Locally Optimal Designs for Generalized Linear Models with Applications to Gamma Models

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von MSc. Osama Idais

geb. am 05.02.1987 in Yatta, Palestine

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe
Prof. Dr. Norbert Gaffke
Prof. Dr. Dankmar Böhning

eingereicht am: 27.05.2019

Verteidigung am: 18.10.2019

Abstract

Locally optimal designs for generalized linear models are derived at certain values of the regression parameters. In the present thesis analytic solutions for optimal designs are mostly developed. In particular situations numerical methods are employed. We restrict to D-, A- and Kiefer Φ_k-optimality criteria.

For general setup of the generalized linear model, by means of The General Equivalence Theorem, necessary and sufficient conditions in term of intensity values are obtained to characterize locally optimal designs. In this context, linear predictors with binary factors are assumed constituting first order models, models with interactions and models without intercept. Additionally, a particular approach is developed to identify locally D- or A-optimal design for the model with intercept from that for the model without intercept and vice versa.

Gamma models with a power link function are considered constituting a particular class of generalized linear models. Relevant structures for the linear predictor are employed based on quantitative factors. The notions of locally essentially complete classes and locally complete classes of designs are introduced and such classes are established. On that basis locally D- and A-optimal designs are derived. In certain cases, the obtained results under generalized linear models with binary factors can be transferred to gamma models with quantitative factors. The explicit impact of the model parameters on the optimality of the designs is investigated. Furthermore, product type designs are derived for gamma models with product-type interactions. Moreover, gamma models having a linear predictor without intercept are considered. For a specific scenario sets of locally Φ_k-optimal designs are developed. Further, by a suitable transformation between gamma models with and without intercept optimality results are transferred from one model to the other. Additionally with the aid of The General Equivalence Theorem optimality are characterized for multiple regression by a system of polynomial inequalities which can be solved analytically or by computer algebra. The robustness of the derived designs for gamma models with respect to misspecifications of the initial parameter values is examined by means of their local efficiencies.

Optimal designs for multivariate generalized linear models are investigated. The components of the multivariate response might be combined with linear predictors via distinct link functions. We found that the locally optimal design for the univariate generalized linear models remains the same in the multivariate structure. In particular, product type designs are developed for the multivariate gamma model.

Zusammenfassung

Lokal optimale Versuchspläne für verallgemeinerte lineare Modelle werden für vorgegebene Werte der Regressionsparameter hergeleitet. In der vorliegenden Arbeit werden zumeist analytische Lösungen für optimale Versuchpläne entwickelt. In speziellen Situationen werden auch numerische Methoden verwendet. Wir beschränken unsere Untersuchungen auf das D- und A-Kriterium sowie Kiefers Φ_k-Optimalitätskriterien.

Im allgemeinen Rahmen der verallgemeinerten linearen Modelle werden mittels des allgemeinen Äquivalenzsatzes notwendige und hinreichende Bedingungen erhalten, die Intensitätswerte verwenden und lokal optimale Versuchspläne charakterisieren. In diesem Zusammenhang werden für lineare Prädiktoren mit binären Faktoren Modelle erster Ordnung, Modelle mit Wechselwirkungen und Modelle ohne Interzept (konstanten Term) betrachtet. Darüber hinaus wird eine spezielle Methode entwickelt, um lokal D- oder A-optimale Versuchspläne für ein Modell *mit* Interzept aus solchen für ein Modell *ohne* Interzept, und umgekehrt, zu konstruieren.

Im Weiteren werden Gamma-Modelle mit einer Potenzfunktion als Link-Funktion (Power Link) betrachtet, die eine spezielle Klasse verallgemeinerter linearer Modelle bilden. Hierzu werden relevante Strukturen für lineare Prädiktoren verwendet, die auf quantitativen Faktoren basieren. Die Begriffe einer lokal wesentlich vollständigen Klasse und einer lokal vollständigen Klasse von Versuchsplänen werden eingeführt, und derartige Klassen werden für verallgemeinerte lineare Modelle mit binären Faktoren erhaltene Resultate. In geeigneten Fällen können die für verallgemeinerte lineare Modelle mit binären Faktoren erhaltene Resultate auf Gamma-Modelle mit quantitativen Faktoren übertragen werden. Zur Messung der Qualität wird der Einfluss der Modellparameter auf die Optimalität der Versuchspläne untersucht. Weiterhin werden Versuchspläne mit Produkt-Struktur als optimal für Gamma-Modelle mit produktartigen Wechselwirkungen identifiziert. Darüber hinaus werden auch Gamma-Modelle mit linearem Prädiktor ohne Interzept betrachtet. Für ein spezielles Szenario werden Mengen lokal Φ_k-optimaler Versuchspläne gefunden. Durch eine geeignete Transformation werden Optimalitätsresultate für Gamma-Modelle *mit* Interzept auf Gamma-Modelle *ohne* Interzept, und umgekehrt, übertragen. Außerdem wird mit Hilfe des allgemeinen Äquivalenzsatzes die Optimalität für multiple Regression charakterisiert durch ein System polynomialer Ungleichungen, die analytisch oder mittels Computer Algebra gelöst werden können. Die Robustheit der hergeleiteten Versuchspläne für Gamma-Modelle bezüglich Fehlspezifikation der Parameter wird mittels ihrer lokalen Effizienzen überprüft.

Schließlich werden optimale Versuchspläne für multivariate verallgemeinerte lineare Modelle untersucht. Dabei können die Komponenten der multivariaten Regressionsfunktion mit linearen Prädiktoren über verschiedene Link-Funktionen kombiniert werden. Es kann gezeigt werden dass der lokal optimale Versuchsplan für das univariate verallgemeinerte lineare Modell such für die multivariate Struktur optimal bleibt. Insbesondere werden Versuchspläne mit Produkt-Struktur für multivariate Gamma-Modelle entwickelt.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor Professor Dr. Rainer Schwabe for the continuous support of my Ph.D study and the respective research, for his patience, motivation, comments and wide knowledge. Without his endless help, I would not have had the chance to be in Germany.

I would like to express my deepest appreciation to Professor Dr. Norbert Gaffke for sharing his knowledge with me, for his time and helpful suggestions and for his unlimited personal support.

I am very grateful for my time at the Institute for Mathematical Stochastics. My deep thank to all its members, especially Professor Dr. Alexandra Carpentier and the secretary Kerstin Altenkirch as well as the colleagues Dr. Maryna Prus, Dr. Kerstin Reckrühm, Frank Röttger, Parisa Parsamaram, Philipp Klein, Joseph Lam, Helmi Shat and Sajad Safarveisi.

I must also thank my colleague Eric Nyarko from Ghana who has taken this Ph.D journey with me.

I gratefully acknowledge the funding "Landesgraduierten-stipendium" received towards my Ph.D from the Graduate Promotion Act - GradFG of the Land of Saxony-Anhalt "Graduiertenförderungsgesetzes - GradFG des Landes Sachsen-Anhalt". I would like to say a big thank you to Frau Heidrun Wölke.

Last but not the least, I would especially like to thank my wife Ikhlas Idais for supporting me spiritually throughout writing this thesis and my life in general.

To my children

ASIL and OMAR

Contents

Chapter 1

Introduction

The generalized linear model (GLM) was developed by Nelder and Wedderburn (1972). It is viewed as a generalization of the ordinary linear regression which allows continuous or discrete observations from one-parameter exponential family distributions to be combined with explanatory variables (factors) via proper link functions. Therefore, wide applications can be addressed by GLMs such as social and educational sciences, clinical trials, insurance, industry. In particular; logistic and probit models are used for binary observations whereas Poisson models and gamma models are used for count and nonnegative continuous observations, respectively (Walker and Duncan (1967), Myers and Montgomery (1997), Fox (2015), Goldburd, Khare, and Tevet (2016)). Methods of likelihood are utilized to obtain the estimates of the model parameters. The precision of these maximum likelihood estimates (MLEs) is measured by their variance-covaraince matrix. In ordinary regression models for which normality assumption is realized the variance-covariance matrix is exactly (proportional to) the inverse of the Fisher information matrix. In contrast, for the GLMs the observations are often non-normal, and therefore large sample theory is demanded for the statistical inference. In this context, the variance-covariance matrix is approximately the inverse of the Fisher information matrix. It should, however, be emphasized that the Fisher information matrix for GLMs depends on the model parameters. The theory of generalized linear models is presented carefully in McCullagh and Nelder (1989) and Dobson and Barnett (2018).

Statistical inference is the procedure of drawing significant conclusions from the maximum likelihood estimates in the statistical models. The performance of the statistical inference is governed by research designs (studies). That is in observational designs like survey designs and cross-sectional designs the values of the explanatory variables are observed by the researcher, along with the values of the response variable without affecting them. In the other hand, in experimental designs like factorial designs the values of the explanatory variables are under the direct control of the researcher. More precisely, the values of explanatory variables are assigned (not observed) by the researcher to the values of the response variable or equivalently, the values of the response variable are allocated to specific values of the explanatory variables. For more details see Oehlert (2000), Fox (2015), Montgomery (2017). It is worthwhile mentioning that the essential ideas and concepts of experimental designs were developed in the books by Fisher (1937) and Cochran and Cox (1957). In the theory of optimal designs a more powerful inference is the main purpose that realizes through minimizing the estimates variation based on certain criteria and therefore, under optimal designs the most precise estimates are achieved.

The initial contribution in optimal experimental designs was made 101 years ago by Smith (1918). Her proposed method was later called G-criterion. Around 25 years later, the next contribution was introduced by Wald (1943) which explicitly includes the idea of the frequently-applied D-criterion. Rapid developments in the theory of optimal designs had been done until the outstanding papers by Kiefer (1959), Kiefer and Wolfowitz (1960) and Kiefer (1961) where the notion of the continuous (approximate) design was proposed which then allowed to employ convex optimization theory to obtain solutions of optimal designs leading to the celebrated Kiefer-Wolfowitz General Equivalence Theorem. Moreover, the alphabet labels referring to the optimality criteria were essentially proposed in the aforementioned works. Besides, several optimality criteria like A, E, I, V, L, c, Φ_k have also been developed. One can follow a variety of published works from the literature in Wynn (1984), Silvey (1980), Schwabe (1996b), Atkinson, Donev, and Tobias (2007) and Fedorov and Leonov (2013).

While deriving optimal designs is obtained by minimizing the variance-covariance matrix there is no loss of generality to concentrate on maximizing the Fisher information matrix. For generalized linear models the optimal design cannot be found without a prior knowledge of the parameters (Khuri et al. (2006), Atkinson and Woods (2015)). One approach which so-called local optimality was proposed by Chernoff (1953) aiming at deriving a locally optimal design at a given parameter value (best guess). This approach is widely employed, for instance; for count data with Poisson models and Rasch Poisson model see Wang et al. (2006), Russell et al. (2009) and Graßhoff, Holling, and Schwabe (2013, 2015, 2018). For binary data: see Abdelbasit and Plackett (1983) and Mathew and Sinha (2001) under logistic models and Biedermann, Dette, and Zhu (2006) under dose-response models whereas under logit, log-log and probit models see Yang, Mandal, and Majumdar (2012). In particular, optimal designs for GLMs without intercept have not been considered carefully. Recently, Kabera, Haines, and Ndlovu (2015) provided analytic proofs of D-optimal designs for zero intercept parameters of a two-binary-factor logistic model with no interaction.

Locally optimal designs for a general setup of generalized linear models have received some attention. Geometrically, Ford, Torsney, and Wu (1992) considered only one continuous factor. Atkinson and Haines (1996) presented a study of optimal designs for nonlinear model including GLMs. Yang (2008) provided optimal designs for

GLMs with applications to logistic and probit models. Also Yang and Stufken (2009) gave a general solution for GLMs. Analytic solutions under D-criterion obtained by Tong, Volkmer, and Yang (2014) for particular limitations.

The gamma model is a generalized linear model with gamma-distributed response variables. Mostly, it is employed for outcomes that are nonnegative, continuous, skewed and heteroscedastic specifically, when the variances are proportional to the square of the means. The gamma model with its canonical link (reciprocal) is appropriate for many real life data. For example; in ecology and forestry (Gea-Izquierdo and Cañellas (2009)), medicine (Grover, Sabharwal, and Mittal (2013)), air pollution studies (Kurtoğlu and Özkale (2016)), psychology (Ng and Cribbie (2017)), car insurance (McCullagh and Nelder (1989), Goldburd, Khare, and Tevet (2016), Section 2.1.3) and for Pharmacokinetic data (Lindsey et al. (2000)). Dette et al. (2013) considered gamma models with identity, inverse and log links. Gamma models with log-link are mostly used in cost data analysis (Barber and Thompson (2004), Moran et al. (2007) and Manning and Mullahy (2001)). However, although the gamma model is used in many applications, but it has no considerable attention for optimal designs. Geometric approaches were employed to derive locally D-optimal designs for a gamma model with a single factor (Ford, Torsney, and Wu (1992)), with two factors without intercept (Burridge and Sebastiani (1992)) and for multiple factors (Burridge and Sebastiani (1994)).

Optimal designs for multivariate linear models have been studied carefully (Fedorov (1971), Krafft and Schaefer (1992), Kurotschka and Schwabe (1996), Schwabe (1996a), Imhof (2000), Huang et al. (2006), Liu, Yue, and Hickernell (2011)). Recently, Rodríguez-Díaz and Sánchez-León (2019) introduced analogous result to that in Kurotschka and Schwabe (1996) for multiresponse models assuming double covariance structure (intra-correlation and inter-correlation). On the other hand, the research contributions in optimal designs for multivariate nonlinear models are limited (Heise and Myers (1996), Zocchi and Atkinson (1999), Fedorov and Leonov (2013), Liu and Colditz (2017)). In multivariate generalized linear model (MGLM) the marginal models are addressed within GLM framework. Mukhopadhyay and Khuri (2008) discussed response surface designs for MGLMs and they mentioned that very little is known about designs for such MGLMs. Das and Mukhopadhyay (2012) compared designs for MGLMs using quantile dispersion graphs when the linear predictor is misspecified.

In the present thesis, we are motivated to derive locally optimal designs in more complex, more realistic generalized linear models with several explanatory variables (multiple factors) and potentially several dependent variables (multivariate). Therefore, with the aid of The General Equivalence Theorem we focus on analytic solutions for optimal designs for a wide class of generalized linear models that are having similar form of the Fisher information matrix. We assume various setups of the linear predictor highlighting the impact of the presence or absence of the intercept term and the existence of interactions. We then concentrate on the gamma model as a particular application for GLMs. We provide outstanding and novel solutions for optimal designs for gamma models under different linear predictor taking into account the impact of the model parameters on the optimality solutions. Moreover, we propose an approach to reduce the complexity of deriving optimal designs for a multivariate generalized linear model to its univariate counterparts.

The thesis is organized as follows. In chapter 2 literature review of generalized linear models and the optimal design theory are presented. In Chapter 3 we introduce locally D- and A-optimal designs for a general setup of the generalized linear model having various linear predictors. In the subsequent sections some auxiliary results are developed and then optimal designs are derived for a one-factor model and a two-factor model with particular extensions to multiple-factor models. Further optimal designs are obtained under models with interactions. For non-intercept models we give a solution for a class of Φ_k-optimal designs. We also establish a relation of models with and without intercept under certain assumptions. In Chapter 4 we deal with the gamma model. We introduce the model considering a class of link functions as well as introducing the notations of locally complete or essentially complete classes of designs. Some relevant cases of linear predictors are considered and locally complete classes and locally essentially complete classes of designs are found leading to a considerable reduction of the problems of locally D- and A-optimal designs. Based on these results locally D- and A-optimal designs are determined. We begin with a one-factor model then models without interactions considering particular linear predictors with and without intercept. Additionally, models with interactions are discussed taking into account the existence and absence of the intercept. Finally, the performance of some derived locally D-optimal designs compared with particular non-optimal designs are examined. In Chapter 5 we concentrate on optimal designs for multivariate generalized linear models under various model structures. This thesis is closed with a summary of the results and suggestions for extensions with possible future topics in Chapter 6.

Chapter 2

Model specification and optimal designs

In the current chapter we present the fundamental concepts and notations in the theory of optimal experimental designs and generalized linear models that are required throughout our research. In Section 2.1 we introduce the model, the link functions, the intensity functions, the Fisher information matrix and the variance-covariance matrix. In Section 2.2 we introduce approximate designs, optimality criteria and The General Equivalence Theorem.

For more details about generalized linear models see the books by McCullagh and Nelder (1989) and Dobson and Barnett (2018). The essential theory of optimal designs and related topics are explained in the books by Fedorov (1972), Silvey (1980), Pukelsheim (1993), Schwabe (1996b), Atkinson, Donev, and Tobias (2007), Berger and Wong (2009) and Fedorov and Leonov (2013).

2.1 Univariate model

Let $Y_1, ..., Y_n$ be independent response variables for n experimental units. Consider the experimental region $\mathcal{X} \subseteq \mathbb{R}^\nu, \nu \geq 1$, to which the covariate value $\boldsymbol{x}$ belongs. Denote by $\boldsymbol{\beta} \in \mathbb{R}^p$ the parameter vector in a particular statistical model of interest. Let $\boldsymbol{f}_\beta(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^p$ be a vector of known functions at a given parameter point $\boldsymbol{\beta}$. The image (induced experimental region) $\boldsymbol{f}_\beta(\mathcal{X}) \subset \mathbb{R}^p$ is assumed to be a compact set where $\boldsymbol{f}_\beta(\mathcal{X}) = \{\boldsymbol{f}_\beta(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$. The Fisher information matrix, for a single observation at a point $\boldsymbol{x} \in \mathcal{X}$, is given by

$$\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_\beta(\boldsymbol{x})\, \boldsymbol{f}_\beta^\top(\boldsymbol{x}), \tag{2.1}$$

which of course depends on the model parameter $\boldsymbol{\beta}$. This form of information matrix appears when the model is nonlinear in $\boldsymbol{\beta}$. The information matrix of the whole experimental points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ reads as

$$\boldsymbol{M}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{\beta}). \tag{2.2}$$

In the context of the generalized linear models the observations (responses) belong to a one-parameter exponential family. The probability density function of Y defined as

$$p(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(\phi, y)\right), \tag{2.3}$$

where $a(\cdot), b(\cdot)$ and $c(\cdot)$ are known functions whereas θ is a canonical parameter and ϕ is a dispersion parameter. A common computational method for fitting the models to data are provided in the GLM framework. That is the expected mean is given by $E(Y) = \mu = b'(\theta)$, and the variance is given by $\mathrm{var}(Y) = a(\phi)b''(\theta)$. The quantity $b''(\theta)$ is called the mean-variance function or equivalently, the variance function of the expected mean, i.e., $V(\mu) = b''(\theta)$. Thus we may write $\mathrm{var}(Y) = a(\phi)V(\mu)$ which depends on the values of $\boldsymbol{x}$ (see McCullagh and Nelder (1989), Section 2.2.2).

Let $\boldsymbol{f}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^p$ be a p-dimensional regression function with components $f_1(\boldsymbol{x}), \dots, f_p(\boldsymbol{x})$. Here, the real-valued regression functions $f_1, \dots, f_p$ are continuous and linearly independent. The generalized linear model can be introduced as

$$\eta = g(\mu) \quad \text{where } \eta = \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}, \tag{2.4}$$

where g is a link function that relates the expected mean μ to the linear predictor $\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}$. It is assumed that g is one-to-one and differentiable. Table 2.1 gives the common link functions including the canonical links with the corresponding one-parameter exponential family distribution where Φ is the normal cumulative distribution and κ is the shape parameter of a gamma distribution (see Nelder and Wedderburn (1972), McCullagh and Nelder (1989), Myers and Montgomery (1997)).

One can realize that $\mu = \mu(\boldsymbol{x}, \boldsymbol{\beta}) = g^{-1}\big(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}\big)$ and $\mathrm{d}\eta/\mathrm{d}\mu = g'\Big(g^{-1}\big(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}\big)\Big)$ and therefore, we can define the intensity function at a point $\boldsymbol{x} \in \mathcal{X}$ as

$$u(\boldsymbol{x}, \boldsymbol{\beta}) = \left(\mathrm{var}(Y)\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}\right)^2\right)^{-1} \tag{2.5}$$

which is positive and depends on the value of linear predictor $\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}$. The intensity function is regarded as the weight for the corresponding unit at the point $\boldsymbol{x}$ (Atkinson and Woods (2015)). Table 2.2 shows the intensity functions with the corresponding link functions (Rodríguez, Ortiz, and Martínez (2016)).

TABLE 2.1: The link functions with the corresponding densities.

Density	Name	Link $g(\mu)$	Variance $V(\mu)$
Normal	Identity	μ	1
Poisson	Log	$\log \mu$	μ
Gamma	Reciprocal	κ/μ	μ^2
Gamma	Power Family	μ^ρ	μ^2
Gamma	Box-Cox	$(\mu^\lambda - 1)/\lambda$	μ^2
Inverse Gaussian	Inverse-Square	$1/\mu^2$	μ^3
Binomial	Logit or Logistic	$\log(\mu/(1-\mu))$	$\mu(1-\mu)/n$
Binomial	Probit	$\Phi^{-1}(\mu)$	$\mu(1-\mu)/n$
Binomial	Complementary log log	$\log\{-\log(1-\mu)\}$	$\mu(1-\mu)/n$

TABLE 2.2:
The intensity functions with the corresponding link functions.

Link	Intensity $u(\boldsymbol{x}, \boldsymbol{\beta})$
Identity	1
Log	μ
Reciprocal	$\kappa\mu^2$
Power Family	$\frac{1}{\rho^2}\frac{\kappa}{\mu^{2\rho}}$
Box-Cox	$\frac{\kappa}{\mu^{2\lambda}}$
Logit or Logistic	$\mu(1-\mu)$
Probit	$\frac{\Phi^2(\eta)}{(\Phi(\eta)(1-\Phi(\eta))}$
Complementary log log	$\frac{1-\mu}{\mu}(\log(1-\mu))^2$

In this context, the Fisher information matrix for a GLM at $\boldsymbol{x} \in \mathcal{X}$ (see Fedorov and Leonov (2013), Subsection 1.3.2) has the form

$$\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = u(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{f}^\top(\boldsymbol{x}), \tag{2.6}$$

which can be explicitly represented in the form (2.1) when $\boldsymbol{f}_\beta(\boldsymbol{x})$ is written as

$$\boldsymbol{f}_\beta(\boldsymbol{x}) = \sqrt{u(\boldsymbol{x}, \boldsymbol{\beta})}\boldsymbol{f}(\boldsymbol{x}), \tag{2.7}$$

The information matrix of the form (2.6) is appropriate for other nonlinear models, e.g., survival times observations which depend on the proportional hazard model (see Schmidt and Schwabe (2017)). Moreover, under homoscedastic regression models the intensity function is constant equal to 1 whereas, under heteroscedastic regression models we get intensity that is equal to $1/\text{var}(Y)$ which depends on $\boldsymbol{x}$ only and thus we have information matrix of form $\boldsymbol{M}(\boldsymbol{x}) = u(\boldsymbol{x})\, \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{f}^\top(\boldsymbol{x})$ that does not depend on the model parameters. The latter case was discussed in Graßhoff et al. (2007) and in the book by Fedorov and Leonov (2013), p.13.

Remark 2.1.1. *In the thesis we will deal with generalized linear models with and without intercept. The generalized linear model includes* ***explicitly*** *an intercept term* β_0 *if the regression function* $\boldsymbol{f}(\boldsymbol{x})$ *includes the constant* 1*, whereas the model includes* ***implicitly*** *an intercept term if there exists a constant vector* $\boldsymbol{c}$ *such that* $\boldsymbol{c}^\top \boldsymbol{f}(\boldsymbol{x}) = 1$ *for all* $\boldsymbol{x} \in \mathcal{X}$*.*

It is worthwhile mentioning that unlike the case of normally distributed response variables, the sampling distributions for MLEs $\hat{\boldsymbol{\beta}}$ in GLMs that used for inference cannot be determined exactly. Therefore, the statistical inferences for GLMs are conducted for large sample sizes under mild regularity assumptions on the probability density (2.3). Hence,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\text{d}} \mathcal{N}_p\Big(\boldsymbol{0}, \boldsymbol{M}^{-1}\Big)$$

where $\boldsymbol{M} = \lim\limits_{n\to\infty} \frac{1}{n}\boldsymbol{M}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{\beta})$ (Fahrmeir and Kaufmann (1985), Theorem 3). Moreover, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is approximately given by the inverse of the Fisher information matrix (2.2), see Fedorov and Leonov (2013), Section 1.5.,

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{M}^{-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{\beta}). \tag{2.8}$$

2.2 Optimal design

In the theory of optimal designs there are three main parts that should be taken into account. The statistical model which relates the response (observation) to the explanatory variables (factors), the experimental region which represents the range of these factors and the optimality criterion under which a design for the proposed model on an experimental region is optimal. The quality of the design is measured by the variance-covariance matrix of parameter estimates (2.8) or equivalently by its inverse, the Fisher information matrix. The solution of optimal designs for generalized linear models is difficult since it is affected by the values of the model parameters $\boldsymbol{\beta}$ that appear in the information matrix.

In the literature of optimal designs there are various approaches to manage the dependence of the model parameters, see Mukhopadhyay and Khuri (2008) and Yang and Mandal (2015). These approaches can be listed as below.

- Local optimality approach in which the unknown parameters are replaced by assumed values.
- Bayesian approach that considers a prior belief on unknown parameters.
- Maximin approach that maximizes the minimum efficiency over certain range of values of the unknown parameters.
- Sequential approach where the estimates of the design parameters are updated in an iterative way.

Throughout we restrict to the local optimality approach which was introduced by Chernoff (1953). A locally optimal design is derived at a certain best guess of the model parameter w.r.t. a particular optimality criterion.

2.2.1 Approximate design

Throughout the present work we will deal with the approximate (continuous) design theory, i.e., a design ξ is a probability measure with finite support on the experimental region $\mathcal{X}$,

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_r \\ \omega_1 & \omega_2 & \dots & \omega_r \end{pmatrix}, \tag{2.9}$$

where $r \in \mathbb{N}$, $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r \in \mathcal{X}$ are pairwise distinct points and $\omega_1, \omega_2, \dots, \omega_r > 0$ with $\sum_{i=1}^r \omega_i = 1$. The set $\text{supp}(\xi) = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r\}$ is called the support of ξ and $\omega_1, \dots, \omega_r$ are called the weights of ξ, see Silvey (1980), p.15. The information matrix of a design ξ from (2.9) at a parameter point $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \int_{\mathcal{X}} \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta})\, \xi(\mathrm{d}\boldsymbol{x}) = \sum_{i=1}^{r} \omega_i \boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{\beta}). \tag{2.10}$$

One might recognize $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ as a convex combination of all information matrices for all design points of ξ. Another representation of the information matrix (2.10) can be utilized based on the $r \times p$ design matrix $\boldsymbol{F} = [\boldsymbol{f}(\boldsymbol{x}_1), \dots, \boldsymbol{f}(\boldsymbol{x}_r)]^\top$ and the $r \times r$ weight matrix $\boldsymbol{V} = \text{diag}(\omega_i u(\boldsymbol{x}_i, \boldsymbol{\beta}))_{i=1}^r$ and hence,

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F}.$$

Remark 2.2.1. *A particular type of designs appears frequently when the support size equals the dimension of* $\boldsymbol{f}_{\boldsymbol{\beta}}$ *(or* $\boldsymbol{f}$*), i.e.,* $r = p$*. In such a case the design is minimally supported and it is often called a minimal-support or a saturated design.*

2.2.2 Optimality criteria

Let Ξ be the convex set of all designs on $\mathcal{X}$. Since we deal with local optimality we define, for a given parameter point $\boldsymbol{\beta}$, the set $\mathbb{M}_{\boldsymbol{\beta}} = \{\boldsymbol{M}(\xi, \boldsymbol{\beta}) : \xi \in \Xi\}$ which is convex and contains symmetric nonnegative definite $p \times p$ moment matrices. Define the criterion function

$$\Phi : \mathbb{M}_{\boldsymbol{\beta}} \to \mathbb{R}$$

which assumed to be convex and differentiable over $\mathbb{M}_{\boldsymbol{\beta}}$. The criterion function Φ depends on the design ξ only through the moment matrix $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ (see Pukelsheim (1993), Section 4.1.). Given a parameter point $\boldsymbol{\beta}$, a design ξ^* is said to be locally Φ-optimal (at $\boldsymbol{\beta}$) if its information matrix at $\boldsymbol{\beta}$ is nonsingular and

$$\Phi\Big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})\Big) = \min_{\xi} \Phi\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big),$$

where the minimum on the r.h.s. is taken over all designs ξ whose information matrix at $\boldsymbol{\beta}$ is nonsingular.

Remark 2.2.2. *The set of designs for which the information matrix is nonsingular does not depend on* $\boldsymbol{\beta}$ *(when* $u(\boldsymbol{x}, \boldsymbol{\beta})$ *is strictly positive on* $\mathcal{X}$*). In particular it is just the set of designs for which the information matrix is nonsingular in the corresponding ordinary regression model (ignoring the intensity* $u(\boldsymbol{x}, \boldsymbol{\beta})$*). That is the singularity depends on the support points of a design* ξ *because its information matrix* $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F}$ *is full rank if and only if* $\boldsymbol{F}$ *is full rank.*

In this research, we mostly concentrate on D- and A-optimal designs so the notions of locally D- and A-optimality will be introduced here in detail (see Fedorov and Leonov (2013), Section 2.2).

D-optimal designs are constructed to minimize the determinant of the variance-covariance matrix of the estimates or equivalently to maximize the determinant of the information matrix. The D-criterion is typically defined by the convex function

$$\Phi\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big) = -\log\det\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big)$$

where $\det(\boldsymbol{A})$ denotes the determinant of a $p \times p$ matrix $\boldsymbol{A}$. Geometrically, the volume of the asymptotic confidence ellipsoid is inversely proportional to $\sqrt{\det\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big)}$ where $\det\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big)$ can be determined by the inverse of the product of the squared lengths of the axes. Therefore, the D-optimal designs minimize the volume of the asymptotic confidence ellipsoid.

A-optimal designs are constructed to minimize the trace of the variance-covariance matrix of the estimates, i.e., to minimize the average variance of the estimates. The A-criterion is defined by

$$\Phi\Big(\boldsymbol{M}(\xi,\boldsymbol{\beta})\Big) = \mathrm{tr}\Big(\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta})\Big)$$

where $\mathrm{tr}(\boldsymbol{A})$ denotes the trace of a $p \times p$ matrix $\boldsymbol{A}$. The A-criterion aims at minimizing the sum of the squared lengths of the axes of the asymptotic confidence ellipsoid.

An advantage of D-optimality is that the optimal designs do not depend on the scale of the factors, even though the value of $\boldsymbol{M}(\xi^*,\boldsymbol{\beta})$ does. A one-to-one linear transformation of $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ leaves the optimal design unchanged, which is not, in general, the case for A-optimal designs.

In a certain part of the current research a family of optimal designs and, more generally, under Kiefer Φ_k-criteria (Kiefer (1975)) is introduced, in particular for models without intercept. Kiefer Φ_k-criteria aim at minimizing the k-norm of the eigenvalues of the variance-covariance matrix. The Φ_k-criteria include the above D- and A-criteria as well as the E-criterion. Note that for a given parameter point $\boldsymbol{\beta}$ a design ξ^* is locally E-optimal if and only if it maximizes the smallest eigenvalue of $\boldsymbol{M}(\xi,\boldsymbol{\beta})$ among all designs $\xi \in \Xi$. The E-criterion minimizes the squared length of the 'largest' axis of the asymptotic confidence ellipsoid.

Denote by $\lambda_i(\xi,\boldsymbol{\beta})$ for all $(1 \le i \le p)$ the eigenvalues of a nonsingular information matrix $\boldsymbol{M}(\xi,\boldsymbol{\beta})$. The Φ_k- criteria are defined by

$$\begin{aligned}
\Phi_k(\xi,\boldsymbol{\beta}) &= \Big(\frac{1}{p}\mathrm{tr}\Big(\boldsymbol{M}^{-k}(\xi,\boldsymbol{\beta})\Big)\Big)^{\frac{1}{k}} = \Big(\frac{1}{p}\sum_{i=1}^{p}\lambda_i^{-k}(\xi,\boldsymbol{\beta})\Big)^{\frac{1}{k}},\ 0 < k < \infty,\\
\Phi_0(\xi,\boldsymbol{\beta}) &= \lim_{k\to 0+}\phi_k(\xi,\boldsymbol{\beta}) = \Big(\det(\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta}))\Big)^{\frac{1}{p}},\\
\Phi_\infty(\xi,\boldsymbol{\beta}) &= \lim_{k\to\infty}\phi_k(\xi,\boldsymbol{\beta}) = \max_{1\le i\le p}\Big(\lambda_i^{-1}(\xi,\boldsymbol{\beta})\Big).
\end{aligned}$$

Note that $\Phi_0(\xi,\boldsymbol{\beta})$, $\Phi_1(\xi,\boldsymbol{\beta})$ and $\Phi_\infty(\xi,\boldsymbol{\beta})$ are the D-, A- and E-criteria, respectively.

Remark 2.2.3. *By the strict convexity of the function Φ on $\mathbb{M}_{\boldsymbol{\beta}}$ the information matrix of a locally Φ-optimal design (at $\boldsymbol{\beta}$) is unique. That is, if ξ^* and ξ^{**} are two locally Φ-optimal designs (at $\boldsymbol{\beta}$) then $\boldsymbol{M}(\xi^*,\boldsymbol{\beta}) = \boldsymbol{M}(\xi^{**},\boldsymbol{\beta})$. Of course this is achieved under D- and A-optimality and in particular under Kiefer Φ_k-criteria for $0 \le k < \infty$.*

Remark 2.2.4. *In general, it is assumed that the experimental region is compact and at a given $\boldsymbol{\beta}$ the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ is continuous. This entails existence of a locally*

D-, A- or Φ_k-optimal design for any given parameter point $\boldsymbol{\beta}$. Although the experimental region $\mathcal{X}$ is the main objective in application more than $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X})$, but the latter region is more realistic analytically in mathematical development. Because of the obvious correspondence between $\boldsymbol{f}_{\boldsymbol{\beta}}(\xi)$ and ξ, no ambiguity will arise (see Pukelsheim (1993), Section 1.25). Therefore, choosing a point $\boldsymbol{x}$ in an experimental region $\mathcal{X}$ is equivalent to choosing $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ in an induced experimental region $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X})$ at a given $\boldsymbol{\beta}$. The compactness of $\mathcal{X}$ is demanded to guarantee that $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X})$ is compact thus the set of all nonnegative definite matrices $\mathbb{M}_{\boldsymbol{\beta}}$ is so. In fact, the compactness of $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X})$ is necessarily demanded which might occur while $\mathcal{X}$ is non-compact.

2.2.3 The General Equivalence Theorem

In order to verify the local optimality of a design The General Equivalence Theorem is usually employed. It provides necessary and sufficient conditions for a design to be Φ-optimal and thus the optimality of a suggested design can be easily verified or disproved. The most generic one is the celebrated Kiefer-Wolfowitz equivalence theorem under D-criterion (see Kiefer and Wolfowitz (1960)).

The Equivalence Theorem is established by making use of the directional derivatives of the optimality criterion Φ at a given parameter point. Denote by $\xi_{\boldsymbol{x}}$ a design which assigns unit mass at the design point $\boldsymbol{x}$ and let ξ' be defined as

$$\xi' = (1-\alpha)\xi + \alpha\xi_{\boldsymbol{x}} \quad \text{for} \quad 0 \le \alpha \le 1,$$

then we have

$$\boldsymbol{M}(\xi', \boldsymbol{\beta}) = (1-\alpha)\boldsymbol{M}(\xi, \boldsymbol{\beta}) + \alpha \boldsymbol{M}(\xi_{\boldsymbol{x}}, \boldsymbol{\beta}),$$

where $\boldsymbol{M}(\xi_{\boldsymbol{x}}, \boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}^{\top}_{\boldsymbol{\beta}}(\boldsymbol{x})$. Due to convexity of both Ξ and $\mathbb{M}_{\boldsymbol{\beta}}$ we observe that $\xi' \in \Xi$ and $\boldsymbol{M}(\xi', \boldsymbol{\beta}) \in \mathbb{M}_{\boldsymbol{\beta}}$. The directional (Frechet) derivative of Φ at $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ in the direction $\boldsymbol{M}(\xi_{\boldsymbol{x}}, \boldsymbol{\beta})$ is given by

$$F_{\Phi}\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta}), \boldsymbol{M}(\xi_{\boldsymbol{x}}, \boldsymbol{\beta})\Big) = \lim_{\alpha \to 0^+} \frac{1}{\alpha}\left[\Phi\Big((1-\alpha)\boldsymbol{M}(\xi, \boldsymbol{\beta}) + \alpha\boldsymbol{M}(\xi_{\boldsymbol{x}}, \boldsymbol{\beta})\Big) - \Phi\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big)\right].$$

The following theorem provides necessary and sufficient conditions for optimality.

Theorem 2.2.1. *(Silvey (1980),Theorem 6.1.2, p. 54) Given a parameter point $\boldsymbol{\beta}$. Let Φ be convex on $\mathbb{M}_{\boldsymbol{\beta}}$ and differentiable at $\boldsymbol{M}(\xi^*, \boldsymbol{\beta})$. Then ξ^* is locally Φ-optimal (at $\boldsymbol{\beta}$) if and only if*

$$F_{\Phi}\Big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta}), \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}^{\top}_{\boldsymbol{\beta}}(\boldsymbol{x})\Big) \ge 0 \quad \textit{for all } \boldsymbol{x} \in \mathcal{X}.$$

Remark 2.2.5. *Under the assumptions of Theorem 2.2.1 and if ξ^* is locally Φ-optimal, note that*

$$F_\Phi\Big(\boldsymbol{M}(\xi^*,\boldsymbol{\beta}), \boldsymbol{f}_\beta(\boldsymbol{x})\boldsymbol{f}_\beta^\top(\boldsymbol{x})\Big) = 0 \quad \textit{for all } \boldsymbol{x} \in \operatorname{supp}(\xi^*).$$

The General Equivalence Theorem is characterized for the locally D-, A- and Φ_k-optimal designs in the following theorem.

Theorem 2.2.2. *Let $\boldsymbol{\beta}$ be a given parameter point and let ξ^* be a design with nonsingular information matrix $\boldsymbol{M}(\xi^*,\boldsymbol{\beta})$.*

(a) The design ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) if and only if

$$\boldsymbol{f}_\beta^\top(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\boldsymbol{f}_\beta(\boldsymbol{x}) \le p \textit{ for all } \boldsymbol{x} \in \mathcal{X}. \tag{2.11}$$

(b) The design ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) if and only if

$$\boldsymbol{f}_\beta^\top(\boldsymbol{x})\boldsymbol{M}^{-2}(\xi^*,\boldsymbol{\beta})\boldsymbol{f}_\beta(\boldsymbol{x}) \le \operatorname{tr}\Big(\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\Big) \textit{ for all } \boldsymbol{x} \in \mathcal{X}. \tag{2.12}$$

(c) The design ξ^* is locally Φ_k-optimal (at $\boldsymbol{\beta}$) if and only if

$$\boldsymbol{f}_\beta^\top(\boldsymbol{x})\boldsymbol{M}^{-k-1}(\xi^*,\boldsymbol{\beta})\boldsymbol{f}_\beta(\boldsymbol{x}) \le \operatorname{tr}(\boldsymbol{M}^{-k}(\xi^*,\boldsymbol{\beta})) \textit{ for all } \boldsymbol{x} \in \mathcal{X}. \tag{2.13}$$

Remark 2.2.6.
If ξ^* is a locally D-, A- or Φ_k-optimal design (at $\boldsymbol{\beta}$) then for each support point $\boldsymbol{x}$ of ξ^* the inequality in (a), (b) or (c), respectively of the theorem is an equality (cp. Remark 2.2.5).

Remark 2.2.7.
In each condition of The General Equivalence Theorem, Theorem 2.2.2, the left hand side of the inequality is called the sensitivity function.

Chapter 3

Generalized linear models

In this chapter we deal with a wide class of generalized linear models. In Section 3.1 we develop some approaches to determine the optimal weights for particular designs under D-, A- and Φ_k-criteria which will be used later. Throughout, with the aid of The Equivalence Theorem (Theorem 2.2.2) we establish a necessary and sufficient condition for a design to be locally D-, A- or Φ_k-optimal designs. We begin with the single-factor model by Section 3.2. In Section 3.3 we consider a model without interaction whereas a model with interaction is studied briefly in Section 3.4. In Section 3.5 we focus on Kiefer Φ_k-criteria for models without intercept. In Section 3.6 a relation of models with and without intercept according to D- and A-optimal designs is developed.

3.1 Auxiliary results

Some saturated designs will appear as candidates for local D- and A-optimality. If points $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^* \in \mathcal{X}$ are given such that the vectors $\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_p^*)$ are linearly independent and $\boldsymbol{\beta}$ is a given parameter point, an interesting question is the choice of locally D- and A-optimal weights ω_i^* $(1 \leq i \leq p)$ to obtain the (saturated) design ξ^* with support $\{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*\}$ and weights ω_i^* $(1 \leq i \leq p)$ which yields the minimum value of $-\log\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big)$ and the minimum value of $\mathrm{tr}\big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\big)$ over all saturated designs with the same support $\{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*\}$. For the A-criterion the answer was given in Pukelsheim (1993), Section 8.8, which is part of our following auxiliary lemma.

Lemma 3.1.1. *Let $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^* \in \mathcal{X}$ be given such that the vectors $\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_p^*)$ are linearly independent and let $\boldsymbol{\beta}$ be a given parameter point. The design ξ^* which achieves the minimum value of $\mathrm{tr}\big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\big)$ over all designs ξ with $\mathrm{supp}(\xi^*) = \{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*\}$ is given by*

$$\xi^* = \begin{pmatrix} \boldsymbol{x}_1^* & \ldots & \boldsymbol{x}_p^* \\ \omega_1^* & \ldots & \omega_p^* \end{pmatrix}, \quad \text{with} \;\; \omega_i^* = c^{-1}\Big(\frac{c_{ii}}{u(\boldsymbol{x}_i^*, \boldsymbol{\beta})}\Big)^{1/2} \; (1 \leq i \leq p)\,, \;\; c = \sum_{k=1}^{p}\Big(\frac{c_{kk}}{u(\boldsymbol{x}_k^*, \boldsymbol{\beta})}\Big)^{1/2},$$

where c_{ii} $(1 \le i \le p)$ *are the diagonal entries of the matrix* $\boldsymbol{C} = (\boldsymbol{F}^{-1})^\top \boldsymbol{F}^{-1}$, *and* $\boldsymbol{F} = \Big[\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_p^*)\Big]^\top$. *Moreover, the design* ξ^* *is locally A-optimal (at* $\boldsymbol{\beta}$*) if and only if*

$$\Big(\boldsymbol{U}^{-1/2}(\boldsymbol{F}^{-1})^\top \boldsymbol{f}(\boldsymbol{x})\Big)^\top \boldsymbol{C}^* \Big(\boldsymbol{U}^{-1/2}(\boldsymbol{F}^{-1})^\top \boldsymbol{f}(\boldsymbol{x})\Big) \le 1/u(\boldsymbol{x}, \boldsymbol{\beta}) \;\; \forall\, \boldsymbol{x} \in \mathcal{X} \setminus \{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*\}, \tag{3.1}$$

where $\boldsymbol{C}^* = \operatorname{diag}\Big(c_{11}^{-1/2}, \ldots, c_{pp}^{-1/2}\Big)\, \boldsymbol{C} \operatorname{diag}\Big(c_{11}^{-1/2}, \ldots, c_{pp}^{-1/2}\Big)$ *and* $\boldsymbol{U} = \operatorname{diag}\Big(u(\boldsymbol{x}_1^*, \boldsymbol{\beta}), \ldots, u(\boldsymbol{x}_p^*, \boldsymbol{\beta})\Big)$.

Proof. The formula for the A-optimal weights ω_i^* $(1 \le i \le p)$ is due to the corollary in Section 8.8 of Pukelsheim (1993). Denoting the weight matrix by $\boldsymbol{V} = \boldsymbol{\Omega}\boldsymbol{U}$ where $\boldsymbol{\Omega} = \operatorname{diag}\Big(\omega_1^*, \ldots, \omega_p^*\Big)$, we can write $\boldsymbol{M}(\xi^*, \boldsymbol{\beta}) = \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F} = \boldsymbol{F}^\top \boldsymbol{\Omega}\boldsymbol{U}\boldsymbol{F}$ and

$$\begin{aligned}
\operatorname{tr}\Big(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\Big) &= \operatorname{tr}\Big(\boldsymbol{F}^{-1}\boldsymbol{U}^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{F}^{-1})^\top\Big) = \operatorname{tr}\Big(\boldsymbol{\Omega}^{-1}\boldsymbol{U}^{-1}\boldsymbol{C}\Big) = \sum_{i=1}^{p}\Big(\omega_i^* u(\boldsymbol{x}_i^*, \boldsymbol{\beta})\Big)^{-1} c_{ii} \\
&= c\sum_{i=1}^{p}\Big(c_{ii} u(\boldsymbol{x}_i^*, \boldsymbol{\beta})\Big)^{-1/2} c_{ii} = c\sum_{i=1}^{p}\bigg(\frac{c_{ii}}{u(\boldsymbol{x}_i^*, \boldsymbol{\beta})}\bigg)^{1/2} = c^2,
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta}) &= \boldsymbol{F}^{-1}\boldsymbol{U}^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{F}^{-1})^\top \boldsymbol{F}^{-1}\boldsymbol{U}^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{F}^{-1})^\top \\
&= c^2\, \boldsymbol{F}^{-1}\boldsymbol{U}^{-1/2}\boldsymbol{C}^*\boldsymbol{U}^{-1/2}(\boldsymbol{F}^{-1})^\top \;\text{ since } \boldsymbol{U}^{-1/2}\boldsymbol{\Omega}^{-1} = c \operatorname{diag}\Big(c_{11}^{-1/2}, \ldots, c_{pp}^{-1/2}\Big).
\end{aligned}$$

So, together with The Equivalence Theorem (Theorem 2.2.2, condition (2.12)) and Remark 2.2.6 the asserted characterization of local A-optimality (at $\boldsymbol{\beta}$) of ξ^* follows. □

For the D-criterion the well-known answer is $\omega_i^* = 1/p$ $(1 \le i \le p)$, see Lemma 5.1.3 of Silvey (1980). That is the locally D-optimal saturated design assigns equal weights to the support points. On the other hand, there is no unified formulas for the optimal weights of a non-saturated design with respect to D-criterion. However, let the model be given with parameter vector $\boldsymbol{\beta}$ of dimension $p = 3$, i.e., $\boldsymbol{\beta} \in \mathbb{R}^3$. The next lemma provides the optimal weights of a design with four support points $\xi^* = \{(\boldsymbol{x}_i^*, \omega_i^*), i = 1, 2, 3, 4\}$ under certain conditions.

Lemma 3.1.2. *Let* $p = 3$. *Let the design points* $\boldsymbol{x}_1^*$, $\boldsymbol{x}_2^*$, $\boldsymbol{x}_3^*$, $\boldsymbol{x}_4^* \in \mathcal{X}$ *be given such that any three of the four vectors* $\boldsymbol{f}(\boldsymbol{x}_1^*), \boldsymbol{f}(\boldsymbol{x}_2^*), \boldsymbol{f}(\boldsymbol{x}_3^*), \boldsymbol{f}(\boldsymbol{x}_4^*)$ *are linearly independent. Denote*

$$\begin{aligned}
d_1 &= \det\Big[\boldsymbol{f}(\boldsymbol{x}_2^*), \boldsymbol{f}(\boldsymbol{x}_3^*), \boldsymbol{f}(\boldsymbol{x}_4^*)\Big], \;\; d_2 = \det\Big[\boldsymbol{f}(\boldsymbol{x}_1^*), \boldsymbol{f}(\boldsymbol{x}_3^*), \boldsymbol{f}(\boldsymbol{x}_4^*)\Big], \\
d_3 &= \det\Big[\boldsymbol{f}(\boldsymbol{x}_1^*), \boldsymbol{f}(\boldsymbol{x}_2^*), \boldsymbol{f}(\boldsymbol{x}_4^*)\Big], \;\; d_4 = \det\Big[\boldsymbol{f}(\boldsymbol{x}_1^*), \boldsymbol{f}(\boldsymbol{x}_2^*), \boldsymbol{f}(\boldsymbol{x}_3^*)\Big]
\end{aligned}$$

such that $d_i \neq 0$, $i = 1, 2, 3, 4$. *For a given parameter point* $\boldsymbol{\beta}$ *denote* $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta})$, $i = 1, 2, 3, 4$. *Assume that* $u_2 = u_3$ *and* $d_2^2 = d_3^2$ *and let*

$$\begin{aligned}
\omega_1^* &= \frac{3}{8} + \frac{1}{4}\left(1 + \frac{d_1^2}{d_4^2}\frac{u_1}{u_4} - 4\frac{d_2^2}{d_4^2}\frac{u_1}{u_2}\right)^{-1}, \\
\omega_2^* = \omega_3^* &= \frac{1}{2}\left(4 - \frac{d_4^2}{d_2^2}\frac{u_2}{u_1} - \frac{d_1^2}{d_2^2}\frac{u_2}{u_4}\right)^{-1}, \\
\omega_4^* &= \frac{3}{8} + \frac{1}{4}\left(1 + \frac{d_4^2}{d_1^2}\frac{u_4}{u_1} - 4\frac{d_2^2}{d_1^2}\frac{u_4}{u_2}\right)^{-1}.
\end{aligned}$$

Assume that $\omega_i^* > 0, i = 1, 2, 3, 4$. *Then the design* ξ^* *which achieves the minimum value of* $-\log\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big)$ *over all designs* ξ *with* $\operatorname{supp}(\xi) = \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*\}$ *is given by* $\xi^* = \{(\boldsymbol{x}_i^*, \omega_i^*), i = 1, 2, 3, 4\}$.

Proof. Let $\boldsymbol{f}_\ell = \boldsymbol{f}(\boldsymbol{x}_\ell^*) = \big(f_{\ell 1}, f_{\ell 2}, f_{\ell 3}\big)^\top$ $(1 \leq \ell \leq 4)$. The 4×3 design matrix is given by $\boldsymbol{F} = \big[\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4\big]^\top$. Denote $\boldsymbol{V} = \operatorname{diag}\big(\omega_\ell u_\ell\big)_{\ell=1}^4$. Then $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F}$ and by the Cauchy-Binet formula the determinant of $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ is given by the function $\varphi(\omega_1, \omega_2, \omega_3, \omega_4)$ where

$$\varphi(\omega_1, \omega_2, \omega_3, \omega_4) = \sum_{\substack{1 \leq i < j < k \leq 4 \\ h \in \{1,2,3,4\} \setminus \{i,j,k\}}} d_h^2 u_i u_j u_k \, \omega_i \omega_j \omega_k. \tag{3.2}$$

By assumptions $u_2 = u_3$, $d_2^2 = d_3^2$ the function $\varphi(\omega_1, \omega_2, \omega_3, \omega_4)$ is invariant w.r.t. permuting ω_2 and ω_3, i.e., $\varphi(\omega_1, \omega_2, \omega_3, \omega_4) = \varphi(\omega_1, \omega_3, \omega_2, \omega_4)$ and thus minimizing (3.2) has similar solutions for ω_2 and ω_3. Hence, $\omega_4 = 1 - \omega_1 - 2\omega_2$ and (3.2) reduces to

$$\varphi(\omega_1, \omega_2) = \alpha_1 \omega_2^3 + \alpha_2 \omega_2^2 + \alpha_3 \omega_1^2 \omega_2 + \alpha_4 \omega_2^2 \omega_1 + \alpha_5 \omega_1 \omega_2,$$

where α_ℓ $(1 \leq \ell \leq 5)$ are given by

$$\begin{aligned}
\alpha_1 &= -2\,\alpha_2 = -2\, d_4^2\, u_2^2\, u_4, \\
\alpha_3 &= -\alpha_5 = -4\, d_2^2\, u_1\, u_2\, u_4, \\
\alpha_4 &= u_2^2 \left(d_1^2\, u_1 - d_4^2\, u_4\right) - 4\, d_2^2\, u_1\, u_2\, u_4.
\end{aligned}$$

Thus we obtain the system of two equations $\partial\varphi(\omega_1, \omega_2)/\partial\omega_1 = 0$, $\partial\varphi(\omega_1, \omega_2)/\partial\omega_2 = 0$. Straightforward computations show that the solution of the above system is the optimal weights ω_ℓ^* $(1 \leq \ell \leq 4)$ presented by the lemma. Hence, these optimal weights minimizing $\varphi(\omega_1, \omega_2)$. □

Remark 3.1.1. *As a consequence of Lemma 3.1.2, let* $h, i, j, k \in \{1, 2, 3, 4\}$ *be pairwise distinct. Assume there are two design points* $\boldsymbol{x}_i^*$ *and* $\boldsymbol{x}_j^*$, *say, such that* $u_i = u_j$ *and* $d_i^2 = d_j^2$. *Then the optimal weights are given in the following*

$$\omega_h^* = \frac{3}{8} + \frac{1}{4}\left(1 + \frac{d_h^2 u_h}{d_k^2 u_k} - 4\frac{d_i^2 u_h}{d_k^2 u_i}\right)^{-1},$$
$$\omega_i^* = \omega_j^* = \frac{1}{2}\left(4 - \frac{d_k^2 u_i}{d_i^2 u_h} - \frac{d_h^2 u_i}{d_i^2 u_k}\right)^{-1},$$
$$\omega_k^* = \frac{3}{8} + \frac{1}{4}\left(1 + \frac{d_k^2 u_k}{d_h^2 u_h} - 4\frac{d_i^2 u_k}{d_h^2 u_i}\right)^{-1}.$$

Remark 3.1.2. *Note that Lemma 3.1.1 and Lemma 3.1.2 can be applied even for generalized linear models without intercept.*

Moreover, saturated designs under Kiefer Φ_k-criteria for a GLM without intercept are of our interest, in specific, under the first order model $\boldsymbol{f}(\boldsymbol{x}) = (x_1, \ldots, x_\nu)^\top$ and parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\nu)^\top$. Therefore, the choice of locally Φ_k-optimal weights which yields the minimum value of $\Phi_k(\xi, \boldsymbol{\beta})$ over all saturated designs with the same support are given by the next lemma.

