Statistical Papers manuscript No.
(will be inserted by the editor)

Optimal design of experiments for hypothesis testing on ordered treatments via Intersection-Union Tests

Belmiro P.M. Duarte · Anthony C. Atkinson · Satya P. Singh · Marco S. Reis

Received: date / Accepted: date

Abstract We find experimental plans for hypothesis testing when a prior ordering of experimental groups or treatments is expected. Despite the practical interest of the topic, namely in dose finding, algorithms for systematically calculating good plans are still elusive. Here, we consider the Intersection-Union principle for constructing optimal experimental designs for testing hypotheses about ordered treatments. We propose an optimization-based formulation to handle the problem when the power of the test is to be maximized. This formulation yields a complex objective function which we handle with a surrogate-based optimizer. The algorithm proposed is demonstrated for several ordering relations. The relationship between designs maximizing power for the Intersection-Union Test (IUT) and optimality criteria used for linear regression models is analyzed; we demonstrate that IUT-based designs are well approximated by C–optimal designs and maximum entropy sampling designs while D_A-optimal designs are equivalent to balanced designs. Theoretical and numerical results supporting these relations are presented.

Keywords Optimal design of experiments, Hypothesis testing, Ordered treatments, Surrogate optimization, Power function, Alphabetic optimality.

Belmiro P.M. Duarte
Instituto Politécnico de Coimbra, Instituto Superior de Engenharia de Coimbra, Department of Chemical and Biological Engineering, Rua Pedro Nunes, Quinta da Nora, 3030−199 Coimbra, Portugal, and CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. Tel.: +351-239-790200, Fax: +351-239-790201, E-mail: bduarte@isec.pt

Anthony C. Atkinson
Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom. E-mail: A.C.Atkinson@lse.ac.uk

Satya P. Singh
Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur 208016, India. E-mail: singhsp@iitk.ac.in

Marco S. Reis
CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. E-mail: marco@eq.uc.pt

Mathematics Subject Classification (2000) 62K05 · 90C47

1 Motivation

Researchers in different areas often have prior beliefs about the order or direction of the parameters in comparisons. For example, a researcher might anticipate that a clinical treatment h_2 performs better than another (h_1), and simultaneously that both are better than a control. Confirming these beliefs corresponds to testing the hypotheses that μ_2, the expected outcome of h_2, is larger than that of μ_1 from h_1, which in turn outperforms the control with expectation μ_0. Specifically, the hypotheses to be tested are $\mathrm{H}_1 : \mu_0 \leq \mu_1 \leq \mu_2$ vs. $\mathrm{H}_0 : \mu_0 = \mu_1 = \mu_2$ with at least one strict inequality in H_1. H_1 is an order-constrained hypothesis, and includes more information than that of the simple alternative $\mathrm{H}_1 : \mu_i \neq \mu_j$ for at least one pair of i, j where $i \neq j \in \{0, 1, 2\}$. A major advantage of testing such one-sided hypotheses is that power can be increased or equivalently, that a smaller sample size is needed for equivalent power.

The problem of testing the homogeneity of the means of K groups against an ordered alternative was first addressed by Bartholomew (1959a,b). The incorporation of order constraints allows improving the precision of the estimators, as measured by their mean squared errors, and increasing the power of the associated tests (Davidov et al., 2014; Davidov and Herman, 2012; Farnan et al., 2014).

Despite the large body of literature on optimal design of experiments for parameter estimation and model discrimination, the optimal design of experiments for testing among groups is rarely addressed. An exception is that of finding optimal designs for comparing test treatments with a control, first introduced by Dunnett (1955, 1964). Later, the optimal allocation problem was solved by Bechhofer and Turnbull (1971); Bechhofer (1969); Bechhofer and Nocturne (1972).

Papers addressing the optimal design of experiments for ordered treatments are scarce. They are typically based on Likelihood Ratio Tests, being designated *Restricted Likelihood Ratio Test* (RLRT) designs if they explicitly incorporate the ordering relations and *Unrestricted Likelihood Ratio Test* (ULRT) designs otherwise. Hirotsu and Herzberg (1987) demonstrated that the optimal design allocates weights only to extreme groups, see also Antognini et al. (2021). An alternative formulation, using the weights of Abelson and Tukey (1963), circumvents this problem, with some weight being given to all groups. Singh et al. (1993) and Singh et al. (2008) evaluated the power function for various ordering schemes and found the optimal designs for three and five subgroups. Vanbrabant et al. (2015) investigated the effect of sample size reduction, when an increasing number of constraints is included into the hypothesis and obtained tables for a specified power level via Monte-Carlo sampling. Recently, Singh and Davidov (2019) proposed a minimax formulation for finding experimental designs for testing in the presence of order restrictions. The approach allows obtaining designs with more power than those of Dunnett (1955) and Singh et al. (1993). However, the authors noted that the designs obtained, although maximizing power, do not allocate any observation to intermediate groups, if any. Singh and Davidov (2019) also noted that, unlike Likelihood Ratio Tests, Intersection-Union Tests

(IUT) lead to optimal designs in which observations are allocated to all groups. The authors derived theoretical results for designs for some order relations but pointed out the complexity of generalizing to other orderings. Our methodology uses IUT to provide a general systematic approach to find experimental designs for ordered treatments.

This paper contains four elements of novelty: i. an optimization-based formulation to find optimal (exact) experimental designs for ordered treatments using the IUT–principle; ii. the use of surrogate-based optimization (SBO) to handle the complexity of the optimal design problem; we believe this to be the first paper that uses SBO to handle problems in the optimal design of experiments for IUT tests; iii. the application of the proposed methods to different ordering relations and treatments; and iv. the demonstration that IUT–based optimal designs are close to exact C–optimal and maximum entropy designs while the balanced designs are equivalent to exact $\mathrm{D_A}$–optimal designs.

The paper is organized as follows. Section 2 provides the background and the notation used to formulate the optimal design problem and solve it with SBO. Section 3 introduces the formulation used to solve the IUT design problem. Comparisons for different ordering schemes and distances between groups are presented in §4. Section 5 analyzes the relation between IUT-based designs and designs using alphabetic optimality criteria when the focus is on the parameters of the model. Section 6 reviews the formulation and offers a summary of the results obtained.

2 Notation and background

This section establishes the nomenclature used in the representation of the models. In §2.1 we overview the ANOVA model used to describe the ordered treatments test and introduce its equivalent graph-based representation. In §2.2 the IUT fundamentals and their use in the context of optimal design of experiments are introduced. Finally, §2.3 overviews the fundamentals of SBO which serve for solving the optimal design problem for the IUT criterion.

In our notation, bold face lowercase letters represent vectors, bold face capital letters stand for continuous domains, blackboard bold capital letters are used to denote discrete domains and capital letters are adopted for matrices. Finite sets containing ι elements are compactly represented by $[\![\iota]\!] \equiv \{1, \cdots, \iota\}$. The transpose operation of a matrix or vector is represented by "$^\intercal$". The cardinality of a vector is represented by $\mathrm{card}(\bullet)$, the trace of a matrix by $\mathrm{tr}(\bullet)$, and $\mathrm{ldet}(\bullet)$ represents $\ln[\det(\bullet)]$. The n-element row vector of ones is represented by $\mathbf{1}_n$ and the square identity matrix of size n is represented by I_n.

2.1 Ordered treatments ANOVA model

The sequence of (partially) ordered means can be represented as an order graph (Hwang and Peddada, 1994). Examples of the most common ordering schemes are shown in Figure 1. The vertices (or nodes) represent group means and an arrow from vertex μ_j to μ_i signifies that $\mu_j \geq \mu_i$. Vertices are called *roots* when there are only arrows leaving them, *leaves* when there are only arrows arriving, and *intermediate* when leaving and arriving arrows are involved. Let $\mathcal{R}$ be the set of roots in a ordering scheme, $\mathcal{L}$ the set of leaves, and $\mathcal{P}$ the set of ordering relations (corresponding to directed arrows) $\mu_i \leq \mu_j,\ i,\ j \in \{1, \cdots, p\}$. $r = \mathrm{card}(\mathcal{R})$ is the number of roots, $l = \mathrm{card}(\mathcal{L})$ the number of leaves and $p = \mathrm{card}(\mathcal{P})$ the number of ordering relations (i.e., pairs (i, j) in $\mathcal{P}$).

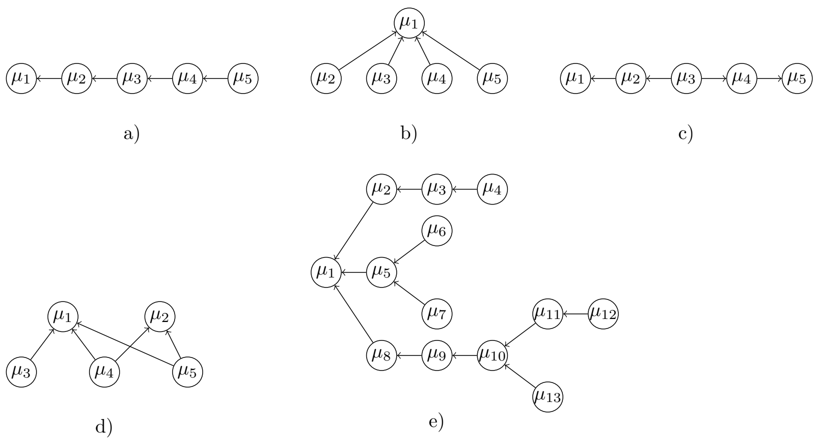

Figure 1 Examples of ordering schemes: a) simple ordering (SO); b) tree ordering (TO); c) umbrella ordering (UO); d) bipartite ordering (BO); and e) complex tree ordering (CTO).

