

RADOSLAV HARMAN AND GUILLAUME SAGNOL

**Computing D-optimal experimental
designs for estimating treatment
contrasts under the presence of a
nuisance time trend¹**

¹appeared in: *Springers Proceedings in Mathematics & Statistics Volume 122*, pp 83-9, 2015, for the proceedings of the SMSA Workshop held in Wrocław, Poland, February 2015.

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: <http://www.zib.de>

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

Computing D-optimal experimental designs for estimating treatment contrasts under the presence of a nuisance time trend

Radoslav Harman¹ and Guillaume Sagnol²

¹ Comenius University, Mlynská dolina, Bratislava, Slovakia

² Zuse Institut Berlin, Germany

February 12, 2015

Abstract

We prove a mathematical programming characterisation of approximate partial D-optimality under general linear constraints. We use this characterisation with a branch-and-bound method to compute a list of all exact D-optimal designs for estimating a pair of treatment contrasts in the presence of a nuisance time trend up to the size of 24 consecutive trials.

1 Introduction

Consider the linear regression model $Y = \mathbf{F}\beta + \varepsilon$, where $Y = (Y_1, \dots, Y_n)^T$ is a vector of observations, $\beta \in \mathbb{R}^m$ is an unknown parameter, $\mathbf{F} = (\mathbf{f}(x_1), \dots, \mathbf{f}(x_n))^T$ is an $n \times m$ design matrix, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is a vector of random errors with $E(\varepsilon) = \mathbf{0}_n$, $\text{Var}(\varepsilon) = \sigma^2 \mathbf{I}_n$, $\sigma^2 \in (0, \infty)$. Suppose that the function $\mathbf{f} : \mathfrak{X} \rightarrow \mathbb{R}^m$, where \mathfrak{X} is a finite design space, is known and fixed, but the design points x_1, \dots, x_n can be chosen in \mathfrak{X} according to the objective of the experiment, see, e.g., [6, 8, 2].

A typical objective is to estimate a linear parameter subsystem $\mathbf{A}^T \beta$, where \mathbf{A} is a full-rank $m \times s$ matrix, $s \leq m$. It is a well-known fact that an unbiased linear estimator of $\mathbf{A}^T \beta$ exists if and only if the estimability condition $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M})$ is satisfied, where $\mathbf{M} = \mathbf{F}^T \mathbf{F}$ is the moment matrix and \mathcal{C} denotes the linear space generated by the columns of a matrix. In this case, the best linear unbiased estimator of $\mathbf{A}^T \beta$ is $\hat{\beta}_{\mathbf{A}} = \mathbf{A}^T \mathbf{M}^- \mathbf{F}^T Y$, and the estimator does not depend on the choice of the generalized inverse \mathbf{M}^- . Moreover, $\text{Var}(\hat{\beta}_{\mathbf{A}}) = \sigma^2 \mathbf{A}^T \mathbf{M}^- \mathbf{A}$ is non-singular.

Let $\mathbf{N}_{\mathbf{A}}(\mathbf{M}) = \min_{\mathbf{L}} \mathbf{L} \mathbf{M} \mathbf{L}^T$, where the minimum is taken on the set of all $s \times m$ matrices \mathbf{L} such that $\mathbf{L} \mathbf{A} = \mathbf{I}_s$, with respect to the Loewner ordering \preceq (for symmetric $s \times s$ matrices $\mathbf{N}_1, \mathbf{N}_2$ we define $\mathbf{N}_1 \preceq \mathbf{N}_2$ iff $\mathbf{N}_2 - \mathbf{N}_1$ is non-negative definite). The matrix $\mathbf{N}_{\mathbf{A}}(\mathbf{M})$ can be interpreted as the amount of information that the experiment conveys about $\mathbf{A}^T \beta$, see [8, Chapter 3]. Hence, $\mathbf{N}_{\mathbf{A}}(\mathbf{M})$ is called the information matrix for $\mathbf{A}^T \beta$. The information matrix is non-singular if and only if

the estimability condition is satisfied, in which case $\mathbf{N}_A(\mathbf{M}) = (\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A})^{-1}$. For estimating the entire parameter β , we have $\mathbf{N}_{\mathbf{I}_m}(\mathbf{M}) = \mathbf{M}$.

