Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

BELMIRO P.M. DUARTE, GUILLAUME SAGNOL

Approximate and exact $D-$optimal designs for 2^k factorial experiments for Generalized Linear Models via SOCP

ZIB Report 18-02 (December 2017)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Approximate and exact $D-$optimal designs for 2^k factorial experiments for Generalized Linear Models via SOCP

Belmiro P.M. Duarte[a,b], Guillaume Sagnol[c]

[a]*Polytechnic Institute of Coimbra, ISEC, Dep. of Chemical and Biological Engineering, Portugal.*
[b]*CIEPQPF, Dep. of Chemical Engineering, University of Coimbra, Portugal.*
[c]*Technische Universität Berlin, Institut für Mathematik, Germany.*

Abstract

We propose (Mixed Integer) Second Order Cone Programming formulations to find approximate and exact $D-$optimal designs for 2^k factorial experiments for Generalized Linear Models (GLMs). Locally optimal designs are addressed with Second Order Cone Programming (SOCP) and Mixed Integer Second Order Cone Programming (MISOCP) formulations. The formulations are extended for scenarios of parametric uncertainty employing the Bayesian framework for *log det* $D-$optimality criterion. A quasi Monte-Carlo sampling procedure based on the Hammersley sequence is used for integrating the optimality criterion in the parametric region. The problems are solved in `GAMS` environment using `CPLEX` solver. We demonstrate the application of the algorithm with the logistic, probit and complementary log-log models and consider full and fractional factorial designs.

Keywords: $D-$optimal designs, 2^k factorial experiments, Exact designs, Second Order Cone Programming, Generalized linear models.

*Correspondence to: Polytechnic Institute of Coimbra, ISEC, Dep. of Chemical and Biological Engineering, R. Pedro Nunes, 3030–199 Coimbra, Portugal.

Email addresses: `bduarte@isec.pt` (Belmiro P.M. Duarte), `sagnol@math.tu-berlin.de` (Guillaume Sagnol)

1. Motivation

We consider the problem of determining model-based $D-$optimal designs of 2^k factorial experiments for Generalized Linear Models where k is the number of factors considered in the study. Our setup is that we have a given Generalized Linear Model defined on a design space formed by combinations of $\{-1, +1\}$ of n_p covariates, and a given total number of observations, N, available for the study. The design problem is to find the number of replicates (if any) at each of these design points, $\boldsymbol{n} \in \mathbb{N}_0$, subject to the requirement that they sum to N A standard approach to deal with the problem is to compute approximate optimal designs, which can be seen as a continuous relaxation of the M-bODE problem, and can be interpreted as the optimal proportions w_i of trials to perform on the design point $\mathbf{x}_i \in \{-1, +1\}^k$. The practical implementation of the approximate design requires approximating $N \times w_i$, $\forall i$, to integer values such that $\sum_{i=1}^{2^k} N \times w_i = N$. In contrast to the exact designs, approximate designs can be computed by using convex programming techniques in polynomial time (with respect to the number of design points).

Recently, it has been shown that the $D-$optimality criterion can be represented by using Second Order Cone (SOC)-inequalities. Let us designate the design obtained by maximizing the expectation of the determinant of the Fisher Information Matrix (FIM) raised to the power $1/n_p$, where n_p is the number of parameters in the model over their prior distribution as *det-root-n_p Bayesian D−optimal designs*. The Bayesian $D-$optimal designs obtained by maximizing the expectation of the logarithm of the determinant of the FIM over the prior distribution of the model parameters proposed by Chaloner and Larntz (1989) are called *log-det Bayesian D−optimal designs*. Both criteria assume that the uncertainty of the parameters can be adequately captured in the prior distribution. Typically, there are several different forms of the Bayesian $D-$optimality criterion, see Berger and Wong (2005) but the $log - det$ criterion is the most common one for non-linear models for three reasons: i. it arises naturally from the expected gain of the Shannon information (Chaloner and Verdinelli, 1995; Lindley, 1956); ii. this criterion is concave, so a global equivalence theorem hold (Firth and Hinde, 1997); and finally iii. its numerical results are the most satisfactory, see Atkinson et al. (2007, pag. 297).

Sagnol and Harman (2015) proposed a MISOCP formulation to find exact $D-$optimal designs for linear models (Sagnol and Harman, 2015). The expansion of the SOCP and MISOCP formulations to handle *det-root-n_p Bayesian D−optimal designs* employing the ideas of Duarte and Wong (2015) was already discussed in

Sagnol and Harman (2015, Appendix). Following the same research line, Harman and Filová (2016) developed the `R` package *OptimalDesign* which relies on the SOCP and MISOCP formulations to find optimal designs for linear models assuming a pre-defined set of candidate points in the design space. This setup was tested with several models and design spaces, and among them is the optimal design of 2^k factorial plans for linear models. Here we use their ideas to develop formulations to: i. find exact locally $D-$optimal designs for generalized linear models on the discrete design space $\mathbf{x}_i \in \{-1, +1\}^k$, i.e. 2^k factorial designs; and ii. compute exact *log-det Bayesian D−optimal designs* for generalized linear models on the same design space. The formulation for computing locally $D-$optimal designs is similar to that proposed by Sagnol and Harman (2015) for linear models, and is considered here as the departure point in the use of MISOCP for computing factorial designs for GLMs.

From our knowledge, this is the first paper to present a MISOCP formulation to find exact *log-det Bayesian D−optimal designs* and using a branch-and-cut algorithm to find the solution of the problems. Furthermore, it is the first work addressing the design of factorial plans of experiments for generalized linear models via MISOCP.

1.1. Prior work

Mathematical programming (MP) algorithms have improved substantially over the last two decades and they can be used to systematically solve the M-bODE problem. The basic paradigm is to reformulate the M-bODE problem to canonical forms that can be handled by specialized solvers, examples are semidefinite programming (SDP) (Vandenberghe and Boyd, 1996) and SOCP (Sagnol, 2011). The strengths of MP algorithms are that i. they rely on efficient software which commonly requires mild computational times to find the optimum; ii. they can be addressed in polynomial time using interior point methods; and iii. they can easily integrate additional constraints imposed to the designs. Among the applications of MP algorithms for finding M-bODE is Linear Programming (Gaivoronski, 1986; Harman and Jurík, 2008), SOCP (Lu and Pong, 2013; Sagnol, 2011), MISOCP (Sagnol and Harman, 2015) and SDP (Duarte and Wong, 2015; Fedorov and Lee, 2000; Filová et al., 2011; Papp, 2012).

The interest in GLMs increased during the latest decades, since it expands the classic linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function of nonlinear nature. The model allows for the dependent variable to have a non-normal distribution, and covers

different statistical models, such as linear regression for normally distributed responses, logistic models for binary data, log-linear models for count data, and complementary log-log models for interval-censored survival data, among others. Several authors addressed the theoretical issues of the optimal design of experiments for GLMs (Li and Majumdar, 2008; Wang et al., 2006; Yang and Stufken, 2009; Zhang and Ye, 2014). Woods et al. (2006), Dror and Steinberg (2008), Waterhouse et al. (2008) and Woods and van de Ven (2011) developed numerical methods to find optimal designs for GLMs with quantitative covariates and, consequently, continuous design spaces. Other authors used the problem as benchmark for testing different algorithms (e.g. King and Wong (2000) and Duarte and Wong (2015)).

In this paper we consider the optimal design of 2^k factorial experiments for GLMs assuming that the covariates are qualitative factors with two levels coded as -1 and $+1$, see Yang et al. (2016, Sec. 1) for a motivating example. Theoretical results for the 2^2 problem were obtained by Graßhoff and Schwabe (2008) and Yang et al. (2012), while Dorta-Guerra et al. (2008) studied the general case of the 2^k factorial design. To find locally $D-$optimal designs for 2^k factorial experiments for GLMs (Yang et al., 2016) proposed a numerical algorithm based on functional analysis, subsequently expanded to find $EW-D-$optimal designs and adapted for fractional factorial setups, where EW stands for *expectation weighted* and is characterized by replacing the expectation of the information matrix under the prior by the information matrix at $\bar{\boldsymbol{\theta}}$ where $\bar{\boldsymbol{\theta}} = \int_{\Theta} \boldsymbol{\theta}\, \pi(\boldsymbol{\theta})\, \mathrm{d}(\boldsymbol{\theta})$, and $\pi(\boldsymbol{\theta})$ is the prior that represents the parameter uncertainty. In contrast, our proposed algorithm is based on SOCP and MISOCP formulations and is a general framework that can be used for all GLMs and other generalized additive models, local and Bayesian optimal setups, full and fractional factorial designs. Our formulations can also be used for computing constrained optimal designs since they can easily accommodate additional constraints; e.g. budget constraints (Harman et al., 2016).