Lemma 3.1.3. *Consider a GLM without intercept with $\boldsymbol{f}(\boldsymbol{x}) = (x_1, \ldots, x_\nu)^\top$ on the experimental region $\mathcal{X}$. Denote by $\boldsymbol{e}_i$ for all $(1 \le i \le \nu)$ the ν-dimensional unit vectors. Let $\boldsymbol{x}_i^* = a_i\, \boldsymbol{e}_i$, $a_i > 0$ for all $(1 \le i \le \nu)$ be design points in $\mathcal{X}$ such that the vectors $\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_\nu^*)$ are linearly independent. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\nu)^\top$ be a given parameter point. Let $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta})$ for all $(1 \le i \le \nu)$. For a given positive real vector $\boldsymbol{a} = (a_1, \ldots, a_\nu)^\top$ the design $\xi_{\boldsymbol{a}}^*$ which achieves the minimum value of $\Phi_k(\xi_{\boldsymbol{a}}, \boldsymbol{\beta})$ over all designs $\xi_{\boldsymbol{a}}$ with $\mathrm{supp}(\xi_{\boldsymbol{a}}^*) = \{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_\nu^*\}$ assigns weights*

$$\omega_i^* = \frac{(a_i^2 u_i)^{\frac{-k}{k+1}}}{\sum\limits_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}} \quad (1 \le i \le \nu)$$

to the corresponding design points in $\{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_\nu^*\}$.

For D-optimality $(k = 0)$, $\omega_i^* = 1/\nu \quad (1 \le i \le \nu)$.

For A-optimality $(k = 1)$, $\omega_i^* = \frac{(a_i^2 u_i)^{-1/2}}{\sum\limits_{j=1}^{\nu} (a_j^2 u_j)^{-1/2}} \quad (1 \le i \le \nu)$.

For E-optimality $(k \to \infty)$, $\omega_i^* = \frac{(a_i^2 u_i)^{-1}}{\sum\limits_{j=1}^{\nu} (a_j^2 u_j)^{-1}} \quad (1 \le i \le \nu)$.

Proof. Define the $\nu \times \nu$ design matrix $\boldsymbol{F} = \mathrm{diag}(a_i)_{i=1}^{\nu}$ with the $\nu \times \nu$ weight matrix $\boldsymbol{V} = \mathrm{diag}(u_i\omega_i)_{i=1}^{\nu}$. Then we have $\boldsymbol{M}\big(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}\big) = \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F} = \mathrm{diag}(a_i^2 u_i \omega_i)_{i=1}^{\nu}$ and $\boldsymbol{M}^{-k}\big(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}\big) = \mathrm{diag}\big((a_i^2 u_i \omega_i)^{-k}\big)_{i=1}^{\nu}$ with $\mathrm{tr}\big(\boldsymbol{M}^{-k}(\xi_{\boldsymbol{a}}, \boldsymbol{\beta})\big) = \sum\limits_{i=1}^{\nu} (a_i^2 u_i \omega_i)^{-k}$. Note that the eigenvalues of $\boldsymbol{M}^{-k}\big(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}\big)$ are given by $\lambda_i(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}) = (a_i^2 u_i \omega_i)^{-k} \quad (1 \le i \le \nu)$. Thus the Kiefer Φ_k-criteria are defined as

$$\Phi_k(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}) = \Big(\frac{1}{\nu}\sum_{i=1}^{\nu}(a_i^2 u_i \omega_i)^{-k}\Big)^{\frac{1}{k}} \quad (0 < k < \infty).$$

Now we aim at minimizing $\Phi_k(\xi_{\boldsymbol{a}}, \boldsymbol{\beta})$ such that $\omega_i > 0$ and $\sum\limits_{i=1}^{\nu} \omega_i = 1$. Then we write $\omega_\nu = 1 - \sum\limits_{i=1}^{\nu-1} \omega_i$ and thus we obtain

$$\Phi_k(\xi_{\boldsymbol{a}}, \boldsymbol{\beta}) = \frac{1}{\nu^{1/k}}\Big((a_\nu^2 u_\nu)^{-k}(1 - \sum_{i=1}^{\nu-1}\omega_i)^{-k} + \sum_{i=1}^{\nu-1}(a_i^2 u_i \omega_i)^{-k}\Big)^{\frac{1}{k}}.$$

It follows that the equation $\frac{\partial \Phi_k(\xi_{\boldsymbol{a}}, \boldsymbol{\beta})}{\partial \omega_i} = 0$ is equivalent to

$$\Big(1+(a_\nu^2 u_\nu)^k(1-\sum_{i=1}^{\nu-1}\omega_i)^k \sum_{i=1}^{\nu-1}(a_i^2 u_i \omega_i)^{-k}\Big)^{\frac{1}{k}-1}\left(\frac{(a_i^2 u_i)^k \omega_i^{k+1} - (a_\nu^2 u_\nu)^k(1 - \sum\limits_{i=1}^{\nu-1}\omega_i)^{k+1}}{a_\nu^2 u_\nu\,(a_i^2 u_i)^k \omega_i^{k+1}(1 - \sum\limits_{i=1}^{\nu-1}\omega_i)^2}\right) = 0.$$

The l.h.s. of the above equation is a multiplication of two quantities. The first one as an equation; $\Big(1 + (a_\nu^2 u_\nu)^k(1 - \sum\limits_{i=1}^{\nu-1}\omega_i)^k \sum\limits_{i=1}^{\nu-1}(a_i^2 u_i \omega_i)^{-k}\Big)^{\frac{1}{k}-1} = 0$ has no solution. From the other one we get

$$(a_i^2 u_i)^k \omega_i^{k+1} - (a_\nu^2 u_\nu)^k(1 - \sum_{i=1}^{\nu-1}\omega_i)^{k+1} = 0$$

which gives $\omega_i = \Big(a_\nu^2 u_\nu/(a_i^2 u_i)\Big)^{\frac{k}{k+1}}\omega_\nu$ $(1 \le i \le \nu-1)$, thus $\omega_i\,(a_i^2 u_i)^{\frac{k}{k+1}} = \omega_\nu\,(a_\nu^2 u_\nu)^{\frac{k}{k+1}}$ $(1 \le i \le \nu-1)$. This means $\omega_i\,(a_i^2 u_i)^{\frac{k}{k+1}}$ $(1 \le i \le \nu)$ are all equal, i.e., $\omega_i\,(a_i^2 u_i)^{\frac{k}{k+1}} = c$ $(1 \le i \le \nu)$, where $c > 0$. It implies that $\omega_i = c\,(a_i^2 u_i)^{\frac{-k}{k+1}}$ $(1 \le i \le \nu)$. Since $\sum\limits_{i=1}^{\nu}\omega_i = 1$ we get $\sum\limits_{i=1}^{\nu} c\,(a_i^2 u_i)^{\frac{-k}{k+1}} = c\sum\limits_{i=1}^{\nu}(a_i^2 u_i)^{\frac{-k}{k+1}} = 1$, and thus $c = \Big(\sum\limits_{i=1}^{\nu}(a_i^2 u_i)^{\frac{-k}{k+1}}\Big)^{-1}$. So we finally obtain $\omega_i = (a_i^2 u_i)^{\frac{-k}{k+1}}/\Big(\sum\limits_{j=1}^{\nu}(a_j^2 u_j)^{\frac{-k}{k+1}}\Big)$ for all $(1 \le i \le \nu)$ which are the optimal weights given by the lemma. □

3.2 Single-factor model

In this section we concentrate on the simplest case for which the model is composed by a single factor through the linear predictor

$$\eta(x, \boldsymbol{\beta}) = \boldsymbol{f}^\top(x)\boldsymbol{\beta} = \beta_0 + \beta_1 x,\ x \in \mathcal{X}.$$

We begin with the discrete experimental region $\mathcal{X} = \{a, b\}$, $a, b \in \mathbb{R}$, i.e., the factor x is binary. In another situation, we consider the continuous experimental region given by the unit interval $\mathcal{X} = [0, 1]$. In each situation we provide locally D- and A-optimal designs.

Theorem 3.2.1. *Consider model* $\boldsymbol{f}(x) = \big(1, x\big)^\top$ *and experimental region* $\mathcal{X} = \{a, b\}$ *with real numbers* a, b*. Let a parameter point* $\boldsymbol{\beta} = (\beta_0, \beta)^\top$ *be given. Let* $u_a = u(a, \boldsymbol{\beta})$ *and* $u_b = u(b, \boldsymbol{\beta})$*. Then:*

(i) The unique locally D-optimal design (at $\boldsymbol{\beta}$*) is the two-point design supported by* a *and* b *with equal weights* $1/2$*.*

(ii) The unique locally A-optimal design (at $\boldsymbol{\beta}$*) is the two-point design supported by* a *and* b *with weights*

$$\omega_a^* = \frac{u_a^{-1/2}\sqrt{1+b^2}}{u_a^{-1/2}\sqrt{1+b^2} + u_b^{-1/2}\sqrt{1+a^2}}, \quad \omega_b^* = 1 - \omega_a^*.$$

Proof. Any D-optimal design and any A-optimal design must have support equal to $\{a, b\}$. In particular, they are saturated designs. Hence the unique D-optimal design gives equal weights $1/2$ to a and b. The weights of the A-optimal design, by Lemma 3.1.1, are obtained from the diagonal entries c_{11} and c_{22} of the matrix

$$\boldsymbol{C} = (\boldsymbol{F}^{-1})^T \boldsymbol{F}^{-1}, \text{ where } \boldsymbol{F} = \begin{bmatrix} 1 & a \\ 1 & b \end{bmatrix}.$$

From $\boldsymbol{F}^{-1} = \dfrac{1}{b-a}\begin{bmatrix} b & -a \\ -1 & 1 \end{bmatrix}$ we obtain $\boldsymbol{C} = \dfrac{1}{(b-a)^2}\begin{bmatrix} 1+b^2 & -(1+ab) \\ -(1+ab) & 1+a^2 \end{bmatrix}$.

So, by Lemma 3.1.1, $\omega_1^* = \sqrt{1+b^2}\,u_a^{-1/2}(b-a)^{-1}/c$, where $c = (b-a)^{-1}\big(\sqrt{1+b^2}\,u_a^{-1/2} + \sqrt{1+a^2}\,u_b^{-1/2}\big)$, which is the same as ω_a^* stated in the theorem. □

The locally D-optimal design given in the previous theorem is independent of the intensities, i.e., it is the same for all generalized linear models. Similar results for Poisson models were indicated in Wang et al. (2006). Furthermore, for each setup (or each intensity form) of a generalized linear model there is a locally A-optimal design that even varies with parameter values. Since a and b are the only design points there is a locally D- or A-optimal design at any parameter value in the parameter space of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$.

Let the experimental region is taken to be the continues unit interval $\mathcal{X} = [0, 1]$. In the following we introduce, for a fixed $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, the function

$$h(x) = \frac{1}{u(x, \boldsymbol{\beta})}, \ x \in [0, 1],$$

which will be utilized for the characterization of the optimal designs. Consider the following conditions:
(i) $u(x, \boldsymbol{\beta})$ is positive and twice continuously differentiable on $[0, 1]$.
(ii) $u(x, \boldsymbol{\beta})$ is strictly increasing on $[0, 1]$.
(iii) $h''(x)$ is an injective (one-to-one) function on $[0, 1]$.
Recently, Lemma 1 in Konstantinou, Biedermann, and Kimber (2014) showed that under the above conditions (i)-(iii) with $h(x) = 2/u(x, \boldsymbol{\beta})$ a locally D-optimal design on $[0, 1]$ is only supported by two points a and b where $0 \leq a < b \leq 1$. In what follows analogous result is presented for locally optimal designs under various optimality criteria.

Lemma 3.2.1. *Consider model $\boldsymbol{f}(x) = (1, x)^\top$ and experimental region $\mathcal{X} = [0, 1]$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \beta)^\top$ be given. Let conditions (i)-(iii) be satisfied. Denote by $\boldsymbol{A}$ a positive definite matrix and let c be constant. Then if the condition of The General Equivalence Theorem is of the form*

$$u(x, \boldsymbol{\beta})\boldsymbol{f}^\top(x)\boldsymbol{A}\boldsymbol{f}(x) \leq c$$

then the support points of a locally optimal design ξ^* is concentrated on exactly 2 points a and b where $0 \leq a < b \leq 1$.

Proof. Let $\boldsymbol{A} = [a_{ij}]_{i,j=1,2}$. Then let $p(x) = \boldsymbol{f}^\top(x)\boldsymbol{A}\boldsymbol{f}(x) = a_{22}x^2 + 2a_{12}x + a_{11}$ which is a polynomial in x of degree 2 where $x \in \mathcal{X}$. Hence, by The Equivalence Theorem ξ^* is locally optimal (at $\boldsymbol{\beta}$) if and only if

$$p(x) \leq h(x) \text{ for all } x \in [0, 1].$$

The above inequality is similar to that obtained in the proof of Lemma 1 in Konstantinou, Biedermann, and Kimber (2014) and thus the rest of our proof is analogous to that. □

Accordingly, for D-optimality we have $c = 2$, $\boldsymbol{A} = \boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})$ and equal weights 1/2. For A-optimality, $c = \mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})) = \left(\sqrt{(a^2+1)/u_b} + \sqrt{(b^2+1)/u_a}\right)/(b-a)^2$ where $u_a = u(a, \boldsymbol{\beta})$ and $u_b = u(b, \boldsymbol{\beta})$ with $\boldsymbol{A} = \boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta})$ and optimal weights as what are given in part (ii) of Theorem 3.2.1. In general, under Kiefer's Φ_k-criteria we denote $c = \mathrm{tr}(\boldsymbol{M}^{-k}(\xi^*, \boldsymbol{\beta}))$ and $\boldsymbol{A} = \boldsymbol{M}^{-k-1}(\xi^*, \boldsymbol{\beta})$ where the optimal weights might be obtained by minimizing the Kiefer Φ_k-criteria. Moreover, the Generalized D-criterion and L-criterion can be applied (Atkinson and Woods (2015), Chapter 10).

Remark 3.2.1. *For a GLM of multiple factors and experimental region given by a polytope, Schmidt (2019), Lemma 2, showed that the support points of an optimal design are located at the edges of the experimental region.*

As a consequence of Lemma 3.2.1, we next provide sufficient conditions for a design whose support is the boundaries of $[0,1]$, i.e., 0 and 1 to be locally D- or A-optimal on $\mathcal{X} = [0,1]$ at a given $\boldsymbol{\beta}$. Let $q(x) = 1/u(x,\boldsymbol{\beta})$, $q_0 = q^{\frac{1}{2}}(0)$ and $q_1 = q^{\frac{1}{2}}(1)$.

Theorem 3.2.2. *Consider model $\boldsymbol{f}(x) = \big(1, x\big)^\top$ and experimental region $\mathcal{X} = [0,1]$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \beta)^\top$ be given. Let $q(x)$ be positive, twice continuously differentiable. Then:*

(i) The unique locally D-optimal design (at $\boldsymbol{\beta}$) is the two-point design supported by 0 and 1 with equal weights 1/2 if

$$q_0^2 + q_1^2 > q''(x)/2 \text{ for all } x \in (0,1). \tag{3.3}$$

(ii) The unique locally A-optimal design (at $\boldsymbol{\beta}$) is the two-point design supported by 0 and 1 with weights

$$\omega_0^* = \frac{\sqrt{2}q_0}{\sqrt{2}q_0 + q_1} \text{ and } \omega_1^* = \frac{q_1}{\sqrt{2}q_0 + q_1}, \text{ respectively}$$

if

$$q_0^2 + q_1^2 + \sqrt{2}q_0 q_1 > q''(x)/2 \text{ for all } x \in (0,1). \tag{3.4}$$

Proof. Ad (*i*) Employing condition (2.11) of The Equivalence Theorem (Theorem 2.2.2) implies that ξ^* is locally D-optimal if and only if

$$(1-x)^2 q_0^2 + x^2 q_1^2 - q(x) \le 0 \ \forall x \in [0,1]. \tag{3.5}$$

Since the support points are $\{0,1\}$, the l.h.s. of the above inequality equals zero at the boundaries of $[0,1]$. Then it is sufficient to show that the aforementioned l.h.s. is convex on the interior $(0,1)$ and this convexity realizes under condition (3.3) asserted in the theorem. Now to show that ξ^* is unique at $\boldsymbol{\beta}$ assume that ξ^{**} is locally D-optimal at $\boldsymbol{\beta}$. Then $\boldsymbol{M}(\xi^*,\boldsymbol{\beta}) = \boldsymbol{M}(\xi^{**},\boldsymbol{\beta})$ and therefore, the condition of the equivalence theorem under ξ^{**} is equivalent to (3.5) and this is an equation only at the support of ξ^*, i.e., 0 and 1.

Ad (*ii*) This case can be shown in analogy to case (*i*) by employing condition (2.12) of The Equivalence Theorem (Theorem 2.2.2) with $\mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})) = (\sqrt{2}q_0 + q_1)^2$. □

3.3 Model without interaction

In this section we consider the model with multiple factors without interactions. More precise, a first order model is employed

$$\boldsymbol{f}(\boldsymbol{x}) = \left(1, \boldsymbol{x}^\top\right)^\top, \quad \boldsymbol{x} \in \mathcal{X}; \tag{3.6}$$

where the linear predictor is determined by $\eta(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} = \beta_0 + \sum_{i=1}^{\nu} \beta_i x_i$ with binary factors. That is a discrete experimental region is considered and has the form $\mathcal{X} = \{0,1\}^\nu, \nu \geq 2$. We aim at constructing locally D- and A-optimal designs for a given parameter point $\boldsymbol{\beta}$ adopting particular analytic solutions.

To this end, we firstly begin with a two-factor model

$$\boldsymbol{f}(\boldsymbol{x}) = \left(1, x_1, x_2\right)^\top \text{ where } \boldsymbol{x} = (x_1, x_2)^\top \in \mathcal{X} = \{0,1\}^2. \tag{3.7}$$

The experimental region can be written as $\mathcal{X} = \{(0,0)^\top, (1,0)^\top (0,1)^\top, (1,1)^\top\}$. Let us denote the design points by $\boldsymbol{x}_1^* = (0,0)^\top$, $\boldsymbol{x}_2^* = (1,0)^\top$, $\boldsymbol{x}_3^* = (0,1)^\top$, and $\boldsymbol{x}_4^* = (1,1)^\top$.

Theorem 3.3.1. *Consider model (3.7) and experimental region $\mathcal{X} = \{0,1\}^2$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ be given. Denote $u_k = u(\boldsymbol{x}_k^*, \boldsymbol{\beta})$, $\boldsymbol{x}_k^* \in \mathcal{X}$ $(1 \leq k \leq 4)$, and denote by $u_{(1)} \leq u_{(2)} \leq u_{(3)} \leq u_{(4)}$ the intensity values u_1, u_2, u_3, u_4 rearranged in ascending order. Then:*

(o) The locally D-optimal design ξ^* (at $\boldsymbol{\beta}$) is unique.

(i) If $u_{(1)}^{-1} \geq u_{(2)}^{-1} + u_{(3)}^{-1} + u_{(4)}^{-1}$ then ξ^* is a three-point design supported by the three design points whose intensity values are given by $u_{(2)}$, $u_{(3)}$, $u_{(4)}$, with equal weights $1/3$.

(ii) If $u_{(1)}^{-1} < u_{(2)}^{-1} + u_{(3)}^{-1} + u_{(4)}^{-1}$ then ξ^* is a four-point design supported by the four design points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$ with weights $\omega_1^*, \omega_2^*, \omega_3^*, \omega_4^*$ which are uniquely determined by the condition

$$\omega_k^* > 0 \ (1 \leq k \leq 4), \ \sum_{k=1}^{4} \omega_k^* = 1, \ \textit{and} \ \ u_k \omega_k^* \Big(\tfrac{1}{3} - \omega_k^*\Big) \ (1 \leq k \leq 4) \ \textit{are equal}. \tag{3.8}$$

Proof. Ad (o) We know that the information matrix $\boldsymbol{M}(\xi^*, \boldsymbol{\beta})$ of a locally D-optimal design ξ^* (at $\boldsymbol{\beta}$) is unique. To show uniqueness of the locally D-optimal design (at $\boldsymbol{\beta}$) it suffices to show that if $\sum_{k=1}^{4} \omega_k u_k \boldsymbol{f}(\boldsymbol{x}_k^*) \boldsymbol{f}^\top(\boldsymbol{x}_k^*) = \sum_{k=1}^{4} \widetilde{\omega}_k u_k \boldsymbol{f}(\boldsymbol{x}_k^*) \boldsymbol{f}^\top(\boldsymbol{x}_k^*)$ then $\omega_k = \widetilde{\omega}_k$ $(1 \leq k \leq 4)$. Since the intensities u_k $(1 \leq k \leq 4)$ are positive it suffices to show that the four information matrices $\boldsymbol{M}_k = \boldsymbol{f}(\boldsymbol{x}_k^*) \boldsymbol{f}^\top(\boldsymbol{x}_k^*)$ $(1 \leq k \leq 4)$ are linearly independent. This is straightforward to verify in view of

$$\boldsymbol{M}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \boldsymbol{M}_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \boldsymbol{M}_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \ \boldsymbol{M}_4 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

For the proof of (i) and (ii) we will use the following three auxiliary statements (3.9), (3.10) and (3.11).

$$\boldsymbol{f}(\boldsymbol{x}_1^*) - \boldsymbol{f}(\boldsymbol{x}_2^*) - \boldsymbol{f}(\boldsymbol{x}_3^*) + \boldsymbol{f}(\boldsymbol{x}_4^*) \ = \ \boldsymbol{0}; \tag{3.9}$$

$$\det\Big([\boldsymbol{f}(\boldsymbol{x}_h^*), \boldsymbol{f}(\boldsymbol{x}_i^*), \boldsymbol{f}(\boldsymbol{x}_j^*)]\Big) = \pm 1 \ \text{ for all pairwise distinct } h, i, j \in \{1,2,3,4\}; \tag{3.10}$$

$$\text{For } \boldsymbol{A} \text{ positive definite } p \times p \text{ and } \boldsymbol{b} \in \mathbb{R}^p\text{:} \ \ \boldsymbol{b}^\top\Big(\boldsymbol{A} + \boldsymbol{b}\boldsymbol{b}^\top\Big)^{-1}\boldsymbol{b} = \frac{\boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}}{1 + \boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}}. \tag{3.11}$$

Equations (3.9) and (3.10) are straightforward to verify, and (3.11) is obtained from the Sherman-Morrison formula on the inverse of $\boldsymbol{A} + \boldsymbol{b}\boldsymbol{b}^\top$, see Bartlett (1951).
<u>Ad (i), (ii):</u> We show that the locally D-optimal design ξ^* is a three-point design if and only if

$$u_{(1)}^{-1} \geq u_{(2)}^{-1} + u_{(3)}^{-1} + u_{(4)}^{-1} \tag{3.12}$$

in which case ξ^* is supported by the three design points whose intensity values are given by $u_{(2)}$, $u_{(3)}$ and $u_{(4)}$, and their weights are all equal to 1/3. To this end consider a three-point design ξ supported by $\boldsymbol{x}_h^*$, $\boldsymbol{x}_i^*$, $\boldsymbol{x}_j^*$ for some $1 \leq h < i < j \leq 3$ and with equal weights 1/3. Then

$$\begin{aligned} &\boldsymbol{M}(\xi^*, \boldsymbol{\beta}) = (1/3)\, \boldsymbol{F}^\top \boldsymbol{U} \boldsymbol{F}, \quad \text{where} \\ &\boldsymbol{F}^\top = \Big[\boldsymbol{f}(\boldsymbol{x}_h^*)\,,\ \boldsymbol{f}(\boldsymbol{x}_i^*)\,,\ \boldsymbol{f}(\boldsymbol{x}_j^*)\Big] \ \text{ and } \ \boldsymbol{U} = \operatorname{diag}(u_h, u_i, u_j). \end{aligned}$$

So, by The Equivalence Theorem (Theorem 2.2.2, condition (2.11)) ξ is locally D-optimal (at $\boldsymbol{\beta}$) if and only if

$$u_\ell \Big(\Big(\boldsymbol{F}^\top\Big)^{-1} \boldsymbol{f}(\boldsymbol{x}_\ell^*)\Big)^\top \boldsymbol{U}^{-1} \Big(\boldsymbol{F}^\top\Big)^{-1} \boldsymbol{f}(\boldsymbol{x}_\ell^*) \ \leq 1 \ \forall\ \ell = 1,2,3,4. \tag{3.13}$$

For $\ell \in \{h, i, j\}$ one has $\Big(\boldsymbol{F}^\top\Big)^{-1}\boldsymbol{f}(\boldsymbol{x}_\ell^*) = \boldsymbol{e}_\ell$ where $\boldsymbol{e}_\ell$ denotes the ℓ-th elementary unit vector in $\mathbb{R}^3$, and hence the left hand side of (3.13) is equal to 1. So (3.13) is equivalent to the single inequality for the index $\ell \in \{1,2,3,4\} \setminus \{h, i, j\}$. From (3.9) it follows that $\boldsymbol{f}(\boldsymbol{x}_\ell^*) = \boldsymbol{F}^\top \boldsymbol{z}$ for some vector $\boldsymbol{z} = (z_1, z_2, z_3)^\top \in \mathbb{R}^3$ with $z_s \in \{\pm 1\}$ $(1 \leq s \leq 3)$. Hence $\boldsymbol{z} = \Big(\boldsymbol{F}^\top\Big)^{-1}\boldsymbol{f}(\boldsymbol{x}_\ell^*)$ and the l.h.s. of (3.13) reads

$$u_\ell \boldsymbol{z}^\top \boldsymbol{U}^{-1} \boldsymbol{z} = u_\ell \Big(u_h^{-1} + u_i^{-1} + u_j^{-1}\Big),$$

and we see that (3.13) is equivalent to $u_\ell^{-1} \ge u_h^{-1} + u_i^{-1} + u_j^{-1}$. From this we see that ξ^* is a three-point design whose support in $\mathcal{X}$ if and only if (3.12) holds. Now assume $u_{(1)}^{-1} < u_{(2)}^{-1}+u_{(3)}^{-1}+u_{(4)}^{-1}$. Then ξ^* is a four-point design with support points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$. It remains to show that a four-point design ξ with support points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$ and positive weights $\omega_1, \omega_2, \omega_3, \omega_4$ is locally D-optimal (at $\boldsymbol{\beta}$) if and only if $u_k\omega_k\left(\frac{1}{3} - \omega_k\right)$ are equal for $k = 1, 2, 3, 4$. Again, by The Equivalence Theorem, local D-optimality of ξ is equivalent to

$$u_\ell\, \boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\, \boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\, \boldsymbol{f}(\boldsymbol{x}_\ell^*) \;=\; 3 \;\; \forall\; \ell = 1, 2, 3, 4. \tag{3.14}$$

Note that the equality in (3.14) is due to Remark 2.2.6. For a given $\ell \in \{1, 2, 3, 4\}$ let $\boldsymbol{A}_\ell = \sum_{k=1,\, k\neq\ell}^{4} \omega_k u_k \boldsymbol{f}(\boldsymbol{x}_k^*)\boldsymbol{f}^\top(\boldsymbol{x}_k^*)$ which is positive definite by (3.10), and $\boldsymbol{b}_\ell = \sqrt{\omega_\ell u_\ell}\,\boldsymbol{f}(\boldsymbol{x}_\ell^*)$. Then

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{A}_\ell + \boldsymbol{b}_\ell \boldsymbol{b}_\ell^\top,$$

and hence, using formula (3.11),

$$u_\ell \boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\boldsymbol{f}(\boldsymbol{x}_\ell^*) = (1/\omega_\ell)\boldsymbol{b}_\ell^\top\left(\boldsymbol{A}_\ell + \boldsymbol{b}_\ell\boldsymbol{b}_\ell^\top\right)^{-1}\boldsymbol{b}_\ell = \frac{u_\ell \boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\boldsymbol{A}_\ell^{-1}\boldsymbol{f}(\boldsymbol{x}_\ell^*)}{1 + \omega_\ell u_\ell \boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\boldsymbol{A}_\ell^{-1}\boldsymbol{f}(\boldsymbol{x}_\ell^*)}.$$

Let $1 \le h < i < j \le 4$ be such that $\{\ell, h, i, j\} = \{1, 2, 3, 4\}$. Then

$$\boldsymbol{A}_\ell = \left[\boldsymbol{f}(\boldsymbol{x}_h^*), \boldsymbol{f}(\boldsymbol{x}_i^*), \boldsymbol{f}(\boldsymbol{x}_j^*)\right]\boldsymbol{V}\left[\boldsymbol{f}(\boldsymbol{x}_h^*), \boldsymbol{f}(\boldsymbol{x}_i^*), \boldsymbol{f}(\boldsymbol{x}_j^*)\right]^\top, \quad \text{where } \boldsymbol{V} = \mathrm{diag}(\omega_h u_h, \omega_i u_i, \omega_j u_j).$$

From (3.9) we get $\left[\boldsymbol{f}(\boldsymbol{x}_h^*), \boldsymbol{f}(\boldsymbol{x}_i^*), \boldsymbol{f}(\boldsymbol{x}_j^*)\right]\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}_\ell^*)$ for some $\boldsymbol{z} \in \mathbb{R}^3$ having components in $\{\pm 1\}$. Hence

$$\begin{aligned}
&\boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\boldsymbol{A}_\ell^{-1}\boldsymbol{f}(\boldsymbol{x}_\ell^*) = \boldsymbol{z}^\top\boldsymbol{V}^{-1}\boldsymbol{z} = \sum_{\substack{k=1\\k\neq\ell}}^{4}(\omega_k u_k)^{-1} = \lambda(\xi) - (\omega_\ell u_\ell)^{-1},\\
&\text{where} \quad \lambda(\xi) = \sum_{k=1}^{4}(\omega_k u_k)^{-1},
\end{aligned} \tag{3.15}$$

and we have obtained

$$u_\ell \boldsymbol{f}^\top(\boldsymbol{x}_\ell^*)\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\boldsymbol{f}(\boldsymbol{x}_\ell^*) = \frac{u_\ell\Big(\lambda(\xi) - (\omega_\ell u_\ell)^{-1}\Big)}{1 + \omega_\ell u_\ell\Big(\lambda(\xi) - (\omega_\ell u_\ell)^{-1}\Big)} = \frac{u_\ell\lambda(\xi) - \omega_\ell^{-1}}{\omega_\ell u_\ell \lambda(\xi)}.$$

So (3.14) is equivalent to

$$\frac{u_\ell\lambda(\xi) - \omega_\ell^{-1}}{\omega_\ell u_\ell \lambda(\xi)} \;=\; 3 \;\; \forall\; \ell = 1, 2, 3, 4 \text{ (with } \lambda(\xi) \text{ from (3.15))}. \tag{3.16}$$

(3.16) rewrites equivalently as $u_\ell\omega_\ell\big(\frac{1}{3}-\omega_\ell\big)=1/\big(3\lambda(\xi)\big)$ for all $\ell=1,2,3,4$. Utilizing that $\omega_k>0$ $(1\le k\le 4)$, $\sum_{k=1}^{4}\omega_k=1$, and the definition (3.15) of $\lambda(\xi)$ one obtains that these equations can equivalently be stated as

$$u_\ell\omega_\ell\Big(\frac{1}{3}-\omega_\ell\Big)\ \text{ are equal for } \ell=1,2,3,4.$$

□

Remark 3.3.1.

1. *It is already seen from the optimality conditions asserted in part (i) of Theorem 3.3.1 that the design points with highest intensities perform as a support of a locally D-optimal design at a given parameter value.*
2. *The optimality condition asserted in part (ii) of Theorem 3.3.1 applies only when the optimality conditions for the three-point (saturated) designs in (i) cannot be satisfied.*

Theorem 3.3.1 covers various results in the literature. For examples; see Yang, Mandal, and Majumdar (2012) for binary responses with several link functions and see Graßhoff, Holling, and Schwabe (2013) for count data in item response theory.

In analogy to Theorem 3.3.1 we introduce locally A-optimal designs in the next theorem where also the design points with highest intensities perform as a support of a locally A-optimal design at a given parameter value.

Theorem 3.3.2. *Consider the assumptions and notations of Theorem 3.3.1. Denote* $q_i=u_i^{-1/2}$ $(1\le i\le 4)$. *Then:*

(o) The locally A-optimal design ξ^* *(at* $\boldsymbol\beta$*) is unique.*

(i) If $q_1^2\ge q_2^2+q_3^2+q_4^2+q_2q_3+2\sqrt{\frac{2}{3}}q_2q_4+2\sqrt{\frac{2}{3}}q_3q_4$ *then*

$$\xi^*=\begin{pmatrix}\boldsymbol{x}_2^* & \boldsymbol{x}_3^* & \boldsymbol{x}_4^*\\ \sqrt{2}q_2/c & \sqrt{2}q_3/c & \sqrt{3}q_4/c\end{pmatrix}.$$

(ii) If $q_2^2\ge q_1^2+q_3^2+q_4^2+q_1q_3+\sqrt{2}q_3q_4$ *then*

$$\xi^*=\begin{pmatrix}\boldsymbol{x}_1^* & \boldsymbol{x}_3^* & \boldsymbol{x}_4^*\\ \sqrt{2}q_1/c & \sqrt{2}q_3/c & q_4/c\end{pmatrix}.$$

(iii) If $q_3^2\ge q_1^2+q_2^2+q_4^2+q_1q_2+\sqrt{2}q_2q_4$ *then*

$$\xi^*=\begin{pmatrix}\boldsymbol{x}_1^* & \boldsymbol{x}_2^* & \boldsymbol{x}_4^*\\ \sqrt{2}q_1/c & \sqrt{2}q_2/c & q_4/c\end{pmatrix}.$$

(iv) If $q_4^2 \geq q_1^2 + q_2^2 + q_3^2 + \frac{2}{\sqrt{3}} q_1 q_2 + \frac{2}{\sqrt{3}} q_1 q_3$ *then*

$$\xi^* = \begin{pmatrix} \boldsymbol{x}_1^* & \boldsymbol{x}_2^* & \boldsymbol{x}_3^* \\ \sqrt{3} q_1/c & q_2/c & q_3/c \end{pmatrix}.$$

For each case (i) – (iv), the constant c appearing in the weights equals the sum of the numerators of the three ratios. If none of the cases (i) – (iv) applies then ξ^* is supported by the four design points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$.

Proof.
Ad (o): Similar to proof of part (o) of Theorem 3.3.1.
For the proof of (i) – (iv), we employ Lemma 3.1.1 where the experimental region $\mathcal{X} = \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*\}$ is employed. In each of the cases (i) – (iv) the design ξ^* stated in the theorem is a saturated design with $\text{supp}(\xi^*) = \mathcal{X} \setminus \{\boldsymbol{x}_\ell^*\}$, $\ell = 1, 2, 3, 4$ where the index ℓ corresponds to the case label. We will show that in each case the design ξ^* coincides with that of Lemma 3.1.1 and the inequality stated in the theorem coincides with the equivalent condition for local A-optimality of the design stated in Lemma 3.1.1. To this end, for each case (i) – (iv), we report the matrices $\boldsymbol{F}$, $\boldsymbol{F}^{-1}$, and $\boldsymbol{C} = (\boldsymbol{F}^{-1})^\top \boldsymbol{F}^{-1}$ from Lemma 3.1.1. Then the weights of ξ^* are easily verified to be those from the lemma and the condition for local A-optimality of ξ^* from the lemma gives that of the theorem.

$$(i): \quad \boldsymbol{F} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{F}^{-1} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -2 \\ -2 & -2 & 3 \end{pmatrix};$$

$$(ii): \quad \boldsymbol{F} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{F}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix};$$

$$(iii): \quad \boldsymbol{F} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{F}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix};$$

$$(iv): \quad \boldsymbol{F} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{F}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

□

In the following we consider model (3.6) for a general number of factors, $\nu \geq 2$, and with the experimental region $\mathcal{X} = \{0, 1\}^\nu$. Here, we are interested in providing an extension of particular locally D- and A-optimal designs given in the preceding theorems under a two-factor model. In specific, the three-point designs with support $(0,0)^\top, (1,0)^\top, (0,1)^\top$ are to be extended.

Theorem 3.3.3. *Consider model (3.6) with experimental region* $\mathcal{X} = \{0,1\}^\nu$*, where* $\nu \geq 2$*. Denote the design points by*

$$\boldsymbol{x}_1^* = (0,\ldots,0)^\top, \quad \boldsymbol{x}_2^* = (1,\ldots,0)^\top, \ \ldots, \ \boldsymbol{x}_{\nu+1}^* = (0,\ldots,1)^\top.$$

For a given parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_\nu)^\top$ *let* $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta})\,(1 \leq i \leq \nu+1)$*. Then the design* ξ^* *which assigns equal weights* $1/(\nu+1)$ *to the design points* $\boldsymbol{x}_i^*$ *for all* $(1 \leq i \leq \nu+1)$ *is locally D-optimal (at* $\boldsymbol{\beta}$*) if and only if*

$$u_1^{-1}\Big(1 - \sum_{j=1}^{\nu} x_j\Big)^2 + \sum_{i=1}^{\nu} u_{i+1}^{-1} x_i^2 \leq u^{-1}(\boldsymbol{x}, \boldsymbol{\beta}) \ \ \textit{for all} \ \ \boldsymbol{x} \in \Big\{0,1\Big\}^\nu. \tag{3.17}$$

Proof. Define the $(\nu+1)\times(\nu+1)$ design matrix $\boldsymbol{F} = \Big[\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_{\nu+1}^*)\Big]^\top$, then

$$\boldsymbol{M}(\xi^*, \boldsymbol{\beta}) = \frac{1}{\nu+1}\boldsymbol{F}^\top \boldsymbol{U} \boldsymbol{F}, \quad \text{where } \boldsymbol{U} = \operatorname{diag}\Big(u_i\Big)_{i=1}^{\nu+1}.$$

We have

$$\boldsymbol{F} = \begin{bmatrix} 1 & \boldsymbol{0}_{1\times\nu} \\ \boldsymbol{1}_{\nu\times 1} & \boldsymbol{I}_\nu \end{bmatrix}, \quad \text{hence } \boldsymbol{F}^{-1} = \begin{bmatrix} 1 & \boldsymbol{0}_{1\times\nu} \\ -\boldsymbol{1}_{\nu\times 1} & \boldsymbol{I}_\nu \end{bmatrix}, \tag{3.18}$$

where $\boldsymbol{0}_{1\times\nu}$, $\boldsymbol{1}_{\nu\times 1}$, and $\boldsymbol{I}_\nu$ denote the ν-dimensional row vector of zeros, the ν-dimensional column vector of ones, and the $\nu\times\nu$ unit matrix, respectively. So, by The Equivalence Theorem (Theorem 2.2.2, condition (2.11)) the design is locally D-optimal if and only if

$$u(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}^\top(\boldsymbol{x})\, \boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\, \boldsymbol{f}(\boldsymbol{x}) \leq \nu + 1 \ \ \forall \boldsymbol{x} \in \{0,1\}^\nu. \tag{3.19}$$

The l.h.s. of (3.19) reads as

$$u(\boldsymbol{x}, \boldsymbol{\beta})\,(\nu+1)\boldsymbol{f}^\top(\boldsymbol{x})\,\boldsymbol{F}^{-1}\boldsymbol{U}^{-1}\Big(\boldsymbol{F}^{-1}\Big)^\top \boldsymbol{f}(\boldsymbol{x}) =$$
$$(\nu+1)u(\boldsymbol{x}, \boldsymbol{\beta})\Big(u_1^{-1}\Big(1 - \sum_{j=1}^{\nu} x_j\Big)^2 + \sum_{i=1}^{\nu} u_{i+1}^{-1} x_i^2\Big),$$

and hence it is obvious that (3.19) is equivalent to (3.17). □

Remark 3.3.2. *The D-optimal design under a two-factor model with support* $(0,0)^\top$*,* $(1,0)^\top$*,* $(0,1)^\top$ *from Theorem 3.3.1 is covered by Theorem 3.3.3 for* $\nu = 2$ *where condition (3.17) is equivalent to the inequality* $u_4^{-1} \geq u_1^{-1} + u_2^{-1} + u_3^{-1}$ *that is asserted in part (i) of Theorem 3.3.1. Moreover, Theorem 3.3.3 covers various results in literature. For example; Russell et al. (2009) provided for the Poisson model a locally D-optimal* *saturated design on the continuous experimental region* $[0,1]^\nu, \nu \geq 2$ *that is supported by* $(0,\ldots,0)^\top$, $(1,\ldots,0)^\top$, $\ldots$, $(0,\ldots,1)^\top$ *at* $\beta_i = -2, (1 \leq i \leq \nu)$.

In analogy to Theorem 3.3.3 we introduce locally A-optimal designs in the next theorem.

Theorem 3.3.4. *Consider the assumptions and notations of Theorem 3.3.3. Denote* $q_i = u_i^{-1/2}$ $(1 \leq i \leq \nu+1)$. *Then the design* ξ^* *which is supported by* $\boldsymbol{x}_i^*$ $(1 \leq i \leq \nu+1)$ *with weights*

$$\omega_1 = \sqrt{\nu+1}q_1/c \quad \textit{and } \omega^*_{i+1} = q_{i+1}/c,\ i = 1,\ldots,\nu,\ c = \sqrt{\nu+1}q_1 + \sum_{i=2}^{\nu} q_i$$

is locally A-optimal (at $\boldsymbol{\beta}$*) if and only if for all* $\boldsymbol{x} = (x_1,\ldots,x_\nu)^\top \in \{0,1\}^\nu$

$$q_1^2\Big(1-\sum_{j=1}^{\nu}x_j\Big)^2 + \sum_{i=1}^{\nu}q_{i+1}^2x_i^2 + \frac{2q_1}{\sqrt{\nu+1}}\Big(\sum_{j=1}^{\nu}x_j - 1\Big)\sum_{i=1}^{\nu}q_{i+1}x_i \;\leq\; u^{-1}(\boldsymbol{x},\boldsymbol{\beta}). \qquad (3.20)$$

Proof. We show that the design ξ^* coincides with that from Lemma 3.1.1. As in the proof of Theorem 3.3.3 the design matrix $\boldsymbol{F}$ and its inverse are given by (3.18) and we obtain

$$\boldsymbol{C} = \left(\boldsymbol{F}^{-1}\right)^\top \boldsymbol{F}^{-1} = \begin{bmatrix} \nu+1 & -\mathbf{1}_{1\times\nu} \\ -\mathbf{1}_{\nu\times 1} & \boldsymbol{I}_\nu \end{bmatrix}.$$

This yields $\sqrt{c_{11}/u(\boldsymbol{x}_1^*,\boldsymbol{\beta})} = \sqrt{\nu+1}q_1$ and $\sqrt{c_{ii}/u(\boldsymbol{x}_i^*,\boldsymbol{\beta})} = q_i$ for $i = 2,\ldots,\nu+1$, and an elementary calculation shows that the weights given by Lemma 3.1.1 coincide with the ω_i^* $(1 \leq i \leq \nu+1)$ as stated in the theorem. Straightforward calculation shows that condition (3.1) that provides a characterization of local A-optimality of ξ^* is equivalent to (3.20). □

Remark 3.3.3. *Theorem 3.3.4 with* $\nu = 2$ *covers the result stated in case (iv) of Theorem 3.3.2. It can be checked that, with the notations of Theorem 3.3.2, the inequality* $q_4^2 > q_1^2 + q_2^2 + q_3^2 + \frac{2}{\sqrt{3}}q_1q_2 + \frac{2}{\sqrt{3}}q_1q_3$ *is equivalent to assumption (3.20) of Theorem 3.3.4 for* $\nu = 2$.

3.4 Model with interaction

We consider a first order model augmented by interaction terms which are products of two or more binary variables x_i $(1 \leq i \leq \nu)$,

$$\boldsymbol{f}(\boldsymbol{x}) \;=\; \left(1, \boldsymbol{x}^\top, g_{S_1}(\boldsymbol{x}),\ldots,g_{S_q}(\boldsymbol{x})\right)^\top, \qquad (3.21)$$

$$\text{where } S_j \subseteq \{1,\ldots,\nu\},\ \#S_j \geq 2, \text{ and } g_{S_j}(\boldsymbol{x}) = \prod_{i\in S_j} x_i \quad (1 \leq j \leq q).$$

Of course, we assume that the subsets S_j $(1 \leq j \leq q)$ in (3.21) are pairwise distinct which implies that the components of $\boldsymbol{f}$ are linearly independent functions on $\mathcal{X}$. Consider the full interaction model, i.e., the collection of subsets S_j $(1 \leq j \leq q)$ provide all subsets of $\{1, \ldots, \nu\}$ of size at least 2 (hence $q = 2^\nu - \nu - 1$). E.g., the full interaction models for $\nu = 2$ and $\nu = 3$ read as

$$\nu = 2: \quad \boldsymbol{f}(\boldsymbol{x}) = \Big(1\,,\, x_1\,,\, x_2\,,\, x_1x_2\Big)^\top, \quad \boldsymbol{x} = (x_1, x_2)^\top;$$
$$\nu = 3: \quad \boldsymbol{f}(\boldsymbol{x}) = \Big(1\,,\, x_1\,,\, x_2\,,\, x_3\,,\, x_1x_2\,,\, x_1x_3\,,\, x_2x_3\,,\, x_1x_2x_3\Big)^\top, \quad \boldsymbol{x} = (x_1, x_2, x_3)^\top.$$

First note that for full interaction models the dimension of $\boldsymbol{f}$ is $p = 2^\nu$. For a full interaction model a design supported by all the 2^ν vertices of a hyperrectangle $\mathcal{X}$ is minimally supported. The following result is immediate.

Theorem 3.4.1. *Consider the full interaction generalized linear model on the experimental region $\mathcal{X} = \{a, b\}^\nu$, $\nu \geq 2$. For any given parameter point $\boldsymbol{\beta}$ the unique locally D-optimal design (at $\boldsymbol{\beta}$) is supported by $\mathcal{X}$ with equal weights $2^{-\nu}$.*

As an example; Theorem 3.4.1 covers a result in Section 3 of Yang, Mandal, and Majumdar (2012) where a two-factor model with interaction $\boldsymbol{f}(\boldsymbol{x}) = \Big(1\,,\, x_1\,,\, x_2\,,\, x_1x_2\Big)^\top$ was considered for binary observations with logit, probit and complementary log-log link functions.

Remark 3.4.1. *For the full interaction model and local A-optimality the unique locally A-optimal design (at $\boldsymbol{\beta}$) is again supported by $\mathcal{X}$ but the weights will depend on $\boldsymbol{\beta}$ according to Lemma 3.1.1. For $\nu = 2$ and the set $\{0, 1\}^2$ as experimental region the locally A-optimal design is given explicitly by the following theorem.*

Theorem 3.4.2. *Consider the full interaction generalized linear model and $\mathcal{X} = \{0, 1\}^2$ as an experimental region, i.e., $\boldsymbol{f}(\boldsymbol{x}) = \big(1, x_1, x_2, x_1x_2\big)^\top$ for all $\boldsymbol{x} = (x_1, x_2)^\top \in \{0, 1\}^2$. Denote the design points by $\boldsymbol{x}_1^* = (0, 0)^\top$, $\boldsymbol{x}_2^* = (1, 0)^\top$, $\boldsymbol{x}_3^* = (0, 1)^\top$, and $\boldsymbol{x}_4^* = (1, 1)^\top$. For any given parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ denote $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta}), i = 1, 2, 3, 4$ and let $q_i = u_i^{-1/2}, i = 1, 2, 3, 4$. The unique locally A-optimal design (at $\boldsymbol{\beta}$) is supported by the points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$ with weights*

$$\omega_1^* = 2q_1/c, \quad \omega_2^* = \sqrt{2}q_2/c, \quad \omega_3^* = \sqrt{2}q_3/c, \quad \omega_4^* = q_4/c,$$

where $c = 2q_1 + \sqrt{2}q_2 + \sqrt{2}q_3 + q_4$.

Proof. The design ξ^* is saturated. Hence the unique locally A-optimal design (at $\boldsymbol{\beta}$) is given by the design ξ^* from Lemma 3.1.1 where the prescribed support points are given by $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*, \boldsymbol{x}_4^*$. For the matrices $\boldsymbol{F}$ and $\boldsymbol{C}$ in that lemma we obtain

$$\boldsymbol{F} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and } \boldsymbol{F}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

$$\text{hence } \boldsymbol{C} = \left(\boldsymbol{F}^{-1}\right)^{\top}\boldsymbol{F}^{-1} = \begin{bmatrix} 4 & -2 & -2 & 1 \\ -2 & 2 & 1 & -1 \\ -2 & 1 & 2 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

The weights $c^{-1}\left(c_{ii}/u(\boldsymbol{x}_i^*, \boldsymbol{\beta})\right)^{1/2}$ $(1 \le i \le 4)$ from Lemma 3.1.1 coincide with the ω_i^*, $i = 1, 2, 3, 4$ stated in the theorem. □

3.5 Model without intercept

In this section we consider GLMs having a first order linear predictor without intercept, i.e.,

$$\boldsymbol{f}(\boldsymbol{x}) = (x_1, \dots, x_\nu)^\top \text{ for all } \boldsymbol{x} = (x_1, \dots, x_\nu)^\top \in \mathcal{X}.$$

Here, $\boldsymbol{f}(\boldsymbol{0}) = \boldsymbol{0}$. We focus on locally optimal designs derived under Kiefer Φ_k-criteria and thus, our results implicitly cover the D- and A-optimal designs. In the following we provide a necessary and sufficient condition for constructing Φ_k-optimal designs. The optimal weights are obtained according to Lemma 3.1.3.

Theorem 3.5.1. *Consider the experimental region $\mathcal{X}$. Given a vector $\boldsymbol{a} = (a_1, \dots, a_\nu)^\top$ where $a_i \in \mathbb{R}$, $a_i > 0$ $(1 \le i \le \nu)$. Let $\boldsymbol{x}_i^* = a_i \boldsymbol{e}_i$ $(1 \le i \le \nu)$ denote the design points which are assumed to belong to $\mathcal{X}$. For a given parameter point $\boldsymbol{\beta}$ denote $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta})$ $(1 \le i \le \nu)$. Let $\xi_{\boldsymbol{a}}^*$ be the saturated design whose support is $\boldsymbol{x}_i^*$ $(1 \le i \le \nu)$ with the corresponding weights*

$$\omega_i^* = \frac{(a_i^2 u_i)^{\frac{-k}{k+1}}}{\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}} \quad (1 \le i \le \nu).$$

Then $\xi_{\boldsymbol{a}}^*$ is locally Φ_k-optimal (at $\boldsymbol{\beta}$) if and only if

$$u(\boldsymbol{x}, \boldsymbol{\beta}) \sum_{i=1}^{\nu} u_i^{-1} a_i^{-2} x_i^2 \le 1 \quad \text{for all } \boldsymbol{x} = (x_1, \dots, x_\nu)^\top \in \mathcal{X}. \tag{3.22}$$

Proof. Define the $\nu \times \nu$ design matrix $\boldsymbol{F} = \operatorname{diag}(a_i)_{i=1}^{\nu}$ with the $\nu \times \nu$ weight matrix

$$\boldsymbol{V} = \operatorname{diag}(u_i \omega_i^*)_{i=1}^{\nu} = \left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{-1} \operatorname{diag}\left((a_i^{-2k} u_i)^{\frac{1}{k+1}}\right)_{i=1}^{\nu}.$$

Then we have

$$\begin{aligned}
\boldsymbol{M}\left(\xi_{\boldsymbol{a}}^*, \boldsymbol{\beta}\right) &= \boldsymbol{F}^\top \boldsymbol{V} \boldsymbol{F} = \left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{-1} \operatorname{diag}\left((a_i^2 u_i)^{\frac{1}{k+1}}\right)_{i=1}^{\nu}, \\
\boldsymbol{M}^{-k-1}\left(\xi_{\boldsymbol{a}}^*, \boldsymbol{\beta}\right) &= \left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{k+1} \operatorname{diag}\left(a_i^{-2} u_i^{-1}\right)_{i=1}^{\nu}, \text{ and} \\
\operatorname{tr}\left(\boldsymbol{M}^{-k}\left(\xi_{\boldsymbol{a}}^*, \boldsymbol{\beta}\right)\right) &= \left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{k+1}.
\end{aligned}$$

Adopting these formulas simplifies the l.h.s. of The Equivalence Theorem (Theorem 2.2.2, condition (2.13)) to $u(\boldsymbol{x}, \boldsymbol{\beta})\left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{k+1} \sum_{i=1}^{\nu} u_i^{-1} a_i^{-2} x_i^2$ which is hence, bounded by $\left(\sum_{j=1}^{\nu} (a_j^2 u_j)^{\frac{-k}{k+1}}\right)^{k+1}$ if and only if condition (3.22) holds true. □

In particular, Theorem 3.5.1 states that for a given parameter point $\boldsymbol{\beta}$ the locally D-optimal design ($k = 0$) has wights $\omega_i^* = 1/\nu$ $(1 \le i \le \nu)$ and the locally A-optimal design ($k = 1$) has weights $\omega_i^* = \frac{(a_i^2 u_i)^{-1/2}}{\sum_{j=1}^{\nu} (a_j^2 u_j)^{-1/2}}$ $(1 \le i \le \nu)$.

Theorem 3.5.1 might be applicable for a wide class of GLMs on appropriate experimental regions. In case of gamma models the relevant aspect will be discussed in Chapter 4, Subsection 4.4.2.

For a Poisson model with intensity $u(\boldsymbol{x}, \boldsymbol{\beta}) = \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta}\right)$ and experimental region $\mathcal{X} = \{0, 1\}^\nu, \nu \ge 2$ let us restrict to the case of $a_i = 1$ $(1 \le i \le \nu)$, i.e., the design points are the unit vectors $\boldsymbol{e}_i$ $(1 \le i \le \nu)$. As a result, condition (3.22) is simplified in the following corollary.