To measure the quality of estimation of $\mathbf{A}^T \beta$, we will use the criterion of D -optimality defined by $\Phi(\mathbf{N}) = (\det(\mathbf{N}))^{1/s}$ for $\mathbf{N} \in \mathcal{S}_+^s$, where \mathcal{S}_+^s is the set of all non-negative definite $s \times s$ matrices. Note that Φ is concave, continuous, positive homogeneous and Loewner-isotonic on \mathcal{S}_+^s , i.e., it is an information function [8, Chapter 5]. The composition $\Phi(\mathbf{N}_A(\cdot)) : \mathcal{S}_+^m \rightarrow [0, \infty)$ is again an information function, and it is called the criterion of partial D -optimality, or D_A -optimality (e.g., [6, Section IV.3] and [2, Section 10.2]). Explicitly, the criterion of D_A -optimality is $\Phi_A(\mathbf{M}) = (\det(\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A}))^{-1/s}$ if $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M})$ and $\Phi_A(\mathbf{M}) = 0$ otherwise.

Since the moment matrix $\mathbf{M} = \sum_i \mathbf{f}(x_i) \mathbf{f}^T(x_i)$ does not depend on the order of x_1, \dots, x_n , we can represent an exact experimental design by a function $\xi : \mathcal{X} \rightarrow \{0, 1, 2, \dots\}$ such that $\xi(x)$ means the number of trials to be performed in $x \in \mathcal{X}$. We will denote the set of all exact designs by Ξ^E . Note that the moment matrix corresponding to $\xi \in \Xi^E$ can be written as $\mathbf{M} = \mathbf{M}(\xi) = \sum_{x \in \mathcal{X}} \xi(x) \mathbf{f}(x) \mathbf{f}^T(x)$.

An approximate experimental design is *any* function $\xi : \mathcal{X} \rightarrow [0, \infty)$, which we understand as a relaxation of an exact design. The set of all approximate designs will be denoted by Ξ^A . The moment matrix $\mathbf{M}(\xi)$ of any $\xi \in \Xi^A$ is defined by the same formula as for the exact designs. For all $\xi \in \Xi^A$, let $\mathbf{N}_A(\xi) := \mathbf{N}_A(\mathbf{M}(\xi))$.

The usual constraint on an experimental design is that we are given a required size n of the experiment, i.e., we restrict the search to the designs ξ satisfying $\sum_{x \in \mathcal{X}} \xi(x) = n$. Such designs will be called size- n -constrained. However, in practice the designs must often satisfy additional constraints, which can represent restrictions on the experimental budget and the availability of material, see, e.g., [3]. Moreover, the ability to compute approximate optimal designs under more general constraints may be used as a key component of algorithms for computing optimal exact designs, such as the branch-and-bound (BNB) method that we propose in Section 3.1.

In this paper, we will consider general linear constraints of the form

$$\sum_{x \in \mathcal{X}} c(j, x) \xi(x) \leq b(j) \text{ for all } j = 1, \dots, k, \quad (1)$$

where $c : \{1, \dots, k\} \times \mathcal{X} \rightarrow \mathbb{R}$ and $b : \{1, \dots, k\} \rightarrow \mathbb{R}$. Given any fixed ordering on \mathcal{X} , the coefficient $c(j, x)$ can be arranged to a matrix \mathbf{C} of type $k \times |\mathcal{X}|$ and $b(j)$ can be arranged to a vector \mathbf{b} of length k . Note that each design can be represented by an $|\mathcal{X}|$ -dimensional vector with non-negative components. Accordingly, Ξ^E corresponds to $\{0, 1, 2, \dots\}^{|\mathcal{X}|}$ and Ξ^A corresponds to $[0, \infty)^{|\mathcal{X}|}$.

The designs satisfying (1) will be called (\mathbf{C}, \mathbf{b}) -designs. We will assume that there exists at least one exact (\mathbf{C}, \mathbf{b}) -design ξ , such that $\mathbf{A}^T \beta$ is estimable under ξ , and the set of all approximate (\mathbf{C}, \mathbf{b}) -designs is bounded. The sets of all exact and approximate (\mathbf{C}, \mathbf{b}) -designs will be denoted by $\Xi_{\mathbf{C}, \mathbf{b}}^E$ and $\Xi_{\mathbf{C}, \mathbf{b}}^A$ respectively. Evidently, $\Xi_{\mathbf{C}, \mathbf{b}}^E = \Xi_{\mathbf{C}, \mathbf{b}}^A \cap \mathbb{Z}^{|\mathcal{X}|}$ is a finite subset of the compact and convex set $\Xi_{\mathbf{C}, \mathbf{b}}^A$.

