Optimal designs for multiple treatments with unequal variances

Yaping Wang, Mingyao Ai*

LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China

Abstract

The response of a patient in a clinical trial usually depends on both the selected treatment and some latent covariates, while its variance varies across the treatment groups. A general heteroscedastic linear additive model incorporating the treatment effect and the covariate effects is often used in such studies. In this paper, under D- and D_A-optimality criteria, it is shown that the product of an optimal treatment allocation and an optimal design for covariates is also optimal among all possible designs for this linear additive model. Moreover, the optimal treatment allocation is characterized by a unique set of solutions to a system of equations. The connection between D- and D_A-optimal designs is also revealed. Several examples are presented to illustrate the applications of the above results to selected models.

Keywords: Optimal treatment allocation, D-optimality, D_A-optimality, Treatment contrast, Product design, Variance heterogeneity.

1. Introduction

Consider a K-treatment ($K \geq 2$) experiment consisting of a set of independent runs, where in each run one treatment is assigned. Suppose the mean value of the response of each run is determined by the effect of the chosen treatment $t \in \mathcal{T} = \{1, \ldots, K\}$ and also by the effects of m covariates $\boldsymbol{z} = (z_1, \ldots, z_m)^T \in \mathcal{Z}$, where $\mathcal{Z}$ is a compact subset of $\mathbb{R}^m$. The variance of the response varies across the treatment groups and depends only on t. Let $\boldsymbol{f}(\boldsymbol{z}) = (f_1(\boldsymbol{z}), \ldots, f_J(\boldsymbol{z}))^T$ denote a vector of J regression functions defined

*Corresponding author. Email: myai@math.pku.edu.cn.

on $\mathcal{Z}$ satisfying $\{1, f_1(\boldsymbol{z}), \ldots, f_J(\boldsymbol{z})\}$ is a linearly independent set. Then the heteroscedastic linear additive model is

$$y(t, \boldsymbol{z}) = \alpha_t + \sum_{j=1}^{J} \gamma_j f_j(\boldsymbol{z}) + \sigma_t \varepsilon, \tag{1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)^T$ are the vectors of treatment effects and covariate effects, respectively. The unequal variances $\sigma_1^2, \ldots, \sigma_K^2$ are assumed to be known and positive, and ε's are independent random variables, each with mean 0 and unit variance.

For simplicity, rewrite model (1) as $y(t, \boldsymbol{z}) = \boldsymbol{\beta}^T \boldsymbol{g}(\boldsymbol{x}) + \sigma_t \varepsilon$, where $\boldsymbol{x} = (t, \boldsymbol{z}) \in \mathcal{X}$, $\mathcal{X} = \mathcal{T} \times \mathcal{Z}$, $\boldsymbol{g}(t, \boldsymbol{z}) = (\boldsymbol{e}_{K,t}^T, \boldsymbol{f}^T(\boldsymbol{z}))^T$ and $\boldsymbol{\beta} = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$. Here $\boldsymbol{e}_{K,t}$ is the vector of length K with its t-th entry equal to one and all other entries equal to zero. Throughout all designs will be treated as approximate designs, i.e., probability measures on the design region with finite support points. A center problem is to find optimal designs for model (1) under some optimality criterion. When there is no covariate effects in model (1), Wong and Zhu (2008), Sverdlov and Rosenberger (2013) obtained optimal treatment allocation designs for different inferential purposes. Recently, Atkinson (2015) studied D- and D_A-optimal designs for model (1) with $K = 2$, $J = m$, $f_j(\boldsymbol{z}) = z_j$ and $\mathcal{Z} = [0, 1]^m$, i.e., only the treatment effects and all the linear main effects of m continuous covariates are considered.

The aim of this paper is to generalize the work of Atkinson (2015), by providing a theoretical insight into the design optimality for the general model (1) with multiple treatments. We note that model (1) can be regarded as a multi-factor model, for which optimal designs are usually obtained by the method of product design. See Schwabe (1996), Rodríguez and Ortiz (2005) and Graßhoff et al. (2007) for examples. We will show certain product design is D- or D_A-optimal for model (1) and present a further investigation of the optimal treatment allocation rules.

The remainder of this paper will unfold as follows. Section 2 proves that the product of an optimal treatment allocation and an optimal design for covariates is D-optimal for model (1). The characterization for the optimal treatment allocation of any D-optimal design is established, some numerical results are also presented. When the goal is to estimate some treatment contrasts and certain covariate effects, parallel results are obtained with respect to D_A-optimality in Section 3. Moreover, the connection between the two optimal treatment allocations under D- and D_A-optimality criteria is built. Applications of the theories to selected models are given in Section 4. Section 5 concludes this paper with some remarks.

2. $\boldsymbol{D}$-optimal designs for model (1)

For model (1), the information matrix of a given design ξ on $\mathcal{X}$ is

$$M(\xi) = \int_{\mathcal{X}} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}^T(\boldsymbol{x})/\sigma_t^2 \,\mathrm{d}\xi. \tag{2}$$

Define $\Xi = \{\xi \mid \det M(\xi) > 0\}$, i.e., the set of all designs on $\mathcal{X}$ with non-singular information matrix. Typically we are going to find optimal designs over Ξ which maximize some concavity criterion function of the information matrix, see Pukelsheim (2006) and Atkinson et al. (2007) for examples. A design is said to be D-optimal for model (1) if it maximizes $\det M(\xi)$ over Ξ. Any D-optimal design minimizes the volume of the confidence ellipsoid for $\boldsymbol{\beta}$, the vector of total unknown parameters in model (1).

The D-optimal designs found by Atkinson (2015) are essentially special product designs (see Example 2 in Section 4). In this section a further characterization of D-optimal designs for the more general linear model (1) with multiple treatments will be presented by using the techniques in the theory of optimal product designs.

