Title
A Semi-Infinite Programming based algorithm for determining T-optimum designs for model discrimination

Journal
Journal of Multivariate Analysis, 135

ISSN
0047259X

Authors
Duarte, Belmiro P.M.
Wong, Weng Kee
Atkinson, Anthony C

Publication Date
2015-03-01

DOI
10.1016/j.jmva.2014.11.006

Peer reviewed

Journal of Multivariate Analysis 135 (2015) 11–24

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A Semi-Infinite Programming based algorithm for determining T-optimum designs for model discrimination

Belmiro P.M. Duarte [a,b,*], Weng Kee Wong [c], Anthony C. Atkinson [d]

[a] GEPSI – PSE Group, CIEPQPF, Department of Chemical Engineering, University of Coimbra, Pólo II, R. Sílvio Lima, 3030-790 Coimbra, Portugal

[b] Department of Chemical and Biological Engineering, ISEC, Polytechnic Institute of Coimbra, R. Pedro Nunes, 3030-199 Coimbra, Portugal

[c] Department of Biostatistics, Fielding School of Public Health, UCLA, 10833 Le Conte Ave., Los Angeles, CA 90095-1772, USA

[d] Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom

ARTICLE INFO

Article history:
Received 5 September 2013
Available online 11 December 2014

AMS subject classifications:
62K05
90C22

Keywords:
Continuous design
Equivalence theorem
Global optimization
Maximum likelihood design
Minimax program
Semi-Infinite Programming

ABSTRACT

T-optimum designs for model discrimination are notoriously difficult to find because of the computational difficulty involved in solving an optimization problem that involves two layers of optimization. Only a handful of analytical T-optimal designs are available for the simplest problems; the rest in the literature are found using specialized numerical procedures for a specific problem. We propose a potentially more systematic and general way for finding T-optimal designs using a Semi-Infinite Programming (SIP) approach. The strategy requires that we first reformulate the original minimax or maximin optimization problem into an equivalent semi-infinite program and solve it using an exchange-based method where lower and upper bounds produced by solving the outer and the inner programs, are iterated to convergence. A global Nonlinear Programming (NLP) solver is used to handle the subproblems, thus finding the optimal design and the least favorable parametric configuration that minimizes the residual sum of squares from the alternative or test models. We also use a nonlinear program to check the global optimality of the SIP-generated design and automate the construction of globally optimal designs. The algorithm is successfully used to produce results that coincide with several T-optimal designs reported in the literature for various types of model discrimination problems with normally distributed errors. However, our method is more general, merely requiring that the parameters of the model be estimated by a numerical optimization.

1. Introduction

Professor Jack Kiefer was an early proponent of using a rigorous mathematical framework to find optimal experimental designs for solving practical problems. In Kiefer [24,25] he introduced continuous designs, in which the design is represented by a measure. As a result, the problems of the dependence of the structure of the design on sample size are avoided. He advocated using such designs for practical reasons; research in this area has continued, largely motivated by rising experimental costs and the need to use resources more efficiently. Book length treatments of this topic include Pukelsheim [34], Fedorov and Hackl [16], Uciński [41], Atkinson et al. [4], Berger and Wong [7] and Fedorov and Leonov [17]. Early

* Correspondence to: Department of Chemical Engineering, University of Coimbra, Pólo II, R. Sílvio Lima, 3030-790 Coimbra, Portugal.
E-mail addresses: bduarte@isec.pt (B.P.M. Duarte), wkwong@ucla.edu (W.K. Wong), a.c.atkinson@lse.ac.uk (A.C. Atkinson).

http://dx.doi.org/10.1016/j.jmva.2014.11.006
0047-259X/

applications of optimal designs were concentrated in the engineering, manufacturing and industrial sectors but applications are increasingly also seen in the biomedical and social sciences.

Optimal designs can depend sensitively on the assumed model. They can lose substantial efficiency if the assumed model is wrong. In practice, the underlying model is unknown and frequently a few plausible alternative models are considered for studying the problem at hand. An optimal discrimination design provides the best strategy for collecting observations to identify the true model among those postulated. Optimal design problems for estimating model parameters are quite well studied but the search for the optimal discrimination design has received considerably less attention. One reason is that finding an optimal discrimination design is an appreciably more difficult task than finding a D-optimal design for estimating model parameters [45]. Unlike D-optimality, we now have an optimality criterion that requires two levels of optimization. To date, effective algorithms for finding these optimal designs for a general regression model remain elusive.

The theoretical framework for experimental design for model discrimination was established in a series of papers, such as Fedorov and Malyutov [18], Atkinson and Cox [3], Atkinson and Fedorov [5,6]. The criterion used for model discrimination is commonly known as T-optimality. The typical setup assumes that we want to discriminate between two parametric models, one of which is a fully parameterized "true model" and the other a "test model" with unknown parameters. The T-optimal design maximizes the lack of fit sum of squares for the second model by maximizing the minimal lack of fit sum of squares arising from a set of plausible values of the unknown parameters. Additional theoretical developments can be found in Ponce de Leon and Atkinson [33], Dette [9], Fedorov and Hackl [16], Wiens [46] and Dette and Titoff [12]. López-Fidalgo et al. [29] extend the method to models in which the errors of observation do not follow a normal distribution. T-optimality has been applied to discriminate among various classes of models, ranging from polynomial models [5,11], to Fourier regression models [10], Michaelis–Menten kinetic models [30], enzyme kinetics [2] and dynamic systems described by sets of ordinary differential equations [41,27,39].

There are analytical descriptions of T-optimal designs for only the simplest situations because of the complexity of the optimization problem. The algorithms commonly used to find T-optimal designs are based on modifications of the Wynn–Fedorov algorithm, which were initially proposed for D-optimal designs; see for example, Atkinson and Fedorov [5]. The method requires a user-selected starting design to initiate the search process before it iterates by sequentially adding one or more selected new points from the design space to the current design. At each iteration a new design is formed by mixing the new point or points appropriately chosen with the current design. The generated design accumulates many points or clusters of points over time and a judicious collapsing of these points into a smaller number of distinct points is periodically required. These are the core steps in the Wynn–Fedorov algorithm formed by aggregating ideas of Wynn [47] and Fedorov [15] and commonly used in computer algorithms for finding different types of optimal designs such as D_s-optimal designs for estimating a selected subset of the model parameters or L-optimal designs for estimating a selected linear function of the model parameters.

Two other approaches have been employed for determining optimal discrimination designs. Dette and Titoff [12] suggest the Remes algorithm from numerical approximation theory and demonstrated the method for problems with a single explanatory variable. Atkinson [2] employs a Quasi-Newton algorithm for convex optimization after applying a transformation on the design region and design weights to ensure that all constraints are satisfied. See also Atkinson et al. [4, Section 9.5] where more details on the method and examples can be found. However, both methods seem somewhat specialized and may not extend to find optimal discrimination designs for more general problems.

Algorithms based on Semi-Infinite Programming (SIP), a branch of mathematical programming, are becoming increasingly popular for solving the minimax programs in computer science, engineering and economics [37]. Several algorithms belonging to exchange methods, discretization methods and local reduction methods have been developed [36]. Coupled with global nonlinear programming (NLP) solvers, they are able to solve minimax programs of moderate dimension. Interestingly, there are only a couple of applications of mathematical programming SIP-based approaches to find minimax-type optimal designs even though the approach provides a general framework and a systematic approach that is guaranteed find such optimal designs. Our goal in this paper is to apply SIP-based algorithms to systematically find optimal discrimination designs and demonstrate their effectiveness using several examples for a variety of situations. Only non-sequential experimentation is considered here; readers interested in a sequential approach to design a study for model discrimination can refer to Atkinson and Fedorov [5,6].

Gribik and Kortanek [21] established a theoretical and general framework for searching minimax designs via SIP. Žakovíc and Rustem [48] found minimax D-optimal designs and Duarte and Wong [14] found various types of minimax optimal designs using SIP based on an exchange method. Kuczewski [27] and Skanda and Lebiedz [39] used a SIP algorithm to find T-optimal designs for dynamic models using algorithms similar to that proposed by Žakovíc and Rustem [48] for general minimax problems.

