Metrika (2021) 84:1–26
https://doi.org/10.1007/s00184-019-00760-3

Analytic solutions for locally optimal designs for gamma models having linear predictors without intercept

Osama Idais[1] · **Rainer Schwabe**[1]

Received: 18 July 2019 / Published online: 3 January 2020

Abstract

The gamma model is a generalized linear model for gamma-distributed outcomes. The model is widely applied in psychology, ecology or medicine. Recently, Gaffke et al. (J Stat Plan Inference 203:199–214, 2019) established a complete class and an essentially complete class of designs for gamma models to obtain locally optimal designs in particular when the linear predictor includes an intercept term. In this paper we extend this approach to gamma models having linear predictors without intercept. For a specific scenario sets of locally D- and A-optimal designs are established. It turns out that the optimality problem can be transformed to one under gamma models with intercept leading to a reduction in the dimension of the experimental region. On that basis optimality results can be transferred from one model to the other and vice versa. Additionally by means of the general equivalence theorem optimality can be characterized for multiple regression by a system of polynomial inequalities which can be solved analytically or by computer algebra. Thus necessary and sufficient conditions can be obtained on the parameter values for the local D-optimality of specific designs. The robustness of the derived designs with respect to misspecification of the initial parameter values is examined by means of their local D-efficiencies.

Keywords Generalized linear model · Optimal design · Gamma model · Intercept term · Interaction

✉ Osama Idais
osama.idais@ovgu.de

Rainer Schwabe
rainer.schwabe@ovgu.de

[1] Institute for Mathematical Stochastics, Otto-von-Guericke-University Magdeburg, PF 4120, 39016 Magdeburg, Germany

1 Introduction

The gamma model is employed for outcomes that are non-negative, 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. In ecology and forestry, Gea-Izquierdo and Cañellas (2009) mentioned that gamma models offer a great potential for many forestry applications and they used gamma models to analyze plant competition. In a medical context, Grover et al. (2013) fitted a gamma model with duration of diabetes as a response variable and predictors like the rate of rise in serum creatinine (SrCr) and number of successes (number of times SryCr values exceed its normal range (1.4 mg/dl)). For a study about air pollution, Kurtoğlu and Özkale (2016) employed a gamma model to analyze nitrogen dioxide concentrations considering some weather factors (see also Chatterjee 1988, Section 8.7). In psychological studies, recently, Ng and Cribbie (2017) used a gamma model for modeling the relationship between negative automatic thoughts (NAT) and socially prescribed perfectionism (SPP).

Although the canonical link is frequently employed in the gamma model, there is always a doubt about the suitable link function for outcomes. Therefore, a class of link functions might be employed. The common alternative links mostly come from the Box–Cox family and the power link family (see Atkinson and Woods 2015) that includes the canonical link. In the theory of optimal designs, the information matrix of a generalized linear model depends on the model parameters through the intensity function. Locally optimal designs can be derived through maximizing a specific optimality criterion at certain values of the parameters. Although the gamma model is used in many applications, optimal designs in this model have not received a wide attention. Geometric approaches were employed to derive locally D-optimal designs for a gamma model with a single factor (Ford et al. 1992) and with multiple factors (Burridge and Sebastiani 1994). For gamma models with two factors and without intercept a geometric approach was also utilized in Burridge and Sebastiani (1992). Recently, in Gaffke et al. (2019) we provided analytic solutions for optimal designs under gamma models and locally complete class and essentially complete class of designs were established under certain assumptions. Therefore, the complexity of deriving locally optimal designs is reduced and one can look for the optimal design in these classes.

The intercept term in generalized linear models (gamma models) characterizes the expected mean when all the explanatory variables are equal to zero. In this case, the linear predictor represents an impact of all the unobserved fixed variables in the model. When the intercept is significantly zero the average impact of all the unobserved fixed variables is also significantly zero and the model includes probably most variables which explain the outcome. In this paper, we will focus on the gamma models when the linear predictor does not significantly include the intercept. Our main goal is to develop various approaches to obtain locally optimal designs w.r.t. D- and A-optimality criteria.

This paper is organized as follows. In Sect. 2, the proposed model, the information matrix and the locally optimal design are presented. In Sect. 3, locally D- and A-optimal designs are derived. In Sect. 4, a two-factor model with interaction is considered for which locally D-optimal designs are derived. The performance of some derived D-optimal designs is examined in Sect. 5. Finally, a brief discussion and conclusions are given in Sect. 6.

2 Model, information and design

Let $Y_1, \ldots, Y_n$ be independent gamma-distributed response variables for n experimental units, where for each Y_i the density is given by

$$p(y_i; \kappa, \lambda_i) = \frac{\lambda_i^\kappa}{\Gamma(\kappa)} y_i^{\kappa-1} e^{-\lambda_i y_i} \;, \kappa,\; \lambda_i,\; y_i > 0,\;\; i = 1, \ldots, n. \tag{2.1}$$

The shape parameter κ of the gamma distribution is assumed to be known and 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 (2.1) is the reciprocal (inverse) link,

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

Here $\boldsymbol{f} = (f_1, \ldots, f_p)^T$ is a given $\mathbb{R}^p$-valued function on the experimental region $\mathcal{X} \subset \mathbb{R}^\nu$, $\nu \geq 1$, with linearly independent component functions $f_1, \ldots, f_p$, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the 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 at a point $\boldsymbol{x} \in \mathcal{X}$ (see Atkinson and Woods 2015) is given by

$$u_0(\boldsymbol{x}, \boldsymbol{\beta}) = \left(\mathrm{var}(Y)\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}\right)^2\right)^{-1} = \kappa\left(\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta}\right)^{-2}. \tag{2.2}$$

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

$$\boldsymbol{f}^T(\boldsymbol{x})\,\boldsymbol{\beta} > 0 \;\; \text{for all} \;\; \boldsymbol{x} \in \mathcal{X}. \tag{2.3}$$

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 $u_0(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{f}^T(\boldsymbol{x})$. Note that the positive factor κ is the same for all $\boldsymbol{x}$ and $\boldsymbol{\beta}$ and will not affect any design optimization below. We will ignore that factor and consider a normalized version of the Fisher information matrix at $\boldsymbol{x}$ and $\boldsymbol{\beta}$,

$$\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = u(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{f}^T(\boldsymbol{x}) \;\; \text{where}\; u(\boldsymbol{x}, \boldsymbol{\beta}) = \left(\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta}\right)^{-2}. \tag{2.4}$$

For a given parameter value $\boldsymbol{\beta}$ we denote by $\boldsymbol{f}_{\boldsymbol{\beta}}$ the local regression function

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \left(\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta}\right)^{-1}\boldsymbol{f}(\boldsymbol{x}) \;\; \text{for all} \;\; \boldsymbol{x} \in \mathcal{X}. \tag{2.5}$$

Then the Fisher information matrix (2.4) can be written as $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}_{\boldsymbol{\beta}}^T(\boldsymbol{x})$.

We will make use of approximate designs with finite support on the experimental region $\mathcal{X}$. An approximate design ξ on $\mathcal{X}$ is defined as

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m \\ \omega_1 & \omega_2 & \cdots & \omega_m \end{pmatrix}, \tag{2.6}$$

where $m \in \mathbb{N}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathcal{X}$ are pairwise distinct points and $\omega_1, \omega_2, \ldots, \omega_m > 0$ denote the proportions of observations to be made at the settings $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$, respectively, with $\sum_{i=1}^m \omega_i = 1$. The set $\text{supp}(\xi) = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ is called the support of ξ and $\omega_1, \ldots, \omega_m$ are called the weights of ξ (see Silvey 1980, p. 15). A design ξ is minimally supported if the number of support points is equal to the number of model parameters (i.e., $m = p$). A minimally supported design which is also called a saturated design will appear frequently in the current work. The information matrix of a design ξ at a parameter vector $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{m} \omega_i \boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{\beta}). \tag{2.7}$$

Another representation of the information matrix (2.7) can be considered by defining the $m \times p$ design matrix $\boldsymbol{F} = [\boldsymbol{f}(\boldsymbol{x}_1), \ldots, \boldsymbol{f}(\boldsymbol{x}_m)]^T$ and the $m \times m$ weight matrix $\boldsymbol{V} = \text{diag}(\omega_i u(\boldsymbol{x}_i, \boldsymbol{\beta}))_{i=1}^m$ and hence, $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}^T \boldsymbol{V} \boldsymbol{F}$.

A locally optimal design minimizes a convex criterion function of the information matrix at a given parameter vector $\boldsymbol{\beta}$. Denote by ”det” and ”tr” the determinant and the trace of a matrix, respectively. We will employ the popular D-criterion and the A-criterion. More precisely, a design ξ^* is said to be locally D-optimal (at $\boldsymbol{\beta}$) if its information matrix $\boldsymbol{M}(\xi^*, \boldsymbol{\beta})$ at $\boldsymbol{\beta}$ is nonsingular and $\det\big(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\big) = \min_\xi \det\big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\big)$ where the minimum on the r.h.s. is taken over all designs ξ whose information matrix at $\boldsymbol{\beta}$ is nonsingular. Similarly, a design ξ^* is said to be locally A-optimal (at $\boldsymbol{\beta}$) if its information matrix at $\boldsymbol{\beta}$ is nonsingular and $\text{tr}\big(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\big) = \min_\xi \text{tr}\big(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})\big)$ where, again, the minimum is taken over all designs ξ whose information matrix at $\boldsymbol{\beta}$ is nonsingular.

Remark 2.1 The set of designs for which the information matrix is nonsingular does not depend on $\boldsymbol{\beta}$ (since $u(\boldsymbol{x}, \boldsymbol{\beta})$ 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 $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \boldsymbol{F}^T \boldsymbol{V} \boldsymbol{F}$ has full rank if and only if $\boldsymbol{F}$ has full rank.

Remark 2.2 If the experimental region is a compact set and the functions $\boldsymbol{f}(\boldsymbol{x})$ and $u(\boldsymbol{x}, \boldsymbol{\beta})$ are continuous in $\boldsymbol{x}$ then the set of all nonnegative definite information matrices is compact. Therefore, there exists a locally D- or A-optimal design for any given parameter vector $\boldsymbol{\beta}$.

In order to verify the local optimality of a design the general equivalence theorem is commonly employed. It provides necessary and sufficient conditions for a design to be optimal with respect to the optimality criterion, in particular to the D- or A-criterion, and thus the optimality of a suggested design can easily be verified or disproved. In the following we present equivalent characterizations of locally D- and A-optimal designs (see Silvey 1980, p. 40, p. 48 and p. 54).

Theorem 2.1 *Let* $\boldsymbol{\beta}$ *be a given parameter point and let* ξ^* *be a design with nonsingular information matrix* $\boldsymbol{M}(\xi^*, \boldsymbol{\beta})$.

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

$$u(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\boldsymbol{f}(\boldsymbol{x}) \leq p \quad \text{for all } \boldsymbol{x} \in \mathcal{X}.$$

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

$$u(\boldsymbol{x}, \boldsymbol{\beta})\, \boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta})\boldsymbol{f}(\boldsymbol{x}) \leq \operatorname{tr}\big(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\big) \quad \text{for all } \boldsymbol{x} \in \mathcal{X}.$$

Remark 2.3 The inequalities given by part (a) or part (b) of Theorem 2.1 become equations at support points of any D- or A-optimal design ξ^*, respectively.