Corollary 3.5.1. *Consider a non-intercept Poisson model with* $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ *on the experimental region* $\mathcal{X} = \{0, 1\}^\nu$, $\nu \ge 2$ *and intensity* $u(\boldsymbol{x}, \boldsymbol{\beta}) = \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta}\right)$. *For a given parameter point* $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\nu)^\top$ *define* $\lambda_i = \exp(\beta_i)$ $(1 \le i \le \nu)$ *and denote by* $\lambda_{[1]} \ge \lambda_{[2]} \ge \cdots \ge \lambda_{[\nu]}$ *the descending order of* $\lambda_1, \lambda_2, \ldots, \lambda_\nu$. *Let the locally* Φ_k*-optimal design* $\xi_{\boldsymbol{a}}^*$ *(at* $\boldsymbol{\beta}$*) from Theorem 3.5.1 be supported by the unit vectors* $\boldsymbol{e}_i$ $(1 \le i \le \nu)$ *with weights* $\omega_i^* = \frac{\lambda_i^{\frac{-k}{k+1}}}{\sum_{j=1}^{\nu} \lambda_j^{\frac{-k}{k+1}}}$ $(1 \le i \le \nu)$. *Then condition (3.22) is equivalent to*

$$\lambda_{[1]} + \lambda_{[2]} \le 1. \tag{3.23}$$

Proof. For intensity $u(\boldsymbol{x}, \boldsymbol{\beta}) = \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta}\right)$ and $a_i = 1$ $(1 \le i \le \nu)$ condition (3.22) reduces to

$$\exp(\sum_{i=1}^{\nu} \beta_i x_i) \sum_{i=1}^{\nu} \exp(-\beta_i) x_i^2 \le 1 \;\; \forall \boldsymbol{x} \in \mathcal{X}. \tag{3.24}$$

For any $\boldsymbol{x} = (x_1, \dots, x_\nu) \in \{0,1\}^\nu, \nu \geq 2$ define the index set $S \subseteq \{1, \dots, \nu\}$ such that $x_i = 1$ if $i \in S$ and $x_i = 0$ else. So for $\boldsymbol{x}$ described by $S \subseteq \{1, \dots, \nu\}$ and $s = \#S$, if $s = 0$ (i.e., $S = \emptyset$) then the l.h.s. of (3.24) is zero. If $s = 1$, inequality (3.24) becomes an equality. However, the l.h.s. of (3.24) is equal to $\exp(\sum_{i \in S} \beta_i) \sum_{i \in S} \exp(-\beta_i)$ which thus rewrites as $\prod_{i \in S} \lambda_i \sum_{i \in S} \lambda_i^{-1}$ or equivalently as $\sum_{i \in S} \prod_{j \in S \setminus \{i\}} \lambda_j$. By the the descending order $\lambda_{[1]} \geq \lambda_{[2]} \geq \dots \geq \lambda_{[\nu]}$ of $\lambda_1, \lambda_2, \dots, \lambda_\nu$ we obtain for all subsets $S \subseteq \{1, \dots, \nu\}$ of same sizes $s \geq 2$,

$$\sum_{i=1}^{s} \lambda_{[i]}^{-1} \prod_{i=1}^{s} \lambda_{[i]} = \sum_{i=1}^{s} \prod_{i \neq j=1}^{s} \lambda_{[j]} \geq \sum_{i \in S} \prod_{j \in S \setminus \{i\}} \lambda_j.$$

Denote $T_s = \sum_{i=1}^{s} \lambda_{[i]}^{-1} \prod_{i=1}^{s} \lambda_{[i]}$. Hence, inequality (3.24) is equivalent to $T_s \leq 1$ for all $s = 2, \dots, \nu$. Then it is sufficient to show that

$$\lambda_{[1]} + \lambda_{[2]} \leq 1 \iff T_s \leq 1 \;\; \forall s = 2, \dots, \nu.$$

For "$\Longleftarrow$", let $s = 2$ then $T_2 = \lambda_{[1]} + \lambda_{[2]} \leq 1$. For "$\Longrightarrow$", firstly, note that $T_2 = \lambda_{[1]} + \lambda_{[2]}$ thus $T_s \leq 1$ is true for $s = 2$. Now assume $T_s \leq 1$ is true for some $s = q < \nu$, i.e., $T_q \leq 1$ and we want to show that it is true for $s = q + 1$. We can write

$$\begin{aligned} T_{q+1} &= \Big(\sum_{i=1}^{q} \lambda_{[i]}^{-1} + \lambda_{[q+1]}^{-1}\Big)\Big(\prod_{i=1}^{q} \lambda_{[i]}\Big)\lambda_{[q+1]} \\ &= T_q \lambda_{[q+1]} + \prod_{i=1}^{q} \lambda_{[i]} = T_q \lambda_{[q+1]} + T_q \Big(\sum_{i=1}^{q} \lambda_{[i]}^{-1}\Big)^{-1} = T_q\Big(\lambda_{[q+1]} + \Big(\sum_{i=1}^{q} \lambda_{[i]}^{-1}\Big)^{-1}\Big) \\ &\text{since } \Big(\sum_{i=1}^{q} \lambda_{[i]}^{-1}\Big)^{-1} \leq \frac{1}{q}\lambda_{[1]} \text{ and } \lambda_{[q+1]} + \frac{1}{q}\lambda_{[1]} \leq T_2 = \lambda_{[1]} + \lambda_{[2]} \leq 1 \text{ we have} \end{aligned}$$

$$T_{q+1} \leq T_q\Big(\lambda_{[q+1]} + \frac{1}{q}\lambda_{[1]}\Big) \leq T_q T_2 \leq 1.$$

□

In analogy to Corollary 3.5.1 we introduce the next corollary for logistic models on the experimental region $\mathcal{X} = \{0,1\}^\nu, \; \nu \geq 2$.

Corollary 3.5.2. *Consider a non-intercept logistic model with $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ on the experimental region $\mathcal{X} = \{0,1\}^\nu$, $\nu \geq 2$ and intensity $u(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}^\top \boldsymbol{\beta})}{(1+\exp(\boldsymbol{x}^\top \boldsymbol{\beta}))^2}$. For a given parameter point $\boldsymbol{\beta} = (\beta_1, \dots, \beta_\nu)^\top$ define $\lambda_i = \exp(\beta_i)$ with $u_i = \lambda_i/(1+\lambda_i)^2$ $(1 \leq i \leq \nu)$. Let the locally Φ_k-optimal design $\xi_{\boldsymbol{a}}^*$ (at $\boldsymbol{\beta}$) from Theorem 3.5.1 be supported by the unit vectors $\mathbf{e}_i$ $(1 \leq i \leq \nu)$ with weights $\omega_i^* = \frac{u_i^{\frac{-k}{k+1}}}{\sum_{j=1}^{\nu} u_j^{\frac{-k}{k+1}}}$ $(1 \leq i \leq \nu)$. Then condition (3.22) is* *equivalent to*

$$\frac{\prod_{i\in S}\lambda_i}{(1+\prod_{i\in S}\lambda_i)^2}\Big(\sum_{i\in S}\lambda_i+\sum_{i\in S}\lambda_i^{-1}+2s\Big)\le 1\ \ \forall\emptyset\ne S\subseteq\{1,\ldots,\nu\}, s=\#S. \tag{3.25}$$

Proof. For $a_i = 1$ $(1 \le i \le \nu)$ condition (3.22) under a logistic model is equivalent to

$$\frac{\exp(\sum_{i=1}^{\nu}\beta_i x_i)}{\Big(1+\exp(\sum_{i=1}^{\nu}\beta_i x_i)\Big)^2}\sum_{i=1}^{\nu}\frac{(1+\lambda_i)^2}{\lambda_i}x_i^2\le 1\ \ \forall\boldsymbol{x}\in\mathcal{X}. \tag{3.26}$$

So for $\boldsymbol{x}$ described by $S \subseteq \{1, \ldots, \nu\}$ with $\lambda_i = \exp(\beta_i)$ $(1 \le i \le \nu)$, (3.26) rewrites as

$$\frac{\exp(\sum_{i\in S}\beta_i)}{(1+\exp(\sum_{i\in S}\beta_i))^2}\sum_{i\in S}\frac{(1+\lambda_i)^2}{\lambda_i}=\frac{\prod_{i\in S}\lambda_i}{(1+\prod_{i\in S}\lambda_i)^2}\sum_{i\in S}(\lambda_i+\lambda_i^{-1}+2)\le 1\ \ \forall\emptyset\ne S\subseteq\{1,\ldots,\nu\},$$

which is equivalent to (3.25). □

Remark 3.5.1. *One can slightly highlight on Φ_k-optimality under the multiple linear model without intercept $\boldsymbol{f}(\boldsymbol{x}) = (x_1, \ldots, x_\nu)^\top$ on the continuous experimental region $\mathcal{X} = [0, 1]^\nu$, $\nu \ge 2$. Here, $u(\boldsymbol{x}, \boldsymbol{\beta}) = 1$ for all $\boldsymbol{x} \in \mathcal{X}$ so the information matrices in a linear model are independent of $\boldsymbol{\beta}$. Note that Theorem 3.5.1 does not cover a non-intercept linear model on $\mathcal{X}$ since condition (3.22) cannot hold true for $\nu \ge 2$. However, the l.h.s. of condition (2.13) of The Equivalence Theorem (Theorem 2.2.2) under a linear model, i.e., $u(\boldsymbol{x}, \boldsymbol{\beta}) = 1$, is strictly convex and of course it attains its maximum at some vertices of $\mathcal{X}$. Thus the support of any Φ_k(or D, A)-optimal design is a subset of $\{0, 1\}^\nu$. As a result, in particular for D- and A-optimality, one might apply the results of Theorem 3.1 in Huda and Mukerjee (1988) which were obtained under a linear model on $\{0, 1\}^\nu$.*

- *For odd numbers of factors $\nu = 2q + 1$, $q \in \mathbb{N}$, the equally weighted designs ξ^* supported by all $\boldsymbol{x}^* = (x_1, \ldots, x_\nu) \in \{0, 1\}^\nu$ such that $\sum_{i=1}^{\nu} x_i = q + 1$ is D- and A-optimal.*
- *For even numbers of factors $\nu = 2q$, $q \in \mathbb{N}$, the equally weighted design ξ^* supported by all $\boldsymbol{x}^* = (x_1, \ldots, x_\nu) \in \{0, 1\}^\nu$ such that $\sum_{i=1}^{\nu} x_i = q$ or $\sum_{i=1}^{\nu} x_i = q + 1$ is D-optimal. Moreover, the design ξ^* which assigns equal weights to all points $\boldsymbol{x}^* = (x_1, \ldots, x_\nu) \in \{0, 1\}^\nu$ such that $\sum_{i=1}^{\nu} x_i = q$ is A-optimal.*

3.6 Relation of models with and without intercept

In this section we develop a particular approach to reduce the solution of locally optimal designs for generalized linear models. Our approach can be utilized under D- and A-criteria. We will show that under certain assumptions the locally D-optimal design for the model with intercept $\mathcal{M}$ can be obtained from the locally D-optimal design for the corresponding model without intercept $\widetilde{\mathcal{M}}$ by adding the origin to its support points. Conversely, the locally D-optimal design of the model without intercept $\widetilde{\mathcal{M}}$ can be obtained from the locally D-optimal design for the corresponding model with intercept $\mathcal{M}$ by removing the origin from its support points. Analogous result will be achieved for the A-criterion.

For that purpose we modify our notations and thus these models; $\widetilde{\mathcal{M}}$ and $\mathcal{M}$ are (without loss of generality) characterized in the following.

$$\widetilde{\mathcal{M}}\;:\;\;\tilde{\eta}=\boldsymbol{f}^\top(\boldsymbol{x})\tilde{\boldsymbol{\beta}},\quad \boldsymbol{x}\in\widetilde{\mathcal{X}}$$

where $\tilde{\boldsymbol{\beta}}=(\beta_1,\dots,\beta_\nu)^\top$ with intensity function $\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})$. Here we assume there is no constant (intercept) term explicitly involved in the present model, i.e., none of the regression components of the ν real-valued function $\boldsymbol{f}(\boldsymbol{x})$ is constant equal to 1. In the current situation, denote $\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})=\tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\boldsymbol{f}(\boldsymbol{x})=(f^{(1)}_{\tilde{\boldsymbol{\beta}}},\dots,f^{(\nu)}_{\tilde{\boldsymbol{\beta}}})^\top$. The information matrix of ξ on $\widetilde{\mathcal{X}}$ under model $\widetilde{\mathcal{M}}$ is thus written as

$$\tilde{\boldsymbol{M}}\big(\xi,\tilde{\boldsymbol{\beta}}\big)=\int_{\widetilde{\mathcal{X}}}\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})\boldsymbol{f}^\top_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})\,\xi(\mathrm{d}\boldsymbol{x}).$$

The corresponding model $\mathcal{M}$ is obtained by including the constant 1 and the intercept parameter β_0 into the linear predictor of the generalized linear model as in the following.

$$\mathcal{M}\;:\;\;\eta=\Big(1,\boldsymbol{f}^\top(\boldsymbol{x})\Big)\boldsymbol{\beta}=\beta_0+\boldsymbol{f}^\top(\boldsymbol{x})\tilde{\boldsymbol{\beta}},\quad \boldsymbol{x}\in\mathcal{X}$$

where $\boldsymbol{\beta}=(\beta_0,\tilde{\boldsymbol{\beta}}^\top)^\top$ with intensity function $u(\boldsymbol{x},\boldsymbol{\beta})$. Let $u_0=u(\boldsymbol{0},\boldsymbol{\beta})$.
Denote the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})=u^{\frac{1}{2}}\Big(\boldsymbol{x},\boldsymbol{\beta}\Big)\boldsymbol{f}(\boldsymbol{x})=(f^{(1)}_{\boldsymbol{\beta}},\dots,f^{(\nu)}_{\boldsymbol{\beta}})^\top$. So we can write $u^{\frac{1}{2}}\Big(\boldsymbol{x},\boldsymbol{\beta}\Big)\Big(1,\boldsymbol{f}^\top(\boldsymbol{x})\Big)^\top=\Big(u^{\frac{1}{2}}\Big(\boldsymbol{x},\boldsymbol{\beta}\Big),\boldsymbol{f}^\top_{\boldsymbol{\beta}}(\boldsymbol{x})\Big)^\top$.

Define Ξ_0 to be the set of all designs on $\mathcal{X}$ for model $\mathcal{M}$ such that $\boldsymbol{0}\in\mathrm{supp}(\xi)$ and there exists a constant vector $\boldsymbol{c}$ such that $\boldsymbol{c}^\top\boldsymbol{f}(\boldsymbol{x})=1$ for all $\boldsymbol{x}\in\mathrm{supp}(\xi)\setminus\{\boldsymbol{0}\}$, i.e.,

$$\Xi_0=\Big\{\xi:\xi\text{ on }\mathcal{X}\text{ with }\boldsymbol{0}\in\mathrm{supp}(\xi)\text{ and }\exists\,\boldsymbol{c}\in\mathbb{R}^\nu\ni\boldsymbol{c}^\top\boldsymbol{f}(\boldsymbol{x})=1\,\forall\boldsymbol{x}\in\mathrm{supp}(\xi)\setminus\{\boldsymbol{0}\}\Big\}.$$

Then the information matrix of $\xi\in\Xi_0$ under model $\mathcal{M}$ reads as

$$\boldsymbol{M}\big(\xi,\boldsymbol{\beta}\big)=\int_{\mathcal{X}}\Big(u^{\frac{1}{2}}\Big(\boldsymbol{x},\boldsymbol{\beta}\Big),\boldsymbol{f}^\top_{\boldsymbol{\beta}}(\boldsymbol{x})\Big)^\top\Big(u^{\frac{1}{2}}\Big(\boldsymbol{x},\boldsymbol{\beta}\Big),\boldsymbol{f}^\top_{\boldsymbol{\beta}}(\boldsymbol{x})\Big)\,\xi(\mathrm{d}\boldsymbol{x}).$$

Lemma 3.6.1. *Consider design $\xi^* \in \Xi_0$ for model $\mathcal{M}$ such that $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathbf{0}) = \mathbf{0}$. Then if the design ξ^* is locally D-optimal (at $\boldsymbol{\beta}$), the weight of the origin $\mathbf{0}$ is $\omega = (\nu+1)^{-1}$.*

Proof. Denote $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}) = (a_{i,j})_{i,j=1,\ldots,\nu+1}$. Let $\boldsymbol{A}_{11}$ be the submatrix of $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})$ formed by deleting the first row and the first column. Let $\boldsymbol{a} = (a_{1,2}, \ldots, a_{1,\nu+1})^\top$. Then the sensitivity function obtained from condition (2.11) of The Equivalence Theorem (Theorem 2.2.2) is given by

$$\begin{aligned}\psi(\boldsymbol{x}, \xi^*) &= u\big(\boldsymbol{x}, \boldsymbol{\beta}\big)\big(1, \boldsymbol{f}^\top(\boldsymbol{x})\big)\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\big(1, \boldsymbol{f}^\top(\boldsymbol{x})\big)^\top \\ &= u\big(\boldsymbol{x}, \boldsymbol{\beta}\big)\Big(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{A}_{11}\boldsymbol{f}(\boldsymbol{x}) + 2\boldsymbol{a}^\top \boldsymbol{f}(\boldsymbol{x}) + \big(\omega u_0\big)^{-1}\Big).\end{aligned}$$

According to Remark 2.2.6 ξ^* is locally D-optimal if $\psi(\mathbf{0}, \xi^*) = \nu + 1$. It implies that $u_0\big(\omega u_0\big)^{-1} = \nu + 1$ which holds true if $\omega = (\nu+1)^{-1}$. □

Lemma 3.6.2. *Consider design $\xi^* \in \Xi_0$ for model $\mathcal{M}$ such that $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathbf{0}) = \mathbf{0}$. Then if the design ξ^* is locally A-optimal (at $\boldsymbol{\beta}$), the weight of the origin $\mathbf{0}$ is*

$$\omega = \sqrt{\frac{\boldsymbol{c}^\top\boldsymbol{c} + 1}{u_0 \mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}))}}.$$

Proof. Denote $\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta}) = (a_{i,j})_{i,j=1,\ldots,\nu+1}$. Let $\boldsymbol{A}_{11}$ be the submatrix of $\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta})$ formed by deleting the first row and the first column. Let $\boldsymbol{a} = (a_{1,2}, \ldots, a_{1,\nu+1})^\top$. Then the sensitivity function obtained from condition (2.12) of The Equivalence Theorem (Theorem 2.2.2) is given by

$$\begin{aligned}\psi(\boldsymbol{x}, \xi^*) &= u\big(\boldsymbol{x}, \boldsymbol{\beta}\big)\big(1, \boldsymbol{f}^\top(\boldsymbol{x})\big)\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta})\big(1, \boldsymbol{f}^\top(\boldsymbol{x})\big)^\top \\ &= u\big(\boldsymbol{x}, \boldsymbol{\beta}\big)\Big(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{A}_{11}\boldsymbol{f}(\boldsymbol{x}) + 2\boldsymbol{a}^\top \boldsymbol{f}(\boldsymbol{x}) + (\boldsymbol{c}^\top\boldsymbol{c} + 1)\big(\omega u_0\big)^{-2}\Big).\end{aligned}$$

According to Remark 2.2.6 ξ^* is locally A-optimal if $\psi(\mathbf{0}, \xi^*) = \mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}))$. It implies that $u_0(\boldsymbol{c}^\top\boldsymbol{c} + 1)\big(\omega u_0\big)^{-2} = \mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}))$ which holds true if
$\omega = \sqrt{(\boldsymbol{c}^\top\boldsymbol{c} + 1)/(u_0 \mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}))}$. □

In the following we give sufficient conditions under which the locally D- resp. A-optimal design at a parameter point $\tilde{\boldsymbol{\beta}}$ for model $\widetilde{\mathcal{M}}$ can be obtained from the locally D- resp. A-optimal design from Ξ_0 at a parameter point $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ for the corresponding model $\mathcal{M}$ by simply removing the origin point from its support points and renormalizing the weights of the remaining support points or vice versa. To this end, for a design $\xi \in \Xi_0$ define $\xi_{-\mathbf{0}}$ on $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$ to be the conditional measure of ξ given $\boldsymbol{x} \neq \mathbf{0}$. It is noted that $\mathrm{supp}(\xi) = \mathrm{supp}(\xi_{-\mathbf{0}}) \cup \{\mathbf{0}\}$. Let $\xi_{\mathbf{0}}$ denotes the one point design supported by the origin point $\mathbf{0}$, then let us set

$$\xi = \omega\,\xi_{\mathbf{0}} + (1-\omega)\,\xi_{-\mathbf{0}}.$$

Assume that for a given parameter point $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ we have $u(\boldsymbol{x}, \boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})$ and thus $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})$ and let $\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\mathbf{0}) = \mathbf{0}$. It follows that

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \begin{pmatrix} m_{1,1}(\xi, \tilde{\boldsymbol{\beta}}) & (1-\omega)\,\tilde{\boldsymbol{m}}^\top(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) \\ (1-\omega)\,\tilde{\boldsymbol{m}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) & (1-\omega)\,\tilde{\boldsymbol{M}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) \end{pmatrix}$$

where

$$m_{1,1}(\xi, \tilde{\boldsymbol{\beta}}) = \int_{\mathcal{X}} \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})\,\xi(\mathrm{d}\boldsymbol{x}),\;\; \tilde{\boldsymbol{m}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \int_{\widetilde{\mathcal{X}}} \tilde{u}^{\frac{1}{2}}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})\,\xi_{-\mathbf{0}}(\mathrm{d}\boldsymbol{x}),$$
$$\text{and}\;\; \tilde{\boldsymbol{M}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \int_{\widetilde{\mathcal{X}}} \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}^\top(\boldsymbol{x})\,\xi_{-\mathbf{0}}(\mathrm{d}\boldsymbol{x}),$$

where the submatrix $\tilde{\boldsymbol{M}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})$ is the information matrix of $\xi_{-\mathbf{0}}$ for model $\widetilde{\mathcal{M}}$. Note that $m_{1,1}(\xi, \tilde{\boldsymbol{\beta}}) = \omega u_0 + (1-\omega)\widetilde{m}^\circ(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})$ where $\widetilde{m}^\circ(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \int_{\widetilde{\mathcal{X}}} \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})\,\xi_{-\mathbf{0}}(\mathrm{d}\boldsymbol{x})$. Since there exists a constant vector $\boldsymbol{c}$ such that $\boldsymbol{c}^\top \boldsymbol{f}(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \mathrm{supp}(\xi^*) \setminus \{\mathbf{0}\}$, it is straightforward to verify the following

$$\boldsymbol{c}^\top \tilde{\boldsymbol{m}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \widetilde{m}^\circ(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) \;\text{ and }\; \tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{m}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \boldsymbol{c} \;\text{ thus}$$
$$\tilde{\boldsymbol{m}}^\top(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{m}}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) = \widetilde{m}^\circ(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}).$$

If $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ is nonsingular, we can get

$$\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}) = \begin{pmatrix} \frac{1}{\omega\,u_0} & -\frac{\boldsymbol{c}^\top}{\omega\,u_0} \\ -\frac{\boldsymbol{c}}{\omega\,u_0} & \frac{1}{1-\omega}\,\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}}) + \frac{\boldsymbol{c}\boldsymbol{c}^\top}{\omega\,u_0} \end{pmatrix}. \tag{3.27}$$

Theorem 3.6.1. *Consider design $\xi^* \in \Xi_0$ for model $\mathcal{M}$. Let the design $\xi^*_{-\mathbf{0}}$ on $\widetilde{\mathcal{X}}$ be the conditional measure of ξ^* given $\boldsymbol{x} \neq \mathbf{0}$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ be given such that $u(\boldsymbol{x}, \boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$. Assume that $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$ and $\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\mathbf{0}) = \mathbf{0}$. Let $\xi^* = (1/(\nu+1))\,\xi_{\mathbf{0}} + (\nu/(\nu+1))\,\xi^*_{-\mathbf{0}}$. Then*
(1) If ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$ then $\xi^*_{-\mathbf{0}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$.
(2) If $\xi^*_{-\mathbf{0}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$ and

$$\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}^\top(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) \leq \nu\Big(1 - \frac{(\boldsymbol{c}^\top \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) - \tilde{u}^{\frac{1}{2}}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}}))^2}{u_0}\Big)\;\forall \boldsymbol{x} \in \mathcal{X} \tag{3.28}$$

then ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$.

Proof. Ad (1) Let $\xi^* = (1/(\nu+1))\,\xi_{\mathbf{0}} + (\nu/(\nu+1))\,\xi^*_{-\mathbf{0}} \in \Xi_0$ be locally D-optimal (at $\boldsymbol{\beta}$) on $\mathcal{X}$ for model $\mathcal{M}$. We want to prove that $\xi^*_{-\mathbf{0}}$ on $\widetilde{\mathcal{X}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$. By condition (2.11) of The Equivalence Theorem (Theorem 2.2.2) we guarantee at $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ that

$$u(\boldsymbol{x},\boldsymbol{\beta})\big(1,\boldsymbol{f}^\top(\boldsymbol{x})\big)\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\big(1,\boldsymbol{f}^\top(\boldsymbol{x})\big)^\top \leq \nu+1 \ \forall \boldsymbol{x} \in \mathcal{X}, \tag{3.29}$$

where, at $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$, $u(\boldsymbol{x},\boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})$ and $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$ with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$. Note also $\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})$ is given by (3.27) with $\omega = 1/(\nu+1)$. Then inequality (3.29) is equivalent to

$$\boldsymbol{f}^\top_{\tilde{\beta}}(\boldsymbol{x})\Big(\frac{\nu+1}{\nu}\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\boldsymbol{\beta}) + \frac{(\nu+1)\boldsymbol{c}\boldsymbol{c}^\top}{u_0}\Big)\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) - \frac{2(\nu+1)\boldsymbol{c}^\top\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) + (\nu+1)\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})}{u_0} \leq \nu+1 \quad \forall \boldsymbol{x} \in \widetilde{\mathcal{X}}.$$

Elementary computations show that the above inequality is equivalent to

$$\boldsymbol{f}^\top_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\boldsymbol{\beta})\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) + \frac{\nu\,(\boldsymbol{c}^\top\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) - \tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}}))^2}{u_0} \leq \nu \quad \forall \boldsymbol{x} \in \widetilde{\mathcal{X}}. \tag{3.30}$$

Since $\dfrac{\nu\,(\boldsymbol{c}^\top\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) - \tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}}))^2}{u_0} \geq 0$, (3.30) implies that

$$\boldsymbol{f}^\top_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\boldsymbol{\beta})\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) \leq \nu \quad \forall \boldsymbol{x} \in \widetilde{\mathcal{X}}.$$

and so $\xi^*_{-\mathbf{0}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) by The Equivalence Theorem (Theorem 2.2.2, condition (2.11)) .

Ad (2) Let $\xi^*_{-\mathbf{0}}$ on $\widetilde{\mathcal{X}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$. Under the assumptions stated in the theorem we want to show that ξ^* from Ξ_0 on $\mathcal{X}$ is locally D-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$. To this end, we investigate condition (2.11) of The Equivalence Theorem (Theorem 2.2.2) which is given above by (3.29) and is also equivalent to (3.30) at $\boldsymbol{\beta}$. Hence, (3.30) holds true by condition (3.28). Of course, because $\xi^*_{-\mathbf{0}}$ is locally D-optimal inequality (3.28) becomes an equality at each design point of $\xi^*_{-\mathbf{0}}$ which surely is a design point of ξ^* and since $\omega = 1/(\nu+1)$ the equality also holds at point $\mathbf{0}$. □

Next we introduce analogous result for the A-optimality. From (3.27), $\boldsymbol{M}^{-2}(\xi, \tilde{\boldsymbol{\beta}})$ is equal to

$$\begin{pmatrix} \frac{\boldsymbol{c}^\top \boldsymbol{c}+1}{\omega^2 u_0^2} & -\frac{(\boldsymbol{c}^\top \boldsymbol{c}+1)\boldsymbol{c}^\top}{\omega^2 u_0^2} - \frac{\boldsymbol{c}^\top \tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})}{(1-\omega)\omega u_0} \\ -\frac{\boldsymbol{c}(\boldsymbol{c}^\top \boldsymbol{c}+1)}{\omega^2 u_0^2} - \frac{\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}}{(1-\omega)\omega u_0} & \frac{(\boldsymbol{c}^\top \boldsymbol{c}+1)\boldsymbol{c}\boldsymbol{c}^\top}{\omega^2 u_0^2} + \frac{\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^\top + \boldsymbol{c}\boldsymbol{c}^\top \tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})}{(1-\omega)\omega u_0} + \frac{\tilde{\boldsymbol{M}}^{-2}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})}{(1-\omega)^2} \end{pmatrix}. \tag{3.31}$$

Lemma 3.6.3. *Let $\xi \in \Xi_0$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^\top)^\top$ be given such that $u(\boldsymbol{x}, \boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$. Assume that ξ is locally A-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$. Then the optimal weight ω of the origin $\mathbf{0}$ is given by*

$$\omega = \frac{\sqrt{\boldsymbol{c}^\top \boldsymbol{c} + 1}}{\sqrt{\boldsymbol{c}^\top \boldsymbol{c} + 1} + \sqrt{u_0 \operatorname{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\Big)}}. \tag{3.32}$$

Moreover, we have

$$\operatorname{tr}\Big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\Big) = \frac{1}{u_0}\left(\sqrt{\boldsymbol{c}^\top \boldsymbol{c} + 1} + \sqrt{u_0 \operatorname{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\Big)}\right)^2. \tag{3.33}$$

Proof. As $\operatorname{tr}\Big(\boldsymbol{c}\boldsymbol{c}^\top\Big) = \boldsymbol{c}^\top \boldsymbol{c}$ we obtain from (3.27)

$$\operatorname{tr}\Big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\Big) = \frac{\boldsymbol{c}^\top \boldsymbol{c} + 1}{\omega u_0} + \frac{1}{1-\omega} \operatorname{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\Big). \tag{3.34}$$

From Lemma 3.6.2 the ω is given by

$$\omega = \sqrt{\frac{\boldsymbol{c}^\top \boldsymbol{c} + 1}{u_0 \operatorname{tr}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}))}}.$$

We can write

$$\frac{1}{\omega} = \sqrt{\frac{u_0 \operatorname{tr}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}))}{\boldsymbol{c}^\top \boldsymbol{c} + 1}}.$$

Substituting (3.34) in the r.h.s. of the above equation leads to

$$\frac{1}{\omega} = \sqrt{\frac{1}{\omega} + \frac{u_0 \operatorname{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\Big)}{\boldsymbol{c}^\top \boldsymbol{c} + 1}\Big(\frac{1}{1-\omega}\Big)}$$

thus

$$\Big(\frac{1}{\omega}\Big)^2 = \frac{1}{\omega} + \frac{u_0 \operatorname{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}}, \tilde{\boldsymbol{\beta}})\Big)}{\boldsymbol{c}^\top \boldsymbol{c} + 1}\Big(\frac{1}{1-\omega}\Big)$$

so we get

$$\Big(\frac{1}{\omega^2}-\frac{1}{\omega}\Big)(1-\omega)=\frac{u_0\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)}{\boldsymbol{c}^\top\boldsymbol{c}+1}$$

The l.h.s. of the above equation is equal to the square $(\frac{1}{\omega}-1)^2$. Straightforward computations imply that

$$\frac{1}{\omega}=\frac{\sqrt{u_0\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)}+\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}}{\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}}$$

and it follows that the ω from (3.32) can be given by the inverse of the r.h.s. of above equation. Now it remains to proof (3.33). To this end, substitute (3.32) in (3.34) and hence we get

$$\begin{aligned}\mathrm{tr}\Big(\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta})\Big)&=\frac{\boldsymbol{c}^\top\boldsymbol{c}+1}{u_0}+2\sqrt{\frac{(\boldsymbol{c}^\top\boldsymbol{c}+1)\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)}{u_0}}+\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)\\&=\Bigg(\sqrt{\frac{\boldsymbol{c}^\top\boldsymbol{c}+1}{u_0}}+\sqrt{\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)}\Bigg)^2.\end{aligned}$$

Then (3.33) follows. □

Theorem 3.6.2. *Consider the assumptions and notations of Theorem 3.6.1. Denote* $\tilde{\tau}=\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)$ *and let*

$$\xi^*=\left(\frac{\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}}{\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}+\sqrt{u_0\tilde{\tau}}}\right)\xi_{\mathbf{0}}+\left(\frac{\sqrt{u_0\,\tilde{\tau}}}{\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}+\sqrt{u_0\,\tilde{\tau}}}\right)\xi^*_{-\mathbf{0}}.$$

Denote the following equations

$$\begin{aligned}T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}})=&\frac{(\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}+\sqrt{u_0\tilde{\tau}})^2(\boldsymbol{c}^\top\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})-\tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}}))^2}{u_0^2}\\&+\frac{(\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}+\sqrt{u_0\tilde{\tau}})^2}{u_0\sqrt{\tilde{\tau}u_0(\boldsymbol{c}^\top\boldsymbol{c}+1)}}\Bigg(\tilde{\boldsymbol{f}}^\top_{\tilde{\beta}}(\boldsymbol{x})\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^\top+\boldsymbol{c}\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\\&-4\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\Bigg),\\T_2(\boldsymbol{x},\tilde{\boldsymbol{\beta}})=&\sqrt{\frac{\tilde{\tau}}{u_0(\boldsymbol{c}^\top\boldsymbol{c}+1)}}\Bigg(\tilde{\boldsymbol{f}}^\top_{\tilde{\beta}}(\boldsymbol{x})\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^\top+\boldsymbol{c}\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\\&-2\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\Bigg).\end{aligned}$$

Then
(1) If ξ^* *is locally A-optimal (at* $\boldsymbol{\beta}$*) for model* $\mathcal{M}$ *and* $T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\geq 0$ *for all* $\boldsymbol{x}\in\widetilde{\mathcal{X}}$ *then*

$\xi^*_{-\mathbf{0}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$.
(2) If $\xi^*_{-\mathbf{0}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$ and

$$\tilde{\boldsymbol{f}}^{\top}_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-2}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x}) \leq \tilde{\tau}\left(1 - \frac{(\boldsymbol{c}^{\top}\boldsymbol{f}^{\top}_{\tilde{\beta}}(\boldsymbol{x}) - \tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}}))^2}{u_0}\right) + T_2(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\ \forall \boldsymbol{x} \in \mathcal{X} \tag{3.35}$$

then ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$.

Proof. Ad (1) Let $\xi^* = (\frac{\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1}}{\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1}+\sqrt{u_0\,\tilde{\tau}}})\,\xi_{\mathbf{0}} + (\frac{\sqrt{u_0\,\tilde{\tau}}}{\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1}+\sqrt{u_0\,\tilde{\tau}}})\,\xi^*_{-\mathbf{0}} \in \Xi_0$ on $\mathcal{X}$ be locally A-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$. We want to prove that $\xi^*_{-\mathbf{0}}$ on $\widetilde{\mathcal{X}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$. Considering (3.33) then condition (2.12) of The Equivalence Theorem (Theorem 2.2.2) guarantees at $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^{\top})^{\top}$ that for all $\boldsymbol{x} \in \mathcal{X}$

$$u(\boldsymbol{x},\boldsymbol{\beta})\big(1,\boldsymbol{f}^{\top}(\boldsymbol{x})\big)\boldsymbol{M}^{-2}(\xi^*,\boldsymbol{\beta})\big(1,\boldsymbol{f}^{\top}(\boldsymbol{x})\big)^{\top} \leq \frac{1}{u_0}\left(\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1} + \sqrt{\mathrm{tr}\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\Big)}\right)^2, \tag{3.36}$$

where, at $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^{\top})^{\top}$, $u(\boldsymbol{x},\boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})$ and $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$ with $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$, and $\boldsymbol{M}^{-2}(\xi^*,\boldsymbol{\beta})$ is given by (3.31) with $\omega = (\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1})/(\sqrt{\boldsymbol{c}^{\top}\boldsymbol{c}+1}+\sqrt{u_0\,\tilde{\tau}})$. Then the l.h.s. of inequality (3.36) equals

$$\tilde{\boldsymbol{f}}^{\top}_{\tilde{\beta}}(\boldsymbol{x})\left(\frac{(\boldsymbol{c}^{\top}\boldsymbol{c}+1)\boldsymbol{c}\boldsymbol{c}^{\top}}{\omega^2u_0^2} + \frac{\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^{\top} + \boldsymbol{c}\boldsymbol{c}^{\top}\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})}{\omega(1-\omega)u_0} + \frac{1}{(1-\omega)^2}\tilde{\boldsymbol{M}}^{-2}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\right)\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})$$
$$-\,2\left(\frac{\boldsymbol{c}^{\top}(\boldsymbol{c}^{\top}\boldsymbol{c}+1)}{\omega^2\,u_0^2} + \frac{\boldsymbol{c}^{\top}\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})}{\omega(1-\omega)u_0}\right)\tilde{u}^{\frac{1}{2}}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x}) + \frac{(\boldsymbol{c}^{\top}\boldsymbol{c}+1)\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})}{\omega^2\,u_0^2},$$

and it is straightforward to see that (3.36) is equivalent to

$$\boldsymbol{f}^{\top}_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-2}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) + T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) \leq \tilde{\tau}\quad \forall \boldsymbol{x} \in \tilde{\mathcal{X}}. \tag{3.37}$$

By the assumption $T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) \geq 0$ for all $\boldsymbol{x} \in \tilde{\mathcal{X}}$, (3.37) implies that
$\boldsymbol{f}^{\top}_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-2}(\xi^*_{-\mathbf{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) \leq \tilde{\tau}\quad \forall \boldsymbol{x} \in \tilde{\mathcal{X}}$.

and so $\xi^*_{-\mathbf{0}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) by The Equivalence Theorem (Theorem 2.2.2, condition (2.12)) .

Ad (2) Let $\xi^*_{-\mathbf{0}}$ on $\tilde{\mathcal{X}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for model $\widetilde{\mathcal{M}}$. Under the assumptions stated in the theorem we want to show that ξ^* from Ξ_0 on $\mathcal{X}$ is locally A-optimal (at $\boldsymbol{\beta}$) for model $\mathcal{M}$. To this end, we investigate condition (2.12) of The Equivalence Theorem (Theorem 2.2.2) which is given above by (3.36) and is also equivalent to (3.37) at $\boldsymbol{\beta}$ for all $\boldsymbol{x} \in \mathcal{X}$. Hence, it is straightforward to see that (3.37) for all $\boldsymbol{x} \in \mathcal{X}$ holds true by condition (3.35). Of course, because $\xi^*_{-\mathbf{0}}$ is locally A-optimal and $T_2(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) = 0$ for all $\boldsymbol{x} \in \mathrm{supp}(\xi^*_{-\mathbf{0}})$ inequality (3.35) becomes an equality at each design point of $\xi^*_{-\mathbf{0}}$ which surely is a design point of ξ^*. Since

$\omega = (\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1})/(\sqrt{\boldsymbol{c}^\top\boldsymbol{c}+1}+\sqrt{u_0\,\tilde{\tau}})$ and $T_2(\mathbf{0},\tilde{\boldsymbol{\beta}})=0$ the equality also holds at the origin point $\mathbf{0}$. □

Remark 3.6.1. *The results of this section might be viewed as a generalization of the results of both Li, Lau, and Zhang (2005) and Zhang and Wong (2013) that were derived under linear models, i.e., when the intensities are constants equal to 1.*

Remark 3.6.2. *Note that the assumption $\boldsymbol{c}^\top\boldsymbol{f}(\boldsymbol{x})=1$ for all $\boldsymbol{x}\in\mathrm{supp}(\xi^*)\setminus\{\mathbf{0}\}$ is equivalent to that $\boldsymbol{f}(\boldsymbol{x})$ for all $\boldsymbol{x}\in\mathrm{supp}(\xi^*_{-\mathbf{0}})$ lies on a hyperplane not containing the origin. Thus every saturated design for generalized linear models without intercept satisfies that assumption. Moreover, the assumption $\boldsymbol{c}^\top\boldsymbol{f}(\boldsymbol{x})=1$ for all $\boldsymbol{x}\in\widetilde{\mathcal{X}}$ is satisfied when $\widetilde{\mathcal{X}}$ is given by the $(\nu-1)$-dimensional unit simplex, i.e., $\widetilde{\mathcal{X}}=\{\boldsymbol{x}=(x_1,\dots,x_\nu)^\top, 0\le x_i\le 1\ \forall i, \sum_{i=1}^{\nu}x_i=1\}$. In such a case the mixture constraint of $\widetilde{\mathcal{X}}$ which is given by $\sum\limits_{i=1}^{\nu}x_i=1$ entails that $\boldsymbol{c}=(1,\dots,1)^\top$.*

Example 3.6.1. Here, we consider a first order Poisson model where $\boldsymbol{f}(\boldsymbol{x})=(1,\boldsymbol{x}^\top)^\top$. The intensity functions under $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ are given by

$$u(\boldsymbol{x},\boldsymbol{\beta})=\exp(\beta_0+\boldsymbol{x}^\top\tilde{\boldsymbol{\beta}})\quad\text{and}\quad\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})=\exp(\boldsymbol{x}^\top\tilde{\boldsymbol{\beta}}),$$

respectively. It is noted that $u(\boldsymbol{x},\boldsymbol{\beta})$ factorizes; i.e., $u(\boldsymbol{x},\boldsymbol{\beta})=\exp(\beta_0)\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})$. Therefore, $\boldsymbol{M}(\xi,\boldsymbol{\beta})=\exp(\beta_0)\boldsymbol{M}(\xi,\tilde{\boldsymbol{\beta}})$ for any given parameter point $\boldsymbol{\beta}=(\beta_0,\tilde{\boldsymbol{\beta}}^\top)^\top$. That means the design ξ is independent of β_0 and hence, locally optimal designs for a Poisson model with intercept is governed by $\tilde{u}(\boldsymbol{x},\tilde{\boldsymbol{\beta}})$. Similar situation holds under the Rasch Poisson-Gamma counts model (Graßhoff, Holling, and Schwabe (2013)) in item response theory and the Rasch Poisson counts model (Graßhoff, Holling, and Schwabe (2018)).

In the current thesis, a Poisson model with two binary factors is addressed as a generalized linear model in Section 3.3. Part (*i*) in Theorem 3.3.1 and Part (*iii*) in Theorem 3.3.2 introduce D- and A-optimal saturated designs ξ^*, respectively, which belong to Ξ_0. Hence, by part (1) in Theorem 3.6.1 or Theorem 3.6.2 the design $\xi^*_{-\mathbf{0}}$ is locally D- or A-optimal, respectively, and is equivalent to the corresponding design given by Corollary 3.5.1 for $\nu=2$ (see Section 3.5 for Φ_k-optimality for GLMs without intercept).

A relevant work from the literature includes Russell et al. (2009) who derived a locally D-optimal saturated design ξ^* for a first order Poisson model with intercept on $\mathcal{X}=[0,1]^\nu$ where $\nu\ge 2$ at $\beta_i=-2$ $(1\le i\le\nu)$. The support is given by $\boldsymbol{x}_0^*=(0,0,\dots,0)^\top$ and the ν-dimensional unit vectors $\boldsymbol{x}_i^*=\boldsymbol{e}_i$ $(1\le i\le\nu)$ with equal weights $(\nu+1)^{-1}$. So under the assumptions of Theorem 3.6.1, part (1), the design $\xi^*_{-\mathbf{0}}$ on $\mathcal{X}$ is locally D-optimal at $\beta_i=-2$ $(1\le i\le\nu)$ for the corresponding model without intercept.

Furthermore, Schmidt (2019) constructed a class of locally D- and A-optimal designs for a general setup of generalized linear models with intercept where the assumptions of Theorem 3.6.1 and Theorem 3.6.2 can be satisfied in some of his results. On that basis it is possible to determine the locally optimal designs for the corresponding models without intercept. □

Example 3.6.2. Consider a first order logistic model with $\boldsymbol{f}(\boldsymbol{x}) = (1, \boldsymbol{x}^\top)^\top$. The intensity functions under $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ are given by

$$u(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \boldsymbol{x}^\top \tilde{\boldsymbol{\beta}})}{(1 + \exp(\beta_0 + \boldsymbol{x}^\top \tilde{\boldsymbol{\beta}}))^2} \quad \text{and} \quad \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}}) = \frac{\exp(\boldsymbol{x}^\top \tilde{\boldsymbol{\beta}})}{(1 + \exp(\boldsymbol{x}^\top \tilde{\boldsymbol{\beta}}))^2},$$

respectively. Note that $u(\boldsymbol{x}, \boldsymbol{\beta}) = \tilde{u}(\boldsymbol{x}, \tilde{\boldsymbol{\beta}})$ at $\boldsymbol{\beta} = (0, \tilde{\boldsymbol{\beta}}^\top)^\top$.

A relevant work from the literature includes Kabera, Haines, and Ndlovu (2015) in which Theorem 3.2 in that work provided a three-point locally D-optimal saturated design ξ^* at $(0, \tilde{\boldsymbol{\beta}}^\top)^\top$, $\tilde{\boldsymbol{\beta}} \in (0, \infty)^2$ for the two-factor logistics model on the experimental region $\mathcal{X} = [0, \infty)^2$. The support is given by $(0, 0)^\top, (0, u^*)^\top, (u^*, 0)^\top$ where $u^* > 0$ is the unique solution for u to the equation $2 + u + 2e^u - ue^u = 0$. Hence, the assumptions of Theorem 3.6.1, part (1), are satisfied so the design $\xi^*_{-\mathbf{0}}$ on $\mathcal{X}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) with equal weights 1/2 for the corresponding model without intercept.

See also Example 1 and Example 3 in Schmidt (2019) where product type designs are locally D-optimal at $\boldsymbol{\beta} = (0, \tilde{\boldsymbol{\beta}}^\top)^\top$ for Poisson and logistic models with intercept, respectively, which are relevant to our results in this section. □

Chapter 4

Applications to gamma models

In the present chapter the gamma model with continuous (quantitative) factors is considered. There are wide applications where the gamma model with its canonical link can be fitted. Nevertheless, there is always a doubt about the suitable link function for outcomes. The common alternative links may come from the power link family that includes the canonical link therefore it is a favorite choice for employment in the thesis.

In section 4.1, we introduce the gamma model highlighting on the related assumptions. Additionally, the notions of locally complete classes and locally essentially complete classes are presented. In section 4.2, locally complete classes and locally essentially complete classes of designs are found leading to a considerable reduction of the problems of locally optimal designs for gamma models. From those classes locally D- and A-optimal designs are derived. Besides, as a gamma model is recognized as a particular generalized linear model the results that are obtained in Chapter 3 for a general setup of the generalized linear model will be applied in relevant cases here. The optimality conditions will be intuitively characterized by the model parameters and hence, those conditions cover relevant subregions of the parameter space. So, our results on locally D- or A-optimality are applicable for the majority of possible parameter points.

In Section 4.3, we consider a model with a single continuous factor. In section 4.4, we deal with a model without interactions whereas a model with interactions is employed in Section 4.5. In both sections, we distinguish between models with and without intercept. Finally, in Section 4.6 the performance of some derived locally D-optimal designs compared with particular non-optimal designs are examined.

The numerical computations are conducted by computer algebra with the aid of the software packages **R** (R Core Team (2018)) and Wolfram Mathematica 11.3 (Wolfram Research, Inc. (2018)).

Some of the results in this chapter are provided in Gaffke, Idais, and Schwabe (2019) and Idais and Schwabe (2019).

4.1 Model specification

Let $Y_1, ..., Y_n$ be independent gamma-distributed response variables for n experimental units, where the density for each Y_i is written as

$$p(y_i; \kappa, \lambda_i) = \frac{\lambda_i^{\kappa}}{\Gamma(\kappa)} y_i^{\kappa-1} e^{-\lambda_i y_i} \quad , \kappa, \lambda_i, y_i > 0, (1 \le i \le n), \tag{4.1}$$

where the shape parameter κ of the gamma distribution is the same for all Y_i but the expectations $\mu_i = E(Y_i)$ depend on the values $\boldsymbol{x}_i$ of a covariate $\boldsymbol{x}$. The canonical link for a gamma distribution (4.1) is reciprocal (inverse),

$$\eta_i = \kappa/\mu_i, \quad \text{where} \quad \eta_i = \boldsymbol{f}^\top(\boldsymbol{x}_i)\boldsymbol{\beta} \;\; \text{is the linear predictor,} \;\; (1 \le i \le n).$$

Here $\boldsymbol{f} = (f_1, \dots, f_p)^\top$ is a given $\mathbb{R}^p$-valued vector of regression functions on the experimental region $\mathcal{X} \subset \mathbb{R}^\nu$, $\nu \ge 1$ with linearly independent component functions $f_1, \dots, f_p$, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is a parameter vector (see McCullagh and Nelder (1989), Section 2.2.4). In this case the mean-variance function is $V(\mu) = \mu^2$ and the variance of a gamma distribution is thus given by $\mathrm{var}(Y) = \kappa^{-1}\mu^2$. Therefore, the intensity function (2.5) at a point $\boldsymbol{x} \in \mathcal{X}$ reads as

$$u(\boldsymbol{x}, \boldsymbol{\beta}) = \kappa \left(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}\right)^{-2}.$$

The power link family which is considered throughout presents the class of link functions as in Burridge and Sebastiani (1994), see also Atkinson and Woods (2015), Section 2.5,

$$\eta_i = \mu_i^{\rho}, \quad \text{where} \quad \eta_i = \boldsymbol{f}^\top(\boldsymbol{x}_i)\boldsymbol{\beta}, \;\; (1 \le i \le n). \tag{4.2}$$

The exponent ρ of the power link function is a given nonzero real number. The intensity function under this family is defined as

$$u_0(\boldsymbol{x}, \boldsymbol{\beta}) = \kappa\rho^{-2}\left(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}\right)^{-2} \;\; \text{for all} \;\; \boldsymbol{x} \in \mathcal{X}. \tag{4.3}$$

Gamma-distributed responses are continuous and nonnegative and thus for a given experimental region $\mathcal{X}$ we assume throughout that the parameter vector $\boldsymbol{\beta}$ satisfies

$$\boldsymbol{f}^\top(\boldsymbol{x})\,\boldsymbol{\beta} > 0 \;\; \text{for all} \;\; \boldsymbol{x} \in \mathcal{X}. \tag{4.4}$$

The Fisher information matrix for a single observation at a point $\boldsymbol{x} \in \mathcal{X}$ under a parameter vector $\boldsymbol{\beta}$ is given by $u_0(\boldsymbol{x}, \boldsymbol{\beta})\,\boldsymbol{f}(\boldsymbol{x})\,\boldsymbol{f}^\top(\boldsymbol{x})$. Note that the positive factor $\kappa\rho^{-2}$ is the same for all $\boldsymbol{x}$ and $\boldsymbol{\beta}$ and will not affect any design consideration below. We will ignore that factor and consider a normalized version of the information matrix at $\boldsymbol{x}$ and $\boldsymbol{\beta}$,

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) = \left(\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}\right)^{-2} \boldsymbol{f}(\boldsymbol{x})\,\boldsymbol{f}^\top(\boldsymbol{x}). \tag{4.5}$$

Modifying notions of Ehrenfeld (1956) we introduce the notions of a locally complete class of designs and of a locally essentially complete class of designs. They are based on the Loewner semi-ordering, " $\leq$ ", of information matrices or, more generally, of nonnegative definite $p \times p$ matrices. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonnegative definite $p \times p$ matrices we write $\boldsymbol{A} \leq \boldsymbol{B}$ if and only if $\boldsymbol{B} - \boldsymbol{A}$ is nonnegative definite.

Definition 4.1.1. *Let $\boldsymbol{\beta}$ be a given parameter point. Denote by Ξ the set of all designs and let $\widetilde{\Xi} \subseteq \Xi$.*

(i) The subset $\widetilde{\Xi}$ is called a locally essentially complete class (at $\boldsymbol{\beta}$) if for each design $\xi \in \Xi \setminus \widetilde{\Xi}$ there exists a design $\tilde{\xi} \in \widetilde{\Xi}$ such that $\boldsymbol{M}(\xi,\boldsymbol{\beta}) \leq \boldsymbol{M}(\tilde{\xi},\boldsymbol{\beta})$.

(ii) The subset $\widetilde{\Xi}$ is called a locally complete class (at $\boldsymbol{\beta}$) if for each design $\xi \in \Xi \setminus \widetilde{\Xi}$ there exists a design $\tilde{\xi} \in \widetilde{\Xi}$ such that $\boldsymbol{M}(\xi,\boldsymbol{\beta}) \leq \boldsymbol{M}(\tilde{\xi},\boldsymbol{\beta})$ and $\boldsymbol{M}(\xi,\boldsymbol{\beta}) \neq \boldsymbol{M}(\tilde{\xi},\boldsymbol{\beta})$.

In particular, the D-criterion as well as the A-criterion are strictly decreasing on the set of all positive definite $p \times p$ matrices w.r.t. the Loewner semi-ordering, i.e., the functions $\Phi_{\mathrm{D}}(\boldsymbol{A}) = -\log\det\big(\boldsymbol{A}\big)$ and $\Phi_{\mathrm{A}}(\boldsymbol{A}) = \mathrm{tr}\big(\boldsymbol{A}^{-1}\big)$ defined on the set of all positive definite $p \times p$ matrices $\boldsymbol{A}$ satisfy the following.

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive definite $p \times p$ and $\boldsymbol{A} \leq \boldsymbol{B}$, $\boldsymbol{A} \neq \boldsymbol{B}$ then $\Phi(\boldsymbol{A}) > \Phi(\boldsymbol{B})$,

for $\Phi = \Phi_{\mathrm{D}}$ and $\Phi = \Phi_{\mathrm{A}}$. So, if $\widetilde{\Xi}$ is a locally essentially complete class (at $\boldsymbol{\beta}$) then there exists a design $\tilde{\xi}^* \in \widetilde{\Xi}$ which is locally D-optimal (at $\boldsymbol{\beta}$). If $\widetilde{\Xi}$ is a locally complete class (at $\boldsymbol{\beta}$) then any locally D-optimal design ξ^* (at $\boldsymbol{\beta}$) must be a member of $\widetilde{\Xi}$. In other words, if a locally essentially complete or a locally complete class (at $\boldsymbol{\beta}$) is given then the search of a locally D-optimal design (at $\boldsymbol{\beta}$) may be restricted to that class of designs. In case of a locally complete class (at $\boldsymbol{\beta}$) it is guaranteed that there are no other locally D-optimal designs (at $\boldsymbol{\beta}$) outside that class whereas in the weaker case of a locally essentially complete class there may be other locally D-optimal designs outside that class. Analogous statements are true for the A-criterion.

4.2 Complete class results

We consider the case of a ν-dimensional covariate $\boldsymbol{x} = (x_1,\ldots,x_\nu)^\top$. So the experimental region $\mathcal{X}$ is a subset of $\mathbb{R}^\nu$. Below, $\mathcal{X}$ will be a polytope, i.e.,

$$\mathcal{X} = \mathrm{Conv}\big\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\big\} \quad \text{with} \quad K \in \mathbb{N},\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_K \in \mathbb{R}^\nu, \tag{4.6}$$

where 'Conv' denotes convex hull operation. That is, $\mathrm{Conv}\big\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\big\}$ consists of all linear combinations $\sum_{k=1}^{K}\alpha_k\boldsymbol{v}_k$ with coefficients $\alpha_k \geq 0$ $(1 \leq k \leq K)$ such that $\sum_{k=1}^{K}\alpha_k = 1$. Usually, the generating vectors $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K$ will constitute the set of vertices of the polytope $\mathcal{X}$. A particular case frequently occurring in applications is that of a ν-dimensional hyperrectangle, i.e., each component x_i may range over a given compact interval $[a_i, b_i]$, $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $i = 1,\ldots,\nu$, so

$$\mathcal{X} = \Big\{\boldsymbol{x} = (x_1,\ldots,x_\nu)^\top \in \mathbb{R}^\nu \,:\, a_i \leq x_i \leq b_i \;\forall\, i = 1,\ldots,\nu\Big\}. \tag{4.7}$$

Clearly, $\mathcal{X}$ from (4.7) is a special case of (4.6) with $K = 2^\nu$ vertices given by the points whose ith coordinates are either a_i or b_i for all $i = 1,\ldots,\nu$. Even more special is the case of a hypercube $\mathcal{X} = \big[a,b\big]^\nu$ with $a, b \in \mathbb{R}$ and $a < b$, which emerges from (4.7) when $a_i = a$ and $b_i = b$ for all $i = 1,\ldots,\nu$.