Uciński and Bogacka [42] used a SIP based algorithm to find T-optimal designs for dynamic models. The SIP procedure relies on the relaxation paradigm proposed by Shimizu and Aiyoshi [38] for minimax problems. All optimization problems included in the SIP procedure are solved with a global solver employing a stochastic NLP solver with an adaptive random search scheme to generate initial solutions. There seems to be no application of SIP to finding T-optimal designs for discriminating between algebraically specified models. Uciński and Bogacka [42], Kuczewski [27] and Skanda and Lebiedz [39] deal with dynamic models and aim to determine the optimal discrimination design in the time domain (time instants where samples are to be gathered). Our paper aims to present and test a SIP based algorithm for finding T-optimal designs for algebraic models, both linear and nonlinear. It shares several properties with the procedure proposed by Uciński and

Bogacka [42]. We include a check from the equivalence theorem which allows us to automate the finding of the optimal number of support points.

Section 2 provides the background, and introduces the T-optimality criterion along with a practical tool for checking whether a design is optimal among all designs on the given design space. It also presents the conceptualization of the minimax program representing the T-optimality criterion as a SIP, and briefly reviews the exchange method for handling semi-infinite programs. Section 3 applies the SIP based algorithm to find T-optimal designs and an automated procedure for confirming the optimality of the SIP-generated design. We report these T-optimal designs for various discrimination problems in Section 4 and offer a conclusion in Section 5.

2. Background

This section is divided into two parts. The first discusses the statistical setup and use of continuous designs as a practical tool to solve general design problems. The second part provides background on SIP-based methods and how they relate to finding an optimal discrimination design and more generally solving minimax design problems.

2.1. Continuous designs

In this paper, we focus on continuous designs on a given compact design space of the regressors $\mathcal{X} \subset \mathbb{R}^{n_x}$. A continuous design is characterized by the number of design points it has from the design space $\mathcal{X}$, the locations of the points and the proportions of the total number of observations n to be taken at each of the design points. Let $x_i \in \mathcal{X}$ be the ith design point or support point of the design, let k be the number of design points and let w_i be the proportion of observations to be taken at x_i, $i = 1, \ldots, k$. Clearly, w_i is positive and less than unity (unless $k = 1$) and $w_1 + w_2 + \cdots + w_k = 1$. The total sample size n is usually predetermined by cost considerations. Continuous designs have continuous weights in $w_i \in [0, 1]$ which lead naturally to the formulation of the optimal design problem as a mathematical program with convex properties. Advantages of working with continuous designs are that they are easier to find and understand than exact designs that depend on n. We denote such a continuous design with k points by

$$\xi = \begin{pmatrix} x_1 & \cdots & x_i & \cdots & x_k \\ w_1 & \cdots & w_i & \cdots & w_k \end{pmatrix}$$

and denote the set of all continuous designs with k points on $\mathcal{X}$ by $\Xi \equiv \mathcal{X}^k \times [0, 1]^k$.

For exact designs, we require that all $n \times w_i$'s are positive integers. In this case, we would have to solve a much harder non-convex optimization problem. Pukelsheim and Rieder [35] describe an efficient method for rounding a continuous design to obtain a nearly optimum exact design of size n. Goos and Jones [20] give examples of finding exact D-optimal designs using a coordinate-exchange algorithm.

For model discrimination design problems, we seek a continuous design that is efficient for identifying the best fitting model from a given class of models. When there are two models and the outcome variable is Y, we designate one as the "true model" $\eta_t(x, \theta_1) = E(Y|x, \theta_1)$ and the other as the "test model" $\eta_2(x, \theta_2) = E(Y|x, \theta_2)$. The vectors of model parameters θ_1 and θ_2 may have different dimensions, but lie in known sets Θ_1 and Θ_2, i.e. $\theta_1 \in \Theta_1 \subset \mathbb{R}^{p_1}$ and $\theta_2 \in \Theta_2 \subset \mathbb{R}^{p_2}$. Following convention, we assume the "true model" is fully parameterized and so the dependence on θ_1 can be discarded and we may write its mean function simply as $\eta_t(x)$.

A common design criterion called T-optimality for model discrimination was proposed by Atkinson and Fedorov [5] and Atkinson et al. [4]. The T-optimal design is defined by:

$$\begin{aligned} \xi_T &= \arg\max_{\xi \in \Xi} \min_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} [\eta_t(x) - \eta_2(x, \theta_2)]^2 \ \xi(\mathrm{d}x) \\ &= \arg\min_{\xi \in \Xi} \max_{\theta_2 \in \Theta_2} - \int_{\mathcal{X}} [\eta_t(x) - \eta_2(x, \theta_2)]^2 \ \xi(\mathrm{d}x). \end{aligned} \tag{1}$$

Employing results from Rustem and Howe [37], problem (1) is equivalent to the bilevel program

$$\begin{aligned} &\xi_T = \arg\min_{\xi \in \Xi} - \int_{\mathcal{X}} \left[\eta_t(x) - \eta_2(x, \theta_2^*)\right]^2 \ \xi(\mathrm{d}x) \\ &\text{s.t.} \sum_{i=1}^{k} w_i = 1 \\ &\theta_2^* = \arg\max_{\theta_2 \in \Theta_2} - \int_{\mathcal{X}} [\eta_t(x) - \eta_2(x, \theta_2)]^2 \ \xi(\mathrm{d}x), \end{aligned} \tag{2}$$

showing that the T-optimality criterion can be equivalently viewed as a maximin, a minimax or a bilevel optimization problem with the outer program having convex properties and the inner problem being concave or convex. An important quantity in the above definition is the least favorable parametric configuration θ_2^* in Θ_2, which is frequently problematic to determine numerically and presents a constant source of difficulty for finding the optimal discrimination design, and more generally for minimax or maximin optimal designs in practice.

The search for the optimal discrimination design ξ_T is nested within the number of support points of the design. To avoid the complexity of simultaneously find the design and the number of support points, a nonconvex optimization problem, we fix k and start the search over all k-point designs. The resulting design ξ_T^k may or may not be optimal among all designs on Ξ. An equivalence theorem similar to those given in Kiefer and Wolfowitz [26] and Kiefer [25] is then used to check whether $\xi_T = \xi_T^k$. The mathematical program to solve the problem is:

$$\begin{aligned} &\Delta(\xi_T^k) = \min_{\xi\in\Xi} \max_{\theta_2\in\Theta_2} -\sum_{i=1}^{k} [\eta_t(x_i) - \eta_2(x_i,\theta_2)]^2 \ w_i \\ &\text{s.t.} \sum_{i=1}^{k} w_i = 1. \end{aligned} \tag{3}$$

A common choice for initializing k is the number of parameters in the model plus one. A theoretical justification for the choice of the value of k is possible only in specialized settings. For example, Dette and Titoff [12] proved that, for nested polynomials in one variable, $k = p_2 + 1$. Our numerical results in Section 4 support such a value for k. For T-optimality, the theorem asserts that the design ξ_T^k is optimal among all designs on $\mathcal{X}$ if and only if

$$\left[\eta_t(x) - \eta_2(x,\theta_2^k)\right]^2 \le -\Delta(\xi_T^k), \quad \forall x \in \mathcal{X}, \tag{4}$$

with equality at the support points of ξ_T^k and θ_2^k is defined similarly as θ_2^* [4]. The function on the left hand side of the above inequality is called the sensitivity function. Of course if the trial value of k is indeed the number of support points of the optimal discrimination design, the equivalence theorem holds and we have $\xi_T^k = \xi_T$ and $\theta_2^k = \theta_2^*$. The theorem applies to continuous designs, but not to exact designs.

