

Optimal Design of Controlled Experiments for Personalized Decision Making in the Presence of Observational Covariates

Yezhuo Li¹, Qiong Zhang¹, Amin Khademi², and Boshi Yang¹

¹School of Mathematical and Statistical Sciences, Clemson University

²Department of Industrial Engineering, Clemson University

Abstract

Controlled experiments are widely applied in many areas such as clinical trials or user behavior studies in IT companies. Recently, it is popular to study the experimental design problem to facilitate personalized decision making. In this paper, we investigate the problem of optimal design of multiple treatment allocation for personalized decision making in the presence of observational covariates associated with experimental units (often, patients or users). We assume that the response of a subject assigned to a treatment follows a linear model which includes the interaction between covariates and treatments to facilitate precision decision making. We define the optimal objective as the maximum variance of estimated personalized treatment effects over different treatments and different covariates values. The optimal design can be obtained by minimizing this objective. Under a semi-definite program reformulation of the original optimization problem, we use a YALMIP and MOSEK based optimization solver to provide the optimal design. Numerical studies are provided to assess the quality of the optimal design.

Keywords: E-optimal Design; Multiple Treatments Allocation; Semi-definite Program.

1 Introduction

Optimal designs of experiments are developed to reduce the variance of estimated model parameters (Pukelsheim, 2006) by optimizing a function of the information matrix usually under a (generalized) linear model assumption. Examples include the determinant, the trace, and the minimum eigenvalue of the information matrix, which result in D-, A-, and E-optimal designs, respectively. In many areas, such as clinical trials or user behavior studies in IT companies, the experiments are conducted to investigate the treatment effects over different users by designing controlled experiments, i.e., divide the users into different groups and assigning each group with a treatment or control. For example, optimal design of experiments is applied to design of clinical trials in the era of precision medicine, where the patient response may depend on some biomarkers (covariates). In fact, there is a significant growing evidence that the response of the patients may depend on their covariates in a wide range of applications. For instance, Schork (2015) showed that the top ten highest-grossing drugs in the US are only effective in 4% to 25% of the patient population, a shocking result for the current clinical practice. Biswas and López-Fidalgo (2013) applied a modified D-optimal design to early stage clinical trials when the covariates are uncontrollable. Lee and Wason (2019) applied a weighted L-optimal design for late stage Phase III clinical trials with observed covariates. As another emerging application, Tech Giants such as Google, Facebook, and LinkedIn frequently use A/B testing for marketing, web design, and data-driven decision making (Siroker and Koomen, 2013). In fact, applying controlled experiments such as A/B testing experiments in business has improved the performance of companies (Koning et al., 2022). Also, there are some methodological contributions for optimal design of experiments applied for A/B testing. For example, Bhat et al. (2020) studied an offline and online standard A/B testing with the goal of maximizing the precision of least square estimations.

The literature of optimal design of experiments for linear models, which we also use in this study, is vast. From the optimal design perspective, Atkinson (2015) investigated the problem

with two treatments, and developed the D-optimal design for both covariates and treatment allocation under the assumption of a linear model with interactions between covariates and treatments. Wang and Ai (2016) generalized the work of Atkinson (2015) to experiments with multiple treatments. Following the same model setting, Rosa (2020) developed the A- and E-optimal designs for treatments and covariates.

In practice, the user covariates may not be controllable, due to the restriction in the user recruiting process. Thus, an alternative problem is to allocate the treatments to a given set of users with *observed* covariates. From the optimal design perspective, the aim is to obtain an exact design by optimizing a design objective with the treatment allocation as the design variable given fixed covariates, which often results in an optimization problem with binary decision variables. For two treatment allocation, under the linear model assumption without interaction between covariates and treatments, Hore et al. (2014) proposed computationally tractable algorithms to search D- and A- optimal designs. Also, with a particular interest of minimizing the variance of the estimated treatment effect (i.e., a D_s optimal design criterion), Bhat et al. (2020) demonstrated that the problem can be solved efficiently based on a generalization of the MAX-CUT semi-definite programming (SDP) relaxation of Goemans and Williamson (1995). Given the linear model with interaction between treatments and covariates, Zhang et al. (2022) proposed an optimal design objective for personalized decision making with two treatments and a corresponding approximation solution approach to obtain near optimal allocations efficiently.

In this paper, we generalize the work in Zhang et al. (2022) to multiple treatment allocation for personalized decision making with observational covariates. Specifically, we consider the situation that there are a finite number of treatment options and subjects with given covariates. We assume that the response of a subject assigned to a treatment follows a linear model which includes the interaction between covariates and treatments. Accordingly, we can define the best treatment option for each subject groups, which can be estimated with data. This setup facilitates precision decision making. However, by including more than

two treatments, the resulting design objective can not be simplified by the approximation approach developed for the two treatment cases in [Zhang et al. \(2022\)](#). Since the design objective matches the E-optimal design criterion, we reformulate the optimization problem as a SDP with binary decision variables. We use a YALMIP and MOSEK based optimization solver for SDP in a branch-and-bound scheme to provide the optimal design. In our numerical study, we assess the quality of the optimal designs provided by the solver.