Throughout, we consider gamma models that do not contain a constant (intercept) term (neither implicitly nor explicitly). More precisely, we assume $\boldsymbol{f}(\boldsymbol{0}) = \boldsymbol{0}$. In particular, we restrict to a first order model with

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}, \quad \text{where} \quad \boldsymbol{x} = (x_1, \ldots, x_\nu)^T, \ \ \nu \geq 1, \tag{2.8}$$

and the two-factor model with interaction

$$\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2, x_1x_2)^T, \quad \text{where} \quad \boldsymbol{x} = (x_1, x_2)^T. \tag{2.9}$$

In this context, condition (2.3), i.e., $\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$ implies that $\boldsymbol{0} \notin \mathcal{X}$. Therefore, as an experimental region $\mathcal{X} = [0, \infty)^\nu \setminus \{\boldsymbol{0}\}$ may be considered. Note that this experimental region is no longer compact therefore the existence of optimal designs is not assured automatically and has to be checked separately, e.g., by the compactness of the induced experimental region $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X}) = \{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$.

In contrast, in the paper we often consider a compact experimental region that is a ν-dimensional hypercube

$$\mathcal{X} = \big[a, b\big]^\nu, \ \nu \geq 2 \text{ with } a, b \in \mathbb{R} \text{ and } 0 < a < b, \tag{2.10}$$

with vertices $\boldsymbol{v}_k$, $k = 1, \ldots, K = 2^\nu$ given by the points whose i-th coordinates are either a or b for all $i = 1, \ldots, \nu$.

In Gaffke et al. (2019), Theorem 3.1, we considered a gamma model with regression function $\boldsymbol{f}(\boldsymbol{x})$ as in (2.8) or (2.9) and experimental region (2.10), i.e., $\mathcal{X} = [a, b]^\nu$, $\nu \geq 2$, $0 < a < b$. In that theorem we showed that for any design $\tilde{\xi}$ which has at least one support point not being a vertex from $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ there exists a design ξ that is supported only on the vertices and which is at least as good as $\tilde{\xi}$ w.r.t. the Loewner semi-ordering of nonnegative definite $p \times p$ matrices. That is 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. The set of all designs ξ such that $\text{supp}(\xi) \subseteq \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ under the proposed models (2.8) and (2.9) is, hence, an essentially complete class of designs. As a result, there exists a design ξ^* that is only supported by vertices of $\mathcal{X}$ which is locally optimal (at $\boldsymbol{\beta}$) w.r.t. D- or A-criterion. On that basis, throughout, we restrict to designs whose support is a subset of the vertices of $\mathcal{X}$ given by the hypercube (2.10).

Remark 2.4 Let us denote by $\psi(\boldsymbol{x})$ the left hand side of the equivalence theorems, Theorem 2.1, part (a) or part (b). Often $\psi(\boldsymbol{x})$ is called the sensitivity function. In non-intercept gamma models $\psi(\boldsymbol{x})$ is invariant with respect to simultaneous scale transformation of $\boldsymbol{x}$, i.e., $\psi(\lambda\boldsymbol{x}) = \psi(\boldsymbol{x})$ for any $\lambda > 0$. This comes from the fact that the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ given by (2.5) is invariant with respect to simultaneous rescaling of all components of $\boldsymbol{x}$, i.e., $\boldsymbol{f}_{\boldsymbol{\beta}}(\lambda\boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$. This property is transferred to the information matrix (2.4) since it can be represented in the form $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}_{\boldsymbol{\beta}}^T(\boldsymbol{x})$, and hence $\boldsymbol{M}(\lambda\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta})$. We will use this property to derive optimal designs.

3 First order gamma model

In this section we consider a gamma model with

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x} \text{ where } \boldsymbol{x} = (x_1, \ldots, x_\nu)^T \in \mathcal{X} \subset \mathbb{R}^\nu,\ \nu \geq 2. \tag{3.1}$$

Then

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \frac{1}{\beta_1 x_1 + \cdots + \beta_\nu x_\nu} \begin{pmatrix} x_1 \\ \vdots \\ x_\nu \end{pmatrix} = \Big(\sum_{i=1}^{\nu} \beta_i x_i\Big)^{-1} \boldsymbol{x}. \tag{3.2}$$

Remark 3.1 For a single-factor gamma model with a linear predictor $\eta = \beta x$ where $x \in \mathcal{X} \subseteq \mathbb{R} \setminus \{0\}$ and a parameter β such that $\beta x > 0$ for all $x \in \mathcal{X}$ the Fisher information is given by $\boldsymbol{M}(x, \beta) = 1/\beta^2$ and is constant in x. Hence every design is optimal.

3.1 Optimal designs on an orthant

Firstly let the experimental region $\mathcal{X} = [0, \infty)^\nu \setminus \{\boldsymbol{0}\}$ be considered. For $i = 1, \ldots, \nu$ denote by $\boldsymbol{e}_i$ the ν-dimensional unit vectors. The parameter space is determined by condition (2.3) as $\boldsymbol{x}^T\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$ which is equivalent to $\beta_i > 0$ for all $i = 1, \ldots, \nu$. The induced experimental region $\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X}) = \{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$ is compact since

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\mathcal{X}) = \text{Conv}\{\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{e}_i) : \boldsymbol{e}_i \in \mathcal{X}, i = 1, \ldots, \nu\},$$

because of the invariance with respect to simultaneous scaling mentioned above. Here, 'Conv' denotes the convex hull. That means that for all $\boldsymbol{x} \in \mathcal{X}$ each point $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ can be written as a convex combination of $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{e}_i)$, $i = 1, \ldots, \nu$, i.e., we obtain $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) =$

$\sum_{i=1}^{\nu} \alpha_i \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{e}_i)$ where $\alpha_i = \beta_i x_i / \sum_{j=1}^{\nu} \beta_j x_j$. Obviously, $\alpha_i \geq 0, i = 1, \ldots, \nu$ and $\sum_{i=1}^{\nu} \alpha_i = 1$. As a consequence, the set of all nonnegative definite information matrices is compact and the existence of a locally optimal design is assured (cp. Remark 2.2).

Theorem 3.1 *Consider model (3.1) on the experimental region $\mathcal{X} = [0, \infty)^{\nu} \setminus \{\mathbf{0}\}$. Let $\boldsymbol{x}_i^* = \boldsymbol{e}_i, i = 1, \ldots, \nu$. Let a parameter vector $\boldsymbol{\beta}$ be given such that condition (2.3) is satisfied, i.e., $\beta_i > 0$ for all $i = 1, \ldots, \nu$. Then*

(a) The minimally supported design ξ^* that assigns equal weight $\omega_i^* = 1/\nu$ to the support points $\boldsymbol{x}_i^*$ for all $i = 1, \ldots, \nu$ is locally D-optimal (at $\boldsymbol{\beta}$).
(b) The minimally supported design ξ^* that assigns the weights $\omega_i^* = \beta_i / \sum_{j=1}^{\nu} \beta_j$ for all $i = 1, \ldots, \nu$ to the support points $\boldsymbol{x}_i^*$ for all $i = 1, \ldots, \nu$ is locally A-optimal (at $\boldsymbol{\beta}$).

Proof Define the $\nu \times \nu$ design matrix $\boldsymbol{F} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{\nu}] = \boldsymbol{I}_{\nu}$ where $\boldsymbol{I}_{\nu}$ is the $\nu \times \nu$ identity matrix and the $\nu \times \nu$ weight matrix $\boldsymbol{V} = \mathrm{diag}(\omega_i^*/\beta_i^2)_{i=1}^{\nu}$. Then we have $\boldsymbol{M}(\xi^*, \boldsymbol{\beta}) = \boldsymbol{F}^T \boldsymbol{V} \boldsymbol{F} = \mathrm{diag}(\omega_i^*/\beta_i^2)_{i=1}^{\nu}$ and $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}) = (\boldsymbol{F}^T \boldsymbol{V} \boldsymbol{F})^{-1} = \mathrm{diag}(\beta_i^2/\omega_i^*)_{i=1}^{\nu}$.
In part (a) for D-optimality $\omega_i^* = 1/\nu \, \forall i$, $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}) = \nu \, \mathrm{diag}(\beta_i^2)_{i=1}^{\nu}$ and $\boldsymbol{f}^T(\boldsymbol{x}) \mathrm{diag}(\beta_i^2)_{i=1}^{\nu} \boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{\nu} \beta_i^2 x_i^2$. Hence, by the equivalence theorem (Theorem 2.1, part (a)) ξ^* is locally D-optimal (at $\boldsymbol{\beta}$) if and only if $\left(\sum_{i=1}^{\nu} \beta_i x_i\right)^{-2} \left(\sum_{i=1}^{\nu} \beta_i^2 x_i^2\right) \leq 1$ for all $\boldsymbol{x} \in \mathcal{X}$ which is equivalent to $\sum_{i=1}^{\nu-1} \sum_{j=i+1}^{\nu} \beta_i \beta_j x_i x_j \geq 0$ for all $\boldsymbol{x} \in \mathcal{X}$. The latter inequality holds true by the model assumptions $\beta_i > 0, x_i \geq 0, i = 1, \ldots, \nu$.
In part (b) for A-optimality $\omega_i^* = \beta_i / \sum_{j=1}^{\nu} \beta_j \, \forall i$, $\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta}) = \left(\sum_{i=1}^{\nu} \beta_i\right) \mathrm{diag}(\beta_i)_{i=1}^{\nu}$, $\mathrm{tr}\left(\boldsymbol{M}^{-1}(\xi^*, \boldsymbol{\beta})\right) = \left(\sum_{i=1}^{\nu} \beta_i\right)^2$, $\boldsymbol{M}^{-2}(\xi^*, \boldsymbol{\beta}) = \left(\sum_{i=1}^{\nu} \beta_i\right)^2 \mathrm{diag}(\beta_i^2)_{i=1}^{\nu}$ and $\boldsymbol{f}^T(\boldsymbol{x}) \left(\sum_{i=1}^{\nu} \beta_i\right)^2 \mathrm{diag}(\beta_i^2)_{i=1}^{\nu} \boldsymbol{f}(\boldsymbol{x}) = \left(\sum_{i=1}^{\nu} \beta_i\right)^2 \sum_{i=1}^{\nu} \beta_i^2 x_i^2$. Hence, straightforward computations show that applying the equivalence theorem (Theorem 2.1, part (b)) leads to a conclusion analogous to that in part (a). □

Remark 3.2 The locally D-optimal design provided by part (a) of Theorem 3.1 does not depend on $\boldsymbol{\beta}$ and is, hence, not affected by misspecification of the model parameter.

While the information matrix is invariant w.r.t. to simultaneous rescaling of the components separately for each $\boldsymbol{x}$ as it is mentioned in Remark 2.4, the results of Theorem 3.1 can be extended:

Corollary 3.1 *Consider model (3.1) on the experimental region $\mathcal{X} = [0, \infty)^{\nu} \setminus \{\mathbf{0}\}$. Let a constant real vector $\boldsymbol{a} = (a_1, \ldots, a_{\nu})^T$ be given such that $a_i > 0, i = 1, \ldots, \nu$. Let $\boldsymbol{x}_i^* = a_i \boldsymbol{e}_i, i = 1, \ldots, \nu$. Let a parameter vector $\boldsymbol{\beta}$ be given such that condition (2.3) is satisfied, i.e., $\beta_i > 0$ for all $i = 1, \ldots, \nu$. Then*

(i) *The minimally supported design $\xi_{\boldsymbol{a}}^*$ that assigns equal weight $\omega_i^* = 1/\nu$ to the support points $\boldsymbol{x}_i^*$ for all $i = 1, \ldots, \nu$ is locally D-optimal (at $\boldsymbol{\beta}$).*
(ii) *The minimally supported design $\xi_{\boldsymbol{a}}^*$ that assigns the weights $\omega_i^* = \beta_i / \sum_{j=1}^{\nu} \beta_j$ for all $i = 1, \ldots, \nu$ to the support points $\boldsymbol{x}_i^*$ for all $i = 1, \ldots, \nu$ is locally A-optimal (at $\boldsymbol{\beta}$).*

The derived locally D- and A-optimal designs at a given $\boldsymbol{\beta}$ are not unique. We observe that designs with larger support can be optimal which may be obtained as convex combinations of locally optimal designs given in Corollary 3.1 w.r.t. D- or A-criterion. In the following we characterize sets of locally D-optimal designs and locally A-optimal designs.

