A Semidefinite Programming based approach for finding Bayesian optimal designs for nonlinear models

Belmiro P.M. Duarte

Department of Chemical and Biological Engineering, ISEC, Polytechnic Institute of Coimbra, R. Pedro Nunes, 3030-199 Coimbra, Portugal.
GEPSI, CIEPQPF, Department of Chemical Engineering, University of Coimbra, R. Sílvio Lima - Pólo II, 3030-790 Coimbra, Portugal.
e-mail: bduarte@isec.pt

and

Weng Kee Wong

Department of Biostatistics, Fielding School of Public Health, UCLA, 10833 Le Conte Ave., Los Angeles, California 90095-1772, U.S.A.
e-mail: wkwong@ucla.edu

Abstract: This paper uses Semidefinite Programming (SDP) to construct Bayesian optimal design for nonlinear regression models. The setup here extends the formulation of the optimal designs problem as a SDP problem from linear to nonlinear models. Gaussian Quadrature Formulas (GQF) are used to compute the expectation in the Bayesian design criterion, such as D-, A- or E-optimality. As an illustrative example, we demonstrate the approach using the power logistic model and compare results in the literature. Additionally, we investigate how the optimal design is impacted by different discretizing schemes for the design space, different amount of uncertainty in the parameter values, different choices of GQF and different prior distributions for the vector of model parameters, including normal priors with and without correlated components. Further applications to find Bayesian D-optimal designs with two regressors for a logistic model and a two-variable generalized linear model with a gamma distributed response are discussed and some limitations of our approach are noted.

AMS 2000 subject classifications: Primary 62K05, 90C22; secondary 65D32.
Keywords and phrases: Approximate Designs, Semidefinite Programming, Gaussian Quadrature Formulas, Nonlinear models.

1. Introduction

The field of optimal experimental designs dates back as early as the formal work of Smith (1918) and probably beyond. Early researchers worked sporadically in this area from the theoretical approach. Kiefer (1959) recognized the general design problem for a regression model was very difficult to solve even for relatively simple problems. As an example, the design problem for optimally estimating the three parameters in a homoscedastic quadratic polynomial model was only

solved in Gaffke and Kraft (1982). In a voluminous series of papers collected in Kiefer (1985), Kiefer proposed using approximate designs to solve general design problems for regression models. Approximate designs are essentially probability measures on the design space and are characterized by its design points and the weights at these points. The approach assumes the objective function is a convex function of the design to take advantage of the results from convex analysis. In particular, there are algorithms for finding different types of designs and equivalence theorems can be employed to verify the optimality of a design using the directional derivative of the design criterion. Equivalence theorems are widely discussed in design monographs and we provide more details in the discussion section.

A key result in the field is from Kiefer and Wolfowitz (1960), where they established two very different design criteria always lead to the same optimal design when the models are homoscedastic. Specifically, the D-optimal design for estimating all model parameters and the G-optimal design for minimizing the maximal variance of the fitted response surface across the design space are the same. Nowadays, optimal experimental designs are increasingly applied in different disciplines to find efficient approximate designs in the biological sciences, engineering, food science, pharmaceutical studies and education arena. A sample of applications of optimal design ideas to solve various real problems are given in Berger and Wong (2009).

Nonlinear models seem to be the norm in the biological sciences with typical applications ranging from modeling kinetic reaction velocity when the enzyme concentration varies (Dette, Melas and Wong, 2005) to the prediction of morbidity after lung resection (López-Fidalgo and Garcet-Rodriguez, 2004). Following convention, the worth of a design is measured by the Fisher Information Matrix (FIM) defined in the next section. One useful property of the FIM is that it is inversely proportional to the covariance matrix of the estimated model parameters. For nonlinear models, the FIM depends on the parameters and so all design criteria formulated in terms of the FIM also depends on the unknown parameters which we want to estimate. This seems to be a roundabout problem, prompting Cochran (1973) to remark "You tell me the values of the parameters and I promise to find the best experiment to estimate the values of the parameters". In practice, there are a few different ways to handle this problem.

The first approach is the simplest and follows from the above idea. Nominal values of the model parameters are required from previous studies or experiments and they are then treated as known so that the FIM does not contain unknown parameters in the design criterion. As such, the resulting optimal design is termed locally optimal (Chernoff, 1953). For example, if interest is in estimating model parameters and nominal values are given, the generalized variance is minimized by careful choice of the design points and weights. Data from the designed experiment are then used to re-estimate the parameters and they are then used as new nominal values to construct another locally optimal design. Usually a few iterations result in convergence of the values of the estimated parameters and the resulting design is termed a locally optimal design. Further discussion of this procedure is in Sitter and Wu (1999).

Another way to overcome the dependence of the FIM on the parameters in the model is to adopt a minimax or maximin approach. Minimax optimal designs minimize the worst possible performance of the design over all possible designs defined on the given design space (Wong, 1992). For example, if we wish to design a study to predict the response over a certain region and we are not sure exactly where the precise location in the region is, we may want to first consider the variance of a point in the region and then design to minimize the largest variance among the predicted responses in the region of interest. This method however is notoriously difficult and analytical solution for a nonlinear model is rarely possible except for very simple problems. The two major reasons are that the minimax criterion is non-differentiable and there are no algorithm that we know of that will converge to a minimax optimal design for a general regression model (with heteroscedastic errors).

The third method is Bayesian and this approach seems like an intermediate one between the locally and minimax paradigms in terms of information required before the design can be constructed. The conceptual framework behind the Bayesian optimal design is the assumption that the prior information of the parameters of interest and their uncertainty can be adequately captured in the prior distribution. This prior density averages out the parameter values and the criterion is no longer dependent on the parameters. The Bayesian optimal design is then found by optimizing the expectation of the design criterion. A review of work in Bayesian optimal designs is given in Chaloner and Verdinelli (1995).

The theory for the construction of Bayesian optimal design depends uniquely on the model and the criterion, and the mathematics required to solve the optimization problem is challenging even for linear models (Dette and Wong, 1996, 1998). In practice, Bayesian optimal designs are determined numerically using various types of algorithms such as those discussed in Fedorov (1972), Wynn (1972), Chaloner and Larntz (1989), Molchanov and Zuyev (2002) and Chang and Lin (2007). For instance, Chaloner and Larntz (1989) used the Nelder-Mead method which is a simplex based approach to find Bayesian D-optimal designs for the logistic model, and Molchanov and Zuyev (2002) used a steepest-ascent algorithm that guarantees convergence to the optimum but can become slow in its vicinity. Nonlinear programming (NLP) stochastic approaches such as genetic algorithms (GA) were also employed to find optimal designs (Heredia-Langner et al., 2004). Zhang (2006) used a hybrid approach by combining GA and a local NLP solver relying on General Reduced Gradient (GRG) or Square Quadratic Programming (SQP) algorithms to increase the convergence rate to the global optimum. Another way to increase the convergence rate is to include a procedure that removes design points that cannot support a D-optimal design measure (Harman and Pronzato, 2007). A review of such algorithms for finding optimal designs is Pronzato (2008).

Many algorithms require that the design space be discretized before the optimization process begins. After the initial grid on the design space is selected, many problems for finding an optimal design can be treated as a problem of minimizing a convex function. This suggests that any convex optimization algorithm such as SQP and recent Interior Point (IP) based algorithms are suitable

candidates (Ye, 1997).

In this paper, we focus on use of Semidefinite Programming to find Bayesian optimal designs for nonlinear models. This method as an optimization tool is often used in engineering and other applied fields, but seems greatly underutilized in statistical research. The SDP approach first requires that the design space be discretized into a finite set of points. The FIM at each candidate point is computed and assuming observations are independent, we then sum these information matrices to obtain the total FIM. The design criterion is formulated as a function of the total FIM before application of SDP to find the optimal design. For example, Vandenberghe and Boyd (1996) developed a SDP formulation to find optimal designs for estimating parameters in linear models using D-, A- and E-optimality criteria discussed in Pukelsheim (1980). Further applications of the SDP based framework for linear models include (i) finding optimal designs for multi-response linear models (Boyd and Vandenberghe, 2004), (ii) finding maximin efficient designs (Filová, Trnovská and Harman, 2011), (iii) use of a generalization of the Elfving's theorem to transform the c-, A- and D-optimality SDP formulations into more efficient Second Order Cone Programming (SOCP) formulations (Sagnol, 2011), (iv) finding sparse c-optimal designs for single-response trigonometric regression models (Qi, 2011), and (v) finding support points of optimal design for model with a mean response given by a rational regression function (Papp, 2012). Collectively, these papers emphasize the simplicity and efficiency of using the SDP based approach to find a solution to the optimization problem even though the optimal design may depend on the discretization scheme on the design space.

The goal of this paper is to extend the SDP formulation for linear models to find Bayesian optimal designs for nonlinear models. A unique feature of our SDP approach is that it will lead to a conic programming problem which is solved via a semidefinite programming solver to produce the globally Bayesian optimal design. The method allows for arbitrary prior distributions and the integration is handled using multidimensional Gaussian Quadrature Formulas (GQF). Section 2 provides background and discusses different types of Bayesian optimal designs. Section 3 formulates our design problem as a semidefinite program to determine Bayesian optimal designs for nonlinear models. Section 4 applies SDP to an illustrative case when we wish to generate a variety of Bayesian optimal designs for the logistic model and compares our results to those available in the literature. In section 5 we determine various optimal designs for the more complicated power logistic model and models with two regressors where we note that in one case, SDP can encounter some numerical problems in the search of an optimal design. We end with a discussion in Section 6.

2. Background

We focus on approximate designs which are probability measures defined on the given compact design space X. Given X, a statistical model and a given design criterion, the goal is to find an optimal approximate design. The setup assumes

that the total sample size, n, is fixed and the criterion is convex as a function of the FIM. If an approximate design ξ has k support points at x_i with weight w_i, $i = 1, \ldots, k$, the implemented design takes roughly $n \times w_i$ observations at the design point x_i, subject to $n \times w_1 + \ldots + n \times w_k = n$. Consequently the implemented designs may not be unique. The main advantages of working with approximate designs are that they are easier to find and understand. In addition, if the criterion is differentiable, there are algorithms that will iterate from a non-singular starting design and converge to the optimal design.