2 Problem Description

A decision-making problem in practice can often be simplified to choose a treatment from a finite candidate set $\{1, \dots, K\}$ that has the best performance measure μ_k , i.e.,

$$k^* \in \operatorname{argmax}_{k \in \{1, \dots, K\}} \mu_k. \quad (1)$$

The values of μ_k 's are unknown, which can be estimated based on collected user responses to their treatment allocations. Particularly, given n users, the experimenter allocates one of the K treatments to each of the users. The response of the i -th user is denoted by y_i . Assume that y_i is a continuous scalar that follows a model given by

$$y_i = \sum_{k=1}^K x_{ik} \mu_k + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \quad (2)$$

where ε_i is an additive error term, and $x_{ik} \in \{0, 1\}$ with $x_{ik} = 1$ indicating that the k -th treatment is allocated to the i -th user. We require that $\sum_{k=1}^K x_{ik} = 1$ to guarantee that each user is only exposed to a single treatment. Under the model assumption in (2), the performance measure μ_k represents the average treatment performance over the target user population, which can be estimated by $\hat{\mu}_k$ given responses y_i 's and treatment allocation x_{ik} 's.

Then a data-driven solution to (1) is given by

$$\hat{k}^* \in \operatorname{argmax}_{k \in \{1, \dots, K\}} \hat{\mu}_k. \quad (3)$$

The accuracy of this decision is associated with the accuracy of the estimates $\hat{\mu}_k$'s. To reduce the uncertainty of the decision, we can reduce the variances of the estimates $\hat{\mu}_k$'s.

Under the context of personalized decision making, a user is associated with observed covariates $\mathbf{z} = (z_1, \dots, z_p)^\top \in \mathcal{Z} \subset \mathbb{R}^p$, where \mathcal{Z} is the covariates space of the target population. Examples of the covariates include the demographic information, social behavior and network connections. The treatment performance can often be related to the covariates values. Let $\mu_k(\mathbf{z})$ be a function of the users' covariates \mathbf{z} representing the personalized treatment performance measure. In this paper, we further assume that $\mu_k(\mathbf{z})$ is a linear function over \mathbf{z} . Therefore, the linear model in (2) is represented by

$$y_i = \sum_{k=1}^K x_{ik} \mu_k(\mathbf{z}_i) + \varepsilon_i = \sum_{k=1}^K x_{ik} \mathbf{z}_i^\top \boldsymbol{\beta}_k + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \quad (4)$$

where $\boldsymbol{\beta}_k = (\beta_{1k}, \dots, \beta_{pk})^\top$ is a vector of linear coefficients. The first feature in each \mathbf{z}_i is fixed to be one as the linear intercept. Similar to Wang and Ai (2016), this model incorporates the interaction between treatment and covariates, which enables the estimation of heterogeneity of treatment performance over covariates \mathbf{z} .

Under this model, the problem of personalized decision making becomes

$$k^*(\mathbf{z}) = \operatorname{argmax}_{k \in \{1, \dots, K\}} \mu_k(\mathbf{z}) = \operatorname{argmax}_{k \in \{1, \dots, K\}} \left\{ \mathbf{z}^\top \boldsymbol{\beta}_k \right\} \quad \text{for any } \mathbf{z} \in \mathcal{Z}. \quad (5)$$

With estimated parameters, a data-driven solution to (5) can be obtained for any $\mathbf{z} \in \mathcal{Z}$, i.e.,

$$\hat{k}^*(\mathbf{z}) = \operatorname{argmax}_{k=1, 2, \dots, K} \left\{ \mathbf{z}^\top \hat{\boldsymbol{\beta}}_k \right\} \quad (6)$$

Similar to (3), to improve the accuracy of personalized decisions, it is desired to reduce the

variance of estimated $\mu_k(\mathbf{z}) = \mathbf{z}^\top \boldsymbol{\beta}_k$ for each $k \in \{1, \dots, K\}$ and each $\mathbf{z} \in \mathcal{Z}$. Following the optimal design perspective in Zhang et al. (2022), we can reduce the the variances by searching an optimal allocation of the treatments x_{ik} 's in the presence of fixed covariates of recruited users as illustrated in Figure 1. Next, we investigate the design criterion that characterizes the variance of estimated $\mu_k(\mathbf{z}) = \mathbf{z}^\top \boldsymbol{\beta}_k$ for each $k \in \{1, \dots, K\}$ and any $\mathbf{z} \in \mathcal{Z}$.