Firstly, in addition to the full model (1) we consider two reduced marginal models: the heteroscedastic one-way layout for treatment effects

$$y_1(t) = \alpha_t + \sigma_t\varepsilon, \tag{3}$$

and the homoscedastic marginal model for covariate effects with an explicit intercept term

$$y_2(\boldsymbol{z}) = \gamma_0 + \sum_{j=1}^{J} \gamma_j f_j(\boldsymbol{z}) + \varepsilon. \tag{4}$$

Let ξ_1 and ξ_2 be designs on $\mathcal{T}$ and $\mathcal{Z}$, respectively, i.e., ξ_1 is a treatment allocation design and ξ_2 is a design for covariates. Since a treatment allocation design always provides K nonnegative weights $w_1, \ldots, w_K$ for the K treatments with $\sum_{k=1}^K w_k = 1$, ξ_1 can be equivalently described by a $K \times 1$ vector of weights $\boldsymbol{w} = (w_1, \ldots, w_K)^T$. For the two marginal models, the corresponding information matrices of ξ_1 and ξ_2 are

$$M_1(\xi_1) = \operatorname{diag}\left\{w_1\sigma_1^{-2}, \ldots, w_K\sigma_K^{-2}\right\}$$

and

$$M_2(\xi_2) = \begin{pmatrix} 1 & \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\,\mathrm{d}\xi_2 \\ \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\,\mathrm{d}\xi_2 & \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\boldsymbol{f}^T(\boldsymbol{z})\,\mathrm{d}\xi_2 \end{pmatrix},$$

respectively.

Given ξ_1 on $\mathcal{T}$ and ξ_2 on $\mathcal{Z}$, the product design is defined as the product measure $\xi_1 \otimes \xi_2$ on $\mathcal{X} = \mathcal{T} \times \mathcal{Z}$. Hence $\xi_1 \otimes \xi_2$ assigns the weight $\xi_1(t)\xi_2(\boldsymbol{z})$ to every point $(t, \boldsymbol{z})$ in the Cartesian product of the supports of ξ_1 and ξ_2. And the information matrix (2) of $\xi_1 \otimes \xi_2$ for model (1) can be rewritten as

$$M(\xi_1 \otimes \xi_2) = \begin{pmatrix} M_{11}(\xi_1 \otimes \xi_2) & M_{12}(\xi_1 \otimes \xi_2) \\ M_{12}^T(\xi_1 \otimes \xi_2) & M_{22}(\xi_1 \otimes \xi_2) \end{pmatrix}, \tag{5}$$

where

$$\begin{aligned} M_{11}(\xi_1 \otimes \xi_2) &= M_1(\xi_1), \\ M_{12}(\xi_1 \otimes \xi_2) &= (w_1\sigma_1^{-2}, \ldots, w_K\sigma_K^{-2})^T \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\,\mathrm{d}\xi_2, \\ M_{22}(\xi_1 \otimes \xi_2) &= \Big(\sum_{k=1}^{K} w_k\sigma_k^{-2}\Big) \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\boldsymbol{f}^T(\boldsymbol{z})\,\mathrm{d}\xi_2. \end{aligned}$$

Furthermore, we suppose that $M_1(\xi_1)$ and $M_2(\xi_2)$ are non-singular. Then the formula for the determinant of a partitioned matrix (see, e.g., Lemma A.2 in Schwabe (1996)) yields

$$\det M(\xi_1 \otimes \xi_2) = \Big(\sum_{k=1}^{K} w_k\sigma_k^{-2}\Big)^J \det M_1(\xi_1) \det M_2(\xi_2) > 0,$$

which means $\xi_1 \otimes \xi_2 \in \Xi$.

The theorem below shows that when a product design is D-optimal for model (1) and gives the optimal treatment allocation, which covers the related results of Atkinson (2015) as special cases. The proofs of all theorems and corollaries are given in Appendix A.

Theorem 1. *Let ξ_1^* be the design on $\mathcal{T}$ with the vector of weights $\boldsymbol{w}^* = (w_1^*, \ldots, w_K^*)^T$ that solves the following K equations*

$$w_k^{-1} + J\Big(\sigma_k^2 \sum_{t=1}^{K} w_t \sigma_t^{-2}\Big)^{-1} = K + J, \ \textit{for } k = 1, \ldots, K. \tag{6}$$

Let ξ_2^* be a D-optimal design on $\mathcal{Z}$ for the marginal model (4). Then the product design $\xi_1^* \otimes \xi_2^*$ is D-optimal for model (1).

Using Theorem 1, we can obtain a product D-optimal design for model (1) once a D-optimal design for the marginal model (4) is provided. However, not all the D-optimal designs for model (1) have the product structure. For example, sometimes it would be desirable to construct optimal designs with minimal number of support points. Typically they are not product designs. For any design ξ on $\mathcal{X} = \mathcal{T} \times \mathcal{Z}$, define the marginal design of ξ on $\mathcal{T}$ by $\xi_{(1)} = \int_{\mathcal{Z}} \mathrm{d}\xi$. Then $\xi_{(1)}$ is a treatment allocation design. The following theorem tells us that although a D-optimal design may not be uniquely determined, the optimal marginal treatment allocation is.

Theorem 2. *Let ξ^* be any D-optimal design for model (1). Then the marginal treatment allocation design $\xi_{(1)}^*$ has the vector of weights $\boldsymbol{w}^*$ which is the unique set of solutions of (6) in Theorem 1 with the constrain $0 < w_k^* < 1$ for $1 \le k \le K$.*

It follows from Theorems 1 and 2 that for model (1), regardless of the choice of regression functions in $\boldsymbol{f}(\boldsymbol{z})$, the optimal allocation for the K treatments under D-criterion is unique, and only depends on the number J of regression functions for covariates and the ratios of variances $\sigma_2^2/\sigma_1^2, \ldots, \sigma_K^2/\sigma_1^2$.

To obtain $\boldsymbol{w}^*$ in Theorem 1, the nonlinear system of K equations (6) can be solved by the following procedure. Comparing the last $K-1$ equations with the first yields

$$\sigma_k^2 w_k^{-1} = \sigma_1^2 w_1^{-1} + (K+J)(\sigma_k^2 - \sigma_1^2), \ \text{for } k = 2, \ldots, K. \tag{7}$$

Substituting them into the first equation in (6), we get a univariate equation

$$\sigma_1^2 w_1^{-1} + J\Big(w_1\sigma_1^{-2} + \sum_{k=2}^{K} \left[\sigma_1^2 w_1^{-1} + (K+J)(\sigma_k^2 - \sigma_1^2)\right]^{-1}\Big)^{-1} - (K+J)\sigma_1^2 = 0, \tag{8}$$

which is essentially a polynomial of w_1 whose roots can be obtained analytically for small K or numerically for large K. After w_1^* is found, $w_2^* \ldots, w_K^*$ are obtained immediately from (7). The optimization problem is, thus, reduced to a one-dimensional root finding problem which can be solved easily. In addition, from the above solving process the following corollary can be derived.