The goal of experimental design for hypothesis testing is maximizing the power of rejecting the null hypothesis, H_0, in favor of an alternative hypothesis, H_1, through the allocation of individuals to treatments. Let the number of individuals included in the study be N, with K being the number of treatments; the first is reserved to be the control group. Further, let $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K)^\intercal$ be the vector of means of the K treatments; $\varPi_0 = \{\boldsymbol{\mu} \in \mathbb{R}^K : \ \mu_1 = \mu_2 = \cdots = \mu_K\}$ is the set of parameter (equality) relations under H_0, and $\varPi_1 = \{\boldsymbol{\mu} \in \mathbb{R}^K : \ Q\ \boldsymbol{\mu} \geq \mathbf{0}_p^\intercal\}$ the parameter inequalities under H_1 where $Q \in \mathbb{R}^{p \times K}$ is an ordering matrix (also known as a contrast matrix), $\mathbf{0}_p$ is the $p-$element row vector of zeros and p is the number of ordering relations. Consequently, we have $\varPi_0 \subset \varPi_1$. In subsequent sections we use $\varPi_\delta :\ \{\boldsymbol{\mu} \in \mathbb{R}^K :\ Q\ \boldsymbol{\mu} \geq \delta\ \mathbf{1}_p^\intercal\}$ to generically represent a larger class of tests where the distance of means is located at $\delta (> 0)$ from the null. Here, δ is the difference between treatment means which for simplicity we assume equal for all pairs (i, j) in $\mathcal{P}$. Matrix Q is formed by elements $Q_{i,j} \in \{-1, 0, +1\}$ where -1 is associated with groups with dominated means and $+1$ with groups with dominant means, 0 to the absence of a relationship, and p is the number of ordering restrictions or, equivalently, of arrows in the graph. In this paper we consider that the matrix of contrasts is known *a priori* and is fixed. Problems where the initial ordering is not confirmed by the experimental design are out of the scope of the paper, as they require treating the values of Q as additional parameters to be inferred from experiments.

The one-way Analysis of Variance (ANOVA) model considered in this study is represented as

$$y_{i,j} = \mu_i + \epsilon_{i,j}, \tag{1}$$

where $y_{i,j}$ is response of i^{th} experimental group to j^{th} experiment where $i \in \{1, \cdots, K\}$ and $j \in \{1, \cdots, n_i\}$. The mean of group i is $\hat{y}_i,\ i \in [\![K]\!]$, n_i is the number of individuals allocated to group i, $\sum_{i=1}^{K} n_i = N$ and N is the total number of individuals tested. The errors $\epsilon_{i,j}$ are assumed i.i.d. with normal distribution $\mathcal{N}(0, \sigma^2)$, where σ is the standard deviation.

Herein, ξ is a $K-$point design supported at $1, \cdots, k, \cdots, K$ treatments with n_k replicates allocated to treatment k, subject to $\sum_{k=1}^{K} n_k = N$. In what follows, let $\mathbf{n}$ be the vector of all possible replicates at the design points, with $\Omega_K^N = \{n_k \in \mathbb{Z}_{\geq 0} :\ \sum_{k=1}^{K} n_k = N,\ k \in [\![K]\!]\}$ being a $K-1$-dimensional standard simplex (containing K-groups allocation) where the superscript stands for the total number of individuals to allocate and the subscript for the number of groups; $\mathbb{Z}_{\geq 0}$ is the set of non-negative integers. An experimental design is compactly represented by

$$\xi = \begin{pmatrix} 1 & \cdots & k & \cdots & K \\ n_1 & \cdots & n_k & \cdots & n_K \end{pmatrix},$$

where the first line is for group ordering, and the second for the number of individuals allocated to each group. Thus, $\Xi_K^N \equiv [\![K]\!] \times \Omega_K^N$ is the set of all $K-$group feasible (ordered) exact designs constrained to Ω_K^N. This paper addresses the calculation of *exact* optimal designs, where by exact we mean small sample designs where the numbers of observations at design points are integers that sum to N. The optimization problem is complex and finding optimal exact designs is computationally challenging, especially when the IUT principle is used.

2.2 Intersection Union Tests

In this section we review the fundamentals of Intersection-Union Tests, a common alternative to Likelihood Ratio Tests, which is appropriate when the null hypothesis is expressed as a union of sets. A seminal version of IUT was proposed by Lehmann (1952), and later named by Gleser (1973). Applications of IUT to quality control problems were discussed by Berger (1982) and Saikali and Berger (2002). Berger and Hsu (1996) uses IUT to formalize bioequivalence tests, and Xiong et al. (2005) consider the application to two-arm clinical trials.

In our context, the IUT is used to test

$$\mathrm{H}_0 = \bigcup_{(i,j)\in\mathcal{P}} \mathrm{H}_0^{(i,j)} \quad \text{vs.} \quad \mathrm{H}_1 = \bigcap_{(i,j)\in\mathcal{P}} \mathrm{H}_1^{(i,j)} \tag{2}$$

where $\mathrm{H}_0^{(i,j)}$ is the null hypothesis for the $(i,j)^{\text{th}}$ pair of treatments (i.e., $\mu_i = \mu_j,\ (i,j) \in \mathcal{P}$) and $\mathrm{H}_1^{(i,j)}$ is the alternative hypothesis (i.e., $\mu_j - \mu_i \geq \delta$ ($>$

$0), (i,j) \in \mathcal{P})$. The rationale behind an IUT is that the overall null hypothesis, H_0, can be rejected only if each of the individual null hypotheses, $\mathrm{H}_0^{(i,j)}$ can be rejected.

Each pair of hypotheses $\mathrm{H}_0^{(i,j)}$ vs $\mathrm{H}_1^{(i,j)}$ can be tested using the statistic

$$g_{(i,j)} = \frac{(\hat{y}_j - \hat{y}_i)}{\sigma} \sqrt{\frac{n_i\, n_j}{n_i + n_j}}, \tag{3}$$

where $\mathbf{g}$ is a column vector with p elements $g_{i,j},\ (i,j) \in \mathcal{P}$. The null hypothesis for pairs (i,j) requires $\mu_i = \mu_j$; consequently $g_{(i,j)}$ will follow a standard normal distribution for all pairs $(i,j) \in \mathcal{P}$. The global null hypothesis is rejected if $g_{(i,j)} > c_\alpha,\ (i,j) \in \mathcal{P}$ with $c_\alpha = \Phi^{-1}(1-\alpha, 0, 1)$ where $\Phi^{-1}(1-\alpha, 0, 1)$ is the inverse of the $100 \times (1-\alpha)$ % percentage point of the standard normal distribution. It is noteworthy that only one critical value (c_α) is used for comparing all the pairs of treatments considered. When σ in (3) is unknown, it can be replaced by the usual mean squared error estimator, s, and the normal cdf is replaced by a noncentral t-distribution with the ratio $(\hat{y}_j - \hat{y}_i)/s$ being the measure of the effect size (Cohen, 1988).

Intersection-union tests differ from union-intersection tests in not requiring multiplicity adjustment (Tamhane, 1996, Section 3.3). Consequently, the design problem for intersection-union tests is simpler than that for union-intersection tests.

Now, let $\mathbf{c} = c_\alpha\, \mathbf{1}_p^\intercal$ be a p-elemental vector populated with the critical values c_α. Vector $\mathbf{g}$ follows a p-dimensional multivariate (non-central) normal distribution with mean $\boldsymbol{\nu}$ and a $p \times p$ correlation matrix R, i.e. $\mathcal{N}_p(\boldsymbol{\nu}, R)$. The elements of $\boldsymbol{\nu} \in \mathbb{R}^p$ are represented as follows

$$\nu_{(i,j)} = \frac{\mu_i - \mu_j}{\sigma} \sqrt{\frac{n_i\, n_j}{n_i + n_j}} = \frac{\delta}{\sigma} \sqrt{\frac{n_i\, n_j}{n_i + n_j}},\ (i,j) \in \mathcal{P}.$$

The matrix R contains the correlation between pairs $(i,j) \in \mathcal{P}$, each term depending on $\mathbf{n}$. The sample size as well as the effect size increase the power of a statistical test (Cohen, 1988). Herein, we consider the most inefficient scenario where the differences of means under analysis are equal to δ.