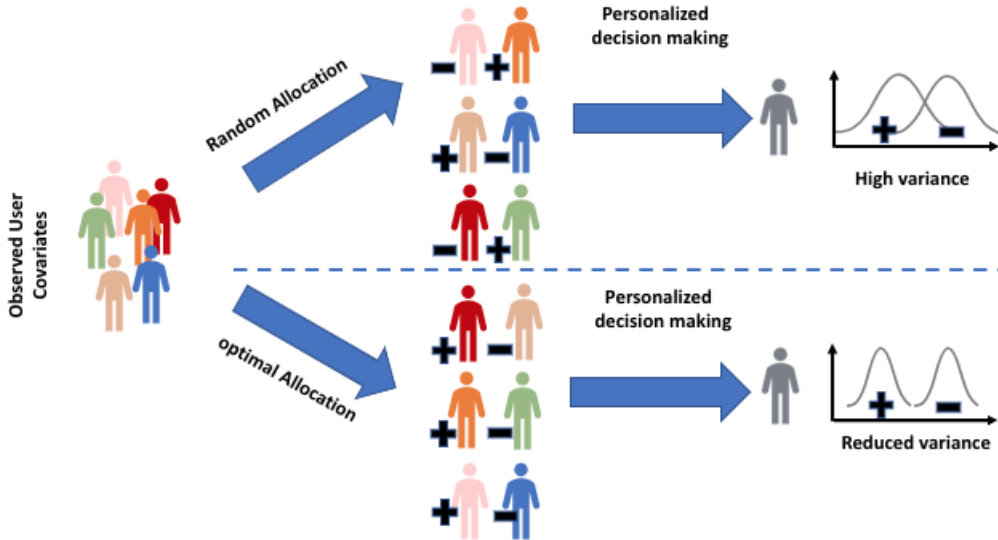


Figure 1: An illustration of the motivation of the proposed approach for two treatment allocation (i.e., + and - in the figure): We aim to obtain an optimal design of the treatment allocations to reduce the variances of estimated personalized treatment effects for users with any covariates \mathbf{z} . By reducing these variances, it is clearer how to make the decision accurately.

3 Design Criterion of Controlled Experiments for Personalized Decision Making

Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be the responses of n users. Correspondingly, let Z be an $n \times p$ matrix, whose rows are the users' covariates $\mathbf{z}_1^\top, \dots, \mathbf{z}_n^\top$ (the first entry is loaded by one as the intercept). The treatment allocation is recorded in an $n \times K$ matrix X , where the i -th row of X is denoted by $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{iK})^\top$ representing the treatment allocation of the i -th user. We assume that the responses are generated under the model in (4). The model

covariates matrix is then given by

$$\begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix}$$

which we assume has a full column rank. The notation \otimes denotes the Kronecker product. By stacking β_k 's in a vector $\mathbf{b} = (\beta_1^\top, \dots, \beta_K^\top)^\top$ of size Kp , we can express the least squares estimator of \mathbf{b} by

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} = \left(\begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix}^\top \mathbf{y}. \quad (7)$$

We assume that the additive error term ε_i 's in (4) are independent and identically distributed random variables with mean zero and variance σ^2 . The variance-covariance matrix of the estimated \mathbf{b} in (7) can be expressed by

$$\text{var}(\hat{\mathbf{b}}) = \sigma^2 \left(\begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix} \right)^{-1}.$$

Notice that

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{x}_1^\top \otimes \mathbf{z}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \otimes \mathbf{z}_n^\top \end{bmatrix} &= \sum_{i=1}^n (\mathbf{x}_i \otimes \mathbf{z}_i)(\mathbf{x}_i^\top \otimes \mathbf{z}_i^\top) = \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top) \otimes (\mathbf{z}_i \mathbf{z}_i^\top) \\ &= \begin{bmatrix} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{i1} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{iK} \end{bmatrix}, \end{aligned}$$

which is the result of that $\mathbf{x}_i \mathbf{x}_i^\top$ is a $K \times K$ diagonal matrix with diagonal entries x_{i1}, \dots, x_{iK} .

Then the variance-covariance matrix of $\hat{\mathbf{b}}$ can be simplified to

$$\text{var}(\hat{\mathbf{b}}) = \sigma^2 \text{diag} \left\{ \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{i1} \right)^{-1}, \dots, \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{iK} \right)^{-1} \right\}. \quad (8)$$

For any $k \in \{1, \dots, K\}$ and $\mathbf{z} \in \mathcal{Z}$, we have that

$$\text{var}(\mathbf{z}^\top \hat{\boldsymbol{\beta}}_k) = \mathbf{z}^\top \text{var}(\hat{\boldsymbol{\beta}}_k) \mathbf{z} = \sigma^2 \mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z}.$$

As noted earlier, we assume that the experiments are conducted with n existing users. Thus, the personalized information matrix Z is observed, and the experimental design problem is to determine the treatment allocation matrix X . A necessary condition for positive definiteness of the variance-covariance matrix requires that $\sum_{i=1}^n x_{ik} \geq p$ for any $k \in \{1, \dots, K\}$ and $n \geq Kp$.