2.2. Semi-Infinite Programming

Hettich and Kortanek [23] and López and Still [28] provide surveys of the theory, applications and recent development of SIP methodology. Broadly speaking, the numerical methods employed to solve SIP problems fall into three classes: exchange methods, discretization based methods and local reduction based methods [22]. Here we use an exchange based procedure similar to the one proposed by Blankenship and Falk [8], and further expounded in Žakovíc and Rustem [48] among others. To this end, consider the general minimax program formalization used by Rustem and Howe [37] and Žakovíc and Rustem [48]:

$$\begin{aligned} \min_{y}\max_{z} \quad & f(y,z) \\ \text{s.t.} \quad & g_{l_1}(y,z) \le 0, \quad l_1 \in \{1,\dots,N_I\} \\ & h_{l_2}(y,z) = 0, \quad l_2 \in \{1,\dots,N_E\} \\ & y \in \mathcal{Y}, \qquad z \in \mathcal{Z}, \end{aligned} \tag{5}$$

where $y \in \mathcal{Y} \subset \mathbb{R}^{n_y}$ are the outer problem decision variables and $z \in \mathcal{Z} \subset \mathbb{R}^{n_z}$ are the decision variables of the inner problem. The set $\mathcal{Y} \equiv \{y : g_{l_1}(y,z) \le 0,\ h_{l_2}(y,z) = 0,\ l_1 \in \{1,\dots,N_I\},\ l_2 \in \{1,\dots,N_E\}\}$ encapsulates all constraints involving y and the set $\mathcal{Z}$ encapsulates all constraints involving z, with $g_{l_1}(y,z)$ representing the inequality constraints and $h_{l_2}(y,z)$ the equality constraints. Both $\mathcal{Y}$ and $\mathcal{Z}$ are compact sets, all the functions $g_{l_1}(y,z)$ and $h_{l_2}(y,z)$ are differentiable and $\mathcal{Z}$ is a set dependent on y. The function $f(y,z)$ is assumed to be differentiable in y and z and convex as a function of the outer problem decision variables y. No assumptions relative to the convexity properties of $f(y,z)$ with respect to inner level decision variables are considered. This formulation has an outer problem (i.e. the min problem) and an inner problem (i.e. the max problem) and we solve the minimax program in two phases, Phase 1 and Phase 2 iteratively, until a convergence condition is satisfied.

At the nth iteration, there exists $\tau^n \in \mathbb{R} : \max_{z\in\mathcal{Z}} f(y,z) \le \tau^n$ if and only if $f(y,z) \le \tau^n,\ \forall z \in \mathcal{Z}$. Accordingly, we may formulate an equivalent semi-infinite program using a relaxation procedure to find the solution of the minimax problem as follows [38]:

$$\begin{aligned} \min_{y\in\mathcal{Y},\tau^n\in[\tau^L,\tau^U]} \quad & \tau^n \\ \text{s.t.} \quad & f(y,z) \le \tau^n \\ & g_{l_1}(y,z) \le 0, \quad l_1 \in \{1,\dots,N_I\} \\ & h_{l_2}(y,z) = 0, \quad l_2 \in \{1,\dots,N_E\} \\ & y \in \mathcal{Y}, \qquad z \in \mathcal{Z}. \end{aligned} \tag{6}$$

Here τ^L and τ^U are finite values bounding τ^n and since they are unknown, we may consider τ^L equal to a finite large negative value and τ^U equal to a finite large positive constant. The problem (6) involves a finite number of variables and an infinite number of constraints as a result of the dependency of $\mathcal{Z}(y)$.

The reformulation of problem (6) to an equivalent problem with a finite number of constraints requires that we replace $\mathcal{Z}$ with a discrete set. At the first iteration, we denote this set by $\mathcal{Z}^1 = \{z_0\}$ where z_0 is feasible solution of the inner program prescribed in Section 3. At the nth iteration, this set is $\mathcal{Z}^n$ and has n elements originating in previous iterations. At the next

iteration, this set becomes $\mathcal{Z}^{n+1}$ with $n+1$ elements formed by augmenting $\mathcal{Z}^n$ with a solution for the Phase 2 problem (9), denoted by z_n, following the rule:

$$\mathcal{Z}^{n+1} = \mathcal{Z}^n \cup \{z_n\}. \tag{7}$$

The Phase 1 program, denoted as $\mathcal{P}_{1,A}$, to solve is therefore:

$$\begin{aligned} \min_{y\in\mathcal{Y},\tau^n\in[\tau^L,\tau^U]} \quad & \tau^n \\ \text{s.t.} \quad & f(y,z) \le \tau^n \\ & g_{l_1}(y,z) \le 0, \quad l_1 \in \{1,\dots,N_I\} \\ & h_{l_2}(y,z) = 0, \quad l_2 \in \{1,\dots,N_E\} \\ & y \in \mathcal{Y}, \qquad z \in \mathcal{Z}^n. \end{aligned} \tag{8}$$

The problem $\mathcal{P}_{1,A}$ solves the outer level of (5) and each solution y minimizes the objective function for a set of discrete points $z \in \mathcal{Z}^n$. Afterwards, we fix y and solve the following program corresponding to the inner program of the problem (5), denoted by $\mathcal{P}_{1,B}$:

$$\begin{aligned} \zeta^n = \max_{z\in\mathcal{Z}} \quad & f(y,z) \\ \text{s.t.} \quad & g_{l_1}(y,z) \le 0, \quad l_1 \in \{1,\dots,N_I\} \\ & h_{l_2}(y,z) = 0, \quad l_2 \in \{1,\dots,N_E\} \\ & y \text{ fixed}, \quad z \in \mathcal{Z}. \end{aligned} \tag{9}$$

The solution of (9), z_n, with the subscript n representing the iteration counter, are stationary/Karush–Kuhn–Tucker (KKT) points of the inner problem and are appended to the set $\mathcal{Z}^n$ employing (7). Then we repeat the cycle and keep iterating between the outer problem corresponding to Phase 1 and the inner problem, corresponding to Phase 2, until convergence occurs. The discrete set $\mathcal{Z}^n$ contains the accumulating successive KKT points of the inner program that produce successively tighter relaxations of (8).

We observe that the number of constraints $f(y,z) \le \tau^n$ for the problem (8) increases by one per iteration as a result of the increase in the number of elements forming the set of discrete points $\mathcal{Z}^n$. Solving problem $\mathcal{P}_{1,A}$ provides a global lower bound to the minimax problem and solving problem $\mathcal{P}_{1,B}$ produces a local upper bound (obtained for a particular point y). Therefore, $\tau^n \ge \tau^{n-1}$ but no conclusion can be drawn for ζ^n in successive iterations, ζ^n being the optimum of problem $\mathcal{P}_{1,B}$. The convergence test checks the condition $|(\zeta^n - \tau^n)/\tau^n| \le \epsilon_1$, where ϵ_1 is a positive small constant provided by the user to assess the relative error. When the condition is satisfied the solution has been found. Theoretical results prove that the procedure described above converges in a finite number of iterations for ϵ_1-optimal solutions [8,27].

Here we assume all constraints in the problem (5) are decoupled. This assumption is reasonable since in the optimal design problem the constraints are functions of the regressors or of the parameters and not on both types of variables. Strategies for this more complicated case are provided by Polak [32, Ch. 3], Mitsos et al. [31] and Tsoukalas et al. [40].

3. Algorithms

In this section we describe the SIP algorithms for finding T-optimal designs. This approach assumes that we want to find a k-point T-optimal design where k is pre-specified. In our algorithm k is initialized to the number of parameters in the problem plus one. If at convergence, the T-optimal design found by SIP is not optimal according to the equivalence theorem, we will repeat the search among designs with $k+1$ points. Our experience is that usually a couple of such iterations will produce the SIP-generated T-optimal design that is optimal among all designs on the design space.