In this work we assume the design space is *intrinsically discrete*, and each factor can have only two possible values $\{-1, +1\}$. To handle the dependence of the FIM over the model parameters we use a Bayesian framework where each parameter follows an *a priori* established probability function.

1.2. Paper organization

In what follows, Section 2 presents mathematical background for the SOCP and MISOCP formulations and the details about the integration method. Section 3 presents the SOCP and MISOCP formulations for locally $D-$optimal designs for 2^k factorial experiments for GLMs, and Section 4 extends them to scenarios with

parametric uncertainty. Section 5 demonstrates the application of the algorithm for obtaining optimal designs for fractional factorial experiments, and Section 6 offers the conclusions.

2. Mathematical background

This section provides the background material required by the mathematical formulation for obtaining optimal designs for 2^k factorial experiments using SOCP and MISOCP. In subsection 2.1 we review the fundamentals of GLMs and respective FIMs. Section 2.2 introduces the conceptual basis for Second Order Cone (SOC)-representability of sets and functions. Subsection 2.3 introduces Bayesian optimal designs, and subsection 2.4 presents the numerical approach used for integration of the optimality criterion over the parameters domain. In the following, we use bold face lowercase letters to represent vectors, bold face capital letters for continuous domains, blackboard bold capital letters for discrete domains, and capital letters for matrices. Finite sets containing ι elements are compactly represented by $[\iota] = \{1, \cdots, \iota\}$.

2.1. Generalized Linear Models

This section introduces the fundamentals of GLMs. In this work we focus on the use of GLMs in problems where the response of interest is binary, i.e the data is assumed to follow a Bernoulli distribution.

Consider an experiment, conducted with the purpose of constructing the dose-response curve for identifying the relationship between the dose of the drug and its effect on patients, where the measured outcome of interest is either a "success" or "failure", coded as 1 and 0, respectively. In our experimental setup, the probability of "success" or "failure" depends on the combination of k qualitative factors and on a set of unknown parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{n_p}$. A *treatment* is represented by a binary vector $\mathbf{x} \in \{-1, +1\}^k$, which indicates a combination of factor levels: $x_i = +1$ or $x_i = -1$ indicates that the i^{th} factor is set to its high or low level in the treatment $\mathbf{x}$, respectively. In this work we assume that the unknown parameter $\boldsymbol{\theta}$ belongs to the n_p-dimensional cartesian box $\boldsymbol{\Theta} = \times_{j=1}^{n_p} [\theta_j^{LO}, \ \theta_j^{UP}]$, with each interval $[\theta_j^{LO}, \ \theta_j^{UP}]$ representing the plausible range of values for the j^{th} parameter. However, it would be straightforward to generalize our approach to non-rectangular parameter regions.

The classic statistical modeling paradigm can not be applied to the probability of "success" because the response is by definition constrained to fall between 0 and 1. Instead, the outcome is modeled in terms of conditional expectation

$\mathbb{E}[y|\mathbf{x},\boldsymbol{\theta}]$. For binary outcomes taking the values 0 and 1, the probability of "success" is modeled by a function $f : \{-1,1\}^k \times \boldsymbol{\Theta} \to [0,1]$, such that the treatment $\mathbf{x}$ induces a binary response $y \in \{0,1\}$, satisfying

$$\mathbb{E}[y|\mathbf{x},\boldsymbol{\theta}] = \mathbb{P}(y=1|\mathbf{x},\boldsymbol{\theta}) = f(\mathbf{x},\boldsymbol{\theta}). \tag{1}$$

Let us first consider an *approximate design* ξ^a, which allocates a weight $w_i \geq 0$ on the i^{th} treatment $\mathbf{x}_i \in \{-1,1\}^k,\ i \in [2^k]$, with $\sum_{i=1}^{2^k} w_i = 1$. For a given numbering $\mathbb{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_{2^k}\}$ of all 2^k treatments in $\{-1,+1\}^k$, such a design can simply be represented by a vector of nonnegative weights $\mathbf{w}$ summing to 1. Hence, we identify the set Ξ^a of all feasible approximate designs with the (2^k-1)-dimensional simplex $\Xi^a \equiv \left\{\mathbf{w} \in \mathbb{R}^{2^k} :\ w_i \geq 0,\ \forall i \in [2^k],\ \sum_{i=1}^{2^k} w_i = 1\right\}$.

Similarly, an $N-$*exact design* ξ^e is defined by the number of replications $n_i \in \mathbb{N}_0$ for the i^{th} treatment $\mathbf{x}_i \in \mathbb{X}$, with $\sum_{i=1}^{2^k} n_i = N$. We identify the set Ξ_N^e of feasible $N-$exact designs $\xi^e = \{(\mathbf{x}_i, n_i)\},\ i \in [2^k]$, with

$$\Xi_N^e \equiv \left\{\mathbf{w} \in \mathbb{R}^{2^k} :\ \exists n_1,\ldots,n_{2^k} \in \mathbb{N}_0,\ w_i = \frac{n_i}{N},\ \sum_{i=1}^{2^k} n_i = N\right\}.$$

The worth of the design ξ is measured by a convex functional of its FIM. The elements of the normalized FIM obtained after adjusting for the sample size are the negative of the expectation of the second order derivatives of the log-likelihood of (1), $\mathcal{L}(\xi,\boldsymbol{\theta})$, with respect to the parameters. This matrix is proportional to

$$\begin{aligned}\mathcal{M}(\xi,\boldsymbol{\theta}) &= -\,\mathbb{E}\left(\frac{\partial\mathcal{L}(\xi,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\ \frac{\partial\mathcal{L}(\xi,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\intercal}\right) = \int_{\mathbf{x}\in\mathbb{X}} M(\mathbf{x},\boldsymbol{\theta})\ \mathrm{d}\,\xi(\mathbf{x}) \\ &= \sum_{i=1}^{2^k} w_i\ \boldsymbol{h}(\mathbf{x}_i,\boldsymbol{\theta})\ \boldsymbol{h}^\intercal(\mathbf{x}_i,\boldsymbol{\theta}),\end{aligned} \tag{2}$$

where $\mathcal{M}(\xi,\boldsymbol{\theta})$ is the *global* FIM from the design ξ at $\boldsymbol{\theta}$, $M(\mathbf{x},\boldsymbol{\theta})$ is the *elemental* FIM from point $\mathbf{x}$, and $\boldsymbol{h}(\mathbf{x}_i,\boldsymbol{\theta}) \in \mathbb{R}^{n_p}$ is the vector of derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$ at point $\mathbf{x}_i \in \{-1,+1\}^k$. For the GLM model (1), the terms $\boldsymbol{h}(\mathbf{x}_i,\boldsymbol{\theta})$ used to construct the FIM for GLMs are (Atkinson et al., 2007):

$$\boldsymbol{h}(\mathbf{x}_i,\boldsymbol{\theta}) = \frac{1}{\sqrt{\mathbb{E}[y|\mathbf{x}_i,\boldsymbol{\theta}]\ (1-\mathbb{E}[y|\mathbf{x}_i,\boldsymbol{\theta}])}}\ \frac{\partial\mathbb{E}[y|\mathbf{x}_i,\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}}. \tag{3}$$

The volume of the asymptotic confidence region of $\boldsymbol{\theta}$ to determine from the design is proportional to $\det[\mathcal{M}^{-1/2}(\xi,\boldsymbol{\theta})]$, and consequently the maximization of

the determinant (or its geometric mean) of the FIM leads to the smallest possible volume. For a given nominal value of $\boldsymbol{\theta}$, a locally $D-$optimal approximate design, ξ_D^a, is obtained from

$$\xi_D^a \in \arg\max_{\xi^a \in \Xi^a} \{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\}^{1/n_p}, \tag{4}$$

and for some $N \geq n_p$, a locally $D-$optimal $N-$exact design, ξ_D^e solves

$$\xi_D^e \in \arg\max_{\xi^e \in \Xi_N^e} \{\det[\mathcal{M}(\xi^e, \boldsymbol{\theta})]\}^{1/n_p}. \tag{5}$$

When locally $D-$optimal designs are considered, maximizing $\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\}^{1/n_p}$ is equivalent to maximizing $\log\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\}$. Along the paper we address and present numerical experiments for the most common case, in which the function $f(\mathbf{x}, \boldsymbol{\theta}) = \mathbb{P}(y = 1|\mathbf{x}, \boldsymbol{\theta})$ has the form

$$f(\mathbf{x}, \boldsymbol{\theta}) = \ell(\kappa(\mathbf{x})^\intercal \boldsymbol{\theta}),$$

in which $\kappa : [-1, 1]^k \to \mathbb{R}^{n_p}$ is a set of n_p known regression functions, typically polynomials, and $\ell : \mathbb{R} \to [0, 1]$ is a so-called *link function*. This framework includes GLMs for the linear predictor $f(\mathbf{x}, \boldsymbol{\theta}) = \ell(\mathbf{x}^\intercal \boldsymbol{\theta})$, but also more complicated models with an intercept and interaction terms such as Scheffé's mixture models. For example, we will give numerical results for a model of the form $\ell((1, x_1, x_2, x_3, x_4, x_1 \times x_3)^\intercal \boldsymbol{\theta})$ in Section 5. Table 1 lists common GLMs used for modeling binary outcome processes.