Corollary 3.2 *Consider the assumptions of Corollary 3.1 and let $\xi_{\boldsymbol{a}}^*$ be the locally D-resp. A-optimal designs at $\boldsymbol{\beta}$ from Corollary 3.1. Let*

$$\Xi^* = Conv\{\xi_{\boldsymbol{a}}^* : \boldsymbol{a} = (a_1, \ldots, a_\nu)^T, a_i > 0\ \forall i = 1, \ldots, \nu\}.$$

Then Ξ^* is a set of locally D- resp. A-optimal designs (at $\boldsymbol{\beta}$). Obviously, any $\xi^* \in \Xi^*$ can be written as $\xi^* = \sum_{t=1}^{s} \alpha_t \xi_{\boldsymbol{a}_t}^*$ such that $\alpha_t \geq 0, t = 1, \ldots, s$ and $\sum_{t=1}^{s} \alpha_t = 1$ where "s" is an arbitrary positive integer number such that the support may become arbitrarily large.

3.2 Optimal designs on a hypercube

In what follows we consider hypercubes $\mathcal{X} = [a, b]^\nu$, $\nu \geq 2$, $0 < a < b$, as the experimental regions. As mentioned in Remark 2.4, we have $\boldsymbol{f}_{\boldsymbol{\beta}}(\lambda \boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ for all $\lambda > 0$. In particular, as already used above the scaling $\lambda = \lambda_{\boldsymbol{x}}$ may be performed for any value of $\boldsymbol{x}$ individually. By choosing $\lambda_{\boldsymbol{x}} = x_1^{-1}$ we can transform a gamma model without intercept in ν variables $x_1, \ldots, x_\nu$ into a gamma model with intercept in $\nu - 1$ variables $t_1 = x_2/x_1, \ldots, t_{\nu-1} = x_\nu/x_1$. This reduction in the dimension of the covariate $\boldsymbol{x} = (x_1, \ldots, x_\nu)^T$ is useful to determine the candidate support points of a design. Another reduction can be obtained on the parameter space by using the scaling equivariance $\boldsymbol{f}_{\delta\boldsymbol{\beta}}(\boldsymbol{x}) = \delta^{-1} \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ on the parameter space for all $\delta > 0$, where, in particular, $\delta = \delta_{\boldsymbol{\beta}}$ can be chosen as $\delta_{\boldsymbol{\beta}} = \beta_1^{-1}$.

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 3.2 *Consider model (3.1) on the experimental region $\mathcal{X} = [a, b]^2$, $0 < a < b$. Let $\boldsymbol{x}_1^* = (a, b)^T$ and $\boldsymbol{x}_2^* = (b, a)^T$. Let $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ be given according to (2.3) or equivalently $\boldsymbol{\beta}^T \boldsymbol{x}_i^* > 0$ for $i = 1, 2$. Then, the locally D-optimal design ξ_{D}^* and the locally A-optimal design ξ_{A}^* are the following*

$$\xi_{\mathrm{D}}^* = \begin{pmatrix} \boldsymbol{x}_1^* & \boldsymbol{x}_2^* \\ 0.5 & 0.5 \end{pmatrix} \quad and \quad \xi_{\mathrm{A}}^* = \begin{pmatrix} \boldsymbol{x}_1^* & \boldsymbol{x}_2^* \\ \frac{\beta_1 b + \beta_2 a}{(\beta_1 + \beta_2)(a+b)} & \frac{\beta_1 a + \beta_2 b}{(\beta_1 + \beta_2)(a+b)} \end{pmatrix}.$$

Proof Because of the invariance with respect to rescaling we can write

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \left(\beta_1 x_1 + \beta_2 x_2\right)^{-1} \left(x_1\, ,\, x_2\right)^T = \left(\beta_1 + \beta_2 t\right)^{-1} \left(1\, ,\, t\right)^T,$$
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 $t = t(\boldsymbol{x})$ is $\mathcal{T} = t(\mathcal{X}) = \big[(a/b)\,,\,(b/a)\big]$ as $\boldsymbol{x}$ ranges over $\mathcal{X} = [a, b]^2$. Note also that the end points $t_1 = a/b$ and $t_2 = b/a$ arise uniquely from the vertices $\boldsymbol{x}_1^* = (a, b)^T$ and $\boldsymbol{x}_2^* = (b, a)^T$, respectively. Following the proof of Theorem 4.1 in Gaffke et al. (2019) yields the stated results on the locally D- and A-optimal designs in the theorem, where in the case of A-optimality the weight at $t_1 = t(\boldsymbol{x}_1^*) = a/b$ is given by

$$\frac{\big(\beta_1 + \beta_2 t_1\big)\sqrt{1 + t_2^2}}{\big(\beta_1 + \beta_2 t_1\big)\sqrt{1 + t_2^2} + \big(\beta_1 + \beta_2 t_2\big)\sqrt{1 + t_1^2}}$$

and it is straightforward to verify that this quantity is equal to the weight at $\boldsymbol{x}_1^*$. □

Note that the optimal weights for the A-optimal design ξ_A^* given in Theorem 3.2 depend only on the ratios a/b and β_2/β_1. In particular, for $\beta_1 = 0$ or $\beta_2 = 0$ the weights of ξ_A^* do not depend on the respective other parameter.

Remark 3.3 The transformation specified above can also be used for higher dimensions $\nu \geq 3$ to show that the information matrix in a ν-dimensional model without intercept is equivalent to that in a corresponding $(\nu - 1)$-dimensional model with intercept. For $\nu \geq 3$, an analogous transformation of the model as in the proof of Theorem 3.2 is given by

$$\begin{aligned} \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) &= \big(\beta_1 + \beta_2 t_1 + \beta_3 t_2 + \cdots + \beta_\nu t_{\nu-1}\big)^{-1}\big(1, t_1, \ldots, t_{\nu-1}\big)^T, \\ &\text{where } \ t_j = t_j(\boldsymbol{x}) = x_{j+1}/x_1,\ j = 1, \ldots, \nu - 1 \ \text{ for } \boldsymbol{x} = (x_1, x_2, \ldots, x_\nu)^T \\ &\quad \in [a, b]^\nu, 0 < a < b, \end{aligned}$$

leading thus to a first order model with intercept employing a $(\nu - 1)$-dimensional factor $\boldsymbol{t} = (t_1, \ldots, t_{\nu-1})^T$. The range $\big\{\boldsymbol{t}(\boldsymbol{x}) : \boldsymbol{x} \in [a, b]^\nu\big\} \subseteq \mathbb{R}^{\nu-1}$ of $\boldsymbol{t}$ is not a cube but a more complicated polytope. E.g., for $\nu = 3$ and the experimental region $\mathcal{X} = [a, b]^3$ denote the vertices by $\boldsymbol{v}_1 = \big(a, a, a\big)^T$, $\boldsymbol{v}_2 = \big(b, a, a\big)^T$, $\boldsymbol{v}_3 = \big(a, b, a\big)^T$, $\boldsymbol{v}_4 = \big(a, a, b\big)^T$, $\boldsymbol{v}_5 = \big(a, b, b\big)^T$, $\boldsymbol{v}_6 = \big(b, a, b\big)^T$, $\boldsymbol{v}_7 = \big(b, b, a\big)^T$, $\boldsymbol{v}_8 = \big(b, b, b\big)^T$. Then we have $\boldsymbol{t}_i = \boldsymbol{t}(\boldsymbol{v}_i), i = 1, \ldots, 8$, i.e., $\boldsymbol{t}_1 = \boldsymbol{t}_8 = (1, 1)^T$, $\boldsymbol{t}_2 = (a/b, a/b)^T$, $\boldsymbol{t}_3 = (b/a, 1)^T$, $\boldsymbol{t}_4 = (1, b/a)^T$, $\boldsymbol{t}_5 = (b/a, b/a)^T$, $\boldsymbol{t}_6 = (a/b, 1)^T$, $\boldsymbol{t}_7 = (1, a/b)^T$. It can be seen that the induced experimental region $\boldsymbol{t}(\mathcal{X})$ is given by $\{\boldsymbol{t}(\boldsymbol{x}) : \boldsymbol{x} \in [a, b]^3\} = \mathrm{Conv}\{\boldsymbol{t}_2, \boldsymbol{t}_3, \boldsymbol{t}_4, \boldsymbol{t}_5, \boldsymbol{t}_6, \boldsymbol{t}_7\}$ is a hexagon which is depicted in the right panel of Fig. 1 for the special case $a = 1$ and $b = 2$. In Gaffke et al. (2019) it is shown that for the present model the support of a locally D- or A-optimal design is a subset of the vertices of the ploytope. Note that for the vertices $(a, a, a)^T$ and $(b, b, b)^T$ of the cube $[a, b]^3$ we have $\boldsymbol{t}\big((a, a, a)^T\big) = \boldsymbol{t}\big((b, b, b)^T\big) = (1, 1)^T$ which lies in the interior of the convex hull. Hence, the vertices $(a, a, a)^T$ and $(b, b, b)^T$ of the cube cannot be support points of an optimal design for the model without intercept. For illustration the original and the induced experimental region are depicted in Fig. 1 in the case $a = 1$ and $b = 2$.

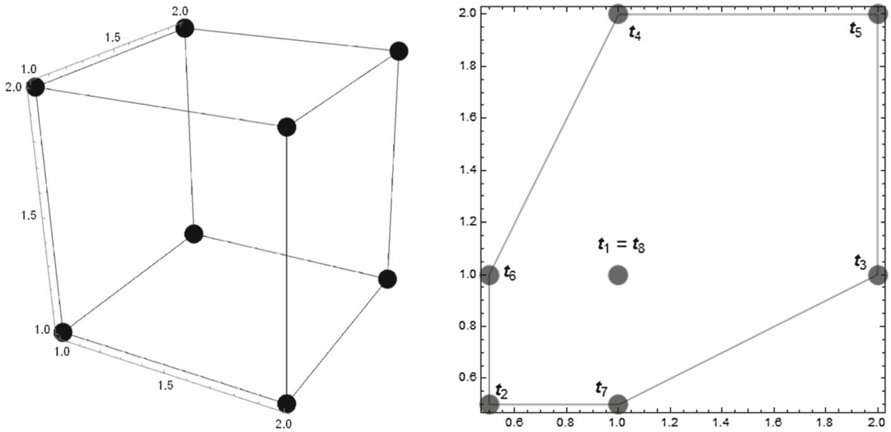


Fig. 1 Left panel: original experimental region $\mathcal{X} = [1, 2]^3$. Right panel: induced experimental region $\boldsymbol{t}(\mathcal{X})$

In the following we will restrict to the standardized experimental region $\mathcal{X} = [1, 2]^3$ for illustrative purposes. 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 $\beta_2 = \beta_3 = \beta$. Then the set of all parameter points of this type satisfying condition (2.3), i.e., $\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 > 0$ for all $\boldsymbol{x} = (x_1, x_2, x_3)^T \in \mathcal{X}$, is characterized by

$$\big(\beta_1 \le 0 \text{ and } \beta > -\beta_1\big) \text{ or } \big(\beta_1 > 0 \text{ and } \beta > -\beta_1/4\big)$$

which is the region above the curve shown in Fig. 2.