To fix ideas, consider the power logistic model proposed by Prentice (1976) for modeling binary responses in a dose response study using the mean function:

$$p(x, \theta) = \frac{1}{\{1 + \exp\left[-\beta\ (x - \mu)\right]\}^s}, \quad x \in\ X, \quad \theta^{\mathrm{T}} = [\mu, \beta, s] \in\ \Theta. \tag{1}$$

Here $\Theta \subset \mathbb{R}^3$ is a known compact region containing all possible values for the 3×1 vector of parameters $\theta^{\mathrm{T}} = [\mu, \beta, s]$. The set Θ is sometimes referred to as the plausible set of values for θ. Typically the dose x is confined in a compact interval X representing the range of doses of interest in the study and the binary outcome is coded as 1 for response and 0 otherwise. The probability of a response at dose x is $p(x, \theta)$. When $s = 1$, we have the logistic model and $\theta^{\mathrm{T}} = [\mu, \beta] \in \Theta \equiv [\mu^L, \mu^U] \times [\beta^L, \beta^U] \in \mathbb{R}^2$ with $\times$ representing the cartesian product. Here, μ^L is the known lower bound of μ and μ^U is its known upper bound. Similarly, β^L is the known lower bound of β and β^U is its known upper bound. More generally, if m is number of parameters to estimate, we have $\Theta \equiv \times_{i=1}^{m} [\theta_i^L, \theta_i^U] \in \mathbb{R}^m$, with θ_i^U representing its upper bound and θ_i^L its lower bound for the ith parameter θ_i. Similarly if X is multi-dimensional, we assume it also has a cartesian product structure.

Suppose our design ξ^n takes n independent observations from $x_1, x_2, ..., x_k$ and there are r_i responses from the n_i subjects randomly assigned to dose x_i, $i = 1, \cdots, k$. Subject to $n_1 + \cdots + n_k = n$, the log likelihood function is

$$\mathcal{L}(\xi^n, \theta) = \sum_{i=1}^{k} \log\left[\frac{n_i!}{(n_i - r_i)!\ r_i!}\right] + r_i\ \log\left[p(x_i, \theta)\right] + (n_i - r_i)\ \log\left[1 - p(x_i, \theta)\right].$$

Let $\mathcal{M}(\xi, \theta)$ denote the FIM using design ξ^n. The elements of $\mathcal{M}(\xi, \theta)$ are the expectation of the negative of the second derivatives of the log likelihood with respect to the parameters. Specifically, the FIM is proportional to

$$\mathcal{M}(\xi, \theta) = -\ \mathbb{E}\left\{\frac{\partial}{\partial \theta}\left(\frac{\partial \mathcal{L}(\xi, \theta)}{\partial \theta^{\mathrm{T}}}\right)\right\}.$$

Approximate designs require that we work with the weights w_i's that may not be ratios of two positive integers as in the case for exact designs like ξ^n where $w_i = n_i/n, i = 1, \ldots, k$. Accordingly, when responses are independent, the FIM of an approximate design ξ with weight w_i at x_i, $i = 1, \ldots, k$ is proportional to

$$\mathcal{M}(\xi, \theta) = \int_X M(x, \theta)\ \mathrm{d}\xi(x) = \sum_{i=1}^{k} w_i\ M(x_i, \theta),$$

where $M(x_i, \theta) = h(x_i, \theta)\ h(x_i, \theta)^{\mathrm{T}}$ and $h(x_i, \theta)$ is the mean response. If we have a binary response model with mean response $p(x, \theta)$ and θ is m-dimensional, the FIM of the design at the point x_i is $M(x_i, \theta) = h(x_i, \theta)\ h(x_i, \theta)^{\mathrm{T}}$, where

$$h(x_i, \theta) = \frac{1}{\sqrt{p(x_i, \theta)\ (1 - p(x_i, \theta))}} \left(\frac{\partial p(x_i, \theta)}{\partial \theta} \right), \quad \frac{\partial p(x_i, \theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial p(x_i, \theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial p(x_i, \theta)}{\partial \theta_m} \end{pmatrix}.$$

When we approximate X by a finite set $\mathcal{X}^q \in \mathbb{R}^{q \times d}$ populated with q discrete points of X uniformly distanced Δ units apart in each of the d design spaces in the cartesian product, the FIM is approximated by

$$\mathcal{M}(\xi, \theta) \approx \sum_{x \in \mathcal{X}^q} M(x, \theta)\ \chi(x),$$

where χ is a selected weight function on $\mathcal{X}^q$ used to approximate the integral using weight $\chi(x)$ at $x \in \mathcal{X}^q$.

Common design criteria are formulated in terms of the FIM and include D- and A-optimality for estimating model parameters. They can be formulated as $\Phi(\mathcal{M}(\xi, \theta))$, where Φ is a function that maximizes the information obtained from experiments in a certain way. When errors are independent and normally distributed, the D-optimality minimizes the generalized variance by minimizing the volume of the confidence ellipsoid of the model parameters. For A-optimality, we minimize the (squared) diagonal of the bounding box of the confidence ellipsoid. For E-optimality, we minimize the squared in-ball radius geometrically by maximizing the minimum eigenvalue of the FIM (Dette and Studden, 1993). More specifically, let Ξ be the set of all feasible designs on X. Then for D-optimality, we seek a design ξ_D that satisfies

$$\xi_D = \arg\min_{\xi \in \Xi} \left\{ -\log \left(\det[\mathcal{M}(\xi, \theta)] \right) \right\}.$$

For A-optimality, we seek a design ξ_A that satisfies

$$\xi_A = \arg\min_{\xi \in \Xi} \left\{ \mathrm{tr}[\mathcal{M}(\xi, \theta)^{-1}] \right\},$$

and for E-optimality, we seek a design ξ_E that satisfies

$$\xi_E = \arg\max_{\xi \in \Xi} \left\{ \lambda_{min}[\mathcal{M}(\xi, \theta)] \right\}$$

where λ_{min} is the minimum eigenvalue of the FIM for a given θ value. The Bayesian paradigm assumes a prior density $\pi(\theta)$ is available for θ and the Bayesian D-optimal design ξ_{BayesD} is defined by

$$\xi_{BayesD} = \arg\min_{\xi \in \Xi} \int_{\Theta} \left\{ -\log \left(\det[\mathcal{M}(\xi, \theta)] \right) \right\}\ \pi(\theta)\ \mathrm{d}\theta.$$

Similar representations apply to Bayesian versions for A- and E-optimality criteria. When Θ is approximated by a finite discrete set $\mathcal{T}^{\iota} \in \mathbb{R}^{\iota \times m}$ containing ι parameter combinations in the space Θ, the discrete version of the Bayesian D-optimal design becomes

$$\xi_{BayesD} \approx \arg\min_{\xi \in \Xi} \sum_{\theta \in \mathcal{T}^{\iota}} \{-\log\left(\det\left[\mathcal{M}(\xi, \theta)\right]\right)\} \ \pi(\theta) \ \gamma(\theta),$$

where $\gamma(\theta)$ is the weight of θ in the integral approximation.

The SDP based strategy treats the design problem as a general conic program where the variables are the weights of the points generated from the discretized set $\mathcal{X}^q$ and associated with the degenerate information matrices $M(x_i, \theta)$. SDP minimizes a linear function of a matrix in the positive semidefinite matrix cone subject to affine constraints. Depending on the objective function, this results in a polyhedral or a non-polyhedral feasibility region but in either case, we have a convex optimization problem (Ye, 1997) that allows us to generalize that any locally optimal design found is a globally optimal one (Boyd and Vandenberghe, 2004). The optimization problems so formulated are solved using accurate efficient solvers, such as `SeDuMi` (Sturm, 1999).

2.1. Semidefinite Programming

Semidefinite Programming is an extension of linear programming where some vector variables are replaced by matrices, and some of the non negativity elementwise constraints are replaced by positive semidefiniteness constraints (Wolkowicz, Saigal and Vandenberghe, 2000). Furthermore, SDP is a class of convex optimization involving linear objective functions subject to constraints requiring that an affine combination of symmetric matrices is positive definite (Boyd and Vandenberghe, 2004). The primal formulation of a general SDP problem, proposed by Nesterov and Nemirovskiĭ (1994), is as follows:

$$\begin{aligned} \min_{Z \in S^n} \ & \langle C, Z \rangle \\ \text{s.t } & Z \succeq 0 \\ & \begin{bmatrix} \langle A_1, Z \rangle \\ \cdots \\ \langle A_m, Z \rangle \end{bmatrix} = b \end{aligned}$$

where $\langle \bullet, \bullet \rangle$ represents the inner (Frobenius) product, $C, \ A_1, \cdots, A_m \in S^n$ are constant matrices, S^n is the space of $n \times n$ symmetric matrices, $Z \in S^n$ is the matrix of decision variables, $b \in \mathbb{R}^m$ is also a vector of constants, $Z \succeq 0$ indicates that Z belongs to the cone of positive semidefinite matrices, and $\langle C, Z \rangle = \mathrm{tr}(C^{\mathrm{T}} \ Z)$. A dual formulation, frequently employed in numerical algorithms, can be found in Ye (1997). A book length volume on SDP is Boyd and Vandenberghe (2004). General SDP formulations for local D-, A- and E-optimal designs of experiments can be found in Vandenberghe, Boyd and Wu (1998) and Vandenberghe and Boyd (1999).