Corollary 1. *Consider the optimal weights for the K treatments $w_1^*, \ldots, w_K^*$ in Theorem 1.*
(i) If $\sigma_{k_1}^2 = \sigma_{k_2}^2$, $1 \leq k_1 < k_2 \leq K$, then $w_{k_1}^* = w_{k_2}^*$. In particular, if all the variances $\sigma_1^2, \ldots, \sigma_K^2$ are equal, then $w_1^* = \ldots = w_K^* = 1/K$.
(ii) If all the variances take only two different values, say σ_a^2 and σ_b^2, with $\sigma_b^2/\sigma_a^2 = \tau \neq 1$. Without loss of generality, assume $\sigma_1^2 = \ldots = \sigma_{K_1}^2 = \sigma_a^2$, and $\sigma_{K_1+1}^2 = \ldots = \sigma_K^2 = \sigma_b^2$, where $1 \leq K_1 < K$, then $w_1^* = \ldots = w_{K_1}^* = w_a^*/K_1$ and $w_{K_1+1}^* = \ldots = w_K^* = (1 - w_a^*)/(K - K_1)$. Here w_a^* is the unique root of the following quadratic equation

$$(K + J)(1 - \tau)w_a^2 - [(K_1 + J)(1 - \tau) + K]w_a + K_1 = 0 \tag{9}$$

satisfying $0 < w_a^* < 1/K_1$.

Corollary 1 implies that the optimal weights are same for all the treatments with same variance. When all the K variances are equal, model (1) degenerates to a homoscedastic model, and in such case the equal allocation is always optimal. Note that when $K = 2$ and $K_1 = 1$, equation (9) becomes equation (4) in Section 3 of Atkinson (2015) for the case of two treatments. Furthermore, from the quadratic equation (9) we get the lower and upper bounds of the optimal weight w_1^* for extreme values of τ: $w_1^* \to 1/(K + J)$ as $\tau \to 0$ and $w_1^* \to (K_1 + J)/[K_1(K + J)]$ as $\tau \to +\infty$. The two bounds depend on J, the number of regression functions for covariates, and approach to 0 and $1/K_1$ respectively as J increases to infinity.

For convenience of use, we provide an R code based on the package rootSolve as supplementary material in Appendix B. The code can return the desired treatment allocation design under D- or D_A-optimality criterion once parameters are inputted. Also, for selected numbers J of covariate effects and the variances ratios $\sigma_1^2 : \sigma_2^2 : \ldots : \sigma_K^2$, optimal treatment allocations are presented in Tables 1 and 2, respectively, for $K = 3$ and $K = 4$. The corresponding result for $K = 2$ can be found in Table 1 of Atkinson (2015). These numerical results show that equal weights are assigned to the

Table 1: Optimal treatment weights w_1^*, w_2^*, w_3^* for $K = 3$ in (6)

Variances	Number J of covariate effects					
$\sigma_1^2 : \sigma_2^2 : \sigma_3^2$	1	2	4	6	8	10
$1:1:1$	.333 .333 .333	.333 .333 .333	.333 .333 .333	.333 .333 .333	.333 .333 .333	.333 .333 .333
$1:1:2$	.354 .354 .293	.370 .370 .260	.395 .395 .210	.412 .412 .175	.425 .425 .150	.435 .435 .131
$1:1:4$	.364 .364 .271	.386 .386 .227	.415 .415 .171	.432 .432 .136	.443 .443 .113	.452 .452 .097
$1:1:8$	.370 .370 .261	.393 .393 .213	.422 .422 .156	.439 .439 .123	.450 .450 .101	.457 .457 .086
$1:1:16$	.372 .372 .255	.397 .397 .206	.425 .425 .149	.442 .442 .117	.452 .452 .096	.459 .459 .081
$1:1:\frac{1}{2}$	.305 .305 .390	.276 .276 .447	.227 .227 .547	.189 .189 .623	.160 .160 .679	.139 .139 .722
$1:1:\frac{1}{4}$	.280 .280 .440	.237 .237 .527	.177 .177 .645	.141 .141 .715	.117 .117 .767	.099 .099 .801
$1:1:\frac{1}{8}$	.266 .266 .469	.218 .218 .565	.159 .159 .683	.124 .124 .751	.102 .102 .795	.087 .087 .826
$1:1:\frac{1}{16}$	.258 .258 .485	.209 .209 .582	.150 .150 .699	.117 .117 .765	.096 .096 .807	.082 .082 .836
$1:2:4$	.412 .311 .277	.483 .283 .234	.593 .230 .176	.669 .191 .140	.722 .161 .116	.761 .140 .099
$1:2:8$	.422 .314 .263	.498 .285 .216	.610 .232 .158	.685 .191 .124	.736 .162 .102	.773 .140 .087
$1:2:16$	.428 .316 .257	.506 .287 .208	.618 .232 .150	.691 .191 .117	.742 .162 .096	.778 .140 .082
$1:4:8$	.454 .282 .265	.545 .238 .217	.664 .178 .158	.735 .141 .124	.781 .117 .102	.814 .099 .087
$1:4:16$	.460 .282 .257	.554 .238 .208	.672 .178 .150	.742 .141 .117	.787 .117 .096	.819 .099 .082
$1:8:16$	.477 .266 .258	.574 .218 .208	.691 .159 .150	.758 .124 .117	.802 .102 .096	.832 .087 .082

Table 2: Optimal treatment weights $w_1^*, w_2^*, w_3^*, w_4^*$ for $K = 4$ in (6)