Let $\boldsymbol{f} = (f_1,\ldots,f_p)^\top$ be an $\mathbb{R}^p$-valued function that is defining a linear predictor $\eta(\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta}$, where the set of feasible parameter points $\boldsymbol{\beta}$ is given by (4.4), i.e., $\boldsymbol{f}^\top(\boldsymbol{x})\,\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$. We call $\boldsymbol{f}$ affine-linear if each component function f_j is affine-linear, i.e., f_j has the form

$$f_j(\boldsymbol{x}) \;=\; c_{0j} + \boldsymbol{c}_j^\top\boldsymbol{x} \;\;\forall\, \boldsymbol{x} \in \mathcal{X},$$

with constants $c_{0j} \in \mathbb{R}$ and $\boldsymbol{c}_j \in \mathbb{R}^\nu$, $1 \leq j \leq p$. A weaker condition on $\boldsymbol{f}$ is the following. We call $\boldsymbol{f}$ affine-multilinear if each component function f_j $(1 \leq j \leq p)$ satisfies

$$f_j \in \mathrm{span}\Big\{g_S \,:\, S \subseteq \{1,\ldots,\nu\}\Big\}, \tag{4.8}$$

$$\text{where}\;\; g_S(\boldsymbol{x}) = \prod_{i\in S} x_i \;\;\forall\, \boldsymbol{x} = (x_1,\ldots,x_\nu)^\top \in \mathcal{X}, \tag{4.9}$$

and, by convention, for $S = \emptyset$ the empty product $\prod_{i\in\emptyset} x_i$ is equal to 1. We mean by $\mathrm{span}\Big\{g_S \,:\, S \subseteq \{1,\ldots,\nu\}\Big\}$ in (4.8) the linear space consisting of all linear combinations of the functions g_S $(S \subseteq \{1,\ldots,\nu\})$. Clearly, this space contains in particular all the affine-linear real-valued functions. Hence the condition of affine-multilinearity of $\boldsymbol{f}$ is weaker than that of affine-linearity. A popular example of an affine-linear function $\boldsymbol{f}$ is given by

$$\boldsymbol{f}(\boldsymbol{x}) = \Big(1, x_1,\ldots,x_\nu\Big)^\top \;\;\forall\, \boldsymbol{x} = (x_1,\ldots,x_\nu)^\top \in \mathcal{X}. \tag{4.10}$$

When pairwise interaction terms $x_h x_i$ $(1 \leq h < i \leq \nu)$ are included, e.g. if $\nu = 2$,

$$\boldsymbol{f}(\boldsymbol{x}) = \Big(1, x_1, x_2, x_1x_2\Big)^\top \;\;\forall\, \boldsymbol{x} = (x_1,x_2)^\top \in \mathcal{X}, \tag{4.11}$$

one has an example of an affine-multilinear function $\boldsymbol{f}$ which is not affine-linear (unless

$\mathcal{X}$ is a suitably degenerated set). Further examples are obtained in case $\nu \geq 3$ where also interaction terms of third order, $x_h x_i x_j$ $(1 \leq h < i < j \leq \nu)$, or even higher order may be included.

In what follows, for a given parameter point $\boldsymbol{\beta}$ (satisfying (4.4)), we denote

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \left(\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta}\right)^{-1}\boldsymbol{f}(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathcal{X}. \tag{4.12}$$

Clearly, the information matrices given by (4.5) can be written in form (2.1), i.e., $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{x})$. So there is no loss of generality to restrict our attention to $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$. Next we derive an auxiliary result of geometric type.

Lemma 4.2.1. *Assume one of the following conditions (i) or (ii).*

(i) The experimental region $\mathcal{X}$ is a polytope (4.6) and $\boldsymbol{f}$ is affine-linear.

(ii) $\mathcal{X}$ is a ν-dimensional hyperrectangle (4.7) and $\boldsymbol{f}$ is affine-multilinear.

Then, for each parameter point $\boldsymbol{\beta}$ according to (4.4),

$$\left\{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \,:\, \boldsymbol{x} \in \mathcal{X}\right\} \subseteq \mathrm{Conv}\left\{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_1)\,,\,\ldots\,,\,\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_K)\right\},$$

where in case (i) the $\boldsymbol{v}_k$ $(1 \leq k \leq K)$ are from (4.6), whereas in case (ii) the $\boldsymbol{v}_k$ $(1 \leq k \leq K = 2^{\nu})$ are the vertices of the hyperrectangle (4.7).

Proof. We will use similar arguments as in Gaffke, Graßhoff, and Schwabe (2014) when proving their Lemma 4.2. Let $\boldsymbol{\beta}$ be given. We have to show that

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \in \mathrm{Conv}\left\{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_1)\,,\,\ldots\,,\,\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_K)\right\} \quad \text{for all } \boldsymbol{x} \in \mathcal{X}. \tag{4.13}$$

A result from convex analysis (see Rockafellar (1970), Corollary 13.1.1) tells us that (4.13) holds true if and only if

$$\boldsymbol{a}^{\top}\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \leq \max_{1\leq k\leq K} \boldsymbol{a}^{\top}\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_k) \quad \text{for all } \boldsymbol{a} \in \mathbb{R}^p \text{ and all } \boldsymbol{x} \in \mathcal{X}. \tag{4.14}$$

Case 1: Condition (i) is satisfied.

To show that (4.14) holds true consider, for any given $\boldsymbol{a} \in \mathbb{R}^p$, the real valued function $h_{\boldsymbol{a}}$ on $\mathcal{X}$ defined by

$$h_{\boldsymbol{a}}(\boldsymbol{x}) = \boldsymbol{a}^{\top}\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \frac{\boldsymbol{a}^{\top}\boldsymbol{f}(\boldsymbol{x})}{\boldsymbol{\beta}^{\top}\boldsymbol{f}(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{X}.$$

Then $h_{\boldsymbol{a}}$ is quasi-convex, i.e., for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ and all $\alpha \in [0, 1]$ one has

$$h_{\boldsymbol{a}}\left(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\right) \leq \max\left\{h_{\boldsymbol{a}}(\boldsymbol{x}_1)\,,\,h_{\boldsymbol{a}}(\boldsymbol{x}_2)\right\}. \tag{4.15}$$

This can be seen as follows. By the affine-linearity of $\boldsymbol{f}$ we have

$$\boldsymbol{f}\Big(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\Big) = \alpha\boldsymbol{f}(\boldsymbol{x}_1) + (1-\alpha)\boldsymbol{f}(\boldsymbol{x}_2),$$

and thus the l.h.s. of (4.15) rewrites as

$$\frac{\alpha\,\boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_1)\; + \;(1-\alpha)\,\boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_2)}{\alpha\,\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_1)\; + \;(1-\alpha)\,\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_2)} = \frac{c_0 + c_1\alpha}{d_0 + d_1\alpha},$$

where

$$\begin{aligned} c_0 &= \boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_2), \quad c_1 = \boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_2), \\ d_0 &= \boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_2), \quad d_1 = \boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_2). \end{aligned}$$

Note that $d_0 + d_1\alpha > 0$ for all $\alpha \in [0, 1]$ due to (4.4). By the monotonicity of the ratio $(c_0 + c_1\alpha)/(d_0 + d_1\alpha)$ as a function of α on $[0, 1]$, i.e., nondecreasing or nonincreasing, we have

$$\frac{c_0 + c_1\alpha}{d_0 + d_1\alpha} \le \max\left\{\frac{c_0}{d_0}, \frac{c_0 + c_1}{d_0 + d_1}\right\} \;\; \forall\, \alpha \in [0, 1].$$

From $c_0/d_0 = \boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_2)\Big/\Big(\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_2)\Big) = h_{\boldsymbol{a}}(\boldsymbol{x}_2)$ and $(c_0 + c_1)/(d_0 + d_1) = \boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x}_1)\Big/\Big(\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x}_1)\Big) = h_{\boldsymbol{a}}(\boldsymbol{x}_1)$ inequality (4.15) follows. By induction one obtains from (4.15) that for all $r \ge 2$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \in \mathcal{X}$, and $\alpha_1, \ldots, \alpha_r \ge 0$ with $\sum_{k=1}^r \alpha_k = 1$ one has

$$h_{\boldsymbol{a}}\left(\sum_{k=1}^r \alpha_k\boldsymbol{x}_k\right) \le \max_{1\le k\le r} h_{\boldsymbol{a}}(\boldsymbol{x}_k). \tag{4.16}$$

Now, by (4.6), every $\boldsymbol{x} \in \mathcal{X}$ can be written as $\boldsymbol{x} = \sum_{k=1}^K \alpha_k\boldsymbol{v}_k$ for some $\alpha_k \ge 0$ $(1 \le k \le K)$ with $\sum_{k=1}^K \alpha_k = 1$, and applying (4.16) to $r = K$, $\boldsymbol{x}_k = \boldsymbol{v}_k$ $(1 \le k \le K)$ we obtain

$$h_{\boldsymbol{a}}(\boldsymbol{x}) \le \max_{1\le k\le K} h_{\boldsymbol{a}}(\boldsymbol{v}_k)$$

for all $\boldsymbol{a} \in \mathbb{R}^p$ and all $\boldsymbol{x} \in \mathcal{X}$ which is (4.14).

Case 2: Condition (ii) is satisfied.

Let $\boldsymbol{a} \in \mathbb{R}^p$ be given. Consider

$$\boldsymbol{a}^\top\boldsymbol{f}_\beta(\boldsymbol{x}) = \frac{\boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x})}{\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{X}.$$

By the affine-multilinearity of $\boldsymbol{f}$ both $\boldsymbol{a}^\top\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{\beta}^\top\boldsymbol{f}(\boldsymbol{x})$ are affine-multilinear functions of $\boldsymbol{x} \in \mathcal{X}$. So for any fixed $i \in \{1, \ldots, \nu\}$ and fixed components $x_j \in [a_j, b_j]$ for all $j \neq i$, the function

$$x_i \mapsto \frac{\boldsymbol{a}^\top \boldsymbol{f}(x_1, \ldots, x_\nu)}{\boldsymbol{\beta}^\top \boldsymbol{f}(x_1, \ldots, x_\nu)}, \quad a_i \leq x_i \leq b_i,$$

has the form

$$\frac{c_0 + c_1 x_i}{d_0 + d_1 x_i}, \quad a_i \leq x_i \leq b_i, \tag{4.17}$$

with constants c_0, c_1, d_0, d_1 (depending on the x_j, $j \neq i$) such that $d_0 + d_1 x_i > 0$ for all $x_i \in [a_i, b_i]$. By the monotonicity of the ratio (4.17) on $[a_i, b_i]$, it attains its maximum at one of the end points a_i or b_i of the interval. So, in particular, we have shown that for every $\boldsymbol{x} \in \mathcal{X}$ and $i \in \{1, \ldots, \nu\}$ there is an $\widetilde{\boldsymbol{x}} = (\widetilde{x}_1, \ldots, \widetilde{x}_\nu) \in \mathcal{X}$ such that $\widetilde{x}_i = a_i$ or $\widetilde{x}_i = b_i$, and $\boldsymbol{a}^\top \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \leq \boldsymbol{a}^\top \boldsymbol{f}_{\boldsymbol{\beta}}(\widetilde{\boldsymbol{x}})$. Using this iteratively for $i = 1, \ldots, \nu$, we conclude that for every $\boldsymbol{x} \in \mathcal{X}$ there is a vertex $\boldsymbol{v}_{k_0}$ of the ν-dimensional hyperrectangle $\mathcal{X}$ for some $k_0 \in \{1, \ldots, 2^\nu\}$ such that $\boldsymbol{a}^\top \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \leq \boldsymbol{a}^\top \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_{k_0})$. Hence (4.14) follows. □

Remark 4.2.1. *For the case that $\boldsymbol{f}$ is given by (4.10) and $\mathcal{X}$ is the unit hypercube $[0,1]^\nu$ it was shown in Burridge and Sebastiani (1994) that equality of both sets in Lemma 4.2.1 holds.*

Theorem 4.2.1. *Under the assumptions of Lemma 4.2.1 consider the subset $\widetilde{\Xi}$ of all designs $\widetilde{\xi}$ with $\operatorname{supp}(\widetilde{\xi}) \subseteq \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$. Let a parameter point $\boldsymbol{\beta}$ with (4.4) be given. Then*

- *$\widetilde{\Xi}$ is a locally essentially complete class (at $\boldsymbol{\beta}$).*
- *$\widetilde{\Xi}$ is a locally complete class (at $\boldsymbol{\beta}$) if the function $\boldsymbol{f}_{\boldsymbol{\beta}}$ from (4.12) is injective.*

Proof. Let $\xi \in \Xi \setminus \widetilde{\Xi}$ be given,

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_r \\ \omega_1 & \ldots & \omega_r \end{pmatrix}.$$

Take any support point $\boldsymbol{x} = \boldsymbol{x}_{i_0}$ of ξ which does not belong to $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ and let $\omega = \omega_{i_0}$. The information matrix of ξ at $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \omega \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) + \overline{\omega} \boldsymbol{M}(\xi_{-\boldsymbol{x}}, \boldsymbol{\beta}), \quad \text{where } \overline{\omega} = 1 - \omega \tag{4.18}$$

and the design $\xi_{-\boldsymbol{x}}$ is obtained from ξ by removing $\boldsymbol{x}$ from the support of ξ and renormalizing the weights of the remaining support points, i.e.,

$$\xi_{-\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_{i_0-1} & \boldsymbol{x}_{i_0+1} & \ldots & \boldsymbol{x}_r \\ \omega_1/\overline{\omega} & \ldots & \omega_{i_0-1}/\overline{\omega} & \omega_{i_0+1}/\overline{\omega} & \ldots & \omega_r/\overline{\omega} \end{pmatrix}.$$

We will show that there is a design $\widetilde{\xi}_{\boldsymbol{x}} \in \widetilde{\Xi}$ such that

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) \leq \boldsymbol{M}(\widetilde{\xi}_{\boldsymbol{x}},\boldsymbol{\beta}); \tag{4.19}$$

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) \neq \boldsymbol{M}(\widetilde{\xi}_{\boldsymbol{x}},\boldsymbol{\beta}) \;\text{ if } \boldsymbol{f}_\beta \text{ is injective.} \tag{4.20}$$

Then the design $\xi'_{\boldsymbol{x}} = \omega\widetilde{\xi}_{\boldsymbol{x}} + \overline{\omega}\xi_{-\boldsymbol{x}}$ satisfies

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) \leq \boldsymbol{M}(\xi'_{\boldsymbol{x}},\boldsymbol{\beta}); \tag{4.21}$$

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) \neq \boldsymbol{M}(\xi'_{\boldsymbol{x}},\boldsymbol{\beta}) \;\text{ if } \boldsymbol{f}_\beta \text{ is injective.} \tag{4.22}$$

Note that the number of support points which do not belong to $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\}$ was diminished by 1 when going from ξ to $\xi'_{\boldsymbol{x}}$. If $\xi'_{\boldsymbol{x}} \in \widetilde{\Xi}$ then $\widetilde{\xi} = \xi'_{\boldsymbol{x}}$ has the desired properties. Otherwise, i.e., $\xi'_{\boldsymbol{x}} \notin \widetilde{\Xi}$ the above arguments can be applied to $\xi'_{\boldsymbol{x}}$ instead of ξ, and so on. After a finite number of repetitions the process must stop with a design $\widetilde{\xi} \in \widetilde{\Xi}$ and, since at each step an analogue to (4.21), (4.22) holds, the design $\widetilde{\xi}$ satisfies

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) \leq \boldsymbol{M}(\widetilde{\xi},\boldsymbol{\beta});$$
$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) \neq \boldsymbol{M}(\widetilde{\xi},\boldsymbol{\beta}) \;\text{ if } \boldsymbol{f}_\beta \text{ is injective.}$$

It remains to show that for any given $\boldsymbol{x} \notin \{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\}$ there exists a design $\widetilde{\xi}_{\boldsymbol{x}} \in \widetilde{\Xi}$ such that (4.19), (4.20) hold. To this end we write $\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\,\boldsymbol{f}_{\boldsymbol{\beta}}^\top(\boldsymbol{x})$ and by using Lemma 3.1,

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) \;=\; \sum_{k=1}^{K} \alpha_k \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_k) \tag{4.23}$$

for some $\alpha_k \geq 0$ $(1 \leq k \leq K)$ with $\sum_{k=1}^K \alpha_k = 1$. Employing a lemma in Pukelsheim (1993), Section 8.4, see also Theorem 4.2 in Gaffke and Krafft (1982), one gets from (4.23)

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) \;\leq\; \sum_{k=1}^{K} \alpha_k \boldsymbol{M}(\boldsymbol{v}_k,\boldsymbol{\beta}). \tag{4.24}$$

The Loewner inequality in (4.24) is an equality if and only if all vectors $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_k)$ with $\alpha_k > 0$ $(1 \leq k \leq K)$ coincide which by (4.23) implies that $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{v}_{k_0})$ for some $k_0 \in \{1,\ldots,K\}$. If $\boldsymbol{f}_{\boldsymbol{\beta}}$ is injective the latter is not possible since $\boldsymbol{x} \notin \{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\}$. Hence it follows that

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) \;\neq\; \sum_{k=1}^{K} \alpha_k \boldsymbol{M}(\boldsymbol{v}_k,\boldsymbol{\beta}) \;\text{ if } \boldsymbol{f}_{\boldsymbol{\beta}} \text{ is injective.} \tag{4.25}$$

Now consider the design $\widetilde{\xi}_{\boldsymbol{x}}$ with support points $\boldsymbol{v}_k$ and weights α_k for all $k \in \{1,\ldots,K\}$ with $\alpha_k > 0$. Clearly, $\widetilde{\xi}_{\boldsymbol{x}} \in \widetilde{\Xi}$, and observing that $\boldsymbol{M}(\widetilde{\xi}_{\boldsymbol{x}},\boldsymbol{\beta})$ is equal to the r.h.s. of (4.24) as well as of (4.25), statements (4.19), (4.20) follow. □

4.3 Single-factor model

In this section we consider the simplest case ($\nu = 1$) determined by a single factor model

$$\boldsymbol{f}(x) = (1, x)^\top,\ x \in \mathcal{X} = [a, b],\ a < b. \tag{4.26}$$

By assumption (4.4), i.e., $\beta_0 + \beta_1 x > 0$ for all $x \in \mathcal{X} = [a, b]$, the parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ is such that $\beta_0 + \beta_1 a > 0$ and $\beta_0 + \beta_1 b > 0$. Note that the function (4.12), i.e., $\boldsymbol{f}_{\boldsymbol{\beta}}(x) = (\beta_0 + \beta_1 x)^{-1}(1, x)^\top$ is injective, and hence we utilize the results of complete class of designs from the previous section, so we only restrict to designs supported by $\{a, b\}$. Consequently, the result of Theorem 3.2.1 can be transferred.

Corollary 4.3.1. *Consider model (4.26) and experimental region $\mathcal{X} = [a, b]$ with real numbers a, b, $a < b$. Let a parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ be given. Then:*

(i) The unique locally D-optimal design (at $\boldsymbol{\beta}$) is the two-point design supported by a and b with equal weights $1/2$.

(ii) The unique locally A-optimal design (at $\boldsymbol{\beta}$) is the two-point design supported by a and b with weights

$$\omega_a^* = \frac{(\beta_0 + \beta_1 a)\sqrt{1 + b^2}}{(\beta_0 + \beta_1 a)\sqrt{1 + b^2} + (\beta_0 + \beta_1 b)\sqrt{1 + a^2}},\quad \omega_b^* = 1 - \omega_a^*.$$

Proof. The locally complete class $\widetilde{\Xi}$ from Theorem 4.2.1 consists all one- or two-point designs with support points a or b. Hence any D-optimal design and any A-optimal design must have support equal to $\{a, b\}$. Thus the rest of the proof follows from Theorem 3.2.1 where $u_a = (\beta_0 + \beta_1 a)^{-2}$ and $u_b = (\beta_0 + \beta_1 b)^{-2}$. □

Remark 4.3.1. *The D-optimal design ξ_{D}^* from the above corollary can be approved by condition (2.11) of The Equivalence Theorem (Theorem 2.2.2). Let $\delta(x)$ be a function in x evaluated from the difference of the l.h.s. (sensitivity function) and the r.h.s. of condition (2.11), hence condition (2.11) is equivalent to $\delta(x) \le 0$ for all $x \in [a, b]$. We have*

$$\begin{aligned}\delta(x) =& \Big(2\big(\beta_0 + \beta_1 a\big)\big(\beta_0 + \beta_1 b\big)\Big)x^2\\ &- 2\Big(a\big(\beta_0 + \beta_1 b\big)^2 + b\big(\beta_0 + \beta_1 a\big)^2 + \beta_0\beta_1(a - b)^2\Big)x\\ &+ \Big(a^2\big(\beta_0 + \beta_1 b\big)^2 + b^2\big(\beta_0 + \beta_1 a\big)^2 - \beta_0^2(a - b)^2\Big).\end{aligned}$$

The function $\delta(x)$ is a polynomial of degree 2. Since $\beta_0 + \beta_1 x > 0$ for all $x \in [a, b]$ the leading coefficient $2\big(\beta_0 + \beta_1 a\big)\big(\beta_0 + \beta_1 b\big)$ is positive. Hence, by the strict convexity of $\delta(x)$ we get $\delta(x) < 0$ for all $x \in (a, b)$ where $\delta(a) = \delta(b) = 0$.

The A-optimal design ξ_A^* from the above corollary depends on the model parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ and on the values of a and b. So the A-optimal design on a specific experimental region varies with the parameter value. In order to examine the effect of $\boldsymbol{\beta}$ on ξ_A^* let us consider, for simplicity, the experimental region $\mathcal{X} = [0,1]$ and thus, the parameter space of $\boldsymbol{\beta} = (\beta_0, \beta)^\top$ obtained from assumption (4.4) is determined by $\beta_1 > -\beta_0$, $\beta_0 > 0$. Define the ratio $\gamma = \beta_1/\beta_0$ whose range is $(-1, \infty)$. The optimal weight as a function of γ rewrites as $\omega^* = \frac{\sqrt{2}}{1+\sqrt{2}+\gamma}$. The curve of ω^* is depicted in Figure 4.1. Clearly, ω^* is monotonic decreasing. The design ξ_A^* is equally weighted ($\omega^* = 1/2$) at $\gamma = \sqrt{2} - 1$. If $\gamma = 1$ then $\omega^* = \frac{\sqrt{2}}{2+\sqrt{2}} = \sqrt{2} - 1$. If $\gamma = 0$ we get $\omega^* = \frac{\sqrt{2}}{1+\sqrt{2}}$ which is identical to the case under simple linear models. Of course the designs is not A-optimal at the limits where in such a case the design is supported only by 0 if $\gamma \to -1$ since $\omega^* = 1$ or by 1 if $\gamma \to \infty$ since $\omega^* = 0$.

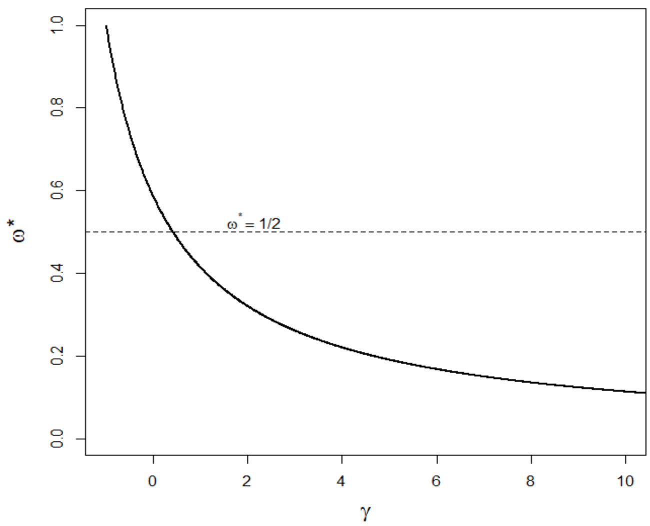


FIGURE 4.1: Effect of $\gamma = \beta_1/\beta_0, \beta_0 > 0, \gamma > -1$ on the optimal weight ω^* of the locally A-optimal design given in Corollary 4.3.1 where $\mathcal{X} = [0, 1]$,

Remark 4.3.2. *The model in Corollary 4.3.1 was also addressed in Aminnejad and Jafari (2017), Section 2 of that reference. The authors report some numerical results on locally D- or A-optimal designs (Subsection 2.1 of that paper). However, they claim the experimental region to be defined by* $-\beta_0/\beta_1 < x \leq 1$ *where* $\beta_0 > 0$ *and* $\beta_1 > 0$ *are given. This half open interval does not make sense as an experimental region since no locally D- or A-optimal design (at* $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$*) exists. E.g., for* $\beta_0 = \beta_1 = 3$ *the authors report a design* ξ_D *supported by* -0.980 *and* 1.000 *with equal weights* 0.5 *as a locally D-optimal design, and a design* ξ_A *supported by* -0.998 *and* 1.000 *with weights* 0.001 *and* 0.999 *as a locally A-optimal design, resp. In fact, by our Corollary*

4.3.1, ξ_{D} is the locally D-optimal design on the experimental region $[-0.980\,,\,1]$ and ξ_{A} is the locally A-optimal design on the experimental region $[-0.998\,,\,1]$. Consider any experimental region $[a,1]$ with $-\beta_1/\beta_0 < a < 1$ and the locally D- and A-optimal designs ξ_{D}^* and ξ_{A}^*, resp., from Corollary 4.3.1. Letting a decrease to $-\beta_1/\beta_0$ entails that ξ_{D}^* converges to the two point design on $-\beta_0/\beta_1$ and 1 with equal weights $1/2$ and ξ_{A}^* converges to the one-point design on 1 (as it is indicated in Figure 4.1), whereas the information matrices $\boldsymbol{M}(\xi_{\mathrm{D}}^*,\boldsymbol{\beta})$ and $\boldsymbol{M}(\xi_{\mathrm{A}}^*,\boldsymbol{\beta})$ diverge. The information matrices of both limiting designs are undefined.

4.4 Model without interaction

In this section we deal with first order gamma models with arbitrary number of quantitative factors without interaction terms. We begin with models with intercept by Subsection 4.4.1 whereas models without intercept are utilized in Subsection 4.4.2.

Here, $\boldsymbol{f}$ is affine-linear as defined in Section 4.2 and the experimental region $\mathcal{X}$ is a ν-dimensional hyperrectangle from (4.7). So by Theorem 4.2.1 we will look for the optimal designs in the set of vertices of $\mathcal{X}$.

4.4.1 Model with intercept

Consider the model

$$\boldsymbol{f}(\boldsymbol{x}) = \big(1\,,\,\boldsymbol{x}^{\top}\big)^{\top}, \quad \boldsymbol{x} \in \mathcal{X}; \tag{4.27}$$

that is, the linear predictor $\eta(\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta}$ is assumed to be an affine-linear function of the ν-dimensional covariate with coefficient vector $\boldsymbol{\beta}$. The experimental region is assumed to be the ν-dimensional unit hypercube $\mathcal{X} = [0,1]^{\nu}$, $\nu \geq 2$. For D-optimality, as pointed out in Burridge and Sebastiani (1994), this is no loss of generality since the case of an arbitrary ν-dimensional hyperrectangle can be transformed to that standard case. However, for A-optimality it is a special case to which we restrict in order to reduce the technical effort.

For a given parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_\nu)^{\top}$ with (4.4), i.e., $(1, \boldsymbol{x}^{\top})\,\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} = (x_1, \ldots, x_\nu)^{\top} \in \mathcal{X}$, the function $\boldsymbol{f}_{\boldsymbol{\beta}}$ is injective in view of

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \big((1, \boldsymbol{x}^{\top})\,\boldsymbol{\beta}\big)^{-1}\big(1\,,\,\boldsymbol{x}^{\top}\big)^{\top} \text{ for all } \boldsymbol{x} \in \mathcal{X}.$$

So by Theorem 4.2.1 the set $\widetilde{\Xi}$ of those designs which are supported only by the vertices of $\mathcal{X}$ is a locally complete class of designs (at $\boldsymbol{\beta}$), and hence any locally D- or A-optimal design (at $\boldsymbol{\beta}$) must be a member of that class.

Let us now focus on the model with $\nu = 2$ factors. In view of Theorem 3.3.1 we next provide necessary and sufficient conditions for the D- and A-optimality.

Corollary 4.4.1. *Consider model (4.27) with $\nu = 2$ and experimental region $\mathcal{X} = [0,1]^2$. Denote the vertices by $\boldsymbol{v}_1 = (0,0)^\top$, $\boldsymbol{v}_2 = (1,0)^\top$, $\boldsymbol{v}_3 = (0,1)^\top$, and $\boldsymbol{v}_4 = (1,1)^\top$. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ be a parameter point, i.e., $\beta_0 > 0$, $\beta_0 + \beta_i > 0$ ($i = 1, 2$), and $\beta_0 + \beta_1 + \beta_2 > 0$. Then the unique locally D-optimal design ξ^* is as follows.*

(i) If $\beta_0^2 - \beta_1\beta_2 \le 0$ then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

(ii) If $(\beta_0 + \beta_1)^2 + \beta_1\beta_2 \le 0$ then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4$.

(iii) If $(\beta_0 + \beta_2)^2 + \beta_1\beta_2 \le 0$ then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4$.

(iv) If $\beta_0^2 + \beta_1^2 + \beta_2^2 + \beta_1\beta_2 + 2\beta_0(\beta_1 + \beta_2) \le 0$ then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$.

(v) If none of the cases (i) – (iv) applies then ξ^* is supported by the four vertices

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \omega_1^* & \omega_2^* & \omega_3^* & \omega_4^* \end{pmatrix}.$$

where the optimal weights in case (v) are uniquely determined by the condition

$$\omega_k^* > 0 \ (1 \le k \le 4), \ \sum_{k=1}^{4} \omega_k^* = 1, \ \text{and} \ \ u_k\omega_k^*\left(\tfrac{1}{3} - \omega_k^*\right) \ (1 \le k \le 4) \ \text{are equal.} \tag{4.28}$$

Proof. By Theorem 4.2.1 the support of a locally D-optimal design must be a subset of the set of vertices $\widetilde{\mathcal{X}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. So we may restrict to designs on the reduced experimental region $\widetilde{\mathcal{X}}$. Denote $u_k = u(\boldsymbol{v}_k, \boldsymbol{\beta})$ $(1 \le k \le 4)$, i.e.,

$$u_1 = 1/\beta_0^2, \ u_2 = 1/(\beta_0 + \beta_1)^2, \ u_3 = 1/(\beta_0 + \beta_2)^2, \ u_4 = 1/(\beta_0 + \beta_1 + \beta_2)^2.$$

Then the proof follows that of Theorem 3.3.1. Hence, straightforward computations can show that the conditions of parts (i)–(iv) in the corollary are equivalent to the following conditions, respectively, derived from part (i) of Theorem 3.3.1.

$$\begin{aligned} u_4^{-1} &\ge u_1^{-1} + u_2^{-1} + u_3^{-1}, \\ u_3^{-1} &\ge u_1^{-1} + u_2^{-1} + u_4^{-1}, \\ u_2^{-1} &\ge u_1^{-1} + u_3^{-1} + u_4^{-1}, \\ u_1^{-1} &\ge u_2^{-1} + u_3^{-1} + u_4^{-1}. \end{aligned}$$

For case (v) of the corollary, however, it is essentially obtained from the condition in part (ii) of Theorem 3.3.1 which argues that there is no $\boldsymbol{\beta}$ value satisfies the conditions mentioned above of saturated designs. Thus from Remark 2.2.4 we guarantee that at any parameter point that does not satisfy the conditions of the saturated designs there exists a four-point design which is locally D-optimal at that parameter point. □

Each condition provided in parts (i)–(iv) of Corollary 4.4.1 characterizes a subregion of the parameter space ($\beta_0 > 0,\ \beta_1 > -\beta_0,\ \beta_2 > -\beta_0,\ \beta_1 + \beta_2 > -\beta_0$) where the corresponding saturated design is D-optimal. As examples of parameter points for which Corollary 4.4.1 applies and hence the locally D-optimal design (at $\boldsymbol{\beta}$) is a three-point design are the following.

(i) $\beta_1 = \beta_2 = -\beta < 0$ where $\beta_0/3 \leq \beta \leq \beta_0$; three-point design on $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$.

(ii) $\beta_1 = \beta > 0,\ \beta_2 = -\beta$ where $\beta_0/2 \leq \beta < \beta_0$; three-point design on $\boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4$.

Remark 4.4.1. *The subregion where a four-point design given in Corollary 4.4.1 is D-optimal has been determined by computer algebra and is given below.*

- $-\beta_0 < \beta_1 < 0$ *and* $\frac{1}{2}\left(\sqrt{-(3\beta_1^2 + 4\beta_0\beta_1)} - (\beta_1 + 2\beta_0)\right) < \beta_2 < -(\beta_1 + \beta_0)^2/\beta_1$.
- $\beta_1 = 0$ *and* $\beta_2 > -\beta_0$.
- $\beta_1 > 0$ *and* $\frac{1}{2}\left(\sqrt{4\beta_0\beta_1 + \beta_1^2} - (\beta_1 + 2\beta_0)\right) < \beta_2 < \beta_0^2/\beta_1$.

On this subregion the optimal weights of a D-optimal design depend on the parameter values and thus there are various locally D-optimal designs with four design points.

Define the ratios $\gamma_1 = \beta_1/\beta_0$ and $\gamma_2 = \beta_2/\beta_0$ such that $\beta_0 > 0$ and $\gamma_1 + \gamma_2 > -1$. Without loss of generality the conditions of the D-optimal designs given in Corollary 4.4.1 can be determined in terms of γ_1 and γ_2. In Figure 4.2 the parameter subregions of γ_1 and γ_2 are depicted where the designs given by Corollary 4.4.1 are locally D-optimal. The design with support $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ is locally D-optimal over the larger subregion in particular for positive larger values of γ_1 and γ_2. It is clear in case of $\gamma_1 = \gamma_2 = 0$ the design assigns equal weights $1/4$ to $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. This case is equivalent to ordinary regression models with two binary factors.

Let us focus on the case of equally effect sizes; i.e., $\beta_1 = \beta_2 = \beta$ or equivalently, $\gamma_2 = \gamma_1 = \gamma$ where $\gamma = \beta/\beta_0$. This case corresponds to the diagonal dashed line in Figure 4.2 at which the D-optimality is achieved for only two saturated designs supported by $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ and $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and for four-point designs. The D-optimal design has four design points in the range $-1/3 < \gamma < 1$, i.e., $-(1/3)\beta_0 < \beta < \beta_0$. The next theorem gives explicit formulas for the weights of locally D-optimal four-point designs at parameter points $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ with $-(1/3)\beta_0 < \beta_1 = \beta_2 < \beta_0$.

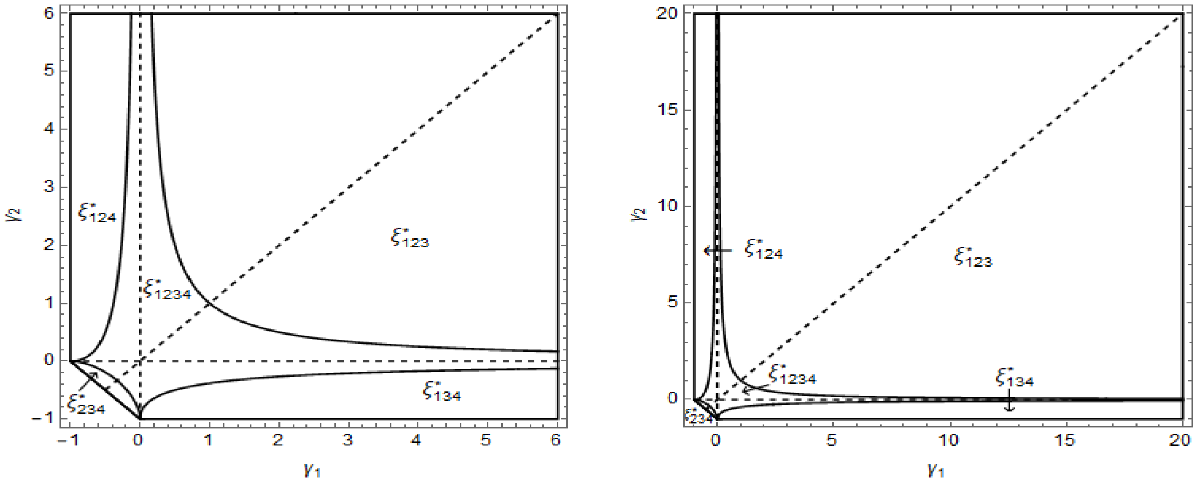


FIGURE 4.2: D-optimal designs on the respective subregions of model parameters where $\text{supp}(\xi^*_{ijk}) = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k\} \subset \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and $\text{supp}(\xi^*_{1234}) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. The diagonal dashed line is $\gamma_2 = \gamma_1$.

Theorem 4.4.1. *Under the assumptions of Corollary 4.4.1 let the parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ be such that $\beta_1 = \beta_2 = \beta$ where $-(1/3)\beta_0 < \beta < \beta_0$. Then the locally D-optimal design (at $\boldsymbol{\beta}$) is supported by the four vertices $\boldsymbol{v}_1 = (0,0)^\top$, $\boldsymbol{v}_2 = (1,0)^\top$, $\boldsymbol{v}_3 = (0,1)^\top$, $\boldsymbol{v}_4 = (1,1)^\top$ with weights*

$$\omega_1^* = \frac{3\gamma+1}{4(2\gamma+1)}, \quad \omega_2^* = \omega_3^* = \frac{(\gamma+1)^2}{4(2\gamma+1)}, \quad \omega_4^* = \frac{1-\gamma}{4}, \quad \textit{where } \gamma = \frac{\beta}{\beta_0}.$$

Proof. By Corollary 4.4.1 for a given $\boldsymbol{\beta}$ a four-point design is D-optimal if and only if the design is not minimally supported (not saturated). Therefore, under the parameter assumption $-(1/3)\beta_0 < \beta < \beta_0$ the design is only supported by four design points with positive weights determined by condition (4.28). Denote $u_k = u(\boldsymbol{v}_k, \boldsymbol{\beta})$ $(1 \le k \le 4)$, i.e.,

$$u_1 = 1/\beta_0^2, \; u_2 = u_3 = \beta_0^2/(1+\gamma)^2, \; u_4 = \beta_0^2/(1+2\gamma)^2.$$

For the weights ω_k^* stated in the theorem, elementary calculations yield

$$u_k\omega_k^*\Big(\tfrac{1}{3} - \omega_k^*\Big) = \frac{1}{\beta_0^2}\,\frac{(3\gamma+1)(1-\gamma)}{48(2\gamma+1)^2} \text{ for } k = 1,2,3,4,$$

and the result follows. □

The optimal weights ω_1^*, ω_2^* and ω_4^* of ξ^* given in Theorem 4.4.1 depend on γ. The range of γ is given by $(-1/3, 1)$. Figure 4.3 exhibits these weights as functions of γ. Clearly, the weights are positive over the respective domain $(-1/3, 1)$ and note that $1/4 \le \omega_2^* = \omega_3^* \le 1/3$. The design ξ^* at $\gamma = 0$ assigns uniform weights to the set of vertices $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. At the limits of $(-1/3, 1)$ the D-optimal four-point design becomes a D-optimal saturated design. That is at $\gamma = -1/3$ we have $\omega_1^* = 0$ and at $\gamma = 1$ we have $\omega_4^* = 0$.

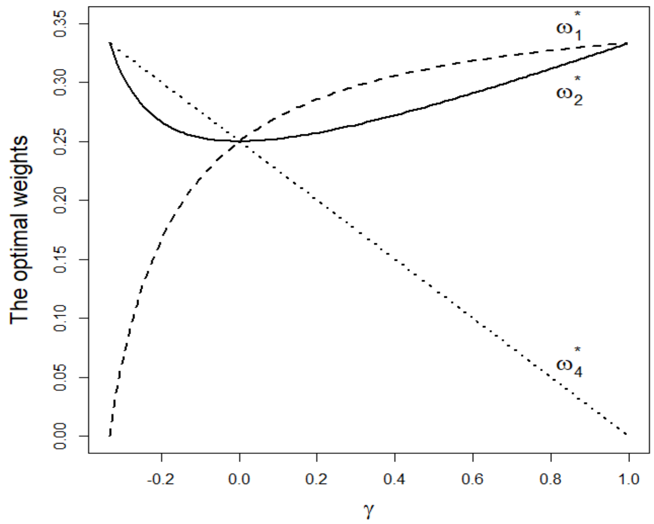


FIGURE 4.3: Effect of γ on the optimal weights ω_1^*, ω_2^* and ω_4^* of the locally D-optimal design given in Theorem 4.4.1.

Remark 4.4.2. *Actually, we did not succeed in evaluating condition (4.28) from Corollary 4.4.1 to find the explicit formulas for the weights ω_k^* $(1 \leq k \leq 4)$ given in Theorem 4.4.1. We derived them from Lemma 3.1.2 by utilizing an explicit representation of $\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big)$ when ξ is supported by the four vertices $\boldsymbol{v}_k$ $(1 \leq k \leq 4)$ with positive weights ω_k. So we have $u_2 = u_3$ and $d_k^2 = 1\,(1 \leq k \leq 4)$ where*

$$\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big) = \sum_{1 \leq h < i < j \leq 4} u_h u_i u_j \, \omega_h \omega_i \omega_j.$$

Next we study the problem of A-optimal designs for $\nu = 2$ and the unit square $[0, 1]^2$ as the experimental region. In view of Theorem 3.3.2 the following corollary presents the cases in which the locally A-optimal design (at $\boldsymbol{\beta}$) is saturated, i.e., a three-point design and moreover, the case of four-point designs.

Corollary 4.4.2. *Under the assumptions and notations of Corollary 4.4.1. The unique locally A-optimal design is as follows.*

(i) If $(3+4\sqrt{2/3})(\beta_0^2 + \beta_1\beta_2) + 2(1+\sqrt{2/3})(\beta_1^2 + \beta_2^2) + (5+6\sqrt{2/3})\beta_0(\beta_1 + \beta_2) \leq 0$ *then*

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \sqrt{2}(\beta_0 + \beta_1)/c & \sqrt{2}(\beta_0 + \beta_2)/c & \sqrt{3}(\beta_0 + \beta_1 + \beta_2)/c \end{pmatrix}.$$

(ii) If $(3+\sqrt{2})\beta_0^2 + (2+\sqrt{2})(\beta_2^2+\beta_1\beta_2) + (5+2\sqrt{2})\beta_0\beta_2 + \sqrt{2}\beta_0\beta_1 \le 0$ *then*

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \sqrt{2}\beta_0/c & \sqrt{2}(\beta_0+\beta_2)/c & (\beta_0+\beta_1+\beta_2)/c \end{pmatrix}.$$

(iii) If $(3+\sqrt{2})\beta_0^2 + (2+\sqrt{2})(\beta_1^2+\beta_1\beta_2) + (5+2\sqrt{2})\beta_0\beta_1 + \sqrt{2}\beta_0\beta_2 \le 0$ *then*

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_4 \\ \sqrt{2}\beta_0/c & \sqrt{2}(\beta_0+\beta_1)/c & (\beta_0+\beta_1+\beta_2)/c \end{pmatrix}.$$

(iv) If $(1+2/\sqrt{3})\beta_0^2 + (1/\sqrt{3})\beta_0(\beta_1+\beta_2) - \beta_1\beta_2 \le 0$ *then*

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \\ \sqrt{3}\beta_0/c & (\beta_0+\beta_1)/c & (\beta_0+\beta_2)/c \end{pmatrix}.$$

For each case (i) – (iv), the constant c appearing in the weights equals the sum of the numerators of the three ratios. If none of the cases (i) – (iv) applies then ξ^* *is supported by the four vertices* $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$.

Proof. In analogy to proof of Corollary 4.4.1 together with Theorem 4.2.1 the support of a locally A-optimal design must be a subset of $\widetilde{\mathcal{X}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. Denote $u_k = u(\boldsymbol{v}_k, \boldsymbol{\beta})$ $(1 \le k \le 4)$ and let $q_i = u_i^{-1/2}$ $(1 \le i \le 4)$. Then the proof follows that of Theorem 3.3.2 in which each condition provided by parts (i)–(iv) is equivalent to those in the corresponding parts of the corollary. □

As the optimal weights of the A-optimal designs depend on the model parameters so each condition provided in Corollary 4.4.2 characterizes a subregion of the parameter space where the corresponding designs with the same support are A-optimal. Accordingly, for each subregion there is a wide class of A-optimal designs that vary with parameter values but the support is similar.

In fact, for each of the cases *(i) – (iv)* of Corollary 4.4.2 it is not difficult to construct particular parameter points meeting the respective condition. Examples of parameter points for which case *(iii)* of the theorem applies are $\boldsymbol{\beta} = (\beta_0, -\beta, \beta)^\top$ with $\frac{3+\sqrt{2}}{5+\sqrt{2}}\beta_0 \le \beta < \beta_0$.

Again, the conditions of A-optimal designs can be written in terms of the defined ratios $\gamma_1 = \beta_1/\beta_0$ and $\gamma_2 = \beta_2/\beta_0$, $\beta_0 > 0$. In Figure 4.4 the parameter subregions of γ_1 and γ_2 are depicted where the designs given by Corollary 4.4.2 are locally A-optimal. Comparing to Figure 4.2 under D-optimality, similar interpretation might be observed. In particular, the majority of the parameter points is for A-optimal designs with support $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$.

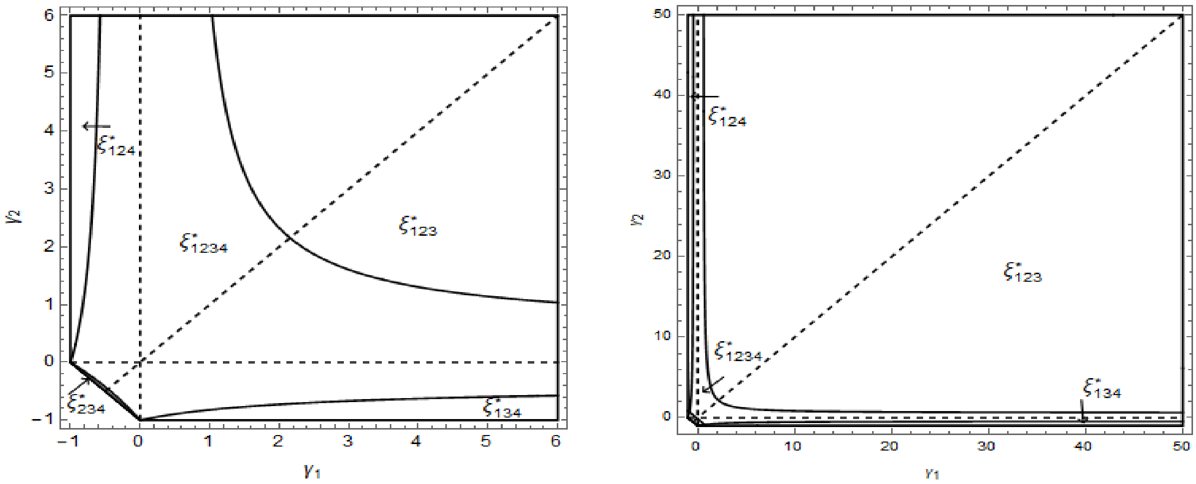


FIGURE 4.4: A-optimal designs at given parameter values $\text{supp}(\xi^*_{ijk}) = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k\} \subset \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and $\text{supp}(\xi^*_{1234}) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. The diagonal dashed line is $\gamma_2 = \gamma_1$ where $\gamma_i = \beta_i/\beta_0, i = 1, 2$.

Locally A-optimal four-point designs (for $\nu = 2$) can be computed numerically. Here, we assume $\gamma_1 = \gamma_2 = \gamma$ and thus $\gamma \in (-1/2, \infty)$. This assumption implies that the conditions in parts *(ii)* and *(iii)* of Corollary 4.4.2 are not fulfilled and thus the corresponding designs are not A-optimal. In contrast, the condition in part *(i)* is fulfilled by $\gamma \in \left(-1/2, -(9+4\sqrt{6})/(21+8\sqrt{6})\right]$ and the condition in part *(v)* is fulfilled by $\gamma \in \left[(3+\sqrt{2})/3, \infty\right)$. Then it turns out that the A-optimal four-point designs are given at parameter points γ in $\left(-(9+4\sqrt{6})/(21+8\sqrt{6}), 1+2\sqrt{3}\right)$. On that basis, the multiplicative algorithm (see Yu (2010) and Harman and Trnovská (2009)) can be employed. Table 4.1 shows some numerical results at particular parameter points γ in $\left(-(9+4\sqrt{6})/(21+8\sqrt{6}), 1+2/\sqrt{3}\right)$.

TABLE 4.1: $\nu = 2$, $\mathcal{X} = [0,1]^2$; locally A-optimal designs at parameter points $\boldsymbol{\beta} = \beta_0(1, \gamma, \gamma)^\top$ where $-\frac{9+4\sqrt{6}}{21+8\sqrt{6}} < \gamma < 1 + 2/\sqrt{3}$.

γ	$\boldsymbol{v}_1$	$\boldsymbol{v}_2$	$\boldsymbol{v}_3$	$\boldsymbol{v}_4$
−0.45	0.1136	0.3983	0.3983	0.0898
0	0.3561	0.2250	0.2250	0.1938
1	0.2700	0.3000	0.3000	0.1300
2	0.2210	0.3805	0.3805	0.0180

Now consider model (4.27) for a general number of factors, $\nu \geq 2$, where the experimental region is the ν-dimensional unit hypercube $[0,1]^\nu$. Note that condition

(4.4) characterizing the feasible parameter points $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_\nu)^\top$ is equivalent to

$$\beta_0 + \sum_{i \in S} \beta_i > 0 \quad \text{for all subsets } S \subseteq \{1, \ldots, \nu\}. \tag{4.29}$$

In Burridge and Sebastiani (1994) the following result was obtained for which we give a shorter and more transparent proof than that in the reference.

Theorem 4.4.2. (Burridge and Sebastiani (1994))
Consider model (4.27) with experimental region $\mathcal{X} = [0,1]^\nu$*, where* $\nu \geq 2$*. Let the parameter point* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_\nu)^\top$ *be given with (4.29). Then the design* ξ^* *that assigns equal weights* $1/(\nu+1)$ *to the support points*

$$\boldsymbol{x}_1^* = (0, \ldots, 0)^\top, \quad \boldsymbol{x}_2^* = (1, \ldots, 0)^\top, \quad \ldots, \quad \boldsymbol{x}_{\nu+1}^* = (0, \ldots, 1)^\top$$

is locally D-optimal (at $\boldsymbol{\beta}$*) if and only if* $\beta_0^2 \leq \beta_i \beta_j$ *for all* $1 \leq i < j \leq \nu$*.*

Proof. The complete class $\widetilde{\Xi}$ from Theorem 4.2.1 consists of all designs whose support points are vertices of the unit cube $\big[0,1\big]^\nu$, i.e., points with components from $\big\{0,1\big\}$. So it suffices to consider the reduced experimental region $\widetilde{\mathcal{X}} = \big\{0,1\big\}^\nu$ and to show that the stated design ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for model (4.27) on the experimental region $\widetilde{\mathcal{X}}$. By Theorem 3.3.3 the design ξ^* is D-optimal if and only if condition (3.17) holds true. Let $u(\boldsymbol{x}, \boldsymbol{\beta}) = (\beta_0 + \sum_{i=1}^{\nu} \beta_i x_i)^{-2}$ and denote $u_1 = u(\boldsymbol{x}_1^*, \boldsymbol{\beta}) = \beta_0^{-2}$ and $u_{i+1} = u(\boldsymbol{x}_{i+1}^*, \boldsymbol{\beta}) = (\beta_0 + \beta_i)^{-2}\,(1 \leq i \leq \nu)$. Then condition (3.17) is equivalent to (4.30) given below;

$$u_1^{-1}\big(1 - \sum_{j=1}^{\nu} x_j)^2 + \sum_{i=1}^{\nu} u_{i+1}^{-1} x_i^2 \leq (\beta_0 + \sum_{i=1}^{\nu} \beta_i x_i)^2 \quad \text{for all} \quad \boldsymbol{x} \in \big\{0,1\big\}^\nu. \tag{4.30}$$

Every $\boldsymbol{x} = (x_1, \ldots, x_\nu)^\top \in \{0,1\}^\nu$ is described by a subset $S \subseteq \{1, \ldots, \nu\}$ via

$$x_i = 1 \quad \text{if } i \in S, \quad \text{and } x_i = 0 \text{ else}.$$

So for $\boldsymbol{x}$ described by $S \subseteq \{1, \ldots, \nu\}$, and denoting $s = \#\, S$, the l.h.s. of (4.30) under model (4.27) rewrites as

$$\beta_0^2(1-s)^2 + \sum_{i \in S} (\beta_0 + \beta_i)^2 = \beta_0^2(s-1)^2 + \beta_0^2 s + 2\beta_0 \sum_{i \in S} \beta_i + \sum_{i \in S} \beta_i^2,$$

and the r.h.s. of (4.30) rewrites as

$$\Big(\beta_0 + \sum_{i \in S} \beta_i\Big)^2 = \beta_0^2 + 2\beta_0 \sum_{i \in S} \beta_i + \Big(\sum_{i \in S} \beta_i\Big)^2 = \beta_0^2 + 2\beta_0 \sum_{i \in S} \beta_i + \sum_{i \in S} \beta_i^2 + \sum_{i,j \in S,\; i \neq j} \beta_i \beta_j.$$

Hence (4.30) is equivalent to

$$\beta_0^2 s(s-1) \leq \sum_{i,j\in S,\ i\neq j} \beta_i\beta_j \ \text{ for all } S \subseteq \{1,\dots,\nu\} \ \text{ (where } s = \#S),$$

or, equivalently,

$$\sum_{i,j\in S,\ i\neq j} \Big(\beta_0^2 - \beta_i\beta_j\Big) \leq 0 \ \text{ for all } S \subseteq \{1,\dots,\nu\}. \tag{4.31}$$

By the assumption that $\beta_0^2 \leq \beta_i\beta_j$ $(1 \leq i < j \leq \nu)$, condition (4.31) holds true, i.e., the design ξ^* satisfies the condition of The Equivalence Theorem (Theorem 2.2.2, condition (2.11)) and hence ξ^* is locally D-optimal. □

Remark 4.4.3. *For $\nu = 2$ the result of Theorem 4.4.2 is covered by case (i) of Corollary 4.4.1.*

The next corollary deals with local A-optimality and may be viewed as an analogue to Theorem 4.4.2 for D-optimality.