2.2. Gaussian Quadrature Formulas

Gaussian Quadrature Formulas are a class of methods that use appropriate weights and nodes to numerically integrate a complex function $f(x)$ to a high degree of accuracy. For a one dimension integral over an arbitrary compact interval $[a, b]$, the formula is:

$$\int_a^b w(x)\ f(x)\mathrm{d}x \doteq \sum_{j=1}^{n} w_{j,n}\ f(x_{j,n})$$

where $w(x)$ is a weighting function, and n is the number of points, also designated as nodes, used in the integration. The accuracy of the approximation of the integral as a sum depends on the selected weight $w_{j,n}$ at the nodes $x_{j,n}$. A major advantage of GQF is that with judicious choices of the nodes and weights, it needs only n points to exactly integrate polynomials of degree $2\ n-1$ or less. This means that only n evaluations of the function $f(x)$ are required (Gerald and Wheatley, 1994). For $w(x) = 1$, $a = -1$ and $b = 1$ the nodes correspond to the zeros of the nth order Legendre polynomials; see, for example, (Atkinson, 1989). For $w(x) = 1$ and an arbitrary compact interval on the real line, the weights and nodes are determined from recursive algorithms such as those presented in Davis and Rabinowitz (1984). Multiple dimension regular domain based integrals may be determined employing one GQF in each dimension or Gaussian Cubature Formulas (Bernardo, Pistikopoulos and Saraiva, 1999). For example, if we follow the former strategy to integrate $f(x)$ with $x \in \mathbb{R}^p$, the weight at each node is the product of the weights for the one dimension integration and the nodes correspond to the intersection of one dimension nodes in $\mathbb{R}^p$ space.

3. SDP based formulation for Bayesian optimal designs

Let $x_1, \cdots, x_q$ be a set of points $\mathcal{X}^q$ in $X \in \mathbb{R}$. $\mathcal{X}^q$ is constructed from a discretization scheme, say with a constant step, where $x_1 = \min(X)$ and the other points determined recursively with a rule $x_j = x_{j-1} + \Delta x,\ j = 2, \cdots, q$. When X is a compact d-dimensional Euclidean subspace discretized into $x_1, \cdots, x_q$, each point $x_j,\ j \in \{1, \cdots, q\}$ is characterized by a vector with d components resulting from discretizing the space X with a rectangular mesh.

We use GQF to approximate the expectation integral of the optimality criterion by first discretizing the parameter space $\Theta \subset \mathbb{R}^m$. If ι is the number of points used in the integral approximation, the discretization points of the set $\mathcal{T}^\iota \equiv \{\theta_p : p = 1, \cdots, \iota\}$ correspond to the set of m-dimension combinations of roots of the $(\kappa - 1)^{th}$ order Legendre polynomials in each dimension of the space Θ and $\gamma_p,\ p \in \{1, \cdots, \iota\}$ are the corresponding weights (Abramowitz and Stegun, 1972). It follows that each discretization point $\theta_p \in \mathbb{R}^m,\ p = \{1, \cdots, \iota\}$ of $\mathcal{T}^\iota$ is obtained by the cartesian product of the sets containing GQF points from each one of the dimensions of Θ. Let $\rho \in \mathbb{R}^\kappa$ be the vector of roots of the

$(\kappa-1)^{th}$ order Legendre's polynomial on the interval $[-1,1]$ and let $\mathbf{1}$ be the $\kappa \times 1$ vector of all ones. Then assuming κ is equal for all dimensions of Θ, we have $\iota = \kappa^m$ and

$$\mathcal{T}^{\iota} = \times_{i=1}^{m} \left\{ \rho \left(\frac{\theta_i^U - \theta_i^L}{2} \right) + \left(\frac{\theta_i^U + \theta_i^L}{2} \right) \mathbf{1} \right\}.$$

Similarly, the ι-dimensional vector of weights γ is given by

$$\gamma = \prod_{i=1}^{m} \left[\times_{i=1}^{m} \omega \ \left(\frac{\theta_i^U - \theta_i^L}{2} \right) \right],$$

where $\omega \in \mathbb{R}^{\kappa}$ is the vector of weights of the Legendre polynomials on the interval $[-1,1]$. We choose GQF because the method tends to provide more accurate approximations than Monte Carlo or Hammersley sequence sampling schemes for an equal number of points (Bernardo, Pistikopoulos and Saraiva, 1999; Reber, 2004).

We now extend the SDP formulation proposed by Vandenberghe and Boyd (1999) to construct Bayesian D-, E- and A-optimal designs. To this end, define the FIM for a single point θ_p by $\mathfrak{M}(\chi, \theta_p) = \sum_{j=1}^{q} M(x_j, \theta_p)\ \chi_j$ and note that (i) the formulations of Vandenberghe and Boyd (1999) are for linear models and can be used to find locally optimal designs for nonlinear models; (ii) a Bayesian design can be interpreted as a convex linear combination of local designs with θ varying in the domain of Θ (Fedorov and Hackl, 1997), and (iii) the use of GQF to numerically represent the expectation of a Bayesian criterion over a pre-defined parameter domain guarantees the design criterion is convex and the weight of each locally optimal design can be represented in a normalized domain with positive values that sum to 1. It follows that if one applies SDP to a linear model, one obtains a global solution and if the model is nonlinear but the parameters are fixed, one obtains a locally optimal design. Further, if one seeks a Bayesian optimal design for a nonlinear model by using GQF to compute the expectation, one first obtains several locally optimal designs before they have to be averaged via the weights of GQF to obtain the Bayesian optimal design. The weights can be scaled to the interval domain [0,1].

For D-optimality, the formulation is:

$$\begin{aligned} \max_{\chi} \ & \sum_{p=1}^{\iota} \log \left\{ \det \left[\mathfrak{M}(\chi, \theta_p) \right] \right\} \ \pi(\theta_p)\ \gamma_p \\ \text{s.t. } & \chi_j \geq 0, \quad j = 1, \cdots, q \\ & \sum_{j=1}^{q} \chi_j = 1 \\ & \theta_p \in \mathcal{T}^{\iota}, \quad x_j \in \mathcal{X}^q \end{aligned} \tag{2}$$

The formulation in (2) is pseudo-SDP because of the log term in the objective function. The `cvx` solver used to solve all SDP problems supports the

log(det) function which is concave and non-monotonic, but uses a sequence of SDP problems to successively approximate the original problem.

It is helpful to reformulate the optimization problem (2) to produce an SDP-representable problem that can be addressed exactly with a SDP solver, thus improving the numerical efficiency. We base our reformulation upon the theoretical results from Ben-Tal and Nemirovskiĭ (2001) who proved that $-[\det(\mathcal{B})]^{1/m}$ is semidefinite representable for a matrix $\mathcal{B} \in \mathbb{R}^{m\times m}$ if $\mathcal{B} \succeq 0$. Accordingly, one can use a Linear Matrix Inequality (LMI) to represent the hypograph (set of points lying on or below a function) of $[\det(\mathcal{B})]^{1/m}$, ς, by:

$$\begin{bmatrix} \mathcal{B} & \mathcal{C} \\ \mathcal{C}^{\mathrm{T}} & \mathrm{diag}(\mathcal{C}) \end{bmatrix} \succeq 0 \tag{3}$$

$$\varsigma \leq \left(\prod_{i=1}^{m} \delta_i \right)^{1/m}. \tag{4}$$

Here $\mathcal{C}$ is a lower-triangular matrix and $\mathrm{diag}(\mathcal{C})$ is the $m \times m$ diagonal matrix with diagonal elements $\delta_i,\ i \in \{1, \cdots, m\}$ obtained from the diagonal elements of $\mathcal{C}$. It follows that the function to be maximized in (2) can be rewritten as $m\ \sum_{p=1}^{\iota} \log(\varsigma_p)\ \pi(\theta_p)\ \gamma_p$ where ς_p represents the hypograph of the geometric mean for each point θ_p. Exponentiating, we can equivalently maximize the left hand side of:

$$\prod_{p=1}^{\iota} \varsigma_p^{m\pi(\theta_p)\gamma_p} \simeq \prod_{p=1}^{\iota} \varsigma_p^{\alpha_p/2^{\ell}}. \tag{5}$$

where the right hand side is its approximation, having had $m\pi(\theta_p)\gamma_p$ replaced by a rational number via dyadic fractions (Reznik, 2008). Having the power now as a ratio of two integers provides a pure SDP representation of the Bayesian D-optimal design. This approximation is always possible because we can find integer α_p for all p and an integer $\ell \in \mathbb{N}$ such that

$$\sum_{p=1}^{\iota} \frac{\alpha_p}{2^{\ell}} \leq 1, \quad \alpha_p = [m\pi(\theta_p)\gamma_p\ 2^{\ell}],$$

and $[\cdot]$ represents the rounding to the nearest integer. The right hand side of (5) is a concave monomial, and recalling that κ is the number of points in the GQF, we may choose $\ell = m + \kappa$ to ensure $\sum_{p=1}^{\iota} \frac{\alpha_p}{2^{\ell}} \leq 1$ holds and so a LMI can be used to find the hypograph (Ben-Tal and Nemirovskiĭ, 2001, Ch. 2):

$$\psi \leq \prod_{p=1}^{\iota} \varsigma_p^{\alpha_p/2^{\ell}} \tag{6}$$

If $\alpha_p/2^{\ell}$ is already a rational fraction for each p, no approximation is required. The monomial in (6) reduces to a weighted geometric mean which can also be represented by a LMI, and consequently has an exact SDP representation

(Boyd and Vandenberghe, 2004). The upshot is that the SDP reformulation of the problem (2) is:

$$\begin{aligned}
\max_{\chi,\Delta_p,\varsigma,\psi}\ & \psi \\
\text{s.t. } & \begin{bmatrix} \mathfrak{M}(\chi,\theta_p) & \Delta_p \\ \Delta_p^{\mathrm{T}} & \operatorname{diag}(\Delta_p) \end{bmatrix} \succeq 0, \quad p \in \{1,\cdots,\iota\} \\
& \varsigma_p \le \left(\prod_{i=1}^{m} \delta_{i,p}\right)^{1/m}, \quad p \in \{1,\cdots,\iota\} \\
& \psi \le \prod_{p=1}^{\iota} \varsigma_p^{\alpha_p/2^\ell} \\
& \chi_j \ge 0, \quad j = 1,\cdots,q \\
& \sum_{j=1}^{q} \chi_j = 1 \\
& \theta_p \in \mathcal{T}^\iota, \quad x_j \in \mathcal{X}^q, \quad \alpha_p = [m\pi(\theta_p)\gamma_p\ 2^\ell]
\end{aligned} \tag{7}$$

where Δ_p are lower-triangular $m \times m$ matrices, and $\delta_{i,p}$ are the diagonal elements of each one. The inequalities involving monomial terms are geometric constraints that can be expressed as LMI to produce an exact SDP formulation for Bayesian D-optimal design.