Our goal is to minimize the variance of $\mathbf{z}^\top \hat{\boldsymbol{\beta}}_k$ for any k and \mathbf{z} to improve the accuracy of the data-driven decision in (6). Instead of minimizing all the variances simultaneously, we formulate the optimization problem by minimizing the worst case, i.e., the maximum variance of any $k \in \{1, \dots, K\}$ and $\mathbf{z} \in \mathcal{Z}$. Therefore, the design criterion is given by

$$\max_{\mathbf{z} \in \mathcal{Z}} \max_{k=1,2,\dots,K} \text{var}(\mathbf{z}^\top \hat{\boldsymbol{\beta}}_k) = \max_{\mathbf{z} \in \mathcal{Z}} \max_{k=1,2,\dots,K} \sigma^2 \mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z}. \quad (9)$$

The optimal design is then given by solving the problem

$$\min_X \max_{\mathbf{z} \in \mathcal{Z}} \max_{k=1,2,\dots,K} \mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z} \quad (10a)$$

$$\text{s.t.} \quad \sum_{i=1}^n x_{ik} \geq p, \quad k = 1, \dots, K, \quad (10b)$$

$$\sum_{k=1}^K x_{ik} = 1, \quad i = 1, 2, \dots, n, \quad (10c)$$

$$X \in \{0, 1\}^{n \times K}, \quad (10d)$$

where $\sum_{i=1}^n x_{ik} \geq p$ is required for the positive definiteness of the matrix $\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik}$, and the constraint $\sum_{k=1}^K x_{ik} = 1$ ensures that each user is exactly assigned to one treatment.

Although the optimal design criterion has a closed-form expression, the corresponding optimization problem is still a challenge to solve directly as a mixed-integer minimax problem.

We assume that the covariates space \mathcal{Z} is given by a unit ball, i.e., $\mathcal{Z} = \{\mathbf{z} \in \mathbb{R}^p \mid \|\mathbf{z}\| \leq 1\}$.

For any X that gives a positive definite variance-covariance matrix, we have that

$$\begin{aligned} & \max_{\|\mathbf{z}\| \leq 1} \max_{k=1,2,\dots,K} \left[\mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z} \right] = \max_{k=1,2,\dots,K} \max_{\|\mathbf{z}\| \leq 1} \left[\mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z} \right] \\ &= \max_{k=1,2,\dots,K} \max_{\|\mathbf{z}\|=1} \left[\mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \mathbf{z} \right] = \max_{k=1,2,\dots,K} \lambda_{\max} \left[\left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right)^{-1} \right] \\ &= \max_{k=1,2,\dots,K} \left\{ 1 / \lambda_{\min} \left[\left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) \right] \right\} = \max_{k=1,2,\dots,K} \left\{ 1 / \min_{\|\mathbf{z}\|=1} \left[\mathbf{z}^\top \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) \mathbf{z} \right] \right\}, \end{aligned}$$

where λ_{\max} and λ_{\min} produce the maximum and minimum eigenvalues of a matrix, respectively. Accordingly, the optimization problem in (10) is equivalent to

$$\max_X \min_{k=1,2,\dots,K} \lambda_{\min} \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) \quad (11a)$$

$$\text{s.t.} \quad \sum_{i=1}^n x_{ik} \geq p, \quad k = 1, \dots, K, \quad (11b)$$

$$\sum_{k=1}^K x_{ik} = 1, \quad i = 1, 2, \dots, n, \quad (11c)$$

$$X \in \{0, 1\}^{n \times K}. \quad (11d)$$

This design objective is known as the E-optimal design in the literature (e.g., [Pukelsheim \(2006\)](#) and [Dette and Studden \(1993\)](#)), i.e.,

$$\min_{k=1,2,\dots,K} \lambda_{\min} \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) = \lambda_{\min} \left(\begin{bmatrix} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{i1} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{iK} \end{bmatrix} \right).$$

Next, we provide a SDP-based solution approach to solve (11) with observed covariates Z and a binary decision X .

4 A SDP-based Solution Approach

Following the reformulation technique for E-optimal design in [Boyd and Vandenberghe \(2004\)](#), we cast problem (11) as a SDP with binary decision variables. The reformulation is based on the following observation. Let A be a real $p \times p$ symmetric matrix. It is well known in matrix analysis that for any $\delta \in \mathbb{R}$, $\lambda_{\min}(A - \delta I) = \lambda_{\min}(A) - \delta$, where I is the $p \times p$ identity matrix. Therefore, $\lambda_{\min}(A)$ can be expressed as the largest value of δ such that $\lambda_{\min}(A - \delta I) \geq 0$, i.e.,

$$\lambda_{\min}(A) = \max \{ \delta \mid A - \delta I \succeq 0 \},$$

where $A - \delta I \succeq 0$ means that $A - \delta I$ is a positive semi-definite matrix.

We apply the above observation to problem (11). For $k = 1, \dots, K$,

$$\lambda_{\min} \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) = \max \left\{ \delta_k \mid \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} - \delta_k I \succeq 0 \right\}.$$

As a result,

$$\begin{aligned} \min_{k=1, \dots, K} \lambda_{\min} \left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} \right) &= \min_{k=1, \dots, K} \max \left\{ \delta_k \mid \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} - \delta_k I \succeq 0 \right\} \\ &= \max \left\{ \lambda \mid \lambda \leq \delta_k, \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} - \delta_k I \succeq 0, k = 1, \dots, K \right\} \\ &= \max \left\{ \lambda \mid \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} - \lambda I \succeq 0, k = 1, \dots, K \right\}. \end{aligned}$$