Table 1: GLMs used for testing the algorithm.

Model	Designation	Model
Logistic	*logit*	$\mathbb{E}[y\|\mathbf{x}, \boldsymbol{\theta}] = \dfrac{\exp(\boldsymbol{\theta}^\intercal \kappa(\mathbf{x}))}{1 + \exp(\boldsymbol{\theta}^\intercal \kappa(\mathbf{x}))}$
Probit	*probit*	$\mathbb{E}[y\|\mathbf{x}, \boldsymbol{\theta}] = \Phi(\boldsymbol{\theta}^\intercal \kappa(\mathbf{x}))^\dagger$
Complementary log-log	*CLL*	$\mathbb{E}[y\|\mathbf{x}, \boldsymbol{\theta}] = 1 - \exp[-\exp(\boldsymbol{\theta}^\intercal \kappa(\mathbf{x}))]$

† - $\Phi(x)$ is the cdf of the standard normal distribution at x.

2.2. *SOC-representability*

In this section we review the concept of SOC-representability and apply it for reformulating the M-bODE problem as a SOCP. Following, we use the theoretical results of Ben-Tal et al. (2009):

Definition 2.1. (SOC-representability of a set). A convex set $S \subset \mathbb{R}^n$ is second-order cone representable if S is the projection of a set in a higher-dimensional space that can be described by a set of second-order cone inequalities. That is, S is SOC-representable if and only if there exist $A_i \in \mathbb{R}^{n_i \times (n+m)}$, $\lambda_i \in \mathbb{R}^{n_i}$, $\upsilon_i \in \mathbb{R}^{n+m}$ and $\delta_i \in \mathbb{R}$ for $i \in [N_c]$ such that:

$$\mathbf{x} \in S \iff \exists \mathbf{y},\ \forall i \in [N_c]:\ \left\| A_i \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \lambda_i \right\| \leq \upsilon_i^{\intercal} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \delta_i$$

A direct application of Definition (2.1) is the rotated cone of inequalities (Alizadeh and Goldfarb, 2001; Lobo et al., 1998) where the set $S = \{(\mathbf{x}, t, u) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R} : \|\mathbf{x}\|^2 \leq t\, u,\ t \geq 0,\ u \geq 0\} \subset \mathbb{R}^{n+2}$ is SOC-representable. Specifically, a constraint of the form $\|\mathbf{x}\|^2 \leq t\, u,\ t \geq 0,\ u \geq 0$ is equivalent to the second-order cone constraint

$$\left\| \begin{pmatrix} 2\,\mathbf{x} \\ t - u \end{pmatrix} \right\| \leq t + u. \tag{6}$$

The concept of SOC-representability can be extended to functions according to Ben-Tal et al. (2009):

Definition 2.2. (SOC-representability of a function). A convex or concave function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ is SOC-representable if and only if the epigraph of f, $\{(t, \mathbf{x}) : f(\mathbf{x}) \leq t\}$ or the hypograph $\{(t, \mathbf{x}) : f(\mathbf{x}) \geq t\}$], respectively, is SOC-representable.

Based on the Definitions (2.1-2.2), the problem of maximizing a concave SOC-representable function (or minimizing a convex one) over a SOC-representable set can be cast as a SOC program. This property can find application, for example, in constructing the epigraph of the geometric mean of q positive variables $\mathbf{x} \in \mathbb{R}^n_+$, that is $f(\mathbf{x}) = \prod_{i=1}^{q} x_i^{1/q}$, used in §3-4. In the following we demonstrate the construction of SOC-representations for the geometric mean assuming a general number of q variables x_i.

Let us introduce an auxiliary variable $\mathbf{u} \in \mathbb{R}_+^{l_1 \times l_2}$ that serves to store products of two variables of the previous level of aggregation where $l_2 = \lceil \log(q)/\log(2) \rceil - 1$ is the number of aggregation levels required by the representation, $\lceil \bullet \rceil$ is the operator that rounds the result of the internal operation to the upper integer, and l_1 is the number of different variables required in each level of aggregation $l \in [l_2]$:

$$\begin{aligned} l_{1,1}(l) &= \lfloor q/2^l \rfloor,\ l \in [l_2] \\ l_{1,2}(l) &= q \bmod (2^l),\ l \in [l_2] \\ l_1(l) &= l_{1,1}(l) + l_{1,2}(l),\ l \in [l_2], \end{aligned}$$

and $\lfloor \bullet \rfloor$ is the result of the approximation for the lower integer value, and $a \bmod(b)$ is the reminder of the division of a by b. For $t \in \mathbb{R}_+$, $\mathbf{x} \in \mathbb{R}_+^q$ we have $t^q \leq \prod_{i=1}^q x_i$ such that

$$\exists \mathbf{u} \in \mathbf{R}_+^{l_1 \times l_2} : \begin{cases} u_{l,1}^2 \leq x_{2l-1}\ x_{2l},\ l \in [l_{1,1}(1)] \\ u_{l,1}^2 \leq x_{2l-1}\ u_{l,1},\ l \in [l_{1,2}(1)] \\ u_{l,m}^2 \leq u_{2l-1,m-1}\ u_{2l,m-1},\ l \in [l_{1,1}(m)],\ m \in \{2, \cdots, l_2\} \\ u_{l,m}^2 \leq u_{2l-1,m-1}\ u_{l,m},\ l \in [l_{1,2}(m)],\ m \in \{2, \cdots, l_2\} \\ t^2 \leq u_{2l-1,l_2}\ u_{2l,l_2} \end{cases} \tag{7}$$

The inequalities in (7) can then be expressed as second-order cone inequalities, see (6). The concave monomial function $f(\mathbf{x}) = \prod_{i=1}^q x_i^{\vartheta_i}$, required by the Bayesian optimal design problem, is also SOC-representable if $\mathbf{x} \in \mathbb{R}_+^n$, $\vartheta_i \in \mathbb{Q}$, $i \in [n]$, $\sum_{i=1}^n \vartheta_i \leq 1$ where $\mathbb{Q}$ is the set of rational numbers (Ben-Tal et al., 2009). Consequently, it can also be represented by a set of second-order cone inequalities similar to (6).

2.3. Locally and Bayesian optimal designs

GLMs are nonlinear by construction, and for that class of models, the FIM depends on the parameters and consequently all design criteria depend on the unknown parameters that we want to estimate. When nominal values are assumed for the parameters, the resulting designs are called locally optimal. The design strategies commonly used to handle the dependence noticed above include the use of: i. a sequence of locally optimal designs, each computed using the latest experimental results (Ford et al., 1989); ii. Bayesian designs that optimize the expectation of the optimality criterion over the prior distribution of the model parameters (Chaloner and Larntz, 1989), which is assumed to be known; and iii. minimax designs that minimize the worst design inefficiency from the unknown values of the model parameters (Imhof, 2001; Wong, 1992). In subsequent sections we focus on the Bayesian optimal design setup, which assumes that a prior density $\pi(\boldsymbol{\theta})$ is available for $\boldsymbol{\theta}$. The approximate *log-det Bayesian D−optimal design*, ξ_{BayesD}^a, solves

$$\xi_{BayesD}^a = \arg\max_{\xi^a \in \Xi^a} \int_{\boldsymbol{\Theta}} \log\left\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\right\}\ \pi(\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}, \tag{8}$$

and a similar representation is used for the exact design.