Corollary 4.4.3. *Consider model (4.27) with experimental region $\mathcal{X} = [0,1]^\nu$, where $\nu \geq 2$. Let $\boldsymbol{\beta} = (\beta_0,\beta_1,\dots,\beta_\nu)^\top$ be a parameter point satisfying (4.29) and denote $\gamma_j = \beta_j/\beta_0$ $(1 \leq j \leq \nu)$. Then the design ξ^* that is supported by*

$$\boldsymbol{x}_1^* = (0,\dots,0)^\top, \quad \boldsymbol{x}_2^* = (1,\dots,0)^\top, \ \dots, \ \boldsymbol{x}_{\nu+1}^* = (0,\dots,1)^\top,$$

with the corresponding weights

$$\omega_1^* = \sqrt{\nu+1}/c, \ \ \omega_j^* = (1+\gamma_{j-1})/c, \ (j = 2,\dots,\nu+1), \ \ \textit{where } c = (\sqrt{\nu+1}+\nu)+\sum_{j=1}^{\nu}\gamma_j.$$

is locally A-optimal (at $\boldsymbol{\beta}$) if and only if

$$\gamma_i\gamma_j - \frac{1}{\sqrt{\nu+1}}(\gamma_i+\gamma_j) \geq \Big(1 + \frac{2}{\sqrt{\nu+1}}\Big) \ \textit{ for all } 1 \leq i < j \leq \nu \tag{4.32}$$

Proof. As in the proof of Theorem 4.4.2 we may reduce the experimental region to $\widetilde{\mathcal{X}} = \{0,1\}^\nu$. By Theorem 3.3.4 the design ξ^* is A-optimal if and only if condition (3.20) holds true. As in the proof of Theorem 4.4.2 we describe every $\boldsymbol{x} \in \{0,1\}^\nu$ by a subset $S \subseteq \{1,\dots,\nu\}$. After some elementary calculations and denoting $s = \#S$, condition (3.20) under model (4.27) rewrites as

$$\beta_0^2\, s(s-1)(1+\frac{2}{\sqrt{\nu+1}})+\frac{2\,(s-1)}{\sqrt{\nu+1}}\beta_0\sum_{i\in S}\beta_i - \sum_{i,j\in S, i\neq j}\beta_i\beta_j \leq 0 \ \forall\, S \subseteq \{1,\dots,\nu\},\ \#S \geq 2,$$

which is equivalent to

$$\sum_{i,j\in S,\, i\neq j}\gamma_i\gamma_j-\frac{2\,(s-1)}{\sqrt{\nu+1}}\sum_{i\in S}\gamma_i-s(s-1)(1+\frac{2}{\sqrt{\nu+1}})\geq 0\ \ \forall\ S\subseteq\{1,\ldots,\nu\},\ \#S\geq 2. \tag{4.33}$$

and (4.33) is equivalent to

$$\sum_{i,j\in S,\, i\neq j}\Big(\gamma_i\gamma_j-\frac{1}{\sqrt{\nu+1}}\big(\gamma_i+\gamma_j\big)-(1+\frac{2}{\sqrt{\nu+1}})\Big)\geq 0\ \ \forall\ S\subseteq\{1,\ldots,\nu\},\ \#S\geq 2.$$

The above system of inequalities is equivalent to (4.32). Thus the design ξ^* satisfies the condition of The Equivalence Theorem (Theorem 2.2.2, condition (2.12)) and hence ξ^* is locally A-optimal. □

Remark 4.4.4. *For $\nu=2$ the result is covered by case (iv) of Corollary 4.4.2. In fact, it can easily be checked that, with the notations of Corollary 4.4.2, the inequality $(1+2/\sqrt{3})\beta_0^2+(1/\sqrt{3})\beta_0(\beta_1+\beta_2)-\beta_1\beta_2\leq 0$ is equivalent to assumption (4.32) of Corollary 4.4.3 for $\nu=2$.*

4.4.2 Model without intercept

In this subsection we restrict to the first order model without intercept,

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{x},\quad \text{where}\quad \boldsymbol{x}=(x_1,\ldots,x_\nu)^\top,\ \nu\geq 2,\ \ \boldsymbol{x}\in\mathcal{X}, \tag{4.34}$$

with

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})=\frac{1}{\beta_1x_1+\cdots+\beta_\nu x_\nu}\begin{pmatrix}x_1\\ \vdots\\ x_\nu\end{pmatrix}. \tag{4.35}$$

Clearly, $\boldsymbol{f}(\mathbf{0})=\mathbf{0}$ and thus condition (4.4), i.e., $\boldsymbol{x}^\top\boldsymbol{\beta}>0$ for all $\boldsymbol{x}\in\mathcal{X}$ requires $\mathbf{0}\notin\mathcal{X}$.

Firstly consider the experimental region $\mathcal{X}=[0,\infty)^\nu\setminus\{\mathbf{0}\}$. The proposed experimental region is no longer compact therefore the existence of optimal designs is not assured and has to be checked separately. To this end, denote by $\boldsymbol{e}_i$ for all $(1\leq i\leq\nu)$ the ν-dimensional unit vectors. The parameter space is determined by condition (4.4) which implies that $\boldsymbol{\beta}\in(0,\infty)^\nu$, i.e., $\beta_i>0$ for all $(1\leq i\leq\nu)$. In view of Theorem 3.5.1 the next result is immediate.

Corollary 4.4.4. *Consider model (4.34) with the experimental region $\mathcal{X}=[0,\infty)^\nu\setminus\{\mathbf{0}\}$. Given a vector $\boldsymbol{a}=(a_1,\ldots,a_\nu)^\top$ where $a_i\in\mathbb{R},\ a_i>0\ (1\leq i\leq\nu)$. Let $\boldsymbol{x}_i^*=a_i\boldsymbol{e}_i$ for all $i=1,\ldots,\nu$ denote the design points which are assumed to belong to $\mathcal{X}$. For a given parameter point $\boldsymbol{\beta}\in(0,\infty)^\nu$ let $\xi_{\boldsymbol{a}}^*$ be the saturated design whose support is $\boldsymbol{x}_i^*$*

$(1 \le i \le \nu)$ with the corresponding weights

$$\omega_i^* = \frac{\beta_i^{\frac{2k}{k+1}}}{\sum\limits_{j=1}^{\nu} \beta_j^{\frac{2k}{k+1}}} \ (1 \le i \le \nu).$$

Then $\xi_{\boldsymbol{a}}^*$ is locally Φ_k-optimal (at $\boldsymbol{\beta}$).

Proof. The corollary covers the result of Theorem 3.5.1 under gamma model (4.34). For a given $\boldsymbol{\beta} \in (0, \infty)^\nu$ under model (4.34) let $u_i = u(\boldsymbol{x}_i^*, \boldsymbol{\beta})$ $(1 \le i \le \nu)$. Thus $u_i = (a_i\beta_i)^{-2}$ $(1 \le i \le \nu)$. Then condition (3.22) of Theorem 3.5.1 is equivalent to $-2 \sum\limits_{i<j=1}^{\nu} \beta_i\beta_j x_i x_j \le 0$ for all $\boldsymbol{x} \in \mathcal{X}$. Since $\beta_i > 0, x_i > 0$ $(1 \le i \le \nu)$ the condition holds true for any $\boldsymbol{x} \in \mathcal{X}$ at any given $\boldsymbol{\beta} \in (0, \infty)^\nu$. □

Note that for any point $\boldsymbol{\beta}$ the locally Φ_k-optimal design given by Corollary 4.4.4 is not unique. That is at $\boldsymbol{\beta}$ any set of ν design points that are located at all distinct edges of $\mathcal{X}$ is a support of a locally Φ_k-optimal design. Given two constant vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ such that the designs $\xi_{\boldsymbol{a}_1}^*$ and $\xi_{\boldsymbol{a}_2}^*$ from Corollary 4.4.4 are locally Φ_k-optimal at a given $\boldsymbol{\beta}$. Then their convex combination $\alpha\xi_{\boldsymbol{a}_1}^* + (1-\alpha)\xi_{\boldsymbol{a}_2}^*$ where $(0 \le \alpha \le 1)$ is locally Φ_k-optimal according to Remark 2.2.3 which asserts $\boldsymbol{M}(\xi_{\boldsymbol{a}_1}^*, \boldsymbol{\beta}) = \boldsymbol{M}(\xi_{\boldsymbol{a}_2}^*, \boldsymbol{\beta}) = \boldsymbol{M}(\alpha\xi_{\boldsymbol{a}_1}^* + (1-\alpha)\xi_{\boldsymbol{a}_2}^*, \boldsymbol{\beta})$. The next corollary is immediate.

Corollary 4.4.5. *Under assumptions of Corollary 4.4.4 let a parameter point $\boldsymbol{\beta}$ be given. Then $\Xi^* = \mathrm{Conv}\{\xi_{\boldsymbol{a}}^* : \boldsymbol{a} = (a_1, \ldots, a_\nu)^\top, a_i > 0 \ \forall i = 1, \ldots, \nu\}$ is a set of locally Φ_k-optimal designs (at $\boldsymbol{\beta}$).*

Example 4.4.1. Let $\nu = 3$ and take ξ_1^* with support $(1,0,0)^\top$, $(0,1,0)^\top$, $(0,0,7)^\top$ and ξ_2^* with support $(1,0,0)^\top$, $(0,6,0)^\top$, $(0,0,1)^\top$. Let ξ_1^* and ξ_2^* be locally Φ_k-optimal (at $\boldsymbol{\beta}$) under the assumptions of Corollary 4.4.4. Then $\xi_3^* = \alpha\xi_1^* + (1-\alpha)\xi_2^*$ has support $(1,0,0)^\top$, $(0,1,0)^\top$, $(0,0,7)^\top$, $(0,6,0)^\top$, $(0,0,1)^\top$ and is locally Φ_k-optimal (at $\boldsymbol{\beta}$).

Remark 4.4.5. *Let us denote by $\psi(\boldsymbol{x}, \xi)$ the left hand side (the sensitivity function) of The Equivalence Theorems, Theorem 2.2.2. Actually, under non-intercept gamma models $\psi(\boldsymbol{x}, \xi)$ is invariant with respect to simultaneous scale transformation of $\boldsymbol{x}$, i.e., $\psi(\lambda\boldsymbol{x}, \xi) = \psi(\boldsymbol{x}, \xi)$ for any $\lambda > 0$. This essentially comes from the fact that the function $\boldsymbol{f}_\beta(\boldsymbol{x}) = \left(\boldsymbol{x}^\top\boldsymbol{\beta}\right)^{-1}\boldsymbol{x}$ from (4.35) is invariant with respect to simultaneous rescaling of the components of $\boldsymbol{x}$, i.e., $\boldsymbol{f}_\beta(\lambda\boldsymbol{x}) = \boldsymbol{f}_\beta(\boldsymbol{x})$. This property is explicitly transferred to the information matrix (4.5) since it can be represented in form $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_\beta(\boldsymbol{x})\boldsymbol{f}_\beta^\top(\boldsymbol{x})$, and hence $\boldsymbol{M}(\lambda\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta})$. In fact, this property plays a main rule in the solution of the forthcoming optimal designs.*

In what follows we consider a hypercube $\mathcal{X} = [a, b]^\nu, \nu \ge 2$, $0 < a < b$, as an experimental region. As the function $\boldsymbol{f}_\beta(\boldsymbol{x})$ from (4.35) is not injective we know from

Theorem 4.2.1 that the set $\tilde{\Xi}$ of those designs which are supported only by the vertices of $\mathcal{X}$ is an essentially locally complete class of designs (at $\boldsymbol{\beta}$), and hence there exists a locally D- or A-optimal design (at $\boldsymbol{\beta}$) in that class.

As pointed out in Remark 4.4.5, we have $\boldsymbol{f}_{\boldsymbol{\beta}}(\lambda \boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$, $\lambda > 0$ and thus a transformation of a gamma model without intercept to a gamma model with intercept can be obtained if, in particular, $\lambda = x_1^{-1}$, $x_1 > 0$. This reduction is useful to determine precisely the candidate support points of a design.

Let us begin with the simplest case $\nu = 2$. A transformation of a two-factor model without intercept to a single-factor model with intercept is employed. Based on that D- and A-optimal designs are derived.

Theorem 4.4.3. *Consider the experimental region $\mathcal{X} = [a, b]^2$, $0 < a < b$. Let $\boldsymbol{x}_1^* = (a, b)^\top$ and $\boldsymbol{x}_2^* = (b, a)^\top$. Let $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ be given such that $\boldsymbol{\beta}^\top \boldsymbol{x}_i^* > 0$ for all $i = 1, 2$ (which is equivalent to condition (4.4)). Then, the unique locally D-optimal design ξ_{D}^* (at $\boldsymbol{\beta}$) is the two-point design supported by $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ with equal weights $1/2$. The unique locally A-optimal design ξ_{A}^* (at $\boldsymbol{\beta}$) is the two-point design supported by $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ with weights $\omega_1^* = \frac{\beta_1 b + \beta_2 a}{(\beta_1+\beta_2)(a+b)}$ and $\omega_2^* = \frac{\beta_1 a + \beta_2 b}{(\beta_1+\beta_2)(a+b)}$.*

Proof. Since $\boldsymbol{f}_{\boldsymbol{\beta}}(x_1^{-1}\boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ for all $\boldsymbol{x} = (x_1, x_2)^\top \in [a, b]^2$, we write

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \Big(\beta_1 x_1 + \beta_2 x_2\Big)^{-1} \Big(x_1\,,\,x_2\Big)^\top = \Big(\beta_1 + \beta_2 t\Big)^{-1} \Big(1\,,\,t\Big)^\top, \quad \text{where} \;\; t = t(\boldsymbol{x}) = x_2/x_1.$$

So the information matrices coincide with those from a single-factor gamma model with intercept. The range of $t = t(\boldsymbol{x})$, as $\boldsymbol{x}$ ranges over $[a, b]^2$ is the interval $\Big[(a/b)\,,\,(b/a)\Big]$. Note also that the end points a/b and b/a come from the unique points $\boldsymbol{x}_1^* = (a, b)^\top$ and $\boldsymbol{x}_2^* = (b, a)^\top$, respectively. This together with Corollary 4.3.1 yields the stated results on the locally D- and A-optimal designs, where for local A-optimality we get

$$\omega_1^* = \frac{\Big(\beta_1 + \beta_2 \frac{a}{b}\Big)\sqrt{1 + (\frac{b}{a})^2}}{\Big(\beta_1 + \beta_2 \frac{a}{b}\Big)\sqrt{1 + (\frac{b}{a})^2} + \Big(\beta_1 + \beta_2 \frac{b}{a}\Big)\sqrt{1 + (\frac{a}{b})^2}}$$

and it is straightforward to verify that the above quantity is equal to $\frac{\beta_1 b + \beta_2 a}{(\beta_1+\beta_2)(a+b)}$. □

Remark 4.4.6. *Actually, in case $\nu \geq 3$ an analogous transformation of the model as in the proof of Theorem 4.4.3 is obvious,*

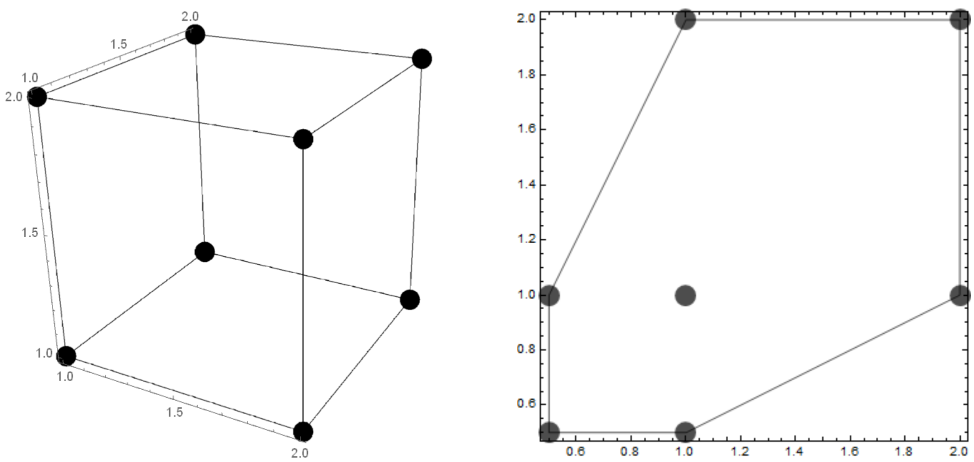
$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \Big(\beta_1 + \beta_2 t_1 + \beta_3 t_2 + \ldots + \beta_\nu t_{\nu-1}\Big)^{-1} \Big(1, t_1, \ldots, t_{\nu-1}\Big)^\top$$

where $t_j = t_j(\boldsymbol{x}) = x_{j+1}/x_1$ $(1 \leq j \leq \nu - 1)$ for $\boldsymbol{x} = (x_1, \ldots, x_\nu)^\top \in [a, b]^\nu, 0 < a < b$, leading thus to a first order model with intercept employing a $(\nu-1)$-dimensional factor

$\boldsymbol{t} = (t_1, \ldots, t_{\nu-1})^\top$. *However, its range* $\left\{\boldsymbol{t}(\boldsymbol{x}) : \boldsymbol{x} \in [a,b]^\nu\right\} \subseteq \mathbb{R}^{\nu-1}$ *is not a cube but a more complicated polytope. E.g., for* $\nu = 3$ *it can be shown that*

$$\left\{\boldsymbol{t}(\boldsymbol{x}) : \boldsymbol{x} \in [a,b]^3\right\} = \mathrm{Conv}\left\{\begin{pmatrix} a/b \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ a/b \end{pmatrix}, \begin{pmatrix} a/b \\ a/b \end{pmatrix}, \begin{pmatrix} b/a \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ b/a \end{pmatrix}, \begin{pmatrix} b/a \\ b/a \end{pmatrix}\right\}$$

where for each $\boldsymbol{x} \in [a,b]^3$ *we get* $\boldsymbol{t}(\boldsymbol{x}) = (x_2/x_1, x_3/x_1)^\top$ *as it is depicted in Figure 4.5 for, in specific,* $a = 1$ *and* $b = 2$*. One notes that for each vertex* $\boldsymbol{v} \in \{(a,a,a)^\top, (b,b,b)^\top\}$ *we get* $\boldsymbol{t}(\boldsymbol{v}) = (1,1)^\top$ *which lies in the interior of the convex hull above, i.e.,* $(1,1)^\top$ *is a proper convex combination of the vertices of the polytope. Thus this reduction on the vertices implies that both vertices* $(a,a,a)^\top$ *and* $(b,b,b)^\top$ *of the hupercube* $[a,b]^3$ *are out of consideration as support points of any optimal design.*



(a) (b)

FIGURE 4.5: Panel (a): The experimental region $\mathcal{X} = [1,2]^3$. Panel (b): The transformed experimental region $\mathrm{Conv}\{(1/2,1/2)^\top, (1/2,1)^\top, (1,1/2)^\top, (2,1)^\top, (1,2)^\top, (2,2)^\top\}$. The interior point is $(1,1)^\top$ which represents the original points $(1,1,1)^\top$ and $(2,2,2)^\top$ in $[1,2]^3$.

Let us concentrate on the experimental region $\mathcal{X} = [1,2]^3$. The linear predictor of a three-factor gamma model is given by $\eta(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Assume that $\beta_2 = \beta_3 = \beta$, so the set of all parameter points under condition (4.4), i.e., $\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 > 0$ for all $\boldsymbol{x} = (x_1, x_2, x_3)^\top \in \mathcal{X}$ is characterized by

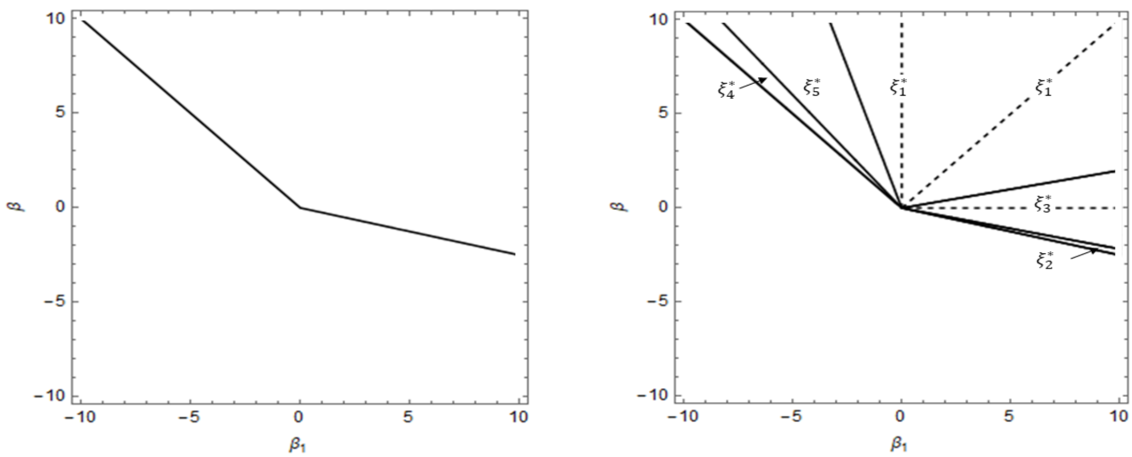
$$\beta_1 \leq 0,\ \beta > -\beta_1 \ \text{ or } \ \beta_1 > 0,\ \beta > -\frac{1}{4}\beta_1$$

which is shown by Panel (a) of Figure 4.6. We aim at finding locally D-optimal designs at a given parameter point in this space. Let the vertices of $\mathcal{X} = [1,2]^3$ be denoted by $\boldsymbol{v}_1 = \big(1,1,1\big)^\top$, $\boldsymbol{v}_2 = \big(2,1,1\big)^\top$, $\boldsymbol{v}_3 = \big(1,2,1\big)^\top$, $\boldsymbol{v}_4 = \big(1,1,2\big)^\top$, $\boldsymbol{v}_5 = \big(1,2,2\big)^\top$, $\boldsymbol{v}_6 = \big(2,1,2\big)^\top$, $\boldsymbol{v}_7 = \big(2,2,1\big)^\top$, $\boldsymbol{v}_8 = \big(2,2,2\big)^\top$ with intensities $u_i = u(\boldsymbol{v}_i, \boldsymbol{\beta}),\ i = 1, \ldots, \nu$.

In Theorem 4.4.4, below, we introduce analytic solutions for the locally D-optimal designs on respective optimality subregions. The results are also shown in Panel (b) of Figure 4.6 where, in particular, the solution of locally D-optimal designs of type ξ_5^* at a point $\boldsymbol{\beta}$ from the subregion $-3\beta_1 < \beta < -\frac{6}{5}\beta_1$, $\beta_1 < 0$ cannot be developed analytically so that numerical results are to be derived (cp. Remark 4.4.7). Table 4.2 presents the order of the intensities in all optimality subregions and the corresponding D-optimal designs that determined by Theorem 4.4.4 . The intensities for both vertices $\boldsymbol{v}_1$ and $\boldsymbol{v}_8$ are ignored due to the reduction (cp. Remark 4.4.6). It is noted that on each subregion the vertices of highest intensities perform mostly as a support of the corresponding D-optimal design.

TABLE 4.2: The order of intensity values according to subregions correspond to D-optimal designs

Subregions	Intensities order	D-optimal design
$\beta > 0, \beta_1 = 0$	$u_2 > u_3 = u_4 = u_6 = u_7 > u_5$	ξ_1^*
$\beta \geq -3\beta_1$, $\beta_1 < 0$	$u_2 > u_6 = u_7 \approx u_3 = u_4 > u_5$	ξ_1^*
$\beta > \frac{1}{5}\beta_1$, $\beta_1 > 0$	$u_2 > u_3 = u_4 > u_6 = u_7 > u_5$	ξ_1^*
$-\frac{1}{4}\beta_1 < \beta \leq -\frac{5}{23}\beta_1$, $\beta_1 > 0$	$u_5 > u_3 = u_4 > u_6 = u_7 > u_2$	ξ_2^*
$-\frac{5}{23}\beta_1 < \beta < \frac{1}{5}\beta_1$, $\beta_1 > 0$	$u_3 = u_4 \geq u_5 > u_2 \geq u_6 = u_7$	ξ_3^*
$-\beta_1 < \beta \leq -\frac{6}{5}\beta_1$, $\beta_1 < 0$	$u_2 > u_6 = u_7 > u_3 = u_4 > u_5$	ξ_4^*
$-3\beta_1 < \beta < -\frac{6}{5}\beta_1$, $\beta_1 < 0$	$u_2 > u_6 = u_7 > u_3 = u_4 > u_5$	ξ_5^*



(a) (b)

FIGURE 4.6: Panel (a): The parameter space of $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ such that $\beta_2 = \beta_3 = \beta$. Panel (b): Dependence of locally D-optimal designs from Theorem 4.4.4 on $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ such that $\beta_2 = \beta_3 = \beta$. The dashed lines are; diagonal: $\beta = \beta_1$, vertical: $\beta_1 = 0$, horizontal: $\beta = 0$.

Theorem 4.4.4. *Consider the experimental region* $\mathcal{X} = [1, 2]^3$*. Let a parameter point* $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ *be given such that* $\beta_2 = \beta_3 = \beta$ *with either* $\beta > -\beta_1,\ \beta_1 \leq 0$ *or* $\beta > -\frac{1}{4}\beta_1,\ \beta_1 > 0$*. Then the following designs are locally D-optimal (at* $\boldsymbol{\beta}$*).*

(i) If $\beta > 0, \beta_1 = 0$ *or* $\beta \geq -3\beta_1, \beta_1 < 0$ *or* $\beta > \frac{1}{5}\beta_1, \beta_1 > 0$ *then*

$$\xi_1^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

(ii) If $-\frac{1}{4}\beta_1 < \beta \leq -\frac{5}{23}\beta_1,\ \beta_1 > 0$ *then*

$$\xi_2^* = \begin{pmatrix} \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_5 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

(iii) If $-\frac{5}{23}\beta_1 < \beta < \frac{1}{5}\beta_1,\ \beta_1 > 0$ *then*

$$\xi_3^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_5 \\ \omega_1^* & \omega_2^* & \omega_3^* & \omega_4^* \end{pmatrix}.$$

where

$$\omega_1^* = \frac{5 + 23\,\gamma}{16\,(1 + 4\,\gamma)},\ \omega_2^* = \omega_3^* = \frac{9\,(1 + 3\gamma)^2}{32\,(1 + \gamma)(1 + 4\,\gamma)},\ \omega_4^* = \frac{1 - \gamma - 20\,\gamma^2}{8\,(1 + \gamma)(1 + 4\,\gamma)},\ \gamma = \frac{\beta}{\beta_1}.$$

(iv) If $-\beta_1 < \beta \leq -\frac{6}{5}\beta_1,\ \beta_1 < 0$ *then*

$$\xi_4^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_6 & \boldsymbol{v}_7 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Proof. The proof is obtained by making use of condition (2.11) of the Equivalence Theorem (Theorem 2.2.2). So that we develop a system of feasible inequalities evaluated at the vertices $\boldsymbol{v}_i$ for all $(1 \leq i \leq 8)$. For simplicity in computations, when $\beta_1 \neq 0$ we utilize the ratio $\gamma = \beta/\beta_1$ of which the range is given by $(-\infty, -1) \cup (-\frac{1}{4}, \infty)$. It turns out that some inequalities are equivalent and thus a resulted system is reduced to an equivalent system of a few inequalities. The intersection of the set of solutions of each system with the range of γ leads to the optimality condition (subregion) of the corresponding optimal design. For saturated designs given in cases (i), (ii), (iv) we report the 3×3 design matrix $\boldsymbol{F}$ with $\boldsymbol{F}^{-1}$ and the 3×3 weight matrix $\boldsymbol{V}$. Note that for $\beta_1 \neq 0$,

$$u_1 = \beta_1^{-2}\Big(1 + 2\,\gamma\Big)^{-2},\; u_2 = \beta_1^{-2}\Big(2 + 2\,\gamma\Big)^{-2},\; u_3 = u_4 = \beta_1^{-2}\Big(1 + 3\,\gamma\Big)^{-2},$$
$$u_5 = \beta_1^{-2}\Big(1 + 4\,\gamma\Big)^{-2},\; u_6 = u_7 = \beta_1^{-2}\Big(2 + 3\,\gamma\Big)^{-2},\; u_8 = \beta_1^{-2}\Big(2 + 4\,\gamma\Big)^{-2}.$$

Ad (i) The 3×3 design matrix $\boldsymbol{F} = [\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4]^\top$ is given by

$$\boldsymbol{F} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{with} \quad \boldsymbol{F}^{-1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} \quad \text{and } \boldsymbol{V} = \operatorname{diag}\Big(u_2, u_3, u_4\Big).$$

Hence, the condition of The Equivalence Theorem is given by

$$\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{F}^{-1}\boldsymbol{V}^{-1}\Big(\boldsymbol{F}^\top\Big)^{-1}\boldsymbol{f}(\boldsymbol{x}) \leq \Big(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3\Big)^2 \quad \forall \boldsymbol{x} \in \{1,2\}^3. \tag{4.36}$$

For case $\beta > 0, \beta_1 = 0$, condition (4.36) is equivalent to

$$4\Big(3x_1 - (x_2 + x_3)\Big)^2 + 9\Big(\Big(3x_2 - (x_1 + x_3)\Big)^2 + \Big(3x_3 - (x_1 + x_2)\Big)^2\Big) \leq 16\Big(x_2 + x_3\Big)^2$$

for all $\boldsymbol{x} \in \{1,2\}^3$, which is independent of β and is satisfied by $\boldsymbol{v}_i$ for all $(1 \leq i \leq 8)$ with equality holds for the support. For the other cases, i.e., $\beta \geq -3\beta_1, \beta_1 < 0$ or $\beta > \frac{1}{5}\beta_1, \beta_1 > 0$ condition (4.36) is equivalent to

$$\begin{aligned} &\Big(3x_1 - (x_2 + x_3)\Big)^2(2 + 2\,\gamma)^2 + \Big(\Big(3x_2 - (x_1 + x_3)\Big)^2 \\ &+ \Big(3x_3 - (x_1 + x_2)\Big)^2\Big)(1 + 3\,\gamma)^2 \leq 16\Big(x_1 + \gamma(x_2 + x_3)\Big)^2 \; \forall \boldsymbol{x} \in \{1,2\}^3. \end{aligned} \tag{4.37}$$

After some lengthy but straightforward calculations, the above inequalities reduce to

$$15\gamma^2 + 2\gamma - 1 \geq 0 \tag{4.38}$$
$$3\gamma^2 + 10\gamma + 3 \geq 0 \tag{4.39}$$

where (4.38) arises from the vertex $\boldsymbol{v}_5$ and (4.39) arises from the vertices $\boldsymbol{v}_6$ and $\boldsymbol{v}_7$. The l.h.s. of each of (4.38) and (4.39) above is a polynomial in γ of degree 2 and thus the sets of solutions are given by $(-\infty, -\frac{1}{3}] \cup [\frac{1}{5}, \infty)$ and $(-\infty, -3] \cup [-\frac{1}{3}, \infty)$, respectively. Note that the interior bounds are the roots of the respective polynomials. Hence, by considering the intersection of both sets with the range of γ, the design ξ_1^* is locally D-optimal if $\gamma \in (-\infty, -3] \cup [\frac{1}{5}, \infty)$ which is equivalent to the optimality subregion $\beta \geq -3\beta_1, \beta_1 < 0$ or $\beta > \frac{1}{5}\beta_1, \beta_1 > 0$ given in part (i) of the theorem.

Ad (*ii*) The 3×3 design matrix $\boldsymbol{F} = [\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5]^\top$ is given by

$$\boldsymbol{F} = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \quad \text{with} \quad \boldsymbol{F}^{-1} = \begin{pmatrix} 2 & 2 & -3 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \operatorname{diag}\Big(u_3, u_4, u_5\Big).$$

Hence, the condition of The Equivalence Theorem is equivalent to

$$\Big(\big(2\,x_1 - x_2\big)^2 + \big(2\,x_1 - x_3\big)^2\Big)\,\big(1 + 3\,\gamma\big)^2 \\ + \big(x_3 + x_2 - 3\,x_1\big)^2\,\big(1 + 4\,\gamma\big)^2 \leq \big(x_1 + \gamma(x_2 + x_3)\big)^2 \;\; \forall\, \boldsymbol{x} \in \{1, 2\}^3,$$

which reduce to

$$69\gamma^2 + 38\gamma + 5 \leq 0 \tag{4.40}$$

which arises from the vertex $\boldsymbol{v}_2$. Again, the set of solutions of the polynomial determined by the l.h.s. of inequality (4.40) is given by $[-\frac{1}{3}, -\frac{5}{23}]$. By considering the intersection with the range of γ, the design ξ_2^* is locally D-optimal if $\gamma \in (-\frac{1}{4}, -\frac{5}{23}]$.
Ad (*iii*) Consider design ξ_3^*. Note that $\omega_1^* > 0$ for all $\gamma > -5/23$, $\omega_2^* > 0$ for all $\gamma \in \mathbb{R}$ and $\omega_4^* > 0$ for all $\gamma \in (-\frac{1}{4}, \frac{1}{5})$, and thus it is obvious that $\omega_1^*, \omega_2^*, \omega_4^*$ are positive over $(-\frac{5}{23}, \frac{1}{5})$ and $\sum_{i=1}^{4} \omega_i^* = 1$. The 4×3 design matrix is given by $\boldsymbol{F} = [\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5]^\top$ with weight matrix $\boldsymbol{V} = \operatorname{diag}\big(s_2, s_3, s_4, s_5\big)$ where $s_i = \omega_i^* u_i, i = 2, 3, 4, 5$ and $s_3 = s_4$. The information matrix is given by

$$\boldsymbol{M}\Big(\xi_3^*, \boldsymbol{\beta}\Big) = \begin{pmatrix} 4\,s_2 + 2\,s_3 + s_5 & 2\,s_2 + 3\,s_3 + 2\,s_5 & 2\,s_2 + 3\,s_3 + 2\,s_5 \\ 2\,s_2 + 3\,s_2 + 2\,s_5 & s_2 + 5\,s_3 + 4\,s_5 & s_2 + 4\,s_3 + 4\,s_5 \\ 2\,s_2 + 3\,s_3 + 2\,s_5 & s_2 + 4\,s_3 + 4\,s_4 & s_2 + 5\,s_3 + 4\,s_5 \end{pmatrix}$$

and one calculates $\det \boldsymbol{M}\Big(\xi_3^*, \boldsymbol{\beta}\Big) = 16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5$. Define the following quantities

$$c_1 = \frac{s_3(2\,s_2 + 9\,s_3 + 8\,s_5)}{16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5}, \;\; c_2 = \frac{-s_3(2\,s_2 + 3\,s_3 + 2\,s_5)}{16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5}, \\ c_3 = \frac{10\,s_2\,s_3 + 9\,s_2\,s_5 + s_3^2 + s_3\,s_5}{16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5}, \;\; c_4 = \frac{-6\,s_2\,s_3 + 9\,s_2\,s_5 - s_3^2}{16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5}.$$

The inverse of the information matrix $\boldsymbol{M}\Big(\xi_3^*, \boldsymbol{\beta}\Big)$ is given by

$\boldsymbol{M}^{-1}\big(\xi_3^*, \boldsymbol{\beta}\big) = \begin{pmatrix} c_1 & c_2 & c_2 \\ c_2 & c_3 & c_4 \\ c_2 & c_4 & c_3 \end{pmatrix}$. Hence, the condition of The Equivalence Theorem is equivalent to

$c_1\, x_1^2 + c_3\,(x_2^2 + x_3^2) + 2\, c_2\,(x_1\, x_2 + x_1\, x_3) + 2\, c_4\, x_2\, x_3 \leq 3\,\Big(x_1 + \gamma\,(x_2 + x_3)\Big)^2\ \forall\ \boldsymbol{x} \in \{1, 2\}^3$

which is equivalent to the following system of inequalities

$$\begin{aligned} c_1 + 4c_2 + 2c_3 + 2c_4 &\leq 3\,(1 + 2\gamma)^2 \\ 4c_1 + 12c_2 + 5c_3 + 4c_4 &\leq 3\,(2 + 3\gamma)^2 \end{aligned}$$

where the first inequality arises from the vertices $\boldsymbol{v}_1$ and $\boldsymbol{v}_8$ and the second inequality comes from the vertices $\boldsymbol{v}_6$ and $\boldsymbol{v}_7$. However, due to the complexity of the system above we employed computer algebra using Wolfram Mathematica 11.3 (see Wolfram Research, Inc. (2018)) to obtain the solution for γ.

Ad *(iv)* The 3×3 design matrix $\boldsymbol{F} = [\boldsymbol{v}_2, \boldsymbol{v}_6, \boldsymbol{v}_7]^\top$ is given by

$$\boldsymbol{F} = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix} \quad \text{with} \quad \boldsymbol{F}^{-1} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \mathrm{diag}\big(u_2, u_6, u_7\big).$$

Hence, the condition of The Equivalence Theorem is equivalent to

$$\left(\left(x_2 - \frac{x_1}{2}\right)^2 + \left(x_3 - \frac{x_1}{2}\right)^2\right)(2 + 3\gamma)^2 + \left(\frac{3x_1}{2} - x_2 - x_3\right)^2 (2 + 2\gamma)^2 \leq \Big(x_1 + \gamma(x_2 + x_3)\Big)^2$$

for all $\boldsymbol{x} \in \{1, 2\}^3$, which reduce to

$$90\,\gamma^2 + 168\gamma + 72 \leq 0 \tag{4.41}$$

$$6\gamma^2 + 16\gamma + 8 \leq 0 \tag{4.42}$$

where (4.41) arises from the vertices $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ and (4.42) arises from the vertex $\boldsymbol{v}_8$. In analogy to parts (i) and (ii) the sets of solutions of (4.41) and (4.42) are given by $[-1.2, -\frac{2}{3}]$ and $[-2, -\frac{2}{3}]$, respectively where the interior bounds are the roots of the respective polynomials. Hence, by considering the intersection of both sets with the range of γ, the design ξ_4^* is locally D-optimal if $\gamma \in [-1.2, -1)$. □

In Panel (b) of Figure 4.6 the optimality subregions of ξ_1^*, ξ_2^*, ξ_3^* and ξ_4^* from Theorem 4.4.4 are depicted. Note that each design of ξ_1^*, ξ_2^* and ξ_4^* denotes a single design whereas ξ_3^* determines a certain type of designs with weights depend on the parameter values. A well known form of ξ_3^* is obtained at $\beta = (-1/7)\beta_1$ which represents the uniform design on the vertices $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$. Additionally, along the horizontal dashed line, i.e., $\beta = 0$, ξ_3^* assigns the weights $\omega_1^* = 5/16$, $\omega_2^* = \omega_3^* = 9/32$, $\omega_4^* = 1/8$ to $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$, respectively. For equally size of parameters, i.e., $\beta_1 = \beta$ the diagonal dashed line in Panel (b) represents a case where ξ_1^* is D-optimal.

Remark 4.4.7. *Deriving a locally D-optimal design at a given parameter point from the subregion $-3\beta_1 < \beta < -(6/5)\beta_1,\ \ \beta_1 < 0$ is not available analytically. Therefore, employing the multiplicative algorithm (see Yu (2010) and Harman and Trnovská (2009)) in the software package* ***R*** *(see R Core Team (2018)) provides numerical solutions which show that the locally D-optimal design on that subregion is of form*

$$\xi_5^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_6 & \boldsymbol{v}_7 \\ \omega_1^* & \omega_2^* & \omega_2^* & \omega_3^* & \omega_3^* \end{pmatrix}$$

which is supported by five vertices with weights may depend on $\boldsymbol{\beta}$. The equal weights are due to the symmetry. Table 4.3 shows some numerical results in terms of the ratio $\gamma = \beta/\beta_1$ where $\gamma \in (-3, -6/5)$.

Table 4.3: D-optimal designs on $\mathcal{X} = [1,2]^3$ at $\gamma \in (-3, -6/5)$ where $\gamma = \beta/\beta_1$ and $-3\beta_1 < \beta < -(6/5)\beta_1,\ \ \beta_1 < 0$.

γ	$\boldsymbol{v}_2$	$\boldsymbol{v}_3$	$\boldsymbol{v}_4$	$\boldsymbol{v}_6$	$\boldsymbol{v}_7$
−2.9	0.3312	0.3285	0.3285	0.0059	0.0059
−2.5	0.3225	0.3051	0.3051	0.0336	0.0336
−2	0.3125	0.2604	0.2604	0.0833	0.0833
−1.5	0.3125	0.1701	0.1701	0.1736	0.1736
−1.23	0.3297	0.0325	0.0325	0.3027	0.3027

It is worthwhile to consider another parameter constellation such that $\beta_1 \neq 0$ and $\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} > 0\ \forall \boldsymbol{x} \in \mathcal{X} = [1,2]^3$. Here it is not necessary that $\beta_2 = \beta_3$. Denote $\gamma_1 = \frac{\beta_2}{\beta_1}$ and $\gamma_2 = \frac{\beta_3}{\beta_1}$ such that $x_1 + \gamma_1 x_2 + \gamma_2 x_3 > 0$, i.e., $\gamma_2 > -(\frac{x_1}{x_3} + \gamma_1 \frac{x_2}{x_3})\ \forall \boldsymbol{x} \in \mathcal{X}$. Let $\gamma_1 = 1$ thus $\gamma_2 > -1$. Also by employing the multiplicative algorithm locally D-optimal designs were numerically computed on $\mathcal{X}$. Table 4.4 shows some numerical results at particular parameter points and it turns out that ξ^* has different performances as follows.

(i) If $-1 < \gamma_2 < -0.833$ then ξ^* is saturated and supported by $\boldsymbol{v}_4$, $\boldsymbol{v}_5$, $\boldsymbol{v}_6$.

(ii) If $-0.833 \leq \gamma_2 < -0.335$ then ξ^* is supported by $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_4$, $\boldsymbol{v}_5$, $\boldsymbol{v}_6$.

(iii) If $-0.335 \leq \gamma_2 < 5$ then ξ^* is saturated and supported by $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_4$.

(iv) If $\gamma_2 \geq 5$ then ξ^* is supported by $\boldsymbol{v}_2$, $\boldsymbol{v}_3$, $\boldsymbol{v}_4$, $\boldsymbol{v}_7$.

TABLE 4.4: D-optimal designs on $\mathcal{X} = [1,2]^3$ at $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ where $\gamma_2 = \beta_3/\beta_1, \gamma_1 = \beta_2/\beta_1 = 1$.

γ_2	$\boldsymbol{v}_1$	$\boldsymbol{v}_2$	$\boldsymbol{v}_3$	$\boldsymbol{v}_4$	$\boldsymbol{v}_5$	$\boldsymbol{v}_6$	$\boldsymbol{v}_7$	$\boldsymbol{v}_8$
−0.9	0	0	0	0.3333	0.3333	0.3333	0	0
−0.5	0	0.2604	0.2604	0.3126	0.0833	0.0833	0	0
0	0	0.3333	0.3333	0.3333	0	0	0	0
1	0	0.3333	0.3333	0.3333	0	0	0	0
100	0	0.2840	0.2840	0.3143	0	0	0.1175	0

In general, for gamma models without intercept, finding optimal designs for a model with multiple factors, i.e., $\nu \geq 4$ is not an easy task. The optimal design given by part (i) of Theorem 4.4.4 might be extended for arbitrary number of factors under a sufficient and necessarily condition on the parameter points:

Theorem 4.4.5. *Consider the experimental region $\mathcal{X} = [a,b]^\nu, \nu \geq 3$, $0 < a < b$. Let $\boldsymbol{\beta}$ be a parameter point such that $\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$. Define $T(\boldsymbol{x}) = \sum_{i=1}^{\nu} x_i$, $q = \frac{a}{(\nu-1)a+b}$ and $c_j = (b-a)\beta_j + a\sum_{i=1}^{\nu}\beta_i$ $(1 \leq j \leq \nu)$. Then the design ξ^* which assigns equal weights ν^{-1} to the support*

$$\boldsymbol{x}_1^* = \Big(b, a, \ldots, a\Big)^\top, \boldsymbol{x}_2^* = \Big(a, b, \ldots, a\Big)^\top, \ldots, \boldsymbol{x}_\nu^* = \Big(a, a, \ldots, b\Big)^\top$$

is locally D-optimal (at $\boldsymbol{\beta}$) if and only if for all $\boldsymbol{x} = (x_1, \ldots, x_\nu)^\top \in \{a,b\}^\nu$

$$\sum_{j=1}^{\nu}\Big(x_j - qT(\boldsymbol{x})\Big)^2 c_j^2 \leq (b-a)^2\Big(\sum_{j=1}^{v}\beta_j x_j\Big)^2. \tag{4.43}$$

Proof. Define the $\nu \times \nu$ design matrix $\boldsymbol{F} = [\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_\nu^*)]^\top$ which is thus presented as $\boldsymbol{F} = (b-a)\boldsymbol{I}_\nu + a\mathbf{1}\mathbf{1}^\top$ with $\boldsymbol{F}^{-1} = \frac{1}{(b-a)}\Big(\boldsymbol{I}_\nu - q\mathbf{1}\mathbf{1}^\top\Big)$ where $\boldsymbol{I}_\nu$ is the $\nu \times \nu$ identity matrix and $\mathbf{1}$ is a $\nu \times 1$ vector of ones. The information matrix of ξ^* is given by $\boldsymbol{M}\Big(\xi^*, \boldsymbol{\beta}\Big) = \frac{1}{\nu}\boldsymbol{F}^\top\boldsymbol{V}\boldsymbol{F}$ where $\boldsymbol{V} = \text{diag}\Big(u(\boldsymbol{x}_j^*, \boldsymbol{\beta})\Big)_{j=1}^{\nu}$ is the $\nu \times \nu$ weight matrix. Note that $u(\boldsymbol{x}_j^*, \boldsymbol{\beta}) = c_j^{-2}$ for all $(1 \leq j \leq \nu)$. Hence, the l.h.s. of the condition (2.11) of the Equivalence Theorem (Theorem 2.2.2) is equal to

$$\begin{aligned}
&\Big(\sum_{j=1}^{v}\beta_j x_j\Big)^{-2}\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{M}^{-1}\Big(\xi^*, \boldsymbol{\beta}\Big)\boldsymbol{f}(\boldsymbol{x}) = \nu\Big(\sum_{j=1}^{v}\beta_j x_j\Big)^{-2}\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{F}^{-1}\boldsymbol{V}^{-1}\boldsymbol{F}^{-1}\boldsymbol{f}(\boldsymbol{x}) \\
&= \nu\Big((b-a)\sum_{j=1}^{v}\beta_j x_j\Big)^{-2}\Big(\boldsymbol{f}^\top(\boldsymbol{x}) - qT(\boldsymbol{x})\mathbf{1}^\top\Big)\text{diag}\Big(c_j^2\Big)_{j=1}^{\nu}(\boldsymbol{f}(\boldsymbol{x}) - qT(\boldsymbol{x})\mathbf{1}) \\
&= \nu\Big((b-a)\sum_{j=1}^{v}\beta_j x_j\Big)^{-2}\sum_{j=1}^{\nu}\Big(x_j - qT(\boldsymbol{x})\Big)^2 c_j^2.
\end{aligned} \tag{4.44}$$

By The Equivalence Theorem design ξ^* is locally D-optimal if and only if (4.44) is less than or equal to ν for all $\boldsymbol{x} \in \{a, b\}^\nu$ leading to resulting inequalities that are equivalent to assumption (4.43). □

Note that the D-optimal design given in part (i) of Theorem 4.4.4 is a special case of Theorem 4.4.5 when $\nu = 3$ where condition (4.43) is equivalent to condition (4.37) in the proof of part (i) of Theorem 4.4.4. Actually it can be seen that already in the general case of Theorem 4.4.5 the optimality condition (4.43) depends only on the ratios $\beta_j/(\sum_{i=1}^{\nu}\beta_i)$ for all $(1 \leq j \leq \nu)$. Similarly note that condition (4.43) depends on a and b only through their ratio a/b. However, assuming the model parameters are having equal size implies that the D-optimality of a design is independent of the model parameters but it depends on the ratio a/b as it is shown in the next corollary.

Corollary 4.4.6. *Consider the experimental region $\mathcal{X} = [a, b]^\nu, \nu \geq 3,\ 0 < a < b$. Let $\boldsymbol{\beta}$ be a parameter point such that $\beta_j = \beta_{j'} = \beta > 0$ $(1 \leq j < j' \leq \nu)$. Then the design ξ^* which assigns equal weights ν^{-1} to the support $\boldsymbol{x}_1^* = \big(b, a, \ldots, a\big)^\top$, $\boldsymbol{x}_2^* = \big(a, b, \ldots, a\big)^\top$, $\ldots$, $\boldsymbol{x}_\nu^* = \big(a, a, \ldots, b\big)^\top$ is locally D-optimal (at $\boldsymbol{\beta}$) if and only if*

$$\left(\frac{b}{a}\right)^2 \geq \frac{\big(\nu - 1\big)\big(\nu - 2\big)}{2}. \tag{4.45}$$

Proof. Let $\beta_j = \beta_{j'} = \beta$ $(1 \leq j < j' \leq \nu)$ then condition (4.43) of Theorem 4.4.5 reduces to

$$\Big((\nu - 1)a^2 + b^2\Big)\left(\sum_{j=1}^{\nu} x_j\right)^2 - ((\nu - 1)a + b)^2 \sum_{j=1}^{\nu} x_j^2 \geq 0 \quad \forall \boldsymbol{x} \in \{a, b\}^\nu. \tag{4.46}$$

For $\boldsymbol{x} = (x_1, \ldots, x_\nu) \in \{a, b\}^\nu$ let $r = r(\boldsymbol{x}) \in \{0, 1, \ldots, \nu\}$ denote the number of coordinates of $\boldsymbol{x}$ that are equal to b. Then we have $\sum_{j=1}^{\nu} x_j^2 = (\nu - r)\, a^2 + r\, b^2$ and $\big(\sum_{j=1}^{\nu} x_j\big)^2 = ((\nu - r)\, a + r\, b)^2$. Hence, condition (4.46) is equivalent to

$$(a - b)^2\, \tau\, r^2 + (a - b)((b + a) - 2\, a\, \nu\, \tau)\, r + \nu\, a^2(\nu\, \tau - 1) \geq 0\ \forall r \in \{0, 1, \ldots, \nu\},\ \nu \geq 2 \tag{4.47}$$

where $\tau = \frac{(\nu-1)a^2+b^2}{((\nu-1)a+b)^2}$. The l.h.s. of inequality (4.47) is a polynomial in r of degree 2 with positive leading term. The polynomial attains 0 at $r = 1$ ($r_1 = 1$ indicates the support of ξ^*) and at $r_2 = \frac{\nu\,(\nu-1)\, a^2}{(\nu-1)a^2+b^2}$. Note that the polynomial is positive and increasing for all $r > 2$ (i.e., (4.47) holds true) when $r_2 \leq 2$ or, equivalently, $\frac{\nu\,(\nu-1)\, a^2}{(\nu-1)a^2+b^2} \leq 2$ which coincides with condition (4.45). □

Remark 4.4.8. Actually, condition (4.45) is obviously fulfilled for $\nu = 2$ (compare Theorem 4.4.3). For the case $\nu = 3$ the bound of l.h.s. of condition (4.45) is 1 and, hence always fulfilled.

4.5 Model with interaction

In this section we deal with gamma models with arbitrary number of quantitative factors with interactions. In Subsection 4.5.1 we consider full interactions models with intercept. In Subsection 4.5.2 we concentrate on a three-factor model without intercept. The gamma models of complete product-type interactions are employed in Subsection 4.5.3.

Here, $\boldsymbol{f}$ is affine-multilinear as defined in Section 4.2 and the experimental region $\mathcal{X}$ is a ν-dimensional hyperrectangle from (4.7). So by Theorem 4.2.1 we will look for the optimal designs in the set of vertices of $\mathcal{X}$.

4.5.1 Model with intercept

In this situation, we briefly discuss locally D- and A-optimal designs for gamma models with intercept where $\boldsymbol{f}(\boldsymbol{x})$ collects interaction terms as given by (3.21) with $\mathcal{X} = [a, b]^\nu$.

Following the results in Section 3.4, the locally D-optimal design for a full interaction gamma model is implicitly obtained from Theorem 3.4.1 which is thus supported by all vertices $\{a, b\}^\nu$ with equal weights. Moreover, from Theorem 3.4.2 we introduce a locally A-optimal design in the next corollary.

Corollary 4.5.1. *Consider the full interaction gamma model on the unit square $\mathcal{X} = [0, 1]^2$, i.e., $\boldsymbol{f}(\boldsymbol{x}) = \big(1, x_1, x_2, x_1x_2\big)^\top$ for all $\boldsymbol{x} = (x_1, x_2)^\top \in [0, 1]^2$. Denote the vertices of $\mathcal{X}$ by $\boldsymbol{v}_1 = (0, 0)^\top$, $\boldsymbol{v}_2 = (1, 0)^\top$, $\boldsymbol{v}_3 = (0, 1)^\top$, and $\boldsymbol{v}_4 = (1, 1)^\top$. For any given parameter point $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ with (4.4) the unique locally A-optimal design (at $\boldsymbol{\beta}$) is supported by the vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ with weights*

$$\omega_1^* = 2\beta_0/c, \quad \omega_2^* = \sqrt{2}(\beta_0 + \beta_1)/c, \quad \omega_3^* = \sqrt{2}(\beta_0 + \beta_2)/c, \; \omega_4^* = (\beta_0 + \beta_1 + \beta_2 + \beta_3)/c,$$

where $c = (3 + 2\sqrt{2})\beta_0 + (\sqrt{2} + 1)(\beta_1 + \beta_2) + \beta_3$.