Variances	Number J of covariate effects			
$\sigma_1^2 : \sigma_2^2 : \sigma_3^2 : \sigma_4^2$	1	3	5	9
$1:1:1:1$	.250 .250 .250 .250	.250 .250 .250 .250	.250 .250 .250 .250	.250 .250 .250 .250
$1:1:1:2$	.258 .258 .258 .225	.271 .271 .271 .187	.280 .280 .280 .159	.293 .293 .293 .122
$1:1:1:4$	.262 .262 .262 .213	.279 .279 .279 .163	.290 .290 .290 .131	.302 .302 .302 .095
$1:1:1:\frac{1}{2}$	.237 .237 .237 .290	.207 .207 .207 .378	.179 .179 .179 .463	.136 .136 .136 .592
$1:1:1:\frac{1}{4}$	.222 .222 .222 .333	.173 .173 .173 .479	.139 .139 .139 .582	.099 .099 .099 .703
$1:1:2:2$	.270 .270 .230 .230	.305 .305 .195 .195	.333 .333 .167 .167	.373 .373 .128 .128
$1:1:4:4$	.284 .284 .216 .216	.333 .333 .167 .167	.365 .365 .135 .135	.404 .404 .096 .096
$1:1:8:8$	.292 .292 .208 .208	.346 .346 .154 .154	.378 .378 .122 .122	.414 .414 .086 .086
$1:1:2:4$	.277 .277 .232 .215	.319 .319 .197 .166	.349 .349 .169 .134	.388 .388 .128 .096
$1:1:2:8$	.280 .280 .233 .207	.324 .324 .198 .154	.355 .355 .169 .122	.393 .393 .129 .086
$1:1:4:8$	.288 .288 .217 .208	.339 .339 .167 .154	.372 .372 .135 .122	.409 .409 .097 .086
$1:2:4:8$	.322 .247 .221 .210	.454 .217 .172 .156	.553 .185 .139 .123	.677 .138 .099 .087
$1:2:4:16$	.326 .248 .221 .205	.460 .218 .173 .149	.559 .185 .139 .117	.681 .138 .099 .081
$1:2:8:16$	.334 .250 .211 .205	.474 .220 .157 .149	.573 .186 .124 .117	.694 .138 .087 .081
$1:4:8:16$	.358 .225 .212 .206	.519 .174 .157 .150	.619 .140 .124 .117	.733 .099 .087 .081

treatment groups having equal variances. When variances are different, the smaller the variance is, the larger the optimal weight will be. Moreover, the treatment allocation becomes more skewed as J increases when the ratios of variances are fixed.

3. D_A-optimal designs for model (1)

If our primary interest is only in some linear combinations of parameters in a model, D_A-optimality can be used instead of D-optimality (Silvey (1980); Atkinson et al. (2007)). Let A be a $(K+J)\times s$ matrix with full column rank s. A design is said to be D_A-optimal for model (1) if it maximizes $\det[(A^T M^{-1}(\xi)A)^{-1}]$ over Ξ. Note that when $s = K+J$, i.e., A is a non-singular square matrix, D_A-optimality is equivalent to D-optimality since $\det[(A^T M^{-1}(\xi)A)^{-1}] = \det M(\xi)/\det(A^T A)$.

Throughout this section we discuss D_A-optimal designs for model (1) in the following scenario. Suppose we aim to estimate a system of s_1 ($s_1 < K$) treatment contrasts $A_1^T\boldsymbol{\alpha}$ and a set of s_2 ($s_2 \le J$) linear combinations of covariate effects $A_2^T\boldsymbol{\gamma}$, where A_1 is a $K\times s_1$ matrix with full rank s_1 satisfying $A_1^T\mathbf{1}_K = \mathbf{0}_{s_1}$, and A_2 is a $J\times s_2$ matrix with full rank s_2. Here $\mathbf{1}_n$ and $\mathbf{0}_n$ are column vectors of n ones and zeros, respectively. When $s_2 = J$, A_2 is a non-singular square matrix, and we can without loss of generality let $A_2 = I_J$. Moreover, each row of A_1 is assumed to be nonzero, i.e., we are interested in all treatments in $\mathcal{T}$. Then the linear combinations of parameters $A^T\boldsymbol{\beta}$ depend separately on $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, where

$$A = \begin{pmatrix} A_1 & \mathbf{0}_{K\times s_2} \\ \mathbf{0}_{J\times s_1} & A_2 \end{pmatrix}$$

with rank $s = s_1 + s_2$, and $\mathbf{0}_{m\times n}$ is an $m\times n$ matrix of zeros. In such cases, the following theorem presents a characterization of D_A-optimal designs for model (1).

Theorem 3. *Define an auxiliary optimality criterion for designs on $\mathcal{T}$ with non-singular information matrix for the marginal model (3)*

$$\Phi_{\lambda,A}(\xi_1) = s_2 \ln\left(\sum_{k=1}^{K} w_k\sigma_k^{-2}\right) - \ln\det[(A_1^T M_1^{-1}(\xi_1)A_1)].$$

(i) Let ξ_1^* on $\mathcal{T}$ maximize $\Phi_{\lambda,A}$ and ξ_2^* be a D_{A_3}-optimal (in particular, D-optimal when $s_2 = J$) design on $\mathcal{Z}$ for the marginal model (4), where $A_3 = (\mathbf{0}_{s_2\times 1}, A_2^T)^T$. Then the product design $\xi_1^* \otimes \xi_2^*$ is D_A-optimal for model (1).
(ii) For any D_A-optimal design ξ^* for model (1), the marginal treatment allocation design $\xi_{(1)}^*$ maximize $\Phi_{\lambda,A}$.

In many experiments, especially in clinical trials, there has been a ubiquitous interest in the comparison of treatments with a control. For example, in a clinical trial it is common to set a placebo group and some treatment groups which are expected to provide improvements over the placebo group. Besides, by the similar arguments in Atkinson (2015) for personalized medicine, the estimation of some linear combinations of the covariate effects is also of interest. Without loss of generality, suppose the first treatment is the control, then the above problem can be formulated by estimating $\boldsymbol{A}^T\boldsymbol{\beta}$ with $A_1 = \big(-\mathbf{1}_{(K-1)\times 1}, I_{K-1}\big)^T$, where I_n is the identity matrix of order n. The optimal weights for the K treatments can be obtained by applying Theorem 3.

Corollary 2. *In Theorem 3, suppose $A_1 = \big(-\mathbf{1}_{(K-1)\times 1}, I_{K-1}\big)^T$, then all the D_A-optimal designs for model (1) share the same marginal treatment allocation design ξ_1^*, whose vector of weights $\boldsymbol{w}^* = (w_1^*, \ldots, w_K^*)^T$ is the unique set of solutions of the following K equations*

$$w_k^{-1} + (s_2 - 1)\Big(\sigma_k^2 \sum_{t=1}^{K} w_t \sigma_t^{-2}\Big)^{-1} = K + s_2 - 1, \ \textit{for } k = 1, \ldots, K \qquad (10)$$

with the constrain $0 < w_k^* < 1$ for $1 \le k \le K$.