The vertices of $\mathcal{X} = [1, 2]^3$ are given by $\boldsymbol{v}_1 = \big(1, 1, 1\big)^T$, $\boldsymbol{v}_2 = \big(2, 1, 1\big)^T$, $\boldsymbol{v}_3 = \big(1, 2, 1\big)^T$, $\boldsymbol{v}_4 = \big(1, 1, 2\big)^T$, $\boldsymbol{v}_5 = \big(1, 2, 2\big)^T$, $\boldsymbol{v}_6 = \big(2, 1, 2\big)^T$, $\boldsymbol{v}_7 = \big(2, 2, 1\big)^T$, $\boldsymbol{v}_8 = \big(2, 2, 2\big)^T$ with intensities $u_i = u(\boldsymbol{v}_i, \boldsymbol{\beta})$, $i = 1, \ldots, 8$. Note that the region shown in Fig. 2 is the parameter space of $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ restricted to the case $\beta_2 = \beta_3 = \beta$. We aim at finding locally D-optimal designs for each given parameter point in that space. The expression " optimality subregion" will be used to refer to a subset of parameter points where a minimally supported design or, generally, designs with similar support are locally D-optimal.

In the next theorem we present designs which are locally D-optimal on the corresponding optimalty subregions.

Theorem 3.3 *Let a gamma model be given by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ on the experimental region $\mathcal{X} = [1, 2]^3$. Then the following designs are locally D-optimal for the specified values of $\boldsymbol{\beta} = (\beta_1, \beta, \beta)^T$.*

(i) $\xi_1^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ *for* $(\beta > 0,\ \beta_1 = 0)$ *or* $(\beta \ge -3\beta_1,\ \beta_1 < 0)$ *or* $(\beta \ge \beta_1/5,\ \beta_1 > 0)$.

(ii) $\xi_2^* = \begin{pmatrix} \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_5 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ *for* $-\beta_1/4 < \beta \le -5\beta_1/23$, $\beta > 0$.

Table 1 Order of intensity values on the optimality subregions corresponding to the D-optimal designs of Theorem 3.3

Optimality subregions	Order of intensities	D-optimal design
$\beta_1 = 0, \beta > 0$	$u_2 > u_3 = u_4 = u_6 = u_7 > u_5$	ξ_1^*
$\beta_1 < 0, \beta \geq -3\beta_1$	$u_2 > u_6 = u_7 \approx u_3 = u_4 > u_5$	ξ_1^*
$\beta_1 > 0, \beta \geq \beta_1/5$	$u_2 > u_3 = u_4 > u_6 = u_7 > u_5$	ξ_1^*
$\beta_1 > 0, -\beta_1/4 < \beta \leq -5\beta_1/23$	$u_5 > u_3 = u_4 > u_6 = u_7 > u_2$	ξ_2^*
$\beta_1 > 0, -5\beta_1/23 < \beta < \beta_1/5$	$u_3 = u_4 \geq u_5 > u_2 \geq u_6 = u_7$	ξ_3^*
$\beta_1 < 0, -\beta_1 < \beta \leq -6\beta_1/5$	$u_2 > u_6 = u_7 > u_3 = u_4 > u_5$	ξ_4^*
$\beta_1 < 0, -3\beta_1 < \beta < -6\beta_1/5$	$u_2 > u_6 = u_7 > u_3 = u_4 > u_5$	ξ_5^*

(iii) $\xi_3^* = \xi_3^*(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_5 \\ \frac{5+23\gamma}{16(1+4\gamma)} & \frac{9(1+3\gamma)^2}{32(1+\gamma)(1+4\gamma)} & \frac{9(1+3\gamma)^2}{32(1+\gamma)(1+4\gamma)} & \frac{1-\gamma-20\gamma^2}{8(1+\gamma)(1+4\gamma)} \end{pmatrix}$ *where* $\gamma = \frac{\beta}{\beta_1}$ *for* $-5\beta_1/23 < \beta < \beta_1/5$, $\beta_1 > 0$.

(iv) $\xi_4^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_6 & \boldsymbol{v}_7 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ *for* $-\beta_1 < \beta \leq -6\beta_1/5$, $\beta_1 < 0$.

(v) $\xi_5^* = \xi_5^*(\boldsymbol{\beta})$ *supported on* $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5, \boldsymbol{v}_6, \boldsymbol{v}_7$ *for* $-3\beta_1 < \beta < -6\beta_1/5$, $\beta_1 < 0$.

The proof of Theorem 3.3 is provided in the "Appendix". Table 1 presents the order of the intensities in all optimality subregions and the corresponding D-optimal designs introduced in the theorem. The intensities for the vertices $\boldsymbol{v}_1$ and $\boldsymbol{v}_8$ are immaterial because of the reduction in Remark 3.3. It can be noted that on each optimality subregion the vertices with highest intensities appear in most cases as a support of the corresponding D-optimal design.

Remark 3.4 For the subregion $-3\beta_1 < \beta < -6\beta_1/5$, $\beta_1 < 0$ in part (v) of Theorem 3.3 explicit analytic solutions cannot be obtained for the characterization of locally D-optimal designs $\xi_5^* = \xi_5^*(\boldsymbol{\beta})$. In that case optimal weights have to be derived numerically (cp. Example 3.1).

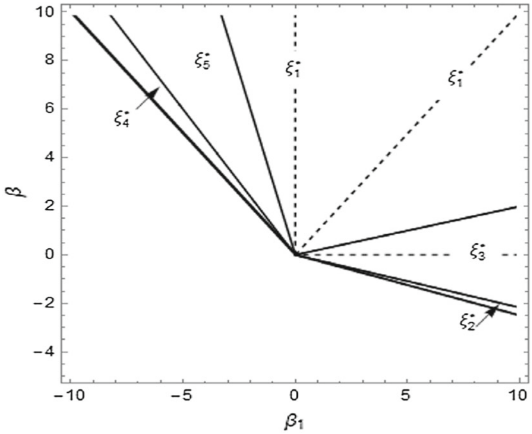
Remark 3.5 (i) Theorem 3.3 constitutes a complete case-by-case analysis of all possible parameter values satisfying $\boldsymbol{x}^T\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$.

(ii) All conditions on β_1 as well as the weights in case (iii) (and in case (v)) can be rephrased as conditions on $\gamma = \beta/\beta_1$.

(iii) Note that in case (iii) the weights of the vertices $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ are equal by symmetry considerations w.r.t. to permutation of the second and third entry x_2 and x_3 in $\boldsymbol{x}$ which is possible because of $\beta_2 = \beta_3$ (see also for similar arguments ξ_5^* below). Moreover, it can be seen that in case (iii) the weights are continuous in $\boldsymbol{\beta}$ where, in particular for $\gamma \to 1/5$ the weight of $\boldsymbol{v}_4$ tends to 0 and for $\gamma \to -5/23$ the weight of $\boldsymbol{v}_2$ becomes 0.

In Fig. 2 the optimality subregions of $\xi_1^*, \ldots, \xi_5^*$ specified in Theorem 3.3 are depicted. Note that each design of ξ_1^*, ξ_2^* and ξ_4^* denotes a single design whereas ξ_3^*

Fig. 2 Optimality subregions for the locally D-optimal designs of Theorem 3.3. The dashed lines are; diagonal: $\beta = \beta_1$, vertical: $\beta_1 = 0$, horizontal: $\beta = 0$



and ξ_5^* determine certain types of designs with weights depending on the parameter values. A particular form of ξ_3^* is obtained at $\beta = (-1/7)\beta_1$ (i.e., $\gamma = -1/7$). In this case ξ_3^* is the uniform design on the vertices $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$ with equal weights 1/4. Note that, in general, the optimal weights are constant along each ray $\beta = \gamma\beta_1$ for fixed γ. In particular, along the horizontal dashed line, i.e., $\beta = 0$, ξ_3^* assigns the weights $\omega_2^* = 5/16, \omega_3^* = \omega_4^* = 9/32, \omega_5^* = 1/8$ to $\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5$, respectively. For the particular case of equal size of the parameter values, i.e., $\beta_1 = \beta$ represented by the diagonal dashed line in Fig. 2 the minimally suppported design ξ_1^* is D-optimal which is supported by those vertices for which all but one component is on the lower setting $a = 1$.

Example 3.1 For a given parameter value in the subregion $-3\beta_1 < \beta < -6\beta_1/5$, $\beta_1 < 0$ specified by part (v) of Theorem 3.3 the weights for the locally D-optimal design ξ_5^* cannot be obtained analytically. Therefore, employing the multiplicative algorithm (see Yu 2010; Harman and Trnovská 2009) in the software package **R** (see R Core Team 2019) provides numerical solutions which show that the vertex $\boldsymbol{v}_5$ does not perform as a support of the locally D-optimal design ξ_5^* on that subregion and it is only supported by the other five vertices with weights depending on $\boldsymbol{\beta}$, where

$$\xi_5^* = \begin{pmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_6 & \boldsymbol{v}_7 \\ \omega_2^* & \omega_3^* & \omega_3^* & \omega_6^* & \omega_6^* \end{pmatrix}.$$

Here the optimal weights of $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ resp. $\boldsymbol{v}_6$ and $\boldsymbol{v}_7$ coincide in view of symmetry considerations w.r.t. to permutation of the second and third component x_2 and x_3 in $\boldsymbol{x}$ which can be applied because of $\beta_2 = \beta_3 = \beta$. By the above mentioned reduction in the parameter space by dividing the vector $\boldsymbol{\beta}$ by its first component it is obvious that the weights only depend on $\gamma = \beta/\beta_1$. Table 2 shows some numerical results for the optimal weights at various values of $\gamma \in (-3, -6/5)$. As a result, we can conjecture that the weight of $\boldsymbol{v}_5$ equals 0 for all $-3 < \gamma < -6/5$.

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

γ	$\boldsymbol{v}_2$	$\boldsymbol{v}_3$	$\boldsymbol{v}_4$	$\boldsymbol{v}_6$	$\boldsymbol{v}_7$
– 3.00	0.3333	0.3333	0.3333	0.0000	0.0000
– 2.90	0.3312	0.3285	0.3285	0.0059	0.0059
– 2.50	0.3225	0.3051	0.3051	0.0336	0.0336
– 2.00	0.3125	0.2604	0.2604	0.0833	0.0833
– 1.50	0.3125	0.1701	0.1701	0.1736	0.1736
– 1.23	0.3297	0.0325	0.0325	0.3027	0.3027
– 1.25	0.3333	0.0000	0.0000	0.3333	0.3333

Remark 3.6 The results of Theorem 3.3 and Example 3.1 can be directly transformed to general cubes $[a, b]^3$ when the ratio b/a equals 2. Because of the invariance of $\boldsymbol{f}_{\boldsymbol{\beta}}$ with respect to rescaling by a the D-optimal designs will also be obtained by rescaling the optimal designs in Theorem 3.3 and Example 3.1 by a, i.e., the support points of the D-optimal design stay at the same relative position of the cube and the corresponding weights are kept fixed. For general a and b D-optimal designs can be obtained similarly to Theorem 3.3 and Example 3.1 which depend on the sign of β_1 and the ratios $\gamma = \beta/\beta_1$ and b/a only. The structure of the optimality subregions apparently remains similar to that exhibited in Fig. 2.