Consequently, problem (11) can be equivalently written as

$$\max_{X, \lambda} \quad \lambda \tag{12a}$$

$$\text{s.t.} \quad \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top x_{ik} - \lambda I \succeq 0, \quad k = 1, 2, \dots, K, \tag{12b}$$

$$\sum_{i=1}^n x_{ik} \geq p, \quad k = 1, \dots, K, \tag{12c}$$

$$\sum_{k=1}^K x_{ik} = 1, \quad i = 1, 2, \dots, n, \tag{12d}$$

$$X \in \{0, 1\}^{n \times K}. \tag{12e}$$

Note that, the reformulation technique is the same as used in the SDP reformulation proposed for the approximate E-optimal design in Section 7.5 of [Boyd and Vandenberghe \(2004\)](#). However, the discrete feature in problem (11) cannot be averted by simply considering a continuous relaxation as in [Boyd and Vandenberghe \(2004\)](#), even when n is large. Thus, the implementation of a SDP problem with continuous decisions cannot solve our problem. As a mixed binary SDP, the reformulation (12) can be solved by a branch-and-bound (BnB) algorithm, where each relaxation problem in a BnB node is solved by an SDP solver ([Gally et al., 2018](#)). A built-in solver in YALMIP ([Löfberg, 2004](#)) provides such a BnB implementation for MATLAB ([MATLAB, 2021](#)), and MOSEK ([ApS, 2019](#)) can be used as the embedded SDP solver.

In practice, the experimenter often requires a (roughly) balanced allocation across the K treatments. If K can be divided by n , we can include balanced constraints

$$\sum_{i=1}^n x_{ik} = n/K, \quad k = 1, \dots, K, \tag{13}$$

which also potentially accelerates the computation time of solving this problem by reducing the number of feasible solutions. If K can not be divided by n , a set of roughly balanced constraints can be added similarly. For convenience, we set n as a multiple of K and include

the balanced constraints in our numerical study.

5 Numerical Study

In this section, we illustrate the numerical performance of the proposed solution approach based on synthetic datasets. Our aim is to demonstrate the quality of the resulting optimal design under different situations.

Implementation of the proposed solution approach. We solve the optimization problem (12) with the YALMIP-MOSEK implementation in MATLAB. In our implementation, we include the balanced constraint in (13). For simplicity, we set the number of users n as a multiple of the number of treatments K . To make the computation tractable, we set the maximum time to solve the optimization problem to 300 seconds. Therefore, the resulting design is not the exact optimal design. The matlab code of this optimization model is provided in the Appendix.

Quality measures of optimal designs. We propose two quality measures to compare the optimal design with random designs. We generate 1000 random realizations of X that satisfy the constraints in the optimization model (12) as well as the balanced constraints in (13). We compute the objective values as in (11) for the 1000 random realizations of X . Given the objective values $\lambda^1, \dots, \lambda^{1000}$ of random designs, we provide two quality measures: the percentile among objectives of random designs and the relative improvement with respect to the median objective of random designs. Let λ_{opt} be the objective value of the optimal design and λ_{med} be the median of $\lambda^1, \dots, \lambda^{1000}$. The ‘‘Percentile’’ within random designs is defined by

$$P = \frac{1}{1000} \sum_{i=1}^{1000} I\{\lambda_{opt} \geq \lambda^i\} \cdot 100\%, \quad (14)$$

where $I\{\cdot\}$ denotes the indicator function, and the “Relative Improvement (RI)” is defined by

$$\text{RI} = \frac{\lambda_{\text{opt}} - \lambda_{\text{med}}}{\lambda_{\text{med}}} \cdot 100\%. \quad (15)$$

The two quality measures compare the objective given by the proposed solution approach with the population and the median of the objectives given by random designs. The higher values of “Percentile” and “RI” indicate better quality of the optimal design, i.e., the value of adopting the proposed optimization approach. Next, we investigate the performance of the optimal design under two common examples of fixed covariates Z .

Example I. In this example, we generate the entries of the covariates Z independently from the standard normal distribution. We consider a batch of cases with all the combinations of $n \in \{120, 240, 360\}$, $p \in \{3, 5\}$ and $K \in \{2, 3, 4\}$. For each combination of n , p and K , we generate a set of covariates $\mathbf{z}_1, \dots, \mathbf{z}_n$, compute the quality measures Percentile and RI, and replicate this process for 100 times. The resulting 100 copies of Percentile and RI are depicted as boxplots in Figures 2 and 3, respectively. The results give the following observations:

1. The values of Percentile and RI increase with K . For $K = 2$, the proposed solution approach does not always give better design than the random designs.
2. The values of Percentile increase with n , whereas the values of RI decrease with n . For $n = 120$, the objective values given by the proposed approach may not outperform the random designs that give median or higher quantiles of objectives.
3. The values of Percentile and RI increase slightly as we increase p from 3 to 5.

These observations indicate that the cases with $K = 3$ or 4 or $n = 240$ or 360 provide solutions of higher quality than the case with $K = 2$ and $n = 120$, which is possibly due to that the difficulty in solving the problem caused by the binary constraints is more significant when K and n are small.