Comparing Theorem 1 and Corollary 2, we derive an interesting and somewhat surprising property of the optimal treatment allocation rules. The equations in (6) turn to the equations in (10) just with J replaced by $s_2 - 1$. For illustration, let $A_2 = I_J$ and thus $s_2 = J$. Then the optimal treatment allocation design for a model incorporating J covariate effects under D_A-criterion with $A_1 = \big(-\mathbf{1}_{(K-1)\times 1}, I_{K-1}\big)^T$, is exactly the same as the optimal treatment allocation design for a model incorporating one less covariate effect under D-criterion. Especially, if $J = 1$, the equal allocation is optimal under D_A-criterion. Hence we establish the connection between the optimal treatment allocation rules under D- and D_A-optimality criteria.

In addition, parallel to Corollary 1, we have the following.

Corollary 3. *Consider the optimal weights for the K treatments $w_1^*, \ldots, w_K^*$ in Corollary 2.*
(i) If $\sigma_{k_1}^2 = \sigma_{k_2}^2$, $1 \le k_1 < k_2 \le K$, then $w_{k_1}^* = w_{k_2}^*$. In particular, if all the variances $\sigma_1^2, \ldots, \sigma_K^2$ are equal, then $w_1^* = \ldots = w_K^* = 1/K$.
(ii) If all the variances take only two different values, say σ_a^2 and σ_b^2, with $\sigma_b^2/\sigma_a^2 = \tau \neq 1$. Without loss of generality, assume $\sigma_1^2 = \ldots = \sigma_{K_1}^2 = \sigma_a^2$, and $\sigma_{K_1+1}^2 = \ldots = \sigma_K^2 = \sigma_b^2$, where $1 \le K_1 < K$, then $w_1^* = \ldots = w_{K_1}^* = w_a^*/K_1$ and $w_{K_1+1}^* = \ldots = w_K^* = (1 - w_a^*)/(K - K_1)$. Here w_a^* is the unique root of the quadratic equation (9) with J replaced by $s_2 - 1$ satisfying $0 < w_a^* < 1/K_1$.

Note that if we are only interested in the treatment contrasts $A_1^T \boldsymbol{\gamma}$ without any covariate effects in model (1), i.e., $A = (A_1^T, \mathbf{0}_{s_1 \times J})^T$, the conclusions in Theorem 3, Corollaries 2 and 3 still apply with $s_2 = 0$. In such cases, the D_A-optimality of a design for model (1) relates only to the marginal weights for the K treatments, and Corollaries 2 and 3 are consistent with the results of Wong and Zhu (2008). In fact, some stronger results can be obtained. We refer to a recent work by Rosa and Harman (2015) where the optimal designs for estimating treatment contrasts in a model like (1) but with homoscedastic errors were studied systematically, with respect to a wide range of optimality criteria. Let ϕ be an information function (see Section 5.8 of Pukelsheim (2006)) defined on the set of s_1-order non-negative definitive matrices. A design $\xi^* \in \Xi$ is said to be ϕ_A-optimal for model (1) if it maximizes $\phi\left([A^T M^{-1}(\xi) A]^{-1}\right)$. In particular, when $\phi = \ln\det(\cdot)$ the definition reduces to D_A-optimality. We conclude this section by providing a theorem which can be viewed as extensions of Theorems 1 and 2 in Rosa and Harman (2015), for model (1).

Theorem 4. *Let $A = (A_1^T, \mathbf{0}_{s_1 \times J})^T$ and ϕ be an information function.*
(i) If ξ_1^* on $\mathcal{T}$ is ϕ_{A_1}-optimal for the marginal model (3) and ξ_2 is any design on $\mathcal{Z}$ with non-singular information matrix for the marginal model (4), then the product design $\xi_1^* \otimes \xi_2$ is ϕ_A-optimal for model (1).
(ii) For any ϕ_A-optimal design ξ^* for model (1), the marginal treatment allocation design $\xi_{(1)}^*$ is ϕ_{A_1}-optimal for the marginal model (3).

4. Applications

The results in the previous sections can be applied to determine the optimal designs for some commonly used models.

Example 1. Suppose there is one qualitative covariate, say z, taking levels in $\mathcal{Z} = \{1, \ldots, L\}$. The model, which is essentially a two-way layout without interactions, can be written as

$$y(t, z) = \alpha_t + \gamma_z + \sigma_t \varepsilon,$$

where $z \in \{1, \ldots, L\}$ and $\gamma_1, \ldots, \gamma_L$ are the effects of the covariate.

For the identifiability of the parameters, we request a baseline constrain $\gamma_1 = 0$. Then the model belongs to the family of (1) with $\boldsymbol{f}(z) = \boldsymbol{e}_{L-1,z}$. The marginal model (3) for covariate effects is a one-way layout with an explicit intercept term, for which the D-optimal design ξ_2^* on $\mathcal{L}$ assigns equal weights $1/L$ to each level. Hence by Theorem 1, the product design $\xi_1^* \otimes \xi_2^*$ which assigns weights w_k^*/L to each of the level combinations (k, ℓ), $k = 1, \ldots, K$, $\ell = 1, \ldots, L$, is D-optimal for the full model, where $\boldsymbol{w}^* = (w_1^*, \ldots, w_K^*)^T$ is the unique set of solutions of (6) with $J = L - 1$. When the estimation of treatment comparisons and all the covariate effects is of interest, it follows from Corollary 2 that $\tilde{\xi}_1^* \otimes \xi_2^*$ is D_A-optimal, where $\tilde{\boldsymbol{w}}^* = (\tilde{w}_1^*, \ldots, \tilde{w}_K^*)^T$ solves (6) with J replaced by $L - 2$. If we are only concern with treatment comparisons without any covariate effects, by Corollary 2 and Theorem 4, all D_A-optimal designs share the same marginal treatment allocation design whose vector of weights $\boldsymbol{w}^*$ solves (6) with $J = -1$.

Example 2. Consider model (1) with $J = m$, $f_j(\boldsymbol{z}) = z_j$ and $\mathcal{Z} = [0, 1]^m$, i.e., the model incorporating all the linear main effects of m continuous covariates

$$y(t, \boldsymbol{z}) = \alpha_t + \sum_{j=1}^{m} \gamma_j z_j + \sigma_t \varepsilon.$$

When $K = 2$, this is just the model in Section 2 of Atkinson (2015). The marginal model (3) for covariate effects is an ordinary linear regression model. Assume $A_1 = \left(-\mathbf{1}_{(K-1)\times 1}, I_{K-1}\right)^T$ and $A_2 = I_m$. To get product D- or D_A-optimal designs by using Theorem 1 or Corollary 2, one can choose ξ_2^* to be an m-factor two-level orthogonal array (OA) of strength two with $\{\pm 1\}$ levels (see e.g., Hedayt et al. (1999)), i.e., ξ_2^* assigns its weights to each row of the orthogonal array uniformly. In fact, Atkinson's choice is two level full or fractional factorials, which are special OAs. The optimal treatment allocation design ξ_1^* is uniquely determined by (6) with $J = m$ or $m - 1$, for D- or D_A-criterion, respectively.