In general, for gamma models without intercept, finding optimal designs for a model with multiple factors, i.e., $\nu > 3$ is not an easy task. The optimal design given by part (i) of Theorem 3.3 can be generalized to an arbitrary number of factors under sufficient and necessary conditions on the parameter points:

Theorem 3.4 *Let a gamma model be given by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ on the experimental region $\mathcal{X} = \big[a, b\big]^\nu$, $\nu \geq 2$, $0 < a < b$. Let $\boldsymbol{\beta}$ be a parameter point such that $\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta} > 0$ for all $\boldsymbol{x} \in \mathcal{X}$. Then the design ξ^* which assigns equal weights $\omega_i = 1/\nu,\ i = 1, \ldots, \nu$ to the vertices $\boldsymbol{x}_1^* = \big(b, a, \ldots, a\big)^T$, $\boldsymbol{x}_2^* = \big(a, b, \ldots, a\big)^T$, $\ldots$, $\boldsymbol{x}_\nu^* = \big(a, a, \ldots, b\big)^T$ is locally D-optimal (at $\boldsymbol{\beta}$) if and only if*

$$\sum_{j=1}^{\nu}\Big(x_j - \frac{a\sum_{i=1}^{\nu} x_i}{(\nu-1)a+b}\Big)^2\Big((b-a)\beta_j + a\sum_{i=1}^{\nu}\beta_i\Big)^2 \leq (b-a)^2\Big(\sum_{j=1}^{\nu}\beta_j x_j\Big)^2. \quad (3.3)$$

for all $\boldsymbol{x} = (x_1, \ldots, x_\nu)^T \in \{a, b\}^\nu$.

Proof Let $T(\boldsymbol{x}) = \sum_{i=1}^{\nu} x_i$, $q = \frac{a}{(\nu-1)a+b}$ and $c_j = (b-a)\beta_j + a\sum_{i=1}^{\nu}\beta_i,\ j = 1, \ldots, \nu$. Define the $\nu \times \nu$ design matrix $\boldsymbol{F} = [\boldsymbol{f}(\boldsymbol{x}_1^*), \ldots, \boldsymbol{f}(\boldsymbol{x}_\nu^*)]^T$. Thus we have $\boldsymbol{F} = (b-a)\boldsymbol{I}_\nu + a\boldsymbol{1}\boldsymbol{1}^T$ and $\boldsymbol{F}^{-1} = \frac{1}{(b-a)}\big(\boldsymbol{I}_\nu - q\boldsymbol{1}\boldsymbol{1}^T\big)$ where $\boldsymbol{1}$ is a $\nu \times 1$ vector of ones. The information matrix of ξ^* is given by $\boldsymbol{M}\big(\xi^*, \boldsymbol{\beta}\big) = \frac{1}{\nu}\boldsymbol{F}^T\boldsymbol{V}\boldsymbol{F}$ where $\boldsymbol{V} = \mathrm{diag}\Big(u(\boldsymbol{x}_j^*, \boldsymbol{\beta})\Big)_{j=1}^{\nu}$ is the $\nu \times \nu$ weight diagonal matrix. Note that $u(\boldsymbol{x}_j^*, \boldsymbol{\beta}) = c_j^{-2}$

for all $j = 1, \ldots, \nu$. Hence, the l.h.s. of the condition of the equivalence theorem (Theorem 2.1, part (a)) is equal to

$$\big(\sum_{j=1}^{\nu} \beta_j x_j\big)^{-2} \boldsymbol{f}^T(\boldsymbol{x}) \boldsymbol{M}^{-1}\big(\xi^*, \boldsymbol{\beta}\big) \boldsymbol{f}(\boldsymbol{x}) = \nu \big(\sum_{j=1}^{\nu} \beta_j x_j\big)^{-2} \boldsymbol{f}^T(\boldsymbol{x}) \boldsymbol{F}^{-1} \boldsymbol{V}^{-1} \boldsymbol{F}^{-1} \boldsymbol{f}(\boldsymbol{x})$$

Since $\boldsymbol{F}^{-1} \boldsymbol{f}(\boldsymbol{x}) = (b-a)^{-1} \boldsymbol{f}(\boldsymbol{x}) - qT(\boldsymbol{x})\mathbf{1}$, the above is equal to

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

By the equivalence theorem the design ξ^* is locally D-optimal if and only if the above expression is less than or equal to ν for all $\boldsymbol{x} \in \{a, b\}^{\nu}$ which is equivalent to condition (3.3). □

Note that the D-optimal design given in part (i) of Theorem 3.3 is a special case of Theorem 3.4 when $\nu = 3$. In this case condition (3.3) is equivalent to condition (6.4) in the proof of part (i) of Theorem 3.3 (see the "Appendix"). It can be seen that in Theorem 3.4 the optimality condition (3.3) depends only on the ratios $\beta_j / (\sum_{i=1}^{\nu} \beta_i)$ for all $j = 1, \ldots, \nu$. Similarly note that already condition (3.3) depends on a and b only through their ratio a/b. However, assuming the model parameters have equal size implies that the D-optimality of a design is independent of the model parameters whereas it depends on the ratio a/b as it is shown in the next corollary.

Corollary 3.3 *Let a gamma model be given by* $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$ *on the experimental region* $\mathcal{X} = \big[a, b\big]^{\nu}$, $\nu \geq 2$, $0 < a < b$. *Let* $\boldsymbol{\beta}$ *be a parameter point such that all* β_j *are equal, i.e.,* $\beta_j = \beta > 0$ *say, for all* $j = 1, \ldots, \nu$. *Then the design* ξ^* *which assigns equal weights* $\omega_i = 1/\nu$, $i = 1, \ldots, \nu$ *to the support points* $\boldsymbol{x}_1^* = \big(b, a, \ldots, a\big)^T$, $\boldsymbol{x}_2^* = \big(a, b, \ldots, a\big)^T, \ldots, \boldsymbol{x}_\nu^* = \big(a, a, \ldots, b\big)^T$ *is locally D-optimal (at* $\boldsymbol{\beta}$*) if and only if*

$$\Big(\frac{b}{a}\Big)^2 \geq \frac{\big(\nu - 1\big)\big(\nu - 2\big)}{2}. \tag{3.4}$$

Proof For equal components $\beta_j = \beta > 0$, $j = 1, \ldots, \nu$ condition (3.3) of Theorem 3.4 reduces to

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

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 $\sum_{j=1}^{\nu} x_j^2 = (\nu - r)\, a^2 + r\, b^2$ and

$\left(\sum_{j=1}^{\nu} x_j\right)^2 = ((\nu - r)\, a + r\, b)^2$. Hence, the condition above is equivalent to

$$(a-b)^2\, \tau\, r^2 + (a-b)((b+a) - 2\, a\, \nu\, \tau)\, r + \nu\, a^2(\nu\, \tau - 1) \ \geq 0 \ \forall r \in \{0, 1, \ldots, \nu\}, \tag{3.5}$$

where $\tau = \frac{(\nu-1)a^2+b^2}{((\nu-1)a+b)^2}$. The l.h.s. of inequality (3.5) is a polynomial in r of degree 2 with positive leading term. The polynomial attains 0 at $r = 1$ which indicates the support of ξ^* and $r = \frac{\nu\,(\nu-1)\,a^2}{(\nu-1)a^2+b^2}$. Hence the condition in (3.5) holds when for the second root $\frac{\nu\,(\nu-1)\,a^2}{(\nu-1)a^2+b^2} \leq 2$ which coincides with condition (3.4). □

Remark 3.7 For $\nu = 2$ and $\nu = 3$ the right hand side of condition (3.4) equals 0 and 1, respectively. Hence, in these cases the condition is obviously fulfilled for all $0 < a < b$.

4 Gamma model with interaction

In this section we consider a two-factor model without intercept but with an additional interaction term such that $\boldsymbol{x} = (x_1, x_2)^T$, $\boldsymbol{f}(\boldsymbol{x}) = \left(x_1, x_2, x_1 x_2\right)^T$ and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \beta_3\right)^T$. The experimental region is given by a square $\mathcal{X} = [a, b]^2$, $0 < a < b$ and the vertices are denoted by $\boldsymbol{v}_1 = (b, b)^T$, $\boldsymbol{v}_2 = (b, a)^T$, $\boldsymbol{v}_3 = (a, b)^T$, $\boldsymbol{v}_4 = (a, a)^T$. We aim at deriving locally D-optimal designs. To this end we develop a solution by removing the interaction term $x_1 x_2$ by a transformation of the present model to a two-factor model with intercept and without interaction. This transformation can be accomplished by using the structure of the underlying gamma models. As it was pointed out in Remark 2.4 the function $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ is invariant w.r.t. simultaneous scaling of $\boldsymbol{x}$, i.e., $\boldsymbol{f}_{\boldsymbol{\beta}}(\lambda \boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})$ for any $\lambda > 0$ where λ may be chosen for each $\boldsymbol{x}$ individually. Let here $\lambda = \lambda_{\boldsymbol{x}} = 1/(x_1 x_2)$ then we obtain

$$\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \left(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2\right)^{-1} (x_1, x_2, x_1 x_2)^T \tag{4.1}$$

$$= \left(\beta_1 t_2 + \beta_2 t_1 + \beta_3\right)^{-1} \, (t_2, t_1, 1)^T \tag{4.2}$$

where $\boldsymbol{t} = \left(t_1, t_2\right)^T$, $t_j = 1/x_j$, $j = 1, 2$. The range $\mathcal{T} = \boldsymbol{t}(\mathcal{X})$ of $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$ is a cube given by $\mathcal{T} = \left[(1/b)\,,\ (1/a)\right]^2$ as $\boldsymbol{x}$ ranges over $\mathcal{X} = [a, b]^2$. By simultaneous permutation of the components in the regression function corresponding to $\boldsymbol{t}$ and the components in $\boldsymbol{\beta}$ we obtain a gamma model with regression function $\tilde{\boldsymbol{f}}(\boldsymbol{t}) = (1, t_1, t_2)^T$ and parameter vector $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)^T = (\beta_3, \beta_2, \beta_1)^T$ with corresponding intensity function $\tilde{u}(\boldsymbol{t}, \tilde{\boldsymbol{\beta}}) = (\tilde{\beta}_1 + \tilde{\beta}_2 t_1 + \tilde{\beta}_3 t_2)^{-2}$ as in the two-factor model with intercept and without interaction. In the following the locally D-optimal designs are given.

Theorem 4.1 *Consider* $\boldsymbol{f}(\boldsymbol{x}) = \left(x_1, x_2, x_1 x_2\right)^T$ *on* $\mathcal{X} = [a, b]^2$, $0 < a < b$. *Let* $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ *be a parameter point such that condition* (2.3) *is satisfied. 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* $(b, b)^T$, $(b, a)^T$ *and* $(a, b)^T$.

(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* $(b, b)^T$, $(b, a)^T$ *and* $(a, a)^T$.

(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* $(b, b)^T$, $(a, b)^T$ *and* $(a, a)^T$.

(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* $(b, a)^T$, $(a, b)^T$ *and* $(a, a)^T$.