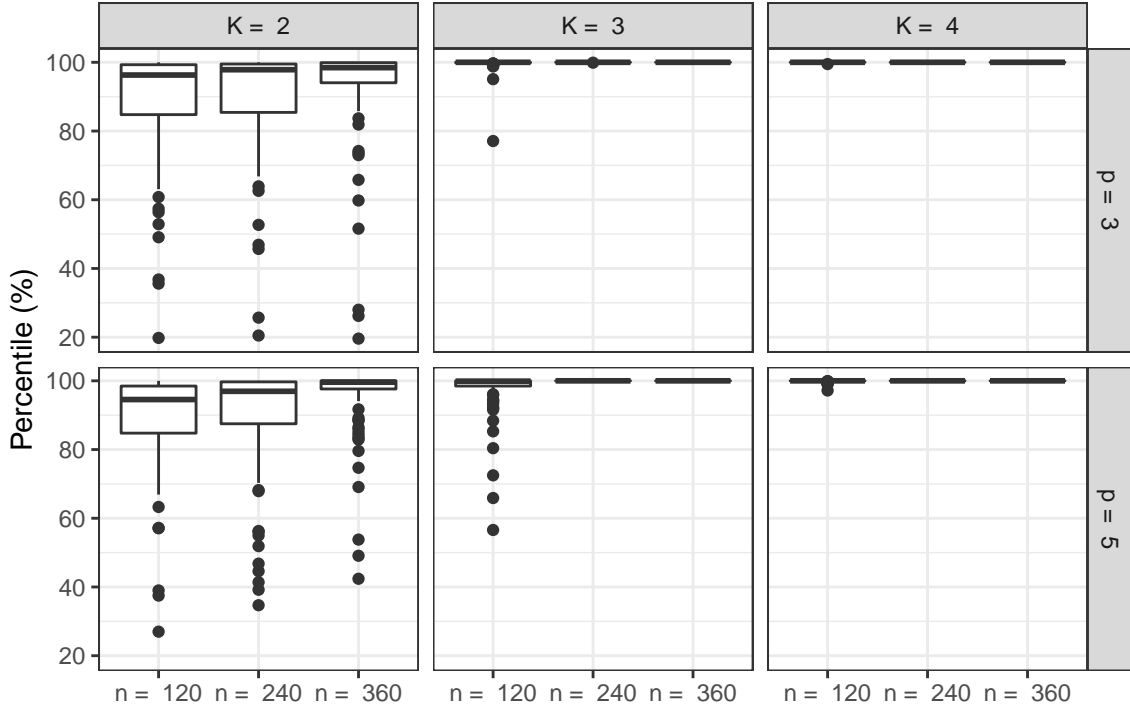


Figure 2: The quality measure Percentile (given by (14)) of different cases in Example I, i.e., Percentile of the optimal objective λ_{opt} in the population of objectives given by 1000 random designs.

Example II. In this example, we generate the entries of the covariates Z independently from a 0-1 Bernoulli distribution with the proportion of ones equal to r . The aim of this example is to demonstrate the impact of imbalance in covariates to the quality of optimal design. Therefore, we fix $n = 200$, $K = 4$ and $p = 3$, and consider $r \in \{0.1, 0.5\}$, where $r = 0.5$ gives balanced covariates and 0.1 gives imbalanced covariates. The values of Percentile and RI are depicted as boxplots in Figure 4. The results show that, for balanced covariates, the optimal design given by the proposed solution approach can often reach to the optimal value, but the values of RI are often around 20%. However, for imbalanced covariates, the relative improvement can often reach 50%-200%, but the Percentile may stand as low as 75% for some instances. This example demonstrates that the optimization problem under balanced covariates provides higher-quality solutions than that under imbalanced covariates, whereas, the benefit of optimal design under balanced covariates may not be as significant as that under imbalanced covariates.