To assess the performance of the computed exact locally $D-$optimal designs ξ^e, we use an optimal approximate design ξ^a as reference and express the effi-

ciency as

$$\text{Eff}_D = \left(\frac{\det[\mathcal{M}(\xi^e, \boldsymbol{\theta})]}{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]} \right)^{1/n_p} \tag{9}$$

To construct a measure to assess the efficiency of the exact *log-det Bayesian D−optimal designs* let $\phi(\xi) = \exp\left[\Phi(\xi)^{\frac{1}{n_p}}\right]$ be a rescaled criterion so that $\phi(\xi)$ coincides with the positvely homogeneous *det-root-n_p criterion* when $\boldsymbol{\Theta}$ is a singleton, and $\Phi(\xi) := \int_{\boldsymbol{\Theta}} \log\{\det[\mathcal{M}(\xi, \boldsymbol{\theta})]\}\ \pi(\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}$. Now, the *log-det Bayesian D−optimal efficiency* is the ratio $\phi(\xi^e)/\phi(\xi^a)$ which reads

$$\begin{aligned} \text{Eff}_D = \Bigg[\exp\Bigg(& \int_{\boldsymbol{\Theta}} \log(\det[\mathcal{M}(\xi^e, \boldsymbol{\theta})])\ \pi(\boldsymbol{\theta})\ \mathrm{d}(\boldsymbol{\theta}) - \\ & - \int_{\boldsymbol{\Theta}} \log(\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})])\ \pi(\boldsymbol{\theta})\ \mathrm{d}(\boldsymbol{\theta}) \Bigg) \Bigg]^{1/n_p}. \end{aligned} \tag{10}$$

Notice the efficiency of the exact designs are computed relatively to optimal approximate designs, and can be seen as a measure of the loss of efficiency due to imposing integer constraints on the design weights for a given N. So if a design ξ^e has an efficiency of 90%, it does not mean that ξ^e is not truly $N-$exact optimal, but only that 10% of efficiency was lost in rounding the optimal approximate design ξ^a.

2.4. Numerical integration scheme

The problem of finding Bayesian $D-$optimal designs requires integrating the optimality criterion over the parametric domain, see (8). In this section we introduce the numerical approach used in our work. Duarte and Wong (2015) use Gaussian Quadrature Formulas (GQF) for computing Bayesian designs via SDP, having found that they are accurate but as the dimensionality of the integration domain increases the size of the MP problem generated becomes computationally prohibitive. One possible strategy to overcome this issue is decreasing the number of collocation points which, however, reduces the accuracy of the integration scheme.

To keep the MP problem tractable without compromising the accuracy, we use a quasi-random Monte Carlo sampling scheme for integrating a general function $\gamma(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$ in the closed domain $\boldsymbol{\Theta} \subset \mathbb{R}^{n_p}$ (Caflisch, 1998). This approach is less accurate than GQF but avoids the exponential increase of the MP problem as

the dimensionality of the parameter's domain increases. The integral approximation is

$$\int_{\boldsymbol{\Theta}} \gamma(\boldsymbol{\theta}) \ \mathrm{d}\,\boldsymbol{\theta} \approx \frac{\prod_{i=1}^{m}(\theta_i^{UP} - \theta_i^{LO})}{N_s} \sum_{j=1}^{N_s} \gamma(\boldsymbol{\theta}_j), \tag{11}$$

where N_s is the number of sampling points and the n_p-dimensional points $\boldsymbol{\theta}_j$ are generated using a quasi-random number generator. We use the Hammersley number sequence, a multi-dimensional low-discrepancy sequence (Hammersley and Handscomb, 1964) which denotes advantages over purely random approaches because it distributes the points more adequately and sample the most important points more frequently. Diwekar and Kalagnanam (1997) demonstrates the advantages of the Hammersley sequence for sampling in stochastic optimization problems such as (8).

3. Locally $D-$optimal designs

In this section, we use the SOCP and MISOCP formulations to find, respectively, locally approximate and exact $D-$optimal designs for 2^k factorial experiments for the GLMs listed in §2.1. In subsection 3.1 we present the formulations for obtaining $D-$optimal designs and in subsection 3.2 we present results for the case of a link function applied to a linear predictor $f(\mathbf{x}, \boldsymbol{\theta}) = \ell(\mathbf{x}^\intercal\boldsymbol{\theta})$, i.e., $k = n_p$. The extension of the formulations to more complex models is considered in Section 5.

3.1. SOCP and MISOCP formulations for locally D−optimal 2^k factorial experiments

This subsection presents the formulations for locally $D-$optimal designs for 2^k factorial experiments. First, we address the approximate optimal design problem and subsequently extend the formulation for finding exact optimal designs. In both cases we use the formulations presented in Sagnol and Harman (2015).

Let us first consider the approximate optimal design problem and recall that the goal is to find $D-$optimal designs for 2^k factorial plans. The set of points of the design is denoted by $\mathbb{X}$ where each treatment is a combination of values -1 and $+1$. For compactness we use $\boldsymbol{h}_i \in \mathbb{R}^{n_p}$ to represent $\boldsymbol{h}(\mathbf{x}_i, \boldsymbol{\theta})$ where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}\}$ is a singleton and $\boldsymbol{h}(\mathbf{x}_i, \boldsymbol{\theta})$ is given by (3). Standard basis vectors with the value 1 in j^{th} position are denoted by $\boldsymbol{e}_j \in \mathbb{R}^{n_p}$ and vectors of size j with all entries being 1's

are represented by $\mathbf{1}_j$. The SOCP formulation for the approximate design is:

$$\max_{\substack{\boldsymbol{z}_i \in \mathbb{R}^{n_p}, \boldsymbol{T} \in \mathbb{R}_+^{2^k \times n_p}, \\ \boldsymbol{J} \in \mathbb{R}^{n_p \times n_p}, \mathbf{w} \in \mathbb{R}_+^{2^k}}} \prod_{j=1}^{n_p} \boldsymbol{J}_{j,j}^{1/n_p} \tag{12a}$$

$$\text{s.t.} \sum_{i=1}^{2^k} \boldsymbol{h}_i \, \boldsymbol{z}_i^\intercal = \boldsymbol{J} \tag{12b}$$

$$\boldsymbol{J}_{j,j'} = 0, \; j' \in [n_p], \; j' \geq j+1 \tag{12c}$$

$$\|\boldsymbol{z}_i^T \intercal \, \boldsymbol{e}_j\|^2 \leq \boldsymbol{T}_{i,j} \, w_i, \; i \in [2^k], \; j \in [n_p] \tag{12d}$$

$$\sum_{i=1}^{2^k} \boldsymbol{T}_{i,j} \leq \boldsymbol{J}_{j,j}, \; j \in [n_p] \tag{12e}$$

$$\mathbf{w}^\intercal \, \mathbf{1}_{2^k} = 1 \tag{12f}$$

where the vectors $\boldsymbol{z}_i$'s and the matrices $\boldsymbol{T}$ and $\boldsymbol{J}$ contain auxiliary variables, and the weights of each design point are stored in vector $\mathbf{w}$. Equation (12d) is SOC-representable, c.f. Equation (6), and the complete SOCP problem is obtained after replacing the geometric mean in Equation (12a) by the corresponding SOC-representation (Equations (7)) with $q = n_p$.

The formulation for the exact optimal design problem is similar to (12) except for the continuous weight variables that are replaced by quotients n_i/N, $n_i \in \mathbb{N}_0$, which only affects Equations (12d) and (12f):

$$\max_{\substack{\boldsymbol{z}_i \in \mathbb{R}^{n_p}, \boldsymbol{T} \in \mathbb{R}_+^{2^k \times n_p}, \\ \boldsymbol{J} \in \mathbb{R}^{n_p \times n_p}, \boldsymbol{n} \in \mathbb{N}_0^{2^k}}} \prod_{j=1}^{n_p} \boldsymbol{J}_{j,j}^{1/n_p} \tag{13a}$$

$$\text{s.t.} \sum_{i=1}^{2^k} h_i \, \boldsymbol{z}_i^\intercal = \boldsymbol{J} \tag{13b}$$

$$\boldsymbol{J}_{j,j'} = 0, \; j' \in [n_p], \; j' \geq j+1 \tag{13c}$$

$$\|\boldsymbol{z}_i^\intercal \, \boldsymbol{e}_j\|^2 \leq \boldsymbol{T}_{i,j} \, \frac{n_i}{N}, \; i \in [2^k], \; j \in [n_p] \tag{13d}$$

$$\sum_{i=1}^{2^k} \boldsymbol{T}_{i,j} \leq \boldsymbol{J}_{j,j}, \; j \in [n_p] \tag{13e}$$

$$\boldsymbol{n}^\intercal \, \mathbf{1}_{2^k} = N \tag{13f}$$

The problems (12) and (13) as well as those formulated in Section 4 were coded in a high-level modeling system for mathematical programming, GAMS (GAMS Development Corporation, 2013). Subsequently, they were solved employing CPLEX that implements a barrier interior point algorithm (IBM, 2015) and supports conic constraints. To handle the integer variables in the MISOCP problem, CPLEX uses a branch and cut algorithm. The tolerance required for convergence in all problems addressed in subsequent sections is 10^{-5}. All computation in this paper were carried using on an Intel Core i7 machine running 64 bits Windows 10 operating system with 2.80 GHz.