4.5.2 Model without intercept

In this subsection we consider a model without intercept, in particular, a two-factor model with interaction where $\boldsymbol{f}(\boldsymbol{x}) = \big(x_1, x_2, x_1x_2\big)^\top$ and $\boldsymbol{\beta} = \big(\beta_1, \beta_2, \beta_3\big)^\top$. The experimental region is given by $\mathcal{X} = [a, b]^2$, $0 < a < b$ and we aim at deriving a locally D-optimal design. Our approach is employing a transformation of the proposed model to a model with intercept by removing the interaction term x_1x_2. As it was pointed out in Remark 4.4.5 the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ is invariant w.r.t. simultaneously scaling of $\boldsymbol{x}$, i.e., $\boldsymbol{f}_{\boldsymbol{\beta}}(\lambda\boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$. Let $\lambda = 1/(x_1x_2)$ then we obtain

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \Big(\beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2\Big)^{-1}\begin{pmatrix} x_1 \\ x_2 \\ x_1x_2 \end{pmatrix} \tag{4.48}$$

$$= \Big(\beta_1t_2 + \beta_2t_1 + \beta_3\Big)^{-1}\begin{pmatrix} t_2 \\ t_1 \\ 1 \end{pmatrix} = \boldsymbol{f}^{\circ}_{\boldsymbol{\beta}}(\boldsymbol{t}) \tag{4.49}$$

where $\boldsymbol{t} = \Big(t_1, t_2\Big)^{\top}, t_j = 1/x_j,\ j = 1, 2$. The range of $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$, as $\boldsymbol{x}$ ranges over $\mathcal{X} = [a, b]^2$ is a cube given by $\mathcal{T} = \Big[(1/b)\,,\,(1/a)\Big]^2$ with $\boldsymbol{f}^{\circ}(\boldsymbol{t}) = (t_2, t_1, 1)^{\top}$. One can rearrange the terms of (4.49) by making use of the 3×3 anti-diagonal transformation matrix $\boldsymbol{Q}$. So we have $\tilde{\boldsymbol{f}}(\boldsymbol{t}) = \boldsymbol{Q}\boldsymbol{f}^{\circ}(\boldsymbol{t}) = (1, t_1, t_2)^{\top}$ and $\tilde{\boldsymbol{\beta}} = \Big(\boldsymbol{Q}^{\top}\Big)^{-1}\boldsymbol{\beta} = \Big(\beta_3, \beta_2, \beta_1\Big)^{\top}$. Hence, $\tilde{\boldsymbol{f}}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{t}) = \Big(\tilde{\boldsymbol{f}}^{\top}(\boldsymbol{t})\tilde{\boldsymbol{\beta}}\Big)^{-1}\tilde{\boldsymbol{f}}(\boldsymbol{t})$ and rewrites as

$$\tilde{\boldsymbol{f}}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{t}) = \Big(\beta_3 + \beta_2t_1 + \beta_1t_2\Big)^{-1}\begin{pmatrix} 1 \\ t_1 \\ t_2 \end{pmatrix},\ \boldsymbol{t} \in \Big[(1/b)\,,\,(1/a)\Big]^2. \tag{4.50}$$

Since (4.50) coincides with that for a gamma model with intercept the D-criterion is equivariant (see Radloff and Schwabe (2016)) with respect to a one-to-one transformation from $\mathcal{T} = \Big[(1/b)\,,\,(1/a)\Big]^2$ to $\mathcal{Z} = [0, 1]^2$ where

$$t_j \to z_j = \frac{1}{(1/a) - (1/b)}t_j - \frac{1/b}{(1/a) - (1/b)},\ j = 1, 2. \tag{4.51}$$

For a given transformation matrix

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{-(1/b)}{(1/a)-(1/b)} & \frac{1}{(1/a)-(1/b)} & 0 \\ \frac{-(1/b)}{(1/a)-(1/b)} & 0 & \frac{1}{(1/a)-(1/b)} \end{pmatrix} \quad \text{with} \quad \boldsymbol{B}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{b} & \frac{1}{a} - \frac{1}{b} & 0 \\ \frac{1}{b} & 0 & \frac{1}{a} - \frac{1}{b} \end{pmatrix}$$

we have $\tilde{\boldsymbol{f}}(\boldsymbol{z}) = \boldsymbol{B}\tilde{\boldsymbol{f}}(\boldsymbol{t}) = (1, z_1, z_2)^{\top}$ with $\tilde{\tilde{\boldsymbol{\beta}}} = \Big(\boldsymbol{B}^{\top}\Big)^{-1}\tilde{\boldsymbol{\beta}} = (\tilde{\tilde{\beta}}_0,\ \tilde{\tilde{\beta}}_1,\ \tilde{\tilde{\beta}}_2)^{\top}$ where $\tilde{\tilde{\beta}}_0 = \beta_3 + (1/b)(\beta_1 + \beta_2)$, $\tilde{\tilde{\beta}}_1 = \beta_2((1/a) - (1/b))$ and $\tilde{\tilde{\beta}}_2 = \beta_1((1/a) - (1/b))$. It follows that $\tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z}) = \Big(\tilde{\boldsymbol{f}}^{\top}(\boldsymbol{z})\tilde{\tilde{\boldsymbol{\beta}}}\Big)^{-1}\tilde{\boldsymbol{f}}(\boldsymbol{z})$ which rewrites as

$$\tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z}) = \left(\tilde{\tilde{\beta}}_0 + \tilde{\tilde{\beta}}_1 z_1 + \tilde{\tilde{\beta}}_2 z_2\right)^{-1} \begin{pmatrix} 1 \\ z_1 \\ z_2 \end{pmatrix}, \boldsymbol{z} \in [0,1]^2. \tag{4.52}$$

Let $\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}_{\boldsymbol{\beta}}^\top(\boldsymbol{x})$, $\tilde{\boldsymbol{M}}(\boldsymbol{t},\tilde{\boldsymbol{\beta}}) = \tilde{\boldsymbol{f}}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{t})\tilde{\boldsymbol{f}}_{\tilde{\boldsymbol{\beta}}}^\top(\boldsymbol{t})$ and $\tilde{\tilde{\boldsymbol{M}}}(\boldsymbol{z},\tilde{\tilde{\boldsymbol{\beta}}}) = \tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z})\tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}^\top(\boldsymbol{z})$ be the information matrices for the models which corresponding to (4.48), (4.50) and (4.52), respectively. It is easy to observe that

$$\boldsymbol{M}(\boldsymbol{x},\boldsymbol{\beta}) = \boldsymbol{Q}^{-1}\tilde{\boldsymbol{M}}(\boldsymbol{t},\tilde{\boldsymbol{\beta}})\boldsymbol{Q}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{Q}^{-1}\tilde{\tilde{\boldsymbol{M}}}(\boldsymbol{z},\tilde{\tilde{\boldsymbol{\beta}}})\boldsymbol{Q}^{-1}\boldsymbol{B}^{-1},$$

thus the derived D-optimal designs on $\mathcal{X}$, $\mathcal{T}$ and $\mathcal{Z}$, respectively are equivariant. According to the mapping of $\boldsymbol{x}$ to $\boldsymbol{t}$ in the line following (4.49) and the mapping from $\boldsymbol{t}$ to $\boldsymbol{z}$ in (4.51) each component is mapped separately: $x_j \to t_j \to z_j$ without permuting them. Therefore, one modifies the direct one-to-one transformation $g : \mathcal{X} \to \mathcal{Z}$ where

$$x_j \to z_j = \frac{1/x_j}{(1/a)-(1/b)} - \frac{1/b}{(1/a)-(1/b)}, j = 1,2. \tag{4.53}$$

Let ξ_g^* be a design defined on $\mathcal{Z}$ that assigns the weights $\xi(\boldsymbol{x})$ to the mapped support points $g(\boldsymbol{x})$, $\boldsymbol{x} \in \text{supp}(\xi^*)$. In fact, ξ^* on $\mathcal{X}$ is locally D-optimal (at $\boldsymbol{\beta}$) if and only if ξ_g^* on $\mathcal{Z}$ is locally D-optimal (at $\tilde{\tilde{\boldsymbol{\beta}}}$). Note that the function $\tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z})$ is injective and thus the optimal design on $\mathcal{Z}$ is only supported by the vertices (cp. Theorem 4.2.1). It is worth noting by transformation (4.53) we obtain

$$(b,b)^\top \to (0,0)^\top,\ (b,a)^\top \to (1,0)^\top,$$
$$(a,b)^\top \to (0,1)^\top,\ (a,a)^\top \to (1,1)^\top.$$

Corollary 4.5.2. *Consider* $\boldsymbol{f}(\boldsymbol{x}) = \left(x_1, x_2, x_1x_2\right)^\top$ *on* $\mathcal{X} = [a,b]^2$, $0 < a < b$. *Denote the vertices by* $\boldsymbol{v}_1 = (b,b)^\top$, $\boldsymbol{v}_2 = (b,a)^\top$, $\boldsymbol{v}_3 = (a,b)^\top$, $\boldsymbol{v}_4 = (a,a)^\top$. *Let* $\boldsymbol{\beta} = (\beta_1,\beta_2,\beta_3)^\top$ *be a parameter point satisfying (4.4). Then the unique locally D-optimal design* ξ^* *(at* $\boldsymbol{\beta}$*) is as follows.*

(i) If $\beta_3^2 + \frac{1}{b^2}(\beta_1^2+\beta_2^2) + (\frac{1}{b^2} - \frac{1}{a^2} + \frac{2}{a\,b})\beta_1\beta_2 + \frac{2}{b}\beta_3(\beta_1+\beta_2) \le 0$ *then* ξ^* *assigns equal weights* $1/3$ *to* $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

(ii) If $\beta_3^2 + \frac{1}{b^2}\beta_1^2 + \frac{1}{a^2}\beta_2^2 + \frac{2}{b}\beta_3\beta_1 + \frac{2}{a}\beta_3\beta_2 + (\frac{1}{b^2} + \frac{1}{a^2})\beta_1\beta_2 \le 0$ *then* ξ^* *assigns equal weights* $1/3$ *to* $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4$.

(iii) If $\beta_3^2 + \frac{1}{b^2}\beta_2^2 + \frac{1}{a^2}\beta_1^2 + \frac{2}{b}\beta_3\beta_2 + \frac{2}{a}\beta_3\beta_1 + (\frac{1}{b^2} + \frac{1}{a^2})\beta_1\beta_2 \le 0$ *then* ξ^* *assigns equal weights* $1/3$ *to* $\boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4$.

(iv) If $\beta_3^2 + \frac{1}{a^2}(\beta_1^2+\beta_2^2) + (\frac{1}{a^2} - \frac{1}{b^2} + \frac{2}{ab})\beta_1\beta_2 + \frac{2}{a}\beta_3(\beta_1+\beta_2) \le 0$ *then* ξ^* *assigns equal weights* $1/3$ *to* $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$.

(v) If none of the cases (i) – (iv) applies then ξ^* *is supported by the four vertices*

$$\xi^* = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \omega_1^* & \omega_2^* & \omega_3^* & \omega_4^* \end{pmatrix}, \quad \text{where } \omega_\ell^* > 0\ (1 \le \ell \le 4),\ \textstyle\sum_{\ell=1}^4 \omega_\ell^* = 1.$$

Proof. The proof is obtained by verifying the D-optimality of ξ_g^* on $\mathcal{Z} = [0,1]^2$ under transformation g given by (4.53). The regression vector $\tilde{\boldsymbol{f}}_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z})$ given by (4.52) coincides with that for the two-factor gamma model with intercept on $\mathcal{Z} = [0,1]^2$ whose intensity function is defined as $u_{\tilde{\tilde{\boldsymbol{\beta}}}}(\boldsymbol{z}) = (\tilde{\tilde{\beta}}_0 + \tilde{\tilde{\beta}}_1 z_1 + \tilde{\tilde{\beta}}_2 z_2)^{-2}$ for all $\boldsymbol{z} \in \mathcal{Z}$. Denote

$$\begin{aligned}
c_1 &= u_{\tilde{\tilde{\boldsymbol{\beta}}}}((0,0)^\top) = \tilde{\tilde{\beta}}_0^{-2} = (\beta_3 + \frac{1}{b}(\beta_1+\beta_2))^{-2},\\
c_2 &= u_{\tilde{\tilde{\boldsymbol{\beta}}}}((1,0)^\top) = (\tilde{\tilde{\beta}}_0 + \tilde{\tilde{\beta}}_1)^{-2} = (\beta_3 + \beta_1\frac{1}{b} + \beta_2\frac{1}{a})^{-2},\\
c_3 &= u_{\tilde{\tilde{\boldsymbol{\beta}}}}((0,1)^\top) = (\tilde{\tilde{\beta}}_0 + \tilde{\tilde{\beta}}_2)^{-2} = (\beta_3 + \beta_1\frac{1}{a} + \beta_2\frac{1}{b})^{-2},\\
c_4 &= u_{\tilde{\tilde{\boldsymbol{\beta}}}}((1,1)^\top) = (\tilde{\tilde{\beta}}_0 + \tilde{\tilde{\beta}}_1 + \tilde{\tilde{\beta}}_2)^{-2} = (\beta_3 + \frac{1}{a}(\beta_1+\beta_2))^{-2}.
\end{aligned}$$

Let $h, i, j, k \in \{1,2,3,4\}$ are pairwise distinct such that $c_k = \min\{c_1, c_2, c_3, c_4\}$ then it follows from Theorem 3.3.1 that if $c_k^{-1} \ge c_h^{-1} + c_i^{-1} + c_j^{-1}$ then ξ^* is a three-point design supported by the three vertices $\boldsymbol{v}_h$, $\boldsymbol{v}_i$, $\boldsymbol{v}_j$, with equal weights $1/3$. Hence, straightforward computations show that the condition in case (i) of the corollary is equivalent to $c_4^{-1} \ge c_1^{-1} + c_2^{-1} + c_3^{-1}$. Analogous verifying is obtained for the cases (ii), (iii), (iv). For case (v) the four-point design according to Theorem 3.3.1 is locally D-optimal if $c_k^{-1} < c_h^{-1} + c_i^{-1} + c_j^{-1}$ which applies implicitly if non of the conditions (i) – (iv) of saturated designs is fulfilled by a given $\boldsymbol{\beta}$ (cp. Remark 2.2.4). □

It is noted that the optimality conditions provided in parts (i)–(iv) of Corollary 4.5.2 depend on the values of a and b. Changing these values might affect the D-optimality of a design. To see that, more specifically, let $a = 1$ and $b = 2$, i.e., the experimental region is $\mathcal{X} = [1,2]^2$ and define $\gamma_1 = \beta_1/\beta_3$ and $\gamma_2 = \beta_2/\beta_3$, $\beta_3 \ne 0$. Here, the parameter space which is depicted in Panel (a) of Figure 4.7 is characterized by $\gamma_2 + \gamma_1 > -1$, $2\gamma_2 + \gamma_1 > -2$ and $\gamma_2 + 2\gamma_1 > -2$. It is observed from Panel (a) of Figure 4.7 that the design given by part (i) of Corollary 4.5.2 is not locally D-optimal because the corresponding optimality condition in part (i) of the corollary; $\frac{1}{4}(\gamma_1^2 + \gamma_2^2) + \frac{1}{4}\gamma_1\gamma_2 + \gamma_1 + \gamma_2 \le -1$ can not be satisfied.

Let us consider another experimental region with a higher length by fixing $a = 1$ and taking $b = 4$, i.e., the experimental region is $\mathcal{X} = [1,4]^2$. The parameter space which is depicted in Panel (b) of Figure 4.7 is characterized by $\gamma_2+\gamma_1 > -1$, $4\,\gamma_2 + \gamma_1 > -4$ and $\gamma_2 + 4\,\gamma_1 > -4$. In this case all designs given by Corollary 4.5.2 are locally D-optimal at particular values of γ_2 and γ_1 as it is observed from the figure. It is obvious that along the diagonal dashed line, $\gamma_2 = \gamma_1$, there exist at most three different types of locally D-optimal designs.

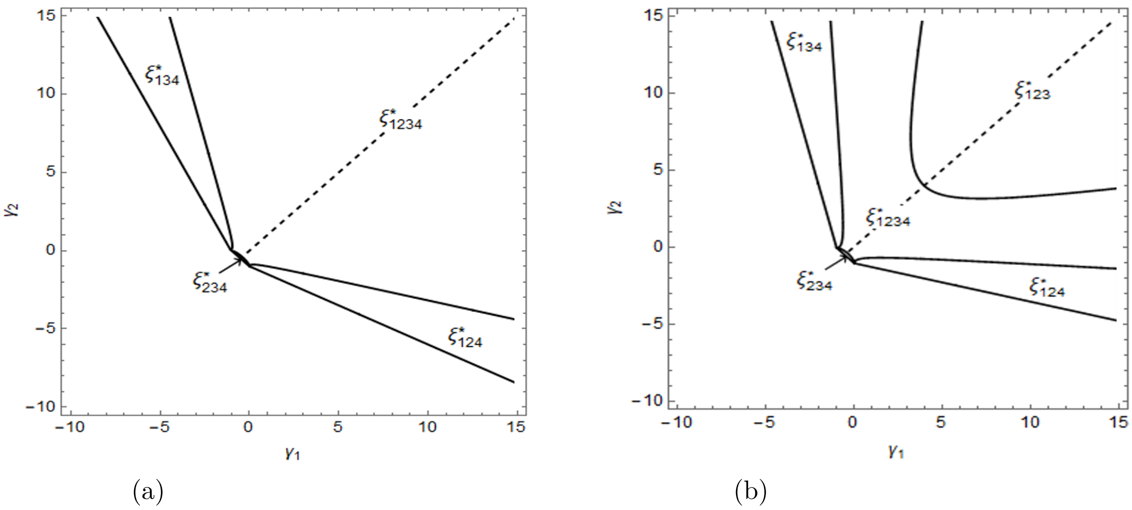


FIGURE 4.7: Dependence of locally D-optimal designs on $\gamma_1 = \beta_1/\beta_3$ and $\gamma_2 = \beta_2/\beta_3$ where for Panel (a) $\mathcal{X} = [1,2]^2$ and for Panel (b) $\mathcal{X} = [1,4]^2$. The diagonal dashed line represents the case $\gamma_2 = \gamma_1$. Note that $\mathrm{supp}(\xi^*_{ijk}) = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k\} \subset \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and $\mathrm{supp}(\xi^*_{1234}) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$.

.

For arbitrary values of a and b, $0 < a < b$ let us restrict to case $\gamma_2 = \gamma_1 = \gamma$, i.e., $\beta_1 = \beta_2 = \beta$, $\beta_3 \neq 0$ and the next corollary is immediate.

Corollary 4.5.3. *Consider $\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2, x_1x_2)^\top$ on an arbitrary square $\mathcal{X} = [a,b]^2$, $0 < a < b$ in the positive quadrant. Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ be a parameter point satisfying (4.4) with $\beta_1 = \beta_2 = \beta$ and $\beta_3 \neq 0$. Define $\gamma = \frac{\beta}{\beta_3}$. Then the locally D-optimal design ξ^* (at $\boldsymbol{\beta}$) is as follows.*

(i) If $-\frac{a}{2} < \gamma \leq -\frac{ab}{3b-a}$, then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_2, \boldsymbol{v}_3$, $\boldsymbol{v}_4$.

(ii) If $b - 3a > 0$ and $\gamma \geq \frac{ab}{b-3a}$, then ξ^* assigns equal weights $1/3$ to $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

(iii) If $b-3a > 0$ and $-\frac{ab}{3b-a} < \gamma < \frac{ab}{b-3a}$ then the design ξ^* is supported by $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ with optimal weights given by

$$\omega_1^* = \frac{ab - (a-3b)\gamma}{4b(a+2\gamma)}, \quad \omega_2^* = \omega_3^* = \frac{\big(ab + (a+b)\gamma\big)^2}{4ab(b+2\gamma)(a+2\gamma)}, \quad \omega_4^* = \frac{ab - (b-3a)\gamma}{4a(b+2\gamma)}.$$

Proof. Consider the experimental region $\mathcal{X} = [a, b]^2$, $0 < a < b$. By assumption $\beta_1 = \beta_2 = \beta$, $\beta_3 \neq 0$ the range of $\gamma = \frac{\beta}{\beta_3}$ is given by $(-a/2, \infty)$. Assumption $b - 3a > 0$ implies that $-\frac{a}{2} < -\frac{ab}{3b-a} < \frac{ab}{b-3a}$. According to Corollary 4.5.2 we show the following under the assumptions of Corollary 4.5.3. Both conditions provided in parts (ii) and (iii) of Corollary 4.5.2 are not fulfilled by any parameter point thus the corresponding designs are not D-optimal. In contrast, the design ξ^* given in (i) of Corollary 4.5.3 is locally D-optimal if the condition provided in part (iv) of Corollary 4.5.2 holds true. That condition is equivalent to

$$(3b^2 + 2ab - a^2)\gamma^2 + 4ab^2\gamma + a^2b^2 \leq 0.$$

The l.h.s. of above inequality is polynomial in γ of degree 2 and thus the inequality is fulfilled by $-\frac{a}{2} < \gamma \leq -\frac{ab}{3b-a}$.
Similarly, the design ξ^* in (ii) of Corollary 4.5.3 is locally D-optimal if the condition provided in part (i) of Corollary 4.5.2 holds true. That condition is equivalent to

$$(3a^2 + 2ab - b^2)\gamma^2 + 4a^2b\gamma + a^2b^2 \leq 0.$$

The l.h.s. of above inequality is polynomial in γ of degree 2 and thus the inequality is fulfilled by $\gamma \geq \frac{ab}{b-3a}$ if $b - 3a > 0$.
The four-point design given in (iii) has positive weights on $-\frac{ab}{3b-a} < \gamma < \frac{ab}{b-3a}$ if $b - 3a > 0$ and hence it is locally D-optimal in view of Remark 2.2.4. □

Remark 4.5.1. *One may note that from Corollary 4.5.3 when $\beta = 0$ the uniform design on the vertices $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ is locally D-optimal.*

4.5.3 Model of complete product-type interactions

Schwabe (1996b) developed an approach to construct optimal designs for linear models of complete product-type interactions by making use of optimal designs under marginal models. This approach is independent of the actual structure of the influence of the single factors and, hence, covers models with both qualitative and quantitative factors as well as purely qualitative or purely quantitative models (see Schwabe (1996b), p.35). In this subsection we extend that approach for gamma models of complete product-type interactions. We will show that locally optimal designs for gamma models of complete product-type interactions can be obtained from locally optimal designs under the marginal counterparts.

We consider K marginal models each is containing ν_k factors, $k = 1, \ldots, K$. The marginal ν_k-factor model is defined with a power link as in (4.2) where

$$\mu_k^{\rho}(\boldsymbol{x}_k) = \boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)}, \;\; \boldsymbol{x}_k = (x_{k1}, \ldots, x_{k\nu_k})^{\top} \in \mathcal{X}_k \subseteq \mathbb{R}^{\nu_k} \tag{4.54}$$

with exponent $\rho \in \mathbb{R}$, $\rho > 0$. So all marginal gamma models are having equal exponent ρ, or, equivalently, all link functions of the marginal models are the same. The positivity assumption (4.4) of the expected mean μ_k is to be satisfied, i.e., $\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)} > 0$ for all $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{k\nu_k})^{\top} \in \mathcal{X}_k$ where

$$\boldsymbol{f}^{(k)} : \mathcal{X}_k \to \mathbb{R}^{p_k}, \;\; \boldsymbol{\beta}^{(k)} = (\beta_1^{(k)}, \ldots, \beta_{p_k}^{(k)})^{\top} \in \mathbb{R}^{p_k} \;\; (1 \leq k \leq K).$$

For each k, the marginal model (4.54) has intensity function

$$u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)}) = \left(\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)}\right)^{-2}, \;\; \boldsymbol{x}_k \in \mathcal{X}_k \;\; (1 \leq k \leq K). \tag{4.55}$$

Let ν denotes the total number of factors in all marginal models, i.e., $\nu = \sum_{k=1}^{K} \nu_k$. The resulting ν-factor gamma model of complete product-type interactions with exponent ρ is thus defined by

$$\begin{aligned} \mu^{\rho}(\boldsymbol{x}) &= \boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta} \\ &= \sum_{i_1=1}^{p_1} \cdots \sum_{i_K=1}^{p_K} f_{i_1}^{(1)}(\boldsymbol{x}_1) \cdot \ldots \cdot f_{i_K}^{(K)}(\boldsymbol{x}_K)\beta_{i_1,\ldots,i_K} \end{aligned} \tag{4.56}$$

where $\boldsymbol{x} = (\boldsymbol{x}_1^{\top}, \ldots, \boldsymbol{x}_K^{\top})^{\top} \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$, in which $\boldsymbol{f}(\boldsymbol{x})$ collects all $K-$fold products of the components $f_i^{(k)}(\boldsymbol{x}_k)$ which belong to the regression functions $\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)$, $k = 1, \ldots, K$ and $\boldsymbol{x} = (x_{11}, \ldots, x_{1\nu_1}, \ldots, x_{K1}, \ldots, x_{K\nu_K})^{\top}$ is a ν-tuple. The unknown parameter $\beta_{i_1,\ldots,i_K}$ is equal to $\prod_{k=1}^{K} \beta_{i_k}^{(k)}$ and $\boldsymbol{\beta}$ is a p-dimensional parameter vector, i.e., $\boldsymbol{\beta} \in \mathbb{R}^p$ where $p = \prod_{k=1}^{K} p_k$. Note that $\boldsymbol{\beta}$ collects the parameters $\beta_{i_1,\ldots,i_K}$, $i_k = 1, \ldots, p_k$, $k = 1, \ldots, K$ and in lexicographic order $\boldsymbol{\beta}$ rewrites as

$$\boldsymbol{\beta} = (\beta_{1,\ldots,1,1}\,,\; \beta_{1,\ldots,1,2}\,,\; \ldots\,,\; \beta_{1,\ldots,1,p_K}\,,\; \beta_{1,\ldots,2,1}\,,\; \ldots\,,\; \beta_{p_1,\ldots,p_{K-1},p_K})^{\top}.$$

Note that $\boldsymbol{f}(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^p$ and can be described by Kronecker products "$\otimes$" as in the following;

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}^{(1)}(\boldsymbol{x}_1) \otimes \cdots \otimes \boldsymbol{f}^{(K)}(\boldsymbol{x}_K) = \bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k), \tag{4.57}$$

$$\text{with} \;\; \boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} \otimes \cdots \otimes \boldsymbol{\beta}^{(K)} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}. \tag{4.58}$$

Therefore, model (4.56) rewrites as

$$\mu^{\rho}(\boldsymbol{x}) = \left(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\right)^{\top} \left(\bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}\right) \tag{4.59}$$

and of course $\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$. The latter positivity assumption is obtained from that in the marginal models; that is because

$$\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta} = \left(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\right)^{\top} \left(\bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}\right) = \prod_{k=1}^{K} \boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)} > 0.$$

The intensity function $u(\boldsymbol{x}, \boldsymbol{\beta})$ for model (4.56) is determined by the product of the intensity functions (4.55) in the marginal ν_k-factor models (4.54).

Lemma 4.5.1. *The intensity function for model (4.56) is given by*

$$u(\boldsymbol{x}, \boldsymbol{\beta}) = \prod_{k=1}^{K} u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)}).$$

Proof. In general, the intensity function of gamma models with linear predictor $\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta}$ is defined by $u(\boldsymbol{x}, \boldsymbol{\beta}) = (\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta})^{-2}$. In view of (4.57) and (4.58) it follows that

$$\begin{aligned} u(\boldsymbol{x}, \boldsymbol{\beta}) &= \left(\left(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\right)^{\top} \left(\bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}\right)\right)^{-2} = \left(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)}\right)^{-2} \\ &= \bigotimes_{k=1}^{K} \left(\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{\beta}^{(k)}\right)^{-2} = \prod_{k=1}^{K} u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)}). \end{aligned}$$

□

Our aim is deriving an optimal design for model (4.56) as a product type design which is supported by the cross-product of the finite sets of design points of the designs under marginal ν_k-factor models and the weights are given by the product of the weights of those designs. To be more specific, denote by ξ_k a design defined on $\mathcal{X}_k$ for a marginal ν_k-factor model (4.54) $(1 \le k \le K)$. We introduce ξ_k as in (2.9);

$$\xi_k = \begin{pmatrix} \boldsymbol{x}_{k1} & \boldsymbol{x}_{k2} & \dots & \boldsymbol{x}_{kr_k} \\ \omega_{k1} & \omega_{k2} & \dots & \omega_{kr_k} \end{pmatrix}, \tag{4.60}$$

where ξ_k has r_k design points $\boldsymbol{x}_{kj}$ and corresponding weights $\omega_{kj}, j = 1, \dots, r_k$. Then the product type design $\xi = \bigotimes_{k=1}^{K} \xi_k$ is defined on $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_K$ and has $r = \prod_{k=1}^{K} r_k$ design points $\boldsymbol{x}_{i_1,\dots,i_K} = (x_{1i_1}, \dots, x_{Ki_K})^{\top}$ with corresponding weights $\omega_{i_1,\dots,i_K} = \prod_{k=1}^{K} \omega_{ki_k}, i_k = 1, \dots, r_k, k = 1, \dots, K$.

In order to reduce the problem of locally optimal designs at a given $\boldsymbol{\beta}$ for model (4.56) to the locally optimal designs for the marginal ν_k-factor models at a given $\boldsymbol{\beta}_k$ it is required to use a factorized information matrix. The information matrix and, hence, the variance-covariance matrix of a product type design ξ factorizes into its marginal counterparts as it is given by the next lemma.

Lemma 4.5.2. *The information matrix of a product type design* $\xi = \bigotimes_{k=1}^{K} \xi_k$ *for a gamma model of complete product-type interactions (4.56) is*

$$\boldsymbol{M}\big(\xi, \boldsymbol{\beta}\big) = \bigotimes_{k=1}^{K} \boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)$$

Proof. In general, the information matrix of ξ for the model (4.56) is given by $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \int_{\mathcal{X}} u(\boldsymbol{x}, \boldsymbol{\beta})\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}^\top(\boldsymbol{x})\,\xi(\mathrm{d}\boldsymbol{x})$. In view of Lemma 4.5.1 with (4.57) and (4.58) it follows that

$$\begin{aligned}
\boldsymbol{M}\big(\xi, \boldsymbol{\beta}\big) &= \int_{\mathcal{X}} u(\boldsymbol{x}, \boldsymbol{\beta})\Big(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)\Big(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)^\top \bigotimes_{k=1}^{K} \xi_k(\mathrm{d}\boldsymbol{x}_k)\\
&= \int_{\mathcal{X}_1}\cdots\int_{\mathcal{X}_K} \Big(\prod_{k=1}^{K} u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)})\Big)\Big(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\Big)\prod_{k=1}^{K} \xi_k(\mathrm{d}\boldsymbol{x}_k)\\
&= \bigotimes_{k=1}^{K}\Big(\int_{\mathcal{X}_k} u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\ \xi_k(\mathrm{d}\boldsymbol{x}_k)\Big)\\
&= \bigotimes_{k=1}^{K} \boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big).
\end{aligned}$$

□

As a consequence of Lemma 4.5.2 we get

$$\det\Big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\Big) = \det\Big(\bigotimes_{k=1}^{K} \boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)\Big) = \prod_{k=1}^{K} \det\Big(\boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)\Big)^{\prod_{j=1, j\neq k}^{K} p_j}.$$

For example; let $\xi = \xi_1 \otimes \xi_2 \otimes \xi_3$ then

$$\det(\boldsymbol{M}(\xi, \boldsymbol{\beta})) = \det(\boldsymbol{M}_1(\xi_1, \boldsymbol{\beta}^{(1)}))^{p_2 p_3} \det(\boldsymbol{M}_2(\xi_2, \boldsymbol{\beta}^{(2)}))^{p_1 p_3} \det(\boldsymbol{M}_3(\xi_3, \boldsymbol{\beta}^{(3)}))^{p_1 p_2}.$$

Therefore, for a given parameter point $\boldsymbol{\beta}$ that is evaluated from the parameter points of the marginal models, i.e., $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$ the best product type design $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ with respect to the D-criterion is generated from the locally D-optimal designs ξ_k^* at $\boldsymbol{\beta}^{(k)}$ for the marginal ν_k-factor models $(1 \leq k \leq K)$.

Theorem 4.5.1. *Let ξ_k^* be a locally D-optimal design (at $\boldsymbol{\beta}^{(k)}$) for a marginal ν_k-factor model (4.54) on the experimental region $\mathcal{X}_k$ $(1 \leq k \leq K)$. Then $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ is a* *locally D-optimal design (at* $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$*) for model (4.56) on the experimental region* $\mathcal{X} = \times_{k=1}^{K} \mathcal{X}_k$*.*

Proof. The proof is obtained by making use of condition (2.11) of The Equivalence Theorem (Theorem 2.2.2). To this end, denote $\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) = \sqrt{u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)})}\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)$ and $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \sqrt{u(\boldsymbol{x}, \boldsymbol{\beta})}\boldsymbol{f}(\boldsymbol{x})$. Since ξ_k^* is locally D-optimal (at $\boldsymbol{\beta}^{(k)}$) we guarantee that $\boldsymbol{f}^{(k)\top}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-1}(\xi_k^*, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) \leq p_k$ for all $\boldsymbol{x}_k \in \mathcal{X}_k$. Thus in view of Lemma 4.5.2 with (4.57) and (4.58) we obtain

$$\begin{aligned}
\boldsymbol{f}^{\top}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) &= u(\boldsymbol{x}, \boldsymbol{\beta})\Big(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)^{\top}\Big(\bigotimes_{k=1}^{K} \boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)\Big)^{-1}\Big(\bigotimes_{k=1}^{K} \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big) \\
&= \bigotimes_{k=1}^{K}\Big(u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-1}\big(\xi_k^*, \boldsymbol{\beta}^{(k)}\big)\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big) \\
&= \prod_{k=1}^{K} \boldsymbol{f}^{(k)\top}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-1}(\xi_k^*, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) \leq \prod_{k=1}^{K} p_k = p
\end{aligned}$$

for all $\boldsymbol{x} \in \mathcal{X}$. The Equivalence Theorem, thus, proves the local D-optimality of the product design $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ at a parameter point $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$. □

In the following we focus on the local A-optimality. From Lemma 4.5.2 we obtain the next straightforward factorization for every product type design $\xi = \bigotimes_{k=1}^{K} \xi_k$.

$$\mathrm{tr}\Big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\Big) = \mathrm{tr}\Bigg(\Big(\bigotimes_{k=1}^{K}\Big(\boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)\Big)\Big)^{-1}\Bigg) = \prod_{k=1}^{K} \mathrm{tr}\Big(\boldsymbol{M}_k^{-1}(\xi_k, \boldsymbol{\beta}^{(k)})\Big).$$

Hence, in analogy to the previous case of local D-optimality for a given parameter point $\boldsymbol{\beta}$ that is evaluated from the parameter points of the marginal models, i.e., $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$ the best product type design $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ with respect to the A-criterion is generated from the locally A-optimal designs ξ_k^* at $\boldsymbol{\beta}^{(k)}$ for the marginal ν_k-factor models $(1 \leq k \leq K)$.

Theorem 4.5.2. *Let* ξ_k^* *be a locally A-optimal design (at* $\boldsymbol{\beta}^{(k)}$*) for a marginal* ν_k*-factor model (4.54) on the experimental region* $\mathcal{X}_k$ $(1 \leq k \leq K)$*. Then* $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ *is a locally A-optimal design (at* $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$*) for model (4.56) on the experimental region* $\mathcal{X} = \times_{k=1}^{K} \mathcal{X}_k$*.*

Proof. The proof is obtained by making use of condition (2.12) of The Equivalence Theorem (Theorem 2.2.2). Since ξ_k^* is locally A-optimal (at $\boldsymbol{\beta}^{(k)}$) we guarantee that $\boldsymbol{f}^{(k)\top}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-2}(\xi_k^*, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) \leq \mathrm{tr}(\boldsymbol{M}_k^{-1}(\xi_k^*, \boldsymbol{\beta}^{(k)}))$ for all $\boldsymbol{x}_k \in \mathcal{X}_k$. Thus in view of

Lemma 4.5.2 with (4.57) and (4.58) we obtain

$$\begin{aligned}
&\boldsymbol{f}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{x})\boldsymbol{M}^{-2}(\xi^*,\boldsymbol{\beta})\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = u(\boldsymbol{x},\boldsymbol{\beta})\Big(\bigotimes_{k=1}^{K}\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)^{\top}\Big(\bigotimes_{k=1}^{K}\boldsymbol{M}_k\big(\xi_k,\boldsymbol{\beta}^{(k)}\big)\Big)^{-2}\Big(\bigotimes_{k=1}^{K}\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)\\
&= \bigotimes_{k=1}^{K}\Big(u_k(\boldsymbol{x}_k,\boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-2}\big(\xi_k^*,\boldsymbol{\beta}^{(k)}\big)\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)\\
&= \prod_{k=1}^{K}\boldsymbol{f}_{\boldsymbol{\beta}^{(k)}}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-2}(\xi_k^*,\boldsymbol{\beta}^{(k)})\boldsymbol{f}_{\boldsymbol{\beta}^{(k)}}^{(k)}(\boldsymbol{x}_k)\\
&\le \prod_{k=1}^{K}\mathrm{tr}(\boldsymbol{M}_k^{-1}(\xi_k^*,\boldsymbol{\beta}^{(k)})) = \mathrm{tr}\Big(\Big(\bigotimes_{k=1}^{K}\Big(\boldsymbol{M}_k\big(\xi_k^*,\boldsymbol{\beta}^{(k)}\big)\Big)^{-1}\Big) = \mathrm{tr}\Big(\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\Big)
\end{aligned}$$

for all $\boldsymbol{x}\in\mathcal{X}$. The Equivalence Theorem, thus, proves the local A-optimality of the product design $\xi^* = \bigotimes_{k=1}^{K}\xi_k^*$ at a parameter point $\boldsymbol{\beta} = \bigotimes_{k=1}^{K}\boldsymbol{\beta}^{(k)}$. □

Example 4.5.1. Marginal single-factor models.
Here we consider model (4.54) with one factor, $\nu_k = 1$, and two parameters, $p_k = 2$,. We restrict ourselves to the case of two marginal models $k = 2$ where

$$\mu_1^{\rho}(x_1) = \boldsymbol{f}^{(1)\top}(x_1)\boldsymbol{\beta}^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}x_1,\ \ x_1\in\mathcal{X}_1 = [0,1], \tag{4.61}$$

$$\mu_2^{\rho}(x_2) = \boldsymbol{f}^{(2)\top}(x_2)\boldsymbol{\beta}^{(2)} = \beta_0^{(2)} + \beta_1^{(2)}x_2,\ \ x_2\in\mathcal{X}_2 = [0,1] \tag{4.62}$$

such that $\beta_0^{(k)} > 0,\ \beta_0^{(k)} + \beta_1^{(k)} > 0, k = 1,2$. The resulting 2-factor gamma model of complete product-type interactions is thus written as

$$\mu^{\rho}(\boldsymbol{x}) = \boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2 \tag{4.63}$$

where $\boldsymbol{x} = (x_1,x_2)^{\top}\in\mathcal{X} = \mathcal{X}_1\times\mathcal{X}_2 = [0,1]^2, \boldsymbol{\beta}\in\mathbb{R}^4, \boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}\otimes\boldsymbol{\beta}^{(2)}$, hence

$$\boldsymbol{\beta} = \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = \begin{pmatrix}\beta_0^{(1)}\beta_0^{(2)}\\ \beta_0^{(2)}\beta_1^{(1)}\\ \beta_0^{(1)}\beta_1^{(2)}\\ \beta_1^{(1)}\beta_1^{(2)}\end{pmatrix}, \tag{4.64}$$

where $\boldsymbol{\beta}$ satisfies $\boldsymbol{f}^{\top}(\boldsymbol{x})\boldsymbol{\beta} > 0$ for all $\boldsymbol{x}\in\mathcal{X}$. Note that the 2-factor model with interaction (4.63) was considered in Subsection 4.5.1 where locally D- and A-optimal designs were derived. However, from Corollary 4.3.1 the marginal design ξ_k^* on $\mathcal{X}_k = [0,1]$ with support $\{0,1\}$ is equally weighted D-optimal (at $\boldsymbol{\beta}^{(k)}$). Thus the design $\xi^* = \xi_1^*\otimes\xi_2^*$ is locally D-optimal (at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}\otimes\boldsymbol{\beta}^{(2)}$) on the experimental region $\mathcal{X} = [0,1]^2$ for model (4.63). The design ξ^* assigns uniform weights $1/4$ to the support $(0,0)^{\top},(1,0)^{\top},(0,1)^{\top},(1,1)^{\top}$.

Moreover, from Corollary 4.3.1 the marginal design ξ_k^* on $\mathcal{X}_k = [0,1]$ with support $\{0,1\}$ is A-optimal (at $\boldsymbol{\beta}^{(k)}$) with weights $\omega_{k0}^* = \sqrt{2}\beta_0^{(k)}/((\sqrt{2}+1)\beta_0^{(k)} + \beta_1^{(k)})$ and $\omega_{k1}^* = (\beta_0^{(k)} + \beta_1^{(k)})/((\sqrt{2}+1)\beta_0^{(k)} + \beta_1^{(k)})$. For a given $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} \otimes \boldsymbol{\beta}^{(2)}$ from (4.64) the product type design $\xi^* = \xi_1^* \otimes \xi_2^*$ on the experimental region $\mathcal{X} = [0,1]^2$ for model (4.63) is locally A-optimal and assigns weights $\omega_{10}^*\omega_{20}^*$ to $(0,0)^\top$, $\omega_{11}^*\omega_{20}^*$ to $(1,0)^\top$, $\omega_{10}^*\omega_{21}^*$ to $(0,1)^\top$ and $\omega_{11}^*\omega_{21}^*$ to $(1,1)^\top$. Clearly, the product type design w.r.t. A-criterion coincides with that from Corollary 4.5.1 where for instance; the optimal weight $\omega_{10}^*\omega_{20}^*$ of the point $(0,0)^\top$ leads to the identity

$$\frac{2\beta_0^{(1)}\beta_0^{(2)}}{(3+2\sqrt{2})\beta_0^{(1)}\beta_0^{(2)} + (1+\sqrt{2})(\beta_0^{(1)}\beta_1^{(2)} + \beta_0^{(2)}\beta_1^{(1)}) + \beta_1^{(1)}\beta_1^{(2)}} = \frac{2\beta_0}{c}$$

where $c = (3+2\sqrt{2})\beta_0 + (1+\sqrt{2})(\beta_1+\beta_2) + \beta_3$.

Example 4.5.2. Marginal two-factor models.
Here we consider model (4.54) with two factors, $\nu_k = 2$, and three parameters, $p_k = 3$,. The case of two marginal models $k = 2$ is also adopted where

$$\mu_1^\rho(\boldsymbol{x}_1) = \boldsymbol{f}^{(1)\top}(\boldsymbol{x}_1)\boldsymbol{\beta}^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}x_{11} + \beta_2^{(1)}x_{12}, \ \boldsymbol{x}_1 = (x_{11}, x_{12})^\top \tag{4.65}$$

$$\mu_2^\rho(\boldsymbol{x}_2) = \boldsymbol{f}^{(2)\top}(\boldsymbol{x}_2)\boldsymbol{\beta}^{(2)} = \beta_0^{(2)} + \beta_1^{(2)}x_{21} + \beta_2^{(2)}x_{22}, \ \boldsymbol{x}_2 = (x_{21}, x_{22})^\top \tag{4.66}$$

where $\boldsymbol{x}_1 \in \mathcal{X}_k = [0,1]^2, k = 1,2$ such that $\beta_0^{(k)} > 0$, $\beta_0^{(k)} + \beta_1^{(k)} > 0$, $\beta_0^{(k)} + \beta_2^{(k)} > 0$, $\beta_0^{(k)} + \beta_1^{(k)} + \beta_2^{(k)} > 0, k = 1,2$. The resulting 2-factor gamma model of complete product-type interactions is thus written as

$$\begin{aligned}\mu^\rho(\boldsymbol{x}) = &\beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{21} + \beta_4 x_{22} + \beta_5 x_{11}x_{21} + \beta_6 x_{11}x_{22} \\ &+ \beta_7 x_{12}x_{21} + \beta_8 x_{12}x_{22}\end{aligned} \tag{4.67}$$

where $\boldsymbol{x} = (x_{11}, x_{12}, x_{21}, x_{22})^\top \in \mathcal{X} = [0,1]^4, \boldsymbol{\beta} \in \mathbb{R}^9, \boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} \otimes \boldsymbol{\beta}^{(2)}$. Note that $\beta_0 = \beta_0^{(1)}\beta_0^{(2)}$. From part (i) of Corollary 4.4.1 the design ξ_k^* on $\mathcal{X}_k = [0,1]^2$ with support $\{(0,0)^\top, (1,0)^\top, (0,1)^\top\}$ for the kth marginal model is equally weighted D-optimal (at $\boldsymbol{\beta}^{(k)}$) if and only if $(\beta_0^{(k)})^2 \leq \beta_1^{(k)}\beta_2^{(k)}$, $k = 1,2$. The latter optimality conditions transfer to the product type design $\xi^* = \xi_1^* \otimes \xi_2^*$ which is thus locally D-optimal on the experimental region $\mathcal{X} = [0,1]^4$ for model (4.67) at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} \otimes \boldsymbol{\beta}^{(2)}$ if $(\beta_0^{(1)}\beta_0^{(2)})^2 \leq (\beta_1^{(1)}\beta_1^{(2)})(\beta_2^{(1)}\beta_2^{(2)})$, i.e., $\beta_0^2 \leq \beta_5\beta_8$ and ξ^* assigns weights $1/9$ to
$\{(0,0)^\top, (1,0)^\top, (0,1)^\top\} \times \{(0,0)^\top, (1,0)^\top, (0,1)^\top\}=$
$\{(0,0,0,0)^\top$,$(0,0,1,0)^\top$, $(0,0,0,1)^\top$, $(1,0,0,0)^\top$, $(1,0,1,0)^\top$, $(1,0,0,1)^\top$,$(0,1,0,0)^\top$, $(0,1,1,0)^\top$, $(0,1,0,1)^\top\}$.

For A-optimality, from part (iv) of Corollary 4.4.2 the design ξ_k^* on $\mathcal{X}_k = [0,1]^2$ with support $\{(0,0)^\top, (1,0)^\top, (0,1)^\top\}$ and the following weights

$$\begin{aligned}
\xi_k^*\big((0,0)^\top\big) &= \sqrt{3}\beta_0^{(k)}/c, \\
\xi_k^*\big((1,0)^\top\big) &= (\beta_0^{(k)} + \beta_1^{(k)})/c, \\
\xi_k^*\big((0,1)^\top\big) &= (\beta_0^{(k)} + \beta_2^{(k)})/c \ \text{ where } c = (\sqrt{3}+2)\beta_0^{(k)} + \beta_1^{(k)} + \beta_2^{(k)}, k = 1,2,
\end{aligned}$$

is locally A-optimal (at $\boldsymbol{\beta}^{(k)}$) if and only if $(1+2/\sqrt{3})(\beta_0^{(k)})^2 + (1/\sqrt{3})\beta_0^{(k)}(\beta_1^{(k)} + \beta_2^{(k)}) - \beta_1^{(k)}\beta_2^{(k)} \le 0$, $k = 1,2$. Hence, the product type design $\xi^* = \xi_1^* \otimes \xi_2^*$ which is thus locally A-optimal (at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)} \otimes \boldsymbol{\beta}^{(2)}$) on the experimental region $\mathcal{X} = [0,1]^4$ for model (4.67) assigns weights $\xi_1^*(\boldsymbol{x}_1)\xi_2^*(\boldsymbol{x}_2)$ to the point $(\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top)^\top \in \{(0,0)^\top, (1,0)^\top, (0,1)^\top\} \times \{(0,0)^\top, (1,0)^\top, (0,1)^\top\}$.

Another result of this type treats the Kiefer Φ_s-optimality (for notational convenience we use index s instead of index k) which covers D-optimality at $s = 0$, A-optimality at $s = 1$ and E-optimality at $s \to \infty$. Therefore, we obtain the next lemma.

Lemma 4.5.3. *For every product type design* $\xi = \bigotimes_{k=1}^{K} \xi_k$ *we have the next factorization*

$$\Phi_s(\xi, \boldsymbol{\beta}) = \prod_{k=1}^{K} \Phi_s(\xi_k, \boldsymbol{\beta}^{(k)}),\ (0 \le s < \infty)$$

Proof. Employing the definition of Kiefer Φ_s-criterion yields

$$\begin{aligned}
\Phi_s(\xi, \boldsymbol{\beta}) &= \Big(\frac{1}{p}\mathrm{tr}\Big(\boldsymbol{M}^{-s}(\xi, \boldsymbol{\beta})\Big)\Big)^{\frac{1}{s}} = \left(\frac{1}{p}\mathrm{tr}\Big(\bigotimes_{k=1}^{K}\Big(\boldsymbol{M}_k(\xi_k, \boldsymbol{\beta}^{(k)})\Big)\Big)^{-s}\right)^{\frac{1}{s}} \\
&= \left(\prod_{k=1}^{K}\frac{1}{p_k}\mathrm{tr}\Big(\boldsymbol{M}_k^{-s}(\xi_k, \boldsymbol{\beta}^{(k)})\Big)\right)^{\frac{1}{s}} = \prod_{k=1}^{K}\Phi_s(\xi_k, \boldsymbol{\beta}^{(k)}).
\end{aligned}$$

□

Theorem 4.5.3. *Let* ξ_k^* *be a locally* Φ_s*-optimal design (at* $\boldsymbol{\beta}^{(k)}$*) for a marginal* ν_k*-factor model (4.54) on the experimental region* $\mathcal{X}_k$ $(1 \le k \le K)$*. Then* $\xi^* = \bigotimes_{k=1}^{K} \xi_k^*$ *is a locally* Φ_s*-optimal design (at* $\boldsymbol{\beta} = \bigotimes_{k=1}^{K} \boldsymbol{\beta}^{(k)}$*) for model (4.56) on the experimental region* $\mathcal{X} = \times_{k=1}^{K}\mathcal{X}_k$.

Proof. The proof is obtained by making use of condition (2.13) of The Equivalence Theorem (Theorem 2.2.2). Since ξ_k^* is locally Φ_s-optimal (at $\boldsymbol{\beta}^{(k)}$) we guarantee that

$\boldsymbol{f}^{(k)\top}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-s-1}(\xi_k^*, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) \leq \mathrm{tr}(\boldsymbol{M}_k^{-s}(\xi_k^*, \boldsymbol{\beta}^{(k)}))$ for all $\boldsymbol{x}_k \in \mathcal{X}_k$. Thus in view of Lemma 4.5.2 with (4.57) and (4.58) we obtain

$$\begin{aligned}
\boldsymbol{f}_{\boldsymbol{\beta}}^\top(\boldsymbol{x})\boldsymbol{M}^{-s-1}(\xi^*, \boldsymbol{\beta})\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) &= u(\boldsymbol{x}, \boldsymbol{\beta})\Big(\bigotimes_{k=1}^K \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big)^\top \Big(\bigotimes_{k=1}^K \boldsymbol{M}_k\big(\xi_k, \boldsymbol{\beta}^{(k)}\big)\Big)^{-s-1}\Big(\bigotimes_{k=1}^K \boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big) \\
&= \bigotimes_{k=1}^K \Big(u_k(\boldsymbol{x}_k, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)\top}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-s-1}\big(\xi_k^*, \boldsymbol{\beta}^{(k)}\big)\boldsymbol{f}^{(k)}(\boldsymbol{x}_k)\Big) \\
&= \prod_{k=1}^K \boldsymbol{f}^{(k)\top}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k)\boldsymbol{M}_k^{-s-1}(\xi_k^*, \boldsymbol{\beta}^{(k)})\boldsymbol{f}^{(k)}_{\boldsymbol{\beta}^{(k)}}(\boldsymbol{x}_k) \\
&\leq \prod_{k=1}^K \mathrm{tr}(\boldsymbol{M}_k^{-s}(\xi_k^*, \boldsymbol{\beta}^{(k)})) = \mathrm{tr}\Big(\Big(\bigotimes_{k=1}^K \Big(\boldsymbol{M}_k\big(\xi_k^*, \boldsymbol{\beta}^{(k)}\big)\Big)\Big)^{-s}\Big) = \mathrm{tr}\Big(\boldsymbol{M}^{-s}(\xi^*, \boldsymbol{\beta})\Big)
\end{aligned}$$

for all $\boldsymbol{x} \in \mathcal{X}$. The Equivalence Theorem, thus, proves the local Φ_s-optimality of the product design $\xi^* = \bigotimes_{k=1}^K \xi_k^*$ at a parameter point $\boldsymbol{\beta} = \bigotimes_{k=1}^K \boldsymbol{\beta}^{(k)}$. □

4.6 Design efficiency and simulation

It is interesting to study the adequacy of locally optimal designs for generalized linear models because typically, the performance of a locally optimal design is affected by the initial parameter values. Misspecified values may lead to a poor performance of the locally optimal design and thus a highly sensitivity of the statistical inference might occur. From our results, in this chapter, each locally optimal design refers to a specific subregion of the parameter space where the design is optimal. In this section we discuss the potential benefits of the derived designs for gamma models with intercept and without intercept. We restrict ourselves to the optimal designs that have been derived with respect to the D-criterion specifically from Corollary 4.4.1, Theorem 4.4.4 and Corollary 4.5.3. Our objective is to examine the overall performance of some of the locally D-optimal designs as well as investigating the performance for finite sample sizes.

4.6.1 Design efficiency

The overall performance of any design ξ is described by its D-efficiencies, as a function of $\boldsymbol{\beta}$,

$$\mathrm{Eff}(\xi, \boldsymbol{\beta}) = \left(\frac{\det \boldsymbol{M}(\xi, \boldsymbol{\beta})}{\det \boldsymbol{M}(\xi_{\boldsymbol{\beta}}^*, \boldsymbol{\beta})}\right)^{1/p} \tag{4.68}$$

where $\xi_{\boldsymbol{\beta}}^*$ denotes the locally D-optimal design at $\boldsymbol{\beta}$ and p is the dimension of $\boldsymbol{\beta}$.

Example 4.6.1. In the situation of Corollary 4.4.1 we consider the D-optimal designs for gamma models with intercept. The experimental region is given by $\mathcal{X} = [0,1]^2$ with the vertices $\boldsymbol{v}_1 = (0,0)^\top$, $\boldsymbol{v}_2 = (1,0)^\top$, $\boldsymbol{v}_3 = (0,1)^\top$, and $\boldsymbol{v}_4 = (1,1)^\top$. For simplicity we restrict to case $\beta_1 = \beta_2 = \beta$, say. That is, the parameter vector is of the form $\boldsymbol{\beta} = (\beta_0, \beta, \beta)^\top$. We utilize the ratio $\gamma = \beta/\beta_0, \beta_0 > 0$. Note that the positivity condition (4.4) implies that $\gamma \in (-1/2, \infty)$. Our interest is in the saturated and equally weighted designs ξ_1 and ξ_2 where $\text{supp}(\xi_1) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, $\text{supp}(\xi_2) = \{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ which by Corollary 4.4.1 are locally D-optimal at $\gamma \geq 1$ and $\gamma \in (-1/2, -1/3]$, respectively. In particular, ξ_1 and ξ_2 are robust against misspecified parameter values in their respective subregions. Additionally, for $\gamma \in (-1/3, 1)$ we consider the locally D-optimal design $\xi_3(\gamma)$ given by Theorem 4.4.1. Note that $\text{supp}(\xi_3(\gamma)) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and the weights depend on $\boldsymbol{\gamma}$.