If $K = 2$ and $\sigma_1^2 = \sigma_2^2$, then the optimal treatment weights are $w_1^* =$

$w_2^* = 1/2$ by Corollary 1. In such a case, a minimal size $(m+1)$-factor two-level OA is D- and D_A-optimal, where the treatment variable is treated as a two-level factor. Typically it is not a product design, but the marginal treatment weights are always $w_1^* = w_2^* = 1/2$, which provides an illustration of Theorem 2 and Theorem 3 (ii).

Example 3. In Example 2, if all the two-factor interactions of the m covariates are non-neglectable, the model becomes

$$y(t, \boldsymbol{z}) = \alpha_t + \sum_{j=1}^{m} \gamma_j z_j + \sum_{j_1=1}^{m-1} \sum_{j_2=j_1+1}^{m} \gamma_{j_1,j_2} z_{j_1} z_{j_2} + \sigma_t \varepsilon.$$

Now the total number of covariate effects is $J = m(m+1)/2$. An m-factor two-level OA of strength r can be chosen as ξ_2^*, where $r = 4$ for $m \geq 4$ and $r = m$ for $m < 4$. Then product D- and D_A-optimal designs can be obtained similarly.

Note that if for the covariates the design region is $\mathcal{Z} = \{\sum_{j=1}^{m} z_j = 1,\, z_j \geq 0\}$, i.e., the $(m-1)$-order simplex, then the above model is the analogue of second-order Scheffé polynomial for mixture experiments. Again the product D- and D_A-optimal designs can be obtained, see Donev (1989) for example.

Example 4. Consider the following model

$$y = \alpha_t + \gamma_1 z + \gamma_2 z^2 + \sigma_t \varepsilon,$$

where $z \in \mathcal{Z} = [-1, 1]$, γ_1 and γ_2 are linear and quadratic effect of the covariate.

The marginal model (4) is a quadratic regression model in one variable, for which the D-optimal design ξ_2^* with the minimal number of support points consists of weights 1/3 at -1, 0 and 1 (see e.g., Section 9.2 of Atkinson et al. (2007)). By Theorem 1, the product design $\xi_1^* \otimes \xi_2^*$ is D-optimal for the full model, where ξ_1^* is uniquely determined by (6) with $J = 2$.

If we are only interested in estimating treatment comparisons and the quadratic covariate effect γ_2, by letting $A_1 = \left(-\mathbf{1}_{(K-1)\times 1}, I_{K-1}\right)^T$ and $A_2 = (0, 1)^T$, Theorem 3 can be applied. Let $A_3 = (0, 0, 1)^T$ and $\tilde{\xi_2}^*$ be a design which assigns weights 1/4, 1/2, 1/4 to -1, 0 and 1, respectively. Then $\tilde{\xi_2}^*$ is D_{A_3}-optimal for the marginal quadratic regression model (Section 10.3 of Atkinson et al. (2007)). By Corollary 2, the product design $\tilde{\xi_1}^* \otimes \tilde{\xi_2}^*$ is D_A-

optimal for the full model, where $\tilde{\xi}_1^*$ assigns its weights equally to the K treatments.

5. Concluding remarks

We generalize the related work of Atkinson (2015) to obtain D- and D_A-optimal optimal designs for a general heteroscedastic linear additive model (1) with multiple treatments and covariate effects. The product structure of optimal designs is systematically revealed and studied. Moreover, the optimal treatment allocations for different inferential purposes are also investigated. The results can be applied to construct optimal designs for specific models in many fields including clinical trials for personalized medicine, agricultural, environmental and industry experiments. Note that the optimal designs in this paper provide unbalanced optimal treatment allocations when the variances are unequal. They are more efficient than equal allocations or other designs where variance heterogeneity among treatment groups and covariate effects is ignored.

Before winding up this paper, it should be mentioned that in our setup, the error ε in model (1) is assumed to have variance 1. When ε has a general variance σ^2 and σ^2 is unknown, the variance of the outcome from the k-th group becomes $\sigma_k^2\sigma^2$ for $k = 1, \ldots, K$. Obviously, the ratio of variances remains unchanged. Therefore, all the results obtained in this paper still hold for this general case.

Acknowledgements

The work is partially supported by NSFC grants 11271032 and 11331011, and BCMIIS.

Appendix A. Proofs of all theorems and corollaries

Proof of Theorem 1

Note that although the model (1) that we are interested in does not contain an intercept term, we can transform it into an equivalent model with an intercept term by letting $\tilde{\alpha}_1 = \alpha_1$, $\tilde{\alpha}_k = \alpha_k - \alpha_1$, $k = 2, \ldots, K$, and $\boldsymbol{\gamma}$ remain unchanged. It is clear that the above reparametrization does not affect the D-optimality, i.e., a design is D-optimal for the transformed model if and only if it is D-optimal for the original model (1).

The transformed model belongs to the family of model (1) in Section 2.1 of Graßhoff et al. (2007) with $\boldsymbol{x}_1 = \boldsymbol{z}$, $\boldsymbol{x}_2 = t$, $\mathcal{X}_1 = \mathcal{Z}$, $\mathcal{X}_2 = \mathcal{T}$, $\boldsymbol{f}_1(\boldsymbol{x}_1) = \boldsymbol{f}(\boldsymbol{z})$, $\boldsymbol{f}_2(\boldsymbol{x}_2) = \boldsymbol{e}_{K-1,t-1}$ and $\lambda_2(\boldsymbol{x}_2) = \lambda_2(t) = 1/\sigma_t^2$. Here we let $\boldsymbol{e}_{K-1,0} = \boldsymbol{0}_{K-1}$ for ease of expression. The parameters are $\beta_0 = \tilde{\alpha}_1$, $\boldsymbol{\beta}_1 = \boldsymbol{\gamma}$ with dimension $p_1 = J$ and $\boldsymbol{\beta}_2 = (\tilde{\alpha}_2, \ldots, \tilde{\alpha}_K)^T$ with dimension $p_2 = K - 1$. Besides, for the reparametrized model, the corresponding marginal model for treatment effects is a reparametrization of model (3) with $\tilde{\alpha}_1 = \alpha_1$, $\tilde{\alpha}_k = \alpha_k - \alpha_1$, $k = 2, \ldots, K$, and this reparametrization does not change the determinant of the information matrix for any design ξ_1 on $\mathcal{T}$. And the marginal model for covariate effects coincides with model (4). Hence by Theorem 1 of Graßhoff et al. (2007), the product design $\xi_1^* \otimes \xi_2^*$ is D-optimal for model (1) if ξ_1^* maximizes the auxiliary criterion