3.1. SIP formulation for T-optimal designs

In this section, we apply the general techniques in Section 2 to solve Problem (3) by finding the optimal discrimination design supported at k points. Accordingly we include a superscript k in the variables in the mathematical codes below. At the nth iteration of the SIP-based procedure, the generated design $\xi^{k,n}$ has $x_i^{k,n}$ as its ith support point with corresponding weight $w_i^{k,n}$, $i = 1,\dots,k$, and they are found by solving the preceding optimization problem. This formulation corresponds to a direct application of the Phase 1 problem (8):

$$\begin{aligned} \min_{\xi^{k,n}\in\Xi,\tau^{k,n}\in[\tau^L,\tau^U]} \quad & \tau^{k,n} \\ \text{s.t.} \quad & -\sum_{i=1}^{k}[\eta_t(x_i^{k,n}) - \eta_2(x_i^{k,n},\theta_2^k)]^2\, w_i^{k,n} \le \tau^{k,n} \\ & \sum_{i=1}^{k} w_i^{k,n} = 1 \\ & \theta_2^k \in \Theta_2^{k,n}. \end{aligned} \tag{10}$$

For a fixed design $\xi^{k,n}$, the program $\mathcal{P}_{1,B}$ for Phase 2 is:

$$\zeta^{k,n} = \max_{\theta_2 \in \Theta_2} - \sum_{i=1}^{k} [\eta_t(x_i^{k,n}) - \eta_2(x_i^{k,n}, \theta_2)]^2 \, w_i^{k,n}. \tag{11}$$

Let $\theta_2^{k,n}$ solve problem (11) and let $\Theta_2^{k,n+1} = \Theta_2^{k,n} \cup \{\theta_2^{k,n}\}$ be the set of stationary/KKT points. The iteration between problems ((10)–(11)) allows convergence to the solution of the minimax problem and determines the optimal design, ξ_T^k. When convergence occurs $\Delta(\xi_T^k) = \zeta^{k,n}$ (at ϵ_1-error optimality) and the parametric combination at which the design is least efficient is $\theta_2^{k,o} = \theta_2^{k,n}$, the argument of problem (11), corresponding to the least favorable parametric configuration in Θ_2 found at the nth iteration.

The algorithm proposed requires an initial feasible instance of θ_2 employed to form the discrete set $\Theta_2^{k,1}$. We generate a feasible initial solution by solving the above program, denoted by $\mathcal{P}_0^k$. The formalization assumes that a feasible solution can be found when we maximize the sum of squares for the lack of fit. The design obtained is T-optimal in distinguishing models for the most efficient combination of parameters for model discrimination.

$$\begin{aligned} \min_{\xi^{k,0} \in \Xi, \theta_2 \in \Theta_2} \quad & - \sum_{i=1}^{k} [\eta_t(x_i^{k,0}) - \eta_2(x_i^{k,0}, \theta_2)]^2 \, w_i^{k,0} \\ \text{s.t.} \quad & \sum_{i=1}^{k} w_i^{k,0} = 1. \end{aligned} \tag{12}$$

Here $\theta_2^{k,0}$ is the argument of the problem (12) and $\Theta_2^{k,1} = \{\theta_2^{k,0}\}$.

The above search for the T-optimal design using the SIP-based method is summarized in Algorithm 1 below. To initiate the algorithm, the following input parameters are required: the tolerance level for testing convergence, the number of support points k, bounds for $\tau^{k,n}$ which may be large negative and positive constants, bounds for $\mathcal{X}$ and bounds for the parameters included in the "test model". The algorithm then iterates to find ξ_T^k, the optimal discrimination design supported at k points, the least favorable parametric configuration vector of parameters, $\theta_2^{k,o}$ and the minimal value of the T-optimality criterion, $\Delta(\xi_T^k)$.

Algorithm 1 Algorithm to find a T-optimal design for k support points.

```
procedure FindDesign(ε_1, k, τ^L, τ^U, θ^L, θ^U, ξ_T^k, θ_2^{k,o}, Δ(ξ_T^k))
    Solve P_0^k                                          ▷ Find an initial feasible solution
    n ← 1
    θ_2^{k,0} ← arg(P_0^k),  Θ_2^{k,1} ← {θ_2^{k,0}}
    while |(ζ^{k,n} − τ^{k,n})/τ^{k,n}| ≤ ε_1 do
        Solve P_{1,A}^k                                  ▷ Solve the Phase 1 program
        τ^{k,n} ← arg(P_{1,A}^k)
        Fix ξ^{k,n}
        Solve P_{1,B}^k                                  ▷ Solve the Phase 2 program
        ζ^{k,n} ← arg(P_{1,B}^k),  θ_2^{k,n} ← arg(P_{1,B}^k)
        Θ_2^{k,n+1} ← Θ_2^{k,n} ∪ {θ_2^{k,n}}            ▷ Update the set of KKT points
        n ← n + 1                                        ▷ Update the iteration counter
    end while
    ξ_T^k ← ξ^{k,n},  θ_2^{k,o} ← θ_2^{k,n},  Δ(ξ_T^k) ← ζ^{k,n}
end procedure
```

3.2. Checking the T-optimality of the k-support point design

The algorithm to find k-point T-optimal designs was discussed in Section 3.1. The task now is to have an automated process to check whether the k-point design obtained by solving the minimax program (3) is globally optimal among all designs on the design space. This task is addressed by iteratively solving a new nonlinear program $\mathcal{P}_2^k$. If the design space is one-dimensional, a univariate graphical plot will suffice to ascertain whether the equivalence theorem is satisfied. When the design space is multi-dimensional, with each dimension representing a different regressor, this task is harder because it is not easy to visually ascertain whether a high-dimensional plot satisfies the conditions of the equivalence theorem. Our procedure conducts a numerical search over $\mathcal{X}$ automatically and so facilitates the decision making process whether the T-optimal design is found.

In practice, we start with k equal to $p_2 + 1$ and increase it by one sequentially until the SIP-generated T-optimal design satisfies the conditions in the equivalence theorem. Problem $\mathcal{P}_2^k$ below finds the maximum value of the sum of squares for

the lack of fit of the models in $\mathcal{X}$ for the k-point design ξ_T^k, and the optimal singleton parametric combination $\theta_2^{k,o}$:

$$\alpha^k = \max_{x \in \mathcal{X}} \; [\eta_t(x) - \eta_2(x, \theta_2^{k,o})]^2 + \Delta(\xi_T^k). \tag{13}$$

The equivalence theorem is validated if and only if $\alpha^k \leq \epsilon_2$, ϵ_2 being a small positive constant supplied by the user. Algorithm 2 presents the complete procedure, that finds the T-optimal design and automates the search for the optimal number of support points.

Algorithm 2 Algorithm to find a T-optimal design

```
procedure OptimalTDesign(ε2, p2, ε1, τ^L, τ^U, θ^L, θ^U, ξ_T^k, θ_2^{k,o}, Δ(ξ_T^k))
    k ← p2 + 1                                                    ▷ Initialization of k
    Solve FindDesign(ε1, k, τ^L, τ^U, θ^L, θ^U, ξ_T^k, θ_2^{k,o}, Δ(ξ_T^k))
    Solve P_2^k                                                   ▷ Check T-optimality for k = p2 + 1
    α^k ← arg(P_2^k)
    while α^k ≥ ε2 do
        k ← k + 1
        Solve FindDesign(ε1, k, τ^L, τ^U, θ^L, θ^U, ξ_T^k, θ_2^{k,o}, Δ(ξ_T^k))
        Solve P_2^k                                               ▷ Solve the NLP problem to check T-optimality
        α^k ← arg(P_2^k)
    end while
end procedure
```

Another indication that k is well chosen is that when the number of support points is greater than that for the optimal design there are one or more points with null weights. However, this indication can only be found if we solve the problem for such a k.

To implement the Semi-Infinite Programming based algorithm we require global NLP solvers to guarantee the global optimality of the solutions of problems ((10)–(13)). Here one uses the solver OQNLP from the GAMS package, which is formed by a general modeling language and a battery of mathematical programming solvers for several types of problems [19]. OQNLP is a multistart heuristic algorithm designed to find global optima of nonlinear programs. The algorithm calls a NLP solver from multiple starting points, saves all the feasible solutions found and picks the best as the optimum of the problem [43]. The starting points are computed with a random sampling driver that uses normal independent probability distribution functions for each decision variable. OQNLP does not guarantee that the final solution is a global optimum but it has been successfully tested on a large set of problems. To build the initial sampling points the variables need to be bounded, which we have since the design space and the region of plausible values are all compact by assumption.

The NLP solver called by OQNLP is CONOPT, which in turn uses the Generalized Reduced Gradient (GRG) algorithm [13]. The maximum number of starting points allowed is set to 3000 and the procedure terminates when 100 consecutive NLP solver calls result in a only fractional improvement of less than 10^{-4} in the criterion value. In all our problems the absolute and relative tolerances of the solvers were all respectively set equal to 10^{-8} and 10^{-7}. The tolerances ϵ_1 and ϵ_2 required by Algorithms 1 and 2 were both set equal to 10^{-5}. In all cases the bounds for $\tau^{k,n}$ were set to -10^4 and $+10^4$.