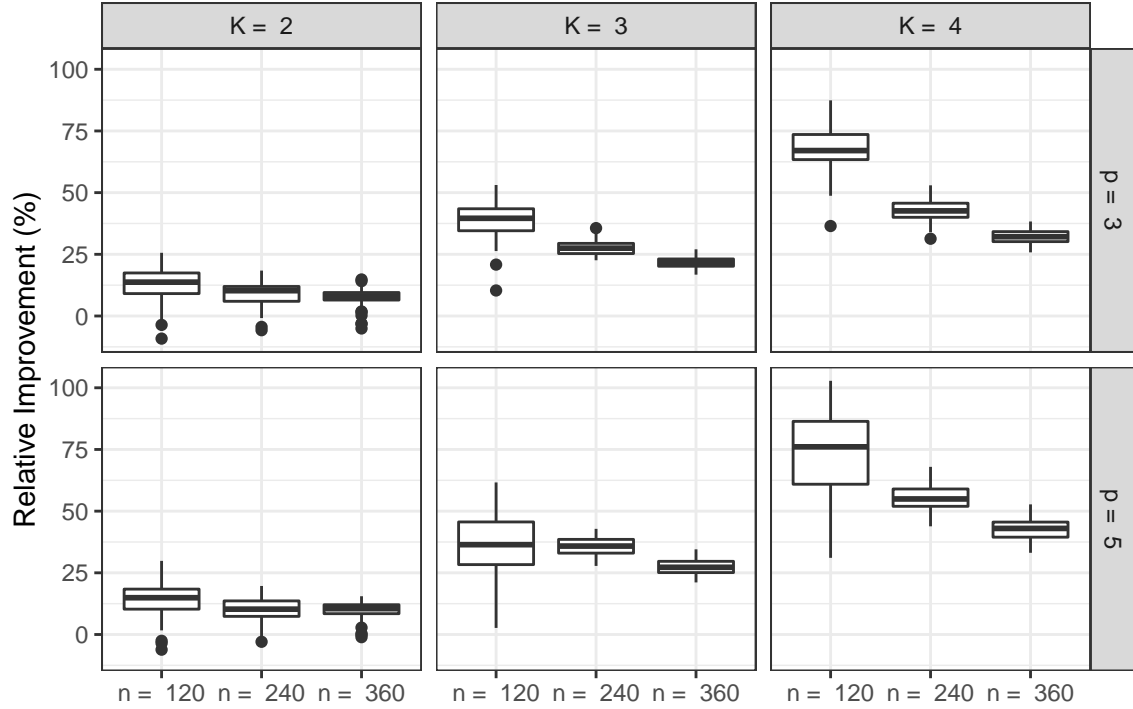


Figure 3: The quality measure RI (given by (15)) of different cases in Example I, i.e., Relative Improvement of the optimal objective λ_{opt} with respect to the median of objectives given by 1000 random designs.

After 300 seconds, the solver outputs the optimality gap (0%-100%), with a smaller optimality gap indicating that the output solution is closer to the true optimal solution. For both examples, the average optimality gaps of solving for optimal design is reduced to 10% or lower after 300 seconds. This shows that, for the sizes of the above examples, the proposed approach is appropriate to solve and returns a solution that is close to the exact optimal design. We also point out that for cases with K over 10 and/or n over 400, the proposed solution approach did not return a feasible solution within 300 seconds. For those cases, a heuristic and/or randomized solution approach can be a reasonable choice.

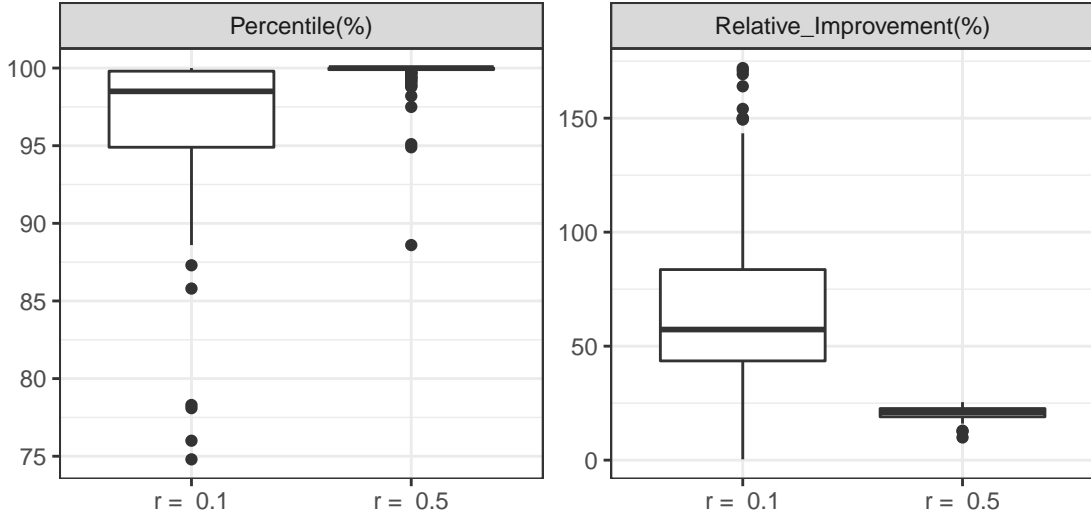


Figure 4: The quality measures Percentile (given by (14)) and RI (given by (15)) with balanced ($r = 0.5$) and imbalanced ($r = 0.1$) covariates in Example II.

6 Conclusion

To improve the accuracy of personalized decision, this paper investigates the optimal design to minimize the maximum variance of estimated personalized treatment effects over different treatments and different covariates values. The resulting design objective matches the E-optimal design criterion. To provide the optimal design of multiple treatment allocation in the presence of observed covariates, a SDP solution approach is applied to solve the optimization problem. The proposed solution approach is able to provide optimal designs efficiently for n from 100-300 as shown in our numerical results. We point out potential future directions. First, it is desired to develop more efficient reformulations or optimization algorithms to handle the proposed design objective with larger sizes. Second, in practice, it is useful to investigate online optimal allocation assuming that the users are recruited in an online dynamic fashion under the proposed design objective.