Let ξ_E^* be an exact (\mathbf{C}, \mathbf{b}) -design, $\mathbf{A}^T \beta$ be estimable under ξ_E^* and $\Phi(\mathbf{N}_A(\xi_E^*)) = \sup\{\Phi(\mathbf{N}_A(\xi_E)) : \xi_E \in \Xi_{\mathbf{C}, \mathbf{b}}^E\}$, where $\tilde{\Xi}_{\mathbf{C}, \mathbf{b}}^E := \{\xi_E \in \Xi_{\mathbf{C}, \mathbf{b}}^E : \mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M}(\xi_E))\}$. We will call ξ_E^* a D_A -optimal exact (\mathbf{C}, \mathbf{b}) -design. Analogously, we define a D_A -optimal approximate (\mathbf{C}, \mathbf{b}) -design and the symbol $\tilde{\Xi}_{\mathbf{C}, \mathbf{b}}^A$. Compactness of $\Xi_{\mathbf{C}, \mathbf{b}}^A \neq \emptyset$

and the continuity of $\Phi_{\mathbf{A}}$ imply the existence of a $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -design $\xi_{\mathbf{A}}^*$. The value $\Phi(\mathbf{N}_{\mathbf{A}}(\xi_{\mathbf{A}}^*))$ will be called the $D_{\mathbf{A}}$ -optimal value of $\Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}$.

Computing $D_{\mathbf{A}}$ -optimal approximate designs is a problem of convex optimization. The size- n -constrained $D_{\mathbf{A}}$ -optimal approximate designs can be computed, to any given precision, by a Fedorov-Wynn vertex direction algorithm or by a multiplicative algorithm (e.g., [15]). However, it is difficult to use the classical algorithms to compute $D_{\mathbf{A}}$ -optimal approximate designs under multiple linear constraints (cf. [3]). In this paper, we show that the problem of $D_{\mathbf{A}}$ -optimal approximate designs under (1) can be cast as a max-det programming problem ([13]), which can be efficiently solved by readily available software.

Computing $D_{\mathbf{A}}$ -optimal exact designs with constraints is in general a challenging problem of discrete optimization. To find the provably $D_{\mathbf{A}}$ -optimal exact designs in small to medium-size problems, it is possible to use the complete enumeration of all permissible designs, or more efficient enumeration methods. For instance, in [14] a BNB method is used for computing $D_{\mathbf{I}_m}$ -optimal exact size- n -constrained designs and in [11] a BNB method is used to compute $D_{\mathbf{I}_m}$ -optimal exact directly constrained designs. In this paper, we propose a specific BNB algorithm for computing $D_{\mathbf{A}}$ -optimal exact designs for the estimation of a set of treatment contrasts.

For large problems it is unrealistic to expect a rapid algorithm that always provides perfectly $D_{\mathbf{A}}$ -optimal exact designs. To find an efficient size- n -constrained exact design, it is possible to use an exchange heuristic, cf. [2, Chapter 12].

2 A mathematical programming characterization of $D_{\mathbf{A}}$ -optimal constrained approximate designs.

The following lemma is a simple consequence of the Schur complement characterization of positive semidefinite matrices (e.g. [8, Section 3.12]). Recently, this lemma has also been used to compute the support points of optimal designs [7, Section 6]. The symbol \mathcal{S}_{++}^s denotes the set of all positive definite $s \times s$ matrices.

Lemma 2.1. *Let \mathbf{A} be an $m \times s$ matrix of full rank, $m \geq s$, let $\mathbf{M} \in \mathcal{S}_{++}^m$, and let $\tilde{\mathbf{N}} \in \mathcal{S}_{++}^s$. Then the following two statements are equivalent: (i) $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M})$ and $\tilde{\mathbf{N}} \preceq (\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A})^{-1}$ for any¹ generalized inverse \mathbf{M}^{-} of \mathbf{M} ; (ii) $\mathbf{A} \tilde{\mathbf{N}} \mathbf{A}^T \preceq \mathbf{M}$.*

Let $\mathcal{I} = \{\mathbf{N} \in \mathcal{S}_{++}^s : \mathbf{N} = (\mathbf{A}^T \mathbf{M}(\xi)^{-1} \mathbf{A})^{-1} \text{ for some } \xi \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}\}$ and let $\tilde{\mathcal{I}} = \{\tilde{\mathbf{N}} \in \mathcal{S}_{++}^s : \mathbf{A} \tilde{\mathbf{N}} \mathbf{A}^T \preceq \mathbf{M}(\xi) \text{ for some } \xi \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}\}$.