To employ (4.68) we put $\xi^*_{\boldsymbol{\beta}} = \xi_1$ if $\gamma \geq 1$, $\xi^*_{\boldsymbol{\beta}} = \xi_2$ if $\gamma \in (-1/2, -1/3]$ and $\xi^*_{\boldsymbol{\beta}} = \xi_3(\gamma)$ if $\gamma \in (-1/3, 1)$. We select for examination the designs ξ_1, ξ_2, $\xi_3(0)$ and, moreover as a natural competitor ξ_4 which assigns uniform weights to the grid $\{0, 0.5, 1\}^2$. Note that $\xi_3(0)$ assigns uniform weights to the set of vertices $\{0,1\}^2$.

In Figure 4.8, the D-efficiencies of the four designs ξ_1, ξ_2, $\xi_3(0)$ and ξ_4 are depicted. The efficiencies of ξ_1 and ξ_2 are, of course, equal to 1 in their optimality subregions $\gamma \in [1, \infty)$ and $\gamma \in (-1/2, -1/3]$, respectively. However, for γ outside but fairly close to the respective optimality subregion both designs perform quite well; the efficiencies of ξ_1 and ξ_2 are greater than 0.80 for $0.07 \leq \gamma < 1$ and $-1/3 < \gamma \leq -0.06$, respectively. However, their efficiencies decrease towards zero when γ moves away from the respective optimality subregion. So, the overall performance of ξ_1 and ξ_2 cannot be regarded as satisfactory. The design $\xi_3(0)$, though locally D-optimal only at $\gamma = 0$, does show a more satisfactory overall performance with efficiencies range between 0.8585 and 1. The design ξ_4 turns out to be uniformly worse than $\xi_3(0)$ and its efficiencies range between 0.6598 and 0.7631.

In addition, we studied the performance of optimal designs of the form $\xi_3(\gamma)$ for various $\gamma \in (-1/3, 1)$. The efficiencies of some of these designs are shown in Figure 4.9 for $\gamma \in \{-0.2, 0, 0.6\}$. We observe that the performance of the design $\xi_3(\gamma)$ comes closer to the performance of ξ_1 or ξ_2 when γ approaches 1 or $-1/3$, respectively.

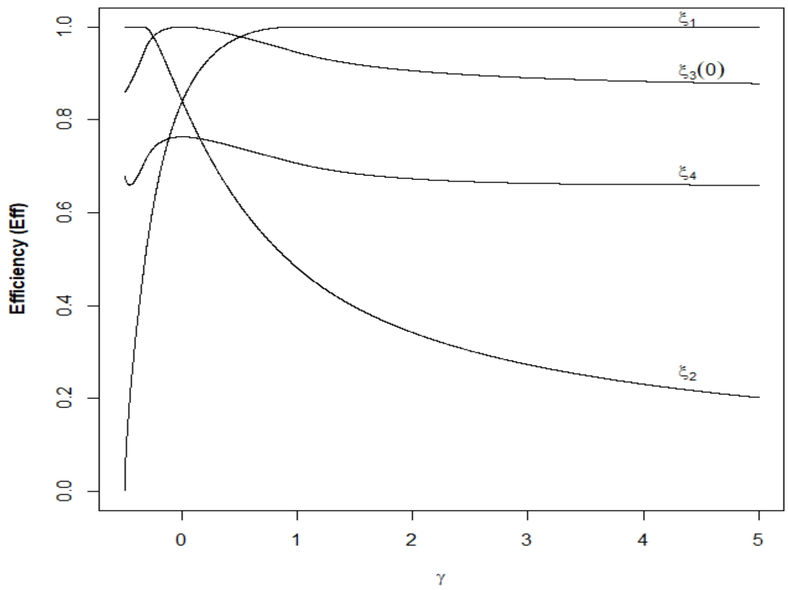


FIGURE 4.8: Example 4.6.1. D-efficiencies from (4.68) over the region $-1/2 < \gamma < \infty$, $\gamma = \beta/\beta_0$, $\beta = \beta_1 = \beta_2, \beta_0 > 0$.

,

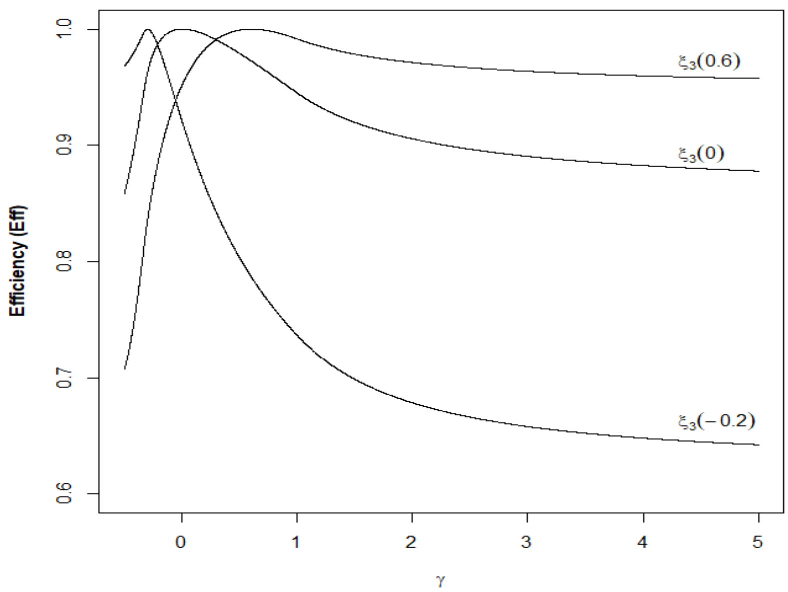


FIGURE 4.9: Example 4.6.1. D-efficiencies of $\xi_3(-0.2)$, $\xi_3(0)$, $\xi_3(0.6)$ from (4.68) over the region $-1/2 < \gamma < \infty$, $\gamma = \beta/\beta_0, \beta = \beta_1 = \beta_2$, $\beta_0 > 0$.

Example 4.6.2. In the situation of Theorem 4.4.4 the experimental region is given by $\mathcal{X} = [1,2]^3$ with the vertices $\boldsymbol{v}_1 = \big(1,1,1\big)^\top$, $\boldsymbol{v}_2 = \big(2,1,1\big)^\top$, $\boldsymbol{v}_3 = \big(1,2,1\big)^\top$, $\boldsymbol{v}_4 = \big(1,1,2\big)^\top$, $\boldsymbol{v}_5 = \big(1,2,2\big)^\top$, $\boldsymbol{v}_6 = \big(2,1,2\big)^\top$, $\boldsymbol{v}_7 = \big(2,2,1\big)^\top$, $\boldsymbol{v}_8 = \big(2,2,2\big)^\top$. We restrict only to the case $\beta_1 > 0$, $\beta_2 = \beta_3 = \beta$ and hence we utilize the ratio $\gamma = \beta/\beta_1$ with range $(-1/4, \infty)$. Our interest is in the saturated and equally weighted designs ξ_1 and ξ_2 where $\text{supp}(\xi_1) = \{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and $\text{supp}(\xi_2) = \{\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$ which by Theorem 4.4.4 are locally D-optimal at $\gamma \geq 1/5$ and $\gamma \in (-1/4, -5/23]$, respectively. In particular, ξ_1 and ξ_2 are robust against misspecified parameter values in their respective subregions. Additionally, for $\gamma \in (-5/23, 1/5)$ we consider the locally D-optimal designs of type $\xi_3(\gamma)$ given by the theorem. Note that $\text{supp}(\xi_3(\gamma)) = \{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$ and the weights depend on γ.

To employ (4.68) we put $\xi^*_{\boldsymbol{\beta}} = \xi_1$ if $\gamma \geq 1/5$, $\xi^*_{\boldsymbol{\beta}} = \xi_2$ if $\gamma \in (-1/4, -5/23]$ and $\xi^*_{\boldsymbol{\beta}} = \xi_3(\gamma)$ if $\gamma \in (-5/23, 1/5)$. We select for examination the designs ξ_1, ξ_2, $\xi_3(-1/7)$. Moreover, as natural competitors we select various uniform designs supported by specific vertices. That is ξ_4 with support $\{\boldsymbol{v}_i : i = 1, \ldots, 8\}^3$ and the two half-fractional designs ξ_5 and ξ_6 supported by $\{\boldsymbol{v}_1, \boldsymbol{v}_5, \boldsymbol{v}_6, \boldsymbol{v}_7\}$ and $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_8\}$, respectively. Additionally, we consider ξ_7 which assigns uniform weights to the grid $\{1, 1.5, 2\}^3$.

In Figure 4.10, the D-efficiencies of the designs ξ_1, ξ_2, $\xi_3(-1/7)$, ξ_4, ξ_5, ξ_6 and ξ_7 are depicted. In analogy to Example 4.6.1 similar interpretation can be presented. The efficiencies of ξ_1 and ξ_2 are, of course, equal to 1 in their optimality subregions $\gamma \in [1/5, \infty)$ and $\gamma \in (-1/4, -5/23]$, respectively. However, for γ outside but fairly close to the respective optimality subregion both designs perform quite well; the efficiencies of ξ_1 and ξ_2 are greater than 0.80 for $-0.15 \leq \gamma < 1/5$ and $-1/4 < \gamma \leq -0.28$, respectively. However, their efficiencies decrease towards zero when γ moves away from the respective optimality subregion. So, the overall performance of ξ_1 and ξ_2 cannot be regarded as satisfactory. The design $\xi_3(-1/7)$, though locally D-optimal only at $\gamma = -1/7$, does show a more satisfactory overall performance with efficiencies range between 0.8585 and 1. The efficiencies of the half-fractional design ξ_6 are greater than 0.80 only for $\gamma > -0.049$, otherwise the efficiencies decrease towards zero. The design ξ_4 turns out to be uniformly worse than $\xi_3(-1/7)$ and its efficiencies range between 0.5768 and 0.7615. The worst performance is shown by the designs ξ_5 and ξ_7.

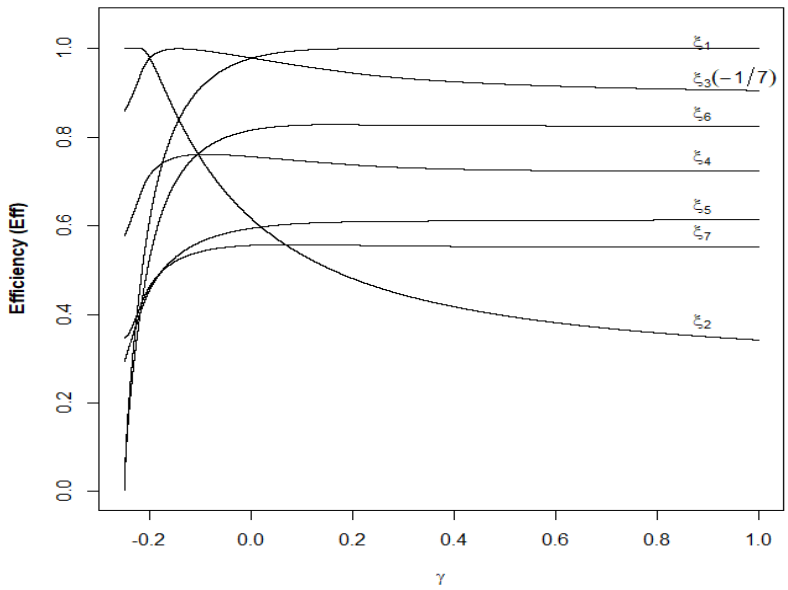


FIGURE 4.10: Example 4.6.2. D-efficiencies from (4.68) over the region $-1/4 < \gamma \leq 1,\ \gamma = \beta/\beta_1,\ \beta = \beta_2 = \beta_3, \beta_1 > 0$.

Example 4.6.3. In the situation of Corollary 4.5.3 we consider the experimental region $\mathcal{X} = [1,4]^2$ where condition $b - 3a > 0$ is satisfied. The vertices are denoted by $\boldsymbol{v}_1 = \big(4,4\big)^\top$, $\boldsymbol{v}_2 = \big(4,1\big)^\top$, $\boldsymbol{v}_3 = \big(1,4\big)^\top$, $\boldsymbol{v}_4 = \big(1,1\big)^\top$. We restrict to $\beta_3 \neq 0$, $\beta_1 = \beta_2 = \beta$, and the range of $\gamma = \beta/\beta_3$ is $(-1/2, \infty)$. In analogy to Example 4.6.2 denote by ξ_1 and ξ_2 the saturated and equally weighted designs with support $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ and $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$, respectively. By the corollary ξ_1 and ξ_2 are locally D-optimal at $\gamma \geq 4$ and $\gamma \in (-1/2, -4/11]$, restrictively. Denote by $\xi_3(\gamma)$ the design given in part *(iii)* of Corollary 4.5.3 which is locally D-optimal at $\gamma \in (-4/11, 4)$. Note that from (4.68) we put $\xi^*_{\boldsymbol{\beta}} = \xi_1$ if $\gamma \geq 4$, $\xi^*_{\boldsymbol{\beta}} = \xi_2$ if $\gamma \in (-1/2, -4/11]$ and $\xi^*_{\boldsymbol{\beta}} = \xi_3(\gamma)$ if $\gamma \in (-4/11, 4)$. For examination we select ξ_1, ξ_2, $\xi_3(0)$. As a natural competitor we select ξ_4 that assigns uniform weights to the grid $\{1, 2.5, 4\}^2$. The efficiencies are depicted in Figure 4.11. We observe that the performance of ξ_1 and ξ_2 is similar to that of the corresponding designs in Example 4.6.2. Moreover, the design $\xi(0)$ show a more satisfactory overall performance. The efficiencies of ξ_4 vary between 0.77 and 0.83 for $\gamma > -4/11$.

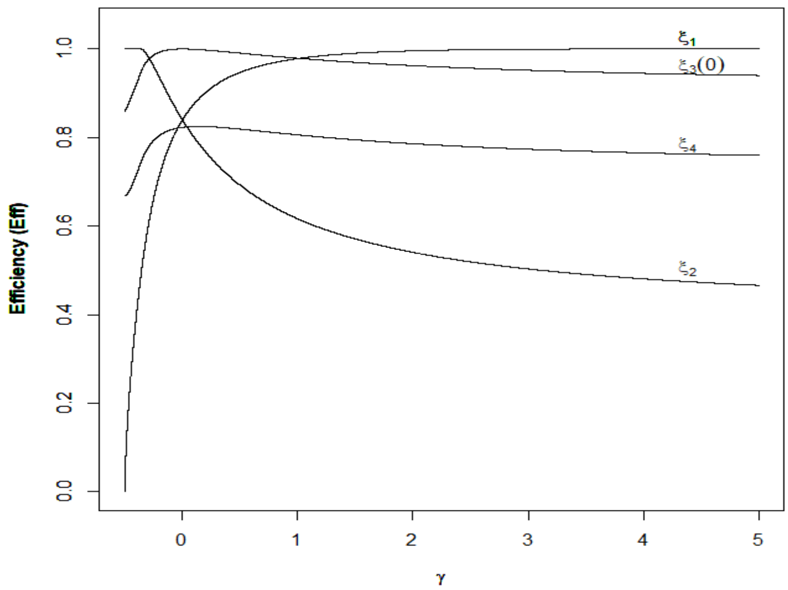


FIGURE 4.11: Example 4.6.3. D-efficiencies from (4.68) over the region $-1/2 < \gamma \leq 5,\ \gamma = \beta/\beta_3,\ \beta = \beta_1 = \beta_2, \beta_3 \neq 0$.

4.6.2 Simulation

It is worth to examine also the performance of a locally D-optimal design for finite sample sizes, in particular, under Example 4.6.1 in the preceding subsection. So, for the locally D-optimal design ξ_1 we compare the precision of the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\beta}}(\xi_1, n)$ under ξ_1 and for sample size n with that of the MLE $\hat{\boldsymbol{\beta}}(\xi, n)$ under another non-optimal design ξ for the same sample size n. Note that here the true parameter vector $\boldsymbol{\beta}$ is chosen from the optimality subregion of ξ_1, i.e., $\beta \geq 1$. Denote by $\boldsymbol{V}(\hat{\boldsymbol{\beta}}(\xi, n), \boldsymbol{\beta})$ the variance-covariance matrix of the MLE at $\boldsymbol{\beta}$ for a given design ξ. Since there is no analytic formula for this quantity its (approximate) numerical computation is done by simulations. Our simulations showed in particular, that the biases of the MLE's are small and hence only the variance-covariance matrices are relevant. In analogy to (4.68) consider the D-efficiencies,

$$\text{Eff}(\xi, \xi_1, n, \boldsymbol{\beta}) = \left(\frac{\det \boldsymbol{V}(\hat{\boldsymbol{\beta}}(\xi_1, n), \boldsymbol{\beta})}{\det \boldsymbol{V}(\hat{\boldsymbol{\beta}}(\xi, n), \boldsymbol{\beta})} \right)^{1/p} \tag{4.69}$$

which give the relative precision of the MLE under ξ, relative to the precision of the MLE under ξ_1, at the parameter point $\boldsymbol{\beta}$ and for the sample size n. In fact, for large sample size n the efficiencies given by (4.68) for $\xi^*_\beta = \xi_1$ will be close to those from (4.69). In our simulation study we generated independent gamma-distributed observations according to the designs under consideration with shape parameter $\kappa = 1$

and expectations $\mu_i = 1/(1 + \beta x_{1i} + \beta x_{2i})$, $i = 1, \ldots, n, \beta \in \{1, 3, 5\}$. For each sample size $n \in \{36, 72, 108, 360, 720, 1080, 1800, 3600\}$ and each $\beta \in \{1, 3, 5\}$ $s = 10000$ simulation runs were carried out. For each instance we calculated the approximate expectation $\mathrm{E}(\hat{\boldsymbol{\beta}}(\xi, n)) \approx \frac{1}{s}\sum_{i=1}^{s} \hat{\boldsymbol{\beta}}_i(\xi, n)$ and the approximate variance-covariance matrix $\boldsymbol{V}(\hat{\boldsymbol{\beta}}(\xi, n), \boldsymbol{\beta}) \approx \frac{1}{s}\sum_{i=1}^{s} \left(\hat{\boldsymbol{\beta}}_i(\xi, n) - \mathrm{E}(\hat{\boldsymbol{\beta}}(\xi, n))\right)\left(\hat{\boldsymbol{\beta}}_i(\xi, n) - \mathrm{E}(\hat{\boldsymbol{\beta}}(\xi, n))\right)^{\top}$.

In Table 4.5 the computed values of the efficiencies (4.69) are reported for the designs $\xi = \xi_3(0)$ and $\xi = \xi_4$. The table shows the benefit of the locally D-optimal design ξ_1 compared to its competitors even for moderate finite sample sizes. Of course, for large sample sizes the reported efficiencies nearly coincide with those from (4.68) which are addressed to in Table 4.5 as $n = \infty$.

TABLE 4.5: Example 4.6.1. D-efficiencies of $\xi_3(0)$ and ξ_4 from (4.69). The employed built-in R-algorithm did not yield results for $n = 36$ under ξ_4. For $n \to \infty$ the values are equal to efficiencies from (4.68)

	$\mathrm{Eff}(\xi_3(0), \xi_1, n, \boldsymbol{\beta})$			$\mathrm{Eff}(\xi_4, \xi_1, n, \boldsymbol{\beta})$		
n	$\beta = 1$	$\beta = 3$	$\beta = 5$	$\beta = 1$	$\beta = 3$	$\beta = 5$
36	1.0431	0.9412	0.9189	...	...	...
72	0.9832	0.9046	0.8736	0.7614	0.6604	0.6306
108	0.9607	0.8891	0.8965	0.7297	0.6739	0.6522
360	0.9501	0.8997	0.8971	0.7061	0.6702	0.6575
720	0.9328	0.8998	0.8782	0.7078	0.6601	0.6589
1080	0.9311	0.9053	0.8880	0.7063	0.6630	0.6548
1800	0.9399	0.8903	0.8843	0.7032	0.6643	0.6572
3600	0.9551	0.9002	0.8720	0.7128	0.6673	0.6666
∞	0.9449	0.8904	0.8778	0.7061	0.6634	0.6598

Chapter 5

Extensions to multivariate generalized linear models

The purpose of this chapter is to study the design optimality for the multivariate generalized linear models, MGLMs. We will use the results obtained in Chapter 3 and Chapter 4 under univariate models to derive optimal designs in the multivariate structure. In Section 5.1 we introduce the model with related assumptions. In Section 5.2 we develop a particular solution of optimal designs for a general setup of MGLMs. In Section 5.3 we concentrate on a MGLM with univariate gamma models.

5.1 Model specification

Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ be independent m-dimensional response variables for n experimental units. There are n observations taken for each one of the m components of the experimental unit i, $i = 1, \ldots, n$. Let a compact experimental region $\mathcal{X}$ be given. Denote by

$$\boldsymbol{Y} = \left(Y_1, \ldots, Y_m\right)^\top$$

the vector of responses for a particular unit at a point $\boldsymbol{x} \in \mathcal{X}$, i.e., an m-dimensional real valued vector is observed instead of a single real valued random variable at each point $\boldsymbol{x} \in \mathcal{X}$.

The distribution of a single response Y_j is assumed to belong to a one-parameter exponential family distribution $p(Y_j; \theta_j, \phi_j)$ from (2.3). Therefore, the approach of the generalized linear model, GLM, that was introduced in Section 2.1 is utilized and to be extended here. Each jth component has expected mean $E(Y_j) = \mu_j = b_j'(\theta_j)$ and variance function $V_j(\mu_j) = b_j''(\theta_j)$ and thus, $\mathrm{var}(Y_j) = a_j(\phi_j)V_j(\mu_j)$. The expected mean μ_j is combined to the linear predictor $\boldsymbol{f}_j^\top(\boldsymbol{x})\,\boldsymbol{\beta}_j$ by a proper link function g_j as in (2.4);

$$\eta_j = g_j\left(\mu_j\right) \quad \text{where} \quad \eta_j = \boldsymbol{f}_j^\top(\boldsymbol{x})\,\boldsymbol{\beta}_j \ (1 \le j \le m),$$

and $\boldsymbol{f}_j(\boldsymbol{x})$ is the p_j-dimensional vector of known regression functions $f_{j1}, \ldots, f_{jp_j}$ with the vector of unknown parameters $\boldsymbol{\beta}_j = \left(\beta_{j1}, \ldots, \beta_{jp_j}\right)^\top \in \mathbb{R}^{p_j}$. Note that in the jth component $\boldsymbol{f}_j^\top(\boldsymbol{x})\boldsymbol{\beta}_j = \sum_{l=1}^{p_j} f_{jl}(\boldsymbol{x})\beta_{jl}$. The total number of MGLM parameters is denoted by p, i.e., $p = \sum_{j=1}^{m} p_j$. The link functions $g_j\,(1 \le j \le m)$ are not necessarily similar and thus the single responses $Y_j\,(1 \le j \le m)$ may belong to distinct one-parameter probability distributions.

Note that $\mu_j = \mu_j(\boldsymbol{x}, \boldsymbol{\beta}) = g_j^{-1}\left(\boldsymbol{f}_j^\top(\boldsymbol{x})\boldsymbol{\beta}_j\right)$ and $\mathrm{d}\eta_j/\mathrm{d}\mu_j = g_j'\left(g_j^{-1}\left(\boldsymbol{f}_j^\top(\boldsymbol{x})\boldsymbol{\beta}_j\right)\right)$ so the intensity function in the jth component is given from (2.5) as

$$u_j(\boldsymbol{x}, \boldsymbol{\beta}_j) = \left(\mathrm{var}(Y_j)\left(\frac{\mathrm{d}\eta_j}{\mathrm{d}\mu_j}\right)^2\right)^{-1} \text{ for all } \ \boldsymbol{x} \in \mathcal{X}\ (1 \le j \le m).$$

Let $\mathbf{f}(\boldsymbol{x}) = \mathrm{diag}\left(\boldsymbol{f}_1(\boldsymbol{x}), \ldots, \boldsymbol{f}_m(\boldsymbol{x})\right)$ denotes the $p \times m$ block diagonal multivariate regression function and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ is the stacked parameter vector of dimension $p \times 1$. Denote by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^\top$ the vector of expected means of a unit at a point $\boldsymbol{x} \in \mathcal{X}$. The MGLM for each unit at a point $\boldsymbol{x} \in \mathcal{X}$ is defined by

$$\boldsymbol{\eta} = \boldsymbol{g}\left(\boldsymbol{\mu}\right) \quad \text{where} \quad \boldsymbol{\eta} = \mathbf{f}^\top(\boldsymbol{x})\boldsymbol{\beta} \tag{5.1}$$

with $\boldsymbol{g}\left(\boldsymbol{\mu}\right) = \left(g_1(\mu_1), \ldots, g_m(\mu_m)\right)^\top$ and $\mathbf{f}^\top(\boldsymbol{x})\boldsymbol{\beta} = \left(\boldsymbol{f}_1^\top(\boldsymbol{x})\boldsymbol{\beta}_1, \ldots, \boldsymbol{f}_m^\top(\boldsymbol{x})\boldsymbol{\beta}_m\right)^\top$. To assure estimability it is assumed that the components $f_{j1}, \ldots, f_{jp_j}$ of $\boldsymbol{f}_j(\boldsymbol{x})$ are linearly independent functions on $\mathcal{X}$ and thus, the components of $\mathbf{f}(\boldsymbol{x})$ are linearly independent functions on $\mathcal{X}$.

The simplest situation can be taken under identiy links, i.e., $\boldsymbol{g}\left(\boldsymbol{\mu}\right) = \boldsymbol{\mu}$ for which the intensities $u_j(\boldsymbol{x}, \boldsymbol{\beta}_j)$, $j = 1, \ldots, m$ are constants equal to 1 for any $\boldsymbol{x} \in \mathcal{X}$. Hence, the design problems are addressed under the multivariate linear model, e.g. see Chang (1994) and Yue, Liu, and Chatterjee (2014). However, Liang and Zeger (1986) mentioned that there is a lack of a rich class of distributions for the multivariate non-normal outcomes. Therefore, they proposed the method of generalized estimating equations (GEEs) to estimate the model parameters. GEEs are considered as an extension of the score function for the GLM. However, in optimal design theory GEEs were used to obtain optimal blocked designs for correlated binary data in Woods and Ven (2011) and then used in Ven and Woods (2014) to find optimal blocked minimum-support designs for non-linear models.

To employ GEEs method we assume that the observations $\boldsymbol{Y}_i\,(1 \le i \le n)$ are uncorrelated across the units while the components are correlated within each unit. That is for the observation $\boldsymbol{Y}$ let $\boldsymbol{R}$ be the $m \times m$ true correlation matrix which is independent of $\boldsymbol{x}$ and $\boldsymbol{\beta}$. The correlation matrix $\boldsymbol{R}$ is assumed to be positive definite and might rewrite as $\boldsymbol{R} = \left(\rho_{jh}\right)_{j=1,\ldots,m}^{h=1,\ldots,m}$ where $\rho_{jj} = 1\,(1 \leq j \leq m)$ and $-1 \leq \rho_{jh} < 1\ (1 \leq j < h \leq m)$. Denote also the inverse of the correlation matrix by $\boldsymbol{R}^{-1} = \left(\rho^{(jh)}\right)_{j=1,\ldots,m}^{h=1,\ldots,m}$.

Remark 5.1.1. *In general, for a square matrix $\boldsymbol{B}$ if there exists a matrix $\boldsymbol{C}$ such that $\boldsymbol{C}\boldsymbol{C}^\top = \boldsymbol{B}$, then we call $\boldsymbol{C}$ a square root of the matrix $\boldsymbol{B}$. If $\boldsymbol{B}$ is a diagonal matrix given by $\boldsymbol{B} = \mathrm{diag}(b_1, \ldots, b_m)$ then we can define its square root as $\boldsymbol{C} = \mathrm{diag}(b_1^{\frac{1}{2}}, \ldots, b_m^{\frac{1}{2}})$ and we denote $\boldsymbol{B}^{\frac{1}{2}} = \boldsymbol{C}$.*

Define $\boldsymbol{A}(\boldsymbol{x}, \boldsymbol{\beta}) = \mathrm{diag}\left(\mathrm{var}(Y_j)\right)_{j=1}^{m}$ for all $\boldsymbol{x} \in \mathcal{X}$. Then the observation $\boldsymbol{Y}$ at a point $\boldsymbol{x} \in \mathcal{X}$ has the covariance structure $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma}\,(\boldsymbol{x}, \boldsymbol{\beta})$, see Liang and Zeger (1986), where

$$\boldsymbol{\Sigma}\,(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{x}, \boldsymbol{\beta})\,\boldsymbol{R}\,\boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{x}, \boldsymbol{\beta}). \tag{5.2}$$

Let $\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{\beta}) = \mathrm{diag}\left(\frac{\mathrm{d}\mu_j}{\mathrm{d}\eta_j}\right)_{j=1}^{m}$ for all $\boldsymbol{x} \in \mathcal{X}$. In the context of GEEs we define the quasi-score function as $\boldsymbol{U}\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{n} \mathbf{f}(\boldsymbol{x}_i)\boldsymbol{\Delta}(\boldsymbol{x}_i, \boldsymbol{\beta})\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i, \boldsymbol{\beta})\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)$ where $\boldsymbol{U}\,(\boldsymbol{\beta})$ is a $p \times 1$ quasi-score vector. The maximum quasi-likelihood estimates $\hat{\boldsymbol{\beta}}$ is the solution of the generalized estimating equations $\boldsymbol{U}\,(\boldsymbol{\beta}) = \mathbf{0}_p$, where $\mathbf{0}_p$ is a $p \times 1$ vector of zeros, see Crowder (1995).

The quasi-Fisher information matrix for the MGLM at a single point $\boldsymbol{x}$ is given by $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \mathbf{f}(\boldsymbol{x})\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{x}, \boldsymbol{\beta})\,\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{\beta})\mathbf{f}^\top(\boldsymbol{x})$.

By modifying function (2.7) for each component j we write

$$\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}) = \sqrt{u_j(\boldsymbol{x}, \boldsymbol{\beta}_j)}\boldsymbol{f}_j(\boldsymbol{x}),\ j = 1, \ldots, m,$$

which then constitute the $p \times m$ matrix $\mathbf{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \mathrm{diag}\Big(\boldsymbol{f}_{1,\boldsymbol{\beta}_1}(\boldsymbol{x}), \ldots, \boldsymbol{f}_{m,\boldsymbol{\beta}_m}(\boldsymbol{x})\Big)$. It is straightforward to obtain

$$\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}, \boldsymbol{\beta})\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{\beta}) = \mathrm{diag}\left(u_j^{\frac{1}{2}}(\boldsymbol{x}, \boldsymbol{\beta}_j)\right)_{j=1}^{m}\boldsymbol{R}^{-1}\,\mathrm{diag}\left(u_j^{\frac{1}{2}}(\boldsymbol{x}, \boldsymbol{\beta}_j)\right)_{j=1}^{m},$$

and thus the quasi-score function rewrites as $\boldsymbol{U}\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{n} \mathbf{f}_{\boldsymbol{\beta}}(\boldsymbol{x}_i)\boldsymbol{R}^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)$ whereas the quasi-Fisher information matrix reads as

$$\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \mathbf{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{R}^{-1}\mathbf{f}_{\boldsymbol{\beta}}^\top(\boldsymbol{x}).$$

For the whole experiment we introduce the information matrix

$$\boldsymbol{M}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{\beta}) \tag{5.3}$$

which rewrites in a block representation as

$$\boldsymbol{M}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{\beta}) = \left(\rho^{(jh)}\sum_{i=1}^{n}\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_i)\boldsymbol{f}^\top_{h,\boldsymbol{\beta}_h}(\boldsymbol{x}_i)\right)_{j=1,\ldots,m}^{h=1,\ldots,m}. \tag{5.4}$$

The variance-covariance matrix $\mathrm{var}(\hat{\boldsymbol{\beta}})$ of the estimated parameters is approximately

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{M}^{-1}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{\beta}).$$

Multi-dimensional observations are rearranged in matrix form in different ways. For the design point of view, particularly, under our assumptions we are to emphasize the relation of MGLM to its univariate GLM as for the linear case in Zellner (1962), Krafft and Schaefer (1992) and Kurotschka and Schwabe (1996). The observational vector of the whole experiment is obtained by vectorization of the data (design) matrix, i.e., by stacking the columns on top of each other which represent the components. Therefore, let $\boldsymbol{Y}_j = \big(Y_j(\boldsymbol{x}_1),\ldots,Y_j(\boldsymbol{x}_n)\big)^\top$ be the observations of the jth component of the whole experiment $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$. The stacked vector of responses for all units at the whole experiment is thus denoted by $\boldsymbol{Y} = \big(\boldsymbol{Y}_1^\top,\ldots,\boldsymbol{Y}_m^\top\big)^\top$.

In this context, the design matrix $\boldsymbol{F}$ for the multivariate model is written in component wise. So let $\boldsymbol{F}_j = [\boldsymbol{f}_j(\boldsymbol{x}_1),\ldots,\boldsymbol{f}_j(\boldsymbol{x}_n)]^\top$ be the $n\times p_j$ design matrix for the jth marginal model, then we obtain $\boldsymbol{F} = \mathrm{diag}\big(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_m\big)$ which represents the stacked $mn\times p$ design matrix for the MGLM. As a result the stacked vector of linear predictors is given by $\boldsymbol{H} = [\boldsymbol{\eta}_1^\top,\ldots,\boldsymbol{\eta}_m^\top]^\top = \boldsymbol{F}\boldsymbol{\beta}$, where $\boldsymbol{\eta}_j = \big(\eta_j(\boldsymbol{x}_1,\boldsymbol{\beta}_j),\ldots,\eta_j(\boldsymbol{x}_n,\boldsymbol{\beta}_j)\big)^\top$, $j=1,\ldots,m$.

For notational simplicity let $Y_{ji} = Y_j(\boldsymbol{x}_i)$ denote the ith observation of the jth component at the point $\boldsymbol{x}_i$ and μ_{ji} denote the value of the jth marginal expected mean at the point $\boldsymbol{x}_i$, i.e., $\mu_{ji} = \mu_j(\boldsymbol{x}_i,\boldsymbol{\beta}_j)$ with $\eta_{ji} = \eta_j(\boldsymbol{x}_i,\boldsymbol{\beta}_j)$. Then define the $n\times n$ diagonal matrices $\boldsymbol{D}_j = \mathrm{diag}\big(\mathrm{var}(Y_{ji})\big)_{i=1}^{n}$ and $\boldsymbol{E}_j = \mathrm{diag}\left(\left(\frac{\partial\mu_{ji}}{\partial\eta_{ji}}\right)^2\right)_{i=1}^{n}$, $j=1,\ldots,m$. Then we obtain the $mn\times mn$ matrices $\boldsymbol{D} = \mathrm{diag}\big(\boldsymbol{D}_j\big)_{j=1}^{m}$ and $\boldsymbol{E} = \mathrm{diag}\big(\boldsymbol{E}_j\big)_{j=1}^{m}$. It can be seen that $\boldsymbol{D}_j^{-1}\boldsymbol{E}_j = \mathrm{diag}\big(u_j(\boldsymbol{x}_i,\boldsymbol{\beta}_j)\big)_{i=1}^{n}$, $j=1,\ldots,m$.

By the Kronecker product "$\otimes$" the $mn\times mn$ variance-covariance matrix of $\boldsymbol{Y}$ is obtained by

$$\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{D}^{\frac{1}{2}}(\boldsymbol{R}\otimes\boldsymbol{I}_n)\boldsymbol{D}^{\frac{1}{2}} = \begin{pmatrix} \rho_{11}\boldsymbol{D}_1 & \rho_{12}\boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{D}_2^{\frac{1}{2}} & \ldots & \rho_{1m}\boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{D}_m^{\frac{1}{2}} \\ \rho_{21}\boldsymbol{D}_2^{\frac{1}{2}}\boldsymbol{D}_1^{\frac{1}{2}} & \rho_{22}\boldsymbol{D}_2 & \ldots & \rho_{2m}\boldsymbol{D}_2^{\frac{1}{2}}\boldsymbol{D}_m^{\frac{1}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1}\boldsymbol{D}_m^{\frac{1}{2}}\boldsymbol{D}_1^{\frac{1}{2}} & \rho_{m2}\boldsymbol{D}_m^{\frac{1}{2}}\boldsymbol{D}_2^{\frac{1}{2}} & \ldots & \rho_{mm}\boldsymbol{D}_m \end{pmatrix}$$

where $\boldsymbol{I}_n$ is an $n\times n$ identity matrix. The overall $mn\times mn$ weight matrix $\boldsymbol{W}$ is defined as $\boldsymbol{W} = \boldsymbol{E}^{\frac{1}{2}}\Big(\mathrm{Cov}(\boldsymbol{Y})\Big)^{-1}\boldsymbol{E}^{\frac{1}{2}} = \boldsymbol{E}^{\frac{1}{2}}\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{R}\otimes\boldsymbol{I}_n)^{-1}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\frac{1}{2}}$. Hence, the information matrix (5.3) can be represented in the form

$$\boldsymbol{M}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{\beta}) = \boldsymbol{F}^\top\boldsymbol{W}\boldsymbol{F}.$$

Lemma 5.1.1. *Consider the MGLM (5.1) and the whole experiment $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$. Let $\boldsymbol{F}_j = [\boldsymbol{f}_j(\boldsymbol{x}_1),\dots,\boldsymbol{f}_j(\boldsymbol{x}_n)]^\top$. For a given parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top,\dots,\boldsymbol{\beta}_m^\top)^\top$ define $\boldsymbol{F}_{j,\beta_j} = \boldsymbol{D}_j^{-\frac{1}{2}}\boldsymbol{E}_j^{\frac{1}{2}}\boldsymbol{F}_j = [\boldsymbol{f}_{j,\beta_j}(\boldsymbol{x}_1),\dots,\boldsymbol{f}_{j,\beta_j}(\boldsymbol{x}_n)]^\top$ for all $j=1,\dots,m$ and denote $\boldsymbol{F}_\beta = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\frac{1}{2}}\boldsymbol{F} = \mathrm{diag}\Big(\boldsymbol{F}_{1,\beta_1},\dots,\boldsymbol{F}_{m,\beta_m}\Big)$. Then the information matrix (5.3) has the form*

$$\boldsymbol{M}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{\beta}) = \boldsymbol{F}_\beta^\top\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{I}_n\Big)\boldsymbol{F}_\beta. \tag{5.5}$$

Proof. Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top,\dots,\boldsymbol{\beta}_m^\top)^\top$ be given. Straightforward steps imply that

$$\begin{aligned}\boldsymbol{M}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{\beta}) &= \boldsymbol{F}^\top\boldsymbol{W}\boldsymbol{F} = \boldsymbol{F}^\top\boldsymbol{E}^{\frac{1}{2}}\Big(\mathrm{Cov}(\boldsymbol{Y})\Big)^{-1}\boldsymbol{E}^{\frac{1}{2}}\boldsymbol{F}\\ &= \boldsymbol{F}^\top\boldsymbol{E}^{\frac{1}{2}}\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{R}\otimes\boldsymbol{I}_n)^{-1}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\frac{1}{2}}\boldsymbol{F}\\ &= \boldsymbol{F}_\beta^\top\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{I}_n\Big)\boldsymbol{F}_\beta.\end{aligned}$$

□

The multivariate version of The Equivalence Theorem (Theorem 2.2.2, part (a), part (b)) for checking the D- and A-optimality of a given design (see Fedorov, Gagnon, and Leonov (2002)) can be used. Denote by $\mathrm{tr}(\boldsymbol{A})$ the trace of a $p\times p$ matrix $\boldsymbol{A}$.

Theorem 5.1.1. *Let $\boldsymbol{\beta}$ be a given parameter point and let ξ^* be a design with nonsingular information matrix $\boldsymbol{M}(\xi^*,\boldsymbol{\beta})$.*

- *A design ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for the MGLM if and only if*

$$\mathrm{tr}\Big(\boldsymbol{R}^{-1}\mathbf{f}_\beta^\top(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\mathbf{f}_\beta(\boldsymbol{x})\Big) \le p \;\; \forall\boldsymbol{x}\in\mathcal{X} \tag{5.6}$$

- *A design ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) for the MGLM if and only if*

$$\mathrm{tr}\Big(\boldsymbol{R}^{-1}\mathbf{f}_\beta^\top(\boldsymbol{x})\boldsymbol{M}^{-2}(\xi^*,\boldsymbol{\beta})\mathbf{f}_\beta(\boldsymbol{x})\Big) \le \mathrm{tr}\Big(\boldsymbol{M}^{-1}(\xi^*,\boldsymbol{\beta})\Big) \;\; \forall\boldsymbol{x}\in\mathcal{X} \tag{5.7}$$

where at the support points of ξ^* both inequalities (5.6) and (5.7) are equations.

5.2 Reduction to univariate models

The locally optimal design for a MGLM is derived at a given parameter point $\boldsymbol{\beta}$ under known correlation matrix $\boldsymbol{R}$. Throughout we focus on approximate designs ξ with support $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_r\} \subseteq \mathcal{X}$ defined by (2.9). The information matrix (5.3) of ξ is given by

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) = \sum_{i=1}^{r} \omega_i \boldsymbol{M}(\boldsymbol{x}_i,\boldsymbol{\beta}). \tag{5.8}$$

In this context, in each jth component the functions $\boldsymbol{f}_j(\boldsymbol{x}_1),\dots,\boldsymbol{f}_j(\boldsymbol{x}_r)$ are linearly independent which constitute the $r \times p_j$ design matrix $\boldsymbol{F}_j = [\boldsymbol{f}_j(\boldsymbol{x}_1),\dots,\boldsymbol{f}_j(\boldsymbol{x}_r)]^\top$. Here, $Y_{ji} = Y_j(\boldsymbol{x}_i)$, $\mu_{ji} = \mu_j(\boldsymbol{x}_i,\boldsymbol{\beta}_j)$ and $\eta_{ji} = \eta_j(\boldsymbol{x}_i,\boldsymbol{\beta}_j)$, $i = 1,\dots,r$. Thus we write $\boldsymbol{D}_j = \operatorname{diag}\Big(\operatorname{var}(Y_{ji})\Big)_{i=1}^{r}$ and $\boldsymbol{E}_j = \operatorname{diag}\left(\left(\frac{\partial\mu_{ji}}{\partial\eta_{ji}}\right)^2\right)_{i=1}^{r}$. For the MGLM we get the $mr \times p$ design matrix $\boldsymbol{F} = \operatorname{diag}\Big(\boldsymbol{F}_1,\dots,\boldsymbol{F}_m\Big)$ with the $mr \times mr$ matrices $\boldsymbol{D} = \operatorname{diag}\Big(\boldsymbol{D}_j\Big)_{j=1}^{m}$ and $\boldsymbol{E} = \operatorname{diag}\Big(\boldsymbol{E}_j\Big)_{j=1}^{m}$. Let $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1,\dots,\omega_r)$ be a diagonal matrix of the design weights. Denote $\boldsymbol{F}_{j,\boldsymbol{\beta}_j} = \boldsymbol{D}_j^{-\frac{1}{2}}\boldsymbol{E}_j^{\frac{1}{2}}\boldsymbol{F}_j = [\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_1),\dots,\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_r)]^\top$ with $\boldsymbol{F}_{\boldsymbol{\beta}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\frac{1}{2}}\boldsymbol{F} = \operatorname{diag}\Big(\boldsymbol{F}_{1,\boldsymbol{\beta}_1},\dots,\boldsymbol{F}_{m,\boldsymbol{\beta}_m}\Big)$ then by Lemma 5.1.1 the information matrix (5.8) of a design ξ rewrites

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) = \boldsymbol{F}_{\boldsymbol{\beta}}^\top\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{\Omega}\Big)\boldsymbol{F}_{\boldsymbol{\beta}}. \tag{5.9}$$

Furthermore, $\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta}) = \Big(\boldsymbol{F}_{\boldsymbol{\beta}}^\top\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{\Omega}\Big)\boldsymbol{F}_{\boldsymbol{\beta}}\Big)^{-1}$, which factorizes if $\boldsymbol{F}_{\boldsymbol{\beta}}$ is square, i.e.,

$$\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta}) = \boldsymbol{F}_{\boldsymbol{\beta}}^{-1}\Big(\boldsymbol{R}\otimes\boldsymbol{\Omega}^{-1}\Big)\Big(\boldsymbol{F}_{\boldsymbol{\beta}}^\top\Big)^{-1}.$$

In general, a block representation of the information matrix (5.8) is of the form

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) = \Big(\rho^{(jh)}\boldsymbol{M}_{jh}(\xi,\boldsymbol{\beta}_j,\boldsymbol{\beta}_h)\Big)_{j=1,\dots,m}^{h=1,\dots,m} \tag{5.10}$$

where $\boldsymbol{M}_j(\xi,\boldsymbol{\beta}_j) = \boldsymbol{M}_{jj}(\xi,\boldsymbol{\beta}_j) = \boldsymbol{F}_{j,\boldsymbol{\beta}_j}^\top\boldsymbol{\Omega}\boldsymbol{F}_{j,\boldsymbol{\beta}_j} = \sum_{i=1}^{r}\omega_i\boldsymbol{f}_{j,\boldsymbol{\beta}_j}\Big(\boldsymbol{x}_i\Big)\boldsymbol{f}_{j,\boldsymbol{\beta}_j}^\top\Big(\boldsymbol{x}_i\Big)$ is the $p_j \times p_j$ information matrix for the jth marginal model $(1 \le j \le m)$, whereas the $p_j \times p_h$ submatrices $\boldsymbol{M}_{jh}(\xi,\boldsymbol{\beta}_j,\boldsymbol{\beta}_h) = \boldsymbol{F}_{j,\boldsymbol{\beta}_j}^\top\boldsymbol{\Omega}\boldsymbol{F}_{h,\boldsymbol{\beta}_h} = \sum_{i=1}^{r}\omega_i\boldsymbol{f}_{j,\boldsymbol{\beta}_j}\Big(\boldsymbol{x}_i\Big)\boldsymbol{f}_{h,\boldsymbol{\beta}_h}^\top\Big(\boldsymbol{x}_i\Big)$ $(1 \le j \ne h \le m)$ which are not necessarily square.

Remark 5.2.1. *All the submatrices $\boldsymbol{M}_j(\xi,\boldsymbol{\beta}_j)$, $j = 1,\dots,m$ are nonsingular if for any design ξ with r design points we have $r \ge \max_{(1\le j\le m)} p_j$.*

Lemma 5.2.1. *Consider a design ξ defined on $\mathcal{X}$ with information matrix (5.10). Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_m^\top)^\top$ be given. Assume that all submatrices $\boldsymbol{M}_{jh}(\xi, \boldsymbol{\beta}_j, \boldsymbol{\beta}_h)$, $j, h = 1, \dots, m$ are square, i.e., $p_1 = \cdots = p_m = p_0$ and nonsingular where the parameters for each component are not necessarily equal. If $\sum_{k=1}^{m} \rho_{jk}\rho^{(hk)}\boldsymbol{M}_{jk}(\xi, \boldsymbol{\beta}_j, \boldsymbol{\beta}_k)\boldsymbol{M}_{hk}^{-1}(\xi, \boldsymbol{\beta}_h, \boldsymbol{\beta}_k) = 0$ for all $(1 \le j \ne h \le m)$ then $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ given by (5.10) is nonsingular and a block representation of its inverse is given by*

$$\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}) = \left(\rho_{hj}\boldsymbol{M}_{hj}^{-1}(\xi, \boldsymbol{\beta}_h, \boldsymbol{\beta}_j)\right)_{j=1,\dots,m}^{h=1,\dots,m} \tag{5.11}$$

Proof. The proposed assumption $\sum_{k=1}^{m} \rho_{jk}\rho^{(hk)}\boldsymbol{M}_{jk}(\xi, \boldsymbol{\beta}_j, \boldsymbol{\beta}_k)\boldsymbol{M}_{hk}^{-1}(\xi, \boldsymbol{\beta}_h, \boldsymbol{\beta}_k) = 0$ for all $(1 \le j \ne h \le m)$ is explicitly describes the multiplication of the off-diagonal submatrices and therefore, $\boldsymbol{M}(\xi, \boldsymbol{\beta})\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})$ is an identity matrix. □

Corollary 5.2.1. *Let a design ξ on $\mathcal{X}$ be given for a bivariate GLM ($m = 2$) such that $p_1 = p_2 = p_0$. Given the correlation matrix $\boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\boldsymbol{R}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$. Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ be given. Then we have $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \frac{1}{1-\rho^2}\begin{pmatrix} \boldsymbol{M}_1(\xi, \boldsymbol{\beta}_1) & -\rho\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta}) \\ -\rho\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta}) & \boldsymbol{M}_2(\xi, \boldsymbol{\beta}_2) \end{pmatrix}$ where $\boldsymbol{M}_j(\xi, \boldsymbol{\beta}_j) = \boldsymbol{F}_{j,\beta_j}^\top \boldsymbol{\Omega} \boldsymbol{F}_{j,\beta_j}, j = 1, 2$ and $\boldsymbol{M}_{jh}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}_{j,\beta_j}^\top \boldsymbol{\Omega} \boldsymbol{F}_{h,\beta_h} (1 \le j \ne h \le 2)$. If $\boldsymbol{M}_j(\xi, \boldsymbol{\beta}_j), j = 1, 2$, $\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta})$ and $\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta})$ are nonsingular then the inverse of $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ (see Lu and Shiou (2002) for inverses of 2×2 block matrices) is denoted by $\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}) = (1 - \rho^2)\begin{pmatrix} \boldsymbol{m}_{11} & \boldsymbol{m}_{12} \\ \boldsymbol{m}_{21} & \boldsymbol{m}_{22} \end{pmatrix}$ where*

$$\begin{aligned}
\boldsymbol{m}_{11} &= \left(\boldsymbol{M}_1(\xi, \boldsymbol{\beta}_1) - \rho^2\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2)\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta})\right)^{-1}, \\
\boldsymbol{m}_{12} &= \rho\,\boldsymbol{m}_{11}\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2), \\
\boldsymbol{m}_{22} &= \boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2) + \rho^2\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2)\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta})\boldsymbol{m}_{11}\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2),
\end{aligned}$$

Moreover, the elements $\boldsymbol{m}_{11}, \boldsymbol{m}_{12}, \boldsymbol{m}_{21}, \boldsymbol{m}_{22}$ can be determined as in the following.

(i) $\boldsymbol{m}_{11} = \frac{1}{1-\rho^2}\boldsymbol{M}_1^{-1}(\xi, \boldsymbol{\beta}_1)$ *and* $\boldsymbol{m}_{12} = \frac{\rho}{1-\rho^2}\boldsymbol{M}_{21}^{-1}(\xi, \boldsymbol{\beta})$ *if*

$$\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2)\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta}) = \boldsymbol{M}_1(\xi, \boldsymbol{\beta}_1).$$

(ii) $\boldsymbol{m}_{22} = \frac{1}{1-\rho^2}\boldsymbol{M}_2^{-1}(\xi, \boldsymbol{\beta}_2)$ *and* $\boldsymbol{m}_{21} = \frac{\rho}{1-\rho^2}\boldsymbol{M}_{12}^{-1}(\xi, \boldsymbol{\beta})$ *if*

$$\boldsymbol{M}_{21}(\xi, \boldsymbol{\beta})\boldsymbol{M}_1^{-1}(\xi, \boldsymbol{\beta}_1)\boldsymbol{M}_{12}(\xi, \boldsymbol{\beta}) = \boldsymbol{M}_2(\xi, \boldsymbol{\beta}_1).$$

and hence under the above assumptions we get

$$\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{M}_1^{-1}(\xi,\boldsymbol{\beta}_1) & \rho\boldsymbol{M}_{21}^{-1}(\xi,\boldsymbol{\beta}) \\ \rho\boldsymbol{M}_{12}^{-1}(\xi,\boldsymbol{\beta}) & \boldsymbol{M}_2^{-1}(\xi,\boldsymbol{\beta}_2) \end{pmatrix}$$

Proof. Under assumptions (i) and (ii), straightforward multiplication yields $\boldsymbol{M}(\xi,\boldsymbol{\beta})\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta}) = \boldsymbol{I}_p$. □

The previous situation can be simplified under saturated designs. Let Ξ_{p_0} denotes the set of all saturated designs under each jth univariate GLM ($1 \le j \le m$), i.e. $\Xi_{p_0} = \{\xi : \text{supp}(\xi) \subseteq \mathcal{X}, r = p_0\}$. Clearly, for any design $\xi \in \Xi_{p_0}$ the design matrix $\boldsymbol{F}$ (or $\boldsymbol{F}_\beta$) of the MGLM is square. In particular, under Corollary 5.2.1 with $\xi \in \Xi_{p_0}$ both matrices $\boldsymbol{F}_{\beta_1}$ and $\boldsymbol{F}_{\beta_2}$ are square and nonsingular. Hence, the submatrices $\boldsymbol{M}_j(\xi,\boldsymbol{\beta}_j), j = 1,2$, $\boldsymbol{M}_{jh}(\xi,\boldsymbol{\beta})$ ($1 \le j \ne h \le 2$) factorize so that the assumptions (i) and (ii) in Corollary 5.2.1 will be implicitly satisfied as it is given by the next corollary.

Corollary 5.2.2. *Consider the notations presented in Corollary 5.2.1. Let $\xi \in \Xi_{p_0}$. Then the assumptions (i) and (ii) given in Corollary 5.2.1 are satisfied. That is*

(i) $\boldsymbol{M}_{12}(\xi,\boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi,\boldsymbol{\beta}_2)\boldsymbol{M}_{21}(\xi,\boldsymbol{\beta}) = \boldsymbol{M}_1(\xi,\boldsymbol{\beta}_1)$

(ii) $\boldsymbol{M}_{21}(\xi,\boldsymbol{\beta})\boldsymbol{M}_1^{-1}(\xi,\boldsymbol{\beta}_1)\boldsymbol{M}_{12}(\xi,\boldsymbol{\beta}) = \boldsymbol{M}_2(\xi,\boldsymbol{\beta}_1)$.