References

- ApS, M. (2019), *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*
- Atkinson, A. (2015), “Optimum designs for two treatments with unequal variances in the presence of covariates,” *Biometrika*, 102, 494–499.
- Bhat, N., Farias, V. F., Moallemi, C. C., and Sinha, D. (2020), “Near-Optimal A-B Testing,” *Management Science*, 66, 4477–4495.
- Biswas, A. and López-Fidalgo, J. (2013), “Compound designs for dose-finding in the presence of nondesignable covariates,” *Pharmaceutical Statistics*, 12, 92–101.
- Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge university press.
- Dette, H. and Studden, W. J. (1993), “Geometry of E-optimality,” *The Annals of Statistics*, 21, 416–433.
- Gally, T., Pfetsch, M. E., and Ulbrich, S. (2018), “A framework for solving mixed-integer semidefinite programs,” *Optimization Methods and Software*, 33, 594–632.
- Goemans, M. X. and Williamson, D. P. (1995), “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,” *Journal of the ACM (JACM)*, 42, 1115–1145.
- Hore, S., Dewanji, A., and Chatterjee, A. (2014), “Design issues related to allocation of experimental units with known covariates into two treatment groups,” *Journal of Statistical Planning and Inference*, 155, 117–126.
- Koning, R., Hasan, S., and Chatterji, A. (2022), “Experimentation and start-up performance: Evidence from A/B testing,” *Management Science*.
- Lee, K. M. and Wason, J. (2019), “Design of experiments for a confirmatory trial of precision medicine,” *Journal of Statistical Planning and Inference*, 199, 179–187.

- Löfberg, J. (2004), “YALMIP: A Toolbox for Modeling and Optimization in MATLAB,” in *In Proceedings of the CACSD Conference*, Taipei, Taiwan.
- MATLAB (2021), *9.10.0.1710957 (R2021a) Update 4*, Natick, Massachusetts: The Math-Works Inc.
- Pukelsheim, F. (2006), *Optimal design of experiments*, SIAM.
- Rosa, S. (2020), “Optimal experimental designs for treatment contrasts in heteroscedastic models with covariates,” *Journal of Statistical Planning and Inference*, 209, 280–291.
- Schork, N. J. (2015), “Personalized medicine: time for one-person trials,” *Nature News*, 520, 609.
- Siroker, D. and Koomen, P. (2013), *A/B testing: The most powerful way to turn clicks into customers*, John Wiley & Sons.
- Wang, Y. and Ai, M. (2016), “Optimal designs for multiple treatments with unequal variances,” *Journal of Statistical Planning and Inference*, 171, 175–183.
- Zhang, Q., Khademi, A., and Song, Y. (2022), “Min-max optimal design of two-armed trials with side information,” *INFORMS Journal on Computing*, 34, 165–182.

Appendix

We provide the Matlab code of the YALMIP-MOSEK implementation to obtain the optimal design.

Optimization Function

With a given covariates matrix Z and the number of treatments K , the optimization function below gives the design matrix X and optimal objective value opt in optimal design model in(at) 300s.

The optimization function

```
function [x,opt]= optdesign300(z,k)

[n,p]=size(z); %%%the size of matrix z.
zopt=[(ones(1,n))' z]; %%%add an all 1 column vector to z.

for i=1:n
    zz{i}= (zopt(i,:))'*zopt(i,:);
    %%%ith matrix zz = (ith row of zopt)'*(ith row of zopt).
end

x=binvar(n,k); %%%define x as an n*k binary matrix.
d=sdpvar(1); %%%define d as a symbolic decision variable.

%%%setup constraints of optimaization function here:
J=cell(1,k);
```

```

for j=1:k
    H=0;
    for i=1:n
        H=H+(zz{i}*x(i,j));
        %%%with a given j, H = the sum of [(ith matrix zz) * (xi)],
        %where i is from 1 to n.
    end
    J{j}=H-d*eye(p+1); %%%with a given j, J{j}= H - d*(identity matrix).
end

constraint=[J{1}>=0];

for i=2:k
    constraint=[constraint,J{i}>=0];
end

%%%constraints: J{i}>=0, for i from 1 to k.

for i=1:n
    constraint=[constraint,sum(x(i,:))=1];
end

%%%add the constraints: the sum of each row of x equals to 1.

for i=1:k
    constraint=[constraint,sum(x(:,i))=n/k];
end

%%%add the constraints: the sum of each column of x equals to n/k.

options=sdpsettings('solver','bnb','bnb.maxtime',300,'savesolveroutput',1);

```

```

%%%set the max running time that maxtime=300s.

diagnostics = optimize(constraint,-d,options);
%%%function for solving optimization problems with defined constraints,
%%%objective and options.

x=value(x); %%% x matrix
opt=value(d); %%% optimal value

end

```

Example

In order to generate a different sequence of random numbers each time, 'shuffle' is used to initialize generator. Create a 10-by-4 matrix of normal random numbers from the normal distribution with mean 0 and standard deviation 1. Then set the number of treatments $k = 2$, and find solution and optimal value in optimal design model by using optimization function.

```

rng shuffle
z=normrnd(0,1,[10, 4]);
k=2;
[x,opt]=optdesign300(z,k)

```