Lemma 2.2. *$\mathcal{I} \subseteq \tilde{\mathcal{I}}$ and for each $\tilde{\mathbf{N}} \in \tilde{\mathcal{I}}$ there is some $\mathbf{N} \in \mathcal{I}$ such that $\tilde{\mathbf{N}} \preceq \mathbf{N}$.*

Proof. If $\mathbf{N} \in \mathcal{I}$ then, trivially, condition (i) of Lemma 2.1 is satisfied for some $\mathbf{M} = \mathbf{M}(\xi)$ such that $\xi \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}$. Hence, the part (ii) of Lemma 2.1 holds, i.e., $\mathbf{A} \mathbf{N} \mathbf{A}^T \preceq \mathbf{M}$, which implies $\mathbf{N} \in \tilde{\mathcal{I}}$. Thus $\mathcal{I} \subseteq \tilde{\mathcal{I}}$. Let $\tilde{\mathbf{N}} \in \tilde{\mathcal{I}}$ and let $\mathbf{M} = \mathbf{M}(\xi)$, $\xi \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}$, be such that $\mathbf{A} \tilde{\mathbf{N}} \mathbf{A}^T \preceq \mathbf{M}$. Note that by Lemma 2.1 we have $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M})$, i.e., we can set $\mathbf{N} = (\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A})^{-1} \in \mathcal{S}_{++}^s$. Hence, also by Lemma 2.1, we have $\tilde{\mathbf{N}} \preceq (\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A})^{-1} = \mathbf{N}$, which was to be proved. \square \square

¹If $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M})$, then $\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A}$ does not depend on the choice of the generalized inverse of \mathbf{M} .

Theorem 2.3. (i) *The set of information matrices of all $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -designs is identical to the set of all solutions \mathbf{N}^* of the problem*

$$\max \Phi(\mathbf{N}) \text{ s.t. } \mathbf{N} \in \mathcal{S}_{++}^s, \mathbf{A}\mathbf{N}\mathbf{A}^T \preceq \mathbf{M}(\xi), \xi \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}. \quad (2)$$

(ii) *If \mathbf{N}^* is any solution of (2), and $\xi^* \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}$ is any design satisfying $\mathbf{A}\mathbf{N}^*\mathbf{A}^T \preceq \mathbf{M}(\xi^*)$, then ξ^* is a $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -design.*

Proof. (i). As $\Phi : \mathcal{S}_+^s \rightarrow [0, \infty)$ vanishes on $\mathcal{S}_+^s \setminus \mathcal{S}_{++}^s$ and is positive on \mathcal{S}_{++}^s , the set of information matrices of all $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -designs is the set of all solutions of the problem $\max \Phi(\mathbf{N})$ s.t. $\mathbf{N} \in \mathcal{S}$. Moreover, note that Φ is strictly Loewner isotonic in the sense that if $\tilde{\mathbf{N}}, \mathbf{N} \in \mathcal{S}_{++}^s$ satisfy $\tilde{\mathbf{N}} \preceq \mathbf{N}$ and $\tilde{\mathbf{N}} \neq \mathbf{N}$, then $\Phi(\tilde{\mathbf{N}}) < \Phi(\mathbf{N})$. Therefore, we see from Lemma 2.2 that the set of information matrices of all $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -designs is equal to the set of all solutions of the problem $\max \Phi(\tilde{\mathbf{N}})$ s.t. $\tilde{\mathbf{N}} \in \mathcal{S}$.