4. Results

We now apply the SIP-based algorithm to find the optimal discrimination design for discriminating among models. The selected applications are from the literature where we can compare results and verify the usefulness of the SIP-approach for finding T-optimal designs.

In Section 4.1 we first consider a benchmark problem to test our proposed SIP-based approach. Accordingly we provide more details, including analysis of the convergence properties of the SIP algorithm. In Section 4.2, we consider 8 optimal discrimination design problems; the first 4 concern univariate polynomial models all defined on a common design space $\mathcal{X} = [-1, +1]$ and the next 3 concern nonlinear models with one or two regressors. In the last problem, case (8) has 3 models for discrimination by design and so the optimization setup is more complicated.

Table 1 displays the T-optimal designs for cases (1)–(7) and Table 2 displays the T-optimal design for case (8). For each case, we list the optimal discrimination design, the least favorable parametric configuration vector of parameters, $\theta_2^{k,o}$, that gives the largest lack of fit sum of squares, the optimum value for the design criterion and the CPU time in seconds (secs) required to solve the optimization problem. All results were obtained using a 12 GB RAM Intel Core i7 machine running 64 bits Windows 7 operating system with 2.80 GHz. The solvers used to handle all NLP programs are from GAMS package, version 23.5. Throughout, we let η_t denote the mean function of the "true model" and let η_1 and η_2 denote the mean functions of the test or alternative models. We assume the $p_2 \times 1$ vector of unknown parameters θ_2 in η_2 is from a user-selected set Θ_2 which we assume is a Cartesian product of the parameter space for each component.

Table 1
T-optimal designs for Cases (1)–(7) and their characteristics.

Case	k	T-optimal design	$\theta^{k,o}$	optimum value	CPU (s)
1	2	$\begin{pmatrix} -0.5000 & 1.0000 \\ 0.5003 & 0.4997 \end{pmatrix}$	$\{1.8742\}$	-1.265625	7.526
2	3	$\begin{pmatrix} -1.0000 & 0.0000 & 1.0000 \\ 0.2500 & 0.5000 & 0.2500 \end{pmatrix}$	$\{1.5000, 1.0000\}$	-0.250000	207.324
3	5	$\begin{pmatrix} -1.0000 & -0.5432 & 0.1803 & 0.7731 & 1.0000 \\ 0.0555 & 0.1594 & 0.2580 & 0.3408 & 0.1864 \end{pmatrix}$	$\{0.8936, 0.5416, 1.9550, 2.4591\}$	-0.022747	361.575
4	5	$\begin{pmatrix} -1.0000 & -0.8090 & -0.3090 & 0.3090 & 0.8090 \\ 0.2004 & 0.3586 & 0.2642 & 0.1376 & 0.0391 \end{pmatrix}$	$\{0.9997, 0.6882, 0.9989, 2.2480\}$	-0.003906	335.324
5	3	$\begin{pmatrix} 0.9808 & 3.6603 & 5.0000 \\ 0.4548 & 0.3665 & 0.1787 \end{pmatrix}$	$\{2.0837, 2.8740\}$	-0.231310	252.643
6	4	$\begin{pmatrix} 3.0580 & 5.4390 & 30.0000 & 30.0000 \\ 0.0000 & 11.6506 & 22.7304 & 0.0000 \\ 0.2498 & 0.4415 & 0.2496 & 0.0590 \end{pmatrix}$	$\{11.8718, 7.6432, 12.7019\}$	-0.533095	187.869
7	4	$\begin{pmatrix} 1.8152 & 4.0914 & 30.0000 & 30.0000 \\ 0.0000 & 4.1462 & 0.0000 & 10.1666 \\ 0.0461 & 0.5498 & 0.0666 & 0.3375 \end{pmatrix}$	$\{8.3470, 2.1013, 0.6554\}$	-0.867212	222.576

k — number of support points.

Table 2
T-optimal design for Case (8) with its characteristics.

Case	k	T-optimal design	$\theta^{k,o}$	Optimum value	CPU (s)
8	5	$\begin{pmatrix} -1.0000 & -0.7364 & -0.0989 & 0.6247 & 1.0000 \\ 0.2022 & 0.3306 & 0.2263 & 0.1664 & 0.0744 \end{pmatrix}$	$\left\{ \begin{matrix} 1.0284, 0.5634, -1.9201, \\ -0.8252, 0.5930, 1.8928, -0.1876 \end{matrix} \right\}$	0.003195	804.695

k—number of support points.

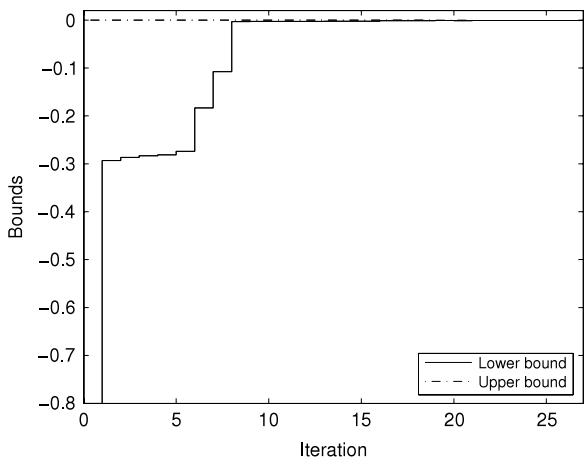

Fig. 1. Evolution of the upper and lower bounds of the 4-point T-optimal design in case (1).

4.1. Test case

We first consider the problem discussed in Atkinson and Fedorov [5] and Atkinson et al. [4] where

$$\eta_t = 4.5 - 1.5\ \exp(x_1) - 2.0\ \exp(-x_1) \tag{14}$$

$$\eta_2 = \theta_1 + \theta_2\, x_1 + \theta_3\, x_1^2 \tag{15}$$

$$\mathcal{X} = [-1, +1], \qquad \Theta = [-10, 4] \times [-10, 4] \times [-10, 4]$$

and compare their T-optimal design with the one from our SIP based algorithm. The alternative model is clearly linear with respect to the parameters $\theta_i,\ i = \{1, 2, 3\}$ and nonlinear in the regressors. For this application, we set the initial number of support points to be 4 and find the optimal discrimination design among all 4-point designs. The SIP requires 28 iterations; Fig. 1 presents the evolution of the lower and upper bounds. The absolute error, defined in Algorithm 1 as $|\zeta^{k,n} - \tau^{k,n}|$,

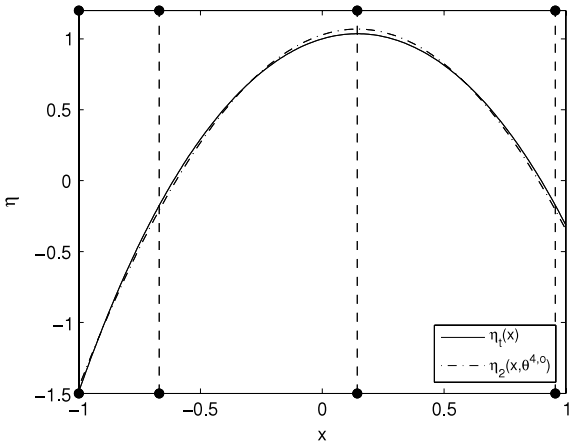

Fig. 2. Mean responses from the "True model" and the "test model" in case (1) with the vertical lines representing the support points of the T-optimal design.