The power function measuring the probability that the test (2) rejects H_0 when H_1 is true is

$$\pi(\mathbf{g}|\mathbf{c}, \boldsymbol{\nu}, R) = \mathbb{P}\left[\bigcap_{(i,j)\in\mathcal{P}} \{g_{(i,j)} > c_\alpha\}\right] = \Phi(\mathbf{c}, \boldsymbol{\nu}, R), \tag{4a}$$

where $\Phi(\mathbf{c}, \boldsymbol{\nu}, R)$ is the cumulative multivariate normal distribution function for the $p-$dimensional domain $\otimes_{i=1}^{p} [c_\alpha, +\infty(\in \mathbb{R}^p$, given by

$$\Phi(\mathbf{c}, \boldsymbol{\nu}, R) = \int_{c_1}^{+\infty} \cdots \int_{c_K}^{+\infty} \phi(\mathbf{z}, \boldsymbol{\nu}, R)\, \mathrm{d}\mathbf{z}, \tag{5}$$

R is the correlation matrix between pairs of ordering relations, say (i,j) and (k,l), and

$$\phi(\mathbf{z}, \boldsymbol{\nu}, R) = \frac{1}{\sqrt{2^p\ \det(R)}}\ \exp\left[-\frac{(\mathbf{z}-\boldsymbol{\nu})^\intercal\, R^{-1}\, (\mathbf{z}-\boldsymbol{\nu})}{2}\right] \tag{6}$$

is the multivariate normal distribution function on $\mathbf{z}$. R is a positive definite matrix formed by elements $\rho_{(i,j),(k,l)}$, with $(i,j),\ (k,l) \in \mathcal{P}$ relating the pairs of ordering relations (Bretz, 1999; Dunnett, 1955; Dunnett and Sobel, 1954; Lee and Spurrier, 1995):

$$\rho_{(i,j),(k,l)} = \begin{cases} 1 & \text{if } i = k \wedge j = l \\ -\sqrt{\frac{n_i\, n_l}{(n_i+n_j)\,(n_k+n_l)}} & \text{if } (j = k \wedge i \neq l) \vee (i = l \wedge j \neq k) \\ \sqrt{\frac{n_j\, n_l}{(n_i+n_j)\,(n_k+n_l)}} & \text{if } (i = k \wedge j \neq l) \vee (j = l \wedge i \neq k) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

When R is not positive definite, which may occur in some initial iterations of SBO, we use the nearest symmetric positive definite (nspd) matrix (in the sense of Frobenius norm) computed with the algorithm of Higham (1988). The multivariate normal cdf is numerically computed with adaptive quadrature methods for bivariate and trivariate cases (Drezner, 1994; Genz, 2004), and a quasi-Monte Carlo integration scheme for more than 3-dimensions (Genz and Bretz, 2002).The positive definiteness of R is required, and is checked in each iteration before the computation of the multivariate normal cdf. The positive definiteness of R is checked by: i. finding the respective minimum eigenvalue ($\lambda_{\min}(R)$); and ii. deciding whether the property holds (or not). When $\lambda_{\min}(R)$ is larger than a small constant ϵ, the matrix is considered to be positive definite otherwise the positive definiteness validation check fails, and it is replaced by the corresponding nspd matrix. Here, we use $\epsilon = 1 \times 10^{-8}$.

The optimal design aims at maximizing (4a) by choice of the number of replicates of each of the K treatments under analysis, $\mathbf{n}$, in the space of feasible designs Ξ_K^N. We note the objective function is computationally challenging as it involves computing $\Phi(\mathbf{c}, \boldsymbol{\nu}, R)$ and the nspd of the correlation matrix, if needed. Apart from the complexity of constructing the gradient and the Hessian information, the problem is non-convex due to i. the decision variables ($\mathbf{n}$) being integer; ii. the necessity of approximating R by the nspd when required; and iii. the possible existence of multiple optima. The statistical approximations of numerically expensive objective functions in continuous Bayesian experimental designs, or for integrals in likelihood expressions, are considered by Overstall and Woods (2017) and Waite and Woods (2015) among others.

2.3 Surrogate-based optimization

In this Section we introduce the fundamentals of SBO which is used for solving the problem outlined in §2.2.

Surrogate-based optimization falls into the class of polynomial response surface methods and is typically used to handle problems involving complex and black-box functions, say $r(\mathbf{x})$, where the cost of fitting and evaluating the surrogate model is much less than a function evaluation and there are no algebraic expressions for the gradient nor for the Hessian matrix (Bhosekar and Ierapetritou, 2018; Kim and Boukouvala, 2020). The approach involves three stages: i. simulate the “real (complex) model", which may or may not be a black box model, for a limited number of well chosen data points; ii. construct an "approximate model" – a surface model – based on generated data; and iii. solve (optimize) the approximate model (also designated surrogate model) to generate a new set of points that emulate the "real model" but whose computation is much faster. Then iterate the three stages until convergence of the response of $f(\mathbf{x})$ to $r(\mathbf{x})$ is attained for a point $\mathbf{x}$ (Müller and Woodbury, 2017).

The models are generally formulated as

$$\min_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x}) \tag{8a}$$

$$\text{s.t. } \mathbf{r}(\mathbf{x}) \leq 0 \tag{8b}$$

$$x_\iota \in \mathbb{Z}_{\geq 0} \text{ for } \iota \in \mathbb{I}, \tag{8c}$$

where $f(\bullet)$ is the computationally cheap objective function that approximates the more complex one $r(\mathbf{x})$, (8b) denote the set of computationally expensive black-box inequality constraints, $\mathbf{X}$ is the finite domain of decision variables. Equation (8c) accounts for problems involving integer variables, say ι variables $x_\iota,\ \iota \in \mathbb{I}$; $\mathbb{I}$ is the set of integer variables.

The surrogate model is created from an initial number of simulations generated according to a sampling plan. Among the techniques used for generating initial sampling points the most common are the Latin Hypercube (LHC) designs (Müller and Day, 2019). Among the surrogate models, i.e. $f(\bullet)$, the most commonly used are interpolating models such as kriging (Martin and Simpson, 2005) and Radial Basis Functions (RBFs) (Buhmann, 2009; Powell, 1992). Both model types have been used for optimizing problems with computationally expensive objective functions, see Müller et al. (2013) for an example. Polynomial regression models and multivariate adaptive regression splines can also be used but they are non-interpolating surrogate models.

The iterative part of the algorithm has a sequence of steps: i. fit/update the surrogate model $f(\mathbf{x})$ using the set of sampling points available, i.e. $\mathcal{B}_n = \{(\mathbf{x}_i, r(\mathbf{x}_i)) \ :\ i \in \{1,\cdots,n\}\}$; ii. determine the "best point", $\mathbf{x}^{\text{best}} = \arg\min_{\mathbf{x}} m(\mathbf{x})$ since the last surrogate reset, where $m(\mathbf{x})$ is a merit function that includes both the surrogate function and a distance from existing points; iii. generate a set of ℓ trial points, $\mathcal{D}_{n,\ell} = \{\mathbf{x}_{n,\jmath}^{\text{trial}} = \mathbf{x}_n^{\text{best}} + \mathbf{e}_\jmath :\ \mathbf{e}_\jmath \in \mathbb{R}^d,\ \jmath \in [\![\ell]\!]\}$ by adding normal random perturbations scaled by the bounds in each dimension $i \in [\![d]\!]$ to $\mathbf{x}^{\text{best}}$; iv. determine the merit function at trial points and find the optimum (also designated the "adaptive point"), $\mathbf{x}^{\text{adap}}$; v. evaluate $r(\mathbf{x}^{\text{adap}})$, then update $\mathcal{B}_{n+1} \equiv \mathcal{B}_n \cup \mathbf{x}^{\text{adap}}$ with this new point and update the surrogate function, $f(\mathbf{x})$; vi. if $r(\mathbf{x}^{\text{adap}}) < r(\mathbf{x}^{\text{best}})$, the "best solution" is replaced by the adaptive point and the procedure iterated from step i.; vii. otherwise, the adaptive point is not included in $\mathcal{B}_n$; viii. the scale length is updated and the procedure iterated from step i. (Regis and Shoemaker, 2013). When integer variables are included in the problem, as here, the algorithm is similar, except for the computation of the minimum of the merit function where three different methods of sampling random points are used. Here, the merit function balances exploration – filling the gaps between the existing sample points by sampling in different zones of the optimization domain – and exploitation – using the available sample points to find an optimum (Regis and Shoemaker, 2007). Alizadeh et al. (2020) provide a recent review of the application of surrogate models in optimization. There are various tools for surrogate optimization available; see, for example, Eriksson et al. (2019); Le Digabel (2011); Müller (2014, 2016); Müller and Woodbury (2017). In §3 we use the algorithm proposed by Regis and Shoemaker (2007) which in turn uses a cubic RBF with a linear tail as the surrogate model (Gutmann, 2001).

3 Formulation for optimal design of experiments

In this section we introduce optimization formulations for finding $K-$treatment designs for ordered relations.

The optimization problem is as follows:

$$\max_{\mathbf{n}} \varPhi(\mathbf{c}, \boldsymbol{\nu}, R) \tag{9a}$$

$$\text{s.t. } c_i \geq \varPhi^{-1}(1-\alpha, 0, 1), \quad i \in [\![p]\!] \tag{9b}$$

$$\nu_{(i,j)} = \frac{\delta}{\sigma}\sqrt{\frac{n_i\, n_j}{n_i + n_j}}, \quad (i,j) \in \mathcal{P} \tag{9c}$$

$$\text{Equation (7)} \tag{9d}$$

$$R = \{\rho_{(i,j),(k,l)}\}, \quad (i,j),\ (k,l) \in \mathcal{P} \tag{9e}$$

$$n_K = N - \sum_{k=1}^{K-1} n_k \tag{9f}$$

$$\mathbf{n} \in \mathbb{Z}_{\geq 0},\ \mathbf{n} \leq N\,\mathbf{1}_K. \tag{9g}$$

Equation (9a) is the objective function, (9b) is used to construct $\mathbf{c}$, (9c) finds the mean difference for all pairs of treatments, (9d) computes the elements of the correlation matrix and (9e) estimates the correlation matrix between ordered pairs. To reduce the degrees of freedom of the problem by one and simultaneously avoid the need to include an integer equality constraint (which may cause additional problems for the optimization solver), the simplex condition that guarantees that the summation of replicates to all groups is N is reformulated; the last treatment, K, receives any trials not previously allocated, see (9f). Finally, Eq. (9g) sets the domain of decision variables. The problem falls into the general form presented in (8) where (9b–9f) form the set of Equations represented by (8b), and (9g) corresponds to Eq (8c); the complexity of evaluating the objective function is notorious. Furthermore, the problem may have multiple optima. However, the equality constraints in (9) are explicit relations that can be computed sequentially with the objective function being a function of previously evaluated quantities.