Proof. Note that $\boldsymbol{M}_j(\xi,\boldsymbol{\beta}_j) = \boldsymbol{F}_{j,\beta_j}^\top \boldsymbol{\Omega}\boldsymbol{F}_{j,\beta_j}, j = 1,2$ and $\boldsymbol{M}_{jh}(\xi,\boldsymbol{\beta}) = \boldsymbol{F}_{j,\beta_j}^\top \boldsymbol{\Omega}\boldsymbol{F}_{h,\beta_h}$ ($1 \le j \ne h \le 2$). Then

(i) $\boldsymbol{M}_{12}(\xi,\boldsymbol{\beta})\boldsymbol{M}_2^{-1}(\xi,\boldsymbol{\beta}_2)\boldsymbol{M}_{21}(\xi,\boldsymbol{\beta}) = \boldsymbol{F}_{1,\beta_1}^\top\boldsymbol{\Omega}\boldsymbol{F}_{2,\beta_2}\boldsymbol{F}_{2,\beta_2}^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{F}_{2,\beta_2}^\top)^{-1}\boldsymbol{F}_{2,\beta_2}^\top\boldsymbol{\Omega}\boldsymbol{F}_{1,\beta_1} = \boldsymbol{F}_{1,\beta_1}^\top\boldsymbol{\Omega}\boldsymbol{F}_{1,\beta_1}$

(ii) $\boldsymbol{M}_{21}(\xi,\boldsymbol{\beta})\boldsymbol{M}_1^{-1}(\xi,\boldsymbol{\beta}_1)\boldsymbol{M}_{12}(\xi,\boldsymbol{\beta}) = \boldsymbol{F}_{2,\beta_2}^\top\boldsymbol{\Omega}\boldsymbol{F}_{1,\beta_1}\boldsymbol{F}_{1,\beta_1}^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{F}_{1,\beta_1}^\top)^{-1}\boldsymbol{F}_{1,\beta_1}^\top\boldsymbol{\Omega}\boldsymbol{F}_{2,\beta_2} = \boldsymbol{F}_{2,\beta_2}^\top\boldsymbol{\Omega}\boldsymbol{F}_{2,\beta_2}$. □

Lemma 5.2.2. *The locally D-optimal design $\xi^* \in \Xi_{p_0}$ at a given parameter point $\boldsymbol{\beta}$ for a MGLM (5.1) is independent of correlation matrix $\boldsymbol{R}$.*

Proof. The determinant of the information matrix $\boldsymbol{M}(\xi^*,\boldsymbol{\beta})$ from (5.9) is given by

$$\begin{aligned} \det \boldsymbol{M}(\xi^*,\boldsymbol{\beta}) &= \det \boldsymbol{F}_\beta^\top\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{\Omega}\Big)\boldsymbol{F}_\beta \\ &= \det\Big(\boldsymbol{F}_\beta^\top\boldsymbol{F}_\beta\Big)\det\Big(\boldsymbol{R}^{-1}\otimes\boldsymbol{\Omega}\Big) \\ &= \det\Big(\boldsymbol{F}_\beta^\top\boldsymbol{F}_\beta\Big)\Big(\det\boldsymbol{\Omega}\Big)^m\Big(\det\boldsymbol{R}^{-1}\Big)^r \end{aligned}$$

where $\Big(\det\boldsymbol{\Omega}\Big)^m = \Big(\prod_{i=1}^r \omega_i\Big)^m = p_0^{-rm}$ since ξ^* is saturated. It follows that $\det \boldsymbol{M}(\xi^*,\boldsymbol{\beta})$ is proportional to $\det\Big(\boldsymbol{F}_\beta^\top\boldsymbol{F}_\beta\Big)$. Thus the design ξ^* is D-optimal if and only if it maximizes $\det\Big(\boldsymbol{F}_\beta^\top\boldsymbol{F}_\beta\Big)$ and hence, ξ^* is independent of $\boldsymbol{R}$. □

Theorem 5.2.1. *Given a parameter point* $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$. *Let the design* $\xi^* \in \Xi_{p_0}$ *be locally D-optimal (at* $\boldsymbol{\beta}_j$ *) for each* j*th marginal model* $(1 \le j \le m)$*. Then* ξ^* *is locally D-optimal (at* $\boldsymbol{\beta}$*) for the MGLM (5.1) in the set* Ξ_{p_0}*.*

Proof. From the proof of Lemma 5.2.2 we have

$$\begin{aligned}
\det \boldsymbol{M}(\xi^*, \boldsymbol{\beta}) &= \det \Big(\boldsymbol{F}_\beta^\top \boldsymbol{F}_\beta\Big)\Big(\det \boldsymbol{\Omega}\Big)^m \Big(\det \boldsymbol{R}^{-1}\Big)^r \\
&= p_0^{-rm} (\det \boldsymbol{F})^2 \Big(\det \boldsymbol{E}^{\frac{1}{2}} \boldsymbol{D}^{-\frac{1}{2}}\Big)^2 \Big(\det \boldsymbol{R}^{-1}\Big)^r \\
&= p_0^{-rm} (\det \boldsymbol{F})^2 \Big(\prod_{j=1}^m \prod_{i=1}^r u_j(\boldsymbol{x}_i, \boldsymbol{\beta}_j)\Big)\Big(\det \boldsymbol{R}^{-1}\Big)^r
\end{aligned}$$

where $\det \boldsymbol{F} = \prod_{j=1}^m \det \boldsymbol{F}_j$. Moreover, the determinant of the information matrix for the jth marginal models is $\det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j) = p_0^{-r} (\det \boldsymbol{F}_j)^2 \prod_{i=1}^r u_j(\boldsymbol{x}_i, \boldsymbol{\beta}_j)$, $j = 1, \ldots, m$. Thus $\prod_{i=1}^r u_j(\boldsymbol{x}_i, \boldsymbol{\beta}_j) = p_0^r \left(\det \boldsymbol{F}_j\right)^{-2} \det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j)$. It follows that

$$\begin{aligned}
\det \boldsymbol{M}(\xi^*, \boldsymbol{\beta}) &= p_0^{-rm} \Big(\prod_{j=1}^m \det \boldsymbol{F}_j\Big)^2 \Big(\prod_{j=1}^m p_0^r \left(\det \boldsymbol{F}_j\right)^{-2} \det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j)\Big)\Big(\det \boldsymbol{R}^{-1}\Big)^r \\
&= p_0^{-rm} p_0^{rm} \Big(\prod_{j=1}^m \det \boldsymbol{F}_j\Big)^2 \Big(\prod_{j=1}^m \det \boldsymbol{F}_j\Big)^{-2} \prod_{j=1}^m \det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j)\Big(\det \boldsymbol{R}^{-1}\Big)^r \\
&= \Big(\det \boldsymbol{R}^{-1}\Big)^r \prod_{j=1}^m \det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j)
\end{aligned}$$

Since ξ^* is locally D-optimal for the jth marginal model it maximizes $\det \boldsymbol{M}_j(\xi, \boldsymbol{\beta}_j)$ on Ξ_{p_0}. Thus $\prod_{j=1}^m \det \boldsymbol{M}_j(\xi^*, \boldsymbol{\beta}_j) \ge \prod_{j=1}^m \det \boldsymbol{M}_j(\xi, \boldsymbol{\beta}_j)$ for all $\xi \in \Xi_{p_0}$. As a result, ξ^* maximizes $\det \boldsymbol{M}(\xi, \boldsymbol{\beta})$ on Ξ_{p_0}. Hence, $\xi^* \in \Xi_{p_0}$ is locally D-optimal (at $\boldsymbol{\beta}$) for a MGLM. □

Next we will deal with the local A-optimality and the following lemma is immediate.

Lemma 5.2.3. *The locally A-optimal design* $\xi^* \in \Xi_{p_0}$ *at a given parameter point* $\boldsymbol{\beta}$ *for a MGLM (5.1) is independent of correlation matrix* $\boldsymbol{R}$*.*

Proof. The inverse of the information matrix of $\xi^* \in \Xi_{p_0}$ is given by $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}) = \boldsymbol{F}_\beta^{-1}\Big(\boldsymbol{R} \otimes \boldsymbol{\Omega}^{-1}\Big)\Big(\boldsymbol{F}_\beta^\top\Big)^{-1}$ which has the block representation (5.11). Thus $\mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})) = \sum_{j=1}^m \mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi^*, \boldsymbol{\beta}_j))$. It is clear that $\mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}))$ does not depend on $\boldsymbol{R}$. □

Theorem 5.2.2. *Given a parameter point* $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$. *Let the design* $\xi^* \in \Xi_{p_0}$ *be locally A-optimal (at* $\boldsymbol{\beta}_j$ *) for each* j*th marginal model* $(1 \le j \le m)$*. Then* ξ^* *is locally A-optimal (at* $\boldsymbol{\beta}$*) for the MGLM (5.1) in the set* Ξ_{p_0}*.*

Proof. For the design $\xi^* \in \Xi_{p_0}$ we have $\mathrm{tr}(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})) = \sum_{j=1}^{m} \mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi^*, \boldsymbol{\beta}_j))$. As ξ^* is locally A-optimal for the jth marginal model then $\mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi^*, \boldsymbol{\beta}_j)) \leq \mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi, \boldsymbol{\beta}_j))$ for all $\xi \in \Xi_{p_0}$. Thus $\sum_{j=1}^{m} \mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi^*, \boldsymbol{\beta}_j)) \leq \sum_{j=1}^{m} \mathrm{tr}(\boldsymbol{M}_j^{-1}(\xi, \boldsymbol{\beta}_j))$ for all $\xi \in \Xi_{p_0}$. As a result, ξ^* minimizes $\mathrm{tr}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}))$ on Ξ_{p_0}. Hence, ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) for a MGLM. □

Remark 5.2.2. *It is well known from Lemma 3.1.1 in Section 3.1 that the optimal weights of a locally A-optimal design under a univariate generalized linear model depends on the model parameters through the intensity functions. Therefore, the locally A-optimal design for marginal univariate models is A-optimal for the MGLM if all intensities in all components have the same form and the A-optimality is derived at equal parameter points. This guarantees that the A-optimal design is the same for all marginal models.*

In the following corollary we provide a solution of locally optimal designs for a multivariate GLM under identical components.

Corollary 5.2.3. *Consider the MGLM (5.1) such that $\boldsymbol{f}_1(\boldsymbol{x}) = \cdots = \boldsymbol{f}_m(\boldsymbol{x}) = \boldsymbol{f}_0(\boldsymbol{x})$. Let the parameter point $\boldsymbol{\beta}$ be given such that $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_m = \boldsymbol{\beta}_0$, i.e., $\boldsymbol{\beta} = \mathbf{1} \otimes \boldsymbol{\beta}_0$. Assume that $u_1(\boldsymbol{x}, \boldsymbol{\beta}_1) = \cdots = u_m(\boldsymbol{x}, \boldsymbol{\beta}_m) = u_0(\boldsymbol{x}, \boldsymbol{\beta}_0)$ for all $\boldsymbol{x} \in \mathcal{X}$. Hence, $\boldsymbol{F}_{1,\boldsymbol{\beta}_1} = \cdots = \boldsymbol{F}_{m,\boldsymbol{\beta}_m} = \boldsymbol{F}_{0,\boldsymbol{\beta}_0}$. Let the design ξ^* be locally D- resp. A-optimal (at $\boldsymbol{\beta}_0$) for each jth marginal model ($1 \leq j \leq m$). Then the design ξ^* is locally D- resp. A-optimal (at $\boldsymbol{\beta} = \mathbf{1} \otimes \boldsymbol{\beta}_0$) for the MGLM (5.1).*

Proof. The proposed assumptions in the corollary implies that the information matrix of a design ξ and its inverse factorize as in the following

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}_0) = \boldsymbol{R}^{-1} \otimes \boldsymbol{M}(\xi, \boldsymbol{\beta}_0) \text{ and thus } \boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}_0) = \boldsymbol{R} \otimes \boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}_0).$$

Therefore, we obtain

$$\det \boldsymbol{M}(\xi, \boldsymbol{\beta}_0) = \left(\det \boldsymbol{R}^{-1}\right)^{p_0} \left(\det \boldsymbol{M}(\xi, \boldsymbol{\beta}_0)\right)^{m},$$
$$\mathrm{tr}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}_0)) = \mathrm{tr}(\boldsymbol{R})\mathrm{tr}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}_0)).$$

The optimization with respect to D- and A-criteria reduces to the corresponding univariate optimization problem. □

5.3 Seemingly unrelated univariate gamma models

In this section we assume that the MGLM is constituted by seemingly unrelated gamma models. That is the univariate responses Y_j $(1 \le j \le m)$ come from gamma distributions that were determined in Section 4.1. Besides, we allow different factors belong to different experimental regions and different regression functions in the linear predictors that are related to the expected means by power link functions,

$$\eta_j = \mu_j^{k_j} \quad \text{where} \quad \eta_j = \boldsymbol{f}_j^\top(\boldsymbol{x}_j)\boldsymbol{\beta}_j = \sum_{l=1}^{p_j} f_{jl}(\boldsymbol{x}_j)\beta_{jl} \;\; (1 \le j \le m),$$

where k_j denotes the exponent of the power link in the jth model (for notational convenience we use the exponents k_j instead of ρ_j). Here, for the jth component $\boldsymbol{f}_j(\boldsymbol{x}_j)$ is the p_j-dimensional vector of linearly independent known regression functions $f_{j1}, \ldots, f_{jp_j}$ and $\boldsymbol{\beta}_j = \big(\beta_{j1}, \ldots, \beta_{jp_j}\big)^\top \in \mathbb{R}^{p_j}$. Denote by ν_j the number of factors in component j and ν denotes the total number of factors in the MGLM, i.e., $\nu = \sum_{j=1}^m \nu_j$. The point $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{j\nu j})^\top$ may differ across the components of a unit and is chosen from an experimental region $\mathcal{X}_j \subseteq \mathbb{R}^{\nu_j}$. The intensity function in component j is given at $\boldsymbol{x}_j \in \mathcal{X}_j$ as $u_j(\boldsymbol{x}_j, \boldsymbol{\beta}_j) = \big(\boldsymbol{f}_j^\top(\boldsymbol{x}_j)\boldsymbol{\beta}_j\big)^{-2}$.

The experimental region for the multivariate model is given by $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$. Denote $p = \sum_{j=1}^m p_j$. The $p \times m$ block diagonal multivariate regression is given by $\mathbf{f}(\boldsymbol{x}) = \mathrm{diag}\big(\boldsymbol{f}_1(\boldsymbol{x}_1), \ldots, \boldsymbol{f}_m(\boldsymbol{x}_m)\big)$ where $\boldsymbol{x} \in \mathcal{X}$. Note that $\boldsymbol{x}$ is a ν-tuple, i.e., $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_m^\top)^\top = (x_{11}, \ldots, x_{1\nu_1}, \ldots, x_{m1}, \ldots, x_{m\nu_m})^\top$. Let the stacked parameter p-vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ be given. The MGLM with univariate gamma models for each unit at a point $\boldsymbol{x} \in \mathcal{X}$ is defined by

$$\boldsymbol{\eta} = \big(\mu_1^{k_1}, \ldots, \mu_m^{k_m}\big)^\top \quad \text{where} \quad \boldsymbol{\eta} = \mathbf{f}^\top(\boldsymbol{x})\boldsymbol{\beta}.$$

In particular, for multivariate models with seemingly unrelated linear models i.e., $\eta_j = \mu_j$, $j = 1, \ldots, m$ product type designs (see Subsection 4.5.3) were developed by Soumaya, Gaffke, and Schwabe (2015). In the following we will develop analogous results under the MGLM with seemingly unrelated gamma models. To this end for the univariate gamma model in component j we define

$$\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_j) = \big(\boldsymbol{f}_j^\top(\boldsymbol{x}_j)\boldsymbol{\beta}_j\big)^{-1}\boldsymbol{f}_j(\boldsymbol{x}_j), \;\; \boldsymbol{x}_j \in \mathcal{X}_j, \;\; j = 1, \ldots, m. \tag{5.12}$$

The function $\boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_j)$ given above in (5.12) involves implicitly the intercept term, i.e., there exist constant vectors $\boldsymbol{c}_j$ such that $\boldsymbol{c}_j^\top \boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_j) = 1$ for all $\boldsymbol{x}_j \in \mathcal{X}_j$ with $\boldsymbol{c}_j = \boldsymbol{\beta}_j$, $j = 1, \ldots, m$. Therefore, under this model property we are able to develop a product structure of a locally optimal design for our MGLM.

In analogy to Lemma 4.1, Lemma 4.2 in Soumaya, Gaffke, and Schwabe (2015) the information matrix of a product type design for the MGLM with seemingly unrelated gamma models and its inverse are given in the following lemmas where the proof is similar to that in the reference.

Lemma 5.3.1. *Let $\xi = \bigotimes_{j=1}^{m} \xi_j$ be a product-type design on the experimental region $\mathcal{X} = \times_{j=1}^{m} \mathcal{X}_j$. Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ be given. For each marginal design ξ_j denote by $\boldsymbol{M}_j(\xi_j, \boldsymbol{\beta}_j) = \int_{\mathcal{X}_j} \boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_j)\boldsymbol{f}_{j,\boldsymbol{\beta}_j}^\top(\boldsymbol{x}_j)\,\xi_j(\mathrm{d}\boldsymbol{x}_j)$ the information matrix and by $\boldsymbol{m}_j(\xi_j, \boldsymbol{\beta}_j) = \int_{\mathcal{X}_j} \boldsymbol{f}_{j,\boldsymbol{\beta}_j}(\boldsymbol{x}_j)\,\xi_j(\mathrm{d}\boldsymbol{x}_j)$ the moment vector. Then the information matrix of design ξ has the form*

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \mathrm{diag}\Big(\rho^{(jj)}\left(\boldsymbol{M}_j(\xi_j, \boldsymbol{\beta}_j) - \boldsymbol{m}_j(\xi_j, \boldsymbol{\beta}_j)\boldsymbol{m}_j^\top(\xi_j, \boldsymbol{\beta}_j)\right)\Big) + \boldsymbol{m}(\xi, \boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{m}^\top(\xi, \boldsymbol{\beta})$$

where $\boldsymbol{m}(\xi, \boldsymbol{\beta}) = \mathrm{diag}\Big(\boldsymbol{m}_j(\xi_j, \boldsymbol{\beta}_j)\Big)_{j=1}^{m}$ is the block diagonal matrix of the marginal moments.

Lemma 5.3.2. *Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ be given. Let $\xi = \bigotimes_{j=1}^{m} \xi_j$ be a product type design on the experimental region $\mathcal{X} = \times_{j=1}^{m} \mathcal{X}_j$ such that for each j the information matrix $\boldsymbol{M}_j(\xi_j, \boldsymbol{\beta}_j)$ of ξ_j in the jth marginal model is nonsingular. Then the information matrix $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ of ξ in the MGLM with seemingly unrelated gamma models is nonsingular and*

$$\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}) = \mathrm{diag}\left(\frac{1}{\rho^{(jj)}}\left(\boldsymbol{M}_j^{-1}(\xi_j, \boldsymbol{\beta}_j) - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^\top\right)\right)_{j=1}^{m} + \boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^\top$$

where $\boldsymbol{B} = \mathrm{diag}\left(\boldsymbol{\beta}_j\right)_{j=1}^{m}$ is the block diagonal matrix of the parameter vectors $\boldsymbol{\beta}_j$.

Next we provide the locally D- and A-optimal designs in the product type. The proof is derived by the conditions of The Equivalence Theorem given in (5.6) and (5.7) which is analogous to the proof of Theorem 4.1 and Theorem 4.2 in Soumaya, Gaffke, and Schwabe (2015). The proof is obtained on the fact that the sensitivity function in condition (5.6) or (5.7) is the sum of the marginal sensitivity functions in condition (2.11) or (2.12), respectively, under univariate models.

Theorem 5.1. *Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ be given. For each $j = 1, \ldots, m$, let ξ_j^* be a locally D-optimal design (at $\boldsymbol{\beta}_j$) for the jth univariate gamma model on the experimental region $\mathcal{X}_j$. Then the product type design $\xi^* = \bigotimes_{j=1}^{m} \xi_j^*$ is a locally D-optimal design (at $\boldsymbol{\beta}$) for the MGLM with seemingly unrelated gamma models on the experimental region $\mathcal{X} = \times_{j=1}^{m} \mathcal{X}_j$.*

Theorem 5.2. *Let a parameter point $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$ be given. For each $j = 1, \ldots, m$, let ξ_j^* be a locally A-optimal design (at $\boldsymbol{\beta}_j$) for the jth univariate gamma model on the experimental region $\mathcal{X}_j$. Then the product type design $\xi^* = \bigotimes_{j=1}^{m} \xi_j^*$ is a locally A-optimal design (at $\boldsymbol{\beta}$) for the MGLM with seemingly unrelated gamma models on the experimental region $\mathcal{X} = \times_{j=1}^{m} \mathcal{X}_j$.*

Chapter 6

Discussion and Outlook

In this chapter the summary of the obtained results is presented in the first section. In the second section potential extensions of the work and future topics are suggested.

6.1 Summary

Optimal experimental designs aim to find an optimal choice of the experimental settings that achieve the most precise statistical inferences of the estimates of the model parameters. There is a considerable amount of literature for constructing optimal designs for generalized linear models. In the current thesis, Chapter 3, we firstly focused on analytic solutions of the optimal designs for generalized linear models considering various setups of the linear predictor. On that basis locally D-, A- and Φ_k-optimal designs were derived. The results of Chapter 3 cover many works in literature specifically Poisson models, logistic models, probit models, survival models and gamma models. The main results are in the following according to the corresponding sections in Chapter 3.

- In Section 3.1, particular approaches were established to obtain the optimal weights of some structures of designs under D-, A- and Kiefer Φ_k-criteria that appeared throughout the thesis.
- In Section 3.2, for a first order model with a single binary factor the D- and A-optimal designs were derived. In contrast, in case of a continuous single factor in a one-dimensional experimental region, $\mathcal{X} = [0, 1]$, we provided a condition on a design to be supported by only two design points a and b such that $0 \leq a < b \leq 1$.
- In Section 3.3, for a first order model with two binary factors on $\mathcal{X} = \{0, 1\}^2$ as an experimental region the locally D- and A-optimal saturated designs as well as D-optimal four-point designs were derived followed by an extension of particular D- and A-optimal saturated designs for multiple-factor models.

- In Section 3.4, the D-optimal design were derived for models with complete interactions whereas we restricted ourselves to a two-factor model with interaction for deriving A-optimality.
- In Section 3.5, for a model without intercept; optimal saturated designs with respect to Kiefer Φ_k-criteria were obtained on a general experimental region. The result guarantees that any point located at the edges of the given experimental region $\mathcal{X}$ is a support of a Φ_k-optimal design. Of course, D- and A-optimal designs are included in that class.
- In Section 3.6, we imposed specific assumptions to allow optimal designs for models with intercept to be obtained from optimal designs for models without intercept and vice versa.

The generalized linear models for gamma-distributed outcomes were adopted in Chapter 4 of the thesis. These so-called gamma models are appropriate for many real life data from psychology, ecology or medicine. Despite of that, much attention has not been given to gamma models in optimal designs consideration. In the literature geometric approaches were only used. In the present thesis, gamma models with the family of power links were considered on a polytope as an experimental regions, particularly a ν-dimensional hypercube, i.e., $\mathcal{X} = [a, b]^\nu$. We aimed at providing outstanding and novel solutions for optimal designs for gamma models under various setups of linear predictor. On that basis we obtained the following results according to the corresponding sections in Chapter 4.

- In Section 4.2, a locally complete class of designs and a locally essentially complete class of designs were developed based on the Loewner semi-ordering of information matrices. These classes contain only the designs which are supported by the vertices. So that the support points of any design for the relevant gamma models are chosen from set of all vertices of the experimental region.
- Consequently, the solutions of particular optimal designs for gamma models with quantitative (continuous) factors were transferred from those under the generalized linear models in Chapter 3. We provided illustrative analysis for D- and A-optimal designs for gamma models with continuous factors on a hypercube $\mathcal{X} = [a, b]^\nu$ as an experimental region. We started with Section 4.3 considering one factor, i.e., $\nu = 1$, then Subsection 4.4.1 considering two factors, i.e., $\nu = 2$ with extension of D- and A-optimal saturated designs to models with multiple factors, i.e., $\nu \geq 2$.

- Moreover, gamma models having linear predictors without intercept were considered and thus the models are undefined on the origin point. Therefore, experimental regions that do not include the origin are allowed. In Subsection 4.4.2 for models without interactions Φ_k-optimal designs were obtained on experimental region $[0, \infty]^\nu \setminus \{\mathbf{0}\}$. However, finding the optimal designs for a gamma model without intercept on the experimental region $\mathcal{X} = [a, b], 0 < a < b$ is complicated and thus various technical approaches were introduced to solve the optimal designs in Subsection 4.4.2 for models without interaction and in Subsection 4.5.2 for a two-factor model with interaction:
  - By a suitable transformation of gamma models without intercept to models with intercept the optimality results thus were transferred.
  - By means of The General Equivalence Theorem optimality were characterized for multiple regression by a system of polynomial inequalities which were solved analytically or by computer algebra.
- In Subsection 4.5.3, we developed an approach to construct locally optimal designs for gamma models of complete product-type interactions by making use of optimal designs under marginal models. The product type designs were derived with respect to D-, A- and Φ_k-criteria.
- In Section 4.6, A comprehensive discussion of the potential benefits of the derived locally D-optimal designs for gamma models with and without intercept were presented. It showed how the performance of a locally optimal design is affected by the initial parameter values and how misspecified values may lead to a poor performance of the locally optimal design.

In Chapter 5 we developed locally optimal designs for multivariate generalized linear models.

- A reduction of design problems for MGLM to the univariate GLM was provided. The locally D- and A-optimal saturated designs are independent of the correlation coefficients. If the saturated design is optimal for the univariate GLMs then it is also optimal for its multivariate extension with respect to D- and A-criteria in the set of all saturated designs. The marginal models are not necessary similar, for instance; a gamma model and a Poisson model may be adopted as a bivariate GLM.
- Locally optimal product-type designs with respect to D- and A-criteria were derived for MGLM when the univariate models are seemingly unrelated gamma models. If designs are optimal under the corresponding marginal gamma models then their product is optimal under the multivariate structure of the marginal models.

6.2 Further topics and extensions

Some of our result might be applicable under another nonlinear models which have similar structure of the information matrix (2.1). A nonlinear model is given by

$$Y = h(\boldsymbol{x}, \boldsymbol{\beta}) + \varepsilon \quad \text{where} \quad \varepsilon \text{ is the error term.} \tag{6.1}$$

In this context $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ is defined as the gradient vector of $h(\boldsymbol{x}, \boldsymbol{\beta})$, i.e.,

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \nabla h(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\partial h(\boldsymbol{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \Big(\frac{\partial h(\boldsymbol{x}, \boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial h(\boldsymbol{x}, \boldsymbol{\beta})}{\partial \beta_p}\Big)^{\top}. \tag{6.2}$$

Actually, nonlinear models of form (6.1) were discussed carefully in the literature (see Ford, Titterington, and Kitsos (1989), Atkinson and Haines (1996)). In Dette et al. (2008) some dose–response nonlinear models with intercept were listed, e.g., exponential models and $E_{\max}$ model. Here, a nonlinear model includes explicitly an intercept term if the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ includes the constant 1 (see Schwabe (1995), Li and Balakrishnan (2011), Rodríguez, Ortiz, and Martínez (2015), He (2018)).

In the situation of Section 3.6 extended results can be obtained for nonlinear models of form (6.1). Here, the information matrix of ξ on $\widetilde{\mathcal{X}}$ under non-intercept model reads as

$$\tilde{\boldsymbol{M}}(\xi, \tilde{\boldsymbol{\beta}}) = \int_{\widetilde{\mathcal{X}}} \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}^{\top}(\boldsymbol{x})\, \xi(\mathrm{d}\boldsymbol{x}),$$

while the information matrix of ξ on $\mathcal{X}$ under model with intercept is

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \int_{\mathcal{X}} \Big(1, \boldsymbol{f}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{x})\Big)^{\top} \Big(1, \boldsymbol{f}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{x})\Big)\, \xi(\mathrm{d}\boldsymbol{x}).$$

Corollary 6.2.1. *Let the design ξ^* be defined on $\mathcal{X}$ such that $\boldsymbol{0} \in \operatorname{supp}(\xi^*)$. Let the design $\xi^*_{-\boldsymbol{0}}$ on $\widetilde{\mathcal{X}}$ be the conditional measure of ξ^* given $\boldsymbol{x} \neq \boldsymbol{0}$ such that $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$. Given a parameter point $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}}^{\top})^{\top}$ such that $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$ with $\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{0}) = \boldsymbol{0}$. Then assume there exists a constant vector $\boldsymbol{c}$ such that $\boldsymbol{c}^{\top} \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \operatorname{supp}(\xi^*) \setminus \{\boldsymbol{0}\}$. Let $\xi^* = (1/(\nu+1))\, \xi_{\boldsymbol{0}} + (\nu/(\nu+1))\, \xi^*_{-\boldsymbol{0}}$. Then*
(1) If ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for model with intercept then $\xi^*_{-\boldsymbol{0}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for the corresponding model without intercept.
(2) If $\xi^*_{-\boldsymbol{0}}$ is locally D-optimal (at $\tilde{\boldsymbol{\beta}}$) for model without intercept and

$$\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}^{\top}(\boldsymbol{x}) \tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}}, \tilde{\boldsymbol{\beta}}) \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) \leq \nu\Big(1 - (\boldsymbol{c}^{\top} \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{x}) - 1)^2\Big)\ \forall \boldsymbol{x} \in \mathcal{X} \tag{6.3}$$

then ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) for the corresponding model with intercept.

Corollary 6.2.2. *Under assumptions and notations of Corollary 6.2.1 Let*

$$\xi^* = \left(\frac{\sqrt{\boldsymbol{c}^\top \boldsymbol{c}+1}}{\sqrt{\boldsymbol{c}^\top \boldsymbol{c}+1}+\sqrt{\widetilde{\tau}}}\right)\xi_{\boldsymbol{0}} + \left(\frac{\sqrt{\widetilde{\tau}}}{\sqrt{\boldsymbol{c}^\top \boldsymbol{c}+1}+\sqrt{\widetilde{\tau}}}\right)\xi^*_{-\boldsymbol{0}}.$$

Denote the following equations

$$\begin{aligned} T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) &= (\sqrt{\boldsymbol{c}^\top \boldsymbol{c}+1}+\sqrt{\widetilde{\tau}})^2(\boldsymbol{c}^\top \tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})-1)^2 \\ &+ \frac{(\sqrt{\boldsymbol{c}^\top \boldsymbol{c}+1}+\sqrt{\widetilde{\tau}})^2}{\sqrt{\widetilde{\tau}(\boldsymbol{c}^\top \boldsymbol{c}+1)}}\Bigg(\tilde{\boldsymbol{f}}^\top_{\tilde{\beta}}(\boldsymbol{x})\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^\top + \boldsymbol{c}\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\Big)\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x}) \\ &- 4\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\Bigg), \\ T_2(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) &= \sqrt{\frac{\widetilde{\tau}}{\boldsymbol{c}^\top \boldsymbol{c}+1}}\Bigg(\tilde{\boldsymbol{f}}^\top_{\tilde{\beta}}(\boldsymbol{x})\Big(\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{c}\boldsymbol{c}^\top + \boldsymbol{c}\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\Big)\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x}) \\ &- 2\boldsymbol{c}^\top\tilde{\boldsymbol{M}}^{-1}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{f}}_{\tilde{\beta}}(\boldsymbol{x})\Bigg). \end{aligned}$$

Then
(1) If ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) for a model with intercept and $T_1(\boldsymbol{x},\tilde{\boldsymbol{\beta}}) \geq 0$ for all $\boldsymbol{x} \in \widetilde{\mathcal{X}}$ then $\xi^*_{-\boldsymbol{0}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for the corresponding model without intercept.
(2) If $\xi^*_{-\boldsymbol{0}}$ is locally A-optimal (at $\tilde{\boldsymbol{\beta}}$) for a model without intercept and

$$\boldsymbol{f}^\top_{\tilde{\beta}}(\boldsymbol{x})\tilde{\boldsymbol{M}}^{-2}(\xi^*_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\boldsymbol{f}_{\tilde{\beta}}(\boldsymbol{x}) \leq \widetilde{\tau}\left(1-(\boldsymbol{c}^\top\boldsymbol{f}^\top_{\tilde{\beta}}(\boldsymbol{x})-1)^2\right) + T_2(\boldsymbol{x},\tilde{\boldsymbol{\beta}})\ \forall \boldsymbol{x} \in \mathcal{X}$$

then ξ^* is locally A-optimal (at $\boldsymbol{\beta}$) for the corresponding model with intercept.

Remark 6.2.1. *In view of the assumptions of the previous corollaries $\boldsymbol{M}^{-1}(\xi,\boldsymbol{\beta})$ is given by (3.27) ($\tilde{u}_0$ vanishes). That is $\boldsymbol{c}^\top\tilde{\boldsymbol{m}}(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}}) = 1$, $\tilde{\boldsymbol{M}}^{-1}(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{m}}(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}}) = \boldsymbol{c}$ thus $\tilde{\boldsymbol{m}}^\top(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{M}}^{-1}(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{m}}(\xi_{-\boldsymbol{0}},\tilde{\boldsymbol{\beta}}) = 1$.*

In addition, specific results given in Chapter 5 can be extended under the multivariate nonlinear model with univariate models given by (6.1). In particular, as presented in Section 5.3, for the jth component with nonlinear model (6.1) and function $\boldsymbol{f}_{\boldsymbol{\beta}_j}(\boldsymbol{x}_j)$ from (6.2) if there exist constant vectors $\boldsymbol{c}_j$ such that $\boldsymbol{c}_j^\top\boldsymbol{f}_{\boldsymbol{\beta}_j}(\boldsymbol{x}_j) = 1$ for all $\boldsymbol{x}_j \in \mathcal{X}_j$ $(1 \leq j \leq m)$ then a product-type design can be derived. E.g. Nonlinear models that were discussed in Dette et al. (2008) can provide such a result.

Furthermore, in many applied aspects for gamma models, the log-link function is considered as a main alternative to the canonical one (see Kilian et al. (2002), Wenig et al. (2009), Gregori et al. (2011), McCrone, Knapp, and Fombonne (2005), Montez-Rath et al. (2006)). In that case the intensity function $u(\boldsymbol{x}, \boldsymbol{\beta}) = 1$ and thus the information matrix under gamma models is equivalent to that under ordinary regression models. For that reason, the optimal designs for a gamma model are identical to those for an ordinary regression model with similar linear predictor. However, in Hardin and Hilbe (2018) gamma models were fitted considering various link functions, for example; the Box-Cox family of link functions that is given by

$$\boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} = \begin{cases} \big(\mu^\lambda - 1\big)/\lambda & (\lambda \neq 0) \\ \log \mu & (\lambda = 0) \end{cases} \tag{6.4}$$

which involves the log-link at $\lambda = 0$ (see Atkinson and Woods (2015)). The intensity function is thus defined as

$$u(\boldsymbol{x}, \lambda\boldsymbol{\beta}) = \Big(\lambda \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} + 1\Big)^{-2}, \boldsymbol{x} \in \mathcal{X}. \tag{6.5}$$

Here, the positivity condition (4.4) of the expected mean $\mu = E(Y)$ of a gamma distribution is modified to $\lambda \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} > -1$ for all $\boldsymbol{x} \in \mathcal{X}$. Therefore, in specific, for a gamma model without intercept the experimental region might be considered as $\mathcal{X} = [0,1]^\nu$. As an example, consider $\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2)^\top$ on $\mathcal{X} = [0,1]^2$ with vertices $\boldsymbol{v}_1 = (0,0)^\top$, $\boldsymbol{v}_2 = (1,0)^\top$, $\boldsymbol{v}_3 = (0,1)^\top$, $\boldsymbol{v}_4 = (1,1)^\top$. Let $u_k = u(\boldsymbol{v}_k, \lambda\boldsymbol{\beta})$ for all $(1 \leq k \leq 4)$. The Equivalence Theorem (Theorem 2.2.2, condition (2.11)) approves the D-optimality of the design ξ^* which assigns equal weights $1/2$ to the vertices $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ at the point $\lambda\boldsymbol{\beta}$. This result might be extended for a multiple-factor model without intercept. However, the expression $\lambda \boldsymbol{f}^\top(\boldsymbol{x})\boldsymbol{\beta} + 1$ under non-intercept model could be viewed as a linear predictor of a gamma model with known intercept (i.e. $\beta_0 = 1$). Adopting the Box-Cox family as a class of link functions for gamma models could be a topic of future research.

There are still various optimality criteria that might be employed under gamma models together with the complete class of designs provided in Chapter 4. In specific, the integrated mean squared error (IMSE)-criterion was not considered carefully under generalized linear models and therefore, it is highly recommended as the next challenging topic.

Bibliography

Abdelbasit, K. M. and Plackett, R. L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, **78**, 90–98.

Aminnejad, M. and Jafari, H. (2017). Bayesian A- and D-optimal designs for gamma regression model with inverse link function. *Communications in Statistics – Similation and Computation*, **46**, 8166–8189.

Atkinson, A. and Haines, L. (1996). 14 Designs for nonlinear and generalized linear models. In: *Design and Analysis of Experiments*. Ed. by S. Ghosh and C. Rao. **13**, Handbook of Statistics. Elsevier, Amsterdam, 437–475.

Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. **34**, Oxford University Press, Oxford.

Atkinson, A. C. and Woods, D. C. (2015). Designs for generalized linear models. In: *Handbook of Design and Analysis of Experiments*. Ed. by J. S. Angela Dean Max Morris and D. Bingha. Chapman & Hall/CRC Press, Boca Raton, 471–514.

Barber, J. and Thompson, S. (2004). Multiple regression of cost data: Use of generalised linear models. *Journal of Health Services Research & Policy*, **9**, 197–204.

Bartlett, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, **22**, 107–111.

Berger, M. P. and Wong, W.-K. (2009). *An Introduction to Optimal Designs for Social and Biomedical Research*. **83**, Wiley, Chichestery.

Biedermann, S., Dette, H., and Zhu, W. (2006). Optimal designs for dose–response models with restricted design spaces. *Journal of the American Statistical Association*, **101**, 747–759.

Burridge, J. and Sebastiani, P. (1992). Optimal designs for generalized linear models. *Journal of the Italian Statistical Society*, **1**, 183–202.

Burridge, J. and Sebastiani, P. (1994). D-Optimal designs for generalised linear models with variance proportional to the square of the mean. *Biometrika*, **81**, 295–304.

Chang, S. I. (1994). Some properties of multiresponse D-optimal designs. *Journal of Mathematical Analysis and Applications*, **184**, 256–262.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586–602.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*. Second edition. Wiley, New York.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82**, 407–410.

Das, I. and Mukhopadhyay, S. (2012). Robust designs for multivariate generalized linear models with misspecification in the linear predictor: A quantile dispersion graphs approach. In: *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, 1–6.

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association*, **103**, 1225–1237.

Dette, H., Hoyden, L., Kuhnt, S., and Schorning, K. (2013). Optimal designs for multi-response generalized linear models with applications in thermal spraying. *arXiv e-prints*. arXiv: 1312.4472 [stat.AP].

Dobson, A. J. and Barnett, A. G. (2018). *An Introduction to Generalized Linear Models*. Fourth edition. CRC press, Boca Raton.

Ehrenfeld, S. (1956). Complete class theorems in experimental design. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability: Contributions to the Theory of Statistics*. **1**, University of California Press, Berkeley, California, 57–67.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13**, 342–368.

Fedorov, V. V., Gagnon, R. C., and Leonov, S. L. (2002). Design of experiments with unknown parameters in variance. *Applied Stochastic Models in Business and Industry*, **18**, 207–218.

Fedorov, V. V. and Leonov, S. L. (2013). *Optimal Design for Nonlinear Response Models*. CRC Press, Boca Raton.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Fedorov, V. V. (1971). The design of experiments in the multiresponse case. *Theory of Probability & Its Applications*, **16**, 323–332.

Fisher, R. A. (1937). *The Design of Experiments*. Oliver And Boyd, Edinburgh.

Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-Linear problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**, 569–583.

Ford, I., Titterington, D. M., and Kitsos, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics*, **31**, 49–60.

Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Third edition. SAGE Publications, Los Angeles.

Gaffke, N., Graßhoff, U., and Schwabe, R. (2014). Algorithms for approximate linear regression design with application to a first order model with heteroscedasticity. *Computational Statistics and Data Analysis*, **71**, 1113–1123.

Gaffke, N., Idais, O., and Schwabe, R. (2019). Locally optimal designs for gamma models. *Journal of Statistical Planning and Inference*. DOI: https://doi.org/10.1016/j.jspi.2019.04.002.

Gaffke, N. and Krafft, O. (1982). Matrix inequalities in the Loewner ordering. In: *Modern Applied Mathematics*. Ed. by B. Korte. North-Holland, Amsterdam, 595–622.

Gea-Izquierdo, G. and Cañellas, I. (2009). Analysis of holm oak intraspecific competition using gamma regression. *Forest Science*, **55**, 310–322.

Goldburd, M., Khare, A., and Tevet, D. (2016). Generalized Linear Models for Insurance Rating. In: *Casualty Actuarial Society, Virginia*.

Graßhoff, U., Großmann, H., Holling, H., and Schwabe, R. (2007). Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor. *Journal of Statistical Planning and Inference*, **137**, 3882–3893.

Graßhoff, U., Holling, H., and Schwabe, R. (2018). D-Optimal design for the Rasch counts model with multiple binary predictors. *ArXiv e-prints*. arXiv: 1810.03893 [stat.ME].

Graßhoff, U., Holling, H., and Schwabe, R. (2013). Optimal design for count data with binary predictors in item response theory. In: *mODa 10-Advances in Model-Oriented Design and Analysis*. Ed. by D. Ucinski, A. C. Atkinson, and M. Patan. Springer International Publishing, Heidelberg, 117–124.

Graßhoff, U., Holling, H., and Schwabe, R. (2015). Poisson model with three binary predictors: when are saturated designs optimal? In: *Stochastic Models, Statistics and Their Applications*. Ed. by A. Steland, E. Rafajłowicz, and K. Szajowski. Springer International Publishing, Cham, 75–81.

Gregori, D., Pagano, E., Merletti, F., Petrinco, M., Bo, S., and Desideri, A. (2011). Regression models for analyzing costs and their determinants in health care: an introductory review. *International Journal for Quality in Health Care*, **23**, 331–341.

Grover, G., Sabharwal, A. S. A., and Mittal, J. (2013). An application of gamma generalized linear model for estimation of survival function of diabetic nephropathy patients. *International Journal of Statistics in Medical Research*, **2**, 209–219.

Hardin, J. W. and Hilbe, J. M. (2018). *Generalized Linear Models and Extensions*. Fourth edition. Stata Press, Texas.

Harman, R. and Trnovská, M. (2009). Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices. *Mathematica Slovaca*, **59**, 693–704,

He, L. (2018). Optimal designs for multi-factor nonlinear models based on the second-order least squares estimator. *Statistics & Probability Letters*, **137**, 201–208.

Heise, M. A. and Myers, R. H. (1996). Optimal designs for bivariate logistic regression. *Biometrics*, **52**, 613–624.

Huang, M.-N. L., Chen, R.-B., Lin, C.-S., and Wong, W. K. (2006). Optimal designs for parallel models with correlated responses. *Statistica Sinica*, **16**, 121–133.

Huda, S. and Mukerjee, R. (1988). Optimal weighing designs: approximate theory. *Statistics*, **19**, 513–517.

Idais, O. and Schwabe, R. (2019). Analytic solutions for locally optimal designs for gamma models having linear predictor without intercept. *arXiv e-prints*. arXiv: 1904.09232 [math.ST].

Imhof, L. (2000). Optimum designs for a multiresponse regression model. *Journal of Multivariate Analysis*, **72**, 120–131.

Kabera, G. M., Haines, L. M., and Ndlovu, P. (2015). The analytic construction of D-optimal designs for the two-variable binary logistic regression model without interaction. *Statistics*, **49**, 1169–1186.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statist. Sci.* **21**, 376–399.

Kiefer, J (1961). Optimum designs in regression problems, II. *The Annals of Mathematical Statistics*, 298–325.

Kiefer, J. (1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, **62**, 277–288.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, **12**, 363–366.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, **21**, 272–304.

Kilian, R., Matschinger, H., Löeffler, W., Roick, C., and Angermeyer, M. C. (2002). A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. *The journal of mental health policy and economics*, **51**, 21–31.

Konstantinou, M., Biedermann, S., and Kimber, A. (2014). Optimal designs for two-parameter nonlinear models with application to survival models. *Statistica Sinica*, **24**, 415–428.

Krafft, O. and Schaefer, M. (1992). D-Optimal designs for a multivariate regression model. *Journal of Multivariate Analysis*, **42**, 130–140.

Kurotschka, V. G. and Schwabe, R. (1996). The reduction of design problems for multivariate experiments to univariate possibilities and their limitations. *Research Developments in Probability and Statistics. Festschrift in Honor of Madan L. Puri on the Occasion of his 65th Birthday VSP, Utrecht*. Ed. by E. Brunner and M. Denker, 193–204.

Kurtoğlu, F. and Özkale, M. R. (2016). Liu estimation in generalized linear models: application on gamma distributed response variable. *Statistical Papers*, **57**, 911–928.

Li, G. and Balakrishnan, N. (2011). Optimal designs for tumor regrowth models. *Journal of Statistical Planning and Inference*, **141**, 644–654.

Li, K.-H., Lau, T.-S., and Zhang, C. (2005). A note on D-optimal designs for models with and without an intercept. *Statistical Papers*, **46**, 451–458.

Liang, K.-Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lindsey, J. K., Byrom, W. D., Wang, J., Jarvis, P., and Jones, B. (2000). Generalized nonlinear models for pharmacokinetic data. *Biometrics*, **56**, 81–88.

Liu, J. and Colditz, G. A. (2017). Optimal design of longitudinal data analysis using generalized estimating equation models. *Biometrical Journal*, **59**, 315–330.

Liu, X., Yue, R.-X., and Hickernell, F. J. (2011). Optimality criteria for multiresponse linear models based on predictive ellipsoids. *Statistica Sinica*, **21**, 421–432.

Lu, T.-T. and Shiou, S.-H. (2002). Inverses of 2×2 block matrices. *Computers & Mathematics with Applications*, **43**, 119–129.

Manning, W. G. and Mullahy, J. (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics*, **20**, 461–494.

Mathew, T. and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, **93**, 295–307.

McCrone, P., Knapp, M., and Fombonne, E. (2005). The Maudsley long-term follow-up of child and adolescent depression. *European Child & Adolescent Psychiatry*, **14**, 407–413.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Second edition. Chapman & Hall, London.

Montez-Rath, M., Christiansen, C. L., Ettner, S. L., Loveland, S., and Rosen, A. K. (2006). Performance of statistical models to predict mental health and substance abuse cost. *BMC Medical Research Methodology*, **6**, 53–63.

Montgomery, D. C. (2017). *Design and Analysis of Experiments*. Ninth edition. Wiley, New York.

Moran, J. L., Solomon, P. J., Peisach, A. R., and Martin, J. (2007). New models for old questions: generalized linear models for cost prediction. *Journal of Evaluation in Clinical Practice*, **13**, 381–389.

Mukhopadhyay, S. and Khuri, A. (2008). Comparison of designs for multivariate generalized linear models. *Journal of Statistical Planning and Inference*, **138**, 169–183.

Myers, R. H. and Montgomery, D. C. (1997). A tutorial on generalized linear models. *Journal of Quality Technology*, **29**, 274–291.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, **135**, 370–384.

Ng, V. K. and Cribbie, R. A. (2017). Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. *Current Psychology*, **36**, 225–235.

Oehlert, G. (2000). *A First Course in Design and Analysis of Experiments*. W. H. Freeman.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radloff, M. and Schwabe, R. (2016). Invariance and equivariance in experimental design for nonlinear models. In: *mODa 11-Advances in Model-Oriented Design and Analysis*. Ed. by A. C. A. Joachim Kunert Christine H. Müller. Springer, 217–224.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, New Jersey.

Rodríguez, C., Ortiz, I., and Martínez, I. (2015). Locally and maximin optimal designs for multi-factor nonlinear models. *Statistics*, **49**, 1157–1168.

Rodríguez-Díaz, J. and Sánchez-León, G. (2019). Optimal designs for multiresponse models with double covariance structure. *Chemometrics and Intelligent Laboratory Systems*, **189**, 1–7.

Rodríguez, C., Ortiz, I., and Martínez, I. (2016). A-optimal designs for heteroscedastic multifactor regression models. *Communications in Statistics-Theory and Methods*, **45**, 757–771.

Russell, K. G., Woods, D. C., Lewis, S. M., and Eccleston, J. A. (2009). D-optimal designs for Poisson regression models. *Statistica Sinica*, **19**, 721–730.

Schmidt, D. (2019). Characterization of c-, L- and ϕ_k-optimal designs for a class of non-linear multiple-regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **81**, 101–120.

Schmidt, D. and Schwabe, R. (2017). Optimal design for multiple regression with information driven by the linear predictor. *Statistica Sinica*, **27**, 1371–1384.

Schwabe, R. (1995). Designing experiments for additive nonlinear models. In: *mODa 4- Advances in Model-Oriented Data Analysis*. Ed. by C. P. Kitsos and W. G. Müller. Physica-Verlag HD, Heidelberg, 77–85.

Schwabe, R. (1996a). Optimal design of experiments for multivariate response in two-factor linear models. *Multidimensional Statistical Analysis and Theory of Random Matrices. Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, OH. VSP, Utrecht*. Ed. by A. K. Gupta and V. L. Girko, 235–242.

Schwabe, R. (1996b). *Optimum Designs for Multi-Factor Models*. Springer, New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall, London.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, **12**, 1–85.

Soumaya, M., Gaffke, N., and Schwabe, R. (2015). Optimal design for multivariate observations in seemingly unrelated linear models. *Journal of Multivariate Analysis*, **142**, 48–56.

Tong, L., Volkmer, H. W., and Yang, J. (2014). Analytic solutions for D-optimal factorial designs under generalized linear models. *Electron. J. Statist.* **8**, 1322–1344.

Ven, P. van de and Woods, D. (2014). Optimal blocked minimum-support designs for non-linear models. *Journal of Statistical Planning and Inference*, **144**, 152–159.

Wald, A. (1943). On the efficient design of statistical investigations. *The Annals of Mathematical Statistics*, **14**, 134–140.

Walker, S. H. and Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, **54**, 167–179.

Wang, Y., Myers, R. H., Smith, E. P., and Ye, K. (2006). D-optimal designs for Poisson regression models. *Journal of Statistical Planning and Inference*, **136**, 2831–2845.

Wenig, C. M., Schmidt, C. O., Kohlmann, T., and Schweikert, B. (2009). Costs of back pain in Germany. *European Journal of Pain*, **13**, 280–286.

Wolfram Research, Inc. (2018). *Mathematica*. Version 11.3. Champaign, IL.

Woods, D. C. and Ven, P. Van de (2011). Blocked designs for experiments with correlated non-normal response. *Technometrics*, **53**, 173–182.

Wynn, H. P. (1984). Jack Kiefer's contributions to experimental design. *Ann. Statist.* **12**, 416–423.

Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics - Simulation and Computation*, **44**, 2264–2277.

Yang, J., Mandal, A., and Majumdar, D. (2012). Optimal designs for two-level fractional experiments with binary response. *Statistica Sinica*, **22**, 885–907.

Yang, M. (2008). A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference*, **138**, 624–641.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics*, **37**, 518–541.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, 1593–1606.

Yue, R.-X., Liu, X., and Chatterjee, K. (2014). D-optimal designs for multiresponse linear models with a qualitative factor. *Journal of Multivariate Analysis*, **124**, 57–69.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, **57**, 348–368.

Zhang, C. and Wong, W. K. (2013). Optimal designs for mixture models with amount constraints. *Statistics & Probability Letters*, **83**, 196–202.

Zocchi, S. S. and Atkinson, A. C. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics*, **55**, 437–444.

List of Abbreviations

GLM	**G**eneralized **L**inear **M**odel
MGLM	**M**ultivariate **G**eneralized **L**inear **M**odel
MLE	**M**aximum **L**ikelihood **E**stimate

List of Symbols

Y	univariate observation
$p(Y;\theta,\phi)$	one-parameter exponential family distribution
μ	expected mean
$V(\mu)$	mean-variance function
$\mathrm{var}(Y)$	variance of Y
$\boldsymbol{f}$	vector regression function
f_j	component regression function
$\boldsymbol{\beta}$	vector of model parameters
$\mathcal{X}$	experimental (design) region
$\widetilde{\mathcal{X}}$	reduced experimental region
$\mathbb{N}$	set of natural numbers
$\mathbb{R}^p$	set of p-dimensional real numbers
p	dimension of model parameter $\boldsymbol{\beta}$
ν	dimension of $\mathcal{X}$
$\boldsymbol{v}$	vertex
η	linear predictor
ξ	design
$\mathrm{supp}(\xi)$	support of ξ
r	support size of ξ
ω	weight of ξ
$\xi_{\boldsymbol{x}}$	one-point design of $\boldsymbol{x}$
ξ^*	optimal design
$\boldsymbol{x}^*$	design point of ξ^*
ω^*	optimal weight of ξ^*
Ω	diagonal matrix of optimal weights
Ξ	set of designs
$\mathbb{M}_{\boldsymbol{\beta}}$	convex set of symmetric nonnegative definite $p \times p$ moment matrices
$\widetilde{\Xi}$	essentially/ complete class of designs
$u(\boldsymbol{x},\boldsymbol{\beta})$	intensity function
$\boldsymbol{f}_{\boldsymbol{\beta}}$	square root of intensity multiplied by vector regression function
$\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X})$	induced experimental region

$\boldsymbol{I}$	identity matrix
$\mathbf{1}$	vector of ones
$\mathbf{0}$	vector of zeros
$\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta})$	Fisher information matrix at one point $\boldsymbol{x}$
$\boldsymbol{M}(\xi, \boldsymbol{\beta})$	Fisher information matrix of design ξ
Φ	optimality criterion function
$\mathrm{tr}(\boldsymbol{A})$	trace of matrix $\boldsymbol{A}$
$\det(\boldsymbol{A})$	determinant of matrix $\boldsymbol{A}$
$\lambda(\xi, \boldsymbol{\beta})$	eigenvalue of $\boldsymbol{M}(\xi, \boldsymbol{\beta})$
Φ_k	Kiefer criteria
Conv	convex hull operation
S	index set
$\boldsymbol{e}$	ν-dimensional unit vector
$\boldsymbol{F}$	design matrix
$\boldsymbol{V}$	weight matrix
γ	ratio of specific parameters
$\hat{\boldsymbol{\beta}}$	parameter estimators
$\otimes$	Kronecker product
$\mathrm{var}(\hat{\boldsymbol{\beta}})$	approximated variance-covariance matrix of $\hat{\boldsymbol{\beta}}$
ρ	exponent/ correlation
$\xi^*_{\boldsymbol{\beta}}$	locally optimal design at $\boldsymbol{\beta}$
$\mathrm{Eff}(\xi, \boldsymbol{\beta})$	efficiency of design ξ with respect to ξ^* at $\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}(\xi, n)$	MLE under ξ and for sample size n
$\boldsymbol{V}(\hat{\boldsymbol{\beta}}(\xi, n), \boldsymbol{\beta})$	covariance matrix of $\hat{\boldsymbol{\beta}}(\xi, n)$ at $\boldsymbol{\beta}$ for design ξ
$\mathrm{Eff}(\xi, \xi_1, n, \boldsymbol{\beta})$	efficiency of design ξ with respect to ξ_1

List of Figures

List of Tables