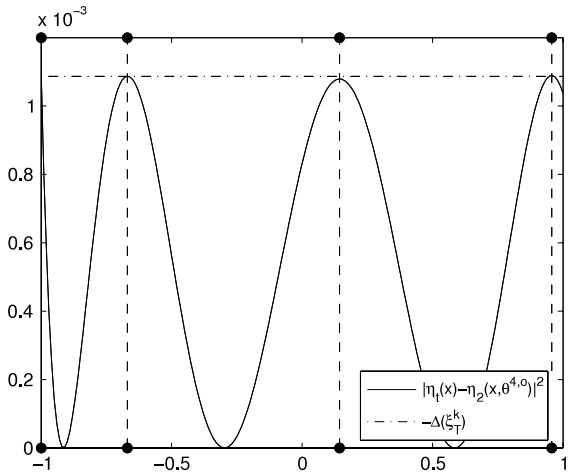

Fig. 3. Sensitivity function of the T-optimal design with the vertical lines representing the support points in case (1).

corresponds to the absolute difference between the upper and the lower bounds at the nth iteration. A value of magnitude 10^{-3} is reached in 9 iterations; the subsequent iterations are necessary to converge to a relative error, defined as $|(\zeta^{k,n} - \tau^{k,n})/\tau^{k,n}|$, of 10^{-5}. It is noteworthy that the convergence condition used in Algorithm 1, which relies on the relative error, is responsible for an important fraction of the iterations. This trend cannot be generalized, but depends on the specific problem. The analysis of the behavior of the upper bound reveals that its value is close to 10^{-6} during the first 20 iterations, due to the ability of the inner program to determine a parametric combination that nearly produces a null lack of fit sum of squares for the design produced in Phase 1. Thereafter, the value of the upper bound decreases to the optimum value of -0.003906 (see Case 4 in Table 1). The behavior of the lower bound, always increasing, is in agreement with the theoretical result presented in Section 2.1. Another aspect to mention is that the algorithm for the T-optimal design ξ_T^4 converges in 20 iterations but the convergence of $\theta^{4,o}$ requires 8 additional iterations. This trend cannot be generalized to all problems, but in this case the least favorable parametric configuration is more difficult to find than the optimal design.

The optimal design found is:

$$\begin{pmatrix} -1.0000 & -0.6693 & 0.1438 & 0.9570 \\ 0.2536 & 0.4250 & 0.2497 & 0.0718 \end{pmatrix}$$

for $\theta^{4,o} = \{1.0288, 0.5550, -1.9292\}$. The optimum of the semi-infinite program is -0.001087, requiring a CPU time of 613.3970 s. The design obtained is in strong agreement with the results of Atkinson et al. [4], and satisfies the conditions for the equivalence theorem. Therefore the T-optimal design found, based on 4 support points, is globally optimal. Fig. 2 illustrates the lack of fit of the models for the best fitting combination of parameters, and the support points. Fig. 3 illustrates graphically the validation of the equivalence theorem.

4.2. Additional test cases

We now consider 8 more cases for further testing the effectiveness of our proposed algorithms. The first 4 concern only univariate polynomial models all defined on a common design space $\mathcal{X} \in [-1, +1]$; the next 3 cases are nonlinear models with one or two regressors. Unlike the first 7 cases where only 2 models are involved, the last example, case (8), is to discriminate between 3 linear models, two of them considered as "test models".

Case (1): This example is taken from Atkinson and Fedorov [5] and Atkinson et al. [4], where interest is to find the T-optimal design to discriminate between a "true" quadratic model and a constant model. We have

$$\eta_t = 1 + 1\,x_1 + 1\,x_1^2 \tag{16}$$
$$\eta_2 = \theta_1 \tag{17}$$
$$\mathcal{X} \in [-1, +1], \qquad \Theta \in [0, 4].$$

Case (2): We now wish to find the T-optimal design to discriminate between the quadratic polynomial from a linear model on the same design space.

$$\eta_t = 1 + 1\,x_1 + 1\,x_1^2 \tag{18}$$
$$\eta_2 = \theta_1 + \theta_2\,x_1 \tag{19}$$
$$\mathcal{X} = [-1, +1], \qquad \Theta \in [0, 4] \times [0, 4].$$

We note that with nested linear models that differ by a single parameter, the optimal discrimination design problem is the same as that for finding the best design to estimate that single parameter, i.e. we want to find the D_s-optimal design.

Case (3): Dette et al. [11] considered finding T-optimal designs for discriminating among polynomials. The true model is a polynomial of degree n and the test models are polynomials of degree $(n-2)$. A particular example of such a situation is when $n = 5$ with

$$\eta_t = 1 + 1\,x_1 + 1\,x_1^2 + 1\,x_1^3 + 1\,x_1^4 + 1\,x_1^5 \tag{20}$$
$$\eta_2 = \theta_1 + \theta_2\,x_1 + \theta_3\,x_1^2 + \theta_4\,x_1^3 \tag{21}$$
$$\mathcal{X} = [-1, +1], \qquad \Theta = [0, 4] \times [0, 4] \times [0, 4] \times [0, 4].$$

Case (4): A special instance of Case 3 was considered by Atkinson [1] and Dette et al. [11] where the coefficient of the term x_1^{n-1} in the "true model" is 0. We have

$$\eta_t = 1 + 1\,x_1 + 1\,x_1^2 + 1\,x_1^3 + 0\,x_1^4 + 1\,x_1^5 \tag{22}$$
$$\eta_2 = \theta_1 + \theta_2\,x_1 + \theta_3\,x_1^2 + \theta_4\,x_1^3 \tag{23}$$
$$\mathcal{X} = [-1, +1], \qquad \Theta = [0, 4] \times [0, 4] \times [0, 4] \times [0, 4].$$

The interesting feature of this problem is that the optimal design is not symmetric and not unique. The 5-point optimal design found by SIP is shown in Table 1. It has support points that range from -1 to 0.8090 and so is not supported at one of the extreme ends of the design space. The symmetrized design with a support range from -0.8090 to 1 is also optimal, as is any convex linear combination of the two designs. The sensitivity plot based on the equivalence theorem in Fig. 4 supports this claimed behavior, since any 5 of the 6 points with the same peaks can be selected as the equivalence theorem is satisfied not only at the points of the design, but also at $x_1 = 1$. This is therefore a potential support point of the optimal design. The convex linear combination of the two designs can be chosen to maximize some secondary design criterion.

Case (5): This example is taken from López-Fidalgo et al. [30] who were interested in finding the T-optimal design to discriminate between a Michaelis–Menten kinetic rate model augmented with a linear term and the simple Michaelis–Menten model.

$$\eta_t = \frac{x_1}{1.0 + x_1} + 0.1\,x_1 \tag{24}$$
$$\eta_2 = \frac{\theta_1\,x_1}{\theta_2 + x_1} \tag{25}$$
$$\mathcal{X} = [10^{-5}, 5.0], \qquad \Theta = [10^{-3}, 30] \times [10^{-3}, 10].$$

The test model is linear in θ_1 but nonlinear in θ_2.

Case (6): This optimal discrimination design problem was discussed and addressed by Atkinson [2]. Here the interest is to find an optimal design to discriminate between two models for studying enzyme activity in the presence of chemical

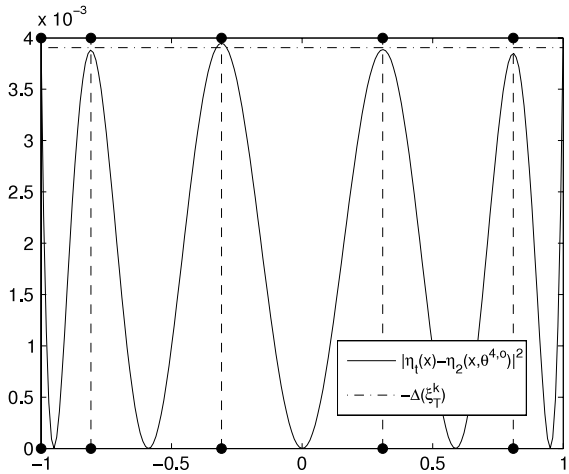

Fig. 4. Sensitivity function of the T-optimal design with the vertical lines representing the support points in case (4).

inhibitors. Both models are direct extensions of the Michaelis–Menten model discussed in Case (5). The null or "true model" assumes there is competitive inhibition and the "test model" assumes there is non-competitive inhibition.

$$\eta_t = \frac{25.8\, x_1}{4.36\,(2.58 + x_2) + 2.58\, x_1} \tag{26}$$

$$\eta_2 = \frac{\theta_1\, \theta_3\, x_1}{(\theta_2 + x_1)\,(\theta_3 + x_2)} \tag{27}$$

$$\mathcal{X} = [10^{-5}, 30.0] \times [10^{-5}, 40.0], \qquad \Theta = [10^{-3}, 100] \times [10^{-3}, 18] \times [10^{-3}, 18].$$

The test model is nonlinear in the parameters θ_2 and θ_3, but not in θ_1.