$$\Phi_\lambda(\xi_1) = J \ln\Big(\sum_{k=1}^{K} w_k \sigma_k^{-2}\Big) + \ln\Big(\prod_{k=1}^{K} w_k \sigma_k^{-2}\Big)$$

and ξ_2^* is D-optimal for model (4). To maximize $\Phi_\lambda(\xi_1)$ as a function of $\boldsymbol{w}$ with the constrain $\sum_{k=1}^{K} w_k = 1$, the method of Lagrange multiplier can be applied. Let λ_0 be a Lagrange multiplier and

$$L(\boldsymbol{w}, \lambda_0) = \sum_{k=1}^{K} \ln w_k + J \ln\Big(\sum_{k=1}^{K} w_k \sigma_k^{-2}\Big) - \lambda_0\Big(\sum_{k=1}^{K} w_k - 1\Big).$$

Equating the partial derivatives of L with respect to $(\boldsymbol{w}^T, \lambda_0)^T$ to zero we have

$$\left\{ \begin{array}{c} w_k^{-1} + J\big(\sigma_k^2 \sum_{t=1}^{K} w_t \sigma_t^{-2}\big)^{-1} - \lambda_0 = 0, \text{ for } k = 1, \ldots, K, \\ \sum_{k=1}^{K} w_k - 1 = 0. \end{array} \right.$$

From the above equations we obtain $\lambda_0^* = \big(\sum_{k=1}^{K} w_k\big)\lambda_0^* = K + J$. Substituting λ_0^* into the first K equations we get (6). The conclusion follows. $\square$

Proof of Theorem 2

By the strict concavity of the transformed D-criterion function $\phi = \ln\det(\cdot)$ considered as a function of $M(\xi)$, the information matrix of a D-optimal design for model (1) is uniquely determined. Hence by Theorem 1, all the D-optimal designs for model (1) share the same information matrix $M(\xi_1^* \otimes \xi_2^*)$ which has the form of (5). Noting that $M_{11}(\xi_1^* \otimes \xi_2^*) =$

$M_1(\xi_1^*) = M_1(\xi_{(1)}^*)$ and $M_1(\xi_1^*) = \mathrm{diag}\left\{w_1^* \sigma_1^{-2}, \ldots, w_K^* \sigma_K^{-2}\right\}$, the vector of optimal weights $\boldsymbol{w}^*$ is uniquely determined and can be solved by the equations in (6). Hence, the conclusion follows. $\square$

Proof of Corollary 1

It follows directly from (7) that part (i) is true. For part (ii), first we have $w_1 = \ldots = w_{K_1}$ and $w_{K_1+1} = \ldots = w_K$ from (i). Define $w_a = K_1 w_1$, then $w_{K_1+1} = (1 - w_a)/(K - K_1)$ and equation (8) becomes the quadratic equation (9) after some algebra. The proof of Corollary 1 is complete. $\square$

Proof of Theorem 3

By assumption the product design $\xi_1^* \otimes \xi_2^*$ has non-singular information matrix for model (1). Applying the formula for the inverse of a partitioned matrix (see e.g., Lemma A.3 in Schwabe (1996)) and noting that

$$A_1^T M_1^{-1}(\xi_1^*) M_{12}(\xi_1^* \otimes \xi_2^*) = A_1^T \mathbf{1}_K \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\, \mathrm{d}\xi_2^* = \mathbf{0}_{s_1 \times J},$$

we have

$$A^T M^{-1}(\xi_1^* \otimes \xi_2^*) A = \begin{pmatrix} A_1^T M_1^{-1}(\xi_1^*) A_1 & \mathbf{0}_{s_1 \times s_2} \\ \mathbf{0}_{s_2 \times s_1} & \left(\sum_{k=1}^K w_k/\sigma_k^2\right)^{-1} A_2^T M_{22\cdot 1}^{-1}(\xi_2^*) A_2 \end{pmatrix},$$

where $M_{22\cdot 1}(\xi_2^*) = \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z}) \boldsymbol{f}^T(\boldsymbol{z})\, \mathrm{d}\xi_2^* - \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\, \mathrm{d}\xi_2^* \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\, \mathrm{d}\xi_2^*$. We also note that for the marginal model (4), the inverse of information matrix $M_2(\xi_2^*)$ can be written as

$$M_2^{-1}(\xi_2^*) = \begin{pmatrix} 1 + \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\, \mathrm{d}\xi_2^* \, M_{22\cdot 1}^{-1}(\xi_2^*) \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\, \mathrm{d}\xi_2^* & -\int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z})\, \mathrm{d}\xi_2^* \, M_{22\cdot 1}^{-1}(\xi_2^*) \\ -M_{22\cdot 1}^{-1}(\xi_2^*) \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z})\, \mathrm{d}\xi_2^* & M_{22\cdot 1}^{-1}(\xi_2^*) \end{pmatrix}.$$