For finding A-optimal designs, the formulation yields:

$$\begin{aligned}
\min_{\chi,a}\ & \sum_{p=1}^{\iota}\sum_{j=1}^{q} a_{i,p}\ \pi(\theta_p)\ \gamma_p \\
\text{s.t. } & \begin{bmatrix} \mathfrak{M}(\chi,\theta_p) & u_i \\ u_i^{\mathrm{T}} & a_{i,p} \end{bmatrix} \succeq 0, \quad i = 1,\cdots,m,\ p = 1,\cdots,\iota \\
& \chi_j \ge 0, \quad j = 1,\cdots,q \\
& \sum_{j=1}^{q} \chi_j = 1 \\
& \theta_p \in \mathcal{T}^\iota, \quad x_j \in \mathcal{X}^q
\end{aligned} \tag{8}$$

where $u_i \in \mathbb{R}^m$ is a unit vector, $a \in \mathbb{R}^{m\times\iota}$ is the matrix of decision variables $a_{i,p}$ with each representing an eigenvalue of a singular FIM from a point of $\mathcal{T}^\iota$.

For E-optimal designs, the formulation is:

$$\begin{aligned} \min_{\chi,e} \; & -\sum_{p=1}^{\iota} e_p \; \pi(\theta_p) \; \gamma_p \\ \text{s.t. } & \mathfrak{M}(\chi,\theta_p) - e_p \; I \succeq 0, \quad p = 1,\cdots,\iota \\ & \chi_j \geq 0, \quad j = 1,\cdots,q \\ & \sum_{j=1}^{q} \chi_j = 1 \\ & \theta_p \in \mathcal{T}^{\iota}, \quad x_j \in \mathcal{X}^q \end{aligned} \tag{9}$$

where I is the $q \times q$ identity matrix and $e^{\mathrm{T}} = [e_1, e_2, \cdots, e_\iota] \in \mathbb{R}^\iota$ is the vector of decision variables with each e_p representing the minimum eigenvalue of all FIM evaluated at the pth. point of $\mathcal{T}^\iota$. In practice, the problems (2), (7), (8) and (9) were first codified using `cvx`, a Matlab compatible environment that supports a particular approach to convex optimization (Grant, Boyd and Ye, 2012) before employing `SeDuMi`, a SDP solver. This solver in turn uses the IP method with a primal-dual predictor-corrector scheme and a self dual embedding (Sturm, 1999). The tolerance used in all problems is a user-selected constant $\mathrm{tol} = 10^{-8}$.

After SDP solves the design problem, a pruning procedure is usually required to select points from the discretized design space as support points for the optimal design. A common rule is to include them if their weights are not very small. This means that support points of the SDP-generated optimal design are selected from the set $C = \{x_j \in \mathcal{X}^q : \chi_j \geq \epsilon\}$, where ϵ is a user-selected small positive constant and to discard points with weights smaller than ϵ. The number of support points of the design is $k = \mathrm{card}(C)$ and the optimal design ξ found by SDP is formed from the points in the set C along with the ith support point having weight w_i equal to χ_i.

4. Bayesian optimal designs for the logistic model

In this section we apply the SDP formulation in (7) to solve all design problems in this section. As a start, we consider some Bayesian design problems in Chaloner and Larntz (1989) and compare our optimal designs with their results. We then generate and compare Bayesian optimal designs using different sizes of the region Θ, different discretization schemes and different integration schemes. In addition, we construct new Bayesian D-optimal designs using bivariate normal prior densities with varying degrees of correlation, and A- and E-optimal designs not discussed in Chaloner and Larntz (1989).

The model is the simple logistic model with $s = 1$ in Equation (1) commonly used to study binary outcomes. For example in dose response study, we wish to model whether the subject responds or not to different doses of a drug. Typically the doses are appropriately scaled to the interval $X \in [-1, 1]$. The two parameters $\theta = [\mu, \beta]^{\mathrm{T}}$ have meaningful interpretation and we assume they have

nominal values in a user-selected plausible region, say with $\theta \in \Theta = [-0.1, 0.1] \times [6.9, 7.1]$. The FIM for this model is given in various bibliographic references, such as King and Wong (2000). Following Chaloner and Larntz (1989), we used independent uniform prior densities for μ and β for comparison purposes but also use bivariate normal prior densities with varying correlation coefficients to show the flexibility of the SDP method. Unless otherwise stated, GQF were all based on 6 points for each parameter space, resulting in a total of 36 points needed to compute the bivariate integral. The grid employed for discretizing X is equally spaced with $\Delta x = 0.01$, and in all examples we consider $\epsilon = 10^{-5}$.

The Bayesian D-optimal designs in Table 1 show good agreement with those reported in Chaloner and Larntz (1989). We note that when $\Theta = [-0.1, 0.1] \times [6.9, 7.1]$, Chaloner and Larntz (1989) reported the Bayesian D-optimal design has only 2 support points, one between -0.23 and -0.22, and the other between 0.22 and 0.23. Our designs have 4 points with the weights shared at two adjacent points. This is due to the discrete grid we employed to search over the design space, implying that the discretization scheme can have an effect on the optimal design. The optimal design found for the case when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ has the same 3 support points reported in Chaloner and Larntz (1989).

In general, one observes that our Bayesian optimal designs require more support points when the plausible region Θ is wider, a phenomenon already observed by several other authors. The designs obtained are symmetric even if no symmetry constraints are included in the optimization problem. The CPU time in seconds (secs) required to solve each problem is shown in Table 1 and other tables. Our computer has a Intel Core i7 machine running 64 bits Windows 7 operating system with 2.80 GHz. In all cases, our reported CPU times are relatively short compared with our earlier experience with other algorithms.

Table 2 compares D-optimal designs when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ and various discretization schemes were applied to the design space X. For this particular setup and a tighter grid, the SDP-generated design has 5 points and the one resulting from $\Delta x = 0.01$ has 3 support points. One may expect that different solutions might be originated in some cases when the grid becomes thinner. Let us consider that a grid with $\Delta x^{(1)}$ leads to a design with a support point located at $x_i^{(1)}$. For a different grid, say constructed with $\Delta x^{(2)} = \Delta x^{(1)}/2$, the optimal support point obtained $x_i^{(2)} \in \{x_i^{(1)} - \Delta x^{(2)}, x_i^{(1)}, x_i^{(1)} + \Delta x^{(2)}\}$, but might indeed be different from $x_i^{(1)}$. However, both grids guarantee very similar designs, and design efficiencies.

Table 3 presents Bayesian D-optimal designs when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ and GQF of 7th, 9th and 11th orders in each dimension were used, resulting in a total of 16, 25 and 36 integration points, respectively. The optimal designs are similar suggesting that different polynomial orders may not matter a lot and that integration errors using GQF based on the fourth, fifth and sixth order polynomials are quite comparable for practical applications. The CPU times required to produce the optimal designs always increases when higher order polynomials are used.

Table 4 presents Bayesian D-optimal designs using 3 different prior distri-

TABLE 1

Bayesian D-optimal designs with independent uniform prior densities for the simple logistic model using different regions Θ when $X \in [-1, 1]$ and $\Delta x = 0.01$.

	μ		
β	$[-0.1, 0.1]$	$[-0.3, 0.3]$	$[-1.0, 1.0]$
$[6.9, 7.1]$	(-0.2300,0.1385)	(-0.3100,0.2520)	(-0.9500,0.1023)
	(-0.2200,0.3615)	(-0.3000,0.1183)	(-0.8200,0.0136)
	(0.2200,0.3615)	(0.0000,0.2593)	(-0.8100,0.0309)
	(0.2300,0.1385)	(0.3000,0.1183)	(-0.4400,0.2294)
		(0.3100,0.2520)	(0.0000,0.2479)
			(0.4400,0.2294)
			(0.8100,0.0309)
			(0.8200,0.0136)
			(0.9500,0.1023)
CPU (secs)	8.0965	7.4568	6.9264
$[6.0, 8.0]$	(-0.2300,0.1193)	(-0.3100,0.3666)	(-0.9600,0.0940)
	(-0.2200,0.3807)	(0.0000,0.2668)	(-0.8100,0.0552)
	(0.2200,0.3807)	(0.3100,0.3666)	(-0.4400,0.2264)
	(-0.2300,0.1193)		(0.0000,0.2487)
			(0.4400,0.2264)
			(0.8100,0.0552)
			(0.9600,0.0940)
CPU (secs)	7.3476	6.8484	6.3024
$[4.0, 10.0]$	(-0.2200,0.5000)	(-0.3200,0.3562)	(-1.0000,0.0749)
	(0.2200,0.5000)	(0.0000,0.2876)	(-0.7800,0.0938)
		(0.3200,0.3562)	(-0.4300,0.0584)
			(-0.4200,0.1519)
			(0.0000,0.2421)
			(0.4200,0.1519)
			(0.4300,0.0584)
			(0.7800,0.0938)
			(1.0000,0.0749)
CPU (secs)	7.8313	7.0044	4.8048

$(x.xxxx, w.wwww)$ = (design point, weight).

butions when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ and 6 points were employed in the GQF. The first set of priors is independent uniform priors on Θ with $U[\theta^L, \theta^U]$ and the other two sets of priors are bivariate normal distributions given by $N[(\theta^L + \theta^U)/2, \Sigma_j],\ j \in \{1, 2\}$, with

$$\Sigma_1 = \begin{bmatrix} 0.3 & 0.0 \\ 0.0 & 0.1 \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} 0.3 & 0.075 \\ 0.075 & 0.1 \end{bmatrix}.$$

Table 4 shows the computing time required to generate the optimal designs are about the same for the different priors and the optimal designs obtained under different priors are not too different after rounding. Table 5 extends Table 4 and displays Bayesian D-optimal designs under the same bivariate normal prior distribution but with different values of the covariance, $\sigma_{\mu,\beta}$, between the two components.

Table 6 shows the D-, A-, and E-optimal designs when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$. All designs are symmetric. We observe A-optimal designs have support points further apart than D and E-optimal designs and E-optimal designs have

TABLE 2

Bayesian D-optimal designs with independent uniform prior densities for the simple logistic model using different discretization schemes on the design interval X when $X \in [-1, 1]$ and $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$.