The initial sample provided to the solver is formed by a set of $\max(20,\ 2^K)$ points generated with a LHC sampling algorithm on the integer domain of interest. Then, the objective function (9a) is evaluated at the initial sample of points. The results are used to construct and optimize an approximate model, and new "improvement" points are added to the initial sample. This procedure is iterated until convergence. We use two stopping criteria in the numerical solution: i. reaching the maximum number of function evaluations, which was set to 700 in all problems solved; and ii. the tolerance of the objective function. To stop we require absolute and relative improvements of the objective function below 1×10^{-6} and 1×10^{-7}, respectively, in 150 consecutive iterations. The procedures that support the examples presented in this study were coded in Matlab® and call the SBO solver available on this platform – `surrogateopt` – and `MISO`, a solver developed by Müller (2016) for Mixed Integer Surrogate Optimization problems. All computations in this paper were carried using an AMD 8-Core processor machine running 64 bits Windows 10 operating system with 3.80 GHz.

4 Results

This Section presents optimal designs obtained by employing the formulation derived in §3. All the results were obtained with $\sigma = 1$ and $\delta = 0.7$ except when explicitly stated otherwise. We call a design *uniformly distributed* (or uniform) when the number of individuals allocated to each treatment is equal to N/K.

To help in the interpretation of the tables of results, each of the columns of the optimal designs is for a treatment; the first line is the treatment identifier (i) and the second line gives the respective value of $n_i,\ \forall i$ in the order graph (see Fig. 1). In §4.1 we study the influence of significance level, N and δ on optimal designs obtained for simple ordering. In Section 4.2 we find optimal designs for other ordering relations. All examples presented in the following sections require less than $2\,\mathrm{min}$ of CPU time.

4.1 The impact of significance level, sample size and difference between treatment means on optimal designs for simple ordering

In this Section we analyze the impact of the significance level (α), N and δ on optimal designs obtained for simple ordering with $K \in \{3, 4, 5, 6, 7\}$. As an example, the ordering matrix Q for $K = 3$ is

$$Q = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

First, we study the impact of the significance level and find the optimal designs for $\alpha = 0.05$ and $\alpha = 0.025$, with $N = 60$ and $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$. To avoid small values of power in the results for $K \in \{6,\ 7\}$, for those cases δ is increased to 1.5. The symbol Δ is used to measure the percentage improvement of the power of the IUT-based designs relative to the equivalent balanced designs. The results are presented in Table 1, and are in good agreement with the theoretical results derived by Singh and Davidov (2019, Theorem 7). The optimal designs found for both α's are close, but not necessarily equal. Although the displayed designs are equal, for other settings they may not be so. Further, as expected, the designs obtained for higher significance levels ensure higher power. For constant δ, the power of the optimal designs decreases with the number of ordering relations, and the designs become almost symmetric with respect to the middle treatment. Small distortions are observed relative to symmetry which are attributable to the integer nature of the decision variables, $\mathbf{n}$.

Now, we study the influence of N on optimal designs; α is fixed to 0.05 and $\delta = 1.0$. The optimal designs obtained for $N = \{30,\ 45\}$ are in Table 8 in Appendix A, and allow comparison with those obtained for $N = 60$ in Table 1. The comparison reveals, as expected, that increasing N increases the power. The relative optimal allocations are similar to those obtained for $N = 60$ (see Table 1). The designs are also nearly symmetric where the point of symmetry is the middle group.

Finally, we analyze the impact of δ on optimal designs. Table 9 in Appendix A contains the designs obtained for $\delta = \{0.9,\ 1.1\}$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = \{1.4,\ 1.6\}$ for $K \in \{6,\ 7\}$ assuming $N = 60$ and $\alpha = 0.05$. To get a clearer picture of the influence of δ, these designs can be analyzed together with those obtained for $\delta = 1.0$ and $\delta = 1.5$ in Table 1. The values of δ used for simulation were obtained by the addition and subtraction of 0.1 to reference values. The designs follow the trends found before and are equal to those in Table 1. Similarly, the designs are symmetrical, and one notices that the power increases with δ.

We now consider in more detail the optimal design obtained for $K = 3$, $N = 60$, $\sigma = 1.0$, $\delta = 1.0$ and $\alpha = 0.05$ (first line of Table 1). Figure 2 displays the power of designs obtained by varying n_1 and n_2 within the integer set $[\![58]\!]$ such that $n_3 = N - n_1 - n_2,\ n_3 > 0$. The response surface is convex, the maximum coinciding with the optimal design found in Table 1. Finally, we note that all IUT-based designs are more powerful than the equivalent balanced designs, the increment ranging from 0 to about $3.7\,\%$. Thus the loss of power from use of balanced designs is small. Further, the exact designs obtained from rounding the approximate designs of Singh and Davidov (2019) will also perform well as they are better than balanced designs.

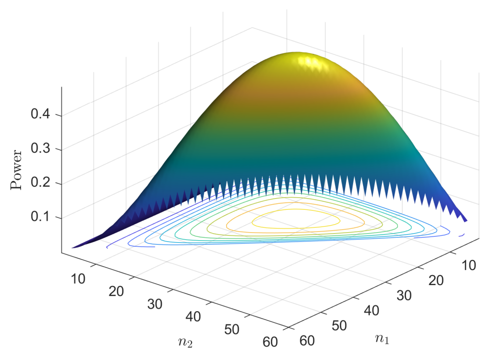

Figure 2 Objective function for experimental designs obtained varying n_1 and n_2 for $K = 3$, $N = 60$, $\sigma = 1.0$, $\delta = 0.7$ and $\alpha = 0.05$.

4.2 Optimal designs for other ordering relations

In this Section we find optimal designs for the other ordering arrangements in Figure 1 except complex tree ordering which is practically uncommon. All cases were solved for $N = 60$ and two values of α; i. 0.05; and ii. 0.025.

First we consider the umbrella ordering scheme and find designs for $K = \{3,\ 5,\ 7\}$ where the middle treatment is allocated to the maximum μ. Specifically, when $K = 3$ the dominant treatment is allocated to $k = 2$ and $\mu_2 - \mu_1 = \mu_2 - \mu_3 = \delta$. Similar approaches were followed for $K = \{5,\ 7\}$. For $K = 3$ the ordering matrix is

$$Q = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Here, we consider $\delta = 1.0$ for $K \in \{3,\ 5\}$ and $\delta = 1.5$ for $K \in \{7\}$. Table 2 presents the resulting optimal designs, which are symmetric. As expected, the power of the designs for $\alpha = 0.05$ are larger than those for $\alpha = 0.025$. The symmetrical allocation is independent of the significance level.

Now, we consider the tree ordering. The treatment allocated to $k = 1$ (first column in the contrast matrix) corresponds to the control group in many-to-one hypothesis testing. Specifically, for $K = 3$,

$$Q = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

The optimal designs for tree ordering with $K = \{3,\ 4,\ 5,\ 6,\ 7\}$ are in Table 3. For comparison we set $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$. We note that i. as with other ordering schemes, the power of the optimal designs decreases as K increases; and ii. more individuals are allocated to the control group than to other groups. As for previous ordering schemes, the power increases with α but the designs are not substantially affected by the significance level. For $K = \{3,\ 4\}$ these designs are in good agreement with those of Dunnett (1955).

Finally, for the bipartite ordering (see Figure 1) we find optimal designs for $K = 5$ and $p = \{5,\ 6\}$ corresponding to the ordering matrices

$$Q_1 = \begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} \text{ and } Q_2 = \begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix},$$

respectively. Matrix Q_2 includes an additional ordering relation between μ_1 and μ_3, and p is 6; the number of ordering relations for Q_1 is 5. Table 4 shows the optimal designs found for the two ordering matrices. The designs are the same for both values of α, with the designs obtained for Q_2 being slightly less powerful.

5 Relating the IUT criterion to other optimality criteria

In this Section we analyze the relation between the IUT-based designs of previous sections and the optimal designs obtained from other criteria such as those from alphabetic optimality. Because of the similarity of the ANOVA model to a multivariate linear regression model, there is interest in criteria that can be used for parameter estimation in regression. We first consider the $\mathrm{D_A}$-optimality criterion (see §5.1), then C–optimality, also known as $\mathrm{A_A}$–optimality, is considered (see §5.2); finally, in §5.3 the maximum entropy criterion is considered. Optimal designs are obtained for all of these criteria and compared with IUT-based designs.