3.2. Numerical experiments

This section demonstrates the application of the formulations in §3 for obtaining locally $D-$optimal designs for 2^k factorial experiments. We consider $k = 4$ (4 factors) and the linear predictor $\boldsymbol{\theta}^\intercal \mathbf{x}$, i.e., $\kappa(\mathbf{x}) = \mathbf{x}$. The plausibility region is a singleton set, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \equiv \{0.15\} \times \{0.20\} \times \{0.25\} \times \{0.2\}$. Table 2 presents the approximate and exact designs for this setup. For *logit* and *probit* models the optimal approximate designs are uniform where by *uniform* we mean designs with uniform allocation on its support points. Factorial models admit a large class of $D-$optimal designs (with the same FIM), and the uniform design is one of them. In contrast, the optimal design for the *CLL* model only requires one half of the experiments with all the remaining points equally replicated, practically being a fractional 2^{4-1} design.

The CPU times required by the algorithm for approximate designs are noticeably shorter compared to those reported for other problems. The main reasons contributing for such a trend are i. the efficiency of the solver; ii. the efficiency of the SOCP formulation, specifically for relatively small size MP problems; and iii. the small number of points in the design.

The last three columns of Table 2 display the optimal exact designs for the same singleton set and $N = 20$. The exact designs found are consistent with the approximate, with a few minor differences emerging from the comparison. First, the exact designs for *logit* and *probit* models are slightly different in the latest case. Second, the CPU time increases on average about 10 times relatively to that required by approximate designs due to the complexity of solving the MISOCP problems. The exact design for the *CLL* model still includes one half of the points, and some of the treatments are to be replicated 3 times. The efficiencies of the exact $D-$optimal designs obtained with equation (9) are in the last line of Table 2. The values are very close to 1.0 with a slightly less efficient design obtained for

the *CLL* model. For *logit* and *probit* models the inefficiency of the designs is almost zero, so the exact designs we have computed are as good as the approximate designs.

Table 2: Approximate and exact locally $D-$optimal designs for 2^4 factorial experiments on GLMs in Table 1, $\boldsymbol{\Theta} \equiv \{0.15\} \times \{0.20\} \times \{0.25\} \times \{0.2\}$ and $N = 20$.

Factor levels				Approximate designs			Exact designs		
x_1	x_2	x_3	x_4	*logit*	*probit*	*CLL*	*logit*	*probit*	*CLL*
-1	-1	-1	-1	0.0625	0.0625	0.1250	1	2	3
-1	-1	-1	1	0.0625	0.0625	0.0000	1	1	0
-1	-1	1	-1	0.0625	0.0625	0.0000	1	1	0
-1	-1	1	1	0.0625	0.0625	0.1250	1	2	3
-1	1	-1	-1	0.0625	0.0625	0.0000	2	1	0
-1	1	-1	1	0.0625	0.0625	0.1250	1	1	2
-1	1	1	-1	0.0625	0.0625	0.1250	1	1	2
-1	1	1	1	0.0625	0.0625	0.0000	1	1	0
1	-1	-1	-1	0.0625	0.0625	0.0000	2	1	0
1	-1	-1	1	0.0625	0.0625	0.1250	1	2	3
1	-1	1	-1	0.0625	0.0625	0.1250	1	2	3
1	-1	1	1	0.0625	0.0625	0.0000	1	1	0
1	1	-1	-1	0.0625	0.0625	0.1250	1	1	2
1	1	-1	1	0.0625	0.0625	0.0000	2	1	0
1	1	1	-1	0.0625	0.0625	0.0000	2	1	0
1	1	1	1	0.0625	0.0625	0.1250	1	1	2
CPU (s)				0.011	0.014	0.013	0.151	0.321	0.117
Eff_D				1.0000	1.0000	1.0000	0.9999	0.9999	0.9870

4. Bayesian $D-$optimal designs

In this section we extend the formulations in §3 for Bayesian optimal designs that maximize the $log - det$ criterion. In subsection 4.1 we introduce the formulations for approximate and exact Bayesian designs, and in subsection 4.2 we present the results for a 2^4 factorial plan.

4.1. SOCP and MISOCP formulations for Bayesian D−optimal designs for 2^k factorial experiments

Here, we extend the SOCP and MISOCP formulations derived for locally optimal designs to Bayesian setups where the FIM depends on the parameters, and the plausibility region as well the respective priors are known.

Contrarily to local $D-$optimal setups where the optimal designs obtained maximizing the *det-root-n_p* and *log-det* criteria are equal, here solving

$$\max_{\xi^a \in \Xi^a} \int_{\boldsymbol{\Theta}} \log\left\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\right\} \, \pi(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}$$

may produce different designs from those obtained for

$$\max_{\xi^a \in \Xi^a} \int_{\boldsymbol{\Theta}} \left\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta})]\right\}^{1/n_p} \, \pi(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta},$$

and the former criterion is more natural when the focus is on the inference of nonlinear models, cf. section 1. Sagnol and Harman (2015, Appendix) introduced SOCP and MISOCP formulations for the *det-root-n_p* criterion. Here, we consider the *log-det* criterion and present new formulations.

Let us recall to the *log-det Bayesian D−optimal design* problem (8). The numerical approach used to compute the expectation uses the quasi-random Monte Carlo sampling scheme presented in §2.4, and is equivalent to

$$\xi^a_{BayesD} \in \arg\max_{\xi^a \in \Xi^a} \sum_{m=1}^{N_s} \log\left\{\det[\mathcal{M}(\xi^a, \boldsymbol{\theta}_m)]\right\} \, \pi(\boldsymbol{\theta}_m), \tag{14}$$

where N_s is the number of samples initially set by the user.

To demonstrate how the expectation is computed let us consider the simplest case where the prior is the uniform distribution over $\boldsymbol{\Theta}$. Then, each vector $\boldsymbol{\theta}_m$ sampled from $\boldsymbol{\Theta}$ has equal probability, and $\pi(\boldsymbol{\theta}_m) = \frac{1}{N_s}$, $\forall m \in [N_s]$. The random number generator follows the Hammersley number sequence scaled to $\boldsymbol{\Theta}$. All examples in sections 4.2 and 5 consider this setup, but other priors as such the normal distribution $\mathcal{N}(\hat{\boldsymbol{\theta}}, \Sigma)$ (truncated over $\boldsymbol{\Theta}$) can be used. In the latest case, the Monte Carlo sampling scheme is used to pick samples $\boldsymbol{\theta}_m$, and their weights are then computed as $\pi(\boldsymbol{\theta}_m) = \phi(\boldsymbol{\theta}_m|\hat{\boldsymbol{\theta}}, \Sigma)$ where $\phi(\boldsymbol{a}|\boldsymbol{b}, \Gamma)$ is the n_p-dimensional normal probability distribution function at $\boldsymbol{a}$ when the average is $\boldsymbol{b}$ and the covariance matrix is Σ with $\boldsymbol{b} = \hat{\boldsymbol{\theta}}$ and Σ set by the user. Next, the values of $\pi(\boldsymbol{\theta}_m)$ are scaled so that they sum to 1.