Case (7): This problem was also addressed by Atkinson [2] and corresponds to the opposite of Case (6). It is assumed that the inhibition in the "true model" is non-competitive and the inhibition in the "test model" is competitive.

$$\eta_t = \frac{51.6\, x_1}{(4.36 + x_1)\,(5.16 + x_2)} \tag{28}$$

$$\eta_2 = \frac{\theta_1\, \theta_3\, x_1}{\theta_2\,(\theta_3 + x_2) + \theta_3\, x_1} \tag{29}$$

$$\mathcal{X} = [10^{-5}, 30.0] \times [10^{-5}, 40.0], \qquad \Theta = [10^{-3}, 100] \times [10^{-3}, 18] \times [10^{-3}, 18].$$

Table 1 presents the T-optimal designs found for the above cases. The results obtained agree with the literature, and in many cases the two sets of results are identical. This shows that the SIP-based strategy works for these examples and is flexible enough to handle polynomials and non-linear models in a model discrimination problem. Fig. 5 is the 2-dimensional plot of the sensitivity function for Case (6), and confirms that the 4-point design found by SIP is optimal for discrimination among all designs on the design space.

Case (8): This case tests our algorithm in a higher-dimensional problem where we want to find the T-optimal design for discriminating among three rival models considered in Atkinson and Fedorov [6]. We have

$$\eta_t = 4.5 - 1.5\ \exp(x) - 2\ \exp(-x) \tag{30}$$

$$\eta_1 = \theta_1 + \theta_2\, x + \theta_3\, x^2 \tag{31}$$

$$\eta_2 = \theta_4 + \theta_5\ \sin(0.5\,\pi\, x) + \theta_6\ \cos(0.5\,\pi\, x) + \theta_7\ \sin(\pi\, x) \tag{32}$$

$$\mathcal{X} = [-1, 1], \qquad \Theta = [-10, 4]^7.$$

The two alternative models have a total of 7 unknown parameters with $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7]^{\mathrm{T}}$ and there is only one regressor $x \in X$. Atkinson and Fedorov [6] formulated the problem as a minmaxmin program where the outer program is to find optimal weights for the lack of fit sum of squares associated with each alternative model. Here, we view the problem differently and interpret the weights as known probabilities, with each representing the probability that the alternative model is the true model. To fix ideas, we assume that the weights for this example are both equal to 0.5, but other values

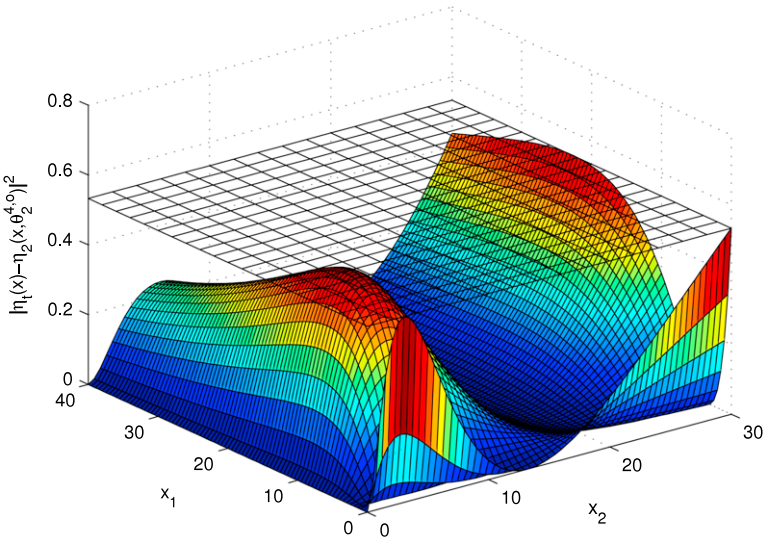

Fig. 5. Sensitivity function for the T-optimal design in Case (6).

can also be used in the same way. The minimax program corresponding to the T-optimal design for this case is:

$$\begin{aligned} \min_{\xi\in\Xi}\max_{\theta\in\Theta} \quad & -\sum_{i=1}^{k}\left(0.5\ [\eta_t(x_i)-\eta_1(x_i,\theta_1,\theta_2,\theta_3)]^2+0.5\ [\eta_t(x_i)-\eta_2(x_i,\theta_4,\theta_5,\theta_6,\theta_7)]^2\right)\ w_i \\ \text{s.t} \quad & \sum_{i=1}^{k} w_i = 1. \end{aligned} \tag{33}$$

The results in Table 2 are in close agreement with those from Atkinson and Fedorov [6] verifying once again that our SIP-based algorithm is able to solve more complex discrimination design problems.

There are two observations from the above examples. First, the CPU times required by our algorithm to find the optimal discrimination design can differ substantially from problem to problem. A main reason is that each problem requires differing amounts of time to solve the program $\mathcal{P}_{1,A}^{k}$ in each iteration. This problem is computationally intensive because it finds the globally optimal solutions by validating the conditions listed in Section 3 and in many cases, the algorithm requires between 3000 and 5000 calls of the NLP solver. Further this solution depends on the assumed dimension of the cartesian regions $\mathcal{X}$ and Θ_2, and on the efficiency of the sampling driver used to generate initial solutions for the NLPs solved.

Second, the T-optimal designs for the first seven cases have $k = p_2 + 1$ support points. This number is that obtained on theoretical grounds by Dette and Titoff [12] for T-optimal designs when all models of interest are polynomials. Case (8) in Table 2 has 7 parameters but leads to an optimal discrimination design with 5 support points. This number also validates the theoretical result if we adopt the rule that when there are 2 or more test models, p_2 should refer to the number of parameters in the mean function of the model with the most parameters among the alternative models. In this example, the model with more parameters among the two models is $\eta_2(x_i, \theta_4, \theta_5, \theta_6, \theta_7)$ with 4 parameters, resulting in the T-optimal design having 5 points. The implication from our numerical results suggest that the theoretical result from Dette and Titoff [12] on the number of support points in a T-optimal design may apply to a larger class of models than polynomials.

5. Conclusions

Traditional statistical methods of finding optimal designs, such as Wynn–Fedorov algorithms, can run into convergence problems when the optimization problem becomes complex as in minimax or maximin optimization. The goal of this paper was to explore alternative, and potentially more flexible and powerful, methods to find optimal designs for discriminating between two or more models. We first formulated the minimax optimization problem as one belonging to a class of non-smooth mathematical programs before applying Semi-Infinite Programming tools for numerical solution of the minimax optimization problem using an exchange based algorithm. To do this, the continuous domain of the parameters is discretized so that we can reformulate the original SIP problem with an infinite number of constraints as a relaxation program with a finite number of constraints.

The discretization points that populate the discrete set of parametric combinations are solutions of the inner program; the solution of the SIP is obtained through convergence of the solutions of the outer and the inner programs. Our approach includes an initial step to find a feasible initial parametric combination to start the iterative procedure. We automate the search for the global T-optimal design and confirm the optimality of the SIP-generated design using an equivalence theorem.

If the SIP-generated design is not T-optimal, the next search will be among designs with one more support point. Our experience is that usually just a couple of such increases will suffice before the process finds the T-optimal design. All programs were solved using a stochastic global NLP solver.

We tested our SIP approach with a number of discrimination design problems for linear and nonlinear models. In all cases, our results from the SIP approach were in agreement with the results reported in the literature. The main advantages of our proposed approach are its flexibility and the relatively low computational resources required to find the T-optimal designs when compared with traditional methods. We are encouraged by SIP methodology and plan to test its ability to find optimal discrimination designs more broadly in more complicated problems. For example, a good test case is to find robust T-optimal designs where we do not require one of the two models to be fully known [44]. Mathematical programming tools such as SIP and Semi-definite programming have already been commonly and effectively used by engineers for decades but not by statisticians. We hope that our work here will encourage statisticians to explore such programming tools in their research work.