5.1 $\mathrm{D_A}$–optimal designs

Here we analyze the relation between IUT designs and $\mathrm{D_A}$–optimal designs. $\mathrm{D_A}$–optimality is the generalization of D–optimality when interest lies in estimating only s linear combinations of the parameters, represented by $A^\intercal\,\boldsymbol{\mu}$ (Atkinson et al., 2007; Sibson, 1974). In our context $A = Q^\intercal$, $s = p$ and the number of parameters to be estimated is K. Here, the set of contrasts of interest is $\mathbb{E}(\boldsymbol{\theta}) = Q\,\boldsymbol{\mu}$. The variance-covariance matrix of the estimates $\hat{\boldsymbol{\theta}}$ is $C(\xi) = Q\,[\mathcal{M}(\xi)]^{-1}\,Q^\intercal$, where $\mathcal{M}(\xi)$ is the Fisher Information Matrix (FIM) for the model (1); $[\mathcal{M}(\xi)]^{-1} = \operatorname{diag}(1/n_1, \ldots, 1/n_K)$ is a $K \times K$ matrix, n_i being the number of individuals allocated to treatment i. We note that $C(\xi)$ depends on the design which also affects the correlation matrix resulting from the standardization of $C(\xi)$, here denoted as $R(\xi)$. The D-optimality criterion is applied to $C(\xi)$.

The uniform design is $\mathrm{D_A}$–optimal for any model $Q\,\boldsymbol{\mu}$ when Q has rank $K-1$. This follows from the invariance of the ordering induced by D–optimality with respect to any regular reparameterization, see Pukelsheim (1993, Section 6.2), corroborated by Rosa (2018, Section 3.2). Thus, approximate $\mathrm{D_A}$–optimal designs for $\boldsymbol{\theta}$ are uniform, that is balanced, designs. The extension of the result to exact $\mathrm{D_A}$–optimal designs is straightforward, only requiring that N/K be integer. When N/K is non-integer the designs allocate $\lfloor N/K \rfloor$ to each group and the remaining $N - K\,\lfloor N/K \rfloor$ are allocated indifferently, one to each different group; here $\lfloor \bullet \rfloor$ is the floor operator. Since balanced designs were used in Tables 1-4 for comparing power, we omit further presentation here. We recall that balanced designs have less power than IUT designs (the difference is $2.15\,\%$ on average). Consequently, the $\mathrm{D_A}$–optimality criterion produces designs that under perform IUT designs when the purpose is hypothesis testing.

5.2 C–optimal designs

In this Section we relate IUT-based designs to C–optimal designs. The C–optimality criterion is used when several linear combinations of parameters are of interest and we minimize $\operatorname{tr}\{Q\,[M(\xi)]^{-1}\,Q^\intercal\}$ where Q is the matrix of contrasts.

In our settings, C–optimality (see Silvey (1980, p. 48) and Atkinson et al. (2007, p. 143)) provides designs which are almost powerful as IUT designs. An approximate C–optimal design for model (1) is obtained by solving the following optimization problem

$$\min_{\xi\in\Xi_{\mathrm{K}}^{\mathrm{N}}} \mathrm{tr}[C(\xi)] = \min_{\xi\in\Xi_{\mathrm{K}}^{\mathrm{N}}} \mathrm{tr}[Q\ [\mathcal{M}(\xi)]^{-1}\ Q^\intercal]. \tag{10}$$

For evidence that the design criterion (10) is connected to IUT designs, we consider a tree ordering relation. For tree order the mean vector of $\mathbf{z} = (z_{(1,\,1)},\ldots,z_{(1,\,K)})^\intercal$ is $\boldsymbol{\mu} = \lambda\ g(\beta)\ \mathbf{1}_{K-1}$, where $g(\beta) = \sqrt{\beta\,(1-\beta)/[\beta\,(K-2)+1]}$, $\beta = n_1/N$, $\lambda = \sqrt{N}\,\delta/\sigma$ (Singh and Davidov, 2019). Since the power function is an increasing function of $g(\beta)$, for large λ the power is maximized when $g(\beta)$ is also maximized. It can be shown that $g(\beta)$ attains its maximum when $\beta = \beta_{\mathrm{IUT}} = \beta_{\text{C-opt}} = 1/(\sqrt{K-1}+1)$. Therefore, for large λ's, the proportion assigned by the IUT design to the control group is $\beta_{\mathrm{IUT}}(= \beta_{\text{C-opt}})$ and is $(1-\beta_{\mathrm{IUT}})/(K-1)$ to each treatment group. For these designs the ratio of control to treatment allocation is $\sqrt{K-1}$ which coincides with Dunnett's allocations for control versus multiple treatments comparisons. See especially Figures 1 and 2 of Dunnett (1955). Theorem 1 establishes properties of C–optimal designs which we compare with IUT-based designs.

Theorem 1 *For a contrast matrix Q an approximate C–optimal design is given by $\xi_{\textit{C-opt}} = (w_{\textit{C-opt},1},\ldots,w_{\textit{C-opt},K})^\intercal$, where*

$$w_{\textit{C-opt},i} = \frac{\sqrt{\mathbf{q}_i{}^\intercal\mathbf{q}_i}}{\sum_{k=1}^{K}\sqrt{\mathbf{q}_k{}^\intercal\mathbf{q}_k}} \quad \textit{for} \quad i\in[\![K]\!], \tag{11}$$

$\mathbf{q}_i$ is the i^{th} column of Q.

The proof follows from Pukelsheim (1993, Corollary 8.8) by assuming that $X = I_p$ and $K = Q^\intercal$ (to simplify the comparison we follow the original nomenclature with K being the matrix containing the set of linear parametric combinations of interest). Two immediate corollaries follow from (11).

Corollary 1 *Approximate C–optimal allocations for simple ordering are given by $w_{\textit{C-opt},1} = w_{\textit{C-opt},K} = 2/[2+(K-2)\sqrt{2}]$ and $w_{\textit{C-opt},2} = w_{\textit{C-opt},K-1} = (1 - 2\,w_{\textit{C-opt},1})/(K-2)$.*

Corollary 2 *For a bipartite ordering relation, approximate C–optimal allocations are given by*

$$\begin{aligned} w_{\textit{C-opt},i} &= \frac{1}{\mathrm{card}(\mathcal{R}) + \sqrt{\mathrm{card}(\mathcal{L})\ \mathrm{card}(\mathcal{R})}} \quad \textit{for} \quad i\in[\![\mathrm{card}(\mathcal{R})]\!] \quad \textit{and} \\ w_{\textit{C-opt},j} &= \frac{1}{\mathrm{card}(\mathcal{L}) + \sqrt{\mathrm{card}(\mathcal{L})\ \mathrm{card}(\mathcal{R})}} \quad \textit{for} \quad j\in[\![\mathrm{card}(\mathcal{L})]\!]. \end{aligned}$$

Now, we formulate the optimization problem to determine exact C–optimal problems in the design space Ξ_K^N. The optimal design problem is

$$\min_{\xi}\ \mathrm{tr}[C(\xi)] \tag{12a}$$

$$\text{s.t. } C(\xi) = Q \ [\mathcal{M}(\xi)]^{-1} \ Q^\intercal \tag{12b}$$

$$[\mathcal{M}(\xi)]^{-1} = \begin{pmatrix} 1/n_1 & & \\ & \ddots & \\ & & 1/n_K \end{pmatrix} \tag{12c}$$

$$\mathbf{1}_K^\intercal \ \mathbf{n} = N \tag{12d}$$

$$\mathbf{n} \in \mathbb{Z}_{\geq 0}^K. \tag{12e}$$

This problem was solved with a MINLP formulation proposed by Duarte et al. (2020) using the GAMS environment (GAMS Development Corporation, 2013). Specifically, a MINLP global solver based on the branch-and-reduce algorithm – BARON (Sahinidis, 2014) – is used.

Table 5 presents the C–optimal designs for the setups used for computing IUT designs for simple and tree ordering relations, i.e. $N = 60$, $\sigma = 1$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$. The results show that C–optimum designs have power only very slightly less than those of the IUT designs. Further, C–optimal designs are in good agreement with i. IUT designs (see Tables 1 and 3); ii. the designs found by Dunnett (1955) for tree ordering relations for $K \in \{3,\ 4\}$; iii. approximate designs predicted by Corollary 1; and iv. the maximum entropy designs to be described in §5.3. The results for umbrella ordering for $K \in \{3,\ 4,\ 5\}$ and bipartite ordering for both contrast matrices (Q_1 and Q_2) in Table 6 show the same trends. The designs are again similar to IUT designs and the approximate designs of Corollary 2 for biregular ordering.

5.3 Maximum entropy designs

Finally, we consider maximum entropy designs. Shewry and Wynn (1987) introduced the notion of sampling by maximum entropy when the design space is discrete. They showed that the expected change in information provided by an experiment is maximized by the design that maximizes the entropy of the observed responses since entropy is the negative of information. This kind of experimental design has been considered for certain spatial models, as well as in the selection of computer experiments (Currin et al., 1991) and for finding Bayesian optimal experimental designs (Sebastiani and Wynn, 2000).

