Pi}\operatorname{sgn}(\pi)\;A_{1\pi(1)}(t)\cdot\ldots\cdot A_{N\pi(N)}(t)\\
&= \sum_{n=1}^{N}A_{kn}(t)\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)\\
&= \sum_{n=1}^{N}A_{kn}(t)\;\alpha_{kn}(t)
\end{aligned}$$

for all $k = 1,\ldots,N$, where

$$\alpha_{kn}(t) = \sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)$$

is the cofactor of $A(t)$ with respect to (k,n). It follows

$$\begin{aligned}
\frac{\partial}{\partial t}\,\ln\det A(t)\Big|_{t=t_0} &= \frac{1}{\det A(t_0)}\,\frac{\partial}{\partial t}\,\det A(t)\Big|_{t=t_0}\\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{\pi\in\Pi}\operatorname{sgn}(\pi)\left(\frac{\partial}{\partial t}\,A_{k\pi(k)}(t)\Big|_{t=t_0}\right)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)\\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\left(\frac{\partial}{\partial t}\,A_{kn}(t)\Big|_{t=t_0}\right)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)\\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial t}\,A_{kn}(t)\Big|_{t=t_0}\right)\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)\\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial t}\,A_{kn}(t)\Big|_{t=t_0}\right)\alpha_{kn}\\
&= \operatorname{tr}\left(A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)\right)
\end{aligned}$$

since $A(t_0)^{-1} = \frac{1}{\det A(t_0)} (\alpha_{kn})_{k,n=1,\dots,N}$. □

8.5 Galois fields

Galois fields are fields of p elements, where p is prime and addition and multiplication over the set $\{0, 2, p-1\}$ is done modulo p.

8.5.1 Lemma
$mod_p(a) = mod_p(b) \Longleftrightarrow mod_p(a + (p-1)b) = 0$ for any $a, b \in \mathbb{Z}$.

Proof. Let $a = pL + l$ and $b = pM + m$ with $L, l, M, m \in \mathbb{Z}$. Then $l = \text{mod}_p(a) = Mod_p(c) = m$ implies

$$\text{mod}_p(a + (p-1)b) = \text{mod}_p(l + (p-1)m) = \text{mod}_p(m + (p-1)m) = \text{mod}_p(pm)) = 0.$$

Conversely, $\text{mod}_p(a + (p-1)b) = 0$ implies $\text{mod}_p(l + (p-1)m) = 0$ and thus $l + (p-1)m = pN$ for $N \in \mathbb{Z}$. This implies with $N = n + m$

$$l = pN - (p-1)m = pn + pm - (p-1)m = pn + m$$

and thus $\text{mod}_p(a) = \text{mod}_p(l) = \text{mod}_p(pn + m) = \text{mod}_p(m) = \text{mod}_p(b)$. □

8.5.2 Lemma
$mod_p(a + b) = mod_p(mod_p(a) + mod_p(b))$ and $mod_p(a \cdot b) = mod_p(mod_p(a) \cdot mod_p(b))$for any $a, b \in \mathbb{Z}$.

Proof. Let $a = pL + l$ and $b = pM + m$ with $L, l, M, m \in \mathbb{Z}$. Then

$$\text{mod}_p(\text{mod}_p(a) + \text{mod}_p(b)) = \text{mod}_p(l + m) = \text{mod}_p(a + b)$$

and

$$\begin{aligned} & \text{mod}_p(a \cdot b) = \text{mod}_p((pL + l) \cdot (pM + m)) \\ = \quad & \text{mod}_p(p^2 LM + lpM + pLm + lm) = \text{mod}_p(lm) = \text{mod}_p(\text{mod}_p(a) \cdot \text{mod}_p(b)). \end{aligned}$$

□

Chapter 9

References

1. Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs.* Clarendon Press, Oxford.

2. Bunke, H. and Bunke, O. (1989). *Nonlinear Regression, Functional Relations and Robust Methods. Statistical Methods of Model Building, Volume II.* Wiley, New York.

3. Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models.* Springer, Berlin.

4. Faraway, J.J. (2006). *Extending the Linear Model with R.* Chapman & Hall/CRC.

5. Federer, W.T. (1993). *Statistical Design and Analysis for Intercropping Experiments. Volume I Two Crops.* Springer, Berlin.

6. Federer, W.T. (1999). *Statistical Design and Analysis for Intercropping Experiments. Volume II Three or More Crops.* Springer, Berlin.

7. Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E. (1996). *Handbook of Small Data Sets.* Chapman & Hall, London.

8. Mukerjee, R., und Wu, C.F.J. (2006). *A Modern Theory of Factorial Design.* Springer, Berlin.

9. Müller, Ch.H. (1997). *Robust Planning and Analysis of Experiments.* Lecture Notes in Statistics 124, Springer, New York.

10. Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

11. Pistone, G., Riccomagno, E., and Wynn, H.P. (2001). *Algebraic Statistics: Computational Commutative Algebra in Statistics.* Chapman & Hall/CRC.

12. Pukelsheim, F. (1993). *Optimal Design of Experiments.* John Wiley, New York.

13. Rasch, D., Verdooren, L. R. und Gowers, J. (2007). *Planung und Auswertung von Versuchen und Erhebungen.* Oldenbourg, München.

14. Rencher, A.C. (1998). *Multivariate Statistical Inference and Applications.* Wiley, New York.

15. Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression.* Wiley, New York.

16. Shah, K.R., and Sinha, B.K. (1989). *Theory of Optimal Designs.* Springer New York.

17. Silvey, S.D. (1980). *Optimal Design.* Chapman and Hall, London.

18. Toutenburg, H. (2002). *Lineare Modelle. Theorie und Anwendungen.* Physica-Verlag, Berlin.

19. http://en.wikipedia.org/wiki/Fractional_factorial_designs

20. http://www1.uni-hamburg.de/RRZ/Software/Statistica/Handbuch/stexdes.html