(ii). Let \mathbf{N}^* be any solution of (2), and let $\xi^* \in \Xi_{\mathbf{C}, \mathbf{b}}^{\mathbf{A}}$ satisfy $\mathbf{A}\mathbf{N}^*\mathbf{A}^T \preceq \mathbf{M}(\xi^*)$. By Lemma 2.1 we have $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{M}(\xi^*))$ and $\mathbf{N}^* \preceq (\mathbf{A}^T \mathbf{M}^{-1}(\xi^*) \mathbf{A})^{-1} = \mathbf{N}_{\mathbf{A}}(\xi^*)$. Since $\mathbf{N}_{\mathbf{A}}(\xi^*)$ satisfies the constraints in (2) (cf. Section 3.14 in [8]), and since Φ is Loewner isotonic, we conclude that $\Phi(\mathbf{N}_{\mathbf{A}}(\xi^*)) \leq \Phi(\mathbf{N}^*) \leq \Phi(\mathbf{N}_{\mathbf{A}}(\xi^*))$. \square \square

By Theorem 2.3, computing $D_{\mathbf{A}}$ -optimal approximate (\mathbf{C}, \mathbf{b}) -designs can be cast as a problem of max-det programming, see [13]. A similar result appeared in [5, Section V.E] (with a different max-det formulation), for the case of a Bayesian framework with a prior density $\beta \sim \mathcal{N}(0, \Sigma)$. Max-det programs can be automatically reformulated as semidefinite programs (SDPs) by user-friendly interfaces such as PICOS [10]. SDPs is a class of optimization problems that are efficiently solvable by algorithms implemented in freely available software, for instance CVXOPT [12].

It is also worth mentioning that in the proof of Theorem 2.3 we did not use specific properties of D -optimality, except for the fact that Φ is a strictly Loewner isotonic information function vanishing on singular information matrices. Hence, we can use Theorem 2.3 also with many other criteria.

3 $D_{\mathbf{A}}$ -optimal designs for estimating a set of contrasts of treatment effects under the presence of a time trend

Suppose that we intend to perform n trials in a time sequence. For each trial $t \in \{1, \dots, n\}$ we select a treatment $u(t) \in \{1, \dots, v\}$, $v \geq 2$, with its effect $\tau_{u(t)} \in \mathbb{R}$. In addition to the treatment effects, the mean value of the response may depend on a nuisance time trend, which can be approximated by a polynomial of degree d . Hence, a natural model for the responses is

$$Y_t = \tau_{u(t)} + \theta_1 p_0(t) + \dots + \theta_{d+1} p_d(t) + \varepsilon_t; t = 1, \dots, n, \quad (3)$$

where $\theta_1, \dots, \theta_{d+1}$ are the parameters of the trend, p_0, \dots, p_d are polynomials of degrees $0, \dots, d$, and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. errors with zero mean and variance $\sigma^2 \in (0, \infty)$. In this model, we have $\mathfrak{X} = \{1, \dots, v\} \times \{1, \dots, n\}$ and $\mathbf{f}(u, t) = (\mathbf{e}_u^T, \mathbf{p}^T(t))^T$, where $\mathbf{e}_u \in \mathbb{R}^v$ is the standard u -th unit vector and $\mathbf{p}(t) = (p_0(t), \dots, p_d(t))^T$. The vector of model parameters is $\beta = (\tau_1, \dots, \tau_v, \theta_1, \dots, \theta_{d+1})^T$, but we suppose

that only $s = v - 1$ contrasts $\tau_2 - \tau_1, \dots, \tau_v - \tau_1$ are of interest. Hence, we will focus on $\mathbf{A}^T \boldsymbol{\beta}$, where $\mathbf{A}^T = (-\mathbf{1}_s, \mathbf{I}_s, \mathbf{0}_{s \times (d+1)})$, $s = v - 1$.

The moment matrix of any design ξ can be expressed in the form

$$\mathbf{M}(\xi) = \begin{pmatrix} \mathbf{M}_{11}(\xi) & \mathbf{M}_{12}(\xi) \\ \mathbf{M}_{12}^T(\xi) & \mathbf{M}_{22}(\xi) \end{pmatrix},$$

where $\mathbf{M}_{11}(\xi)$ is diagonal with $\sum_t \xi(1, t), \dots, \sum_t \xi(v, t)$ on the diagonal, $\mathbf{M}_{12}(\xi) = (\sum_t \xi(1, t) \mathbf{p}(t), \dots, \sum_t \xi(v, t) \mathbf{p}(t))^T$ and $\mathbf{M}_{22}(\xi) = \sum_t (\sum_u \xi(u, t)) \mathbf{p}(t) \mathbf{p}^T(t)$. If ξ is exact, then the diagonal elements of the matrix $\mathbf{M}_{11}(\xi)$ can be interpreted as replication numbers of individual treatments.