If the regression parameters are fixed, as they are for Q, the entropy criterion reduces to $\max_\xi \text{ldet}[R(\xi)]$ where $R(\xi)$ is the correlation matrix (Jin et al., 2005; Koehler and Owen, 1996). Since $\det[R(\xi)] = \det[C(\xi)]/\prod_{k=1}^K C_{i,i}$, where $C_{i,i}$ are the diagonal elements of $C(\xi)$, the problem is equivalent to $\max_\xi \text{ldet}[C(\xi)] + \text{ldet}\{[I_p \circ C(\xi)]^{-1}\}$ (Anstreicher et al., 2001; Cover and Thomas, 2006). Here $I_p \circ C(\xi)$ provides the diagonal matrix formed by the diagonal elements of the matrix $C(\xi)$ and $\circ$ stands for the Hadamard (or elementwise) product. Thus, the MINLP problem to find maximum entropy designs is given by:

$$\max_\xi \ \text{ldet}[C(\xi)] + \text{ldet}\{[I_p \circ C(\xi)]^{-1}\} \tag{13a}$$

$$\text{s.t. } C(\xi) = Q\ [\mathcal{M}(\xi)]^{-1}\ Q^\intercal \tag{13b}$$

$$[\mathcal{M}(\xi)]^{-1} = \begin{pmatrix} 1/n_1 & & \\ & \ddots & \\ & & 1/n_K \end{pmatrix} \tag{13c}$$

$$\mathbf{1}^\intercal\ \mathbf{n} = N \tag{13d}$$

$$\mathbf{n} \in \mathbb{Z}^K_{\geq 0}. \tag{13e}$$

Table 7 presents the optimal maximum entropy designs obtained for simple and tree ordering relations with (13). A MINLP global solver was also used to assure global optimality. The designs obtained are similar to those produced by the IUT criterion (see the results in Tables 1 and 3 and C–optimal designs in Table 5), and are independent of α. We compared the power of the optimal maximum entropy designs for $\alpha = \{0.05, 0.025\}$ and observed that they are slightly less powerful than the IUT, equivalent to C-optimal designs, although more powerful than uniform designs. However, the relative differences are small.

6 Conclusions

In this paper we consider the optimal design of experiments for hypothesis testing of ordered treatments employing the Intersection-Union Test framework. The optimal design problem was formalized as a Mixed Integer Nonlinear Programming problem. Given the complexity of the objective function, a Surrogate-Based Optimization solver was used for the solution. The results obtained are in good agreement with previous theoretical results which are available for only a few cases. We tested the formulation to study the influence of i. the confidence level; ii. the sample size; and iii. the difference between treatment means (i.e., the effect size) for simple ordering relations (see §4.1). Optimal designs for other ordering relations are in §4.2. Typically, the optimal designs found are more powerful than balanced designs and ensure at least equal power to those of Dunnett (1955) for tree ordering relations.

Singh and Davidov (2019) developed theoretical results supporting the construction of optimal experimental designs using the Intersection-Union Test framework for ordered treatments. Their results are limited to some ordering relations and number of groups. They noted that the generalization is problematic due to the need of integrating a complex multivariate cdf. Here we have introduced a systematic way to handle the problem of constructing exact designs, a problem which is both more challenging than that of finding approximate designs and of immediate applicability. We have formulated all our numerical design problems as Mixed Integer Nonlinear Programmes. Given the complexity of the objective function, we use SBO to handle the resulting formulation for IUT designs. We believe this is the first paper where this technique has been used for the construction of exact designs. Our numerical approach allows addressing more complex ordering schemes and more groups than those of Singh and Davidov (2019). Although of the influence of the sample size on standardized mean difference of pairs of treatments, the approximate optimal designs based on IUT provide good estimates to exact optimal designs, see Singh and Davidov (2021). The main reason is that they maximize the power function and that occurs when all values of c_i in (5) are equal. This requirement, in turn, is independent of the group size since all of the c_i's are limited from above by c_α.

Our MINLP formulation enabled us to compare the IUT designs with designs from alphabetic optimality criteria used for model fitting. The theoretical results available for C–optimality for ordered treatments are limited to simple and bipartite ordering (the corollaries to Theorem 1). With the numerical formulation we have been able to construct optimal designs for other ordering schemes, for example the tree ordering results in Table 5. Finally, there are no theoretical results available for maximum entropy designs, so that the numerical treatment is the only approach.

Our results show that IUT-based designs are well approximated by C–optimal and maximum entropy designs which are superior to $\mathrm{D_A}$–optimal designs that correspond to uniform allocation schemes. The IUT–based designs are systematically slightly more powerful than alphabetic designs while the increase in terms of complexity of computation is marginal. While the former requires SBO to address the complexity and non-convexity of the objective function, the latter criteria require a global MINLP optimizer to guarantee the optimum is achieved.

Author statement

B.P.M. Duarte: Research, Conceptualization, Methodology, Writing original draft preparation. A.C. Atkinson: Research, Validation, Reviewing and editing. S.P. Singh: Validation, Reviewing and editing. M.S. Reis: Validation, Reviewing and editing.

References

1. Abelson, R.P., Tukey, J.W.: Efficient utilization of non-numerical information in quantitative analysis general theory and the case of simple order. The Annals of Mathematical Statistics **34**(4), 1347–1369 (1963)
2. Alizadeh, R., Allen, J.K., Mistree, F.: Managing computational complexity using surrogate models: a critical review. Research in Engineering Design **31**, 275–298 (2020)
3. Anstreicher, K.M., Fampa, M., Lee, J., Williams, J.: Maximum-entropy remote sampling. Discrete Applied Mathematics **108**(3), 211–226 (2001). DOI https://doi.org/10.1016/S0166-218X(00)00217-1
4. Antognini, A.B., Frieri, R., Novelli, M., Zagoraiou, M.: Optimal designs for testing the efficacy of heterogeneous experimental groups. Electronic Journal of Statistics **15**(1), 3217 – 3248 (2021). DOI 10.1214/21-EJS1864
5. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)
6. Bartholomew, D.J.: A test of homogeneity for ordered alternatives. Biometrika **46**, 36 – 48 (1959a)
7. Bartholomew, D.J.: A test of homogeneity for ordered alternatives. II. Biometrika **46**, 328 – 335 (1959b)

8. Bechhofer, R., Turnbull, B.: Optimal allocation of observations when comparing several treatments with a control (III): Globally best one-sided intervals for unequal variances. In: Gupta, S.S., Yackel, J. (eds.) Statistical Decision Theory and Related Topics, pp. 41 – 78. Academic Press (1971)
9. Bechhofer, R.E.: Optimal allocation of observations when comparing several treatments with a control. In: Krishnaiah, P.R. (ed.) Multivariate Analysis II, pp. 673 – 685. Academic Press (1969)
10. Bechhofer, R.E., Nocturne, D.J.M.: Optimal allocation of observations when comparing several treatments with a control, II: 2-sided comparisons. Technometrics **14**(2), 423 – 436 (1972)
11. Berger, R.L.: Multiparameter hypothesis testing and acceptance sampling. Technometrics **24**(4), 295–300 (1982)
12. Berger, R.L., Hsu, J.C.: Bioequivalence trials, intersection-union tests and equivalence confidence sets. Statistical Science **11**(4), 283–302 (1996)
13. Bhosekar, A., Ierapetritou, M.: Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers & Chemical Engineering **108**, 250–267 (2018)
14. Bretz, F.: Powerful modifications of Williams' test on trend. Ph.D. thesis, University of Hannover (1999)
15. Buhmann, M.D.: Radial Basis Functions - Theory and Implementations, vol. 12. Cambridge University Press (2009)
16. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Lawrence Erlbaum Associates, Publishers, New York (1988)
17. Cover, T.M., Thomas, J.A.: Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience (2006)
18. Currin, C., Mitchell, T., Morris, M., Ylvisaker, D.: Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. Journal of the American Statistical Association **86**(416), 953–963 (1991). DOI 10.1080/01621459.1991.10475138
19. Davidov, O., Fokianos, K., Iliopoulos, G.: Semiparametric inference for the two-way layout under order restrictions. Scandinavian Journal of Statistics **41**(3), 622–638 (2014)
20. Davidov, O., Herman, A.: Ordinal dominance curve based inference for stochastically ordered distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **74**(5), 825–847 (2012)
21. Drezner, Z.: Computation of the trivariate normal integral. Mathematics of Computation **62**(205), 289–294 (1994)
22. Duarte, B.P.M., Granjo, J.F.O., Wong, W.K.: Optimal exact designs of experiments via Mixed Integer Nonlinear Programming. Statistics and Computing **30**, 93–112 (2020)
23. Dunnett, C.W.: A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association **50**(272), 1096–1121 (1955)
24. Dunnett, C.W.: New tables for multiple comparisons with a control. Biometrics **20**, 482–491 (1964)