The reformulation of the $log - det$ Bayesian criterion (14) uses the theoretical result from Sagnol and Harman (2015), which shows that $\xi \mapsto \det(\mathcal{M}(\xi, \boldsymbol{\theta}))^{1/n_p}$ is SOC-representable for all $\boldsymbol{\theta}$. After exponentiation, the function to be maximized in (14) can be rewritten as

$$\xi^a_{BayesD} \in \arg\max_{\xi^a \in \Xi^a} \prod_{m=1}^{N_s} \left\{[\det[\mathcal{M}(\xi^a, \boldsymbol{\theta}_m)]]^{1/n_p}\right\}^{n_p \pi(\boldsymbol{\theta}_m)}. \tag{15}$$

After raising to the power $1/n_p$, the problem is equivalent to

$$\xi^a_{BayesD} \in \arg\max_{\xi^a \in \Xi^a} \prod_{m=1}^{N_s} \varsigma_m^{\pi(\boldsymbol{\theta}_m)}, \tag{16a}$$

$$\text{s.t.} \quad [\det(\mathcal{M}(\xi^a, \boldsymbol{\theta}_m))]^{1/n_p} \geq \varsigma_m, \quad m \in [N_s] \tag{16b}$$

where the concave monomial $\prod_{m=1}^{N_s} \varsigma_m^{\pi(\boldsymbol{\theta}_m)},\ \varsigma_m \geq 0,\ \forall m \in [N_s]$ is SOC-representable provided the $\pi(\boldsymbol{\theta}_m)$'s are rational, cf. §2.2.

Finally, the formulation for approximate *log-det Bayesian D−optimal designs* is obtained by replacing the constraints (16b) by their equivalent SOC-representation:

$$\max_{\substack{\boldsymbol{z}_{m,i} \in \mathbb{R}^{n_p}, \boldsymbol{T} \in \mathbb{R}_+^{N_s \times 2^k \times n_p}, \\ \boldsymbol{J} \in \mathbb{R}^{N_s \times n_p \times n_p}, \mathbf{w} \in \mathbb{R}_+^{2^k}, \boldsymbol{\varsigma} \in \mathbb{R}_+^{N_s}}} \quad \prod_{m=1}^{N_s} \varsigma_m^{\pi(\boldsymbol{\theta}_m)} \tag{17a}$$

$$\text{s.t.} \ \sum_{i=1}^{2^k} \boldsymbol{h}_{i,m}\, \boldsymbol{z}_{m,i}^{\mathsf{T}} = \boldsymbol{J}_m,\ m \in [N_s] \tag{17b}$$

$$\boldsymbol{J}_{m,j,j'} = 0, m \in [N_s],\ j' \in [n_p],\ j' \geq j+1 \tag{17c}$$

$$\|\boldsymbol{z}_{m,i}^{\mathsf{T}}\, \boldsymbol{e}_j\|^2 \leq \boldsymbol{T}_{m,i,j}\, w_i,\ m \in [N_s], i \in [2^k],\ j \in [n_p] \tag{17d}$$

$$\sum_{i=1}^{2^k} \boldsymbol{T}_{m,i,j} \leq \boldsymbol{J}_{m,j,j},\ m \in [N_s],\ j \in [n_p] \tag{17e}$$

$$\varsigma_m \leq \prod_{j=1}^{n_p} \boldsymbol{J}_{m,j,j}^{1/n_p},\ m \in [N_s] \tag{17f}$$

$$\mathbf{w}^{\mathsf{T}}\, \mathbf{1}_{2^k} = 1 \tag{17g}$$

where (17f) represents the hypograph inequality for each $\varsigma_m,\ m \in [N_s]$, $\boldsymbol{h}_{i,m} := \boldsymbol{h}(\mathbf{x}_i, \boldsymbol{\theta}_m)$ is the vector of derivatives of the likelihood wrt. the parameters at $\boldsymbol{\theta}_m$, and the SOC-representation of the product of eigenvalues in (17f) is constructed applying the equation (7).

The construction of the MISOCP formulation for exact *log-det Bayesian D−optimal*

designs from (17) is straightforward

$$\max_{\substack{\boldsymbol{z}_{m,i}\in\mathbb{R}^{n_p},\boldsymbol{T}\in\mathbb{R}_+^{N_s\times 2^k\times n_p},\\ \boldsymbol{J}\in\mathbb{R}^{N_s\times n_p\times n_p},\boldsymbol{n}\in\mathbb{N}_0^{2^k},\boldsymbol{\varsigma}\in\mathbb{R}_+^{N_s}}} \quad \prod_{m=1}^{N_s} \varsigma_m^{\pi(\boldsymbol{\theta}_m)} \tag{18a}$$

$$\text{s.t.} \sum_{i=1}^{2^k} \boldsymbol{h}_{i,m}\, \boldsymbol{z}_{m,i}^{\mathsf{T}} = \boldsymbol{J}_m,\ m \in [N_s] \tag{18b}$$

$$\boldsymbol{J}_{m,j,j'} = 0, m \in [N_s],\ j' \in [n_p],\ j' \geq j+1 \tag{18c}$$

$$\|\boldsymbol{z}_{m,i}^{\mathsf{T}}\, \boldsymbol{e}_j\|^2 \leq \boldsymbol{T}_{m,i,j}\, \frac{n_i}{N},\ m \in [N_s], i \in [2^k],\ j \in [n_p] \tag{18d}$$

$$\sum_{i=1}^{2^k} \boldsymbol{T}_{m,i,j} \leq \boldsymbol{J}_{m,j,j},\ m \in [N_s],\ j \in [n_p] \tag{18e}$$

$$\varsigma_m \leq \prod_{j=1}^{n_p} \boldsymbol{J}_{m,j,j}^{1/n_p},\ m \in [N_s] \tag{18f}$$

$$\boldsymbol{n}^{\mathsf{T}}\, \mathbf{1}_{2^k} = N \tag{18g}$$

4.2. Numerical experiments

To demonstrate the application of the formulations in §4.1 for finding 2^k factorial experiments the same setup used in subsection 3.2 is considered, i.e $k = 4$. Here, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \equiv [0.0, 0.3] \times [0.0, 0.4] \times [0.0, 0.5] \times [0.0, 0.4]$, and the parameters are uniformly distributed in $\boldsymbol{\Theta}$, i.e $\pi(\boldsymbol{\theta}) \approx \mathbb{U}(\boldsymbol{\Theta})$, which yields $\pi(\boldsymbol{\theta}_m) = \frac{1}{N_s},\ \forall \boldsymbol{\theta}_m \in \boldsymbol{\Theta}$. In the numerical experiments we use $N_s = 256$.

The results in Table 3 agree with those obtained for local designs. The approximate optimal designs for all three models are equal and differences occur in the exact optimal designs. The exact optimal design for *probit* model with $N = 20$ has only 15 treatments, while that obtained for the *CLL* model includes 7 different treatments. The results may be different for other values of N. To corroborate, this idea we tested the algorithm with $N = 16$ and obtained designs where, for *logit* and *probit* models, all treatments are replicated once, and for the *CLL* model the treatments with $w_i = 0.1250$ in the approximate design are replicated twice. The efficiency of the obtained designs is computed with the Equation (10), and is in the last line of Table 3. Similarly to local optimal setups, the exact *log-det Bayesian D−optimal designs* found have an efficiency close to 1.0 with a smaller performance observed for the *CLL* model. As for the local designs, the MISOCP

problem requires about 10 times more CPU time than that required by the SOCP formulation for finding approximate designs.

Table 3: Approximate and exact Bayesian $D-$optimal designs for 2^4 factorial experiments on GLMs in Table 1, $\boldsymbol{\Theta} \equiv [0.0, 0.3] \times [0.0, 0.4] \times [0.0, 0.5] \times [0.0, 0.4]$, $\pi(\boldsymbol{\theta}) \approx \mathbb{U}(\boldsymbol{\Theta})$, $N_s = 256$ and $N = 20$.

Factor levels				Approximate designs			Exact designs		
x_1	x_2	x_3	x_4	*logit*	*probit*	*CLL*	*logit*	*probit*	*CLL*
-1	-1	-1	-1	0.0625	0.0625	0.1250	1	1	3
-1	-1	-1	1	0.0625	0.0625	0.0000	1	1	0
-1	-1	1	-1	0.0625	0.0625	0.0000	1	0	0
-1	-1	1	1	0.0625	0.0625	0.1250	1	1	5
-1	1	-1	-1	0.0625	0.0625	0.0000	1	2	0
-1	1	-1	1	0.0625	0.0625	0.1250	1	1	3
-1	1	1	-1	0.0625	0.0625	0.1250	1	1	4
-1	1	1	1	0.0625	0.0625	0.0000	2	1	0
1	-1	-1	-1	0.0625	0.0625	0.0000	1	2	0
1	-1	-1	1	0.0625	0.0625	0.1250	1	1	1
1	-1	1	-1	0.0625	0.0625	0.1250	1	1	2
1	-1	1	1	0.0625	0.0625	0.0000	2	1	0
1	1	-1	-1	0.0625	0.0625	0.1250	1	1	0
1	1	-1	1	0.0625	0.0625	0.0000	2	3	0
1	1	1	-1	0.0625	0.0625	0.0000	2	2	0
1	1	1	1	0.0625	0.0625	0.1250	1	1	2
CPU (s)				7.915	8.935	8.925	73.510	62.932	199.324
Eff_D				1.0000	1.0000	1.0000	0.9999	0.9999	0.9587

5. Optimal designs for more complex models

Now we use the formulations presented in subsections 3.1 and 4.1 for obtaining locally and Bayesian optimal designs for the case where $n_p > k$.