Now we will show the D_A-optimality of $\xi_1^* \otimes \xi_2^*$ through the general equivalence theorem. For any $\boldsymbol{x} = (t, \boldsymbol{z}) \in \mathcal{X}$, denote by

$$d_A(\boldsymbol{x}, \xi) = \boldsymbol{g}^T(\boldsymbol{x}) M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) \boldsymbol{g}(\boldsymbol{x}) / \sigma^2(t).$$

the sensitivity function of model (1). The general equivalence theorem for D_A-optimality, similar to the D-optimality one (Kiefer and Wolfowitz (1960)), says that a design $\xi^* \in \mathit{\Xi}$ is D_A-optimal for model (1) if and only if $d_A(\boldsymbol{x}, \xi^*) \leq s$ for all $\boldsymbol{x} \in \mathcal{X}$, see e.g., Section 5.2 of Silvey (1980). Through a direct calculation, we derive

$$\begin{aligned}
d_A(\boldsymbol{x}, \xi_1^* \otimes \xi_2^*) = {} & \boldsymbol{e}_{K,t}^T M_1^{-1}(\xi_1^*) A_1 (A_1^T M_1^{-1}(\xi_1^*) A_1)^{-1} A_1^T M_1^{-1}(\xi_1^*) \boldsymbol{e}_{K,t} / \sigma_t^2 \\
& + \Big(\boldsymbol{f}^T(\boldsymbol{z}) - \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z}) \, \mathrm{d}\xi_2^* \Big) M_{22\cdot 1}^{-1}(\xi_2^*) A_2 (A_2^T M_{22\cdot 1}^{-1}(\xi_2^*) A_2)^{-1} \\
& \times A_2^T M_{22\cdot 1}^{-1}(\xi_2^*) \Big(\boldsymbol{f}(\boldsymbol{z}) - \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z}) \, \mathrm{d}\xi_2^* \Big) \Big/ \Big(\sigma_t^2 \sum_{k=1}^K w_k \sigma_k^{-2} \Big).
\end{aligned}$$

The general equivalence theorem for the D_{A_3}-optimal design ξ_2^* for the marginal model (4) yields

$$\begin{aligned}
\Big(\boldsymbol{f}^T(\boldsymbol{z}) - \int_{\mathcal{Z}} \boldsymbol{f}^T(\boldsymbol{z}) \, \mathrm{d}\xi_2^* \Big) & M_{22\cdot 1}^{-1}(\xi_2^*) A_2 (A_2^T M_{22\cdot 1}^{-1}(\xi_2^*) A_2)^{-1} \\
& \times A_2^T M_{22\cdot 1}^{-1}(\xi_2^*) \Big(\boldsymbol{f}(\boldsymbol{z}) - \int_{\mathcal{Z}} \boldsymbol{f}(\boldsymbol{z}) \, \mathrm{d}\xi_2^* \Big) \leq s_2
\end{aligned}$$

for all $\boldsymbol{z} \in \mathcal{Z}$. In particular, when $A_2 = I_J$ ($s_2 = J$) it is well known that D_{A_3}-optimality is equivalent to D-optimality for the marginal model (4), see e.g., Section 1.4.2 of Goos (2002). Furthermore, by the concavity of the auxiliary criterion $\Phi_{\lambda,A}$, the general equivalence theorem for the $\Phi_{\lambda,A}$-optimal design ξ_1^* yields

$$\boldsymbol{e}_{K,t}^T M_1^{-1}(\xi_1^*) A_1 (A_1^T M_1^{-1}(\xi_1^*) A_1)^{-1} A_1^T M_1^{-1}(\xi_1^*) \boldsymbol{e}_{K,t} / \sigma_t^2 + s_2 \Big/ \Big(\sigma_t^2 \sum_{k=1}^K w_k \sigma_k^{-2} \Big) \leq s_1 + s_2$$

for all $t \in \boldsymbol{T}$ (see e.g., Section 3.6 of Silvey (1980)). Thus we have $d_A(\boldsymbol{x}, \xi_1^* \otimes \xi_2^*) \leq s_1 + s_2$ for all $\boldsymbol{x} \in \mathcal{X}$, which proves the part (i).

Part (ii) follows by the strict concavity of $\ln\det(\cdot)$ as a function on the set $\{C = (A^T M^{-1}(\xi) A)^{-1} \mid \xi \in \Xi\}$ and a similar argument as in the proof of Theorem 2. □

Proof of Corollary 2

By adopting the same procedure for maximizing $\Phi_{\lambda,A}(\xi_1)$ as in the proof of Theorem 1, the equations in (10) can be obtained. Moreover, for any D_A-optimal design ξ^* for model (1) with marginal treatment allocation design

$\xi^*_{(1)}$, since

$$\begin{aligned} A_1^T M_1^{-1}(\xi^*_{(1)})A_1 &= A_1^T M_1^{-1}(\xi^*_1)A_1 \\ &= (\sigma_1^2/w_1^*)\mathbf{1}_{K-1}\mathbf{1}_{K-1}^T + \mathrm{diag}\left\{\sigma_2^2/w_2^*, \ldots, \sigma_K^2/w_K^*\right\} \end{aligned}$$

is uniquely determined by the strict concavity of $\ln\det(\cdot)$ as a function on the set $\{C = (A^T M^{-1}(\xi)A)^{-1} \mid \xi \in \Xi\}$, the vector $\boldsymbol{w}^*$ is unique. □

Proof of Theorem 4

By noting that $A^T M^{-1}(\xi_1^* \otimes \xi_2)A = A_1^T M_1^{-1}(\xi_1^*)A_1$ and $A_1^T M_1^{-1}(\xi^*_{(1)})A_1 = A^T M^{-1}(\xi^*)A$, the conclusion follows. □

Appendix B. Supplementary material

The R code based on the package rootSolve is available.

References

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

Atkinson, A. C. (2015). Optimum designs for two treatments with unequal variances in the presence of covariates. *Biometrika*, **102**, 494–499.

Donev, A. N. (1989). Design of experiments with both mixture and qualitative factors. *Journal of the Royal Statistical Society. Series B*, **50**, 297–302.

Goos, P. (2002). *The optimal design of blocked and split-plot experiments.* Springer, New York.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2007). Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor. *Journal of Statistical Planning and Inference*, **137**, 3882–3893.

Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications.* Springer, New York.

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

Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM, Philadelphia.

Rodríguez, C. and Ortiz, I. (2005). D-optimum designs in multi-factor models with heteroscedastic errors. *Journal of Statistical Planning and Inference*, **128**, 623–631.

Rosa, S. and Harman, R. (2015). Optimal approximate designs for estimating treatment contrasts resistant to nuisance effects. preprint arXiv: 1504.06079.

Schwabe, R. (1996). *Optimum designs for multi-factor models*. Springer, New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Sverdlov, O. and Rosenberger, W. F. (2013). On recent advances in optimal allocation designs in clinical trials. *Journal of Statistical Theory and Practice*, **7**, 753–773.

Wong, W. K. and Zhu, W. (2008). Optimum treatment allocation rules under a variance heterogeneity model. *Statistics in medicine*, **27**, 4581–4595.