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

$$\xi^* = \begin{pmatrix} (b, b)^T & (b, a)^T & (a, b)^T & (a, a)^T \\ \omega_1^* & \omega_2^* & \omega_3^* & \omega_4^* \end{pmatrix},$$

$$\text{where } \omega_\ell^* > 0 \ (1 \le \ell \le 4),\ \textstyle\sum_{\ell=1}^4 \omega_\ell^* = 1.$$

Proof The optimality problem can be transformed to that under the function $\tilde{\boldsymbol{f}}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{t})$ with the intensity function $\tilde{u}(\boldsymbol{t}, \tilde{\boldsymbol{\beta}}) = (\beta_3 + \beta_2 t_1 + \beta_1 t_2)^{-2}$ on the experimental region $\mathcal{T} = \left[(1/b)\,,\,(1/a)\right]^2$. The vertices are given by $\boldsymbol{t}_1 = (1/b, 1/b)^T$, $\boldsymbol{t}_2 = (1/a, 1/b)^T$, $\boldsymbol{t}_3 = (1/b, 1/a)^T$ and $\boldsymbol{t}_4 = (1/a, 1/a)^T$ with corresponding intensities $\tilde{u}_1 = (\beta_3 + \frac{1}{b}(\beta_1 + \beta_2))^{-2}, \tilde{u}_2 = (\beta_3 + \beta_1\frac{1}{a} + \beta_2\frac{1}{b})^{-2}, \tilde{u}_3 = (\beta_3 + \beta_1\frac{1}{b} + \beta_2\frac{1}{a})^{-2}$ and $\tilde{u}_4 = (\beta_3 + \frac{1}{a}(\beta_1 + \beta_2))^{-2}$. According to part (i) of the lemma in the "Appendix", straightforward computations can show that the condition in case (i) of the theorem is equivalent to $u_4^{-1} \ge u_1^{-1} + u_2^{-1} + u_3^{-1}$. Analogous verifying is obtained for other cases (ii), (iii), (iv). Case (v) follows from part (ii) of that lemma. □

It is noted that the optimality conditions (i)–(iv) provided by Theorem 4.1 depend on the values of a and b. Changing these values might affect the D-optimality of a design since its optimality condition will no longer be fulfilled. 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 $\gamma_2 = \beta_2/\beta_3$, $\beta_3 \neq 0$. Here, the parameter space which is depicted in the left panel of Fig. 3 is characterized by $\gamma_2 + \gamma_1 > -1$, $2\,\gamma_2 + \gamma_1 > -2$ and $\gamma_2 + 2\,\gamma_1 > -2$. It is observed that from the left panel that the design given by part (i) of Theorem 4.1 is not locally D-optimal because the corresponding optimality condition $\frac{1}{4}(\gamma_1^2 + \gamma_2^2) + \frac{1}{4}\gamma_1\gamma_2 + \gamma_1 + \gamma_2 \le -1$ cannot be satisfied.

Let us consider another experimental region as a square with a higher length by fixing $a = 1$ and taking $b = 4$, i.e., $\mathcal{X} = [1, 4]^2$. The parameter space which is depicted in the right panel of Fig. 3 is characterized by $\gamma_2 + \gamma_1 > -1$, $4\,\gamma_2 + \gamma_1 > -4$ and $\gamma_2 + 4\,\gamma_1 > -4$. In this situation all cases of designs given by Theorem 4.1 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.

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 4.1 *Consider* $\boldsymbol{f}(\boldsymbol{x}) = (x_1, x_2, x_1x_2)^T$ *on* $\mathcal{X} = [a, b]^2$, $0 < a < b$. *Let* $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ *be a parameter point according to condition* (2.3) *such that* $\beta_1 =$

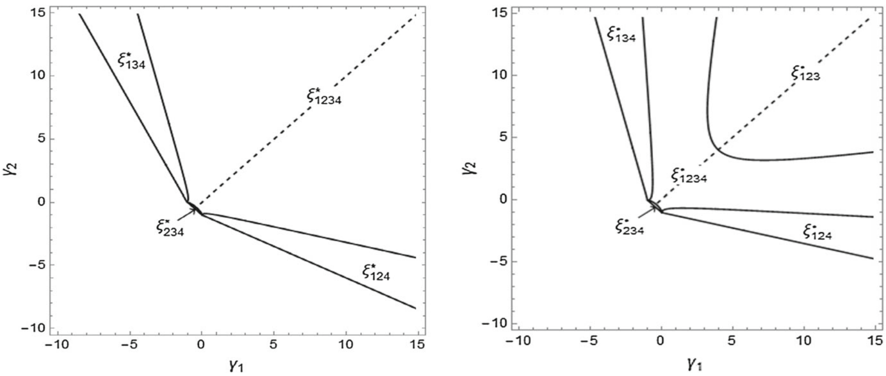


Fig. 3 Optimality subregions for locally D-optimal designs of Theorem 4.1 with $\mathcal{X} = [1, 2]^2$ (left panel) and $\mathcal{X} = [1, 4]^2$ (right panel) where $\gamma_1 = \beta_1/\beta_3$ and $\gamma_2 = \beta_2/\beta_3$. Note that $\mathrm{supp}(\xi^*_{ijk}) = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_k\} \subset \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ and $\mathrm{supp}(\xi^*_{1234}) = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$

$\beta_2 = \beta$ and $\beta_3 \neq 0$. Define $\gamma = \frac{\beta}{\beta_3}$. Then the following designs are locally D-optimal for the specified values of γ.

(i) *If $b - 3a > 0$ and $\gamma \geq \frac{ab}{b-3a}$, then ξ^*_1 assigns equal weights $1/3$ to $(b, b)^T$, $(b, a)^T$ and $(a, b)^T$.*
(ii) *If $-\frac{a}{2} < \gamma \leq -\frac{ab}{3b-a}$, then ξ^*_2 assigns equal weights $1/3$ to $(b, a)^T$, $(a, b)^T$ and $(a, a)^T$.*
(iii) *If $b - 3a > 0$ and $-\frac{ab}{3b-a} < \gamma < \frac{ab}{b-3a}$ then*

$$\xi^*_3 = \xi^*_3(\boldsymbol{\beta}) = \begin{pmatrix} (b, b)^T & (b, a)^T & (a, b)^T & (a, a)^T \\ \frac{ab-(a-3b)\gamma}{4b(a+2\gamma)} & \frac{(ab+(a+b)\gamma)^2}{4ab(b+2\gamma)(a+2\gamma)} & \frac{(ab+(a+b)\gamma)^2}{4ab(b+2\gamma)(a+2\gamma)} & \frac{ab-(b-3a)\gamma}{4a(b+2\gamma)} \end{pmatrix}.$$

Proof 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 Theorem 4.1 we show the following under the assumptions of Corollary 4.1. Both conditions provided in parts (*ii*) and (*iii*) of Theorem 4.1 are not fulfilled by any parameter point thus the corresponding designs are not D-optimal. In contrast, the design ξ^*_1 in (*i*) of Corollary 4.1 is locally D-optimal if the condition provided in part (i) of Theorem 4.1 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 a polynomial in γ of degree 2 and thus the inequality is fulfilled for $\gamma \geq \frac{ab}{b-3a}$ if $b - 3a > 0$. Similarly, the design ξ^*_2 given in (*ii*) of Corollary 4.1 is locally D-optimal if the condition provided in part (iv) of Theorem 4.1 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 a polynomial in γ of degree 2 and thus the inequality is fulfilled for $-\frac{a}{2} < \gamma \leq -\frac{ab}{3b-a}$.
The four-point design ξ_3^* given in (iii) has positive weights on $-\frac{ab}{3b-a} < \gamma < \frac{ab}{b-3a}$ if $b - 3a > 0$ and hence it is implicitly locally D-optimal in view of Remark 2.2. □

Remark 4.1 By Corollary 4.1 for $\beta_1 = \beta_2 = 0$, the uniform design on the vertices $(b, b)^T$, $(b, a)^T$, $(a, b)^T$ and $(a, a)^T$ is locally D-optimal.

5 Design efficiency

The D-optimal design for gamma models depends on a given value of the parameter $\boldsymbol{\beta}$. Misspecified values may lead to a poor performance of the locally optimal design. By the above results the designs are locally D-optimal in a specific subregion of the parameter space. In this section we discuss the potential merits of the derived designs, in particular, the D-optimal designs from Theorem 3.3 for a three-factor gamma model without interactions and from Corollary 4.1 for a two-factor gamma model with interaction. Our objective is to examine the overall performance of some of the locally D-optimal designs. The overall performance of any design ξ is described by its D-efficiencies, as a function of $\boldsymbol{\beta}$,

$$\mathrm{Eff}(\xi, \boldsymbol{\beta}) = \left(\frac{\det \boldsymbol{M}(\xi, \boldsymbol{\beta})}{\det \boldsymbol{M}(\xi^*_{\boldsymbol{\beta}}, \boldsymbol{\beta})} \right)^{1/p}$$

where in both cases $p = 3$ and $\xi^*_{\boldsymbol{\beta}}$ denotes the locally D-optimal design at $\boldsymbol{\beta}$.

Example 5.1 In the situation of Theorem 3.3 the experimental region is given by $\mathcal{X} = [1, 2]^3$. We restrict to the case $\beta_1 > 0$, $\beta_2 = \beta_3 = \beta$ and use the ratio $\gamma = \beta/\beta_1$ with range $(-1/4, \infty)$. Our interest is in the minimally supported and equally weighted designs $\xi_1 = \xi_1^*$ and $\xi_2 = \xi_2^*$ where $\mathrm{supp}(\xi_1) = \{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ with $\mathrm{supp}(\xi_2) = \{\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$ which by Theorem 3.3 parts (i) and (ii) are locally D-optimal for $\gamma \geq 1/5$ and $\gamma \in (-1/4, -5/23]$, respectively. In particular, ξ_1 and ξ_2 are robust against misspecification of the parameter values in their respective subregions. Additionally, for $\gamma \in (-5/23, 1/5)$ we consider the locally D-optimal designs of type $\xi_3(\gamma) = \xi_3^*(\gamma)$ given by Theorem 3.3 part (iii). Note that $\mathrm{supp}(\xi_3(\gamma)) = \{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$ and the weights depend on γ.

For calculating the efficiency we use $\xi^*_{\boldsymbol{\beta}} = \xi_1$ 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 $\gamma \in (-5/23, 1/5)$. For examination of the efficiency we select the designs $\xi_1, \xi_2, \xi_3(-1/7)$ which are uniform on $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$, $\{\boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$ and $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5\}$, respectively. Moreover, as further natural competitors we choose various uniform designs supported by specific subsets of the vertices. Those are the full factorial ξ_4 which is uniform on all eight vertices $\{1, 2\}^3$ and the two corresponding half-fractions; 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 the design ξ_7 which is uniform on the equidistant grid $\{1, 1.5, 2\}^3$.

In the left panel of Fig. 4, the D-efficiencies of the designs $\xi_1, \xi_2, \xi_3(-1/7), \xi_4, \xi_5, \xi_6$ and ξ_7 are depicted. The efficiencies of ξ_1 and ξ_2 are as to be expected equal to 1 in their optimality subregions $\gamma \in [1/5, \infty)$ and $\gamma \in (-1/4, -5/23]$, respectively. Moreover, 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 larger 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 far away from the respective optimality subregion. So, in total, the overall performance of ξ_1 and ξ_2 cannot be regarded as satisfactory if no prior knowledge is available for the parameter values. The design $\xi_3(-1/7)$, though locally D-optimal only at $\gamma = -1/7$, does show a more satisfactory overall performance with efficiency ranging between 0.8585 and 1. The efficiency of the half-fractional design ξ_6 exceeds 0.80 only for $\gamma > -0.049$, while for smaller values of γ the efficiency decreases to zero. The full-factorial design ξ_4 turns out to be uniformly worse than $\xi_3(-1/7)$ and its efficiency ranges between 0.5768 and 0.7615. The worst performance is shown by the half-fraction ξ_5 and the uniform design ξ_7 on the grid.

Example 5.2 In the situation of Corollary 4.1 we consider the experimental region $\mathcal{X} = [1, 4]^2$ where condition $b - 3a > 0$ is satisfied in parts (i) and (iii). The vertices are denoted as before by $\boldsymbol{v}_1 = (4, 4)^T$, $\boldsymbol{v}_2 = (4, 1)^T$, $\boldsymbol{v}_3 = (1, 4)^T$, $\boldsymbol{v}_4 = (1, 1)^T$. 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 5.1 denote by $\xi_1 = \xi_1^*$ and $\xi_2 = \xi_2^*$ the minimally supported 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 $\xi_3(\gamma) = \xi_3^*(\gamma)$ the design given in part (iii) of Corollary 4.1 which is locally D-optimal at $\gamma \in (-4/11, 4)$. For the calculation of the efficiency the optimal designs are given by $\xi_{\boldsymbol{\beta}}^* = \xi_1$ 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 and the design $\xi_3(0)$ which is uniform on the set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ of all four vertices. As a natural competitor we choose the design ξ_4 which is uniform on the equidistant grid $\{1, 2.5, 4\}^2$. The efficiencies are depicted in the right panel of Fig. 4. We observe that the performance of ξ_1 and ξ_2 is similar to that of the corresponding designs in Example 5.1. Moreover, the design $\xi_3(0)$ shows a more satisfactory overall performance. The efficiency of ξ_4 varies between 0.77 and 0.83 for $\gamma > -4/11$. The worst performance is shown by the design ξ_4.