Let us consider a simple example where $k = 4$, $n_p = 6$ and a GLM in which the link function is applied to the predictor $\theta_0 + \sum_{i=1}^{k} \theta_i\, x_i + \theta_5\, x_1 \times x_3$, i.e., $\boldsymbol{\kappa}(\mathbf{x}) = (1.0, x_1, x_2, x_3, x_4, x_1 \times x_3)^\intercal$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^\intercal$.

First we consider the locally optimal designs for $\boldsymbol{\Theta} \equiv \{0.10\} \times \{0.15\} \times \{0.20\} \times \{0.25\} \times \{0.2\} \times \{-0.05\}$ and $N = 20$, shown in Table 4. We notice the approximate designs for this setup are uniform for all GLMs. The exact designs obtained for $N = 20$ only require 12 different treatments on all GLMs. The efficiency of the exact designs is above 0.98 for all GLMs. This value is below those obtained earlier for four-parameter models, and a plausible reason is that we used the same number of sampling points to compute the expectation in a parameter region with two additional dimensions. Because the performance of the quasi-random Monte

Carlo sampling depends on the number of samples used, which is required to increase as the dimension of the integration domain increases, setting $N_s = 256$ may degrade the numerical accuracy of the integral in this experiment. However, increasing the number of samples used to estimate the integral would lead to untractable problems.

A noticeable aspect is that the CPU time of the exact designs is more than 1000 times larger than that of the approximate designs which is due to the higher complexity of the the MISOCP problem. This difference in CPU time may be attenuated for other values of N, and the numerical test for N equal to a multiple of 2^k that produces uniform designs only requires a few seconds to solve.

Table 4: Approximate and exact locally $D-$optimal designs for 2^4 factorial experiments on GLMs in Table 1, $\kappa(\mathbf{x}) = (1, x_1, x_2, x_3, x_4, x_1 \times x_3)^\intercal$, $\boldsymbol{\Theta} \equiv \{0.10\} \times \{0.15\} \times \{0.20\} \times \{0.25\} \times \{0.2\} \times \{-0.05\}$ and $N = 20$.

Factor levels						Approximate designs			Exact designs		
1	x_1	x_2	x_3	x_4	$x_1 \times x_3$	*logit*	*probit*	*CLL*	*logit*	*probit*	*CLL*
1	-1	-1	-1	-1	1	0.0625	0.0625	0.0625	1	0	3
1	-1	-1	-1	1	1	0.0625	0.0625	0.0625	1	3	0
1	-1	-1	1	-1	-1	0.0625	0.0625	0.0625	2	0	1
1	-1	-1	1	1	-1	0.0625	0.0625	0.0625	0	2	1
1	-1	1	-1	-1	1	0.0625	0.0625	0.0625	1	2	0
1	-1	1	-1	1	1	0.0625	0.0625	0.0625	1	0	3
1	-1	1	1	-1	-1	0.0625	0.0625	0.0625	0	3	1
1	-1	1	1	1	-1	0.0625	0.0625	0.0625	2	0	1
1	1	-1	-1	-1	-1	0.0625	0.0625	0.0625	2	3	0
1	1	-1	-1	1	-1	0.0625	0.0625	0.0625	1	0	2
1	1	-1	1	-1	1	0.0625	0.0625	0.0625	0	2	1
1	1	-1	1	1	1	0.0625	0.0625	0.0625	3	0	2
1	1	1	-1	-1	-1	0.0625	0.0625	0.0625	1	0	2
1	1	1	-1	1	-1	0.0625	0.0625	0.0625	2	2	0
1	1	1	1	-1	1	0.0625	0.0625	0.0625	3	0	2
1	1	1	1	1	1	0.0625	0.0625	0.0625	0	3	1
CPU (s)						0.017	0.021	0.016	676.384	654.672	652.110
Eff_D						1.0000	1.0000	1.0000	0.9848	0.9835	0.9880

Now, we focus on *log-det Bayesian D−optimal designs*. Let us assume that the plausibility region for parameters is $\boldsymbol{\Theta} \equiv [0.05, 0.15] \times [0.0, 0.3] \times [0.0, 0.4] \times [0.0, 0.5] \times [0.0, 0.4] \times [-0.075, -0.025]$ that encloses the singleton set used for the locally optimal design. Similarly to previous numerical experiments we use $N_s = 256$ and $N = 20$. The prior is also the uniform distribution, $\pi(\boldsymbol{\theta}) \approx \mathbb{U}(\boldsymbol{\Theta})$ with $\Theta \subset \mathbb{R}^6$. Table 5 displays the results for this setup. The approximate designs obtained are also equal for all GLMs and they are also equal to locally optimal designs. Naturally, the CPU time increases relatively to that required by the locally

optimal designs. Finally, the MISOCP formulation for Bayesian optimal designs is computationally very challenging, and one observe that after 1000 s of CPU time the tolerance imposed has not been achieved. The solution displayed in the latest three columns is the one found at 1000 s of CPU time. The penultimate line of Table 5 presents the gap of the integer solution obtained after 1000 s of CPU time to that of the relaxed solution (expressed in percentage), and is an indication of the Bayesian $D-$efficiency of the exact design found (see the last line of the Table). Apart the *CLL* model, the efficiency of the designs found is close to 1.0.

Because the performance of the quasi-random Monte Carlo sampling scheme depends on the number of samples used, which is required to increase as the dimension of the integration domain increases, setting $N_s = 256$ may degrade the numerical accuracy of the integral in this experiment compared to that of examples presented in §4.2 where $n_p = 4$. On turn, the inaccuracy of the integral may decrease the efficiency of the designs. However, higher values of N_s increase the CPU time and the difficulty in converging the optimization problems, especially the MISOCP that computes the exact Bayesian optimal designs.

Table 5: Approximate and exact Bayesian $D-$optimal designs for 2^4 factorial experiments on GLMs in Table 1, $\kappa(\mathbf{x}) = (1, x_1, x_2, x_3, x_4, x_1 \times x_3)^\intercal$, $\boldsymbol{\Theta} \equiv [0.05, 0.15] \times [0.0, 0.3] \times [0.0, 0.4] \times [0.0, 0.5] \times [0.0, 0.4] \times [-0.075, -0.025]$, $\pi(\boldsymbol{\theta}) \approx \mathbb{U}(\boldsymbol{\Theta})$, $N_s = 256$ and $N = 20$.

Factor levels						Approximate designs			Exact designs		
1	x_1	x_2	x_3	x_4	$x_1 \times x_3$	*logit*	*probit*	*CLL*	*logit*	*probit*	*CLL*
1	-1	-1	-1	-1	1	0.0625	0.0625	0.0625	0	2	0
1	-1	-1	-1	1	1	0.0625	0.0625	0.0625	2	0	3
1	-1	-1	1	-1	-1	0.0625	0.0625	0.0625	2	1	3
1	-1	-1	1	1	-1	0.0625	0.0625	0.0625	1	2	0
1	-1	1	-1	-1	1	0.0625	0.0625	0.0625	2	0	2
1	-1	1	-1	1	1	0.0625	0.0625	0.0625	0	3	0
1	-1	1	1	-1	-1	0.0625	0.0625	0.0625	1	1	0
1	-1	1	1	1	-1	0.0625	0.0625	0.0625	2	1	2
1	1	-1	-1	-1	-1	0.0625	0.0625	0.0625	1	1	3
1	1	-1	-1	1	-1	0.0625	0.0625	0.0625	2	1	0
1	1	-1	1	-1	1	0.0625	0.0625	0.0625	2	1	0
1	1	-1	1	1	1	0.0625	0.0625	0.0625	0	2	3
1	1	1	-1	-1	-1	0.0625	0.0625	0.0625	2	2	0
1	1	1	-1	1	-1	0.0625	0.0625	0.0625	1	1	3
1	1	1	1	-1	1	0.0625	0.0625	0.0625	0	2	2
1	1	1	1	1	1	0.0625	0.0625	0.0625	2	0	0
CPU (s)						16.562	19.181	20.635	>1000.0†	>1000.0†	>1000.0†
Gap (%)									1.272	0.717	10.792
Eff_D						1.0000	1.0000	1.0000	0.9872	0.9928	0.8858

†After 1000 s of CPU the convergence condition is not satisfied.