A usual assumption of an experiment modelled by (3) is that exactly one treatment is assigned to each time, therefore any permissible design ξ must satisfy

$$\sum_{u=1}^v \xi(u, t) = 1; t = 1, \dots, n. \quad (4)$$

If ξ satisfies (4), then $\mathbf{M}_{22}(\xi)$ does not depend on ξ . Moreover, it is simple to show that $\mathbf{N}_{\mathbf{A}}(\xi)$ does not depend on the choice of p_0, \dots, p_d (provided that p_i has degree i), but a suitable choice of the polynomials can lead to simpler computations. We used discrete orthogonal polynomials which makes $\mathbf{M}_{22}(\xi)$ diagonal.

For any system of $v - 1$ independent contrasts of treatments, the $D_{\mathbf{A}}$ -optimal designs are the same; the system $\tau_2 - \tau_1, \dots, \tau_v - \tau_1$ is chosen only for technical convenience. Also, note that for $\xi_1, \xi_2 \in \Xi^E$ we have $\Phi(\mathbf{N}_{\mathbf{A}}(\xi_1)) = \Phi(\mathbf{N}_{\mathbf{A}}(\xi_2))$ if the designs differ only in the labelling of treatments or if ξ_1 is only the time-reversed version of ξ_2 . Hence, we can call such designs isomorphic.

The study of “trend-resistant” or “systematic” designs dates back to Cox ([4]), who studied sequences involving two or three treatments under the presence of quadratic and cubic time trends. The extensive combinatorial theoretical results that followed are usually restricted to the so-called trend-free orthogonal designs, with a focus on selecting a suitable permutation of treatments, or combinations of factor levels, see, e.g., [9] for a survey. In [1] the authors propose a more universal approach based on optimality criteria. Compared to the combinatorial design results, the optimal design approach covers many more practical situations, e.g., it can be applied if the orthogonality is not attainable, the time points are not evenly spaced, or the time trend is non-polynomial.

In contrast to [1], we will use a BNB algorithm. This algorithm always results in a catalogue of perfectly $D_{\mathbf{A}}$ -optimal exact designs, although the time requirements are higher compared to the heuristic used in [1], especially for large n .

3.1 The BNB algorithm.

Let $a_1, \dots, a_n \in \{0, 1, \dots, v\}$. We define the template $[a_1, \dots, a_n]$ to be the set $\{\xi \in \Xi^A : \text{Eq. (4) holds and } (a_t \neq 0 \Rightarrow \xi(a_t, t) = 1) \text{ for all } t\}$. In particular, $[0, \dots, 0]$ corresponds to the set of all approximate designs satisfying (4). We call a template $[a_1, \dots, a_n]$ exact if $a_t > 0$ for all t , and composite if $a_t = 0$ for some t . Note that exact templates are singletons, whose only element is the design using treatments a_1, \dots, a_n for the times $1, \dots, n$. Clearly, for any selection $a_1, \dots, a_n \in \{0, 1, \dots, v\}$ the template $[a_1, \dots, a_n]$ is equal to $\Xi_{\mathbf{C}, \mathbf{b}}^A$ for some \mathbf{C} and \mathbf{b} . Thus, the $D_{\mathbf{A}}$ -optimal approximate design on any template can be computed using Theorem 2.3.

By a complete enumeration tree we will call any directed v -ary tree, with vertices corresponding to templates, satisfying: 1) The root template is $[0, \dots, 0]$; 2) Any non-terminal template $[a_1, \dots, a_n]$ is composite, and its v child nodes are $[a_1, \dots, a_{t^*-1}, u, a_{t^*+1}, \dots, a_n]$, $u = 1, \dots, v$, where t^* is selected such that $a_{t^*} = 0$. Therefore, the set of all permissible exact designs of a parent template is a disjunctive union of the sets of all permissible exact designs of its child templates; 3) The v^n terminal templates are the exact templates $[a_1, \dots, a_n]$, $a_1, \dots, a_n \in \{1, \dots, v\}$.