	$\Delta x = 0.02$	$\Delta x = 0.01$	$\Delta x = 0.005$
	(-0.3200,0.1225)	(-0.3100,0.3666)	(-0.3100,0.2696)
	(-0.3000,0.2479)	(0.0000,0.2668)	(-0.3050,0.0984)
	(0.0000,0.2594)	(0.3100,0.3666)	(0.0000,0.2641)
	(0.3000,0.2479)		(0.3050,0.0984)
	(0.3200,0.1225)		(0.3100,0.2696)
CPU (secs)	5.7252	6.8484	6.0996

$(x.xxxx, w.wwww)$ = (design point, weight).

TABLE 3

Bayesian D-optimal designs with independent uniform prior densities for the simple logistic model using different integration schemes on the parameter region Θ when $X \in [-1, 1]$ and $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$.

	GQF based on 4 points	GQF based on 5 points	GQF based on 6 points
	(-0.3100,0.3662)	(-0.3100,0.3665)	(-0.3100,0.3666)
	(0.0000,0.2676)	(0.0000,0.2670)	(0.0000,0.2668)
	(0.3100,0.3662)	(0.3100,0.3665)	(0.3100,0.3666)
CPU (secs)	3.4320	3.9936	6.8484

$(x.xxxx, w.wwww)$ = (design point, weight).

larger weights at the non-zero support points compared with the A and D-optimal designs.

Table 7 compares the Bayesian D-optimal designs obtained using different SDP formulations in section 3. We denote results from the pseudo-SDP formulation (2), the SDP formulation with α_p rational (7) and the SDP formulation with α_p allowed to be irrational by p-SDP, rα-SDP and irα-SDP, respectively in the table. The first requires approximate successive algorithms, the latter treats the powers of the monomial terms as a sequence of LMI's which may extend the dimension of the problem considerably, and the second requires the approximation of α_p by rational terms given by quotients of integer numbers. One may observe that for this particular setup the designs produced by all the formulations are somewhat similar, with rα-SDP offering the best efficiency.

TABLE 4

Bayesian D-optimal designs with uniform and normal prior densities for the simple logistic model when $X \in [-1, 1]$ and $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$.

	$U[\theta^L, \theta^U]$	$N[(\theta^L + \theta^U)/2, \Sigma_1]$	$N[(\theta^L + \theta^U)/2, \Sigma_2]$
	(-0.3100,0.3666)	(-0.3000,0.1150)	(-0.2900,0.3548)
	(0.0000,0.2668)	(-0.2900,0.2727)	(-0.2800,0.0353)
	(0.3100,0.3666)	(0.0000,0.2247)	(0.0000,0.2148)
		(0.2900,0.2727)	(0.2900,0.3951)
		(0.3000,0.1150)	
CPU (secs)	6.8484	6.8640	7.2540

$(x.xxxx, w.wwww)$ = (design point, weight).

TABLE 5
Bayesian D-optimal designs for the simple logistic model with $X \in [-1, 1]$, $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ *and bivariate normal prior densities with different covariances* $\sigma_{\mu,\beta}$.

	$\sigma_{\mu,\beta} = 0.05$	$\sigma_{\mu,\beta} = 0.1$	$\sigma_{\mu,\beta} = 0.15$
	(-0.2900,0.3904)	(-0.2900,0.3645)	(-0.2900,0.2613)
	(0.0000,0.2161)	(-0.2800,0.0244)	(-0.2800,0.1276)
	(0.2900,0.3935)	(0.0000,0.2154)	(0.0000,0.1718)
		(0.2900,0.3956)	(0.0100,0.0407)
			(0.2900,0.3725)
			(0.3000,0.0260)
CPU (secs)	6.8172	6.4740	6.5676

$(x.xxxx, w.wwww)$ = (design point, weight).

TABLE 6
Bayesian D-, A- and E-optimal designs with independent uniform prior densities for the simple logistic model when $X \in [-1, 1]$ *and* $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$.

	D-optimal design	A-optimal design	E-optimal design
	(-0.3100,0.3666)	(-0.4300,0.3865)	(-0.4100,0.4174)
	(0.0000,0.2668)	(0.0000,0.2271)	(0.0000,0.1651)
	(0.3100,0.3666)	(0.4300,0.3865)	(0.4100,0.4174)
CPU (secs)	6.8484	4.0404	2.7612

$(x.xxxx, w.wwww)$ = (design point, weight).

5. Extensions to More Complicated Models

We now apply SDP to find various Bayesian optimal designs for the more complicated power logistic model where we now permit the power s to additionally vary over a known interval. We also find Bayesian D-optimal designs for the exponential growth model with homoscedastic error with a single regressor. Additionally, we find Bayesian D-optimal designs for two models each with two regressors. The first is a logistic model with two regressors and two known parameters in the meann function and the second is a generalized linear model with a Gamma distributed response and a reciprocal link function.

5.1. Power-logistic model

The power-logistic model has an additional power parameter s that enables it to fit skewed binary data. Different values of s signify varying degrees of skewness

TABLE 7
Bayesian D-optimal designs with different SDP formulations for the simple logistic model when $X \in [-1, 1]$, $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$.

	p-SDP	rα-SDP	irα-SDP
	(-0.3100,0.3666)	(-0.3100,0.3666)	(-0.3100,0.3666)
	(0.0000,0.2668)	(0.0000,0.2668)	(0.0000,0.2668)
	(0.3100,0.3666)	(0.3100,0.3666)	(0.3100,0.3666)
CPU (secs)	19.6093	6.8484	15.7093

$(x.xxxx, w.wwww)$ = (design point, weight).

in the data. As before, let $\theta^{\mathrm{T}} = [\mu, \beta, s]$ and assumed that each parameter has a known range of possible values with $s \in S \equiv [s^L, s^U],\ \beta \in [\beta^L, \beta^U],\ \mu \in [\mu^L, \mu^U]$ and $\theta \in \Theta$ with Θ being a cartesian product of the three ranges of nominal values for μ, β and s. This setup is more complicated than the one considered in King and Wong (2000) where they allowed s to be a singleton set only.

For this model, a direct calculation shows the elements in the function $h(x, \theta)$ for an observation at x_i are:

$$h(x_i, \theta) = \begin{pmatrix} \frac{-s\ (x_i-\mu)\ \exp(-\beta(x_i-\mu))}{(1+\exp(-\beta(x_i-\mu)))\ \sqrt{(1+\exp(-\beta(x_i-\mu)))^s-1}} \\ \frac{\beta\ s\ \exp(-\beta(x_i-\mu))}{(1+\exp(-\beta(x_i-\mu)))\ \sqrt{(1+\exp(-\beta(x_i-\mu)))^s-1}} \\ \frac{\log(1+\exp(-\beta(x_i-\mu)))}{\sqrt{(1+\exp(-\beta(x_i-\mu)))^s-1}} \end{pmatrix}$$

We use the same setup employed in section 4 and suppose that $s \in [0.5, 1.0]$ and the priors for all three parameters are independent uniform densities over their range. The design interval is discretized using $\Delta x = 0.01$ and the integration scheme has a 5-point GQF for each parameter space, thus rendering a total of $\iota = 125$ points to evaluate the expectation. Selected Bayesian D-optimal designs are shown in Table 8.

The results show that the Bayesian optimal designs for the power logistic model have more support points than the corresponding number of support points required when the model is logistic. The additional points are needed to estimate an additional parameter. Generally, it is very difficult to determine the exact relationship between the number of support points in the optimal design and the size of S and Θ. We also note that (i) D-optimal Bayesian designs for the power logistic model are generally non symmetric unless $s = 1$ and, (ii) the notable increase in the CPU time required to find the optimal design for the power logistic model versus the logistic model when $s = 1$. This is also due to the larger number of discretized points required for the additional parameter in the power logistic model to evaluate the expectation which now also has more number of constraints in the SDP problem than when the model is logistic.

We note that when there is greater uncertainty on the power parameter s, as signified by a larger domain of S (e.g., $S \equiv [0.2, 1.0]$), we observe that the optimal design requires additional support points, and the smallest two of them are located at smaller dose values to capture the increasingly skewed trend. Table 9 compares the D-, A- and E-optimal designs found by SDP when $\Theta = [-0.3, 0.3] \times [6.0, 8.0] \times [0.5, 1.0]$.

5.2. Exponential growth model

We now apply SDP to find Bayesian D-optimal designs for the exponential growth model with normally distributed homoscedastic errors (Braess and Dette, 2007). The model has two parameters, β_1 and β_2 and a single regressor:

$$y(x, \theta) = \beta_0 + \exp(-\beta_1\ x), \quad x \in X, \quad \theta^{\mathrm{T}} = [\beta_0, \beta_1] \in\ \Theta \tag{10}$$

TABLE 8
Bayesian D-optimal designs with independent uniform prior densities for the simple logistic model ($s = 1$) and the power logistic model when $X \in [-1, 1]$, $\Theta = [-0.3, 0.3] \times [6.0, 8.0] \times [0.5, 1.0]$.

	$s = 1$	$s \in [0.5, 1]$	$s \in [0.2, 1]$
	(-0.3100,0.3666)	(-0.7000,0.2638)	(-0.8200,0.2451)
	(0.0000,0.2668)	(-0.2400,0.2474)	(-0.8100,0.0238)
	(0.3100,0.3666)	(0.0900,0.0118)	(-0.2800,0.0095)
		(0.1000,0.2412)	(-0.2700,0.2378)
		(0.4600,0.2357)	(0.0800,0.2478)
			(0.4400,0.1841)
			(0.4500,0.0520)
CPU (secs)	6.8484	46.5195	52.9467

$(x.xxxx, w.wwww)$ = (design point, weight).

TABLE 9
Bayesian D-, A- and E-optimal designs with independent uniform prior densities for the power logistic model when $X \in [-1, 1]$, $\Theta = [-0.3, 0.3] \times [6.0, 8.0] \times [0.5, 1.0]$.