25. Dunnett, C.W., Sobel, M.: A bivariate generalization of student's t–distribution, with tables for certain special cases. Biometrika **41**(1-2), 153–169 (1954). DOI 10.1093/biomet/41.1-2.153
26. Eriksson, D., Bindel, D., Shoemaker, C.A.: pySOT and POAP: An event-driven asynchronous framework for surrogate optimization (2019)
27. Farnan, L., Ivanova, A., Peddada, S.D.: Linear mixed efects models under inequality constraints with applications. PLOS ONE **9**(1) (2014)
28. GAMS Development Corporation: GAMS - A User's Guide, GAMS Release 24.2.1. GAMS Development Corporation, Washington, DC, USA (2013)
29. Genz, A.: Numerical computation of rectangular bivariate and trivariate normal and t probabilities. Statistics and Computing **14**, 251–260 (2004)
30. Genz, A., Bretz, F.: Comparison of methods for the computation of multivariate t probabilities. Journal of Computational and Graphical Statistics **11**(4), 950–971 (2002)
31. Gleser, L.J.: On a theory of intersection-union tests. Institute of Mathematical Statistics Bulletin **2**, 233–233 (1973)
32. Gutmann, H.M.: A radial basis function method for global optimization. Journal of Global Optimization **19**, 201–227 (2001)
33. Higham, N.J.: Computing a nearest symmetric positive semidefinite matrix. Linear Algebra and its Applications **103**, 103–118 (1988). DOI https://doi.org/10.1016/0024-3795(88)90223-6
34. Hirotsu, C., Herzberg, A.M.: Optimal allocation of observations for inference on k ordered normal population means. Australian Journal of Statistics **29**(2), 151–165 (1987)
35. Hwang, J.T.G., Peddada, S.D.: Confidence interval estimation subject to order restrictions. Ann. Statist. **22**(1), 67–93 (1994)
36. Jin, R., Chen, W., Sudjianto, A.: An efficient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inference **134**(1), 268–287 (2005). DOI https://doi.org/10.1016/j.jspi.2004.02.014
37. Kim, S.H., Boukouvala, F.: Surrogate-based optimization for mixed-integer nonlinear problems. Computers & Chemical Engineering **140**, 106,847 (2020)
38. Koehler, J.R., Owen, A.B.: Computer experiments. In: Gosh, S., Rao, C.R. (eds.) Handbook of Statistics, Vol. 13, Design and Analysis of Experiments, pp. 261–308. Elsevier, Amsterdam (1996)
39. Le Digabel, S.: Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. ACM Trans. Math. Softw. **37**(4) (2011)
40. Lee, R.E., Spurrier, J.D.: Successive comparisons between ordered treatments. Journal of Statistical Planning and Inference **43**(3), 323–330 (1995). DOI https://doi.org/10.1016/0378-3758(95)91803-B
41. Lehmann, E.L.: Testing multiparameter hypotheses. Ann. Math. Statist. **23**(4), 541–552 (1952)
42. Martin, J.D., Simpson, T.W.: Use of kriging models to approximate deterministic computer models. AIAA Journal **43**(4), 853–863 (2005)
43. Müller, J.: MATSuMoTo: The Matlab surrogate model toolbox for computationally expensive black-box global optimization problems. arXiv **abs/1404.4261**(1404.4261v1) (2014)

44. Müller, J.: MISO: mixed-integer surrogate optimization framework. Optimization and Engineering **17**, 177–203 (2016)
45. Müller, J., Day, M.: Surrogate optimization of computationally expensive black-box problems with hidden constraints. INFORMS Journal on Computing **31**(4), 689–702 (2019)
46. Müller, J., Shoemaker, C.A., Piché, R.: SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. Computers & Operations Research **40**(5), 1383–1400 (2013)
47. Müller, J., Woodbury, J.D.: GOSAC: global optimization with surrogate approximation of constraints. Journal of Global Optimization **69**, 117–136 (2017)
48. Overstall, A.M., Woods, D.C.: Bayesian design of experiments using approximate coordinate exchange. Technometrics **59**(4), 458–470 (2017). DOI 10.1080/00401706.2016.1251495
49. Powell, M.J.D.: The theory of radial basis function approximation in 1990. In: Light, W.A. (ed.) Advances in Numerical Analysis II: Wavelets, Subdivision, and Radial Functions, pp. 105–210. Oxford University Press, Oxford (1992)
50. Pukelsheim, F.: Optimal Design of Experiments. SIAM, Philadelphia (1993)
51. Regis, R.G., Shoemaker, C.A.: A stochastic radial basis function method for the global optimization of expensive functions. INFORMS Journal on Computing **19**(4), 497–509 (2007)
52. Regis, R.G., Shoemaker, C.A.: Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. Engineering Optimization **45**(5), 529–555 (2013). DOI 10.1080/0305215X.2012.687731
53. Rosa, S.: Optimal designs for treatment comparisons represented by graphs. AStA Advances in Statistical Analysis **102**(4), 479–503 (2018)
54. Sahinidis, N.: BARON 14.3.1: Global Optimization of Mixed-Integer Nonlinear Programs, *User's Manual*. The Optimization Firm, LLC, Pittsburgh, PA, USA. (2014)
55. Saikali, K.G., Berger, R.L.: More powerful tests for the sign testing problem. Journal of Statistical Planning and Inference **107**(1), 187–205 (2002)
56. Sebastiani, P., Wynn, H.P.: Maximum entropy sampling and optimal bayesian experimental design. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62**(1), 145–157 (2000). DOI https://doi.org/10.1111/1467-9868.00225
57. Shewry, M.C., Wynn, H.P.: Maximum entropy sampling. Journal of Applied Statistics **14**(2), 165–170 (1987). DOI 10.1080/02664768700000020
58. Sibson, R.: D_A-optimality and duality. In: Gani, J., Sarkadi, K., Vincze, I. (eds.) Progress in Statistics, Vol.2 – Proc. 9th European Meeting of Statisticians, Budapest, pp. 677–692. North-Holland, Amsterdam (1974)
59. Silvey, S.D.: Optimal Design. Chapman & Hall, London (1980)
60. Singh, B., Halabi, S., Schell, M.J.: Sample size selection in clinical trials when population means are subject to a partial order: one-sided ordered alternatives. Journal of Applied Statistics **35**(5), 583–600 (2008)
61. Singh, B., Schell, M.J., Wright, F.T.: The power functions of the likelihood ratio tests for a simple tree ordering in normal means: unequal weights. Communications in Statistics: Theory and Methods **22**(2), 425–449 (1993)

62. Singh, S.P., Davidov, O.: On the design of experiments with ordered treatments. Journal of the Royal Statistical Society Series B **81**(5), 881–900 (2019)
63. Singh, S.P., Davidov, O.: On efficient exact experimental designs for ordered treatments. Computational Statistics & Data Analysis **164**, 107,305 (2021). DOI https://doi.org/10.1016/j.csda.2021.107305
64. Tamhane, A.C.: Multiple comparisons. In: Gosh, S., Rao, C.R. (eds.) Handbook of Statistics, Vol. 13, Design and Analysis of Experiments, pp. 587–630. Elsevier, Amsterdam (1996)
65. Vanbrabant, L., Van De Schoot, R., Rosseel, Y.: Constrained statistical inference: sample-size tables for ANOVA and regression. Frontiers in Psychology **5**, 1565 (2015)
66. Waite, T.W., Woods, D.C.: Designs for generalized linear models with random block effects via information matrix approximations. Biometrika **102**(3), 677–693 (2015). DOI 10.1093/biomet/asv005
67. Xiong, C., Yu, K., Gao, F., Yan, Y., Zhang, Z.: Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. Clinical Trials **2**(5), 387–393 (2005)

Appendix A: Optimal designs for simple order relation

Here we present the optimal designs for tree ordering resulting from varying N and δ.

Table 1 Optimal designs for simple order relation ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$).

	$\alpha = 0.05$				$\alpha = 0.025$			
K	Design	Power	Power†	Δ (%)	Design	Power	Power†	Δ (%)
3	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 25 & 17 \end{pmatrix}$	0.8821	0.8710	1.11	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 25 & 17 \end{pmatrix}$	0.7896	0.7720	1.76
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.6393	0.6122	2.71	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.4542	0.4159	3.83
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 9 & 14 & 14 & 14 & 9 \end{pmatrix}$	0.3526	0.3167	3.59	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 9 & 14 & 14 & 14 & 9 \end{pmatrix}$	0.1659	0.1295	3.64
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	0.8062	0.7930	1.32	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	0.6552	0.6326	2.26
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 9 & 9 & 10 & 7 \end{pmatrix}$	0.6370	0.6155	2.15	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 9 & 9 & 10 & 7 \end{pmatrix}$	0.4252	0.3973	2.79

† - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design.

Table 2 Optimal designs for umbrella ordering ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 5\}$ and $\delta = 1.5$ for $K \in \{7\}$).

	$\alpha = 0.05$				$\alpha = 0.025$			
K	Design	Power	Power[†]	Δ (%)	Design	Power	Power[†]	Δ (%)
3	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 24 & 18 \end{pmatrix}$	0.8948	0.8883	0.65	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 24 & 18 \end{pmatrix}$	0.8185	0.8101	0.84
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 10 & 14 & 12 & 14 & 10 \end{pmatrix}$	0.4056	0.3879	1.77	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 10 & 14 & 12 & 14 & 10 \end{pmatrix}$	0.2317	0.2143	1.70
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 10 & 9 & 8 & 9 & 10 & 7 \end{pmatrix}$	0.6517	0.6306	2.11	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 10 & 9 & 8 & 9 & 10 & 7 \end{pmatrix}$	0.4534	0.4253	2.81

[†] - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design.

Table 3 Optimal designs for tree ordering ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$).

	$\alpha = 0.05$					$\alpha = 0.025$				
K	Design	Power	Power†	Δ (%)	Power‡	Design	Power	Power†	Δ (%)	Power‡
3	$\begin{pmatrix} 1 & 2 & 3 \\ 24 & 18 & 18 \end{pmatrix}$	0.8948	0.8883	0.65	0.8948	$\begin{pmatrix} 1 & 2 & 3 \\ 24 & 18 & 18 \end{pmatrix}$	0.8185	0.8101	0.84	0.8185
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 21 & 13 & 13 & 13 \end{pmatrix}$	0.7332	0.7146	1.86	0.7332	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 21 & 13 & 13 & 13 \end{pmatrix}$	0.6016	0.5838	1.78	0.6010
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 20 & 10 & 10 & 10 & 10 \end{pmatrix}$	0.5563	0.5387	1.76		$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 20 & 10 & 10 & 10 & 10 \end{pmatrix}$	0.4080	0.3940	1.40	
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 17 & 9 & 9 & 9 & 8 & 8 \end{pmatrix}$	0.8910	0.8516	3.94		$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 17 & 9 & 9 & 9 & 8 & 8 \end{pmatrix}$	0.8020	0.7511	5.09	
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 17 & 7 & 7 & 7 & 7 & 7 & 8 \end{pmatrix}$	0.8067	0.7569	4.98		$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 17 & 7 & 7 & 7 & 7 & 7 & 8 \end{pmatrix}$	0.6774	0.6264	5.10	

† - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design;
‡ - Power of the corresponding Dunnett (1955) design.