6 Discussion

In the present paper we considered gamma models without intercept for which locally D- and A-optimal designs have been developed. The positivity of the expected means entails a positive linear predictor and the absence of the intercept term requires additionally an experimental region which does not contain the origin $\mathbf{0}$. The information matrix for the non-intercept gamma model is invariant w.r.t. simultaneous scaling of the components of $\boldsymbol{x}$ and equivariant w.r.t. to simultaneous scaling of the components of $\boldsymbol{\beta}$. Various approaches were utilized to derive locally optimal designs. The optimality problem under gamma models without intercept was transformed to that under

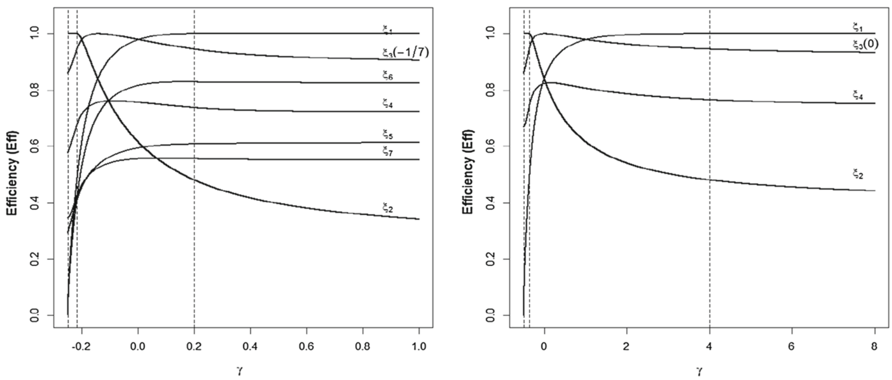


Fig. 4 D-efficiency for Examples 5.1 (left panel) and 5.2 (right panel) in dependence on the ratio $\gamma = \beta/\beta_1$ and $\gamma = \beta/\beta_3$, respectively. Vertical dotted lines indicate the boundaries of the subregions

corresponding gamma models which involve an intercept and defined on a transformed experimental region. This approach simplified the optimality problem and thus known results could be applied. In this context we considered in particular two-factor models without and with interaction in Theorems 3.2 and 4.1, respectively.

In another context, as in Theorem 3.3 we made use of the equivalence theorem to establish the local D-optimality under a non-intercept gamma model with three factors. Due to the complexity of this approach the optimization problem was identified by a system of inequalities which was solved analytically or by employing computer algebra. In contrast to that, the optimality problem could be transformed for a two-factor gamma model with intercept in view of Remark 3.3 where the experimental region $\mathcal{X} = [1, 2]^3$ as considered in Theorem 3.3 can be reduced to the polytope $\mathcal{T} = \text{Conv}\big\{(1/2, 1)^T, (1, 1/2)^T, (1/2, 1/2)^T, (2, 1)^T, (1, 2)^T, (2, 2)^T\big\}$. Rescaling and shifting $\mathcal{T}$ yields $\mathcal{Z} = \text{Conv}\big\{(0, 1/3)^T, (1/3, 0)^T, (0, 0)^T, (1, 1/3)^T, (1/3, 1)^T, (1, 1)^T\big\}$. Consequently, the linear predictor is reparameterized as $\tilde{\beta}_0 + \tilde{\beta}_1 z_1 + \tilde{\beta}_2 z_2$ where $(z_1, z_2)^T \in \mathcal{Z}$ and $\tilde{\beta}_0 = \beta_1 + (1/2)(\beta_2 + \beta_3)$, $\tilde{\beta}_1 = (3/2)\beta_2$, $\tilde{\beta}_2 = (3/2)\beta_3$, where D-optimal designs are equivariant w.r.t. these transformations.

Practically, there are various link functions that can be considered to fit gamma-distributed observations. The power link family which presents the class of link functions (see Burridge and Sebastiani 1994; Atkinson and Woods 2015, Section 2.5) can be employed. In this case $\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta} = \mu^\rho$ for a given exponent $\rho \in \mathbb{R}$ and $\rho \neq 0$. The intensity function under this family reads as $u_0(\boldsymbol{x}, \boldsymbol{\beta}) = \kappa\rho^{-2}\big(\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta}\big)^{-2}$. Note that $\kappa\rho^{-2}$ is a positive constant and can be ignored. Therefore, the results obtained in this paper still apply to the family of power links.

Additionally, the log-link function can be considered as a main alternative to the canonical one (see Kilian et al. 2002; Wenig et al. 2009; Gregori et al. 2011; McCrone et al. 2005; Montez-Rath et al. 2006). In that case the intensity function is constantly equal to 1 and thus the information matrix under gamma models is equivalent to that under ordinary linear models. For this reason, the optimal designs for a gamma model are identical to those for an ordinary linear model with the same linear predictor. In

Hardin and Hilbe (2018) gamma models were fitted considering the Box–Cox family of link functions;

$$f^T(\boldsymbol{x})\boldsymbol{\beta} = \begin{cases} \left(\mu^\delta - 1\right)/\delta & (\delta \neq 0) \\ \log \mu & (\delta = 0) \end{cases} \tag{6.1}$$

which involves the log-link at $\delta = 0$ (see Atkinson and Woods 2015). The intensity function is thus defined as

$$u_0(\boldsymbol{x}, \delta\boldsymbol{\beta}) = \left(\delta f^T(\boldsymbol{x})\boldsymbol{\beta} + 1\right)^{-2}, \boldsymbol{x} \in \mathcal{X}. \tag{6.2}$$

Here, the positivity condition (2.3) of the expected mean $\mu = E(Y)$ of a gamma distribution is modified to $\delta f^T(\boldsymbol{x})\boldsymbol{\beta} > -1$ for all $\boldsymbol{x} \in \mathcal{X}$. Therefore, the experimental region is relaxed and might be considered as $\mathcal{X} = [0, 1]^\nu$. As an example, consider $f(\boldsymbol{x}) = \boldsymbol{x}$ where $\boldsymbol{x} = (x_1, x_2)^T \in \mathcal{X} = [0, 1]^2$ with vertices $\boldsymbol{v}_1 = (0, 0)^T$, $\boldsymbol{v}_2 = (1, 0)^T$, $\boldsymbol{v}_3 = (0, 1)^T$, $\boldsymbol{v}_4 = (1, 1)^T$. For a given value of δ let $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ satisfy $\delta f^T(\boldsymbol{x})\boldsymbol{\beta} > -1$ for all $\boldsymbol{x} \in \mathcal{X}$ or equivalently $\delta\beta_1 > -1, \delta\beta_2 > -1$, $\delta(\beta_1+\beta_2) > -1$. Let $u_i = u(\boldsymbol{v}_i, \delta\boldsymbol{\beta})$ for all $i = 1, 2, 3, 4$. The equivalence theorem (Theorem 2.1, part (a)) proves the D-optimality of the design ξ^* which assigns equal weights 1/2 to the vertices $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ at $\delta\boldsymbol{\beta}$. This result may be extended for a multiple-factor model as in Theorem 3.1. However, the expression $\delta f^T(\boldsymbol{x})\boldsymbol{\beta} + 1$ could be viewed as a linear predictor of a gamma model with known positive intercept. Adopting the Box–Cox family as a class of link functions for gamma models will be a topic of future research.

Acknowledgements The authors are thankful to Prof. Dr. Norbert Gaffke for many helpful discussions and suggestions. The work of the first author was supported by a scholarship of the state of Saxony-Anhalt.

Appendix

Lemma *Consider a two-factor gamma model with intercept such that* $f(\boldsymbol{x}) = (1, x_1, x_2)^T$ *on the experimental region* $\mathcal{X} = [a, b]^2$, $a, b \in \mathbb{R}$, $a < b$ *with the intensity function* $u(\boldsymbol{x}, \boldsymbol{\beta}) = 1/(\beta_0 + \beta_1 x_1 + \beta_2 x_2)^2$. *The vertices of* $\mathcal{X}$ *are given by* $(a, a)^T$, $(b, a)^T$, $(a, b)^T$ *and* $(b, b)^T$ *with the corresponding intensities* $u_1 = 1/(\beta_0 + \beta_1 a + \beta_2 a)^2$, $u_2 = 1/(\beta_0 + \beta_1 b + \beta_2 a)^2$, $u_3 = 1/(\beta_0 + \beta_1 a + \beta_2 b)^2$ *and* $u_4 = 1/(\beta_0 + \beta_1 b + \beta_2 b)^2$. *Let* $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ *be a parameter point such that* $f^T(\boldsymbol{x})\boldsymbol{\beta} > 0$ *for all* $\boldsymbol{x} \in \mathcal{X}$, *or equivalently,* $\beta_0+\beta_1 a+\beta_2 a > 0$, $\beta_0+\beta_1 b+\beta_2 a > 0$, $\beta_0 + \beta_1 a + \beta_2 b > 0$ *and* $\beta_0 + \beta_1 b + \beta_2 b > 0$. *Then the unique locally D-optimal design* ξ^* *(at* $\boldsymbol{\beta}$*) is as follows.*

(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 vertices 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 vertices* $(a, a)^T$, $(b, a)^T$, $(a, b)^T$, $(b, b)^T$ *with corresponding weights* $\omega_1^*, \omega_2^*, \omega_3^*, \omega_4^* > 0$ *and* $\sum_{k=1}^4 \omega_k^* = 1$.