	D-optimal design	A-optimal design	E-optimal design
	(-0.7000,0.2638)	(-0.8400,0.1369)	(-0.8500,0.0363)
	(-0.2400,0.2474)	(-0.2700,0.1418)	(-0.8400,0.0965)
	(0.0900,0.0118)	(-0.0200,0.0287)	(-0.2600,0.0871)
	(0.1000,0.2412)	(-0.0100,0.0083)	(-0.2500,0.0505)
	(0.4600,0.2357)	(0.1500,0.1721)	(0.1200,0.2035)
		(0.1600,0.0198)	(0.5900,0.5262)
		(0.5700,0.3474)	
		(0.5800,0.1450)	
CPU (secs)	46.5195	34.7882	7.0668

$(x.xxxx, w.wwww)$ = (design point, weight).

To fix ideas $m = 2$, $X \in [0, 1]$, and $\Theta \equiv [\beta_0^L, \beta_0^U] \times [0, \beta_1^U]$. The FIM for any design for this model is given in Braess and Dette (2007),where they showed this matrix is does not depend on β_0 and consequently the optimal design depends only on the prior density for β_1 and its domain. We assumed the the prior density for β_1 is uniform distribution over its domain which has the form domain $[0, \beta_1^U]$. To implement SDP, we discretized the design interval using $\Delta x = 0.01$ and an integration scheme based on 7 points GQF over $[0, \beta_1^U]$. This results in a total of $\iota = 7$ points to evaluate the expectation with selected Bayesian D-optimal designs for different values of β_1^U shown in Table 10. The optimal designs were able to capture a phenomenon commonly observed in Braess and Dette (2007) and several others that more uncertainty in prior information for the model parameter as reflected by a larger domain in this case requires more support points in the optimal design. We note that in Table 10, the extreme ends of the design interval are always support points of the optimal design and the weight of the optimal design at $x = 0.0000$ decreases when β_1^U increases, as will be the case when prior information on β_1 becomes increasingly vague.

TABLE 10
Bayesian D-optimal designs with a uniform prior on $[0.0, \beta_1^U]$ *for* β_1 *for the exponential growth model with one regressor when* $X \in [0, 1]$.

	$\beta_1^U = 1.0$	$\beta_1^U = 5.0$	$\beta_1^U = 20.0$	$\beta_1^U = 50.0$
	(0.0000,0.5000)	(0.0000,0.4435)	(0.0000,0.3152)	(0.0000,0.0792)
	(1.0000,0.5000)	(0.3300,0.3965)	(0.0900,0.4442)	(0.0400,0.4544)
		(1.0000,0.1600)	(0.5300,0.1215)	(0.1800,0.1128)
			(1.0000,0.1191)	(1.0000,0.3536)
CPU (secs)	4.3836	5.7408	5.3196	5.2572

$(x.xxxx, w.wwww)$ = (design point, weight).

5.3. Simple logistic model with two regressors

In this section we consider a logistic model with $m = 3$ parameters, $\beta_i,\ i \in \{0, 1, 2\}$, and two regressors discussed in Haines et al. (2007):

$$p(x, \theta) = \frac{1}{1 + \exp(\beta_0 + \beta_1\ x_1 + \beta_2\ x_2)}, \quad x_1 \in X_1,\ x_2 \in X_2,$$
$$\theta^{\mathrm{T}} = [\beta_0, \beta_1, \beta_2] \in\ \Theta \tag{11}$$

Following Haines et al. (2007), we assumed β_1 and β_2 are known (set equal to 1) and β_0 is the unknown intercept parameter with plausible values between two known limits β_0^L and β_0^U. The design problem is to find a Bayesian design to estimate β_0 assuming $\Theta = [\beta_0^L, \beta_0^U] \times \{1\} \times \{1\}$ using various uniform prior densities for β_0 on $[\beta_0^L, \beta_0^U]$. The FIM for any design for this model can be directly worked out and is given in Haines et al. (2007).

To fix ideas, we set $X_1 \equiv X_2 \equiv [0, 6]$ and discretize each of the design spaces using a step size of 0.06. . The integration in the design criterion was evaluated using a 6-point GQF. Table 11 presents Bayesian D-optimal designs when the prior density is (i) degenerate at $\beta_0 = -4.0$; (ii) uniform on $[-4.0, -1.0]$; (iii) uniform on $[-4.0, 2.0]$. The results for (i) agree closely with those presented by Haines et al. (2007), noting that the differences arise mainly from the discretization scheme employed by SDP. We note that as the uncertainty of the value of θ_0 increases, the plausible region Θ grows in size and the CPU time required to compute the Bayesian optimal design increases. This is a common trend observed here and in other examples reported in the literature (Chaloner and Larntz, 1989). One also observes that the computing time to find A-optimal designs is similar to that for D-optimal designs.

Table 12 further compares Bayesian D-, A- and E-optimal designs when β_0 is uniformly distributed on the interval $[-4.0, 2.0]$ and $X \equiv [0, 6] \times [0, 6]$ for the logistic model with two regressors. One observes that the E-optimal designs has the largest weight at the point $(x_1, x_2) = (0.0000, 0.0000)$, followed by A- and D-optimal designs, respectively. The A-optimal design has 9 support points, the E-optimal design has 5 points and the D-optimal design has 7 points.

TABLE 11
Bayesian D-optimal designs for the logistic model with two regressors when $X \in [0,6] \times [0,6]$ *and uniform prior distributions over different* Θ *regions.*

	$\Theta = \{-4.0\}$	$\Theta \in [-4.0, -1.0]$	$\Theta \in [-4.0, 2.0]$
	(0.0000,2.7000,0.1896)	(0.0000,1.0800,0.1184)	(0.0000,0.0000,0.2887)
	(0.0000,5.3400,0.3104)	(0.0000,1.1400,0.0588)	(0.0000,2.4600,0.1663)
	(2.7000,0.0000,0.1896)	(0.0000,4.0200,0.0263)	(0.0000,3.7200,0.0495)
	(5.3400,0.0000,0.3104)	(0.0000,4.0800,0.2965)	(0.0000,3.7800,0.1399)
		(1.0800,0.0000,0.1184)	(2.4600,0.0000,0.1663)
		(1.1400,0.0000,0.0588)	(3.7200,0.0000,0.0495)
		(4.0200,0.0000,0.0263)	(3.7800,0.0000,0.1399)
		(4.0800,0.0000,0.2965)	
CPU (secs)	69.0616	69.4048	75.1769

$(x1.xxxx, x2.xxxx, w.wwww)$ = (design point, weight).

TABLE 12
Bayesian D-, A- and E-optimal designs when $\beta_0 \approx U[-4.0, 2.0]$ *for the logistic model with two regressors when* $\Theta = [-4.0, 2.0]$ *and* $X \in [0,6] \times [0,6]$.

	D-optimal design	A-optimal design	E-optimal design
	(0.0000,0.0000,0.2887)	(0.0000,0.0000,0.3944)	(0.0000,0.0000,0.6244)
	(0.0000,2.4600,0.1663)	(0.0000,2.0400,0.1782)	(0.0000,3.1200,0.1481)
	(0.0000,3.7200,0.0495)	(0.0000,2.1000,0.0391)	(0.0000,3.1800,0.0397)
	(0.0000,3.7800,0.1399)	(0.0000,4.5600,0.0385)	(3.1200,0.0000,0.1481)
	(2.4600,0.0000,0.1663)	(0.0000,4.6200,0.0470)	(3.1800,0.0000,0.0397)
	(3.7200,0.0000,0.0495)	(2.0400,0.0000,0.1782)	
	(3.7800,0.0000,0.1399)	(2.1000,0.0000,0.0391)	
		(4.5600,0.0000,0.0385)	
		(4.6200,0.0000,0.0470)	
CPU (secs)	75.1769	67.4860	68.5936

$(x1.xxxx, x2.xxxx, w.wwww)$ = (design point, weight).

5.4. Two-dimensional model with a Gamma distributed response

Here we consider a model with six parameters with two predictors to further test the ability of SDP to find the optimal design. The generalized linear model is defined on the design space $X \in [0,1] \times [0,1]$ and has a Gamma distributed response with a two-dimensional mean function given by

$$\mu = \mathbb{E}(y|x) = g^{-1}(\beta_0 + \beta_1\ x_1 + \beta_2\ x_2 + \beta_3\ x_1^2 + \beta_4\ x_2^2 + \beta_5\ x_1\ x_2) \qquad (12)$$

where $g(\mu) = 1/\mu$ is the link function and the region for the parameter values is $\Theta \equiv [\beta_0^L, \beta_0^U] \times [\beta_1^L, \beta_1^U] \times [\beta_2^L, \beta_2^U] \times [\beta_3^L, \beta_3^U] \times [\beta_4^L, \beta_4^U] \times [\beta_5^L, \beta_5^U]$. The FIM for this model is given in Dette et al. (2013). To implement SDP, we discretized the design space using $\Delta x_1 = \Delta x_2 = 0.1$ and used an integration scheme based on a 3-point GQF for each of the six components of Θ. This requires evaluation of the expectation using $\iota = 3^6 = 729$ points. Following Dette et al. (2013), we employed independent uniform prior for each parameter with $\beta_0^L = 0.5$, $\beta_1^U = 2.0$, $\beta_1^L = \beta_2^L = \beta_3^L = \beta_4^L = \beta_5^L = 0.0$, and $\beta_1^U = \beta_2^U = \beta_3^U = \beta_4^U = \beta_5^U = 1.0$. The SDP-generated Bayesian optimal designs for different criteria are shown in Table 13.

In this example, the SDP solver we employed experienced difficulty in handling the large number of constraints and variables to optimize. A main reason

is that the integration schemes used to compute the expectation require a large number of points ι to minimize the numerical error that in turn requires SDP representations to have a large number of LMIs and equality constraints. Consequently, this presents a more and more challenging task for for the IP solvers. Table 13 reports our SDP- generated designs and the two last lines in the table show the large number of variables and the large number of equality constraints involved in the optimization problem. For this problem, the high dimensionality of the problem will only increase exponentially if higher order GQFs are used. To avoid this problem, we may reduce the number of discretization points in the space of regressors, however this strategy might lead to designs with low efficiency since the small number of candidate points initially considered may not adequately capture the features of the problem. One way to overcome this limitation might be to use of grid adaptive algorithms combined with SDP or use a Monte-Carlo sampling scheme to perform the integration. We do not discuss these latter two possibilities here.