Table 4 Optimal designs for bipartite ordering ($N = 60$, $K = 5$, $\delta = 1.5$).

	$\alpha = 0.05$				$\alpha = 0.025$			
K	Design	Power	Power†	Δ (%)	Design	Power	Power†	Δ (%)
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 14 & 12 & 10 & 12 & 12 \end{pmatrix}$	0.9200	0.9148	0.52	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 14 & 12 & 10 & 12 & 12 \end{pmatrix}$	0.8473	0.8396	0.77
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 14 & 13 & 11 & 11 & 11 \end{pmatrix}$	0.9056	0.9029	0.27	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 14 & 13 & 11 & 11 & 11 \end{pmatrix}$	0.8240	0.8201	0.39

† - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design.

Table 5 Optimal designs for simple and tree order relation based on C–optimality criterion ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$).

	Simple order						Tree order					
K	Design	Optimum	Power†	Power‡	Power*	Power*	Design	Optimum	Power†	Power‡	Power*	Power*
3	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 25 & 17 \end{pmatrix}$	0.194	0.8821	0.7896	0.8821	0.7896	$\begin{pmatrix} 1 & 2 & 3 \\ 25 & 17 & 18 \end{pmatrix}$	0.194	0.8943	0.8176	0.8948	0.8185
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.389	0.6393	0.4542	0.6393	0.4542	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 22 & 13 & 13 & 12 \end{pmatrix}$	0.374	0.7308	0.5976	0.7332	0.6016
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 10 & 13 & 14 & 13 & 10 \end{pmatrix}$	0.651	0.3477	0.1576	0.3526	0.1659	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 20 & 10 & 10 & 10 & 10 \end{pmatrix}$	0.600	0.5535	0.3992	0.5563	0.4080
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	0.977	0.8061	0.6552	0.8062	0.6552	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 19 & 8 & 8 & 9 & 8 & 8 \end{pmatrix}$	0.874	0.8905	0.8001	0.8910	0.8020
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 10 & 9 & 9 & 7 \end{pmatrix}$	1.375	0.6362	0.4251	0.6370	0.4252	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 18 & 7 & 7 & 7 & 7 & 7 & 7 \end{pmatrix}$	1.190	0.8051	0.6755	0.8067	0.6774

† - Power computed for $\alpha = 0.05$; ‡ - Power computed for $\alpha = 0.025$
* - Power computed for IUT design for $\alpha = 0.05$; * - Power computed for IUT design $\alpha = 0.025$.

Table 6 Optimal designs for umbrella and bipartite order relations based on C–optimality criterion ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 5\}$ and $\delta = 1.5$ for $K \in \{7\}$).

	Umbrella order							Bipartite order						
K	Design	Optimum	Power†	Power‡	Power*	Power*		Contrast	Design	Optimum	Power†	Power‡	Power*	Power*
3	$\begin{pmatrix} 1 & 2 & 3 \\ 18 & 25 & 17 \end{pmatrix}$	0.194	0.8943	0.8176	0.8948	0.8185		Q_1	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 15 & 12 & 9 & 12 & 12 \end{pmatrix}$	0.811	0.9191	0.8457	0.9200	0.8473
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 10 & 13 & 14 & 13 & 10 \end{pmatrix}$	0.651	0.3994	0.2242	0.4056	0.2317		Q_2	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 14 & 13 & 11 & 11 & 11 \end{pmatrix}$	0.991	0.9056	0.8239	0.9056	0.8240
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 10 & 9 & 9 & 7 \end{pmatrix}$	1.375	0.6462	0.4446	0.6517	0.4534								

† - Power computed for $\alpha = 0.05$; ‡ - Power computed for $\alpha = 0.025$
* - Power computed for IUT design for $\alpha = 0.05$; * - Power computed for IUT design $\alpha = 0.025$.

Table 7 Optimal designs for simple and tree order relation based on maximum entropy criterion ($N = 60$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$).

	Simple order						Tree order					
K	Design	Optimum	Power†	Power‡	Power*	Power*	Design	Optimum	Power†	Power‡	Power*	Power*
3	$\begin{pmatrix} 1 & 2 & 3 \\ 17 & 25 & 18 \end{pmatrix}$	1.786	0.8821	0.7896	0.8821	0.7896	$\begin{pmatrix} 1 & 2 & 3 \\ 25 & 17 & 18 \end{pmatrix}$	1.786	0.8943	0.8176	0.8948	0.8185
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.759	0.6393	0.4542	0.6393	0.4542	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 22 & 13 & 13 & 12 \end{pmatrix}$	1.040	0.7308	0.5976	0.7332	0.6016
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 10 & 13 & 14 & 13 & 10 \end{pmatrix}$	−1.444	0.3477	0.1576	0.3526	0.1659	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 20 & 10 & 10 & 10 & 10 \end{pmatrix}$	−0.516	0.5535	0.3992	0.5563	0.4080
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	5.603	0.8061	0.6552	0.8062	0.6552	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 19 & 8 & 8 & 8 & 8 & 9 \end{pmatrix}$	6.642	0.8905	0.8001	0.8910	0.8020
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 9 & 10 & 10 & 10 & 9 & 6 \end{pmatrix}$	4.807	0.6296	0.4228	0.6370	0.4252	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 18 & 7 & 7 & 7 & 7 & 7 & 7 \end{pmatrix}$	6.526	0.8050	0.6755	0.8067	0.6774

† - Power computed for $\alpha = 0.05$; ‡ - Power computed for $\alpha = 0.025$
* - Power computed for IUT design for $\alpha = 0.05$; * - Power computed for IUT design $\alpha = 0.025$.

Table 8 Optimal designs for simple order relation ($\alpha = 0.05$, $\delta = 1.0$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = 1.5$ for $K \in \{6,\ 7\}$).

	$N = 30$				$N = 45$			
K	Design	Power	Power†	$\Delta\,(\%)$	Design	Power	Power†	$\Delta\,(\%)$
3	$\begin{pmatrix} 1 & 2 & 3 \\ 8 & 14 & 8 \end{pmatrix}$	0.4969	0.4711	2.58	$\begin{pmatrix} 1 & 2 & 3 \\ 13 & 19 & 13 \end{pmatrix}$	0.7464	0.7282	1.82
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 6 & 9 & 9 & 6 \end{pmatrix}$	0.1650	0.1320	3.30	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 9 & 13 & 14 & 9 \end{pmatrix}$	0.4176	0.3743	3.33
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 7 & 8 & 7 & 4 \end{pmatrix}$	0.0269	0.0160	1.09	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 11 & 11 & 11 & 6 \end{pmatrix}$	0.1571	0.1234	3.37
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 5 & 5 & 5 & 5 & 5 \end{pmatrix}$	0.1835	0.1835	0.00	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 8 & 8 & 8 & 8 & 7 \end{pmatrix}$	0.5667	0.5413	2.54
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 4 & 5 & 4 & 5 & 4 & 4 \end{pmatrix}$	0.0403	0.0372	0.31	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 7 & 7 & 7 & 7 & 7 & 5 \end{pmatrix}$	0.3542	0.3177	3.65

† - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design.

Table 9 Optimal designs for simple order relation ($N = 60$, $\delta = \{0.9,\ 1.1\}$ for $K \in \{3,\ 4,\ 5\}$ and $\delta = \{1.4,\ 1.6\}$ for $K \in \{6,\ 7\}$).

	$\alpha = 0.05$				$\alpha = 0.025$			
K	Design	Power	Power†	Δ (%)	Design	Power	Power†	Δ (%)
3	$\begin{pmatrix} 1 & 2 & 3 \\ 17 & 25 & 18 \end{pmatrix}$	0.7876	0.7715	1.61	$\begin{pmatrix} 1 & 2 & 3 \\ 17 & 25 & 18 \end{pmatrix}$	0.9404	0.9333	0.71
4	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.4772	0.4427	3.45	$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 12 & 18 & 18 & 12 \end{pmatrix}$	0.7697	0.7510	1.87
5	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 9 & 14 & 14 & 14 & 9 \end{pmatrix}$	0.2006	0.1651	3.55	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 9 & 14 & 14 & 14 & 9 \end{pmatrix}$	0.5153	0.4848	3.05
	$\delta = 1.4$				$\delta = 1.6$			
6	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	0.7042	0.6853	1.89	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 11 & 11 & 11 & 11 & 8 \end{pmatrix}$	0.8794	0.8711	0.83
7	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 9 & 9 & 10 & 7 \end{pmatrix}$	0.5003	0.4759	2.44	$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 9 & 9 & 9 & 9 & 10 & 7 \end{pmatrix}$	0.7501	0.7328	1.73

† - Power of the corresponding balanced design; Δ - percentage increase of the power of IUT-based design.