Our results for both local and Bayesian $D-$optimal designs in sections 4.2 and 5 show that the exact designs can not be obtained by straightforward rounding of the approximate designs, and this finding is especially observed for *CLL* model. Consequently, the development of accurate algorithms to specifically compute exact designs is required to optimally allocate the experiments.

6. Conclusions

We use SOCP and MISOCP formulations similar to those proposed by Sagnol and Harman (2015) for finding, respectively, approximate and exact locally $D-$optimal designs for 2^k factorial experiments on Generalized Linear Models (logistic, *probit* and complementary log-log models) with binary response. We extend the approximate and exact formulations to *log-det Bayesian D−optimal* setups employing an integration procedure based on a quasi-random Monte Carlo sampling scheme using the Hammersley sequence. We test the formulations for obtaining locally and Bayesian optimal designs for 2^4 factorial experiments. Finally, we also consider the case where the number of factors is different of the number of the parameters. From all numerical computations we observed that the in most cases the optimal designs for *logit* and *probit* are uniform over the 2^k vertices the hypercube, with exceptions observed for *CLL* model. The computational time required by the approximate designs is one order of magnitude lower than that used by the MISOCP formulation for finding $N-$exact designs. However, our study reveals that in many cases, the $N-$exact optimal design cannot be trivially obtained by rounding the approximate optimal design. An advantage of SOCP and MISOCP formulations for $D-$optimal designs for 2^k factorial experiments is that they can incorporate explicitly additional constraints, e.g. the requirement that some of the treatments must have a low replication.

Acknowledgments

The authors thank Radoslav Harman of Comenius University in Bratislava for valuable comments and advice on an earlier draft of the manuscript.

References

Alizadeh, F., Goldfarb, D., 2001. Second-order cone programming. Mathematical Programming 95, 3–51.

Atkinson, A. C., Donev, A. N., Tobias, R. D., 2007. Optimum Experimental Designs, with SAS. Oxford University Press, Oxford.

Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. Princeton Series in Applied Mathematics. Princeton University Press.

Berger, M. P. F., Wong, W. K., 2005. Applied Optimal Designs. John Willey & Sons.

Caflisch, R. E., 1998. Monte Carlo and quasi-Monte Carlo methods. Acta Numerica 7, 1–49.

Chaloner, K., Larntz, K., 1989. Optimal Bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference 59, 191–208.

Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review. Statist. Science 10, 273–304.

Diwekar, U. M., Kalagnanam, J. R., 1997. An efficient sampling technique for optimization under uncertainty. AIChE Journal 43, 440–449.

Dorta-Guerra, R., González-Dávila, E., Ginebra, J., 2008. Two-level experiments for binary response data. Computational Statistics & Data Analysis 53 (1), 196 – 208.

Dror, H. A., Steinberg, D. M., 2008. Sequential experimental designs for generalized linear models. Journal of the American Statistical Association 103, 288–298.

Duarte, B. P. M., Wong, W. K., 2015. Finding Bayesian optimal designs for nonlinear models: A semidefinite programming-based approach. International Statistical Review 83 (2), 239–262.

Fedorov, V., Lee, J., 2000. Design of experiments in statistics. In: Handbook of Semidefinite Programming: Theory, Algorithms, and Applications. Springer Netherlands, Netherlands, pp. 511–532.

Filová, L., Trnovská, M., Harman, R., 2011. Computing maximin efficient experimental designs using the methods of semidefinite programming. Metrika 64 (1), 109–119.

Firth, D., Hinde, J. P., 1997. On bayesian $D-$optimum design criteria and the equivalence theorem in non-linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59 (4), 793–797.

Ford, I., Titterington, D. M., Kitsos, C. P., 1989. Recent advances in nonlinear experimental design. Technometrics 31 (1), 49–60.

Gaivoronski, A., 1986. Linearization methods for optimization of functionals which depend on probability measures. In: Prékopa, A., Wets, R. J.-B. (Eds.), Stochastic Programming 84 Part II. Vol. 28 of Mathematical Programming Studies. Springer Berlin Heidelberg, pp. 157–181.

GAMS Development Corporation, 2013. GAMS - A User's Guide, GAMS Release 24.2.1. GAMS Development Corporation, Washington, DC, USA.

Graßhoff, U., Schwabe, R., 2008. Optimal design for the Bradley–Terry paired comparison model. Statistical Methods & Applications 17 (3), 275–289.

Hammersley, J. M., Handscomb, D. C., 1964. Monte Carlo Methods. Methuen, London.

Harman, R., Bachratá, A., Filová, L., 2016. Construction of efficient experimental designs under multiple resource constraints. Applied Stochastic Models in Business and Industry 32 (1), 3–17.

Harman, R., Filová, L., 2016. Package "OptimalDesign". https://cran.r-project.org/web/packages/OptimalDesign/OptimalDesign.pdf.

Harman, R., Jurík, T., 2008. Computing $c-$optimal experimental designs using the Simplex method of linear programming. Comput. Stat. Data Anal. 53 (2), 247–254.

IBM, 2015. IBM ILOG CPLEX Optimization Studio - CPLEX User's Manual. Version 12, Release 6.

Imhof, L. A., 2001. Maximin designs for exponential growth models and heteroscedastic polynomial models. The Annals of Statistics 29 (2), 561–576.

King, J., Wong, W. K., 2000. Minimax $D-$optimal designs for logistic model. Biometrics 56, 1263–1267.

Li, G., Majumdar, D., 2008. $D-$optimal designs for logistic models with three and four parameters. Journal of Statistical Planning and Inference 138 (7), 1950 – 1959.

Lindley, D. V., 1956. On a measure of the information provided by an experiment. Ann. Math. Statist. 27 (4), 986–1005.

Lobo, M. S., Vandenberghe, L., Boyd, S., Lebret, H., 1998. Applications of second-order cone programming. Linear Algebra and its Applications 284 (1), 193 – 228.

Lu, Z., Pong, T. K., 2013. Computing optimal experimental designs via interior point method. SIAM Journal on Matrix Analysis and Applications 34 (4), 1556–1580.

Papp, D., 2012. Optimal designs for rational function regression. Journal of the American Statistical Association 107, 400–411.

Sagnol, G., 2011. Computing optimal designs of multiresponse experiments reduces to second-order cone programming. Journal of Statistical Planning and Inference 141 (5), 1684–1708.

Sagnol, G., Harman, R., 2015. Computing exact $D-$optimal designs by mixed integer second order cone programming. Annals of Statistics 43 (5), 2198–2224.

Vandenberghe, L., Boyd, S., 1996. Semidefinite programming. SIAM Review 8, 49–95.

Wang, Y., Myers, R. H., Smith, E. P., Ye, K., 2006. $D-$optimal designs for poisson regression models. Journal of Statistical Planning and Inference 136 (8), 2831 – 2845.

Waterhouse, T., Woods, D., Eccleston, J., Lewis, S., 2008. Design selection criteria for discrimination/estimation for nested models and a binomial response. Journal of Statistical Planning and Inference 138 (1), 132 – 144.

Wong, W., 1992. A unified approach to the construction of minimax designs. Biometrika 79, 611–620.

Woods, D. C., Lewis, S. M., Eccleston, J. A., Russell, K. G., 2006. Designs for generalized linear models with several variables and model uncertainty. Technometrics 48 (2), 284–292.

Woods, D. C., van de Ven, P. M., 2011. Blocked designs for experiments with correlated non-normal response. Technometrics 53 (2), 173–182.

Yang, J., Mandal, A., Majumdar, D., 2012. Optimal design for two-level factorial experiments with binary response. Statistica Sinica 22 (2), 885–907.

Yang, J., Mandal, A., Majumdar, D., 2016. Optimal designs for 2^k factorial experiments with binary response. Statistica Sinica 26, 381–411.

Yang, M., Stufken, J., 2009. Support points of locally optimal designs for nonlinear models with two parameters. Annals of Statistics 37 (1), 518–541.

Zhang, Y., Ye, K., 2014. Bayesian $D-$optimal designs for Poisson regression models. Communications in Statistics - Theory and Methods 43 (6), 1234–1247.