TABLE 13

Bayesian optimal designs for the quadratic mean model with 2 *regressors and a Gamma distributed response when* $X \in [0,1] \times [0,1]$ *and independent uniform prior density for each of the* 6 *parameters* ($\beta_0^L = 0.5$, $\beta_0^U = 2.0$, $\beta_1^L = \beta_2^L = \beta_3^L = \beta_4^L = \beta_5^L = 0$, *and* $\beta_1^U = \beta_2^U = \beta_3^U = \beta_4^U = \beta_5^U = 1$).

	D-optimal design	A-optimal design	E-optimal design
	(0.0,0.0,0.1532)	(0.0,0.0,0.1393)	(0.0,0.0,0.1238)
	(0.0,0.4,0.1333)	(0.0,0.4,0.1668)	(0.0,0.4,0.1659)
	(0.0,1.0,0.1651)	(0.0,1.0,0.0942)	(0.0,1.0,0.0581)
	(0.4,0.0,0.1333)	(0.4,0.0,0.1668)	(0.4,0.0,0.1659)
	(0.4,0.4,0.0880)	(0.4,0.4,0.1593)	(0.4,0.4,0.1985)
	(1.0,0.0,0.1651)	(0.4,1.0,0.0386)	(0.4,1.0,0.0099)
	(1.0,1.0,0.1619)	(1.0,0.0,0.0942)	(1.0,0.0,0.0581)
		(1.0,0.4,0.0386)	(1.0,0.3,0.0646)
		(1.0,1.0,0.1022)	(1.0,0.4,0.0099)
			(1.0,1.0,0.0807)
CPU (secs)	260.1629	134.8941	61.4176
# vars	73049	56984	15431
# eqls	21271	15430	850

$(x.x, x.x, w.wwww)$ = (design point, weight)
vars - number of variables involved in the SDP problem
eqls - number of equality constraints involved in the SDP problem

6. Discussion

Our work here illustrates how to formulate an optimal design problem into a SDP problem and find optimal designs for nonlinear models. We focus on Bayesian optimal designs and show SDP was able to verify published results in the literature. We produced new Bayesian A-, E- and D-optimal designs for the logistic model using bivariate prior normal distributions and explored how the correlation coefficient affects the optimal design, along with the choice of the discretized design space and the number of points used in the GQF approximation to the integration problem. We also applied SDP to find Bayesian

D-optimal designs for the more complicated power logistic model, a logistic model with two variables and a gamma distributed response model with two variables.

SDP provides a general framework for determining locally optimal designs for linear and nonlinear models with several regressors and parameters. We showed here that extensions of SDP to find Bayesian optimal designs for nonlinear models can be systematically carried out. The problem formulation relies on robust and accurate solvers that guarantee a global optimal solution that depends on the grid set. For SDP-generated optimal design to be globally optimal, the discretization scheme should be sufficiently fine to capture all the true design points. The global optimality of the SDP-generated design can be verified using an equivalence theorem available for each convex design criterion. For instance, if we let δ_x be the point mass design at x, the equivalence theorems for D- and A-optimality are as follow:

ξ_D is globally D-optimal if and only if

$$\operatorname{tr}\left\{[\mathcal{M}(\xi_D,\theta)]^{-1}\ \mathcal{M}(\delta_x,\theta)\right\} - \operatorname{card}(\theta) \le 0, \quad \forall x \in X,$$

and

ξ_A is globally A-optimal if and only if

$$\operatorname{tr}\left\{[\mathcal{M}(\xi_A,\theta)]^{-2}\ \mathcal{M}(\delta_x,\theta)\right\} - \operatorname{tr}\left\{[\mathcal{M}(\xi_A,\theta)]^{-1}\right\} \le 0, \quad \forall x \in X.$$

E-optimal designs do not have a differentiable criterion and the equivalence theorem is more complicated. We do not report it here but its details, along with derivation of the other equivalence theorems are available in design monographs, such as Fedorov (1972) and Pukelsheim (1993). The equivalence theorems can be extended to their Bayesian versions directly. For instance, for D-optimality, ξ_{BayesD} is Bayesian D-optimal among all designs on X if and only if:

$$\int_\Theta \operatorname{tr}\left\{[\mathcal{M}(\xi,\theta)]^{-1}\ \mathcal{M}(\delta_x,\theta)\right\}\ \pi(\theta)\ \mathrm{d}\theta - \operatorname{card}(\theta) \le 0, \quad \forall x \in X.$$

Corresponding equivalence conditions can be stated similarly for Bayesian A- and E-optimality. We note that in all the above equivalence theorems, equality holds at the support points of the optimal design. In practice, one may verify the optimality of the SDP-generated design by checking whether the above inequality is satisfied with equality at the support points. For other convex Bayesian design criterion, one can similarly calculate the directional derivative of the criterion at the SDP-generated design to obtain an equivalence theorem and perform the integration using GQF. For a high level of precision in the integration, one may have to use higher degree Legendre polynomials and refine the discretized design space to verify optimality.

Figure 1(a) displays the directional derivative of the Bayesian D-optimality criterion for the SDP-generated design when $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$. The plot

satisfies the conditions in the equivalence theorem and confirms that the 3-point design is optimal. Figures 1(b) and 1(c) show the directional derivatives of the A- and E-optimality criteria at the SDP-generated designs and the plots also confirm their optimality. The corresponding plot for the Bayesian D-optimal design for the power logistic model with the additional parameter $s \in [0.5, 1.0]$ is shown in Figure 1(d).

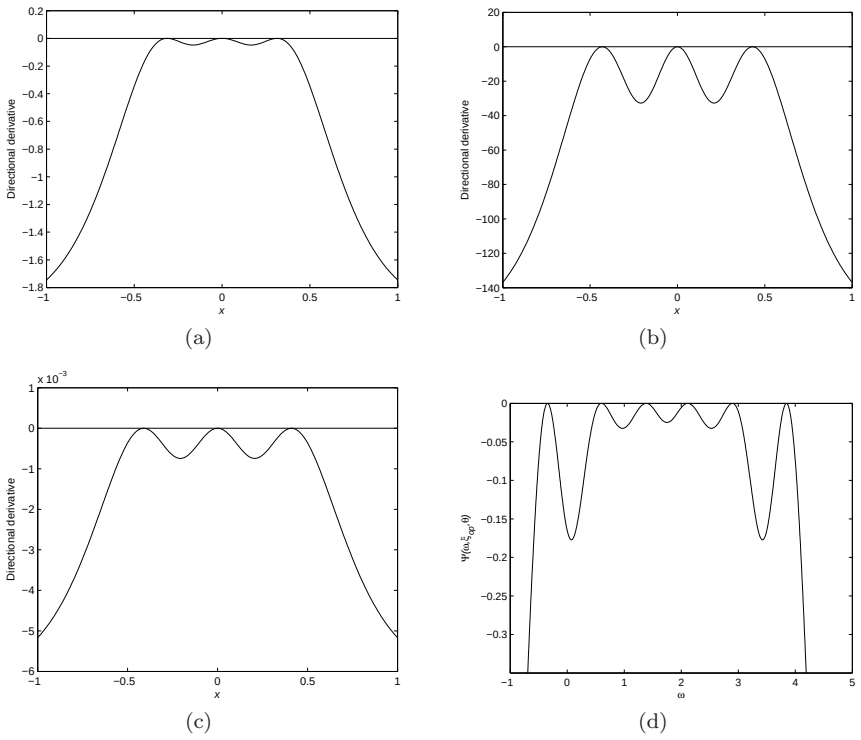

FIG 1. *Directional derivative for the Bayesian criterion of the SDP-generated design on* $X = [-1, 1]$ *with independent uniform priors on* $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$ *for the logistic model: (a) D-optimality, (b) A-optimality, (c) E-optimality; and D-optimality with independent uniform priors on* $\Theta = [-0.3, 0.3] \times [6.0, 8.0] \times [0.5, 1.0]$ *for the power logistic model.*

Figure 2 displays the directional derivative plots for the SDP-generated designs using bivariate normal priors with varying degrees of correlation between the two components. Likewise, Figure 3 displays the directional derivative for the SDP-generated Bayesian D-optimal design for the logistic model with two regressors and a unknown intercept with a uniform prior distribution on $\beta_0 \in [-4.0, -1.0]$ and the design space is $X = [0, 6] \times [0, 6]$.

In summary, our experience is that if one is willing to discretize the design space to find locally or Bayesian optimal designs, SDP is an effective tool. Our last example with 2 regressors and a Gamma distributed response shows that SDP can encounter problems when the user-selected solver cannot solve the ap-

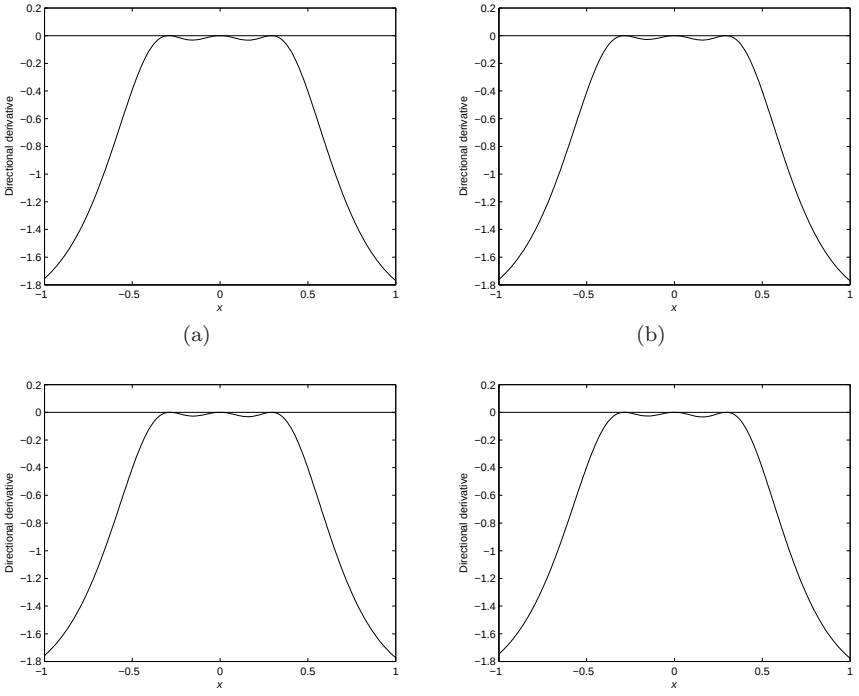

FIG 2. *Directional derivatives for the Bayesian D-optimal criterion of the SDP-generated designs for the simple logistic model when* $\Theta = [-0.3, 0.3] \times [6.0, 8.0]$, $X = [-1, 1]$ *and bivariate normal prior densities in Table 5 with different covariances: (a)* $\sigma_{\mu,\beta} = 0.05$*; (b)* $\sigma_{\mu,\beta} = 0.075$*; (c)* $\sigma_{\mu,\beta} = 0.10$*; (d)* $\sigma_{\mu,\beta} = 0.15$.

proximated optimization problem with thousands of variables to be optimized and hundreds of constraints that have to be satisfied. Otherwise, SDP is generally effective for finding locally and Bayesian optimal designs when there are not too many variables for the solver cvx to handle. Despite the general effectiveness of SDP, it is a curiosity that SDP is not used more often in mainstream statistical applications. We hope that this paper encourages more statisticians to consider using SDP in their future work.