Acknowledgments

Work on this paper began at the Isaac Newton Institute for Mathematical Sciences in Cambridge, England, during the 2011 programme on the Design and Analysis of Experiments. Professors Atkinson and Wong thank the institute for support during their visits. The research of Wong reported in this paper was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content of the paper is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

References

[1] A.C. Atkinson, The non-uniqueness of some designs for discriminating between two polynomial models in one variable, in: A. Giovagnoli, A.C. Atkinson, B. Torsney (Eds.), moDA 9 - Advances in Model-Oriented and Analysis, Proceedings of the 9th International Workshop in Model-Oriented Design and Analysis, Springer-Verlag, Heidelberg, 2010, pp. 9–16.
[2] A.C. Atkinson, Optimum experimental designs for choosing between competitive and non competitive models for enzyme inhibition, Comm. Statist. Theory Methods 41 (2012) 2283–2296.
[3] A.C. Atkinson, D.R. Cox, Planning experiments for discrimination between models (with discussion), J. R. Stat. Soc. Ser. B 36 (1974) 321–348.
[4] A.C. Atkinson, A.N. Donev, R.D. Tobias, Optimum Experimental Designs, with SAS, Oxford University Press, New York, 2007.
[5] A.C. Atkinson, V.V. Fedorov, The design of experiments for discriminating between two rival models, Biometrika 62 (1975) 57–70.
[6] A.C. Atkinson, V.V. Fedorov, Optimal design: experiments for discriminating between several models, Biometrika 62 (1975) 289–303.
[7] M.P.F. Berger, W.K. Wong, An Introduction to Optimal Designs for Social and Biomedical Research, in: Statistics in Practice, Wiley, New York, 2009.
[8] J.W. Blankenship, J.E. Falk, Infinitely constrained optimization problems, J. Optim. Theory Appl. 19 (1976) 261–281.
[9] H. Dette, Discrimination designs for polynomial regression on compact intervals, Ann. Statist. 22 (1994) 890–903.
[10] H. Dette, G. Haller, Optimal designs for the identification of the order of the Fourier regression, Ann. Statist. 26 (1998) 196–1521.
[11] H. Dette, V.B. Melas, P. Shpilev, T-optimal designs for discrimination between two polynomial models, Ann. Statist. 40 (2012) 188–205.
[12] H. Dette, S. Titoff, Optimal discriminating designs, Ann. Statist. 37 (2009) 2056–2081.
[13] A. Drud, CONOPT: a GRG code for large sparse dynamic nonlinear optimization problems, Math. Program. 31 (1985) 153–191.
[14] B.P.M. Duarte, W.K. Wong, A Semi-Infinite Programming based algorithm for finding minimax optimal designs for nonlinear models, Stat. Comput. 24 (2013) 1063–1080.
[15] V.V. Fedorov, Theory of Optimal Experiments, Academic Press, New York, 1972.
[16] V.V. Fedorov, P. Hackl, Model-oriented Design of Experiments, Springer, New York, 1997.
[17] V.V. Fedorov, S.L. Leonov, Optimal Design for Nonlinear Response Models, Chapman and Hall/ CRC Press, Boca Raton, 2014.
[18] V.V. Fedorov, M.B. Malyutov, Optimal designs in regression problems, Math. Oper.forsch. Stat. 3 (1972) 281–308.
[19] GAMS Development Corporation, GAMS: The Solver Manuals, GAMS Development Corporation, Washington, 2013.
[20] P. Goos, B. Jones, Optimal Design of Experiments: A Case Study Approach, Wiley, New York, 2011.
[21] P.R. Gribik, K.O. Kortanek, Equivalence theorems and cutting plane algorithms for a class of experimental design problems, SIAM J. Appl. Math. 32 (1977) 232–259.
[22] R. Hettich, A. Kaplan, R. Tichatschke, Semi-Infinite Programming: numerical methods, in: Encyclopedia of Optimization, Vol. 5, Kluwer, Amsterdam, 2001, pp. 112–117.
[23] R. Hettich, K.O. Kortanek, Semi-infinite programming: theory, methods and applications, SIAM Rev. 35 (1993) 380–429.
[24] J.C. Kiefer, Optimum experimental designs, J. R. Stat. Soc. Ser. B 21 (1959) 272–319.
[25] J. Kiefer, General equivalence theory for optimum design (approximate theory), Ann. Statist. 2 (1974) 849–879.
[26] J. Kiefer, J. Wolfowitz, The equivalence of two extremum problem, Canad. J. Math. 12 (1960) 363–366.
[27] B. Kuczewski, Computational Aspects of Discrimination between Models of Dynamic Systems (Ph.D. thesis), University of Zielona Góra, Zielona Góra, Poland, 2005.
[28] M. López, G. Still, Semi-Infinite Programming, European J. Oper. Res. 180 (2007) 491–518.
[29] J. López-Fidalgo, C. Trandafir, C. Tommasi, An optimal experimental design criterion for discriminating between non-normal models, J. R. Stat. Soc. Ser. B 69 (2007) 231–242.
[30] J. López-Fidalgo, C. Tommasi, C. Trandafir, Optimal designs for discriminating between some extensions of the Michaelis–Menten model, J. Statist. Plann. Inference 138 (2008) 3797–3804.
[31] A. Mitsos, P. Lemonidis, P.I. Barton, Global solution of bilevel programs with a nonconvex inner program, J. Global Optim. 42 (2008) 475–513.
[32] E. Polak, Optimization: Algorithms and Consistent Approximations, Springer, New York, 1997.
[33] A.C. Ponce de Leon, A.C. Atkinson, Optimum experimental design for discriminating between two rival models in the presence of prior information, Biometrika 78 (1991) 601–608.
[34] F. Pukelsheim, Optimal Design of Experiments, SIAM, Philadelphia, 1993.
[35] F. Pukelsheim, S. Rieder, Efficient rounding of approximate designs, Biometrika 79 (1992) 763–770.
[36] R. Reemtsen, J.J. Rückman, Semi-Infinite Programming, Kluwer, Amsterdam, 1998.
[37] B. Rustem, M. Howe, Algorithms for Worst-case Design and Applications to Risk Management, Princeton University Press, Princeton, NJ, 2002.
[38] K. Shimizu, E. Aiyoshi, Necessary conditions for min-max problems and algorithms by a relaxation procedure, IEEE Trans. Automat. Control 25 (1980) 62–66.

[39] D. Skanda, D. Lebiedz, A robust optimization approach to experimental design for model discrimination of dynamical systems, Math. Program. A (2012) 1–29.
[40] A. Tsoukalas, B. Rustem, E.N. Pistikopoulos, A global optimization algorithm for generalized semi-infinite, continuous minimax with coupled constraints and bi-level problems, J. Global Optim. 24 (2009) 235–250.
[41] D. Uciński, Optimal Measurement Methods for Distributed Parameter System Identification, CRC Press, Boca Raton, 2005.
[42] D. Uciński, B. Bogacka, T-optimum designs for discriminating between two multiresponse dynamic models, J. R. Stat. Soc. Ser. B 67 (2005) 3–18.
[43] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, R. Martí, A multistart scatter search heuristic for smooth NLP and MINLP problems, in: Metaheuristic Optimization via Memory and Evolution, Springer, New York, 2005, pp. 25–51.
[44] P. Vajjah, S.B. Duffull, A generalisation of T-optimality for discriminating between competing models with an application to pharmacokinetic studies, Pharmaceutical Stat. 11 (2012) 503–510.
[45] T.H. Waterhouse, J.A. Eccleston, S.B. Duffull, Optimal design criteria for discrimination and estimation in nonlinear models, J. Biopharm. Statist. 19 (2009) 386–402.
[46] D.P. Wiens, Robust discrimination designs, J. R. Stat. Soc. Ser. B 71 (2009) 805–829.
[47] H.P. Wynn, The sequential generation of D-optimum experimental designs, Ann. Math. Stat. 41 (1970) 1655–1664.
[48] S. Žakovíc, B. Rustem, Semi-Infinite Programming and applications to minimax problems, Ann. Oper. Res. 124 (2003) 81–110.