Suppose that we have constructed a sub-tree of a complete enumeration tree. Let Ξ correspond to a terminal template of the sub-tree. If the $D_{\mathbf{A}}$ -optimal value of Ξ is smaller than the value of the $D_{\mathbf{A}}$ -optimality criterion of the best available exact design then Ξ does not contain any $D_{\mathbf{A}}$ -optimal exact design. Hence, we can “prune” the branch of the tree, i.e., skip constructing the descendants of Ξ . This idea allows us to circumvent the construction of most of the complete enumeration tree, yet the terminal exact nodes of the final tree contain all $D_{\mathbf{A}}$ -optimal exact designs.

To fully specify the BNB method, we need to define the branching rule, that is, how to select the next composite template to be “refined” by constructing its child nodes, and which child nodes should be created. At each step, we choose the composite template $\Xi^* = [a_1^*, \dots, a_n^*]$ with the highest $D_{\mathbf{A}}$ -optimal value. Then we form its v child templates $[a_1^*, \dots, a_{t^*-1}^*, u, a_{t^*+1}^*, \dots, a_n^*]$, $u = 1, \dots, v$, such that t^* is selected to maximize the entropy of the distribution given by the pmf $\xi^*(\cdot, t)$, where ξ^* is an optimal approximate design on Ξ^* . The rationale behind this selection is that if the marginal design $\xi^*(\cdot, t^*)$ has a high “uncertainty”, fixing the treatments for the time t^* often makes the $D_{\mathbf{A}}$ -optimal values of the child templates smaller than the threshold given by the $D_{\mathbf{A}}$ -optimal value of the best available exact design.

In the actual implementation of our algorithm, all templates are stored in one of the lists `closed`, `open`, and `exact` (that is, the BNB tree itself is constructed only implicitly). The list `closed` is continually updated to include the templates that cannot contain any $D_{\mathbf{A}}$ -optimal exact design, since their $D_{\mathbf{A}}$ -optimal values are smaller than the $D_{\mathbf{A}}$ -optimal value Φ_{\max}^* of the best available exact design. The list `open` comprises the composite templates that still can contain an optimal exact design. Finally, the list `exact` contains the exact templates with the criterial value exactly Φ_{\max}^* . The algorithm terminates once the list `open` is empty, which corresponds to the moment when the list `exact` contains all $D_{\mathbf{A}}$ -optimal designs.

4 Example

In this section, we will show the results of the BNB algorithm when applied to the problem of computing $D_{\mathbf{A}}$ -optimal exact designs from Section 3 for $v = 3$ treatments, the cubic ($d = 3$) time trend, and $n = 6, \dots, 24$ trials.

For some n , there may exist $D_{\mathbf{A}}$ -optimal exact designs that differ with respect to a secondary criterion. From the set of all $D_{\mathbf{A}}$ -optimal exact designs computed by the BNB algorithm, we generated the complete list of all $D_{\mathbf{A}}$ -optimal exact designs, and from each class Ξ^* of mutually isomorphic $D_{\mathbf{A}}$ -optimal exact designs we selected:

1. A design ξ^A that is A -optimal in Ξ^* (see [2, Section 10.1]).
2. A design ξ^\dagger that is symmetric in the sense $\text{Var}(\hat{\beta}_{\mathbf{C}_2}(\xi^\dagger)) = \text{Var}(\hat{\beta}_{\mathbf{C}_3}(\xi^\dagger))$, provided that Ξ^* contains a design with this property. For $j \in \{2, 3\}$, the

symbol $\hat{\beta}_{e_j}(\xi^\dagger)$ denotes the BLUE of $\tau_j - \tau_1$ under the design ξ^\dagger .

3. A design ξ^{\leftrightarrow} that is symmetric with respect to the reversal of time, i.e., satisfying $\xi^{\leftrightarrow}(u, t) = \xi^{\leftrightarrow}(u, n+1-t)$ for all $u \in \{1, 2, 3\}$ and $t \in \{1, \dots, n\}$, provided that Ξ^* contains a design with this property.

The selected $D_{\mathbf{A}}$ -optimal exact designs ξ^A , ξ^\dagger , ξ^{\leftrightarrow} are not necessarily distinct. The results are summarized in the following list.