7. Acknowledgements

This paper was initiated when Wong was a visiting fellow at the Isaac Newton Institute of Mathematical Sciences in Cambridge where he also served as scientific advisor for a 6-month workshop on design and analysis of experiments in 2011. Dr. Wong wishes to express his thanks to the institute for the support for his repeated visits to the workshop, to Professor Rosemary Bailey for hosting the workshop and to Dr. Radoslav Harman for stimulating discussion

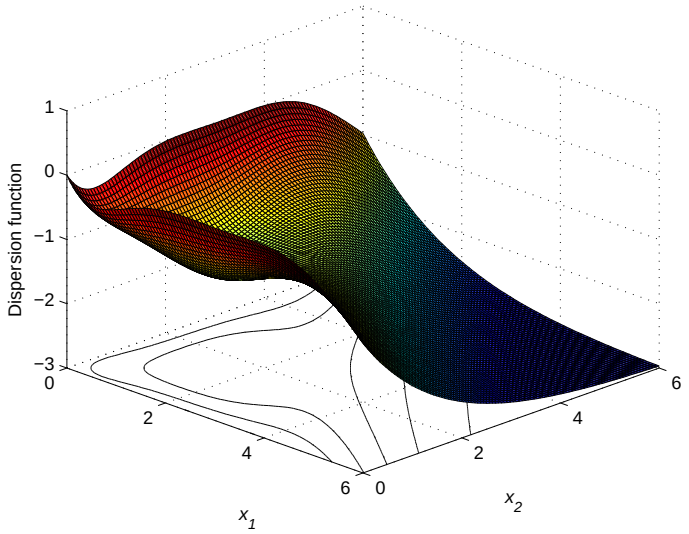

FIG 3. *Directional derivative of the Bayesian D-optimal criterion at the SDP-generated design for the simple logistic model with two regressors on* $X = [0, 6] \times [0, 6]$ *when* $\beta_1 = \beta_2 = 1$ *and a uniform prior on* $[-4.0, 2.0]$ *is used for the intercept* β_0.

at the institute on use of semidefinite programming for finding optimal experimental designs. We also thank Professor Lieven Vandenberghe at UCLA for his comments on an earlier version of this paper and are also grateful to the entire editorial team for their many helpful comments and valuable suggestions to improve the paper.

References

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. Dover Pulications, Inc., New York.

ATKINSON, K. E. (1989). *An Introduction to Numerical Analysis*, 2nd. ed. John Willey & Sons, New York.

BEN-TAL, A. and NEMIROVSKIĬ, A. S. (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Society for Industrial and Applied Mathematics, Philadelphia.

BERGER, M. P. F. and WONG, W. K. (2009). *An Introduction to Optimal Designs for Social and Biomedical Research*. John Wiley & Sons.

BERNARDO, F. P., PISTIKOPOULOS, E. N. and SARAIVA, P. M. (1999). Integration and computational issues in stochastic design and planning optimization problems. *Ind. Eng. Chemistry Research* **38** 3056-3068.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. University Press, Cambridge.

Braess, D. and Dette, H. (2007). On the number of support points of maximin and Bayesian optimal designs. *The Annals of Statistics* **35** 772-792.

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

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statist. Science* **10** 273-304.

Chang, F. and Lin, H. (2007). On minimally-supported D-optimal designs for polynomial regression with log-concave weight function. *Metrika* **65** 227-233.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* **24** 586–602.

Cochran, W. G. (1973). Experiments for nonlinear functions (r.a. Fisher Memorial Lecture). *Journal of American Statistical Association* **68** 771–781.

Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*, 2nd. ed. Academic Press, New York.

Dette, H., Melas, V. B. and Wong, W. K. (2005). Optimal gesign for goodness-of-fit of the Michaelis-Menten enzyme kinetic function. *Journal of the American Statistical Association* **100** 1370-1381.

Dette, H. and Studden, W. J. (1993). Geometry of E-optimality. *The Annals of Statistics* **21** 416–433.

Dette, H. and Wong, W. K. (1996). Bayesian optimal designs for models with partially specified heteroscedastic structure. *The Annals of Statistics* **24** 2108-2127.

Dette, H. and Wong, W. K. (1998). Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models. *Biometrika* **85** 869–882.

Dette, H., Hoyden, L., Kuhnt, S. and Schorning, K. (2013). Optimal designs for additional day effects in generalized linear models with gamma distributed response. Discussion Paper. `http://www.statistik.tu-dortmund.de/fileadmin/user_upload/SFB_823/discussion_papers/2013/DP_0513_SFB823_Dette_Hoyden_Kuhnt__Schorning.pdf`.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press.

Fedorov, V. V. and Hackl, P. (1997). *Model-oriented Design of Experiments.* Springer, New York.

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

Gaffke, N. and Kraft, O. (1982). Exact D-optimum designs for quadratic regression. *Journal of Royal Statist. Soc. Ser. B* **44** 394-397.

Gerald, C. F. and Wheatley, P. O. (1994). *Applied Numerical Analysis*, 5th. ed. Addison-Wesley Publishing Comp., New York.

Grant, M., Boyd, S. and Ye, Y. (2012). `cvx` Users Guide for `cvx` version 1.22, 1104 Claire Ave., Austin, TX 78703-2502.

Haines, L. M., Kabera, G., Ndlovu, P. and O'Brien, T. E. (2007). D-optimal designs for logistic regression in two variables. In *mODa 8 - Advances in Model Oriented Design and Analysis* (J. López-Fidalgo, J. M. Rodríguez-

Díaz and B. Torsney, eds.) 91-98. Physica Verlag, Heidelberg.

Harman, R. and Pronzato, L. (2007). Improvements on removing non-optimal support points in D-optimum design algorithms. *Statistics and Probability Letters* **77** 90–94.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M. and Borror, C. M. (2004). Model-robust optimal designs: A Genetic Algorithm approach. *Journal of Quality Technology* **36** 263–279.

Kiefer, J. C. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society, Series B* **21** 272–319.

Kiefer, J. C. (1985). Design of Experiments. In *Jack Carl Kiefer Collected Papers. III* (L. D. Brown, I. Olkin, J. Sacks and H. P. Wynn, eds.) Springer-Verlag and the Institute of Mathematical Statistics, Netherlands.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problem. *Canadian Journal of Mathematics* **12** 363-366.

King, J. and Wong, W. K. (2000). Minimax D-optimal designs for logistic model. *Biometrics* **56** 1263-1267.

López-Fidalgo, J. and Garcet-Rodriguez, S. A. (2004). Optimal experimental designs when some independent variables are not subject to control. *Journal of the American Statistical Association* **99** 1190-1199.

Molchanov, I. and Zuyev, S. (2002). Steepest descent algorithm in a space of measures. *Statistics and Computing* **12** 115–123.

Nesterov, Y. and Nemirovskiĭ, A. (1994). *Interior-point polynomial algorithms in convex programming*. Society for Industrial and Applied Mathematics, Philadelphia.

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

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics* **32** 761-768.

Pronzato, L. (2008). Optimal experimental design and some related control problems. *Automatica* **44** 303-325.

Pukelsheim, F. (1980). On linear regression designs which maximize information. *Journal of Statistical Planning and Inference* **4** 339–364.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. SIAM, Philadelphia.

Qi, H. (2011). A semidefinite programming study of the Elfving theorem. *Journal of Statistical Planning and Inference* **141** 3117–3130.

Reber, G. S. (2004). Gaussian Quadrature for Computer Aided Robust Design PhD thesis, Massachussets Institute of Technology.

Reznik, Y. A. (2008). Continued fractions, Diophantine approximations, and design of color transforms. In *Proc. SPIE, Applications of Digital Image Processing XXXI* **7073**.

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

Sitter, R. R. and Wu, C. J. (1999). Two-stage design of quantal response studies. *Biometrics* **55** 396–402.

Smith, K. (1918). On the standard deviation of adjusted and interpolated

values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* **12** 1–85.

Sturm, J. (1999). Using `SeDuMi` 1.02, a Matlab toolbox for optimization oversymmetric cones. *Optimization Methods and Software* **11** 625–653.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Review* **8** 49-95.

Vandenberghe, L., Boyd, S. and Wu, S. (1998). Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Analysis and Applications* **19** 499–533.

Vandenberghe, L. and Boyd, S. (1999). Applications of semidefinite programming. *Applied Numerical Mathematics* **29** 283-299.

Wolkowicz, H., Saigal, R. and Vandenberghe, L. (2000). *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Springer Netherlands, Netherlands.

Wong, W. K. (1992). A unified approach to the construction of minimax designs. *Biometrika* **79** 611-620.

Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. *Journal of Royal Statistics Soc. - Ser. B* **34** 133-147.

Ye, Y. (1997). *Interior Point Algorithms: Theory and Analysis*. Wiley.

Zhang, Y. (2006). Bayesian D-Optimal Design for Generalized Linear Models PhD thesis, Virginia Polytechnic Institute and State University.