Proof The proof can be demonstrated by making use of the results of Gaffke et al. (2019) who derived the D-optimal designs for a gamma model with intercept on the standardized experimental region $\mathcal{Z} = [0, 1]^2$. To this end, denote $\boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x}) = (\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta})^{-1}\boldsymbol{f}(\boldsymbol{x})$ so we have $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{f}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{f}_{\boldsymbol{\beta}}^T(\boldsymbol{x})$. Based on equivariance (see Radloff and Schwabe 2016) a D-optimal design ξ^* on $\mathcal{X}$ given in the lemma can be derived by transformation of a respective D-optimal design ξ^{**} on $\mathcal{Z} = [0, 1]^2$. Here, $x_j \to z_j = \frac{x_j}{b-a} - \frac{a}{b-a}$, $j = 1, 2$. For the transformation matrix

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

we have $\boldsymbol{f}(z) = \boldsymbol{B}\boldsymbol{f}(\boldsymbol{x}) = (1, z_1, z_2)^T$ with $\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{B}^T\right)^{-1}\boldsymbol{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)^T$ where $\tilde{\beta}_0 = \beta_0 + (-a/(b-a))\,(\beta_1 + \beta_2)$, $\tilde{\beta}_1 = \beta_1(1/(b-a))$ and $\tilde{\beta}_2 = \beta_2(1/(b-a))$. It follows that $\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(z) = \left(\boldsymbol{f}^T(z)\tilde{\boldsymbol{\beta}}\right)^{-1}\boldsymbol{f}(z)$ and the information matrix is given by $\tilde{\boldsymbol{M}}(z, \tilde{\boldsymbol{\beta}}) = \boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}(z)\boldsymbol{f}_{\tilde{\boldsymbol{\beta}}}^T(z)$. It is easily seen that $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{B}^{-1}\tilde{\boldsymbol{M}}(z, \tilde{\boldsymbol{\beta}})\boldsymbol{B}^{-1}$, thus the derived D-optimal designs on $\mathcal{X}$ and $\mathcal{Z}$, respectively are equivariant. Then the results follow from Theorem 4.2 in Gaffke et al. (2019). □

Proof of Theorem 3.3 First we give an outline of the proof. The proof is obtained by making use of the condition of the equivalence theorem (Theorem 2.1, part (a)). By that we develop a system of feasible inequalities evaluated at the vertices $\boldsymbol{v}_i$ for all $i = 1, \ldots, 8$. For simplification of presentation in the case $\beta_1 \neq 0$ we use the ratio $\gamma = \beta/\beta_1$ for which the range is given by $(-\infty, -1) \cup (-1/4, \infty)$. It turns out that some of the inequalities are implied by some others and thus the resulting system is reduced to an equivalent system of only a few inequalities. The intersection of the set of solutions of the system with the range of γ leads to the optimality condition (subregion) of the corresponding optimal design. For minimally supported designs given in cases (i), (ii), (iv) we display the 3×3 design matrix $\boldsymbol{F}$, its inverse $\boldsymbol{F}^{-1}$ and the 3×3 weight matrix $\boldsymbol{V}$. Note that for $\beta_1 \neq 0$ the intensities $u_i = u(\boldsymbol{v}_i, \boldsymbol{\beta})$, $i = 1, \ldots, 8$ are equal to $u_1 = \beta_1^{-2}(1+2\gamma)^{-2}$, $u_2 = \beta_1^{-2}(2+2\gamma)^{-2}$, $u_3 = u_4 = \beta_1^{-2}(1+3\gamma)^{-2}$, $u_5 = \beta_1^{-2}(1+4\gamma)^{-2}$, $u_6 = u_7 = \beta_1^{-2}(2+3\gamma)^{-2}$, $u_8 = \beta_1^{-2}(2+4\gamma)^{-2}$, respectively.

Proof of part (i): The 3×3 design matrix $\boldsymbol{F} = [\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4]^T$ is given by

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

Hence, the condition of the equivalence theorem is given by

$$\boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{F}^{-1}\boldsymbol{V}^{-1}\left(\boldsymbol{F}^T\right)^{-1}\boldsymbol{f}(\boldsymbol{x}) \leq \left(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3\right)^2 \quad \forall \boldsymbol{x} \in \{1, 2\}^3. \tag{6.3}$$

For the case $\beta > 0, \beta_1 = 0$, condition (6.3) is equivalent to

$$4\big(3x_1-(x_2+x_3)\big)^2+9\Big(\big(3x_2-(x_1+x_3)\big)^2+\big(3x_3-(x_1+x_2)\big)^2\Big)$$
$$\le 16\big(x_2+x_3\big)^2\ \forall \boldsymbol{x}\in\{1,2\}^3,$$

which is independent of β and is satisfied for all vertices $\boldsymbol{v}_i$ for all $i=1,\dots,8$ and equality holds for the support. For the other cases, i.e., $\beta\ge -3\beta_1, \beta_1<0$ or $\beta>\beta_1/5, \beta_1>0$ condition (6.3) is equivalent to

$$\big(3x_1-(x_2+x_3)\big)^2(2+2\gamma)^2+\Big(\big(3x_2-(x_1+x_3)\big)^2+\big(3x_3-(x_1+x_2)\big)^2\Big)(1+3\gamma)^2$$
$$\le 16\big(x_1+\gamma(x_2+x_3)\big)^2\ \forall \boldsymbol{x}\in\{1,2\}^3. \tag{6.4}$$

By some lengthy but straightforward calculations, the above inequalities can be reduced to

$$15\gamma^2+2\gamma-1\ge 0 \text{ and } 3\gamma^2+10\gamma+3\ge 0$$

where the first inequality comes from vertex $\boldsymbol{v}_5$ and the second inequality comes from the vertices $\boldsymbol{v}_6$ and $\boldsymbol{v}_7$. The l.h.s. of each inequality is a polynomial in γ of degree 2 and the sets of solutions are given by $(-\infty,-1/3]\cup[1/5,\infty)$ and $(-\infty,-3]\cup[-1/3,\infty)$, respectively. Note that the 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 $\gamma\in(-\infty,-3]\cup[1/5,\infty)$ which is equivalent to the optimality subregion $\beta\ge -3\beta_1, \beta_1<0$ or $\beta>\beta_1/5, \beta_1>0$ given in part (*i*) of the theorem.
Proof of part (*ii*): The 3×3 design matrix $\boldsymbol{F}=[\boldsymbol{v}_3,\boldsymbol{v}_4,\boldsymbol{v}_5]^T$ is given by

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

Hence, the condition of the equivalence theorem is equivalent to

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

and similar to part (*i*) the above inequalities reduce to

$$69\gamma^2+38\gamma+5\le 0$$

which arises from vertex $\boldsymbol{v}_2$. The set of solutions of the polynomial determined by the l.h.s. is given by $[-1/3,-5/23]$. By considering the intersection with the range of γ, the design ξ_2^* is locally D-optimal if $\gamma\in(-1/4,-5/23]$.
Proof of part (*iii*): Consider design ξ_3^*. Let the associated weights be denoted as $\omega_2^*=(5+23\gamma)/(16(1+4\gamma))$, $\omega_3^*=\omega_4^*=9(1+3\gamma)^2/(32(1+\gamma)(1+4\gamma))$, $\omega_5^*=(1-\gamma-20\gamma^2)/(8(1+\gamma)(1+4\gamma))$ where $\gamma=\beta/\beta_1$. Note that $\omega_2^*>0$ for all $\gamma>-5/23$, $\omega_3^*=\omega_4^*>0$ for all $\gamma\in\mathbb{R}$ and $\omega_5^*>0$ for all $\gamma\in(-1/4,1/5)$, and it is obvious that $\omega_2^*,\omega_3^*,\omega_5^*$ are positive over the interval $(-5/23,1/5)$ and satisfy $\omega_2^*+\omega_3^*+\omega_4^*+\omega_5^*=1$. The 4×3 design matrix is given by $\boldsymbol{F}=[\boldsymbol{v}_2,\boldsymbol{v}_3,\boldsymbol{v}_4,\boldsymbol{v}_5]^T$ with

weight matrix $\boldsymbol{V} = \mathrm{diag}\big(s_2, s_3, s_4, s_5\big)$ 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 hence $\det \boldsymbol{M}\big(\xi_3^*, \boldsymbol{\beta}\big) = 16\,s_2\,s_3^2 + 18\,s_2\,s_3\,s_5 + s_3^2 s_5$. Define the following quantities

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

Then the inverse of the information matrix 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 can be rewritten as

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

which is equivalent to the following system of inequalities

$$c_1 + 4c_2 + 2c_3 + 2c_4 \le 3\,(1 + 2\gamma)^2 \text{ and } 4c_1 + 12c_2 + 5c_3 + 4c_4 \le 3\,(2 + 3\gamma)^2$$

where the first inequality arises form the vertices $\boldsymbol{v}_1$ and $\boldsymbol{v}_8$ and the second inequality comes from the vertices $\boldsymbol{v}_6$ and $\boldsymbol{v}_7$. Because of the complexity of the system above we employed computer algebra using Wolfram Mathematica 11.3 (see Wolfram Research, Inc. 2018) to obtain the set of solutions for γ as given in part (*iii*).

Proof of part (*iv*): The 3×3 design matrix $\boldsymbol{F} = [\boldsymbol{v}_2, \boldsymbol{v}_6, \boldsymbol{v}_7]^T$ is given by

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

The condition of the equivalence theorem is equivalent to

$$\left(\left(x_2 - \frac{x_1}{2}\right)^2 + \left(x_3 - \frac{x_1}{2}\right)^2\right)(2+3\gamma)^2 + \left(\frac{3x_1}{2} - x_2 - x_3\right)^2$$
$$(2+2\gamma)^2 \le \big(x_1 + \gamma(x_2 + x_3)\big)^2 \ \forall \boldsymbol{x} \in \{1,2\}^3,$$

and the above inequalities reduce to

$$90\,\gamma^2 + 168\gamma + 72 \le 0 \text{ and } 6\gamma^2 + 16\gamma + 8 \le 0$$

where the first inequality arises form the vertices $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ and the second inequality comes from vertex $\boldsymbol{v}_8$. In analogy to parts (*i*) and (*ii*) the sets of solutions are given by $[-1.2, -2/3]$ and $[-2, -2/3]$, respectively where the 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 $\gamma \in [-1.2, -1)$. □

References

Atkinson AC, Woods DC (2015) Designs for generalized linear models. In: Dean A, Morris M, Stufken J, Bingham D (eds) Handbook of design and analysis of experiments. Chapman & Hall/CRC Press, Boca Raton, pp 471–514

Burridge J, Sebastiani P (1992) Optimal designs for generalized linear models. J Ital Stat Soc 1:183–202

Burridge J, Sebastiani P (1994) D-optimal designs for generalised linear models with variance proportional to the square of the mean. Biometrika 81:295–304

Chatterjee S (1988) Sensitivity analysis in linear regression. Wiley, New York

Ford I, Torsney B, Wu CFJ (1992) The use of a canonical form in the construction of locally optimal designs for non-linear problems. J R Stat Soc Ser B (Methodol) 54:569–583

Gaffke N, Idais O, Schwabe R (2019) Locally optimal designs for gamma models. J Stat Plan Inference 203:199–214

Gea-Izquierdo G, Cañellas I (2009) Analysis of holm oak intraspecific competition using gamma regression. For Sci 55:310–322

Gregori D, Pagano E, Merletti F, Petrinco M, Bo S, Desideri A (2011) Regression models for analyzing costs and their determinants in health care: an introductory review. Int J Qual Health Care 23:331–341

Grover G, Sabharwal ASA, Mittal J (2013) An application of gamma generalized linear model for estimation of survival function of diabetic nephropathy patients. Int J Stat Med Res 2:209–219

Hardin JW, Hilbe JM (2018) Generalized linear models and extensions, 4th edn. Stata Press, Texas

Harman R, Trnovská M (2009) Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices. Math Slov 59:693–704

Kilian R, Matschinger H, Löeffler W, Roick C, Angermeyer MC (2002) A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. J Mental Health Policy Econ 51:21–31

Kurtoğlu F, Özkale MR (2016) Liu estimation in generalized linear models: application on gamma distributed response variable. Stat Pap 57:911–928

McCrone P, Knapp M, Fombonne E (2005) The Maudsley long-term follow-up of child and adolescent depression. Eur Child Adolesc Psychiatry 14:407–413

McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Chapman & Hall, London

Montez-Rath M, Christiansen CL, Ettner SL, Loveland S, Rosen AK (2006) Performance of statistical models to predict mental health and substance abuse cost. BMC Med Res Methodol 6:53–63

Ng VK, Cribbie RA (2017) Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. Curr Psychol 36:225–235

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Radloff M, Schwabe R (2016) Invariance and equivariance in experimental design for nonlinear models. In: Kunert J, Müller CH, Atkinson AC (eds) mODa 11-Advances in model-oriented design and analysis. Springer, Berlin, pp 217–224

Silvey SD (1980) Optimal design. Chapman & Hall, London

Wenig CM, Schmidt CO, Kohlmann T, Schweikert B (2009) Costs of back pain in germany. Eur J Pain 13:280–286

Wolfram Research, Inc. (2018) Mathematica. Wolfram Research, Inc., Champaign

Yu Y (2010) Monotonic convergence of a general algorithm for computing optimal designs. Ann Stat 38:1593–1606