(6) 212313^{A,†}; (7) 1231231^{A,†}; (8) 12311231^{A,†}; (9) 123121321^{A,↔}; (10) 1232113231^{A,†}; (11) 23113221312^A; (12) 312213312213^{A,↔}; (13) 1233211123321^{A,↔}; (14) 31212331312213^A; (15) 123322111332231^{A,†}; (16) 1233212113132231^{A,†}; (17a) 31221133233112213^{A,↔}; (17b) 12332121312123321^{A,↔}; (18a) 132232131131232231^{A,↔}; (18b) 231131232232131132^{†,↔}; (18c) 231132132231231132^{A,↔}; (18d) 123321321123123321^{†,↔}; (19a) 3121223131313221213^{A,↔}; (19b) 1233211321231123321^{A,↔}; (20) 12323132112213132321^A; (21a) 123233112132122331312^A; (21b) 231311223213233112123[†]; (22) 1233122311231123312231^{A,†}; (23a) 12331222311311322213321^{A,↔}; (23b) 23112333122122133321132^{†,↔}; (23c) 31221313231213231312213^{A,↔}; (23d) 23113232123132123231132^{†,↔}; (24a) 231132132321123231231132^{A,↔}; (24b) 123321321213312123123321^{†,↔}; (24c) 123321321123321123123321^{A,↔}; (24d) 312213213312213312312213^{†,↔}.

Note that the $D_{\mathbf{A}}$ -optimal designs for the model with the cubic time trend are model-robust in the sense that they are either perfectly optimal, or very efficient for the models with polynomial trends of degrees 0 (i.e., if there is no trend), 1 as well as 2. For $n \geq 7$ ($n \geq 13$, $n \geq 17$), these efficiencies are higher than 0.9 (0.99, 0.999).

5 Conclusions

We described a mathematical programming characterisation of $D_{\mathbf{A}}$ -optimal approximate designs under linear constraints and a BNB method for computing $D_{\mathbf{A}}$ -optimal exact designs for estimating a set of treatment contrasts in the presence of a nuisance trend. In the illustrative example the nuisance parameters specify an unknown cubic time trend, but the algorithm can be analogously used for a large variety of other models, where the nuisance parameters represent any time trend, any spatial trend or the effects of blocks.

The main advantage of the BNB algorithm, compared to heuristic local-search methods is that it provides a complete list of perfectly optimal exact designs. The list can be used to select the best design according to a secondary criterion. Moreover, the list can motivate (or disprove) theoretical conjectures, or support the intuition about the desirable properties of designs. Note also that the mathematical programming approach permits adding linear constraints on the design that can represent, for example, constraints on the total cost of the experiment.

acknowledgement The research of the first author was supported by the VEGA 1/0163/13 grant of the Slovak Scientific Grant Agency.

References

- [1] Atkinson AC and Donev AN (1996) Experimental designs optimally balanced for trend. *Technometrics* 38(4):333–341

- [2] Atkinson AC, Donev AN and Tobias RD (2007) *Optimum Experimental Designs, with SAS*. Oxford University Press, New York
- [3] Cook D and Fedorov V (1995) Constrained optimization of experimental design. *Statistics* 26(2):129–148
- [4] Cox DR (1951) Some systematic experimental designs. *Biometrika* 38(3/4):312–323
- [5] Joshi S and Boyd S (2009) Sensor selection via convex optimization. *IEEE Trans Signal Process* 57(2):451–462
- [6] Pázman A (1986) *Foundations of Optimum Experimental Design*. D. Reidel Publishing Company, Dordrecht
- [7] Papp D (2012) Optimal designs for rational function regression. *J Am Stat Assoc* 107(497):400–411
- [8] Pukelsheim F (2006) *Optimal design of experiments*. SIAM, Philadelphia
- [9] Tack L and Vandebroek M (2001) (\mathcal{D}_t, c) -optimal run orders. *J Stat Plan Inference* 98(1-2):293–310
- [10] Sagnol G (2012) Picos, a python interface to conic optimization solvers. Technical Report 12-48, ZIB: <http://picos.zib.de>
- [11] Uciński D and Patan M (2007) D-optimal design of a monitoring network for parameter estimation of distributed systems. *J Glob Optim* 39(2):291–322
- [12] Vandenberghe L (2010) The CVXOPT linear and quadratic cone program solvers. <http://cvxopt.org/documentation/coneprog.pdf>
- [13] Vandenberghe L, Boyd S and Wu SP (1998) Determinant maximization with linear matrix inequality constraints. *SIAM J Matrix Anal Appl* 19(2):499–533
- [14] Welch WJ (1982) Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics* 24(1):41–48
- [15] Yu Y (2010) Monotonic convergence of a general algorithm for computing optimal designs. *Ann Stat* 38(3):1593–1606